Full paper for 6th Humanist Conference, The Hague, Netherlands, 13-14 June 2018

Heart Rate Analysis for Human Factors: Development and Validation of an Open Source Toolkit for Noisy Naturalistic Heart Rate Data

Van Gent, P., Delft University of Technology, Netherlands, *p.vangent@tudelft.nl*, **Farah, H.,** Delft University of Technology, Netherlands, **van Nes, N.,** SWOV, Netherlands, **van Arem, B.,** Delft University of Technology, Netherlands.

Keywords:

Human factors, heart rate analysis, physiological signals, signal analysis, open source

Abstract

Heart rate data are collected often in human factors studies. Advances in open hardware platforms and off-the-shelf photoplethysmogram (PPG) sensors allow the non-intrusive collection of heart rate data at very low cost. However, the signal is not trivial to analyse, since the morphology of PPG waveforms differs from electrocardiogram (ECG) waveforms and shows different noise patterns. PPG is often preferable because it can be collected less intrusively. However, few validated open source available algorithms exist that handle PPG data well, as most of these algorithms are specifically designed for ECG data. We have developed a novel algorithm specifically for PPG data collected in noisy field- or simulator-based settings. The main aim of this paper is to present the validation of a novel algorithm on a PPG dataset collected in a recent driving simulator experiment. The dataset was manually annotated, and performance of the algorithm compared to two other popular open source available algorithms. We show that the algorithm performs well and displays superior performance on the PPG dataset. Implications and further steps are discussed.

1. Introduction

Physiological data is collected in many human factors studies. Among the physiological data, heart rate data is usually included as it is sensitive to for example changes in (driver) workload (Mehler et al., 2010), stress (Healey et al., 2005), and general driver state such as drowsiness (Danisman et al., 2010). However, capturing heart rate in the often noisy conditions of either a driving simulator or an on-road field test, and subsequently analysing the complex signals, can be difficult or costly (Brookhuis et al., 2010). Low-cost commercial devices are available, but these are generally designed for sporting contexts and not specifically for scientific research. Furthermore, the proprietary nature of the firmware and software used in these devices creates problems with data reliability and reproducibility of results.

One potential solution lies in the recent advances in wearable technology and open hardware platforms, such as Arduino[1] and Raspberry Pi[2]. The advancement in available open hardware and software provides researchers with validated, transparent and open source data collection and processing tools. There is, however, a lack of open source available heart rate analysis algorithms that are validated, easy to use and able to handle noisy data from low-cost sensors. Implementations of heart rate analysis algorithms described in research papers are often not available, poorly documented, or require substantial technical expertise to implement properly.

Since we found the open source available heart rate analysis software unsuited for the noisy field- and simulator-based PPG data we were collecting using low-cost sensors, our aim was to develop a novel algorithm

[1] See http://www.arduino.cc

[2] See http://www.raspberrypi.org

that (i) functions better on this type of data, and (ii) provides an easy-to-use way of analysing heart rate data collected in the field or in simulators. The main aim of this paper is to describe the validation of the developed algorithm on a noisy dataset collected in a driving simulator. It was designed to be resistant to typical noise patterns (e.g. motion artefacts, momentary signal loss) of participants engaged in other tasks (driving simulator, on-road car experiment, bike experiment), to be capable of handling signals from low-cost off-the-shelf sensors, and to be user friendly. It has been designed to run on both wearable devices (Arduino, Raspberry Pi), as well as on Desktop computers. For a technical overview of the algorithm, its development and its availability, please see (van Gent et al., 2018).

In the rest of the paper, we first describe basic properties of the heart rate signal as they relate to data collection and analysis. This is followed by a discussion of our methods, results and concluding remarks.

1.1 Measuring Heart Rate in Naturalistic or Simulated Settings

There are two major approaches to measuring heart rate in (naturalistic) on-road or in simulated settings, which mainly differ in the physiological properties they measure.

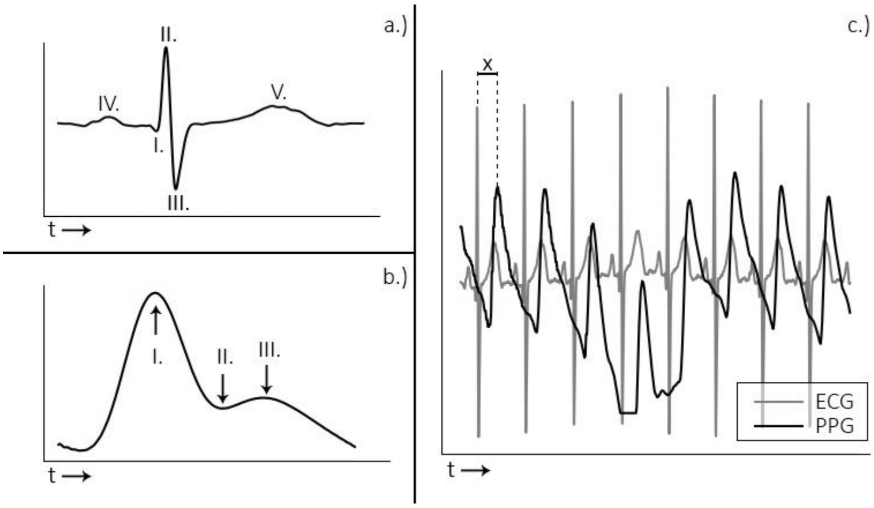

Figure 1 – The differences in morphology of the ECG wave (a) and PPG wave (b), and the time lag 'x' between both waves (c). The ECG (a) wave consists of most notably the Q-R-S complex (I-III). The P (IV) and T (V) waves are also marked in the plot.

Electrocardiogram recordings (ECG) are collected by placing electrodes on the chest near the heart. These electrodes measure the electrical activation of the heart during each cardiac cycle. The defining feature in the ECG signal is the QRS complex (Figure 1a I-III). Advantages of the ECG signal are that it directly measures the heart's electrical activation and that it presents a strong QRS complex presence in the resulting signal (Figure 1 a). A common source of noise in ECG signals are motion artefacts resulting from sensor displacement due to participant movement. These tend to fall in the same frequency range as the QRS-complexes, which can make it difficult to filter them without deforming the QRS complex (Kirst et al., 2011). In traffic related studies, ECG recordings have been used in for example (Brouwer et al., 2015; Fallahi et al., 2016; Miyaji et al., 2008).

Photoplethysmogram (PPG) recordings offer a less invasive method of assessing the cardiac cycle. These devices employ an optical sensor to measure the discoloration of the skin as blood perfuses through the arteries and capillaries with each heartbeat. PPG is typically measured at the fingertip or through wrist bracelets. The PPG signal consists of a systolic peak (Figure 1b-I), a dicrotic notch (1b-II), and a secondary peak called a diastolic peak (1b-III). The secondary peak may be absent in some recordings or of very low amplitude. Advantages of the PPG method are that it is low cost, easy to set up, and non-invasive (Elgendi, 2012; Millasseau et al., 2000). Ways of obtaining the PPG signal contactless through cameras have been proposed, further reducing intrusiveness (Sun et al., 2012). However, PPG tends to display more amplitude variation over short time-intervals (Figure 1c), more variation in waveform morphology, as well as contain more noise from various sources when compared to ECG measurements. This makes analysis more difficult. In the traffic domain PPG sensors have been used by for example (Jarvis et al., 2011; van Gent et al., in press; Zhai et al., 2006).

1.2 Analysing Heart Rate Data

The heart signal is often split into heart rate (HR) and heart rate variability (HRV) measures. These measures are calculated using the distance between the detected heart beats (the RR-intervals, named because in the ECG, the largest amplitude peak is called the R-wave). The heart beats are represented by the peaks in both signals (Figure 1a, b). Despite the different underlying physiological constructs that are measured, a high correlation (median 0.97) between RR-intervals extracted from ECG and PPG signals has been reported (Selvaraj et al., 2008). This makes the PPG a valid alternative for human factors studies that require non-intrusive heart rate measurements, given that validated analysis algorithms exist.

2. Methods

The algorithm used in this study was validated using a dataset collected with a PPG sensor in a driving simulator experiment (van Gent et al., in press). The dataset contained approximately 20.7 hours of PPG recordings.

The data were split into segments of one minute each. The R-peak positions in the segments were annotated manually to serve as a basis of comparison. We compared the algorithm performance to the annotated data on four variables: detected peak position, mean of the RR-intervals for the analysed segment, beats per minute computed by the algorithm, and a common heart rate variability (HRV) measure: the standard deviation of successive differences (SDSD). To quantify the accuracy of the algorithms predictions, we used the Root Mean Squared Error (RMSE), defined as:

$$Eq.1 \qquad RMSE = \sqrt{\frac{\sum(y - \hat{y})^2}{n}}$$

Where y is the ground truth value, ŷ is the value predicted by the algorithm, and n the number of comparisons.

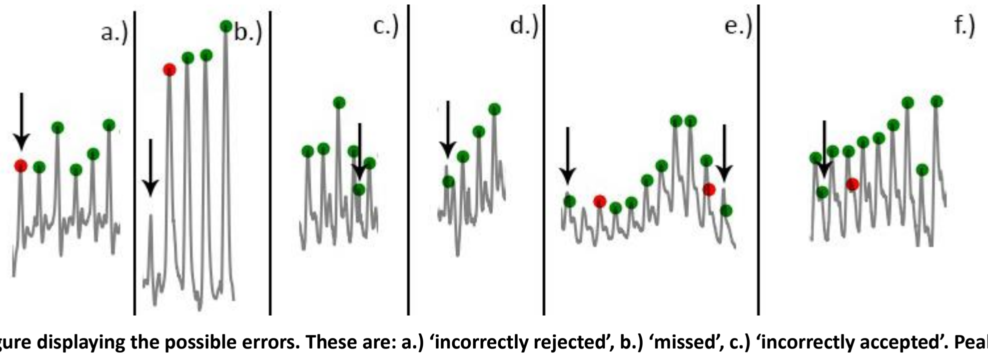

Figure 2 – Figure displaying the possible errors. These are: a.) 'incorrectly rejected', b.) 'missed', c.) 'incorrectly accepted'. Peaks marked on a correct QRS complex but not on its R-peak maximum, are also counted as 'incorrectly accepted'. This type of error is shown in d.). Other possible mistakes are marking an R-peak at a non-maximum position (e), or incorrectly marking a diastolic peak (f).

Additionally, the results of the one-minute segments were plotted and three types of errors annotated (shown in Figure 2): 'Incorrectly rejected' means a correct R-peak has been marked as low confidence (Figure 2a). 'Missed' indicates a peak is present but not marked (Figure 2b). 'Incorrectly accepted' indicates a peak is marked where no R-peak is considered present by the human annotator (Figure 2c). Figure 2d shows another example classified as 'incorrectly accepted': cases where an R-peak was marked at a non-maximum position. The HRV measures are sensitive to outliers: marking an R-peak on an anomalous position affects these measures since they rely on the intervals computed between R-peak positions. Marking a peak at an incorrect time position creates a deviation in the interval length. The algorithm has been designed to minimise the 'incorrectly accepted' error type.

3. Results

The used dataset represents 20.7 hours of PPG recordings split into 1,240 one-minute segments. The signals were recorded at the tip of the finger as participants were driving in a driving simulator. The sensor placement did not interfere with driving. Participants were instructed to drive as they normally would.

The data set was fully annotated. A total of 92,304 peaks were detected by the algorithm. Of these, 88,830 (96%) were correctly accepted, and 2,808 (3.04%) were correctly rejected automatically. This indicates that for 99.04% of all detections, the algorithm correctly labelled R-peaks. 666 (0.72%) peaks were incorrectly rejected. 229 (0.25%) R-peaks were incorrectly accepted. A total of 304 R-peaks were annotated as missed. Most of the incorrectly accepted peaks occur either because an R-peak was marked not at the maximum value (Figure 2, e), or because a diastolic (secondary) peak is marked as an R-peak (Figure 2, f). Future updates of the algorithm aim to further reduce these error rates. Overall, the error rates were low.

We compared the performance of our algorithm with an implementation of the Pan-Tompkins QRS algorithm (Pan et al., 1985), as well as with an open source algorithm HRVAS ECGViewer[3]. The latter was chosen because it is one of the first hits when searching for open source heart rate analysis software on Google, and it shows high usage statistics. It is designed for Matlab, but a standalone version is also available. The Pan Tompkins algorithm is a computationally efficient algorithm widely used in ECG analysis.

The comparison results are displayed in Table 1. They indicate that our algorithm significantly outperforms the other two open source algorithms on PPG data. The RMSE of peak position is 1.64035 (milliseconds), indicates that the standard deviation of the errors between the actual peaks and the predicted peaks was low compared to the other two algorithms. The resulting RR-intervals were also more accurate compared to the other algorithms, likely due to less missed and less incorrectly accepted beats. Differences in BPM error are not very large. Since the BPM uses the mean of all RR-intervals in a segment, it is relatively robust to a few incorrectly placed peak positions. However, effects on heart rate variability measures are large. The Standard Deviation of Successive Differences (SDSD), which is a less outlier-resistant measure for how the intervals between the heart beats vary over time, shows a large error in the other two algorithms. This shows the importance of correctly identifying R-peak positions as well as identifying incorrectly labelled peaks, as deviations risk introducing substantial error to the output measures.

[3] See: https://github.com/jramshur/ECG_Viewer

Table 1 – Comparison of algorithm performance on PPG dataset (N=1,240)

	RMSE peak location	RMSE RR-intervals	RMSE BPM	RMSE SDSD
Developed algorithm	1.64	38.87	4.18	217.41
Pan-Tompkins	16.52	171.32	4.76	364.74
HRVAS ECGViewer	16.43	272.97	6.71	1067.82

4. Discussion and Conclusion

In this paper we have described the validation of a novel, robust heart rate analysis algorithm developed for use in human factors studies. The motivation to develop such an algorithm was that what is available is often highly technical or expensive to implement, and because low-cost commercial measurement devices offer no suitable solution for scientific purposes.

We have evaluated the algorithm's performance on a manually annotated PPG data set and compared the performance to two popular available algorithms. Results showed superior performance on this type of data. It must be noted that these results reflect lower performance of the algorithms only on this specific type of data: PPG data collected in the field using low-cost sensors has quite different signal and noise properties compared to ECG data, for which many available open source algorithms are designed. The evaluation does show, however, that for many human factors studies our algorithm will outperform the other available open-source methods, especially when less intrusive measurements are desired or when low-cost sensors are used.

By offering human factors researchers an openly available and validated toolkit for heart rate analysis, we aim to increase their research possibilities, as well as the reliability and reproducibility of results obtained. Future steps include increasing the accuracy and functionality of the algorithm further.

References

Brookhuis, K. A., & de Waard, D. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. *Accident Analysis and Prevention*, *42*(3), 898–903. https://doi.org/10.1016/j.aap.2009.06.001

Brouwer, A., & Dijksterhuis, C. (2015). Physiological correlates of mental effort as manipulated through lane width during simulated driving. In *Proceedings of the 2015 International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 42–48).

Danisman, T., Bilasco, I. M., Djeraba, C., & Ihaddadene, N. (2010). Drowsy driver detection system using eye blink patterns. *2010 International Conference on Machine and Web Intelligence, ICMWI 2010 - Proceedings*, 230–233. https://doi.org/10.1109/ICMWI.2010.5648121

Elgendi, M. (2012). On the Analysis of Fingertip Photoplethysmogram Signals. *Current Cardiology Reviews*, *8*(1), 14–25. https://doi.org/10.2174/157340312801215782

Fallahi, M., Motamedzade, M., Heidarimoghadam, R., Soltanian, A. R., & Miyake, S. (2016). Effects of mental workload on physiological and subjective responses during traffic density monitoring: A field study. *Applied*

Ergonomics, *52*, 95–103. https://doi.org/10.1016/j.apergo.2015.07.009

Healey, J. A., & Picard, R. W. (2005). Detecting Stress During Real-World Driving Tasks Using Physiological Sensors. *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS*, *6*(2), 156–166.

Jarvis, J., Putze, F., Heger, D., & Schultz, T. (2011). Multimodal person independent recognition of workload related biosignal patterns. *Proceedings of the 13th International Conference on Multimodal Interfaces - ICMI '11*, 205. https://doi.org/10.1145/2070481.2070516

Kirst, M., Glauner, B., & Ottenbacher, J. (2011). Using DWT for ECG motion artifact reduction with noise-correlating signals. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 4804–4807. https://doi.org/10.1109/IEMBS.2011.6091190

Mehler, B., Reimer, B., Coughlin, J. F., & Dusek, J. a. (2010). Impact of Incremental Increases in Cognitive Workload on Physiological Arousal and Performance in Young Adult Drivers. *Transportation Research Record: Journal of the Transportation Research Board*, *2138*(1), 6–12. https://doi.org/10.3141/2138-02

Millasseau, S. C., Guigui, F. G., Kelly, R. P., Prasad, K., Cockcroft, J. R., Ritter, J. M., & Chowienczyk, P. J. (2000). Noninvasive Assessment of the Digital Volume Pulse: Comparison With the Peripheral Noninvasive Assessment of the Digital Volume Pulse Comparison With the Peripheral Pressure Pulse. *Hypertension*, *36*, 952–956. https://doi.org/10.1161/01.HYP.36.6.952

Miyaji, M., Danno, M., Kawanaka, H., & Oguri, K. (2008). Driver's cognitive distraction detection using adaboost on pattern recognition basis. *Proceedings of the 2008 IEEE International Conference on Vehicular Electronics and Safety, ICVES 2008*, 51–56. https://doi.org/10.1109/ICVES.2008.4640853

Pan, J., & Tompkins, W. J. (1985). A simple real-time QRS detection algorithm. *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, *BME-32*(3), 230–236. https://doi.org/10.1109/IEMBS.1996.647473

Selvaraj, N., Jaryal, A., Santhosh, J., Deepak, K. K., & Anand, S. (2008). Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography. *Journal of Medical Engineering and Technology*, *32*(6), 479–484. https://doi.org/10.1080/03091900701781317

Sun, Y., Hu, S., Azorin-Peris, V., Kalawsky, R., & Greenwald, S. (2012). Noncontact imaging photoplethysmography to effectively access pulse rate variability. *Journal of Biomedical Optics*, *18*(6), 61205. https://doi.org/10.1117/1.JBO.18.6.061205

van Gent, P., Farah, H., van Nes, N., & van Arem, B. (2018). Analysing Noisy Driver Physiology Real-Time Using Off-the-Shelf Sensors: Heart rate analysis software from the Taking the Fast Lane Project. *Manuscript in preparation*.

van Gent, P., Melman, T., Farah, H., van Nes, N., & van Arem, B. (in press). Multi-Level Driver Workload Prediction Using Machine Learning and Off-The-Shelf Sensors. *Transportation Research Record: Journal of the Transportation Research Board*.

Zhai, J., & Barreto, A. (2006). Stress Recognition Using Non-invasive Technology. *Proceedings of the 19th International Florida Artificial Intelligence Research Society Conference (FLAIRS).*, 395–400.