Testing Validity of Heart Rate Measuring Devices During a Sustained Attention Task to Predict Attentional Failure

Rover A. Willemars

S3323579

Department of Psychology, University of Groningen

PSB3E-BT15: Bachelor Thesis

Group 45

Supervisor: Prof. Dr. Mark Span

Second Evalutor: Prof. Dr. Dick de Waard

In collaboration with: Nienke Buist, Lisanne Zondag,

Theres Patzelt, and Harmien Tamsma

June 20th 2022

Abstract

Occupational drivers require long term sustained attention to be able to act vigilantly to stimuli in their environment. Due to the long hours that they work, fatigue has a great impact on the level of vigilance they can portray at any given moment. In previous research, it has been found that fatigue can be predicted through measures of heart rate variability, namely an increase of LF (0.04-0.15Hz) contribution in a heart rate frequency analysis (Chua et al., 2012). In the current study, we aimed to determine the validity of the Empatica E4 and Polar H10 to see if they could be used as alternatives to gold standard heart rate measuring devices, such as the TMSI REFA amplifier. This has been done by comparing the IBI's collected from participants (N=28), by the two devices, to the TMSI REFA amplifier, while the participants undergo a psychomotor vigilance task. We found that the Polar H10 was highly correlated with the TMSI REFA amplifier (r=.993, p<.001), with 97.4% of the IBI's being in a ±0.002sec of the TMSI REFA amplifier. This suggested that the Polar H10 would be valid enough to predict attentional failure. However, the Empatica E4 suffered from motion artefacts, leading to 81.2% of the IBI's not being registered. This suggested that the Empatica E4 was not sufficiently valid to predict attentional failure in occupational drivers. Limitations of the study include that HRV frequency analysis has not been compared for the devices, as the paper assumes that the IBI's have to be precise to determine valid results for the LF component of the HRV frequency analysis.

Keywords: Sustained attention, heart rate monitors, validity

Testing Validity of Heart Rate Measuring Devices During a Sustained Attention Task to Predict Attentional Failure

Addressing problems caused by fatigue is not only a concern for governmental interventions and employer responsibilities, but also the software and hardware industry (Chacon-Murguia & Prieto-Resendiz, 2015). With current body-worn technologies being more accessible to the masses due to a decrease in cost, it begs the question of how valid these devices are and what application portable physiological measuring devices can bring. One of these applications is the use of heart rate measuring devices while performing occupations that require a high state of vigilance to ensure safety. According to Chua et al. (2012), the heart rate variability derived from an electrocardiogram (ECG), can be used to predict when an individual is at risk of attentional failure. This implies that if consumer market heart rate measuring devices can accurately and precisely measure heart rate, they could be incorporated into safety devices to warn operators when their performance is impaired due to fatigue.

Furthermore, numerous occupations require the ability to sustain attention for long periods of time to avoid dangerous scenarios; this includes jobs such as long haul truck and taxi drivers. Williamson and Boufous (2007) found that work-related crashes were largely due to attentional failures caused by fatigue, with three times the amount of casualties compared to non fatigued crashes. Subsequently, according to survey data collected by Meng et al. (2015), both taxi and long haul truck drivers often indicated that they were aware of the effects that fatigue had on their driving abilities, but still decided to continue driving. Although people are often aware of when they are tired, (McDonald, 1984) there are work pressures present that may contribute to the decision to drive for longer hours. In addition, Meng et al. (2015) suggests that individuals are subject to optimism bias, where they are led to believe that the negative effects of fatigue affect others more than themselves, in essence overestimating their own abilities.

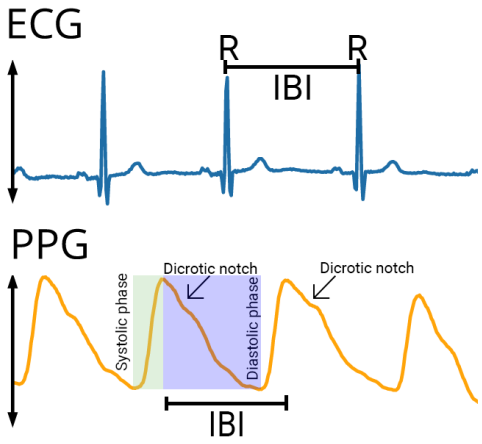
Heart Rate Measuring Devices

Previous studies have shown that the type of heart rate measuring device matters in the validity of the data. For example, a study conducted by Terbizan et al. (2002) has found that wrist-worn heart rate monitors can accurately assess heart rate, but suffered when heart rate became too high or external movement was creating interference with the sensors (Schuurmans et al., 2020; Terbizan et al., 2002; Wang et al., 2017). For situations where vigorous movements were present, Wang et al. (2017) concluded that chest-strapped heart rate monitors are more accurate than wrist-worn due to there being less artefacts in the data. However, the wrist-worn design does give the possibility to wear the heart rate monitor for longer periods of time, as the comfort of the wrist-worn device resembles that of a watch. This offers opportunities for long term heart rate monitoring, such as in the aforementioned occupations.

The current study will be examining the validity of the Empatica E4 wrist-worn and the Polar H10 chest-strapped heart rate monitors. These two devices differ in back-end technology. The Empatica E4 uses photoplethysmography (PPG) signals, which consists of an optical sensor that determines the difference in arterial volume for each heart beat to derive the R-peaks. On the other hand, the Polar H10 uses ECG signals, which measures the bio-potential generated by electrical signals that control the expansion and contraction of the heart chambers. Even though the ECG takes a more direct physiological measure of the heart, it was found that PPG data was sufficient to measure heart rate accurately when stationary (Blok et al., 2021).

Figure 1

Difference in ECG and PPG Signal



To establish the validity of these devices, the inter-beat interval (IBI) data derived

from these two devices will be compared to a reference heart rate measuring device, which will be the TMSI REFA amplifier. IBI is defined as the time between two heart beats and is known to be slightly different each time the heart beats, which is termed eart rate variability (HRV). For ECG signals, the IBI is established by measuring the time between two R-peaks. However, for PPG signals in the Empatica E4, the two consecutive points where the systolic phase ends (known as systolic peak) is used for calculating the IBI. See Figure 1 for an example of how the IBI is calculated for ECG and PPG signals. Determining HRV from a PPG signal can be challenging, as an algorithm is used to define the systolic peak as the R-peak. This comes with drawbacks, as the PPG signal is prone to movement artefacts that can make the diastolic peak (slight increase in signal right after the dicrotic notch) greater in amplitude than the systolic peak (Alqaraai et al., 2016). This would result in invalid IBI detection. In addition, according to Alqaraai et al. (2016), calculations of the HRV statistics requires precise and accurate peak detection. This is because HRV frequency analysis is sensitive to any small error in the identification of peaks.

Psychomotor Vigilance Task

To examine the validity of the heart rate measuring devices in the setting of an occupational driver, a cognitive task will be performed. The psycho-motor vigilance task (PVT) will operationalise the sustained attention that is required to avoid dangerous scenarios while driving. During the PVT, there are random interval periods (of 60-240 sec) where the participant has to be sufficiently vigilant to press a button when the target stimulus is presented. For a normal PVT task, if the participant fails to react in a specific time window (>400ms), the trial would be considered a lapse. This would indicate that the participant's vigilance state was impaired due to fatigue (Basner & Dinges, 2011), making them unsuited to operate a vehicle safely. However, as this study does not expect to include fatigued participants, the reaction time and amount of lapses will be eliminated from our research. Instead, exclusively the heart rate measures will be analysed to see if

the Polar H10 and Empatica E4 are sufficiently valid in the setting of a PVT.

As the PVT will be conducted alongside various Stroop tasks, the decision was made to use auditory target stimuli rather than visual stimuli. This was done because the performance on PVT's is greatly impacted by visual distractions (Anderson & Horne, 2006), which would be difficult to control for, as the participants would be simultaneously performing visual Stroop tasks. However, auditory PVT's have been shown to have similar effects as visual PVT's, namely that the amount of lapses that occur increase when the participant is affected by fatigue (Jung et al., 2011).

Predicting Attentional Failure

Given that fatigue has effect on the ability to sustain attention, researchers have been trying to find objective measures to predict fatigue. This includes methods that analyse the behaviour of the driver, such as: (i) lateral position on lane, (ii) steering-wheel movements, (iii) speed (iv) acceleration and brake patterns (Chacon-Murguia & Prieto-Resendiz, 2015). The drawback of these systems is that they are manipulable once the driver is aware of them. On the other hand, physiological reactions happen without conscious control, making it difficult for drivers to manipulate. Amongst others, the physiological measures that have been researched include: (i) HR, HRV and blood pressure, (ii) eye features (i.e. PERCLOS, blink rate) (iii) brain waves and (iv) skin conductance (Chacon-Murguia & Prieto-Resendiz, 2015).

In the present study, we intend to extend the findings of Chua et al. (2012), who found that attentional failure was strongly correlated with an increase in the low-frequency (LF, 0.04-0.15Hz) component of the HRV spectrum. If the Empatica E4 and the Polar H10 can precisely determine the IBI's, the HRV statistics will be valid enough to be used to predict attentional failure (Alqaraai et al., 2016; Chua et al., 2012). This resulted in our research question; are the Empatica E4 and the Polar H10 heart rate measuring devices sufficiently valid to predict attentional failure during a PVT?

Method

Participants

For this study, we recruited 49 participants (X% male, 65% female) from the University of Groningen. Out of the recruited participants, 21 were excluded, which resulted in 28 participants' data being analysed. Participants were excluded on the basis of missing data for various reasons, (i) 17 were excluded for either having a loose electrode or unusable data and (ii) 4 did not show up. The participants were from the first year of the Psychology bachelor programme and were recruited using a convenience sampling method. As compensation, the participants were rewarded 2 SONA credits, which assisted in the completion of their propaedeutic year.

Monitors and Apparatus

The collection of all streams of data was conducted by Lab Streaming Layer's (LSL) Lab Recorder. As reference heart rate monitor, the TMSI REFA amplifier (ECG) was used with a wired connection to the data acquisition computer. The software used to collect the data was OpenViBE Acquisition Server and OpenViBE Designer, which directly streamed to LSL. Furthermore, the two heart rate monitors that were compared to the reference monitor are the Polar H10 chest strap (ECG) and the Empatica E4 wristband (PPG). Both the Polar H10 and the Empatica E4 sent their data through a Bluetooth connection to the data acquisition computer. To make the Polar H10 and the Empatica E4 stream to LSL, a custom script was created by Span (2021). This resulted in three heart rate data streams in LSL. In addition to the three streams for the collection of heart rate data, two streams were used to generate event markers for the Stroop and PVT task.

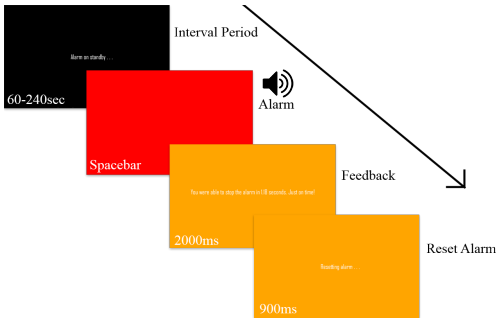
The Stroop and PVT task were created using OpenSesame 3.3.11 (Mathôt et al., 2011) using PyGame (Legacy) as backend. The participants carried out the tasks in an EEG room, in effort to reduce auditory and visual distractions. In the EEG room, there was one monitor, keyboard and button box for running the Stroop task and a laptop

(connected to speakers), placed slightly out of arms reach for the PVT task. For the Stroop task, the monitor showed the stimuli and the button box was used for user input. The button box consisted of four LED buttons that illuminated red, blue, green and yellow (from left to right). This was used to indicate the colour of the font in the Stroop task.

For the PVT task, the laptop showed a white text ("Alarm on standby...") with a black background. With random intervals between 60-240 sec, the laptop played a siren sound on the speakers and display a red screen, instigating a space-bar response from the participant on the laptop. Each time the siren went off, the pitch was increased by 10% of the original siren sample. For an example of the test sequence for the auditory PVT, see Figure 2.

Figure 2

Test Sequence of Auditory PVT



Research Design and Procedure

Initially, the participants signed up through the online SONA platform, where they were directly prompted with a screening. In the screening, participants were asked their gender, preferred language, consent, if they have a driver's license, if they had normal vision and which hand preference they had. All questions had the option to decline to answer, which did not have an effect on whether or not they could participate. In addition, the information and data gathered for each participant was treated with confidentiality, making it unrecognizable which data set was associated with which participant.

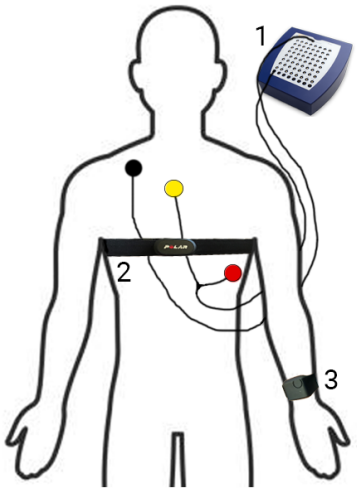
Furthermore, the experiment took place in an EEG room at the University of Groningen. Upon arrival, the participants were informed about what was expected of them during the study and a consent form was signed. In a separate room, the researcher assisted the participant with the sticking on three electrodes for the TMSI ECG (after

cleaning the skin), putting on the Polar H10 chest strap (with a bit of water on the inside of the strap) and the Empatica E4 wristband (tightly fitted, so that the sensor touches the skin). For an illustration of the simultaneous testing of the Polar H10, TMSI ECG and Empatica E4, see Figure 3.

Following, the participant was asked to sit in front of the monitor inside the EEG room and the researcher connected the three heart rate measuring devices to LSS lab recorder and started recording the data. The participant was prompted to start and the EEG room was closed. The experiment started with two conditions for the participants; standing and sitting for 60 seconds. After that, the participant completed the Stroop task, with the PVT task activating the sound stimulus every 60-240 sec. When the PVT task activated, it was the goal of the participant to press the space-bar key on the laptop as fast as possible. As the Stroop task took approximately 20 minutes, each participants went through 5-20 cycles of PVT trials. After the completion of the task, the participant was disconnected from the heart rate measuring devices and asked about their experience. Anything of interest that could have affected the data was noted down in a document. Participants were thanked for their participation and prompted to leave on their own account.

Figure 3

Simultaneous Testing of Devices



Note. (1) TMSI REFA amplifier (2) Polar-H10 (3)Empatica E4.

Results

Comparing the Polar H10 and the Empatica E4 to the gold standard of the TMSI REFA amplifier, it was found that the Polar H10 was positive correlated with the TMSI

amplifier (r=.942, p<.001, CI [.941, .943]) at a 95% confidence interval.

However, for the Empatica E4, the IBI's obtained were considered non-viable for a correlational analysis for two reasons. (1) There were too many missing values and (2) matching the IBI's between devices was impractical due to an uncertain delay in the Empatica E4 data. The missing IBI's were partially due to movement artefacts, as it seemed that the Empatica E4 only measured the IBI's when the arm was completely still. This resulted in the Empatica E4 only obtaining 7,758 IBI's (82.7% missing) compared to the 44,854 IBI's collected by the other two devices (as shown in Table 1). However, during the sitting condition, the Empatica E4 measured a missing rate of 49.5%.

Table 1

Descriptive Statistics of IBI's

	TMSI REFA Amp.	Polar H10	Empatica E4
Valid	44854	44854	7758
Missing	0	0	37096
Mean	0.700	0.700	0.728
Std. Deviation	0.107	0.107	0.135
Minimum	0.330	0.327	0.406
Maximum	2.419	1.704	1.844

Note. Includes IBI's for the whole experiment.

Following, the descriptive statistics gathered from each device resulted in identical values for the Polar H10 (M=0.700, SD=0.107) and TMSI REFA amplifier (M=0.700, SD=0.107). The Empatica E4 differed from the other two devices (M=0.728, SD=0.135).

Psychomotor Vigilance Task

To analyse the validity of the three heart rate measuring devices during the PVT trials, the IBI's that were measured in the time periods of -10,000ms to +2,000ms from the time that the each siren started, were extracted. This resulted in 154 PVT trials across 28 participants, leading to an average of 5.5 PVT trials per participant. In total, there were 2,643 IBI's collected for both the TMSI REFA amplifier and the Polar H10. For the Empatica E4, there were 497 IBI's (81.2% missing).

In the extracted PVT data, the IBI's collected by the Polar H10 were also positively correlated with the IBI's collected by the TMSI REFA amplifier (r=.993, p<.001, CI [.993, .994]) at a 95% confidence interval. The Empatica E4 suffered from the same issues mentioned prior, which led to an absence for the correlational value between the Empatica E4 and TMSI REFA amplifier. However, the descriptive statistics of the TMSI REFA amplifier (M=0.698, SD=0.106), Polar H10 (M=0.699, SD=0.106) and Empatica E4 (M=0.740, SD=0.139) will be used to identify the validity of the Empatica E4 during the PVT (see Table 2). For the Polar H10, it was found that during the PVT, 97.4% of the IBI's fell in a ±0.002sec range of the IBI measured by the TMSI REFA amplifier.

Table 2

Descriptive Statistics of IBI's During PVT

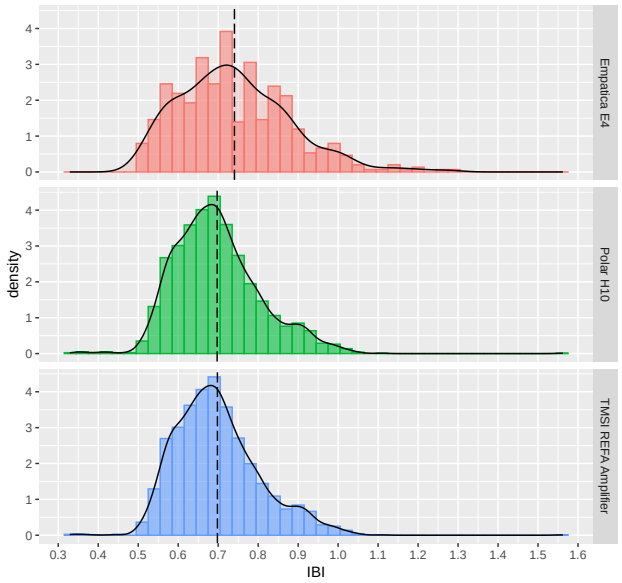
	TMSI REFA Amp.	Polar H10	Empatica E4
Valid	2643	2643	497
Missing	0	0	2146
Mean	0.698	0.699	0.740
Std. Deviation	0.106	0.106	0.139
Minimum	0.336	0.339	0.500
Maximum	1.560	1.561	1.281

Note. Includes IBI's for all PVT trials (-2,000ms - 10,000ms.)

In addition, a density plot (see Figure 4) was made to examine the relative frequency distribution of IBI's collected by each device. In the plot, the mean of the Empatica E4 deviates slightly from the other two devices. Additionally, it can be seen that the Empatica E4 has a different density curve than the other two devices. The Polar H10 and TMSI REFA amplifier are more right-skewed than the Empatica.

Figure 4

Density Plot of IBI's Split by Device



Note. The IBI's included are during the exclusively during the PVT task.

Discussion

In this study, we investigated the validity of the Polar H10 and the Empatica E4 during a PVT, by comparing the IBI values collected by each device to the TMSI REFA amplifier. This was driven by our research question; are the Empatica E4 and the Polar H10 heart rate measuring devices sufficiently valid to predict attentional failure during a PVT? To answer this question, Alqaraai et al. (2016) found that precise IBI detection is essential to determine HRV statistics, as slight errors in IBI's can lead to erroneous HRV statistics. Our research resulted in strong positive correlations between the Polar H10 and the TMSI REFA amplifier (r=.993, p<.001). This means that the IBI's collected for by the Polar H10 were significantly corresponding to the IBI's collected by the TMSI REFA Amplifier. However, a correlation could not be established for the Empatica E4 and the TMSI REFA amplifier due to there being too many missing IBI's and an unknown delay in the Empatica E4 data. This resulted in the absence of matched IBI's pairs between the Empatica E4 and the TMSI REFA amplifier.

Empatica E4

The large amount of missing data collected by the Empatica E4 during our experiment was in line with research conducted by Schuurmans et al. (2020), Van Lier et al. (2020) and Ollander et al. (2016), who resulted in 37.5%, 45% and 61.9% missing IBI's respectively. Our study resulted in 49.5% missed IBI's during the sitting task, 82.7% for the full experiment and 81.2% during the PVT tasks. The miss rate in our task trials could be higher due to the arm movements that were required to press the button box and space bar for the PVT, making it difficult for the Empatica E4 to register clear systolic peaks. That suggests that artefacts in the PPG signal are common, leading to many IBI's being missed, which would lead to uncertain HRV statistics (Alqaraai et al., 2016). This is in line with a study conducted by Schuurmans et al. (2020), where they found that the Empatica E4 was not valid enough to determine the LF component of the HRV frequency

analysis. However, they did find that the Empatica E4 was a strongly correlated with the VU-AMS while people were at rest, making it a good estimation of average heart rate. As HRV analysis requires precise IBI's and Schuurmans et al. (2020) suggests that the Empatica can be used as a good estimate of average heart rate, it shows that depending on the application, validity required from a device can differ. In the application of analysing HRV statistics, the Empatica E4 seems to be lacking. However, if the application were for measuring average heart rate for a stationary task (e.g. meditation), the Empatica E4 could suffice.

The missed IBI's in the Empatica E4 is due to the algorithm that the wrist-worn device employs. According to Schuurmans et al. (2020), the Empatica E4 algorithm only sends the IBI data when the software knows that the IBI is calculated correctly. However, even knowing that the calculations between the two systolic peaks is correct, the IBI's collected seem to be differently distributed than in the TMSI REFA amplifier. This can be seen in the density plot and descriptive statistics. This suggests that the HRV frequency analysis using the Empatica E4 would not result in identical results as the TMSI REFA amplifier (Alqaraai et al., 2016). The effect that this could have is that the Empatica E4 could determine different conclusions regarding attentional failure of the occupational driver, compared to the TMSI REFA amplifier. However, to extend this study, it would be of importance to examine the HRV frequency analysis produced by the Empatica E4 and the TMSI REFA amplifier to see if they come to the same conclusion based on Chua et al. (2012) research.

Polar H10

The Polar H10 had a strong correlation with the TMSI REFA amplifier during the full experiment (r=.942, p<.001) and during the PVT trials (r=.993, p<.001). Additionally, the mean and standard deviation are identical for the full experiment (TMSI: M=0.700, Polar M=0.700), while the mean of the IBI's collected during the PVT only

differed slightly (TMSI: M=0.698, Polar: M=0.699). Additionally, the density plots look identical, suggesting that the IBI's that are being collected by the Polar H10 are valid.

The results gathered from our experiment suggest that the Polar H10 is valid measure for HR, as the IBI's corresponded significantly to the IBI's collected from the TMSI REFA amplifier. Additionally, 97.4% of the IBI's measured by the Polar H10 fell in a range of ±0.002sec of the TMSI REFA amplifier. This is in line with research conducted by Gilgen-Ammann et al. (2019), who found that the Polar H10 was strongly correlated with a Medilog AR12plus and measured 97.1% of the IBI's within a range of 2% of the Medilog. This implies that the Polar H10 is sufficiently valid to get similar results to the TMSI REFA amplifier when HRV statistics are calculated (Alqaraai et al., 2016). However, as our experiment dos not directly measure this, no conclusion can be made regarding the effectiveness of the Polar H10 band in the setting of an occupational driver. With further research that includes HRV analysis, a better insight could be made.

Limitations

To answer if the Empatica E4 and the Polar H10 can accurately predict attentional failure in the context of an occupational driver, it would be of importance to increase the construct validity of the experiment. This could be done by including fatigued, as well as well-rested participants in the study and by including the lapse count in the data collection process. With this information, the amount of LF (0.04-0.15Hz) component of the HRV frequency analysis could be examined to view its predictive nature as well as evaluate the validity of the Polar H10 and the Empatica E4 in that context.

Lastly, the validity of the Empatica E4 was difficult to conclude, as the pairs could not be matched between the IBI's collected from the Empatica E4 and the TMSI REFA amplifier. This made it impractical to generate a correlational value for the Empatica E4, which made our conclusion less definitive.

Conclusion

Findings of the current study indicate that the Empatica E4 cannot be used in the setting of an occupational driver to predict attentional failure. In line with what other researchers have found, the Empatica E4 is subject to motion artefacts, leading to a majority of missed IBI values. However, the study does indicate that the Polar H10 can be used in the setting of an occupational driver to predict attentional failure, as the Polar H10 seems to be strongly correlated with the TMSI REFA amplifier. Although, further research would have to be conducted to see if the increase of LF power density in a HRV frequency analysis is correlated with the amount of lapses in a PVT when using the Polar H10 chest strap.

References

Alqaraai, A., Alwosheel, A., & Alasaad, A. (2016). Heart rate variability estimation in photoplethysmohraphy signals using bayesian learning approach. *Healthcare Technology Letters*, *3*(2), 136–142. https://doi.org/10.1049/htl.2016.0006

Anderson, C., & Horne, J. A. (2006). Sleepiness Enhances Distraction During a Monotonous Task. *Sleep*, *29*(4), 573–576. https://doi.org/10.1093/sleep/29.4.573

Basner, M., & Dinges, D. F. (2011). Maximizing Sensitivity of the Psychomotor Vigilance Test (PVT) to Sleep Loss. *Sleep*, *34*(5), 581–591. https://doi.org/10.1093/sleep/34.5.581

Blok, S., Piek, M., Tulevski, I., Somsen, G., & Winter, M. (2021). The accuracy of heartbeat detection using photoplethysmography technology in cardiac patients. *Journal of Electrocardiology*, *67*, 148–157. https://doi.org/https://doi.org/10.1016/j.jelectrocard.2021.06.009

Chacon-Murguia, M. I., & Prieto-Resendiz, C. (2015). Detecting driver drowsiness: A survey of system designs and technology. *IEEE Consumer Electronics Magazine*, *4*(4), 107–119. https://doi.org/10.1109/MCE.2015.2463373

Chua, E. C.-P., Tan, W.-Q., Yeo, S.-C., Lau, P., Lee, I., Mien, I. H., Puvanendran, K., & Gooley, J. J. (2012). Heart rate variability can be used to estimate sleepiness-related decrements in psychomotor vigilance during total sleep deprivation. *Sleep*, *35*(3), 325–334. https://doi.org/10.5665/sleep.1688

Gilgen-Ammann, R., Schweizer, T., & Wyss, T. (2019). Rr interval signal quality of a heart rate monitor and an ecg holter at rest and during exercise. *European Journal of Applied Physiology*, *119*(7), 1525. https://doi.org/10.1007/s00421-019-04142-5

Jung, C. M., Ronda, J. M., Czeisler, C. A., & Wright JR, K. P. (2011). Comparison of sustained attention assessed by auditory and visual psychomotor vigilance tasks prior to and during sleep deprivation. *Journal of Sleep Research*, *20*(2), 348–355. https://doi.org/https://doi.org/10.1111/j.1365-2869.2010.00877.x

Mathôt, S., Schreij, D., & Theeuws, J. (2011). Opensesame: An open-source, graphical experiment builder for social sciences. https://doi.org/https://doi.org/10.3758/s13428-011-0168-7

McDonald, N. (1984). *Fatigue, safety, and the truck driver*. Taylor; Francis.

Meng, F., Li, S., Cao, L., Li, M., Peng, Q., Wang, C., & Zhang, W. (2015). Driving fatigue in professional drivers: A survey of truck and taxi drivers. *TRAFFIC INJURY PREVENTION*, *16*(5), 474–483. https://doi.org/10.1080/15389588.2014.973945

Ollander, S., Godin, C., Campagne, A., & Charbonnier, S. (2016). A comparison of wearable and stationary sensors for stress detection, 004362–004366. https://doi.org/10.1109/SMC.2016.7844917

Schuurmans, A., de Looff, P., Nijhof, K. S., Rosada, C., Scholte, R. H. J., Popma, A., & Otten, R. (2020). Validty of the empatica e4 wristband to measure heart rate variability (hrv) parameters: A comparison to electrocardiography (ecg). *Journal of Medical Systems*, *44*(190). https://doi.org/10.1007/s10916-020-01648-w

Span, M. (2021). Polarband2lsl.

Terbizan, D. J., Dolezal, B. A., & Albano, C. (2002). Validity of seven commercially available heart rate monitors. *Measurement in Physical Education Exercise Science*, *6*(4), 243–247. http://search.ebscohost.com.proxy-ub.rug.nl/login.aspx?direct=true&db=aph&AN=8742003&site=ehost-live&scope=site

van Lier, H. G., Pieterse, M. E., Garde, A., Postel, M. G., de Haan, H. A., Vollenbroek-Hutten, M. M. R., Schraagen, J. M., & Noordzij, M. L. (2020). A standardized validity assessment protocol for physiological signals from wearable technology: Methodological underpinnings and an application to the e4 biosensor. *Behavior Research Methods*, *52*(2), 607–629. https://doi.org/10.3758/s13428-019-01263-9

Wang, R., Blackburn, G., Desai, M., Phelan, D., Gillinov, L., Houghtaling, P., & Gillinov, M. (2017). Accuracy of Wrist-Worn Heart Rate Monitors. *JAMA Cardiology*, *2*(1), 104–106. https://doi.org/10.1001/jamacardio.2016.3340

Williamson, A., & Boufous, S. (2007). A data-matching study of the role of fatigue in work-related crashes. *Transportation Research Part F: Traffic Psychology and Behaviour*, *10*(3), 242–253. https://doi.org/https://doi.org/10.1016/j.trf.2006.10.002