Validation of a Commercial Android Smartwatch as an Activity Monitoring Platform

James D. Amor, *Member, IEEE*, and Christopher J. James, *Senior Member, IEEE*

***Abstract*—Activity monitoring (AM) is an established technique for the assessment of a person's physical activity. With the rapid rise of smartwatch technology, this platform presents an interesting opportunity to use a device for AM that has both the ability to monitor activity and also the ability to interface seamlessly with other healthcare systems. There are questions however around the suitability of smartwatches as monitoring devices. This paper presents a validation of one smartwatch, the ZGPAX S8, for use as an activity monitor. Two experiments are presented: a physical manipulation test and a co-location test. In the physical manipulation test, three S8s are compared to a reference accelerometer under human physical manipulation. In the co-location test, the smartwatch is used alongside a reference device for a period of three hours by four participants to assess both the accelerometer data and the results of processing on data from both devices. Findings from these experiments show that the S8 accelerometer has a good correlation and limits of agreement in the physical manipulation test ($r^2 \sim 0.95$, CR ~ 2.5 m/s^2), and excellent correlation and limits of agreement in the analysis of processed data from the co-location experiment ($r^2 \sim 0.99$, CR ~ 0.23). From these results, the S8 is evaluated to be a suitable device for AM. Some specific limitations in the S8 are identified such as data range clipping, time drift and sample rate consistency, but these are not found to impact on the suitability of the device once algorithmic processing is applied to the data.**

***Index Terms*—Accelerometers, activity monitoring, evaluation framework, physical activity, smartwatch.**

I. Introduction

ACTIVITY monitoring (AM) is an established technique for the objective, quantitative assessment of the amount of physical activity (PA) undertaken by a person over time. AM has been used in a number of fields such as obesity [1], COPD [2] and mental health [3] and is often used instead of, or alongside traditional questionnaire or interview based methods to reduce the inaccuracies, such as recall bias, inherent in these qualitative techniques [4], [5]. AM is also well suited to longitudinal applications outside of the laboratory, where an assessment of a user's activity over several days or weeks is required [6], [7] and increasingly, activity monitors are being integrated into wider health monitoring systems and applications to provide input into a larger data analysis engine to assist in healthcare [8], [9].

The devices used for AM are often based around accelerometers, which measure the acceleration (in m/s^2 or g; 1 g = 9.801 m/s^2) of the body part to which they are attached.

There are several well established and well validated AM platforms available to researchers such as ActiGraph [10], [11] and ActivPAL [11]. These systems, whilst well validated are often standalone systems with limited ability to integrate into wider systems or automatically extract and transmit raw data without user involvement over longitudinal monitoring periods. These drawbacks are not significant for short-term laboratory based studies but can be significant when devices need to be deployed outside the laboratory for an extended duration. In particular, the limited interoperability of these devices is a significant limiting factor to their usage outside of laboratory and research domains. In these sorts of scenarios, commercial fitness monitors, such as the FitBit are becoming increasingly used [12] but these suffer from similar drawbacks around integration at the data level; whilst data can be obtained and transmitted from the device it is often into a closed, proprietary 'ecosystem' and subjected to undisclosed algorithmic processing prior to being made available.

Smartphones and smartwatches are a way around this problem as they offer good communication, sensing and user feedback potential [13]—although it should be noted that smartphones have some issues in this regard, particularly their size, indeterminate location relative to the user's body [14] and the low percentage of time (<23%) that they are typically "on the user" [15]. It is smartwatches that the authors use for their own work in this field [16], [17]. However, one critical question has yet to be fully investigated in regards to the use of smartwatches; are they actually suitable and fit for this purpose, in terms of accelerometer accuracy, time interval accuracy and timing drift, when compared to other, well validated instruments? Indeed, a search of Web of Science returns very few papers that address this question.

This paper sets out to answer this question by examining one specific smartwatch and presenting an evaluation framework that others can use. The smartwatch investigated in this paper is the ZGPAX S8 (Shenzhen PGD Digital Technology Co., Ltd., Shenzhen, China), which is an Android 4.4 platform and which the authors have significant experience using The S8 is a reasonably old device, having been on the market for a number

Manuscript received October 3, 2016; revised May 11, 2017 and June 29, 2017; accepted July 21, 2017. Date of publication October 22, 2017; date of current version June 29, 2018. This work was supported in part by EPSRC under Grant EP/M025543/1. The data that support this work are available in IEEE Xplore alongside the paper. *(Corresponding author: James David Amor.)*

The authors are with the School of Engineering, University of Warwick, Coventry CV31 2BT, U.K. (e-mail: J.D.Amor@warwick.ac.uk; C.James@warwick.ac.uk).

Digital Object Identifier 10.1109/JBHI.2017.2732678

TABLE I
SUMMARY TECHNICAL SPECIFICATION FOR THE S8

Size	58 × 42.5 × 13 mm
Weight	67 g
Screen type	Capacitive touch screen; 240 × 240 resolution
Battery	3.7 V, 470 mA·h, Li-ion
CPU	MTK6572; dual core 1.2 GHz
Memory	12 GB
RAM	512 MB
Operating system	Android 4.4
Chipset	MTK6572 (MediaTek)
Accelerometer	KXTJ2-1009 3-axis accelerometer (Kionix)

TABLE II
SUMMARY TECHNICAL SPECIFICATION FOR KXTJ2 ACCELEROMETER EMBEDDED IN THE MTK6572 SUBJECT TO THE ANDROID OS

Accelerometer chip	KXTJ2-1009 3-axis accelerometer (Kionix)
Sampling rate options	5 Hz, 16 Hz, 50 Hz
Range	$\pm 19.61\ \mathrm{m/s^2}$ (± 2 g)
Sensitivity	$\sim 0.005\ \mathrm{m/s^2}$ (~ 0.05 g)
Noise	$3.43 \times 10^{-3}\ \mathrm{m/s^2}/\sqrt{\mathrm{Hz}}\,(350\,\mu g/\sqrt{Hz})$

of years, but it has two key advantages compared with newer devices on the market. First and foremost it is significantly cheaper at around £50 when compared to newer smartwatches such as the Samsung Gear and the Huawei Watch, both of which cost over £200. Additionally, the S8 offers full standalone functionality including 3G and WiFi connectivity, which is by no means universally available in the smartwatch market. The Samsung Gear for example requires pairing with a smartphone and uses the connectivity of the smartphone. These two advantages make the S8 an attractive option for activity monitoring studies outside of the laboratory environment and are the reasons for its use in the present work.

This paper firstly examines the smartwatch accelerometer as an accelerometer in and of itself and the ability of the smartwatch to detect motion compared with a reference accelerometer. Secondly, this paper examines the algorithmic output obtained by running some analysis routines on the smartwatch data compared with the same analysis applied to data from a reference accelerometer.

The paper is organized as follows; Section II presents the experimental methodology for the two experiments undertaken; Section III presents the data analysis techniques used and the mathematics that underpins this; Section IV presents the results with a discussion of these in Section V; finally, Section VI provides some concluding remarks.

II. EXPERIMENTAL METHODOLOGY

This section describes the methodology used to assess the smartwatch against established alternatives. Two experiments were carried out. Firstly a physical manipulation experiment to test the mechanical ability of the S8 to detect acceleration compared to a reference and secondly, a co-location experiment where the S8 and a reference device (RD) were worn by participants for three hours to test the S8 against an RD under realistic conditions and provide data on which to test algorithmic output. The specifications of the S8 are provided in Table I and the specifications of the accelerometer as made available through the Android OS and smartwatch chip (which restrict some of the options natively available on the accelerometer chip) are provided in Table II.

For these experiments two different RDs were used, an Actigraph GT3X+ (Actigraph LLC, Pensacola, FL, USA) for the physical manipulation experiment due to the ease of mounting is on the back of the S8, and a GeneActiv Original (GA) (Activinsights Ltd., Kimbolton, Cambridgeshire, UK), due to ease of wearing on the wrist. Both devices are well validated and routinely used for research purposes [18]–[20].

Prior to each experiment the clock on the S8 is manually synchronized to the computer that the reference is initialized from (taking the time from the computer system time), which results in a slight time misalignment that is adjusted for in the data analysis preliminary step.

A. Physical Manipulation Experiment

For this experiment 3 S8s (S8-1, S8-2 and S8-3) were tested to assess their physical-mechanical ability to detect acceleration. The smartwatches used a modified version of the app that had previously been developed [16] to record raw data for ten minutes. Data were recorded at 16.6 Hz with a range of ± 2 g. The sampling rate of 16.6 Hz was chosen as it is sufficient to capture PA whilst not overly draining of the battery and is the setting the authors have used with this device in the past. The ± 2 g limit is not alterable within the Android framework and could potentially lead to saturation of the accelerometer during vigorous activity. Nevertheless, the range should be sufficient for the majority of daily tasks and is not considered to be problematic. An RD (ActiGraph GT3X+) was mounted on the reverse of the smartwatch in axial alignment using sticky pads to ensure a secure attachment. Prior to mounting, the GT3X+ was calibrated to ensure accuracy and was set to record data to ± 8 g at 50 Hz. The difference in sample rate between the RD and the S8 is of little concern as data are subsequently interpolated and aligned to match timings – indeed even were the sample rates the same it would be almost impossible to align the exact sample times by hand so a degree of interpolation and alignment would be necessary in either case. The higher range of the RD allows for an assessment of the degree of saturation experienced by the S8 accelerometer.

The smartwatch and attached RD (SRD) was subsequently manipulated by hand to cover a range of possible human-induced movement as well as steady state resting, as shown in Table III. The total duration of each test was about 7 minutes.

Following each test, data were retrieved from both devise and processed as shown in Section III.

B. Co-location Experiment

For this experiment four participants were asked to wear both the S8 and a GA on their non-dominant arm with the smartwatch closest to the wrist. The S8 was set to ± 2 g at 16 Hz and the GA set to ± 8 g at 50 Hz. Both devices were worn for a period of 3 hours and 5 minutes, of which the middle 3 hours is used to trim out any discrepancy from putting on or removing the

TABLE III
MOTIONS APPLIED TO DEVICE FOR PHYSICAL MANIPULATION TESTING

Manipulation	Description
Steady state	The SRD was left on a flat surface in each of 6 possible orientations. For vertical alignment the smartwatch screen was abutted to a vertical surface with a weight placed on the strap to prohibit movement.
Shaking	The SRD was shaken in each of the three axial directions at differing intensities and speeds to cover a range of motion.
Rolling	The SRD was held and rotated, principally through rotation of the wrist.
Swinging	The SRD was held and swung as the researcher walked. The device was swung at different speeds and frequencies to cover a broad range of arm swing.
Tapping	The SRD was tapped with a finger from all six sides.

devices. Participants were asked to wear the devices over their lunchtime and go about their normal activity to try to capture as much natural variation as possible. Following completion of the monitoring period raw data were collected from both devices and analyzed as shown in Section III.

III. DATA ANALYSIS

This section details the data analysis undertaken to support both experiments. In both cases, the data from the RD (GA or GT3X+) are assumed to be the gold standard and the timestamps provided by these data assumed to be correct. The data processing chain for the S8 data shares some preliminary steps from both experiments to align the two datasets in time and correct for the slight loss of time in the S8 compared to the RDs. Subsequently the analysis differs between the experiments.

A. Raw Data

Data from the accelerometers are collected in the X, Y and Z axis such that each data point has a timestamp and a value reading. Data are contained in vectors $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$, and all share a set of timestamps, contained in the vector $\mathbf{T}$, which are all assumed to be timestamps in milliseconds from a common epoch. For convenience, the notation D is used to refer to the whole dataset when necessary and includes the nominal sample rate f. To distinguish between datasets from different sources at different nominal sampling rates a lower case Greek superscript identifier is used on both the dataset, vectors and on each component and the corresponding uppercase letter denotes the total number of samples in the dataset. The dataset D^α with A samples in the dataset would be expanded as

$$D^\alpha = \{\mathbf{X}^\alpha, \mathbf{Y}^\alpha, \mathbf{Z}^\alpha, \mathbf{T}^\alpha, f^\alpha\}, \tag{1}$$

$$\mathbf{X}^\alpha = (x_1^\alpha, x_2^\alpha, x_3^\alpha, \ldots, x_{\mathrm{A}}^\alpha), \tag{2}$$

$$\boldsymbol{Y}^\alpha = (y_1^\alpha, y_2^\alpha, y_3^\alpha, \ldots, y_{\mathrm{A}}^\alpha), \tag{3}$$

$$\mathbf{Z}^\alpha = (z_1^\alpha, z_2^\alpha, z_3^\alpha, \ldots, z_{\mathrm{A}}^\alpha), \tag{4}$$

$$\mathbf{T}^\alpha = (t_1^\alpha, t_2^\alpha, t_3^\alpha, \ldots, t_{\mathrm{A}}^\alpha), \tag{5}$$

where vector components with a particular index are all at the same time instant.

B. Preliminary Processing

A simple alignment in time between data gathered from the S8 and from an RD shows both a global time-misalignment (due to the difficulty of synchronizing the clock on the S8 exactly) and a slight drift over time, such that if the start of the signal is synchronized, the end of the signal is out by a few tens of milliseconds. Ordinarily this would not be enough to worry about but for the purposes of these experiments, particularly the physical manipulation, it is important.

Synchronization and time-drift adjustment between the reference data, D^α, and the S8 data, D^β, uses the RMS values of the data in the data structures D_R^α and D_R^β along with corresponding timestamps such that

$$D_R^\alpha = \{\mathbf{R}^\alpha, \mathbf{T}^\alpha\}, \tag{6}$$

$$\boldsymbol{R}^\alpha = (r_1^\alpha, r_2^\alpha, r_3^\alpha, \ldots, r_{\mathrm{A}}^\alpha), \tag{7}$$

$$r_i^\alpha = \sqrt{\frac{x_i^\alpha + y_i^\alpha + z_i^\alpha}{3}}. \tag{8}$$

Correction of time misalignment is then essentially an optimization problem with an objective function $F()$ given by

$$\begin{aligned}\underset{\tau^P,\tau^G}{\operatorname{argmax}}\ \mathrm{F}\left(D^\alpha, D^\beta,\ \tau^P,\ \tau^G\right) \\ = \underset{\tau^P,\tau^G}{\operatorname{argmax}}\left(\left(D_R^\alpha \star D_R^\gamma\right)\left(\tau^G\right)\right),\end{aligned} \tag{9}$$

where τ^P the partial time drift in milliseconds per hour in the timestamps from the S8 and τ^G is the time lag at the maximum of the convolution function, given in this case as

$$\left(D_R^\alpha \star D_R^\gamma\right)\left(\tau^G\right) \underline{\underline{\mathrm{def}}} \sum_{i=1}^{\mathrm{A}} r_i^\alpha r_{i+\tau^G}^\delta \tag{10}$$

Since all values in both D_R are real the usual complex conjugate operation is omitted. There is also zero padding of $\boldsymbol{R}^\gamma$ to match length A such that all $r_{i>\Gamma}^\gamma = 0$. The cross-correlation is implemented using Matlab's (Matlab 2016a, The Mathworks, Inc., Natick, MA, USA) xcorr(X, Y) function. The data structure D_R^γ is a manipulated version of D_R^β that has had the time drift corrected and then been linearly interpolated to the same sampling frequency as D_R^α. These manipulations are defined as

$$D_R^\gamma = \mathrm{P}\left(D_R^\beta,\ \tau^P\right). \tag{11}$$

The function $\mathrm{P}(D_R, \tau^P)$ corrects for the partial time-shift such that $\mathbf{T}^\gamma$ is uniformly spaced at the same frequency as $\mathbf{T}^\alpha$ with

$$t_i^\gamma = t_i^\beta + \left(h^\beta \tau^P\right) \times \frac{\left(t_i^\beta - t_1^\beta\right)}{t_{\mathrm{B}}^\beta - t_1^\beta}, \tag{12}$$

where h is the length of the data D^β in hours, obtained as

$$h^B = \frac{t_{\mathrm{B}}^\beta - t_1^\beta}{1000 \times 60 \times 60}, \tag{13}$$

and $\mathbf{R}^{\gamma}$ is a resampled version of $\mathbf{R}^{\beta}$ linearly interpolated with an anti-aliasing filter applied at the time points $\mathbf{T}^{\gamma}$ using Matlab's resample(X, Tx, Fs) function.

A particle swarm optimizer [21] is used to find the solution to $F(D^{\alpha}, D^{\beta}, \tau^{P}, \tau^{G})$ subject to changes in τ^{P}, and will, in the process of doing so, find the corresponding optimum of τ^{G} as the global time-shift, through the use of xcorr.

Once optimum values of τ^{P} and τ^{G} have been found, they are used to adjust the timestamps in D^{β} to give D^{δ} such that the data in D^{δ} are correctly aligned with those in D^{α}. The partial time-shift is corrected by the function $\mathrm{P}()$ in equation (11) using the optimized τ^{P} and subsequently adding τ^{G} to all of the timestamp values. The datasets used moving forwards are the reference data D^{α} and the time-corrected S8 data D^{δ}.

C. Comparison of Physical Manipulation Output

The data from the physical manipulation experiment are compared to determine the accuracy with which the time-corrected S8 data, D^{δ}, capture the movement of the SRD, using the GT3X+ data, D^{α}, as the ground truth. In order to do this an initial step trims one or both datasets so that D^{α} overlaps D^{δ} by the smallest number of samples such that

$$t_1^{\alpha} < t_1^{\delta}, \tag{14}$$

and

$$t_{\mathrm{A}}^{\alpha} > t_{\Delta}^{\delta}. \tag{15}$$

Subsequently, an interpolated subset of D^{α} is created, as D^{ε}, such that the timestamps in D^{ε} match those in D^{δ}. The data vectors $\mathbf{X}^{\varepsilon}$, $\mathbf{Y}^{\varepsilon}$, $\mathbf{Z}^{\varepsilon}$ and the timestamp vector $\mathbf{T}^{\varepsilon}$ are created by linear interpolation of the vectors in D^{α} at the timestamps $\mathbf{T}^{\delta}$ using Matlab's interp1(X, V, Xq) function.

Once the dataset D^{ε} has been constructed a direct comparison between the points in D^{δ} and D^{ε} can be undertaken to evaluate the accuracy of D^{δ}, and by extension D^{α}, from which D^{δ} is created. This comparison is performed using both linear regression and Bland-Altman analysis [22] and the results are presented in Section IV-B. An identical analysis is also performed on the data from the co-location test, using data from the GA as the reference instead, results from this analysis are presented in Section IV-C.

D. Comparison of Algorithmic Output

The algorithm applied to the data from the GA and the S8 breaks the data down into epochs and calculates an activity score (AS) for each epoch. The AS algorithm is applied to the co-location test data, as such the reference data comes from the GA and is denoted D^{α}. The time-corrected S8 data is used and denoted as D^{δ}.

The algorithm has been developed and used in prior work [17], [23] and the output, when applied to data from a well validated accelerometer, has been shown to broadly correlate to calorific energy expenditure [23]. The mathematics of the algorithm, whilst already published across two papers [17], [23] are presented here in one place for the sake of completeness.

For a given dataset D, the AS dataset D_V is calculated. The AS dataset D_V is defined as a set containing vectors of AS values, $\mathbf{V}$, and corresponding epoch start times, $\mathbf{E}$, both of length N. The Greek superscript denotes the dataset from which D_V has been derived and α is used here as an example. The AS dataset is thus defined as

$$D_V^{\alpha} = \{\mathbf{V}^{\alpha}, \mathbf{E}^{\alpha}\}, \tag{16}$$

$$\mathbf{V}^{\boldsymbol{\alpha}} = (v_1^{\alpha}, v_2^{\alpha}, v_3^{\alpha}, \ldots, v_N^{\alpha}), \tag{17}$$

$$\mathbf{E}^{\alpha} = (e_1^{\alpha}, e_2^{\alpha}, e_3^{\alpha}, \ldots, e_N^{\alpha}). \tag{18}$$

The AS dataset D_V^{α} is calculated in three steps. Firstly the RMS value of the data is calculated to give D_R^{α} in accordance with equations (8), (9) and (10); secondly, the DC component of the signal is removed; and finally, the signal is broken up into epochs and an AS calculated for each epoch.

The DC component of the signal is removed using a 3 second sliding window to calculate and subtract the mean of the signal. The sliding window length, W, and the half-window parameter, h, both defines in number of samples, are calculated as

$$W = \begin{cases} \|3f\|, & \text{if } \|3f\| \bmod 2 = 0, \\ \|3f\| + 1, & \text{otherwise}, \end{cases} \tag{19}$$

$$h = \frac{W-2}{2}. \tag{20}$$

where f is the sampling frequency. These parameters are used to calculate the DC-removed dataset, $D_R^{'\alpha}$ from D_R^{α} as

$$D_R{}'\alpha = \left\{\mathbf{R}'^{\alpha}, \mathbf{T}^{\alpha}\right\} \tag{21}$$

$$\mathbf{R}'^{\boldsymbol{\alpha}} = (r_1'^{\alpha}, r_2'^{\alpha}, r_3'^{\alpha}, \ldots, r_{\mathrm{A}}'^{\alpha}), \tag{22}$$

$$r_i'^{\alpha} \begin{cases} r_i^{\alpha} - \frac{1}{W}\sum_{j=1}^{W} r_j^{\alpha}, & 1 \le i \le h, \\ r_i^{\alpha} - \frac{1}{W}\sum_{j=i-h}^{i+h} r_j^{\alpha}, & h \le i \le (\mathrm{A}-\mathrm{h}), \\ r_i^{\alpha} - \frac{1}{W}\sum_{j=\mathrm{A}-\mathrm{W}+1}^{\mathrm{A}} r_j^{\alpha}, & (\mathrm{A}-\mathrm{h}) \le i \le \mathrm{A}, \end{cases} \tag{23}$$

whilst maintaining the corresponding timestamps $\mathbf{T}^{\alpha}$. Subsequently, the vector $\mathbf{R}'^{\alpha}$ is broken into epochs with the epoch timings, $\mathbf{E}^{\alpha}$, calculated as

$$e_j^{\alpha} = T_s + (j-1) \times L, \tag{24}$$

where T_s is the start time for the AS epochs, typically $\max(t_1^{\alpha}, t_1^{\delta})$, and can be used to synchronize AS vectors across different datasets; and L is the epoch length in milliseconds. The corresponding AS scores are calculated with the use of some intermediate calculations where

$$\mathbf{D}_j''^{\alpha} \subseteq \mathbf{D}'^{\alpha} \tag{25}$$

such that the vector subsets $\mathbf{T}_j''^{\alpha}$ and $\mathbf{R}_j''^{\alpha}$ are defined as

$$I_j''^{\alpha} = \left\{i | e_j^{\alpha} \le t_i^{\alpha} < e_{j+1}^{\alpha}\right\}, \tag{26}$$

$$\mathbf{T}_j''^{\alpha} = \left\{t_i^{\alpha} | i \in I_j''^{\alpha}\right\}, \tag{27}$$

$$\mathbf{R}_j''^{\alpha} = \left\{r_i'^{\alpha} | i \in I_j''^{\alpha}\right\}. \tag{28}$$

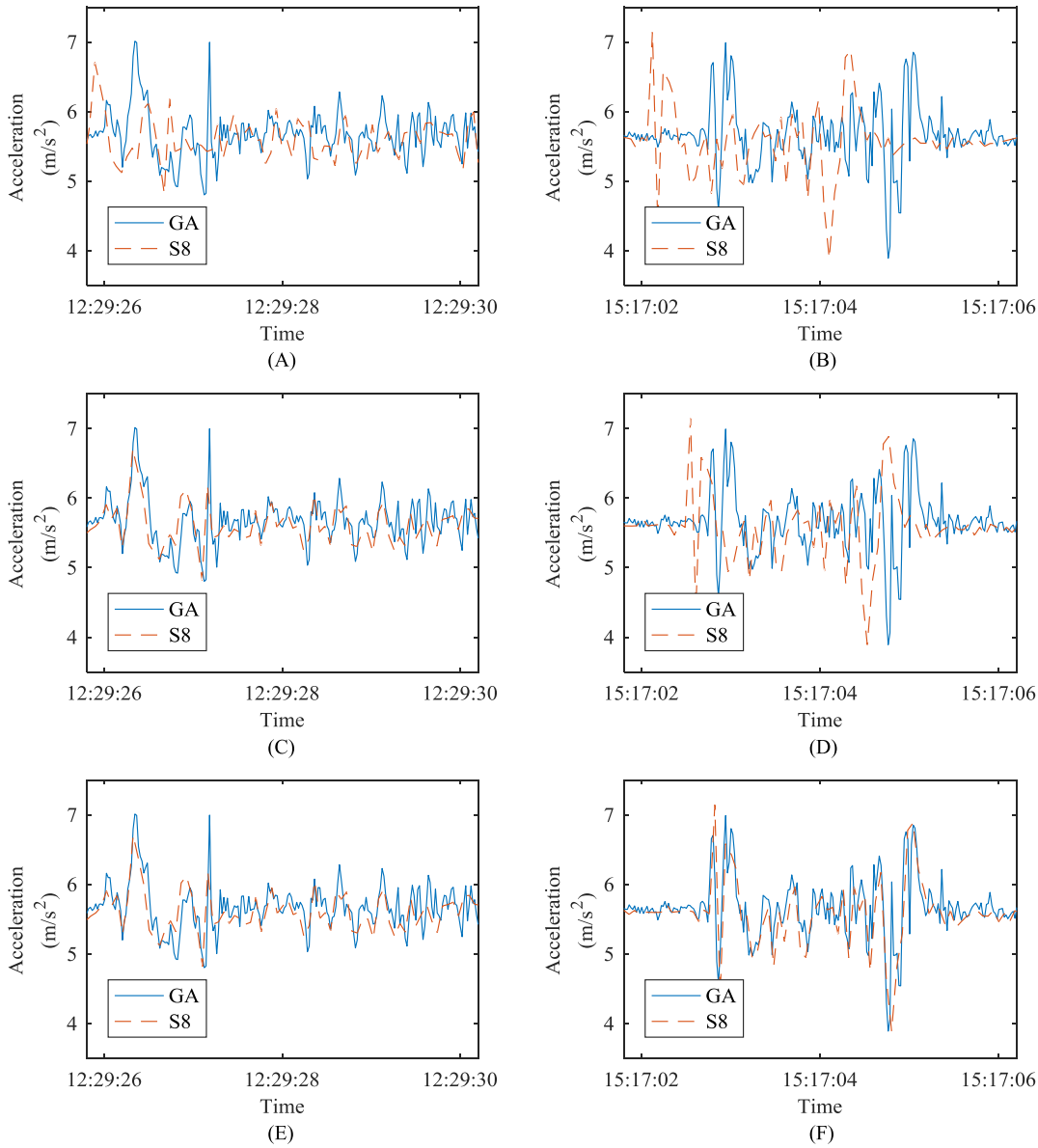

Fig. 1. Plots of the GA reference data and S8 data during the pre-processing step from Participant 1 in the co-location test showing (A) and (B) uncorrected data, (C) and (D) data corrected with only a global shift and (E) and (F) fully corrected data. Two sections of data are shown: (A), (C), and (E) one from the start of the data and (B), (D), and (F) one near the end, in which the difference between global shifting and full adjustment can be clearly seen.

Subsequently, the calculation of v_j^α is

$$v_j^\alpha = \begin{cases} 4u_j^\alpha, & \text{if } u_j^\alpha \leq 0.5 \\ 2u_j^\alpha + 1, & \text{if } u_j^\alpha > 0.5 \end{cases} \tag{29}$$

where

$$u_j^\alpha = \frac{1}{n} \sum_{j}^{\mathbf{R}''^\alpha} \tag{30}$$

and n is the total number of values in the vector subset $\mathbf{R}_j''^\alpha$.

In this way, AS datasets D_V^α and D_V^δ can be constructed from the reference data D^α and time corrected test data D^δ respectively, such that the epoch timings in both AS datasets correspond to each other. This allows both pairwise regression analysis and Bland-Altman analysis to be applied.

IV. Results

A. Pre-processing

Fig. 1 shows the results of the pre-processing step, highlighting two short sections of data from Participant 1, one section from the beginning of the data (A, C, E) and one from the end of the recording (B, D, F). The figure shows the uncorrected (A, C), globally time-shifted (B, D) and fully corrected data (E, F). It can be seen in the figure that the global time-shift acts to correct the alignment of the data at the start of the signal, shown in the contrast between A and C, but that at the end of the

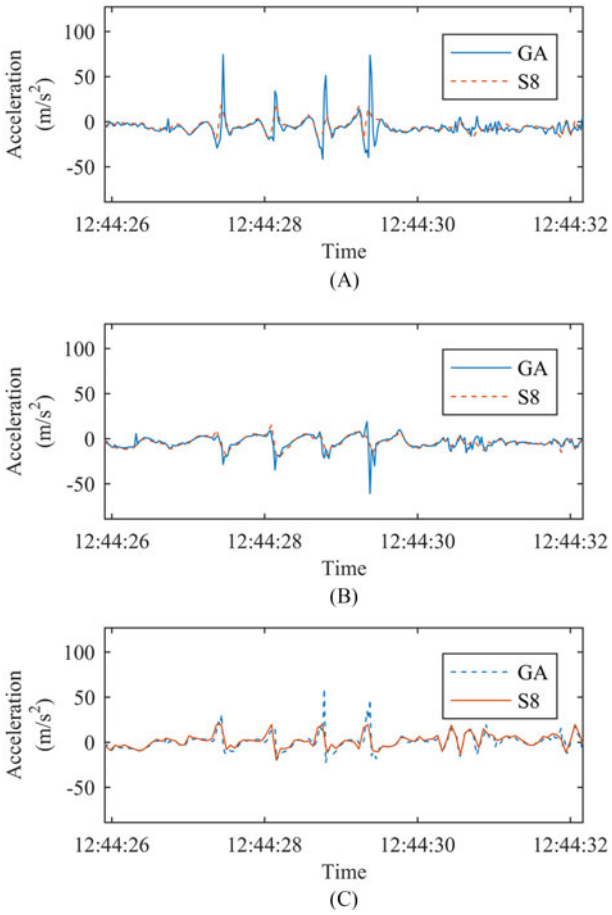

Fig. 2. Example of data from the physical manipulation testing showing a period of rolling. The acceleration recorded from the GT3X+ and the S8 can be seen in the X, Y and Z axis.

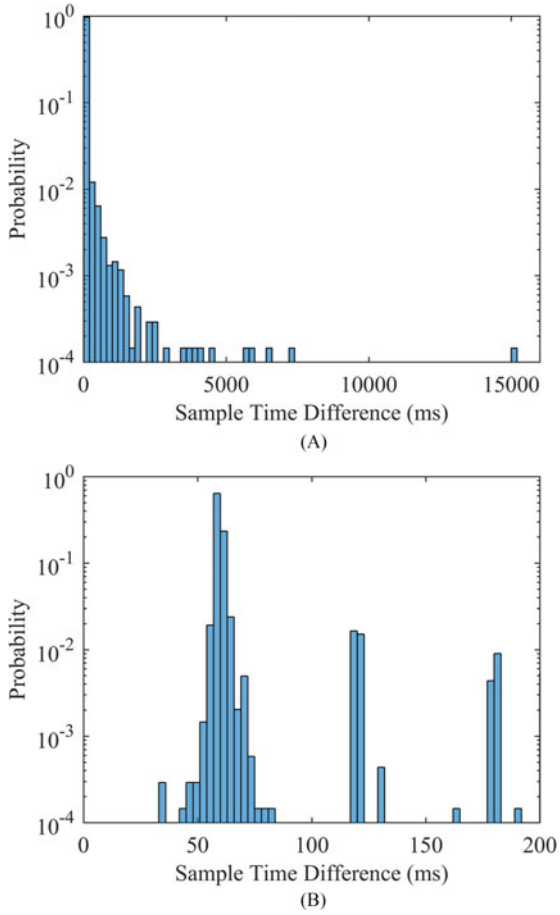

Fig. 3. Probabilistic histogram plots of the sample to sample time differences from the S8 for the physical manipulation test for S8-1. (A) The full range of differences is shown along with (B) close up of the time differences between 0 and 200 ms. The nominal sample rate was 16.6 Hz with an expected time difference of 60 ms.

signal, the GA and S8 data are misaligned, shown in B and D. The full time-shift correction, shown in E and F fully corrects the misalignment.

B. Physical Manipulation Testing

Fig. 2 shows plots of data in the three axis of measurement from the physical manipulation testing at the two different frequencies. It can be seen from this plot that the S8 data closely mirrors the GT3X+ data, albeit not perfectly. Some element of data range clipping can also be seen in this figure, particularly in the X-Axis around 12:44:28 where the GT3X+ data goes outside of the range of the S8 the S8 data is clipped at its maximal range.

Fig. 3 shows probabilistic histogram plots of the times in milliseconds between each sample from one of the runs of the physical manipulation test. Some summary statistics from all the runs are presented in Table IV. From the figure and the table, it can be noted that the sample rate of the device is not consistent and that the actual time differences occur around multiples of the expected difference (in this case 60 ms). It can be seen that while the S8 operates within ±5 ms of the requested sample rate approximately 90% of the time, there are noticeable deviations from this, which, in some cases, are several orders of magnitude above the expected delay between samples of 60 ms.

TABLE IV
TABLE OF SUMMARY STATISTICS FROM SAMPLE-TO-SAMPLE TIME DIFFERENCES FROM THE PHYSICAL MANIPULATION TEST

Device	Mean	Median	Mode	Skew.	55–65 ms
S8-1	87.03	60.00	60.00	30.27	90.55%
S8-2	99.90	60.00	60.00	24.71	88.86%
S8-3	120.76	60.00	60.00	32.52	89.85%
Mean	102.52	60.00	60.00	29.16	90.00%

Fig. 4 shows an example of the regression analysis and Bland-Altman analysis from the physical manipulation experiment from the X axis of device S8-1. Tables V and VI present the summary statistics for the whole physical manipulation test. It can be seen from the figure and the table that the S8 performs relatively well when compared to the GT3X+. All three devices show a consistent CR across their three axis, and the r2 values are high across the board, with regression slopes close to 1. The exception to the good performance is the calibration offset of the devices and the data range clipping.

The calibration offsetis shown in the mean values for the Bland-Altman analysis and the intercept values in the regression analysis. Whilst these are generally below ±0.5 m/s^2 they are

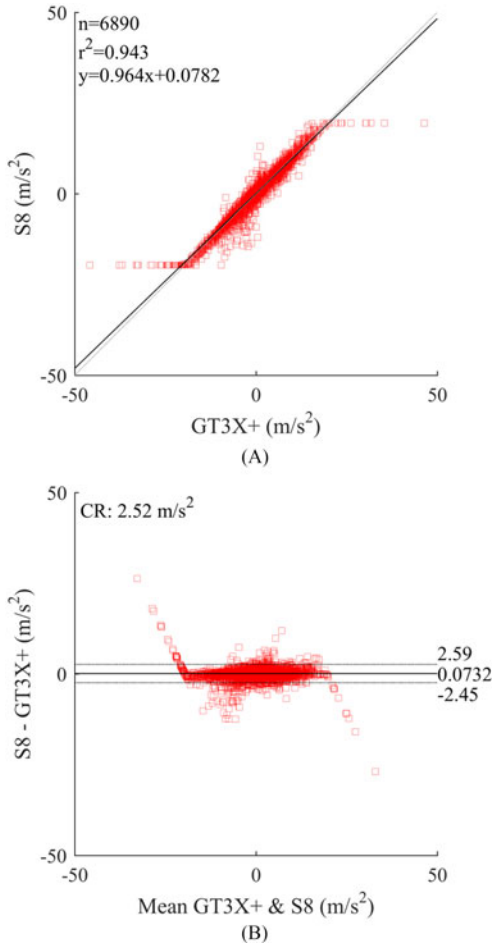

Fig. 4. (A) Regression analysis and (B) Bland-Altman analysis of data from the X-axis at 16 Hz for the physical manipulation test for S8-1. The regression analysis shows the number of data points (n), the r^2 value; and the fitted regression line and its formula. The Bland-Altman plot shows the mean, CR and limits of agreement.

highly variable and the Z-Axis from S8-3 shows a very large offset. This is shown graphically in Fig. 5, where the offset can clearly be seen. It is highly probable that this is some form of factory calibration error as the offset is close to 9.801 m/s^2 (1 g).

The data clipping can clearly be seen in Fig 4, as the GA data extends past the ± 2 g range, the S8 data caps at this value.

C. Co-location Data Equivalency Test

Fig. 6 shows an example of the regression analysis and Bland-Altman analysis for proximal-distal axis data from Participant 1 in the co-location test. As with the physical manipulation tests, a strong linear correlation can be seen in the scatter diagram and a broadly good agreement between the two devices is shown in the Bland-Altman diagram.

Table VII shows statistics derived from scatter plots of the time corrected S8 data and the interpolated GA data from the co-location test in the three axis of measurement. It can be seen from this data that, in general, there is a strong positive correlation between GA and S8, demonstrated by the r^2 and slope values. The linear regression through the data, shown by the slope and intercept values shows that the relationship between the data is almost one-for-one, albeit subject to a slight offset. The high r^2 values indicate that the linear model is a good fit for the data.

Table VIII shows the statistics derived from the Bland-Altman analysis of the time corrected S8 data and the interpolated GA data. This table confirms the constant offset between the two datasets as generally between 0.75 and 0.83 m/s^2 (with the exception of Participant 2) and also shows a fairly small coefficient of repeatability (CR) which indicates that the two methods are in good agreement with one another.

Taken together, these two sets of statistics show that, on the co-location test data, the two devices are in good, but not perfect agreement. The exception to this is data for Participant 2, for which the results are significantly worse, particularly the r^2 and

TABLE V
STATISTICAL RESULTS OF THE REGRESSION ANALYSIS ON THE THREE S8 DEVICES USED IN THE PHYSICAL MANIPULATION TEST

Device	Slope	Int.	r^2	F-stat	P-val.
X-axis					
S8-1	0.96	0.08	0.94	1.14×10^5	$\ll$0.0001
S8-2	1.00	0.27	0.95	1.14×10^4	$\ll$0.0001
S8-3	0.96	1.14	0.92	5.79×10^4	$\ll$0.0001
Mean	0.97	0.50	0.94	9.54×10^4	$\ll$0.0001
(SD)	(0.02)	(0.56)	(0.02)	(3.25×10^5)	(0.00)
Y-axis					
S8-1	0.97	−0.46	0.96	1.66×10^5	$\ll$0.0001
S8-2	0.99	−0.13	0.97	1.73×10^5	$\ll$0.0001
S8-3	0.97	−0.93	0.97	1.54×10^5	$\ll$0.0001
Mean	0.98	−0.51	0.97	1.64×10^5	$\ll$0.0001
(SD)	(0.01)	(0.40)	(0.01)	(9.33×10^3)	(0.00)
Z-axis					
S8-1	0.95	0.31	0.97	2.33×10^5	$\ll$0.0001
S8-2	0.93	−0.02	0.96	1.32×10^5	$\ll$0.0001
S8-3	0.97	9.31*	0.98	2.46×10^5	$\ll$0.0001
Mean	0.95	3.2	0.97	2.04×10^5	$\ll$0.0001
(SD)	(0.02)	(5.29)*	(0.01)	(6.22×10^4)	(0.00)

*S8-3 factory calibration error unduly influence mean and SD.

TABLE VI
SUMMARY STATISTICS FROM THE BLAND-ALTMAN ANALYSIS OF THE PHYSICAL MANIPULATION TEST DATA FOR THE THREE DEVICES

Device	M	CR	ULA	LLA
X-axis				
S8-1	0.07	2.52	2.59	−2.45
S8-2	0.27	2.39	2.66	−2.13
S8-3	1.13	2.53	3.66	−1.40
Mean (SD)	0.49 (0.56)	2.48 (0.08)	2.97 (0.60)	−2.00 (0.53)
Y-axis				
S8-1	−0.41	2.59	2.18	−3.00
S8-2	−0.12	2.37	2.25	−2.48
S8-3	−0.96	2.61	1.65	−3.57
Mean (SD)	−0.50 (0.43)	2.52 (0.13)	2.03 (0.33)	−3.02 (0.55)
Z-axis				
S8-1	0.33	2.89	3.21	−2.56
S8-2	0.18	3.75	3.93	−3.57
S8-3	9.36*	2.30	11.67*	7.06*
Mean (SD)	3.29 (5.26)*	2.98 (0.73)	6.27 (4.69)*	0.31 (5.87)*

*S8-3 factory calibration error unduly influence mean and SD.

TABLE VII
RESULTS FROM PAIRWISE REGRESSION ANALYSIS BETWEEN GA AND S8 DATA POINTS FROM CO-LOCATION DATA IN THE THREE AXIS OF MEASUREMENT

P.	Slope	Int.	r^2	F-stat.	P-val.
Radioulnar axis					
1	0.94	−1.02	0.94	2.24×10^5	≪0.0001
2	0.70	−2.62	0.59	1.97×10^4	≪0.0001
3	0.89	−1.41	0.91	1.51×10^5	≪0.0001
4	0.92	−1.17	0.91	1.42×10^5	≪0.0001
Mean	0.86	−1.55	0.84	1.34×10^5	≪0.0001
(SD)	(0.11)	(0.73)	(0.17)	(8.47×10^4)	(0.00)
Proximal-distal axis					
1	1.00	0.99	0.98	8.85×10^5	≪0.0001
2	0.94	−0.88	0.98	5.41×10^5	≪0.0001
3	0.97	0.17	0.97	4.60×10^5	≪0.0001
4	0.97	0.10	0.98	7.15×10^5	≪0.0001
Mean	0.97	0.10	0.98	6.50×10^5	≪0.0001
(SD)	(0.02)	(0.77)	(0.01)	(1.89×10^5)	(0.00)
Palmar-dorsal axis					
1	0.73	−1.23	0.90	1.23×10^5	≪0.0001
2	0.88	0.87	0.86	8.24×10^4	≪0.0001
3	0.87	0.39	0.91	1.44×10^5	≪0.0001
4	0.80	−0.76	0.94	2.41×10^5	≪0.0001
Mean	0.82	−0.18	0.90	1.47×10^5	≪0.0001
(SD)	(0.07)	(0.98)	(0.04)	(6.73×10^4)	(0.00)

TABLE VIII
RESULTS OF BLAND-ALTMAN ANALYSIS BETWEEN GA AND S8 DATA POINTS FOR CO-LOCATION DATA IN THE THREE AXIS OF MEASUREMENT

Participant	Mean	CR[a]	ULA[b]	LLA[c]
Radioulnar axis				
1	−0.86	1.43	0.56	−2.29
2	−1.65	4.25	2.59	−5.90
3	−0.81	1.97	1.16	−2.78
4	−0.80	2.15	1.35	−2.95
Mean (SD)	−1.03 (0.41)	2.45 (1.24)	1.41 (0.85)	−3.48 (1.64)
Proximal-distal axis				
1	0.99	1.78	2.78	−0.79
2	−0.74	1.69	0.95	−2.43
3	0.18	1.92	2.10	−1.73
4	0.12	1.65	1.78	−1.53
Mean (SD)	0.14 (0.71)	1.76 (0.12)	1.90 (0.76)	−1.62 (0.68)
Palmar-dorsal axis				
1	−0.15	3.30	3.16	−3.45
2	1.42	3.30	4.73	−1.88
3	0.69	2.72	3.41	−2.04
4	−0.31	2.72	2.41	−3.02
Mean (SD)	0.41 (0.80)	3.01 (0.34)	3.43 (0.96)	−2.60 (0.76)

[a]Coefficient of repeatability, [b] upper limit of agreement, [c] lower limit of agreement.

CR values for the radioulnar axis, which could be due to device slippage on the participant's wrist as the strap could not be tightened enough on the S8.

D. Algorithmic Output Test

Fig. 7 shows a line plot of the processed data for Participant 1 from the co-location experiment, analyzed with a 30 second epoch. From the figure it can be seen that the processed S8 data tracks the GA data very well, except in the upper end of activity score, where the S8 data overestimates the GA data by a small margin.

Fig. 8 shows an example of the regression analysis and Bland-Altman analysis performed on the processed data from the co-location test for Participant 1. The figure shows a very tight correlation in the regression analysis, which is mirrored in the

TABLE IX
RESULTS FROM PAIRWISE REGRESSION ANALYSIS BETWEEN ALGORITHMIC OUTPUT FROM THE GA AND S8 DATA FROM THE CO-LOCATION EXPERIMENT PROCESSED USING DIFFERENT EPOCH LENGTHS

P.	Slope	Int.	r^2	F-stat.	P-val.
10 s Epoch (N = 1080)					
1	0.93	0.00	1.00	3.49×10^5	≪0.0001
2	0.90	0.01	0.98	5.59×10^4	≪0.0001
3	0.98	−0.01	0.98	6.26×10^4	≪0.0001
4	0.95	−0.01	0.99	1.09×10^5	≪0.0001
Mean	0.94	−0.00	0.99	1.44×10^5	≪0.0001
(SD)	(0.03)	(0.01)	(0.01)	(1.39×10^5)	(0.00)
30 s Epoch (N = 360)					
1	0.93	−0.00	1.00	2.27×10^5	≪0.0001
2	0.90	0.00	0.99	2.50×10^4	≪0.0001
3	0.97	−0.01	0.99	3.44×10^4	≪0.0001
4	0.95	−0.01	0.99	5.52×10^4	≪0.0001
Mean	0.94	−0.01	0.99	8.53×10^4	≪0.0001
(SD)	(0.03)	(0.01)	(0.01)	(9.50×10^4)	(0.00)
60 s Epoch (N = 180)					
1	0.93	−0.00	1.00	1.14×10^5	≪0.0001
2	0.90	0.00	0.99	1.48×10^4	≪0.0001
3	0.97	−0.02	0.99	1.87×10^4	≪0.0001
4	0.95	−0.01	1.00	4.31×10^4	≪0.0001
Mean	0.94	−0.01	0.99	4.78×10^4	≪0.0001
(SD)	(0.03)	(0.01)	(0.00)	(4.62×10^4)	(0.00)

TABLE X
RESULTS BLAND-ALTMAN ANALYSIS BETWEEN ALGORITHMIC OUTPUT FROM THE GA AND S8 DATA FROM THE CO-LOCATION EXPERIMENT PROCESSED USING DIFFERENT EPOCH LENGTHS

Participant	Mean	CR[a]	ULA[b]	LLA[c]
10 s Epoch				
1	−0.03	0.12	0.09	−0.15
2	−0.01	0.06	0.05	−0.07
3	−0.02	0.09	0.08	−0.11
4	−0.02	0.10	0.08	−0.12
Mean (SD)	−0.02 (0.01)	0.09 (0.02)	0.08 (0.02)	−0.11 (0.03)
30 s Epoch				
1	−0.04	0.11	0.07	−0.14
2	−0.01	0.05	0.04	−0.07
3	−0.02	0.06	0.04	−0.09
4	−0.03	0.08	0.06	−0.11
Mean (SD)	−0.02 (0.01)	0.08 (0.02)	0.05 (0.01)	−0.10 (0.03)
60 s Epoch				
1	−0.04	0.10	0.07	−0.14
2	−0.02	0.05	0.03	−0.06
3	−0.03	0.06	0.03	−0.08
4	−0.03	0.07	0.04	−0.10
Mean (SD)	−0.03 (0.01)	0.07 (0.03)	0.04 (0.02)	−0.10 (0.03)

[a]Coefficient of repeatability, [b]upper limit of agreement, [c]lower limit of agreement.

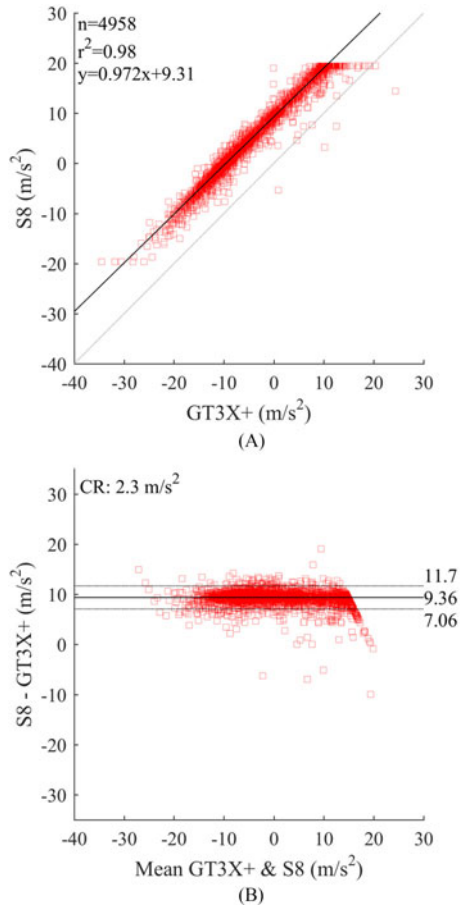

Fig. 5. (A) Regression analysis and (B) Bland-Altman analysis of the Z-axis from S8-3 in the physical manipulation experiment showing the calibration offset. The regression analysis shows the number of data points (n), the r^2 value; and the fitted regression line and its formula. The Bland-Altman plot shows the mean, CR and limits of agreement.

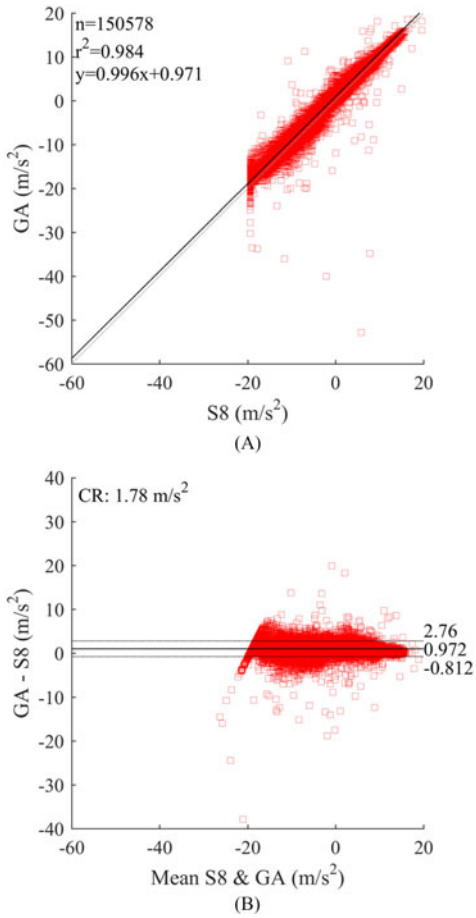

Fig. 6. (A) Example of regression analysis and (B) Bland-Altman analysis for co-location data from the proximal-distal axis for Participant 1. The regression analysis shows the number of data points (n), the r^2 value; and the fitted regression line and its formula. The Bland-Altman plot shows the mean, CR and limits of agreement.

Bland-Altman analysis, which shows fairly tight limits of agreement for the data.

Tables IX and X show the results of the regression analysis and Bland-Altman analysis on the algorithmic output of the data from the co-location experiment. The statistical tables show that all four participants exhibit a very good agreement between the GA and S8 data once the data are processed as discussed in Section III-D. This is evidenced by the high r^2 scores across the board and by the low values of CR in the Bland-Altman analysis. It should be noted however, that there is a small but constant estimation bias in the data resulting in a consistent offset intercept in the regression analysis and off-zero mean score in the Bland-Altman analysis.

V. Discussion

The results presented in Section IV, taken in their entirety show the S8 as a device which is well suited to being used as a platform for AM, even if it is not very good as an accelerometer, if analyzed specifically as such, as evidenced by the data in Tables V–VIII, particularly the large SDs calculated showing high inter-device variability.

The results from the pre-processing, physical manipulation analysis and data equivalency analysis on the co-location data show that the S8 accelerometer is functional and broadly accurate, whilst suffering from a number of specific issues. In particular are the time-shifting, sample rate consistency, measurement range-clipping, and calibration accuracy.

The time-shifting, shown in Fig. 1 is a very minor issue and unlikely to be a factor if the S8 is used in isolation, and even when used in conjunction with other systems, will only be an issue if millisecond time alignment is required. In isolation, and where millisecond time accuracy is not required, the S8 keeps time well enough. Additionally, for long-running tasks, where the time-shifting could add up to around two seconds a week, the S8 can be set to synchronize its time to a timeserver to eliminate the issue. Were the S8 to be used in conjunction with a smartphone, then both devices could be set to synchronize to the same timeserver. The end result of the processing shown in this work is an activity score over a 30 second epoch and the errors introduced into this metric by millisecond level timing issues will be vanishingly small.

The sample-rate consistency, shown in Fig. 3 is potentially a concern and stems from the S8 being an Android platform.

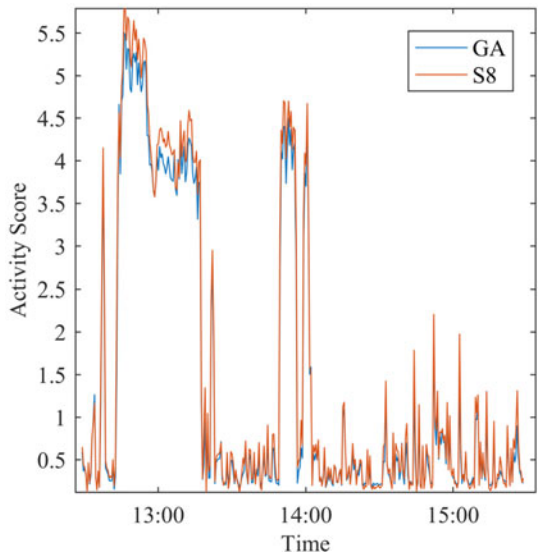

Fig. 7. Line plot of the processed data from the co-location experiment for Participant 1 at the 30 second epoch level.

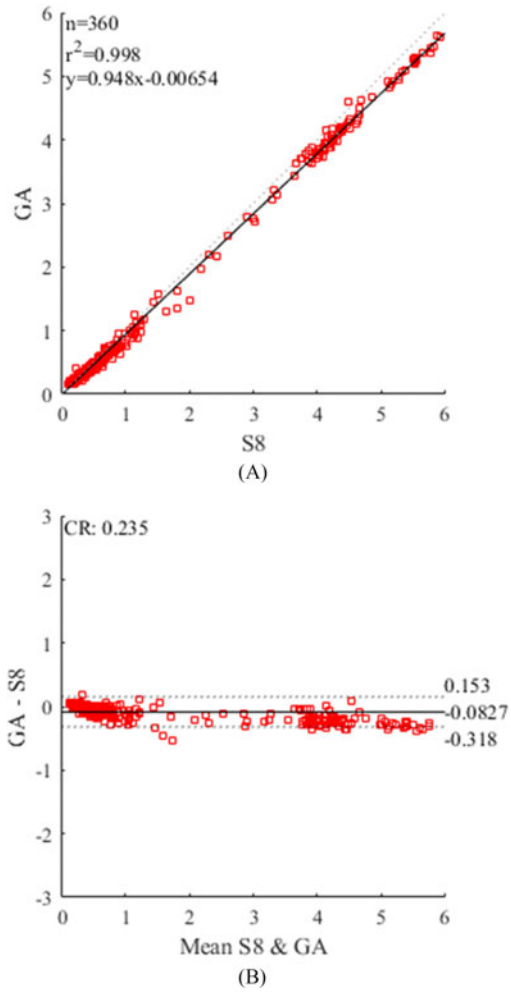

Fig. 8. (A) Regression analysis and (B) Bland-Altman analysis of processed co-location data from Participant 1. The regression analysis shows the number of data points (n), the r^2 value; and the fitted regression line and its formula. The Bland-Altman plot shows the mean, CR and limits of agreement.

Android is, by its very nature, designed to be a smart-phone platform and concordantly, the accelerometer is not given priority by the operating system. As such, when the OS has higher priority tasks to do, fewer system resources are devoted to sensor sampling, resulting in sample rate inconsistency. This is compounded by the fact that an accelerometer on an Android device will only report changes in detected value resulting in potentially large gaps between readings if the device is stationary, which can result in some very large time differences such as those in Fig. 3. For applications where sampling rate consistency is important this renders the S8 (and other Android platforms – the authors cannot comment as to iOS) unsuitable. For long-term AM, as presented in this paper, the sample rate inconsistency is not an issue as the processing step to obtain AS is robust to differing data counts in each epoch.

The data-clipping issue occurs on the S8 to limit the data to the ± 2 g range. This is an inherent limitation of the base chipset used in the hardware and is not controllable from the app level on the device. As can be seen in Fig. 4, recorded acceleration was mostly contained within ± 2 g, although it did regularly exceed this.

The presence of clipping indicates that the S8 is not suited to capture of highly vigorous activity as the clipping effect at high acceleration values would limit the ability of the accelerometer and subsequent processing to distinguish between different intensity levels once the saturation point was reached. However, the agreement in the processed data between the Geneactive (± 8 g) and S8 (± 2 g) methodologies, shown in Tables IX and X, was incredibly strong. This indicates that there is little practical difference between monitoring general activity at ± 2 g compared to ± 8 g once downstream processing is applied and shows the S8 to be suitable for this purpose.

As with the other device limitations, for applications where a high acceleration range is essential this would be an issue, but for the applications presented in this paper it is a limitation that is worked around and that does not adversely affect the results.

Calibration inaccuracy is to be expected to a certain extent in all accelerometers and is present to a small degree in the S8's that were tested. Whilst the results shown in this work indicate that a small calibration error will not adversely affect the results of algorithmic processing, the poor result from the Z-Axis of S8-3 is of some concern and may point to manufacturing quality control issues. It would be sensible therefore to suggest that all devices of this type, that is all commercially available tools for which activity sensing is not the primary goal, be calibrated and tested to ensure any calibration errors are within acceptable bounds; and calibrated and corrected if necessary and possible.

In spite of the device limitations however, the results from the algorithmic output comparison show a very good agreement between the processed values obtained from the S8 compared with those from the GA across different epoch lengths. The Bland-Altman analysis shows that the S8 is usually within $\sim \pm 0.23$ of the GA for the AS and the line graph shows the two measures in good agreement. In the judgement of the authors (absent any 3rd party gold standard to compare against), this is sufficiently

accurate for long-term AM. This is an encouraging result that shows that using the S8 is a suitable device for this level of analysis and that the issues with the S8 accelerometer do not adversely affect the overall result. The one thing of note in the algorithmic comparison results is that the results from the S8 are offset slightly compared to the GA – as shown in the off-zero intercept and mean scores. This could be a result of the overestimation in the S8 data during high-intensity activity as shown in Fig. 7. This is not a significant issue however, as it does not detract from the ability of the S8 to show activity levels over time. It should be noted though, that whilst the S8 is able to serve as an activity monitor through the calculation of AS, if accurate accelerometer data is required a smartwatch is not an appropriate device.

In terms of the generalizability of these results to other devices, this is, perhaps obviously, limited. However, the chipset family (MT65xx) that the S8 is based on is used in a number of other smartphones and the authors would expect a similar level of performance from devices which use the same chipset family. Furthermore, it is the expectation of the authors that many devices of a similar nature, that is those that have accelerometers on them but that are not specifically designed as such, will behave in a similar fashion and that even if they are not very good accelerometers, that they will be "good enough" to serve as activity monitors.

It is important that new devices are tested and evaluated, especially where the accelerometer data is hidden behind different layers of hardware and software. The work presented in this paper has shown an approach for testing a device for general purpose activity monitoring. The authors would suggest that this approach could be applied by others in the future for the evaluation of other devices.

Finally, it should be noted that the work presented in this paper used healthy normal controls to put the devices through their test regimes and that these results may not be valid under all patterns of usage and that for some specific applications (monitoring of sports exercise for example when a lot of high-intensity activity is expected) an independent accuracy evaluation with the target user group may be required.

VI. Conclusions

This paper has presented a comprehensive analysis of the S8 as a data gathering platform for monitoring a user's activity compared to other well validated devices for the same task. The results have shown that whilst the accelerometer on the S8 is far from perfect, the algorithmic processing employed is not adversely affected by this and that the data returned from the S8 is sufficiently accurate, compared to the same process applied to data from a well validated device, that the S8 is a suitable platform. This is an important result as it opens the way to using the S8 to provide features not commonly found on other devices at the same price point, such as easy system integration and user interaction.

Future work will investigate the capabilities of newer devices as they become available as well as the validity of the monitoring approach presented in different population groups.

References

[1] A. G. Bonomi and K. R. Westerterp, "Advances in physical activity monitoring and lifestyle interventions in obesity: A review," *Int. J. Obesity*, vol. 36, pp. 167–77, Feb. 2012.

[2] B. Waschki *et al.*, "Physical activity monitoring in COPD: Compliance and associations with clinical characteristics in a multicenter study," *Respiratory Med.*, vol. 106, pp. 522–530, Apr. 2012.

[3] C. A. Janney, A. Fagiolini, H. A. Swartz, J. M. Jakicic, R. G. Holleman, and C. R. Richardson, "Are adults with bipolar disorder active? Objectively measured physical activity and sedentary behavior using accelerometry," *J. Affective Disorders*, vol. 152–154, pp. 498–504, Jan. 2014.

[4] J. J. Reilly, V. Penpraze, J. Hislop, G. Davies, S. Grant, and J. Y. Paton, "Objective measurement of physical activity and sedentary behaviour: Review with new data," *Arch. Dis. Childhood*, vol. 93, pp. 614–619, Jul. 1, 2008.

[5] L. M. Reiser and E. A. Schlenk, "Clinical use of physical activity measures," *J. Amer. Acad. Nurse Practitioners*, vol. 21, pp. 87–94, 2009.

[6] R. W. Taylor, S. M. Williams, V. L. Farmer, and B. J. Taylor, "Changes in physical activity over time in young children: A longitudinal study using accelerometers," *PLoS One*, vol. 8, p. e81567, 2013.

[7] J. T. Cavanaugh, T. D. Ellis, G. M. Earhart, M. P. Ford, K. B. Foreman, and L. E. Dibble, "Capturing ambulatory activity decline in Parkinson's disease," *J. Neurol. Phys. Therapy*, vol. 36, pp. 51–57, Jun. 2012.

[8] A. Pantelopoulos and N. G. Bourbakis, "A survey on wearable sensor-based systems for health monitoring and prognosis," *IEEE Trans. Syst. Man, Cybern. C, Appl. Rev.*, vol. 40, no. 1, pp. 1–12, Jan. 2010.

[9] S. C. Mukhopadhyay, "Wearable sensors for human activity monitoring: A review," *IEEE Sens. J.*, vol. 15, no. 3, pp. 1321–1330, Mar. 2015.

[10] D. John and P. Freedson, "ActiGraph and actical physical activity monitors: A peek under the hood," *Med. Sci. Sports Exercise*, vol. 44, pp. S86–S89, Jan. 2012.

[11] M. Schneller, M. Pedersen, N. Gupta, M. Aadahl, and A. Holtermann, "Validation of five minimally obstructive methods to estimate physical activity energy expenditure in young adults in semi-standardized settings," *Sensors*, vol. 15, pp. 6133–6151, 2015.

[12] Y. Huang, J. Xu, B. Yu, and P. B. Shull, "Validity of FitBit, Jawbone UP, Nike+ and other wearable devices for level and stair walking," *Gait Posture*, vol. 48, pp. 36–41, Jul. 2016.

[13] J. D. Amor and C. J. James, "Setting the scene: Mobile and wearable technology for managing healthcare and wellbeing," in *Proc. 2015 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2015, pp. 7752–7755.

[14] R. Rawassizadeh, B. A. Price, and M. Petre, "Wearables: Has the age of smartwatches finally arrived?" *Commun. ACM*, vol. 58, pp. 45–47, 2014.

[15] K. Van Laerhoven, M. Borazio, and J. H. Burdinski, "Wear is your mobile? Investigating phone carrying and use habits with a wearable device," *Front. ICT*, vol. 2, 2015.

[16] V. Ahanathapillai, J. D. Amor, and C. J. James, "Assistive technology to monitor activity, health and wellbeing in old age: The wrist wearable unit in the USEFIL project," *Technol. Disability*, vol. 27, pp. 17–29, 2015.

[17] V. Ahanathapillai, J. D. Amor, Z. Goodwin, and C. J. James, "Preliminary study on activity monitoring using an android smart-watch," *Healthcare Technol. Lett.*, vol. 2, pp. 34–39, 2015.

[18] A. H. K. Montoye, J. M. Pivarnik, L. M. Mudd, S. Biswas, and K. A. Pfeiffer, "Validation and comparison of accelerometers worn on the hip, thigh, and wrists for measuring physical activity and sedentary behavior," *AIMS Public Health*, vol. 3, pp. 298–312, 2016.

[19] T. G. Pavey, S. R. Gomersall, B. K. Clark, and W. J. Brown, "The validity of the GENEActiv wrist-worn accelerometer for measuring adult sedentary time in free living," *J. Sci. Med. Sport*, vol. 19, pp. 395–399, May 2016.

[20] A. Santos-Lozano, P. J. Marin, G. Torres-Luque, J. R. Ruiz, A. Lucia, and N. Garatachea, "Technical variability of the GT3X accelerometer," *Med. Eng. Phys.*, vol. 34, pp. 787–790, Jul. 2012.

[21] J. Kennedy, "Particle swarm optimization," in *Encyclopedia of Machine Learning*, C. Sammut and G. I. Webb, Eds. Boston, MA, USA: Springer, 2010, pp. 760–766.

[22] J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet*, vol. 1, pp. 307–310, Feb. 8, 1986.

[23] J. D. Amor, J. G. Hattersley, T. M. Barber, and C. J. James, "Characterization of wrist-wearable activity measurement using whole body calorimetry in semi-free living conditions," in *Proc. 2015 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2015, pp. 3735–3738.