

UBFC-Phys: A Multimodal Database For Psychophysiological Studies of Social Stress

Rita Meziati Sabour[✉], Yannick Benezeth[✉], Pierre De Oliveira, Julien Chappé, and Fan Yang[✉]

Abstract—As humans, we experience social stress in countless everyday-life situations. Giving a speech in front of an audience, passing a job interview, and similar experiences all lead us to go through stress states that impact both our psychological and physiological states. Therefore, studying the link between stress and physiological responses had become a critical societal issue, and recently, research in this field has grown in popularity. However, publicly available datasets have limitations. In this article, we propose a new dataset, UBFC-Phys, collected with and without contact from participants living social stress situations. A wristband was used to measure contact blood volume pulse (BVP) and electrodermal activity (EDA) signals. Video recordings allowed to compute remote pulse signals, using remote photoplethysmography (RPPG), and facial expression features. Pulse rate variability (PRV) was extracted from BVP and RPPG signals. Our dataset permits to evaluate the possibility of using video-based physiological measures compared to more conventional contact-based modalities. The goal of this article is to present both the dataset, which we make publicly available, and experimental results of contact and non-contact data comparison, as well as stress recognition. We obtained a stress state recognition accuracy of 85.48 percent, achieved by remote PRV features.

Index Terms—Psychophysiology, stress recognition, remote photoplethysmography, pulse rate variability, electrodermal activity

1 INTRODUCTION

PSYCHOPHYSIOLOGY is a branch of psychology that links human psychological states with physiological responses. Our psyche is constantly influenced by internal or external stimuli, inducing us to feel different emotions. During this process, the human body reacts through physiological signals that are controlled by the autonomic part of the peripheral nervous system. Psychophysiology is concerned with the *Autonomic Nervous System* (ANS) activity that is induced when an emotional state is experienced [1], [2].

Living in society confronts us with various situations that imply human interactions. Social psychology analyzes social phenomena that affect human behavior. Social psychophysiology combines social psychology with physiological signal analysis. In other words, it investigates the relationship between social behavior and physiological responses emitted along with. One particular issue that social psychology is interested in is social stress, which arises from the resistance a person lives when handling social situations. In our everyday life, we are constantly exposed to social stressors, such as giving a speech, talking to a person who has authority and meeting new people.

When a human is coping with a stressful event, his body transmits involuntary responses, all controlled by the part of the ANS called the *Sympathetic Nervous System* (SNS). Along with the SNS, the *Parasympathetic Nervous System* (PNS) works on balancing and regulating physiological signals. Heart rate variations, high blood pressure, sweating, are all examples of physiological reactions to stress. These phenomena can be quantified using several signals used in psychophysiological experiments. Among most frequently used signals we can cite the *Heart Rate* (HR), *Blood Volume Pulse* (BVP), *Heart Rate Variability* (HRV), and the *ElectroDermal Activity* (EDA), also called *Galvanic Skin Response* (GSR). Interested readers may refer, for example, to [3] where an interesting review of experimental studies related to emotional effects on physiological signals that are controlled by the ANS is realized.

The BVP is a quasiperiodic signal that is constituted of successive pulse peaks, generated with the pumping activity of the heart. Pulses naturally occur at non-regular times [4]. Computing the time intervals between the BVP pulses, leads to a signal called the *Pulse Rate Variability* (PRV). The reference method to obtain the BVP is *Photoplethysmography* (PPG), which is a technique that uses a light source and a receptor to quantify the light that is absorbed or reflected by human skin. In fact, the blood volume in our tissues impacts the amount of light that skin absorbs or reflects. PPG offers the possibility to estimate heart rate in an easy to use, inexpensive and less-invasive way than *Electrocardiograms* (ECG). ECG allows to obtain the heart rate, and further deduce the HRV, by measuring the electrical activity of the heart. Several studies focused on the analogy between PRV and HRV, leading to the result that both are linked to the ANS activity, and that the same features can be extracted from the two signals [5].

• Rita Meziati Sabour, Yannick Benezeth, and Fan Yang are with the ImViA Laboratory, University Bourgogne Franche-Comté, 21000 Dijon, France.
E-mail: rita_meziati-sabour@etu.u-bourgogne.fr, {Yannick.Benezeth, fanyang}@u-bourgogne.fr.

• Pierre De Oliveira and Julien Chappé are with the Psy-DREPI Laboratory, University Bourgogne Franche-Comté, 21000 Dijon, France.
E-mail: {Pierre.De-Oliveira, julien.chappe}@u-bourgogne.fr.

Manuscript received 11 May 2020; revised 8 Dec. 2020; accepted 18 Jan. 2021.

Date of publication 0 . 0000; date of current version 0 . 0000.

(Corresponding author: Rita Meziati Sabour.)

Recommended for acceptance by G. Chanel.

Digital Object Identifier no. 10.1109/TAFFC.2021.3056960

Recent researches have introduced *Remote Photoplethysmography* (RPPG), which offers the advantage of measuring the same parameters as PPG in a completely remote way. In fact, RPPG is the non-contact equivalent to the reflective mode of PPG using ambient light as a source and a camera as a receptor. The light reflected by the skin is then estimated by capturing subtle skin color variations by the camera as blood volume changes. Several image and signal processing steps allow to obtain a pulse signal, also called the RPPG signal. State-of-the-art RPPG extraction methods include *Blind Source Separation* (BSS), chrominance-based and deep learning-based approaches. *Independent Component Analysis* (ICA) and *Principal Component Analysis* (PCA) are well-known BSS techniques that have been widely used for remote PPG extraction [6], [7], [8]. Both methods aim at isolating RPPG signal from the input information captured by the camera, considered as a mixture of intensity, specular and pulse components. De Haan *et al.* define RPPG signal as a combination of two orthogonal chrominance signals that they build based on RGB information [9]. Complementary information about baseline RPPG algorithms can be found in [10] and [11]. RPPG estimation from facial video sequences using spatiotemporal deep neural networks has been newly proposed as in [12], [13] and [14]. From an RPPG signal, PRV can be computed following the same definition as for the BVP.

A few publicly shared datasets are adapted to psychophysiological studies, and, to the best of our knowledge, no publicly available datasets analyze social stress using remote physiological measures. In fact, there exist datasets that are intended for emotion analysis, such as CASME [15] and CASME² [16], but they target facial expression detection and recognition. Zhang *et al.* [17] introduce a multimodal dataset that involved 140 participants who were exposed to emotion eliciting videos. They use various imaging techniques, extract facial expression features and collect several physiological data, such as EDA and heart rate, in order to study the human emotional behavior. Other examples of datasets allow to obtain RPPG signals. This is the case, for instance, of the UBFC-RPPG dataset [18] which proposes 42 videos along with corresponding contact pulse signals measured from participants who played a time-sensitive mathematical game. VIPL-HR dataset [19] was created to train a deep network to estimate heart rate, based on a total number of 3130 visible light and near-infrared videos. Stricker *et al.* present the PURE dataset [20] that is constituted of 60 image sequences of 1 min and contact pulse signals, collected from 10 subjects who were asked to perform head motions while they were filmed. In [21], 204 uncompressed videos and ECG signals are gathered from 17 participants to test video-based HR estimation robustness to illumination and motion variations. However, these datasets are adapted for remote PPG computation but do not cover emotion or stress analysis. Some researchers that were interested in stress analysis through physiological responses had to create their own datasets, such as Kurniawan *et al.* who detect stress using speech and EDA, gathered from 10 participants [22]. The only studies that have used remotely acquired physiological signals to measure or recognize stress states use private and quite small databases. For example, Bousefsaf *et al.* worked on mental stress detection among 12 subjects using remote heart rate and EDA [23]. In [24], McDuff *et al.* detect cognitive stress based on contact and

non-contact physiological measurement (the heart rate, the breathing rate and the HRV), from 10 participants.

This article introduces a dataset, named UBFC-Phys,¹ that was collected to analyze the impact of social stress on physiological responses. Participants experienced stressful situations while they were filmed and were wearing a contact bracelet, in order to measure contact and non-contact physiological data. A Form based on the *Competitive State Anxiety Inventory* (CSAI) [25], allowed to quantify for every participant three dimensions of self-reported anxiety, namely cognitive anxiety, somatic anxiety and self-confidence. The experience followed a rigorous protocol inspired from the well-known *Trier Social Stress Test* (TSST) [26], and was conducted through three steps: a rest, a speech and an arithmetic tasks. The two latest tasks were organized following two levels of difficulty, as explained in Section 2. Physiological signals that were measured for the study are RPPG, BVP, remote and contact PRV and EDA. Facial expressions were also extracted from collected videos. This multimodal, large, and publicly shared dataset proposes a total number of 168 videos, contact BVP and EDA signals, as well as self-reported state anxiety scores, and can serve for studies related to affective computing and psychophysiology.

In this study, we show that physiological responses can be assessed during stress using the non-contact RPPG technique. Actually, non-contact PRV measurement can substitute contact reference measurement (correlations up to 99.83 percent) in similar experience conditions. Compared to existing datasets, this study proposes a further non-contact PRV analysis, through statistical and classification results. PRV extracted features were compared to EDA features and facial expressions. We obtained interesting results that show that non-contact PRV features can separate statistically the rest state from the stress state, which is supported by an 85.48 percent classification accuracy. In addition to this result, contact PRV and EDA narrowly surpassed remote PRV in recognizing the experience three steps (respective accuracies of 65.71 and 63.09 percent). Remote PRV performed best when it came to classifying the two levels of difficulty (with an accuracy of 69.73 percent). In all the classification tests, remote PRV gave better results than contact PRV.

The rest of the article is organized as follows: data collection, experiment protocol and dataset organization are explained in Section 2. Physiological signal processing as well as PRV, EDA and facial feature extraction are detailed in Section 3. In Section 4 the obtained results are presented: correlations between non-contact and contact PRV signals and features are computed in Section 4.1, stress state and experiment task recognition results using PRV, EDA and facial expression features are compared in Section 4.2, while stress level recognition based on physiological data and self-reported anxiety scores are explored and discussed in Section 4.3. A conclusion, accompanied by future work ideas, is given in Section 5.

2 DATASET PRESENTATION

2.1 Data Collection

The experience took place in a laboratory room. 68 undergraduate psychology students participated in the experience.

1. <https://sites.google.com/view/ybenezeth/ubfc-phys>

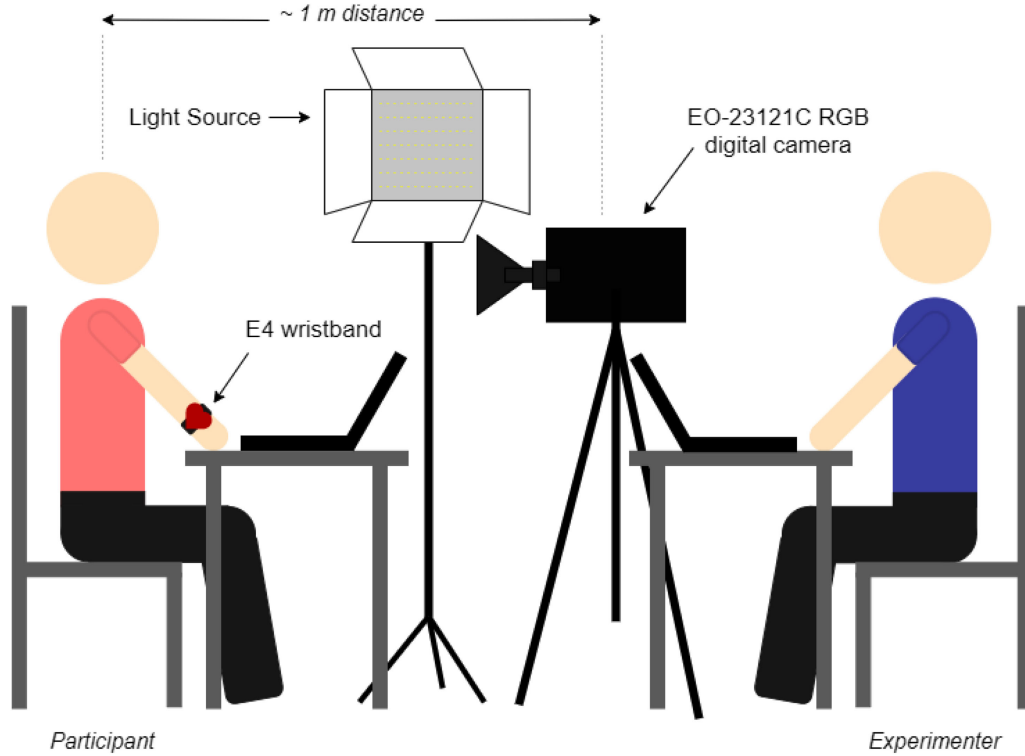


Fig. 1. Schema of the experience setting up.

Participants were filmed during the experience with an EO-23121C RGB digital camera by *Edmund Optics*, with a Motion JPEG compression and a 35 frame per second rate. The frame resolution was of 1024×1024 pixels. An artificial light source was used to ensure uniform lighting conditions for all the participants. Participants were seated around 1m away from the camera and the light source. In front of each participant an experimenter was seated. The experimenter used a laptop to start and stop video recordings. Another laptop was needed to simulate a Skype call as explained in Section 2.2. Having one experimenter (three experimenters run the sessions, all were students aged between 20 and 25) constitutes a novelty in comparison with the original TSST protocol. This choice is motivated by our concern for experimental condition stabilization.

Contact measurements were realized using the Empatica E4 wristband,² which records BVP, skin temperature and EDA responses. The E4 bracelet has also an accelerometer and computes the *Inter-Beat Intervals* (IBI) from the BVP signal, which constitute the PRV. E4 wristband performances have been validated in several studies [27], [28], and relate to various research areas such as sleep monitoring [29], driving safety [30], and emotion arousal assessment [31]. Accelerometer data and the IBI given by the E4 wristband were not used in this study. In fact, the IBI provided by the E4 bracelet are strongly filtered using a proprietary algorithm, which limits the reproducibility of this research. In this study, PRV extraction and filtering were realized based on standard algorithms as explained in Section 3.2, making it possible to compare PRV obtained from contact and remote pulse signals.

Self-reported data were collected using forms that were given to the participants at the beginning and at the end of the experience. These forms were built following the CSAI, and aimed at mirroring their state anxiety through three dimensions: cognitive anxiety, somatic anxiety and self-confidence. Forms contained 7 items for each of cognitive and somatic anxiety, which respectively designate mental and physical expressions of anxiety. Additional 9 self-confidence items captured how confident participants felt before and after the experience session. Each item was partitioned into four possible responses, presented as numbers from 1 to 4. These numbers were intended to indicate a level of the experienced anxiety item, 1 meaning *not at all* and 4 *extremely*. For cognitive and somatic anxiety, low scores mean the participant estimates they deal pretty well with the item concerned and feels decreased anxiety, and high scores indicate the participant proves difficulty handling the anxiety aspect expressed by the item. This does not apply for self-confidence, as score values are proportional to participant confidence degree in succeeding the experience tasks. Fig. 1 shows the experiment setting-up.

2.2 Experience Process

Before data were collected, the objective and process of the experience were presented to every participant. It was explained that the purpose of this experience was to study the effect of social stress on our physiological responses. The use of the camera and the wristband was justified to each subject. A consent form was given to the participants, allowing them to choose whether to share their data with the scientific research community. The consent form was signed by both the experimenter and the participant.

2. <https://www.empatica.com/research/e4/>



Fig. 2. Examples of frames extracted from videos recorded during the rest task $T1$ (first column), the speech task $T2$ (second column) and the arithmetic task $T3$ (third column).

The experience consisted of three tasks: a rest task, a speech task and an arithmetic task. During the rest task, participants were asked to stay quiet and not to talk. This phase constituted a baseline for physiological responses, and allowed to remove the effects of prior stimuli, such as coming late to the experience session. This was an important step in order to facilitate comparisons between participants' responses. The rest task lasted 10 minutes.

The speech and arithmetic tasks were interactive and had two possible scenarios, conceived so that two stress levels can be treated: a hard scenario (that we called *test*) and an easier one (named *ctrl* in our dataset). Participants were randomly assigned to one of the two versions. In the *test* scenario, the speech task was a simulation of a job interview. Participants had to imagine their dream job and convince the experimenter to hire them. In the *ctrl* scenario, the subjects had to either recall a positive holiday memory or imagine a dream vacation, and persuade the experimenter to wish to have the same holidays. In both scenarios, participants had the possibility to prepare a draft for their speech before the task started. In the *test* version, the experimenter picked up the speech draft, and a fake Skype call video played on the participant laptop (see Fig. 1). Participants were told that the Skype call permitted an additional jury member who would not intervene to watch their performance. Introducing this artificial intervener, presented as a non-verbal communication expert, constitutes another

novelty compared to the original TSST protocol, and supports our concern in holding the experimental conditions as constant as possible across all the sessions. The speech task lasted 6 minutes, subjects were free to choose the duration of their speech. Almost all of them did not exceed 3-minute speeches; hence the experimenter asked them extra questions to fulfill the task devoted time.

In the arithmetic task, participants were asked to perform a countdown starting from 2025 in steps of 10 in the *ctrl* scenario, while it had to start at 2023 in the *test* version and respect steps of 17. Subjects had to pronounce the countdown numbers out loud, and were stopped whenever they gave the wrong number. When this happened, they had to start over the countdown. The arithmetic task lasted 4 minutes. For the rest of this article, we denote the rest task $T1$, the speech task $T2$ and the arithmetic task $T3$. Fig. 2 gives examples of frames belonging to videos acquired during the three experience tasks. The three subjects show typical attitudes for each task: gaze directed outside of the camera lens in $T1$ and $T3$ while subjects more often look at the experimenter in the eye in $T2$, limited motion in $T1$, postures that describe a reflection effort in $T3$.

The experimenter announced the beginning and ending of each task, and participants were asked to click on the E4 wristband button wrist faced to the camera. This allowed to use the bracelet time marker function and facilitate the synchronization between the wristband and the video

recordings. The different tasks were explained to participants before each task started. The experimenter launched the video recording before the rest task started and stopped it once the arithmetic task finished.

At the end of the experience, subjects were free to ask further questions and discuss with the experimenter about their opinion, comments and feelings regarding the experiment protocol.

As stated earlier, our experiment is inspired by the *Trier Social Stress Test* and aims at assessing social stress by comparing measured physiological data. The stress induced by the TSST was qualified as *social* because the protocol combines key elements of social evaluative threat and uncontrollability to produce physiological and psychological stress responses in humans [32], [33]. Within the original protocol [26], [34], a high level of social-evaluative threat was induced by a public speaking in front of an unresponsive audience and completing a surprise mental arithmetic test. In the present study, we used a slightly modified version, the participants performed a stressful public speaking (a self-presentation and an unexpected arithmetic task) in front of an audience composed of one experimenter. In the *test* scenario, the audience included another experimenter presented as an expert in behavioral analysis by video-conference during the speech task, with the purpose to enhance the social-evaluative dimension of the protocol. As in the original TSST version, participants were evaluated by the *expert* without any signs of support and were informed that their performance will be recorded in audio/video.

2.3 Dataset Organization

The final UBFC-Phys dataset proposes data collected from 56 healthy subjects (12 participants were eliminated due to technical problems or data sharing refusal). Participants are all aged between 19 and 38 (mean age is 21.8 and standard deviation is 3.11). Among these participants 46 are female and 10 male. Data is organized into 56 folders, corresponding to every participant. Thus, for each subject, three videos are available, one video per task. For concerns of equalizing the duration of the three tasks and alleviating the dataset size, only three minutes were kept. For data related to the rest task, the 3 minutes started from the middle, while they started at the beginning for the speech and the arithmetic tasks. Alongside the videos, contact blood volume pulse and electrodermal activity signals obtained from the E4 wristband are given. There are three BVP and three EDA signal *.csv* files for each participant, respective to the three experience tasks. A *.txt* file contains information relative to the experiment, such as the subject associated number (from 1 to 56), his sex, the date and time the video recording began. In the same file the experience scenario (*test* or *ctrl*) is indicated. Pre and post-session self-reported state anxiety scores are given in a *.csv* file.

Fig. 3 summarizes the conduct of the experience sessions. It shows selected timeframes (indicated by a red broken line) that define available data extraction. Pink windows refer to introduction, transition and ending times (respectively denoted as 1, 2 and 3 in Fig. 3). During introduction time, which lasted approximately 5min, the experimenter first explained the objective of the experience and the equipment used to collect data, then gave consent and pre-session forms, before presenting the rest task *T1*. Transition times

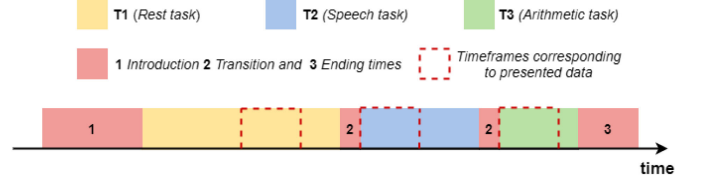


Fig. 3. Global timeline of the experience sessions.

(2 in Fig. 3) designate periods where the experimenter gave instructions of tasks *T2* and *T3*, both lasted around 1min. Ending time (3 in Fig. 3) lasted nearly 3min, and allowed participants to ask further questions about the experience after post-session forms were filled.

3 DATA PROCESSING

In this section, we explain how collected data were processed before validation tests were applied. In Section 3.1, we present the steps that allowed RPPG signal extraction from recorded videos. Next, PRV estimation and filtering, as well as PRV-based feature extraction, are detailed in Section 3.2. In Sections 3.3 and 3.4 we correspondingly cite EDA and facial expression features estimated for this study.

3.1 RPPG

To extract RPPG signals, the face is detected on the input video frames, then skin pixels are selected since they contain blood volume information. Next, the RGB values over the skin pixel area are spatially averaged. The concatenation of successive averaged RGB yields RGB temporal traces that are thereafter processed before an RPPG estimation algorithm is applied.

We used the face detection deep learning-based OpenCV model, which relies on the Single Shot MultiBox Detector method proposed by Liu *et al.* [35]. Skin pixels were selected following the Conaire *et al.* algorithm [36]. RGB traces pre-processing consisted of two steps: a detrending step that was realized by dividing samples by their mean over a 1s temporal interval, followed by a band-pass filtering step which was obtained using a 0.7 and 3.5 Hz cut-off frequency Butterworth filter. For the following, let us denote R_n , G_n and B_n the detrended and filtered RGB traces, and $(RGB)_n$ the vector space comprised of R_n , G_n and B_n . Among the existing RPPG extraction methods, we used the *Plane-Orthogonal-to-Skin* (POS) algorithm, introduced by Wang *et al.* in [11], mainly because of its efficiency and execution speed. Wang *et al.* remove intensity variations induced by motion, which equally impact the three RGB channels. To do so, they consider a plane P that is orthogonal to the unit vector $u = (1, 1, 1)^T$ in the $(RGB)_n$ space. Temporally normalized RGB are projected onto the plane P , leading to two signals Y_1 and Y_2 that are linear combinations of R_n , G_n and B_n . Y_1 and Y_2 are defined as

$$\begin{cases} Y_1 = G_n - B_n \\ Y_2 = G_n + B_n - 2R_n. \end{cases} \quad (1)$$

The pulse signal Y is obtained as

$$Y = Y_1 + \alpha Y_2. \quad (2)$$

The α factor is set as de Haan *et al.* propose in [9].

3.2 PRV

After resampling RPPG (initially sampled at the frame per rate, i.e., 35) and BVP (initially sampled at 64 Hz) signals to 128 Hz using shape-preserving piecewise cubic interpolation, remote PRV and contact PRV were extracted. From both contact and remote pulse signals, peaks were detected. Then, *Pulse-to-Pulse* (PP) intervals, which are time differences between successive peaks expressed in *seconds* (s), were computed to constitute PRV signals. Two steps have been followed in PRV signal processing: signal filtering (explained in Section 3.2.1) and feature extraction (detailed in Section 3.2.2).

3.2.1 PRV Filtering

We were inspired by the Kubios HRV software threshold-based artefact correction algorithm [37] for PRV signal filtering. PRV signals were median filtered using a 51 sample wide median filter (let us denote this width as L , thus $L = 51$), yielding a new signal $medPP$. Next, PRV samples $PP(i)$ with $i \in [1, N]$, N being PRV signal length, were compared to the corresponding local median value $medPP(i)$. If

$$|PP(i) - medPP(i)| > \tau, \quad (3)$$

i.e., the difference exceeded a threshold value τ , PRV samples were replaced by the respective median value. We applied a 0.15 s threshold, which corresponds to a strong filtering in [37].

To minimize the influence of zero-padding introduced by median filtering, we included a mirror flip at PRV signal boundaries. Two symmetries were realized: a symmetry of the first $(\frac{L-1}{2} + 1)$ PP values with respect to the first sample $PP(1)$, and a symmetry of the last $(\frac{L-1}{2} + 1)$ PP values with respect to the last sample $PP(N)$. In other words, for

$$\begin{cases} i \in [2; \frac{L-1}{2} + 1] \\ i \in [N - \frac{L-1}{2}; N - 1], \end{cases}$$

we flipped $PP(i)$ samples to obtain two vectors that were respectively concatenated at the beginning and at the end of PRV signals. Therefore, median filtering generated an $M = (N + L - 1)$ long pulse signal PP_f . Only $PP_f(j)$ for $j \in [\frac{L-1}{2} + 1; M - \frac{L-1}{2}]$ were kept before applying threshold-based filtering.

3.2.2 PRV Feature Extraction

Several features were extracted from PRV signals, all reflecting the PNS and SNS activity. Temporal features that were computed are: pulse-to-pulse mean value \overline{PP} , heart rate mean value \overline{HR} , PP standard deviation value $SDPP$ and root mean square of successive differences $RMSSD$. \overline{HR} , $SDPP$ and $RMSSD$ were computed as follows:

$$\overline{HR} = \frac{60}{\overline{PP}} \quad (4)$$

$$SDPP = \sqrt{\frac{1}{N} \sum_{i=1}^N (PP(i) - \overline{PP})^2} \quad (5)$$

$$RMSSD = \sqrt{\frac{1}{N-1} \sum_{i=2}^N (PP(i) - PP(i-1))^2}. \quad (6)$$

Building PP histogram allowed to compute the Baevski stress index SI as presented in [38]. Frequency domain-related features were extracted from the PP series Lomb-Scargle power spectral density. Low and high frequency (LF and HF) PRV components were obtained by summing power spectral density over $[0.04; 0.15]$ Hz and $[0.15; 0.4]$ Hz intervals respectively. LF is assumed to reflect both the SNS and PNS activity, while HF is related to PNS reactions. The ratio $LFHF$, given by LF over HF , is supposed to describe interactions between the SNS and the PNS, also called the sympathovagal balance. Geometric features $SD1$, $SD2$ and S resulted from PRV two-dimensional Poincaré plot, which was obtained by plotting each PP sample $PP(i)$ as a function of the preceding PP value $PP(i-1)$. Poincaré plot can be fitted into an ellipse [39] of minor axis $SD1$ and of major axis $SD2$. The ellipse surface S was computed as

$$S = \pi \cdot SD1 \cdot SD2. \quad (7)$$

In total, 11 features were extracted from contact and non-contact PRV signals.

3.3 EDA

Skin electrodermal activity signal reflects the influence of the SNS during stress. It is characterized by two components: tonic *Skin Conductance Level* (SCL), and phasic *Skin Conductance Response* (SCR). The first type is the smooth baseline level, and the latter represents rapid reactions to an external stimulus. EDA is measured in *microSiemens* (μS) and sampled at 4 Hz.

Multiple EDA signals corresponding to the rest task presented no phasic SCR responses, which led us to only calculate tonic SCL. Tonic skin conductance level was computed using *Continuous Decomposition Analysis*, introduced in [40], using the Ledalab Matlab toolbox. Therefore, EDA signal mean \overline{eda} and standard deviation $stdEda$ values were computed, as well as SCL mean \overline{scl} and standard deviation $stdScl$. Besides, we extracted the SCL minimum and maximum values (respectively $minScl$ and $maxScl$), and subtracted the 10 first sample mean value from the 10 last sample mean value (we denote this difference as $diffScl$) to characterize the SCL variation during the experience tasks. In sum, 7 features were estimated from EDA signals. Fig. 4 shows subject 1's electrodermal activity signal, recorded during the second experiment task, as well as extracted tonic conductance level.

Supplemental Material I, which can be found on the Computer Society Digital Library at <http://doi.ieeecomputersociety.org/10.1109/TAFFC.2021.3056960>, gives standard statistical data of contact and remote PRV as well as EDA features cited in Sections 3.2 and 3.3, extracted from PRV and EDA signals of all 56 subjects during the three experience tasks.

3.4 Facial Expressions

From videos recorded during experience sessions, we extracted facial expression features using the OpenFace

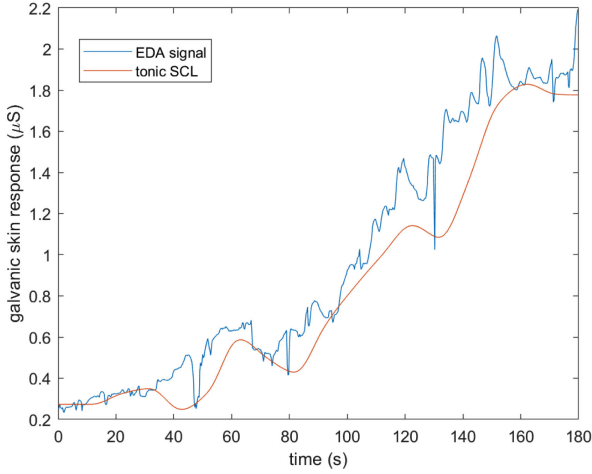


Fig. 4. EDA signal and extracted SCL for subject 1 during the speech task.

FeatureExtraction estimator.³ For each frame, eye gaze direction vectors, location of the head with respect to camera (and euclidean norm of location coordinates), head rotation angles around X, Y and Z axes were retrieved, as well as 17 *Action Unit* (AU) intensity (from 0 to 5) and presence (0 if the AU is absent and 1 otherwise) in the frame (AUs extracted are 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, and 45). Action units are defined in the *Facial Action Coding System* (FACS), first introduced by Hjortsjö [41] and developed by Ekman and Friesen [42]. FACS includes facial muscle movements related to emotions humans can feel. Next, mean and standard values of all the features extracted along input video frames were computed, except for AU intensity and presence. We only computed AU intensity mean value, and a presence percentage was calculated by normalizing the number of frames with a 1 presence score over the total number of video frames. 52 final features were estimated from video recordings to describe the participants' facial expressions.

4 EXPERIMENTATION VALIDATION AND RESULTS

In this section, we validate the proposed experimentation protocol through several demonstrations. First, we prove in Section 4.1 that remote PRV can substitute contact PRV features in similar experimental conditions. We estimated continuous heart rate signals from remote and contact PRV, and obtained satisfying correlations reaching 99.83 percent. In Section 4.2, we aim at detecting stress using statistical ANOVA test and machine learning. ANOVA shows that several contact and remote PRV features, as well as all EDA features, succeed in recognizing the stress state. In addition to that, remote PRV features surpass the rest of physiological modalities in classifying rest and stress states, with a 85.48 percent accuracy. A task recognition classification was also applied, and whereas EDA performs best with an accuracy of 75.71 percent, remote PRV outpaces contact PRV. ANOVA and machine learning classification were also used to recognize the two stress levels proposed in the experience, as presented in

Section 4.3. While physiological features fail in separating the levels using ANOVA, self-reported somatic anxiety allows stress level recognition. Besides, remote PRV gives the best accuracy in terms of stress level classification (69.73 percent accuracy).

4.1 Remote and Contact PRV Comparison

To compare remote and contact PRV signals, we built heart rate signals for each pair of remote/contact PRV, by computing the mean HR over PRV sliding windows of 30 seconds. We chose HR since it is a robust and simple-to-calculate feature. For each task, we computed *Pearson Correlation Coefficient* (PCC) between contact and remote HR signal estimations.

PCC estimates the linear correlation between two variables, and can range from -1 to 1 . Obtained PCC values for the three tasks were respectively 0.83, 0.42 and 0.61. These values reflect the level of noise that impacted both contact and remote pulse signal measurements. Highest value obtained for $T1$ can be justified by the fact that in the rest task subjects were still, whereas during the speech and the arithmetic tasks they acted spontaneously, moving their hands and/or face. This caused noise in BVP and RPPG signals, leading remote and contact PRV traces to be noisy, mainly for the second and third tasks.

Since the camera and the wristband have different sensitivities to noise, it is difficult to determine when a single pulse signal (BVP or RPPG) is reliable. For this reason, we worked with the hypothesis that BVP and RPPG signals are reliable if they highly correlate, and have close values. To validate the experience and obtain reliable results, we eliminated contact and remote PRV traces that failed in respecting this hypothesis. To do so, We selected pairs of remote/contact PRV based on two criteria: *Pearson Correlation Coefficient*, of HR signals and *Mean Absolute Error* (MAE) of remote and contact HR. MAE allows to compare two variables that have the same scale by evaluating their difference. A PCC value of 1 combined with a null MAE signifies that two variables are the same. Two signals may have a high PCC value, implying they correlate, yet their values can be very different. An acceptable PCC value, combined with a controlled MAE, would allow to select signals close in terms of correlation and values. For this reason, for HR signals of each task T_i with $i \in \{1, 2, 3\}$, we tolerated PCCs of at least 0.40 (40 percent), and MAE values lower than the mean MAE value of all signals of that task $\overline{MAE_{T_i}}$. This can be expressed by the two following conditions:

$$\begin{cases} PCC_{HR}(HR_R, HR_C) > 0.40 \\ MAE_{HR}(HR_R, HR_C) < \lceil \overline{MAE_{T_i}} \rceil \end{cases}, \quad (8)$$

where HR_R and HR_C are respectively remote and contact HR signals, and $\lceil \cdot \rceil$ is the ceiling function. PCC_{HR} is expressed in % and MAE_{HR} in *beats per minute* (bpm).

Hence, respecting these conditions led 14 subjects to be eliminated from $T1$, 33 subjects from $T2$ and 28 subjects from $T3$. A visual inspection permitted to retrieve participants with signals that did not respect one of the conditions given in Equation (8) and showed visual similarity in HR traces. No subject was added for $T1$, 3 subjects were

3. <https://github.com/TadasBaltrusaitis/OpenFace/>

TABLE 1
Remote/Contact HR Signal Pearson Correlation
Coefficient and Mean Absolute Error Mean, Minimum
and Maximum Values Per Task

	PCC _{HR} (%)		MAE _{HR} (bpm)	
	before select.	after select.	before select.	after select.
T1				
mean value	67.64	87.00	3.55	0.95
min value	-47.45	40.67	0.10	0.10
max value	99.83	99.83	33.95	2.15
T2				
mean value	32.00	74.14	9.26	3.96
min value	-75.66	29.67	0.23	0.23
max value	98.26	98.26	27.80	8.04
T3				
mean value	36.35	66.13	5.99	2.36
min value	-77.00	17.48	0.51	0.51
max value	97.31	97.31	34.81	4.93

retrieved for $T2$ and 5 for $T3$. We obtained a total of 101 signals per modality (contact or remote) for the three tasks, distributed as follows: 42 signals for $T1$, 26 for $T2$ and 33 for $T3$. Supplemental Material II, available online, gives the list and number of subjects eliminated in each task. Statistical data of physiological features extracted from selected subject PRV and EDA signals are also given in Supplemental Material I, available online. Selected HR signals showed a PCC mean value of 87 percent for $T1$, 74 percent for $T2$ and 66 percent for $T3$, while values before selection were of 68, 32 and 36 percent correspondingly for the three tasks. MAE had average values of 3.55 bpm for $T1$, 9.26 bpm for $T2$ and 5.99 bpm for $T3$. Subject selection allowed to reach values of 0.95 bpm, 3.96 bpm and 2.36 bpm for the respective tasks. Exact values are detailed in Table 1. Fig. 5 shows remote and contact heart rate (HR_R and HR_C) plots corresponding to maximum Pearson correlation values obtained for each task. Maximum PCC was obtained with subject 17's HR signals for $T1$, subject 18 for $T2$ and subject 12 for $T3$.

Noise may alter the accuracy of heart rate measurements based on PRV signals. In fact, it has been shown that very little heart period artifacts can lead to errors of heart rate variability features that are larger than the typical effect size in psychological studies [43]. Therefore, for each task, we validated PRV features by computing correlations between contact and remote features. Pearson correlation coefficient was computed for each couple of vectors that contained all values of a given feature f_i with $i \in [1; 11]$. We obtained PCC mean values of 0.85 for $T1$, 0.66 for $T2$ and 0.73 for $T3$. Best PCC obtained for $T1$, followed by $T3$ and $T2$, shows once more the influence of noise on PRV features, since in $T1$ movements were very limited while in $T2$ and $T3$ participants moved freely. Highest PCC values for the three tasks were obtained by \overline{PP} (same results were given by \overline{HR} , since they are inversely proportional). $LFHF$ gave lowest PCC values. Figs. 6 and 7 show correlation plots of \overline{PP} (denoted as $meanPRV$) and $LFHF$.

4.2 Stress State Recognition

After valid data selection, we sought to investigate the relationship between physiological data and stress induction. Within this context, we analyzed whether stress can be detected.

To detect a stress state, we kept only participants with signals in $T1$ and at least in one other task among signals selected as explained in Section 4.1. This allowed to define a non-stress state (represented by $T1$), and a stress state estimated based on $T2$ and $T3$. For subjects with either valid $T2$ or $T3$ signals, corresponding PRV and EDA features were considered as defining the stress state. Stress state was expressed as the average $T2$ and $T3$ PRV and EDA feature values for subjects who showed valid signals for both tasks. The non-stress state was defined by $T1$ PRV and EDA feature values.

To differentiate between non-stress and stress states, two approaches were considered: *Analysis of Variance* (ANOVA) and machine learning, both tested on PRV and EDA features. ANOVA analysis is a statistical tool that determines whether a variable has different behaviors in changing levels of a given factor. This is realized by comparing variations between group data to variations within these groups. A probability p -value and a score F are defined to decide on

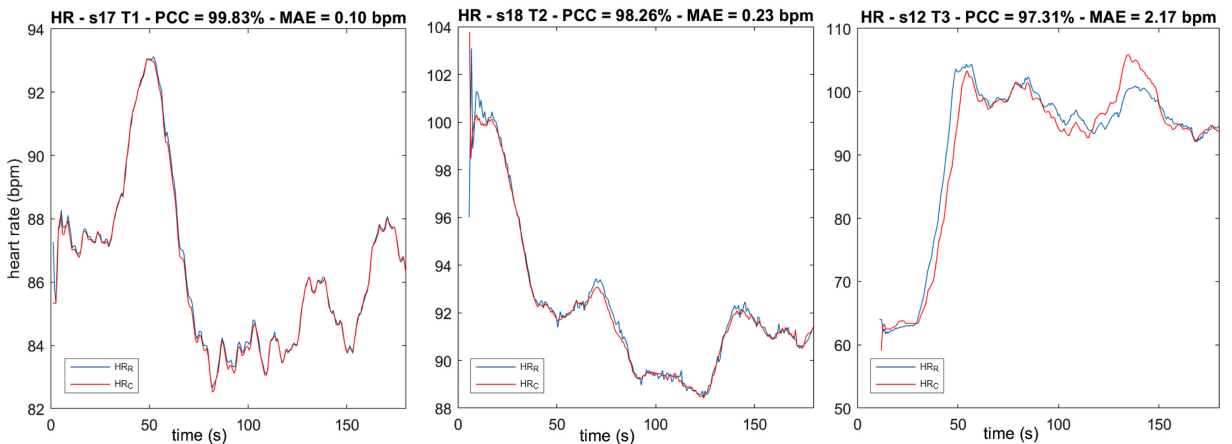


Fig. 5. Remote and contact heart rate (HR_R and HR_C) plots with maximum PCC values in $T1$, $T2$, and $T3$.

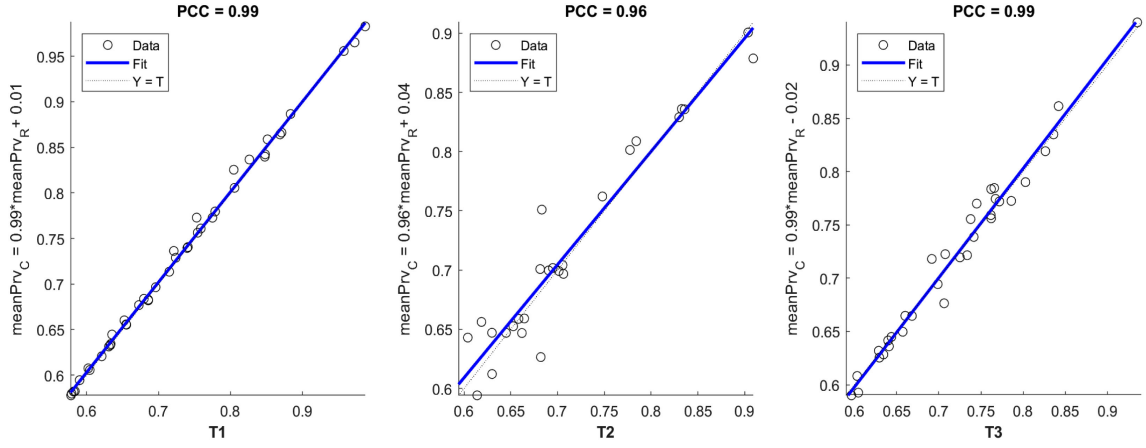


Fig. 6. meanPRV Correlation Plots for tasks $T1$, $T2$, and $T3$, showing very high correlation between contact and remote average HR estimations.

the significance of this variability. In our study, the ANOVA test was applied to non-stress and stress groups for each feature separately, and the p -value was compared to a significance level of 0.05. Hence, for p -values lower to the significance level, it was concluded that the group means are different, and the groups separable.

PP , HR , and $LFHF$ failed to separate non-stress and stress states for both modalities (remote and contact), as well as remote LF ($p = 0.06$) and contact HF ($p = 0.25$). Features yielding p -values between 0.05 and 0.1 were considered as close to be significant in terms of group separation, which is the case for remote LF . Other remote and contact PRV features succeeded in differentiating between the two groups. Fig. 8 shows ANOVA test results using PRV features S and HR . HR fails in separating non-stress and stress ($p = 0.80$ for contact HR and $p = 0.70$ for remote HR , denoted as $meanHR_C$ and $meanHR_R$ in Fig. 8). It can be observed in Fig. 8 that HR group means are close for both modalities: non-stress mean values are 83.31 bpm and 83.68 bpm for contact and remote HR respectively, while stress mean-values are 83.95 bpm for contact HR and 84.63 bpm for remote HR .

ANOVA tests on all EDA features accomplished non-stress and stress group separation. Test result on EDA signal mean value $meanEda$ and standard deviation $stdEda$ are given by Fig. 9. For both features, group means considerably increase from non-stress to stress states ($meanEda = 0.48 \mu S$

and $stdEda = 0.06 \mu S$ for non-stress group *vs* $meanEda = 1.55 \mu S$ and $stdEda = 0.24 \mu S$ for stress group).

In the machine learning approach, two classifications have been realised: non-stress *vs* stress states, and $T1$ *vs* $T2$ *vs* $T3$. Both tests were applied based on contact PRV features (PRV_C), remote PRV features (PRV_R) and EDA features. Results obtained with contact PRV alongside with EDA are also presented, in order to highlight the comparison between contact and non-contact modalities. It is important to note here that features related to facial expressions do not make sense for $T1$ and are not used in this experiment. Contact PRV and EDA feature combination was obtained by reducing their concatenation dimensionality to a chosen number of features. Dimensionality reduction can usually be achieved by selecting a number k best scoring features following a given metric, which we chose to be ANOVA F -score, since it expresses whether a variable can distinguish between different groups. Hence, ANOVA analysis was performed for each facial expression feature using classification labels (*Non-stress / stress* in stress state classification, and $T1 / T2 / T3$ in task recognition) to define the groups. Features with the highest resulting F -scores (i.e., the greatest capacity to separate the groups) were kept (we fixed $k = 10$).

Four classifiers were considered: *Support Vector Machine* (SVM) with a linear kernel, SVM with *Radial Basis Function* (RBF), *Logistic Regression* (Log Reg) and *K-Nearest*

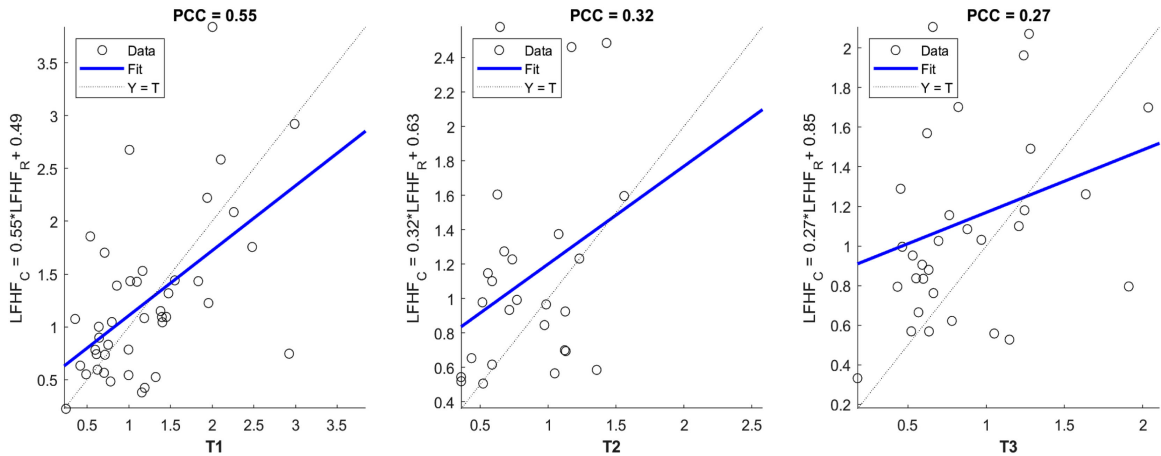


Fig. 7. LFHF Correlation Plots for tasks $T1$, $T2$, and $T3$. LFHF is the feature that gives the lowest PCC in this study.

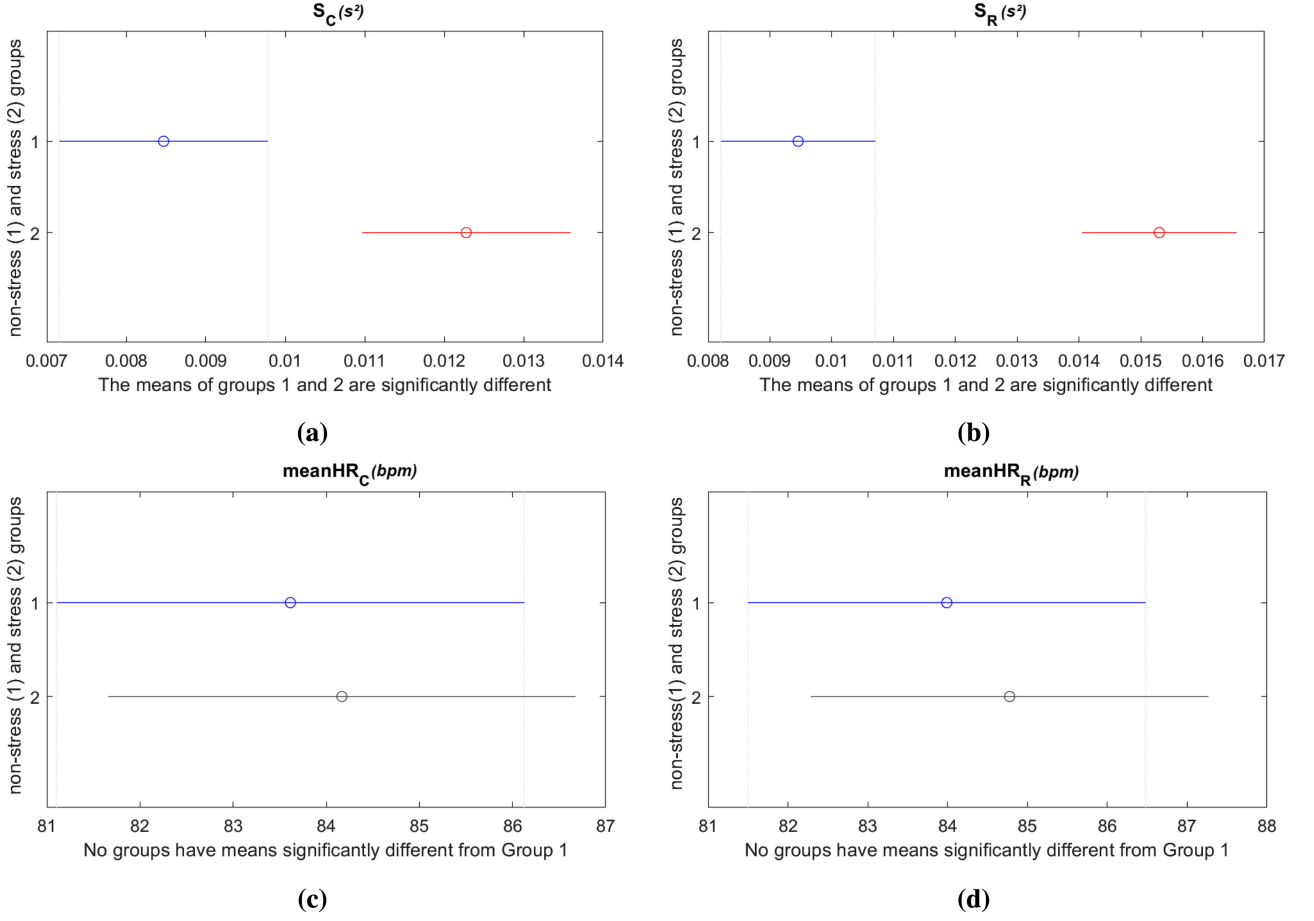


Fig. 8. Non-stress and stress group comparison with ANOVA for: (a) contact S , (b) remote S , (c) contact \overline{HR} and (d) remote \overline{HR} . Non-stress and stress data groups are represented in blue and red colors when ANOVA test succeeds in separating them. A data group is coloured in gray when considered as similar to the other group by ANOVA test. This colour interpretation applies for all ANOVA test figures presented in the article.

Neighbors (KNN). Classifier models were validated using stratified KFold (with a number of folds equal to 7) cross-validation. This way, the entire data is divided into 7 subsets, each subset serves iteratively as a test set and the rest constitutes the training set, and a classification accuracy score is computed for each fold. Constituted folds are stratified, meaning they contain the same percentage of samples for each label. Accuracy represents the percentage of correct class predictions over the total number of predictions. Mean accuracy over the 7 folds was

calculated and retained as a classification result for all classification tests presented in this article.

Table 2 gives non-stress *vs* stress state classification results using features previously cited. Maximum accuracy is written in bold for each feature modality. Absolute maximum value of 85.48 percent (written in red) is achieved by remote PRV (PRV_R) features, followed by EDA features (82.38 percent), and contact PRV + EDA combination (79.52 percent). Contact PRV (PRV_C) features give the lowest accuracy (75.00 percent).

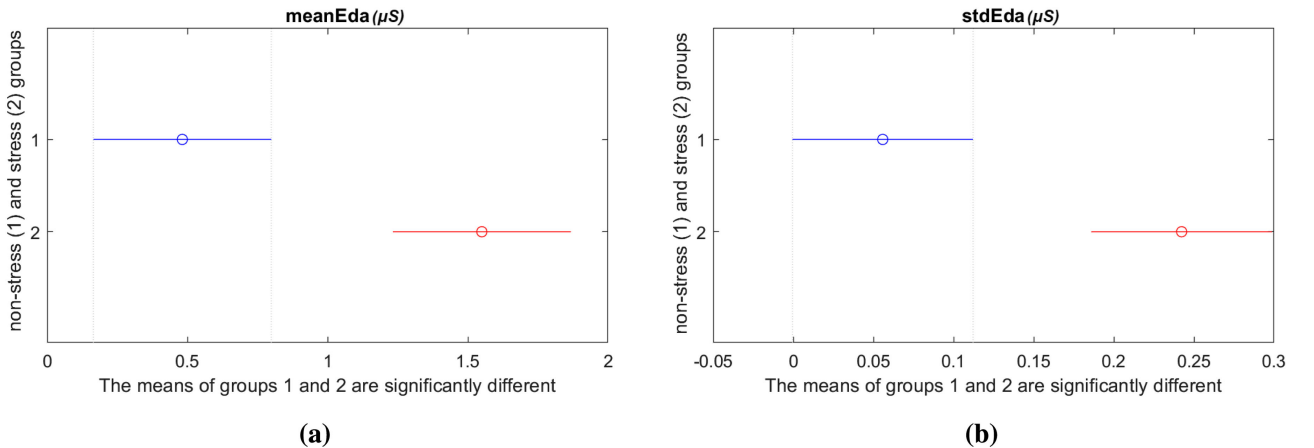


Fig. 9. Non-stress and stress group comparison with ANOVA for (a) meanEda and (b) stdEda .

TABLE 2
Non-Stress Versus Stress State Classification Results
Based on Contact and Remote PRV Features
(respectively denoted as PRV_C and PRV_R), EDA Features
and Combination of Contact Features

Features	Classifiers	Accuracy (%)
PRV_C	SVM - linear kernel	75.00
	SVM - RBF kernel	71.90
	Log Reg	73.57
	KNN	70.71
PRV_R	SVM - linear kernel	84.29
	SVM - RBF kernel	85.48
	Log Reg	83.81
	KNN	78.81
EDA	SVM - linear kernel	79.29
	SVM - RBF kernel	82.38
	Log Reg	77.86
	KNN	81.66
$PRV_C + EDA$	SVM - linear kernel	79.52
	SVM - RBF kernel	74.29
	Log Reg	78.33
	KNN	72.86

Table 3 shows task classification results, where EDA features reach a 75.71 percent accuracy. Remote PRV and combined contact features give close accuracy values (respectively 63.09 and 65.71 percent). PRV_C features show the lowest accuracy (51.43 percent).

After we showed in Section 4.1 that the correlation between contact and non-contact PRV features was high, we show here that the results obtained with PRV features estimated from the video signals were even higher than those obtained with PRV features measured from the contact wristband device. We observed in Section 4.1 that the BVP signals from the bracelet were at least as noisy as the RPPG signals. Actually, it is well known that the PPG technology is very sensitive to motion, and this severely limits its exploitation in psychophysiological studies such as ours, where participants' movements are not restricted. Based on stress state and task recognition results given by Tables 2 and 3, EDA modality seems more resilient to disturbances since EDA features give better results than BVP features. One hypothesis for this gap may be the nature of sensors that measure EDA and BVP, as electrodes are used for the former, while the latter uses an optical device. Besides, video-based PRV modality has proven to be reliable in recognizing the stress state generated following the experience. It is also important to note that a duration of 5 minutes is usually recommended to calculate PRV features [44], making the 3 minutes considered in our study for each task seem relatively short. However, emergent works propose to use significantly shorter durations, as in [45] and [46]. The 3min duration can explain why, for example, *LFHF* does not allow to differentiate the states of stress and non-stress. Besides, we can mention that the correlation between remote and contact *LFHF* was very low, which suggests that this feature was not stable enough and therefore too noisy to discriminate the stress states.

TABLE 3
Task Recognition Results on Contact and Remote PRV Features (Respectively Denoted as PRV_C and PRV_R), EDA Features and Combination of Contact Features

Features	Classifiers	Accuracy (%)
PRV_C	SVM - linear kernel	51.43
	SVM - RBF kernel	50.48
	Log Reg	51.43
	KNN	49.29
PRV_R	SVM - linear kernel	62.62
	SVM - RBF kernel	61.90
	Log Reg	63.09
	KNN	57.62
EDA	SVM - linear kernel	75.71
	SVM - RBF kernel	64.52
	Log Reg	71.19
	KNN	60.24
$PRV_C + EDA$	SVM - linear kernel	65.71
	SVM - RBF kernel	60.95
	Log Reg	64.29
	KNN	55.95

4.3 Stress Level Recognition

In Section 4.2, we proved that physiological measures, particularly remote PRV, allowed detecting stress. In this subsection, we assess collected data's capacity to recognize stress levels. ANOVA and machine learning algorithms were applied to determine whether the different features succeeded in separating the experience stress levels, linked to *test* and *ctrl* scenarios. ANOVA tests were applied to measured physiological features and self-reported data. Average item scores were computed based on participants' responses for each state anxiety dimension set of items. This led to a total number of 2 scores per dimension (cognitive anxiety, somatic anxiety and self-confidence), corresponding to the times participants filled the form (i.e., before the experience session started and after it finished). Post-session scores were normalized using pre-session scores for the three dimensions. Physiological features were normalized by non-stress state (i.e., $T1$) values. All contact and remote PRV features, as well as EDA features, failed in differentiating the two groups (i.e., all p -values were superior to 0.05). Regarding PRV, we give as an example ANOVA test results for contact and remote *SDPP* in Fig. 10. Group separation results based on ANOVA are shown in Fig. 10 for tonic SCL mean and maximum values (respectively named meanSCL and maxSCL). It can be observed that the difference between the groups is slightly more visible when using the EDA features, even if still not statistically significant.

Before ANOVA test was applied to self-reported data in order to determine whether it manages to separate *test* and *ctrl* groups, the reliability of anxiety levels indicated by participants had to be verified. This was achieved by computing *Cronbach's alpha* α , also called *tau-equivalent reliability*. Coefficient α estimates the consistency of a psychometric test. Introduced in [47], it expresses the correlation of items that measure the same phenomenon. Thus, coefficient α was computed for each anxiety dimension, namely cognitive anxiety, somatic anxiety and self-confidence, based on

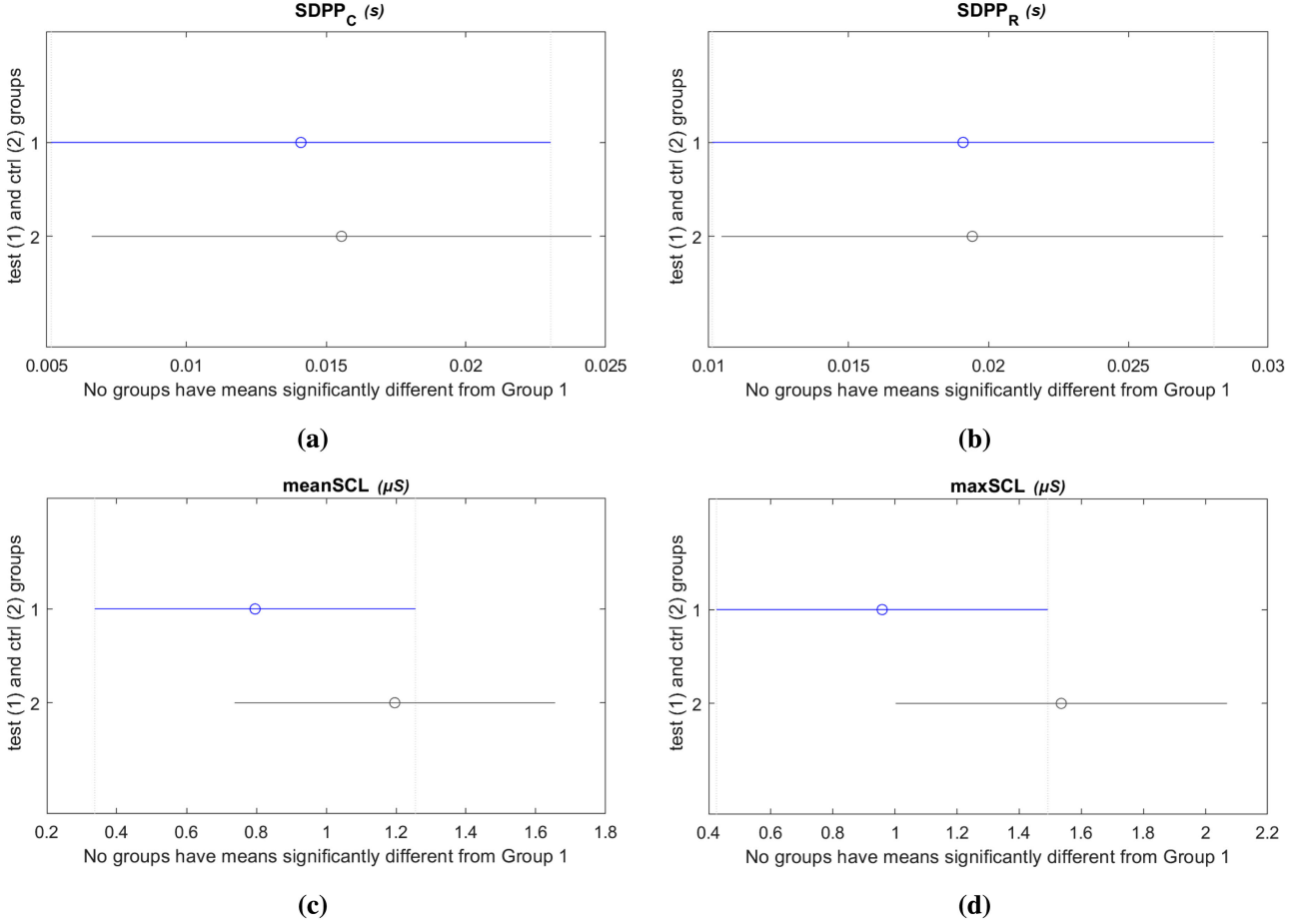


Fig. 10. *test* and *ctrl* group comparison with ANOVA for: (a) contact S , (b) remote S , (c) contact meanHR and (d) remote meanHR.

participants' answers before (pre-session) and after (post-session) experience sessions. As explained in [48], α values below 0.7 lead items selected to be questioned, whereas values starting from 0.7 involve items that have reliable consistency. All α values obtained ranged between 0.81 and 0.94, which implies items measuring the three anxiety dimensions are reliable. Pre-session α values were 0.85, 0.81 and 0.88 for the cognitive anxiety, somatic anxiety and self-confidence respectively, which indicate good item consistency according to [48]. Corresponding post-session values were 0.94, 0.90 and 0.91, pointing an excellent item consistency.

Somatic anxiety scores allowed to distinguish the two stress levels ($p = 0.03$), while cognitive anxiety and self-confidence scores failed in separating the levels (obtained p -values were respectively 0.12 and 0.33). Fig. 11 shows *test* and *ctrl* group comparison based on somatic anxiety scores. It can be noticed that there is nearly no evolution from pre to post-session for *ctrl* group (mean value of 0.01), meanwhile *test* group score mean value is 0.35. Positive mean values indicate somatic anxiety scores increase from pre to post-session. *test* mean value being superior to *ctrl* mean value leads us to assume the *test* scenario is more stressful. Cognitive anxiety mean values are also positive, and evolve the same way as for somatic anxiety. Self-confidence mean values for *test* and *ctrl* are both negative, meaning participants feel less confident with regard to their performance during the tasks at the end of the experience. Obtained mean value for *test* group is lower to *ctrl*, which supports

the hypothesis that *test* scenario induces higher state anxiety than *ctrl* scenario.

Machine learning classifiers used for stress state classification and task recognition were used for *test* and *ctrl* scenario classification. The four classifiers were applied to contact and remote PRV, EDA, combination of contact PRV and EDA features, as well as facial expression extracted as explained in Section 3.4. Contact PRV and EDA combination was obtained following the same procedure as for stress state classification and task recognition, with the exception of ANOVA groups that were defined according

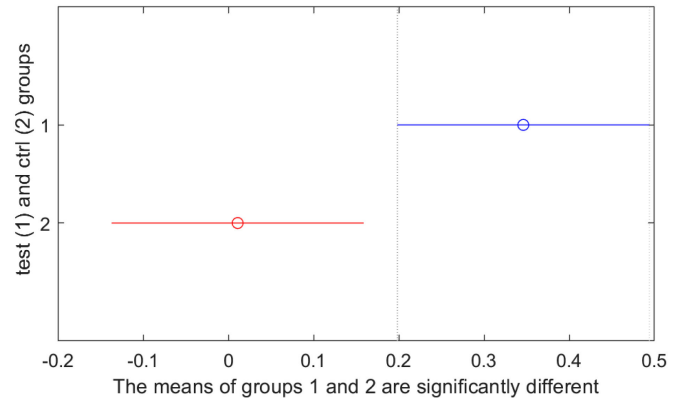


Fig. 11. *test* and *ctrl* group comparison for somatic anxiety normalized scores with ANOVA.

TABLE 4

Stress Level Recognition Results on Contact and Remote PRV Features (respectively denoted as PRV_C and PRV_R), EDA Features, Combination of Contact Features and Facial Expressions

Features	Classifiers	Accuracy (%)
PRV_C	SVM - linear kernel	58.19
	SVM - RBF kernel	62.27
	Log Reg	62.27
	KNN	54.56
PRV_R	SVM - linear kernel	69.73
	SVM - RBF kernel	57.48
	Log Reg	69.73
	KNN	58.76
EDA	SVM - linear kernel	63.60
	SVM - RBF kernel	61.56
	Log Reg	58.25
	KNN	57.99
$PRV_C + EDA$	SVM - linear kernel	63.60
	SVM - RBF kernel	60.43
	Log Reg	59.78
	KNN	58.50
Facial exp.	SVM - linear kernel	54.42
	SVM - RBF kernel	53.29
	Log Reg	54.68
	KNN	55.07

to *test* and *ctrl* class labels. Similarly, facial expression feature dimensionality was reduced and facial features that achieving the 20 best *F*-scores were kept. Physiological features were normalized by non-stress state (i.e., *T1*) values. Results, detailed in Table 4, show that all physiological data outperformed facial expression features, which had a maximum accuracy of 55.07 percent. This result may be explained by the fact that contrary to physiological responses which are spontaneous, facial expressions can be voluntary [49], [50]. Besides, Ekman made the hypothesis that some individuals do not display facial expression even when their physiological signals prove they are experiencing an emotional state [51]. These aspects may complicate characterizing stress based on facial activity. The best accuracy was reached by remote PRV (69.73 percent). EDA features as well as EDA and contact PRV combination yielded the same accuracy (63.60 percent).

Stress level classification results are globally more limited than those obtained for stress state recognition. A possible explanation would be that the speech and the arithmetic tasks imply a high stress level in both scenarios. In other words, talking about one's ideal holidays or doing arithmetic operation, even simple ones, may be experienced as very stressful in front of a stranger. Therefore, the difference between the two stress levels may have not present enough significance. Furthermore, the *test* version of the arithmetic task may be too difficult and produce a paradoxical demotivation effect on participants. Nevertheless, results obtained with physiological data are still promising, and somatic anxiety scores show a strong variation between *ctrl* and *test* groups.

In the three classification tests, contact PRV features give lower results than EDA. Combining contact PRV and EDA features does not give better accuracies than separated EDA

features. We assume that this may be due to the fact that information contained in contact PRV features do not complete EDA's. Besides, although we applied a subject selection step in order to maximize data exploitation, contact PRV features may present noisy information that worsen the combination's performance.

5 CONCLUSION AND FUTURE WORKS

This article presents a multimodal dataset involved with social stress effect on contact and remote physiological responses. Social stress was induced following a rigorous and well-prepared protocol, based on the commonly-used in psychophysiology TSST test. Electrodermal activity and blood volume pulse signals, as well as videos and self-reported anxiety scores of 56 participants are proposed. In this study, we validated the experimentation protocol through several tests and particularly proved that remote pulse rate variability could substitute contact PRV in similar experience conditions. This was supported by satisfying extracted HR comparison results. Moreover, stress state, task and level classifications showed remote PRV performed better than contact PRV. High accuracy of 85.48 percent was obtained for stress state recognition based on remote PRV features, surpassing all other modalities.

However, to obtain and validate our results, several BVP and RPPG signals were put aside due to noise occurring especially in the speech and arithmetic tasks. Future works will evaluate to what extent these signals can be used or, more specifically, whether the parts of these signals that are not noisy can be exploited. It will then be necessary to automatically and reliably determine the portions of the signal that are noise-free, and study the possibility to extract relevant information from these sequences. These exciting prospects pave the way for a more widespread use of non-contact physiological measurements towards a better understanding of our psyche.

ETHICAL AND TRANSPARENCY STATEMENT

The present study was conducted in accordance with the 1964 Helsinki declaration (World Medical Organization, 1964) and its later amendments, the French legislation on research involving human participants, the ethical principles of the French Code of Ethics for Psychologists (Commission Nationale Consultative de Déontologie des Psychologues, 2012), and the 2016 APA Ethical Principles of Psychologists and Code of Conduct (American Psychological Association, 2017).

ACKNOWLEDGMENTS

This work was supported by the Conseil Régional de Bourgogne Franche-Comté, France and the Fond Européen de Développement Régional (FEDER).

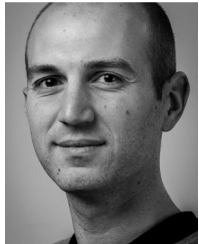
REFERENCES

- [1] P. Ekman, R. W. Levenson, and W. V. Friesen, "Autonomic nervous system activity distinguishes among emotions," *Science*, vol. 221, pp. 1208–1210, 1983.
- [2] S. D. Kreibug, "Autonomic nervous system activity in emotion: A review," *Biol. Psychol.*, vol. 84, pp. 394–421, 2010.

- [3] S. D. Kreibitz, "Autonomic nervous system activity in emotion: A review," *Biol. Psychol.*, vol. 84, pp. 394–421, 2010.
- [4] R. McCraty, M. Atkinson, and D. Tomasino, "Science of the heart - Exploring the role of the heart in human performance," *HeartMath Research Center, Institute of Heart Math, Publication No. 01-001*. Boulder Creek, CA, USA, 2001.
- [5] A. Schäfer and J. Vagedes, "How accurate is pulse rate variability as an estimate of heart rate variability? A review on studies comparing photoplethysmographic technology with an electrocardiogram," *Int. J. Cardiol.*, vol. 166, pp. 15–29, 2013.
- [6] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Opt. Express*, vol. 18, pp. 10762–10774, 2010.
- [7] R. Macwan, Y. benezeth, K. Nakamura, R. Gomez, Y. Wu, and A. Mansouri, "Parameter-free adaptive step-size multiobjective optimization applied to remote photoplethysmography," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat.*, 2018, pp. 267–270.
- [8] M. Lewandowska, J. Rumiński, and T. Kocejko, "Measuring pulse rate with a webcam - A non-contact method for evaluating cardiac activity," in *Proc. Federal Conf. Comput. Sci. Inf. Syst.*, 2011, pp. 405–410.
- [9] G. de Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 10, pp. 2878–2886, Oct. 2013.
- [10] Y. Sun and N. Thakor, "Photoplethysmography revisited: From contact to noncontact, from point to imaging," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 463–477, Mar. 2016.
- [11] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Algorithmic principles of remote-PPG," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1479–1491, Jul. 2017.
- [12] W. Chen and D. McDuff, "DeepPhys: Video-based physiological measurement using convolutional attention networks," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 356–373.
- [13] Z. Yu, W. Peng, X. Li, X. Hong, and G. Zhao, "Remote heart rate measurement from highly compressed facial videos: An end-to-end deep learning solution with video enhancement," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 151–160.
- [14] T. Laguev, D. Seuß, and J.-U. Garbas, "Deep learning based affective sensing with remote photoplethysmography," in *Proc. 54th Annu. Conf. Inf. Sci. Syst.*, 2020, pp. 1–4.
- [15] W. J. Yan, Q. Wu, Y. J. Liu, S.-J. Wang, and X. Fu, "CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces," in *Proc. 10th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit.*, 2013, pp. 1–7.
- [16] F. Qu, S. Wang, W. Yan, H. Li, S. Wu, and X. Fu, "CAS(ME)²: A database for spontaneous macro-expression and micro-expression spotting and recognition," *IEEE Trans. Affective Comput.*, vol. 9, no. 4, pp. 424–436, Fourth Quarter 2018.
- [17] Z. Zhang et al., "Multimodal spontaneous emotion corpus for human behavior analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 3438–3446.
- [18] S. Bobbia et al., "Unsupervised skin tissue segmentation for remote photoplethysmography," *Pattern Recognit. Lett.*, vol. 124, pp. 82–90, 2019.
- [19] X. Niu et al., "VIPL-HR: A multi-modal database for pulse estimation from less-constrained face video," in *Proc. Asian Conf. Comput. Vis.*, 2018, pp. 562–576.
- [20] R. Stricker, S. Müller, and H. M. Gross, "Non-contact video-based pulse rate measurement on a mobile service robot," in *Proc. 23rd IEEE Int. Symp. Robot Hum. Interactive Commun.*, 2014, pp. 1056–1062.
- [21] R. Špetlík, V. Franc, and J. Matas, "Visual heart rate estimation with convolutional neural network," in *Proc. Brit. Mach. Vis. Conf.*, 2018, pp. 1–12.
- [22] H. Kurniawan, A. V. Maslov, and M. Pechenizkiy, "Stress detection from speech and skin response signals," in *Proc. 26th IEEE Int. Symp. Comput.-Based Med. Syst.*, 2013, pp. 209–214.
- [23] F. Bousefsaf, C. Maaoui, and A. Pruski, "Remote assessment of the heart rate variability to detect mental stress," in *Proc. 7th Int. Conf. Pervasive Comput. Technol. Healthcare Workshops*, 2013, pp. 348–351.
- [24] D. McDuff, S. Gontarek, and R. Picard, "Remote measurement of cognitive stress via heart rate variability," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Medicine Biol. Soc.*, 2014, pp. 2957–2960.
- [25] R. Martens et al., "Development and validation of the competitive state anxiety inventory-2," in *Competitive Anxiety in Sport*, Champaign, IL, USA: Human Kinetics, 1990.
- [26] C. Kirschbaum, K. M. Pirke, and D. Hellhammer, "The 'Trier Social Stress Test' - A tool for investigating psychobiological stress responses in a laboratory setting," *Neuropsychobiology*, vol. 28, pp. 76–81, 1993.
- [27] C. McCarthy, N. Pradhan, C. Redpath, and A. Adler, "Validation of the empatica E4 wristband," in *Proc. IEEE EMBS Int. Student Conf.*, 2016, pp. 1–4.
- [28] J. Pietilä et al., "Evaluation of the accuracy and reliability for photoplethysmography based heart rate and beat-to-beat detection during daily activities," in *Proc. Eur. Med. Biol. Eng. Conf.*, 2017, pp. 145–148.
- [29] J. A. Onton, D. Y. Kang, and T. P. Coleman, "Visualization of whole-night sleep EEG from 2-channel mobile recording device reveals distinct deep sleep stages with differential electrodermal activity," *Front. Hum. Neurosci.*, vol. 10, 2016, Art. no. 605.
- [30] T. Kundinger, N. Sofra, and A. Riener, "Assessment of the potential of wrist-worn wearable sensors for driver drowsiness detection," *Sensors*, vol. 20, 2020, Art. no. 1029.
- [31] M. Matsubara et al., "Emotional arousal estimation while reading comics based on physiological signal analysis," in *Proc. 1st Int. Workshop coMics Analysis Process. Understanding*, 2016, Art. no. 7.
- [32] A. P. Allen et al., "The trier social stress test: Principles and practice," *Neurobiol. Stress*, vol. 6, pp. 113–126, 2017.
- [33] B. M. Kudielka, D. H. Hellhammer, and C. Kirschbaum, "Ten years of research with the trier social test-revisited," *Soc. Neurosci.: Integrating Biol. Psychol. Explanations Soc. Behav.*, vol. 56, pp. 56–83, 2007.
- [34] C. Kirschbaum, S. Wüst, and D. Hellhammer, "Consistent sex differences in cortisol responses to psychological stress," *Psychosom. Med.*, vol. 54, pp. 648–657, 1992.
- [35] W. Liu et al., "SSD: Single shot MultiBox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.
- [36] C. O. Conaire, N. E. O'Connor, and A. F. Smeaton, "Detector adaptation by maximising agreement between independent data sources," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2007, pp. 1–6.
- [37] M. P. Tarvainen et al., "Kubios HRV (ver. 3.3) users's guide," *Kubios Oy*, 2019.
- [38] R. M. Baevsky and A. P. Berseneva, "Methodical recommendations use kardivar system for determination of the stress level and estimation of the body adaptability standards of measurements and physiological interpretation," Moscow - Prague, pp. 1–42, 2008.
- [39] M. P. Tulppo et al., "Quantitative beat-to-beat analysis of heart rate dynamics during exercise," *Amer. J. Physiol.*, vol. 271, pp. 244–252, 1996.
- [40] M. Benedek and C. Kaernbach, "A continuous measure of phasic electrodermal activity," *J. Neurosci. Methods*, vol. 190, pp. 80–91, 2010.
- [41] C. H. Hjortsjö, *Man's Face and Mimic Language*. Sweden: Studentlitteratur, 1969.
- [42] P. Ekman and W. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto, CA, USA: Consulting Psychol. Press, 1978.
- [43] G. Berntson and J. R. Stowell, "ECG artifacts and heart period variability: Don't miss a beat!," *Psychophysiology*, vol. 35, pp. 127–132, 1998.
- [44] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *Eur. Heart J.*, vol. 93, no. 5, pp. 1043–1065, 1996.
- [45] R. Castaldo et al., "Ultra-short term HRV features as surrogates of short term HRV: A case study on mental stress detection in real life," *BMC Med. Informat. Decis. Making*, vol. 19, 2019, Art. no. 12.
- [46] L. Pecchia et al., "Are ultra-short heart rate variability features good surrogates of short-term ones? State-of-the-art review and recommendations," *Healthcare Technol. Lett.*, vol. 5, pp. 94–100, 2018.
- [47] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika*, vol. 16, pp. 297–334, 1951.
- [48] D. George and P. Mallery, *SPSS for Windows Step by Step: A Simple Guide and Reference*. Boston, MA, USA: Allyn and Bacon, 2003.
- [49] P. Gosselin, M. Perron, and M. Beaupré, "The voluntary control of facial action units in adults," *Emotion*, vol. 10, pp. 266–271, 2010.
- [50] K. L. Schmidt et al., "Movement differences between deliberate and spontaneous emotion facial expressions: Zygomaticus major action in smiling," *J. Nonverbal Behav.*, vol. 30, pp. 37–52, 2006.
- [51] P. Ekman, "Facial expression and emotion," *Amer. Psychol.*, vol. 48, pp. 384–392, 1993.



Rita Meziati Sabour received the engineering degree in electronics with a specialization in signal and image processing, from the ENSEIRB-MAT-MECA Engineering School, Bordeaux, France, and the MS degree from the College of Science and Technology, University of Bordeaux, Bordeaux, France. She is currently working toward the PhD degree with the ImViA Laboratory, University of Bourgogne Franche-Comté, Dijon, France. Her research interests include emotion recognition and remote physiological signal measurement and analysis.



Yannick Benezeth received the engineering degree from the ENSI de Bourges, the MS degree from the University of Versailles-Saint-Quentin-en-Yvelines, France, in 2006, and the PhD degree in computer science from the University of Orléans, France, in 2009. He is currently an associate professor in computer science with the University Bourgogne Franche-Comté, Dijon, France. He worked as a research fellow with the Orange Labs and INRIA, Rennes, France, between 2009 and 2011. He has co-supervised or is supervising eight

PhD students since 2011. He was a reviewer of significant scientific journals the *IEEE Transation on Image Processing*, *IEEE Transation on Circuits and Systems for Video Technology*, *Pattern Recognition letters*, and international conferences. He was a visiting researcher with the University de Sherbrooke, Canada, NECTEC, Thailand, University of Magdeburg, Germany and Boston University, Boston, Massachusetts. He participated in several French and international projects in the field of computer vision (CNRS, ANR, PHC, etc.). His research interests include biomedical engineering, affective computing, image processing, and video analytics. His application areas include video health monitoring and endoscopy. Since 2009, he has published 17 international journals and more than 40 international conferences and patents.



Pierre De Oliveira received the PhD degree in psychology from the LAPSCO-UMR UBP-CNRS 6024, Clermont-Ferrand, France, in 2009. He is currently an associate professor in social psychology with the University of Bourgogne Franche-Comté, Dijon, France. In 2011, he completed a doctoral fellowship at the Political Psychology Research Center, University of Belfast, U.K. Since 2011, he has been working at the Psy-DREPI Laboratory (Laboratoire de Psychologie: Dynamiques Relationelles Et Processus Identitaires, EA 7458). His research is organized around two main lines of research. The first aims at examining and understanding cognitive and motivational processes involved in the maintenance and legitimization of social inequalities (e.g., social dominance, stigmatization, hierarchy threat, etc.). The second focuses more specifically on collective and personal behavior in regulating stress and uncertainty situations (e.g., stress mindset, social cure, collective resilience, etc.). Since 2007, he has published several international journals and communicated in more than 30 international conferences.



Julien Chappé received the PhD degree in psychology from the University of Paris X Nanterre, France, in 2007. He is currently an associate professor in social psychology with the University of Bourgogne Franche-Comté, Dijon, France. In 2008, he completed a doctoral fellowship at the Institut National de Recherche sur les Transports et leur Sécurité (INRETS). In 2010, he was research associate at the INSEAD Business School, Centre de Recherche en Sciences Sociales (ISSRC), in Paris. Between 2011 and

2013, he worked as researcher and consultant at the Institut Français d'Action sur le Stress (IFAS) in Paris. Since 2013, he is an associate professor in social psychology at the Psy-DREPI Laboratory (Laboratoire de Psychologie : Dynamiques Relationelles Et Processus Identitaires, EA 7458). His research is organized around two axes. The first axis deals with persuasive communication and non-conscious influences (message framing, health halo effect, priming). The second axis focuses on organizational health and more specifically on quality of life and stress regulation (stress mindset). His work has been the subject of national and international publications and communications.



Fan Yang received the BS degree from the University of Lanzhou, China, and the MS and PhD degrees from the University of Burgundy, France, in 1994 and 1998, respectively. She is currently a full professor with the University Bourgogne Franche-Comté, Dijon, France. She worked as an associate professor with the University of Lanzhou, China between 1982 and 1999, then at the University Institute of Technology UIT of Dijon, France between 2000 and 2007. Since 2008, she works as a full professor at the UIT of

Dijon, France. She has supervised or co-supervised 20 PhD students since 2000. From the beginning of her career, she published 43 papers in international peer-reviewed journals, three book chapters, and authored more than 90 conference papers. She was a reviewer of significant scientific journals the *IEEE Transation on Neural Networks*, *IEEE Transation on Circuits and Systems for Video Technology*, *IEEE Transation on Circuits and Systems*, *Pattern Recognition*, *Pattern Recognition Letters*) and international conferences. She was visiting researcher at the University of Montreal, Canada, in 2010. Her research interests include pattern recognition, neural networks, multispectral imaging, parallelism and real-time implementation, and more specifically, biometric image processing.

► For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/csdl.