Sensing Fine-Grained Hand Activity with Smartwatches

Gierad Laput   **Chris Harrison**
Carnegie Mellon University, Human-Computer Interaction Institute
5000 Forbes Ave, Pittsburgh, PA 15213
{gierad.laput, chris.harrison}@cs.cmu.edu

Figure 1. In this work, we investigate the feasibility of sensing 25 hand activities using commodity smartwatches, which are uniquely positioned to capture such fine-grained activity. Activity names are provided in Figure 3. The 25th hand activity we evaluated, brushing teeth (Y), is not shown here.

ABSTRACT

Capturing fine-grained hand activity could make computational experiences more powerful and contextually aware. Indeed, philosopher Immanuel Kant argued, "the hand is the visible part of the brain." However, most prior work has focused on detecting whole-body activities, such as walking, running and bicycling. In this work, we explore the feasibility of sensing hand activities from commodity smartwatches, which are the most practical vehicle for achieving this vision. Our investigations started with a 50 participant, in-the-wild study, which captured hand activity labels over nearly 1000 worn hours. We then studied this data to scope our research goals and inform our technical approach. We conclude with a second, in-lab study that evaluates our classification stack, demonstrating 95.2% accuracy across 25 hand activities. Our work highlights an underutilized, yet highly complementary contextual channel that could unlock a wide range of promising applications.

CHI 2019, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5970-2/19/05...$15.00
https://doi.org/10.1145/3290605.3300568

Author Keywords

Bio-acoustics; context-sensing; activity recognition;

CCS Concepts

Human-centered computing~ Ubiquitous and mobile computing systems and tools.

ACM Reference format:
Gierad Laput and Chris Harrison. 2019. Sensing Fine-Grained Hand Activity with Smartwatches. In *2019 CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland, UK.* ACM, New York, NY, USA. 10 pages. https://doi.org/10.1145/3290605.3300568

1 INTRODUCTION

Human activity sensing has been an area of active research for several decades (see *e.g.,* [6][9][14][23][71]). The advent of robust mobile sensing platforms (like the Intel Mobile Sensing Platform (MSP) [12]) and the ubiquity of smartphones has served to further accelerate research in this domain. In just the past few years, wearables have emerged, affording researchers a beachhead on the body, offering improved fidelity and new sensed dimensions. Today, many consumer smartphones and smartwatches include activity sensing capabilities that can distinguish between *e.g.,* walking, biking, driving and sleeping [23].

Most often, classified data is exposed to users for fitness and personal informatics applications, and thus is chiefly focused on locomotion (or the lack thereof). However, this ignores a rich and expansive landscape of fine-grained human actions, especially those undertaken by the hands. We call these *hand activities*, and they are often independent of body activity. For example, one can *type on their smartphone* (hand activity) while *walking (*body activity); or take a *sip* of water from a bottle while *jogging;* or *flip through a book* while *lying* in bed. Wilson [76] offers an eloquent portrayal of these diverse uses:

"After tugging at the covers and sheets and rolling yourself into a more comfortable position, you realized that you really did have to get out of bed. Next came the circus routine of noisy bathroom antics: the twisting of faucet handles, opening and closing of cabinet and shower doors, putting the toilet seat back where it belongs. There were slippery things to play with: soap, brushes, tubes, and little jars with caps and lids to twist or flip open; [...] Each morning begins with a ritual dash through our own private obstacle course – objects to be opened or closed, lifted or pushed, twisted or turned, pulled, twiddled, or tied. The hands move so ably over this terrain that we think nothing of the accomplishment. [...] Our lives are so full of commonplace experience in which the hands are so skillfully and silently involved that we rarely consider how dependent upon them we actually are."

If computing systems could know the activity of both the body *and* the hands, applications could be more context sensitive and assistive to immediate, ongoing tasks. State-of-the-art activity detection has been largely stuck at ambulatory states (walking, standing, sleeping, etc.) for decades. We envision smartwatches (slowly becoming more pervasive) as a unique beachhead on the body for capturing rich everyday actions. This could unlock many applications, ranging from personal informatics, health and skills assessment, and broadly, context-awareness. For example, a system that knows what your hands are doing can intelligently avoid interruptions. Hand activity detection can also be used to identify the onset of harmful patterns (*e.g.,* repetitive strain injury or hand-arm vibration syndrome), or for building healthy habits (*e.g.,* regular hand washing).

In this work, we show that hand activity can be sensed robustly from a commodity, off-the-shelf smartwatch, without any external infrastructure or instrumentation of objects, opening a new and practical means for achieving this vision. In addition to tracking coarse movement and orientation of the hand, the wrist is also the perfect vantage point to capture bio-acoustic information produced as a byproduct of most hand activities (*e.g.,* typing, brushing teeth). Here, we define bio-acoustics as body-coupled micro-vibrations propagating through the user's arm. These signals are inherently diverse, owing to user variance, innumerable tools and accessories, and differences in environment. To overcome this, we developed a flexible processing pipeline that demonstrates surprising robustness, underscoring the feasibility of our approach.

We also draw an important distinction between ***hand actions*** versus ***hand activities***. Specifically, we define a hand activity as a sustained series of related hand actions, typically lasting seconds or minutes. For example, a single clap would be a hand action, whereas a series of claps would be the activity of clapping. The decision to focus on hand activities was both *practical* (there is more data from a continuous signal to enable robust classification) and *functional* (instantaneous hand events are rarely indicative of a user's activity, and thus offer less opportunities for computational enhancement).

As we will discuss in detail, we started our investigations with a 50 participant, in-the-wild, experience sampling (ESM) study [17]. This yielded a trove of real-world data that informed our machine learning efforts. Secondly, it gave us a working set of how people employ their hands in the modern world, as there were no contemporary "hand ethnographies" to draw upon. We categorized participants' labels and selected 25 routine, yet interesting hand activities (Figures 1 and 4) to study for a second, in-lab feasibility evaluation. Employing an "obstacle course" methodology [6], we tested our full pipeline, which demonstrated 95.2% classification accuracy. We also ran a series of supplemental experiments to investigate specific questions, such as false positive rejection. Overall, we believe this work demonstrates practical sensing of fine-grained hand activities using just a commodity smartwatch, opening new possibilities for responsive and context-sensitive applications.

2 RELATED WORK

Hands and their activities are the subject of study in many fields; here we concentrate on prior work directly related to our immediate efforts, with a focus on the HCI literature.

2.1 Coarse Activity Recognition

Most activity recognition efforts have inferred user-state through worn sensors [6][9][16][43][58]. These systems are generally constrained to a limited set of coarse, whole-body activities, such as walking, running and bicycling. In general, systems must strike a balance between signal richness and instrumentation unobtrusiveness. Activity recognition systems have seen some success in the market, including Nike+ shoe sensors, Apple Watch, FitBit, and Garmin armbands. These products sense human activity via a combination of inertial measurement units (IMU), heart rate sensors and GPS. However, these products require explicit selection of pre-planned activities (*i.e.*, user selects a new run session from Nike+) to function reliably.

2.2 Fine-Grained Activity Recognition

One approach for fine-grained, human activity sensing is to deploy sensors and tags in the environment. Methods include acoustic monitoring [13][55], computer vision [70], electromagnetic sensing [60][80], and tagging objects of interest with markers [44][57] and sensors [46]. Infrastructure-mediated [15][26][27] and general-purpose sensing approaches [40][41] have attempted room- and building-scale activity recognition. Alternatively, activity sensing can be achieved through worn sensing systems (see [14] for a survey). Wearables with electromagnetic [[38]], magneto-inductive [71], inertial [49][50][51] and acoustic sensing [72] have all been used to recognize fine-grained activity sensing, including recognition of tool and appliance use. Also, worn cameras, in glasses [66] or on wrists [35][47][54], have been used extensively.

2.3 Hand Pose and Gesture Detection

Hand pose and motion sensing technologies can also be used to infer a user's context and activity (*e.g.*, typing, playing a musical instrument, grasping a cup). A wide variety of approaches have been demonstrated, including computer vision [20][35][54][74], electromyography [61], ultrasonics [73], bioacoustics [28], anatomical tomography [82], high-frequency radio [45][81] and motion sensing [1][7][56][75]. One of the major uses of such sensing is automatic sign language translation [63][79].

2.4 ViBand; Hand Actions vs. Activities

Most related to this research is our previous work on ViBand [39], which also overclocked the accelerometer in a smartwatch to 4kHz, allowing it to capture high-fidelity, bio-acoustic information. We directly build on top of this enabling capability, though our research focus is different: ViBand looked at explicit hand *gestures* for interactive control — waves, flicks, snaps and the like – which are different than *hand activities* performed by users to achieve a task – *e.g.,* typing, cutting, pouring, writing. Moreover, gestures tend to have exaggerated motions (as they are used for communication) and are well segmented — this is rarely true of hand activities, which can be subtle, discontinuous and of varying durations.

Enabling hand activity recognition required different innovations, and as a consequence, our machine learning pipeline is significantly more advanced than that used in ViBand. Further, ViBand's largest hand gesture set contained six actions, which achieved 96.0% accuracy in its user study. This stands in contrast to our user study's 25 hand activities demonstrated at 95.2% accuracy. Our experience sampling and obstacle course studies also move significantly beyond our initial work in ViBand.

3 PROOF-OF-CONCEPT SMARTWATCH

As previously noted, a wide array of methods have been considered for detecting hand actions and activities. In contrast to almost all prior work in hand sensing, we purposely selected a commodity platform for our explorations and studies. On one hand, this is constraining, limiting the worn locations and breadth of sensors we can bring to bear on this challenging problem. On the other hand, if successful, it offers an immediate and practical means to achieve our vision; manufacturers could deploy such sensing functionality with little more than a software update.

As a proof-of-concept platform, we use the smartwatch and high-speed sampling mode identified in our previous ViBand [39] research. This is a LG W100 smartwatch running Android Wear (Figure 2). By modifying the publicly available kernel [5], it is possible to configure the built-in MPU6515 IMU to stream three-axis accelerometer data at 4kHz [30]. This data stream captures coarse hand movement and orientation, as well as bio-acoustic data (up to the 2kHz Nyquist limit).

4 POWER CONSUMPTION

In worn systems, with small batteries, it is important to consider how changes in operation will affect battery life. The MPU6515 datasheet details power consumption rates. At 200Hz, power draw is 147μW, while at 4kHz sampling, power draw is 2719μW. While a ~18x difference is substantial, both are small values and it is important to consider it in context. The LG G watch contains a 1520mWh battery, which means the difference of 2572μW consumes <0.2% of battery life per hour.

Harder to estimate is total power load, which includes *e.g.,* waking the main application processor, moving data around in memory, and saving data to persistent storage. As a real-world test, we configured five LG G watches to *continuously* capture high-speed accelerometer. We gave

Figure 2. Our experience sampling watch app. At random intervals, wearers are prompted for activity labels (A). They select a hand activity (B), followed by a body activity (C).

these watches to five participants, who wore them all day, and charged them at night. Over the course of five days, we recorded battery statistics from when the watches powered on to when they ran out of power.

Across five days and five devices, the average battery life was 7.1 hours (SD=2.5). Given that our application kept the main application processor awake, we believe all day battery life is immediately achievable in a commercial implementation. It is now standard practice in the mobile industry to use low-power coprocessors (*i.e.*, "sensor hubs") for reading, buffering and processing continuous sensor data (for functions such as step counting, lift to unlock, and spoken keyword detection).

5 OPEN SOURCE DATA AND CODE

To facilitate future work in this area, we provide the source code for high-speed accelerometer acquisition on compatible Android devices. We also make available our study data and model code. *http://github.com/FIGLAB/hand-activities*

6 CONTEMPORARY HAND ACTIVITIES

Hands are central to the human experience, and as such, have been the focus of inquiry across many fields, including paleontology and anatomy [77], linguistics [48] and neuroscience [76], to name just a few. Ethnographic work has studied how hands are employed in everything from domestic life to industrial settings [3][37]. Many hand taxonomies have been proposed, most commonly organized by grip or communication primitives [18], which roughly correlate to functional or expressive uses respectively (see *e.g.*, [21] for a survey of taxonomies). Unfortunately, much of this seminal research was completed in a time before computing was common. Thus, as a starting point to our research, we wished to know two key questions:

1) *What activities do humans perform with their hands in the modern world?* Armed with such a list, we hoped to focus our technical efforts and better understand how recognition of these activities could be valuable in a computationally-enhanced setting.

2) *Do different hand activities generate characteristic signals?* In other words, are hand activities distinct and separable? Does a commodity sensor in a smartwatch provide sufficient fidelity to enable robust classification?

6.1 Experience Sampling Study

To explore these questions, we sought to collect hand activity data, in the wild, from a random cross-section of participants going about their daily routines. Although retrospective data collection methods (*e.g.,* surveys, interviews) are relatively easy to deploy, they are also subject to self-selection and recall bias [36], especially for something as unexceptional as hand activities. We also considered observational methods, but this was impractical for the scale of deployment we wished to achieve.

Instead, we employed an experience sampling method (ESM) [17], which reduces biases by collecting data in situ [10]. Using a fleet of ten smartwatches, we deployed a custom application to 50 participants over the course of two weeks. We used a participant pool drawn from the local population to cover a variety of ages, genders and professions (25 female, mean age of 26.3).

Our smartwatches ran a custom background application that we developed (Figure 2). After a random sleep interval between 7 and 15 minutes, our application surreptitiously captures ten seconds of accelerometer data. The app then activates the screen and vibration motor to catch the wearer's attention. A simple labeling interface is displayed. The initial screen offers three options: ignore the prompt, mark the activity as ill-defined (*e.g.,* indistinct, between activities), or proceed to label the hand activity. Selecting either of the first two options causes the application to return to sleep. If "*label activity*" is selected, the next screen asks: "*what were your hands doing?"* A pre-populated list of activities is provided, as well as the ability to add custom labels (using a companion smartphone application for ease of typing), which are added to the list for future use. If no user input was received on any screen for more than 30 seconds, the application returns to sleep.

Before deployment, participants completed a one-hour setup and orientation. The pre-populated hand activity labels were reviewed for understanding. Participants could also add additional labels as they desired. Participants also specified when they did not wish to be disturbed by the experiment (*e.g.,* 10pm – 8am). Following this orientation, participants wore the smartwatch for two days on their dominant arm (removed at night for recharging). Participants were paid $10 per day, plus $0.25 per label, up to a maximum of $15 on top of the base pay (*i.e.*, max $25 per day). The study concluded with a 30-minute open-ended interview. Participants often elaborated on hand activities they noticed but were never captured by the watch's random sampling interval.

6.2 Results

Cumulatively, our watches were deployed for 100 days (950 worn hours), during which time they captured 5830 instances. Of these, 765 instances (13.1%) were labeled as ill-defined. The remaining 5065 instances contained 120 unique labels. To regularize participant labels, a pair of human coders used a consensus merging scheme. For example, "hand

RANK	HAND ACTIVITY	CATEGORY	COUNT
1	Hands Still / Idle (a) • ‡	atomic	1797
2	Scrolling on Trackpad / Phone (b) • ‡	atomic	615
3	Typing on Keyboard (c) • ‡	atomic	480
4	Swaying (while locomoting) • ‡	atomic	346
5	Typing on Phone (e) ‡	atomic	281
6	Moving/Clicking Mouse (d) ‡	atomic	266
7	Eating ‡	compound	241
8	Gesturing (while speaking) ‡	compound	236
9	Carrying Object ‡	ambiguous	233
10	Writing (with implement) (i) • ‡	atomic	127
11	Drinking (s) • ‡	atomic	61
12	Cooking ‡	compound	53
13	Steering (while driving) ‡	atomic	38
14	Turning Pages	atomic	32
15	Smoking ‡	ambiguous	25
16	Washing Hands (x) •	atomic	23
17	Exercising (on elliptical) ‡	atomic	19
18	Brushing Teeth (y) • ‡	atomic	19
19	Stocking Items ‡	ambiguous	19
20	Using Hand Tools ‡	compound	9
21	Grasping Bicycle Exercise Machine ‡	atomic	8
22	Playing Piano (f) •	atomic	8
23	Operating Weight Machine ‡	ambiguous	8
24	Sign Language ‡	compound	7
25	Washing Dishes	compound	7
26	Putting on Clothes	compound	7
27	Showering	compound	5
28	Dancing	compound	5
29	Cleaning	compound	4
30	Putting Away Clothes	compound	4
31	Brushing Hair (g) •	atomic	4
32	Folding Napkins	ambiguous	4
33	Scratching (o) •	atomic	3
34	Doing Makeup	compound	3
35	Using Scissors (j) •	atomic	3
36	Pushing	ambiguous	2
37	Operating Microscope	compound	2
38	Petting (m) •	atomic	2
39	Drying Hair	ambiguous	2
40	Using Remote / Game Controller (l) •	atomic	2
41	Clapping (n) •	atomic	2
42	Folding Clothes	compound	2
43	Opening Door (p) •	ambiguous	2
44	Closing Door	ambiguous	2
45	Reaching for Object	ambiguous	2
46	Giving Massage	compound	2
47	Tying Shoes	atomic	2
48	Adjusting Watch	ambiguous	2
49	Kickboxing	compound	2
50	Operating Hand Drill (k) • ‡	atomic	2
51	Pilates	compound	2
52	Wiping (cleaning) (v) •	atomic	2
53	Selecting Clothes	compound	1
54	Exercising	compound	1
55	Shaving	ambiguous	1
56	Tying Hair	compound	1
57	Counting Cash	ambiguous	1
58	Holding Phone (on call)	atomic	1
59	Grating (food) (t) •	atomic	1
60	Chopping Vegetables (u) •	atomic	1
61	Using Spoon (eating)	atomic	1
62	Using Knife (eating)	atomic	1
63	Yoga	compound	1
64	Washing Utensils (w) •	atomic	1
65	Scrubbing Counter	atomic	1
66	Operating Vacuum	atomic	1
67	Putting on Lotion	ambiguous	1
68	Stretching	ambiguous	1
69	Searching Pocket	ambiguous	1
70	Screwing Bolt	atomic	1
71	Opening Bottle	ambiguous	1
72	Opening Jar (q) •	atomic	1
73	Operating Scanner	compound	1
74	Putting on Jacket	ambiguous	1
75	Grooming Beard	ambiguous	1
76	Shifting Gears (while driving)	atomic	1
77	Tapping Screen (e) •	atomic	1
78	Pouring Drink (r) •	atomic	1
79	Blowing Nose	ambiguous	1
80	Playing Tennis	compound	1
81	Sorting Paper	compound	1
82	Lifting Free Weights ‡	ambiguous	1
83	Putting on Chapstick / Lipstick	atomic	1
		Total Count	**5065**

Table 1. All hand activities captured and labeled during our experience sampling study (50 users, 950 worn hours). ‡ denotes activities that were pre-populated on our deployed watches (*i.e.,* not entered manually by users). Bulleted (•) items were incorporated into the subsequent obstacle course user study (letter key also used in Figures 1 and 3).

in pocket" and "hand on hips" were ultimately merged into a unified "hand still" label, which is the fundamental hand activity. This reduced the number of unique labels to 83, provided in Table 1.

The insights and implications from our experience sampling study were multifold. Foremost, it confirmed our assumption that human hands engage in an incredible diversity of activities. However, a few activities dominate: 35% of our labels are of the hands still or idle, and the next 4 most frequent labels are more common than the remaining 78 hand activities combined. We believe this bodes well for hand activity sensing, as detecting a small class of common actions could encompass most hand activities over the course of a day (an easier classification problem). However, it is also apparent there is an extraordinarily long tail of less frequent activities. Some of these may be rare, but others may be common and are simply short in duration, so as to be infrequently captured by our random sampling method.

We also found many labels that participants did not decompose into *atomic* hand activities. By atomic, we mean events that cannot be broken down into distinct stages. For example, eating and cooking were common labels, but these are *compound* activities that encompass a variety of atomic hand activities (*e.g.,* washing, chopping, mixing). Our coders also encountered labels that were *ambiguous.* For example, "open bottle" might mean twisting a cap off or using a bottle opener, which we view as distinct hand activities (though with a similar goal). These categorizations are included in Table 1.

Obviously, this result is just a small window onto the diverse landscape of hand activities, and much future work remains to be done in both HCI and beyond. Nonetheless, this result was sufficient to ground our assumptions and guide our subsequent technical efforts.

7 HAND ACTIVITY CLASSIFICATION

Informed by the findings from our experience sampling study, we proceeded to build a hand activity sensing pipeline for evaluation. This is comprised of three key stages: sensing, signal processing, and machine learning.

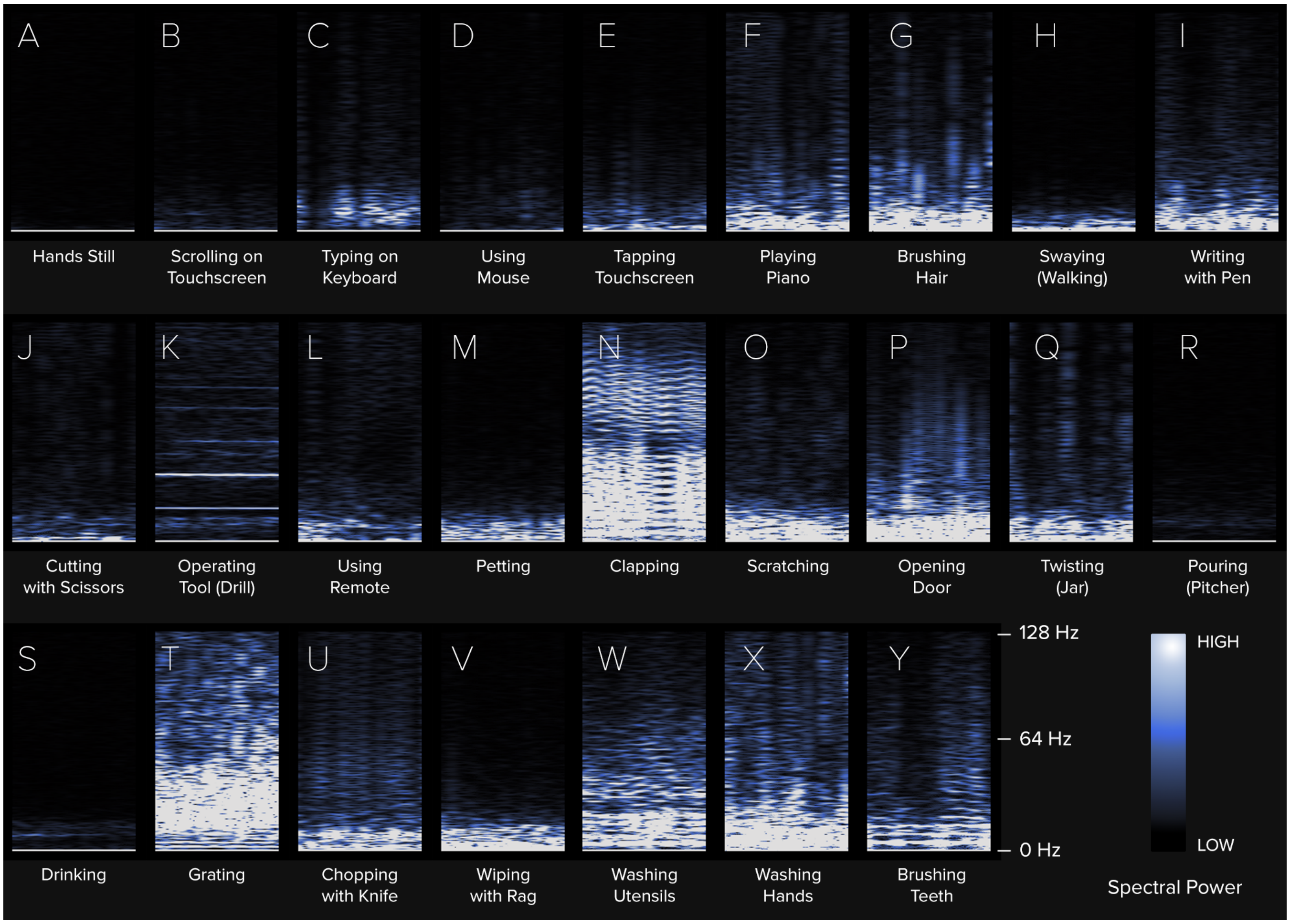

Figure 3. Example spectrograms of the 25 hand activities used in our obstacle course study (max of accelerometer axes shown). Y-axis is spectral power from 0 to 128 Hz. X-axis is time (3 seconds). Photos of these hand activities are shown in Figure 1, while Table 1 offers a rough estimation of how frequent these activities occur.

7.1 Sensing

As mentioned earlier, our software for the LG G watch captures both gross orientation and movement of the hands, as well as higher-fidelity, bio-acoustic information resulting from hand activities. A dedicated background process reads IMU data and fills three, 256-length circular buffers with accelerometer readings (X, Y and Z axes) at 4kHz. These buffers are sent to a laptop over Bluetooth at ~16 FPS, which maintains an even larger buffer and performs additional processing operations.

7.2 Signal Processing

A sampling rate of 4kHz in combination with a large buffer (8192 samples) allows our system to compute very high resolution Fourier transforms (4096 bins with a 0.5Hz resolution) within a short period – just over two seconds worth of data, which is about how fast hands transition to new activities. We utilize only the lower 256 FFT bins representing frequencies from 0-128Hz, which we found best characterized most human activities in our ESM study. Finally, these 256 bins are saved into a 48-frame rolling spectrogram, representing ~3 seconds of data (Figure 3). These spectrograms are maintained for each of the three accelerometer axes.

7.3 Machine Learning

The next stage of the pipeline is extracting and modeling patterns from the signal. From our experiments, we noticed that important spatial-temporal relationships are encoded in the accelerometer's three axes. For instance, when wiping a table, the Z-axis is mostly unperturbed (chiefly bio-acoustic noise resulting from friction, but little coarse motion), while the X and Y channels experience low frequency oscillations as the hand slides on the surface, often in a linear or circling motion. Indeed, we found many hand activities generated similarly distinctive activation patterns, which can be automatically learned with sufficient data.

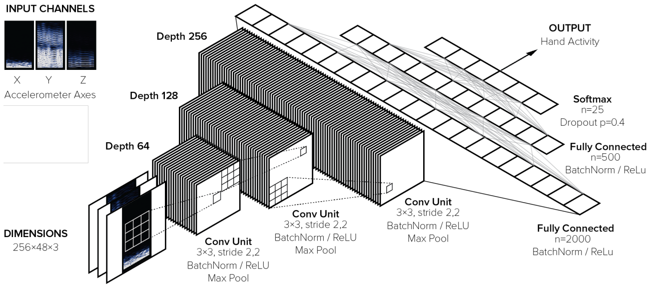

Figure 4. Our convolutional neural network (CNN) architecture, comprised of several convolutional units (three shown here), two fully connected layers, a dropout layer, and a softmax. We also apply batch normalization between nonlinear layers (*i.e.*, activations).

To learn from our data, we leverage a convolutional neural network (CNN) architecture [2]. Specifically, we use a variant of VGG16 [64] with modified fully connected layers (Figure 4, last two layer sizes set to 2000 and 500). CNNs have been widely used for visual datasets (*i.e.*, width × height × color channel), and in our case, we represent hand activities as spatio-temporal patches of bio-acoustic data. Specifically, we stack accelerometer spectrograms as 256 frequency bins × 48 frames × 3 orientation channels, which serves as input to the CNN. Because of the strongly coupled nature of our channels, this setup forces our architecture to learn cross-axis relationships.

Each convolutional unit in our VGG16 architecture is comprised of four sub-layers: 1) the convolutional operator, 2) a batch normalization layer [31], 3) a rectified linear unit (ReLu) activation layer [53], and 4) a pooling layer [62]. We also added a dropout layer [64] to the output of the second fully connected layer (p=0.4) to mitigate overfitting. An illustration of our network architecture is offered in Figure 4 (showing 3 of 5 convolutional units). The network was implemented using TensorFlow (tensorflow.org) and Keras (keras.io). To encourage replication, readers are welcome to download our dataset and model code (see Section 5).

8 EVALUATION

To quantify the feasibility and robustness of our hand activity classifier, we conducted a second user study. To properly validate our system, a reliable ground truth was needed. Because of the unsupervised nature of our earlier experience sampling study, it was not possible to use that dataset for an accuracy evaluation (though we use it to study false positives, described later). Instead, we employed an "obstacle course" methodology [6] – a technique that has been reliably used in past research to provide ground truth data collection, while emulating natural activities and settings. For this, we selected 25 atomic hand activities (Figures 1 and 3) from classes identified in our experience sampling study (Table 1, bulleted items). We dropped several frequent hand activities that were impractical to capture experimentally, including as eating, cooking, and steering a vehicle. We integrated our final hand activity set into a series of physical tasks that participants completed while wearing our smartwatches.

We recruited 12 people from a public participant pool (9 female, mean age 26.6), who were compensated $20 for the 90-minute study. Participants were asked to wear our smartwatch on their dominant arm. Once comfortable, the "obstacle course" began. Each "lap" of the course consisted of visiting four stations with physical activities that incorporated the 25 hand activities (random order). Participants performed each hand activity for at least 15 seconds, and they were free to perform them however they saw fit, capturing natural user variation.

In total, participants completed four laps of our course, with three-minute breaks in between. This ensured temporal separation between data collection rounds. Additionally, in between laps three and four, participants were asked to remove and then re-wear the smartwatch, again to capture variation and to mitigate overfitting (common in worn sensing systems). A trained observer labeled data using a laptop interface immediately after each hand activity was performed. This process yielded 2500 labeled instances per session, per user, resulting in a total of 120K instances.

9 RESULTS

9.1 Per-User Accuracy

To assess whether accelerometers provide sufficient information power to distinguish between dozens of hand actions, we trained a model using data from laps one and two, and tested it with data collected from lap three. Across all

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
A	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
B	0%	98%	0%	0%	0%	0%	0%	0%	0%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
C	0%	0%	99%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
D	0%	0%	6%	92%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	0%	0%	0%	0%	0%	0%	0%
E	0%	0%	0%	4%	95%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
F	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
G	0%	0%	0%	0%	0%	0%	97%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	0%
H	0%	0%	0%	0%	0%	0%	0%	99%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
I	0%	0%	0%	0%	0%	0%	0%	0%	93%	6%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
J	0%	1%	0%	0%	0%	0%	0%	0%	0%	90%	0%	4%	0%	0%	0%	0%	0%	4%	0%	0%	0%	0%	0%	0%	0%
K	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	99%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L	1%	8%	0%	0%	0%	0%	0%	0%	0%	12%	0%	77%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
M	0%	0%	3%	3%	2%	0%	0%	0%	0%	0%	0%	0%	91%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
N	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	99%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
O	0%	0%	0%	2%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	86%	0%	1%	0%	0%	0%	0%	0%	8%	0%	0%
P	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	97%	2%	0%	0%	0%	0%	0%	0%	0%	0%
Q	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	96%	0%	0%	0%	0%	0%	0%	3%	0%
R	0%	0%	0%	0%	3%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	92%	4%	0%	0%	0%	0%	0%	0%
S	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	99%	0%	0%	0%	0%	0%	0%
T	0%	0%	0%	0%	0%	0%	0%	0%	0%	5%	0%	0%	0%	0%	0%	0%	0%	0%	0%	94%	0%	0%	0%	0%	0%
U	0%	0%	0%	0%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	95%	0%	0%	0%	0%
V	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	98%	0%	0%	0%
W	0%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	96%	0%	0%
X	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	91%	4%
Y	1%	0%	0%	0%	0%	7%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	2%	0%	0%	0%	0%	0%	86%

Figure 5. Per-user-trained model confusion matrix. Mean accuracy is 95.2% across 25 activities and 12 users.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
A	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
B	6%	75%	0%	0%	0%	0%	0%	0%	0%	10%	0%	0%	0%	0%	0%	0%	0%	0%	8%	0%	0%	0%	0%	0%	0%
C	0%	0%	93%	0%	7%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
D	0%	0%	1%	84%	14%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
E	0%	0%	9%	1%	89%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
F	0%	0%	0%	0%	0%	90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	8%	0%	0%	0%	1%
G	0%	0%	0%	0%	4%	0%	94%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	1%	0%	0%	0%	0%
H	8%	0%	0%	0%	0%	0%	0%	91%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
I	0%	0%	0%	0%	0%	0%	2%	0%	78%	8%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	9%	0%	1%	0%	0%
J	0%	0%	0%	0%	0%	0%	0%	0%	0%	82%	0%	18%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
K	0%	0%	6%	0%	0%	0%	0%	0%	0%	0%	91%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L	1%	0%	0%	0%	0%	0%	0%	0%	1%	10%	0%	78%	0%	0%	0%	0%	0%	5%	4%	0%	0%	0%	0%	0%	0%
M	0%	0%	1%	2%	3%	0%	2%	0%	1%	0%	0%	0%	77%	0%	0%	0%	5%	0%	0%	0%	0%	6%	1%	0%	0%
N	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	98%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
O	0%	0%	0%	0%	2%	1%	0%	0%	7%	2%	0%	0%	11%	0%	75%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%
P	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	90%	0%	0%	0%	0%	0%	0%	0%	8%	0%
Q	0%	0%	0%	0%	0%	0%	0%	0%	7%	0%	0%	0%	0%	0%	0%	0%	88%	0%	0%	0%	3%	0%	0%	1%	0%
R	8%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	89%	0%	0%	0%	0%	0%	0%	0%
S	6%	3%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	89%	0%	0%	0%	0%	0%	0%
T	0%	0%	0%	0%	0%	0%	0%	0%	0%	8%	0%	0%	0%	1%	0%	0%	0%	0%	0%	89%	0%	0%	0%	0%	0%
U	0%	0%	0%	0%	0%	2%	0%	0%	0%	5%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	91%	0%	0%	0%	0%
V	0%	0%	0%	0%	0%	0%	1%	2%	0%	0%	0%	0%	0%	0%	2%	0%	1%	0%	0%	0%	0%	92%	0%	0%	0%
W	5%	0%	0%	0%	0%	0%	0%	0%	3%	0%	0%	0%	8%	0%	0%	0%	0%	0%	1%	0%	0%	1%	80%	1%	0%
X	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	4%	2%	0%	0%	0%	0%	0%	0%	92%	0%
Y	0%	0%	0%	0%	4%	0%	0%	0%	3%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	90%

Figure 6. Post-watch removal confusion matrix. Mean accuracy is 88.3% across 25 activities and 12 users.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
A	98%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
B	0%	92%	0%	0%	0%	0%	0%	0%	0%	2%	0%	1%	0%	0%	0%	0%	0%	2%	2%	0%	0%	0%	0%	0%	0%
C	0%	0%	90%	3%	1%	0%	0%	0%	0%	0%	0%	0%	5%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%
D	0%	0%	3%	83%	6%	0%	0%	0%	0%	0%	0%	0%	6%	0%	1%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%
E	0%	0%	1%	4%	88%	0%	0%	0%	0%	0%	0%	0%	5%	0%	1%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%
F	0%	0%	0%	0%	0%	96%	0%	0%	3%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%
G	0%	0%	0%	0%	0%	0%	92%	1%	0%	1%	0%	0%	0%	0%	0%	0%	2%	0%	0%	0%	0%	0%	0%	2%	0%
H	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
I	0%	0%	0%	0%	0%	0%	1%	0%	94%	1%	0%	1%	1%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%
J	0%	1%	0%	0%	0%	0%	0%	0%	0%	85%	0%	13%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%
K	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L	0%	2%	0%	0%	0%	0%	0%	0%	1%	14%	0%	80%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	1%
M	0%	1%	8%	5%	3%	0%	0%	0%	2%	0%	0%	0%	71%	0%	7%	1%	0%	0%	0%	0%	0%	0%	0%	0%	1%
N	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
O	0%	0%	0%	1%	1%	0%	0%	0%	0%	0%	0%	0%	6%	0%	82%	0%	3%	0%	0%	0%	0%	2%	4%	0%	0%
P	0%	0%	0%	0%	0%	0%	2%	0%	0%	0%	0%	0%	0%	0%	0%	92%	3%	0%	0%	0%	0%	0%	0%	2%	0%
Q	0%	0%	0%	0%	0%	0%	3%	0%	2%	0%	0%	0%	1%	0%	0%	2%	86%	0%	0%	1%	3%	0%	0%	1%	0%
R	1%	2%	0%	2%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	93%	0%	0%	0%	0%	0%	0%	0%
S	0%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	98%	0%	0%	0%	0%	0%	0%
T	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	94%	2%	0%	2%	0%	0%
U	0%	0%	1%	0%	0%	1%	0%	0%	2%	2%	0%	1%	0%	0%	0%	0%	4%	0%	0%	0%	89%	0%	0%	0%	0%
V	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	1%	0%	0%	0%	0%	0%	0%	96%	0%	0%	0%
W	0%	0%	0%	0%	0%	0%	4%	0%	1%	0%	0%	0%	2%	0%	1%	2%	3%	0%	0%	0%	0%	1%	84%	0%	1%
X	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	1%	1%	0%	0%	0%	1%	0%	3%	89%	4%
Y	0%	0%	1%	0%	1%	3%	0%	0%	0%	1%	0%	1%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	90%

Figure 7. Across-user performance confusion matrix. Mean accuracy is 90.7% across our 25 activities and 12 users.

participants and 25 hand activities, our system achieved a mean per-user accuracy of 95.2% (SD=4.1, max=98.8%, chance=4%). See Figure 5 for the confusion matrix.

9.2 Accuracy Post-Removal

Too often, worn sensing systems are trained (or calibrated) and then tested having never been removed from the user. This is artificial, as most wearables are removed daily. Owing to placement sensitivity for most worn sensors, it also tends to lead to artificially impressive results. This experimental effect can be allayed by explicitly including a post-removal collection round, which not only offers for a more realistic estimate of accuracy, but also lets one assess the accuracy drop-off.

Using the same model as before (*i.e.,* trained on laps one and two), we evaluated accuracy using data collected from lap 4 (*i.e.,* post removal). Overall, our system achieved an average accuracy of 88.3% (SD=16.5, max=98.9%, chance=4%). The confusion matrix is offered in Figure 6. The 6.9% drop in accuracy from pre-to-post watch removal was much less than we expected and suggests that our signals and approach are fairly robust to placement variation. We strongly suspect that if additional laps of data were collected following a similar watch removal/replacement procedure, accuracy would rebound.

9.3 All-Users Accuracy

To answer the central question of whether a commodity smartwatch accelerometer provides sufficient information power to distinguish between a variety of hand activities, we ran a lap-fold, cross validation study. For example, we trained a model on all user data from on laps 1, 2 and 3, and then tested on lap 4 (*i.e.,* 90,000 train, 30,000 test instances). We repeated this process for all lap combinations and averaged the results. Across all participants, this "all users" model achieves a mean accuracy of 90.7% (SD=2.2, chance=4%). Figure 7 shows the confusion matrix.

9.4 Leave-One-User-Out Accuracy (Across User)

Finally, we ran a leave-one-user-out analysis to investigate performance across users. Here, data from one participant (laps 1-4) served as a hold-out set, while data from all remaining participants are used for training. We repeat this process for all users and average the results. Across all participants, mean leave-one-out accuracy was 79.2% (SD=6.4, max=84.8%). This is a 10% drop compared to the previous result (90.7%), which simulated a general model seeded with some per-user calibration data (*i.e.,* 1/12th of corpus).

9.5 False Positive Rejection

In a worn input system – especially one that is hand-centric – it is vital to consider mechanisms for rejecting false positive events. For this, we take advantage of the per-class confidence scores output from our classifier's softmax layer. When participants performed a (known) hand activity, the top ranked class had an average confidence of 98.0%, while the second highest ranked class had a mean confidence of 2%. This significant drop-off suggested that confidence could be a good predictor of "unknown-ness". For example, our software could label events as unknown if the most confident class was below 50%. To identify a reasonable confidence threshold, we ran a simulation using our study data varying this threshold from 0 to 100%. The results, plotted in Figure 8, suggest rejecting events when the top-ranked class is under 90% confident, offering a balance between false positive rejection and missed detections.

In addition to the simulation above, we ran another experiment where we trained a model with "negative"

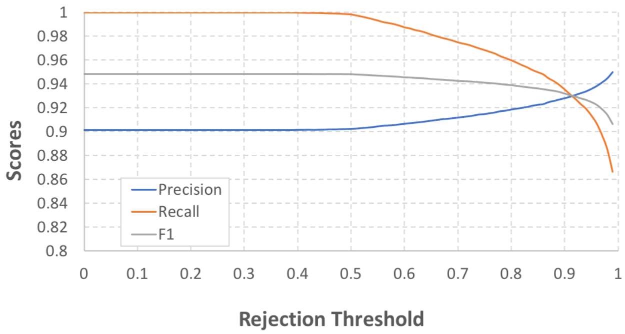

Figure 8. Precision vs. recall characterizations of our model. We can prevent false positive occurrences by setting a confidence value cutoff, but at the expense of missing events.

example data extracted from our experience sampling study. More specifically, we randomly selected 30K data instances that participants had labeled with hand activities not included in our test set of 25 (which included *e.g.,* driving, smoking and doing makeup). We labeled these diverse instances as an "unknown" hand activity class. Next, we performed a random 80/20 train-test split on the unknown class dataset, and then added this to our all-users model's train and test datasets. After retraining, our model correctly predicted (*i.e.,* rejected) 76.0% of the unknown activities, while the overall accuracy was 87.9%. If we include the confidence threshold identified in the previous simulation (confidence > 90%), the overall accuracy is 92.2% (with unknown detection of 86.3%). Figure 9 provides this confusion matrix; note the confusion along the Z column, which we use for the unknown class.

Finally, we also ran a clustering experiment (t-SNE; based on the top-3 PCA components of our input data; Figure 10) to visualize the discriminability of our signals, as it may be possible to employ clustering techniques to mitigate false positives, where events that are "distant" from known hand activity clusters are rejected. This distance-based method could also be used to capture negative example data, or prompt wearers for labels for future recognition.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	99%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%
B	0%	95%	0%	0%	0%	0%	0%	0%	0%	3%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%
C	0%	0%	93%	0%	1%	0%	0%	0%	0%	0%	0%	0%	5%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%
D	0%	0%	3%	84%	0%	2%	0%	0%	0%	0%	0%	0%	7%	0%	0%	0%	0%	2%	0%	0%	0%	0%	0%	0%	0%	1%
E	0%	0%	2%	5%	88%	0%	0%	0%	0%	0%	0%	0%	4%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%
F	0%	0%	0%	0%	0%	99%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
G	0%	0%	0%	0%	0%	0%	93%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	1%	0%	5%
H	0%	0%	0%	0%	0%	0%	0%	99%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%
I	0%	0%	0%	0%	0%	0%	0%	0%	96%	1%	0%	1%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	1%
J	0%	0%	0%	0%	0%	0%	0%	0%	0%	87%	0%	9%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%
K	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L	0%	1%	0%	0%	0%	0%	0%	0%	1%	13%	0%	84%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%
M	0%	0%	6%	4%	2%	0%	0%	0%	1%	0%	0%	0%	71%	0%	6%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	8%
N	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
O	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	4%	0%	88%	0%	1%	0%	0%	0%	0%	3%	3%	0%	0%	0%
P	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	93%	4%	0%	0%	0%	0%	0%	0%	2%	0%	0%
Q	0%	0%	0%	0%	0%	1%	3%	0%	1%	0%	0%	0%	1%	0%	0%	0%	88%	0%	0%	1%	2%	0%	0%	0%	0%	2%
R	1%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	97%	0%	0%	0%	0%	0%	0%	0%	1%
S	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	98%	0%	0%	0%	0%	0%	0%	0%
T	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	96%	0%	0%	2%	0%	0%	0%
U	0%	0%	1%	0%	0%	0%	0%	0%	2%	2%	0%	2%	0%	0%	0%	0%	3%	0%	0%	0%	90%	0%	0%	0%	0%	0%
V	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	1%	0%	0%	0%	0%	0%	0%	97%	0%	0%	0%	0%
W	0%	0%	0%	1%	0%	0%	2%	0%	0%	0%	0%	0%	1%	0%	0%	0%	1%	0%	0%	0%	0%	0%	88%	0%	1%	3%
X	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	2%	85%	2%	9%
Y	0%	0%	0%	0%	0%	4%	0%	0%	0%	1%	0%	1%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	93%	0%
Z	0%	1%	0%	0%	0%	0%	0%	2%	0%	0%	0%	0%	2%	0%	8%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	86%

Figure 9. Confusion matrix for a cross-users model with unknown class detection. Using confidence thresholds, global accuracy is 92.2%, while unknown rejection is 86.3%.

9.6 Sampling Frequency vs. Accuracy

We ran a final post hoc experiment to investigate the effect of accelerometer sampling frequency on classification accuracy. For this, we created downsampled versions of our original 4kHz data to simulate lower sampling rates: 2kHz, 1kHz, 500Hz, 250Hz, and 125Hz. As with our 4kHz data, this was featurized into three-axis, 0-128 Hz, three-second spectrograms. We performed cross-lap validation for each sample rate (train on one round, test on remaining rounds, four rounds total), and combined the results, shown in Figure 11. There is a clear, monotonic decrease in accuracy as sample rate decreases, with a marked cliff around 500Hz.

10 LIMITATIONS

The most immediate limitation of our technique is the need for smartwatches to be worn on the active arm. Most often, this will be a wearer's dominant arm, whereas it is more common for watches to be worn on the passive arm. However, detection still works for two-handed activities such as

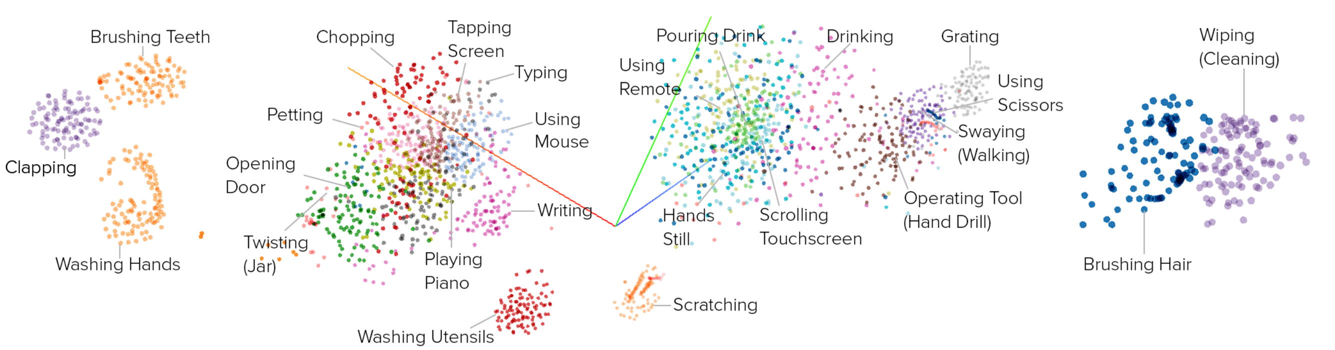

Figure 10. Clustering results from a t-SNE nonlinear dimensionality reduction (perplexity=40, 5000 iterations, projector.tensorflow.org) on a random subset of our 25 hand activities. Note how less intense activities (*e.g.,* pouring drink, scrolling touch screen) cluster together, while more vigorous hand activities (*e.g.,* clapping, scratching and wiping) emerge as distinct groups.

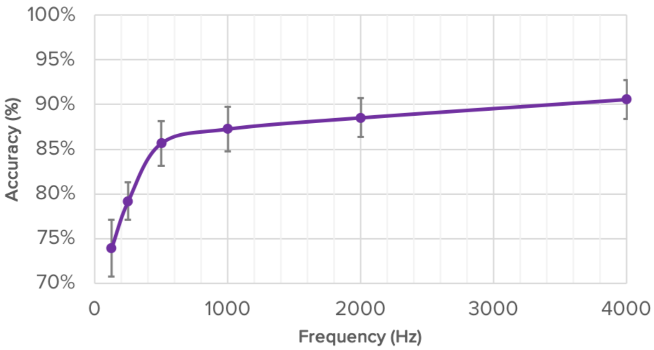

Figure 11. Post-hoc analysis of simulated sampling frequency vs. hand activity classification accuracy.

clapping, washing hands, or typing. Detecting events on the passive arm is an area we plan to explore in future work. We also note that we did not explore simultaneous hand activities, though these appear rare for a single hand.

We also acknowledge that the 25 hand activities we evaluated, though large for a recognition study, are a small fraction of the ways we engage our arms and hands in the real world. As reported earlier, there is an exceptionally long tail of hand actions and activities that will certainly prove challenging to distinguish. Thus we suggest that future work focus on specific activities that can especially benefit from computational support (*e.g.,* contextual aware assistance [14][38][39][44][58], smoking secession [11], elder care [19][59][68], hand-arm vibration syndrome (HAVS) [68][69], typing RSI [8]). Fortunately, as classifiers become more robust (perhaps through mass adoption of consumer smartwatches), over-the-air updates could unlock recognition of new classes incrementally.

Finally, as discussed earlier, our FFTs use long windows, both to mitigate noise and capture high-resolution spectral data for lower frequencies. As a consequence, we incur a latency penalty of a few seconds in recognizing events. For extended activities, like eating a meal, where classification might trigger devices to enter a "do not disturb" mode, a few seconds of latency is acceptable. However, for short-duration activities, like operating a TV remote control, users will want relevant information to be pulled up quickly (*e.g.*, to make a decision regarding entertainment). Future systems will likely want to use smaller windows with variable class confidence thresholds in order to support hand activities of longer and shorter durations.

11 EXAMPLE USE DOMAINS

Using commodity smartwatches for hand activity recognition is applicable to a wide range of application scenarios that have been well motivated in prior research. We believe our work points towards a more practical means to bring these use cases closer to feasibility.

One obvious use for fine-grained hand activity sensing is personal informatics [33]. Hand activities are a fine-grained, contextual channel that naturally complements ambulatory state [9]. Fine-grained activity tracing (e.g., washing hands, brushing teeth – Figure 12, A & B) has been shown to nudge users towards more healthy lifestyles [16], and spark personal reflection and social facilitation [22]. A system that knows what hands are doing could also have many health-related applications. For example, a smartwatch could track a user's typing behavior to prevent repetitive strain injury (RSI) [8]. Likewise, a smartwatch might be able to track smoking as part of a cessation regime [11], or monitor a construction workers tool use to prevent hand-arm vibration syndrome (HAVS) [68][69]. Eldercare monitoring systems [19][59][68] could also make use of this new and nuanced information source.

There has also been interesting research into automatic skill assessment [34][67]. Prior work has looked at musical skill acquisition [78], sports performance [52], and rehabilitation [4]. Automatic assessment could create opportunities for *in situ* feedback and skill-level evaluation [67]. It may even be possible to detect skill degradation overtime, and the onset of motor impairments such as Parkinson's.

Most generally, hand activity recognition could unlock richer, context-sensitive applications [14][38][39][44][58]. Sequences of fine-grained hand activities could also be used to infer higher-level human activities. For example, filling a kettle, turning on the stove, and then later pouring the kettle, can be indicative of the user making a cup of tea. However, if hand activities occur out of order (*e.g.,* pouring

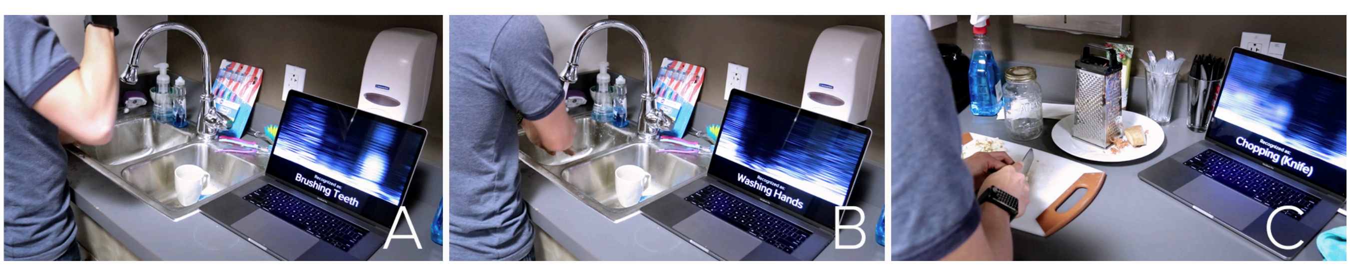

Figure 12. Real-time detection of hand activities can unlock many applications, ranging from personal informatics to skills assessment, and in general, power richer context-aware applications. Please also see Video Figure.

water before boiling it), it could suggest *e.g.,* the onset of dementia [40]. Hand activities could also be valuable in augmenting methods that gauge human interruptibility [25][29], for example delaying a notification while a user is chopping with a knife (Figure 12C).

12 CONCLUSION

We believe there is great value in knowing what activities the hands are engaged in to support assistive computational experiences. In this paper, we investigated the feasibility of such sensing using commodity smartwatches, which are an immediately practical means for achieving this vision. Our explorations started with an in-the-wild deployment and culminated with a controlled lab study. Our classification pipeline demonstrates 95.2% accuracy across 25 hand activities, and can reject unknown hand activities at 86.3% accuracy. At a high level, we believe these results bring the promise of contextually responsive applications much closer to reality, especially as our approach requires no external infrastructure or instrumentation of objects.

ACKNOWLEDGEMENTS

This work was generously supported with funds from the Packard Foundation, Sloan Foundation, and the Google Ph.D. Fellowship. We would also like to thank Prof. Robert Xiao for his early work on kernel driver integration.

REFERENCES

[1] Ahmad Akl, Chen Feng and Shakroh Valaee. 2011. A Novel Accelerometer-Based Gesture Recognition Sys-tem. In *IEEE Transactions on Signal Processing,* vol. 59, no. 12, pp. 6197-6205, Dec. 2011.

[2] Neena Aloysius and M. Geetha. 2017. A review on deep convolutional neural networks. 2017 *International Conference on Communication and Signal Processing (ICCSP),* Chennai, 2017, pp. 0588-0592.

[3] Ethel J. Alpenfels. 1955. The Anthropology and Social Significance of the Human Hand. In *Artificial Limbs. Volume 2*, Number 2. 4-21.

[4] Swamy Ananthanarayan, Miranda Sheh, Alice Chien, Halley Profita, and Katie Siek. 2013. Pt Viz: towards a wearable device for visualizing knee rehabilitation exercises. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13).* 1247-1250. https://doi.org/10.1145/2470654.2466161

[5] Android Wear 5.0.2 for Qualcomm chipsets, Inven-Sense MPU 6515 driver https://android.googlesource.com/kernel/msm.git/+/android-wear-5.0.2_r0.1/drivers/iio/imu/inv_mpu6515

[6] Ling Bao and Stephen S. Intille. 2004. Activity Recognition from User-Annotated Acceleration Data. In *Proceedings of Pervasive Computing '14. 1-17. DOI: 10.1007/978-3-540-24646-6_1*

[7] Yannick Bernaerts, Matthias Druwé, Sebastiaan Steensels, Jo Vermeulen, and Johannes Schöning. 2014. The office smartwatch: development and design of a smart-watch app to digitally augment interactions in an office environment. In *Proceedings of the 2014 companion publication on Designing interactive systems (DIS Companion '14).* ACM, New York. 41-44. DOI=http://dx.doi.org/10.1145/2598784.2602777

[8] Dave Berque, Jimmy Burgess, Alexander Billingsley, ShanKara Johnson, Terri L. Bonebright, and Brad Wethington. 2011. Design and evaluation of persuasive technology to encourage healthier typing behaviors. In *Proceedings of the 6th International Conference on Persuasive Technology: Persuasive Technology and De-sign: Enhancing Sustainability and Health (PERSUASIVE '11).* ACM, New York. Article 9, 10 pages. DOI: https://doi.org/10.1145/2467803.2467812

[9] J. B. J. Bussmann, W. L. J. Martens, J. H. M. Tulen, F. C. Schasfoort, H. J. G. van den Berg-Emons, and H. J. Stam. 2001. Measuring daily behavior using ambulatory accelerometry: The Activity Monitor. In *Behavior Research Methods, Instruments, & Computers.* August 2001, Volume 33, Issue 3, pp 349-356.

[10] Yung-Ju Chang, Gaurav Paruthi, and Mark W. New-man. 2015. A field study comparing approaches to collecting annotated activity data in real-world settings. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15).* ACM, New York, NY, USA, 671-682. DOI: http://dx.doi.org/10.1145/2750858.2807524

[11] Soujanya Chatterjee, Karen Hovsepian, Hillol Sarker, Nazir Saleheen, Mustafa al'Absi, Gowtham Atluri, Emre Ertin, Cho Lam, Andrine Lemieux, Motohiro Nakajima, Bonnie Spring, David W. Wetter, and Santosh Kumar. 2016. mCrave: continuous estimation of craving during smoking cessation. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16).* ACM, New York, NY, USA, 863-874. DOI: http://dx.doi.org/10.1145/2971648.2971672

[12] Tanzeem Choudhury, Gaetano Borriello, Sunny Consolvo, Dirk Haehnel, Beverly Harrison, Bruce Heming-way, Jeffrey Hightower, Predrag "Pedja" Klasnja, Karl Koscher, Anthony LaMarca, James A. Landay, Louis LeGrand, Jonathan Lester, Ali Rahimi, Adam Rea, and Danny Wyatt. 2008. The Mobile Sensing Platform: An Embedded Activity Recognition System. In *IEEE Pervasive Computing 7, 2* (April 2008), 32-41. DOI=http://dx.doi.org/10.1109/MPRV.2008.39

[13] Brian Clarkson and Nitin Sawhney and Alex Pentland. 1998. Auditory Context Awareness via Wearable Computing. In *Proceedings of the 1998 Workshop on Perceptual User Interfaces (PUI '98).*

[14] Brian Clarkson. 2002. Life Patterns: Structure from Wearable Sensors. *Ph.D. Thesis*, Massachusetts Institute of Technology.

[15] Gabe Cohn, Sidhant Gupta, Jon Froehlich, Eric Larson, and Shwetak N. Patel. 2010. GasSense: appliance-level, single-point sensing of gas activity in the home. In *Proceedings of the 8th international conference on Pervasive Computing (Pervasive'10).* 265-282.

[16] Sunny Consolvo, David W. McDonald, Tammy Toscos, Mike Y. Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony LaMarca, Louis LeGrand, Ryan Libby, Ian Smith, and James A. Landay. 2008. Activity sensing in the wild: a field trial of ubifit garden. In *Proceedings of the SIGCHI Conference on Hu-man Factors in Computing Systems (CHI '08).* ACM, New York, NY, USA, 1797-1806. DOI=http://dx.doi.org/10.1145/1357054.1357335

[17] Mihaly Csikszentmihalyhi and Reed Larson. 1987. Validity and Reliability of the Experience-Sampling Method. *Journal of Nervous and Mental Disease.* 175:526-536.

[18] Mark R. Cutkosky. 1989. On grasp choice, grasp models, and the design of hands for manufacturing tasks. In *IEEE Transactions on Robotics and Automation*, vol. 5, no. 3. 269-279.

[19] Siddharth Dalal, Majd Alwan, Reza Seifrafi, Steve Kell, and Donald Brown. A rule-based approach to the analysis of elders activity data: Detection of health and possible emergency conditions. In *Proceedings of AAAI Fall 2005 Symposium.* 2545-2552, 2005.

[20] Huong Giang Doan, Hai Vu, and Thanh Hai Tran. 2015. Recognition of hand gestures from cyclic hand movements using spatial-temporal features. In *Proceedings of the Sixth International Symposium on Information and Communication Technology (SoICT 2015).* ACM, New York, NY, USA, 260-267. DOI=http://dx.doi.org/10.1145/2833258.2833301

[21] Aaron M. Dollar. 2014. Classifying Human Hand Use and the Activities of Daily Living. The Human Hand as an Inspiration for Robot Hand Development. Volume 95, Springer *Tracts in Advanced Robotics.* 201-216.

[22] Daniel A. Epstein, Jennifer H. Kang, Laura R. Pina, James Fogarty, and Sean A. Munson. 2016. Reconsidering the device in the drawer: lapses as a design opportunity in personal informatics. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16).* ACM, New York, NY, USA, 829-840. DOI: http://dx.doi.org/10.1145/2971648.2971656

[23] FitBit. Healt and Activity Index. Last Accessed: January 2, 2019. https://www.fitbit.com/activity-index

[24] F. Foerster, M. Smeja, J. Fahrenberg. 1999. Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring. In *Computers in Human Behavior,* Volume 15, Issue 5, 1 September 1999, Pages 571-583, ISSN 0747-5632, http://dx.doi.org/10.1016/S0747-5632(99)00037-0.

[25] James Fogarty, Scott E. Hudson, and Jennifer Lai. 2004. Examining the robustness of sensor-based statistical models of human interruptibility. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04).* ACM, New York. 207-214. DOI=http://dx.doi.org/10.1145/985692.985719

[26] Jon E. Froehlich, Eric Larson, Tim Campbell, Conor Haggerty, James Fogarty, and Shwetak N. Patel. 2009. HydroSense: infrastructure-mediated single-point sensing of whole-home water activity. In *Proceedings of the 11th international conference on Ubiquitous computing (UbiComp '09).* ACM, New York, NY, USA, 235-244. DOI=http://dx.doi.org/10.1145/1620545.1620581

[27] Sidhant Gupta, Daniel Morris, Shwetak Patel, and Desney Tan. 2012. SoundWave: using the doppler effect to sense gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12).* ACM, New York, NY, USA, 1911-1914. DOI: http://dx.doi.org/10.1145/2207676.2208331

[28] Chris Harrison, Desney Tan, and Dan Morris. 2010. Skinput: appropriating the body as an input surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10).* ACM, New York, NY, USA, 453-462. DOI=http://dx.doi.org/10.1145/1753326.1753394

[29] Scott Hudson, James Fogarty, Christopher Atkeson, Daniel Avrahami, Jodi Forlizzi, Sara Kiesler, Johnny Lee, and Jie Yang. 2003. Predicting human interruptibility with sensors: A Wizard of Oz feasibility study. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03).* 257-264. DOI=http://dx.doi.org/10.1145/642611.642657

[30] InvenSense, Inc. MPU-6500 Register Map and Descriptions Revision 2.1. September 16, 2013. https://www.invensense.com/wp-content/uploads/2015/02/MPU-6500-Register-Map2.pdf

[31] Sergey Ioffe and Christian Szegedy. 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In *ICML, volume 37 of JMLR Workshop and Conference Proceedings.* 448-456.

[32] Howell Istance, Richard Bates, Aulikki Hyrskykari, and Stephen Vickers. 2008. Snap clutch, a modeled approach to solving the Midas touch problem. In *Proceedings of the 2008 symposium on Eye tracking research & applications (ETRA '08).* 221-228. DOI: https://doi.org/10.1145/1344471.1344523

[33] Elisabeth T. Kersten-van Dijk, Joyce H.D.M. Westerink, Femke Beute & Wijnand A. IJsselsteijn. 2017. Personal Informatics, Self-Insight, and Behavior Change: A Critical Review of Current Literature. In *Journal of Human-Computer Interaction.* Taylor & Francis Publishing. 1-29. DOI http://dx.doi.org/10.1080/07370024.2016.1276456

[34] Aftab Khan, Sebastian Mellor, Eugen Berlin, Robin Thompson, Roisin McNaney, Patrick Olivier, and Thomas Plötz. 2015. Beyond activity recognition: skill assessment from accelerometer data. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15).* ACM, New York, NY, USA, 1155-1166. DOI: http://dx.doi.org/10.1145/2750858.2807534

[35] David Kim, Otmar Hilliges, Shahram Izadi, Alex D. Butler, Jiawen Chen, Iason Oikonomidis, and Patrick Olivier. 2012. Digits: freehand 3D interactions any-where using a wrist-worn gloveless sensor. In *Proceedings of the 25th annual ACM symposium on User inter-face software and technology (UIST '12).* ACM, New York, NY, USA, 167-176. DOI: http://dx.doi.org/10.1145/2380116.2380139

[36] Jacek A. Kopec, John M. Esdaile. 1990. Bias in case-control studies: A review. *Journal of epidemiology and community health.* 44 (3): 179-86

[37] Wilton M. Krogman. 1942. The Anthropology of the hand. *Ciba Symposia*, Vol. 4 (4). 1294.

[38] Gierad Laput, Chouchang Yang, Robert Xiao, Alanson Sample, and Chris Harrison. 2015. EM-Sense: Touch Recognition of Uninstrumented, Electrical and Electromechanical Objects. *In Proc. of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15).* 157-166. DOI: https://doi.org/10.1145/2807442.2807481

[39] Gierad Laput, Robert Xiao, and Chris Harrison. 2016. ViBand: High-Fidelity Bio-Acoustic Sensing Using Commodity Smartwatch Accelerometers. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16).* 321-333. DOI: https://doi.org/10.1145/2984511.2984582

[40] Gierad Laput, Yang Zhang, and Chris Harrison. 2017. Synthetic Sensors: Towards General-Purpose Sensing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI '17). ACM, New York, NY, USA, 3986-3999. DOI: https://doi.org/10.1145/3025453.3025773

[41] Gierad Laput, Karan Ahuja, Mayank Goel, and Chris Harrison. 2018. Ubicoustics: Plug-and-Play Acoustic Activity Recognition. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology (UIST '18).* ACM, New York, NY, USA, 213-224. DOI: https://doi.org/10.1145/3242587.3242609

[42] Walter S. Lasecki, Young Chol Song, Henry Kautz, and Jeffrey P. Bigham. 2013. Real-time crowd labeling for deployable activity recognition. In *Proceedings of the 2013 conference on Computer supported cooperative work (CSCW '13).* 1203-1212. DOI=http://dx.doi.org/10.1145/2441776.2441912

[43] Seon-Woo Lee and Kenji Mase. 2002. Activity and Location Recognition Using Wearable Sensors. In *IEEE Pervasive Computing* 1, 3 (July 2002), 24-32. DOI=http://dx.doi.org/10.1109/MPRV.2002.1037719

[44] Hanchuan Li, Can Ye, and Alanson P. Sample. 2015. IDSense: A Human Object Interaction Detection System Based on Passive UHF RFID. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15).* ACM, New York, NY, USA, 2555-2564. DOI: https://doi.org/10.1145/2702123.2702178

[45] Jaime Lien, Nicholas Gillian, M. Emre Karagozler, Patrick Amihood, Carsten Schwesig, Erik Olson, Hakim Raja, and Ivan Poupyrev. 2016. Soli: ubiquitous gesture sensing with millimeter wave radar. *ACM Trans. Graph.* 35, 4, Article 142 (July 2016), 19 pages. DOI: https://doi.org/10.1145/2897824.2925953

[46] Beth Logan, Jennifer Healey, Matthai Philipose, Emmanuel Munguia Tapia, and Stephen Intille. 2007. A long-term evaluation of sensing modalities for activity recognition. In *Proceedings of the 9th international conference on Ubiquitous computing (UbiComp '07).* Springer-Verlag, Berlin, Heidelberg, 483-500.

[47] Takuya Maekawa, Yasue Kishino, Yutaka Yanagisawa and Yasushi Sakurai. 2012. WristSense: Wrist-worn sensor device with camera for daily activity recognition. In *IEEE Pervasive Computing.* 510-512.

[48] David McNeill. 1985. So You Think Gestures Are Nonverbal? In Psychological Review 92, No. 3. 350-371.

[49] Dan Morris, T. Scott Saponas, Andrew Guillory, and Ilya Kelner. 2014. RecoFit: using a wearable sensor to find, recognize, and count repetitive exercises. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14).* ACM, New York, NY, USA, 3225-3234. DOI: https://doi.org/10.1145/2556288.2557116

[50] Alessandra Moschetti, Laura Fiorini, Dario Esposito, Paolo Dario, and Filippo Cavallo. 2016. Recognition of Daily Gestures with Wearable Inertial Rings and Bracelets. In *Sensors* (Basel, Switzerland) vol. 16,8 1341. 22 Aug. 2016. DOI:10.3390/s16081341

[51] Alessandra Moschetti, Laura Fiorini, Dario Esposito, Paolo Dario, and Filippo Cavallo. 2017. Daily activity recognition with inertial ring and bracelet: An unsupervised approach. In *Proceedings of IEEE International Conference on Robotics and Automation (ICRA),* Singapore, 2017, pp. 3250-3255. DOI: 10.1109/ICRA.2017.7989370

[52] Josip Music and Roderick Murray-Smith. 2010. Virtual hooping: teaching a phone about hula-hooping for fitness, fun and rehabilitation. In *Proceedings of Mobile HCI (MobileHCI '10).* ACM, New York, NY, USA, 309-312. DOI=http://dx.doi.org/10.1145/1851600.1851654

[53] Vinod Nair and Geoffrey E. Hinton. 2010. Rectified Linear Units Improve Restricted Boltzmann Machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10).* 807-814.

[54] Masa Ogata and Michita Imai. 2015. SkinWatch: skin gesture interaction for smart watch. In *Proceedings of the 6th Augmented Human International Conference (AH '15).* ACM, New York, NY, USA, 21-24. DOI: http://dx.doi.org/10.1145/2735711.2735830

[55] Vesa Peltonen, Juha Tuomi, Anssi Klapuri, Jyri Huopaniemi and Timo Sorsa. 2002. Computational auditory scene recognition. In *IEEE International Conference on Acoustics, Speech, and Signal Processing,* Orlando, FL, USA, 2002, pp. II-1941-II-1944.

[56] John K. Perng, Brian Fisher, Seth Hollar and Kristofer S. J. Pister. 1999. Acceleration sensing glove (ASG)". In *Proceedings of the Third International Symposium on Wearable Computers,* San Francisco, CA, USA, 1999, pp. 178-180.
[57] Matthai Philipose. 2005. Large-Scale Human Activity Recognition Using Ultra-Dense Sensing. In *The Bridge, National Academy of Engineering.*
[58] Cliff Randell and Henk Muller. 2000. Context Aware-ness by Analyzing Accelerometer Data. In *Proceedings of the 4th IEEE International Symposium on Wearable Computers (ISWC '00).* IEEE, Washington, DC, USA, 175.
[59] Marilyn J. Rantz et al. 2011. Using sensor networks to detect urinary tract infections in older adults. In *IEEE 13th International Conference on e-Health Networking, Applications and Services,* Columbia, MO, 2011, pp. 142-149.
[60] Ruth Ravichandran, Elliot Saba, Ke-Yu Chen, Mayank Goel, Sidhant Gupta and Shwetak N. Patel. 2015. Wi-Breathe: Estimating respiration rate using wireless signals in natural settings in the home. In *IEEE International Conference on Pervasive Computing and Communications (PerCom),* St. Louis, MO, 2015, pp. 131-139.
[61] T. Scott Saponas, Desney S. Tan, Dan Morris, Ravin Balakrishnan, Jim Turner, and James A. Landay. 2009. Enabling always-available input with muscle-computer interfaces. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology (UIST '09).* 167-176. DOI=http://dx.doi.org/10.1145/1622176.1622208
[62] Dominik Scherer, Andreas C. Müller, and Sven Behnke. 2010. Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition. In *Artificial Neural Networks (ICANN),* 20th International Conference on. Thessaloniki, Greece: Springer. 92-101.
[63] Frank M. Shipman, Ricardo Gutierrez-Osuna, and Caio D. D. Monteiro. 2014. Identifying Sign Language Videos in Video Sharing Sites. In *ACM Trans. Access. Comput. 5, 4, Article 9 (March 2014),* 14 pages. DOI: http://dx.doi.org/10.1145/2579698
[64] Karen Simonyan and Andrew Zisserman. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. In arXiv. eprint arXiv:1409.1556. 2014arXiv1409.1556S
[65] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Drop-out: a simple way to prevent neural networks from over-fitting. *In J. Mach. Learn. Res.* 15, 1 (January 2014), 1929-1958.
[66] Thad Starner, Bernt Schiele and Alex Pentland. 1998. Visual contextual awareness in wearable computing". In *Proc. of the Second International Symposium on Wearable Computers (Cat. No.98EX215),* 1998, pp. 50-57.
[67] Todor Stojanov and Xue-Mei Ding. 2015. Operators Skill Level Evaluation Method for Balancing of an Apparel Assembly Line. *Int. J. Productiv. Manage. As-sess. Technol.* 3, 1 (January 2015), 1-12. DOI=http://dx.doi.org/10.4018/IJPMAT.2015010101
[68] James S. Taylor. 1985. Vibration syndrome in industry: Dermatological viewpoint. *Am. J. Ind. Med.*, 8: 415-432. doi:10.1002/ajim.4700080423
[69] Vibration Syndrome. Current Intelligence Bulletin #38. National Institute for Occupational Safety and Health (NIOSH), Publication No. 83-110. March 29, 1983. Cincinnati, OH.
[70] Christian Vogler and Dimitris Metaxas. 1998. ASL recognition based on a coupling between HMMs and 3D motion analysis. In *Proc. of the Sixth International Conference on Computer Vision,* IEEE Cat. No.98CH36271, Bombay, 1998, pp. 363-369.
[71] Edward J. Wang, Tien-Jui Lee, Alex Mariakakis, Ma-yank Goel, Sidhant Gupta, and Shwetak N. Patel. 2015. MagnifiSense: inferring device interaction using wrist-worn passive magneto-inductive sensors. In Proceedings of the 2015 ACM *International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15).* ACM, New York, NY, USA, 15-26. DOI: http://dx.doi.org/10.1145/2750858.2804271
[72] Jamie. A. Ward, Paul Lukowicz, Gerhard Troster and Thad E. Starner. 2006. Activity Recognition of Assembly Tasks Using Body-Worn Microphones and Accelerometers. In *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 28, no. 10, pp. 1553-1567, Oct. 2006.
[73] Hiroki Watanabe, Tsutomu Terada, and Masahiko Tsukamoto. 2013. Ultrasound-based movement sensing, gesture-, and context-recognition. In *Proceedings of the 2013 International Symposium on Wearable Computers (ISWC '13).* ACM, New York, NY, USA, 57-64. DOI=http://dx.doi.org/10.1145/2493988.2494335
[74] David Way and Joseph Paradiso. 2014. A Usability User Study Concerning Free-Hand Microgesture and Wrist-Worn Sensors. In *Proceedings of the 2014 11th International Conference on Wearable and Implantable Body Sensor Networks (BSN '14).* IEEE, Washington, DC, USA, 138-142. DOI=http://dx.doi.org/10.1109/BSN.2014.32
[75] Hongyi Wen, Julian Ramos Rojas, and Anind K. Dey. 2016. Serendipity: Finger Gesture Recognition using an Off-the-Shelf Smartwatch. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16).* 3847-3851. DOI: https://doi.org/10.1145/2858036.2858466
[76] Frank R. Wilson. 1998. The Hand: How its use shapes the brain, language, and human culture. Random House, Inc.
[77] Charlotte Wolf. 1942. The Human Hand. Methuen, London.
[78] Iwan G. J. H. Wopereis, Jeroen J. G. van Merriënboer, and Paul A. Kirschner. 2010. Improvising in music: a learning biography study to reveal skill acquisition. In *Proceedings of the 9th International Conference of the Learning Sciences - Volume 2 (ICLS '10),* Vol. 2. International Society of the Learning Sciences 419-420.
[79] Jie Yang, Jiang Gao, Ying Zhang, and Alex Waibel. 2001. Towards automatic sign translation. In *Proceedings of the first international conference on Human language technology research (HLT '01).* 1-6. DOI=http://dx.doi.org/10.3115/1072133.1072223
[80] Hui-Shyong Yeo, Gergely Flamich, Patrick Schrempf, David Harris-Birtill, and Aaron Quigley. 2016. Radar-Cat: Radar Categorization for Input & Interaction. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16).* ACM, New York, NY, USA, 833-841. DOI: https://doi.org/10.1145/2984511.2984515
[81] Ouyang Zhang and Kannan Srinivasan. 2016. Mudra: User-friendly Fine-grained Gesture Recognition using WiFi Signals. In *Proceedings of the 12th International on Conference on emerging Networking Experiments and Technologies (CoNEXT '16).* ACM, New York, NY, USA, 83-96. DOI: https://doi.org/10.1145/2999572.2999582
[82] Yang Zhang and Chris Harrison. 2015. Tomo: Wearable, Low-Cost Electrical Impedance Tomography for Hand Gesture Recognition. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15).* 167-173. DOI: https://doi.org/10.1145/2807442.2807480