Towards the improvement of event-related potential techniques in developmental populations

Estefanía Domínguez Martínez, M.Sc.

Thesis submitted for the degree of
Doctor of Philosophy

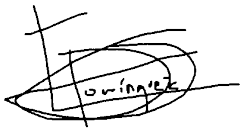
Department of Psychology

Lancaster University

July 2017

Declaration

This thesis is my own work and no portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification at this or any other institute of learning.



Estefanía Domínguez Martínez 22^{nd} of July 2017

Towards the improvement of event-related potential techniques in developmental populations, Estefanía Domínguez Martínez, M.Sc.

Submitted for the degree of Doctor of Philosophy, July 2017

Abstract

Visual event-related potentials (ERPs) have been widely used during the past two decades to study cognition in developmental populations. One of the main challenges related to current infant ERP practices is the need to obtain good-quality ERPs that provide reliable results with which to examine research issues. There are two main aspects that play a role in obtaining a valid ERP trial: the infant's attention to the stimuli and an EEG signal free of artefacts—primarily eye and body movement artefacts—. The aims of this thesis were to explore alternative methodological approaches that could potentially improve current ERP practice as well as to assess the validity of the current ERP editing methods. Specifically, the first aim was to study the use of eye tracking techniques to improve current ERP data collection methods and ERP final data quality. In Chapter 2, eye tracking technology was used as a mean to study how advanced ERP components are negatively affected by the fixation location distance to the stimulus. Chapter 3 evaluated the changes in attention and visual behaviour of a gaze-contingent paradigm used during ERP data collection in 10-month-old infants. The second aim was to investigate to what extent the current infant ERP editing methods are reliable and to assess the need for their further standardization within the field. Chapter 3 focused on understanding the current editing methods used in ERP infant research for selecting ERP trials free of artefacts. The study analyzed the agreement between three experienced infant

ERP human editors and one algorithm for selecting artefact-free ERP trials as well as the EEG signal characteristics that influenced the selection of valid trials. The implications of the results of the three studies presented in this thesis for current infant ERP practices are discussed in Chapter 5. Overall, this thesis highlights the importance and need of the field to work towards the improvement of current methodologies for collecting and analyzing infant ERP data.

Acknowledgements

First of all, I want to express my gratitude to my supervisor, Dr. Vincent Reid for giving me the opportunity of starting a PhD under his supervision when my knowledge in the field of developmental psychology was still very limited. Thanks to your support and guidance, I've been able to acquire new skills and knowledge that had never thought I would be able to learn. I also appreciate that you always adapted to my remote situation and did your best to be available to meet whenever I needed, and always cared not only about my research work progress, but about my work-life balance. I would also want to thank to Tommy Strandvall, who was involved in my supervision during the first half of my thesis and who also contributed to my work with his eye tracking expertise and support.

Even though I conducted most of my work outside Lancaster, I always felt a part of the psychology department, especially during my stays at Lancaster. I was very lucky to have such good colleagues (Malen, James and Claire to name a few) that always welcomed me and involved me both in work discussions and social events. A especial mention to Katherina, who I have to thank for being the person who taught me how to properly conduct an EEG experiment. I also want to have a especial word for Áine, who, together with Kath, have always taken special care of me, have shared all their extensive knowledge and have made my stay at Lancaster enjoyable and

fun. Finally, my journey of learning about psychology having a technical background would have been much harder if I did not have the support of my colleague and friend Óscar. I always enjoyed our trips around Europe where we had time to work, to enjoy together our passion for signal analysis and outdoor activities, and to discuss about life.

I spent most of my period as a PhD student at Tobii. I want to thank my Tobii colleagues for making my day-to-day office life fun and for teaching me so many things about eye tracking, industry and ways of having fun in Stockholm. Especial mention to Tommy Strandvall for his supervision and Nina for being my research mate at Tobii. We have experienced and learned together many things about being a PhD student far away from the home institution, and also enjoyed some fun Nordic outdoors activities together. Of course, I cannot forget to thank Arzu, my friend, my very best friend, for all her psychological support during these years, for saving me from a dull life in Stockholm, and for sharing with me thousands of good moments.

I want to thank my mum, dad and sister for always being there, for believing in me, and supporting me in all the decisions that I have taken throughout my life, even when that meant leaving Spain and seeing me a few times a year. I know that my dad, who is not here anymore to see this accomplishment, would have used his particular language to express how proud he is of me, something like "I am not surprised, I knew that you could do it". Thank you for always believing in me so blindly, that has helped me and will help me a lot in my life. And mum, thank you for being the best mother one can ask for, I would not be half the person I am if it were not for you. Mum, dad, sister, Eric, Paula: I love you all.

Finally, my partner Nuria deserves to have a final and special remark. She has

been in my life only for half of my PhD student period, but she has contributed in many ways to the successfulness of this thesis. First of all, making my life happier and better in all senses, which has given me all the mental strength that I needed to finish it. Also, for always listening to me when I had doubts about my work, many times, and for encouraging me to achieve my goals. I look forward to many more achievement journeys together with you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction to infant ERP research

Electroencephalography (EEG) is a method for recording the electrical activity of the brain from electrodes placed on the scalp. EEG measures voltage fluctuations and primarily reflects the sum of the postsynaptic potentials of neurons in the cortex (Davidson, Jackson, & Larson, 2000). The main advantages of EEG when contrasted with other measures of brain activity are its non-invasiveness and its high temporal resolution. Event-related potential (ERP) is a methodology that uses EEG to study the cognitive responses caused by specific external events. ERPs have been used for decades in infant research to understand different cognitive functions as they develop in human infants in typical and atypical developmental populations (Nelson & McCleery, 2008; Reynolds & Guy, 2012). The main research areas where ERP has been extensively applied to date are attention, memory, action understanding, face processing, language development and emotions (for a review, see De Haan, 2007a). Despite all the advantages that ERP offers, it also comes with a number of challenges when applied to infant samples. These challenges are mainly related with

the impossibility of giving instructions during the experiment and infants' attention span. The current thesis' aim is to identify and address some of these important issues as well as study its consequences.

One of the reasons why ERP has become popular in infant research is because it is one of the few non-invasive methodologies that allows to study healthy infants' brain activity. A cap or net must be placed on the head of the infant that typically contains the electrodes that will record the electrical activity. Some EEG manufacturers (e.g., EGI®, Brain Products®, Biosemi®) have developed specific EEG systems adapted to infants. They are designed to fit infants' head and facilitate a quick setup as infants do not tolerate long preparation times. Once the cap is on the infant's head and the recording has started, the ERP paradigm begins. The infant is exposed to the same environmental events many times (from dozens to hundreds of times). The general ERP methodology requires several repetitions of the same stimuli because its principle is based on averaging the time-locked brain activity to the same events. By doing this, the brain response caused by the event—considered the event-related signal—is enhanced, and the rest of the brain activity that is not associated to the event —considered random signal—is diminished (Luck, 2005; Otten & Rugg, 2005).

An ERP consists of the sum of different components that are related to brain processes (Luck, 2005). In both adults and infants, the ERP components can be divided in early components and late components. The former, related to rapid processing of the stimulus (e.g., visual processing), are those within 200 ms after the onset of the stimulus. The latter, typically related to more complex cognitive processing of the stimulus, can be observed from 200 ms up to few seconds after the onset of the stimulus (e.g., semantic processing). Each component is related with

one or more cognitive processes. Thus, for example, the visual N290 and P400 components are sensitive to face processing (H. Halit, De Haan, & Johnson, 2003) and the positive slow wave has been related to the updating of memory representation in infants (Nelson, 1994). Amongst the most studied of the developmental ERP components is the Negative Central (Nc) component. It is a negative deflection observed mainly over the frontal and central electrodes. The latency varies with age, peaking around 400-800 ms at 6 months of age. This component is thought to reflect attentional orienting to interesting stimuli (Courchesne, Ganz, & Norcia, 1981; Richards, Reynolds, & Courage, 2010) and sustained attention and attention arousal (Richards, 2003). The Nc is elicited by any type of visual stimuli, thus it has been suggested that it may reflect "automatic" orienting (De Haan, 2007b).

Infant ERP components change during the first postnatal year and differ in amplitude and latency from adult ERP components (Picton & Taylor, 2007; Webb, Long, & Nelson, 2005). In general, infants' ERP components are greater in amplitude and latencies than in adults. Some infant ERP components are suggested to evolve in terms of amplitude and latency with increasing age and become adult ERP components. This is the case of the face-sensitive components N290 and P400, which have been associated as possible precursors of the also face-sensitive component N170 in adults (Halit, Csibra, Volein, & Johnson, 2004; Leppänen, Moulson, Vogel-Farley, & Nelson, 2007) or the N400, which keeps the same name in infants and adults although it is delayed in infants. This component has been related to semantic processing of language and action and appears in unanticipated conditions (Kaduk et al., 2016; Reid et al., 2009).

ERP components vary its amplitude, latency and duration with the stimulus

properties (e.g., color, congruency, luminosity) and cognitive tasks (e.g., attention, passive listening, semantic categorization). The general approach in ERP research is to compare the same ERP component under study elicited by two or more different conditions. Conditions are supposed to vary only those properties of the stimulus that are related to the research question, leaving other factors as similar as possible (e.g., luminosity, color if the research question involves a change in a cognitive task; Luck, 2005). Any difference in the ERP components can then be assumed to be produced by the stimulus property that varied between conditions. The most common measures are the amplitude and latency differences between ERP components. There are widely applied paradigms that are known to create a difference in specific ERP components. This is the case with the oddball paradigm, which has a standard condition shown frequently (more than 50% of the trials) and an oddball condition for the rest of the trials that elicits differences in ERP components (for a review, see Friedman, Cycowicz, & Gaeta, 2001). Another example are the paradigms where there are two conditions: congruent and incongruent. As in the oddball paradigm, the incongruent condition will create a different cognitive process and therefore differences in the ERP components can be measured (e.g., Reid et al., 2009; Reid & Striano, 2008). In infant ERP research, these types of paradigms are used to study whether infants have already developed the ability to understand or process certain forms of information. That is, if infants are able to distinguish between the conditions under study at a certain age their cognitive processes will be different and therefore the ERPs will reflect this. If, on the contrary, at a certain age infants do not show differences in the ERP elicited by different conditions, it can be inferred that they have not developed the ability to understand or process the perceptual or cognitive differences between conditions. The ERP technique has the

advantage of being suitable from very young infants to older children and adults. This enhances the possibilities of comparing the cognitive responses of different developmental groups to the same stimulus and, therefore, provide valuable insights into the cognitive changes that occur with development.

ERP in infants are typically limited to passive viewing or listening paradigms. The nature of the ERP stimuli can be purely auditory (e.g., Choudhury & Benasich, 2011; C. He & Trainor, 2009; Tew, Fujioka, He, & Trainor, 2009), purely visual (e.g., Bakker, Kaduk, Elsner, Juvrud, & Gredebäck, 2015; Kaduk et al., 2016; Leppänen et al., 2007) or a combination of both (e.g., Friedrich & Friederici, 2010; Hyde, Jones, Porter, & Flom, 2010; Junge, Cutler, & Hagoort, 2012). Because of the EEG high temporal resolution and, consequently, the importance of time-locking the brain response elicited by the stimulus with a high timing precision, the stimuli are usually controlled via computers since they are able to accurately mark the onset of the stimuli in the EEG data. Auditory events are usually presented via loudspeakers. In the case of visual ERP experiments, there have been some interesting attempts to show real world objects to infants (e.g., Carver, Meltzoff, & Dawson, 2006) but the most broadly used setup is still to present the stimuli on monitors in the form of still images or videos. One of the issues addressed by this thesis is related to how visual attention affects the ERP quality and the study of how to improve infants' attention to the stimuli with the use of eye tracking. Therefore, from now on, the emphasis of this thesis will be on visual ERP experiments.

1.2 Current infant ERP methodology

The main principles of the ERP methodology are common to both infant and adult participants, particularly, the stimuli repetition and the inclusion of trials without EEG artifacts in the final ERP. In order to obtain a visual ERP in any population, it is necessary to show the participant the same stimulus dozens of times (i.e., ERP trials) and average the time-locked brain responses. Also, only clean ERP trials without artifacts should be included in the final sample to compute the ERP. To ensure a good-quality EEG signal, the participant should move as little as possible, minimize eye movements and pay attention to the stimulus. This is relatively easy to accomplish in practice with adult participants since they can receive instructions and behave accordingly during an ERP experiment. When the tested population are infants or young children, avoiding body and eye movements and paying attention to the stimuli become a challenge. The impossibility of giving instructions during an experiment is a common issue across infancy research techniques, but it is particularly critical for ERP studies due to the number of repetitions of the same stimulus, which causes boredom and a rapid lack of interest to the stimuli.

General ERP methodology literature describes the guidelines related to how to record and analyze ERPs (e.g., Luck, 2005; Picton et al., 2000). They mainly focus on adult studies and point out that special measures need to be adopted in clinical and developmental studies, although do not elaborate in detail on these topics. Several authors have contributed to the methodological literature specifically focusing on infant ERP studies, providing guidance and good practice related to how to record and analyze infant ERP data (Csibra, Kushnerenko, & Grossmann, 2008; DeBoer, Scott, & Nelson, 2007; Hoehl & Wahl, 2012; M. H. Johnson et al.,

2001; Nelson & McCleery, 2008). These guidelines typically identify the challenges that arise from having infants as test subjects, the common approaches applied in literature to deal with them and practical advice for researchers new to the infant ERP field.

1.2.1 Data collection methodology

The biggest challenges identified by the literature during an infant ERP experiment are: body movements and the limited attention span of the infant participants. Body movement activity can either produce an electrical signal that will be picked up by the electrodes (i.e., muscle activity or eye movements) or can widely distort the EEG signal by rapid impedance changes (i.e., electrode or cable movements). Eye movements and blinks are especially a risk because of their near location to EEG electrodes. The general method is to avoid eliciting these as much as possible during the trial presentation since they widely distort the EEG signal. In an infant ERP experiment, a common approach to limit eye movements is to present the stimuli at a relatively small size. Stimuli sizes vary between studies. For infants older than 6 months of age, usually the stimuli sizes range from 5° to 15° of visual angle (e.g., Grossmann, Striano, & Friederic, 2006; Hoehl & Striano, 2008; Kaduk et al., 2016; Otte, Donkers, Braeken, & Van den Bergh, 2015; Webb, Dawson, Bernier, & Panagiotides, 2006).

Paying attention to the stimuli is particularly critical in visual paradigms because a brain response will only be generated if the infant is visually attending to the stimuli. Besides general lab recommendations such as creating a non-distractive

environment (Hoehl & Wahl, 2012), there are some methods used in infant ERP experimental design that aim to maintain infants' attention to the stimuli for as long as possible. First, the most obvious one, is to use, when possible, infant-attractive stimuli like geometric shapes, human faces or toys (e.g., "Visual statistical learning of shape sequences: An ERP study", 2009; Farroni, Massaccesi, Menon, & Johnson, 2007; De Haan & Nelson, 1997; Hoehl, Reid, Mooney, & Striano, 2008). Second, a complementary method used whenever possible is to create different stimuli that depict the same condition. A suitable case is when each condition represents a broad category. For example, Jeschonek, Marinovic, Hoehl, Elsner, and Pauen (2010) used photographs of 59 different animals and 59 different furniture items to represent the animal and furniture category, respectively. There are also examples of having more than one stimuli, representing a condition when the categories are not as generic as furniture or animals. The stimuli in Reid et al. (2009) included a congruent and an incongruent condition of an eating action. Each condition was presented with two different stimuli that included a different actor and a different food element. Increasing the variety of stimuli has been shown to increase the infant attention span (Stets, Burt, & Reid, 2013). However, when different stimuli depict the same condition, it is important that the changes in the stimulus will not modify the brain response. Third, infants engage more with complex and dynamic stimuli than with static stimuli (Courage, Reynolds, & Richards, 2006; Shaddy & Colombo, 2004). Dynamic elements are not generally used as stimuli in ERP studies due to a potential increase in eye movement artifacts and the difficulty of time-locking a dynamic stimulus (a few exceptions are biological motion studies such as Reid, Hoehl, Landt, & Striano, 2008; Reid, Hoehl, & Striano, 2006). However, what is a common practice in infant ERP methodology is to use short dynamic elements with or without sound

between ERP trials, known as "attention grabbers", to keep or regain attention when an infant disengages from the stimulus presentation. Last, another common practice is to pause the stimulus presentation when the infant becomes inattentive or starts to show symptoms of distress, which are usually described as "fussiness". Having breaks can help the infant get back to a more relaxed state and restore the interest of the infant in the presentation of more ERP trials after the pause. These are the principal tools available for experimenters to try to overcome the primary challenge present in a visual ERP experiment: keeping the infant's attention to the stimulus presentation for as long as possible and therefore increasing the likelihood of obtaining valid ERP trials.

There are no guidelines accepted as standard in the infant ERP field about the most suitable strategy of presenting stimuli in order to keep an infant's attention during an experiment. The sequence of trials, attention grabbers and pauses that are used in order to maximize the infant's attention to the screen are typically employed at the experimenter's discretion (Hoehl & Wahl, 2012). A common approach is to show the ERP trials in a continuous mode and present attention grabbers or make pauses when the experimenter considers that the infant is disengaging with the stimuli and needs to recapture the attention (e.g., Bakker et al., 2015; van Hoogmoed, van den Brink, & Janzen, 2013). Another approach is to present a trial when the experimenter considers that the infant is looking at the monitor (e.g., H. Halit et al., 2003; Richards, 2003). For example, H. Halit et al. (2003) showed a sequence of trials where each trial consisted of a central fixation point followed by an upright or inverted monkey or human face. The fixation point remained on the screen until the experimenter saw the infant looking at the fixation point via a video camera. Attention grabbers were also presented anytime the experimenter

considered it necessary in between trials. The last common approach found in the literature is to show an attention grabber or animation by default prior to each trial to attract the infant's attention towards the part of the screen where the trial will be presented (e.g., Csibra, Tucker, Volein, & Johnson, 2000; Parise & Csibra, 2013) or every certain number of trials, usually defined as block (e.g., Kouider et al., 2013).

New methods for keeping an infant's attention to the ERP presentation to get as many valid trials as possible have recently been proposed. In her research, Stets et al. (2013) investigated a new presentation procedure based on the knowledge that infants engage more with complex stimuli. The proposed testing procedure essentially consisted in increasing the variety of stimuli shown to the infant during the presentation. The authors used stimuli from three previous visual ERP studies and presented all stimuli randomly intermixed during the same ERP sequence. Their findings confirmed that this stimulus-variability sustained infants' attention longer, enabling the acquisition of more valid trials when contrasted with studies that only present two conditions. They also partly replicated the results from the original studies. The main restriction of the new method proposed was, as the authors pointed out in their study, that the stimuli presented from each different experiment had to be conceptually and perceptually different from each other in order to avoid the risk of interference of information processing between the conditions. This contamination risk is not only specific to the proposed methodology by Stets et al. (2013), it is also important to take it into account when presenting other type of engaging elements during an ERP experiment, such as attention grabbers between trials (Hoehl & Wahl, 2012).

1.2.2 Data reduction methodology

Data reduction is the first pre-processing phase in infant ERP data and comprises the selection of ERP trials that will be included in the final ERP average. This data editing process detects and rejects trials that contain artifacts. Causes to reject a trial are, amongst others: movement artifacts, muscle artifacts, eye movement artifacts or lack of attention to the stimuli (Hoehl & Wahl, 2012). This process can be split into two steps, which can be applied together or consecutively, depending on the methodology used to select ERP trials:

The first step is intended to select the trials where the infant is visually fixating on the stimuli. As explained earlier, this is the primary condition to generate a cognitive response to a visual stimulus and therefore this step is critical in order to maximize as much as possible the quality of the ERP signal, i.e., Signal-to-Noise Ratio (SNR). The most used approach within the infant literature is to record with a video camera the infant's behaviour during the experiment. Following data acquisition, the researcher visually inspects the video recording trial by trial to select the ERP trials where the infant was looking at the stimulus (Hoehl & Wahl, 2012; e.g., Altvater-Mackensen, Jessen, & Grossmann, 2016; Guy, Roberts, Tonnsen, & Richards, 2017; Kushnerenko, Teinonen, Volein, & Csibra, 2008; Striano, Kopp, Grossmann, & Reid, 2006). There are a few challenges associated with this method, both technical—synchronization of the video recording with the stimulus presentation onset times—and human related—researcher's time spent and the ability to discern visual attention to the stimulus in some cases—. Nevertheless, this is still regarded as the most reliable method to select visually attended ERP trials in an infant population. Another approach that is used less often involves the inclusion of

trials based on the inspection of infant's behaviour during the experiment (e.g., De Haan & Nelson, 1999; Webb et al., 2005). This approach avoids the technical and time challenges described with the video recording but increases the risk of including trials where the infant is not looking at the stimulus during the trial, especially because of the impossibility of reviewing the decision after the experiment.

The second step is the detection of EEG channels that contain unacceptable levels of noise due to movements or other artifacts that distort the EEG signal. There is extensive literature about how to complete this detection in adults utilising automatic algorithms (e.g., Delorme, Sejnowski, & Makeig, 2007; C. A. Joyce, Gorodnitsky, & Kutas, 2004; Luck, 2005) as well as software analysis tools that ease the use of them (e.g., Delorme & Makeig, 2004; Mognon, Jovicich, Bruzzone, & Buiatti, 2011; Nolan, Whelan, & Reilly, 2010). However, the automatic algorithms for rejecting artifacts are generally created for adult EEG signals. They are not always appropriate for infant EEG amplitudes and artifacts because they are usually too strict for infant data (Hoehl & Wahl, 2012). Therefore, the common approach in infant ERP literature is to conduct a manual trial-by-trial editing process to detect the trials and channels where the EEG signal is contaminated with artifacts. This process is performed based on visual inspection of the EEG waveforms, or partially assisted by running an automatic algorithm before the inspection to facilitate the detection of contaminated trials (e.g., Jeschonek et al., 2010; Leppänen et al., 2007; Reid, Csibra, Belsky, & Johnson, 2007; Righi, Westerlund, Congdon, Troller-Renfree, & Nelson, 2014). There are a few cases in the infant ERP literature where the selection of trials has been done fully automatically. For example, Kouider et al. (2013) utilized a generic automatic algorithm designed for adult EEG data based on amplitude rules. The amplitude thresholds were adapted (i.e. increased) to cope

with the generally larger amplitudes of signal and noise of the infant EEG. The authors reported in their study that the automatic algorithm successfully eliminated trials with noise. This statement was based on the late slow wave pattern observed on the grand average ERP across the scalp. According to the authors, this late slow wave pattern in the final ERP is not compatible with the presence of eye movements or with motion artifacts in the included trials.

The number of artifact-free trials per condition needed for ERP analysis will depend on the SNR of the component under study. For adult participants this goes from a minimum of 40 trials with a typical number of around 75-100 trials (Picton et al., 2000). Luck (2005) also states in his introductory book to ERP that *"the quality of the resulting ERP, known as signal-to-noise ratio (SNR), will increase with the number of trials included in the average. Only clean trials without artifacts should be included since artifacts extremely distort the EEG signal and will also distort the ERP"*. In infants, the number of artifact-free trials that are left after the data editing process is usually much lower than in adults. In fact, one of the main reasons for exclusion of a participant from the final sample in an infant ERP study is not reaching the minimum number of artifact-free trials. The attrition rate in infant ERP studies is typically between 50%-75% of the tested infants (for a review, see Stets, Stahl, & Reid, 2012). This is due to a combination of the fewer number of trials seen during the experiment and the larger amount of trials discarded during the editing process. A minimum of 10 artifact-free trials per condition is usually considered sufficient in an infant visual ERP study (e.g., Bakker et al., 2015; DeBoer et al., 2007; Reid et al., 2007; Reynolds & Guy, 2012). A factor that contributes to this small number of trials per infant and condition is the amplitude of infants' electrical brain activity measured on the scalp. It is usually

higher than in adults due to reduced impedances, such as a thinner skull, in addition to larger postsynaptic activity as a result of a larger number of synapses (DeBoer et al., 2007; Thierry, 2005). Another aspect to take into account is the within-subject variability in infant ERP data. Snyder, Webb, and Nelson (2002) found that infants' brain response changes during the time-course of an experiment. The Nc and slow wave components of the ERP significantly changed their morphology when including trials from either the first or second half of the experiment. One hypothesis is that these changes are associated with familiarity to the stimulus. Stets and Reid (2011) also found that including different amount of trials in the final ERP can significantly modify the Nc morphology and even change the differences in amplitude between conditions. They attributed the changes in their experiment to changes in the allocated attention to stimuli during the experiment. The authors also studied the possibility of reducing the criterion of minimum number of trials per condition as this facilitates the inclusion of a larger sample of infants in the grand average ERP. The inclusion criterion of a minimum of 3 to 5 artifact-free trials per condition has been recently reported in several studies (e.g., Kaduk et al., 2016; Kaduk, Elsner, & Reid, 2013; Missana, Rajhans, Atkinson, & Grossmann, 2014; Yrttiaho, Forssman, Kaatiala, & Leppänen, 2014). Reducing the number of minimum artifact-free trials per condition has the advantage that more infants can be included in the final ERP because they reach the criteria. This potentially decreases the SNR of the average ERP of each individual but increases the power of the effect by increasing the sample.

1.3 Eye tracking and EEG

Eye tracking is one of the methodologies most used amongst infant researchers. The most widespread method is the corneal reflection eye tracking technique, which estimates the pupil dimensions and gaze direction using near-infrared light and image processing algorithms (Holmqvist et al., 2011). Eye tracking based on corneal reflection is an unobtrusive technique and can be used to study infants' eye movements from a very early age. Eye tracking techniques with infants have been widely described in the literature (Aslin, 2012; Feng, 2011; Gredebäck, Johnson, & von Hofsten, 2010; Oakes, 2010, 2012). Eye tracking has been used, as can be expected, to study eye movements during the development of oculomotor control (e.g., Karatekin, 2007; Rosander & Von Hofsten, 2002; Shea & Aslin, 1990), although this is just a very small field in developmental research where eye tracking methodology has been applied. The ability to measure with timing and spatial precision where infants look at a visual stimulus can tell much about their perception and interpretation of the world at a specific moment of development. There are well established eye tracking paradigms that are currently used to study different aspects of development. Some examples are the occlusion paradigm to measure infants' understanding and expectations (Johnson, 2003), the face-distractor paradigm to measure the latency to disengage from a central target, such as a face (Ahtola et al., 2014) or the visual expectation paradigm to evaluate the speed of information processing and expectancy formation (Haith, Hazan, & Goodman, 1988). Eye tracking can also ease the analysis process and give more accurate results in general developmental paradigms, such as the relatively common preferential looking behavioural paradigm (e.g., Gredebäck et al., 2010; Chong, Richmond, Wong, Qiu, & Rifkin-Graboi, 2015; S. P. Johnson,

Slemmer, & Amso, 2004; Richmond & Nelson, 2009).

An additional use of eye tracking is to analyze the eye movement information during the experiment to update the stimulus presentation based on the participant's looking behaviour, known as gaze contingency (Holmqvist et al., 2011). In infancy research, some gaze-contingent paradigms have been used in recent research and are allowing the exploration of new methods to understand development, as well as disentangle more complex infant behaviour that is not possible to study only with eye tracking as a post-processing tool (Deligianni, Senju, Gergely, & Csibra, 2011; Marraffa & Sperati, 2012; Miyazaki, Takahashi, Rolf, Okada, & Omori, 2014; Shukla, Wen, White, & Aslin, 2011; Wang et al., 2012; Wass, Porayska-Pomsta, & Johnson, 2011). Wang et al. (2012) demonstrated in their novel paradigm that infants can learn that they could trigger a stimulus (considered as a reward) more rapidly than most of the tested adults, just by looking at a red circle presented on the monitor. The hypothesis of why infants were able to learn this gaze-triggering behaviour and adults generally were not, is that infants do not have a prebuilt idea of what is possible and what is not as robustly as adults have. This allows them to learn and understand new concepts more rapidly than adults. Also, Wass et al. (2011) and Forssman, Wass, and Leppänen (2014) have shown that infants are able to get better at controlling attention when they have been previously trained with gaze-contingent stimuli that reacted to their looking behaviour (i.e., a butterfly started flying whenever an infant looked at it). Finally, gaze contingency has also been used extensively, especially in adults, as a method to control the stimulus presentation (Holmqvist et al., 2011), this is, presenting the next trial only when the eye tracker estimates that the participant is looking at the right place.

The combination of eye tracking and EEG within one experiment has increased in recent years, especially in adult research, as both knowledge and technology have evolved. Simultaneous recordings of eye tracking and EEG are currently used to get new insights about cognition and human behaviour that are not possible without the combination of the two sources of information (e.g., Kulke, Atkinson, & Braddick, 2016a) and also for methodological purposes (e.g., Ahtola, Stjerna, Stevenson, & Vanhatalo, 2017). One of the fastest recent methodological developments has occurred in the area of eye-fixation-related potentials (EFRPs). EFRPs overcome the challenge of avoiding eye movements during ERP studies and thus, allow the measurement of brain responses with free-viewing stimuli. Eye tracking data are used to calculate fixation locations and onset times. Each fixation onset becomes the onset of an EFRP trial, measuring the brain response during the fixation time (i.e., eyes are still). This allows the acquisition of EFRP epochs that are not contaminated by eye movements while allowing the participant to perform more natural behaviour during the experiment. This methodology first started in reading ERP research (Dimigen, Sommer, Hohlfeld, Jacobs, & Kliegl, 2011) and, in the last years, other areas of research such as visual search (Kamienkowski, Ison, Quiroga, & Sigman, 2012; Kaunitz et al., 2014; Weaver, Clayton Hickey, & van Zoest, 2017) or affective processing of natural scenes (Simola, Le Fevre, Torniainen, & Baccino, 2015) have also used this methodology successfully. According to the knowledge of the author, EFRP has not been applied to infants to date. This is most likely due to the complexity of the methodology and the requirement of obtaining high-quality eye tracking data in order to accurately compute the fixation onset times.

Another potential application of the simultaneous recording of eye tracking and EEG is the use of eye tracking data as an artifact rejection tool to detect EEG data

contaminated with eye movement artifacts. Eye movements and blinks contaminate frontal EEG electrodes and it is crucial in ERP studies to detect them to either reject the EEG data where the artifact occurs or correct them to eliminate the noise from the EEG activity (Luck, 2005). There are algorithms that are used to detect blinks and eye movements from the frontal electrodes of an EEG recording or the electrooculogram (EOG) electrodes. These can be based on basic rules that detect sudden changes in EEG amplitudes, characteristic of these type of artifacts, or in more advanced algorithms based on source separation such as ICA or PCA (e.g., Jung et al., 2000). Nevertheless, the use of eye tracking to detect eye movements has the advantage of the high spatial accuracy achieved with eye tracking, which can potentially help detect not only eye movements precisely, but also the specific location of the gaze at a certain point in time (e.g., fixation location outside the fixation cross shown prior the stimulus). Only two studies have been found in the literature that investigate the use of eye movement data from an eye tracker to improve the eye movement artifact detection in combination with ICA algorithms (Plöchl, Ossandón, & König, 2012) and to detect valid EEG epochs based on the gaze behaviour (Ahtola et al., 2017).

In developmental research, the combination of eye tracking and EEG is still less common than in adult research. This is in a large extent due to the methodological and technical difficulties of recording simultaneously two sources of data such as EEG and eye tracking with infants. In some cases, these difficulties can be overcome by recording one method after the other with two experiments during the same session. Righi et al. (2014) investigated how 7-month-old infants processed female and male faces. The authors designed two experimental paradigms: an ERP where they found differences in the N290 component, demonstrating that infants at that age

differentiated gender in human faces. They also conducted an eye tracking study, which further demonstrated this fact. Interpreting both eye tracking and ERP results together, allowed the authors to conclude that the differential processing for female faces may be a signature of expert-level processing. There are also a few studies where eye tracking and EEG have been successfully combined simultaneously (Agyei, Holth, van der Weel, & van der Meer, 2015; Kulke, Atkinson, & Braddick, 2016b; Yrttiaho et al., 2014). Kulke et al. (2016b) recorded eye tracking and EEG simultaneously with infants from 1 to 8 months of age. The authors applied the fixation shift paradigm and were able to relate the evolution of visual attention shifts with the evolution of neural responses during the first months of postnatal development.

The current disadvantages when combining these techniques with infants is still a limiting factor of their use. Southgate and Vernetti (2014) reported in their study that the authors could not record simultaneously eye tracking and EEG data with infants (as they did with adults) because the time spent in setting up both methods together negatively affected the number of valid trials that they were able to acquire due to the limited attention span of infants. Kulke et al. (2016b) also reported in their study that the combination of techniques led to the exclusion of more trials, as some trials were excluded because they did not meet either the eye tracking data quality criteria or the EEG data quality criteria. Nevertheless, the advantages of combining eye tracking and EEG can be potentially significant. The most relevant advantage is the use of eye movements and brain responses to get more accurate information about development mechanisms in infants. This is particularly the case in instances where the use of only one technique cannot provide enough evidence for a strong scientific conclusion. There are also methodological advantages, such

as the possibility of using the eye tracker to automate the beginning of trials (i.e., stimulus control with gaze contingency), or using the eye tracking data as an EEG artifact detection and rejection method.

1.4 Main limitations of current infant ERP techniques

Being able to get insights about cognitive development with a non-invasive and well-established technique is without doubt the main advantage and reason why ERP has become a popular method in developmental research. However, it is clear that the ERP methodology that is currently used with infants has a number of challenges that increase the difficulty of obtaining good quality data and, therefore, can jeopardize the conclusions of the research questions under study. The challenges are, to a major extent, a result of the infants' attention span and behaviour during the experiment, which have a direct impact on the EEG data quality and thus on the number of artifact-free ERP trials that can be included in the average ERP. One of the main consequences derived from these challenges is the attrition rate that is usually reported in infant ERP studies, which varies between 25% - 75% across studies (DeBoer et al., 2007; Hoehl & Striano, 2008; Stets et al., 2012). An extreme example is the case of De Haan, Pascalis, and Johnson (2002). The authors tested 175 6-month-old infants and only included 34 of those in the final ERP analysis. The high attrition rate of 80% was attributed to the strict criteria followed to select artifact-free trials. The most obvious consequence of having high attrition rates is the amount of time spent by the experimenter testing infants and analyzing data that will not contribute to the final ERP results. There are, however, other more important questions to address that are related to excluding infants from the final sample that could even have an impact on the results. For example: are the most temperamentally "patient" infants likely to behave positively during the ERP experiment and therefore more likely to be the ones that will be included in

the final ERP sample? Could infants who are developmentally more advanced than the majority of their cohort be more likely to contribute enough trials for an ERP? Either of these possibilities could potentially create a bias in the results, which could, in some cases, not represent the age group under study.

It is common in current ERP methodology that participants are discarded from the final ERP analysis if they have not contributed to all the conditions with at least a minimum number of trials (Luck, 2005). This criterion becomes especially significant in infant ERP where a relatively reduced amount of artifact-free trials are generally acquired, consequently increasing the attrition rate. The attrition rate increases even more when using common infant paradigms where conditions are presented at different frequencies, such as the odd-ball paradigm, with one condition presented significantly fewer times than the other e.g., 20%/80% ratio: Karrer, Karrer, Bloom, Chaney, and Davis (1998). A high attrition rate means increasing the tested sample in order to have statistical results with enough power to answer the research question. This is a clear limitation since testing more infants means spending more time in looking for infant candidates, in recording sessions and in analyzing the ERP dataset to determine whether the infant can be included in the final analysis sample of the study. Besides these practical consequences, excluding a high number of infants from a study has also the potential risk of obtaining results not representative of the population under examination. Overall, decreasing the attrition rate in infant ERP studies is an important question that needs to be addressed.

Stets et al. (2012) conducted a meta-analysis to investigate the possible connections between study-related features, such as the nature of the stimuli (visual or

auditory), the duration of the session or the number of conditions, with the attrition rate reported by the corresponding infant ERP study. The variable that had the greatest influence in the attrition rate was the nature of the stimuli, with visual stimuli having higher attrition rates than auditory stimuli or combinations of auditory and visual stimuli. Remarkably, neither the number of conditions presented nor the mean number of artifact-free trials per condition included in the analysis were found to have impact on the attrition rates reported. As the authors conclude, there might be more factors—difficult to measure—that were not taken into account in the meta-analysis, such as the amount of experience of the experimenter or the social factors related with the end of maternity leave (i.e., an infant's behaviour could be altered after their mother's maternity leave period has ended and the infant has started daycare). Indeed, a naïve interpretation of the variables that can influence attrition rate in infant ERP experiment would be to hypothesize that with fewer conditions, the number of trials presented per condition will increase with a consequent reduced risk of having a high attrition rate. However, the research to date on this matter suggests that there are many variables that can influence attrition rate and it is the combination of different variables that will produce a specific attrition rate. For example, the authors of the meta-analysis, also examined the potential benefits of including more conditions in the same ERP session belonging to different ERP paradigms (Stets et al., 2013). The hypothesis here was that the presentation of more variation of stimuli would increase the attention span of infants. Indeed, the increment of conditions in this experiment resulted in a decrease in the attrition rate. In this study, the variety of stimuli presented was the most probable cause of decrease in the attrition rate, mainly due to the attention span increase and therefore the number of trials presented to infants. This contradicts on one hand the

simplified assumptions that fewer conditions will increase the likelihood of collecting more artifact-free trials but, on the other hand, demonstrates how complex it is to assess the partial contribution of each factor to the attrition rate in infant ERP experiments.

Attention span of infants during ERP experiments is another limitation of the current infant ERP techniques. Keeping an infant's attention during a experiment is a general challenge in infant research, but in visual ERP paradigms it is a factor that can seriously limit the amount of artifact-free trials that are obtained, with possible implications for the data quality. Usually, the experimenter is able to recognize that the infant's attention span has ended or is near the end because the infant shows one or more of the following behaviours: loss of interest in the stimuli starting to move, to yawn or to sob, amongst other displays of distress. This state of the infant, which creates big artifacts in the EEG signal, is usually defined as "fussiness" and it is one of the main causes of attrition rate reported in behavioural visual experiments with infants (Slaughter & Suddendorf, 2007). When an infant becomes "fussy" during the stimulus presentation, the experimenter can take a break and try to continue with the experiment a few minutes later, show attention grabbers to try to regain infants' visual attention to the screen or to stop the experiment. It is the experimenter who usually decides how to proceed when fussiness is detected. This is likely to be a subjective and fluctuating decision that can, therefore, introduce variability in the amount and quality of acquired infant ERP data.

Apart from these methodological difficulties related to the experimental design and the recording session, there are also other factors that are related with the data analysis. The main limitation is the lack of robust methods during the editing pro-

cess of the infant ERP data. First, the selection of trials where the infant is visually fixating on the stimulus is mainly based on human judgements. When it relies on technology such as the video recording of infant's behaviour, it might be sometimes unclear whether the infant was looking at the important part of the stimulus due to the resolution and quality of the video recording. This can potentially decrease the SNR of the average ERP if trials with no (or weak) ERP responses are included. Second, the artifact detection usually has the same associated problem. To the knowledge of the author, there are no robust automatic algorithms that can cope with the high within- and inter-variability of infant EEG data. Most of the time, it is the researcher who takes the final decision of accepting or rejecting a trial in terms of EEG noise. This creates a potential source of variability due to "the human factor," whereby there is variability in terms of judgments made across editors regarding the quality of obtained EEG data. Hoehl and Wahl (2012) recommend in their methodological guidelines to reduce this variability in artifact rejection by training new editors with an already analyzed data set to compare results and learn from experienced editors. It is uncertain how much this practice is followed and also to what extent it avoids the potential human bias.

The last limitation that is worth highlighting is the potential variability in the final infant ERP morphology based on the amount of trials included and their characteristics. Based on adult ERP methodology theory (Luck, 2005; Picton et al., 2000), the more trials included in the average ERP, the better the SNR. This statement is considered to be true as long as the ERP trials included are artifact-free so that there is an increase in the SNR and therefore in the quality of the ERP signal. However, recent studies suggest that infant ERP response quality could be also influenced by characteristics related with the temporal dynamics of the course

of the experiment session. Stets et al. (2012) investigated in two of her studies the implications of including different amounts of trials to the average ERP. In both studies, the authors demonstrated that including different number of trials in the average ERP can modulate amplitude attention-related ERP components such as the Nc. The authors concluded that these changes were explained by changes in attention allocation during the experiment. This means that the SNR of a single trial might not only be influenced by the amount of EEG signal noise, but also by the amount of ERP response due to the attention allocation to the stimuli and other factors.

1.5 Thesis objectives

Can eye tracking help improve infant ERP techniques and ERP data quality?

Are current infant ERP editing methods in need of further standardization?

The methodological limitations explained in the previous section highlight the most important challenges presented by current infant ERP practices. The number and diversity of these challenges, as well as their potential impact on the final ERP results demonstrate the need for improving them. To date, there have been few attempts to propose new strategies that improve infant ERP techniques. This introduction has outlined different strategies to record and analyze ERP data, but there has been no systematic investigation of whether these strategies bring any improvement or if they are actually having an impact on the final results. Thus, there is a need to further study the validity of the current ERP methods as well as exploring other approaches that could potentially improve current practice.

Researching and working towards an improvement in the infant ERP methodologies would bring several crucial advantages to the field:

First and most important, an improvement in the ERP data quality. A methodology that would ensure that the ERP data have the appropriate quality to be analyzed and to extract conclusions with confidence is vital in these types of studies. This goal could be approached from different perspectives. One strategy would

be to improve the data collection methodology so that the infants are more attentive to the stimulus. To achieve larger attention span from an infant, either the stimulus presentation could be more salient and attractive for infants or procedural improvements could be employed. For example, the experimenter could have strategies to bring back the infant's attention to the stimulus presentation when the first symptoms of fussiness are detected. Another strategy would be to work toward a more optimal selection of the ERP trials that will be included in the final sample and that will contribute positively to the different ERP components. An important aspect that should be common to any of the strategies proposed is the concept of standardization of practice. Any current or future ERP methodology should be standardized to the point where it can be applied through different labs and experimenters and derive similar results.

Second, it would be beneficial to develop strategies to decrease the high attrition rate registered in visual ERP paradigms with infants. This could be correlated with the first advantage because it is quite likely that an improvement in the data quality will be associated with a decrease in the attrition rate, i.e., better quality of the ERP data potentially implies more trials accepted and that probably means more infants included in the final sample. In general, these two advantages are related with obtaining more certainty and more robust results of an infant ERP experiment.

The third advantage of working towards improvements in infant ERP techniques is the time invested by the researcher. This is not a limitation that will impact on the quality of the results a priori but it is worth highlighting because it is probably the most practical aspect that ERP researchers will detect as a limitation in this technique. In fact, the amount of time spent particularly in data analysis could impact

on the final results due to the high content of manual and repetitive work needed when applying the main current techniques for analyzing infant ERPs, namely the visual trial-by-trial inspection of the EEG data to detect artifacts.

The first goal of this thesis is to explore whether the use of eye tracking techniques during the ERP data collection could improve current ERP data collection methods and ERP data quality. This direction of research is motivated by the potential benefit of having precise information of the infant's gaze behaviour (i.e., where the infant is looking) throughout the stimulus presentation. Knowing the specific location of the infant's gaze on the stimulus in real time during the experiment opens new possibilities of data collection techniques that have not yet been studied in the context of infant ERP data collection. The closest exploration found to date in the literature reports a successful use of gaze-contingent techniques when applied in a visual ERP experiment to 7- and 8-year old children (Maguire, Magnon, & Fitzhugh, 2014). In their study, the authors where able to include more valid ERP trials in the final sample when using gaze-contingent techniques during the ERP experiment as compared to when they used a standard ERP presentation of the stimulus. Although the context of this study was slightly different as it was applied to a different developmental age with different behavioural challenges in the collection of EEG data, it helps support the hypothesis that eye tracking techniques can improve infant ERP data collection techniques and ERP data quality.

Having precise spatial and temporal information of the infant's gaze can be associated with a number of advantages that could help ERP data quality. First, the gaze information can be used in real time to present the ERP trials only when the infant is looking at the intended place (usually at a fixation cross or at an image

prior to the ERP trial). Second, the gaze data can also be used after the experiment as an objective and precise measure of the ERP trials that are valid because the infant was looking at the stimulus and did not perform any large eye movement. This thesis focuses on the first advantage and explores the possibility of using gaze-contingent techniques during the stimulus presentation to ensure that the infant is looking at the right place before presenting an ERP stimulus (Chapter 3). The hypothesis that motivated this work is that controlling the stimulus presentation sequence based on infant's looking behaviour would improve the infant ERP data quality. This could be caused by both an increase in the infants' attention span and an increase in the number of trials that the infant looks at the right place, with the consequence of an increase in the number of valid trials included in the sample. Another aim for the use of gaze-contingent techniques during an infant ERP study is to explore a new automatic and objective methodology for the ERP data collection. Utilising eye tracking has the advantage of avoiding human biases as much as possible, which may influence the data quality. This approach may well help to increase the standardization of the methodology. This thesis will discuss both the advantages mentioned and the current limitations found when using eye tracking techniques during an infant ERP experiment.

An important part of the use of gaze-contingent techniques is to define what "looking at the right place before presenting a stimulus" means when recording visual ERPs. It is unclear how visual acuity affects cognitive processes and thus ERP responses when presenting visual stimuli. Visual ERP components can be modulated by attention (Luck, Woodman, & Vogel, 2000; Mangun & Hillyard, 1988), but it is unknown how visual acuity affects late ERP components, such as the widely studied infant Nc or N400 components. In infant ERP experiments with visual

stimuli, the stimulus is usually presented at the center of the screen and it is usually limited in size in order to avoid as much as possible the infants' eye movements. Also, the general practice is to present before each trial a fixation cross or image that is located just in the middle of the stimulus. The intention of this practice is that the infant is looking at the center of the stimulus at the trial onset. It is unclear how this practice will affect the ERP response when the visual information needed to process the stimulus is not within the highest visual acuity area of the infant. In adults, visual acuity has its highest value at the fovea, which is around 2 degrees in size, and decreases abruptly in parafoveal and peripheral areas of the visual field (Rayner & Inhoff, 1981). In typically developing infants, visual acuity increases quickly in the first six months of life and then stabilizes slowly until it reaches adult values (Dobson & Teller, 1978; Salomão & Ventura, 1995). The ideal situation would be that the infant is looking at the center of the monitor when the stimulus is presented and the visual acuity allows to perceive and process all the visual information without eye movements. It is unknown how the ERP response is affected when the visual information is outside the high visual acuity limits. In Chapter 2, this is investigated by analyzing the influence of the fixation distance to the stimulus in the N400 ERP component. One requirement of the methodology used in this experiment was that participants followed correctly the instructions and looked at the correct place in each trial (moving fixation cross along the display). This important requirement for obtaining successful results required that the tested sample were adults. Based on the visual acuity development in infants (Dobson & Teller, 1978), the assumption that is made in this thesis is that the results obtained with adults are comparable with the ones that would have been obtained with infants. The second assumption that is made is that the rest of advanced ERP

components will be influenced to a similar amount as the results obtained with N400 component. These considerations and the results of Chapter 2 were taken into account when defining the gaze-contingency theoretical framework of the study carried out in Chapter 3.

The second goal of this thesis is to investigate to what extent the current infant ERP editing methods are reliable and to assess the need for their further standardization. As already explained in this introduction chapter, the trial selection process of infant ERP is a real challenge due to the infant's behaviour and the consequent EEG signal complexity with usually high amount of EEG artifacts. The infant ERP studies related to methodology found in the literature have been mainly centered around the important topic of the minimum number of trials needed to obtain a valid ERP waveform and the evolution of the ERP waveform with the number of trials included (e.g., Stets et al., 2012). However, it is still unknown whether the current methods used during the infant ERP editing process can alter the final ERP waveform and therefore the results obtained. Since the methods are mainly the use of trial-by-trial visual inspection and selection by the human experimenter, there is a level of subjective criteria used that could potentially influence the final number of trials that are accepted in the final sample. Another important aim that may help move forward the current infant ERP method is to understand what EEG signal characteristics human editors take into account when editing data using the current methodology. Having more information on this matter can benefit future research toward the understanding on the complexity of infant EEG and the detection of artifacts in the EEG signal.

Chapter 4 of this thesis focuses on these important questions. The first part of

the study evaluates the reliability of the current infant ERP editing methods. It investigates the agreement between three human editors and one automatic algorithm that analyzed the same infant ERP dataset. The scope of this study is limited to the trial editing on the level of EEG noise. The editing process usually carried out by the researchers to select the trials where the infant is visually fixating the stimulus were not included in this study to reduce the sources of variability. The infant attention to the stimulus editing was therefore coded beforehand and was common to all editors. The selection of trials of each of the editors as well as the final ERP waveform obtained with each of the three editing processes was analyzed to understand whether the editing processes led to any significant differences in the final ERP waveforms that could affect the final ERP results. The second part of the study in Chapter 4 focuses on the exploration of the variables that are taken into account by human editors when selecting ERP trials based on the level of noise in the EEG signal. Advanced statistical methods are used to model the EEG characteristics that each of the editors used when editing the ERP data. Overall, Chapter 4 seeks to understand the validity of the current editing methods and evaluate the need of further work toward more standardized infant ERP editing methods.

In summary, the objective of the thesis is to contribute to the improvement of infant ERP data quality and of the methods that are currently used. For that, the approach taken is to explore a new method for collecting ERP data based on eye tracking that can potentially improve the quality and quantity of ERP trials (Chapters 2 and 3), and to analyze the current editing methods techniques to increase the knowledge about their validity and limitations (Chapter 4). The work presented in this thesis is also intended to help future infant ERP methodological research focused on a better understanding of the infant EEG signal and disentangling its

complexity in order to improve the current infant ERP techniques.

Chapter 2

The Fixation Distance to the Stimulus Influences ERP Quality: An EEG and Eye Tracking N400 Study

Text as it appears in Domínguez-Martínez, E., Parise, E., Strandvall, T., & Reid, V. M. (2015). The Fixation Distance to the Stimulus Influences ERP Quality: An EEG and Eye Tracking N400 Study. PloS One, 10(7).

Abstract

In a typical visual Event Related Potential (ERP) study, the stimulus is presented centrally on the screen. Normally an ERP response will be measured provided that the participant directs their gaze towards the stimulus. The aim of this study was to assess how the N400 component of an ERP was affected when the stimulus was presented in the foveal, parafoveal or peripheral vision of the participant's visual field. Utilizing stimuli that have previously produced an N400 response to action incongruities, the same stimuli sequences were presented at 0°, 4°, 8° and 12° of visual angle from a fixation location. In addition to the EEG data, eye tracking

data were recorded to act as a fixation control method and to allow for eye artifact detection. The results show a significant N400 effect in the right parieto-temporal electrodes within the 0° visual angle condition. For the other conditions, the N400 effect was reduced (4°) or not present (8° and 12°). Our results suggest that the disappearance of the N400 effect with eccentricity is due to the fixation distance to the stimulus. However, variables like attentional allocation could have also had an impact on the results. This study highlights the importance of presenting a stimulus within the foveal vision of the participant in order to maximize ERP effects related to higher order cognitive processes.

2.1 Introduction

When visual gaze remains relatively still, known as a fixation, the spatial resolution of the human visual field changes as a function of the distance from the center of the fixation point (Duchowski, 2007). Researchers often divide the visual field into three areas: foveal, parafoveal and peripheral. Foveal vision is the area with the highest visual acuity, extending approximately 2° around the fixation point. In the parafoveal area the visual acuity decreases. It extends between 2° and 5° around the fixation point. The peripheral vision extends from 5° around the fixation point until the edge of the field of view. In this area, visual acuity decreases abruptly (Hirsch & Curcio, 1989; Rayner & Inhoff, 1981). We usually move our eyes to place the fovea on the part of any object or location that we want to see clearly (Rayner, 1998). Parafoveal and peripheral vision also play a role when planning eye movements and extracting visual information (for a review see Strasburger, Rentschler, & Jüttner, 2011). For example, it has been shown to be useful for gist recognition in natural scenes (e.g., Antes, Penland, & Metzger, 1981; Fei-Fei, Iyer, Koch, & Perona, 2007; Sanocki & Epstein, 1997) and when examining familiar objects (Biederman & Cooper, 1992).

Usually during a fixation the attention is directed to the fovea. It is also possible to fixate on one location in space, yet to attend to other spatial positions in parafoveal or peripheral visual areas. These are known as overt and covert attention, respectively (Posner, 1980). It is known that attention modulates the ERP waveform (Luck et al., 2000). Thus, the use of overt or covert attention will influence the ERP when a visual stimulus is presented to peripheral vision. In a visual experiment, covert attention usually occurs when the participant has been previously

instructed to do so or has a cue that indicates that this should be done. In ERP studies of selective spatial attention, attended and unattended stimuli presented at different eccentricities have been found to differ in the magnitude of the early ERP components P1, posterior N1 and anterior N1, with all three having a larger amplitude for attended stimuli when compared with those that were non-attended (Clark & Hillyard, 1996; Eason, 1981; Mangun & Hillyard, 1988).

The fixation distance to the stimulus, together with attention, alters the early components of an ERP. Little is known, however, about how advanced cognitive processes are altered when the stimulus is presented outside the foveal area. Advanced cognitive processes are related to the middle and late epoch periods of an ERP after stimulus onset (Rugg & Coles, 1995). One of these components is the N400, a widely investigated ERP component sensitive to semantic integration/violation in language research (for a review see Kutas & Federmeier, 2011) and also exploited in action-related studies (for a review see Amoruso et al., 2013). This component can be observed as a negative deflection in the ERP waveform at frontal, parietal and temporal areas of the scalp—depending on the nature of the stimuli and task—at around 400 ms from the onset of the stimulus.

The main aim of the present study was to investigate whether the fixation location alters the cognitive processes that are generated by a visual stimulus. This was studied by analyzing the N400 response when participants were fixating at different distances from the stimulus location. We hypothesized that there would be a spatial distance from the stimuli where the N400 ERP component would start to become affected. This would cause a decrease in the amplitude of the N400, which could in turn affect final conclusions. These changes in the N400 component are expected

based on the decrease of visual acuity with distance. To avoid changes in the ERP waveform due to differences in overt and covert attention, participants were asked to always attend to the stimulus. They were asked to maintain their fixation on a fixation cross that moved to different locations on the screen. They were also asked to attend to the stimulus without moving their eyes irrespective of where it appeared on the screen.

This study utilized stimuli that were validated in an earlier experiment conducted by Reid et al. (2009). In that experiment, simple sequences of actions were presented with two possible outcomes, one anticipated and one not anticipated. Results showed an N400 component in the unanticipated condition, whereas no N400 was observed in the anticipated condition. We aimed to replicate these results when presenting foveal stimuli. We also included three more conditions where the same stimuli were presented at 4°, 8° and 12° of visual angle from the participant's fixation location. We calculated the averaged ERPs and compared the N400 component obtained in each of them with the foveal condition.

Gaze-contingent stimulus presentation was used during ERP recordings to ensure that participants were directing their gaze to the right location on the screen at the beginning of each trial. We also used the eye tracking data during the ERP analysis to discard trials that contained eye movements or blinks. As a secondary aim of the study, we assessed the use of eye tracking data as a means to reject ERP trials that contain eye artifacts. We compared these results with the more normative detection of eye artifacts via the EEG amplitude information of the frontal electrodes (Luck, 2005). We anticipated that the eye tracking data may be more sensitive to eye movement data than the algorithms related to eye artifact detection via EEG data.

2.2 Materials and Methods

2.2.1 Participants

Twenty-eight adult (13 females) aged from 18 to 49 years (m=24.7, sd=8.02) volunteered in the experiment. All participants were right handed and had normal or corrected-to-normal vision. All participants were given oral information about the experimental procedures and gave their written informed consent before participation. The study was approved by the Lancaster University Research Ethics Committee. For inclusion in the ERP grand average, participants were required to have at least 40 artifact-free trials per condition. A final data sample of eighteen participants was included in the analysis. The data of three participants had to be rejected because of technical problems during the experiment, one participant due to poor quality of eye tracking data and six participants because the minimum amount of artifact-free trials was not reached for one or more of the conditions.

2.2.2 Stimuli

The stimuli were the same set of photographs used in Experiment 1 of Reid et al. (2009), which involved adult participants. Specifically, these were two sets of photographs depicting a male actor eating with a spoon or holding food. Each set consisted of three photographs. The first photograph displayed the general context of the action. The second displayed the initiation of the action. Each set of photographs finished with either an anticipated conclusion of the action with the spoon or food directed to the mouth, or, an unanticipated conclusion to the action was

shown where the spoon proceeded to the forehead or food was positioned near the ear (Figure 2.1).

Figure 2.1: Action sequence stimuli used in the study. Each row corresponds to a different stimulus set. First column: context of action; Second column: initiation of action; Third column: anticipated conclusion of action; Fourth column: unanticipated conclusion of action. The stimuli are the same as those used by Reid et al. (2009) in the adult study.

2.2.3 Apparatus

During the experiment, both eye movements and EEG were recorded simultaneously.

Eye tracking: Eye movements were recorded at a sampling rate of 300Hz using a Tobii TX300 eye tracker (Tobii technology, Danderyd, Sweden). The eye tracking data were acquired using Tobii SDK 3.0 in Matlab® (Mathworks, Natick, MA, USA) on the presentation computer.

EEG: EEG data were acquired by a 128-channel Sensor Net (Electrical Geodesic, Eugene, USA) with Net Station® acquisition software running on a different computer. The EEG data were sampled at 250Hz. Electrode impedance was kept below 50KΩ. Raw EEG data were recorded with the vertex (Cz) as the online reference

and re-referenced offline to an average reference.

Stimulus presentation: The stimuli were displayed on a 20-inch CRT monitor with a refresh rate of 60Hz. Psychtoolbox-3 for Matlab and custom made scripts were used for stimuli presentation (Brainard, 1997; Kleiner et al., 2007).

2.2.4 Procedure and experimental paradigm

Participants were seated at a fixed distance of 60cm from the eye tracker in accordance with the eye tracker requirements and in order to keep a similar field of vision across participants. First, the EEG net was positioned following the manufacturer instructions. When the electrodes' impedance level was below the threshold, a five-point calibration routine of the eye tracker was performed followed by a validation of the calibration. The validation consisted of seven dots presented sequentially at different locations of the screen. Five of these locations corresponded with the potential locations of the fixation cross in the study. Before the experiment started, participants were instructed to fixate a fixation cross at the beginning of each trial and keep fixating on that location during the presentation of the stimuli. The fixation cross changed its location every trial and participants were asked to move their gaze to the fixation cross in order to start the next trial. The main purpose of changing the location of the fixation cross was to ensure that participants were engaged with the experimental design. Participants were also instructed to move as little as possible, to blink during the inter-stimulus interval and to pay attention to the stimuli without moving their eyes from the location of the fixation cross. The aim of the last instruction was to ensure that participants deliberately attended to

the stimulus irrespective of the eccentricity of the image to the fixation location.

Before each trial, a fixation cross of 0.4° of visual angle was presented on the screen. The position was selected randomly from five possible locations along the horizontal axis. The cross remained on the screen until the participant fixated it for at least 500 ms. When this requirement was met, the trial started after a random period of 200 – 400 ms. The three images were presented sequentially. The first two images were on the screen for 500 ms each and the third image was on the screen for 1000 ms. The stimuli were presented at a size of 10°x7.5° of visual angle. The experiment consisted of eight (2x4) different conditions: two levels corresponding to congruency (anticipated or unanticipated action) and four eccentricities corresponding to the distance of the image to the fixation cross (0°, 4°, 8° or 12° of visual angle). For the 0° condition, the mid-point of the main action in the anticipated and unanticipated third image was centered on the fixation cross. The main action for both congruent conditions was within the fovea of the participant when fixating the fixation cross. For the rest of the conditions, the stimuli were presented at the same vertical level, but the images depicting the action were horizontally shifted by the corresponding number of degrees of visual angle (Figure 2.2). The shifts were made randomly to the left or to the right for each trial and balanced by the end of the presentation. Each condition was presented 75 times, with a division of the two sets of stimuli of approximately half each. For each trial, the condition was pseudo-randomly chosen with the constraints that the same type of condition should not be presented three times consecutively.

The stimulus presentation was split into four segments of around 10 minutes each. Between each segment, a five-minute pause was given to the participants to

rest, re-calibrate the eye tracker and rehydrate the EEG net when necessary.

2.2.5 Data analysis

Eye tracking: Eye tracking data were analyzed using Matlab. First, a fixation filter was applied to detect fixations. The fixation filter was implemented based on the Tobii I-VT fixation filter (Olsen & Ricardo, 2012), with a fixed velocity threshold of 30°/s for all the recordings. Interpolation of missing data, the merging of adjacent fixations and the discarding of short fixations were completed following the standard values suggested in the Tobii I-VT fixation filter description (Olsen, 2012). Eye tracking data were segmented into epochs that comprised 1200 ms before the onset of the third image in the stimulus sequence and 1200 ms following onset. Fixations were corrected every epoch using the validation data. The correction values were calculated independently for each cross location. For each validation location, the X and Y mean value of the samples collected was subtracted from the location of the validation dot. For every epoch, the X and Y values of the fixations were corrected using the correction value that corresponded with the trial location.

Eye tracking data were used to detect and reject the trials that did not meet the ERP analysis requirements. Trials were marked for rejection when: an eye movement larger than 1° was made during the ERP epoch, the participant's point of gaze was not within the 1° around the fixation cross or a blink was detected. A trial was considered to contain a blink when there was a period of 75 ms to 350 ms of missing data preceded or followed by a saccade. Saccades before and after the blinks are considered eye tracking artifacts which are generated by the eyelid going

up or down during a blink (Holmqvist et al., 2011).

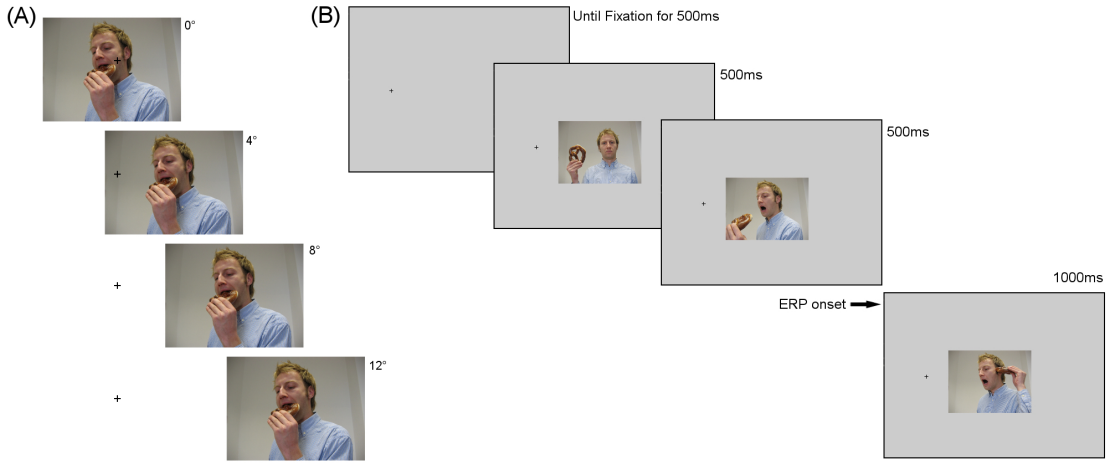


Figure 2.2: Location conditions and example of stimuli sequence. (A): The image appeared on the screen shifted horizontally depending on each specific location condition. (B): Example of a 8° condition trial. The stimuli sequence started by showing the cross on the corresponding location. Once the participant fixated it for 500 ms, plus a random time between 200 and 400 ms, the sequence of three images appeared on the screen. The cross remained on the screen during the trial. The ERP was time-locked to the presentation onset of the final image in the sequence.

EEG data pre-processing: EEG data were analyzed using the Matlab toolbox EEGLAB v13.0.1 (Delorme & Makeig, 2004). EEG data were first band-pass filtered between 0.3 and 30 Hz. The EEG recordings were segmented into epochs and baseline corrected. Each epoch comprised a 100 ms baseline before the onset of the third image (namely from the last 100 ms of the second image in the sequence) and 1000 ms of the third image, displaying the final state of the action. Trials marked as rejected during the eye tracking analysis were automatically discarded. For the remaining trials, electrodes were marked as bad when the absolute amplitude exceeded 100μV. An electrode was marked as bad for the entire recording when it had been marked as bad for more than 40% of the trials. A trial was marked as bad and rejected when more than 10% of the electrodes were marked as bad

during the trial. The remaining electrodes that were marked as bad for individual trials were interpolated. Electrodes marked as bad for the entire recording were completely interpolated. After the rejection of trials and interpolation of electrodes, the EEG data were re-referenced to the average reference and segmented into the eight conditions described above.

In order to evaluate the new technique for rejecting trials based on eye tracking data, we also conducted the same ERP analysis using the EEG analysis software Net Station. The EEG data analysis was the same as described above except for the trial rejection procedure. A standard artifact detection procedure, that is based on EEG amplitude, was used to detect eye movements and artifacts (Luck, 2005).

Measuring the ERP N400 effect: In the 0° condition, a clear negative peak was observed in the unanticipated condition in the right parieto-temporal area in the general time window expected for an N400 component, whereas this was not the case for the anticipated condition. In agreement with N400 literature (Amoruso et al., 2013; Kutas & Federmeier, 2000), a time window was chosen around the amplitude peak of the N400 component in the right parieto-temporal area from 300-450 ms after the stimulus onset. Due to a constant delay of EEG data inherent in the anti-aliasing filters of the GES 300 amplifier that was used, the data are offset by 36 ms in all reported ERP waveforms and time windows. This information was communicated by EGI on August 29, 2014. A cluster of 10 electrodes including P4 and T6 where the N400 effect was observed were averaged together for each of the eight conditions. The corresponding electrodes in the left part of the scalp were also averaged together for each condition (Figure 2.3).

In the area where the effect was observed, one condition displayed a defined

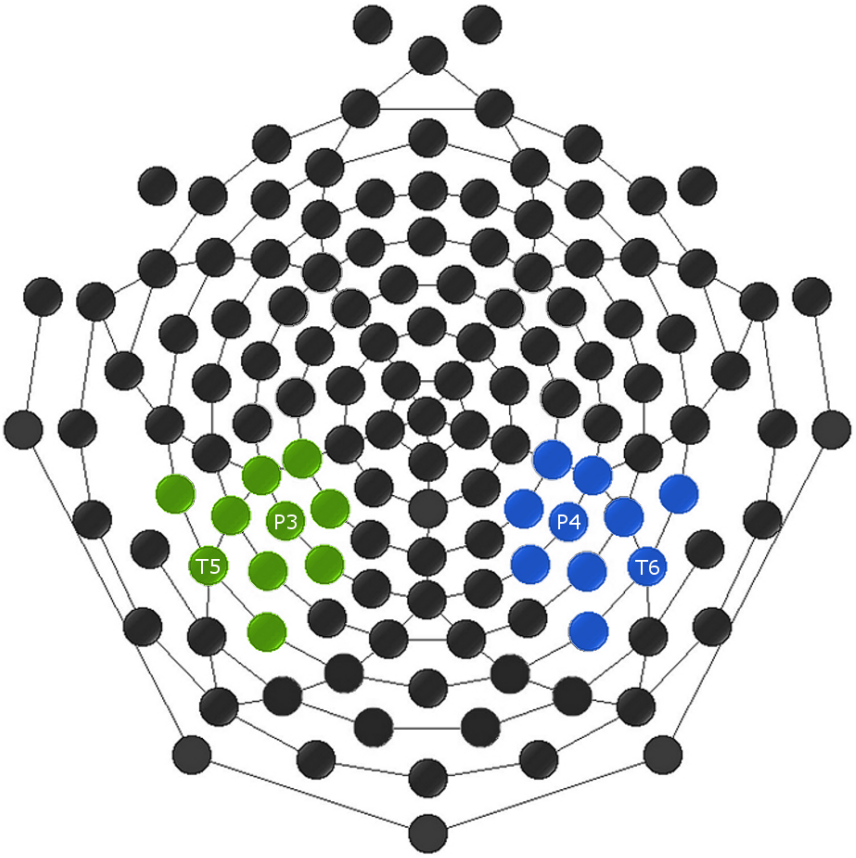


Figure 2.3: Clusters of electrodes selected for the analysis of the N400 component. A cluster of ten electrodes was selected based on where the N400 component was observed in the right parieto-temporal region of the scalp (blue). Symmetrically, a cluster with the corresponding electrodes was chosen in the left parieto-temporal region (green).

peak, whereas the other did not. This effect had the same morphology as the one observed in the original study (Reid et al., 2009) for the adult data, and therefore the same window analysis technique of Hoormann, Falkenstein, Schwarzenau, and Hohnsbein (1998) was utilized. In this analysis, a repeated measures analysis of variance (ANOVA) is conducted with time as an additional within-subjects factor. A significant interaction of condition with time means that the waveform evolves differently for each condition and therefore the effects are different. Using time as a

factor can be problematic in an ANOVA as adjacent sample points correlate more highly than more distant points. Although the conservative Greenhouse-Geisser method is intended to account for violation of sphericity, this method is also the standard approach used to deal with correlations between observations (Hoormann et al., 1998). Thus, all the statistical results reported were corrected with the conservative Greenhouse-Geisser method.

To study whether the N400 component changed significantly between eccentricity conditions, we analyzed the differences in ERP amplitudes by a 4 x 2 x 2 x 13 general linear model repeated measures ANOVA analyzing eccentricity (0°, 4°, 8° and 12°), congruency (anticipated, unanticipated), hemisphere (parieto-temporal right and left area) and time (13 samples at one sample per 12 ms). We were interested in the triple interaction effect between eccentricity, condition and hemisphere. Follow up analyses with the same factors but pairing between eccentricity conditions showed which eccentricity pairs had a significant difference in their N400 component. No effects with time in these analyses indicate that the evolution of the waveform is always similar within and between most of the conditions, even if the amplitudes are different and significant between eccentricities, hemisphere and congruency.

We also ran four independent statistical analyses to compare the results of the N400 differences between congruency conditions for each of the eccentricity conditions. Variances of ERP were analyzed by a 2 x 2 x 13 general linear model repeated measures ANOVA analyzing congruency (anticipated, unanticipated), hemisphere (parieto-temporal right and left area) and time (13 samples at one sample per 12 ms). The aim of this approach was to apply the standard statistical methodology used to report results in ERP studies. This is, whether the component of interest is

different between conditions. We were interested in the interaction effect between hemisphere and congruency. Follow up time window analyses showed which eccentricity and hemisphere had a significant difference in the N400 component between congruency conditions by the interaction of congruency and time.

In addition, the visual ERP component N1 (channels 71, 72, 75 and 76) was analyzed between eccentricity conditions. A repeated measures ANOVA analysis was conducted with the eccentricity condition as a within-subjects factor. The analysis was based on the average amplitude measured during the time window where the N1 peak was observed for each eccentricity condition. For this analysis, both congruency conditions were collapsed into a unique condition for each subject since both of them had a similar N1 visual component. All the analyses were conducted using IBM SPSS statistics 19®.

2.3 Results

The time window analysis that included the factors hemisphere (right and left), congruency (anticipated and unanticipated), eccentricity (0°, 4°, 8° and 12°) and time indicated the following significant effects: There were main effects of eccentricity, $F(2.41, 41.04) = 12.75$, $p < 0.001$, and congruency, $F(1.0, 17.0) = 9.01$, $p = 0.008$. There was an interaction effect of eccentricity and hemisphere, $F(2.64, 44.90) = 3.67$, $p = 0.022$. More important for the hypothesis of the study, there was also an interaction effect of eccentricity, hemisphere and congruency, $F(2.06, 35.02) = 4.04$, $p = 0.025$. Six follow up time window analyses including the same factors but pairing between eccentricity conditions showed that there were interaction effects of

eccentricity, hemisphere and congruency only between the eccentricity pairings of 0° - 8°, $F(1.0, 17.0) = 7.51$, $p = 0.014$ and 0° - 12°, $F(1.0, 17.0) = 6.60$, $p = 0.02$.

Four individual statistical analyses, one for each of eccentricity condition, showed the effect that the fixation distance to the stimuli had on the N400 component of the ERPs. The 0° condition displayed an N400 component in the right parieto-temporal region of the scalp in the unanticipated condition but not in the anticipated condition. No N400 effect was observed in the left region (Figure 2.4 - 0°). For the 0° condition, a time window analysis including hemisphere, congruency and time indicated an interaction effect of hemisphere and congruency, $F(1.0, 17.0) = 6.25$, $p = 0.023$. Two follow up time windows analyses were conducted, one for each hemisphere, including congruency and time as within-subject factors. In the right parieto-temporal region of the scalp, the results indicated a congruency by time interaction in the repeated measures ANOVA, $F(3.51, 59.77) = 2.84$, $p = 0.038$. This means that the N400 component evolved significantly different in each congruency condition in the right region for the 0° condition. No effects were observed in the left parieto-temporal region of the scalp.

For the 4°, 8° and 12° conditions, the ERP waveform decreased in amplitude between eccentricities for all the components across the scalp compared to the 0° condition. In the case of the N400 component, the effect showed a decrease in amplitude for the 4° condition and no N400 component was observed for the 8° and 12° conditions (Figure 2.5). Three more time window analyses were conducted, one for each of the remaining eccentricities (4°, 8° and 12°), including hemisphere, congruency and time. The results showed no interaction effects of hemisphere and congruency for any of the three eccentricities. Six individual time window analyses

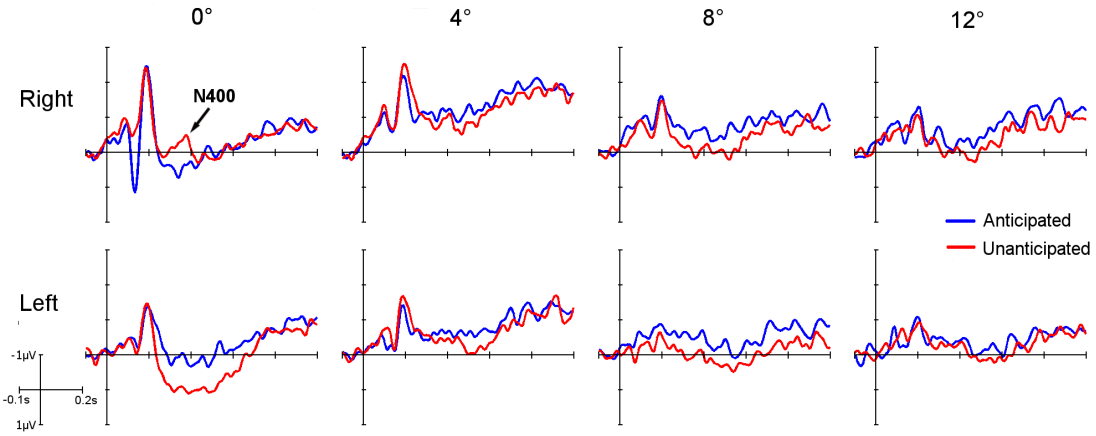


Figure 2.4: ERP averaged waveforms of the different location conditions in the parieto-temporal right and left clusters. The colors indicate the two different congruent conditions: anticipated (blue line) and unanticipated (red line). The N400 effect can be mainly observed in the anticipated condition for the 0° condition in the right parieto-temporal region of the scalp (arrow), and decreased in amplitude for the 4° condition.

for the eccentricities 4°, 8° and 12° for left and right hemisphere of the scalp indicated no significant interactions: in the 4° condition, the congruency by time interaction ANOVA was not significant in the right parieto-temporal region, $F(4.03, 68.47) = 0.73$, $p = 0.575$, as the N400 effect in the unanticipated condition was reduced in amplitude (Figure 2.4 – 4°). In the 8° and 12° conditions, the right parieto-temporal region of the scalp did not display an N400 effect in the unanticipated condition (Figure 2.4 – 8° and 12°). This was also shown in the congruency by time interaction in the ANOVA conducted for each of the two remaining eccentricities, where no significant differences were obtained in the right parieto-temporal region ($F(4.50, 76.38) = 0.50$, $p = 0.759$ and $F(4.51, 76.77) = 0.95$, $p = 0.449$, 8° and 12° respectively). No N400 effect was observed in the left parieto-temporal region for any of the eccentricities (Figure 2.4). This observation was confirmed by the results of the congruency by time interaction in the three repeated measures ANOVA analyses

conducted, one for each eccentricity (4°, 8° and 12°).

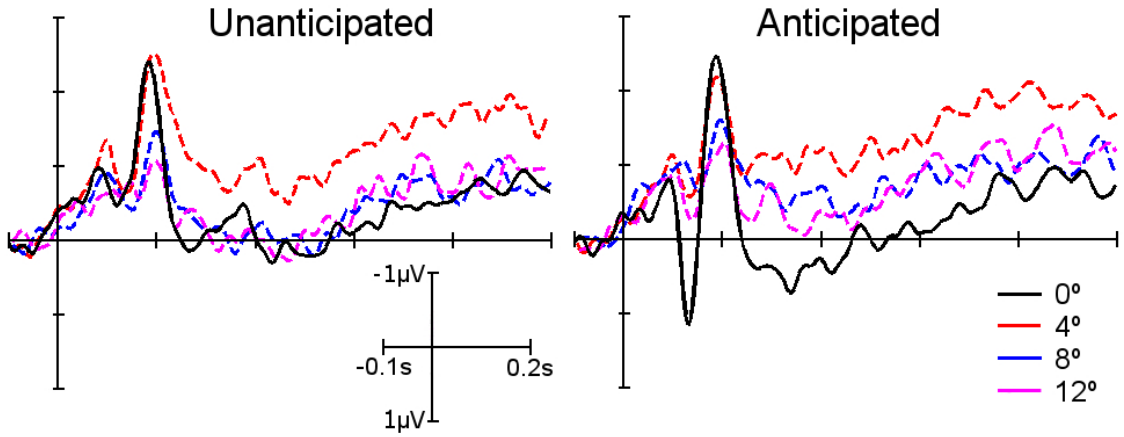


Figure 2.5: ERP averaged waveforms of the unanticipated condition (left) and anticipated condition (right) in the right parieto-temporal cluster. The different colors indicate the four different location conditions: 0° (solid black line), 4° (dashed red line), 8° (dashed blue line) and 12° (dashed magenta line). The N400 component can be observed in the unanticipated condition at 0° and decreased in amplitude at 4°.

The amplitude of the early visual components, including P100 and N100, decreased proportionally along the scalp, being reduced in amplitude as a function of the increase in visual degrees away from the displayed image. The highest mean amplitude of the measured visual N100 ERP component was observed when the stimulus was within foveal limits: 0° (-1.68μV, SE = 0.30). The mean amplitude of this component was reduced in amplitude when showing the stimulus within parafoveal limits: 4° (-1.09μV, SE = 0.31). The lowest mean amplitude was observed when the stimulus was presented within peripheral vision: 8° (-0.60μV, SE = 0.20) and 12° (-0.81μV, SE = 0.25). These differences were confirmed in the main effect of eccentricity in the repeated measures ANOVA of the visual component N100, $F(3, 51) = 7.12$, $p = 0.003$. Analyses pairing between eccentricities showed significant effects between the 0° and 8° condition, $F(1, 17) = 17.21$, $p = 0.001$ and 0° and 12°

condition, $F(1, 17) = 6.67$, $p = 0.02$. No significant effect was obtained between the pairing 0° and 4° condition.

2.3.1 Comparison of artifact detection techniques

The same results were obtained when applying the statistical analyses to the EEG data that was pre-processed using a standard ERP procedure based on EEG amplitudes. A significant difference in the congruency by time interaction was shown in the ANOVA, $F(3.75, 63.884) = 2.8$, $p = 0.035$ in the 0° condition for the right parieto-temporal region of the scalp. There were no differences in the left hemisphere in the 0° condition. Also, no differences were found for 4°, 8° and 12° conditions in parieto-temporal right and left regions.

2.4 Discussion

The aim of this study was to investigate how the N400 component of an ERP is affected by visual eccentricity when images are presented as stimuli. For this purpose, we showed N400 inducing stimuli from Reid et al. (2009) at different eccentricities from the fixation point of the participant. We expected to observe the same N400 effect in the 0° condition in our study as the participants' foveal vision was within the critical element of the stimulus.

In the original study, an N400 component was observed in the unanticipated but not in the anticipated condition across the frontal, central and parietal regions of

the scalp. We partially replicated these results, finding an N400 component with the same morphology in the unanticipated but not in the anticipated condition but with a moderately different scalp distribution. Our results show this effect across the right parieto-temporal region. There are at least two aspects related to the experimental design that could have influenced these spatial differences. First, there were technical differences between the studies as they employed different EEG systems. In the original study, the EEG system consisted of 23 electrodes located according to a 10-20 system. During the analysis, the data were re-referenced to the linked mastoid electrodes. In the present study, a high-density array of 128 electrodes was used. The linked mastoid reference is not ideal with high-density EEG systems. The recommended reference method is the average across all the electrodes since a good approximation of a zero potential of the head is obtained (Bertrand, Perrin, & Pernier, 1985; Dien, 1998). This aspect could have contributed to the differences observed as ERP waveforms are known to change in amplitude depending on the reference method applied (C. Joyce & Rossion, 2005; Yao et al., 2005). However, a visual inspection of the grand average obtained after re-referencing the data to the linked mastoid revealed that the N400 effect did not change substantially, having the same scalp distribution and morphology as with the average reference.

A second factor to explain the different scalp distribution of the N400 component, is the changes made to the paradigm with respect to the original study. Few changes were made in order to modify the paradigm to fit the research questions within the present study. Those changes that did exist comprised (a) higher number of times that the stimuli were presented to the participant, (b) the presence of the fixation cross also during the presentation of the stimuli and (c) the given task of fixating

at the cross but attending to the stimuli irrespective of the location of the stimuli.

The N400 effect that is obtained during the viewing of images is commonly reported in frontal and central locations (Amoruso et al., 2013). However, Shibata, Gyoba, and Suzuki (2009) reported a parietal N400 effect when studying cooperative actions using pictures. Some authors also report a right lateralization of the N400 due to action images (Sitnikova & Holcomb, 2008; West & Holcomb, 2002). Moreover, the N400 effect found in the present study matches with the N400 effect usually found in language studies (Kutas & Federmeier, 2011).

Regardless of the N400 distribution across the scalp, the most important result of the current study shows that this component was affected as soon as the main part of the stimulus was not within foveal vision. In the 4° condition, the fixation cross was still on the critical stimulus element within parafoveal limits and the N400 component could be observed in the unanticipated condition in the right parieto-temporal ERP cluster (Figure 2.4 - 4°). However, at 4° the N400 had decreased in amplitude, leading to a non-significant statistical result when applying the same time window analysis as the 0° condition. Additionally, the N400 effect appeared later in time, suggesting that the stimulus required more time to be processed. A shift in the time window in order to index the shifted N400 effect within the window in this condition did not lead to significant results. When the critical element of the stimulus was within the limits of peripheral vision (8° and 12° condition) no N400 effect for the unanticipated condition could be observed (Figure 2.4 – 8° and 12°). In these conditions, the fundamental aspect of the stimulus was still within visual field limits and therefore eccentricity clearly influenced the cognitive capacity to process this information.

Taking into account the results obtained in the present study regarding eccentricity, we hypothesized that the N400 component would be affected when computing an averaged ERP combining trials presented at various eccentricities, which might happen in canonical EEG studies not taking eye tracking information into account. We confirmed this hypothesis by means of a simulation analysis. We intermixed trials from across stimulus location conditions and applied the same N400 analysis to the new sets of ERPs. The N400 effect was consistently nullified when valid foveal trials were intermixed with at least 10% of the trials from other locations. The risk of intermixing trials from different eccentricities might not be high in adult ERP studies since fixating a specific part of the screen is an easy instruction to follow. Some experimental participants, however, are not able to follow instructions, for example, children and infants. This means that they may not fixate at the critical part of the stimulus on all trials, with the result that the stimulus is presented at a mix of eccentricities. As the results of the present study show, this will increase the risk of obtaining a diminished N400 effect, or even of nullifying the N400 differences between conditions.

To our knowledge, this is the first study that examines how the late components of an ERP are affected by the fixation distance to the stimuli. The current results show that there are effects on the N400 as a function of distance to the target location. The nature of the mechanisms that cause this to happen are still unknown. One candidate is spatial attention. Studies of spatial attention in word processing demonstrate that the N400 is modulated or even eliminated by selective spatial attention (Cristescu & Nobre, 2008; McCarthy & Nobre, 1993). The design of the present study was intended to exclude the variable of spatial attention as participants were instructed to always attend to the stimulus, irrespective of its location.

However, the lack of a more specific and demanding task during the experiment could may have influenced the participants' allocation of attention.

The fixation distance to the stimuli has already been studied in relation with attention on early components of an ERP (Clark & Hillyard, 1996; Eason, 1981; Luck et al., 2000; Mangun & Hillyard, 1988). Our results are in line with previous findings, with a significant proportional decrease in the visual N1 component as the fixation distance to the stimuli increases. From the results obtained in this study, we can hypothesize that there is a relationship in terms of how the early and late components of an ERP are affected by the fixation distance to the stimuli. How different components interact is an area of research that requires further attention in the future.

The rate of exclusion was higher than it is usually reported in other ERP studies. Approximately 20% of the participants were excluded because they did not reach the minimum number of trials required for one or more of the conditions of the experiment. The main reason behind this was the nature of the task, with the presentation of the image outside foveal vision. Roughly 30% of the trials seen by the rejected participants were excluded because saccades were made towards the image when its location was not at zero degrees.

Eye tracking data were recorded simultaneously with EEG data. In addition to using eye tracking data to ensure that the participant fixated on the correct location at the beginning of each trial, the data were also used to detect trials that were contaminated with blinks and saccades. This approach was shown to be as valid as the traditional methodology based on assessing changes of the amplitude of the EEG data. However, the blink algorithm used was not fully robust, and

visual inspection was required to detect false positives and false negatives. Some studies have used blink detection algorithms for different purposes (e.g., Bonifacci, Ricciardelli, Lugli, & Pellicano, 2008; Brouwer, van Ee, & Schwarzbach, 2005; Geng, Ruff, & Driver, 2009; Karatekin, 2007). In these studies, the quality of the method used was not reported, most likely because no other physiological measurements, such as EOG, were compared with the obtained results. The present study suggests that in order to have reliable automatic blink information from eye tracking data, the improvement of blink detection algorithms would be necessary.

Having access to the participant's gaze information during an ERP experiment has some advantages that can be used to ensure the quality of the ERP data. First, gaze-contingent techniques can be applied during the experiment to make sure that the participant fixates on the critical part of the stimulus. In our study, the fixation cross remained on the screen as long as the participant was not fixating on it for a minimum amount of time. This ensured that the participant was looking at the right location at the beginning of the trial and gave more control to the participant in terms of pacing the experiment. Additionally, eye tracking data can be used to discard invalid trials due to eye movements or blinks. This might not be necessary in studies with adult participants since the use of common detection techniques based on EEG amplitude are reliable. For more complex populations, such as infants, the selection of valid ERP trials is usually manually performed (De Haan, 2007a; Hoehl & Wahl, 2012). In this context, eye tracking could improve the quality of the final ERP waveform via utilizing the gaze data to help select those ERP trials where the infant attended to the stimuli.

In sum, we have shown that when viewing still images that induce N400 re-

sponses, the critical aspect of the stimulus needs to be within foveal vision. The inclusion of non-foveal trials that contain a reduced N400 effect can rapidly modify the ERP waveform, decreasing quality and potentially masking effects that have been generated by the stimulus. Our results suggest that the disappearance of advanced cognitive components from the ERP response when a participant is not fixating within the main part of the stimulus is related with the fixation distance to the stimulus. Further research able to control covert attention would help understand if there are more variables involved in this process, like the allocation of attention in the stimulus.

2.5 Prelude to Chapter 3

The preceding study shows that the visual distance to the stimuli has a negative impact in the amount of ERP response that is related to high order cognitive processes, and specifically to the ERP component under study: the N400. Importantly, in the parafoveal condition, which showed a diminished N400 ERP component, the fixation location was still within the stimulus area. This suggests that the center of the stimulus might not always be the best location to present the fixation cross that normally precedes a trial in a visual ERP experiment. The location of the fixation cross should be chosen so that the foveal vision of the participant falls within the main part of the stimulus, i.e. where the main visual information concerning the research question is contained.

The results obtained in Chapter 2 have two main implications for visual ERP techniques:

The first is a practical implication that needs to be taken into consideration when designing a visual ERP experiment: Either stimuli need to be designed so that the main visual information is centered on the image, or the fixation cross needs to be positioned at the closest location to the main visual information. According to the results of Chapter 2, both strategies will maximize the ERP response due to distance of the fixation image or cross to the stimulus. With the former strategy, the fixation cross can be kept at the center, which is the most common practice in ERP research. With the latter strategy, the stimulus can be designed freely without restrictions of visual information location.

The second implication is related to ERP data quality. Including trials in the

ERP average where the participant's foveal area was not within the main visual information of the stimulus and, therefore, the ERP response is diminished or not present, will have a negative contribution to the averaged ERP of that participant. This second implication is of great importance, especially in experimentally difficult populations such infants, since it is challenging to control and have certainty where an infant looked during an ERP trial.

Chapter 2 partially helped address the first research question of this thesis about the potential usefulness of using eye tracking technology to improve infant ERP techniques. Although the conclusions are based on adult populations, it is reasonable to hypothesize that eye tracking technology could help improve infant ERP data quality by assisting during the presentation of trials with infant visual behaviour information. As outlined in the introduction chapter, the current methodologies used to control whether an ERP trial is valid in terms of infant visual behaviour are mainly based on live observation during the experiment or offline observation during data editing based on video recordings of the infants behaviour. The resolution and precision of these techniques is uncertain, particularly if researchers need to differentiate between foveal or parafoveal distances to the main part of the stimulus, as the previous results suggested. In this sense, eye tracking is a much more precise technique to measure and evaluate the validity of an ERP trial in terms of distance to the main part of the stimulus when contrasted with current approaches. There are two methods using eye tracking that could help maximize the number of valid ERP trials in terms of visual distance. The first is to use eye tracking technology as stimulus presentation control, i.e. use it to present an ERP stimulus only when the infant is looking at the fixation image prior to the stimulus. The second is to use eye tracking technology after the experiment, during the data editing in order

to select only the ERP trials where the gaze distance to the stimulus is found to be within the foveal range. The approach that will be explored in the next chapter is the first method. The hypothesis is that stimulus presentation control techniques based on eye tracking technology could help maximize the chances that the infant's foveal vision is within the main part of the stimulus when a trial is presented and, therefore, maximize the ERP response due to the distance to the stimulus with it.

In Chapter 3, this hypothesis is explored by proposing and testing a new gaze-contingent methodology that is used during the presentation of an ERP experiment. The approach taken when defining the new stimulus presentation methodology was to use the online gaze data obtained from the eye tracker to control the presentation of the ERP trials. Specifically, to show trials only when the infant was looking at the fixation image presented prior each ERP stimulus. The aim of this approach was twofold: first, to ensure that the infant's foveal vision was within the main part of the stimulus, and second to obtain a higher engagement of the infant to the presentation by showing an attention grabber whenever the infant was distracted from the stimulus presentation. This information was obtained in real time from the eye tracker which tracked the infant's eye movements throughout the ERP stimulus presentation. To analyze and validate the methodology proposed in Chapter 3, a between-subject study was designed where half of the infants experienced a regular ERP stimulus presentation without any gaze contingency control of the ERP trials, and the other half experienced the stimulus presentation with the methodology proposed.

Chapter 3

Gaze-Contingent Techniques Help Sustain Infants' Attention and Improve Infant ERP Morphology

Text in preparation for submission

Abstract

The event-related potential (ERP) is a common technique used in infancy research to gain information about cognitive development. This technique, however, requires that the stimuli are presented in repetition dozens of times. Keeping an infant's attention to the stimulus during a visual ERP experiment is one key factor in obtaining valid results. We have used a gaze-contingent stimulus presentation to guide infants' attention to the important part of the stimulus. We compared the eye tracking and ERP results with another group of infants that participated in the same experiment but experienced a standard technique of presenting ERP stimuli to infants. We found differences in the visual behaviour of each group. This was shown in the distribution of gaze on the screen in the eye tracking data. We also found a

difference in the ERP waveform of the negative central (Nc) component, which is a marker of visual attention and further indicates that the infants assigned to the gaze-contingent group were more attentive to the stimuli. Moreover, we found an enhancement in the overall morphology of the ERP waveform in all scalp locations, suggesting that the improvement in attention also led to an improvement in the signal-to-noise ratio of the final ERP.

3.1 Introduction

Event-related potential (ERP) methodologies are a common practice in developmental research as they can provide insights into how perceptual and cognitive faculties change during early development. One of the advantages of this technique is that infant ERP paradigms do not need an active response from the participant. The experiments can therefore be designed in a way that no instructions are required for the participant during the experiment (De Haan, 2007a). Usually, differences between conditions can be interpreted within a framework of prior studies that focus on different ERP components, each capable of informing developmental scientists on the specific processing capacities of the sample. However, in the case of visual ERP paradigms, one of the most important requirements to obtain an appropriate ERP brain response is that the infant pays attention to the stimuli (Hoehl & Wahl, 2012). This is indeed one of the main challenges as the repetition of the same stimuli many times within an experimental session, as it is the case of ERP methodology, maximizes the likelihood of reducing infant attention.

During an ERP recording, trials are typically presented in a pseudo-random order and in a continuous mode—uninterruptedly without breaks—until the infant loses interest in the stimuli and starts to attend to other aspects of the environment (De Haan, 2007a). The current attrition rate when conducting an infant ERP experiment is expected to be around 50% of all infants tested (De Haan, 2007a; Stets et al., 2012). One of the main causes for excluding infants from the final ERP analysis is the lack of enough valid trials. In order to include an infant into the final analysis, a common criterion is at least 10 artifact-free trials per condition (e.g., Hoehl, Reid, et al., 2008; Reid et al., 2009, 2007; Striano et al., 2006). The reason for discarding

trials is commonly reported to be movement artifacts and lack of attention to the stimuli, also described as 'fussiness'. Therefore, sustaining the infant's attention for as long as possible during a visual ERP experimental session appears to be one of the keys to increasing the number of artifact-free trials that can be obtained.

Controlling the stimulus presentation based on the infant's attention to the screen is a widespread practice in infancy research. One common approach to recover the infant's attention to the stimuli is to present 'attention grabbers'. These commonly consist of a combination of dynamic stimuli including audio with high saliency that are likely to attract the infant's attention back to the stimulus location. The decision to deploy an attention grabber is a matter of experimenter discretion. Depending on the stimulus presentation and the experimenter, the decision can vary considerably between experiments and labs. In some experiments, the attention grabbers are shown when the experimenter considers that the infant is not attending to the stimuli (e.g., Nelson & De Haan, 1996; van Hoogmoed et al., 2013). In other experiments, the attention grabbers are shown automatically before each trial or block of trials (e.g., Junge et al., 2012; Yrttiaho et al., 2014). Methods to control the stimulus presentation based on the infant's attention have also occurred. First attempts to control infant's attention were conducted by the experimenter observing the infant in real time with a modification of the stimulus presentation depending on the infant's behaviour (e.g., Fantz, 1964; Horowitz, Paden, Bhana, & Self, 1972; M. H. Johnson, Posner, & Rothbart, 1991). More recent technology has allowed the development of more complex and precise gaze-contingent experiments via the use of eye tracking technology (e.g., Deligianni et al., 2011; Shukla et al., 2011; Wang et al., 2012; Wass et al., 2011). In general, the studies report that gaze-contingent techniques have positive results in the level of engagement of the infant in the experiment.

Taking the previous studies into account, we hypothesize that a gaze-contingent presentation of ERP trials would help sustain an infant's attention to visual stimuli. This would be reflected in a more pronounced difference between conditions, with a relative increase of the negative central (Nc) component in the average ERP compared to the Nc component obtained with a standard ERP presentation technique. The Nc component has been related to allocation of attention in infants and is usually observed over fronto-central areas of the scalp (De Haan, 2007a; Richards, 2003). We also expected that the use of a gaze-contingent presentation of ERP trials would result in an increase of the number of trials accepted and in the quality of the final ERP morphology. These hypotheses were tested comparing two groups of infants: a group that experienced stimuli presented via a gaze-contingent paradigm, and a control group that encountered a traditional continuous mode stimulus presentation. The only difference between the two groups was the presence of gaze-contingent attention grabbers based on the infants' looking behaviour prior to the presentation of each stimulus.

3.2 Methods

3.2.1 Participants

Twenty-eight infants (12 males, 16 females) were tested at ten months of age (322 ± 26 days). Before the experiment started, the infants were assigned randomly to one of two possible groups: continuous or gaze contingent. Five of these infants had to be excluded in the ERP analysis due to a technical problem with the video recording of the infants' behaviour during the experiment. The video recording was necessary in the methodology used for analysing the infant ERP data. Thus, the final sample included in the eye tracking analysis consisted of 28 infants (continuous group = 13, gaze-contingent group = 15) and the final sample included in the ERP analysis consisted of 23 infants (Continuous group = 10, Gaze contingent group = 13). Another eight infants were tested but excluded from the final sample: Four infants due to tracking difficulties of the eye tracker and four infants because of technical problems with the EEG system.

All infants tested were born full-term without complications and with normal birth weight. The parents were informed about the experimental procedure before it started and gave written consent for their infant to participate in accordance with institutional protocols. The study was approved by the Lancaster University Research Ethics Committee.

3.2.2 Apparatus

Stimulus presentation

The visual stimuli were presented on a 20-inch CRT monitor (Mitsubishi Diamond Pro 2070SB) with a refresh rate of 60Hz. The monitor resolution was 800x600 pixels. All visual stimuli were shown using Psychtoolbox 3 for Matlab® (Kleiner et al., 2007).

EEG

EEG activity was measured via a 124-channel Sensor Net (Electrical Geodesic, Eugene, USA) with a sampling rate of 500Hz and recorded using the software Net Station. Raw EEG data were recorded with the vertex electrode (Cz) as the online reference. All infants were videotaped and the video was synchronized with the onset of every stimulus for later analysis of the infants' behaviour during the ERP analysis.

Eye tracker

Eye movements were recorded using a Tobii TX300 eye tracker (Tobii AB, Danderyd, Sweden) with a sampling rate of 120Hz. Eye tracking data were recorded using the Tobii SDK 3.0 in Matlab on the presentation computer.

3.2.3 Visual Stimuli

The experimental procedure used was primarily designed to investigate the neural correlates of statistical learning of action sequences in infants (Monroy et al., 2017). The experiment consisted of two phases: a training phase with video stimuli, and an ERP phase with a sequence of static stimuli. Given that these were visually presented stimuli, an Nc component was expected to be elicited by specific stimuli during the test ERP phase.

Training phase

A training phase, not analysed in the present study, comprised videos of an actor's hand performing actions on a box. The box had six unique button objects equally positioned around its edges. Each object afforded a distinct manual action. In total, four video blocks were shown to the infants. The videos were presented at 35°x 20°. The average length of each video during the training session was 91 seconds.

ERP phase

Nine different pictures were used as stimuli. The pictures were single frames extracted from the training videos, depicting mid-action execution movements. The pictures were presented with a size of 35°x 20°.

3.2.4 Procedure

The infants sat on their parents lap at a distance of 65 cm from the monitor. Prior to the start of the experiment, the EEG net was positioned on the infants' heads and a five-point infant calibration routine of the eye tracker was performed. During the training phase, the same procedure was followed for all the infants, irrespectively of the group they were assigned to. Therefore, the results of the training phase were only used in the present study to make sure that no quality bias in the eye tracking data existed between the groups. The videos were presented in a continuous mode. The experiment was paused or an attention grabber was shown whenever the infant did not attend to the screen.

The ERP phase consisted of a maximum of 144 trials. Regardless of the continuous or gaze-contingent group, the experiment continued until all the trials were presented or until the experimenter considered that the infant had lost interest in the stimuli. In both groups, the experimenter paused the presentation or showed an attention grabber whenever the infant was distracted from the stimuli for a number of sequential trials. Each trial started with a fixation image showing a picture of a duck on a black background. The fixation image was presented at a size of 3°x 3.7°. The image of the duck changed its location for each trial so that it matched with the location of the object being acted upon in the next trial (Figure 3.1). Six different positions of the duck were possible. All the positions were distributed on the screen with an average distance of 10° to the centre of the screen. None of the positions corresponded with the centre of the screen.

The ERP presentation procedure was different for the continuous and the gaze-

contingent group:

Continuous group

In the continuous group the fixation image was shown for one second and, immediately after, the corresponding stimulus was shown for one second. That is, the presentation of trials was conducted in a continuous fashion (Figure 3.1-B).

Gaze-contingent group

In the gaze-contingent group, the fixation image was also shown for one second. The eye tracking data were used to check the infant's point of gaze (PoG) at the end of the presentation of the fixation image. Three different alternatives were possible from this moment, depending on the infant's gaze behaviour (Figure 3.1-A): (1) First attempt: if the infant's PoG was within 4° around the fixation image at the moment of the first gaze check, the stimulus image of the corresponding trial was shown for one second. This resulted in the same trial procedure as the continuous group. (2) Second attempt: if the infant's PoG was not within the area of interest, the fixation duck would wiggle together with a jingling sound to capture the infant's attention. When the infant's gaze was measured to be within the area of interest, the movie stopped and the fixation image was shown again for one second. Immediately after, without checking the infant's PoG again, the corresponding stimulus was shown for one second. (3) Third attempt: If the video of the moving duck was presented for four seconds and the infant's PoG was still not within the area of interest, the trial started automatically. As in (2), the fixation image was shown for

one second, followed immediately by the next trial stimulus. The third option was implemented to allow the experiment to continue without significant disruption, and also to increase the chance of regaining the infants' attention as a consequence of changes in the stimuli. The gaze-contingent process was repeated every trial. For each trial, the stimulus was triggered by one of the three options described above, depending on the infant's looking behaviour.

3.2.5 Analysis

Eye tracking data processing

Eye tracking data were processed using custom-made Matlab scripts. The analyses of eye tracking data focused on the percentage of valid eye tracking data and the distribution of infants' gaze on the screen during the ERP phase for the two groups. Raw eye tracking data were pre-processed prior to data analysis. The pre-processing was based on the previous steps reported in some common fixation filters: data smooth applying a moving average, interpolation of missing points up to 100 ms and eye averaging (Olsen, 2012; Saez de Urabain, Johnson, & Smith, 2014).

Percentage of valid eye tracking data. The percentage of valid eye tracking samples during the presentation of stimuli was calculated for each presentation phase and group to evaluate the quantity of eye tracking data collected. We considered valid eye tracking samples the ones with a validity code other than 4 (value assigned by the Tobii SDK to samples where the eyes are not found). We included in the analysis the periods of time where the videos where presented during the training phase and the static stimuli during the ERP phase.

Distribution of gaze. We examined the looking behaviour during the presentation of the stimuli in the ERP phase for each group. For that, the root mean square (RMS) was calculated, which gives a measure of distribution of gaze during a period of time (Gredebäck, Eriksson, Schmitow, Laeng, & Stenberg, 2012). RMS provides an estimate, in degrees of visual angle, of the distribution of gaze relative to the centre of the fixation image of each trial, and therefore this measure is independent from the fixation image location. For each trial, the RMS $\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}\theta_i^2}\right)$ of the distances to the fixation image centre was calculated by (1) computing the square of the angular distance (θ), in visual degrees, of each sample (i) of the 1-sec trial to the centre of the fixation image and (2) taking the root of the averaged resulting squared distances of the trial. Only trials with at least 75% of valid samples were included in this analysis. On average, data from 70.2 trials (SD = 23.7) per participant in the continuous group and 90.2 trials (SD = 23.4) per participant in the gaze-contingent group were included in the analysis.

ERP data processing

EEG data were analysed using the Matlab toolbox EEGLAB v13.0.1 (Delorme & Makeig, 2004). First, EEG data were filtered using a band-pass filter from 0.3 to 30 Hz. The data were then segmented into epochs that comprised a 100 ms baseline and 1000 ms after the onset of the stimulus. The first 100 ms of each epoch corresponded to the last period of the fixation image on the screen.

EEG epochs were examined off-line together with the video recording of the infant's behaviour by a trained infant ERP data editor, as is described in infant ERP literature (Hoehl & Wahl, 2012). Trials that contained artifacts (i.e. movements,

blinks) or in which infants did not look towards the stimulus were marked for rejection. Trials marked as rejected during the manual editing of the EEG data were automatically discarded. For the remaining trials, an automatic algorithm based on amplitude thresholds that had already used with infant data was applied to detect and interpolate bad electrodes (Kouider et al., 2013). Next, the EEG data were re-referenced to the average reference. ERPs were calculated for every participant by averaging all the remaining trials to create one single condition. We collapsed all the valid trials to obtain only one ERP condition per infant because we were interested in the general level of attention during the ERP experiment. All the infants with valid EEG data were included in our analysis since all of them accomplished our criteria about minimum number of 10 valid trials. On average, infants contributed with 65.70 trials (SD = 29.15) in the continuous group and 60.61 trials (SD = 24.25) in the gaze-contingent group. Both groups contained a similar number of trials from each of the nine different stimuli shown during the experiment ($p = 0.684$).

ERP grand averages were created for each group. An Nc component was identified on the basis of morphology and previous literature as a negative peak over the frontal and central electrodes centred approximately at 480 ms after the stimulus onset (Csibra et al., 2008; Reynolds & Richards, 2005). We selected groups of electrodes of the EGI sensor net around locations of the 10-20 international system for each area of the scalp where the Nc was present. These were the left frontal region (F3 and surrounding electrodes), the right frontal region (F4 and surrounding electrodes) and the central region (Cz and surrounding electrodes). We also selected a cluster of electrodes around the occipital region (Oz and surrounding electrodes; Figure 3.4).

Statistical analyses

The statistical analyses focused on determining whether the difference in the ERP experimental procedure (i.e. continuous or gaze-contingent) had an impact on the eye tracking and EEG data that could be associated with a difference in infants' behaviour and attention during the experiment. Thus, the analyses were performed with the ERP presentation condition (continuous or gaze-contingent) as a between-subjects factor. For the percentage of valid eye tracking data analysis, the phase of the experiment (training or ERP) was also added as a within-subjects factor. Although the training phase data are not analysed in this study, we used them to ensure that there were no eye tracking data quality differences between groups not related to the gaze-contingent manipulation. All the variables tested were normally distributed. All statistical analyses were conducted in IBM SPSS Statistics 19®.

3.3 Results

3.3.1 Experimental session results

Table 3.1 shows the descriptive statistics on the characteristics of the experimental session split by the presentation condition during the ERP phase (continuous or gaze-contingent). The ERP phase was longer for the infants assigned to the gaze-contingent group. In contrast, the mean number of trials shown during the ERP phase was lower for the gaze-contingent group (maximum possible number of trials shown was 144). A slightly higher percentage of trials were accepted for further ERP analysis in the gaze-contingent group (57.35%, SD = 21.63%) compared to the continuous group (52.45%, SD = 21.29%). This difference was not significant.

With respect to the triggering modes in the gaze-contingent group, approximately 70% of the trials were triggered at the first attempt. That is, the infants' PoG was within the area of interest when the fixation image was shown for the first time during a trial. The second and third attempts, which showed the stimulus after the attention grabber movie, were both used on average for around 15% of the ERP trials. Regarding the distribution of trial acceptance by triggering mode, the first and second attempt had higher percentage of trials accepted (69% and 52%, respectively) than the third attempt (29%).

	Continuous group (SD)	Gaze-contingent group (SD)
Length of the experiment		
Average length of the training phase (seconds)	394.71 (56.71)	371.26 (20.17)
Average length of the ERP phase (seconds)	269.36 (46.86)	349.14 (87.68)
Number of trials during ERP phase		
Mean number of ERP trials shown (maximum 144)	134.23 (20.05)	113.46 (31.55)
Accepted for ERP analysis (%)	*52.45 (21.29)*	*57.35 (21.63)*
Triggering mode		
Mean number of trials triggered by first attempt	-	76.93 (26.78)
Accepted for ERP analysis (%)	-	*69.29 (24.85)*
Mean number of trials triggered by second attempt	-	17.33 (10.08)
Accepted for ERP analysis (%)	-	*52.81 (36.17)*
Mean number of trials triggered by third attempt	-	19.20 (8.23)
Accepted for ERP analysis (%)	-	*29.64 (21.61)*

Table 3.1: Experimental session characteristics split by ERP presentation group (continuous and gaze-contingent). The variables included for each group are: the average length of the training and ERP phase, mean number of ERP trials shown during the ERP phase and mean number of trials triggered by first, second and third attempt (this applies only to gaze-contingent presentation group). The mean percentage of trials accepted after the manual editing procedure is also included for each group and triggering mode.

3.3.2 Eye tracking results

Percentage of valid eye tracking data

The repeated measures ANOVA results revealed a main effect of presentation phase: $F(1, 26) = 20.58$, $p < 0.01$, $\eta_p^2 = 0.442$ as well as an interaction effect between presentation phase and group: $F(1, 26) = 16.17$, $p < 0.01$, $\eta_p^2 = 0.384$. Post-hoc independent-sample tests confirmed that the percentage of valid eye tracking data was different between groups only during the ERP phase. Specifically, the percentage was higher for the gaze-contingent group ($M = 84.94$, $SD = 7.59$) than for the continuous group ($M = 60.86$, $SD = 13.73$): $t(26) = -5.845$, $p < 0.001$ (Figure 3.2-A).

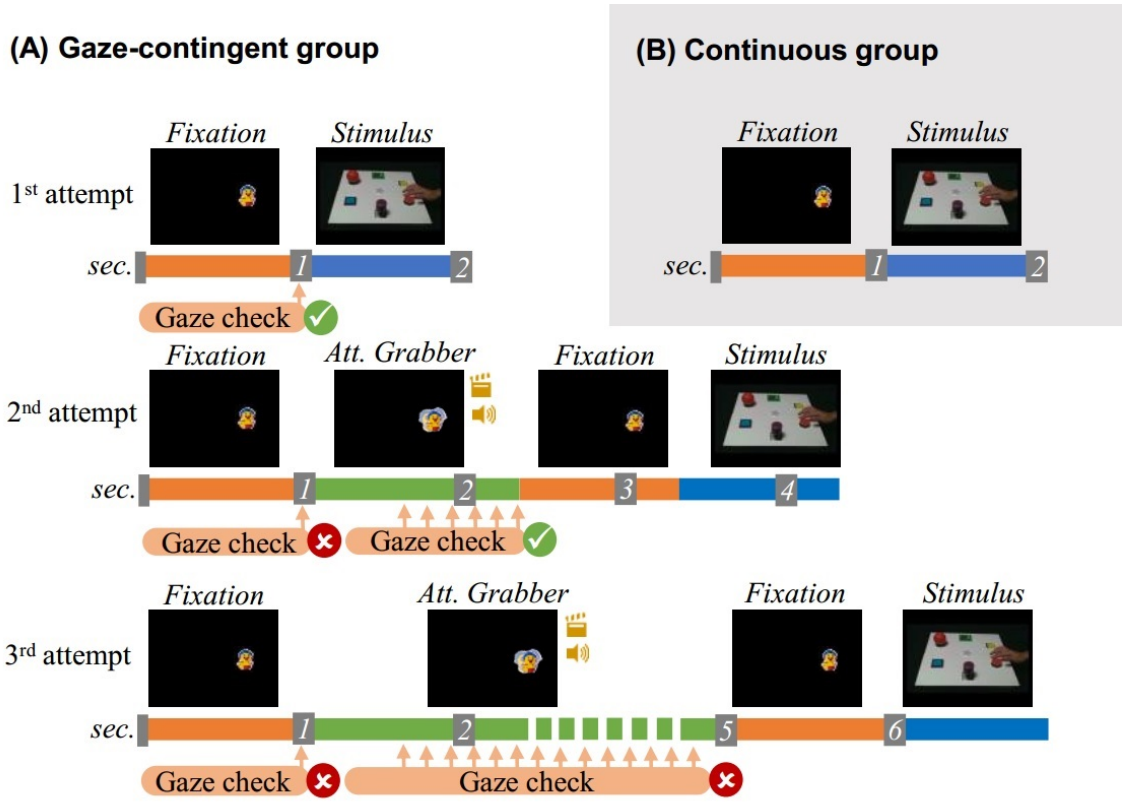


Figure 3.1: A representation of the ERP trial sequence used for each of the two groups. (A) Gaze-contingent. Depending on the infant's looking behaviour, each trial followed one of the three sequences. First attempt: The infant's PoG was within the fixation image area of interest after showing the fixation image for one second. Second attempt: The infant's PoG was within the attention grabber area of interest after starting showing the attention grabber movie. Third attempt: The infant's PoG was not within the attention grabber area of interest after showing the movie for four seconds. (B) Continuous. Every trial was triggered in the same manner: The stimulus was shown right after showing for one second the fixation image.

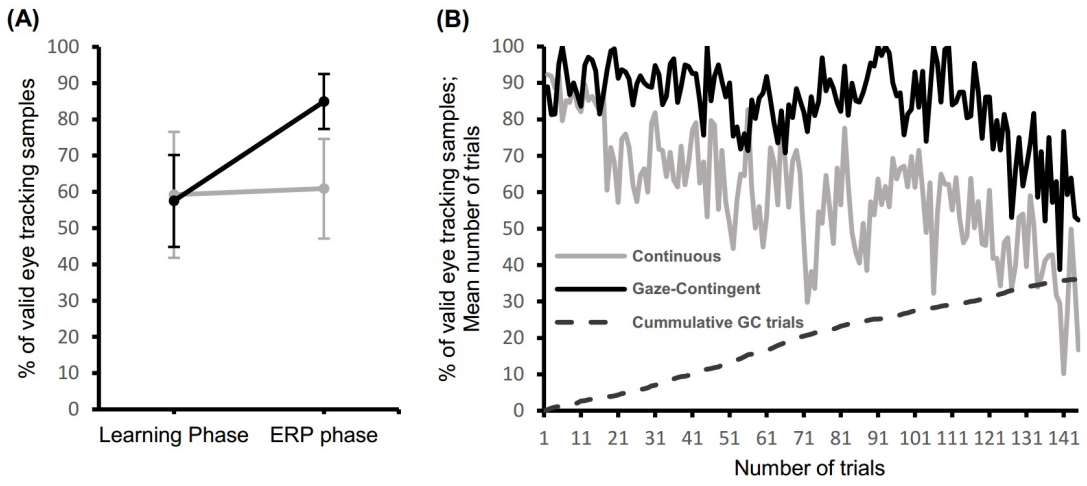


Figure 3.2: Mean percentage of valid eye tracking data during the two phases of the experiment (Learning phase and ERP phase) separately for the continuous group (grey line) and the gaze-contingent group (black line). Error bars represent standard deviation. (B): Evolution of the mean percentage of valid eye tracking data collected each trial of the ERP phase separately for continuous group (grey line) and gaze-contingent group (black line). The dashed line shows the cumulative mean number of trials triggered by second or third attempt—attention grabber was shown—in the gaze-contingent group.

The reason for the difference found in the percentage of valid eye tracking data between both groups during the ERP phase can be seen in Figure 3.2-B. It shows the evolution of the mean percentage of valid eye tracking samples over the trials during the ERP phase. The mean number of valid samples is approximately 85% for both groups at the beginning of the ERP phase. Only around one minute after the ERP phase started, the mean number of valid samples decreased abruptly for the continuous group whereas it was more stable over the trials for the gaze-contingent group. Also in Figure 3.2-B, the mean cumulative number of trials triggered by gaze-contingent during the experiment can be observed. This value ascends mostly linearly over the trials, with no specific flat area or steep slope.

Distribution of gaze

The distribution of gaze during each ERP trial was mainly expected to be around the fixation image. The RMS analysis showed that infants in the gaze-contingent group distributed their gaze nearer the centre of the fixation image ($M = 4.73$, $SD = 0.79$) than the continuous group ($M = 6.97$, $SD = 1.48$): $t(26) = 5.09$, $p < 0.001$ (Figure 3.3-A). This difference in looking behaviour during the ERP phase can also be seen in Figure 3.3-.B, which shows the mean distribution of gaze on the screen during the ERP trial period. The infants assigned to the gaze-contingent group directed their gaze mainly around the location of the fixation image and a smaller amount of time around the centre of the stimulus. The infants assigned to the continuous group had a more spread gaze distribution along the first 12 degrees from the centre of the fixation image, with no clear peaks in any areas of the stimulus.

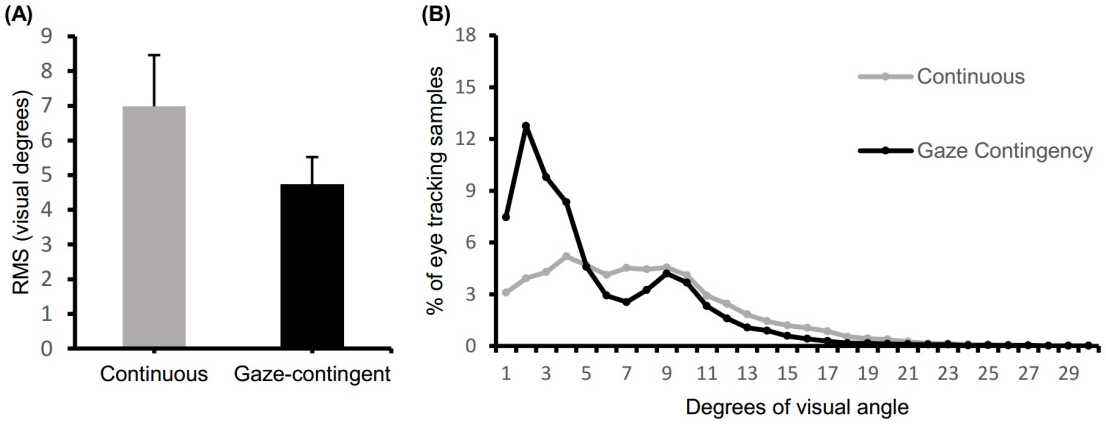


Figure 3.3: (A): Mean distance between individual gaze data samples to the centre of the fixation image (RMS) during the ERP phase separately for the continuous group (grey line) and the gaze-contingent group (black line). Error bars represent standard deviation. (B): Mean distribution of gaze on the screen during ERP stimuli separately for the continuous group (grey line) and the gaze-contingent group (black line). The vertical axis corresponds to the mean percentage of eye tracking samples. The horizontal axis corresponds to the rounded distance, measured in degrees of visual angle, of the infant's PoG to the centre of each fixation image.

3.3.3 EEG results

Figure 3.4 displays the grand average waveforms of the four electrode clusters calculated across the scalp. The Nc component was more negative for the gaze-contingent group than for the continuous group for the three clusters computed in frontal and central regions. A time window around the peak of the Nc was chosen from 400 ms to 575 ms after stimulus onset for statistical analysis. The mean amplitude of the window was calculated for each participant and frontal-left, frontal-right and central clusters of electrodes. We performed a repeated-measures ANOVA including the location of clusters (frontal-right, frontal-left and central) as a within-subjects factor, and the two presentation conditions (continuous and gaze-contingent) as a

between-subjects factor. The mean amplitude was found to be significantly lower for the gaze-contingent group (M = -5.31 μV, S.E. = 1.41) than for the continuous group (M = -1.24 μV, S.E. = 1.04) for the main effect of presentation condition: F(1, 21) = 4.787, p < 0.05, $\eta_p^2 = 0.186$.

There were no main effects on location. Also, no interaction effects between location and condition were found, suggesting that the same effect was found across the frontal and central region, as can be seen in Figure 3.4.

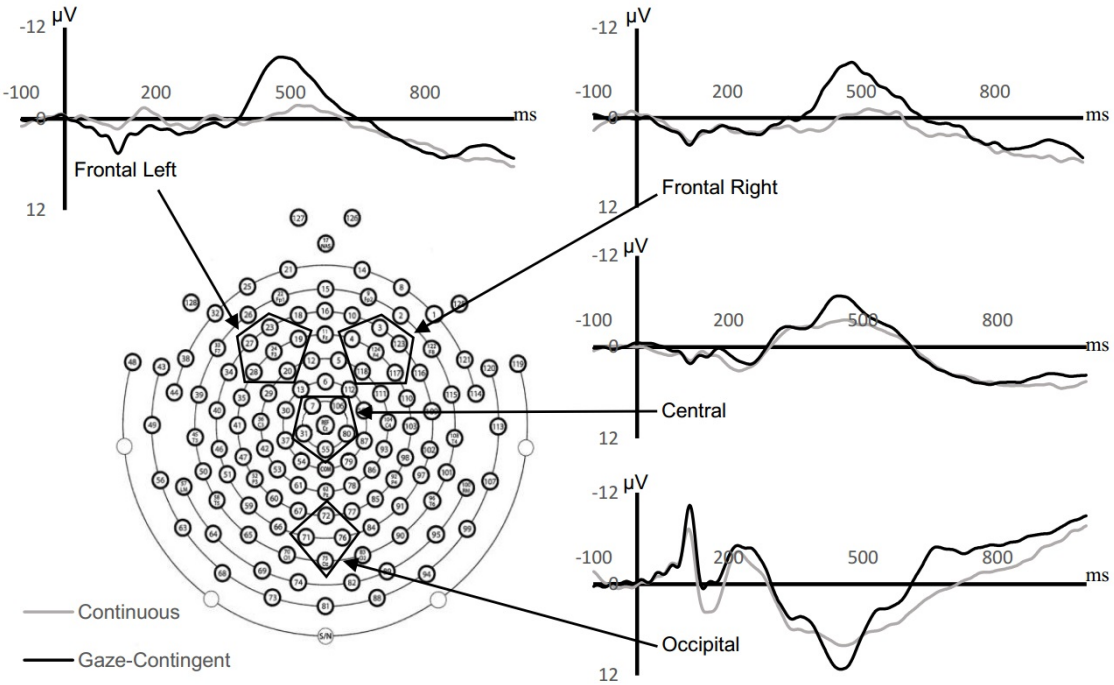


Figure 3.4: Grand average ERP waveforms for the different clusters. Grey line: Continuous group. Black line: Gaze-contingent group. Top left: Frontal-Left cluster including F3 and surrounding electrodes. Top right: Frontal-Right cluster including F4 and surrounding electrodes. Central right: Central cluster including Cz and surrounding electrodes. Bottom right: Occipital cluster including Oz and above electrodes. An Nc effect was present in the frontal and central areas of the scalp, larger in amplitude for the gaze-contingent group. Note that the negative is plotted up.

3.4 Discussion

The aim of the present study was to investigate whether gaze-contingent techniques could help sustain infants' attention during an ERP study and whether this could engender a higher quality of obtained data. We tested two groups of 10-month-old infants, one with a gaze-contingent presentation and one with a continuous presentation. The results showed that the number of trials accepted for the final ERP analysis were similar. Both the eye tracking data and ERP data show, however, that there was a clear difference in behavior and attention of the infant during the ERP phase as a consequence of contingent stimulus presentation. First, the distribution of gaze of the gaze-contingent group was more concentrated near the fixation image. Also, the Nc component of the ERP response in the gaze-contingent group was significantly larger in amplitude over the frontal and central electrodes.

Based on previous studies of attention in infants, the larger Nc component found in the gaze-contingent group can be related with sustained attention during the ERP phase (Richards et al., 2010; Richards, 2001). Also, previous studies have already shown that gaze contingency improves attention in infants (Forssman et al., 2014; Wass et al., 2011). The present results support this notion and add behavioral and electrophysiological evidence that indicate that gaze contingency helps sustain an infant's attention to the stimuli during an ERP experiment. One factor for this attentional effect seems to be increasing the complexity of the elements shown during the presentation. Animated stimuli and variability have been shown to sustain infant's attention longer than static stimuli (Cohen, 1972; Courage et al., 2006; Stets et al., 2013). Thus, our hypothesis of why gaze contingency sustained infants' attention is that we showed animations with movement and sound when the infants

started to diminish their attention towards the screen. Often this would recover the engagement of the infant to the stimuli, with more attention allocated towards the subsequent presentation of trials. Showing animated attention grabbers during an infant ERP experiment is a common practice to regain infant's attention. However, they are usually triggered manually by the researcher or automatically by the stimulus presentation in a regular basis every certain number of trials. In fact, the continuous group in our study was also shown attention grabbers following these common procedures whenever the infant was distracted for a number of trials. Presenting attention grabbers more consistently whenever the infant started to diminish the attention—as the gaze-contingent presentation did—seems to have made a difference in sustaining the infant's attention to stimuli.

There are some studies that suggest that gaze contingency can have more consequences for infants than just sustain their attention. First, in Wass et al. (2011), it is concluded that gaze contingency acted as an attention trainer for the infants. Our data suggest, however, that in our experiment gaze contingency helped sustain infants' attention but did not act as an attentional trainer. The mean cumulative number of trials that used an attention grabber as a consequence of the infant disengaging from the screen increased through the test session in a linear fashion. In our understanding, an attentional training process would be associated with a decrease in the number of trials that needed gaze contingency to grab the infant's attention back to the stimuli. There are significant differences in the experimental design between Wass et al. (2011) and the present study related to the type of stimuli and procedure, but also to the number of sessions. In Wass et al. (2011), the attentional improvement was tested in a second session after applying a gaze-contingent training session. Second, it has also been shown that infants are able to learn that their

gaze behaviour can modify the presentation of stimuli in gaze-contingent paradigms (Wang et al., 2012). This would be a negative consequence in the proposed ERP methodology because it would have presumably made the infants look off the screen more often to trigger an attention grabber. The constant increase of the mean cumulative number of trials that triggered an attention grabber also suggest that infants did not learn the consequences of gaze contingency. We believe that infants did not learn the consequences of gaze contingency in the present experiment (presentation of the next trial) because they experienced them on average only during 15% of the total presented trials.

Gaze contingency was found to positively affect the ERP response in terms of engagement and attention. It is unknown whether it modified the cognitive response to the stimuli in any other way that could affect the ERP response. This question applies to showing attention grabbers during an experiment, with or without gaze contingency. This issue should be of interest to the developmental research community as attention grabbers are a common practice in infant ERP studies. In this regard, our intention was to study whether the ERP response was different in the gaze-contingent group between the trials that were shown right after an attention grabber (second and third attempt) and the trials that were shown without it (first attempt). However, the lack of enough valid ERP trials shown after an attention grabber made that addressing this important question was unfeasible. On the basis of our data, it is unlikely that attention grabbers affect the ERP response as the animations were presented between trials and only when the infant was distracted. Further research would shed light on this important topic.

At least one previous study has found that gaze contingency helps retain more

trials during an ERP visual experiment (Maguire et al., 2014). The study was conducted with a similar experimental setup but on six to eight year-old children. Based on this and after examining the difference in the distribution of gaze on the screen between groups (Figure 3.3), we did not expect to have similar number of ERP trials accepted between both groups in our experiment. The procedure followed to assess the validity of ERP trials for both groups was normative for infant ERP studies: an ERP editor visually inspected each ERP trial both in terms of EEG noise and looking behavior of the infant. Looking behavior is coded by analyzing the session recordings frame-by-frame and marking trials in which the infant was not looking to the screen as invalid. A t-test analysis applied to the N1 visual component of the occipital area of the ERP revealed no significant differences in this component between the two presentation modes. This suggests that the trials included in both groups had similar amount of the visual N1 component in response to the stimuli. However, in our experience, the quality of the video could lead to uncertainties in the decision of whether the infant is looking at a certain location of the screen. The infant could seem to look at the screen but the precise location of fixation is ambiguous. In this regard, and considering the different pattern of distribution of gaze on the screen observed in the eye tracking data, it is possible that there were ERP trials included in the continuous group where the infants were looking at the screen but not at the fixation image. This would have introduced trials with a weaker cognitive response to the stimulus, which would have affected the final ERP quality in the continuous group (Domínguez-martínez, Parise, Strandvall, & Reid, 2015).

The lack of difference in the percentage of valid eye tracking data between groups during the training phase confirms that the differences that were found during the

ERP phase are due to the gaze contingency manipulation. It is notable that both groups of infants began the ERP phase with a similar average percentage of valid eye tracking data and, after only 20 trials, the difference in the quantity of valid eye tracking data started to increase between groups. Since the number of accepted trials were similar in both conditions, we have two different hypotheses that could have contributed to this difference: (1) the eye tracker had more difficulties in tracking infants assigned to the continuous group as they were less engaged with the stimuli and moved more. Such behavior is known to cause a decrease in the eye tracking data quality (Holmqvist et al., 2011). (2) The infant behaved in the same way in both groups but the infants assigned to the continuous group were more often outside the virtual trackable box of the eye tracker. The fact that the presentation of trials in the gaze-contingent group was dependent on the correct functioning of the eye tracker could have created an operator bias. This is, the operator would have been more attentive to the location of the infant with respect to the eye tracker to make sure that the infant was inside the trackable box and therefore the trial was triggered properly. A visual inspection of the video recordings did not show any clear behavioral bias towards any of the two hypotheses stated. Thus, we believe that the reason for the decrease in quality of the eye tracking data in the continuous group could be due to a combination of the two hypotheses. These two factors make the use of gaze contingency advantageous to maximize the eye tracking data quality of any visual infant paradigm.

The gaze-contingent presentation also appeared to induce an improvement in the quality of the final ERP morphology. All the components of the ERP—not only the Nc—in different locations of the scalp seem to be enhanced in the gaze-contingent group. Our hypothesis is that the difference in the quality of the ERP morphology is

directly related to the general infant behaviour during the experiment. Being more attentive to the stimulus presentation—as shown by the Nc component—would have also reduced the amount of movement in the gaze-contingent group. Head and body movements are one of the main causes of artifacts in infant EEG signals which may lead to a reduction in the signal-to-noise ratio of the final ERP (De Haan, 2007a; Hoehl & Wahl, 2012). An ERP with a low signal-to-noise ratio could mask cognitive responses and potentially even change final results. For that reason, obtaining the best signal quality should be the main goal of any ERP methodology. The method proposed in this study based on gaze-contingent presentation is one example of an ERP methodology that is worth applying when considering the potential improvements in ERP quality.

3.5 Conclusions

In this experiment, a gaze-contingent technique acted as an attention guide to direct infants' attention to the part of screen displaying the manipulation in the visual stimuli. Our results show that the use of gaze-contingent techniques during the ERP experiment helped sustain infants' attention to the stimuli. The change in behavior of the infants assigned to the gaze-contingent group also affected the eye tracking data quality and helped collect as many valid eye tracking samples as possible. We believe that this is a strong rationale for using gaze-contingent techniques in any kind of eye tracking experiment with infants. Importantly, the quality of the ERP data was improved as determined by the amplitude of the Nc component of the resulting grand average ERP.

3.6 Prelude to Chapter 4

The preceding experiment investigated the use of gaze-contingent techniques to sustain infants' attention during an ERP presentation. The differences in the gaze patterns observed in the continuous and gaze-contingency group suggest that the gaze-contingent presentation facilitated that infants assigned to this condition fixated the important part of the stimulus. Also, the Nc component was found to be significantly more negative for the gaze-contingent group than for the continuous group. Since the Nc component is related with attention, the Nc amplitude difference indicates that infants were more attentive to the stimuli when using the gaze-contingent technique. Finally, the general ERP morphology enhancement observed in the grand average ERP of the infants assigned to the gaze-contingent group shows that the visual and general behavioural differences had a positive impact in the general ERP data quality.

Chapter 3 further explored the potential advantages of using eye tacking during an infant ERP study, specifically, whether eye tracking increases infants' attention when used during the presentation of ERP stimuli as a gaze control tool. Our results suggest that the use of eye tracking during the presentation of repetitive trials, as it is the case of ERP stimuli, can help sustain infants' attention longer than conventional methods of presentation of stimuli. These results are consistent with other studies that have reported the positive use of gaze contingency as an attentional training tool (Forssman et al., 2014; Wass et al., 2011). Additionally, using eye tracking as a gaze control tool during the presentation of ERP stimuli seems to help direct infants' visual attention to the important part of the stimulus. Based on the results obtained in Chapter 2, this is of great importance to maximize the brain response

related to the stimuli and obtain a good quality ERP. Taken together, the new ERP presentation methodology proposed in Chapter 3 based on eye tracking techniques appears to have a positive effect in the final ERP data quality.

As a next step, the objective of this thesis was to further explore the potential benefits of using eye tracking in an infant ERP experiment. Particularly, the aim was to investigate whether the use of eye tracking data during the ERP data editing process would help assess the validity of ERP trials based on the gaze behaviour of an infant (i.e., was the infant's gaze behaviour "good enough" to include a given trial in the grand average ERP?). However, during the study definition process we realized that the data analysis limitations of current infant ERP techniques, explained in the introduction chapter (see page 21), would not allow us to obtain valid conclusions. The approach that we intended to take to evaluate the use of eye tracking during the ERP data editing was to propose an algorithm that, based on the eye tracking data collected during the experiment, could give an objective measure of the validity of a given ERP trial. To validate such an algorithm and evaluate the performance, it would have been necessary to compare it with the current gold standard in ERP editing techniques. Currently, as also explained in the introduction, the most common methodology to select valid trials is mainly based on visual inspection and manual selection. A literature search revealed that these practices have never been assessed and validated. This means that there is no validated gold standard in ERP editing methodology that could be used to compare it with a potential new editing methodology approach based on eye tracking and evaluate its performance. We considered it to be critical that an assessment of the current standard ERP editing methods was conducted, before proposing a new infant ERP editing method that could not be properly validated. The following

chapter will, therefore, try to address the second research question of this thesis: Are current infant ERP editing methods in need for further standardization?

The aim of Chapter 4 was to evaluate the validity of the current infant ERP editing practices. The hypothesis was that the current methods to select valid infant ERP trials, which are mainly based on visual and subjective methods, could have a low level of agreement between infant ERP expert editors. The second hypothesis was that, if the low agreement between ERP editors was confirmed, differences in the ERP trials selected as valid by various ERP editors could lead to differences in the final ERP morphology and could even affect the final results in terms of statistical analyses. The following chapter focuses in the selection and agreement of ERP trials for inclusion and exclusion for further analysis, based on EEG noise. Three ERP human expert editors and one automatic algorithm were chosen to select the valid trials of the same infant ERP dataset. Each human editor used the standard methods for trial selection that they commonly use in their infant ERP studies. Once the level agreement was determined, we further investigated the causes of the level of agreement. We created a statistical model of the EEG characteristics that each of the human editors took into account when selecting the valid ERP trials. The aim of this work was to better understand the differences in current ERP trial selection strategies.

Chapter 4

Understanding the Causes and Consequences of Low Agreement in Infant ERP Editing Practices

Text in preparation for submission

Abstract

During an infant event-related potential (ERP) study, the most common approach used by developmental researchers to analyse infant ERPs is via a trial-by-trial editing process of the data. This process detects EEG channels that contain noise due to movement and other artifacts and rejects trials that are judged to contain an excessive level of noise. This editing process is either carried out manually by experienced researchers, partially assisted by specialized software or completely performed by an automatic artifact detection algorithm. The validity of current infant ERP data editing methods was assessed by analysing the agreement between four editing methods: three expert human editors and an automatic algorithm. The level of agreement found between editors was low in terms of the number of tri-

als included and number of channels interpolated. The agreement between editors had an impact on the ERP morphology as well as in the statistical results of the ERP component under study. The causes of disagreement were also analysed by estimating the EEG characteristics that each of the human editors took into consideration for accepting an ERP trial. We conclude that there is a need for further standardization of infant ERP editing methods.

4.1 Introduction

Event-related potentials (ERPs) allow for the measurement of brain responses related to external stimuli without the need for overt behavioural responses from a participant. This technique can be applied from birth to study typical and atypical development (Nelson & McCleery, 2008). ERPs have been increasingly used over the last two decades to study how infants develop (De Haan, 2007a; Thierry, 2005). There are, however, some limitations when using this technique with developmental populations. One of them is that the automatic processing algorithms typically used to detect artifacts in the adult electroencephalogram (EEG) are often not suitable for infant EEG. To overcome this challenge, the most common approach is to manually edit the EEG on a trial-by-trial basis to select artifact-free data for inclusion in the final dataset (Hoehl & Wahl, 2012). Importantly, it is unknown how this activity alters the makeup of the final dataset, particularly in terms of number of included trials, and replicability of ERP effects and waveform morphologies between individual data editors.

During an ERP experiment there are a number of aspects that need to be taken into account to ensure that the recorded data are of a good quality. First, the EEG signal recorded is sensitive to body and eye movement. In order to obtain artifact-free trials, the participant should be as still as possible during the experiment, as muscle activity and movements usually contaminate or mask brain signal (Luck, 2005). Second, in the case of visual stimuli, it is essential that the participant directs attention towards the stimulus to obtain a brain response to the stimulus. Adults can be instructed to move as little as possible and to pay attention to the stimuli. When the participants tested are infants, these factors, together with their

limited attention span, become a substantial challenge when recording EEG (Hoehl & Wahl, 2012).

The ERP methodology typically requires that the brain response to the stimuli must be recorded a number of times and averaged to obtain a final ERP. The number of artifact-free trials per condition needed for ERP analysis will depend on the signal-to-noise ratio of the component under study. For adult participants this goes from a minimum of 40 trials with a typical number of around 75-100 trials (Picton et al., 2000). The amplitude of infants' electrical brain activity measured on the scalp is usually higher than in adults due to reduced impedances, such as a thinner skull, in addition to larger postsynaptic activity as a result of a larger number of synapses (DeBoer et al., 2007; Thierry, 2005). This, together with the difficulty of getting artifact-free trials in infants, are the main reasons why a minimum of 10 trials per condition is usually reported (DeBoer et al., 2007; e.g., Bakker et al., 2015; Reid et al., 2007; Reynolds & Guy, 2012), although some studies have suggested that using three or five trials per condition results in a larger sample size and a similar ERP morphology, which compensates for the reduced signal-to-noise ratio (Kaduk et al., 2016; Stets & Reid, 2011; Yrttiaho et al., 2014). One of the main reasons for exclusion of a participant from the final sample in an infant ERP study is not reaching the minimum number of artifact-free trials. The attrition rate in infant ERP studies is typically between 50%-75% of the tested infants (Stets et al., 2012).

The current approach used by developmental researchers is to analyse infant ERPs via a trial-by-trial editing process of the data. This process detects and rejects trials that contain artifacts. Causes to reject a trial are, amongst others movement artifacts, muscle artifacts, eye movement artifacts or lack of attention to

the stimuli (Hoehl & Wahl, 2012). During the editing process, an individual trial can be directly accepted or rejected from those that will go on to form the ERP average, or it can be accepted but with the interpolation of channels that contain noise. This editing process is usually carried out manually by experienced researchers but can also be partially assisted by specialized software. It usually comprises two steps: (1) rejection of trials due to lack of attention. This is generally based on the visual inspection of the video recording of the infant's behaviour during the experiment. The researcher makes sure that the infant is visually fixating on the stimulus and therefore the ERP response is present (Hoehl & Wahl, 2012). (2) Detection of EEG channels that contain noise due to movement and other artifacts. In adults, this detection is usually completed utilising automatic algorithms (Luck, 2005). In infants, this practice has been applied in some studies (e.g., Kouider et al., 2013; Kulke et al., 2016a). However, the automatic algorithms for rejecting artifacts are generally created for adult EEG signals, and are not always appropriate for infant EEG amplitudes and artifacts because they are usually too strict for infant data (Hoehl & Wahl, 2012). Therefore, this automatic detection is usually followed or fully replaced by a manual editing process to detect the trials and channels where the EEG signal is contaminated with artifacts (e.g., Jeschonek et al., 2010; Leppänen et al., 2007; Reid et al., 2007; Righi et al., 2014).

One limitation of the current infant ERP editing process is the potential source of variability due to "the human factor," whereby there is variability in terms of judgments made across editors regarding the quality of obtained EEG data. Infant ERP methodology literature usually describes changes in amplitude as a way to identify artifacts in EEG infant data (DeBoer et al., 2007; Hoehl & Wahl, 2012). However, due to the complexity of infant EEG data and the high inter-subject

variability, there are no clear standards in terms of what exact EEG characteristics and what thresholds should be taken into account when editing ERP trials. There are some software packages available for editing infant ERPs that synchronize the EEG data with the video recorded (e.g., NetStation®, Brain Vision Analyzer® or the graphical user interface for infant ERP analysis based on EEGLAB ; Delorme & Makeig, 2004; Kaatiala, Yrttiaho, Forssman, Perdue, & Leppänen, 2013). These tools ease the editing process but they do not automatize it. It is therefore possible that the judgment and training of researchers could have an effect in the number of trials accepted for further analysis.

In the present study we examined the reliability of current infant ERP data editing processes for trial acceptance. To reduce the sources of variability, we focused on the second step in the ERP editing process: the detection of artifacts in EEG channels. The same infant EEG dataset was edited four times: three times by human editors and once by an automatic algorithm. The editors applied methods representative of the current processes for editing infant ERP data. We divided our study in three parts. First, we assessed the level of agreement between the editors in terms of trial acceptance and trial interpolation. Based on the complexity of the infant EEG signal and the potential limitation of the "human factor", we hypothesized a rather low agreement between editing processes on selection of valid trials and trials with channels marked for interpolation. Second, we studied whether the level of agreement found could have any consequences in the final results of an infant ERP study (i.e. statistical analyses on the average ERP waveforms). The aim of the second part was to evaluate the importance of the agreement level found between the editing processes. We computed a grand-averaged ERP for each of the editing processes and analysed whether the differences in these processes had

an impact on the final ERP morphology. Third, we examined the causes of the agreement level obtained. The aim of the third part of the study is to shed light on which EEG variables human editors take into account when selecting clean ERP epochs. We did this by estimating the EEG signal characteristics that each of the three human editors took into consideration for accepting an ERP trial. The questions that we wanted to answer with the last part of the study are: (1) if there is an agreement in the EEG characteristics taken into consideration between the human editors and (2) if these EEG characteristics are similar to the ones used by automatic algorithms.

4.2 Materials and Methods

4.2.1 EEG dataset

EEG data from 10-month-old infants were selected from a pre-existing infant EEG dataset (Monroy et al., 2017). The authors used a visual ERP paradigm to study infants' sensitivity to statistical structure in action sequences. Each ERP trial consisted of a fixation image shown for one second followed by the stimulus, which was a still image, displayed for one second. In total, nine different still images were used as stimuli. EEG data were recorded using the Electrical Geodesic Incorporated (Eugene, Oregon, USA) 124-channel recording system with a sampling rate of 500Hz and Cz as an online reference. A video of each participant was recorded and synchronized with the onset of every stimulus. The raw EEG data were filtered using a band-pass filter from 0.3Hz to 30Hz and segmented into epochs that comprised a 200 ms baseline and 1000 ms after the onset of the stimulus.

The video recordings were visually inspected to identify the ERP trials where the infants had been attentive to the screen in order to control gaze decisions across the editors, thereby removing this variable as a source of potential variation between edited datasets. Nineteen EEG recordings were selected for the present study. In total, an average of 42.47 (SD = 4.4) trials per participant were included in the study. The trials were split into two conditions that differed from the ones used in the original study: condition 1 (M = 17.47) and condition 2 (M = 25). This was done to have a larger number of trials per condition to be edited. An ERP analysis based on the original data editing indicated that there was an Nc component over the frontal area of the scalp in both conditions, with one of the conditions under

study having a more negative Nc than the other condition.

4.2.2 ERP editors

Four editors: three human editors from three universities (Lancaster University, Birkbeck University and Bangor University) and an automatic algorithm edited the same ERP dataset. The three human editors were developmental researchers with substantial experience in editing infant EEG data, with at least three EEG derived papers published per editor. They applied the methodology and criteria used in their labs to edit infant ERP data (see Table 4.1 for details). Two editors used a similar manual approach to edit the data based on trial-by-trial visual inspection. The third editor used a semi-automatic approach to edit the data: an automatic algorithm was applied to detect artifacts followed by a trial-by-trial visual inspection. The automatic algorithm chosen as the fourth editor had been previously utilised to edit infant ERP data (Kouider et al., 2013). The algorithm was implemented in Matlab. It automatically marked channels for interpolation and marked each trial as accepted or rejected based on EEG amplitude levels (for detailed information about the algorithm, see the supplementary material in Kouider et al., 2013).

4.2.3 ERP data editing procedure

The editors' task was to assess the EEG signal for each of the ERP trials and accept or reject each trial for inclusion in the individual ERP average. Additionally, for the trials accepted, the editors had to report any channel that should be interpolated

Editor	*Editing software*	*Editing method*	*Criteria to include a trial*	*Criteria to include a subject*
Editor 1	NetStation®	- Visual Inspection trial by trial	- No more than 10 channels with eye movement or other artifacts detected. - No cluster of 3 or more nearby channels with artifacts detected.	- At least 5 trials per condition
Editor 2	Matlab® (ERPLab)	- Automatic algorithm to detect eye and slow wave artifacts. - Visual inspection trial by trial.	- Not many channels with eye artifacts or slow wave artifacts detected by the algorithm.	- At least 10 trials per condition. - 35% or more of trials accepted.
Editor 3	NetStation®	- Visual inspection trial by trial.	- No eye artifacts detected. - No alpha waves or noise over frontal channels detected. - Less than 20% channels marked for interpolation.	- At least 10 trials per condition. - Not clear drowsiness shown by alpha waves or not extremely fidgety.
Editor 4 (automatic algorithm)	Matlab® (EEGLab)	- Automatic algorithm to detect eye and movement artifacts.	- Less than 35% channels marked for interpolation.	- At least 10 trials per condition.

Table 4.1: Summary of the methodology used by all the editors that participated in the study.

during the ERP data pre-processing. Finally, they decided whether they would include each individual participant in the final ERP sample.

Prior to the start of the ERP data editing, the three human editors were given the same information and guidelines: (1) a general introduction to the ERP experiment, including the type of stimuli and the paradigm. (2) Hypothesis about the Nc effects that were expected over the frontal area of the scalp. (3) A template document with a list containing the trials to be analysed for each participant. The list of trials was blind for condition. (4) A written explanation of the task as described in the beginning of this section. The editors were also given the nineteen ERP data files already filtered and segmented and the corresponding video recordings of the infants with the stimulus onset information embedded.

Each of the four editors returned the following information: (1) a list with the trials accepted and rejected by participant. (2) Only for the accepted trials, a list with the channels, if any, marked for interpolation. (3) The decision, when possible, of inclusion of each participant in the final ERP sample.

Part 1: Agreement between editors

We used the Krippendorff's alpha coefficient (α) for nominal data as a main measure to evaluate the level of agreement between editors. It is a reliability coefficient designed to measure agreement between independent coders and, among other advantages when compared to other reliability methods, can be applied regardless of the number of observers (Hayes & Krippendorff, 2007). The values of α range from -1 to 1, with 1 representing perfect agreement, 0 representing no reliability between

coders and values below 0 representing disagreement that exceeds what can be expected by chance. There is not a minimum acceptable value of α coefficient, but a suggested threshold of $\alpha \geq 0.667$ is considered to indicate that the coded data is reliable for subsequent analyses (Krippendorff, 2004). We used the SPSS macro KALPHA to compute all the agreement values reported (Hayes & Krippendorff, 2007).

The agreement among editors was calculated on: (1) trial assessment (N = 804), (2) participants included in the final sample (N = 19) and (3) channels marked for interpolation on trials accepted by all editors (N = 218). To reduce complexity in the agreement on channels marked for interpolation, we recoded the trials in a binary format: trials with one or more channels marked for interpolation (1) and trials with no channels marked for interpolation (0). As a secondary agreement measure, the percentage agreement on all trials accepted and rejected was calculated. We calculated the agreement among the four editors as well as in groups of three editors to detect any possible editor outlier.

Part 2: ERP data analysis

We pre-processed the original ERP dataset four times to obtain four grand-averages ERP, one for each of the editors. The only difference in the ERP analyses was the selection of trials accepted for further processing and the channels that were interpolated. These were based on the responses obtained from each of the editors. For each of the ERP analyses, the following steps were applied: first, the trials marked as rejected were discarded. Next, a trial-by-trial channel interpolation was

applied to the channels that had been marked for interpolation. ERP trials were re-referenced to the average of all channels and baseline corrected. Finally, the ERP trials were split into conditions 1 and 2. Participants that did not comply with the inclusion criteria given by the editors (Table 4.1; i.e. minimum number of trials per condition) were excluded.

The Nc component was analysed to explore differences due to the data editing process. The most logical analysis given our experimental design would be to run a repeated measures ANOVA with the factors of condition (1, 2) and editors (1, 2, 3, 4) as between-subject variables for a specific location of the scalp. However, this was not possible because different editors accepted different participants (see Table A.1 of the supplementary material) which led to only 6 infants accepted by all the editors. We confirmed that such ANOVA did not yield any significant main or interaction effect for the mean amplitude Nc over the frontal-left area, which was to be expected because of the small sample size. Also, running an ANOVA analysis only with the participants accepted by all editors is probably not representative of the entire sample accepted by each editor and it is therefore likely to be biased.

The only way to understand if different editing processes produced different results was to run four independent statistical analyses on the Nc component, one for each of the four final ERP datasets. By doing this, we examined each final ERP as if they were four different studies studying the same data set. If the differences in the editing process had no or little impact, then the statistical results of the Nc component would be the same for all editors. We selected three clusters of electrodes over the frontal area of the scalp: left (four electrodes, including F3), central (six electrodes, including Fz) and right (four electrodes, including F4; see

Figure 4.2). The clusters were chosen based on the Nc component observed in the original grand average ERP that was calculated prior to this study. We selected two Nc analyses based on visual inspection of the four grand-averages: a mean amplitude analysis within a time window between 300 and 500 ms and a peak latency analysis within a time window between 300 and 600 ms. For each type of analysis and each final ERP, a repeated measures ANOVA was applied with location (left, central, right) and condition (1, 2) as within-subject variables. Also, a paired-sample t-test statistical analysis was applied for each location with the experiment condition (1, 2) as a within-subject variable. SPSS 22® was used to run all the statistical analyses.

Part 3: Estimation of EEG characteristics that affected editors' agreement

EEG characteristics: We identified 22 EEG characteristics that have been used in the literature to characterize an EEG signal and identify noise (Delorme et al., 2007; Hoehl & Wahl, 2012; Inuso, La Foresta, Mammone, & Morabito, 2007; Junghöfer, Elbert, Tucker, & Rockstroh, 2000; Luck, 2005; Nolan et al., 2010). The selection contained both time-domain EEG characteristics and frequency-domain characteristics. Each EEG characteristic was calculated for each of the 804 ERP trials and each of the 124 electrodes. The reduction of characteristics was based on Pearson correlation analyses. For each group of two or more variables with a correlation higher than | 0.7 | only one of the variables was selected. The final set contained 11 EEG characteristics. The initial and final set of EEG characteristics used can be found in the supplementary material, Table A.2.

Statistical model: The main statistical challenge with edited infant EEG data

is that the electrode level quality ratings are not observed, we only observe the trial level decision (i.e., trial accepted or rejected). For that reason and to reduce the complexity of the statistical model, we first eliminated the electrode level information. To do so while losing as little information as possible, we constructed a new one-dimensional EEG characteristic for each trial that was based on all the electrodes. This new one-dimensional EEG characteristic described how unusual the EEG characteristic was during an ERP trial compared to a measure of that same characteristic for trials accepted by all editors. Our one-dimensional EEG characteristics $v \sim$ij were computed as:

$$v_{ij} = \log\left[(\underline{\mathrm{v}}_{ij} - \acute{v})^T \Sigma_v^{-1} (\underline{\mathrm{v}}_{ij} - \acute{v})\right]$$

where $\underline{\mathrm{v}}_{ij}$ are the vectors of each EEG characteristic for each participant i and trial j, and $\acute{v}$ and Σ_v is the average EEG characteristic, calculated as the arithmetic mean across participants of the within-participant empirical mean and covariance of trials accepted by all the editors. In other words, to compute the mean we first averaged the response across accepted trials within each individual, then took the arithmetic mean of those averages across individuals, and similarly for the covariance. To avoid issues of potentially very large EEG characteristics in certain electrodes and trials—which would affect our estimates of the mean and covariance—we replaced extreme observations, defined as those below the lower 2.5% or above the upper 97.5% quantiles respectively, by the 2.5% and 97.5% quantile for that particular property across all individuals. We used the log transform because according to both the AIC and BIC (Gelman, Hwang, & Vehtari, 2014), the fit of our model (see detail below) was better compared with the non-logged measure.

Finally, we further standardized each v ∼ij to allow us to directly compare effect sizes using the estimated regression coefficients. We fitted the following mixed-effect logistic regression model to the resulting data:

$$\log(\tfrac{p_i}{(1-p_i)}) = \tilde{V}_{ij}\beta + \epsilon_i$$

where p_i is the editor's decision for each participant i and trial j, $\tilde{V}_{ij}$ is a vector with the fixed effects containing the $\tilde{v}_{ij}$ for each property, β is a vector of fixed effect sizes, ϵ_i is the random effect at the participant level. We used backward selection, at each step excluding the least significant variable whose estimated p-value was over 0.05 to arrive at the final model. We fitted three models, one for each of the editors, and obtained the significant variables that influenced each editor's decision. We also obtained, for each model, the estimated fixed effects for each significant variable.

4.3 Results

4.3.1 Part 1: Agreement between editors

The first difference between the four editing processes can be observed in the number of participants included by each editor in the final ERP sample: editor 1 included 10 infants (M = 6.8 trials per condition), Editor 2 included 12 participants (M = 14.5 trials per condition), Editor 3 included 7 participants (M = 10.8 trials per condition) and the automatic algorithm, from now on called Editor 4, included 16 participants (M = 18.5 trials per condition; Figure 4.1). This difference is a direct consequence of each editor's inclusion criteria and the number of trials accepted by each editor, which ranged from 32% of accepted trials by Editor 1 to 87% of accepted trials by Editor 4. Regarding interpolation, the percentage of trials with at least one channel marked for interpolation varied from 67% of the accepted trials by Editor 2 to 100% of the accepted trials by Editor 4. Also, the average number of channels marked for interpolation included high variability, from 1.8 channels on average for Editor 2 to 16.05 channels on average for Editor 4 (Figure 4.1).

These differences were confirmed by the Krippendorff's alpha coefficients (Table 4.2). The overall α agreement among editors on trial assessment was 0.275. The overall agreement on participants included in the final ERP sample was within the same range of low reliability data (0.409). Very low agreement—close to chance—was obtained for the trials marked for interpolation (0.061). Small differences of α values were obtained when excluding one editor from the calculations. Only a slightly higher agreement was found among the human editors for all the variables, although in any of the cases it was near the reliability threshold.

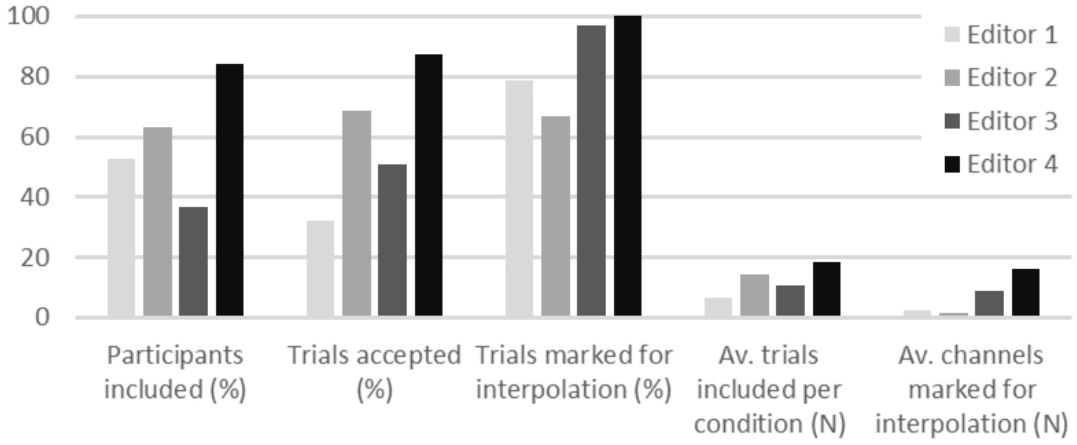


Figure 4.1: Summary of the editing process results for each editor. From left to right: (1) The percentage of participants included in the final ERP sample. (2) The percentage of trials marked as accepted. (2) The percentage of trials marked for interpolation. This percentage was based only on the trials marked as accepted by each editor. (4) The average number of trials per condition that were included in the final ERP sample. This was calculated only from the participants included. (5) The average number of channels marked for interpolation. This average was calculated only from trials that contained at least one channel marked for interpolation.

The percentage agreement also reveals low agreement on trials accepted and rejected. 27.08% of the trials were accepted by all the editors, whereas 10.06% were rejected by all of them (Table 4.2). When observing the percentage agreement results by groups of three editors, it can be noted that Editor 1 substantially affected this result, which reaches 46.21% when not taking Editor 1 into account. For the agreement on trials rejected, Editor 4 has disproportionately affected this result which reaches 26.46% when not taking Editor 4 into account.

Editors	Krippendorff's Alpha			Percentage agreement	
	Trial assessment $\alpha(95\%CI)$	Participants included $\alpha(95\%CI)$	Trials interpolated $\alpha(95\%CI)$	Trials accepted	Trials rejected
All	0.275 (0.085, 0.460)	0.409 (0.224, 0.591)	0.061 (-0.209, 0.318)	27.08%	10.06%
ED1, ED2 & ED3	0.381 (0.200, 0.560)	0.517 (0.310, 0.724)	0.146 (-0.091, 0.366)	27.20%	26.46%
ED1, ED2 & ED4	0.139 (-0.069, 0.338)	0.457 (0.186, 0.690)	0.041 (-0.222, 0.283)	29.57%	10.06%
ED1, ED3 & ED4	0.201 (0.002, 0.389)	0.293 (0.046, 0.540)	0.061 (-0.358, 0.469)	29.07%	12.17%
ED2, ED3 & ED4	0.320 (0.113, 0.533)	0.345 (0.091, 0.600)	-0.132 (-0.436, 0.171)	46.21%	10.19%

Table 4.2: Agreement results among all editors and by groups of three. From left to right: (1) Krippendorff's alpha agreement values and confidence intervals (CI) for trial assessment (accepted, rejected), participants (included, excluded) and number of trials marked for interpolation (one or more channels marked, no channels marked). The agreement on trials interpolated was calculated taking only into account the trials accepted. Alpha confidence intervals at the 95% level were calculated by applying bootstrap analysis of 10.000 samples (Hayes & Krippendorff, 2007). (2) Percentage agreement on trials accepted and rejected where all the editors of each group agreed on the same decision.

4.3.2 Part 2: ERP data analysis

Nc results

Figure 4.2 shows the frontal left, central and right clusters of electrodes of the four grand average ERP waveforms calculated for each editor. The final sample that Editor 2 used for the statistical analyses was 11 participants. One participant, originally accepted by Editor 2, had to be excluded because the averaged ERP contained high amount of eye artifacts and substantially altered the resulting ERP. An Nc component was observed in the four grand averages over the frontal electrodes, but with variability in their amplitude levels as well as in latency to peak and the amplitude difference between conditions (Figure 4.2).

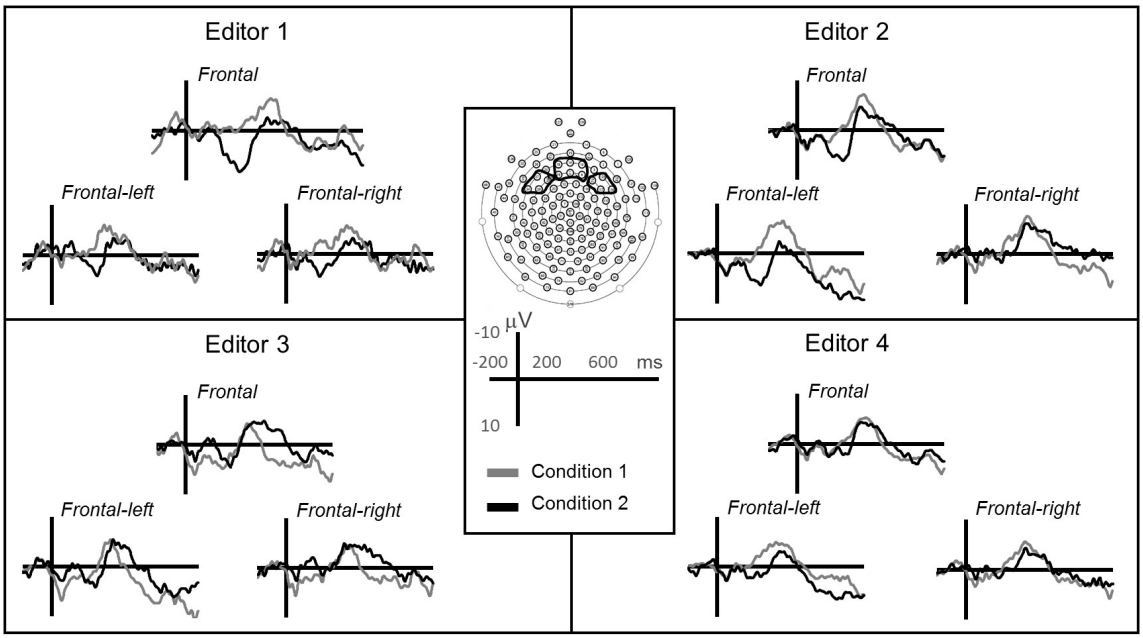


Figure 4.2: Grand average of each editor across the frontal area split by condition. For each editor, the same three clusters of electrodes are displayed: Frontal-left, Frontal and Frontal-right. The variability in amplitude and latency of the ERP components is notable when comparing the grand averages.

Mean amplitude analysis: The repeated measures ANOVA analyses indicated no effects in any of the editors' ERP data for condition or location. However, Editor's 2 ERP data had a trend interaction effect between condition and location (p = 0.09). When each location was examined independently, there was a significant effect for condition in the frontal left cluster as showed by the paired-sample t-test, t(10) = -2.29, p = 0.045; d =0.69, where the amplitude in condition 1 was significantly more negative (M = -5.03) than condition 2 (M = 0.88). None of the other editors had any significant effects for condition in any of the locations.

Peak latency analysis: There were two significant effects in the ERP data via Editor 1 as indicated by the repeated measures ANOVA analysis. First, there was a significant main effect for condition, F(1,9) = 5.37, p = 0.046; $\eta_p^2 = 0.38$. The peak latency of the Nc component in condition 2 was significantly longer (M = 685.73) compared to condition 1 (M = 636.75). Second, there was a significant main effect for location, $F(2, 18) = 6.67, p = 0.007; \eta_p^2 = 0.43$. The peak latency of Nc in the frontal central cluster was significantly longer (M = 690.2 ms) compared to the frontal left cluster (M = 636 ms), t(9) = -3.737, p = 0.005; d =1.18. None of the other editors had any significant effects in the peak latency for condition or location.

4.3.3 Part 3: Estimation of EEG characteristics that affected editors' agreement

Table 4.3 shows the results of the three mixed-effect logistic regression models, one for each human editor. The fixed effects in these models measure the relationship

between the EEG characteristics and the probability of accepting the trial. The significant EEG characteristics can be interpreted as the characteristics that influenced each editor's decision.

Our main finding is that the EEG characteristics that were significant vary across editors. Two characteristics—kurtosis and SNR—were significant in the three models. Thus, it is estimated that the three editors took them into consideration when editing the ERP data. Two characteristics—number of local maxima and power at 0-4Hz—were significant in two of three models. Five characteristics—amplitude range, linear trend, power at 30-60Hz, spectral edge frequency and peak frequency—were significant in only one of the three models. Two characteristics—deviation from the channel mean and power at 8-13Hz—were not significant in any of the three models. All the regression coefficients except one were estimated to be below 1. This can be interpreted as the reduction in odds of accepting a trial given a unit increase in the specific EEG characteristic. That is, the larger an EEG characteristic value, the higher the chance of the trial being rejected by the editor with that EEG characteristic estimated as significant; recalling that a higher EEG characteristic means a more unusual signal. There is only one EEG characteristic for editor 1, peak frequency, with its estimated regression coefficient above 1. The interpretation of this is that this editor tended to accept trials with which the peak frequency was more unusual.

EEG characteristic	Model 1 (Editor 1)	Model 2 (Editor 2)	Model 3 (Editor 3)
Amplitude range	Ns	Ns	0.547 (0.414; 0.722)
Linear trend	Ns	0.720 (0.539; 0.964)	Ns
Deviation from channel mean	Ns	Ns	Ns
Signal-to-noise ratio	0.369 (0.270; 0.502)	0.553 (0.400; 0.763)	0.505 (0.375; 0.680)
Number of local maxima	Ns	0.711 (0.519; 0.975)	0.621 (0.459; 0.839)
Kurtosis	0.416 (0.300; 0.577)	0.666 (0.520; 0.851)	0.569 (0.442; 0.731)
Power at 0-4Hz	0.443 (0.337; 0.582)	0.549 (0.399; 0.756)	Ns
Power at 8-13Hz	Ns	Ns	Ns
Power at 30-60Hz	Ns	Ns	0.713 (0.523; 0.972)
Spectral edge frequency	Ns	0.757 (0.573; 0.999)	Ns
Peak frequency	1.378 (1.113; 1.70)	Ns	Ns

Table 4.3: List of final EEG characteristics included (first column) and EEG characteristics estimated by the three mixed-effect logistic regression models that influenced the decision of each human editor. Significant EEG characteristics of an editor include their estimated regression coefficients and, in brackets, their confidence intervals at 95%. The estimated regression coefficients can be interpreted as odds ratios. Thus, a coefficient below 1 can be interpreted as the reduction in odds of accepting a trial given a unit increase in the associated EEG characteristic. A coefficient above 1 can be interpreted as the increase in odds of accepting a trial given a unit increase in the associated variable. Ns: not significant.

4.4 Discussion

The aim of the present study was to evaluate the reliability of the current infant ERP editing methods. Four editors, three experienced humans and one algorithm, edited the same infant ERP dataset to detect artifact-contaminated ERP trials. We found low agreement between the editing methods in the number of participants included in the final sample, the trials accepted for further analysis and the channels marked for interpolation. As a consequence of the low agreement, the morphology of the grand averages contained variability in the amplitude and latency of the resulting ERP components. This variability was substantial enough to produce inconsistent statistical results in the amplitude and latency difference between conditions of the Nc component.

The three human editors relied on a trial-by-trial visual inspection of the EEG data to identify the channels and trials with unacceptable values of noise in the EEG signal. The criteria described by the editors to accept a trial was similar between them (Table 4.1). They described a focus primarily on detecting physiological artifacts such as eye movements, blinks, body movement, slow waves or alpha waves. However, the results of the statistical models applied to each editor suggest that one of the causes of the low agreement may be the different EEG characteristics taken into account by each of them to evaluate a trial. We believe that the differences in the significant characteristics are explained by the level of complexity of infant ERP data and highlight the lack of standardized methods and definitions to evaluate the noise in infant EEG data.

There are some EEG characteristics included in the model (such as amplitude

range or linear trend) that seemed easier to be measured or evaluated by visual inspection with the consequence that they were more likely to be included as significant characteristics taken into account by the editors. These characteristics were, however, significant only for individual editors. The two EEG characteristics that were significant for the three editors were kurtosis and signal-to-noise ratio. These EEG characteristics are not easily assessed by visual inspection. It is unlikely that the editors assessed the level of noise in the EEG data by consciously using the significant EEG characteristics that the model has predicted for each of them. In our view, the main value of the results of the statistical model is not related to the specific significant EEG characteristics obtained for each editor but, rather, the evidence of the existence of different criteria and thresholds of what is considered noise within infant EEG by the current infant EEG editing methods.

The results of this study indicate that one variable to consider when humans edit complex data is the human error factor, which is inherent to any human process. An example of this is a participant that Editor 2 included in the final sample with a high amount of eye movement artifacts. As explained in the results section, the participant had to be excluded from the final sample of Editor 2 for this reason, which is a common practice in ERP methodology when an individual ERP average is distorted due to artifacts attributable to one participant (Luck, 2005). However, Editor 1 and Editor 4 included the participant in their final samples and the individual ERP averages were not contaminated with eye movements for those editors. This type of human error contributes to the decrease in agreement between editing methods and, importantly, contributes to the differences found in the grand averages.

The agreement between editors was slightly reduced when the automatic algo-

rithm (Editor 4) was included. From the number of trials accepted and channels interpolated by each editor, the human editors and the automatic algorithm seem to have two editing strategies: the human editors tended to reject trials when few EEG channels where contaminated with noise, and usually marked only a small number of channels for interpolation. The automatic algorithm tended to accept trials with a higher amount of contaminated EEG channels but it interpolated them first. Both strategies resulted in grand average with expected ERP morphologies and amplitudes. Since there are no gold standards in infant ERP data editing, none of the editing strategies can be considered better than the others a priori. However, the interpolation of channels creates a new EEG signal based mainly on the nearby channels (Luck, 2005). Therefore, an editing strategy where many channels are interpolated needs to be applied carefully to make sure that the EEG signal is not being altered excessively by creating correlation among channels.

The automatic algorithm used in the present study was initially intended for adult EEG data and was adapted to infant levels of noise (Kouider et al., 2013). The main limitation of utilising this type of algorithm with infant EEG data is that the algorithm is based on general amplitude level rules. However, infant EEG data have a high inter-individual variance of infant EEG and greater delta and theta band activity than adult EEG data (Thierry, 2005). Finding a unique amplitude threshold that is valid for all the infant participants may not be possible. As the results show in terms of trials accepted, the algorithm thresholds used in this study were rather flexible, especially when compared with the human editors' criteria. The higher tolerance of the automatic algorithm suggests that it could have included noisy trials into the averaged ERP. Interestingly, the grand average ERP created by the algorithm does not reflect any effect of potential noise. Rather, the higher number

of trials and participants accepted seem to have had a positive effect in the signal-to-noise ratio of the grand average ERP. It is unknown whether the higher amount of channels interpolated could have influenced the ERP components' morphology and caused false negative results in the algorithm dataset (Editor 4). It is also important to highlight that current algorithms based only on one EEG characteristic (such as the amplitude threshold of the algorithm used in the present study) do not capture the complexity of infant EEG data. The present study reinforces this view, where different EEG characteristics play a role in the detection of noise in infant EEG data for different editors.

The low agreement observed between editors had consequences for the morphology of the final ERPs. There was a notable variability in the amplitude and latency of the ERP components between the four grand averages. Also, what is commonly more important for ERP studies, the grand averages of each editor showed variability in the amplitude and latency between conditions for the Nc component. The direct explanation of this variability between grand averages is the different set of accepted trials and participants that were included by each editor. Each editor selected varying levels of different trials that most likely contained different signal-to-noise ratios: (1) because of higher amount of noise or (2) because of a higher amount of response to the stimuli. Regarding the first reason, the results suggest that the current methods for editing infant data do not have a common threshold of noise to reject a trial, with the consequence that the grand averages are calculated with trials that contain different amounts of noise. Regarding the second reason, the changes in the infant's brain response to the stimulus over time during an ERP experiment could be also playing a role in the variability of the ERP morphology (Stets & Reid, 2011).

It should be pointed out that we are not able to compare the results obtained in this study with any "gold standard" related to data preparation. There are no right or wrong results for any of the editing methods, and therefore we can only speak about the variability that exists between editing methods. Consequently, it is not possible to come to any conclusion as regards of what editing method to choose or how to standardize the editing methods yet. Also, in an infant study the sample size is quite variable but it is uncommon to find significant effects with fewer than ten participants due to effect sizes. In this study, the final sample size of some of the editors would not be considered suitable to extract conclusions. It is quite probable that were this a standard experimental study, the researchers would have continued testing infants until getting a larger sample size could be reported (as indicated in Stets et al., 2012). It should also be pointed out that this study had no strong rationale about the expected Nc differences between conditions. It is possible that the variability found in the current editing methods have less impact on the final ERP morphology and differences between conditions of an ERP study where the expected effect size of the component of interest is greater. Despite this, these issues do not change the conclusions inferred from the results about the existing low agreement between current editing processes in infant ERP studies.

There are many issues that need to be better understood before improved data editing procedures can be adopted by the field. Some examples are whether some infant participants are easier to edit than others because of behaviour or EEG signal (see Appendix A, Table A.1 for a description of the agreement between editors on a participant level, and Appendix A, Table A.3 for a description of the participant level random effects from the models calculated for each editor). Another issue is how the level of noise in the infant EEG evolves or changes during an experiment. Until

these factors are more robustly understood, it is difficult to determine appropriate pathways for methodological improvement. In addition, a more comprehensive exploration should be made regarding how interpolation of channels affects the final ERP and its components. Using statistical models to address these questions could be a valuable way to learn more about infant EEG data and to continue exploring which EEG characteristics and values are important to consider when assessing noise in infant EEG data. We believe that there would be large benefits for the infant ERP community from examining such methodological questions that may help to underpin a more reliable editing procedure in the future.

4.5 Conclusion

This study highlights the low level of agreement of the current infant ERP editing methods. They introduce high levels of variability in the selection of ERP trials as a consequence of noise in the EEG signal. As demonstrated in this study, this variability induced by the editing processes can have an impact on the final ERP morphology and on the amplitude and latencies of ERP components. Our conclusion is that there is a need for a more standardized method when editing infant ERP data. This will need to be preceded by a better understanding of the infant EEG signal and noise characteristics in general, most likely on a component-by-component basis.

Chapter 5

General Discussion

5.1 Introduction to the discussion: Why study infant ERP techniques?

The ERP technique requires a very specific paradigm in terms of the experimental design, data collection and analysis of the data in order to achieve valid results. Methodological guidelines have been exhaustively described in the literature for adult ERP research (e.g., Luck, 2005; Picton et al., 2000). However, as reviewed in Chapter 1 of this thesis, infant ERP methodologies feature a number of limitations that make it unfeasible for infant ERP researchers to follow many of the general adult ERP guidelines, particularly during the data collection and analysis of infant ERP data. The limitations are mainly due to the inability of infants to understand and follow instructions and pay attention to the stimulus presentation for a sufficient period of time. There have been some significant contributions to the infant ERP methodological literature where some challenges have been described (DeBoer et al., 2007; Hoehl & Wahl, 2012; M. H. Johnson et al., 2001). Although in some cases advices on how to overcome certain limitations are suggested in these guidelines, they

also capture the variability that currently exist in the number of strategies that are adopted by infant ERP researchers during data collection and data analysis.

The aim of this thesis was twofold. Firstly, to assess the validity and potential improvements of infant ERP data when including eye tracking during the data collection. Secondly, to study the current variability of infant ERP editing methods and assess their need for standardization. In this chapter, the key findings of the three studies conducted as part of this thesis will be summarized, and their implications for infant ERP methodology will be discussed. Also, a broader analysis about the use of eye tracking in infant ERP research will be made, as well as about infant ERP editing methods. This will be followed by a discussion of the limitations of these findings and directions for future research.

5.2 Summary of findings

5.2.1 The fixation distance to the ERP stimulus influences higher order cognitive responses of an ERP

The aim of Chapter 2 was to study to what degree higher-order cognitive components ERP components are affected by the distance of the fixation point to the center of the main part of a stimulus. Adult participants were presented with ERP stimuli within foveal, parafoveal and peripheral area of their field of view. To achieve this, four ERP conditions were created, with the only difference between them the distance to the center of the main part of the stimulus to the fixation location of the participant: 0°- foveal vision, 4° - parafoveal vision, 8° and 12° - peripheral vision. The ERP stimuli consisted of a series of three images that depicted either a congruent or an incongruent action. The same stimuli had been used before in an ERP study and produced an N400 response for the incongruent action in adults and 9-month-old infants (Reid et al., 2009). We hypothesized that there is a spatial limit away from the stimuli that would cause a decrease in the effect measured in the N400 component in the averaged ERP. This decrease could, in turn, affect final conclusions of an ERP study.

EEG and eye tracking data were recorded simultaneously. Eye tracking data were used to control the stimulus presentation throughout the experiment based on the participants' gaze behaviour. As a secondary aim of the study, the use of eye tracking data as an eye artifact detection tool was also assessed.

The results showed a significant N400 effect in the right parietotemporal elec-

trodes within the 0° visual angle condition. The N400 effect was reduced in amplitude within the parafoveal condition (4°) and not present within the peripheral condition (8° and 12°). The same results were found with the standard artifact rejection technique based on EEG amplitude and the artifact rejection technique based on eye tracking data, suggesting that the standard artifact rejection technique is good enough to discard trials with eye artifacts in adult EEG data. The main conclusion of this study is that, when conducting visual ERP experiments, it is important to ensure that the stimulus is presented within the foveal vision of the participant to maximize ERP effects related to higher order cognitive processes. This is especially crucial when the participants come from difficult populations that are unable to follow instructions, like certain clinical conditions or infants.

5.2.2 Gaze contingency as a tool to sustain infant's attention during an ERP presentation

Chapter 3 described an experiment that evaluated the use of a gaze-contingent stimulus presentation as a way to increase infants' attention during an ERP experiment. Two groups of 10-month-old infants were shown the same ERP stimuli with different presentation strategies. The first group, called the continuous group, was shown the ERP stimuli in a standard manner with the ERP trials presented in a continuous fashion, with the experimenter presenting attention grabbers and making pauses when the infant diminished the attention. The second group, called the gaze-contingent group, was shown the ERP stimuli in a continuous fashion until the infant's gaze was measured to be outside the important part of the stimulus, measured with an eye tracker. At that moment, an attention grabber was automatically

shown to regain infant's attention to the screen, moment when the ERP stimulus presentation was resumed. EEG and eye tracking data were recorded in both groups of infants. The infants' ERP response to the stimuli and their gaze behaviour were compared between the two groups.

The results showed no significant difference in the number of trials shown or accepted for the ERP average between the continuous and the graze-contingent group. In terms of gaze behaviour, the gaze-contingent group showed a difference in gaze behaviour when compared to the continuous group. The gaze-contingent group looked mainly around the fixation image and the stimulus, whereas the continuous group had a more spread looking behaviour along the first 12 degrees from the center of the fixation image. The amount of valid eye tracking data was also significantly different, being higher for the gaze-contingent group. Our hypothesis of this difference in eye tracking data quality would be due to a combination of a better overall infant behaviour or an operator bias due to the gaze contingency procedure. Importantly, we also found a difference in the ERP waveform of the negative central (Nc) component, which is a marker of visual attention (Richards, 2001; Richards et al., 2010) and further indicates that the infants assigned to the gaze-contingent group were more attentive to the stimuli. Our hypothesis of why gaze contingency sustained infants' attention for a longer period of time is that the dynamic attention grabbers, shown at specific moments of the presentation when infants started to diminish their attention to the stimuli, helped them keep their visual attention on the area of the fixation image. These results are also in line with previous studies that have already shown that gaze contingency improves attention in infants (Forssman et al., 2014; Wass et al., 2011).

5.2.3 Low agreement of current ERP editing practices

In the last study described in this thesis, in Chapter 4, the validity of the current infant ERP editing practices was evaluated. Currently, the selection of valid trials in an infant ERP study is commonly either carried out manually by experienced researchers, partially assisted by specialized software or completely performed by an automatic artifact detection algorithm (Hoehl & Wahl, 2012). To study the reliability of the current editing approaches, the agreement between four different editors was analyzed in terms of trial acceptance. We hypothesized that a low level of agreement between editors would affect the final ERP morphology obtained—since different trials would be included in the final sample—and that could, in turn, modify the results of an ERP study in extreme cases. The same infant ERP dataset was analyzed by four editors: three expert human editors and one automatic algorithm.

The level of agreement found between editors was low in terms of the number of trials included and number of channels interpolated. The agreement between editors had an impact on the final ERP morphology. The difference in morphology was large enough to cause disagreement in the final results of the ERP study. Only editor 2 found a significant Nc amplitude difference between conditions, and only editor 1 found a significant Nc latency difference between conditions. These results highlight the impact of the current infant ERP editing techniques disagreement in the final ERP morphology.

As a second part of the current study, the causes of disagreement between human editors were examined. The aim of this part of the study was to understand which EEG variables human editors take into account when selecting clean ERP epochs.

We did this by estimating with statistical models the EEG signal characteristics that each of the three human editors took into consideration for accepting an ERP trial. The three models estimated that the three human editors have used mostly different EEG characteristics to evaluate the validity of an ERP trial. Also, in most of the cases, the EEG characteristics were not the same ones that the human editors reported to have used during the selection of valid trials, nor were these EEG characteristics easily assessed by visual inspection. We believe that this result highlights the evidence of the existence of different criteria and thresholds of what is considered noise within infant EEG by the current infant EEG editing methods and that it is most probably complicated for an expert human editor to be consistent in the criteria used to assess the validity of infant ERP trials. The main conclusion that can be extracted from this last study is that there is a need for further standardization of infant ERP editing methods.

5.3 Eye tracking as a research tool to improve infant ERP techniques

5.3.1 Help sustain infants' attention

Sustaining an infant's attention during an ERP study appears to be one of the fundamental keys to obtain a good quality ERP in terms of number of trials, and morphology of the ERP individual average. When infants are in a sustained attention phase, they are less likely to be distracted from the primary stimulus (Casey & Richards, 1988) or to perform body or eye movements. Body movements and gaze shifts are one of the most frequent causes of EEG artifacts, which contributes to a large extent to the final quality of an infant ERP. Previous research about infant's visual attention has also shown that attention has positive effects on infant's recognition memory (Frick & Richards, 2001) and peripheral stimulus localization (Richards & Hunter, 1997). Furthermore, it has been suggested that the positive effects observed during attention may be caused by an enhancement of areas of the brain that are related to information acquisition or recognition (Richards, 2003). In his study, Richards (2003) classified periods of attention and inattention of infants during an ERP study by measuring variability of infants' heart rate over time. Stimuli presented during periods of attention correlated with a larger Nc component than the periods of inattention, as well as an enhancement of the slow wave, the ERP component under study. Based on these indications, an ERP stimulus presentation methodology that keeps infants' attention during the experiment for as long as possible is most likely one of the possible directions of research that could fundamentally improve current infant ERP techniques.

The results presented in Chapter 3 indicated that gaze-contingent techniques have helped sustain infants' attention to the ERP stimulus presentation when compared to a common method of presenting ERP stimuli. The main difference between the two evaluated methods was the presentation of attention grabbers as soon as the infant started to diminish attention to the stimuli, measured as the visual disengagement from the stimulus presentation. The concept of gaze contingency is not exclusive to eye tracking technology. Gaze control approaches have been used before to sustain infants' attention to an ERP presentation. An ERP experiment that can serve as an example is Leppänen et al. (2007). The approach used by Leppänen and colleagues to present an ERP trial was also based on a gaze control technique but, instead of using an eye tracker to assess the infant's visual attention, one of the experimenters assessed it based on the live monitoring of the infant's behaviour. Also, instead of using video attention grabbers to regain infant's visual attention, another experimenter tapped the screen to attract infant's attention to the screen. The advantages of using a gaze-contingent presentation based on an eye tracker instead of human assessment are (1) that the procedure is more automatized and therefore less prone to human biases and errors and (2) that fewer human resources are needed during the experiment.

Point (1) above requires one important question to be addressed, namely, the pros and cons of having a totally standardized and controlled stimulus presentation methodology when the participants are infants. One of the advantages that eye tracking technology can offer as a gaze control method during an ERP presentation is the consistency of the experimental procedure across infants, e.g., show attention grabbers consistently under the same infant behaviour. This is, in principle, seen as an advantage since it increases the number of variables under control in

an experiment. It is uncertain whether infants benefit from the same method for sustaining their attention or it would be better to have an individualized approach. Another question to take into account when standardizing a stimulus presentation methodology is its predictability. A gaze-contingent presentation method with a very consistent outcome has the potential of being discovered by the infant and of modifying the infant's behaviour (Wang et al., 2012). A sign of learning process in the study reported in Chapter 3 could have been a significant increase during the ERP presentation in the number of times that an infant looked away from the fixation image, which could have been associated with an intention of the infant to make the attention grabber play on the screen. The results of Chapter 3 do not indicate any sign of learning process by the infants of the gaze-contingency mechanism. However, getting further insights about the learning process and if that has an effect in the infant cognitive responses is a crucial question that should be examined in further investigations when proposing any type of ERP stimulus presentation methodology with predictable events that are not related with the research question.

5.3.2 Help improve ERP data quality

One significant question to address in order to improve infant ERP data quality is the inclusion of only those trials where the infant was fixating at the main part of the stimulus in the ERP average. As shown in Chapter 2, including trials where the main part of the stimulus is not within the participant's foveal vision diminishes the cognitive response to the stimulus and can affect the overall quality of the resulting ERP. In this regard, eye tracking is a tool that could improve the main current method of selecting ERP trials in terms of visual attention (i.e., based on

a video recording of the infant's behaviour during the experiment; see Chapter 1 – Introduction, page 12).

The use of eye tracking as a tool to select the valid ERP trials in terms of visual attention has already been assessed in some studies with adult participants (Ahtola et al., 2017; Plöchl et al., 2012). In Chapter 2 of this thesis, eye tracking data was also used as a method for selecting ERP trials where the participant was looking at the right part of the stimulus and there were no blinks or significant eye movements during the trial. This method was compared with the more traditional ERP rejection methods based on EEG amplitude algorithms. Our results showed no significant difference between the two methods in the ERP quality obtained. These results reaffirm that automatic algorithms based on EEG characteristics are, in the case of healthy adult participants, generally good enough to ensure a good ERP quality. A previous study proposed a method that used eye tracking data to detect and correct eye movement artifacts in EEG data (Plöchl et al., 2012). In this case, more advanced methodologies based on Independent Component Analysis (ICA) where used, and eye tracking data was found to increase the eye movement detection sensitivity and help remove ocular artifacts from EEG data. Also, a recent study has investigated the use of eye tracking as an artifact rejection tool designed for selecting valid epochs in infant EEG clinical studies (Ahtola et al., 2017). The authors used the eye movement information to obtain a gaze quality index that was used in the classification of trials with an evoked visual response. The results obtained in that study showed that eye movement information helped detect visual evoked responses to complex stimuli in the EEG of infants. It is unknown, however, whether this positive result found by Ahtola et al. (2017) applied to other infant ERP experimental paradigms would also benefit from using artifact

detection techniques based on eye tracking. The main limitations found are: (1) the infant EEG variability and complexity, which increase substantially the difficulty of applying advanced artifact rejection or correction methods based on automatic techniques and (2) the eye tracking data quality that is usually obtained when participants are infants, i.e., usually less accuracy, precision and track robustness than in adult datasets (Wass, Forssman, & Leppänen, 2014).

5.3.3 Limitations of the combination:

Combining EEG and eye tracking techniques has many potential benefits but also a number of limitations:

A limitation is the potential increase in time to make the experimental setup ready to start acquiring data. The preparation time needed once the infant is in the lab is especially important in infant research due to the limited attention span of infants. Adding eye tracking to an ERP study usually requires ensuring that the infant is within the virtual track box of the eye tracker[1] and the calibration time, which may result in an increase of the experiment preparation time. A way to overcome this limitation would be to validate the approach already taken in a small number of studies, where the authors have tried to reduce the eye tracking setup time by calibrating the eye tracker with an adult before the infant participant arrives (Kulke, Atkinson, & Braddick, 2015; Kulke et al., 2016b).

A second limitation involves eye tracking data quality. Combining EEG and eye tracking techniques increases the risk of exclusion of ERP trials, especially if

[1]The track box of an eye tracker is a virtual three-dimensional space in front of the eye tracker where the eyes of the participant can be detected and tracked.

the results are dependent of the good data quality of both techniques. In the case of gaze-contingent techniques, the correct functioning of the stimulus presentation depends on the quality of the eye tracking data. If the eye tracker cannot track the eyes at a certain moment of the presentation (e.g., the infant is outside of the virtual track box), the gaze control functionality will not be able to work as it is primarily expected. Thus, it is crucial to consider the abovementioned track loss risk when designing any experiment that relies on eye tracking gaze-contingent techniques. Alternatives should be defined that ensure that the ERP trials will be displayed to the infant and the EEG data quality will not be diminished.

Lastly, the hardware and software resources that are currently available to make the EEG and eye tracking combination possible in research are limited to researchers with technical knowledge. A good time synchronization of the EEG and eye tracking data stream is typically crucial to obtain valid results (Dimigen et al., 2011). With the current hardware and software tools, ensuring an accurate time alignment of the data streams requires certain level of technical expertise and a thorough synchronization validation of the setup (Xue, Quan, Li, Yue, & Zhang, 2017). Also, most of the software tools that allow the simultaneous recording of eye tracking and EEG data require a moderate level of programming skills, especially if gaze-contingent techniques are to be employed (e.g., Matlab®, E-Prime® or Python).

Most of the identified limitations have a technical nature and therefore improvements that overcome these limitations can be expected in the near future. Having EEG and eye tracking devices better tailored to infant participants (e.g., EEG systems with faster setup or eye tracking systems with larger track boxes or that are more robust to sudden head movements) and more software tools that facilitate the

combination, will decrease the number and severity of the limitations and will likely open up new research possibilities that rely on eye movements and EEG measurements.

5.3.4 Opportunities for new infant research methodologies.

One of the main themes of this thesis has been to explore the potential benefits and viability of using eye tracking as a tool to improve infants' attention and data quality during an infant ERP experiment. It is important to note that eye tracking data could be used not only as a tool to improve current methodology but as a research method to measure visual attention information of the participant in order to obtain further insights in an infant ERP experiment. In other words, the combination of eye tracking and EEG within infant research could be used to gain a better understanding of the *mechanisms* of cognitive development. An example of the potential of this combination is the study conducted by Kulke et al. (2016b). Neural responses and saccadic reaction times were measured simultaneously to understand the development of attention shifts during infancy. The combination allowed the authors to show a neural lateralization response with age that was associated with improving ability to shift visual attention. In line with the previous example, eye tracking could be used in infant ERP studies to, for instance, classify ERP trials based on the infant's visual behaviour (e.g., saccade onset, duration or amplitude).

Gaze-contingent techniques during an EEG study could be used not only as a method to sustain infant's attention but as a method to obtain information about cognitive processes (i.e., novel paradigms where the stimulus presentation process

is contingent on the infant visual behaviour). They could potentially open new methods of obtaining further insights into human cognition. For instance, the studies that show how gaze-contingent techniques can help control attention could benefit from the addition of EEG as a method to understand better the neural mechanisms of attention in such situations (Forssman et al., 2014; Wass et al., 2014, 2011).

In addition to eye movement data, an eye tracker usually provides pupil size data. Pupil dilation is closely linked with the sympathetic and parasympathetic systems and it has been broadly used in the literature with infants and adults to study cognitive and emotional processes (for a review, see Laeng & Endestad, 2012; Sirois & Brisson, 2014; Hepach & Westermann, 2016). The combination of ERP methodology and pupil dilation could provide new evidence for neural mechanisms associated with specific functions and their development. This combination has been already explored with adults, where a correlation was found between the amplitude of the N400 ERP component and pupil size change (Kuipers & Thierry, 2011). The same authors have further explored the relation between pupil size and cognitive process in toddlers. Their results suggest similar links between pupil dilation and cognitive processes associated with language acquisition (Kuipers & Thierry, 2013).

Finally, there is an increasing interest in the study of cognitive processes using EFRP methodologies in adults (see Chapter 1 – Introduction, page 17). To the knowledge of the author, EFRPs have not to date been applied to infant participants. This is probably due to the eye tracking high temporal resolution and data quality that is required in order to measure valid EFRP signals. However, allowing the participant to move their eyes freely during the trial presentation would be especially advantageous for a difficult experimental group such as infants. Paradigms that

would allow more natural behaviour of the infants, while measuring event-related cognitive responses, would open a totally new window of research possibilities.

5.4 Need for standardization of infant ERP editing practices

5.4.1 Current editing practices in infant ERP research

Chapter 4 highlighted the variability that currently exists in infant ERP editing practices, and the consequences for the ERP morphologies and results that this variability can have. The question of why ERP editing techniques have not been standardized to date may have diverse explanations. The most likely is that ERP methodologies started to be applied in healthy adults and this has been the wider and most studied population group. As a consequence, most of the signal processing techniques to detect and remove noise from the EEG signal have been studied and developed based on adult EEG signals (e.g., Croft & Barry, 2000; Delorme et al., 2007; Luck, 2005; Vos et al., 2010). Developmental researchers are aware that these EEG processing tools, that are mainly based on automatic algorithms, are not directly applicable to infant EEG due to its higher complexity (Thierry, 2005). The relatively small and new infant EEG field of research might be the reason why not many scientific resources have been dedicated to study the most appropriate and reliable method to edit infant EEG data. Thus, the main approach taken by the developmental research community to increase the reliability of the EEG artifact detection has been a manual edition approach to remove the trials with excessive noise (Hoehl & Wahl, 2012).

In Hoehl and Wahl (2012), the following is pointed out: *"One major problem of manual trial rejection (or inclusion) is that criteria for rejection may be very*

subjective and vary from researcher to researcher". Indeed, the differences found between human ERP editors in the study presented in Chapter 4 are believed to be due mainly to a subjective rejection criteria. The general rejection criteria described by the 3 human editors was very similar (see Chapter 4, table 4.1). The rejection criteria were mainly based on the detection of the characteristic patterns that artifacts create on the EEG signal (e.g., large EEG amplitudes due to electrode or cable movements, large waves on frontal and central electrodes due to alpha waves or high-frequency noise due to muscle activity). These patterns are to some extent described in the literature (Hoehl & Wahl, 2012; Luck, 2005). However, one of the problems seems to be to visually detect them and decide with an objective criterion when they contain unacceptable levels of noise.

Although the general aspects of the current infant ERP editing techniques are very much in agreement within the developmental research community as explained in the previous paragraph, each lab or even each individual researcher seems to have slightly different thresholds in some of the analysis variables. The variables with different criterion found in Chapter 4 of this thesis where: minimum number of trials to include an infant in the final sample and per condition, and maximum number of channels that can be interpolated in total and in the same area. The minimum number of trials to include an infant in the final ERP could be one of the key factors to ensure a good quality ERP. To date, the most common threshold used in infant literature has been a minimum of 10 trials per condition (for references see Chapter 1, page 13). There are other studies that have used a lower minimum of trials per condition, such as 3 or 5 trials (for references see Chapter 1, page 14). To the knowledge of the author, these thresholds do not have any scientific rationale and seem to have been chosen arbitrarily through a trial and error method. Importantly,

they seem more like a standard threshold that each researcher uses rather than adapted to each particular type of ERP component under study or experiment type—as recommended by the guidelines for human ERP studies (Picton et al., 2000). It is unknown to what extent each of these variables modify the final ERP waveform but, based on the results presented in this thesis, it is likely that they contribute a variable amount ranging from a small to a great extent.

Two statistic measurements are used to validate a method that relies on human judgements and consider it as reliable and reproducible: (1) consistency of the method between different raters (inter-rater reliability) and (2) consistency of the method within the same rater (intra-rater reliability; Mchugh, 2012). Reliability is simpler to achieve when the method is carried out by an algorithm that follows certain rules. When the method relies on human decisions, it becomes crucial to assess the reliability of the method to ensure valid results (Kirwan, 1992). In the EEG field, several agreement studies have been conducted to assess the validity of different human methods, especially in clinical applications (e.g., Benbadis et al., 2009; Gaspard, Hirsch, LaRoche, Hahn, & Brandon, 2014; Nordli, Bazil, Scheuer, & Pedley, 1997). In the ERP research field, there are also some studies that aim to assess the reliability of raters. An example is Manresa, Arguissain, Redondo, Morch, and Andersen (2015). In this study, the agreement between manual and automatic methods for single-trial detection and estimation of features from ERP generated by electrical stimulation was assessed. The authors found large differences between detection and estimation of quantitative EEG features among methods and concluded that the method selected for the selection and estimation should be chosen carefully as it can lead to considerably different results. The authors also highlighted that several factors that can modify the signal-to-noise ratio of the EEG signal (e.g.,

the intensity of the stimulation in the case of this study) can play a role in the size of these differences. These results and conclusions are very much in agreement with the ones found in Chapter 4, and they stress the importance of finding a reliable editing method for infant ERP that is consistent among infant researchers in order to avoid as much as possible any influence on the final ERP morphology and the results due to the editing process.

There are some actions that could be taken as a first step towards an infant ERP editing technique standardization. Firstly, it would be beneficial if there was a more detailed description of the criteria used to edit the infant ERP data in the methodology section of any published study. This would emphasize the need of being consistent with the method described (intra-rater agreement) and probably would help the infant ERP research community to acknowledge the lack of common methods and the need for a standardization (inter-rater agreement). And secondly, if the visual inspection of the EEG data to select valid trials is the method used, it would be good practice to have at least two editors reviewing the EEG data to ensure an agreement between editors. This may well be a time-consuming task which may impede this in becoming regular practice for every participant for all ERP studies. However, there could be other ways that are less time consuming that could increase the changes of improving the agreement or, at least, make editors more aware of disagreement. Hoehl and Wahl (2012) recommend that editors learn how to edit infant ERP data by editing an already analyzed dataset and compare the results with those obtained by experienced editors. An add-on to this good practice would be to do the same for every ERP experiment but only for a few randomly chosen ERP trials or infants. This would help to ensure consistency of the ERP editing techniques at the very least within the researchers working at the same lab.

Nonetheless, the infant ERP research field would benefit from working towards an standardization of the ERP editing practices that would ensure obtaining reliable and reproducible results across specific individuals and across labs.

5.4.2 Could automatic algorithms be the solution to standardize infant ERP editing practices?

There are several limitations to the manual editing of infant ERP data where automatic algorithms may well improve current practice. As explained in section 5.4.1, the main limitation of manual editing practices is the variability that human subjective criteria introduce in trial selection process. Also, manual editing is time consuming for the researcher as it requires a trial-by-trial visual inspection. Finally, the detection of EEG artifacts by visual inspection is limited by temporal characteristics of the EEG signal, which are mainly based on amplitude changes of the signal. Automatic algorithms would be a way to improve the three main disadvantages of manual editing practices. First, they ensure a reliable and reproducible way of detecting artifacts in the EEG data as the mathematical rules applied would be the same across participants and ERP studies. Second, it is likely that the editing time would substantially decrease. This would depend on the level of automatization of the algorithm and the need for a visual inspection after applying the algorithm to confirm the results. For example, when an automatic algorithm is applied currently to edit infant ERP data, it is common that it is followed by a manual visual inspection to validate the algorithm results (Hoehl, Wiese, & Striano, 2008; Kaduk et al., 2016). Third, automatic algorithms can apply complex mathematical transformations of the EEG signal that may be able to help the detection

of EEG characteristics. An example is Independent Component Analysis (ICA) algorithms that separate independent sources that are linearly mixed in several EEG channels (e.g., oculomotor artifacts; see Delorme et al., 2007) or wavelet analysis that uses joint time-frequency features to extract information from a signal (see Unser & Aldroubi, 1996).

Automatic algorithms based on advanced mathematical transformations are widely applied to many biomedical signals, both in clinical and research applications. In almost all cases, these algorithms have been designed for adult EEG signals. The type, amount and magnitude of the artifacts that most of the algorithms are designed for are very different from the artifacts present in an infant EEG signal. There have been some attempts to apply some of these techniques to infant EEG signals, such as ICA. Some infant EEG studies where ICA techniques were used to detect artifacts—predominantly, oculomotor artifacts—can be found in the literature (e.g., Marshall, Young, & Meltzoff, 2011; Orekhova, Stroganova, Posikera, & Elam, 2006; Saby, Marshall, & Meltzoff, 2012). However, ICA techniques are not appropriate when non-stereotyped artifacts (e.g., head movement and electrode movements) are present in the EEG data. ICA decompositions may even be compromised when such artifacts are contained in the data (Delorme et al., 2007). In their study, Delorme et al. (2007) highlights the importance of visually identifying and discarding periods with such artifacts from the data before running ICA . In the case of infant EEG data, abrupt moves that induce electrode artifacts in the data are very common, which may in turn make the application of ICA techniques to infant data ineffective.

Only one study has been found in the literature that studied and compared dif-

ferent artifact detection methods for infant EEG data applied to the extraction of ERP signals (Fujioka, Mourad, He, & Trainor, 2011). The authors synthetized an ERP signal into a real 4-month-old infant EEG dataset and compared three artifact removal techniques: conventional trial rejection based on visual inspection, the Artifact Blocking technique (AB; Mourad, Reilly, De Bruin, Hasey, & MacCrimmon, 2007) and the Independent Channel Rejection technique (ICR; C. He, Hotson, & Trainor, 2007). Both AB and ICR techniques were found to be significantly better in identifying the noise and isolating the ERP signal than conventional trial rejection. A thing to note is that the study simulated an auditory ERP signal, which means that the synthetized ERP signal contained more than 500 trials. This is a common number of trials in an auditory ERP experiment. Also, the infants watched a silent movie with a puppet show during the EEG recording. This is probably a more engaging stimulus than repeated images, which it is typically the case during visual ERP stimuli. It is unknown whether the AB and ICR techniques are equally valid with EEG recordings containing more noise as it is likely to be the case of a visual ERP experiment and with participants older than 4 months.

Infant EEG data are, in general terms, significantly more complex than EEG data recorded from participants that can follow instructions. Some of the variables that contribute to the complexity in the EEG signal are likely to be due to the specific infant's behaviour and they are likely to vary between different participants (Thierry, 2005). Conventional automatic artifact rejection methods based on simple characteristics—like the amplitude level algorithm applied in Chapter 4—might not be able to reflect this complexity. An automatic algorithm suited for infant EEG data would need to take into account this complexity and be capable of adapting to the existing signal variability. An example of an algorithm that tries to capture this

variability is described in Kulke et al. (2016b). In this study, the EEG noisy data was defined based on individual infant's EEG data, following the median absolute deviation procedure (Hampel, 1974) . To create and validate an automatic algorithm that aimed to capture the infant's EEG complexity, it would be necessary to first better understand the variables that generate the complexity in the infant EEG signal. In this sense, the developmental field would benefit from collaborating with researchers from more technical fields, like mathematicians or statisticians. They could investigate and apply more advanced methods to help understand the infant EEG signal and find a suitable automatic algorithm for noise rejection or correction. Due to the magnitude of this task, currently a fully automatic noise detection algorithm seems to be a long-term solution. Semi-automatic methods that provide researchers objective measures about the level of noise of the EEG signal and help them make a more informed decision about valid trials could be a medium-term solution as a first step towards the standardization of the editing methods.

5.5 Wider methodological implications

One of the goals of improving infant ERP methodologies is to maximize the number of trials and infants that are included in the final sample of an ERP study. The meta-analysis conducted by Stets et al. (2012) investigating the factors underlying attrition rates indicated that the main factor that influences attrition rate in infant ERP studies is the nature of the stimuli. In experiments with visual stimuli, a combination of audio and visual stimuli had significantly lower attrition rates than experiments with purely visual stimuli. The authors' explanation of this result included a direct relation of infants' attrition rate with the ability of the stimulus presentation to maintain infants' attention to the ERP stimuli. Studies of selective attention in infants that proposed and studied the intersensory redundancy hypothesis are also in line with these results (Bahrick & Lickliter, 2000; Bahrick, Walker, & Neisser, 1981). These studies found evidences of higher attentional resources towards a visual stimulus which was accompanied by a synchronous audio than towards a silent visual stimulus. On the other hand, Stets et al. (2013) supported the direct relation between attention and attrition rate not by including audio during the presentation of stimulus but by increasing the complexity of the visual stimuli by means of mixing visual stimuli from three different ERP experiments in the same session—as explained in the introduction chapter (see page 23)—although in that study full replications of prior studies did not occur for some of the experiments.

These studies suggest that stimulus sets that include audiovisual or more complex visual elements are more engaging for infant participants, which could, in turn, help create longer periods of attention. While an infant is engaged with the stimulus presentation, the infant is less likely to move and create artifacts in the EEG signal.

Artifact-free EEG data from a steady participant increases the EEG data quality and therefore it is likely that the number of valid trials for the final sample increases as well. Results of Chapter 3 support the theory of the increase in attention when the stimulus presentation combined visual with audiovisual elements—as the dynamic attention grabbers used consistently based on infants' gaze behaviour. However, the increase in attention did not lead to a decrease in attrition rate when compared to standard ERP stimulus presentation that mainly used visual stimuli. The reason why attention did not lead to more trials included in the final sample might be the increased length of the trials due to the attention grabbers that were introduced in the gaze-contingent condition. Stets et al. (2012) also highlighted in their meta-analysis study the negative impact on the attrition rate that dynamic stimuli can produce due to the increase in trial length.

Overall, these results suggest that just the use of a more engaging stimulus presentation that increases the attention is not sufficiently robust to decrease attrition rate on its own. It is likely that the sum of factors such as the nature of the stimuli, the length of the trials, and other factors still unmeasured or unknown that are responsible for the final attrition rate. The unknown factors may be related with the lab environment, the experimenter's interaction with the infant and parent or with the specific conditions of the infant at the moment of data collection such as time since last fed, sleep, age, or general health situation. Further research that helps to understand how and at what level these and other variables influence the attrition rate would be highly beneficial for the improvement of data collection methodologies in infant ERP. The improvements could go from practical aspects such as booking the appointments with parents based on the sleeping or feeding patterns of the infants—if these factors happen to have an effect in the likelihood of an infant being

included in the final sample—, to more theoretical aspects such as guidelines to help design a more engaging stimulus presentation with the appropriate type of stimuli, length and breaks to maximize infants' attention span.

Infant ERP data collection is not the only part of an infant ERP study that determines attrition rate. Infant ERP editing practices can also have a direct impact on the number of trials that are valid and the number of infants that are included in the final sample. A clear example of how the editing practices influence the attrition rate of a study has been shown in Chapter 4. Each of the four editors that assessed the same infant ERP dataset included a different number of infants. The use of different editing methodologies by each of the editors created a variance in the attrition rate that ranged from the 16% for the automatic algorithm to 63% for one of the human editors. As seen in Chapter 4, the difference in the editing practices was not only observed in the agreement of the number of infants and trials included in the final sample, but also in the final morphology of the grand average and ERP component under study that each of the editing methods created. Stets and Reid (2011) also found similar results when they reanalyzed an EEG dataset and changed the criteria of the minimum number of trials to only the first three trials. With this practice, they were able to include more infants into the final sample—the attrition rate decreased—and they also found contradicting results in the morphology of the ERP component under study when different amount of trials and infants where included in the final sample.

Attrition rate may have not only practical consequences, which are mainly related to the time spent by the researcher in collecting and analyzing infant data. Attrition rate may also have crucial consequences in terms of the ERP results when trying to

get insights into cognitive development at a specific age. In most cases, the exclusion of infants from the final sample is due to an insufficient number of valid trials per condition. The practice of excluding infants that do not contribute a sufficient number of valid trials is important in order to ensure a good quality ERP but, it is unknown to what extent there might be a relation between the infants excluded and other variables that may influence results, such as their developmental stage. For example, there could be a bias in the final ERP results if only the infants that are at a more advanced developmental stage are able to sustain attention during the ERP paradigm and therefore contribute to the final sample. In general, a high attrition rate could be a sign of not having a representative sample of the population under study. The potential bias that attrition might cause in ERP developmental research would need further investigation and it raises an essential question: Is it more important to work towards infant ERP methodologies that prioritize a decrease in the attrition rate or towards ERP methodologies that prioritize that only infants that provided good enough EEG signal quality are included in the final sample? This question might be able to be answered when there is more information about how attrition rate may influence and bias the final ERP results.

5.6 Limitations and directions of future research

5.6.1 Limitations of the research

This thesis has focused on the study of infant ERP methodology with the aim of expanding upon the understanding of how data collection and editing practices can be improved to obtain better data quality and more robust results. Whilst the research conducted in this thesis has contributed with some valuable insights to the infant ERP methodology field, there are also a number of limitations that need to be acknowledged and taken into consideration when examining them.

A general limitation of the present thesis is that the conclusions of the methodological implications are drawn based on a specific developmental age. The EEG data were recorded from 10-month-old infants and was used for both experiments presented in Chapters 3 and 4. It is unknown whether or not the results and methodological implications would have been the same if another developmental age were tested. The main difference that could have arisen from testing a different age would probably have been the infant behaviour and engagement with the stimulus presentation (Chapter 3). A difference in the engagement would have probably led to a difference in the EEG artifacts due to movements and therefore a difference in the validity of the ERP (Chapter 4). However, infant ERPs are a common technique used with infants from 3 to 12 months of age. This age range is considered to be the least problematic in terms of attention span and with a comparable set of behaviours during the experiment (Hoehl & Wahl, 2012). Also, Stets et al. (2012) did not find a significant relationship between age group tested and the attrition rate reported in a meta-analysis that studied the causes of attrition rate.

A limitation is derived from the wide scope of the first research question: Can eye tracking help improve infant ERP and ERP data quality? The research question asked was too broad to be answered with certainty with data from one infant study that combined EEG and eye tracking. The research conducted in Chapter 2 and 3 shows evidence that partially answered the research question, but also generated further questions rather than specific conclusions. The main questions that arise are: to what extent can we directly relate visual attention to eye movements? Can gaze contingency or attention grabbers affect the cognitive response of the stimulus? Is it better for the ERP data quality to control the stimulus presentation in the same way for every infant—e.g., with gaze-contingent techniques—to avoid biases or to adapt the presentation of stimulus to each infant's need and mood? These questions are not only relevant when using eye tracking as a method to improve infant ERP methodologies, they are more general questions, and the answers would help improve infant ERP data collection methodologies.

An eye tracker measures with high accuracy and precision where the fovea of the participant is located on a screen. When the fovea is relatively statically pointing at a location, it is known as fixation. The non-direct link between fixation and spatial attention has been a limitation both in Chapter 2 and 3. Studies of attention have shown that the ERP response is modulated by spatial attention. In their review, Luck et al. (2000) provided a compilation of several studies that show evidence for how the attentional components of an ERP are modified when the participant is instructed to allocate the attention in a location outside the fovea or fixation point. The experimental design of Chapter 2 intended to exclude the variable of spatial attention as the main purpose of the study was to investigate how a physical variable, fixation distance to the stimulus, affected late components. This limitation

was potentially compensated for by instructing participants to always attend to the stimulus, irrespective of its location, depending on the trial condition. However, there was no mechanism to control that the attention was indeed allocated to the stimulus on every condition. That creates uncertainties about the cause of N400 depletion being only due to distance to the stimulus or also due to attention allocation. This uncertainty applies as well in Chapter 3, where eye tracking data indicated that infants assigned to the gaze-contingent condition significantly fixated longer at the area where the main part of the stimulus was located. It cannot be known, however, what level of attention was allocated to the location where the fovea was pointing.

Similarly, the Nc results of Chapter 3 indicate that the level of attention to the stimuli was increased when using gaze-contingent techniques. Several studies have shown how the level of attention can be indirectly measured by the heart rate variability as periods of sustained attention are accompanied by a deceleration of the heart rate (Casey & Richards, 1988; Richards, 2001, 2003; Richards & Hunter, 1997; Richards et al., 2010). Richards (2003) also showed a correlation between the level of attention, measured by the heart rate variability, and the amplitude of the Nc component. This provides evidence that the Nc results obtained in Chapter 3 can be associated with an increase of the level of attention to the stimulus presentation of the infants assigned to the gaze-contingent condition. Nevertheless, the measurement of the heart rate could have helped increase the certainty of the results and could have provided more information to classify periods of sustained attention and termination of attention.

The experimental design of the study conducted in Chapter 2, where participants needed to follow instructions to fixate at a fixation cross and allocate the attention to

the stimulus, limited the possibility of using infants as participants. The conclusions of Chapter 2, which served for the experimental design of Chapter 3, were based on adults due to this limitation. The assumption was that the results of Chapter 2 drawn with adult participants apply in the same way to the 10-month-old infant participants of the experiment conducted in Chapter 3. It is currently conjectured that the visual system develops primarily during the first postnatal year, including abrupt changes in the development of visual acuity during the first six postnatal months, both in central vision and peripheral vision (Allen, Tyler, & Norcia, 1996; Sireteanu, Kellerer, & Boergen, 1983; Sireteanu, Fronius, & Constantinescu, 1994). Therefore, the assumption taken in the present research work can be considered reasonable and the potential errors in the experimental design that may have arisen from differences in visual acuity between infancy to adulthood are unlikely to modify the results or conclusions of the present work.

The main limitation of Chapter 4 was the lack of a method to measure which editing technique performed better. Having information about the best performing editing technique could have not only highlighted the current inconsistency between editing methods as the present results do, but could have provided further insights about how to improve infant ERP editing methods. Signal-to-noise ratio would probably be the best method known to date to measure the proportion of signal and noise that an ERP signal has as a consequence of the editing method used. Also, one of the goals of the experimental design of the study conducted in Chapter 4 was to give the editors dataset as realistic as possible, and so an existing infant EEG dataset from a previous infant ERP experiment was used. The estimation of signal-to-noise ratio with real ERP data is a real challenge that is difficult to overcome (Spencer, 2005). In real ERP data, there are many factors present that can contribute to

noise, and the amount of cognitive response to the stimulus (signal) is not known and can vary between trials, subjects and environmental conditions. Some authors have proposed to calculate signal-to-noise ratio in real ERP data as the division of the individual ERP amplitude by the standard deviation of the ERP baseline (e.g., Hu, Mouraux, Hu, & Iannetti, 2010; Spencer, 2005). This approach to estimate signal-to-noise ratio was applied to the results of Chapter 4 to the four editors but led to no significant results. There are two possible explanations of the lack of significant results. One is that the four editing methods included similar amounts of signal and noise and therefore the estimation of the signal-to-noise ratio was correct. This is unlikely since the four editing methods clearly changed the grand average ERP and led to different results of the component of interest. The second possibility is that this approach to estimate the signal-to-noise ratio is not sensitive enough to capture the difference in ERP signal quality that each of the editors caused. The hypothesis for the reason why the signal-to-noise ratio estimation was not significantly different between editors is currently a combination between the two aforementioned possibilities.

One way to overcome this limitation would be to use synthetic ERP signals instead of real ERP signals. The advantage would be that the amount of noise and signal would then be known and the estimation of the signal-to-noise ratio would be straightforward. There are many examples in the literature that have used this method, mainly to evaluate artifact detection algorithms. They are usually applied to adult data (e.g., Kayser & Tenke, 2003; Romero, Mañanas, & Barbanoj, 2008; Yeung, Bogacz, Holroyd, & Cohen, 2004). At least one study has used this approach with infant data (Fujioka et al., 2011). The authors embedded an auditory ERP signal in a real infant EEG recordings to compare artifact rejection methods in infant

data. It would be necessary to assess the ecological validity of simulated infant ERP when studying the editing methods, especially with human editors, but it may be a reliable option to have a better estimation of the quality of the editing method under study.

5.6.2 Future directions

There are some specific questions and directions of future research arising from the work of this thesis. Their main objective being the further understanding the infant EEG signal, as well as infants' behaviour during an ERP study. These questions' ultimate aim would be to improve current infant ERP techniques.

As described in section 5.6.1, one of the possible directions of future research would be to apply the same paradigms used in the experiments conducted as part of this thesis to other age groups. Comparing the results of different age groups would help get insights into what degree the results of this work and their implications with infant ERP methodology are the same for all developmental ages or, on the contrary, whether infant ERP practices would benefit from their adaptation to different developmental ages. In terms of data collection methods, this becomes a rather complex research question that may need to include additional variables besides age to obtain a full picture of what influences the general behaviour and attention of an infant during an ERP experiment. Variables such as time since last sleep or feeding, parental leave status of their progenitors or teeth could modify the behaviour and attention span of infants. This new knowledge would contribute to the development of new data collection methodologies. Specifically, knowing if it is

more beneficial for the final ERP data quality and attrition rate to apply the same stimulus presentation strategy to every infant or it is better to modulate the stimulus presentation strategy to the emotional and physical status of each infant. The use of gaze-contingent techniques such as the one used in the present work could also benefit from this knowledge since the level of automatic control could be adapted in the paradigm once this new knowledge is acquired.

In terms of infant data editing, this thesis has contributed to highlight the variability that current editing methods have in visual infant ERP studies. There are some future directions of research that could be taken to work towards the standardization that is required in this field. One direction towards standardization would be to look into more advanced, non-linear automatic algorithms that have been already used in other EEG applications, and that can capture the complexity of infant EEG. Some examples may be neural networks (e.g., Srinivasan, Eswaran, & Sriraam, 2007; Subasi & Ercelebi, 2005), adaptive filters (e.g., Correa, Laciar, Patiño, & Valentinuzzi, 2007; P. He, Wilson, & Russell, 2004) or wavelet analysis (e.g., Adeli, Ghosh-Dastidar, & Dadmehr, 2007; Krishnaveni, Jayaraman, Aravind, Hariharasudhan, & Ramadoss, 2006). This line of research would require the collaboration of developmental researchers with experts in biomedical signal processing.

Increasing the knowledge of the nature of the infant EEG signal would be essential before starting to develop any automatic algorithm. The mathematical model applied in Chapter 4 to understand the causes of disagreement showed that the editors took different EEG characteristics into account. The characteristics that the model predicted were not the same ones that the editors claimed to have used in most of the cases. This result emphasizes the need for a better understating of infant

EEG data. A good characterization of infant EEG would facilitate the classification of signal and noise in an infant EEG and therefore have a better judgment when selecting ERP trials with valid levels of noise. This could improve the editing methods for human editors and could be used for the development of automatic algorithms.

Lastly, another important line of research towards an improvement of the current editing methods would be to investigate whether or not the cognitive response of the infant is modulated during an ERP session and what is the minimum number of valid trials that is needed for each ERP component to obtain an appropriate signal-to-noise ratio and robust results. Picton et al. (2000) provides some guidelines to calculate the minimum number of trials to achieve certain signal-to-noise ratio. However, the differences between adult and infant EEG signal, the differences in cognitive processes, and the complexity of the infant EEG signal makes necessary to further investigate this relation in infant EEG. Stets and Reid (2011) already started this line of research and obtained interesting results about the differences in the ERP waveform with different amount of average trials from different parts of an experimental session. Increasing this knowledge would definitely help develop new infant ERP editing techniques that could maximize the trials where the cognitive response is present. An example of how this knowledge could be applied to infant ERP editing methodologies would be to use a weighted ERP average such as the one suggested by Davila and Mobin (1992).

5.7 Conclusions

The research conducted as part of this thesis seeks to understand how current infant ERP techniques can be improved to overcome some of the data quality challenges that are associated with the recording and analysis of EEG data on developmental populations. The first part of the thesis focused on the ERP data collection phase of an ERP experiment. It explored the possible improvement of ERP methods when using eye tracking techniques during the ERP data collection. The results suggest that eye tracking can potentially help infant ERP techniques to improve the ERP data quality. However, further development of hardware and software tools that allow a more efficient combination of EEG and eye tracking techniques would be needed. That would also facilitate the creation of new lines of research that use this combination in infant populations. The second part of the thesis focused on the evaluation of the current ERP data editing methods. The results show the low agreement and low reliability of the current infant ERP editing methods. Overall, the results of this thesis highlight that there are fundamental issues in current infant ERP techniques that should be addressed. This thesis supports the need of further research towards more robust and standardized infant ERP techniques.

References

Adeli, H., Ghosh-Dastidar, S., & Dadmehr, N. (2007). A wavelet-chaos methodology for analysis of EEGs and EEG subbands to detect seizure and epilepsy. *IEEE Transactions on Biomedical Engineering*, *54*(2), 205–211. doi: 10.1109/TBME.2006.886855

Agyei, S. B., Holth, M., van der Weel, F. R. R., & van der Meer, A. L. H. (2015). Longitudinal study of perception of structured optic flow and random visual motion in infants using high-density EEG. *Developmental Science*, *18*(3), 436–451. doi: 10.1111/desc.12221

Ahtola, E., Stjerna, S., Stevenson, N., & Vanhatalo, S. (2017). The use of eye tracking improves the detection of evoked responses to complex visual stimuli in the EEG of infants. *Clinical Neurophysiology Practice*, *2*, 81–90. doi: 10.1016/j.cnp.2017.03.002

Ahtola, E., Stjerna, S., Yrttiaho, S., Nelson, C. A., Leppänen, J. M., & Vanhatalo, S. (2014). Dynamic eye tracking based metrics for infant gaze patterns in the face-distractor competition paradigm. *PLoS one*, *9*(5), 1–11. doi: 10.1371/journal.pone.0097299

Allen, D., Tyler, C. W., & Norcia, A. M. (1996). Development of grating acuity and contrast sensitivity in the central and peripheral visual field of the human infant. *Vision Research*, *36*(13), 1945–1953. doi: 10.1016/0042-6989(95)00257-X

Altvater-Mackensen, N., Jessen, S., & Grossmann, T. (2016). Brain responses reveal that infants' face discrimination is guided by statistical learning from distributional information. *Developmental Science*, *20*(2). doi: 10.1111/desc.12393

Amoruso, L., Gelormini, C., Aboitiz, F., Alvarez González, M., Manes, F., Cardona, J. F., & Ibanez, A. (2013). N400 ERPs for actions: building meaning in context. *Frontiers in Human Neuroscience*, *7*. doi: 10.3389/fnhum.2013.00057

Antes, J., Penland, J., & Metzger, R. (1981). Processing global information in briefly presented pictures. *Psychological Research*, *43*(3), 277–292.

Aslin, R. (2012). Infant eyes: A window on cognitive development. *Infancy*, *17*(1), 1–11. doi: 10.1111/j.1532-7078.2011.00097.x.Infant

Bahrick, L. E., & Lickliter, R. (2000). Intersensory Redundancy Guides Attentional Selectivity and Perceptual Learning in Infancy. *Developmental Psychology*, *36*(2), 190–201. doi: 10.1037//0012-1649.36.2.190.Intersensory

Bahrick, L. E., Walker, A. S., & Neisser, U. (1981). Selective looking by infants. *Cognitive Psychology*, *13*(3), 377–390. doi: 10.1016/0010-0285(81)90014-1

Bakker, M., Kaduk, K., Elsner, C., Juvrud, J., & Gredebäck, G. (2015). The neural basis of non-verbal communication-enhanced processing of perceived give-me gestures in 9-month-old girls. *Frontiers in Psychology*, *6*, 1–6. doi: 10.3389/fpsyg.2015.00059

Benbadis, S. R., Lafrance, W. C., Papandonatos, G. D., Korabathina, K., Lin, K., & Kraemer, H. C. (2009). Interrater reliability of EEG-video monitoring. *Neurology*, *73*(11), 843–846. doi: 10.1212/WNL.0b013e3181b78425

Bertrand, O., Perrin, F., & Pernier, J. (1985). A theoretical justification of the average reference in topographic evoked potential studies. *Electroencephalography and Clinical Neurophysiology*, *62*(6), 462–464. doi: 10.1016/0168-5597(85)90058-9

Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 121–133. doi: 10.1037/0096-1523.18.1.121

Bonifacci, P., Ricciardelli, P., Lugli, L., & Pellicano, A. (2008). Emotional attention: effects of emotion and gaze direction on overt orienting of visual attention. *Cognitive Processing*, *9*(2), 127–35. doi: 10.1007/s10339-007-0198-3

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–6. doi: 10.1163/156856897x00357

Brouwer, G. J., van Ee, R., & Schwarzbach, J. (2005). Activation in visual cortex correlates with the awareness of stereoscopic depth. *The Journal of Neuroscience*, *25*(45), 10403–13. doi: 10.1523/JNEUROSCI.2408-05.2005

Carver, L. J., Meltzoff, A. N., & Dawson, G. (2006). Event-related potential (ERP) indices of infants' recognition of familiar and unfamiliar objects in two and three dimensions. *Developmental Science*, *9*(1), 51–62. doi: 10.1111/j.1467-7687.2005.00463.x

Casey, B. J., & Richards, J. E. (1988). Sustained Visual Attention in Young Infants Measured with an Adapted Version of the Visual Preference Paradigm. *Child Development*, *59*(6), 1514–1521. doi: 10.2307/1130666

Chong, H. J., Richmond, J. L., Wong, J., Qiu, A., & Rifkin-Graboi, A. (2015). Looking Behavior at Test and Relational Memory in 6-Month-Old Infants. *Infancy*, *20*(1), 18–41. doi: 10.1111/infa.12067

Choudhury, N., & Benasich, A. A. (2011). Maturation of auditory evoked potentials from 6 to 48 months: Prediction to 3 and 4 year language and cognitive abilities. *Clinical Neurophysiology*, *122*(2), 320–338. doi: 10.1016/j.clinph .2010.05.035

Clark, V., & Hillyard, S. (1996). Spatial selective attention affects early extrastriate but not striate components of the visual evoked potential. *Journal of Cognitive Neuroscience*, *8*(5), 387–402.

Cohen, L. B. (1972). Attention-getting and attention-holding processes of infant visual preferences. *Child Development*, *43*(3), 869–79. doi: 10.2307/1127638

Correa, A. G., Laciar, E., Patiño, H. D., & Valentinuzzi, M. E. (2007). Artifact removal from EEG signals using adaptive filters in cascade. In *Journal of physics: Conference series* (Vol. 90, p. 12081).

Courage, M. L., Reynolds, G. D., & Richards, J. E. (2006). Infants' Attention to Patterned Stimuli: Developmental Change from 3 to 12 Months of Age. *Child Development*, *77*(3), 680–695. doi: 10.1037/a0013572.An

Courchesne, E., Ganz, L., & Norcia, A. M. (1981). Event-related brain potentials to human faces in infants. *Child Development*, *52*(3), 804–811. doi: 10.2307/ 1129080

Cristescu, T. C., & Nobre, A. C. (2008). Differential modulation of word recognition by semantic and spatial orienting of attention. *Journal of Cognitive Neuroscience*, *20*(5), 787–801. doi: 10.1162/jocn.2008.20503

Croft, R. J., & Barry, R. J. (2000). Removal of ocular artifact from the EEG: A review. *Neurophysiologie Clinique*, *30*(1), 5–19. doi: 10.1016/S0987-7053(00) 00055-1

Csibra, G., Kushnerenko, E., & Grossmann, T. (2008). Electrophysiological Methods in Studying Infant Cognitive Development. In *Handbook of developmental cognitive neuroscience* (pp. 1–50). MIT press.

Csibra, G., Tucker, L. A., Volein, A., & Johnson, M. H. (2000). Cortical development and saccade planning: the ontogeny of the spike potential. *Neuroreport*, *11*(5),

1069–73. doi: 10.1097/00001756-200004070-00033

Davidson, R. J., Jackson, D. C., & Larson, C. L. (2000). Human Electroencephalography. In J. Caccioppo, L. Tassinary, & G. Berntson (Eds.), *Principles of psychophysiology* (2nd editio ed., chap. Human elec). Cambridge, England: Cambridge University Press.

Davila, C., & Mobin, M. (1992). Weighted averaging of evoked potentials. *IEEE Transactions on Biomedical Engineering*, *39*(4), 338–345. doi: 10.1109/10.126606

De Haan, M. (2007a). *Infant EEG and Event-Related Potentials*. Psychology Press; 1 edition.

De Haan, M. (2007b). Visual attention and recognition memory in infancy. *Infant EEG and event-related potentials*, 101–143.

De Haan, M., & Nelson, C. A. (1997). Recognition of the Mother ' s Face by Six-Month-Old Infants : A Neurobehavioral Study. *Child Development*, *68*(2), 187–210.

De Haan, M., & Nelson, C. A. (1999). Brain activity differentiates face and object processing in 6-month-old infants. *Developmental Psychology*, *35*(4), 1113–1121. doi: 10.1037/0012-1649.35.4.1113

De Haan, M., Pascalis, O., & Johnson, M. H. (2002). Specialization of neural mechanisms underlying face recognition in human infants. *Journal of Cognitive Neuroscience*, *14*(2), 199–209. doi: 10.1162/089892902317236849

DeBoer, T., Scott, L. S., & Nelson, C. A. (2007). Methods for acquiring and analyzing infant event-related potentials. *Infant EEG and event-related potentials*, 5–37.

Deligianni, F., Senju, A., Gergely, G., & Csibra, G. (2011). Automated gaze-contingent objects elicit orientation following in 8-month-old infants. *Developmental Psychology*, *47*(6), 1499–503. doi: 10.1037/a0025659

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Delorme, A., Sejnowski, T., & Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage*, *34*(4), 1443–1449. doi: 10.1016/j.neuroimage.2006.11.004

Dien, J. (1998). Issues in the application of the average reference: Review, critiques, and recommendations. *Behavior Research Methods, Instruments, &*

Computers, *30*(1), 34–43. doi: 10.3758/BF03209414

Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., & Kliegl, R. (2011). Coregistration of eye movements and EEG in natural reading: Analyses and review. *Journal of Experimental Psychology*, *140*(4), 552–572. doi: 10.1037/a0023885

Dobson, V., & Teller, D. Y. (1978). Visual acuity in human infants: A review and comparison of behavioral and electrophysiological studies. *Vision Research*, *18*(11), 1469–1483. doi: 10.1016/0042-6989(78)90001-9

Domínguez-martínez, E., Parise, E., Strandvall, T., & Reid, V. M. (2015). The Fixation Distance to the Stimulus Influences ERP Quality : An EEG and Eye Tracking N400 Study. *PloS one*, *10*(7), 1–14. doi: 10.1371/journal.pone.0134339

Duchowski, A. (2007). *Eye tracking methodology: Theory and practice*. Springer London.

Eason, R. (1981). Visual evoked potential correlates of early neural filtering during selective attention. *Bulletin of the Psychonomic Society*, *18*(4), 203–206. doi: 10.3758/BF03333604

Fantz, R. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science*, *146*(3644), 668–670. doi: 10.1126/science.146.3644.668

Farroni, T., Massaccesi, S., Menon, E., & Johnson, M. H. (2007). Direct gaze modulates face recognition in young infants. *Cognition*, *102*(3), 396–404. doi: 10.1016/j.cognition.2006.01.007

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision*, *7*(1), 10. doi: 10.1167/7.1.10

Feng, G. (2011). Eye Tracking: A Brief Guide for Developmental Researchers. *Journal of Cognition and Development*, *12*(1), 1–11. doi: 10.1080/15248372.2011.547447

Forssman, L., Wass, S., & Leppänen, J. (2014). Training non-social attention control improves infants' socio-cognitive abilities. *Poster session presented at the Biennial International Conference on Infant Studies, Berlin, Germany*.

Frick, J. E., & Richards, J. E. (2001). Individual Differences in Infants' Recognition of Briefly Presented Visual Stimuli. *Infancy*, *2 (3)*(772814175), 311–330. doi: 10.1207/S15327078IN0203

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: An event-related brain potential (ERP) sign of the brain's evaluation of nov-

elty. *Neuroscience and Biobehavioral Reviews*, *25*(4), 355–373. doi: 10.1016/S0149-7634(01)00019-7

Friedrich, M., & Friederici, A. D. (2010). Maturing brain mechanisms and developing behavioral language skills. *Brain and Language*, *114*(2), 66–71. doi: 10.1016/j.bandl.2009.07.004

Fujioka, T., Mourad, N., He, C., & Trainor, L. J. (2011). Comparison of artifact correction methods for infant EEG applied to extraction of event-related potential signals. *Clinical Neurophysiology*, *122*(1), 43–51. doi: 10.1016/j.clinph.2010.04.036

Gaspard, N., Hirsch, L. J., LaRoche, S. M., Hahn, C. D., & Brandon, M. (2014). Interrater agreement for Critical Care EEG Terminology. *Epilepsia*, *55*(9), 1366–1373. doi: 10.1111/epi.12653

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*(6), 997–1016. doi: 10.1007/s11222-013-9416-2

Geng, J., Ruff, C., & Driver, J. (2009). Saccades to a remembered location elicit spatially specific activation in human retinotopic visual cortex. *Journal of Cognitive Neuroscience*, *21*(2), 230–245. doi: 10.1162/jocn.2008.21025.Saccades

Gredebäck, G., Eriksson, M., Schmitow, C., Laeng, B., & Stenberg, G. (2012). Individual differences in face processing: Infants' scanning patterns and pupil dilations are influenced by the distribution of parental leave. *Infancy*, *17*(1), 79–101. doi: 10.1111/j.1532-7078.2011.00091.x

Gredebäck, G., Johnson, S., & von Hofsten, C. (2010). Eye tracking in infancy research. *Developmental Neuropsychology*, *35*(1), 1–19. doi: 10.1080/87565640903325758

Grossmann, T., Striano, T., & Friederic, A. D. (2006). Crossmodal integration of emotional information from face and voice in the infant brain. *Developmental Science*, *9*(3), 309–315. doi: 10.1111/j.1467-7687.2006.00494.x

Guy, M., Roberts, J. E., Tonnsen, B. L., & Richards, J. E. (2017). Neural correlates of face processing in etiologically-distinct 12-month-old infants at high-risk of autism spectrum disorder. *Developmental Cognitive Neuroscience*. doi: 10.1016/j.dcn.2017.03.002

Haith, M. M., Hazan, C., & Goodman, G. S. (1988). Expectation and anticipation of dynamic visual events by 3.5-month-old babies. *Child Development*, *59*(2), 467–479. doi: 10.2307/1130325

Halit, Csibra, G., Volein, Á., & Johnson, M. H. (2004). Face-sensitive cortical processing in early infancy. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *45*(7), 1228–1234. doi: 10.1111/j.1469-7610.2004.00321.x

Halit, H., De Haan, M., & Johnson, M. H. (2003). Cortical specialisation for face processing: Face-sensitive event-related potential components in 3- and 12-month-old infants. *NeuroImage*, *19*(3), 1180–1193. doi: 10.1016/S1053-8119(03)00076-4

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, *69*(346), 383–393. doi: 10.1080/01621459.1974.10482962

Hayes, A. F., & Krippendorff, K. (2007). Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures*, *1*(1), 77–89. doi: 10.1080/19312450709336664

He, C., Hotson, L., & Trainor, L. J. (2007). Mismatch responses to pitch changes in early infancy. *Journal of Cognitive Neuroscience*, *19*(5), 878–92. doi: 10.1162/jocn.2007.19.5.878

He, C., & Trainor, L. J. (2009). Finding the Pitch of the Missing Fundamental in Infants. *The Journal of Neuroscience*, *29*(24), 7718–8822. doi: 10.1523/JNEUROSCI.0157-09.2009

He, P., Wilson, G., & Russell, C. (2004). Removal of ocular artifacts from electroencephalogram by adaptive filtering. *Medical and Biological Engineering and Computing*, *42*(3), 407–412. doi: 10.1007/BF02344717

Hepach, R., & Westermann, G. (2016). Pupillometry in Infancy Research. *Journal of Cognition and Development*, *17*(3), 359–377. doi: 10.1080/15248372.2015.1135801

Hirsch, J., & Curcio, C. a. (1989). The spatial resolution capacity of human foveal retina. *Vision Research*, *29*(9), 1095–101. doi: 10.1016/0042-6989(89)90058-8

Hoehl, S., Reid, V., Mooney, J., & Striano, T. (2008). What are you looking at? Infants' neural processing of an adult's object-directed eye gaze. *Developmental Science*, *11*(1), 10–6. doi: 10.1111/j.1467-7687.2007.00643.x

Hoehl, S., & Striano, T. (2008). Neural Processing of Eye Gaze and Treat-Related Emotional Facial Expressions in Infancy. *Child Development*, *79*(6), 1752–1760. doi: 10.2307/27563590

Hoehl, S., & Wahl, S. (2012). Recording infant ERP data for cognitive research. *Developmental Neuropsychology*, *37*(3), 187–209. doi: 10.1080/87565641.2011

.627958

Hoehl, S., Wiese, L., & Striano, T. (2008). Young infants' neural processing of objects is affected by eye gaze direction and emotional expression. *PLoS one*, *3*(6). doi: 10.1371/journal.pone.0002389

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

Hoormann, J., Falkenstein, M., Schwarzenau, P., & Hohnsbein, J. (1998). Methods for the quantification and statistical testing of ERP differences across conditions. *Behavior Research Methods, Instruments, & Computers*, *30*(1), 103–109. doi: 10.3758/BF03209420

Horowitz, F. D., Paden, L., Bhana, K., & Self, P. (1972). An infant-control procedure for studying infant visual fixations. *Developmental Psychology*, *7*(1).

Hu, L., Mouraux, A., Hu, Y., & Iannetti, G. D. (2010). A novel approach for enhancing the signal-to-noise ratio and detecting automatically event-related potentials (ERPs) in single trials. *NeuroImage*, *50*(1), 99–111. doi: 10.1016/j.neuroimage.2009.12.010

Hyde, D. C., Jones, B. L., Porter, C. L., & Flom, R. (2010). Visual stimulation enhances auditory processing in 3-month-old infants and adults. *Developmental Psychobiology*, *52*(2), 181–189. doi: 10.1002/dev.20417

Inuso, G., La Foresta, F., Mammone, N., & Morabito, F. C. (2007). Brain activity investigation by EEG processing: Wavelet analysis, kurtosis and Renyi's entropy for artifact detection. *Proceedings of the 2007 International Conference on Information Acquisition, ICIA*, 195–200. doi: 10.1109/ICIA.2007.4295725

Jeschonek, S., Marinovic, V., Hoehl, S., Elsner, B., & Pauen, S. (2010). Do animals and furniture items elicit different brain responses in human infants? *Brain & Development*, *32*(10), 863–71. doi: 10.1016/j.braindev.2009.11.010

Johnson. (2003). Development of object concepts in infancy: Evidence for early learning in an eye-tracking paradigm. *Proceedings of the National Academy of Sciences*, *100*(18).

Johnson, M. H., de Haan, M., Oliver, A., Smith, W., Hatzakis, H., Tucker, L. A., & Csibra, G. (2001). Recording and analyzing high-density event-related potentials with infants. Using the Geodesic sensor net. *Developmental Neuropsychology*, *19*(3), 295–323. doi: $10.1207/S15326942DN1903_4$

Johnson, M. H., Posner, M. I., & Rothbart, M. K. (1991). Components of visual

orienting in early infancy: contingency learning, anticipatory looking, and disengaging. *Journal of Cognitive Neuroscience*, *3*(4), 335–44. doi: 10.1162/jocn.1991.3.4.335

Johnson, S. P., Slemmer, J. A., & Amso, D. (2004). Where infants look determines how they see: Eye movements and object perception performance in 3-month-olds. *Infancy*, *6*(2), 185–201. doi: $10.1207/s15327078in0602_3$

Joyce, C., & Rossion, B. (2005). The face-sensitive N170 and VPP components manifest the same brain processes: the effect of reference electrode site. *Clinical Neurophysiology*, *116*(11), 2613–31. doi: 10.1016/j.clinph.2005.07.005

Joyce, C. A., Gorodnitsky, I. F., & Kutas, M. (2004). Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. *Psychophysiology*, *41*(2), 313–325. doi: 10.1111/j.1469-8986.2003.00141.x

Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., Mckeown, M. J., Iragui, V., & Sejnowski, T. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, *37*(2), 163–178. doi: 10.1017/S0048577200980259

Junge, C., Cutler, A., & Hagoort, P. (2012). Electrophysiological evidence of early word learning. *Neuropsychologia*, *50*(14), 3702–3712. doi: 10.1016/j.neuropsychologia.2012.10.012

Junghöfer, M., Elbert, T., Tucker, D. M., & Rockstroh, B. (2000). Statistical control of artifacts in dense array EEG/MEG studies. *Psychophysiology*, *37*(4), 523–532. doi: 10.1111/1469-8986.3740523

Kaatiala, J., Yrttiaho, S., Forssman, L., Perdue, K., & Leppänen, J. (2013). A graphical user interface for infant ERP analysis. *Behavior Research Methods*, *46*(3), 745–757. doi: 10.3758/s13428-013-0404-4

Kaduk, K., Bakker, M., Juvrud, J., Gredebäck, G., Westermann, G., Lunn, J., & Reid, V. M. (2016). Semantic processing of actions at 9months is linked to language proficiency at 9 and 18months. *Journal of Experimental Child Psychology*, *151*, 96–108. doi: 10.1016/j.jecp.2016.02.003

Kaduk, K., Elsner, B., & Reid, V. M. (2013). Discrimination of animate and inanimate motion in 9-month-old infants: An ERP study. *Developmental Cognitive Neuroscience*, *6*, 14–22. doi: 10.1016/j.dcn.2013.05.003

Kamienkowski, J. E., Ison, M. J., Quiroga, R. Q., & Sigman, M. (2012). Fixation-related potentials in visual search: A combined EEG and eye tracking study. *Journal of Vision*, *12*(7), 4. doi: 10.1167/12.7.4

Karatekin, C. (2007). Eye tracking studies of normative and atypical development. *Developmental Review*, *27*(3), 283–348. doi: 10.1016/j.dr.2007.06.006

Karrer, J. H., Karrer, R., Bloom, D., Chaney, L., & Davis, R. (1998). Event-related brain potentials during an extended visual recognition memory task depict delayed development of cerebral inhibitory processes among 6-month-old infants with Down syndrome. *International Journal of Psychophysiology*, *29*(2), 167–200. doi: 10.1016/S0167-8760(98)00015-4

Kaunitz, L. N., Kamienkowski, J. E., Varatharajah, A., Sigman, M., Quiroga, R. Q., & Ison, M. J. (2014). Looking for a face in the crowd: Fixation-related potentials in an eye-movement visual search task. *NeuroImage*, *89*, 297–305. doi: 10.1016/j.neuroimage.2013.12.006

Kayser, J., & Tenke, C. E. (2003). Optimizing PCA methodology for ERP component identification and measurement: theoretical rationale and empirical evaluation. *Clinical Neurophysiology*, *114*(12), 2307–2325. doi: https://doi.org/10.1016/S1388-2457(03)00241-4

Kirwan, B. (1992). Human error identification in human reliability assessment. Part 1: Overview of approaches. *Applied Ergonomics*, *23*(5), 299–318. doi: http://dx.doi.org/10.1016/0003-6870(92)90292-4

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3? *Perception 36 ECVP Abstract Supplement*, 14.

Kouider, S., Stahlhut, C., Gelskov, S. V., Barbosa, L. S., Dutat, M., de Gardelle, V., ... Dehaene-Lambertz, G. (2013). A neural marker of perceptual consciousness in infants. *Science*, *340*(6130), 376–80. doi: 10.1126/science.1232509

Krippendorff, K. (2004). Reliability in Content Analysis: Some Common Misconceptions and Recommendations. *Human Communication Research*, *30*(3), 411–433. doi: 10.1093/hcr/30.3.411

Krishnaveni, V., Jayaraman, S., Aravind, S., Hariharasudhan, V., & Ramadoss, K. (2006). Automatic identification and removal of ocular artifacts from EEG using wavelet transform. *Measurement Science Review*, *6*(4), 45–57.

Kuipers, J. R., & Thierry, G. (2011). N400 amplitude reduction correlates with an increase in pupil size. *Frontiers in Human Neuroscience*, *5*, 61. doi: 10.3389/fnhum.2011.00061

Kuipers, J. R., & Thierry, G. (2013). ERP-pupil size correlations reveal how bilingualism enhances cognitive flexibility. *Cortex*, *49*(10), 2853–2860. doi: 10.1016/j.cortex.2013.01.012

Kulke, L., Atkinson, J., & Braddick, O. (2015). Automatic detection of attention shifts in infancy: Eye tracking in the fixation shift paradigm. *PLoS one*, *10*(12), 1–14. doi: 10.1371/journal.pone.0142505

Kulke, L., Atkinson, J., & Braddick, O. (2016a). Neural Differences between Covert and Overt Attention Studied using EEG with Simultaneous Remote Eye Tracking. *Frontiers in Human Neuroscience*, *10*, 592. doi: 10.3389/fnhum.2016.00592

Kulke, L., Atkinson, J., & Braddick, O. (2016b). Neural mechanisms of attention become more specialised during infancy: Insights from combined eye tracking and EEG. *Developmental Psychobiology*, *59*(2), 1–11. doi: 10.1002/dev.21494

Kushnerenko, E., Teinonen, T., Volein, A., & Csibra, G. (2008). Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(32), 11442–11445. doi: 10.1073/pnas.0804275105

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463–470. doi: 10.1016/S1364-6613(00)01560-6

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. doi: 10.1146/annurev.psych.093008.131123

Laeng, B., & Endestad, T. (2012). Bright illusions reduce the eye's pupil. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(6), 2162–7. doi: 10.1073/pnas.1118298109

Leppänen, J. M., Moulson, M. C., Vogel-Farley, V. K., & Nelson, C. A. (2007). An ERP study of emotional face processing in the adult and infant brain. *Child Development*, *78*(1), 232–245. doi: 10.1111/j.1467-8624.2007.00994.x.An

Luck, S. J. (2005). *An introduction to the event-related potential technique.* MIT press.

Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences*, *4*(11), 432–440. doi: 10.1016/S1364-6613(00)01545-X

Maguire, M. J., Magnon, G., & Fitzhugh, A. E. (2014). Improving data retention in EEG research with children using child-centered eye tracking. *Journal of Neuroscience Methods*, *238*, 78–81. doi: 10.1016/j.jneumeth.2014.09.014

Mangun, G. R., & Hillyard, S. A. (1988). Spatial gradients of visual attention: be-

havioral and electrophysiological evidence. *Electroencephalography and Clinical Neurophysiology*, *70*(5), 417–28. doi: 10.1016/0013-4694(88)90019-3

Manresa, J. A. B., Arguissain, F. G., Redondo, D. E. M., Morch, C. D., & Andersen, O. K. (2015). On the agreement between manual and automated methods for single-trial detection and estimation of features from event-related potentials. *PLoS one*, *10*(8), 1–21. doi: 10.1371/journal.pone.0134127

Marraffa, R., & Sperati, V. (2012). A bio-inspired attention model of anticipation in gaze-contingency experiments with infants. *Development and Learning and Epigenetic Robotics (ICDL), 2012 IEEE International Conference on*.

Marshall, P. J., Young, T., & Meltzoff, A. N. (2011). Neural correlates of action observation and execution in 14-month-old infants: An event-related EEG desynchronization study. *Developmental Science*, *14*(3), 474–480. doi: 10.1111/j.1467-7687.2010.00991.x

McCarthy, G., & Nobre, A. C. (1993). Modulation of semantic processing by spatial selective attention. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *88*(3), 210–219. doi: 10.1016/0168-5597(93)90005-A

Mchugh, M. L. (2012). Interrater reliability : the kappa statistic. *Biochemia Medica*, *22*(3), 276–282.

Missana, M., Rajhans, P., Atkinson, A. P., & Grossmann, T. (2014). Discrimination of fearful and happy body postures in 8-month-old infants: an event-related potential study. *Frontiers in Human Neuroscience*, *8*, 1–7. doi: 10.3389/fnhum.2014.00531

Miyazaki, M., Takahashi, H., Rolf, M., Okada, H., & Omori, T. (2014). The image-scratch paradigm: a new paradigm for evaluating infants' motivated gaze control. *Scientific Reports*, *4*, 1–6. doi: 10.1038/srep05498

Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2011). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*, *48*(2), 229–240. doi: 10.1111/j.1469-8986.2010.01061.x

Monroy, C. D., Gerson, S. A., Domínguez-Martínez, E., Kaduk, K., Hunnius, S., & Reid, V. (2017). Sensitivity to structure in action sequences: An infant event-related potential study. *Neuropsychologia*, –. doi: https://doi.org/10.1016/j.neuropsychologia.2017.05.007

Mourad, N., Reilly, J. P., De Bruin, H., Hasey, G., & MacCrimmon, D. (2007). A simple and fast algorithm for automatic suppression of high-amplitude artifacts

in EEG data. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *1*. doi: 10.1109/ICASSP.2007.366699

Nelson, C. A. (1994). Neural correlates of recognition memory in the first postnatal year. *Human Behavior and the Developing Brain*.

Nelson, C. A., & De Haan, M. (1996). Neural correlates of infants' visual responsiveness to facial expressions of emotion. *Developmental Psychobiology*, *29*(7), 577–595. doi: 10.1002/(SICI)1098-2302(199611)29:7\<577::AID-DEV3\>3.0.CO;2-R

Nelson, C. A., & McCleery, J. (2008). Use of event-related potentials in the study of typical and atypical development. *Journal of the American Academy of Child Adolesc Psychiatry*, *47*(11), 1252–1261. doi: 10.1097/CHI.0b013e318185a6d8.Use

Nolan, H., Whelan, R., & Reilly, R. B. (2010). FASTER: Fully Automated Statistical Thresholding for EEG artifact Rejection. *Journal of Neuroscience Methods*, *192*(1), 152–162. doi: 10.1016/j.jneumeth.2010.07.015

Nordli, D. R., Bazil, C. W., Scheuer, M. L., & Pedley, T. a. (1997). Recognition and classification of seizures in infants. *Epilepsia*, *38*(5), 553–60. doi: 10.1111/j.1528-1157.1997.tb01140.x

Oakes, L. M. (2010). Infancy Guidelines for Publishing Eye-Tracking Data. *Infancy*, *15*(1), 1–5. doi: 10.1111/j.1532-7078.2010.00030.x

Oakes, L. M. (2012). Advances in Eye Tracking in Infancy Research. *Infancy*, *17*(1), 1–8. doi: 10.1111/j.1532-7078.2011.00101.x

Olsen, A. (2012). *The Tobii I-VT Fixation Filter Algorithm description* (Tech. Rep.). Stockholm: Tobii technology AB.

Olsen, A., & Ricardo, M. (2012). Identifying parameter values for an I-VT fixation filter suitable for handling data sampled with various sampling frequencies. In *Symposium on eye tracking research and applications* (Vol. 1, pp. 317–320).

O'Regan, S. (2013). *Artefact Detection and Removal Algorithms for EEG Diagnostic Systems* (Unpublished doctoral dissertation). PhD Thesis. University College Cork.

O'Regan, S., & Marnane, W. (2013). Multimodal detection of head-movement artefacts in EEG. *Journal of Neuroscience Methods*, *218*(1), 110–120. doi: 10.1016/j.jneumeth.2013.04.017

Orekhova, E. V., Stroganova, T. A., Posikera, I. N., & Elam, M. (2006). EEG theta rhythm in infants and preschool children. *Clinical Neurophysiology*, *117*(5),

1047–1062. doi: 10.1016/j.clinph.2005.12.027

Otte, R. A., Donkers, F. C. L., Braeken, M. A. K. A., & Van den Bergh, B. R. H. (2015). Multimodal processing of emotional information in 9-month-old infants I: Emotional faces and voices. *Brain and Cognition*, *95*, 99–106. doi: 10.1016/j.bandc.2014.09.007

Otten, L. J., & Rugg, M. D. (2005). Interpreting Event-Related Brain Potentials. *Event Related Potentials: A Methods Handbook*(1997), 3–16. doi: 10.1016/j.biopsycho.2005.11.012

Parise, E., & Csibra, G. (2013). Neural responses to multimodal ostensive signals in 5-month-old infants. *PloS one*, *8*(8). doi: 10.1371/journal.pone.0072360

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. a., Johnson, R., ... Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology*, *37*(2).

Picton, T. W., & Taylor, M. J. (2007). Electrophysiological evaluation of human brain development. *Developmental Neuropsychology*, *31*(3), 249–278. doi: 10.1080/87565640701228732

Plöchl, M., Ossandón, J. P., & König, P. (2012). Combining EEG and eye tracking: identification, characterization, and correction of eye movement artifacts in electroencephalographic data. *Frontiers in Human Neuroscience*, *6*, 278. doi: 10.3389/fnhum.2012.00278

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. doi: 10.1080/00335558008248231

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. doi: 10.1080/13803395.2011.639298

Rayner, K., & Inhoff, A. (1981). Masking of foveal and parafoveal vision during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance,*, *7*(1), 167–79. doi: 10.1037/0096-1523.7.1.167

Reid, V. M., Csibra, G., Belsky, J., & Johnson, M. H. (2007). Neural correlates of the perception of goal-directed action in infants. *Acta Psychologica*, *124*(1), 129–38. doi: 10.1016/j.actpsy.2006.09.010

Reid, V. M., Hoehl, S., Grigutsch, M., Groendahl, A., Parise, E., & Striano, T. (2009). The neural correlates of infant and adult goal prediction: evidence for semantic processing systems. *Developmental Psychology*, *45*(3), 620–9. doi:

10.1037/a0015209

Reid, V. M., Hoehl, S., Landt, J., & Striano, T. (2008). Human infants dissociate structural and dynamic information in biological motion: evidence from neural systems. *Social Cognitive and Affective Neuroscience*, *3*(2), 161–7. doi: 10 .1093/scan/nsn008

Reid, V. M., Hoehl, S., & Striano, T. (2006). The perception of biological motion by infants: An event-related potential study. *Neuroscience Letters*, *395*(3), 211–214. doi: 10.1016/j.neulet.2005.10.080

Reid, V. M., & Striano, T. (2008). N400 involvement in the processing of action sequences. *Neuroscience Letters*, *433*(2), 93–7. doi: 10.1016/j.neulet.2007.12 .066

Reynolds, G. D., & Guy, M. W. (2012). Brain-behavior relations in infancy: integrative approaches to examining infant looking behavior and event-related potentials. *Developmental Neuropsychology*, *37*(3), 210–25. doi: 10.1080/ 87565641.2011.629703

Reynolds, G. D., & Richards, J. E. (2005). Familiarization, Attention, and Recognition Memory in Infancy: An Event-Related Potential and Cortical Source Localization Study. *Developmental Psychology*, *41*(4), 598–615. doi: 10.1016/j.biotechadv.2011.08.021.Secreted

Richards, J. E. (2001). Attention in young infants: A developmental psychophysiological perspective. *Handbook of Developmental Cognitive Neuroscience*, 321–338.

Richards, J. E. (2003). Attention affects the recognition of briefly presented visual stimuli in infants: an ERP study. *Developmental Science*, *6*(3), 312–328. doi: 10.1016/j.micinf.2011.07.011.Innate

Richards, J. E., & Hunter, S. K. (1997). Peripheral stimulus localization by infants with eye and head movements during visual attention. *Vision Research*, *37*(21), 3021–3035. doi: 10.1016/S0042-6989(97)00082-5

Richards, J. E., Reynolds, G. D., & Courage, M. L. (2010). The Neural Bases of Infant Attention. *Current Directinon in Psychology Science*, *19*(1), 41–46. doi: 10.1177/0963721409360003.The

Richmond, J., & Nelson, C. A. (2009). Relational memory during infancy: Evidence from eye tracking. *Developmental Science*, *12*(4), 549–556. doi: 10.1111/ j.1467-7687.2009.00795.x

Righi, G., Westerlund, A., Congdon, E. L., Troller-Renfree, S., & Nelson, C. a.

(2014). Infants' experience-dependent processing of male and female faces: Insights from eye tracking and event-related potentials. *Developmental Cognitive Neuroscience*, *8*, 144–152. doi: 10.1016/j.dcn.2013.09.005

Romero, S., Mañanas, M. A., & Barbanoj, M. J. (2008). A comparative study of automatic techniques for ocular artifact reduction in spontaneous EEG signals based on clinical target variables: A simulation case. *Computers in Biology and Medicine*, *38*(3), 348–360. doi: 10.1016/j.compbiomed.2007.12.001

Rosander, K., & Von Hofsten, C. (2002). Development of gaze tracking of small and large objects. *Experimental Brain Research*, *146*(2), 257–264. doi: 10.1007/s00221-002-1161-2

Rugg, M. D., & Coles, M. G. (1995). *Electrophysiology of mind: Event-related brain potentials and cognition.* Oxford University Press.

Saby, J. N., Marshall, P. J., & Meltzoff, A. N. (2012). Neural correlates of being imitated: An EEG study in preverbal infants. *Social Neuroscience*, *7*(6), 650–661. doi: 10.1080/17470919.2012.691429

Saez de Urabain, I. R., Johnson, M. H., & Smith, T. J. (2014). GraFIX: A semiautomatic approach for parsing low- and high-quality eye-tracking data. *Behavior Research Methods*, *47*(1), 53–72. doi: 10.3758/s13428-014-0456-0

Salomão, S. R., & Ventura, D. F. (1995). Large sample population age norms for visual acuities obtained with Vistech-Teller Acuity Cards. *Investigative Ophthalmology & Visual Science*, *36*(3), 657–70.

Sanocki, T., & Epstein, W. (1997). Priming spatial layout of scenes. *Psychological Science*, *8*(5), 374–378.

Shaddy, D. J., & Colombo, J. (2004). Developmental Changes in Infant Attention to Dynamic and Static Stimuli. *Infancy*, *5*(3), 355–365. doi: 10.1207/s15327078in0503_6

Shea, S. L., & Aslin, R. N. (1990). Oculomotor responses to step-ramp targets by young human infants. *Vision Research*, *30*(7), 1077–1092. doi: 10.1016/0042-6989(90)90116-3

Shibata, H., Gyoba, J., & Suzuki, Y. (2009). Event-related potentials during the evaluation of the appropriateness of cooperative actions. *Neuroscience Letters*, *452*(2), 189–93. doi: 10.1016/j.neulet.2009.01.042

Shukla, M., Wen, J., White, K. S., & Aslin, R. N. (2011). SMART-T: a system for novel fully automated anticipatory eye-tracking paradigms. *Behavior Research Methods*, *43*(2), 384–398. doi: 10.3758/s13428-010-0056-6

Simola, J., Le Fevre, K., Torniainen, J., & Baccino, T. (2015). Affective processing in natural scene viewing: Valence and arousal interactions in eye-fixation-related potentials. *NeuroImage*, *106*, 21–33. doi: 10.1016/j.neuroimage.2014.11.030

Sireteanu, R., Fronius, M., & Constantinescu, D. H. (1994). The development of visual acuity in the peripheral visual field of human infants: binocular and monocular measurements. *Vision Research*, *34*(12), 1659–1671. doi: 10.1016/0042-6989(94)90124-4

Sireteanu, R., Kellerer, R., & Boergen, K. P. (1983). The development of peripheral visual acuity in human infants. A preliminary study. *Human Neurobiology*, *3*(2), 81–85.

Sirois, S., & Brisson, J. (2014). Pupillometry. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(6), 679–692. doi: 10.1002/wcs.1323

Sitnikova, T., & Holcomb, P. (2008). Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *Journal of Cognitive Neuroscience*, *20*(11), 2037–2057. doi: 10.1162/jocn.2008.20143.Two

Slaughter, V., & Suddendorf, T. (2007). Participant loss due to "fussiness" in infant visual paradigms: A review of the last 20 years. *Infant Behavior and Development*, *30*(3), 505–514. doi: 10.1016/j.infbeh.2006.12.006

Snyder, K., Webb, S. J., & Nelson, C. A. (2002). Theoretical and methodological implications of variability in infant brain response during a recognition memory paradigm. *Infant Behavior and Development*, *25*(4), 466–494. doi: 10.1016/S0163-6383(02)00146-7

Southgate, V., & Vernetti, A. (2014). Belief-based action prediction in preverbal infants. *Cognition*, *130*(1), 1–10. doi: 10.1016/j.cognition.2013.08.008

Spencer, K. M. (2005). Averaging, Detection, and Classification of Single-Trial ERPs. *Event-Related Potentials: A Methods Handbook*, 209.

Srinivasan, V., Eswaran, C., & Sriraam, N. (2007). Approximate entropy-based epileptic EEG detection using artificial neural networks. *IEEE Transactions on information Technology in Biomedicine*, *11*(3), 288–295. doi: 10.1109/TITB.2006.884369

Stets, M., Burt, M., & Reid, V. M. (2013). Infants Need More Variety - Increased Data Acquisition with Reduced Participant Attrition in Infant ERP Studies. *Frontiers in Psychology*, *4*, 117. doi: 10.3389/fpsyg.2013.00117

Stets, M., & Reid, V. M. (2011). Infant ERP amplitudes change over the course of an

experimental session: Implications for cognitive processes and methodology. *Brain and Development*, *33*(7), 558–568. doi: 10.1016/j.braindev.2010.10.008

Stets, M., Stahl, D., & Reid, V. M. (2012). A meta-analysis investigating factors underlying attrition rates in infant ERP studies. *Developmental Neuropsychology*, *37*(3), 226–52. doi: 10.1080/87565641.2012.654867

Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern recognition: a review. *Journal of Vision*, *11*(5), 13. doi: 10.1167/11.5.13

Striano, T., Kopp, F., Grossmann, T., & Reid, V. M. (2006). Eye contact influences neural processing of emotional expressions in 4-month-old infants. *Social Cognitive and Affective Neuroscience*, *1*(2), 87–94. doi: 10.1093/scan/nsl008

Subasi, A., & Ercelebi, E. (2005). Classification of EEG signals using neural network and logistic regression. *Computer Methods and Programs in Biomedicine*, *78*(2), 87–99.

Tew, S., Fujioka, T., He, C., & Trainor, L. (2009). Neural representation of transposed melody in infants at 6 months of age. *Annals of the New York Academy of Sciences*, *1169*(1), 287–290. doi: 10.1111/j.1749-6632.2009.04845.x

Thierry, G. (2005). The use of event related potentials in the study of early cognitive development. *Infant and Child Development*, *14*(1), 85–94. doi: 10.1002/icd.

Unser, M., & Aldroubi, A. (1996). A review of wavelets in biomedical applications. *Proceedings of the IEEE*, *84*(4), 626–638. doi: 10.1109/5.488704

van Hoogmoed, A. H., van den Brink, D., & Janzen, G. (2013). Infants' object location and identity processing in spatial scenes: An ERP study. *Brain and Behavior*, *3*(6), 729–737. doi: 10.1002/brb3.184

Visual statistical learning of shape sequences: An ERP study. (2009). *Neuroscience Research*, *64*(2), 185–190. doi: 10.1016/j.neures.2009.02.013

Vos, D. M., Riès, S., Vanderperren, K., Vanrumste, B., Alario, F. X., Huffel, V. S., & Burle, B. (2010). Removal of muscle artifacts from EEG recordings of spoken language production. *Neuroinformatics*, *8*(2), 135–150. doi: 10.1007/s12021-010-9071-0

Wang, Q., Bolhuis, J., Rothkopf, C. a., Kolling, T., Knopf, M., & Triesch, J. (2012). Infants in control: rapid anticipation of action outcomes in a gaze-contingent paradigm. *PloS one*, *7*(2), e30884. doi: 10.1371/journal.pone.0030884

Wass, S., Forssman, L., & Leppänen, J. (2014). Robustness and precision: How data quality may influence key dependent variables in infant eye-tracker analyses. *Infancy*, *19*(5), 427–460. doi: 10.1111/infa.12055

Wass, S., Porayska-Pomsta, K., & Johnson, M. H. (2011). Training attentional control in infancy. *Current Biology*, *21*(18), 1543–1547. doi: 10.1016/j.cub .2011.08.004

Weaver, M. D., Clayton Hickey, & van Zoest, W. (2017). The impact of salience and visual working memory on the monitoring and control of saccadic behaviour: An eye-tracking and EEG study. *Psychophysiology*, *54*(4), 544–554. doi: 10.1111/psyp.12817

Webb, S. J., Dawson, G., Bernier, R., & Panagiotides, H. (2006). ERP evidence of atypical face processing in young children with autism. *Journal of Autism and Developmental Disorders*, *36*(7), 881–890. doi: 10.1007/s10803-006-0126-x

Webb, S. J., Long, J. D., & Nelson, C. A. (2005). A longitudinal investigation of visual event-related potentials in the first year of life. *Developmental Science*, *8*(6), 605–616. doi: 10.1111/j.1467-7687.2005.00452.x

West, W. C., & Holcomb, P. J. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Brain Research. Cognitive Brain Research*, *13*(3), 363–75. doi: 10.1016/S0926-6410(01)00129-X

Xue, J., Quan, C., Li, C., Yue, J., & Zhang, C. (2017). A crucial temporal accuracy test of combining EEG and Tobii eye tracker. *Medicine*, *96*(13). doi: 10.1097/MD.0000000000006444

Yao, D., Wang, L., Oostenveld, R., Nielsen, K. D., Arendt-Nielsen, L., & Chen, A. C. N. (2005). A comparative study of different references for EEG spectral mapping: the issue of the neutral reference and the use of the infinity reference. *Physiological Measurement*, *26*(3), 173–84. doi: 10.1088/0967-3334/26/3/003

Yeung, N., Bogacz, R., Holroyd, C. B., & Cohen, J. D. (2004). Detection of synchronized oscillations in the electroencephalogram: An evaluation of methods. *Psychophysiology*, *41*(6), 822–832. doi: 10.1111/j.1469-8986.2004.00239.x

Yrttiaho, S., Forssman, L., Kaatiala, J., & Leppänen, J. M. (2014). Developmental precursors of social brain networks: The emergence of attentional and cortical sensitivity to facial expressions in 5 to 7 months old infants. *PLoS one*, *9*(6). doi: 10.1371/journal.pone.0100811

Appendix A

Supplementary tables for Chapter 4

Participant	Trial assessment		Participants included			
ID	α	95% CI	ED1	ED2	ED3	ED4
1	0.116	(-0.078, 0.301)	x	x	x	x
2	0.237	(0.044, 0.431)	-	x	x	x
3	0.147	(-0.083, 0.374)	x	x	-	x
4	0.075	(-0.126, 0.263)	x	x	x	x
5	0.455	(0.176, 0.686)	-	-	-	x
6	0.549	(0.382, 0.701)	x	x	x	x
7	0.319	(0.121, 0.501)	x	x	-	x
8	0.146	(-0.083, 0.385)	x	x	-	x
9	0.085	(-0.113, 0.284)	x	x	x	x
10	0.257	(0.019, 0.470)	-	-	-	x
11	0.169	(-0.035, 0.371)	x	-	-	x
12	0.326	(0.116, 0.524)	-	x	-	x
13	0.285	(0.061, 0.483)	-	x	-	x
14	0.188	(-0.008, 0.383)	x	x	x	x
15	0.311	(0.109, 0.500)	-	-	-	x
16	0.034	(-0.159, 0.221)	-	-	-	-
17	0.012	(-0.179, 0.207)	-	-	-	-
18	0.231	(0.013, 0.446)	x	x	x	x
19	0.505	(0.320, 0.671)	-	-	-	-

Table A.1: Agreement results among editors at a participant level and participants included by each editor. From left to right: (1) Agreement results among all editors (Krippendorff's alpha values and confidence intervals (CI)) for trial assessment, divided by participants. Alpha confidence intervals at the 95% level were calculated by applying bootstrap analysis of 10.000 samples (Hayes & Krippendorff, 2007). (2) Inclusion (x) or exclusion (-) of each participant in the final ERP sample by each of the editors.

EEG Characteristic	Included	Description
Time domain characteristics		
Amplitude range	Yes	Peak-to-peak amplitude
Linear trend	Yes	Absolute value of the slope fit of the EEG signal to an oblique straight line (Delorme et al., 2007)
Deviation from the channel mean	Yes	Deviation from the channel's average value as described in Nolan et al. (2010)
SNR	Yes	Signal-to-Noise Ratio defined as the ratio of the range and standard deviation
Number of local maxima	Yes	Number of times that the first derivative crosses x-axis (O'Regan, 2013)
Kurtosis	Yes	4th order moment of the data distribution as described in Delorme et al. (2007)
Sliding standard deviation	No	Maximum value of the standard deviation of a sliding time window of 500 ms (Hoehl & Wahl, 2012)
RMS	No	Root mean Square (RMS) amplitude. Quadratic mean of the amplitude.
Line length	No	Sum of distances between consecutive points of the EEG epoch (O'Regan, 2013)
NLE	No	Non-linear Energy (NLE) of the EEG signal as described in O'Regan (2013) and O'Regan and Marnane (2013)
Entropy	No	Reny's entropy. Level of randomness in the EEG signal (Inuso et al., 2007)

Table continues on the next page

Table begins on the previous page

EEG Characteristic	**Included**	**Description**
Frequency domain EEG characteristics		
Power at 0 – 4Hz	Yes	Average power of the Power Spectral Density (PSD) between the frequencies 0-4Hz
Power at 8 – 13Hz	Yes	Average power of the PSD between the frequencies 8-13Hz
Power at 30 - 60Hz	Yes	Average power of the PSD between the frequencies 30-60Hz
Spectral edge frequency	Yes	Frequency under which the 80% of the power in the PSD lies. (O'Regan, 2013)
Peak frequency	Yes	Frequency corresponding to the largest amplitude of the PSD. (O'Regan, 2013)
Power at 5 – 7Hz	No	Average power of the Power Spectral Density (PSD) between the frequencies 5-7Hz
Power at 14 – 29Hz	No	Average power of the Power Spectral Density (PSD) between the frequencies 14-29 Hz

Table A.2: List of the EEG characteristics that were initially calculated for each EEG epoch and electrode. Based on the Pearson correlation between all the EEG characteristics, 11 independent variables were chosen as fixed effects for the mixed-effect logarithmic model.

Participant ID	Model 1 (Editor 1)	Model 2 (Editor 2)	Model 3 (Editor 3)
1	1.28 (0.95, 1.73)	0.28 (0.19, 0.42)	0.58 (0.39, 0.88)
2	0.18 (0.12, 0.28)	0.38 (0.29, 0.51)	0.87 (0.64, 1.18)
3	0.24 (0.19, 0.31)	0.15 (0.12, 0.19)	0.40 (0.32, 0.50)
4	3.06 (2.48, 3.77)	4.81 (3.80, 6.09)	5.66 (4.53, 7.06)
5	0.42 (0.27, 0.64)	0.96 (0.78, 1.18)	0.80 (0.62, 1.03)
6	4.62 (3.69, 5.78)	1.64 (1.24, 2.17)	1.78 (1.40, 2.27)
7	0.60 (0.44, 0.82)	0.30 (0.24, 0.37)	0.44 (0.35, 0.56)
8	1.30 (1.02, 1.67)	0.95 (0.69, 1.32)	0.89 (0.66, 1.18)
9	2.67 (2.17, 3.28)	1.28 (1.04, 1.57)	1.70 (1.38, 2.09)
10	1.39 (1.06, 1.82)	0.35 (0.29, 0.43)	1.26 (1.00, 1.58)
11	2.84 (2.25, 3.58)	0.93 (0.75, 1.17)	2.54 (1.99, 3.24)
12	0.58 (0.37, 0.90)	1.34 (1.07, 1.68)	0.33 (0.25, 0.43)
13	0.65 (0.47, 0.90)	4.35 (3.49, 5.44)	0.42 (0.28, 0.62)
14	1.81 (1.47, 2.24)	1.83 (1.34, 2.50)	0.96 (0.78, 1.18)
15	0.73 (0.47, 1.15)	1.34 (1.03, 1.73)	0.97 (0.75, 1.26)
16	2.60 (1.84, 3.67)	2.66 (2.17, 3.26)	3.72 (2.87, 4.82)
17	0.30 (0.19, 0.47)	0.82 (0.59, 1.15)	0.36 (0.26, 0.51)
18	2.40 (1.85, 3.11)	6.82 (4.61, 10.08)	3.55 (2.73, 4.61)
19	0.52 (0.38, 0.71)	0.22 (0.18, 0.27)	0.48 (0.36, 0.63)

Table A.3: Participant-level random effects from each model. The table shows participant level random effects of each model and confidence intervals (CI). Random effects above 1 indicate greater variation in the outcome for that individual compared to that which would be predicted by the fixed effects (and may indicate an especially restless child) and random effects below 1 indicate lesser variation in the outcome for that individual compared to that which would be predicted by the fixed effects (and may indicate that the child is comparatively calm).