Recognizing Mimicked Autistic Self–Stimulatory Behaviors Using HMMs

Tracy Westeyn, Kristin Vadas, Xuehai Bian, Thad Starner, and Gregory D. Abowd
College of Computing and GVU Center; Georgia Institute of Technology
Atlanta, GA 30332-0280 USA
{turtle, vadas, bxh, thad, abowd}@cc.gatech.edu

Abstract

Children with autism often exhibit self-stimulatory (or "stimming") behaviors. We present an on-body sensing system for continuous recognition of stimming activity. By creating a system to recognize and monitor stimming behaviors, we hope to provide autism researchers with detailed, quantitative data. In this paper, we compare isolated and continuous recognition rates of emulated autistic stimming behaviors using hidden Markov models (HMMs). We achieved an overall system accuracy 68.57% in continuous recognition tests. However, the occurrence of stimming events can be detected with 100% accuracy by allowing minor frame–level insertion errors.

1 Introduction

Autism is a developmental disorder affecting a child's social development and ability to communicate. Children with autism will often exhibit behaviors, such as vocal stutters and brief bouts of vigorous activity (*e.g.*, violently striking the back of the hands) to cope with everyday life. Depending on the child's level of functioning, these highly individualized, self–stimulatory ("stimming") behaviors can be disruptive and socially awkward. Caregivers and researchers would like to explore the correlation between these stimming behaviors and environmental factors, behavioral treatments, mood, or other physiological markers.

To assist in this analysis, we aim to automate the recording and analysis of these behaviors. Although it is impractical for a researcher or therapist to monitor a given child continuously for episodes of stimming, an intelligent monitoring system could collect data daily from the child and filter it to highlight just the stimming episodes. These episodes could then be analyzed, or even replayed alongside other captured information such as video. An automated data collection system may provide insight into a given child's mental and physiological state, and may provide detailed, quantitative data for researchers in the field (which is currently rare).

The system does not need to be 100% accurate. A therapist or researcher can ignore a certain low percentage of false positives. In this sense, the recognition system should err on the side of including too many episodes as opposed to incorrectly ignoring episodes that should have been recorded.

Because we are interested in monitoring the everyday lives of children with autism, the monitoring system must be mobile and non-invasive. We believe such a system could be implemented with small 3-axis accelerometer modules, possibly embedded in the child's shoes, belt, wristwatch, or clothing. Data from the sensor modules would be transmitted wirelessly to a Personal Server located in the child's pocket or backpack [3]. Both the sensor modules and server could be charged daily, and the data downloaded for analysis.

Analysis of this data presents interesting research challenges. Our approach is to use hidden Markov models (HMMs) to distinguish between types of stimming behavior as well as to determine when stimming is not occurring. In this paper, we present a proof-of-concept prototype of such a system.

2 Related Work

Bao and Intille recently compared several sensor positions and methods for recognizing full-body activities. Results showed the use of two accelerometers positioned at the waist and upper arm were sufficient to recognize 20 distinct activities using the C4.5 decision tree learning algorithm, with overall accuracy rates of 84% [1]. The purpose of our system is to monitor stimming behavior so that it can be correlated with various factors (i.e. therapies, varied stress levels, etc.). Similarly, researchers at the Groden Center investigated correlating telemetric heart rate measurements with video for the assessment of stress responses for children with autism [2].

3 Data Set and Collection System

We collected data from a single, neurotypical adult mimicking some anecdotally common autistic self-stimulatory

Proceedings of the 2005 Ninth IEEE International Symposium on Wearable Computers (ISWC'05)
0-7695-2419-2/05 $20.00 © 2005 IEEE

behaviors. Autism is a highly individual disorder. The self-stimulatory behaviors of one child with autism are often dramatically different from those of another child with autism. For this reason, our system is designed to be trained for an individual. However, a common factor among children with autism is that physical self–stimulatory behaviors are often dramatic, abrupt motions, departing from that of normal activity. For this reason, we use 3–axis accelerometers to collect the data.

3.1 Data Set

Our data set consists of acceleration data generated from mimicking autistic stimming behaviors while performing unscripted activities. Seven stimming behaviors and an intermediary activities"garbage" class are defined below.

- **Drumming (dr)**: Repeatedly tapping a surface.
- **Hand Flapping (hf)**: Repeatedly shaking the hands.
- **Hand Striking (hs)**: Repeatedly striking the top of one hand.
- **Pacing (pc)**: Walking a path in a confined space.
- **Rocking (rk)**: Swaying back and forth while standing.
- **Spinning (sp)**: Spinning in circles while standing.
- **Toe Walking (tw)**: Walking with shortened steps up on the toes. Pacing can often occur while toe walking.
- **Garbage**: Any activity not listed above. For most data collection sessions this involved sitting, standing, writing, hand motions involved with talking, and jumping.

The individual was adorned with three wireless Bluetooth accelerometers positioned on the right wrist, the back of the waist, and the left ankle. The individual was instructed to partake in normal daily activities (*e.g.* walking, writing, skipping, etc...), while intermittently performing one of the stimming actions upon receiving an auditory cue. The researcher performed this action until notified to stop. The stimming behaviors were cued in cyclic order once every one to three minutes. The cueing system used to prompt the subject also automatically provided ground truth labeling information for the behaviors of interest. Each sequence of stimming behaviors and intermediary activities was collected over 15 trials — resulting in 15 x 7 = 105 stimming examples and 105 examples of non–relevant activities.

3.2 Data Collection Hardware

Figure 1 shows one of three wireless sensors used for data collection [1]. The sensor contains two perpendicularly mounted dual–axis ADXL202JE accelerometers which are sampled at 100Hz by a PIC microcontroller (via an on–chip analog–to–digital converter). Data is sent to a remote receiving device using a Bluetooth link.

For all experiments, the sensors established a connection via a USB D-Link Bluetooth adaptor to an IBM x31 Pentium M laptop with 256MB of RAM. An hour of data collection produces approximately 36MB of ASCII text data.

[1](http://www.cc.gatech.edu/ccg/resources/btacc/)

Figure 1. Wireless–bluetooth accelerometer.

The laptop was placed on the supports of a false ceiling in the center of the room. This placement allowed for maximum coverage and helped prevent data transmission failures resulting from the occlusion of radio signals by the body. Data was also successfully collected when the laptop was carried in a backpack for additional experimentation in mobile settings.

4 Experiments and Results

For this paper we make minor assumptions allowing us to use HMMs to model the data. We assume that the data and model of the world are stationary. We further assume that the acceleration of human body movements is generated from a single hidden variable representing the behavioral activity of the person (*e.g.*, swaying, spinning, or pacing). For each experiment, eight HMMs were constructed, trained, and tested using the Georgia Tech Gesture Toolkit [4]. These models represent each of the seven stimming behaviors and one "garbage" model (for all other intermediary activities). In all experiments, we used left–right topologies with self transitions, a single loop back to a previous state, and one allowable skip at each state. At each state, a nine–element vector, representing the 3–axis readings from each of the three sensors can be observed. The garbage model consisted of 20 states while six of the seven stimming behaviors were modeled with identical topologies of 10 states. Due to the limited amount of data, models were trained and tested using leave–one–out validation using every permutation of the data set with one example "left out" of the training set. The results of each iteration are then tallied to compute overall statistics of the models' performance.

Two types of experiments were conducted: isolated recognition using HMMs, and continuous recognition using HMMs. With isolated recognition, the ground truth labeling produced during data collection is used to create single–activity examples. Recognition is then performed only on these isolated examples, one at a time, aligning the single observation to the most likely model. Continuous recognition involves aligning a stream of observations (including numerous relevant and irrelevant activities) with multiple models. Alignment is performed by constructing the most probable path through all possible sequences of the models in parallel. Recognition (alignment) accuracy is defined as: $Accuracy = \frac{N - S - D - I}{N}$ where: Substitution errors (S) occur when the system incorrectly classifies a behavior; Insertion errors (I) occur when the system recognizes an in-

Table 1. Confusion matrix for isolated recognition using HMMs. The vertical axis represents ground truth

	dr	gb	hf	hs	pc	rk	sp	tw	accuracy
dr	14	0	0	1	0	0	0	0	93.33%
gb	7	92	1	2	0	0	1	2	87.62%
hf	0	1	12	2	0	0	0	0	80.00%
hs	0	0	1	14	0	0	0	0	93.33%
pc	0	0	0	0	15	0	0	0	100.00%
rk	0	0	0	0	0	15	0	0	100.00%
sp	0	0	0	0	0	0	15	0	100.00%
tw	0	0	0	0	5	0	0	11	73.33%

stance of a behavior that did not occur; Deletion errors (D) arise when the system fails to recognize the occurrence of a behavior within a stream of data; and N represents the total number of examples. It should be noted that insertion and deletion errors can only occur during continuous recognition.

4.1 HMM Isolated Recognition Experiments

To test the validity of the HMM models, we ran isolated recognition on the segmented raw sensor data. Results showed an accuracy of 90.95%. Table 1 provides the overall confusion matrix (with ground truth as the vertical axis) and accuracy breakdown.

Walking-related activities, such as toe walking (tw) and pacing (pc), were sometimes confused. Similarly, wrist–articulated hand motions, such as hand flapping, hand striking, and drumming, were also confused. The HMM models accurately classified all data samples for the pacing (pc), rocking (rk), and spinning (sp) classes. Only one example (of the hand flapping class) is misclassified as garbage. Given our application, it is important that the occurrence of self–stimulatory events be identified, even if they are incorrectly labeled. Misclassifying a self–stimulatory behavior as garbage (*ie:* ignoring it) is worse than misclassifying it as another self–stimulatory behavior.

4.2 Continuous Recognition Experiments

For continuous recognition, we trained the models using 15 sequences. Each sequence contained instances representing 14 activities (7 distinct self–stimulatory behaviors and 7 irrelevant activities interspersed). Because our system acts as a guide, an activity is considered to be correctly identified if over half of the data samples generated from that activity (referred to as frames) are properly classified. Our models detected all 105 instances of the seven self–stimulatory activities (100% correct positive rate) and correctly identify 92.86% of the events when irrelevant activities (garbage) were also considered. However, the models were unable to accurately identify the exact boundaries of the behaviors (*i.e.*, the exact frame where an activity started and stopped). Errors made by the system are not reflected in the event detection accuracy, but rather by the overall system accuracy (as defined in Section 4).

Table 2. Activity level confusion matrix for continuous recognition. The table reflects the results of all permutations of leave–one–out validation. The vertical axis represents ground truth.

	dr	gb	hf	hs	pc	rk	sp	tw	Del
dr	15	0	0	0	0	0	0	0	0
gb	0	90	0	2	2	1	2	8	0
hf	0	0	15	0	0	0	0	0	0
hs	0	0	0	15	0	0	0	0	0
pc	0	0	0	0	15	0	0	0	0
rk	0	0	0	0	0	15	0	0	0
sp	0	0	0	0	0	0	15	0	0
tw	0	0	0	0	0	0	0	15	0
Ins	59	233	280	185	45	14	72	53	

Table 2 shows the continuous recognition confusion matrix. Note the high number of insertion errors (displayed in the bottom row). The majority of insertion errors resulted from sporadic single–frame insertions, though some resulted from larger contiguous, multi–frame insertions. We reduced the number of insertion errors by imposing an insertion penalty on the model alignment process — a capability provided by the Georgia Tech Gesture Toolkit which helps reduce the likelihood of insertion errors. Figure 2 shows the typical results of first imposing an insertion penalty and then smoothing the results with a Gaussian kernel to eliminate remaining, minor insertion errors.

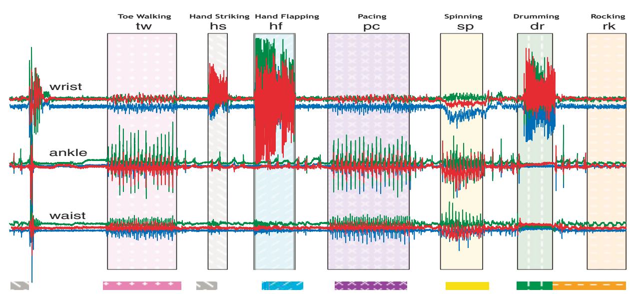

Figure 2. Results of continuous recognition with an insertion penalty and smoothed results. Results are indicated by the thin horizontal stripe below the signal.

An inverse relationship exists between correct activity detection and the insertion penalty. For example, increasing the insertion penalty to 500 during the alignment process decreases the correct positive rate (activity detection) to 70.95%. Insertion errors are reduced and overall system accuracy increases to 68.57%. The bottom half of Figure 3 shows that, as the penalty increases, the number of insertions decreases exponentially, while the number of deletions increases almost linearly. As a result, the event detection rate also decreases. The insertion penalty does not dramatically effect the substitution rate. The top half of Figure 3 shows as the insertion penalty increases, the accuracy

Proceedings of the 2005 Ninth IEEE International Symposium on Wearable Computers (ISWC'05)
0-7695-2419-2/05 $20.00 © 2005 IEEE

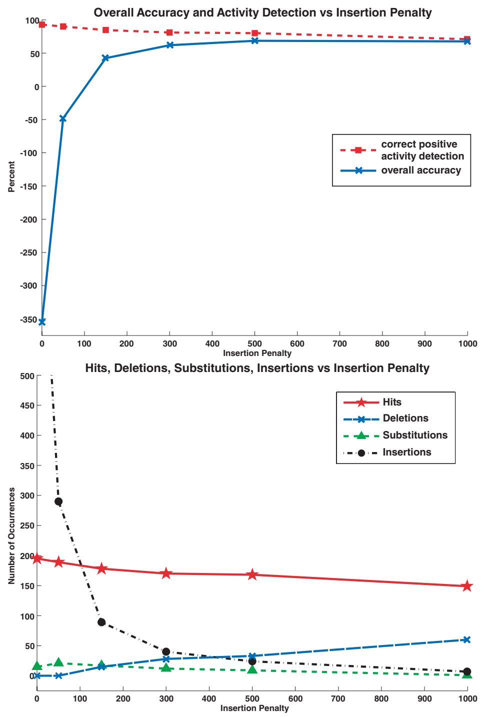

Figure 3. *Top:* **The effects of the insertion penalty on correctness of event identification (dashed red squares) and overall accuracy (solid blue crosses).** *Bottom:* **The effects of the insertion penalty on the number of hits, deletions, substitutions, and insertions.**

rate logarithmically increases. However, as the penalty increases, the correctness of activity detection decreases due to an increase in deletion errors.

4.3 Discussion

Figure 2 shows that our prototype correctly identified all stimming events in a stream of data, even though alignment was not frame–level accurate. While deletion and insertion errors occurred on a frame–by–frame basis, HMMs aligned well enough to avoid deletion errors at the macro level. Even when a high number of frame–level insertion errors exist, a mild insertion penalty and post–alignment smoothing can allow for correct indexing of the data stream. This mild insertion penalty helps reduce insertion errors while minimizing the risk of introducing deletion errors during the alignment process. A correct balance of this trade–off can ensure that no stimming behaviors are ignored.

Our results show that the system can automatically index data in a way that is useful for caregivers and researchers. While the alignment will not be exact, the indexing can still guide the caregiver to the approximate position of a stimming behavior instance within a data stream. Searching the immediate area of a stream will still be faster than searching the entire stream unassisted. With no deletion errors, the caregiver will only need to view data indexed by the system, rather than searching through the stream for missed events. While the system will sometimes identify an uninteresting event as a stimming instance a therapist or researcher can ignore most of these false positives and still potentially benefit from the system's guidance.

5 Future Work

In the next phase, we will collect more authentic data by outfitting a child with autism with our sensors. We will evaluate the ability of our system to index the more authentic data, and will use the results to drive improvements to the system.

Additional future work involves experimenting with wearable platform design for the data collection device (i.e. shrinking the device so it can be stored in the child's backpack or pocket). We also wish to investigate ways to capture and index vocalized self–stimulatory behaviors (*e.g.*, screaming, clicking the tongue, and giggling).

6 Conclusions

In this paper we presented a proof-of-concept system capable of collecting, modeling, and recognizing self–stimulatory behaviors using on–body wireless accelerometers. Our initial results indicate an automatic indexing system for stimming activity is feasable. Seven self–stimulatory behaviors, mimicking those of a child with autism, were performed by a neurotypical adult and modeled using Hidden Markov Models. We explored the performance of these models in both the isolated and continuous settings. The isolated HMM experiments assumed slight noise in data segmentation and achieved accuracy rates of 90.95%. In the continuous recognition experiments, detection of the self–stimulatory event was achieved with 100% correctness (92.86% correctness including identification of non–self–stimulatory activities). While exact segmentation was not possible due to frame–level insertion errors, we were able to improve segmentation by using insertion penalties and smoothing.

References

[1] L. Bao and S. S. Intille. Activity recognition from user-annotated acceleration data. In *Pervasive Computing*, pages 1–17, 2004.

[2] M. Goodwin. Telemetric assessment of stress in individuals with autism. In *Spectrum Disorders. American Psychological Association*, 2004.

[3] G. R. Hayes, J. A. Kientz, K. N. Truong, D. R. White, G. D. Abowd, and T. Pering. Designing capture applications to support the education of children with autism. In *UbiComp: 2004.*, pages 161–178, 2004.

[4] T. Westeyn, H. Brashear, A. Atrash, and T. Starner. Georgia Tech Gesture Toolkit: Supporting experiments in gesture recognition. In *Fifth International Conference on Multimodal Interfaces*, pages 85–92, 2003.

Proceedings of the 2005 Ninth IEEE International Symposium on Wearable Computers (ISWC'05)
0-7695-2419-2/05 $20.00 © 2005 **IEEE**