



HHS Public Access

Author manuscript

J Meas Phys Behav. Author manuscript; available in PMC 2021 July 23.

Published in final edited form as:

J Meas Phys Behav. 2019 December ; 2(4): 268–281. doi:10.1123/jmpb.2018-0068.

An Open-Source Monitor-Independent Movement Summary for Accelerometer Data Processing

Dinesh John, Qu Tang

Northeastern University

Fahd Albinali

QMedic Medical Alert Systems

Stephen Intille

Northeastern University

Abstract

Background: Physical behavior researchers using motion sensors often use acceleration summaries to visualize, clean, and interpret data. Such output is dependent on device specifications (e.g., dynamic range, sampling rate) and/or are proprietary, which invalidate cross-study comparison of findings when using different devices. This limits flexibility in selecting devices to measure physical activity, sedentary behavior, and sleep.

Purpose: Develop an open-source, universal acceleration summary metric that accounts for discrepancies in raw data among research and consumer devices.

Methods: We used signal processing techniques to generate a Monitor-Independent Movement Summary unit (MIMS-unit) optimized to capture normal human motion. Methodological steps included raw signal harmonization to eliminate inter-device variability (e.g., dynamic g-range, sampling rate), bandpass filtering (0.2–5.0 Hz) to eliminate non-human movement, and signal aggregation to reduce data to simplify visualization and summarization. We examined the consistency of MIMS-units using orbital shaker testing on eight accelerometers with varying dynamic range (± 2 to ± 8 g) and sampling rates (20–100 Hz), and human data ($N = 60$) from an ActiGraph GT9X.

Results: During shaker testing, MIMS-units yielded lower between-device coefficient of variations than proprietary ActiGraph and ENMO acceleration summaries. Unlike the widely used ActiGraph activity counts, MIMS-units were sensitive in detecting subtle wrist movements during sedentary behaviors.

Conclusions: Open-source MIMS-units may provide a means to summarize high-resolution raw data in a device-independent manner, thereby increasing standardization of data cleaning and analytical procedures to estimate selected attributes of physical behavior across studies.

Keywords

activity count; activity monitor; physical activity measurement

Accelerometer-based devices offer unprecedented potential to study the impact of human physical behavior patterns on health (Doherty et al., 2017; Wright, Brown, Collier, & Sandberg, 2017). Thus, the National Institutes of Health (NIH) and the Food and Drug Administration aim to advance “digital health,” the scope of which includes wearable and mobile health technologies (U.S. Food and Drug Administration, 2017). Despite the promise, procedural variability in data collection, processing, and analyses impedes progress (Wijndaele et al., 2015). Experts in physical activity monitoring have emphasized the need to standardize data collection, processing, and analyses using raw acceleration data and non-proprietary algorithms (Wijndaele et al., 2015). However, inconsistency and transparency among popular research-grade devices (John, Morton, Arguello, Lyden, & Bassett, 2018; John, Sasaki, Hickey, Mavilia, & Freedson, 2014; Lee, Macfarlane, & Cerin, 2010; Sasaki, John, & Freedson, 2011), and the lack of uniformity among devices that are typically targeted at end-consumers (e.g., smartwatches/phones/fitness monitors), contribute to inter-device output variability. Proprietary considerations further complicate efforts to permit device-agnostic data collection and analyses.

Access to raw data may allow researchers to interchangeably use consumer and research-grade devices and enable valid, cross-study comparisons of findings. For example, Fitbit, which is the market leader in wearables, now allows access to raw data from their newer smartwatch devices via a software development kit (Accelerometer sensor guide, 2019). Consumer-oriented devices such as smartwatches and phones can now store high resolution raw data and offer new possibilities for researchers (Althoff et al., 2017; Wang et al., 2015). Device manufacturers, however, make different engineering tradeoffs to optimize device performance based on considerations such as the purpose of the device (e.g., detecting steps vs. sitting), battery-life, and raw data-storage issues associated with data compression or wireless transmission. Such variability changes the raw data and limits subsequent inter-device comparison of both raw and summary output. One key factor that causes inter and intra-device inconsistency in raw and summarized acceleration data is variability in accelerometer dynamic range among devices (e.g., ± 2 to ± 16 g). Low-range sensors may not capture signals necessary to optimally characterize moderate to vigorous activities. Other factors include the resolution of the sensor (e.g., 8 vs. 12-bit) and sampling rate (e.g., 10 to 100 Hz). Inter-device output heterogeneity may increase device-dependency, impact cross-study uniformity in accelerometer data processing, and, thus, limit the interpretation of findings across studies.

Raw accelerometer data processing to detect specific activity type, intensity, and duration is an active area of research (Mannini, Rosenberger, Haskell, Sabatini, & Intille, 2017; Pavey, Gilson, Gomersall, Clark, & Trost, 2017). However, summarized acceleration measures continue to have a place in accelerometer-based research. The use of raw vs. summarized acceleration may be influenced by the duration and type of study, which impact the volume of accelerometer data, its manageability, and the required level of interpretation. For

instance, large-scale epidemiological studies with several thousand participants providing week-long raw acceleration data have used summarized acceleration output to simplify various steps in accelerometer data processing and interpretation (Doherty et al., 2017; Lee et al., 2018). NIH intends to release both summary and raw data from the 2011–12/2013–14 National Health and Nutrition Examination Survey (NHANES). Accelerometer data processing in a study commonly involves data cleaning to establish data integrity and quality, followed by the determination of relationships between physical behaviors and health using cut-points or machine learning algorithms. Data cleaning typically involves visualization and checking of hourly-to-daily chunks of data to establish data integrity and quality (Shiroma, Kamada, Smith, Harris, & Lee, 2015). Using raw acceleration to visualize and clean multiple days of data has low feasibility due to a high burden on human and computing resources arising from the sheer volume of the high sampling rate raw signal. Thus, large epidemiological studies (e.g., Women’s Health Study, NHANES 2011–14) have used acceleration summaries to perform such tasks associated with data cleaning (Lee et al., 2018; Shiroma, Cook, et al., 2015; Shiroma, Kamada, et al., 2015). However, inter-device variability in raw and summary output may yield cross-study differences in monitor wear-time, detection of faulty data (e.g., sensor malfunction), and the detection of valid minutes for analyses.

In addition to physical activity studies that commonly employ batch processing of accelerometer data, another application of summarized acceleration is in the increasingly popular area of just-in-time intervention studies (Nahum-Shani, Hekler, & Spruijt-Metz, 2015; Nahum-Shani et al., 2014). Real to near-real-time automated decisions to gather physical behavior information (e.g., determine behavioral context via ecological momentary assessment) or/and to intervene with motivational behavior change messaging/strategies, have been made using thresholds that indicate active vs. inactive states (Dantzig, Geleijnse, & Halteren, 2013; Hardeman, Houghton, Lane, Jones, & Naughton, 2019). Just-in-time interventions are usually deployed for weeks-to-months, during which they may use small, consumer-grade devices to constantly monitor behavior. Such devices must process raw data in real-time, with low power draw and memory, and often wirelessly transmit movement metrics to partner devices (e.g., phone). Such studies may use “steps” to estimate the volume of physical activity and when steps occur to determine the delivery of real-time cues (Klasnja et al., 2015). However, step-counting algorithms may filter out movement based on temporal and movement frequency considerations (John et al., 2018) and underestimate movement (Mendoza et al., 2019; Toth et al., 2019). Comparatively, acceleration summaries of gross movement may be more inclusive of overall movement and a better indicator of an individual’s current state of physical behavior in real-time studies. Additionally, inter-device differences in how steps are detected cause output variability, which will influence cue delivery (Bassett, Toth, LaMunion, & Crouter, 2016; John et al., 2018).

Much of the current knowledge on relationships between physical activity and health is tied to the approach of using ActiGraph’s acceleration summary thresholds (Wijndaele et al., 2015) and the use of sum total movement derived from summarized acceleration (Wolff-Hughes, Fitzhugh, Bassett, & Churilla, 2015). However, ActiGraph’s popular activity counts has several flaws. This summary metric demonstrates variable relationships with increasing frequency of movement based on wear location (e.g., hip vs. wrist) (John, Tyo, & Bassett,

2010; LaMunion, Bassett, Toth, & Crouter, 2017), yield dissimilar activity-counts for the same signal at variable sampling rates (Brond & Arvidsson, 2016), and may eliminate information just above the accelerometer's noise threshold that may help differentiate low-motion behaviors such as sedentary behavior and sleep from non-wear. Hence, when using acceleration summaries, the ActiGraph activity count may not be optimal to represent movement behavior.

Until the achievement of required improvements in battery life, data processing speed, and high-resolution behavior detection methodologies that allow a convenient use of raw acceleration data by an end-user researcher, the application of summary acceleration metrics to perform different tasks in various types of accelerometer-based physical activity research may continue to serve as a reasonable alternative. This paper describes an open-source acceleration summary that is designed to maximally capture human movement and accounts for key factors that cause variability in raw output, to yield a device-independent universal summary metric that can be used in various stages of accelerometer data processing. Such a summary metric may not only be better suited for the development of methods to improve the distinction of specific outcomes (e.g., sleep/sedentary behavior vs. non-wear) that may not require raw data, but may also enable cross-study comparisons of total movement regardless of the variability introduced by different types of devices (e.g., consumer vs. research-grade), models, and brands.

Methods

We propose a method that first uses digital signal processing techniques to harmonize raw data from devices with different dynamic range and sampling rates, and then aggregates the raw data to yield a *Monitor-Independent Movement Summary unit* (MIMS-unit) that captures normal human motion. We outline key algorithmic considerations, delineate the algorithm step-by-step, and then describe algorithm testing used to optimize algorithm output consistency. Figure 1 illustrates each step of the algorithm. The first row labeled "Raw Data" shows a one-second acceleration signal captured from one axis of a wrist-worn triaxial sensor. The reference "Signal 1" has a dynamic range of ± 8 g at the commonly used sampling rate of 80 Hz. Signals 2, 3, and 4 are manipulated versions of the reference signal and have a lower g-range and sampling rate. Signals 2, 3, and 4 provide a visually distinct representation of how decreasing sampling rate and dynamic range diminishes the smoothness of the curve and cuts off (i.e., "maxes out") the signal, respectively. The signals represent a difficult test condition, containing (i) an artifact from the sensor being tapped on a hard surface (first peak) and (ii) meaningful, but intense motion (the start of a jumping jack).

Key Algorithmic Considerations: Dynamic Range and Frequency Spectrum

Accelerometers capture both voluntary and involuntary human spatial movement, artifacts associated with movement (e.g., vibration from heel/foot strike during ambulation; drug- or disease-related involuntary tremors), and artifacts in the environment (e.g., vibrations while travelling in a car, bumping a sensor against a hard surface). MIMS-units are designed to

capture meaningful human movement that may impact health, while minimizing environmental and movement artifacts via signal filtering.

Peak acceleration during normal human physical activities can exceed 6 g at the wrist and hip (e.g., running and jumping) (Rowlands & Stiles, 2012). Thus, a device with a dynamic range of ± 2 g may underestimate meaningful movement (see raw data for signals 3 and 4 in Figure 1). Rare cases of extreme peak wrist acceleration due to spatial movement during sport include transient (<0.5 s) forces up to 15 g during a bat swing, or 80 g when pitching in baseball (Berkson, Aylward, Zachazewski, Paradiso, & Gill, 2006; King, Hough, McGinnis, & Perkins, 2012). Movements during voluntary physical activities of daily living typically generate average acceleration within ± 2 g (in a single axis). Some activities may generate up to ± 3 g on average, but activity outside an average of ± 4 g from sensors at the wrist and hip may be less frequent (e.g., Supplementary Table 1 [available online]) (Mannini et al., 2017).

Dominant frequencies for most voluntary movements lie between 0.3 and 5 Hz (Supplementary Table 1 [available online]). Small movements when motionless during sitting, standing, and lying may fall below 0.3 Hz; examples of such movements include fluctuations in acceleration due to changes in orientation of devices relative to the ground. Therefore, to retain information on normal human voluntary movement while eliminating other signal artifacts (Supplementary Table 2 [available online]), the MIMS-unit algorithm preserves acceleration signals within a frequency spectrum of 0.2 to 5 Hz.

Stepwise Description of the MIMS-Unit Algorithm

To derive MIMS-units from raw acceleration, interpolation, extrapolation, bandpass filtering, and aggregation are conducted independently for each axis of triaxial acceleration data using empirically determined time and frequency-domain parameters. Processing data for each individual axis is conducted separately to retain some information about the direction of motion, which in conjunction with other information such as the sensor's body location may be useful during data cleaning to infer if sensor response is appropriate for known activities or if the sensor was accidentally misplaced or mis-oriented. Processed data from each axis can then be aggregated into a single motion summary. Access to both individual axis and aggregated triaxial motion summaries provides researchers with greater flexibility in the application of MIMS-units. The MIMS-unit algorithm code and all data used to test and refine the algorithm are publicly available at: <http://mhealthgroup.org/data-mims-unit>.

Interpolation.—Interpolation accounts for inter-device variability in sampling rate. The MIMS-unit algorithm first interpolates data to a consistent sampling rate. Interpolation changes the sampling density of the acceleration signal by fitting a continuous function to discrete points in the signal by using a weighted average of known neighboring samples to estimate unknown/required samples (Figure 1, Step 1). The MIMS-unit algorithm uses a natural cubic spline interpolator (the “spline” function in the standard R package), which generally yields an improved approximation of low-sampling-rate signals over linear interpolation (De Boor, 1978). Our empirical testing found that changing sampling density to 100 Hz increases the robustness of signal reconstruction (described below), over lower sampling rates.

Extrapolation.—Extrapolation accounts for inter-device variability in dynamic range. When detected acceleration exceeds a sensor’s dynamic range, signals are maxed-out and create undesirable artifact patterns in the raw data (Figure 1, Raw Data, Signals 3 and 4). To harmonize raw acceleration output across devices and to obtain a truer representation of actual movement, the algorithm extrapolates maxed-out signals from narrow g-range sensors in three steps.

In Step 1, we estimate the probability that an individual sample is on the edge of a maxed-out condition. This step models the probability (P) that a sample at n is maxed-out based on its distance (x) to the sensor’s dynamic range ceiling as a gamma cumulative distribution function (F). The mathematical model is depicted in Equation 1, where E is the event that sample n is maxed-out, $|d|$ is the absolute acceleration value, d_0 is $r - 3 \times \delta$ representing the boundary of the maxed-out ceiling (r = dynamic range ceiling; δ = standard deviation of measured static noise), x is the distance between the sample’s acceleration value and the boundary at d_0 , k is the shape parameter, and θ is the scale parameter of gamma cumulative distribution. We set $\theta = 1$ and optimize k so that $P(E|x = 3 \times \delta) \approx 0.95$ (Zar, 1984). A visual representation of the result of this step can be seen in Figure 1, Step 2A—oval inset in signals 2, 3, and 4, where a darker point indicates a higher probability that the sample is on the edge of a maxed-out condition.

$$P_n(E|x = |d| - d_0) = F(x = |d| - d_0; k, \theta) = \frac{\gamma(k, \theta x)}{\Gamma(k)} \text{ if } x < 0, P_n(E|x) = 0 \quad (1)$$

In Step 2, we determine the maxed-out edge. The maxed-out regions in the signal are determined from the difference between confidence probabilities of adjacent samples as $Q(n - 1, n) = P_n - P_{n-1}$, where Q is the difference of maxed-out probability between sample $n-1$ and n , and P is the maxed-out probability. If $|Q|$ is greater than a maxed-out threshold C (set at 0.5), then sample n is treated as the boundary of a maxed-out region. The signs of $Q(n - 1, n)$ and P_n determine whether the boundary sample is at a “valley” (–) or “hill” (+) in the acceleration signal and if it is on the left or right side of the region. The algorithm then picks samples that are within T seconds outside from the left and right boundary samples of a maxed-out region for signal reconstruction. Preceding extrapolation, interpolating the signal up to 100 Hz ensures that when $T = 0.05$, there will be at least five samples on each side of the maxed-out boundary samples for regression fitting. This improves robustness during fitting.

In Step 3, we extend the maxed-out signal. The samples selected that are within T seconds from the left and right boundaries of a maxed-out region are modeled using weighted spline regression on each side, where weights are set as the probability of not maxing-out: $W = 1 - P$. Neighborhood duration (T) and smoothing parameter (S) (range: 0 to 1) are tunable parameters that increase the robustness of the regression model (Equation 2). A smaller S ensures fitting that is smooth and close to a straight line. This can be observed in the oval insets in Figure 1, Step 2A, Signals 2, 3, and 4 where the solid lines are regression curves, extending from the left and right boundary samples of the maxed-out signal. The mean position of the point of intersection of the two regression curves $f_{left}(t)$ and $f_{right}(t)$ in the middle of the maxed-out region: $t_{middle} = \frac{t_{left} + t_{right}}{2}$ is computed to determine the

extrapolated peak (Equation 2) (markers in Figure 1, Step 2A). The original maxed out samples are then replaced by spline-interpolated points derived from the extrapolation peak (t_{middle} , y_{middle}) and non-maxed-out samples to yield a consistent sampling interval sequence (Equation 3), where m and $m+N$ are indices of the edges of the maxed-out region (Figure 1, Step 2B). We empirically determined the values for parameters S and T during the process of optimizing MIMS-units (see section titled “Simulated Sinusoidal Signal Testing” in results).

$$y_{middle} = \frac{f_{left}(t_{middle}) + f_{right}(t_{middle})}{2} \quad (2)$$

$$y_{extrapolated} = \text{SplineInterpolation}(y_1, \dots, y_m, y_{middle}, y_{m+N}, \dots, y_n) \quad (3)$$

Bandpass Filter.—This step filters out artifacts from the acceleration signal that do not pertain to voluntary human movement, minimizes unwanted tremors and vibrations (Figure 1, Step 3), and eliminates the gravity vector. The bandpass filter in the MIMS-unit algorithm is a fourth-order Butterworth Infinite Impulse Response (IIR) filter (0.2–5 Hz) (Oppenheim & Schaffer, 1975).

The Butterworth filter is commonly used in signal processing because it has a flat gain in the passband and lends optimal control of the cut-off frequencies (Oppenheim & Schaffer, 1975). Compared to other bandpass filters such as the Finite Impulse Response filter, the Butterworth IIR filter has a superior frequency response (i.e., reduced signal attenuation in the transition band at a low filter order) (Oppenheim & Schaffer, 1975). However, compared to other IIR filters, the Butterworth also has a longer transition band, which may include more noise from vibration/tremors that may inflate the time-domain signal (Oppenheim & Schaffer, 1975). A Butterworth filter with a 5 Hz upper bound will have a transition band that may include noise between 5 and 18 Hz, which is wider than that in other IIR filters such as the Elliptical filter (transition band = 5–10 Hz). Evaluation of the Butterworth filter is described below.

Aggregation: Single Axis and Triaxial.—Aggregating the filtered acceleration signal into MIMS-units for a single axis is additive, where aggregated values from shorter epochs can be summed to yield aggregations from longer epochs. The additive property allows an estimation of both total movement and movement intensity. Thus, we choose area under curve instead of the Euclidean norm for aggregation. Signals from each channel are aggregated using numerical integration based on the trapezoid rule (“trapz” in the R package “caTools”) to yield values in epochs (Figure 1, Step 4). The epoch can be arbitrary in length. In the experiments described below, we used 1 min epochs.

Rather than computing vector magnitude to aggregate individual axis data from a triaxial sensor, the MIMS-unit algorithm sums integrated values from each of the three axes for an epoch. While a vector has both magnitude and direction, bandpass filtering and signal integration eliminates direction and returns a scalar quantity. Thus, summation is appropriate

to represent physical movement when aggregating triaxial signals after filtering and integration.

MIMS-Unit Truncation.—Truncated zero counts resulting from non-linear zeroing of values below an arbitrary proprietary threshold may eliminate useful information in the signal. Non-movement-related variability in the signal attributable to inherent imperfections in filter transition bands, and limits in floating number precision, were eliminated where output was less than or equal to 10^{-3} via truncation of MIMS-unit values to zero. This threshold was empirically determined by generating MIMS-units from simulated signals with zero variance and real signals from stationary sensors placed on a horizontal surface.

Algorithm Optimization and Testing

Experimental testing included simulated sinusoidal data, controlled shaker testing, and human data from ambulatory and activities of daily life. Simulated sinusoidal signal testing was aimed at optimizing the step of signal extrapolation. Shaker and human testing aimed to compare the performance of MIMS-units to two popular output types—the ActiGraph activity count and “Euclidean Norm Minus One” (ENMO) summary used in the UK Biobank cohort study (Doherty et al., 2017). We picked the two latter outputs for comparison as they are widely used in physical activity research.

Simulated Sinusoidal Signal Testing.—Simulating sinusoidal signals allows control over signal frequency and amplitude during testing and were hence used to optimize the process of extrapolating a maxed out signal. Simulated signals were generated using MATLAB (R2016B, MathWorks, Natick, MA). A maxed-out signal was defined as $y = \min(\hat{y}, D)$, where the ground truth signal \hat{y} is cut-off at a peak dynamic range D . The algorithm was tested on simulated signals with dominant frequencies between 1 to 8 Hz, with cut-off D ranging between ± 2 to ± 8 g (in 1 Hz and 1 g increments). These dynamic range cut-offs simulate extrapolation performance for sensors with varying dynamic range. The ± 2 g range represents the lower end of the dynamic range used in current accelerometers including those in several consumer wearables and phones and is a challenging test case for the extrapolation step of the MIMS-unit algorithm. Additionally, we included frequencies beyond the upper bound of the bandpass filter (5 Hz) to determine performance in edge cases where there was a greater possibility of the presence of artifacts (e.g., vibration, tremors).

Orbital Shaker Testing.—Orbital shaker testing subjected devices with variable dynamic range and sampling rates to incremental testing frequencies and acceleration amplitudes. This protocol allowed a comparison of intra-output consistency among MIMS-units, ActiGraph counts, and ENMO.

For research grade devices, we selected different generations of ActiGraph devices for testing because they facilitated variability in both dynamic range and sampling rate, which increased scientific rigor in testing the consistency of MIMS-units. Additionally, we tested the ActivPal3 as it would be a difficult test case for MIMS-units due to the device’s default setting of 2 g with a low sampling rate of 20 Hz. We did not use the ± 8 g Axivity AX3 or

GENEActiv devices because these would be similar test cases to those in our experiments and would likely yield limited new information. Both of these devices use the same MEMS sensor (Analog Devices ADXL345 at a higher dynamic g-range) as the ActivPal3 and also belongs to the same family (ADXL) of sensors used in ActiGraph devices (except the ± 6 g GT3X+). Two commercial devices were selected to introduce additional variability in sensor specifications. Thus, eight accelerometers (brand name [g range, sampling rate]): ActivPal3 [± 2 g, 20 Hz], GT3X [± 3 g, 30 Hz], GT3X+ [± 6 g, 40 Hz], GT3X+ [± 6 g, 80 Hz], GT9X [± 8 g, 60 Hz], GT9X [± 16 g, 100 Hz], LG watch Urbane R [± 2 g, 100 Hz], and moto g⁶ play [± 2 g, 50 Hz]) underwent controlled orbital shaker testing (Model 1231A89, Thomson Scientific, London, UK) at frequencies from 1 to 5 Hz (in 1 Hz increments) and a fixed radius of 8.5 cm, which generated maximum acceleration between 0.2 to 8.6 g. Devices were fastened to the shaker such that there was minimal movement artifacts and the z-axis was sensitive to gravity. Three minutes of steady state data at each frequency were analyzed in this experiment.

Human Testing.—Human testing aimed to compare inter-output behavior patterns among MIMS-units, ActiGraph activity counts, and ENMO during human movement that exposed devices to a broad range of signal frequency and amplitude characteristics. Two sets of data from a total of 60 participants (Age = 23.6 ± 2.5 y; BMI = 24.4 ± 5.1 kg/m²) were used to evaluate MIMS-units. The first dataset was from 10 participants who wore a GT9X on the dominant hip and wrist and performed an ambulation protocol at increasing speeds between 2.9 and 20.1 km/h (i.e., 1.8 to 12.5 mph) (John, Miller, Kozey-Keadle, Caldwell, & Freedson, 2012; John et al., 2010). This protocol aimed to evaluate the signal processing methodologies on both low and vigorous rhythmic human movements. These data were used to compare patterns of ActiGraph activity counts (ActiLife v6.13.4) and the “Euclidean Norm Minus One (ENMO)” summary used in the UK Biobank cohort study, to patterns of MIMS-units. While this protocol has identified inadequacies of the ActiGraph bandpass filter in detecting high frequency movement (John et al., 2012; John et al., 2010), the impact of this protocol on MIMS-units and ENMO is unknown. Moreover, a recent study reported converse findings for hip and wrist output during this protocol where wrist ActiGraph counts continued to increase with speed (LaMunion et al., 2017). Variability in output patterns during human movement with similar dominant frequencies may have implications on sensor wear location, signal filtering, and in selecting device dynamic range both when summarizing signals or using complex techniques to process raw data.

The second data set was from 50 participants who wore a GT9X on the non-dominant wrist and dominant hip while performing an uncontrolled multi-activity protocol involving indoor and outdoor ambulatory and activities of daily living (Supplementary Tables 1 and 3 [available online]). These included six different activities performed when sitting, five when walking, and three when standing. Participants were only told what activity was to be performed with no instructions on how to perform the same. This protocol aimed to subject sensor signal processing to both rhythmic and unrhythmic movements of various frequencies and amplitudes. A researcher recorded start/stop times for all activities (2–4 min each). All participants provided written informed consent approved by the Northeastern University Institutional Review Board.

Data Analyses

Analyses were aimed at examining specific aspects and overall performance of the MIMS-unit algorithm including overall algorithm consistency and comparing patterns of MIMS-units to popular acceleration summary metrics.

Bandpass Filter.—To determine the impact of filter type, we compared the performance of the Butterworth filter to the Elliptical filter during outdoor bicycling. During this activity, there is limited spatial movement of the upper body, but a significant proportion of acceleration detected by a wrist sensor may pertain to continuous vibration artifacts that are transferred from the wheels rolling on the pavement to the accelerometer. We report the proportional difference in the gain factor of the original acceleration signal at -3 dB for the 5 Hz cut-off. In signal processing, it is typical to examine the rate of signal power attenuation on a logarithmic scale at a gain factor of 3 dB, at which power changes by a factor of 0.5 (i.e., half-power) (Oppenheim & Schaffer, 1975).

Simulated Sinusoidal Signal Testing.—Optimizing extrapolation involved defining best values for the two tunable parameters: neighborhood duration T and smoothing parameter S (see section describing extension of the maxed-out signal). A grid search on the simulated sinusoidal signals was used to determine optimal values for T and S . Extrapolation performance on simulated signals of varying sampling frequencies and dynamic range cut-offs (detailed in the testing protocol above) were computed per Equations 4.1 and 4.2., where E_0^i is the difference between the rectified area under the curve of the original maxed-out signal i (subject to a dynamic range cut-off) and the ground truth (not subject to a dynamic range cut-off); i.e., the “true error.” True error is 0 when the original signal is identical to the ground truth, and >0 when the signal is maxed-out. E^i is the difference between the extrapolated signal i and ground truth; i.e., the extrapolation error. If $E_0^i > E^i > 0$, the extrapolated signal underestimates the ground truth but is closer to the ground truth than the original signal. If $E^i = 0$, the extrapolated signal matches the ground truth. If $-E_0^i < E^i < 0$, the extrapolated signal overestimates the ground truth but is closer to the ground truth than the original signal. Correspondingly, the optimal value for extrapolation rate E_j is 1, where performance degrades from 1 (best performance) to 0 (extrapolated signal is the same as the original maxed-out signal). $E_j < 0$ means extrapolation performs worse than not applying it. Extrapolation performance was assessed by computing the average extrapolation rate E of the true signal at different sampling rates in 10 Hz increments.

$$E_0^i = \frac{AUC_{\text{truth}}^i - AUC_{\text{original}}^i}{AUC_{\text{truth}}^i} \quad (4.1)$$

$$E^i = \frac{AUC_{\text{truth}}^i - AUC_{\text{extrapolated}}^i}{AUC_{\text{truth}}^i} \quad (4.2)$$

Orbital Shaker Testing.—Acceleration signals from the eight devices during shaker testing at orbital frequencies from 1 to 5 Hz (in 1 Hz increments) were processed to compute five types of motion summaries: (i) output from the MIMS-unit algorithm; (ii) output from the MIMS-unit algorithm without extrapolation, to examine the need for extrapolation; (iii) output from the MIMS-unit algorithm with a lower passband of 0.25 to 2.5 Hz, to examine the effect of a narrower passband; (iv) UK Biobank ENMO output (Doherty et al., 2017); and (v) ActiGraph activity counts. The ENMO-based summary, which was averaged over 5-s epochs in the UK Biobank study, was derived using open-source software released by the authors of the paper on the analyses of wrist-worn accelerometry in the study (Doherty et al., 2017). Prior to computing UK Biobank ENMO output in the current study, free-living data were collected using all eight sensors over a period of 48 h, which enabled the generation of calibration correction coefficients for each sensor (Doherty et al., 2017; van Hees et al., 2014). Data from the simulated protocol were sufficient to generate calibration correction factors for UK Biobank ENMO output. With respect to ActiGraph counts, previous work using a similar approach has examined activity counts generated by ActiLife software using raw acceleration from a non-ActiGraph device (Brond, Andersen, & Arvidsson, 2017). These analyses allowed a direct comparison of MIMS-unit to the popular proprietary ActiGraph activity count algorithm, and to the UK Biobank ENMO-based summary. Inter-device consistency for each of the abovementioned five different acceleration summary output types at each testing frequency (1–5 Hz in 1 Hz increments) during the shaker testing protocol was assessed by computing the coefficient of variation (CV) among outputs. Using CV enables a fair comparison of inter-device consistency for sensors generating the same type of output with different inputs because CV is independent of measurement units and is appropriate for absolute values such as summarized acceleration.

Human Testing.—For the ambulatory and multi-activity protocols, we examine visualizable patterns of MIMS-units, ActiGraph activity counts, and UK Biobank ENMO summary from (i) an ± 8 g signal, and (ii) the same signal that was “cut-off” at ± 2 g with added Gaussian noise (i.e., random fluctuations). The comparison of outputs from a cut-off ± 2 g signal was only made between MIMS-units and UK Biobank ENMO because the ActiGraph activity count algorithm computes counts after clipping signals at ± 2 g. Signal clipping is performed by ActiGraph’s ActiLife software to maintain consistency of output from current devices having a higher dynamic range (e.g., GT9X: ± 8 to 16 g) with that from ± 2 g legacy devices. Comparing outputs from a cutoff ± 2 g signal to the original ± 8 g is relevant because manufacturers of wearables and mobile phones often use sensing with lower g-ranges (e.g., ± 2 g) to obtain higher sensitivity in motion for applications that pertain to user interaction (e.g., device orientation detection to rotate display screen) and for detection of subtle movements, such as those during sleep. For all activities in the multi-activity protocol, paired-sample t-tests ($p < .05$) between MIMS-units from the ± 8 and ± 2 g signals were conducted to detect inter-output differences at the same site for each activity. Similar, analyses were conducted for ENMO outputs from ± 2 and ± 8 g signals at each location.

Results

Findings from the analyses of specific aspects and overall performance of the MIMS-unit algorithm are described below.

Bandpass Filter Testing

The Fourier transform of acceleration during bicycling revealed substantial vibration artifacts beyond 5 Hz. Artifacts preserved by the Butterworth (1.7% of original gain) and Elliptical filters (1.0% of original gain) were marginal (Supplementary Figure 1 [available online]). Given that the Butterworth filter preserved only a small proportion of artifacts and has other known strengths (described above), use of the Butterworth filter in MIMS-units is preferable.

Simulated Sinusoidal Signal Testing

The grid search on the simulated signals yielded values of 0.05 and 0.6 for T and S , respectively. These were used to optimize extrapolation of a ± 8 g signal that was cut-off at ± 2 g. Figure 2a shows the average extrapolation rate (E) from 20 to 100 Hz. The average was computed from signals cut off at ± 2 g with different signal frequencies ranging from 1 to 8 Hz (in 1 Hz increments). Increasing sampling rate improved E and decreased its variability. E improved greatly from 20 to 40 Hz, followed by gradual improvements at higher sampling rates. Figure 2b details extrapolation performance for the simulated signals across cutoff amplitudes ranging from ± 2 g to ± 8 g with frequencies from 1 to 8 Hz at sampling rates of 20, 50, and 100 Hz. At 50 Hz, for most scenarios that may need extrapolation (movement between ± 2 and ± 4 g and < 3 Hz), E is roughly 65 to 75%, which may be sufficient to achieve acceptable inter-device consistency. E is over 75% at a sampling rate of 100 Hz. MIMS-units are computed from an interpolated signal of 100 Hz, which maximizes extrapolation rate of the signal and thereby, reduces inter-device discrepancies in the output. Figures 2c–2f depict signals extrapolated from maxed-out signals of ± 2 g and the corresponding ground truth from a ± 8 g signal during shaker testing (5 Hz), running at 8.95 (ankle) and 12.1 km/h (wrist), and jumping jacks (wrist), respectively. Extrapolation enabled a more complete representation of movement of the monitor during the activity.

Orbital Shaker Testing

Figure 3a shows CV among acceleration summary outputs from five different algorithms, each considering data from eight different accelerometers (with varying specifications) acquired at different shaker testing frequencies. CV worsened as testing frequency increased, which is understandable because differences between sensor capabilities become increasingly apparent when measuring motion with higher intensity. A lower CV demonstrates improved algorithmic capability in overcoming inter-device hardware and signal processing discrepancies to enable greater comparability of the same output from different devices (with variable dynamic range and sampling rates). At testing frequencies up to 4 Hz, the increase in CV was lowest for MIMS-units (~average: 6%; range: 2–14%). CV increased to 25% at 5 Hz. In comparison, up to 4 Hz, the ENMO-based summary displayed a consistently high CV of ~31% on average (~range: 18 to 50%) and the

ActiGraph activity count had an average CV of ~26% (~range: 18–36%). At 5 Hz, CV increased to ~60% for the ENMO-based summary. Figure 3b shows MIMS-units from the eight different devices that underwent shaker testing. Quantitatively, the outputs at each frequency have a desirable close-knit pattern up to 3 Hz. At 4 Hz, outputs from the ActivPal3 (± 2 g, 20 Hz) and moto g⁶ play (50 Hz, ± 2 g) deviate from that of other devices with a decrease in MIMS-units. At 5 Hz, deviations were additionally observed in the LG Urbane smartwatch (100 Hz, ± 2 g) and the GT3X (30 Hz, ± 3 g). Deviations from these sensors caused observed increases in CV at 4 Hz (14%) and 5 Hz (25%). While we did not conduct shaker tests at frequencies >5 Hz (due to limitations in shaker setup), MIMS-units will plateau at 5 Hz (bandpass cut-off), and then decrease with increasing frequency. Supplementary Figure 2 (available online) shows ENMO and ActiGraph count outputs from the eight different devices that underwent shaker testing.

Human Testing

Figure 4 shows MIMS-units, ActiGraph activity counts, and UK Biobank ENMO output from a GT9X at the hip and wrist during the ambulatory protocol, respectively. We do not report data for running at the highest speed because four out of the 10 participants were unable to run at this speed. Unlike hip ActiGraph activity counts, MIMS-units and UK Biobank ENMO demonstrate an upward slope with increasing speed. However, for the ± 2 g signal from the hip, the UK Biobank ENMO output started to incrementally deviate from the ± 8 g output at 7 km/h (4.3 mph; difference with ± 8 g signal = ~5%) and continued to widen with increasing speed (~50% difference at 18 km/h). Comparatively, MIMS-unit at the hip demonstrated widening later, at 10 km/h (6.3 mph), where the ± 2 g signal was ~3% lower than the ± 8 g signal, which increased to approximately 9% at 18 km/h (11.2 mph). Similar widening for the ENMO-based output was observed for the wrist signal (± 2 g signal lower than the ± 8 g signal by ~28% and 62% at 8 km/h and 18 km/h, respectively). However, MIMS units derived from the ± 2 g signal at the wrist tracked consistently with the ± 8 g signal with percent differences within 2% and 4% between 10 km/h and 18 km/h, respectively. Figure 5 contains boxplots for ActiGraph counts (A), MIMS-units (B), and UK Biobank ENMO (C) output computed from data acquired during the multi-activity protocol. Activities are arranged in the order of increasing MET values derived from the Compendium of Physical Activities (Ainsworth et al., 2011). Selected activities from Figure 5 are depicted in Figure 6, which shows activities with instances where ActiGraph activity counts were equal to zero in few participants (depicted by the lower whisker). This inability of ActiGraph activity counts to detect subtle movements is overcome by MIMS-units and UK Biobank ENMO output. There were no significant differences ($p < .05$) between intra-location (e.g., wrist ± 2 g vs. ± 8 g output) MIMS-units during any activity (Figure 5). However, there was a significant intra-location difference ($p < .05$) between ± 2 and ± 8 g ENMO outputs at both the wrist and hip during the Frisbee and running at 8.8 km/h activities (see Figure 5). Additionally, ENMO outputs from ± 2 and ± 8 g signals at the hip were significantly different from each other during walking downstairs ($p < .05$).

Computation Time and Memory Use

It takes about 18 min on a Dell Precision 7810 workstation (64 GB memory, 24-core 3.0 GHz CPU; disabled multi-core processing; Dell, Round Rock, TX) to process a standard-

size (12 MB per hour), week-long dataset from a single triaxial accelerometer (80 Hz, ± 8 g) in 1 min epochs. Aggregating MIMS-units into specific epochs (1 s vs. 60 s) has a relatively small impact on computational duration. We do not advocate for a specific epoch length or methodology when using MIMS-units.

Discussion

We first discuss the key algorithmic step of signal extrapolation optimization performed using simulated sinusoidal signal testing. This is followed by an interpretation of findings from the shaker and human testing protocols.

Simulated Sinusoidal Signal Testing

When a ± 2 g device is subjected to acceleration between ± 2 to ± 3 g (typically observed during human movement), interpolating the resulting signal to 100 Hz improved the robustness of the local spline regression during extrapolation (Figures 2a and 2b). However, when devices do not have battery/storage constraints and sampling-rate-dependent software inconsistencies (e.g., ActiGraph counts; Brond & Arvidsson, 2016), collecting data at the highest sampling rate possible is preferable. At low sampling rates (i.e., < 30 Hz, Figure 2b), extrapolation may be unsatisfactory with increasing frequency and amplitude because parameters T and S (described in R1) were optimized to maximize overall performance across a variety of sampling rates. At low-sampling rates there may be insufficient information to extrapolate accurately.

Orbital Shaker Testing

Output similarity across devices is crucial to maintain uniformity in various analytical steps to enable cross-study comparisons. In response to the same acceleration, inter-device MIMS-units displayed a low CV within the filter passband (2% at 1 Hz to 14% at 4 Hz; Figure 3a). Additionally, CV is $< 7\%$ up to a frequency of 3 Hz. Typical spatial movement during common activities of daily living typically fall below 3 Hz (Supplementary Table 2 [available online]). CV increases with oscillation frequency due to a concurrent increase in the number of devices that require extrapolation. Predictably, eliminating extrapolation and a narrow filter passband increased CV (Figure 3a).

The proprietary ActiGraph count displayed increasingly high CVs (average of $\sim 26\%$) at testing frequencies from 1 to 4 Hz. Unavailability of information on the exact data processing steps in this proprietary algorithm prevents a clear understanding of the factors that yield a high CV for this output. A factor that may impact inter-device output CV at higher frequencies may be a flaw in the algorithm that allows signal frequencies above 5 Hz to escape the bandpass filter when the sampling rate is not a multiple of 30 Hz (Brond & Arvidsson, 2016).

One factor that potentially elevates CV for the UK Biobank ENMO-based summary (average of $\sim 32\%$ for 1–4 Hz) may be the use of first-degree polynomial linear interpolation to resample the acceleration signal to 100 Hz. The assumption that motion between consecutive acceleration data points is linear, may minimally impact the signal curve at higher sampling rates. However, decreasing sampling rate leads to increasingly spread-out

data points, which may cause linear interpolation to produce an increasingly irregular signal. It is important to note that the ENMO-based summary used in the UK Biobank study was computed on a signal that was collected at a sampling rate of 100 Hz, which is further interpolated to the same frequency to account for fluctuations in sampling rate (Biobank UK, 2016). While interpolation improves signal density and may minimize variability in output, versions of ENMO prior to that used in the UK Biobank study, did not perform interpolation to improve sampling density (Doherty et al., 2017; van Hees et al., 2013), which may induce further variability when comparing ENMO outputs from low-sampling rate end-consumer devices to those from higher sampling rate research grade devices. Another factor that further elevates CV among the UK Biobank ENMO-based summaries for frequencies above 3 Hz is variability in the dynamic range of the 8 accelerometers that were tested (± 2 to ± 16 g). Shaker testing at 3 Hz produces a peak acceleration of $\sim \pm 3.1$ g in the x - and y -axis, with minimal acceleration (vibration) in the z -axis that detects gravity. Thus, at 3 Hz, devices with a dynamic range of less than ± 3 g will demonstrate minimal increase in ENMO (i.e., devices will max-out at their set dynamic range). These devices contribute to increasing CV in the UK Biobank ENMO-based summary with increasing testing frequency. Inevitable variability in both sampling rate and dynamic range among the variety of devices that are available may impact inter-device comparability of the UK Biobank ENMO-based summary.

A considerably lower CV across shaker testing frequencies for the MIMS-unit algorithm may be attributable to the use of (i) a third-degree polynomial cubic spline interpolator that produces a comparably regular acceleration curve during interpolation and (ii) signal extrapolation that minimizes the issue of a low dynamic range in a device.

The low CV for MIMS-units from different devices corresponds with our observation of similar visual patterns of output (Figure 3b) during shaker testing from the eight different accelerometers. A tight pattern exists up to 3 Hz, with widening at higher frequencies. A notable deviation is the ActivPal3 (20 Hz, ± 2 g) and the moto g⁶ play (50 Hz, ± 2 g), which plateau between 3 and 4 Hz (acceleration generated during testing $> \pm 3.1$ g), and then continue to decline. This may be due to a combination of a low sampling rate of the original signal (20 Hz for the ActivPal) and a low dynamic range of ± 2 g that may have impacted the process of extrapolation. Compared to these, MIMS-units from the GT3X (30 Hz ± 3 g) and LG Urbane (100 Hz, ± 2 g) plateau later between 4 and 5 Hz (acceleration generated during testing $> \pm 5.5$ g). Most current research and consumer devices have sampling rates ≥ 30 Hz. Increasing CV and a widening of inter-device output at 4 and 5 Hz may also be related to variance in local weighted spline regression during signal extrapolation, which may manifest as noise that is not eliminated entirely during signal filtering.

Human Testing

Here we separate the discussion based on findings from the ambulatory and multi-activity protocols. Discussion is limited to the performance of the three acceleration summaries derived from a 2 g and 8 g signal, in detecting motion characterized by a wide range of signal frequencies and amplitude, and in minimizing the inclusion of extraneous artifacts that may not pertain to human movement.

Ambulatory Protocol (2.9–20.1 km/h).—We first discuss findings related to bandpass signal filtering. This is limited to ActiGraph activity counts and MIMS-units because other than a low-pass filter of 20 Hz, the UK Biobank ENMO-based output does not apply filtering that is specific to eliminating non-human movement. We then discuss findings related to dynamic range, which is limited to ENMO and MIMS-units because ActiGraph counts are generated after raw acceleration is clipped at ± 2 g.

Similar to previous work, our study demonstrated flaws in the ActiGraph's narrow passband of 0.25 to 2.5 Hz on hip-activity counts (John et al., 2010) (Figure 4). Conversely, the 5 Hz passband cutoff for hip-based MIMS-units yielded an increase in output and may be appropriate to detect normal human movement using a hip-worn device. We base this on current findings from the ambulatory protocol, and on evidence from existing literature on various activities, which suggest that the rate of spatial movement of the hip during normal unrestricted human ambulation is not likely to exceed 3.5 Hz in the general population (and is within 5 Hz even in world class sprinters) (John et al., 2012; Krzysztof & Mero, 2013). Hip-based MIMS-units will plateau at 5 Hz and then decrease with increasing frequency.

Both MIMS-units and ActiGraph counts at the wrist increased with running speed (Figure 4). Compared to the hip, we observed higher signal power (i.e., signal density) at the wrist, which may be one reason for increasing output at the wrist for ActiGraph activity counts (Supplementary Figure 3 [available online]). A Fourier transform of raw acceleration for each axis of data from the hip and the wrist revealed that both signal amplitude and power for the 1st and 2nd dominant frequencies for wrist signals demonstrated a greater rate of increase and magnitude than those at the hip. This finding reinforces the potential of the wrist location in capturing most whole-body movement and also rapid movements of the upper body when processing both raw wrist data or summarizing the same.

Both MIMS-units and the UK Biobank ENMO-based acceleration summaries derived from the ± 8 g signal increased with speed at the hip and the wrist. While marginal differences ($<4\%$) were observed between MIMS-units derived from a ± 8 and a ± 2 g signal at the wrist, incrementally large percent differences between the ± 2 and ± 8 g signals for the UK Biobank ENMO-based output demonstrates limited transferability of the latter when using devices with a dynamic range that is lower than ± 8 g. Thus, while this output may be adequate for typical research grade sensors that offer a high dynamic range of $\sim \pm 8$ g, it may be inappropriate for a wide variety of consumer based devices such as smartwatches and phones that typically have a lower dynamic range and are increasingly being used in research (Althoff et al., 2017; Wang et al., 2015).

Multi-Activity Protocol.—To examine the effect of the lower bound 0.2 Hz passband cutoff of MIMS-units, we conducted Fourier transform analyses on wrist and hip acceleration detected during sedentary activities such as lying and sitting still (data not presented). These analyses verified the existence of subtle movement acceleration during motionless sitting and lying with dominant frequencies between 0.2 and 0.25 Hz. Such movements are particularly evident at the wrist. Both MIMS-units and the ENMO-based summary were sensitive to capturing acceleration due to subtle wrist movements during stationary sitting and lying, which was not evident in ActiGraph counts (Figure 6). Thus, the

MIMS-units and the UK Biobank ENMO-based summary may be useful in developing new methods that potentially yield improved distinction of non-wear from sedentary behavior and/or sleep. However, the UK Biobank ENMO-based summary may also be susceptible to capturing various environmental artifacts detected by the sensor, particularly when it is worn at an extremity such as the wrist. For example, this output during outdoor cycling was higher for the wrist than the hip, which may not be reflective of the actual work performed during an activity. Both ActiGraph and MIMS-units demonstrated the opposite trend. Higher ENMO output may be attributable to the retention of non-human movement vibration artifacts arising from the front-tire of the bicycle rolling on the pavement that are transferred through the rigid handlebars and captured by the wrist sensor. Vibrations may be lower at the hip due to a greater damping of these artifacts when transferred through the lower limbs from the pedals and from the saddle through the pelvic region. Supplementary Figure 4 (available online) shows filtered and unfiltered signals from the hip and wrist during outdoor cycling. Vibration artifacts detected by the wrist sensor are substantially higher than that at the hip. Supplementary Figure 1 (available online) shows that a significant proportion of vibration artifacts lie between 5 and 20 Hz. These non-movement artifact signals are retained by the 20 Hz low-pass filter in the UK Biobank ENMO-based summary and likely result in wrist output that is higher than that from the hip (Figure 6). Thus, any approach that summarizes raw acceleration may need to consider potential applications of the metric when determining the parameters that include or eliminate information from the raw signal during the data reduction process.

Given that the UK Biobank ENMO retains signals under 20 Hz, including artifacts, we quantified the impact of such signals on this metric. ActiGraph GT3X+ data from 15 free-living older women (60+ y) were used to compute ENMO using a 5 Hz and a 20 Hz low-pass cut-off. We previously justified 5 Hz as the upper limit cut-off that retains most human movement. ENMO output using the 20 Hz cut-off was ~9% higher than that using the 5 Hz cut-off. We conducted similar analyses on grouped activities (ambulation, sedentary, household, standing, bicycling) from the multi-activity protocol (Supplementary Figure 5 [available online]) to further understand possible physical behaviors that contribute to the observed output differences in free-living data. Compared to UK Biobank ENMO generated using a 5 Hz low-pass cutoff, the 20 Hz low-pass cutoff for sedentary behaviors yielded ~10% higher output. This is similar to the ~9% difference noted in the FL sample. Sedentary behaviors dominate FL waking behavior (Matthews et al., 2008), which may explain the similarity in output differences observed in free-living and multi-activity protocol.

Study Limitations, Practical Applicability, and Conclusions

A potential limitation in the experiments used to test the performance of MIMS-units is the absence of free-living data collected simultaneously from devices with different configurations. Some work has shown free-living acceleration signals to yield superior results when calibrating machine learning algorithms to estimate physical activity outcomes such as type and intensity (Ellis et al., 2014; Sasaki et al., 2016). However, unlike the impact of free-living activity context in predicting specific physical behavior labels when training/using a machine learning algorithm, we do not expect free-living activity context to influence the aggregation of the accelerometer signal into the three examined acceleration

summaries (MIMS, ENMO, and counts) per se and thus, limit our findings on inter-metric comparisons. Additionally, the likelihood that free-living behavior in the general population will generate (i) sufficient testing sample signals of (ii) a variety of frequency and amplitude characteristics (including edge conditions), is low. Our testing procedure ensured the availability of signals with a wide variety of such characteristics.

The study has some other limitations. One is that MIMS-units cannot be derived from devices that do not provide raw data, although, this is becoming increasingly available among consumer devices. Processing of raw data using advanced methodologies that harness important information within the acceleration time and frequency components may represent physical behavior at a much higher resolution. Simpler summarizations of acceleration into epochs using MIMS-units or other accelerations summaries will result in the loss of such information. Another limitation is that we did not include devices such as the AX3 and GeneACTIV. As stated previously, our selection criteria for devices was aimed at inducing maximal variability in device specification. Our study did not quantify and compare associations between health and various acceleration summaries and/or other metrics such as steps. However, this was beyond the scope of the study.

Surveillance studies and longitudinal real-time interventions increasingly rely on understanding behavior patterns using large volumes of accelerometer data from wearable sensors or mobile phones. MIMS-units may be used to simplify the process of visualizing and verifying the existence of anomalous data prior to subsequent scrutiny of the signal and the application of simple to complex data processing techniques to determine relationships between physical behavior and health. Given that the MIMS-unit algorithm accounts for inter-monitor discrepancies in sampling rate and dynamic range, using MIMS-units in surveillance and longitudinal studies may eliminate the reliance on one specific brand of monitor and enable cross-study uniformity in various stages of data processing. Low device-dependency will facilitate access to a significantly larger recruitment pool and valid cross-comparison of output across a wider array of studies that use personal health technologies, which provide access to raw data or may do so in future. From a generalizability standpoint, we do not expect any performance impairment of MIMS-units in study samples with a wide age range. Parametric specifications for MIMS-units were developed, justified via previous literature, optimized, and tested on capturing motion that occur in a wide spectrum of the frequency of human movement and sedentary behavior, while minimizing artifacts that are inevitably captured by a sensor in daily life. Importantly, our findings demonstrate that MIMS-units are generalizable across devices with dissimilar manufacturer specifications.

MIMS-units may be used to develop cutoffs for activity intensity categories or to estimate wear-time. These may be used to determine and compare associations between physical behavior and health outcomes across surveillance and longitudinal studies where the application of more complex methodologies using raw data may presently have low feasibility. Other summary metrics such as ActiGraph activity counts and ENMO continue to be used to develop cut-point methodologies aimed at determining attributes of physical behavior (Sasaki, John, & Freedson, 2011; van Loo et al., 2018). While we do not encourage cut-point-based methodologies due to inherent flaws in the approach, it is up to sensor calibration researchers to overcome current barriers in not only developing accurate

methodologies, but also those that hinder method-deployment. These may be necessary to allow end-user researchers that currently use cut-points to transition into using improved methods. Additionally, compared to bouted MVPA derived using thresholding methodologies, total volume of movement derived using ActiGraph counts have stronger associations with several cardiometabolic biomarkers (Wolff-Hughes et al., 2015). MIMS-units may be used similarly to explore associations between health outcomes and total movement. It is not possible to examine such associations using the UK Biobank ENMO metric since it is not additive and is averaged over an epoch (Hildebrand, van Hees, Hansen, & Ekelund, 2014).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by funding from the National Cancer Institute of the NIH under award number 261201300055C-0-0-1. Enabling technology used in this work was made possible by funding from the National Heart Lung and Blood Institute of the NIH under award number UO1HL09173. We thank Diego Arguello and Alvin Morton for assisting with data collection in the study.

References

- Accelerometer Sensor Guide (2019). Retrieved from <https://dev.fitbit.com/build/guides/sensors/accelerometer/>
- Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett DR Jr., Tudor-Locke C, ... Leon AS (2011). 2011 Compendium of physical activities: A second update of codes and MET values. *Medicine & Science in Sports & Exercise*, 43(8), 1575–1581. doi:10.1249/MSS.0b013e31821ece12 [PubMed: 21681120]
- Althoff T, Sosic R, Hicks JL, King AC, Delp SL, & Leskovec J (2017). Large-scale physical activity data reveal worldwide activity inequality. *Nature*, 547(7663), 336–339. doi:10.1038/nature23018 [PubMed: 28693034]
- Bassett DR, Toth LP, LaMunion SR, & Crouter SE (2016). Step counting: A review of measurement considerations and health-related applications. *Sports Medicine*, 47(7), 1303–1315. doi:10.1007/s40279-016-0663-1
- Berkson E, Aylward R, Zachazewski J, Paradiso J, & Gill T (2006). IMU Arrays: The biomechanics of baseball pitching. *Orthopaedic Journal at Harvard Medical School*, 8, 90–94.
- Biobank UK. (2016). Physical activity monitor (accelerometer). Retrieved from <https://biobank.ctsu.ox.ac.uk/crystal/docs/PhysicalActivityMonitor.pdf>
- Brond JC, Andersen LB, & Arvidsson D (2017). Generating ActiGraph counts from raw acceleration recorded by an alternative monitor. *Medicine & Science in Sports & Exercise*, 49(11), 2351–2360. doi:10.1249/MSS.0000000000001344 [PubMed: 28604558]
- Brond JC, & Arvidsson D (2016). Sampling frequency affects the processing of ActiGraph raw acceleration data to activity counts. *Journal of Applied Physiology* (1985), 120(3), 362–369. doi:10.1152/jappphysiol.00628.2015
- Dantzig S, Geleijnse G, & Halteren AT (2013). Toward a persuasive mobile application to reduce sedentary behavior. *Personal and Ubiquitous Computing*, 17(6), 1237–1246. doi:10.1007/s00779-012-0588-0
- De Boor C (1978). *A practical guide to splines* (Vol. 27). New York, NY: Springer-Verlag.
- Doherty A, Jackson D, Hammerla N, Plotz T, Olivier P, Granat MH, ... Wareham NJ (2017). Large scale population assessment of physical activity using wrist worn accelerometers: The UK

- Biobank study. *PLoS One*, 12(2), e0169649. doi:10.1371/journal.pone.0169649 [PubMed: 28146576]
- Ellis K, Kerr J, Godbole S, Lanckriet G, Wing D, & Marshall S (2014). A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. *Physiological Measurement*, 35(11), 2191–2203. doi:10.1088/0967-3334/35/11/2191 [PubMed: 25340969]
- Hardeman W, Houghton J, Lane K, Jones A, & Naughton F (2019). A systematic review of just-in-time adaptive interventions (JITAs) to promote physical activity. *International Journal of Behavioral Nutrition and Physical Activity*, 16(1), 31. doi:10.1186/s12966-019-0792-7
- Hildebrand M, van Hees V, Hansen BH, & Ekelund U (2014). Age group comparability of raw accelerometer output from wrist- and hip-worn monitors. *Medicine & Science in Sports & Exercise*, 46(9), 1816–1824. doi:10.1249/MSS.000000000000289 [PubMed: 24887173]
- John D, Miller R, Kozey-Keadle S, Caldwell G, & Freedson P (2012). Biomechanical examination of the ‘plateau phenomenon’ in ActiGraph vertical activity counts. *Physiological Measurement*, 33(2), 219–230. doi:10.1088/0967-3334/33/2/219 [PubMed: 22260902]
- John D, Morton A, Arguello D, Lyden K, & Bassett D (2018). “What is a step?” Differences in how a step is detected among three popular activity monitors that have impacted physical activity research. *Sensors (Basel)*, 18(4), E1206. doi:10.3390/s18041206 [PubMed: 29662048]
- John D, Sasaki J, Hickey A, Mavilia M, & Freedson PS (2014). ActiGraph activity monitors: “The firmware effect”. *Medicine & Science in Sports & Exercise*, 46(4), 834–839. doi:10.1249/MSS.000000000000145 [PubMed: 24042309]
- John D, Tyo B, & Bassett DR (2010). Comparison of four ActiGraph accelerometers during walking and running. *Medicine & Science in Sports & Exercise*, 42(2), 368–374. doi:10.1249/MSS.0b013e3181b3af49 [PubMed: 19927022]
- King K, Hough J, McGinnis R, & Perkins N (2012). A new technology for resolving the dynamics of a swinging bat. *Sports Engineering*, 15(1), 41–52. doi:10.1007/s12283-012-0084-9
- Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, & Murphy SA (2015). Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. *Journal of Health Psychology*, 34S, 1220–1228. doi:10.1037/hea0000305
- Krzysztof M, & Mero A (2013). A kinematics analysis of three best 100 m performances ever. *Journal of Human Kinetics*, 36(1), 149–160. doi:10.2478/hukin-2013-0015 [PubMed: 23717364]
- LaMunion SR, Bassett DR, Toth LP, & Crouter SE (2017). The effect of body placement site on ActiGraph wGT3X-BT activity counts. *Biomedical Physics & Engineering Express*, 3(3), 035026. doi:10.1088/2057-1976/aa777c
- Lee I-M, Shiroma EJ, Evenson KR, Kamada M, LaCroix AZ, & Buring JE (2018). Using devices to assess physical activity and sedentary behavior in a large cohort study: The Women’s Health Study. *Journal for the Measurement of Physical Behaviour*, 1(2), 60–69. doi:10.1123/jmpb.2018-0005 [PubMed: 30666321]
- Lee KY, Macfarlane D, & Cerin E (2010). Do three different generations of the ActiGraph accelerometer provide the same output? *Medicine & Science in Sports & Exercise*, 42(5), 476. doi:10.1249/01.MSS.0000385060.86877.ee
- Mannini A, Rosenberger M, Haskell WL, Sabatini AM, & Intille SS (2017). Activity recognition in youth using single accelerometer placed at wrist or ankle. *Medicine & Science in Sports & Exercise*, 49(4), 801–812. doi:10.1249/MSS.0000000000001144 [PubMed: 27820724]
- Matthews CE, Chen KY, Freedson PS, Buchowski MS, Bettina MB, Pate RR, & Troiano RP (2008). Amount of time spent in sedentary behaviors in the United States. *American Journal of Epidemiology*, 167(7), 875–881. doi:10.1093/aje/kwm390 [PubMed: 18303006]
- Mendoza AR, Lyden K, Sirard J, Staudenmayer J, Tudor-Locke C, & Freedson PS (2019). Step count and sedentary time validation of consumer activity trackers and a pedometer in free-living settings. *Journal for the Measurement of Physical Behaviour*, 2(2), 109–117. doi:10.1123/jmpb.2018-0035
- Nahum-Shani I, Hekler EB, & Spruijt-Metz D (2015). Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework. *Health Psychology*, 34S, 1209–1219. doi:10.1037/hea0000306 [PubMed: 26651462]

- Nahum-Shani I, Smith SN, Tewari A, Witkiewitz K, Collins LM, Spring B, ... Murphy S (2014). Just in time adaptive interventions (JITAI): An organizing framework for ongoing health behavior support. *Methodology Center Technical Report*, 52(6), 446–462.
- Oppenheim AV, & Schafer RW (1975). *Digital signal processing* (1st ed.). Boston, MA: Pearson.
- Pavey TG, Gilson ND, Gomersall SR, Clark B, & Trost SG (2017). Field evaluation of a random forest activity classifier for wrist-worn accelerometer data. *Journal of Science and Medicine in Sport*, 20(1), 75–80. doi:10.1016/j.jsams.2016.06.003 [PubMed: 27372275]
- Rowlands AV, & Stiles VH (2012). Accelerometer counts and raw acceleration output in relation to mechanical loading. *Journal of Biomechanics*, 45(3), 448–454. doi:10.1016/j.jbiomech.2011.12.006 [PubMed: 22218284]
- Sasaki JE, Hickey AM, Staudenmayer JW, John D, Kent JA, & Freedson PS (2016). Performance of activity classification algorithms in free-living older adults. *Medicine & Science in Sports & Exercise*, 48(5), 941–950. doi:10.1249/MSS.0000000000000844 [PubMed: 26673129]
- Sasaki JE, John D, & Freedson PS (2011). Validation and comparison of ActiGraph activity monitors. *Journal of Science and Medicine in Sport*, 14(5), 411–416. doi:10.1016/j.jsams.2011.04.003 [PubMed: 21616714]
- Shiroma EJ, Cook NR, Manson JE, Buring JE, Rimm EB, & Lee IM (2015). Comparison of self-reported and accelerometer-assessed physical activity in older women. *PLoS One*, 10(12), e0145950. doi:10.1371/journal.pone.0145950 [PubMed: 26713857]
- Shiroma EJ, Kamada M, Smith C, Harris TB, & Lee IM (2015). Visual inspection for determining days when accelerometer is worn: Is this valid? *Medicine & Science in Sports & Exercise*, 47(12), 2558–2562. doi:10.1249/MSS.0000000000000725 [PubMed: 26110697]
- Toth LP, Park S, Pittman WL, Sarisaltik D, Hibbing PR, Morton AL, & Bassett DR (2019). Effects of brief intermittent walking bouts on step count accuracy of wearable devices. *Journal for the Measurement of Physical Behaviour*, 2(1), 13–21. doi:10.1123/jmpb.2018-0050
- U.S. Food and Drug Administration. (2017). Medical devices, digital health. Retrieved from <https://www.fda.gov/MedicalDevices/DigitalHealth/default.htm>
- van Hees VT, Fang Z, Langford J, Assah F, Mohammad A, da Silva IC, & Brage S (2014). Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: An evaluation on four continents. *Journal of Applied Physiology*, 117(7), 738–744. doi:10.1152/jappphysiol.00421.2014 [PubMed: 25103964]
- van Hees VT, Gorzelniak L, Dean Leon EC, Eder M, Pias M, Taherian S, & Brage S (2013). Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. *PLoS One*, 8(4), e61691. doi:10.1371/journal.pone.0061691 [PubMed: 23626718]
- van Loo CMT, Okely AD, Batterham MJ, Hinkley T, Ekelund U, Brage S, ... Cliff DP (2018). Wrist acceleration cut points for moderate-to-vigorous physical activity in youth. *Medicine & Science in Sports & Exercise*, 50(3), 609–616. doi:10.1249/MSS.0000000000001449 [PubMed: 29023358]
- Wang JB, Cadmus-Bertram LA, Natarajan L, White MM, Madanat H, Nichols JF, & Pierce JP (2015). Wearable sensor/device (Fitbit One) and SMS text-messaging prompts to increase physical activity in overweight and obese adults: A randomized controlled trial. *Telemedicine and e-Health*, 21(10), 782–792. doi:10.1089/tmj.2014.0176 [PubMed: 26431257]
- Wijndaele K, Westgate K, Stephens SK, Blair SN, Bull FC, Chastin SF, & Healy GN (2015). Utilization and harmonization of adult accelerometry data: Review and expert consensus. *Medicine & Science in Sports & Exercise*, 47(10), 2129–2139. doi:10.1249/MSS.0000000000000661 [PubMed: 25785929]
- Wolff-Hughes DL, Fitzhugh EC, Bassett DR, & Churilla JR (2015). Total activity counts and bouted minutes of moderate-to-vigorous physical activity: Relationships with cardiometabolic biomarkers using 2003–2006 NHANES. *Journal of Physical Activity & Health*, 12(5), 694–700. doi:10.1123/jpah.2013-0463 [PubMed: 25109602]
- Wright SP, Brown TSH, Collier SR, & Sandberg K (2017). How consumer physical activity monitors could transform human physiology research. *American Journal of Physiology, Regulatory, Integrative and Comparative Physiology*, 312(3), R358–R367. doi:10.1152/ajpregu.00349.2016
- Zar JH (1984). *Biostatistical analysis* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

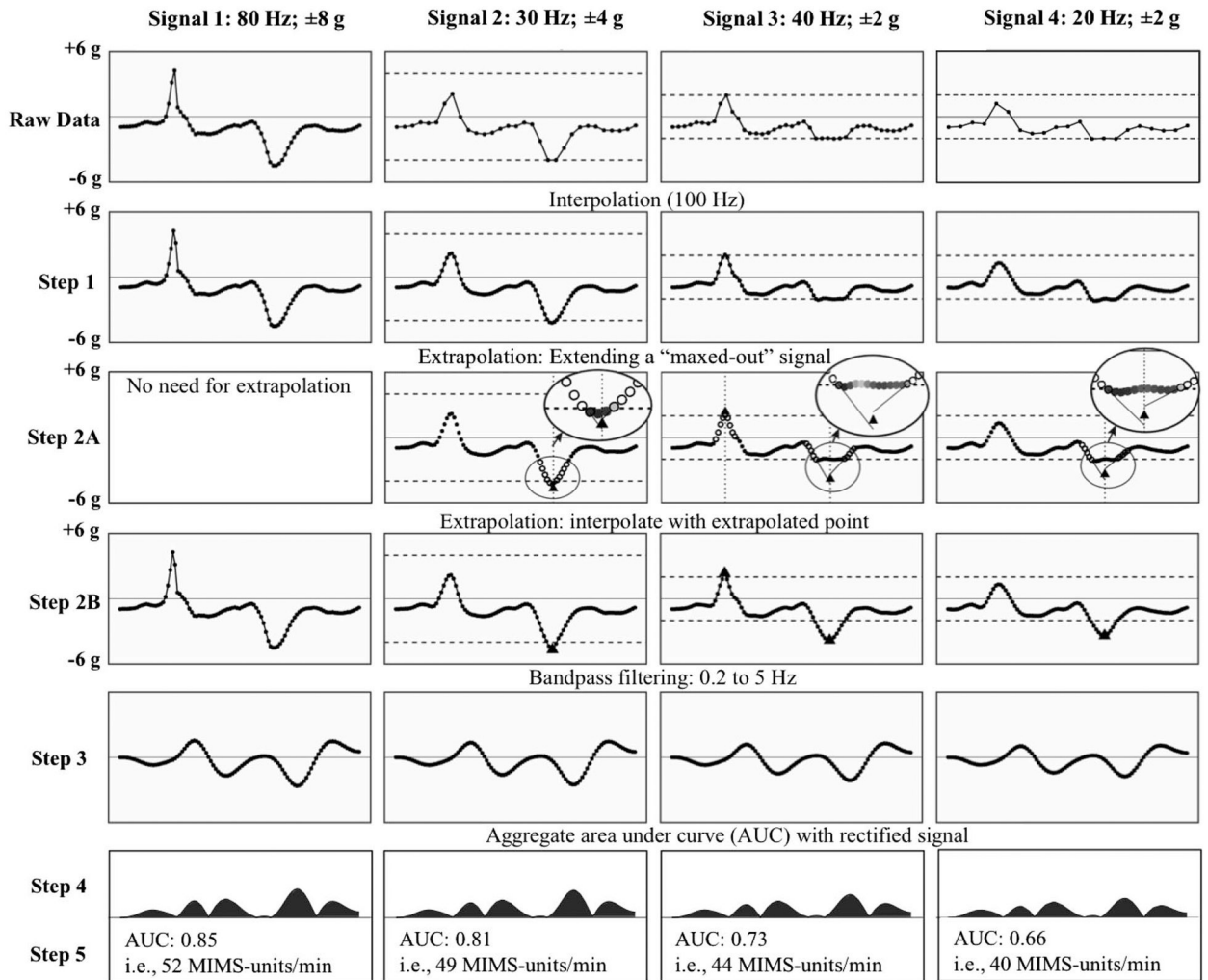


Figure 1 —. Conceptual illustration depicting the steps involved in computing Monitor-Independent Movement Summary units (MIMS-units). The original signal (1 s in duration) is depicted in the top-left quadrant, which was collected using a wrist-worn GT9X from a participant who tapped the sensor with a finger and then performed the start of a jumping jack. Signals 2 to 4 were generated from the original signal to represent variability in raw signals due to differences in sampling rate and dynamic range, which is possible when using different devices. In Step 1, the signal is interpolated to 100 Hz. In Step 2, the signal is extrapolated (where needed: steps 2A and B), which involves estimating the edge of a maxed-out condition, estimating the extrapolated point to which the signal needs to be extended, and extending the signal to the extrapolated point (e.g., refer to oval inset in Step 2, signal 3 and to step 2B, signal). In Step 3, the signal is bandpass filtered to eliminate signal components that lie outside 0.2 to 5 Hz. In Steps 4 and 5, the signal is rectified and then integrated (area under the curve is computed for the rectified signal over a desired time-period), respectively, to obtain MIMS-units for that period. A more detailed description of each step can be found in the main body of the paper. In this example, Signal 4 yields comparably low MIMS-units/min due to the low sampling rate of the original signal.

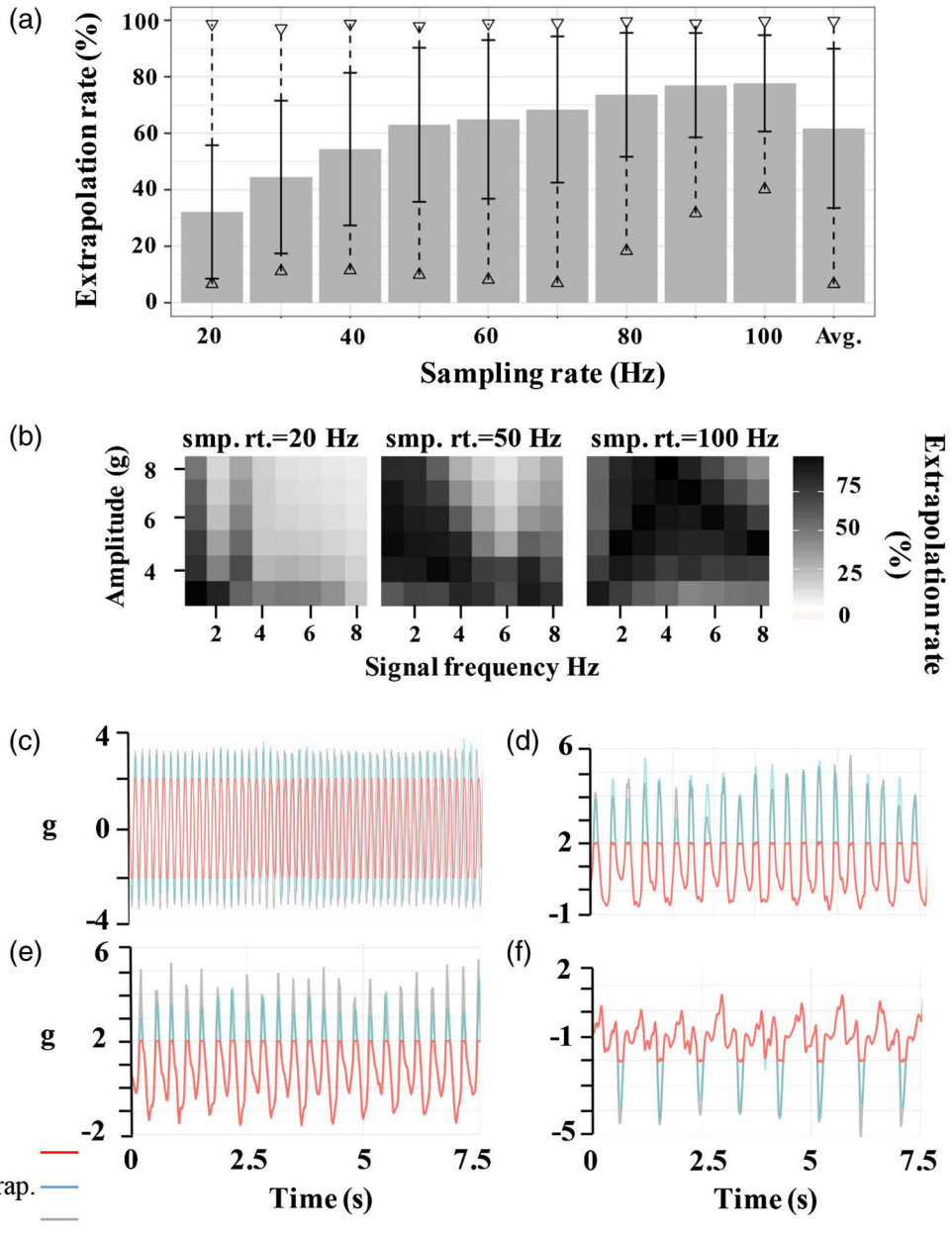


Figure 2 —
 (a) Average and variance of extrapolation rate for simulated sinusoidal signals cut off at ± 2 g with different signal frequencies ranging from 1 to 8 Hz (1 Hz increments) with sampling rates from 20 to 100 Hz (10 Hz increments); (b) grid representations of extrapolation performance tested on simulated signals with amplitudes from ± 2 to ± 8 g and frequencies from 2 to 8 Hz at sampling rates of 20, 50, and 100 Hz (left to right); (c) extrapolation of a ± 2 g (100 Hz sampling rate) signal during shaker testing (5 Hz); (d) running at 8.95 km/h (5.5 mph-ankle, 80 Hz sampling rate); (e) running at 12.1 km/h (7.5 mph-wrist, 40 Hz sampling rate); (f) and jumping jacks (wrist, 100 Hz sampling rate). In C–F, red, blue, and grey lines indicate the cut-off ± 2 g signal, the extrapolated signal, and a ground-truth ± 8 g signal, respectively. *Note.* Extrap. = extrapolated signal; smp. rt. = sampling rate.

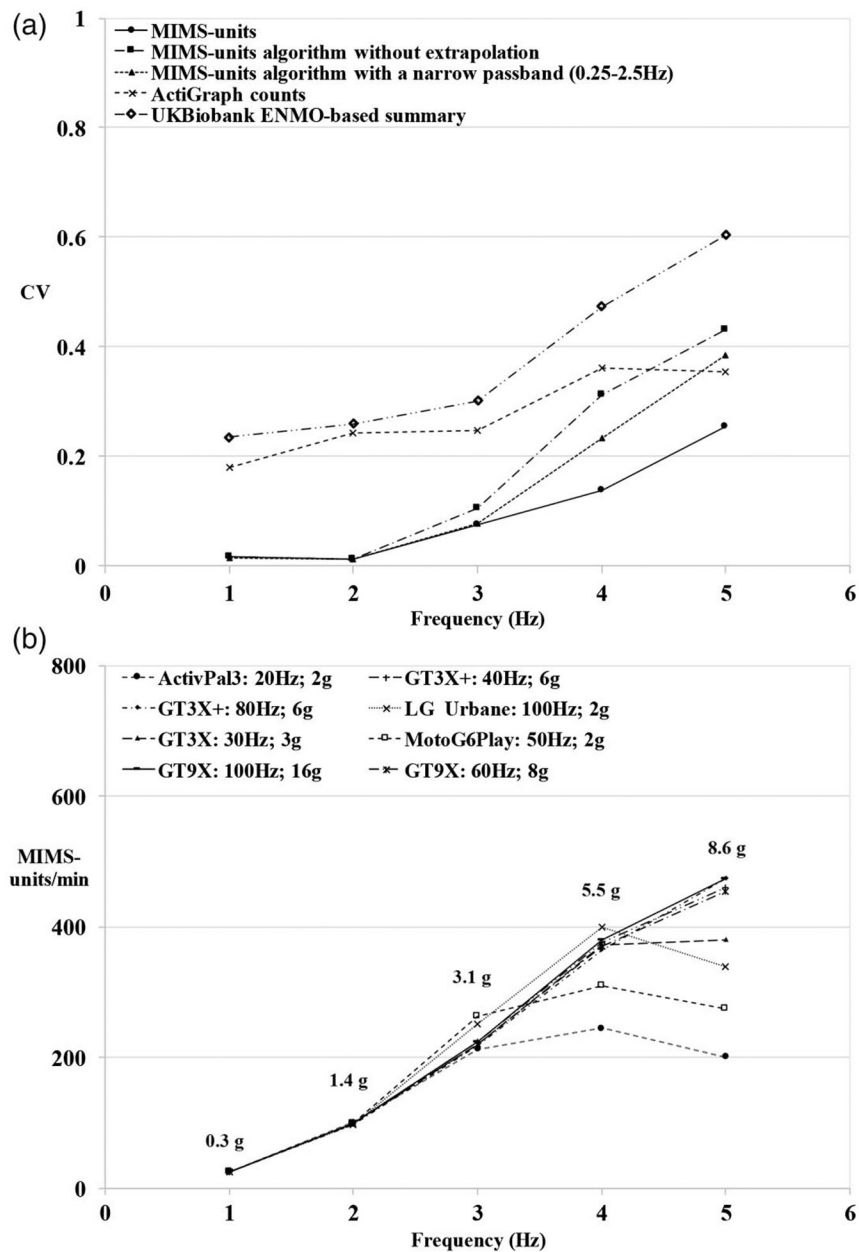


Figure 3 —.

(a) Coefficient of variation among acceleration summaries (i.e., Monitor-Independent Movement Summary units [MIMS-units] algorithm involving all steps outlined in Figure 1, MIMS-units algorithm without the extrapolation step, MIMS-units algorithm with a narrower passband of 0.25–2.5 Hz, ActiGraph counts, and UK Biobank Euclidean Norm Minus One-based summary), obtained from eight different acceleration sensors (each with different sampling rates and g range) at different frequencies during shaker testing; (b) MIMS-units from different devices during the shaker testing protocol. Numerical values in graph indicate peak g-forces generated during shaker testing.

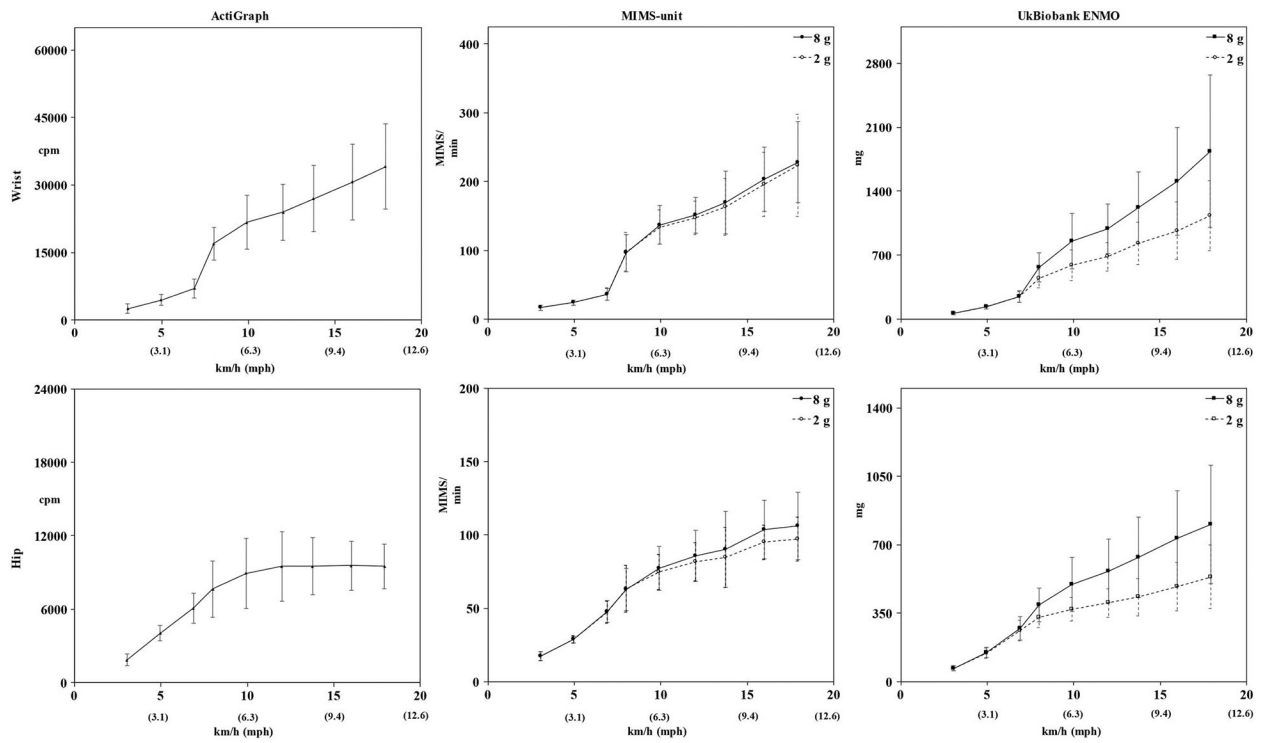


Figure 4 —. ActiGraph counts, Monitor-Independent Movement Summary units (MIMS-units), and UK Biobank Euclidean Norm Minus One (ENMO) output from a GT9X worn on the hip and wrist during the treadmill ambulation protocol. MIMS-units and UK Biobank ENMO are derived from both an ± 8 and a ± 2 g signal. Horizontal panels show patterns in different output types for the same wear location. Y-axes of graphs in the two horizontal panels are scaled to minimize visual bias that may arise due to differences in the units of the three different types of outputs.

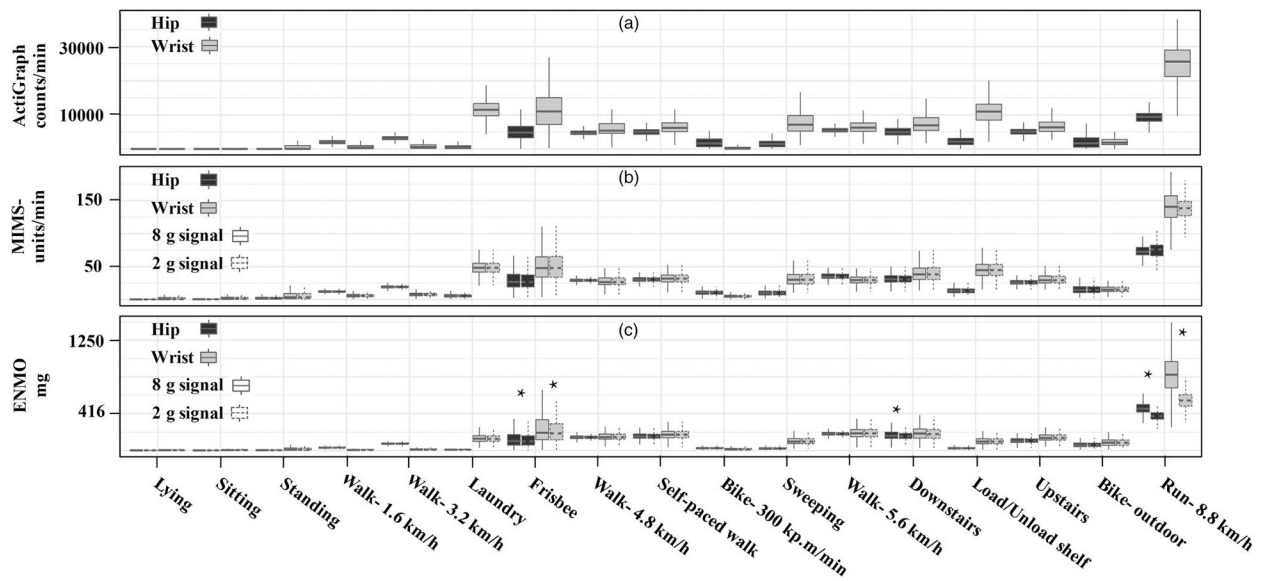


Figure 5 —.

Boxplots for (a) total ActiGraph activity counts derived from an ± 8 g signal, (b) total Monitor-Independent Movement Summary units (MIMS-units) derived, and (c) average UK Biobank Euclidean Norm Minus One (ENMO) output for the hip and wrist acceleration for various activities. MIMS-units and UK Biobank ENMO are derived from both an ± 8 and a ± 2 g signal. Activities are arranged in order of increasing MET values based on representative activities from the Compendium of Physical Activities. * indicates significant intra-location difference in outputs derived using a ± 2 and ± 8 g signal. Note: Sitting, standing, and walking at 4.8 and 5.6 km/h include grouped activities in the same posture or speed (Supplementary Table 3 [available online]).

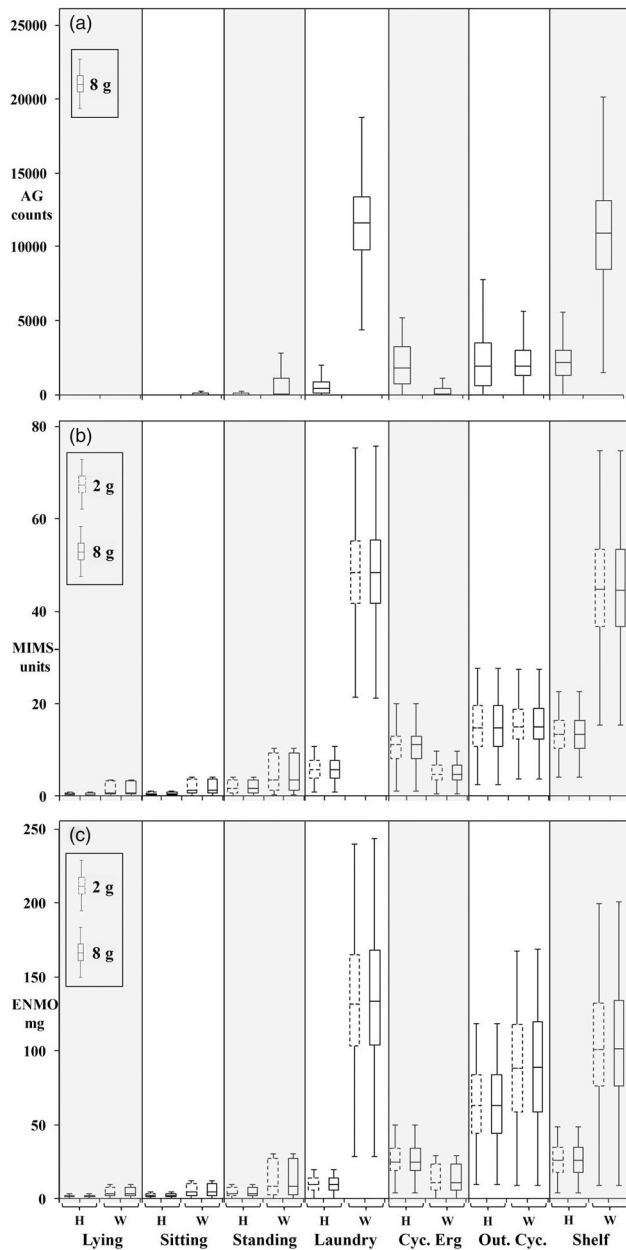


Figure 6 —.

(a) Boxplot for selected activities where at least one participant returned zero total ActiGraph activity counts derived from a ± 8 g signal; (b) and (c) corresponding total Monitor-Independent Movement Summary units and average UK Biobank Euclidean Norm Minus One output (mg) derived from an ± 8 and ± 2 g signal. *Note.* H = hip; W = wrist; Cyc. Erg. = cycle ergometry at 300 kpm/m; Out. Cyc. = outdoor cycling; AG = ActiGraph.