Pervasive and Mobile Computing 50 (2018) 137–147

SAND2019-11632B

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

journal homepage: www.elsevier.com/locate/pmc

Physiological state in extreme environments

Glory Emmanuel-Aviña [a,*], Kristin Divis [b], Robert G. Abbott [b]

[a] Sandia National Laboratories-California, Livermore, CA, USA
[b] Sandia National Laboratories-New Mexico, Albuquerque, NM, USA

ARTICLE INFO

Article history:
Received 7 March 2018
Received in revised form 13 August 2018
Accepted 15 August 2018
Available online xxxx

Keywords:
Wearable devices
Extreme environments
Physiology
Performance

ABSTRACT

Commercial off-the-shelf (COTS) wearable devices are used to quantify physiology during physical activities to monitor levels of fitness and to prevent overexertion. We argue that there are limitations and challenges to measuring physiological data with current state-of-the-art wearable devices, both with the hardware as well as the data itself. These limitations and challenges are exacerbated when wearable devices are used in extreme climate environments We discuss these through empirical findings from our study where hikers are suited with wearable technologies as they cross the Grand Canyon. We discuss the performance of various wearable technologies in the extreme environment of the canyon as well as the concerns with downloaded data. These findings highlight the needs and opportunities for the wearable devices market, specifically how wearable technologies could mature to quantify performance and fatigue through real-time data collection and analysis.

1. Introduction

The use of wearable devices has exploded over the past decade, especially over the past five years. The wearable technology industry alone is approximately a 20 billion dollar enterprise with new smart wearable coming out every day [1]. The wearable industry is expected to double in the next five years and be worth $34 billion in 2020, selling about 411 million wearable smart devices [1]. There is a strong attraction towards wearable devices because they provide a way to quantify human performance. While conducting various activities such as running, hiking, or even sleeping, wearable devices quantify our performance. This in turn helps us become better versions of ourselves because we can measurably work towards personal goals and prevent overexertion. And the market keeps enhancing our ability to quantify performance. Each new wearable presents a feature better than its previous version, intended to make the device more intuitive, easier to wear, and/or more integrated with the activities being performed. Smart watches are able to track optical heart rate, pulse oximetry and sleep patterns. Bioharnesses are able to collect core body temperature, respiration rates, and electrocardiogram (ECG) data. Smart clothing can track electromyography, the electrical activity produced through muscle movement. Not only can we collect a wealth of performance data, it can be done with minimal disruption to the activity being performed.

However, empirical data is limited on how wearable technologies perform in extreme climate environments where both hardware and software are tested to their limits. It is also unclear if the data that is being collected helps to enhance actual performance as well as mitigate against health events that may occur in harsh climates. The purpose of this article is to discuss the current state of wearables and explore the path forward based on what we have learned so far from research conducted in an extreme environment. We focus on how devices have been used in a study conducted at the Grand Canyon, what the data tells us where it is lacking, and what additional data sources are needed.

* Corresponding author.
E-mail address: gremman@sandia.gov (G. Emmanuel-Aviña).

https://doi.org/10.1016/j.pmcj.2018.08.003
1574-1192/

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

2. The benefits of wearable technologies

One can easily argue that there are many benefits to using wearable technologies while engaging in a range of activities. In 2017 alone, empirical articles have discussed how wearable technologies can aid in enhancing the areas of education [2], cognitive research [3], healthy living [4], medicine [5], and mental health [6] to name a few. The Huffington Post lists that the major benefits of wearables are the improvement to personal safety, the ability to track health and fitness, and staying connected at all times [7]. As the 21st century becomes increasingly interested in gathering more data and quantitatively knowing how to become better versions of ourselves physically, physiologically, neurocognitively, and biologically, the wearable technology industry will continue to increase exponentially.

More investment and research will continue to be poured into enhancing wearable technologies and their benefits to users. While wearables have many benefits, the technology is far from serving us exactly how we need it to. The following section provides a generalized overview of wearable technology limitations. We focus on the use of wearables that collect physiological data, but also the potential for improvement through research, development, and innovation.

3. The limitations of physiological wearable technologies

The constant release of new wearable technologies highlights that there are limitations to what each one can do. Users continue to upgrade devices because a wearable will only last for so long and there is always room for better data to be collected. Issues with devices include the hardware, the data collected by devices, how the device works with particular activities, the longevity and robustness, how the data is downloaded and processed, if the data is useful, and if the data is enough. Each of these areas for wearables that collect physiological data are discussed.

For the hardware of wearable technologies, the greatest barrier to overcome is making a product that will fit and conform to the body type of a large population. Many smart watches have a large and small version to accommodate varying wrist sizes. However, various size availability does not account for how the body changes (e.g., gains or loses water) while performing an activity. Wearable clothing has even stronger fit requirements than a watch or bioharness due to the concentrated movement it is measuring. Therefore, clothing can seem very snug and restrictive as well as be uncomfortable and lead to chaffing. Due to the hindrances with device fit, there is a push towards having "invisibles" be the next generation of wearables, meaning that sensors are integrated with the person [8]. To be fully integrated, the technology would not need to move from device to clothing, but to the human biome. This could be through microneedle patches, temporary tattoos, and even device implantation.

Even when the device hardware is functional, the data collected by a device may not be. The data collected by devices should be both valid and reliable, but many times this is not the case [9,10]. Inaccurate data leads to uninformed decision-making. Regarding validity, if a watch captures Global Positioning System (GPS) but GPS struggles with running switchbacks, under bridges, up an incline, et cetera, then the watch may not accurately report mileage, causing a user to run more or less than they intend to. If heart rate (HR) is inaccurate, even by small amounts, this could pose a health risk with a user over-exerting to hit a target heart rate or not realizing how low their heart rate has become. For reliability, many devices create a user profile and baseline for performance. If data variably measures a user's performance, the user will be unable to set a baseline and devices will not be able to reliably track performance.

It is normal for users to wear multiple devices because it is rare that one or two devices tracks all the data desired. For example, to capture HR, cadence, breathing rate, GPS, and ambient temperature, one could wear a fitness watch, footpod, wearable clothing, and chest strap. This also leads to complications when downloading the data, especially if each device has its own software and user interface. Users not only desire easy uploading, but to see all their data in one location to analyze collectively. That also leads to limitations in longevity and robustness. A device battery may not be able to last the entire duration of the activity, such as a daylong hike. The device may also give out due to the screen or interface breaking, battery dying, water damage, et cetera. The limitations of data collected, longevity, and robustness may be why 40% of millennials replace their wearable devices every 6 to 12 months [11].

We have already touched on whether the data is useful by being accurate, but there is also the question of, "is the data enough?" Users rely on wearables to provide the earliest predictor possible for a health event, to advance their performance, and to achieve short- and long-term goals, but human performance involves more than HR, cadence, body temperature and other physiological markers. Cognitive data, such as how fatigued one feels, how motivated they are to complete the task as well as neurocognitive data, such as neural activity and brain chemistry, are also indicators of how well one might perform an activity. Biological data such as change in glucose, sodium chloride, electrolytes during an activity also quantifies level of effort and resources available to continue the activity. Because we know that human performance has not been fully quantified through current wearable devices, the industry continues to search for innovative ways to collect performance data. We want useful data and more of it.

4. Empirical study informing wearable use

4.1. Wearable data from extreme environments

When conducting research or completing a performance task in an extreme environment, one quickly learns to "expect the unexpected", knowing that such environments often provide unforeseen challenges. One example of an extreme

Fig. 1. View of the Grand Canyon from South Kaibab to North Kaibab. Intended to illustrate the terrain, length, peaks/valleys, and ruggedness of the hike.

environment where visitors regularly use wearable devices to track their performance is the Grand Canyon, specifically the Grand Canyon Rim-to-Rim (R2R) hike.

The Grand Canyon R2R is not a typical walk in the park—the 33.7 km course crosses the canyon, with temperature changes up to 10 °C common for hikers starting early in the morning at the top of the canyon and crossing the bottom during the afternoon (see Fig. 1). The trail includes elevation changes of approximately 1500 m going down into the canyon and then 1750 m coming up out of it, and the notorious "box" section of the trail has geological properties that lead to dangerously high temperatures in the summer. The National Parks Service actively discourages visitors from attempting to complete the R2R in a single day due to the challenging environment, but hikers still daily attempt the R2R. Over 250 hikers are airlifted out of the canyon each year, many with heat-stress related symptoms [12].

The R2R is challenging not only for hikers but also for most commercially available wearable devices. The usual time to finish the hike is 10–15 h, but those struggling often take far longer (assuming they do make it out in a single day). With no opportunity to recharge or swap batteries, that time limit stretches or exceeds the capacity of many commercially available devices locked in "activity mode". The Grand Canyon is also far removed from a sterile, clean lab environment—devices must be able to withstand a combination of sweat, dirt, submersion in the Colorado River if a hiker jumps in to cool off, freezing temperatures, extended direct exposure to sunlight, and dangerously high temperatures. GPS also functions poorly in the steep sections of the canyon, with insufficient access to the requisite satellites. Finally, there is virtually no mobile connectivity at the canyon, so all devices need to be able to function and store data without an active mobile or wireless network connection.

4.2. The WATCH study

Ultimately, the best way to understand the application of wearable devices in extreme environments is to actually use and study wearable devices in such settings. The Wearables At The Canyon for Health (WATCH) study is an ongoing research study funded by the Defense Threat Reduction Agency (Project number CB10359). WATCH uses wearable devices to track performance while participants attempt to complete the grueling 33.7 km, R2R hike at the Grand Canyon in a single day, which makes it a natural testbed for evaluating wearable device performance in an extreme environment.

Data is collected from volunteer hikers as they complete the R2R. Each hiker is provided with a suite of commercial-off-the-shelf wearable devices to wear during the hike including a smart watch, fitness chest strap, core temperature sensor, footpod, and ambient temperature cube. Each hiker is also given a smart phone to help connect devices and collect data. Once the hikers reach the end of the hike, a team of researchers stop the devices and downloads the data. This study follows a strict protocol to prevent confounds (e.g., taking different paths from start to finish, using personal wearable devices, different device configurations, etc.) and has been reviewed and approved through Sandia National Laboratories' human subjects board. For a complete write-up of the study's methodology, see referenced publications [13,14].

4.3. Data collected to date

The WATCH study has completed three of four rounds of data collection to date. Both civilian and military hikers have participated in the study, with a total of 150 hikers wearing devices as they crossed the canyon. The hikers ranged in age from 18 to 73 and included those who have completed the R2R multiple times along with those who were hiking it for the first time. See Table 1 for more details on participant demographics.

Hikers were outfitted with a variety of packages of wearable devices—some lightweight options only contained a few items whereas more advanced packages contained up to eight items. See Table 2 for a breakdown of functionality of devices by packages, including counts of how many were sent out.

5. Initial results on wearable device performance

Prior to collecting data from volunteer hikers doing the Rim-to-Rim, the WATCH team tested a suite of devices through "bench tests". These bench tests involved individual team members taking the devices on local hikes, doing the Rim-to-Rim hike, and assessing the data and hardware for overall quality. The bench tests quickly ruled out many devices that did not meet battery life requirements, durability requirements, or data quality requirements. For example, elevation can be measured a number of ways: some low-end devices measure elevation via "flights of stairs", most devices rely on GPS, and high-end models rely on barometric altimeters. Our bench tests revealed that only elevation measured via a barometric altimeter gave consistent, accurate data in the canyon. Elevation profiles reliant on GPS data show extreme jumps in elevation during steep sections of the trail where satellite signals are impaired—a reading that would only be valid if the testers decided to go bungee jumping.

Those early tests also revealed some unexpected challenges. The electronic components in the devices do not function well in the near-freezing temperatures encountered early in the morning at the start of the hike. Another challenge is to achieve reliable skin-to-sensor connections on cold hikers who did not want to wet the sensors and were not yet sweaty. Watches and other devices are easily bumped out of recording when the "locking" functions are not distinct from common, accidental interactions (e.g., devices that allowed one to press any button to unlock led to more problems than those that required a specific button to be pressed, held, and confirmed). This is true even when hikers are familiar with the devices and are being careful—simply taking a jacket off as the day warmed up could adversely affect the recording via an accidental button bump.

The large samples of data collected on "real" hikers during data collection events revealed further insights into using wearable devices in extreme environments. ECG heart rate monitors (e.g., chest straps) consistently outperformed a number of optical heart rate monitors. The optical heart rate monitors (either wrist-based or forehead-based) tended to have more dropouts, more inaccurate readings, and showed a pattern of "fading" in the latter part of the hike (i.e., the last few kilometers out of the canyon would show an unexplained decline in heart rate). One of the biggest lessons learned is the importance of onboard logging. Devices that paired with others (e.g., a temperature sensor that paired with a watch or a heart rate monitor that paired with an iPod app) often had fatal errors during the data collection event. Devices would lose their pairing, some could only be paired when the activity was started (making it more likely that the wrong devices would be paired, if other hikers or sensors were in proximity of the device), and some would pair correctly at the beginning of the hike only to drop out for unknown reasons later in the hike. Devices that included onboard logging consistently provided more reliable data. Finally, comfort and fit of the wearable device is imperative for extended wear in extreme environments. Watches must be tight enough to collect usable data but be expanded if wrists swell during the activity or tightened if falling temperatures lead to a loser fit. This point is particularly important for wearable clothing items—clothing imbedded with sensors by necessity has a tight fit, which can lead to an unacceptable level of chaffing with extended use.

Finally, not all devices are created equally when it comes to accessing the data. Some devices must have the data uploaded within a few days or it will be deleted from memory (or overwritten for devices that are continuously "on" and recording such as some environmental sensors). Some vendors give their users access to the raw data via an easy-to-download file. Others only provide access to curated summary statistics or require users to scrape web pages in order to access the information (rather than providing downloadable files). Some delete the information off the device once it is downloaded, others allow one to keep multiple copies in case one gets corrupted. Bugs and poor interfaces abound when it comes to accessing recorded data—especially in "cutting edge" devices, making confirmation of data accessibility a necessary when choosing wearable devices.

6. Wearable device performance at the Grand Canyon

The R2R hike tests the hiker and devices in many ways as described in previous sections. This section highlights the tradeoffs encountered for a variety of sensors that provide data streams pertaining to location and heart rate. These are needed to capture either hiker performance or the level of effort required to attain that level of performance. These are not the *only* important data streams (we record several others), but these are interesting cases because each can be measured using significantly different types of sensors.

Each wearable device is a compromise among numerous design considerations including cost, appearance, weight, size, battery life, data quality, durability, convenience, and comfort. Thus a device that is optimized for any single use case is necessarily sub-optimal for other uses. Our analysis is not intended to discount the potential usefulness of any device or type of sensor in general, but only insofar as the prioritization of these considerations resembles those of our study.

Our findings are generally consistent with informal observations made by consumers of fitness equipment, but by systematically collecting data on a relatively large number of participants and devices, the R2R study enables more quantitative analysis.

Fig. 2. GPS Tracks from 173 Hikers on the North Kaibab Trail in the Grand Canyon. North is to the right. The clouds of yellow at some points along the trail indicate locations where GPS reception is poor. Overhead imagery courtesy of Bing Maps.. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

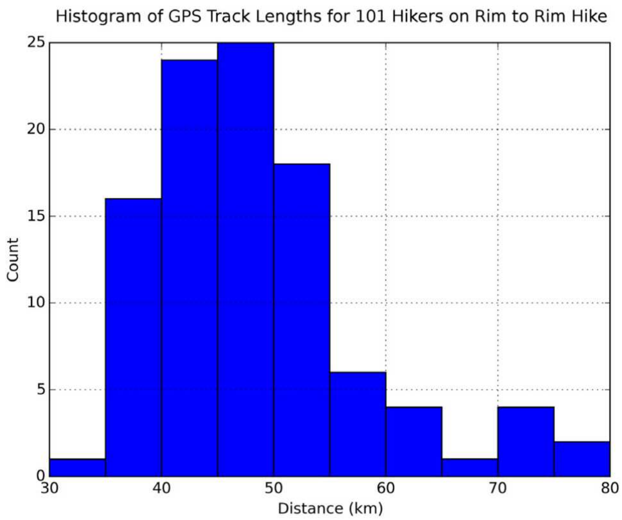

Fig. 3. Histogram of GPS track lengths for 108 hikers on R2R Hike. The wide variability in track length implies that other measures derived from GPS (e.g. speed) will not be reliable. The median track length is 46.3 km, whereas the actual length of the Kaibab trail according to the National Park Service is 33.7 km.

6.1. Location tracking

Location tracking is not of direct interest to our study, since fatigue is the product of effort over time, rather than absolute position on the earth. However, from location data can be derived several measures of interest: speed, distance, climb rate, elevation (which is tied to barometric pressure and thus aerobic performance), and even group dynamics (if multiple hikers are tracked). With GPS receivers now almost ubiquitous (for example in all but the least expensive smartphones), location tracking might seem to be a solved problem. The data we have collected confirms several known issues with GPS. First, GPS receivers still consume a large share of the power available to wearable devices. Second, GPS reception is poor to nonexistent indoors or outdoors under cover (e.g. tall trees or between buildings) and in canyons. We first examine the accuracy and consistency of GPS tracking in the R2R hike, and then compare it to alternate sensors for measuring elevation and hiking speed.

Fig. 2 shows the GPS data from 173 hikers on the North Kaibab Trail in the Grand Canyon. It is a distance of 8 km from the south to north (left to right) on this image. With minor exceptions the trail is non-branching, so under ideal tracking conditions all the tracks would coincide. Instead, the tracks agree in some locations but are widely dispersed at other times. A few locations are obvious outliers, falling kilometers away from the point only seconds before and after. The areas of greatest dispersion coincide to portions of the trail through steep canyons.

Fig. 3 is a histogram of the lengths of 108 GPS tracks. This is the subset of our GPS tracks that are complete (i.e. they start at the South Kaibab trailhead and end at the North trailhead, meaning the watch battery did not run out, the hiker did not

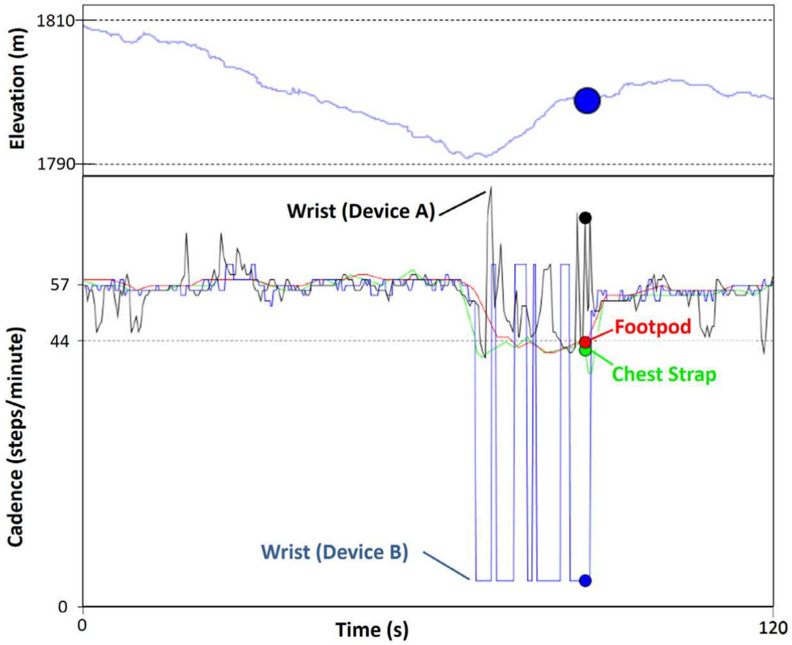

Fig. 4. Elevation (top) and Cadence (bottom) data from 4 sensors on the same individual during a short hike. The two wrist-based cadence sensors are less accurate than the footpod and chest strap, particularly while walking more slowly up a hill.

push the button to stop tracking, etc.). The lengths are the sum of distances between successive points using the Haversine formula, using latitude and longitude only (i.e. disregarding elevation). The median track length is 46.3 km, whereas the length of the Kaibab Trail according to the National Park Service is 33.7 km [15]. The mean track length is 48.4 km and the standard deviation is 10.9 km. Track lengths should be marginally more than the trail length, since any deviation from the trail (e.g. using an outhouse) adds to the distance actually walked, but this cannot account for a median extra distance of 12.6 km. GPS tracking error usually increases the track length, as demonstrated by previous theoretical and empirical research [16], but could also reduce it slightly by cutting corners and switchbacks that exist in the trail.

Location tracking error impacts derived metrics such as distance and speed. In the absence of accurate fine-grained speed data, we focused on identifying breaks. This enables us to study whether fatigue affects heart rate dynamics during short breaks.

To identify breaks, we used cadence data, which is measured in steps per minute. Cadence does not directly correspond to speed, since steps may be shorter or longer, but cadence should be near zero during breaks. Compared to GPS, cadence sensing requires less power and does not require a satellite signal so it works indoors or in canyons. However, our initial analysis of the cadence data we had collected from wrist wearables revealed that it was also very inaccurate. In particular, a cadence of 0 is often reported when hikers are moving slowly uphill. A parallel recording of multiple cadence sensors on the same individual confirmed that wrist-based cadence is not sensitive enough to capture this data. Fig. 4 shows this data, which includes two wrist-based cadence sensors, one integrated into a chest strap, and a footpod. At typical walking speed on flat ground, cadence is approximately 57 steps per minute and all the sensors are in fairly close agreement. But walking uphill, cadence drops to approximately 44 steps per minute and the sensors undergo softer accelerations. The wrist-based sensors both exhibit more noise, with one brand especially frequently and incorrectly dropping to zero. These results led us to integrate footpods into subsequent data collection events rather than relying on wrist-based cadence measures.

Elevation data is also crucial to measuring hiking performance in the Grand Canyon. More than the length of the hike, it is the ascent and decent that make the R2R hike unique and difficult. Fine-grained elevation data is required since speed and heart rate data must be evaluated in the context of the slope of the trail.

Unfortunately, GPS elevation data is less accurate than positional data (latitude/longitude), even with clear satellite reception. Fig. 5 shows an example of tracking error in GPS elevation data. The reading from the barometric altimeter (in green) is consistent with the observed elevation profile of the trail—the first part mostly level, the second part a steady incline. The GPS elevation profile shows large oscillations (approximately 15 m) every couple of minutes. Such errors greatly disrupt calculations of slope and total distance climbed, and therefore effort expended.

Fig. 6 shows a histogram of total vertical ascent, in meters, for 119 hikers on the R2R trail. Twenty of the tracks are from GPS, the other 99 are from barometric data. In both cases, we computed differences between 30 s median values. The total vertical ascent values collected from GPS for different hikers vary widely (standard deviation = 3133 m), which is not plausible given that all hikers followed the same path. In contrast, the measurements of total climb from barometric

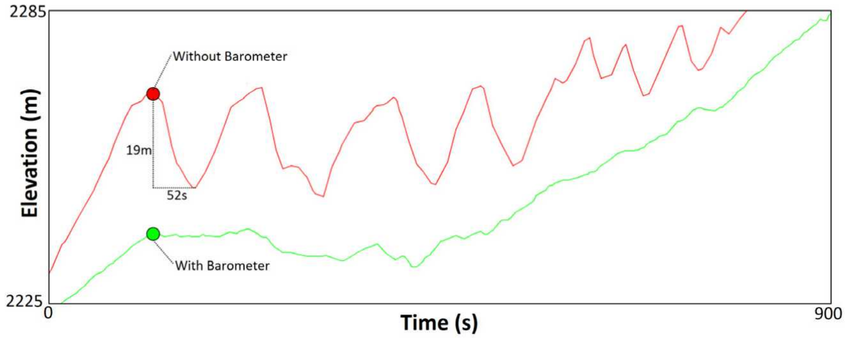

Fig. 5. Fifteen minutes of simultaneous elevation readings from GPS vs Barometric altimeters. The subject wearing the altimeters was on a steadily ascending trail, consistent with the barometric altimeter reading. The GPS elevation reading repeatedly oscillates by approximately 10 to 20 m each two minutes.

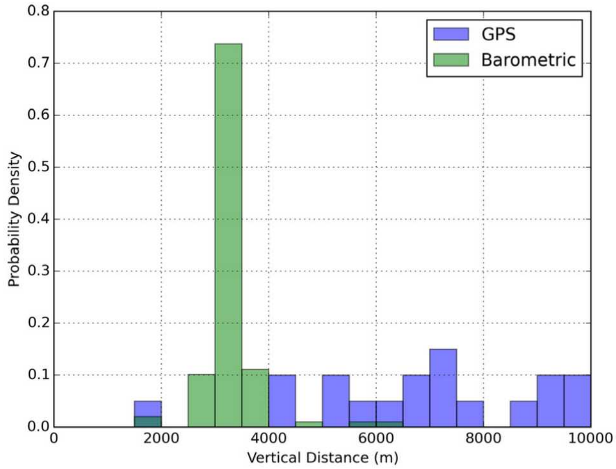

Fig. 6. Histogram of total vertical ascent of 119 GPS tracks on R2R Hike. Twenty tracks were collected from GPS, the remaining 99 from Barometric altimeters. Approximately 75% of the barometric altimeters (in green) accumulated between 3000 m and 3500 m, whereas the GPS altimeters (in blue) varied widely.. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

altimeters are much more consistent (standard deviation = 492 m). The median total climb recorded is 7044 m for GPS and 3363 m for barometric. For reference, the vertical distance from the Phantom Ranch to the North Rim is 1736 m according to the National Park Service [17]. The total ascent of hikers is somewhat more than this, since the descent down the South Kaibab trail includes a small amount of ascent, and the climb back up has a small amount of descent, which is offset by yet more ascent. The values from both GPS and barometric data are too large, but much more for the GPS data.

6.2. Heart rate

Heart rate is closely related to level of effort and thus of vital interest to our study. Until recently, the only practical continuous measure of heart rate during physical activity was ECG (electrocardiography) which (outside laboratory or clinical settings) typically requires an elastic strap with two or more electrically-conductive pads worn around the chest. The pads sense the electrical signals that cause the heart muscles to contract. This is a relatively simple and reliable sensor, but is less than ideal with respect to comfort and convenience, particularly since the strap must be worn snugly to achieve good conductivity and stay in place. In recent years optical heart rate sensors have arrived in consumer devices. These devices illuminate the skin and measure changes in light reflection corresponding to blood flow due to the cardiac cycle. Compared to ECG, optical sensing may increase convenience by using alternative locations for measurement.

We tested optical heart rate sensors at the wrist and forehead. To facilitate direct comparisons, we equipped each hiker with up to 3 heart rate sensors (ECG strap, wrist, and forehead sensors). Unfortunately, the heart rate data provided by the optical sensors was found to be inconsistent with the ECG data. Fig. 7 shows example data recorded from six participants each wearing three heart rate sensors throughout the R2R hike. None of these six cases illustrated the three sensors recording matching values throughout the hike. At *almost* all times when both optical forehead and ECG data are present, they agree closely; the optical forehead device (blue) usually stops reporting values when it cannot get a reading. See for example subject

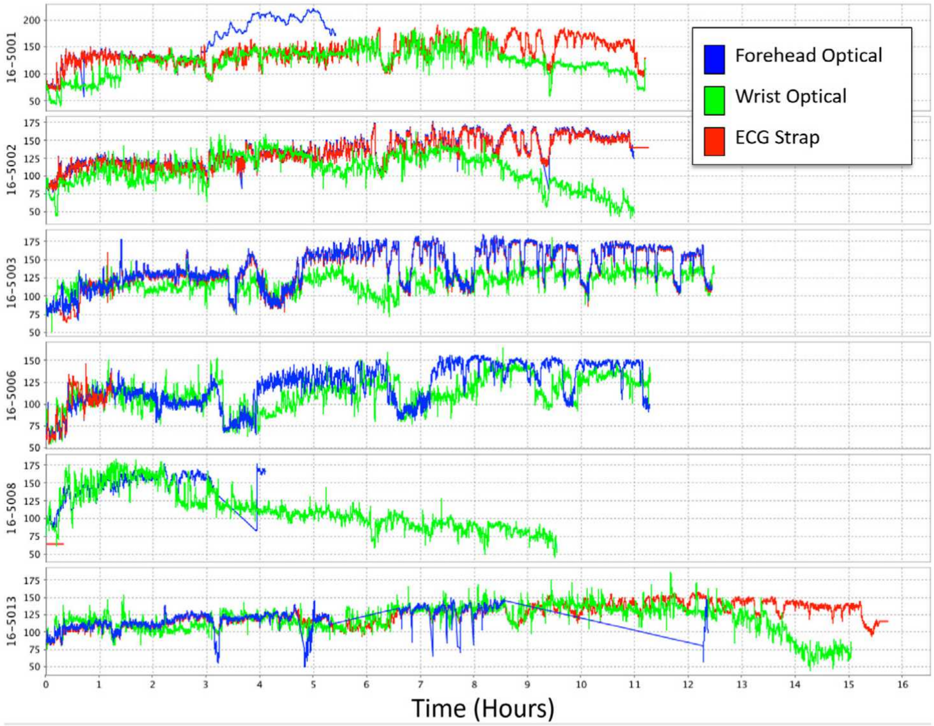

Fig. 7. Parallel recordings of heart rate using 3 types of sensors on each of 6 participants. Their completion times for the hike range from 9.5 to 15.5 h.

Table 1
Demographic information for initial three data collection events for WATCH study.

Count	150 hikers
Type	67.3% civilian
Gender	56.0% male
Age	40.8 years (*stdev* = *10.4*)
Elevation (residence)[a]	446.0 m (*stdev* = *652.4*)
Weight[b]	76.5 kg (*stdev* = *13.0*)
SpO2[b]	95.1% (*stdev* = *4.7*)
Heart Rate[b]	78.0 bpm (*stdev* = *4.7*)
Sleep (previous night)	6.3 hrs (*stdev* = *1.3*)
Longest distance (prior to hike)	51.4 km (*stdev* = *37.2*)
R2R (completed previously)	35.6% yes

[a]Elevation along R2R trail ranges from 750 to 2500 m.
[b]Measured right before hike.

16-5013 (bottom row), where the blue curve contains long straight segments that are interpolated when no samples are recorded. (The exception is for Subject 16-5001, top row, with a reported heart rate of over 200 for two straight hours, which is not plausible and which disagrees with the other two sensors.) In contrast, the wrist optical data (green) unfortunately reports values that appear plausible, but are inconsistent with the other two sensors (Table 3). Moreover, the wrist optical data becomes steadily biased downwards later in the hike (the green curve lying below the other two by a growing amount).

It is not entirely clear whether these failure modes are tied to sensing modality, or simply artifacts of different processing methods on the specific models of devices used. It is better to report no value than make a bad guess, but the wrist sensor may have the weakest signal to work with because the wristwatch does not fit as snugly as a hat or chest strap, and the wrist constantly accelerates back and forth through the motion of walking. If these are indeed factors contributing to the errors we encountered, then accurate sensing of pulse at the wrist may be extremely difficult to achieve. Another (unproven) explanation of the 'drooping' heart rate reporting by the wrist device is that it may not adequately regulate the intensity of the LED (to illuminate the wrist) as its battery discharges. This would be a design flaw more amenable to improvement in future models of such devices.

Table 2
Functionality and counts of items in packages of wearable devices sent across the Grand Canyon.

Functionality	Count
Optical HR (watch)	146
Optical HR (hat)	42
ECG HR (chest strap)	69
Barometric Altimeter (watch)	57
Altimeter (watch)	98
Accelerometer (watch)	155
GPS	154
Environment Sensor	164
Mobile (apps)	150
Footpod	57
Body Temperature	60
EMG (shorts)	11

Table 3
Summary of wrist and forehead sample count and error.

Device type	Number of samples	Median unsigned error	Mean signed error (Bias)	Root mean squared error
Wrist	346,316	12.0	−12.2	27.6
Forehead	108,393	2.0	3.4	13.1

7. Conclusion

Overall, it is important to consider different aspects when choosing wearable technologies: the environment in which they will be used, the performed activity or activities, the level of accuracy required, how many devices need to be worn to achieve the quality and amount of data desired, and if the devices worn will actually measure what they say they can measure. This is especially true when using wearables in extreme environments because devices are exposed to conditions that may hinder the use and quality of the data collected. These factors are important in theory, but they could be a matter of dire consequence when applied to extreme environments.

Our research in the Grand Canyon empirically demonstrates how wearable devices that perform well in everyday fitness activities break down and provide faulty data when in harsh conditions. Extended duration of the R2R hike, switchbacks, battery limitations, altitude, temperature change, and other environmental factors impact the quality of data as well as how much trust one can put in their wearable devices. For example, one hiker in the WATCH study explicitly used GPS on a smartwatch to time his water breaks. Unbeknownst to the hiker, the GPS was off by about 8 km and he drank all of his water after the last water station, thinking he was very close to the end of the hike. He then had to climb the steep switchbacks without any water and was not in a good physiological or cognitive state when he reached the check-in tent at the end of the hike. Another hiker associated with WATCH, received a message from her wearable device 10 min into the hike indicating that she "will fail today". The profile stored in the device was based on an individual who regularly went on long runs. When the individual started the hike at a slower pace and different cadence, this threw the profile estimates off, which in turn lead the wearable device to predict a sharp decrease in performance. However, while the prediction was based on physiological data, the impact to this individual's confidence, motivation, and self-esteem could not be quantified, although surely affected with such a negative message.

Misleading or inaccurate data and feedback from a device is not just a nuisance—it can have dire effects on performance outcomes. Relying on a faulty GPS signal or elevation measure (e.g., without a barometric altimeter) could lead to a hiker becoming lost in the wilderness if he were relying solely on the validity of that information. This highlights the point that while wearable technologies are constantly improving and have many benefits, it is wise to still supplement them with redundant or more reliable information (e.g., topographical map and compass). On a related note, as available wearable technologies expand, it is important to verify that the reputed features actually provide the promised information. For example, cutting edge EMG clothing may be designed to give detailed information on muscle activity for a range of activities but only hold up well in lab conditions and not the activity the user had in mind. We propose that a hopeful but ultimately evidence-driven mindset is ideally suited for interacting with today's wearable technology market.

Although wearable data has been found to be highly useful for augmenting behavior to maintain and enhance performance, the data does not provide the whole story. Cognitive performance is another aspect that is fully engaged during physical activity and extreme fatigue. Memory, accurate decision-making, reaction time, attention, and cognitive executive functions are impaired by fatigue and stress [15,18] but required for performance. However, the connection between cognitive state and physiological performance is difficult to quantify. In the WATCH study, cognitive data was collected from participants taking a cognitive "game" measuring focused attention and short term memory on an iPod Touch every 5 miles [10], but this data is limited and low-fidelity compared to the real-time physiological data collected by wearable devices. And the statistical findings have not been as strong as hoped due to the inability to capture cognitive activity in real time. This is a significant area of opportunity where wearable technologies should expand. The ability to non-invasively and continuously measure neurocognitive activity would further help to quantify human performance. There is still room to understand how physiology and neurocognition interact, especially in extreme environments. Capturing both data streams could help to enhance human performance as well as anticipate consequential health events. As initially tested in the WATCH study, we know that even basic cognitive activity relates to physiological performance and can serve as a useful measure of performance.

Overall, the use of wearable technologies presents an opportunity to enhance human performance. These devices are being further developed to serve multiple functions, from every day, common health monitoring to applications in intense conditions. However, challenges remain regarding the use of wearables for physiological monitoring in extreme contexts. Devices must obtain valid, useful data and remain powered for long periods of time. They must be robust enough to withstand extreme environments, multiple types of terrains, temperature swings, and variable climates. They must also enhance human performance without added distraction, weight, or discomfort.

As the wearable technology industry continues to grow, there is a need to develop hardware and software features to deal with the challenges and limitations that devices face, especially in extreme environments. We hope that we will not only see better devices, but different types of devices. We hope for wearable technologies to take on a new look and form, collecting not only physiological data but also neurocognitive, biological, and other data streams that holistically quantify human performance. As the need to gather real-time data increases, the goal is for these devices to collect and analyze all data streams while performing an activity to enhance performance as well as mitigate consequential health events.

Acknowledgments

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. Funding for Wearables At The Canyon for Health is provided by the Defense Threat Reduction Agency, Chemical and Biological Defense Program (CB10359).

This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

Approved for unlimited release, SAND2018-0866B.

Conflict of interest

The authors confirm there are no conflicts of interest.

References

[1] P. Lamkin, Wearable tech market to be worth $34 billion by 2020 [Internet] 2016. Forbes, Tech Available from: https://www.forbes.com/sites/paullamkin/2016/02/17/wearable-tech-market-to-be-worth-34-billion-by-2020/#1bda01b73cb5. (Accessed 12 December 2017).

[2] C. Mouza, Strengthening the impact, novelty and diversity of research on technology and teacher education, Contemp. Issues Technol. Teacher Educ. 17 (2017) 154–159.

[3] I. Carpinella, D. Cattaneo, G. Bonora, T. Bowman, L. Martina, A. Montesano, M. Ferrarin, Wearable sensor-based biofeedback training for balance and gait in parkinson disease: a pilot randomized controlled trial, Arch. Phys. Med. Rehabil. 98 (2017) 622–630, http://dx.doi.org/10.1016/j.apmr.2016.11.003.

[4] K. Karola, M. Cho, D. Magnus, Consent and engagement, security, and authentic living using wearable and mobile health technology, Nature Biotechnol. 35 (2017) 617, http://dx.doi.org/10.1038/nbt.3887.

[5] J. Pevnick, K. Birkeland, R. Zimmer, Y. Elad, I. Kedan, Wearable technology for cardiology: An update and framework for the future, Trends Cardiovasc. Med. 28 (2018) 144–150, http://dx.doi.org/10.1016/j.tcm.2017.08.003.

[6] N. Newbutt, C. Sung, H. Kuo, M. Leahy, The acceptance, challenges, and future applications of wearable technology and virtual reality to support people with autism spectrum disorders, in: Recent Advances in Technologies for Inclusive Well-Being, Springer, Cham, 2017, pp. 221–241, http://dx.doi.org/10.1007/978-3-319-49879-9_11.

[7] J. Revesencio, Exploring the benefits of wearable technology [Internet]. Huffington Post. 2017. Available from https://www.huffingtonpost.com/jonha-revesencio/exploring-the-benefits-of_b_7910662.html. (Accessed 12 December 2017).

[8] L. Bell, Are 'invisibles' the next wearables? Smart apparel, shoes, and jewelry, the future is integrated [Internet] 2017. Forbes. Available from: https://www.forbes.com/sites/leebelltech/2017/10/31/are-invisibles-the-next-wearables/#6c6b643520f8. (Accessed 11 December 2017).

[9] P. Düking, F.K. Fuss, H.C. Holmberg, B. Sperlich, Recommendations for assessment of the reliability, sensitivity, and validity of data provided by wearable sensors designed for monitoring physical activity, JMIR MHealth UHealth 6 (4) (2018) http://dx.doi.org/10.2196/mhealth.9341.

[10] K.R. Evenson, M.M. Goto, R.D. Furberg, Systematic review of the validity and reliability of consumer-wearable activity trackers, Int. J. Behav. Nutrit. Phys. Act. 12 (1) (2015) 159.

[11] LexisNexis Risk Solutions, Millennial study: Privacy vs. customer experience [Internet]. 2017. Available from: https://www.lexisnexis.com/risk/downloads/news/Millennials-Global-Summary.pdf. (Accessed 13 December 2017).

[12] M. Ghiglieri, T. Myers, Over the edge: Death in the Grand Canyon, Puma Press, Flagstaff, 2001.

[13] G. Emmanuel-Aviña, R. Abbott, C. Anderson-Bergman, C. Branda, K.M. Divis, L. Jelinkova, V. Newton, E. Pearce, J. Femling, Rim-to-Rim wearables at the canyon for health (R2R WATCH): Experimental design and methodology, in: International Conference on Augmented Cognition, 9–13 2017, Springer, Cham, 2017, pp. 263–274.

[14] K. Divis, C. Anderson-Bergman, R. Abbott, V. Newton, G. Emmanuel-Aviña, Physiological and cognitive factors related to human performance during the Grand Canyon Rim-to-Rim hike, Hum. Perform. Extreme Environ. (2018) 14, http://dx.doi.org/10.7771/2327-2937.1095.

[15] L.E. Bourne, R.A. Yaroush, Stress and Cognition: A Cognitive Psychological Perspective. (NAG2-1561), National Aeronautics and Space Administration, 2003.

[16] P. Ranacher, R. Brunauer, W. Trutschnig, S. VanderSpek, S. Reich, Why GPS makes distances bigger than they are, Int. J. Geogr. Inf. Sci. 30 (2) (2015) 316–333, http://dx.doi.org/10.1080/13658816.2015.1086924.

[17] National Park Services. Plan your visit image [Internet]. Available from https://www.nps.gov/grca/planyourvisit/trail-distances.htm. (Accessed 04 January 2018).

[18] I.N. Karatsoreos, B.S. McEwen, Stress and allostatis, in: Handbook of Behavioral Medicine: Methods and Applications, Springer, New York, NY, 2010, pp. 649–658.