POLITECNICO DI MILANO
Corso di Laurea in Ingegneria Informatica
Dipartimento di Elettronica e Informazione
AIRLab - Laboratorio di Intelligenza Artificiale e Robotica

User fatigue analysis from keyboard and mouse usage

Relatore:
Prof. Andrea Bonarini
Correlatori:
Ing. Simone Tognetti
Ing. Maurizio Garbarino

Tesi di Laurea di:
Goran Burek, matricola 707497

Anno Accademico 2008-2009

Contents

Chapter 1

Introduction

"Zoon politikon" was the written on the blackboard on the first philosophy lesson I had in high school. The teacher wrote Aristotle's famous definition of man as a political being. We had to remember well that phrase and understand its meaning. It still stands what great Aristotle said, but more importantly human is an emotional being. The emotional being lies under the political one and in fact it predicates the political being. This is the starting point of this thesis to show and explain how deeply the emotional being runs in humans, and the affect the emotion has on all aspects of human life.

Artificial intelligence is a branch of computer science usually defined as a study trying to create an intelligent agent, that is capable of perceiving it's environment and taking actions in order to maximize it's goal (what ever it may be). In other words the main objective of artificial intelligence is to design the agent to act in every situation in the most effective and most desirable way.

Affective computing is a branch of artificial intelligence. It's occupation is to design computer systems that are capable of recognizing, interpreting and processing human emotions.

It finds its beginning in 1995 when Rosalind Picard's paper Affective computing came out. It suggested that in order to develop real artificial intelligence the machine has to be able to recognize human emotions, to be more empathic. In human-computer interaction we may perceive computer as an agent and human emotions as agent's environment. So the main goal of affective computing is to make machine more responsive to user's emotions and changes in emotions. Doing so the interaction could be more natural like the way two humans interact.

In this work it will be presented the research we done, trying to explore the possibility of decoding user emotional state, in this case fatigue, from the behavior he/she demonstrates using the mouse and the keyboard. In chapter 2 is described the state of the art in emotion recognition, diverse techniques that produced promising results. In chapter 3 it will be explained the motivation of this research. After the introduction of the main hypothesis it will be explained in detail the experiments we proposed to confirm our theory. Chapter 4 brings the detailed descriptions of data collected executing experiments and interesting features. The chapter finishes demonstrating results of preliminary analysis.

Chapter 2

State of the art

Emotions and mood take major part in human life. They are omnipresent, during the whole lifetime person switches from one mood to an other and from an emotional state to an other. Emotions and moods are both feeling states, differentiated by time and strength. Moods are general feeling states that last for long time from several hours up to several days, and aren't so expressive. That's why when asked people usually answer "I'm OK" or "I'm a bit down" but cannot exactly verbalize their state or be more specific on what are they feeling in that moment. Emotions last for much shorter period of time but are stronger and more expressive. Emotion is triggered by noticeable events and are immediate reactions to these [36].

Since emotions and moods are both feeling states, for the rest of this thesis they will be united and called emotion. The mood will be represented as a neutral emotional state. The neutral state is the state where person feels very mild and tends to stay in that state for a long time given that no external events are trying to change it. In neutral states person doesn't show any facial expressions that are specific for other emotional states, has bio-signals on normal level (normal heart rate, no sweating, normal skin conductance level etc.)

Psychologists agree that humans have a group of basic emotions, and that all other emotions that we may experience or perceive are the combination of the basic ones. However they have not yet agreed on the number of basic emotions or what exactly they are, but more or less these six can be found on most of the lists: neutral, happy, surprised, angry, sad, disgusted and afraid [12, 37, 33].

Recognizable features of the emotions are arousal and the valence. Arousal is the criterion that distinguishes positive emotions from negative ones, and valence is the level that shows whether the emotion is strong or weak. Those

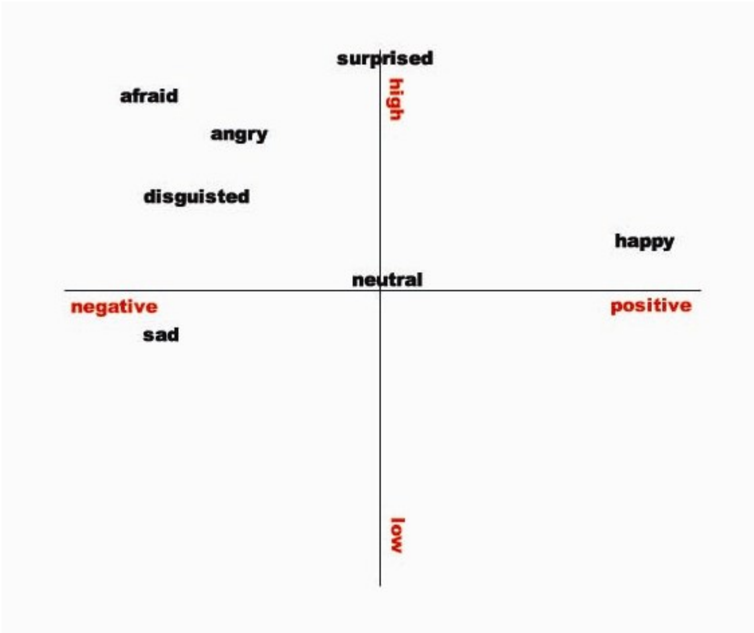

Figure 2.1: Emotions chart

two features can be represented on a two dimensional chart and used to map human emotions.

In this chart one dimension represents the valence and the other dimension represents the arousal level. Figure 2.1 shows the emotion chart with basic emotions. Due to insufficient knowledge of human emotions the values of arousal and valence are arbitrary, it just shows the conceptual mapping.

In the graph it is clearly visible that neutral state takes the center of the graph, while other states surrounds it. When an emotion is further from the center of the graph it becomes more strong and evident, but while in the neutral zone a person is not able to verbalize it's state, furthermore he is still not conscious toward what extreme his state is gravitating.

Human emotions in a way follow Newton's third law of motion: for every action there is a reaction. On every day basis everything a man does is predicated by his emotions: the way he perceives his environment, the way he chooses to act, to say, to think. There are so many ways that emotions influence every aspect of our lives, so let's take a look on a brief overview.

Let's take for example a man playing checks. It is an interesting example because it can demonstrate the variety of aspects how emotions impact on humans. To be able to play a good game of checks, a person has to learn the rules, the basic openings and offensive and defensive strategies. Since the rules are not so straightforward, a person can't learn them if he is tired, stressed out or agitated, because he finds it hard to concentrate. He has to be in proper mood to be able to learn the new concepts.

Then after learning phase, a person is prepared to play a good match. However, the knowledge itself is not enough to win a game; beside the experience, emotional state influences the outcome of the game. If the player is troubled or upset he might overlook an important move, he might not see a threatening figure and leave his own figure exposed.

After many matches played people learn some tricks and can remember well played situations, so that if on the board a similar situation occurs he can simply replay his moves and triumph over his opponent. Yet again the brain can play tricks on him, depriving him of recognizing the situation on board and not allow him to remember what the next prudent move is.

This example with checks served just to give a general overview, but every person had found himself in similar situations. Students often have difficulties while learning subjects that they don't like or don't enjoy, when looking in a hurry people find it hard to find a thing lying in front of their nose, or when agitated people do things unintentionally later excusing themselves saying "I'm sorry for what I have said/done I was angry and wasn't myself at the time". This just shows how emotions are important in our lives and their influence on our perception, memory, cognitive processing and behavior.

In the next sections there will be more explanations about the way emotions influence the four aforementioned processes.

2.1 The role of emotions

2.1.1 Perception

One of the central assumptions of the constructivist approach to perception is that perception is not determined entirely by external stimuli. As a consequence, it is assumed that emotional and motivational states, together with expectation and culture, may influence people's perceptual hypotheses and thus their visual perception. This notion that perception is influenced by various factors is often referred to as perceptual set. This is "a perceptual bias or predisposition or readiness to perceive particular features of a stimulus" [2]. Basically, it is the tendency to perceive or notice some aspects of the available sense data and ignore others.

Many researchers suggest that our emotional state will affect the way that we perceive. For example there is a term "perceptual defense" which refers to the effects of emotion on perception. Findings from a number of experiments show that subliminally-perceived words which evoke unpleasant emotions take longer to perceive at a conscious level than neutral words. It is almost as if our perceptual system is defending us against being upset or offended and it does this by not perceiving something as quickly as it should. McGinnies [24] investigated perceptual defense by presenting subjects with eleven emotionally neutral words (such as “apple”, “broom” and “glass”) and seven emotionally arousing, critical words (such as “whore”, “penis”, “rape”). Subjects were attached to a microammeter to measure their galvanic skin response. Each word was presented for increasingly long durations until it was named. There was a significantly higher recognition threshold for critical words. On 16 subjects the mean difference between two groups of words was 1.98 microampers with t-value of 5.10 and confidence of .01. And the time difference of the recognition was 0.045 seconds with t-value of 3.96 and confidence level of .01. It confirmed that perceptual defence was in operation and that it was causing alterations in perception.

Consistent with other studies, the study by Bhavin and Pham [34] finds that the emotional content of an image is an important factor guiding the transition from unconscious to conscious perception. That is to say, in addition to differences in formal stimulus measures such as luminance, contrast, and complexity, differences in the emotional content of rivaling images also influence which image ends up dominating perception. Specifically, the arousal level of an image determines the level of its access to visual awareness; the more arousing image of an iso-valence image pair dominates conscious perception; for arousing images duration of perceptual dominance was 22.93 ± 1.14 s, against non-arousing images whose duration was 18.24 ± 0.86 s. Arousal also interacts with valence (affect): For images of identical, low arousal level, the more pleasant image of the iso-arousal pair dominates perception (20.57 ± 1.63 s opposed to 16.97 ± 1.63 s), whereas for images of high arousal value, the less pleasant image of the iso-arousal pair dominates perception(18.63 ± 1.57 s opposed to 16.52 ± 1.39 s). The effects were remarkably consistent across observer (e.g. the more arousing image of the iso-valence pair dominated the conscious perception on 11 out of 12 observers).

2.1.2 Memory

First of all, there are two, quite different, elements to the question how does our memory work. The first concerns the emotional content of the information man wants to remember. The second concerns the effect of one’s emotional state on learning and remembering. It does seem clear that, as a general rule, people remember emotionally charged events better than boring ones. One of the latest researches [25] suggests that it is the emotional arousal, not the personal significance of the event, that makes such events easier to remember. The memory of strongly emotional images and events may be at the expense of other information. Thus, it may be less likely to forget information if it is followed by something that is strongly emotional. This effect appears to be stronger for women. It does seem that memories are treated differently depending on whether they are associated with pleasant emotions or unpleasant ones, and that this general rule appears to be affected by age and other individual factors. Specifically, pleasant emotions appear to fade more slowly from our memory than unpleasant emotions, but among those with mild depression, unpleasant and pleasant emotions tend to fade evenly, while older adults seem to regulate their emotions better than younger people, and may encode less information that is negative. An investigation of autobiographical memories [4] found that positive memories contained more sensorial and contextual details than neutral or negative memories (which didn't significantly differ from each other in this regard). This was true regardless of individual's personal coping styles.

A research focusing on the mood [25] showed that the emotional state at the time of encoding or retrieving plays important role in remembering. It is clear that mood affects what is noticed and encoded. This is reflected in two (similar but subtly different) effects:

- Mood congruence: whereby we remember events that match our current mood (thus, when we're depressed, we tend to remember negative events)
- Mood dependence: which refers to the fact that remembering is easier when your mood at retrieval matches your mood at encoding (thus, your chances of remembering an event or fact are greater if you evoke the emotional state you were in at the time of experiencing the event or learning the fact).

The complete explanation of how does our memory work has not yet been reached, but there appears to be two main aspects to answering what mechanisms are relevant for the memory processes. One is that stress hormones, such as cortisol, interact with the amygdala. The other is that the amygdala (an almond-sized structure deep in the medial temporal lobe) can alter the activity of other brain regions. One of the ways in which it does this is by acting on consolidation processes (principally in the hippocampus). It is perhaps this effect on consolidation that is reflected in a study using facial stimuli (involving inversion of eyes and mouth to change the emotional impact of a face without significantly changing its visual features), that indicated that the emotional load of a stimulus does not in fact affect the way we perceive it but does have an effect on how we become used to it if we see it many times. Notwithstanding this study, however, it does seem clear that, in some circumstances and for some types of stimuli, at least, the emotional attributes of a stimulus do affect the way we perceive it and process it - that is, the encoding of the memory. One of the ways in which it might do this is through the involvement of different brain regions depending on the nature of the emotion experienced. A recent imaging study [25] found that positive emotional contexts evoked activity in the right fusiform gyrus (among other regions), and negative emotional contexts evoked activity in the right amygdala. Another way in which emotions might affect memory encoding is through working memory. It has been suggested that, in the case of anxiety, part of working memory may be taken up with our awareness of fears and worries, leaving less capacity available for processing. In support of this theory, one study found that math-anxious people have working memory problems as they do math.

The common experience (typically with emotions less intense) confirms that emotional events tend to be well remembered, and extensive scientific evidence confirms anecdotal observations that emotional arousal can strengthen memory. In the past decade, neuroscience has witnessed a compelling convergence of evidence from human and animal subject studies regarding the neurobiology of emotion-enhanced memory. Key among the neurobiological players so far identified are endogenous stress hormones (in particular, the adrenergic hormones epinephrine and norepinephrine) and the amygdala. The evidence suggests that endogenous stress hormones, released by emotionally arousing events, feed back directly or indirectly to the amygdala to amplify long-term memory storage of the events that induced their release [5, 28].

When it comes to memory, however, emotion is a double-edged sword. It may enhance or impair memory, depending on many factors. Even different aspects of a single emotional experience may be retained very well by the memory, or seemingly lost from it entirely. The amnestic potential of emotional arousal has received less attention from neurobiologists than has its memory-enhancing potential [6].

Strange et al. [14] began by establishing an experimental procedure producing both memory-enhancing and memory-impairing effects reasonably attributable to emotional arousal. Subjects viewed brief lists of nouns presented for 1 s every 3 s. The lists were composed mostly of semantically related, emotionally neutral words. "Emotional oddballs" (aversive words such as "scream", "murder" and "morgue") determined by independent judges to be significantly more emotional than the neutral words were interspersed within the lists. After a short delay, subjects freely recalled as many words as they could from the just-seen list, before continuing with the next list. Results from the lists were pooled. Subjects recalled the emotional words ("E" words) significantly better than they did the neutral words. Furthermore, subjects recalled the words presented immediately before the E words (neutral "E-1" words) significantly worse than they recalled the other neutral words. Control procedures substituting "perceptual oddballs" (words in the same semantic category, but in another font) or "semantic oddballs" (words independently judged to be semantically unrelated to the other words in the list) for the emotional oddballs provided reasonable evidence that both the enhancing effect for the E words and the impairing effect for the E-1 words resulted from the emotional character of the E words. Thus, the experimental procedure establishes both an emotion-related memory enhancement (for the E words) and an emotion-related retrograde amnesia of a few seconds' duration.

It is the strong parallel to prior studies of emotionally influenced, long-term memory that is perhaps most surprising about the new results. In addition to involving relatively short-term memory processes, the experiments of Strange et al. involved far less-arousing stimuli than are typically used in studies of amygdala and adrenergic mechanisms in long-term memory in humans. Also, unlike these prior studies, they involved encoding instructions to the subjects and used procedures making subjects aware throughout that their memory would be tested. Each of these important variables could well influence the outcomes. Thus, it is not at all necessary that propranolol administration and amygdala lesions should produce such similar effects in this study as in the previous studies of their role in long-term, emotionally influenced memory. But they do, and this simple fact should force us to carefully reexamine our ideas about how exactly β-adrenergic and amygdala-based mechanisms participate in memory for emotionally arousing events.

Emotion's influence on memory is highly multifaceted, no doubt with correspondingly intricate neural underpinnings. By focusing on emotion's effects on memory in the short term, and on its ability to impair memory, Strange et al. have with a single shot widened the field's current focus considerably. In so doing, they have moved us closer to understanding both the beneficial and harmful effects of emotion on memory.

2.1.3 Learning

Emotions play an important role in cognitive processes and specially in learning tasks. Moreover, there are some evidences that the emotional state of the learner is correlated with his performance. Researches in neuroscience and psychology [7, 3] have shown that emotions exert influences in various behavioral and cognitive processes, such as attention, long-term memorizing, decision-making, etc. Additionally, positive affects are fundamental in cognitive organization and thought processes; they also play an important role to improve creativity and flexibility in problem solving. However, negative affects can block thought processes; people who are anxious have deficit in inductive reasoning, slow decision latency and reduced memory capacity. This is not new to teachers involved in traditional learning; students who are bored or anxious could not retain knowledge and think efficiently.

The impetus to engage in and persist with any learning activity is directly linked to the emotional state of the learner (how he feels about himself) and his motivation (how he feels about the subject). Positive emotions allow efficient acquisition and creation of knowledge, while negative emotions reduce and inhibit retrieval and memory. Learning is a complex process and involves both a cognitive and emotional process. In fact, emotional intelligence is recognized as an essential component to the success of human learning processes. Goleman [10] argues that "emotional intelligence" is of much greater importance than "academic intelligence" in developing a well-rounded person, stating that "at best, IQ contributes about 20 percent to the factors that determine life success, which leaves 80 percent to other forces".

Positive and negative emotions can go as far as affecting the way the brain processes and retrieves information. An "emotional high" will provoke the release of endorphin in the brain, which in turn, trigger the flow of acetylcholine, the vital neuro-transmitter that orders new memories to be imprinted in various parts of the brain. Ronald Kotulak [20] describes acetylcholine as "the oil that makes the memory machine function. When it dries up, the machine freezes.", often resulting in Alzheimer.

The involvement of the emotional-right brain with the cognitive left-brain can dramatically improve learning. It is suggested that a synergistic principle operates between the hemispheres, hence, a functioning whole becomes significantly greater than the sum of its parts. Researcher Colin Rose [9] provides us with a clear example of this phenomenon: "If you're listening to a song, the left brain would be processing the words and the right brain would be processing the music. So it's no accident that we learn the words of popular songs very easily."

2.1.4 Behavior

There are two theories on the relationship between emotions and behavior. The first one that is widely accepted, simple and drives from intuitive understanding of emotions and behavior. This theory is known as direct causation theory. It holds that emotion directly causes behavior. Actions can be explained by citing the emotional state that gave rise to them: someone did something "because he was angry" or "because he was happy" or "because he was afraid" or "because he was sad". The evolved purpose and function of emotions was to cause people to act in particular ways. On the other hand recently a new theory showed up that disagrees, it holds that conscious emotion tends to come after behavior and operates as a kind of inner feedback system that prompts the person to reflect on the act and its consequences, and possibly learn lessons that could be useful on future occasions. People may choose their actions based on the emotional outcomes they anticipate. The influence of emotion on behavior is thus indirect [38].

The direct causation theory has, until recently, been very well accepted theory. Mostly due to its theoretical and intuitive appeal that depending on how a person feel he will react in a predefined way. New studies have however shaken this theory and put a doubt on the results confirming it.

Firstly it is not a well established theory, it does sound very convincing but it doesn't offer actual evidence. Even though some authors claimed that "the idea that emotions exert a direct and powerful influence on behavior receives ample support in the psychological literature" they never provided a list of such references. Some researchers who were trying to supply arguments to reinforce the theory ended up providing just more evidence on how emotions affect cognition, but much, much less its effect on behavior.

However there is some literature that offers proofs to support direct causality theory, but the results they provide are often less than optimal and sometimes downright counterproductive. Some studies showed that emotions do cause people to do irrational, destructive and even self-destructive things. That is named self-defeating behavior. When people find themselves in intense emotional states they sometimes do things that bring suffering, harm, or failure to themselves. When upset some people take foolish risks, that have no real chance of a good payoff. At first blush, it might seem that such findings could support the direct causation theory. But evolution would not likely build the psyche with mechanisms that cause it to harm itself. So, even the observations about self-defeating behavior could support the idea that emotion does sometimes affect behavior, but they contradict the idea that is its main function. Self-defeating behaviors are almost by definition unwanted side effect of processes that serve other, adaptive functions. If emotions do cause behavior in the form of self-defeating behavior, that indicates that their main function lies elsewhere.

Other studies that tried to support the direct causation theory didn't bring particularly significant results. In over four thousand articles and four hundred tests for mediation by emotion. Half of those looked for effects on behavior, of them, only 17% were significant at the .05 level. The remaining studies examined effects on judgments and only 18% reached significance.

The other objections concern the inequality between the number of human emotions and number of possible behaviors in a given situation. There are many behaviors but not nearly as many emotions. Emotions are thus not specific enough to give rise to specific behaviors, as the direct causality theory requires. Specific behaviors depend on the situation and its structure of opportunities, constraints and affordances. Therefore behavior cannot be driven by the emotion alone, at most, emotions might activate broad tendencies toward approach and avoidance, but what specific form the behaviors would take would depend on the situation.

Concentrating on individual emotions and searching for its direct influence on behavior haven't offered strong or significant results. On some occasion it might appear that emotion leads to a certain behavior, but on reexamination it shows that the behavior was predicated on anticipated emotion, not on the current emotional state. For example, sadness had been shown to lead to helping. But findings indicated that sadness does not directly cause helping, rather sadness makes people look for some opportunity to escape from sadness and they strategically decide to do good deeds in order to achieve this goal. The operative relevant effect of emotion is that emotion is the goal and the outcome of the behavior, not its direct cause.

Another usually accepted stereotype is that anger causes aggression. Even though this fact is widely remarked, in practice aggression researchers found it nearly impossible to get laboratory participants to behave aggressively. Similarly to sadness study angry people who tend to aggressive behavior when they believe they can change their emotional state. The implication is that angry aggression is a strategic effort to improve one's mood.

In all of the cases in which emotion does seem to cause behavior, further study with appropriate control groups again disconfirms the direct causation theory. What looks at the first glance like emotion causing behavior is in fact behavior pursuing emotional outcomes. This brings us to the second, feedback theory, which proposes precisely that: Emotion functions as the outcome of behavior.

The second theory states that emotion is the goal rather than driver of behavior. This theory is known as theory of emotion as a feedback system. The core idea is that full-fledged, conscious emotion serves mainly to provide feedback after behavior, by stimulating the person to reflect on recent actions and their consequences and possibly to learn lessons for the future. This approach overpasses the pitfalls of direct causation theory. First just by its definition the focus is to learn the consequences after certain behavior, and it is already established that emotions are in strong relationship with cognitive processes. Second emotions are by nature slow-arising therefore in most cases the behavior happens before the emotion comes to presence. That again shows that emotion does not trigger behavior because most of the times it appears after the behavior has happened, on the other hand it goes on the wagon of feedback theory because the emotion is present during the consequences of behavior and thus helps "the lesson to be learned".

As it has been stated previously, researching emotions like fear and sadness some scientists have come to conclusion that although it appears that people who are sad often are willing to offer to help others, they are doing so because they are sad and this drives them to help others. On a better inspection it shows that they are doing so because they have come to learn that when they help somebody they will feel better about themselves. They didn't act because they felt sad, but they have chosen actions strategically to produce emotional outcomes they desired. Anticipation of the emotion was the driver for behavior, not emotion itself.

Several important sets of findings about anticipated emotion lend credence to the feedback theory. One is the evidence that anticipated regret, in particular, can influence decisions, mostly in beneficial, advantageous way. That is, people make choices based on anticipating what will bring them regret, and the impact of this anticipation on the choices is generally to steer people to choose in ways that will benefit them.

Research on affective forecasting is also relevant. Affective forecasting refers to people's predictions of how they will feel under future or hypothetical circumstances. People tend to predict that their emotional reactions to future events will be relatively long-lasting, whereas when such events occur the emotions tend to dissipate. In a sense, people overpredict their emotions. The overprediction of emotion indicates the importance of anticipation. If people underpredict their emotional reactions, it would be very difficult to suggest that anticipation of emotion is important, because anticipation would tend to be small and trivial whereas the experienced reality would be relatively large and impactful.In a sense, then, the biggest emotion is the expected one, rather than the actually experienced one. Emotion looms larger and thus presumably has more impact in anticipation than in actual experience.

It is important to stress out that the behavior described in this theory is the set of actions that people would consciously undertake, not the unconscious behavior that people demonstrates: biological reactions like heart hate, blood pressure, muscular contractions, skin conductance. For the purpose of our research we focus on the physiological effects caused by emotions such as change in heart rate and muscular tension and vibrations that should be visible on the dynamics of typing or movement of the mouse.

All in all, in [38] there were introduced two interesting theories, mutually not excluding: the causation theory and the feedback theory. The latter is more general than the former. In this work we will focus on the effect of the emotion on behavior of keyboard typing and mouse movement.

2.2 Human-computer interaction

The computer era has begun long time ago. Personal computers are more and more present in our every day lives. People use them to work, for personal needs, for entertainment, to get informed. We all got used to them and certain tasks seem impossible to achieve with out them. Their great advantage is the easy-to-use approach and intuitiveness, since it takes little time to get used to keyboard and mouse as input components. Still more advanced machines can use video camera and microphone as input devices, which offers new ways of interaction.

A basic goal of Human-computer interaction (HCI) is to improve the interactions between users and computers by making computers more usable and receptive to the user's needs. Specifically, HCI is concerned with: methodologies and processes for designing interfaces, methods for implementing interfaces (e.g. software toolkits and libraries; efficient algorithms), techniques for evaluating and comparing interfaces, developing new interfaces and interaction techniques and developing descriptive and predictive models and theories of interaction.

A long term goal of HCI is to design systems that minimize the barrier between the human's cognitive model of what they want to accomplish and the computer's understanding of the user's task. Professional practitioners in HCI are usually designers concerned with the practical application of design methodologies to real-world problems. Their work often revolves around designing graphical user interfaces and web interfaces. Researchers in HCI are interested in developing new design methodologies, experimenting with new hardware devices, prototyping new software systems, exploring new paradigms for interaction, and developing models and theories of interaction.

The interaction sofar is a very intuitive one, and very similar to a natural interaction, what one part does affects the other part, that than conducts some actions to affect the first part again. But the emotional component is being neglected entirely [32, 31]. Although everybody is conscious that computer feels no emotions and does not provide any emotional response, people cannot run away from their nature. Depending on their emotional state and results they obtain from computer, people are affected and change their state as a result of interaction. Knowing that their emotion does not affect the computer, it often happens that agitated users scream on their computers, engage violent behavior, or even silently pray to it to output an positive result.

There are efforts to render human-computer interaction more smooth, to minimize possible aggravation of user. Interfaces are designed to be as much user-friendly as possible, functions to be good organized and ambiguities reduced to the minimum. Some designers have come up with friendly mascots that offer services to users, but usually do not reach desired effect. The classical example is Microsoft's clipper, a smiling happy dancing clipper offering suggestions. In most cases people are annoyed and irritated with it and shut it down as quickly as possible. And the most provoking fact is that even if you closed it every time at the moment it has appeared, it comes back with the same enthusiasm hoping that this time is the charm.

That kind of operability offers affective computing, the branch of artificial intelligence which aims designing computer systems that are capable of recognizing, interpreting and processing human emotions. Affective computing tries to create intelligent agent that will perceive users emotional state as its environment and take actions in order to maximize human satisfaction and minimize stress and tension. Together with human-computer interaction designers, affective computing is trying to create systems that minimize the chances of irritating the user and recognize user's irritation in run-time so that the system could react and meet the user's emotional state trying to satisfy him again.[30]

The goal of affective computing to approximate the human-computer interaction to human-human interaction, but not so drastic to create a new friend or offer psychological therapy. The goal is go get rid of cold text based paradigm that is deprived of tinniest color of emotion. Humans do not only feel their emotions, they also recognize emotions of others (empathy), and they act and speak differently when interacting with a person depending on the emotional state they encounter (emotional intelligence). Some people lack some emotional intelligence skills, so in interaction they, even if it wasn't their intention, can worsen the emotional state of the other person. And usually it is not the problem of "what" they said but the problem was in approach, "how" they phrased it, the intonation they used, gestures etc.

Those skills are exactly what lacks in computers. To show they care for user's problems, to "feel" empathy and communicate messages in more appropriate way. It is obvious that computers will never learn to care or feel for real, and no one will be fooled in thinking so. Nevertheless some studies made tests with interactive agents, that tried to build social-emotional relationship with users who were undergoing a month-long program to increase their exercise levels. The agent was designed to converse about user's feelings and show occasional empathetic caring and concern with both text and boldly expressions. Even though users doubted the character had any feelings or really cared, they rated the agent significantly high on likability, trust, respect, feelings it cared for them and willingness to continue interacting with it. Thus even with limited affective abilities this experiment improved the quality of experience for users, especially over long-term interaction.

There have been similar studies with more advanced expression abilities of agent. The agent was presented as and animated avatar that could change facial expressions. And that study also gave very promising result, demonstrating users satisfaction and better scores on test [29, 16, 19, 27].

2.3 Emotion recognition and their role in HCI

In order to provide natural ways for humans to use computer as aides, the computer needs to have communication skills of humans. One of this skills is the ability to understand the emotional state of the person. Up to now many methods have been developed and provided significant results. Such methods are facial expression recognition, speech recognition, gesture recognition and recognition from biological signals.

The emotion recognition can contribute to HCI on a large scale. Once a computer is able to recognize the emotional state of the user it could adopt to operate in the way that is best suitable for the user. The computer agent should observe human reactions to different situations and learn to repeat the behavior that was agreeable to people and stop doing the actions that irritate the user.The learning process could be developed through machine learning techniques that use input information to reinforce the good results and penalize the bad outcomes.

User satisfaction is one of the main characteristics of a good product. If computer is trained to recognize human emotions and is able to operate in order to keep the user in good emotional state or perform actions to suppress negative emotions it would not only provide greater satisfaction with the product but even raise the productivity of user's work. It is a well established fact that when in positive emotional states people do better their jobs, opposed to when in negative states, while then person's concentration and productiveness drops.

2.3.1 Facial expression recognition

The most natural way of emotion recognition that comes to mind is the analysis of the facial expression. Facial expressions are very fast, usually people express facially their emotions before that they can verbalize them or realize them. It is also a reasonable choice: Ekman and al. [13] have established that facial expressions for the 7 basic emotions (neutral, happy, sad, angry, afraid, surprised and disgusted) are "species-specific" and not "culture-specific". Even more, babies from their earliest days show facial expressions which clearly shows that they haven't been thought, but that the expressions are innate. Therefore, once a good classifier is built it could be universal.

Human face has 20 groups of facial muscles, having a rich combination of gripped and relaxed muscles there is a vast number of possible expressions one can produce. Studies have shown that the vector space is defined on 46 of so called action units that keep track of movements of the key muscles that tell on the current emotion. Some of those action units are: the center of the lips, corners of the mouth, corners of eyes, lifting of cheeks, blinking of the eyes.

The recognition of the emotions can be static or dynamic. Static classifiers use features from a single image of the face, while dynamic tries to find the temporal pattern in a video sequence. The results are good both for static and dynamic analysis of facial expression. One of the best methods is Facial Emotion Tree Structure (FETS) that scored over 85% accuracy recognizing 7 different emotions, both for known and unknown subject.[11]

Even though this method presents very good results there are some drawbacks. Improvements are needed on the cameras to be more robust on lighting conditions and partial occlusion. If subject face is not well lit or if it is partially occluded the system fails to detect face at all. In most cases

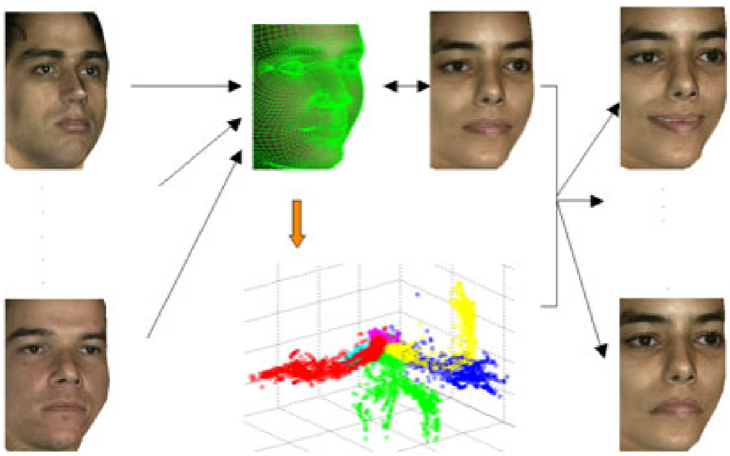

Figure 2.2: Facial expression recognition

the face detection is still not automatized, it is needed to chose manually the points to trace. Another problem is related to the ethical aspect of the method: not everybody is happy with the idea to be monitored over video camera during the time spent on computer.

2.3.2 Speech recognition

The way people express themselves verbally also serves to indicate their emotional state. During the interaction one can pick up on the emotion just through the audio channel. The emotional content in speech is a major information carrier along the explicit verbal content. The emotional expression, which is partially voluntary and partially involuntary, affects multiple linguistic and acoustical features of speech. Some of the many indicators of the emotion in speech are speed, pauses one takes while speaking, how loud one speaks, the way he/she articulates and other cues.

A number of studies have been conducted in this area. Still it is not determined no widely acknowledged set of speech signal characteristic. Usual feature set consists of pitch, log energy, formant, mel-band energies, mel frequency cepstral coefficients (MFCCs), velocity/acceleration of pitch and MFCCs to form feature streams. Empirically is shown that pitch and energy wave are primary factors in emotion recognition from speech [21].

Speech characteristics that are commonly used in emotion recognition can be grouped into four different categories. The first one includes frequency characteristics (for example a pitch and pitch-derived measure) which are related to voiced speech generation mechanism and vocal tract formation. The second group contains various energy descriptors that are related to speech production processes (such as mean or standard deviation of energy of an utterance). The third group comprises temporal features, which are related to behavioral speech production processes (such as utterance duration, pauses). The fourth group consists of energies in sub-bands of a speech signal spectrogram. Some of emotions can correlate better with energies embedded in specific frequency intervals than with a total energy of a speech signal.

Once feature set is extrapolated emotion recognition is typically performed using neural network, SVM, Gaussian SVM (GSVM), LDA, QDA or binary-tree based recognition strategy. Out of the typical classifiers GSVM produced the best results, an overall accuracy of 42.3% for five emotions: angry, bored, happy, neutral, sad [21]. Another research based on binary-trees managed to score 76.3% for speaker-dependent emotion recognition and 72.04% for speaker-independent [18]. Latter research was realized for recognition of six emotions: neutral, sadness, fear, anger, joy, boredom.

Even though these results are promising and advances are observed emotion recognition from speech signals has some deeper obstacles. The classifiers are language and cultural dependent. Because of linguistics and phonetics, language apparatus creates differences and unique trademarks for distinct language speakers. This differs from the emotion recognition from facial expression where it is established that facial expressions are universal. Nonetheless this provides a good reason to build a hybrid classifier that will process both speech and facial expressions. It is even natural, humans tend to recognize emotions better through audio and video channel together, rather than just from listening to speech or watching the mute image.

2.3.3 Physiological signals

It is possible to recognize emotion from physiological signals. When a person enters an emotional state, the change is not just psychological but the whole body responds to the emotional shift. Even though science doesn't have all the answers on dynamics and processes about how emotions are triggered and how are they managed, most likely the explanation hides in hormones that are released when a certain emotional state is reached. The hormones are responsible for changes in hear rate, respiration, perspiration, muscular tension etc.

For a long time there have been a serious doubt whether it would be possible to recognize emotion from physiological signals, due to number of possible signals, number of emotions, poor technology and little knowledge. But these obstacles didn't shake down enthusiasm. Taken in consideration the complexity and variety of human emotions researchers concentrate on smaller subsets of basic emotions, leaving fine-grained emotions behind. With advances in technology and machines to collect physiological signals and many experiments done, the research has obtained accurate recognition methods that can classify seven basic emotions with a rate significantly higher than random guessing [30].

The most successful methods acquire four signals, electrocardiogram (ECG), skin temperature (SKT), respiration and skin conductance (SC), also called electodermal activity (EDA). From those four signals a number of features is extracted that serve to create a classifier. There is still no winning classifier, but they all have very high performance, the difference is that one classifier performs better for one emotion while the others prefer another emotion.

Electrocardiogram(ECG) presents electrical activity of the heart. Analyzing the power spectrum from heat's inner-beat intervals and dividing them into three bands: low frequency band (0.01 to 0.08 Hz), middle frequency band (0.09 to 0.14Hz) and high frequency band(0.15 to 0.5Hz). Computing integral of power spectrum in these bands and comparing the relative magnitude of these bands it can be used in a number of ways, as to determine the emotional state. Mental stress and ager increase the activity in low frequency band rather than middle or high frequency bands. Also higher activity in low frequency band indicates the sympathetic activity whereas higher activity in high frequency band indicates the parasympathetic activity. The main problem with ECG is sensitivity. In most cases it gives rather noisy output and therefore does not provide clear data to analyze [22, 23].

Skin conductance (SC) is a measure of alteration of electrical resistance of the skin associated with sympathetic nerve discharge. It characterizes changes in electrical properties of the skin due to the activity of sweat glands and is physically interpreted as conductance. Sweat glands distributed on the skin only receive output from the sympathetic nervous system, and thus SC is a good indicator of arousal level due to cognitive stimulus [22, 23].

Specific emotion expressions, such as crying, laughing, or shouting, have unique respiratory signatures. A detailed quantification of volume, timing and shape parameters in the respiratory pattern waveform can map into different emotional states along the dimensions of calm-excitement, relaxation-tension, and active vs. passive coping. Other evidence indicates that respiratory parameters also map into affective space dimensions of valence and arousal [22].

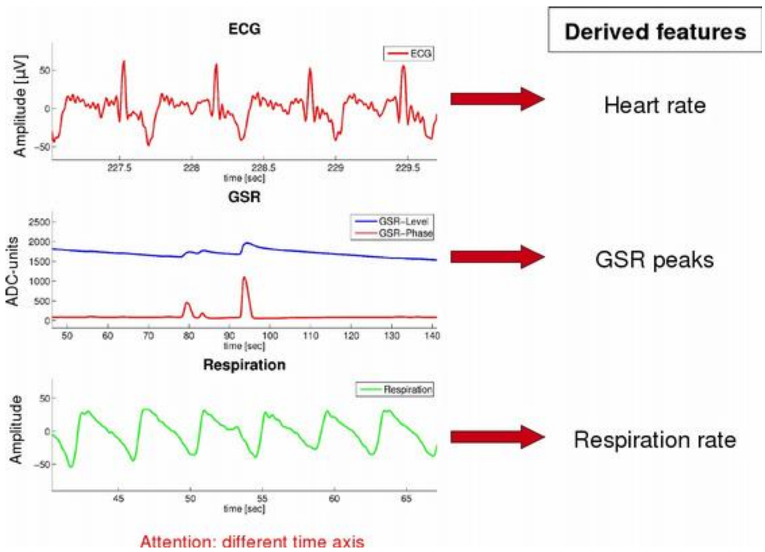

Figure 2.3: Example of physiological signals

Some good results were obtained with electroencephalogram (EEG) and facial electromyograms, but the two aren't so popular because electrodes have to be attached on the scalp and the face that are not tolerable for practical use. There is a need for practicality in wearable devices.

Researches have demonstrated that anger, fear and sadness increase the heart rate more than happiness and surprise do, while disgust decreased the heart rate. Anger increases skin temperature more than happiness and sadness do, while fear, surprise and disgust decrease the temperature. SC increases with arousal, decreases in neutral state and stays the same during the sadness.

In the research by [26] using KNN algorithm the recognition accuracy they obtained was: 78.85% for anger, 75.00% for sadness, 70.00% for fear, 66.67% for surprise, 58.33% for frustration and 43.75% for amusement. Using DFA algorithm they got 90.00% for fear, 87.50% for sadness, 78.58% for anger, 56.25% for amusement, 53.33% for surprise and 50.00% for frustration.

2.4 Applications

The emotion recognition is just a first step into building an affective interface. To increase quality and productivity of human-computer interaction interfaces should be modeled in such manner that after the detection of emotion the system should respond either to encourage the positive emotion or to engage procedures to diminish negative emotion. Or to exploit the given emotional state to complete specific tasks, because it has been determined that specific emotional states are key factors to better accomplish specific problems [15, 1].

Recent studies have developed many models of HCI that take in consideration human emotion and respond in particular way thus increasing the performance. Here are represented some of them.

Rehabilitation

To support rehabilitation of lost functionalities robots can be exploited to allow specialized therapists to follow many patients at the same time. Measuring the user's emotional state provides additional information to give an exhaustive description of the subject state and his engagement in the rehabilitation task: in particular, monitoring stress during an exercise can be useful to reduce frustration and improve interest.

Psychology

Affective computing can be used to incentive the emotion communication in case of autistic people. An approach to create a teacher is presented: it is a persistent aid for the children that progressively introduces basic emotional expression, guides recognition development through machine and records the child's success.

Learning

Other applications in this field include improving the learning process. With skills of affect perception a computer that detects the learner making a mistake while appearing curious and engaged could leave the learner alone since mistakes can be important for facilitating learning and exploration; however, if the learner is frowning, fidgeting, and looking around while making the same mistake, then the computer might use this affective feedback to encourage a different strategy. A computer able to understand the affective state could understand what interests the user too: in a scholastic scenario for example, instead of system recording everything the user hear, it might play back just those places in lecture that the user missed because of its mind wondered or because he is bored.

Multimedia/games

Integrating wearable biosensors with consumer products, for example a MP3 player, even if it is impossible to exactly predict what song the user wants to listen, it is possible to predict, basing on his mood, what kind of music he would like more. Moreover, application of affective computing to video games has the focus on the sensing and recognition of the players' emotions and on tailoring the game response to these emotions.

Automotive

Integrate with cars a system able to recognize the affective state is an argument of big interest. Information on emotions could be used automatically by adaptive systems in various ways to help the driver better cope with stress. Some examples of this might include automatic management of non-critical in-vehicle information system.

2.5 Typewriting skills

In modern computers input devices have become numerous and advanced, like speech recognition and handwriting recognition systems. But the main input device still is the keyboard. Since the era of typewriters the keyboard has been the main communication device between man and machine. Taken that in consideration the performance of keyboard usage will be in the focus of this thesis.

The closer look on sofar developed models of typewriting skills is a valuable help to determine and model the user's work on his computer. The keyboard in present day computers is not only a device for typing and copying text but it is used for a number of different communication protocols. It is used for passing commands to computer using only one key at a time, using arrow keys for navigation, or pressing certain control keys to perform a special activity. That's why it is important to detect the different behaviors of the keyboard. Most helpful is to determine the pauses in typing on the keyboard. Analyzing them could provide us with the patterns to categorize the actual writing from occasional key pressing.

In the works from Salthouse [35] and Wu & Liu [8] there are a number of phenomena that describe the typewriting and help to distinguish it from other keyboard usage. They came up with the list of phenomena in typewriting taking in consideration concurrent perceptual, cognitive and motoric processes involved in transcription writing. Following is the list of phenomena that could serve in the typewriting analysis. Fist part of the list consists of the basic phenomena and they are related to the major factors affecting the interkey time. Interkey time refers to interval between two adjacent keystrokes, and is regarded as the basic measurement of human performance in transcription typing.

- The rate of typing is nearly the same for random words as it is for meaningful text.
- The rate of typing is slowed as the material approaches random. The difference form the previous phenomena is that the former refers to the order of words being randomized, while the latter refers to the order of characters within each word being randomized. It was found that the average interkey time in typing increased to 454ms when subjects are typing materials composed of words with random characters
- The rate of typing is severely impaired by a restricted preview of the material to be typed. Decreasing the number of characters to be typed in the restricted preview increased the interkey time and severely impaired the typing rate.
- Alternate-hand keystrokes are faster than the same-hand keystrokes (called the alternate-hand advantage). Successive keystrokes from fingers on alternate hands are 30-60ms faster than successive keystrokes from fingers on the same hand.
- Digram (letter pairs) that occur more frequently in normal language are typed faster than less frequent digram (called the digram frequency effect).
- Interkey time is independent of word length.
- The first keystroke in the word is slower than the subsequent keystrokes (the word initiation effect). Mostly the interval before the first keystroke in a word is approximately 20%(45ms) longer than that between the later keystrokes in the word.
- The time for a keystroke is dependent on the specific context in which the character appears, especially for the topography of the keyboard (the context phenomenon).
- A concurrent task does not affect typing performance, especially for highly skilled typists, a concurrent activity can be performed with little or no effect on the speed of typing.

The next four phenomena are four categories of most common mistakes in typing.

- Many substitution errors (e.g., "work" for "word") involve adjacent keys. Experimental results from highly skilled typists indicated that 30.1% of substitution errors involved horizontally or vertically adjacent keys.
- Many intrusion errors (e.g., "worrd" for "word") involve extremely short interkey time in the immediate vicinity of the error. Nearly 38% of the intrusion error keystrokes had ratios (interkey time of an error keystroke divided by that of the regular interkey time) less than 0.1 of the average interkey time and over 54% of intrusion errors involved an adjacent key in the same row or the same column.
- Many omission errors (e.g., "wrd" for "word") are followed by a keystroke interval approximately twice of the overall median. Interkey time of the keystroke right after the omission error was 1.54 times longer than that of the average interkey time.
- Transposition errors (e.g., "wrod" for "word") mostly occur cross-hand, 80% of the transposition errors were typed by the opposite hands.

These phenomena describe in detail characteristics of typewriting behavior. They will give a better understanding of the data that will be collect by keylogger. It also narrows the search space to those key phenomena to search for their differences under distinct emotional states. Additionally we expect higher error rate while user is undergoing stress or feels tired, the quantitative analysis of listed errors could uncover some patterns for discrete emotions.

Chapter 3

Problem Definition and Design of Experiments

The goal of this thesis is the search of the model representing user behavior while working on computer, taking in consideration user emotional state and interaction with keyboard and mouse. As presented in the previous chapter, emotions play important role in human everyday activities. Depending on the emotional state people will perceive, learn, memorize and behave in different ways. The effects of the emotion on the user behavior in human-computer interaction is the focus of this research. It is an innovative research therefore it's mostly based on our hypothesis and attempts to prove them.

The behavior we are interested in is the dynamics of keyboard typing and mouse movement. These two devices are the most common input devices used in human-computer interaction. The two are almost in constant touch with the user, and therefore suitable as a long term data collection agents. Under different emotional states we expect to find patterns in typing and mouse movement dynamics, because in every emotional state biological manifestation of the body differs, significant changes are evident in heart rate, body temperature, skin conductance, muscle tension. We expect those manifestations to have influence on the muscles in the hands, thus causing different speed of typing, changes in duration of pressure of keys, changes in mouse movements etc.

3.1 User behavior model

The graphical representation of the user model is shown in figure 3.2. The first set of variables is Activity. For a discrete moment "t" A(t) depicts what user has been doing from the moment "t-1" until the moment "t".

Figure 3.1: Common computer input devices

In that time window activity is everything that user has been doing, either on his computer or any activities besides the computer. From the point of view of input devices it can be only determined what the user has been interacting with computer. Whenever he/she stopped being in contact with input devices for some period of time we will assume that the user has taken a pause and relaxed. Due to impossibility of knowing what user is doing besides the work on computer we will consider Activity variable generally hidden, but we will try to observe the activity on the computer in as much detail as possible. Two methods we have chosen will be explained later under sections Experiment 1 and Experiment 2, because their design is based on the user behavior model.

Other set of hidden variables is Emotion, E(t). People change their emotional state throughout the day. Between moments "t-1" and "t" it is quite possible for user to pass a series of emotional states, but we are interested only in the emotion the user is experiencing in the observed moment "t", because the current emotional state is the one that is directly influencing the observed variables. As it has been stated in previous chapter, surrounding is stimulating emotion, so the work user is doing influences his/her emotional state. This is the other reason why it is important to find good way of representing user activity, to be able to find the best behavioral patterns

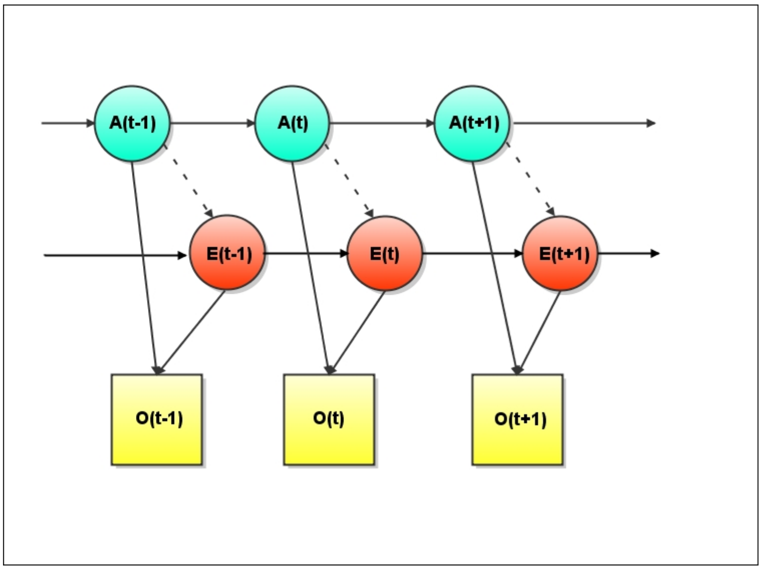

Figure 3.2: User behavior model

causing distinct emotion.

The observation set is composed of data and features gathered from the input devices, namely mouse and keyboard. In specific moments the user will be performing some control activities (tasks), that were beforehand designed, in order to have more relevant data on Observation variable. Features that will be of our interest are writing speed, key pressure duration, mouse movement speed, continuity of mouse movement direction etc. Detailed description od data and features will be given in the chapter Data and Features.

3.2 Application elements

Both experiments we designed consist of similar elements with minor changes between them. In this subsection the building blocks will be described and later under Experiment 1 and Experiment 2 it will be explained how the blocks are used and if there were some specific changes from the general model given in this section.

3.2.1 Data Loggers

The observation we are interested in is described with the data of usage of the keyboard and mouse. Data collection was performed using keylogger and mouselogger, two executable programs adopted from an earlier, similar thesis by Alberto Franciosi [17].

Specific Keylogger

Keylogger is a hook based executable program that *hooks*the keyboard using functionality provided by the operating system for applications to subscribe to keyboard events. The operating system notifies the keylogger each time a key is pressed or released and the keylogger simply records it. On each key event keylogger returns to the prompt a row composed of 4 integer values separated by a blank space. The first value of the row is the unix timestamp in seconds representing the time when the event (keypressed or keyreleased) occurred. Second integer value is the rest of the second up to microsecond of the event occurrence. That is not the exact time when user has pressed the key but the moment when event was recorded in the central processing unit. Depending on the operating frequency of the processor it is possible to establish the absolute maximal error of the time representation (on the machine we tested the experiment the error was 6.125 ms). In average man does not write that fast so the errors will not be of great importance. However sometimes when it happens we will assume that the two keys that have identical timestamps are pressed simultaneously.

Third data returned by the keylogger is a 0 or a 1. It just states what event is being described within the row values. Value 1 stands for keypressed event and 0 for keyreleased event.

The last data provided by the logger is the code of the key being pressed. This code is dependent on software the keyboard is using. The code differs depending on what language for the keyboard is used. For example using *qwerty* keyboard the key to the right from the key ‘p’ in English keyboard is ‘[’ , in French is ‘ˆ’ , and in Italian is ‘é’ their values are 219, 221 and 186 respectively. Therefore while processing keyboard data we designed it to process Italian keyboard given that the users will be Italians or people that frequently use Italian keyboard.

Keylogger also keeps track of mouse buttons being clicked. Just as for keyboard keys it returns same four integer values, with specific codes for each of mouse buttons.

Generally keyloggers are treated as malware. They represent a threat to personal privacy, because the one who collects the data has a total insight into user’s keyboard activity and can easily reconstruct what user has been writing during the whole session: personal e-mails user has been writing, all the user names and passwords that were written, everything user was

characters	specific keylogger	whole-day keylogger
h	72	100
o	39	100
	32	32
2	50	50
5	53	50
	32	32
a	65	100
n	78	100
n	78	100
i	73	100

Table 3.1: Difference between specific and whole-day logger

writing for his/her own work. Therefore we will be using specific keylogger only in moments known to user, when he is doing the tasks relevant for this research in order to have a transparent observation his/her actions.

Whole-day Keylogger

The specific keylogger is a threat to user privacy, on the other hand we are interested in what user is doing through out the day, to get a minimal description of the kind of activities he/she was doing outside of observation window. The whole-day logger introduced modification that solves this ethical problem. Instead of logging exact character codes, we will hide them with a masking code: all character codes will be masked with value 100, and all number codes will be masked with value 50. Remaining codes of function, control and other keys will be logged as previously, because the usage of those keys is not reveling or invading privacy and we are interested in distribution and frequency on using those keys. We hypothesize that usage of those keys might benefit the analysis of activity period and reveal some patterns that indicate a certain emotion, because when behavior changes it usually leads to emotional response.

In the table 3.1 are shown the examples of specific keylogger and whole-day keylogger. Both have simultaneously recorded writing of the string “Ho 25 anni”. The second column shows specific key logger, every character, number or sign has it’s distinct value. While in the third column whole-day logger recognizes that there were some characters pressed, two spaces, two numbers.

Mouselogger

Mouselogger returns also four integer values to the prompt. Values are: unix timestamp in seconds, rest of the second in microseconds, x and y coordinates of the mouse pointer position on the screen.

Mouselogger is also treated as malware, but it is not as revealing as keyloggers. Usually mouse logs are not usable without the keylog, because they just record the coordinates where mouse pointer was moving on the screen. Without knowing what was on the screen or what user has written after specific movement, mouse logs are useless. Nevertheless we explained to users how are working both keyloggers and mouselogger before applying them for the experiment and waited for their consent.

3.2.2 Tasks

Having data provided by keylogger and mouselogger we achieve partial observation of user's behavior. The exact user activity is still hidden. Only by knowing what is he/she doing the observation will be complete. To do so we designed a series of tasks in form of web pages, created in PHP. Task is a short exercise user has to solve, designed to be solvable specifically with a mouse or with a keyboard. There are two types of tasks: activity and induction tasks.

Activity tasks

Activity tasks are short and simple problems that user has to solve. They have to be simple so that user doesn't loose focus, if task is complex or difficult user will spent time in thinking and reflecting and want be using the input devices. We want him to continuously execute the task. From the similar reason we required tasks to be short, if they take to much time to be solved user will begin to feel bored, which is a undesired state. Activity tasks should not influence user's emotional state. Besides the data from mouse and keyboard, we are interested in performance of each task, we expect user to manifest different results in different emotional states. Following is the list of activity tasks.

- Copy The task is designed for keyboard observation only. User is given a text, three of four phrases long. He/she has to rewrite it in a below given textbox. It should be copied as correctly as possible. To escape "cheating" a javascript function is used that disables possibility to select the text with mouse and then simply "copy/paste" it.

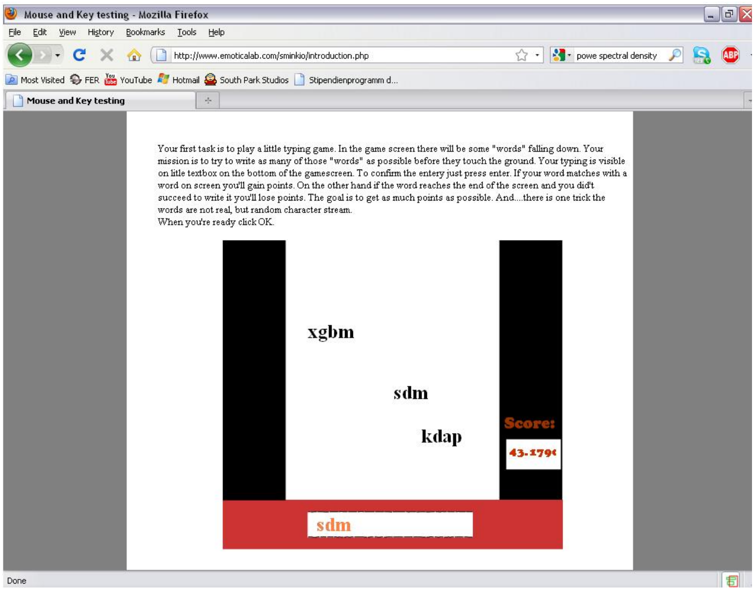

Figure 3.3: Keygame task

- Keygame Keyboard game was designed to measure writing speed of unknown words. In this game a word is very relative term. Words were not real words from any known language, but strings of random letters 3 to 6 characters long. When the game starts those words come falling from the top of the screen to the bottom of the screen. The falling speed of words variate between 5 and 15 pixels per second. User can follow his/her progress on a scoring label. If user managed to write the word before it hits the ground, the word vanishes from the screen and user earns some points. If however the word hits the ground user looses some points. Scoring is introduced to motivate user not to give up on playing, consequently not to stop providing data. When 60 words have been generated and fallen down the screen the game is ended.

- Navigation The task is designed for mouse use only. User is supposed to correctly find three links on a web page. The task page is a copy of Repubblica.it homepage, because it is big enough to scroll up and down, but not too big to get lost, and it has a significant number of links that could be clicked on. The copy of Repubblica page needed some editing. All the links that are either hidden or non textual were removed. From the rest of textual links duplicate links (links that

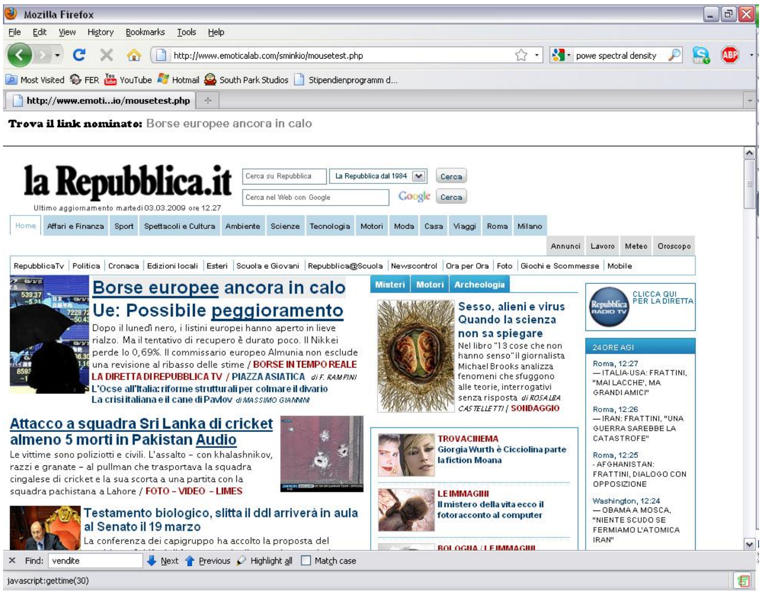

Figure 3.4: Navigation task

have the same name) were also removed. The remaining links were modified in such way that all point to a same URL.

Names of all links are stored in the data base. The link that user has to find on the web page is randomly chosen from the data base before loading the task page. On the top of the page a frame was added that advertises what link is user supposed to find. If user clicks on an incorrect link browser signals that it was a wrong answer, then chooses another link and repeats the task. If he/she clicks on the correct link, browser increases hit counter by 1. The task loops until the counter reaches 3 correct hits, in that case the task is over.

- Mouse game Another mouse task. It is a fast game but not too straightforward. It is just an adaptation of classical "shoot-em-up" game. The goal is to click with a mouse on target appearing on the screen. Mouse game was intended to monitor user speed and reactions on limited space, since the navigation task covers a big web page that needs scrolling and attentive search. Inside the game screen appear 3 targets at random locations on the screen. Every target has a numerical value. The goal is to click on the target that has the lowest value.

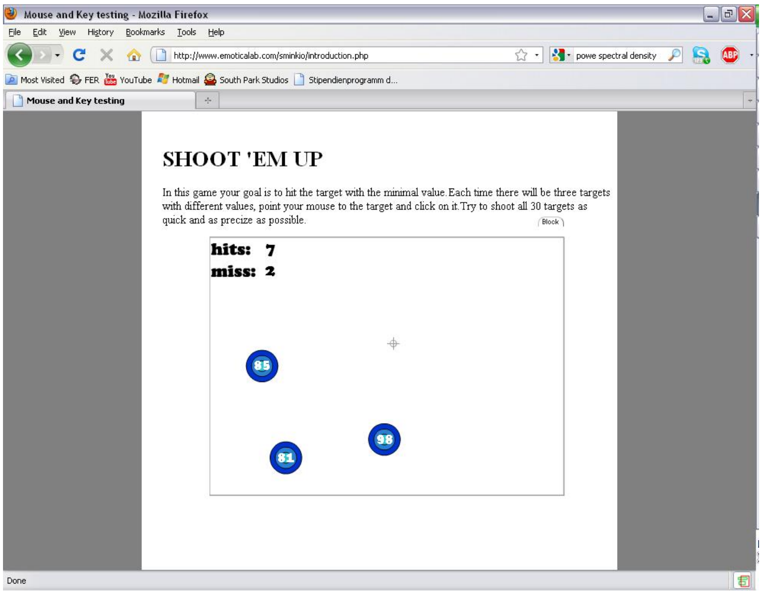

Figure 3.5: Mousegame task

Induction Tasks

Induction tasks have a dual mission. First one is the same as in activity tasks: have the knowledge of what kind of work user is doing, collect data and evaluate performance. Second is to additionally induce emotion. Cognitive problem solving can in general elicit two emotions: higher attention and stress. Higher attention is achieved when the problem requires thinking before solving it, but basically is not complex. When the problem is hard, complex and seemingly impossible to solve, it induces stress.

For this purpose we adopted a game called Strimko. It is a kind of game like Sudoku, it has a grid of 4x4 cells (can be also 5x5 or 6x6). All cells have to be filled with numbers form 1 to 4, so that in every row and in every column there are all four numbers without repeating. Differently from Sudoku there are no subcelles but there are four lines that zig-zag through the grid and cells on each line have to be filled with numbers 1 to 4 without repeating. User plays the game moving and clicking the mouse. On the right form the grid there is the list of numbers, user click on the one he wants to place in the cell. When he clicks on the number the cursor changes in a small circle with currently chosen number written in the center. Now when he clicks on a cell that cell takes that value. If an error was made user can change the number he desires to put in the cell or by clicking on the

C cell, he can erase values form the cells. The values that are assigned on the beginning of the game can not be changed. When user puts all correct values the game stops and pops a message that he has just won the game.

For the elicitation of attention state user is given a 4x4 Strimko (easy Strimko). For it's simplicity and, in most cases, fast wining. While for elicitation of stressful state the game is a 6x6 Strimko (hard Strimko), it is much harder and requires lot of thinking and is much less obvious to solve.

Data Base

For implementation Microsoft MySQL was used. The data base collected all the data throughout the session. Data from the on line tasks were collected directly, while the data from the keylogs, mouselogs and windowlogs were input manually after user has uploaded asked logs on the server. The following describes the tables found in the data base:(after the table name there is the list of its attributes)

User: id_user, username, password, sex, age and hand (character L or R indicating wether the user is lefthanded or righthanded). User table contains the data describing general information of the user. Attributes are pretty straightforward so additional explanation is not needed. Id_user is auto-increment unique key for the table user.

Session: id_session, date and id_user. Session table indicates what session which day user performed. Id_session is auto-increment unique key for this table. Id_user is on the other hand foreign key defining N-to-1 relation with user table.

Test: id_test, id_session. For every session user can complete a number of tests. This table takes note of each test user did in one session. Id_test is auto increment unique key for the table, and id_session is the foreign key.

Testlist: id_test, timestamp, type. Every test consists of four tasks (keygame, copy, mousegame and navigation). id_test is the foreign key to Test table. Table updates the type of the event, start and end of each task. And the time when the event occurred. Events are coded with numbers in rage 1-8. Odd numbers represent the begin event of the task, and even number the end event.

Keylog: id_session, presssec, pressmicro, releasesec, relesemicro, btn. This table collects the keyboard data during the whole session. btn is the numerical code of the button being pressed. In this table all the characters are coded with the value 100, and all numbers with value 50. The other keys are coded normally. Presssec is the unix timestamp in seconds when the keypress event occurred, pressmicro is the time in microseconds. Releasesec is the unix timestamp in seconds when the keyrelease event occurred, relesemicro is the time in microseconds.

Mouselog: id_session, time, micro, cordx, cordy. Time and micro are unix timestamps in seconds and microseconds. Cordx and cordy are x and y coordinates of mouse pointer on the screen in the given moment.

Keylogspec: id_test, presssec, pressmicro, releasesec, releasemiro, btn. This is the log of the keyboard activity during the test phase. Data are teh same as in Keylog table, but here btn does not hide any codes, all alphanumeric keys are written in the table.

Tkeygame: id_test, timestart, startmicro, word, speed, flag, timeend, endmicro. In this table data describe the going-ons in the game. All 60 words generated in the game are gathered together with the timestamp when it was created on the screen(timestart and startmicro), the word created by the game is recorded in the word attribute, the length of the word, it's velocity, flag's value is 1 if it was matched with user's writing, 0 otherwise. If the flag is 1 then timeend and endmicro's value is the time when the user matched the word, when flag is 0 timeend and endmicro's value matches the time when the word has "hit the floor". This and other tables that have id_test attribute use that attribute as foreign key to the test table enabling linking with the test to which it belongs.

Tmousetargets: id_test, time, micro, x1,y1,v1,x2,y2,v2,x3,y3,v3. This and the next table describe the development during the "shoot-em-up" game. Time and micro are the timestamps when the game has generated new three targets on the screen. x and y are the screen coordinates of the centers of each target and vs are the values on each of the target.

Tmouseshots: id_test, time, micro, x, y. This table takes the time of user's click and the screen coordinates where the user has clicked.

Textfiles: id_text, text. This table has 20 rows with predetermined texts for users to copy during the copy task. Each time user has to rewrite another text. This table consults the copy page to pick the text to put on screen, and later the text is used for comparison with the user's to evaluate the correctness.

Tcopy: id_test, id_text, textc. id_text is the foreign key to Textfiles table, showing what was the original text user had to rewrite. Textc is the text of what the user has written.

Links: id_link, link. In this table is the list of links on the Repubblica web page. The page chooses one of the links to be the task to find.

Tnavigation: id_test, link_wanted, timestart, microstart, link_clicked, timeend, microend. link_wanted is the id_link form the Links table. This is the link the user is looking for. Timestart and microstart are the timestamps when the page has been loaded. link_clicked is the id of the link the user has clicked on. By comparing link_clicked and link_wanted it will be visible is the user got it right. timeend and microend are the timestamps in the moment when user has clicked on the link.

3.3 Experiment 1

This experiment is based on the behavior model introduced in the beginning of the chapter. In general model Activity and Emotion variables are hidden, in the experiment we will try to make them observable. To do so the experiment will be composed of the activity tasks to have the insight of what kind of work the user is performing. To observe the Emotion variable we will elicit user emotion applying visual, auditive and cognitive techniques. Three emotions will be induced during one experiment session: attention, stress and fatigue.

A window application Emotion tracker is developed in Java (Eclipse Europe). It is the central application of the experiment, it embeds specific keylogger and mouselogger, coordinates the sequence of tasks and communicates the data to the data base. Sequence diagram of the application is given in figure 3.6

We wanted to collect several sessions for each user in order to have an individual analysis for each user and a general inter-user analysis. To ensure that there is no mixing of data every user upon starting the Emotion tracker application first has to register or log in. While registering we required some data about him/her, namely sex, age and to nominate the stronger hand (lefthanded or righthanded). Those data could lead to some correlations depending on the age or sex and depending on the stronger hand there might be some different patterns in mouse movement behavior or maybe even in usage of keyboard. But this hypothesis is just a second grade hypothesis, we are more interested in behavior independent of those data.

Once user is logged applications starts specific keylogger and mouselogger. From this point all the user activity on keyboard and mouse is recorded. It was mentioned earlier that specific keylogger is considered to invade privacy, but during the conduction of this experiment no privacy is invaded, because user will not be doing any private work, just the set of predesigned tasks on the laboratory computer.

Task Cycles

After successful log in and initiation of data loggers, user has to complete the three task cycles. Each cycle is composed of emotion induction and activity tasks (copy and navigation) see figure 3.6. Activity task are the same for each cycle, in that manner the user will perform the same activity under different conditions. Differences in performance of same task in each cycle should be influenced by the effects of the emotional states.

Emotion induction varies in each cycle. Instead of inducing the desired emotion before the task solving, elicitation will be continuous, because emotions tend to wear off quickly. Using visual and auditive techniques emotional states can be stimulated. Visual stimulus is created changing background color of the task pages, because colors are easily perceived and directly animate the brain. Every color provides unique effect and puts the observer in desired state. Auditive stimulus operates in similar way, depending on the sounds and music being played people demonstrate changes in feeling states.

Attention: During the first cycle user will be put in higher attention state. Since this is the first task he/she will be doing we already assume that user is already relaxed and attentive. For visual stimulation blue color is chosen. It is considered beneficial to the mind and body and it slows human metabolism and produces a calming effect. Classical music affects the brain and helps the listener to compete with problems with a higher attention. During the first cycle Beethoven's bagatelle in A minor (Für Elise) will be playing. Before solving the activity tasks, user will be solving an induction task; the easy version of Strimko game. While successfully solving cognitive problems person's brain is stimulated and attention is rased.

Stress: Red is a very emotionally intense color. It enhances human metabolism, increases respiration rate, and raises blood pressure. These features

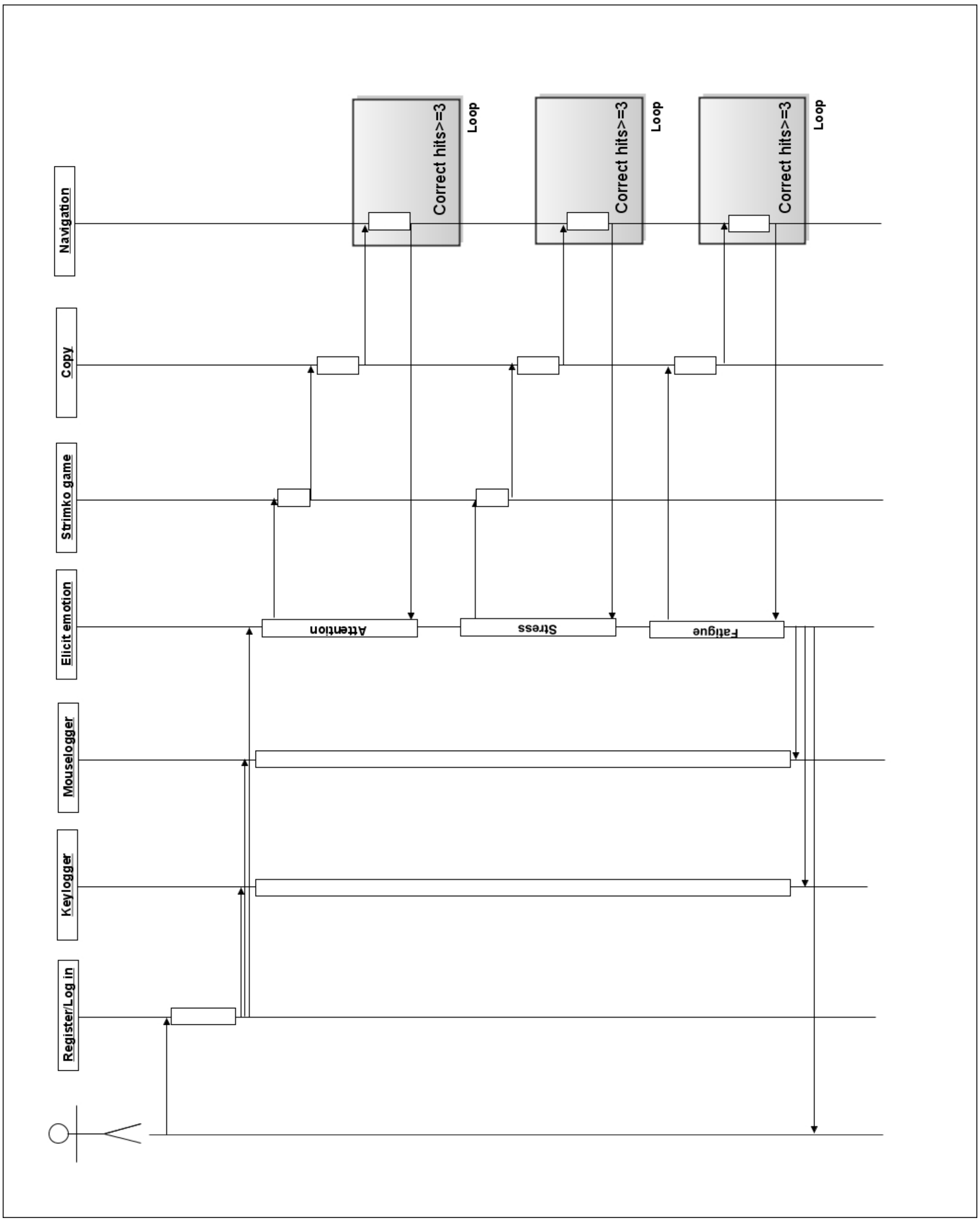

Figure 3.6: Emotion tracker sequence diagram

are associated with stress, there fore the page background color for the second cycle is chosen to be red. High-pitched sound is considered as a common stressor. During the second cycle such a sound will continuously be playing in the background, provoking user and forcing him to make mistakes. Also a time constrain is put for each task, user is timed to complete each task under 60 seconds. Time constrains when solving cognitive problems resuts in... With the above mentioned stressors the cycle includes an induction task, it is again a Strimko game, but the hard version. It is more complex and more difficult to solve and it is also time constrained. It is virtually impossible to solve it in 60 seconds.

Fatigue: Last cycle induces the fatigue. There are very few stimulus that could quickly make user to feel tired. One of the inductors is repetition, therefore the fatigue cycle is the last because it has the same activity tasks that were seen in previous cycles. Also a sideeffect of stress is that people tend to feel tired after stressful situations, therefore this cycle is immediately after the stress cycle. According to psychology of colors, brown is a calming color, mostly neutral. We used this color because it doesn't actually provoke people to feel tired, but it is not stimulative, thus it won't induce any other unwanted feeling state. In this cycle there is no induction task, from the same reason, not to stimulate brain in any way.

To verify wether the emotional states were properly induced we decided to use ProComp Infiniti. It is a 8 channel, multi-modality encoder that has ability for real-time, computerized biofeedback and data acquisition of the physiological signals. Connecting users to ProComp Infinti, we could record their physiological signals and compare them with results from the research by [22, 26, 23]. In that way we can identify emotional state of the user and verify how good the induction was.

3.3.1 Experiment problems

Experiment 1 was never executed. When approaching finishing stages of development we had some doubts about the quality of the experiment and the possibility of getting desired results. First concern was about the repetitive nature of tasks. The copy task is in every cycle and in every session the same. Also Strimko game, we didn't find the formula for creating dynamical puzzle, but we used the same template. Minimally attentive user could quickly memorize the solution pattern and/or memorize the text in the copy

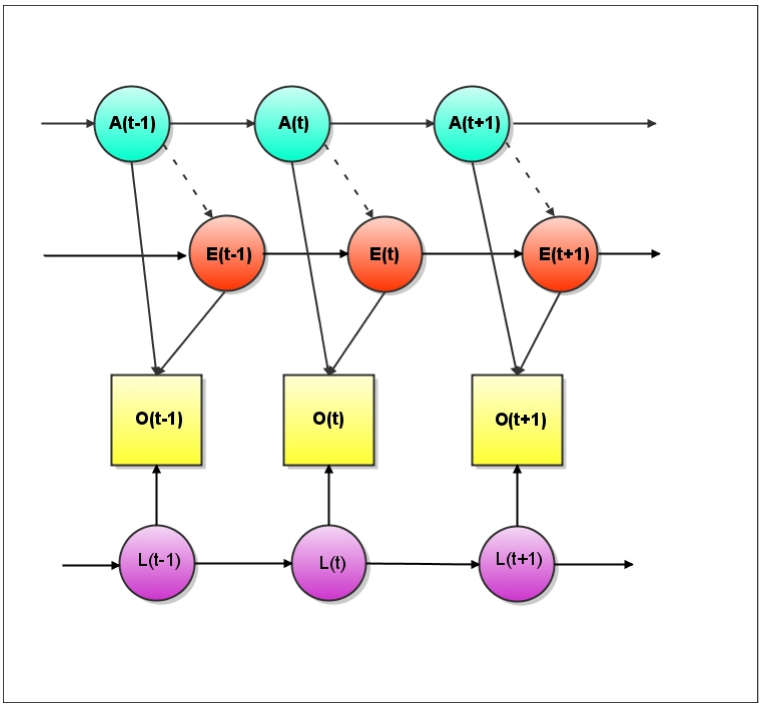

Figure 3.7: Learning affecting observation

task. In our suggested behavior model we forgot to include the user learning rate, that could have a significant impact on observed variables, see figure 3.7.

Further reflection of the proposed method introduced doubts in the quality of elicited emotions. The elicitation is artificial, and possibly rare to encounter during normal interaction on the computer. Besides the change in emotional states is too sudden and in short period of time user is forced to change several, very distinct and diverse emotions. This is no natural behavior and it might result in providing us with nonconclusive data.

3.4 Experiment 2

After abandoning the experiment 1 we designed a new method that surpassed the lacks and drawbacks found in the earlier method. Main difference in the experiment 2 is dropping of the artificial emotion elicitation and concentration on just two emotional states: neutral and tired.

Without emotion elicitation or any invasive and secure method of monitoring the user level of fatigue we constructed a new method on a new

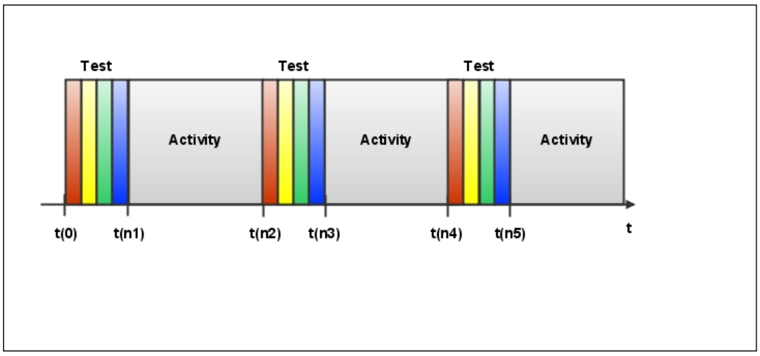

Figure 3.8: Session diagram

hypothesis. Since fatigue is usually result of long, concentrated work with minimal time spent on pauses, we decided that we should monitor users during the whole work day on their computers. It will be assumed that user starts the session on the beginning of day and works up to 8 hours. As time progresses user should become more and more tired and the fatigue should result in his performance.

This experiment operates on the same behavior model defined in the beginning of the chapter (figure 3.2). Emotion variables are hidden and the scope of the research is to determine them analyzing Activity and Observation variables. On the diagram in figure 3.8 it is demonstrated the time distribution of user free activity and test phases that user is asked to complete. User free activity in this scenario is the Activity variable. During activity phase user is doing his personal work. The applications and the kind of work will remain hidden, but using the whole-day keylogger and mouselogger it will be recorded the usage of keyboard and mouse. Data gathered with those two loggers will serve to describe the workload user has done during the activity phase. Features of interest are: amount of keys being pressed during one activity phase, length of the path mouse pointer passed, time distribution of keyboard and mouse activity. Using those features it will be gained partial observation of activity variable. On the other hand the Observation variable will be constructed during the test phase, during which user will be solving keygame, copy, mousegame and navigation tasks. Observation is guaranteed because tasks are predesigned and consulting the appropriate tables in the data base it is ensured to have an overview of what exactly was expected of user to do. Mouselogger and specific keylogger will record the data from input devices during the test phase.

3.4.1 Modification of Application Elements

Formalization: It is not an actual application element, these are definitions of terms session and test used in this experiment. Session is the time period of the data and feature collection during the run time of Emotion Tracker. A session is considered valid if it's duration is longer than 4 hours, shorter than that is not enough to perceive fatigue effects. Test is the subset of session, it is the phase during which user is solving the four tasks. The test is not complete if a task is not completed or when there are some data missing.

Database: In experiment 1 database was stored on the localhost server on the computer that ran the emotion tracker application. Being on the localhost server the direct access to the database was enabled. Because of the intention of experiment 2 to have users operating from their own computers at home or in the office, database needed to be transferred on a public server to make data collection possible. This change introduced some problems for Emotion Tracker application because the direct access to database is not allowed for external applications. The resolution of those problems is described under User registration and login and Logger data storing.

Emotion Tracker: In experiment 2 Emotion Tracker (ET) runs for 8 hours and 15 minutes. On the start of the session it triggers whole-day logger and mouse logger as a background processes in order to collect user activity data. Whole-day logger is required to keep hidden user's personal data. Every hour from the beginning of session ET opens computer's default web browser and invokes the test, simultaneously it runs the specific keylogger to record the keyboard data during the test.

Copy Task: To avoid memorizing the text user has to copy, it changes every time user starts this task. In the data base's table Text there is a collection of 20 different text. Each time another text is chosen.

User registration and login: Official registration and login is dropped, because of ET's disabled direct access to the data base. Therefore it is achieved in indirect manner. Since we assume that every user will be working each time on his own computer ET will read it's local area network adapter's MAC address, because it is unique for every computer. When ever ET starts the test it passed MAC address as a

PHP argument of the test page URL. That way all the task data will be correctly stored in the data base.

Logger data storing: Except the problem of accessing data base on public server, it still remained the ethical problem of user consenting to provide us with his personal data no matter how hidden they are. To resolve those problems all loggers will write the data in a temporary folder on user's hard disk. When ET terminates it will open a web page with upload interface, user will be instructed to upload requested logs and give us his/her final agreement to authorize us to freely use his/her data.

3.4.2 X server

To cover bigger number of users we developed a Emotion Tracker application for systems under X server (namely Linux and Mac OS). The application in general works the same way as the one designed for MS Windows operation system. Applications on user computer communicate with the X server using the classical client-server protocol. Every event raised in the client application is exchanged over a network channel.

The ET application is developed in C language with support of *Xlib*, X server library. By default creating an window application under X server event handling is required. There is a big number of events; keypressed, keyreleased, mouse movement, focus in, focus out, resize, etc. When ever an event occurs inside defined borders of the window the event is put in window buffer and inside the application the event is handled or ignored.

Additionally to *Xlib* the *Xrecord* extension library is needed. Using it's methods it is supported the recording and reporting of all core X protocol and arbitrary X extension protocol without interrupting functionality of any program. The design approach of the extension is to record core X protocol and arbitrary X extension protocol entirely within the X server itself. When the extension has been requested to record specific protocol by one or more recording clients, the protocol data is formatted and returned to the recording clients. The extension provides a mechanism for capturing all events, including input device events that do not go to any clients.

The keylogger and mouselogger were united under same process, the recording client. *XRecordRange* defines what events, out of the list of all possible events, we want to monitor . Events of our interest are *KeyPress* and *KeyRelease* events (for the keyboard), *ButtonPress* and *ButtonRelease* (for mouse buttons) and *MotionNotify* for mouse movement. The function *XRecordClientSpec* can define what clients on the server we would like to monitor. Since we are interested in all what is going on the display we assigned *XRecordAllClents* value. In that way we keep count of all events happening, even those events that are not going to any client.

To handle the events a dispatch function has to be declared. The arguments of dispatch function are predefined. For our needs it is enough the *XRecordDatum*. It is a large structure composed of all information of all events in X systems. Depending on the type of the event that arrived in que the data describing it will be found inside *XRecordDatum* structure, all other fields will have value null. For keyboard events we gather the data of the key that has been pressed or released the unix timestamp when it occurred. For mouse button events we do the same thing and for mouse movement we sign x and y coordinates and also the timestamp of the event occurrence.

All those data are stored in the list. From time to time, when the list becomes too big i.e. when number of elements reaches a threshold, the list is cleared and the events are written on the hard drive in text log files.

Additionally to the application it has been added the windowlogger, another logging system that keeps track of the active clients on the screen. This logger will make the Activity variable completely observable, because we will have the insight of what kind of applications user is running in every moment. The logger is prepaid on InFocus events, which activates when the focus on the screen changes. Every window on the screen has it's unique id number, so when the id number changes the logger searches the window name and notes it down together with the timestamp when the change occurred.

This application ends upon expiring 8 hours and 15 minutes from the start, or when user manually closes the application. When that happens some editing of the log files is needed. This is again caused by the privacy policy. The keylog file has again all the data collected from the keyboard along with the exact codes of the keys being pressed throughout the day. Just before exiting application changes all the character codes with code 101 and all number codes with 102. And it creates keylogspec file, because of the need to have the exact characters and numbers that have been pressed during the tests. Knowing the timestamps of start and end of every test application just extracts form the original log data occurring between start and end of tests.

When the editing of log files is done, application opens browser on the upload page asking user to upload log files onto our server.

3.5 Experiment results

We gathered 13 interested users who agreed to allow us insight of they personal data collected using keylogger and mouselogger. All users confirmed that they spend most of their day in front of computer and are willing to participate in the experiment. They were supposed to run 7 session each in course of two weeks. However their enthusiasm didn't last for long. Out of 13 users only 4 have ever started Emotion Tracker application. One user completed one whole session in duration of 8 hours. One user provided us with a 4 hours long session and remaining two users exited the session after 2 hours. Data collection campaign resulted with practically no data to analyze. Incomplete sessions, having duration less then 5 hours, are of no use in our analysis, because we expected the biggest accumulation of fatigue in last few hours of the 8-hour-long work.

Chapter 4

Data and Features

Since users haven't provided us significant data, we weren't able to perform data and feature analysis and draw some substantial conclusions. However during the testing phase of the ET application we recorded data from 5 complete sessions of my work on the computer. Using these data we created a set of preliminary analysis to get an overview of promising features and confirmation whether the observations are in accordance with our initial hypothesis.

In this chapter we will present the data that were collected during the session and define features of interest. Later we will exhibit analysis results of the 5 sessions. The analysis had too little data to offer some significant results, it is just a qualitative analysis that gave us preliminary confirmation of our hypothesis, namely whether the features of interest are behaving in the way we previewed.

4.0.1 Keylog features

In course of one session keylogger accumulates up to 15.000 rows of data. One row consist of keypressed unix timestamp, keypressed microsecond timestamp, keyreleased unix timestamp, keyreleased microsecond timestamp and the key code. Features are the same for whole-day keylogger and specific keylogger. Whole-day keylogger features are analyzed during the activity phase for the partial observation of Activity variable, while specific keylogger features are analyzed during the test phase for the observation of Observation variable.

Writing speed Term speed is not technically correct, because it is just time difference between two keypressed events. It is better concept than calculating exact speed (number of keys per second) because occasionally it happens that two keys have been pressed simultaneously (virtually simultaneously) then the time difference is 0 and we would face the problem of dividing by zero. The speed is calculated from the keylog for every two consecutive keys being pressed.

Writing acceleration Acceleration is defined as a difference of speed over time. For this purpose the speed is expressed in formal way. The problem of division by zero is escaped by defining a maximal acceleration and putting that value when a division by zero occurs.

Pressure duration Time duration a key has been hold down. When tired user reflexes and movements tend to slow down, we expect to observe longer key pressure duration when user is tired.

Special keys frequency Besides characters and numbers on the keyboard are to be found other function ar control keys. Recording the frequency of their usage might be relevant, especially the usage of backspace and delete key: the more user is getting tired the more errors is writing occur, to cancel errors the two keys are usually used.

Mouse button activity Keylogger also records the mouse button press and release events the same way as for keyboard keys. Mouse button features are the same as the keyboard features.

Key amount To express user workload during the session we keep count of how much keys did he/she pressed during some period of time. There are several time periods of our interest, depending on the analysis method. The way time moments of the period are chosen and analysis methods will be given under Analysis section.

4.1 Mouselog features

During one session mouselogger records up to 250.000 rows of data. One row consists of X and Y coordinate of mouse pointer position on user screen, unix timestamp in seconds when the position was recorded and unix timestamp in microseconds. Features extracted from mouselog will be used in analysis of Activity and Observation variables.

Movement speed For every two consecutive entries in mouselog speed is expressed as the distance between two points over time difference of their occurrence. The movement speed is expected to drop as user becomes more tired.

Movement acceleration Acceleration is calculated as the difference of speed over time, for every two consecutive speed features.

Angle For every two consecutive points mouse has passed over it is expressed the angle that the line between those points closes with the upper horizontal border of the screen.

Angle difference Angles for them selves are not so informative. More significant information we get by analyzing the difference of two consecutive angles. If the difference is close to 0 it describes the continuous motion in one direction, or graduate motion on a curve. We assume that linear or graduate motion if more present when user is attentive and calm. When angle difference is bigger than 0 then it exhibits the sudden, sharp change in movement direction.

Path length A measure of the workload done with mouse is the length of the path that has been crossed. Workload is defined for a time period we we analyze the user behavior. As for the Key amount the method of choosing the the interest time period will be given under Analysis section.

4.2 Pause features

During the activity phase, user is working in his/her own manner. Besides the work on the computer user can do something else, not on the computer or he/she can take a pause to relax. We are able to monitor only when user is active on computer, the lack of activity we will consider as taking a pause, a period when he/she is not engaged in any mental or physical activity, but he/she relaxes and thus lowering the level of fatigue.

We defined a pause as a time period longer than 60 seconds during which there is no occurrence of key or mouse events. Pause periods are found by examining keylog and mouse log. For each log pause periods are found by comparing time distances between consecutive events, when the time distance is longer than 60 seconds it is noted as a pause recording the pause begin timestamp and pause end timestamp. This gives us set of keyboard pauses and set of mouse movement pauses, the two sets are intersected and the resulting set is the set of computer inactivity.

For each activity phase in session pause features are determined.

Number of pauses How many times did user stop interacting with input devices. Having more pauses during one activity phase suggests that user spend some time relaxing and no significant change in the observation variables should be noted.

Average pause duration Number of pauses during one activity phase is not enough to estimate their effect on the observations. It is more informative to know how long those pauses were. Longer pauses have more relaxing effect.

Maximal pause Additional description of the pauses. One big pause in combination with several shorter pauses could have mislead understanding of the average pause duration

4.3 Task features

Task features together with keylog and mouselog features enable the complete observation of the task phase. Task features are specific for each task in the experiment and measure user's performance. Better task performance is expected in the beginning of the session, and lower results approaching the end of the session.

4.3.1 Keygame features

In the keygame task features are: the number of words that user has written correctly, and the score gained in the game.

The game generates 60 words of variable length of 3 to 6 characters. Users is challenged to write in the textbox as much words as possible. In this task we measure the performance of writing the unfamiliar random strings. In chapter 2, section Typewriting skills is described the general performance of writing this kind of stings.

For every correct word a counter is increased.

The score is worked out in this way: when user writes correctly the word, he receives the points in amount that equals to the product of the word length and the speed of the word. If the word drops on the floor user looses length of the word times 15 and divided by the speed. 15 is chosen because for normalization, it is the maximal speed of the falling word. The score might be a good indicator of user's concentration, we expected higher scores on the beginning of the session than on the end of the session.

4.3.2 Copy features

In the copying task user is rewriting an unknown sensefull text. Sensefull text writing has some known features (chapter 2, section Typewriting skills), those features were taken in controlled environment without taking emotional state in consideration. Besides comparing the performance of this task on the beginning of the session (when we expect the level of fatigue to be minimal) and on the end of the session (when we expect the level of fatigue to be maximal), we will compare the results with those from [35] and [8].

Text writing speed Since this task engages the text copying the speed will be defined in standardized unit; word per second. It is defined as a time needed for writing 5 characters. In this way the speed diagram is more smooth and less noisy.

Text writing acceleration The difference in text writing speed over time.

Total writing error This error is a result of comparison of the original text and the text user produced. Comparison is performed using the longest common subsequence algorithm. The algorithm produces result in terms of the longest possible string that was matched between the two input string. The total error therefore is the length of the original text minus the length of the longest common subsequence.

Effective writing error The total writing error is a static value obtained in off line analysis. The effective error is the on line analysis of the user's input, it takes into account the mistakes user committed during writing, but corrected them before submitting the final solution. The effective error is always greater or equal to total error because it takes all possible errors into account.

Typing nature Keyboard is not only used for the textual input, but it is used for navigation in certain applications, or user may use some control keys while operating. Since text typing requires more attention than other kinds of keyboard usage we recognize two typical typing behaviors: text typing and key stroking. We defined text typing as a continuous sequence without the pause in typing longer than 20 seconds when at least one word has been written (5 characters), anything other than that will be classified as key stroking.

4.3.3 Mousegame features

Score The number of correctly clicked targets.

Path ratio The shortest distance between the mouse pointer and the center of the correct target is the straight line between the pointer and the center of the target. During the movement towards the target user won't cross perfectly straight, because the hand is never perfectly still. When tired user will exhibit more trammer than when attentive and fresh. Also when not concentrated user might start going in the wrong direction what will result in increased path. The ratio between the path user traveled with the mouse and the shortest distance between is calculated for the normalization.

Precision When user clicks on the correct target a distance between the clicked point and the target's center is calculated. The smaller the distance is the user is evaluated to be more precise. We expect higher precision when user is in attentive state.

Reaction speed The speed is evaluated as the total distance user crossed with the mouse over the time the set of three targets appeared until he/she clicked on the target.

4.3.4 Navigation features

Number of cycles Navigation task doesn't have the predetermined number of cycles, it loops until user clicks three times on the right link.

Search duration Time is measured from the moment the task page is loaded in the browser.

Path length The navigation task page doesn't fit on one screen, it is a rather long page and requires scrolling up and down to find the requested link. Therefore links are divided in four categories depending on their distance from the top of the page. Further from the page they are they get placed in the higher category. The path user crosses with the pointer is normalized over the mean distance of the link's category.

4.4 Preliminary Analysis and Results

Given that no user have provided any data there are no results to be analyzed or consulted to draw conclusions. The miss was probably that for the experiment it is expected from users to spend 8 hours in front of the computer what can be stressful, even though the candidates we found, confirmed that due to the projects and work they spend most of the day working on their computers (many of them said even more than 8 hours). The compassion for the fellow student was just not such a stimulus.

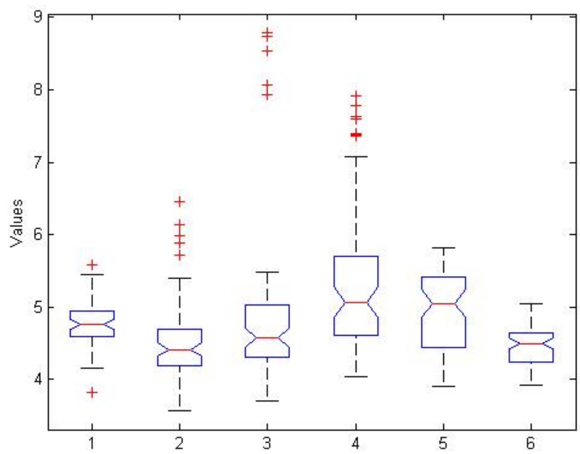

Figure 4.1: Anova results for writing speed

However we did have some data, five sessions of my own that were gathered during the testing phase of the experiment. Those data are probably influenced by the expectations of the results and knowledge of the tasks, so they cannot be credentials for the conclusive analysis. Never the less some analysis was performed to verify the plausibility of our hypothesis and serve as the guideline for the real data.

4.4.1 Anova Analysis

Fist set of analysis was conducted using anova one-way analysis of the writing speed variance during the keygame over one session. As we have expected during the time the writing speed is getting reduced, best visible on figure 4.2 for the tests 4 to 7. Means of the writing speed tend to fall down when user is continually working on the computer, but the data are not statistically independent. However these are just some first results that gave us confirmation that we are looking in the right direction.

Another interesting observation was found, in figure 4.1 between tests 3 and 4 and in the figure 4.2 between test 3 and 4 there is a visible jump of the mean value of the writing speed. In both cases user's activity stopped, because of a lunch brake. This is another plausible proof that resting reflects good on the work performance, and it could be useful in the future work to take a better look on the correlations between pauses and the user's typing speed.

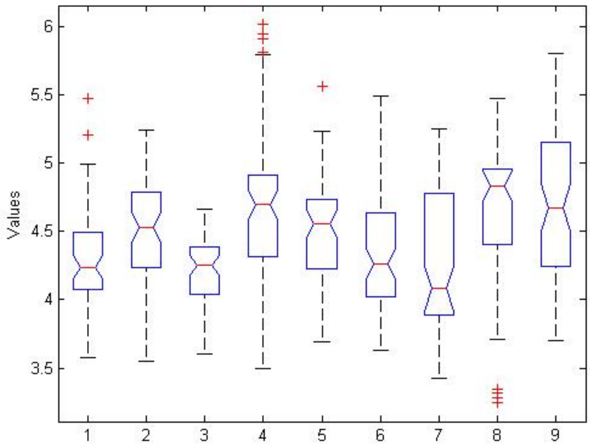

Figure 4.2: Anova results for writing speed

Next similar observations were found analyzing the pressure duration of the keyboard keys. As the time and workload progresses it is noted that key pressure duration is rising. User tend to hold the key down longer as he gets tired. The best example is shown in figure 4.3 from test 1 to test 4, and then from test 5 to the end of the session. The mean of the key pressure duration is constantly rising, and similar as with the writing speed, after the test 4 user made an hour long pause for lunch and then came back to the computer. This pause resulted in immediate drop of the pressure duration, and then, continuing to work, it started rising again.

4.4.2 Linear regression

After reaching first confirmations from the anova results, the idea came to mind if there is some possibility to model some kind of *work formula*, an empirical linear model that would take in consideration workload and rest to get from it some evaluation on the future observations. For this analysis we made a new assumption. Since we are monitoring the fatigue caused by the work on computer, then the Activity variable directly influences Emotion and Observation variables. But Emotion variable is not influenced by external stimulus, so we will drop the correlation between Emotion and Observation, since they are both dependant directly on Activity.

The Activity variable is partially observable using keylog, mouselog and pause features. From those features we describe a workload for the entire

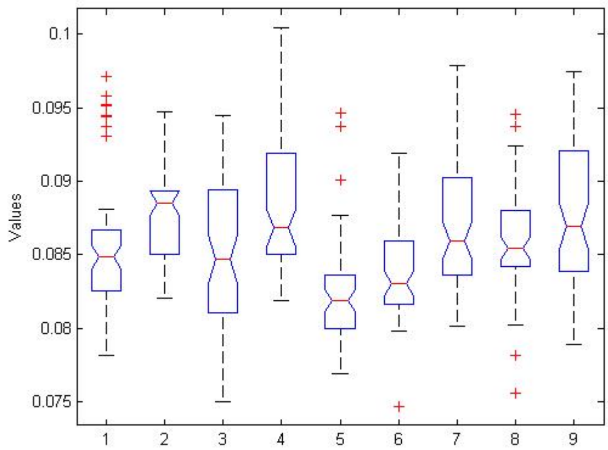

Figure 4.3: Anova results for key pressure duration

session. Workload includes the amount of keys being pressed, the length crossed with the mouse pointer and pauses.We modeled two methods of Activity description: the absolute activity method and the differential activity method.

Absolute activity This method takes into account the whole workload done from the session beginning until the observation time, because the whole history of the day's work contributes to the user performance in the test phase.

Differential activity Assuming that in a certain point we established the level of user fatigue, then the additional workload will be responsible of the difference in performance between two observations. Therefore this method takes into account the workload done between two observations and it doesn't look for the correlation to absolute values of observed variables, but their difference.

The idea was to express fatigue as a linear function dependant on active time, pause time, written amount and length of mouse movement. The very basic behavior we expected is that activity time, write amount and movement length linearly increase to the fatigue level, while the pause time is linearly decreasing the fatigue level. Increase and decrease of fatigue level will be exhibited in observation variables.

Using linear regression on the session data we tried to get some empirical values of the coefficients of the aforementioned variables. Regressor was based on the solving of the systems of the linear equations. The system's matrix was build from the observations of the keyboard and mouse behavior during one session. Number of rows was equal to the number of tests during one session. Every row of the matrix was composed of the 4 workload variables. The measured variable of the linear system is chosen to be mean value of every feature observed during the test phase. The ability to change the system's measured variable is very important, in this way we can have a large number of research and find if there is some particular feature that shows linear dependency.

Here is one example of the linear regression analysis on the typing speed during the copy task, using the absolute activity method. The values seen in the figure 4.4 we got in 4 out of 6 sessions, although diversely scaled. These results didn't offer any direct conclusion, even though they are repeated. They were didn't confirm our hypothesis that; values of coefficients for activity duration, key amount and movement length were expected to be positive, because those variables should elicit fatigue thus reducing typing speed. On the other hand value of coefficient for pause duration was expected to be negative, since pause should reduce the signs of fatigue, in this case raise typing speed. Result values were positive for activity duration and negative for key amount, movement length and pauses.

Again the problem of this analysis is the lack of data, if there were enough data we would have considered a normalization scale, because some of the values differ significantly for order of magnitude. More probably the lack of success of this method is the assumption that variables are linearly dependant on workload variables.

Analysis based on differential activity method was also performed, but it didn't provide any coherent data. Coefficient values were very different for each session, not even seemingly similar.

4.4.3 Rank Correlation Analysis

In the previous analysis keylog and mouselog features of task phase were defined as mean values of collected data. Even though the task are short and require little time to be solved, the data are very noisy, thus defining features as mean value is not very informative. It was needed to preprocess keylog and mouselog data in other way to obtain a more informative observation values. It was decided to preprocess data using power spectral density (PSD) using Welch's method. It provides much fluent description of the features in terms of frequency distributions. In this way the obtaining results might be more significant and point out the key frequencies.

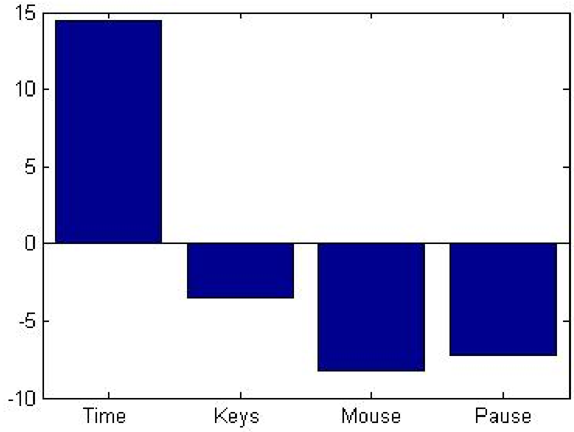

Figure 4.4: Linear regression coefficients

Matlab implementation of Welch's method allows to freely change the parameters; the scanning window size, the overlapping rate and the length of fast Furier transform. Values of those parameters we have chosen empirically, after experimenting with random values until we obtained the smoothest spectrum.

Not all keylog and mouselog features were considered for analysis, just mouse speed, mouse acceleration, mouse angle difference, typing speed and key pressure duration. They have been chosen because they are the best observed elements of the Observation variable. After applying the PSD analysis we obtained 65 elements long vector for each feature.

The rank correlation analysis is calculated for the set of PSD features and features of Activity variable. Latter features were not preprocessed because, even though they are noisy, they were collected over much longer period of time (about an hour) so the size of the resulting vector should be much bigger than 65 elements what would bring additional confusion in analysis. The correlation is executed for diverse sets of activity features separately, mostly to have a better overview of visual representation of results.

Figure 4.5 shows the correlation values between the pause features during activity phase and PSD features of each task. A high value correlation between the mean value of pauses duration (value 3 of the y axis) and key pressure duration (last 65 values on x-axis) is noted. It goes up to 0.78. Remanning correlations are not reaching significant values.

In the figure 4.6 are reported correlation values of writing features during activity phase and task features. Correlation values are not so high

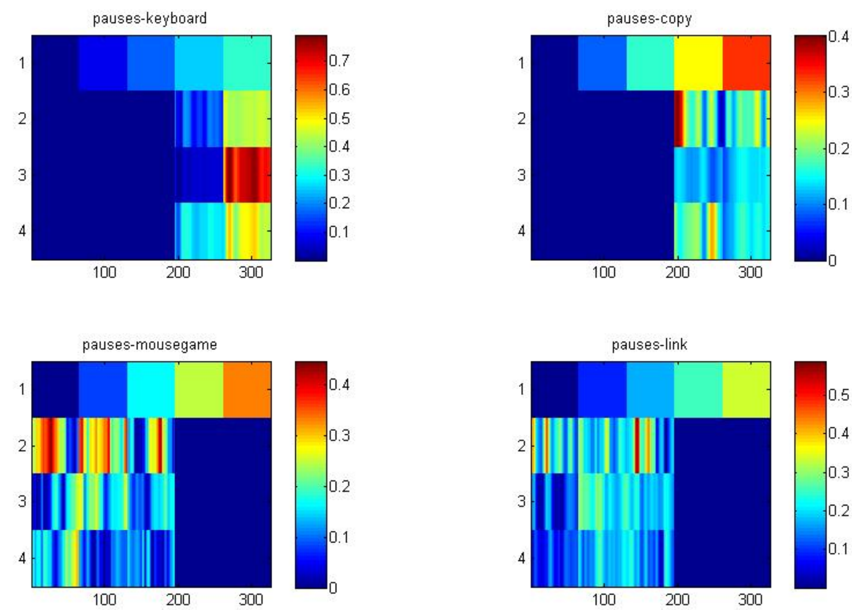

Figure 4.5: Pauses - tasks rank correlation

for key based tasks, but it is interesting that PSD features of key pressure and almost all writing features result in above average correlation. Another interesting remark is that writing features show some relatively high correlation with mouse oriented tasks. We didn't expect such high results between keyboard and mouse features.

Further analysis of such interesting results is needed. Also it might be a coincidence of some sort because there we performed this analysis on rather small amount of data.

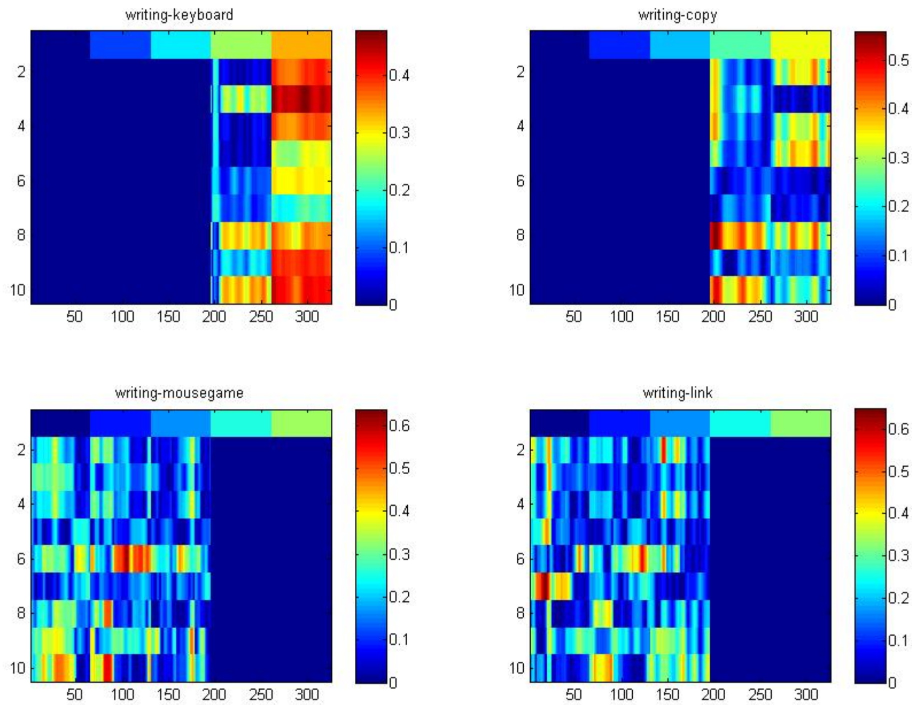

Figure 4.6: Writing - tasks rank correlation

Bibliography

[1] Takeuchi A. and Nagao K. Communicative facial displays as a new conversational modality, 1993.

[2] Gordon Allport. Becoming: Basic considerations for a psychology of personality, 1955.

[3] Orly Benchetrit and Claude Frasson. Controlling emotional conditions for learning.

[4] D'ArembeauA. Comblain C. and Van der Linden M. Phenomenal characteristics of autobiographical memories for positive, negative and neutral events, 2002.

[5] Cahill and McGaugh. Trends neurosci 21, 1998.

[6] Larry Cahill. Similar neural mechaniscms for emotion-induced memory impairment and enhancement, 2003.

[7] Soumaya Chaffar and Claude Frasson. Inducing optimal emotional state for learning in intelligent tutoring systems.

[8] Wu Changxu and Liu Yili. Queing network modeling of transcription writing, 2008.

[9] Rose Colin. Accelerated learning, 1985.

[10] Goleman D. Emotional intelligence, 1995.

[11] Julien A. Deonna. Emotion, perception and perspective, 2005.

[12] Friesen W. V. & Ellsworth P. Ekman, P. What emotion categories or dimensions can observers judge from facial behavior?, 1982.

[13] P. Ekman. Strong evidence for universala in facial expressions, 1994.

[14] Strange et al., 2003.

[15] Walker et al. Using a human face in an interface, 1994.

[16] Lorne Falk. The affective interface, 1997.

[17] Alberto Franciosi. Classificazione di caratteristiche affetive di un utente che interagisce con calcolatore tramite tastiera e mouse, 2008.

[18] Cichosz J. and Slot K. Emotion recognition in speech signal using emotion-extraction binary decision trees, 2008.

[19] Dongkyu Lee Sonya S. Kwak Myungsuk Kim. Application of biological signal to the expressions of robotic emotions, 2007.

[20] R. Koutlak. Inside the brain, 1997.

[21] Oh-Wook Kwon Kwokleung Chan Jiucang Hao Te-Won Lee. Emotion recognition by speech signals, 2003.

[22] Lan Li and Ji hua Chen. Emotion recognition using physiological signals, 2006.

[23] Ruiqing Yangand Guangyuan Liu. Emotion feature selection from physiological signal based on bpso, 2006.

[24] E. McGinnies. Emotionality and perceptual defense, 1949.

[25] F. McPhersor. The role of emotion in memory, 2004.

[26] Nasoz F. Lisetti C.L. Alvarez K. Finkelstein N. Emotion recognition from physiological signals for user modeling of affect.

[27] Rui Figueriredo Ana Paiva. Watch and feel: an affective interface in virtual storytelling environment, 2003.

[28] Elisabeth Phelps. Human emotion and memory: interactions of amygdala and hippocampal complex, 2004.

[29] Rosalind Picard and Jonathan Klein. Computers that recognize and respond to user emotion: theoretical and practical implications.

[30] R.W. Picard. Affective computing: challenges.

[31] R.W. Picard. Affective computing, 1997.

[32] Wexelblat A. Picard R.W. Future interfaces: Social and emotional.

[33] R. Plutchik. A general psychoevolutionary theory of emotion, 1980.

[34] Bhavin r. Sheth Thuan Pham. Howemotional arousal and valence influence access to awareness, 2008.

[35] T.A. Salthouse. Perceptual, cognitive, and motoric aspects of transciption writing, 1986.

[36] L.J. Schinnerer. Guide to self: The beginner's guide to managing emotion and thought, 2007.

[37] S. S. Tomkins. Affect theory, 1984.

[38] Roy F. Baumeister C. Nathan DeWall Kathleen D. Vohs and Jessica L. Alquist. Does emotion cause behavior (apart from making people do stupid, destructive things).