Science Arts & Métiers (SAM) is an open access repository that collects the work of Arts et Métiers ParisTech researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: https://sam.ensam.eu
Handle ID: http://hdl.handle.net/10985/7878

To cite this version:

Any correspondence concerning this service should be sent to the repository Administrator: archiveouverte@ensam.eu
This article proposes a method based on wavelet transform and neural networks for relating pupillary behavior to psychological stress. The proposed method was tested by recording pupil diameter and electrodermal activity during a simulated driving task. Self-report measures were also collected. Participants performed a baseline run with the driving task only, followed by three stress runs where they were required to perform the driving task along with sound alerts, the presence of two human evaluators, and both. Self-reports and pupil diameter successfully indexed stress manipulation, and significant correlations were found between these measures. However, electrodermal activity did not vary accordingly. After training, the four-way parallel neural network classifier could guess whether a given unknown pupil diameter signal came from one of the four experimental trials with 79.2% precision. The present study shows that pupil diameter signal has good discriminating power for stress detection.

1. INTRODUCTION

Stress detection and measurement are important issues in several human–computer interaction domains such as Affective Computing, Adaptive Automation, and Ambient Intelligence. In general, researchers and system designers seek to estimate the psychological state of operators in order to adapt or redesign the working environment accordingly (Sauter, 1991). The primary goal of such adaptation is to enhance overall system performance, trying to reduce workers’ psychophysical detriment (e.g., Czaja & Sharit, 1993; Dennerlein, Becker, Johnson, Reynolds, & Picard, 2003; Fujigaki & Mori, 1997). One key aspect of stress measurement concerns the recording of physiological parameters, which are known to be modulated by the autonomic nervous system (ANS). However, despite technological progress in biological signal acquisition, inferring psychological significance from physiological signals is still a major challenge as biological signal analysis has progressed less intensively (Cacioppo & Tassinary, 1990), and it can be stated that affect recognition has not reached a satisfying level yet (Mauss & Robinson, 2009; Van den Broek, Janssen, & Westerink, 2009).

This study describes a new method for stress measurement using pupil diameter (PD) signal analysis. It is well known that pupillometry is a reliable tool for studying cognitive and emotional processes (Granholt & Steinbauer, 2004; Kuhlmann & Böttcher, 1999). The pupil is the aperture of the iris, the pigmented structure containing two antagonistic muscle groups—the sphincter and the dilator muscles. The former constrict the pupil; the latter dilate the pupil. The human pupil is known to reflect the activity of the ANS: In particular, it has been shown that the pupil enlarges (mydriasis) as a consequence of mental effort exertion and various sources of psychological stress (see Beatty, 1982; Beatty & Lucero-Wagoner, 2000; Bradley, Miccoli, Escrig, & Lang, 2008; Goldwater, 1972; Partala & Surakka, 2003; Vo et al., 2008). After dilation, the pupil naturally tends to constrict (myosis) back to previous diameter. Thus, we formulated the hypothesis that overall pupillary activity (i.e., PD oscillations) should be more intense under stressful conditions: Phasic changes should follow stressful events. Moreover, mean signal amplitude should also increase, indicating higher tonic level.

However, the use of PD as a dependent variable in psychological studies has important methodological implications. The main concern stems from the primary function of the pupil itself, that is, the regulation of the amount of light that enters the retina. The so-called light reflex occurs to avoid overexposure and retinal damage (Loewenfeld & Lowenstein, 1993, p. 136). Such a constriction is rapid (latency within 250 ms from stimulus onset) and proportional to stimulus intensity, and it is affected by individual differences. The return to prestimulus diameter (dark reflex) is much slower

Address correspondence to Marco Pedrotti, CHART/LUTIN (EA 4004), Cité des Sciences et de l’Industrie de la Villette, 30 Avenue Corentin Cartou, Paris 75930, France. E-mail: marco.pedrotti@hotmail.it

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hihc.
(see Beatty, 1986; Bergamin & Kardon, 2003; Ellis, 1981; Lanting et al., 1990). The near reflex (or accommodation response), that is, a near object requiring foveal focusing, causes pupil constriction, accompanied by eyes convergence and lenses curvature. Although this phenomena could regulate human pupil size from 1 up to 10 mm, very small (up to 0.01 mm) pupillary dilations can be elicited by various psychological manipulations (see Beatty & Luccero-Wagoner, 2000): The psychosensory reflex denotes pupil dilations evoked by sensory stimulation, whether auditory, tactile, gustatory, olfactory, or noxious. Beatty (1986) pointed out that this phenomena is a sort of bridge between the well-understood light and near reflexes and the more complex pupillary variations associated with cognitive processing. Many other factors can lead to pupil diameter variations: Janisse (1977) identified 23 sources of pupil variation, to which the effect of verbalization (Bernick & Oberlander, 1968) could be added. Therefore, extreme care should be taken to control as many potential bias sources as possible. Among these, illumination requires particular attention, although most studies fail to report such measurements.

Bearing these considerations in mind, there is a substantial interest and potential benefit in using PD for stress detection in applied studies: Unlike other physiological measures (e.g., cardiovascular activity, electrodermal activity, electroencephalography, etc.), the pupil can be measured unobtrusively. With modern devices, one camera is sufficient for pupil tracking, and there is no need for physical contact. Remote eye trackers have such properties, and recent research has demonstrated their suitability for pupillometry studies (Klingner, Kumar, & Hanrahan, 2008; Palinko & Kun, 2010, 2011; Palinko, Kun, Shyrokov, & Heeman, 2012). Besides PD, other eye-movement metrics (e.g., saccade parameters) are known to reflect stress-related variations (see Benedetto, Pedrotti, & Bridgeman, 2011; Di Stasi, Catena, Cañas, Macknik, & Martinez-Conde, 2013). Efforts are also being made to unobtrusively measure skin temperature and other ANS measures with imaging techniques (see, e.g., Nhan & Chau, 2010; Shastri, Merla, Tsiamyrtzis, & Pavlidis, 2009).

The present study concerns stress detection in a simulated driving task. With the aim of validating our results by means of triangulation (Van den Broek et al., 2009), we recorded—besides PD—participants’ self-reported stress levels and electrodermal activity (EDA; see section 3.5), a sensitive psychophysiological indicator of stress (Boucsein, 2012, p. 459).

The article is organized as follows: Existing PD data analysis techniques are introduced in section 1.1. Experimental setup and hypotheses of the present study are described in section 2. Section 3 describes the whole process of data acquisition and analysis. Statistical results are presented in section 4. The framework of automatic stress detection based on Wavelet-Neural Network is outlined in section 5: Wavelet multiresolution decomposition and Neural Networks are used for feature extraction and classification, respectively. Classification results and future challenges are discussed in section 6.

1.1. Methods for PD Data Analysis

Over the last decades, several methods have been used for analyzing PD data. The signals coming from PD recordings have been analyzed in both the time and frequency domains: State-of-the-art is briefly reviewed in this section. Regardless of the method employed for the analysis, eye-blink artifacts represent a common problem in video-pupillography: Most systems measure pupil size upon eye image processing (see Holmqvist et al., 2010; Wyatt, 2010). During eye-blinks the lid covers the eye, and the camera cannot detect the pupil. Because eye-tracking systems deal with this problem (loss of information) in a variety of ways (Gitelman, 2002), it is impossible to create a universal procedure to recover missing information. Several algorithms for eye-blink detection have been proposed, by both researchers directly interested in the eye-blink phenomenon and researchers faced with eye-blink artifacts (Pedrotti, Lei, Dzaack, & Rötting, 2011). Once blink onset and offset have been identified, missing/corrupted pupil data are usually estimated using linear (or cubic) interpolation, or even more sophisticated techniques such as moving average or support vector regression (see Nakayama, Yamamoto, & Kobayashi, 2012). Overall, pupil data preprocessing is necessary, because it is known that eye-blink artifacts have an impact on analysis results, both in the time and frequency domains (Nakayama, 2006; Nakayama & Shimizu, 2002, 2004).

The Task-Evoked Pupillary Response (TEPR) is a useful tool for PD signal analysis in the time domain. The main contributions on TEPR come from cognitive psychology. The rationale underlying this method is the same as the event-related potential (ERP) in electroencephalography (EEG) measurements. Because the magnitude of psychologically induced pupillary responses can be in the order of tenths—even hundredths—of millimeters, we can be more confident in associating such responses to a given stimulus if we know the exact time point of stimulus appearance. Like for event-related potentials, a time window (e.g., 500 ms) before stimulus onset is used as baseline, that is, the average PD $\bar{x}$ for the prestimulus time window is calculated. Subsequently, $\bar{x}$ is subtracted from each data point in the poststimulus time window (e.g., 5 s after stimulus onset). The resulting waveform—usually an average of several measurements (e.g., Figure 3A)—indicates the pupillary reaction to the stimulus, and parameters such as peak dila-
to temporally align PD time windows. In addition, TEPRs have been analyzed using principal component analysis and independent component analysis in an attempt to reduce the large number of time points to a smaller set—usually two or three factors (see Granholm & Verney, 2004; Jainta & Baccino, 2010; Kuchinke, Vo, Hofmann, & Jacobs, 2007; Verney, Granholm, & Marshall, 2004).

To the authors’ knowledge, Lüdtke, Wilhelm, Adler, Schaeffel, and Wilhelm (1998) first introduced the use of Fast Fourier Transform for pupil signal analysis in the frequency domain. Analyzing a signal in such domain allows to know, for example, whether significant changes happen within specific frequency bands. With the aim of detecting low-frequency fatigue-related pupillary oscillations, Lüdtke and colleagues evaluated the mean value of the amplitude spectrum for all the frequencies at 0.8Hz or lower. Nakayama and Shimizu (2004) found that the power spectrum density of pupil signals increases within certain band intervals (0.1–0.5 Hz and 1.6–3.5Hz) as a function of cognitive task difficulty. Lew, Dyre, Werner, Wotring, and Tran (2008) analyzed PD signals using the Short-Time Fourier Transform, which allows to extract the frequency information yet preserving the time domain.

One promising technique for reducing data complexity in recorded PD signals is wavelet analysis. Marshall (2000, 2002) first proposed the use of wavelet analysis for analyzing PD time series, and associated the occurrence of high-frequency variations (faster than 20 ms, i.e., > 50Hz) to instances of cognitive load. Since then, to our knowledge, few studies have applied wavelet transforms to PD signals: Shi et al. (2006) analyzed pupillary behavior in relation to a user’s visual ability; Pinzon-Morales and Hirata (2012) evaluated PD oscillations to estimate sleepiness levels. In the present study, we used the Discrete Wavelet Transform (DWT) as a tool to extract relevant signal features (i.e., low-frequency approximation), discarding the noise that appears in the high-frequency part of the signal (see section 3.4.4). In this respect, our approach is opposite to the one proposed by Marshall.

2. METHOD

2.1. Stimuli

Our rationale for stress manipulation implies the repeated performance of a simple driving task, to which we added different external stressful stimuli (see section 2.3). The protocol was approved by the French National Board of Informatics and Freedom (declaration n. 0727429; http://www.cnil.fr). Participants performed a simulated Lane Change Test (LCT), which consists of driving on a traffic-free straight three-lane road (see Figure 1A), changing lanes according to the information appearing on two identical road signs displayed concurrently every 150 m, on both sides of the road (ISO, 2010; Minin, Benedetto, Pedrotti, Re, & Montanari, 2011; Mitsopoulos-Rubens, Trotter, & Lenne, 2011).

The driving simulator consisted of seat (Playseat Evolution), steering wheel and pedals (Logitech G25, no gear-shift was used), and a 32-in. LCD monitor (Thomson 32LB220B4, 70 × 39cm, 1366 × 768px). The LCT software (http://www.corys.com) limited the maximum speed at 60 km/hr, so that participants could maintain this speed by simple flat-out. Each trial consisted of 18 lane changes, accomplished over 180 s (ISO, 2010). The average distance between participant and screen was 130 cm.

2.2. Participants

Thirty-three healthy people (all with valid driving license) participated in the experiment. Seventeen people were allocated to the experimental group, that is, the group that underwent stress manipulation, and the remaining 16 were assigned to the control group (see section 2.3). The experimental group contained nine women ($M_{age} = 38$ years, $SD = 15$) and eight men ($M_{age} = 43$ years, $SD = 9$). Eight women ($M_{age} = 42$ years, $SD = 8$) and eight men ($M_{age} = 41$ years, $SD = 13$) were assigned to the control group. All participants read and signed an informed consent and received a reward for every hour spent inside the laboratory. Participants were informed that they could leave the experiment at any time and for any reason. One participant from the control group quit the experiment during the second session because of simulator sickness.

Participants’ stress trait was measured with the State-Trait Anxiety Inventory (STAY-B; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983, translated in French by Schweitzer & Paulhan, 1990). To disclose possible social desirable responding, we asked participants to fill in the Social Desirability Scale (DS36; Tournois, Mesnil, & Kop, 2000). The experimental and control groups did not differ in terms of STAY-B.
scores, \(t(30) = 1.03, \text{ns}\), or on DS36 scores, \(t(30) = 0.1, \text{ns}\), for self-deception, and \(t(30) = 0.23, \text{ns}\), for other-deception.

2.3. Experimental Design and Procedure

Upon arrival at the lab, participants sat in a quiet room wherein they installed the electrodes for electrodermal measurement on their forehead (see section 3.5), read and signed an informed consent, and filled in the STAY-B and DS36 questionnaires (see section 2.2). Thereafter, participants moved to the simulation room, where they received onscreen instructions for the LCT (see ISO, 2010, Annex A) and performed a 1-min LCT training to disclose any potential issue (misunderstandings, simulator sickness, technical equipment, etc.).

Before beginning the LCT sessions, we recorded baseline physiological signals during a 90-s rest period (\(t_0\)) in which participants looked at a static picture of the LCT scenario. Subsequently, participants carried out four LCT trials (Figure 1B). The control group performed the four driving trials without any disturbing factor. The experimental group underwent two types of stressors, announced by means of screen instructions displayed before the LCT trial started.

Before the second trial (\(t_2\)), participants were informed that they would hear a sound if their driving behavior was not appropriate (i.e., beginning a lane change as soon as the symbols appear on a sign, but not before; see ISO, 2010, Annex A). Indeed, a sound was presented every 20 s, regardless of driving behavior. To avoid excessive habituation, three different sounds were used: A 8.6 Hz white noise, a police siren, and a 4 KHz tone (.wav files are available from the corresponding author). All sounds lasted 1 s, and they were presented in a pseudorandom order: each sound was presented three times, totaling nine sound presentations in the second trial.

Before the third trial (\(t_3\)), participants were informed that their driving performance would have been evaluated by two experts. After this announcement, two experimenters entered the test room. These “fake” experimenters (one man and one woman) wore white lab coats, they held a copybook which they used for taking notes, and they stood on the participant’s right (90° of visual angle) so that their presence could be perceived without disturbing the execution of the driving task. Moreover, a previously floor-pointed camera was turned toward the participant, so that she or he would believe she or he was being filmed. No stressful sounds were presented during the third trial.

Before the fourth trial (\(t_4\)), participants were informed that they would have been evaluated by the experimenters and alerted with sounds in case of incorrect driving behavior.

After each LCT trial, participants reported their perceived stress level (see section 3.3). Each trial lasted 3 min.

2.4. Hypotheses

For the experimental group, stress level should be lowest at \(t_1\)—where simple LCT performance is demanded—and highest at \(t_4\), where two concurrent stressors are associated with the LCT task. We cannot predict whether stress level will significantly differ between \(t_2\) and \(t_3\), that is, we do not know a priori whether alert sounds are more stressful than observers (or vice versa). What we know is that they are two different types of stressors, and they could be then classified.

For the control group, we expect stress level to significantly decrease from \(t_1\) to \(t_4\) as an effect of habituation.

The two groups should exhibit the same stress level at \(t_1\), as they bear exactly the same conditions until that point. We expect to find significant between-groups differences at \(t_2, t_3, t_4\).

We expect PD changes to reflect increased stress levels for the experimental group and decreased stress levels for the control group. Electrodermal and self-report measures are expected to correlate with pupillary behavior measures.

3. DATA ACQUISITION AND ANALYSIS

3.1. Apparatus Synchronization

The equipment used in the present study includes an eye tracker (RED 4, http://www.smivision.com) for PD measurement, an A/D converter (MP36R, http://www.biopac.com) for measuring illumination and EDA, a PC for stimulus (LCT) presentation, and a PC running Matlab (http://www.mathworks.com) for synchronization and auditory stimulus delivery (psyctoolbox.org; Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). All the systems were connected on a local area network, and synchronization was achieved using a combination of custom software written in Matlab and free software (synergy-foss.org). Each single experimental event (e.g., beginning of a LCT trial, presentation of a LCT sign, presentation of a sound, etc.) was marked—via TCP/IP messages—in the eye tracker’s and A/D converter’s log files for conjoint offline analysis. A detailed technical description goes beyond the scope of the present article, and interested readers may refer to the corresponding author for further information.

3.2. Illumination

Illumination was measured with an Extech 403125 luxmeter (http://www.extech.com). With the aim of recording the amount of light that impacted on the participant’s eyes, the light sensor was attached to a ceiling-mounted holder, placed 5 cm above the participant’s head, laterally centered with respect to her or his nose. The luxmeter was connected to a Biopac MP36R A/D converter which stored illumination values in lux at 50Hz sampling rate.

3.3. Visual Analog Scale

After each LCT trial (\(t_1, t_2, t_3, t_4\)) participants rated their perceived stress level using three on-screen visual analog scales (VASs). The VAS ranged from 0 (not at all) to 100 (maximum). The three dimensions were stress (“How much stress were you feeling during task performance?”), anxiety (“How much anxiety were you feeling during task performance?”), and avoidance
(“During task performance, to what extent were you willing to leave the situation?”). Correlation analysis revealed strong correlations between the three dimensions, except between stress $t_1$ and avoidance $t_1$, and between anxiety $t_1$ and avoidance $t_1$ (see Table 1).

### 3.4. Pupil Diameter

PD was recorded at 50Hz sampling rate using a SMI RED 4 remote video eye tracker. This system measures pupil size and eye movements by means of pupil and corneal reflection tracking (see Holmqvist et al., 2010), and has a precision of 0.01 mm for PD. PD signal was treated according to the following procedure:

**Step 1. Preprocessing:** Eye-blink artifacts were identified and replaced by linear interpolation (see section 3.4.1).

**Step 2. TEPRs extraction:** Pupillary responses following sound presentations (in $t_2$ and $t_4$) were evaluated to confirm the existence of pupillary reactions to stressful stimuli (see section 3.4.2).

**Step 3. Normalization:** PD values in $t_1$, $t_2$, $t_3$, $t_4$ were normalized for each participant, according to her or his average PD at rest (see section 3.4.3).

**Step 4. Analysis of variance (ANOVA):** We tested the hypothesis that our stress manipulation had an effect on average PD (see section 4.2).

**Step 5. Signal approximation extraction:** The normalized PD signals from $t_1$, $t_2$, $t_3$, $t_4$ were transformed by means of DWT. The Haar wavelet was used to decompose and transfer the signal into multiresolution representation (see section 3.4.4).

**Step 6. Classification with neural networks:** PD signal approximation was used as an input vector (feature) during the training and test stages (see section 5).

The first step (Preprocessing) is generally necessary regardless of the aim of any study. Concerning our study, we consider steps 2 and 4 (TEPRs extraction, ANOVA) mandatory for theoretically justifying the use of PD as a stress index, as they demonstrate the sensitivity of PD to stress manipulations. Normalization (step 3) is necessary because each person has her or his own PD at rest. Steps 5 and 6 (Signal approximation extraction and Classification) are an attempt to use PD for automatic stress measurement in applied contexts. With the aim of being able to measure stress levels over relatively short periods, we extracted and analyzed (in steps 4, 5, and 6) only the first 80s of PD data from each trial.

**3.4.1. Pupil diameter preprocessing.** The RED 4 provides two types of pupil measurement: (a) in pixel and (b) in millimeters (Figure 2). In the present study, millimeter values were preferred for PD signal analysis because they are calculated taking into account the distance between participant and camera, providing more reliable data than raw pixel measurement. Nonetheless, the RED 4 outputs useful information for blink detection in the pixel data. When a blink occurs, zeros—along with other physiologically impossible values—are recorded in the pixel output (see Figure 2, bottom line, right-hand scale). In a statistical perspective, such values can be considered as outliers of the PD distribution of a given data set. Blinks were detected as contiguous sets of outliers. Moreover, other blink markers in the eye-tracking protocol—such as the momentary loss of gaze position during blink—were combined to foster correct blink detection percentage (for a detailed description of the algorithm, see Pedrotti et al., 2011). Blink onset was defined as the third sample (60 ms) preceding the first zero observation: At this point, the lid starts its descent until the pupil is covered (in 79% of blinks; see Pedrotti et al., 2011). Blink offset was defined as the first valid sample after a blink: At this point, the pupil is visible to the eye tracker camera.

After identifying blink onsets and offsets in the pixel data, we replaced blink data in the millimeter output by means of linear interpolation, using blink onset as starting point, and two samples after blink offset as ending point. An example of the result of this preprocessing procedure is shown in Figure 2 (top line, left-hand scale).

### TABLE 1

<table>
<thead>
<tr>
<th></th>
<th>anxiety $t_1$</th>
<th>anxiety $t_2$</th>
<th>anxiety $t_3$</th>
<th>anxiety $t_4$</th>
<th>avoidance $t_1$</th>
<th>avoidance $t_2$</th>
<th>avoidance $t_3$</th>
<th>avoidance $t_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>stress $t_1$</td>
<td>.93**</td>
<td>.87**</td>
<td>.95**</td>
<td>.61**</td>
<td>.76**</td>
<td>.51**</td>
<td>.74**</td>
<td></td>
</tr>
<tr>
<td>stress $t_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stress $t_3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stress $t_4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>anxiety $t_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.13</td>
<td>.76**</td>
<td>.51**</td>
<td>.74**</td>
</tr>
<tr>
<td>anxiety $t_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>anxiety $t_3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>anxiety $t_4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. N = 32 (one participant quit the experiment after $t_2$ because of simulator sickness). *$p < .05. **p < .005.
3.4.2. Task-Evoked Pupillary Responses. TEPRs were extracted—from the experimental group—for each sound presentation in t2 and t4 (see Figure 1B and section 2.3). Baseline pupil diameter was computed as the mean PD in the 200-ms prestimulus time window. Figure 3A shows the average pupillary response (solid line, left-hand scale) from 288 waveforms (16 participants × 9 sounds × 2 trials). The dashed line (right-hand scale) shows the average illumination, measured in synchrony with PD (see section 3.1). A typical phasic pupillary reaction—dilation (mydriasis) followed by constriction (myosis)—occurred. With the aim of reducing the 150 data points (50 points × 3 s, 50Hz sampling rate) to a smaller set of factors, the data from the 288 TEPRs extracted were analyzed with a factor analysis using Statistica (http://www.statsoft.com). The 150 data points were treated as dependent variables serving as input for factor analysis. Factor loadings were extracted following a varimax rotation. Two factors could explain 82.34% of variance in the data, that is, pupillary response shapes are consistent across individuals and trials. Factor loadings plotted against time (Figure 3B) clearly show the separation between the rising (factor 1, mydriasis, 62.91% of explained variance) and falling (factor 2, myosis, 19.43% of explained variance) part of the pupillary response.

1One participant’s data were excluded because of poor recording quality.
depicted in Figure 3A. Moreover, the absence of measured relevant illumination changes allows us to associate the recorded waveform to the stress elicited by experimental manipulation (sound delivery associated with poor driving performance).

3.4.3. Pupil diameter normalization. The TEPR does not require previous data normalization, as the baseline PD is recalculated for each event, based on a short prestimulus time window (200 ms in our case): This procedure can be viewed as a sort of normalization, in that a participant- and moment-specific PD value is subtracted from absolute PD values. However, before performing any interindividual comparison (such as an ANOVA), PD values should be normalized, as it is known that PD at rest differs between people. For each participant, we calculated mean PD (μPD₀) at rest (t₀ in Figure 1B). Subsequently, μPD₀ was subtracted from each PD data point in t₁, t₂, t₃, t₄. This procedure allows for later comparisons of PD between participants.

3.4.4. Signal approximation extraction. Combination of wavelet and neural networks has been accepted as an accurate method for feature extraction and classification (Minu, Lineesh, & Jessy John, 2010). Any noisy signal imposes some uncertainty to the calculation and, consequently, to the results. Therefore, before using PD signal as input for the neural network classifier, we need to remove noise from the signal. Denoising could improve classification performance, as it would increase the signal-to-noise ratio. Moreover, it would reduce computational costs, that is, a shorter time to obtain results: This latter aspect is important in a real-time application perspective, although we focus on offline analysis for the present study.

Several mathematical approaches could be used for this purpose: We chose the DWT because (a) it allows removing noise yet preserving the original shape of the signal and (b) it encompasses a down-sampling procedure, which reduces computation time.

Mathematically, a signal (time series) \( x(t) \in L(R) \) can be decomposed into linear combination of a set of \( n \) base functions \( \{\phi₀, \phi₁, \ldots, \phi_n\} \) if the signal is in the space spanned by the basis. Then, \( x(t) \) can be decomposed into a linear combination of the base functions (Mallat, 1989):

\[
x(t) = \sum_{k} a_k \phi_k(t) \quad k \rightarrow n (k \rightarrow \infty), k \in \mathbb{Z}
\]

where \( k \) is an integer index of the finite or infinite sum and \( a_k, \phi_k(t) \) are expansion coefficients and functions, respectively. This representation is the most common form of multispectrum decomposition. Consider two sets of base functions:

1. \( \phi_{j,k}(t) = 2^{-\frac{j}{2}} \phi \left( 2^{-j}t - kx \right), j > 0, k \in \mathbb{Z} \)
2. \( \phi_k(t) = e^{\frac{-2\pi i}{n}} \)

If Item 1 or 2 are substituted in Equation 1, the wavelet and Fourier decompositions will be achieved, respectively. These are two well-known examples of decomposition of a signal into primitive or fundamental constituents of their spaces. In fact, the Fourier series decomposes a signal into a set of sine and cosine functions. By DWT in multiresolution analysis, a signal is represented by a sum of a set of more flexible functions called mother wavelet, which are localized both in time and frequency.

Any wavelet decomposition consists of two parts: approximation and detail. Approximation refers to the overall, general form of the signal (e.g., low-frequency component), whereas detail better explains the high-frequency information such as edges, discontinuities, sharp points, and so on. Approximation and detail coefficients of a given discrete signal \( x[n] \) can be extracted by low-pass and high-pass filtering, respectively. Figure 4B shows an example of signal approximation extraction using the Haar wavelet as mother wavelet, and five decomposition levels (i.e., the output of level \( n \) is used as input for level \( n+1 \)). It seems evident that approximation preserves the original shape of the signal, whereas noise is discarded: In most cases, noise resides in the high-frequency part (Hamid, Nawi, & Ghazali, 2011). In the original signal (Figure 4A) 4,000 data points (coefficients) are needed to describe 80 s of pupil diameter (sample rate is 50Hz). Each wavelet decomposition level encompasses a down-sampling by a 2 factor. After five decompositions, sampling rate is reduced from 50Hz to 1.5625Hz, and 125 coefficients are sufficient to describe 80 s of pupil diameter. The vectors containing the 125 coefficients will be used as input for a neural network classifier (see section 5).

3.5. Electrodermal Activity

Skin conductance (SC) was recorded using the exosomatic method with Direct Current (Boucsein, 2012). Two Biopac EL507 disposable circular electrodes (Ag/AgCl, 1cm diameter circular contact area, 0.5% Chloride) were attached to the participant’s forehead. Although palmar and plantar zones would be preferable for EDA recording, as they have higher sweat glands density (Dawson, Schell, & Filion, 2000; Sato, Kang, Saga, & Sato, 1989), the driving task employed in the present study required both hands and feet to be completely free. The electrodes were fixed to the skin upon participant’s arrival at the lab, assuring a minimum delay of 15 min before the recordings started. This time is sufficient to allow good electrical contact between the skin and the electrode surface. An elastic headband was used to prevent artifacts due to wire movements (Boucsein et al., 2012). The electrodes were connected to the Biopac MP36R A/D converter (50Hz sampling rate). For the EDA recording channel, a low-pass filter was applied (35Hz cutoff frequency). SC signals were treated with the following procedure:

Step 1. Filtering: A low-pass filter was applied, with 2Hz cutoff frequency.
Step 2. Down-sampling: Sampling rate was reduced from 50Hz to 10Hz.
Step 3. Transformation: Data were transformed with the formula \( y = \log ( 1 + x ) \) (Boucsein, 2012). The respective units are labeled as log μS.
Step 4. SC Response (SCR) extraction: SCRs were extracted following sound presentations in $t_2$ and $t_4$ (experimental group). The average SC in the 200-ms prestimulus time window was subtracted from each SC data point in the 10-s poststimulus time window. Figure 5A shows the average SCR from 288 sound presentations (16 participants $^2$ $\times$ 9 sounds $\times$ 2 trials). Like for PD data (see section 3.4.2), we used SCR time-series as input for factor analysis: Two factors could explain 75.52% of variance (54.24% Factor 1, 21.28% Factor 2). Figure 5B shows factor loadings plotted against time. Unlike for TEPRs, the temporal separation between the two factors (roughly at 4 s) does not match the peak of the average SCR (Figure 5A): Factor interpretation is harder in this case; however, the overall proportion of explained variance tells that SCRs, like TEPRs, have a uniform structure across participants and trials.

$^2$ One participant was excluded because of too many artifacts in the SC record.

Step 5. Non-Specific Electrodermal Response Frequency (NS.EDR freq.) extraction (see section 4.3.1)

Step 6. EDA Area extraction (see section 4.3.2)
4. RESULTS

4.1. Visual Analog Scale

VAS stress scores from 32 participants were analyzed with a repeated-measures ANOVA (rmANOVA) using trial (t1, t2, t3, t4) as within-factor and group (experimental, control) as between-factor (see Figure 6). A Trial × Group interaction effect was found, $F(3, 90) = 5.41, p < .005, \eta^2_p = .15$. The effect of group is also significant, $F(1, 30) = 7.25, p < .05, \eta^2_p = .19$, with higher scores for the experimental group.

The difference on VAS stress scores between the experimental and control group is not significant at $t_1$, $F(1, 30) = 0.03, ns$. VAS stress scores are significantly higher for the experimental group at $t_2$, $F(1, 30) = 10.06, p < .005, \eta^2_p = .25$; $t_3$, $F(1, 30) = 7.79, p < .01, \eta^2_p = .21$; $t_4$, $F(1, 30) = 5.64, p < .05, \eta^2_p = .16$.

A trial effect, $F(3, 48) = 4.6, p < .01, \eta^2_p = .22$, was found within the experimental group: VAS stress scores are significantly higher at $t_2$ with respect to $t_1$, $F(1, 16) = .872, p < .01, \eta^2_p = .15$; at $t_3$ with respect to $t_1$, $F(1, 16) = 9.78, p < .01, \eta^2_p = .38$; at $t_4$ with respect to $t_1$, $F(1, 16) = 4.46, p = .05, \eta^2_p = .22$. The differences between $t_2$ and $t_3, t_2$ and $t_4, t_3$ and $t_4$ are not significant.

Within the control group, there is no trial effect.

4.2. Pupil Diameter

Before carrying out any between-groups comparison at $t_1, t_2, t_3, t_4$, we verified that average PD at rest ($t_0$) did not differ between the experimental and control groups. Preprocessed average PD was calculated for 16 participants of the experimental group (one participant was excluded because of poor recording quality) and for 13 participants of the control group (two participants were excluded because of poor recording quality, and one participant was excluded because she quit the experiment after $t_2$). The data were analyzed in an independent samples $t$ test, which confirmed no difference on average PD at $t_0$ between the two groups, $t(27) = 1.52, ns$.

We then tested the hypothesis that stress manipulation had an effect on mean PD across the experimental trials $t_1, t_2, t_3, t_4$. Normalized average pupil diameters were calculated for the experimental group (16 participants; one was excluded because of poor recording quality) and the control group (13 participants; two were excluded because of poor recording quality, one was excluded because she quit the experiment after $t_2$) at $t_1, t_2, t_3, t_4$. Data were analyzed with a rmANOVA, using trial ($t_1, t_2, t_3, t_4$) as within-factor and group (experimental, control) as between factor. A significant Trial × Group interaction effect was found, $F(3, 81) = 9.14, p < .001, \eta^2_p = .25$ (see Figure 7). The effect of group was also present, $F(1, 27) = 4.01, p = .05, \eta^2_p = .13$, in that normalized average pupil diameter was higher for the experimental with respect to the control group. When $t_1$ was removed from the within factors (as stress manipulation effectively started at $t_2$), the effect of group emerged more clearly, $F(1, 27) = 6.31, p < .05, \eta^2_p = .19$.

There was no difference between the experimental and control group at $t_1$, $F(1, 27) = 0.05, ns$, whereas PD was significantly larger for the experimental group at $t_2$, $F(1, 27) = 6.93, p < .05, \eta^2_p = .25$; $t_3$, $F(1, 27) = 5.35, p < .05, \eta^2_p = .16$; $t_4$, $F(1, 27) = 5.31, p < .05, \eta^2_p = .16$.

![FIG. 6. Average VAS scores (stress dimension) for each group and experimental trial. Note. Vertical bars denote 95% confidence intervals (mean ± 2SE). $N = 32$ (17 experimental + 15 controls).]

---

3One participant quit the experiment after $t_2$ because of simulator sickness.
The effect of trial was significant, $F(3, 45) = 3.45, p < .05$, $\eta^2_p = .19$, within the experimental group: PD was significantly larger at $t_2$ with respect to $t_1$, $F(1, 15) = 12.63, p < .005$, $\eta^2_p = .46$; $t_3$ with respect to $t_1$, $F(1, 15) = 7.42, p < .05$, $\eta^2_p = .33$. The differences between $t_1$ and $t_4$, $t_2$ and $t_3$, $t_2$ and $t_4$, $t_3$ and $t_4$ were not significant.

Within the control group, the effect of trial was significant, $F(3, 36) = 7.22, p < .001$, $\eta^2_p = .37$: PD significantly decreased at $t_2$ with respect to $t_1$, $F(1, 12) = 7.7, p < .05$, $\eta^2_p = .39$; at $t_2$ with respect to $t_1$, $F(1, 12) = 13.74, p < .005$, $\eta^2_p = .53$; at $t_4$ with respect to $t_1$, $F(1, 12) = 12.83, p < .005$, $\eta^2_p = .52$. The differences between $t_2$ and $t_3$, $t_2$ and $t_4$, $t_3$ and $t_4$ were not significant.

4.3. Electrodermal Activity

4.3.1. Nonspecific Electrodermal Response Frequency. NS.EDR freq., that is, the number of SCRs in absence of apparent stimulation, is thought to be an indicator of negatively tuned emotional states such as stress (Boucsein, 2012), in that NS.EDR freq. should increase under stressful conditions. NS.EDR freq. scores were analyzed with an rmANOVA using trial ($t_1$, $t_2$, $t_3$, $t_4$) as within factor and group (experimental, control) as between factor. No significant effects were found.

4.3.2. EDA area. The area (computed as time-integral) below a SC waveform can be used as a measure of emotional arousal (see Bach, Friston, & Dolan, 2010; Boucsein, 2012). Because SC level has great interindividual variability, we subtracted the estimated tonic level from each SC time series before computing area measures. Tonic level was estimated by means of deconvolution using the Ledalab package (Benedek & Kaernbach, 2010; http://www.ledalab.de). Figure 8A shows an example of the tonic-level estimation used in this study. The SC waveform—after subtraction of the estimated tonic level—shows a zero baseline (Figure 8B), making it possible to compare SC between individuals. These SC vectors were used as input for calculating area scores on a trial-by-trial basis. Areas were calculated by trapezoidal numerical integration using the Matlab trapz function. Area scores were then analyzed with an rmANOVA using trial ($t_1$, $t_2$, $t_3$, $t_4$) as within factor and group (experimental, control) as between factor. No significant effects were found.

4.4. Bivariate Analysis

4.4.1. Correlation between subjective and physiological measures. It appears that normalized average VAS and PD scores have a similar pattern across the experimental trials, for both the experimental and control groups (see Figures 6 and 7). For the experimental group, both stress measures show a steep increase at $t_2$ with respect to $t_1$, followed by a decrease at $t_3$. Finally, another increase occurs at $t_4$. For the control group, both measures decrease at $t_2$ and maintain a relatively stable level until $t_4$. Significant correlations were found between normalized average PD and VAS stress scores at $t_1$ ($r = .4, p < .05$), $t_2$ ($r = .41, p < .05$), $t_3$ ($r = .44, p < .05$), and $t_4$ ($r = .4, p < .05$).

4.4.2. Correlation between PD and illumination. Illumination is an important factor influencing PD (see section 1). In this experiment, both stimulus and ambient illumination were kept constant during the whole experiment. Despite this, illumination data analysis revealed a slight illumination decrease for the experimental group at $t_3$ and $t_4$ (see Figure 7). Postexperiment investigations revealed that the cause...
could be attributed to the presence of the “fake” experimenters in the test room. Any displacement in the test room—even apparently insignificant ones like removing a chair—could cause an illumination change, indexed by the high sensitivity of the luxmeter. Although such subtle changes are below visual threshold and could not influence pupil size (Loewenfeld & Lowenstein, 1993), we tested whether normalized average PD was correlated with average illumination at $t_1, t_2, t_3, t_4$. No significant correlations were found.

5. CLASSIFICATION OF PD WITH NEURAL NETWORKS

After verifying the sensitivity of PD to stress manipulations (see section 4.2), we used PD signal approximation as input for a classifier. Statistical analyses support our choice of PD as a stress measure, in that average PD is significantly larger for the experimental group—with respect to the control group—at $t_2, t_3, t_4$, whereas there is no difference at $t_1$. This is in line with our predictions, as there was no stress manipulation until the end of $t_1$, that is, the two groups were exactly in the same conditions before $t_2$. Further support for this consideration comes from the subjective stress ratings (VAS; see section 4.1).

The aim of this analysis stage is automatic stress classification using normalized pupil diameter as the only information source: for this purpose, we use only PD data from the experimental group, that is, the group that underwent stress manipulation.

Following our experimental plan (Figure 1B), four classes should be used—one class for $t_1$, one class for $t_2$, one class for $t_3$, and one class for $t_4$. The hypothesis underlying the experimental plan was that participants in the experimental group would feel more stressed as the experiment went on, with $t_4$ being the most stressful trial because of the cumulative effect of stressful sounds and human observers. Indeed, statistics revealed that PD significantly increased only at $t_2$ and $t_3$ with respect to $t_1$, that is, differences between the different types of stressors (sound at $t_2$, observation at $t_3$, and their combination at $t_4$) could not be revealed using statistical linear models (see section 4.2). Such statistics rely on mean and variance as basic features for discrimination. In contrast, neural networks have a nonlinear characteristic, which is imposed by nonlinear transfer functions such as logsig, tansig, and so on. Such a more sophisticated classifier could improve discrimination performance using the whole signal (or its approximation) as input. Specifically, PD signal approximations (see section 3.4.4) were used as input features for classification.

The classification procedure involves two stages. In Stage 1 (training), four binary neural network classifiers are trained. Each of these classifiers operates in one-versus-all mode, that is, the aim of the training here is to maximize recognition precision of one class with respect to all the other classes (e.g., maximize recognition of $t_1$ with respect to $t_2, t_3, t_4$).

In Stage 2 (test), the four binary classifiers are put in parallel. An unknown, unlabeled PD signal approximation $\vec{x}$ is given as input to each of the four binary classifiers. Each classifier returns a score $y$ (between 0 and 1), which can be interpreted as the probability that $\vec{x}$ belongs to a certain class (i.e., the degree to which an instance is a member of a class; see Fawcett, 2006). The final decision is made according to the highest score attributed to $\vec{x}$ by each of the four binary classifiers.

Data from 10 participants (randomly selected) were used in the training stage (10 participants $\times$ 4 classes, totaling
40 signals). Data from the remaining six participants were used in the test stage (6 participants × 4 classes, totaling 24 signals). Implementation details are outlined in the following sections.

5.1. Binary Classifiers Architectures

Figure 9 shows a schematic representation of the one-versus-all classification procedure: an 80-s artifact-free PD signal \( \vec{x}' \) (preprocessed normalized PD) is decomposed by means of DWT (see section 3.4.4) using the Haar mother wavelet. Signal approximation \( \vec{x}' \) is extracted and given as input to a binary neural network classifier. The classifier returns a score \( y \). In an ideal situation, the binary classifier “\( t_1 \) vs. \( t_2, t_3, t_4 \)” (i.e., the classifier specialized for recognizing PD signals coming from \( t_1 \) trials) assigns a score \( y = 1 \) to an input signal \( \vec{x}' \) recorded during a \( t_1 \) trial, and a score \( y = 0 \) to an input signal \( \vec{x}' \) recorded during a \( t_2, t_3, \) or \( t_4 \) trial.

Table 2 summarizes architectural details of the four one-versus-all binary classifiers.

5.2. Binary Classifiers Training

Each of the four binary classifiers was trained separately using the Matlab Neural Network Training tool (nntool). The Levenberg-Marquardt algorithm, which updates weight and bias values according to gradient descent and other conjugate gradient methods (Moré, 1978), was selected. Parameter values are reported in Table 3.

5.3. Four-Way Parallel Classifier Architecture

Figure 10 depicts the scheme of the four-way parallel classification procedure. An unknown, unlabeled PD signal approximation \( \vec{x}' \) (i.e., \( \vec{x}' \) has never been used in the training stage) is given as input to each of the four binary classifiers. Each classifier returns a score \( y \). Scores are stored in the 4-D vector \( \vec{y} \). In the example in Figure 10, the binary classifier “\( t_1 \) vs. \( t_2, t_3, t_4 \)” assigned a score of 0.9 to \( \vec{x}' \). All the scores assigned to \( \vec{x}' \) from the other binary classifiers are lower than 0.9; thus we conclude that \( \vec{x}' \) comes from a \( t_1 \) trial.

5.4. Four-Way Parallel Classifier Test

Data from six participants were used for test, totaling 24 (6 participants × 4 classes) signal approximations \( \vec{x}' \). The four-way parallel classifier has a precision of 79.2%, that is, five misclassifications out of 24 signals. Detailed confusion matrix is shown in Table 4.

6. DISCUSSION

Among several psychophysiological correlates of stress—such as cardiovascular activity, electrodermal activity, respiration—we focused on PD because it can be measured in a completely unobtrusive manner. This makes PD particularly attractive in a real-life implementation perspective, where stress level could be measured automatically, by using video cameras. We proposed a method for relating PD behavior to psychological stress and tested its validity in a simulated driving experiment. For ethical reasons, experimental stressors...
TABLE 2
Neural Network Classifiers Architectures

<table>
<thead>
<tr>
<th>Network</th>
<th>Hidden Layer I No. of Neurons</th>
<th>Hidden Layer I Transfer Function</th>
<th>Hidden Layer II No. of Neurons</th>
<th>Hidden Layer II Transfer Function</th>
<th>Output Layer No. of Neurons</th>
<th>Output Layer Transfer Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>t1 vs. t2, t3, t4</td>
<td>15</td>
<td>tan-sigmoid</td>
<td>8</td>
<td>log-sigmoid</td>
<td>1</td>
<td>pureline</td>
</tr>
<tr>
<td>t2 vs. t1, t3, t4</td>
<td>14</td>
<td>tan-sigmoid</td>
<td>7</td>
<td>log-sigmoid</td>
<td>1</td>
<td>pureline</td>
</tr>
<tr>
<td>t3 vs. t1, t2, t4</td>
<td>11</td>
<td>tan-sigmoid</td>
<td>6</td>
<td>log-sigmoid</td>
<td>1</td>
<td>pureline</td>
</tr>
<tr>
<td>t4 vs. t1, t2, t3</td>
<td>11</td>
<td>tan-sigmoid</td>
<td>6</td>
<td>log-sigmoid</td>
<td>1</td>
<td>pureline</td>
</tr>
</tbody>
</table>

Note. Architecture details of the four binary classifiers (see Figure 9). Each classifier is designed to maximize recognition of one class with respect to all the other classes (e.g., maximize recognition of $t_1$ with respect to $t_2$, $t_3$, $t_4$).

TABLE 3
Parameter Values of the Neural Network Training Algorithm

<table>
<thead>
<tr>
<th>Parameter</th>
<th>epochs</th>
<th>time</th>
<th>goal</th>
<th>grad$_{min}$</th>
<th>$\mu$</th>
<th>$\mu$$_{dec}$</th>
<th>$\mu$$_{inc}$</th>
<th>$\mu$$_{max}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>1000</td>
<td>infinite</td>
<td>0</td>
<td>1e-08</td>
<td>0.001</td>
<td>0.1</td>
<td>10</td>
<td>1e10</td>
</tr>
</tbody>
</table>

Note. epochs = maximum number of iterations; time = time limit before algorithm stops; goal = target gradient value; grad$_{min}$ = minimum gradient magnitude; $\mu$ = convergence factor (see Barman & Chowdhury, 2012).

FIG. 10. Schematic representation of the four-way parallel neural network classifier.
were conceived to elicit moderate stress levels, as confirmed by subjective ratings (VAS) results (see Figure 6). The fact that PD could index such mild variations is encouraging for further developments in which higher stress levels could be detected.

We proved the sensitivity of PD as a stress concomitant, in both event-related and general state paradigms. For the event-related part, we showed how pupillary responses (TEPRs) follow the presentation of auditory stimuli associated with poor task performance: Participants in the experimental group were told that they would hear a sound alert if their driving behavior was not appropriate. Because of the simplicity of the driving task (LCT), participants were expected to feel disoriented, irritated, frustrated at every sound presentation: During task (LCT), participants were expected to feel disoriented, irritated which one is more stressful: Within the experimental group, significant differences were found only between the nonstress trials (t1) and stress trials (t2, t3); that is, we could not discriminate between t2 and t3, t2 and t4, t3 and t4. Moreover, although average PD is higher at t4 with respect to t1 (see Figure 7), the difference is not statistically significant: Two factors could explain this. First, humans are likely influenced by habituation: At t4, participants have already had some experience with both sounds (since t2) and observers (since t3); thus it is reasonable that they feel less stressed at t4. Moreover, at t4 they perform the LCT for the fourth time, which should also lower the stress induced by the LCT itself. We suggest that habituation (to both the LCT and the stressors) played a major role than the summation of two stressors (these two stressors are no more a novelty at t4). Second, we are dealing with linear statistical models (ANOVA) and it should be clear that psychophysiological phenomena cannot be completely described in this domain.

With the aim of improving discrimination performance in a real-life oriented context, we devised an automated classifier. The results provided by the classifier are promising, yet we underline that they come from PD data recorded under controlled illumination. Designing such an autonomous stress measurement system, which relies solely on a short period of a time series (or a real-time signal), raises at least one question: What would happen if the level of noise increases, for example, because of environmental effects? In the case of PD, we can regard environmental illumination as a major noise source. Because in our experiment illumination was controlled, we cannot answer this question with empirical evidence. Neural networks were essentially inspired from studies of brain structure (Widrow, Gupta, & Maitra, 1973), and they are known to be remarkably tolerant to noise in input data. However, further research is needed to integrate—in our system—pupillary light reflex information, which could be estimated by measuring illumination and other factors (e.g., Watson & Yellot, 2012).

Concerning EDA measures, we obtained contrasting results: Event-related concomitants of stressful sounds (i.e., SCRs) were found, and factor analysis confirmed their relatively stable response behavior (75.52% of explained variance with two factors). However, NS.EDR freq.—an indicator of adverse emotional states in human–computer interaction (Boucsein, 2012)—did not return the expected results, in that it did not increase with stress level. It is known that NS.EDR freq. calculation can provide different results depending on event-detection algorithms: Overlapping SCRs are likely, especially in applied contexts such as the present experiment. We tried to overcome this problem by extracting an area measure of EDA, but the results remain unclear: It might be that the stress level in our experiment is too low—according to subjective
measures—to be indexed by electrophysiological measures. Another possible and more “technical” explanation could be the fact that we placed the electrodes on the participants’ foreheads—because of experimental constraints (i.e., driving task)—instead of using palms and soles as recording sites: Dawson et al. (2000) suggested that emotion-evoked sweating is more evident in palmar and plantar zones because of higher sweat gland density (600/cm² for palms, 700/cm² for soles, 181/cm² for the forehead; see Sato et al., 1989). Finally, recent studies suggested that EDA signals have less discriminating power—compared to PD signals—for stress classification (Ren, Barreto, Gao, & Adjouadi, 2013; Zhai & Barreto, 2006).

Subjective ratings showed moderate yet significant correlations with normalized average PD. Although this result is encouraging, we remark that relying only on ANS measures is not the key for automatic stress measurement. A widely accepted perspective states that emotions are organized according to two principal dimensions, that is, arousal and valence (Mauss & Robinson, 2009). The former ranges from states of low activation (e.g., calm) to states of high activation (e.g., excited), whereas the latter counters positively tuned states (e.g., happy) versus negatively tuned ones (e.g., angry). Whereas ANS measures are known to be reliable indexes of arousal, they don’t give us indications about valence (see, e.g., Janisse, 1977). Thus, pupil diameter could increase because of positive stress (eustress) or negative stress, in the same way. In the present study, we examined the valence dimension by means of subjective ratings. However, this requires some active intervention by the user’s side, which is not suitable for an automatic stress measurement system. In this perspective, automatic valence indexes will be investigated in future research, with particular attention toward facial expressions: Like for PD, these measures can be acquired unobtrusively by using video cameras.

FUNDING
This research was supported by the French Minister of Economy, Finances and Industry, under contract n. FI105041V (MASSAI).

ACKNOWLEDGEMENT
We thank the anonymous reviewers for their helpful comments.

REFERENCES


ABOUT THE AUTHORS

Marco Pedrotti received his Ph.D. in Applied Psychology and Ergonomics from the Università di Torino in 2011. He is currently a post-doc researcher at the LUTIN lab (http://www.lutin-userlab.fr) for Université Pierre et Marie Curie, Paris. His research focuses on unobtrusive sensing of emotions and operator’s state.

Mohammad Ali Mirzaei is a Ph.D. Student in Arts et Métiers ParisTech, France. He received his MS in microelectronics engineering (signal processing and SoC systems) in Imperial College, London. His research focuses on signal/image processing and real-time systems.

Adrien Tedesco is a Ph.D. Student in the Laboratoire d’étude des mécanismes cognitifs at Université Lyon 2. He works as research engineer for Scientific Brain Training (Villeurbanne, France) and coordinator of the MASSAI project (www.projet-massai.com). His research investigates measures and effects of anxiety on cognitive performance.

Jean-Rémy Chardonnet received his Ph.D. degree in robotics from the Université de Montpellier II, France, in 2009. He was then with INRIA Grenoble, France. He is currently Assistant Professor at Arts et Métiers ParisTech and CNRS Le2i Institut Image, Chalon-sur-Saône, France. His research interests include virtual reality and multimodal interaction.

Frédéric Mérienne is full professor at Arts et Métiers ParisTech and head of the research team in Virtual Immersion of the Le2i laboratory (Burgundy University, Arts et Métiers ParisTech, CNRS). His research topics concern multisensory immersive techniques for interacting with a digital mock-up.

Simone Benedetto received his Ph.D. in Applied Psychology and Ergonomics from the Università di Torino in 2011. He was then with IFSTTAR, France. He is currently a post-doc researcher at the LUTIN lab (http://www.lutin-userlab.fr) for Université Paris 8. His research is focused on cognitive ergonomics.

Thierry Baccino is full professor of Cognitive Psychology at Université Paris 8 and scientific director of the LUTIN lab (http://www.lutin-userlab.fr). His research aims at interpreting and modeling visual search in real-life contexts such as reading, information retrieval, and image viewing.