This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2021.3061417, IEEE Transactions on Affective Computing

Towards Participant-Independent Stress Detection Using Instrumented Peripherals

Dennis R. da C. Silva, Zelun Wang, and Ricardo Gutierrez-Osuna, *Senior* Member, IEEE

Abstract—Methods to measure work stress generally rely on subjective measures from questionnaires or require dedicated sensors that are cumbersome to wear and interfere with the task. To address this problem, we propose a method to detect stress unobtrusively using commodity devices (keyboards, mice) instrumented with pressure sensors. We propose a minimalist design that can be easily replicated by other researchers using off-the-shelf and low-cost hardware. We validate the design in a laboratory experiment that simulates office tasks and mild stressors while avoiding methodological limitations of previous studies. We compare stress-detection performance when using conventional features reported in the literature (keystroke dynamics, mouse trajectories) augmented with information from pressure sensors. Our results indicate that pressure provides additional information for stress discrimination; adding pressure information to keystroke dynamics and mouse trajectories improves classification performance by 6% and 3%, respectively. These results show how devices that are already part of the modern workplace may be used and enhanced to automatically and unobtrusively detect stress.

Index Terms— Stress detection; pressure-sensitive keyboard; pressure-sensitive mouse; keystroke dynamics; mouse dynamics; affective computing

1 INTRODUCTION

Work stress is dramatically increasing as a result of rising competitiveness, more intense workloads, and longer and harder working hours [1, 2]. Although stress can help people stay focused and motivated, excessive physical and psychological demands can lead to severe stress episodes, putting employees at a higher risk for health problems [3]. For example, acute stress exacerbates negative coping behaviors, such as smoking [4] and substance abuse [5], and can also lead to depression [6].

Monitoring stress levels throughout the day may allow employees to identify stress triggers and stressful episodes early on and develop healthier coping strategies [7]. The gold standard for monitoring stress objectively is by measuring stress hormones (e.g., cortisol, alpha-amylase) from saliva samples [8]. However, this method is impractical to be deployed in the workplace setting and only provides a single-point measurement rather than a continuous measure. Another approach for measuring stress levels consists of using self-report instruments [9, 10]. Unfortunately, these instruments are sensitive to subjective biases and also only provide single point measurements. More recently, wearable sensors are being used to measure physiological markers such as heart rate variability and skin conductance, which correlate with stress [11, 12]. However, the most common among these measures (wrist-based heart rate and skin conductivity) are sensitive to motion artifacts, which can be caused by physical activity (e.g., walking) or even typing. Contactless measures, such as facial expression analysis from webcams [13], can also be used but are subject to changes in illumination, differences in skin tones, among others.

Several studies have explored the possibility of monitoring stress indirectly by analyzing keyboard and mouse use patterns [14-16]. Keystroke and mouse dynamics have long been used for user authentication [17, 18] and recently to infer emotional state [15, 19, 20]. Most of these studies use timing and latency information, which can be easily obtained from off-the-shelf devices. Studies have also explored the use of experimental keyboards and mice to predict stress. For example, Hernandez et al. [21] found greater typing pressure and mouse grip pressure when subjectively-rated stress and electrodermal activity levels were higher. This suggests that additional stress-related information may be obtained by instrumenting keyboards and mice with sensors. Another benefit of using peripherals is their ubiquity, with around 90 million new desktops shipped in 2018 alone [22].

This study presents two low-cost designs to measure typing pressure and mouse-grip pressure from off-the-shelf devices. Our designs use force-sensitive resistors placed on keyboards and mice to record changes in pressure. To evaluate our design, we conducted a user study aimed at detecting stress while participants completed two conventional tasks in knowledge work: typing texts and filling out multiple-choice questionnaires. Then, we trained binary classifiers to discriminate stress vs. neutral states using features derived from keystroke and mouse dynamics, and from our pressure measurements. We obtained higher classification accuracy when combining keystroke and mouse dynamics with their corresponding pressure features.

- *D.R. da Cunha Silva is with the Computer Science and Engineering Department at Texas A&M University, College Station, TX 77843. E-mail: silva.dennis@tamu.edu.*
- *Z. Wang is with the Computer Science and Engineering Department at Texas A&M University, College Station, TX 77843. E-mail: wang14359@tamu.edu*
- *R. Gutierrez-Osuna is with the Computer Science and Engineering Department at Texas A&M University, College Station, TX 77843. E-mail: rgutier@cse.tamu.edu.*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2021.3061417, IEEE Transactions on Affective Computing

The rest of the paper is organized as follows. First, we discuss related work on using keystroke and mouse information to recognize emotion. Next, we present our keyboard and mouse designs, as well as the experimental protocol. Finally, we present results from the user studies, followed by a discussion of findings and conclusions.

2 RELATED WORK

Various sensing modalities have been used for emotion recognition, including facial expression and speech [23], physiological sensors [24], and thermal and visual imaging [11, 25]. Alternatively, some approaches have relied on changes in behavior (e.g., keyboard and mouse usage [26], linguistics [15], posture [27]) that may be affected by the user's emotional state. In an early study, Zimmerman et al. [28] provided a rationale for assessing user affect using keyboard and mouse. Following this seminal work, dozens of publications have investigated how these commodity devices can be used to infer user affect.

2.1 Emotion Detection Using Keystroke and Mouse Dynamics

A number of features from keystroke dynamics have been explored, including typing speed, latency, and pause frequency, to mention a few. Banerjee et al. [29] found that individual keystroke patterns vary over time and are also affected by the user's emotional and cognitive states (e.g., reduced typing speed when in a negative emotional state). Motivated by these findings, Tsihrintzis et al. [30] used keystroke features to improve visual-facial emotion recognition. They showed that recognition of anger and sadness was greatly improved by adding keystroke features.

Several studies have used keystroke dynamics to differentiate among a wider number of emotions. In most cases, the studies were conducted in a laboratory setting, but a few studies sought to capture natural behaviors while participants performed daily tasks in the wild [19, 31]. As an example of an in-situ experiment, Epp et al. [19] used keystroke dynamics features to model data collected from 15 different emotional states. In their user studies, 26 participants had keystroke information logged for an average of four weeks. For each sample logged, participants also reported their emotion using self-reports. Their best models achieved accuracies of 77-88% when classifying confidence, hesitance, nervousness, relaxation, sadness, and tiredness.

While in-situ studies can capture more realistic interactions, they are subject to uncontrolled external factors. For this reason, the majority of emotion-detection literature has relied on lab studies. Khanna and Sasikumar [32] used keystroke features to differentiate between positive, negative, and neutral emotional states. According to their findings, most people tend to type more slowly when in a negative emotional state and faster while in a positive emotional state. In a related study on typing patterns, Bixler and D'Mello [33] used task appraisals and stable traits to differentiate bored, engaged, and neutral emotional states. Their model achieved 56% accuracy.

Several studies have focused on differentiating between low and high cognitive load conditions based on keystroke and mouse dynamics [34-40]. As an example, Lim et al. [26] used both keystroke and mouse features to detect cognitive load induced by time pressure and mental-arithmetic problems. They found that when problem difficulty increases, task error, task duration, stress perception, and mouse idle duration also increase, whereas mouse speed, left mouse click rate, and typing speed decrease. Brizan et al. [36] have explored the use of keystroke dynamics combined with linguistics to predict cognitive load levels. In their experiments, participants were asked to type freely when asked to answer questions that elicited six different levels of cognitive load. Their models were able to differentiate the six cognitive load levels with above-chance accuracy, and their best performing models achieved 72% classification accuracy when differentiating behavior elicited by the more extreme cognitive load inducing prompts (level one vs. level six).

A good number of studies that use peripherals to detect cognitive load have focused on the sole use of mouse dynamics [35, 37-39]. For example, Chen et al. [35] studied the effects of cognitive load while participants performed the primary task of screening participants for a fictitious human resource department. Cognitive load was elicited by a secondary task, which popped-up on the user's screen and required a classification action. They reported that when under high cognitive load, participants presented more frequent contemplation (i.e., from 1 to 5 seconds) and hesitation (i.e., from 0.5 to 1 second) pauses in mouse activity, which was attributed to hesitant/cautious behavior. Grimes and Valacich [38] have used mouse dynamics to detect low, medium, and high levels of cognitive load. The authors elicited medium and high cognitive load using lag-1 and lag-2 number recall tasks, respectively. They observed higher mouse distance traveled, more frequent direction changes, and lower mouse speed during tasks performed under higher cognitive loads.

In contrast with laboratory experiments that rely on emotion-elicitation procedures, Gunawardhane et al. [14] collected non-stress behaviors when participants (college students) had no exam pressure, and during exam week. In their study, keystroke features were extracted while participants solved arithmetic problems. The authors found significant differences in several features, such as the duration of certain bigraphs and trigraphs, when comparing stressed and non-stressed emotional states. In a recent work, Lau [20] compared the efficacy of personalized and generic models to predict stress from keystroke dynamics. The author used a baseline-stressor-recovery design, where stress was elicited using multi-tasking and social evaluative threats. The personalized models obtained accuracies in the range of 83-92%, while the generic models achieved chance-level accuracy.

Although most of the studies reported in the literature employ a single-day experimental procedure, a few works analyzed how keystroke features generalize over multiple sessions [15, 41]. Vizer and Sears, for example, compared personalized and generic models in discriminating high and low cognitive demand using keystroke and linguistic features [41]. In their study, participants were asked to write freely about any topic either in the presence of a stressor (mental arithmetic tasks) or without it. Their generic model achieved 66% accuracy while their personalized model reached accuracies in the range of 65-93%.

Some studies have focused exclusively on mouse dynamics to perform emotion recognition. Yamauchi [42] investigated the relation between mouse activity and state anxiety. In the study, participants performed a task where they had to select and click geometric figures based on their similarities. The author extracted mouse features such as velocity and directional change, and fed them to a support-vector-regression model to predict state anxiety scores measured from questionnaires. He found that correlation coefficients between predicted and observed state anxiety scores were significantly higher than zero. Sun et al. [16] modeled the arm-hand dynamics as a mass-spring-damper system to study muscle stiffness during mouse movement. Their participants performed a set of abstract mouse tasks that involved pointing and clicking, dragging and dropping, and steering the mouse cursor through a tunnel. The authors used mental arithmetic to induce stress and mindfulness meditation to induce relaxation. They found higher damping frequency and lower damping ratio when participants were stressed.

In a recent study, Hibbeln et al. [43] investigated the relationship between mouse movement and negative emotion. They induced negative emotion by introducing delays and errors into time-limited tasks. The authors found that mouse movement distance increased and mouse speed decreased during the tasks. They explained this phenomenon in terms of attentional control theory, which suggests that negative emotion decreases attention control, shifting cognitive resources from goals to distractions.

Keystroke and mouse dynamics features are easy to extract and require no specialized hardware. For this reason, they have been used extensively in emotion recognition, and show promise as an approach to measure stress in the workplace.

2.2 Emotion Detection Using Instrumented Keyboard and Mouse

Interestingly, researchers have found that the pressure the user applies to the keyboard and mouse can provide additional emotion-related information. In a study by Tsihrintzis et al. [30], 65% of the participants reported typing harder when angry, whereas Karunaratne et al. [44] found that 15% of participants reported an increase in typing pressure when under stress. A few works have also observed variations in mouse grip pressure when experiencing different emotions. Picard et al. [45], for example, observed an increase in mouse grip pressure when participants were frustrated.

Prior studies have found that mental stress increases arm muscle activity and muscle tension [46, 47]. As such, pressure sensors could capture these changes and provide additional features to assist with automatic emotion detection. However, to the best of our knowledge, there are currently no keyboards or mice embedded with pressure sensors available on the market and little research has been reported regarding this type of device.

To our knowledge, the first work on instrumenting a computer mouse with pressure sensors dates back to 1993 [48]. In this work, the authors built a force-sensing mouse to investigate injuries related to intensive mouse use. The authors used foil strain gauges to measure finger forces applied to the mouse sides and buttons. They analyzed the applied force to distinguish between different activities, such as holding, moving, and dragging. In a later study [49], the authors recruited 16 subjects to test their proposed force-sensing mouse. They collected mouse data while participants performed their daily work in a field setting, and standardized tasks (e.g., pointing, dragging) in a lab setting. The authors observed that changes in applied force were task- and setting-dependent, but not time-dependent. In a subsequent study [50], the authors delivered stress by using time pressure and verbal provocation during a text editing task. They collected finger forces with their mouse, and physiological measures and subjective ratings of stress. They found higher forces applied to mouse buttons and more repetitive wrist movements during stress compared to a control condition.

In 2001, Qi et al. [45] instrumented a computer mouse with eight pressure sensors. They asked participants to fill out a web form and delivered a fictitious data-loss problem at submission time by erasing all the content they had filled out, with the goal of inducing frustration. Since participants had limited time to complete the task, they also experienced time pressure the second time they filled out the form. Initial tests on a limited number of participants were promising, achieving 88% classification accuracy.

In 2009, Dietz et al. [51] proposed an experimental keyboard design capable of sensing the force level at every depressed key by means of a pressure-sensitive membrane. In subsequent work, Hernandez et al. [21] used that experimental keyboard as well as a Microsoft Touch Mouse (a mouse with capacitance sensors on its surface) to analyze how typing pressure and mouse grip pressure change under stress. The authors collected data from 24 participants performing typing tasks and mouse-clicking tasks under relaxed and stressed conditions. They observed significantly higher typing pressure when comparing the stressful condition to the relaxed condition, for around 85% percent of participants. They also found increased capacitance value on the mouse for 75% of the participants, which indicates an increased hand contact area on the mouse surface. However, they did not report how these results compare to using traditional keystroke analysis for stress detection.

2.3 Emotion Detection Using Mobile Devices

It is estimated that currently 3.5 billion people use smartphones [52]. While first-generation smartphones and tablets were only equipped with simple applications (e.g., email, limited web browsers), current mobile computing power and capability are comparable to personal

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2021.3061417, IEEE Transactions on Affective Computing

computers. As a result, mobile devices have become an integral part of modern life [53]. For this reason, researchers have also investigated how to recognize emotion and mood disorders from mobile device usage data [54-61].

A number of studies have investigated how typing behavior on mobile devices can be used to recognize emotions [54, 60]. For example, Ghosh et al. [54] conducted a field study where they recorded participants' keystrokes on their smartphones during daily activities. In this work, participants used typing-intensive apps (e.g., instant messaging, email) and self-reported their affect right after the typing session. The authors obtained an average classification accuracy of 73% when differentiating between stressed, happy, sad, and relaxed states. Lee et al. [60] developed a Twitter-like application that logged participants' keystrokes and some additional contextual information such as illuminance, location, and weather. Their models obtained 67.5% average classification accuracy when differentiating happiness, surprise, anger, disgust, sadness, fear, and neutral emotions. In 2019, Sarsenbayeva et al. [57] investigated the effects of stress on several daily life-like tasks, including a text entry task in which participants were asked to type both easy and difficult texts, under neutral and stressed states. Mental stress was elicited utilizing the Trier Social Stress Test (TSST) [62] and mental arithmetic tasks. In their analysis, the authors reported that participants had a tendency to make more errors when under stress, but the effect was not significant. However, the authors observed a significant effect between the text difficulty and number of errors.

Other studies have taken advantage of additional built-in sensing capabilities (e.g., accelerometer, pressure-sensing screen) when recognizing emotion on mobile devices. As an example, Carneiro et al. [61] collected a multimodal dataset while participants performed tasks under neutral and stressed mental states, elicited by means of time pressure, sounds, and vibration. The dataset included accelerometer data, touch intensity and duration, video recordings, and others. The authors performed participant-specific statistical analysis and observed significant differences in at least one feature group when comparing stressed and unstressed behavior. They reported that acceleration, and mean and maximum touch intensity were the most successful features for recognizing stressed behavior. In recent work, Exposito et al. [55] investigated how stress is manifested in touch intensity. In their user studies, participants performed expressive writing, where they were asked to write about neutral and stressful memories. The authors observed a significant positive correlation between the increase in touch intensity and self-reported stress across the two conditions.

2.4 Limitations of Previous Work

A number of the above studies have reported high accuracies, even when performing multi-emotion classification. We believe that some of these results are optimistic, owing to their experimental design and data analysis, which we discuss below.

One of most common type of stressor in the above studies is time pressure (e.g., [21, 34, 46, 49, 63]). Time pressure is an effective stressor, but its use is problematic when combined with keystroke and mouse timing features. Since time pressure is confounded with stress, it is not clear whether an algorithm is predicting stress or simply detecting the natural changes in behavior caused by the time pressure, since the analyses rely on timing and latency features.

A second problem is the lack of multi-day protocols. In some cases [45, 64], classification results were obtained by splitting data from the same session into a training set and a testing set. This inevitably overestimates the accuracy of the classification models due to the highly correlated nature of the time-series data. To demonstrate that the models are robust, we feel that they must be tested across different sessions.

As noted by Lau [20], several works lack a vetted emotion-induction procedure. For example, some studies elicited emotions by asking participants to read a text [64] or watch a video clip [28], but these emotion-elicitation methods were not validated with physiological measures or subjective ratings. Another problem in prior studies is the lack of sufficient details about the experimental procedures, which can make it difficult to replicate a study or compare results across studies [30, 32, 64, 65].

To the best of our knowledge, two studies by Vizer et al. [15, 41] are the only ones to have employed multiday protocol with a vetted stress induction procedure (mental arithmetic tasks). However, these studies only involved keystroke and linguistic feature analysis. Our paper aims to address all the limitations discussed here.

3 DESIGN OF THE PRESSURE-SENSITIVE DEVICES

Since there are no pressure-sensitive keyboards or mice readily available on the market, we propose a simple and low-cost method that researchers may adopt to measure pressure with off-the-shelf keyboards and mice.

3.1 Keyboard Design

Our experimental keyboard uses an array of force-sensitive resistors (FSRs) to measure typing pressure. FSRs can be used to detect physical pressure, squeezing, and weight. This type of sensor is easy to use and is low cost, making it ideal for our design. However, most FSRs suffer from signal drift, i.e., a monotonic decrease in resistance when they are subject to a static load. Drifting is especially problematic in our design because, when a keyboard is standing on a surface, its weight naturally applies pressure to the sensors, causing drift. To address this issue, our design uses ShuntMode FSRs manufactured by Sensitronics[1], shown in Fig. 1.a, which have low-drift characteristics.

The FSRs are arranged in a voltage-divider configuration, with one terminal connected to a 5V power

[1] https://www.sensitronics.com

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2021.3061417, IEEE Transactions on Affective Computing

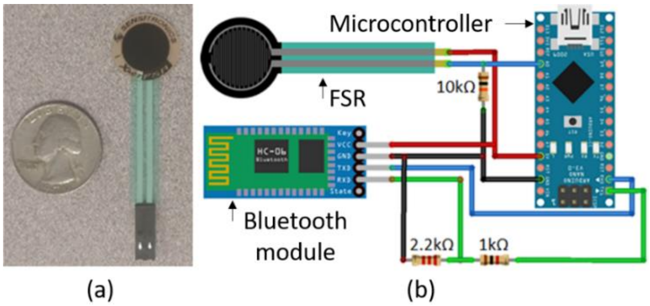


Fig. 1: Schematic of the circuit used in our instrumented keyboard. (a) The size of the pressure sensor used when compared to a quarter dollar. (b) Shows how the pressure sensor is connected to an Arduino microcontroller, which streams data to a Bluetooth module that can be paired with any Bluetooth device.

source and the other connected to an analog input to a microcontroller, as well as to ground by means of a 10kΩ pull-down resistor. To stream data, we use an HC-06 Bluetooth module manufactured by KEDSUM, which is also connected to the microcontroller. Wiring is shown in Fig. 1.b. The HC-06's RX pin expects a 3.3V input, so we used a voltage divider to reduce the input voltage from the microcontroller from 5V to 3.3V.

Our design uses an off-the-shelf keyboard (Dell model KB212-B). We chose this specific keyboard because it has a flat underside, most of its feet are close to corners of the case, and it has enough room to route the sensors to the microcontroller. In addition, the keyboard is comfortable and low-cost (note, though, that our design could be easily adapted to many other keyboard models, including laptops). We placed four FSRs on the underside of the keyboard, near the four corners, and connected them to analog inputs on the microcontroller, as shown in Fig. 2.

No changes were made to the upper side of the keyboard. In addition, we attached gel bumpers to the FSRs to distribute the pressure more efficiently across the sensor surface. When the user types, pressure is applied to the keyboard, which in turn presses the bumpers that apply pressure to the FSRs, generating a response. We attached the FSRs to the keyboard using their built-in adhesive tape, secured the cables with duct tape, and connected them to the microcontroller. Finally, we connected the keyboard's internal ground and 5V pins to the microcontroller and Bluetooth module, eliminating the need for an external battery. The sensors' sampled pressure data at 100 Hz.

3.2 Mouse Design

During the early stages of the mouse design, we compared two sensor choices: capacitive sensors and FSRs. Capacitive sensors have been used to detect and measure position and force because of capacitance coupling [66]. In our first prototype (Fig. 3.a), we used copper tape to build a conductive surface as a capacitive sensor. We attached copper tape to the mouse shell surface and covered it with electrical tape to protect the sensor from abrasion and prevent signal saturation. The sensors were placed on the mouse buttons (one sensor for each button) and on either side of the mouse. We used the same microcontrollers as in the keyboard design. The entire circuit (except for the sensor itself) is invisible to the users since it is small enough to fit inside a regular computer mouse and is powered from the mouse's own power line. We drilled four holes in the mouse shell to connect the sensors placed on the outer part of the mouse to the microcontroller inside the mouse shell.

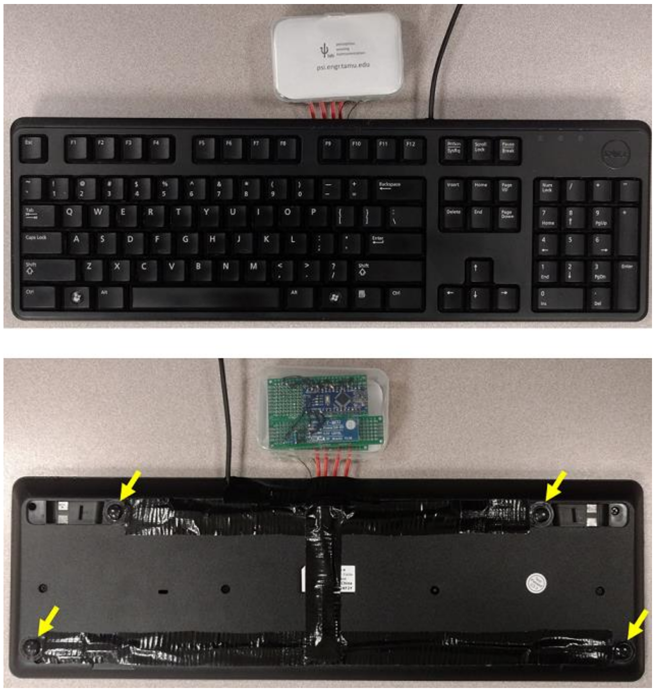


Fig. 2. Top and back view of the instrumented keyboard. Four FSRs (indicated by the yellow arrows) are placed on the back of the keyboard and are connected to the analog inputs of an Arduino microcontroller, which is used to interface with a computer.

Our second prototype (Fig. 3.b) also used capacitive sensors. This time, however, we replaced the copper tape with conductive paint. The advantage of conductive paint is that the shape of the sensor is more flexible and can be placed inside the mouse, underneath its shell, hiding it completely from the user. We tested these two prototypes and found that both sensors behaved similarly:

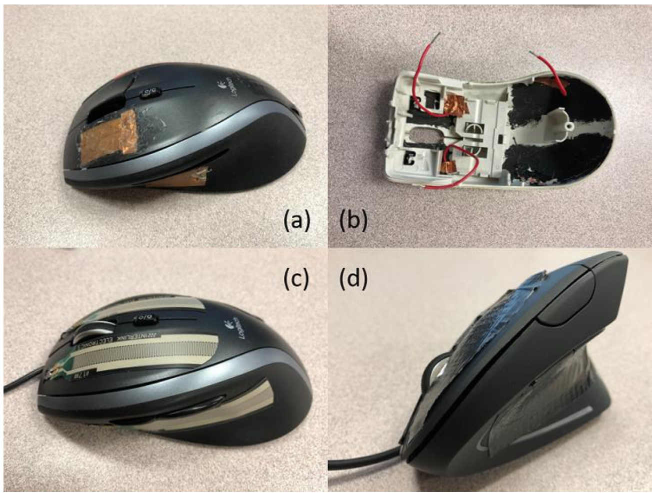


Fig. 3. Various pressure-sensitive mice prototyped. (a) A mouse with copper tape on the surface. (b) A mouse with conductive paint underneath the shell. (c) A regular mouse with four FSRs. (d) A vertical mouse with four FSRs covered by black tape.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2021.3061417, IEEE Transactions on Affective Computing

capacitance values increased as the user made more skin contact with the mouse. However, we could only observe an increase in capacitance when the users held the mouse unrealistically tightly.

This result led us to investigate the use of FSRs to measure grip pressure. We compared FSR and conductive paint by applying different weights to the sensors and recording the corresponding responses. Results in Fig. 4 show a linear relationship between weight and FSR response, whereas the capacitance sensor saturates rather quickly. Based on these results, we decided to use FSRs for our final mouse design. Namely, we used an Interlink 408 FSR, a 0.6-inch wide strip that can be cut to length.

As with the two capacitive prototypes, we attached four sensors, two on the L/R buttons and two on the sides of the mouse. Microcontrollers and circuits were able to fit inside the mouse shell, and sensors were connected to the microcontrollers through four holes drilled in the plastic shell. The measurement circuit for these sensors is the same as the one proposed for the pressure keyboard (Fig. 1.b). An example of the FSR-based prototype is shown in Fig. 3.c. During pilot studies, we observed that people used a variety of grip patterns (e.g., palm grip, claw grip, tip grip) with this mouse, which introduced undesired variability into the sensor data. To overcome this issue, we created a fourth design using a vertical mouse (Anker Ergonomic). The ergonomic design of this mouse encourages users to grip the mouse consistently, thus reducing variability. After attaching the FSRs and protecting them with duct tape, we obtained the final design of the proposed pressure-sensitive mouse shown in Fig. 3.d. As in the keyboard design, we set the FSRs' sampling rate to 100 Hz.

4 EXPERIMENTAL PROTOCOL

We conducted a user study to investigate whether the proposed pressure devices could be used to detect stress. We were particularly interested in determining how features extracted from the pressure devices compared to traditional keystroke and mouse dynamics analysis. During the experiment, software running in the background logged the typing pressure, mouse pressure, keystrokes, and mouse event-related information.

In this work, we adhere to Lazarus and Folkman's definition of stress [67], which states that stress is experienced when a person perceives that the "demands exceed the personal and social resources the individual is able to mobilize." Thus, mental distress (i.e., negative stress) is caused when the mental resources cannot appropriately deal with the demands posed. In our experiments, the demands we impose upon our participants are delivered by means of cognitive interference, cognitive load, and rapid decision making – explained in more detail throughout this section. As such, we sought to elicit and capture changes in behavior when participants experience mental distress, which is often associated with an increase of arousal and decrease of valence.

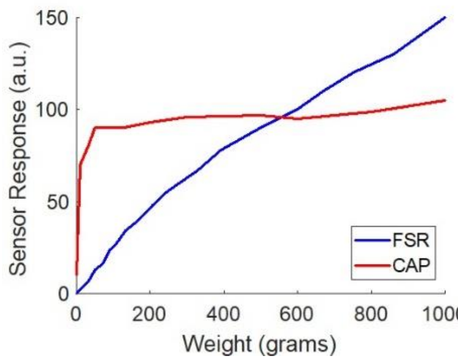


Fig. 4. Weight vs. FSR sensor response (blue curve) and conductance response (red curve) in arbitrary units (a.u.).

4.1 Overview

The user study consisted of four sessions, each session performed on a different day. Fig. 5 shows the structure of each session. First, we asked participants to fill out a questionnaire about their arousal and valence at that moment. If it was their first session, we also asked them to provide information about computer use (how long they have been using computers and how frequently they use them). After filling out the pre-experiment questionnaire, we instructed participants to proceed to the study desk and start the experiment. Next, participants started either the control or experimental block (counterbalanced). In each block, participants performed a priming task for 5 minutes, followed by a 10-minute writing task. After completing the priming and writing tasks, participants reported their perceived valence, arousal, and workload by filling out a questionnaire using the mouse (details to follow). During the control block, participants performed the tasks in an easier mode, while in the experimental block they performed a more challenging version of the tasks designed to induce stress. We provide details of both tasks in the next section. Once participants finished the first block, they were asked to watch a 3-minute transitional video with images from nature and calming background music. Next, participants started the second block (either the control or experimental block, depending on the first block completed), which also lasted 15 minutes. At the end of each session, we thanked and dismissed participants. At the end of the last session on day 4, participants were debriefed and compensated with a $30 gift card.

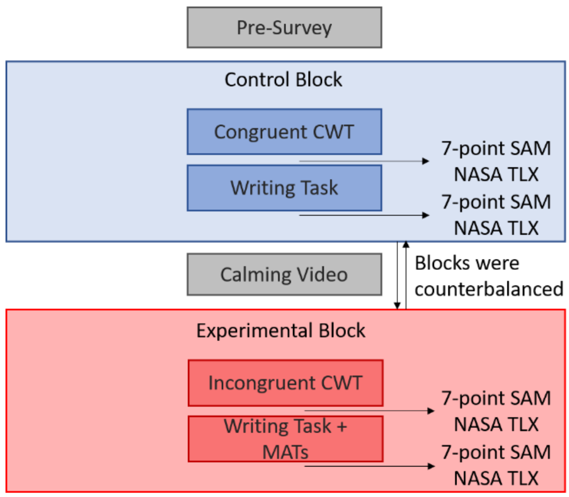


Fig. 5. Procedure of the experiment. The order of control and experimental blocks were counterbalanced.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2021.3061417, IEEE Transactions on Affective Computing

4.2 Priming Task: Stroop Color-Word Test

The *priming task* was designed to influence the participants' behavior during the subsequent questionnaire and writing task. Namely, participants were asked to complete the Stroop Color-Word Test (CWT), a computer-based cognitive task that is commonly used to elicit stress due to cognitive interference and rapid decision making [24, 68-72]. In particular, Tulen et al. [70] have shown that the Stroop CWT simultaneously induces four types of reactions that are required for a suitable stress test: 1) psychological changes that indicate increased distress, 2) physiological changes that indicate sympathoadrenal activation, 3) muscular exertion as part of the fight-flight defense reaction, and 4) hormonal changes, reflected in plasma and urinary catecholamines, and plasma cortisol and prolactin.

For our study, we developed a version of the CWT where participants make their choices by selecting one of four options positioned at the corners of the screen. The objective of the task was to choose the correct font color or the text of the word, depending on what was asked. An example of a trial is shown in Fig. 6. In this particular trial, the font color (orange) does not match the text (blue) and the instructions ask the participant to choose word (i.e., blue). If the instructions had asked to select color, the correct choice would have been orange. We implemented two versions of the CWT: difficult and easy. In the difficult mode, participants were presented with incongruent stimuli in which the font color did not match the text; see Fig. 6. Participants need to select the correct answer from the four options (shown on the corners), which were shown in white font color. In the easy mode, participants were presented with congruent stimuli, i.e., the target word's font color and its text always matched. In addition, the four options were shown with their respective font colors. In either mode, whenever the participant selected the wrong option or took more than 5 seconds, the CWT played a loud buzzer sound and displayed a visual message as an extra stressor. Note that the sole purpose of this task was to elicit stressed or neutral behavior prior to the subsequent tasks. The tasks used in our keyboard and mouse analyses are described, respectively, in Sections 4.3 and 4.4.

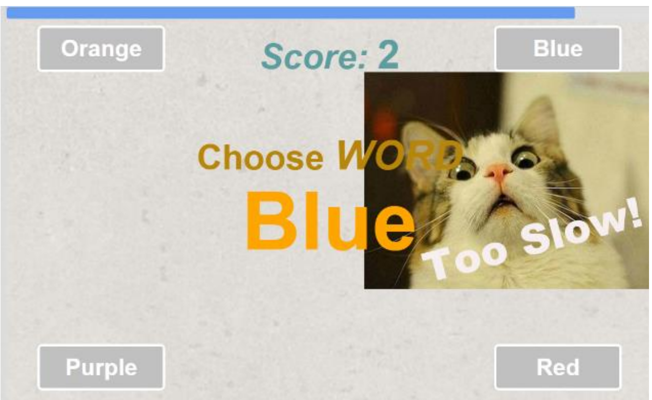


Fig. 6. Stroop Color-Word test variant used in our experiments. Every round, the participants must select their choices using the mouse. The four options are positioned on the corners of the screen.

4.3 Writing Task

In this task, participants were presented with various classical paintings and were asked to describe them (i.e., how characters are dressed, what activities they are performing). We also encouraged participants to come up with a story behind that picture; see Fig. 7.a for an example of a description for the Story of Golden Locks painting, by Seymour Joseph Guy [73]. For each painting, participants had to write at least 200 words before submission. Within each block, we presented up to three paintings to the participants, depending on how fast they completed each description. To finish the task, they had to either finish describing the three paintings, either writing the minimum number of words or spending ten minutes writing, whichever happened first. In total, we used 24 paintings in our experiments, which were never repeated for a participant.

To make the writing task more stressful, participants had to perform mental arithmetic tasks (MATs) during the experimental block. MATs have been extensively used to create a stress response due to high cognitive load, intensive mental demand, and rapid decision making [41, 46, 69, 74-76]. In particular researchers have observed that when under mental distress elicited utilizing MATs, participants presented higher self-reported stress, systolic and diastolic blood pressure, heart rate, urinary catecholamines, salivary cortisol, and electromyogram activity [46]. Our interface is shown in Fig. 7.b. While describing the paintings, our software prompted MATs at intervals specified by sampling a Poisson distribution with

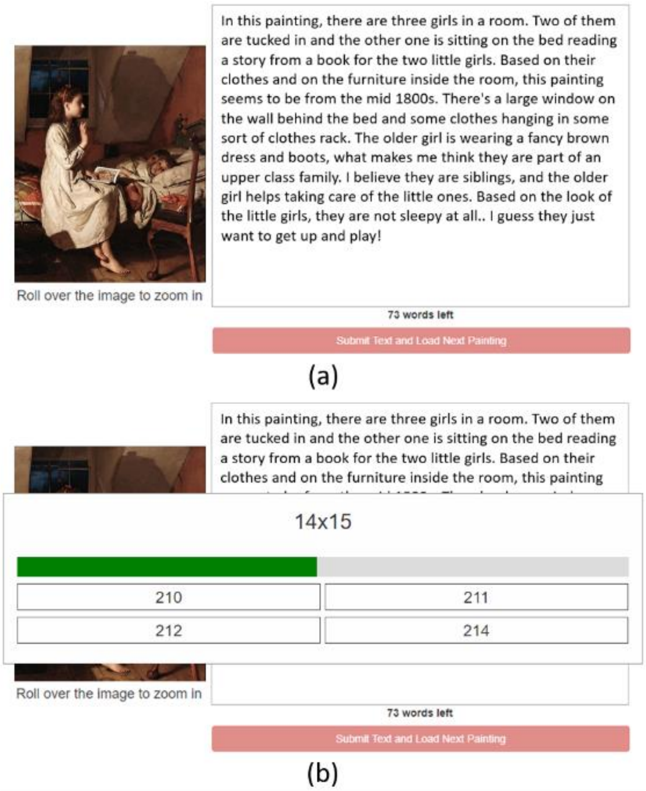


Fig. 7. An example of a painting and its description is shown in (a), (b) shows the same painting and description overlaid with a mental arithmetic task during the writing task. Here, the submit button was deactivated because the participant/t has only written 127 words.

a mean of 30 seconds. When answering a MAT, the participant had to choose one of the four provided options within 5 seconds. If the participant failed to select the correct option or ran out of time, a loud buzzer was played.

4.4 Self-Reported Emotional State and Workload

All participants were asked to complete a questionnaire, in which they reported their perceived valence, arousal, and perceived workload after finishing each task in both the control and experimental blocks. The questionnaire served two purposes. First, it allowed us to determine whether the stressors delivered were successful. Second, it provides an opportunity to analyze changes in mouse behavior elicited by the prior priming task. To do so, we compared the mouse data logged during the questionnaire after the easy (control) CWT and after the difficult (experimental) CWT. We expected changes in mouse behavior after the CWT to be more pronounced than those after the writing task.

Fig. 8 shows the user interface of the self-reported questionnaire. For self-reported valence and arousal, we used the 7-Point Self-Assessment Manikin [77], which has been extensively used for self-reporting arousal and valence. We expected participants to report a lower valence score and a higher arousal score in the tasks performed during the experimental block, as compared to the control block.

To assess task workload, we used the NASA Task Load Index (NASA-TLX), a survey instrument that asks participants to report their perceived mental demand, physical demand, temporal demand, frustration, effort, and performance on the tasks they just finished [78]. We expected higher values of mental demand, physical demand, temporal demand, frustration, and effort, and

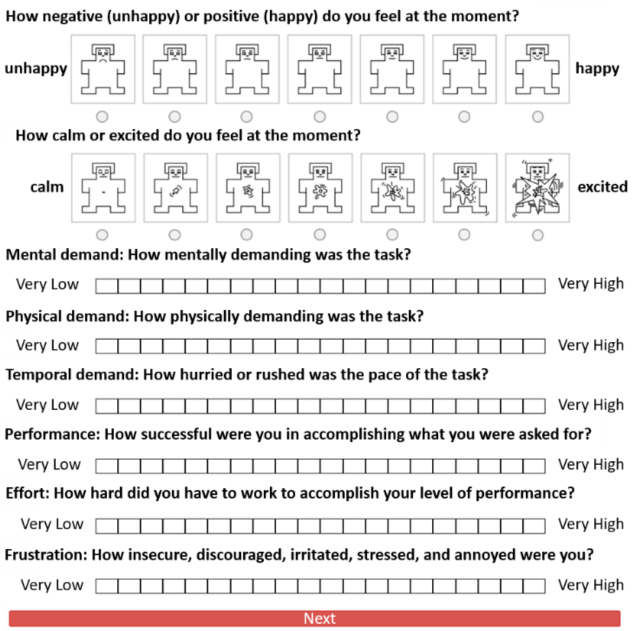


Fig. 8. Self-report questionnaire. Self-Assessment Manikin questions are shown in the top of the questionnaire, and NASA-TLX questions in the bottom.

lower values of performance reported for the experimental block when compared to those of the control block.

4.5 Participants

We invited participants using our institution's bulk mail system, which sends the invitations to student and staff mailing lists. The inclusion criteria were that participants should be 18 years of age or older and fluent in English. We received approval from the Texas A&M University Institutional Review Board (study #IRB2017-0183D) prior to the study. We obtained written consent from each participant before the first session started.

In total, 25 participants (9 male and 16 female) participated in this study. One of the participants was left-handed, so we decided not to consider his data in the mouse analysis. Participants had an average age of 22 (standard deviation (SD): 8.1). All participants reported using computers for at least 2 years (average: 13 years, SD: 6.8 years) and at least 5 hours of weekly usage (average: 28 hours, SD: 15.3 hours). One participant decided to drop out after the second session for personal reasons unrelated to the experiments, but we were able to use the data from her first two sessions in our analysis.

5 DATA ANALYSIS METHODS

5.1 Keyboard Features

We extracted two types of features from the keyboard data: keystroke dynamics features[2] and pressure features; see Table 1. We chose keystroke dynamics features that have been used extensively in the affect-recognition and user-authentication domains [19, 28, 79]; see Fig. 9 for an illustration of these features. To define the set of pressure features, we initially referred to the works of Hernandez et al. [21], Lv et al. [64], and Carneiro et al. [61]. From these works, we used the features mean pressure, maximum pressure (referred to as peak pressure), and pressure standard deviation, and combined them with additional pressure features we designed. As summarized in Table 1, we extracted six features aiming to capture the pressure signature.

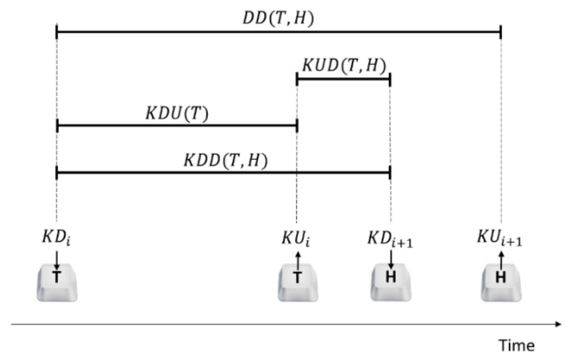


Fig. 9. Keystroke features computed over consecutive key events. Respectively, $\boldsymbol{KD_i}$ and $\boldsymbol{KU_i}$ represents the i-th keydown and keyup events.

[2] In the keystroke dynamics literature, a key press is called a *keydown* event, and a key release is called a *keyup* event.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2021.3061417, IEEE Transactions on Affective Computing

TABLE 1
KEYBOARD FEATURES USED, WHERE $\boldsymbol{KD_i}$ STANDS FOR A KEYDOWN AT TIME i, $\boldsymbol{KU_i}$ STANDS FOR A KEYUP AT TIME i, AND $\boldsymbol{K_i}$ REPRESENTS A KEYSTROKE AT TIME i

Feature	Acronym	Description
Keystroke Dynamics		
Keydown-Keydown	$KDD(KD_i, KD_{i+1})$	Time between two consecutive keydown events
Keydown-Keyup	$KDU(KD_i, KU_i)$	Duration of key being pressed. Also known as dwell time
Keyup-Keydown	$KUD(KU_i, KD_{i+1})$	Time between releasing a key and pressing the next one. Also known as flight time
Digraph duration	$DD(KD_i, KU_{i+1})$	Time between pressing a key and releasing the consecutive one
Pressure features		
Mean Pressure	$MP(K_i)$	Pressure value
Peak Pressure	$PP(K_i)$	Peak pressure value
Standard Deviation	$STD(K_i)$	Standard deviation of a pressure response
Pressure Difference	$PD(K_i, K_{i+1})$	Difference between two consecutive pressure values
AUC Difference	$AUCD(K_i, K_{i+1})$	Difference between the area under the curve (AUC) of two consecutive pressure samples
Pressure Time Difference	$PTD(K_i, K_{i+1})$	Time difference of two consecutive pressure samples

To extract pressure features, we sampled the pressure time series only when keydown events occurred. This allowed us to discard pressure measurements when there was no keyboard activity. In a first step, we subtracted the static load (i.e., keyboard weight) from each sensor's raw pressure time series, which helped normalize sessions from different days and different participants. Then, we assigned a pressure measurement to each keypress by choosing the maximum pressure value between the current and the next keydown event, which we refer to as Peak Pressure (PP); see Fig. 10. To compute the features Pressure Difference (PD) and Pressure Time Difference (PTD), we considered the sampled pressure (PP) as the reference value, as described above. The Mean Pressure (MP) and Standard Deviation (STD) features represent, respectively, the mean and standard deviation of each pressure response. Finally, the feature Area Under the Curve Difference (AUCD) is obtained by computing the AUC of each pressure response, and then calculating the difference in AUC between consecutive keys.

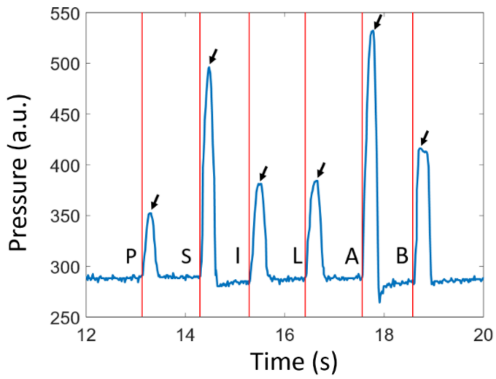


Fig. 10. A segment of a pressure time series in arbitrary units (a.u.), along with keystroke information. The pressure time series are shown in blue; red vertical lines represent keydown events; black arrows point to the pressure values chosen to represent the pressure of each keystroke, which we refer to as the Peak Pressure (PP) feature. The characters shown close to the keydown events represent which key was typed at that moment.

We considered keydown and keyup events only for keys in the range A, B, …, Z. Hence, the features considered are calculated for either each single key (A, B, …, Z) or pairs of keys ([A,A], [A,B], …, [Z,Z]), depending on whether the feature involves a single key or a pair of keys. For each feature, we used its average value across the entire session block. In instances where a key or pair of keys was not observed during a session, the corresponding features were assigned a value of zero. In summary, the features KDU, MP, PP, and STD have dimensionality 26 (26 keys), and the features KDD, KUD, DD, PD, AUCD, and PTD have dimensionality 676 (26 keys × 26 keys).

As mentioned previously, we used mental arithmetic tasks (MATs) during the writing task as a stress-elicitation procedure. One of the drawbacks of this procedure, however, is that it interrupts the participants' writing process. As these interruptions could lead to exaggerated features calculations, we preprocessed the keyboard data in order to minimize the effect caused by the MATs on the keyboard features. Namely, we observed that typing speed decreases to zero while participants were answering the MATs (as expected) and that, once participants resume typing, it took an average of two seconds (or five keystrokes) for their typing speed to return to its average level. Hence, in our approach, we ignored any eventual keystroke logged during the MAT and five additional keystrokes after each MAT. Further, our analysis showed that the sensor placed at the bottom-left corner (i.e., close to the Z key) was the most sensitive of the four sensors; this was likely because the bottom-left sensor was the closest sensor for 60% of the keys examined in our study (the 26 alphabetical keys). Therefore, all preprocessing methods and data analyses are based on the pressure time series obtained by the bottom-left sensor[3].

5.2 Mouse Features

We extracted two types of features from the mouse data:

[3] In a separate experiment not reported here, we compared performance when using a single sensor vs. using the four sensors, and the results were virtually identical.

mouse dynamics and pressure measurements from the FSR sensors. As with the keyboard dynamics, we chose mouse dynamics features that have been used in the related literature [26, 42], with the exception of the pressure features, which we needed to design on our own. The mouse features are listed in Table 2. We extracted six features from mouse dynamics: two mouse trajectory features (travel distance and direction change), two speed features (overall speed and moving speed), and two timing features (dwell duration and moving duration). Note that the mouse features were calculated considering the entire session duration. For example, for the Travel Distance (TD) feature, we summed up the distance covered by each mouse stroke performed during a session. We extracted four pressure features from the FSR sensors. Two of these pressure features were from the FSR on the left click button: mean and standard deviation of the click forces (since no right-click was required during our experiments, we did not extract any features from the FSRs on the right-click button). The other two features were the mean and standard deviation of the grip force, measured from the two FSRs on the sides; see Fig. 3.

As with typing pressure, we only considered pressure values during periods of mouse activity. We used the maximum peak value immediately after the click event as the clicking force. As for the grip force features, we sampled the FSR time series whenever a user interaction event occurred (e.g., cursor movement, click).

TABLE 2
MOUSE FEATURES USED IN OUR ANALYSES

Feature	Acron.	Description
Mouse trajectories features		
Dwell duration	DD	How long the mouse remains idle
Moving duration	MD	How long the mouse is being moved
Travel distance	TD	Cumulated distance in pixel (px) that the mouse cursor moved
Overall speed	OS	OS = TD / (DD + MD)
Moving speed	MS	The speed only during mouse movement. Given by MS = TD / MD
Direction change	DC	Cumulative direction change (in radian) that the mouse cursor traveled
Pressure features		
Click force mean	CFM	Mean of the click peak values from the FSR on the left click button (our task never requires using a right click)
Click force std	CFS	Standard deviation of the click peak values from the FSR on the left click button
Grip force mean	GFM	Mean of the grip forces (grip force is defined as the sum of left-side and right-side FSR)
Grip force std	GFS	Standard deviation of the grip forces

5.3 Classifier Design

Once the time series were preprocessed (as described in the previous two sections), we executed our feature extractor module to convert raw data into feature sets, which were then passed to a binary classifier trained to discriminate between neutral and stress conditions, as described below. Due to the large number of features relative to the number of samples in our dataset, we used linear discriminant analysis (LDA) [80] to reduce the dimensionality of the feature vector. LDA projects the features in a way that maximizes the ratio of between-class scatter to within-class scatter, leading to more pronounced differences between neutral and stressed samples. In addition, since we only have eight samples per participant, we pooled data from multiple participants to train subject-independent classifiers (i.e., generic classifiers) using a leave-one-participant-out procedure. For illustration purposes, assume we are considering four feature groups in our keyboard analysis: Keydown-Keydown (676 dimensions: 26 keys × 26 keys), Keydown-Keyup (676 dimensions), Mean Pressure (26 dimensions), and Pressure Time Difference (676 dimensions); see Fig. 11. The dimensionality of this combined set would be equal to 2,054. Our dimensionality-reduction procedure projects each feature group (i.e., KDD, KDU, MP, PTD) into a single dimension (i.e., a two-class problem has one LDA projection), resulting in four projections – one projection per feature group. The procedure shown in Fig. 11 is detailed next.

For each run, we split the dataset into a training and a test set. The test set contains data from a single participant (8 samples), while the training set contains data of the remaining participants (180 samples). We use the training set to compute an LDA eigenvector for each feature group, as illustrated in Fig. 11. Then, we use the resulting eigenvectors to project the test set. As such, the test data is never used to compute the LDA eigenvectors. Once the training set and test set are projected into the LDA subspace, we use a classifier to generate class labels for the test samples. We repeat this procedure for each participant

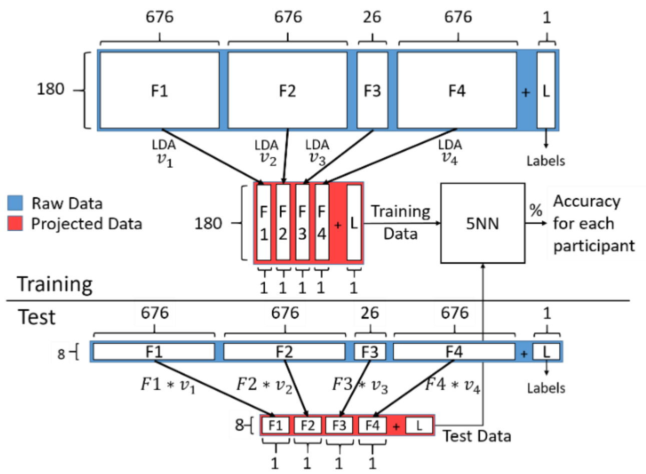


Fig. 11. Generating a low-dimensional projection from four feature sets (F1-F4; blue block) using Linear Discriminants analysis. The reduced feature vector (four dimensions; red block) is passed to a nearest neighbor classifier to generate a class prediction into neutral vs. stressed states.

and report the mean classification accuracy obtained by each run of the leave-one-participant-out analysis. We compared three classifiers for this purpose: 5-nearest-neighbors (5-NN)[4], support vector machine (SVM), and naïve bayes (NB) using their corresponding optimized set of features. On keyboard data, the best-performing classifier was 5-NN, achieving 74% classification accuracy, whereas SVM and NB classifiers achieved 73% and 69% classification accuracy, respectively. On mouse data, 5-NN also yielded the highest classification rate (73%), compared to SVM (70%) and NB (72%). We expand on the results achieved by 5-NN in the following section.

6 RESULTS

In this section, we show how the stressors delivered affected the participants' perceived arousal, valence, and workload with respect to the control block. Then, we present the results obtained by the automated classifiers.

6.1 Stress Elicitation (SAM)

As described earlier, we used four questionnaires in each session to rate the participants' stress levels at different time points. We administered a questionnaire after the easy CWT (Easy CWT Questionnaire, or ECQ for short), and another after the difficult CWT (DCQ). We also administered questionnaires for the easy and difficult typing tasks (ETQ and DTQ, respectively). Presenting both the Self-Assessment Manikin (SAM) and NASA Task Load Index (NASA-TLX) questionnaires after each task ensured that participants faced the same questionnaire page every time, for a fair mouse analysis comparison during the questionnaire. Fig. 12 shows boxplots for the arousal and valence ratings, with each session as one sample. Since each of the 24 participants completed 4 sessions (except one who only finished two sessions), we have 94 pairs of samples in total. We used paired t-test for statistical purposes.

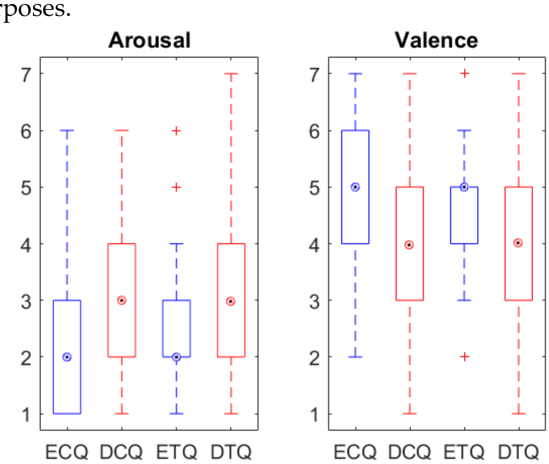


Fig. 12. Box plots for self-reported valence and arousal. The x-axis labels represent questionnaire after easy CWT (ECQ), questionnaire after difficult CWT (DCQ), questionnaire after easy typing task (ETQ), and questionnaire after difficult typing task (DTQ).

First, we examined if the perceived stress level was different between the two versions of the CWT. A comparison of ECQ to DCQ indicates that arousal ratings during the difficult CWT were significantly higher (mean increase of 1.05, $t(93) = -7.63$, $p << 0.01$) and valence ratings were significantly lower (mean decrease of 0.81, $t(93) = 6.12$, $p << 0.01$) than those during the easy CWT. This indicates that the difficult CWT increased participants' stress levels, as expected. Next, we examined whether stress levels were different between the two versions of the typing task. A comparison of ETQ to DTQ indicates that arousal ratings during the difficult typing task were significantly higher (mean increase of 0.44, $t(93) = -4.23$, $p << 0.01$), and the valence was significantly lower (decreased by 0.26, $t(93) = 2.03$, $p = 0.04$) than those during the easy typing task. These results confirm that the two tasks were able to manipulate the participants' stress levels as we had intended.

6.2 NASA TLX

Analyzing the TLX results also served as a validity check to determine whether the nature of the tasks performed during the experimental block was more difficult than those during the control block. Indeed, during the difficult version of the CWT/typing task, participants reported significantly higher mental demand, higher physical demand, higher temporal demand, lower performance, higher effort, and higher frustration than during the easy version of the CWT/typing task – see Fig. 13. The only

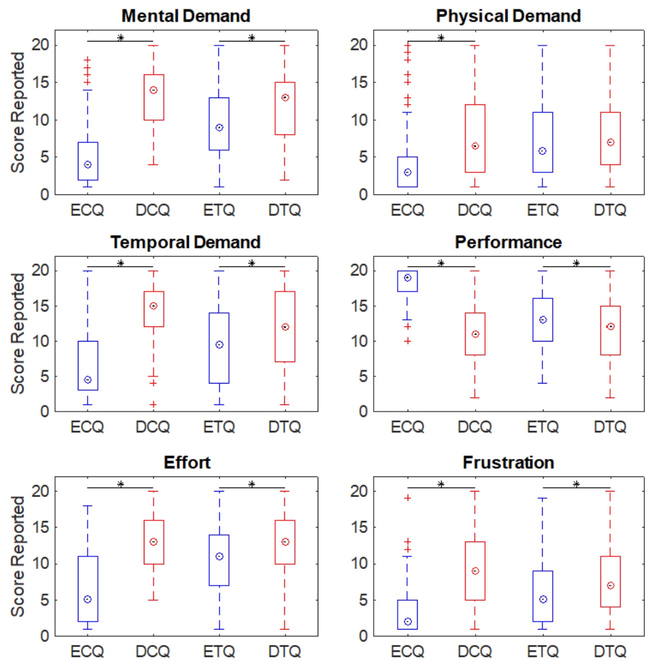


Fig. 13. Box plots for self-reported NASA-TLX scores. The x-axis labels represent questionnaire after easy CWT (ECQ), questionnaire after difficult CWT (DCQ), questionnaire after easy typing task (ETQ), and questionnaire after difficult typing task (DTQ). * represents statistically-significant differences for the scores reported for each task.

[4] To optimize the number of neighbors (k), we varied k from 1 to 10 and did the following. In each iteration of the leave-one-participant-out analysis, we randomly selected eight samples from the training data and used them for validation purposes. We then trained a k-NN classifier with the specific value of k using the remaining training samples. Next, we evaluated the models trained using the validation data. We repeated this analysis for each participant and considered the average classification accuracy obtained with each k to decide the final configuration.

exception was the self-reported physical demand for the typing task (mean increase of 0.62, $t(93) = 1.92$, $p = 0.056$). However, we still observed a trend towards the expected direction (difficult typing task leading to higher physical demand) and a p-value close to significance. These results suggest that the tasks were successful in eliciting stress.

6.3 Keyboard Analysis

In total, we collected 188 samples (a sample contains all features computed during a block), as every participant but one went through four control sessions and four experimental sessions. Given that the number of samples in the control and experimental sessions are the same, a random classifier would achieve 50% classification accuracy. For the remainder of the manuscript, the classification accuracy obtained by such a random classifier will be referred to as a chance-level classification accuracy.

To identify the best subset of features for each type (keystroke only, pressure only, and keystroke + pressure), we performed exhaustive search on the feature sets, i.e., we evaluated our models on every possible combination of features, for a total of 1023 ($2^{10} - 1$) feature subsets. Average classification accuracies are shown in Table 3. Using all (timing and pressure) features as input performs slightly worse than selecting a subset of them. When using timing features alone, the classifier obtained an accuracy of 68% using the feature groups DD and KDU. Using pressure features alone, our classifier obtained an accuracy of 71% using the feature groups PP, MP, AUCD, and PDT. When both timing and pressure features were combined, the optimal feature subset contained the feature groups KDD, KUD, DD, KDU, PP, AUCD, and PTD and achieved 74% classification accuracy. Thus, adding pressure information to timing features led to a 6% absolute improvement in average classification accuracy (i.e., from 68% to 74%). Hence, combining timing and pressure features provides higher classification accuracy than using each feature type in isolation. Next, we expand on the results achieved by the optimal feature subset, obtained with the combination of keystroke dynamics and pressure features.

TABLE 3
GENERIC-CLASSIFIER CLASSIFICATION ACCURACY WHEN USING DIFFERENT SETS OF KEYBOARD FEATURES

Feature Set	Accuracy (St. dev)
Keystroke Only	
Full: [DD, KDU, KUD, KDD]	65.1 % (16.1 %)
Optimal: [DD, KDU]	67.7 % (13.7 %)
Pressure Only	
Full: [MP, PP, STD, PD, AUCD, PTD]	67.1 % (14.2 %)
Optimal: [PP, MP, AUCD, PTD]	71.3 % (13.5%)
Keystroke and Pressure	
Full:[KDD,KDU,KUD,DD,MP,PP,STD,PD,AUCD,PTD]	72.4 % (14.2 %)
Optimal: [KDD,KUD,DD,KDU,PP,AUCD,PTD]	**74.5 % (14.7 %)**

Results per participant for the optimal feature subset model are shown in Fig. 14.a. The classifier obtained accuracies of 60% or higher for all but three participants, and accuracy of 85% or higher for ten participants. It is important to note that these classification results were obtained using a leave-one-participant-out procedure; in other words, the classifiers are *subject independent*. Classification performance would likely increase if the classifier were to be adapted to match the characteristic typing patterns of each user.

The confusion matrix for the optimal keyboard data feature subset is shown in Table 4. As it can be seen, there is no significant correct class prediction imbalance, as the number of samples correctly classified do not differ by much (67 vs. 73). The same happened when the prediction did not agree with the actual class label (21 vs. 27). Finally, Fig. 15 shows the ROC curves of the optimal feature subset model trained with keyboard data and that of a null classifier[5]. The optimal feature subset obtained an AUC equal to 0.77, outperforming the null classifier (AUC: 0.50).

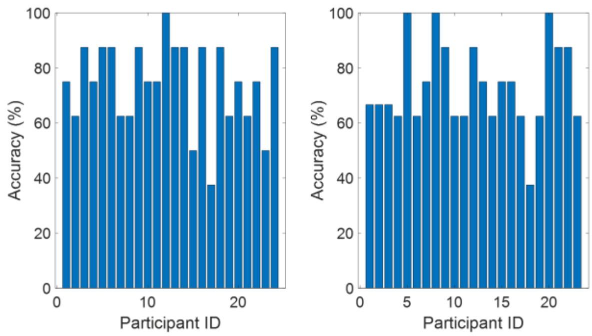


Fig. 14. Individual participant classification accuracies using the best-performing input set from a) the keyboard device, b) the mouse device.

6.4 Mouse Analysis

Mouse pressure data was lost due to Bluetooth connection problem for three sessions, and as mentioned earlier, one participant dropped out after the second session and one was left-handed. Thus, we ended up with mouse data for 87 sessions, totaling 174 samples (87 sessions x 2 blocks).

It is tempting to compare mouse features between the easy and difficult CWT. However, this comparison would yield overly optimistic results since the difficult CWT naturally results in longer response times. In fact, we observed differences in mouse usage patterns (e.g., faster

TABLE 4
CONFUSION MATRIX SHOWING THE ACTUAL VS. PREDICTED OUTPUT OF THE OPTIMUM FEATURE SUBSET FOR THE KEYBOARD ANALYSIS

		Actual	
		Neutral	*Stressed*
Predicted	*Neutral*	67	27
	Stressed	21	73

[5] To generate this plot, we adapted our 5NN model to classify a stressed sample for different minimum number of neighbors (e.g., classify as stressed if at least two out five neighbors are also stressed).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2021.3061417, IEEE Transactions on Affective Computing

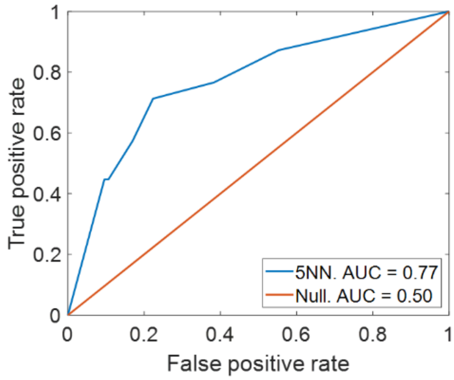


Fig. 15. ROC curves achieved by the optimal feature subset using keyboard data.

mouse movements during the difficult CWT), but these are likely due to the nature of the task, rather than due to the participants' stress levels.

Instead, to examine how stress affected mouse behavior without a time-pressure confounder, we used the mouse data collected during the NASA TLX questionnaire. This phase is ideal for stress analysis because (1) the questionnaire page is the same for the two blocks, (2) no time pressure was applied during the task, and (3) answering a questionnaire is a more realistic task than completing a lab task (i.e., CWT).

Following the procedures outlined for the keyboard analysis, we perform a leave-one-participant-out analysis with the mouse data. Then, we computed LDA projections for three different combinations of features: (1) trajectory features only, (2) pressure features only, and (3) trajectory features and pressure features combined. All these features were projected into a one-dimensional feature and fed to a classifier.

Classification results are reported in Table 6. As with the keyboard analysis, we used exhaustive search to find the optimal set of features when building our models. Trajectory features (70%) outperformed pressure features (61%), both performing performed above chance levels.

TABLE 6
GENERIC-CLASSIFIER CLASSIFICATION ACCURACY WHEN USING DIFFERENT SETS OF MOUSE FEATURES

Feature Set	**Accuracy (St. dev.)**
Trajectories Only	
Full: [DD, MD, TD, OS, MS, DC]	64.1 % (12.1 %)
Optimal: [MD, TD, OS]	70.2 % (17.8 %)
Pressure Only	
Full: [CFM, CFS, GFM, GFS]	56.5 % (15.2 %)
Optimal: [CFM]	61.1 % (19.8 %)
Trajectories and Pressure	
Full: [DD,MD,TD,OS,MS,DC,CFM,CFS,GFM,GFS]	70.1 % (16.1 %)
Optimal: [DD, OS, DC, CFM, CFS, GFS]	**73.3 % (15.5 %)**

More importantly, combining both types of features into a single vector yielded higher classification performance (73%) than either feature alone – a 3% absolute improvement in classification accuracy from using trajectory features alone. Classification rates per participant are shown in Fig. 14.b. Our models obtained classification accuracies above 60% for all but one participant, and 80% classification accuracy or higher for seven participants. As with the keyboard analysis, it is important to note that these classification models are *subject-independent*. It is likely that higher performance may be obtained by adapting a generic classifier to fit the individual mouse behaviors of each user.

Table 5 shows the confusion matrix of the actual vs. predicted class label for the optimum feature subset trained using trajectories and pressure features. As in the keyboard analysis, there is neither significant imbalance between the elements of the main diagonal nor of the anti-diagonal. This indicates the best performing classifier did not obtain the highest classification rate by mainly predicting one class over the other. Fig. 16 shows the ROC curves of the classifier trained with the optimum feature subset using mouse data and that of the null classifiers. As in the keyboard analysis, optimum feature set trained using mouse data obtained an AUC (0.75) superior to that of the null classifier (0.50).

7 DISCUSSION

This paper presents an approach to monitor work stress by analyzing subtle changes in keyboard and mouse usage during knowledge work tasks. The approach involves instrumenting computer peripherals that are already part of modern workplace settings with low-cost external sensors. We developed an experimental protocol to

TABLE 5
CONFUSION MATRIX SHOWING THE ACTUAL VS. PREDICTED OUTPUT OF THE OPTIMUM FEATURE SUBSET FOR THE MOUSE ANALYSIS

		Actual	
		Neutral	***Stressed***
Predicted	***Neutral***	62	21
	Stressed	25	66

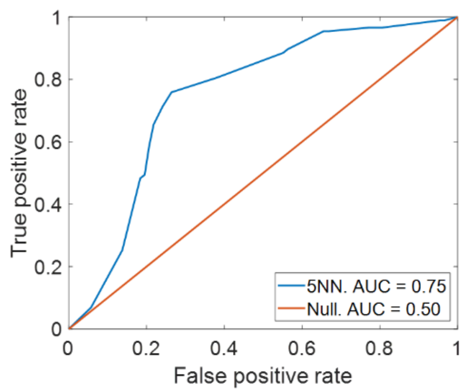


Fig. 16. ROC curves achieved by the optimum feature set using mouse data.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2021.3061417, IEEE Transactions on Affective Computing

simulate two typical tasks in knowledge work (completing questionnaires and writing reports) that require keyboard and mouse interaction. With our instrumented peripherals, we are able to detect consistent changes in behavior caused by mild stressors.

We designed a protocol that addresses the limitations found in the literature, as discussed in Section 2.4. First, we used vetted stressors (Stroop effect and mental arithmetic) in our emotion-induction procedure and validated their effects by analyzing changes in arousal and valence through self-report measures. Second, we carefully avoided confounding factors that may yield overly optimistic results, such as time pressure, one of the most widely used stressors in affective computing. Third, we carried out a multiday user study, totaling four sessions for each participant, and showed that our method is robust to inter-session variability. Finally, we provided detailed instructions about our procedure, to enable other researchers to replicate our study and compare their methods against ours.

To analyze whether we could correctly classify data from neutral and stressed conditions, we designed participant-independent models and trained them with keystroke or mouse dynamics and pressure features from the respective devices. We believe that classification accuracies could have been even higher if we had trained participant-specific classifiers, but the limited number of samples per participant was not sufficient to successfully build a model and test it. Although the recent literature [21, 31, 65] raises awareness that personalized models can lead to higher performance, we showed how to create participant-independent classifiers using simple and robust methods. In addition, we believe that deploying a participant-independent classifier would be more beneficial for workplace settings since it could be trained with a much larger number of samples and would be readily available for new workers.

A major challenge when building participant-independent classifiers is how to account for individual differences. For example, when under stress some people move the mouse cursor faster; others more slowly. In our analysis, we did not explicitly apply any type of feature normalization to account for these individual differences[6]. Instead, our classification approach projects the features onto the LDA subspace to minimize within-class scatter (i.e., intra and inter-individual differences) while maximizing between-class scatter (i.e., due to the stress manipulation). This step makes the classifier more robust against individual differences.

Our results indicate that combining keyboard and mouse dynamics with their respective pressure features improves discrimination between neutral and stressed states. This suggests that features extracted from the two modalities (i.e., time vs. pressure) provide complementary information. However, since using all features during training is not necessarily beneficial, we used exhaustive search to find the set of features that provided the highest discrimination power for the trained classifiers. Exhaustive search was helpful in both the keyboard and mouse analysis, where we obtained the highest classification rates when using a reduced set of features.

7.1 Limitations of our Work

One of the challenges in affective computing research consists of labeling behavioral data with the proper emotional state. In our work, our classification models were trained on the tasks' labels (i.e., the intended effect of the tasks), rather than on the participants' actual stress levels. While the questionnaires we administered confirm that our experiments were successful in manipulating the participants' stress levels, objective measures of stress by means of physiological stress responses would have provided additional validation. However, gathering these measurements is difficult using existing technology. The most reliable physiological measure of stress, electrodermal activity (EDA), requires placing electrodes at the fingers or the palms, which interferes with typing tasks. While measuring EDA from the wrist or the sole/feet is possible, it also has drawbacks; see Tsiamyrtzis et al. [81] for a recent guide comparing the accuracy of different EDA sensors and measurement configurations. Alternatively, perinasal perspiration, a measure known to correlate with EDA, can be captured from image sequences collected from thermal cameras [82], but this requires specialized hardware.

Even though our experimental protocol was designed to be realistic (filling out questionnaires and writing descriptions), performing tasks in a laboratory setting can still cause participants to behave differently than when they are in their usual work environments. Thus, replicating our findings with a field study could help us understand how our methods would work with real-life stressors. This field study would have participants perform their daily computer tasks at work, and would use ecological momentary assessment (EMA) to provide the ground truth emotional state at the time of work – a similar approach is described in [19]. This could lead to an even more realistic keyboard and mouse usage dataset. A field study would also help us collect more data per participant, which could help us build more robust prediction models or adapt generic models to each user.

One potential limitation of our work was the use of desktop computers as opposed to mobile devices, such as laptops and tablets. Projections show that there should be approximately four times as many new laptops and tablets as desktops by 2023 [22]. However, there will also be 80 million desktop shipments by the same year, a number that is far from negligible [22]. More importantly, there is nothing inherent to our approach that would prevent it from being used in laptop keyboards and touchpads, other than we would need to design new features (e.g., specific to touchpads) and adjust the classification models accordingly. Yet, the rise of popularity of laptops and tablets cannot be ignored and we strongly recommend

[6] In an early stage of our analysis, we tried normalizing the data for each subject by computing the z-score of each feature across the eight samples, but the results were largely identical when compared to the ones presented.

future efforts on the detection of stress using these devices – some of which already provide touch intensity [55, 56, 61].

7.2 Future Work

Our work may find application in the domain of user authentication [17, 18], where instrumented devices could be used to gather additional biometric information and train existing user authentication techniques to differentiate between valid users and imposters. One situation where current user authentication methods might fall short is when changes in keyboard and mouse due to stress are recognized as an anomaly (i.e., potential imposter). To address this shortcoming, user authentication algorithms would require training on both unstressed and stressed data from multiple users. This could lead to more robust continuous user-authentication methods.

Most office tasks are computer-based and involve significant mouse and keyboard usage (e.g., writing an e-mail, filling out a spreadsheet). As such, future work on stress detection could benefit from considering tasks which use the keyboard and mouse simultaneously, and build stress detection models using the combined set of features. Such a set of features can be used to train classifiers with even higher discriminatory power, since complementary information can be extracted from both mouse and keyboard usage, potentially leading to higher stress detection rates.

The ultimate goal of stress-detection methods is to help people suffering from stress. One potential future application of our work is to deploy our method in a live workplace, where a software gathers data to detect moments of stress and then recommends just-in-time relaxation interventions to employees (e.g., perform deep breathing exercises, go for a walk, play a relaxation game) to help them better cope with acute stress. While it can be difficult to deploy our stress detection system at a *large scale* since pressure-sensing peripherals are rare, we think the simplicity and low-cost of our design would not be barriers.

8 CONCLUSIONS

In this paper, we investigated whether keyboard and mouse pressure, combined with keystroke dynamics and mouse dynamics, could be used to predict users' stress levels. We designed a simple and cost-effective pressure-sensitive augmentation for keyboard and mouse using force-sensitive resistors (FSRs) and low-cost microcontrollers. To test our approach, we recruited 25 participants to perform two sets of tasks under neutral and stressed conditions. We built a generic classifier by projecting keyboard and mouse features with LDA and fed into a nearest neighbor classifier. Our leave-one-participant-out analysis showed that combining pressure features with keystroke and mouse dynamics improves classification rates. We achieved *subject-independent* classification rate of 74% with the keyboard device and 73% with the mouse device, an average absolute improvement of 6% and 3%, respectively, when adding pressure information to the set of keystroke and mouse dynamics. This work presents the first attempt to build a subject-independent classifier to predict stress with realistic tasks using a pressure-sensitive keyboard and mouse. This is especially important because it is a step closer to providing ways to automatically, continuously, and non-intrusively detect stress in the workplace.

ACKNOWLEDGMENT

We thank all the volunteers who participated in our user study. This work was funded by grants from the Brazilian National Council for Scientific and Technological Development (award #208214/2014-3), and the National Science Foundation (award #1704636).

REFERENCES

[1] M. W. DeVries and B. Wilkerson, "Stress, Work and Mental Health: A Global Perspective," Acta neuropsychiatrica, vol. 15, no. 1, pp. 44-53, 2003.

[2] L. R. Murphy, "Stress Management in Work Settings: A Critical Review of the Health Effects," American Journal of Health Promotion, vol. 11, no. 2, pp. 112-135, 1996.

[3] M. F. Dallman, "Stress-Induced Obesity and the Emotional Nervous System," Trends in Endocrinology & Metabolism, vol. 21, no. 3, pp. 159-165, 2010.

[4] A. Kouvonen, M. Kivimäki, M. Virtanen, J. Pentti, and J. Vahtera, "Work Stress, Smoking Status, and Smoking Intensity: An Observational Study of 46 190 Employees," Journal of Epidemiology and Community Health, vol. 59, no. 1, pp. 63-69, 2005.

[5] R. Sinha, "How Does Stress Increase Risk of Drug Abuse and Relapse?," Psychopharmacology, vol. 158, no. 4, pp. 343-359, 2001.

[6] E. R. Blackmore, S. A. Stansfeld, I. Weller, S. Munce, B. M. Zagorski, and D. E. Stewart, "Major Depressive Episodes and Work Stress: Results from a National Population Survey," American Journal of Public Health, vol. 97, no. 11, pp. 2088-2093, 2007.

[7] A. Alberdi, A. Aztiria, and A. Basarab, "Towards an Automatic Early Stress Recognition System for Office Environments Based on Multimodal Measurements: A Review," Journal of biomedical informatics, vol. 59, pp. 49-75, 2016.

[8] H. M. Burke, M. C. Davis, C. Otte, and D. C. Mohr, "Depression and Cortisol Responses to Psychological Stress: A Meta-Analysis," Psychoneuroendocrinology, vol. 30, no. 9, pp. 846-856, 2005.

[9] S. Cohen, T. Kamarck, and R. Mermelstein, "A Global Measure of Perceived Stress," Journal of health and social behavior, pp. 385-396, 1983.

[10] R. Karasek, C. Brisson, N. Kawakami, I. Houtman, P. Bongers, and B. Amick, "The Job Content Questionnaire (Jcq): An Instrument for Internationally Comparative Assessments of Psychosocial Job Characteristics," Journal of occupational health psychology, vol. 3, no. 4, p. 322, 1998.

[11] D. McDuff, S. Gontarek, and R. Picard, "Remote Measurement of Cognitive Stress Via Heart Rate Variability," in 36th Annual

International Conference of the IEEE Engineering in Medicine and Biology Society, 2014, pp. 2957-2960: IEEE.

[12] J. Choi, B. Ahmed, and R. Gutierrez-Osuna, "Development and Evaluation of an Ambulatory Stress Monitor Based on Wearable Sensors," IEEE transactions on information technology in biomedicine, vol. 16, no. 2, pp. 279-286, 2012.

[13] D. F. Dinges et al., "Optical Computer Recognition of Facial Expressions Associated with Stress Induced by Performance Demands," Aviation, space, and environmental medicine, vol. 76, no. 6, pp. B172-B182, 2005.

[14] S. D. Gunawardhane, P. M. De Silva, D. S. Kulathunga, and S. M. Arunatileka, "Non Invasive Human Stress Detection Using Key Stroke Dynamics and Pattern Variations," in Proceedings of the International Conference on Advances in ICT for Emerging Regions (ICTer), 2013, pp. 240-247: IEEE.

[15] L. M. Vizer, L. Zhou, and A. Sears, "Automated Stress Detection Using Keystroke and Linguistic Features: An Exploratory Study," International Journal of Human-Computer Studies, vol. 67, no. 10, pp. 870-886, 2009.

[16] 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: ACM.

[17] F. Monrose and A. D. Rubin, "Keystroke Dynamics as a Biometric for Authentication," Future Generation computer systems, vol. 16, no. 4, pp. 351-359, 2000.

[18] A. A. E. Ahmed and I. Traore, "A New Biometric Technology Based on Mouse Dynamics," IEEE Transactions on dependable and secure computing, vol. 4, no. 3, pp. 165-179, 2007.

[19] C. Epp, "Identifying Emotional States through Keystroke Dynamics," University of Saskatchewan Saskatoon, 2010.

[20] S.-H. Lau, "Stress Detection for Keystroke Dynamics," Ph.D., Carnegie Mellon University, 2018.

[21] 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: ACM.

[22] A. Holst. (2019, 09/26/2019). Shipment Forecast of Laptops, Desktop Pcs and Tablets Worldwide from 2010 to 2023 (in Million Units). Available: https://www.statista.com/statistics/272595/global-shipments-forecast-for-tablets-laptops-and-desktop-pcs/

[23] C. Busso et al., "Analysis of Emotion Recognition Using Facial Expressions, Speech and Multimodal Information," in Proceedings of the 6th international conference on Multimodal interfaces, 2004, pp. 205-211: ACM.

[24] J. Zhai and A. Barreto, "Stress Detection in Computer Users Based on Digital Signal Processing of Noninvasive Physiological Variables," in 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS'06), 2006, pp. 1355-1358: IEEE.

[25] M. Garbey, N. Sun, A. Merla, and I. Pavlidis, "Contact-Free Measurement of Cardiac Pulse Based on the Analysis of Thermal Imagery," IEEE Transactions on Biomedical Engineering, vol. 54, no. 8, pp. 1418-1426, 2007.

[26] Y. M. Lim, A. Ayesh, and M. Stacey, "Detecting Cognitive Stress from Keyboard and Mouse Dynamics During Mental Arithmetic," in Science and Information Conference (SAI), 2014, 2014, pp. 146-152: IEEE.

[27] B. Arnrich, C. Setz, R. La Marca, G. Tröster, and U. Ehlert, "What Does Your Chair Know About Your Stress Level?," IEEE Transactions on Information Technology in Biomedicine, vol. 14, no. 2, pp. 207-214, 2009.

[28] P. Zimmermann, S. Guttormsen, B. Danuser, and P. Gomez, "Affective Computing—a Rationale for Measuring Mood with Mouse and Keyboard," International journal of occupational safety and ergonomics, vol. 9, no. 4, pp. 539-551, 2003.

[29] S. P. Banerjee and D. L. Woodard, "Biometric Authentication and Identification Using Keystroke Dynamics: A Survey," Journal of Pattern Recognition Research, vol. 7, no. 1, pp. 116-139, 2012.

[30] G. A. Tsihrintzis, M. Virvou, E. Alepis, and I.-O. Stathopoulou, "Towards Improving Visual-Facial Emotion Recognition through Use of Complementary Keyboard-Stroke Pattern Information," in Proceedings of the Fifth International Conference on Information Technology: New Generations (ITNG 2008), 2008, pp. 32-37: IEEE.

[31] I. A. Khan, W.-P. Brinkman, and R. Hierons, "Towards Estimating Computer Users' Mood from Interaction Behaviour with Keyboard and Mouse," Frontiers of Computer Science, vol. 7, no. 6, pp. 943-954, 2013.

[32] P. Khanna and M. Sasikumar, "Recognising Emotions from Keyboard Stroke Pattern," International journal of computer applications, vol. 11, no. 9, pp. 1-5, 2010.

[33] R. Bixler and 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, Santa Monica, California, USA, 2013, pp. 225-234: ACM.

[34] Y. M. Lim, A. Ayesh, and M. Stacey, "Detecting Emotional Stress During Typing Task with Time Pressure," in Science and Information Conference (SAI), 2014, 2014, pp. 329-338: IEEE.

[35] F. Chen et al., "Mouse Based Measures," in Robust Multimodal Cognitive Load Measurement: Springer, 2016, pp. 147-157.

[36] D. G. Brizan, A. Goodkind, P. Koch, K. Balagani, V. V. Phoha, and A. Rosenberg, "Utilizing Linguistically Enhanced Keystroke Dynamics to Predict Typist Cognition and Demographics," International Journal of Human-Computer Studies, vol. 82, pp. 57-68, 2015.

[37] S. Arshad, Y. Wang, and F. Chen, "Interactive Mouse Stream as Real-Time Indicator of User's Cognitive Load," in Proceedings of the 33rd annual ACM conference extended abstracts on human factors in computing systems, 2015, pp. 1025-1030.

[38] M. Grimes and J. Valacich, "Mind over Mouse: The Effect of Cognitive Load on Mouse Movement Behavior," 2015.

[39] A. Khawaji, F. Chen, J. Zhou, and N. Marcus, "Trust and Cognitive Load in the Text-Chat Environment: The Role of Mouse Movement," in Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design, 2014, pp. 324-327.

[40] H. Rheem, V. Verma, and D. V. Becker, "Use of Mouse-Tracking Method to Measure Cognitive Load," in Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2018, vol. 62, no. 1, pp. 1982-1986: SAGE Publications Sage CA: Los Angeles, CA.

[41] L. M. Vizer and A. Sears, "Efficacy of Personalized Models in Discriminating High Cognitive Demand Conditions Using Text-Based Interactions," International Journal of Human-Computer Studies, vol. 104, pp. 80-96, 2017.

[42] T. Yamauchi, "Mouse Trajectories and State Anxiety: Feature Selection with Random Forest," in Proceedings of the 2013

Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII 2013), 2013, pp. 399-404: IEEE.

[43] M. Hibbeln, J. L. Jenkins, C. Schneider, J. S. Valacich, and M. Weinmann, "How Is Your User Feeling? Inferring Emotion through Human–Computer Interaction Devices," Management Information Systems Quarterly, vol. 41, no. 1, pp. 1-47, 2017.

[44] I. Karunaratne, A. S. Atukorale, and H. Perera, "Surveillance of Human-Computer Interactions: A Way Forward to Detection of Users' Psychological Distress," in Colloquium on Humanities, Science and Engineering (CHUSER), 2011, pp. 491-496: IEEE.

[45] Y. Qi, C. Reynolds, and R. W. Picard, "The Bayes Point Machine for Computer-User Frustration Detection Via Pressuremouse," in Proceedings of the 2001 workshop on Perceptive user interfaces, 2001, pp. 1-5: ACM.

[46] U. Lundberg et al., "Psychophysiological Stress and Emg Activity of the Trapezius Muscle," International journal of behavioral medicine, vol. 1, no. 4, pp. 354-370, 1994.

[47] J. Wahlström, "Ergonomics, Musculoskeletal Disorders and Computer Work," Occupational Medicine, vol. 55, no. 3, pp. 168-176, 2005.

[48] P. W. Johnson, W. P. Smutz, and D. M. Rempel, "Computer Mouse Designed to Measure Finger Forces During Operation," in Proceedings of the 15th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Part 3 (of 3), 1993: Publ by IEEE.

[49] P. W. Johnson, M. Hagberg, E. W. Hjelm, and D. Rempel, "Measuring and Characterizing Force Exposures During Computer Mouse Use," Scandinavian journal of work, environment & health, pp. 398-405, 2000.

[50] J. Wahlström, M. Hagberg, P. Johnson, J. Svensson, and D. Rempel, "Influence of Time Pressure and Verbal Provocation on Physiological and Psychological Reactions During Work with a Computer Mouse," European journal of applied physiology, vol. 87, no. 3, pp. 257-263, 2002.

[51] P. H. Dietz, B. Eidelson, J. Westhues, and S. Bathiche, "A Practical Pressure Sensitive Computer Keyboard," in Proceedings of the 22nd annual ACM symposium on User interface software and technology, 2009, pp. 55-58: ACM.

[52] S. O'Dea. (2020, 03/04/2020). Number of Smartphone Users Worldwide from 2016 to 2021. Available: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/

[53] A. Kunst. (2019, 03/04/2020). How Much Time Do You Spend Daily on Your Smartphone? Available: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/

[54] S. Ghosh, N. Ganguly, B. Mitra, and P. De, "Evaluating Effectiveness of Smartphone Typing as an Indicator of User Emotion," in 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), 2017, pp. 146-151: IEEE.

[55] M. Exposito, J. Hernandez, and R. W. Picard, "Affective Keys: Towards Unobtrusive Stress Sensing of Smartphone Users," in Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, 2018, pp. 139-145.

[56] R. Wampfler, S. Klingler, B. Solenthaler, V. R. Schinazi, and M. Gross, "Affective State Prediction Based on Semi-Supervised Learning from Smartphone Touch Data," in Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 2020, pp. 1-13.

[57] Z. Sarsenbayeva et al., "Measuring the Effects of Stress on Mobile Interaction," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 3, no. 1, pp. 1-18, 2019.

[58] A. Leow et al., "247. Biaffect: Passive Monitoring of Psychomotor Activity in Mood Disorders Using Mobile Keystroke Kinematics," Biological Psychiatry, vol. 85, no. 10, pp. S102-S103, 2019.

[59] H. Huang, B. Cao, S. Y. Phillip, C.-D. Wang, and A. D. Leow, "Dpmood: Exploiting Local and Periodic Typing Dynamics for Personalized Mood Prediction," in 2018 IEEE International Conference on Data Mining (ICDM), 2018, pp. 157-166: IEEE.

[60] H. Lee, Y. S. Choi, S. Lee, and I. Park, "Towards Unobtrusive Emotion Recognition for Affective Social Communication," in 2012 IEEE Consumer Communications and Networking Conference (CCNC), 2012, pp. 260-264: IEEE.

[61] D. Carneiro, J. C. Castillo, P. Novais, A. Fernández-Caballero, and J. Neves, "Multimodal Behavioral Analysis for Non-Invasive Stress Detection," Expert Systems with Applications, vol. 39, no. 18, pp. 13376-13389, 2012.

[62] C. Kirschbaum, K.-M. Pirke, and D. H. Hellhammer, "The 'Trier Social Stress Test'–a Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting," Neuropsychobiology, vol. 28, no. 1-2, pp. 76-81, 1993.

[63] P.-M. Lee, W.-H. Tsui, and T.-C. Hsiao, "The Influence of Emotion on Keyboard Typing: An Experimental Study Using Auditory Stimuli," PloS one, vol. 10, no. 6, p. e0129056, 2015.

[64] H.-R. Lv, Z.-L. Lin, W.-J. Yin, and J. Dong, "Emotion Recognition Based on Pressure Sensor Keyboards," in International Conference on Multimedia and Expo, 2008, pp. 1089-1092: IEEE.

[65] A. Kołakowska, "A Review of Emotion Recognition Methods Based on Keystroke Dynamics and Mouse Movements," in 6th International Conference on Human System Interaction (HSI 2013), 2013, pp. 548-555: IEEE.

[66] E. B. Joffe and K.-S. Lock, Grounds for Grounding: A Circuit to System Handbook. John Wiley & Sons, 2011.

[67] R. S. Lazarus and S. Folkman, Stress, Appraisal, and Coping. Springer publishing company, 1984.

[68] C. J. Golden and S. M. Freshwater, "Stroop Color and Word Test," age, vol. 15, p. 90, 1978.

[69] P. Karthikeyan, M. Murugappan, and S. Yaacob, "A Review on Stress Inducement Stimuli for Assessing Human Stress Using Physiological Signals," in 2011 IEEE 7th International Colloquium on Signal Processing and its Applications, 2011, pp. 420-425: IEEE.

[70] J. Tulen, P. Moleman, H. Van Steenis, and F. Boomsma, "Characterization of Stress Reactions to the Stroop Color Word Test," Pharmacology Biochemistry and Behavior, vol. 32, no. 1, pp. 9-15, 1989.

[71] B. Pehlivanoğlu, N. Durmazlar, and D. Balkancı, "Computer Adapted Stroop Colour-Word Conflict Test as a Laboratory Stress Model," Erciyes Medical Journal, vol. 27, no. 2, pp. 58-63, 2005.

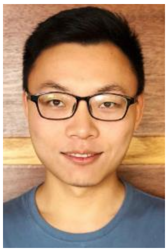
[72] M. Svetlak, P. Bob, M. Cernik, and M. Kukleta, "Electrodermal Complexity During the Stroop Colour Word Test," Autonomic Neuroscience, vol. 152, no. 1-2, pp. 101-107, 2010.

[73] S. J. Guy, "Story of Golden Locks," ed. New York City: The Met Fifth Avenue, 1870.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2021.3061417, IEEE Transactions on Affective Computing

[74] Y. M. Lim, A. Ayesh, and M. Stacey, "Detecting Cognitive Stress from Keyboard and Mouse Dynamics During Mental Arithmetic," in 2014 Science and Information Conference, 2014, pp. 146-152: IEEE.
[75] C. Ring, M. Drayson, D. G. Walkey, S. Dale, and D. Carroll, "Secretory Immunoglobulin a Reactions to Prolonged Mental Arithmetic Stress: Inter-Session and Intra-Session Reliability," Biological psychology, vol. 59, no. 1, pp. 1-13, 2002.
[76] Y. Noto, T. Sato, M. Kudo, K. Kurata, and K. Hirota, "The Relationship between Salivary Biomarkers and State-Trait Anxiety Inventory Score under Mental Arithmetic Stress: A Pilot Study," Anesthesia & Analgesia, vol. 101, no. 6, pp. 1873-1876, 2005.
[77] M. M. Bradley and P. J. Lang, "Measuring Emotion: The Self-Assessment Manikin and the Semantic Differential," Journal of behavior therapy and experimental psychiatry, vol. 25, no. 1, pp. 49-59, 1994.
[78] S. G. Hart and L. E. Staveland, "Development of Nasa-Tlx (Task Load Index): Results of Empirical and Theoretical Research," in Advances in Psychology, vol. 52: Elsevier, 1988, pp. 139-183.
[79] M. Rybnik, M. Tabedzki, and K. Saeed, "A Keystroke Dynamics Based System for User Identification," in 2008 7th Computer Information Systems and Industrial Management Applications, 2008, pp. 225-230: IEEE.
[80] R. A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems," Annals of eugenics, vol. 7, no. 2, pp. 179-188, 1936.
[81] P. Tsiamyrtzis, M. Dcosta, D. Shastri, E. Prasad, and I. Pavlidis, "Delineating the Operational Envelope of Mobile and Conventional Eda Sensing on Key Body Locations," in Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 2016, pp. 5665-5674: ACM.
[82] D. Shastri, M. Papadakis, P. Tsiamyrtzis, B. Bass, and I. T. Pavlidis, "Perinasal Imaging of Physiological Stress and Its Affective Potential," IEEE Trans. Affective Computing, vol. 3, no. 3, pp. 366-378, 2012.

Dennis R. da Cunha Silva received his B.S. degree in computer engineering and M.S. degree in systems engineering from the University of Pernambuco, Recife, Brazil, in 2012 and 2014, respectively. He is now a computer science Ph.D. student at Texas A&M University, College Station. His current research interests include affective computing, machine learning, and wearable sensors.



Zelun Wang received his B.Eng. degree in automation from the Xi’an Jiaotong University, Xi’an, China, in 2014. He is currently a computer science Ph.D. student at Texas A&M University, College Station. His current research interests include machine learning, wearable sensors, and natural language processing.

Ricardo Gutierrez-Osuna received the B.S. degree in electrical engineering from the Polytechnic University of Madrid, Madrid, Spain, in 1992 and M.S. and Ph.D. degrees in computer engineering from North Carolina State University, Raleigh, in 1995 and 1998, respectively. He is a Professor in the Department of Computer Science and Engineering, Texas A&M University, College Station. His current research interests include voice and accent conversion, speech and face perception, wearable physiological sensors, and active sensing.