Aalto University
School of Science

Eero Ahtola

Eye Tracking Based Methods for Evaluation of Infants' Visual Processing

Thesis for the Degree of Licentiate of Science
Helsinki, 6 January 2015

Thesis instructor: Docent Sampsa Vanhatalo
Supervising professor: Prof. Risto Ilmoniemi

Aalto University
School of Science

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of licentiate thesis

Author	Eero Ahtola
Title of thesis	Eye Tracking Based Methods for Evaluation of Infants' Visual Processing
Department	Department of Neuroscience and Biomedical Engineering
Field of research	Biomedical Engineering
Supervising professor	Prof. Risto Ilmoniemi **Code of professorship** SCI017Z
Thesis advisor(s)	Docent Sampsa Vanhatalo
Thesis examiner(s)	Docent Ari Pääkkönen
Number of pages	55+11 **Language** English
Date of submission for examination	19.11.2014

Abstract

Cortical visual processing and mechanism under eye movements and visuospatial attention undergo prominent developmental changes during the first 12 months of infancy. At that time, these key functions of vision are tightly connected to the early brain development in general. Thus, they are favourable targets for new research methods that can be used in treatment, prediction, or detection of various adverse visual or neurocognitive conditions.

This Thesis presents two eye tracker assisted test paradigms that may be used to evaluate and quantify different functions of infants' visual processing.

The first study concentrates on the analysis of the gaze patterns in classic face–distractor competition paradigm known to tap mechanisms under infant's attention disengagement and visuospatial orienting. A novel, dynamic analysis metric is devised to measure allocation of attention between the presented stimuli over a given period of time. In further evaluation, the metric is shown to be sensitive to developmental changes in infants' face processing between 5 and 7 months of age.

The second study focuses on the visual evoked potentials (VEPs) elicited by orientation reversal, global form, and global motion stimulation known to measure distinct aspects of visual processing at the cortical level. To improve the reliability of such methods, an eye tracker is integrated to the recording setup, which can be used to control the stimulus presentation to capture the attention of the infant, and in the analysis to exclude the electroencephalography (EEG) segments with disoriented gaze. With this setup, VEPs can be detected from the vast majority of the tested 3-month-old infants (N=39) using circular variant of Hotelling's T2 test statistic and two developed power spectrum based metrics.

After further development already in progress, the presented methods are ready to be used clinically in assessments of neurocognitive development, preferably alongside other similar biomarker tests of infancy.

Keywords Eye tracker, gaze tracking, infant, visual processing, attention, face preference, visual evoked potential, EEG, analysis method

Aalto-yliopisto
Perustieteiden
korkeakoulu

Aalto-yliopisto, PL 11000, 00076 AALTO
www.aalto.fi

Lisensiaatin tutkimuksen tiivistelmä

Tekijä Eero Ahtola		
Työn nimi Menetelmiä lasten näkötiedon käsittelyn arvioimiseksi katseseurannan avulla		
Laitos Neurotieteen ja lääketieteellisen tekniikan laitos		
Tutkimusala Lääketieteellinen tekniikka		
Vastuuprofessori Prof. Risto Ilmoniemi		**Professuurikoodi** SCI017Z
Työn ohjaajat Dos. Sampsa Vanhatalo		
Työn tarkastajat Dos. Ari Pääkkönen		
Jätetty tarkastettavaksi 19.11.2014	**Sivumäärä** 55+11	**Kieli** englanti

Tiivistelmä

Näkötiedon käsittely aivokuorella sekä silmänliikkeiden ja visuospatiaalisen tarkkaavaisuuden mekanismit kehittyvät valtavasti lapsen ensimmäisen 12 elinkuukauden kuluessa. Nämä näön avaintoiminnot ovat tiukasti sidoksissa aivojen yleiseen varhaiskehitykseen, jonka vuoksi ne ovat suotuisia kohteita uusille tutkimusmenetelmille käytettäväksi visuaalisten tai neurologisten ongelmien hoidossa, ennustuksessa tai löytämisessä.

Tämä työ esittelee kaksi katseenseurantaa hyödyntävää koeasetelmaa, joita voidaan käyttää lasten kortikaalisen näkötiedon käsittelyn arvioinnissa ja kvantifioinnissa.

Ensimmäisessä tutkimuksessa kehitettiin mitattujen katsekuvioiden analyysiä klassisessa kasvokuva–distraktori-koeasetelmassa, jonka tiedetään koskettavan lasten tarkkaavaisuuden vapauttamiseen ja katseen siirtoon liittyviä mekanismeja. Työssä kehitetyllä laskennallisella mittarilla pystytään määrittämään tarkkaavaisuuden jakautuminen ruudun keskellä ja reunalla esitettyjen ärsykkeiden välillä haluttuna aikana. Jatkotarkastelu osoittaa mittarin olevan herkkä kasvokuvien käsittelyn kehityksen muutoksille 5 ja 7 kuukauden ikäisten lasten välillä.

Toinen osatyö keskittyy näkötiedon kortikaalista käsittelyä heijastavien, suunnan kääntämisen, globaalin muodon tai liikkeen tuottamien näköherätepotentiaalien mittaamiseen ja analyysiin. Parantaakseen menetelmien luotettavuutta laitteistoon liitetään silmänliikekamera, joka mahdollistaa sekä ärsyketoiston ohjaamisen lapsen tarkkaavaisuuden mukaisesti että kerätyn aivosähkökäyrän karsimisen niiltä osin, jolloin lapsen katse oli harhautunut esityksestä. Käyttäen muunnelmaa Hotellingin T2 statistiikasta ja kahta työssä kehitettyä, tehospektriin pohjautuvaa analyysimenetelmää herätevasteet pystytään löytämään valtaosalta testatuista 3 kuukauden ikäisistä lapsista (*N*=39).

Meneillään olevan jatkokehityksen jälkeen esitetyt menetelmät ovat valmiita kliiniseen käyttöön neurokognitiivisen kehityksen arvioinnissa muiden vastaavien biomarkkeritutkimusten rinnalla.

Avainsanat Silmänliikekamera, katseenseuranta, lapsi, näkötiedon käsittely, tarkkaavaisuus, kasvopreferenssi, näköherätepotentiaali, aivosähkökäyrä, analyysimenetelmä

Acknowledgements

First I want to thank my instructor, Sampsa Vanhatalo, for providing me the opportunity to work on this interesting research topic and contributing endless number of ideas throughout the process.

I want to express my gratitude to supervising professor Risto Ilmoniemi at Department of Neuroscience and Biomedical Engineering at Aalto University School of Science and Sauli Savolainen, Chief Hospital Physicist at Helsinki University Central Hospital.

I want to thank Susanna Stjerna for performing all the data recordings included in this work and also other co-authors of the publication: Jukka Leppänen, Santeri Yrttiaho, and Charles Nelson. Regarding the second yet unpublished study, I would like to thank Nathan Stevenson for assistance with the signal processing, Juha Voipio for help with the eye tracker synchronization, and late John Wattam-Bell for providing the original scripts for VEP stimulus presentation and T2circ data analysis.

Finally, I would like to thank my family for all their support that I truly appreciate, and most of all, my dear Anna for all the love, patience, and encouragement during the past years.

Financially work on this Thesis was supported by Helsinki University Central Hospital, Finnish Foundation for Pediatric Research, and Sigrid Jusélius Foundation.

Helsinki, 6 January 2015
Eero Ahtola

Contents

List of Abbreviations

AD	Analog-to-Digital
AOI	Area of Interest
DP	Disengagement Probability
DR	(Average) Dynamic Response
DT	Disengagement Time
E	(Spectral) Energy
EEG	Electroencephalogram
EP	Evoked Potential
FDR	False Discovery Rate
fMRI	Functional MRI
GF	Global Form
GM	Global Motion
IQR	Interquartile range
ISCEV	International Society for Clinical Electrophysiology of Vision
IR	Infrared
LGN	Lateral geniculate nucleus
M	Mean
MD	Mean Deviance
MRI	Magnetic Resonance Imaging
OR	Orientation Reversal
PCCR	Pupil Center and Corneal Reflections (eye tracking)
PR	Pattern Reversal
PSD	Power Spectral Density, Power spectrum
SC	Superior colliculus

SD	Standard deviation
SNR	Signal-to-noise ratio
SSVEP	Steady-state VEP
T2circ	Circular variant of Hotelling's T2 test statistic
TTL	Transistor–transistor logic
V1	Primary visual area
V2–V5/MT	Visual association areas
VEP	Visual Evoked Potential

Author's Contribution

This Thesis consists of two interlinked studies. The first work presenting dynamic eye tracking based metrics for common face–distractor competition paradigm has been published in an established peer-reviewed journal.

Publication 1: Ahtola, E., Stjerna, S., Yrttiaho, S., Nelson, C. A., Leppänen, J. M., and Vanhatalo, S. (2014). Dynamic eye tracking based metrics for infant gaze patterns in the face–distractor competition paradigm. *PloS One,* 9(5).

The second part of the Thesis, presenting a study of eye tracking integrated visual evoked potentials in Chapter 5, is yet to be published.

The author's contribution to the studies has been significant throughout the process from planning and implementation of the laboratory setup to the final data analyses. He optimized the inherited test paradigms for the new environment and was responsible for the technical quality of the measurements. He created the novel analysis methods and performed all presented data calculations. All MATLAB and E-Prime codes needed in the stimulus presentation and in the analyses were programmed by the author. Finally, he was the principal writer of the publication mentioned above.

1. Introduction

1.1 Motivation

Babies are not born with their full visual abilities. The different aspects of cortical visual processing, as well as the ability to control eye movements accurately and engage (or disengage) visual attention on targets, take time to develop. The most prominent physiological changes in these mechanisms occur during the first 12 months. At that time the development of these key functions is tightly connected to the early brain development in general. Thus, they provide a unique functional window into the (typically or atypically) developing infant brain. (Braddick and Atkinson 2011)

Naturally, there is constant need for ever more reliable and practical research methods and study paradigms targeted to evaluate the key cortical functions such as visual processing. The new methods could be used in multiple ways in treatment, prediction, or detection of various adverse visual or neurocognitive conditions. Potential applications in fields of pediatrics and child neurology include the assessment of infant neurocognition, perinatal brain injuries, very premature infants (Atkinson et al. 2002), and various neurodevelopmental disorders for instance. (Braddick and Atkinson 2011)

The challenges typically associated with almost all infant studies are, however, holding back the ultimate breakthrough of some experimental, yet highly promising, methods into clinical routines. Lack of subject's co-operation prevents one from using highly demanding task designs, and babies cannot be asked to perform 'circus tricks'. The simpler test design is often the more practical and helps to minimize the unwanted external effects possibly contributing to the test result.

Comparing to adult studies, infant measurements and recordings (such as EEG) are also generally more vulnerable to artefacts, which are typically caused by excessive movement, eye blinks, etc., depending on the nature of the experiment. These challenges (co-operation and artefacts) are closely linked to the infant's level of vigilance. Fatigue and anxiety may also effectively constrain the duration of the test, especially if the paradigm required any active participation.

Even if the artefact pitfalls were successfully escaped, the natural inter- and intra-subject variabilities are generally high in infant studies. The results are also highly dependent on the age of the subject and obviously differ drastically from adults taking the same tests. All these things must be taken into account

within the whole diagnostic chain including measurements, data analysis, and result interpretation.

Because of the issues described above, the number of suitable physiological mechanisms that can be conveniently studied is regrettably limited compared to adult studies. Likewise, the repertoire of different test paradigms in clinical routine is low although the requirements of accuracy and sensitivity are often compromised for the sake of practicality.

For this challenging territory, this Thesis introduces a set of test paradigms measuring visual cortical processing, attention, and face preference of infants aged from 3 to 7 months. What unites these methods is the use of eye tracking – a modern technique that can be used to investigate patterns of visual attention during a specific task by continuously recording subject's point of gaze on the screen. These gaze data may be helpful in three ways: i) providing feedback and interactive control during the test sessions, ii) allowing optimization and automation of the present data analysis methods, and iii) paving way for the development of novel analysis tools and metrics.

The versatility of the eye tracking advantages is well manifested in the two presented studies that constitute this Thesis. The first work is concentrated specifically on the analysis of the measured gaze patterns to characterize the functioning of visuospatial attention. On the other hand, the focus of the second work is in the analysis of EEG (electroencephalogram), and the eye tracking is used in a minor, yet crucial, role to control stimulus presentation and analysis in a visual evoked potential experiment. Notably, these objectives are very generic, and thus the developed ideas could be easily exploited in other similar experiments as well.

As a technology, eye tracking is not that new an invention. Yet, the technique is constantly developing, and the devices are nowadays more accurate, more usable, and more affordable than in the past. Consequently, the number of possible eye tracking applications has rocketed as the technique has broken into new research areas. In the field of clinical neurophysiology studies exploiting eye tracking are still rare. Nevertheless, there should be no reason to reject these new opportunities just because an eye tracker does not belong to the standard equipment of the laboratories, as long as the necessary steps of background research and validation have been adequately arranged before the entry into the clinical diagnostics. One goal of this Thesis was to bridge this translational gap between the research and the clinical routines and to take the developed tools closer to the actual use in clinical neurophysiology unit. (Holmqvist et al. 2011)

1.2 Outline

This Thesis is composed of seven chapters. Chapter 2 presents a brief review of the general physiological background behind the work concentrating on the visual system and its development.

Chapter 3 reviews the principles of the eye tracking technique and describes the laboratory setup that was designed and used in the experiments including fixation based stimulus presentation and synchronization with EEG system.

Next, the two studies included in this Thesis are presented separately. Chapter 4 outlines the key parts of the published article “Dynamic eye tracking based metrics for infant gaze patterns in the face–distractor competition paradigm” (Ahtola et al. 2014) adding also further discussion on the subject.

Chapter 5 is dedicated to the eye tracker guided pattern VEP study performed for 3-month-old babies to measure their visual cortical processing. Compared to the previous chapter it is described in more detail since the method is not yet published.

Chapter 6 presents the discussion interlinking the two studies. The focus of the discussion is on future objectives and issues that still need to be addressed before the potential entry into clinical diagnostics.

Chapter 7 draws up the conclusions.

1.3 Aims of the study

The objective was to further develop two existing test paradigms: pattern VEPs measuring visual cortical processing and face–distractor competition paradigm measuring visuospatial attention. In the previous works, these tests have yielded interesting results with infants aged from 3 to 7 months and shown promise for becoming early biomarkers of later visual and other cognitive adversities. Our idea was to implement eye tracker guidance for the paradigms in order to overcome the challenges caused mainly by the poor co-operation of the infants that could prevent the methods to be used clinically.

The specific aims of the study were to:

1. Design and implement the laboratory environment needed for the optimal use of the eye tracked guided tests.
2. Develop study paradigms that take the advantage of the real time information from the eye tracker to allow dynamic adaptation of the recording session to the attention of the infant.
3. Study and evaluate alternatives for previously used data analysis methods, also taking into account the gaze data available from the tracker.

2. Physiological background

Eye and retina

Retina is the inner coat of the eyeball that lays out the light-sensitive projection surface for the entering light to form an inverted image of the visual world. Upon that, the light has been regulated by the size of the pupil as well as refracted in the cornea and in the lens that focuses the image on the retina at different distances (see Fig. 1).

The specialized photoreceptor cells, rods and cones, at the neural layer of retina absorb the incoming photons and convert them into receptor potentials through cascades of electrochemical events. The rod cells have a low light threshold allowing us to see in dim light, whereas the cones can process the colors and are responsible for our sharp central vision (at the fovea). The output from the photoreceptors undergoes considerable further processing by other neurons of the retina and finally takes the form of action potentials in ganglion cells whose axons form the optic nerve at the beginning of the visual pathway.

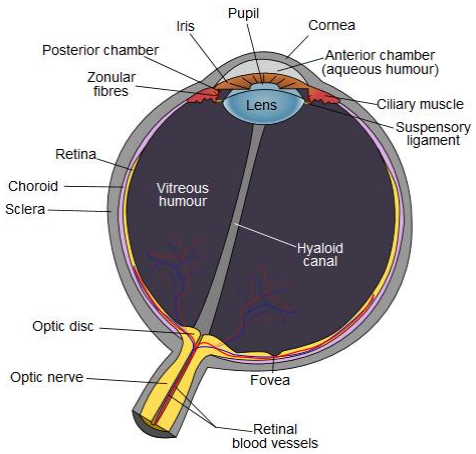

Figure 1. Anatomy of the human eye. The image is adopted from Wikimedia commons (commons.wikimedia.org)

Visual pathway

The visual stimulus terminates at the retinal receptors, but the visual information is carried further by neural impulses in the optic nerves. The nerves partially cross at the optic chiasm; as a result each side of the visual field is ultimately projected on the contralateral hemisphere of the brain. The output axons from the chiasm form the optic tract which terminates in a layered structure

called the lateral geniculate nucleus (LGN) located in the thalamus – the general relay station for most sensory impulses before reaching the cerebral cortex. The neurons are first sorted by the synapses at the LGN so that nearby cells receive inputs from the same region of the visual field (retinotopic mapping), and then they fan out through the white matter of the brain as optic radiations terminating ultimately at the primary visual area. (Tortora and Grabowski 2003)

The pathway described above is, of course, just a simplified model. It is thought that the visual signals are processed by at least three separate systems, each with its own functions devoted to specific properties of the perceived stimulus (shape, color, spatial organization, and movement). Also a minor amount of neurons connect to the visual cortex through the pulvinar nucleus and superior colliculus (SC) constituting a feedback mechanism related to automatic control of the eye movements (e.g., following of a moving object). In addition, SC is connected to frontal and inferior temporal cortical areas that supply the input from the neurons processing movement (see Fig. 2). (Tortora and Grabowski 2003; Nicholls et al. 2001)

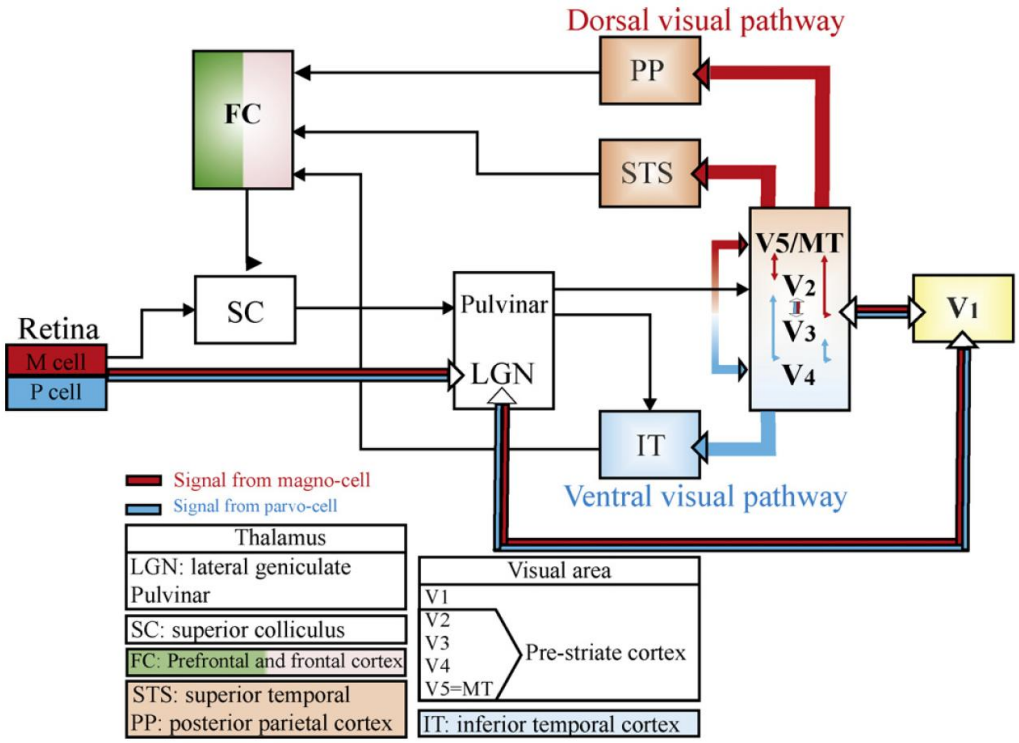

Figure 2. Illustration of a classic visual pathway model. After stimulation of the retina the visual neural signals are conveyed first through the LGN to primary visual cortex and then further to other brain areas. The figure is adopted from Vialatte et al. (2010).

Cortical processing

The primary visual area (also known as V1 or the striate cortex) in the occipital lobe of the brain receives neural impulses from its ipsilateral LGN in retinotopic order. The nearby visual association areas (extrastriate regions V2–V5) receive sensory impulses from V1 and straight from the thalamus. Generally, they are responsible for the complex integrative processing that binds new visual experiences to the past and thus are essential for recognizing and evaluating what is

seen. For instance, a person with damaged primary visual area would be at least partially blind, whereas damage to a visual association area may not necessarily hinder the vision, but causes problems when trying to recognize friends. (Nicholls et al. 2001)

According to a classic model (Fig. 2), V1 connects to extrastriate areas by two major pathways that may be identified already at the retinal (magno- and parvocells) and LGN levels and are associated with different functions. The dorsal processing stream propagates from V1 to V2, V5 (also known as MT area), and finally to posterior parietal cortex. It is often called the "where pathway" due to its relations to motion and depth processing, object localization, and visual tracking. The complementary ventral stream, going from V1 to V3, V4, and inferior temporal cortex, is in turn called the "what pathway" because of its role in object identification and detection of spatial contrast and form. (Vialatte et al. 2010; Goodale and Milner 1992; Braddick and Atkinson 2011)

Eye movements and visual attention

Three pairs of extrinsic eye muscles, innervated by cranial nerves, control the horizontal, vertical, and torsional movements of each eye. The movements that uniquely determine the direction of the gaze and the borders of the visual field are coordinated in relation to the processed visual information by complex circuitry involving brain stem and cerebellum. In normal conditions, the two eyes function binocularly. It means that their movements are synchronized, and that they look at the same location from slightly different angles, providing visual signals needed to process the depth information of the scenery. (Tortora and Grabowski 2003; Holmqvist et al. 2011)

The gaze movements are usually divided into two main categories: fixations and saccades. The fixations are longer pauses in the eye movement, lasting generally for 200–500 ms, typically related to exploring of content-rich scenes. Saccades, in turn, are faster movements, lasting typically 10–100 ms, needed in the repositioning of the fixation in a visual scene. There exist eye movements of other kinds as well. For example, fast tremors and microsaccades and slow drifts occur involuntary during the seemingly stationary fixations and prevent the fading of the visual perception because of neural adaptation. (Duchowski 2007; Holmqvist et al. 2011)

We cannot register everything that happens in front of our eyes, but tend to focus (unconsciously or consciously) only on a selected portion of the visual field using the processing of visual attention. Psychologically, attention is a complex concept, but usually two separate attentional mechanisms are distinguished. Using the overt visual attention process, we move our eye to direct the gaze to the objects of interest, whereas covert attention means that our mind is focused on the peripheral areas of our visual field independent of the eye movements. (Tobii Technology AB 2010)

In the context of the studies within this work, we cut corners slightly by assuming that generally the consecutive fixations follow the shifts of the attention, and thus the gaze coordinates measured by eye tracker can be interpreted as markers of the (overt) visual attention focus.

Development of the visual processing

The visual system of a newborn baby is not fully functional from the beginning. Infants learn to see and observe the environment gradually over first two years with most prominent physiological changes occurring during their first 12 months.

At the age of 3 months, the development of the subcortical oculomotor loop and strengthening of the muscles surrounding the eyes have provided infants the control of their gaze, making the fixations to targets possible. The ability to disengage attention in a 'competition' situation, understood to reflect functioning of a specific cortical system and connection to the superior colliculus, however, has just started to develop. Yet, eye movements are mainly coordinated, and processes related to visual attention and visual searching through saccades have initiated. (Braddick and Atkinson 2011; Hunnius et al. 2006)

The eyes of a 3-month-old infant are already operating together, but the visual acuity is still poor, mostly because of the physical limitations of the eye and still developing retina. The functional processing of cortical ventral and dorsal streams has started to emerge. Though, the neural basis of the streams is still far from its mature anatomical organization. The developmental stages of the two pathways are still partly unclear, but the consensus is that the dorsal processes mature earlier than the ventral functions. For example, orientation-selective processes may be registered after 6 weeks of age, even though the intracortical interactions, on which these depend, continue to develop at least over the first 6 months. Additionally, although the local directional selectivity for motion develops later than that for orientation, the processes of integrative global motion specificity start to emerge faster compared to global form selectivity, respectively at ages of 2 and 3 months. Infants also learn to distinguish colors of the stimuli at the same of age, indicating the development of the ventral pathway. (Braddick and Atkinson 2011)

During the next few months, the core attention processes and preferential attention to social cues continue to develop, establishing foundations for cognitive and social brain networks. (Hunnius 2007) Infants' tendency to focus their attention on faces increases gradually between 3 and 9 months of age. (Frank et al. 2009) At the age of 5–7 months, the ability to visually control reaching and grasping has improved significantly with the developed processing of binocular depth information playing an important role in initiating and guiding the movements. At the same time, infants start to explore visually the environment and sense the size of objects located near and far. Also the acuity begins to reach the normal adult levels around this stage. (Braddick and Atkinson 2011)

3. Eye tracking system

3.1 Eye tracking technique

The modern, remote (meaning the head is allowed to move) eye trackers generally use a technique called pupil center and corneal reflections (PCCR) method. It is based on the interpretation of digital video capture of infrared (or near IR) light illuminated by the tracker and reflected from the cornea that covers the outside of the eye. Theoretically, the tracking could be done from the pupil information alone, but corneal reflections provide an additional reference point for the calculations resulting in improved accuracy. Technically, eye tracking is a process of three phases: image acquisition, image analysis, and gaze mapping. (Holmqvist et al. 2011)

PCCR eye tracking devices can operate in two modes, bright pupil and dark pupil, which differ by the position of the illuminator diodes in respect to the optics. In the former, the illumination source is located close to the optical axis of the imaging device, which causes the pupil to appear bright in the acquired video image. In the latter, the IR source is placed further away from the optical path resulting in the image with a black pupil. (Holmqvist et al. 2011)

The eye image captured using the bright mode is simpler to process because of its better contrast between the iris and the pupil, and hypothetically it should yield more robust tracking results (no data available to verify this). The dark pupil mode is more reliable in environments with generous ambient lighting that decreases the pupil size. In addition, subject's ethnicity may also affect the superiority of the two tracking modes. The device used in this work, Tobii T120 (Tobii Technology AB, Stockholm, Sweden), provides both of these techniques and is capable of switching automatically between the modes to yield optimal results (Tobii Technology AB 2010; Holmqvist et al. 2011)

After the image of the eye is grabbed, the goal of the image analysis phase is to detect the positions of the eyes and segment the pupils and corneal reflections. There exist multiple ways to perform this object recognition, and the details of the image processing algorithms used in the commercial products are rarely available in public. Most likely the segmentation is done either by feature based bottom–up methods (intensity thresholding, clustering, and edge detection) or model based top–down methods (deformable templates and pattern matching). (Holmqvist et al. 2011; Hammoud 2008)

Theoretically, after the image segmentation the necessary geometrical information (eyes, cameras, target) is available for the calculations to reveal the gaze

direction and position on the target. In PCCR, however, it is more practical to exploit the additional reference point from the corneal reflection. When the eye moves, the pupil center moves faster than the corneal reflection, and the gaze mapping can be calculated from the vector between these points. This shortcut requires some a priori information gathered in a calibration procedure that measures the characteristics of the subject's eyes and integrates them into an internal 3D eye model, essentially providing the link between the measurements and the screen geometry. (Holmqvist et al. 2011; Tobii Technology AB 2010)

3.2 Calibration

In this work, each study session began with the calibration that followed the semi-automated workflow within the Tobii Studio software (version 2.2.8) of the eye tracker. The protocol proceeded by showing the infant an audiovisual animation sequentially at five locations on the screen. The outcome of the calibration procedure was read from a qualitative output showing the offset between measured gaze points and the center of the given calibration spot (see Fig. 3). Hence, the final acceptance of calibration was based on subjective assessment by the examiner.

Since our present paradigms did not rely on tracking accuracy at millimetre level, we found it most practical to accept calibrations which were able to track all 10 calibration points (5 for each eye). If one or more marks were missing completely, the calibration procedure was repeated until all (or at least 9) marks were obtained.

Figure 3. Eye tracker calibration results. Example screenshot from a calibration session demonstrating how Tobii Studio software shows the offset of gaze during calibration procedure. The figure is adopted from Ahtola et al. (2014).

3.3 Laboratory design and requirements

Although modern remote devices allow eye tracking to be performed almost anywhere, some general requirements can be compiled for an eye tracking laboratory to provide optimal accuracy and usability.

The laboratory space should be big enough to comfortably fit the equipment, the operator, and the participant attending the experiment. In case of infant studies, it is a good idea to reserve space also for the parent. For instance, in the

studies of this work, the infants were placed sitting in a baby carrier attached on their parent's chest (see Fig. 4).

At minimum, the eye tracking setup consists of a control PC, the eye tracker, and a display for the stimulus presentation. In our case, 17-inch TFT monitor (refresh rate 60 Hz, max. resolution 1280 x 1024 pixels) was an integral part of the eye tracker device (Tobii T120). A second computer could be added to allow the real time monitoring of the session on a secondary display showing the presentation with the superimposed, live gaze data. This was not needed in our experiments, but instead an additional EEG device, required for the VEP studies, was placed in the vicinity and synchronized with the eye tracker.

A silent location or acoustic insulation helps to minimize the influence of external distractions. For the same reason, the ambient light in the experiments was kept dim, not completely dark though, as maximally dilated pupils reduce the data quality (Holmqvist et al. 2011). The participants (baby and caregiver) were separated from the experimenter by light walls (Fig. 4). Despite the isolation, the operator should have the possibility to observe the infant during the session. In our laboratory, this was achieved through a window with one-way transparency created by tinting the window and having higher lighting level on the baby's side.

The lightning conditions should also be considered due to the possible optic artefacts. Reflective surfaces and additional sources of IR-illumination (avoid windows, prefer fluorescent lamps) may cause imprecision and data loss to the recording. Other potential artefact sources include external vibrations (traffic) and electromagnetic fields (MRI). (Holmqvist et al. 2011)

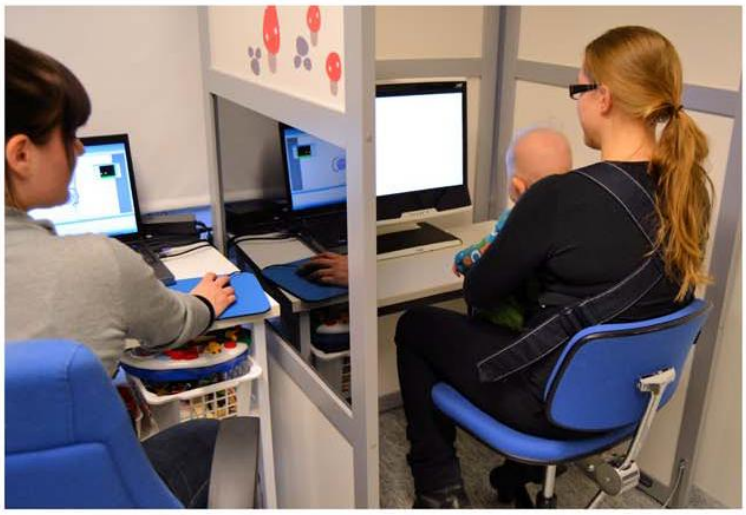

Figure 4. Eye tracking setup. The photograph shows the laboratory setup where infant is placed into a baby carrier attached to the chest of the caregiver (right), and the experimenter (left) is sitting behind a light wall with one-way transparency (not visible on the photograph) that allows observation of the infant during the test. The figure is adopted from Ahtola et al. (2014).

Tobii T120 eye tracker can be used at 50–80 cm from the eyes, and may follow head movements within a window of 30 x 22 cm (70 cm from the screen). According to the manufacturer, the system provides a typical spatial accuracy, i.e.,

difference between measured and actual gaze directions, of 0.5 degrees, which corresponds to 4.4–7.0 mm on screen at the allowed viewing distance. This would be sufficient to fulfill the requirements of the applied paradigms. In order to verify the accuracy measures in the context of our laboratory setup, however, a 'practical' assessment of the spatial accuracy was performed as explained in the supplementary data (Fig. S2) of Ahtola et al. (2014).

3.4 Structure of the eye tracking data.

Our 120-Hz eye tracking system has sampling resolution of 8.3 ms, which allows us to detect the eye movements of the two basic categories: fixations (200–500 ms) and saccades (10–100 ms). Other types of eye movements (tremors, drifts and microsaccades) escape detection of the eye trackers that work without a head restrain, because they would require considerably higher spatial and temporal resolution. (Duchowski 2007)

Eye tracking results are often interpreted through graphical scanpath illustrations where fixations are depicted by dots (dot size indicating the associated fixation time) and saccades by lines between the consecutive fixation spots. These objects are extracted through spatial and temporal 'fixation filters' reducing the amount of collected data. Nevertheless, in order to optimize both spatial and temporal accuracy, the analyses in this work were based on the actual raw data instead. (Tobii Technology AB 2010)

The raw eye tracking data were written into an ASCII file containing multiple time series sampled at 120 Hz. These series include i) *x*- and *y*-coordinates for the point of gaze on the screen, ii) timestamps for each data sample at microsecond accuracy, iii) 'validity codes' for each eye indicating the reliability of tracking at each time point (out of codes 0–4 we only used samples with validity code 0 or 1, which were taken to indicate technically reliable gaze tracking), and iv) additional index that characterizes the exact timing of specified events in the stimulus presentation.

Gaze movements could also be expressed in visual angles (α) determined by the equation

$$\tan\frac{\alpha}{2} = \frac{x}{2d}, \qquad (1)$$

where *x* is the size of the target and *d* is the viewing distance (the equation holds if target is located close to the central line of sight). In this work, however, it was simpler to stick with the screen resolution dependent pixel units due to their benefits in the Area of Interest (AOI) -based analyses.

3.5 Fixation based stimulus presentation

In the experiments of this Thesis, all stimuli were presented electronically using E-Prime software (version 2.0.8.22, Psychology Software Tools, Pittsburgh, PA)

and E-Prime Extensions for Tobii (version 2.0.1.5) interfacing with the eye tracker hardware.

In order to ensure that the starting points for successive trials or stimulus sequences were always equal, a special E-Prime script was programmed by the author to control the presentation. Consequently, each trial began by first attracting the child's attention to the center of the screen using simple audiovisual animations, for example, a gradually expanding red circle with recurring sound. The stimulus presentation was programmed to start automatically only after the eye tracking device had recorded a 600-ms-long segment of continuous fixation onto the predefined 'fixation area' (diameter 4.2°) around the animation stimulus (see later in Fig. 6B). This 'fixation counter' was reset to zero whenever the point of gaze hit outside the defined AOI or could not be tracked.

3.6 Integration with EEG

The VEP paradigm presented in Chapter 5 required an EEG system to record the brain responses evoked by the visual stimulation. As the computational operations on two separate devices may not necessarily take predictable amounts of time, the EEG and eye tracker gaze data had to be synchronized. In our laboratory, this was achieved using a self-built, battery powered light sensor that was attached on the bottom right corner of the eye tracker screen with a piece of double-sided tape.

The current passing through the photodiode component (BPW34 with 100 ns rise and fall times) depended on the brightness of the screen on that location (active size 3x3 mm). In the circuitry, the voltage over the diode was connected to the input of an operational amplifier (LF411) used as a comparator. The output voltage of the op-amp was either -100 µV or 100 µV depending on whether the screen intensity exceeded a preset threshold level that can be adjusted using an appropriate resistor. Finally, the resulting voltage was fed to a spare bipolar input of the EEG amplifier so that all the observed brightness changes would be registered aside the actual EEG recording.

The events in the eye tracker presentation could now be integrated with the EEG by addition of graphical, black and white (off/on) marker boxes to the corner of the presentation. In our VEP paradigms the stimuli consisted of two states reversing at the rate of 4 Hz. Each stimulus reversal was marked to the presentation yielding a square wave output signal where rises and falls indicated changes between the stimulus states (see Fig. 5).

The actual synchronization was performed later, offline in MATLAB environment (version R2010a, The MathWorks, Natick, MA) together with other signal analyses. In the developed semiautomatic reversal detection algorithm programmed by the author, the trigger signal x_n was transformed by calculating the difference between two moving average windows using equation

$$y_n = \frac{1}{M}\left(\sum_{k=1}^{M} x_{n-k} - \sum_{k=1}^{M} x_{n+k} \right), \qquad (2)$$

where y_n is the transformed signal and *M* is the size of the averaging window (in samples) that was adjusted to match the assumed trigger interval. The Locations of the local minima and maxima (reversal moments) were easy to pick from this smoothed signal (depicted with red in Fig. 5) with MATLAB's own peak detection function.

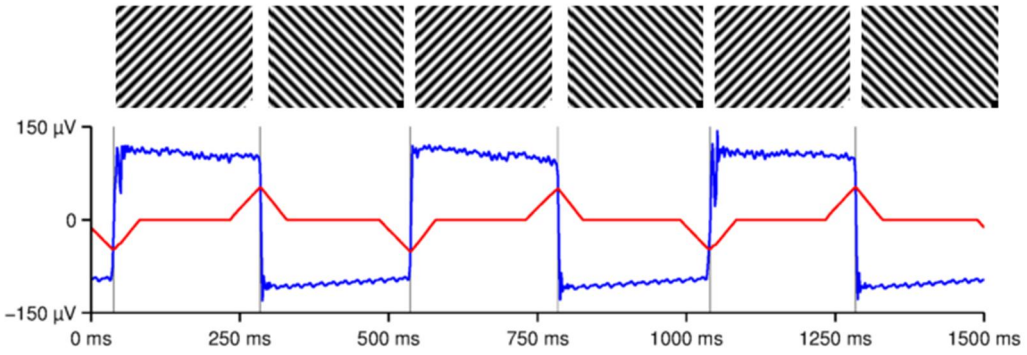

Figure 5. Synchronization of EEG recording and eye tracker presentation. A light sensor is attached on the corner of the eye tracker screen and its output is connected to a bipolar AC input of the EEG amplifier. The output voltage of the sensor (blue) varies between -100 µV and 100 µV depending on the marker brightness (black or white). The raw signal is further processed to yield a smoother signal (red) where each peak (detections marked with grey verticals) indicates a reversal in the stimulus presentation illustrated at the top.

Occasionally, the trigger signal was found to be corrupted by the interferences typically caused by poor attachment of the light sensor. This caused us problems in the pursuit for a fully automated algorithm. Hence, a small interactive MATLAB program was created to take care of the potential issues in the synchronization. It provided user a graphical interface that allowed to i) adjust the parameters of the automatic detection algorithm on the fly, ii) easily correct any missing or redundant trigger markers (false findings) by manual mouse clicks, and iii) see supporting numerical information (amounts and timings) about the detection results.

Of course, there exist alternative ways to perform the transmission of the synchronization triggers. For example, TTL pulses could be sent through the parallel or serial port of the PC using E-Prime commands. However, the light sensor implementation has two clear advantages: i) it is synchronized to the events in the actual screen presentation and thus not affected by possible drifts and latency variations in the computer timing, and ii) it is compatible with all regular EEG amplifiers as no special trigger inputs are needed.

4. Eye tracker based measurement of attention and face preference

4.1 Motivation and background

This part of the Thesis focuses on the development of the eye tracking based assessment of infant cognition in the context of the classic face–distractor competition paradigm (also known as 'disengagement' or 'overlap' paradigm; Aslin and Salapatek 1975). Generally, the experiment consists of two visual stimuli presented successively with a short delay in between. The first centrally presented stimulus is an image of human face that aims to capture the infant's attention, whereas the second overlapping peripheral stimulus tends to distract it (more detailed description in Section 4.2.2).

The adapted paradigm is known to yield a fairly stereotypic gaze pattern in typically developing infants (Hunnius et al. 2006; Hunnius 2007); conventionally, studies have mainly concentrated on the latency and frequency of the gaze shifts between the stimuli. Although the current understanding of the neural bases of infants' visuospatial orienting is still limited, these measures allegedly reflect the functioning of the core components related to the ability to flexibly engage, disengage, and shift visual attention in response to changes in the environment (Posner and Petersen 1990; Johnson et al. 1991). These processes undergo rapid developmental improvements during the first months of life reaching apparent stability at around six months of age (Hunnius et al. 2006; Hunnius 2007; Matsuzava and Shimojo 1997; Blaga and Colombo 2006).

The use of face images as stimuli in the experiment means that the paradigm taps another important neurocognitive function of this age as well – face perception and preferential attention to social cues. According to the prior results in studies using the overlap paradigm, sensitivity to faces and facial expressions emerges between 5 and 7 months of age and is manifested as a delayed latency and reduced probability of gaze shifts, particularly when the face displays a fearful expression (Forssman et al. 2013; Leppänen et al. 2011; Peltola et al. 2013).

Being sufficiently global and targeted to mechanisms that provide fundamental basis for the development of higher cognitive functions, the measures of visual attention have been often proposed as potential surrogate markers of infant's neurocognitive status (Muñoz and Everling 2004; Green et al. 2009; Muñoz 2009). The emergence of the core attention processes and preferential attention to social cues during the first postnatal months provides foundations

for cognitive and social brain networks and may be critical in initiating the developmental process that leads to a full repertoire of human social skills (Johnson 2005). Hence, the presented method holds a strong promise for clinical use as well.

Infant gaze shifts in the overlap paradigm have been analyzed by using manual scoring of eye movements from video records (Peltola et al. 2013), by using electro-oculogram (Csibra et al. 1998), and more recently, by using eye tracking systems (Hunnius et al. 2006; Wass et al. 2011; Morgante et al. 2012). The eye tracking techniques have the advantage of offering the possibility for semi- or fully automated acquisition and analysis of eye movements at a high spatial and temporal resolution.

Goals of the study

The aim of this study was to exploit the new possibilities of the eye tracking by further developing the assessment of infants gaze behavior in the overlap paradigm. The first objective was to re-implement the eye tracking integrated overlap paradigm in clinical environment using automated presentation and data analysis protocols. The second target was to extend previous analyses focusing on infants' first gaze shifts (Aslin and Salapatek 1975; Leppänen et al. 2011) by exploring the dynamics of infants' gaze movements over a longer time period, covering the entire trial time, and ultimately develop a novel eye tracking based metric reflecting the cumulative allocation of attention to the central vs. peripheral stimulus.

4.2 Materials and Methods

4.2.1 Participants

The primary sample consisted of 13 infants tested at the Helsinki University Central Hospital (N = 13; 10 females; age range 7.1–8.0 months, mean 7.50). To test the metrics developed in the analyses of the primary sample, data from additional groups of 7-month-old (N = 32; 6.6–7.6 months, mean 7.0) and 5-month-old infants (N = 22; 4.7–5.6 months, mean 5.1) were also analyzed. The subjects of these additional datasets had participated in a similar, independent study in Boston Children's Hospital. The participants in all datasets were born full term and reported to have typical development. For a detailed description of the participant groups, see the supplementary data (dataset S1) of Ahtola et al. (2014).

4.2.2 Test procedure and stimuli

The test procedure followed the design of the classic face–distractor competition paradigm. It consisted of 32 successive trials that included two phases together lasting for 4000 ms (Fig. 6B). Each trial began by first attracting the child's attention to the center of the screen using simple audiovisual animations as described in Section 3.5; the actual stimulus presentation started automatically after the eye tracking device had reported a stable fixation onto the screen.

For the first 1000 ms, a face image was shown on the center of the screen. During the remaining 3000 ms, a peripheral stimulus (also called 'target') was added into the edge of the screen 10.2° away from the face, equiprobably on the left or right.

The trials alternated randomly within a set of four different face images, each followed by the presentation of a peripheral target (randomly selected from two alternatives). Both sets of face images (Fig. 6A) were copied from a prior study (Leppänen et al. 2010) and consisted of three different facial expressions (i.e., an image of a female model with neutral, happy, or fearful facial expression) and a sham face (i.e., phase-scrambled face that retained the amplitude and colour spectra as well as the contour of the face stimulus but was not identifiable as face). Each face stimulus was presented eight times with the only constraints being that the same face was presented no more than twice in a row, and the target was not shown more than three times in a row on the same side of the screen.

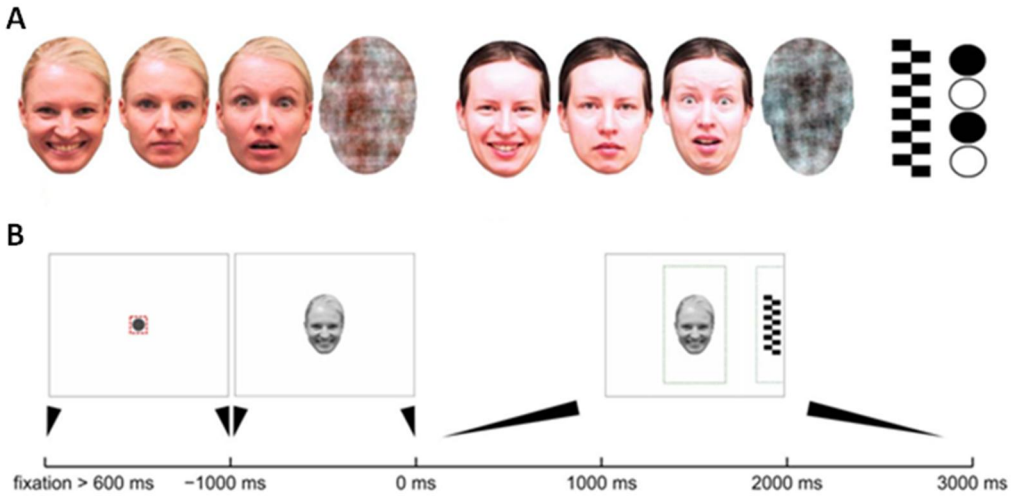

Figure 6. The face–distractor competition paradigm. A) The two sets of face stimuli used in the study. Every infant was presented with stimuli from one set only (chosen randomly). Both sets included a face with happy, neutral, and fearful expressions, as well as a noise (or 'non-face') stimulus. Peripheral distractors consisted of checkerboards and other geometrically simple high-contrast stimuli. B) Content of a single trial. Each trial was preceded by showing a fixation cue in the center of the screen to attract the infant's attention. The trial started only after the infant's fixation in the central area (small square around the dot) had lasted 600 ms. Then, the face stimulus was shown for 1000 ms, followed by adding the target to either side of the screen for the remaining 3000 ms. The dashed lines depict areas of interest (AOI) used later in the analyses. The AOIs were deliberately defined to be larger than the original stimuli to cancel out effects of measurement inaccuracies. The figures are adopted from Ahtola et al. (2014).

4.2.3 Conventional analysis parameters

Disengagement time (DT) and probability (DP) are the two most conventional measures in face–distractor competition paradigm (Peltola et al. 2009; Leppänen et al. 2010), and they were also used in this study to replicate the results of the previous works.

DT is defined as the latency from the target onset to the time when the participant shifts gaze from the face to the target. DP is the proportion of gaze shifts

out of all accepted trials (i.e., trials with a gaze shift and trials in which the gaze shift was not observed within 1000 ms after the onset of the peripheral stimulus). Trials with DT values below 160 ms were excluded from the analyses as 'anticipatory' due to falling below the absolute saccade reaction time (e.g., Kalesnykas and Hallett 1987).

Consistently with prior studies, all infants with ≥ 2 scorable DTs per condition were retained in the DT analysis and all infants with ≥ 3 scorable trials per condition in the DP analysis. For a more detailed description of the statistical procedures in this study, see Ahtola et al. (2014).

4.2.4 Developed dynamic metrics for gaze patterns

Time-varying characteristics of gaze – the novel dynamic response (DR).
The novel metrics developed within this work were based on quantification the amount of time that the infant gazed at the specific areas of the presentation. To this end, the two-dimensional gaze data coordinates of a trial were converted into binary time series for each applied Area of Interest (1 = gaze within the AOI; 0 = gaze outside of AOI, see Fig. 7A). The subsequent averaging of the AOI-transformed data of multiple trials resulted in time series constituting the dynamic response (DR) that reflects the time-varying characteristic of infant average gaze dynamics between the distractor and face AOIs (Fig. 7B).

The DRs for each participant were generated either separately for each stimulus to produce four stimulus-specific responses or jointly for all face stimuli yielding one face response averaged from larger (up to $N = 24$; phase-scrambled face excluded) number of trials. To highlight the differences in the dynamics, the trials where the gaze shift (face to target) did not happen were always omitted from the averaging. The resulting time series were finally median filtered in order to moderate abrupt, mostly technical artefacts.

It is common that a delayed disengagement causes a likewise delayed return from the target. Thus, if trials with different DTs were averaged for a DR, we would often end up blending trials that are in different phases. This would lead into distorted DR dynamics, even though the actual difference between the trials was solely in the disengagement times. To overcome this, the dynamic responses were time-locked to DT before the DR averaging by coaligning the trial data in time according to the moment when the gaze first reached the lateral distractor AOI (Fig. 7C). The data prior to DT were omitted from the analysis, but this data loss was accounted for in the separate DT analysis. This approach offered better representation of gaze changes where DT may be variable across the trials and was crucial especially when the focus of the gaze dynamics assessment was on the latter phases of the trial.

Computation of Mean Deviance (MD).
The dynamic response presents the general time course of the experiment conveniently in a graphical form. However, another novel metric was devised to quantify how much the calculated DRs deviate from the normative responses

during a given time window. This was determined as the average (in time) displacement of DR under the primary sample median DR, and expressed in percentiles of the DR distribution in primary sample (see later in Fig. 8B).

The best matching percentile for each DR value was calculated by interpolating the primary sample distribution (Fig. 8A). Because we were especially interested in downward deviations, all DR values above the 50th percentile (median) would count as 50 in the averaging. This procedure not only eliminates the possibility that displacements in the opposite directions could cancel each other out but also prevents from brief dips in the DR leading into excessive scores in cases where the DR is mostly over the median curve.

Despite the nonlinear operation, the resulting score was called 'Mean Deviance' (MD). It yielded a value of 50 if the participant's response was entirely within the upper 50 percentiles, whereas values decreasing below 50, in turn, implied increasing deviations from the normative range.

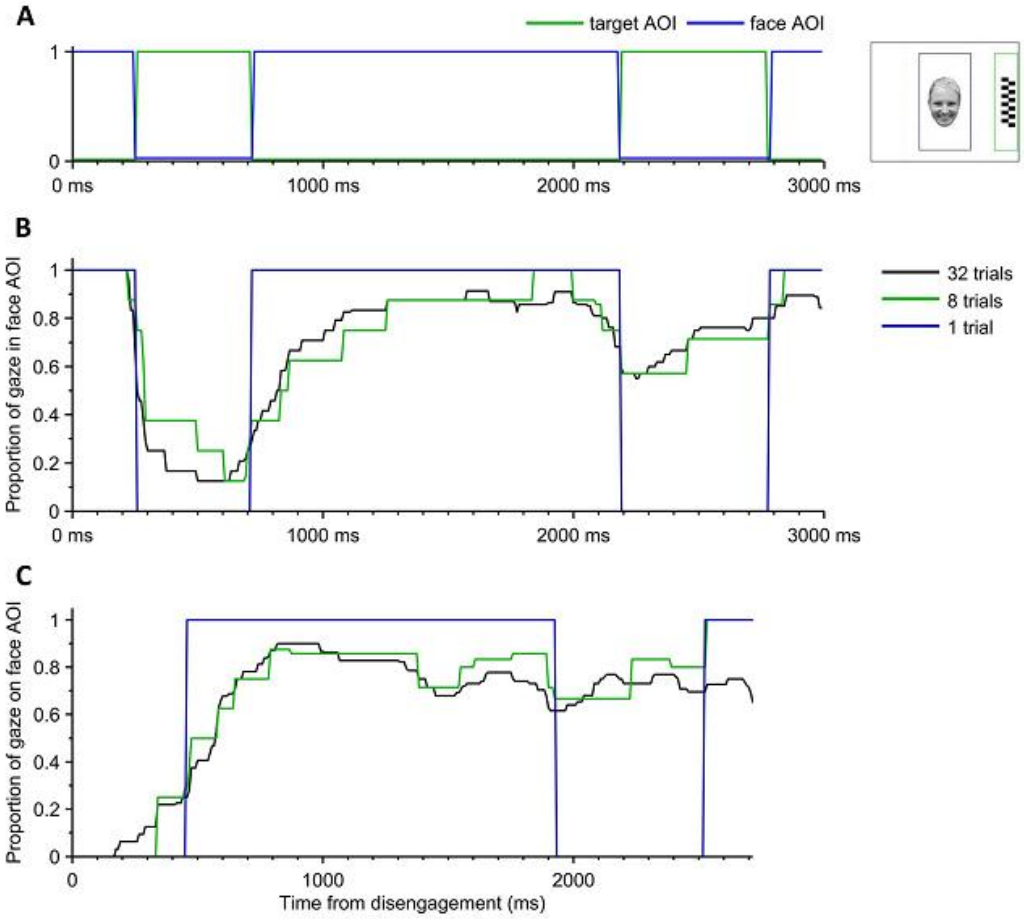

Figure 7. AOI-transformation of the gaze data and derivation of a dynamic response. A) The AOI-transformation of a single trial. The 2D gaze data coordinates of a trial were converted into binary time series according to the AOIs drawn around the central and the peripheral stimuli (shown in the right). In the example case infant's gaze was first directed to the face, but switched to peripheral target on two occasions (at 250–700 ms and 2200–2800 ms). B) A representative example of a dynamic average (face) responses of one participant with different number of trials taken into the averaging (1, 8, and 32). The value at each time point represented the estimated probability of the gaze being at the face AOI. The response resolution (step size) in vertical axis was always inversely proportional to the number of trials included in the averaging. C) An example of disengagement time-locked dynamic responses. Prior to averaging, each trial time series was shifted in time according to the moment when gaze first reached the target AOI, and the data before that were omitted. Note that the data

presented here were from the same experiment as in figures A and B. The figures are adopted from Ahtola et al. (2014).

4.3 Study results

4.3.1 Conventional analyses of initial gaze disengagement

Inspection of the gaze data revealed two distinct phases that reflect the main components of infants' stereotypic gaze pattern during the trial: i) gaze fixation to the face followed by transition to the peripheral target, and ii) possible return of the gaze back to face. Traditionally, studies using the paradigm have focused on measuring the probability (DP) and latency (DT) of the first gaze transition to the peripheral target as the primary dependent variables (Leppänen et al. 2011; Peltola et al. 2009, 2011).

Analyses of these transitions in our primary sample (N = 13) showed that a rapid saccade from the central to the lateral stimulus (Mean = 388 ms, Standard Deviation = 72 ms) was observed on the majority of trials, although the proportion of missing saccades varied substantially between individuals (range 6–78%, M = 44%, SD = 22%).

The results also showed the predicted influence of facial expression on disengagement probability (DP) in a statistical comparison using a chi-squared test: χ^2 = 15.1, df = 3, $p < 0.01$. Consistent with prior studies (e.g., Leppänen et al. 2011), the proportion of missing saccades was lowest in the non-face control condition (M = 27%, SD = 23%) and highest in the fearful (M = 58%, SD = 27%), although none of the pairwise conditions was significant after correction for multiple testing due to the small sample size. Disengagement times (DT) varied significantly by facial expression condition, χ^2 = 8.4, df = 3, $p < 0.05$, with faster DT in the control condition (M = 343 ms, SD = 90 ms) than in face conditions. Yet again, none of the pairwise comparisons was significant after correction for multiple testing.

4.3.2 Comparison of the dynamic responses

Inspection of the gaze dynamics from all infants showed that in 86.3% of the trials with a registered gaze shift to the target a re-engagement of attention to the face followed the initial transition. Nevertheless, there were striking time-varying differences in the shapes and scales of the DR distributions across the time course of the trial (see Figs. 8A–B). This variability provided the motivation to subsequently build an index that reflects the cumulative deviation of gaze from the normative range (MD), and bind the metric to dynamically evolving percentiles rather than some a priori defined limit.

In a stimulus-specific comparison (see Ahtola et al. 2014) the DRs varied slightly between the stimuli, being generally lower for the non-face control stimulus than to faces (especially faces displaying a fearful expression). Also the comparison of the Mean Deviances computed from these traces showed that non-face stimulus yielded significantly lower MDs compared to the combined face condition (p = 0.05; Mann–Whitney U-test). Nevertheless, no significant

differences were found between facial expressions, which may be partly explained by the limited sample size.

These results were consistent with the previous studies (e.g., Leppänen et al. 2011) and the a priori hypothesis of differential patterns of gaze behavior for non-face control stimuli and faces, and motivated us to include only trials with the face stimuli in the subsequent analyses. In addition, the observed stability of the DR distributions even at the end of the trial justified the decision to extend the time window associated with the MD calculations from 0 ms to 2200 ms (from the disengagement) that covers all later phases of the trial.

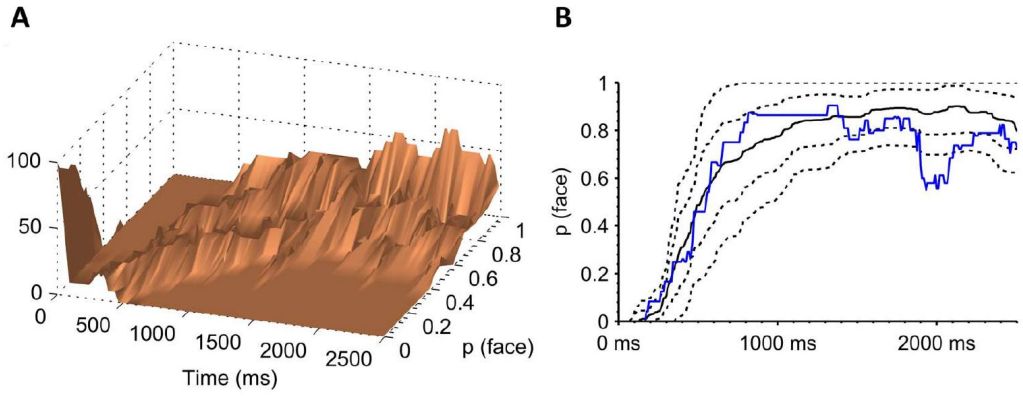

Figure 8. Statistical characteristics of the dynamic eye tracking data. A) Summary graph of DR average responses over the study population that revealed clear changes of distribution within the trial time. The percentiles of this distribution were used in computation of the mean deviance (MD metric) to incorporate the normal variation of the DRs at a specific time. B) The difference between group median response (black solid line) and estimates of 90th, 75th, 25th, and 10th percentiles (dotted lines), as well as an example DR of an individual infant (in blue). Adopted from Ahtola et al. (2014).

4.3.3 Analyses using the novel Mean Deviance metric

Given that our analyses beyond the first gaze shift were highly explorative, we examined whether the results obtained from the analysis of our primary data from 7-month-old infants could be replicated in an independent sample of infants of the same age, and whether the hereby introduced parameter would be sensitive to the known developmental difference in face preference between 7- and 5-month-old infants (Frank et al. 2009; Peltola et al. 2009).

In order to assess the replicability of our measures for the classification of individual infants, we used an independently collected infant sample of same age from the laboratory in Boston (N = 32). These infants were found to be comparable to the infant cohort from Helsinki based on observing no significant differences in the conventional measures, mean DT and DP (all p-values > 0.25; Mann–Whitney U-test). Likewise, the median and the overall distribution of MD values were compatible between the two datasets as there were no statistical differences between the groups (Fig. 9A). In spite of one clear outlier, the group level variance (expressed as IQR in the graph) was smaller in the Boston dataset, which may be readily explained by random effects in small datasets.

Previous research has shown that the preference to faces is strengthened between 6 and 9 months of age (e.g., Frank et al. 2009). We used this prior

knowledge to test the potential of our MD metrics to detect known developmental changes in infants' face processing. Our premise was that a reasonably sensitive test should distinguish younger (5-month-old) infants from older (7-month-old) ones.

The findings in Fig. 9A clearly showed how in a large proportion of the younger infants the MD values fell below the range of the older infants. For demonstration purposes, a tentative MD threshold of 24 was chosen, based on the lower bound of the MD distribution in 7-monts-old infants. Using this threshold, three (7%) out of 45 older (7 months; data from both laboratories combined) infants and ten (45%) out of 22 younger (5 months) infants could be considered deviant.

The observed difference between the age groups was significant ($p < 0.05$; chi-squared test). Closer inspection of the MD results in younger infants suggested that they may fall into two subgroups: one with MD value comparable to the older group and the other with MD value clearly below them. It could be a subject of a further study to determine the neurodevelopmental correlates of these subgroups, yet the data suggested the intriguing possibility of distinct developmental trajectories at this age.

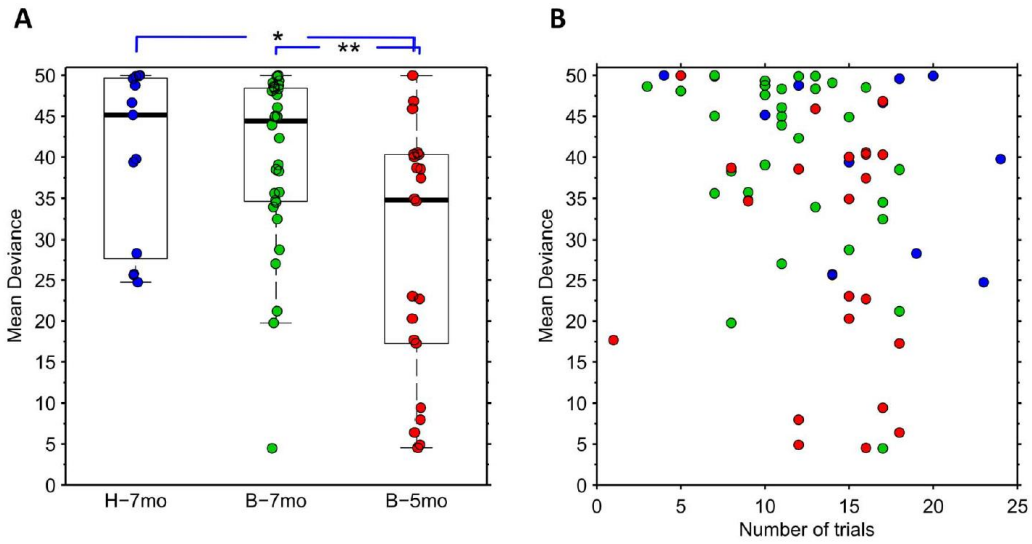

Figure 9. Individual and group level findings of the MD scores. A) Summary of MD findings in all three infant groups: 7-month-old Helsinki (H-7mo) and Boston infants (B-7mo), and 5-month-old Boston infants (B-5mo). Scores of individual infants' are illustrated with coloured circles with the underlying boxplot depicting the median and IQR, and the whiskers showing the total score extent neglecting the outliers (max. length 1.5 x IQR from the box edge). Comparison of MD scores between younger and older infants revealed age-related differences (** and * indicate *p*-values under 0.02 and 0.01 respectively; Mann–Whitney U-test). B) Comparison of MD scores as a function of number of the trials. The figures are adopted from Ahtola et al. (2014).

4.3.4 Stability of the metrics

In addition to the demonstrated ability to distinguish deviations in the test performance, the MD metric seems robust for the variability in the amount of trials included in the DR calculation. MD values represented as a function of number of the trials (Fig. 9B) failed to reveal apparent trends within any participant group.

It was also assessed how much the performance of infants (habituation, fatigue, etc.), as well as our eye tracking based metrics, vary within a single study session. These systematic changes were studied by sign tests between first and second half of experiments and trial-by-trial Pearson correlations after pooling all data together. The correlation analyses revealed that DPs, DTs, and MDs all declined as a function of trial number. Yet, the split-half analyses showed that only DPs were significantly higher in the first half compared to the last half ($p < 10^{-6}$) indicating that there is no consistent direction of change at individual level for DTs ($p = 0.18$) and MDs ($p = 1$).

The observations further suggested that infants' tendency to reduce gaze shifts during the test session might be linked to the measured disengagement times. When the DTs of the seven infants whose DP values declined the most across the session were compared to the seven infants with weakest DP decline, a clear difference was observed; mean DTs were 461 ms (SD 61 ms) and 370 ms (SD 70 ms), respectively.

4.4 Study-specific discussion

The goal of this study was to extend the conventional analyses used in the overlap paradigm by systematically exploring infants gaze behavior during the latter half of the trial period, after the initial disengagement. To parameterize these data, we used dynamic response analysis to produce a graphical distribution model and finally introduced a novel Mean Deviance metric that reflects infants' cumulative allocation of attention to the central stimulus (as opposed to lateral geometric shape). To demonstrate the feasibility of the new methods, it was evaluated how the metrics vary by stimulus condition (emotion) and age, and how robust they are for the variability within and between measurement sessions.

The obvious advantages of this approach are that i) it is insensitive to single outlier values in each trial (removed by median filter of the gaze tracking data and averaging over multiple trials), ii) it can be adjusted to allow detection of very marginal abnormalities (by changing the threshold), and iii) it is able to measure cognitive operation (gaze attention) over extended period of time instead of single time point like the conventional DT.

As for the underlying neurocognitive processes and attentional systems, it can be speculated that the initial disengagement of gaze from the face to the lateral stimulus (DT in our analysis), and subsequent (prolonged) re-engagement with the face (MD in our analyses) reflect partly different aspects of attention. For example, the initial disengagement may be based on subcortical systems important for reflexive orienting responses in infants (Csibra et al. 1998; Corbetta and Shulman 2002), whereas the latter may engage more voluntary and goal-directed attentional control mechanisms, centered in the dorsal parietal and frontal cortex (Corbetta and Shulman 2002; Holmboe et al. 2008). Clearly, further research is needed to tease apart between these explanations, but it is notable that our results showed strikingly similar MD scores in two independent samples of 7-month-old infants, and a clearly differentiated these two groups

from younger, 5-month-old infants. Additionally, our observation (Fig. 9A) suggests that MD may even be able to identify different developmental subgroups among the young infants that are just approaching the age when the tested behavior is developing. Our present data, however, were not able to statistically confirm this or identify its wider correlates, yet the results suggest that MD might be sensitive to developmental differences.

The current results supported the idea that the relative developmental stage of infant's visual attention could be assessed in an objective and automated manner by using DR-based metrics from eye tracking time series. Obviously, MDs alone cannot characterize the whole task performance as parameters related to the initial gaze shift that MD ignores are equally (or even more) important. Although the measured disengagement times failed to show any significant difference between the age groups, the median of trial DTs could perhaps be used as a proxy of the initial disengagement at the individual level. The median DTs could be further transformed into a standard Z-score that compares individual's DT against a priori normal data.

This way, robust classification of individual infants could be performed by two metrics that are independent in principle and reflect partly different cognitive operations: one that measures the early gaze shift and attention disengagement (Z-score of the median DT), and the other that estimates later distribution of gaze between the face and the target (mean deviance, MD). The measured disengagement probabilities (DPs) could also be included in the individual assessment one way or another, at least through a minimum trial criterion needed to guarantee reliable test results.

5. Eye tracking integrated visual evoked potentials

5.1 Motivation and background

An EEG response in cerebral activity following a visual stimulus is called a visual evoked potential (VEP). These potentials are generated continuously in our brains as a result of normal visual processing, but they can be also used as a clinical tool to provide important diagnostic information regarding the functional integrity of the visual system when measured in a controlled laboratory environment.

Evoked potentials are weak signals, often seemingly shrouded by other simultaneous EEG activity. The trick in the technique is to repeat the stimulus numerous times (100–200 according to ISCEV guidelines; Odom et al. 2010), and to average the recorded epochs (repeated measurements) to yield a clean EP with the random noise components cancelling themselves. A typical parameter measured from VEP is the latency of the response, meaning the time from the stimulation onset to the moment the VEP activation reaches its maximum. Obviously, this demands accurate synchronization between the presentation software and the EEG recorder, which is typically achieved by sending trigger pulses between the devices.

VEP techniques are divided in two categories based on the stimulation frequency, i.e., the duration of recurring cycle in the stimulus. When the presentation is accelerated from the transient VEPs (described above) to the point where the activation from to the preceding stimulus has not ended before the new one is elicited, the consecutive individual VEPs begin merging into a continuous stimulation driven waveform called steady-state VEP (SSVEP). (Vialatte et al. 2010)

Instead of time domain averaging used with transient EPs, SSVEPs are usually interpreted by means of spectral analysis. Decomposition with methods such as Fast Fourier Transform (FFT) is used to dissect the components of the activation waveform in the frequency domain. The Fourier-based methods cannot extract the latency from the continuous VEP response, which often limits the clinical usability of the SSVEPs. (Vialatte et al. 2010)

According to the prevalent ISCEV standard for clinical visual evoked potentials (not SSVEPs though; Odom et al. 2010) the most essential EEG electrode

in a VEP recording is Oz in the midline occipital area with the reference electrode located at Fz. Also lateral, occipital electrodes O1, O2 or even PO7 and PO8 are recommended especially for evaluation of chiasmal and postchiasmal function.

There is a wide variety of different stimulus types available in the research field, each designed to target some specific aspect of the visual processing. Clinically, the two by far most used and most studied ones are flash-VEP and pattern-VEP stimuli. The former measures the response to the luminance changes generated by a blinking photic stimulator. The latter is based on contrast changes in the visual field typically created by screen presentation of dynamically reversing black and white checkerboard pattern, and thus often called pattern-reversal (PR) VEP.

Flash-VEP is less used in adult studies due to its high level of intra- and inter-subject variability but since it has no fixation requirements it is a useful tool when testing infants or other subjects with weak co-operation ability or poor visual acuity. Successful PR-VEP produces much more stable responses measuring activation of the neurons at the primary visual cortex responsible for the central area of the visual field. Stimuli like these, however, produce responses also at earlier stages of the visual pathway from the eye to the visual cortex (e.g., in on- and off- ganglion cells of retina). Hence, flash- and PR-VEP responses indicate merely that the visual information is successfully activating the cortex, but not necessarily reflect actual cortical processing at the level of V1. (Atkinson et al. 2002; Lee et al. 2012a)

Orientation-reversal VEPs

In contrast to the traditional luminance or contrast based VEP responses, orientation-specific responses that are essential, for example, in object recognition can be generated only in the primary visual cortex and in further extrastriate visual areas. (Atkinson et al. 2002) Aspects of the cortical orientation selectivity have been evaluated by means of steady-state orientation-reversal (OR) VEP studies which typically use a high contrast sine or bar grating stimulus that switches its orientation continuously back and forth between 45° and 135° angles (Braddick et al. 1986; Braddick et al. 2005).

It has been suggested (Atkinson et al. 2002; Lee et al. 2012b) that the method could be used as a clinical indicator of cerebral development showing diagnostic and prognostic promise in infant studies with children with focal brain injury, hypoxic-ischemic brain, and prematurely born infants with white matter injuries.

Braddick et al. (2005) studied the onset age of the cortical orientation selectivity by OR-VEPs using sine wave gratings with spatial frequency of 0.3 cycle/deg and reversal rate of 4 Hz. In the study, 115 healthy full-term infants aged between 5 and 18 weeks were measured and analyzed in five groups based on their age. According to their data, an orientation-reversal response could be detected from a large majority also within the youngest group aged 5–7 weeks. In total, the response detection rate varied in between 75% and 100% across the groups.

The work by Atkinson et al. (2002) has shown that OR-VEP could be used as a prognostic indicator of early brain development alongside other neurological measures by comparing OR-VEP responses of 24 very low birth weight preterm infants and 31 infants born at term. The measurements were performed at ages of 0–6 weeks and 12–17 weeks after term using square wave grating stimulus with spatial frequency of 0.5 cycle/deg and reversal rates of 4 and 8 Hz. The observed average OR-response detection rate in the control group was 90%. No significant difference, at either age, was found between the preterm group and the control group after the children with abnormal ultrasound findings (N=9) had been excluded from the preterm group. However, four infants with very deviant ultrasound findings failed to produce OR-responses at all and showed neurological abnormalities later at the age of 2–3 years.

Global form and motion VEPs

Steady-state VEP techniques have also been developed to test the specificity of visual cortex for coherent form and motion, functions that are established by around 2 and 3 months of age. (Wattam-Bell et al. 2010)

The stimulus sequence for global form (GF) VEPs consists of short arc segments that alternate between states of coherence, in which the short arc segments are concentrically organized to create a global circular pattern, and non-coherence, in which the same segments are randomly oriented without any global structure. Correspondingly, in global motion (GM) stimulus, arc segments are replaced by dots that in coherent phase move along short tangential trajectories creating global rotational movement, and in non-coherent phase move to random directions.

GF- and GM-VEPs have been used to measure of global processing in the ventral (object recognition) and dorsal (encoding object position and motion) visual streams, respectively. (Wattam-Bell et al. 2010) These findings have also been supported by the results from a separate fMRI study by Braddick et al. (2000) using similar stimuli.

In a work by Wattam-Bell et al. (2010) GF and GM tests were performed for 26 adults and 26 infants aged between 4 and 5 months. It showed that about 90% of the adults produced global responses to both form and motion stimuli. For the infants, the GM response rate was at similar level (92%), but notably only 50% of them yielded responses to the GF stimulus. These observations indicate that the global form sensitivity may be still partly under development at the age of 5 months.

The general topographical distributions of the GF and GM responses were studied in the same work revealing distinct differences between adults and infants, as well as between both stimulus types. In infants, motion stimulus produced strong responses whose focus covered wide area extending to lateral, occipital areas. The form response, on the contrary, was focused near the posterior midline (see Fig. 10). In adults, it was the other way round as motion responses were concentrated close to the midline and form more laterally on the sides.

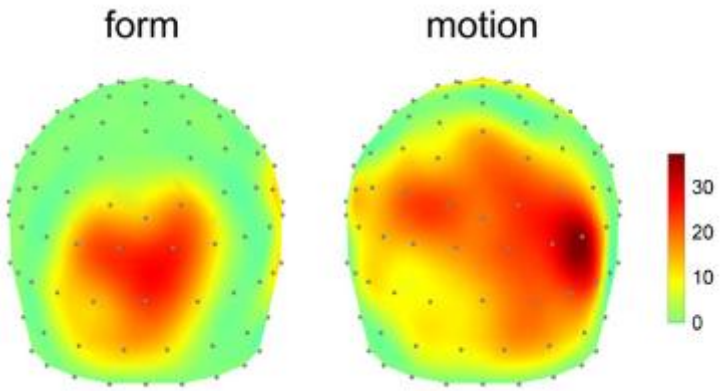

Figure 10. The spatial distribution of infant global form and motion VEPs. Response strength for each electrode position was measured as T2circ statistic amplitude (method described later in Section 5.2.5) that was subsequently projected into a graphical head model (posterior view). The presented data were averaged from recordings of 26 infants aged 4–5 months and tested with GF and GM VEP protocols. Adopted from Wattam-Bell et al. (2010).

There are no published attempts where an eye tracker would have been integrated in SSVEP recordings. In the infant studies presented above (Braddick et al. 2005; Wattam-Bell et al. 2010), the question of how to attract infants' attention on the stimulus, was solved by a small noisy toy that was shaken in front of the screen whenever attention was lost. The averaging process was likewise controlled online by the experimenter who could stop the sampling until the appropriate gaze direction was regained.

Goals of the study

The specific goals of this study were to: i) implement eye tracker guidance to OR, GF, and GM SSVEP recordings, ii) develop and evaluate alternatives for conventionally used data analysis methods, iii) study how the available gaze data can be exploited in the VEP analysis, and iv) determine VEP response detection rates and spatial distributions in a sample of typically developing 3-month-old infants with a view towards the potential diagnostic use in the future.

5.2 Materials and Methods

5.2.1 Participants

The VEP recordings were performed for a sample of 3-month-old infants at the Helsinki University Central Hospital (N = 39; 17 females; age range 3.0–4.0 months, mean 3.44). In two cases, the tests could not be performed successfully at the time, and hence infants were already over 5 months old at the recording date (subject number 23: 5.7 months and number 25: 5.8 months). All babies were born full term and reported to have a typical development.

An objective for the study sessions was to get, for all three stimulus types separately, at least 100 s of EEG with valid gaze focus on the screen (monitored cumulatively during the recording). This ambitious goal was not always possible to achieve because of the fatigue of the infant. The GF and GM stimuli were then prioritized in the recording session, and thus generated generally more data

than OR presentations. In two cases particularly, the conditions for the OR recording were so poor that the stimulus type needed to be omitted from the session completely.

5.2.2 Stimuli and stimulus presentation

In this study, three different pattern VEP stimuli were consecutively shown to the participants.

Orientation reversal

The orientation-reversal stimulus (see Fig. 11) consists of high contrast sine wave gratings with spatial frequency of 0.45 cycle/deg at the measured average viewing distance of 59 cm. The grating pattern switches its orientation back and forth between oblique angles of 45 and 135 degrees at frequency of 4 Hz (every 250 ms). (Braddick et al. 2005)

The orientation reversals are accompanied by random phase shifts in the gratings occurring at faster 24 Hz rate. The purpose of the 'jitters' is to control non-orientation-specific responses to local contrast change. Because of the phase shifts, the local luminance variations in the pattern are statistically similar in every jitter and in the orientation reversal. Thus, the local luminance changes contribute to the response signal at 24 Hz (the 6th harmonic of the OR frequency), but not to the component at the actual stimulus frequency. (Braddick et al. 2005)

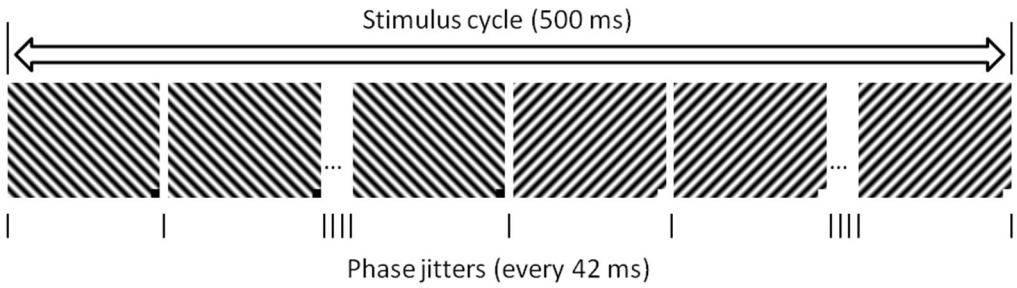

Figure 11. The orientation reversal stimulus (OR). The sine grating switches its orientation continuously between 45° and 135° angles at intervals of 250 ms (stimulus rate 4 Hz). The gratings also undergo faster, random phase-shifts ('jitters') at the rate of 24 Hz (every 42 ms) to isolate non-orientation-specific responses due to local contrast changes.

Global form

The global form stimulus consists of a set of short white arc segments (N = 2000, maximum length 1.3°) that switch between states of global coherence and non-coherence on a black background. In the former state the arcs are concentrically organized to create a global circular pattern, and in the latter the same segments are randomly oriented without any global structure (see Fig. 12). The reversal frequency between the two states was 4 Hz meaning that the change from the non-coherent to the coherent global structure occurred at the intervals of 500 ms.

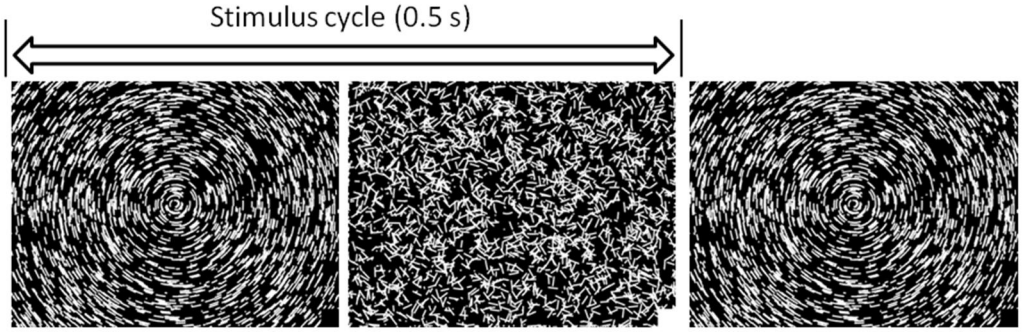

Figure 12. The global form stimulus (GF). The GF stimulus sequence consists of short arc segments alternating between coherent and non-coherent phases at the rate of 4 Hz (full stimulus cycle is 500 ms).

Global motion

The global motion stimulus consists of a set of white dots (*N* = 2000, diameter 0.3°) that move on a black background (see Fig. 13). Similarly to GF presentation, the stimulus alternates between states of global coherence and non-coherence at frequency of 4 Hz. In coherent phase the dots move along short tangential trajectories creating global rotational movement, and in non-coherent phase they move independently to random directions.

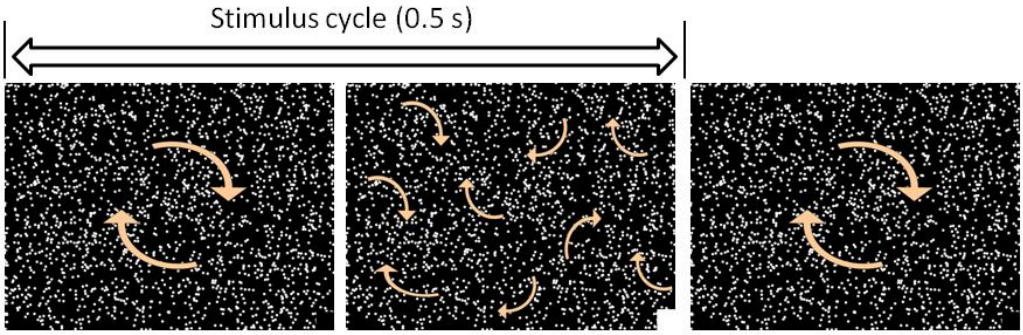

Figure 13. The global motion stimulus (GM). In alternating coherent and non-coherent states white dots move along short tangential trajectories creating global rotational movement (indicated by the arrows) or independently to random directions, respectively.

All stimuli were first generated frame by frame in MATLAB with functions included in Psychophysics toolbox (version 3; Brainard 1997) routinely used in vision and neuroscience research. The frame images were encoded into 10 s/20 s/30 s -long MPEG-4 video files using XVID-codec that was found to provide the smoothest presentation. The frame rates of the video clips were 24 Hz for OR, and 60 Hz for GF and GM stimuli. The resolution was 1024 x 768 pixels for OR and 800 x 600 for GF and GM to ensure flawless playback at the higher frame rate.

The start and end times for the playback of the video files were time-stamped to the gaze datafiles to provide synchronization with the eye tracking results. The integration with the EEG was then achieved using black and white markers at the bottom right corner of the stimuli (visible in Figs. 11–13) as described in Section 3.6.

Use of gaze data in stimulus presentation

In addition to the obvious advantages in the analysis phase, the eye tracking could be exploited already during the stimulus presentation. The real time gaze position information allowed dynamic adaptation of the recording session to the attention of the infant. This did not only help the work of the experimenter but also substantially improved the quality of the recorded data in general.

The playback of the video stimulus started only after child's attention had been attracted appropriately. This was achieved using the programmed E-Prime script described in Section 3.5. A simple attraction animation was presented on the screen, and the video presentation started automatically after the eye tracker had reported a stable fixation on it. During this period the operator could switch the animation or launch the stimulus presentation, ignoring the gaze status, by a key press.

Eye tracking also allowed interaction between infant and stimulus presentation giving real time feedback of the accumulation of the acceptable data, i.e., how well the child's gaze followed the latest stimulus sequence. The position of infant's eyes in the eye tracking camera could also be monitored online through the interface of Tobii Studio software. Based on these mechanisms the operator could change length of the continuous stimulus presentation between 10, 20, and 30 s depending on the child's level of attention. The attraction animations described above were programmed to start automatically always between the successive stimulus sequences.

Finally, the program kept track on the cumulative amount of acceptable data that were recorded during the session in total. This helped the operator to make the final decision when the adequate amount of data for the reliable analysis was achieved, and the experiment could be stopped.

5.2.3 VEP recordings

The EEG measurements of the VEP studies were preformed with NicoletOne V44 EEG-system and ANT WaveGuard electrode caps that provide 22-channel Cz-referenced EEG recording according to the international 10–20 electrode placement scheme. Because of our main interest in the analysis was in the occipital regions, an additional Oz electrode was added between O1 an O2 (see Fig. 14).

The impedance levels of the electrode-skin interface were monitored before the data collection, and the contacts were prepared so that that the desired level below 5 kOhm was achieved. The sampling of the EEG was performed at the rate of 500 Hz with 16 bit AD-transform.

The stimulation periods were pruned from the recording and converted into the standard EDF -file format that can be read in MATLAB. Consistently with prior studies using same VEP stimuli (e.g., Braddick et al. 2005), the raw EEG was digitally bandpass filtered between 0.5 and 30 Hz before any further signal processing. The filtering was performed in MATLAB with order 8^{th}, zero-phase, forward and reverse Butterworth filter.

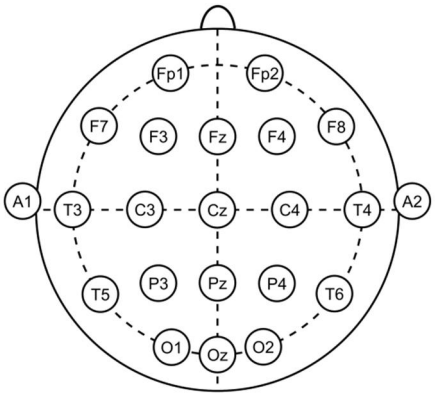

Figure 14. Electrode placements in the VEP recordings. Cz at the vertex was used as the reference electrode.

5.2.4 Use of gaze data in VEP analysis

The gaze data collected by the eye tracker during the stimulus presentation can be used in the epoch division of EEG that identifies the segments with appropriate gaze orientation to the stimulus. Correspondingly, the patches with disoriented attention can be excluded resulting in improved overall sensitivity (see example illustration in Fig. 15).

An automatic MATLAB-based algorithm was developed to perform the epoch selection. There are two ways how the optimized segmentation can be done: i) by setting a fixed threshold of the average gaze fixation quality for the acceptable epochs, or ii) by fixing the number of epochs needed for the analysis, and choosing the segments so that the overall gaze fixation quality of the selection is maximized. The former way was found to be more useful for the developed analysis methods that exploit the baseline EEG data, whereas the latter is preferable for the conventional T2circ analyses because of the observation that the number of the analyzed epochs had a significant influence on the findings (analysis methods described in the next sections).

The gaze fixation quality metric used here is defined as the proportion of the time the gaze (according to the tracker) was directed on the screen during a chosen time interval. This calculation was based on the 'validity codes', measured at 120 Hz, indicating the reliability of tracking (see Section 3.4); out of codes 0–4, only samples with validity code 0 for both eyes were accepted. The tracked gaze coordinates (x and y) were not used in VEP analyses, i.e., the gaze at center of the screen was valued equally to the edge areas.

The gaze fixation quality index was first calculated for each 0.5-s-long reversal cycle in the recurrent stimulation sequence. Afterwards, adjacent reversal cycles were merged to form longer epochs (for example 3.5-s-long epoch equals 7 cycles) needed for the response analysis.

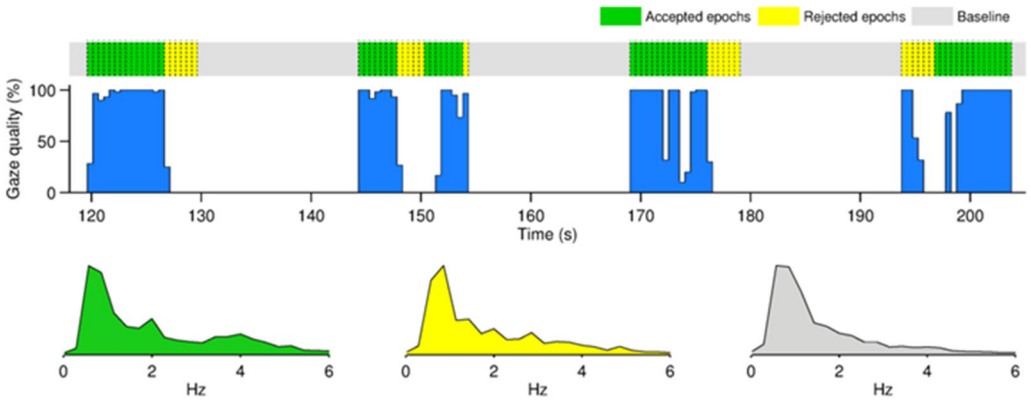

Figure 15. Principles of the automatic epoch selection algorithm. Above: The blue bars depict evolution of gaze fixation quality during four successive 10-s-long GM-stimulation sequences each consisting of 20 reversal cycles (black dotted lines). The occasional gaze fixation quality drops indicate participant's temporary disorientation from the stimulus. The developed selection algorithm identifies 3.5-s-long epochs with good (green) and poor (yellow) gaze fixation for the further analysis (fixed gaze threshold of 45% used here). Below: Power spectrum of good fixation epochs (green) reveals a distinct peak at the stimulus frequency of 2 Hz. The spectral peak is substantially decreased in rejected epochs with poor fixation (yellow) and completely missing from the spectrum calculated from the baseline epochs measured in between the stimulus sequences.

5.2.5 VEP analysis method used in the reference studies

In the reference works done with OR, GF, and GM VEPs (Braddick et al. 2005; Wattam-Bell et al. 2010), the detection of the steady-state VEP responses was performed using T2circ statistic (Victor and Mast 1991) that effectively measures SNR of periodic, phase-locked events in the recorded EEG. Being a statistical metric, T2circ values measure the reliability of the response. The higher statistic values indicate less variability and thus stronger correlation between the stimulation and the recording.

In T2circ analysis the SSVEP recording is processed as epochs that repeat the response to a periodic stimulus. Like many other SSVEP analysis approaches, T2circ is based on the estimates of the Fourier components at the stimulation frequency. Each such estimate is a complex number ($z=x+iy$) and may be interpreted as a vector in the (x,y)-plane. In this graphical representation the length and angle of the vector correspond to the amplitude and phase of the signal, respectively.

When multiple epochs are processed similarly, the resulting vectors form a cluster in the complex plane. The sample mean of the individual vectors provides an estimate of the 'true' response, whereas the scatter of the cluster may be seen as an index of reliability indicating the deviation between the individual observations. The T2circ statistic assumes that the errors in the real and imaginary parts of the measured Fourier components are independent quantities with equal variances. Thus, the amount of scatter may be represented as a circle around the tip of the mean vector with the radius determined by the chosen confidence interval. By continuing the graphical analogy, the recording is determined to be phase locked to the stimulus if the confidence circle does not reach the origin of the (x,y)-plane, indicating presence of a significant VEP response in EEG.

In the article by Victor and Mast (1991), T2circ statistic is defined through the ratio of two independent estimates of the population variance. The first estimate (V_{indiv}) is calculated from the scatter of the individual components (z_j) about their sample mean ($\langle z \rangle_{est}$), whereas the second variance estimate (V_{group}) is derived from the difference between the sample mean and the a priori hypothetical population mean (ζ). Ultimately, the derivation leads to the equation:

$$T^2_{circ} = \frac{1}{N}\frac{V_{group}}{V_{indiv}} = (N\text{-}1)\frac{\left|\langle z \rangle_{est} - \zeta\right|^2}{\sum\left|z_j - \langle z \rangle_{est}\right|^2}, \qquad (3)$$

where *N* is the number of trials.

Moreover, $N \cdot T^2_{circ}$ is statistically distributed according to the $F_{[2,2N-2]}$, that allows determining the confidence region for the population mean using *F*-distribution:

$$T^2_{circ(1-p)} = \frac{1}{N} F_{(1-p)[2,2N-2]}, \qquad (4)$$

where (1-*p*) the chosen confidence level (typically 95%). For the sake of practicality, all T2circ values reported later in this work are actually $N \cdot T^2_{circ}$ values. This choice enabled a straight comparison with the *F*-distribution values, when an observed set of Fourier components was evaluated to consists either of random fluctuations alone ($\zeta = 0$) or contain a significant signal within a certain confidence level.

A thorough theoretical background, the full derivation of the statistic, and analysis on the method can be found from the referred article (Victor and Mast 1991).

5.2.6 Power spectrum based methods developed for VEP analysis

The novel analysis methods developed in this study were also based on splitting the continuous EEG into epochs of fixed length and examination of their spectral characteristics around the stimulus frequency. Compared to the previously used T2circ statistic, however, the method presented here followed a more straightforward path using the estimates of EEG power spectrum (power spectral density, PSD). Furthermore, the evaluation of the response existence (response significance) was no longer based on mathematical assumptions but on the characteristics of actual background EEG measured from the same individual.

The analyzed EEG was first divided into 0.5-s-long cycles that included the both phases of the continuously alternating stimulus. Using the epoch selection algorithm described earlier, the adjacent cycles were further combined into 3.5-s-long epochs. This length was found to provide a good compromise between

the spectral resolution (1/length), temporal accuracy of the epoch selection algorithm, and the resulting number of the epochs for the statistical comparisons.

To avoid analyzing data corrupted by artefacts, any segments with peak-to-peak EEG voltage difference over 200 µV were rejected (similarly as in Braddick et al. 2005 and Wattam-Bell et al. 2010). However, a temporal overlap of 1 s between the neighbouring epochs was allowed to maximize the amount of analyzed data. The epoch selection was performed for all measured EEG channels independently.

To characterize the spectral properties of the background EEG in the measurement, a set of baseline epochs was chosen from the periods of the recording where the visual stimulation was not presented (usually before, after, or in-between the stimulus sequences). This background epoch selection was performed automatically, channel by channel with the same constraints that were applied for the stimulation epochs (length, neighbour overlap, and voltage limit).

Next, the power spectral density was calculated for each stimulus and baseline epoch separately. Two spectral features were extracted from each PSD: i) the height of the spectral peak at the stimulus frequency, and ii) the same peak height divided by the average of the adjacent PSD values in the range of ±1 Hz from the stimulus frequency (neglecting the stimulus component itself). These features are measures of spectral energy (called 'E' or 'Energy' from now on) and signal-to-noise ratio ('SNR'), respectively (see Figs. 16A–B).

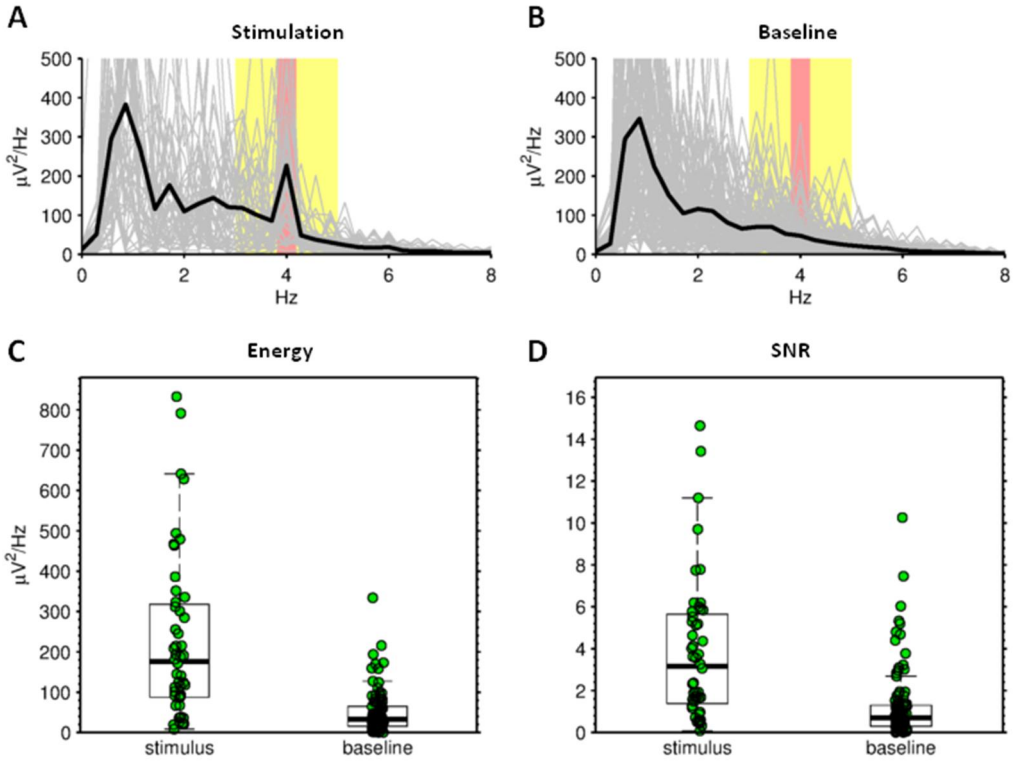

Figure 16. Extraction of the spectral features. A–B) Two averaged spectra of occipital EEG, one with (A) and another without (B) the presence of visual stimulation. The single-epoch PSDs underlying the averages, and depicted with grey, were calculated from a large number of epochs selected across the stimulation session. Despite high inter-epoch variability the presence of a spectral peak at the 4 Hz stimulation frequency is evident in the stimulation data. The presence of this peak (or lack of it) can be characterized by two spectral features

reflecting the energy and SNR of the response. Red and yellow areas highlight the PSD components used in the feature extraction. C–D) Boxplot distributions of the spectral features (Energy and SNR) extracted from the epoch PSDs presented above. In this illustration the difference between the stimulus and baseline conditions is apparent with both features and could be statistically evaluated using the Mann–Whitney U-test.

The selected baseline data were used to constitute individual normal ranges for the characteristics of infant's EEG without the presence of the visual stimulation. Thus, in order to distinguish the possible presence of VEP response at the specific channel, the extracted features from the stimulus and baseline epochs were compared statistically using Mann–Whitney U-test (see Figs. 16C–D). Statistically significant VEP response was detected if the corresponding *p*-value was below 0.05.

The Mann–Whitney U-test (also known as Wilcoxon rank-sum test) is widely used, non-parametric, statistical test. It was chosen to be used here as it does not assume the compared values to be normally distributed (this assumption would not hold, see example Figs. 16C–D), and neither it requires the sizes of compared populations to be equivalent.

Estimation of the response strength

Although the statistical test described above yields reliable results in assessment of the existence of a VEP response, it is not practical when actual response magnitudes are compared, for example, topographically between different electrode locations.

The solution was to develop a new metric for the estimation of the VEP response strength ($E_{strength}$) calculated from the difference of the median of PSD peak energies of stimulation and baseline epochs (E_{stim} and $E_{baseline}$). The difference was further scaled by the average of corresponding interquartile ranges (IQR), resulting in an equation

$$E_{strength} = \frac{\text{median}(E_{stim}) - \text{median}(E_{baseline})}{1/2 \times \left(\text{IQR}(E_{stim}) + \text{IQR}(E_{baseline})\right)}. \quad (5)$$

The same equation could also be applied to spectral signal-to-noise ratio by substituting the energies with SNR values.

5.3 Study results

5.3.1 Individual results: comparison of the analysis methods

39 participants were analyzed separately for OR, GF, and GM responses, except for two OR-VEP recordings there was not enough recording data. The analysis was performed using the two detection methods presented in Section 5.2.6 (Energy and SNR). In both cases, the eight posterior EEG channels that, based on a priori information (Braddick et al. 2005; Wattam-Bell et al. 2010), were the most likely to show responses were included in the process: P3, P4, O1, O2, T5, T6, Pz, and Oz all referenced to Cz electrode at the vertex.

The data on these channels were analyzed independently. A gaze fixation threshold of 20% was applied in the epoch selection (the decision justified in a later section). A VEP response was detected if the *p*-value in any channel comparison was below 0.05 (Mann–Whitney U-test).

VEP-detection results are presented in Fig. 17 individually for each participant. On average, the analyses yield the following response detection rates for the Energy-metric: OR: 89%, GF: 72%, GM: 82% and correspondingly for the SNR-metric: OR: 78%, GF: 82%, GM: 72%.

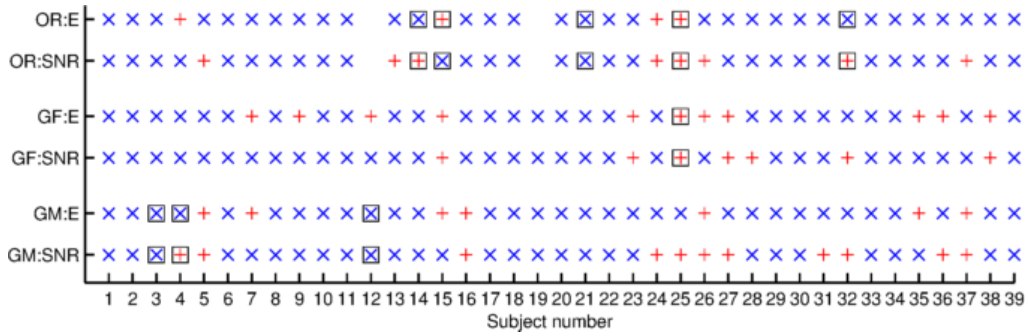

Figure 17. Individual VEP response detection results. VEP responses for 39 participants were calculated using the two analysis methods (E and SNR) for all three stimulus types (OR, GF, and GM). In the graph blue “x” and red “+” markers indicate positive and negative findings, respectively. The measurements where the number of available stimulation epochs was low (below 22) are enclosed by black squares.

It was further studied whether the VEP response detections could have been systematically affected by variations in the recording conditions such as the average gaze quality and the numbers of epochs in the analysis (stimulation and baseline). The results of this inspection are presented in Fig. 18, in which the factors possibly contributing to the significance of each finding are evaluated separately.

In addition to the graphical representations, Spearman's rank correlations were calculated for the data points (stimulus types combined) of each graph in Fig. 18. The response significance values derived using the Energy-method yielded correlation coefficients of 0.02, –0.06, and –0.27* in comparisons with the average gaze quality, the number of stimulation epochs, and the number of baseline epochs, respectively (* indicating that the corresponding Spearman *p*-value was below 0.05, and thus the correlation being significantly different from zero). Correspondingly, the similar comparisons with the response *p*-values generated using the SNR-method produced correlation coefficients of –0.22*, –0.16, and –0.16.

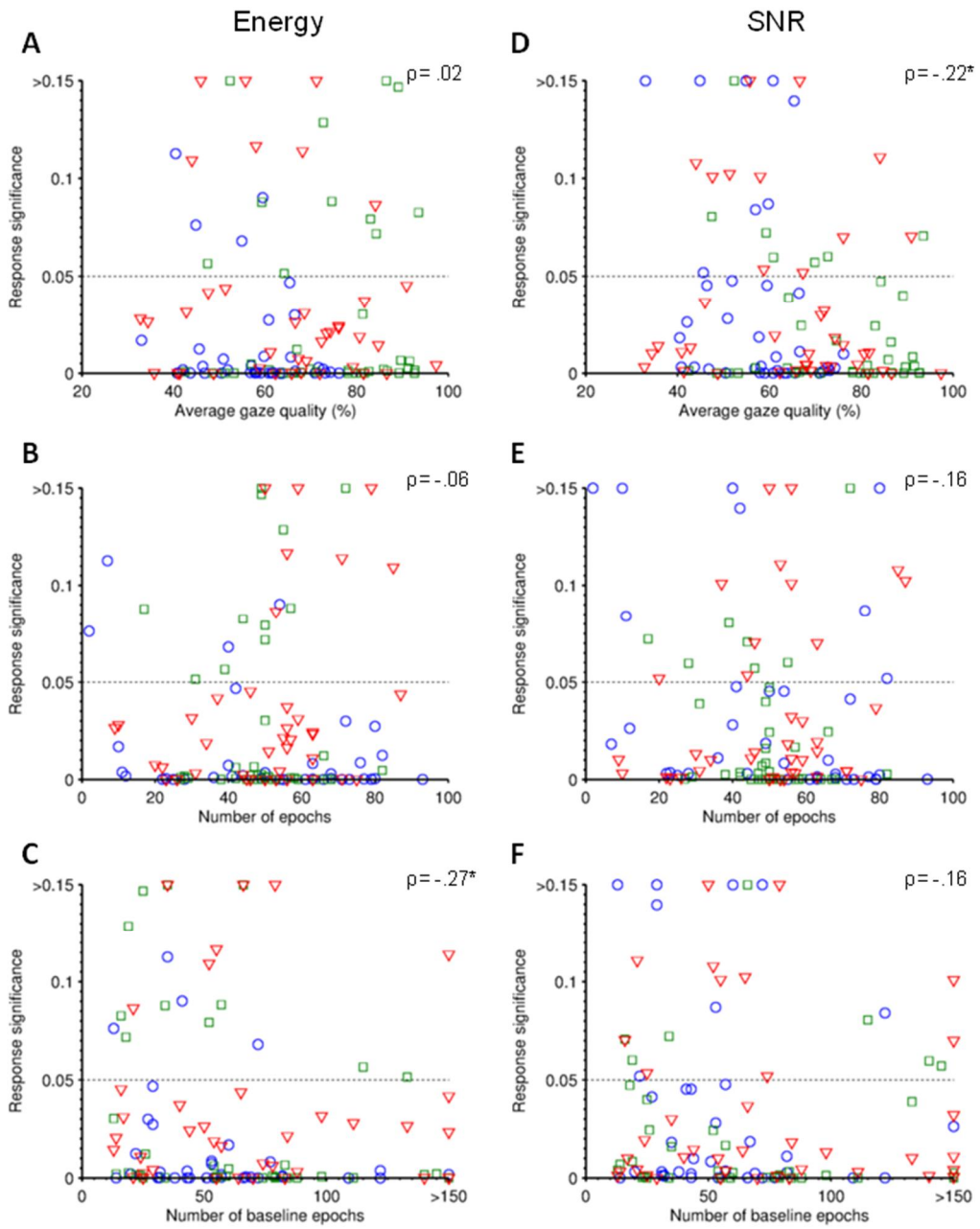

Figure 18. VEP response significance levels in variable recording conditions. The lowest response *p*-value for each participant (*N*=39) is depicted as a function of corresponding average gaze quality, number of analyzed stimulation, and baseline epochs. OR-, GF- and GM-recordings are marked with blue circles, green squares, and red triangles, respectively. The dotted line depicts the 0.05 *p*-value threshold that was used in response detection. Results from both developed VEP detection methods are illustrated separately: A–C) Energy-metric, D–F) SNR-metric.

Visual inspection of the graphs does not reveal any clear trends or dependencies among the highly scattered measurement points. Nevertheless, the two statistically significant Spearman correlations imply that i) the higher level of the fixation on the stimulus presentation yields more prominent responses (supporting the need for the eye tracking control), and ii) the greater amount of baseline data improves the reliability of the statistical response detection.

Although no correlation was observed between the response significances and the number of stimulation epochs, the Figs. 18B and E showed that the recordings with the fewest number of epochs often failed to produce significant responses. Hence, the sensitivity of the findings could be improved by setting a requirement for the quantity of stimulation data needed for the analysis (see Fig. 17). Based on the presented data, a suitable threshold level would be 22 epochs.

This exclusion criterion would increase the measured response percentages slightly: OR: 94%, GF: 74%, and GM: 81% for the Energy-metric and correspondingly OR: 84%, GF: 84%, and GM: 72% for the SNR-metric.

Generally, in studies where multiple independent statistical tests are combined, a significance-level correction method (such as Bonferroni) is used to control the familywise error rate. Fig. 18 showed how large proportion of the response *p*-values (especially GM) was located just below the proposed detection threshold of 0.05. Hence, it is obvious that any correction method would likely have a substantial effect on the observed response detection rates. For example, using the false discovery rate correction technique (FDR; Benjamini and Hochberg 1995), also applied in the reference study (Wattam-Bell et al. 2010), the response percentages decreased to OR: 72%, GF: 66%, and GM: 39% for the Energy-metric and OR: 59%, GF: 61%, and GM: 39% for the SNR.

5.3.2 Spatial distribution of the VEP responses

One target of this VEP study was to investigate the spatial distribution of the responses to OR, GF, and GM stimulation. This topographical analysis was performed in group level by first calculating the channel specific magnitudes of the VEP responses (Eq. 5) and then averaging the individual results across the whole study population.

To improve the specificity of the actual response characteristics, the recordings that did not produce a response with either detection method (before FDR control) were excluded from the topographical analysis. In total, this omitted data included 4 OR, 5 GF, and 4 GM recordings (see Fig. 17). The results of the topographical analysis are presented in Fig. 19 separately for both analysis methods.

According to the spatial response profiles, the presentation of the orientation stimulus produced a strong response on the area covered by the three occipital electrodes, but was hardly distinguishable at any other region. The distribution of the global form response was mostly similar to OR, but expanded further away from the occipital focus and could be observed also at the posterior and temporal areas to a lesser extent. Spatially the global motion produced the broadest response from the three stimuli types covering a wide area ranging from the lateral posterior areas (T5 and T6) to the occipital midline regions (Oz) at even, although quite low, magnitudes.

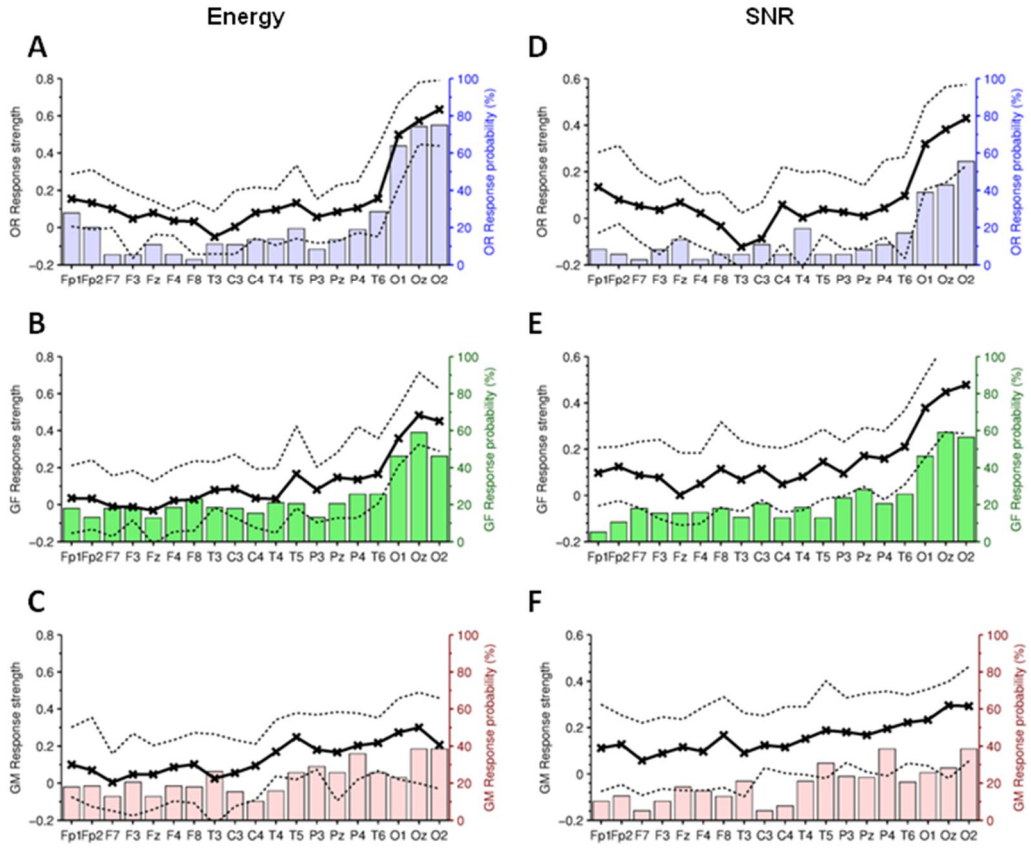

Figure 19. Topographical VEP analysis. The spatial distributions of the responses to OR (A, D), GF (B, E), and GM stimulation (C, F) were calculated by averaging the response strengths at different channels across the whole study population. Results for the two developed analysis methods are illustrated separately: A–C) Energy-metric, D–F) SNR-metric. The intersubject variability of the response strength at each channel is characterized by the interquartile ranges of the measurements and depicted as dotted lines in the figures. The colored bars on the background show the average probability of the response detection at the specific channel (scale on the right side *y*-axes).

Generally, the results from both analysis methods (Figs. 19 A–C vs. D–F) were very similar. The only noticeable difference was observed with the OR stimulation that yielded over 30% weaker occipital response with the SNR-method compared to the Energy-metric (note different scales on the *y*-axes of the figures).

5.3.3 Optimization of the gaze quality threshold

The gaze quality based epoch selection algorithm (Section 5.2.4) provided a way to fine tune the VEP analysis tools by adjusting the level of the gaze quality threshold. Obviously, as the analysis were performed offline after the recording, the higher threshold, while ensuring improved data quality, limited also the number of the EEG epochs taken into the analysis.

Although the assessment of the intersubject variability of the VEP responses presented earlier (Figs. 18A–D) suggested that there is no remarkable correlation between the response significance with either average gaze quality or epoch quantity, more profound investigation was performed aiming to find the optimal value for the gaze threshold.

Both VEP response probability and response strength calculations (as in Sections 5.3.1 and 5.3.2) were repeated for the whole dataset with 20 different epoch segmentations gained from different gaze thresholds between 0 and 95%. The results were averaged over the 39 participants and depicted as functions of the threshold (Figs. 20C–F). The two developed analysis methods (Energy and SNR) and the three stimulus types (OR, GF, and GM) were all processed separately. To further help the reasoning, also gaze quality distributions and epoch quantities were calculated under the same circumstances (Figs. 20A–B).

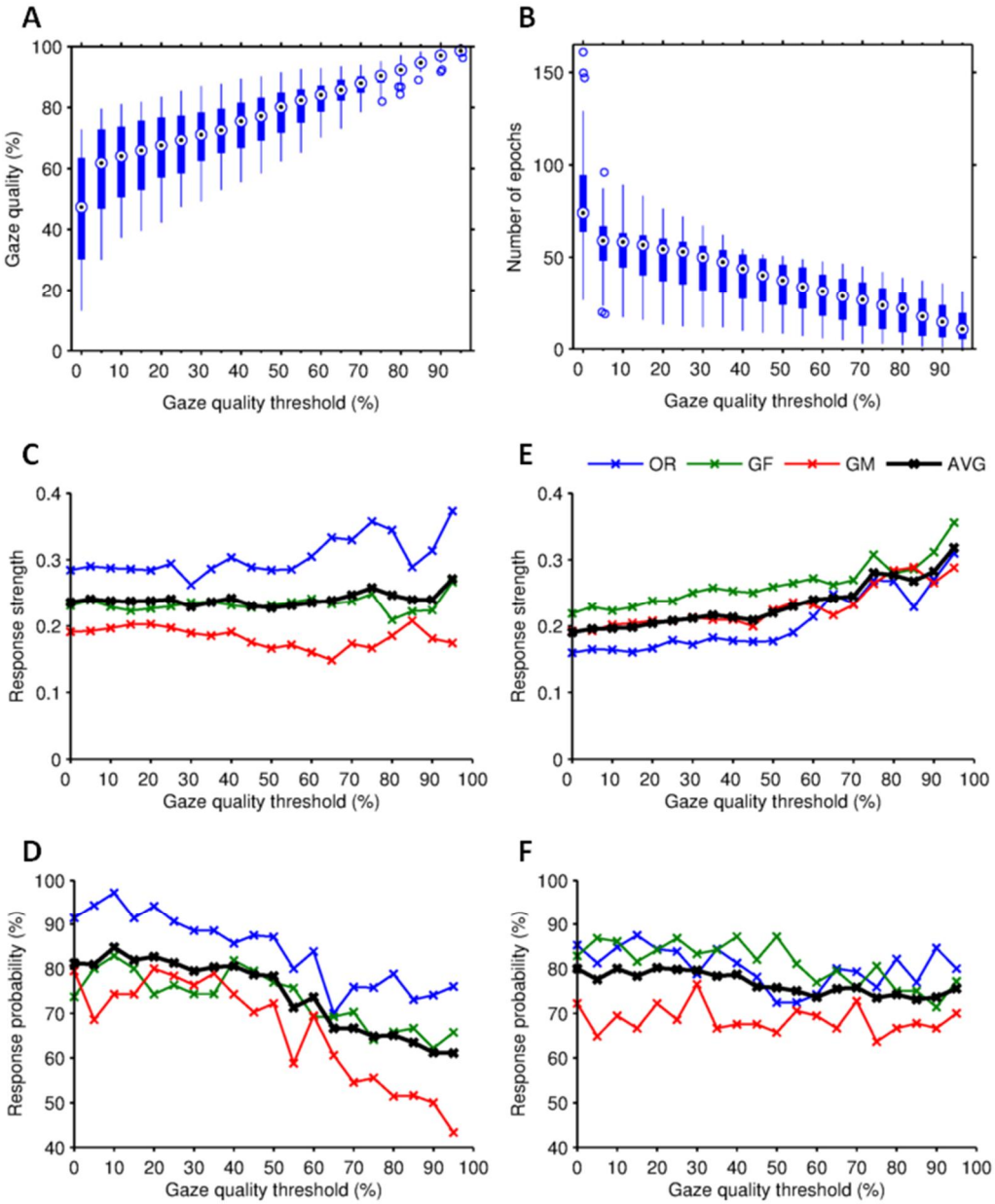

Figure 20. Optimization of the gaze quality threshold. In order to find the optimal threshold level for the gaze quality in the epoch selection, characteristics of the VEP responses were studied as a function of variable gaze threshold ranging from 0 to 95%. The figures A and B first show how the distributions of average gaze quality in the analyzed stimulation epochs and epoch number evolve as functions of the threshold level. The distributions are presented as boxplots depicting the medians (circles) and the IQRs (boxes), with the whiskers showing the total range neglecting the outliers (max length 1.5*IQR from the box edge). Both average response strengths and response probabilities were then calculated for each step as in Sections 5.3.1 and 5.3.2 (channels averaged in the strength calculation). The results from both

developed methods are presented separately: C–D) Energy, E–F) SNR. Recordings using OR, GF, and GM stimuli were processed separately as well and are depicted with blue, green, and red, respectively, with black graph denoting an average of them.

The figure showed that higher gaze threshold generally produces stronger VEP responses, but on the contrary, the response probability tends to decrease at the same time. These observations were further assessed (see Fig. 21) by depicting the average response strength and probability (black curves from Fig. 20C–F) as functions of the median gaze quality and epoch quantity (circles in Fig. 20A–B).

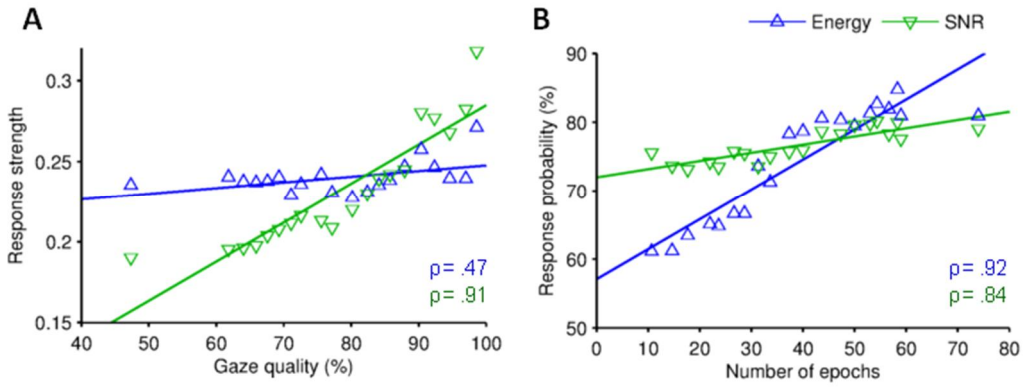

Figure 21. Influence of the gaze quality thresholding. A) The average response strength (stimulus types combined) is depicted as a function of the median gaze quality in the 20 gaze threshold optimization steps (see Fig. 20). B) The average response probability is presented correspondingly as a function of the median epoch quantity. The results from Energy and SNR analyses are illustrated separately with blue and green triangles, respectively. Trend lines calculated using the least-squares method are fitted for the data points to highlight the linear relationships.

The graphs revealed positive correlation (Energy: 0.47 and SNR: 0.91; Pearson) between the average response strength and the median gaze quality, especially when the SNR-method was used. In addition, there was also strong positive correlation (Energy: 0.92 and SNR: 0.84; Pearson) between the average response probability and the median epoch quantity. Putting these together, it meant that by raising the threshold, the epochs taken into the analyses possessed higher gaze quality and stronger VEP responses. Nevertheless, the statistical significance of the averaged responses actually weakened in the epoch rejection process because of the amount of the analyzed epochs diminished.

The assessments conducted above did not reveal one self-evident choice for the optimal threshold level, but they did show that too high threshold is likely to cause problems in the statistical analysis if the total amount of the data is limited. For the analyses presented in this work, a conservative threshold value of 20% was chosen. It yielded strong responses and high detection rate for both analysis methods by rejecting the epochs with none or very poor fixation on the presentation. Moreover, this choice was unlikely to exclude any recording (with generally poor gaze quality) completely from the analysis as a higher threshold would have done.

5.3.4 Evaluation of the reference VEP analysis method

In the reference works done with OR, GF, and GM stimulation (Braddick et al. 2005; Wattam-Bell et al. 2010) the presence of a VEP response was assessed using the Fourier-based T2circ statistic (Victor and Mast 1991; see Section 5.2.5). Although the current work presents two new VEP metrics, the gathered recordings were analyzed also with the conventional method (T2circ). The purpose was to benchmark the new metrics and to reveal how the recorded gaze data could improve the old analysis ways.

Similarly to Section 5.3.1, the OR, GF, and GM recordings for all 39 participants were analyzed separately, and the same eight EEG channels (P3, P4, O1, O2, T5, T6, Pz, and Oz) were taken into consideration and processed independently. The presence of a statistically significant VEP response was detected if the T2circ value within any channel exceeded a threshold level corresponding to false discovery rate corrected (Benjamini and Hochberg 1995) criterion of $p < 0.05$.

With the previously used gaze quality threshold of 20% the calculations yielded response probabilities as high as OR: 92%, GF: 100%, and GM: 95%, and by applying the requirement of 22 epochs for the amount of data the figures elevated even more to OR: 100%, GF: 100%, and GM: 97%.

It was also studied how changing the level of the gaze threshold would affect the results by reproducing the calculations presented earlier in Section 5.3.3 also using the T2circ method. The magnitudes of the T2circ statistic, reflecting the response strengths and the response detection rates were calculated for different gaze thresholds between 0 and 95% (see Fig. 22). The graphs revealed a clear and surprising tendency: the lower the gaze threshold was set the stronger and more frequent the detected VEP responses were. Extremely high response probabilities (> 95%) were registered with all threshold levels between 0% and 70%. Moreover, the most prominent responses were acquired at the gaze threshold of 0% where all epochs were analyzed regardless of the concurrent fixation level, which questions the need for the eye tracking altogether.

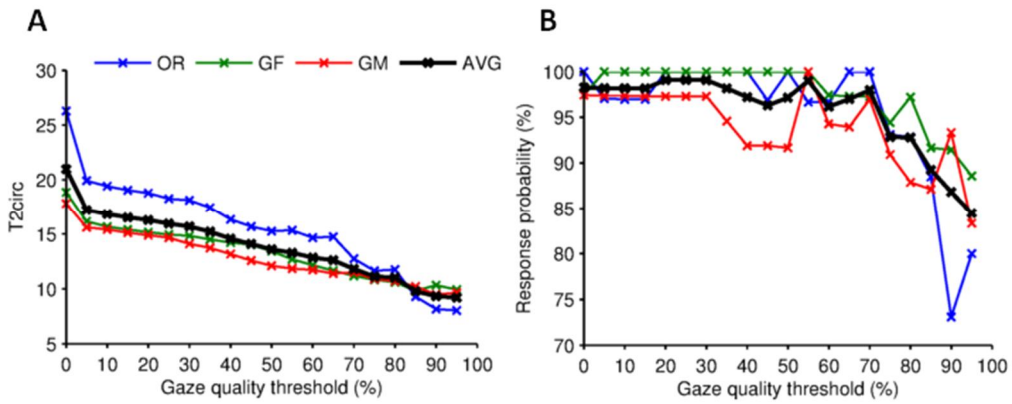

Figure 22. VEP detection results with T2circ statistic. For the benchmarking purposes, the calculations presented earlier in Figs. 20C–F for the novel VEP metrics were reproduced for the T2circ – the VEP detection method used in the reference studies. In figures A and B the magnitude of the T2circ statistic, reflecting the response strength, and the response detection probability were calculated as a function of the gaze quality threshold selection ranging from

0 to 95%. The graphs showed that using the T2circ, the lower the gaze threshold was set the stronger and more frequent the detected VEP responses were.

Although the participants were normally developing babies, and thus always likely to produce VEP responses at high rate, the observations raised questions about the specificity of the reference detection method. Of course, as there was no 'ground truth' information available about the 'true negative' findings at the moment, it was impossible to verify this hypothesis.

To further investigate the characteristics of the T2circ metric, supplementary, practical analyses were performed using simulated datasets instead of the actual recording data. The simulation datasets included variable numbers of 3.5-s-long noise and dummy VEP epochs (see Figs. 23A–B for two averaged examples). The dummy epochs consisted of two additive components: 4-Hz sine wave with 10-µV amplitude, roughly analogous to a VEP response and random pink (1/*f*) noise (V_{rms} 17 µV on average) emulating background EEG. The pure noise epochs in the datasets included only the latter component.

The sensitivity of the detection method was assessed by repetitive calculations of the T2circ statistic values from datasets that included an increasing number of noise epochs (Fig. 23C). As assumed, the T2circ values first decreased because of the added noise. Yet, despite this early decline the values constantly exceeded the threshold level corresponding to $p < 0.05$, even when simulation data consisted of 150 epochs of pure noise in addition to the original 10 dummy VEP epochs. Correspondingly, when the same simulation data were analyzed using the PSD-based metrics developed in this work, more expected results were gained (Fig. 23D). The measured response strengths fell rapidly to zero as the quantity of the noise epochs surpassed the number of the dummy response epochs.

The dependence of the T2circ method on the amount of analyzed data was assessed conversely. When an increasing number of dummy VEP epochs was added to the basis dataset of 10 noise epochs, the measured T2circ values rocketed exponentially (Fig. 23E). Often an addition of just three VEP epochs was enough to generate the response detection ($p < 0.05$). In the corresponding test with the same data using the Energy- and SNR-metrics (Fig. 23F), the measured response strength values increased as well, but at much more moderate rate and began to saturate after addition of 40 dummy epochs. The distinct curves seen in the 'Energy-graph' in this figure resulted from the IQR component of the response strength derivation (Eq. 5).

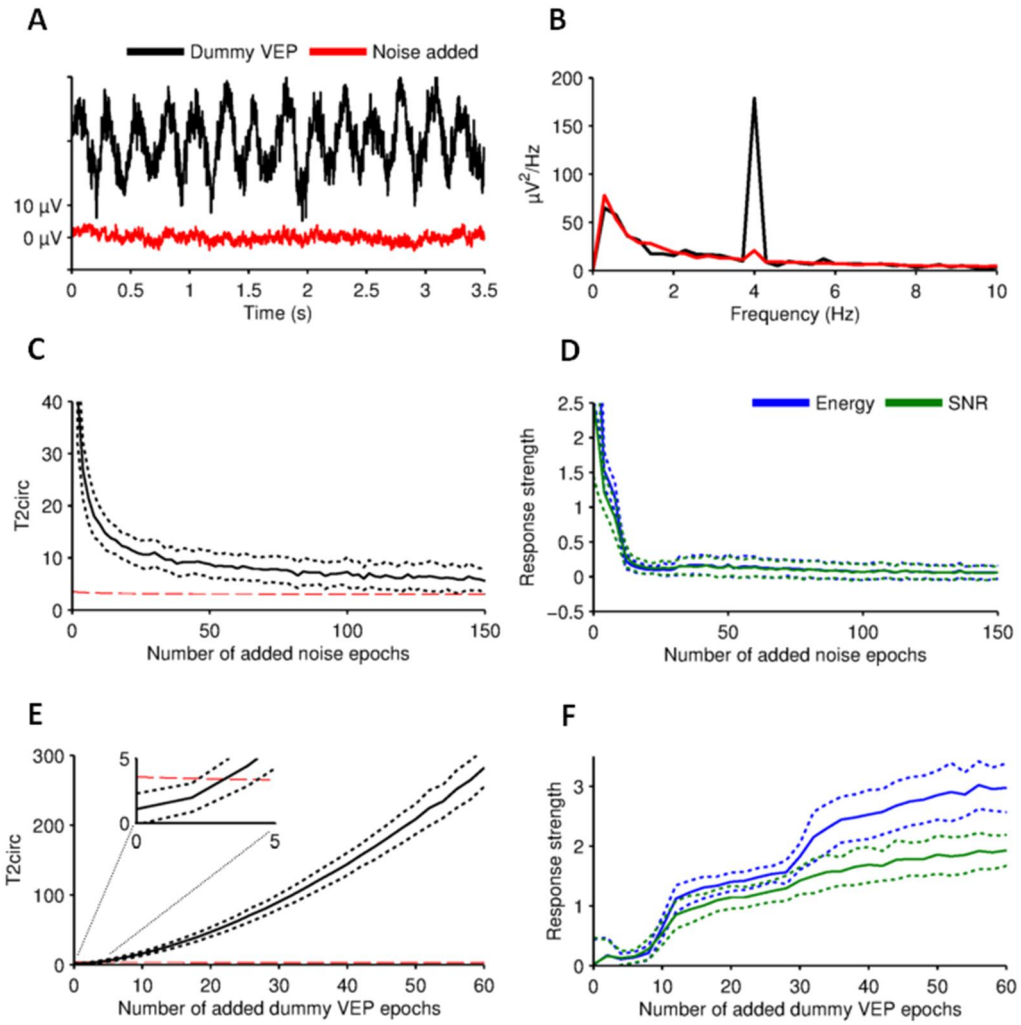

Figure 23. Evaluation of the T2circ analysis method with simulated VEP data. The supplementary T2circ analyses were performed using datasets that included noise and dummy VEP epochs. The dummy epochs consisted of two additive components: 4 Hz sine wave and random pink noise, whereas the noise epochs in the datasets included only the latter component. A–B) Averaged representations in time and frequency domains of two example datasets: one consisting 10 VEP epochs showing a clear response (black) and another where the response is shrouded by an addition of 150 noise epochs (red). C) The sensitivity of T2circ VEP detection method was evaluated by plotting the statistic values of the simulation data as a function of the number of noise epochs added to the dataset of 10 dummy VEP epochs. The red dashed line in the figure corresponds to the 0.05 significance level. Interestingly, despite the early decline the T2circ values constantly exceeded this threshold. D) For the comparison, the same data were similarly analyzed using the PSD-based metrics developed in this work. E) Similar analyses were also performed conversely by adding a number dummy VEP epochs to the basis dataset of 10 noise epochs. The figure shows how T2circ values rise exponentially as a function of the dummy epochs. F) Using the PSD-based metrics the measured response strengths increased as well, but at much slower rate and began to saturate after addition of 40 epochs. To account for the randomness in the simulation data, the calculations described above were repeated multiple times. In the figures C–F the solid and the dashed graphs depict averages and ±SD intervals of these repeated measurements, respectively.

5.4 Study-specific discussion

The starting point for this work was the SSVEP methodology presented in the prior studies by Wattam-Bell et al. (2010) and Braddick et al. (2005). The original idea was to adopt the stimuli and the analysis tools as such, but enhance the

recording setup with an eye tracker integration that could be used to control the stimulus presentation and to help excluding the moments of disoriented attention in the analysis.

The initial results with the new setup were puzzling. The calculations yielded highly promising response detection rates in the region of 95–100% that exceeded the expectations based on the prior works. The effect of the eye tracking based data reduction was, however, far from expected. The lower the gaze fixation threshold was set the more frequent and prominent the detected VEPs were, which questioned the benefits of the eye tracking data for the analysis altogether.

The further inspection with simulated data revealed that the cause for the observed behavior was the used T2circ -analysis method (Victor and Mast 1991). The tests showed that the method is very robust regarding the amount of noise added to the analyzed VEP epochs. In practice it means that the data with poor gaze quality does not significantly deteriorate the VEP detection as long as some brief segments with actual response are recorded as well. Although this is generally desirable in the context of challenging infant VEP studies, it was completely against the rationale behind the integration of the eye tracker and this study in general.

Considering the scope of this study, it was interesting to see how the use of gaze data worked with other SSVEP analysis techniques. Hence, two novel methods were developed that were based on spectral characteristics (energy and SNR) of a VEP signal. Like T2circ, the new metrics used the Fourier transform, but they did not exploit the phase information. The existence of possible response was determined by statistically comparing the spectral features with actual background EEG of the same individual recorded from the phases in between the stimulation sequences. Thus, the metrics did not require assumptions regarding the variance of the analyzed data (like T2circ).

The orientation reversal response rates measured using the developed analysis methods from the sample of 3-month-old infants (E: 94%, SNR: 84%) were in concordance with the reported figures of the previous works (Atkinson et al. 2002; Braddick et al. 2005). More deviation was found when global form and motion results were compared. The study by Wattam-Bell et al. (2010) showed that 92% of the 4–5-month-old infants produced a GM response, whereas for GF the same figure was only 50%. This is not fully consistent with our findings, where the responses were typically stronger for GF than for GM. Statistically the prior results are much closer to the observed response rates produced using Energy-metric (GF: 74%, GM: 81%) than SNR (GF: 84%, GM: 72%).

No multiple comparison correction of any kind has been applied to the statistical tests that yielded the detection rates reported above. The number of analyzed EEG channels was, however, reduced from 22 to 8 in order to control the potential chance of positive findings. Using the FDR correction technique (Benjamini and Hochberg 1995), also applied in the reference study (Wattam-Bell et al. 2010), the detection percentages decreased below the level suitable for the diagnostic use.

Beside the limited amount of the epochs, the main reason why the effect of the FDR correction was so dramatic is probably the dependency between EEG channels within a recording. The VEP responses are always likely be elicited on several electrodes simultaneously rather than strictly on one solitary location. Thus, the multiple comparison correction suitable for this kind of data would need to properly account for the correlations between the channels in order to avoid being overly conservative.

Consistently with the previous works (Wattam-Bell et al. 2010), the spatial distributions revealed distinct topographical response profiles for the three stimuli (compare Figs. 10 and 19). The OR and GF responses were clearly focused on the occipital region indicating that the physiological mechanisms underlying orientation- and form-specific responses could be closely linked. The GM stimulus yielded the spatially broadest response that covered a wide area ranging from the lateral posterior areas to the occipital midline regions at even magnitudes.

The observed inter-subject variability between the channel-specific response strengths was high in all electrode locations. Albeit the peculiar spatial distributions of the findings could well reflect the locations of the cortical generators of specific physiological mechanisms, the differences may as well be caused by the size differences of the neuron population responsible for the potential or their geometry in relation to the electrode positions on the scalp (Braddick et al. 2005). These factors could also explain the earlier discussed observations related to GF and GM responses. A true source localization study would require denser EEG recording than the current 22-channel setup.

The benefits of the gaze information turned out to be less valuable for the analysis than it was expected also when the novel analysis metrics were used (as it was with T2circ). The exclusion of the EEG segments with poor fixation on the stimulus did produce stronger VEP responses, but in the process the same data reduction undermined statistical significance of the remaining epochs remarkably. As a result a conservative gaze fixation threshold of 20% was found to be optimal in this sample of recordings.

Had the initial amount of recorded data been generally greater, the negative side of the eye tracking based data reduction might not have been that significant. In addition, by using shorter division (now 3.5 s) the epoch number could have been artificially increased so that more samples would have been affordable to waste due to the measured gaze fixation. Unfortunately, this adjustment would have probably caused subsequent issues with the weakening spectral resolution.

Finally, it should be also noted that data available in this study did not allow pure comparison of the results with and without the use of the eye tracker. Even when the gaze data based epoch selection was bypassed completely (using 0% threshold), the EEG had been already collected under the guidance of the eye tracker. The developed fixation based stimulus presentation ensured that the infant's attention was on the screen at least in the beginning of each 10–30s-long stimulus sequence (see Fig. 15), which must have substantially improved

the data quality especially with poorly co-operating subjects. This is also probably the reason why the response detection rates gained using the T2circ method exceeded the corresponding figures of the reference works.

Based on the experience of our recording sessions, the role of the developed presentation feedback mechanisms provided by the eye tracker should not be underestimated either. That would be another valuable asset favoring the inclusion of the eye tracking, if one of the presented, or any other, infant pattern VEP technique is eventually taken out of the research laboratory into the clinical routine.

6. Discussion

The two studies included in this Thesis presented two eye tracker assisted research methods that may be used to evaluate the functions of infant visual processing at individual level. The novel metrics developed in the studies were targeted to different physiological mechanisms.

The first study focused on visuospatial attention and face preference using the classic overlap paradigm. It showed that three metrics could be used to characterize the essence of the measured gaze patterns, and tap partly different cognitive operations. Conventional Disengagement Time and Disengagement Probability measures are calculated from the initial gaze shifts driven by reflexive saccade in response to the sudden onset of the peripheral stimulus, and could reflect the functioning of the subcortical systems important for reflexive orienting responses in infants. On the other hand, the novel Mean Deviance does not measure infant's response to experimental triggers, but instead reflects the spontaneous tendency to look back at face after momentary distraction (Carpenter et al. 1998) and cumulative preference for social over non-social stimuli (Pierce et al. 2011). Hence, MD may engage more voluntary and goal-directed attentional control mechanisms, centered in the dorsal parietal and frontal cortex areas.

The focus of the second study was on SSVEP responses elicited by orientation reversal, global form, and global motion stimuli. The purpose of using three stimulus types was, again, to target complementary aspects of integrative cortical visual processing at striate and extrastriate areas. Orientation-specific responses that are essential, for example, in object recognition can be generated only in the primary visual cortex and also partly in further extrastriate visual areas. The absence or late development of the OR response are likely to indicate deviation in the thalamocortical connections and correlate with the neurological outcome at later age. (Mercuri et al. 1999). Additionally, the measurements of GF and GM responses allow direct comparison of the functioning of the extrastriate areas in ventral (e.g., V4) and dorsal (e.g., V5) pathways. The processing of the two stimuli should be very similar regarding the attentional mechanisms and at the level of primary visual cortex, but differ in the interplay with the extrastriate areas. (Braddick and Atkinson 2011)

Despite the evident promise shown in the studies, there still lie several obstacles ahead that need further consideration in order to take the novel methods into the clinical diagnostics. These are discussed in the next sections along with the future objectives and potential methodological improvements.

6.1 Future objectives

In the first study using the classic face–distractor competition paradigm almost half of all the measured trials did not result in a gaze movement away from the center within the first 1000 ms after the onset of the peripheral stimulus. This is particularly concerning as the applied metrics (except DP) concentrated only on the trials with the gaze shift and excluded the rest. It could be argued that this excessive amount of 'unsuccessful' trials may easily increase the random-effects in the derivation of DTs and MDs that are based on the trial averages. The total number of trials (test duration) cannot be raised remarkably as the probability of the saccade misses was observed to generally increase during the course of the test session.

Better alternative for future studies would be to reconsider the nature of the distractor. The geometrically shaped targets used in this work could be made more attractive (e.g., simple animations) to ensure an increased rate of gaze shifts throughout the experiment. The analysis could be made even more robust by simplifying the paradigm to consist of only one face and one non-face condition, which would also reduce heterogeneity in the trials pooled together. Although the role of the phase-scrambled stimulus was minor in this study, it should be retained in the experiment. In future it could be incorporated in the metrics as a baseline reference that would even out potential differences caused by external factors such as fatigue or habituation.

The developed Mean Deviance metric is based on comparing the observed cumulative deviation of gaze against the normative ranges, and also the evaluation of DTs could be done similarly in the future (proposed Z-score method). However, in the study the primary sample that was used to generate this normative data consisted of just 13 infants without information about their ultimate neurocognitive development. A genuine normative dataset should obviously be much larger than that to minimize the chances of data being influenced by extreme values. Although the pilot observations were found encouraging, it is obvious that future studies with standardized cohort characterization will be needed to establish genuine normative values and their developmental trajectories.

The second study of eye tracker integrated visual evoked potentials left also room for future improvements. In the presented setup, the synchronization between EEG and eye tracker was achieved using a self-built light sensor that was attached on the eye tracker screen. The output of the sensor system was a square wave varying between -100 and 100 µV voltage depending on the observed intensity, and further signal processing was needed to extract the timings of the trigger events. Since the dates of the recordings included in the study, our laboratory setup has already been improved. It is now equipped with a new EEG device capable of receiving TTL-level trigger pulses. This possibility is exploited by a new synchronization device (Cedrus StimTracker) designed to couple EEG with external devices. It is also based on similar photodiode technique, but with digital output and a sensor that fits on the screen better, the offline synchronization steps in MATLAB should now require considerably less manual work.

The new solution should also improve the trigger accuracy reducing the potential leakage of the response power to the neighbouring bins in the spectral analysis, especially if the epoch length was to be shortened to increase the eye tracking gain. In future, trigger output from E-Prime controlled LPT/COM port of the eye tracker computer could also be enabled as a backup mechanism for the synchronization.

The display currently integrated to the eye tracker (Tobii T120) may not be perfect for VEP presentation (no data available to verify this) and potential deficits in screen luminance, contrast, or frame rate could affect the stability and reliability of the VEP responses. In future, it would be better to change the eye tracker to a lightweight, portable model (for example Tobii X2-60) that could be freely combined with a monitor known to be suitable for the VEP presentation. Additionally, using the new MATLAB bindings included in the Tobii Analytics SDK the control of the eye tracking and the stimulus presentation could now be migrated from the E-Prime to MATLAB platform. Hence, the cumbersome video encoding/decoding step in the stimulus creation could be replaced by online use of the Psychophysics toolbox that directly commands the graphics accelerator hardware through OpenGL interface.

Finally, the analysis of the VEP recordings often revealed distinct peaks also at the harmonic frequencies (multiples of the stimulation frequency) of the spectrum. This indicates that waveforms of the VEP responses are not likely to be purely sinusoidal, which consequently questions the use of the Fourier-based signal processing methods such as FFT and PSD. More sophisticated techniques like wavelet representations or matched filters calculating correlation between measured signal and preset VEP pattern template could provide better results in the VEP feature extraction. The topographical analyses could be improved, obviously, by increasing the electrode number, but also with new signal processing methods focusing on the source localization or connectivity between different brain areas.

6.2 Prospects of clinical use

The presented studies and results of the prior works suggest that the described methods could eventually turn out to be valuable tools in prospective clinical use in differential diagnostics and developmental assessments. The presented metrics may not correlate uniquely to known neurological mechanisms, but they are clearly reflecting some key functions of the visual processing and thus provide general, yet quantifiable and objective information about the entity that is the cerebral functions of an infant.

In clinical use, it would be fascinating to combine them with a battery of other diagnostic tests that would complement each other by monitoring different aspects of the core infant brain functions. Different tests would be performed at different stages of the development during infantile period for a patient in high-risk groups (prematurity, birth asphyxia, metabolic disturbances, early sepsis, stroke, infections, etc.). These measured biomarkers could then guide the up-

coming decisions related to treatment, prediction, or detection of potential neurological adversities. In our vision, the use of tests of this kind would take place in clinics that now follow the neurodevelopment of children, i.e., specialized centers associated with intensive care units and neuropediatric expertise.

The data analyses in these studies needed special signal processing expertise. The purpose throughout the study, however, was to design the methods so that they could be implemented as automated applications putting less demand for personnel administering the tests. The end product of this laborious work should be a script package or program ready for larger scale studies that not only provides feedback during the test session but also immediately output the analyses at the end of it. Given that the study procedures were already mostly automated, running this system requires less than an hour of training from people with basic computer skills and expertise with EEG recordings.

The practical setup could be similar to what we already use in our work. However, attention should be paid to maximize the usability and mobility of the system considering the general challenges in the conditions of the proposed test locations. The cost of eye tracking devices has been an issue, but recent development of consumer applications using the eye tracking technology has already dropped the pricing to the level of any smaller medical devices.

Applicability of the developed methods to the individual level for clinical evaluation requires validation that is conventionally done by calculation of statistical performance indicators of sensitivity and specificity. The orthodox derivation of these figures would need ‘ground truths’ of the studied features of interest. The attempted use of 5-month-olds as a reference for the deviant findings in the overlap paradigm was partly flawed (being atypical is not equivalent to being younger). On the other hand, in the VEP study all measured infants were, based on their age and history, principally likely to yield responses. Because of these challenges, the validation of the methods will probably come from the test of time by outperforming the conventional behavioral assessments in study cohorts, and providing a perceived added value to studies seeking early developmental biomarkers. Validation in this way will be based on larger scale recruitments of both typically and atypically developing infants, which is already running in our and other laboratories.

7. Conclusion

This Thesis presented two eye tracker assisted test paradigms that may be used to evaluate and quantify the functions of visual cortical processing, attention and face preference from infants aged from 3 to 7 months. After further development planned for the future, the methods are ready to be used clinically in differential diagnostics and assessments of neurocognitive development preferably alongside other similar biomarker tests of infancy.

The presented studies revealed how the eye tracking may be exploited in different ways, and what are the technical constraints required for the use of this technology. The first study was concentrated on the analysis of the measured gaze patterns in classic face–distractor competition paradigm, and showed that three metrics, tapping partly different cognitive functions, could be used to characterize the essence of the test results. One of these metrics was developed by the author to measure cumulative allocation of attention between the central and the peripheral stimuli from the infant's gaze dynamics over a given period of time. In further evaluation, the metric was shown to be sensitive to known developmental changes in infants' face processing between 5 and 7 months of age.

The second study focused on the visual evoked potentials elicited by orientation reversal, global form, and global motion stimulation. The VEP data were recorded from 39 3-month-old infants and analyzed using T2circ statistic (as in the reference works) and with two developed PSD-based metrics. The eye tracking was used to control stimulus presentation to capture the attention of the infant, and later in the analysis to exclude the EEG segments with disoriented gaze. With this setup, VEPs could be reliably detected from the vast majority of the participants with all the stimulus types. The effect of the eye tracking turned out to be remarkable during the recordings, but minor (or even negative with T2circ) in the analysis where the data reduction had a tendency to undermine the statistical significance of the VEP responses.

References

Ahtola, E., Stjerna, S., Yrttiaho, S., Nelson, C. A., Leppänen, J. M., and Vanhatalo, S. (2014). Dynamic eye tracking based metrics for infant gaze patterns in the face-distractor competition paradigm. *PloS One*, 9(5).

Aslin, R. N. and Salapatek P. (1975). Saccadic localization of visual targets by the very young human infant. *Perception & Psychophysics*, 17(3):293–302.

Atkinson, J., Anker, S., Rae, S., Weeks, F., Braddick, O., and Rennie, J. (2002). Cortical visual evoked potentials in very low birthweight premature infants. *Archives of Disease in Childhood. Fetal and Neonatal Edition*, 86(1):F28–31.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300.

Blaga, O. M. and Colombo, J. (2006). Visual processing and infant ocular Latencies in the overlap paradigm. *Developmental Psychology*, 42(6):1069–1076.

Braddick, O. J., Wattam-Bell, J., and Atkinson, J. (1986). Orientation-specific cortical responses develop in early infancy. *Nature*, 320(6063):617–619.

Braddick, O. J., O'Brien, J. M. D., Bell, W. J., Atkinson, J., and Turner, R. (2000). Form and motion coherence activate independent, but not dorsal/ventral segregated, networks in the human brain. *Current Biology*, 10(12):731–734.

Braddick, O., Birtles, D., Wattam-Bell, J., and Atkinson, J. (2005). Motion- and orientation-specific cortical responses in infancy. *Vision Research*, 45(25–26):3169–3179.

Braddick, O. and Atkinson, J. (2011). Development of human visual function. *Vision Research*, 51(13):1588–1609.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4):433–436.

Carpenter, M., Nagell, K., and Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development,* 63(4):1–143.

Corbetta, M. and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience,* 3(3):201–215.

Csibra, G., Tucker, L. A., and Johnson, M. H. (1998). Neural correlates of saccade planning in infants: A high-density ERP study. *International Journal of Psychophysiology*, 29(2):201–215.

Duchowski, A. T. (2007). *Eye Tracking Methodology: theory and practice, 2nd ed.,* London: Springer.

Forssman, L., Peltola M. J., Yrttiaho, S., Puura, K., Mononen, N., Lehtimäki, T., and Leppänen, J. M. (2013). Regulatory variant of the tryptophan hydroxylase 2 gene enhances infants' attention to social signals of fear. *Journal of Child Psychology and Psychiatry,* 55(7):793–801.

Frank, M. C., Vul, E., and Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition*, 110(2):160–170.

Green, C. R., Mihic, A. M., Brien, D. C., Armstrong, I. T., Nikkel, S. M., Stade, B. C., Rasmussen, C., Muñoz, D. P., and Reynolds, J. N. (2009). Oculomotor control in children with fetal alcohol spectrum disorders assessed using a mobile eye-tracking laboratory. *European Journal of Neuroscience*, 29(6):1302–1309.

Goodale, M. A. and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1):20–25.

Hammoud, R. I. (2008). *Passive Eye Monitoring: Algorithms, Applications and Experiments, Series: signals and communication technology.* Berlin: Springer-Verlag.

Holmboe, K., Fearon, R. M., Csibra, G., Tucker, L. A., and Johnson, M. H. (2008). Freezeframe: A new infant inhibition task and its relation to frontal cortex tasks duringinfancy and early childhood. *Journal of Experimental Child Psychology*, 100(2):89–114.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., and van de Weijer, J. (2011). *Eye Tracking – A comprehensive guide to methods and measures.* Oxford: Oxford University Press.

Hunnius, S., Geuze, R. H., and van Geert, P. (2006). Associations between the developmental trajectories of visual scanning and disengagement of attention in infants. *Infant Behavior and Development,* 29(1):108–125.

Hunnius, S. (2007). The early development of visual attention and its implications for social and cognitive development. *Progress in Brain Research,* 164:187–209.

Johnson, M. H., Posner, M. I., and Rothbart, M. K. (1991). Components of visual orienting in early infancy: Contingency learning, anticipatory looking and disengaging. *Journal of Cognitive Neuroscience,* 3(4):335–344.

Johnson, M. H. (2005). Subcortical face processing. *Nature Reviews Neuroscience*, 6(10):766–774.

Kalesnykas, R. P. and Hallett, P. E. (1987). The differentiation of visually guided and anticipatory saccades in gap and overlap paradigms. *Experimental Brain Research,* 68(1):115–121.

Lee, J., Birtles, D., Wattam-Bell, J., Atkinson, J., and Braddick, O. (2012a). Latency measures of pattern-reversal VEP in adults and infants: different information from transient p1 response and steady-state phase. *Investigative Ophthalmology & Visual Science*, 53(3):1306–1314.

Lee, J., Birtles, D., Wattam-Bell, J., Atkinson, J., and Braddick, O. (2012b). Orientation-reversal VEP: comparison of phase and peak latencies in adults and infants. *Vision Research*, 63:50–57.

Leppänen, J. M., Peltola, M. J., Mäntymaa, M., Koivuluoma, M., Salminen, A., and Puura, K. (2010). Cardiac and behavioral evidence for emotional influences on attention in 7-month-old infants. *International Journal of Behavioral Development,* 34(6):547–553.

Leppänen, J. M., Peltola, M. J., Puura, K., Mäntymaa, M., Mononen, N., and Lehtimäki, T. (2011). Serotonin and early cognitive development: Variation in the tryptophan hydroxylase 2 gene is associated with visual attention in 7-month-old infants. *Journal of Child Psychology and Psychiatry,* 52(11):1144–1152.

Matsuzava, M., Shimojo, S. (1997). Infants' fast saccades in the gap paradigm and development of visual attention. *Infant Behavior and Development,* 20(4):449–455.

Mercuri, E., Haataja, L., Guzzetta, A., Anker, S., Cowan, F., Rutherford, M., Andrew, R., Braddick, O., Cioni, G., Dubowitz, L., and Atkinson, J. (1999). Visual function in term infants with hypoxic-ischaemic insults: correlation with neurodevelopment at 2 years of age. *Archives of Disease in Childhood. Fetal and Neonatal Edition*, 80(2):F99–104.

Muñoz, D. P. and Everling, S. (2004). Look away: The anti-saccade task and the voluntary control of eye movement. *Nature Reviews Neuroscience,* 5(3):218–228.

Muñoz, L. C. (2009). Callous-unemotional traits are related to combined deficits in recognizing afraid faces and body poses. *Journal of the American Academy of Child and Adolescent Psychiatry*, 48(5):554–562

Morgante, J. D., Zolfaghari, R., and Johnson, S. P. (2012). A critical test of temporal and spatial accuracy of the tobii T60XL eye tracker. *Infancy*, 17(1):9–32.

Nicholls, J. G., Martin, A. R., Wallace, B. G., and Fuchs, P. A. (2001). *From Neuron to Brain. 4th ed.* Sunderland (MA): Sinauer Associates Incorporated.

Odom, J. V., Bach, M., Brigell, M., Holder, G. E., McCulloch, D. L., Tormene, A. P. P., and Vaegan (2010). ISCEV standard for clinical visual evoked potentials (2009 update). *Documenta Ophthalmologica. Advances in Ophthalmology*, 120(1):111–119.

Peltola, M. J., Leppänen, J. M., Vogel-Farley, V. K., Hietanen, J. K., and Nelson, C. A. (2009). Fearful faces but not fearful eyes alone delay attention disengagement in 7-month-old infants. *Emotion,* 9(4):560–565.

Peltola, M. J., Leppänen, J. M., and Hietanen, J. K. (2011). Enhanced cardiac and attentional responding to fearful faces in 7-month-old infants. *Psychophysiology,* 48(9):1291–1298.

Peltola, M. J., Hietanen, J. K., Forssman, L., and Leppänen, J. M. (2013). The emergence and stability of the attentional bias to fearful faces in infancy. *Infancy,* 18(6):905–926.

Pierce, K., Conant, D., Hazin, R., Stoner, R., and Desmond, J. (2011). Preference for geometric patterns early in life as a risk factor for autism. *Archives of General Psychiatry*, 68(1):101–109.

Posner, M. I. and Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, 13:25–42.

Tobii Technology AB (2010). Tobii eye tracking: An introduction to eye tracking and Tobii Eye Trackers. (White paper). Retrieved October 15, 2014 from www.tobii.com/eye-tracking-research/global/library/white-papers/tobii-eye-tracking-white-paper.

Tortora, G. J. and Grabowski, S. (2003). *Principles of Anatomy and Physiology, 10th ed.* New York (NY): J. Wiley & Sons.

Vialatte, F. B., Maurice, M., Dauwels, J., and Cichocki, A. (2010). Steady-state visually evoked potentials: focus on essential paradigms and future perspectives. *Progress in neurobiology*, 90(4):418–438.

Victor, J. D. and Mast, J. (1991). A new statistic for steady-state evoked potentials. *Electroencephalography and Clinical Neurophysiology*, 78(5):378–388.

Wass, S. V., Porayska-Pomsta, K., and Johnson, M. (2011). Training attentional control in infancy. *Current Biology,* 21(18):1543–1547.

Wattam-Bell, J., Birtles, D., Nyström, P., von Hofsten, C., Rosander, K., Anker, S., Atkinson, J., and Braddick, O. (2010). Reorganization of global form and motion processing during human visual development. *Current Biology*, 20(5):411–415.