Automatic Detection of Acute Mental Stress With Camera-based Photoplethysmography

Hannes Ernst, Hagen Malberg, Martin Schmidt

Institute of Biomedical Engineering, TU Dresden, Germany

Abstract

Autonomic nervous system (ANS) activity reflects in vital signs that can be measured by means of camera-based photoplethysmography (cbPPG). This work investigates the automatic detection of acute mental stress with cbPPG. Data from the Dresden Multimodal Biosignal Dataset for the Mannheim Multicomponent Stress Test (DMBD-MMST) covering $> 40\,\mathrm{h}$ *uncompressed facial RGB videos of 56 healthy participants were used for rest vs. stress classification on the basis of nine cbPPG vital signs with decision tree ensembles. Also, the impact of normalization, measurement duration, and color channel combination was investigated. Best performance for rest (baseline and recovery) vs. stress classification (*$F1 = 0.81$*, Cohen's Kappa* $\kappa = 0.61$*) was achieved with normalization,* $30\,\mathrm{s}$ *measurement duration, and vital signs from the green channel and the color channel combination called O3C. Without recovery (baseline vs. stress), this configuration achieved* $F1 = 0.97$ *and* $\kappa = 0.94$*. Paired t-tests revealed significant changes from rest (baseline and recovery) to stress in eight of the nine vital signs and the maximum effect size amounted to* $d = 0.73$*, indicating sympathetic excitation. Findings from this work are central to the non-contact evaluation of ANS activity. Our results demonstrate that automatic detection of acute mental stress with cbPPG is feasible.*

1. Introduction

Camera-based photoplethysmography (cbPPG) is an optical technique for non-contact assessment of skin perfusion [1]. Pulsatile fluctuations in dermal blood volume lead to modulation of the interaction of light with tissue [1], which allows the measurement of vital signs related to the cardiovascular system [2, 3] and respiration [4].

Such vital signs provide information about the state of the autonomic nervous system. For example, sympathetic excitation as stimulated by physical pain from cold stress alters perfusion, which is detectable with cbPPG [5].

Cameras often provide red, green, and blue color channels (RGB), and photoplethysmographic signals can be derived from different color channel combinations [1]. It has been shown for pulse rate [2], perfusion parameters [3], and breath rate [4] that cbPPG measurement accuracy varies greatly with the color channel combination.

The objective of this contribution is to determine to what extent acute mental stress can be automatically detected by cbPPG vital signs. Moreover, the influence of color channel combinations, measurement durations and individual normalization is investigated.

2. Methods and Materials

To investigate the effects of acute mental stress with cbPPG, we utilized the Dresden Multimodal Biosignal Dataset for the Mannheim Multicomponent Stress Test (DMBD-MMST) [6]. Acute mental stress was induced by an arithmetic task complemented by several other stressors such as noise and negative performance feedback [6]. Figure 1 illustrates the study procedure and timeline. The DMBD-MMST contains not only physiological measurements but also synchronized facial video data during the conditions rest ($c = 0$) and stress ($c = 1$). Synchronized video data was available for 56 participants (see Table 1).

The camera (UI-3060CP-C-HQ Rev.2, IDS GmbH, Obersulm, Germany) recorded 100 uncompressed frames

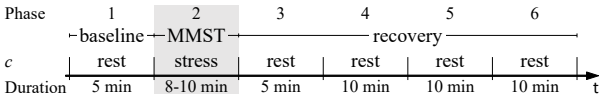

Figure 1: Timeline and condition c of the DMBD-MMST.

Table 1: Demographic data of the 56 participants (24 female, 32 male, all healthy). BMI: Body mass index.

	Unit	Mean ± SD	Range
Age	yr	25.9 ± 4.8	18 – 39
Weight	kg	69.0 ± 12.6	50 – 108
Height	cm	175.4 ± 9.3	158 – 194
BMI	$\mathrm{kg/m^2}$	22.3 ± 2.7	17.4 – 31.6

 ISSN: 2325-887X DOI: 10.22489/CinC.2023.034

per second with a spatial resolution of $320\,\mathrm{px} \times 640\,\mathrm{px}$ and an amplitude resolution of $3 \times 12\,\mathrm{bit}$ (RGB).

The region-of-interest (ROI) was extracted with the Level-Set technique described in [7]. Frame-wise averaging of the pixel intensities within the ROI yielded a signal for each of the color channels red R, green G, and blue B.

Photoplethysmographic signals were derived with color channels combinations G (green channel G), O3C [2], CHROM [8], and POS [9] for separate evaluation. It should be noted that G and O3C are static while CHROM and POS, both deduced from the dichromatic reflection model of skin, are dynamic color channel combinations that normalize their signal and tune it with amplification factors redetermined every $1.6\,\mathrm{s}$ [2, 8, 9]. Optimization of the data-driven approach O3C for the DMBD-MMST was performed by means of a hemispherical surface grid search on the accuracy of pulse rate measurement in comparison to the earlobe photoplethysmogram as described in [2].

Signals were split into segments of $10\,\mathrm{s}$ length with $3\,\mathrm{s}$ step size resulting in $47\,637$ segments for each color channel combination ($41\,597$ rest, 6040 stress). To investigate the effect of measurement duration, signals were also split into segments of $30\,\mathrm{s}$ length with $9\,\mathrm{s}$ step size resulting in $15\,295$ segments ($13\,380$ rest, 1915 stress).

For each segment of each color channel combination, nine vital signs were extracted following the references provided: pulse rate (PR) [2], logarithmized pulse strength (PS) [10], breath rate (BR) by means of the frequency modulation approach [4], pulse arrival time (PAT) [11], average low-frequent intensity (DC) [6], and perfusion index (PI) [12], as well as the product of PR and PI (PRxPI), the product of DC and PS (C1), and the ratio of PR to C1 (C2). Furthermore, we set up a mixed vital sign set (MIX) in which DC was derived from the green channel due to its direct physiological link to photon absorption [3] and all other vital signs were derived with O3C for increased signal robustness [2] without using dynamic signals (CHROM or POS) that normalize signals and frequently alter signal amplification, which potentially corrupts DC and PS. To investigate the effect of inter-individual differences, vital signs were normalized to the individual baseline average.

We trained binary decision tree ensembles (method: LogitBoost, predictor binning: 256 steps, learn rate: 0.1) to predict the condition c (0: rest, 1: stress) from the cbPPG vital signs (rest vs. stress classification). Hyperparameter optimization included the number of splits (5 to 100) and the number of learners (1 to 1000). Data were split in participant-exclusive and stratified manner for 3-fold cross-validation. To gain a balanced dataset, the majority class was randomly subsampled. Classification performance was evaluated with F1 score and Cohen's Kappa (test set average of cross-validation). F1 score:

$$F1 = 2 * (PRE * SEN)/(PRE + SEN) \quad (1)$$

combines precision PRE and sensitivity SEN in a single metric while Cohen's Kappa:

$$\kappa = (ACC - p_0)/(1 - p_0) \quad (2)$$

is a metric that corrects the prediction accuracy ACC by the chance of random guessing $p_0 = 0.5$ (balanced binary classification). It has been proposed to rate values of $\kappa < 0$ as poor, 0–0.2 as slight, 0.21–0.4 as fair, 0.41–0.6 as moderate, 0.61–0.8 as substantial, and 0.81–1 as almost perfect agreement (between classifier and ground truth) [13]. Paired t-tests and effect sizes Cohen's d were calculated from the individual means for each participant in each condition for statistical analyses of the best performing configuration (without normalization).

To summarize, binary decision tree ensemble classifiers were trained and tested for 20 configurations that resulted from the combination of four color channel combinations (G, O3C, CHROM, POS) as well as the mixed vital sign set (MIX), two segment lengths ($10\,\mathrm{s}$, $30\,\mathrm{s}$), and two settings for vital sign normalization (with, without).

3. Results

The hemispherical surface grid search for O3C led to:

$$O3C_{DMBD} = 0.46 \cdot R - 0.83 \cdot G + 0.33 \cdot B \quad (3)$$

with $O3C_{DMBD}$ as the optimal static linear RGB color channel combination for the DMBD-MMST. In comparison to reference pulse rates from the earlobe photoplethysmogram as defined in [2], O3C yielded an accuracy of $97.8\,\%$ (G: $88.3\,\%$, CHROM: $97.0\,\%$, and POS $97.9\,\%$).

Table 2 contains the results for the evaluation metrics $F1$ (Table 2a) and κ (Table 2b) of all configurations. Best rest vs. stress classification performance of $F1 = 0.81$ and $\kappa = 0.61$ was achieved by MIX with normalization and $30\,\mathrm{s}$ segment length. This result corresponds to substantial agreement with the ground truth. Performance differences mainly manifested in κ (maximum difference of $F1$ across all channels within a configuration: 0.04). On average, normalization and longer segments led to relative improvements of $11.2\,\%$ and $3.4\,\%$ for $F1$ and $193.7\,\%$ and $17.0\,\%$ for κ, respectively.

Table 3 provides an overview over the cbPPG vital signs during rest and stress. Paired t-tests yielded significant results for all cbPPG vital signs except PI. Absolute effect sizes ranged from 0.18 (PI) to 0.73 (C2), with relative changes of mean of $+4\,\%$ and $+42\,\%$, respectively. Vital signs indicated positive chronicity (PR), peripheral vasoconstriction (PS and DC), and increased blood pressure (PAT), all consistent with sympathetic excitation. Increased breath rate indicated the processing of acute men-

Table 2: Evaluation metrics F1 score and Cohen's Kappa for the rest vs. stress classification of all 20 configurations.

(a) F1 score.

$F1$	Length	10 s		30 s	
	Normalization	no	yes	no	yes
Channel	G	0.68	0.76	0.69	0.79
	O3C	0.68	0.73	0.69	0.78
	CHROM	0.67	0.72	0.67	0.77
	POS	0.69	0.72	0.68	0.77
	MIX	0.69	0.76	0.70	0.81

(b) Cohen's Kappa.

κ	Length	10 s		30 s	
	Normalization	no	yes	no	yes
Channel	G	0.22	0.49	0.21	0.54
	O3C	0.15	0.41	0.26	0.48
	CHROM	0.10	0.33	0.06	0.51
	POS	0.18	0.36	0.19	0.50
	MIX	0.33	0.48	0.29	0.61

Table 3: Mean, standard deviation (SD), and relative change of mean (ΔM) of the vital signs from the MIX configuration (30 s segment length, without normalization). The conditions rest and stress were compared with paired t-tests of the individual means in each condition ($N = 56$). Cohen's d was calculated with pooled standard deviation.

Vital		Rest ($n = 13\,380$)		Stress ($n = 1915$)				
sign	Unit	Mean	SD	Mean	SD	ΔM	p	d
PR	bpm	68.9	12.8	77.2	17.7	+12 %	3.17e−07 ***	0.62
PS	a. u.	1.63	0.36	1.52	0.40	−6 %	8.52e−04 **	−0.29
DC	a. u.	2006	425	1824	497	−9 %	5.02e−05 ***	−0.42
BR	rpm	15.5	4.1	17.1	5.4	+10 %	7.61e−12 ***	0.36
PAT	ms	223	51	205	50	−8 %	1.87e−04 **	−0.35
PI	a. u.	8.33e−04	1.88e−04	8.67e−04	2.46e−04	+4 %	3.20e−02	0.18
PRxPI	a. u.	5.64e−02	1.21e−02	6.46e−02	1.36e−02	+15 %	3.29e−09 ***	0.67
C1	a. u.	3328	1172	2893	1328	−13 %	1.29e−04 **	−0.36
C2	a. u.	2.44e−02	1.27e−02	3.46e−02	2.15e−02	+42 %	6.86e−07 ***	0.73

p-values reported uncorrected, significance markers with Bonferroni correction: *: $p < 0.05/9$, **: $p < 0.01/9$, ***: $p < 0.001/9$. PAT logarithmized before paired t-test. N: Number of participants. n: Number of segments.

tal stress by the amygdala and the anterior cingulate cortex in line with the literature [6].

4. Discussion

Results for κ and $F1$ correlate strongly ($r = 0.96$, $p < 0.001$), which indicates statistical robustness of the methodological approach even if performance differences manifest to varying degrees in the two metrics.

Classification performance benefits from normalization as it reduces inter-subject variability and facilitates the focus on intra-subject changes. The choice of segment length remains a compromise: Measurements on longer segments benefit from averaging effects and become less prone to transient artefacts, but lead to reduced temporal resolution that may be insufficient to reflect physiological changes [1]. While segments of 10 s length suffice to measure the pulse rate with cbPPG [1], breath rate measurement becomes more accurate for segments of 30 s length [4]. It is reasonable that more accurate measurement results lead to higher classification performance.

Signal normalization of CHROM and POS diminishes the informational content of the vital sign DC [3], and therefore also affects PI. O3C does not require normalization and the low-frequent components of O3C have been shown to contain vital information [3, 4]. In general, if O3C is optimized for metrics targeting the cardiac pulsatile signal component, it converges towards color channel combinations that resemble static versions of CHROM and POS [2,14], which empirically indicates the validity of the application of the dichromatic reflection model. However, if the same kind of optimization is performed with metrics targeting low-frequent signal components (e.g. the baseline modulation caused by respiration), the optimization converges towards the green channel [4]. These observations are consistent with the increased classification performance of MIX that combines low-frequent information from the green channel with other vital signs more robustly extracted by O3C. Accordingly, the use of more than one color channel combination depending on the vital sign to be measured stands to reason.

Because the rest condition distributes over 40 min, a certain physiological variation of the vital signs is to be expected. To investigate the effect on classification performance, we used the best performing configuration to train and test decision tree ensembles following the same procedure as before but with data from a single rest phase instead of all rest phases. Table 4 contains the results of this approach for all five rest phases. It can be seen that the discrimination of stress phase and baseline with $F1 = 0.97$ and $\kappa = 0.94$ achieved almost perfect agreement with the ground truth while classification performance decreased for rest phases after the stress phase (on average $F1 = 0.77$ and $\kappa = 0.52$). There are multiple possible explanations for this: The first phases were closer in time to the set point for normalization. Set points of phys-

iological regulation processes may alter over time (allostasis), and the sympathovagal balance was disturbed prior to the rest phases 3 to 6. This also means that recovery processes may be on-going, e.g. to readjust blood pressure. It could also be hypothesized that participants were most relaxed during the first phase (baseline) because this was the only phase without preceding saliva sampling[1] [6].

Table 4: Evaluation metrics F1 score $F1$ and Cohen's Kappa κ of the rest vs. stress classification by phase (MIX, $30\,\mathrm{s}$ segment length, with normalization).

	Baseline	Recovery			
Rest phase	1	3	4	5	6
$F1$	0.97	0.78	0.78	0.78	0.75
κ	0.94	0.57	0.50	0.57	0.42

PAT, used as a substitute for pulse transit time (PTT), has been proven sensitive to acute mental stress but includes the cardiac pre-ejection period and requires an electrocardiogram [11]. Without PAT, κ decreased by $8\,\%$ from 0.61 to 0.56 (MIX, $30\,\mathrm{s}$ segment length, with normalization). The non-contact approach of cbPPG would benefit from a transition from PAT to purely video-based PTT from facial recordings. However, such measurement of PTT requires extremely high frame rates, which, at present, is impracticable for uncompressed videos of reasonably high resolution. It can be assumed that technological advancements will make the parameter more accessible in the near future.

Finally, it should be noted that the rest vs. stress classification is a general approach, i.e. it omits the participant-specific stress-strain relationship. However, successful stress induction for the DMBD-MMST has already been shown beforehand [6, 15].

5. Conclusion

We present a study on the automatic detection of acute mental stress by means of cbPPG. Our results show substantial to almost perfect ground truth agreement. Our findings are relevant to evaluate the suitability of cbPPG to assess the activity of the autonomic nervous system without contact. Future research should investigate options to increase classification performance during recovery, e.g. by extending segments to include pulse rate variability.

Acknowledgments

This work was supported by grants from the European Union's Horizon 2020 research and innovation program (TIMELY, No. 101017424).

[1] Chewing and spitting out an absorbent roll in front of the study supervisor can be considered an unusual interaction with an authority figure (social-evaluative stress). This may leave participants mentally strained at the start of the next phase even in rest condition.

References

[1] Zaunseder S, et al. Cardiovascular assessment by imaging photoplethysmography - a review. Biomed Eng Biomed Tech jun 2018;63(5):617–634.

[2] Ernst H, et al. Optimal color channel combination across skin tones for remote heart rate measurement in camera-based photoplethysmography. Biomed Signal Process Control jul 2021;68:102644.

[3] Hammer A, et al. Camera-based assessment of cutaneous perfusion strength in a clinical setting. Physiol Meas feb 2022;43(2):025007.

[4] Ernst H, Malberg H, Schmidt M. Non-contact Measurement of Respiration Rate With Camera-based Photoplethysmography During Rest and Mental Stress. In 49th Conf. Comput. Cardiol. Tampere: IEEE, 2022; 1–4.

[5] Fleischhauer V, et al. Impact of Sympathetic Activation in Imaging Photoplethysmography. In 2019 IEEE/CVF Int. Conf. Comput. Vis. Work. IEEE, oct 2019; 1697–1705.

[6] Ernst H, et al. Assessment of the Human Response to Acute Mental Stress – An Overview and a Multimodal Study. PLOS ONE 2023;under review.

[7] Trumpp A, et al. Skin Detection and Tracking for Camera-Based Photoplethysmography Using a Bayesian Classifier and Level Set Segmentation. In Bild. für die Medizin 2017. Berlin, Heidelberg: Springer Vieweg, 2017; 43–48.

[8] de Haan G, Jeanne V. Robust Pulse Rate From Chrominance-Based rPPG. IEEE Trans Biomed Eng oct 2013;60(10):2878–2886.

[9] Wang W, et al. Algorithmic Principles of Remote PPG. IEEE Trans Biomed Eng jul 2017;64(7):1479–1491.

[10] Ernst H, Malberg H, Schmidt M. More Reliable Remote Heart Rate Measurement by Signal Quality Indexes. In 47th Conf. Comput. Cardiol. IEEE, dec 2020; 1–4.

[11] Ernst H, et al. Pulse Arrival Time - A Sensitive Vital Parameter for the Detection of Mental Stress. Curr Dir Biomed Eng oct 2021;7(2):419–422.

[12] Rasche S, et al. Association of remote imaging photoplethysmography and cutaneous perfusion in volunteers. Sci Rep oct 2020;10(1):16464.

[13] Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics mar 1977; 33(1):159.

[14] Ernst H, Malberg H, Schmidt M. Optimierung der kamerabasierten Photoplethysmographie für die kontaktlose Messung kardiovaskulärer Parameter. In Proc. Work. Biosignale 2022. Dresden, 2022; 1–2.

[15] Ernst H, et al. Cardiovascular Effects of Mental Stress in Healthy Volunteers. In 48th Conf. Comput. Cardiol. Brno: IEEE, sep 2021; 1–4.

Address for correspondence:

Dipl.-Ing. Hannes Ernst
Institute of Biomedical Engineering, TU Dresden
Fetscherstr. 29, 01307 Dresden, Germany
hannes.ernst@tu-dresden.de