Quantifying the effects of external factors on individual performance

Davide Carneiro[a,b], Paulo Novais[b]

[a] CIICESI, ESTGF, Polytechnic Institute of Porto, Portugal
[b] Algoritmi Center/Department of Informatics, Minho University, Braga, Portugal

Abstract

Monitoring and managing performance in the workplace is nowadays an important aspect, in a time in which methodologies like Agile push individual and team limits further. Current performance monitoring approaches are either intrusive or based on productivity measures and are thus often dreaded by workers. Moreover, these approaches do not take into account the importance and role of the numerous external factors that influence productivity. We present a non-intrusive performance monitoring environment based on behavioral biometrics and real time analytics. It monitors and analyses 15 features extracted from the workers' interaction with the computer and can provide a measure of performance that is completely transparent. This measure is sensitive to external factors such as mental fatigue, stress or emotional valence. We validate this environment by assessing the effects of musical selection on Human-Computer Interaction. Results show a significant improvement on mouse motion when participants listen to the selected auditory stimuli and a negative effect on typing performance, especially with stimuli with positive tension. This work will enable the development of performance monitoring and management environments, with benefits for both organizations and individuals.

Keywords: Performance Monitoring, Emotional Valence, Human-Computer Interaction

1. Introduction

The Human being is currently under an increasing demand for performance, fruit of a society that is moving faster everyday. Workplaces are particularly "good" examples of this reality. Lack of jobs, decreasing wages, increasing working hours, working in shifts, competitiveness or unrealistic productivity goals result in a constant and increasing pressure on the individual.

Numerous studies highlight the negative effects of this lifestyle. Sparks et al. [1] show positive mean correlations between overall health symptoms, physiological and psychological health symptoms, and hours of work while Dembe et al. [2] analyze the impact of overtime and long work hours on occupational injuries and illnesses, to conclude that these variables depend more on the amount of time worked rather than on the level of risk of the job. In [3], the effects of shift work and extended hours of work are analyzed at different levels, including family and social life, performance, fatigue, productivity and health.

As addressed in detail in [4], there is currently an overwork culture, which is further encouraged by greedy management techniques and job insecurity. While the main objective of management in doing so is to increase production, this does not necessarily happen, nor will it increase productivity.

There is thus the need to improve performance or productivity by other means that do not bring along such negative effects. In this paper we pursue this aim by developing an environment that takes as input individuals' behavioral cues. Indeed, Humans tend to show their personality or their state through their actions, whether in a conscious or unconscious way. The way individuals interact with technological devices is influenced by such factors and can thus be assessed to assess them. Some of these effects, analyzed in preliminary studies, show that our performance tends to decrease with the onset of mental fatigue [5] or that the way we touch or handle a smartphone changes with our level of stress [6].

This means that the performance of each individual in a group (and consequently of the group) can be assessed in real time and continuously, instead of being assessed on a regular basis as happens with traditional approaches (e.g. weekly, monthly, yearly). Moreover, workers are more likely to accept this kind of performance monitoring as opposed to traditional measures based on productivity indicators, which are often dreaded and may even result in decreased performance or productivity due to the pressure of monitoring [7]. The privacy of the worker is thus safeguarded.

This approach also translates into a range of other advantages. Namely, it is non-invasive and non-intrusive as it requires no specific interactions by the workers. In that sense, the approach is also transparent. We validate the proposed environment by assessing performance in real time of three groups of people. Moreover, we assess the effect of musical selection on performance.

We do this aiming at the main goal of this line of research, which is not only to monitor performance but also to manage and improve it. The fact that we pursue performance improvement through the use of music may yield additional advantages as described in section 2.1 (e.g. mood, well-being), as opposed to the culturally established use of punishments (e.g. economic, withdrawal of benefits) or intimidation (e.g. threatening to lose job, preventing career promotion), with negative side effects.

In the overall, this approach is expected to result in better, more efficient and easier to implement performance monitoring initiatives in the workplace, that take into account the worker's well-being and satisfaction.

The remaining of the paper is organized as follows. Section 2 addresses some related work on the field of Human-Computer Interaction and deals with some of the effects of music on the Human being at different levels. Section 3 describes the intelligent environment for real time performance management, including a detailed description of the interaction features considered as well as an analysis of the system's scalability. Section 4 describes in detail the experimental study carried out with the aims to: (1) validate the environment developed and (2) assess the influence of external factors on worker performance. It includes a description of the study design, a thorough characterization of the population (including an emotional and personality characterization) as well as the process followed to validate the auditory stimuli used in the study. The next section details the results of the study in what concerns Human-Computer Interaction. Finally, section 6 discusses the results achieved and summarizes the conclusions.

2. Related Work

In the last years there has been a growing interest in the use of the affective state of computer users, namely to develop applications or hardware that adapt to their users [8]. However, the interest in affective computing goes beyond the sheer acquisition of this information: there is also an interest in how knowledge about affective states can be used to improve the current state of the user. Two main challenges thus exist in this field: how to acquire this kind of information and how to act on the user state. In this paper we look at Behavioral Biometrics for the purpose of information acquisition and at music to influence the affective state.

Behavioral Biometrics defines a field that extracts user's behavioral features from the use of the mouse and the keyboard [9]. Traditional biometrics use human physical or physiological characteristics that are virtually unique for each individual, including fingerprints, iris or face recognition, palm print or veins, among others [10]. These characteristics are used mostly for the purpose of identification. Behavioral Biometrics, on the other hand, rely on behavioral traits of the individual such as typing rhythm, gait, voice, among others. While Behavioral Biometrics can also be used for identification purposes, its features are prone to change according to the inner state of the individual. For example, a stressed individual may show significant differences in speech, reducing the accuracy of identification. Nonetheless, this allows for other interesting applications, namely to assess behavioral changes. Knowing how a individual usually behaves allows to detect significant behavioral changes, which may in turn indicate changes in the inner state of the individual.

In this field, the mouse and keyboard have been used in the last years as the source of valuable inputs for behavioral patterns analysis, known respectively as mouse and keyboard dynamics. These two approaches have been consistently used in the last years for a wide range of different purposes.

Shukla et al. look at the user's typing behavior to identify emotional states [11]. The authors use a total of 8 features: session time, keystroke latency, dwell time, sequence, typing speed, frequency of error, pause rate and capitalization rate. Questionnaires were used to assess the emotional state of the participants. All this data was then used to train classifiers for human emotion recognition from the keyboard typing patterns. In a related approach, the authors of [12] analyze typing behavior against positive/negative emotions. Its main conclusion is that all participants have shown significant differences in typing patterns when under positive and negative emotions, elicited through facial feedback [13]. 15 individuals participated in the study, which only considered the valence of emotion (positive and negative) and two interaction features (keystroke duration and latency). The results support the claim that different emotional valences result in different typing behaviors. Further emotion recognition methods based on keystroke dynamics and mouse movements can be found in [14].

Typing behavior has also been used by [15] to determine the effect of music and induced mental load in a word processing task. The authors measured typing force, typing productivity, and electromyography of the left hand *extensor digitorum* muscle, concluding that overall typing productivity was compromised by music while also observing a reduction of wrong finger touch during typing. Music also resulted in an increased *extensor digitorum* muscle activity for lifting and controlling fingers. Nonetheless, only 8 individuals participated in this study, rendering these results rather limited.

Behavioral Biometrics have also been used for the purpose of user identification. Several research works can be pointed out that Mouse Dynamics for this specific purpose. Both holistic features (single-click statistics, double-click statistics, movement offset and movement elapsed time) and procedural features (speed curve against time and acceleration curve against time) to characterize mouse movement are used in [16]. The authors conducted a study with 37 participants, in which satisfying acceptance rate were obtained with only 11.8 seconds of interaction. Similarly, [17] use 25 participants and 5 features that model clicking rhythm, which quantify different timings between clicks and during clicks. While the previous work used mostly mouse movement, this one uses mouse clicking alone. They could thus be used in conjunction, in an attempt to develop a more precise approach. Other approaches have been analyzed by [18], who review existing authentication approaches based on mouse dynamics and shed light on some important limitations regarding how the effectiveness of these approaches has been evaluated in the past. The authors also present the results of several experiments conducted by them to illustrate their observations and suggest guidelines for evaluating future authentication approaches based on mouse dynamics.

Behavioral Biometric techniques have also been used for assessing people's level of stress and mental fatigue. In the last years we have been studying the interaction patterns of people with computers and smartphones, to build models that can be used in real time for classifying stress and fatigue [19, 6]. The aim is to develop software and hardware that is sensitive to the user's state, adapting accordingly. Other authors have also looked at Behavioral Biometrics for similar purposes. Vizer et al. [20] are able to classify cognitive and physical stress with accuracy rates comparable to those currently obtained using affective computing methods, using keystroke and linguistic features of spontaneously generated text. A case-based approach relying on Behavioural Biometrics is used by [21] to determine a user's stress level. On a somehow related approach, the authors of [22] use keystroke analysis to detect boredom and engagement during writing.

2.1. The Effects of Music

Music has been studied thoroughly in the last years due to its benefits at many levels. One of the most well studied domains is the one of healthcare. Specifically, music has been shown to have positive effects on pain management, namely postoperative pain, as well as on anxiety, depression and disabilities, allowing patients to use music to enhance the effects of analgesics and other drugs, promoting overall feelings of power [23]. Other studies show that intraoperative music may decrease postoperative pain, and that postoperative music therapy may reduce anxiety, pain and morphine consumption [24, 25]. Other effects in health include reducing blood pressure [26], speeding up post-stroke recovery [27] or increasing postpartum well-being [28].

Still concerning well-being, some notable effects of music are in relaxing, inducing sleep and reducing stress. In fact, most of us have, at a certain point in our lives, listened to some type of music with the purpose of relaxing. Studies show that music is one way to beat insomnia, especially if listened to before bedtime [29]. Music also significantly reduces stress on people [30], namely in people working in particularly stressful jobs. Finally, music has also many interesting psychotherapeutic benefits, with an overall positive effect on mood and is even used to address conditions such as depression [31].

Certain types of music, especially classical music, have also been shown to improve higher functions of the brain. Nonetheless, it was also noted that any kind of music can have a similar effect so-long it is enjoyable by the listener [32]. Liking the music is, therefore, one major aspect when seeking positive effects from its listening. Among many other effects, major higher cognitive functions positively affected by music include reading and literacy skills [33], working memory and mathematical abilities [34], memory [35] or concentration and attention [36].

Given the scope of this line of work, we are more interested in addressing the effects of music on performance and productivity. Music has been shown to improve physical performance, namely of athletes, who perform better while exercising to music [37, 38]. Improvements happen at several levels including physical fatigue, motor coordination, motivation and endurance. While part of these effects come from motivational factors associated to certain lyrics, the other undoubtedly comes from the emotional factors associated to the music itself.

In terms of cognitive performance music has also been studied thoroughly, especially regarding its effects in the workplace. Music has been shown to improve productivity and reduce fatigue, especially in cases of monotonous work [39].

This analysis of recent works in the field shows the current interest of the research community on Behavioral Biometrics as well as its possible applications. Unfortunately, the effect of music on interaction features has not been studied with the necessary detail. To the extent of our knowledge, only the previously mentioned study by [15] addresses this subject and only superficially. We believe that the effects of music on Human-Computer Interaction should be more thoroughly studied, for the following reasons:

- Music has such a significant effect in so many aspects of our lives that its potential effect on HCI is worth investigating;
- Failing to acknowledge the potential effects of music on HCI may render existing research efforts useless as listening to music while interacting with the computer may significantly affect the interaction patterns;
- Music can have a very interesting role on the development of user-awareness in Intelligent Environments with effects not only on the interaction itself (e.g. performance, productivity) but also on other important variables (e.g. well-being, stress level, memory).

In this paper we try to address this gap in research. Specifically we validate, for our population, a list of stimuli put forward by [40], and use it during a task that involves interaction with the computer using mouse and keyboard. We use stimuli with positive and negative tension and record a total of 15 interaction features that fully characterize the participant's interaction with the computer for more than one hour. We believe that this work can be important to start to understand the effects that music with different valence and arousal may have in interaction patterns. This is especially true in a time in which so many people work with a computer and listen to music while doing so.

3. An Intelligent Environment for Performance Management

In order to implement the proposed approach, an Intelligent Environment for performance management was developed. The key element in this environment is a data collection tool that registers all the events describing the user's interaction with the computer. These events support the generation of 15 interaction features, described in Section 3.1. These features, describing the interaction of each user with each device, are processed, transformed and sent to a server, which continuously builds a user interaction profile, as described in Section 3.2. The analysis of this profile and its evolution with time are what allows the detection of significant behavioral changes.

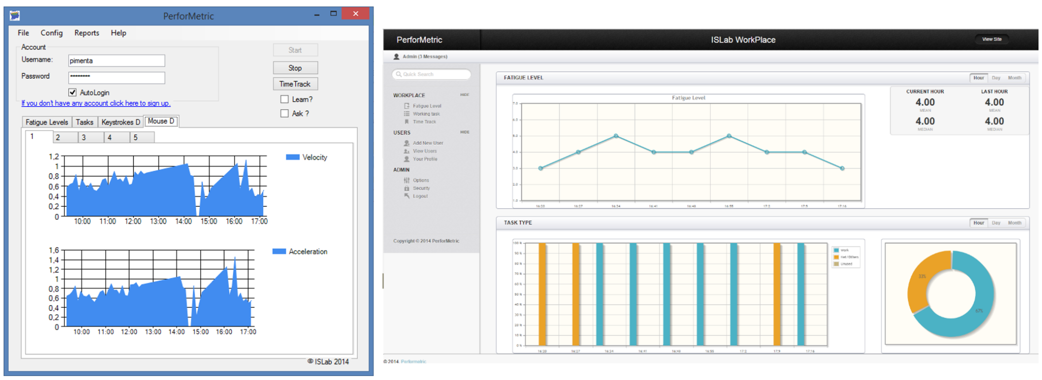

Figure 1: User Interface showing some of the raw data collected by the tool (left) and web interface showing compiled high-level information (right).

3.1. Feature Extraction

The process of feature extraction starts with the acquisition of interaction events, which is carried out by a specifically developed application that is installed in each of the computers or smartphones. This application runs in the background and requires no interaction by the user. It is thus non-intrusive. The following events are acquired by the application and sent to the server for processing:

- MOV, timestamp, posX, posY

  An event describing the movement of the mouse, in a given time, to coordinates (posX, posY) in the screen;

- MOUSE_DOWN, timestamp, [Left | Right], posX, posY

  This event describes the first half of a click (when the mouse button is pressed down), in a given time. It also describes which of the buttons was pressed (left or right) and the position of the mouse in that instant;

- MOUSE_UP, timestamp, [Left | Right], posX, posY

  An event similar to the previous one but describing the second part of the click, when the mouse button is released;

- MOUSE_WHEEL, timestamp, dif

  This event describes a mouse wheel scroll of amount dif, in a given time;

- KEY_DOWN, timestamp, key

  Identifies a given key from the keyboard being pressed down, at a given time;

- KEY_UP, timestamp, key

  Describes the release of a given key from the keyboard, in a given time.

The following example depicts a brief log that starts with some mouse movement (first two lines), contains a click with a little drag (lines 3-5) and ends with some more movement (last two lines).

```
MOV,635296941683402953,451,195
MOV,635296941684123025,451,197
MOUSE_DOWN,635296941684443057,Left,451,199
MOV,635296941685273140,452,200
MOUSE_UP,635296941685283141,Left,452,200
MOV,635296941685723185,452,203
MOV,635296941685803193,454,205
```

The individual logs build by the aforementioned application are then processed in order to compile information that can characterize the behavior of the user while interacting with the computer. This subsection details the features that are extracted from the logs of interaction events.

It is important to note that these features aim at quantifying the user' performance. Taking as example the mouse, its motion between two points is virtually never in a straight line as there is always some deviation. The larger this deviation, the less efficient the movement is. An interesting property of the features described in this section is that, except for mouse velocity and acceleration (as detailed further ahead), an increasing value denotes a decreasing performance (e.g. longer click ⇒ poorer performance, larger average excess of distance ⇒ poorer performance). These relationships with performance have been established in previously conducted research studies [19, 6].

Of the following 15 considered features, 12 are extracted from the mouse and 3 from the keyboard:

Click Duration (CD)
UNITS - milliseconds
Measures the time span between two consecutive MOUSE_DOWN and MOUSE_UP events. The longer the clicks, the less efficient the interaction is.

Time Between Clicks (TBC)
UNITS - milliseconds
The time span between two consecutive MOUSE_UP and MOUSE_DOWN events, i.e., how long did it take the user to perform another click. Similarly to other features, we consider that a smaller time between clicks is representative of a faster working rhythm, thus increased performance.

Time Double Click (TDC)
UNITS - milliseconds
The time span between two consecutive MOUSE_UP and MOUSE_DOWN events when smaller than 200 ms i.e., the duration of a double click. Similarly to other features, a shorter double click time represents increased interaction performance.

Mouse Velocity (MV)
UNITS - pixels/milliseconds
The distance traveled by the mouse (in pixels) over the time (in milliseconds). The velocity is computed for each interval defined by two consecutive MOUSE_UP and MOUSE_DOWN events. Let us assume two consecutive MOUSE_UP and MOUSE_DOWN events, *mup* and *mdo*, respectively in the coordinates $(x1, y1)$ and $(x2, y2)$, that took place respectively in the instants $time_1$ and $time_2$. Let us also assume two vectors *posx* and *posy*, of size n, holding the coordinates of the consecutive MOUSE_MOV events between *mup* and *mdo*. The velocity between the two clicks is given by $r\_dist/(time_2 - time_1)$, in which r_dist represents the distance traveled by the mouse and is given by equation 1. The relationship of this feature with performance is not as straightforward as in previous features. In fact, to a certain extent, a higher velocity may indicate increased performance. How-

ever, after a given threshold that is not true as higher velocity will result in less precision and control, which compromises performance. For that reason, we do not consider this feature for performance assessment. Nonetheless, it may still provide interesting insights about changes in the user's behavior.

Mouse Acceleration (MA)

UNITS - pixels/milliseconds^2

The velocity of the mouse (in pixels/milliseconds) over the time (in milliseconds). A value of acceleration is computed for each interval defined by two consecutive MOUSE_UP and MOUSE_DOWN events, using the intervals and data computed for the Velocity. As with mouse velocity, the relationship of mouse acceleration with performance is not straightforward. For this reason, it is analyzed to assess changes in user behavior but is not considered for the purpose of estimating performance.

Distance Between Clicks (DBC)

UNITS - pixels

Represents the total distance traveled by the mouse between two consecutive clicks, i.e., between each two consecutive MOUSE_UP and MOUSE_DOWN events. Let us assume two consecutive MOUSE_UP and MOUSE_DOWN events, *mup* and *mdo*, respectively in the coordinates $(x1, y1)$ and $(x2, y2)$. Let us also assume two vectors *posx* and *posy*, of size n, holding the coordinates of the consecutive MOUSE_MOV events between *mup* and *mdo*. The total distance traveled by the mouse is given by equation 1. A larger distance between clicks indicates a less efficient movement pattern.

$$r\_dist = \sum_{i=0}^{n-1} \sqrt{(posx_{i+1} - posx_i)^2 + (posy_{i+1} - posy_i)^2} \tag{1}$$

Excess of Distance (ED)

UNITS - pixels

This feature measures the excess of distance that the mouse traveled between each two consecutive MOUSE_UP and MOUSE_DOWN events. Let us assume two consecutive MOUSE_UP and MOUSE_DOWN events, *mup* and *mdo*, respectively in the coordinates $(x1, y1)$ and $(x2, y2)$. To compute this feature, first it is measured the distance in straight line between the coordinates of *mup* and *mdo* as $s\_dist = \sqrt{(x2 - x1)^2 + (y2 - y1)^2}$. Then, it is measured the distance actually traveled by the mouse by summing the distance between each two consecutive MOUSE_MV events. Let us assume two vectors *posx* and *posy*, of size n, holding the coordinates of the consecutive MOUSE_MV events between *mup* and *mdo*. The distance actually traveled by the mouse, *real_dist* is given by equation 1. The Excess of Distance is given by $r\_dist - s\_dist$. A larger excess of distance indicates a lower performance of interaction.

Average Excess of Distance (AED)
UNITS - pixels
This feature measures the average excess of distance that the mouse traveled between each two consecutive MOUSE_UP and MOUSE_DOWN events. The average excess of distance between the two consecutive clicks (Figure 2 (a) is given by r_dist/s_dist, with *r_dist* and *s_dist* computed similarly to the ED feature. Once again, there is an inverse relationship between this feature and performance.

Distance of the Mouse to the Straight Line (DMSL)
UNITS - pixels
This feature quantifies the sum of the successive distances of the mouse to the straight line defined by two consecutive clicks. Let us assume two consecutive MOUSE_UP and MOUSE_DOWN events, *mup* and *mdo*, respectively in the coordinates $(x1, y1)$ and $(x2, y2)$. Let us also assume two vectors *posx* and *posy*, of size n, holding the coordinates of the consecutive MOUSE_MOV events between *mup* and *mdo*. The sum of the distances between each position and the straight line defined by the points $(x1, y1)$ and $(x2, y2)$ is given by 2, in which *ptLineDist* returns the distance between the specified point and the closest point on the infinitely-extended line defined by $(x1, y1)$ and $(x2, y2)$. It results logical that the relationship of this feature with performance is inverse, i.e., the movement of the mouse is more efficient if it is moving closer to the line.

$$s\_dists = \sum_{i=0}^{n-1} ptLineDist(posx_i, posy_i) \tag{2}$$

Average Distance of the Mouse to the Straight Line (ADMSL)
UNITS - pixels
This feature is similar to the previous one in the sense that it will compute the *s_dists* between two consecutive MOUSE_UP and MOUSE_DOWN

events, *mup* and *mdo*, according to equation 2. However, it returns its average rather than its sum during the path. The relationship of this feature with performance is also similar. The average distance of the mouse to the straight (Figure 3 (b)) line defined by two consecutive clicks is thus given by s_dists/n.

Absolute Sum of Angles (ASA)
Units - degrees
This feature seeks to find how much the mouse "turned", independently of the direction to which it turned (Figure 3 (a)). In that sense, it is computed as the absolute of the value returned by function $degree(x1, y1, x2, y2, x3, y3)$, as depicted in equation 3. An efficient mouse movement is characterized by lines of movement that are almost straight. Therefore, higher values of this feature indicate a less efficient movement.

$$rCls\_angle = \sum_{i=0}^{n-2} \mid degree(posx_i, posy_i, posx_{i+1}, posy_{i+1}, posx_{i+2}, posy_{i+2}) \mid \tag{3}$$

Signed Sum of Angles (SSA)
Units - degrees
This feature is very similar to the previously mentioned one, with the exception that it measures to which side the mouse turns more. A negative value indicates that, in the overall, the mouse turns more counter-clockwise, while a positive value indicates that the mouse turns more clockwise. Concerning the relationship of this feature with interaction performance, the same considerations of the previous feature apply, i.e., the higher the value the less efficient the movement.

Time Between Keys (TBK)
Units - milliseconds
The time span between two consecutive KEY_UP and KEY_DOWN events, i.e., how long did it take the user to press another key after releasing the previous one. The lower the value of this feature the higher the performance.

Key Down Time (KDT)
Units - milliseconds

The time span between two consecutive KEY_DOWN and KEY_UP events, i.e., for how long did the user press a given key. Intuitively, the longer the key press the lower the performance of the writing.

Writing Velocity (WV)
Units - keys per minute
The time span between two consecutive KEY_UP and KEY_DOWN events, i.e., how long did it take the user to press another key. The lower the value of this feature the higher the performance.

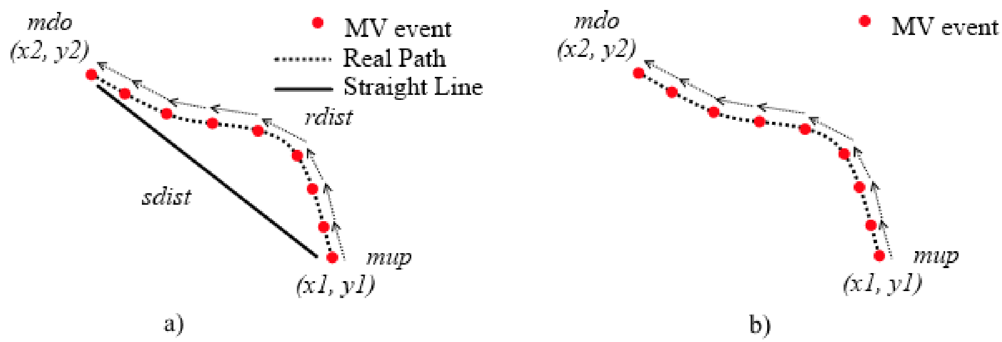

Figure 2: (a) A series of MOV events, between two consecutive clicks of the mouse. The difference between the shortest distance (sdist) and distance actually traveled by the mouse (rdist) is depicted. (b) The real distance traveled by the mouse between each two consecutive clicks is given by summing the distances between each two consecutive MOV events.

3.2. Real Time Analytics for Teams

There is nowadays an undeniable interest in collecting contextual data about people, for individual use as well as for third party objectives. One of the visible results is the number of mobile apps to monitor many different aspects of our behaviors, routines or health with aims such as stress management, fatigue management, improving well being, fall detection among others [41, 42].

In the context of an organization, the gathering and analysis of metrics describing people's behavior, and the providing of tools for visualization (particularly real time analytics) enables better decision making and data-driven actions that consider the state and well-being of each individual worker.

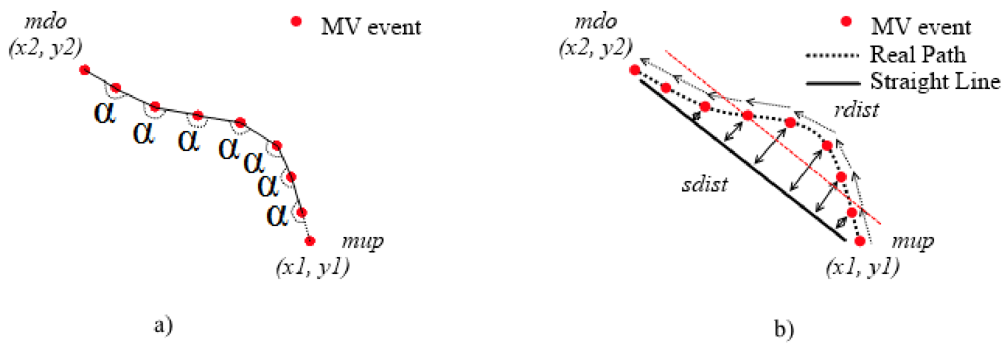

Figure 3: (a) The sum of the angles of the mouse's movement is given by summing all the angles between each two consecutive movement vectors. (b) The average distance at which the mouse is from the shortest line between two clicks is depicted by the straight dashed line.

Such initiatives can nowadays be scaled to hundreds or thousands of workers, through the use of Big Data tools and techniques, without compromising performance and availability.

The features described in Section 3.1 describe the behavior of each individual while interacting with the computer. This behavior, as many others, is affected by factors that influence performance at work, including mental fatigue, stress level or emotional arousal. Specifically, each instance of the behavior is characterized by fifteen values (represented as doubles) that are a result of applying several data summarization techniques (e.g. i.e. aggregation of collected data by calculating values such as mean and variance on the very frequently collected values). Each of these instances also contains a timestamp.

Given that this data is stored in a Mongo databae, each record needs 136 bytes of storage space: 15 times 8 bytes (the MongoDB double size) plus 8 bytes for the timestamp, and 8 bytes for the two keys that describe the application being used and the user. A new record is produced every five minutes, for each user of the environment. Assuming that each individual is expected to work around 8h per day, a production of around 12.75 Kbytes of data per worker is estimated. Table 1 shows the expected data growth projections for different numbers of users and different time-spans.

The architecture of the developed environment (Figure 4) is divided in three major parts. The lower-level is composed by the devices that generate

Table 1: Projections of data growth for different number of users and time frames.

Time \ # Users	1	100	10000	1000000
5 minutes	136 Bytes	13.28 KBs	1.297 MBs	129.7 MBs
1 day	12.75 KBs	1.245 MBs	124.5 MBs	12.159 GBs
1 week	89.25 KBs	8.716 MBs	871.6 MBs	85.115 GBs
1 month	382.5 KBs	37.354 MBs	3.648 GBs	364.8 GBs
1 year	4.545 MBs	454.5 MBs	44.382 GBs	4.438 TBs

the raw data (e.g. computers, smartphones). These devices store the raw data locally in SQLite databases, until it is sychnronized with the web servers in the cloud, which happens at regular intervals.

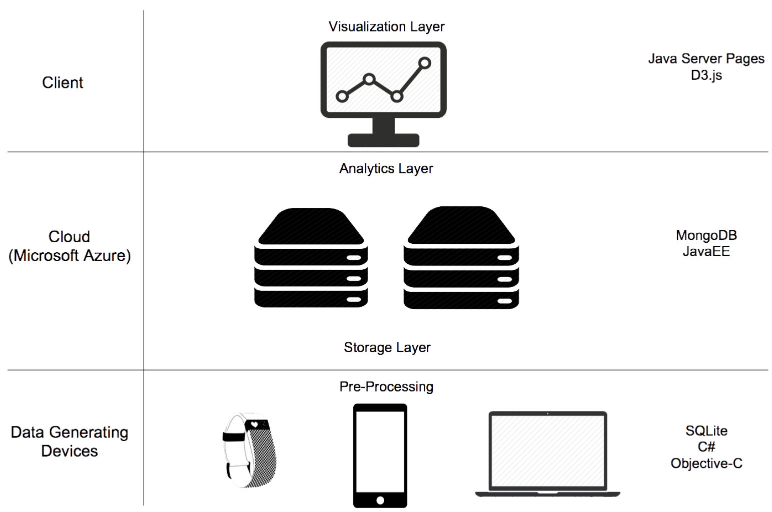

Figure 4: Layered architecture of the environment.

The main element in the middle layer is a Mongo database. MongoDB is half way between relational and non-relational systems. It provides indexes on collections, it is lockless and provides a query mechanism. MongoDB provides atomic operations on fields like relational systems MongoDB supports automatic sharding by distributing the load across many nodes with automatic failover and load balancing, on the other hand CouchDB achieves scalability through asynchronous replication. MongoDB supports replication with automatic failover and recovery. The data is stored in a binary JSON-like format called BSON that supports boolean, integer, float, date, string and binary types. The communication is made over a socket connection (in CouchDB it is made over an HTTP REST interface).

MongoDB is actually more than a data storage engine, as it also provides native data processing tools: MapReduce and the Aggregation pipeline. Both the aggregation pipeline and map-reduce can operate on a shared collection (partitioned over many machines, horizontal scaling). These are powerful tools for performing analytics and statistical analysis in real time, which is useful for ad-hoc querying, pre-aggregated reports, and more. MongoDB provides a rich set of aggregation operations that process data records and return computed results, using this operations in the data layer simplifies application code and limits resource requirements.

In what concerns fault tolerance MongoDB provides master-slave replication and replica sets. Nowadays, replica sets are recommended for most use cases. The standard (and minimum) number of replicas in a set is three: one being the primary (the only one with writes allowed), and two secondaries (can become the primary in an election), since an odd number of members ensures that the replica set is always able to elect a primary. MongoDB also provides pluggable storage engines, namely WiredTiger and MMAPv1. Multiple storage engines can co-exist within a single MongoDB replica set, making it easy to evaluate and migrate engines. Running multiple storage engines within a replica set can also simplify the process of managing the data lifecycle. WiredTiger (default storage engine starting in MongoDB 3.2) will provide significant benefits in the areas of lower storage costs, greater hardware utilization, and more predictable performance 6 and, consequently should be used in this system.

Finally, the visualization layer (top-most layer) is developed as a web app on Java technology and uses the D3 library for graphics and diagrams. It includes a set of intuitive data visualization tools to facilitate decision making and human resources management, with a focus on individual and group performance real time analytics.

4. Experimental Study

In our line of research we have been studying ways to assess HCI performance. Moreover, we have also studied how HCI performance relates to important phenomena such as mental fatigue and stress. In the present work, besides developing and presenting an environment for supporting the whole process, we also analyze the potential influence of external factors on performance. Specifically, we seek to find the potential effect of different types of music on HCI performance, with the aim of developing an Intelligent Environment for group performance management.

This environment will tendentiously be used in the workplace by team managers to, on the one hand, analyze the evolution of the performance of the group during the day or during wider time periods. It will allow to establish each individual's optimum working cycle, improving productivity and well being. Moreover, the work now carried out opens the door to the use of music as a powerful tool to manage performance. In fact, until now, in order to manage performance, the manager could only take actions such as recommending a pause or adjusting working schedules. Bringing music into the equation allows the manager to act not only in reaction (e.g. when a performance drop is detected in one worker) but also in prevention (e.g. selecting music that improves performance throughout the day).

This section describes the experimental study carried out to assess the potential effect of music on HCI. We build on previous work by Tuomas Eerola and Jonna K. Vuoskoski, who compared the discrete and dimensional models of emotion in music [40]. The aim of the authors' work was to systematically compare evaluations of perceived emotions in music using two different theoretical frameworks: the discrete emotion model, and dimensional model of affect. A secondary aim was to introduce a new, improved set of stimuli – consisting of unfamiliar thoroughly tested and validated non-synthetic music excerpts – for the study of music-mediated emotions. A very interesting aspect of their work is that the resulting stimuli include not only the best examples of target emotions but also moderate examples that permit the study of more subtle variations in emotion.

Stimuli consist of excerpts from film music, which are generally composed for the purpose of mediating powerful emotional cues, and can serve as a relatively 'neutral' stimulus material in terms of musical preferences and familiarity. Unfamiliar excerpts were chosen to avoid episodic memories from particular films influencing perceived emotions in the music. With the same aim, the selected excerpts do not have lyrics, dialogue or sound effects. Stimuli (360 audio clips between 10 and 30 seconds long) were selected and validated by a panel of 12 experts as detailed in [40]. The authors finally compile a list of 110 stimuli (50 discrete + 60 dimensional), properly validated and fit for use in academic research. Given that in this study we use the dimensional model of music, the 60 dimensional excerpts were used in our study, as described in the following section.

4.1. Study Design

The participants in the study were 35 voluntaries (mean age 18.6, SD = 1.4 years), mostly students of the first year of the Degree in Computer Science. All the participants thus have computer proficiency. Participants were randomly distributed among four groups. Each group was led to a different room. All the rooms were equipped with similar computers. Each participant was randomly assigned to one computer. The rooms were all inside the same building (the Department of Informatics of the University of Minho) and had similar conditions in terms of lightning, temperature and humidity. The rooms were equipped with Logitech X-540 5.1 surround sound speaker systems with subwoofer.

Participants in Group 1 (n = 8) validated the set of stimuli used. Audio clips of the dimensional model were used that are highly representative of the six extremes of three bipolar dimensions (valence, energy arousal and tension arousal). Thus, participants listened to a total of 30 auditory stimuli. These 30 stimuli had examples of both positive and negative valence, high and low tension arousal, as well as high and low energy arousal. As opposed to the other three groups, Group 1 did not fill in questionnaires as they did not participate in the text-writing task.

Participants received instructions to rate the *perceived* emotion (in other words, emotions that are represented by music and perceived as such by the listener) for the three dimensions using bipolar scales. Each extreme was characterized using three adjectives. For valence, these were pleasant–unpleasant, good–bad, and positive–negative. For the energy dimension the adjectives were awake–sleepy, wakeful–tired, and alert–drowsy. The adjectives used to represent the extremes of the tension dimension were tense–relaxed, clutched up–calm, and jittery–at rest. Familiarity with the excerpts was also rated (0 = unfamiliar, 1 = somewhat familiar, 2 = very familiar). Participants were also asked to mark how much they liked each example (with a preference rating) and how beautiful they considered each example (with a beauty rating). In both cases this was on a scale from 1 (minimum) to 9 (maximum). These additional measures were added to determine if there is a relationship between liking a music and its effect on performance. In total, participants answered to 12 questions for each clip. All clips were played twice and in a random order.

Group 2 (n = 8) was deemed the control group. These participants were led into the room and assigned a computer. They started by filling in an electronic questionnaire (detailed in the following subsection) and, afterwards, instructed to write a text for around 1h. The text was printed and placed on the left side of the keyboard. During the whole time (questionnaire + text typing) the interaction patterns of the students were being recorded. There was no music playing, as well as no external noise, and students were not allowed to communicate during the task. The text was a thorough description of Portuguese geography and weather. It was selected to contain no chronological references as well as no emotional cues or memories.

Groups 3 and 4 (n = 10 and n = 9, respectively) followed a very similar protocol. They were led to their respective rooms and each participant was assigned a computer. As with Group 2, participants filled in the questionnaire and, afterwards, started typing the same text, in the same conditions. Nonetheless, each of these groups did it while listening to a different selection of auditory stimuli extracted from the original dataset compiled by [40]. Group 3 listened to a random mix of auditory stimuli classified as high or moderate examples of positive tension. Group 4 listened to a random mix of auditory stimuli classified as high or moderate examples of negative tension.

4.2. Characterization of the Population

As mentioned in Section 4.1, students of Groups 2-4 filled in a questionnaire before proceeding to the text-typing task. This questionnaire had as main aim to determine if there were any differences between the randomly generated groups. Specifically, we measured the participants' affect in search for predispositions to experience a certain emotion. Similarly, we also measured personality in terms of the Big-Five dimensions [43]. Finally, we measured the participants' preferences in terms of music and movie genres, among other aspects. This section analyzes these results and compares them among groups.

In general, most of the participants do not currently play any musical instrument. Only 22% of the participants of Group 2 play, as well as 8% of Group 3 and 20% of Group 4.

Concerning musical training in Group 2, 63.6% never had any musical training, 9.1% played an instrument for less than 2 years, 18.2% between 2 and 4 years, 9.1% between 4 and 8 years and no participant had musical training for more than 8 years. For Group 3, the values are, respectively, 50% (no training), 37.5% (less than 2 years) and 12.5% (2 - 4 years). Finally, in Group 4, the values are 20% (no training), 40% (less than 2 years), 20% (2-4 years) and 20% (4-8 years). The levels of musical training are thus very similar among groups.

Participants were also asked to rate how much they liked several music genres on a scale from 1 - I hate it, to 5 - I love it, also including a 0 - No Opinion/Never listened to it. The following genres were included in this question: Classic, Blues, Country, Dance/Electronic, Folk, Rap/hip-Hop, Soul/funk, Religious, Alternative, Jazz, Rock, Pop, Heavy Metal and Movie Soundtracks. No significant differences were found between the 3 groups in terms of musical preferences. However, one interesting aspect must be pointed out: Movie Soundtracks are the second most favorite genre, only bested by Rock. Preferences regarding both genres are consensual, having only neutral and positive evaluations, unlike all others. Specifically, 33.3% of participants find soundtracks neutral, 40.7% like them and 22.2% love them, while 3.7% have no opinion on this genre. It is interesting to note that soundtracks (which are the source of stimuli in this study) are the second preferred music genre since the effects of music are largely conditioned by how much we like it [32].

In order to assess the participants' personality we used the Ten Item Personality Measure (TIPI) [44]. This instrument was developed to be used in scenarios where time is limited, in which researchers may be faced with the choice of using an extremely brief measure of the Big-Five personality dimensions or using no measure at all. Although somewhat inferior to standard multi-item instruments, it reached adequate levels in terms of (a) convergence with widely used Big-Five measures in self, observer, and peer reports, (b) test-retest reliability, (c) patterns of predicted external correlates, and (d) convergence between self and observer ratings. On the basis of these tests, a 10-item measure of the Big Five dimensions is offered for situations when very short measures are needed, personality is not the primary topic of interest, or researchers can tolerate the somewhat diminished psychometric properties associated with very brief measures. This short instrument was selected in an attempt to shorten the duration of the study, bearing in mind that participants would already be answering many other questions and also typing the text for around 1 hour.

The Portuguese (European Portuguese) translation of the TIPI instru-

Table 2: Mean and Standard Deviations for each of the five personality measures and each of the 3 groups of participants.

	Group 2			Group 3			Group 4		
	$\bar{x}$	$\widetilde{x}$	Std	$\bar{x}$	$\widetilde{x}$	Std	$\bar{x}$	$\widetilde{x}$	Std
Extraversion	5	4.5	0.87	4.73	5.5	1.93	4.44	3.75	1.68
Agreeableness	5.6	5.5	0.65	5.23	5	0.85	4.81	4.75	1.33
Conscientiousness	4.9	5	1.08	4.41	4.5	0.80	4.63	4.75	1.18
Emotional Stability	4.6	5	1.19	4.27	4	1.46	3.94	3.75	1.72
Openness	4.6	4.5	0.74	4.82	5	1.27	5.63	5.5	0.88

ment was used, developed by São Luís Castro and Cesar Lima[1]. Participants were thus asked to which extent they agree or disagree with 10 statements, on a scale between 1 - Disagree strongly and 7 - Agree strongly. These statements started with the sentence "I see myself as:" and were completed by 10 different pairs of traits: (1) Extraverted, enthusiastic; (2) Critical, quarrelsome; (3) Dependable, self-disciplined; (4) Anxious, easily upset; (5) Open to new experiences, complex; (6) Reserved, quiet; (7) Sympathetic, warm; (8) Disorganized, careless; (9) Calm, emotionally stable; and (10) Conventional, uncreative. The instrument produces, as output, a value between 1 and 7 for Extraversion, Agreeableness, Conscientiousness, Emotional Stability and Openness.

Table 2 summarizes the measured values of personality for the three groups and the five personality measures, which are very similar among the three groups and for all the personality measures, depicted graphically in Figure 5. In order to identify eventual differences, we used the Mann-Whitney Test to compare the distributions of the data of each group with each of the other two, for each personality measure. As Table 3 details, no statistically significant differences were found between the groups for all personality measures, which indicates that all the groups have similar personality predispositions. This is relevant as finding significant personality differences could, at least to some extent, explain eventual differences in the effect of music as participants could feel music differently. This does not happen in these groups.

[1]The Portuguese version of the TIPI instrument is available for download at http://gosling.psy.utexas.edu/scales-weve-developed/ten-item-personality-measure-tipi/

Table 3: p-values of the Mann-Whitney Test when comparing the distribution of the data characterizing the personalities of each group.

	G2 vs. G3	G2 vs G4	G3 vs G4
Extraversion	0.909134	0.17847	0.835094
Agreeableness	0.347551	0.209604	0.450147
Conscientiousness	0.388446	0.656092	0.616875
Emotional Stability	0.491535	0.234366	0.560321
Openness	0.644029	0.0728175	0.181782

We were also interested in measuring emotion in the participants since a given predisposition to a certain emotion (e.g. someone being deeply sad due to some recent event) may significantly alter the way emotions are perceived. For this purpose we used the Positive and Negative Affect Schedule (PANAS) instrument [45]. Specifically, we used the Portuguese adaptation of this instrument for the Portuguese population, put forward and validated by [46].

The PANAS comprises two mood scales, one that measures positive affect and the other that measures negative affect. Used as a psychometric scale, the PANAS can show relations between positive and negative affect with personality stats and traits. Ten descriptors are used for each positive affect scale (interested, alert, attentive, excited, enthusiastic, inspired, proud, determined, strong, active) and negative affect (distressed, upset, guilty, ashamed, hostile, irritable, nervous, jittery, scared, afraid) to define their meanings. Participants in the PANAS are asked to rate the extent to which they experienced each of the 20 emotions on a 5-point Likert Scale ranging from "very slightly" to "very much". Although the time-frame may vary according to the study's goals, in this specific case participants were asked to which extent they experienced the emotions in the last week, including the day of the study.

The results concerning participants' emotions indicate, in the first place, that positive affects are higher than negative affects in both groups (Table 4 and Figure 6). This is what is expected in a healthy population. Moreover, the table also shows that the valence of emotions felt by the participants in the same week are very similar among the three groups. To validate this we used the Mann-Whitney test. The resulting p-values are detailed in Table 5, showing no statistically significant differences between the groups.

The analysis carried out in this section shows that there are no statisti-

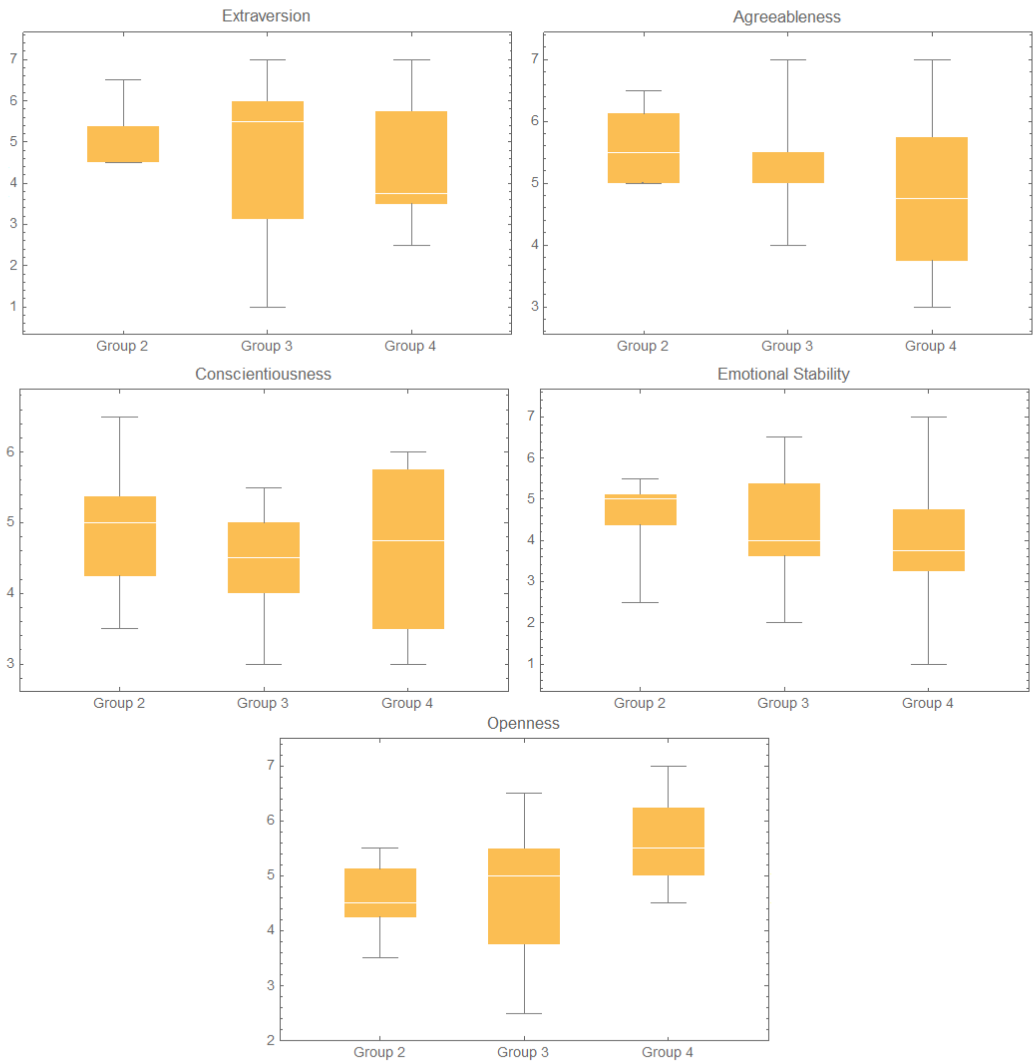

Figure 5: Positive and negative affects for the three groups.

cally significant differences among the groups that constitute the population of this study in terms of, among others, musical preferences, musical training, emotional state and personality. Although such differences were not expected since the population was randomly selected, proving that they do not exist is important to ensure the validity of the results described further ahead on the paper.

Table 4: Mean, median and standard deviation of positive and negative affects for the three groups.

	Group 2			Group 3			Group 4		
	$\bar{x}$	$\widetilde{x}$	Std	$\bar{x}$	$\widetilde{x}$	Std	$\bar{x}$	$\widetilde{x}$	Std
Positive Affect	28.8	29	7.29	27.55	27	6.20	29.5	31	4.96
Negative Affect	19.4	20	4.45	19.55	19	4.44	22	22	7.55

Table 5: p-values of the Mann-Whitney Test comparing the distributions of the positive and negative affects for all the pairs of groups.

	G2 vs. G3	G2 vs G4	G3 vs G4
Positive Affect	0.689776	0.941411	0.319404
Negative Affect	1.	0.604915	0.450147

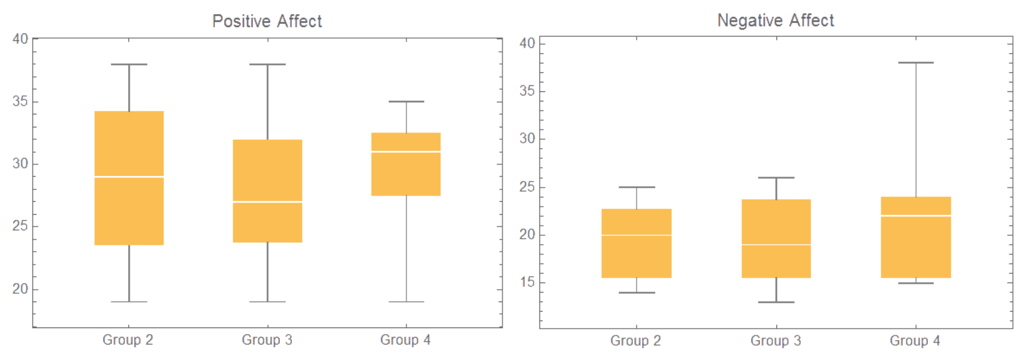

Figure 6: Positive and negative affects for the three groups.

4.3. Validation of the Stimuli

As mentioned in Section 4.1, the aim of Group 1 was to validate the auditory stimuli used. Participants listened to 30 excerpts and classified each one in terms of valence, tension arousal and energy arousal (to a total of 9 adjectives) and in terms of preference, beauty and familiarity. The main aim of this analysis is to determine if results similar to those achieved in [40] are found in the Portuguese population. The fact that we are building on a previously classified and validated dataset allows to perform the following: we analyze the responses of our participants regarding each stimulus in function of what we already know about them. This allows us to determine, for example, if the stimuli classified by [40] as having high positive valence are also classified by our participants as having high positive valence and low negative valence.

Concerning valence indicators (Figure 7), each participant rated each stimulus with a value between 1 and 7 for the following adjectives: unpleasant(1)-pleasant(7), negative(1)-positive(7) and bad(1)-good(7). Positive valence and energy and negative tension are considered, by far, more pleasant. Participants also find that positive valence and energy convey more positive emotions. For tension the results are more blurred. Finally, participants also deem positive valence and energy as good, as well as negative tension.

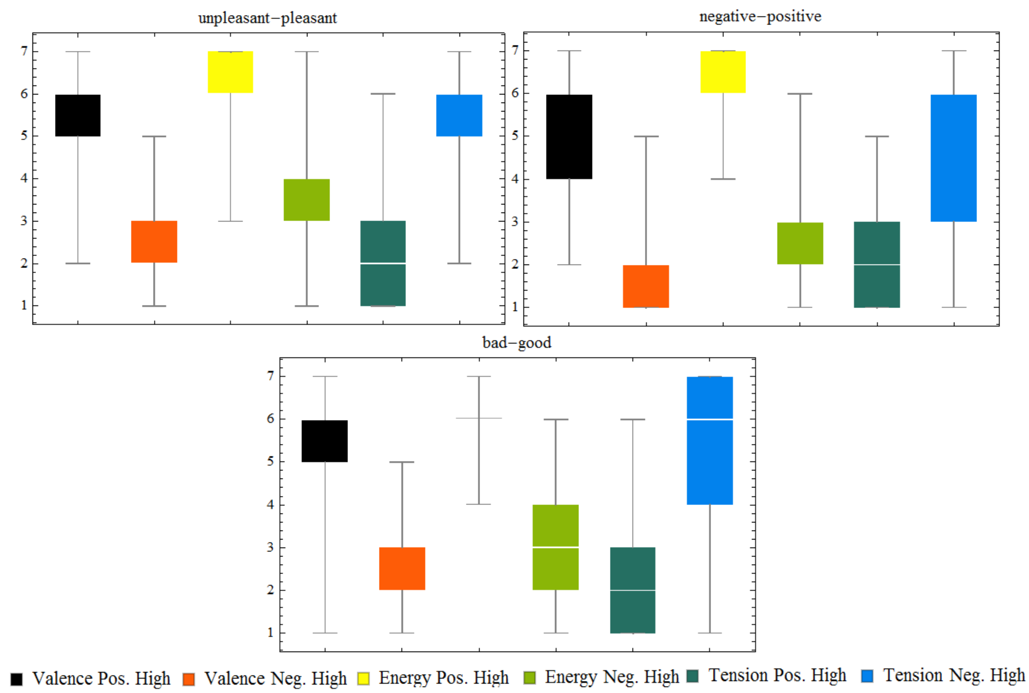

Figure 7: Ratings of the auditory stimuli concerning valence: unpleasant-pleasant, negative-positive and bad-good.

Concerning energy indicators (Figure 8), each participant rated each stimulus as sleepy(1)-awake(7), tired(1)-wakeful(7) and alert(1)-drowsy(7). Clips with negative valence, positive energy and positive tension are characterized as more awaking. The same happens with the other adjectives (tired-wakeful and alert-drowsy).

Finally, concerning tension indicators (Figure 9), each participant rated each stimulus as tense(1)-relaxed(7), clutched(1)-calm(7) and jittery(1)-at rest(7). The stimuli classified by students as more relaxed are the repre-

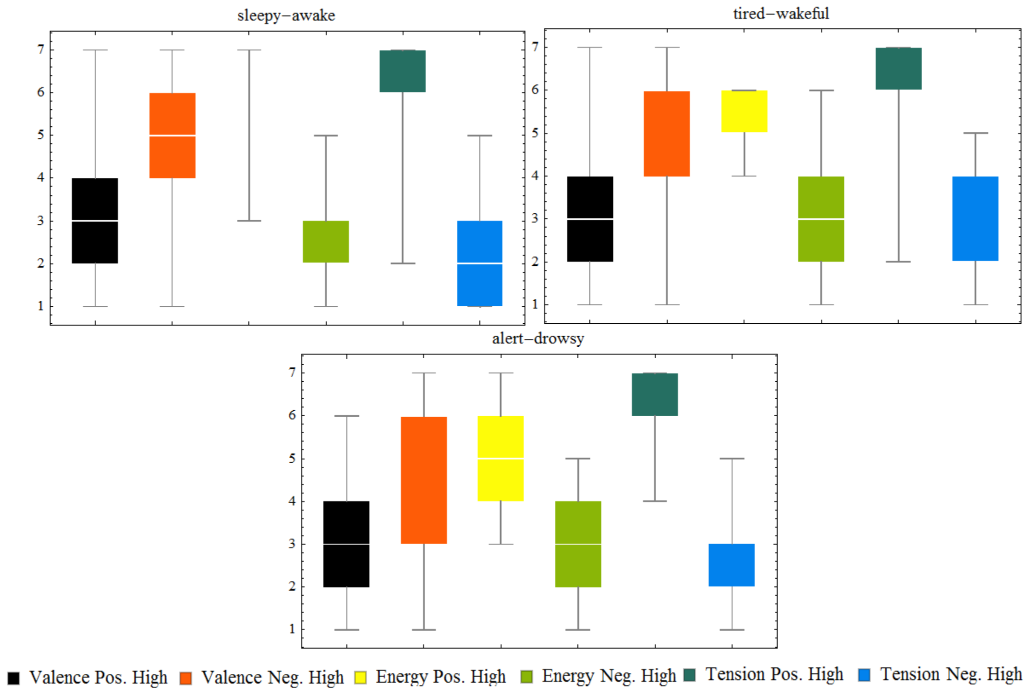

Figure 8: Ratings of the auditory stimuli concerning energy: sleepy-awake, tired-wakeful and alert-drowsy.

sentatives of positive valence, positive energy (but not as conclusive) and negative tension (with a clear agreement). Similar results are achieved for calm and positive valence and negative tension are the ones that make participants feel more at-rest.

Looking at these results, one of the first conclusions is that the Tension dimension is where differences are more visible between the two extremes. Table 6 provides the necessary data for this analysis. The mean values of the Tension adjectives (tense-relaxed, clutched-calm and jittery-at rest) for positive tension are, respectively 1.8, 2.1 and 2.0, while for negative tension these values are 5.8, 6.0 and 5.9. If we do a similar analysis for the other two dimensions the values are 5.8, 5.1 and 5.5 for positive valence, 2.4, 2.1 and 2.3 for negative valence, 5.9, 5.3 and 5.1 for positive energy and 2.6, 2.9 and 2.9 for negative energy. This points out that the Tension dimension is where participants more clearly feel the difference between extremes in music. If we adhere to the principle that dimensions whose differences in extremes

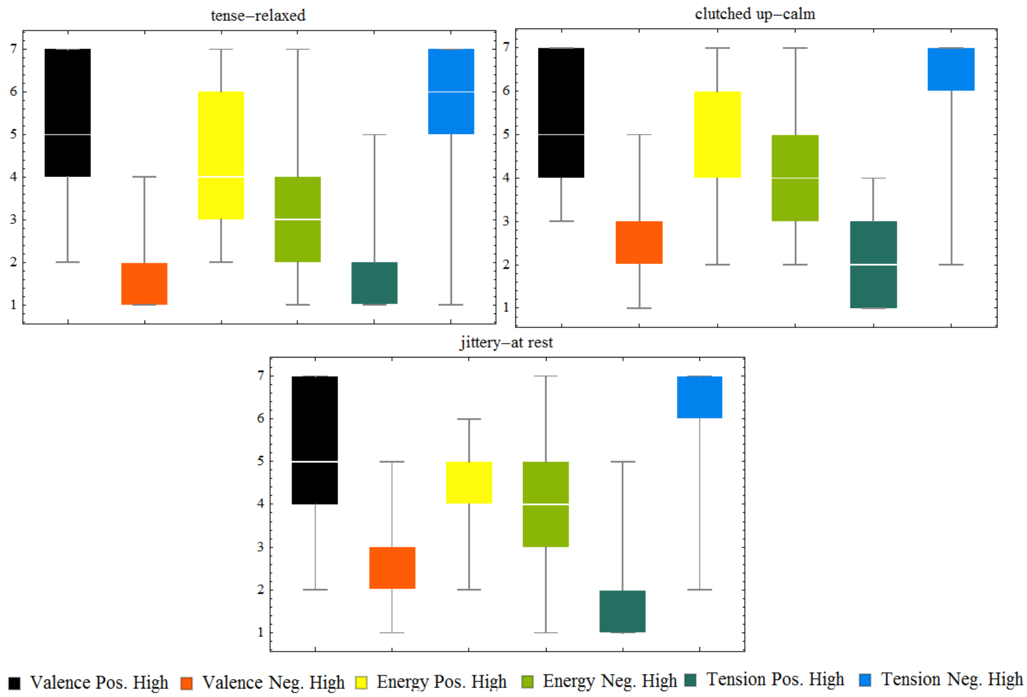

Figure 9: Ratings of the auditory stimuli concerning tension: tense-relaxed, clutched-calm and jittery-at rest.

are more clearly perceived by the participants will most likely also have more significant effects on the listener, selecting Tension as the dimension to explore in the study with the other three groups will increase the probability of finding differences in interaction patterns.

In Table 6 the ratings of the adjectives that belong to their category are highlighted. It is clear that the ratings are in line with the previous classification of the stimuli: positive valence and energy are associated with more positive adjectives, while positive tension is associated with more tense adjectives. In this table, the first 9 lines describe the previously mentioned 9 pairs of adjectives and are represented by their initials: unpleasant-pleasant (u-p), negative-positive (n-p), bad-good (b-g), and so forth. Familiarity is represented by an 'F', preference by a 'P' and Beauty by a 'B'.

In terms of familiarity, participants denote that they are unfamiliar with all the stimuli (with an average value of around 1.5 between 1 and 3). As mentioned before, this is desirable in order to decrease the possibility of influencing perceived emotions with episodic memories.

Table 6: Averages (M) and Standard Deviations (S) of the ratings of the participants, for stimuli of 6 different types: high positive and negative valence (V+H and V-H), high positive and negative energy (E+H and E-H) and high positive and negative tension (T+H and T-H)

	V+H		V-H		E+H		E-H		T+H		T-H	
	M	S	M	S	M	S	M	S	M	S	M	S
u-p	5.8	1	2.4	1	6.0	0.83	3.6	1.3	2.4	1.4	5.5	1.5
n-p	5.1	1.5	2.1	1.0	6.1	0.93	2.8	1.0	2.0	1.2	4.8	2.0
b-g	5.5	1.3	2.3	1.0	6.1	0.69	3.2	1.2	2.2	1.3	5.2	1.7
s-a	3.2	1.7	5.1	1.5	5.9	0.93	2.6	1.1	6.1	0.99	2.3	1.2
t-w	3.2	1.5	5.3	1.6	5.3	0.69	2.9	1.5	6.1	1.2	2.6	1.3
a-d	3.5	1.4	5.0	1.9	5.1	0.97	2.9	1.4	6.4	0.81	2.5	1.2
t-r	5.3	1.4	1.9	1.0	4.4	1.5	3.4	1.5	1.8	1.0	5.8	1.6
c-c	5.5	1.3	2.4	1.1	4.6	1.3	4.1	1.3	2.1	0.88	6.0	1.3
j-r	5.5	1.4	2.3	0.91	4.4	1.2	3.9	1.4	2.0	0.99	5.9	1.4
F	1.5	0.55	1.4	0.63	1.5	0.72	1.4	0.67	1.6	0.74	1.4	0.67
P	6.3	1.3	3.6	1.8	6.7	0.98	4.5	1.7	4.9	2.1	6.8	1.6
B	7.0	1.3	2.8	1.6	7.0	1.2	4.5	1.7	4.0	1.9	7.2	1.7

Finally, another very interesting result concerns preference and beauty. In fact, the most preferred stimuli and the ones deemed as most beautiful are the ones previously classified as high negative tension. Once again, this increases our confidence on the potential effects that will be attained by studying the effects of tension on interaction patterns, especially when we recall that the effects of music are higher when the listener likes the music [32].

In order to assess the consensus between the raters for each emotional extreme the Cronbach's alpha was employed. This is an indicator of the reliability of a questionnaire, evaluating the correlation between the answers in a questionnaire given by the participants. Cronbach's alpha is a function of the number of items in a test, the average covariance between item-pairs, and the variance of the total score.

Most emotion extremes scored relatively high consistency using this procedure, with the exception of negative energy. The values of consistency measured are as follows: positive valence $\alpha = 0.94$, negative valence $\alpha = 0.94$, positive energy $\alpha = 0.95$, negative energy $\alpha = 0.75$, positive tension $\alpha = 0.96$ and negative tension $\alpha = 0.95$.

5. Results

Until now we have analyzed the results of this study in terms of the characteristics of the different groups (e.g. musical preferences, musical training, emotional state, personality). We have also detailed the process carried out to validate the stimuli used for the Portuguese population. In the first case we determined that there are no significant differences between the participants of each group. In the second case we achieved ratings of the stimuli that are in line with the ones achieved by the authors who compiled the dataset, also with high consistency between raters.

This process, although not directly related to the main aim of the paper, was fundamental to establish the validity of the results that will be analyzed in this section. We analyze how music affects Human-Computer Interaction in terms of 15 interaction features, described in detail in Section 3.1. We first compare the control group (group 2, without music) with groups 3 and 4 (positive tension and negative tension, respectively) to determine if music has an effect on the performance of the interaction. Next we compare groups 3 and 4 to determine if opposite extremes of tension have effects on the same variable.

5.1. Control Group vs. Music Groups

In this analysis of the data the main aim is to find differences between the control group and each of the two groups who were listening to music while filling in the questionnaires and typing the text. Specifically, we analyze the performance of the interaction with the computer by considering the 15 interaction features described previously.

A feature-by-feature analysis is performed, supported on the data summarized in Table 7. This table details the mean value of each feature for each of the three groups in the first three columns. Statistically significant differences between the distribution of the data for groups 2 and 3 and groups 2 and 4 were also studied. The Kolmogorov-Sminorv test was used to assess the normality of the distributions. Given that the majority of them are not normal, the Mann-Whitney test was used to test for statistically significant differences. A value of $\alpha = 0.05$ was used. Figure 10 depicts some of the observed differences graphically.

The numbers in each line represent the different interaction features as follows: 1 - Click Duration, 2 - Time Between Clicks, 3 - Time Double Clicks, 4 - Mouse Velocity, 5 - Mouse Acceleration, 6 - Distance Between Clicks, 7 -

Table 7: Mean values of each feature for each of the three groups (columns Mean G2 - Mean G4). p-values of the Mann-Whitney test comparing Group 2 with Group 3 (column G2 vs G3) and comparing Group 2 with Group 4 (column G2 vs G4).

	Mean G2	Mean G3	Mean G4	G2 vs G3	G2 vs G4
1	77.7184	103.074	110.514	$3.00528 * 10^{-10}$	$1.86567 * 10^{-9}$
2	3503.3	2594.23	2989.64	0.370572	0.00750602
3	153.5	145.81	156.63	0.981598	0.426485
4	0.717832	0.311914	0.318445	$2.37754 * 10^{-195}$	$9.54369 * 10^{-190}$
5	0.791762	0.399685	0.397133	$1.87349 * 10^{-182}$	$6.10427 * 10^{-199}$
6	227.671	160.639	161.133	0.00866251	0.0223202
7	300.226	155.573	154.653	0.005846	0.0191563
8	1.84713	1.66645	1.76318	0.835636	0.341505
9	6032.7	2580.51	2193.12	0.0481916	0.00660626
10	38.5546	24.4485	23.4449	0.000608397	0.000661682
11	3510.79	3818.22	3766.5	0.00534491	0.00703587
12	7.45692	-1.68332	-8.14551	0.0409389	0.0114148
13	190.008	171.815	173.742	$2.06629 * 10^{-11}$	$1.05541 * 10^{-10}$
14	86.563	90.8743	88.4992	0.000113984	0.0000222556
15	217.951	178.152	194.202	$5.06863 * 10^{-8}$	0.00126828

Excess of Distance, 8 - Average Excess of Distance, 9 - Distance of the Mouse to the Straight Line, 10 - Average Distance of the Mouse to the Straight Line, 11 - Absolute Sum of Angles, 12 - Signed Sum of Angles, 13 - Time Between Keys, 14 - Key Down Time and 15 - Writing Velocity.

Of the 15 features studied, 12 of them show statistically significant differences when comparing the control group with groups 3 and 4 , which clearly shows that the stimuli used have indeed an influence on the interaction patterns.

In terms of the performance of this interaction, only 12 features of the original 15 can be considered, as for the other 3 there is no direct relationship to performance. Of this 12, 9 features have statistically significant differences when comparing the control group with the other two groups. This allows us to conclude that the stimuli used affect not only the interaction patterns but also the performance of this interaction. We will thus focus our analysis on these 9 features.

In what concerns the motion of the mouse, performance improves in all the 6 features. This indicates that, with music, participants move the mouse in a more efficient way. In what concerns the keyboard, performance improves in only 1 of the 3 features. Results show that participants spend less time between keys but more time pressing them, resulting in an overall slower writing speed.

This difference between mouse and keyboard can be attributed to the nature of the stimuli, especially the ones classified as positive tension. These stimuli, which can be noisy and unpleasant (as pointed out in Section 4.3), were also clearly classified as "awake", "wakeful" and "alert". This means that these stimuli actually activate participants physically, driving them to perform faster, as often happens with activating music and as addressed in Section 2.1. However, this effect is only positive at a physical level. At the level of cognitive performance, namely in what concerns memory, there is a negative effect, which translates into poorer typing performance. This points out to a very interesting conclusion: the effect of music may be dependent on the type of task being performed and there may be the need to clearly distinguish between different objectives before selecting the most appropriate type of music.

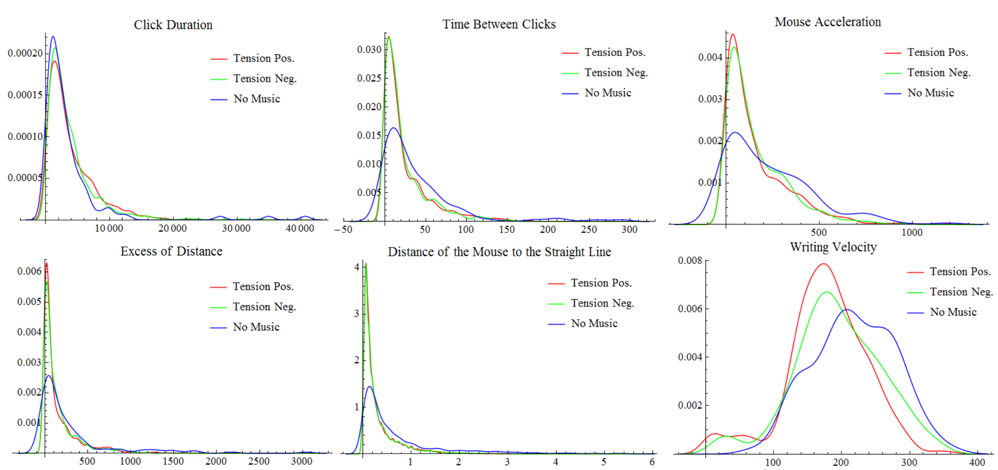

Figure 10: Histograms depicting the differences between the distributions of the data for 6 different features.

5.2. High Tension vs. Low Tension

Next, we analyzed the differences between Groups 3 and 4 alone. In the previous section we determined that listening to music characterized as containing positive and negative tension significantly affects performance. In this section we analyze if these different extremes of tension also have a different effect on the interaction features.

Comparing the distributions of the data from Groups 3 and 4, only 5 of the 15 features show statistically significant differences: Time Between Clicks (p-value = $5.94024 * 10^{-6}$), Mouse Acceleration (p-value = 0.000544135), Average Excess of Distance (p-value = 0.0399531), Signed Sum of Angles (p-value = 0.00107474) and Writing Velocity (p-value = 0.00148261).

Of these features, the Mouse Acceleration and the Signed Sum of Angles are not considered for the purpose of assessing performance, for the reasons mentioned previously. Of the remaining 3 features, the Time Between Clicks and the Average Excess of Distance (both features extracted from the mouse) show a better performance in Group 3 (positive tension). On the other hand, Writing Velocity shows a better performance in Group 4 (negative tension). These results are in line with the conclusions drawn in the previous section: they point out that stimuli with positive tension (very activating) improve mouse motion while impairing text typing. We thus conclude that although only 1 third of the features are significantly affected, they are affected in a coherent way, which strengthens our conclusions.

6. Discussion and Conclusions

The main goal of this paper was to present an approach for the non-intrusive analysis of performance in groups of people. This approach was implemented in the form of a distributed architecture that constantly collects, processes, stores, analyzes and monitors data describing individual behavior, a method in line with the Big Data approach. We have shown the expected requirements of this kind of architecture as the number of users and the time of the data collection grows, which allows for a precise planning of the system in terms of its scalability.

In order to validate this environment and to support the claim that external factors may influence performance and should be considered in performance monitoring approaches, we implemented an experimental study to demonstrate the effect of music selection on worker performance.

In what concerns this validation, the first conclusion is that performance is indeed affected. The results attained may have, however, more far-reaching implications and deserve to be analyzed with care.

In terms of study design, we chose a dimensional model since it is one of the most widely accepted by research. Although it is based on three dimensions, for the purpose of effects on HCI we selected only one of these dimensions (tension) and its two extremes (positive and negative). The decision on using this dimension was made after an analysis of the auditory stimuli in the dataset used, which held the conclusion that the extremes of tension were where music could be called more "different". In fact stimuli with positive tension can be deemed as stressful, tense, activating or even scary. They were extracted from the soundtracks of movies such as The Alien Trilogy, The Missing or Cape Fear. Stimuli with negative tension, on the other hand, can be characterized as very relaxing and calm. They were extracted from soundtracks of movies such as Shakespeare in Love or Pride and Prejudice. Nonetheless, we are now preparing similar studies to assess the potential effect of the other two dimensions of emotion that were not addressed in the current one (energy and valence).

The decision on selecting tension as the dimension to study was to some extent confirmed as a good one by the results, especially in what concerns Group 1. First of all, it is the dimension where differences between the two extremes are more clearly pointed out (i.e. perceived) by the participants. Thus, it should also be the dimension that induces more significant effects on people. Moreover, of the six different types of stimuli (3 dimension * 2 extremes), negative tension was the style that was rated as most beautiful and preferred. As different authors have pointed out, a positive effect of music is more dependent on the listener liking it than on the specific style of the music [32].

In terms of the effects of music on Human-Computer Interaction, the results point out to a clear and generally positive effect on performance. These results must however be analyzed in detail. Music, independently of the extreme of tension, seems to affect overall performance positively, especially in what concerns mouse motion. Participants in general move the mouse faster and in a more efficient way, traveling smaller distances and in more straight lines to perform the same task. This is because the motion of the mouse is essentially a physical task, that people carry out automatically after enough training, much like driving a car or riding a bicycle. The results concerning text typing are not so straightforward.

In fact, listening to music negatively affected writing speed, confirming the conclusions of previous researchers [15]. The explanation is that the task of writing text is largely cognitive: participants read the text from a sheet of paper, which means that they look at the paper to memorize a few words, then type them, and repeat the process until the end. It thus requires prolonged use of memory and concentration. These cognitive functions are affected by music, and are especially negatively affected by the positive dimension of tension, with stimuli that can be characterized as noisy and unpleasant, decreasing the ability of participants to memorize groups of words. The results attained point this out explicitly, with group 3 showing the slowest typing rhythm.

This is one of the most interesting conclusions of this work: the effect of music on performance is complex and must be broken down according to the different emotional dimensions of music and according to the task being performed by the individual. These results point out that while music with both negative and positive tension can be used to improve performance in what concerns mouse motion, the same is not true if one seeks to improve typing performance, in which case music with positive tension shows an especially negative effect. It also shows that Human-Computer Interaction is a complex issue, involving different abilities of the Human being, which are affected differently by music.

Human-Computer Interaction and the effect that music can have on it must thus be studied in more detail. We are convinced that there is still a large space to explore in what concerns the potential positive effects of music on HCI, with results not only on productivity but also on motivation, well-being, stress and other indicators. And this is one of the most interesting aspects of music: its many positive side-effects on the Human being.

In order to reduce complexity, in this study we have focused only on two extremes of tension. We are now preparing to study the effects of the other two dimensions. This will allow to understand, in a thorough way, the effect of the different emotional dimensions of music on HCI. This research may have a very interesting impact, not only in economic terms but also in social terms as it may improve performance and productivity in parallel with other indicators such as well-being. This is something that most of current approaches fail to achieve as most of productivity improving initiatives come at the cost of individual well-being, with negative consequences associated.

Specifically, we envision the use of this kind of system in workplaces or other environments in which people are allowed to work while listening to music, either using headphones or on speakers. In such a scenario, it is the aim of the manager to select the goal of the group at each moment as well as the characteristics/type of the task (e.g. brainstorming, creativity, longterm). Based on this information and on each worker's musical profile, the system will decide which music to play to improve general performance (if there are available speakers) or individual performance (if workers are allowed to use headphones).

This line of research will support the ongoing development of a fully-functional autonomous environment for performance management, with a strong focus on individual well-being. It will result advantageous for both organizations and individuals, economically and socially.

The presented work contributed to this goal with two main aspects:

- An environment for the continuous and transparent acquisition of large amounts of data describing the performance of groups of people, enabling real time analytics and improving decision making on human resources management;

- Some insights on the potential benefits of music as a way to improve performance in the workplace, with benefits at other levels including individual well-being.

Acknowledgments

This work has been supported by COMPETE: POCI-01-0145-FEDER-007043 and FCT – Fundação para a Ciência e Tecnologia within the Project Scope: UID/CEC/00319/2013. The work of Davide Carneiro is supported by a Post-Doctoral Grant by FCT (SFRH/BPD/109070/2015).

References

[1] K. Sparks, C. Cooper, Y. Fried, A. Shirom, The effects of hours of work on health: A meta-analytic review, Journal of occupational and organizational psychology 70 (4) (1997) 391.

[2] A. E. Dembe, J. B. Erickson, R. G. Delbos, S. M. Banks, The impact of overtime and long work hours on occupational injuries and illnesses: new evidence from the united states, Occupational and environmental medicine 62 (9) (2005) 588–597.

[3] J. M. Harrington, Health effects of shift work and extended hours of work, Occupational and Environmental medicine 58 (1) (2001) 68–72.

[4] R. Cowell, Willing slaves: how the overwork culture is ruling our lives by madeleine bunting. london: Harper collins, 2004. 368pp, isbn 0-00-716371-1, Practice Development in Health Care 4 (3) (2005) 167–167.

[5] A. Pimenta, D. Carneiro, P. Novais, J. Neves, Monitoring mental fatigue through the analysis of keyboard and mouse interaction patterns, in: Hybrid Artificial Intelligent Systems, Springer, 2013, pp. 222–231.

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

[7] J. R. Aiello, K. J. Kolb, Electronic performance monitoring and social context: impact on productivity and stress., Journal of Applied Psychology 80 (3) (1995) 339.

[8] J. Rincon, J. Bajo, A. Fernandez, V. Julian, C. Carrascosa, Using emotions for the development of human-agent societies, Frontiers of Information Technology & Electronic Engineering 17 (4) (2016) 325–337.

[9] R. V. Yampolskiy, V. Govindaraju, Behavioural biometrics: a survey and classification, International Journal of Biometrics 1 (1) (2008) 81–113.

[10] R. M. Bolle, J. Connell, S. Pankanti, N. K. Ratha, A. W. Senior, Guide to biometrics, Springer Science & Business Media, 2013.

[11] P. Shukla, R. Solanki, Web based keystroke dynamics application for identifying emotional state, vol 2 (2011) 4489–4493.

[12] W.-H. Tsui, P. Lee, T.-C. Hsiao, The effect of emotion on keystroke: an experimental study using facial feedback hypothesis, in: Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE, IEEE, 2013, pp. 2870–2873.

[13] M. B. Lewis, Exploring the positive and negative implications of facial feedback., Emotion 12 (4) (2012) 852.

[14] A. Kolakowska, A review of emotion recognition methods based on keystroke dynamics and mouse movements, in: Human System Interaction (HSI), 2013 The 6th International Conference on, IEEE, 2013, pp. 548–555.

[15] X. Jiang, A. K. Sengupta, Effect of music and induced mental load in word processing task, in: Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on, IEEE, 2011, pp. 3261–3266.

[16] C. Shen, Z. Cai, X. Guan, Y. Du, R. A. Maxion, User authentication through mouse dynamics, Information Forensics and Security, IEEE Transactions on 8 (1) (2013) 16–30.

[17] C.-J. Tsai, T.-Y. Chang, Y.-J. Yang, M.-S. Wu, Y.-C. Li, An approach for user authentication on non-keyboard devices using mouse click characteristics and statistical-based classification, International Journal of Innovative Computing, Information and Control 8 (11) (2012) 7875–7886.

[18] Z. Jorgensen, T. Yu, On mouse dynamics as a behavioral biometric for authentication, in: Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security, ACM, 2011, pp. 476–482.

[19] A. Pimenta, D. Carneiro, J. Neves, P. Novais, A neural network to classify fatigue from human–computer interaction, Neurocomputing 172 (2016) 413–426.

[20] L. M. Vizer, L. Zhou, A. Sears, Automated stress detection using keystroke and linguistic features: An exploratory study, International Journal of Human-Computer Studies 67 (10) (2009) 870–886.

[21] J. Andrén, P. Funk, A case-based approach using behavioural biometrics to determine a user's stress level., in: ICCBR Workshops, 2005, pp. 9–17.

[22] R. Bixler, S. D'Mello, Detecting boredom and engagement during writing with keystroke analysis, task appraisals, and stable traits, in: Proceedings of the 2013 international conference on Intelligent user interfaces, ACM, 2013, pp. 225–234.

[23] K. D. Allred, J. F. Byers, M. L. Sole, The effect of music on postoperative pain and anxiety, Pain Management Nursing 11 (1) (2010) 15–25.

[24] U. Nilsson, M. Unosson, N. Rawal, Stress reduction and analgesia in patients exposed to calming music postoperatively: a randomized controlled trial, European journal of anaesthesiology 22 (02) (2005) 96–102.

[25] T. P. Hatem, P. I. Lira, S. S. Mattos, The therapeutic effects of music in children following cardiac surgery, Jornal de pediatria 82 (3) (2006) 186–192.

[26] X. Teng, M. Y. M. Wong, Y.-T. Zhang, The effect of music on hypertensive patients, in: Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE, IEEE, 2007, pp. 4649–4651.

[27] T. Särkämö, M. Tervaniemi, S. Laitinen, A. Forsblom, S. Soinila, M. Mikkonen, T. Autti, H. M. Silvennoinen, J. Erkkilä, M. Laine, et al., Music listening enhances cognitive recovery and mood after middle cerebral artery stroke, Brain 131 (3) (2008) 866–876.

[28] S. Simavli, I. Kaygusuz, I. Gumus, B. Usluogulları, M. Yildirim, H. Kafali, Effect of music therapy during vaginal delivery on postpartum pain relief and mental health, Journal of affective disorders 156 (2014) 194–199.

[29] C.-F. Wang, Y.-L. Sun, H.-X. Zang, Music therapy improves sleep quality in acute and chronic sleep disorders: A meta-analysis of 10 randomized studies, International journal of nursing studies 51 (1) (2014) 51–62.

[30] G. G. D. C. Taets, C. J. Borba-Pinheiro, N. M. A. d. Figueiredo, E. H. M. Dantas, Impact of a music therapy program on the stress level of health professionals, Revista brasileira de enfermagem 66 (3) (2013) 385–390.

[31] S. Garrido, E. Schubert, Music and people with tendencies to depression, Music Perception: An Interdisciplinary Journal 32 (4) (2015) 313–321.

[32] E. G. Schellenberg, S. Hallam, Music listening and cognitive abilities in 10-and 11-year-olds: The blur effect, Annals of the New York Academy of Sciences 1060 (1) (2005) 202–209.

[33] M. Besson, D. Schön, S. Moreno, A. Santos, C. Magne, Influence of musical expertise and musical training on pitch processing in music and, Restorative neurology and neuroscience 25 (2007) 399–410.

[34] I. Roden, D. Grube, S. Bongard, G. Kreutz, Does music training enhance working memory performance? findings from a quasi-experimental longitudinal study, Psychology of Music 42 (2) (2014) 284–298.

[35] M. H. Thaut, D. A. Peterson, G. C. McIntosh, V. Hoemberg, Music mnemonics aid verbal memory and induce learning–related brain plasticity in multiple sclerosis, Frontiers in human neuroscience 8.

[36] D. L. Strait, J. Slater, S. O'Connell, N. Kraus, Music training relates to the development of neural mechanisms of selective auditory attention, Developmental cognitive neuroscience 12 (2015) 94–104.

[37] Z. Belford, C. Neher, T. Pernsteiner, J. Stoffregen, Z. Tariq, A. Lokuta, Music and physical performance: the effects of different music genres on physical performance as measured by the heart rate, electrodermal arousal, and maximum grip strength (2013).

[38] J. Edworthy, H. Waring, The effects of music tempo and loudness level on treadmill exercise, Ergonomics 49 (15) (2006) 1597–1610.

[39] M. Korczynski, M. Pickering, E. Robertson, Rhythms of labour: Music at work in Britain, Cambridge University Press, 2013.

[40] T. Eerola, J. K. Vuoskoski, A comparison of the discrete and dimensional models of emotion in music, Psychology of Music.

[41] J. C. Castillo, D. Carneiro, J. Serrano-Cuerda, P. Novais, A. Fernández-Caballero, J. Neves, A multi-modal approach for activity classification and fall detection, International Journal of Systems Science 45 (4) (2014) 810–824.

[42] P. Novais, R. Costa, D. Carneiro, J. Neves, Inter-organization cooperation for ambient assisted living, Journal of Ambient Intelligence and Smart Environments 2 (2) (2010) 179–195.

[43] D. S. Chiaburu, I.-S. Oh, C. M. Berry, N. Li, R. G. Gardner, The five-factor model of personality traits and organizational citizenship behaviors: a meta-analysis., Journal of Applied Psychology 96 (6) (2011) 1140.

[44] D. A. Cobb-Clark, S. Schurer, The stability of big-five personality traits, Economics Letters 115 (1) (2012) 11–15.

[45] D. Watson, L. A. Clark, A. Tellegen, Development and validation of brief measures of positive and negative affect: the panas scales., Journal of personality and social psychology 54 (6) (1988) 1063.

[46] I. C. Galinha, J. L. P. Ribeiro, Contribuição para o estudo da versão portuguesa da positive and negative affect schedule (panas): Ii–estudo psicométrico, Análise Psicológica XXIII (2) (2005) 219–227.