dHealth 2022
G. Schreier et al. (Eds.)

doi:10.3233/SHTI220370

Accuracy of Wearable Photoplethysmography Sensors for Continuous Heart Rate Monitoring in Telehealth Applications

Hannah VINATZER[a,b] , Angelika RZEPKA[a,1] , Dieter HAYN[a,c], Andreas ZIEGL[a,b], Kurt EDEGGER[a], Sandra PRESCHER[d] and Günter SCHREIER[a,b]
[a] *AIT Austrian Institute of Technology GmbH, Graz, Austria*
[b] *Graz University of Technology, Graz, Austria*
[c] *Ludwig Boltzmann Institute for Digital Health and Prevention, Salzburg, Austria*
[d] *Charité –Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt Universität zu Berlin, Medical Department, Division of Cardiology and Angiology, Centre for cardiovascular Telemedicine, Berlin, Germany*

Abstract. The demand for extended care for people suffering from heart failure is omnipresent. Wearables providing continuous heart rate measurement through optical sensors are of great interest due to their ease of use without the need for medical staff and their low cost. In this study, seven wearables were tested in fifteen measurement runs, with a duration of fourteen-hour each, and compared to a reference sensor. By calculating the Pearson correlation and the root mean square error, as well as the graphical representation by a Bland Altman plot, it was found that these wearables lack sufficient accuracy and may not be suitable for medical purposes.

Keywords. heart failure, healthcare monitoring, photoplethysmography, wearable device, mHealth.

1. Introduction

Heart failure is a chronic disease in which the heart pumps an insufficient amount of blood through the human body. As a result, the organs can no longer be supplied with sufficient blood, leading to an undersupply of the organs. One to two percent of the total population in Austria suffers from heart failure [1]. The Institution *Statistik Austria* recorded a total of 21.644 hospitalizations due to heart failure for the year 2020 in Austria [2]. Up to 50 % of patients hospitalized for heart failure die or require hospital readmission by the end of 6 months [3]. To be able to counteract this disease, early detection and close cooperation between the patients and their respective caregivers are the main approaches [1].

In addition to standard care, telemedicine can be used to support this dialogue also outside of the in-clinic environment. Here, the patients regularly measure their vital

[1] Corresponding Author: Angelika Rzepka, AIT Austrian Institute of Technology, Graz, Austria, E-Mail: angelika.rzepka@ait.ac.at

parameters, such as blood pressure, heart rate or body weight by themselves. These parameters are monitored to detect cardiac decompensation at an early stage to set appropriate therapeutic steps by the responsible health care professionals. Among others, heart rate serves as a valuable indicator for the further course of the state of health. With an increased heart rate, it was observed that the total mortality and the mortality due to pump failure of the heart increased. By lowering the heart rate to a normal value, an improvement in the survival rate could accordingly be achieved [4],[5].

This aspect was already discussed in previous studies which all concluded that the development of telehealth systems for patients with heart failure is of great importance and should play an important role in the outpatient treatment of those patients [6]-[9]. For instance, the TIM-HF2 study, conducted in 2018, showed that a reduction in mortality after hospitalization could be achieved through telemedical co-management of heart failure patients [5]. So far, heart rate monitoring was done either only during blood pressure measurements or ECG based with specific event recorders or belts.

Photoplethysmography (PPG) sensors are widely used in clinical applications and in commercially available medical devices. By optical measurement techniques, the volume change of the blood in the tissue can be determined only with a few components such as a light emitting diode (LED) and a detector catching the transmitted light beam from the tissue [10].

PPG sensors are quite sensitive to motion. Three types of movement artefacts are described in literature: the change in tissue caused by movement, the movement of the sensor on the skin and changes in the pressure of the sensor lying on the skin [11].

Currently, the manufacturers of such wearables, designed for everyday life, that measure heart rate using PPG sensors emphasise that their products are not classified as medical devices and are only intended for personal use. Another important factor is that in Europe and elsewhere, medical devices are subject to strict regulation and that the transfer and sale of the recorded medical data to third parties is subject to privacy regulations [12].

In this study, long-term monitoring of heart rate was used to investigate whether wearable photoplethysmography sensors provide reliable values in everyday life so that they can be used for telemedical applications.

2. Methods

2.1. Hardware

The wearables were selected according to the following criteria: Accuracy, battery life and usability. Seven wearables and one reference sensor were selected based on an internet search, all of which are PPG sensors. The selected wearables are listed in Table 1.

Table 1: Listing of wearables, manufacturers, and acronym. Bold: reference sensor

Wearable product name	***Manufacturer***	***Acronym***
Apple Watch Series 6	Apple	AW
Beurer AS97	Beurer	AS97
Vivoactive 4	Garmin	V4
Vantage M	Polar	VM
Verity Sense	***Polar***	***VS***
Rhythm+	Scosche	R+
Rhythm 24	Scosche	R24
Scanwatch	Withings	SW

2.2. Software

To obtain the data from the various wearables, different software had to be used, most of which was provided by the manufacturers themselves. The following software (available version from June 2021) was used for the individual wearables:

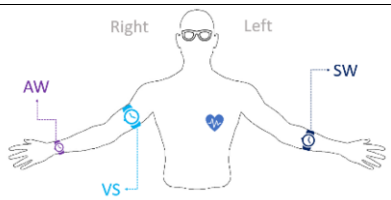
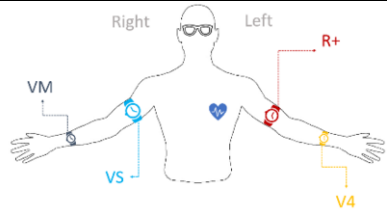
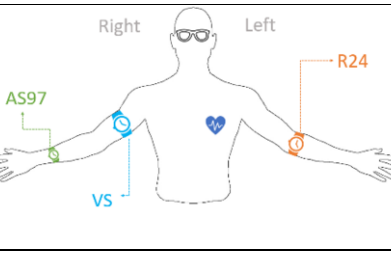
- *Health (v. iOS 12.5.3)* from Apple for the AW and both Scosche devices
- *HerzMobil App (v. 1.72)* and *TeleReha (v. 1.5)* software by AIT for the AS97
- *Garmin Connect (v. 4.43)* by Garmin for the V4
- *Polar Flow (v. 5.3.0)* from Polar for the VS and VM
- *Health Mate (v. 5.5.1)* from Withings for the SW

For analysing the data MATLAB and python scripts were implemented.

2.3. Measurement protocol

In addition to the handling of the devices, the manufacturers also provided information on how to wear the devices to achieve the best possible results. To realise this, the wearables were divided into three groups: measurement variant A, B and C, each containing the reference and two or three additional sensors, as described in Table 2.

Table 2: Overview of the types of measurement performed. Left: wearables and selected mode. Right: graphical representation of the position of the wearables.

Measurement variant A - Reference sensor: VS, in recording mode - Wearable 1: AW default mode (13 measurements) and activity mode (2 measurements) - Wearable 2: SW, in activity mode	
Measurement variant B - Reference sensor: VS, in recording mode - Wearable 3: VM, in activity mode - Wearable 4: R+, in recording mode - Wearable 5: V4, in activity mode	
Measurement variant C - Reference sensor: VS, in recording mode - Wearable 6: AS97, in recording mode - Wearable 7: R24, in recording mode	

The VS proved to be sufficiently reliable when compared to ECG based sensors in previous studies [13], this sensor was used for recording a reference heart rate.

Five test subjects, four females and one male, who did not suffer from heart failure, aged between 20 and 24 years, were provided with the wearables and underwent three measurement runs of 14 hours each for every measurement variant. The test subjects were left unsupervised during the measurements. Detailed information on the operation and handling of the wearables was provide at the beginning of the measurements. The test subjects underwent their daily activities during the 14-hour measurement period.

2.4. Evaluation

All data recorded according to the protocol described in chapter 2.3 were pre-processed in order to:

a) remove obvious outliers (heart rate > 250bpm or heart rate < 20bpm) and
b) to eliminate potential time shifts in between the recorded signals.

To compare each wearable to the reference sensor, the mean and standard deviation of the Pearson correlation (CORR) and the root mean square error (RMSE) were calculated for each wearable which was tested.

To visualize the data, a Bland Altman plot was created for each wearable. All data of the individual measurements were displayed in one plot for each wearable.

3. Results

A total of 13 out of 15 measurements for measurement A, 12 out of 15 measurements for parameter B and 13 out of 15 measurements for parameter C were successfully performed. The remaining measurements needed to be excluded due to failed data transmissions, so that no collected data was available to analyze.

Table 3 summarizes the calculated mean and standard deviation of the RMSE and the CORR for each wearable. The Bland Altman plots for each wearable are shown in Figure 1.

Table 3: Root mean square error (RMSE) and correlation (CORR) between sensor data and reference data for each wearable (mean ± standard deviation).

Sensor	RMSE [bpm]	CORR
VM	2.50±3.54	0.95±0.11
AW	5.51±2.96	0.94±0.07
R24	5.70±2.37	0.91±0.05
R+	6.30±2.36	0.88±0.11
V4	7.69±2.96	0.83±0.14
SW	9.24±2.39	0.79±0.11
AS97	11.09±1.68	0.73±0.12

4. Discussion

From the calculated parameters in Table 3 it can be seen that the lowest RMSE (2.5 bpm) and the highest CORR (0.95) were obtained for the VM, followed by the AW. The AS97, on the other hand, performed worst with a mean RMSE of 11.09 bpm and a correlation coefficient of 0.73.

The VM and the reference sensor stem from the same manufacturer. Therefore, they might apply similar pre-processing algorithms on the data, including outlier removal, moving averaging, etc. This might have been a benefit for the VM. Therefore, although the VM was clearly one of the most accurate sensors in our study, it is hard to say whether the VM actually performs better than the AW or even the R24.

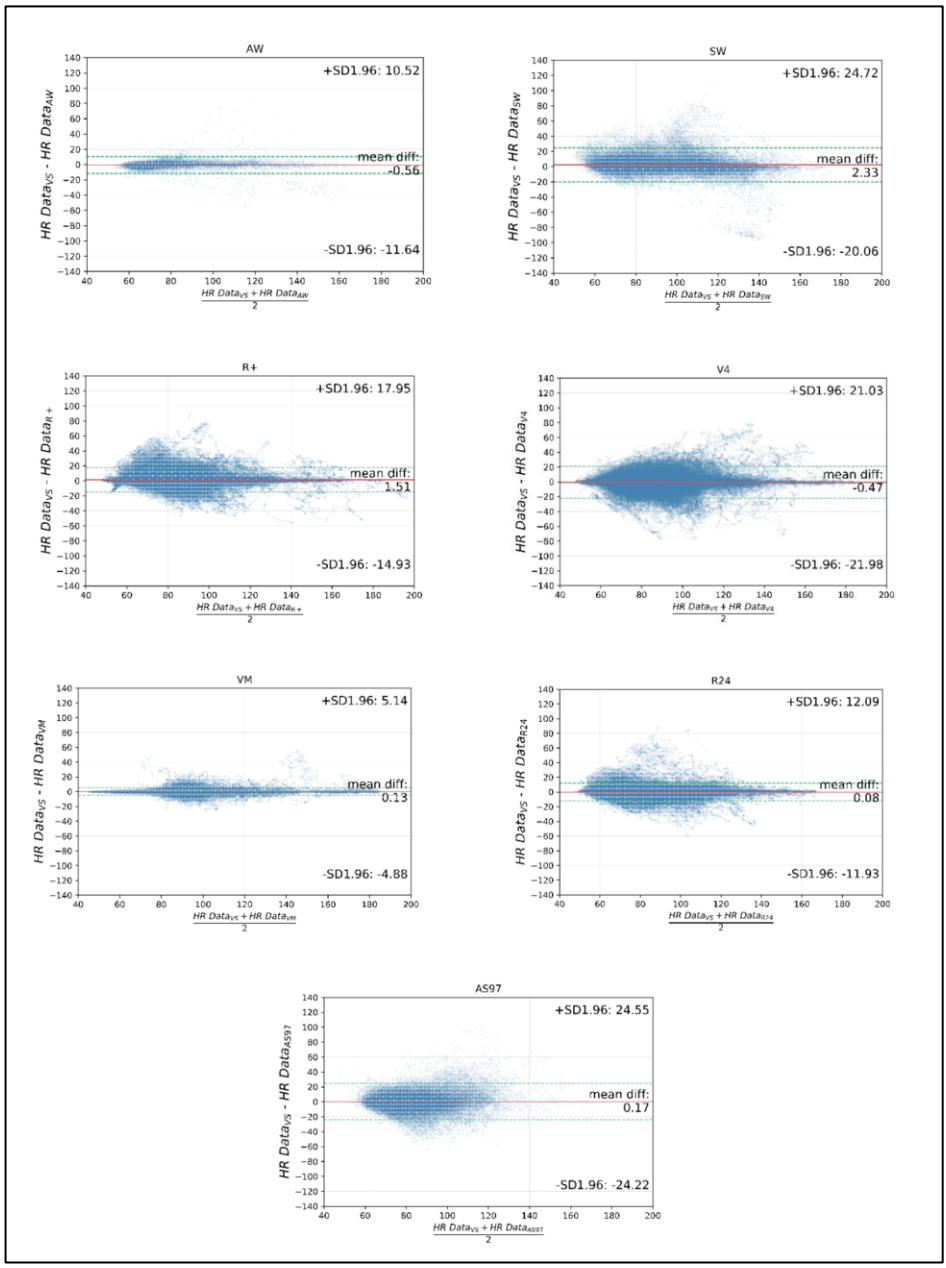


Figure 1: Bland Altman plots for the individual wearables. Green lines mark the confidence interval. Red lines mark the mean deviation. All values displayed in bpm.

The Bland Altman plots (Figure 1) reveal that most sensors over- or underestimates the heart rate in certain situations. E.g., the AW seems to generally underestimate higher heart rates (approx. >130 bpm). For the SW, we identified a region at approx. 100 bpm where the SW overestimates the heart rate and a region at approx. 140 bpm where underestimation was observed. Such regions may stem from double detection of single heart beats or from detection of only every second or every third beat.

Lines in the Bland Altman plots (e.g., for R+ or V4) typically stem from single measurements and may be due to different averaging of the watches: e.g., if the actual heartrate changes from 100 to 140 bpm in an almost stepwise way, due to different moving averaging methods, the reference sensor and the tested sensor typically provide a smoothed heart rate change and differences in the time constants of these step responses may lead to lines as seen in Figure 1.

The confidence intervals of the different sensors were identified in the range of approx. 5 bpm (VM) and 25 bpm (AS97). Deviations of more than 15 to 20 bpm are likely to be unacceptable for most medical application scenarios. 5 bpm, on the other hand, may be sufficient for selected applications. However, even for the VM and the AW, there were sequencies when the sensors significantly differed from the reference sensor.

It is not only essential to know the accuracy of a specific sensor, but also, a) in which setting the sensor is intended to be used (e.g., at rest or during intensive activity), b) whether short term effects or mean values (e.g., over days are considered), and c) which decisions should derived from the sensor data (e.g., change of medication regime vs. classify a patient as rather active or rather inactive).

The parameters obtained in this study should be treated with caution, as even the signals collected from the reference sensor may be susceptible to signal errors. We have manually inspected the reference signal and removed obvious outliers. We assume that the reference sensor provides reliable results, based on comparisons of the reference sensor with an ECG-based recording. However, further studies with an ECG based reference sensor may be necessary, depending on the respective research questions. Nevertheless, for the purpose of the present paper, i.e., to provide insights on the application of photoplethysmography sensors in telehealth scenarios, we assume that the accuracy of the reference sensor was sufficient.

To use wearables for clinical purpose, they must fulfill the requirements of a medical device, which was not the case for the ones used in the present study. Nevertheless, uncertified devices are of interest due to their convenient handling and lower costs. Depending on the application scenario, even uncertified devices may of course be used by patients monitored in telehealth systems (e.g., if they use their fitness tracker when going for a walk). However, for integrating such devices in a telehealth infrastructure, regulatory requirements need to be considered.

It is difficult to find comparable results in other studies for the chosen wearables, as the investigations mainly take place within a precisely defined activity-setting and not in everyday activities as in this study.

Future work will include analysis of subsets of data, e.g., at rest, during medium and during intensive physical activity. So far, the sensors were tested by young healthy subjects. Prior implementation in a telehealth setting, an evaluation with the target group, i.e., heart failure patients, might be required, as well as a larger group of test subjects. Additionally, the effect of potential artifacts of the reference sensor on the results, e.g., by comparison with an ECG based reference sensor, might provide further insights. Since new sensors become available each year, the present study might be repeated with new sensors and at regular intervals in the future.

5. Conclusion

Photoplethysmography sensors provide a user-friendly way to monitor heart rate especially during everyday life. Our results indicate that the reliability of selected state-of-the-art sensors is quite high. However, for each of the sensors we identified sequences with significant deviations from the reference signal. Whether a sensor is accurate enough or not for integration in a telehealth scenario depends on the detailed application scenario.

Acknowledgment

This work is part of the project "Telemed5000" (https://www.telemed5000.de) and was supported by the Federal Ministry for Economic Affairs and Energy on the basis of a decision by the German Bundestag and Austrian Research Promotion Agency (FFG).

References

[1] J. Zirkl. "Herzinsuffizienz (Hl) schwerwiegende Erkrankung im Vormarsch - Österreichischer Herzverband - Landesverband Steiermark". (2021). Online: https://www.herzverband-stmk.at/herzinsuffizienz-hl-schwerwiegende-erkrankung-im-vormarsch

[2] Statistik Austria, Hospital discharges according to selected diagnoses Online: http://statistik.at/web_de/statistiken/menschen_und_gesellschaft/gesundheit/stationaere_aufenthalte/spitalsentlassungen_nach_ausgewaehlten_diagnosen/index.html

[3] G. Hasenfuß, F.Edelmann und R. Wachter. "*Empfehlung zur Herzinsuffizienz*" in: Internist vol.**54**(2013). Online: https://doi.org/10.1007/s00108-013-3312-x

[4] J. R. Kapoor and P. A. Heidenreich. "Role of Heart Rate as a Marker and Mediator of Poor Outcome for Patients with Heart Failure." in: *Current Heart Failure Reports* vol.9.2 (2012). Online: 10.1007/ s11897-012-0086-8

[5] F. Köhler, S. Prescher, and K. Köhler. "Telemedizin bei Herzinsuffizienz." in: *Der Internist* vol.**60**.4 (2019), pp. 331–338. Online: 10.1007/ s00108-019-0570-2

[6] B. Munos, P. C. Baker, B. M. Bot, M. Crouthamel, G. de Vries, I.Ferguson, J. D. Hixson, L. A. Malek, J. J. Mastrototaro, V. Misra, A.Ozcan, L. Sacks, P. Wang "Mobile health: the power of wearables,sensors, and apps to transform clinical trials.", in: *Annals of the New York Academy of Sciences* vol. **1375**.1, (2016). Online: doi.org/10.1111/nyas.13117

[7] A. Mizuno, S. Changolkar, M. S. Patel, "Wearable Devices to Monitor and Reduce the Risk of Cardiovascular Disease: Evidence and Opportunities", in: *Annual Review of Medicine* vol.**72**:459-471, (2020). Online: doi.org/10.1146/annurev-med-050919-031534

[8] F. Freitas, F. S. Silveira, G. E. Conceição-Souza, M.F. Canesin, P. V. Schwartzmann, S. Bernardez-Pereira, R. B. Bestetti, "Emerging Topics in Heart Failure: The Future of Heart Failure: Telemonitoring, Wearables, Artificial Intelligence and Learning in the Post-PandemicEra.", in: *Arquivos Brasileiros de Cardiologia* vol.**115**, (2020). Online: doi.org/10.36660/abc.20201127

[9] SC. Inglis, RA. Clark, FA. McAlister, S. Stewart, JGF. Cleland. "Which components of heart failure programs are effective?", in: *European Journal of Heart Failure* vol.**13**(9), (2011). Online: doi: 10.1093/eurjhf/hfr039.

[10] J. Allen. "Photoplethysmography and its application in clinical physiological measurement." in: *Physiological Measurement* vol. **28**.3, (2007). Online: 10.1088/0967-3334/28/3/r01 (cit. on pp. 3, 6).

[11] E. Sazonov, *"Wearable sensors: fundamentals, implementation and applications 2nd Edition"*. Academic Press (2020) ISBN: 9780128192467

[12] N.Jiangn, J.E. Mück, A.K. Yetsien, "The Regulation of Wearable Medical Devices" in: *Science&Society* vol. **38**, (2020). Online: https://doi.org/10.1016/j.tibtech.2019.06.004

[13] M. Kim, J. Kim, K. Park, H. Kim and D. Yoon, "Comparison of Wristband Type Devices to Measure Heart Rate Variability for Mental Stress Assessment" in *2021 International Conference on Information and Communication Technology Convergence (ICTC)*, 2021, [Online]: 10.1109/ICTC52510.2021.9620772.