Eye-tracking test battery for detecting cognitive impairments in premature children

Andrea De Gobbis1*, Marta Malavolta2, Emiliano Trimarco3, Isabel Benavente-Fernández1,4,5, Simón Pedro Lubián-López3,4, Aleksander Sadikov1,2 and Vida Groznik1,2,6

1NEUS Diagnostics d.o.o., Ljubljana, Slovenia
2University of Ljubljana, Faculty of Computer and Information Science, Ljubljana, Slovenia
3Biomedical Research and Innovation Institute of Cádiz (INiBICA) Research Unit, Puerta del Mar University, Cádiz, Spain
4Division of Neonatology, Department of Paediatrics, Puerta del Mar University Hospital, Cádiz, Spain
5Area of Paediatrics, Department of Child and Mother Health and Radiology, Medical School, University of Cádiz, Cádiz, Spain
6University of Primorska, Faculty of Mathematics, Natural Sciences and Information Technologies, Koper, Slovenia

Abstract
Premature birth exponentially increases the risk for impaired neurological outcomes later in life, and early diagnosis is critical to optimise therapeutic options. There is evidence that oculomotor movements can be used as biomarkers for cognitive impairment (CI) in adults and young children. The aim of this study is to develop a prototype of a test battery using screen-based eye-tracking for detecting early signs of CI in preterm children and monitoring their neurological development. The study will also delve into identifying potential biomarkers of cognitive functions based on oculomotor movements found in medical literature, and provide methods to design explainable features and models. Finally, we summarise the most common experimental design practices, and propose an eye-tracking test battery that, by combining different stimuli, could be able to measure CI in different cognitive domains.

Keywords
Eye-tracking, premature children, neurodevelopment impairment

1. Introduction

An estimated 15 million births in the world every year are preterm, amounting to 9.4% of all live births [1]. Prematurity leads to an increased risk of altered neurodevelopmental outcomes in childhood and adolescence (such as Autism Spectrum Disorder [2], altered brain development [3], or cognitive and motor delays [4]), with many survivor children facing a lifetime of disability [5]. Despite that advances in neonatal care have greatly improved survival of preterm born
infants even at an extremely low gestational age, long-term neurodevelopmental outcomes have not improved significantly. For this reason, early diagnosis is an important strategy that could lead to a quick treatment and a wider array of therapeutic options.

There is evidence that oculomotor movements during specific tasks, such as smooth pursuit [6, 7], reading [8] and dot counting [9] are biomarkers for impaired cognitive processes in adults (e.g. linked to Alzheimer or Parkinson’s disease). In the case of children, eye-tracking has been used to measure sensory [10], cognitive [11] and social [12] functions. Thus, a combined eye-tracking test battery able to summarise the state of the patient’s cognitive development appears feasible; and it would help identifying early signs of altered brain maturation and detect a wider spectrum of symptoms.

The objective of this work is to summarise the current standard practices in test design and data analysis, and to provide the reader with a handbook of eye-tracking-based diagnostics. Moreover, we propose a novel testing paradigm that, by combining existing methodologies, allows to monitor the neurodevelopment of young children. In the Sections 2-3 we describe the state of the art in using a screen-based eye-tracking with children, highlighting the most common issues and challenges of creating a test battery for very young patients. We will also present existing eye-tracking test batteries based on machine learning [8, 13] used as part of a clinical decision support system. In Section 4 we present a prototype of our combined test battery for children (as young as 3 months corrected age) and briefly describe how data may be parsed.

2. Related Work

The non-invasiveness of eye-tracking methods has made them a particularly appealing approach with younger patients and has inspired a variety of works in the last two decades [14]. In Section 3, we build upon the previous work by Venker et al. [15], which presents an overview of using eye-tracking with children afflicted by Autism Spectrum Disorders, and Gredeback et al. [7] which summarises how eye-tracking can be used to monitor neurodevelopment in children. We review the state-of-the-art in diagnosing with oculomotor movements and highlight the challenges of testing younger patients with eye-tracking.

Different test procedures have been used with children to measure cognitive functions such as ability to smooth pursuit [10], attention [11], spatial inhibition [16], memory [17], and social orienting [12], these works offer adaptations of existing cognitive tests to the eye-tracking paradigm. An alternative approach is proposed by Oakes [18], who advises against using the device to adapt tests that could be conducted by medical professionals and instead suggests a more exploratory approach of gaze trajectories during everyday activities. Data disruption is investigated by Wass et al. [19], where they describe how age can impact the quality of eye-tracking data and design some strategies to preprocess raw data. Other factors that have been found to influence data quality are eye colour [20] and head positioning [21]. Moreover, as reported in previous studies [15, 10], standard calibration procedures can prove difficult with younger patients.

Test batteries using oculomotor movements as a biomarker to detect cognitive impairment have been employed with ageing patients for an early diagnosis of dementia [22, 13, 8]. With a
similar approach, Kaul et al. relate eye movements during smooth pursuit to neuropsychological tests taken at 6.5 years. These methodologies, once properly adapted by age group, provide a blueprint for our combined test battery.

3. Eye-tracking tests overview

From a physiological point of view, there are three main types of eye movements: **saccades** are very rapid movements that align a stimulus to the area of highest acuity (fovea), a **fixation** is defined as the moment where gaze position is fixed on the image, usually between two saccades. Finally, **smooth pursuit** is the type of movement where eyes remain fixed on a moving object without saccadic activity and thus the gaze position changes slowly. Smooth pursuit develops early in life [10] and it is a biomarker of cognitive functions [6, 23].

There are two main types of eye-tracking devices used in infancy research: head-mounted and screen-based [24]. In the former case, the device is fitted on a helmet and can be carried as the patient moves in an environment; in the latter case, the eye-tracker is fixed under a screen where the stimuli are presented. In both setups the gaze is recorded by capturing the cornea reflection of a small infrared light with a camera and reconstructing the person’s point of view. Since we aim at creating a test battery that can be applied to patients as young as 3 months old and a head mounted tracker could prove uncomfortable for infants, in this work we focus our attention on the screen-based eye-tracker. This choice allows us to create different types of tests for the same instrument and monitor patients during early development. Nonetheless, both methods present advantages and disadvantages in a clinical setting and for an overview we refer to [24].

In the next section we compare the setup and of medical studies using eye-tracking in children, especially if preterm, describe tests batteries based on eye-tracking, and present how the data can be parsed and analysed.

3.1. Patient setup and calibration procedure

During testing, the patient is seated comfortably in front of the screen from a distance that varies from 60 cm [17, 25, 16] to around 120 cm [26, 10, 27, 11, 28], younger patients can be positioned in either a baby seat by themselves [26, 16] or in their caretaker’s lap [10, 11]. Ben Itzhak et al. [29] compile a set of good practices to follow when setting up the environment (e.g. having natural light coming from the side). The authors warn about having the caretaker behind the patient, which is a very common practice, since the eye-tracking device could erroneously detect their gaze, and suggest to employ sunglasses to solve this problem. Another difference that can influence analysis [19] is the sampling rate of the eye-tracker, which can reach the 300 Hz [26, 17] in a hospital setting but for a widespread application the commercially available 60 Hz sampling device is more affordable due to cost.

Calibration is an essential first step when using an eye-tracker. The participant needs to look at different points spanning the entire screen, this allows the device to adapt to the patient and map camera signals to gaze positions. For adults and older children (≥ 6 years) the procedure poses no issue, the patient can simply be instructed to look at the dots. Thus, a higher (5 to 8) number of dots is used to ensure high precision in the measurements during testing, the
sufficient number of dots is suggested by the device’s manufacturer. In case of younger patients, calibration becomes challenging since the participant cannot be instructed and might not pay attention to the screen. The common solution employed [19, 10] is using a lower number of points (2 to 4), substituting dots with attractive stimuli such as smiling faces and coloured balls, and animating the stimuli and playing a rhythmic sound.

3.2. Type of stimuli and analysis

We divide the stimuli in three macro categories depending on the type of eye-movements the test should elicit:

1. **Smooth pursuit tasks** present an object (usually a dot but sometimes a smile for younger patients [10]) moving in a periodic pattern, usually a sinus wave [10, 22, 9] but some works use in addition triangular waves [26, 6, 23]. The stimulus can move either in one dimension along the horizontal or vertical direction [26, 6, 9, 22, 23] or in a circular pattern to test both directions simultaneously [10]. One approach is to study smooth pursuit from an input/output dynamical system prospective, with the moving stimulus as input term and the gaze position as output [23, 10]. Features encoded within this paradigm are inspired by dynamical systems and time series analysis (e.g. gain ratio, phase shift, cross-correlation, and mean squared error between input and output). It is detected that with high frequency stimuli often the patient starts compensating with anticipatory saccades [9, 10, 23], in this case it is possible to separate the saccadic and smooth pursuit contributions and analyse them separately.

2. **Fixation and saccade tasks** measure how quickly (time to first fixation) and how long (looking time) the patient fixates on a new stimulus. This paradigm covers a wide variety of approaches aimed at monitoring different cognitive functions, depending on the type and timing of the stimuli. Attention tests measure reaction time to a stimulus given different cues [11, 28, 25]. Memory capabilities are measured by presenting a pattern, letting the patient get acclimatised to it, displaying the same image with some differences and measuring the looking time to the novel stimuli [16, 17]. Social interaction is tested by presenting images containing or not human presence and measuring the difference in looking pattern [30, 12, 31]. The study by Oyama et al. [13] proposes an example of a test battery consisting exclusively of fixation tasks, displaying the versatility of this type of tasks.

3. **General tasks** that mimic everyday activities instead of adapting existing neuropsychological tests, and as such can elicit saccades, fixations and smooth pursuit. In contrast with the other categories, in this case the objective is data exploration and the challenge is feature design, since there is no well-defined cognitive ability under scrutiny. The approach then consists in finding differences in gaze behaviours during complex activities, and the challenge lies in designing suitable features and parsing methods without specific domain knowledge. An example is given by paper [8], where the authors show that there is a significant difference in reading behaviour between healthy and cognitively impaired individuals when measuring reading time and the distribution of forward (right) and backward (left) saccades.
4. Combined test battery

Combining the works presented in Section 3 and building up from existing test batteries to diagnose cognitive impairment in adults [8, 9], we present a first prototype of test battery designed to identify and monitor cognitive impairment due to premature birth. The test is appropriate for 4 months old participants and as such it contains no complex tasks and no instructions.

The present study was conducted on 23 babies (10 females and 13 males, 15 term babies ranging from 3 to 24 months of age, and 8 preterm babies ranging from 3 to 20 months of corrected age) at the Hospital Universitario Puerta del Mar, Cadiz (ethical committee code PIEBA 0672-N-22, register number 44.22). The babies’ caretakers voluntarily accepted to participate in the pilot study after routine visits at the hospital. The aim of this pilot study is designing a first version of a test battery to study the feasibility of employing eye-tracking to find statistically significant differences in gaze behaviour of premature children. In particular, we investigate if the tasks can be presented in a single session and how long we could feasibly keep the participant’s attention. The considerations in the current paper are mostly qualitative, and data analysis is left for future work. The setup is inspired by previous works in the field described in Section 3: the procedure is conducted in a small room lighted from the side, the patient is seated on their caretaker’s lap at 60 cm from a 24 inches computer screen. A simple seven points calibration procedure using a sequence of white crosses as fixation stimuli proved to work correctly for the majority of the patients, as such we do not employ different stimuli for the calibration process in this test. All tests were conducted using a Tobii 4C Eye-Tracker working at 90 Hz.

The test battery is designed to measure responses across a wide variety of cognitive functions while at the same time being short enough that an average young child will not become fussy before the end. The test battery is comprised of the following tasks, which are presented on a black background to offer the maximal contrast with the stimuli, and are interspersed with a smile appearing at the centre of the screen to attract the patient’s attention:

- **Sensation task**: a smooth pursuit task with a sinusoidal ∼0.4 Hz one dimensional wave and a smiling face as stimulus. The original study [10] reports that 5 months old children are able to follow a smiling face in a circular movement and the optimal frequency to minimise missing data is between 0.1 and 0.4 Hz. The movement is exclusively horizontal since it develops earlier than vertical smooth pursuit [6, 9]. The stimulus remains on the screen for a total of 8 s. Figure 1a shows an explicative diagram of the task.

- **Attention task**: this task takes inspiration from similar existing methodologies [11, 28] to measure the response of the patient given a cue. First a smile appears in the centre to induce a fixation, then an auditory aid accompanies a visual cue to one side of the screen along the horizontal axis followed by an attractive target (the colourful image of an animal). The cue can appear in the same position of the target (valid anticipation), in the opposite side (invalid anticipation), in both sides (double) or not appear at all (baseline). If the child inhibits correctly then we expect the valid modality to show faster reactions time compared to baseline, while the invalid modality should be slower than both. A summarising picture can be found in Figure 1b. The smile appears for 1.5 s followed by the cue that lasts 100 ms, then after a 100 ms delay the target appears and remains on the screen for 1 s.
• **Memory task**: a task inspired by [17] that checks the predisposition of children to fixate on novel stimuli. Two pictures are shown to the patient for 1 s, followed by a blank screen lasting 500 ms, then the pictures are presented again with one of them substituted with a new image and remain on the screen for 3 s. We use three types of differences to measure what the child is able to identify: colour, shape, and faces. The face pictures are taken from the London Set Dataset [32]. A summarising picture can be found in Figure 1c.

• **Social orienting task**: the aim of this task is measuring if the children displays social responses to human stimuli, and is inspired by the previous similar studies [12, 31]. The patient is presented with two pictures for 5 s, one containing a human face and the other containing the front of a house, and the looking time to the former is measured. The face pictures are taken from the London Set Dataset [32] and the house pictures are taken from the DalHouses Dataset [33]. An example of how the task appears to the participant can be found in Figure 1d.

• **Face exploration task**: in this task the patient is presented with the image of a human face with a neutral expression viewed from the front and taken from the London Set Dataset [32] (see Figure 1e for an example). Telford et al. [12] showed how the gaze trajectories while observing a human face, and in particular the difference of looking time with respect to the eyes and the mouth could be influenced by premature birth. In total the face remains on the screen for 10 s.

Table 1
Order and duration of each task composing the test battery. The block described in the table is repeated four times, for a total duration of 223.2 s.

<table>
<thead>
<tr>
<th>Task</th>
<th>Task duration</th>
<th>Cumulative duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smile</td>
<td>1.5 s</td>
<td>1.5 s</td>
</tr>
<tr>
<td>Sensation</td>
<td>7.5 s</td>
<td>9 s</td>
</tr>
<tr>
<td>Attention</td>
<td>2.7 s</td>
<td>11.7 s</td>
</tr>
<tr>
<td>Attention</td>
<td>2.7 s</td>
<td>14.4 s</td>
</tr>
<tr>
<td>Attention</td>
<td>2.7 s</td>
<td>17.1 s</td>
</tr>
<tr>
<td>Attention</td>
<td>2.7 s</td>
<td>19.8 s</td>
</tr>
<tr>
<td>Smile</td>
<td>1.5 s</td>
<td>21.3 s s</td>
</tr>
<tr>
<td>Memory (colour)</td>
<td>4.5 s</td>
<td>25.8 s</td>
</tr>
<tr>
<td>Smile</td>
<td>1.5 s</td>
<td>27.3 s s</td>
</tr>
<tr>
<td>Memory (shape)</td>
<td>4.5 s</td>
<td>31.8 s</td>
</tr>
<tr>
<td>Smile</td>
<td>1.5 s</td>
<td>33.3 s s</td>
</tr>
<tr>
<td>Memory (face)</td>
<td>4.5 s</td>
<td>37.8 s</td>
</tr>
<tr>
<td>Smile</td>
<td>1.5 s</td>
<td>39.3 s</td>
</tr>
<tr>
<td>Social orienting</td>
<td>5.0 s</td>
<td>44.3 s</td>
</tr>
<tr>
<td>Smile</td>
<td>1.5 s</td>
<td>45.8 s</td>
</tr>
<tr>
<td>Social orienting</td>
<td>10.0 s</td>
<td>55.8 s</td>
</tr>
</tbody>
</table>

Each task is repeated four times and the test battery lasts approximately 223 s, which during the pilot study appeared to be a time frame where we can expect the the children to be able to maintain their attention. The timings of the singular tasks were kept as specified in the studies that inspired them. Gaze trajectories acquired during the test battery are parsed and a first set
of features can be computed following the original works that inspired the different tasks, as found in Section 3.

(a) The sensation task shows a smile moving horizontally in a sinusoidal pattern.

(b) The attention task. From top to bottom: valid, invalid, double, and baseline.

(c) The memory task. From top to bottom: colour, shape, and face differences.
5. Conclusions

In this study, we have reviewed current medical literature on the topic of eye movements and in particular the recent advances in the application of eye-tracking to young children and infants. We supplied a general guideline on how similar tests have been conducted: how the calibration procedure should be modified to address the needs of younger patients, how to avoid fussiness in the participant through a correct experimental setup, and how to optimise the quality of acquired data. Then, we reported which stimuli can be employed to monitor neuropsychological development in children. Finally, we presented a prototype transversal test battery that, by combining and shortening existing experimental paradigms, might supply information on the state of different cognitive functions in developing children.

Future works will consist in analysing the eye-tracking data obtained with the test battery by defining features and comparing different classifiers. This analysis could lead to insights on how to design the tasks, and may result in an improved test battery. Evaluation will be conducted with different metrics (accuracy, Area under the Roc curve, explained variance etc.) and, by integrating modern machine learning methods, we aim at improving current state-of-the-art results. Moreover, we will judge our results by their clinical utility as per hospital requirements,
and compare evaluation metrics to similar studies on adults.

Acknowledgments

The authors would like to thank all the parents and the children for their participation in the pilot study. This research was partly founded by H20202 MSCA-ITN project PARENT, Grant Agreement N° 956394.

References


