An Automated Assessment System for Embodied Cognition in Children: From Motion Data to Executive Functioning

Alex Dillhoff[1], Konstantinos Tsiakas[2], Ashwin Ramesh Babu[1], Mohammad Zakizadehghariehali[1], Benjamin Buchanan[3], Morris Bell[2], Vassilis Athitsos[1] and Fillia Makedon[1]
[1]University of Texas at Arlington, [2]Yale University, [3]University of Connecticut
{alex.dillhoff, ashwin.rameshbabu, mohammad.zakizadehgharie}@mavs.uta.edu,
{konstantinos.tsiakas, morris.bell}@yale.edu, {athitsos, makedon}@uta.edu

ABSTRACT

We present our preliminary data analysis towards an automated assessment system for the Activate Test for Embodied Cognition (ATEC), a test which measures cognitive skills through physical activity. More specifically, we present two core ATEC tasks designed to assess attention, working memory, response inhibition, rhythm and coordination in children: the Sailor Step and the Ball-Drop-to-the-Beat task. These tasks are specifically designed to assess lower and upper body accuracy, response inhibition and rhythm. Motion data were collected through the Kinect camera. This paper presents an overview of the assessment tasks, the data collection, and annotation with a preliminary analysis towards an automated scoring system through machine learning and computer vision methods.

CCS CONCEPTS

• **Human-centered computing**; • **Computing methodologies**;

KEYWORDS

Computer Vision, Motion Capture, Embodied Cognition, Cognitive Assessment

ACM Reference Format:
Alex Dillhoff[1], Konstantinos Tsiakas[2], Ashwin Ramesh Babu[1], Mohammad Zakizadehghariehali[1], Benjamin Buchanan[3], Morris Bell[2], Vassilis Athitsos[1] and Fillia Makedon[1] [1]University of Texas at Arlington, [2]Yale University, [3]University of Connecticut {alex.dillhoff, ashwin.rameshbabu, mohammad.zakizadehgharie}@mavs.uta.edu, {konstantinos.tsiakas, morris.bell}@yale.edu, {athitsos, makedon}@uta.edu . 2019. An Automated Assessment System for Embodied Cognition in Children: From Motion Data to Executive Functioning. In *iWOAR 2019 – 6th International Workshop on Sensor-based Activity Recognition and Interaction, September 16–17, 2019, Rostock, Germany*. ACM, New York, NY, USA, 6 pages. https://doi.org/xxxxxx

1 INTRODUCTION

Executive functions are high-order cognitive processes involved in multitasking, time management, attention, planning, inhibition, self-regulation and memory. Children with Attention-Deficit/Hyperactivity Disorder (ADHD) exhibit weaknesses in executive functions, specifically response inhibition, planning, vigilance, and working memory [Willcutt et al. 2005]. Cognitive impairments in early childhood can lead to poor academic performance and require proper assessment and intervention at the appropriate time [McClelland and Cameron 2012]. Such impairments can also affect the individual well into adulthood. Providing a system for automatic assessment can provide more opportunities for diagnosis, treatment, and the progress of cognitive skills.

The NIH toolbox, a standardized test used for cognitive assessment [Zelazo et al. 2013] and other existing computer-based assessments are extensively used to assess executive functions in children, but they require little body movement and may be less closely related to daily functioning than assessing cognition in motion. The **Activate Test of Embodied Cognition (ATEC)** is an assessment test designed to measure executive functions in children through physically and cognitively demanding tasks and provides measurements for attention, working memory, response inhibition, self-regulation, rhythm and coordination, as well as motor speed and balance. The overall goal of our research is to design a high-fidelity and low-cost automated assessment system which analyzes the movements of the performed tasks and produces reliable cognitive measures.

In this paper, we present our proposed methods to automatically administer and assess two core ATEC tasks; the *Ball-Drop-to-the-Beat* task and the *Sailor Step* task. These tasks are designed to assess upper-body (hands) and lower-body (feet) movements. We describe the two tasks, as well as the experimental approach towards an automated scoring system through computer vision and machine learning methods. Children between the ages of 6-10 were invited to perform the ATEC assessment tasks in classroom environments. Video data were annotated and scored by experts for both presented tasks. Preliminary results for both tasks indicate the efficiency of our proposed methods towards an automated assessment system for embodied cognition in children.

2 RELATED WORK AND MOTIVATION

Physically active behaviors are important in the daily lives of children and have implications for fitness, learning, social interactions, and physical and psychological development [Malina et al. 2016]. Moreover, studies have shown a measurable improvement in cognitive skills and academic performance in children associated with increased physical fitness [Davis and Cooper 2011; Donnelly and Lambourne 2011]. These indicate the strong relation between motor and cognitive development in children and their implications to

iWOAR 2019, September 16–17, 2019, Rostock, Germany

ACM ISBN xxxxxxx.
https://doi.org/xxxxxx

daily functioning [Davis et al. 2011]. While there are existing assessment systems for both motor development [Tieman et al. 2005] and neurocognitive measures (NIH) [Zelazo et al. 2013], as well as for assessing emotional and behavioral problems (CBCL) [Achenbach et al. 2000] and executive function behaviors at home and at school (BRIEF) [Stiffman et al. 1984], these are either computer-based or paper-based in the form of parent/teacher reports and require no movement. Our proposed embodied cognition assessment system, ATEC, utilizes the advances of computer vision and machine learning methods to analyze a child's performance during a set of physical tasks, specifically designed to extract information about the child's cognitive and motor functions and development.

All ATEC tasks involve physical movement. Action recognition methods use image or body key-point data to model the spatial and temporal features of each class-action [Ali and Taylor 2018; Devanne et al. 2014; Li et al. 2018]. Thus, they are the most applicable to our solution. Action recognition often involves classifying high-level events with more variation between classes [Kuehne et al. 2011; Marszałek et al. 2009; Soomro et al. 2012]. [Zhang et al. 2016] report that many image-based action recognition data sets feature low variability amongst actions. Although each individual ATEC task must be performed in a specific manner, there can be large variations between each individual's performance. As such, our approach must follow other methods and data sets with high intra-class variance [Forster et al. 2012; Neidle et al. 2012; Piergiovanni and Ryoo 2018]. The body key-points are the most salient high-level features for these tasks. Recent body pose estimation methods have shown excellent results on benchmarks featuring multiple persons with varying viewpoints and lighting [Cao et al. 2018; Fang et al. 2017; Pavlakos et al. 2018]. Since our participants are relatively close to the cameras and are recorded with good lighting, we are able to get high quality key-point estimates.

We use the DeepGRU [Maghoumi and LaViola Jr 2018] model as a benchmark as it requires fewer parameters than other recurrent models. The authors show good performance even with smaller data sets which is especially important for our current system as we have recorded our own dataset with a relatively small number of examples per class. The goal of our work is to develop efficient and reliable methods for child activity recognition, since detecting and analyzing child movements is challenging due to high variability and large amount of random movements.

3 THE ACTIVATE TEST FOR EMBODIED COGNITION

ATEC consists of 17 physical exercises with different variations and difficulty levels, designed to provide measurements of executive and motor function, including sustained attention, self-regulation, working memory, response inhibition, rhythm, and coordination, as well as motor speed and balance. These measurements are converted to a final ATEC score which describes the level of development (e.g., early, middle, full development).

In this work, we focus on two core ATEC tasks: Sailor Step and Ball Drop. These exercises were prioritized because of their stronger association with self-regulation. They are also related to attention, working memory, response inhibition, and rhythm. The Sailor Step task includes lower-body activity, where children must remember instructions and coordinate their feet movements following visual cues presented during a themed video clip. Ball-Drop-to-the-Beat includes upper body activity which requires children to pass a ball from one hand to another following the presented rules.

3.1 Ball-Drop-to-the-Beat

Ball-Drop-to-the-Beat is a core ATEC task designed to assess both audio and visual cue processing while performing upper-body movements. The child is required to pass a ball from one hand to another, following audio and visual instructions. The task modifies the rules of the Red-Light/Green-Light game in which the participants are required to perform certain movements while they hold a ball. Based on the rules, the child is instructed to pass the ball for Green-Light, keep the ball still for Red-Light, and move the ball up and down with the same hand for Yellow-Light. The light colors are presented both audibly and visually to measure both audio and visual accuracy and response inhibition. The task is assessed at 60 beats per minute and 100 beats per minute.

During this assessment, the stimuli are presented as pictures of traffic lights: red, green and yellow. The child is instructed to do the appropriate movement with the ball when a new picture of a traffic light appears. Apart from accuracy and response inhibition, exercises also assess rhythm. The ATEC on-screen host, Aliza, presents the stimuli in a rhythmic manner by saying "green/red/yellow-light" in two beats; one for the color word and one for the word "light". The children are instructed to perform the movements in two beats. For pass and raise commands, the ball is raised on the first beat and either passed or lowered on the second. To acclimate the participants to the task, they are instructed to pass the ball eight times following the rhythm of the spoken instructions (and ONE, and TWO, ...). Figure 1 visualizes both audio and visual stimuli.

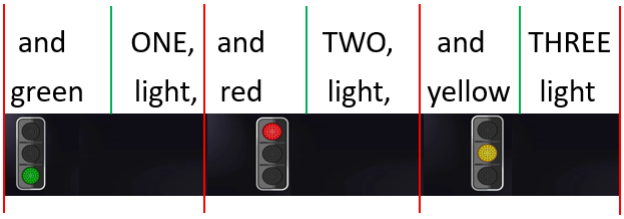

Figure 1: Audiovisual stimuli during the Ball Drop task. Each segment requires a specific activity (red lines). For the audio tasks, each segment includes two beats (green line).

3.2 Sailor Step

The Sailor Step task, another core ATEC task, is designed to assess visual cue processing while performing lower-body movements with accuracy, rhythm and response inhibition. In order to make the task compelling and engaging it is presented as a dance exercise, where the child has to move following the Sailor Step instructional video. The video shows a predefined sequence of three different visual stimuli that appear on the screen for a given time: (a) a Red Crab, (b) a Blue Crab, and (c) a Happy Clam, as shown in figure 2. Based on the rules, the child needs to move one step to the right when the Red Crab appears on the screen, one step to the left for

the Blue Crab, and stand still if the Happy Clam appears. Each (left, right) step is performed in two beats; one for the first foot and one for the second one. The task requires the child to (a) remember the rules, (b) move accurately based on the rules, and (c) move in rhythm with the song.

The scoring approach considers three different scores: (a) visual accuracy, (b) visual response inhibition and (c) visual rhythm. It consists of 13 (blue and red) crabs and 8 Happy Clams, which are presented during the song in a predefined and fixed order to ensure test-retest reliability. The visual accuracy score, $acc = »0; 13¼$, is the amount of correct movements (right or left), following the rules. Even if the child shows delays in performing the correct movement (out of rhythm), the step is considered accurate. The response inhibition score, $res = »0; 16¼$, measures how a child responds during a Happy Clam. For this paper, we consider three possible movements during a Happy Clam: (a) Still: where the child stays completely still, (d) Half-Still: where movements are detected only during the first beat (half) of the segment and (c) Not-Still: where movements are detected during the entire segment. The rhythm score, $rhy = »0; 26¼$, is the amount of beats where the child stayed in rhythm during movement. Each movement has two beats, one for each foot and gets: two rhythm points if the child performs both movements in rhythm (complete), one rhythm point if the child misses the first beat (incomplete) and zero rhythm points if the child fails to complete a movement during both beats.

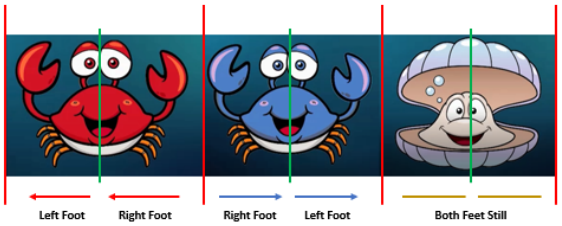

Figure 2: Sailor Step task rules. Children are instructed to perform a specific movement for each presented stimuli. Each segment has two beats; one for each foot movement.

4 DATA COLLECTION AND ANNOTATION

In this section, we describe the data collection procedure for the ATEC administration. Children between the ages of 6-10 were invited to participate to the ATEC assessment, after the required parent consenting and screening procedure required by the study protocol. The ATEC administration includes a recording and administrative interface, which was created for the purposes of streamlining assessments with as little distraction and interruption as possible. The ease of use is paramount as the assessment suite will be used by both experts and non-experts. Video data is preferred as sensor-based data collection can be more expensive and distracting, especially with child participants. Two Microsoft Kinect V2 cameras record a front and side view of the participant. RGB, depth, audio, and skeletal data are stored. The recording modules are connected to the Android-based administrative interface which controls the flow of the assessment. It allows the administrator to select between all the tasks in the ATEC suite. Figure 3 shows a diagram of the recording protocol. Each task has an instructional video and one or more assessment videos, while there are also practice videos to ensure that the child has understood the rules. An instructional video gives a brief demonstration on the current exercise and how it is performed. Selecting an assessment video triggers the recording modules to activate while Aliza, the on-screen instructor, guides the children through each task.

Annotation software was developed to enable both computer science and cognitive experts to visualize and annotate the collected data. The software performs automated segmentation given the time stamps of the presented stimuli for each task. For each assessment recording, an expert evaluates the performance against a set of task-specific criteria. The annotation and scoring guidelines were designed considering both computer vision and cognitive related aspects of the task. This expert annotation is then used as the benchmark for automated approaches based on machine learning and computer vision methods.

5 AUTOMATED SCORING APPROACH

In this section, we present our experimental procedure and results towards an automated scoring system for both tasks following the task rules and scoring guidelines. We highlight the challenges for each task as well as discuss the limitations of our approaches and our future steps.

5.1 Ball-Drop: Upper Body Activity

For this task, there are three main events involved: ball pass, no ball pass and hand raise. The participant is asked to perform one of these actions when instructions are provided through visual or auditory cues, as mentioned before. The aim of the automated system is to detect the actions performed and score them according to the rules. More specifically, Ball-Drop scores include (a) auditory and visual accuracy scores, (b) visual and auditory response inhibition scores and (c) auditory rhythm scores.

5.1.1 Experimental Approach. The complete pipeline to score the ball drop task consists of multiple parts. Videos are recorded at the rate of 30 frames per second. The input video is broken down into image frames which are decoded. As the first step for feature extraction, we extract the body key-points of the participants. The body key-points considered for this experiment are wrist points, elbow points and shoulder points. A Convolutional Neural Network based approach is used to extract the body key-points [Cao et al. 2018]. The system takes in the decoded image as input of size $w \quad h$. The feed forward network predicts 2D confidence maps of the body joint locations and a set of 2D vector fields of part affinity fields which is the degree of association between the parts.

With the key-points extracted for every frame in the segment, more detailed features such as the x-distance, y-distance between the wrist points, elbow points, wrist and the shoulder points were extracted. These features were used to detect various events during the exercise. For every segment, the features were pre-processed to remove noise. Noise includes any wrong detection of body key-points, key-points not being detected etc. First, moving average is computed on the features for every segment followed by applying a low pass filter to remove the high frequency components caused by hand jitters and minor movements.

Figure 3: The ATEC setup includes two Kinect cameras, a large screen and a tablet interface for the administrator. Administration takes place in classroom environments. Annotation software was developed to enhance manual scoring and annotate the collected data, given the task rules and the cognitive measures to be assessed.

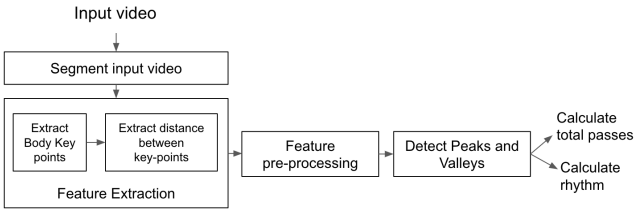

Figure 4: Complete pipeline to compute scores for the ball drop task

A ball pass event occurs when the participant moves the hand holding the ball towards the other hand, makes the transfer and moves back to the original position. In such a scenario, the distance between the wrist points decreases until the transfer happens and increases again. Similarly, a hand raise event occurs when the participant moves the hand holding the ball towards the shoulder of the hand holding the ball and retreats back to the original position where the distance between the wrist and the shoulder joint initially increases and starts to decrease while retreating. A peak is formed every time such an event occurs.

After processing the segmented features, we attempt to detect peaks and valleys in the segment. Mathematically, peaks and valleys represent local maxima and minima. A video segment T which consists of n image frames, with x being the features for every image frame, is defined by

$$T = \{(f_1; x_1); (f_2; x_2); \ldots; (f_n; x_n)\} \quad (1)$$

Peaks (P) and valleys (V) for a segment are defined by,

$$P = \{(f_i; x_i) | (x_{i-1} < x_i > x_{i+1}) \_ (x_1 > x_2) \_ (x_n > x_{n-1})\}$$
$$\text{and}$$
$$V = \{(f_i; x_i) | (x_{i-1} > x_i < x_{i+1}) \_ (x_1 < x_2) \_ (x_n < x_{n-1}); \forall i = 2; 3; \ldots n-1\} \quad (2)$$

With the above equations, peaks and valleys are detected which correspond to the respective events. Figure 5 (left) represents a ball pass event in a video segment. Figure 5 (right) represents the task being divided into different segments. In this approach, noise in the signal was detected as a ball pass when there was no pass. Hence, a threshold on the height of the peak was considered. The height of the peak is the distance between a peak and a valley in the segment. We used 998 segments from our dataset for this experiment. 15 percent (144 segments) of the data was used to identify the right threshold for different events and 85 percent (854 segments) of the data was used for evaluation. For these evaluations we use data from 7 subjects performing 10 ball drop related tasks. Each subject performed the exercises twice, two weeks apart, to determine test-retest reliability.

Other than the proposed method, we also used two Neural network based classifiers. A 1D Convolutional Neural Network (CNN) was used where the input dimension for the CNN for each segment was $Nx16$ with N representing the time steps. The network consisted of two 1D convolutional layers followed by a fully connected layer and a softmax output layer. Similarly, a recurrent neural network based approach was also attempted with Gated Recurrent Units (GRU) as proposed by [Maghoumi and LaViola Jr 2018]. The input to the system was of dimension N, where N represents the time steps and $F \in \mathbb{R}^{25 \times 3}$.

The neural network and DeepGRU models were trained and evaluated using sequence annotations for each subject and task. The sequences were labeled as one of the 3 classes. The models were trained using k-fold cross validation such that each fold is tested in a user independent manner. The accuracy from each of the 7 folds is averaged and reported in table 1.

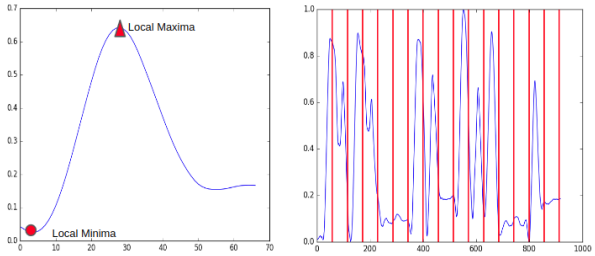

Figure 5: left: Ball pass event in a video segment. Right: Segmentation of features, given the presented stimuli

Method	Overall Accuracy
1D-CNN	0.59
DeepGRU	0.61
Proposed Method	**0.78**

Table 1: Comparison with other Deep Neural Network based methods

	Normalized Confusion Matrix		
	Pass	No Pass	Hand Raise
Pass	0.89	0.11	0.00
No Pass	0.22	0.78	0.00
Hand Raise	0.31	0.00	0.69

Table 2: Confusion matrix for the proposed method

5.1.2 Results. Table 1 lists all of the attempted methods. The proposed method performs highest with 78 percent accuracy. Table 2 represents the confusion matrix of the validation set using the proposed method. Additionally, we evaluate rhythm for that particular event. An event is said to be in rhythm if the peak is within an annotated time frame, determined by task. For a ball pass event, the peak would occur in the middle of the segment as the hands are closest together.

5.1.3 Discussion and future work. We observed that the deep neural network based approach yielded low accuracy when compared to the proposed method. This could be due to insufficient training data and additionally, in the proposed method, we observed that the system sometimes classified a no-pass as a pass. On looking at such samples, the participants initiated a pass, but then retreated without completing the pass. Similarly, a hand raise was also classified as ball pass when the participants raised their hand first and then made a pass. Our current work involves working directly on the RGB image frames. The lower accuracy of the neural network methods can be explained by higher intra-class variance that is not fully captured by our current data set. We suspect that these models will surpass our proposed method as more subjects are recorded.

5.2 Sailor Step: Lower Body Activity

The purpose of the Sailor Step task is to assess both visual cue processing and rhythm, based on the performed activity of the child during each stimuli. As mentioned earlier, there are three main movements involved in this exercise: moving left, moving right, staying still. Scoring refers both to accuracy (which movement was performed) and rhythm (when the movement was performed).

5.2.1 Experimental Approach. notes: Each segment includes the movement performed during a presented stimuli (crab, clam). Following the approach of our preliminary work [Buchanan et al. 2019], and considering the nature of the required movements, the joint coordinates for both feet were extracted from the Kinect skeleton data. Since the duration is the same for each stimuli, each segment is represented as a fixed-length vector of the feet joint coordinates. More specifically, the segment consists of 45 frames, resulting in a 90-dimensional vector for both feet (*left-foot* and *right-foot*). For both visualization and normalization purposes, the gradient for each vector was calculated. The annotation software was used to visualize both video and the gradient plot for each segment to score and annotate the videos of N = 15 recorded assessments, resulting in N = 334 annotated segments.

Considering the scoring guidelines and the different stimuli, the available annotations are [*LeftComplete, LeftIncomplete, RightComplete, RightIncomplete, Still*] given a "crab" segment and [*Still, HalfStill, NotStill*] given a Happy Clam segment. An *Other* label was also used for random or extraneous movements for future analysis. Considering the task rules and the different stimuli (clams, crabs), our experimental approach includes the training of five different models which predict the following classes:

M: [LeftComplete, LeftIncomplete, RightComplete, RightIncomplete, Still, HalfStill, NotStill]
M1: [Left, Right, Still], where *Left = [LeftComplete, LeftIncomplete]* and *Right = [RightComplete, RightIncomplete]*
M2: [LeftComplete, LeftIncomplete]
M3: [RightComplete, RightIncomplete]
M4: [Still, HalfStill, NotStill]

These models were selected considering the presented stimuli and the scoring guidelines both for accuracy and rhythm. One approach would be to use a single model M for all segments (stimuli-required movement). Another approach, as proposed in our previous work [Buchanan et al. 2019], is a hierarchical one which considers the presented stimuli of the segment. Given a red/blue crab, the system uses model $M1$ to predict the direction of the movement (Left, Right, Still). The predicted direction can be used to score accuracy, compared to the required direction (task rules). Given a predicted direction, in order to score for rhythm (Complete = 2, Incomplete = 1, Still = 0), the system uses models $M2$ and $M3$ for left and right, respectively. Given a Happy Clam segment, the system uses model $M4$ which assigns an accuracy score (Still = 2, HalfStill = 1, NotStill = 0).

In order to get an insight of the class distributions and evaluate our classes selection, we applied truncated SVD for dimensionality reduction to visualize a 2D projection of the datapoints (6). We observe that there is a much clearer distinction between Right and Left (complete/incomplete) classes, compared to the Still classes (Still, Half-Still, Not-Still).

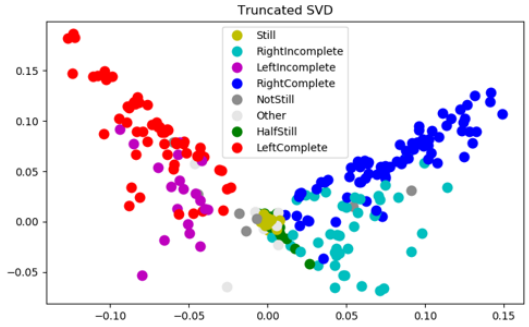

Figure 6: 2D projection of the data using truncated SVD

In order to train and evaluate the proposed models and get an insight towards improvement, we compared different classification and feature extraction approaches. More specifically, we used *K-Nearest Neighbor, Random Forest, Decision Tree* and *Multi-layer Perceptron* classifiers on two datasets, one with the raw data (gradient vectors) and one with manually extracted feature vectors. The extracted features include the statistics [*min, max, mean, median, std*] of the whole vector, of each foot, as well as for each half of each foot segment (two beats), resulting to a set of 35 features. Additionally, we evaluated our data using a PyTorch implementation of DeepGRU [Maghoumi and LaViola Jr 2018]. This model was chosen because of its ability to explicitly model temporal features via a

recurrent design. Instead of using the raw or processed features as described above, we opted to use the body joint locations provided by OpenPose [Cao et al. 2018]. Following [Maghoumi and LaViola Jr 2018], we augmented the data by applying random scaling, translation, and gesture path stochastic resampling [Taranta et al. 2016].

5.2.2 Results and Discussion. Table 3 summarizes our results for the defined models using (a) the gradient data, (b) the extracted features and (c) the DeepGRU implementation. While all approaches result to high accuracy scores, most classification approaches fail to distinguish the classes in model M4, resulting in many false negatives. This may be due to the imbalanced dataset, since there are very few samples for Half-Still and Not-Still, compared to Still. While our manually extracted features and the DeepGRU approach did not result in improved total accuracy, the feature-based models showed better results in M4. Different feature extraction approaches or deep-learning approaches for automatic feature extraction and classification will be further investigated to face the challenge of imbalanced data and intra-class variability.

	gradient data				extracted features				OpenPose + DeepGRU
models	KNN	DT	RF	MLP	KNN	DT	RF	MLP	
M	82.5	85.9	75.3	84.7	81.0	81.3	85.3	84.4	83.7
M1	98.3	96.4	90.1	98.7	97.7	98.0	98.7	97.4	83.1
M2	89.5	92.1	82.9	89.5	88.2	86.8	93.4	90.8	88.7
M3	92.5	92.5	90.8	95.8	88.3	94.2	95.0	93.3	92.9
M4	86.2	86.2	78.5	86.2	90.0	89.2	90.0	86.2	86.2

Table 3: Summary of classification accuracy

6 CONCLUDING REMARKS AND FUTURE WORK

In this paper, we presented our preliminary results towards an automated scoring approach for two core tasks of the Activate Test for Embodied Cognition. We presented the two exercises, Ball-Drop-to-the-Beat and Sailor Step, designed to assess executive functioning in children. Our main research goal is to design a fully-automated, high-fidelity, and low-cost assessment system for embodied cognition. The purpose of this paper is to compare different approaches and models for both tasks towards an improved scoring system.

Our ongoing work includes more data collection and improvement of the current methods, considering the challenges highlighted for each task. More detailed analysis is needed considering rhythm scoring. Further analysis can be used to extract more information related to performance delays and speed, self-correction, and extraneous movements, which will help us identify and model individual differences in child performance.

ACKNOWLEDGMENTS

This work was partially supported by National Science Foundation grants IIS 1565328 and IIP 1719031.

REFERENCES

Thomas M Achenbach, Thomas M Ruffle, et al. 2000. The Child Behavior Checklist and related forms for assessing behavioral/emotional problems and competencies. *Pediatrics in review* 21, 8 (2000), 265–271.

Alaaeldin Ali and Graham W Taylor. 2018. Real-Time End-to-End Action Detection with Two-Stream Networks. In *2018 15th Conference on Computer and Robot Vision (CRV)*. IEEE, 31–38.

Benjamin Buchanan, Konstantinos Tsiakas, and Morris Bell. 2019. Towards an automated assessment for embodied cognition in children: the sailor step task. In *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. ACM, 331–332.

Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2018. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. In *arXiv preprint arXiv:1812.08008*.

Catherine L Davis and Stephanie Cooper. 2011. Fitness, fatness, cognition, behavior, and academic achievement among overweight children: do cross-sectional associations correspond to exercise trial outcomes? *Preventive medicine* 52 (2011), S65–S69.

Emma E Davis, Nicola J Pitchford, and Ellie Limback. 2011. The interrelation between cognitive and motor development in typically developing children aged 4–11 years is underpinned by visual processing and fine manual control. *British Journal of Psychology* 102, 3 (2011), 569–584.

Maxime Devanne, Hazem Wannous, Stefano Berretti, Pietro Pala, Mohamed Daoudi, and Alberto Del Bimbo. 2014. 3-d human action recognition by shape analysis of motion trajectories on riemannian manifold. *IEEE transactions on cybernetics* 45, 7 (2014), 1340–1352.

Joseph E Donnelly and Kate Lambourne. 2011. Classroom-based physical activity, cognition, and academic achievement. *Preventive medicine* 52 (2011), S36–S42.

Hao-Shu Fang, Shuqin Xie, Yu-Wing Tai, and Cewu Lu. 2017. Rmpe: Regional multi-person pose estimation. In *Proceedings of the IEEE International Conference on Computer Vision*. 2334–2343.

Jens Forster, Christoph Schmidt, Thomas Hoyoux, Oscar Koller, Uwe Zelle, Justus H Piater, and Hermann Ney. 2012. RWTH-PHOENIX-Weather: A Large Vocabulary Sign Language Recognition and Translation Corpus.. In *LREC*. 3785–3789.

Hildegard Kuehne, Hueihan Jhuang, Estíbaliz Garrote, Tomaso Poggio, and Thomas Serre. 2011. HMDB: a large video database for human motion recognition. In *2011 International Conference on Computer Vision*. IEEE, 2556–2563.

Chenyang Li, Xin Zhang, Lufan Liao, Lianwen Jin, and Weixin Yang. 2018. Skeleton-based Gesture Recognition Using Several Fully Connected Layers with Path Signature Features and Temporal Transformer Module. *arXiv preprint arXiv:1811.07081* (2018).

Mehran Maghoumi and Joseph J LaViola Jr. 2018. DeepGRU: Deep Gesture Recognition Utility. *arXiv preprint arXiv:1810.12514* (2018).

Robert M Malina, Sean P Cumming, and Manuel J Coelho-e Silva. 2016. Physical Activity and Inactivity Among Children and Adolescents: Assessment, Trends, and Correlates. In *Biological Measures of Human Experience across the Lifespan*. Springer, 67–101.

Marcin Marszałek, Ivan Laptev, and Cordelia Schmid. 2009. Actions in context. In *CVPR 2009-IEEE Conference on Computer Vision & Pattern Recognition*. IEEE Computer Society, 2929–2936.

Megan M McClelland and Claire E Cameron. 2012. Self-regulation in early childhood: Improving conceptual clarity and developing ecologically valid measures. *Child development perspectives* 6, 2 (2012), 136–142.

Carol Neidle, Ashwin Thangali, and Stan Sclaroff. 2012. Challenges in development of the american sign language lexicon video dataset (asllvd) corpus. In *Proc. 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, Language Resources and Evaluation Conference (LREC) 2012*. Citeseer.

Georgios Pavlakos, Luyang Zhu, Xiaowei Zhou, and Kostas Daniilidis. 2018. Learning to estimate 3D human pose and shape from a single color image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 459–468.

AJ Piergiovanni and Michael S Ryoo. 2018. Fine-grained activity recognition in baseball videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 1740–1748.

Khurram Soomro, Amir Roshan Zamir, and Mubarak Shah. 2012. UCF101: A dataset of 101 human actions classes from videos in the wild. *arXiv preprint arXiv:1212.0402* (2012).

Arlene R Stiffman, John G Orme, Deborah A Evans, Ronald A Feldman, and Phoebe A Keeney. 1984. A brief measure of children's behavior problems: The Behavior Rating Index for Children. *Measurement and Evaluation in Counseling and Development* 17, 2 (1984), 83–90.

Eugene M. Taranta, II, Mehran Maghoumi, Corey R. Pittman, and Joseph J. LaViola, Jr. 2016. A Rapid Prototyping Approach to Synthetic Data Generation for Improved 2D Gesture Recognition. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 873–885. https://doi.org/10.1145/2984511.2984525

Beth L Tieman, Robert J Palisano, and Ann C Sutlive. 2005. Assessment of motor development and function in preschool children. *Mental retardation and developmental disabilities research reviews* 11, 3 (2005), 189–196.

Erik G Willcutt, Alysa E Doyle, Joel T Nigg, Stephen V Faraone, and Bruce F Pennington. 2005. Validity of the executive function theory of attention-deficit/hyperactivity disorder: a meta-analytic review. *Biological psychiatry* 57, 11 (2005), 1336–1346.

Philip David Zelazo, Jacob E Anderson, Jennifer Richler, Kathleen Wallner-Allen, Jennifer L Beaumont, and Sandra Weintraub. 2013. II. NIH Toolbox Cognition Battery (CB): Measuring executive function and attention. *Monographs of the Society for Research in Child Development* 78, 4 (2013), 16–33.

Jing Zhang, Wanqing Li, Philip O. Ogunbona, Pichao Wang, and Chang Tang. 2016. RGB-D-based Action Recognition Datasets: A Survey. Wanqing Li (2016). arXiv:1601.05511 http://arxiv.org/abs/1601.05511