

Considerations for Using Fitness Trackers in Psychophysiology Research

Lauren Reinerman-Jones ^(✉), Jonathan Harris, and Andrew Watson

University of Central Florida Institute for Simulation and Training Prodigy Laboratory,
3100 Technology Parkway, Orlando, FL 32826, USA
lreinerm@ist.ucf.edu

Abstract. Wrist worn fitness trackers have become ubiquitous in recent years. Economies of scale have drastically reduced the cost of these devices while concurrent advances in technology have expanded their physiological recording capabilities. These devices now contain numerous sensors capable of monitoring and collecting various physiological attributes. Additionally, some of these devices provide access to the application programming interface (API), allowing researchers direct access to the data. The use of these devices offers a wide-ranging benefit to the scientific research community. However, there are several factors to consider when selecting a fitness tracker for use in research. Data rights, data protection, and data quality are all important considerations that must be addressed. In addition, other factors, such as sensor types, capabilities, and sampling rates, can directly affect the utility of a wearable device for use in research. In this paper, the Microsoft Band 2 fitness tracker was selected to evaluate participant mental workload during task performance in a simulated nuclear power plant (NPP) Main Control Room (MCR) as well as training effectiveness in UH-60A/L simulated environments. The Microsoft Band 2 fitness tracker was selected specifically for its optical heart rate sensor, API access to RR intervals (interval between two continuous heartbeats), and direct access to real-time streaming data from the device. To validate the utility of using the Microsoft Band 2 fitness trackers in scientific research, the RR interval and heart rate sensor readings need to be directly compared to FDA medical approved sensor readings. This paper discusses considerations when using a fitness tracker for psychophysiology research and compares data collected from the Microsoft Band 2 to two different FDA approved medical grade ECG devices.

Keywords: Fitness tracker · Wearables · Psychophysiology · Sensors · Optical heart rate · Photoplethysmography · Microsoft band

1 Introduction

Since the early 1990s on-body computer sensors have been used in research [1]. Early on-body devices used in research were constructed of standard computer components [2]. Recent improvements in sensor technology and manufacturing processes has enabled these devices to shrink in size and weight. This has allowed wearable devices to enter the consumer market. The consumer market for wearable devices has exploded

in recent years and is expected to reach 245 million units sold by 2019 [3]. The term wearable is a widely used term with various definitions. This paper adopts the simple definition that a wearable device is a computer or electronic technology that can be incorporated into clothing or accessories, and be worn comfortably on the body [4]. This definition differentiates wearables from dedicated sensor devices that are worn on-body since a dedicated device is not incorporated into clothes or used as an accessory. Sarah Wilson from CCS Insight estimates that the wearables market will be \$25 billion by 2019 [3]. Fitness trackers are a subset of wearables and account for a substantial share of the wearable market. PC Magazine defines fitness trackers as a wrist-worn device that detects some combination of steps, run distances, heart rate, and/or sleep patterns [5]. By the end of 2015 nearly 33 million fitness trackers had been sold [6]. Fitness trackers come packed with various sensors that monitor the health of the human wearing the device. While mainly developed for health monitoring, fitness trackers are ideal for performing psychophysiology research. However, as with any technology, before incorporating fitness trackers into scientific research, it is important to understand their capabilities and limitations. This paper covers the process that was followed for selecting the most suitable fitness tracker for two field research studies conducted by the Prodigy laboratory at the Institute for Simulation and Training, University of Central Florida. The first study investigated participants' mental workload demands when performing operating procedure tasks in a simulated nuclear power plant main control room environment. The second study conducted a training effectiveness evaluation with UH-60A/L pilots in virtual environments by looking at task engagement.

2 Fitness Tracker Sensors

Fitness trackers provide a low cost and minimally invasive way of sampling certain physiological responses to stimuli [1]. Psychophysiology research is classified as any research where the dependent variable is a physiological measure, and independent variables are behavioral or mental [7]. While sensors in fitness trackers are an ideal psychophysiological instrument for measuring responses to experimental stimuli, the kind of data and its quality depend on the wearable's capabilities. Most commercial fitness trackers such as the Jawbone 4, Fitbit Surge, Garmin vívoactive, and Microsoft Band 2 have accelerometers and gyroscopes to track physical activity, motion, and sleep. The inertial sensors in these devices typically have variable or user configurable sampling rates. In the case of the Microsoft Band, the data stream has three supported sampling frequencies (8/31/62 Hz) for its Accelerometer and Gyroscope [8]. The Microsoft Band 2 also includes digital GPS, barometer, and UVI (ultra violet index) sensors [9]. The data provided by these additional sensors are useful for outdoor operations where environmental conditions have significant impact on the user's vitals.

There are several other common wearable sensors available on consumer fitness trackers. These can monitor temperature (skin and ambient) as well heartbeat. The Microsoft Band 2 uses a more commonly found optical heart-rate sensor. The sensor shines an LED through the skin and measures the light that is reflected back to the sensor. The more blood in a vessel, the more the light gets absorbed. A process called

photoplethysmography is used to translate fluctuations in light to heart beats [10]. Another commonly found physiological sensor is the Galvanic Skin Response (GSR) sensor. This instrument measures electrodermal activity or skin conductance. GSR sensors are commonly implemented as a set of conductive probes that measure the resistance across an area of the user's skin [11].

Wearables like the Myo Band contain more exotic sensors found outside of the consumer fitness market. These devices can measure blood pressure, respiratory rate, and even muscle activity (Electromyography).

3 Data Considerations

It is important to consider the legal aspects of data rights when using wearables. When considering a device, it is important to review the legal documents, such as the end user license agreement (EULA) and product safety guides. The EULA will explain data rights and ownership for the wearable. This is particularly important for a study because the organization executing data collection and/or the agency funding the data collection will want to own all of the data for analyses and reporting.

Related, is data storage and retrieval. Reviewing the product website for other relevant information on the wearable that may affect its utility for a study and FAQs pages are good sources to learn about a wearable's capabilities, data formats, and methods for accessing sensor data. There are two approaches for acquiring the sensor data: 1. access through a web interface or 2. access directly from the device.

Most wearable devices provide cloud based services for synchronizing sensor data to the web. For cloud-based storage, an internet connection is required; this is most often done through a tethered phone or PC connected to the internet. Sensor data stored on the web increases the risk of unauthorized persons accessing sensitive data and anonymizing data becomes important. Another issue to consider when accessing data from the web is how much of a delay uploading adds to the data collection process, which potentially affects the data's value as it pertains to direct connection of the physiological response to the stimulus. Cloud based access is not suitable for real-time closed loop research.

A few wearable devices provide local access to sensor data. These wearable devices provide access to sensor data without requiring an internet connection. This is accomplished by synchronizing a computer with the device's local storage and/or real time sensor streaming over a wireless protocol such as Bluetooth. Whichever method used for acquiring data, it is important to understand data formats, sample rates, and sensor reporting units since there are no industry standards for wearables [12].

4 Selecting a Fitness Tracker

4.1 Considerations

Reading online product reviews and consumer ratings provides a wealth of information on a fitness tracker. When reading the reviews, also look to see if the tracker supports

3rd party app development through an API. Most major fitness trackers provide software libraries to allow software developers to get access to sensor data. Some manufactures allow access to sensor data through a real-time stream while others only allow access to aggregated data post hoc through a web portal and often, that aggregated data is from the entire day. Depending on the nature of the research, real-time sensor streaming may be a requirement. However, if accessing data post hoc, be sure to keep accurate time records for when a participant wore the tracker in order to tie the sensor data to the proper stimuli events of the experiment.

Often physiological recording devices are complex and require specialized training for the researcher to use the equipment properly. Fitness trackers on the other hand are simple to use and can be as easy as putting on a watch. This ease of use can be deceptive though, be cautious with placing fitness trackers on participants to be consistent with both sensor orientation and placement on the wrist. Also, optical heart rate sensors are susceptible 60 Hz light sources, so be sure the tracker is worn with the correct tautness and does not shift throughout the experiment. Field studies, and to some degree laboratory studies, benefit from the mobility that fitness trackers provide. However, some trackers require Bluetooth or Wi-Fi connectivity to record data. When testing, look at wireless signal strengths with paired devices and make sure to stay within range throughout the study.

4.2 Selection of Microsoft Band 2

As mentioned previously, determining the appropriate device was important for two studies. The criteria for both studies were to choose a fitness tracker that provided local access to heart beat data. The fitness tracker ideally needed to provide raw photoplethysmogram data. This would allow for post hoc analysis on algorithms for heart beat detection. At a minimum, the tracker needed to provide an R-peak detection. This would allow the tracker to support heart rate (HR) and heart rate variability (HRV) analysis to assess mental workload or task engagement for the Nuclear Power Plant and UH60A/L studies, respectively. Several fitness trackers provided API access to beats per minute data, but the Microsoft Band 2 was the only commercially available fitness tracker that provided RR intervals. Both of the studies chose to use the Microsoft Band 2 as the fitness tracker because the RR interval supports HR and HRV analysis. However, both studies used additional sensors to augment and verify HR and HRV analysis. GSR and skin temperature were collected for both studies due to their correlates with both mental workload and task engagement.

4.3 Lessons Learned

Most fitness trackers have enough battery life to last several days on a single charge for normal daily use. However, when using a fitness tracker for research, it typically will be streaming and recording multiple sensor feeds for the duration of the study. If the study is several hours, battery life could be of concern. During pre-experiment testing, the Microsoft Band 2 lasted well over four hours with all sensors continuously streaming.

However, as a precautionary measure, the fitness trackers were fully charged before each experimental run. Battery life was never an issue for either of the studies.

Psychophysiology studies often require participants to be connected to multiple sensors. Most of these sensors are only mildly invasive; however, there is not much risk of participants forgetting to remove them before concluding the study. This is not the case with fitness trackers. Since fitness trackers are so natural to wear and non-invasive, it is important to have explicit procedures reminding researchers to remove the trackers from participant's wrists before dismissing them. During the two studies no fitness trackers were lost, but even with a checklist, a few occasions required participants to be tracked down in the parking lot to return a fitness tracker.

The default settings on fitness trackers are setup for the consumer and some of the settings need to be changed before conducting research. It is important to go through each setting and make sure the fitness tracker is in a mode that will not interfere with the study. Between piloting and data collection for training effectiveness evaluation with UH-60A/L pilots, the Microsoft Band 2 had a software update. The update changed all the settings back to defaults (including the activity reminder). During experimentation, two of the UH60A/L pilots had a vibration reminder to get up and move due to inactivity. After that experiment, a checklist item was added to visually verify the fitness tracker was in "do not disturb" mode before each experimental session.

It is critical to understand how to retrieve the data from the fitness tracker. Doron Katz of Programmable Web has a great web page that aggregates the popular APIs for fitness trackers along with links to developer pages for each API [13]. The Microsoft Band 2 API provided direct access to the sensor readings from a paired Bluetooth phone or windows PC. Windows PC development requires c# RT metro libraries. The metro app requirement limits the APIs capability of integrating the sensor data with a native windows application. As a workaround, a custom streaming UDP client/server app was developed for data collection. This allowed the metro app to communicate with a native windows app.

Be careful when purchasing a newly released fitness tracker model because technology changes so fast that manufactures often do not have adequate time to perform sufficient product testing. Be sure to adequately put the technology through its paces before using in an experiment. Several of the Microsoft Band 2 that were purchased for experimentation had to be returned due to a charging issue after less than a month of use. Waiting a few months before ordering a newly released fitness tracker can avoid having to return it because of a minor design flaws. On the flip side, do not order fitness trackers that are near the end of life for production. If the fitness tracker breaks in the middle of a study, replacement availability for the specific model may become impossible. Also, select fitness trackers from manufactures that have a history of making quality fitness trackers. While Microsoft is a large company with a proven track record of quality hardware, their line of fitness trackers was discontinued following completion of these two studies due to lack of consumer interest in their high-end fitness tracker.

5 Data Quality Comparisons

5.1 Nuclear Power Plant Study

The nuclear power plant (NPP) main control room (MCR) study used both the Microsoft band 2 fitness tracker and the B-Alert X10 [14] system to collect psychophysiological responses of participants. The B-Alert system was used to sample the participant’s electrocardiography signal at 256 Hz. Post hoc analysis was run on the B-Alert sampled signal using a bio signal processing toolkit written in Python called BioSPPy [15]. The Hamilton ECG R-peak segmentation algorithm was used to calculate R-peaks [16]. From the R-peaks, three measures were derived: interbeat interval (IBI), beats per minute (BPM) and HRV. The Microsoft Band 2’s R-peaks collected during the same time period allows for a comparison between the fitness tracker and the Hamilton ECG R-peak segmentation algorithm on the B-Alert sampled ECG signal. The table below shows metrics for five participants’ data collected during a 25-minute change detection task (Table 1).

Table 1. Comparison of Microsoft band 2 and Hamilton ECG algorithm using B-Alert sampled ECG signal. The Microsoft band 2 data is in columns with MB headers and the Hamilton data is in columns with the H headers. IBI values are in milliseconds.

Participant	Min IBI		Max IBI		Mean IBI		Median IBI		BPM		HRV	
	MB	H	MB	H	MB	H	MB	H	MB	H	MB	H
1	3	417	7641	1343	836	631	682	625	71.77	94.95	715.17	58.91
2	3	531	5084	757	690	614	649	609	86.91	97.69	385.96	34.66
3	1	351	11297	1355	731	581	655	574	82.00	103.13	607.91	46.21
4	32	289	2626	1070	801	814	827	832	74.87	73.70	225.14	116.72
5	16	578	2928	2359	803	809	792	757	74.68	74.14	252.01	202.06

A visual inspection of the Microsoft Band 2’s R-Peak detection accuracy was done by plotting each participants’ data for both the Hamilton and Microsoft band 2’s R-peak algorithms overlaid on the B-Alert sampled ECG signal. The graph below shows the ECG signal for the first 5 s from the first participant. The vertical dashed lines represent Microsoft Band 2’s detected R-peaks and the solid light gray lines represent the Hamilton detected R-peaks (Fig. 1).

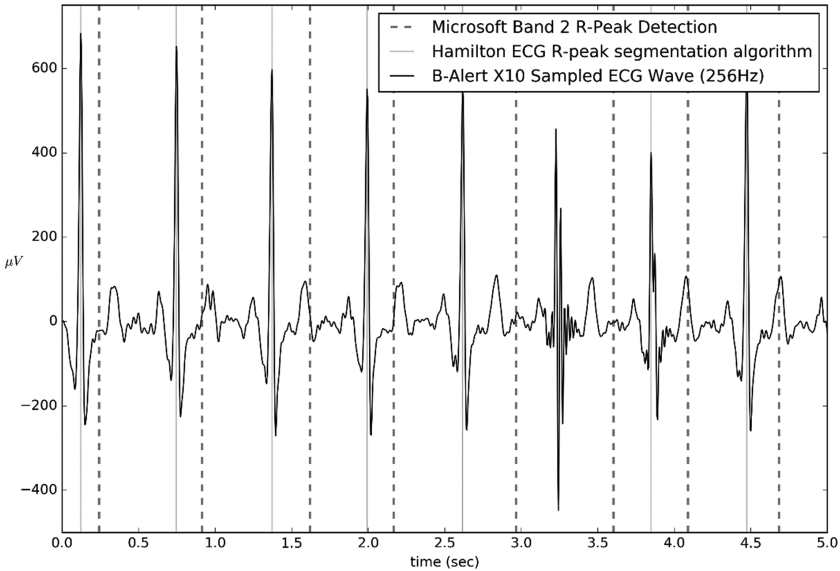


Fig. 1. First 5 s of a former reactor operator performing tasks in an EOP. ECG was collected using B-Alert.

5.2 UH-60A/L Training Effectiveness Study

The UH-60A/L study performed similar analyses, but used a different ECG sensor. The Thought Technologies ProComp Infinity device [17] was used to sample the electrocardiography signal at 2048 Hz. In the same manner as the NPP study, post hoc analysis was run on the ProComp Infinity sampled signal using BioSPPy [15]. The Hamilton ECG R-peak segmentation algorithm was also used to calculate R-peaks [16]. Three measures were derived from the R-peaks: IBI, BPM and HRV. In the same manner as the NPP MCR study, the Microsoft Band 2’s R-peaks were compared to the Hamilton ECG R-peak segmentation algorithm on the ProComp Infinity ECG signal. The table below shows metrics for five UH60A/L pilots’s data collected during a 60-minute simulated CASEVAC scenario. The scenario was conducted in the Operational Flight Trainer [18] (Table 2).

Table 2. Comparison of Microsoft band 2 and Hamilton ECG algorithm using ProComp Infinity sampled ECG signal. The Microsoft band 2 data is in columns with MB headers and the Hamilton data is in columns with the H headers. IBI values are in milliseconds.

Participant	Min IBI		Max IBI		Mean IBI		Median IBI		BPM		HRV	
	MB	H	MB	H	MB	H	MB	H	MB	H	MB	H
1	16	387	2172	1006	753	725	766	716	79.61	82.76	170.88	85.64
2	16	321	2516	907	787	670	812	669	76.25	89.46	235.31	64.25
3	31	287	1797	1045	810	811	828	812	74.03	73.99	159.24	65.38
4	31	277	2016	850	513	432	453	431	116.98	139	187.36	18.15
5	31	550	1828	1238	808	783	828	786	74.24	76.59	199.58	95.42

A visual inspection of the Microsoft Band 2’s R-Peak detection accuracy was done by plotting each participants’ data for the Microsoft band 2’s R-peak algorithms overlaid on the ProComp Infinity sampled ECG signal. The graph below shows the ECG signal for the first 5 s from the first UH-60 pilot. The Vertical dashed lines represent Microsoft Band 2’s detected R-peaks and the solid light gray lines represents the Hamilton detected R-peaks (Fig. 2).

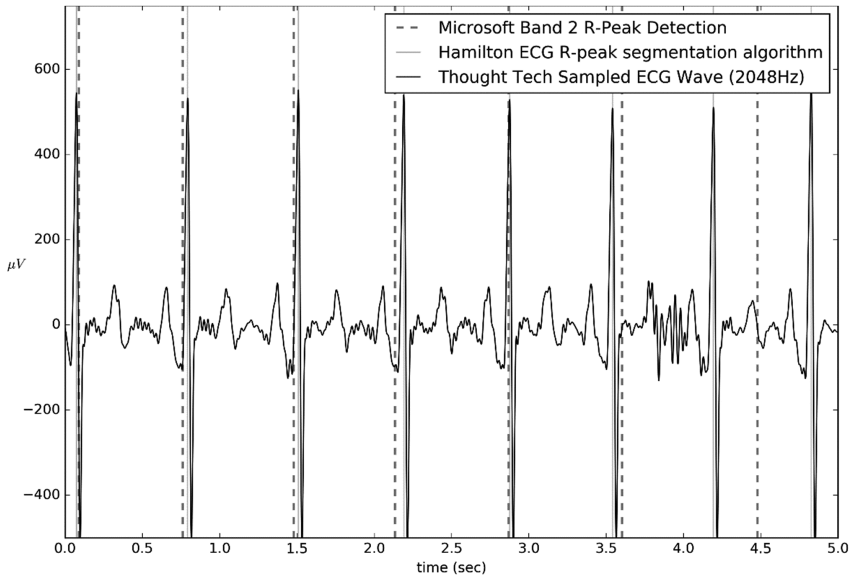


Fig. 2. First 5 s of a pilots ECG during a simulated CASEVAC Mission. ECG was collected using ProComp infinity.

It is clear that the optical heart rate sensor in the Microsoft Band 2 fitness tracker does not provide equivalent R-peak data to electrical sensors designed to provide highly accurate ECG signals. While optical heart rate sensors found in wrist worn fitness trackers provide useful estimates for monitoring health, more work is needed to ensure accurate data is provided for performing certain types psychophysiology research. An incorrect estimate of the R-peak by 100 ms can increase HRV by 100% [19]. While HRV analysis using the band is not practical given the magnitude of error associated with the optical sensors on the wrist, BPM analysis may be accurate enough for some types of analysis. Visual inspection of the Microsoft Band 2 plotted over the raw ECG signal shows roughly one R-peak detected per QRS wave seen in the ECG signal. Also, BPM analysis is more robust to false positives/negatives because one or two missed beats only slightly affects BMP.

Acknowledgments. This report was prepared as an account of work sponsored by (CA# W911NF-14-2-001 and in part (NRC-HQ-12-C-04-0058/NRC-HQ-60-14-T-0001) by an agency of the U.S. Government. Neither the U.S. Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for any third party's use, or the results of such use, of any information, apparatus, product, or process disclosed in this report, or represents that its use by such third party would not infringe privately owned rights. The views expressed in this paper are not necessarily those of the U.S. Army Program Executive Office for Simulation, Training, and Instrumentation or the U.S. Nuclear Regulatory Commission.

References

1. Udovičić, G., Topić, A., Russo, M.: Wearable technologies for smart environments: a review with emphasis on BCI. In: 2016 24th International Conference on Software, Telecommunications and Computer Networks (SoftCOM), pp. 1–9. IEEE (2016)
2. Kim, Y., Wang, H., Mahmud, M.: Wearable body sensor network for health care applications. In: Smart Textiles and Their Applications, p. 161 (2016)
3. Wilson, S.: Wearables Market to Be Worth \$25 Billion by 2019 (2015). Accessed 2017
4. Tehrani, K., Michael, A.: Wearable Technology and Wearable Devices: Everything You Need to Know. *Wearable Devices Magazine* (2014)
5. Definition of: fitness tracker. <http://www.pcmag.com/encyclopedia/term/67469/fitness-tracker>
6. Goode, L.: The Apple Watch didn't slow down sales of fitness trackers last year. *The Verge*, Vox Media (2016)
7. Infantolino, Z., Miller, G.A.: Psychophysiological methods in neuroscience. In: Biswas-Diener, R., Diener, E. (eds.) *Noba Project* (2017)
8. Get started with the Microsoft Band SDK. <https://developer.microsoftband.com/bandsdk>. Accessed 2017
9. Wetherow, O.: Microsoft Band II Teardown (2016). <http://www.chipworks.com/about-chipworks/overview/blog/microsoft-band-ii-teardown>
10. Hayes, T.: What's Inside a Fitness Tracker, Anyway? (2014). <http://www.digitaltrends.com/wearables/whats-inside-fitness-tracker-anyway/>. Accessed 2017
11. bin Syed Noh, S.A.: Galvanic Skin Response (2015)
12. Rhodes, H.B.: Accessing and using data from wearable fitness devices. *J. AHIMA* **85**(9), 48–50 (2014)
13. Katz, D.: Top 10 Fitness APIs: Apple Health, Fitbit and Nike (2015). <https://www.programmableweb.com/news/top-10-fitness-apis-apple-health-fitbit-and-nike/analysis/2015/04/17>. Accessed 2017
14. B-Alert X10 EEG Headset System (2015). <http://www.advancedbrainmonitoring.com/xseries/x10/>. Accessed 2017
15. BioSPPy - Biosignal Processing in Python (2017). <https://github.com/PIA-Group/BioSPPy>
16. Hamilton, P.: Open Source ECG Analysis Software Documentation (2002)
17. ProComp Infiniti System (2016). <http://thoughttechnology.com/index.php/hardware/procomp-infiniti-170.html>. Accessed 2017
18. Training, L.S.: UH-60 M Operational Flight Trainer (2010)
19. Mulder, L.: Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biol. Psychol.* **34**(2), 205–236 (1992)