Detecting delays in motor skill development of children through data analysis of a smart play device

Jörg Sander
University of Amsterdam
Amsterdam
jorg.sander@toologic.com

Antoine de Schipper
Annette Brons
Svetlana Mironcika
University of Applied Sciences
Amsterdam
a.w.de.schipper@hva.nl

Huub Toussaint
University of Applied Sciences
Amsterdam
Huub.Toussaint@hva.nl

Ben Schouten
University of Applied Sciences
Amsterdam
Ben.Schouten@hva.nl

Ben Kröse
University of Applied Sciences
Amsterdam
b.j.a.krose@hva.nl

ABSTRACT

This paper describes experiments with a game device that was used for early detection of delays in motor skill development in primary school children. Children play a game by bi-manual manipulation of the device which continuously collects accelerometer data and game state data. Features of the data are used to discriminate between normal children and children with delays. This study focused on the feature selection. Three features were compared: mean squared jerk (time domain); power spectral entropy (fourier domain) and cosine similarity measure (quality of game play). The discriminatory power of the features was tested in an experiment where 28 children played games of different levels of difficulty. The results show that jerk and cosine similarity have reasonable discriminatory power to detect fine-grained motor skill development delays especially when taking the game level into account. Duration of a game level needs to be at least 30 seconds in order to achieve good classification results.

ACM Classification Keywords

I.5.2 Feature evaluation and selection; K.8 Personal computing; games

Author Keywords

Motor skill assessment, machine learning, signal processing

INTRODUCTION

Children's activity levels have decreased over the past 25 years [4]. The amount of physical activity in the primary school period is associated with the development of motor skills [2].

Less physical active children engage themselves less in new activities that challenge their motor development. Consequently, the motor development is delayed. In Amsterdam alone the percentage of children with a delay in gross motor development is about 15% and this number increases [10].

In order to counter these developments it would be beneficial if the early detection of delayed motor skill development in primary school children could be facilitated with so called *smart play sets* that can be used in physical education. These play sets are capable to provide a high level diagnosis of the children's motor skill development through analyzed sensor data (e.g. accelerometer or gyroscope) and possibly engage children in physical activities that are fun to do. This paper focuses on the use of the *Futurocube*[1], a commercially available toy, that was adapted to detect fine motor skill delays. The cube enables playing games at different levels of difficulty and measures acceleration and game state continuously. The work presented here tries to answer the following research questions:

1. What features are best for detecting delays in fine motor skill development of children: a) established, empirical movement measures for smoothness and coordination and/or b) game state information of the Futurocube?

2. Are there specific levels of difficulty of a game where these detections are best?

The contribution of this research is twofold. The experiments tested the different types of features on a large realistic data set from game play of 28 children. The study also compared two different classifiers for the detection of motor skill deficiencies.

RELATED WORK

Previous research on quantitative assessment of motor abilities mostly focused on wearable technology in physical medicine and rehabilitation where subjects were suffering from severe

---
[1] http://www.futurocube.com/

Figure 1: Futurocube

motor impairments e.g. stroke survivors [8] or cerebral palsy [9]. Many of these studies apply a machine learning approach in which the accelerometer and/or gyroscope signal is used to extract and calculate features from the time and frequency domain that are subsequently used to train and test an algorithm that is able to predict the motor skill performance of impaired subjects.

The work presented in this article differs in three ways from previous research: 1) a discriminatory model that detects children with motor development delays is made instead of a predictive model; 2) the children participating in the experiments do not suffer from severe motor skill impairments due to (birth) injuries and therefore differences in fine motor skills are assumed to be more fine-grained and 3) the analysis presented here will include a *game state* feature e.g. an error measure that will quantify how accurately a child can perform the predefined task.

APPROACH

Game device
The Futurocube uses vibration, audio and colorful LEDs to interact with the player. A photo is shown in figure 1. It uses a low energy communication protocol to wirelessly exchange information with a computer. Each side of the cube is divided into nine squares, each containing a LED that can be identified by a specific *index* number $i \in \{0, \ldots, 53\}$. The Futurocube can be programmed by a simple, typeless 32-bit language called *PAWN*[2]. The experimental data is collected from a built-in 3-axial accelerometer where each axis has a 16 bit data output. The game software was able to capture data samples with a frequency of approximately 20.8 Hz.

Roadrunner game
For the experiments a game application was scripted (referred to as roadrunner hereafter) to challenge and assess the cube handling speed of the child. The child has the task to pursue a white spot i.e. LED (hereafter referred to as *walker*) that moves over the surface of the cube. The player has to make sure that the walker is always on top of the cube.

Each step the walker moves randomly (using uniform probabilities) to an adjacent square in one of three directions: forward, right and left. The delays between the moves will determine the level of difficulty of the game (inverse proportional). A

[2]http://www.compuphase.com/pawn/pawn.htm

game lasts for 180 seconds and is divided into six time segments of 30 seconds (denoted T) where each segment has a fixed level of difficulty (hereafter referred to as LoD). The roadrunner game uses three LoDs (denoted $l \in \{1, 2, 3\}$), each with a different delay (500, 400, and 350 msec respectively). In the game every LoD occurs twice. The game starts with a segment with $l = 1$ followed by five random LoD's in which two subsequent segments can not have the same LoD.

Features
The data analysis used the digitally low-pass filtered (cutoff frequency of 8Hz) accelerometer signal and game state data. All mathematical calculations in this study were performed using Python 2.7 with modules Numpy 1.11.1 and Scipy 0.18.1.

Measures of smooth and coordinated movements
As a basis for the further analysis the *Movement Intensity (MI)*[9] was calculated for each sample as follows;

$$MI(t) = \sqrt{a_x(t)^2 + a_y(t)^2 + a_z(t)^2} \qquad (1)$$

with $a_x(t)$, $a_y(t)$ and $a_z(t)$ being the acceleration measured on the x-, y- and z-axis, respectively at time t.

Common measures of smooth and coordinated movements are jerk measures [6] and spectrum-based measures [7] (e.g. power spectral entropy). This study calculated both measures. Based on the work of [6] the mean squared jerk (MSJ) was calculated.

The power spectral entropy (PSE) can be interpreted as a complexity measure for an uncertain system [11]. This work used the calculation of the PSE as outlined in [11].

Game state measure
In order to measure how accurately a child can pursue the optimal trail of the walker a distance measure, hereafter referred to as Cosine Similarity Measure (CSM), was defined that quantifies the deviation between the walker position and the inverted *gravity cursor* representing the top of the cube.

The center of the Futurocube can be referred to as the origin of a 3D Euclidean space. It is then possible to express each cube square and therefore walker location by means of a vector $\mathbf{w_i} \in \mathbb{R}^3$ where $i \in \{0, \ldots, 53\}$ in this 3D reference framework.

In order to express the vector in the same metric as the embedded accelerometer, the vectors are normalized and subsequently multiplied by -256 which is approximately equal to 1g.

The cursor $\mathbf{c} \in \mathbb{R}^3$ can be represented as the gravity cursor that indicates the top of the cube and is determined by means of the acceleration readings provided by the system[3]. As a final measure the sum of squared distances between the cursor and walker at each timestamp t denoted as $d(t)$ where $\{d(t)|0 \leq d(t) \leq 2\}$ was calculated.

Classifiers
Feature values collected from a time segment are used to detect delays in motor skill development. For this, machine learning

[3]this location was not visible for the player.

classifiers were studied. As a starting point this work used the Gaussian Naive Bayes (GNB) and Support Vector Machine[4] (SVM) classifiers. The use of ensemble (Random Forest) and boosting methods (Extreme Gradient Boosting machine) did not reveal better classification results and therefore, following the Occam's razor principle, only results for GNB and SVM are reported here.

EXPERIMENTS

Experiments with children playing the game were carried out to compare the three different features, their combination, to compare two different classifiers and three different segment lengths.

Participants

A total of 28 children, age 7 to 8 years from a primary school in Amsterdam participated in the experiments. All children were assessed before the study by means of the Movement Assessment Battery for Children (MABC-2) which is designed to identify and describe impairments in motor performance of children and adolescents [1]. The standard MABC-2 scores were converted into percentiles (higher is better) to determine how a child's motor coordination compares to typically developing children of the same age. Children whose performance falls below the 16th percentile and, according to the MABC-2, are labeled as *likely* or *most likely* to have motor skill deficiencies (denoted *deficiency* with class label 1). All other children (equal or above the 16th percentile) were denoted as *normal* with class label 0. 63.6% of the collected experimental data belongs to children with normal and 36.4% can be attributed to children that are likely or most likely showing delayed motor skill performance.

Each child trained for three minutes individually with the Futurocube before performing the experiment. Except for one, all children played two games, resulting in 110 data points for each LoD.

Performance measure

F1 and recall scores were obtained performing five-fold stratified cross-validation (CV) for 50 repetitions on the data set. Performance measures were collected for each level of difficulty and for the complete game. As mentioned in the previous section the used data set was significantly biased towards children with normal motor skill performance and therefore a stratified CV approach was chosen to assure that the data skewness is represented in the bootstrap samples during cross-validation.

RESULTS

Figure 2 visualizes the cross-validation results for the Naive Bayes (figure 2a, 2b, 2c, 2d) and SVM (figure 2e, 2f, 2g, 2h) classifiers for the individual features and the complete feature set (figure 2d, 2h). Each figure presents the F1 and recall scores for the three distinct levels of difficulty of the game together with the overall scores.

For the combined features the GNB performs slightly better than the SVM (F1 54.9 resp. 50.6 and recall 46.1 resp. 38.7).

---

[4] using a non-linear RBF kernel

For both classifiers the best discriminatory performance can be achieved on the third game level (F1 61.7 resp. 58.4 and recall 58.1 resp. 49.0) when using the complete feature set.

The discriminatory performance of the mean squared jerk (MSJ) seems to increase when the level of difficulty of the game raises, whereas the Cosine Similarity Measure (CSM), based on game state information has a clear discriminatory peak performance at game level 2. The cross-validation scores for the PSE do not reveal a clear trend, the SVM scores show an equivalent increasing trend as for the MSJ (better if game level increases) but scores obtained with GNB show a slight increase only at game level 2.

Figure 3 shows the effect of different segmentation lengths (30, 10 and 5 sec) on the GNB classification performance of the complete feature set which significantly decreases with decreased length.

CONCLUSION AND FUTURE WORK

The results indicate that both mean squared jerk and cosine similarity are suitable measures to distinguish between children with normal and delayed motor skill development. The time window on which the feature calculation is based should be at least 30 seconds in order to yield good discriminatory power. Moreover, the predictive power of the features is influenced by the level of difficulty.

The developed method was able to predict the outcome of the MABC assessment, thus we can conclude that it is indeed possible to measure fine motor skills in a fast and playful way. This is very valuable, because conventional tests are time-consuming and not playful.

Since the level of difficulty influences the predictive power of the features, future work should investigate whether it is possible to adapt the level of the game to the motor skill level of a child. Then, the discriminatory power of the features as well as the playfulness can be increased. Another improvement could be to calculate features that capture how accurately children adapt to new situations, for instance to changes in speed or to changes in the direction of the walker.

Future work should apply more complex machine learning models that have the capacity to model temporal sequences. Especially the current success of deep recurrent networks, most notably those that rely on Long Short-Term Memory cells (LSTM) [5], have achieved impressive performance when applied to human activity recognition [3]. Most probably this approach should be combined with the collection of gyroscope data that can in principle be extracted from the used game device.

REFERENCES

1. Ted Brown and Aislinn Lalor. 2009. The movement assessment battery for children-second edition (MABC-2): A review and critique. *Physical & occupational therapy in pediatrics* 29, 1 (2009), 86–103.

(a) Cross-validation score - mean squared jerk with GNB

(b) Cross-validation score - cosine similarity with GNB

(c) Cross-validation score - power spectral entropy with GNB

(d) Cross-validation score - all three features with GNB

(e) Cross-validation score - mean squared jerk with SVM

(f) Cross-validation score - cosine similarity with SVM

(g) Cross-validation score - power spectral entropy with SVM

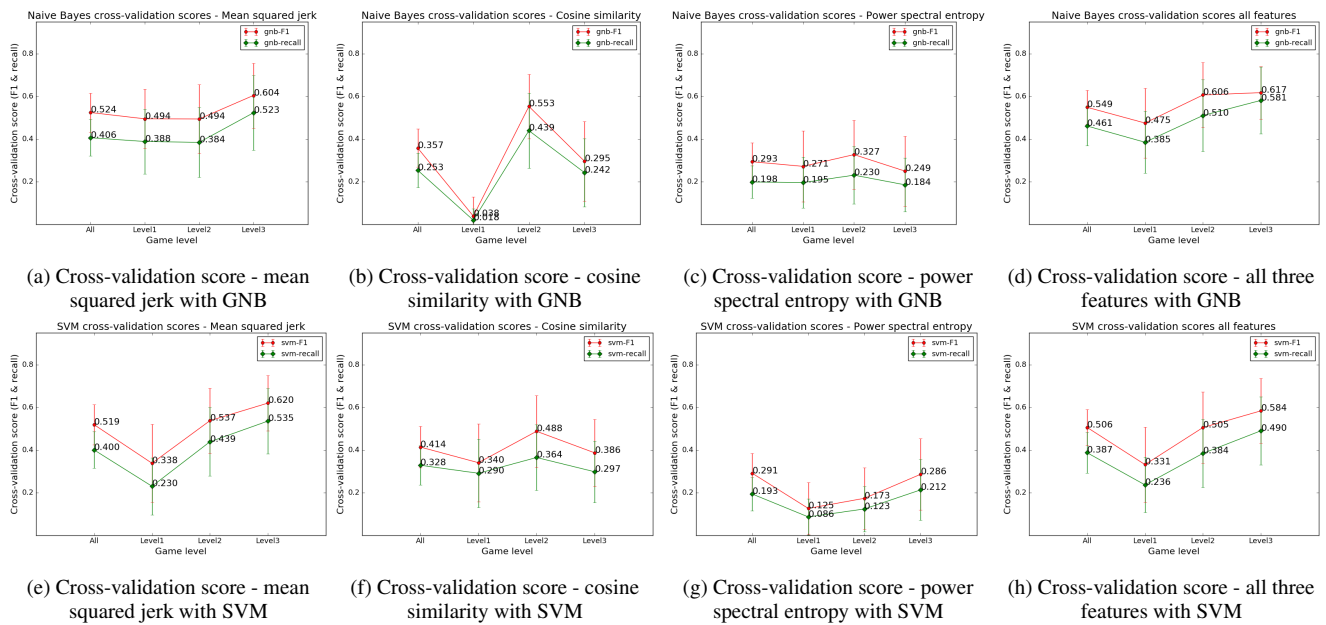(h) Cross-validation score - all three features with SVM

Figure 2: Cross-validation score with Naive Bayes & SVM on individual features and complete feature set
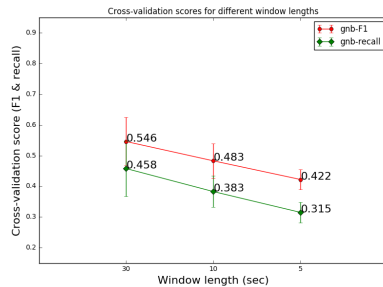

Figure 3: Cross-validation score for GNB with complete feature set for different segment lengths

2. Dylan P Cliff, Anthony D Okely, Leif M Smith, and Kim McKeen. 2009. Relationships between fundamental movement skills and objectively measured physical activity in preschool children. *Pediatric exercise science* 21, 4 (2009), 436–449.

3. Nils Y Hammerla, Shane Halloran, and Thomas Ploetz. 2016. Deep, convolutional, and recurrent models for human activity recognition using wearables. *arXiv preprint arXiv:1604.08880* (2016).

4. VH Hildebrandt, WTM Ooijendijk, and M Hopman-Rock. 2008. Trendrapport bewegen en gezondheid 2006/2007. (2008).

5. Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

6. Neville Hogan and Dagmar Sternad. 2009. Sensitivity of smoothness measures to movement duration, amplitude, and arrests. *Journal of motor behavior* 41, 6 (2009), 529–534.

7. Motonaga Kojima, Shuichi Obuchi, Kousuke Mizuno, Osamu Henmi, and Noriaki Ikeda. 2008. Power spectrum entropy of acceleration time-series during movement as an indicator of smoothness of movement. *Journal of physiological anthropology* 27, 4 (2008), 193–200.

8. Shyamal Patel, Richard Hughes, Todd Hester, Joel Stein, Metin Akay, Jennifer G Dy, and Paolo Bonato. 2010. A novel approach to monitor rehabilitation outcomes in stroke survivors using wearable technology. *Proc. IEEE* 98, 3 (2010), 450–461.

9. Christina Strohrmann, Rob Labruyère, Corinna N Gerber, Hubertus J van Hedel, Bert Arnrich, and Gerhard Tröster. 2013. Monitoring motor capacity changes of children during rehabilitation using body-worn sensors. *Journal of neuroengineering and rehabilitation* 10, 1 (2013), 1.

10. Huub Toussaint, Antoine de Schipper, Tim van Kernebeek, and Ilse Kat. 2015. MAMBO Meten Amsterdamse Motoriek Basis Onderwijs. (2015). `http://www.hva.nl/kc-bsv/projecten/content/projecten-algemeen/monitoring-gezonde-ontwikkelin-kinderen.html`

11. Aihua Zhang, Bin Yang, and Ling Huang. 2008. Feature extraction of EEG signals using power spectral entropy. In *2008 International Conference on BioMedical Engineering and Informatics*, Vol. 2. IEEE, 435–439.