An Evaluation of Wearable Activity Monitoring Devices

Fangfang Guo Yu Li Mohan S. Kankanhalli Michael S. Brown
School of Computing, National University of Singapore
{guofang, liyu1988, mohan, brown}@comp.nus.edu.sg

ABSTRACT
This paper examines an increasingly relevant topic in the multimedia community of wearable devices that record the physical activity of a user throughout a day. While activity and other accelerometry-based data has been shown effective in various multimedia applications -- from context-aware music retrieval to approximating carbon footprint -- the most promising role of these target application for healthcare and personal fitness. Recently, several low-cost devices have become available to consumers. In this paper, we perform an evaluation on the most popular devices available on the market (in particular *Fitbit* and *Nike+*) and report our findings in terms of accuracy, type of data provided, available APIs, and user experience. This information is useful for researchers considering incorporating these activity-based data streams into their research and for getting a better idea of the reliability and accuracy for use in life-logging and other multimedia applications.

Categories and Subject Descriptors
H.3.0 [**Information Storage and Retrieval**]: General;
K.8.2 [**Personal Computing**] Hardware;
B.8.0 [**Performance and Reliability**] General;

General Terms
Measurement, Performance, Experimentation, Standardization

Keywords
Fitbit; Nike+ Fuelband; Fitness Applications; Social Media; Activity Monitoring; Quantified Self; Life Logging

1. MOTIVATION AND INTRODUCTION
Close to 70 years ago, Vannevar Bush published his seminal futuristic essay "As We May Think" [2] that envisioned a day when individuals would have access to a device that could store their entire collection of books, records, and communications in a manner that could be easily retrieved and examined. Bush postulated that this device, termed *Memex* (a combination of memory and index) would not only allow an individual to access their own "lifelog", but also a collective information/knowledge pool which would thereby benefit all of mankind.

PDM'13, October 22, 2013, Barcelona, Spain.

Figure 1: Devices evaluated in our study from left to right: iPhone *Moves* app, *Fitbit*, *Nike+Sportsband* and *Fuelband*, and conventional pedometers.

Bush's vision has had great influence on the development of many aspects of the World Wide Web as well as how personal collections of videos and images are shared and organized [4,7]. Interestingly, however, Bush's prediction did not envisage archiving an individual's daily physical activity as part of one's personal lifelog or for use as public sociality knowledge. This may be due to the lack of importance associated to physical activity 70 years ago. However, in modern times, it is well accepted that physical activity is crucial for both mental and physical well-being [15], and small wearable devices are making this possible to incorporate activity streams into our personal and collective "memory index".

Personal activity data has already been shown to be effective in various multimedia applications. Examples include context-aware music retrieval [14], as well estimating one's carbon footprint [5]. However, the most promising role for this data is in applications targeting health and fitness. In particular, prior work has shown that wearable sensors can benefit individual patient health [1], individual personal fitness [16], and epidemiology studies to assess the large scale activity of populations [13].

Early work examined data collected from specialized or research-grade accelerometry-based devices [3,15]. Seeing the benefits of this technology, several companies have now produced wearable activity monitoring devices at price levels that are attractive to everyday consumers. While various trade magazines periodically review these devices via anecdotal feedback, there has yet to be a systematic evaluation of these devices to examine their accuracy or suitability in terms of research and application development. This paper aims to provide this information by evaluating several devices including the two dominant market products, *Fitbit* and *Nike Fuelband* (see Figure 1). The findings in this paper are useful for researchers interested in incorporating these devices into various multimedia and life-logging applications. We also outline related open problems in the multimedia systems area.

2. DEVICES AND METHODOLOGY

2.1 DEVICES

The following section provides a brief description of the devices evaluated in this work. Our main emphasis is placed on the *Fitbit* and *Nike+ Fuelband*, the two current dominant devices on the market. We also include the *Nike+ Sportsband*, an iPhone *Moves* app and conventional mechanical pedometers.

Fitbit *Fitbit* [6] has four wearable devices on the market, *Ultra*, *One*, *Zip*, and the *Flex* (recently released). With the exception of the *Flex*, all are made to be discreet and wearable either on trousers or shirts. The *Flex* is worn on the wrist. The devices record steps taken, distance travelled, and calories expended. The *Ultra* and *One* have an altimeter that allows the counting of the number of floors walked up. These devices communicate with a host computer using Bluetooth that sends their data directly to a user's account on the *Fitbit* website.

Nike+ Fuelband *Nike+ Fuelband* [10] (the plus is pronounced) is worn on the wrist and records calories, steps, distance, and Nike's own unit of activity terms "Nike Fuel". Accumulative amounts of each item can be retrieved through a small display on the device. The device connects via USB to a host machine which syncs the data to a user's account on the Nike+ website.

Nike+ Sportsband/Motion This device is worn on the wrist and works with the *Nike+ Motion* sensor which is worn on an individual's shoe [11]. This device only records distance, which can be uploaded via USB to Nike website. The Nike+ *Motion* sensor can also be used with a Nike watch and iPhone App.

iPhone *Moves* app The *Moves* iPhone app [9] tracks a user's daily fitness activity through the built in accelerometer and location information from iPhone. The app runs in the background and the user only needs to carry the phone. The app records activity patterns and tracks the distance and the steps. The application is free, however, requires the cost of an iPhone.

Mechanical Pedometer For the sake of completeness, we have included two conventional mechanical pedometers with digital displays, *Omron Steps* and *SM-2000*, both available on Amazon.

Table 1 provides a comparison matrix of the devices evaluated, including estimated costs in US dollars.

2.2 Omitted Devices

Jawbone Up It is worth mentioning the *Jawbone Up* [8], which is the other high-profile consumer level activity device launched in 2011. The device, however, was pulled off the market after its initial launch due to faulty batteries and leaks. The product has only been recently re-launched in Nov 2012, while we were performing the study. As a result, we have omitted this from our study. The *Jawbone Up* provides steps, distance, calories. Currently the *Jawbone up* can only be used with mobile device, drivers for laptop and PCs are not provided.

GPS Watches We have omitted GPS-based watches, such as those made by *Garmin* and *Timex*, because they still represent high-end devices with costs typically exceeding USD$200. Moreover, these devices are not intended to record daily activity, but are used while engaging in specific activities, e.g. running.

Table 1. Summary matrix of devices evaluated.

	Fitbit	***Nike+ FuelBand***	***Nike+ SportsBand***	**iPhone *Moves***	**Pedo-meters**
Steps	✓	✓		✓	✓
Distance	✓		✓	✓	
Calories	✓	✓			
Nike Fuel		✓			
Wireless	✓		✓	✓	
Web Archive	✓	✓	✓		
API	✓	✓[1]	✓		
Price (USD)	**$50-100**	**$150**	**$60**	**Free**	**$5-30**

2.3 METHODOLOGY

We purchased several of each device to be evaluated, with the exception of the iPhone App which could be downloaded. Our goal was to evaluate the accuracy of the devices for recording the number of steps and distance travelled, as well as consistency in the measurements. We enlisted the help of several participants, each of whom wore multiple devices at the same time. Participants walked 400 meters multiple times on a running track. The output of each device was recorded after each lap. Participants used a handheld mechanical clicker to assist in recording the true number of steps taken each 400 meters walked. Figure 2 shows a picture of one of our participants.

In addition, a *Fitbit* and *Fuelband* device was worn by one participant for several weeks. The correlation between the daily activities of each device is reported in Section 3.2.

3. RESULTS

3.1 Device Accuracy

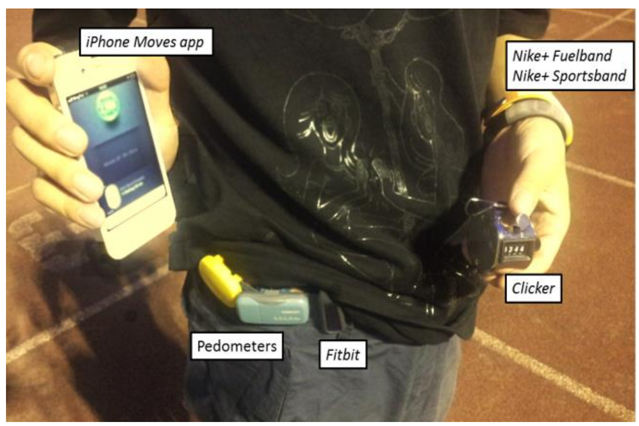

Figure 2. A participant in our study wearing multiple devices. A mechanical clicker was used to record number of steps.

Table 2 shows the results from our evaluations on the track. The table shows the results for each participant as well as the average across participants. The number ***N*** is how many effective readings were recorded, i.e. laps around the track (*e.g.* sometimes participants forgot to start the device or properly record the steps). Note that the number of effective reading may vary across different devices for the same participant. Reported is the mean and standard deviation of the recorded value as well as the error from the ground truth (either distance or steps). For steps, we report the ground truth as the average number of steps taken by each participants (denoted as GT mean). For *Nike+ fuel* we only report the mean and standard deviation since there is no ground truth for comparison.

[1] API for *Fuelband* is only open to Nike partner developers.

Table 2. Accuracy evaluation of tested devices. GT *mean* is the ground truth mean for each participant.

			Steps		Distance (actual dis = 400m)		*Nike Fuel (400m)*
	par (*N*)	***GT mean***	**mean±std**	**error±std**	**mean+std**	**error+std**	**mean+std**
Fitbit	P1 (11)	550.09	551.00±19.18	0.40±0.32%	416.36±16.29	4.09±4.07%	
	P2(15)	556.33	557.40±14.97	0.43±0.37%	410.00±10.69	3.17±1.76%	
	P3(15)	562.33	559.33±25.93	2.14±3.51%	389.33±24.04	4.00±5.16%	
	All		**556.39±20.43**	**1.05±2.26%**	**404.14±21.09**	**3.72±3.84%**	
Fuelband	P1(5)	539.20	503.80±7.53	6.49±3.22%			72.80±4.97
	P2(9)	562.00	540.22±9.86	4.19±1.72%			66.44±5.96
	P3(9)	562.33	492.89±79.01	12.38±13.47%			83.11±18.20
	All		**513.78±52.94**	**7.79±9.17%**			**74.38±13.98**
Sportsband	P1(3)	-			416.67±5.77	4.17±1.44%	
	P2(9)	-			400.00±39.69	8.33±4.51%	
	P3(3)	-			293.33±158.85	26.67±39.71%	
	All	**-**			**382.00±81.61**	**11.17±17.45%**	
Pedometer1 (Omron)	P1(5)	539.20	537.80±16.66	0.94±0.55%			
	P2(5)	564.60	437.00±133.21	23.24±22.4%			
	P3(9)	562.33	504.89±190.06	14.21±32.26%			
	All		**495.68±146.81**	**13.10±25.39%**			
Pedometer2 (SM-2000)	P1(5)	539.20	529.60±20.53	1.80±1.03%			
	P2(5)	564.60	521.40±89.30	8.35±14.82%			
	P3(5)	571.80	573.60±36.73	3.09±2.29%			
	All		**541.53±57.86**	**4.42±8.56%**			
iPhone Moves app	P1(9)	489.33	377.44±214.37	24.25±34.98%			
	P2(9)	554.00	418.67±184.15	30.32±25.77%			
	All		**398.06±195.02**	**27.28±29.97%**			

From our experiments, the *Fitbit* device clearly provided the most accurate results with the least variability. The pedometer (SM-2000, $5) was also very good, however, we note that the pedometers were sensitive to how they were worn and required them to be securely fastened to the beltline (placing them in a pocket gave poor readings which we omitted).

The experiments in Table 2 used the *Fitbit One*. An experiment was performed to check the difference between the *Fitbit One*, *Zip* and *Ultra*. A participant walked three loops with each device worn side by side. As shown in in Table 3, the readings in terms of step errors were all similar and less than 0.5 percent. An additional test was performed to see if the *Fitbit* was affected by where it worn, e.g. clipped on trousers or a shirt pocket. Four participants walked three loops each with the devices clipped on their shirts or shirt pockets as well as on their trousers. Table 4 shows that when worn on the trousers the data was slightly more accurate, but both places were less than 1% error.

Table 3. Comparison between different *Fitbit* products

	Fitbit One	*Fitbit Ultra*	*Fitbit Zip*
step err mean	0.32%	0.38%	0.32%

Table 4. *Fitbit* step errors worn on S=shirt, T=trousers

	P1	P2	P3	P4	Mean
S	0.31%	0.31%	1.83%	1.40%	0.96%
T	0.40%	0.43%	2.14%	0.21%	0.80%

3.2 FuelBand / Fitbit Steps Correlations

One participant wore a *Fitbit One* and *Nike+ Fuelband* for several consecutive weeks and recorded all their daily activity. This information was synced to their *Fitbit* and *Nike+* accounts. Figure 3 shows the two data streams as captured from the respective website. The images are overlaid on each other. The top image shows the entire month of February, 2013 (28 days) and shows daily *Fitbit* steps overlaid on daily *Nike+ fuel*. The bottom image shows an example of a daily stream. Nike shows activity as a smooth curve, and reports activity levels per-hour can be observed. *Fitbit* reports activity in 5-minute intervals. The difference in time granularity of these two devices is evident in Figure 3.

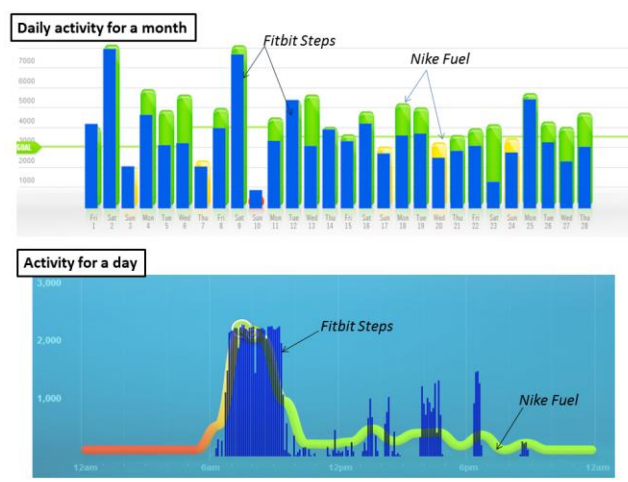

Figure 3: Comparison of *Nike+ fuel* and *Fitbit* step readings over a month (top) and over a day (bottom).

We also performed *normalize cross correlation* of the data gathered from the devices over the entire month. This is reported in Table 5. This table also shows the correlation between an individual devices own data (e.g. *Nike+ fuel* and *Nike* steps, *Fitbit* steps and *Fitbit* distance). We can see that the overall data is highly correlated. The correlation coefficient between *Fitbit's* steps and *Nike+ fuel* is 0.863, while the steps reported from the two devices have a correlation coefficient of 0.97 (i.e. very highly correlated). For the *Nike+ Fuelband*, the correlation coefficient between fuel and steps on the same device was 0.94. This implies that fuel also captures movement that is not directly related to steps. This is not too surprising as the *Nike+ Fuelband* is worn on the wrist and can capture hand movement.

Table 5. Data correlation over a month (F=Fitbit, N=Nike)

	F step	F dis	N step	N dis	N fuel
F step	1	0.9948	0.9707	0.9708	0.8630
F dis		1	0.9647	0.9648	0.8577
N step			1	1	0.9383
N dis				1	0.9381
N fuel					1

4. DISCUSSION AND CONCLUSION

From our experiments, the clear winner among the devices was the *Fitbit*, with a very low error of around 1% for step recording. While distance error was higher, it was highly correlated with steps. Other devices showed significantly more error, with *Nike+ Fuelband* at almost 8% error for steps and *Sportsband* with over 10% for distance - both with significant variations. *Nike+ Fuelband* only provided distance as a daily accumulation on their website and could not be accurately measured per 400m lap walked. These errors are a noteworthy find, especially for communities such as the Quantitative Self (QS) that look to these devices to provide accurate quantitative measurements.

Another positive aspect of the *Fitbit* device is its API. *Fitbit* has an API that allows its information to be extracted with a per-minute step readings (the information is obtained by connecting to the webpage and not the Fitbit itself). While a 3rd party API is available to get data from the *Nike+ Sportsband* webpage, the API for *Fuelband* is currently only open to select developers. Another nice benefit was that *Fitbit* was the Bluetooth connection that made it easier than the *Nike Fuelband* to collect data. We do note, that one downside of the *Fitbit* was that it was relatively easy to lose the device, compared to the Nike Fuelband. A Fitbit's was lost by one participant who did not notice that it had fallen off over the course of the day.

One disappointing finding was how poorly the iPhone app performed. We believe such monitoring may becoming better with time. While these apps do require the additional cost of the host device (as well as resources, such as battery), given the prevalence of smart phones, this is a very promising direction and could be more seamlessly integrated for multimedia applications. We still believe, however, that there will be a demand for smaller devices like *Fitbit* and *Nike+Fuelband*, specifically because they are lightweight which is suitable for wearing thorough the day and can last several days without the need for recharging.

An area we are keen to explore in the future is fusing the different data, e.g. *Nike Fuel* and *Fitbit steps*. While we found the data to already be highly correlated, the complementary nature of walking versus arm movements intuitively should be able to provide better activity monitoring when used together.

Finally, we conclude by noting that activity monitoring is a rapidly increasing market and there are many products coming on the market, e.g. *Fitbit* is taking pre-orders for a wrist form-factor device similar to *Nike+ Fuelband*. While our evaluation shows there is still room for improvement in accuracy and API availability, given the fierce competition in the market, we are hopeful that improvements will be forthcoming. This is also a very promising area for multimedia systems research. Some open problems in this area are:

1. How to fuse complementary and correlated information from multiple activity data streams to obtain better accuracy (e.g. upperbody activity and steps)?
2. How to incorporate the use of cameras and microphones with these devices?
3. How to interpolate for missing activity data using ambient audio-visual sensors?
4. Applications for multimodal healthcare data analytics, and motivation for staying active.

Acknowledgements
This research was carried out at the SeSaMe Centre supported by the Singapore NRF under its IRC@SG Funding Initiative and administered by the IDMPO.

5. REFERENCES

[1] S. Barnason, L. Zimmerman, J. Nieveen, P. Schulz, C. Miller, M. Hertzog, C. Tu. Influence of a symptom management telehealth intervention on older adults' early recovery outcomes after coronary artery bypass surgery. *Heart Lung*, 38, 364-376, 2009

[2] V. Bush. As we may think. *The Atlantic Monthly*, 1945

[3] K. Y. Chen, D. R. Bassett, Jr. The technology of accelerometry-based activity monitors: current and future. *Med. Sci. Sports. Exerc*, 37, 490-500, 2005

[4] P. Dikes. A difference maker. MIT-News (http://web.mit.edu/newsoffice/2011/timeline-bush-0216.html), 2011 (accessed Mar 1, 2013)

[5] A. R. Doherty, Z. Qiu, C. Foley, H. Lee, C. Gurrin, A F. Smeaton. Green Multimedia: Informing people of their carbon footprint through two simple sensors. *ACM Multimedia*, 2010

[6] Fitbit. http://www.Fitbit.com (accessed Mar 6, 2013)

[7] J. Gemmell, G. Bell, R. Lueder, S. Drucker and C. Wong. MyLifeBits: Fulfilling the Memex vision. *ACM Multimedia*, 2002

[8] JawBone Up. https://jawbone.com/up (accessed Mar 6, 2013)

[9] Moves. http://www.moves-app.com/ (accessed Mar 8, 2013)

[10] Nike+ Fuelband. http://www.nike.com/us/en_us/c/nikeplus-fuelband (accessed Mar 6, 2013)

[11] Nike+ Sportsband. http://nikeplus.nike.com/plus/products/sport_band/ (accessed Mar 6, 2013)

[12] J. H. Park, H. J. Kim, S.J. Kang. Validation of the AMP331 monitor for assessing energy expenditure of free-living - physical activity, *Res. Quart. Exerc. Sport*, 77, 2006

[13] D. Martinez-Gómez, G. J. Welk, M. E. Calle, A. Marcos, O. L. Veiga. Preliminary evidence of physical activity levels measured by accelerometer in Spanish adolescents; the AFINOS Study. *Nutr. Hosp*., 24, 226-232, 2009.

[14] X. Wang, D. Rosenblum, and Y. Wang. Context-aware mobile music recommendation for daily activities. *ACM Multimedia*, 2012.

[15] D. E.R. Warburton, C. W. Nicol, S.D. Bredin. Health benefits of physical activity: the evidence. *CMAJ*, 174, 801-809, 2006

[16] C.-C. Yang and Y.-L. Hsu. A Review of Accelerometry-Based Wearable Motion Detectors for Physical Activity Monitoring. *Sensors*, 10(8), 7772-7788, 2010