An Analysis of EEG Signals Present During Target Search

Graham Healy

A dissertation submitted in fulfilment of the requirements for the award of

Doctor of Philosophy (Ph.D.)

to the

Dublin City University
School of Computing

CLARITY: Centre for Sensor Web Technologies

Supervisor: Prof. Alan F. Smeaton

16th January 2012

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: ________________ (Candidate) ID No.: 54430093 Date: 16th Jan 2012

Abstract

Recent proof-of-concept research has appeared highlighting the applicability of using Brain Computer Interface (BCI) technology to utilise a subjects visual system to classify images. This technique involves classifying a users EEG (Electroencephalography) signals as they view images presented on a screen. The premise is that images (targets) that arouse a subjects attention generate distinct brain responses, and these brain responses can then be used to label the images. Research thus far in this domain has focused on examining the tasks and paradigms that can be used to elicit these neurologically informative signals from images, and the correlates of human perception that modulate them. While success has been shown in detecting these responses in high speed presentation paradigms, there is still an open question as to what search tasks can ultimately benefit from using an EEG based BCI system.

In this thesis we explore: (1) the neural signals present during visual search tasks that require eye movements, and how they inform us of the possibilities for BCI applications utilising eye tracking and EEG in combination with each other, (2) how temporal characteristics of eye movements can give indication of the suitability of a search task to being augmented by an EEG based BCI system, (3) the characteristics of a number of paradigms that can be used to elicit informative neural responses to drive image search BCI applications.

In this thesis we demonstrate EEG signals can be used in a discriminative manner to label images. In addition, we find in certain instances, that signals derived from sources such as eye movements can yield significantly more discriminative information.

Acknowledgements

Foremost, I would like to express my gratitude to my supervisor Prof. Alan Smeaton. Without his encouragement and guidance over the past 3 years, this dissertation would not have been possible.

I would also like to thank all of my wonderful friends and colleagues within DCU, as well as the entire CLARITY family for their continuous support. I enjoyed collaborating with each and every one of you.

My sincere thanks to my parents Linda and Frank for having provided a learning environment which allowed me to get to where I am now. A special thanks to my sister Sarah for always questioning whether or not I really knew what I was talking about.

I appreciate all the support from everyone that I have known these past three years. There is, however, a lot to be said about the people and places from my past who have helped shape who I am today. For instance my old secondary school, Ardscoil Rís and its teachers have played a large part in shaping the thought processes that led me to doing this PhD. I would also like to thank Tom, Jo, Stephen, Alan, and all of my friends who have been constant pillars for me throughout my life and for that I am forever grateful.

Finally, I would like to thank my fiancée Liane for her endless support, love and devotion throughout these last few years. There were many difficult days and she was there standing by me through them.

Beep.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Computing technology is now a fundamental enabler for many forms of entertainment. From gaming to movies, much of these forms of entertainment are based around using images and video. In a similar way, much of the development of science in recent decades has been enabled by computing technology and has also been based around image and video information, from astronomy to x-ray imaging. In fact most of our society is now supported and enhanced by computing technology, from space exploration to social networking. Again, the use of image and video is central. Computing technology has enabled the creation, storage, transmission and rendering of image and video data, but we struggle to develop computational approaches to actually *managing* image and video data. At our fingertips are billions of images and millions of hours of video. The greatest challenge, however, is in searching, browsing and finding the right media at the right time. The area of multimedia retrieval/browsing, especially of visual media, remains the focus of a large research effort. Progress in this research has been slower than the rate at which this media is growing in volume. Recent proof-of-concept research has appeared showing the applicability of Brain Computer Interface (BCI) technology to detect and label images. The premise is that detectable responses occur in the brain in response to stimuli such as pictures. While a user may not explicitly express that there is anything significant about a particular event, such as seeing a picture of a loved one or finding a key piece of information in a document, their brain signals can indicate otherwise in a way which is outside the user's control. By placing EEG (Electroencephalography) sensors on the scalp, we can monitor electrical signals generated by the brain so as to identify those that may allow us to label such events (such as looking at an image) as significant, emotionally/attentionally arousing, or unexpected. Traditionally, the problems approached by BCI systems have focused on restoration of functionality and/or communication to people with a variety of impairing disorders such as stroke or brain damage. These systems tended to be cumbersome and the cost of acquisition, set-up, and maintenance were justified by the sheer necessity of the basic communication facilities they could assist in restoring. Recently, however, systems of this type are becoming cheaper and more accessible to the consumer, with research exposing potential applications in domains such as media and entertainment. In this thesis we are concerned with the utilisation of EEG signals in response to the detection of targets in images. An example of this might be searching a fast-paced stream of images displayed on a computer screen for those containing bridges, or searching an image to see if it contains one or more people. Other applications spaces where we might expect this research to be applicable are those involving situations or tasks wherein a subject does not vocalise or explicitly state meaning of events such as in sports, military combat, air flight, and so on.

1.1 Motivation

Research examining the use of EEG BCI for assisting in image search has thus far focused on examining the tasks and paradigms that can be used to elicit and detect neurologically informative signals using images. While success has been shown in detecting these responses in high speed presentation paradigms, there is still an open question as to what search tasks can ultimately benefit from using an EEG-based BCI systems. In this thesis we examine the hypothesis that EEG and Eye Tracking can be used to improve the effectiveness in searching for certain types of targets in images. EEG BCI systems have been demonstrated in being able to augment the speed and proficiency of those engaged in tasks of explicit image labelling, such as intelligence analysts sorting through satellite imagery searching for intelligence related content (Huang et al., 2011). This is useful in an instance such as this as only an intelligence analyst or human may be capable of detecting this type of content. To date, however, little research has been done in examining how these signals can be combined with eye movements to unveil neural correlates of target detection on a fixation by fixation basis. This thesis conducts investigatory work in this regard. In addition to this, we explore with a number of experiments a set of pertinent questions surrounding the application of EEG BCI such as whether or not we can use a reduced number of sensor channels.

This thesis and the content within is shaped by four questions:

1. What neural signals are present during visual search tasks that require eye movements, and how do they inform us of the possibilities for BCI applications utilising eye tracking and EEG *in combination* with each other?

2. How do the temporal characteristics of eye movements give indication of the suitability of a search task to being augmented by an EEG based BCI system?

3. What are the characteristics of paradigms that can be used to elicit informative neural responses to drive image search BCI applications?

4. Can we use a reduced number of EEG channels in EEG BCI search?

1.2 Thesis Structure

In this thesis we explore a number of research questions in conjunction with a central hypothesis to understand how EEG and eye tracking can be utilised in image search applications. We primarily do this with a number of experiments where subjects are required to engage in a variety of search tasks. These experiments and search tasks are intended by their nature to allow us to expose and study the signals that may occur in application spaces. They are in themselves not applications though, and in this thesis we do not develop any final application.

Similarly, it is important to note that we are not augmenting or assisting the human in these experiments. We are primarily concerned with the signals and behaviours surrounding target detection in search tasks from the subjects. Although we explore strategies that may allow a subject to search a body of images or annotate them in a more efficient manner than conventional means, we are not augmenting the subject in a way where they do not need to detect targets for them to be ultimately detected by the system. Any target detected by the system is first detected by the subject. Although this is the case, other research has explored how EEG systems like this may be combined with computer vision algorithms that incorporate strategies wherein the computer aids in the detection and prioritisation of targets utilising the neural signals to ultimately decide on the classification of an image. We do not use such computer vision approaches in this thesis, as we focus primarily on what signals are detected from the subject.

Chapter 2 While EEG represents a single sensor source of activity that can be detected from the brain, it nonetheless contains a rich variety of signals displaying modulations affected by states such as sleep or periods of high levels of concentration. Not only do these signals display indicators of state, but they also show perturbations surrounding events like the presentation of a stimulus, displaying sensitivities to stimulus parameters such as brightness, and in addition, to the content and meaning of the stimulus. These signals are utilised in a variety of paradigms to enable EEG BCI systems. In Chapter 2 I give an overview of EEG, and explain how these signals are utilised in both conventional BCI systems and newer BCI application spaces. In the final section of this chapter having established the scope for this work, I outline a central hypothesis and a set of research questions through which we examine this hypothesis in the thesis.

Chapter 3 Stereotyped EEG responses known as ERPs (Event-Related Potentials) are known to occur in response to the presentation of a stimulus such as an image. Similarly, another class of responses known as FLERPs (Fixation Locked Event Related Potentials) are known to occur relative to the time of eye movements. In Chapter 3 we examine the signals present with regard to eye movements during a variety of search tasks, and examine how we can utilise these signals to aid in target detection.

Chapter 4 In Chapter 4 we explore how we can combine eye tracking and EEG signals to improve search performance. Here we show that it can be optimal to combine both signal sources as they are complimentary. We also show that eye tracking signals tend to demonstrate better discriminative activity than EEG signals to assist in target search.

Chapter 5 In Chapter 5 we explore a number of related questions that contribute to the support of our hypothesis. Firstly, we examine what advantages are realised by using a button press response in combination with EEG signals, and how this effects using a reduced number of EEG channels. Secondly, we explore whether some images have inherent characteristics in a search task that lend them to being correctly labelled/mis-labelled. Thirdly, we examine the effect of presentation speed on our ability to discern target images from EEG signals.

Conclusions The final chapter summarises the contributions of each of the chapters within the thesis and discusses the outcome of this work. We retrospectively discuss our research questions and central hypothesis here. Following this, we discuss future work and speculate on further research questions and application spaces that this work can help direct.

Appendices In addition to these chapters, we have included a number of Appendices in the thesis. Appendix A provides an equipment overview, outlining details of the EEG systems and eye tracking technology that we used for the experiments described in this thesis. Additionally, it outlines the software processing techniques used on the raw signals such as clean-up and digitization. Appendix B outlines some conventions used within the thesis, along with algorithmic parameters not pertinent for discussion within the chapter bodies but necessary for the interested reader to get a full and complete picture of our work. Appendix C provides additional data to supplement the experiments outlined in Chapter 3 and this data is included for the interested reader who may wish to pursue a deeper exploration into our results. Appendix D provides additional data to supplement the experiments outlined in Chapter 5. Appendix E provides documentation on university ethics approval, and other miscellaneous materials.

Chapter 2

Overview of EEG Brain Computer Interfaces: Trends and Methods

It has been long since known that the brain generates electric signals and that changes in these signals can reflect aspects of cognitive and sensory processing. Over the past century from the initial discovery of these signals in humans, their detection has provided a mechanism for us to glean insight into ongoing processes within the brain. Berger (1929) was the first to show that these neurally generated electrical signals existed in humans, and displayed regularities across subjects with respect to behaviours such as closing ones eyes. It was here the process of recording these signals acquired its name Electroencephalography (EEG). Sutton et al. (1965) later revealed that not only did these signals display characteristic patterns indicative of mental states like arousal, but they also showed consistent patterns of deflections in response to sensory stimuli. More interestingly, these deflections could be modulated by events like the presentation of a stimulus as an exception to what was anticipated by the subject. While the study of these signals provided further insights into cognition, Vidal (1973) demonstrated that they could be used to allow direct communication with a computer, calling such systems BCIs (Brain Computer Interfaces). Traditionally the problems approached with BCI systems have focused on providing restoration of functionality and/or communication to people with a variety of impairing disorders such as stroke or brain damage. A question posed by the emergence of brain-computer interface technology is what scope exists for applications that could bring benefit to healthy users. Obviously for the most part enabling an able bodied subject to communicate a sentence or word through a cumbersome and slow interface without needing to move brings no real benefit. In the first section of this chapter we give an overview of BCI systems and the signals and paradigms on which they rely. Following this, we describe an emerging trend in using BCI for applications outside of those to assist disabled users. In the penultimate section, we overview some of the modern computational methods used to investigate EEG signals.

2.1 BCI Systems and their signals

In recent years the potential of using EEG signals to augment able bodied users have become more apparent. Gerson et al. (2006) were the first to demonstrate that EEG signals generated in response to images could be used to assist in sorting them, and that by using these signals they could sort images at a faster rate than say using a button press alone. Blankertz et al. (2010) further outlines a number of application scenarios that can benefit from the use of an EEG BCI including performance and mental state monitoring, as well as in augmenting media and game applications. Although strong distinctions exists between each of us, a commonality is observed in how our brain responds to sensory events and how mental states present idiosyncratic indicators on an EEG. For instance, EEG signals can be seen to change preceding movements. It is this fundamental level of similarity which enables generalized techniques within mainstream BCI to be utilised on nearly anybody. In this section we give an overview of BCI, and describe the neurological phenomena that are utilised by such systems.

2.1.1 EEG Signal Sources

EEG signals are generated by the summation of the post synaptic potentials of thousands of neurons with conducive spatial alignments that in time periods of synchronised firing give rise to potential differences on the scalp. Pyramidal neurons are thought to be the primary contributors to these detected potentials (Luck, 2005). The signals generated are typically small in the order of 0-100μv, and heavily susceptible to noise. They typically display a number of oscillatory components referred to by their frequency bands: delta rhythm (1-3 Hz), theta rhythm (5-7 Hz), alpha rhythm (8-12 Hz), beta rhythm (13-30 Hz), gamma rhythm (above 30 Hz), and mu rhythm (8-13 Hz). Facets of the neural networks responsible for producing certain patterns of these synchronised firings have been implicated in functions such as motor preparation, which for instance shows modulations of the mu-rhythm over sensorimotor areas (Pineda et al., 2000). Additional to these patterns of activity are perturbations within the EEG signals related to specific cognitive and sensory events. Of particular interest to us are those related to sensory events whose timing and content can be controlled, i.e. an image presented on a computer screen. Analysis of the EEG signal in the time domain with respect to the time of display of a particular stimulus is more commonly known as an Event Related Potential (ERP) study (Luck, 2005). Two different types of signal features are often used for BCI systems. Features from oscillatory activity can be extracted by examining amplitudes of sinusoidal components of the EEG signal at particular scalp points, and how they vary referentially and over time (for instance the mu-rhythm over sensorimotor areas). Evoked potentials on the other hand are stereotyped spatio-temporal EEG responses induced by the presentation of a stimulus such as displaying an image on a computer screen to which we measure a response.

2.1.2 Types of BCIs

Of interest to the BCI community is EEG due its relatively low cost, ease of set up, and reliable results. Multiple paradigms are used to allow communication with an EEG BCI based system. These can generally be divided into 2 categories: synchronous and asynchronous BCI. Asynchronous BCI systems are driven by for example a user's ability to modulate the amplitude of a particular frequency band (i.e. mu rhythm) in a particular set of sensors placed on the scalp. These modulations are achieved often through imagining acts such as moving a left or right hand, or perhaps mentally visualizing the rotation of an object within one's mind. These communication paradigms often necessitate some behavioural training. In an asynchronous BCI the user is the driver of a signal, which upon detection carries some explicit intention on behalf of the user for an action to be implemented by the system (display a letter on the screen, select an option, turn on/off a switch, etc). In this sense an asynchronous EEG BCI system is self-paced as users should be able to spontaneously control the system without needing to adhere to a fixed communication cycle. Synchronous BCI systems in contrast measure a user's response to a provided stimulus, wherein a subject is locked into a communication cycle and may only be able to communicate in defined time frames. One popular instantiation of this is the P300 speller system as described by Farwell and Donchin (1988). In this system a 6 x 6 matrix is displayed on screen composed of the alphabet along with other symbols. The rows and columns intensify in a random order (at a rate of 8 intensifications per second). The user is instructed to pay attention to when the letter (or symbol) which he intends to target intensifies regardless of whether it done as part of a row or column intensification. A number of these intensifications will occur, and upon each relevant intensification of the intended letter an electrical signal in response to this will be detectable on the person's scalp through the EEG apparatus. Due to the low SNR (signal to noise ratio) of these responses, it is required within this paradigm for multiple elicitations of this signal to be produced, and then with averaging, the result is that the intensifications relevant to a particular letter will have displayed a differentiated EEG signal relative to the other letters present on the screen. The elicitation of features in the EEG signal allowing communication in this way are facilitated by *a priori* knowledge of the existence of an underlying attentive mechanism that can be controlled by the user in such a way as to give rise to these differentiating signals that convey intent.

2.1.3 ERPs (Event Related Potentials)

In this thesis we are primarily concerned with the analysis of EEG signals time-locked to events such as image presentations, button presses and eye movements. Besides the ongoing oscillatory patterns of EEG activity, there are well-known stereotypical responses to stimuli called ERPs (Event Related Potentials). A time window of these ERP responses is composed typically of a number of positive and negative voltage deflections following a stimulus presentation that adheres to a stereotyped time, amplitude and spatial signature. The earliest of these ERP components are typically involved with sensory processing, with the later occurring components implicated in reflecting higher cognitive processes such as recognition (Johnson and Olshausen, 2003).

Due to the low signal to noise ratio (SNR) of these potentials, a number of signal time windows (epochs) are typically averaged to mitigate noise and reveal the underlying ERP components. Doing this allows us to reveal stable patterns of EEG activity following a stimulus, and by doing so reveal differences in the amplitude or timing of components with regard to various stimulus conditions. A number of ERP components are typically elicited with the presentation of visual stimuli such as the P1, N1, P2, N2 and the P3. The P/N prefix indicates the direction of the voltage deflection as being positive or negative, while the number is a shortened representation of the number of milliseconds the component typically occurs at (i.e. P100 shortened to P1). These component identifiers are not wholly strict in adherence to the exact timing of a component, and serve more so to act as identifiers to the general phenomena of a deflection occurring with a particular spatio-temporal signature in the context of, say, a visual image presentation.

For example, the visual N1 generally occurs around anterior sites of the scalp first and then posterior sites, but differs in latency and amplitude from task to task and from person to person (Makeig et al., 1999). Although the earlier of these ERP components can display different timing/amplitude characteristics across people and tasks, these effects tend to be more accentuated for the later occurring components. For instance, the P3 component as identified by Sutton et al. (1965) can occur in much later time windows up to 1000ms after a stimulus presentation. Typically in such instances the component may be referred to as a LPC (late positive complex), or as a P3b, which refers to a particular subcomponent of the P300 phenomena. The nature of reference and how the component is identified is often dependent on the nature and effects of interest of the experimental paradigm involved with its elicitation.

Although superficially these components may appear with a characteristic spatio-temporal signature, recent advances in computational methods have revealed that they can often be composed of a multitude of overlapping subcomponents sharing similar time and spatial characteristics. The P300, being one such class of ERP component, has had numerous subcomponents identified, and with each being implicated as being involved with different cognitive processes including target recognition and response selection (Makeig et al., 2004). The less spatially and temporally entangled of these components such as the P3a and P3b have a long history of study, and have been shown to be invokable in a number of experimental paradigms sharing a set of common characteristics involving detection of a target stimulus or detection discontinuation of a trend in a series of stimuli (Polich, 2007).

The P3 ERP is typically elicited using the oddball paradigm. This entails using two stimuli with one being less frequent (target) and the other more frequent (standard). In the case of visual stimuli these could be for example the letters X and O respectively. If we were to randomly shuffle the order of these visual stimuli, and present them in a RSVP stream to a subject, we would expect the less frequent target stimuli to elicit an oddball P3. RSVP (rapid serial visual presentation) is the presentation of a sequence of stimuli in a consistently timed fashion. There exists a three-stimulus variation of this paradigm where an additional distractor stimulus that the subject is instructed to ignore is introduced. Although the distractor stimulus in this sequence can be ignored, it nevertheless elicits another P3 subcomponent (P3a). The P300 has been observed in other paradigms involving for instance concealed information tests where subjects are told to conceal any explicit signs of recognition of a stimulus such as a familiar face (Meijer et al., 2007).

Treder and Blankertz (2010) have shown that the P300 speller although initially thought to use only P3 activity as its major source of differentiated activity has recently been shown to utilise time periods of differentiating activity that correspond with other component such as the N2. Steffensen et al. (2008) highlight that differences exist in ERP averages between males and females with regard to the processing of target and distractor stimuli, and purport that these reflect differences in the allocation of attentional resources in response to task demands between males and female. Interestingly in their target search experiment they identify a late occurring negativity typically peaking at 800ms that additionally identifies the target along with the P3. Luck and Hillyard (1994) similarly implicate additional ERP components that differentiate between target and non-target stimuli involving similar array search such as the visual N2pc known to be present when a target item is discriminated in the presence of competing distractors.

A number of studies have shown that depending on the task requirements, a variety of ERP components may be present, and further modulated by attentional strategies to allow differentiation between target and non-target stimuli. It is important to note that while stereotyped ERP responses can be expected to occur in particular experimental paradigms, they can often be accompanied by additional components depending on the task demands. In this thesis we highlight the occurrence of such phenomena with regard to our own experiences.

In addition to these components that can be modulated by recognition and target detection, there exists other earlier visual components such as the N170 that is known to be enhanced in amplitude for the presentation of faces. Although face stimuli may appear without any intended distinction made about their appearance in the context of a task, they will nonetheless elicit this response. Shenoy and Tan (2008) have demonstrated that such signals generated in response to faces can be used to drive a BCI intended to find face images where the subjects are not searching for face targets and are unaware of the true purpose of the task. Furthermore, this N170 component and others show modulations related to the perceived emotional expression of this face such as to whether it is angry (Blau et al., 2007). Olofsson et al. (2008) provide an overview of a number of a studies of ERP phenomena also known to be modulated by affective picture processing involving image sets that differ in valence (unpleasant-to-pleasant) and arousal (low-to-high).

ERP phenomena are also known to be modulated by effects like recognition of familiar objects (Miyakoshi et al., 2007). Shapiro et al. (2009) describe an effect known as the attentional blink where detection of a target may cause lapses of attention in which subsequent targets fail to be detected. Interestingly, it has been shown while certain targets fail to be reported, patterns of differentiating EEG activity occurs in response, indicating they undergo processing but fail to reach consciousness (Luck et al., 1996) What is important to note from this subsection is the wide variety of ERP signals that can be utilised by a BCI system. Some of these signals are the result of the recognition of a target, while on the other hand some are in response to stimuli displaying particular properties.

2.1.4 EEG BCI for Image Search

Sajda et al. (2010), Poolman et al. (2008), and Bigdely-Shamlo et al. (2008) all demonstrate a newer trend emerging in using BCI for image search where such systems assist subjects in search for images. All of these authors demonstrate the capability of using EEG signals to drive image search applications across a variety of search tasks encompassing in some cases learned skills of visual recognition, namely intelligence analysis of satellite imagery. While the general technique has been demonstrated to work on numerous image sets, other work has endeavoured to assess whether greater efficiencies can be achieved when combining the detection of EEG signals with behavioural responses (Huang et al., 2007). Kapoor et al. (2008) show the complementary nature between using computer vision algorithms and EEG signals in tandem. Their work demonstrates that certain visual discriminative information while not adequately captured by the computer vision algorithms can be perceivable to the subject, and thus register responses in the EEG signals. Other work has investigated the use of pupillary features such as TEPR (task-evoked pupillary response) wherein changes in pupillary dilation can be indicative of events such as the detections. Qian et al. (2009) proposes the use of such signals. Pohlmeyer et al. (2011b) describes an enhanced application scenario of using EEG-BCI for target search where a subject's neural responses are used with an adaptive computer vision system wherein images detected as being visually similar to those that arouse the strongest neural responses are prioritised to be viewed. In this way images of a target nature are more efficiently converged on within a database. In work of our own, we have shown that eye movements synchronised with presentation of images can index time periods of EEG signals. We have shown that EEG activity offset to the time of eye fixations on target objects can reveal pattern of ERP-like activity that can be used for a BCI (Healy and Smeaton (2011a)). In other work, we have shown that a reduced number of EEG channels can be used in combination with an overt behavioural response in image search applications ((Healy and Smeaton, 2011b)). Finally, we have also examined the EEG signals present in experts and non-experts in a task involving complex stimuli to which the expert was familiar and accustomed (Izzo et al., 2009a).

2.1.5 Conclusions

In this section we have introduced a number of EEG BCI paradigms whilst describing the signals and phenomena on which they rely. What is evident is that while these phenomena are utilised in traditional BCI applications to enable communication in those with a variety of impeding disorders, there exists a wealth of additional information in these EEG signals that can be utilised to assist in tasks such as labelling images, and perhaps providing further degrees of semantic and emotional interpretations of the user in response to these stimuli.

2.2 EEG BCI Methods

In this section we delve into the methods used to analyse and extract meaningful information from EEG signals in order to enable applications such as image search.

2.2.1 Analysing EEG Signals

A number of methods have been developed for the analysis of EEG signals. For instance techniques like FFT (Fast Fourier Transform) analysis have been used to reveal changes in frequency oscillations surrounding events like real and imagined movements, with techniques of this type used to drive BCI systems. In this thesis, however, we are primarily interested in the signals that occur relative to timed events such as image presentations, and hence our discussion regarding analysis of these signals will converge primarily to those used in this regard. Traditionally, ERP components were revealed by averaging epochs of EEG signal offset to the time of a stimulus presentation to reveal signal perturbations otherwise obscured by ongoing unrelated EEG activity and noise, to elucidate those that differ between conditions of stimuli, so as to disentangle and study the cognitive phenomena that give rise to them. In Figure 2.1 we show an example of such an ERP average. While the ERP averaging method has a long history of use, it has some inherent limitations like failing to adequately reveal cortical dynamics that may have complex temporal-spatial relationships. To mitigate some of these issues, techniques like the ERP image and ICA (independent component analysis) have gained widespread use (Jung et al., 2001) (Makeig et al., 2004). Recently, however, methods using machine learning principles have been demonstrated as being able to elucidate temporal and spatial relationships in EEG activity that differentiate between tasks conditions (Gerson et al., 2005). Methods such as these can capture trial-to-trial variability of components in order to provide regressors for fMRI analysis (Sajda et al., 2009) allowing the study of neural phenomena otherwise obscured by temporal smearing.

2.2.2 Single Trial Detection

Much research has been done with ERP phenomena and the relative paradigms surrounding eliciting such signals for the purpose of BCI, along with detecting these signals in what is known as single-trial detection. In single-trial detection we seek to classify signals as belonging to a particular class without directly relying on methods such as epoch averaging. A number of machine learning methods have been evaluated for the single-trial analysis of EEG data with many having a primary focus on EEG BCI. Bashashati et al. (2007) provide a thorough overview of the processing algorithms used to extract features of EEG signals that can be used to drive BCI. Lotte et al. (2007) furthermore provide a review of the classification algorithms used in BCI. Amongst the most popular methods for single-trial detection are what are known as linear classification methods. Proponents of these methods argue that they are suitable due to the simplicity of their mechanism, which aids for instance in their use in BCI applications requiring near instantaneous classification (Blankertz et al., 2011), (Clay et al., 2005). These classification schemes belong to a broader scope of methods known as supervised machine learning methods. Supervised machine learning methods require the input of class labelled data along with feature vectors. The labelled classes in our case are generally target and non-target stimuli. The feature vectors are the time-varying EEG signal values concatenated across all channels surrounding timed events such as the presentation of an image. In this way, we can train a machine learning model with examples of EEG responses, and then use this model to label unseen examples. By being able to construct models that give prediction accuracies on unseen data we can demonstrate regularities in

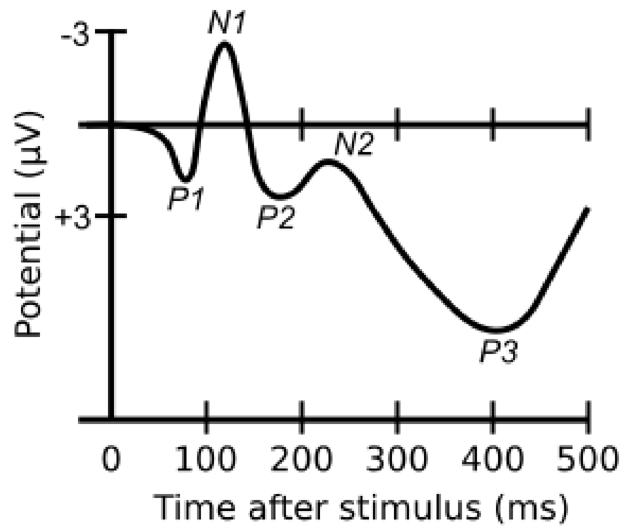

Figure 2.1: Example of ERP average

signals that can differentiate between defined classes. In this thesis, we utilise a machine learning method known as SVM (Support Vector Machine) (Chang and Lin, 2011). SVM has been shown to outperform a number of other methods for the classification of EEG data (Huang et al., 2011). We use cross-validation as part of an evaluation method in this thesis to evaluate whether signals contain discriminative information to allow us to differentiate between classes such as target or non-target. Cross-validation methods rely on a strategy of splitting a collection of labelled instances from 2 or more classes into a non-overlapping training and test sets. On each iteration of this process, a model is trained using the training set, and then evaluated on the test set. With each iteration of this, a score is derived of how well the model performed. These scores are then averaged to give an indicative measure of the level of discriminative information that exists between the instance classes. Largely we use ROC-AUC as the accuracy measure of choice in this thesis as it is insensitive to differences in class imbalances and hence derives a good measure of class separability (Fawcett, 2006).

Further details regarding the preparation of EEG signals for machine learning analysis are outlined in Appendix B.

2.3 Hypothesis and research questions

This thesis has a central hypothesis that we set out below, and explore with a number of research questions that we address in our experiments later in the thesis. This gives our work a focus and structure that allows us to clearly define its contribution in respect of the related work of others.

In this thesis we examine the hypothesis that EEG and Eye Tracking can be used to improve the effectiveness in searching for certain types of targets in images.

In subsection 2.1.4 we outlined work investigating the application of EEG BCI for image search. There are still questions as to what tasks may be fundamentally be possible with this image search methodology. A number of application scenarios have been exemplified highlighting how using EEG signals can not only allow more efficient image annotation, but that they may also allow the capture of more informative signals reflecting subtleties of interpretation.

A central aim underlying our work is to develop faster and more reliable single trial detection mechanisms to progress the goal of high throughput search. In order to do this we must firstly direct our questions at what image sets we intend to utilise this technique upon and what information we intend to extract. Applications of this technique have already been shown in proof of concept scenarios through to specialised applications where the underlying image set is domain specific and requires an expert with specialist image analyst skills. We have a difficulty in identifying which image sets this technique is applicable to, and how we can refine these image sets towards making them suitable to being used with this technique.

There are limits to the speed at which we can process information. For instance Thorpe et al. (1996) has shown that some forms of image processing can take place in less than 150ms. Additionally, some tasks require visual search, where targets may not be saliently detected (Treisman and Gelade, 1980). Eye movements may also be necessitated meaning limits on presentation speeds may need to be imposed to allow time for the deployment of fixations. Conversely, some image types can be rapidly categorised by *gist* (Oliva and Torralba, 2006). Other effects like the attentional blink must be accounted for too (Shapiro et al., 2009). Research of this kind implies that we are capable of a wide variety of visual search and target detection tasks, but also that constraints exist indicating how this might be optimally done.

Outlined below are a number of research questions that we use to explore our hypothesis as there is no particular experiment with which we can prove or disprove it.

Research question 1

What neural signals are present during visual search tasks that require eye movements, and how do they inform us of the possibilities for BCI applications utilising eye tracking and EEG *in combination* with each other?

Research question 2

How do the temporal characteristics of eye movements give indication of the suitability of a search task to being augmented by an EEG based BCI system?

Research question 3

What are the characteristics of paradigms that can be used to elicit informative neural responses to drive image search BCI applications?

Research question 4

Can we use a reduced number of EEG channels in EEG BCI search?

We return to these research questions again in Chapter 6, and discuss them further with respect to the experiments and analysis within the thesis.

2.4 Conclusions

In the first section of this chapter we gave an overview of BCI systems and the signals and paradigms on which they rely. Following this we described an emerging trend in using BCI for applications outside of those to assist disabled users. In the section after this we overviewed some of the modern computational methods used to investigate EEG signals.

In penultimate section we outline our hyptothesis and research questions.

Chapter 3

Neural Correlates of Search Involving Eye Movements

A great variety of visual search tasks exist where we seek to detect something present within an image. Some of these tasks are understood to be easy as a target may be obvious in that we know before we look that it is a target. An example of this might be finding a face in an image. On the other hand, its detection may require searching a complex image using visual cues and other information within the image to guide attention to an optimal location in which to search. An example of this would be a radiographer searching through x-ray images looking for possible tumours where detection may require discriminating between subtle features in tone and texture at various locations within the image. These search behaviours may even be further influenced by factors like expertise and expectation. Efforts to understand the mechanisms that govern the deployment of gaze and visual attention on images in search scenarios like these are ongoing and models to predict the search behaviours are continually evolving as new evidence presents itself (Ehinger et al., 2009).

For instance, it has long been known that a number of basic object features exist that allow for a target item in an image to be efficiently detected (parallel search) from an array of distractor objects (Treisman and Gelade, 1980). By confirming that additional distractor items in these arrays do not effect a subject's reaction time in detecting the target, we can conclude a target object in an image is perceived and detected through a parallel visual mechanism. The authors also contrast this with the case where target detection may require fixations upon individual items in the array (serial search). Experiments of this sort have allowed us to understand the contribution of basic visual features in visual perception, and how they are processed in the visual system.

This style of experiment though has been criticised in that it fails to capture the nature of the environment in which our visual system has evolved to perform within. Torralba et al. (2006) highlights that not just local information within an image can be used for target detection, but typically in real world search, global information such as scene statistics are used too.

Many real world search tasks involve an array of high level cognitive faculties. The integration of both bottom-up visual features coupled with the top-down deployment of attention often manifests itself in visual search scenarios as a number of fixations and eye movements on parts of an image. Previous research into the neurological correlates present during eye movements have typically focused on the early visual components present in EEG for events like eye fixations. These eye movements are known as saccades, where each saccade is a transition from one point of fixation to another on an image.

What has been shown is that there are EEG components offset to the time of these fixations called FLERPs (Fixation Locked Event Related Potentials) (Baccino and Manunta, 2005), also referred to as EFRPs (Eye Fixation Related Potentials) and SRPs (Saccade Related Potentials). Many of the scalp and timing characteristics of these components have been shown to be present across people regardless of the task being completed just so long as it involves eye movements (Ogawa et al., 2005).

Eye fixations generate a distinctive pattern of neural activity with EEG components that show consistent temporal-spatial characteristics in the same way that a visual or audio stimulus does. A lot of similarity has been drawn between these EFRPs and the early visual ERPs. For instance, they both display similar neural generators for an early positive component known as the lambda potential/visual P1 (Kazai and Yagi, 2003). Other such comparisons can be made with components like the N1, and P2 due to the similarity of their spatio-temporal onsets to those of the visual ERPs. Similarities can be drawn between ERP and EFRP components extending to later occurring activity too. The P300 for instance has been shown in attention orientation, and is known to be modulated by effects like expectation and surprise. Later occurring components like these are commonly claimed to be involved with semantic processing of stimuli, the initiation of behavioural responses, along with the detection that a particular stimulus has violated expectation of a trend (Olofsson et al., 2008); (Comerchero and Polich, 1999). It has been shown these components can also be present in EFRPs (Healy and Smeaton, 2011a). What differentiates the nature of neurological activity surrounding events like eye fixations to image presentations is that a subject is understood to be in control of their eye fixations. In an experiment employing an RSVP paradigm, a user has no control of the display time of an image, and hence no control over the timing of information availability.

This means that for visual search tasks, a user might have deployed their gaze at a location because it displayed a salient quality typically indicative of a target in an image, or global scene information or the previous fixations provided information to guide them to that location. The neural activity offset to eye fixations within an image being searched may not be independent of each other in the way that images are in an RSVP paradigm. This chapter explores what EEG activity is present both after and before fixation related events in a variety of visual search tasks involving eye movements. The primary investigation here is to understand the neural signals present in visual search and whether we can better leverage and assess the scope of applications for BCIs that may be driven by them. This chapter will also extend upon previous work by employing an information theoretic approach to quantify and compare these signals using state of the art machine learning techniques.

3.1 Brain Signals of Eye Fixations in Serial Search

The early work by Treisman and Gelade (1980) on visual search revealed that in certain instances a number of basic object features can be modulated to allow the efficient detection of target objects (parallel search) from distractor objects when displayed in array configurations. Similarly shown were object features that necessitated a user to search the array to find the target, and in such cases the task was said to require a serial search. In this section we show that discriminating EEG activity is present in serial search tasks and that it is locked to the time of fixation. In serial search we can examine a scenario where a target may only be detectable when a fixation is deployed on or near it. By elucidating EEG signals for these fixations we show that signal phenomena present in other ERP eliciting paradigms are similarly present, namely the P300. We can draw similarity here with RSVP paradigms in that the availability of the information allowing discrimination is offset to a temporal event. In the case of a RSVP paradigm this is the time an image is displayed whereas with eye movements it is the time of an onset fixation.

To this aim we conducted an experiment which is described below that required subjects to perform a search task on-screen while their EEG and EOG activity were simultaneously recorded. The object images used were constructed to be balanced in feature similarity between their target and non-targets. Using the same subjects in a follow-on experiment, we found that the target stimuli used when presented amongst arrays of increasing sizes of distracters did require longer times to detect, and that detection performance deteriorated when presentation time limitations were imposed. This further verifies that the target stimuli used could not be detected until the time of fixation.

3.1.1 Experiment Outline

In the first part of the experiment, a subject was required to search each of the 4 corners of a 24 inch (1680x1050) screen where either a target or a non-target object

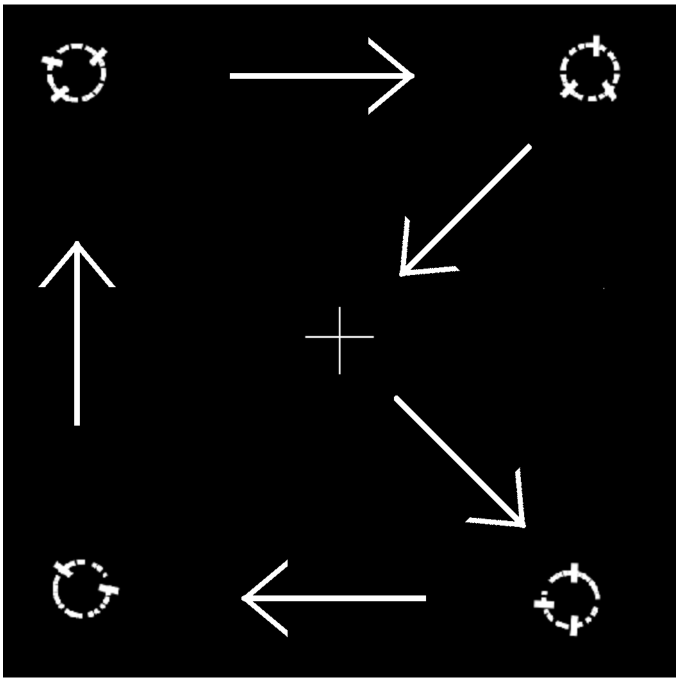

Figure 3.1: Search pattern followed to detect stimuli

was present. The experiment was designed so that when the subject's gaze remained fixed on the central fixation cross (Figure 3.1), they would remain unaware as to whether any of the objects were a target until the time of fixation. The target object to be detected and counted was a broken circle with 2 lines, while the non-targets were broken circles with 3 lines. Examples of these are illustrated in Figure 3.2. By using such stimuli we were able to restrict detection of the target item to the time of fixation. Subjects also confirmed whilst staring at the central fixation cross that they were unaware as to whether any of the corner objects were indeed targets.

The experiment was broken into 16 blocks, with each block containing 16 frames. Preceding each block, a search pattern was presented on-screen for 10 seconds without the object stimuli to indicate the route to be followed to examine the objects for that block (shown by the arrows in Figure 3.1). A white circle then appeared in the centre of the screen to indicate that a fixation cross would appear in 500 milliseconds after which the subject was to follow the given search pattern. At the end of a block, a subject then reported the total number of targets observed. Each frame was displayed for 2,500 milliseconds. Within that time, the subject was ex-

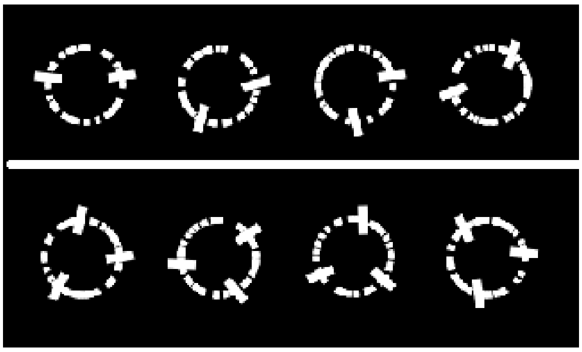

Figure 3.2: Examples of the object stimuli. Targets are shown on top, non-targets on the bottom.

pected to view all 4 corner objects following the outlined pattern, and to then return their focus to the central fixation cross. This central fixation cross would then be replaced by the warning white circle where after 500 milliseconds the fixation cross would reappear, indicating the next frame was about to appear.

The search pattern within each block was kept consistent, but changed from block to block. The arrows used to indicate the search pattern were superimposed over all frames for that block so that the subject would not forget the pattern. With A,B,C,D referencing each corner (Figure 3.1) on the screen (and E as the central fixation) we permuted this sequence to create 8 distinct search sequences, each consisting of 5 movements. 32 frames, each containing 4 corner stimuli, were then generated for that sequence. The probability of any one object stimulus being a target was kept to 0.125. Each of these 8 populated sequences were then cut in half to create the 16 blocks. In this way the target count per block could not be predicted. The order of these blocks for each subject was randomised.

3.1.2 Data collection

For data recording, we used the KT88-1016 EEG system. Signals were recorded from channels F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, Oz. A 2 channel pendant EEG device was used to record vertical and horizontal EOG (Electrooculography) signals. Subjects were seated 1.2m away from the screen. This meant each object stimulus was perceivable within 0.72 degrees.

A total of 7 subjects, 4 males and 3 females were recruited from the postgraduate and staff population with an average age of 27.3, and a standard deviation of 4.7. One of these was left-handed. Ethics approval certificate attached in Appendix E.

3.1.3 Methods

By using the EOG channels, namely VEOG (vertical) and HEOG (horizontal) we were able to find the time indexes of fixations on the object stimuli. Eye movements along one plane (i.e. horizontal) generate signals more prominently on one channel than the other, and the voltage deflections are sensitive to the direction of eye movement. Eye movements in any direction are typically characterised by either positive or negative voltage deflections on both channels. Since search patterns were consistent within blocks, the EOG patterns remained fairly consistent in that they always displayed a stereotyped sequence of deflections, other noisy EOG components were often present though. An example of a subject's eye movement search pattern for one such frame is shown in Figure 3.3. With 8 basic eye movements used across the blocks, we could detect the fixations in the EOG signals using a simple scheme of matching these deflection patterns to the movement most likely to have generated them. Deflections present in the EOG signals not conforming to the stereotyped sequence for that block were discarded. In the case of two consecutive eye movements occurring in the same direction, the second peak was taken as the fixation upon the object (the first assumed to be upon the arrow). The time at which the EOG signal(s) peaked were taken as the index time from which to extract EEG activity. The peak times were detected by finding zero-crossings of the first derivative of the signal. To mitigate against noise in the EOG signals, we disregarded eye movements where the combined absolute value of the peak height(s) fell below 2 standard deviations for that movement.

In ideal circumstances we should have been able to extract 128 target fixations, and 896 non-target fixations in total for each subject. In practice, for each subject

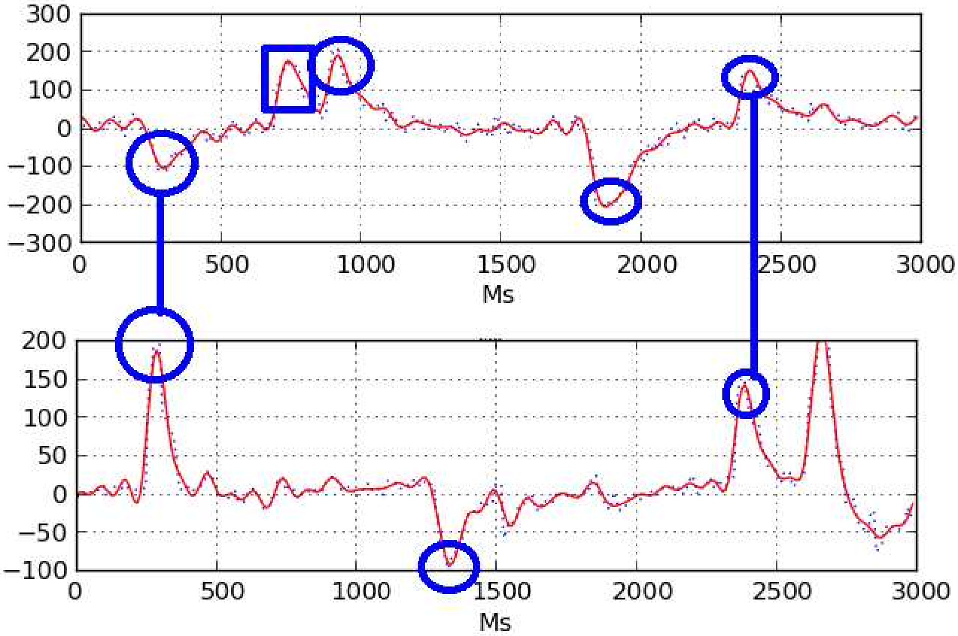

Figure 3.3: EOG Channels: HEOG (horizontal) on top and VEOG (vertical) on the bottom. Shown are peaks related to saccadic eye movements.

(1 to 7) respectively we extracted the following target/non-targets counts: 111/778, 107/768, 117/825, 109/772, 113/761, 118/838, 114/794.

Using these labelled time indexes of fixations, we extracted windows of the EEG signal starting post-fixation 0ms to 1000ms for each of the 16 channels. These were then combined to form a feature vector of length 640 which was then normalised to the range [-1,1]. No distinction was made to the eye movement associated with each target and non-target, only that that feature vector represented a target or non-target fixation.

3.1.4 Results

Both early visual EFRP and later discriminating components are visible in the grand average scalp plots shown in Figure 3.4. The first notable component is the fixation lambda potential (Kazai and Yagi, 2003) (related to the visual P100) which peaks at occipital sites at 80ms (visible on the grand average of channel Oz in Figure 3.5). Oz corresponds to an EEG sensor placement location over the occipital region of the brain. At this time a negative component was also present at frontal sites which subsequently peaked around 120ms, where it then followed a wide spatial and temporal spread continuing to 200ms. Early frontal negativities have been shown to occur in combination with the lambda potential following this time-course (Rama and Baccino, 2010), while the latter activity is consistent with the visual N1. A posterior negative component was seen across subjects typically peaking between 250ms and 350ms, and occurring later and more generally enhanced in amplitude for target objects across subjects. This activity is consistent with a posterior visual N2 in a feature discrimination task (O'Donnell et al., 1997). A positivity was seen far frontally between 280ms and 400ms peaking typically at 320ms for both object classes, and was diminished across users for targets. This diminished activity may be due to the an earlier counterpart anterior negativity related to posterior N2 activity observed for targets.

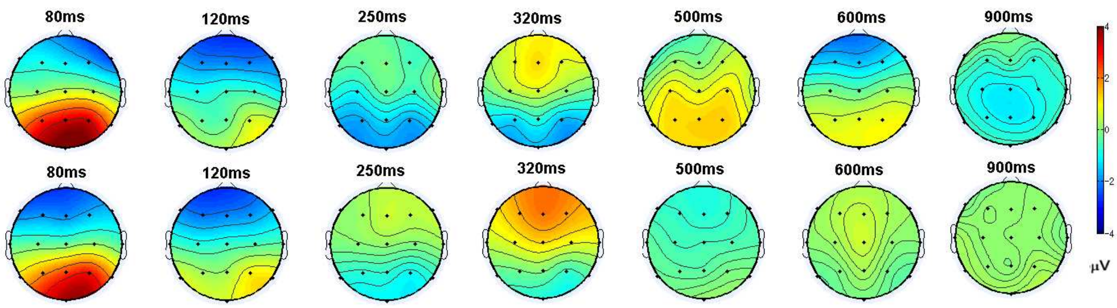

Figure 3.4: Grand average scalp plots – target plots shown on top, non-target plots shown on bottom

Differentiating activity between the detection of the target and non-target objects could be seen emerging at 250ms for most subjects, but prominent differences appear on the grand average scalp maps at 500ms with the presence of a widely distributed positive component present over occipital and parietal regions, which is consistent with P3b activity expected to occur with an oddball task such as this (Comerchero and Polich, 1999). This posterior positivity began for most subjects at 460ms and continued on to 600ms. A frontal negativity emerged for subjects for the target objects at typically 600ms (starting as the p3b activity diminished) and continued for up to 1000ms typically diminishing with a parietal distribution.

Previous work examining target detection in search tasks have shown a similar late occurring component with target detection (Sajda et al., 2010). This component may be reflective of a self-monitoring process.

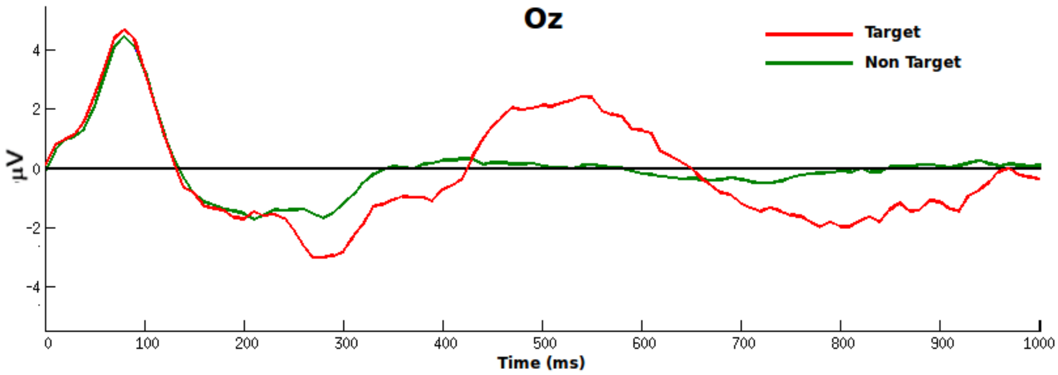

Figure 3.5: Grand average plots for channel Oz across all subjects

To examine and to derive a set of measures of the detectability of the EEG signal (P300) associated with the target fixations, we used a support vector machine (SVM) with radial basis function. A fuller account of the ML strategy described here is outlined in Appendix B. Using 20-fold cross-validation for each subject, we randomly sampled a training set of 80 target and 80 non-target examples, and then used these to train an SVM model. An independent testing set of 27 target and 27 non-target examples were randomly sampled from the remaining feature vectors. The SVM models' gamma and cost parameters were found by using a gridsearch approach on the training data only. The test sets used to generate final results were always kept seperated from the training set. For each iteration of the cross-validation, an ROC (Receiver Operating Characteristic) curve was generated and its AUC (Area Under Curve) calculated. These AUC values were then averaged and are displayed in Table 1 for each subject. The AUC measure provides a ratio independent measure of the general discriminative capability of the constructed classifier. We also formed another 3 separate feature vectors, the first using only signals from anterior nodes (F7, F3, Fz, F4, F8, VEOG, HEOG), the second using posterior nodes only (T5, P3, Pz, Oz, P4, P6) and the third using signals from all 16 channels but only extracting 600ms pre-fixation. We wanted to confirm that

Subject	AUC-All	AUC-Posterior	AUC-Anterior	AUC-PreFix
1	.74	.67	.58	.49
2	.81	.73	.56	.51
3	.79	.73	.66	.51
4	.85	.75	.66	.50
5	.74	.66	.58	.51
6	.68	.68	.55	.48
7	.68	.61	.48	.52
Average	.76	.69	.58	.5

Table 3.1: AUC results from classifiers in Serial Search experiment

the discriminative information learned by the classifier was not largely derived from the EOG activity alone (anterior sites), and that this activity only appeared after fixation. The AUC averages for these are displayed in Table 3.1.

Bootstrapping a significance AUC value for this evaluation process reveals the probability of obtaining an accuracy of .538 being <1% (p <.01). This would confirm a number of our results to be strongly significant as they are above this threshold. This bootstrapping procedure is further outlined in Appendix B.

Using the full features from all channels we obtained an average AUC of .76 across subjects. Using only signals from the frontal nodes we still obtained an above-chance classification rate, however, this lowered rate confirms that a majority of the discriminative information learned by the classifier came from posterior positioned nodes. This behaviour fits with the typical scalp topography of the P3b. No discriminative information was learned in the EEG signals pre-fixation, further confirming object detection was offset to the time of fixation.

3.1.5 Conclusions on EEG signals related to serial search

In this section we demonstrated that eye fixations are accompanied by distinct patterns of EEG activity with signal perturbations related to the processing of visual stimuli, with temporal and spatial components that can differentiate between target and non-target objects. The focus of this section was on the signals present during a search task where the visually discriminative features of the target were not perceptible until the time of fixation. This experiment demonstrated that signals related to target detection offset to the time of eye fixations can be detected, and furthermore on a single-trial basis using machine learning techniques. By being able to do this on a single-trial basis, we can annotate cognitive activity to the granularity of individual fixations.

This is confirmatory evidence of our hypothesis that EEG and Eye Tracking can be used to improve the effectiveness in searching for certain types of targets in images.

In the following sections we will extend upon what was learned in this experiment to examine the EEG signals under conditions where target stimuli are present in complex natural scenes.

3.2 Fixations in Search

While models of search are continually updated to incorporate new empirical findings, the highly diversified nature of visual search often restricts the generalisability of these models. For instance, it has been shown that subjects can learn to pre attentively detect objects with feature configurations initially implicated in necessitating serial search (Sireteanu and Rettenbach, 2000). The adaptability of the visual system in task specialisation has also shown to be modulated by factors like expertise and practice in tasks such as airport security screen (McCarley et al. (2004)) and X-Ray screening (Ericsson and Lehmann (1996)). What has been found from such research is that these increased search performances are often accompanied by more efficient search patterns of fixations.

In order to examine a wider spectrum of signals present in visual search we designed an experiment where subjects were required to search an array of pictures displayed on-screen for those containing people while an EEG and an eye-tracker were used to monitor their behaviour and reaction. In this experiment the subjects were unrestricted in the order they searched the images (free viewing paradigm), hence allowing them to employ their own search strategy. This task was designed using a natural image dataset so that it would encompass a wide variety of visual properties and scene configurations across the target and non-target examples.

In this section we describe this experiment and present results from the eye tracker. In Section 3.3 we evaluate the neural signals present around these events.

3.2.1 Experiment Outline

Subjects were required to search a screen for pictures containing people. In total each person viewed 528 frames, where each frame contained 4 pictures. Of these frames 240 contained a single target image, 268 zero target images, and 24 two target images. The experiment was broken into 24 blocks, with each block containing 22 image frames. An example of a frame is shown in Figure 3.6. All images were sampled from the SUN dataset (Xiao et al., 2010), a 100,000 image dataset containing images annotated across a variety of categories. Target and non-targets were randomly selected from this dataset in an unbiased fashion by selecting an equal number of targets and non-targets from each category. A random sampling of 240 targets and 1,680 non-targets were taken to be used as the sampling for frames containing no targets, and for frames containing one target. Another random sampling of 48 targets and 48 non-targets were taken to be used for the frames containing 2 targets.

Each subject viewed the same 2,112 images, however, the order in which they were displayed was randomised for each subject.

Prior to the start of each block, a countdown timer for 4 seconds appeared. During this time the subject was instructed to fixate centrally on the screen. After this a central white circle appeared centrally on the screen for 0.5 seconds, which was followed by an image frame. Subjects were instructed upon the appearance of the image frame to search and count those images containing people, and to try and do so in an optimal fashion. After 2.5 seconds, the images disappeared, the white circle reappeared, where the subject would redeploy their gaze to the centre of the screen and await the next frame.

Prior to the experiment a number of test runs were performed where a subject was allowed to practice the task. Throughout the length of the experiment, no subject ever viewed the same image twice.

Figure 3.6: Example of a frame where a user would search for an image containing a person

3.2.2 Data collection

For data recording, we used the KT88-1016 EEG system. Signals were recorded from channels F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, Oz. Subjects were seated between 0.5 and 0.7 metres from the screen. Eye tracking signals were also recorded using the Tobii eye tracker system. Further detail of the eye tracking system used can be found in Appendix A.

A total of 8 subjects, 5 males and 3 females were recruited from the postgraduate and staff population with ages in the range of 23 to 45. All of these were right-handed.

3.2.3 Methods

Since this was a free viewing search task a number of significant events and parameters needed to be accounted for from the eye tracking data. These included eye blinks, the onset times of fixations onto images, the offset fixation times from images, whether an image was revisited, and whether the subject began and ended the frame staring at the central fixation cross.

Frames where a user failed to fixate on any image were discarded from the data.

3.2.4 Eye Tracking Results

Shown in Table 3.2 are the total image view counts across subjects. These results indicate that for the most part subjects were able to view all the images within the presentation time.

While subjects viewed the majority of images, there were other significant events that needed to be captured from the data:

- The subject's gaze was already deployed at the location of the image when it appeared
- The subject's gaze remained deployed at the location of the image when it disappeared
- The subject entered the image while blinking
- The subject blinked during the viewing of the image
- The subject blinked leaving the image

We present the total image count across subjects excluding these events in Table 3.3.

Subject	Total	1-T	1-NT	NT	2-T	2NT
1	2069	238	703	1037	48	43
2	2049	238	682	1035	48	46
3	2089	239	706	1051	48	45
4	1808	237	572	915	47	37
5	2105	239	715	1055	48	48
6	1094	208	271	563	39	13
7	2040	232	671	1042	48	47
8	2064	239	692	1043	48	42

Table 3.2: Breakdown of the number of images viewed across the frame types. 1-T is target view counts from single target frames, 1-NT is non-target view counts from single target frames, NT is non-target view counts from non-target frames only, 2-T is target counts from frames containing 2 targets, and 2-NT is non-target view counts from frames containing 2 targets

Subject	Total	1-T	1-NT	NT	2-T	2NT
1	1844	215	617	934	45	33
2	1741	186	588	892	40	35
3	1935	224	630	1001	42	38
4	1416	208	398	749	35	26
5	1942	224	651	982	43	42
6	1019	196	251	524	36	12
7	1749	200	549	922	43	35
8	1907	218	622	996	41	30

Table 3.3: Breakdown of the number of images viewed across the frame types without noise. 1-T is target view counts from single target frames, 1-NT is non-target view counts from single target frames, NT is non-target view counts from non-target frames only, 2-T is target counts from frames containing 2 targets, and 2-NT is non-target view counts from frames containing 2 targets

Subject	1st	2nd	3rd	4th	5th	6th	7th	Total
1	81	85	40	20	0	0	0	226
2	87	59	44	19	2	0	0	211
3	163	40	16	8	0	0	0	227
4	137	55	11	10	0	0	0	213
5	145	50	22	18	0	0	0	235
6	132	43	16	7	0	0	0	198
7	65	97	38	20	1	0	0	221
8	139	51	23	11	0	0	0	224

Table 3.4: Breakdown of the order in which target images were viewed on frames containing one target excluding noisey events

In Table 3.4 we show the number of targets viewed across the order the images were viewed. Similarly, we show the same for the non-target images in Figure 3.5. What can be seen from these results is that subject are more likely to fixate on a target than a non-target with their first image fixation. While these results suggest subjects fixated on salient targets in preference, there is evidence that a search strategy was followed by subjects. In Table 3.6 we show the total counts of first fixations to each of the 4 locations on the screen. In Table 3.7 we recalibrate these values as proportions of the total number of frames in which images were viewed. Here we can see on average almost 50% of the time subjects fixated on the upper left corner as their first fixation. This is not a bias introduced by the locations of the target images in the frames as can be seen in Table 3.8. This evidence suggests that subjects employed a search strategy of following a set search path of the images, unless an image was saliently detected early in the process. Subjects were not given any particular strategy to follow, although the evidence suggests they all followed a similar search strategy.

Although subjects may have followed a search strategy part of the time, there were common target images looked at first between the subjects. These are shown in Table 3.9. This suggests there existed salient target images, and subjects may have known before fixation they were targets. Examples of these are shown in Figure 3.7.

While trends exist in which images were looked at first, little difference exists

(a) (b)

Figure 3.7: (a) Example of one of the lowest ranked targets, and (b) Example of one of the highest ranked targets

in the times to the deployment of gaze on these images from the frame onset time. In Figure 3.8 we summarise these results across subjects with a histogram, and additionally show an AUC measure of discrimination possible between these two distributions.

In Figure 3.9 we show that differences were more clearly evident for the distributions of time spent on the image between target and non-target cases (AUC=0.8, by bootstrapping a significance value we find that an AUC=0.537 has less than a 1% likelihood of chance occurrence - i.e. $p < 0.01$).

Similarly, in Figure 3.10 we show a similar analysis using the distribution of times from the frame onset to the time gaze transferred from the image (AUC=0.82, similarly with bootstrapping a significance value we find with an AUC=0.537 with a $p < 0.01$).

While most subjects followed a similar search strategy, Subject 6 employed a search strategy where they did not scan all locations, and only looked at those where they felt confident a target may lie. This is the reason for reduced images viewed count for this subject.

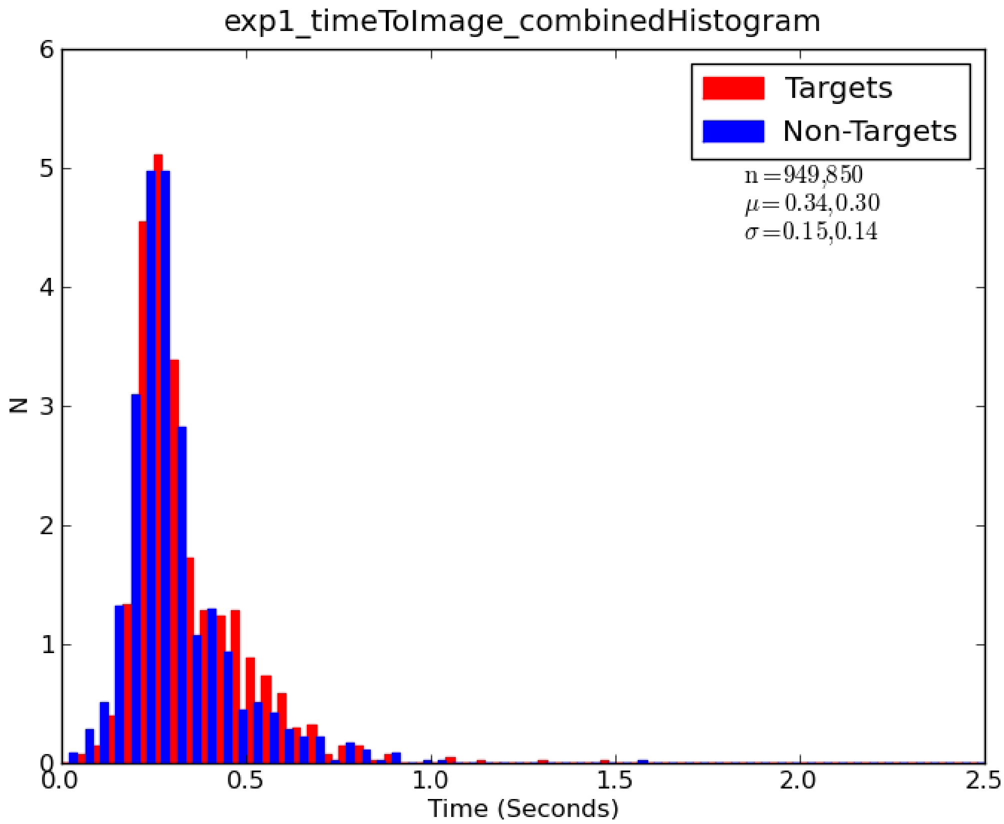

Figure 3.8: Collective histogram across subjects showing the time to the first image fixation for targets and non-targets

Subject	1st	2nd	3rd	4th	5th	6th	7th	Total
1	129	142	175	160	36	4	0	646
2	138	161	161	134	14	0	0	609
3	70	179	207	159	17	4	0	636
4	91	129	122	59	2	1	0	404
5	93	174	190	179	27	4	0	667
6	85	90	53	24	1	0	0	253
7	149	128	155	122	34	12	3	604
8	95	178	191	165	8	0	0	637

Table 3.5: Breakdown of the order in which non-target images were viewed on frames containing one target excluding noisy events

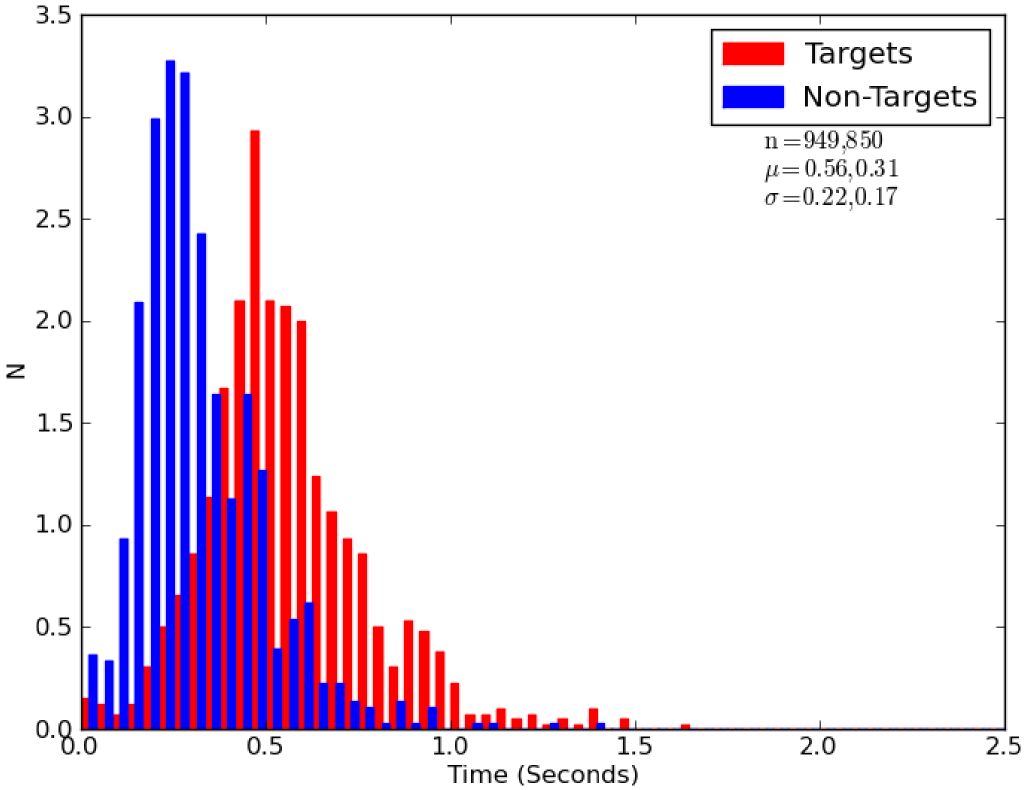

Figure 3.9: Collective histogram across subjects showing the time spent on the first image fixated upon for targets and non-targets

Subject	Location 1	Location 2	Location 3	Location 4	Total
s1	350	80	56	38	524
s2	339	76	69	41	525
s3	203	118	116	89	526
s4	231	110	103	83	527
s5	236	123	80	87	526
s6	158	55	151	124	488
s7	359	66	52	42	519
s8	159	153	107	107	526

Table 3.6: Count of image locations visited first by subjects across all frames

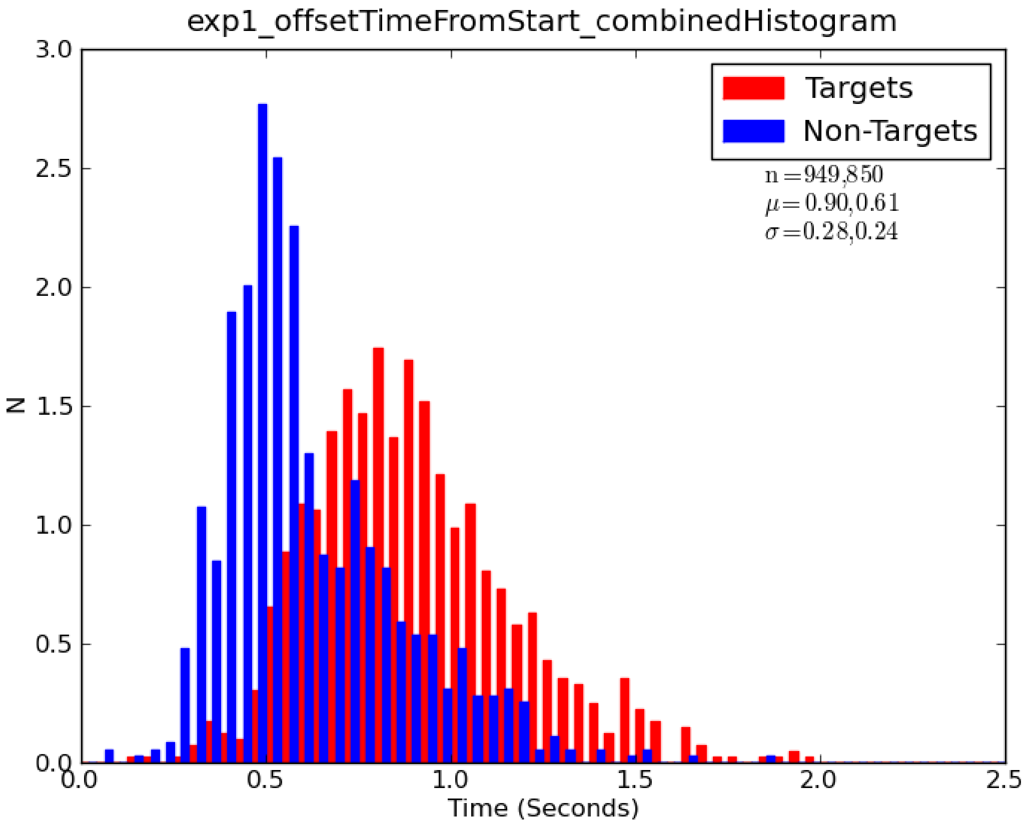

Figure 3.10: Collective histogram across subjects showing the time from frame onset to the time of fixation offset for the first image looked at for both targets and non-targets

Subject	Location 1	Location 2	Location 3	Location 4
s1	0.67	0.15	0.11	0.07
s2	0.65	0.14	0.13	0.08
s3	0.39	0.22	0.22	0.17
s4	0.44	0.21	0.20	0.16
s5	0.45	0.23	0.15	0.17
s6	0.32	0.11	0.31	0.25
s7	0.69	0.13	0.10	0.08
s8	0.30	0.29	0.20	0.20
Average	0.49	0.19	0.18	0.15
Standard Dev.	0.16	0.06	0.07	0.07

Table 3.7: Proportions of image locations viewed first across all frames

Subject	Location 1	Location 2	Location 3	Location 4
s1	70	73	63	82
s2	82	72	71	63
s3	83	74	58	73
s4	65	83	68	72
s5	77	73	73	65
s6	82	58	74	74
s7	56	81	87	64
s8	68	80	64	76
Average	72.88	74.25	69.75	71.13

Table 3.8: Distribution (random) of target image locations across all frames

Shared viewers	Count
8	5
7	19
6	33
5	41
4	57
3	37
2	31
1	14
0	3

Table 3.9: Counts of the number of targets visited in common on first fixation for frames containing one target

3.3 Discriminative Signals Present During Search

In this section we explore the neural signals present with eye fixations during the visual search task outlined in 3.2.1.

3.3.1 Analysis techniques used to identify brain signals

Recently there has been a growing interest in using machine learning techniques to analyse EEG and other neural data to understand and identify significant channels and time portions of the signal (Gerson et al., 2005) and (Philiastides et al., 2006).

In a search task necessitating eye movements there are a variety of signals and timing events which we can analyse the EEG signals in relation to. For the experiment described in Section 3.2 we might analyse EEG signals offset to the time the frame appeared on screen, the first fixation within an image, or the offset time of the saccade when the subject left the image. Furthermore, the signals generated in relation to one event may be present in the other, creating dependencies.

Additionally, eye movements can create artifacts in EEG signals creating a situation where it can be difficult to disentangle what activity comes from neural sources, and what comes from EOG sources. This is especially the case for frontal recording sites. Methods have been proposed to overcome some aspects of these limitations such as using tools like ICA (Independent Component Analysis). ICA is a case of blind source separation, and uses statistical properties of a set of signals to derive a new set of signals that are maximally independent of each other in a first and higher order statistical sense. Here, derived ICA signals implicated in eye movements can be removed, and the ICA signals can be projected back to their original space with the infringing signals removed. While ICA might be able to identify component signals of the EEG related to eye movements, the infringing signals may still remain entangled with their neural counterparts due to their time-dependent nature. In this thesis since we can obtain quantitative measures of discriminative power to aid in disentangling these signals we use a machine learning analysis alone rather than an ICA based approach.

3.3.2 Machine learning on EEG and the eye tracking signal

With the wide variety of events captured, we need a way of identifying the presence of discriminating neural signals that can occur in combination with electrical activity introduced into the EEG signal by eye movements. Differences in the distribution for dwell times on targets in comparison to non-targets exist that in turn introduce artefacts related to eye movement activity into the EEG signal.

In Figure 3.11 we show a combination of graphs offset to the same time index that can be used to identify time regions of discriminative activity and their potential sources in the EEG and eye tracking signals. Here we are looking at the discriminative activity present between targets and non-targets at the first deployment of gaze in frames containing one target. Shown in the first graph is the timeline of discriminative activity present that be detected using a SVM with a linear kernel using all the EEG signals. An overlapping sliding window approach was used, where on each iteration a 200ms time window was extracted from the available target and non-target example signals. A testing set containing 20 examples of each class was made, and the remaining examples were used to train the SVM model. For each of these iterations an AUC (Area Under Curve) measurement was taken of the ROC (Receiver Operating Characteristic) curve as the performance of the model on the test set, indicating the separability of the classes in this time region. An average of these iterations is shown on the second graph. Additionally shown on the second graph are the AUC values using the same outlined process, but for using anterior (frontal) channels: F3, F4, F7, Fz, F8 and posterior (back) channels: P3, P4, T5, Oz, T6.

On the third graph we show an average distance signal for target and non-target image fixations. Here the raw signals that are averaged are derived by measuring the distance between the successive measured points of gaze over time. In this way saccades that are time aligned will average to create deflections indicating periods of eye movement. On the second graph we also show a SVM analysis like that performed on the EEG signal for these eye movements signals (Eye signal).

On the fourth graph we show a pair of histograms plots. In this case they display the first time of fixation relative to the frame start time for the target and non-target cases. Similarly on the fifth graph we show a pair of histograms, but here we are looking at the fixation offset time from the first image viewed relative to the onset time. Expressing the data in this fashion allows us visualise the sources and strengths of differentiating activity present in the EEG and eye tracking signals.

Using the same technique we can visualise patterns of discriminatory activity when the signals are aligned to the time of the image fixation (Figure 3.12). In this case, however, differentiating activity may be present before the onset time. Shown on the fourth graph are the frame onset times, with the fifth graph showing the fixation offset times for this image. Figure 3.13 shows a similar analysis for the same dataset, but aligned to the time of the first image fixation offset. Shown on the third graph are the frame onset times, while shown on the 5th are the fixation onset times for the image.

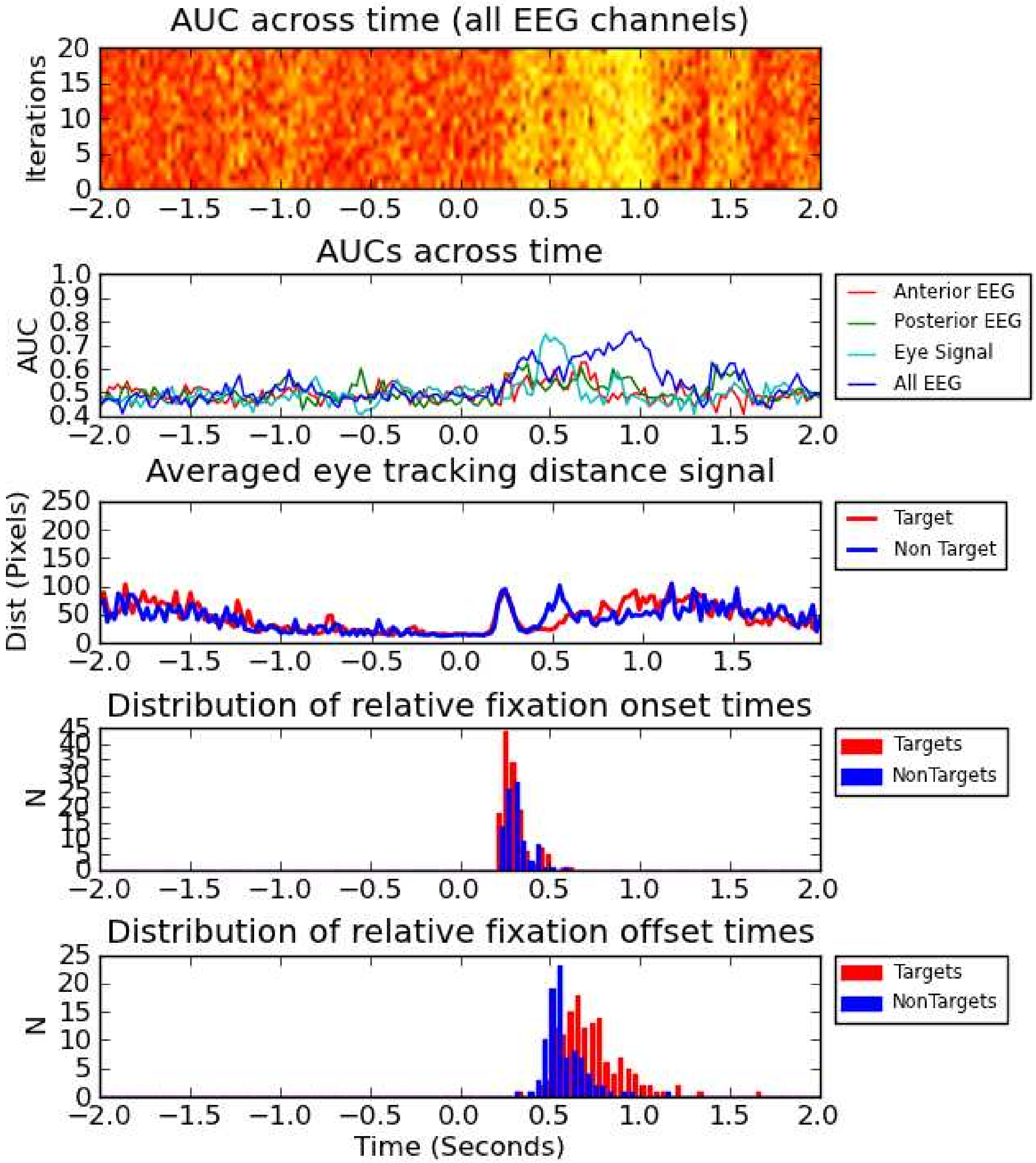

Figure 3.11: Temporally aligned discrimination example graph centred on the frame onset time showing differentiating activity related to target image detection compared to non-target image detection

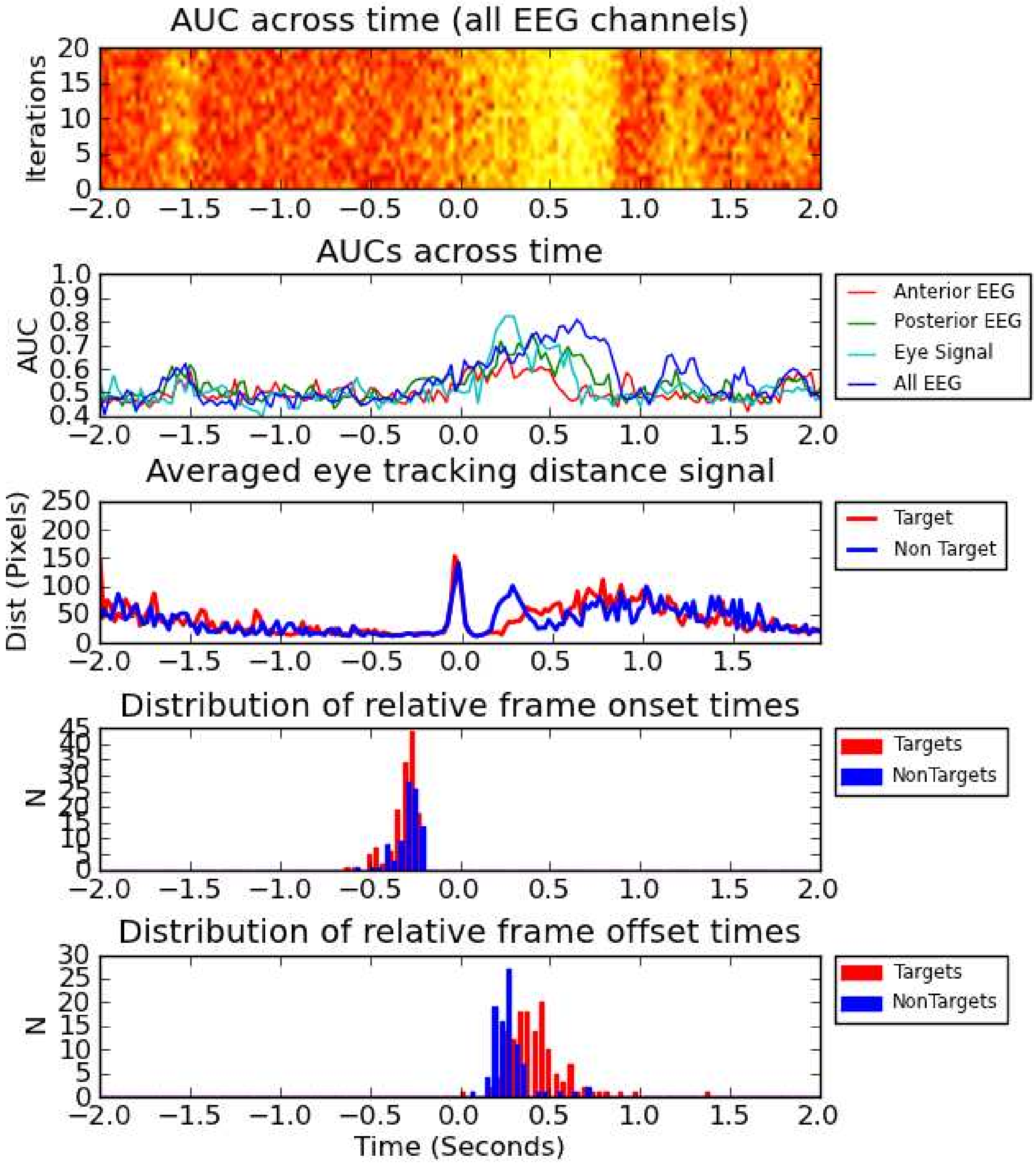

Figure 3.12: Temporally aligned discrimination example graph centred on the fixation onset time showing differentiating activity related to target image detection compared to non-target image detection

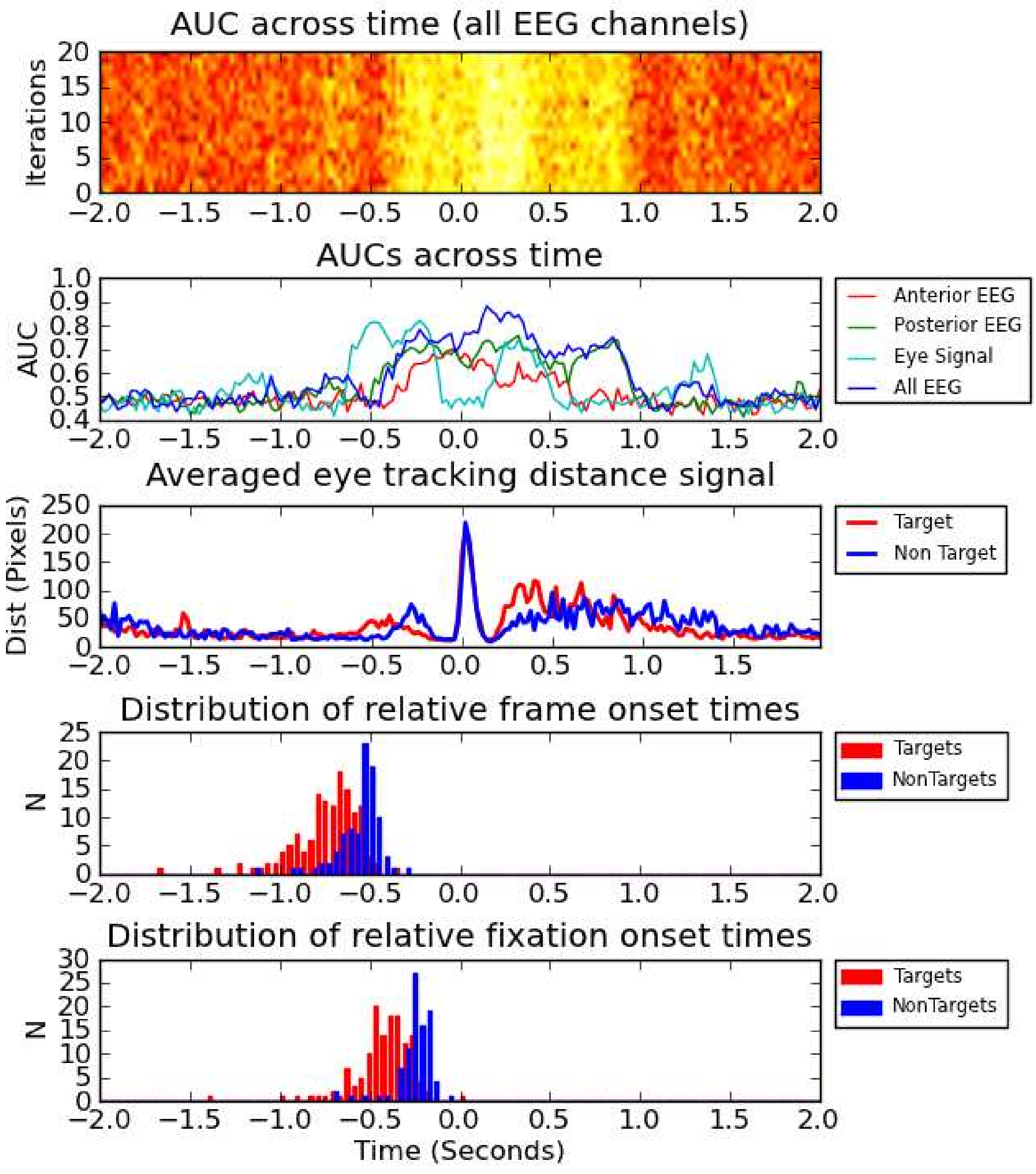

Figure 3.13: Temporally aligned discrimination example graph for subject 4 centred on the fixation offset time showing differentiating activity related to target image detection compared to non-target image detection

	EEG		Anterior		Posterior		Eye	
Subject	Time	AUC	Time	AUC	Time	AUC	Time	AUC
s1	0.45	0.69	0.5	0.74	0.5	0.66	0.4	0.67
s2	1	0.73	0.525	0.68	1	0.76	0.45	0.71
s3	1.625	0.59	0.075	0.56	0.95	0.6	1.4	0.63
s4	0.95	0.76	0.7	0.63	0.375	0.62	0.45	0.74
s5	1.05	0.77	0.825	0.65	1.075	0.7	0.75	0.62
s6	1.175	0.64	1.1	0.6	1.175	0.6	0.5	0.84
s7	0.65	0.6	0.575	0.69	0.375	0.57	0.4	0.62
s8	0.95	0.73	1.275	0.69	0.7	0.73	0.5	0.75
Average	0.98	0.69	0.70	0.66	0.77	0.66	0.61	0.70
Standard Dev.	0.35	0.07	0.37	0.06	0.32	0.07	0.34	0.08

Table 3.10: Subject's peak discrimination times from signal sources within -2,2 seconds locked to the frame onset time. Shown are peaks for anterior EEG channels, posterior EEG channels, all EEG channels, and eye tracker signal

3.3.3 Sources of Discriminative EEG activity

In the previous section we showed the discriminative activity present between targets and non-targets at the first deployment of gaze in frames containing one target. Here we will extend this by showing the time points of the EEG signal that have EEG discriminative activity. In Tables 3.10, 3.11 and 3.12 we summarise the peak times of discriminative activity within the signal sources. The temporally aligned discrimination graphs from which these figures are derived are shown in Appendix C for all subjects, and alignment times.

Across all subjects we can see the accuracy of discriminative EEG activity is highest around the time of the offset of fixation. In Table 3.14 we show that there is a greater dwell time on target images compared to non-target images, implicating motor response initiation as a source of this discriminative activity in the EEG signal. Makeig et al. (2004) provides a decomposition of the EEG signals related to decision making processes occurring in tandem with behavioural responses. Using blind source separation techniques they show that a number of underlying components are present in the EEG signal, but can be masqueraded and seen only as one component due to their overlapping time periods of activity. Using a paired two-

	EEG		Anterior		Posterior		Eye	
Subject	Time	AUC	Time	AUC	Time	AUC	Time	AUC
s1	0.25	0.73	0.25	0.73	0.375	0.71	0.25	0.72
s2	0.4	0.77	0.3	0.77	0.4	0.78	0.25	0.76
s3	0.35	0.66	1.525	0.59	0.325	0.7	0.35	0.8
s4	0.65	0.81	0.275	0.65	0.4	0.75	0.25	0.83
s5	0.325	0.69	1.675	0.66	0.725	0.66	0.5	0.6
s6	0.8	0.66	1.425	0.64	0.9	0.66	0.275	0.87
s7	0.55	0.68	0.35	0.72	0.3	0.66	0.25	0.72
s8	0.625	0.81	1.05	0.68	0.525	0.74	0.3	0.83
Average	0.49	0.73	0.86	0.68	0.49	0.71	0.30	0.77
Standard Dev.	0.19	0.06	0.63	0.06	0.21	0.05	0.09	0.09

Table 3.11: Subject's peak discrimination times from signal sources within -2,2 seconds locked to the first image fixation onset time. Shown are peaks for anterior EEG channels, posterior EEG channels, all EEG channels, and eye tracker signal

	EEG		Anterior		Posterior		Eye	
Subject	Time	AUC	Time	AUC	Time	AUC	Time	AUC
s1	0.05	0.8	-0.15	0.7	-0.15	0.75	-0.2	0.65
s2	0.175	0.78	-0.175	0.69	0.175	0.76	-0.225	0.72
s3	-0.125	0.74	-0.025	0.57	-0.125	0.7	-0.25	0.77
s4	0.15	0.88	-0.05	0.7	0.15	0.74	-0.225	0.82
s5	0.175	0.8	0.2	0.68	0.2	0.79	0.05	0.53
s6	0.2	0.81	0.075	0.53	0.15	0.8	-0.2	0.83
s7	-0.125	0.75	0.025	0.72	-0.15	0.77	-0.25	0.72
s8	0.175	0.84	0.225	0.73	0	0.79	-0.225	0.75
Average	0.08	0.80	0.02	0.67	0.03	0.76	-0.19	0.72
Standard Dev.	0.14	0.05	0.15	0.07	0.16	0.03	0.10	0.10

Table 3.12: Subject's peak discrimination times from signal sources within -.25,.25 seconds to the first image fixation offset time. Shown are peaks for anterior EEG channels, posterior EEG channels, all EEG channels, and eye tracker signal

Subject	Targets	Non Targets	All
1	0.21	0.18	0.195
2	0.28	0.26	0.27
3	0.52	0.56	0.54
4	0.31	0.3	0.305
5	0.41	0.41	0.41
6	0.32	0.28	0.3
7	0.24	0.22	0.23
8	0.28	0.27	0.275
Average	0.32	0.31	0.32
Standard Dev.	0.10	0.12	0.11

Table 3.13: Average time to first image fixation across subjects broken down by target, non-targets and total counts

Subject	Targets	Non Targets	All
1	0.56	0.29	0.425
2	0.42	0.24	0.33
3	0.63	0.35	0.49
4	0.43	0.27	0.35
5	0.56	0.46	0.51
6	0.54	0.25	0.395
7	0.6	0.35	0.475
8	0.68	0.35	0.515
Average	0.55	0.32	0.44
Standard Deviation	0.09	0.07	0.07

Table 3.14: Average time on first image fixation across subjects broken down by target, non-targets and total counts

Subject	Targets	Non Targets	All
1	0.77	0.47	0.62
2	0.7	0.5	0.6
3	1.15	0.91	1.03
4	0.74	0.57	0.655
5	0.97	0.87	0.92
6	0.86	0.53	0.695
7	0.84	0.57	0.705
8	0.96	0.62	0.79
Average	0.87	0.63	0.75
Standard Deviation	0.15	0.17	0.15

Table 3.15: Average time from the frame onset toe the first images fixation offset across subjects broken down by target, non-targets and total counts

Subject	Target counts	Non Target counts
s1	45	129
s2	54	138
s3	84	70
s4	76	91
s5	83	93
s6	77	85
s7	35	149
s8	79	95
Average	66.625	106.25

Table 3.16: The number of fixation related events for target images viewed first in common with at least 5 other people as their first fixation also, shown with the count of non-target images viewed first from frames containing one target

tailed t test on the AUCs of the posterior and anterior channels (Table 3.12), we show discriminative activity was greater across subjects at the posterior channels with t(7)=3.6056, p=0.0087. This further indicates that the sources of discriminative information learned at this time are from neural sources and not purely EOG related signals.

3.4 Salience in Search

In Section 3.3 we examined the scope of signals present around eye fixation events in search. Eye movement patterns existed indicating that some images were salient. In this section we further observe that this is the case in another search task involving a subset of the subjects from the experiment in Section 3.2.

3.4.1 Target salience experiment

In Section. 3.2.4 we presented an analysis of subjects' eye movements showing that they tended to focus their gaze on target images first. Here we further investigate differentiating signal activity surrounding deployments of gaze in salient images using a machine learning analysis similar to earlier in this chapter.

3.4.2 Experiment Outline

Subjects who participated in the experiment outlined in Section 3.2 were given the option to participate in a further experiment involving search. Of the 8 participants, 6 expressed interest and agreed to take part. The subjects were required to search images for those containing people, but unlike the first experiment only one image was presented at a time. In total each subject viewed 200 images, with 100 images containing targets (people). The experiment was broken in 10 blocks, with each block containing 20 images. The image set used for this task was a random sampling of images containing one target or no targets from the 900 images used in a previous study investigating fixations in search using natural images (Ehinger et al. (2009)). Prior to the start of each block a countdown appeared on screen where subjects were instructed to deploy their gaze centrally on the screen and to prepare to perform the task. Each image was presented for 2.5 seconds, where a centrally fixated circle then appeared for 0.5 seconds. Subjects were instructed to redeploy their gaze within this circle upon its appearance to await the next image. At the end of each block subjects reported their target counts. The image orders were randomised across all blocks for each user. While subjects viewed the same 200 images, they did not view them in the same sequence. No subject viewed the same image twice throughout the length of the experiment.

The images were comprised of urban scenes of roads with some containing people. While a number of these images contained elements such as windows, traffic, or other possible locations for targets, subjects were instructed that targets would always be obviously exposed and not obscured by elements in the image.

Targets regions were manually annotated and defined within the image as any point on or within a perimeter drawn around the outline of each target.

Data collection was performed under the same conditions as described in subsection 3.2.2 using the same equipment for EEG and eye track recording.

Figure 3.14: Target example showing geometric overlay to explain ratio distance calculation

3.4.3 Eye Tracking Results

This experiment required a different search strategy from the previous experiment with a primary difference being that frames only contained one target i.e. purposeful driven search within the scope of the task ends once the target is detected. We are primarily interested in the eye behaviour up to the point of the target detection (i.e. fixation on the target).

Shown in Figure 3.15 is a histogram of the angles of the first deployment of gaze relative to the target location for target images across all subjects. The first deployment of gaze is calculated as the first fixation over 40 pixels from the start point of the image presentation. The target point is defined as the closest point on the perimeter of the target to the first point of fixation. From this figure we can see subjects' first eye movements were towards the target locations. Although this was the case, some subjects tended to look towards the target in two stages with the first fixation in the direction of the target, and the second fixation upon the target. In Figure 3.16 we show the ratio distance left to the target on the first deployment of gaze across all subjects. Ratio distance is calculated as a translated distance left to be travelled to the nearest target perimeter point divided by the total distance to be travelled to the nearest target perimeter point from the originating start gaze point. In Figure 3.14 we show a geometric overlay with 4 marked points being S as the gaze start point, E as the first fixation point, T as the closest target point and I. I is calculated as the point of intersection for the perpendicular line from the line segment between T and S. Ratio distance to target is then calculated as the division of line segments TI/TS. This measurement was taken instead of TE/TS as the latter distance measurement was more sensitive to angle changes.

In Figure 3.17 we further summarise these results on a scatter plot generated using the angle to target measurement combined with the ratio distance left to target. Here we can see that while subjects tend look in the direction of the target with their first eye movement, they often fail to look as far as the target. In Figure 3.18 we show a scatter plot showing the trend between ratio distance and the remaining

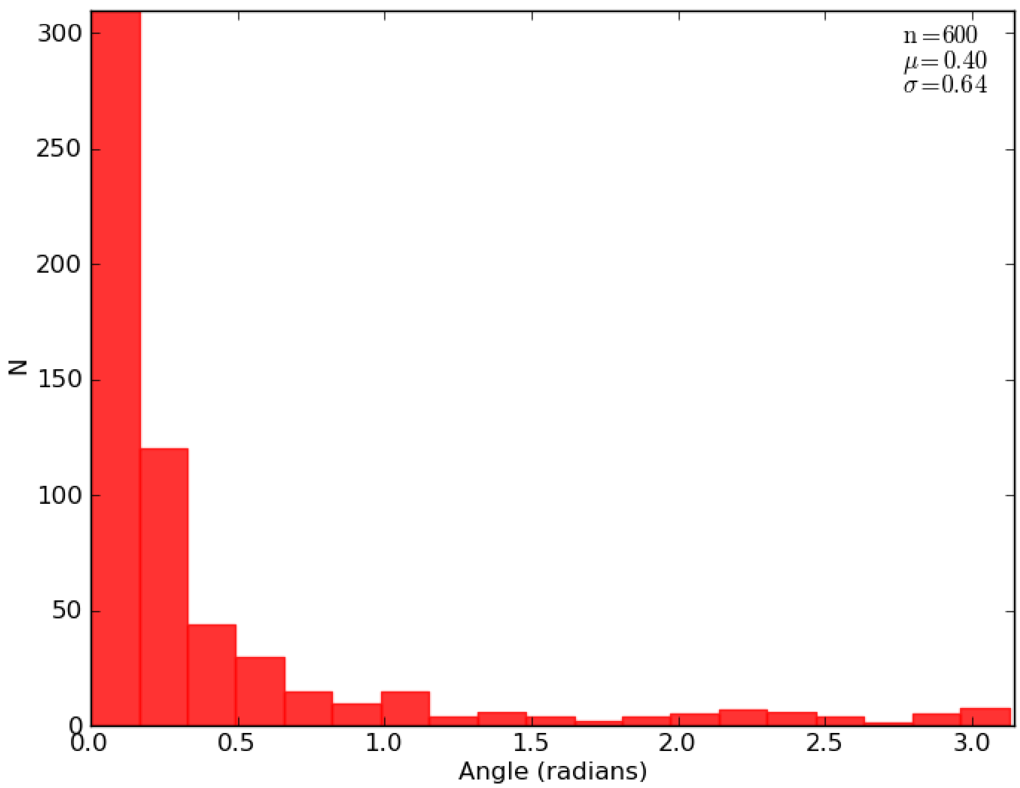

Figure 3.15: Histogram showing the angle of the first deployment of gaze relative to the target location for target images across all subjects

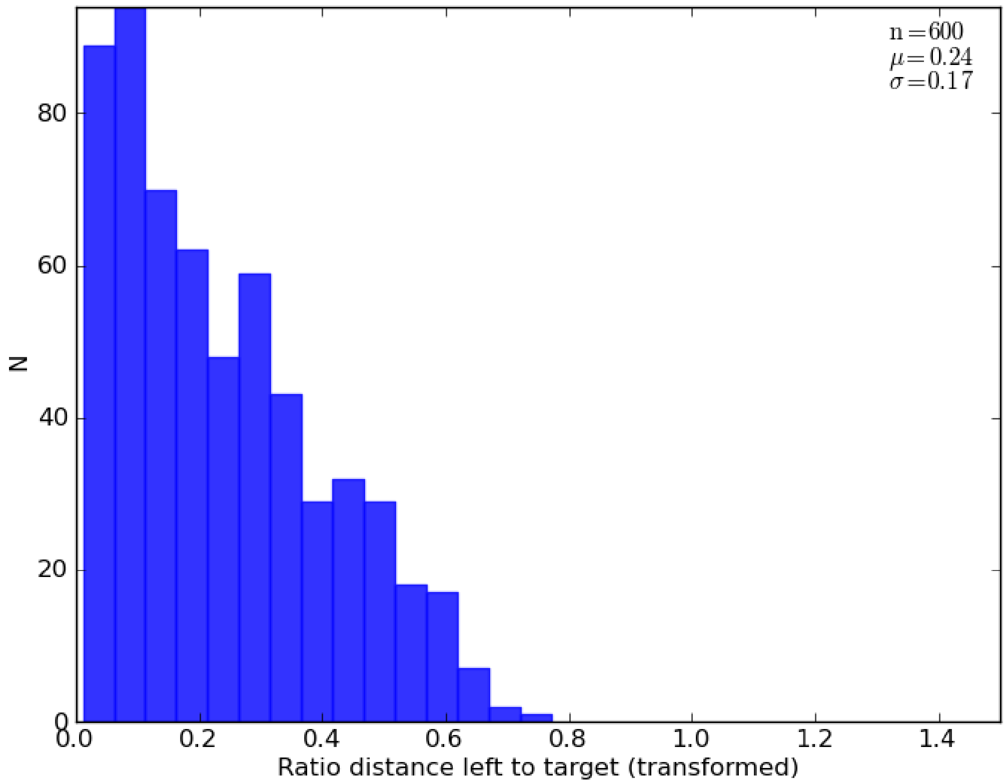

Figure 3.16: Histogram showing the ratio distance left to the target on the first deployment of gaze across all subjects

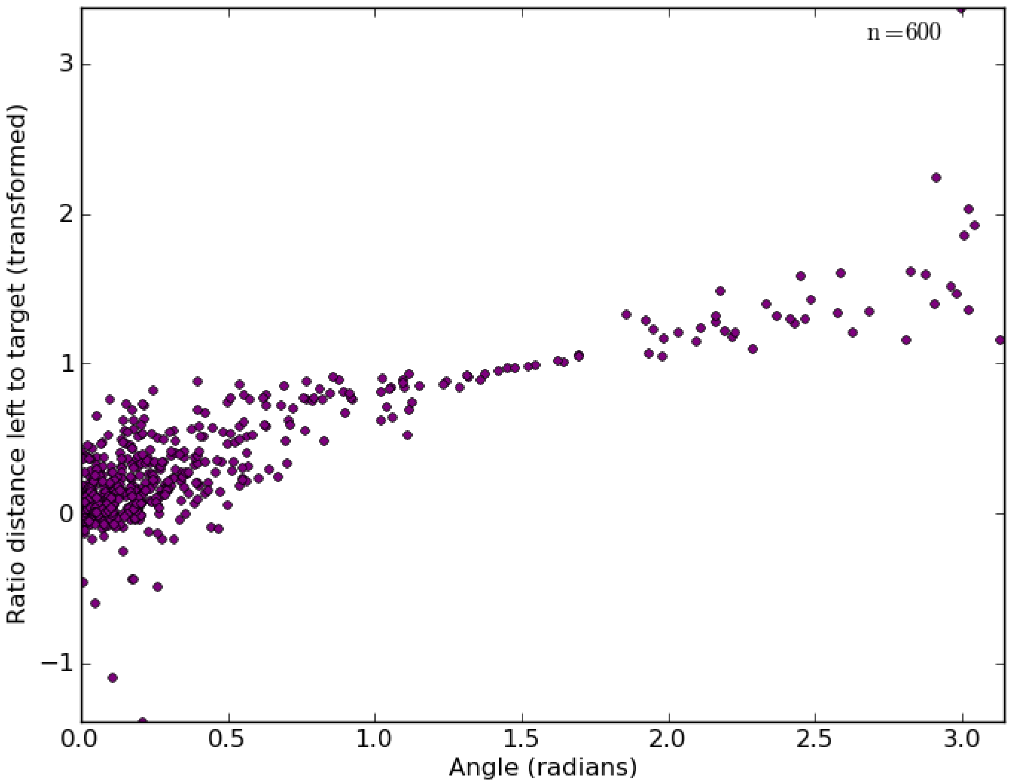

Figure 3.17: Scatter plot showing the trend between angle to target and ratio distance left to target using data from all subjects

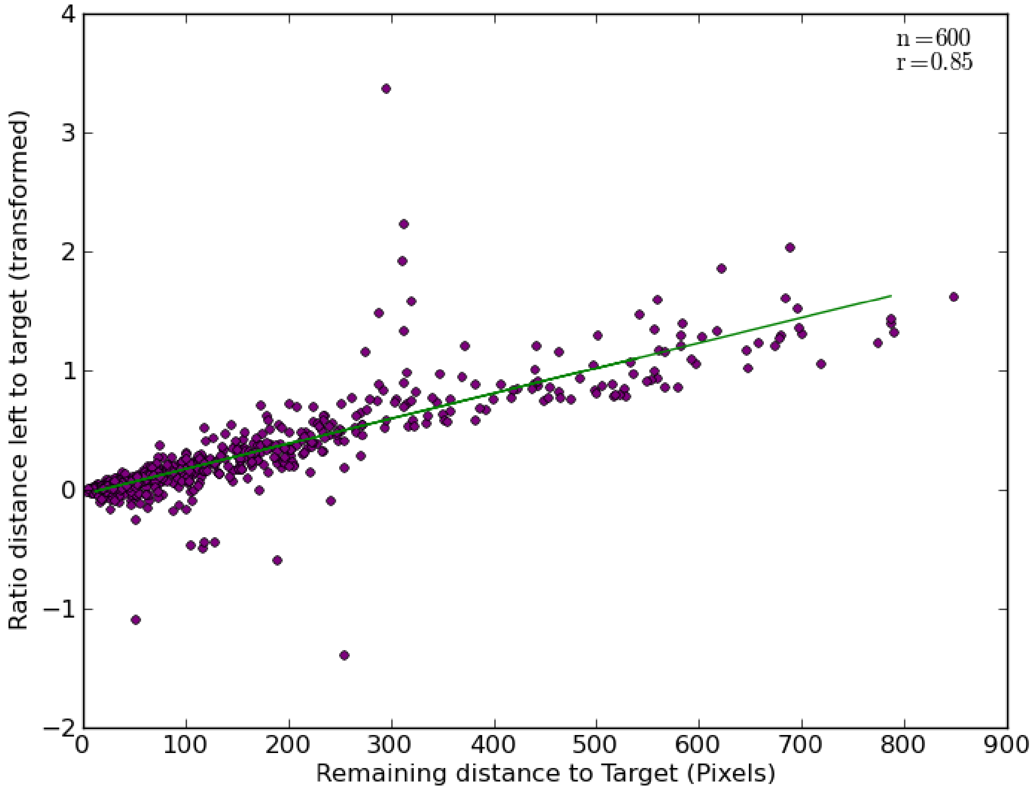

Figure 3.18: Scatter plot showing the trend between ratio distance and the remaining distance left to the target

distance left to the target. In this we can see that an increase in ratio distance has a counterpart increase in remaining distance left to target. This implies that while subjects often look in the right target direction first, they are only looking in the area of where the target is, and then with the subsequent fixations on the target.

In Figure 3.19 we can see a histogram plot showing that the distance left to target on the first deployment of gaze. From this we can see that subjects were in the vicinity of the target with the first deployment of gaze. It is the time of this first fixation we use as the target onset fixation time later when analysing onset time to the target.

In Figure 3.20 we present a histogram plot showing the distance from the start point for first deployment of gaze for target and non-targets. Here we can see that subjects tended to go further with their first fixation in the presence of a target.

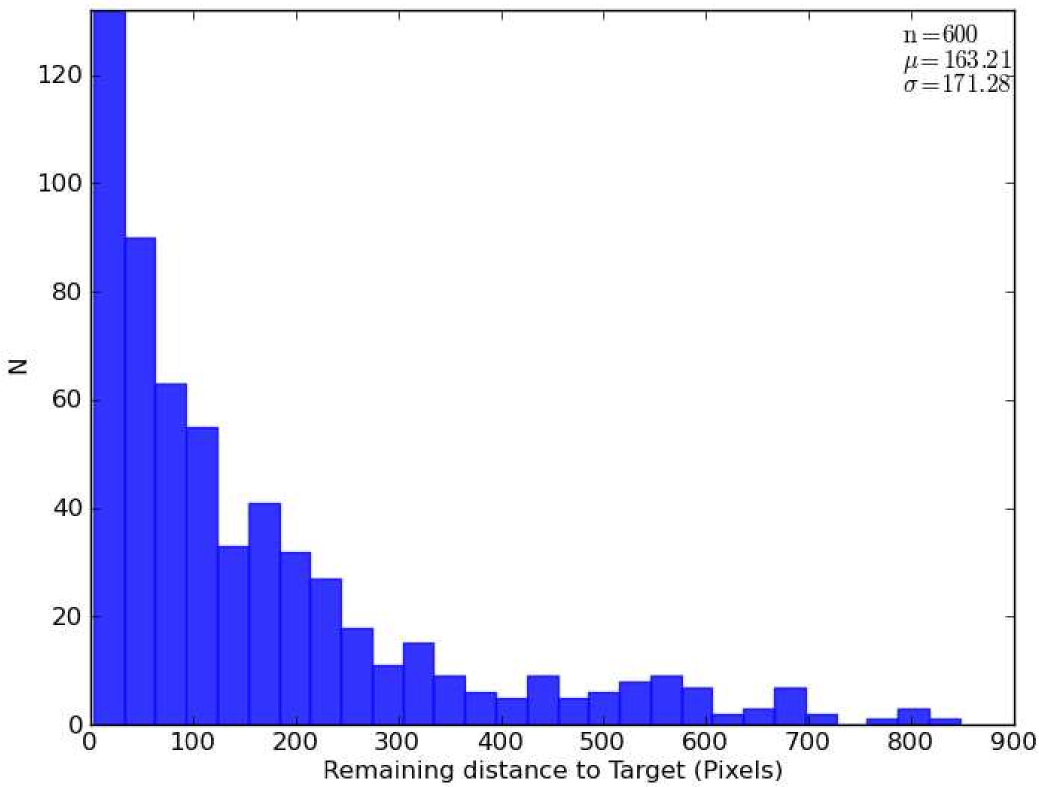

Figure 3.19: Histogram plot showing the distance left to target on the first deployment of gaze

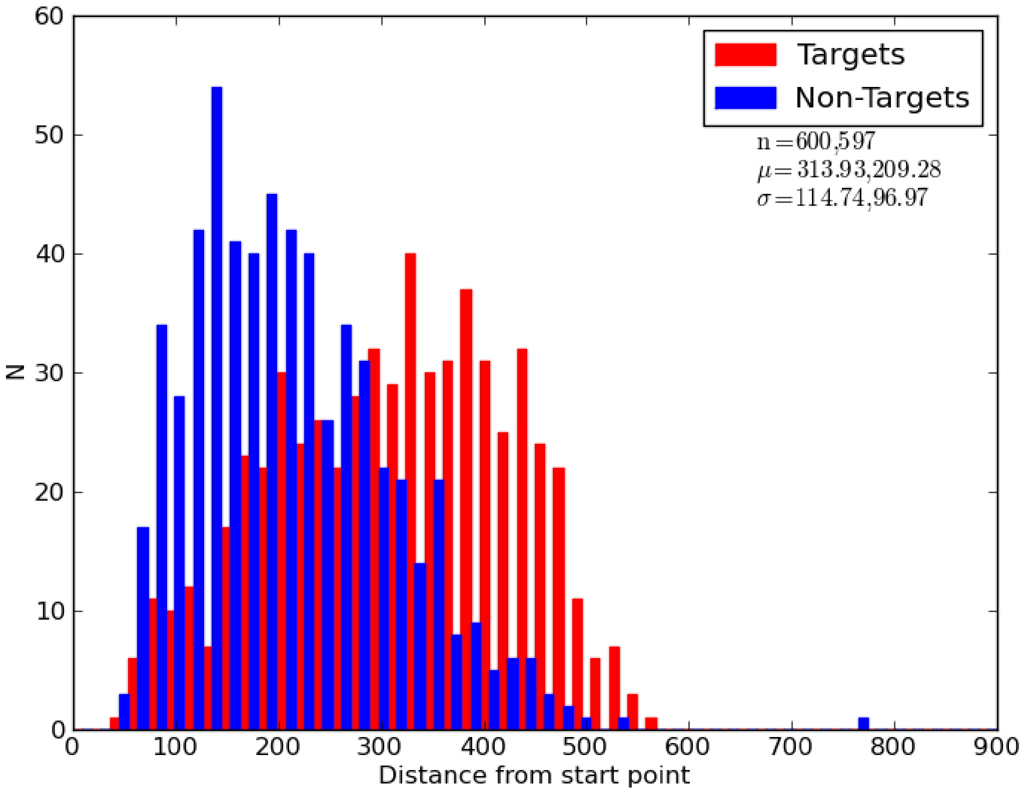

Figure 3.20: Histogram plot showing the distance from the start point for first deployment of gaze for target and non-targets

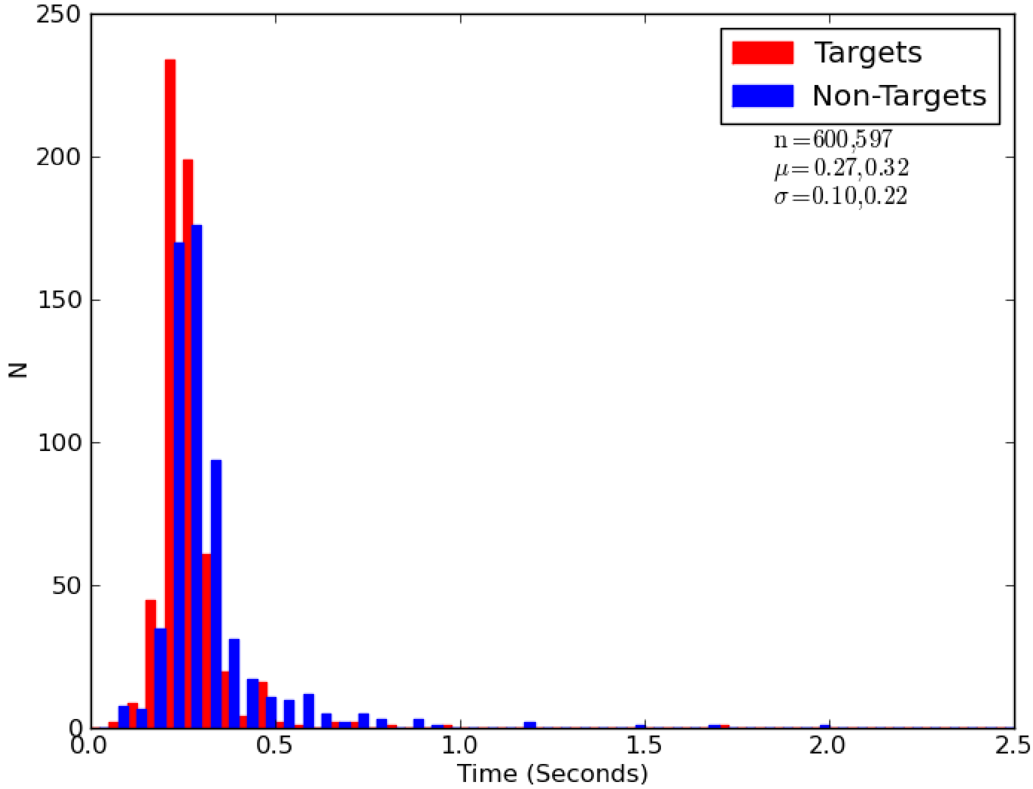

Figure 3.21: Histogram plot showing the time from the start point for first deployment of gaze for target and non-targets

While this is the case interestingly there are no significant differences in time between the time to these fixations as show in 3.21.

In Figure 3.22 we compare the distance from the starting point between the first fixations over 40 pixels in length (Method 1) with the first fixations within 180 pixels of a target point location for target frames (Method 2). What we can see here is that there are differences in the distance between these two different methods of annotating the target onset fixation. In Figure 3.23 we show that there is also a time difference to what might be considered as the target onset fixation.

3.4.4 Discriminative Signals Present During Search Task

In this subsection we show, using the same machine learning analysis used previously in Section 3.3, the temporally defined sources of discriminative information in the

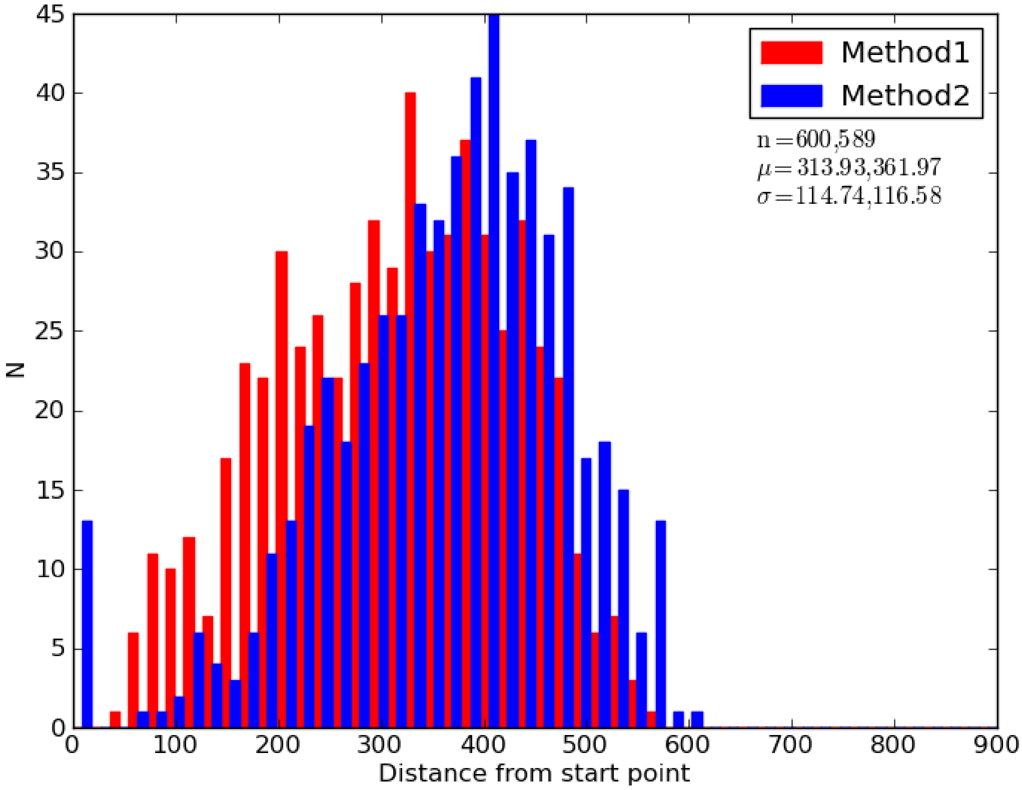

Figure 3.22: Histogram plot showing the difference in distance from the start point for target fixation between two different method to assess target onset fixation

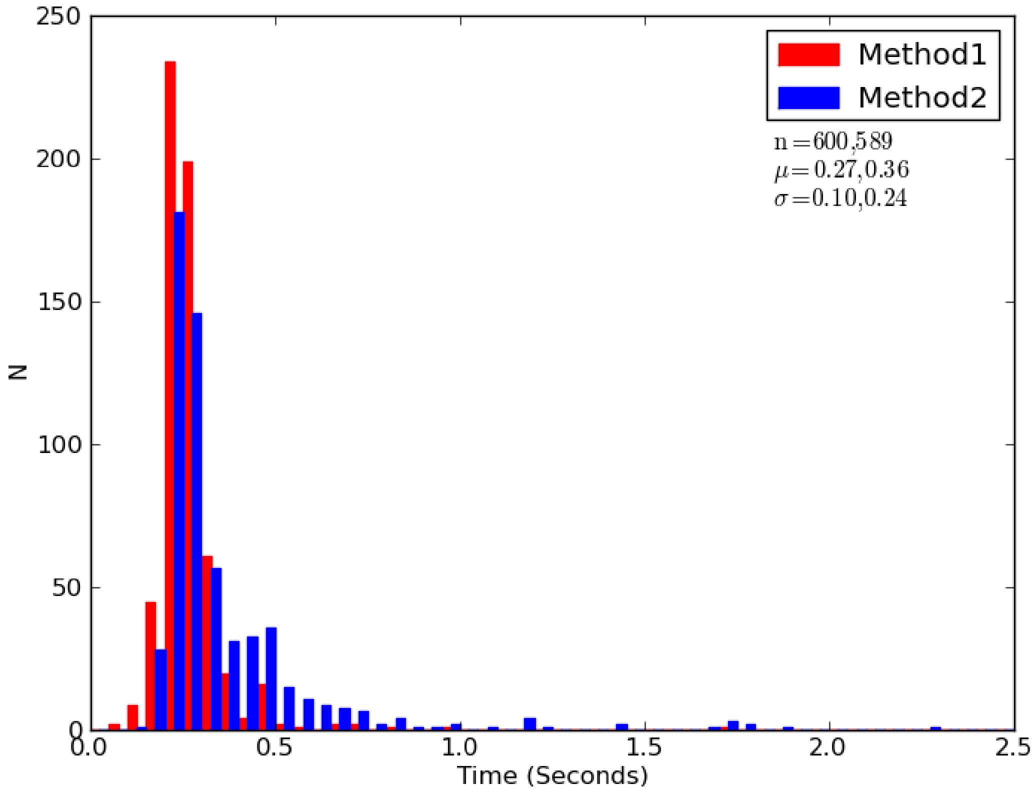

Figure 3.23: Histogram plot showing the difference in the time since frame start for target onset fixations between two different method to assess target onset fixation

EEG and eye tracking signals for this experiment. While the analysis technique used here applies the same parameters and constraints, the method used to extract the fixation onset time parameter is different from that in Section 3.3. In the former experiment we had defined spatial regions where a target could exist (i.e. image boxes) while in this experiment the target may occur at any location within the image. In the previous subsection we outlined 2 methods of extracting an onset fixation. To generate the signal discriminative maps and results in this section we used the time point of the first deployment of gaze as the target/non-target onset fixation time. A reason for this was to choose a later fixation closer to the target further confounds what we are assessing since we are introducing *a posteriori* knowledge into the process (of where the target is). Additionally, unlike the previous experiment where we had a defined offset time of when the subject left the image and deployed their gaze to assess the next image, here we do not. Subjects in the this experiment displayed various behaviours like lingering on the target, or investigating other interesting features of the image. With the wide variety of post-target detection behaviours present, assessing a fixation offset time was difficult as these behaviours were inconsistent within and across subjects. For this reason we only investigate differentiating signal activity with regard to the frame onset times and the first fixation onset times.

Whilst these search pattern behaviours have implications on the meaningful utilisation and development of systems driven by eye tracking and EEG signals, they furthermore confirm that discriminative signals are present in the various sensor modalities.

In Figure 3.24 and Figure 3.25 we show examples of these patterns of discriminative activity mapped in time for the frame onset time, and the first deployment of gaze for one subject. The remainder of these graphs are in a later Appendix C for all subjects. In addition we summarise these graphs in Table 3.17 and Table 3.18 respectively.

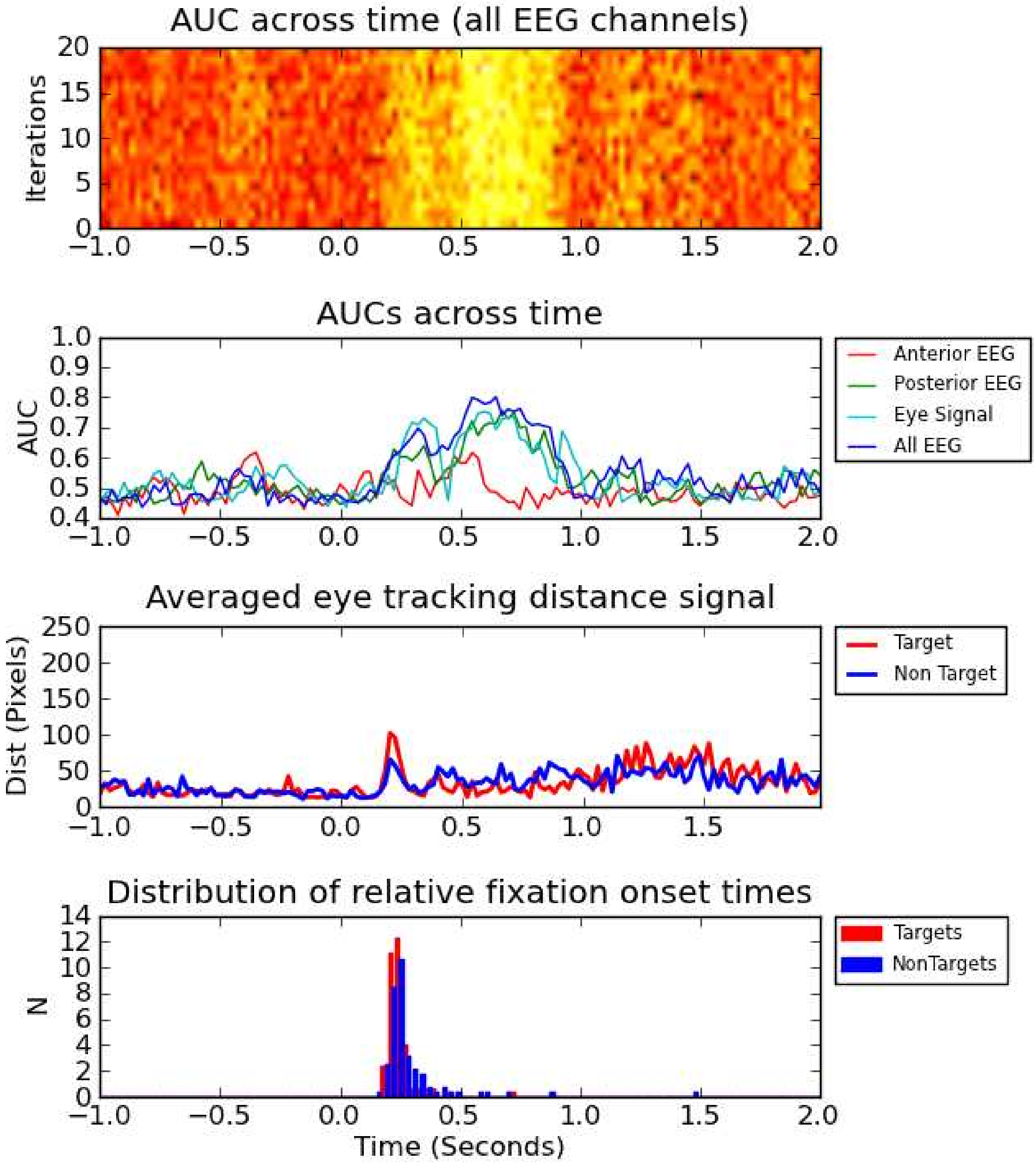

Figure 3.24: Temporally aligned discrimination graph example for subject 4 centred on the frame onset time showing differentiating activity related to target detection compared to non-target detection

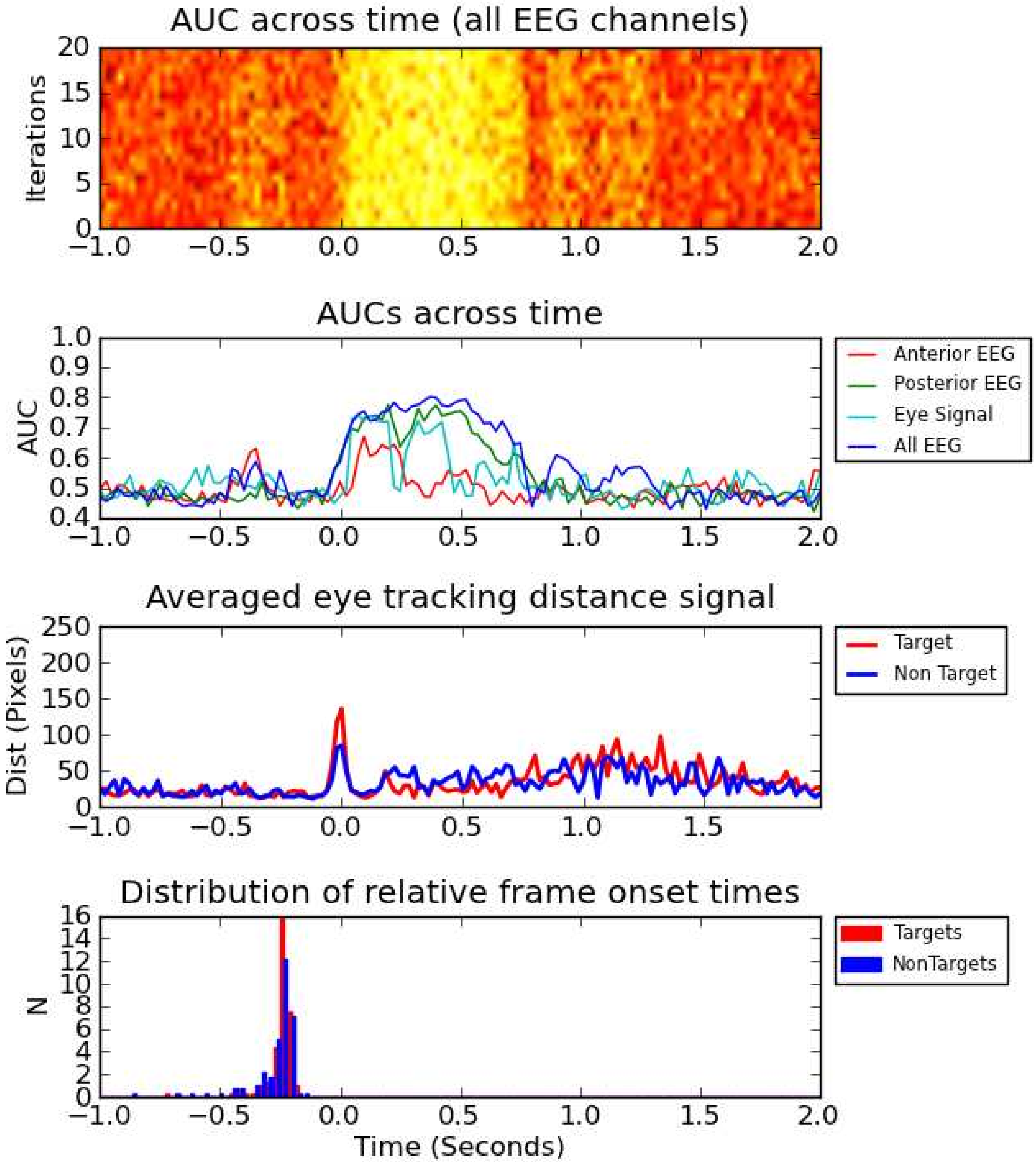

Figure 3.25: Temporally aligned discrimination graph for subject 4 example centred on the first deployment of gaze onset time showing differentiating activity related to target detection compared to non-target detection

	EEG		Anterior		Posterior		Eye	
Subject	Time	AUC	Time	AUC	Time	AUC	Time	AUC
s2	0.775	0.7	0.6	0.59	0.55	0.73	0.875	0.59
s3	0.8	0.68	1.925	0.57	0.875	0.67	0.25	0.71
s4	0.65	0.8	0.55	0.62	0.675	0.76	0.475	0.76
s5	0.825	0.7	0.025	0.53	0.975	0.58	0.85	0.56
s7	0.325	0.67	0.325	0.64	1.3	0.62	0.225	0.72
s8	0.35	0.77	1.3	0.55	0.35	0.75	0.15	0.81
Average	0.62	0.72	0.79	0.58	0.79	0.69	0.47	0.69
Standard Dev.	0.23	0.05	0.70	0.04	0.34	0.07	0.32	0.10

Table 3.17: Subject's peak discrimination times from signal sources within -2,2 seconds locked to the frame onset time. Shown are peaks for anterior EEG channels, posterior EEG channels, all EEG channels, and eye tracker signal.

	EEG		Anterior		Posterior		Eye	
Subject	Time	AUC	Time	AUC	Time	AUC	Time	AUC
s2	0.65	0.71	0.15	0.68	0.1	0.73	-0.075	0.71
s3	0.6	0.66	-0.675	0.57	0.45	0.67	0.6	0.72
s4	0.375	0.8	0.1	0.67	0.2	0.78	-0.075	0.76
s5	0.15	0.62	-1.25	0.61	1.4	0.63	1.475	0.67
s7	-1.05	0.63	0.95	0.58	1.8	0.59	-0.05	0.63
s8	0.65	0.75	-0.5	0.68	-0.025	0.76	0.275	0.79
Average	0.23	0.70	-0.20	0.63	0.65	0.69	0.36	0.71
Standard Dev.	0.66	0.07	0.77	0.05	0.76	0.08	0.61	0.06

Table 3.18: Subjects' discrimination times from signal sources within -2,2 seconds locked to the onset time of the first deployment of gaze. Shown are peaks for anterior EEG channels, posterior EEG channels, all EEG channels, and eye tracker signal.

3.5 Neural Signals in Search

In this chapter we have shown the presence of a variety of signals present in image target search from EEG and eye tracking sensors. The described experiments employed different paradigms and task constraints but all demonstrated differentiated patterns of activity from signal sources for target detection involving search with eye movements.

In the first experiment (Section 3.1) subjects were required to visit the targets in a set order, and had no prior knowledge as to whether the object was a target until fixation. In the second experiment (Section 3.2) subjects were free to search the 4 corner images in any order but many choose to follow a set search pattern. In the third experiment (Section 3.4) there was no set search pattern.

From the temporal discrimination maps presented for these experiments we may attribute sources of differentiating activity as being driven by a neural source, or by eye movements alone.

An evident property of these signals is that the discriminative time periods whether aligned to frame display time, fixation onset, or fixation offset are differentiated across subjects. This may be accounted for by differences between the subject's attentional strategies and employed search behaviours.

The spatio-temporal presence of ERP phenomena such as the P3 have been shown to be modulated by task constraints and attentional engagement strategies (Olofsson et al. (2008)). Similarly, other ERP phenomena are present like the late posterior negativity in tasks involving increased action monitoring demands (Johansson and Mecklinger (2003)). Although stereotyped responses are present for particular experimental paradigms, individual differences in attentional strategy, where allowed within the task constraints, may be responsible for invoking idiosyncratic EEG phenomena particular to that individual and/or task. For instance in Experiment 1 we observe a late posterior negativity occurring in relation to targets. An explanation for this is the task constraint of not being allowed to re-visit an object stimuli which causes a subject to call into question whether the previous object visited was a target, while simultaneously maintaining a the task requirement of visiting the other object stimuli. (Pohlmeyer et al., 2011a) observe a late differentiating component like this in an RSVP search paradigm, and suggested its presence may be due to the response locked nature of a self-monitoring process (i.e. the subject is perhaps similarly calling into question whether a previous image was a target or not).

A number of time regions for Experiment 2 and 3 were implicated as containing neural sources of discriminative information. In Appendix C we show the average scalp plots of EEG activity, along with their temporal discrimination maps.

3.6 Conclusion

In this Chapter we have shown the presence of neural signals that occur in combination with eye movements during during visual search. By understanding the paradigms and scenarios in which these signals can be elicited allows us to make informed decisions in considering the applications that may be ultimately driven by them. EEG signals used in this way can allow us to bridge a semantic gap, providing a way to measure correlates of the implicit and subjective reactions to stimuli that may be otherwise unavailable.

Although the patterns of brain activity vary across subjects, and between tasks, differentiable signals exist related to the detection and recognition of targets that can be used to drive image search applications.

In the next chapter we will evaluate strategies of how these signal sources can be used in combination with each other.

Chapter 4

Application of EEG and Eye Movement signals

In this chapter we show how both EEG and eye tracking signals can be used to assist the detection of targets as they appear in images. In the previous chapter we described a number of signals sources which we found contained discriminative information with respect to events like eye movements that allowed the detection of targets in and amongst images. Here we examine how we can combine such signals together into one integrated detection, and we investigate whether the different signal sources are complimentary when used in tandem.

4.1 Classifying Signal Sources

In the previous chapter we outlined and presented results for 2 experiments (Experiments 2 and 3) involving users in a target search task, in Section 3.2 and Section 3.4. Both EEG and eye tracking signals from our participants in these experiments demonstrated differentiable activity levels when the users viewed either target or non-target images and when these signals were locked to a number time points including frame onset, target eye fixation onset, and target eye fixation offset. In this section we present results on the classification of these signals when taken across larger time windows than what was used in the previous chapter's analysis.

4.1.1 Experiment 2 - Further Analysis

In the analysis given in the previous chapter for experiment 2 we focused on elucidating time periods of differentiating signal activity present in respect to various time events involved with selection and deployments of eye gaze in images. Here we present results for the classification of these signals when taken when we use wider time windows to capture discriminative activity.

Although our subjects had varying levels of success in selecting the target image with their first gaze deployment, some selected the target image second. Thus we think it useful to additionally present results for the second image viewed across subjects.

As a number of signal sources and time indices demonstrated discriminative activity across subjects, we set out to analyse the discriminative information present in each. In order to do this, we extracted a number of feature vectors with each offset to a particular event like a frame offset, and to a signal source like EEG. These are described in Tables 4.1, 4.2 and 4.3. In these tables we summarise the features extracted showing the signal source and time regions.

In order to classify these signals we used a machine learning algorithm known as a Support Vector Machine (SVM) with a linear kernel. In Table 4.4 we show the number of training examples available for this learning process. In order to measure classification accuracy we used a repeated random sub-sampling validation approach where on each iteration we randomly sampled a testing set of 10 target and 10 non-target instances, and then using the remainder of instances we trained the model with the largest possible equal number of target and non-targets. This model was then benchmarked upon the test set to obtain a ROC-AUC score. 20 iterations of this process were carried out, and the results were averaged. In Tables 4.5 we show the average of these ROC-AUC scores for the average of these iterations. Generating a bootstrapped AUC value for p @ .01 for this classification/evaluation

Feature Name	Source	Time Region (ms)
f1	frame offset	0,1000
f2	frame offset	0,500
f3	frame offset	0,2000
f4	image onset	-1000,0
f5	image onset	0,1000
f6	image onset	-500,0
f7	image onset	0,500
f8	image onset	-500,500
f9	image onset	-1000,1000
f10	image onset	0,2000
f11	image offset	-1000,0
f12	image offset	0,1000
f13	image offset	-500,0
f14	image offset	0,500
f15	image offset	-500,500
f16	image offset	-1000,1000
f17	image offset	0,1000

Table 4.1: List of features set names and timing sources taken from EEG signal

Feature Name	Source	Time Region (ms)
f19	frame offset	0,1000
f21	frame offset	0,500
f22	frame offset	0,2000
f23	image onset	-1000,0
f24	image onset	0,1000
f25	image onset	-500,0
f26	image onset	0,500
f27	image onset	-500,500
f28	image onset	-1000,1000
f29	image onset	0,2000
f30	image offset	-1000,0
f31	image offset	0,1000
f32	image offset	-500,0
f33	image offset	0,500
f34	image offset	-500,500
f35	image offset	-1000,1000
f36	image offset	0,1000

Table 4.2: List of features set names and timing sources taken from Eye signal

Feature Name	Source
f37	Time to fixation onset from frame onset
f38	Time spent on image
f29	Time to fixation offset from frame onset

Table 4.3: List of additional features set names and timing sources taken from Eye movement data

Subject	Targets	Non-Targets
1	81	129
2	87	138
3	163	70
4	137	91
5	145	93
6	132	85
7	65	149
8	139	95

Table 4.4: Count of frames available for each subject used for classification analysis for Experiment 2 Comparison 1

process revealed an AUC=.5745. This would indicate the probability of a selected AUC result being above this number by chance being at p=.01. This indicates strongly significant classification results.

Although some subjects performed more poorly than others on the task for Experiment 2, subjects 1 and 7 had greater counts for finding the target image with their second deployment of gaze outside of the first image. Using the described classification analysis applied to the data recorded for the first deployment of gaze, we similarly applied this analysis to the second image fixated upon (we refer to this as comparison 2). The results of this are presented in table 4.7. Non-target examples were extracted from frames containing no target images. The relative counts of training examples are provided in Table 4.6.

4.1.2 Experiment 3 - Further Analysis

In this subsection we present results from Experiment 3 using the same analysis as outlined in the previous subsection. To clarify Experiment 3 refers to the experiment

Feature	S 1	S 2	S 3	S 4	S 5	S 6	S 7	S 8	Average
f1	0.651	0.629	0.468	0.736	0.614	0.593	0.623	0.641	0.619
f2	0.653	0.472	0.497	0.588	0.509	0.596	0.513	0.474	0.538
f3	0.697	0.601	0.465	0.713	0.655	0.591	0.519	0.660	0.612
f4	0.507	0.454	0.496	0.528	0.476	0.560	0.517	0.530	0.508
f5	0.754	0.664	0.649	0.778	0.685	0.689	0.647	0.788	0.706
f6	0.493	0.486	0.466	0.547	0.532	0.483	0.542	0.422	0.496
f7	0.698	0.683	0.518	0.727	0.667	0.710	0.650	0.713	0.670
f8	0.674	0.625	0.529	0.687	0.639	0.473	0.693	0.646	0.620
f9	0.739	0.652	0.572	0.729	0.662	0.506	0.625	0.692	0.647
f10	0.738	0.651	0.591	0.753	0.676	0.578	0.575	0.763	0.665
f11	0.687	0.681	0.702	0.753	0.665	0.655	0.647	0.776	0.696
f12	0.711	0.747	0.692	0.858	0.789	0.785	0.623	0.882	0.761
f13	0.717	0.513	0.746	0.740	0.661	0.679	0.695	0.803	0.694
f14	0.758	0.807	0.595	0.905	0.789	0.810	0.559	0.842	0.758
f15	0.725	0.708	0.735	0.885	0.785	0.784	0.719	0.852	0.774
f16	0.681	0.772	0.711	0.852	0.725	0.732	0.711	0.848	0.754
f17	0.663	0.731	0.653	0.832	0.750	0.687	0.541	0.850	0.713
f19	0.714	0.655	0.503	0.719	0.590	0.839	0.581	0.773	0.671
f21	0.706	0.610	0.500	0.462	0.466	0.724	0.625	0.615	0.588
f22	0.754	0.610	0.560	0.648	0.525	0.822	0.594	0.774	0.661
f23	0.490	0.521	0.556	0.576	0.465	0.483	0.516	0.500	0.513
f24	0.726	0.665	0.781	0.835	0.612	0.889	0.711	0.861	0.760
f25	0.419	0.508	0.587	0.565	0.509	0.454	0.499	0.498	0.505
f26	0.707	0.712	0.848	0.860	0.627	0.899	0.766	0.871	0.786
f27	0.700	0.712	0.807	0.843	0.577	0.889	0.718	0.835	0.760
f28	0.635	0.657	0.733	0.818	0.578	0.851	0.651	0.789	0.714
f29	0.652	0.645	0.665	0.796	0.587	0.863	0.624	0.850	0.710
f30	0.587	0.742	0.782	0.824	0.566	0.892	0.662	0.800	0.732
f31	0.626	0.521	0.579	0.686	0.602	0.568	0.567	0.702	0.606
f32	0.669	0.753	0.730	0.812	0.588	0.865	0.736	0.835	0.748
f33	0.704	0.561	0.610	0.710	0.639	0.603	0.536	0.720	0.635
f34	0.720	0.718	0.695	0.865	0.686	0.824	0.739	0.841	0.761
f35	0.722	0.703	0.735	0.830	0.645	0.832	0.701	0.781	0.743
f36	0.629	0.529	0.683	0.687	0.625	0.556	0.487	0.623	0.602
f37	0.461	0.468	0.447	0.533	0.523	0.472	0.446	0.487	0.479
f38	0.856	0.794	0.861	0.856	0.678	0.917	0.849	0.926	0.842
f39	0.849	0.799	0.687	0.808	0.627	0.861	0.844	0.913	0.798
Max EEG	0.758	0.807	0.746	0.905	0.789	0.810	0.719	0.882	0.774
Max EYE	0.856	0.799	0.861	0.865	0.686	0.917	0.849	0.926	0.842

Table 4.5: AUCs for feature sets across subjects for Experiment 2 using Frame Set 1

Subject	Targets	Non-Targets
1	85	248
2	59	254
3	40	250
4	55	246
5	50	241
6	43	164
7	97	255
8	51	259

Table 4.6: Count of frames available for each subject used for classification analysis for Experiment 2 Comparison 2

carried out in Section 3.4.

In Experiment 3 fixation offset times were not extracted, and subsequently nor were the counterpart EEG and eye tracking signals for these time points. The feature vectors extracted use the same parameters as outlined in Tables 4.1, 4.2 and 4.3, with the exception that offset feature frames were not extracted (the former 2 tables), and only the time to gaze deployment was extracted in the latter table. In Table 4.8 we show the number of instances available for each case for classification.

In 4.9 we present the classification results for each of time locking extraction points within these signal sources.

4.2 Combining Signal Sources

In the previous section we outlined a number of signal sources that each individually provide discriminative information in regard to differentiating between target and non-targets examples in image search. In this section we present and evaluate a method to combine these signal sources.

4.2.1 Method to combine signal sources

A number of methods exist to combine classifier and signal sources with the intent of achieving an information greater than that which could be realised by using these

Feature	S 1	S 2	S 3	S 4	S 5	S 6	S 7	S 8	Average
f1	0.529	0.477	0.521	0.546	0.461	0.477	0.479	0.509	0.500
f2	0.533	0.525	0.472	0.539	0.487	0.538	0.478	0.520	0.511
f3	0.567	0.503	0.488	0.566	0.507	0.504	0.490	0.553	0.522
f4	0.512	0.572	0.535	0.454	0.493	0.527	0.626	0.494	0.527
f5	0.642	0.668	0.614	0.728	0.656	0.501	0.579	0.841	0.653
f6	0.513	0.489	0.508	0.485	0.589	0.548	0.670	0.429	0.529
f7	0.615	0.643	0.570	0.676	0.538	0.535	0.630	0.774	0.622
f8	0.641	0.534	0.532	0.660	0.562	0.493	0.626	0.691	0.592
f9	0.689	0.652	0.580	0.682	0.691	0.484	0.647	0.720	0.643
f10	0.619	0.673	0.641	0.721	0.661	0.462	0.471	0.785	0.629
f11	0.654	0.583	0.522	0.556	0.630	0.489	0.476	0.702	0.576
f12	0.617	0.634	0.497	0.518	0.696	0.505	0.567	0.737	0.596
f13	0.679	0.509	0.579	0.581	0.622	0.443	0.513	0.727	0.582
f14	0.687	0.635	0.534	0.578	0.718	0.521	0.587	0.725	0.623
f15	0.707	0.635	0.559	0.572	0.701	0.470	0.624	0.766	0.629
f16	0.694	0.628	0.515	0.586	0.701	0.438	0.524	0.733	0.602
f17	0.690	0.555	0.500	0.532	0.598	0.528	0.557	0.684	0.580
f19	0.636	0.607	0.538	0.633	0.475	0.659	0.561	0.678	0.598
f21	0.519	0.481	0.513	0.511	0.500	0.500	0.520	0.633	0.522
f22	0.635	0.497	0.529	0.590	0.674	0.591	0.483	0.695	0.586
f23	0.481	0.508	0.560	0.503	0.480	0.535	0.563	0.580	0.526
f24	0.669	0.651	0.782	0.795	0.902	0.729	0.577	0.882	0.748
f25	0.488	0.496	0.510	0.476	0.420	0.451	0.578	0.630	0.506
f26	0.747	0.724	0.824	0.845	0.921	0.749	0.624	0.883	0.789
f27	0.722	0.658	0.773	0.782	0.900	0.699	0.632	0.906	0.759
f28	0.637	0.575	0.699	0.728	0.837	0.675	0.564	0.850	0.695
f29	0.686	0.529	0.885	0.743	0.783	0.677	0.634	0.849	0.723
f30	0.609	0.738	0.801	0.739	0.847	0.665	0.573	0.889	0.732
f31	0.569	0.441	0.471	0.511	0.480	0.482	0.563	0.513	0.504
f32	0.659	0.731	0.803	0.766	0.859	0.731	0.691	0.883	0.765
f33	0.538	0.481	0.421	0.478	0.497	0.454	0.505	0.457	0.479
f34	0.655	0.671	0.713	0.752	0.816	0.658	0.660	0.870	0.724
f35	0.658	0.622	0.763	0.690	0.754	0.657	0.596	0.818	0.695
f36	0.516	0.455	0.555	0.514	0.542	0.548	0.584	0.484	0.524
f37	0.639	0.512	0.542	0.509	0.482	0.473	0.427	0.520	0.513
f38	0.819	0.839	0.833	0.836	0.923	0.853	0.767	0.930	0.850
f39	0.617	0.598	0.802	0.591	0.795	0.479	0.655	0.813	0.668
Max-EEG	0.707	0.673	0.641	0.728	0.718	0.548	0.670	0.841	0.653
Max-EYE	0.819	0.839	0.885	0.845	0.923	0.853	0.767	0.930	0.850

Table 4.7: AUCs for feature sets across subjects for Experiment 2 using frame set 2

Subject	Targets	Non-Targets
s2	100	96
s3	100	96
s4	98	94
s5	98	90
s7	99	96
s8	97	96

Table 4.8: Count of frames available for each subject used for classification analysis for Experiment 3

Feature	S 2	S 3	S 4	S 5	S 7	S 8	Average
f1	0.635	0.647	0.811	0.582	0.608	0.716	0.655
f2	0.610	0.613	0.668	0.478	0.662	0.704	0.626
f3	0.670	0.599	0.792	0.546	0.586	0.684	0.642
f4	0.476	0.507	0.435	0.501	0.467	0.595	0.490
f5	0.711	0.694	0.876	0.566	0.624	0.748	0.694
f6	0.494	0.638	0.572	0.435	0.457	0.695	0.530
f7	0.709	0.575	0.823	0.570	0.553	0.742	0.654
f8	0.629	0.629	0.764	0.559	0.613	0.745	0.647
f9	0.694	0.587	0.796	0.547	0.644	0.784	0.673
f10	0.757	0.623	0.839	0.592	0.502	0.733	0.663
f19	0.593	0.754	0.793	0.494	0.777	0.847	0.703
f21	0.669	0.687	0.708	0.496	0.774	0.721	0.687
f22	0.639	0.758	0.696	0.484	0.728	0.829	0.687
f23	0.506	0.432	0.566	0.519	0.439	0.596	0.500
f24	0.469	0.768	0.631	0.470	0.747	0.815	0.639
f25	0.435	0.544	0.650	0.513	0.480	0.662	0.524
f26	0.437	0.716	0.681	0.540	0.694	0.803	0.625
f27	0.467	0.696	0.740	0.608	0.697	0.803	0.647
f28	0.499	0.759	0.672	0.520	0.759	0.793	0.657
f29	0.484	0.743	0.670	0.500	0.718	0.778	0.636
f37	0.442	0.420	0.444	0.446	0.390	0.379	0.419
Max-EEG	0.757	0.694	0.876	0.592	0.662	0.784	0.694
Max-EYE	0.669	0.768	0.793	0.608	0.777	0.847	0.703

Table 4.9: AUCs for feature sets across subjects for Experiment 3

sources alone. In our experiments we demonstrated a number of sources which can discriminate activity from neural and eye movement signals that allowed us, with varying levels of accuracy, to differentiate between images, or regions of images that contained a search target. Here we present a method for how these can be combined.

For our approach we used a SFFS (Sequential Forward Feature Selection) scheme (Somol et al. (1999)). This iterative scheme finds subsets of features which offer an optimal discriminative capability between two classes, by starting with an empty set and adding the feature (or features) that provide the greatest increase in accuracy on each iteration. This algorithm for each forward iteration also evaluates back-steps by seeing if removing a feature offers an increase in accuracy. In this way local minima are avoided and optimal subsets are found by this floating search method.

Due to the large number of permutations that would have needed to be evaluated by the SFFS algorithm if we had included each feature source as a raw vector, we used a strategy where we jackknifed the classification predictions for each training example using classifiers trained from the remainder of the training pool. In this way we transformed an arbitrary length feature vector for each source to a feature vector containing one element (its predicted value as expressed by classifiers trained without its inclusion).

This scheme was implemented by generating a predicted score for each instance example using linear SVM models trained on the remaining training examples. For each training instance to be evaluated, a random sampling using an equal number of instances for each case of the remaining training examples was taken and used to train a linear SVM model. Since in some cases the number of training instances was low, and in order to bring stabilisation to these values, we repeated this process 5 times and averaged the predicted results for each instance. In this way, each of the newly-generated features are represented on a scale between the two classes with their relative accuracy determined by the level of discriminative information learnt using the models trained on the remaining independent instance examples available in the pool.

With this, we can construct a feature vector for each instance comprised of a single numerical value for each of the information sources (as described in Tables 4.1, 4.2 and 4.3). In order to determine optimal combinations of these features we used the SFFS algorithm on these features from EEG sources and eye tracking signal sources.

We determined AUC values for classification accuracy on the selected features by the SFFS algorithm by applying the algorithm to a training set, and then benchmarking a model generated from this training set on a testing set. Training and testing sets were selected by means of repeated random sub-sampling. Test sets contained 10 instances for each class, with training sets comprised of the maximum equal number of target and non-target instances remaining. Each iteration of the SFFS algorithm kept an additional 10 target and 10 non-target examples for testing each of its feature combinations to be evaluated from the initial training set it was given. By following this process the feature combinations selected were evaluated in an unbiased way on each iteration on an unseen testing set. 20 iterations were carried out for each subject for each profile of features to be evaluated (EEG only, Eye only, EEG + Eye). On each of these iterations the best combination of EEG and eye features were combined and evaluated on the test set. The SFFS algorithm used a linear SVM for all classification.

4.2.2 Results of combined signal sources

In Tables 4.10 and 4.11 we present results comparing the classification accuracy using the SFFS selected features for each signal source (eye and EEG) with the respective maximum AUCs that were achieved by applying a classification analysis using the individual feature sets as profiled in Tables 4.5 and 4.7 respectively for Experiment 2. We similarly present an analysis like this for Experiment 3 in Table 4.12.

From these results we can see that the SFFS algorithm selects combinations of features that in some cases have a higher accuracy than what can be achieved with a single best feature alone. This can be seen for the EEG feature sets for Experiment 2 in Table 4.10 comparing SFFS EEG and Max EEG. Here we show the SFFS algorithm is finding feature combinations that in all cases bring about a higher accuracy, with SFFS EEG having an AUC=.846 and MAX EEG AUC=.802.

In this same table though we can see that the SFFS eye features do not score better than the max SFFS features. Although this difference is small (AUC difference = .002) two additional factors need to be taken into account as to why this might be the case. Selection of a maximum score from a pool of profiled feature sets is a biased approach in that what we are selecting might just be larger by random variance than that of an equal or better feature, with this bias accumulating across subjects.

Since the SFFS algorithm initially evaluates each feature singly as part of its permutation exploration, if a single feature did score better than a combination it would have been selected as the optimal combination. Failing to do this might indicate a secondary problem with the SFFS algorithms internal mechanism needing to validate each permutation on a test set, it is utilising a smaller training set for the evaluation of each feature permutation set.

The small number of training examples available in some cases may be additionally hindering the performance of the algorithm and thus it is selecting feature combinations that may be optimal with this restricted number of training examples, but that perform worse than a single best feature alone with more training examples.

In Tables 4.13, 4.14 and 4.15 we show results for the combined accuracies from Experiment 2 set 1, Experiment 2 set 2, and Experiment 3 respectively. In the first columns we show classification scores obtained with SFFS for EEG and Eye movements. Following this we show the scores for combining the best features from both sources (as per the method outlined in subsection 4.2.1). The fourth column show the max value amongst both the EEG and Eye sources.

For Experiment 2 set 1 we can see that in all but one case the merging of the features selected by SFFS achieve an accuracy greater than either information source

Subject	SFFS EEG	Max EEG	SFFS Eye	Max Eye
1	0.790	0.758	0.859	0.856
2	0.860	0.807	0.774	0.799
3	0.787	0.746	0.878	0.861
4	0.931	0.905	0.876	0.865
5	0.829	0.789	0.759	0.686
6	0.870	0.810	0.898	0.917
7	0.787	0.719	0.779	0.849
8	0.912	0.882	0.919	0.926
Average	0.846	0.802	0.843	0.845

Table 4.10: Table comparing SFFS AUC scores for eye and EEG sources with the maximums achieved without SFFS for Experiment 2 Set 1

alone. Comparing the averaged combined scores across subjects with the average of their respective maximums shows a greater value (AUC=.879 and AUC=.870 respectively). Merging of the feature sources for Subject 2 failed to demonstrate an accuracy increase. This may be due in part to an increase in the number of features used to train the benchmarking model for the combined signal sources, providing more noise than information gained to the model. Guyon and Elisseeff (2003) describes issues of this kind in feature selection problems.

For Experiment 2 set 2 we show the average of the AUC for combined signal sources across subjects being lower than the maximum achieved in either (AUC=.839 and AUC=.845 respectively). This lowered accuracy may be due in part to the issues we describe in subsection 4.2.1 with having a reduced of training examples for the algorithm to learn from and correctly benchmark with. We, however, show demonstrate an increased accuracy with 4 of the 8 subjects.

For Experiment 3 we show the average of the AUC for combined signal sources across subjects being higher than the maximum achieved in either (AUC=.788 and AUC=.777 respectively). While this is the case, only 3 of the 6 subjects show display this increase.

Subject	SFFS EEG	Max EEG	SFFS Eye	Max Eye
1	0.777	0.707	0.824	0.819
2	0.718	0.673	0.820	0.839
3	0.605	0.641	0.881	0.885
4	0.729	0.728	0.814	0.845
5	0.773	0.718	0.940	0.923
6	0.625	0.548	0.768	0.853
7	0.630	0.670	0.781	0.767
8	0.894	0.841	0.932	0.930
Average	0.719	0.690	0.845	0.857

Table 4.11: Table comparing SFFS AUC scores for eye and EEG sources with the maximums achieved without SFFS for Experiment 2 Set 2

Subject	SFFS EEG	Max EEG	SFFS Eye	Max Eye
2	0.748	0.757	0.606	0.669
3	0.664	0.694	0.845	0.768
4	0.849	0.876	0.810	0.793
5	0.585	0.592	0.589	0.608
7	0.664	0.662	0.750	0.777
8	0.816	0.784	0.883	0.847
Average	0.721	0.727	0.747	0.743

Table 4.12: Table comparing SFFS AUC scores for eye and EEG sources with the maximums achieved without SFFS for Experiment 3

Subject	EEG	Eye	Combined	Max
1	0.790	0.859	0.872	0.859
2	0.860	0.774	0.834	0.860
3	0.787	0.878	0.902	0.878
4	0.931	0.876	0.946	0.931
5	0.829	0.759	0.840	0.829
6	0.870	0.898	0.905	0.898
7	0.787	0.779	0.808	0.787
8	0.912	0.919	0.928	0.919
Average	0.846	0.843	0.879	0.870

Table 4.13: Table comparing the SFFS scores for EEG, Eye and same combined for Experiment 2 Set 1

Subject	EEG	Eye	Combined	Max
1	0.777	0.824	0.855	0.824
2	0.718	0.820	0.773	0.820
3	0.605	0.881	0.890	0.881
4	0.729	0.814	0.769	0.814
5	0.773	0.940	0.938	0.940
6	0.625	0.768	0.779	0.768
7	0.630	0.781	0.750	0.781
8	0.894	0.932	0.955	0.932
Average	0.719	0.845	0.839	0.845

Table 4.14: Table comparing the SFFS scores for EEG, Eye and same combined for Experiment 2 Set 2

Subject	EEG	Eye	Combined	Max
2	0.748	0.606	0.744	0.748
3	0.664	0.845	0.837	0.845
4	0.849	0.810	0.898	0.849
5	0.585	0.589	0.618	0.589
7	0.664	0.750	0.744	0.750
8	0.816	0.883	0.885	0.883
Average	0.721	0.747	0.788	0.777

Table 4.15: Table comparing the SFFS scores for EEG, Eye and same combined for Experiment 3

4.3 Conclusion

In this chapter we have shown how signals recorded from EEG and eye tracking sensors can be used to allow us to discriminate images or regions there within containing targets. The results indicate that both of these sensor sources provide discriminative activity, offset to events including image onset, time to deployment of gaze on target, and time spent with gaze deployed in one region.

We have shown in many instances combining these signal sources is advantageous using the SFFS feature selection method to best select features from each source to be combined. Conversely, we have shown in other instances this fails to be the case, indicating no detectable gain is attained by including the EEG signal source. While we failed to detect an increase we highlighted a number of issues such the low number of training examples in some instances that might be attributable as the cause of this increased/decreased accuracy.

From the results and analysis given in this chapter we can conclude combining the signal sources does give an increase in some instances, but in others it may be more advantageous to use only the features derived from the eye tracking signals (as these predominantly seemed to be a more reliable source of information). It should be noted as a point of clarity that where we found that the EEG signals alone provided the greater accuracy, we in fact are analysing the EEG signals with respect to information we attain from the eye tracker, so although we may not use the eye tracking signals directly in classification they are nonetheless needed to extract the EEG signals offset to events like eye fixations.

In this chapter we have further demonstrated that these signals can be used to drive image search, thus furthering our hypothesis that, EEG and Eye Tracking can be used to improve the effectiveness in searching for certain types of targets in images.

Chapter 5

Paradigms in EEG Search

In this chapter we examine how a number of factors affect the performance of EEG augmented target search.

The psychosocial phenomena on which we rely to drive EEG search are known to be modulated by a number of factors such as attentional strategy, target density and target difficulty. Furthermore the signals that we are detecting are known to be generated from a number of neural sources.

We explore a number of related questions in three separate sections in this chapter, each of which contributes to supporting our thesis hypothesis in some way. Firstly, we investigate the importance of the number of channels used i.e. the number of nodes placed on the skulls of our subjects, and the accuracies that can be achieved with their respective placements. This is important because few nodes means a cheaper setup in terms of computational processing power needed, as well as reduced inconvenience for the participant. Secondly, we explore whether some images have inherent characteristics in a search task that lend them to being correctly labelled/mis-labelled by a subject using an EEG augmented image search system. This is important to know something about the nature of such images as we generalise our work to other forms of EEG-augmented image search. Thirdly, we describe an investigation into the relationship between target presentation speed, and detection accuracy, which is important in optimising our overall process so as

Figure 5.1: Examples of the object stimuli. Targets (18,161,373,455) are shown on top, non-targets on the bottom.

to maximise the information provided by our participants.

5.1 Channel reduction

In this section we provide description and results of an experiment carried out utilizing EEG signals to drive an image search task. The primary contribution of the work here is in demonstrating that similar or even better accuracy can be achieved using fewer EEG channels.

5.1.1 Experimental Outline

Images from the ALOI (Amsterdam Library of Object Images) were used in this experiment (Geusebroek et al. (2005)). This image set is comprised of 1,000 objects, each photographed from a number of camera angles and under a number of different lighting conditions. This image set was chosen because it allowed use of a wide variety of non-target images which display visually salient and attentional arousing properties whilst allowing for a large number of different camera angles/lighting conditions for each object. Our target object was represented by a large number of different images and examples of some of these target images are shown in Figure 5.1.

Eight participants were shown a number of images of a target object that they were to search for prior to starting the experiment. The participants were instructed to press a button upon noticing the appearance of this target object. In total 4,800 images were shown to each user at a rate of 10Hz. Amongst these images, 60 target images were randomly distributed accounting for 1.25% of the total. The total duration of the task was thus 8 minutes. Four different targets were randomly selected from the ALOI dataset, with each target searched for by 2 users. Participants 1 & 5, 2 & 6, 3 & 7, 4 & 8 searched for ALOI targets 161, 455, 18 and 373 respectively. Each block sequence was constructed by randomly sampling the pool of available target and non-targets. The images of the target object could be from any of a number of perspectives or lighting conditions, thus ensuring the actual target image would always be different.

5.1.2 Data Collection

Recording of EEG signals was done using the KT88-1016 EEG system with a left mastoid reference and the chin as ground. Ag/AgCl electrodes were used with a 10-20 placement cap at locations 16 locations, namely F7, F3, FZ, F4, F8, T3, C3, CZ, C4, T4, T5, P3, PZ, P4, T6, OZ. Button presses were recorded on the KT88 apparatus to allow for time-stamping of behavioural responses with the EEG data. 5 males and 3 females were recruited with an average age of 27.5 years with standard deviation of 4.5 years. One of the males was left handed.

5.1.3 Analysis

The purpose of EEG-augmented image search is to enhance the detection capabilities of a user searching for a target image within a large database. In this regard we evaluate in this subsection the increased accuracy achieved by using EEG in combination with behavioural responses (button press), and where various trade-offs exist between the number of channels used.

To examine the EEG signals and derive a set of measures of their detectability we used an SVM (Support Vector Machine) with a linear learning kernel. For each image in the stream, we extracted the EEG signals from 16 channels for the 1 second following its presentation. We also extracted an additional channel which recorded the button presses. We set out to examine the effects of using a reduced number of channels on classification accuracy of the EEG signals and behavioural responses. To achieve this we used a SFFS (Sequential Forward Feature Selection) scheme (as described in Section 4.2). This scheme finds subsets of feature sets which offer optimal discriminative capacity between two classes by starting with an empty set and adding the feature (or set of features) that provide the greatest increase in accuracy on each iteration. This algorithm for each forward iteration also evaluates back-steps by seeing if removing a feature (or set) offers an increase in accuracy.

Using this algorithm in combination with a linear kernel SVM we were able to find subsets of channels which offered optimal solutions. We did examine the use of a SVM-RBF kernel with wide gridsearching for cost and gamma parameters, but this provided little gain at the cost of much increased running times of the SFFS algorithm so we do not report those details in the thesis.

Using the SFFS algorithm with a linear SVM we employed an approach where on each iteration a test set of 10/790 and 50/50 non overlapping target/non-targets were randomly selected from the available pool of samples. The training partition of 50/50 targets/non-target were fed into the SFFS algorithm that then evaluated subset combinations of channels. The SFFS algorithm evaluated channel subset combinations by further partitioning its training set into a test and training set of sizes 10/10 and 40/40 respectively. On each iteration, the SFFS algorithm produced a set of the channels for subset sizes 1 to 15 which represent the best found channel combination for that subset size. These subsets were evaluated on the initially removed test set of 10/790. The feature vector corresponding to a channel subset being evaluated was created by combining the EEG signal for those channels.

Additionally a second feature vector was created using only the button press signal channel. SVM models using these two feature sets were trained on the training set of size 50/50, where an additional SVM model to fuse their outputs was created by using their predictions in a 10-fold cross validation on this set. These two models were then used on the originally removed testing set of size 10/790 to produce prediction values for EEG and button presses, where the third model was used to combine the predictions. These predictions were then evaluated using accuracy measurements namely P@n and ROC-AUC for each of the 15 channel subsets.

We repeated this 20 times, and averaged the P@n and ROC-AUC accuracies as identified by their channel subset size (i.e. 20 accuracy values for channel subsets of size 4 were averaged to give an accuracy value for 4 channels). P@n (precision at n) is the proportion of true positives within the first n elements of an ordered list. We set n = 10 since our test set contained 10 targets, as this reflects the target to non-target ratio of the pool data collected (10/790 to 60/4740). This scheme of keeping independent testing sets was necessary to ensure that subset solutions found by the algorithm were not simply biased by random relationships in the training data which did not generalise to the rest of the data. By keeping a test set of size 10/790 separate from the beginning on each iteration, we can ensure the models applied and evaluated are not biased in this way.

5.1.4 Combined EEG and Button Press Results

Shown in Tables 5.2 and 5.3 are the results for all 8 of our participants. The P@n accuracies indicate that the inclusion of EEG in all cases brought about a performance increase. This was not the case though for all participants when using the AUC-ROC accuracy. This may well indicate that P@n is a more stable measurement of accuracy in this situation in comparison to AUC which may be failing to reveal these performance gains.

In Figure 5.2 we graphically show the increase in accuracy achieved when including increasing numbers of EEG channels with the button press response.

We can also see that the inclusion of additional EEG channels in some cases can reduce detection performance (i.e. Subject 3) albeit not very much. This may be

Subject	Max P@n (# of channels)	Button	% Increase
1	.4 (3)	.310	29%
2	.665(3)	.511	30%
3	.6(3)	.375	60%
4	.76 (5)	.466	63%
5	.62 (6)	.315	97%
6	.414(5)	.328	26%
7	.4(3)	.287	39%
8	.435 (5)	.247	76%

Table 5.1: Increases in accuracy obtained using EEG and Button press.

due to that fact that additional channels do not provide any further discriminative information, and only serve to introduce noise. Examining the button press channel following target presentations it was found that some users failed on occasion to respond within one second (i.e they missed the target). Participants 7 and 3 missed 9 and 3 targets respectively, with participants 3 and 6 missing 2. This may explain the lowered accuracy in some cases.

Table 5.1 summaries some of this detail from Figure 5.2 for each subject. In Column 2 (c1) we show the maximum P@n achieved along with the associated number of channels. In Column 3 (c2) we show the P@n achieved using only button presses (x-axis value = 0). In column 4 we show the percentage increase calculated as $((c1 - c2)/c2) * 100$. The average increase by including EEG data was 52.56% over using only the button press.

Of interest to us in this work is examining the effects that a greater/fewer number of EEG channels has on performance of signal detection. In Figure 5.3 we show the average increase across the set of 8 users achieved by adding an additional channel. We can see that by using 4 channels of EEG we achieve nearly 50% of an increase compared to using button press responses alone. The optimum seems to be indicated at 6 channels with a 51.17% increase, but this negligible gain if statistically significant hardly seems worth introducing 2 more nodes for. In Table 5.4 we show the data graphically represented in Figure 5.4.

Comparing the mean P@n accuracies using a paired t-test between all subjects for using only button press, and using 4 channels, reveals a strong significant difference between the means further confirming that EEG in combination with button press provides an accuracy greater than button alone (two tailed t-test, $t(7)=5.6255$, $p=0.0008$).

Channels	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Average
0	0.311	0.511	0.375	0.466	0.315	0.328	0.287	0.247	0.355
1	0.373	0.605	0.582	0.750	0.510	0.373	0.376	0.300	0.484
2	0.367	0.645	0.588	0.735	0.515	0.370	0.388	0.359	0.496
3	0.400	0.665	0.600	0.745	0.575	0.376	0.400	0.388	0.519
4	0.393	0.660	0.600	0.750	0.605	0.400	0.388	0.429	0.528
5	0.380	0.655	0.600	0.760	0.590	0.414	0.406	0.435	0.530
6	0.400	0.635	0.594	0.750	0.620	0.410	0.406	0.429	0.531
7	0.380	0.635	0.588	0.730	0.610	0.410	0.400	0.429	0.523
8	0.367	0.665	0.600	0.745	0.580	0.404	0.400	0.435	0.524
9	0.373	0.660	0.582	0.745	0.580	0.391	0.371	0.412	0.514
10	0.393	0.655	0.594	0.750	0.595	0.385	0.388	0.400	0.520
11	0.380	0.640	0.606	0.745	0.595	0.372	0.388	0.400	0.516
12	0.387	0.640	0.582	0.750	0.615	0.380	0.388	0.388	0.516
13	0.380	0.645	0.582	0.735	0.610	0.393	0.365	0.406	0.514
14	0.393	0.650	0.576	0.745	0.585	0.392	0.388	0.371	0.513
15	0.393	0.655	0.553	0.735	0.595	0.400	0.376	0.388	0.512
16	0.380	0.660	0.547	0.735	0.585	0.390	0.347	0.371	0.502
max	0.400	0.665	0.606	0.760	0.620	0.414	0.406	0.435	0.531

Table 5.2: P@n accuracies across subjects showing the effect of increased EEG channel count on accuracy when combined with behavioural response.

Channels	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Average
0	0.9619	0.9858	0.9638	0.9873	0.9546	0.9577	0.9148	0.9617	0.9610
1	0.9603	0.9944	0.9711	0.9960	0.9722	0.9503	0.8968	0.9741	0.9644
2	0.9582	0.9952	0.9697	0.9963	0.9699	0.9485	0.9014	0.9755	0.9644
3	0.9674	0.9953	0.9731	0.9958	0.9745	0.9456	0.8926	0.9764	0.9651
4	0.9631	0.9952	0.9741	0.9965	0.9736	0.9416	0.8941	0.9784	0.9646
5	0.9687	0.9950	0.9702	0.9964	0.9731	0.9453	0.8961	0.9811	0.9657
6	0.9666	0.9945	0.9724	0.9965	0.9736	0.9454	0.8944	0.9785	0.9652
7	0.9708	0.9944	0.9705	0.9964	0.9731	0.9466	0.8956	0.9788	0.9658
8	0.9638	0.9947	0.9687	0.9965	0.9731	0.9481	0.8974	0.9786	0.9651
9	0.9712	0.9951	0.9704	0.9957	0.9723	0.9492	0.8861	0.9788	0.9648
10	0.9644	0.9946	0.9706	0.9960	0.9724	0.9473	0.8903	0.9774	0.9641
11	0.9623	0.9944	0.9696	0.9959	0.9724	0.9471	0.8846	0.9774	0.9630
12	0.9664	0.9942	0.9700	0.9960	0.9727	0.9483	0.8877	0.9732	0.9636
13	0.9628	0.9946	0.9695	0.9959	0.9730	0.9486	0.8814	0.9771	0.9629
14	0.9615	0.9941	0.9694	0.9957	0.9718	0.9478	0.8896	0.9788	0.9636
15	0.9636	0.9944	0.9689	0.9958	0.9723	0.9537	0.8756	0.9791	0.9629
16	0.9680	0.9947	0.9692	0.9956	0.9727	0.9487	0.8758	0.9783	0.9629
max	0.9712	0.9953	0.9741	0.9965	0.9745	0.9577	0.9148	0.9811	0.9658

Table 5.3: AUC accuracies across subjects showing the effect of increased EEG channel count on accuracy when combined with behavioural response.

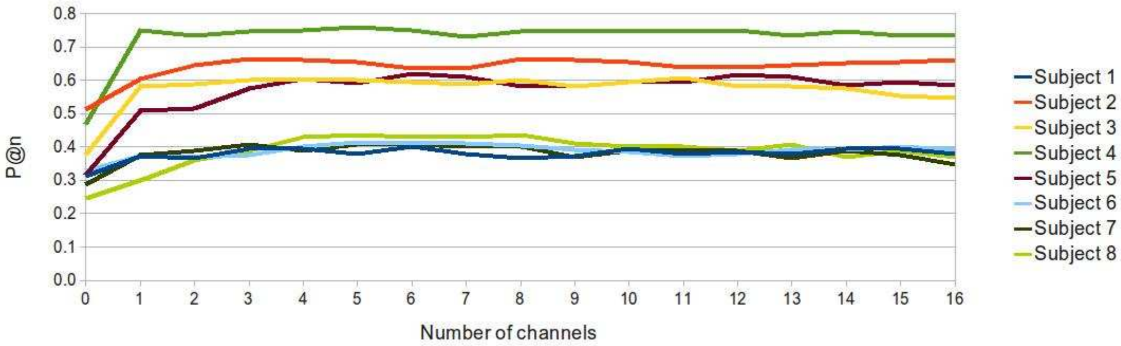

Figure 5.2: Graph showing P@n accuracies across subjects for increasing EEG channel counts when combined with behavioural response.

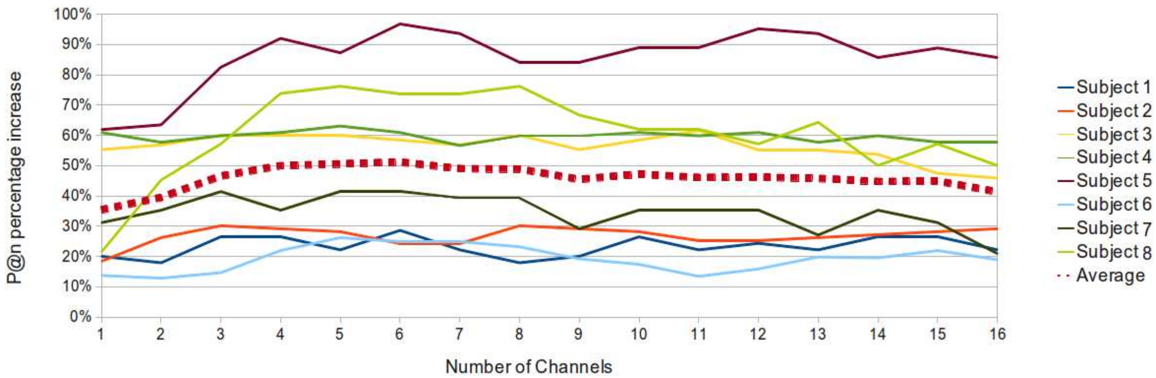

Figure 5.3: Graph showing P@n accuracies and their percentage increase over button press alone across subjects for increasing EEG channel counts.

Channels	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 7	Subject 8	Subject 9	Average
1	20.04%	18.40%	55.29%	60.94%	61.90%	13.72%	31.17%	21.46%	35.37%
2	17.90%	26.22%	56.86%	57.73%	63.49%	12.80%	35.27%	45.27%	39.44%
3	28.62%	30.14%	60.00%	59.87%	82.54%	14.63%	39.37%	57.18%	46.54%
4	26.47%	29.16%	60.00%	60.94%	92.06%	21.95%	35.27%	73.85%	49.96%
5	22.19%	28.18%	60.00%	63.09%	87.30%	26.22%	41.42%	76.23%	50.58%
6	28.62%	24.27%	58.43%	60.94%	96.83%	25.00%	41.42%	73.85%	51.17%
7	22.19%	24.27%	56.86%	56.65%	93.65%	25.00%	39.37%	73.85%	48.98%
8	17.90%	30.14%	60.00%	59.87%	84.13%	23.17%	39.37%	76.23%	48.85%
9	20.04%	29.16%	55.29%	59.87%	84.13%	19.21%	29.12%	66.71%	45.44%
10	26.47%	28.18%	58.43%	60.94%	88.89%	17.38%	35.27%	61.94%	47.19%
11	22.19%	25.24%	61.57%	59.87%	88.89%	13.41%	35.27%	61.94%	46.05%
12	24.33%	25.24%	55.29%	60.94%	95.24%	15.85%	35.27%	57.18%	46.17%
13	22.19%	26.22%	55.29%	57.73%	93.65%	19.82%	27.08%	64.32%	45.79%
14	26.47%	27.20%	53.73%	59.87%	85.71%	19.51%	35.27%	50.04%	44.73%
15	26.47%	28.18%	47.45%	57.73%	88.89%	21.95%	31.17%	57.18%	44.88%
16	22.19%	29.16%	45.88%	57.73%	85.71%	18.90%	20.93%	50.04%	41.32%

Table 5.4: P@n accuracies across subjects showing the effect of increased EEG channel count on accuracy when combined with behavioural response.

5.1.5 Analysis of channels chosen by SFFS algorithm

For each iteration of the SFFS algorithm (evaluating EEG+button) a score was kept of how many times each channel was selected for inclusion. These counts for each channel were then converted to the percentage they accounted for, for that number of channels being evaluated in that level of the SFFS selection scheme. This list is shown in Table 5.5. Analysing this table we can for instance see that the most common single channel selected across subjects when only one was selected was the Pz channel followed by T5. It should be noted that when more than a single channel's score percentage count is being evaluated that these scores represent the number of times the channel was selected across combinations, and thus fails to convey information regarding the specific combinations of channels chosen, and their respective informational relationships. In Figure 5.4 we summarise the extent to which each channel is included over increasing channel counts, and hence give indication of the importance of each channel on each successive increment of the number of channels used.

Channel # of Channels	Pz	Oz	Fz	Cz	C4	P4	P3	F4	F3	C3	T5	T6	T3	T4	F7	F8
1	26.9%	1.9%	3.1%	11.3%	4.4%	1.9%	6.3%	3.1%	3.1%	13.8%	14.4%	6.3%	3.1%	0.6%	0.0%	0.0%
2	15.9%	8.4%	4.4%	12.5%	4.4%	2.8%	9.4%	1.6%	3.1%	7.8%	11.9%	6.6%	4.4%	1.6%	2.2%	3.1%
3	9.8%	10.6%	6.0%	9.2%	4.4%	4.2%	11.0%	1.7%	5.2%	7.9%	9.0%	6.3%	5.6%	2.9%	3.8%	2.5%
4	8.9%	10.2%	4.8%	9.1%	3.9%	5.3%	7.3%	2.5%	5.2%	7.5%	9.5%	7.7%	6.9%	3.1%	3.8%	4.4%
5	8.1%	9.6%	5.1%	8.9%	3.8%	4.6%	7.4%	2.8%	5.5%	7.4%	7.8%	6.5%	8.1%	3.8%	6.1%	4.6%
6	7.4%	8.9%	4.9%	8.5%	4.6%	4.9%	7.1%	3.9%	5.8%	6.8%	6.6%	7.3%	8.3%	4.1%	6.4%	4.7%
7	7.7%	8.6%	4.6%	8.1%	4.7%	5.3%	6.8%	4.6%	5.5%	6.5%	6.5%	7.1%	7.1%	5.4%	6.5%	4.9%
8	7.1%	8.4%	5.5%	7.7%	4.8%	4.7%	6.2%	4.9%	5.5%	6.2%	6.8%	6.9%	7.1%	5.5%	7.3%	5.5%
9	7.0%	7.8%	6.0%	7.4%	5.3%	5.3%	6.4%	4.7%	5.7%	6.8%	6.2%	6.9%	6.5%	5.6%	7.2%	5.3%
10	6.9%	7.3%	5.7%	7.0%	6.1%	5.5%	5.9%	5.2%	5.8%	6.6%	6.4%	6.8%	6.9%	6.1%	6.5%	5.3%
11	6.6%	7.2%	5.5%	6.8%	6.5%	5.7%	5.6%	5.5%	5.9%	6.4%	6.5%	6.4%	7.1%	5.9%	6.7%	5.7%
12	6.1%	6.9%	5.7%	6.8%	6.6%	5.9%	5.5%	5.4%	6.9%	6.7%	6.1%	6.4%	7.0%	5.5%	6.7%	5.8%
13	6.2%	6.8%	5.8%	6.6%	6.6%	5.9%	5.5%	5.8%	6.7%	6.6%	5.8%	6.2%	6.8%	6.1%	6.6%	6.1%
14	6.2%	6.5%	5.9%	6.6%	6.6%	5.9%	5.8%	6.0%	6.6%	6.4%	5.6%	6.3%	6.6%	6.2%	6.7%	6.2%
15	6.3%	6.4%	6.0%	6.5%	6.5%	5.7%	6.2%	6.3%	6.4%	6.3%	5.7%	6.3%	6.5%	6.2%	6.4%	6.3%
16	6.3%	6.3%	6.3%	6.3%	6.3%	6.3%	6.3%	6.3%	6.3%	6.3%	6.3%	6.3%	6.3%	6.3%	6.3%	6.3%
Average	9.0%	7.6%	5.3%	8.1%	5.3%	5.0%	6.8%	4.4%	5.6%	7.2%	7.6%	6.6%	6.5%	4.7%	5.6%	4.8%

Table 5.5: SFFS channel score count percentage representations across subjects

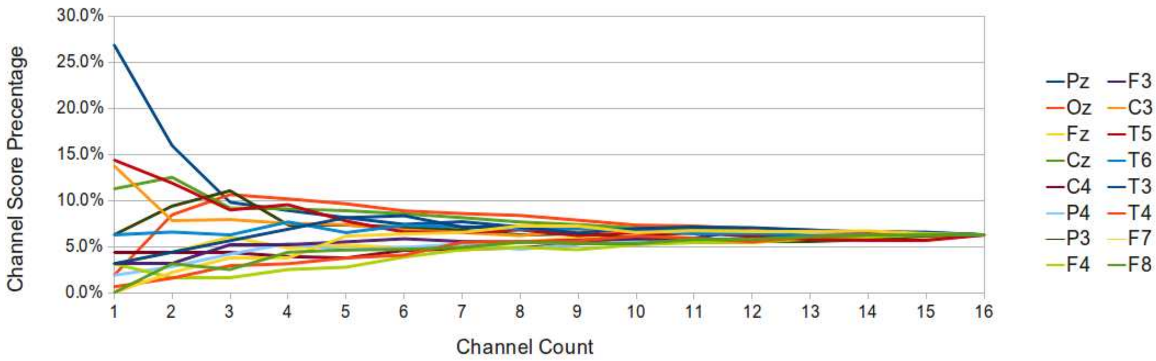

Figure 5.4: Graph showing channel score percentages over each increment of the number of channel used across subjects.

5.1.6 Analysis of classification using EEG signals only

In this subsection we present classification results when using EEG signals only. While these signals, when combined with button press responses, provide an increase in accuracy over using either alone and display effects like trade-offs with the number of channels to be used, they by themselves present a different accuracy profile with each increment of the number of channels used. In Tables 5.6 and 5.7 we present the P@n and AUC accuracies when using these EEG signals in combination with the SFFS algorithm without the button press. In Figures 5.5 and 5.6 we present the results of these tables in terms of respective difference from the initial accuracy score (1 channel) so as to calibrate them to a representation that is comparable between subjects.

The method used to generate these results is like that outlined in subsection 5.1.3 with the exception that the intermediate classifier used in the previous subsections to combine the EEG and button press scores is not needed. Here we just use the EEG classifier directly instead of this combining classifier. All other parameters such as iteration count and training/testing set sizes are consistent with those in the previous subsections.

Channels	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Average
1	0.010	0.265	0.290	0.490	0.230	0.055	0.040	0.195	0.197
2	0.000	0.250	0.290	0.460	0.310	0.060	0.015	0.200	0.198
3	0.005	0.275	0.305	0.505	0.340	0.080	0.040	0.215	0.221
4	0.005	0.270	0.300	0.560	0.350	0.065	0.045	0.235	0.229
5	0.005	0.255	0.275	0.625	0.360	0.125	0.040	0.250	0.242
6	0.000	0.265	0.315	0.615	0.400	0.105	0.050	0.235	0.248
7	0.000	0.295	0.315	0.635	0.400	0.150	0.055	0.245	0.262
8	0.010	0.315	0.310	0.640	0.375	0.135	0.045	0.270	0.263
9	0.000	0.285	0.310	0.635	0.410	0.110	0.065	0.295	0.264
10	0.000	0.300	0.325	0.650	0.400	0.130	0.070	0.290	0.271
11	0.000	0.335	0.325	0.665	0.410	0.130	0.065	0.275	0.276
12	0.005	0.350	0.300	0.655	0.440	0.150	0.065	0.295	0.283
13	0.005	0.335	0.325	0.655	0.420	0.160	0.065	0.305	0.284
14	0.005	0.330	0.350	0.640	0.420	0.135	0.060	0.295	0.279
15	0.000	0.310	0.350	0.635	0.420	0.155	0.065	0.290	0.278
16	0.000	0.355	0.340	0.660	0.405	0.160	0.055	0.305	0.285

Table 5.6: P@n accuracies across subjects showing the effect of increased EEG channel count on accuracy.

Channels	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Average
1	0.5752	0.9089	0.8755	0.9540	0.8893	0.7455	0.7178	0.8677	0.8167
2	0.6020	0.9171	0.8828	0.9371	0.9219	0.7427	0.7159	0.8625	0.8227
3	0.6398	0.9217	0.8906	0.9538	0.9253	0.7358	0.7232	0.8717	0.8327
4	0.6401	0.9262	0.8836	0.9655	0.9168	0.7329	0.7139	0.8741	0.8316
5	0.6287	0.9292	0.8895	0.9737	0.9218	0.7543	0.7251	0.8762	0.8373
6	0.6289	0.9358	0.8949	0.9751	0.9316	0.7705	0.7235	0.8898	0.8438
7	0.6273	0.9464	0.9018	0.9792	0.9290	0.7793	0.7312	0.8866	0.8476
8	0.6291	0.9489	0.8980	0.9813	0.9330	0.7784	0.7305	0.8947	0.8492
9	0.6138	0.9486	0.8986	0.9805	0.9371	0.7880	0.7327	0.9016	0.8501
10	0.6357	0.9468	0.8975	0.9825	0.9421	0.7970	0.7322	0.9100	0.8555
11	0.6285	0.9484	0.8960	0.9822	0.9448	0.7944	0.7398	0.9047	0.8549
12	0.6287	0.9461	0.8960	0.9827	0.9466	0.7955	0.7416	0.9096	0.8558
13	0.6425	0.9494	0.9001	0.9838	0.9487	0.8025	0.7461	0.9087	0.8602
14	0.5797	0.9512	0.9000	0.9833	0.9504	0.8065	0.7411	0.9126	0.8531
15	0.6068	0.9491	0.9013	0.9831	0.9512	0.8096	0.7442	0.9116	0.8571
16	0.5983	0.9535	0.9006	0.9832	0.9510	0.8129	0.7470	0.9131	0.8575

Table 5.7: AUC accuracies across subjects showing the effect of increased EEG channel count on accuracy.

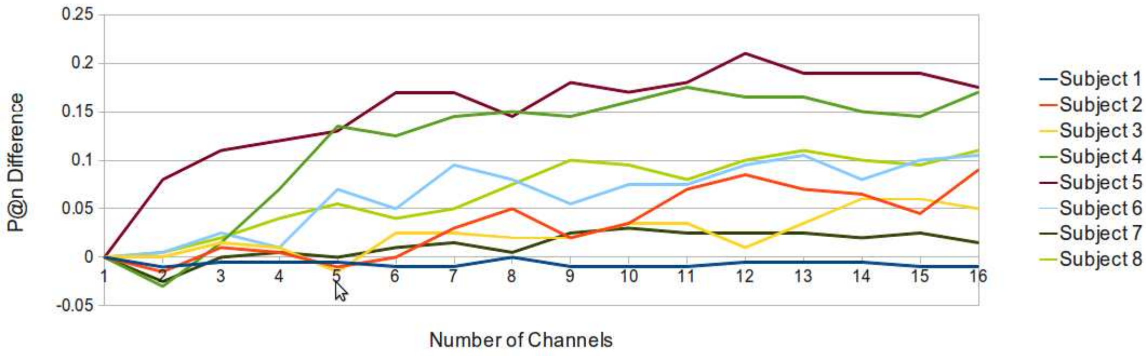

Figure 5.5: Graph showing the effect on accuracy of additional EEG channels as measured by P@n over using one channel.

5.1.7 Conclusions

In this section we have outlined an experiment to assess whether EEG and button press responses can be used to help label images when presented in a RSVP steam. What we found is that while both can be used for this purpose — and when combined provide an accuracy greater than either achieves alone — selecting a subset of EEG channels to be used in tandem with a button press can provide the same gains in accuracy but with the cost of fewer EEG channels. These results further lend support to our hypothesis that EEG and Eye Tracking can be used to improve the effectiveness in searching for certain types of targets in images.

5.2 Non Repeated Search

We now outline and present results of an experiment where subjects were required to search an RSVP stream of images for those containing bridges, and to indicate detection by pressing a button. While this work bears similarity in some regard to the previous section, it differs both in basic parameters like target density and presentation speed, but also in the nature and diversity of the target and non target images used. In previous sections all our target images were single objects with no distracting background whereas here our subjects have to interpret each image to see if it contains an actual bridge, so there is some semantic interpretation needed.

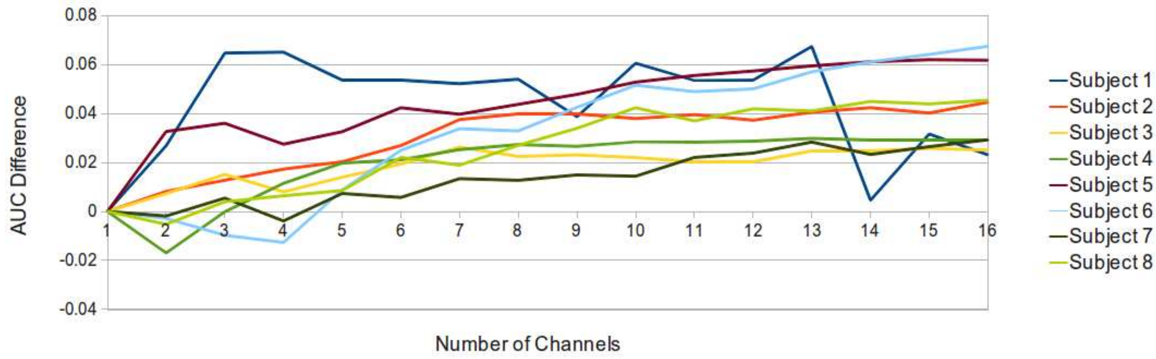

Figure 5.6: Graph showing the effect on accuracy of additional EEG channels as measured by AUC over using one channel.

5.2.1 Experimental Outline

Subjects were required to signal detection of images of bridges within an RSVP stream with a behavioural response, normally a button press. The images used were gathered from the photo sharing website flickr, and subsequently annotated into 6 categories: bridges, churches, fountains, houses, office blocks, and statues. With each of these images having a different aspect ratio, each was rescaled so that the greater of its width or height was 500 pixels (fitting each image to a 500x500 pixel bounding box whilst retaining its original aspect ratio).

The experiment was broken into 2 blocks. 50 target and 780 non-target images were shown per block totalling 100 targets and 1560 non-targets in all ($\tilde{6}$% targets). The same target and non-target images were used for subjects, but their order was randomised for each subject. Images were presented at a rate of 4 Hz.

5.2.2 Data Collection

Following the experiment outlined in subsection 3.1.1 subjects were invited to take part in an additional experiment. All 7 participants from the first experiment agreed to take part. Data recording was performed using the same equipment set-up as described in subsection 3.1.2.

5.2.3 Analysis Technique

A linear SVM was employed to analyse both the EEG and button press signals. Feature vectors were extracted from the EEG and button press channel for the 1 second following each image presentation in the stream. In order to generate metrics of the discriminative capacity of these signals both individually and in combination with respect to their ability to differentiate between targets and non-targets we used a linear SVM. For this analysis we used a repeated random sub-sampling validation approach where on each iteration we selected a training and testing set. The training set was comprised of 75 examples for each class, with the testing set containing 25 targets and 390 non targets (retaining the original target/non-target ratio). We repeated this process 40 times, and averaged the scores for each subject to obtain an ROC-AUC accuracy, along with a P@n accuracy (n=25). The results of this analysis are shown in Tables 5.8 and 5.9.

In order to combine the information sources (EEG and button), for each of the training instances we trained a linear SVM model without that instances (using 99 independent randomly selected instances from each class), and then using this model we generated a predicted score for each of these instances. In this way we establish in a non biased fashion a score for each of the training examples that can be combined between both the EEG and button press, where a classification analysis can then be used to reveal their combined accuracy. With this new feature vector constructed, we performed a classification analysis using the same parameters as were used on these signal sources alone (40 iterations, 75 training example from each class, and a testing set of 25 targets and 390 non-targets). The results of this analysis are presented in the Merged columns of Tables 5.8 and 5.9.

We also transformed the combined scores using the same process to a single score. Using these scores and 0 as a cut-off point we calculated the true positive, false positive, true negative, and false negative counts for the EEG scores, button press scores, and EEG and button press scores combined. These results are presented in Tables 5.10, 5.12 and 5.11. A visual rendering is shown in Figure D.1 of the most

Subject	Merged	Button	EEG
1	0.9362	0.9130	0.8022
2	0.9598	0.9491	0.8464
3	0.9516	0.9443	0.8362
4	0.9475	0.9176	0.8442
5	0.9826	0.9780	0.8819
6	0.9256	0.9228	0.7354
7	0.9179	0.9124	0.7492
Average	0.9459	0.9339	0.8136

Table 5.8: AUC classification accuracies across subjects for EEG, Button press, and EEG and Button combined.

Subject	Merged	Button	EEG
1	0.785	0.721	0.350
2	0.812	0.811	0.450
3	0.775	0.700	0.409
4	0.761	0.690	0.440
5	0.876	0.853	0.463
6	0.747	0.735	0.241
7	0.712	0.683	0.311
Average	0.781	0.742	0.381

Table 5.9: P@n classification accuracies across subjects for EEG, Button press, and EEG and Button combined.

significant of true positives, false positives, true negatives and false negatives. The remainder of these are shown in Appendix D.

Subject	TP	FP	TN	FN
1	88	63	1497	12
2	87	38	1522	13
3	88	64	1496	12
4	88	53	1507	12
5	96	26	1534	4
6	87	53	1507	13
7	80	53	1507	20
Average	87.71	50.00	1510.00	12.29

Table 5.10: Confusion Matrix Scores for classification results on merged EEG and Button sources

Subject	TP	FP	TN	FN
1	88	88	1472	12
2	88	48	1512	12
3	89	77	1483	11
4	88	79	1481	12
5	96	31	1529	4
6	89	59	1501	11
7	83	70	1490	17
Average	88.71	64.57	1495.43	11.29

Table 5.11: Confusion Matrix Scores for classification results on button press

Subject	TP	FP	TN	FN
1	74	419	1141	26
2	72	331	1229	28
3	76	382	1178	24
4	79	346	1214	21
5	82	333	1227	18
6	66	501	1059	34
7	68	464	1096	32
Average	73.86	396.57	1163.43	26.14

Table 5.12: Confusion Matrix Scores for classification results on EEG signals

Figure 5.7: Ranked images from subject 1 prediction scores. First line is top ranked true positives (descending from strongest predictions left to right), second line is true negatives (descending from strongest prediction left to right), third line is false negatives (ascending from worst prediction left to right), and the fourth line in false positives (ascending from worth prediction left to right)

5.2.4 Similarity across subjects

In order to assess similarities across subjects with regard to which images were most easily detected, and those which were not, we utilised a custom scoring method to elucidate prediction similarities with regard to individual images. In the previous subsection we analysed various measures of accuracy with respect to how images were ranked in terms of their classification with measures like true positive, true negative, false positive and false negative. In this subsection we demonstrate that statistically significant similarities exist between the image rankings between subjects, indicating that some images are more easily detected or mistaken as being a target or non-target.

We sorted the prediction scores generated with the merged classifier (EEG+button) for each subject from lowest to highest value, and paired each with an ascending value from 1 to 100 for targets, and 1 to 1560 for non-targets. These ordinally ranked lists were then combined across subjects combining scores into a tuple based on their paired image identity.

On each of these lists we implemented a scoring system to assess whether similarity on rankings existed between subjects. To do this we kept count of the number of instances with 2 or more subjects having the same image ranked within their top N or bottom N. For instance if 5 subjects shared an image within their top/bottom N the score was incremented by 5. By choosing a cut-off such as the top 20 images from the target list, we can derive a score of how well matched this list was between subjects. If the set of images exactly matched between all subjects within say the top 20, this score would be 140 (7*20). If none of these images matched the score would be 0.

In order to interpret these abstract scores we need a method of assessing to what degree they may have occurred by chance. To do this we use a bootstrapping method where we randomise the score orderings and repeat our measurement process on each iteration whilst keeping account of the maximum and minimum scores achieved with random orderings over a given number of iterations. By doing this we can discern that a score falling outside the range of the maximum and minimum scores as calculated through this bootstrapping process has a probability of having occurred by chance below a particular threshold. This probability threshold is calculated as p = 1/(number of iterations).

In Table 5.13 we show these values for the target image predictions between subjects. What we can see here is that all the predicted values fall within the range of being greater than a probability of 0.01, except the bottom 5% (5 targets) corresponding to target images with low overall detection rankings, or incorrectly classified as non-targets. This would indicate common false negatives between subjects.

In Table 5.14 we present an analysis for the non-target image predictions between subjects. Here we can see that significant ranking similarities occur with the highest ranked and the lowest ranked prediction scores as evidenced by the computed value falling outside of the range of bootstrapped significance values in all cases. This would indicate for the top ranked values that there exists significant commonalities corresponding to true negatives, and similarly for the lowest ranked values corresponding with false positives.

While this testing procedure is intended to capture ranking relationships that exist for images across subjects it may be failing to detect these in some cases. The absence of a significant value may not indicate that there is not one, and may simply mean a type 2 error has occurred (believing there to be no significant effect when there is one). This subtlety is important to note.

Range	Value	Max	Min
Top-5%	12	16	0
Top-10%	31	38	16
Top-25%	94	106	83
Bottom-25%	118	122	99
Bottom-10%	49	51	30
Bottom-5%	24	21	5

Table 5.13: Significance analysis of image ordering between subjects for targets. Values falling within Max & Min fail to satisfy a significance of p=.01.

Range	Value	Max	Min
Top-5%	394	174	121
Top-10%	866	571	470
Top-25%	1850	1670	1575
Bottom-25%	1722	1642	1569
Bottom-10%	719	557	459
Bottom-5%	395	175	111

Table 5.14: Significance analysis of image ordering between subjects for non-targets. Values falling within Max & Min fail to satisfy a significance of p=.01.

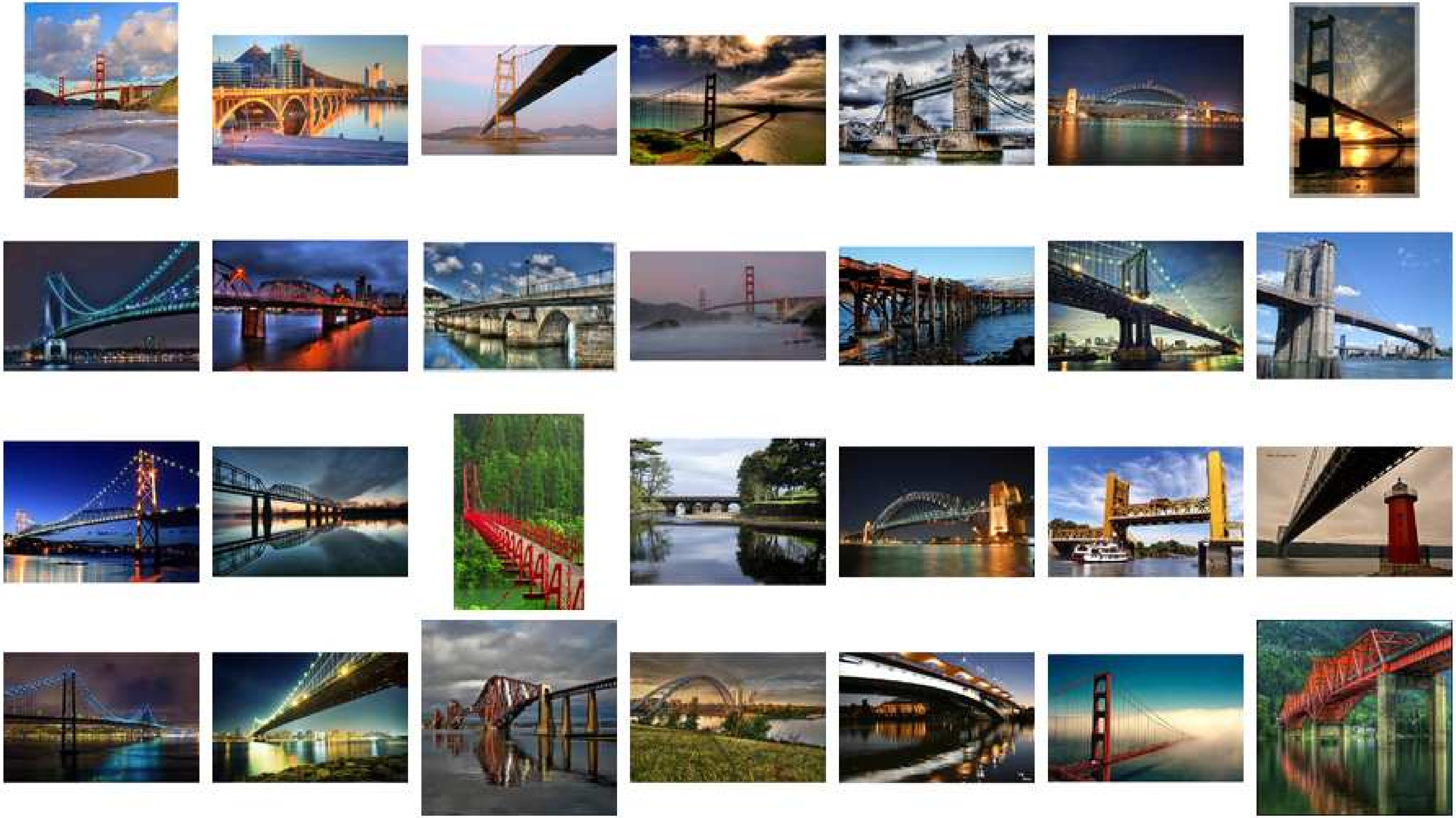

Figure 5.8: Images in order of most highly ranked as targets (left to right, top to bottom) across subjects for merged EEG and Button press prediction scores

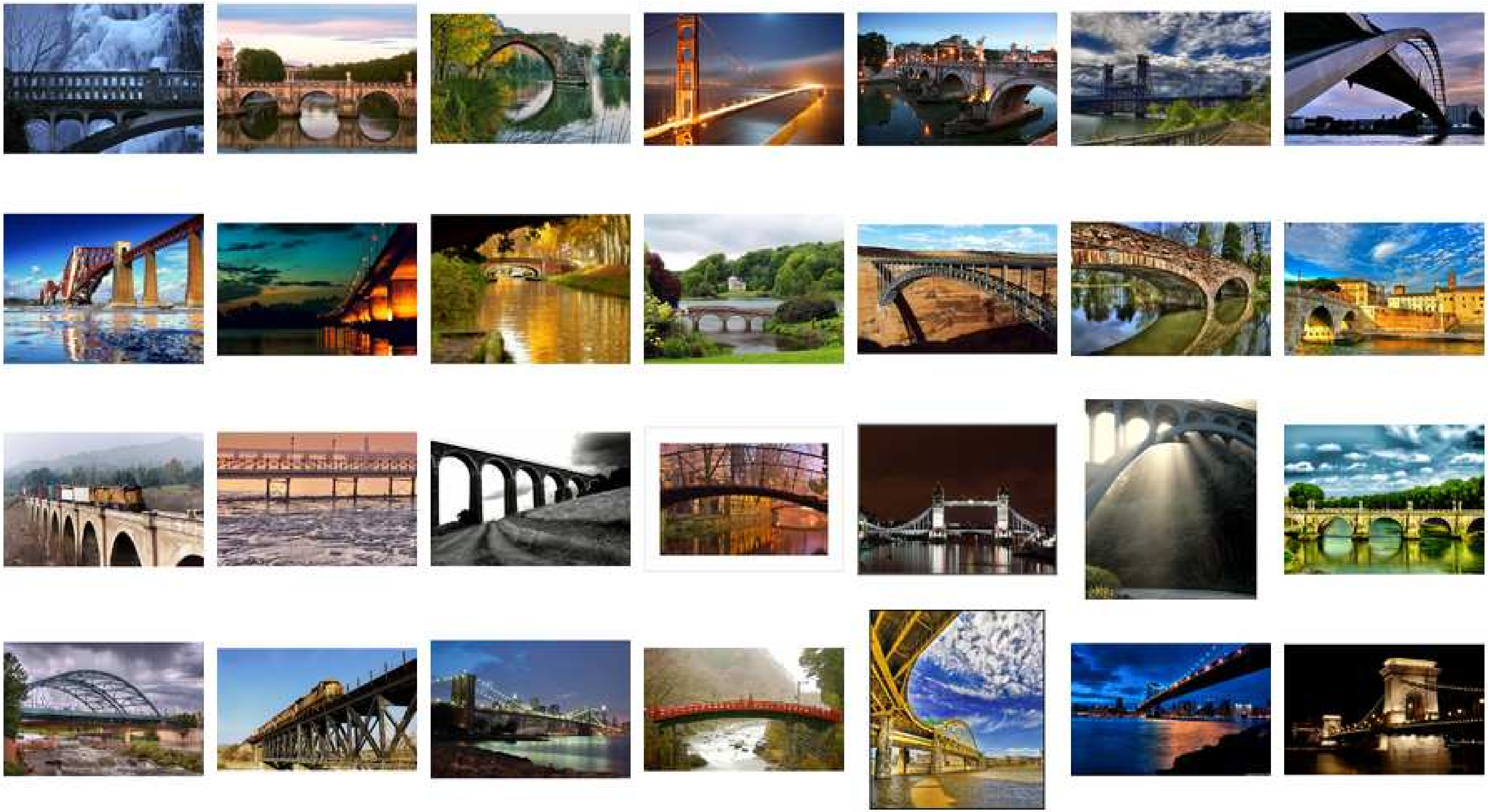

Figure 5.9: Images in order of least highly ranked as targets (left to right, top to bottom) across subjects for merged EEG and Button press prediction scores

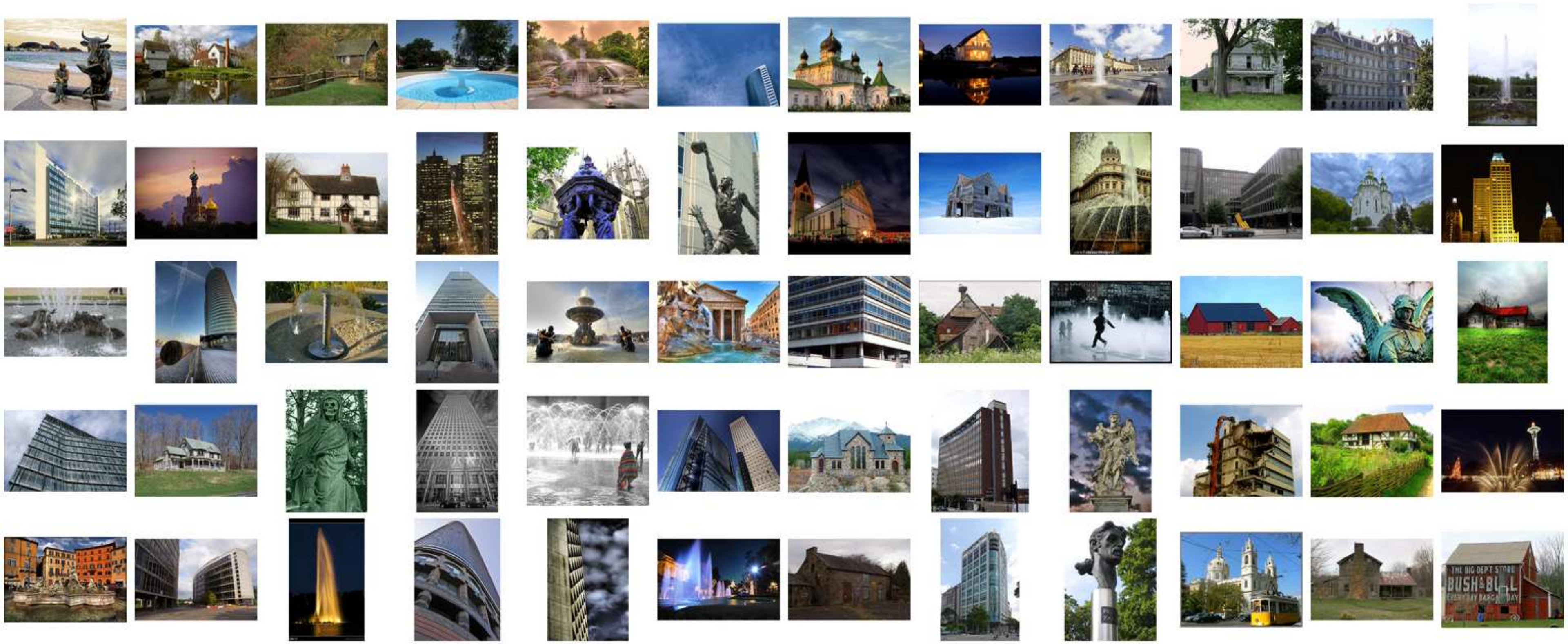

Figure 5.10: Images in order of most highly ranked as non-targets (left to right, top to bottom) across subjects for merged EEG and Button press prediction scores

Figure 5.11: Images in order of least highly ranked as non-targets (left to right, top to bottom) across subjects for merged EEG and Button press prediction scores

5.2.5 Conclusions

In this section we have shown results of an experiment where subjects were required to search a stream of images for those containing bridges. Analysing the neural and behavioural signals recorded we demonstrated that neural and behavioural signals existed that allowed us to use machine learning techniques to differentiate between target and non-target images, and also that these signals when combined, provide an accuracy greater than either one can achieve alone. We also demonstrated that similarities exist across subjects regarding the prediction scores calculated when trained on their behavioural and EEG data. This latter observation may allow us to be aware in the future that while although EEG and button press signals combined can provide an increase in annotation speed, there exist images which may tend to be misclassified across subjects.

5.3 Presentation Speed vs. Accuracy

In this section we outline part of an experiment carried out with the ESA (European Space Agency) to understand the effect of image presentation on discriminative signal detectability from EEG signals.

5.3.1 Outline

In this experiment, subjects were required to count the number of target items (plastic models of space shuttles) appearing in a stream of non-target images (rocks). At the end of each block the user would then input on a nearby keyboard the number of targets that they counted. In total, 4 subjects completed this phase of the experiment. Of interest here and under measurement was whether we were capable of detecting through a subject's EEG responses whether they viewed a target or non-target image without any explicit response, and how our capability to detect this dropped off with faster image presentation speeds.

(a)

(b)

Figure 5.12: Examples of an oddball (a) and non-oddball (b) images from the Simulated Martian Rocks collection.

Images of rocks (non-targets) with some containing a plastic space shuttle (targets) were provided by the ESA (Izzo et al. (2009b)), and used at the stimulus dataset. Examples of each are shown in Figure 5.12.

Each subject completed two repetitions of each of 5 sequences, across four different speeds. Totalling the number of target/non-target training examples across each image presentation speed for each subject this totalled 30/400, 61/670, 164/1330, 230/2000 and 382/4000 for each image presentation speed respectively. These speeds were 500ms, 300ms, 150ms, 100ms, 50ms. In between each image displayed was a gray mask (blank screen) for an equal amount of time. This would mean for a 500ms image presentation, it would be followed by a 500ms grey screen before the next image in the sequence appeared.

5.3.2 Data Collection

Two pendant EEG bluetooth devices were used to record EEG signals in this experiment. The devices were joined by tethering their reference connections, and similarly for their ground connections. In this way we could convert two, two channel devices into a 4 channel recording device. Ag (Silver) electrodes were placed at sites Cz, Pz, P3, P4, with a joint earlobe reference with the chin as ground. Subjects were aged 23-33.

5.3.3 Analysis

Analysis was completed using machine learning tools using a radial basis function SVM kernel.

Feature vectors were constructed of:

- 14 samples are extracted from the signal for the time-window between 220ms and 810ms relative to the image presentation time, low-pass filtered at a cut-off frequency of 14Hz. A time resolution of 40ms (inferior to any IDP) is thus obtained.

- Spectral information –as obtained from the Fast Fourier Transform (FFT)– of the raw signal (the DC component is previously removed) during the time-window ranging from 220ms to 620ms. 5 features are extracted for frequencies from 1hz to 15hz at a spectral resolution of 3Hz, which attempt to capture out differences in the high frequencies over a short time-frame.

- Additional spectral information of the low frequencies between 1Hz and 5Hz for the whole signal (time window between 220ms and 1000ms). 5 attributes are chosen, which thus encode changes at a resolution of 1Hz.

Before classification, samples were normalized into the range [-1,1] using a linear transformation. Finally, for each stimulus (either oddball or non-oddball) 24 features were extracted from each dataset. Since the EEG setup consists of 4 channels, an overall feature vector of 96 features per stimulus was gathered.

We pruned the feature vectors from their original length of 96 attributes to 35 attributes via an SVM attribute evaluator (as implemented in the Weka toolkit **?**).

Stratified cross-validation was then performed to iteratively build the classifier, whereby we instituted an approximate 66/33 split between training and test samples based upon the number of oddballs. Training was undertaken on a balanced dataset as constructed by the modified bagging approach.

The cross-validation methodology was constructed out of 30-folds, and for each

	500ms	300ms	150ms	100ms	50ms
Subject 1	0.8254	0.7997	0.7291	0.6702	0.6276
Subject 2	0.8297	0.8164	0.8012	0.7492	0.6114
Subject 3	0.9043	0.7844	0.6593	0.6282	0.6362
Subject 4	0.6946	0.8072	0.7948	0.7207	0.6524
Average	0.8135	0.8019	0.7461	0.6921	0.6319

Table 5.15: AUC Values across subjects for ESA Speed vs Accuracy Experiment

fold a grid-search optimization was run to determine the best parameters (C,γ) for the SVM.

5.3.4 Results

We can see from the graphs in Figure 5.13 a clear attenuation in both signals as the presentation time becomes faster, which indicates that classification accuracy should similarly deteriorate as the presentation time increases. Presented in Table 5.15 are the AUC values per subject, and the overall averages, whilst Figure 5.13 presents the averaged ROC curves across each of the presentation times.

5.3.5 Conclusions

In this section we have shown using a 4 node EEG system that increased presentation speed has an effect on classifier detection accuracy when searching for targets with an explicit indication of detection. In terms of our hypothesis, that EEG and Eye Tracking can be used to improve the effectiveness in searching for certain types of targets in images, this work shows that discriminating signals can be detected across a number of presentation speeds, however, we may need to calibrate the presentation speed to ensure maximum throughput.

5.4 Conclusions

In this chapter we explored three separate questions. Firstly, we investigated the importance of the number of EEG channels used, and the accuracies that can be

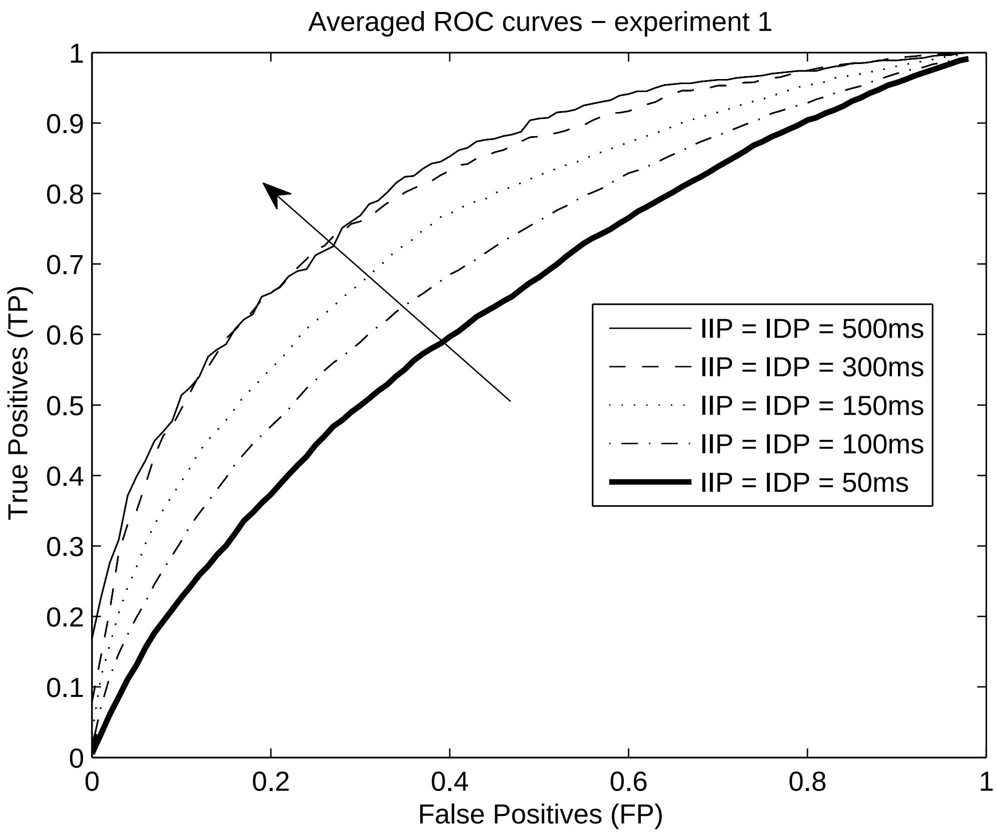

Figure 5.13: ROC curves averaged across subjects showing classification degradation with increased presentation speed

achieved with their respective placements on the scalp. Secondly, we explored whether some images have inherent characteristics in a search task that assist them in being correctly labelled/mislabelled by a subject using an EEG augmented image search system. Thirdly, we investigated the relationship between target presentation speed, and detection accuracy.

In all, these questions and results from experiments support our further hypothesis, that EEG and Eye Tracking can be used to improve the effectiveness in searching for certain types of targets in images.

Chapter 6

Conclusions

In this thesis we examined the hypothesis that EEG and Eye Tracking can be used to improve the effectiveness in searching for certain types of targets in images. We explored this hypothesis through a number of research questions. In this chapter we provide a retrospective overview of our chapters, examine our research questions in light of our experimental results, and discuss future work.

6.1 Chapter Summary

Chapter 1 In Chapter 1 we introduced our thesis, providing a brief overview, hypothesis, motivation and a list of central questions explored throughout the thesis.

These central questions were:

1. What neural signals are present during visual search tasks that require eye movements, and how do they inform us of the possibilities for BCI applications utilising eye tracking and EEG *in combination* with each other?

2. How do the temporal characteristics of eye movements give indication of the suitability of a search task to being augmented by an EEG based BCI system?

3. What are the characteristics of paradigms that can be used to elicit informative neural responses to drive image search BCI applications?

4. Can we use a reduced number of EEG channels in EEG BCI search?

Chapter 2 Traditionally the problems addressed by BCI (Brain Computer Interfaces) focused on the restoration of functionality and/or communication with people suffering from a variety of disorders such as ALS (Amyotrophic lateral sclerosis), stroke, and brain damage to name a few. There are many signals detectable from the brain, and many techniques for capturing these signals which can then be used to drive these systems. In this thesis we are primarily concerned with the analysis of EEG signals time-locked to events such as image presentations, button presses and eye movements. Besides the ongoing oscillatory patterns of EEG activity, there are well-known stereotypical responses to stimuli called ERPs (Event Related Potentials). Here we describe these signals, and give an overview of how they have been used in conventional BCI systems along with describing their significance within this thesis.

In Chapter 2 we give an overview of EEG, and explain how its constituent signals are utilised in both conventional BCI systems and newer BCI application spaces. In the penultimate section of this chapter we outline a central hypothesis and a set of research questions through which we examine this hypothesis in the thesis.

Chapter 3 In Chapter 3 we examined the signals present with regard to eye movements during a variety of search tasks, and examined how we can utilise these signals to aid in target detection. We showed that although the patterns of brain activity vary across subjects, and between tasks, differentiable signals exist related to the detection and recognition of targets that can be used to drive image search applications.

In the first section we explore these signals when subjects are searching for targets that are not discriminatingly perceivable until the time of fixation. This demonstrates that visual search scenarios may exist where a subject does not know the nature of a stimulus until the time of fixation, and that discriminative EEG patterns are present following this fixation.

In the following sections we explore fixations, and the patterns of EEG activity surrounding these fixations when the targets display salient qualities.

Understanding the paradigms and scenarios in which these signals can be elicited allows us to make informed decisions in considering the applications that may be ultimately driven by them. In order to expose these signals, we employ machine learning methods both analysing EEG signal sources and eye movements signal sources so as to disentangle the information sources available from each.

In this chapter we confirm that both EEG and eye movements signals contain discriminative informations that can allow us to identify targets.

Chapter 4 In Chapter 4 we showed how signals recorded from EEG and eye tracking sensors can be used to allow us to discriminate images or regions therein containing targets. The results indicate that both of these sensor sources provide discriminative activity, offset to events including image onset, time to deployment of gaze, and time spent with gaze deployed in one region.

The results presented in this chapter are pertinent to understanding the value of using these signal sources in tandem for real world search scenarios.

In this chapter we demonstrate that by combining EEG and eye tracking signals we can achieve accuracies greater than using either alone. We, however, show instances where this is not so, but provide reasons as to why this may be the case. They include lack of a sufficient number of training examples.

Chapter 5 In Chapter 5 we explored a number of related questions that contribute to the support of our hypothesis.

Firstly, we examined what advantages are realised by using a button press response in combination with EEG signals in a target search tasks involving images of objects displayed at high speed. Additionally we explore strategies of using a reduced number of EEG channels in tandem with the button press, to conclude that when EEG signals are combined with a button press, it is acceptable to use a reduced number of EEG channels. Secondly, we explored whether some images have inherent characteristics in a search task that lend themselves to being correctly labelled/mis-labelled. Here we had subjects respond with a button press to the detection of images of bridges displayed in a RSVP paradigm amongst a number of distractors. What we found were statistically significant relationships between the images that tended to be labelled/mis-labelled by subjects combined EEG and button press scores. Additionally, we show that combining button press responses with EEG signals provides a higher accuracy than using either alone. Thirdly, we examined the effect of presentation speed on our ability to discern target images from EEG signals in an RSVP paradigm.

6.2 Analysis and Discussion of Hypothesis

In this thesis we explored our hypothesis that EEG and Eye Tracking can be used to improve the effectiveness in searching for certain types of targets in images. The results and respective analysis support this hypothesis, and have exposed a number of its aspects. In Chapter 2 (Section 2.3) we outlined a hypothesis, and a number of research experiments with which we could explore this hypothesis. In this section we examine our research questions in respect to the experimental results in the thesis.

Research question 1

What neural signals are present during visual search tasks that require eye movements, and how do they inform us of the possibilities for BCI applications utilising eye tracking and EEG *in combination* with each other?

In Chapter 3 and 4 we explore this question by showing across a number of experiments that both EEG and eye tracking sensor signals provide discriminative information allowing us to differentiate between target and non-target stimuli.

In Chapter 3 (Section 3.1) we explored this question by conducting an experiment where subjects were required to detect object stimuli in a paradigm whereby they could not differentiate between target and non-targets stimuli until the time of deployment of gaze on them. This experiment established not only are there neural signals present that are sensitive to target/non-target detection offset to the time of fixation, but that these signals can be detected using machine learning algorithms, thus indicating they can be used in a BCI. This addresses our first research question, and provides insight into a task scenario where neural signals that can be used to drive a BCI involving eye movements are examined.

In Chapter 3 (Section 3.2) we further explored this research question utilising a more complex image set and task. In this task subjects were instructed to find images containing people. These images encompassed a variety of visual features, thus requiring the subject to perform a detection task using a wider diversity of visual informations. By demonstrating that we could detect differentiable neural signals during this more complex search scenario, we further address our first research question in showing across a variety of search strategies and images that we can detect brain activity related to discrimination between targets and non-targets. This is important as it allows us to know that although people may ultimately utilise different search strategies, we can still detect important neural signals that allow us to drive a BCI application.

In Chapter 3 (Section 3.4) we further explored this research question examining a scenario where we show subjects to be detecting targets as indicated by their high success rate in looking towards the target with their first eye movements. Here we similarly show signals that can be used to enable BCI systems, in tasks that encompass highly similar search behaviour and proficiency across subjects.

These experiments have allowed us to conclude that BCI applications involving target search are possible, and are applicable in both artificially generated and natural image stimuli. This evidence evaluated within the context of this research question further supports our research hypothesis that EEG and eye tracking can be used to improve the effectiveness in searching for certain types of targets in images.

Research question 2

How do the temporal characteristics of eye movements give indication of the suitability of a search task to being augmented by an EEG based BCI system?

In Chapter 3 we explore how indicators like dwell time on a stimulus like an image are indicative of effects like high discriminability for identifying those that are or contain targets. By observing effects like this, we can in part evaluate the gains that may be achieved by combining measures acquired from non-neural sources with an EEG BCI system. In Chapter 4 for instance we extend upon this by examining effects like those observed in Chapter 3 and evaluate them in a more directed way examining the signals when combined and used in tandem.

In Chapter 3 we additionally identify timed neural responses surrounding events like image presentation and eye movements. By identifying time periods of activity like this that provide us with discriminative information, we can discern that there exists patterns of neural activity for subjects for particular search tasks that can be utilised to allow us to drive a EEG BCI based system.

These information sources whether reaction time to look at a target, reaction time to detect a target, time spent processing a stimulus, or time periods of neural activity offset to events such as stimulus presentation or an eye movement, can all display indicative measures that can allow us to understand what tasks are suitable to be augmented by an EEG BCI based system.

Research question 3

What are the characteristics of paradigms that can be used to elicit informative neural responses to drive image search BCI applications?

In Chapters 3, 4, and 5 we identify a number of characteristics of paradigms that can be used in application scenarios to drive image search BCI applications.

In Chapter 3 we show in the first section that target detection can be offset to the time of fixation, and that informative neural signals are generated in response to this.

In Chapter 3 in the latter sections, and in Chapter 4, we show that although effects like salience may influence eye movements, these behaviours can help to provide discriminative information to allow us to drive image search BCI applications.

In Chapter 5 (Section 5.2) we demonstrate that button press reactions in combination with EEG signals are indicators that certain images may be likely to be labelled/mis-labelled as targets or non-targets. Measures derived in this way allow us to fundamentally understand a task and paradigm at a level where we can assess its limitations, and ultimately its benefit in the scope of a being augmented by a BCI system. In Chapter 5 (Section 5.1) we also show that RSVP display paradigms are suitable for driving image search BCI applications, and in addition button presses can augment their efficiency.

In Chapter 5 (Section 5.3) we also demonstrate using a reduced number of EEG channels that a subject does not need to overtly indicate detection of a target stimulus across a variety of speeds, thus demonstrating a variety of paradigm configurations in which EEG BCI allows us to label images by EEG signals.

By understanding the characteristics of paradigms like these, we are more informed to assess and identify what applications exist that can benefit from the use of an EEG BCI system. By examining this research question we also address a fundamental question regarding the scope of the applications that can exist. Ultimately this furthers our hypothesis that EEG and Eye Tracking can be used to improve the effectiveness in searching for certain types of targets in images.

Research question 4

Can we use a reduced number of EEG channels in EEG BCI search? In Chapter 5 (Section 5.1) we analyse the effect of reducing the number of EEG channels used when in combination with a button press. Analysing the discriminative EEG signals present without utilising this button press indicates that a reduction in channels hinders performance. Conversely, however, when combined with an overt behavioural response (button press) this no longer remains the case. Here we find that a reduced number of channels may be utilised, and we give an analysis of what the implicated channels placements are.

In this thesis we have shown that utilising non-neural sensor sources such as a button in combination with EEG signals can be conducive in allowing a better performing system. This further supports our hypothesis.

6.3 Future Work

A number of behaviours serving different motives and encompassing different and often evolving strategies are referred to as search. In this thesis we have examined under controlled circumstances a range of these search behaviours in order to both better understand their behavioural and neural components, and also to understand how we might provide systems that can enable a user to search in a more efficient and meaningful way. Although EEG BCI research has shown promise in a multitude of application scenarios, we focused our efforts in this thesis on evaluating how it relates to target search in images.

Assessing the benefits of EEG BCI (whether coupled or not) with eye tracking in search like tasks is difficult. In this thesis we were concerned with searching for targets in images. We may, however, envision scenarios where the user might not even be considered to be conventionally searching at all. An example of this would be somebody watching a television show where they see an attractive actor arriving on set, or perhaps they find an image amusing while browsing online. Neural signals may indicate correlates of interest or surprise to such events, but the subject may ultimately not explicitly express their meaning or significance. EEG signals interpreted in the context of these events may provide an additional layer of information. For instance, to allow a user to summarise the day's events by their neural responses. The work in this thesis supports the notion that research into applications with this aim may be fruitful.

In the experiments outlined in this thesis, we were able to ascertain the time of events such as image presentation and eye movements, and using these we could index the EEG signals in a meaningful way so as to unveil a further information source. An avenue of future research to build upon this might be to examine whether informative signals of this type exist when a person is mobile and engaged in daily tasks, such as shopping in a supermarket.

We would expect a task like this to entail a number of behaviours such as comparing a product with another, or perhaps deciding whether the product is worth the quoted price. Integral to decisions and behaviours of this type are eye movements in assimilating information such as the quantity of the product, its price and its packaging. Recent technological advances such as a portable eye tracking glasses (Bulling and Gellersen, 2010) that combine video recordings of not only the wearers view but also of their eye movements could be used in tandem with a portable EEG system to enable the capture of neural signals of the wearers as they are engaged in consumer behaviour. Understanding neural signals with this level of granularity and context may provide an avenue of research to better understand consumer behaviour.

EEG BCI type systems may not only assist us in goal directed behaviours such as a search, but they may also allow us to record an additional layer of implicit information surrounding events as we go about our daily lives in order to later not only summarise these events, but perhaps to share and communicate them with others.

Appendices

Appendix A

Equipment Overview

The experiments outlined in this thesis used a variety of apparatuses including an eye tracker, EEG (Electroencephalogram), and EOG (Electrooculogram). In this appendix we describe how these various pieces of equipment were used together.

A.1 EEG Recording

In order to detect the minute electrical signals generated by the brain we need specialised equipment to amplify, filter, and digitise these signals.

To record EEG signals we used Ag/AgCl (silver/silver-chloride) electrodes in an elasticated cap. The electrodes in this cap are arranged into standardised positions using what is known as the 10-20 placement system as shown in Figure A.1.

The potential difference at each of these electrode sites is then recorded in reference to a reference site. Typically the earlobes or the mastoid bone is chosen as a reference site. An additional electrode is typically placed elsewhere on the body as a ground reference site. For each of the signals recorded across the scalp in reference to the reference site, we subtracted from these the reading between the reference and ground site. This is done in order to mitigate noise due to environmental sources such as 50 Hz hum pattern from electrical equipment. These signals are then digitised and passed to a computer to be timestamped and recorded.

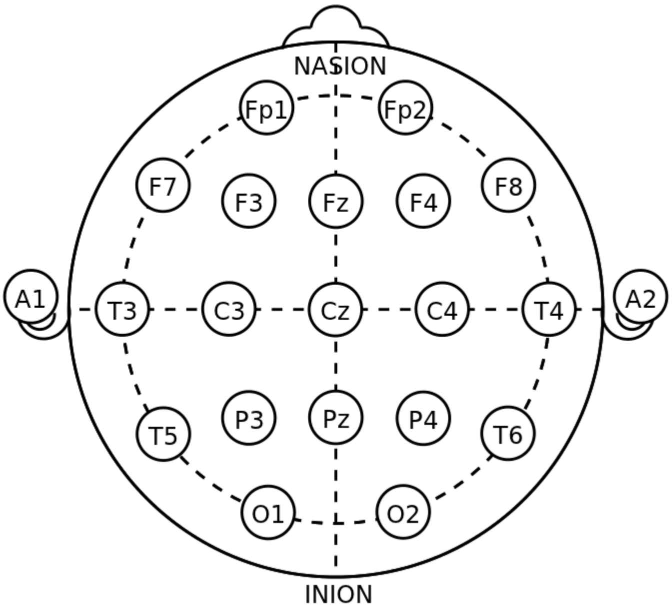

Figure A.1: 10-20 Electrode placement map

Unless otherwise noted, we used a joint mastoid linked reference, except for the experiment outlined in Section 5.1, where we used a left mastoid reference.

The KT88-1016 EEG system was used for signal recording in the experiments outlined with the exception of the experiment described in Section 5.3. In the latter experiment two 2-channel wireless EEG devices were used sharing a common ground and reference electrodes.

A.2 Eye tracking Recording

Two methods of recording eye movements were used for the experiments described in this thesis. In the experiment described in Section 3.1 we used EOG (Electrooculogram) signals acquired from VEOG (vertical) and HEOG (horizontal) channels. To do this we attach electrodes to the lateral canthus on both eyes for the horizontal pair, and above and below the eye for the vertical pair. By using the EOG channels (VEOG and HEOG) we were able to find the time indexes of fixations on the object stimuli. Eye movements along one plane (i.e. horizontal) generate signals more prominently on one channel pair than the other, and the voltage deflections are sensitive to the direction of eye movement. Eye movements in any direction are typically characterised by either positive or negative voltage deflections on both channels. By examining these voltage deflections we were able to identify the time of eye movements. The downside to this method although it provides high temporal accuracy to index the EEG signals is that it is difficult to exactly establish the eyes location.

For all other experiments requiring eye tracking we used the Tobii x50 Eye Tracking System. This system is comprised of a desktop LCD monitor equipped with infrared light emitters, and receiving cameras. The basic principle is that the infrared light accentuates properties of the eye such as the pupils that can be detected by the cameras to track eye movements and the location of gaze. This system provides X & Y pixel coordinate values of where on the screen the user's gaze is located. By referencing the time of these values against the system clock we can attain the time of eye movements and fixations, and thus index the EEG signal to reveal related neural activity. The Tobii x50 samples eye location at 50 Hz.

A.3 EEG Filtering

After EEG signals were received in a digitized from the KT88-1016 apparatus, the signal windows extracted relative to events such as image presentation or eye movement were bandpassed filtered to 0.1 Hz to 20 Hz. These signals were then resampled at 40 samples per second.

Appendix B

Analysis Conventions

In this appendix we outline the parameters and methods of a number of conventions used within the thesis for analysis.

B.1 Machine Learning and evaluation

In this thesis we rely on machine learning to show the presence of discriminating EEG activity and eye movement patterns. Primarily we utilise SVM (Support Vector Machine) which belongs to a class of supervised learning methods. Using a set of training examples, each marked as belonging to one of two classes, an SVM training algorithm constructs a model that can assign unseen examples into one category or the other. The effectiveness of the model relies on the presence of adequate discriminative information being present in the training examples.

Training examples are supplied to the model's training algorithm as belonging to one of two classes each assigned a numeric value such as [-1,1]. Each of these training examples are accompanied by a feature vector. In this thesis our feature vectors are composed of discrete samples taken from the EEG/Eye tracking signals. After post processing procedures such as bandpassing have been applied the relevant signals are re-sampled. These numeric samples are then combined into a feature vector. These feature vectors are then normalised into the range [-1,1].

In order to discern the amount of discriminative information present between two classes we employ a cross validation method. In this thesis we use Repeated Random Sub-sampling Validation. This procedure involves randomly subsampling the available instances into testing and training sets. On each iteration a model is trained using only the training set, and then this model is benchmarked upon the withheld test set. In this way we can examine the model's effectiveness by examining how it classifies unseen examples.

In order to obtain a measure of the effectiveness of the model in correctly discerning the true classes of the instances in each iteration's test set, we use a measure know as AUC (Area Under Curve). AUC is calculated as the area under a ROC (Receiver Operating Characteristic) curve.

This curve is generated by firstly sorting the numeric output for each instance of the test set from the model.

Outputs more closely approaching one of the binary numeric labels can be understood as the model more strongly indicating its belief that the relevant instance belongs to that class. Iterating across each instance in this ordered list, taking all instances above this point as belonging to one class and all those below it as belonging to the other class, we can calculate the true positive vs the false positive rate for each point in this list. The average (area below) this list is the AUC. We finally average the AUC values obtained by this repeated process of randomly sampling the available instance pool into training and test sets, and benchmarking the model trained on the training set upon the test set.

Throughout this thesis we additionally employ bootstrapping methods to verify that the accuracy obtained from an evaluation procedure was unlikely due to chance. In order to do this for any of the machine learning evaluation schemes, we simply randomise the test labels in the test set, and observe over a number of N iterations of this procedure what the highest accuracy achieved by chance was. Repeating this procedure N times, we can obtain a p=1/N measure of the highest accuracy achieved by chance. If we obtain an accuracy from our evaluation scheme on a dataset above the accuracy derived using this bootstrapping procedure, we can say probability of us having obtained our result by chance is below a particular probability threshold.

In this thesis we use the linear/RBF SVM function of the libsvm library (Chang and Lin, 2011). In the cases where we use a RBF kernel we outline the method by which we obtain the cost and gamma parameters using the grid search approach. In the case of us of a linear SVM, we chose the cost parameter at 1 as other values tended to be at best equal in accuracy, but more often detrimental to it.

B.2 Scalp Plots

We utilise averaged scalp plots throughout this thesis as a visual tool to show patterns of brain activity. These scalp plots are generated by averaging the sampled patterns of brain activity at their noted times. It is important to note they are representations of brain activity at that moment only, as activity occurring between the slices is not accounted for (i.e. we do not average across time window).

Appendix C

Supplemental material for experiments outlined in Chapter 3

In this section we provide the temporal discrimination plots obtained for the experiments described in Chapter 3 for experiments 2 and 3. Additionally we provide the averaged scalp plots. These materials are intended to primarily supplement Section 3.3 and 3.4 respectively.

C.1 Temporal Discrimination Plots for Experiment 2

In this section we present the temporal discrimination plots for experiment 2 outlined in 3.3.

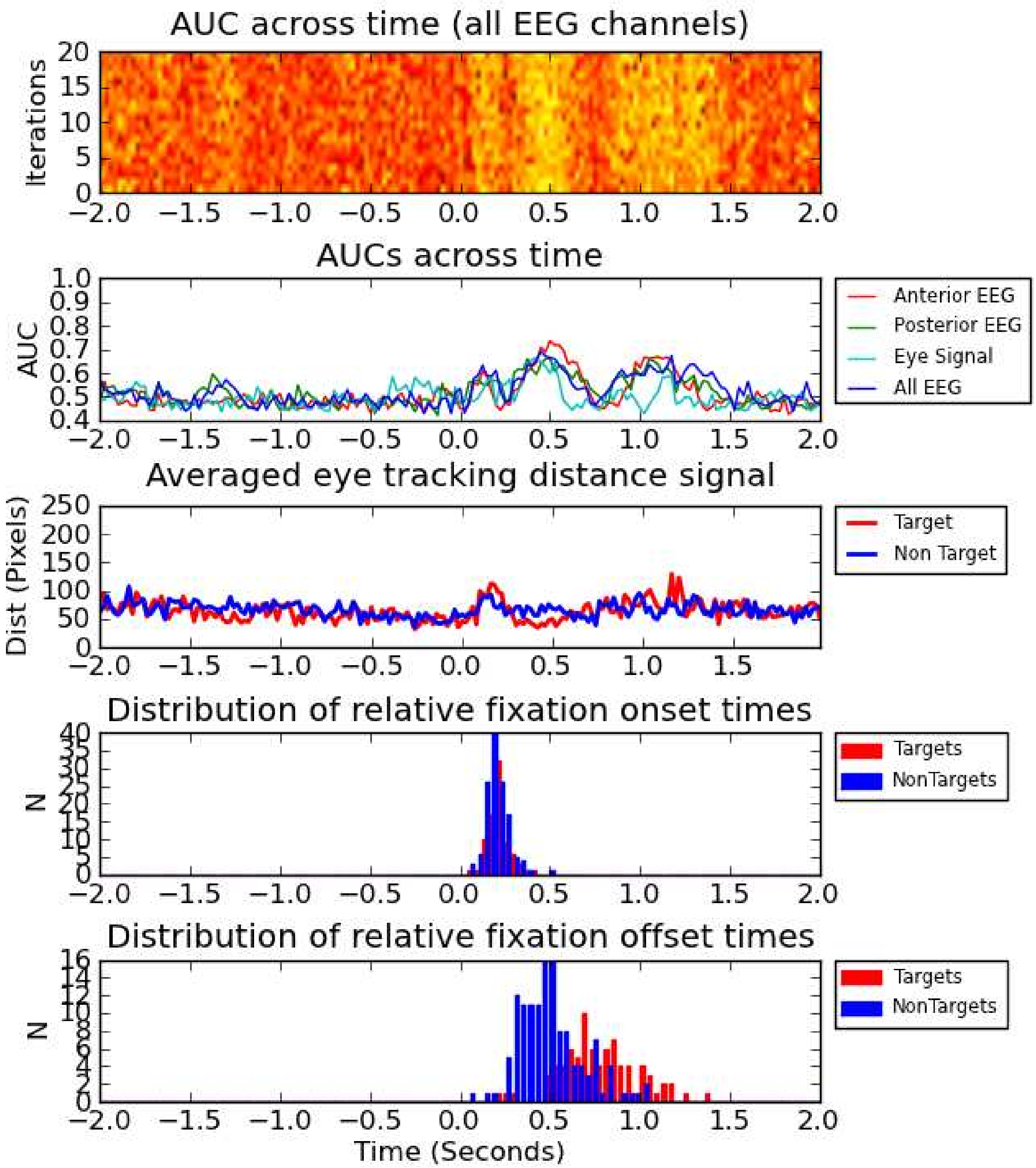

Figure C.1: Subject 1: Temporally aligned discrimination graphs map centred on the frame onset time showing differentiating activity related to target image detection compared to non target image detection

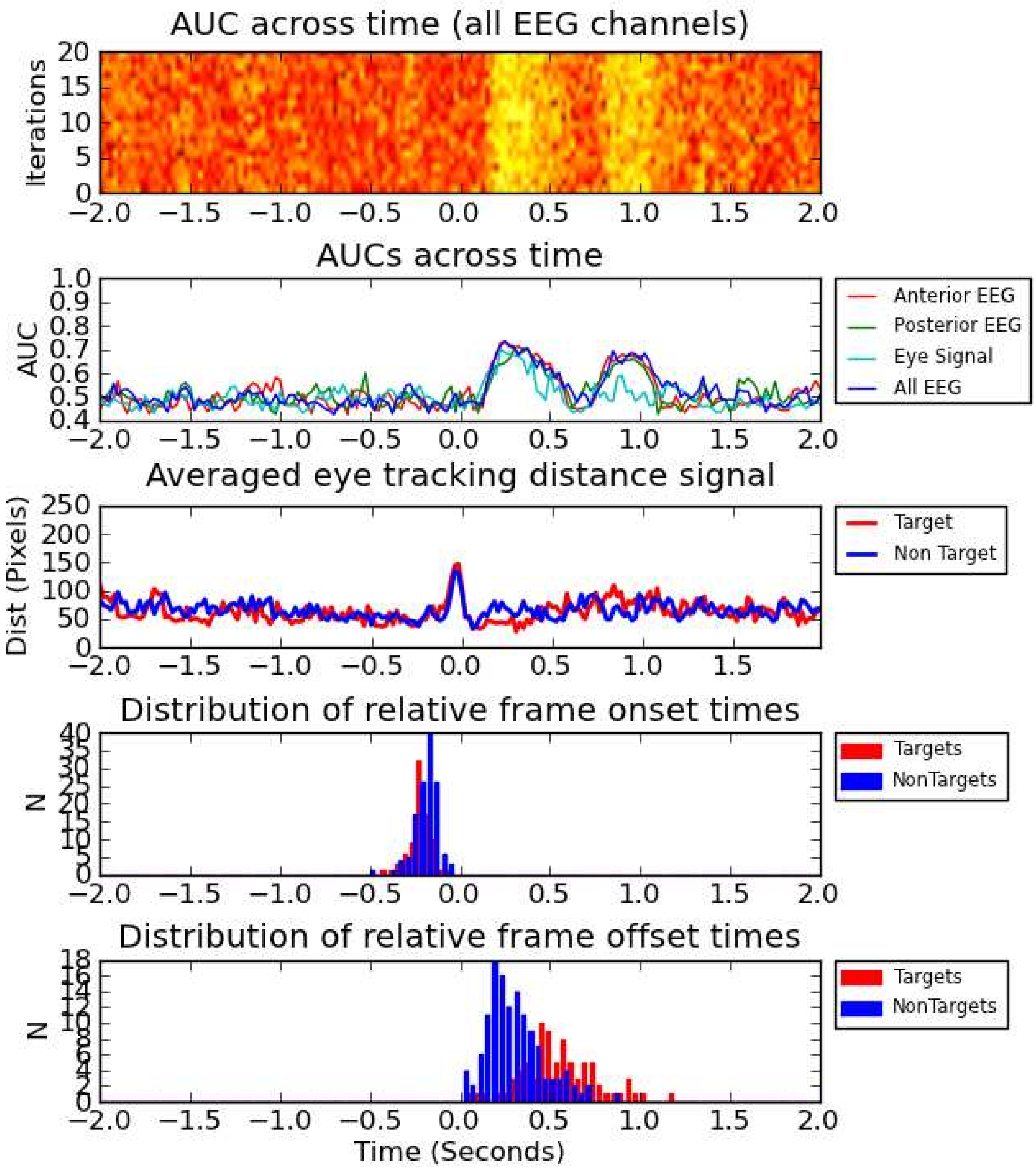

Figure C.2: Subject 1: Temporally aligned discrimination graphs centred on the fixation onset time showing differentiating activity related to target image detection compared to non target image detection

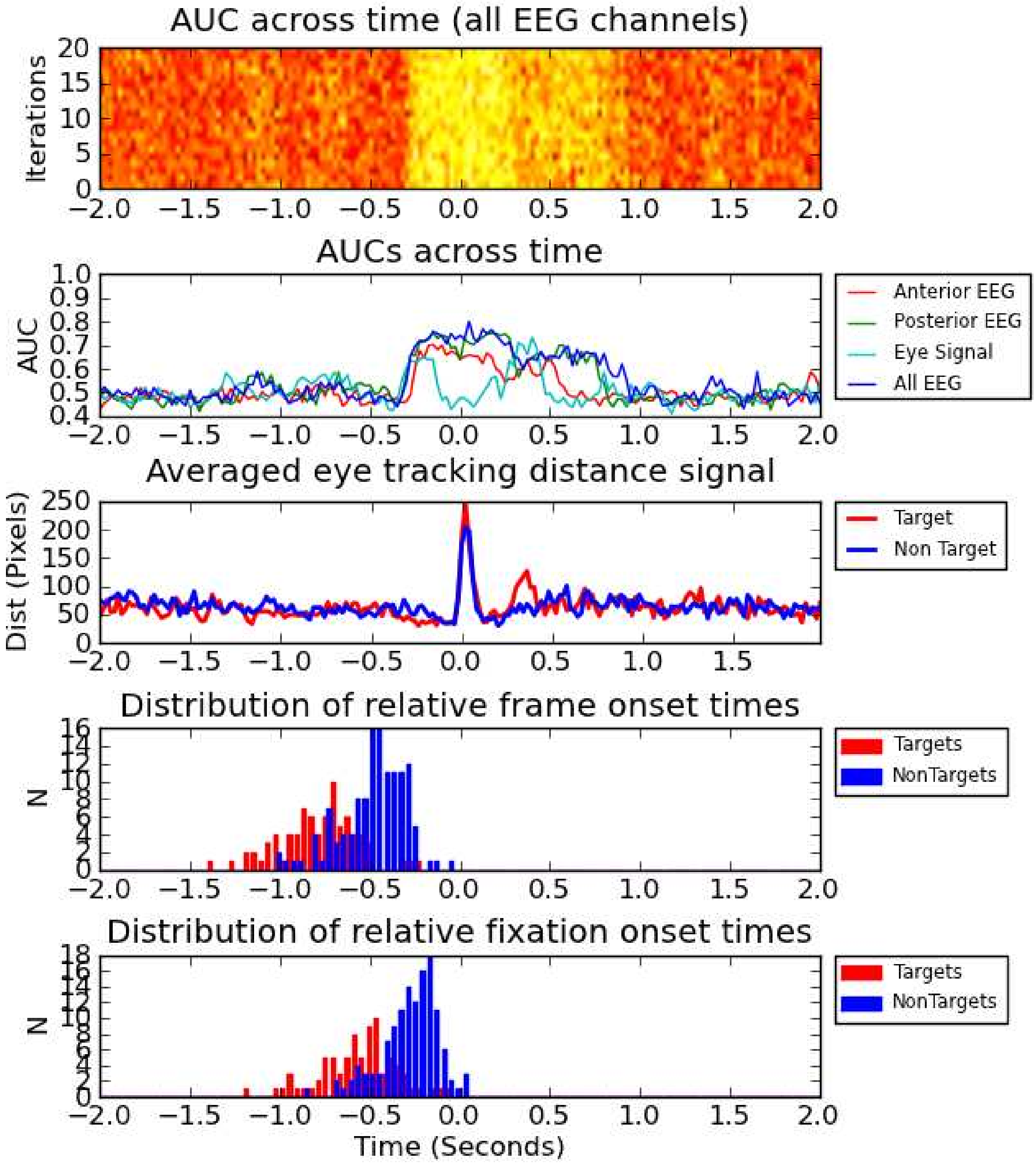

Figure C.3: Subject 1: Temporally aligned discrimination graphs centred on the fixation offset time showing differentiating activity related to target image detection compared to non target image detection

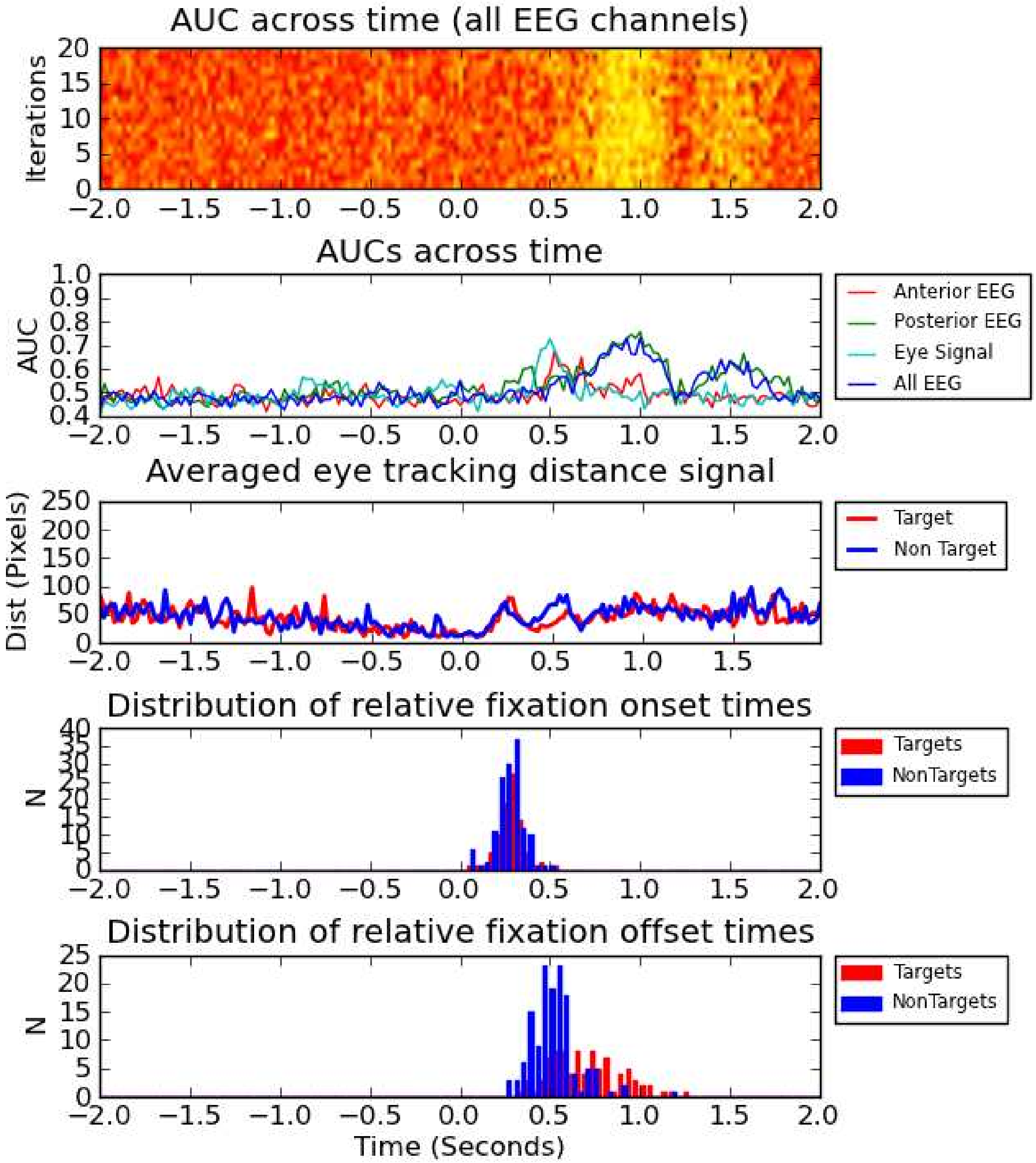

Figure C.4: Subject 2: Temporally aligned discrimination graphs map centred on the frame onset time showing differentiating activity related to target image detection compared to non target image detection

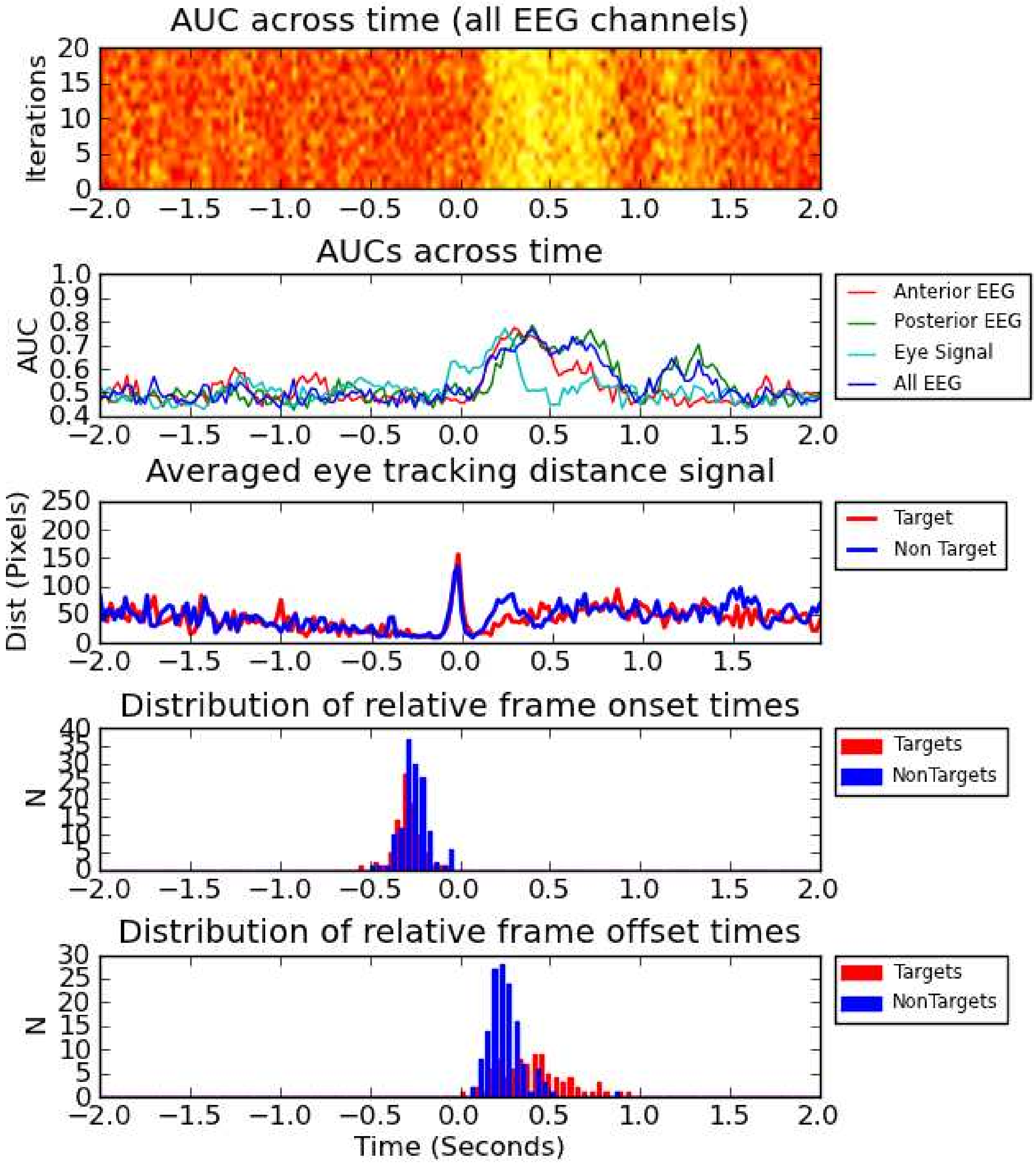

Figure C.5: Subject 2: Temporally aligned discrimination graphs centred on the fixation onset time showing differentiating activity related to target image detection compared to non target image detection

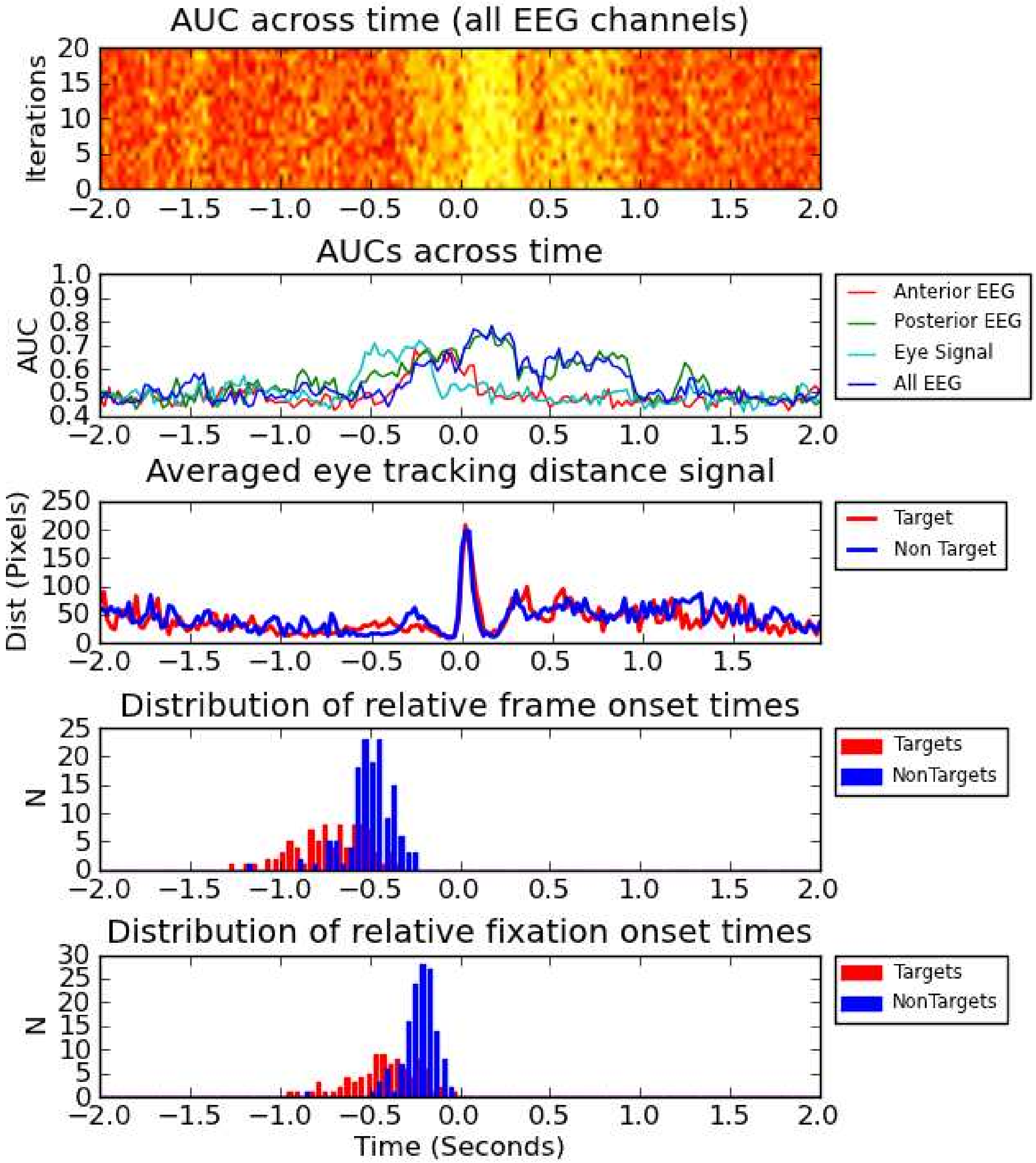

Figure C.6: Subject 2: Temporally aligned discrimination graphs centred on the fixation offset time showing differentiating activity related to target image detection compared to non target image detection

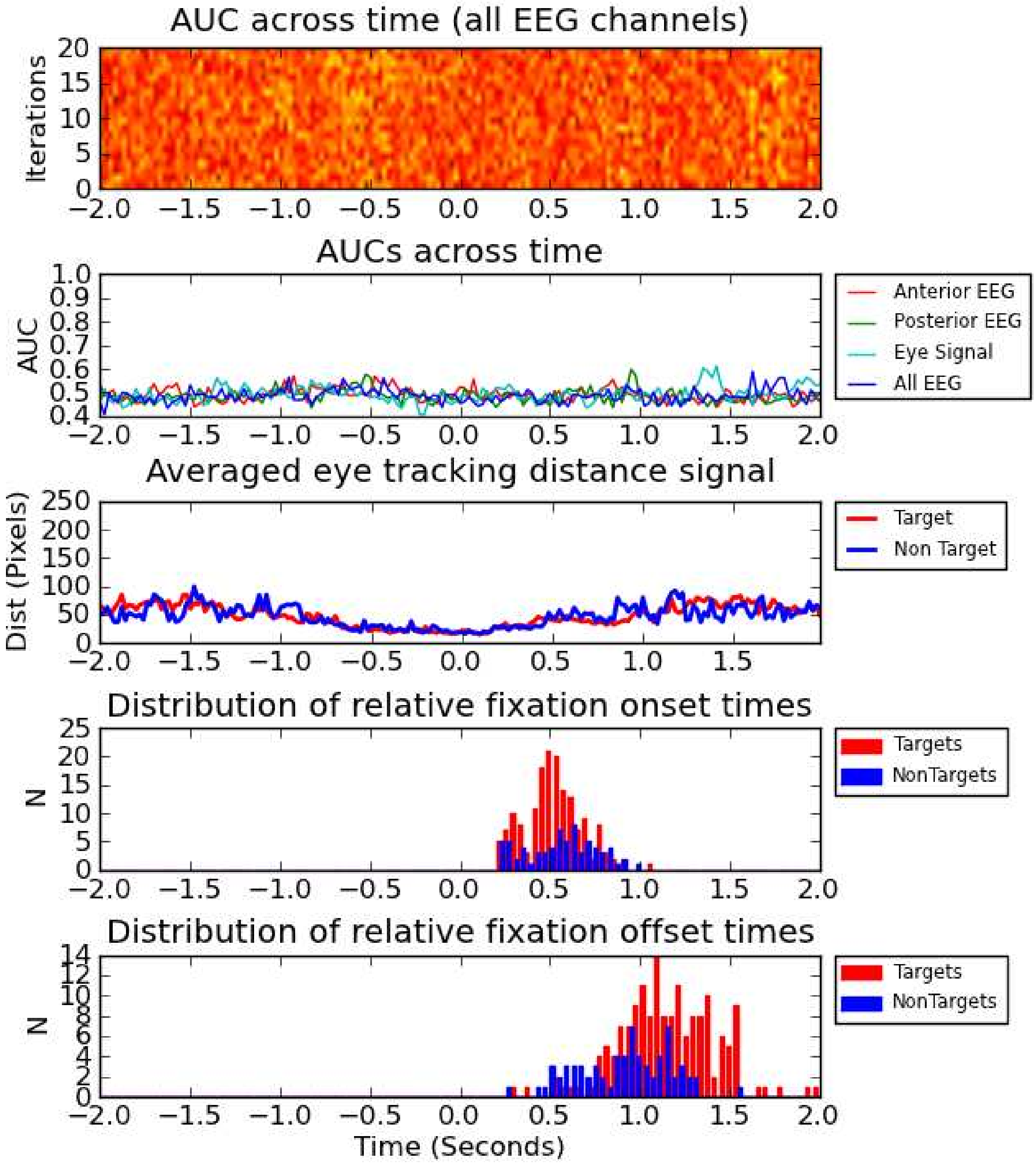

Figure C.7: Subject 3: Temporally aligned discrimination graphs map centred on the frame onset time showing differentiating activity related to target image detection compared to non target image detection

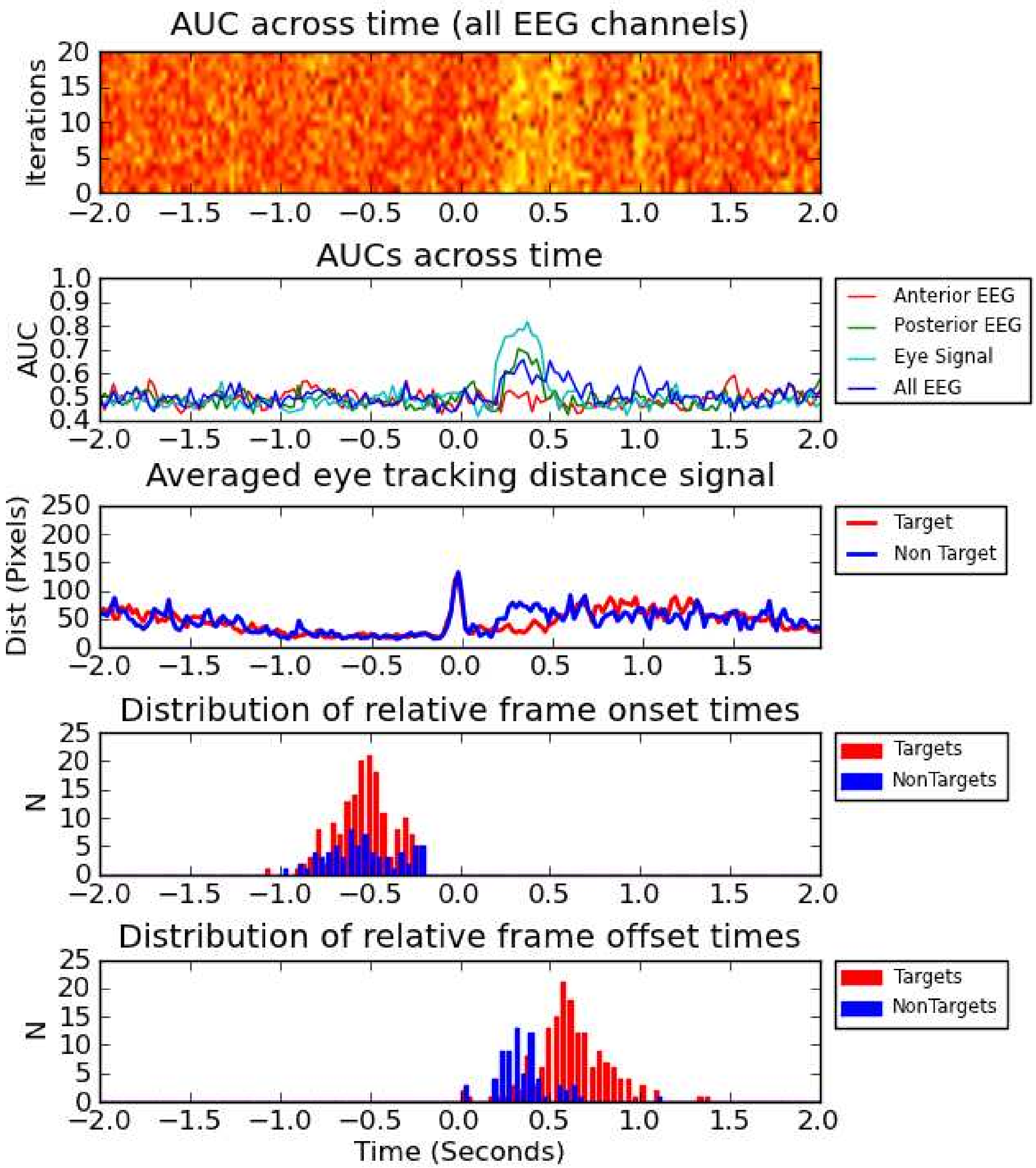

Figure C.8: Subject 3: Temporally aligned discrimination graphs centred on the fixation onset time showing differentiating activity related to target image detection compared to non target image detection

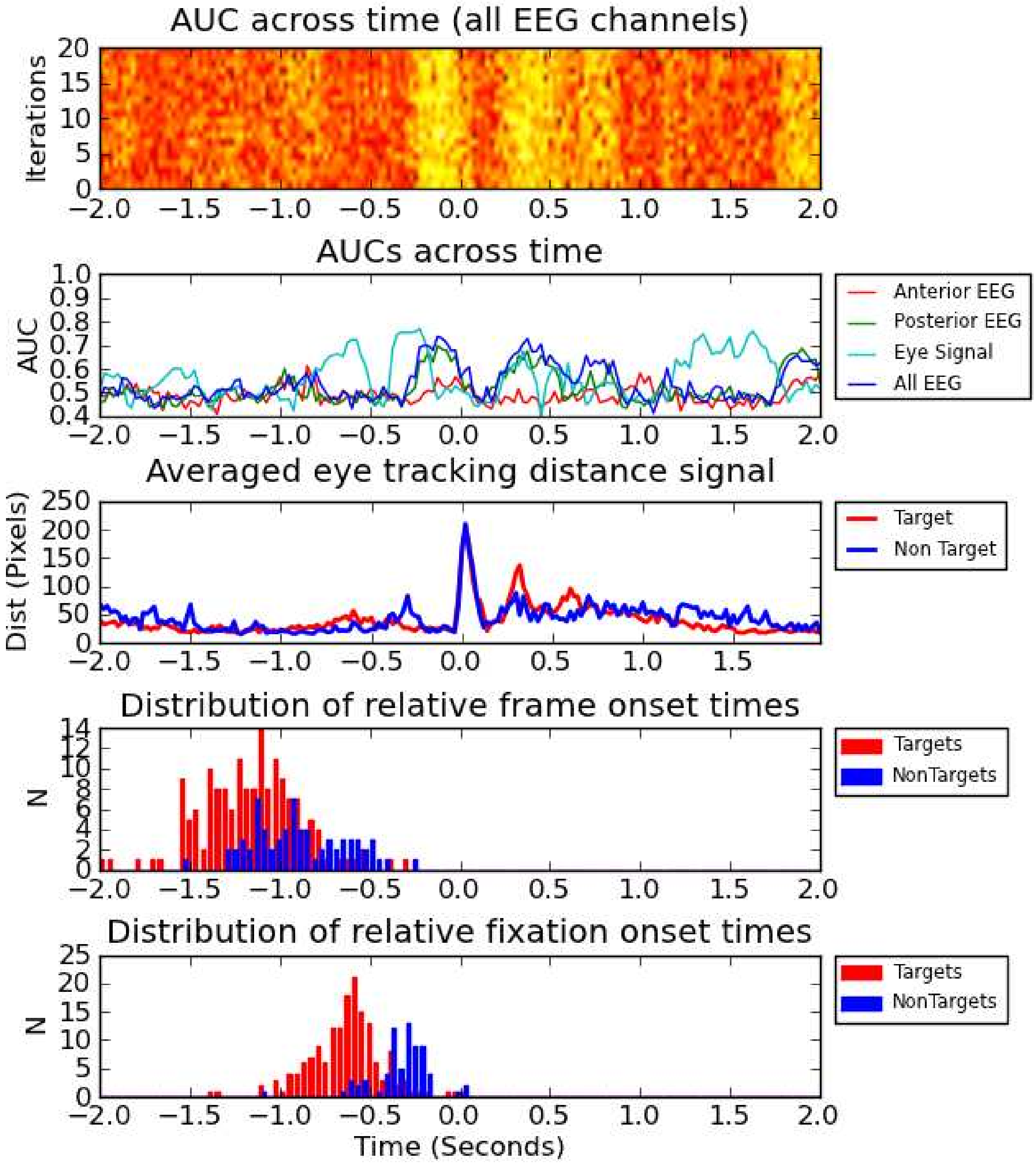

Figure C.9: Subject 3: Temporally aligned discrimination graphs centred on the fixation offset time showing differentiating activity related to target image detection compared to non target image detection

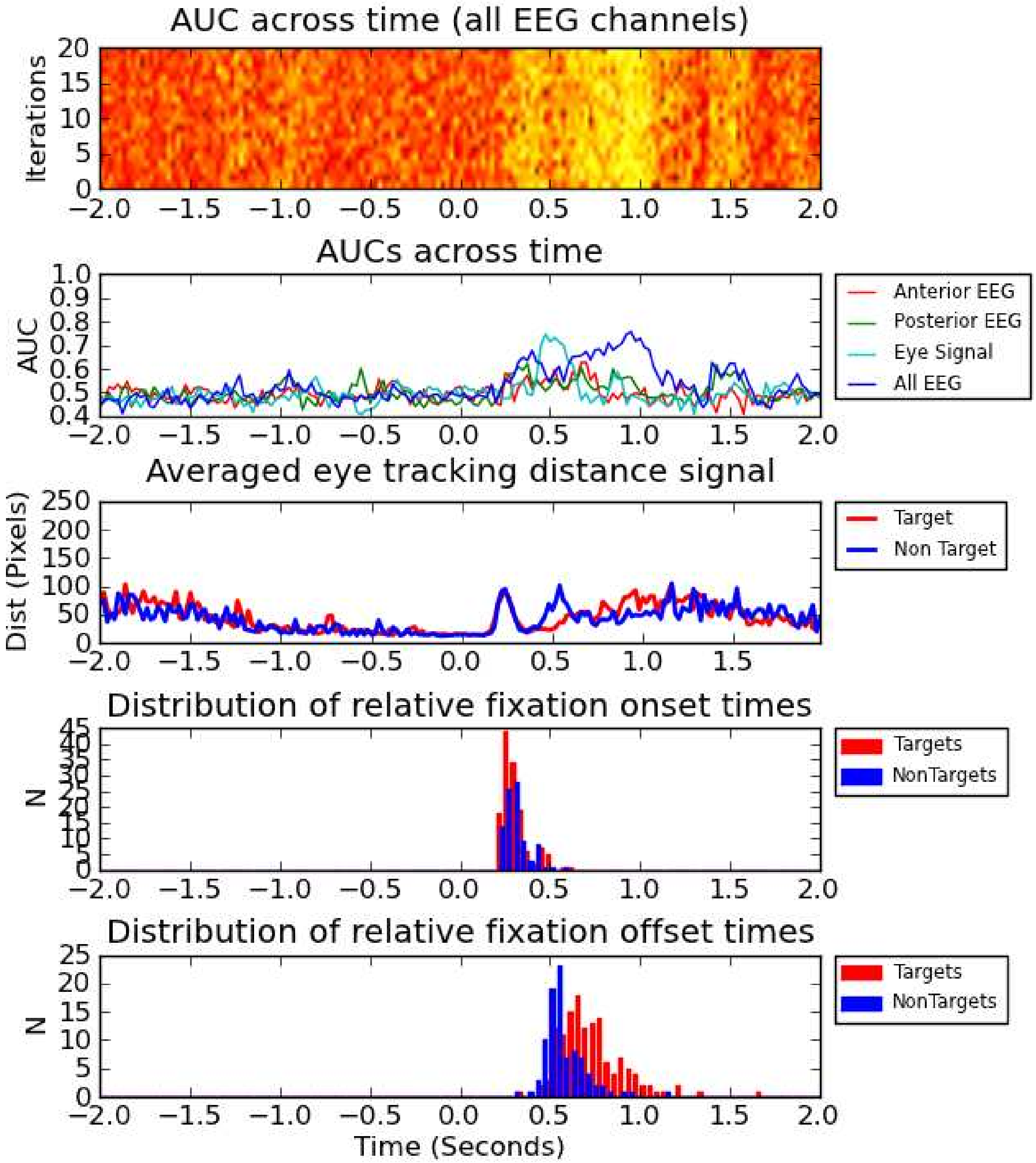

Figure C.10: Subject 4: Temporally aligned discrimination graphs map centred on the frame onset time showing differentiating activity related to target image detection compared to non target image detection

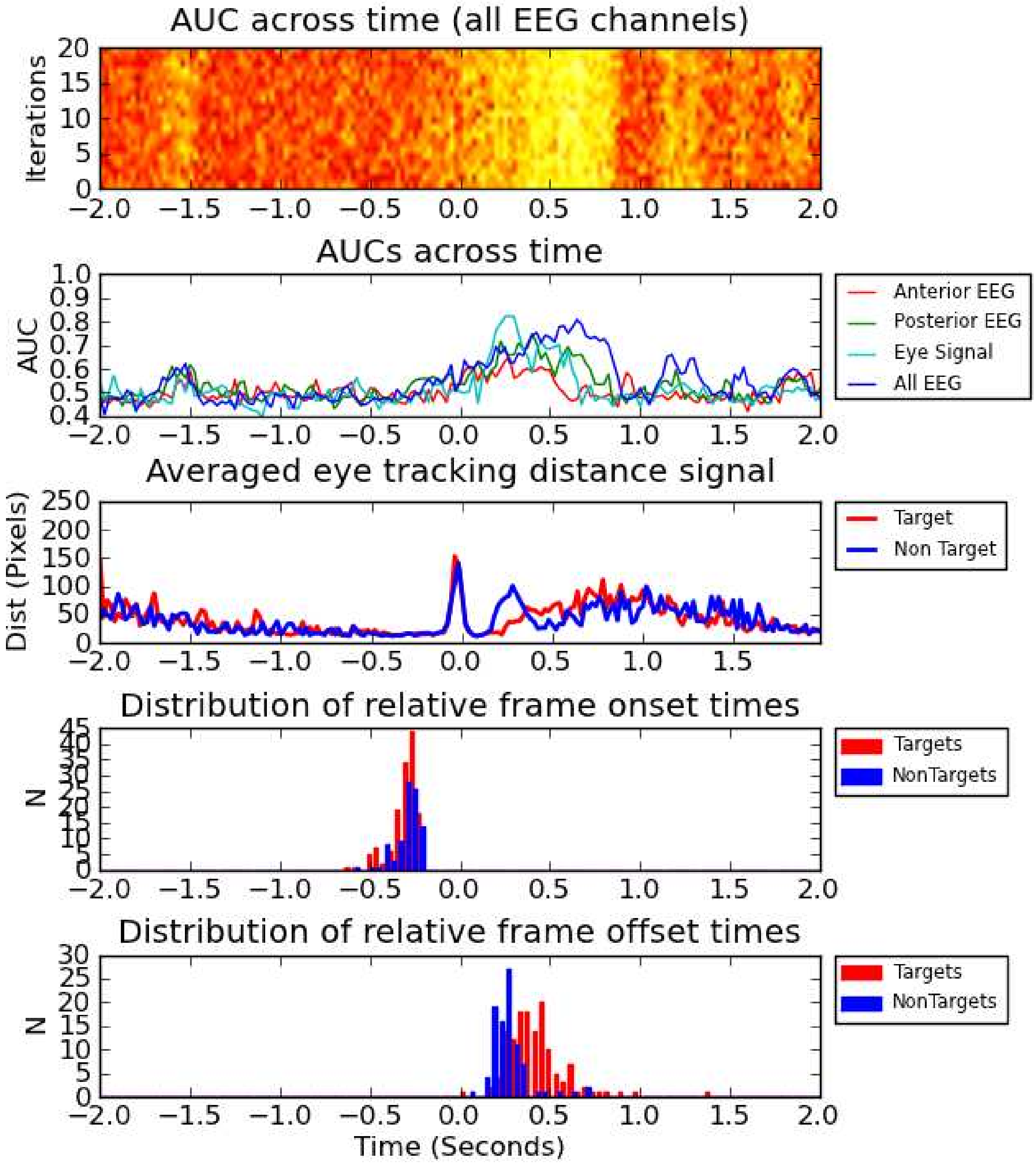

Figure C.11: Subject 4: Temporally aligned discrimination graphs centred on the fixation onset time showing differentiating activity related to target image detection compared to non target image detection

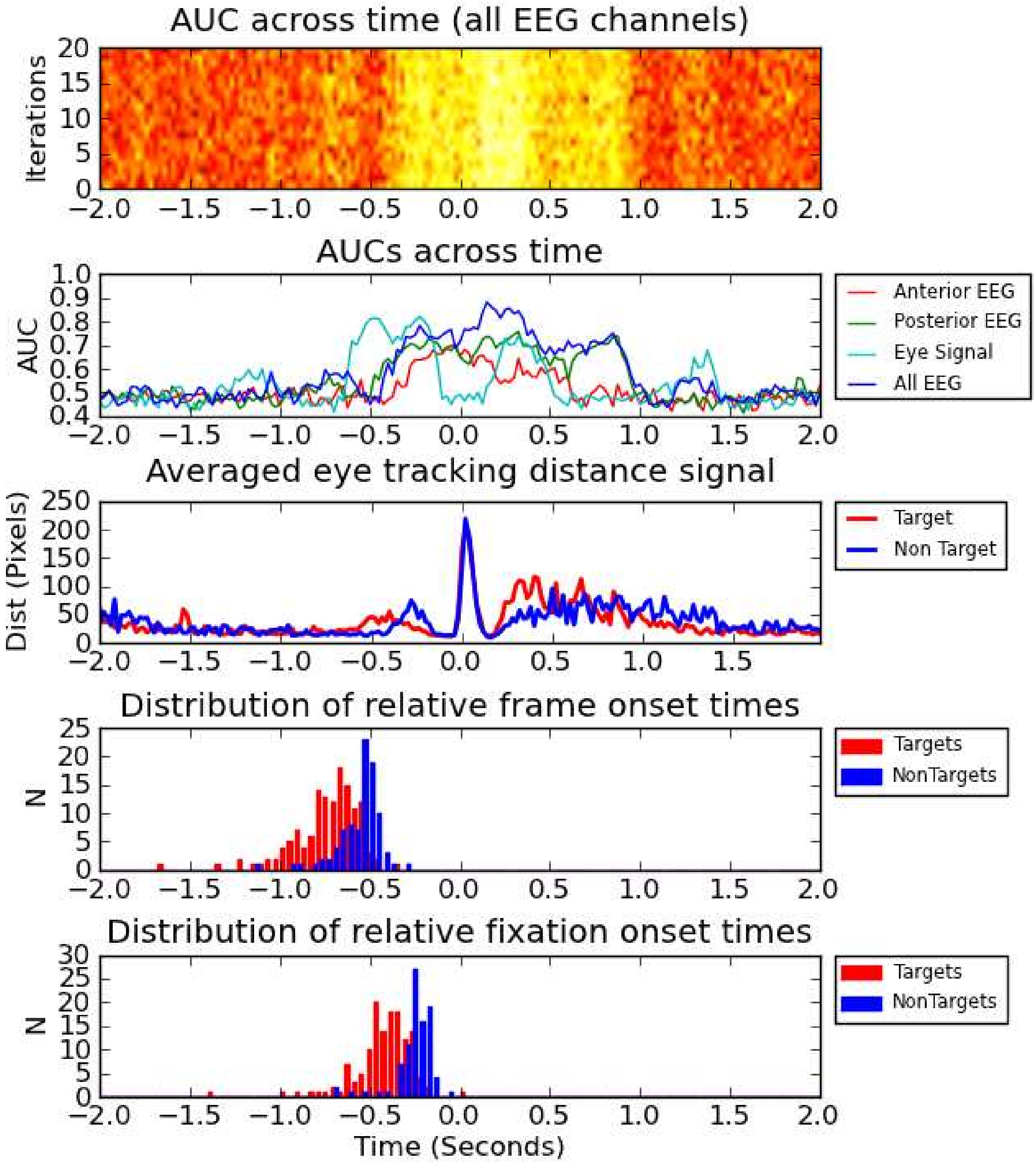

Figure C.12: Subject 4: Temporally aligned discrimination graphs centred on the fixation offset time showing differentiating activity related to target image detection compared to non target image detection

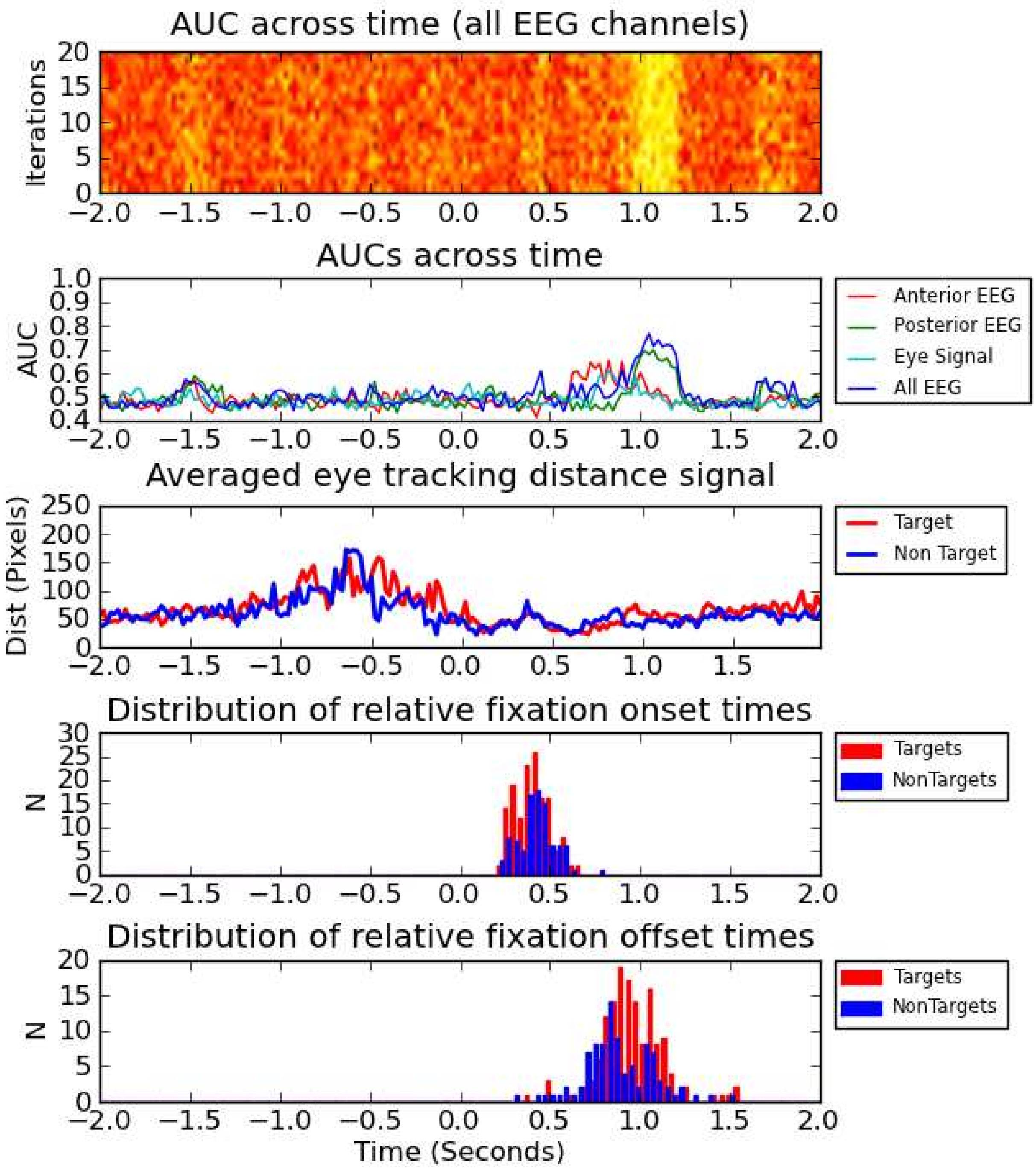

Figure C.13: Subject 5: Temporally aligned discrimination graphs map centred on the frame onset time showing differentiating activity related to target image detection compared to non target image detection

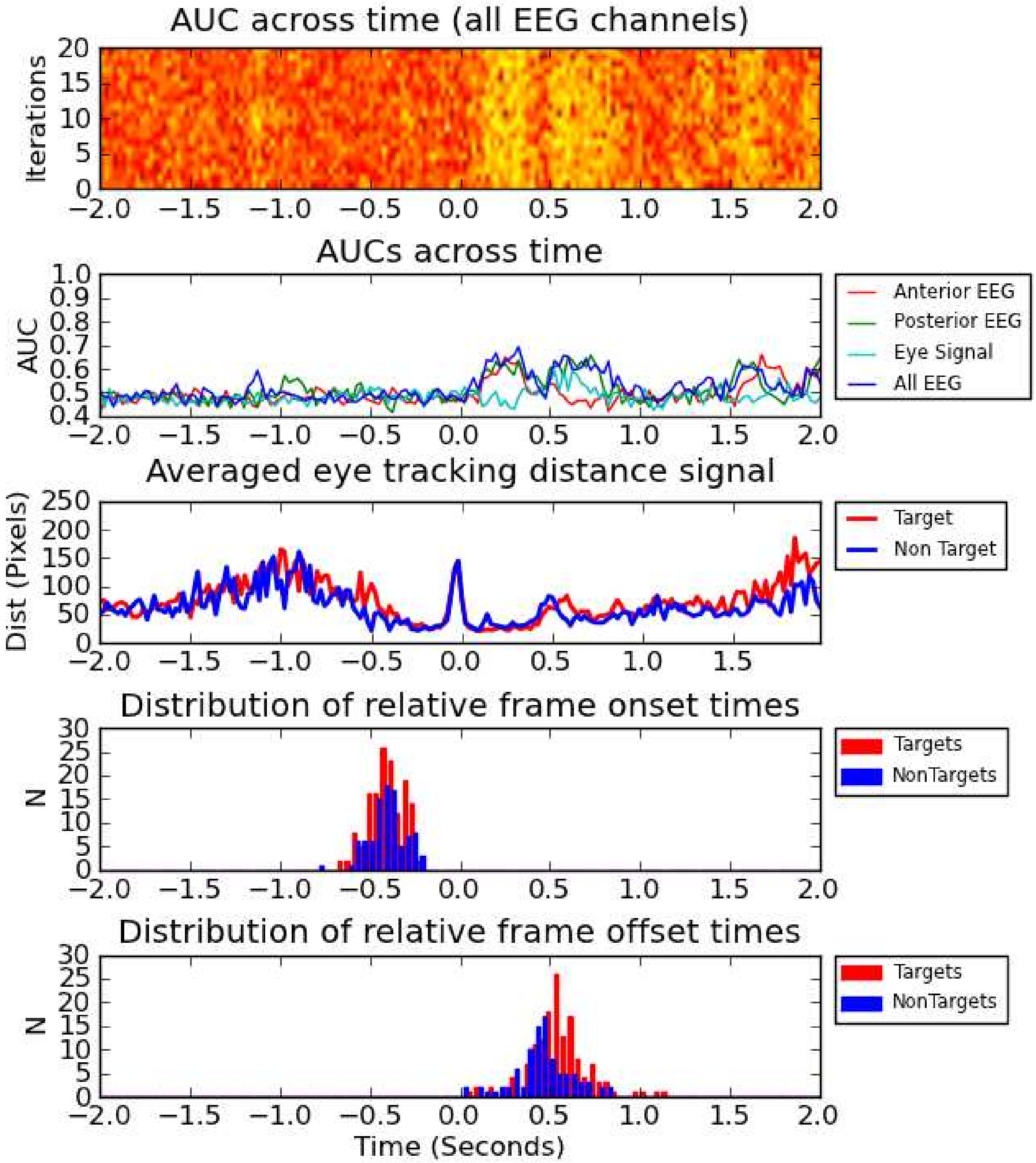

Figure C.14: Subject 5: Temporally aligned discrimination graphs centred on the fixation onset time showing differentiating activity related to target image detection compared to non target image detection

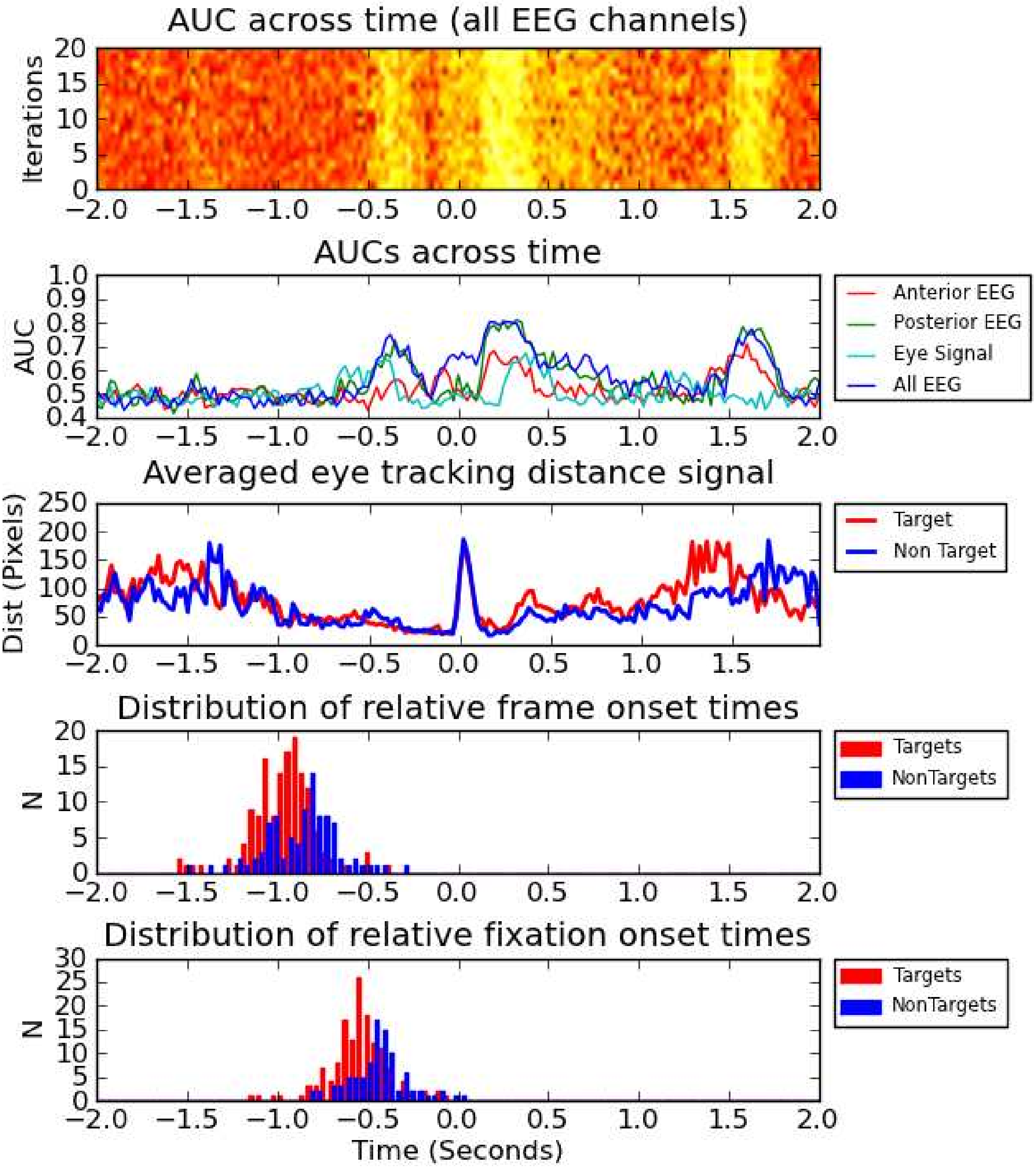

Figure C.15: Subject 5: Temporally aligned discrimination graphs centred on the fixation offset time showing differentiating activity related to target image detection compared to non target image detection

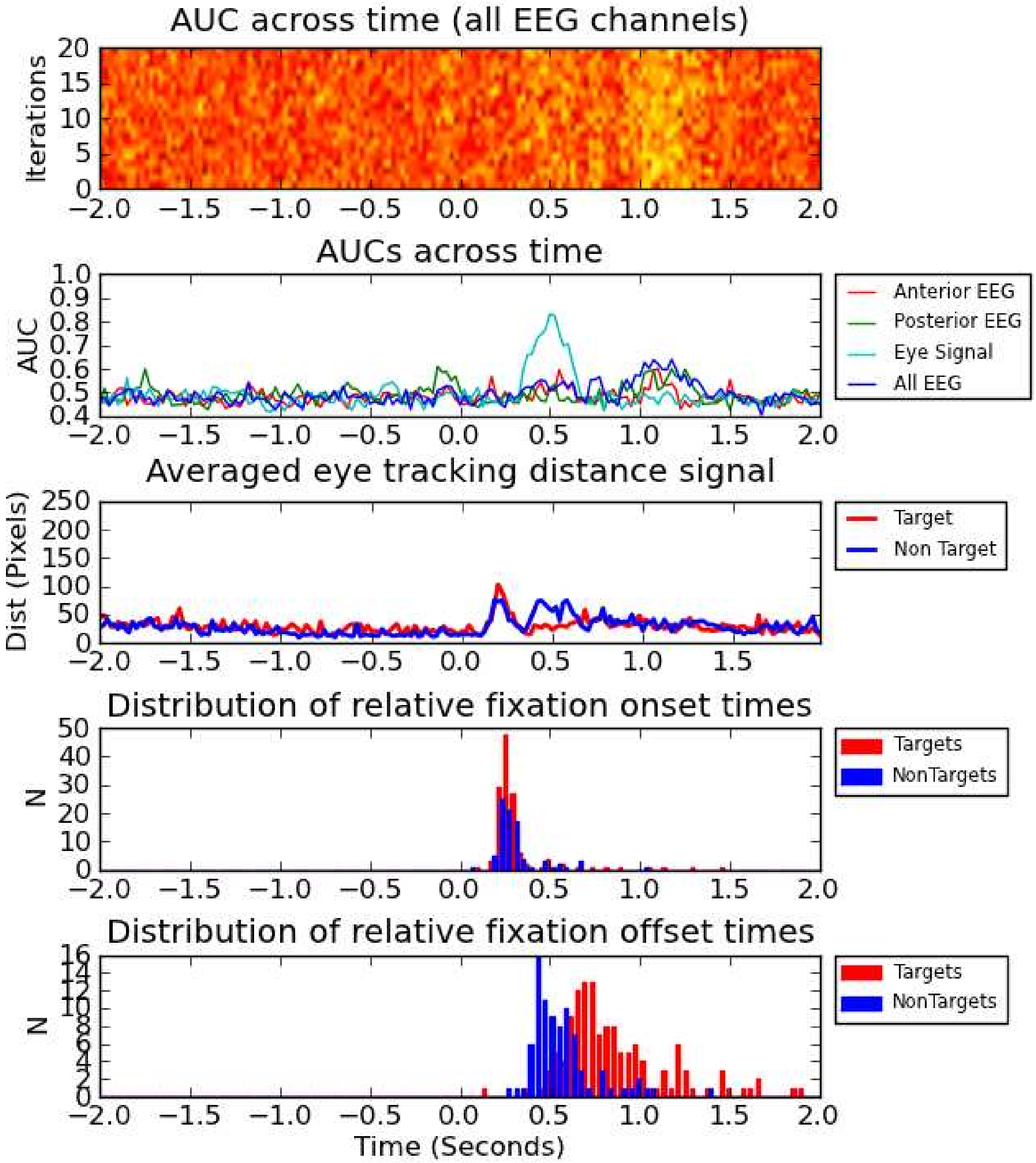

Figure C.16: Subject 6: Temporally aligned discrimination graphs map centred on the frame onset time showing differentiating activity related to target image detection compared to non target image detection

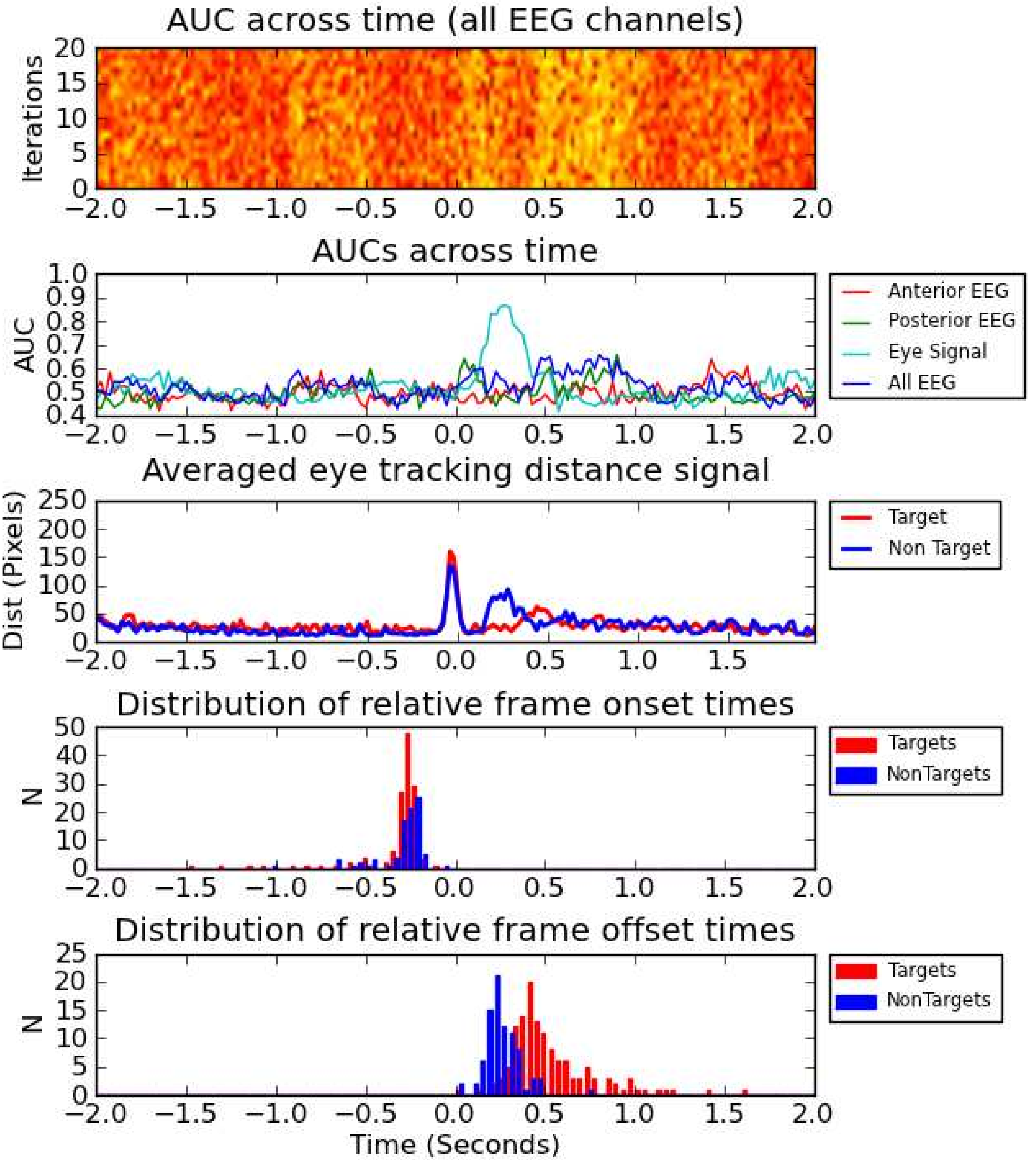

Figure C.17: Subject 6: Temporally aligned discrimination graphs centred on the fixation onset time showing differentiating activity related to target image detection compared to non target image detection

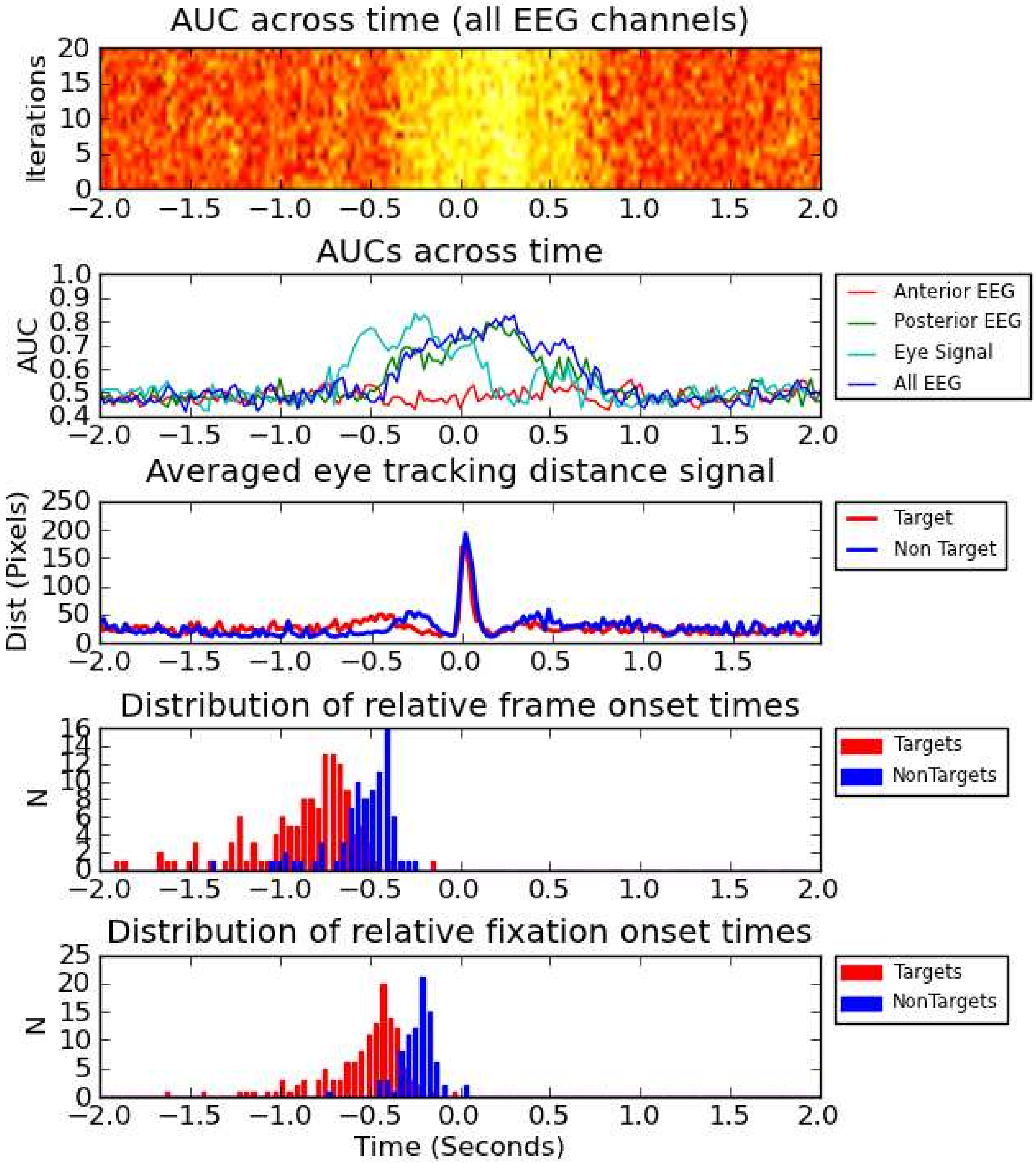

Figure C.18: Subject 6: Temporally aligned discrimination graphs centred on the fixation offset time showing differentiating activity related to target image detection compared to non target image detection

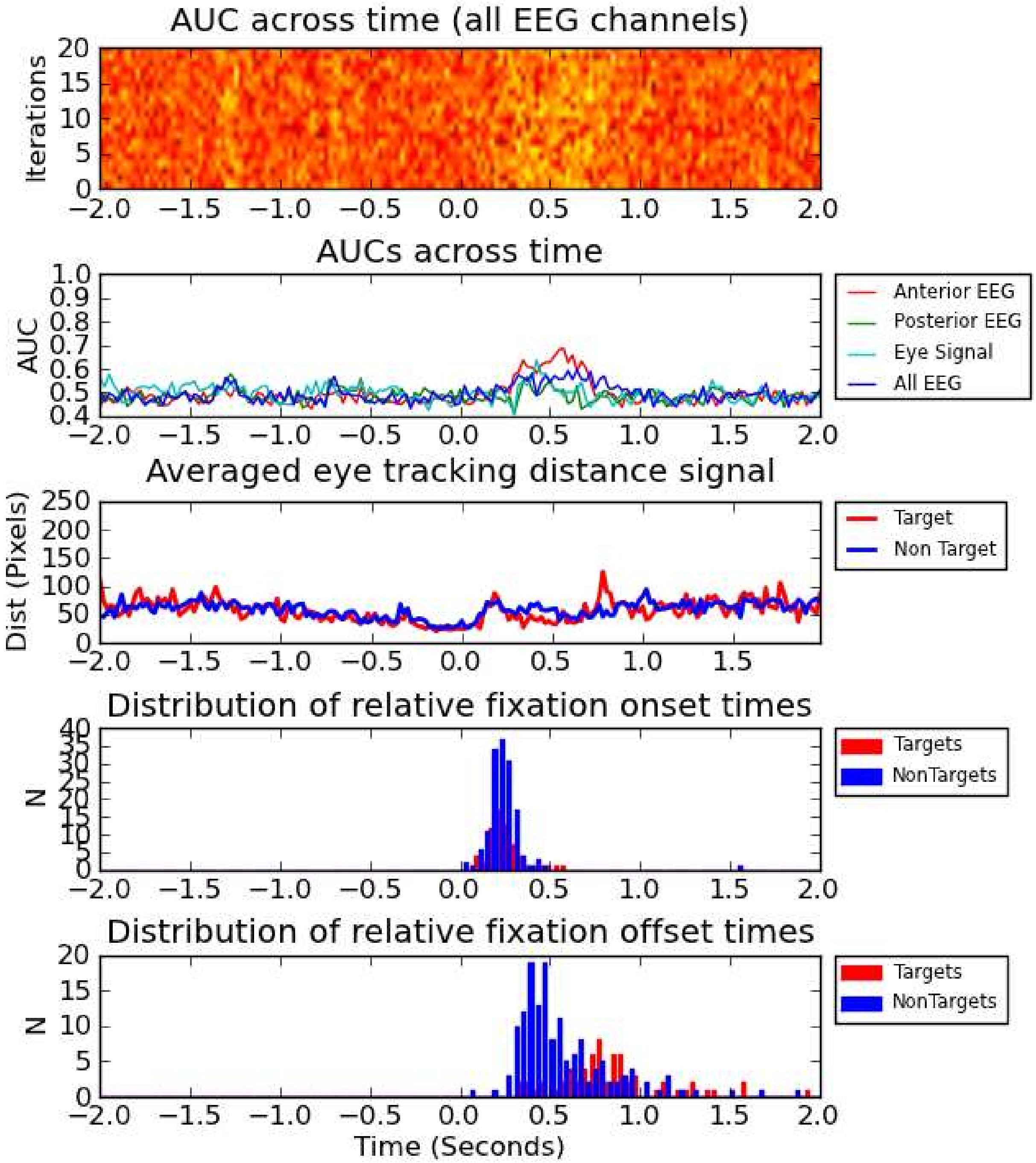

Figure C.19: Subject 7: Temporally aligned discrimination graphs map centred on the frame onset time showing differentiating activity related to target image detection compared to non target image detection

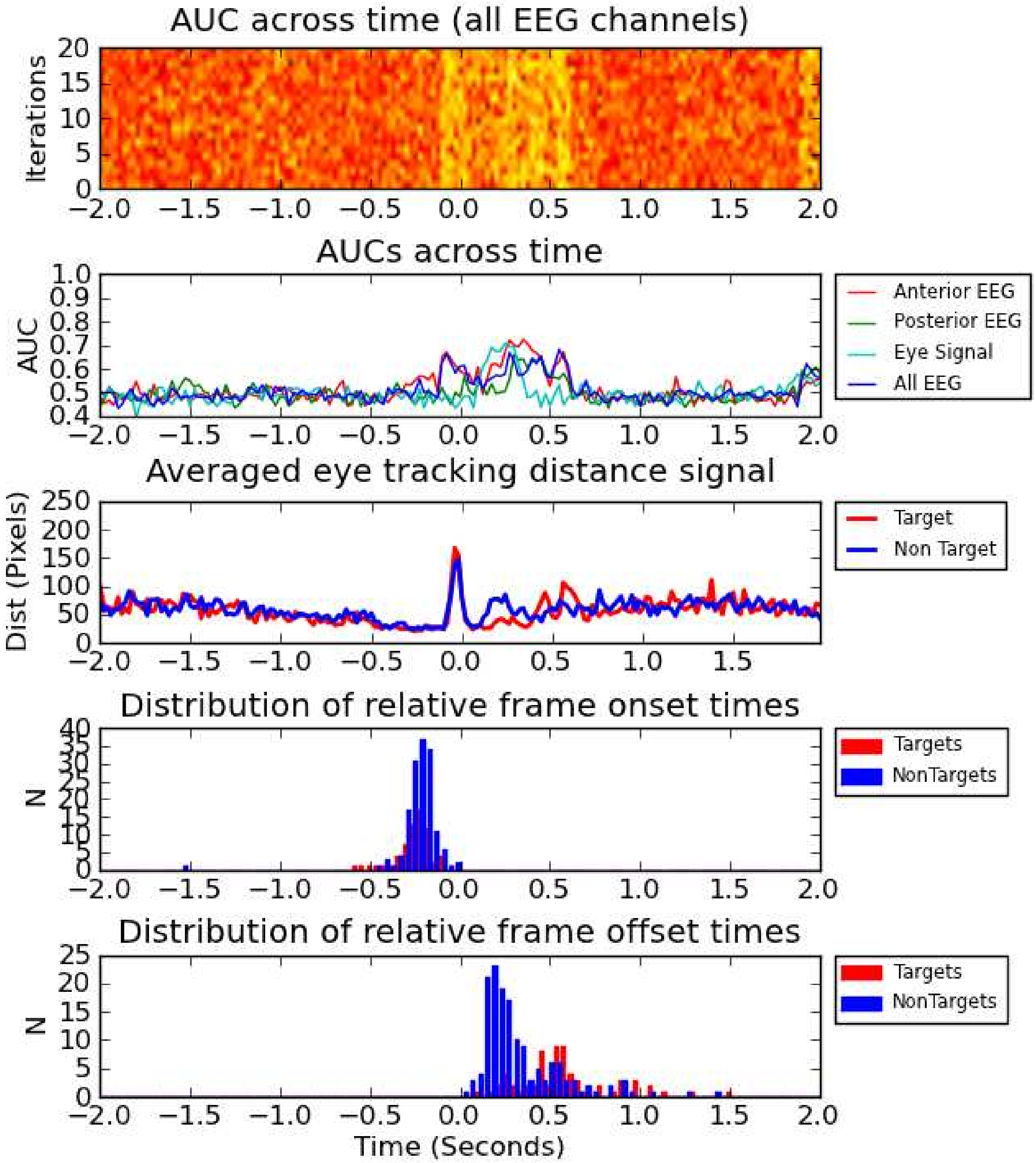

Figure C.20: Subject 7: Temporally aligned discrimination graphs centred on the fixation onset time showing differentiating activity related to target image detection compared to non target image detection

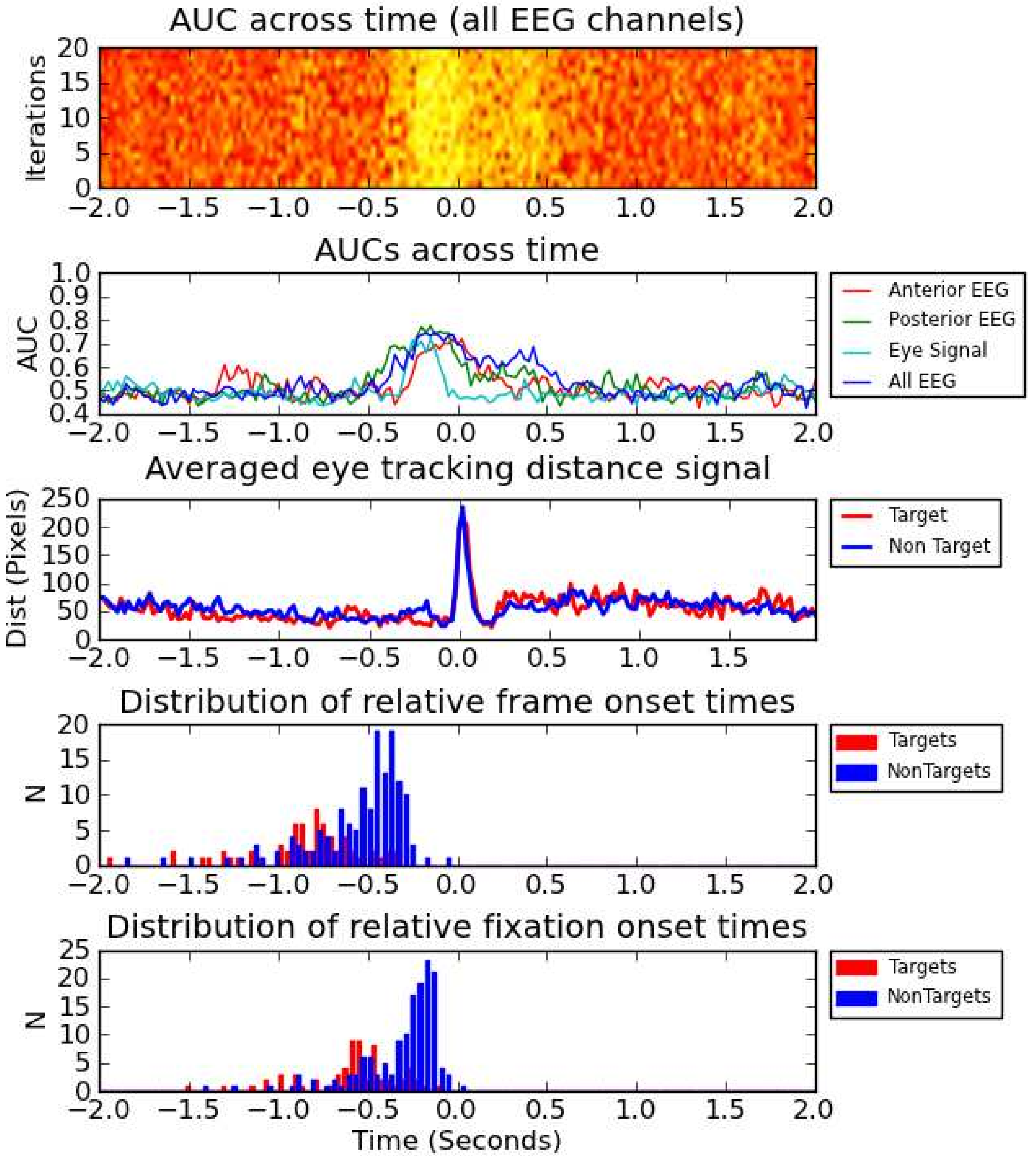

Figure C.21: Subject 7: Temporally aligned discrimination graphs centred on the fixation offset time showing differentiating activity related to target image detection compared to non target image detection

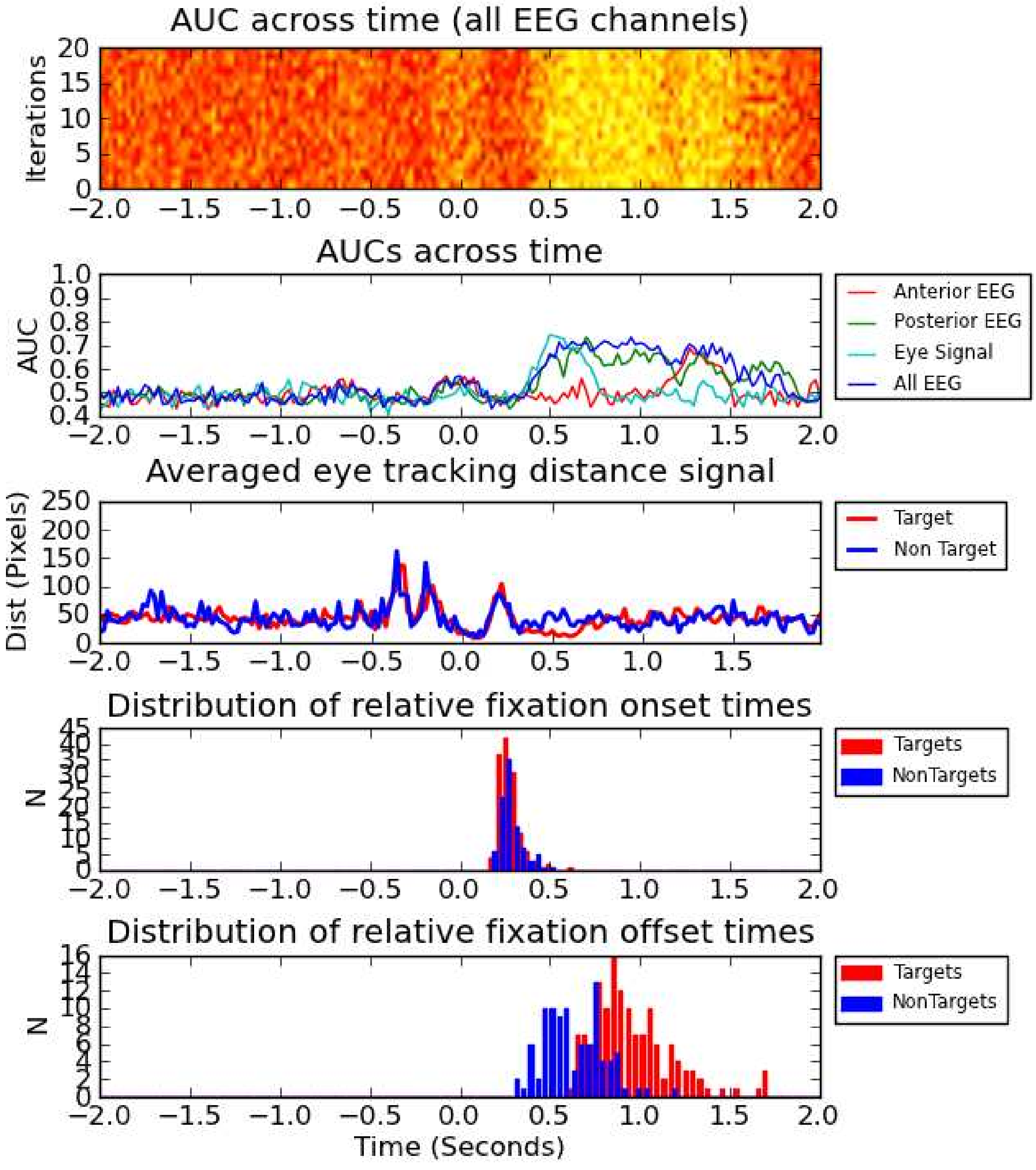

Figure C.22: Subject 8: Temporally aligned discrimination graphs map centred on the frame onset time showing differentiating activity related to target image detection compared to non target image detection

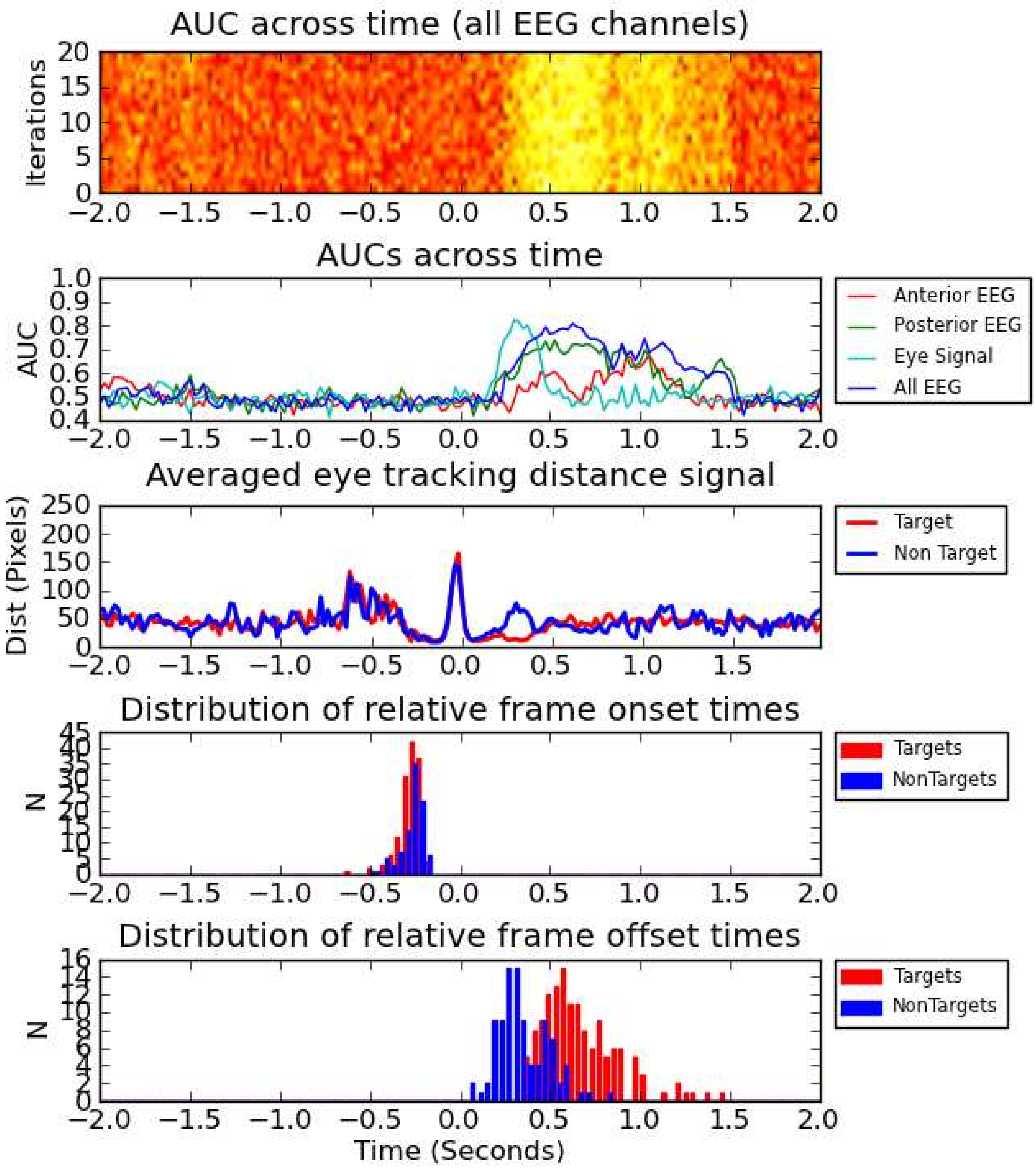

Figure C.23: Subject 8: Temporally aligned discrimination graphs centred on the fixation onset time showing differentiating activity related to target image detection compared to non target image detection

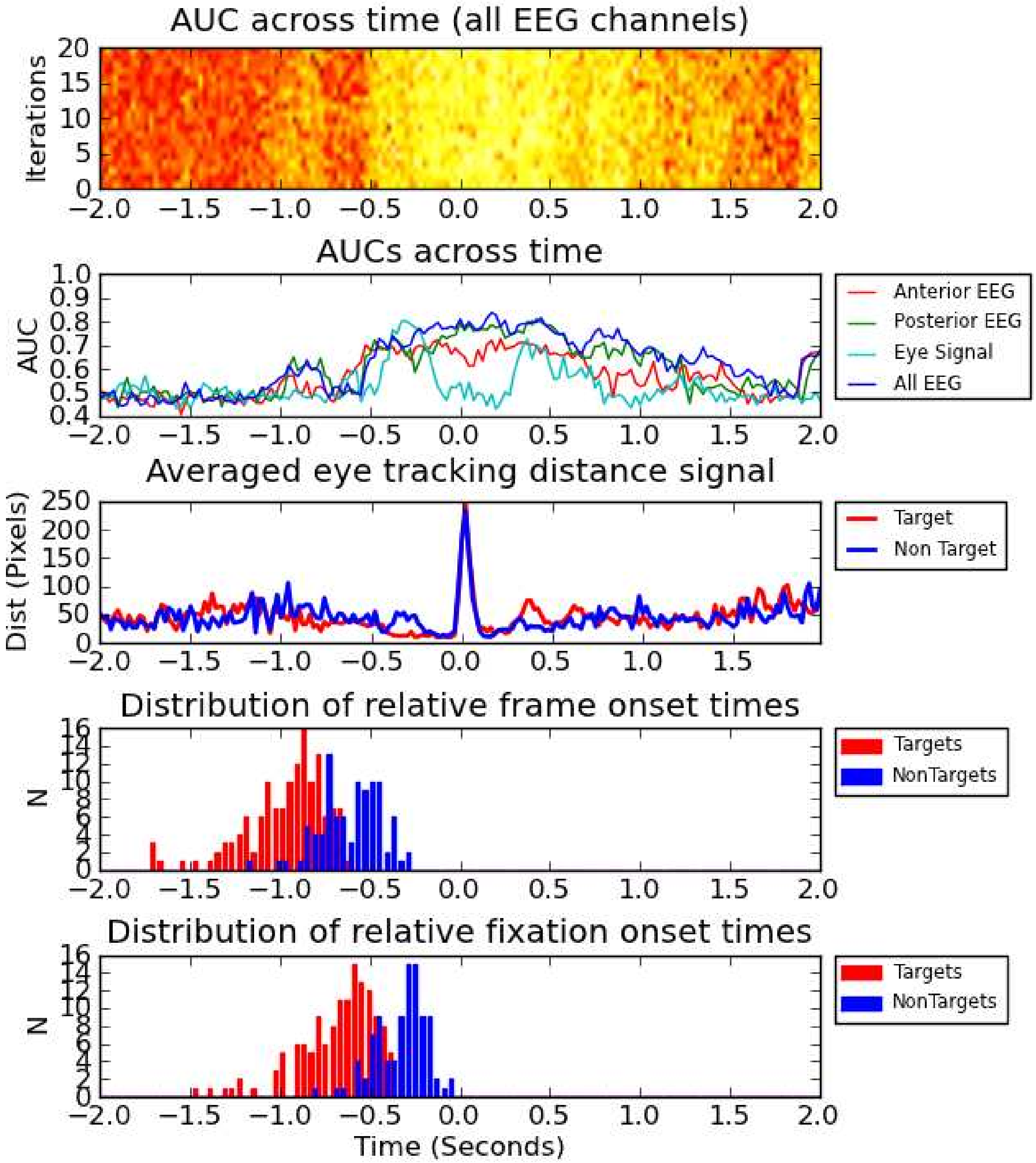

Figure C.24: Subject 8: Temporally aligned discrimination graphs centred on the fixation offset time showing differentiating activity related to target image detection compared to non target image detection

C.2 Scalp Plots for Experiment 2

In this section we provide the scalp plots obtained for experiment 2 across all subjects. These materials are intended to primarily supplement Section 3.3.

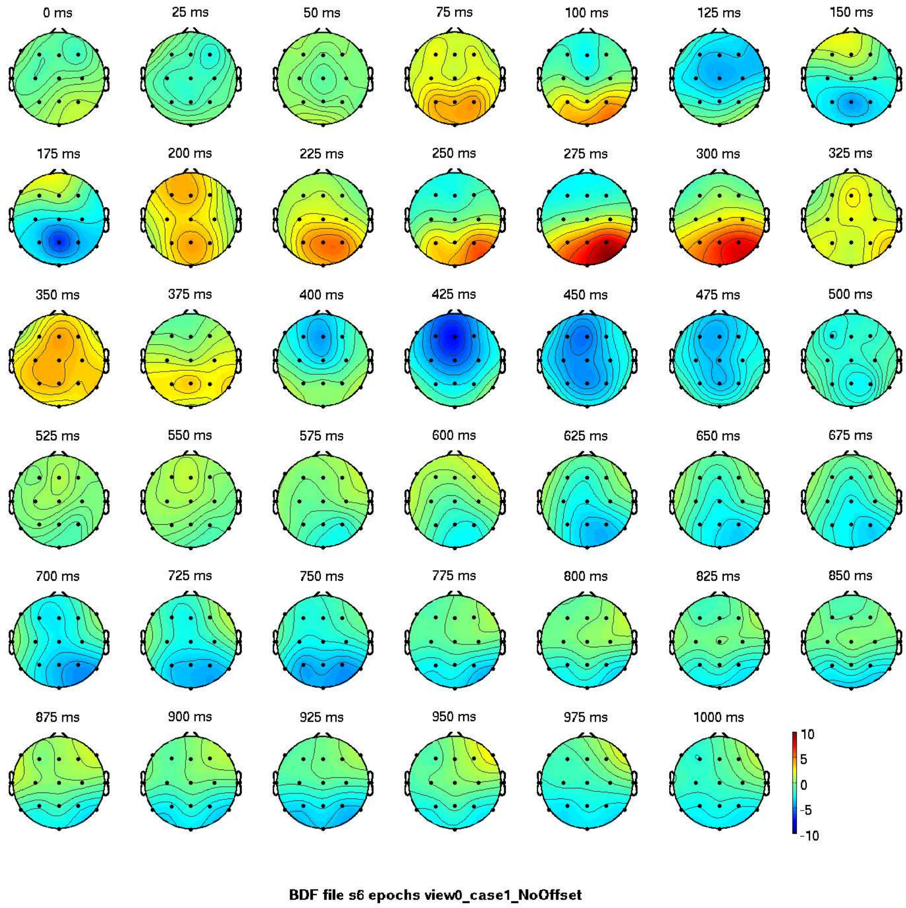

Figure C.25: Subject 1: Averaged scalp plots aligned to frame onset for target frames

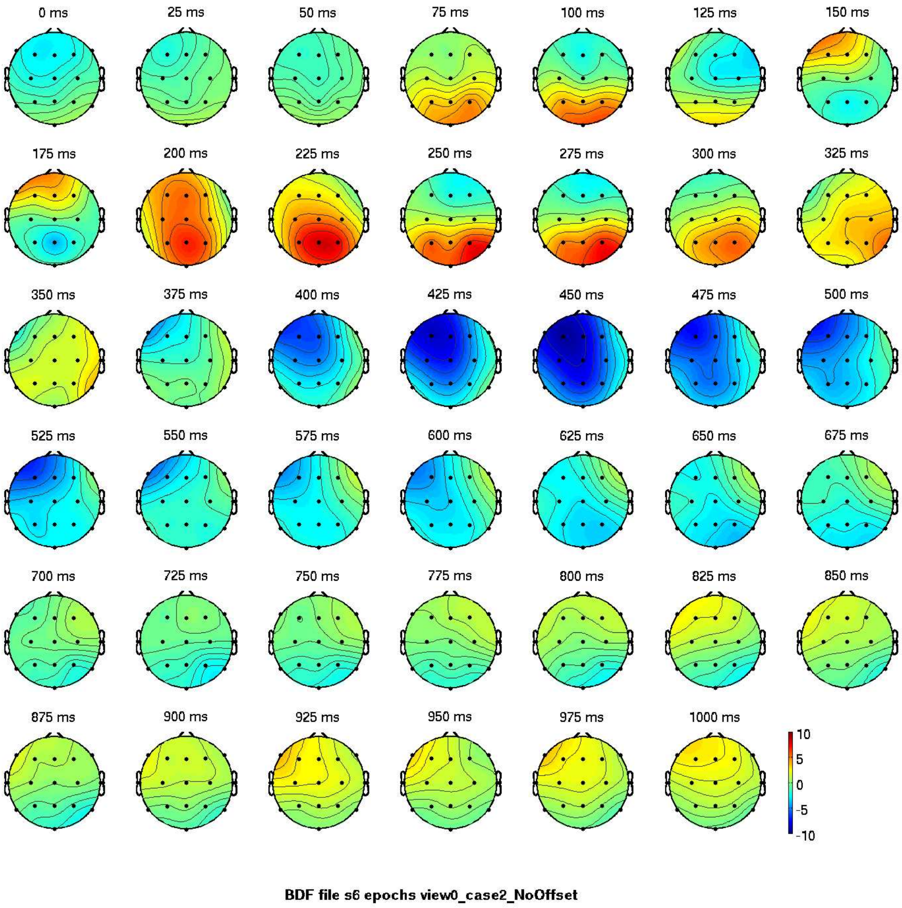

Figure C.26: Subject 1: Averaged scalp plots aligned to frame onset for non-target frames

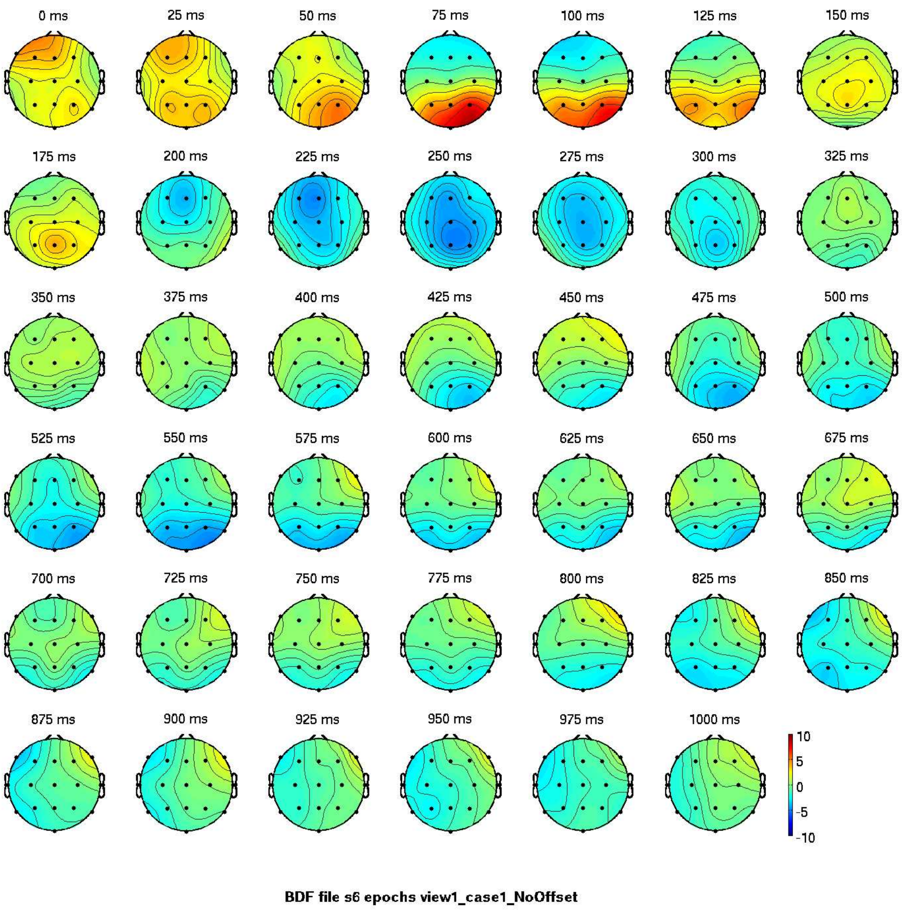

Figure C.27: Subject 1: Averaged scalp plots aligned to fixation onset for target frames

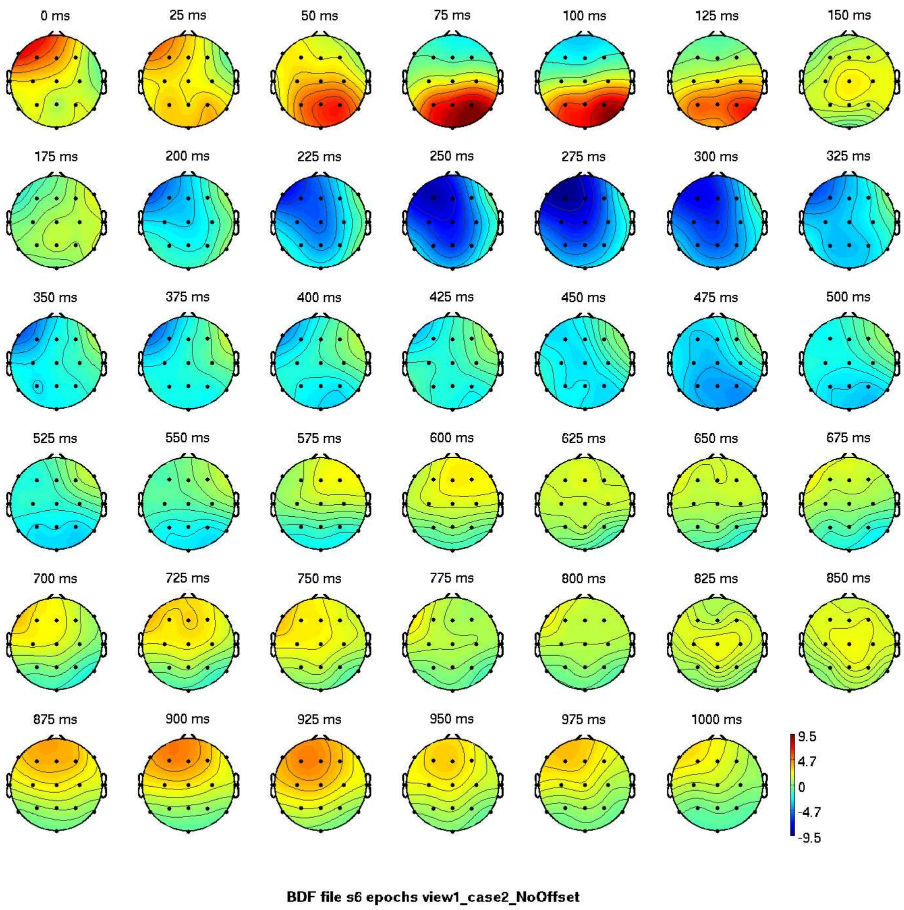

Figure C.28: Subject 1: Averaged scalp plots aligned to fixation onset for non-target frames

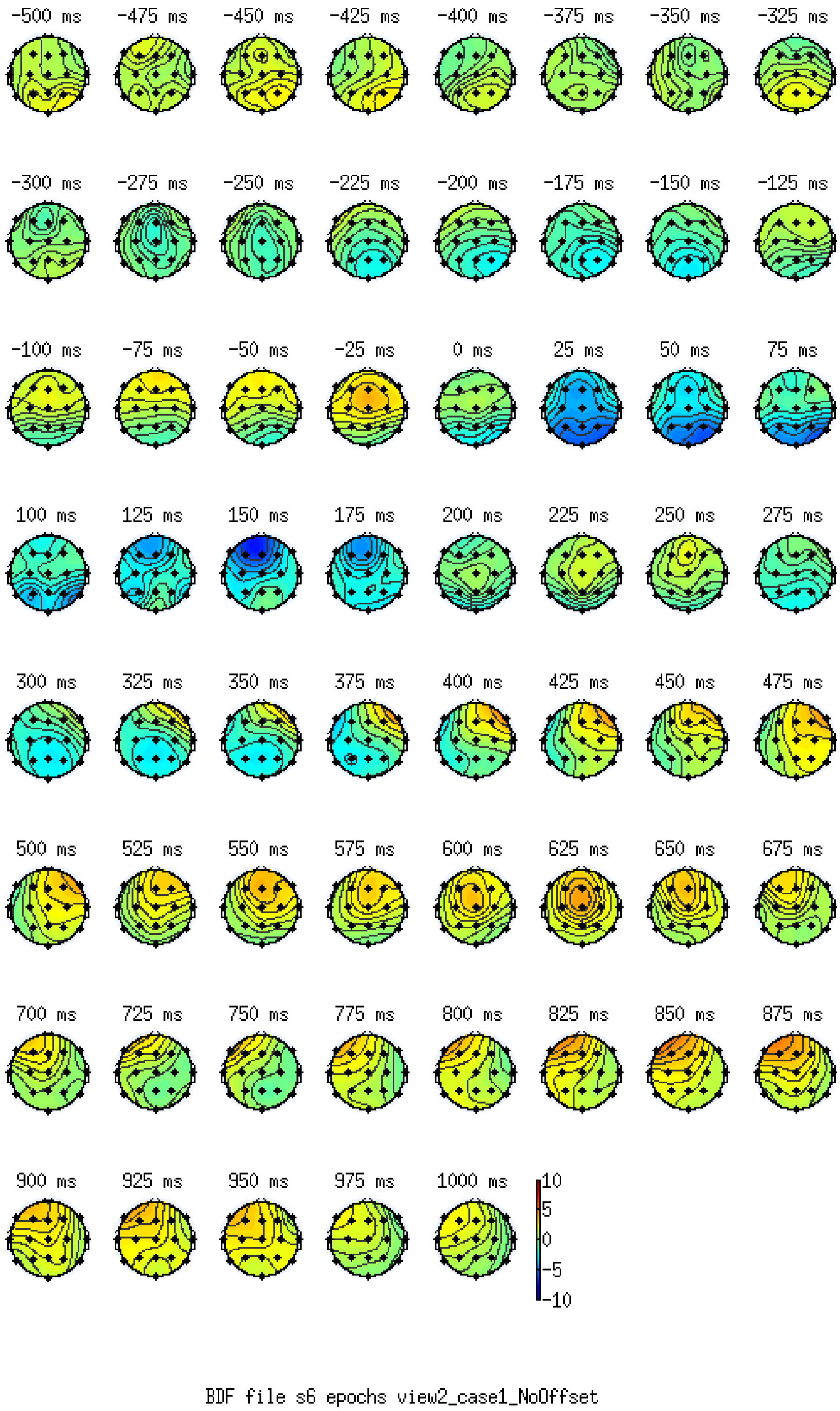

Figure C.29: Subject 1: Averaged scalp plots aligned to fixation offset for target frames

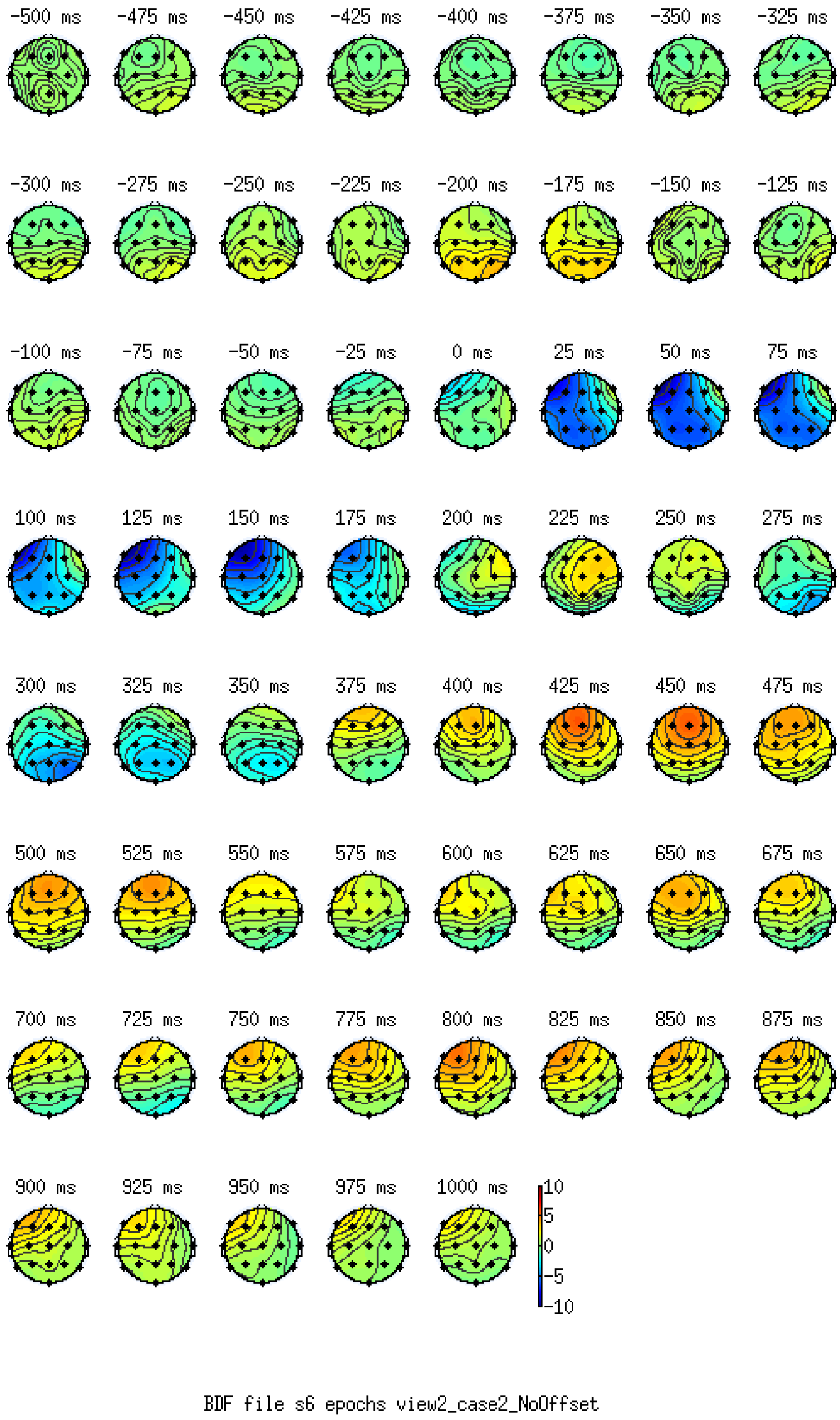

Figure C.30: Subject 1: Averaged scalp plots aligned to fixation offset for non-target frames

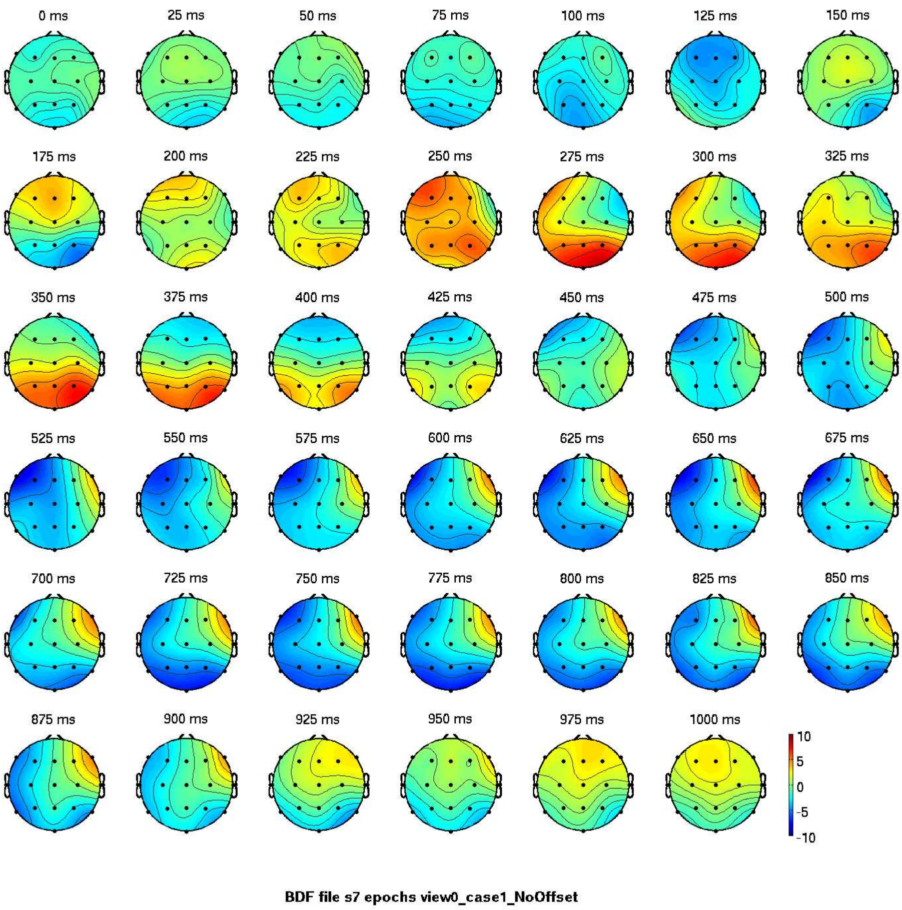

Figure C.31: Subject 2: Averaged scalp plots aligned to frame onset for target frames

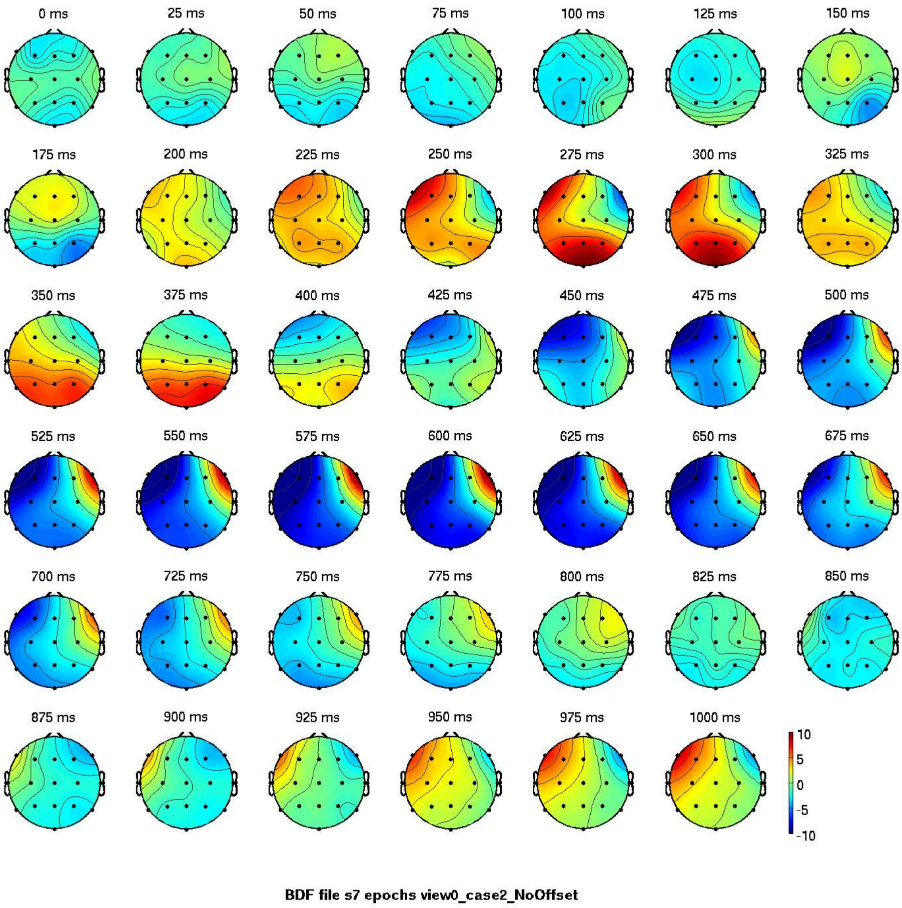

Figure C.32: Subject 2: Averaged scalp plots aligned to frame onset for non-target frames

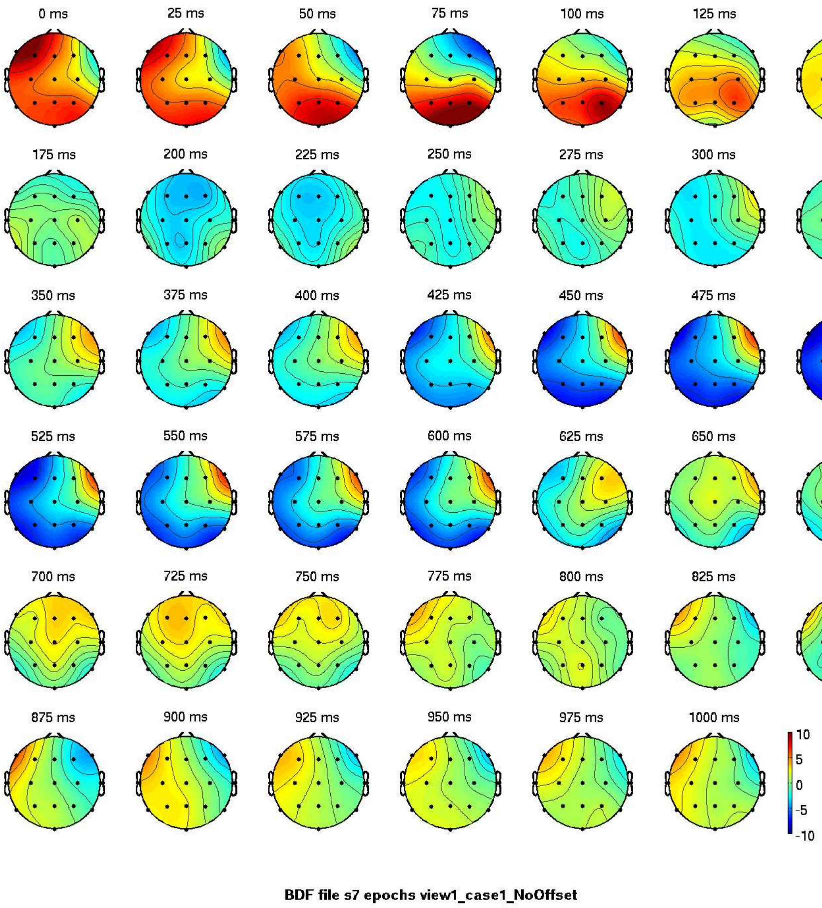

Figure C.33: Subject 2: Averaged scalp plots aligned to fixation onset for target frames

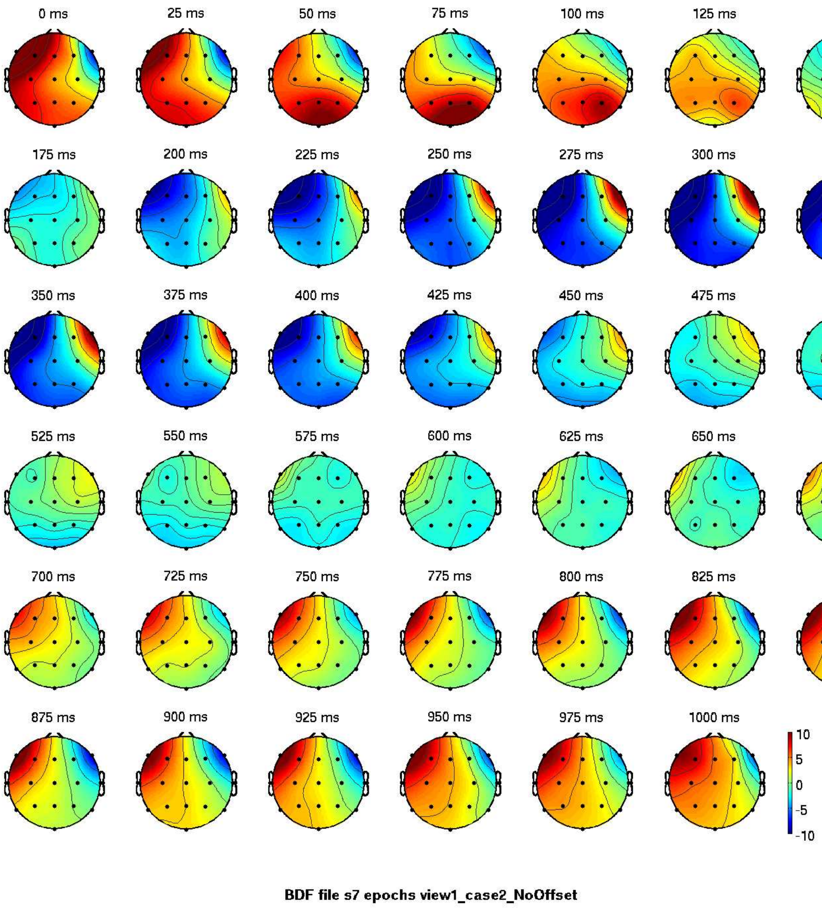

Figure C.34: Subject 2: Averaged scalp plots aligned to fixation onset for non-target frames

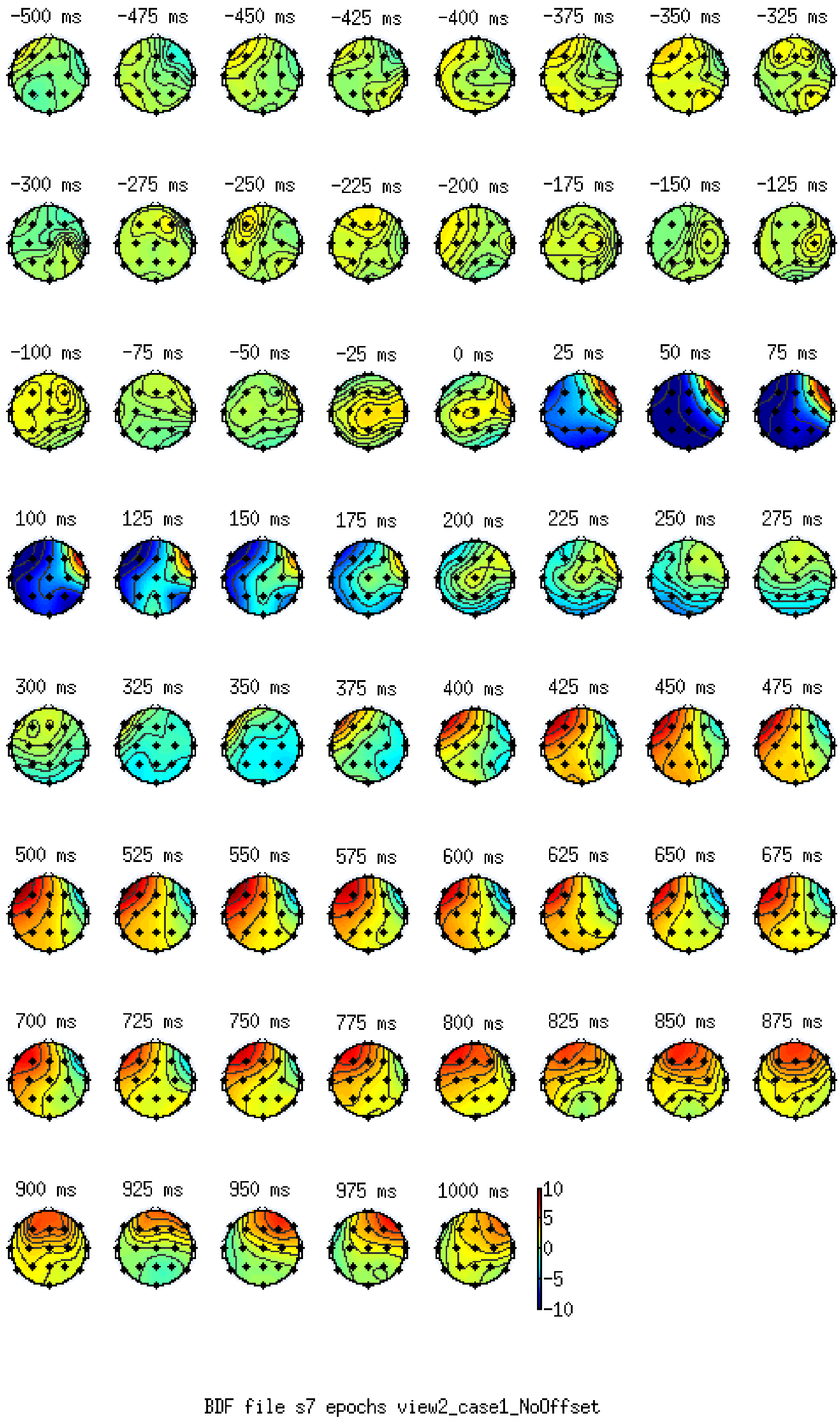

Figure C.35: Subject 2: Averaged scalp plots aligned to fixation offset for target frames

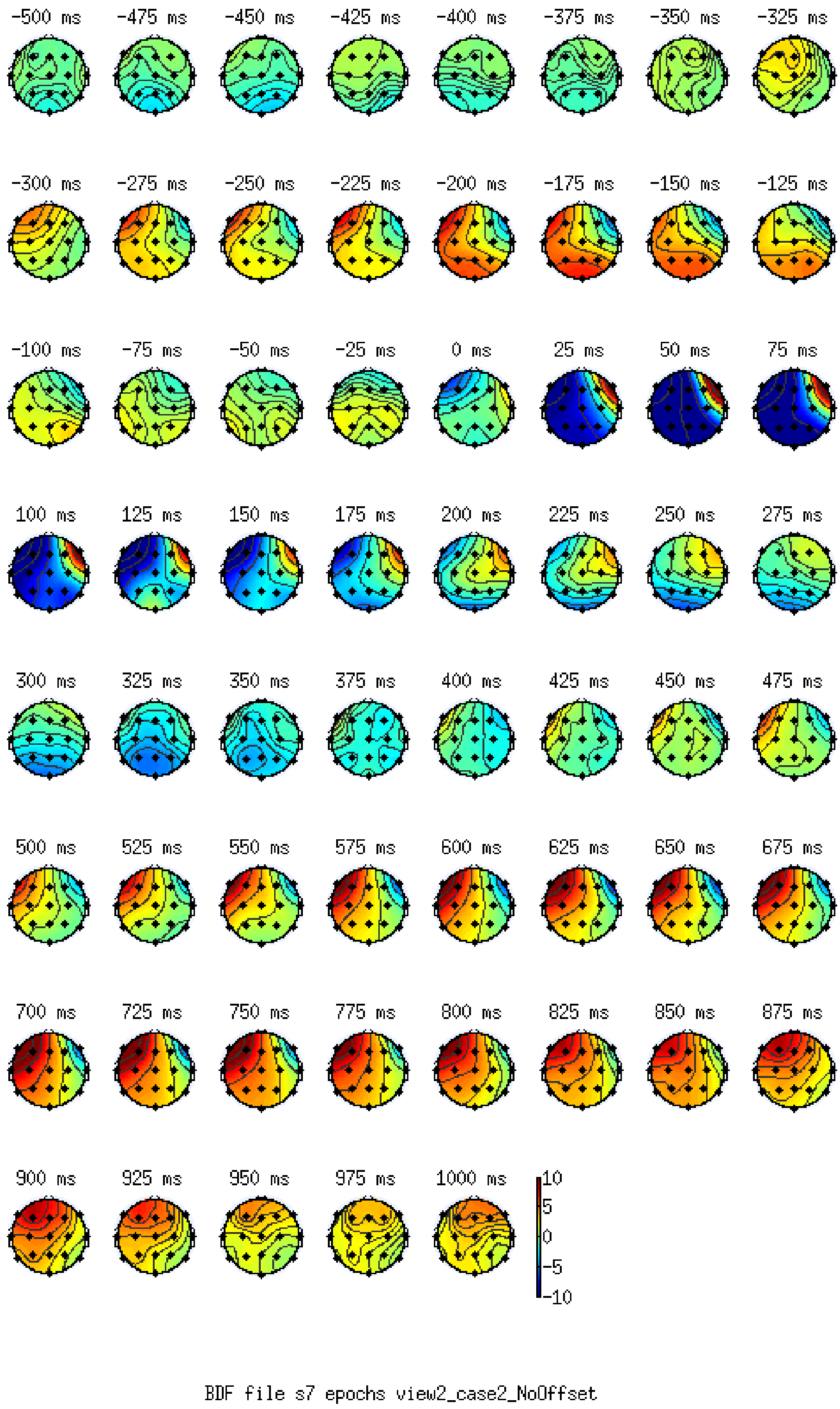

Figure C.36: Subject 2: Averaged scalp plots aligned to fixation offset for non-target frames

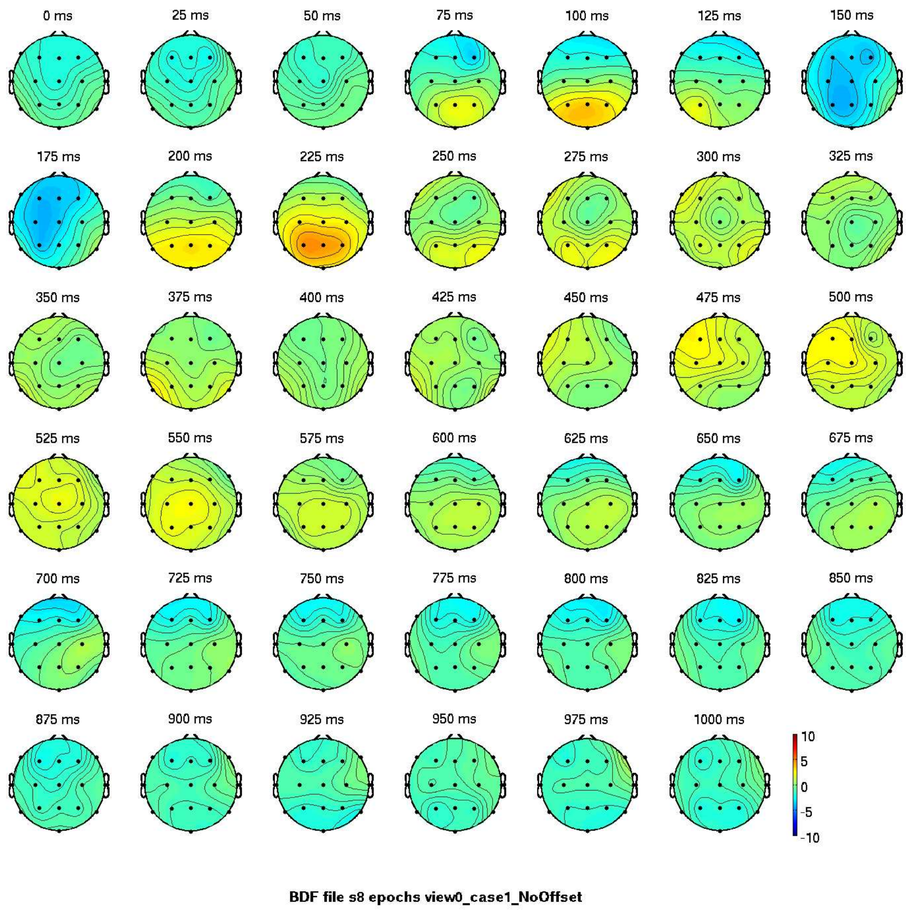

Figure C.37: Subject 3: Averaged scalp plots aligned to frame onset for target frames

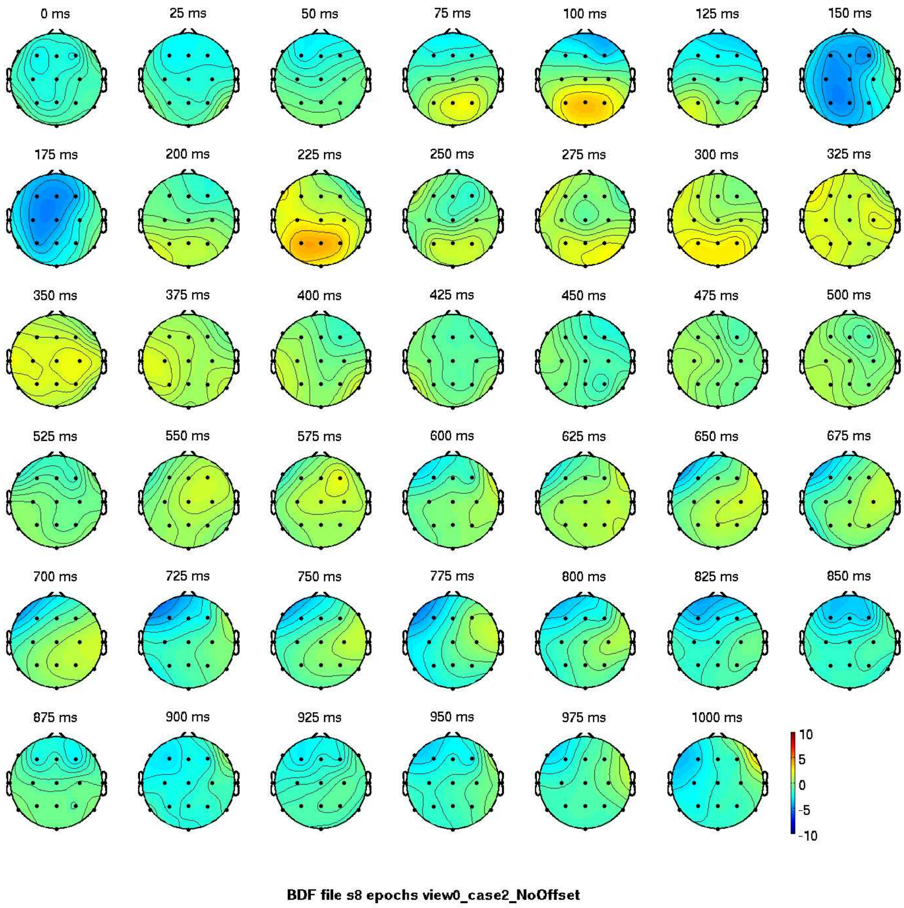

Figure C.38: Subject 3: Averaged scalp plots aligned to frame onset for non-target frames

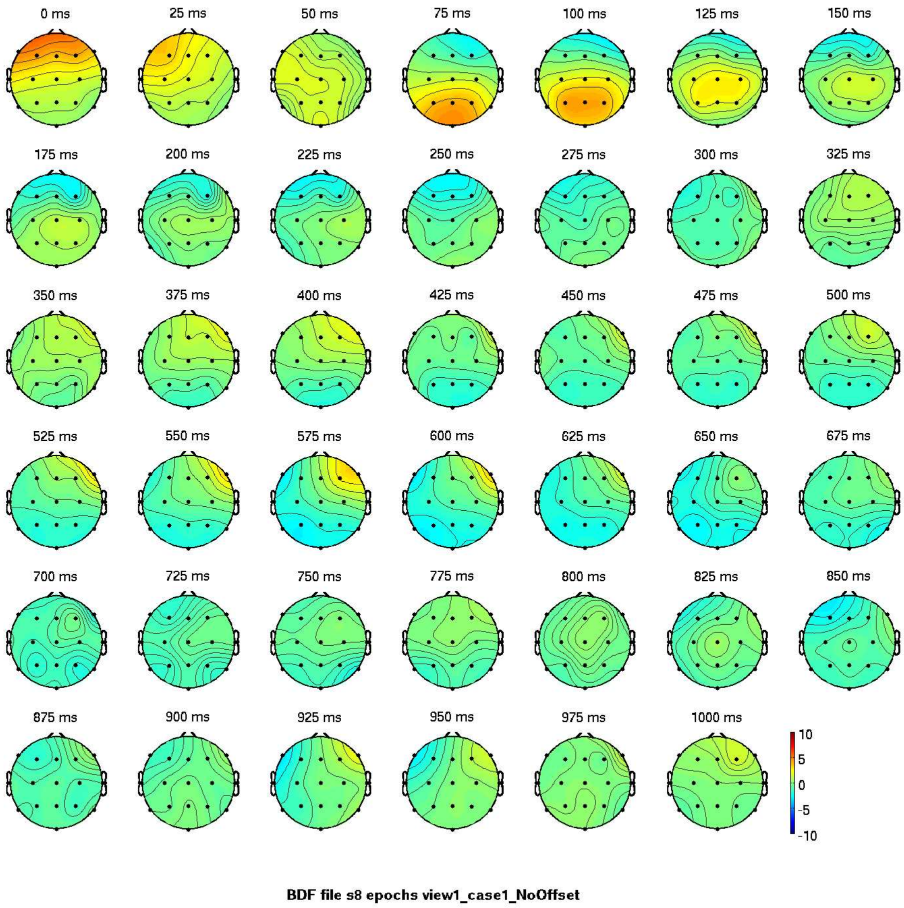

Figure C.39: Subject 3: Averaged scalp plots aligned to fixation onset for target frames

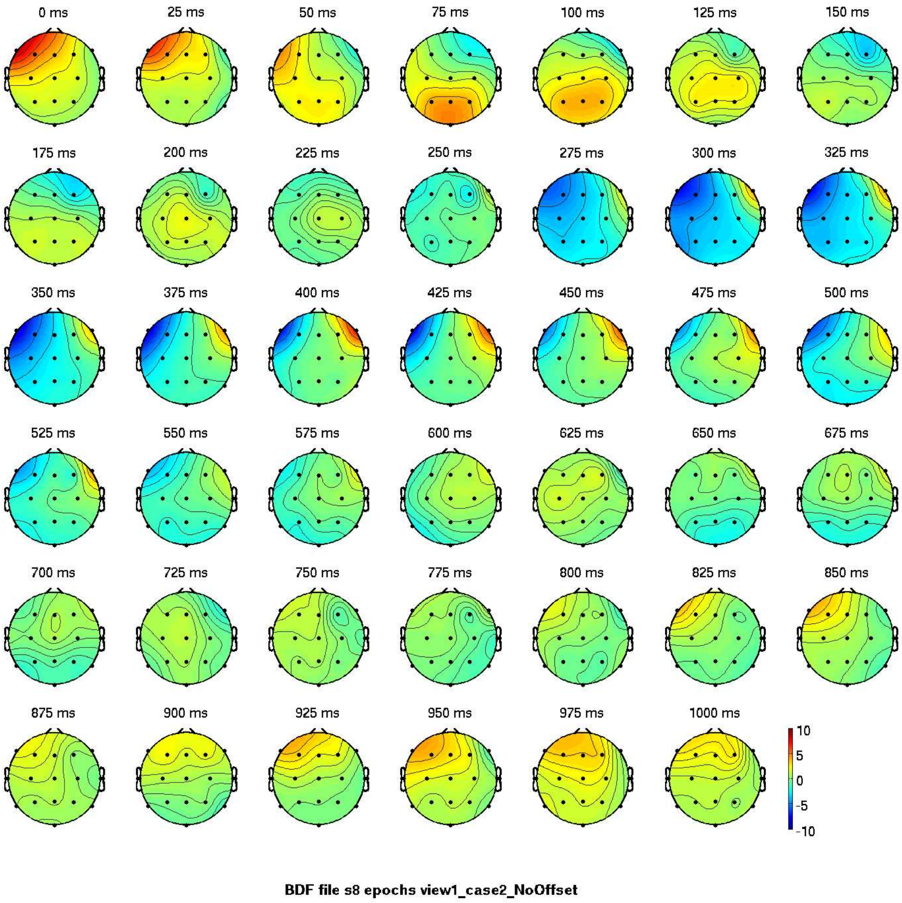

Figure C.40: Subject 3: Averaged scalp plots aligned to fixation onset for non-target frames

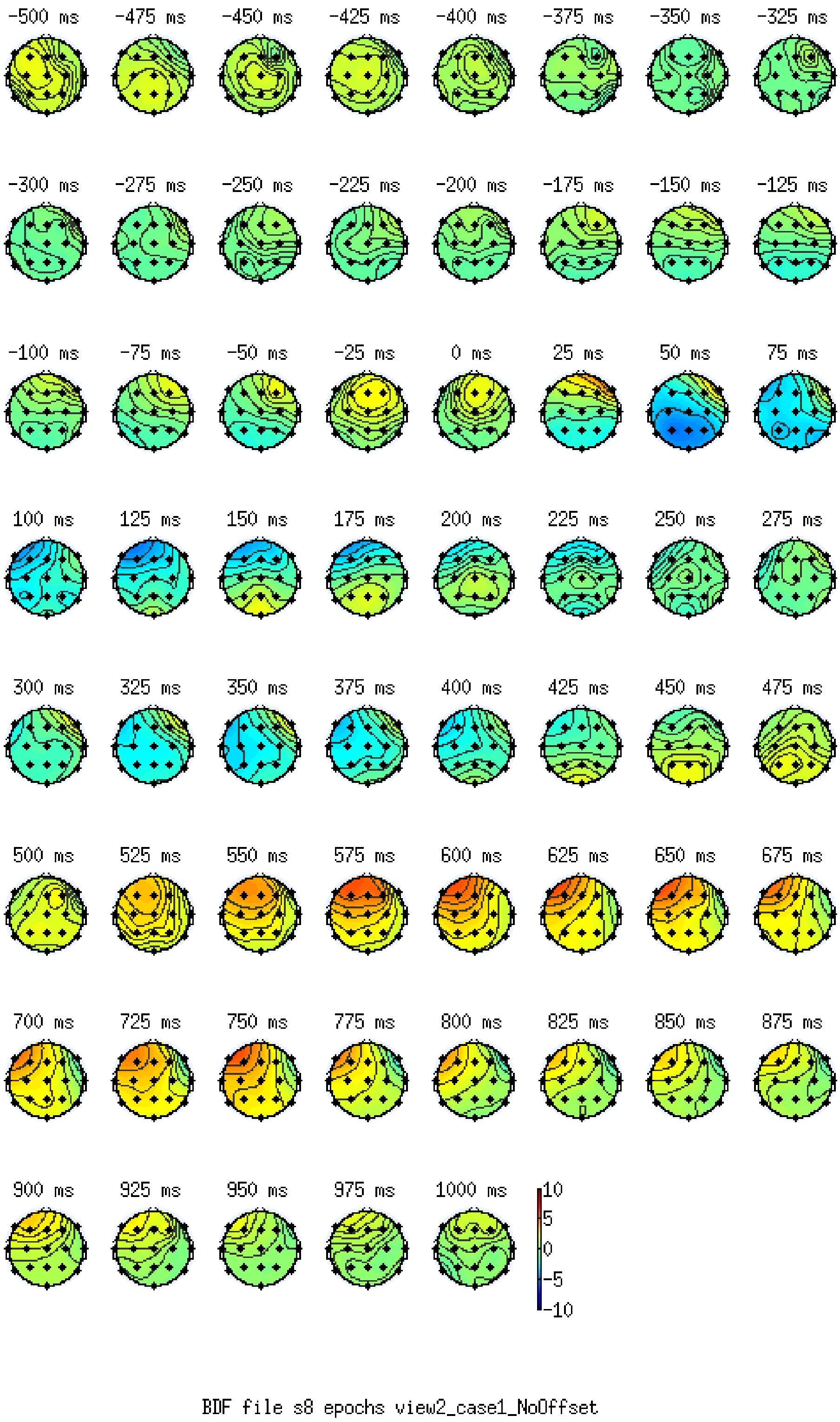

Figure C.41: Subject 3: Averaged scalp plots aligned to fixation offset for target frames

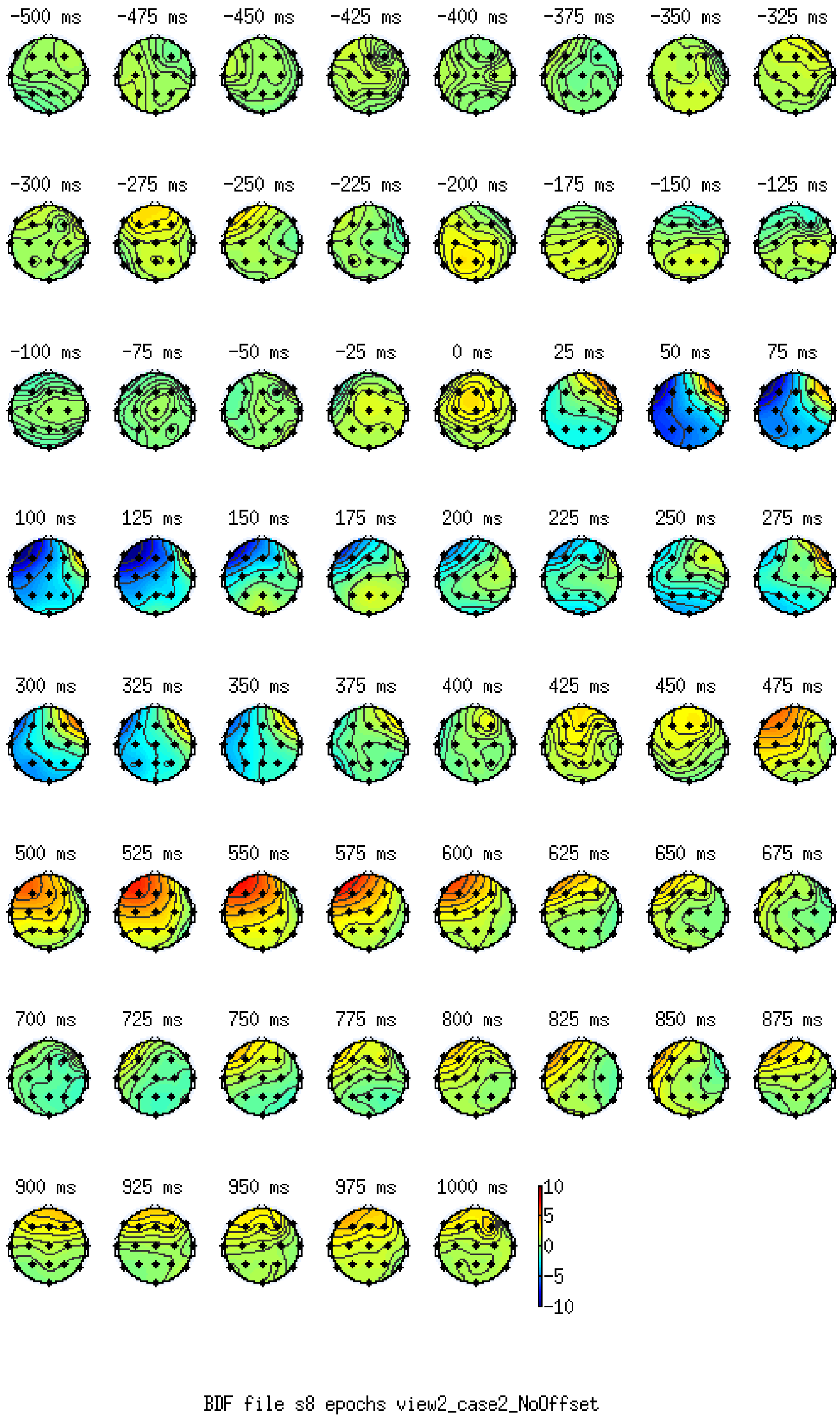

Figure C.42: Subject 3: Averaged scalp plots aligned to fixation offset for non-target frames

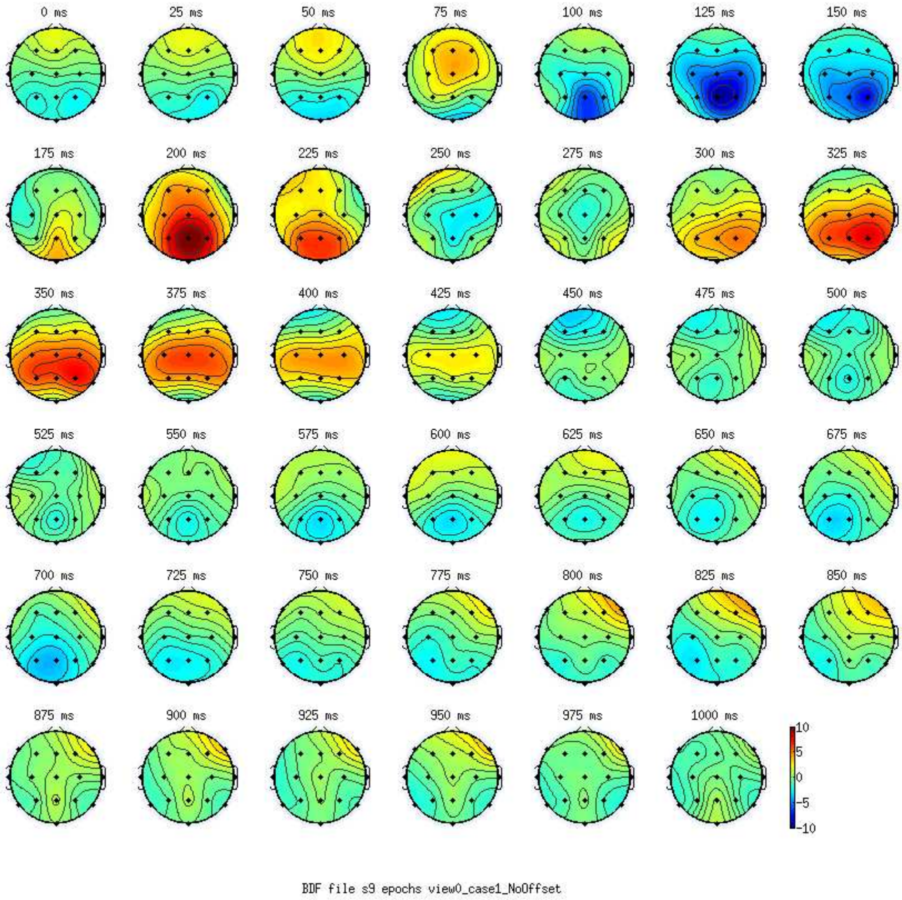

Figure C.43: Subject 4: Averaged scalp plots aligned to frame onset for target frames

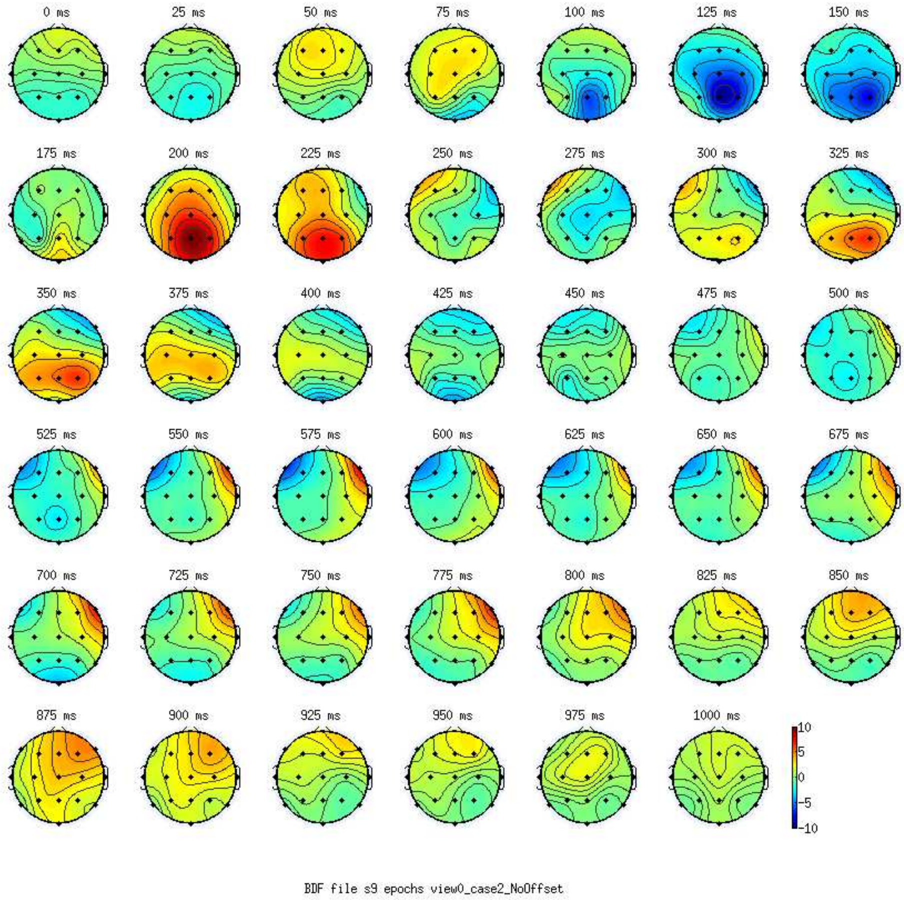

Figure C.44: Subject 4: Averaged scalp plots aligned to frame onset for non-target frames

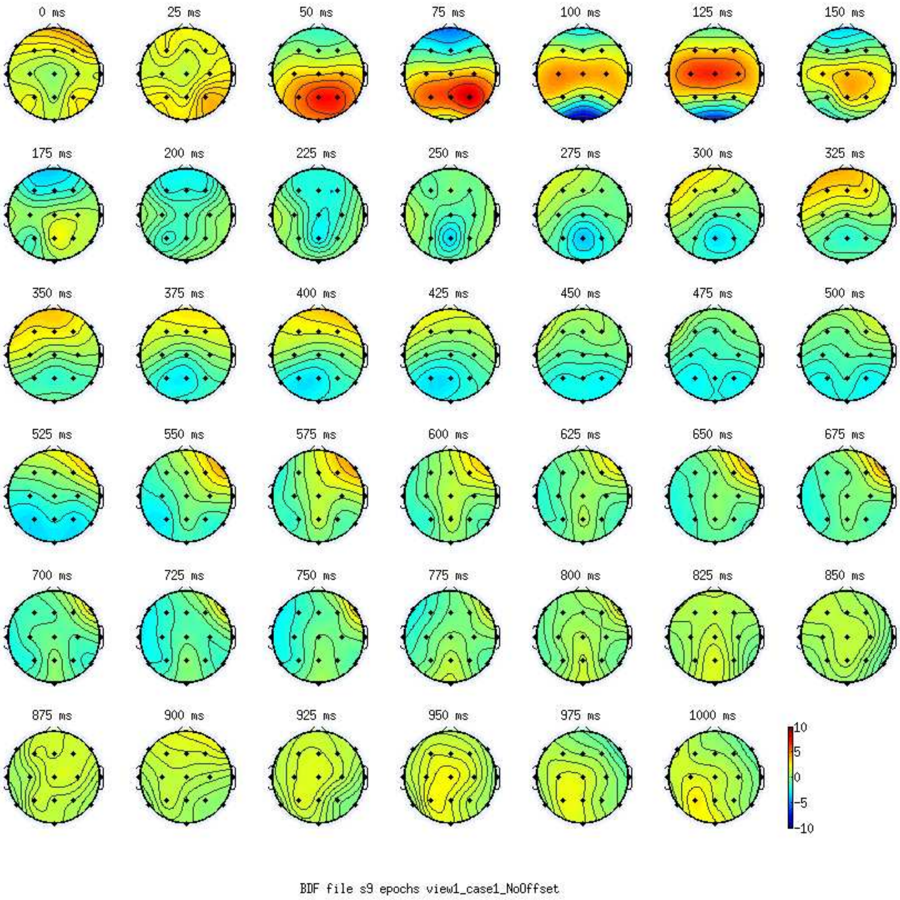

Figure C.45: Subject 4: Averaged scalp plots aligned to fixation onset for target frames

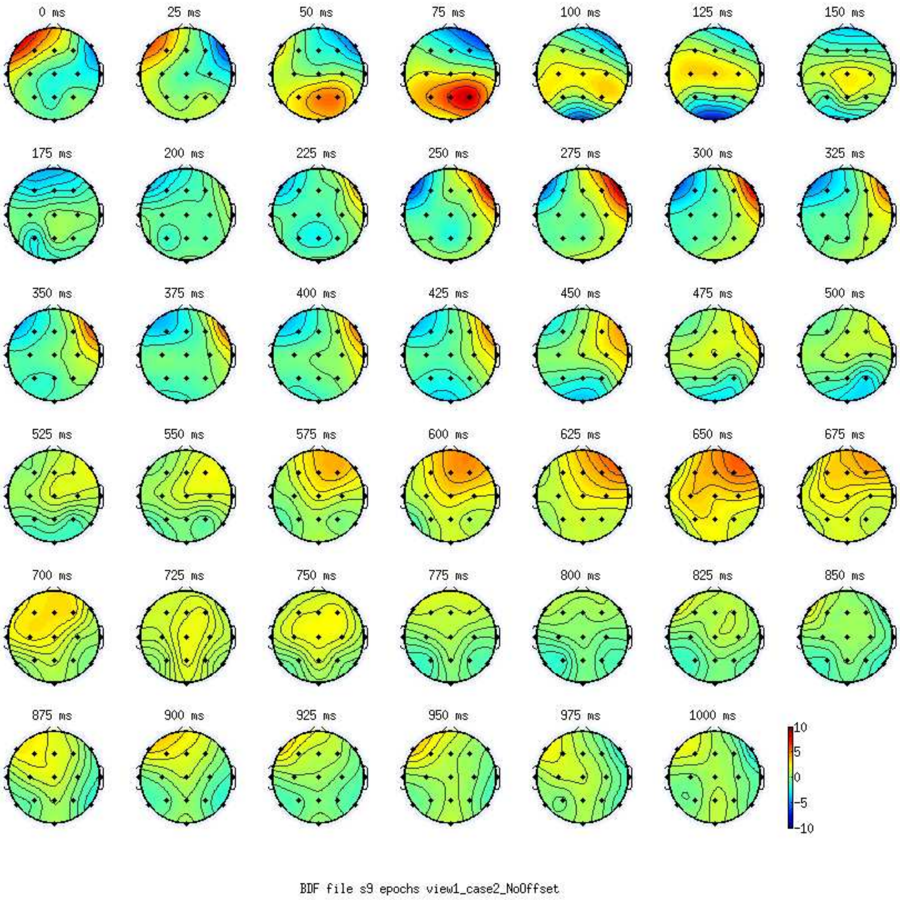

Figure C.46: Subject 4: Averaged scalp plots aligned to fixation onset for non-target frames

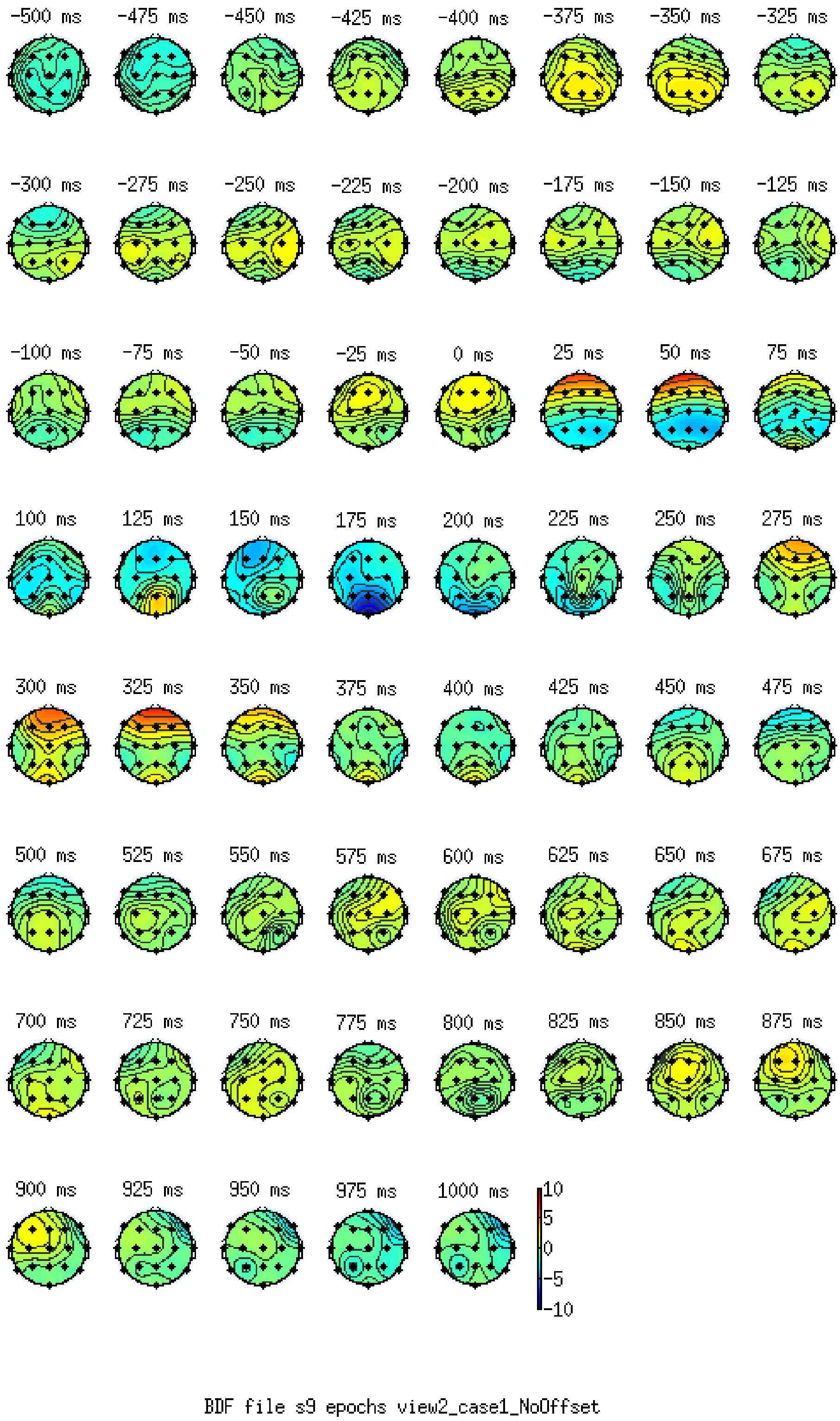

Figure C.47: Subject 4: Averaged scalp plots aligned to fixation offset for target frames

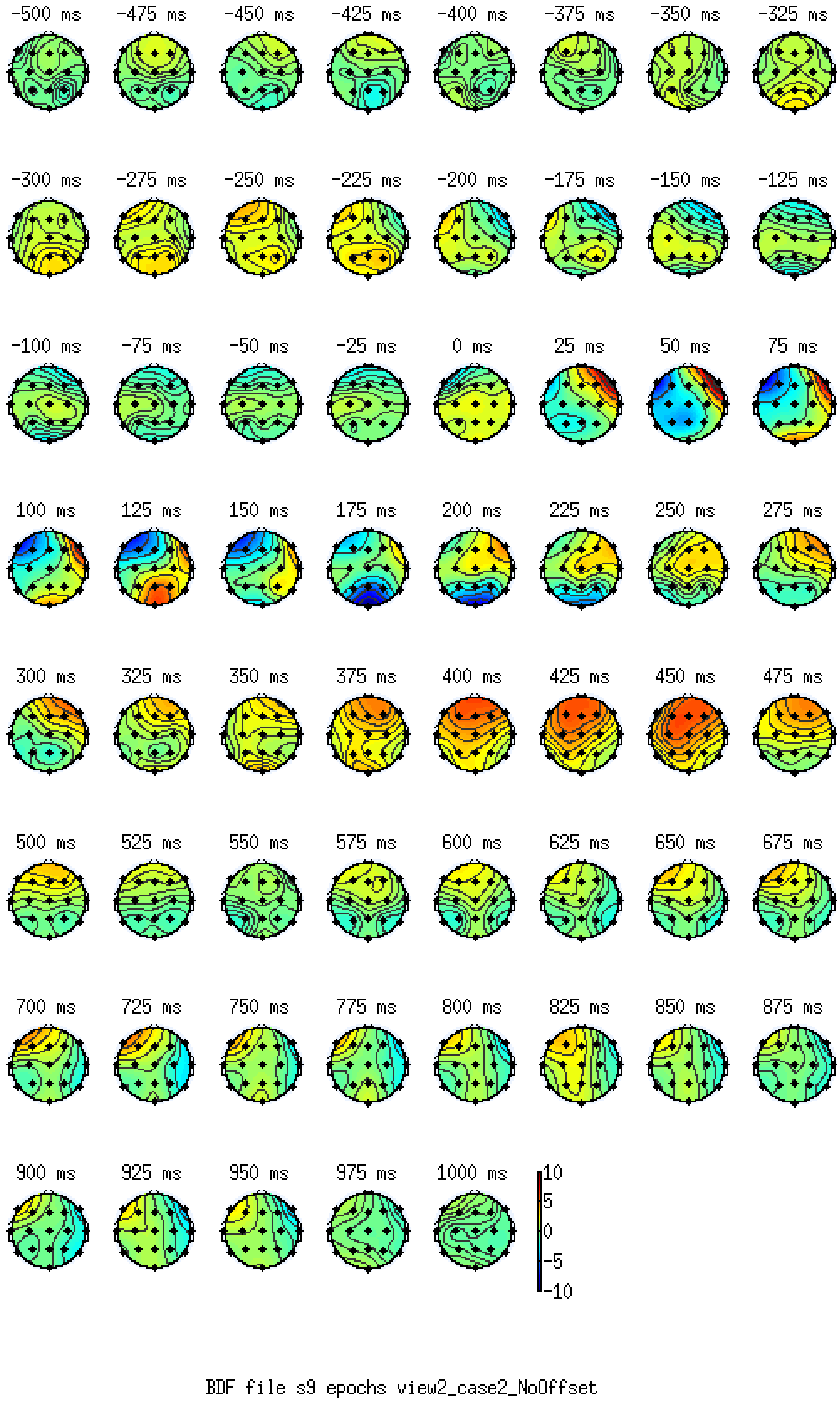

Figure C.48: Subject 4: Averaged scalp plots aligned to fixation offset for non-target frames

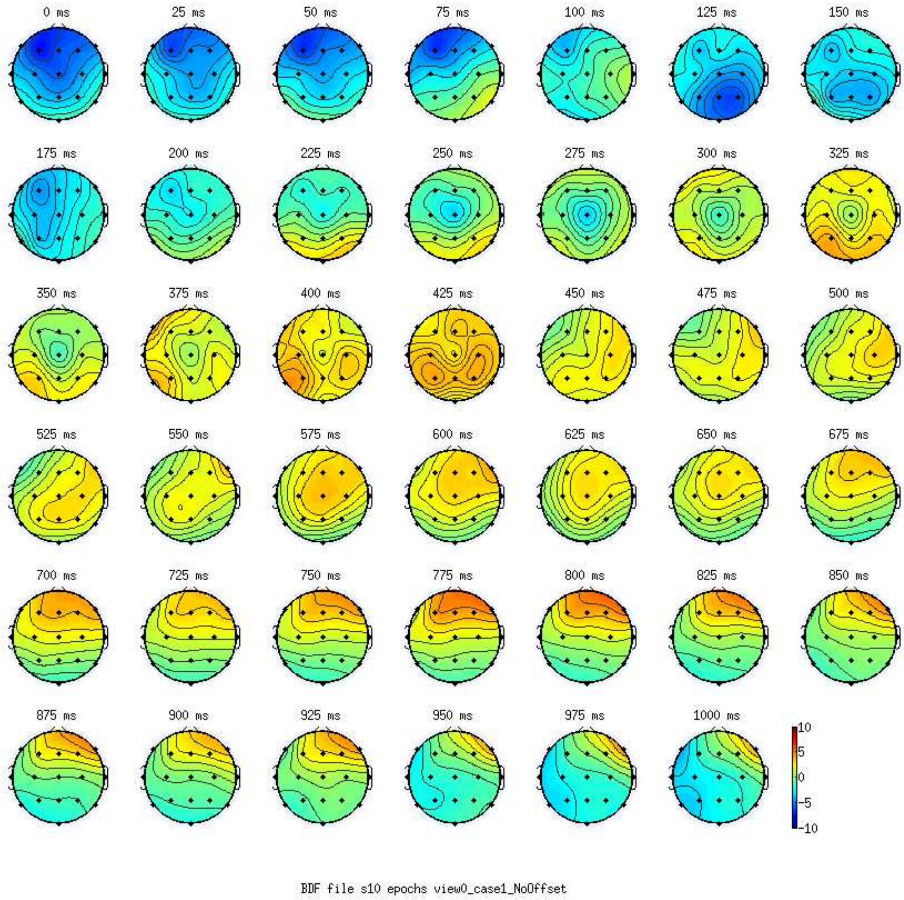

Figure C.49: Subject 5: Averaged scalp plots aligned to frame onset for target frames

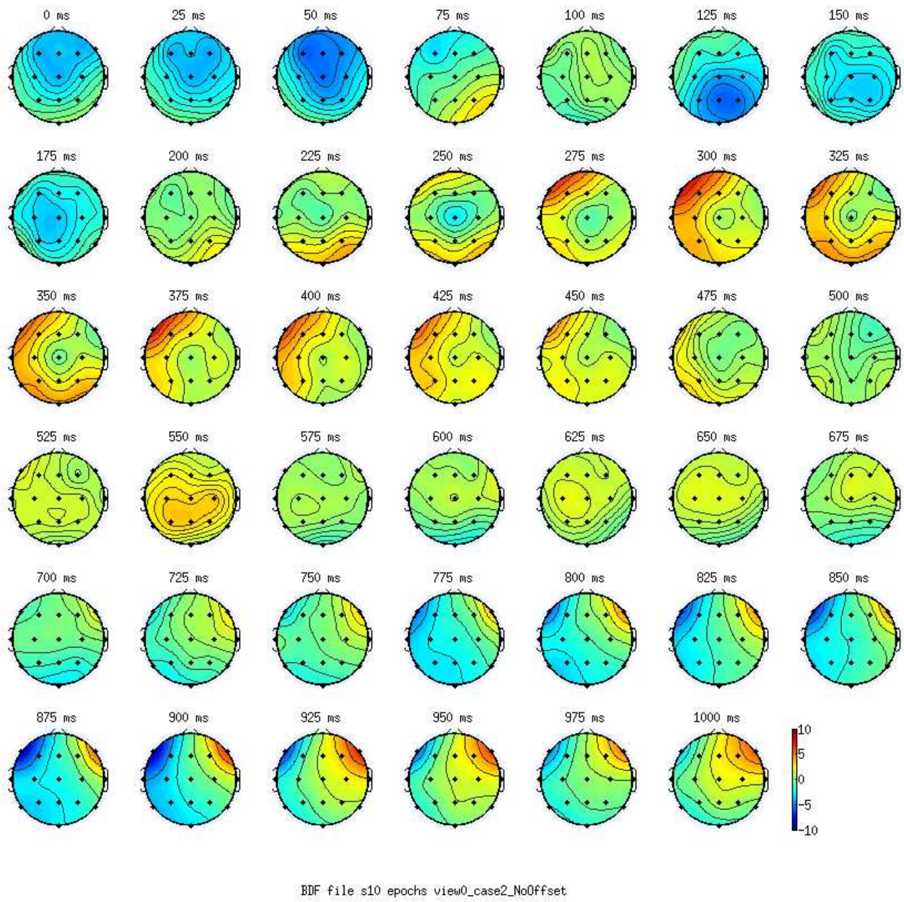

Figure C.50: Subject 5: Averaged scalp plots aligned to frame onset for non-target frames

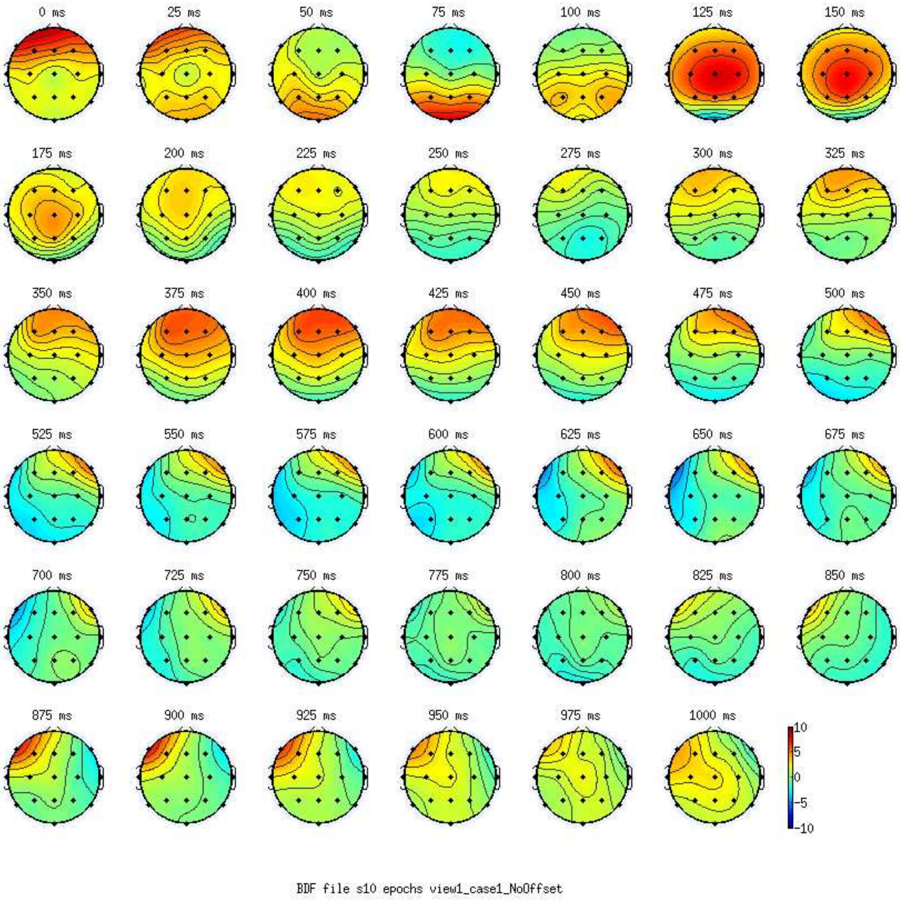

Figure C.51: Subject 5: Averaged scalp plots aligned to fixation onset for target frames

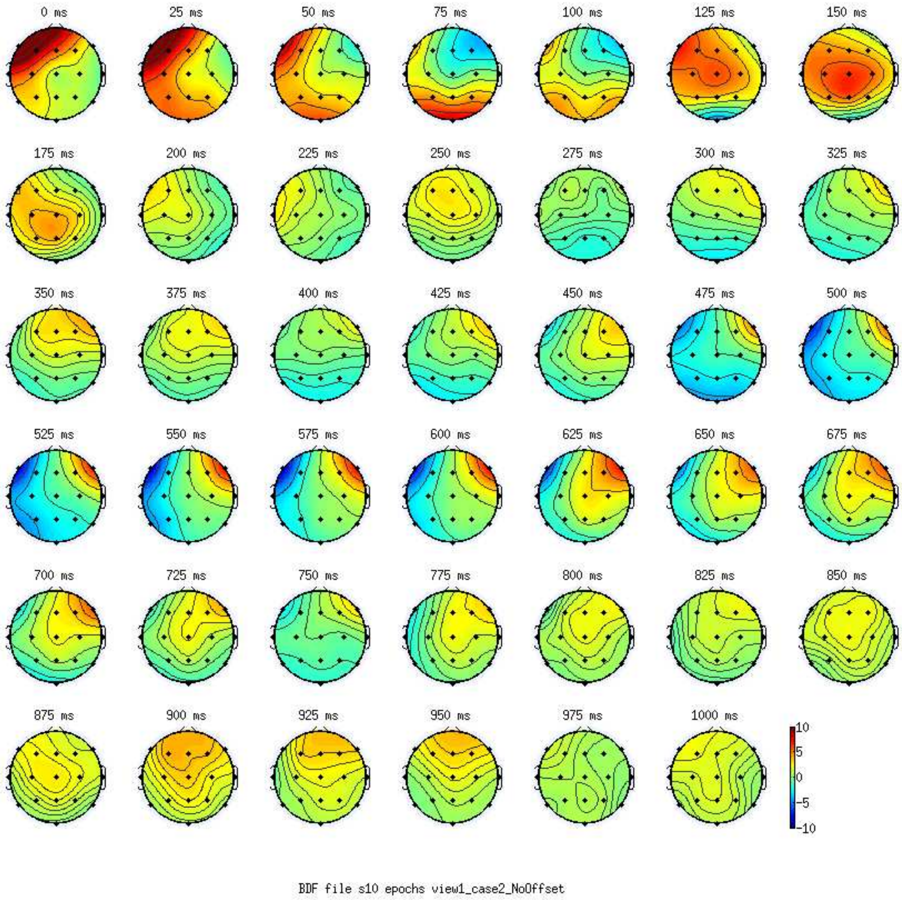

Figure C.52: Subject 5: Averaged scalp plots aligned to fixation onset for non-target frames

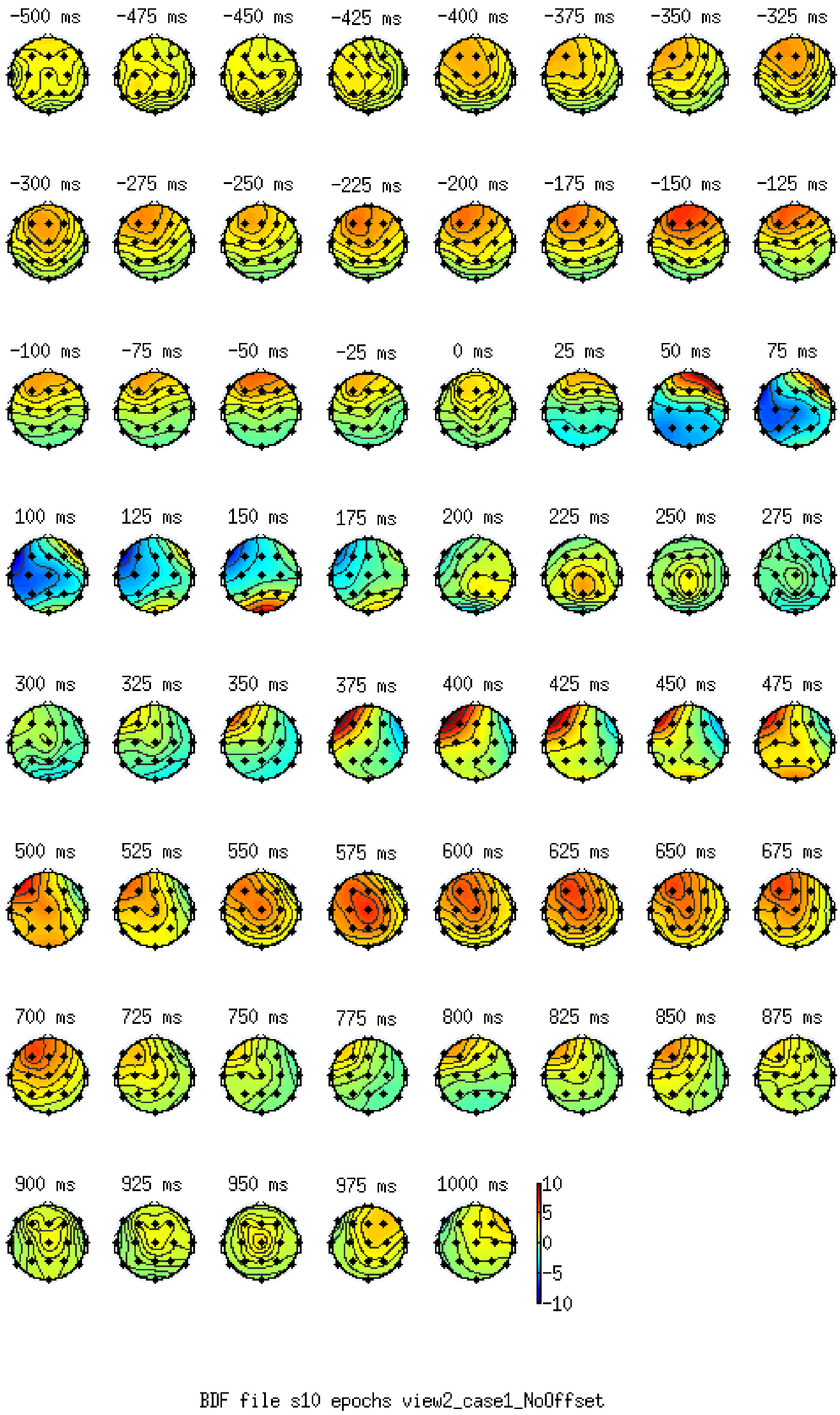

Figure C.53: Subject 5: Averaged scalp plots aligned to fixation offset for target frames

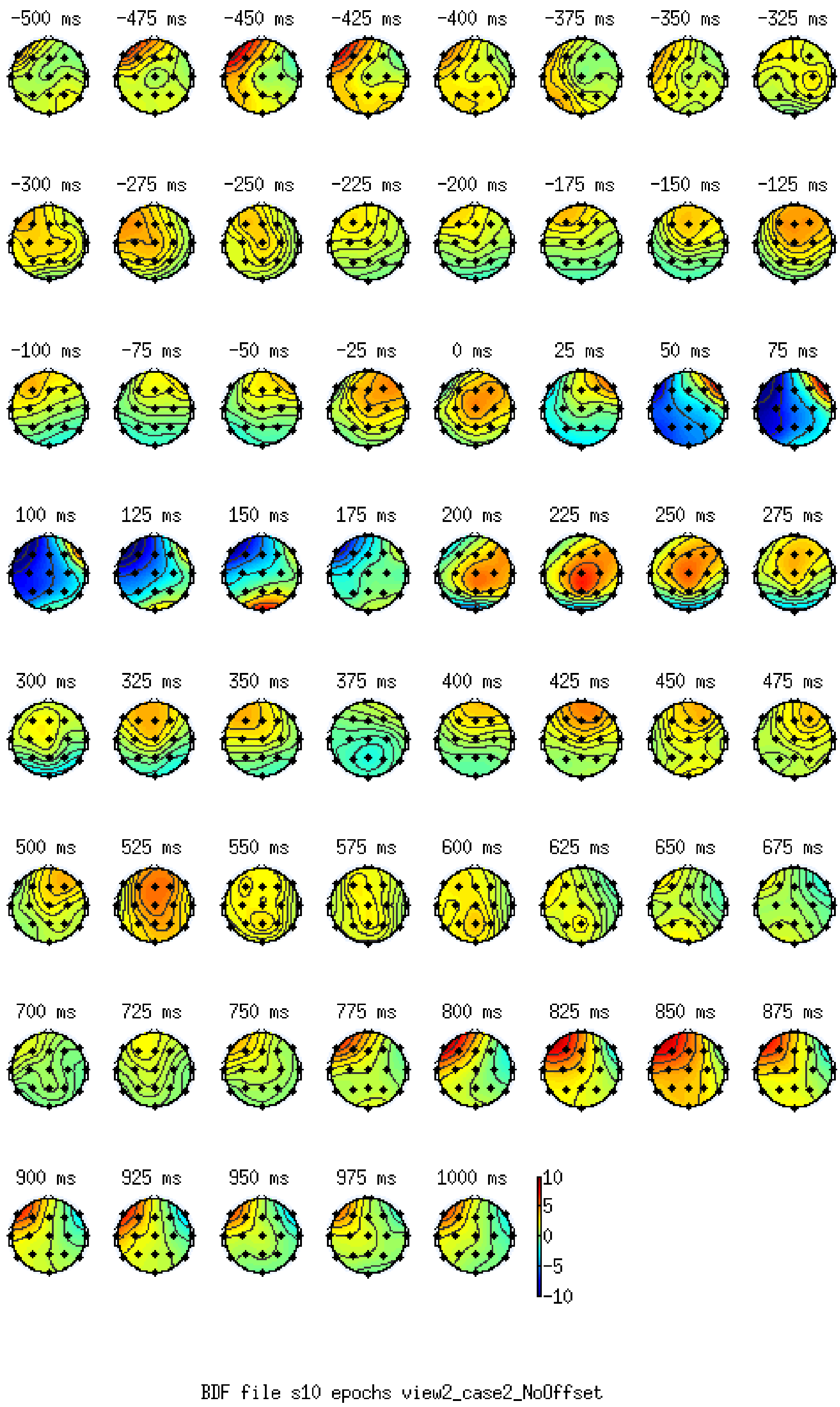

Figure C.54: Subject 5: Averaged scalp plots aligned to fixation offset for non-target frames

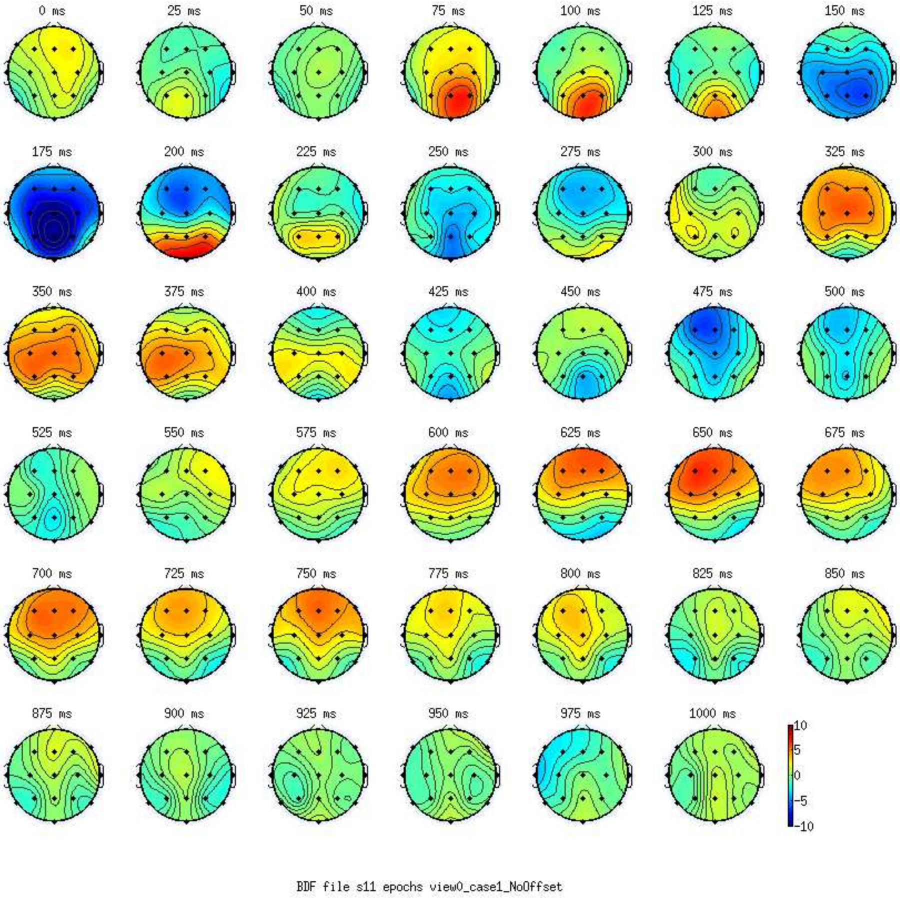

Figure C.55: Subject 6: Averaged scalp plots aligned to frame onset for target frames

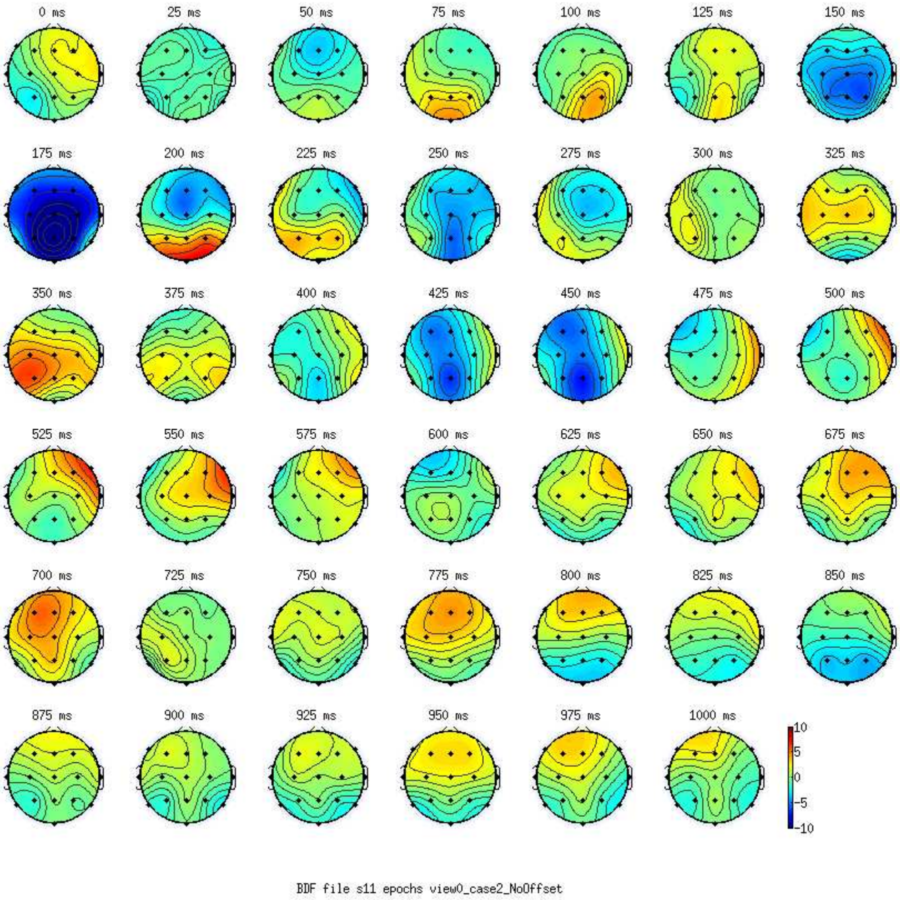

Figure C.56: Subject 6: Averaged scalp plots aligned to frame onset for non-target frames

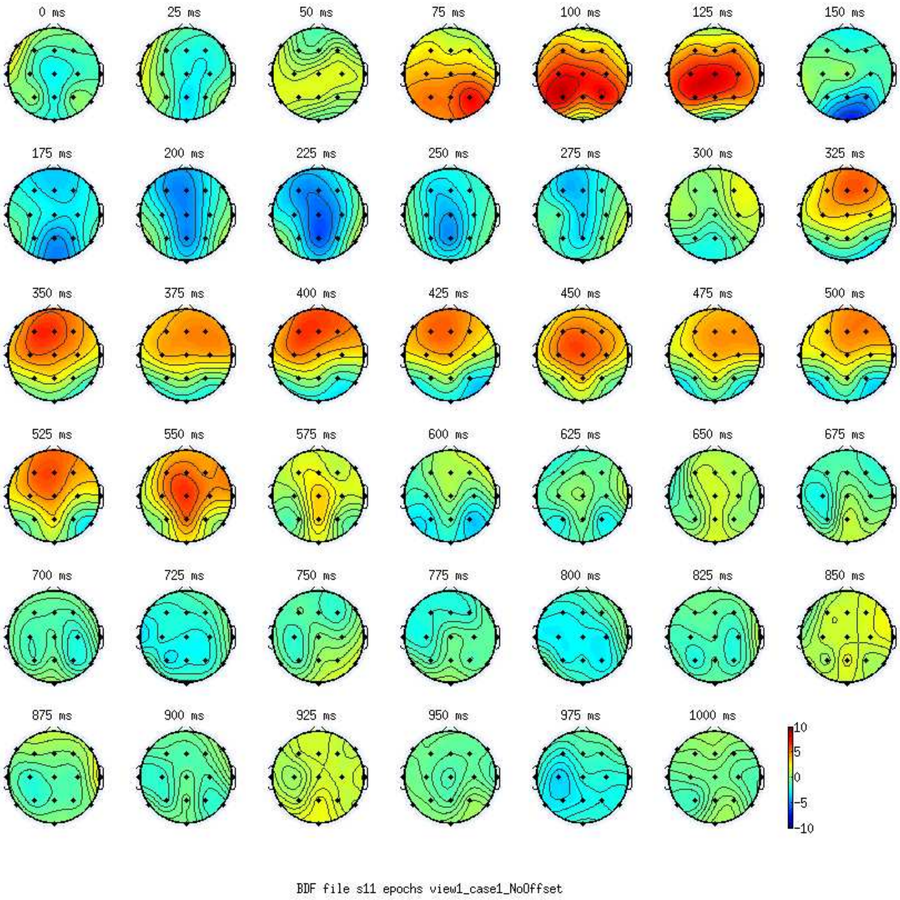

Figure C.57: Subject 6: Averaged scalp plots aligned to fixation onset for target frames

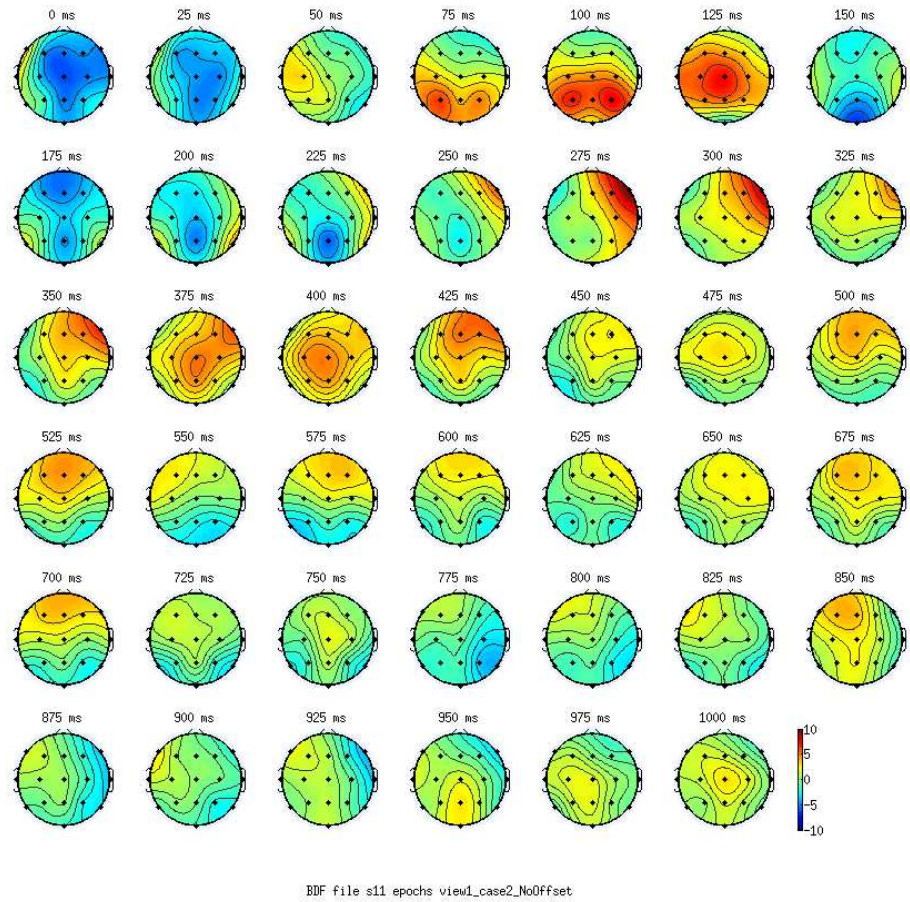

Figure C.58: Subject 6: Averaged scalp plots aligned to fixation onset for non-target frames

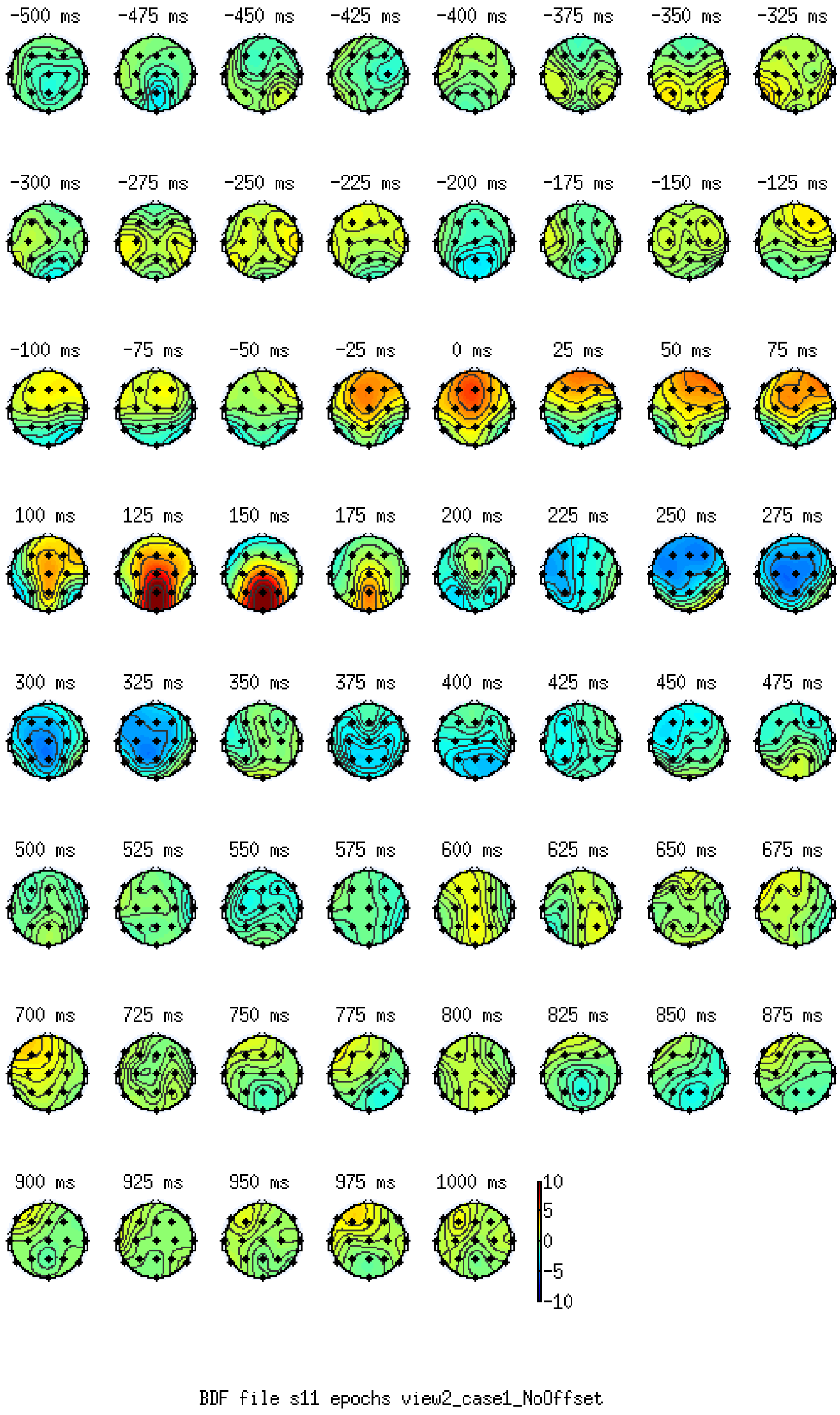

Figure C.59: Subject 6: Averaged scalp plots aligned to fixation offset for target frames

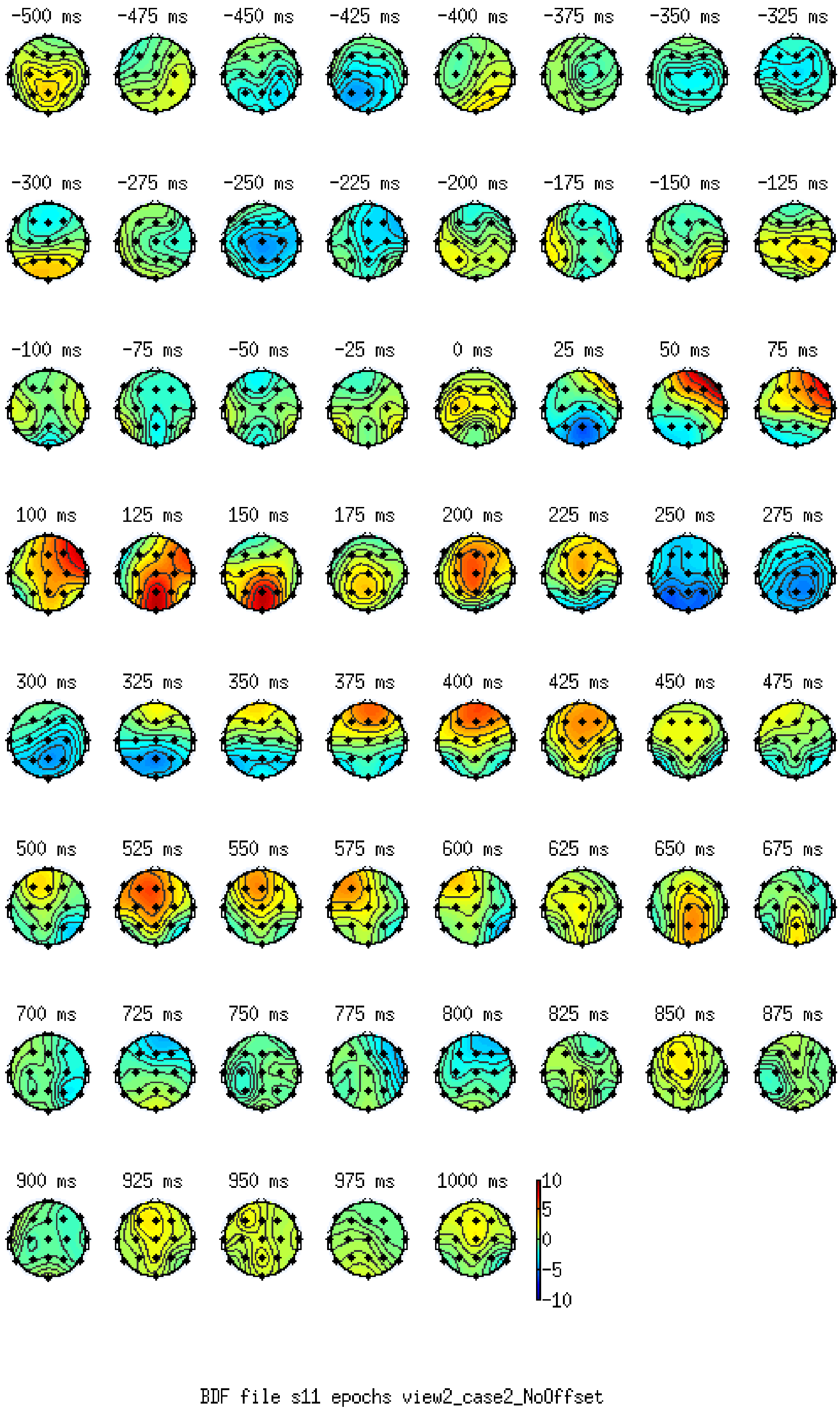

Figure C.60: Subject 6: Averaged scalp plots aligned to fixation offset for non-target frames

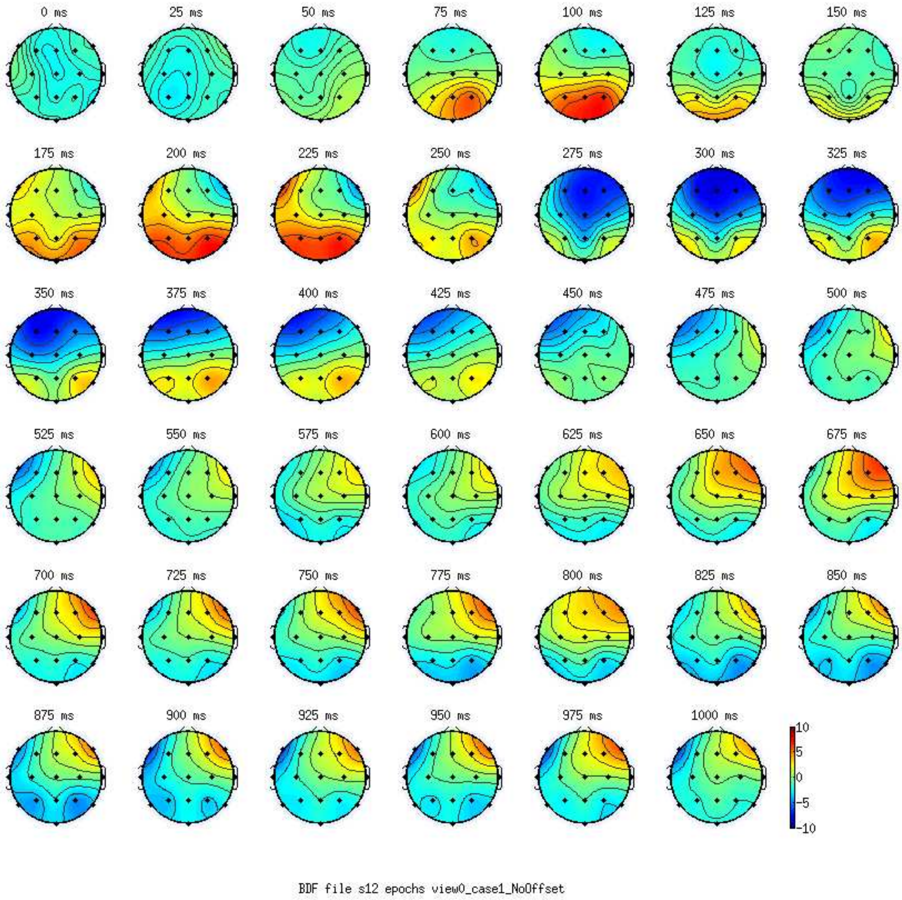

Figure C.61: Subject 7: Averaged scalp plots aligned to frame onset for target frames

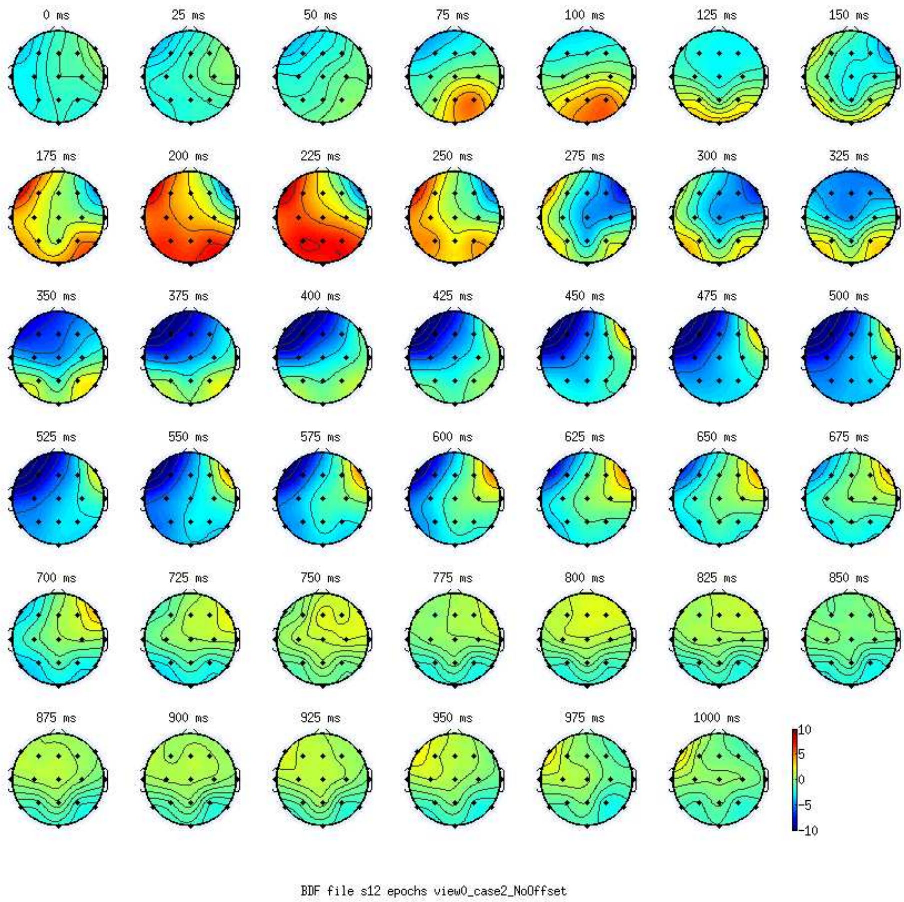

Figure C.62: Subject 7: Averaged scalp plots aligned to frame onset for non-target frames

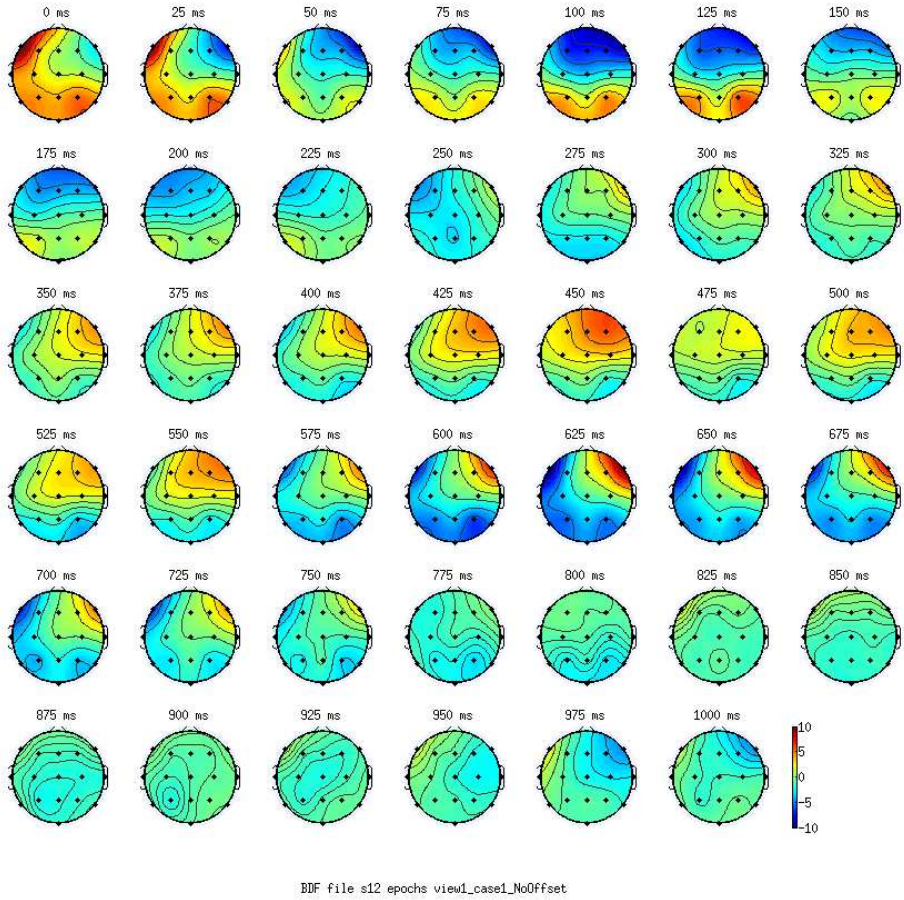

Figure C.63: Subject 7: Averaged scalp plots aligned to fixation onset for target frames

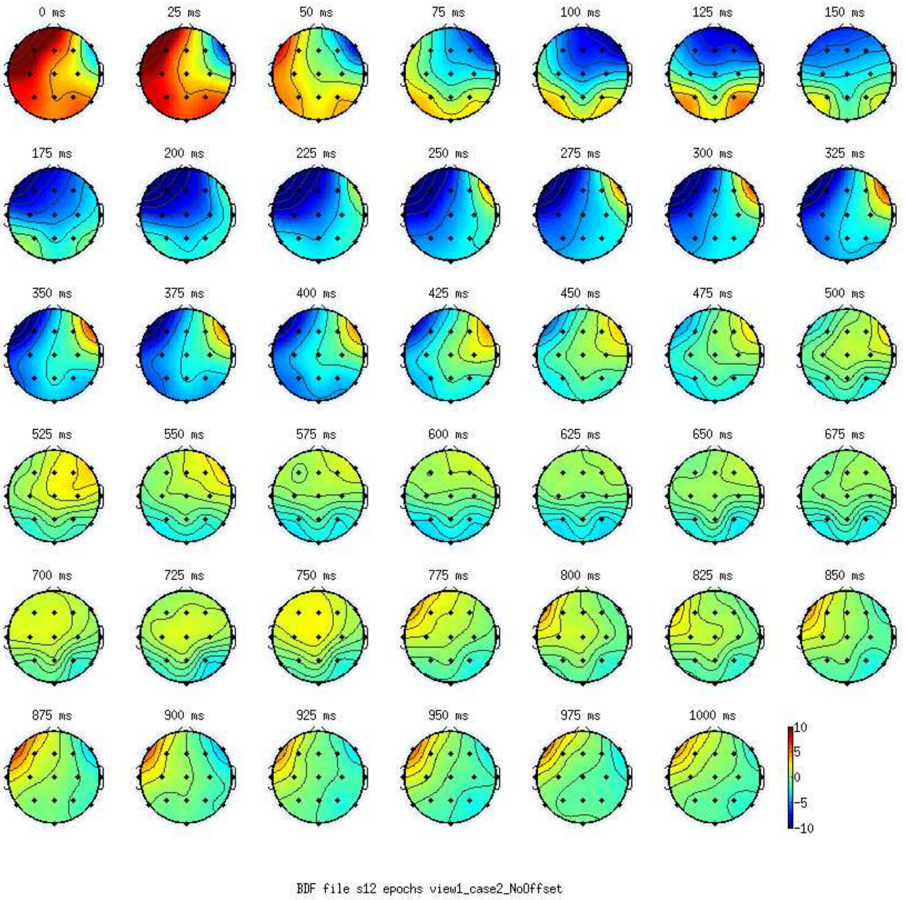

Figure C.64: Subject 7: Averaged scalp plots aligned to fixation onset for non-target frames

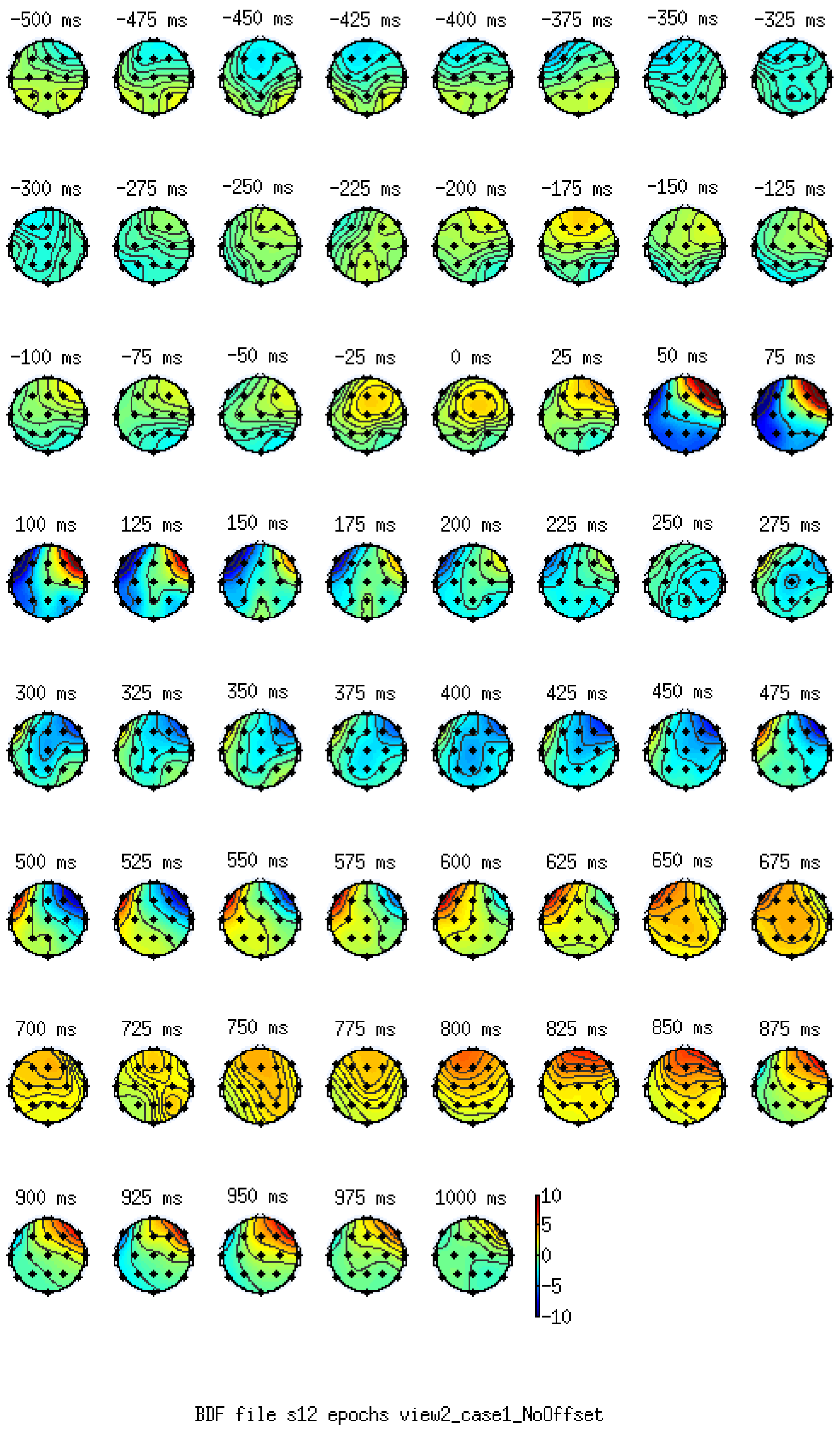

Figure C.65: Subject 7: Averaged scalp plots aligned to fixation offset for target frames

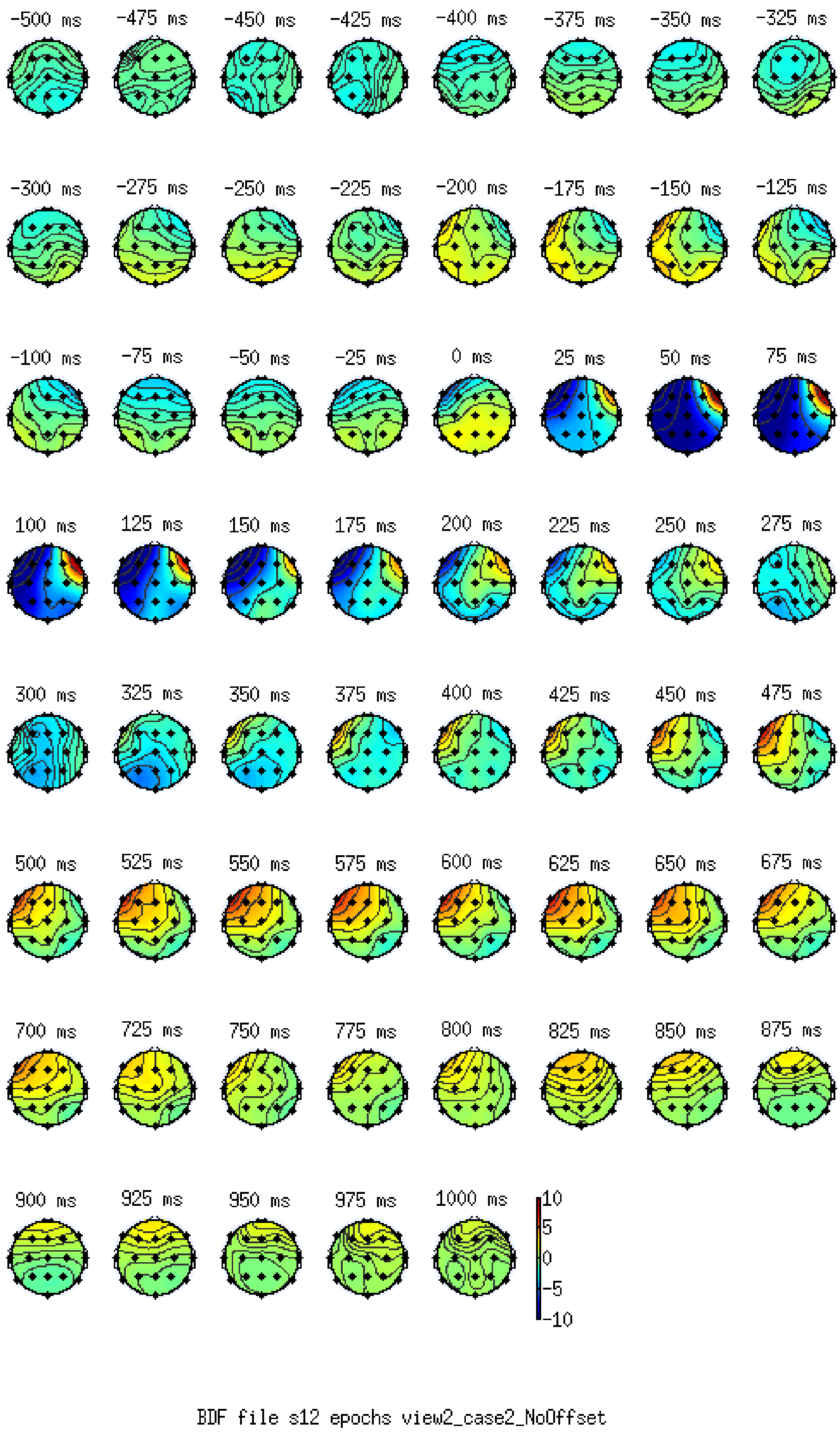

Figure C.66: Subject 7: Averaged scalp plots aligned to fixation offset for non-target frames

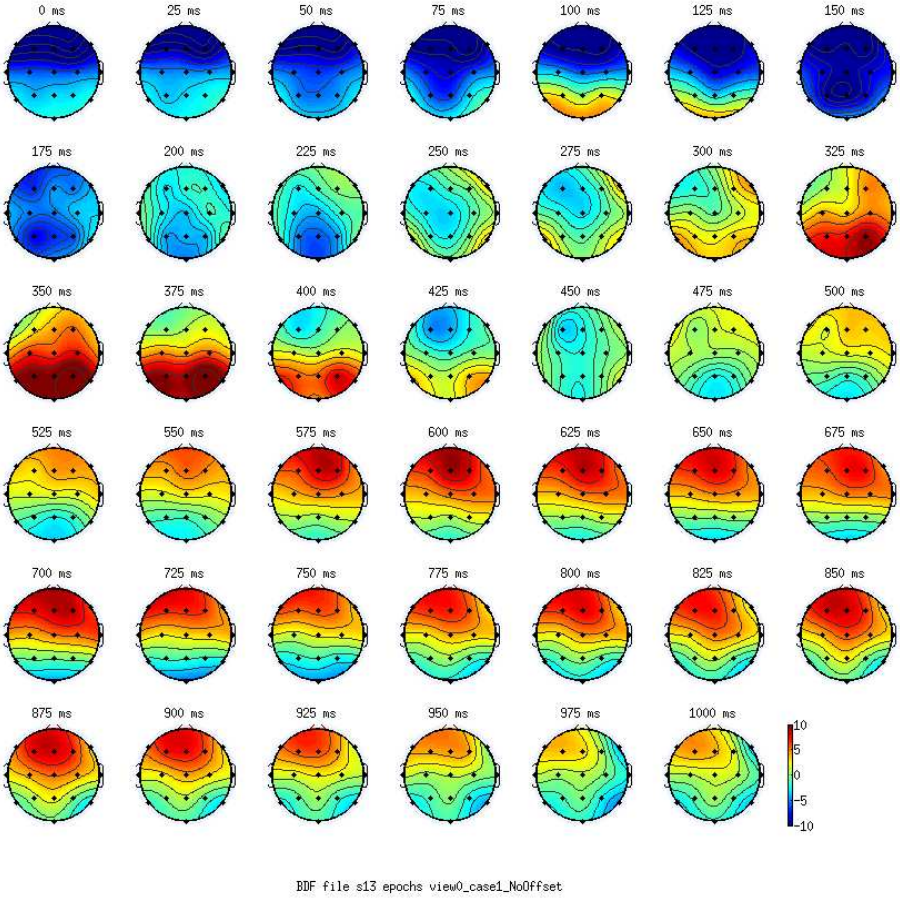

Figure C.67: Subject 8: Averaged scalp plots aligned to frame onset for target frames

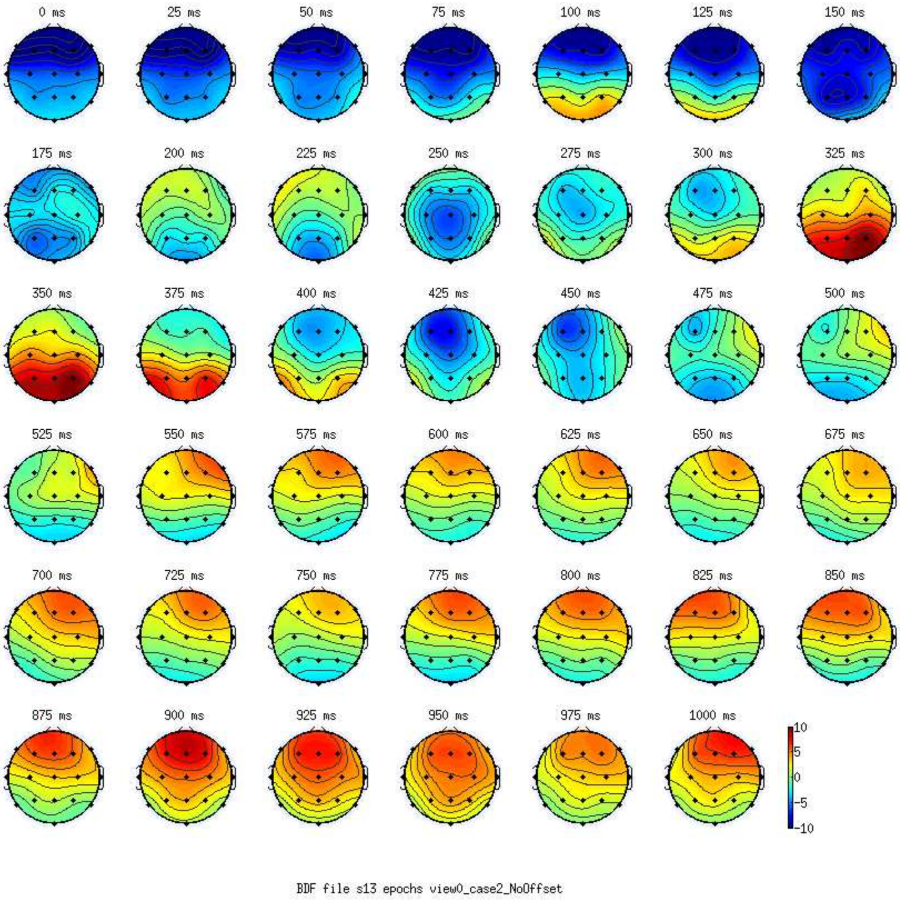

Figure C.68: Subject 8: Averaged scalp plots aligned to frame onset for non-target frames

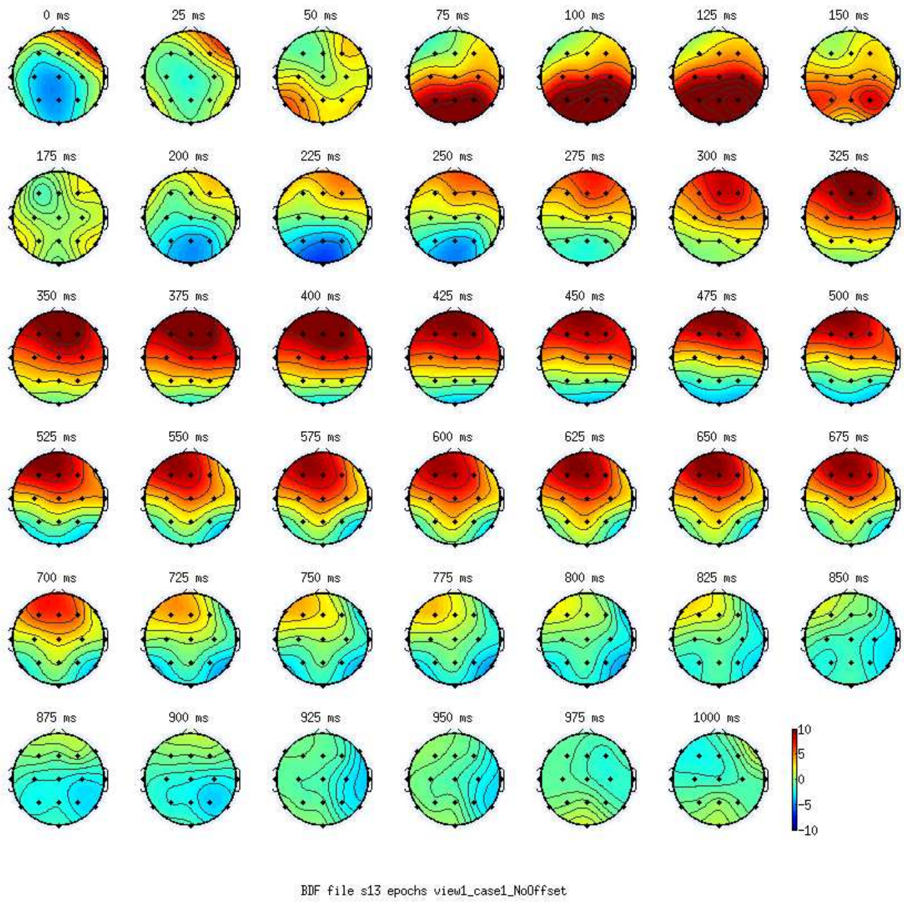

Figure C.69: Subject 8: Averaged scalp plots aligned to fixation onset for target frames

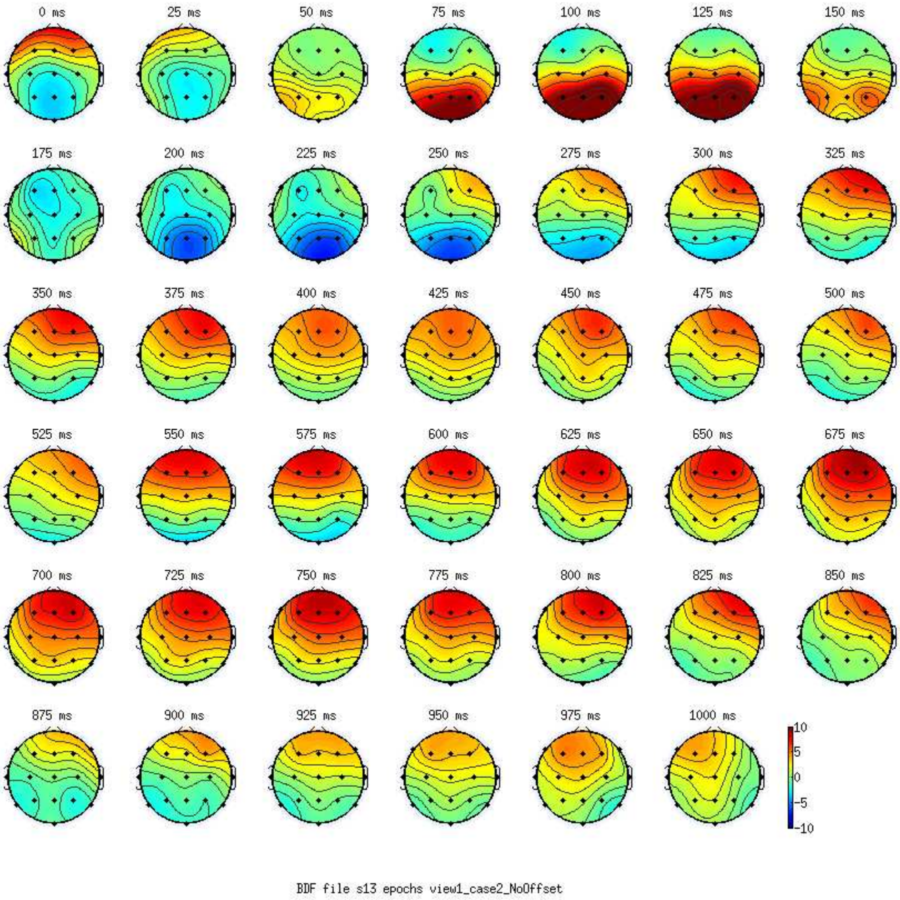

Figure C.70: Subject 8: Averaged scalp plots aligned to fixation onset for non-target frames

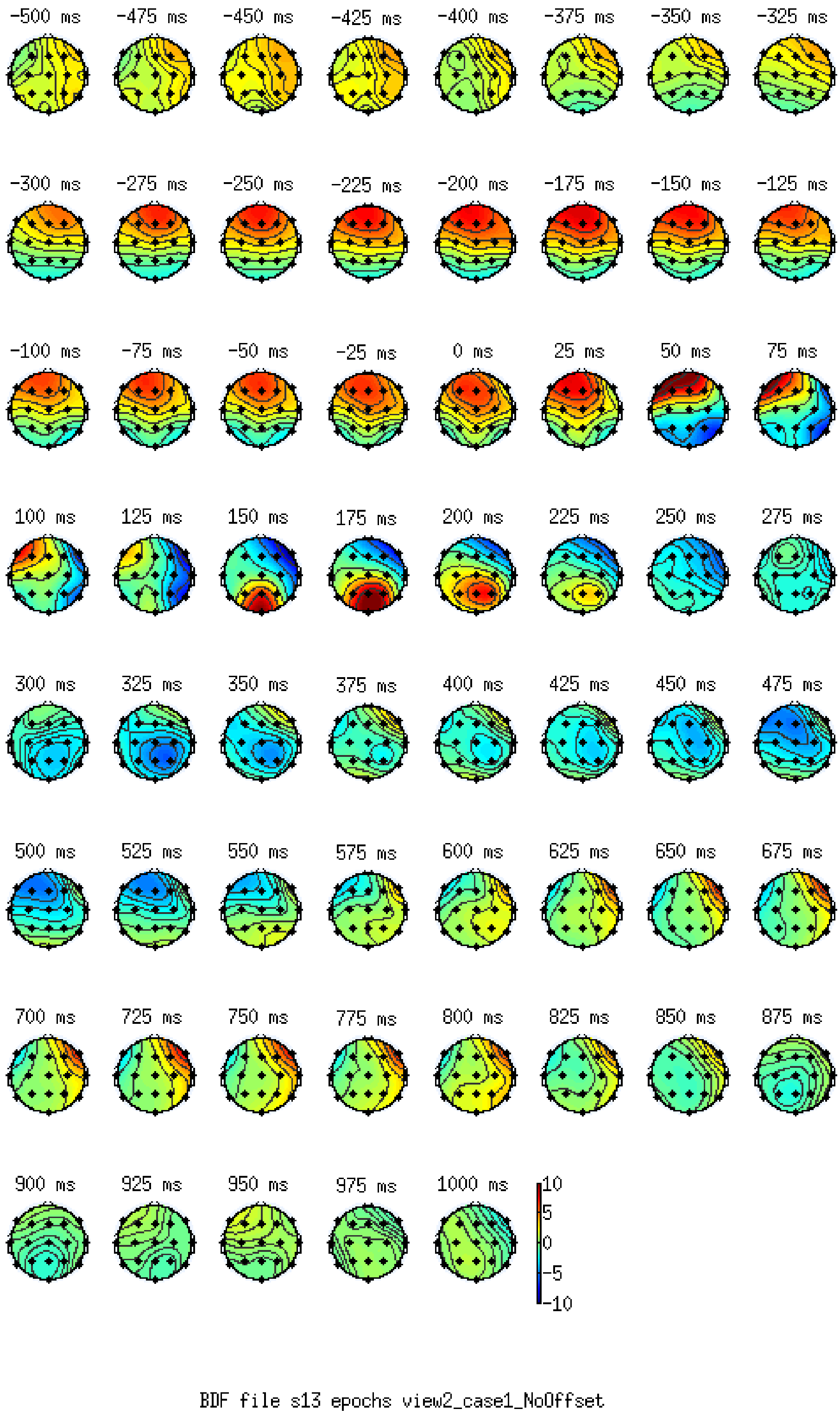

Figure C.71: Subject 8: Averaged scalp plots aligned to fixation offset for target frames

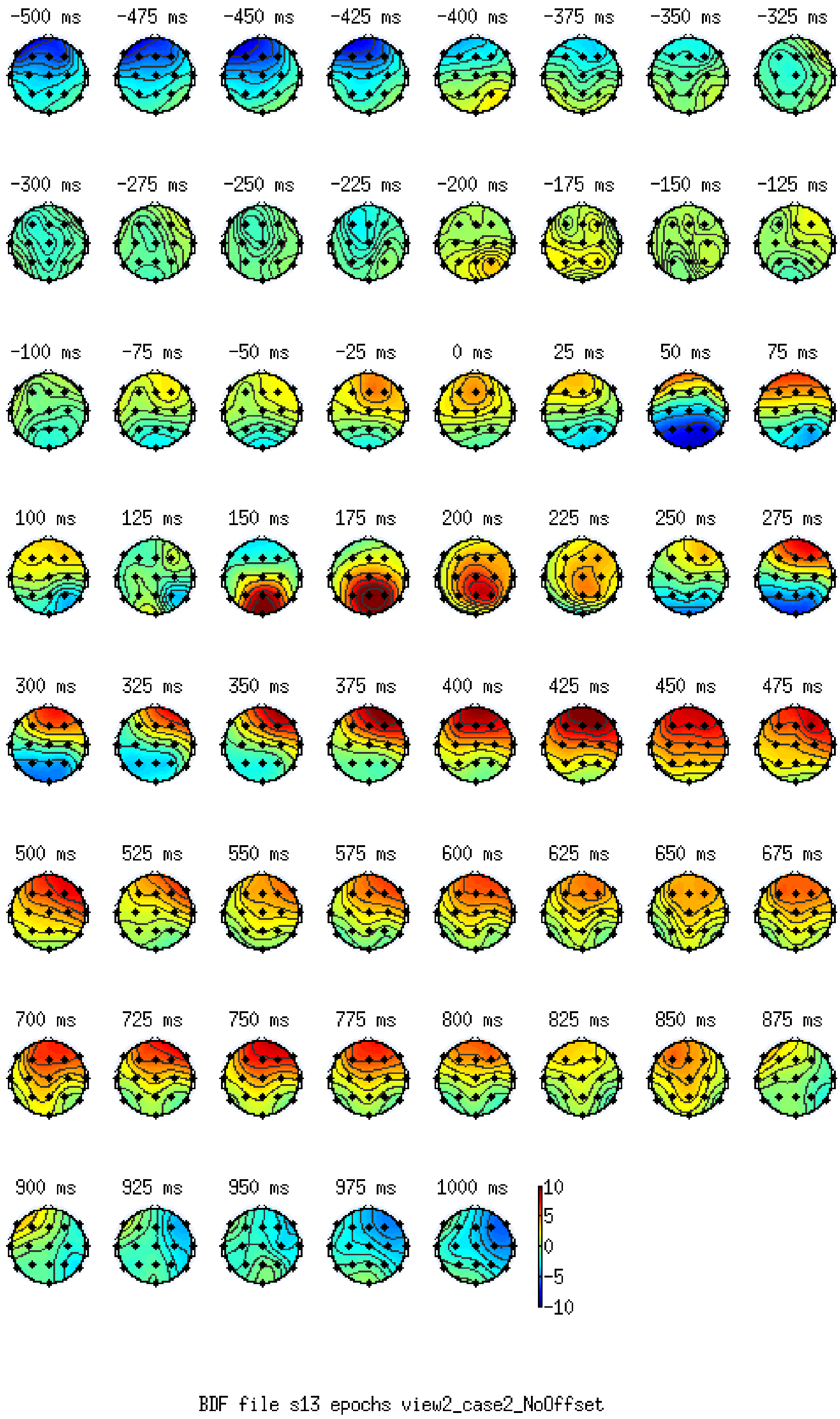

Figure C.72: Subject 8: Averaged scalp plots aligned to fixation offset for non-target frames

C.3 Temporal Discrimination Plots for Experiment 3

In this section we provide the temporal discrimination plots obtained for experiment 3 across all subjects. This materials are intended to primarily supplement Section 3.4.

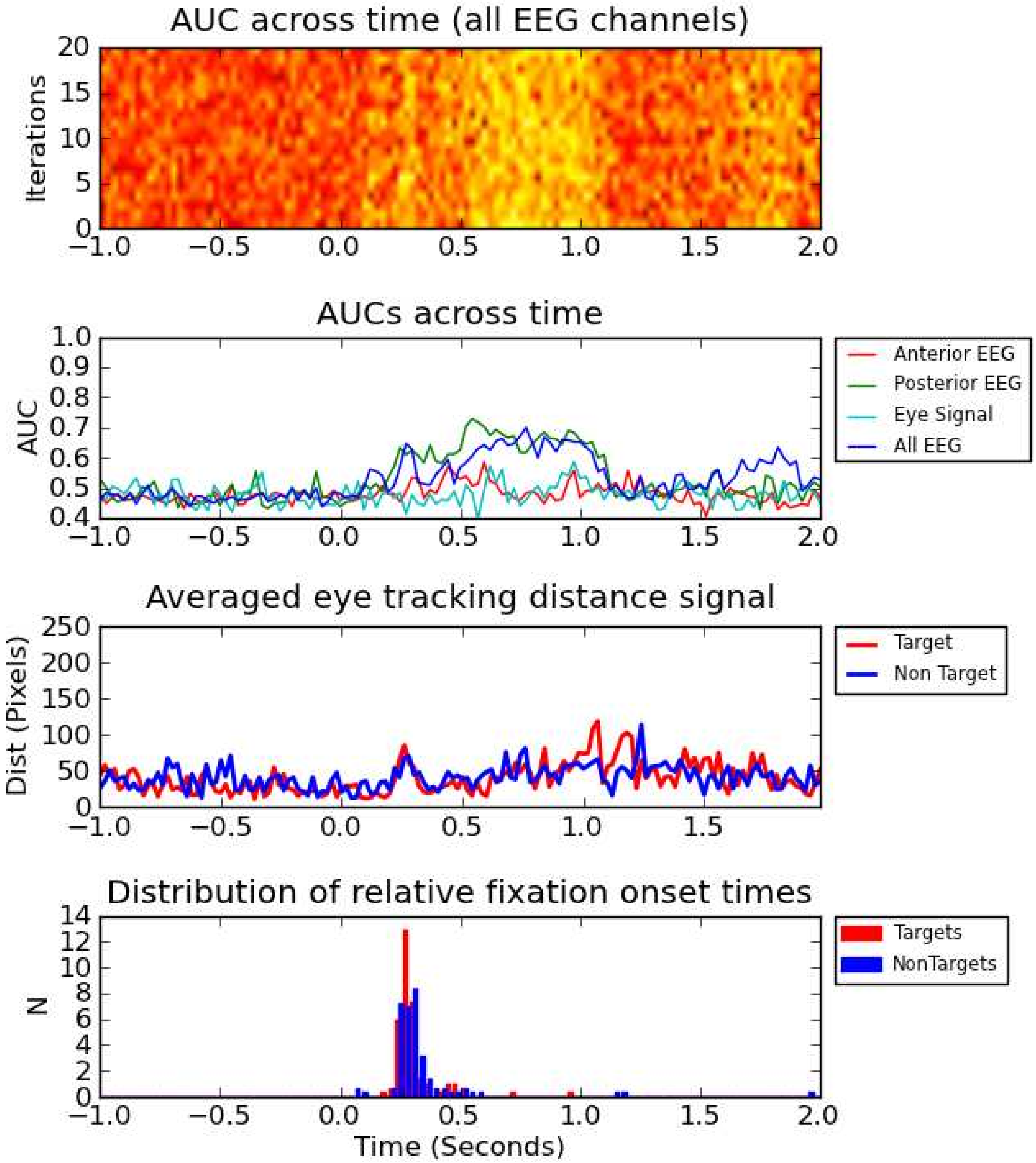

Figure C.73: Subject 2: Temporally aligned discrimination graphs map centred on the frame onset time showing differentiating activity related to target image detection compared to non target image detection

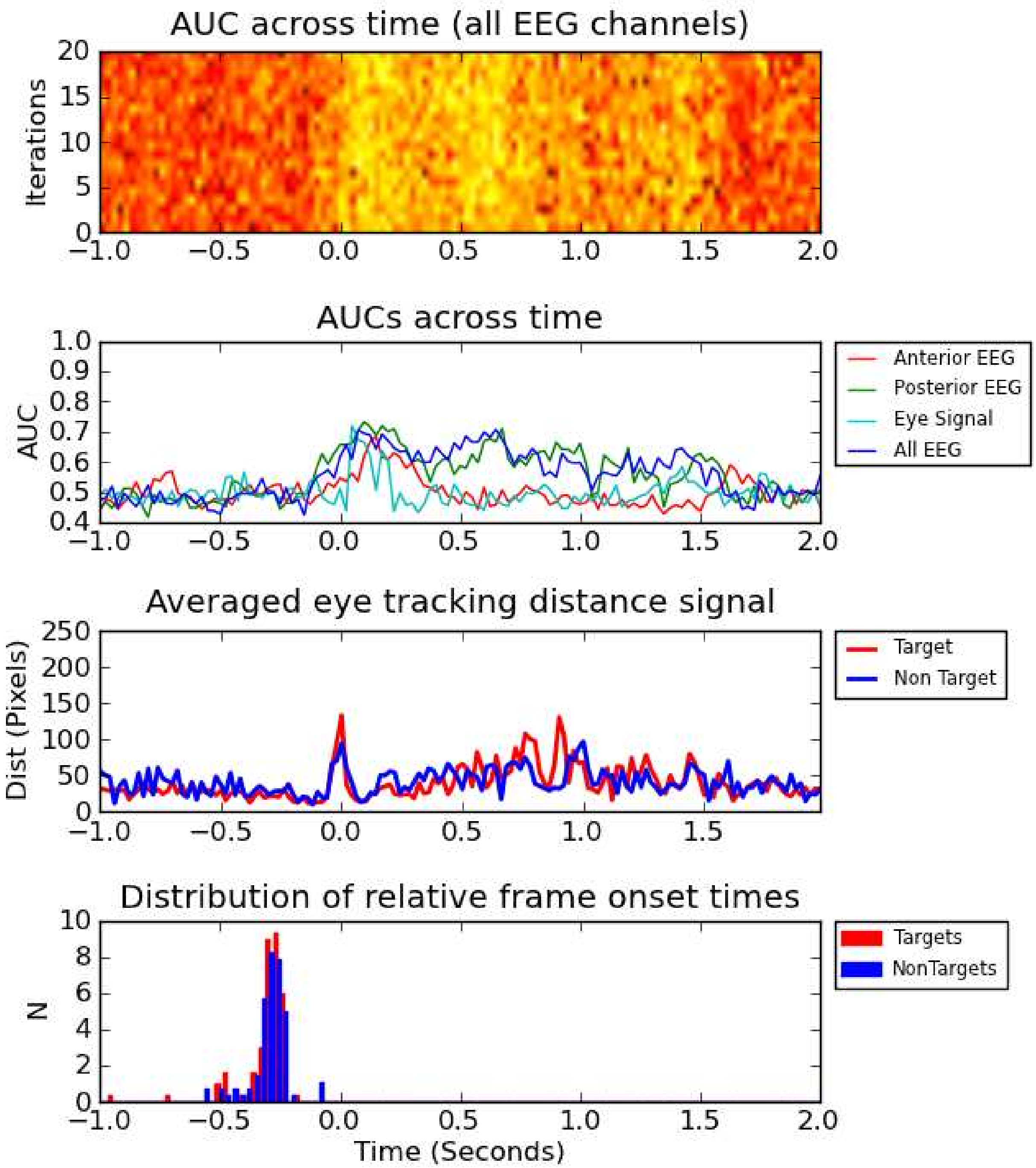

Figure C.74: Subject 2: Temporally aligned discrimination graphs centred on the fixation onset time showing differentiating activity related to target image detection compared to non target image detection

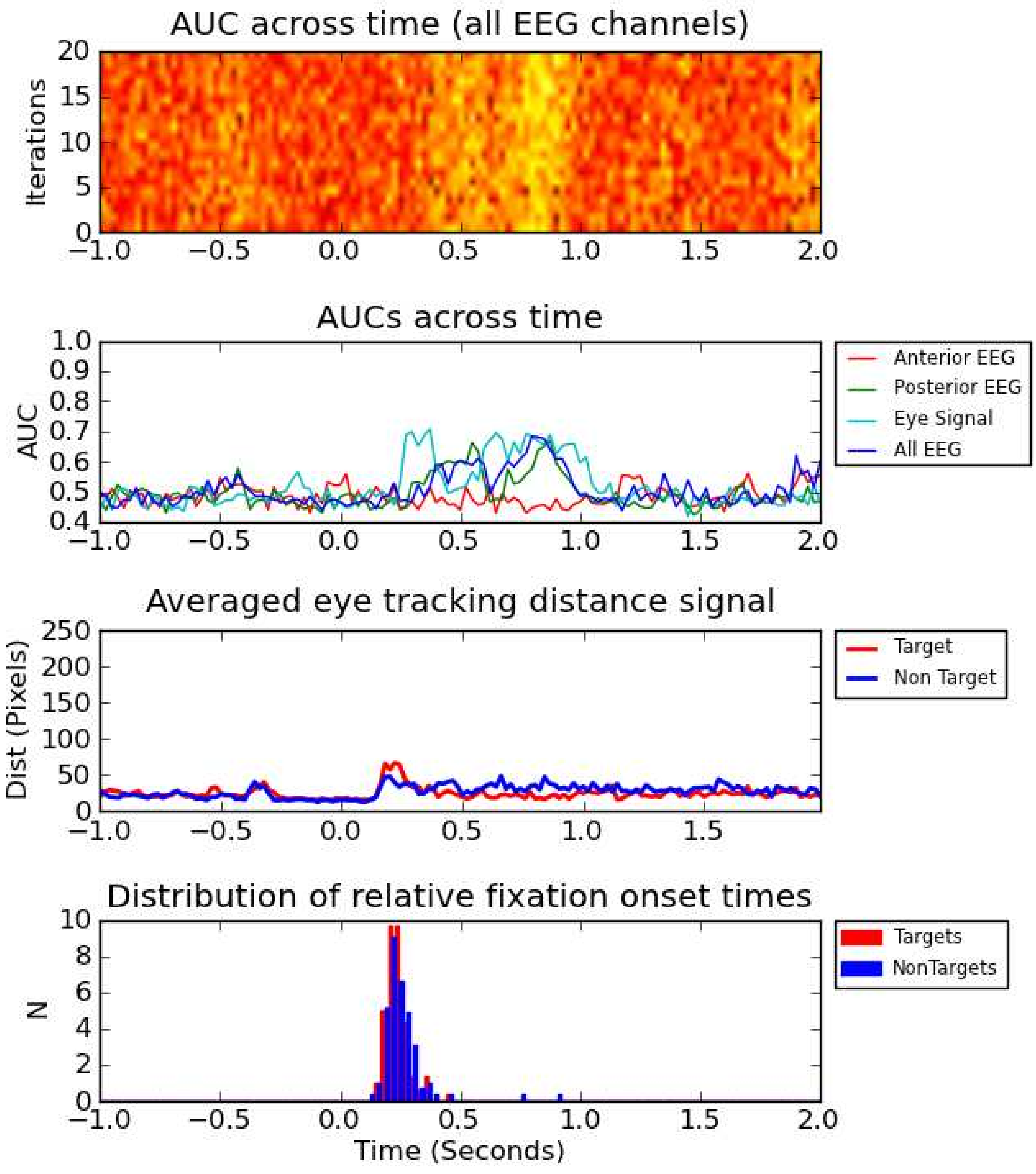

Figure C.75: Subject 3: Temporally aligned discrimination graphs map centred on the frame onset time showing differentiating activity related to target image detection compared to non target image detection

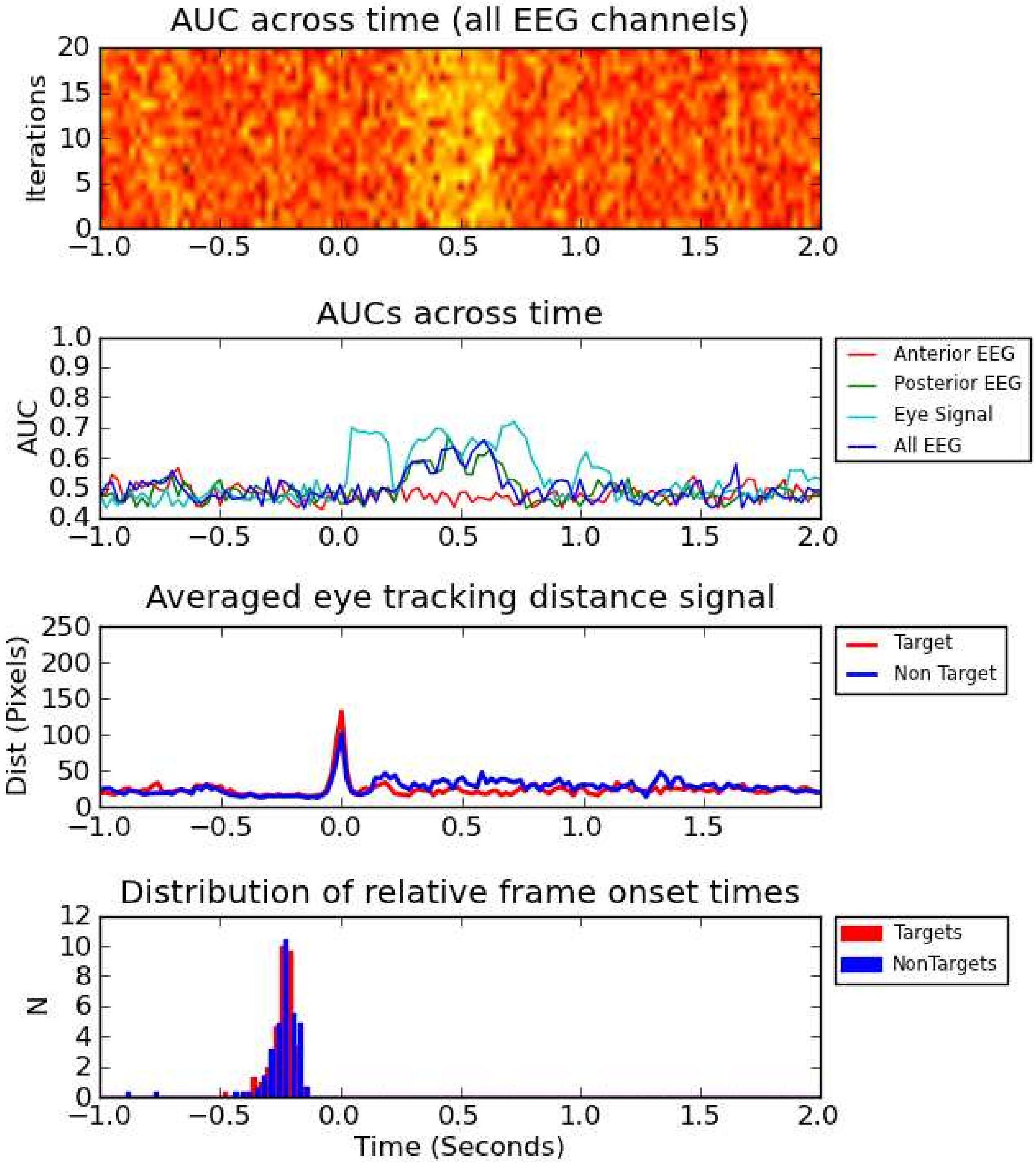

Figure C.76: Subject 3: Temporally aligned discrimination graphs centred on the fixation onset time showing differentiating activity related to target image detection compared to non target image detection

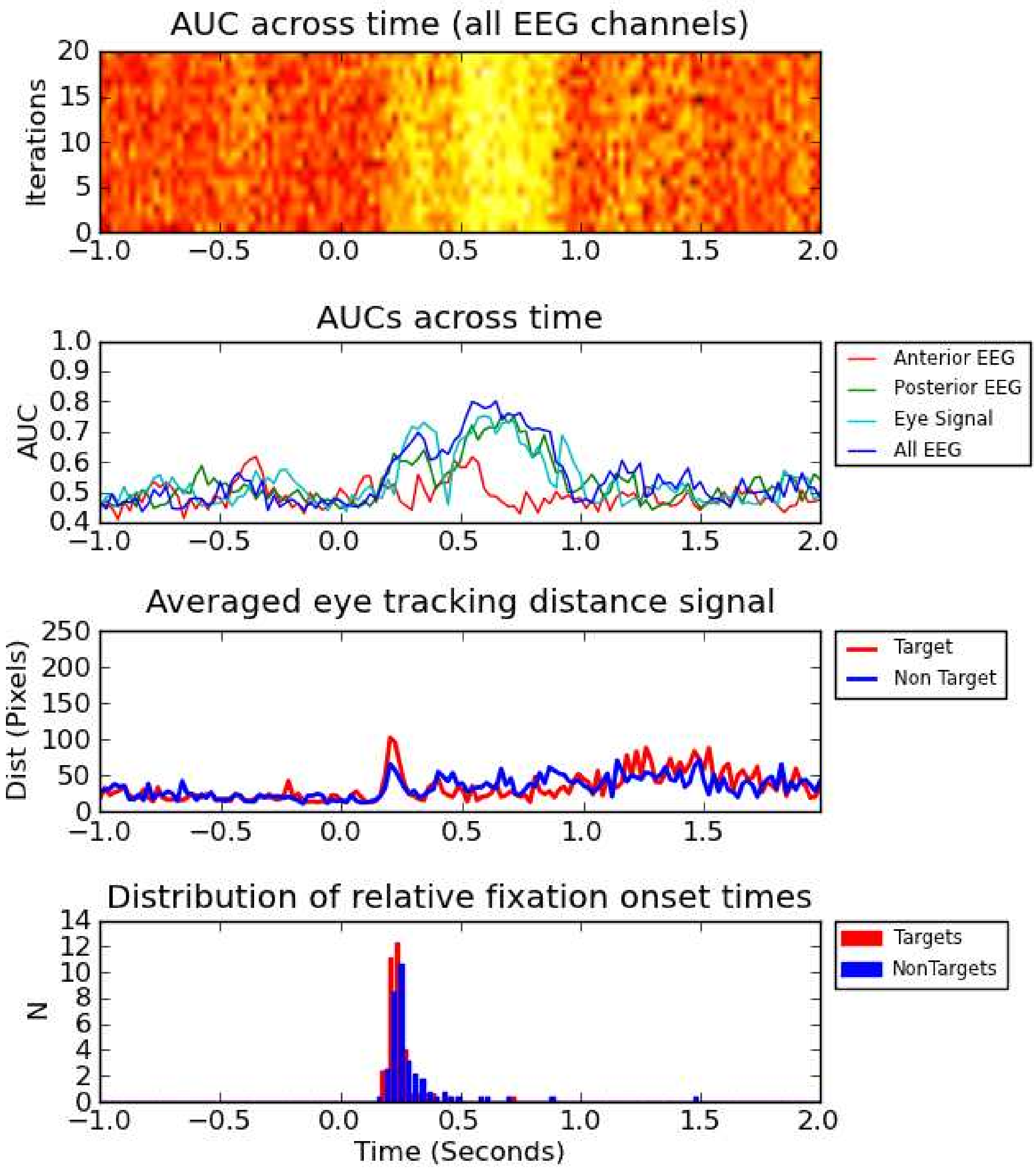

Figure C.77: Subject 4: Temporally aligned discrimination graphs map centred on the frame onset time showing differentiating activity related to target image detection compared to non target image detection

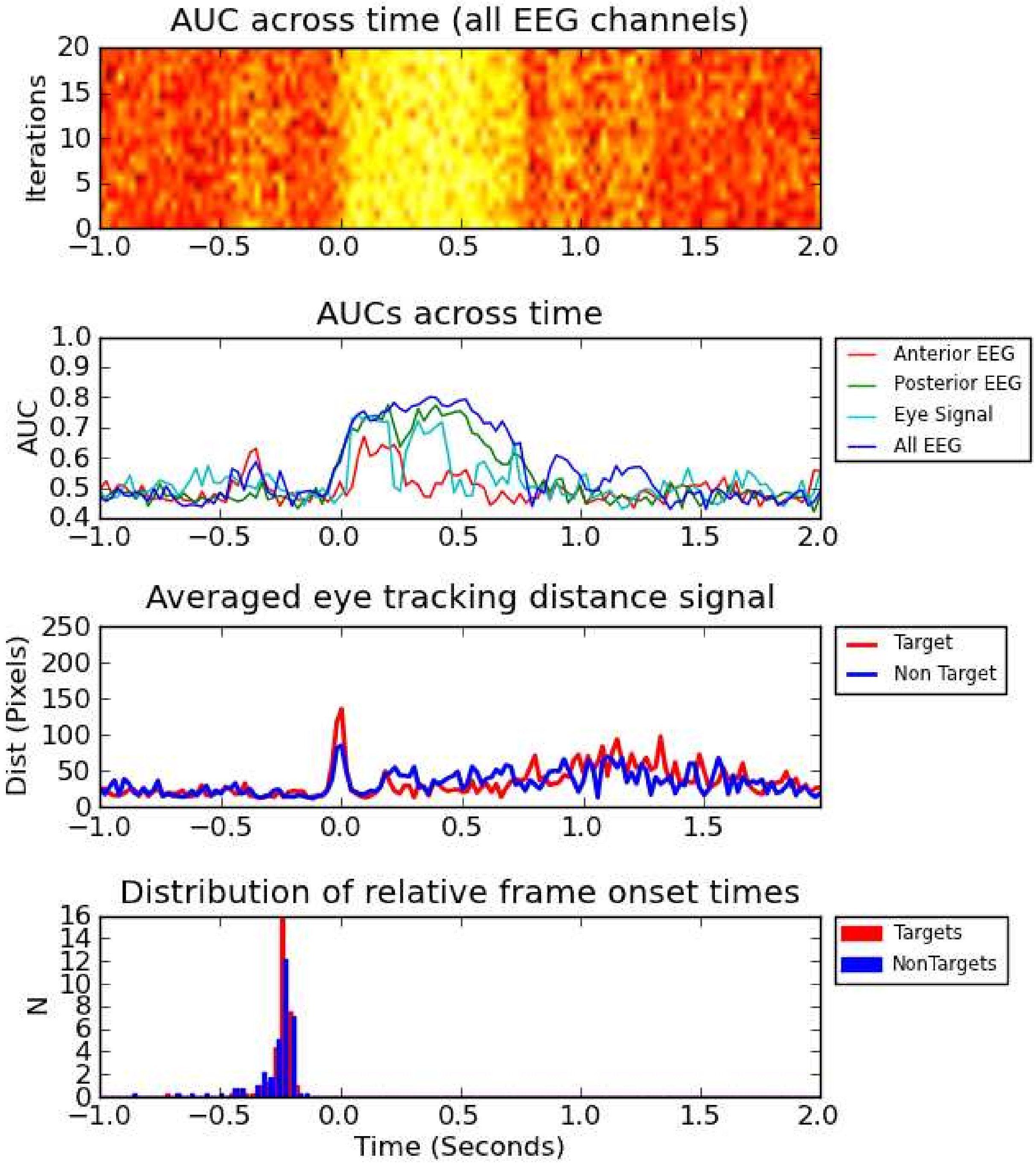

Figure C.78: Subject 4: Temporally aligned discrimination graphs centred on the fixation onset time showing differentiating activity related to target image detection compared to non target image detection

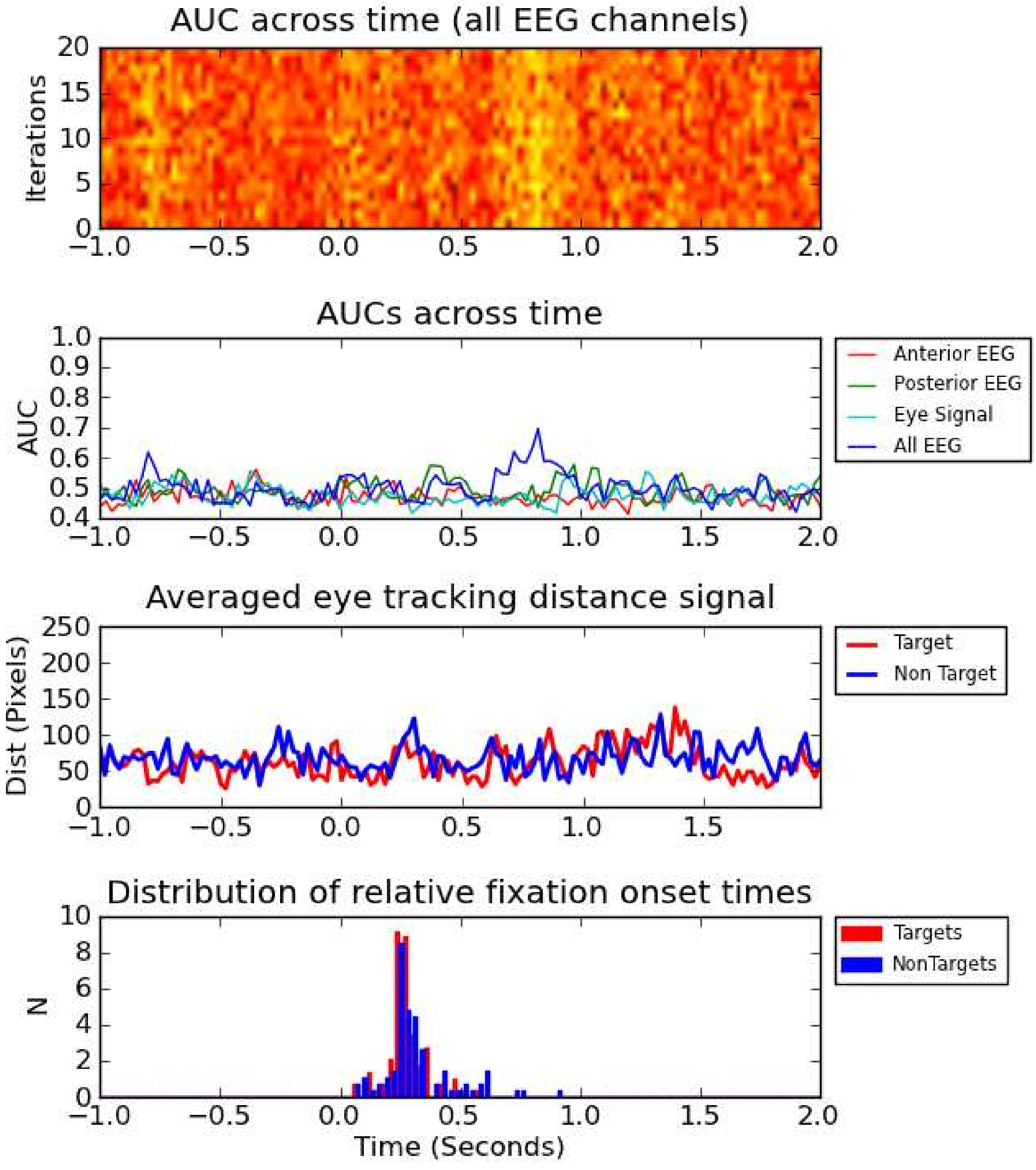

Figure C.79: Subject 5: Temporally aligned discrimination graphs map centred on the frame onset time showing differentiating activity related to target image detection compared to non target image detection

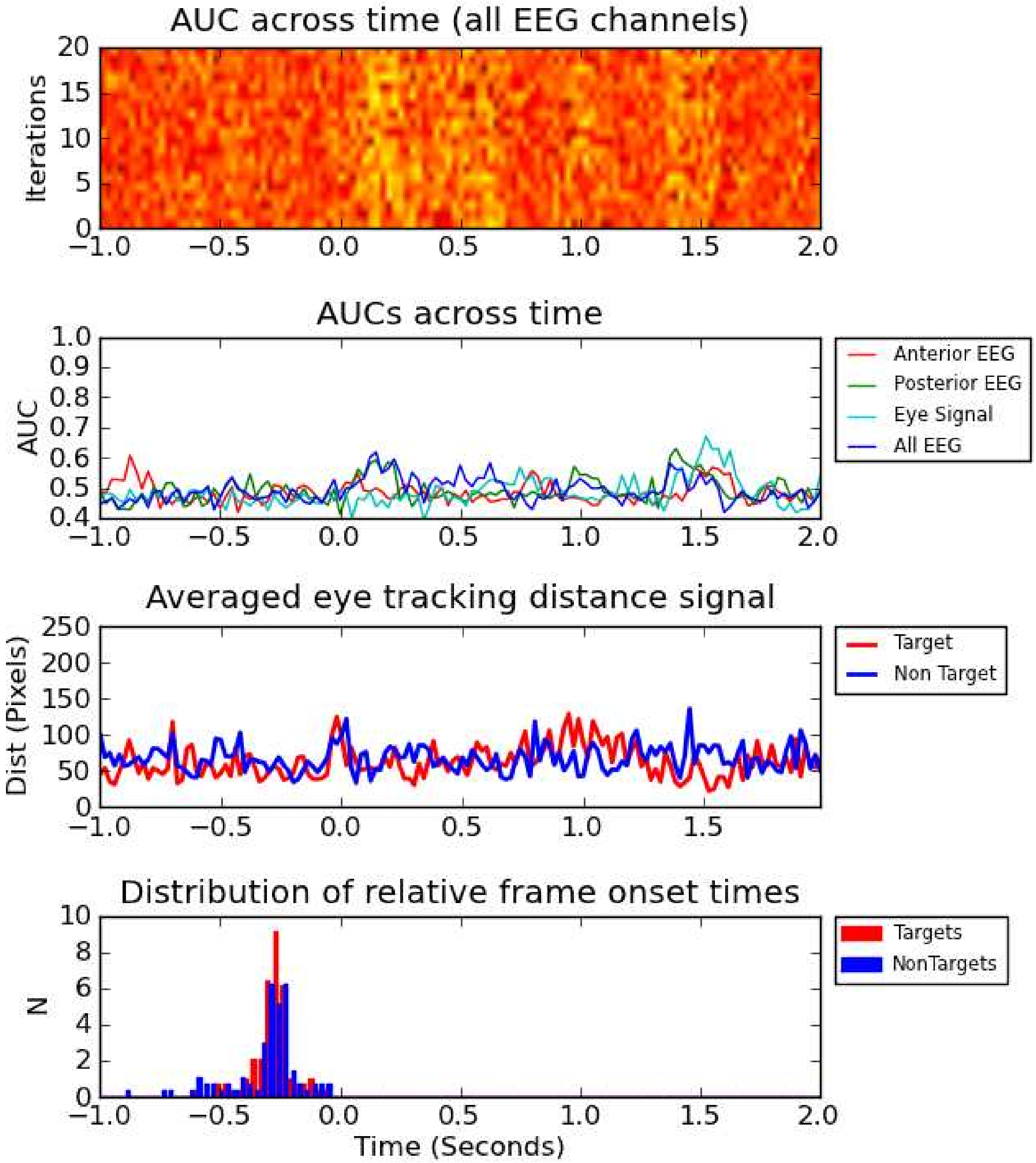

Figure C.80: Subject 5: Temporally aligned discrimination graphs centred on the fixation onset time showing differentiating activity related to target image detection compared to non target image detection

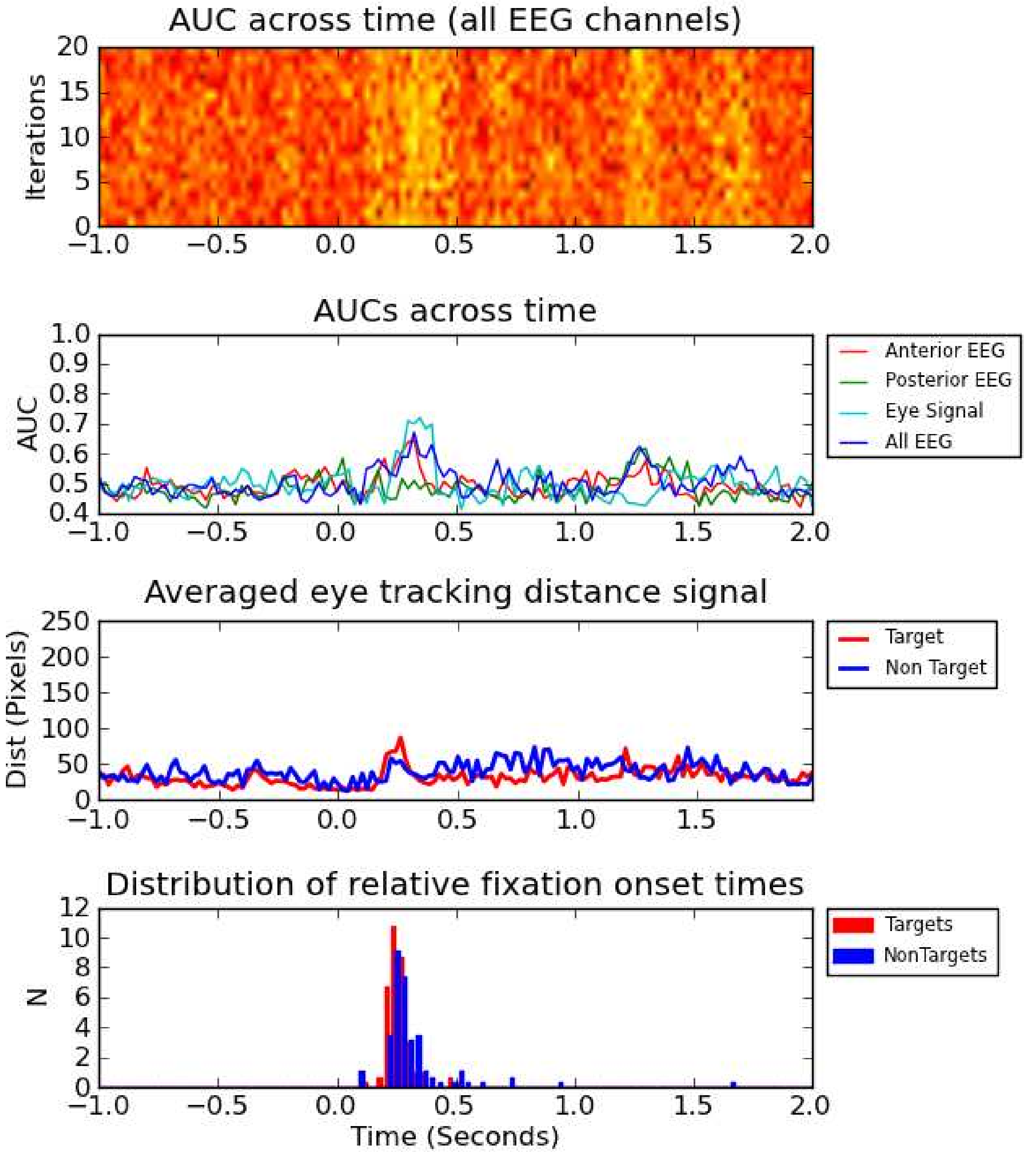

Figure C.81: Subject 7: Temporally aligned discrimination graphs map centred on the frame onset time showing differentiating activity related to target image detection compared to non target image detection

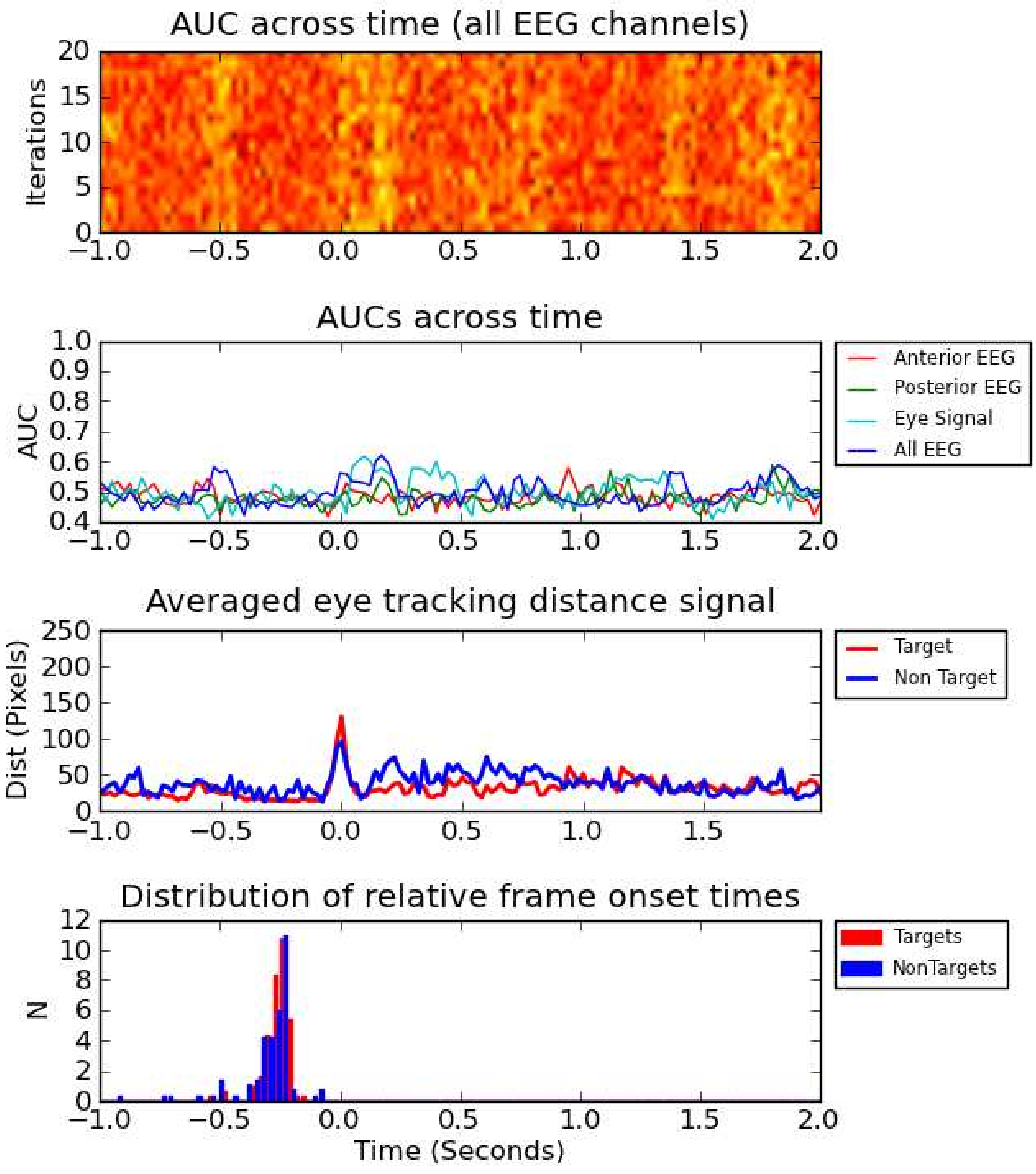

Figure C.82: Subject 7: Temporally aligned discrimination graphs centred on the fixation onset time showing differentiating activity related to target image detection compared to non target image detection

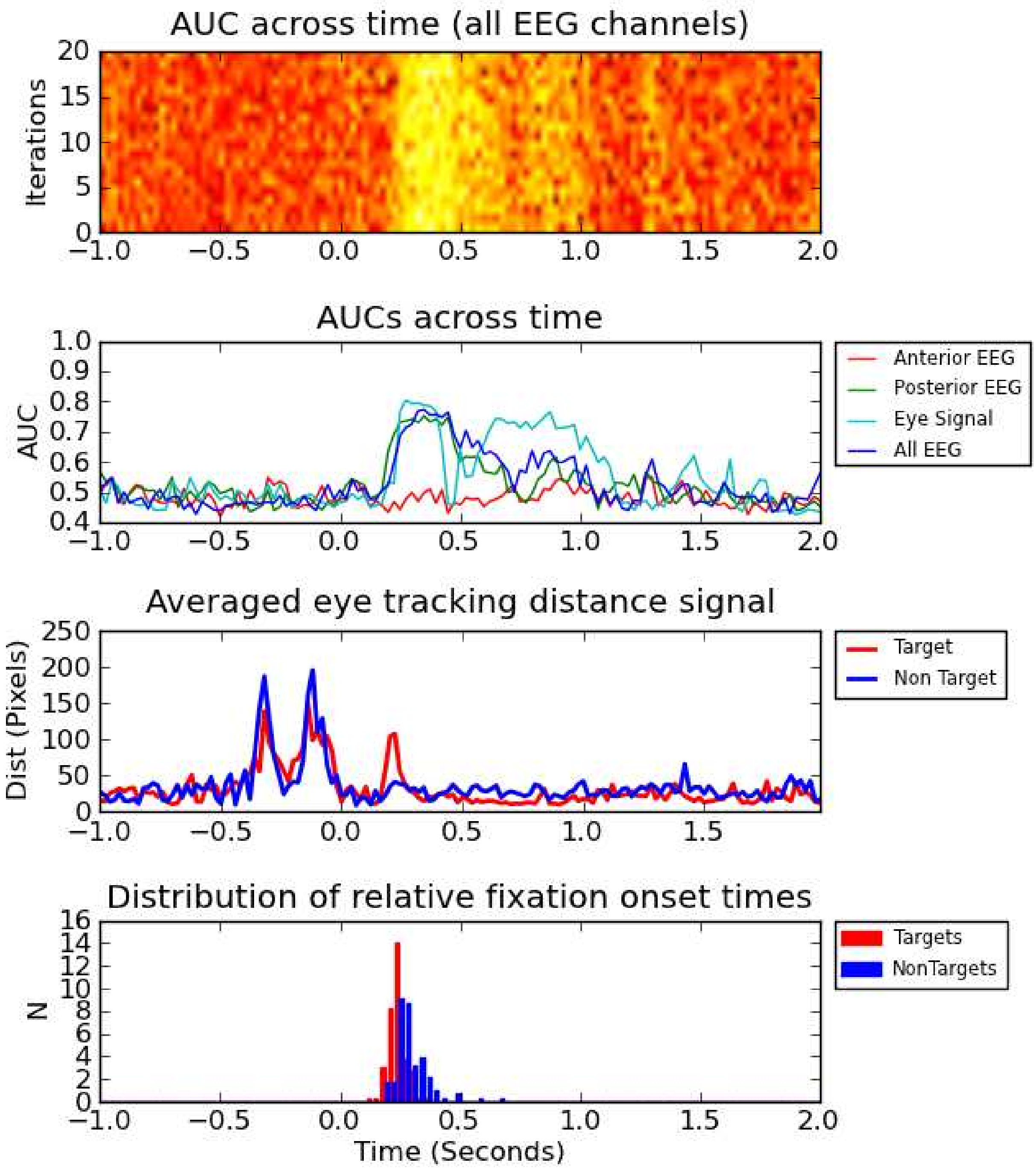

Figure C.83: Subject 8: Temporally aligned discrimination graphs map centred on the frame onset time showing differentiating activity related to target image detection compared to non target image detection

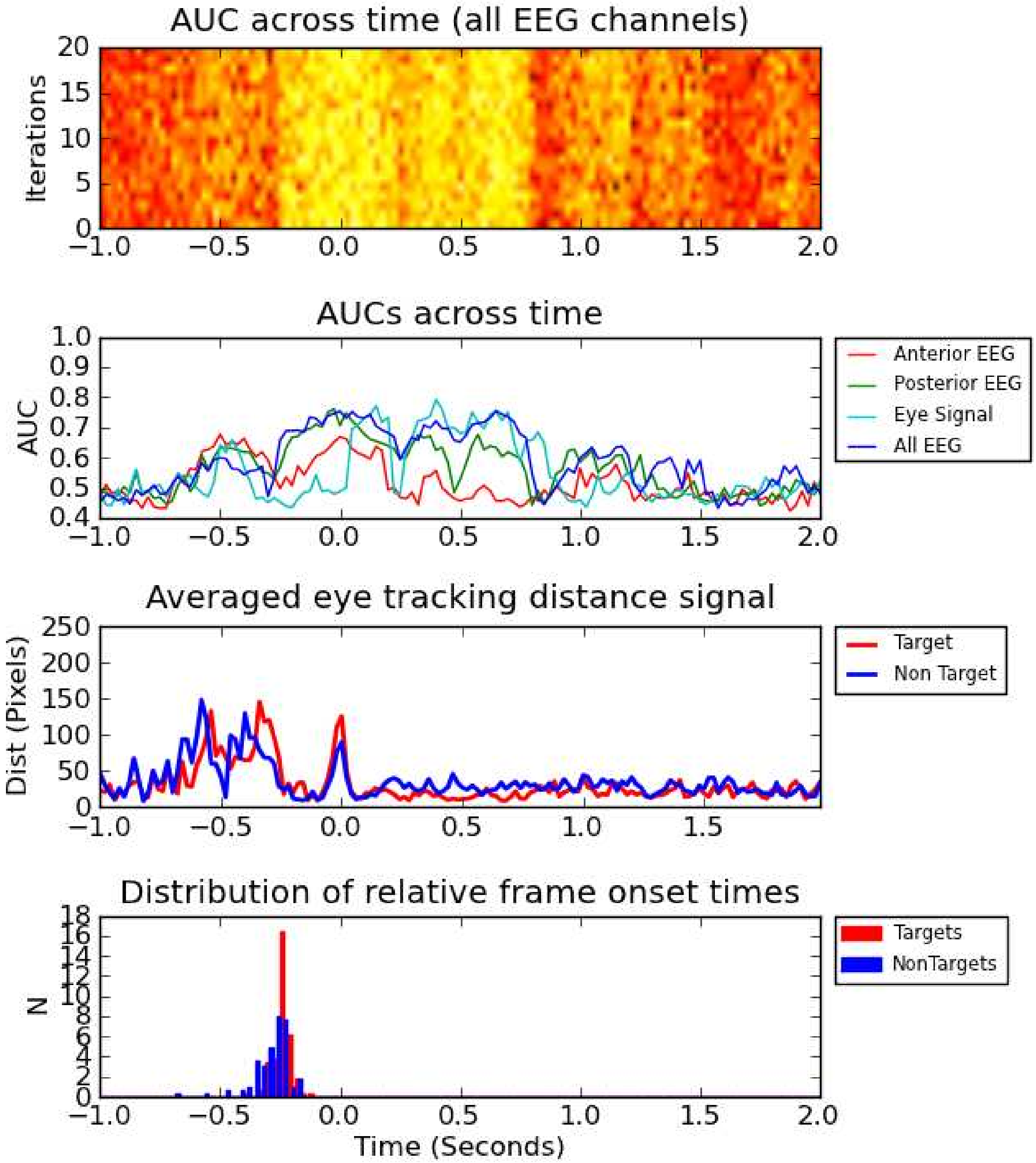

Figure C.84: Subject 8: Temporally aligned discrimination graphs centred on the fixation onset time showing differentiating activity related to target image detection compared to non target image detection

C.4 Scalp Plots for Experiment 3

In this section we provide the scalp plots obtained for experiment 3 across all subjects. These materials are intended to primarily supplement Section 3.4.

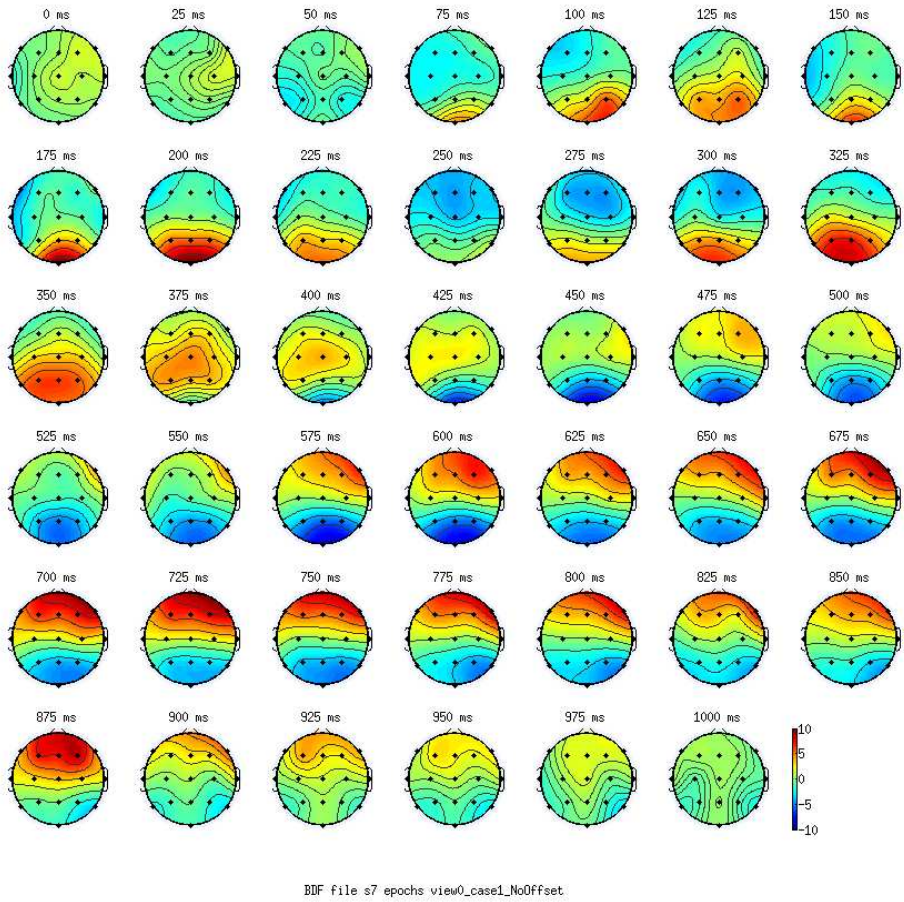

Figure C.85: Subject 2: Averaged scalp plots aligned to frame onset for target frames

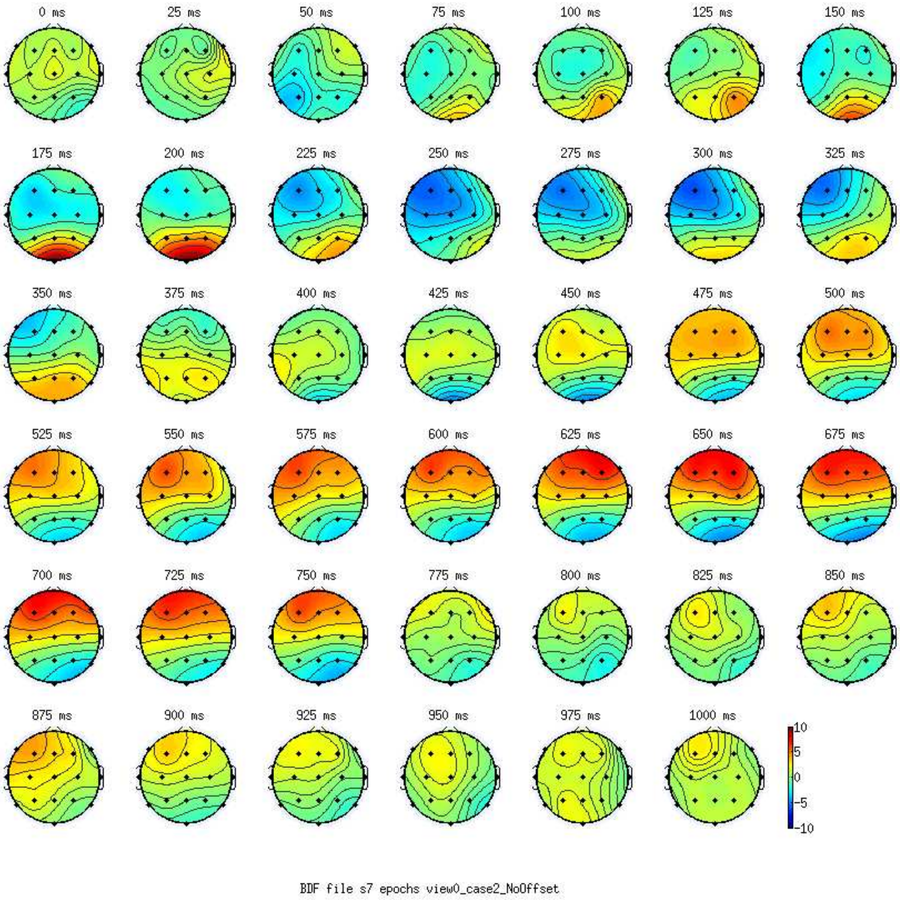

Figure C.86: Subject 2: Averaged scalp plots aligned to frame onset for non-target frames

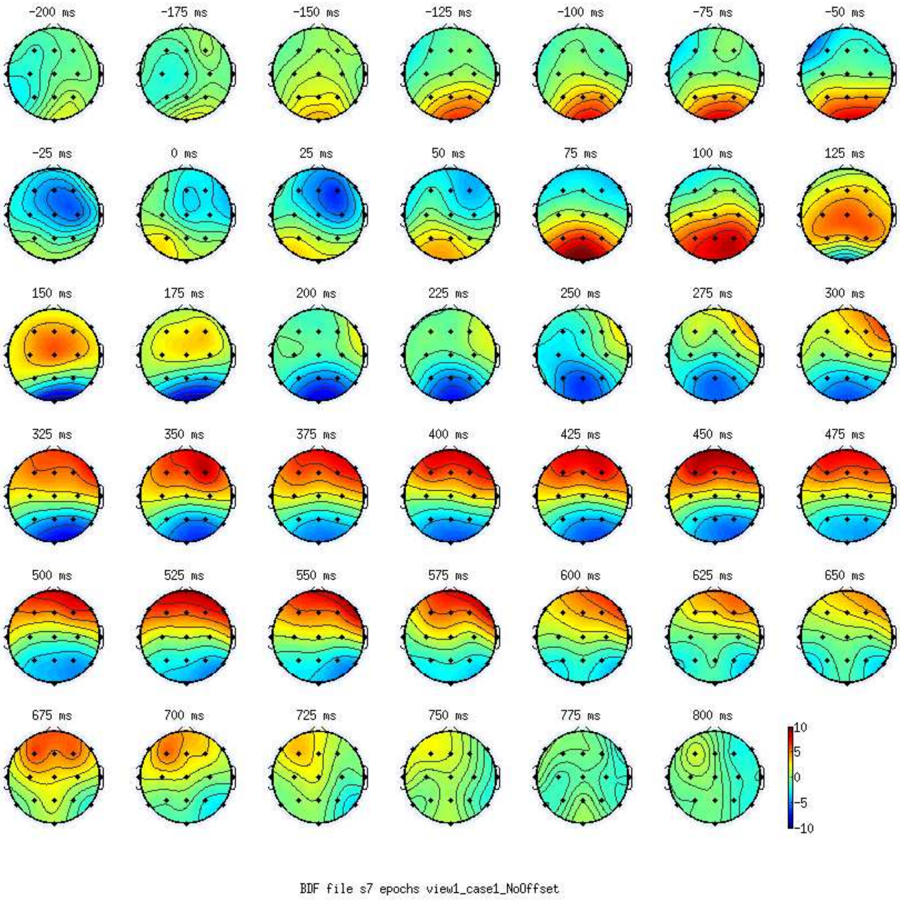

Figure C.87: Subject 2: Averaged scalp plots aligned to fixation onset for target frames

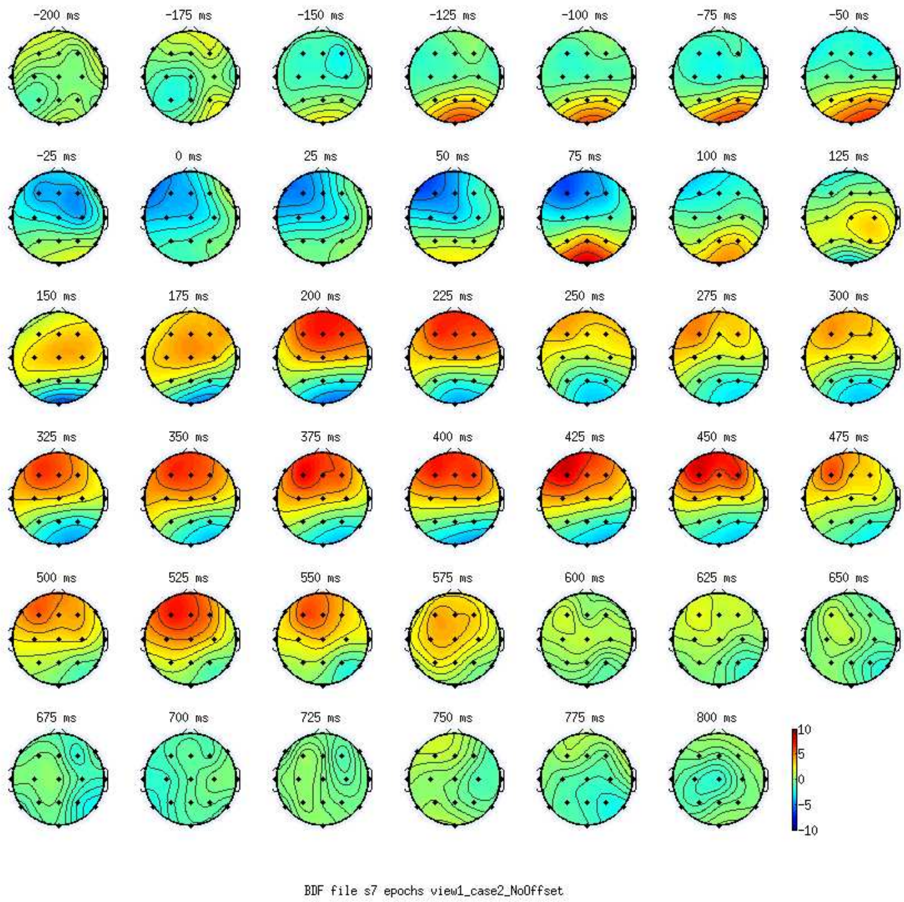

Figure C.88: Subject 2: Averaged scalp plots aligned to fixation onset for non-target frames

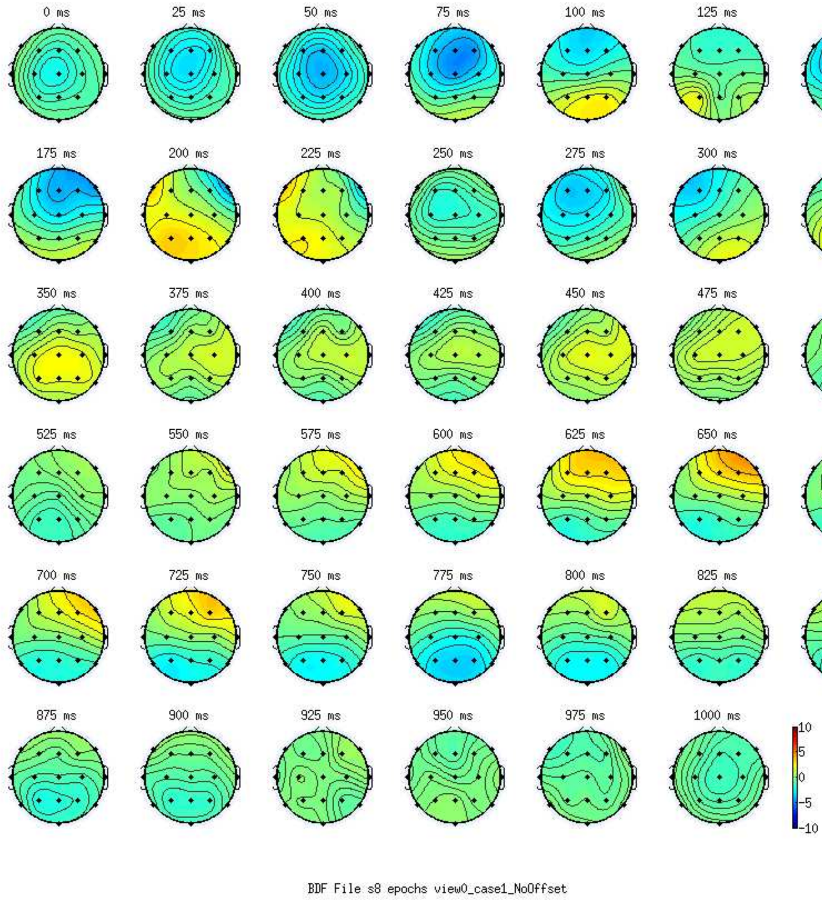

Figure C.89: Subject 3: Averaged scalp plots aligned to frame onset for target frames

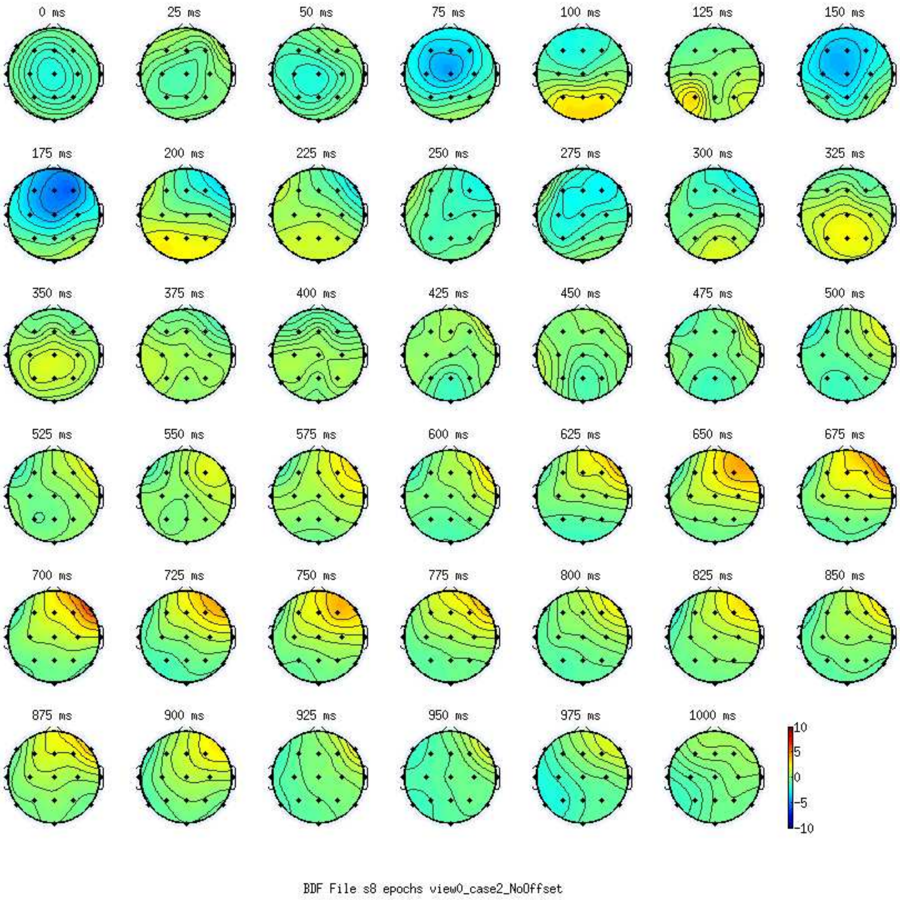

Figure C.90: Subject 3: Averaged scalp plots aligned to frame onset for non-target frames

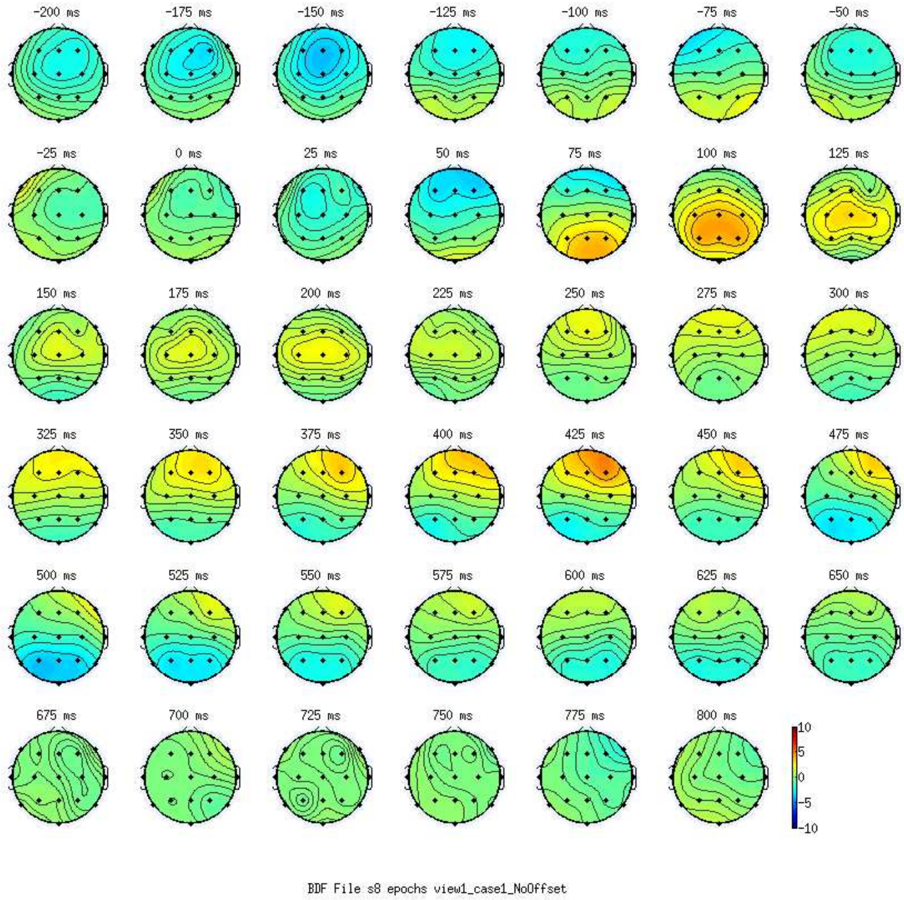

Figure C.91: Subject 3: Averaged scalp plots aligned to fixation onset for target frames

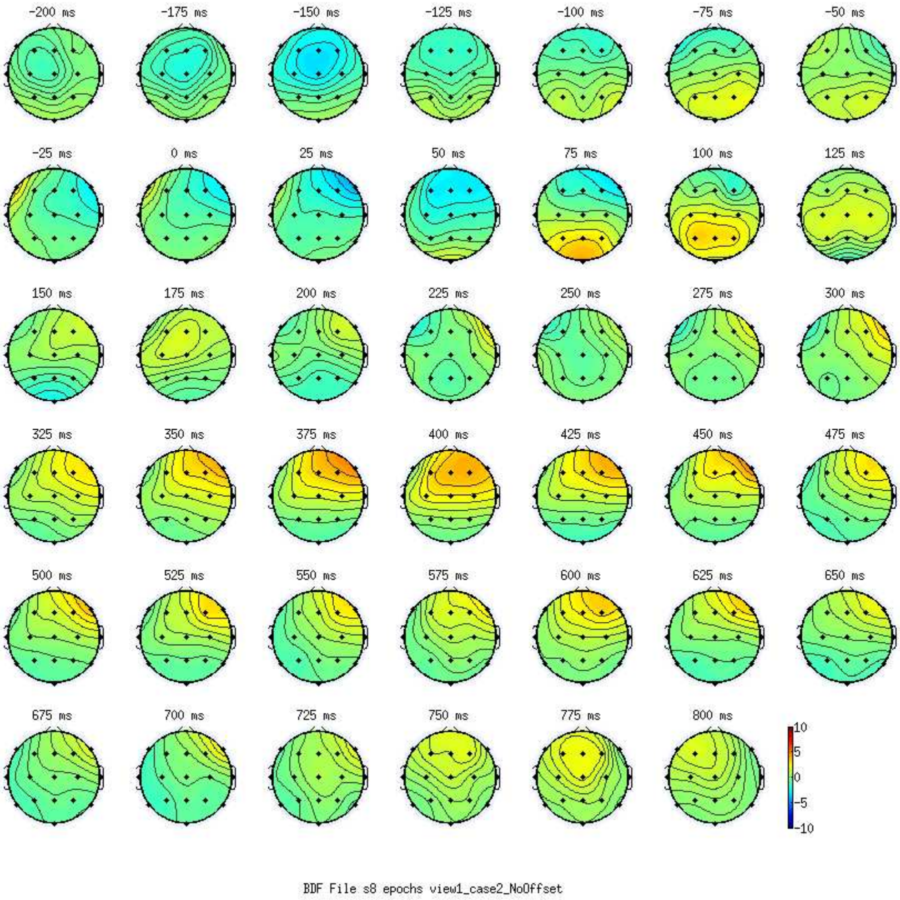

Figure C.92: Subject 3: Averaged scalp plots aligned to fixation onset for non-target frames

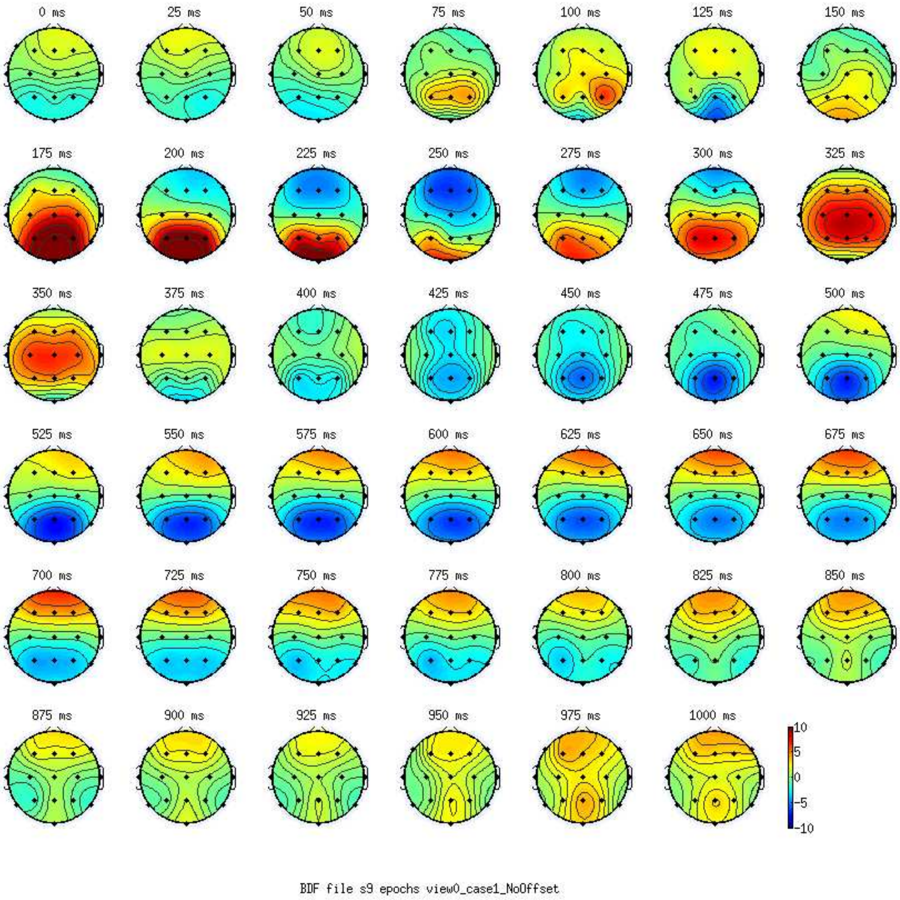

Figure C.93: Subject 4: Averaged scalp plots aligned to frame onset for target frames

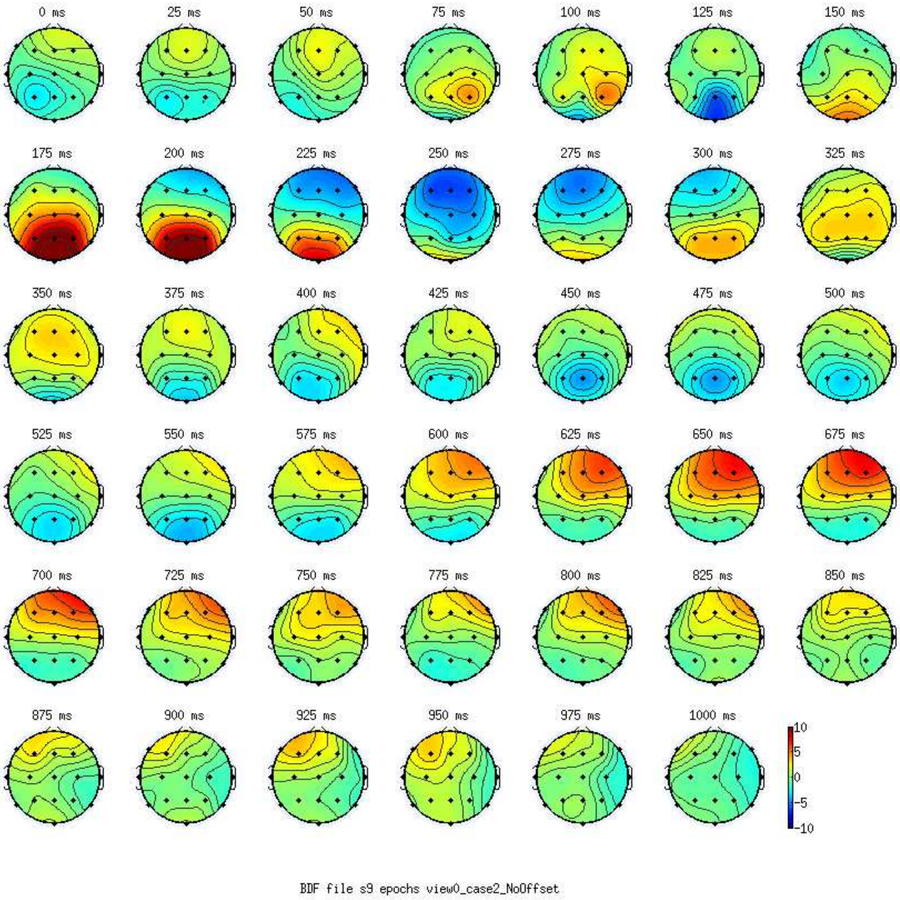

Figure C.94: Subject 4: Averaged scalp plots aligned to frame onset for non-target frames

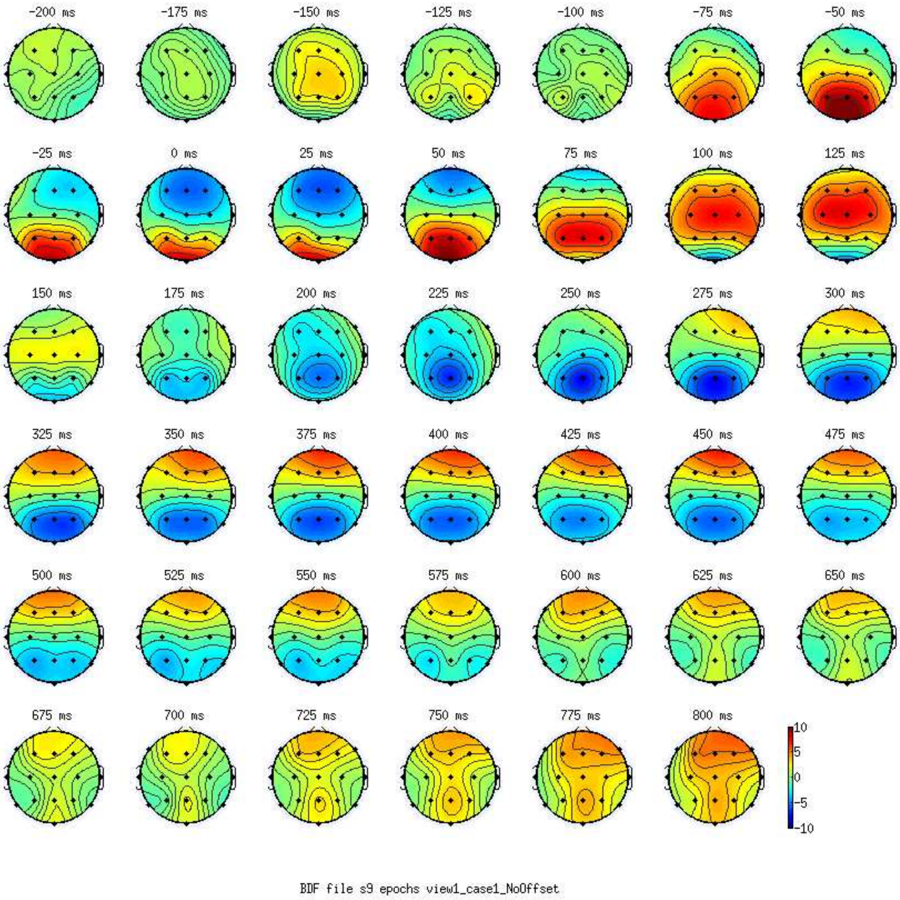

Figure C.95: Subject 4: Averaged scalp plots aligned to fixation onset for target frames

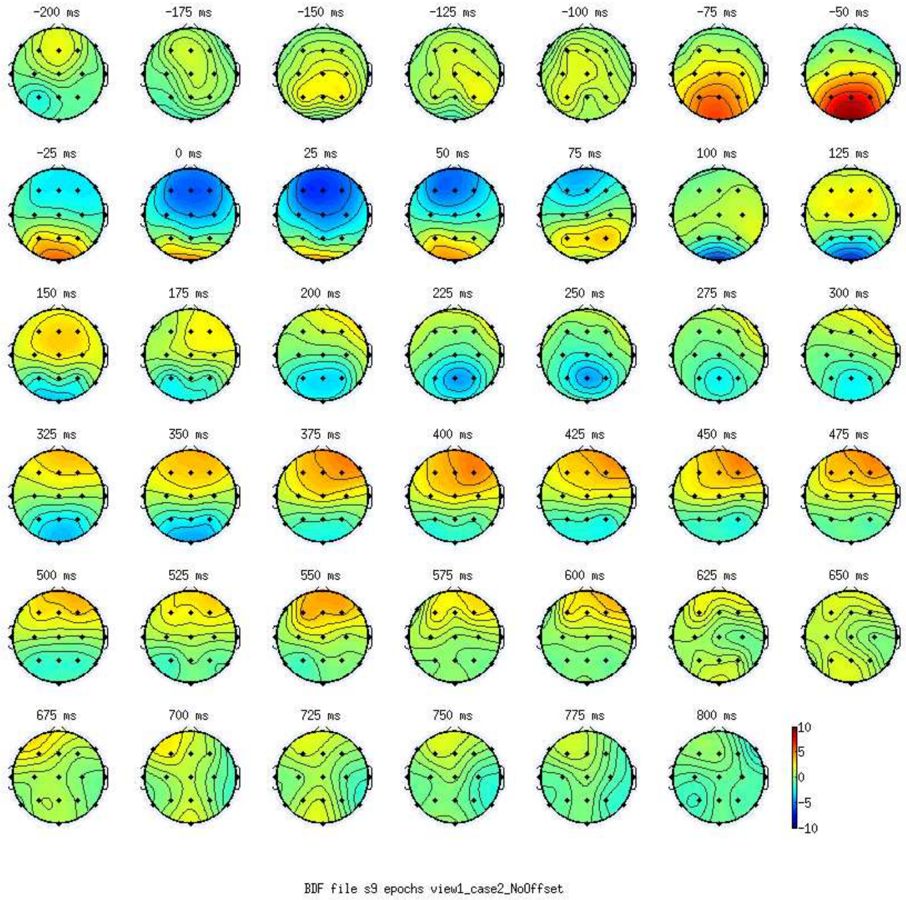

Figure C.96: Subject 4: Averaged scalp plots aligned to fixation onset for non-target frames

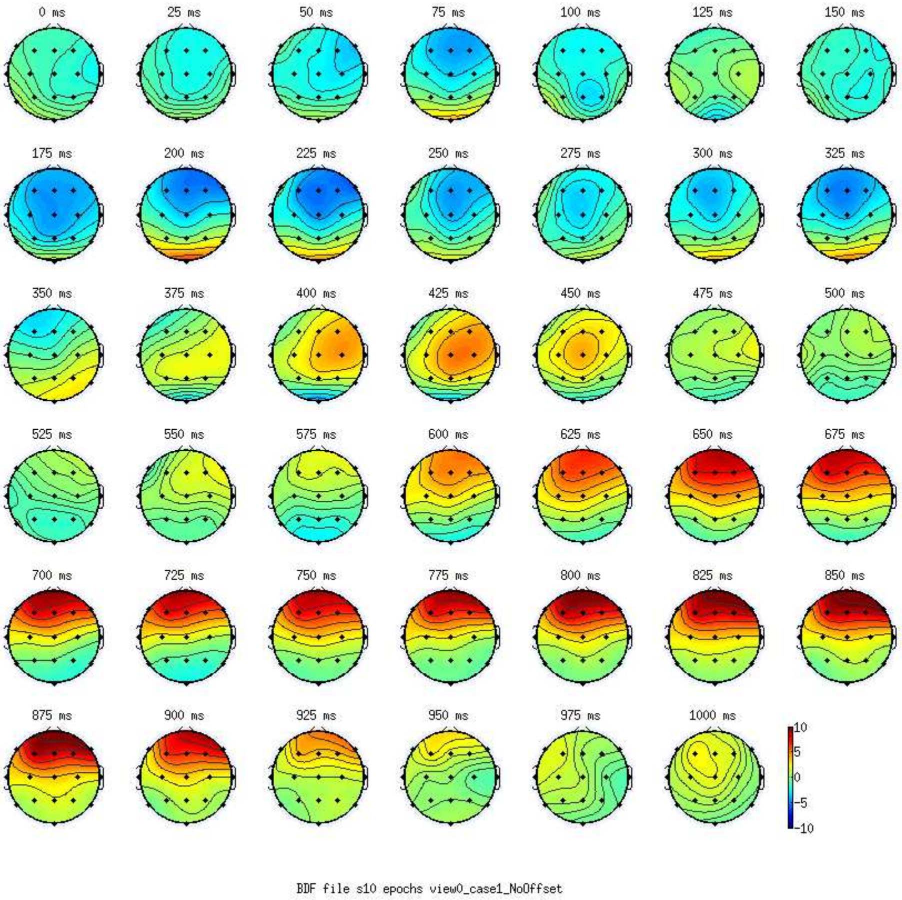

Figure C.97: Subject 5: Averaged scalp plots aligned to frame onset for target frames

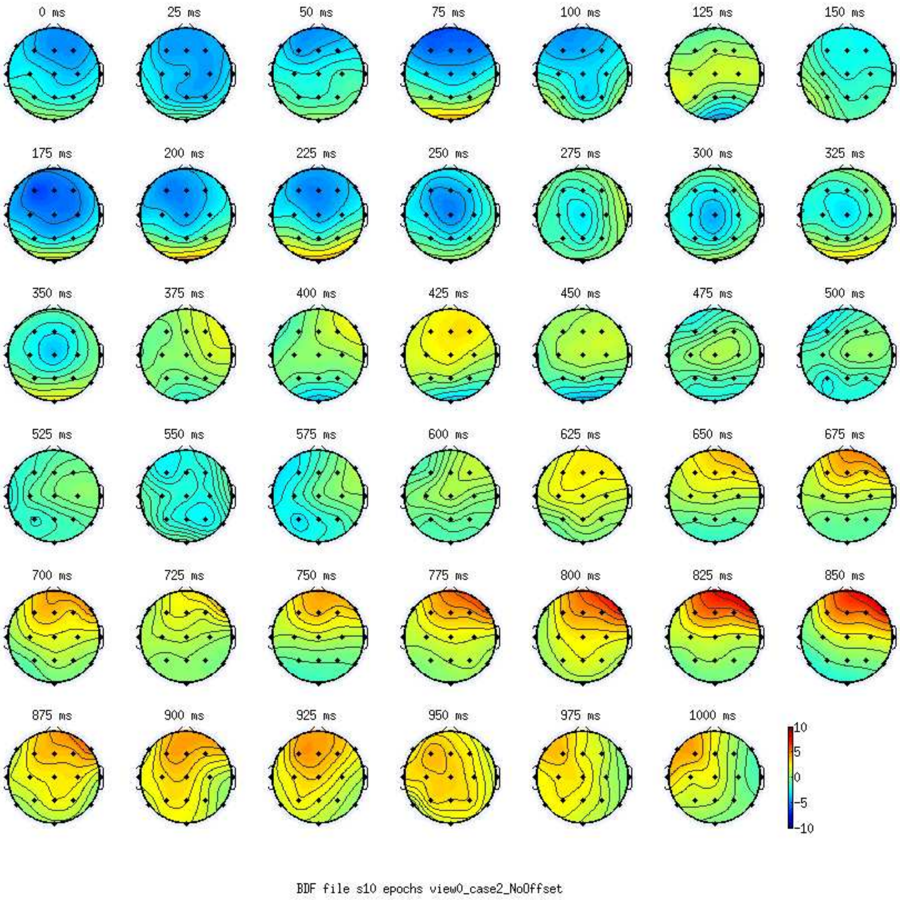

Figure C.98: Subject 5: Averaged scalp plots aligned to frame onset for non-target frames

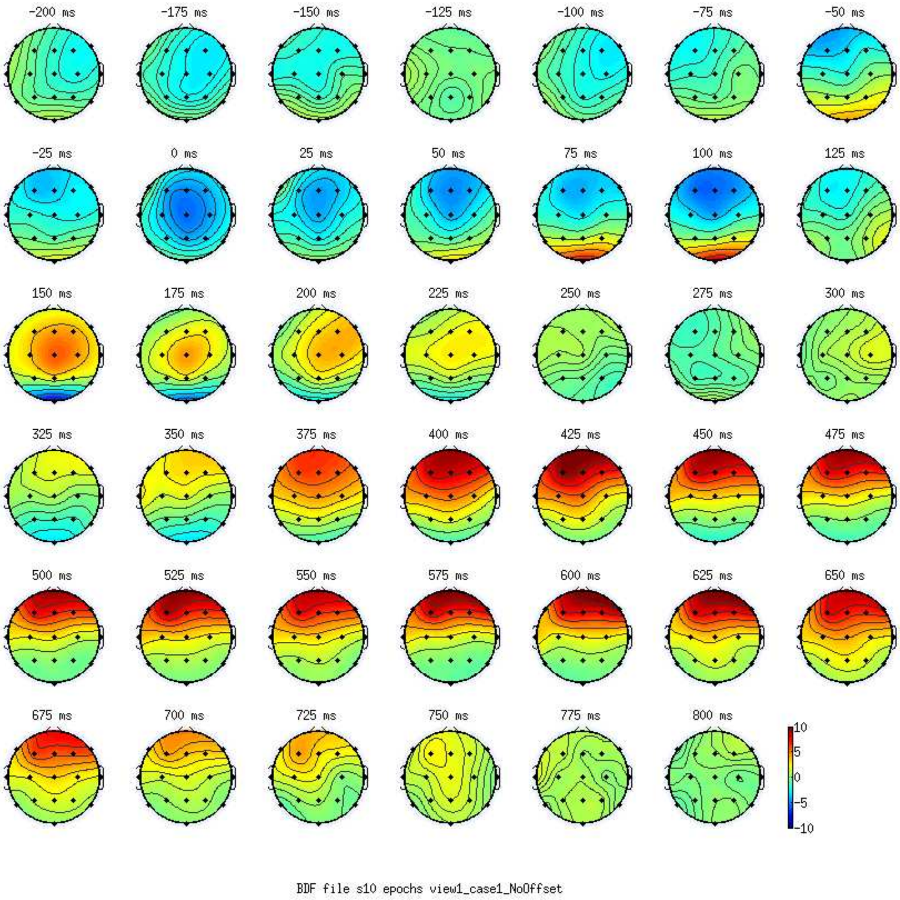

Figure C.99: Subject 5: Averaged scalp plots aligned to fixation onset for target frames

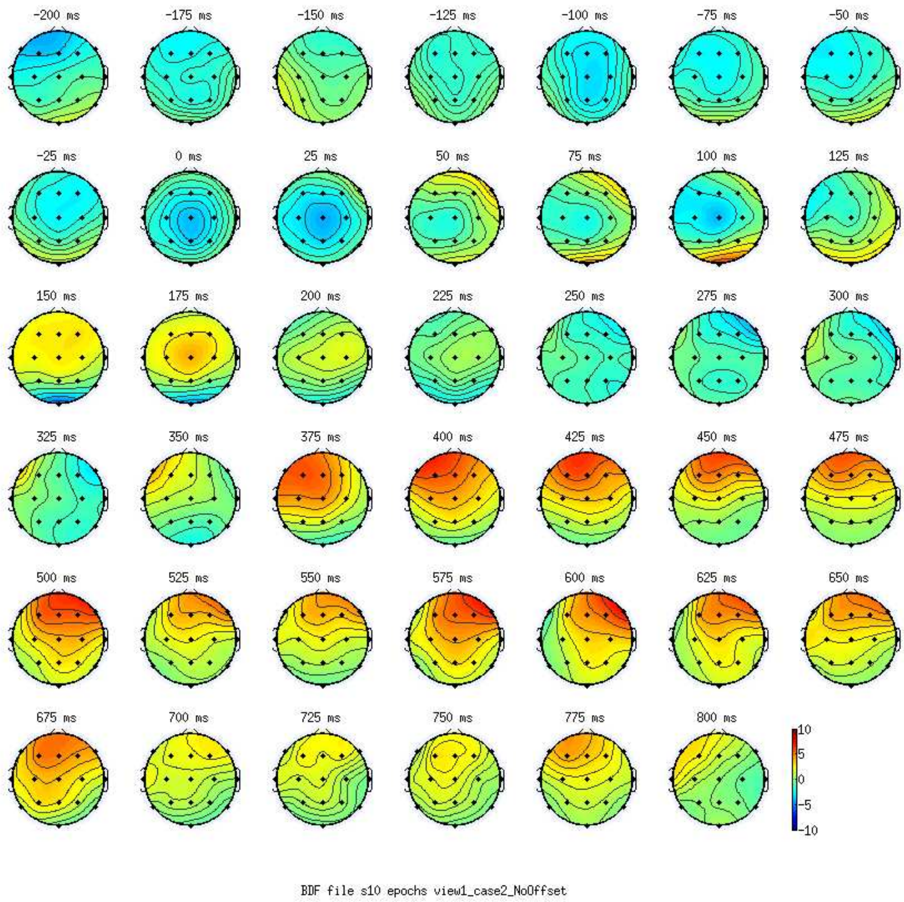

Figure C.100: Subject 5: Averaged scalp plots aligned to fixation onset for non-target frames

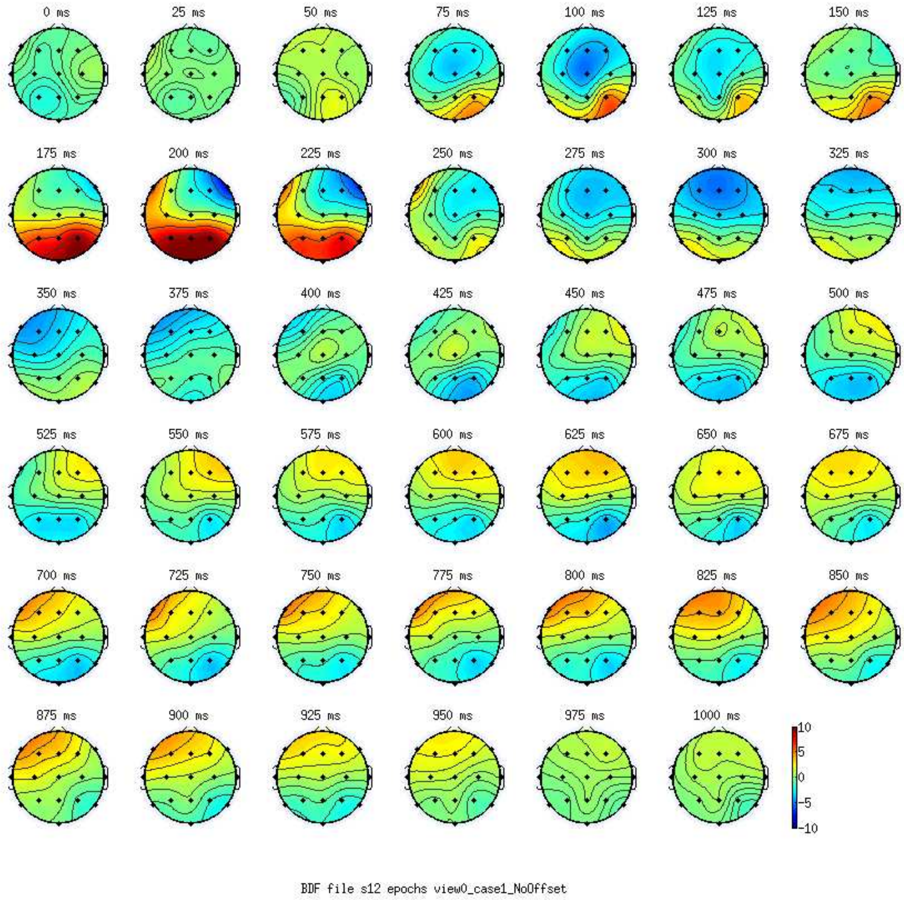

Figure C.101: Subject 7: Averaged scalp plots aligned to frame onset for target frames

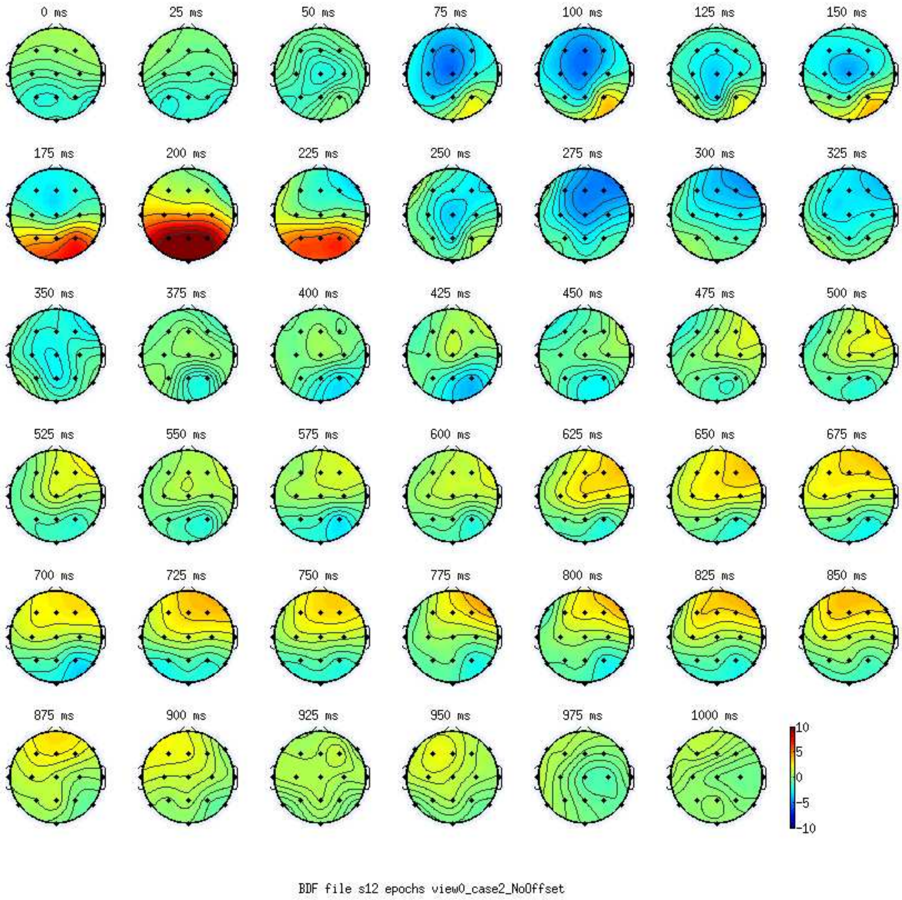

Figure C.102: Subject 7: Averaged scalp plots aligned to frame onset for non-target frames

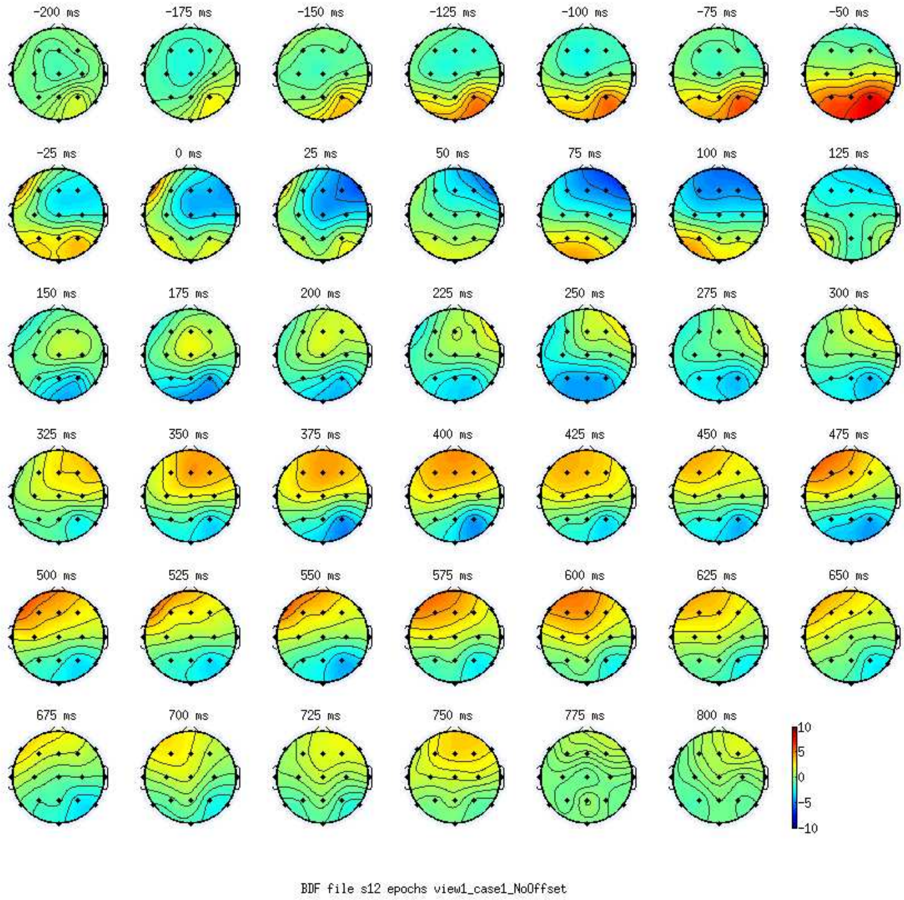

Figure C.103: Subject 7: Averaged scalp plots aligned to fixation onset for target frames

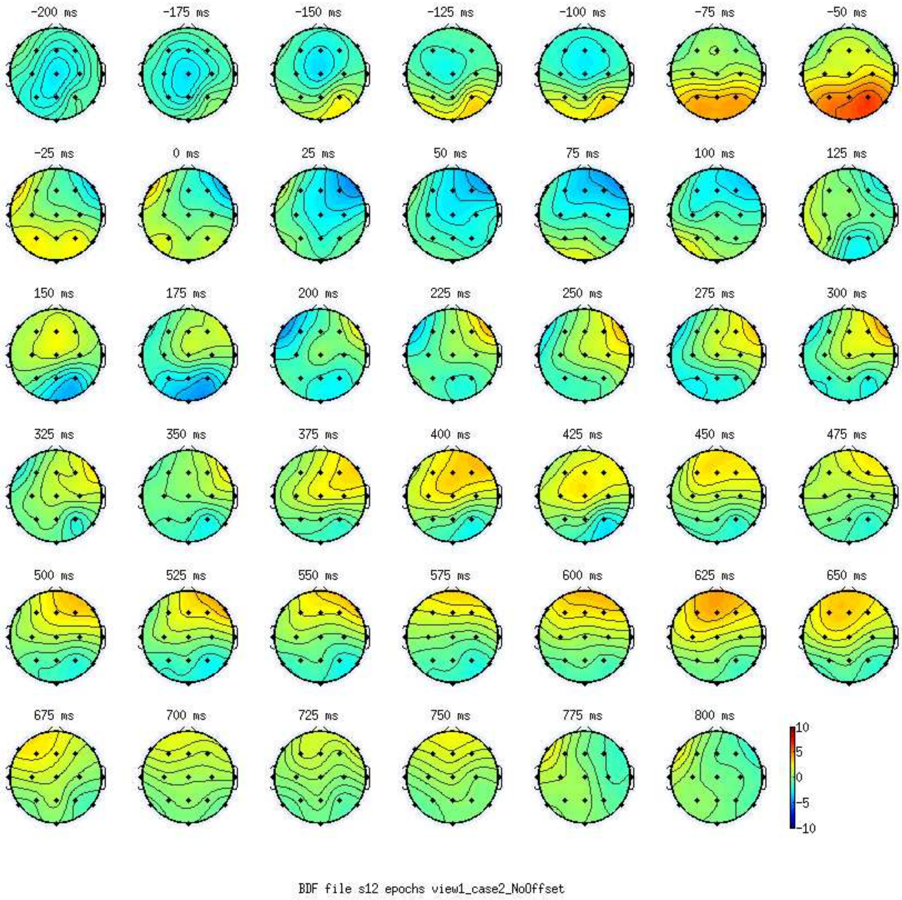

Figure C.104: Subject 7: Averaged scalp plots aligned to fixation onset for non-target frames

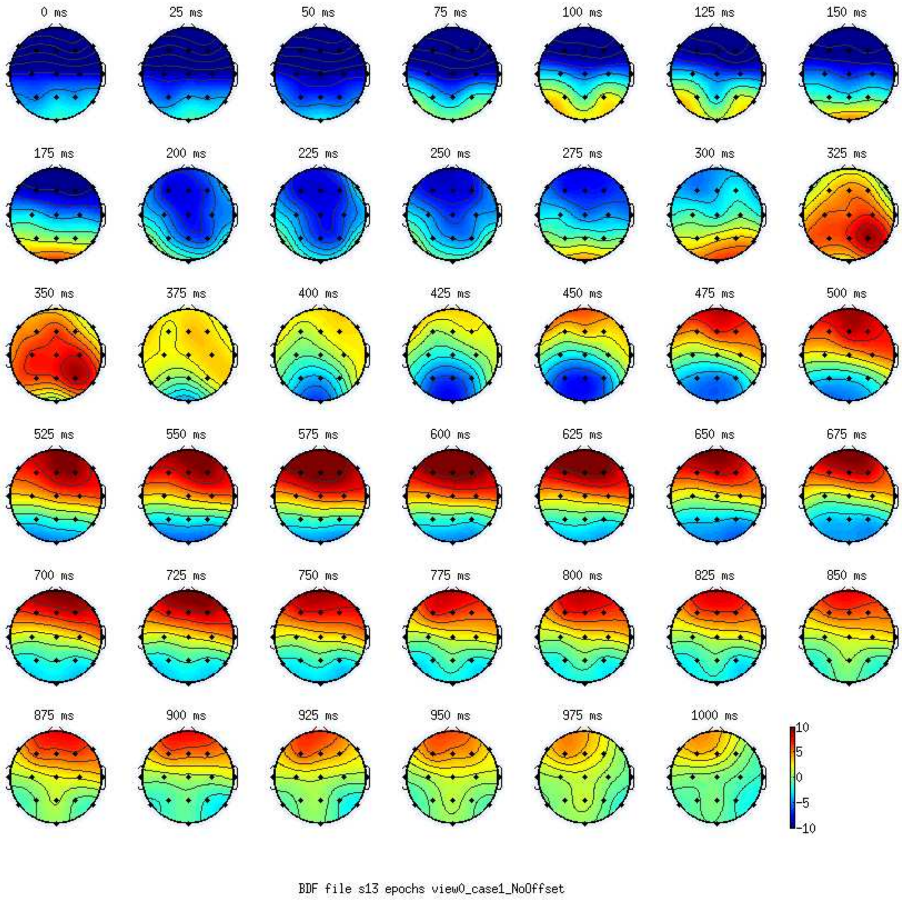

Figure C.105: Subject 8: Averaged scalp plots aligned to frame onset for target frames

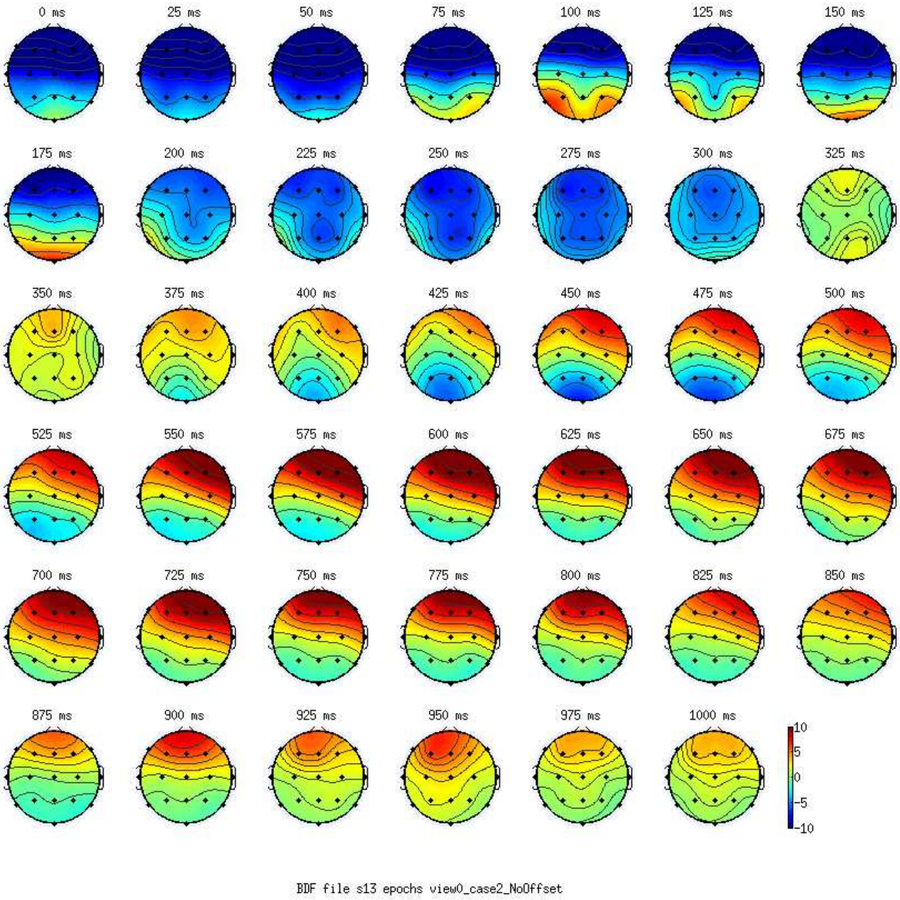

Figure C.106: Subject 8: Averaged scalp plots aligned to frame onset for non-target frames

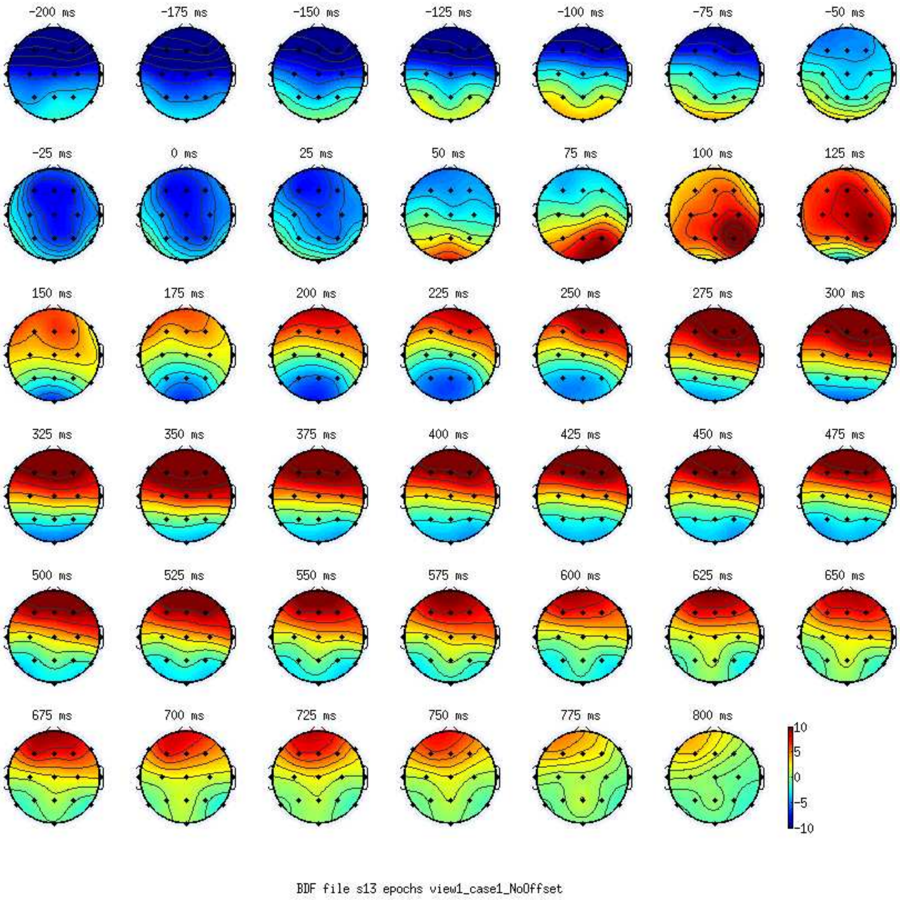

Figure C.107: Subject 8: Averaged scalp plots aligned to fixation onset for target frames

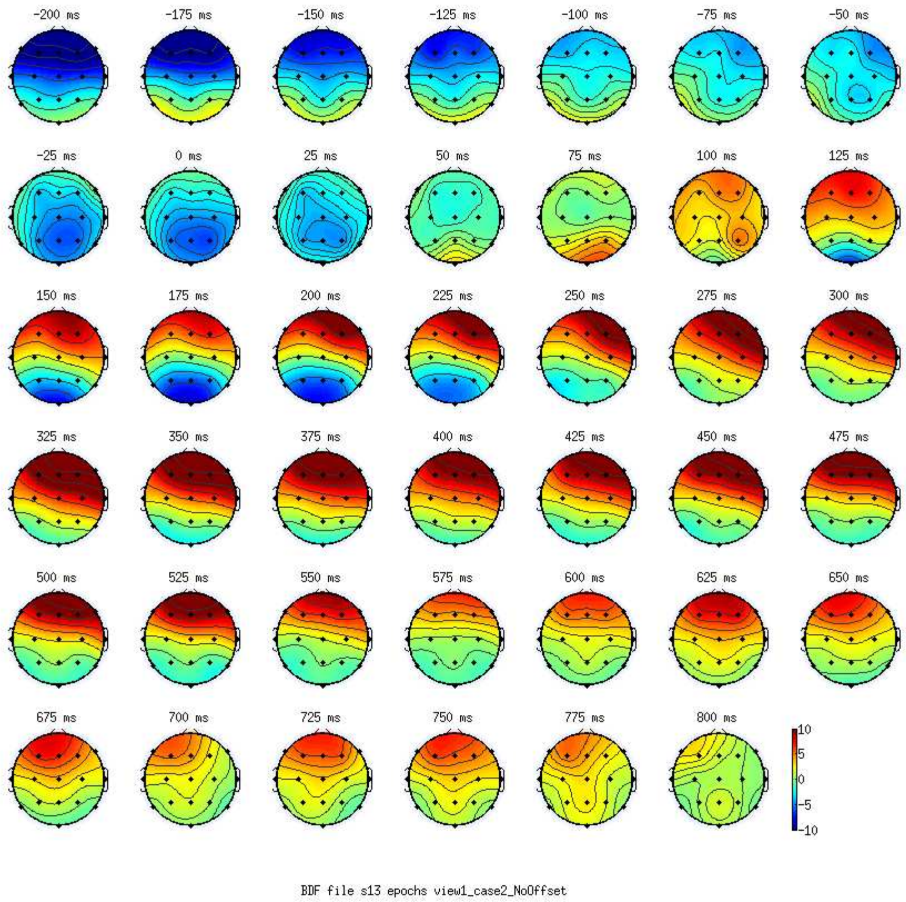

Figure C.108: Subject 8: Averaged scalp plots aligned to fixation onset for non-target frames

Appendix D

Supplemental material for experiments outlined in Chapter 5

In this appendix section we present the supplemental materials to the experiment described in 5.2.

D.1 Ranked images for subjects - Supplement for Section 5.2

Figure D.1: Ranked images from subject 1 prediction scores. First line is top ranked true positives (descending from strongest predictions left to right), second line is true negatives (descending from strongest prediction left to right), third line is false negatives (ascending from worst prediction left to right), and the fourth line in false positives (ascending from worst prediction left to right)

Figure D.2: Ranked images from subject 2 prediction scores. First line is top ranked true positives (descending from strongest predictions left to right), second line is true negatives (descending from strongest prediction left to right), third line is false negatives (ascending from worst prediction left to right), and the fourth line in false positives (ascending from worst prediction left to right)

Figure D.3: Ranked images from subject 3 prediction scores. First line is top ranked true positives (descending from strongest predictions left to right), second line is true negatives (descending from strongest prediction left to right), third line is false negatives (ascending from worst prediction left to right), and the fourth line in false positives (ascending from worst prediction left to right)

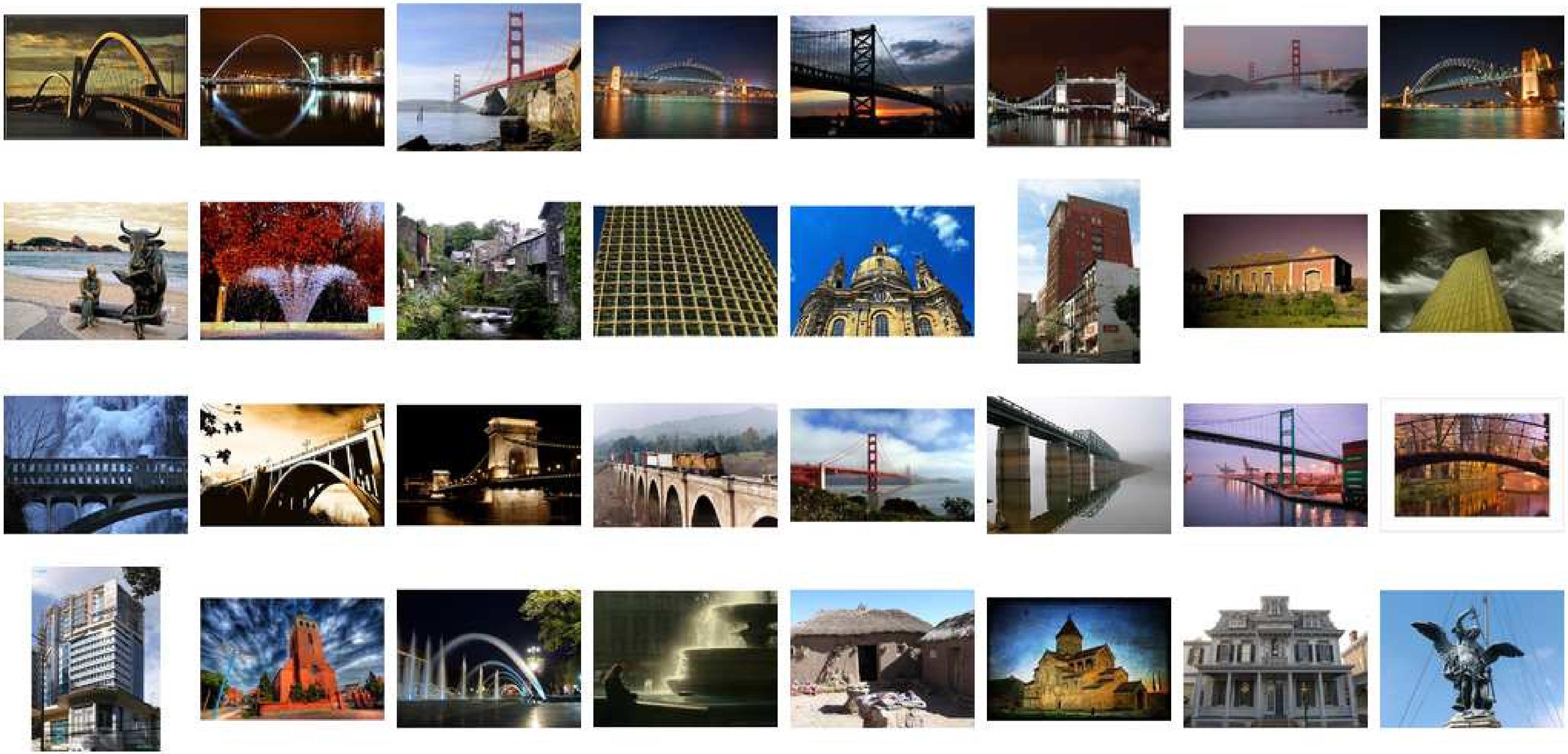

Figure D.4: Ranked images from subject 4 prediction scores. First line is top ranked true positives (descending from strongest predictions left to right), second line is true negatives (descending from strongest prediction left to right), third line is false negatives (ascending from worst prediction left to right), and the fourth line in false positives (ascending from worst prediction left to right)

Figure D.5: Ranked images from subject 5 prediction scores. First line is top ranked true positives (descending from strongest predictions left to right), second line is true negatives (descending from strongest prediction left to right), third line is false negatives (ascending from worst prediction left to right), and the fourth line in false positives (ascending from worst prediction left to right)

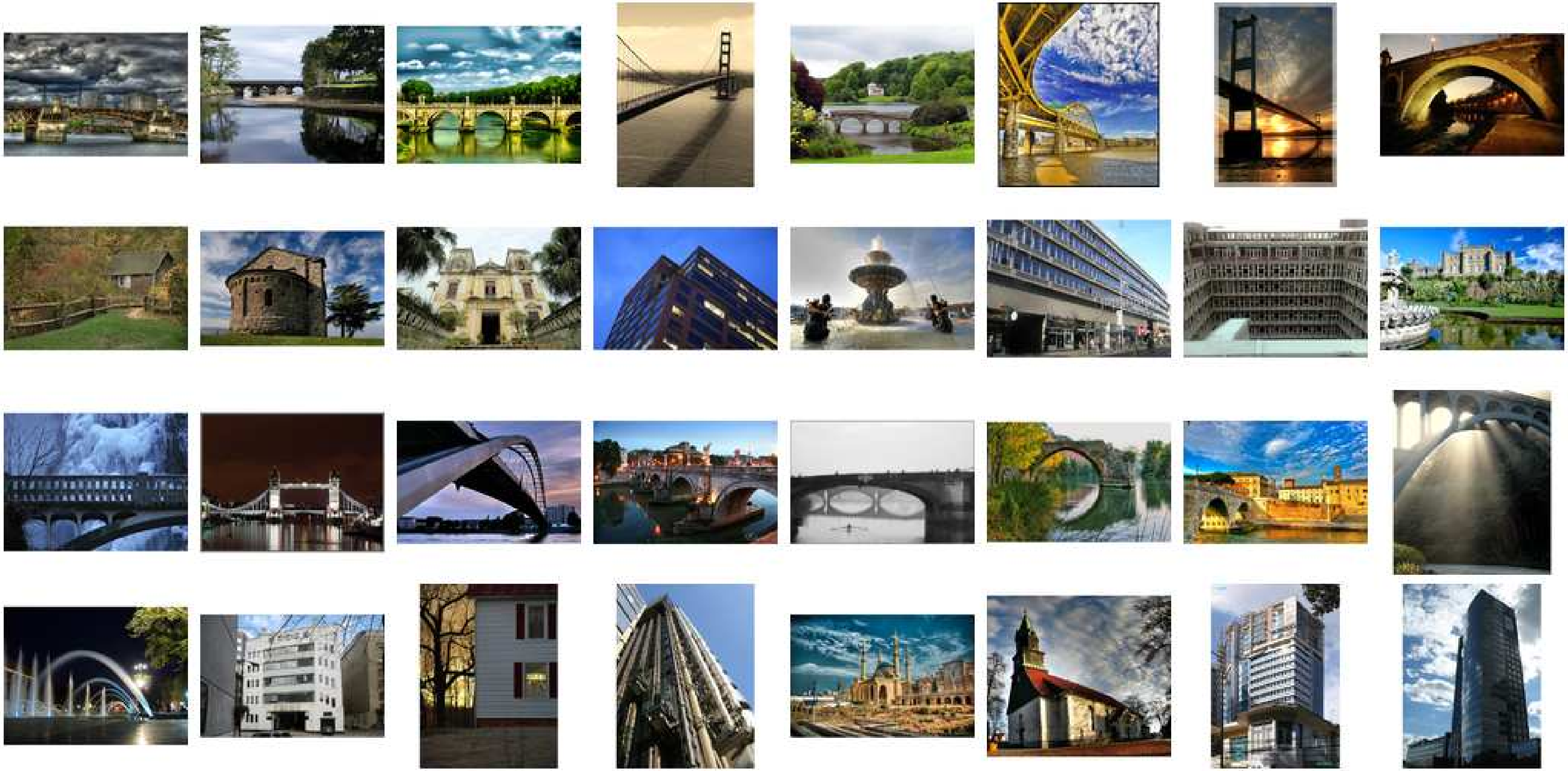

Figure D.6: Ranked images from subject 6 prediction scores. First line is top ranked true positives (descending from strongest predictions left to right), second line is true negatives (descending from strongest prediction left to right), third line is false negatives (ascending from worst prediction left to right), and the fourth line in false positives (ascending from worst prediction left to right)

Figure D.7: Ranked images from subject 7 prediction scores. First line is top ranked true positives (descending from strongest predictions left to right), second line is true negatives (descending from strongest prediction left to right), third line is false negatives (ascending from worst prediction left to right), and the fourth line in false positives (ascending from worst prediction left to right)

Appendix E

Miscellaneous materials

Dublin City University
Ollscoil Chathair Bhaile Átha Cliath

DCU

Prof. Alan Smeaton
CLARITY

19th August 2010

REC Reference: DCUREC/2010/066

Proposal Title: **Using brainwaves detected through EEG to label images as relevant or irrelevant**

Applicants: Prof. Alan Smeaton, Mr. Graham Healy

Dear Alan,

Further to expedited review, this research proposal is approved. Should substantial modifications to the research protocol be required at a later stage, a further submission should be made to the REC.

Yours sincerely,

pp Fiona Brennan

Dr. Donal O'Mathuna
Chair
DCU Research Ethics Committee

DCU
Office of the Vice-President
for Research

Office of the Vice-President
for Research
Dublin City University,
Dublin 9, Ireland

T +353 1 700 8000
F +353 1 700 8002
E research@dcu.ie
www.dcu.ie

Figure E.1: University ethics approval

Publications

Healy, Graham and Smeaton, Alan (2011) Eye fixation related potentials in a target search task. *In: 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 11), Aug 30 - Sept 3, 2011, Boston, USA.*

Healy, Graham and Smeaton, Alan (2011) Optimising the number of channels in EEG-augmented image search. *In: The 25th BCS Conference on Human-Computer Interaction (HCI), 4th - 8th July 2011, Newcastle-upon-Tyne, UK.*

Smeaton, Alan F. and Wilkins, Peter and Healy, Graham and Ampatzis, Christos and Rusinski, M. and Izzo, Dario (2009) Neurological modeling of what experts vs. non-experts find interesting. *In: Neuroscience 2009, 17-21 October 2009, Chicago, USA.*

Healy, Graham and Smeaton, Alan F. (2009) An outdoor spatially-aware audio playback platform exemplified by a virtual zoo. *In: ACM Multimedia 2009, 19-24 October 2009, Beijing, China.*

Gaffney, Mark and O'Flynn, Brendan and Mathewson, A. and Buckley, John and Barton, John and Angove, Philip and Vcelak, J. and O Conaire, Ciaran and Healy, Graham and Moran, Kieran and O'Connor, Noel E. and Coyle, Shirley and Kelly, Philip and Caulfield, Brian and Conroy, Luke (2009) Wearable wireless inertial measurement for sports applications. *In: IMAPS-CPMT - 33rd International Microelectronics and Packaging Poland Conference, 21-24 September, 2009, Gliwice - Pszczyna, Poland.*

Conroy, Luke and O Conaire, Ciaran and Coyle, Shirley and Healy, Graham and Kelly, Philip and Connaghan, Damien and O'Connor, Noel E. and Smeaton, Alan F. and Caulfield, Brian and Nixon, Paddy (2009) TennisSense: a multi-sensory approach to performance analysis in tennis. In: 27th International Society of Biomechanics in Sports Conference, 17-21 August 2009, Limerick, Ireland.

Izzo, Dario and Rossini, Luca and Rucinski, Marek and Ampatzis, Christos and Healy, Graham and Wilkins, Peter and Smeaton, Alan F. and Yazdani, Ashkan and Ebrahimi, Touradj (2009) Curiosity cloning: neural analysis of scientific knowledge. In: IJCAI-09 - International Joint Conference on Artificial Intelligence 2009, Workshop on Artificial Intelligence in Space, 17-18 July 2009, Padadena, California, CA, USA. (ESA SP-673).

Healy, Graham and Smeaton, Alan F. (2009) Spatially augmented audio delivery: applications of spatial sound awareness in sensor-equipped indoor environments. In: ISA 2009: First International Workshop on Indoor Spatial Awareness, 18 May 2009, Taipei, Taiwan. ISBN 978-1-4244-4153-2

Bibliography

Baccino, T. and Manunta, Y. (2005). Eye-fixation-related potentials: Insight into parafoveal processing. *Journal of Psychophysiology*, 19(3):204 – 215.

Bashashati, A., Fatourechi, M., Ward, R. K., and Birch, G. E. (2007). A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *Journal of Neural Engineering*, 4(2):R32–R57.

Berger, H. (1929). Ueber das electroencephalogram des menschen. *Archive fur Psychiatrie und Nervenkrankheitin*, 87:527–570.

Bigdely-Shamlo, N., Vankov, A., Ramirez, R. R., and Makeig, S. (2008). Brain activity-based image classification from rapid serial visual presentation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 16:432–441.

Blankertz, B., Lemm, S., Treder, M., Haufe, S., and Mller, K.-R. (2011). Single-trial analysis and classification of ERP components - a tutorial. *NeuroImage*, 56(2):814–825.

Blankertz, B., Tangermann, M., Vidaurre, C., Fazli, S., Sannelli, C., Haufe, S., Maeder, C., Ramsey, L. E., Sturm, I., Curio, G., and Mueller, K. R. (2010). The Berlin brain-computer interface: Non-medical uses of BCI technology. *Frontiers in Neuroscience*, 4(0).

Blau, V. C., Maurer, U., Tottenham, N., and McCandliss, B. D. (2007). The face-specific n170 component is modulated by emotional facial expression. *Behavioral and Brain Functions*, 3(7):7.

Bulling, A. and Gellersen, H. (2010). Toward mobile eye-based human-computer interaction. *Pervasive Computing, IEEE*, 9(4):8 –12.

Chang, C.-C. and Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27.

Clay, L. P., Parra, L. C., Spence, C. D., Adam, and Paul Sajda, C. (2005). Recipes for the linear analysis of EEG. *NeuroImage*, 28:326–341.

Comerchero, M. D. and Polich, J. (1999). P3a and P3b from typical auditory and visual stimuli. *Clinical Neurophysiology*, 110(1):24 – 30.

Ehinger, K. A., Hidalgo-Sotelo, B., Torralba, A., and Oliva, A. (2009). Modelling search for people in 900 scenes: A combined source model of eye guidance. *Visual Cognition*, 17:945–978.

Ericsson, K. A. and Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, 47(1):273–305.

Farwell, L. and Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 70(6):510 – 523.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861 – 874. ROC Analysis in Pattern Recognition.

Gerson, A., Parra, L., and Sajda, P. (2006). Cortically coupled computer vision for rapid image search. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 14(2):174 –179.

Gerson, A. D., Parra, L. C., and Sajda, P. (2005). Cortical origins of response time variability during rapid discrimination of visual objects. *Neuroimage*, 28:342–353.

Geusebroek, J. M., Burghouts, G. J., and Smeulders, A. W. M. (2005). The amsterdam library of object images. *International Journal of Computer Vision*, 61(1):103–112.

Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. *J. Mach. Learn. Res.*, 3:1157–1182.

Healy, G. and Smeaton, A. F. (2011a). Eye fixation related potentials in a target search task. In *Engineering in Medicine and Biology Society, 2011. EMBC 2011. Annual International Conference of the IEEE*.

Healy, G. and Smeaton, A. F. (2011b). Optimising the number of channels in eeg-augmented image search. In *The 25th BCS Conference on Human Computer Interaction*.

Huang, Y., Erdogmus, D., Mathan, S., and Pavel, M. (2007). A fusion approach for image triage using single trial erp detection. In *Neural Engineering, 2007. CNE '07. 3rd International IEEE/EMBS Conference on*, pages 473 –476.

Huang, Y., Erdogmus, D., Pavel, M., Mathan, S., and Hild, II, K. E. (2011). A framework for rapid visual image search using single-trial brain evoked responses. *Neurocomput.*, 74:2041–2051.

Izzo, D., Rossini, L., Rucinski, M., Ampatzis, C., Healy, G., Wilkins, P., Smeaton, A., Yazdani, A., and Ebrahimi, T. (2009a). Curiosity Cloning: Neural Analysis of Scientific Knowledge. In *Proceedings of the International Joint Conference on Artificial Intelligence 2009, Workshop on Artificial Intelligence in Space*, California, USA.

Izzo, D., Rossini, L., Rucinski, M., Ampatzis, C., Healy, G., Wilkins, P., Smeaton, A., Yazdani, A., and Ebrahimi, T. (2009b). Curiosity Cloning: Neural Analysis of Scientific Knowledge. In *Proceedings of the International Joint Conference on Artificial Intelligence 2009, Workshop on Artificial Intelligence in Space*, California, USA.

Johansson, M. and Mecklinger, A. (2003). The late posterior negativity in erp studies of episodic memory: action monitoring and retrieval of attribute conjunctions. *Biological Psychology*, 64(1-2):91 – 117. Information Processing and Error Analysis: Retrospectives on a Career.

Johnson, J. S. and Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. *Journal of Vision*, 3(7):499–512.

Jung, T.-P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., and Sejnowski, T. J. (2001). Analysis and visualization of single-trial event-related potentials. *Human Brain Mapping*, 14:166–185.

Kapoor, A., Shenoy, P., and Tan, D. (2008). Combining brain computer interfaces with vision for object categorization. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 0:1–8.

Kazai, K. and Yagi, A. (2003). Comparison between the lambda response of eye-fixation-related potentials and the P100 component of pattern-reversal visual evoked potentials. *Cognitive, Affective, and Behavioural Neuroscience*, 3(1):46–56.

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., and Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of neural engineering*, 4(2).

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique (Cognitive Neuroscience)*. The MIT Press, 1 edition.

Luck, S. J. and Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*, 31(3):291–308.

Luck, S. J., Vogel, E. K., and Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature*, 383(6601):616–618.

Makeig, S., Delorme, A., Westerfield, M., Jung, T.-P., Townsend, J., Courchesne, E., and Sejnowski, T. J. (2004). Electroencephalographic brain dynamics following manually responded visual targets. *PLoS Biol*, 2(6):e176.

Makeig, S., Westerfield, M., Townsend, J., Jung, T.-P., Courchesne, E., and Sejnowski, T. J. (1999). Functionally independent components of early event-related potentials in a visual spatial attention task. *Philosophical Transactions of The Royal Society B: Biological Sciences*, 354:1135–1144.

McCarley, J. S., Kramer, A. F., Wickens, C. D., Vidoni, E. D., and Boot, W. R. (2004). Visual skills in airport-security screening. *Psychological Science*, 15(5):pp. 302–306.

Meijer, E. H., Smulders, F. T., Merckelbach, H. L., and Wolf, A. G. (2007). The p300 is sensitive to concealed face recognition. *International Journal of Psychophysiology*, 66(3):231 – 237.

Miyakoshi, M., Nomura, M., and Ohira, H. (2007). An erp study on self-relevant object recognition. *Brain and Cognition*, 63(2):182–189.

O'Donnell, B. F., Swearer, J. M., Smith, L. T., Hokama, H., and McCarley, R. W. (1997). A topographic study of ERPs elicited by visual feature discrimination. *Brain Topography*, 10:133–143. 10.1023/A:1022203811678.

Ogawa, K., Nittono, H., and Hori, T. (2005). Brain potentials before and after rapid eye movements: an electrophysiological approach to dreaming in REM sleep. *Sleep (Rochester)*, 28(9):1077–1082.

Oliva, A. and Torralba, A. (2006). Building the gist of a scene : the role of global image features in recognition. *Brain*, 155(1):23–36.

Olofsson, J. K., Nordin, S., Sequeira, H., and Polich, J. (2008). Affective picture processing: an integrative review of ERP findings. *Biological Psychology*, 77(3):247–265.

Philiastides, M. G., Ratcliff, R., and Sajda, P. (2006). Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram. *J Neurosci*, 26(35):8965–8975.

Pineda, J., Allison, B., and Vankov, A. (2000). The effects of self-movement, observation, and imagination on mu; rhythms and readiness potentials (RPs): toward a brain-computer interface (BCI). *Rehabilitation Engineering, IEEE Transactions on*, 8(2):219 – 222.

Pohlmeyer, E. A., Wang, J., Jangraw, D. C., Lou, B., Chang, S.-F., and Sajda, P. (2011a). Closing the loop in cortically-coupled computer vision: a brain-computer interface for searching image databases. *Journal of Neural Engineering*, 8.

Pohlmeyer, E. A., Wang, J., Jangraw, D. C., Lou, B., Chang, S.-F., and Sajda, P. (2011b). Closing the loop in cortically-coupled computer vision: a braincomputer interface for searching image databases. *Journal of Neural Engineering*, 8(3):036025.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118(10):2128–2148.

Poolman, P., Frank, R. M., Luu, P., Pederson, S. M., and Tucker, D. M. (2008). A single-trial analytic framework for eeg analysis and its application to target detection and classification. *NeuroImage*, 42(2):787 – 798.

Qian, M., Aguilar, M., Zachery, K. N., Privitera, C., Klein, S., Carney, T., and Nolte, L. W. (2009). Decision-level fusion of EEG and pupil features for single-trial visual detection analysis. *IEEE Transactions on Biomedical Engineering*, 56:1929–1937.

Rama, P. and Baccino, T. (2010). Eye fixation-related potentials (efrps) during object identification. *Visual Neuroscience*, 27(5-6):187–192.

Sajda, P., Philiastides, M., and Parra, L. (2009). Single-trial analysis of neuroimaging data: Inferring neural networks underlying perceptual decision-making in the human brain. *Biomedical Engineering, IEEE Reviews in*, 2:97 –109.

Sajda, P., Pohlmeyer, E., Wang, J., Parra, L. C., Christoforou, C., Dmochowski, J., Hanna, B., Bahlmann, C., Singh, M. K., and Chang, S.-F. (2010). In a blink of an eye and a switch of a transistor: Cortically coupled computer vision. *Proceedings of the IEEE*, 98:462–478.

Shapiro, K. L., Raymond, J. E., and Arnell, K. (2009). Attentional blink. *Trends in Cognitive Sciences*, 1(8):291–296.

Shenoy, P. and Tan, D. S. (2008). Human-aided computing: utilizing implicit human processing to classify images. In *Computer Human Interaction*, pages 845–854.

Sireteanu, R. and Rettenbach, R. (2000). Perceptual learning in visual search generalizes over tasks, locations, and eyes. *Vision Research*, 40(21):2925 – 2949.

Somol, P., Pudil, P., Novovičová, J., and Paclík, P. (1999). Adaptive floating search methods in feature selection. *Pattern Recogn. Lett.*, 20:1157–1163.

Steffensen, S. C., Ohran, A. J., Shipp, D. N., Hales, K., Stobbs, S. H., and Fleming, D. E. (2008). Gender-selective effects of the P300 and N400 components of the visual evoked potential. *Vision Research*, 48(7):917–925.

Sutton, S., Braren, M., Zubin, J., and John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, 150:1187–1188.

Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381(6582):520–522.

Torralba, A., Oliva, A., Castelhano, M. S., and Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, 113:766–786.

Treder, M. S. and Blankertz, B. (2010). (C)overt attention and visual speller design in an ERP-based brain-computer interface. *Behavioral and brain functions BBF*, 6(1):28.

Treisman, A. M. and Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136.

Vidal, J. J. (1973). Toward Direct Brain-Computer Communication. *Annual Review of Biophysics and Bioengineering*, 2(1):157–180.

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., and Torralba, A. (2010). Sun database: Large-scale scene recognition from abbey to zoo. In *CVPR'10*, pages 3485–3492.