



## LETTER

# Multimodal data acquisition at SARS-CoV-2 drive through screening centers: Setup description and experiences in Saarland, Germany

Philipp Flotho  | Mayur J. Bhamborae | Tobias Grün | Carlos Trenado |  
David Thinnés | Dominik Limbach | Daniel J. Strauss\* 

Systems Neuroscience and Neurotechnology Unit, Neurocenter, Faculty of Medicine, Saarland University and School of Engineering, htw saar, Saarbrücken, Germany

## \*Correspondence

Daniel J. Strauss, Systems Neuroscience and Neurotechnology Unit, Neurocenter, Faculty of Medicine, Saarland University and School of Engineering, htw saar, Saarbrücken, Germany.  
Email: daniel.strauss@uni-saarland.de

## Abstract

SARS-CoV-2 drive through screening centers (DTSC) have been implemented worldwide as a fast and secure way of mass screening. We use DTSCs as a platform for the acquisition of multimodal datasets that are needed for the development of remote screening methods. Our acquisition setup consists of an array of thermal, infrared and RGB cameras as well as microphones and we apply methods from computer vision and computer audition for the contactless estimation of physiological parameters. We have recorded a multimodal dataset of DTSC participants in Germany for the development of remote screening methods and symptom identification. Acquisition in the early stages of a pandemic and in regions with high infection rates can facilitate and speed up the identification of infection specific symptoms and large-scale data acquisition at DTSC is possible without disturbing the flow of operation.

## KEYWORDS

computer vision, COVID-19, drive through screening, non-contact medical assessment, SARS-CoV-2

## Large-scale data acquisition at SARS-CoV-2 drive through screening centers



## 1 | INTRODUCTION

Research on digital technologies to combat the COVID-19 pandemic includes the computational analysis of video and audio data [1–3]. Due to their contactless nature, such methods are particularly promising and needed for mass

screening purposes [3] and besides fever, there are other atypical and non-severe symptoms (e.g. [4–7]) which allow for non-contact medical assessment. The value of such screening systems would of course be directly related to the achievable sensitivity and specificity for detecting SARS-CoV-2 infections. However, it is challenging to acquire

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Journal of Biophotonics* published by Wiley-VCH GmbH.

homogeneous data sets for the development and assessment of such remote systems without interfering with medical services due to the pressure of the ongoing pandemic.

For the rapid collection of samples for polymerase chain reaction (PCR) based screening, drive-through screening centers (DTSCs), for example, Kwon et al. [8], are already in use in several countries. There are advantages of DTSCs for the acquisition of contactless recordings of the patients: They simulate a scenario, where contactless screening of infectious diseases might be implemented 1 day, such as the entrance of employee parking area. The exposure of equipment and personnel to patients is minimized and at the same time the exposure of healthy participants to contaminated air or equipment is fully controlled and can be completely avoided as the patients stay seated in their own car.

We propose an acquisition system along with a processing pipeline for rapidly acquiring such data at DTSC without disturbing their flow of medical operation and present a multimodal dataset of DTSC users as well as preliminary evaluations.

## 2 | MATERIALS AND METHODS

We recorded our dataset between May and July 2020 at the SARS-CoV-2 DTSC located at the former fairground area in Saarbrücken, State of Saarland, Germany. The study was approved by the responsible ethics committee (ethics commission at the Ärztekammer des Saarlandes, ID No 90/20) and after a detailed explanation of the procedure, all included participants signed a consent form. Admission to and recommendation for the tests was given by the participants' general practitioner if a patient had a potential SARS-CoV-2 infection based on the Robert Koch Institute's guidelines [9]. The PCR-test result for SARS-CoV-2 from the individual nose and throat swab was accessible for us at the responsible public health office. The recordings were done through an opened window with the participants sitting in their car. Our multimodal setup consisted of RGB, NIR, depth and thermal cameras as well as microphones (see Figure 1). We recorded at 120fps (face closeups, RGB), 50fps (thermal camera), 30fps (NIR) and 10fps (high resolution RGB, stereo) and used custom acquisition routines and frame grabbers to minimize the user interaction with the recording systems. The investigators followed the same guidelines for the personal protective equipment (PPE) as the physicians taking the swab samples. Participants waiting for the experiment were regularly informed about the estimated waiting period and had the option to quit the experiment between two recordings, to reduce the contamination with ambient sound during audio

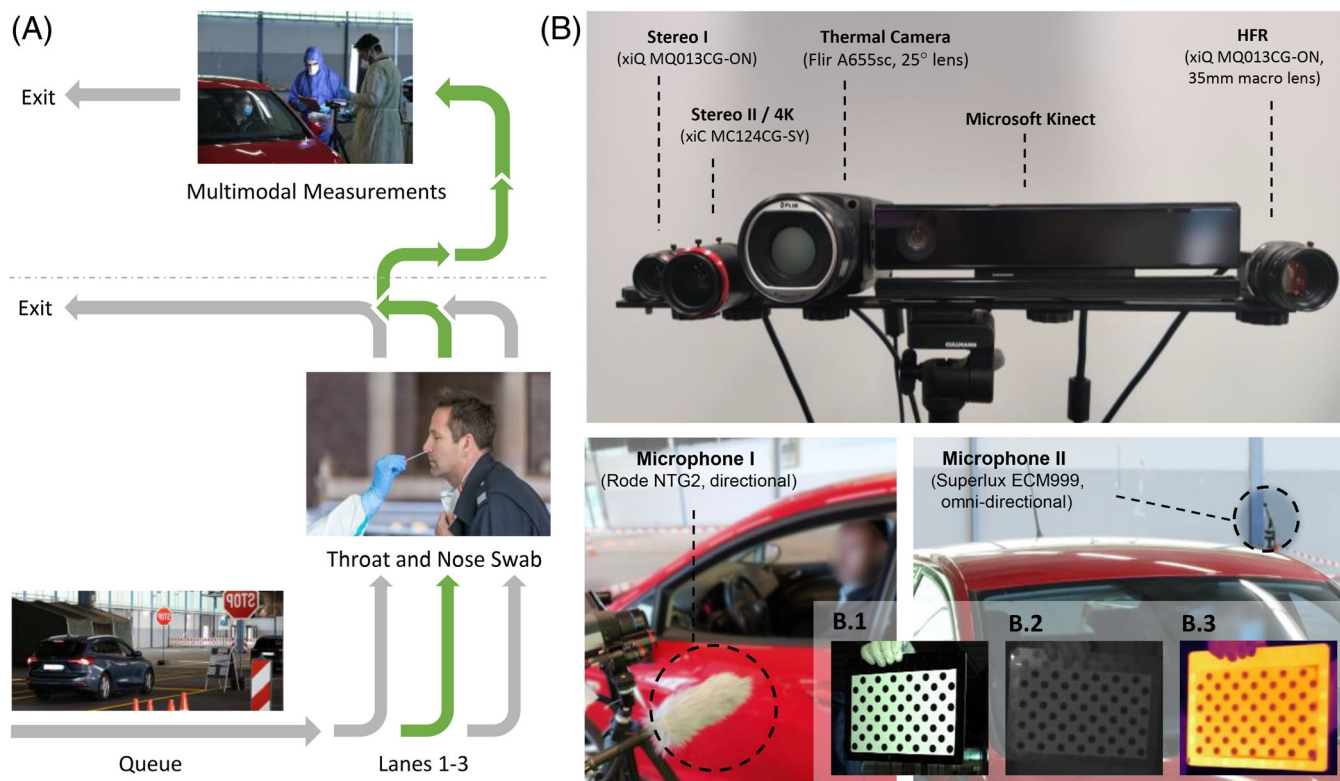
recordings. The experiment began with a set of yes/no/unknown questions with the goal to generate uniform voice samples that can be compared between subjects. These also provided additional medical history. The second segment was a free speech sample, where the participants were asked to tell the circumstances that led to them visiting the DTSC. Subsequently, we asked the participants to briefly present their hands from both sides towards the camera, to catch potential cues to skin rashes. In the final segment, participants were asked to take 10 deep breaths and to breathe normally for 30 seconds afterwards. The entire study/data acquisition took around 6 minutes. A total of 436 participants with signed consent form participated in our study, aged 19–86 (mean age  $45.6 \pm 15.2$ , 215 males, 221 females, 7 participants did not provide their age). Thirty-four subjects reported chronic or acute respiratory diseases or symptoms (see Figure 3). Despite a relatively high participation of 36% of the DTSC users in our study, our data set contained only two subjects with SARS-CoV-2 positive PCR results from the swab tests at the DTSC, see discussion.

## 3 | RESULTS

We have recorded a dataset up to 6 minutes per subject with HFR or high-resolution multimodal cameras and microphones. We applied already available procedures from computer vision and computer audition for assessment of the data quality. The evaluations and respective methodology used as proof of concept are described below. Figure 2 summarizes the results for each of these modalities, in particular how the contactless physiological parameters compare to gender and age specific norm values and grouped them by a presence/absence of symptoms. Due to ethics / privacy conditions, evaluations on the full dataset are limited to our local infrastructure. However, we have recordings of individual participants that agreed to have their recordings made publicly available and offer those for download on our project website. The code for the calibration and reading of the demo data can be found on our GitHub.\*

### 3.1 | System calibration and camera mapping

The camera setup was calibrated with a  $4 \times 13$  circular calibration pattern glued to a cut-out metal plate. The pattern was heated with two heating blankets prior to acquisition and therefore was visible in the thermal, NIR, and RGB cameras (see Figure 1B.1-B.3). Since the macro lens of the high-speed camera required manual adjustment prior to



**FIGURE 1** Schematic depiction of the flow at the SARS-CoV-2 DTSC (A) and of the multimodal acquisition system (B). Measurements were done through the opened window of the car with RGB, NIR, depth and thermal cameras as well as microphones. We recorded from the thermal camera at 50fps ((B), top center), the high framerate (HFR) camera with macro lens for face closeups at 120fps ((B), top right) and the high resolution camera in a stereo setup at 10fps ((B), top left). The stereo cameras covered the same field of view as the thermal camera. The camera setup was calibrated with a circular calibration pattern glued to a metal plate with pattern cut out as seen in the RGB (B.1), near infrared (B.2), and thermal (B.3) image. Of the three possible lanes for the swab tests, the middle one was used during the time of our measurements

each recording and had a drastically different field of view, this camera was not calibrated into the system. We applied pairwise stereo calibration with the goal of mapping from the Kinect and from the stereo pair into the thermal camera. For stereo-thermal calibration, the left camera (see Figure 1, stereo II) was used as origin each and for kinect-thermal, the kinect was the origin. The 4 k camera (stereo II) was spatially downsampled to the resolution of the low resolution right stereo camera (stereo I). Mapping into the thermal camera was implemented via the Kinect intrinsic matrix and the depth channel, as well as triangulation and projection from the stereo cameras.

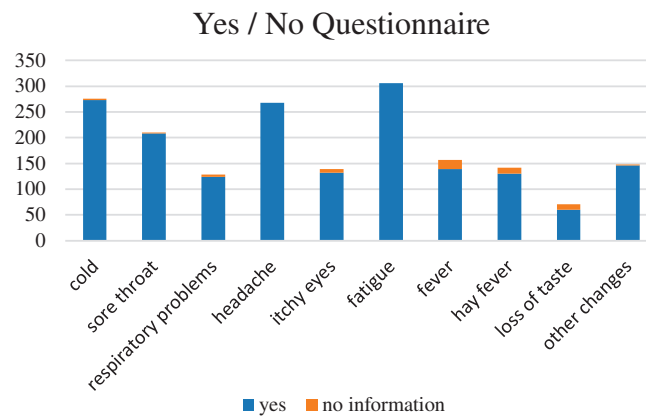
### 3.2 | Remote plethysmography

We applied the method of Wang et al. [10] for the extraction of remote plethysmography (rPPG) signals from skin segmented super-pixels of the HFR and 4 k recordings and used custom scripts for peak detection and analysis. The analysis shows that the mean heart rate (MHR) decreases with age (see Figure 2A). Female

participants show slightly higher values in MHR and lower values related to heart rate variability than male participants. This agrees with results from experiments with gold standard contact based sensors on large sample sizes [11,12].

### 3.3 | Remote eye analysis

We precomputed landmarks for the HFR and 4 k recordings using dlib [13]. On the eye, we analyzed the average blinking rate which can be a marker for drowsiness [14] and redness of the sclera as an indicator for follicular conjunctivitis [15] (see Figure 2B). We calculated the eye-aspect-ratio (Eye-AR) [16] from the pre-computed landmarks and applied custom algorithms to detect / count peaks for blinking detection. We found a significant difference in the blinking rate which was larger for participants reporting itchy eyes as compared to fever. We also extracted a region of interest around the eye and applied segmentation (gray scale based) to calculate the redness index (RI) from the sclera [15].



**FIGURE 2** Counts of the symptoms reported by participants during the yes/no section of our experiment. Because every participant was already admitted to the drive through screening centers (DTSC) based on the RKI recommendations regarding contact to infected persons and symptoms, most of our participants showed different flu-like symptoms or symptoms of a common cold. Most common were fatigue (306), followed by headache (268), cold (273), and sore throat (208). Evaluations had to exclude at most 10% of the measurements due to various reasons linked to modality. 6 subjects did not want to remove their glasses, which aggravated temperature extraction around the orbital and periorbital regions. We measured 5 symptom free subjects that did not participate in the swab test

We found no significant difference of the RI between those with fever vs. itchy eyes.

### 3.4 | Functional thermography

From the thermal recordings, we analyzed static temperatures from the orbital, periorbital, maxillary, and nose region (see Figure 2C). Vanilla landmark detectors did not perform consistently for all participants on the thermal camera, so we developed a stack of pre-processing methods such as image inversion and unsharp masking to a set of 10 randomly sampled frames for each subject, applied a facial landmark detector [17] to each of the images with different pre-processing and averaged the results per frame. Failed detections were manually annotated.

For the two demo recordings (two different recording days and system calibrations), we compared this approach with projecting landmarks from the Kinect NIR image via the Kinect depth channel as well as with triangulation from landmarks on the stereo RGB frames into the thermal camera frame. The root mean square error (RMSE) between stereo and thermal was 2.4 and 5.3 pixels and between Kinect and thermal/stereo a few magnitudes larger. The low performance of the Kinect can partially be explained by outliers due to missing depth values, which could be solved by depth map filtering. Hence, the median error is much lower with 2.7 and 4.6 for Kinect and thermal as compared to 1.8 and 2.5 for stereo and thermal. Due to the generally low RMSE between stereo and thermal and good qualitative performance of the thermal detection approach,

we used our method in the thermal domain for the static evaluations. This has the advantage that it does not depend on the calibration of the recording day and is easy to deploy.

Our results showed a significant difference in temperature for subjects reporting fever vs. no fever in the maxillary, periorbital and nose region.

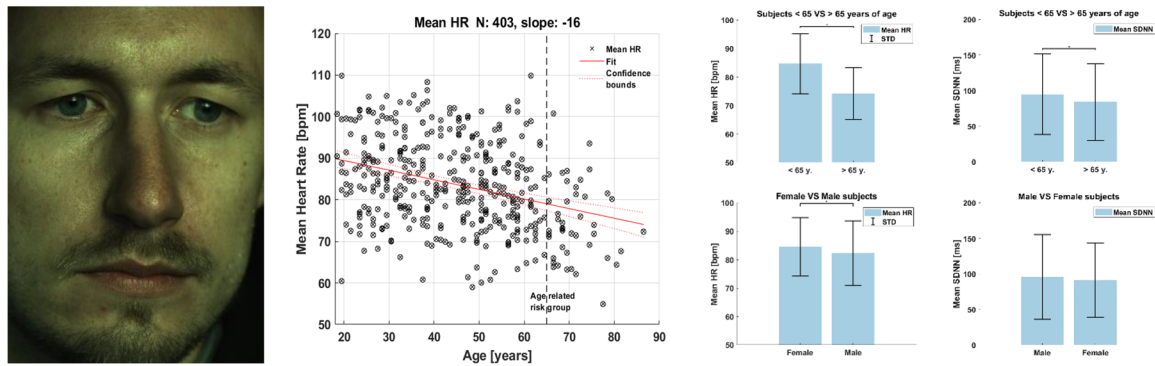
### 3.5 | Computer audition

Schuller et al. argue how audio “in-the wild” recordings under unconstrained conditions with various signal degradations can already have value for COVID-19 computer audition [2]. With our platform, we get reproducible quality audio recordings from uniform hardware (see Figure 3D). For instance, using established speech feature extraction schemes [18], the voice quality of our data is sufficient to solve a gender classification task with above 90% cross-validated accuracy of a support vector machine.

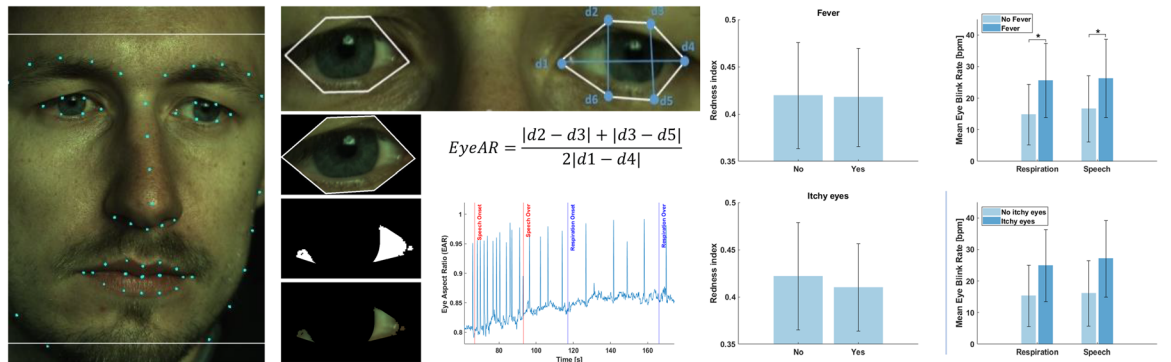
## 4 | DISCUSSION

A major limitation of our study is the marginal number of participants with positive PCR test result for SARS-CoV-2. The reason for this was the generally very low incidence rate in the study period in the region where the DTSC was located. In fact, the positive rate was below 0.4% at the DTSC Saarbrücken in the respective period. Thus, the specificity and sensitivity of the described approach with respect to SARS-CoV-2 infections cannot

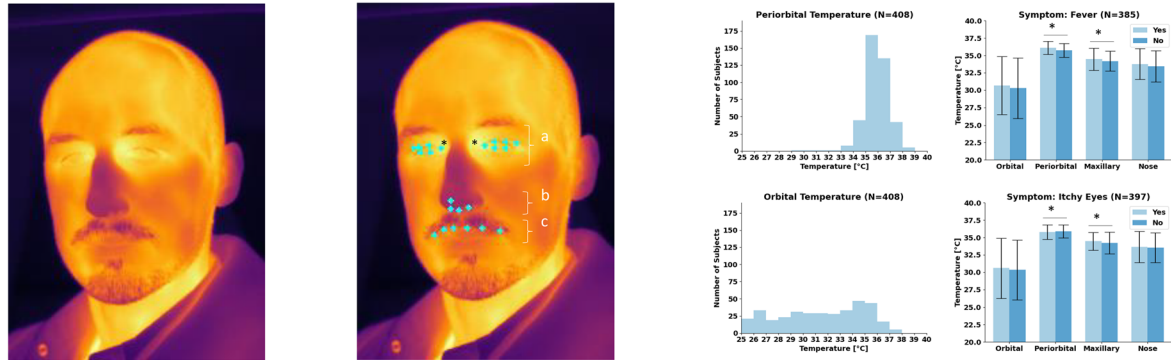
(A) Remote Plethysmography



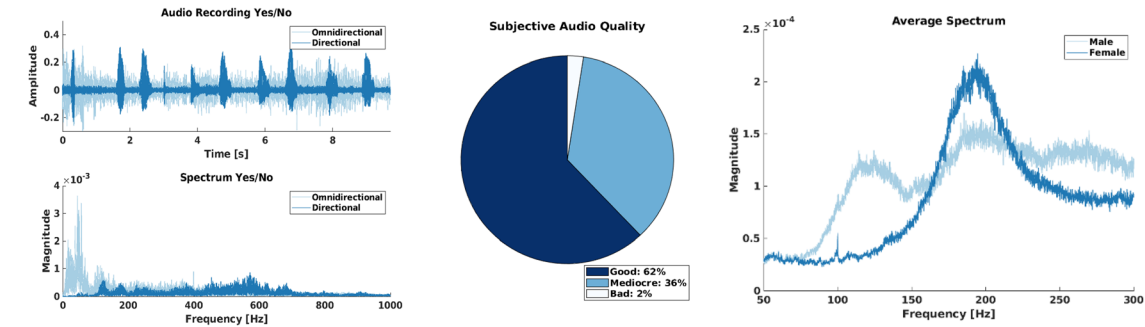
(B) Remote Eye Analysis



(C) Functional Thermography



(D) Computer Audition



**FIGURE 3** Results obtained in the reported evaluation period from the different modalities. rPPG (A) was used for the extraction of HR and HRV. For remote optometry (B) we analyzed EYE-AR and sclera redness. For static temperatures from the face (C), we looked at four points of interest. Manual assessment of our audio recordings (D) shows that 97.6% of our recordings had very good to acceptable sound quality, while for the rest the voice was rated unintelligible

be assessed. However, our proof of concept shows that a remote data acquisition of SARS-CoV-2 infection related symptoms at DTSCs is possible. Our experiences and results enable the installation of similar approaches in regions that do massive DTSC testing.

Additionally, we have recorded an unprecedented, multimodal dataset with a high number of subjects that can be used for the development and refinement of computer vision methods. Many state-of-the-art computer vision studies with isolated modalities have to resort to smaller, publicly available datasets: Considering the development of algorithms for the detection and classification of micro-expressions, Li et al. report between 80–210 subjects for common micro-expression datasets with evoked micro-expressions and they present a dataset with 20 subjects for spontaneous micro-expressions, recorded at 100fps (RGB) and 25fps (NIR) [19]. Davison et al record 32 subjects at 200fps for spontaneous micro-expressions [20]. Our dataset contains recordings of 436 participants at 120fps (RGB), 50fps (thermal) and 30fps (NIR) over 6 minutes. While some subjects moved out of frame during various parts for the experiment for the HFR face close-up recordings, we expect a similar percentage of successful recordings that allow for micro-expression annotation as for our HR evaluations (see Figure 2) and we can additionally report HFR thermal recordings with wider field of view.

For a facial landmarking task in the context of functional thermal imaging, datasets of around 2935 frames from 90 subjects with full manual annotation can be considered among the state-of-the-art [21]. Our 436 recordings at 50 hz of up to 6 minutes potentially enable the generation of a dataset with 7–8 million frames which would be multiple orders of magnitudes larger. Our preliminary results suggest the possibility of partial, automatic annotation with either with an appropriate stack of pre-processing methods and together with tracking approaches or projection from the stereo or Kinect cameras into thermal frame with any landmarking or detection algorithm could be used for a full annotation of the dataset in the future. On top of that, the setup potentially allows for multimodal mapping between NIR and thermal domain using the depth channel of the kinect and between RGB and thermal domain with the stereo setup (e.g. compare Palmero et al [22]).

Computer vision algorithms have different requirements for environmental parameters. In the context of rPPG methods, Wang et al require illumination with constant spectrum to reconstruct PPG signals from videos in talking and static scenarios and of various skin types and then achieve high signal to noise ratio of the reconstructed spectograms [10]. The employed studio illumination in our setup at the DTSC fulfills those requirements and allowed us to record high-quality data for rPPG measurements.

The redundant camera setup allows for investigations of optimal and minimal sensor configurations of similar setups: With the minimal sensor configuration of one high speed camera with macro lens and a thermal camera, the results in this paper can be reproduced. However, the macro lens required manual adjustments for each subject which aggravates calibration. Additionally, the narrow field of view makes it more likely for a participant's face to move out of the frame. Employing a Kinect that is calibrated with the thermal camera is an affordable way to project results from tracking or detection algorithms from the NIR domain to the thermal image. In terms of available algorithms and constancy assumptions, an additional pair of RGB stereo cameras forms the better solution. Using one or two high resolution cameras in the stereo setup could allow for additional analysis of skin and facial parameters.

## 5 | CONCLUSION

We have proposed a setup to record multimodal data and have recorded a unique dataset of DTSC users across all age groups. To our knowledge, this is the first time that a multimodal video and audio dataset has been recorded at a SARS-CoV-2 DTSC. We have shown in our preliminary evaluations, that the data quality is sufficient for the application of already available procedures from computer vision and computer audition. This could allow for SARS-CoV-2 infection related symptoms assessment from such data in the future. While we captured only a marginal number of PCR positive subjects, our dataset can help with the development and refinement of computer vision algorithms beyond the COVID-19 pandemic: After full annotation for various computer vision tasks such as landmarking or micro-expression analysis, it has the potential to rank among the state-of-the-art in terms of the number of participants and age statistics. It has been recorded in an out-of-lab setting with realistic participant interaction, which for example could be encountered at the entrance of an employee parking area.

## ACKNOWLEDGMENTS

We would like to thank the association of statutory health insurance physicians Saarland (KVSaar), in particular, Dr. Joachim Meiser and Michael Schneider for supporting the management with the local medical offices as well as for equipping us with PPE at the DTSC Saarbrücken. We thank the German Armed Forces, Federal State Command Saarland (Landeskommando Saarland) in particular Oberstleutnant Christoph Schacht, for their steady support during the study regarding the logistics and the installation of our data acquisition station at DTSC. We acknowledge the support of the ZF Friedrichshafen AG, in particular Florian

Dauth, Volker Wagner, and Dr. Peter Reitz for supporting us with the research vehicle used for the data acquisition. A special note of thanks goes to the responsible authorities at the Regionalverband (regional association of towns) Saarbrücken, in particular Peter Gillo and Alexander Birk and the responsible ethics commission at the medical council Saarland for helping us to resolve all the data security and ethical issues in an incredible short amount of time. Furthermore, the authors would like to thank Benedikt Buchheit, Maximilian Becker, Dr. Farah I. Corona-Strauss, Adrian Mai, Richard Morsch, Patrick Schäfer, and Elena Schneider from our research unit for supporting the administration as well as the data processing. Finally, we would like to thank Professor Alexander L. Francis from Purdue University for proving valuable feedback on the first version of this manuscript. Open Access funding enabled and organized by ProjektDEAL.

### CONFLICTS OF INTEREST

The authors declare no potential conflict of interests.

### AUTHOR CONTRIBUTIONS


Daniel J. Strauss & Philipp Flotho conceived and designed the study. Mayur J. Bhamborae, Tobias Grün, David Thinnes, and Dominik Limbach conducted experiments. Mayur J. Bhamborae, Tobias Grün implemented the acquisition system. The data were processed, analyzed, and curated by Philipp Flotho, Tobias Grün, Carlos Trenado, Mayur J. Bhamborae, David Thinnes, and Dominik Limbach. The manuscript was written by Philipp Flotho and Daniel J. Strauss.

### DATA AVAILABILITY STATEMENT

The personalized data cannot be made available. See <https://www.snnu.uni-saarland.de/covid19/> for more detailed information on data and source availability.

### ORCID

Philipp Flotho  <https://orcid.org/0000-0002-8480-0085>

Daniel J. Strauss  <https://orcid.org/0000-0001-8481-499X>

### ENDNOTE

\* [https://github.com/phflot/multimodal\\_cam\\_calib](https://github.com/phflot/multimodal_cam_calib).

### REFERENCES

- [1] A. S. Manolis, A. A. Manolis, T. A. Manolis, E. J. Apostolopoulos, D. Papatheou, H. Melita, *Trends Cardiovasc. Med.* **2020**, 30(8), 451.
- [2] B. W. Schuller, D. M. Schuller, K. Qian, J. Liu, H. Zheng, X. Li, *arXiv* **2020**, abs/2003.11117.
- [3] B. J. Quilty, S. Clifford, S. Flasche, R. M. Eggo, *Euro-surveillance* **2020**, 25(5), 2000080.
- [4] Y.-F. Tu, C.-S. Chien, A. A. Yarmishyn, Y.-Y. Lin, Y.-H. Luo, Y.-T. Lin, W.-Y. Lai, D.-M. Yang, S.-J. Chou, Y.-P. Yang, M.-L. Wang, S.-H. Chiou, *Int. J. Mol. Sci.* **2020**, 21(7), 2657.
- [5] R. Wölfel, V. M. Corman, W. Guggemos, M. Seilmaier, S. Zange, M. A. Müller, D. Niemeyer, T. C. Jones, P. Vollmar, C. Rothe, et al., *Nature* **2020**, 581(7809), 465.
- [6] W.-j. Guan, Z.-y. Ni, Y. Hu, W.-h. Liang, C.-q. Ou, J.-x. He, L. Liu, H. Shan, C.-l. Lei, D. S. C. Hui, et al., *N. Engl. J. Med.* **2020**, 382(18), 1708.
- [7] J.-P. O. Li, D. S. C. Lam, Y. Chen, D. S. W. Ting, *Br. J. Ophthalmol.* **2020**, 104(3), 297.
- [8] K. T. Kwon, J.-H. Ko, H. Shin, M. Sung, J. Y. Kim, *J. Korean Med. Sci.* **2020**, 35(11).
- [9] RKI, COVID-19-Verdacht: Maßnahmen und Testkriterien-Orientierungshilfe für Ärzte, **2020**. [https://www.rki.de/DE/Content/InfAZ/N/Neuartiges\\_Coronavirus/Massnahmen\\_Verdachtsfall\\_Infografik\\_Tab.html](https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Massnahmen_Verdachtsfall_Infografik_Tab.html).
- [10] W. Wang, A. C. den Brinker, S. Stuijk, G. de Haan, *IEEE Trans. Biomed. Eng.* **2017**, 64(7), 1479.
- [11] R. Zhao, D. Li, P. Zuo, R. Bai, Q. Zhou, J. Fan, C. Li, L. Wang, X. Yang, *Ann. Noninvasive Electrocardiol.* **2015**, 20(2), 158.
- [12] J. Koenig, J. F. Thayer, *Neurosci. Biobehav. Rev.* **2016**, 64.
- [13] D. E. King, *J. Mach. Learning Res.* **2009**, 10, 1755.
- [14] R. Martins, J. M. Carvalho in *Occupational Safety and Hygiene III*, CRC Press, **2015**.
- [15] I. Sáráncsi, D. P. Claßen, A. Astvatsatourov, O. Pfaar, L. Klimek, R. Mösges, T. M. Deserno, *Methods Inf. Med.* **2014**, 53(4), 238.
- [16] J. Cech, T. Soukupova, Center for Machine Perception, Department of Cybernetics Faculty of Electrical Engineering, Czech Technical University in Prague **2016**, 1–8.
- [17] A. Bulat, G. Tzimiropoulos, in Proceedings of the IEEE International Conference on Computer Vision, **2017**, pp. 1021–1030.
- [18] F. Eyben, M. Wöllmer, B. Schuller, MM'10 - Proceedings of the ACM Multimedia 2010 International Conference **2010**, 1459–1462.
- [19] X. Li, T. Pfister, X. Huang, G. Zhao, M. Pietikäinen, in 2013 10th IEEE International Conference and Workshops on Automatic face and gesture recognition (fg), IEEE, **2013**, pp. 1–6.
- [20] A. K. Davison, C. Lansley, N. Costen, K. Tan, M. H. Yap, *IEEE Trans. Affective Comput.* **2016**, 9(1), 116.
- [21] M. Kopaczka, R. Kolk, J. Schock, F. Burkhard, D. Merhof, *IEEE Trans. Instrum. Meas.* **2018**, 68(5), 1389.
- [22] C. Palmero, A. Clapés, C. Bahnsen, A. Møgelmoose, T. B. Moeslund, S. Escalera, *Int. J. Comput. Vision* **2016**, 118(2), 217.

**How to cite this article:** P. Flotho, M. J. Bhamborae, T. Grün, C. Trenado, D. Thinnes, D. Limbach, D. J. Strauss, *J. Biophotonics* **2021**, 14 (8), e202000512. <https://doi.org/10.1002/jbio.202000512>