Towards an Automated Assessment for Embodied Cognition in Children: The Sailor Step Task

Benjamin Buchanan
Computer Science Department
University of Connecticut
benjamin.buchanan@uconn.edu

Konstantinos Tsiakas
Department of Psychiatry
Yale University
konstantinos.tsiakas@yale.edu

Morris Bell
Department of Psychiatry
Yale University
morris.bell@yale.edu

ABSTRACT

In this paper, we present our preliminary data analysis towards an automated assessment system for the Sailor Step task. The Sailor Step task is an embodied cognition task designed to assess attention, working memory, response inhibition, rhythm and coordination in children. Children must remember instructions and coordinate their movements following visual cues presented during a themed video clip. Skeleton data were collected through the Kinect camera. This paper presents an overview of the assessment task, the data collection and annotation with a preliminary analysis towards an automated scoring system through motion capture analysis methods.

CCS Concepts

•Human-centered computing → Human computer interaction (HCI); Interaction design;

Keywords

Embodied Cognition, Motion Capture, Cognitive Assessment

1. INTRODUCTION AND RELATED WORK

Cognitive impairments in early childhood can lead to poor academic performance and require proper assessment and intervention at the appropriate time [2]. The NIH toolbox, a standardized test used for cognitive assessment [3] and other existing paper or computer-based assessment tasks assess executive functions in children, but they require little body movement and may be less closely related to daily functioning than the measure we are developing, which assesses cognition in motion. This approach is in line with a trend towards assessing cognitive functioning through physical tasks [1]. The Activate Test of Embodied Cognition (ATEC) is an assessment test designed to measure executive function in children through physical cognitively demanding tasks.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

PETRA '19, June 5–7, 2019, Rhodes, Greece

© 2019 ACM. ISBN 978-1-4503-6232-0/19/06...\$15.00

DOI: <https://doi.org/10.1145/3316782.3322744>

2. THE SAILOR STEP TASK

The Sailor Step task is one of the basic ATEC tasks, and it is designed to assess visual cue processing in order to perform lower-body movements with accuracy, rhythm and response inhibition. In order to make the task compelling and engaging, it is presented as a dance task, where the child has to move following the Sailor Step song clip. The song clip 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 1. 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 Clams 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.

Figure 1: The visual stimuli presented during Sailor Step: Red Crab, Blue Crab, and Happy Clam.

3. PRELIMINARY EXPERIMENTS

3.1 Dataset Description

In the context of the larger ATEC study, which is to evaluate a comprehensive battery of tasks for Embodied Cognition, children between ages 5-10 were invited to perform the ATEC assessment on two occasions with at least a week break for the purpose of determining test-retest reliability. Briefly, the ATEC setup includes two Kinect cameras for motion capture and a large screen which visualizes the videos for each task. For the Sailor Step task, there is an instruction video which explains and demonstrates the task rules, highlighting of the importance of moving in the correct direction and in rhythm, or staying still, when needed. The ATEC assessment includes two versions of the song, one slow and one fast. Both versions include a sequence of 21 visual stimuli (13 crabs and 8 clams). For our preliminary analysis, we used the Kinect Skeleton data from $N = 18$ total rounds from the slow version of the Sailor Step task.

3.2 Scoring Approach and Annotation

The Sailor Step task scoring approach considers three different scores: (a) visual accuracy, (b) visual response inhibition and (c) visual rhythm. The visual accuracy score, $acc = [0, 21]$, is the amount of correct movements, 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, $res = [0, 8]$, is the amount of times where the child stayed still during a happy clam. For this paper, we consider self-correction (when the child makes a small movement during a happy clam but returns to the still position), as inaccurate. The rhythm score, $rhy = [0, 26]$, 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 a complete movement (in rhythm), one rhythm point if the child has a delay performing the correct movement and zero rhythm points if the child is totally out of rhythm. Considering this scoring approach, we built a Graphical user Interface (GUI) for video annotation. The timestamps of the visual stimuli were extracted from the presentation video. Based on these timestamps, each video was segmented to a set of 21 clips of the same duration (1s) and frame rate (30fps). In order to follow the scoring guidelines, we defined five main classes $C1 = [LeftComplete, LeftIncomplete, RightComplete, RightIncomplete, Still]$ and one extra class for *Other* type of movements. The annotation procedure led to an annotated library of activity clips.

As a baseline approach, we extracted the Kinect skeleton data from each clip, getting the x -axis coordinates of both ankle joints (left and right) and concatenated them in a single array. Since the video clips have the same length and rate, all arrays have the same dimensions (45 frames per foot - 90 values). These coordinate vectors were grouped by their annotation labels, resulting to a 339×90 dataset of movement labels and a vector including ankle joint coordinates during the clip. Since movements can be done differently (e.g., small vs. big step length), each coordinate vector was modified to a distance-based vector D , where $D(i)$ is the distance of the foot during frame i and frame 0 ($D[0] = 0$). Each array was then normalized to $[0, 1]$.

3.3 Preliminary Experiments and Results

For our initial experiments, we trained a KNN classifier, using two variations of the dataset. $D1$ includes the five main classes $C1$, while $D2$ groups the labels into three categories $C2 = [Left, Right, Still]$, where Left/Right includes both complete and incomplete movements.

Our experiments are conducted following two approaches. A multiclass model which classifies an input vector into one of the five classes $C1$ and a hierarchical approach which consists of two levels of classification; the first one returns $c \in C2$ and the second decides between *complete* and *incomplete* given Left or Right. We used $D1$ to train a KNN model $M1$ for the five classes ($C1$). $D2$ was used to train a model $M2$ for the three classes ($C2$). Moreover, we used subsets of $D1$ to build two models $M3$ and $M4$, which classify right or left movements to complete and incomplete, respectively. A K-fold cross-validation was followed to evaluate the models. Figure 2 shows the hierarchical approach, as well as the confusion matrices for the different models. The higher accuracy scores of the hierarchical-approach models show that two levels of classification is a promising approach.

Figure 2: The hierarchical classification approach and the confusion matrices for the different models

4. ONGOING AND FUTURE WORK

While our preliminary experiments indicate promising results, two main limitations are the imbalanced dataset and that the Kinect skeleton data can often be noisy, thus not preferred for joint detection. We are currently investigating more sophisticated models for joint detection, e.g., OpenPose. This would significantly improve our results. Another limitation is the feature selection approach. For our classifiers, we used the raw position data for each foot as a feature vector. More sophisticated feature selection algorithms will be investigated, as well as Deep Learning approaches to include high-dimensional feature vectors, e.g., including more joints, such as knee and head. Moreover, due to noisy data while in still position, a threshold value needs to be estimated before normalization to distinguish between minor and major movements and other movements of interest (e.g., self-correction during a Happy Clam, balance).

5. ACKNOWLEDGMENTS

This work is based upon research supported by NSF under award number CHS 1565328 and the NSF REU program.

6. REFERENCES

- [1] S. Gattupalli, D. Ebert, M. Papakostas, F. Makedon, and V. Athitsos. Cognilearn: A deep learning-based interface for cognitive behavior assessment. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, pages 577–587. ACM, 2017.
- [2] M. M. McClelland and C. E. Cameron. Self-regulation in early childhood: Improving conceptual clarity and developing ecologically valid measures. *Child development perspectives*, 6(2):136–142, 2012.
- [3] P. D. Zelazo, J. E. Anderson, J. Richler, K. Wallner-Allen, J. L. Beaumont, and S. Weintraub. Ii. nih toolbox cognition battery (cb): Measuring executive function and attention. *Monographs of the Society for Research in Child Development*, 78(4):16–33, 2013.