SAND2018-0866B

Wearable Technologies in Extreme Environments

Glory Emmanuel-Aviña[1], Kristin M. Divis[2], Robert G. Abbott[2]

[1]Sandia National Laboratories-California, [2]Sandia National Laboratories-New Mexico

Corresponding author: Glory Emmanuel-Aviña, gremman@sandia.gov

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 just measuring physiological data, both with the hardware as well as the data itself. While physiological markers (e.g., heart rate, cadence, temperature) are useful, neurocognitive markers such as response time, accuracy, and focused attention are also useful indicators of health and performance. Combining physiological data with other data streams, such as environmental, neurocognitive, and biological data, could better quantify early signs of health decrements as well as prevent catastrophic health events.

In this chapter, we focus on the challenges and lessons learned of measuring physiology in extreme environments. We present empirical findings from a study we lead 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 Grand Canyon. We also present the opportunity wearables have to quantify performance and fatigue by developing wearable technologies that collect and analyze, in real-time, neurocognition and biomarkers.

Table of Contents

The benefits of wearable technologies

The use of wearable devices has exploded over the past decade, especially over the past five years. We are now able to quantify on our productivity by wearing wearable devices while being physically active. We can use this data to measure our performance, work towards personal records, and prevent overexertion. The wearable technology industry alone is approximately a 20 billion dollar enterprise with new smart wearables coming out every day[1]. According to Forbes, the wearable industry will double in the next five years and be worth $34 billion in 2020, selling about 411 million wearable smart devices[1]. Wearable technologies come in various forms, from smart watches to bioharnesses to smart clothing. We can now collect a wealth of performance data without devices interfering with our activities.

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 education[2], cognitive research[3], healthy living[4], medicine[5], and mental health,[6] to name a few. The Huffington Post lists the improvement to personal safety, the ability to track health and fitness, and staying connected at all times as the major benefits of wearables[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.

Although there are many benefits to wearable technologies, more investment and research will continue to pour into enhancing wearable technologies. We realize the benefits of wearables, however, we are far from having technologies that serve us exactly the way needed. The following section provides a generalized overview of wearable technology limitations. We focus on the use of wearables that collect physiological data, but also present the need to collect neurocognitive data (defined in its relative section). These challenges highlight the existing potential in the wearable technology space as well as the long road ahead for research, development, and innovation.

The limitations of physiological wearable technologies

The constant release of new wearable technologies highlights that there are limitations to what they can do. Users will continue to upgrade devices as well as be unsatisfied with device hardware, the

---

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

[2] Mouza, C. (2017). Strengthening the Impact, Novelty and Diversity of Research on Technology and Teacher Education. *Contemporary Issues in Technology and Teacher Education*, *17*(2), 154-159.

[3] Carpinella, I., Cattaneo, D., Bonora, G., Bowman, T., Martina, L., Montesano, A., & Ferrarin, M. (2017). Wearable Sensor-Based Biofeedback Training for Balance and Gait in Parkinson Disease: A Pilot Randomized Controlled Trial. *Archives of physical medicine and rehabilitation*, *98*(4), 622-630.

[4] Kreitmair, Karola V., Mildred K. Cho, and David C. Magnus. "Consent and engagement, security, and authentic living using wearable and mobile health technology." *Nature biotechnology* 35 (2017): 617.

[5] Pevnick, J. M., Birkeland, K., Zimmer, R., Elad, Y., & Kedan, I. (2017). Wearable technology for cardiology: An update and framework for the future. *Trends in cardiovascular medicine*.

[6] Newbutt, N., Sung, C., Kuo, H. J., & Leahy, M. J. (2017). 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* (pp. 221-241). Springer International Publishing.

[7] Revesencio, J. (2015). Exploring the benefits of wearable technology. *Huffington Post.* Accessed online December 12, 2017 at https://www.huffingtonpost.com/jonha-revesencio/exploring-the-benefits-of_b_7910662.html

data collected by devices, how the device works with particular activities, the data collected, 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 smartwatches 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 and therefore can 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. Inaccurate data leads to uninformed decision-making. Regarding validity, if a watch captures Global Positioning System (GPS) but struggles with running switchbacks, under bridges, up an incline, et cetera, then a user may be running 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 usefully alert the user or 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, foot pod, 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[9].

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 don wearables to provide the earliest predictor possible for a health event, advance performance, and achieve goals, but the industry continues to search for innovative ways to collect human performance data. We want more data; more useful data. There are opportunities to merge physiological data with neurocognitive data to better understand the mind-body connection.

[8] Bell, L. (2017). Are 'invisibles' the next wearables? Smart apparel, shoes, and jewelry, the future is integrated.
[9] LexisNexis Risk Solutions (2016). Millennial study: Privacy vs. customer experience. Accessed online December 14, 2017 at https://www.lexisnexis.com/risk/downloads/news/Millennials-Global-Summary.pdf.

Collecting neurocognitive data

Physiological data reigns supreme in the wearable technologies world, but they are not the only data streams of importance when measuring performance. Neurocognitive data, which is how the mind and brain are working while performing and completing a task, is of increasing interest. How does the current activity impact decision-making? How about memory or the ability to inhibit maladaptive responses? Reaction time or level of alertness? How is the mind, brain, and body working together during an activity? All of these questions have underlying neurocognitive components.

Passively measuring neurocognitive abilities is no small challenge. In current research studies, participants are measured before and after performing a task or they are required to stop and take a cognitive battery while performing an activity. It is very difficult to passively capture neurocognitive data. Portable, wearable eye trackers have the potential to lend insight via gaze location, blink rate, and/or pupil dilation. Similarly, portable, reduced-electrode electroencephalogram (EEG) devices might lend insight into brain function. However, these devices have been limited to lab environments and/or activities that have restricted movements (e.g., gaming).

The two largest limitations when considering the application of neurocognitive wearable devices to extreme environments is their ability to provide useful, actionable data and their ability to withstand the environmental challenges of uncontrolled, rougher settings. Will using one of these devices provide information that could not be obtained in a simpler matter (e.g., from another device)? If the data is novel, is it data that can reasonably be acted upon (e.g., limited noise in the data and reasonable post-processing steps)? Will the device function properly when not in the pampered setting of a laboratory?

Eye trackers are often used to gain insight into underlying mental processes, especially attention via gaze location[10]. Much like the wearable technologies field, the eye tracking community is booming, with lightweight, increasingly portable machines becoming more accessible and numerous than ever before (e.g., eye tracking "glasses"). However, these devices tend to provide lower accuracy, precision, and sampling rates than those traditionally used in laboratory settings, which translates to noisier data. Most applications to extreme environments would also suffer from an overabundance of data. One could track gaze location at 60 Hz for a trek up a mountain, but the post-processing involved in coding the data from that uncontrolled setting and attempting to correct for drift would make even the most stoic of research assistants cringe (let alone the lack of usefulness for real-time feedback). Pupil dilation has been linked to neurocognitive factors such as cognitive load[11], but pupil dilation is heavily dependent upon light exposure, which is challenging to properly account for in outdoor environments. Perhaps the most promising and easily implemented metric is measuring eye closure or blink rate to inform levels of fatigue[12], which could be done without the need to code the outside environment.

Portable EEG machines that are quick to set up and have a reduced number of electrodes are now readily available. In relatively uncontrolled settings, these machines are best used for determining general alertness or activity (e.g., the ratio of alpha- to beta-band activity), which has been linked to

[10] Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

[11] Laeng, B., Sirois, S., Gredebäck, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science*, *7*(1), 18-27.

[12] Stern, J. A., Boyer, D., & Schroeder, D. (1994). Blink rate: a possible measure of fatigue. *Human Factors*, *36*(2), 285-297.

fatigue during intense activity in laboratory environments[13]. EEG studies are traditionally run in laboratory settings for multiple reasons: the equipment tends to be quite delicate (and would not stand up well to harsh environments) and the cleanest data comes from settings where activity is limited—even seemingly small movements such as blinking can disrupt the recorded signal. Sweat—a common and expected variable in extreme environments—can severely compromise the data by creating a bridge across multiple electrodes. With that said, EEG devices might first be applied as a tool in extreme environments when relatively coarse data is targeted (e.g., power band analysis rather than event related potentials) and is recorded for set periods of the activity (perhaps triggered by a signal from other wearable devices) rather than the entire activity.

Other wearable devices with a neurocognitive slant such as transcranial direct-current stimulation (tDCS) are becoming increasingly popular. tDCS has been linked to improvements in working memory capacity[14] and is becoming more prevalent in the gaming community. Application of tDCS as part of a wearable device in extreme settings should be considered with caution, as the rough environment could lead to negative effects due to misplaced electrodes or lack of effect due to bridging between electrodes via sweat.

Empirical study informing wearable use

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. 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 empirical study funded by the Defense Threat Reduction Agency (Project number CB10359) that uses wearable devices to track performance while participants complete the grueling 38 kilometers, Rim-to-Rim (R2R) hike at the Grand Canyon. In addition to providing physiological, cognitive, and biological data to predict performance decrement, the WATCH study is also a natural testbed for evaluating wearable device performance in an extreme environment[15,16].

The Grand Canyon R2R is not a typical walk in the park—the 38 kilometer 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 Figure 1). The trail includes elevation changes of approximately one mile going down into the canyon and then again 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

---

[13] Ftaiti, F., Kacem, A., Jaidane, N., Tabka, Z, & Dogui, M. (2010). Changes in EEG activity before and after exhaustive exercise in sedentary women in neutral and hot environments. *Applied Ergonomics*, *41*, 806-811.

[14] Ruf, S. P., Fallgatter, A. J., & Plewnia, C. (2017). Augmentation of working memory training by transcranial direct current stimulation (tDCS). *Scientific Reports, 7*(1), 876.

[15] Emmanuel-Aviña, G., Abbott, R., Anderson-Bergman, C., Branda, C., Divis, K. M., Jelinkova, L., Newton, V., Pearce, E., Femling, J. (2017). Rim-to-Rim Wearables At The Canyon for Health (R2R WATCH): Experimental design and methodology. *Augmented Cognition, Neurocognition, & Machine Learning*, 263-274.

[16] Divis, K. M., Anderson-Bergman, C., Abbott, R., Newton, V. E., & Emmanuel-Aviña, G. (2017, in press). Physiological and cognitive factors related to human performance during the Grand Canyon Rim-to-Rim hike. *Human Performance in Extreme Environments.*

complete the R2R in a single day due to the challenging environment. Over 250 hikers are airlifted out of the canyon each year, many with heat-stress related symptoms[17].

Figure 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.

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 hours, 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 boiling temperatures. 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.

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

[17] Ghiglieri, M. P., & Myers, T. M. (2001). *Over the edge: Death in the Grand Canyon*. Flagstaff, AZ: Puma Press.

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.

Empirical findings of issues and concerns with wearables

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.

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 Global Positioning System (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.

Figure 2 shows the GPS data from 173 hikers on the North Kaibab Trail in the Grand Canyon. It is a distance of 8 kilometers 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.

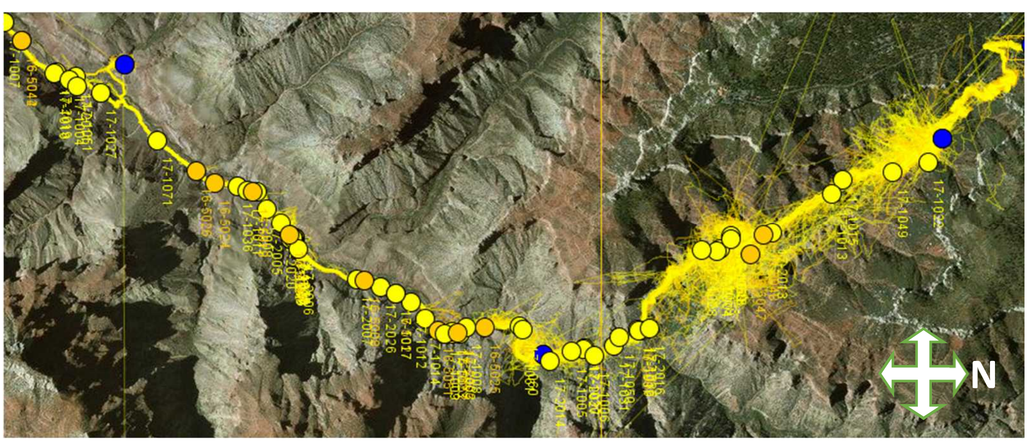

Figure 2: GPS Tracks from 173 Hikers on the North Kaibab Trail in the Grand Canyon. North is to the right. The clouds of yellow at some points along the trail indicate locations where GPS reception is poor. Overhead imagery courtesy of Bing Maps.

Figure 3 is a histogram of the lengths of the 173 GPS tracks. 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 46.2 km, whereas the length of the Kaibab Trail according to the National Park Service is 33.7 km[18]. Track lengths should be marginally more than the trail length, since any deviation from trail can only add to the track length (e.g. using an outhouse, or the Ribbon Falls detour indicated by the leftmost blue dot in Figure 2), but most of the extra track length is due to tracking error. GPS tracking error almost always increases the track length, but could also reduce it slightly by cutting corners and switchbacks that exist in the trail. The mean track length is 47.7 km and the standard deviation 11.1.

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

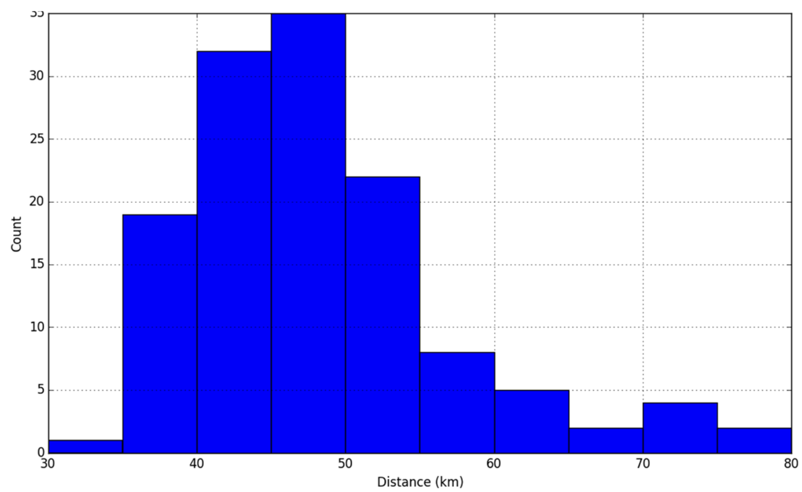

Figure 3: Histogram of GPS track lengths for 134 hikers on R2R Hike. The wide variability in track length implies that other measures derived from GPS (e.g. speed) will not be reliable. The median track length is 46.2 km, whereas the actual length of the Kaibab trail is 33.7 km.

To identify breaks, we are using 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. Figure 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 and all the sensors are in fairly close agreement. But walking uphill, cadence drops to approximately 44 and the sensors undergo softer accelerations. The wrist-based sensors both exhibit more noise, with one brand especially frequently and incorrectly dropping to zero. As a result of this test, we have integrated footpods into our device packages.

[18] https://www.nps.gov/grca/planyourvisit/trail-distances.htm

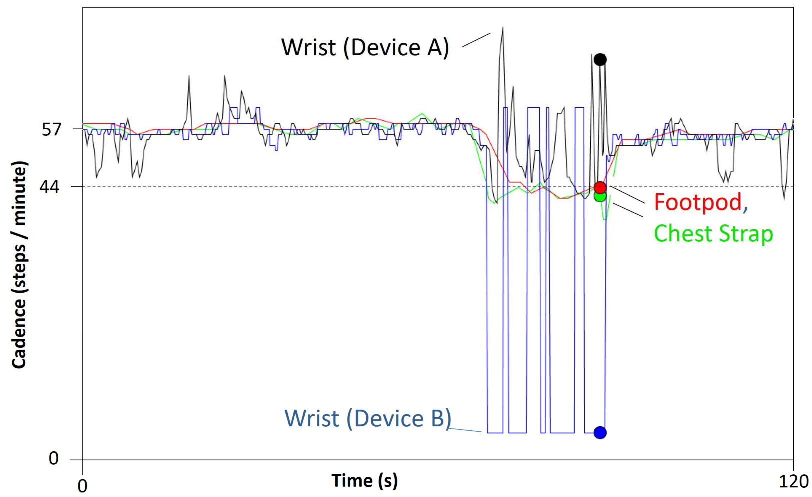

Figure 4: Cadence data from 4 sensors on the same individual during a short hike.

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 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 the latitude/longitude, even with clear satellite reception. Figure 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 meters) every couple of minutes. Such errors greatly disrupt calculations of slope and total distance climbed, and therefore effort expended.

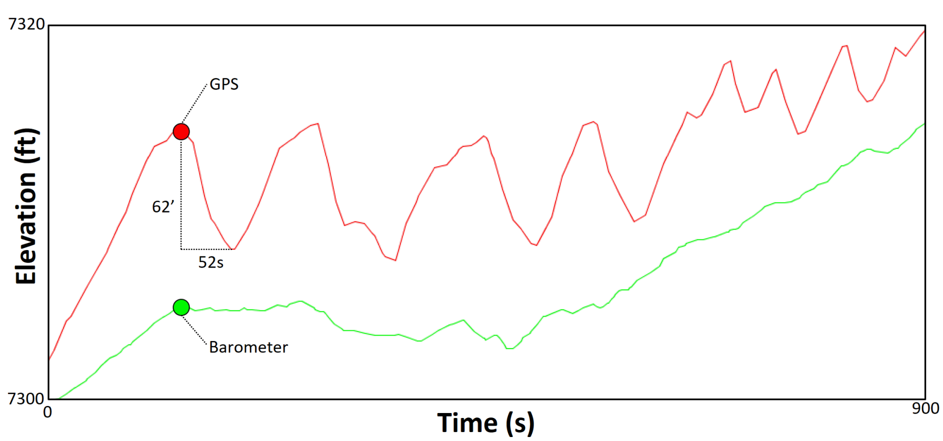

Figure 5: Fifteen minutes of simultaneous elevation readings from GPS vs Barometric altimeters. The subject wearing the altimeters was on a steadily ascending trail, consistent with the barometric altimeter reading. The GPS elevation reading repeatedly oscillates by approximately 15 meters each two minutes.

Figure 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 second 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 altimeters are much more consistent (standard deviation = 492 m). The median total climb recorded is 7044m for GPS and 3363 for barometric. For reference, the vertical distance from the Phantom Ranch to the North Rim is 1736m according to the National Park Service. 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.

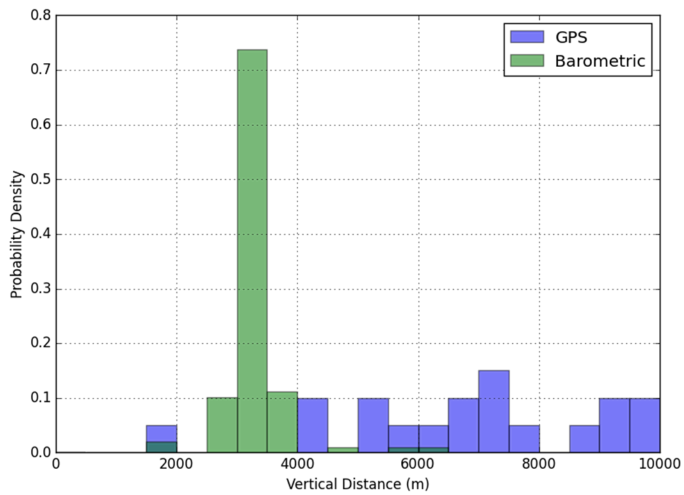

Figure 6: Histogram of total vertical ascent of 119 GPS tracks on R2R Hike. Twenty tracks were collected from GPS, the remaining 99 from Barometric altimeters. Approximately 75% of the barometric altimeters (in green) accumulated between 3000 m and 3500 m, whereas the GPS altimeters (in blue) varied widely.

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. Figure 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 record 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 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. 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).

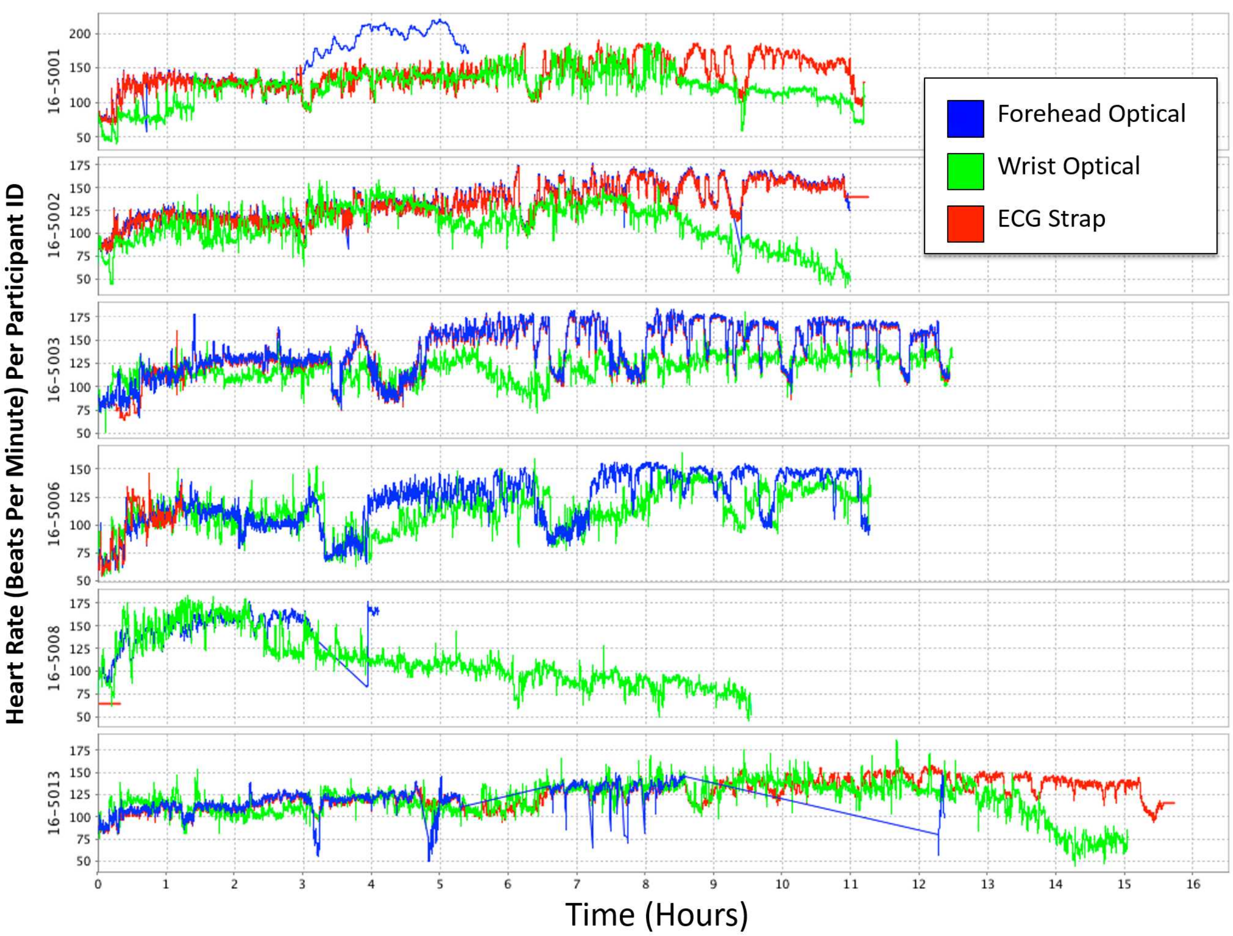

Figure 7: Parallel recordings of heart rate using 3 types of sensors on each of 6 participants. Their completion times for the hike range from 9.5 to 15.5 hours.

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.

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

Path forward for wearable technologies

Overall, it is important to consider different aspects when choosing wearable technologies: the environment in which they'll be used, the performed activity or activities, the level of accuracy required, how many devices needed 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. For example, one hiker in the WATCH study explicitly used GPS on a smartwatch to time their water breaks. Unbeknownst to the hiker, the GPS was off by about 5 miles 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 minutes into the hike indicating that she was "had a high risk of failure." The profile stored in the device was based on this individual going on long runs, not intense hikes, which threw the profile estimates off. A message like that, especially for an individual first attempting a new activity, does affect one's mental performance for the day.

Although we say mental performance is affected by physiological metrics and performance, this is very difficult to quantify. 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.

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 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.

Acknowledgements

The funding for the Wearables At The Canyon for Health (WATCH) study is provided by the Defense Threat Reduction Agency, project number CB10359.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and 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. SAND NO. 2017-XXXXP

References (in order of appearance)

1. Lamkin, P. (2016). Wearable tech market to be worth $34 billion by 2020. Forbes, Tech. Accessed online on December 12, 2017 at https://www.forbes.com/sites/paullamkin/2016/02/17/wearable-tech-market-to-be-worth-34-billion-by-2020/#1bda01b73cb5
2. Mouza, C. (2017). Strengthening the Impact, Novelty and Diversity of Research on Technology and Teacher Education. Contemporary Issues in Technology and Teacher Education, 17(2), 154-159.
3. Carpinella, I., Cattaneo, D., Bonora, G., Bowman, T., Martina, L., Montesano, A., & Ferrarin, M. (2017). Wearable Sensor-Based Biofeedback Training for Balance and Gait in Parkinson Disease: A Pilot Randomized Controlled Trial. Archives of physical medicine and rehabilitation, 98(4), 622-630.
4. Kreitmair, Karola V., Mildred K. Cho, and David C. Magnus. "Consent and engagement, security, and authentic living using wearable and mobile health technology." Nature biotechnology 35 (2017): 617.
5. Pevnick, J. M., Birkeland, K., Zimmer, R., Elad, Y., & Kedan, I. (2017). Wearable technology for cardiology: An update and framework for the future. Trends in cardiovascular medicine.
6. Newbutt, N., Sung, C., Kuo, H. J., & Leahy, M. J. (2017). 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 (pp. 221-241). Springer International Publishing.
7. Revesencio, J. (2015). Exploring the benefits of wearable technology. Huffington Post. Accessed online December 12, 2017 at https://www.huffingtonpost.com/jonha-revesencio/exploring-the-benefits-of_b_7910662.html
8. Bell, L. (2017). Are ‘invisibles’ the next wearables? Smart apparel, shoes, and jewelry, the future is integrated.
9. LexisNexis Risk Solutions (2016). Millennial study: Privacy vs. customer experience. Accessed online December 14, 2017 at https://www.lexisnexis.com/risk/downloads/news/Millennials-Global-Summary.pdf.
10. Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). Eye tracking: A comprehensive guide to methods and measures. OUP Oxford.
11. Laeng, B., Sirois, S., Gredebäck, G. (2012). Pupillometry: A window to the preconscious? Perspectives on Psychological Science, 7(1), 18-27.
12. Stern, J. A., Boyer, D., & Schroeder, D. (1994). Blink rate: a possible measure of fatigue. Human Factors, 36(2), 285-297.
13. Ftaiti, F., Kacem, A., Jaidane, N., Tabka, Z, & Dogui, M. (2010). Changes in EEG activity before and after exhaustive exercise in sedentary women in neutral and hot environments. Applied Ergonomics, 41, 806-811.
14. Ruf, S. P., Fallgatter, A. J., & Plewnia, C. (2017). Augmentation of working memory training by transcranial direct current stimulation (tDCS). Scientific Reports, 7(1), 876.
15. Emmanuel-Aviña, G., Abbott, R., Anderson-Bergman, C., Branda, C., Divis, K. M., Jelinkova, L., Newton, V., Pearce, E., Femling, J. (2017). Rim-to-Rim Wearables At The Canyon for Health (R2R

WATCH): Experimental design and methodology. Augmented Cognition, Neurocognition, & Machine Learning, 263-274.

16. Divis, K. M., Anderson-Bergman, C., Abbott, R., Newton, V. E., & Emmanuel-Aviña, G. (2017, in press). Physiological and cognitive factors related to human performance during the Grand Canyon Rim-to-Rim hike. Human Performance in Extreme Environments.

17. Ghiglieri, M. P., & Myers, T. M. (2001). Over the edge: Death in the Grand Canyon. Flagstaff, AZ: Puma Press.

18. National Park Services. Plan your visit image. Accessed January 4, 2018 at https://www.nps.gov/grca/planyourvisit/trail-distances.htm