

## Chapter 3

---

### Egocentric Sensing of body movements

---

As explained in Chapter 2, people who are blind and visually impaired consider their body mannerisms to be an important component of their social interactions with their sighted counterpart. They are specifically worried about the display of any stereotypic body mannerisms that could potentially become a social distraction and hindrance for them. In Chapter 1, a detailed discussion of such stereotypic body mannerisms is presented and one specific quality of stereotypic body rocking was discussed in detail as body rocking is the most widely studied stereotypy in visually impaired population. In this chapter we introduce the concept of using egocentric motion sensing towards detection such stereotypic behaviors. While the discussion is limited to body rocking, similar concepts can be used to extend the rehabilitation to other body mannerisms.

Recently, human activity detection and recognition using motion sensors have taken a front seat in technology and behavioral research. This is due to the availability of micro mechanized electronic systems (MEMS) that have started to implement complex mechanical systems at a micro scale on integrated circuit chips. These offers advantages like reliability, cheaper cost of production, smaller form factor and above all extremely precise measurement with least or no maintenance. One such sensor is the accelerometer that is capable of measuring the effect of gravity on three perpendicular axes. When mounted on any moving object, the opposing motion (opposing gravity) of the entity allows these sensors to measure the speed and direction of motion. Integrating the magnitude and orientation information over time it is possible to accurately measure the exact motion pattern of the moving entity. These accelerometers have been used by researchers to track motion activity in almost every joint of the human body. Researchers have used single, double or triple orthogonal axis accelerometers to detect various activities of humans. They all follow the same underlying supervised learning architecture with difference in learning algorithm used. A simplified representation of the same is shown in Figure 1.

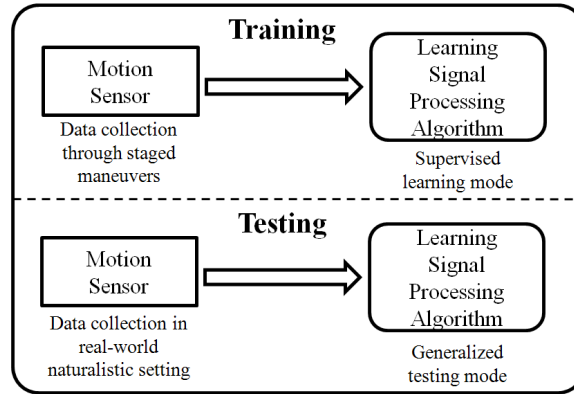


Figure. 1: Training and testing phases of a typical learning framework found in literature.

In a similar framework, we use these motion sensors that are placed on strategically located places on the human body to detect body rocking behaviors.

### 3.1 The Hardware:

In our application, in order to keep the motion detector discrete, we have chosen state-of-the-art tri-axial accelerometer package, ZStar III, marketed by Freescale Semiconductor. The accelerometer is shown in the inset of Figure 2. The device (including a coin battery as a power source) is an inch in diameter and less than eighth of an inch in thickness thereby allowing an elegant integration into everyday clothing. Figure 2 shows the typical use of the accelerometer in the proposed application for detecting body rocking. The accelerometer has a very high sensitivity with protection against excessive g-force damage. The sensors wirelessly connect to a PDA and/or cell phone through IEEE 802.15.4 (ZigBee) wireless standards. The use of low power consumption electronics for both acceleration sensing and wireless communication allows this device to work for hours at length on a single coin battery. Further, the advanced sleep mode implementations allow the device to stay at nano watt power mode during non-operation. The proposed solution allows for prolonged use of the device to the effect of an assistive technology thereby maintaining a longer duration feedback based rehabilitation regimen.

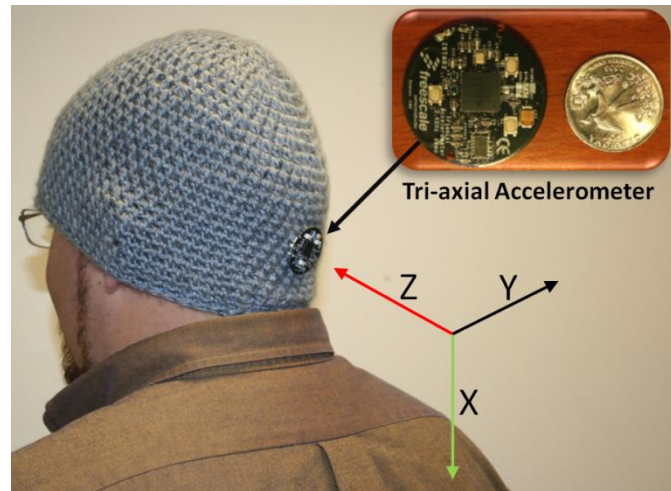


Figure 2. The proposed hardware for use in the detection of body rocking stereotypic behavior. The accelerometer, in comparison with a US quarter, is shown in the inset. The three axes marked in the image shows the orientation of the accelerometer as it is placed on the head.

The processing element for the current study was a Windows Mobile Operating System based PDA running on a 400Mhz XScale processor. The software components (described in detail in Section 3.2 of the proposed solution) were placed on the PDA that could be carried by a user without any extra load. The software component implementation is generic to be ported to most modern cell phones that possess enough processing power, but is always underutilized for its capacity. The feedback (an audio tone) is currently being provided through a Bluetooth headset that is paired with the processing element. The choice of this feedback device was again based on the idea that Bluetooth headset has everyday acceptance among the masses and is no longer seen as a social distraction. In future, we plan to explore the use of delivery modalities that transcends the typical visual and audio medium. We intend to use haptic cues to inform the participant not only their rocking behavior but more complex self-monitoring routines that could allow the user to withdraw from the rocking behavior effectively.

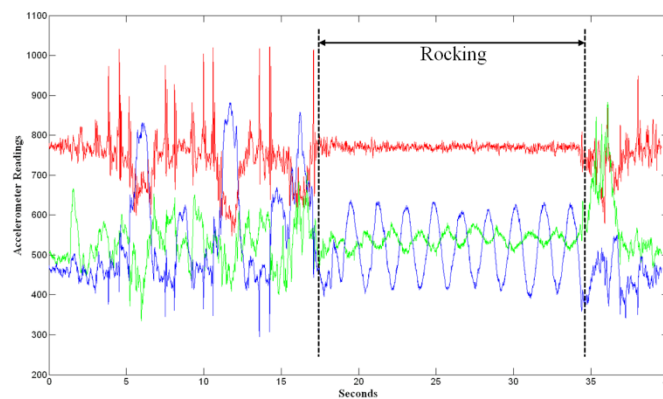


Figure 3. Data stream for the tri-axial accelerometer. The three streams correspond to the three axes. The figure shows non-rocking events followed by rocking and then followed by non-rocking.

Figure 3 shows a typical data stream collected from the accelerometer shown in Figure 2 during rocking and non-rocking functional behavior. The three data streams correspond to the three axis of the accelerometer each sampled at 100 Hz. It can be seen that the data stream under rocking conditions are visually distinguishable when compared to non-rocking functional movements. The following section highlights our choice of learning framework and features we extracted from these data stream in order to achieve reliable rocking and non-rocking discrimination.

### 3.2 Extracting Body Rock Information from Motion Sensor Data

The work presented in this paper builds on top of the work presented in [1] where the authors use two accelerometers placed one at the ankle and the other on the thigh to distinguish between simple activities like walking, running, standing etc. They proved the use of an aggregated AdaBoost classifier system that was built out of simple linear classifiers to achieve activity recognition. Unfortunately, the work does not provide any assessment on the generalization capabilities of their aggregate classifier. We extend their work into the problem of body rock detection using only one accelerometer placed on the back of the person's head. Below, we discuss the various features that we extract from the accelerometer data and introduce the variant of AdaBoost that generalizes on its training set very well (termed Modest AdaBoost). We show results of our experiments and discuss our reasoning to believe how the new AdaBoost framework is able to generalize on body rocking data when compared to classical AdaBoost used by [1].

#### 3.2.1 Features:

Since we are using a tri-axial accelerometer, we obtain three orthogonal axis data through rocking and non-rocking events. In order to capture the temporal variation in the acceleration data, we accumulate the input stream on each axis for a fixed duration  $T$  seconds and all features are extracted on this packet of acceleration data. As a part of the assessment, we determine the best packet length for the task of body rock detection. Further, successive packets are extracted with a fixed duration of overlap between them.

We chose five sets of features that were extracted on the three axes of accelerometer data. For the sake of clarity, we cluster these sets into two groups based on whether they were chosen due to popular use in the accelerometer data processing community or due to the author's insights into the body rocking data.

##### 3.2.1.1 Group 1 – Popular features used by the motion analysis research community [2] [1]:

We choose the following three sets each of which were applied on all three axes of acceleration data, henceforth referred to as x, y, z axis data.

1. Mean of x, y, z data over the duration of packet.
2. Variance of x, y, z data over the duration of packet.
3. Correlation between the three axes (x-y, y-z and z-x) over the duration of packet.

### 3.2.1.2 Group 2 – Authors insights into body rocking data:

Inspecting the accelerometer data shown in Figure 3, it can be seen that the Z axis changes from random signal pattern to more of a sinusoidal pattern when the individual's behavior transitions from non-rocking to rocking. Thus we choose two sets of features which we hope would capture this non-sinusoid to sinusoid transition between events. These features include

4. The first order differential power on all three axes – Sinusoidal signals change gradually over time such that the averaged sum square energy in the temporal first order differential of the signal should be less when compared to a random signal where the first order differential can have very high variations and hence higher power.
5. Fourier Transform variance and kurtosis on the Z-axis only – An effective way to capture power distribution of a signal into sinusoids is by using Fourier Transform. We hypothesize that the non-sinusoid to sinusoid transitions can be captured by quantitatively measuring the power spread spectrum of the Z-axis accelerometer data. We model the power spread to be a Gaussian and extract the variance and kurtosis (peaking) of the spread to determine if there is rocking or not.

Thus, the features used in our study can be categorized as belonging to two groups with three sets in Group 1 and two sets in Group 2. Each set has varying number of features based on what parameter the set is extracting from the temporal accelerometer data. Based on the descriptions above, the entire feature set has a total of 14 features. We identify each of these by their respective Feature Identification Numbers. Table 1 shows the two groups and the different sets under the group with typical values of these features under rocking and non-rocking behavior.

Group 1															
<b>Set 1</b>  <i>Definition:</i> Mean on the temporal dimension.  <i>Axes affected:</i> x, y, z  <i>Number of contributing features:</i> 3  <i>Feature Identification Numbers:</i> 1, 2, 3	$M_x = \frac{1}{N} \sum_{i=1}^N x_i$	1. $M_x$	<table border="1"><caption>Mean Feature Values</caption><thead><tr><th>Feature</th><th>P_pos</th><th>P_neg</th></tr></thead><tbody><tr><td>M_x</td><td>~750</td><td>~750</td></tr><tr><td>M_y</td><td>~550</td><td>~550</td></tr><tr><td>M_z</td><td>~450</td><td>~450</td></tr></tbody></table>	Feature	P_pos	P_neg	M_x	~750	~750	M_y	~550	~550	M_z	~450	~450
		Feature		P_pos	P_neg										
		M_x		~750	~750										
M_y	~550	~550													
M_z	~450	~450													
2. $M_y$															
3. $M_z$															
<b>Set 2</b>  <i>Definition:</i> Variance on the temporal dimension  <i>Axes affected:</i> x, y, z  <i>Number of contributing features:</i> 3  <i>Feature Identification Numbers:</i> 4, 5, 6	$V_x = \frac{1}{N-1} \sum_{i=1}^N (x_i - M_x)^2$	4. $V_x$	<table border="1"><caption>Variance Feature Values</caption><thead><tr><th>Feature</th><th>P_pos</th><th>P_neg</th></tr></thead><tbody><tr><td>V_x</td><td>~1000</td><td>~2000</td></tr><tr><td>V_y</td><td>~1000</td><td>~2000</td></tr><tr><td>V_z</td><td>~2500</td><td>~3000</td></tr></tbody></table>	Feature	P_pos	P_neg	V_x	~1000	~2000	V_y	~1000	~2000	V_z	~2500	~3000
		Feature		P_pos	P_neg										
		V_x		~1000	~2000										
V_y	~1000	~2000													
V_z	~2500	~3000													
5. $V_y$															
6. $V_z$															
<b>Set 3</b>  <i>Definition:</i> Cross correlation between axes  <i>Axes affected:</i> x, y, z  <i>Number of contributing features:</i> 3  <i>Feature Identification Numbers:</i> 7, 8, 9	$C_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - M_x)(y_i - M_y)$	7. $C_{xy}$	<table border="1"><caption>Correlation Coefficient Feature Values</caption><thead><tr><th>Feature</th><th>P_pos</th><th>P_neg</th></tr></thead><tbody><tr><td>C_xy</td><td>~-0.2</td><td>~0.1</td></tr><tr><td>C_yz</td><td>~0.4</td><td>~-0.1</td></tr><tr><td>C_xz</td><td>~-0.2</td><td>~0.0</td></tr></tbody></table>	Feature	P_pos	P_neg	C_xy	~-0.2	~0.1	C_yz	~0.4	~-0.1	C_xz	~-0.2	~0.0
		Feature		P_pos	P_neg										
		C_xy		~-0.2	~0.1										
		C_yz	~0.4	~-0.1											
C_xz	~-0.2	~0.0													
8. $C_{yz}$															
9. $C_{xz}$															
14. $F_{kz}$	<table border="1"><caption>FFT kurtosis Feature Values</caption><thead><tr><th>Feature</th><th>P_pos</th><th>P_neg</th></tr></thead><tbody><tr><td>F_x</td><td>~5.5</td><td>~5.0</td></tr><tr><td>F_z</td><td>~5.0</td><td>~5.0</td></tr></tbody></table>	Feature	P_pos	P_neg	F_x	~5.5	~5.0	F_z	~5.0	~5.0					
Feature	P_pos	P_neg													
F_x	~5.5	~5.0													
F_z	~5.0	~5.0													

Table 1: Group 1 Features. Most popularly used features for extracting motion information. The figures shown in the last column plots mean values of data from positive rocking samples and negative rocking samples as bars. The variance on the same is shown as vertical error lines around the mean. The lighter (blue when viewed in color) shaded bar are values from the positive class, whereas the darker (pink when viewed in color) bar are values from the negative class.

Group 2			
<b>Set 4</b>  <b>Definition:</b> First order differential power.  <b>Axes affected:</b> x, y, z  <b>Number of contributing features:</b> 3  <b>Feature Identification Numbers:</b> 10, 11, 12	$D_x = \sqrt{\sum_{i=2}^N (x_i - x_{i-1})^2}$	10. $D_x$	
		11. $D_y$	
		12. $D_z$	
<b>Set 5</b>  <b>Definition:</b> Gaussian fit power spread spectrum – Variance and Kurtosis  <b>Axes affected:</b> z  <b>Number of contributing features:</b> 2  <b>Feature Identification Numbers:</b> 13, 14	<p>If, <math>Freq_k = \left\{-\frac{\gamma}{2}, \dots, 0, \dots, \frac{\gamma}{2}\right\}</math>, <math>\gamma</math> is the sampling frequency, and</p> $X_k = \sum_{i=1}^N x_n e^{\frac{2\pi i k n}{N}}, k = \{1, \dots, N\},$ <p>then</p> <p><b>FFT Variance:</b> <math>F_v = \sum_{i=1}^N X_k (Freq_k)^2</math></p> <p><b>FFT Kurtosis:</b> <math>F_k = \sum_{i=1}^N X_k (Freq_k)^4</math></p>	13. $F_{v_z}$	
		14. $F_{k_z}$	

Table 2: Group 2 features that were specifically extracted based on the nature of body rocking data. The figures shown in the last column plots mean values of data from positive rocking samples and negative rocking samples as bars. The variance on the same is shown as vertical error lines around the mean. The lighter (blue when viewed in color) shaded bar are values from the positive class, whereas the darker (pink when viewed in color) bar are values from the negative class.

### 3.2.2 Learning Algorithm:

As discussed in introduction of this section, we compare the performance of two AdaBoost learning frameworks to determine which one can generalize the best on the training data. The two algorithms are introduced briefly below. For further details, the reader is referred to appropriate references provided within the subsections.

### 3.2.2.1 Classic AdaBoost Learning Framework:

AdaBoost learns any classification problem by working with a set of weak classifiers. Weak classifiers are those classifiers that use simple decision steps to categorize data into one of two pools – positives or negatives (In all our experiments, we used a three level decision tree [3] as the simple classifier). AdaBoost proceeds by ranking the labeled training data as being simple to complex based on how many weak classifiers are needed to learn each of the examples. The process continues on an iterative manner until all the training examples are learnt or till the allowed number of learning cycles are exhausted. Let,  $X$  be the input to a learning algorithm, in our case the features extracted as explained in the previous step, and  $Y$  be the label of what class the data belongs to, in our case,  $Y = \{1, -1\}$  implying {rocking, non-rocking}, respectively. Values at each dimension of input  $X$  can be considered to characterize the incoming data in some manner and the task of the learning algorithm is to learn these representational values of the input dimensions that allow the algorithm to distinguish between rocking and non-rocking. AdaBoost does this learning by using a large set of simple (weak) learners (or classifiers) that act on each of the dimension of the input data with the determined goal of distinguishing rocking from non-rocking. The final decision of the complete learning module is a combined opinion of all the simple learners that make up the system. The beauty of AdaBoost implementation is that the human intervention into the learning process stops at identifying what simple (weak) learners to use and what feature pool to operate on. Selection of number of weak learners, selection of input dimension on which the weak learners have to act, and the confidence to place on the decision of each of the weak learner is all determined by the algorithm during the training phase. Once the algorithm is trained, the final learnt rocking/non-rocking classifier can be represented as

$$L(x) = \text{sign} \left[ \sum_{i=1}^N w_i f_i(x) \right]$$

where,

$x$ : An instance of all possible rocking patterns  $X$ .

$L$ : The final learnt classifier that can distinguish input  $x$  as rocking or non-rocking.

$f$ : The simple (weak) learner.

$N$ : The total number of weak learners that make up the complete learner  $L$ .

$w$ : Weight associated with each weak learners output. This corresponds to the confidence placed in each weak learner by the Boosted system.

From a learning perspective, in each step of the iterative learning, the AdaBoost algorithm implements a greedy optimization to pick a set of weak learners that minimize exponential classification error of the picked simple classifiers as shown below



$$Error_k = \sum_{i=1}^M e^{-y_i L(x_i)}$$

where,

$y$ : Label of the input instance  $x$

$M$ : Total number of examples in the training set

$k$ : Learning iteration number

Further, based on each iterative step, a distribution ( $D_m$ ) is created over the training set examples to represent their complexity (difficulty to learn). For example, in a given iteration, an example that could be solved is assigned a lower distribution weight while, a sample that was not learnt in that iteration step is assigned a higher weight. The lower weight on the learnt example implies that this example will be stressed less in the next learning iteration while all other examples which could not be solved will become the focus for picking new weak learners. Moving from one iteration to the next, all the weak learners from the past  $k$  iterations are added into a pool of selected weak learners leading up to the final classifier  $L$ .

### 3.2.2.1 Modest AdaBoost

All learning algorithms, including AdaBoost suffer from the problem of over fitting or over learning. This is due to the fact that training sample sets of positives and negatives can never be representative of all the possible samples that the algorithm will face in its operational life span. Since the learning is limited to a restricted set of examples, there is always the problem of over fitting into this small set and thereby loosing the ability to generalize their learnt knowledge to all other possible examples. To this end, many alternatives have been proposed to AdaBoost that will allow the algorithm to generalize better. We introduce Modest AdaBoost [4] which was recently proposed towards better generalization capabilities and has been shown to be powerful on various machine learning datasets. Unlike the classic AdaBoost where the distribution penalizes only examples that are not learnt in the previous iteration, Modest AdaBoost penalizes for examples that are not learnt and also examples that are learnt very well (over fitting). This is done by projecting all the examples in the training pool on to four separate distributions,

1.  $P_m^{+1} = P_{D_m}(y = +1 \cap L(x)) \rightarrow$  Probability of the learner, as measured on  $D_m$ , predicting an input instance  $x$  correctly as being rocking when the label also represents it to be rocking.
2.  $P_m^{-1} = P_{D_m}(y = -1 \cap L(x)) \rightarrow$  Probability of the learner, as measured on  $D_m$ , predicting an input instance  $x$  correctly as being non-rocking when the label also represents it to be non-rocking.
3.  $\bar{P}_m^{+1} = P_{\bar{D}_m}(y = +1 \cap L(x)) \rightarrow$  Probability of the learner, as measured in the inverse distribution ( $\bar{D}_m$ ), predicting an input instance  $x$  correctly as being rocking when the label also represents it to be rocking.

4.  $\bar{P}_m^{-1} = P_{\bar{D}_m}(y = -1 \cap L(x)) \rightarrow$  Probability of the learner, measured in the inverse distribution ( $\bar{D}_m$ ), predicting an input instance  $x$  correctly as being rocking when the label also represents it to be rocking.

Conditions 1 and 2 penalize the classifier on examples that are not learnt during a training iteration, whereas 3 and 4 penalize examples that are already learnt in the previous iteration which was learnt again in the current iteration. Combining these four measures as

$$f_m = (P_m^{+1}(1 - \bar{P}_m^{+1}) - P_m^{-1}(1 - \bar{P}_m^{-1}))(x)$$

provides a means for penalizing the learner for not classifying an example and also for over fitting an example. This provides a means for modest learning of the final combined classifier  $L$ .

We hypothesize that the choice of a learning algorithm that generalizes well will provide the opportunity to allow better non-rocking detection thereby hopefully increasing discrimination ability for the assistive device. This would directly reflect upon the motivation of the user to get feedback only when he/she is rocking and not performing other functional activities.

### 3.3 Data Collection

Two separate data collections were carried out, one in a controlled setting while the other in a more uncontrolled naturalistic everyday research laboratory setting. The controlled setting data collection was used for training and lab testing the device, whereas the uncontrolled naturalistic setting was used to determine how well the learning algorithm was able to generalize when used for an extended period of time as an assistive tool.

#### 3.3.1 Controlled Data Collection:

Data was collected on ten participants who did not have any known stereotype rocking behavior. The goal of the experiments was to collect data for training the system to differentiate rocking from non-rocking behavior. To this end, we devised three separate data collection routines where the subjects were required to do rocking and non-rocking tasks as naturally as possible. The details of the routines are as follows:

##### 3.3.1.1 Routine A:

Rocking data: Participants were allowed to choose from a rocking chair or a stool or sitting on the ground, so they could rock as comfortably and naturally as possible. We found some cultural preferences to the way people choose to rock. The subjects were asked to rock for a total of 20 complete cycles.

##### 3.3.1.2 Routine B:

Non-rocking data: The participants were asked to do activities that did not involve rocking. They moved around the experimental setup reading posters, operating computers, interacting with everyday office equipments and included some functional body motions similar to rocking like,

stooping down to pick up objects, rapidly bending down to pick up objects etc. Data was collected for a total of 30 seconds.

### 3.3.1.3 Routine C:

Test data: Since rocking can happen at any given instance, we collected data where subjects did various activities and interspersed them randomly with rocking. The goal is to determine how fast and accurately our system can detect such rocking occurrences. In all of these data streams, rocking instances were manually identified and marked for the sake of ground truth. Figure 3 shows the combination of rocking and non-rocking activities by the participants. It can be noticed that there is a clear demarcation between the two activity zones.

### 3.3.2 Uncontrolled Data Collection:

The uncontrolled data was collected towards testing the generalization capabilities of the learnt system. To this end, the body rock detection system was worn by the primary author during everyday laboratory activities. Body rock detection was provided as a feedback through a pair of headphones in the form of an audio beep. Five trails of four separate ten minute data collections were done. Two of the four were done with classic AdaBoost whereas the other two were done with Modest AdaBoost. Further, under each of these two classifiers, one data collection measured how many false positives were detected, whereas the second data collection counted how many rocking actions went undetected. During all these data collection the researcher counted the number of false positive or false negatives using a handheld thumb counter. This experiment was conducted purely to test the generalization capability of the learnt classifier.

## 3.4 Experiments

Experiments were carried out for comparing the performance of the classic AdaBoost framework with Modest AdaBoost for the specific tasks of determining

- a. The length of a temporal packet of data needed to effectively distinguish rocking from non-rocking.
- b. The accuracy with which the two classifiers can distinguish between rocking from non-rocking.
- c. The generalization capabilities of the two classifier systems.

To this end the rocking samples collected in Routine A (discussed under Section 3.3.1.1) and Routine B (discussed under Section 3.3.1.2) were used as labeled positive (rocking) and negative (non-rocking) data for training the AdaBoost classifiers. Data collected under Routine C (discussed in Section 3.3.1.3) were used for testing the learnt classifiers. The results from this analysis were used for determining a. and b. above. We varied the packet length on the data stream and determined the recognition rate on the test data. While the packet length was varied, a constant overlap was maintained between successive packets. This overlap was determined empirically to be 0.5 seconds or 50 samples (@ 100 Hz sampling rate). With the ground truth already provided for the test set, we were able to determine the accuracy of the two classifiers.

To determine c., we resorted to using the data collected in Section 3.3.2. The primary author of the paper used the device to collect false positive and false negative data in order to determine how well the classifiers generalized on the training data. Further, we analyzed the working of the two classifiers in a piece wise manner by breaking down the features into individual sets (Sets 1 through 5 as identified in Table 1 and Set 6 that included all 14 features) and understanding the functional ability of the classifiers under individual feature sets. This allowed for an in-depth analysis of the workings of the two classifiers. In Section 3.6, we discuss the generalization capability of the two classifiers by heuristic analysis of the piecewise operational modes.

All our experiments were carried out with the aid of the AdaBoost Matlab library developed by Graphics and Media Lab at the Dept. of Computer Science at Moscow State University [5].

### 3.5 Results

Figures 4 and Figure 5 shows the box plot [6] of packet length ( $T$  secs) versus recognition rate for classic AdaBoost and Modest AdaBoost frameworks, respectively. The abscissa represents the length of the data stream (in seconds) used for the analysis, while the ordinate represents the recognition rate. Training and testing were all carried out on the data collected as depicted in Section 3.3.1. The horizontal line inside the box represents the median (second quartile) of recognition rates over the ten subject's data. The lower end of box presents the first quartile (25 percentile) and the upper end of the box represents the third quartile (75 percentile). Thus the box surrounds the center 50 percentile ranges of recognition results. This box is also called the Inter-Quartile Range ( $IQR = \text{third quartile} - \text{first quartile}$ ). The dotted extremity represents the minimum and maximum recognition rate under a certain packet length among the ten subjects. Any outlier (an outlier is greater than 1.5 IQR from the median in any direction) is marked by an asterisk.

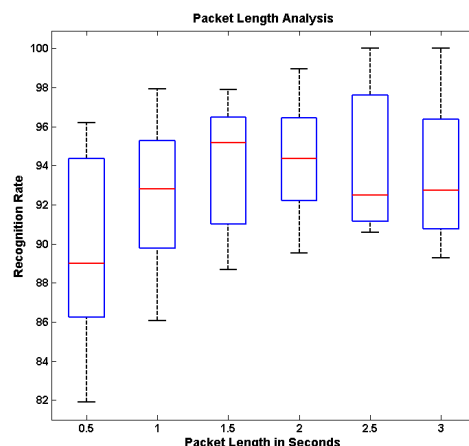


Figure. 4. Packet length to recognition rate comparison under the classic AdaBoost framework.

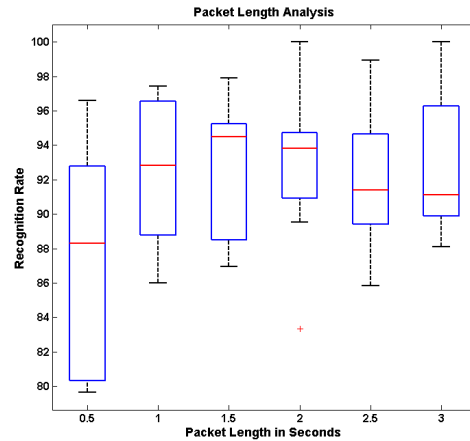


Figure. 5. Packet length to recognition rate comparison under the Modest AdaBoost framework.

Table 2 presents the results from the experiment carried out to determine the generalization capabilities of the two classifiers. The entries in the table are counts as measured by the researchers of the number of false positives and false negatives counted manually while using the device for body rock detection and feedback. Five trails were carried out of 10 minutes each for determining these numbers. False positives represent the number of times the device falsely gave feedback when the user was not involved in rocking. It is important that this rate be minimal as too many false feedbacks would be discouraging for the user to continue using the assistive aid. The false negative represents the number of times the device did not detect that the user was rocking. This metric could be correlated to the failure of the device to perform its functional task.

TABLE 2  
EXPERIMENTS ON NATURALISTIC DATA

Generalization Capabilities	Classic AdaBoost	Modest AdaBoost
<b>False Positives per Minute</b> – Number of false feedback in ten minute <sup>†</sup>	86	44
<b>False Negatives per Minute</b> – Number of times rocking was not detected in ten minute <sup>†</sup>	20	9
<sup>†</sup> These numbers were averaged over 5 trails of 10 minute each		

Figure 6 and Figure 7 shows the piecewise analysis of the classic AdaBoost and Modest AdaBoost frameworks. Subfigure (a) shows the performance of each feature set considered one at a time in detecting body rocking; feature set 6 corresponds to the use of all 14 features together. For example, column 1 in Figure 6 (a) represents the recognition performance using only temporal mean along x, y and z axis tested on all ten subjects. The bar graph in (a) shows the mean performance rate while the superimposed box plot shows the performance at first, second and third quartile as discussed earlier.

Subfigure (b) represents the Receiver Operating Characteristics (ROC) [7] for the same six feature sets as in subfigure (a). ROC is plotted a false positive rate (FPR) versus true positive rate (TPR). The better the performance, the curve moves towards the (1,1) co-ordinate. For example, in Figure 6 (b) Set 6 with all features is performing better than feature set 1 as Set 6 curve is closer towards (1,1) while the feature set 1 curve is almost along the diagonal of the plot. The diagonal of the ROC plot represents a recognition rate of 50% i.e. random pick.

Subfigure (c) is a derivate of the ROC curves in subfigure (b). Each bar in the graph is representing the area under the corresponding curve (AUC) in (b). An area of 1 represents an ideal classifier with no false positive or false negatives, while an area of 0.5 represents randomness in the classifier output. AUC can be used to immediately determine the curve with the best performance.

Subfigure (d) is an understanding of how the aggregated AdaBoost classifier is built. As discussed above, AdaBoost classifier uses a collection of simple classifiers to achieve the final classifier. We plotted the number of times a particular feature is being used by the aggregate classifier. Further, the features are grouped into 5 sets corresponding to the five feature sets identified in Table 1. Columns belonging to the same set have the same top count which corresponds to the total simple classifiers used from that set. Each column within the set represents how many classifiers are used on each feature within that set. The count on the individual feature is represented by the bottom color along each column. For example, consider set 4 in Figure 7 (d), features with identification number 10, 11 and 12 form this set (corresponding to the first order differential power from x, y and z axis of the accelerometer data) and have a top count of 646 simple classifiers. Within the group, the z axis differential power dominates the other two by having a count of 374.

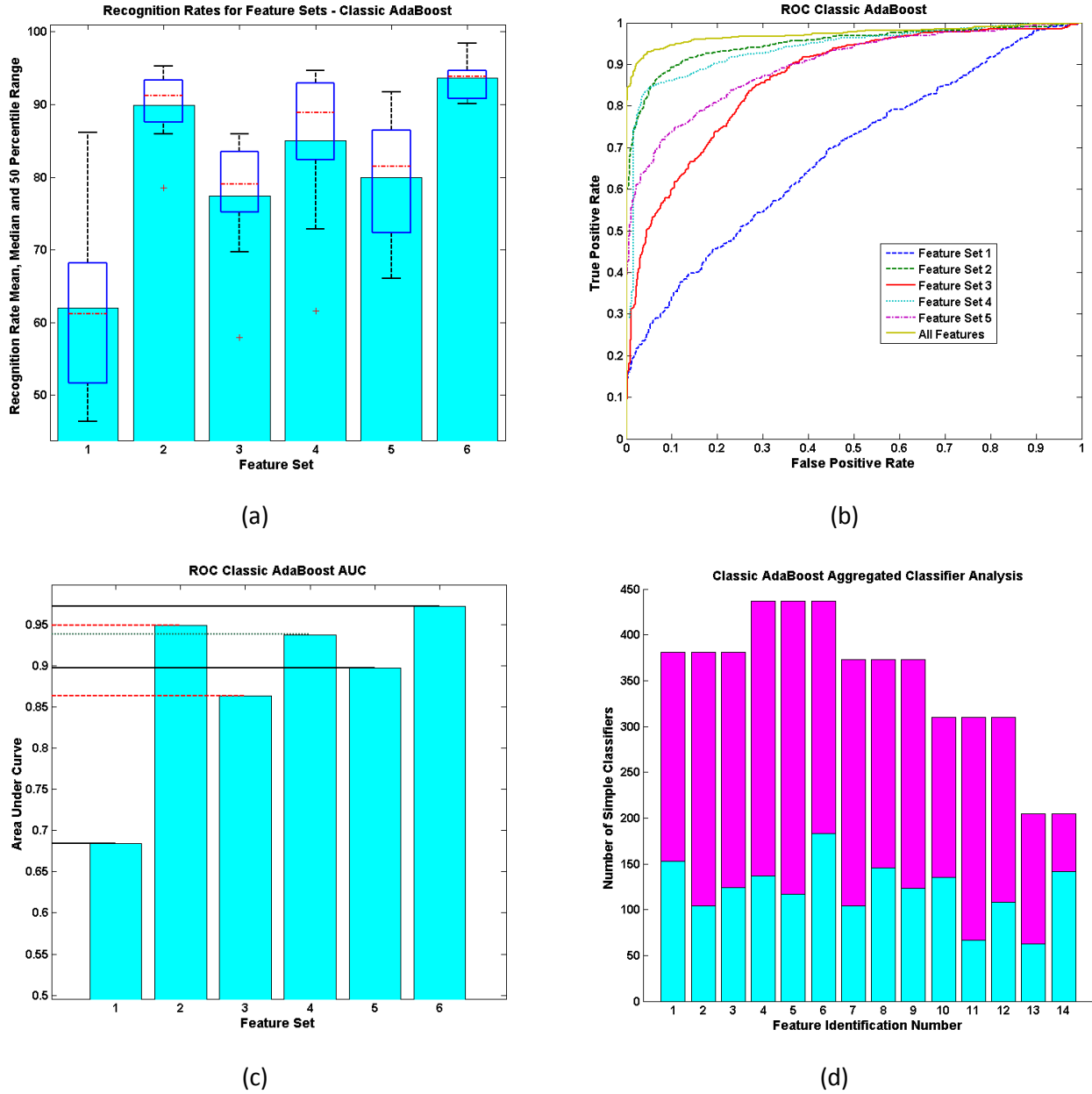


Figure 6: Piecewise performance analysis of the classic AdaBoost classifier framework. (a) Recognition rates under use of individual feature sets. (b) The Receiver Operating Characteristics (ROC) under the use of individual feature sets. (c) Area under the curve (AUC) for each feature set as estimated from the ROC. (d) The number of simple classifiers used by the aggregated AdaBoost classifier. Each set and each feature representation in the classifier pool are separately marked. In all the graphs Set 1 through 5 are as explained by Table 1. Set 6 represents a set containing all 14 features from Table 1.

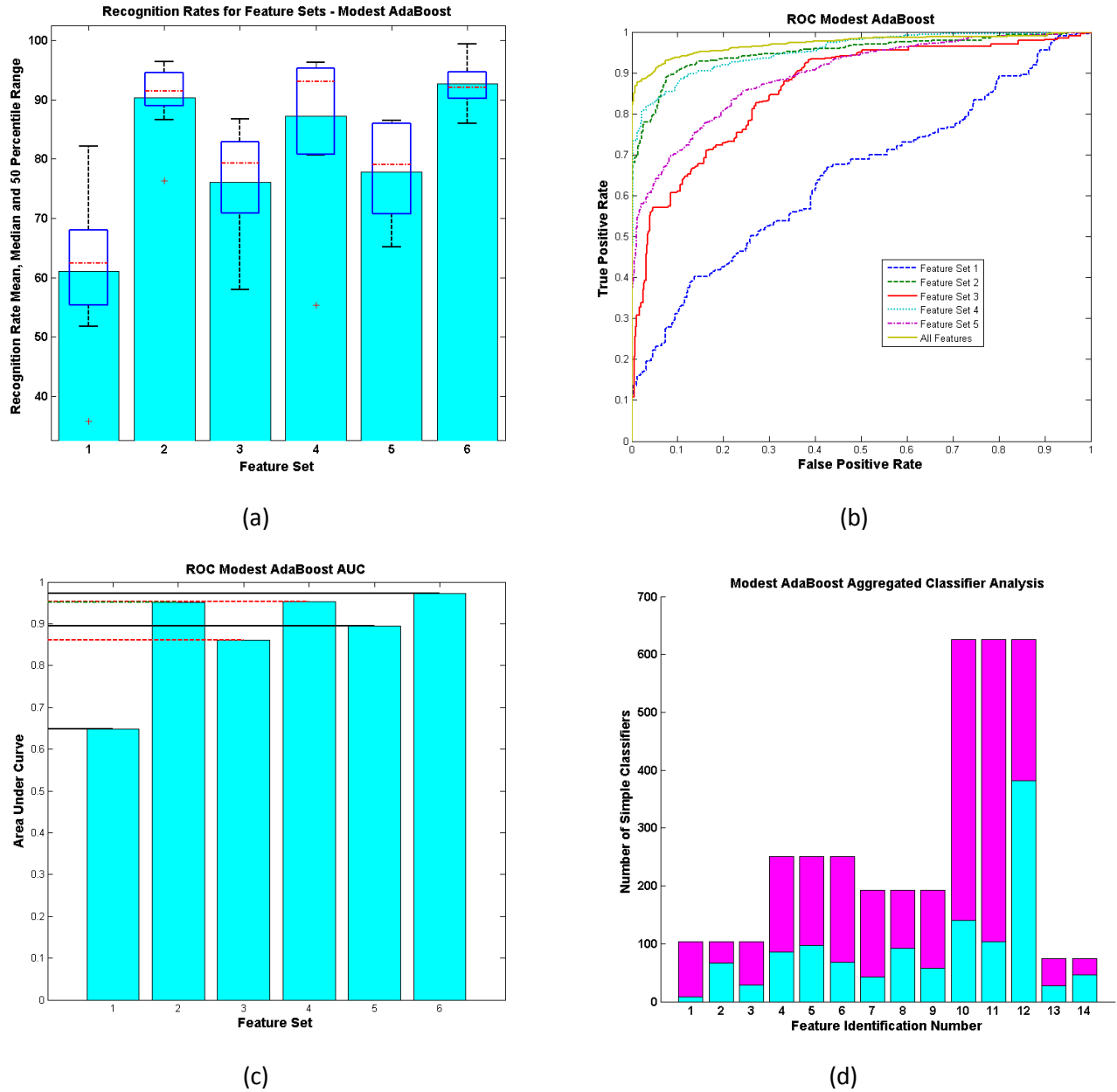


Figure 7. Piecewise performance analysis of the classic AdaBoost framework. (a) Recognition rates under use of individual feature sets. (b) The Receiver Operating Characteristics (ROC) under the use of individual feature sets. (c) Area under the curve (AUC) for each feature set as estimated from the ROC. (d) The number of simple classifiers used by the aggregated AdaBoost classifier. Each set and each feature representation in the classifier pool are separately marked. In all the graphs Set 1 through 5 are as explained by Table 1. Set 6 represents a set containing all 14 features from Table 1.



### 3.6 Discussion of Results

Regarding the research question, what is the state-of-the-art technology available to detect and notify individuals of their rocking behavior? We have identified the state-of-the-art motion sensor that is small enough in form factor to become part of one's everyday clothing. Further, we designed this device to be discrete so that the user does not feel any intrusion into their everyday activities. The software can be run on any mobile processing device that the user already carries like a cell phone or PDA. This allows the users to use the device without carrying any additional load.

Focusing on the research question of, Is it possible to build a device that detects body rocking condition and how well can it distinguish body rocking from other functional activities of daily living? We turn our attention to the various results presented in Section 3.5 to prove the efficiency of our proposed method in detecting body rocking and distinguishing it from other non-rocking behavior.

#### 3.6.1 Packet Length, and Detection Efficiency

From Figure 4 and Figure 5, it is evident that the recognition rates for the two classifiers are comparable and the median recognition rate ranges from 89% to 95%. Based on these numbers, the best performance was achieved at a sample length of 1.5 seconds or 150 samples per packet. Packet length of 150 samples has the highest recognition rate on both the classifiers. Comparing this packet with the 2 seconds packet length or 200 samples per packet, we notice that the 2 seconds packet is very close behind and it has a smaller 1.5 IQR box. Thus, the variance in the recognition rates between 10 subjects is lesser in the 200 samples packet length, implying that the results are more consistent. Further, we noticed that the average natural rocking motion of the 10 subjects was around 27 rocks a minute (i.e. 27 rocks in 60 seconds or 2.22 seconds per rock; this is supported by results from [8]), which implies that a latency of 2 seconds was the closest to the time duration of a single rocking action. As mentioned earlier, all experiments were carried out with an overlap 0.5 seconds or 50 samples between successive packets. Combining these two results, we have

1. Optimum Packet Length: 2 seconds or 200 samples with 0.5 seconds or 50 samples overlap between packets.
2. Best Detection Rate: @ 2 seconds packet length  $\approx 94\%$  under both classifiers

#### 3.6.2 Generalization Capabilities

From Figure 4 and Figure 5, it is very difficult to distinguish any performance benefits between classic AdaBoost and Modest AdaBoost. But analyzing Table 2, we can notice a dramatic difference in the performance of the Modest AdaBoost when compared to classic AdaBoost. The number of false positives is down from 86 to 44 over a ten minute period. That is, the user receives nearly half less number of false feedback with Modest AdaBoost framework when compared to the classic AdaBoost. This was not evident in the detection tests that were carried out with data collected from Routine C (Section 3.3.1.3). We asked the question of why

there is an increased performance in Modest AdaBoost and why there is a discrepancy between the test results from Routine C and the naturalistic data capture (Section 3.3.2). The answer to these questions lies in the generalization capabilities of the two classifiers. We noticed that most of the false feedback provided by classic AdaBoost occurred while the user was sitting and not rocking. In hind sight, we realized a slight discrepancy in our non-rocking (negative class) data collection. While capturing data under Routine B (as explained in Section 3.3.1.2.) the participants were asked to perform various tasks that did not involve rocking to use as negative training set. We realized that most of the participants performed tasks that involved some form of walking or standing activities while they did no activity that involved sitting and not rocking. Thus, just sitting activity was a non-rocking event that was not represented in the training data set. We hypothesize that classic AdaBoost over trained on the non-rocking data while Modest AdaBoost, which is penalized for learning the training set very well, had a better generalization. Extending this heuristic analysis to a more formal analysis, we look at the piecewise performance of the two classifiers. Comparing the ROC curves from Figure 6 (b) with Figure 7 (b), it can be seen that feature set 2 – Variance and feature set 4 – First Order Differential Power performed the best following Set 2 - All features set. Now comparing Figure 6 (d) with Figure 7 (d) it can be seen that Modest AdaBoost distributed its simple classifiers such that there were more classifiers representing the two feature sets 2 and 4. On the other hand, the classic AdaBoost's distribution of simple classifiers is unexplainable as feature set 1 – Mean – seems to have received more representation than set 4. Mean had the worst performance as an individual feature set as can be verified by the ROC curve that comes closest to the diagonal on the plot hinting that the performance is barely above random guess. Contrasting this with Modest AdaBoost selection, Mean is in the bottom two sets among the five feature sets. This bad performance of Mean as a feature set can be understood by looking at the graph shown in the first row and last column of Table 1. It can be seen that the Mean acceleration values between rocking and non-rocking are not significantly different. Table 1, Rows 2 and 4 highlights the capabilities of Variance and First Order Differential Power in distinguishing rocking from non-rocking. This is further confirmed by the ROC graph.

Feature 4 having the highest distribution of simple classifiers under Modest AdaBoost (Figure 7 (d)), within this feature set we can see that the highest number of simple classifier is assigned to feature 12 which corresponds to First Order Differential Power on z axis. As can be verified from Figure 3, the best distinguishing character between non-rocking and rocking patterns seems to be the transformation of a random signal pattern on z-axis to a deterministic sinusoidal waveform. If this can be the true identity of the rocking data stream, feature 12 would capture it in the best possible manner by measuring the power in the first order differential of the temporal signal. Using this feature as the most reliant feature would provide a good basis to support the final classifier selected by Modest AdaBoost.

We are now ready to answer the last research question stating that the use of approximately 2 seconds (or 200 samples @ 100 Hz sampling rate) packet length used with a

learning framework biased towards generalized learning (like Modest AdaBoost) would be a good assistive technology solution for detecting and giving feedback towards stereotypic body rocking. We can extend the same argument to other body mannerisms that involve any form of repetitive body part movement.

### 3.7 Summary:

In this chapter, we have addressed the topic of detecting stereotypic body mannerisms, specifically body rocking, and propose a technology solution for providing an assistive technology that may reduce or control body rocking. We have discussed the hardware and software components of the proposed system in detail and offer a thorough analysis on the learning framework that provides generalization benefits to allow this framework to be extended to detection of any body mannerism. Investigations are in progress to determine how incoming samples of acceleration data can be labeled automatically by the system based on the AdaBoost classifier's classification confidence metrics. This would provide opportunity for self-learning [9] modes where the device can readily understand and learn data points that were not available in the training set. Combining such self-learning into a generalized learner would provide immense opportunities for not only body mannerism detection, but for solving future data mining problems where typical lab setting training data collection would just not be sufficient to train a robust classifier.

From the assistive technology perspective, we plan to integrate a well planned self-monitoring as a part of the proposed device. We are exploring the broad area of human communication to determine the best cognitive self-correction techniques that could augment the proposed solution. We have implemented a rudimentary form of real-time body rock counter, as discussed in [10], but we have not yet tested it for its feedback capabilities.

### 3.8 References:

- [1] N. Krishnan and S. Panchanathan, "Analysis of low resolution accelerometer data for continuous human activity recognition," *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, 2008, pp. 3337-3340.
- [2] L. Bao and S.S. Intille, "Activity recognition from user-annotated acceleration data," 2004, pp. 1-17.
- [3] Y. Yuan and M.J. Shaw, "Induction of fuzzy decision trees," *Fuzzy Sets Syst.*, vol. 69, 1995, pp. 125-139.
- [4] A. Vezhnevets and V. Vezhnevets, "Modest AdaBoost - Teaching AdaBoost to Generalize Better," Novosibirsk Akademgorodok, Russia: 2005.
- [5] A. Vezhnevets and V. Vezhnevets, *GML AdaBoost Matlab Toolbox 0.3*, Graphics and Media Lab, Moscow State University, 2007.
- [6] Y. Benjamini, "Opening the Box of a Boxplot," *The American Statistician*, vol. 42, Nov. 1988, pp. 257-262.
- [7] T. Fawcett, "An introduction to ROC analysis," *Pattern Recogn. Lett.*, vol. 27, 2006, pp.

- 861-874.
- [8] K.M. Newell, T. Incledon, and J.W. Bodfish, "Variability of stereotypic body-rocking in adults with mental retardation," *American Journal on Mental Retardation*, vol. 104, May. 1999, pp. 279 - 88.
  - [9] R. Raina, A. Battle, H. Lee, B. Packer, and A.Y. Ng, "Self-taught learning: transfer learning from unlabeled data," *Proceedings of the 24th international conference on Machine learning*, Corvallis, Oregon: ACM, 2007, pp. 759-766.
  - [10] D.B. McAdam and C.M. O'Cleirigh, "Self-monitoring and verbal feedback to reduce stereotypic body rocking in a congenitally blind adult," *Re:View*, vol. 24, Winter93. 1993, p. 163.