iSenseStress: Assessing stress through human-smartphone interaction analysis

Matteo Ciman
Department of Mathematics
University of Padua
Padua, Italy
mciman@math.unipd.it

Katarzyna Wac
University of Geneva and University of Copenhagen
Quality of Life and HCC group
Geneva, Switzerland and Copenhagen, Denmark
Katarzyna.Wac@unige.ch and wac@di.ku.dk

Ombretta Gaggi
Department of Mathematics
University of Padua
Padua, Italy
gaggi@math.unipd.it

***Abstract*—Stress condition, if experienced for an extended amount of time, can negatively affect individual's health. Several external sensors monitoring different physiological states correlated with stress, or smartphone apps that monitor individuals context, have been leveraged to assess stress state in everyday life. The less intrusive "human-smartphone interaction" have been under-investigated so far. In our research we leverage 'swipe', 'scroll' and 'text input' interactions to assess the stress state of smartphone users. Based on data collected from 13 participants, we leverage 'swipe' and 'scroll' data to assess stress with an average F-measure of 79-85% for a within-subject model, and of 70-80% when building a global model. Moreover, 'text input' via a virtual keyboard has been analyzed, showing how several easy to calculate features enable to differentiate between stress and no-stress state. To the best of our knowledge, this is the first attempt to leverage human-smartphone interaction, and in particular 'swipe', 'scroll' and 'text input' interactions, to accurately assess stress state in individuals without using any external sensor or leveraging privacy-sensitive context information.**

Keywords*—*smartphone; interaction; stress assessment; gesture

I. Introduction

Stress is a mental condition that everybody experiences in his/her life, sometimes even daily. This could come as a short reaction to an experience, i.e., speaking in public, or it could last longer, i.e., during problems in a relationship. Even if everybody will experience this state during his/her life, it is clear that a long exposure to stress that interferes with normal life can become unhealthy. It can lead to an inability to concentrate, irritability, anxiety, depression etc. [1], which may strongly contribute to the increase of health care costs [2]. Furthermore, a recent survey from the American Psychological Association showed that in 2013 even teens reported that they experience stress at unhealthy levels; the first experience with negative stress occurs at an increasingly lower ages [3].

Due to all these negative effects, it can be assumed that early assessment of stress condition, and early suggestions on how to reduce it, may reduce its overall impact and lead to improved health state of individuals. Wearable and ubiquitous devices, i.e., smartphones or wearable physiological sensors, provide the possibility to assess stress conditions in everyday life, without interfering with habits of their users.

Several past research results have shown that stress conditions are correlated with particular physiological states like heart rate variability (HRV), muscle tension or galvanic skin response (GSR) [4] [5]. Moreover, other researchers investigated the possibility to assess stress condition using smartphone as a monitoring device, trying to find differences in individuals behavior leveraging location (GPS), Bluetooth and other sensors [6].

Differently from previous works, in our approach we assess stress state in a less intrusive way; we do not use any privacy-sensitive context information. We rely on the fact that, since our smartphones are integrated part of our lives, we interact with them several times during a day, in different ways and under different conditions to achieve some goals. In this paper we present how by analyzing this interaction it is possible to infer the individual's stress state. We show how features related to 'tap', 'scroll' and 'swipe' interaction are accurate indicators to discriminate between stress and no-stress state. Our solution lowers the intrusiveness of stress monitoring in everyday life, and so can increase individuals' acceptance.

The paper is organized as follows. Section II discusses state of the art in stress assessment. Section III describes our overall protocol and its specific design choices. Section IV presents data acquisition and elaboration phase, while Section V presents our results. We conclude in Section VI providing final remarks and future works areas.

II. Related work areas

Stress detection and monitoring using external sensors and smartphones has gained lot of interest from the research community, due to the increasing diffusion of this mental condition and the possibility to use non-intrusive sensors and unobtrusive devices to monitor people during their everyday life activities.

Sun et al. [4] explored the possibility to identify stress in users analyzing common computer mouse operations and building a model of the hand and muscles with a *mass-spring-damper* system. They asked to users to perform three different activities: *Point-and-click* (click inside two target rectangles), *Drag-and-drop* (move one rectangle over another), and *Steer* (draw a line inside a tunnel as much linear as possible). Each user has to perform this task both in a relaxed and a stressed state. They concluded that stress measures can be more indicative using mouse-derived metrics than electrocardiogram signal analysis, in particular when considering within-subject model. Moreover, they showed that a small number of data

PervasiveHealth 2015, May 20-23, Istanbul, Turkey

DOI 10.4108/icst.pervasivehealth.2015.259280

(about 10 minutes of interaction for the user in a given stress state) can be used to infer stress state with 70% accuracy.

Hernandez et al. [7] evaluated the possibility to use keyboards and mouses to detect stress in users. They used a pressure-sensitive keyboard and a capacitive mouse, focusing analysis on the touch pressure. The tasks that each of the 24 participants had to perform were: text input (pre-specified text, as fast as possible), expressive writing (writing memories for 5 minutes) and mouse clicking (clicking on different horizontal bars of different sizes). The results showed that stress significantly influences the keyboard touching pressure, in more than 83% of the users. Considering the mouse, the pressure intensity was not significant, but the amount of mouse contact area increased during stress moments.

Sano and Picard [8] researched physiological or behavioral markers for stress; they collected data from an accelerometer, GSR sensor and mobile phone usage. In particular, they collected data relative to screen ON/OFF, SMS (length, number of receivers, number of messages etc.), calls (person called, duration, etc.) and the location of the user. With a precision of 75%, they were able to discriminate stress/no-stress state using information about when the screen is ON, mobility data, calls data and user activity (sitting, walking) information. This is however a very privacy-sensitive context.

Bauer and Lukowicz [6] researched behavior change during stress and no-stress period using a smartphone, leveraging data about the location of the user (using GPS and WiFi), his/her social interactions (using Bluetooth sensor), and calls/SMSes patterns. They involved 7 students during two exams weeks (stress period) and the following two holiday weeks (non stress period). By combining all the features from social interaction data, SMSs behavior and call patterns, they were able to detect a change in behavior of about 86% of testers for stress/no-stress situations.

Even if for different purpose, Gao et al. [9] analyzed if touchscreen interaction can be used to know its user's emotional state (Excited, Relaxed, Frustrated and Bored). They collected data about finger stroke behavior during gameplay with an iPod touch game. The main data collected were: coordinates of each stroke, the contact area and the total time. Accuracy in discriminating the four different emotional states was between 69% and 77%, while higher accuracy (89%) was achieved when discriminating between two levels of arousal and two levels of valence.

Differently from other approaches, we leverage only human-smartphone interaction data to assess the stress state of an individual. We analyze only data from 'touch', 'scroll', 'swipe' interactions and 'text input' behavior, without attaching any external device to the user or to the smartphone, thus lowering system intrusiveness. Moreover, no privacy-sensitive information are used with our approach, like user's calls, SMSes or location data. To the best of our knowledge, this is the first research inferring stress state in smartphone users considering only the way the users interact with their smartphone while performing usual tasks.

III. Protocol Design

The purpose of this work is to study the possibility of using information about the interaction that takes place between the user and his/her smartphone and to accurately assess his/her stress state. Smartphone is instrumented for measurement and collection of interaction data. Participants completed several exercises under the relaxed (no stress) and stressed state while being at home or in another comfortable and static (i.e. controlled) environment. This Section provides details about the protocol, the input data collected and what kind of interactions were analyzed.

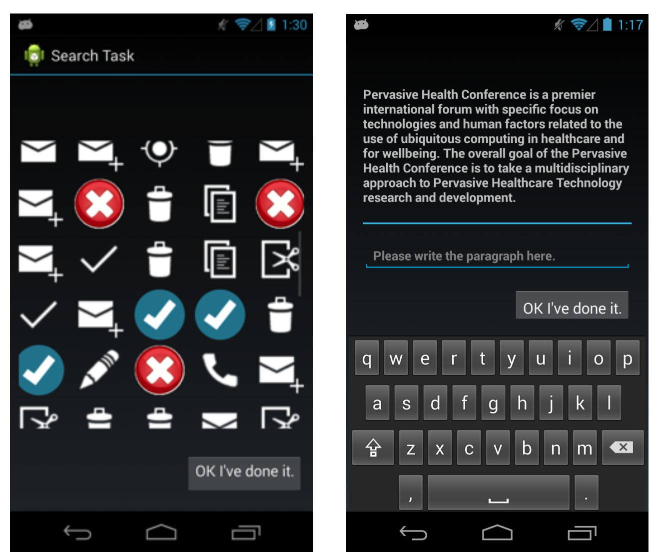
A. Human-Smartphone Interaction

Every time we pick up our smartphone to perform usual everyday activities, e.g., make a call, write a SMS, read/write email, what we essentially do, independently from the goal we are fulfilling, is that we are interacting with the smartphone. A smartphone interaction is defined as any contact with the screen that causes a change in the smartphone interface and its internal state. The main interaction types are: 'tap', 'double tap', 'long press', 'scroll', 'swipe', 'pinch', 'zoom' and 'rotate'. Some of them can be considered as "application dependent", with much lower daily occurrence than others. For this reason, we decided to focus only on the interactions that occur much more frequently and that are independent from the currently running application: 'tap', 'scroll' and 'swipe'.

Another common task performed with the smartphone is 'text input', i.e., writing SMS or email. As shown previously [7], keyboard typing can be leveraged as an indicator of stress. We expand this idea considering smartphone typing.

B. Test Tasks

To acquire information about 'tap', 'scroll', 'swipe' and 'text input' interactions, we developed several exercises (called *Tasks*) requiring the participants to perform pre-defined set of actions: the *Search Task* and the *Write Task*.



(a) Search task (b) Write task

Fig. 1. *Search* and *Write Task*

1) Search Task: In this task, the participant has to search all the repetitions (30) of three randomly chosen icons on the smartphone screen within a total of 300 icons (there are 10 different icons used). Since it is not possible to show all the icons on the screen at the same time, the participant has to 'scroll', i.e., execute a vertical movement on the screen, and 'swipe', i.e., execute an horizontal movement on the screen, to be able to inspect all the icons. Every time he/she clicks on one icon, there is a 'tap' interaction and an immediate feedback is provided about the correctness of the choice. Figure 1a presents a screenshot of the application during the *Search Task*.

2) Write Task: In this task, the participant is required to rewrite a paragraph that is shown on the screen. We decided to use English as text language, even with no native English speakers, since it does not have accents or other language-specific elements. Each sentence was composed only by words, commas and dots; the most used elements when writing SMSes or emails. The participants used our custom virtual keyboard. In this way, we ensure the same test conditions for all participants, i.e., some smartphone users have a custom keyboard to write, instead of the standard QWERTY one. Our keyboard does not provide auto correction or word suggestion, giving us the possibility to collect data about errors and user-based corrections during writing. During this task we collect 'tap' and 'text input' data. A screenshot is provided in Figure 1b.

C. Stressor Task

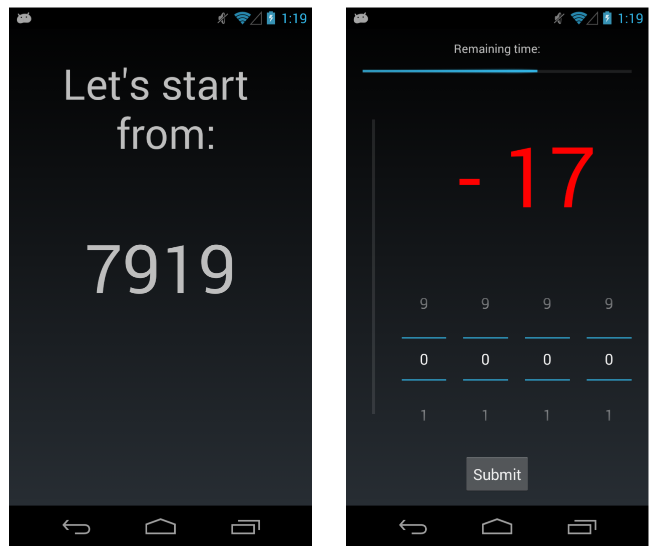
To measure the differences in interaction between a relaxed and a stress state, we stressed the participant through a particular task. Several papers in literature address the problem of how induce stress in people [10] [11]. The main stressors that can be used are:

- *Cognitive stressor*: mathematical and memory tests, e.g. starting from a big prime number and going down by 7 or 13, without the possibility of making notes;
- *Social pressure*: evaluation of the performance of the individual, in particular by an external person, e.g., via public speech;
- *Timing pressure*: giving a maximum amount of time to complete the task;
- *Random events*: generation of random events that could disturb the main user task, i.e., simulation of faults, unexpected results, etc.

We have build a smartphone task that combines all the stressors mentioned above. Therefore, participant is asked to make several mathematical calculations and to input the answer using the provided interface. We start from a random big prime number and the user has to calculate the subtraction of this number by 7 or 13 (randomly). To submit the answer the participant has a limited time. If he/she submits a correct answer, he/she will have to continue the task, and the available time will be further decreased. If the submitted answer is wrong, an annoying sound is played to underline the error, and the smartphone vibrates. After an error, the participant will have to restart from another prime number provided by the task, and we increase the available time to answer. Every correct answer let the participant earn points (that increase with the number of consecutive correct answers submitted), while wrong answers will decrease the current score. There is a predefined amount of points that let the participant finish the task (after a minimum amount of time).

To increase the stress of the individual, several random events can happen. For example, if more than 4 correct answers are submitted, the decrease number will change, randomly choosing over different possibilities, i.e., 5, 19, 21 etc. Furthermore, every two minutes there is a random possibility that the level of earned points necessary to stop the *Stressor Task* and proceed to the next step, are decreased by 10%.

The *Stressor Task* lasts at least five minutes, and ends if the participant has reached the minimum score necessary to proceed, or after 10 minutes, independently from the achieved score (this information is not provided to the participant). We used a maximum time for this task to avoid possible dropouts, i.e., avoid that participants decide not to continue the protocol because they are too frustrated with it. Figure 2 provides two screenshots of the *Stressor Task*.



(a) Starting prime number (b) Stressor task interface

Fig. 2. Stressor task

D. Stress Measurement

In order to evaluate whether the two modalities of the tasks (no-stress and stress mode) elicited the intended stress in participants, we request to report their state after each step of the protocol using the Experience Sampling Method (ESM) [12]. The inputs requested were: Valence, Energy (Arousal) and Stress levels on a 5-point Likert scale [13]. The provided survey is presented in Figure 3. From this survey, we expect that the higher stress values are associated with tasks in "stress mode", while more relaxed ones with the "relaxed tasks". Stress condition is usually associated with negative valence and higher energy.

E. Protocol Overview

The protocol is divided into 4 different phases:

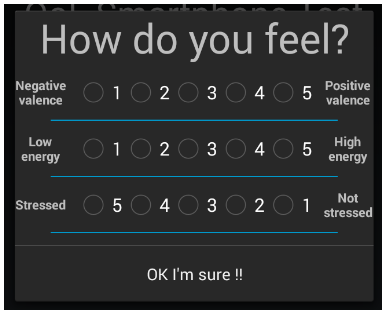


Fig. 3. ESM dialog prompted after each step of the protocol.

- “Relax phase”: participants experience some relaxing music with several relaxing images on the screen [14]. This step is used to destress participants and baseline their stress state;
- “Calm phase”: in this phase participants perform the *Search* and *Write Task* in a relaxed state. Each task is repeated three times to collect a sufficient amount of ‘scroll’, ‘swipe’ and ‘text input’ data. This phase has a variable duration, depending on the speed of the participant in finding the icons or writing the text;
- “Stressor phase”: in this phase participants experience the stress-inducing task. It will last between 5 and 10 minutes, as explained in Section III-C;
- “Stress phase”: the participants are now stressed from the previous task and they repeat the *Search* and *Write Task* in a “stress state”. Additionally, other stressors are applied, i.e., a tic-tac sound is played. The participant has to find all the possible icons, or to write the text, before the time for the task expires. Moreover, for the *Search Task*, if a wrong icon is selected, the available time is decreased and the smartphone vibrates, while if a correct one is selected, the time is increased. The usage of these stressors during the stress phase aim at keeping participants constantly stressed over all the second part of the protocol.

Figure 4 shows the overall protocol.

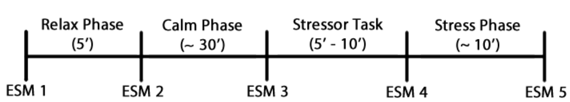


Fig. 4. Overview of the study protocol.

F. *Participants*

Participants for this study were recruited through email sent to several mailing list and to individuals already enrolled in other studies in our lab. Participants were informed that our research focused on building better user interfaces for smartphones in collaboration with Google. We did not inform the participants about stress analysis and recognition to avoid a bias of their attitude and their interaction styles. In the email, we defined the requirements to participate in the study; the only one was to be an Android OS smartphone owner and user and to have one hour of not interrupted time to follow the protocol from the beginning to the end. The approximate duration of the whole experiment was about one hour, depending on the skills/speed of the participant using his/her own smartphone, in particular during the *Write Task*.

In the recruitment email, we have explained all the steps necessary to start the application to execute the protocol and to correctly complete it. Individuals that decided to participate, first needed to download and install the application following the provided link. Before starting, they had to prepare themselves to correctly perform the exercises: be sure that they would not be disturbed for the following hour, sit on a chair where they could be comfortable, hold the smartphone as naturally as possible (keeping it in hand and not putting it on a table) and since there are associated sounds (music) in specific tasks, we have additionally advised participants to use an headset. After the test phase, we extract three different participants to receive a 50 CHF Amazon Gift Card to thank them for their participation.

In total, 13 participants (7 males and 6 females) answered to our participation request and completed the experiment. The average age was 26,38 ($\pm 2, 53$) with a minimum of 22 and a maximum of 32 years old. The participants differed in level of skill in English speaking/writing and in smartphone usage. We let the participants use their own smartphone to avoid that a new one during the experiment could cause stress in them.

IV. Data collection

A. *Object and action*

Along each task of the protocol, the main data we collect is the position of object(s) relevant for the task on the smartphone screen. The information collected is stored as an n-tuple:

(ID, type, x, y, width, height, visibility, text)

where:

- *ID*: is the ID of the object;
- *type*: the type of the object, i.e., button, icon, text etc.;
- *x*: x-coordinate of the top left corner of the object;
- *y*: y-coordinate of the top left corner of the object;
- *width*: the width of the object (in pixels);
- *height*: the height of the object (in pixels);
- *visibility*: if the object is visible or hidden;
- *text*: the text of the object (if any).

Given this information, together with information about the device type, i.e., model, screen size and resolution, we are able to reconstruct the state of the screen while the participant was performing the tasks, and understand for example how a participant’s clicks’ on the screen are related to the position of the objects on the screen.

Every touch interaction, that could be a ‘tap’, ‘scroll’, ‘swipe’ or ‘text input’ is made up by a set of *action* event. The first action event has its flag field set to “ACTION_DOWN”, indicating a press of the finger on the screen. After that, there are zero or more “ACTION_MOVE” actions and the final “ACTION_UP” flag that indicates the end of the interaction. This set is the same for all the possible interactions we are recording. An example of the sequence of *action* readings for the ‘scroll’ interaction is provided in Figure 5.

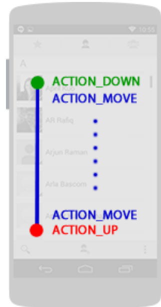


Fig. 5. Example sequence of *action* readings during a 'scroll' interaction.

Each action is correlated with a set of information, and is stored as an n-tuple:

(action, timestamp, x, y, size, pressure, firstTimestamp)

where:

- *action*: a flag to distinguish between "ACTION_DOWN", "ACTION_MOVE" or "ACTION_UP";
- *timestamp*: timestamp of the recorded *action* (in milliseconds);
- x: x-coordinate of the center of the touch point on the screen from the top left corner (in pixels);
- y: y-coordinate of the center of the touch point on the screen from the top left corner (in pixels);
- *pressure*: the pressure applied with the finger, where pressure $\in]0,1] \in R$ (unit less, granularity $0,01$);
- *size*: a measure of the contact surface of the finger with the screen, where size $\in]0,1] \in R$ (unit less, granularity $0,01$);
- *firstTimestamp*: the timestamp of the first action ("ACTION_DOWN") of the interaction.

One may notice that 'text input' and 'tap' interaction should have no "ACTION_MOVE" flags, since there is no movement with the finger between the press (i.e. "ACTION_DOWN") and the release (i.e. "ACTION_UP") of the finger on the screen, while 'scroll' and 'swipe' interactions have a variable number of them. However, we noted that even in case of a 'tap' there is the possibility that the participant makes a small movement, not releasing the finger in the same point where he/she initially started to touch on the screen. As we will present later, we also investigate if this movement is significantly different during stress state for each individual.

B. Feature Extraction

Once all the 'object' and 'action' data was collected from participants, we extracted several features from the *Search* and the *Write Task*.

1) Search Task features: For the *Search Task*, we collect feature both for the 'tap' interaction and for the 'swipe' and 'scroll' ones. For the 'tap' interactions, i.e., when clicking on icons, we consider standard features like *average pressure*, *average size*, *movement* (how much the 'tap' moves when touching on the screen) and *time duration*. For the 'scroll' and 'swipe' interaction, we were interested in features that could describe in more details these kind of interactions with the smartphone. In particular, we calculated information about interaction *time* and *space length*, interaction *speed* ($\frac{space\_movement}{time\_length}$), mean distance of the interaction from the center or the top left corner of the screen and the overall *linearity* of the interaction.

We must note here that a *scroll/swipe interaction* is different from a *scroll/swipe movement*. The first is the action performed by the participant to 'scroll'/'swipe', meaning the sequence of press on the screen ("ACTION_DOWN"), movement on the screen ("ACTION_MOVE") and the final release ("ACTION_UP"). Each *scroll/swipe interaction* generates the sequence of actions and n-tuples presented before, and the features referred to the *scroll/swipe interaction* refers to data calculated from this set of actions.

The *scroll/swipe movement* is the movement that occurs on a scrollable object on the screen, i. e., the movement of the interface, that changes the visible elements on the screen according to the direction and the speed of the 'scroll'/'swipe'. The *scroll/swipe movement* is the resulting interface movement that generates data about how much the interface is changing on the screen.

The linearity of the 'scroll' or 'swipe' interaction is derived in two different ways. The first way is the following one:

$$\text{Scroll linearity} = \frac{|x_l - x_f|}{W} \text{ and Swipe linearity} = \frac{|y_l - y_f|}{H}$$

where:

- x_f and y_f are respectively the x and y coordinates of the first point of the 'scroll'/'swipe' (in pixels);
- x_l and y_l are respectively the x and y coordinates of the last point of the 'scroll'/'swipe' (in pixels);
- W and H are respectively the smartphone screen width and height (in pixels).

Since a 'scroll'/'swipe' interaction is a set of points $S = \{(x_{1..n}, y_{1..n}) | n \in N\}$, it is possible to calculate the linearity of a scroll/swipe interaction as sum of every individual point. Based on which, we can calculate the linearity as:

$$\text{Scroll linearity} = \frac{\sum_{i=1}^{n} |(x_i - x_{i-1})|}{W}$$

and

$$\text{Swipe linearity} = \frac{\sum_{i=1}^{n} |(y_i - y_{i-1})|}{H}$$

Since we let users use their own smartphone, in order to cope with different screen resolutions, all features related to space length of the 'scroll' or 'swipe' interaction are normalized with the width or height resolution of the screen.

Table I presents all the features calculated from the data acquired during the *Search Task*. The first three features are calculated for each 'tap' interaction (when clicking on a particular icon on the screen). For the other rows of the Table, when writing "Scroll/Swipe ..." means that that feature is calculated for both the 'scroll' and the 'swipe" interaction.

2) Write Task features: In addition to the standard features about 'tap' interaction already explained for the *Search Task* (*Pressure*, *Tap size* and *Tap movement*), with the *Write Task* we employ other features that are related to 'text input'.

TABLE I. Features derived for the *Search Task*

Feature	Description
Tap pressure	Pressure applied when 'tapping' on icons on the screen (value $\in]0,1] \in R$)
Tap size	Size of the touch when 'tapping' on screen (value $\in]0,1] \in R$)
Tap movement	Movement of the touch when 'tapping' on icons (in pixels)
Scroll/Swipe Average pressure	Average pressure applied while 'scrolling'/'swiping' (value $\in]0,1] \in R$)
Scroll/Swipe average size	Average size of the touch surface while 'scrolling'/'swiping' (value $\in]0,1] \in R$)
Scroll/Swipe delta	Space movement of the 'scroll'/'swipe' (in pixels)
Scroll/Swipe interaction length	Space length of the interaction used to 'scroll'/'swipe' (in pixels)
Scroll/Swipe delta speed	Speed of the 'scroll'/'swipe' movement (in pixels / milliseconds)
Scroll/Swipe interaction speed	Speed of the interaction movement to 'scroll'/'swipe' (in pixels / milliseconds)
Scroll/Swipe distance from screen center	Distance of the interaction from the center of the screen (in pixels)
Scroll/Swipe distance from top left corner of the screen	Distance of the interaction from the top left corner of the screen (in pixels)
Scroll/Swipe linearity	Linearity of scroll/swipe as horizontal/vertical distance between first and last point (unit less)
Scroll/Swipe linearity as sum of every point	Linearity of scroll/swipe as sum of the horizontal/vertical distance between two consecutive points (unit less)

Specifically, we investigate the *number of errors*, the *number of corrections* and the *digits frequency*. Table II explains the features derived for the *Write Task*.

V. Results

In this Section we provide the statistical analysis of the data collected in our studies, both for the perceived stress data from participants' Experience Sampling Method, and from the features extracted along data collected in the two tasks.

A. Stress induction analysis

The first thing we analyzed is if the *Stressor Task* achieved its goal, i.e., to increase the participants' perceived stress. Moreover, we investigated if there is a significant difference between tasks performed in calm and stress state. Recalling our protocol (Section III-E), participants have to fill a survey after each main step, providing their Valence, Energy and Stress state. Mean values of all the participants are reported in Table III. To understand if during our protocol the perceived stress by participants changed accordingly to the different task, we compared stress readings between ESM 3 and ESM 4 (before and after the *Stressor Task*) and between ESM 3 and ESM 5 (after the "Calm Phase" and after the "Stress Phase"), that we assume are statistically different, and between ESM 4 and ESM 5 (at the beginning and at the end of the "Stress Phase"), that we assume are not statistically different since the perceived stress should not change.

TABLE III. Valence, Energy and Stress perceived by participants among the protocol phases

Phase	Step	Valence (1÷5)	Energy (1÷5)	Stress (1÷5)
Before Relax phase	**ESM 1**	3,8 ± 0,9	3,2 ± 1,1	2,5 ± 1,3
Before Calm phase	**ESM 2**	4,2 ± 0,7	3,7 ± 0,8	1,8 ± 0,8
Before Stressor	**ESM 3**	3,6 ± 1,0	3,2 ± 1,0	2,5 ± 1,2
After Stressor	**ESM 4**	2,6 ± 1,1	3,1 ± 1,0	3,6 ± 1,1
After Stress phase	**ESM 5**	2,8 ± 1,2	2,7 ± 1,0	3,7 ± 1,2

A paired, one-tailed t-test was applied to the "Stress" value of the ESM surveys. Comparison between ESM 3 and ESM 4 and between ESM 3 and ESM 5 have proved to be significantly different, while ESM 4 and ESM 5 were not. Result of the test are provided in Table IV, and significance level is $p = 0.05$.

TABLE IV. Significance test between Stress value in survey of participants at different steps of the Protocol. (*) indicates significance at 0.05

Test	t(13)	*p*-value
ESM 3 vs ESM 4	1,99	0,007 *
ESM 3 vs ESM 5	-2,84	0,009 *
ESM 4 vs ESM 5	-2,74	0,5

As we can see from our results, stress values were significantly different between ESM 3 and ESM 4 and between ESM 3 and ESM 5. That means that the stress state differs from calm state: participants were stressed. Moreover, difference between ESM 4 and ESM 5 was not significantly different (*p-value* $= 0.5$), which means that the perceived stress after the *Stressor Task* and the "Stress Phase" has not changed significantly. It means that our *Stressor Task* was effective and the stress state of the participant was not changed during the "Stress Phase".

B. Features analysis

Once proved that our *Stressor Task* (and the following "Stress Phase") effectively increased the perceived stress in participants, we analyze our features to understand how indicative they are for stress state assessment.

1) Search Task: For the *Search Task*, we collected a total amount of 2937 instances, 1790 for the "scroll" interaction and 1147 or the "swipe" interaction. On average, each participant provided 225 instances. The first analysis we did was to identify features statistically different between no-stress and stress tasks. We built both a within-user model (considering only data from a single participant), and a global model, where we put together all the average values for all participants. However, none of the features that we considered had a significant correlation in the global model (at the significance level of 0.05), but only a weak correlation. For example, the "average swipe pressure" (*p-value*$=0,09$, $t(13) = 1,53$), "scroll distance from center" (*p-value*$=0,065$, $t(13) = 1,65$) and

TABLE II. FEATURES DERIVED FOR THE *Write Task*

Feature	Description
Tap pressure	Pressure applied when 'tapping' on a digit (value $\in]0,1] \in R$)
Tap size	Size of the touch when 'tapping' on a digit (value $\in]0,1] \in R$)
Pressure / Size	Ratio between pressure applied and size of the touch (unit less)
Tap movement	Movement of the touch when clicking on a digit (in pixels)
Tap duration	Time length of the touch when clicking on a digit (in milliseconds)
Tap distance	Time distance between touching two consecutive digits (in milliseconds)
Back / All digits	Ratio between BACK digits and all the clicked digits
Wrong Words / All Words	Ratio between wrongly written words and all the words of the sentence

"scroll distance from top left" (*p-value*=$0,07$, $t(13) = 1,57$). Moreover, when considering significance applied within-user, only "scroll interaction length" was significantly different for 61% of participants. All other features were significantly different for less than 50% of participants, i.e., "pressure", "scroll time length" and "linearity" for about 30%, "average pressure" for about 40% and "scroll linearity as sum of every point" for 45% of participants.

Due to these results, instead of evaluating statistical significance between no-stress and stress state for each feature, we decided to use machine learning techniques to evaluate the precision of a stress prediction model. Even in this case, we have build a within-user model and a global model, both for 'scroll' and 'swipe' interactions. To evaluate our features we used the Weka Software [15] for machine learning with different classifiers: Decision Tree (DT), k-Nearest Neighborhood (kNN), Bayes Network (BN), Support Vector Machine (SVM) and Neural Networks (NN). 10-Fold Cross-Validation was used to evaluate the model build for each participant, while Leave-one-out was used for the global model. Table V and Table VI report F-measure evaluation for each model, both for 'scroll' and 'swipe' interaction. Please note that some models for 'swipe' interaction for some participants are missing since during the first step of the protocol these participants did not understand how to 'swipe', and did not provide a sufficient amount of data for analysis.

TABLE V. EVALUATION OF 'SCROLL' INTERACTION CLASSIFICATION BOTH FOR WITHIN-USER AND GLOBAL MODEL.

	Scroll F-measure				
User / Model	**DT**	**kNN**	**SVM**	**NN**	**BN**
1	0,852	0,874	0,769	0,879	0,91
2	0,771	0,825	0,812	0,869	0,763
3	0,764	0,694	0,71	0,673	0,749
4	0,664	0,779	0,809	0,725	0,675
5	0,77	0,819	0,774	0,774	0,73
6	0,832	0,894	0,868	0,87	0,873
7	0,798	0,769	0,835	0,767	0,835
8	0,645	0,605	0,636	0,633	0,572
9	0,921	0,941	0,875	0,915	0,91
10	0,918	0,932	0,945	0,966	0,897
11	0,653	0,672	0,752	0,672	0,492
12	0,714	0,701	0,789	0,77	0,574
13	0,957	0,9	0,914	0,936	0,986
Global	**0,73**	**0,71**	**0,78**	**0,74**	**0,67**

On average, the F-measure of the 'scroll' within-user interaction model is 0,79 $\pm$ 0,02, while for the swipe interaction is 0,85 $\pm$ 0,03. kNN seems to be the most accurate technique for classification.

TABLE VI. EVALUATION OF 'SWIPE' INTERACTION CLASSIFICATION BOTH FOR WITHIN-USER AND GLOBAL MODEL.

	Swipe F-measure				
User / Model	**DT**	**kNN**	**SVM**	**NN**	**BN**
2	0,811	0,843	0,833	0,858	0,833
5	0,883	0,742	0,68	0,711	0,84
6	0,867	0,865	0,69	0,87	0,859
9	0,989	0,978	0,909	0,989	0,968
10	0,958	0,98	0,79	0,958	0,958
11	0,709	0,874	0,773	0,824	0,669
12	0,711	0,756	0,749	0,825	0,716
13	0,958	0,872	0,906	0,926	0,967
Global	**0,92**	**0,75**	**0,81**	**0,82**	**0,77**

Results from our analysis show that 'scroll' and 'swipe' interaction features are accurate indicators for stress assessment and could be used to implement a real-time stress assessment service that runs on a smartphone in background without affecting the user's behavior and his/her interactions with the device.

Considering the global model for 'swipe' and 'scroll' interaction, we investigate which features are the more predictive ones to build our model. We evaluate them based on their *Information Gain* with respect to the classification problem. Table VII reports the final rank of the most informative features for 'scroll' and 'swipe' interaction respectively.

TABLE VII. RANK OF 'SCROLL' AND 'SWIPE' FEATURES BASED ON THEIR *Information Gain*.

Rank	Scroll Feature	Swipe Feature
1	Interaction length	Interaction length
2	Time length	Swipe delta
3	Mean distance from center	Average touch size
4	Scroll delta speed	Average touch pressure
5	Scroll delta	Mean distance from center
6	Speed interaction	Time length
7	Mean distance from top left	Linearity as sum every point
8	Linearity as sum every point	Mean distance from top left

From Table VII we conclude that "interaction length", "mean distance from center", "time" and "linearity" are predictive across both types of interactions, and are computationally cheap to derive.

2) Write Task: Applying machine-learning techniques to each single 'digit' or 'tap' interaction during the 'text input' task on the smartphone is not a practical operation, since each single 'tap' is a too short operation to extract a sufficient amount of data for analysis. Therefore, we have decided to consider only statistical difference between no-stress and stress data for the features explained above (Table II). In this case we performed both a within-user and a global model test.

Considering within-user model, results show that "digits size" (the area clicked by the finger on the screen) is significantly different for 64% of the participants and the "ratio between pressure and size" for 55% of participants. Averaging data for each participant and building a global model, significance analysis shows that "ratio between wrong words and all the words to write of the paragraph" (*p-value* = $0,028$, $t(13) = -2,15$) and the "digits time distance" (*p-value* = $0,012$, $t(13) = 2,67$) prove to have a significant difference between no-stress and stress condition, meaning that in stress state we tend to write faster but with more errors. Finally, "digits duration" feature, e.g., the time length the finger presses on the screen to tap on the digit, has only weak correlation between stress and no stress condition (*p-value* = 0,08, $t(13) = 1,51$).

The acquired results are promising and motivate us make further analysis of stress vs no-stress state assessment in the future. For example, since the number of misspelled words increases during the stress condition, auto-correction feature of smartphone keyboards would correct much more frequently text messages with respect to a no-stress state. Moreover, simply measuring the time elapsed between two input (digits) may serve as stress assessment method during the text input.

VI. CONCLUSIONS

Ubiquitous stress assessment and monitoring is an increasingly interesting research field, since unhealthy stress condition is uprising and decreasing the intrusiveness of novel technologies for stress assessment would increase individuals acceptance for them.

With this goal in mind, we have researched the possibility of using human-smartphone interaction as an indicator of stress state in users. We investigate the main interactions with a smartphone, i.e., 'scroll', 'swipe', 'touch' and 'text input'. We developed an Android application that implemented a protocol along which individuals interacted with the smartphone both in relaxed and in stress state. We have evaluated data from 13 participants, building both a within-user model and a global one with data from all the participants.

As a result of our investigation, we conclude that 'scroll' and 'swipe' interaction can be used for stress classification with an F-measure between $79\% \pm 0,02$ and $85\% \pm 0,03$ if building a separate model for each user, or a F-measure of about $73\% \pm 0,04$ ('scroll') and $81\% \pm 0,06$ ('swipe') for the global model. Moreover, we have investigated statistical difference in 'text input' features between no-stress and stress state, identifying information like 'the number of errors made during writing' and the 'digits' (tap interactions over a virtual keyboard) frequency as features with a statistical significance for stress assessment. To the best of our knowledge, this is an unique approach focusing on human-smartphone interaction, and especially on 'swipe', 'scroll' and 'text input', proving to accurately enabling to assess stress state in smartphone users.

As future research, we plan to investigate the possibility to assess stress state of the individuals "in the wild", i.e., in their everyday life. In particular, we plan to analyze how their smartphone interaction patterns change between no-stress and stress state, without employing any external sensor and without accessing privacy-sensitive information (e.g., location).

ACKNOWLEDGMENT

This research is supported by the Swiss NSF PCS-OBEY (SNSF-149591) and MIQModel (SNSF-157003), AAL MyGuardian (AAL-2011-4-027), ANIMATE (AAL-2013-6-071) and CoME (AAL-2014-7-127) projects, and COST actions (1003, 1303 and 1304).

REFERENCES

[1] S. Cohen, R. Kessler, and L. Gordon, *Measuring Stress: A Guide for Health and Social Scientists*, ser. Guide for Health and Social Scientists, 1997.

[2] A. Perkins, "Saving money by reducing stress," *Harvard Business Review*, 1994.

[3] "Are teens adopting adults' stress habits?" http://www.apa.org/news/press/releases/stress/2013/teenstress.aspx, accessed: 2014-12-15.

[4] D. Sun, P. Paredes, and J. Canny, "Moustress: Detecting stress from mouse motion," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2014, pp. 61–70.

[5] J. Hernandez, R. R. Morris, and R. W. Picard, "Call center stress recognition with person-specific models," in *Proceedings of ACII*, 2011, pp. 125–134.

[6] G. Bauer and P. Lukowicz, "Can smartphones detect stress-related changes in the behaviour of individuals?" in *Proceedings of IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, 2012, pp. 423–426.

[7] J. Hernandez, P. Paredes, A. Roseway, and M. Czerwinski, "Under pressure: Sensing stress of computer users," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2014, pp. 51–60.

[8] A. Sano and R. W. Picard, "Stress recognition using wearable sensors and mobile phones," in *Proceedings of ACII*, 2013, pp. 671–676.

[9] Y. Gao, N. Bianchi-Berthouze, and H. Meng, "What does touch tell us about emotions in touchscreen-based gameplay?" *ACM Transactions on Computer-Human Interaction*, pp. 31:1–31:30, 2012.

[10] L. Bernardi, J. Wdowczyk-Szulc, C. Valenti, S. Castoldi, C. Passino, G. Spadacini, and P. Sleight, "Effects of controlled breathing, mental activity and mental stress with or without verbalization on heart rate variability," *Journal of the American College of Cardiology*, pp. 1462–1469, 2000.

[11] S. S. Dickerson and M. E. Kemeny, "Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research," *Psychological bulletin*, pp. 355–391, 2004.

[12] R. Larson and Csikszentmihalyi, "The experience sampling method," *New Directions for Methodology of Social & Behavioral Science*, 1983.

[13] R. Likert, "A technique for the measurements of attitudes," *Archives of psychology*, p. 55, 1932.

[14] E. Dan-Glauser and K. Scherer, "The geneva affective picture database (gaped): a new 730-picture database focusing on valence and normative significance," *Behavior Research Methods*, 2011.

[15] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: An update," *SIGKDD Explorations*, 2009.