Pervasive and Mobile Computing 8 (2012) 103–114

Contents lists available at SciVerse ScienceDirect

Pervasive and Mobile Computing

journal homepage: www.elsevier.com/locate/pmc

Fast track article

Detecting stereotypical motor movements in the classroom using accelerometry and pattern recognition algorithms

Fahd Albinali [a,]*, Matthew S. Goodwin [a,b], Stephen Intille [a,c]

[a] *Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[b] *The Groden Center, Inc., Providence, RI 02906, USA*
[c] *Northeastern University, Boston, MA 02115, USA*

ARTICLE INFO

Article history:
Received 16 January 2010
Received in revised form 13 April 2011
Accepted 14 April 2011
Available online 21 April 2011

Keywords:
Autism
Accelerometers
Activity recognition
Pattern recognition

ABSTRACT

Individuals with Autism Spectrum Disorders (ASD) frequently engage in stereotyped and repetitive motor movements. Automatically detecting these movements using comfortable, miniature wireless sensors could advance autism research and enable new intervention tools for the classroom that help children and their caregivers monitor, understand, and cope with this potentially problematic class of behavior. We present activity recognition results for stereotypical hand flapping and body rocking using accelerometer data collected wirelessly from six children with ASD repeatedly observed by experts in real classroom settings. An overall recognition accuracy of 88.6% (TP: 0.85; FP: 0.08) was achieved using three sensors. We also present pilot work in which non-experts use software on mobile phones to annotate stereotypical motor movements for classifier training. Preliminary results indicate that non-expert annotations for training can be as effective as expert annotations. Challenges encountered when applying machine learning to this domain, as well as implications for the development of real-time classroom interventions and research tools are discussed.

1. Introduction

Health researchers in many disciplines lack effective tools for unobtrusively acquiring information about peoples' behavior in natural settings. Ubiquitous computing systems that detect certain behaviors might create new opportunities to improve scientific understanding of the interaction between context, behavior, and health. The goal of the current work is to use ubiquitous monitoring tools for the automated detection of stereotypical motor movements observed in persons with Autism Spectrum Disorders. Autism Spectrum Disorders (ASD) affect as many as 1 in 110 children [1] and are characterized by deficits in socialization and communication, including stereotypical behavior [2]. Stereotyped behaviors are generally defined as repetitive interests and/or motor or vocal sequences that appear to the observer to be invariant in form and without any obvious eliciting stimulus or adaptive function [3]. The current work focuses on stereotypical motor movements. Several stereotypical motor movements have been identified [4], the most prevalent among them being body-rocking, mouthing, and complex hand and finger movements [5]. The majority of research in ASD focuses on social and communication deficits, rather than on restricted and repetitive behavior [4]. A lack of research in stereotypical movements is a potential problem given the high prevalence of stereotypical motor movements reported in individuals with ASD (e.g., [6]).

* Corresponding author.
E-mail address: albinali@mit.edu (F. Albinali).

1574-1192/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.pmcj.2011.04.006

One reason why stereotypical motor movements may not be as thoroughly studied is because appropriate tools for measuring the behavior are not available to the research community. In this work, we present a case study on the automatic identification of stereotypical body rocking and hand flapping activity in children with ASD gathered from wireless accelerometers. Stereotypical body rocking and hand flapping are examples of movements that occur frequently in people with mental retardation and developmental disabilities [4], and less frequently in typically developing children and adults.

1.1. Impact of stereotypical motor movements

When severe, stereotypical motor movements can present several problems for individuals with ASD and their caregivers. First, some persons with ASD often engage in stereotypical motor movements for the majority of their waking hours. Second, if unregulated, stereotypical motor movements can become the dominant behavior in an individual with ASD's repertoire and interfere with the acquisition of new skills and performance of established skills (e.g., [7]). Third, engagement in these movements is socially inappropriate and stigmatizing and can complicate social integration in school settings and the community [8]. Finally, stereotypical motor movements can lead to self-injurious behavior under certain environmental conditions [9].

1.2. Tools for measuring stereotypical motor movements

There are currently no tools for clinicians or caregivers to easily, accurately, and reliably monitor stereotypical motor movements. Traditional measures of stereotypical motor movements rely primarily on paper-and-pencil rating scales, direct observation, and video-based methods [10], all of which have limitations.

Paper-and-pencil rating scales typically involve a global impression of the frequency and/or severity of stereotypical motor movements based on general, non-specific observations. Several paper-and-pencil rating scales have been developed that ask an informant to give a global impression of an individual's stereotypical motor movements (e.g., [4]). From a measurement standpoint, informant rating scales are subjective, can have questionable accuracy, and fail to capture inter-individual variations in the form, amount, and duration of stereotypical motor movements [11].

Direct observation also involves a rating but the focus is on the direct observation of specific behaviors. The observer watches and records a sequence of stereotypical motor movements. According to Sprague and Newell [10], the following factors, among others, can make direct observational measures unreliable: (a) Reduced accuracy in observing and documenting high-speed motor sequences; (b) Difficulty determining when a sequence has started and ended; (c) Limitations in the ability to observe concomitantly occurring stereotypical motor movements; and (d) Limitations in the ability to note environmental antecedents and record stereotypical motor movements at the same time.

Video-based methods involve video capture of behavior and off-line coding of stereotypical motor movements by an expert. The ability to view videos repeatedly and to slow playback speeds makes video-based methods more reliable than paper-and-pencil and direct observation methods. Video-based methods, however, are tedious and time consuming. The necessity to code videos off-line also precludes real-time monitoring. However, combining video recording with other tagging technologies to permit practical, semi-automatic logging is an area of active research [12].

1.3. Goal: Explore the possibility of automating recognition of stereotypical motor movements

The primary aim of the current work is to explore whether wireless accelerometer sensor technology and pattern recognition algorithms can provide an automatic measure of stereotypical motor movements that may be more objective, detailed, and precise than rating scales and direct observation, and more time-efficient than video-based methods. An algorithm that achieves good recognition performance could also operate for much longer periods of time than a human observer. Another goal of our work is to move towards a real-time annotation and recognition system that could be used by teachers and caregivers to annotate and train algorithms while giving real-time feedback on a convenient mobile device such as a phone. To assess the viability of having teachers use our system, we report initial pilot experiments to analyze the quality of non-expert annotations and to assess their impact on the automated recognition of stereotypical motor movements.

In the remainder of this paper we describe analyses we have performed to determine whether pattern recognition techniques using mobile wireless accelerometers that have shown promise in other domains of recognition of posture, mobility, exercise, and everyday activities can be adapted to create a tool for stereotypical motor movement monitoring in children with ASD.

2. Related work

2.1. Using pattern recognition to detect physical activities

A growing body of work shows that wearable accelerometers can be used to detect activities, such as postures, ambulation, exercise, and even household activities (e.g., [13–15]). A variety of methods and models have been used for feature generation and classification. Our focus in this work is not on any particular activity recognition algorithm, per se, but instead on the issues one encounters when trying to apply pattern recognition to the problem of monitoring stereotypical motor movements in naturalistic settings.

2.2. Quality of non-expert annotations

A variety of domains have collected annotations from non-experts to train classification algorithms. For example, von Ahn collected online annotations using games such as ESPGame [16] for labeling images and Verbosity [17] for annotating word relations. Snow et al. [18] showed that non-expert raters could be as effective as expert raters in natural language tasks. Kittur et al. [19] describe the use and necessity of verifiable questions in acquiring accurate ratings of Wikipedia articles from Mechanical Turk users. Although non-expert annotations are often noisier than expert annotations, the exemplary work above suggests that they are more readily available, easier to scale, and often able to achieve high reliability when aggregated.

2.3. Automatically detecting stereotypical motor movements

Other than our own prior work [20] that we extend here, we are aware of only one published attempt to apply Accelerometry and pattern recognition algorithms to this domain [12,21]. While 69% of hand flapping events were automatically and accurately detected in this work using Hidden Markov Models, the data were acquired from individuals mimicking the actual behaviors – the work did not observe children with ASD actually performing the behaviors.

Most prior work in accelerometer-based activity recognition uses supervised learning strategies. Activities are performed by people wearing wired or wireless accelerometers on one or more body locations. Annotators (usually the researchers) then use video or audio to label the start and end points of each behavior of interest. Algorithms are then tested on the datasets using cross-fold validation. We use this same approach with expert and non-expert annotators, and describe the challenges we encountered in the stereotypical motor movement recognition domain.

3. Data collection

The current investigation consisted of a series of six single case studies, with direct replication across participants. For each participant, the study included repeated observations of body rocking, hand flapping, and/or simultaneous body rocking and hand flapping while children wore sensors in a classroom setting (Study 1). To analyze the quality of non-expert annotations, we ran four additional data sessions where an expert and a non-expert annotator encoded the movements of one of the participants in both a laboratory and a classroom setting (Study 2). Results from earlier experiments involved in Study 1 that were undertaken in a laboratory setting are reported in previous work [20].

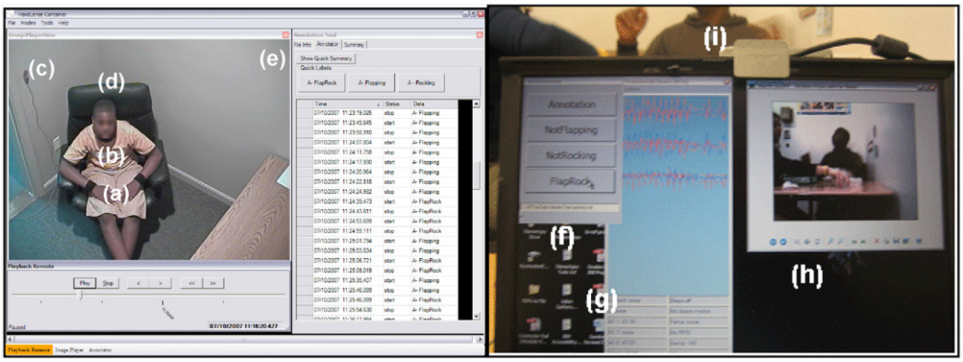
3.1. Participants

Six participants were recruited from The Groden Center, RI, a school for children and young adults with ASD. The study was approved by a human subjects review board and parental consent was obtained for each participant. Children included in the study: (1) Had a documented DSM-IV-TR diagnosis of ASD made by a licensed psychologist familiar with the child; (2) Were between the ages of 12–20 yrs.; (3) Had a clinically significant score on the Stereotyped Behavior subscale of the Repetitive Behavior Scale-Revised (RBS-R; [22]) for body rocking and/or hand flapping; (4) Tolerated the wireless sensors; and (5) Exhibited, on average, at least 10 hand flapping or body rocking incidents per hour. Based on informal observations in the school, we note that our sample appeared to be representative of the population of students at the school, however we do not have data on how the children compare to others elsewhere and not in special environments.

3.2. Sensors

Each participant wore three wireless accelerometers placed simultaneously on the left wrist and right wrist using wristbands, and on the torso using a thin strip of comfortable fabric tied around the chest (see Fig. 1(a)–(b)). The wrists and torso were chosen because stereotypical hand flapping and body rocking are associated with movements in these areas. The sensors were small enough to be worn on these locations comfortably and without restricting movement. All participants tolerated wearing the sensors for the duration of each observation. Also, visual inspection of each participant's acceleration data prior to analysis confirmed that there were no equipment failures or other problems occurring (improper attachment, weak signal strength, unusual amount of signal loss, etc.).

Two different accelerometer systems were used. In Study 1, the devices were set to transmit 3-axis $+/-2$ g motion data at 60 Hz to a nearby receiver (Fig. 1(c)) attached to a desktop computer [23]. It should be noted that the body can block the 2.4 GHz range low-power radio signal, so there is occasional signal loss experienced that the pattern recognition algorithms must compensate for. A receiver was plugged into a standard laptop where data from the three sensors were synchronized and saved to disk. Simultaneously, a video camera attached to the laptop was used to capture video of the scene that could be synchronized with the accelerometer streams and used for annotation of activity. We note that we have not used any audio and the syncing of the video frames was done using the computer clock. In Study 2, because our long-term goal is to construct a mobile recognition system that could be used by parents and teachers, we used a newer generation of our sensors that samples 3-axis $+/-4$ g motion data at 90 Hz and transmits it using Bluetooth directly to the mobile phone [24]. The use



Time	Status	Data
07/10/2007 11:23:19.025	stop	A- Flapping
07/10/2007 11:23:49.845	start	A- Flapping
07/10/2007 11:23:58.910	stop	A- Flapping
07/10/2007 11:24:07.804	start	A- Flapping
07/10/2007 11:24:11.758	stop	A- Flapping
07/10/2007 11:24:17.930	start	A- Flapping
07/10/2007 11:24:20.964	stop	A- Flapping
07/10/2007 11:24:22.818	start	A- Flapping
07/10/2007 11:24:24.932	stop	A- Flapping
07/10/2007 11:24:39.473	start	A- FlapRock
07/10/2007 11:24:43.611	stop	A- FlapRock
07/10/2007 11:24:53.699	start	A- FlapRock
07/10/2007 11:24:59.111	stop	A- FlapRock
07/10/2007 11:25:01.794	start	A- Flapping
07/10/2007 11:25:03.634	stop	A- Flapping
07/10/2007 11:25:06.721	start	A- FlapRock
07/10/2007 11:25:09.019	stop	A- FlapRock
07/10/2007 11:25:35.407	start	A- Flapping
07/10/2007 11:25:46.099	stop	A- Flapping
07/10/2007 11:25:46.099	start	A- FlapRock
07/10/2007 11:25:54.630	stop	A- FlapRock

Fig. 1. (a) A wireless accelerometer placed on each wrist. (b) A wireless accelerometer placed on the chest. (c) Receiver for sensor data. (d) An image of a child in the laboratory setting. (e) The video coding software that allows frame-accurate annotation. (f) The real-time activity annotator. (g) The acceleration data window plotting data streams in real-time. (h) The video window with images being captured. (i) USB camera clipped on to the top of the laptop in the classroom.

Table 1
Summary of participant stereotypical movements.

ID	Total duration (min:s)	% Engaged	*M* (*SD*) (s)	Num Stereo	Total stereo	Consistency
1	47:42	28	7 (6)	3	372	Mild
2	18:18	17.5	3 (2)	2	345	Very
3	10:00	8.5	4 (2)	3	149	Very
4	36:53	45	9 (12)	2	240	Very
5	32:14	48	7 (7)	3	253	Mild
6	67:28	71	21 (23)	2	199	Very

of Bluetooth on our new sensors provides better signal quality and is more robust against data loss. Importantly, the new sensors have been redesigned with comfort and ease of use as a primary requirement. The sensors are thin and waterproof and can be worn for extended periods in flexible fabric bands that dissipate sweat and minimize irritation. When worn, participants got used to them in minutes and did not seem to notice them afterwards.

3.3. Setting and procedure

In Study 1, repeated observations were undertaken in a classroom setting, which included a diverse set of stimuli, demands for shared attention, and other students present. While wearing the sensors, each participant was observed for 10–30 min (over a period of 12 months) during regular school hours on two separate occasions in their classrooms. These observations included typical classroom activities (e.g., eating lunch, spelling program, sorting), with participants working both on their own and with a familiar teacher. This data collection effort resulted in 4.5 h of data.[1]

In Study 2, we ran four 30-min data collection sessions alternating between the laboratory and classroom with one of the participants involved in Study 1. This data collection effort resulted in 2 h of data. The observational sequence in Study 2 consisted of an initial session in the laboratory (Lab1) followed by a session in the classroom (Class1). After an hour break, we conducted a second session in the laboratory (Lab2) followed by a second session in the classroom (Class2). The laboratory sessions differed from classroom sessions in that the laboratory contained limited stimulus materials, one-to-one monitoring by a familiar teacher, and no other students present. Also, in contrast to the classroom, there were no structured activities involved in the laboratory. We undertook data collection in both laboratory and classroom settings to determine the accuracy of annotation and recognition performance across both constrained and real-world environments. Participants were seated during all observations in both environments.

3.4. Stereotypical motor movements

One of the first challenges we encountered was simply the diversity of stereotypical motor movements observed in our participants, and the difficulty associated with annotating those movements.

Table 1 summarizes quantitative and qualitative stereotypical motor movement characteristics of the participants averaged across observation sessions. *Total Duration* is the total time spent engaged in stereotypical motor movements

[1] When we began this work, the primary focus was to collect large numbers of examples of stereotypical movements quickly. Sessions that had infrequent episodes of stereotypical movements (<10 per hour) were aborted to reduce undue burden on participants, and the data were discarded. In hindsight, given the relative difficulty acquiring high volume examples, data from these sessions should have been coded and used in the analysis.

across sessions. *% Engaged* is the percentage of time participants engaged in stereotypical motor movements during data collection sessions. *M* and *SD* are the mean duration and standard deviation of each participant's episodes of stereotypical movements. *Num Stereo* is the number of different types of stereotypical motor movements observed during all sessions and *Total Stereo* is the total number of episodes of those movements for each participant. This includes hand flapping, body rocking, and simultaneous hand flapping and body rocking—dubbed "flaprock". Finally, *Consistency* is a subjective grade (none, mild, or very) assigned by a trained behavioral scientist indicating how consistent each participant's stereotypical motor movement appeared to be.

3.4.1. Annotation

In Study 1, each session involved two observational coding procedures. The first, real-time coding, was undertaken during the sessions to see how well start time, end time, and type of stereotypical motor movement could be documented live by a trained observer. Start time, end time, and type of stereotypical motor movement were coded in real-time using custom annotation software (see Fig. 1(f)). The activity annotator included three buttons that corresponded to the stereotypical motor movements under observation (i.e., hand flapping, body rocking, flaprock). Pressing a button once marked the start of the corresponding stereotypical motor movement. Pressing a button a second time marked the end of the corresponding stereotypical motor movement.

The second observation coding procedure in Study 1, offline coding, was undertaken after the sessions using video records and computerized annotation software. A digital camera (mounted in the ceiling of the laboratory; attached to the front of the laptop in the classroom (Fig. 1(i)) was used to record each session.[2] The camera was placed at an angle with an unobstructed view of the participant to allow for accurate offline annotation. The camera was connected to a computer that synchronized the saved video with the accelerometer data streams. Start time, end time, and type of stereotypical motor movement were coded offline by two independent raters using a custom video coding software application (Fig. 1(e)).

3.4.2. Expert and non-expert mobile annotations

In Study 2, one expert annotator and one non-expert annotator simultaneously coded start time, end time, and type of stereotypical motor movement using a Windows Mobile phone running custom annotation software (see Fig. 3). The expert is a trained behavioral scientist who is familiar with the child and his specific movements and the non-expert is a teacher who is not familiar with the child and is not trained to identify specific stereotypical movements of the child. Before the first session, the non-expert annotator (i.e., Groden Center teacher) was given the phone and received a 3-min tutorial on how to annotate stereotypical motor movements by pressing corresponding labels included on the phone's screen. Pressing a button once marked the start of the corresponding stereotypical motor movement which included three behaviors: flapping, rocking and simultaneous flapping and rocking. Pressing a button a second time marked the end of the corresponding stereotypical motor movement.

4. Recognition evaluation and experiences

In this section we describe in detail our experience applying physical activity pattern recognition to the stereotypical motor movement recognition domain.

4.1. Algorithm

Prior work [23] demonstrates that decision tree classifiers namely C4.5 [25] can be used to effectively recognize a variety of physical activities. We are ultimately interested in creating a real-time recognition tool, and decision trees have a desirable combination of properties: They have performed well in prior experiments reported in the literature on posture and ambulatory recognition, and they are fast to run once trained. To construct the decision trees, we chose C4.5 over its proprietary commercial successor C5.0 because C4.5 is freely available and can be disseminated with our open-source software [26].

We use five time and frequency domain features computed for each acceleration stream. These are: (1) The distances between the means of the axes of each accelerometer to capture sensor orientation for posture; (2) Variance to capture the variability in different directions; (3) Correlation coefficients to capture the simultaneous motion in each axis direction; (4) Entropy to capture the type of stereotypical motor movement; and (5) FFT peaks and frequencies to capture differentiation between stereotypical motor movement intensities. The features are computed for a window of data, assembled into a vector and used as input to the C4.5 classifier in the WEKA toolkit [27]. WEKA is then used to evaluate classification performance using 10-fold cross-validation.

Stereotypical motor movements were labeled as flapping, rocking, or flaprock (i.e., simultaneous flapping and rocking). The mobile interface in Study 2 enforced mutual exclusivity, so that only one label could be active at once. Non-stereotypical motor movements were labeled as unknown segments. Including an unknown class resulted in highly skewed class

[2] Again, only results from the classroom sessions are reported here. Findings from laboratory sessions can be found in our previous report [19].

Table 2
Summary of analyses.

Exp	Description	Goals
#1	Trained using participant-dependent data and offline annotations from the classroom. Tested using cross-validation	1. Measure the performance of the classifier in a naturalistic setting 2. Measure the agreement between 2 offline annotators 3. Measure performance on agreement and disagreement segments
#2	Trained and compared 3 different methods: (1) One-annotator training that uses offline annotations; (2) One-annotator training that uses real-time annotations; and (3) Two-annotator training that uses agreement segments from 2 offline annotators for training	1. Understand the impact of the annotation (e.g. offline, real-time, multiple annotators) on the performance of the classifier 2. Measure the agreement between offline and real-time annotations 3. Determine and compare where errors occur in real-time and offline annotations scenarios
#3	Trained the classifier with data from all the participants but one and tested the performance on the left-out participant	1. Measure the performance of the classifier using participant-independent data 2. Determine if some stereotypical motor movements are more consistent across participants and therefore detectable using participant-independent training
#4	Trained the classifier with expert and non-expert annotations using 1 session and tested on 3 other sessions	1. Measure the agreement and differences between an expert and a non-expert annotator 2. Measure the impact of using non-expert annotations on the classifier 3. Determine the viability of using non-expert annotations

Table 3
Performance of the classifier on 6 participants in a classroom (offline one-annotator training).

Participant ID	Accuracy (%)	Accuracy (Agree) (%)	Accuracy (Disagree) (%)	TP	FP	Precision	Recall	K
1	84.6	85.9	64.9	0.82	0.07	0.58	0.82	0.72
2	87.5	88.9	62.8	0.89	0.12	0.75	0.88	0.83
3	89.7	90.6	47.6	0.74	0.07	0.55	0.74	0.86
4	91.2	93.7	39.8	0.90	0.05	0.88	0.90	0.92
5	88.0	90.2	43.0	0.85	0.08	0.72	0.83	0.83
6	90.7	93.0	58.5	0.90	0.07	0.86	0.90	0.88
Mean	88.6	90.4	52.8	0.85	0.08	0.72	0.84	0.84

distributions, such that the frequencies of stereotypical motor movements were substantially lower than the examples of the unknown class when stereotypical motor movements were not occurring. To reduce skewness in the present data, all classifiers used balanced data for training and natural imbalanced data for testing. Balancing the data was done by randomly under-sampling the majority class (i.e. unknown) and re-sampling minority classes (i.e. stereotypical motor movements) with replacement.

Nine acceleration streams (x, y, and z from three accelerometers) were broken into 50% overlapping sliding windows of length 1 s. Our choice of 1 s was based on pilot work where we changed the window length from 200 ms to 5 s and measured the performance of the C4.5 classifier over pilot datasets. Best performing window sizes were slightly different across subjects depending on individuals' behaviors. A window of 1 s obtained good overall accuracy across all subjects while minimizing classification delay. Windows that are shorter than 1 s had suboptimal performance, while longer windows did not result in a consistent improvement in classifier performance.

Cubic spline interpolation was used to fill in missing data points (e.g. due to wireless signal loss). Windows that lost more than 50% of their expected data points were excluded from the analysis. This amounted to less than 1% of the data across participants in both settings.

We conducted four types of analyses that are summarized in Table 2. The following metrics are used to report on the performance of the activity classifier: *Accuracy*, *True positive rate* (TP), *False positive rate (FP)*, *Precision*, and *Recall*.

In what we will call one-annotator training, we perform 10-fold cross-validation over each participant's data and present averaged results across classroom sessions. In two-annotator training, we train on only agreement segments between two annotators and test on the complete data including both agreement and disagreement segments. For the agreement portion of the data, we perform 10-fold cross-validation. For the disagreement portion, we train on the agreement data and test on the disagreement data. Results are then combined and averaged across sessions for each participant. Finally, we report on the agreement between two offline annotations and real-time-offline annotations using Cohen's Kappa inter-rater reliability statistic.

4.2. Study 1, analysis 1: Performance in classroom

Table 3 shows the overall performance results of the algorithm averaged over multiple sessions for each participant in the classroom. *Accuracy* is the average accuracy of the classifier across all sessions. *Accuracy (Agree)* is the accuracy of the

Table 4
Performance of the classifier using real-time and offline annotations.

Participant ID	Offline (%)	Real-time (%)	Agreement (%)	K
1	86.5	75.8	84.5	0.42
2	86.8	80.4	89.2	0.32
3	95.0	91.1	91.9	0.54
4	83.7	82.2	92.1	0.76
5	81.9	85.9	91.4	0.71
6	84.0	82.6	92.3	0.68
Mean	86.3	83.0	90.2	0.57

classifier on examples where both offline annotators agreed. *Accuracy (Disagree)* is the accuracy of the classifier on examples where both annotators disagreed. *TP* and *FP* are the true and false positive rates, respectively. These are followed by *Precision* and *Recall* [27]. Finally, K is Cohen's Kappa, a statistic representing inter-rater reliability between two offline annotators. Cohen's Kappa values are computed on the stereotypy categorization for every 1 s window.

The performance of the algorithm appears to be directly dependent on at least three factors: (1) The duration of each episode of stereotypical motor movement; (2) The percentage of time participants engaged in stereotypical motor movements; and (3) The consistency with which participants performed these movements.

Participants 1, 2, and 3 had the shortest mean duration for an episode of stereotypical motor movement (7, 3, and 4 s, respectively) and spent the least amount of time engaged in these movements (28%, 17.5%, and 8.5%, respectively). As expected, the recognition performance for participants 1 and 3 with respect to precision (0.58 and 0.55, respectively) and recall (0.82 and 0.74, respectively) is significantly lower than participants 4, 5, and 6 who engaged in stereotypical motor movements more often and for longer periods of time. For participant 2, precision and recall were higher, likely due to the fact that he engaged in very consistent stereotypical motor movements. Participants 4 and 6 showed the highest frequency and the longest duration of stereotypical motor movements and therefore performed the best with respect to accuracy (91.2% and 90.7%, respectively), TP rate (0.90 and 0.90, respectively), precision (0.88 and 0.86, respectively) and recall (0.90 and 0.90, respectively).

A major concern is the high false positive rates averaging 0.08 across all participants. For intervention applications that target specific stereotypical motor movements, the system would incorrectly deliver the intervention 8% of the time when the participant is not engaged in the behavior. A closer look at the distribution of FP errors across the different activities reveals that more than 75% of the FP errors are for the unknown class and less than 25% of the errors are shared between specific stereotypical motor movements. This brings the average FP rate for specific stereotypical motor movements down to approximately 0.03, which is more desirable when an intervention is to be delivered only when a stereotypical motor movement is occurring. Standard smoothing techniques may further reduce these errors.

Finally, the highest agreement between offline annotators is for participant 4 (kappa = 0.92) and the lowest agreement is for participant 1 (kappa = 0.72). Participant 1 exhibited the most inconsistent stereotypical motor movements, displaying a range (i.e., topography, intensity, duration) of different flapping and rocking movements. To determine whether the majority of classification errors occurred on the subset of data where annotators were not in agreement, we evaluated the classifier on agreement and disagreement segments independently. Not surprisingly, the algorithm performed poorly on segments where there was disagreement between the annotators. However, this modestly impacted the overall performance of the classifier because the frequency of disagreement in offline annotation was relatively low, averaging less than 3% of the collected data for each participant.

4.3. Study 1, analysis 2: Comparing performance using real-time and offline annotations

Table 4 compares the impact of real-time and offline annotation on the performance of the classifier in the classroom. The column labeled *Offline* reiterates the overall accuracy reported in the previous section with one-annotator training. *Real-time* describes the results using real-time annotations for training and offline annotations for testing. *Agreement* describes the results from training the algorithm on segments where both offline annotators agreed. Finally, K is Cohen's Kappa inter-rater reliability statistic that measures agreement between real-time and offline annotations.

Our first observation is that a strong association exists between duration of stereotypical motor movements and performance of the real-time annotator. For example, participants 1, 2, and 3 have the shortest mean durations (7, 3, and 4 s, respectively) and the least percentage of engagement in stereotypical motor movements (28%, 17.5%, and 8.5%, respectively). The kappa values between offline and real-time annotations for these participants are also lowest. The real-time annotations and offline annotations differ in at least two ways. First, the real-time annotator frequently missed short episodes of stereotypical motor movements. For participant 3, this constituted approximately 33% of the episodes that were labeled offline. Second, when the real-time annotations overlapped with corresponding offline annotations, the real-time onsets and offsets were shifted in time but biased slightly towards errors in onset. These two factors appear to reduce the performance of the real-time classifier relative to the offline one-annotator classifier, particularly when stereotypical motor movements are of short duration.

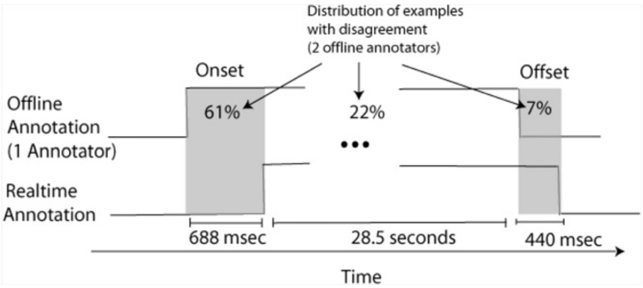

Distribution of examples with disagreement (2 offline annotators)
Onset
Offset
Offline Annotation (1 Annotator)
61%
22%
7%
•••
Realtime Annotation
688 msec
28.5 seconds
440 msec
Time

Fig. 2. Distribution of disagreements between offline annotators with respect to the onset and offset of an activity for participant 11.

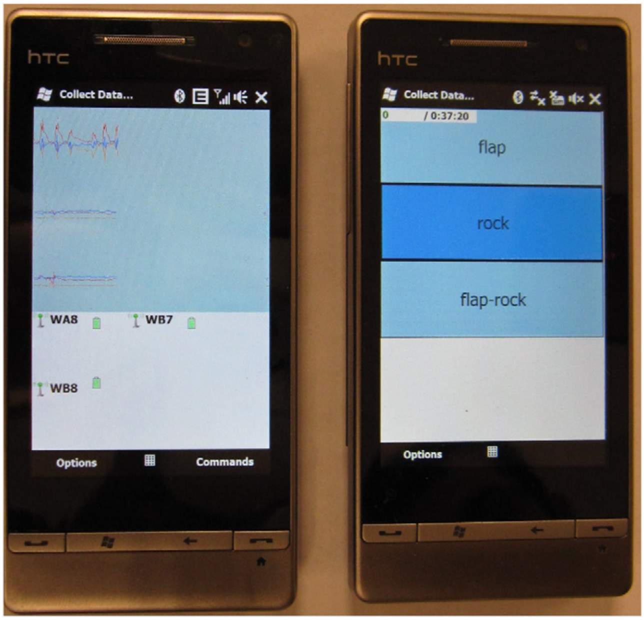


Fig. 3. Screenshot of phone plotting the real-time acceleration data streams received from the wireless sensors (left) and phone running real-time annotation software (right).

Our second observation is that when both the duration of stereotypical motor movements is long (7 s or more) and the percentage of engagement is high (40% or more), there is little difference between real-time and offline one-annotator classifiers. We expected the offline annotations to be of higher quality, with more accurate boundaries than the real-time labels, because real-time labeling is challenging given the differing speeds, frequency, and consistency of stereotypical motor movements. A surprising result, however, is that the performance of the classifier using real-time annotations for participant 5 outperformed the offline one-annotator classifier.

As exemplified by participant 6 (see Fig. 2), our third observation is that the highest frequency of offline disagreements occurs around the boundaries of stereotypical motor movements, particularly in the area that separates real-time and offline onsets (61% for participant 6). The lowest frequency of offline disagreements occurred in the area that separates the real-time and offline offsets (7% for participant 6). The rest of the disagreements were scattered within an episode (22%) or occurred in isolation (10%). As a result, the real-time training data included approximately 29% of the examples where the offline annotators disagreed, whereas the offline annotation included all the examples with disagreement. This may partially explain why the real-time classifier can sometimes outperform the offline one-annotator classifier on participants with longer episodes.

Our final observation is that offline annotation appears to facilitate labeling subtle and transitive variations on stereotypical motor movements, and that with no way to model the uncertainty of the annotator, the algorithm overemphasizes examples that are not particularly good for training. These transitive examples are likely to be missed in real-time annotation and thus not included in training. For stereotypical motor movements of long duration, missing noisy transitive examples in real-time annotation might improve accuracy. For stereotypical motor movements of short duration, real-time annotation seemingly misses both noisy transitive examples and good examples of short duration (e.g., 33% of the episodes were missed for participant 3). In this case, the increase in performance due to loss of noisy transitive examples did not offset the reduction in performance due to loss of good but short examples.

Table 5
Performance across different participants.

Participant ID	Accuracy (%)	TP	FP	Precision	Recall
1	74.3	0.52	0.13	0.48	0.52
2	77.1	0.53	0.23	0.61	0.53
3	72.9	0.48	0.15	0.58	0.48
4	82.3	0.60	0.19	0.62	0.60
5	73.0	0.45	0.14	0.45	0.45
6	83.1	0.67	0.19	0.61	0.67
Mean	77.1	0.54	0.17	0.56	0.54

Table 6
Agreement and disagreement between expert and non-expert annotators.

Session	Cohen's Kappa	Number of onset delays	Mean delay in response onset (s)	Number of offset delays	Mean delay in response offset (s)
Lab1	0.25	8	4.1	16	1.3
Lab2	0.25	5	1.8	7	1.1
Classroom1	0.33	5	1.8	0	0
Classroom2	0.375	3	5.3	14	2.5
Mean	0.30	5.25	3.25	9.25	1.23

4.4. Study 1, analysis 3: Performance across different participants

In this analysis we trained the classifier with data from all the participants but one and tested the performance on the left-out participant. This procedure was repeated across all participants and results were averaged across activities. The purpose of this analysis was to directly assess how a general classifier would perform compared to person-dependent training.

The overall performance is relatively low with an average TP rate of 0.54 and an average FP rate of 0.17. The FP rate is dominated by errors associated with the unknown class. There was considerable variability across participants with respect to topography, duration, frequency, and consistency of the movements. This results in overall low performance using participant-independent training. For example, the duration of stereotypical flapping episodes varied from 1 s to several minutes and involved different hand postures and movements across participants.

We also observed that body rocking is more consistent than hand flapping, likely because hand movements have more degrees of freedom than torso movements. Table 5 shows that the best performance on participant-independent training is for participants 4 (Precision 0.62 and Recall 0.60) and 6 (Precision 0.61 and Recall 0.67). Unlike other participants, both engaged primarily in body rocking (82% and 95% of the time they were observed engaging in stereotypical motor movements, respectively). Because body rocking is more consistent across participants (i.e., less variability in how body rocking is performed), the results for these two participants were higher than other cases.

4.5. Study 2: Performance using non-expert annotations

In real-world environments, expert annotators are rarely available, necessitating annotations from non-experts to train classification algorithms. In this pilot study we analyze the quality of non-expert annotations obtained via a mobile phone and assess the impact these annotations have on algorithm performance.

Table 6 shows agreement and disagreement between expert and non-expert annotators. Cohen's Kappa averages 0.33 across all sessions, which indicates fair agreement between expert and non-expert annotations. Interestingly, agreement between annotators was higher in the classroom than in the laboratory. This is encouraging because the training data needed in this domain is likely to be gathered in real-world settings such as classrooms. The non-expert annotator had an average delay of 3.25 s in labeling onsets and 1.23 s in labeling offsets versus the expert. The frequency of delays in the onset (5.25) was much lower than the frequency of delays in the offset (9.25). However, as discussed below, this does not seem to impact performance of the activity recognition algorithm.

Tables 7 and 8 show results for the recognition algorithm when trained on one session and tested on another session using expert annotations and non-expert annotations in both laboratory and classroom settings. All cases are evaluated against expert annotations. The tables report the following metrics: accuracy (A), precision (P), and recall (R) [27]. The average accuracies, precisions, and recalls for the algorithm when training and testing on different sessions using expert annotations (A: 81%, P: 0.79, R: 0.76) and non-expert annotations (A: 80%, P: 0.8, R: 0.79) are similar. The average performance achieved for different sessions seems close for both expert (Lab1: 79%, Lab2: 89%, Classroom1: 82%, Classroom2: 74%) and non-expert annotations (Lab1: 85%, Lab2: 86%, Classroom1: 81%, Classroom2: 69%). It is important to point out that the expert annotator had significantly more distractions during the second classroom session than in other sessions, this is reflected in the performance results when training on classroom2 data; the performance using expert annotations (Lab1 73%, Lab2: 87%, Classroom1 83%) is lower than the performance using non-expert annotations (Lab1 89%, Lab2: 92%, Classroom1 86%). These types of distractions are typical in classroom environments where there are multiple children. However, developing scalable

Table 7
Performance using expert annotations.

Training set	Num examples	Lab1			Lab2			Classroom1			Classroom2		
		A (%)	P	R	A (%)	P	R	A (%)	P	R	A (%)	P	R
Lab1	1291	97	0.97	0.97	93	0.93	0.93	80	0.75	0.70	70	0.70	0.73
Lab2	941	90	0.91	0.86	97	0.96	0.97	84	0.83	0.75	70	0.70	0.71
Classroom1	582	75	0.77	0.70	88	0.88	0.86	95	0.94	0.93	82	0.81	0.80
Classroom2	1445	73	0.78	0.70	87	0.87	0.65	83	0.80	0.74	94	0.94	0.92
Mean	1064	84	0.86	0.80	91	0.91	0.85	85	0.80	0.78	79	0.79	0.79
Mean without training session	989	79	0.82	0.75	89	0.89	0.81	82	0.79	0.73	74	0.74	0.75

Table 8
Performance using non-expert annotations.

Training set	Num examples	Lab1			Lab2			Classroom1			Classroom2		
		A (%)	P	R	A (%)	P	R	A (%)	P	R	A (%)	P	R
Lab1	1327	95	0.95	0.95	91	0.91	0.90	76	0.69	0.61	53	0.61	0.55
Lab2	968	90	0.90	0.89	95	0.94	0.96	81	0.81	0.78	80	0.77	0.79
Classroom1	557	75	0.77	0.75	75	0.79	0.79	92	0.91	0.89	74	0.74	0.75
Classroom2	1554	89	0.89	0.90	92	0.90	0.94	86	0.82	0.84	89	0.90	0.84
Mean	1101	87	0.88	0.87	88	0.89	0.90	84	0.81	0.78	74	0.76	0.73
Mean without training session	1026	85	0.85	0.85	86	0.87	0.88	81	0.77	0.74	69	0.71	0.70

tools for annotation that non-experts can use will enable us to collect larger datasets which in turn reduces the impact of inaccuracies in annotation on the performance of pattern recognition algorithms. Finally, our preliminary results on the quality of non-expert annotation suggest that non-expert, real-time annotation using mobile phones is a viable approach for gathering person-dependent data needed to train accurate pattern recognition algorithms in this domain.

5. Discussion

The problem of accurately recognizing stereotypical motor movements in children with ASD and creating a real-time monitoring tool is more challenging than it may appear at first due to the complexity of the domain. First, there was considerable variability in the topography, duration, frequency, and consistency with which participants performed stereotypical motor movements. Each child had very specific stereotypical motor movements that required participant-dependent data to train the classifier. Second, both real-time and offline annotations were difficult to generate, even by trained experts. The annotators had more difficulty and disagreement in documenting stereotypical motor movements in real-time than offline. In real-time annotation, annotators often missed activity start and stop times and sometimes missed whole activities altogether. In offline annotation, the annotation tool did not account for the uncertainty of the annotator but rather provided discrete markers for the beginning and end of observed stereotypical motor movements. This resulted in noise around the boundaries of each stereotypical motor movement that appears to especially impact recognition of shorter movement segments. Third, it can be difficult to collect enough training data from participants who engage in infrequent stereotypical motor movements during regular school hours. For some of our participants, the data was sparse, with less than 10% of the data representing specific stereotypical motor movements.

A key challenge in this domain versus other activity recognition domains is the problem of acquiring adequate participant-dependent training data in real-life situations without undue burden on the person under observation, a researcher, or caregiver. Training a classifier on agreement data from two annotators produced the best results when the duration of stereotypical motor movements were long, but acquiring such annotation is unrealistic for all but highly controlled research settings. In this case, most of the examples of transitive behavior were eliminated from the training data by virtue of disagreements between the annotators. A more realistic deployment scenario would involve a caregiver who utilizes a real-time annotation tool on a mobile device to record naturally-evoked stereotypical motor movements and capture uncertainty in the annotations. Our preliminary finding that real-time mobile annotation by a non-expert is comparable to expert annotation suggests that this approach may be viable.

Another encouraging and somewhat surprising result is that classifiers trained on real-time annotations performed slightly better than classifiers trained using offline annotations from one annotator for participants with stereotypical motor movements of long duration (7 s or more) who engaged in the behavior frequently. This suggests that transitive and subtle examples typically missed in real-time annotation are not particularly good for training. However, real-time annotation of stereotypical motor movements of short duration misses a significant number of valid episodes that results in worse performance than a single offline annotator. Thus, our results again suggest that a real-time annotation system may be feasible for training classifiers to recognize stereotypical movements of long duration. However, to deploy such a system for stereotypical motor movements of short duration, research efforts are still needed to: (1) Improve the accuracy of real-time annotation, for example using auto segmenting techniques; and (2) Capture the uncertainty in the annotations, particularly on the stereotypical motor movement onset and offset boundaries.

In sum, we observed good average classifier performance across participants and sessions in a naturalistic classroom (Accuracy: 88.6%; TP: 0.85; FP: 0.08), suggesting that Accelerometry and pattern recognition algorithms can be usefully employed to detect stereotypical motor movements in unconstrained settings. We also observed minimal differences in classification performance between expert and non-expert annotations (83% and 85%, respectively), suggesting that real-time, non-expert annotation may be viable for training. We are hopeful that these results can be improved even further with the use of temporal filtering techniques and other graphical model-based supervised algorithms and unsupervised learning techniques.

It is important to emphasize that automatic, real-time detection of stereotypical motor movements made possible using comfortable, miniature wireless sensors could both advance autism research and enable new intervention tools that help children and their caregivers monitor, understand, and cope with these behaviors. For research, our system has the potential to overcome many of the problems Sprague and Newell [10] associate with direct observational methods mentioned in the introduction. Specifically, using acceleration data, pattern recognition algorithms have the potential to accurately document high-speed motor sequences; indicate when a sequence has started and ended; and handle concomitantly occurring stereotypical motor movements. Automating stereotypical motor movement detection in this way could free a human observer to concentrate on and note environmental antecedents and consequences necessary to determine what functional relations exist for this perplexing and often disruptive class of behavior. Real-time recognition tools could also be used to document and alert caregivers when rates of other repetitive motor movements common in individuals with ASD (e.g., self-injurious behavior, pacing, pica) are escalating. For intervention, mobile classifiers could be integrated into a real-time intervention system where real-time training data are provided by caregivers and feedback is provided to participants when stereotypical motor movements are detected to better manage or even reduce the occurrence of these episodes (e.g. using vibro-tactile feedback). Such a system could facilitate efficacy studies of behavioral and pharmacological interventions intended to decrease the incidence or severity of stereotypical motor movements. Finally, we note that the technology used in Study 2 has been optimized to run for 48 h on a single phone charge using 2 sensors with standard Bluetooth technology, and is capable of automatically uploading data every night to a server. Further improvements in battery life are expected as Bluetooth Low Energy (BLE) technology becomes available. This could enable researchers to continuously monitor participants and to run new types of offline analyses. For instance, the system could be deployed to monitor the effects of known and novel interventions over a period of several months to determine how long it takes for an intervention to take effect, how long it lasts, and to what extent it reduces frequency, duration, and/or intensity of episodes. Long-term monitoring may also enable assessments of wearability and comfort over more extended periods of time.

Acknowledgments

The authors thank Autism Speaks and the Nancy Lurie Marks Family Foundation for funding this work. We also thank the children and parents who graciously agreed to participate in this research. The sensors used in Study 1 of this work were made possible by funding from the NSF (grant #0313065). The sensors used in Study 2 of this work were made possible by funding from the NIH/NHLBI (grant U01 HL091737). The contents of this work are solely the responsibility of the authors and do not represent the official views of the NSF or NIH/NHLBI.

References

[1] US CDC, Prevalence of autism spectrum disorders—autism and developmental disabilities monitoring network, United States, 2009.
[2] A.P.A., Diagnostic and Statistical Manual on Mental Disorders, 4th ed., vol. IV-TR, Amer. Psychiatric Publishing, Washington, DC, 2000.
[3] A. Baumeister, R. Forehand, Stereotyped acts, in: Int'l. Rev. of Res. in Mental Retardation: VI., ed., New York, 1973, pp. 55–96.
[4] M.H. Lewis, J.W. Bodfish, Repetitive behavior disorders in autism, Ment. Retard. Dev. Disabil. Res. Rev. 4 (1998) 80–89.
[5] S.J. LaGrow, A.C. Repp, Stereotypic responding: a review of intervention research, Am. J. Ment. Def. 88 (1984) 595–609.
[6] G. Berkson, R.K. Davenport Jr, Stereotyped movements of mental defectives. I. Initial survey, Am. J. Ment. Def. 66 (1962) 849–852.
[7] O.I. Lovaas, A. Litrownik, R. Mann, Response latencies to auditory stimuli in autistic children engaged in self-stimulatory behavior, Behav. Res. Ther. 9 (1971) 39–49.
[8] R.S. Jones, D. Wint, N.C. Ellis, The social effects of stereotyped behaviour, J. Ment. Defic. Res. 34 (1990) 261–268.
[9] C.H. Kennedy, Evolution of stereotypy into self-injury, in: Self-Injurious Behavior: Gene–Brain–Behavior Relationships, Amer. Psych. Assoc., Washington, DC, 2002, pp. 133–143.
[10] R.L. Sprague, K.M.E. Newell, Stereotyped Movements: Brain and Behavior Relationships, Amer. Psych. Assoc., Washington, DC, 1996.
[11] D.A. Pyles, M.M. Riordan, J.S. Bailey, The stereotypy analysis: an instrument for examining environmental variables associated with differential rates of stereotypic behavior, Res. Dev. Disabil. 18 (1997) 11–38.
[12] J.A. Kientz, G.R. Hayes, T.L. Westeyn, T. Starner, G.D. Abowd, Pervasive computing and autism: assisting caregivers of children with special needs, IEEE Pervasive Comput. 6 (1) (2007) 28–35.
[13] L. Bao, S.S. Intille, Activity recognition from user-annotated acceleration data, in: Proc. of PERVASIVE, 2004, pp. 1–17.
[14] J. Lester, T. Choudhury, N. Kern, G. Borriello, B. Hannaford, A hybrid discriminative/generative approach for modeling human activities, in: Proc. of IJCAI, 2005, pp. 766–722.
[15] P. Lukowicz, J.A. Ward, H. Junker, M. Stager, G. Troster, A. Atrash, T. Starner, Recognizing workshop activity using body worn microphone and accelerometers, in: Proc. of PerCom, 2004, pp. 18–32.
[16] L. von Ahn, L. Dabbish, Labeling images with a computer game, in: ACM Conference on Human Factors in Computing Systems, CHI 2004, 2004.
[17] L. von Ahn, M. Kedia, M. Blum, Verbosity: a game for collecting common-sense knowledge, in: ACM Conference on Human Factors in Computing Systems, CHI Notes 2006, 2006.

[18] R. Snow, B. O'Connor, D. Jurafsky, A. Ng, Cheap and fast—but is it good? Evaluating non-expert annotations for natural language tasks, in: The Proceedings of EMNLP-08, January 2008.
[19] Aniket Kittur, Crowdsourcing user studies with mechanical turk, in: H. Chi and Bongwon Suh (Eds.), Proc. of CHI, 2008, pp. 453–456.
[20] F. Albinali, M. Goodwin, S. Intille, Recognizing stereotypical motor movements in the laboratory and classroom: a case study with children on the autism spectrum, in: Ubicomp 2009.
[21] T. Westeyn, K. Vadas, X. Bian, T. Starner, G.D. Abowd, Recognizing mimicked autistic self-stimulatory behaviors using HMMs, in: Proc. of ISWC, 2005, pp. 164–169.
[22] J.W. Bodfish, F.J. Symons, D.E. Parker, M.H. Lewis, Varieties of repetitive behavior in autism: comparisons to mental retardation, J. Autism. Dev. Disord. 30 (2000) 237–243.
[23] E. Munguia Tapia, S.S. Intille, L. Lopez, K. Larson, The design of a portable kit of wireless sensors for naturalistic data collection, in: Proc. of PERVASIVE, 2006, pp. 117–134.
[24] The Wockets Project, Wockets: open source accelerometers for phones. http://web.mit.edu/wockets/ (accessed on: 15.01.10).
[25] J.R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo, California, 1993.
[26] Rulequest research, See5/C5.0. http://www.rulequest.com/ (accessed on: 10.12.10).
[27] I.H. Witten, E. Frank, Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann, San Francisco, CA, 1999.

Fahd Albinali, Ph.D., is the Chief Technology Officer of EveryFit, Inc., a start-up exercise fitness and measurement company he co-founded in 2010. He was previously a Research Scientist at the House_n Consortium in the MIT Department of Architecture. His research focuses on building and studying interactive technologies that measure behavior. Currently, he is working on building wearable systems that measure physical activity to address high value societal challenges such as preventive health care and support for aging and disabled populations. His work has been published in academic venues including UbiComp, AAAI, CHI and PerCom and has received one best paper award. Dr. Albinali received his Ph.D. from the University of Arizona in 2008 working on activity recognition in domestic environments, an M.Sc. from the University of Arizona in 2002, and a B.Sc. degree in Computer Science from the American University in Cairo in 1999.

Matthew S. Goodwin, Ph.D., received his B.A. in Psychology from Wheaton College in 1998 and his M.A. in 2008 and Ph.D. in 2010, both in Experimental Psychology from the University of Rhode Island. He is currently the Director of Clinical Research at the Massachusetts Institute of Technology, Media Laboratory and Associate Director of Research at the Groden Center—an Institute for Autism Spectrum Disorders in Providence, RI. He is Co-Chair of the Autism Speaks—Innovative Technology for Autism Initiative, has an Adjunct Associate Research Scientist appointment in the Department of Psychiatry and Human Behavior at Brown University, and is an Adjunct Assistant Professor in the Department of Psychology at the University of Rhode Island. He has over 15 years of research and clinical experience working with the full spectrum of children and adults with ASD and has extensive experience developing and evaluating innovative technologies for behavioral assessment, including telemetric physiological monitors, accelerometry sensors, and digital video/facial recognition systems.

Stephen Intille, Ph.D., is an Associate Professor in the College of Computer and Information Science & Dept. of Health Sciences, Bouvé College of Health Sciences at Northeastern University. His research is focused on the development of novel healthcare technologies that incorporate ideas from ubiquitous computing, user-interface design, pattern recognition, behavioral science, and preventive medicine. One area of special interest to him is the development and pilot testing of systems that support healthy aging and well-being in the home setting. Another area of interest is the creation of tools for mobile phones that permit longitudinal measurement of health behaviors for research, especially the type, duration, intensity, and location of physical activity, and sensor-enabled, mobile tools that motivate and assist people in making healthy behavior changes. Dr. Intille received his Ph.D. from MIT in 1999 working on computational vision at the MIT Media Laboratory, an S.M. from MIT in 1994, and a B.S.E. degree in Computer Science and Engineering from the University of Pennsylvania in 1992. He has published research on computational stereo depth recovery, real-time and multi-agent tracking, activity recognition, perceptually-based interactive environments, and technology for healthcare. Dr. Intille has been principal investigator on sensor-enabled health technology grants from the NSF, the NIH, foundations, and industry. In September, 2010 he joined Northeastern University to help establish a new transdiciplinary Ph.D. program in health informatics/technologies.