Using an in-ear wearable to annotate activity data across multiple inertial sensors

Alexander Hölzemann*
University of Siegen,
Ubiquitous Computing
alexander.hoelzemann@uni-siegen.de

Henry Odoemelem
University of Siegen,
Ubiquitous Computing
henry.odoemelem@student.uni-siegen.de

Kristof Van Laerhoven
University of Siegen,
Ubiquitous Computing
kvl@eti.uni-siegen.de

ABSTRACT

Wearable activity recognition research needs benchmark data, which rely heavily on synchronizing and annotating the inertial sensor data, in order to validate the activity classifiers. Such validation studies become challenging when recording outside the lab, over longer stretches of time. This paper presents a method that uses an inconspicuous, ear-worn device that allows the wearer to annotate his or her activities as the recording takes place. Since the ear-worn device has integrated inertial sensors, we use cross-correlation over all wearable inertial signals to propagate the annotations over all sensor streams. In a feasibility study with 7 participants performing 6 different physical activities, we show that our algorithm is able to synchronize signals between sensors worn on the body using cross-correlation, typically within a second. A comfort rating scale study has shown that attachment is critical. Button presses can thus define markers in synchronized activity data, resulting in a fast, comfortable, and reliable annotation method.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing**; • **Computing methodologies** → *Machine learning*.

ACM Reference Format:
Alexander Hölzemann, Henry Odoemelem, and Kristof Van Laerhoven. 2019. Using an in-ear wearable to annotate activity data across multiple inertial sensors. In *1st International Workshop on Earable Computing (EarComp'19), September 9, 2019, London, United Kingdom.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3345615.3361136

*

EarComp'19, September 9, 2019, London, United Kingdom

ACM ISBN 978-1-4503-6902-2/19/09…$15.00
https://doi.org/10.1145/3345615.3361136

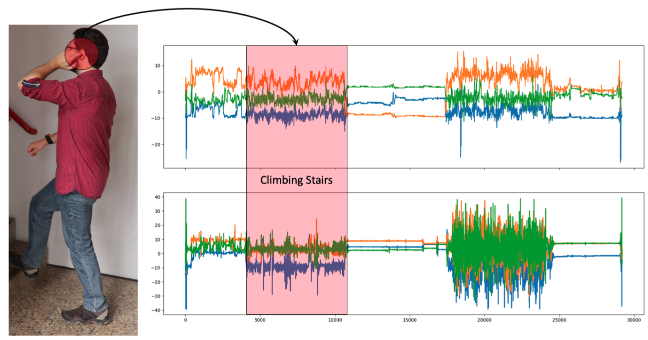

Figure 1: We present an inconspicuous annotation method, in which users can annotate their activity data *in situ* with an in-ear wearable (left), to mark and synchronize inertial data from the ear with all other inertial sensors (right).

1 INTRODUCTION

As wearable sensors have been shrinking and getting less power-hungry, their operation time and places where they can be worn have inadvertently increased accordingly. Nowadays, multiple such sensors can be worn as patches or miniature straps anywhere on the limbs, torso, or even on the head. When doing experiments with such sensor data, however, the annotation of the data has remained a burden, taking a substantial amount of effort. Few methods exist that allow the sensor data to be annotated directly, even fewer methods allow these annotations to be made for any amount of wearable sensor data from the user's body. In this paper, we argue that an in-ear device that is equipped with inertial sensors and a button would be an excellent candidate for user annotation of activity data. It would allow the users to annotate their data without much effort in a socially comfortable way, which also enables 'in the wild' experiments as study volunteers annotate activities in their daily lives. A critical step in our method is the synchronization of sensor data between all wearable sensors: We assume that all sensors contain inertial sensors that show sufficient correlation during everyday activities. The synchronization of different sensor signals plays a decisive role in activity recognition. In most cases, a synchronization gesture is executed at the beginning and end of the measurements to synchronize the two or more time series. This method has the decisive disadvantage that it is time-consuming and error-prone. With this paper we would like to introduce another approach, that helps synchronizing an arbitrary number of sensor signals. These signals only have the basic preconditions that they must be recorded at the same time and that sufficient sedentary phases, greater than 1 minute, are included. The here presented algorithm works with a very few calculation steps, these are the calculation of the vector length, the standard deviation and a binary filter that is used to decide if the acceleration signal represents a sedentary or non-sedentary activity. The given results show an median time mismatch of 1.10 seconds and can be used to synchronize related, but independently captured, sensor signals with a shared time base. In order for the algorithm to work reliably with the raw data, they must first be prepared and preprocessed. Chapter 4 describes the algorithm in detail. Our presented algorithm is fast and easy to implement. This allows researchers to take up this idea and incorporate it into their projects [2].

2 RELATED WORK

When experiments in human activity recognition rely on multiple inputs from an arbitrary number of sensors, a significant hurdle is to synchronize all sensors' data streams in order to attach them to the same time base. Problems that are typical in such cases have been discussed in detail in previous publications, for example in [1] or [6]. Thus far, several works have been published that deal with the synchronization of two or more independently working sensors. Some of these papers aim at explicitly synchronizing the clocks or time stamps and consider distributed systems or wireless networks, e.g. [14], [16], [15]. A large amount of relevant work is also mentioned in patents, [17], [8], [7], [5], [9] or [19].

[19] describes a procedure that synchronizes the recordings of several media devices with each other [19]. However, the sample applications highlighted in this patent refer to audio signals, such as sound recordings of various musical instruments or vocals. The published system is a client-server application and works with manual set markers in the audio signals. Much of the related research is modality-specific. One approach is presented in [5], where the proposed synchronization technique works with markers that need to bet set in the data. The data then gets synchronized based on these markers. Hesch et al. [7] provide a method that uses a set of interrupt triggered markups. In this system, a processor working in parallel to the CPU is responsible for managing this. [18] developed an algorithm with which he was able to synchronize data coming from a network of seismic sensors. The approach calculates the most probable clock offset for the data. The probably most usable global time was determined from all available sensor clocks by calculating the probability distribution of the clock offset measurements. The most probable offset has been chosen as the offset to tie all clocks to.

Work that is situated in wearable activity recognition research encompasses [20], which presents an approach to achieve robust active learning and avoid the typical annotations errors that asks users to solve a relayed related task and and estimates confidence scores from crowd sourcing.

Kunze et al. [13] published 2006 a method that used features calculated for a sliding window to recognize the sensor position on the human body. The experiment is divided into 5 phases. Firstly the walking activity has been recognized frame by frame. To be sure that only the clean walking segments are used for location recognition, only these frames where more then 70% of the data has been recognized as walking were taken into account. The final total accuracy is 82% and shows that it is possible to create a certain context awareness. This context awareness would be useful in the following investigations for the algorithm presented here. In this way the results could be improved.

3 SYSTEM DESIGN

Hardware

The hardware used for collecting labeled data is the eSense-BLE [11] by the Pervasive Systems group at Nokia Bell Labs, Cambridge. It is built with a custom-designed 15 x 15 x 3 mm PCB and composed of a Qualcomm CSR8670, a dual-mode Bluetooth audio system-on-chip (SoC) with a microphone per earbud; a InvenSense MPU6500 six-axis inertial measurement unit (IMU) including a three-axis accelerometer, a three-axis gyroscope, and a two-state button; a circular LED; associated power regulation; and battery-charging circuitry. It is powered by an ultra-thin 40-mAh LiPo battery, but lacks internal storage or real-time clock. Each earbud weights 20g and is 18 x 20 x 20 mm. The left earbud is the one containing the IMU sensor accessible through the BLE and will be used in the remainder of this paper.

The Platypus prototype is a wrist-worn activity sensing platform [10] that is equipped with a number of sensors, including a full MPU9250 IMU, environmental sensors, and several processing units included in an Edison System-on-Chip module that runs an embedded Linux distribution as operating system. We have used this prototype as it can record the IMU data at a relatively high sampling speed of 300 Hz and present the recordings via a Secure Shell (SSH) over the built-in WiFi transceiver.

Mobile Application

To be able to get labeled data for cross correlation, we developed an Android App for data collection on Android Studio by adapting only the needed aspects of the Android library provided by the developers. The Android ScanFilter was used to restrict the scan result to our desired eSense device, using the LOW LATENCY scan mode. Notification was enabled by writing to the descriptor for the push button status and accelerometer data from the on-board IMU, which we set to 50Hz sampling rate. The accelerometer works with the default configuration of +/-4g(sensitivity of 8192 LSB/g). Using the Android onCharacteristicChanged, accelerometer data about three-axis is received and checked for correctness using the CheckSum, then stored in the internal storage of our mobile phone as a CSV file in unit of g and multiplied by 10 to increase amplitude. We also saved a time-stamp in microseconds that have elapsed since January 1, 1970 at 00:00:00 GMT. Additional, on every button push, the current data from the accelerometer is labeled with an ASCII character and stored, as well as displayed on the TextView. We had a challenge of receiving the same data on different time stamps, but this was resolved by keeping the processing time in the onCharacteristicChanged method as low as possible, another problem that we encountered, was that each button push notification caused some accelerometer package index to be skipped on subsequent readings, restarting the IMU sampling on each button push solved this problem. Finding and establishing connection with the eSense (BLE and classic Bluetooth) is a challenge and requires several trials.

4 METHODOLOGY

Beside of our study about the reliability of our proposed algorithm we also asked the participants to fill out a short questionnaire regarding the wearing comfort of the eSense earbud by using the Comfort Rating Scale (CRS) as proposed in [12].

Data Set

A data set of activity data of seven participants has been recorded using the eSense and the platypus. The data recorded by the eSense is sampled with 50 Hz, the data recorded with the platypus is sampled with 300 Hz. The participants are between 20 and 40 years old. We were able to recruit five men and two women for the study. The platypus data set consists of a total amount of round about 2.255.764 samples or 2.08 hours of data. For the eSense we recorded 375.967 samples, which also results in 2.08 hours. The data set contains acceleration data from both sensors for mixed activities: (1) read or desk work, (2) walk, (3) climbing stairs, (4) sit, (5) dribble a basketball and (6) pause or rest phase.

Title	Description
Emotion	I am worried about how I look when I wear this device. I feel tense or on edge because I am wearing the device.
Attachment	I can feel the device on my body. I can feel the device moving.
Harm	The device is causing me some harm. The device is painful to wear.
Perceived change	Wearing the device makes me feel physically different. I feel strange wearing the device.
Movement	The device affects the way I move. The device inhibits or restricts my movement.
Anxiety	I do not feel secure wearing the device.

Table 1: Comfort Rating Scale (CRS) categories as proposed in [12]. The CRS includes 6 categories: Emotion, Attachment, Harm, Perceived change, Movement and Anxiety.

eSense Wearing Comfort

In addition to the evaluation of the reliability of our presented algorithm, it was also very important for us to evaluate how comfortable the provided prototype was perceived by the participants of the study. Table 1 shows the categories and their description. The Emotion, Harm, and Anxiety categories are more about personal and psychological sensations when wearing the device, while the remaining three categories focus on the device's body feel. The participants are able to choose a value between 0 and 10 for every category. 0 means it has low impact, and 10 a high impact.

Data Preparation

For evaluation purposes a ground truth is needed. Therefore we used a synchronization gesture. In our case the synchronization gesture has the requirement that both hand a head needs to follow the same movement. Therefore a simple vertical jump was chosen. This creates a clearly identifiable peak in the acceleration data. The gesture was done at the beginning and at the end of the recording and thus marks both in the data.

To be able to synchronize two independent recorded signals we need to preprocess the data. The first step in the data preparation process was to crop out the data at these marks. Due to sample losses, for example through the the wireless connection (BT or WiFi) or not fully achieved sampling rates, the signals are initially of different length. Furthermore they need to be sampled with the same sampling rates. Therefore the Platypus data must be sampled down to 50 Hz. Since we are only able to set a label for a certain sequence using the eSense, we have to set the timestamps of the eSense as ground truth for all other body sensors involved in the synchronization process. Under these circumstances we are now able to calculate all sensor signals equidistantly. Due to simulate the case that no synchronization gesture is available to synchronize the data, the time series of the eSense was shortened by 10% at the beginning and at the end.

Data Synchronization Method

Since the method for synchronizing signals is of central importance, it takes up most of the work presented here. The following table describes step wise the developed algorithm and the results after every step. When the algorithm finishes we are able to propagate the label throughout the sensors. Parameters like the window-size and window-length, but also the threshold of the binary filter, can be adjusted variably. In the first version of the algorithm a simple ASCII character is written with a button press at the beginning and end of the activity. In the future we plan to use the microphone and voice-to-speech recognition for setting the label.

Step	Name	Description
1	Dimension reduction of raw data	Calculate vector length per sample (dimension reduction from 3D to 1D).
		Result: Signals with reduced dimension.
2	Windowing	Data is divided into windows. Window length and overlap ratio can be set variable.
		Result: Windowed data.
3	Feature calculation	Calculation of the standard deviation.
		Result: Standard deviation per window.
4	Binary Filter	Both standard deviation signals are passed through a binary filter. A threshold is used to decide whether it's a sedentary or a non-sedentary activity. 0 (sedentary) if the current value is smaller than threshold and 1 (non-sedentary) if higher than the threshold.
		Result: Two signals with the values 0 and 1. 0 for sedentary sequences and 1 or non-sedentary activities.
5	Cross-Correlation	Cross-correlation [3] of both binary filtered signals. The eSense signal is cross correlated with the other signals.
		Result: Cross-Correlation coefficiants.
6	Index Selection	The window with the highest correlation coefficient marks the best index synchronize the signal.
		Result: Start window for synchronizing
7	In-Window-Cross-Correlation	Cross correlation for all samples in this window.
		Result: Exact index for synchronization.
8	Label propagation	Labels from the eSense signal can be copied to the other sensor data.
		Result: Labeled data.

Table 2: Step by step explanation of the algorithm. The algorithm is divided into 8 steps. First the dimension of the data is reduced by calculating the vector length, divided into windows and finally the standard deviation is calculated. The standard deviations are now passed through a binary filter, which writes a 0 for sedentary activity and a 1 for non-sedentary activity. Both signals are then cross-correlated. The position of the highest correlation can then be used to deduce the synchronization point in the initial signal.

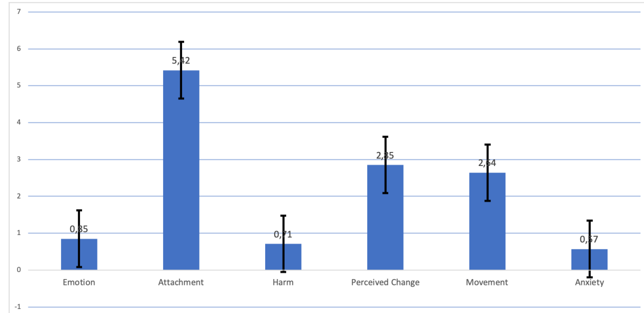

Figure 2: CRS result means and standard deviation. Anxiety: 0.57, 0.6; Movement: 2.64, 3.09; Perceived Change: 2.85, 3.17; Harm: 0.71, 0.8; Attachment: 5.42, 2.38; Emotion: 0.85, 1.31.

5 RESULTS AND DISCUSSION

Comfort Rating Scale

Figure 2 shows the result of our CRS study. To sum the result up we can say that in general the device is comfortable to wear, but sometimes you can feel it moving in your ear. One participant in the study noted that the earplugs tend to fall out of the ear during heavy movement, like dribbling a basketball, even if adjusted correctly. The average values for the Emotion, Harm and Anxiety categories show that users are generally not concerned about their appearance. This is certainly due to the fact that earbuds are very inconspicuous and devices like these have long since found their way into our everyday lives. The results in the other categories vary. This shows the standard deviation. The perceived wearing comfort is strongly user-dependent and is probably also related to the individual shape of the inner ear. This is as unique as the fingerprint [4], which is why it is difficult to develop a shape that everyone feels comfortable with.

Data Synchronization Method

In order to investigate the reliability in terms of automatically synchronizing the inertial data streams, we decided to use the time and sample mismatch between the ground truth and the index used as the synchronization point. To evaluate the performance of our algorithm we first calculated the best working parameters for window size and overlap ratio by using a brute force method. This was possible because of the short computational time and, compared to long-term benchmark data, limited amount of data. The determined parameters from these experiments were found to be:

- **window-size:** 50 samples
- **overlap-ratio:** 85%

With these parameters fixed, we calculated the time mismatch separately for every inertial data recording, as depicted in Table 3. The graphical representation as given out by our algorithm are shown in the Figures 3, 5 and 4.

These figures present two different signals: The top one is the one recorded at the wrist by the Platypus prototype. The bottom plot contains the inertial data from the eSense. The ground truth, as obtained from synchronization gestures before and after the recording (not shown), as a reference is plotted transparently. Overlaying the ground truth the resulting shortened and synchronized signal is depicted, with the vertical red lines marking the beginning and ending of the calculated synchronization point. The black line-plot embedded in each bottom plots shows the correlation signal. The inertial signals are synchronized according to the highest cross-correlation.

Record	Mismatch in Samples	Mismatch in Seconds	Activity
1	15	0.30	1, 3, 5, 6
2	16	0.32	1, 6
3	16	0.32	1, 4, 6
4	20	0.40	1, 6
5	21	0.42	1, 2, 5, 6
6	21	0.42	1, 4, 6
7	23	0.46	1, 3, 5, 6
8	87	1.47	1, 3, 5, 6
9	293	5.86	1, 3, 5, 6
10	386	7,72	1, 4
11	418	8.36	5
12	874	17.48	1, 4
13	1195	23.90	1, 2, 3, 5
14	1742	34.84	1, 4

Table 3: Synchronization error per record in samples and seconds. The records are recordings from 7 participants and overall 6 different activities. Synchronization tends to be within one second for records that contain clear sequences of activities that contain different intensities. (1) read or desk work, (2) walk, (3) climbing stairs, (4) sit, (5) dribble a basketball and (6) pause or rest phase.

In the first version of the algorithm the binary filter was not yet part of it, which resulted in problems synchronizing correctly, if no pause phases has been part of the record, for example record 11 in Table 3. The binary filter sets a very hard boundary between sedentary and non-sedentary activities, decided by a threshold, wherefore we needed to have a closer look on the calculated standard deviation signal. Here we saw that the threshold needs to be between 0.500 mg and 0.515 mg. After setting the boundaries we evaluated that the best working threshold is at 0.508 mg. The mismatch (median) of the algorithm was 61 samples or 1.22 seconds. Due to the usage of the binary filter we were able to improve our results to 55 samples or 1.10 seconds of mismatch. The records that could be rather poorly synchronized with our algorithm are records that mostly consists of sedentary activities as e.g. sitting, reading or desk work, as depicted in Figure 5 or records with heavy movements, but without pause phases, fig. 4. Very well devoted, data sets can be synchronized that

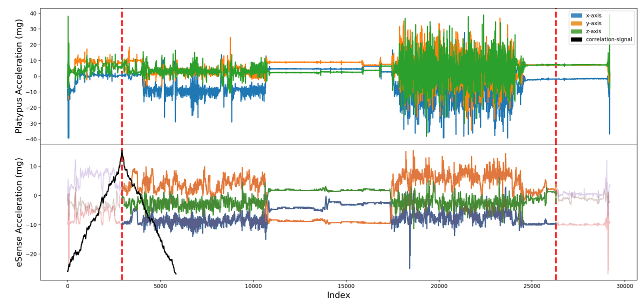

Figure 3: Best synchronization with a mismatch of 0.30 seconds. The figure shows that the synchronization works best with sufficient long periods of sedentary activity. The inertial signal of the wrist (top) compared with the synchronized signals of the head (bottom). The black signal at the bottom left depicts the cross-correlation between the binary filtered signals.

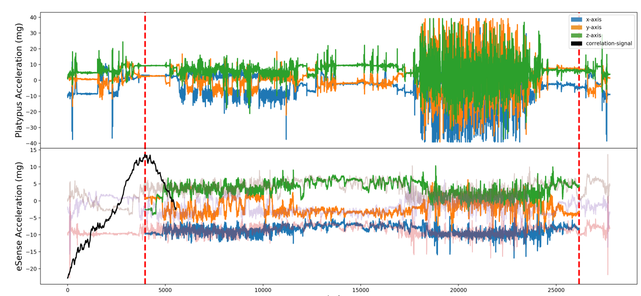

Figure 4: Data without sufficient pause phases, with plots defined as in Figure 3, using Record 13 in table 3. Our algorithm's synchronization was off by 1195 samples or 23.90 seconds.

reflect activities involving a high degree of locomotion as well as sufficient phase of pauses, e.g. figure 3.

- **Minimum time mismatch:** 0.30 seconds or 15 samples
- **Maximum time mismatch:** 34.84 seconds or 1742 samples
- **Median time mismatch:** 1.10 seconds or 55 samples

6 CONCLUSIONS

We presented in this paper a novel annotation method for recording activity recognition benchmark data. Our method relies on users wearing a small earbud-like device in their ear, which is equipped with a button and an inertial measurement unit. The inertial data from the ear-worn sensor are synchronized to all other data via cross-correlation, after which the user-presses serve as labels that annotate all sensor streams. In a preliminary study with 7 users, we investigated how

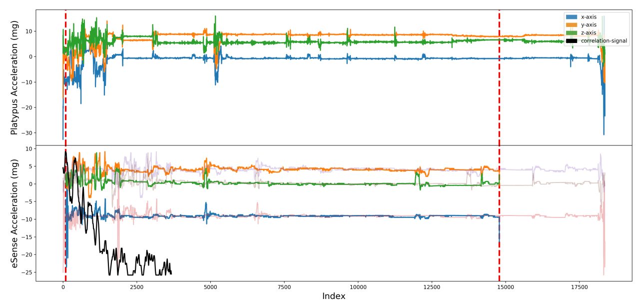

Figure 5: Worst-case synchronization with a mismatch of 34.84 seconds, record 14 in table 3. In this data almost only desk work has been performed. The inertial signal of the wrist (top) compared with the synchronized signals of the head (bottom). The black signal at the bottom left depicts the cross-correlation between the binary filtered signals.

well this synchronization works, as well as how comfortable the earbud-like wearable was to our study volunteers.

This paper offers a first approach to spread the annotations temporally correct over any number of sensors and to synchronize time series that have been recorded at the same time from different devices. If the data contains sequences that can be uniquely assigned to an activity, with sufficient periods of resting activity, the synchronization was found to be sufficiently reliable. However, the algorithm does not work reliably enough if the head and hand movements during an activity do not basically follow the same direction or if they can completely differ from each other. In addition, care must be taken to ensure that the movements follow a pattern that includes rest periods. The evaluation, as in Table 3, has shown that these are essential for reliable synchronization.

In terms of wearing comfort, we found that the used eSense prototype is highly promising as an annotation tool for everyday recordings 'in the wild'. The fact that it can be worn comfortably, with attachment as a weakest link for some participants, and almost hidden in the ear makes it ideal for recording *and* annotating data outside our laboratory. As such devices could be operated simultaneously as wireless headsets, the one remaining hurdle for use of our method in long-term and day-long activity recordings is the eSense's battery.

ACKNOWLEDGMENTS

We would like to thank all study participants, funding was supplied by the FFG's MinIAttention Project (BRIDGE Fruehphase, 3. Ausschreibung 2014, Projekt 851227).

REFERENCES

[1] Manfred Benesch, Helmuth Kubin, and Klaus Kabitzsch. 2017. Processing of Real-time Data in Big Manufacturing Systems.
[2] Ulf Blanke, Diane Larlus, Kristof Van Laerhoven, and Bernt Schiele. 2010. Standing on the Shoulders of Other Researchers A Position Statement. In *Pervasive Workshop*.
[3] The SciPy community. 2019. Numpy Documentation For Cross-Correlation. https://docs.scipy.org/doc/numpy/reference/generated/numpy.correlate.html. Last updated on July 26, 2019, Accessed: 2019-08-11.
[4] A. H. Cummings, M. S. Nixon, and J. N. Carter. 2010. A novel ray analogy for enrolment of ear biometrics. In *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. 1–6.
[5] Brian Derek DeBusschere. U.S. Patent 2016/0338599 A1, Nov. 24, 2016. Synchronizing Cardiovascular Sensors for Cardiovascular Monitoring.
[6] Daniel Roggen et al. 2010. Collecting complex activity datasets in highly rich networked sensor environments. *2010 Seventh International Conference on Networked Sensing Systems (INSS)* (2010), 233–240.
[7] Joel Hesch et al. U.S. Patent 2015/0185054A1, Jul. 2, 2015. Methods and Systems for Synchronizing Data Received From Multiple Sensors of a Device.
[8] Neal Crook et al. U.S. Patent 8,711,238 B2, Apr. 29, 2014. Systems and Methods for Synchronizing and Controlling Multiple Image Sensors.
[9] Radu Pitigoi-Aron et al. U.S. Patent 9,436,214 B2, Sep. 6, 2016. System And Methods of Reducing Energy Consumption by Synchronizing Sensors.
[10] Alexander Hölzemann and Kristof Van Laerhoven. 2018. Using Wrist-Worn Activity Recognition for Basketball Game Analysis. In *Proceedings of the 5th international Workshop on Sensor-based Activity Recognition and Interaction - iWOAR '18*. ACM Press.
[11] Fahim Kawsar, Chulhong Min, Akhil Mathur, and Allesandro Montanari. 2018. Earables for Personal-Scale Behavior Analytics. *IEEE Pervasive Computing* 17, 3 (jul 2018), 83–89.
[12] James F. Knight, Chris Baber, Anthony Schwirtz, and Huw William Bristow. 2002. The comfort assessment of wearable computers. *Proceedings. Sixth International Symposium on Wearable Computers*, (2002), 65–72.
[13] Kai Kunze, Paul Lukowicz, Holger Junker, and Gerhard Tröster. 2005. Where am I: Recognizing On-body Positions of Wearable Sensors. In *LoCA*.
[14] Christoph Lenzen, Philipp Sommer, and Roger Wattenhofer. 2015. PulseSync: An Efficient and Scalable Clock Synchronization Protocol. *IEEE/ACM Transactions on Networking* 23, 3 (Jun 2015), 717–727.
[15] Lin Lin, Chengfeng Yang, Maode Ma, Shiwei Ma, and Hao Yan. 2016. A clock synchronization method for molecular nanomachines in bionanosensor networks. *IEEE Sensors Journal* 16, 19 (2016), 7194–7203.
[16] Aneeq Mahmood, Reinhard Exel, Henning Trsek, and Thilo Sauter. 2016. Clock synchronization over IEEE 802.11 - A survey of methodologies and protocols. *IEEE Transactions on Industrial Informatics* 13, 2 (2016), 907–922.
[17] Tal Mizrahi. 2017. Clock synchronization using multiple network paths. US Patent 9,806,835.
[18] Christoph Sens-Schönfelder. 2008. Synchronizing seismic networks with ambient noise. *Geophysical Journal International* 174, 3 (Sep 2008), 966–970.
[19] Joshua Abraham Tabak. U.S. Patent 9,646,650 B2, May 9, 2017. Automatically Syncing Recordings Between Two Or More Content Recording Devies.
[20] Liyue Zhao, Gita Reese Sukthankar, and Rahul Sukthankar. 2011. Robust Active Learning Using Crowdsourced Annotations for Activity Recognition. In *Human Computation*.