Face Perception and Recognition, on the Fringe of Human Awareness

PhD THESIS
submitted to the
UNIVERSITY of KENT at CANTERBURY
in the subject of Computer Science, for the Degree of
DOCTOR of PHILOSOPHY

Defended by

Omid HAJILOU

January 2020

Abstract

In Rapid Serial Visual Presentation (RSVP), we can present a large volume of information, on the fringe of awareness, whilst observing the brain's electrical signals using an Electroencephalogram (EEG). The vast majority of stimuli are not consciously perceived, but the salient ones breakthrough into awareness, enter into working memory and can be reported by the participant (Bowman, et al., 2013). Deception detection studies have successfully employed this countermeasure resistant *fringe-P3* method, using letters, numbers and words (Bowman, Filetti, Alsufyani, Janssen, & Su, 2014), to differentiate between familiar and unfamiliar information. The inclusion of faces, and their application in Concealed Information Tests (CIT) have yet to be fully explored.

In this thesis, we hypothesised that the *fringe-P3* method could be successfully used to detect intrinsic salience of familiar faces, even when there was no task associated with the stimuli. Using experiments, we investigated the sensitivity of the ERP-based RSVP paradigm, to infer recognition of celebrity, as well as, lecturer faces, and performed statistical tests in the Time and Frequency domains, to differentiate between known and unknown faces, at group and subject levels. Furthermore, we used ground-truth data simulations to explore the viability of using online statistical tests, to focus experimental data collection efforts, on the critical stimulus with the highest significance, in order to improve statistical power (i.e. reduce the risk of Type II errors), without the inflation of Type I errors.

As a result, we introduced new methods of analysis, and a two-part experimental design, where Part II's parameters are independently influenced by Part I's results, using online statistical tests. Finally, we applied our new findings in a concluding experiment, which explored a real-life scenario of revealing participants' familiarity with their lecturers, through the data captured from their brain. Our findings provide evidence that familiar faces are differentially perceived and processed by participants' brains, as compared to novel (unfamiliar) faces. Therefore, we propose our final experiment to be a workable solution for deception detection applications of crime compatriots (e.g. accomplices), using faces in RSVP-based EEG tests.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to express my deepest gratitude and sincerest thanks to my PhD supervisor, Professor Howard Bowman, who guided and supported me, at every step of this journey. His wisdom, attention to detail and selfless care were simply invaluable. I appreciate the contribution of the members of my review panel, who provided helpful comments and support. My thanks to Dr. Alsufyani, who contributed to the first experiment, as our fields of study overlapped. A postumous thanks to my learned friend and mentor, Michael Kaufman, who always had my back.

Finally, I am indebted to my loving wife, *Niloufar*, and our children, beautiful *Yasmin* & perfect *Jubin*, who inspired me to pursue this amazing adventure in cognitive neuroscience, and for not letting me forget that *it's never too late to seek a newer world*.

I dedicate this PhD thesis to my wonderful family.

P A R T I

–

B A C K G R O U N D

Chapter 1:
Introduction

The aim of this chapter is three-fold: firstly, to introduce the reader to the topic and the concepts that are relevant to the field of research; secondly, to consider the scientific questions, objectives and motivation for the research; and finally, to outline the scope of the research, with the aid of a description of each chapter. To that end, the first section (1.1) will provide an overview of the broader research territory and background information. The second section (1.2) will outline the gap that this research will fill, and introduce the central hypotheses and objectives. The next section (1.3) will provide a road-map of the thesis, by summarising the contents of all chapters, and the last section (1.4) will review the relevant collaborations and publications.

1.1 Overview

In this thesis, we will introduce image-based stimuli to the existing fringe-P3 deception detection studies, and utilise Rapid Serial Visual Presentation (RSVP), to study the instance of perceiving *sub/liminal* faces. Further, we will design Electroencephalogram (EEG) experiments that can explore the sensitivity of our ERP-based RSVP paradigm to infer recognition of broadly familiar faces, and introduce analysis methods, which can improve the statistical power of detecting an effect.

The RSVP technique enables us to present information at a very high speed, whilst observing the brain's electrical signals using an EEG. After time-locking and averaging the EEG signals, we will analyse the Event Related Potentials (ERPs), to identify salient components which break through into consciousness, such as the P3 component (Craston, Wyble & Bowman, 2006), and the face related ERP modulations: N400f and P600f (Eimer, 2000).

By exploring implicit perception of salient stimuli within image-based RSVP streams, we will introduce a new dimension to the existing words/names fringe-P3

lie-detection studies (Bowman, Filetti, Alsufyani, Janssen, & Su, 2014). This new feature will further the real-life application of deception detection, since the ability to recognise faces, amongst humans, is an evolutionary based socio-cognitive skill that is considered to be emotionally evocative. Furthermore, the subject of an investigation (e.g. the accused) may not be able to read, or know the name of the other participants in the crime (e.g. the compatriots). To that end, we will explore and develop new methods of designing RSVP-based EEG experiments, as well as, improved techniques for analysing EEG data, to advance future face detection and recognition applications.

1.1.1 – Human Brain

Anatomically, human brains appear to differ very little from their Palaeolithic ancestors who lived more than 30,000 years ago, in the Ardèche region of southern France, as evidenced by their "cognitive ability to create art separate from the body", in the form of paintings on the walls of the Chauvet cave (Morriss-Kay, 2010). Throughout this period, recognition of faces has been an important neurological mechanism for societal interactions. Furthermore, the ability to extract information within milliseconds of viewing a face may have played a critical role in survival (e.g. to deal with major threats, or reproductive opportunities).

Recent studies (Kanwisher & Yovel, 2006) suggest that the area of the brain that discriminates between familiar objects (i.e. the fusiform gyrus) contains a specialised region, called the fusiform face area (FFA), which plays a key role in face perception. Mankind's innate introspection and technical ingenuity, over thirty millennia, has failed to reveal all the secrets of the human brain. Even though we are still struggling to understand how the 'mind' emerges from the brain, advances in brain imaging technology, and increases in research funding, have accelerated the discoveries about the human brain. With around 86 billion neuros in the brain (Azevedo, et al., 2009), each cubic millimetre of cerebral cortex can contain approximately one billion connections, since each neuron is able to process information based on as many as 10,000 neuronal inter-connections (Alonso-Nanclares, Gonzalez-Soriano, Rodriguez, & DeFelipe, 2008). The complex and intricate nature of the brain's structure, as well as,

the fleeting nature of information being processed, makes the task of non-invasive brain imaging a challenging prospect.

1.1.2 – Brain Imaging

Recent advances in imaging technologies have enabled researchers to improve the analysis of brain functions. As such, functional Magnetic Resonance Imaging (fMRI) can accurately measure changes in blood flow within the brian's blood vessels, providing scans with three spatial dimensions, and a granularity of cubic millimeters (Friston, et al., 1998). However, fMRI's temporal resolution is low (i.e. in the range of seconds), and the cost of each test can be prohibitively high. Alternatively, Magnetoencephalography (MEG) benefits from much higher temporal resolution, by measuring microscopic changes in magnetic field, caused directly by neural activity. Albeit, unlike fMRI, MEG's spatial resolution is relatively low (Malmivuo, Suihko, & Eskola, 1997).

Our chosen brain imaging technology, Electroencephalography (EEG), is cheaper and more practical than the others, but EEG does not have a spatial resolution that is as high as fMRI or MEG, since it is limited by the distortions to the measured electrical field, created by the skull. However, it benefits from a high temporal resolution, as electrodes register combined activity of large groups of neurons (Makeig, Debener, Onton, & Delorme, 2004b), in the form of electric potentials released during neuronal activation. In addition to genuine brain activity, electric potentials that are captured by EEG's electrodes are affected by skin conductivity, muscular movement/tissue and ambient interference (Malmivuo, Suihko, & Eskola, 1997), which is colloquially referred to as 'noise'. Fortunately, noise reduction techniques, such as Event-Related Potential (ERP) averaging enables us to replicate the same event multiple times, and observe/measure the resultant brain activity, like the P3 (or P-300) component (Sutton, Braren, Zubin, & John, 1965), which happens to be a feature of interest, in this thesis. If a component of interest occurs consistently across replications, it will survive averaging, while noise, which should not be consistent, will cancel out.

1.1.3 – Perception

In each nanosecond, humans can encounter billions of sensations, and a large portion of the sensory data can be processed by the visual system. Whereas the majority of the data can be ignored by the brain, the ones that are recognised and processed form part of our perception. Studies have shown that the brain uses *visual search* to select the most relevant features from the data and ignore the irrelevant ones, as it allocates attention and/or resources (Luck, Hillyard, Mouloua, & Hawkins, 1996). The brain's perceptual system is constantly performing visual searches, using top-down or bottom-up mechanisms (Melloni, Van-Leeuwen, Alink, & Mueller, 2012), in order to direct attention towards salient information. Whilst the top-down mechanism is goal-orientated and signifies deliberate allocation of attention (e.g. looking for your own face in a group photo), the bottom-up mechanism can be stimulus-driven, as attention is directed automatically to salient information (e.g. unexpected encountering of your face in a random album, which grabs your attention due to the high saliency of your own face).

The three experiments in this thesis have been designed to explore the stimulus-driven mechanisms of visual attention, by comparing the difference between the intrinsic salience of two conditions: unexpected familiar faces and unknown faces. These conditions are presented equally as often and statistically in the same position as one another, using a stimulus presentation paradigm, called Rapid Serial Visual Presentation (RSVP).

1.1.4 – RSVP

The RSVP technique allows a series of items to be shown at the same spatial location, at a high presentation rate (Lawrence, 1971), where multiple distractors (i.e. non-target fillers) are interspersed with salient stimuli (e.g. distinct Targets). Numerous studies ((Potter, 1975) (Lawrence, 1971) (Chun & Potter, 1995)) have shown that the detection of a Target stimuli was successful at high presentation rates of 10 or more items per second. In fact, it is the high presentation rate of the RSVP paradigm that

facilitates (what has been argued as) a countermeasure resistant technique for our experiments (Bowman, Filetti, Alsufyani, Janssen, & Su, 2014).

Due to the high stimulus presentation speed of RSVP, participants are prevented from perceiving every single stimulus in the stream, and only the salient stimuli breakthrough into conscious awareness, rendering the distractors/fillers much harder to identify. As such, presenting images on the fringe of awareness takes advantage of the concept called *sub/liminal salience search* (Bowman, et al., 2013), where the majority of images are not perceived at a level which is considered to be sufficient for encoding into working memory. However, salient images that breakthrough into conscious awareness will generate a unique pattern, which (we believe) can be correlated with the P3 ERP components, providing a method called the *fringe-P3*. This is an essential mechanism for our deception detection investigations.

1.1.5 – Deception Detection

The ancient history of lie/deception detection is mired with techniques that employed torture, but at the end of the 19th century, James McKenzie invented a (relatively harmless) lie detector test that could measure human pulse and detect irregular heartbeats (Trovillo, 1939). Further enhancements in the early-20th century would superimpose other physiological responses, such as, blood pressure, respiratory rate, galvanic skin resistance and various reaction times (Larson, 1932). Whilst the operational principles of these machines (a.k.a. polygraph) remain largely unchanged, additional physiological responses have been incorporated, including voice pitch, pupil size, eye blinks and facial skin temperature (Synnott, Dietzel, & Ioannou, 2015). When conducting polygraph tests, the examiner would commonly use one of the following two predominant types of questioning techniques to induce changes in the subject's physiological responses, which could lead to a diagnosis of deception or non-deception:

i. Control Question Test (CQT) compares 'control' questions with 'relevant' questions about the crime, but the subjective decisions made by the examiner may leave room for human error (Lykken D. , 1984).

ii. Improved and more standardised techniques, such as Guilty Knowledge Test (GKT), or Concealed Information Test (CIT), rely on pieces of information that were not disclosed to the public (e.g. a multiple choice test with items that only a guilty subject could know). Subjects who display a selective reaction to the incriminating information could be considered to have been aware of the information, thus, CIT has been promoted as the ideal paradigm for deception detection (Ben-Shakhar & Elaad, 2003).

Whilst the polygraph industry and its practitioners have a vested interest in promoting their systems, scientific opinion of polygraph tests remains conflicted, due to its weakness to mental and/or physical countermeasures (Honts & Kircher, 1984). However, studies (Abootalebi, Moradi, & Khalilzadeh, 2006) have shown that measuring deception at the source of cognition (i.e. the brain) – rather than the nerve endings, as is the case in polygraph tests – would improve the possibility of detecting deception. Due to its affordability and practicality, EEG has been widely used in deception detection studies (Farwell & Donchin, 1991) (Rosenfeld, Angell, Johnson, & Qian, 1991). These studies exploit the assumption that familiar stimuli generate a larger P-300 (P3) than non-familiar ones (Farwell & Donchin, 1986).

1.1.6 – P3 based CIT

In classical oddball experiments, the P3 is a positive ERP component that occurs approximately 300 to 800 milliseconds after the onset of an infrequently-presented target stimulus that appears within a series of frequently-presented non-target stimuli (Fabiani, Karis, Coles, & & Donchin, 1983). Studies have successfully used the P3 component within CITs (Johnson & Rosenfeld, 1992) (Rosenfeld, Soskins, Bosch, & Ryan, 2004) in an investigative context, to detect deception. However, due to the design of such deception detection studies, where the stimuli are presented at a slow rate (typically one per second), there could be a potential for suspects to confound the tests, using countermeasures (Meixner, Haynes, Winograd, Brown, & Rosenfeld, 2009). One technique that could be employed by the suspect is to choose/identify an irrelevant stimulus and imagine a violent act, every time that the irrelevant stimulus appears, in

order to invalidate the difference between the actual guilty-knowledge stimulus and the irrelevant one.

The key reason for the possibility of such behavioural countermeasure techniques is because the subject possesses sufficient time, between each stimulus, to consciously determine that there is a repeating non-critical irrelevant item. However, if the stimuli were presented at a rapid rate, using the RSVP paradigm (e.g. 10 items per second), the irrelevant stimulus will not be sufficiently perceived to be noticed as repeating. Therefore, by presenting items rapidly, and taking advantage of the concept of Sub/liminal Salience Search (Bowman, et al., 2013), the majority of the items are not perceived at a level which is considered to be sufficient for encoding into working memory. However, according to the fringe-P3 method, items that are salient and breakthrough will generate a unique ERP pattern, which, we believe, can be correlated with a P3 ERP component.

1.2 Central Hypotheses

This thesis is a continuation of the previous studies into deception detection, carried out at the Centre for Cognitive Neuroscience & Cognitive Systems (CCNCS), at the University of Kent. Up until 2014, CCNCS had produced significant evidence for the existence of the countermeasure-resistant P3 component in concealed information experiments that employed numbers, letters and words, as stimuli (Bowman, Filetti, Alsufyani, Janssen, & Su, 2014). The question that lies at the core of this thesis is whether a new category of critical stimuli, in the form of human faces, can be introduced to the ERP-based RSVP paradigm? Furthermore, will familiar faces differentially break through into conscious awareness, and can we detect the breakthrough events in EEG?

To begin with, we will explore the suitability of highly familiar and emotionally evocative faces of celebrities. Our first scientific question is: can we detect the group-level breakthrough of Probe (celebrity) faces, which are differentially perceived and processed, as compared to Irrelevant (unfamiliar) faces? Secondly, can we detect the

individual/subject level breakthrough, using statistical analyses in the Time (ERP) domain, as well as, Frequency (ERSP/ITC) domain?

Having successfully established that famous/celebrity faces can breakthrough into conscious awareness, using an RSVP subliminal search paradigm, we will substitute the highly evocative faces of famous celebrities with familiar faces that are personally known to the participants, in the form of the University's lecturers. Subsequently, we pose the question: can we differentiate between the Probe (familiar University of Kent lecturer) and Irrelevant (unknown lecturers from another university), at group and subject levels, using statistical analyses in the Time and Frequency domains?

The results of the above two studies will suggest that we will be able to apply our findings to the differentiation of deceivers and non-deceivers, in the application of crime compatriots, whereby, a suspect's familiarity with a criminal/terrorist can be established using faces. At this point, we chose to pursue ground-truth data simulations, to improve the statistical power of detecting an effect and enhancing the signal-to-noise ratio (SNR). As a result of these methodological explorations, we should be able to justify the design of a novel two-part experiment, in which Part I of the experiment would independently select the critical stimuli for Part II, using online statistical tests to infer the familiar face that achieves the highest significance.

Finally, we will introduce a key change to the instructions given to the subject, in order to modify the covert nature of presenting familiar faces, as prescribed in the previous studies. Thus, we will reveal the possibility of the subject seeing faces that are personally known to them (without being told who these familiar faces could be), so that we can study a real-life scenario, in which the subject/perpetrator who is being questioned about a crime, will know that the purpose of being shown a series of faces is to ascertain his/her familiarity with an accomplice! We believe that, by bringing together all the findings and improvements in this thesis, our final experiment could demonstrate the closest workable solution for deception detection, using faces in RSVP-based EEG tests. Therefore, we propose that our research will be empirically relevant to the application of real-life deception detection of compatriots, and that our findings offer significant evidence and improvements to the existing work in the CCNCS.

1.3 Organisation of document

This thesis is broken down into four parts: Background, Research, Discussion and Appendix. The following outlines the eight chapters that constitute these four parts:

Part I (Background) contains the Introduction and Literature Review. As such, **chapter 1** outlines the foundation of this thesis, and defines the background material related to this work. **Chapter 2** reviews the essential literature that guides us in our work, and comprises the information about the brain and neuroimaging techniques. Additionally, this chapter outlines the techniques used to test and analyse EEG brain signals, in relation to RSVP-based tests, and how they can be applied to deception detection studies.

Part II (Research) contains five chapters that include the three research studies and the pivotal methodological explorations that comprise the main body of this thesis. **Chapter 3** begins by developing our research ideas, and describes the general framework of all three face recognition studies. Thereafter, this chapter will be used as the standard reference point for the three research studies that will follow, in order to avoid unnecessary repetition of our general research aims, concepts, and methods. **Chapter 4** is the first experiment of its kind to use faces in an RSVP-based EEG experiment, to examine whether famous faces can breakthrough into conscious awareness, and that the breakthrough event can be detected by EEG, on a per individual basis. The objective of this (first) experiment is to advance recent EEG-based studies into concealed information tests, using RSVP-based countermeasure-resistant fringe-P3 method, by introducing a new category of critical stimuli, in the form of human faces. The primary aim of this study is to investigate whether faces can be used in an RSVP subliminal search paradigm, and whether famous faces can breakthrough into conscious awareness. Furthermore, can the breakthrough event be detected by EEG, on a per individual basis, through statistical analyses of the ERP data (in the Time domain) and single-trial data (in the Frequency domain), to determine whether the evoked response by the Probe (celebrity) faces were significantly different from that evoked by the Irrelevant (unknown) faces. In **chapter 5**, the primary change is to substitute the highly evocative faces of famous celebrities with familiar faces that are personally known to

the participants, in the form of the University's lecturers. The objective of this (second) experiment is to investigate whether familiar lecturer faces can breakthrough into conscious awareness, as successfully as the first experiment's famous celebrity faces. **Chapter 6** is dedicated to the exploration of data simulations, in order to justify a new experimental design, which would adopt a novel technique of (online) mid-experiment inference, whereby the most significant critical stimulus, from the first-part of the experiment, will be carried-over and re-used in the second part. With the aid of ground-truth data simulations, we will demonstrate that the improved new experimental design does not inflate type I errors, and reduces the risk of type II errors. **In chapter 7**, we will put the new experimental design into practice, in order to advance the use of faces in RSVP-based EEG tests for deception detection applications. Furthermore, to bring this (third and final) experiment in-line with real-life scenarios, in which the perpetrator who is being questioned about a crime will naturally assume that the purpose of being shown a series of faces is to ascertain his/her familiarity with an accomplice, we will inform the subject of the possibility of seeing familiar faces. We consider this format to be the closest workable solution for deception detection applications of compatriots, using faces in RSVP-based EEG tests.

Part III (Discussion) contains a single concluding chapter. In **chapter 8**, we will present all conclusions, describing how we have addressed the central hypotheses, and go on to discuss future potential developments and research into deception detection.

Part IV (Appendix) contains further material and results of statistical tests, which support our research, but were not critical to the findings. Finally, a glossary of common terms and the bibliography concludes this thesis.

1.4 Collaborations and Publications

In addition to comprehensive guidance throughout this thesis from my supervisor, Professor Howard Bowman (HB), the research experiments (chapters 4, 5 and 7) were designed with his input and direction.

I would like to acknoweledge original contribution to the first (celebrity faces) experiment (chapter 4), from Dr. Abdulmajeed Alsufyani (AA), as our fields of study overlapped. Consequently, we designed this experiment together (with guidance from my supervisor, HB), prepared the image stimuli and jointly performed the experiment on fourteen participants. Although we also collaborated on the analysis (leading to publishing a paper, in 2019), all the results presented in this thesis have been independently analysed and interpreted by me, using the newly proposed standards that will be introduced in chapter 3 (e.g. incorporating the detrending technique).

Whilst the other two experiments (chapters 5 and 7) were designed with input from my supervisor (HB), I was solely responsible for the preparation of the bespoke stimuli and for conducting the experiments on all thirty participants. Furthermore, all statistical analyses, experimental findings and related explorations (e.g. data simulations in chapter 6) were conducted by me, and I presented interim results at departmental meetings (e.g. talks at Computational Intelligence Group, University of Kent, 2014/16) and external conferences (e.g. talks at Institute of Psychiatry, King's College London, 2015; and two posters at the British Association for Cognitive Neuroscience, 2016/17).

Alsufyani, Abdulmajeed, **Hajilou, Omid**, Zoumpoulaki, Alexia, Filetti, Marco, Solomon, Christopher J., Gibson, Stuart J., Alroobaea, Roobaea, Bowman, Howard (2019) Breakthrough Percepts of Famous Faces. Psychophysiology, 56 (1). Article Number 13279. ISSN 0048-5772. (doi:10.1111/psyp.13279) (KAR id:68555)

Chapter 2:
Literature Review

The aim of this chapter is to provide the key background information that will be referenced and used as the foundation, throughout this thesis. We will begin by introducing the reader to the history of neuroscience, and outline the brain's structure and functions. Next, we will describe the applicable brain imaging tools, which have been used to measure brain activity, and outline the techniques that enable researchers to interpret neuronal responses. Finally, we will introduce the concept of deception detection, and describe the relevant approaches that will be used in our research, in order to establish the foundations that support this thesis.

2.1 History of neuroscience

According to Herodotus – the ancient Greek historian, who lived around 484 to 425 BC – early Egyptians tended to dismiss the importance of the brain (Immerwahr, 1985), and instead considered the heart as the seat of intelligence. Indeed, in preparation for mummification, "as much of the brain as possible [was extracted] with an iron hook". Over the ensuing 5,000 years, the Egyptians' misconception that the brain is merely "cranial stuffing" proved hard to shake off. Indeed, even Aristotle – the Greek philosopher and scientist, who lived 384 to 322 BC – favoured the heart as the most important organ, believing that the brain and the lungs existed merely to cool the blood and cushion the heart. However, other notable figures, like Hippocrates – the Greek physician, known as the father of western medicine, who lived around 460 to 370 BC – began to recognise the importance of the brain, and his followers were the first to identify the brain as the locus for speech, consciousness and emotions (Pearce, 2016).

The most notable advance in the understanding of the brain and spinal cord, came about during the last half of the second century (AD), when Galen – the leading physician of the Roman empire, who lived around 129 to 216 AD – concluded that, "the brain controlled cognition and willed action" (Freeman, 1994). As the site of

termination of all five senses, Galen asserted that common sense, memory and knowledge were all functions of the brain; a doctrine that survives to the present day. Although Galen's teachings were widely known throughout the Middle-ages, no significant advance in the understanding of the human brain has been recorded during the ensuing 1,400 years. Thus, a status quo remained until the Renaissance (i.e. a period of "rebirth" and enlightenment in European history, that started around 1350), when dissections of human cadavers and anatomical studies resumed.

The French mathematician and philosopher, Rene Descartes (1596 – 1650), believed that the mind and body are separate (i.e. the dualistic theory), communicating via the brain's pineal gland (Skirry, 2014). As the father of modern neuroscience who coined the phrase "Cogito ergo sum" (*I think, therefore I am*), his early study of the human nervous system would eventually lead to Luigi Galvani's (1737 to 1798) discovery of involuntary muscle contraction, as a result of static electricity coming into contact with the nerve cell of a dead frog's leg (Bresadola, 1998). Since then, most major advances in the understanding of the nervous function have been, broadly, as a result of improved detection of electromagnetic signals, and better analysis techniques.

2.2 The nervous system

The human Nervous system enables us to move and communicate with our environment, and is subdivided into the Central Nervous System (CNS) and the Peripheral Nervous System (PNS). The CNS, which is the centre for processing information, consists of the brain (enclosed in the cranium) and the spinal cord. The PNS consists mainly of nerves that connect the CNS to every part of the body, and controls the autonomic nervous system (i.e. the unconscious control system, which regulates the body's involuntary/vital systems, like heart rate, digestion and respiratory rate). Our Brain controls all of our body's functions, without us giving it a thought. It is the most complex structure we know of in the universe, where everything that makes us human is contained. However, how the 'mind' emerges from the brain's complexity remains debated, and the interactions between the different parts of the brain are yet to be fully understood.

2.2.1 – Brain and its Neurons

The human brain weighs about 1.5Kg and is the size of a small melon. It is composed of approximately 86.1 billion special nerve cells, called "neurons" (Azevedo, et al., 2009), which, incidently, is approximately the same number of trees in the Amazon rain forest. Furthermore, the number of interconnections between individual neurons can be as high as 10,000, resulting in about one billion connections in each cubic millimeter (Alonso-Nanclares, Gonzalez-Soriano, Rodriguez, & DeFelipe, 2008).

As the basic working unit of the brain, neurons communicate through a space/gap (in the order of 20 nanometres, on average) called a 'synapse', using chemical neurotransmitters. When a postsynaptic neuron (i.e. a downstream nerve cell) receives the neurotransmitter signal, it converts it into a small electrical signal (Hodgkin & Huxley, 1952), called a Post Synaptic Potential (PSP). This PSP provides a mechanism for the electrical signal to propagate down the dendrite, to the cell body. If enough of these PSPs occur in similarly aligned neurons and in synchrony, an observable electrical field is generated. Thus, as a by-product of the electrochemical processes that are used by neurons for signalling, our brain tissue generates electrical fields that are large enough to be detected outside the skull. It is possible to categorise the rhythmic and non-rhythmic electrical activity that can be detected from the outer layer of the brain's neural tissues (known as the 'cerebral cortex'), into Frequency bands, measured in Hertz (cycles per second). However, the story of how the above brain oscillations relate to cognition is actually much more complex and, as yet, inconclusive.

2.2.2 – Structure of the Brain

The brain is broadly split into four main structural divisions: the *brain stem* (connecting to the spinal cord, regulating basic body functions, like breathing, heart rate and blood pressure), *cerebellum* (regulates movements, balance and equilibrium), *diencephalon* (interior of the brain, relaying sensory information and controling

autonomic functions) and *cerebrum* (regulating human thought, language, consciousness, etc.). The outer layer of the cerebrum is called the cerebral cortex, and it is split into two hemispheres, separated by the longitudinal fissure in the centre of the cerebrum, and joined together by a thick nerve tract, called the corpus callosum. The human cerebral cortex is folded in a way that allows a large surface area to fit within the confines of the cranium; as such, each bump is known as a gyrus and each groove is known as a sulcus.

Whilst only 2 to 4 millimeters thick, the cerebral cortex makes up about 40% of the brain's mass and contains about 20% of the total number of neurons (i.e. approx. 14 to 16 billion). Along the larger gyri and sulci, the cortex can be divided into four sections (see Figure 2.1): the Frontal lobe (associated with reasoning, motor skills and higher level cognition), the Parietal lobe (associated with processing tactile sensory information, pain and touch), the Temporal lobe (associated with processing sounds, languages and memory) and the Occipital lobe (associated with interpreting visual stimuli and information).

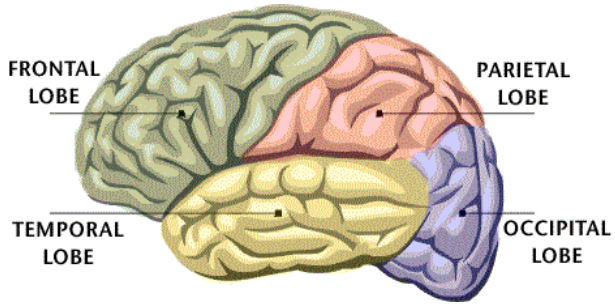

Figure 2.1 – Lobes of the cerebral cortex. Reproduced from Wikimedia.org

A more precise definition of the cortex can be achieved by using the Brodmann areas (Brodmann, 1909), which is a partitioning of the brain (i.e. 52 regions), based solely on the cellular and layer structures (i.e. cytoarchitecture or histological organisation) of the enclosed neurons. Brodmann observed such structures in the cerebral cortex using the Nissl method of cell staining, and published maps of the cortical areas in mammalian cortex. Whilst remaining relevant, and being widely cited for over a century, most brain functions are now seen as being mediated by several distributed systems, which result in functional overlap in multiple areas

(Shepherd, 1988). Furthermore, actual boundaries of Brodmann areas in any individual brain can vary, so without histological examination, we can only approximate the localisation of brain activities.

Whilst it is possible that some Brodmann areas can perform specific functions (e.g. Area 4, which is associated with the motor cortex), most brain functions are mediated by distributed systems, with functional overlap in several areas (Shepherd, 1988). To appreciate the complex nature of how stimuli can produce overlapping brain functions in several areas, we shall briefly consider the visual pathway and the concept of face perception (the latter is related to the subject matter of this thesis).

2.2.3 – The Visual pathway

All sensory information must reach the cerebral cortex to be perceived (Ishai, Underleider, Martin, Schouten, & Haxby, 1999). After leaving the eye via the optic nerve, visual information is sent through the Lateral Geniculate Nucleus (LGN) and the Superior Colliculus (SC). Whereas LGN is part of the conscious pathway, which leads to the visual cortex (V1 or striate Cortex), information that passes to SC is not consciously perceived, and does not lead to the visual cortex. LGN is a sensory relay nucleus of the Thalamus, which is about the size and shape of a walnut (within the Limbic system), and is viewed as the gateway to the visual cortex.

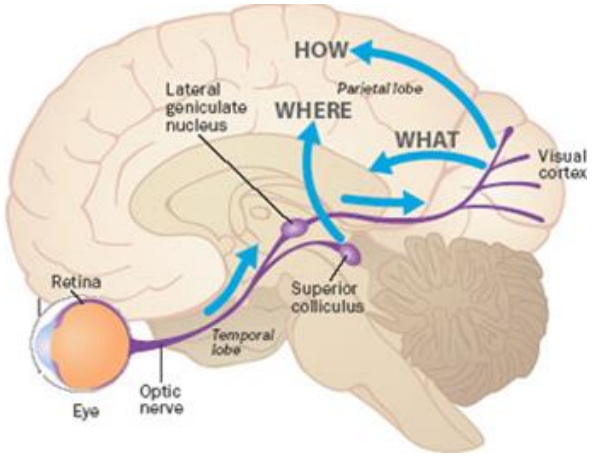

Figure 2.2 – How visual information moves from the eye to the brain. The attribution here of HOW to the dorsal stream, rather than WHERE, is certainly debated. Reproduced from Wikimedia.org.

Our vision consists of both the image that falls onto the retina, and the consequences of moving the eye, in particular, the fast saccadic eye movements (Merriam & Colby, 2005). However, it is notable that only about 10% of the input to the LGN comes from the retina, and the remaining 90% of the input consists of modulatory inputs from the Cortex and the Brainstem (Guillery & Sherman, 2002). Thus, our perception of visual information is highly influenced by our knowledge and expectations.

2.2.4 – Face perception

An individual's interpretation and understanding of the human face is called Face Perception. Humans are highly sensitive to remembering and recognising small differences between facial features (O'Toole, 2005). Many factors go into making each face unique; from proportions, colours, and features to emotional tendencies, health qualities and social information. In the brain, the processing of faces is known as the "sum of parts perception" (Gold, Mundy, & Tjan, 2012), however, it has been argued that, in order to pull all the features of a face together, individual parts must be processed first. Therefore, it has also been argued that early processing uses the Occipital Face Area (OFA), which is located in the inferior occipital gyrus, for single features of the face (e.g. mouth and nose). In contrast, the Fusiform Face Area (FFA), which is located in the lateral fusiform gyrus, is believed to pull all the processed pieces together in a holistic fashion. Although the FFA is employed in face detection and recognition, studies by (Gauthier & Tarr, 1997) have shown that experts at other objects/shapes (e.g. cars, birds, sheep, or even an invented category of stimuli, named 'greebles') will also employ the same fusiform gyrus for recognition of similar visual objects.

Whilst face processing appears to always cause activation in the FFA, encoding and recalling faces utilises extended networks (Nasr & Tootell, 2012), for example: pulvinar nuclei, inferior occipital gyrus, anterior infero-temporal cortex, posterior superior temporal gyrus, and amygdala – all with a pronounced right lateralisation. Furthermore, the means by which we gain familiarity with (or become acquainted to) faces can involve different face areas, whereby, famous/celebrity faces can be processed

differently to familiar/family faces (Sugiura, Mano, Sasaki, & Sadato, 2011). This is an important finding that may inform our observations, when analysing our familiar faces experiments – in particular, we could encounter differences between the brain oscillations in the Celebrity faces experiment (chapter 4) and the Lecturer faces experiment (chapter 5).

A cognitive/neurological disorder of face perception, whereby familiar faces (or even one's own face) are not recognised by an individual is known as Prosopagnosia (or, Face Blindness). Prosopagnosia can very rarely be 'acquired' through brain injury in the occipito-temporal lobe, but as many as 1 in 50 people (around 2% of the population) may suffer from the 'developmental' variety, which is linked to their genes (Grüter T, 2008). As the core of this thesis focuses on face perception, the quality of our research (i.e. experiments in chapters 4, 5 and 7) will depend on participants' lack of neurological defects (e.g. undiagnosed prosopagnosia). Therefore, in addition to careful selection of subjects, we will pay particular attention to measuring each participant's ability to recognise an unknown face that they have been trained to detect (i.e. by using a Target face and recognition questions, unrelated to the deception detection study, which compares a Probe with an Irrelevant).

2.2.5 – Brain as an intelligent machine

As we live in a world of uncertainty, it has been argued that the brain might work like a probabilistic machine (Pouget, Beck, Ma, & Latham, 2013); our noisy environment is filled with ambiguity, which may result in multiple interpretations of the world around us, as a result of the limitations of our sensory receptors. Thus, the best that our brain can do is to try to guess what is present, and what best action to take. This hypothesis is often credited to Hermann Von Helmholtz (Patton, 2018). Whilst studying the human eye, Von Helmholtz (1821–1894) judged it to be a very imperfect optical instrument. He proposed that visual perception was the result of what he called "a process of unconscious inference". Through this process, the brain would complete missing information using past knowledge and construct a hypothesis about our environment. This hypothesis would then be immediately accepted as a reality, and, thus, the brain can be considered to be a very sophisticated guessing machine.

Recent studies into machine learning and statistics (e.g. (Ichisugi, 2007)) have also proposed that the brain works by constantly forming hypotheses (or beliefs) about the environment, and the actions to take. These hypotheses can be described mathematically as *conditional probabilities*, where the conditional probability of an event A is the probability of an event (A), given that another event (B) has already occurred. For example, suppose we are trying to determine whether it is going to rain today, and the data available might be the existence of dark clouds. Statisticians have shown that the best way to compute this probability is to use Bayes' Rule, named after Thomas Bayes (1701–1761):

$$\mathbf{P(A \mid B) = P(B \mid A) * P(A) / P(B)}$$

Bayes' rule states that we can determine the probability (P) of the hypothesis given the data (called the 'posterior' probability), by multiplying two other probabilities: P of the data given the hypothesis (called the 'likelihood' probability), which is based on our knowledge about the probability of the data given the hypothesis (e.g. how probable is it that the clouds look the way they do now when you actually know it is going to rain), multiplied by P of the hypothesis (called the 'prior' probability), which represents our knowledge about the hypothesis before we collect any new information. The denominator, P of the data, is only there to ensure the resulting probability is normalized to be between zero and one, and can often be disregarded in the computation.

For the brain to be doing something similar to Bayesian inference, it must first combine information from different sensory modalities. For example, we use our hearing and sight to judge whether someone is following us in a quiet street at night. We may dimly see and hear some movement, and, thus, we will use both sensory modalities to assess (and react to) the situation. However, our assessment/reaction will depend on the reliability of the information available to each of our senses. Bayesian Inference predicts that the optimal way to combine information from both modalities will depend on the reliability of each information stream. Therefore, if the visual information is much clearer than the auditory information, it should have much more impact on our experience. This can lead to illusions in situations where there is a conflict between the two modalities. As a result, we can occasionally be fooled by what appears to be the dominant sensory modality, which overpowers other sensory input, as demonstrated by

the McGurk effect (McGurk & MacDonald, 1976). Therefore, it seems that the Bayesian predictions are qualitatively correct; the brain appears to combine information from different modalities (like an intelligent prediction validation machine), in a way that depends on their uncertainty, and weighs the predicted outcome with prior experiences and assumptions, before forming unconscious expectations. These prior experiences and assumptions may also influence our research, since participants' recognition of unknown faces may be influenced by memories and emotional perceptions.

The aim of this section (2.2) was to highlight the complexity of the brain, and how most brain functions rely on multiple distributed systems, for possessing information (i.e. functional overlap of several areas that activate in the brain). To simulate or decipher a certain pattern of activation in the brain, it is necessary to replicate patterns of activation by formulating a hypothesis, and testing it under strict experimental conditions, which reduce the number of sporadic activations to a minimum. Then, by recording the brain activity, and using the relevant analytical methods/tools, the previously formulated hypothesis can be tested. The next section (2.3) will focus on imaging tools and techniques that enable us to conduct our research.

2.3 Imaging tools and techniques

Events captured by human's sensory nerves (such as: touch, taste, sight, smell and sound), travel between the brain's neurons, in the form of all-or-nothing electrical pulses, called 'action potentials' or 'spikes'. The binary-paradigm of 'spike' or 'no spike' may imply that the analogue sensory data has been transformed into digital signals (Azevedo, et al., 2009). However, we are unsure of exactly how information is encoded in a spike train, as there are at least two different encoding protocol theories. On the one hand, the transfer of information can be attributed to the number of spikes in a given time interval, and on the other, the information can be encoded in the precise timing between each spike. Either way, the resultant binary neural code that occupies the brain, exists in the form of electrical activity, which refers to the neuron either firing, or not firing (i.e. 'spike' or 'no spike'), and can be recorded using several different brain imaging techniques.

2.3.1 – Brain signal Imaging

Invasive brain imaging techniques that are employed in clinical settings (e.g. Positron Emission Tomography (PET) or Electrocorticography (EcoG)) can provide more accurate measurements of brain signals (Zumsteg & Wieser, 2005), but the ethical consequences and adverse medical issues limit/prohibit their use in research settings, such as ours. Whilst invasive techniques have not been used in deception detection experiments, non-invasive techniques, such as Electroencephalography (EEG), Magnetoencephalography (MEG) and functional Magnetic Resonance Imaging (fMRI) are in common use.

The choice of technique depends on several factors, like cost, portability and spatial/temporal resolution. Whereas EEG and MEG have a similar (high) temporal resolution (less than 1ms), they suffer from low spatial resolution and lack of sensitivity to depth (see Figure 2.3). On the other hand, fMRI possesses high spatial resolution (up to 1mm) and good sensitivity to depth, but the prohibitively high cost of the equipment (as well as, the high operational cost), and low temporal resolution (i.e. delay in seconds, as it detects changes in blood oxygenation and flow) are major drawbacks.

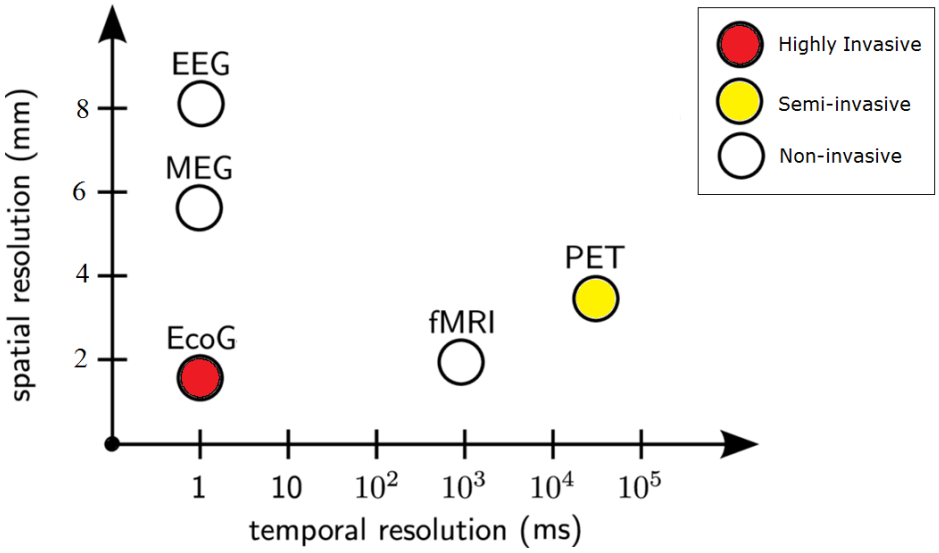

Figure 2.3 – Spatial and temporal resolutions for five different functional brain imaging techniques (Olivi, 2011). As cognitive neuroscientisit would only consider non-invasive techniques, the options are EEG, fMRI and MEG (in order of availability/cost).

The above considerations have guided us to focus our research studies on non-invasive techniques and independent neural measurements, which are captured in response to a stimulus. Namely, our goal is to use time-locked and averaged EEG signals – better known as Event Related Potentials (ERPs) – to capture and study neural activity in the brain. Despite EEG's poor spatial resolution, its precise temporal resolution (to the order of millisecond time-scale) and affordability, is the reason why EEG (and the ERP method) remains the most popular, non-invasive measure of microscopic cognitive activity, taking place in the human cortex (Luck S. , 2005).

2.3.2 – Electroencephalography (EEG)

The existence of electrical currents in the brains of rabbits and monkeys was discovered by an English physician, named Richard Caton (Caton, 1875). German psychiatrist, Hans Berger (1873 – 1941), used ordinary radio equipment to demonstrate the first non-invasive method for recording human brain activity (Berger, 1929), in the form of electrical signals. Berger laid the foundations for the use of the Electroencephalogram (EEG), as the tool for recording brain activity in humans. Early studies (Davis, 1939) examined the changes in the raw EEG activity, during simple detection and sensory processing tasks. By the mid-60s, scientists began to focus on signal averaging to generate the Event Related Potential (ERP), as the main research tool in human cognitive neuroscience (Sutton, Braren, Zubin, & John, 1965) (Donchin, 1969); (Walter, 1964).

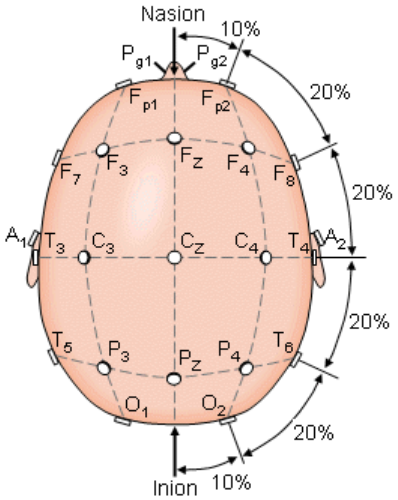

Figure 2.4 – The 10-20 International system of electrode placement (viewed top-side). Electrodes are spread over frontal, central, parietal, occipital and temporal (as denoted by letters F, C, P, O and T). Furthermore, even numbers refer to the right-side and odd refers to the left-side. Letter z (for zero) refers to the mid-line (Malmivuo, Suihko, & Eskola, 1997).

As the primary data acquisition tool in this thesis, all of our EEG experiments will employ the International 10-20 system (Jasper,

1958) for the positioning of the scalp electrodes (see Figure 2.4). Electrodes placed on the scalp are capable of capturing electrical current from the combined activity of a large numbers of similarly oriented neurons – more accurately, the synaptic excitations of the dendrites of many pyramidal neurons in the cerebral cortex (Niedermeyer & Lopes da Silva, 1993). Thus, EEG recordings reflect summed post-synaptic activity of large cell ensembles. Operationally, a change in voltage over time, between two or more electrodes placed on the scalp, can be defined as an EEG recording.

However, because the voltage fluctuations produced by brain activity are extremely small, EEG must be amplified (by a factor of 10,000 – 50,000), in order to be accurately measured. During the EEG recording, it is essential that the noise of the environment is reduced as much as possible, as there is no substitute for good data (Hansen's Axiom). Once captured, the signals can be processed, in order to improve their signal-to-noise ratio, using artefact rejection techniques, which eliminate physiological noise (e.g. eye blinks or heartbeats), as well as, environmental sources (e.g. mains power line or electrode pop/movement). Finally, the resultant EEG pattern (or waveform) will enable further analysis, along the lines of morphology and distribution. Waveforms can be measured by their Frequency (recorded in Hertz, cycles per second), Polarity (positive or negative), Phase (temporal position in an oscillation relative to a reference oscillation) and Spatial Distribution (position of the electrical currents flowing through the scalp).

2.3.3 – EEG Interpretation

Electrical currents, arising (almost exclusively) from inhibitory and excitatory cortical postsynaptic potentials, that pass from the synaptic cleft of neurons to the scalp (also known as "Volume Conduction"), consist of simultaneous summation of large cortical groups of neurons. The resultant activation of electrical current is viewed as a reliable source for our EEG waveform recordings (Atwood & MacKay, 1989).

Note that, as a result of neural activities being oriented in the opposite direction, activation may not be detected on the surface (Makeig, Debener, Onton, & Delorme, 2004b). Furthermore, since the electrodes are placed outside the brain (i.e. on the scalp),

it is conceivable that different neural activities taking place at the same time will produce little or no recordable output on the surface. The reason for such phenomena is because multiple neural activities might oscillate at opposite phases, and, thus, cancel each other out (Makeig S. D., 2004a). Additionally, electrical activity generated by the cortex is not focused onto the immediately overlaying scalp area, since the signal spreads out as it meets different layers (e.g. dura, skull and skin), before being detected by the electrode.

2.3.4 – Event-Related Potentials (ERP)

Isolating specific neural processes, using raw EEG data, can be difficult due to the vast amount of random noise. But, identifying an event/stimulus, and associating it to a pattern of activation (i.e. event-locking), makes it possible to time-lock and average the signal, in order to filter out all brain activity that is not related to the event/stimulus (Makeig, Debener, Onton, & Delorme, 2004b).

To demonstrate this averaging technique in a real-world example, we can set-up a camera on a tripod – largely, to ensure that photos can be taken from the same position, and to avoid any shaking – and then take multiple photos of the same scene; one every 15 seconds. After 20 or more shots, we can download the images to a photo editing software (in this example, we used an open source photo editor, called GIMP), and open all photos together, as layers. The 20+ layers of real-world photos (see Figure 2.5) will possess prominent features that appear in all images (e.g. mountains and trees), but there will also exist non-stationary items (e.g. moving people and animals), which only appear in some of the images, and their positions in those images are different, due to the time-lag between each photo. Next, if all the layers/photos are averaged, using the '*Median*' filter (see Figure 2.6), the GIMP photo editor combines all the images and removes the unwanted noise (or in this case, non-stationary people and animals).

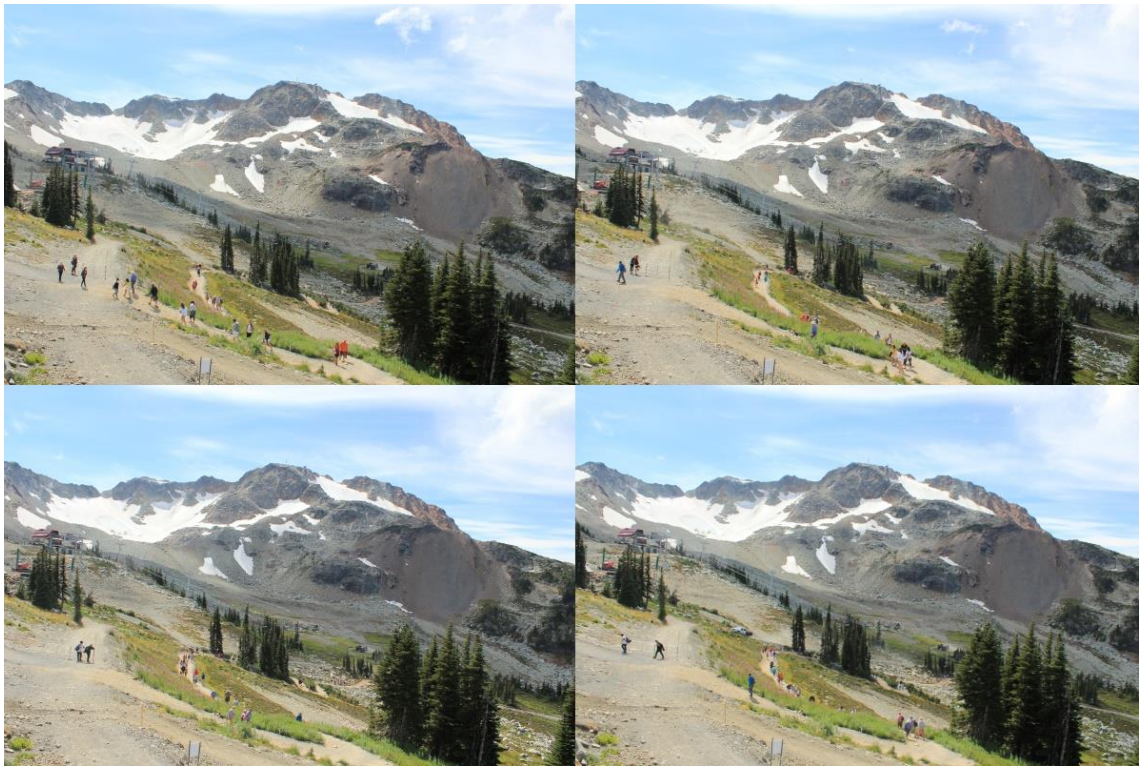

Figure 2.5 – Real-world demonstration of the averaging technique: four of 20 photos, taken from the same position, showing walkers passing by a mountain scene. Original photos reproduced from toomanyadapters.com website.

As a result of averaging, we are left with an image of the prominent features (e.g. the mountain scenery), which excluded the noisy/non-stationary items (see Figure 2.6).

Figure 2.6 – Results of the real-world demonstration of the averaging technique: having averaged 20+ photos taken from the same position (using the Median filter of the graphics package, GIMP), the unwanted tourists (likened to 'noise') are removed, leaving the the mountain scenery (i.e. the prominent features). Figure 2.5 shows 4 of the original photos.

Similarly, stimuli that are presented often in raw EEG data must first be separated into individual trials, by time-locking them to the stimulus, and then averaged into Event Related Potentials (ERP), as shown in Figure 2.7 (below). Indeed, as random noise varies across trials, averaging will reduce the noise, however, deflections that are consistent across trials will remain. Typically, the resulting ERP waveforms will show a series of positive and negative components, which can be identified using their polarity and time-point (e.g. P3, which is a positive ERP component that occurs approximately 300 to 800 milliseconds after the onset of stimulus).

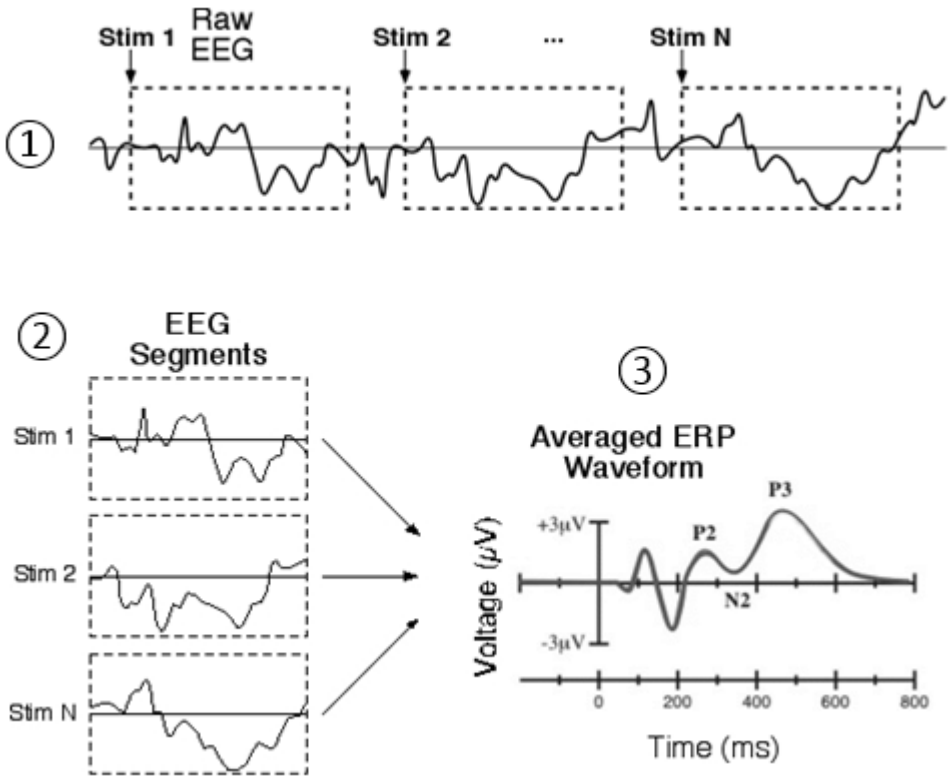

Figure 2.7 – Step 1: Raw EEG data is split into stimulus and time-locked trials. Step 2: Trials are averaged together, in order to subtract brain activity that is not related to the appearance of the stimulus, to form the ERP waveform. Step 3: Averaged waveform (ERP) contains positive and negative components (reproduced from erpinfo.org site).

Averaging is a popular signal processing technique, which is employed to clear time-locked noisy signal components (i.e. artefacts). An ERP is considered to be a good tool to delineate psychiatric and neurological conditions, such as schizophrenia and

ADHD (Ford et al., 1999) (Van der Stelt, Van der Molen, Boudewijn, & Kok, 2001), as well as, studies into human attention (Mangun & Hillyard, 1995), and lie-detection (Farwell & Donchin, 1991). Its widespread applicability is because ERPs reflect the summed activity of postsynaptic potentials, which are produced when thousands or millions of similarly oriented cortical pyramidal neurons fire in synchrony, whilst processing information (Peterson et al., 1995). It has been argued that ERPs are broadly divided into 2 categories, according to latency and amplitude of the waveforms:

> **Sensory** (or exogenous) components, which peak within approx. 100 milliseconds post-stimulus, and depend on the properties of the stimulus.
>
> **Cognitive** (or endogenous) components, which appear later, and reflect the manner in which the information is evaluated and processed.

From a group-level analysis point of view, we must acknowledge a potential flaw in the averaging technique: if the amplitude of a component varies from trial-to-trial, then the ERP waveform will reflect the average amplitude from all the trials. Although the trial-to-trial variability in *amplitude* does not pose any issues, such variability in *latency* can result in failure to identify a neural response, or can lead to false conclusions; especially, as most studies (including ours) present a grand average ERP, which is produced by averaging the individual ERPs with the objective to minimise the variability across several subjects (see Figure 2.8).

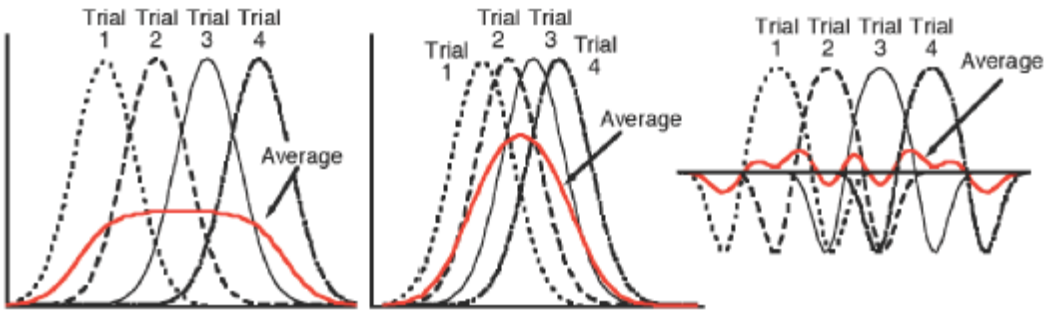

Figure 2.8 – Latency differences across single trials in the ERP waveform, which can result in failure to identify a neural response (reproduced from (Luck, 2005), p.136).

One way to overcome the above limitations is to complement the statistical analysis in the Time domain (using ERPs), with additional analysis in the Frequency domain (using single trials), as we have done in all three research experiments, presented in this thesis.

2.3.5 – ERP Components

Stimulus-locked ERP waveforms consist of peaks and troughs, which correspond to cognitive processing that is time-locked to the sensory presentation of a stimulus. Major functional areas of the waveform are known as ERP components; they are defined by their polarity, timing and scalp distribution. It is often possible to infer specific features of ERP waveforms, as markers for correlated cognitive processes. Furthermore, cognitive processes that differentiate two conditions (e.g. attended familiar 'Probe' stimuli/faces and unattended unknown 'Irrelevant' stimuli/faces) will elicit differing stimulus-locked ERP waveforms, which relate to the functional characteristics of selective attention.

Because sensory and cognitive processes overlap, both in time and space, the peaks and troughs of the resultant waveforms can be associated with the summation of several contributing sources. Other than the 'physiological' and 'functional' approaches to component classification, there is no universally accepted definition of what constitutes an ERP component. The 'physiological' approach (Naatanen & Picton, 1987) defines ERP components in terms of their anatomical source within the brain, and the 'functional' approach (Duncan-Johnson & Kopell, 1981) defines the functional process with which it is associated. A combination of both approaches (i.e. functional significance and their underlying neural sources) is often adopted in the definition of ERP components. Some of the common family of components that can be observed in ERP waveforms, and are of interest in our cognitive studies, are outlined below:

N170 components reflect the neural processing of faces, with an increased negative deflection between 130 – 200ms post-stimulus (Bentin, Truett, Puce, Perez, & McCarthy, 1996).

N250 components are associated with repeated presentation of faces, with more negativity in response to familiar faces, at approximately 250ms after stimulus onset (Zimmermann & Eimer, 2013); (Pierce, et al., 2011).

P3 (or P-300) component is related to indexing working memory (Vogel & Machizawa, 2004) followed by components elicited during the selection and preparation of motor responses. As the 3rd positivity found in ERPs, it has become one of the most prominent patterns, since its discovery in 1965 (Sutton, Braren, Zubin, & John, 1965). P3 has been studied extensively, in relation to cognitive functions (Craston, Wyble, Chennu, & Bowman, 2009); (Wyble, Bowman,, & Nieuwenstein, 2009), decision making (Rohrbaugh, Donchin, & Eriksen, 1974); (Radlo, Janelle, Barba, & Frehlich, 2001), and working memory (Gaspar, et al., 2011). This component can be elicited in response to an attended (i.e. task-relevant) stimulus, and it can further be sub-divided into **P3a** (or the 'novelty' P3), which is related to the engagement of attention and the processing of novelty, whilst **P3b** relates to task-relevant (and thus, not novel) improbable events (Kok, 2001). Classically, both the P3a and P3b require the stimulus to be infrequent (i.e. an oddball paradigm), such that the stimulus frequency and P3 amplitude appear to be inversely proportional (Verleger, 1988). Albeit, this requirement may not be relevant to P3s generated from RSVP experiments.

N400 component is a negative-going deflection, approximately 300 – 500ms post-stimulus, which reflects the identification of anomalous endings to semantic processing of words, images & sounds (Kutas & Hillyard, 1980).

P600 component reflects language processing's syntactic violation, non-preferred syntactic structure, or complex syntactic structure (Osterhout & Holcomb, 1992). It is characterized as a positive-going deflection, with onset around 500ms after the stimulus that elicits it.

Whilst the above N400 and P600 components relate to semantic language processing, their equivalent components to face familiarity, which possess roughly similar latency and topography, are called **N400f** and **P600f** , where 'f' denotes face

(Taylor, Shehzad, & McCarthy, 2016). The N400f appears to be associated with the activation of the episodic memory of the face, and the P600f is considered to reflect explicit recognition of a particular individual (Sun, Chan, & Lee, 2012). For the purposes of our three research experiments (see Chapters 4, 5 and 7), it must be noted that, whilst N170 will always accompany neural encoding of faces, N400f and P600f appear to "indicate subsequent processes involved in face recognition" (Eimer, 2000). Therefore, the key aim of this thesis is to focus on the identification and ERP analysis of the face related N400f and P600f components.

2.3.6 – ERP (Time Domain) Analysis

ERP *amplitude* (which measures the size of the component, and, thus, the strength of response) and ERP *latency* (which measures the timing of the component, and, thus, the speed of response), along with scalp distribution, are considered to be the most common ERP component measurement techniques in neuroscience studies (Polich & Kok, 1995). Indeed, the main challenge is the quantification and interpretation of ERP differences across conditions, in order to obtain an accurate result, without distortions caused by noise and overlapping components.

2.3.6.1 – Window Placement

To determine the size of the ERP components, as a measurement of their amplitude, we have previously employed the *Mean amplitude* method (Bowman, Filetti, Alsufyani, Janssen, & Su, 2014), in which the average voltage over a pre-selected time window (i.e. the region that is supposed to contain the component of interest) is calculated, to provide the mean amplitude measure. Mean amplitude measurements are robust against high frequency noise, since, instead of using a single point, it is based on a range of time points. However, the correct selection of the Region Of Interest (ROI) requires a careful balance between the detection of effects without increasing false positive (type I error) rates (Kilner, 2013), and reducing false negative (type II error) rates.

According to (Brooks, Zoumpoulaki, & Bowman, 2017), a data-driven ROI selection technique can be used safely in ERP experiments, which typically increases statistical power relative to a-priori fixed-window placement (Luck S. , 2014). As such, group-level ROI selection depends on the Aggregate-Grand-Average-from-Trials (AGAT), which is similar to the use of orthogonal contrasts for ROI selection in fMRI research (Friston, Rotshtein, Geng, Sterzer, & Henson, 2006). The AGAT is computed by aggregating all of the individual trails from all subjects and conditions (e.g. the Probe and Irrelevant), before averaging them into a single time-series. Next, an algorithm searches automatically, to find the minimal/maximal 100ms interval average, where the start-and-end of this minimal/maximal 100ms ROI defined the group-level features/components, for both conditions. In Chapter 3 (see section 3.3.3), we will describe the time domain data analysis method in more detail, as it relates to group-level and subject-level analyses.

2.3.6.2 – Statistical Test

To determine whether the observed values of ERP components are significant, and to draw conclusions, analysis of variance (ANOVA) can be used (amongst other statistical tools) to determine whether variability of means is extreme relative to the error variance (Dien & Santuzzi, 2004). However, in previous deception detection experiments (Bowman, Filetti, Alsufyani, Janssen, & Su, 2014), to show the significance across the whole set of individuals (i.e. group-level), a *t*-test was used to determine whether there is a statistically significant difference between two conditions (e.g. a paired *t*-test of the mean amplitudes of the Probe and Irrelevant) for the whole group of subjects.

Alternatively, at individual/subject-level, to determine whether the difference between two conditions is significant (e.g. to draw conclusions as to whether the EEG data evoked by the familiar face was significantly different from that evoked by the unfamiliar face), we have previously applied a randomisation (also known as, Monte Carlo Permutations) test (Bowman, Filetti, Alsufyani, Janssen, & Su, 2014). A randomisation test is a technique to determine whether the difference between two conditions is significant, using the following steps:

1. The trials of two experimental conditions are collected into a single set;

2. Two random partitions are created from the combined set (i.e. by randomly choosing trials from the set of combined data, until there are as many trials in the first partition as in the original condition, and then placing the remaining trials in the second partition);

3. The difference between the new sets is calculated.

4. Steps 2 and 3 are repeated many times (e.g. 1000, or more), resulting in a histogram of test statistics;

5. The *p*-value is calculated, as the proportion of the test statistics that were greater than the one obtained from the original condition;

6. If the *p*-value is smaller than a critical alpha-level (normally, 0.05), then the data in the two experimental conditions are significantly different.

2.3.7 – Single Trial (Frequency Domain) Analysis

So far, we have only considered the Time Domain (ERP) analysis of EEG data, but (as noted earlier) if there is a large variability across single trails (with regards to the latency or amplitude), or the phase of ongoing oscillations are not reset, the ERP waveform might not accurately reflect the individual brain activity waveforms recorded in single trials (Bressler & Ding, 2006). Therefore, we will now consider the time-frequency transforms that enable us to switch the perspective of our analysis from the time to the frequency domain (Delorme & Makeig, 2004), as provided by EEGLAB's implementation. The '*newtimef'* function, which is used in our experiments, provides both power and phase-locking (also referred to as, coherence) information (Makeig, Debener, Onton, & Delorme, 2004b). *Power* changes at each time-point, against a pre-stimulus baseline, are computed by Event Related Spectral Perturbations (ERSP), and

Coherence changes are calculated by the Inter-Trial Coherence (ITC), which reflects the synchronisation of phase across trials.

Whereas, ERPs can detect evoked responses (i.e. average of multiple evoked brain responses that are time and phase locked to the same event), Time Frequency analysis (i.e. ERSP and ITC) can detect any induced or evoked response (i.e. induced changes in the spectrum of ongoing EEG that are time-locked, but not phase-locked). Indeed, averaging oscillations across a set of trials that consist of random phase potentials may yield a feature-less flat (ERP) line. However, brain activity of single trials (ERSP/ITC) may be able to interpret such oscillations. Therefore, our research will augment our ERP analysis with time-frequency transforms, which carry more information. For further information on Frequency Domain Analysis and time-frequency window placement, see chapter 3's section 3.3.5.

2.4 Research Objectives

Having described the relevant imaging tools and the statistical analysis methods, we will now consider the objectives of our research, by introducing the techniques and procedures that make our face recognition experiments possible.

2.4.1 – Rapid Serial Visual Presentation (RSVP)

Introduced in 1971 (Lawrence, 1971), the Rapid Serial Visual Presentation (RSVP) technique was used to examine the temporal characteristics of attention. RSVP allows a series of visual items (i.e. letters, words or images) to be presented, in a fixed position (see Figure 2.9), at a speed of between 6 – 20 items per second (Raymond, Shapiro, & Arnell, 1992). Studies ((Potter, 1975) (Lawrence, 1971) (Chun & Potter, 1995)) have shown that the detection of a Target stimulus was successful at high presentation rates of 10 or more items per second.

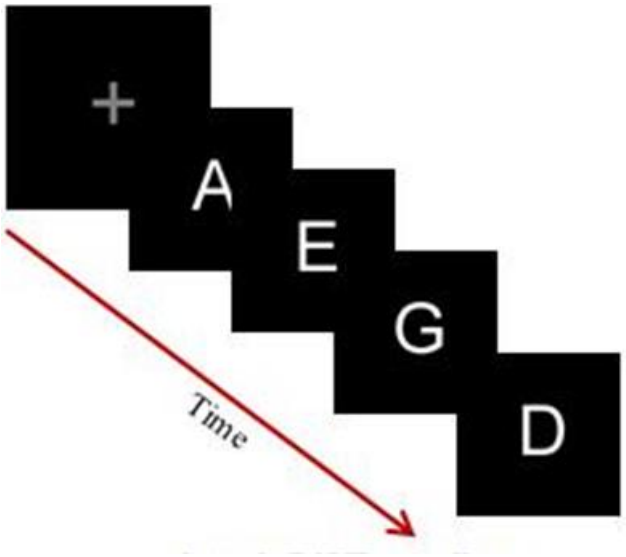

Figure 2.9 – RSVP stream, consisting of letters, where letters are presented in a fixed position on the screen.

RSVP contains a combination of Distractors (also known as, Fillers) and one (or more) meaningful Target(s). In addition to the ability to display a large number of items, very quickly, RSVP's non-dependency on eye-gaze is extremely useful, since all items are presented in the same location. However, the central point that makes RSVP a useful tool for our EEG-based cognitive research has to do with the ability to measure the timing of rapidly presented salient stimuli – because the brain cannot process (to the point of encoding into memory) all the items in the RSVP stream that is being presented at a high speed, only salient stimuli will be processed by the Ventral Visual Processing pathway (Bowman & Wyble, 2007), and in turn may elicit a P3 component. Indeed, it is the high presentation rate of the RSVP paradigm that facilitates (what has been argued as) a countermeasure resistant technique for our experiments (Bowman, Filetti, Alsufyani, Janssen, & Su, 2014). This is the ideal scenario for our studies, as we are targeting the events and stimuli that appear on the 'fringe of awareness'.

Despite RSVP's advantages (e.g. gaze independence), some considerations must be observed, in order to ensure that the RSVP streams are not affected by *Masking* effects. For example, the Attentional Blink (AB) may occur if the time that lapses between two salient stimuli is less than 500ms (Luck, Vogel & Shapiro, 1996); (Raymond, Shapiro, & Arnell, 1992). Similarly, Repetition Blindness may affect the salient stimuli (Kanwisher, 1990). Note that the Target item(s) within the RSVP stream,

must not be placed in the starting-region (i.e. the first 5 items) because the P1-N1 complex, which is generated by the transition from a blank to non-blank screen, can interfere with the onset of the P3 component. Furthermore, we must also avoid the finishing-region of the RSVP stream (i.e. at least the last 5 items), in order to allow enough time for the P3 component to start, reach its peak, and finish. Thus, an allowance of up to 1000ms may be required, before components marking the end of stream can be shown (e.g. the expectation of the behavioural question, at the end of the stream, may generate an unintended preparation effect).

In our face recognition experiments, each item in the RSVP stream will be presented for a very short time; this is called the Stimulus Duration (SD). Sometimes, a blanking interval may be required in between each item; this is called the Inter-Stimulus Interval (ISI). However, the key indicator is the amount of time that elapses between the presentation of each stimulus (i.e. SD + ISI); this is called the Stimulus Onset Asynchrony (SOA), which has been fixed, for all of our experiments, to 133ms.

As we will show in chapter 3, the use of RSVP for our face recognition experiments will enable us to present images on the fringe of awareness, and take advantage of the concept of sub/liminal salience search (Bowman, et al., 2013), where the majority of images are not perceived at a level that is considered to be sufficient for encoding into working memory. However, salient images that breakthrough into conscious awareness will generate a P3 ERP component pattern, which underlies the fringe-P3 method, introduced in (Bowman, Filetti, Alsufyani, Janssen, & Su, 2014). Indeed, the fringe-P3 method provides the possibility of a reliable ERP deception detection test that is resistant to countermeasure that had confounded previous methods (Rosenfeld, Soskins, Bosch, & Ryan, 2004).

2.4.2 – Fringe-P3 in Concealed Information Tests

As we have noted earlier (in chapter 1), the classical oddball experiments, which uses the P3 (i.e. the ERP component that occurs after the onset of a target stimulus that appears within a series of non-target stimuli) has been successfully used, in an investigative context, within Concealed Information Tests (Johnson & Rosenfeld, 1992)

(Rosenfeld, Soskins, Bosch, & Ryan, 2004) to detect deception. However, where the stimuli are presented at a slow rate, there could be a potential for suspects to confound the tests, using countermeasures (Meixner, Haynes, Winograd, Brown, & Rosenfeld, 2009); (Lukács, Weiss, Dalos, Kilencz, Tudja, Csifcsák, 2016). A key reason for the possibility of behavioural countermeasures was previously attributed to the availability of sufficient time between each stimulus, to consciously determine that there is a repeating non-critical irrelevant item. However, (Bowman, Filetti, Alsufyani, Janssen, & Su, 2014) observed that if the stimuli were presented at a rapid rate using the RSVP paradigm, the irrelevant stimulus will not be sufficiently perceived to be noticed as repeating. Therefore, by presenting items rapidly, the majority of the items are not perceived at a level that is considered to be sufficient for encoding into working memory.

As noted earlier, the cognitive state that the subject assumes, whilst attending to the RSVP stream, is called Sub/liminal Salience Search (SSS). The term 'Search' refers to the rapid perceptual search for a Target item in the RSVP stream; the term 'Salience' refers to the Target which is salient to the subject; and the term 'Sub/liminal' refers to the fact that the majority of the items in the RSVP stream are not consciously perceived by the subject, even though, the subject is 'unconsciously' processing for salience of the items in the stream. Additionally, in RSVP, whilst the subject's brain is searching for salient stimuli, at a very high presentation rate, it is possible to detect an electrophysiological marker (e.g. the P3 component), when the salient stimulus is detected.

As a result, according to the fringe-P3 method (Bowman, Filetti, Alsufyani, Janssen, & Su, 2014), salient items that breakthrough will generate a unique ERP pattern, which, we believe, can be correlated with a P3 ERP component. Due to these characteristics, SSS has been proposed as "a novel deception detection system based on RSVP" (Bowman, et al., 2013). This proposed method is more robust in the context of deceivers, who aim to confound the test using countermeasures.

2.4.3 – Deception Detection

Despite our existential reliance on the ability to be able to distinguish between the truth and a lie, humans can rarely outperform chance (Bond & DePaulo, 2006). The earliest recorded causal effect, between a physiological indicator and deceit, has been attributed to the Greek physician, Erasistratus (304 – 250 BC), who posited that an increase in a subject's pulse rate is an indicator of guilt (Trovillo P. , 1939).

For over two millennia, the pulse-indicator remained the only objective way to measure deception, until instruments for blood pressure measurements were invented in the late-nineteenth century (Trovillo P. , 1939). In the early-twentieth century, an instrument that combined heart-rate, blood pressure and respiration was invented for the U.S. law enforcement agencies (Larson., 1923), with the primary purpose of detecting deception. With the addition of the galvanic skin response, the modern-day 'Polygraph' (Greek for 'many writings') was born (Lykken., 1959). Although polygraph machines look scientific, and measure real physiological reactions to stimuli, the methodology suffers from an unacceptably large number of "False Positives" (Lykken D. , 1984); (Adelson, 2004), which contributed to the polygraph test being discredited in many legal proceedings.

The decline in the scientific validity of the polygraph, as a reliable aid in detecting lies, has accelerated the search for new methods and techniques, which can provide non-invasive and reliable insight into the human psyche. Despite its unwavering popularity as a cultural icon, which purports to expose the liar, the idea that we can detect deception, by monitoring psychophysiological changes, is more myth than reality (Saxe, 1991). Polygraphs typically record three indicators of autonomic arousal: the heart's blood pressure and heart rate (using blood pressure cuffs around the arm), respiration rate (using pneumographs around the chest), and skin conductivity (using electrodes attached to the fingertips). In criminal incident investigations, the reaction to a Control Question Test (CQT), that is broad in scope and appears threatening (e.g. 'Have you ever betrayed a person who trusted in you?'), will be compared to a crime-relevant question (e.g. 'Did you steal the £500?'). If the subject is innocent, they will fear the control questions more than the relevant questions, because the former arouses concerns about past truthfulness, whilst the latter is related to a crime that they know

they did not commit. The examiner may conclude that, the opposite physiological reaction to these questions symbolises deception (Larson., 1923). However, a host of mental states (e.g. nervousness, embarrassment, anger and fear), or medical conditions (e.g. headaches, constipation, colds and neurological/muscular problems) can be causal factors in altering a subject's heart rate, blood pressure, skin conductance and respiration.

Another popular polygraph procedure is named the Guilty Knowledge Test (GKT); it requires the subject to take a multiple-choice test (e.g. "Was there £200, £500 or £800 stolen?"). Such a psychometric test is designed to promote a larger psychological reaction to the correct choice for the guilty subject, albeit, the reaction might just be due to familiarity. However, GKT is limited by the factual information that is available to the examiner (i.e. it is not possible to ask questions that only the guilty party has the answers to). The 2004 findings of the American Psychological Association asserted that, there is no empirical evidence that any pattern of physiological reactions is unique to deception, as an honest person may be nervous when answering truthfully, and a dishonest person may use countermeasures to stave off anxiety (APA, 2004), according to several studies (Saxe & Ben-Shakhar, 1999); (Kozel, Padgett, & George, 2004). The reality is that, at best, polygraphs are a good test of a subject's physical reactions to certain questions, but nobody knows how the subject's nervous system acts when they might be lying.

The ongoing search for a scientifically viable alternative has not escaped the attention of scientists who specialise in cognitive neuroscience. Empirical techniques for measuring deception at the source of human cognition (Abootalebi, Moradi, & Khalilzadeh, 2006) have profound legal, moral and clinical implications, and as such, this is considered to be the 'holy grail' for an increasing community of neuroscientists. The practicality and affordability of EEG has meant that it is widely used in deception detection studies (Farwell & Donchin, 1991) (Rosenfeld, Angell, Johnson, & Qian, 1991). These studies exploit the assumption that familiar stimuli generate a larger P3 ERP than non-familiar ones (Farwell & Donchin, 1986), but, so far, only the fringe-P3 method (Bowman, Filetti, Alsufyani, Janssen, & Su, 2014) has demonstrated the advantage of being countermeasure resistant. This had opened up the possibility of designing reliable RSVP-based EEG concealed information tests, with real-life

deception detection applications, using *numbers, words* and *names*. And now, in the first dedicated research of its kind, we have used *human faces* exclusively, to demonstrate the broader applicability of the fringe-P3 method, and to enhance the validity of RSVP-based EEG face recognition tests.

2.4.4 – Face Identification

Face perception and recognition is extremely difficult, and yet, most of us can seamlessly recognise thousands of faces, often without much effort. The debate surrounding the question of whether the recognition of faces is an automatic process, or a task-dependent one remains controversial (Tanaka & Simonyi, 2016). It has been claimed that the pre-emptive nature of face perception stops short of face recognition, as the latter requires selective attention of facial cues that define the individual's identity (Palermo & Rhodes, 2002). If this were correct, individual faces may only be recognised in a task-relevant context, but (Ellis, Young, & Flude, 1990) have argued that "the identity of familiar faces is impossible to ignore", and can be recalled in a task-independent (automatic) fashion. Furthermore, notwithstanding the exception of repetition priming effects with novel faces (Goshen-Gottstein & Ganel, 2000), identity-related visual cues for unfamiliar faces do not appear to be encoded when they were not task-relevant (Ellis, Young, & Flude, 1990). In other words, familiar faces may be recognised regardless of current task demands, whereas, unfamiliar faces require a task-relevant context (Zimmermann & Eimer, 2014).

The contrast between familiar and unfamiliar face recognition can be attributed to the fact that familiar faces have been repeatedly revealed to an individual, over numerous perceptual episodes, and are thus likely to be well established in visual memory. Whereas the transient nature of unfamiliar faces means that recognition is based on a (very) limited number of prior encounters. As we have noted earlier, it is our intention to introduce human faces to the countermeasure resistant fring-P3 method, and by using RSVP-based EEG recordings, we aim to identify precise information about the timing of neural events (in Time and Frequency domains), and focus on the specific ERP components that can be present during face perception and recognition tests.

2.5 Conclusion

For as long as humans have been able to think, they have sought to understand the brain. As the most complicated organ in the human body, considerable work needs to be done before it will be fully understood. And yet, recent advances in imaging technologies and analysis techniques have accelerated our understanding, and led to significant findings. Imaging techniques, such as EEG, have enhanced our knowledge of various cognitive tasks, as researchers can record changes in brain activity during engagement of mental tasks. Interpretation of such recordings, using innovative techniques, in Time (ERP) and Frequency (ERSP/ITC) domains, has facilitated new discoveries between brain responses to cognitive tasks.

Novel techniques, such as RSVP, and unique analysis methods (e.g. countermeasure resistant fring-P3) have made it possible to study the instance of perceiving otherwise sub/liminal items. As a result, EEG experiments that can explore the sensitivity of ERP-based RSVP paradigm have been successfully used in deception detection studies. Their success will ultimately lead to a viable alternatives to the (controversial) polygraph-based methods.

The introduction of human faces to the fringe-P3 method, combined with the use of RSVP-based EEG experiments, to infer recognition of broadly familiar faces, could be an important step in the adoption of brain recording systems, in real-life settings. We propose that the success of our research will extend the exploration of image-based RSVP solutions, and promote future applications of deception detection of crime compatriots (e.g. a scenario where relevant authorities can establish a suspect's familiarity with a criminal/terrorist, using identification of sub/liminal faces).

P A R T II

–

R E S E A R C H

Chapter 3:
Research Design Framework

3.1 Introduction

In Part Two of this thesis, we will begin by developing our research ideas (in this, chapter 3), and by describing the general design and framework of all three face recognition studies. Thereafter, we will use chapter 3 as the standard reference point for each experiment (see chapters 4, 5 and 7), in order to avoid unnecessary repetition of our general research aims, concepts, and methods.

3.1.1 – Background

All three experiments, in this thesis, employed the Rapid Serial Visual Presentation (RSVP) paradigm, whereby, salient images – in the form of familiar faces – can breakthrough into conscious awareness and become encoded into memory. RSVP enables us to present a critical stimulus (e.g. a famous face) within a series of distractor images (e.g. unknown/anonymous faces), at a rate that is considered to be on the fringe of awareness, so that only the salient stimuli would breakthrough. By presenting images rapidly, and taking advantage of the concept of Sub/liminal Salience Search (Bowman, et al., 2013), the majority of the images are not perceived at a level which is considered to be sufficient for encoding into working memory. However, images that are salient and breakthrough will generate a unique ERP pattern, which, we believe, can be correlated with a P3 ERP component (also referred to as the *fringe-P3*).

The evolution of our research began with the first study (Chapter 4), which involved the comparison of two categories of critical stimuli: those that were highly familiar to the subjects who participated in the experiment (hereafter referred to as the *Probes*), and novel faces that were believed to be unknown to the subjects (hereafter referred to as the *Irrelevants*). The familiar stimuli consisted of the most famous celebrities of the day, based on the highest ranked Yahoo searches of famous people, in

2014. However, participants were not informed of the presence of famous faces within the experiment, as this enabled us to study their sub/liminal reactions, to these highly recognisable stimuli. In fact, subjects were trained to look for a third critical stimulus (hereafter referred to as the *Target*), in the form of a single image that was believed to be unknown to subjects, prior to the experiment. Thus, making the Target task-relevant, as subjects were instructed to look out for (and respond to) this image, within the RSVP streams.

The key change in the second experiment (Chapter 5) was to replace the famous/celebrity critical stimuli with familiar faces from our University, in the form of lecturers (and/or supervisors) that each subject had a close working relationship with. For this reason, we performed the experiment on PhD students only, so that we could be assured of a long-term relationship/familiarity between subjects and their lecturers' faces. Also, subjects were chosen on the basis of never having been included in a similar EEG/RSVP experiment, and all participants were instructed to avoid discussing the experiment with their colleagues, in order to avoid any priming of future participants. In the final experiment (Chapter 7), we maintained the use of the same critical stimuli (i.e. familiar lecturer faces), but this time, the notable change was that we revealed the fact that the experiment would contain lecturers that subjects would be highly familiar with. Once again, only PhD students were invited to the third experiment, and none of them were included in previous/similar EEG/RSVP experiments.

Through the evidence gathered in each of our experiments, we planned to evolve our face perception and recognition methods and hypotheses, towards a scientifically robust framework, which would facilitate effective EEG tests on individuals (utilising independent measures to obtain orthogonal contrasts) that could reveal the hidden information/knowledge that is contained within the human mind. Our research is considered to be the *first* systematic attempt at employing Rapid Serial Visual Presentation (RSVP) tasks, to study the breakthrough of familiar faces, on the fringe of human awareness (i.e. using the fringe-P3 method). This research will inform future studies into face recognition and concealed information tests, as well as, potential applications in EEG-based deception detection of compatriots/deceivers.

3.1.2 – Face Perception and Recognition

Face perception has been important in the understanding of perceptual and cognitive aspects of human neurodevelopment, as they convey essential information regarding identity, intent, emotion and social interaction. Ever since Darwin's book, Expression of the Emotions in Man and Animals (1872), processing of faces by the human brain has remained at the centre of the nature-nurture debate, in regards to *phylogeny* (species adaptations) versus *ontogeny* (experience-based individual development). Fantz demonstrated that new-born babies generally preferred looking at a schematic-face rather than a bull's-eye pattern (Fantz, 1963). Johnson and Morton proposed the 'two-process' model of CONSPEC (tendency of new-borns to orient to faces) and CONLEARN (acquired specialisation of cortical circuits for face processing), in which sub-cortical processing guided the behaviour of new-borns in favour of face-like patterns (Johnson & Morton, 1991). Farroni demonstrated the tendency of new-borns to maintain mutual gaze, thus, developing the face-sensitive areas within the cortex (Farroni, 2002). Studies indicate that although infant face recognition tends to develop rapidly, adult-like maturity and proficiency takes much longer to develop (Carey, 1977). The authors predicted that by about the age of 10, children's dependency on featural/piecemeal strategies in perceiving faces rapidly evolves into more configural/holistic strategies. During face perception in adults, excepting those with neurological/cognitive disorders (e.g. prosopagnosia, which has been described in section 2.2.4 – *Face Perception*), neural networks are activated within the brain to recall memories and to process information.

ERP studies into face perception and recognition, in adults, have reliably reported neural activity, with specific components that indicate sub-conscious and conscious processing of faces. Numerous studies have reported a face-specific N170 component – a negative deflection, elicited in the ERP within 140 and 200ms (peaking at around 170ms post-stimulus) over lateral occipito-temporal areas and posterior fusiform gyrus – which is thought to reflect an activation of "person identity nodes" (i.e. structural encoding of faces) in the subject's semantic memory (Bruce & Young, 1986); (Bentin, Truett, Puce, Perez, & McCarthy, 1996). However, the specificity of the N170 to faces remains inconclusive, as studies (Gauthier & Tarr, 1997) have shown that it is

possible to observe a similar neural activity when the participant is presented with objects/shapes that are highly familiar to them (e.g. cars, birds, sheep or even *greebles*).

Whilst the N170 component does not appear to be affected by the difference between famous and unfamiliar faces (Bentin & Deouell, 2000), an enhanced negativity called the N400 component (also referred to as N400face, or N400f) appears to be associated with the subsequent activation of the episodic memory of the face; this interpretation is consistent with, and supported by, studies of similar semantic processing of linguistic material (Kutas & Hillyard, 1980). As a negative deflection, the N400f is elicited in the ERP within a time window of approx. 250 to 500ms, post stimulus. Amongst others, (Eimer, 2000) observed a subsequent/late positivity for famous faces, referred to as the P600 component (or P600f), which has also been compared with the P3 component (Sutton, Braren, Zubin, & John, 1965). This P3/P600f neural activity is considered to reflect explicit recognition of a particular individual, elicited as a positive deflection, between 300 to 900ms, post stimulus.

3.1.3 – Aim of Research

Scientific enquiries into 'Lie Detection' and 'Concealed Information Test' through the use of the P3 Oddball paradigm – such as, EEG-based studies by (Rosenfeld, Soskins, Bosch, & Ryan, 2004) and (Labkovsky & Rosenfeld, 2012) – have exhibited an ERP based vulnerability to countermeasures. Other studies (Meijer, 2009) and (Lefebvre, 2007), into Concealed Information Test (CIT), demonstrated successful application of face stimuli in ERP-based experiments, albeit, they employed the classic P3-oddball paradigm, which is considered to be vulnerable to countermeasures (Rosenfeld, Soskins, Bosch, & Ryan, 2004). However, recent studies into RSVP (Bowman, et al., 2013); (Bowman, Filetti, Alsufyani, Janssen, & Su, 2014) have demonstrated what they argue is a countermeasure-resistant P3 component, which could be used to differentiate between deceivers and non-deceivers. Consequently, due to the Sub/liminal Salience Search (SSS) effect, salient stimuli that are presented in RSVP streams (e.g. a famous face, as hypothesised in this thesis) will breakthrough into awareness, whilst non-salient stimuli (e.g. an unknown face) will remain non-conscious (i.e. sub/liminal), making it more difficult for subjects to use countermeasure strategies.

Having provided significant evidence for the existence of the countermeasure-resistant P3 component in concealed information experiments that employ numbers, letters and words as stimuli, the next logical (and unchartered) frontier had been to explore the suitability of image-based stimuli, in ERP-based RSVP paradigm, to wit: do familiar faces differentially break into conscious awareness, on an individual basis, and can we detect the breakthrough events in EEG? Thus, we began by studying the effects of presenting famous faces (i.e. Probe critical stimuli) and unknown faces (i.e. Irrelevant critical stimuli), using the RSVP technique. Notably, in the first study (see Chapter 4), we did not provide instructions about the presence of the Probes because we were interested to find out if the brain will select the salient stimuli (i.e. the famous faces), even when there is no explicit task associated with them. Furthermore, subjects were instructed to look for a Target stimulus (i.e. a face that was unknown prior to the experiment), that would become the only task-relevant context. This enabled us to make direct comparisons between the Probe and Irrelevant conditions, which were presented equally as often and statistically in the same position in the RSVP streams.

As outlined in the 'Introduction' of this chapter, in the second study (see Chapter 5), we retained the majority of the experimental parameters, but swapped the famous celebrity faces with highly familiar faces, in the form of University of Kent lecturers (i.e. the single/key change between the two experiments). Whereas the participants in the first (Celebrity faces) experiment could be selected from the larger pool of all University of Kent student, the second (Lecturer faces) experiment was limited to senior PhD students, at the School of Computing, in order to ensure greater familiarity with the chosen Lecturers. In keeping with the first experiment, the presence of Lecturer faces was not divulged to the participants in the second experiment, and the Target stimulus was task-relevant. In the third and final experiment (see Chapter 7), we enhanced the Lecturer faces experiment by improving the design (see Chapter 6), and made a single change to the experiment: whilst retaining the task-relevant objective (i.e. to look for a Target stimulus), we revealed the presence of Lecturer faces in the RSVP streams. This extra information brings our experiments closer to the real-life scenario, whereby, the accused subject of an investigation will be made aware of the fact that compatriot faces may appear in the experiment.

3.2 Blueprint of Research

3.2.1 – Participants

This PhD study consisted of three experiments, and in each of our experiments we selected 14 (fourteen) participants, who were all students at the University of Kent, free from neurological disorders, and with normal, or corrected-to-normal vision. All subjects signed a consent form before participating in the experiment. The first experiment was advertised publicly, but the other two experiments were limited to School of Computing's PhD students, who were hand-picked, according to their level of familiarity with the department's lecturers. All subjects were given a monetary reward for participating in the experiments, and the Sciences Research Ethics Advisory Group, at the University of Kent, approved each study.

3.2.2 – Stimuli

For all three experiments, the instructions, stimuli and questions were presented on the same 20-inch LCD monitor, with a refresh rate of 60Hz, and a resolution of 1600x1200 pixels. The screen was placed at a comfortable position for each subject, at a distance of approximately 60 to 80cm. All stimuli were scaled to 280x320 pixels, and presented in the centre of the screen, using the Rapid Serial Visual Presentation (RSVP) method.

For all three experiments, the stimuli were split into two groups: a) **Distractor** images and b) **Critical** images, as described below:

a) **Distractor** images (i.e. the first group) were photographs of unfamiliar faces, which were obtained from an open-source, online database of faces (Minear & Park & Park, 2004) from the University of Texas at Dallas – all of these faces were frontal views. After the removal of unwanted images (e.g. those with significant facial expressions like wide grins), the resulting 524 faces were converted to monochrome (i.e. black-and-white), and scaled to 280x320 pixels. In all experiments, distractors were used as fillers.

b) **Critical** images (i.e. the second group) consisted of the following three categories:

i) **Target** face, which was a single image, chosen by us from the Distractors database. The Target was task-relevant (i.e. Subjects were instructed to look out for, and respond to this image).

ii) **Irrelevant** (a.k.a. *unknown*) faces consisted of several images that were not familiar to the subjects (note that for the first experiment only, Irrelevants were chosen, at random, from the Distractors database).

iii) **Probe** (a.k.a. *familiar*) faces consisted of several images of familiar faces, in the form of famous celebrity faces (in the first experiment), and our University's lecturer faces (in the next two experiments).

Great care was taken to ensure that all the images used in the three experiments would conform to our compatibility criteria. Accordingly, images with incongruous features (e.g. angry or smiley faces), which could breakthrough into conscious awareness due to their dissonant features, were avoided. As a result, the large database of Distractor faces (with over 1000 images) was carefully scrutinised and reduced to 573 'neutral' images, without significant facial expressions or features. Furthermore, all Distractor faces were centred and converted to monochrome (i.e. greyscales or black-and-white). The remaining 573 Distractor images were available as fillers for RSVP streams, albeit, five Distractors were randomly chosen as Irrelevants, in the first experiment, and one was selected as the Target, for all experiments.

The Probes for the first experiment (i.e. celebrity images) required careful selection and manipulation, to assure compatibility with the Distractor images. Having collected multiple pictures for each of the five pre-selected celebrities, we narrowed our selection to a single image (for each celebrity) which would conform to the same standards that had been applied to Distractor images. However, the Probes for the second and third experiments (i.e. familiar Lecturer faces from the University of Kent's School of Computing), as well as, the Irrelevants for the same two experiments (i.e. unknown Lecturer faces from Christ Church University) were taken using the same SLR

camera (Canon PowerShot G11). All lecturers consented to their image being used in our EEG experiments; these photos were taken from the same position/distance, under similar lighting conditions, and with neutral poses.

All images of celebrities and lectures were manually edited, in order to remove any non-conforming distinguishing features, and to obtain similar brightness and contrast. First, each image was centred by aligning the eye-line to the same horizontal position, and then resized, to occupy the same space/size as the Distractor images. Next, the background of each celebrity/lecturer face was removed (i.e. borders were carefully highlighted/selected, using the photo-editing tool GIMP, before being cropped out), and then the background colour was changed to light-grey (Hex colour: #e7e7e7). Next, the contrast of the images were reduced, wherever necessary, and all Probe images were resized to 280x320 pixels and converted to monochrome (i.e. grey-scale, or black-and-white). Due to the high quality of the Probe (celebrity) images, it was necessary to further reduce the contrast of the original photos (i.e. to bring them in-line with the contrast of the Irrelevant photos, which were taken from the Distractor database). Furthermore, a 'blur' tool was used, wherever necessary, to smear the edges of the celebrity's head/shoulder/hair, to reduce the sharp contrast with the cut-out background.

After the above exercise to select-and-edit our Probe images, we decided to further reduce the Distractor images to 524 faces, in order to approximately match the age range of the Probe faces (i.e. by excluding Distractor images of very young and very old individuals). As there were 524 possible Distractors, which could be used as fillers for RSVP streams, the probability that one would be randomly selected for each stream was 0.032, and equal for all of them.

3.2.2.1 – Probe/Irrelevant comparison in the first experiment

As explained above, the second and third experiments' Probe and Irrelevant critical stimuli were lecturer faces that were photographed by us, using the same camera and in accordance with a strict set of standards. However, the first experiment's Probes and Irrelevants came from different sources; the former was carefully selected from various online celebrity websites, whereas, the latter was randomly selected from our Distractor database. To demonstrate that there was no significant difference between the brightness and contrast properties of the Critical images (e.g. the Probes and Irrelevants

in the first experiment), we performed statistical analysis of the pixel intensities for each Probe and Irrelevant image. The mean, variance, skewness and kurtosis (i.e. the first four probability theory Moments of pixel intensities) for each image was evaluated, in both groups. As we had standardised our critical images to monochrome (i.e. black-and-white images), each pixel would be characterised by an intensity value of up to 256 different possible intensities, in order to represent the brightness of the pixel. Statistical tests were performed, in the form of two independent sample *t*-tests, between the pixel intensities of our Probe and Irrelevant critical stimuli. At an alpha level of 0.05, no significant differences were found between the physical properties of faces in the Probe and Irrelevant images of the first experiment (see Table 3.1, below).

Moment	Probe	Irrelevant	Two sample independent *t*-test
Mean	*M = 158.33* *SD = 12.02*	*M = 166.81* *SD = 12.30*	*T = 1.55* *P = 0.134*
Variance	*M = 3.04e+03* *SD = 1.2e+03*	*M = 2.5e+03* *SD – 1.0e+03*	*T = -0.96* *P = 0.349*
Skewness	*M = -1.06* *SD = 0.526*	*M = -1.47* *SD = 0.322*	*T = 2.07* *P = 0.117*
Kurtosis	*M = 4.21* *SD = 2.06*	*M = 3.16* *SD = 1.16*	*T = -1.4* *P = 0.178*

Table 3.1 – Outcome of statistical analysis of Pixel intensity of the first experiment's images for Probe (celebrity) and Irrelevant (Unknown/Distractor) stimui, confirming that no significant differences could be found between the physical properties of the two conditions that were being compared.

3.2.3 – Design

All stimuli were presented using the *Psychophysics* toolbox version 3, running under MATLAB 2012a. All RSVP stream items were presented in the same location (centre of the screen), at an SOA of 133ms (see Figure 3.1), and without an Inter-stimulus-Interval (ISI). Each RSVP stream contained 18 faces, 17 of which were

Distractors (i.e. unknown fillers), and only one was a Critical item (i.e. either a Probe, Irrelevant or Target). The Probes (i.e. famous faces in the first experiment, and lecturer faces in the other two experiments) were paired with Irrelevants (i.e. unknown faces), in order to make direct comparisons, and to perform statistical tests on the evidence gathered. To ensure that neither the Probes nor the Irrelevants were task-relevant, we instructed all subjects to look for a single Target (i.e. an unknown face that subjects were trained to detect), throughout the experiment. This task-relevant Target was repeated as many times as each of the other two Critical stimuli (i.e. the Target, Probe and Irrelevant conditions were repeated equal number of times).

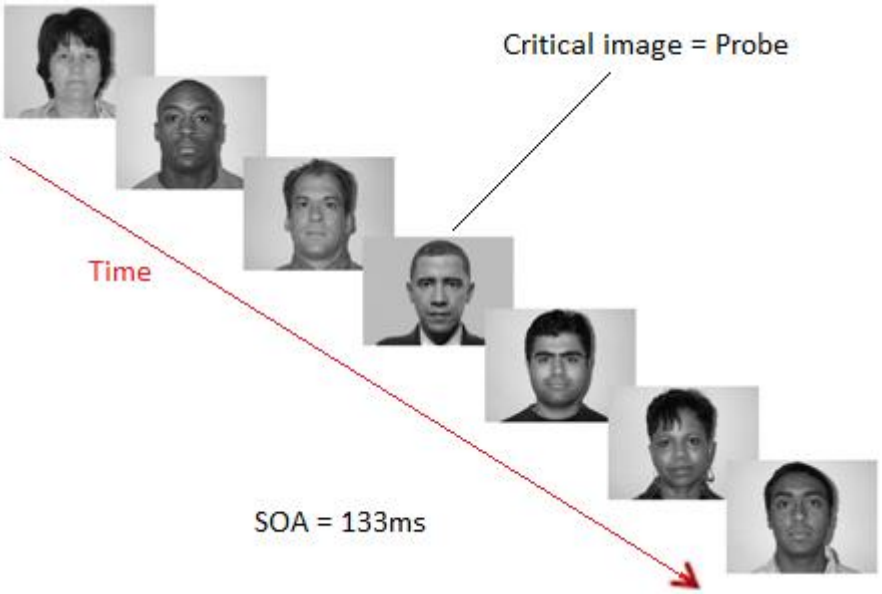

Figure 3.1 – RSVP stream, showing 7 of the 18 faces that could be presented in a trial, where each trial consists of 17 Distractors (i.e. unfamilar faces), and one Critical image (i.e. a Probe, Irrelevant or Target). In the above example from the first experiment, Barak Obama is the Probe that is presented as the Critical image. Note that in the second and third experiments, we used lecturer faces as Probes, instead of celebrity faces.

The position of the single Critical item within each RSVP trial, which contained 17 Distractors as fillers, was selected pseudo-randomly by the application, so that it had equal probability of appearing anywhere in the 5th to 9th position of the stream of images. The starting boundary (i.e. the first 4 items of the stream) was avoided because of onset transients, which produce overlapping EEG effects (Crevits L, 1982), as a

result of going from nothing to something on the screen. Similarly, the ending boundary (i.e. the last 9 items of the stream) was also avoided because of anticipatory transients, relating to the subject's anticipation of the end of stream item-and-question, such that, a transient component may overlap with the ongoing waveform that was produced by the Critical stimuli.

In addition to the 18 images, each RSVP stream contained a starting and finishing item, which would improve the subjects' focus, from the beginning to the end of each stream (i.e. to know when the stream is about to start, and when it will end). A starting item "+ + + + + + +" was presented for 800ms, to position the subject's focus on the presentation area of the screen. After the last image of the RSVP stream (i.e. the 18th face), a random finishing item (a.k.a. attention-checker image), which could either be "- - - - - - -" or "= = = = = = =", was presented for 133ms; this end-of-stream image required the subject to remain attentive until the end of the stream. Therefore, if the Critical item was perceived in the middle of the stream (i.e. randomly, between item 5 and 9 of the 18-item stream), the subject was expected to observe the remaining images, until they could see and identify the finishing item. Using a standalone keypad, which was positioned under the subject's preferred left or right hand, the subject was asked to report the attention-checker item, using '1' and '2' keys (1 signifying "- - - - - - -" and 2 signifying "= = = = = = =").

As soon as the subject responded to the attention-checker question, a task-relevant question was shown, in order to confirm the detection of the Target image, using the question: "Did you see the *Target* image within the stream?". In response, the subject could select the keys '4' (for "Yes") or '5' (for "No"). In all three experiments, the above two questions were repeated at the end of each RSVP stream, however, in the first experiment (only), an additional *recognition-question* was asked at the end of each block, to determine if the subject had observed and/or recognised the Probe or Irrelevant critical stimulus. In the other two experiments, we moved the end-of-block recognition-questions to the end of the experiment, in order to mitigate the subjects' inference that the Probe could also be task-relevant (i.e. in addition to the Target). Therefore, even if the subject perceived the Probe (or Irrelevant) images, they would not receive a mid-experiment hint of their relevance, until the end of the experiment.

In addition to briefing the subject on the required etiquette for EEG tests (e.g. no eye blinks during trials, as well as, sitting comfortably and very still, in order to avoid muscle artefacts), subjects were given one (or more) training session(s). Each training session consisted of 20 RSVP trials, to ensure that the subject is comfortable observing the rapid presentation of images (which may take a little time to get used to), and to make sure that they could identify the Target. Note that during the training session(s), the RSVP streams did not contain any images that were assigned to the Probe or Irrelevant category of critical stimuli, since we wanted the subject to remain naïve, with regards to the possibility of seeing the famous-or-Lecturer faces (as well as, the paired Irrelevant/unknown faces).

3.3 Analysis of Research

3.3.1 – Data acquisition

All three experiments were recorded using a BioSemi ActiveTwo system (*BioSemi, Amsterdam, The Netherlands; see www.biosemi.com*). The Electroencephalographic (EEG) data was filtered at recording, with a low-pass of 100 Hz, and digitised at 2,048 Hz, for offline analysis, and the impedance was kept below 10 kΩ. In accordance with the standard 10-20 system (Jasper, 1958), the following 8 scalp electrodes were used in the first two experiments: Fz, Cz, Pz, P3, P4, Oz, A1 and A2. However, in the third/final experiment, all 32 scalp electrodes were used (see Figure 3.2).

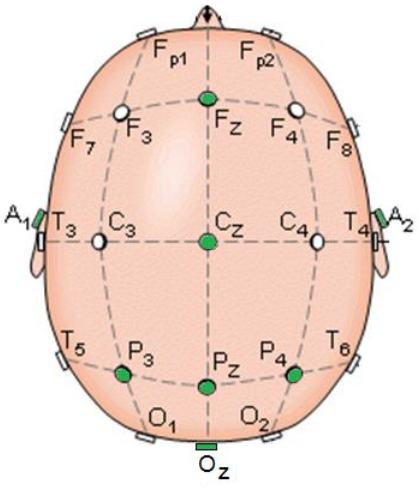

Figure 3.2 – Position of EEG electrodes, on the human scalp. Fz, Cz, Pz, P3, P4, Oz, A1 and A2 are highlighted in green.

During recording, data was referenced to a ground formed from a Common Mode Sense (CMS) active electrode, and Driven Right Leg (DRL) passive electrode. These two electrodes form a feedback loop, which drives the average potential of the

subject, as close as possible, to the Analogue to Digital Converter (ADC). Also, Electrooculograms (EOG) generated from eye blinks and eye movements, were recorded from the subject's left and right eyes, using two bipolar Horizontal EOG (HEOG) and Vertical EOG (VEOG) electrodes.

3.3.2 – EEG data

For all three experiments, the recorded data, which was analysed using EEGLAB (version 12.02.4b), under MATLAB 2012a (Delorme & Makeig, 2004), was resampled at 512 Hz. Before analysis, we filtered the data with a low-pass, and high-pass, filters of 45 Hz to 0.5 Hz, respectively. Furthermore, in order to remove the steady state Visually Evoked Potential (ssVEP) oscillations (Wang & Jung, 2010), notch filters were applied between 7 Hz and 9 Hz. Then, we off-line referenced the data to the average of the combined mastoids (i.e. A1 and A2 electrodes), and generated ERPs by separately averaging all trials for each Critical condition (i.e. Target, Probe, and Irrelevant). Each EEG trial was generated by epoching the data, using -200ms to 1200ms stimulus-locked window
(i.e. -200ms before the appearance of the Critical stimulus, and 1200ms after the occurrence of the Critical stimulus), and all ERPs were time-locked to the onset of a Critical item (i.e. time-point zero marks the appearance of the Critical face). Although baseline correction could be applied at trial level (i.e. mean of -100ms to 0ms window subtracted from each trail), the new standard for applying the detrending technique (see 4.4.2.3 – Application of Detrending) required us to baseline correct each trial after the adjustment for any errant drift.

Eye blinks and muscle movements were detected, by marking any activity below -100μV or above +100μV in the EOG channels (reflecting eye blinks and horizontal/vertical movements). Furthermore, we rejected any trials containing electrical activity below -50μV or above +50μV, in a time window from -200ms to 1200ms (reflecting other physiological and environmental artefacts), with respect to the Critical stimulus Onset. Finally, we performed manual inspection of the resulting ERP data, to verify that the rejected trials were accurately detected.

3.3.3 – Time Domain (ERP) Analysis

Each experiment's trails can be broadly split into three categories: those with the *Target* critical stimulus (i.e. trials that contained the task-specific face that the subject was asked to detect and report), ones with *Probe* critical stimulus (i.e. trials that contained a Celebrity face) and the ones with *Irrelevant* critical stimulus (i.e. trials that contained an Unknown/Distractor face, which was paired with the Probe and appeared the same number of times). As typical of ERP-based deception detection studies, our analyses were performed at the ERP-level, and the primary goal was to compare the EEG responses to-and-between familiar (Probe) and unfamiliar (Irrelevant) faces.

Within the Time Domain (i.e. study of ERPs), we have used mean amplitude measurement for our analysis, as it is more robust against high frequency noise (Luck S. , 2005), and it has been used in previous studies into familiar and unfamiliar faces (Curan & Hancock, 2007); (Eimer, 2000); (Touryan, 2011). As previously discussed, these and other studies, like (Bentin & Deouell, 2000), into recognition of familiar faces have reported two features/components, prominently observed within ERP patterns:

i) an enhanced (early) negativity called **N400** – also referred to as **N400f** (Curan & Hancock, 2007) – within a 250ms to 500ms search range;

ii) a late positivity called **P3** – also referred to as **P600f** (Trenner, Jentzsch, & Sommer, 2004) – within a 300ms to 900ms search range.

The time window associated with the above two ERP components should be identified based on an independent contrast, rather than eye-balling the ERP plot. Specifically, the window placement must be made independently of the contrast that is statistically tested (Kilner, 2013). Thus, for subject-level analysis, we selected the time window using the subject's aggregated ERP, generated from all trials (hereafter called **aERPt**), within the combined Probe-and-Irrelevant conditions; this is also called the

aggregated ERP of trials (Brooks, Zoumpoulaki, & Bowman, 2017). Similarly, for group-level analysis, we selected the time window using an Aggregated Grand Average from all Trials (**AGAT**), belonging to all subjects' combined Probe-and-Irrelevant conditions.

Using the above techniques for time window placement (i.e. aERPt for subject-level, and AGAT for group-level), there is no question of 'looking' for the conditions that show a big effect (i.e. fishing in the Probe), which would inflate the false positive rate. Studies (Kilner, 2013) have pointed out this inflation in Type I errors, which increases the probability of detecting an effect under the null. The aERPt and AGAT techniques, for data-driven (safe) window selection, have been fully investigated and justified in (Brooks, Zoumpoulaki, & Bowman, 2017), and thus resolve the problem of non-orthogonality arising from trial count asymmetry, identified in (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009). Having established this precedent, the size of the window (i.e. mean amplitude of the region of interest) would be quantified by searching for, and finding, a window with the highest (for positive features) and/or the lowest (for negative features) mean amplitude, when the paired Probe and Irrelevant conditions are combined, using the aERPt (aggregated ERP of trials, at subject-level), or AGAT (aggregated grand average of trials of all subjects, for group-level) methods. The following describes both methods, in greater detail:

3.3.3.1 - Subject-level (aERPt) window placement

For each subject, their aggregated ERP of all trials for both conditions, Probe and Irrelevant, were collected. This aERPt was then used to identify the time window of the two components of interest (i.e. N400f and P600f). For the N400f component, an algorithm searched from the lower boundary to the upper boundary (i.e. 300ms to 500ms post-stimulus, or the entire ERP window, if we wanted an independent, non a-priori contrast), to find the *minimal* 100ms interval average. Similarly, for the P600f component, the algorithm searched automatically, from the lower boundary to the upper boundary (i.e. 500ms to 800ms post-stimulus, or the entire ERP window, if we wanted an independent, non a-priori contrast), to find the *maximal* 100ms interval

average. The start and the end of this minimal/maximal 100ms Region of Interest (ROI) defined the face related N400f and P600f time features/components.

Although the search windows that we could have employed, for the minimal/maximal interval averages, were available from studies that measured ERPs to familiar and unfamiliar faces (Touryan, 2011); (Eimer, 2000), we elected to expand our search parameters, to the entire ERP window (i.e. from 0ms to 1200ms), which would give us a non a-priori/independent contrast, that would present no limits to the automatic selection of the true (and unbiased) ROI.

After defining the time windows for each component (e.g. the ROI for N400f, and the ROI for P600f), the mean amplitude measure was applied separately to each condition within the defined time window – in other words, for each component, one mean amplitude value for the Probe, and another for the Irrelevant, was calculated using the same time window that was independently found when Probe and Irrelevant trials were combined. It could be said that the 'True Observed' difference of each component (i.e. N400f and P600f) in their respective ROI, is the difference between this measure for each condition:

- ***∴ True Observed difference = mean of Probe (minus) mean of Irrelevant***

Having found the True Observed difference for N400f and P600f, we were able to perform statistical analyses of the ERP data, to determine whether the evoked response by the Probe (e.g. the famous face) was significantly different from that evoked by the Irrelevant (i.e. the unknown face). Individual, or subject-level, analysis is based on analysing each experimental participant separately; that is, to determine whether there was a significant difference for that subject alone. The null hypothesis (H0) was that there is no difference between the Probe and Irrelevant patterns, for each subject. Our experimental hypothesis is that H0 can be rejected. In this analysis, a randomisation (i.e. Monte Carlo Permutation) test was used to define a *p*-value for each subject (see section 3.3.3.3, below). A null hypothesis distribution for each subject was generated in order to calculate the individual's *p*-value; the *p*-value would determine the probability that the observed pattern could have arisen if the null hypothesis was true. This is a reliable way to assess each subject's patterns individually, and to determine that subject's significance (i.e. is s/he guilty?).

3.3.3.2 – Group-level (AGAT) window placement

Time window placement for the group starts with the collation of all the trials for all subjects, in both conditions (Probe and Irrelevant). The resultant Aggregated Grand Average of Trials (AGAT) would then be used to identify the time window of the two components of interest (i.e. N400f and P600f). For the N400f component, an algorithm searched automatically, to find the *minimal* 100ms interval average, and for the P600f component, the algorithm searched automatically, to find the *maximal* 100ms interval average. The start and the end of this minimal/maximal 100ms Region of Interest (ROI) defined the group-level face related N400f and P600f features/components, for both conditions.

In keeping with the aERPt (subject-level) method, the search windows that we could have employed, for the minimal/maximal interval averages, were available from studies that measured ERPs to familiar and unfamiliar faces (Touryan, 2011); (Eimer, 2000). However, we elected to expand our search parameters, to the entire ERP window (i.e. from 0ms to 1200ms), for a non a-priori/independent contrast, which would present no limits to the automatic selection of the true (and unbiased) ROI. Finally, the mean amplitude measure was applied separately to each condition, within the defined time window, and the 'True Observed' difference of each component (i.e. N400f and P600f) was obtained by finding the difference between each condition (i.e. same as the aERPt *True Observed difference* calculation, described in section 3.3.3.1).

Having found the True Observed difference for N400f and P600f, we were able to perform statistical analyses of the data, to determine whether the evoked response by the Probe (e.g. the famous face) was significantly different from that evoked by the Irrelevant (i.e. the unknown face). Of course, group-level analysis denotes the significance across the whole set of individuals, so a *t*-test was used to determine whether there is a statistically significant difference between Probe and Irrelevant patterns, for the whole group of subjects. The null hypothesis (H0) was that there is no difference between the two patterns. Our experimental hypothesis is that H0 can be

rejected. In this analysis, a paired *t*-test of the mean amplitudes of Probe N400f/P600f and Irrelevant N400f/P600f was used, across all participants.

3.3.3.3 – Randomisation test

At individual/subject-level, to determine whether the difference between two conditions (Probe and Irrelevant) is significant – in other words, to draw conclusions as to whether the evoked EEG data by the familiar face was significantly different from that evoked by the unfamiliar face – we applied a randomisation (i.e. Monte Carlo Permutations) test. This was done separately for N400f and P600f components, in order to generate a null hypothesis distribution, for each subject. Before applying the test at subject-level, the smallest number of trials between the Probe and Irrelevant conditions was determined and denoted by '*m*' (note that, the Probe and Irrelevant conditions could contain different numbers of trials due to their respective artefact rejection). Thus, only '*m*' trials were selected (at random without replacement) from the Probe condition, and '*m*' trials from the Irrelevant condition. Notably, if a direct comparison was to be made between individual blocks (as relating to a single famous/celebrity face), we made sure that the pairing of the Probe and Irrelevant conditions were maintained. Next, we calculated the difference between the mean amplitude values of Probe and Irrelevant ERPs (*Probe [minus] Irrelevant*), in order to obtain a mean amplitude difference measure. This mean amplitude difference became the **True Observed Value**.

The randomisation test was applied by populating a matrix of size (*2.m* × *number of time points*) with *2.m* selected trials; row position was randomised in the matrix. Under the null hypothesis, the Irrelevant and Probe trials are samples from the same distribution (i.e. the null distribution), and would thus be exchangeable. This justifies the randomisation of position in the matrix. Next, a pair of datasets were generated: the first, the surrogate Probe, was generated from the first half of the matrix rows, and the second, the surrogate Irrelevant, was generated from the remaining half. The desired analysis (i.e. mean amplitude measure) was then applied to each of the two randomised data sets, and the mean amplitude value was calculated (referred to as the **Surrogate Values**), in the same way that the True Observed Value was calculated. In

other words, mean amplitude of N400f in surrogate Probe minus mean amplitude of N400f in surrogate Irrelevant; also, mean amplitude of P600f in surrogate Probe minus mean amplitude of P600f in surrogate Irrelevant.

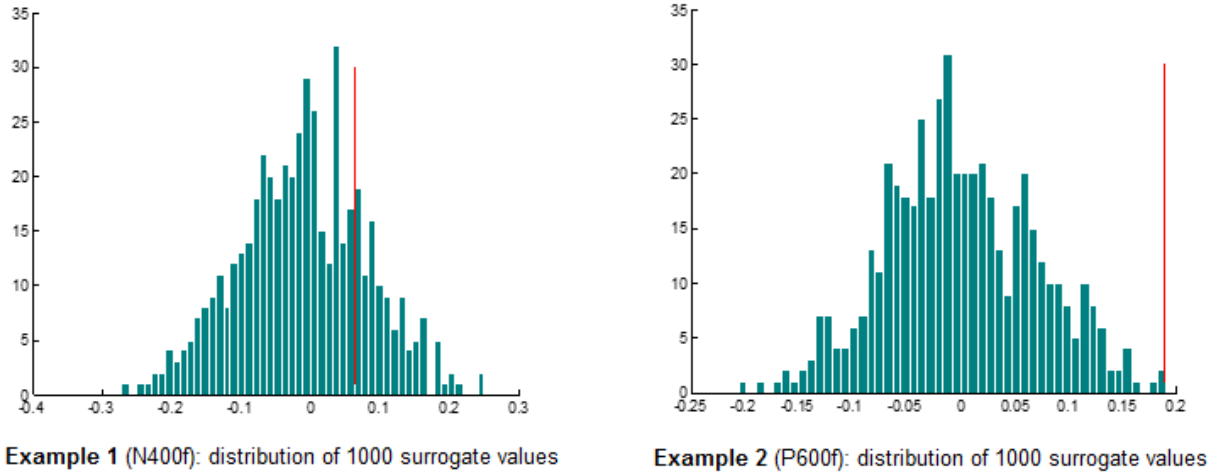

Figure 3.3 – Two examples of representing the null hypothesis distributions of randomisation procedure's 1,000 Surrogate Values (split into 50-bins, and shown as turquoise bars, for each plot), and the corresponding True Observed Value (red line), for N400f and P600f components (x-axis represents the count, and y-axis represents the probability). Example 1 (left plot) shows that the True Observed Value could not reject the null hypothesis (p = 0.206), whereas, example 2 (right plot) shows that the True Observed Value falls outside the null hypothesis distribution, resulting in a significant p-value (p = 0.02).

The above randomisation procedure was repeated 1,000 times. In each iteration, a new mean amplitude difference was obtained, resulting in 1,000 *Surrogate Values*, which act as a contrast to the single *True Observed Value*. The *p*-value was then calculated as the proportion of randomised results that were greater than the true observed value (see Figure 3.3). Finally, if this *p*-value is smaller than a critical alpha-level (0.05), then the data in the two experimental conditions are significantly different, thus, the null Hypothesis (H0) can be rejected. Note that in each resampling, the randomised mean amplitude difference (i.e. surrogate value) was measured for both the N400f and the P600f components (i.e. these values were calculated from the same random sample, rather than being calculated form two separate randomisations).

3.3.4 – Combined probability test (Fisher's)

Having determined the *p*-values for both evoked components (i.e. N400f and P600f), of each individual subject, we were minded to generate a single measurement that could be used to infer the subject's recognition of the familiar face – after all, in real-life applications (e.g. concealed information test), we would have to judge whether the subject is guilty or not. In fact, we have used similar procedures to combine EEG data from multiple electrodes (e.g. Fz, Cz and Pz), resulting in a single combined *p*-value, that could be employed to distinguish between deceivers and non-deceivers (Bowman, et al., 2013). Thus, we used the Fisher combined probability test (Fisher, 1932) to calculate a joint *p*-value across N400f and P600f, for each subject. As discussed in (Hayasaka, 2004), the Fisher method treats the different dimensions consistently, since combining *p*-values in this way automatically normalises into a common comparable measure. The Fisher procedure can be viewed as a non-parametric statistical inference method for handling multi-variate dependent measures. Parametric methods (e.g. Hotelling's T-squared, or multivariate analysis of variance, MANOVA) are unsuitable for subject-level ERP analysis because it is often difficult to robustly measure the variable of interest from a single trial, due to excessive noise (i.e. low signal-to-noise ratio). Therefore, the use of a resampling method, such as a Permutation test, combined with the Fisher procedure would safely extend the analysis to the multivariate case.

Using the Fisher combining function (Hayasaka, 2004), we calculated the True Observed Fisher Value (TOFV), by multiplying the two *p*-values obtained from the N400f (`p_N400f`) and P600f (`p_P600f`) null hypothesis distributions, for each individual subject (*sub*) in the experiment:

```
TOFV(sub) = (-2 * (log ( p_N400f(sub) * p_P600f(sub) ) ) );
```

Then, to determine the combined *p*-value of a single subject, across both N400f and P600f components, we calculated 1,000 points of randomised *p*-values for each of

these two components. The resultant 2,000 *p*-values were used to calculate the Fisher Value of Randomisation (FVR) for the number of randomisations (denoted by 'i') that were performed:

```
FVR(sub).i = (-2 * (log ( p_N400f(sub).i * p_P600f(sub).i ) ) );
```

Finally, the number of FVRs that were greater than the single TOFV, divided by 1,000 would be the Fisher *p*-value. Note that when calculating the Fisher score, values of *p* that equal to zero were replaced with the smallest legitimate *p*-value (i.e. 0.001), to avoid the formula returning infinity.

3.3.5 – Frequency Domain Analysis

Although ERP averaging is very useful in mitigating the excessive noise in single trials, it has its weaknesses, as out-of-phase increases in power across single trials (i.e. induced responses) may be missed by ERP analysis (Makeig, Debener, Onton, & Delorme, 2004b). However, Time Frequency analysis does not have this weakness, as power and coherence are analysed across trials and less information is lost (Van Vugt, Sederberg, & Kahana, 2007). To analyse EEG data in the time-frequency domain, the following two transforms were used: Event Related Spectral Perturbations (ERSP) and Inter-Trial Coherence (ITC). These were calculated, using a fast Fourier transform, with a baseline correction of -100ms to 0ms.

ERSP calculates the average changes, relative to baseline, in the frequency power spectrum at each time point, across all individual trials that are time-locked to the same stimulus (Delorme & Makeig, 2004). ITC measures phase consistency between trials, determining the extent to which individual trials are phase-locked, at each time point and frequency range (Makeig, Debener, Onton, & Delorme, 2004b).

Previous studies into familiar faces (Bentin & Deouell, 2000) have reported ongoing oscillations in ERPs, from about 100ms to 500ms post stimulus onset, over

parietal and occipital sites. In the famous faces experiment reported here (see chapter 4), we also observe multi-cycle oscillations in grand-averaged ERPs of the Probe (celebrity face), which are not present in the Irrelevant (unknown face), and, importantly, not in the task-critical Target (i.e. the unknown face that the subject was trained to respond to). Because classic ERP analysis methods, like peak-to-peak or base-to-peak, would not fully reflect or measure the Probe's multi-cycle oscillations, time frequency analyses (ERSP and ITC) were used, as outlined below.

3.3.5.1 – Time Frequency Window Placement

Time frequency analyses were measured, over two time windows, using orthogonal contrast time window placement, in relation to the contrast that is statistically tested (Kilner, 2013); (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009). In a similar way to our ERP analysis of the time domain, the time window for the Region of Interest that we used to measure ERSP and ITC, was identified based on aggregated power and coherence.

For group-level Time Frequency analysis, the placement of the critical time window (i.e. the highest 100ms interval in the broader time window of 0ms to 1200ms, post-stimulus) for measuring ERSP/ITC was calculated using the average of power/coherence of all single trials of all subjects (i.e. the aggregated grand average of all trials, across all subjects) from both Probe and Irrelevant conditions. For subject-level Time Frequency analysis, the placement of the critical time window (i.e. the highest 100ms interval in the broader time window of 0ms to 1200ms, post-stimulus) for measuring ERSP/ITC was calculated using each individual subject's average of power/coherence (i.e. the aggregated ERP of all trials, for a single subject) from both Probe and Irrelevant conditions. Thus, both methods for time window placement were calculated independently of the contrast that is statistically tested.

Next, these orthogonally derived time windows could be employed to measure ERSP and ITC separately, in the Probe and Irrelevant conditions. In keeping with previous studies (Delorme & Makeig, 2004), the EEGLAB time-frequency function *newtimef* was used to calculate the ERSP and ITC for each condition. Each condition (i.e. Probe and Irrelevant) would supply this function with a matrix that contains its

respective time-points and trials. The *newtimef* function would process these two input matrices and calculate two output matrices, which represent the difference in the power and coherence; the first output matrix comprised the difference in power (i.e. ERSP) between Probe and Irrelevant conditions, and the second comprised the difference in coherence (i.e. ITC) between Probe and Irrelevant conditions.

3.3.5.2 – Time Frequency Statistical test

By taking the sum of all the values that were greater than zero, in the available frequency range, a single difference measurement was obtained for the ERSP and ITC transforms. Note, the assumption exists that high values of ERSP and ITC indicate the existence of evoked and induced activity that the procedure aimed to detect. The above summation process resulted in two difference measures – one for power (i.e. ERSP) and another for coherence (i.e. ITC) – which would become the True Observed Values of our transforms, and used to statistically calculate *p*-values for ERSP and ITC. Just as we had done in time domain analysis of ERPs, we used a randomisation (Monte Carlo permutation) procedure to generate two Null hypothesis distributions for power and coherence transforms (as calculated by the summation process, outlined above, across the orthogonally derived time windows). For each subject, we calculated *p*-values for both power and coherence transforms, and then utilised the Fisher combining procedure to combine them into a single *p*-value for that individual.

3.4 Conclusion

Having described, in detail, the blueprint for the design and analysis of all three face recognition studies, we will, hereafter, reference and apply the above general research aims, concepts, and methods, in chapters 4, 5 and 7.

Chapter 4:

EEG study 1 – Breakthrough of Celebrity Faces in RSVP

4.1 Introduction

The objective of this chapter was to reconcile recent EEG-based studies into concealed information tests (Labkovsky & Rosenfeld, 2012), with our own RSVP-based countermeasure-resistant fringe-P3 methods, which could be used to differentiate between deceivers and non-deceivers (Bowman, Filetti, Alsufyani, Janssen, & Su, 2014), by introducing a new category of critical stimuli, in the form of human faces. The aim of this work was to provide a proof of concept, which could be further refined and developed into a scientifically robust framework, in the pursuit of a means by which a suspect's familiarity with compatriots/deceivers can be demonstrated using EEG tests.

By referencing the standard design and analysis methods described in Chapter 3, we will begin by outlining the celebrity faces experiment (i.e. hypotheses, design and behavioural results), and then summarise the group-level analysis; at that point, we will introduce a *detrending* technique for dealing with EEG drift, as an alternative to high-pass filtering, and justify its application throughout our research. Next, we will describe our in-depth group and subject level analyses, in the Time (ERP) domain, as well as, the Frequency (ERSP/ITC) domain. Finally, we will discuss the results and draw conclusions to our hypotheses, based on the evidence gathered.

4.2 Experiment's Hypotheses

In pursuit of exploring the suitability of the RSVP paradigm, to infer the recognition of familiar/compatriot's faces, in real-life EEG-based deception detection tests, we started by substituting the existing *words/letters* based lie-detection studies with famous celebrity *faces*, in order to test the following hypotheses, experimentally:

i) Human faces can be used in a broader range of stimuli, to infer the recognition of familiar faces, using the RSVP paradigm, and the fringe-P3 method can be employed to detect the group-level breakthrough of Probe (celebrity) faces, which are differentially perceived and processed, as compared to Irrelevant (unfamiliar) faces;

ii) In addition to the breakthrough of Probe faces at group-level, we can use ERPs to detect the breakthrough events on an individual basis, even though, subjects were not instructed to look for the Probe conditions (i.e. only the Target was task-relevant);

iii) In keeping with previous ERP-based RSVP experiments, the strongest brain responses to the familiar (Probe) faces are recorded at the Pz electrode site.

4.3 Design of the first Experiment

4.3.1 – Experiment's Participants

Fourteen participants were tested and none were excluded. Out of 14 subjects, 5 were male (36%) and 9 female (64%). The ages of the subjects ranged from 18 to 26 (M = 20.5 years, SD = 2.029); 13 of them were right-handed (93%), and one was left-handed (7%). All subjects were students at the University of Kent, who responded to a public advert to participate in our EEG experiment. The duration of each experiment was (approx.) 1 hour and 30 minutes, and each subject was paid £10 (ten pounds) for their time.

4.3.2 – Experiment's Stimuli

As outlined in Chapter 3 (see section 3.2.2), the stimuli were split into two groups: **Distractors** (i.e. 524 unknown faces) and **Critical** images. The Critical group was further split into 3 categories: **Target** image (a single face that became task-relevant), **Irrelevant** images (five unknown faces) and **Probe** images (five famous

celebrity faces). The Probe faces were hand-picked, as they were widely known to the general public, in accordance with the top-20 searched celebrities in 2014 (using Yahoo's UK-specific search engine). We chose our five famous people from this top-20 list, based on the suitability of the available photographs of the celebrities, which would conform to our database of Distractors (i.e. ensuring that celebrity's images did not have significant facial expressions). The chosen five famous faces were: Angelina Jolie, Barack Obama, David Beckham, Justin Bieber and Leonardo DiCaprio (see Figure 4.1).

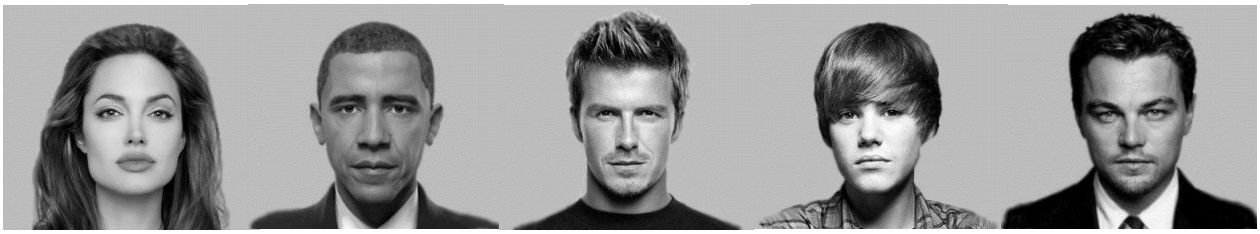

Figure 4.1 – Probes faces (from left to right):
Angelina Jolie, Barack Obama, David Beckham, Justin Bieber & Leonardo DiCaprio

Other than the subject's prior familiarity with the Probe (i.e. famous/celebrity) faces, it should be noted that our chosen celebrities were collectively considered to be more attractive than the average face in our Distractor database. It has been argued that stimuli with attractive features have an (evolutionary) attentional capture-and-processing advantage over unattractive stimuli (Silva, 2016). Even though the RSVP method was not used in the referenced study that compared the attentional advantage of attractive faces, this and other studies (Olson & Marshuetz, 2005) & (Willis, 2006) have shown that facial attractiveness can be extracted with minimum conscious endeavour, in as little as 13ms of presentation. Whilst acknowledging that our first experiment's *attractive* faces may benefit from a breakthrough advantage, future experiments (see Chapters 5 and 7) will counter any criticism, as they will not include celebrities.

4.3.3 – Experiment's design

As outlined in Chapter 3 (see section 3.2.3), each RSVP stream's 18 faces included a single Critical stimulus and 17 Distractors (with an SOA of 133ms). The

Critical stimuli in each RSVP stream could either be a Probe (i.e. one of five celebrity faces), or an Irrelevant (i.e. one of five unknown faces), or the Target (i.e. the same face that is task-relevant).

In total, Probes, Irrelevants and Targets were presented an equal number of times, and (in a statistical sense) in the same position in streams. In this (first) experiment, each Probe was repeated 15 times, resulting in 75 Probe-trials (i.e. 15 times for each of the 5 Probes), and each Irrelevant was also repeated 15 times, resulting in 75 Irrelevant-trials (i.e. 15 times for each of the 5 Irrelevants). The single Target was, therefore, repeated 75 times, to equal the number of times that the other two Critical Stimuli category were included in RSVP streams. The resultant 225 RSVP trials were divided into 5 blocks, each block comprising 45 trials (i.e. 15 Probe trials, 15 Irrelevant trials, and 15 Target trials), and the order of the three Critical stimuli were randomised within the blocks. However, each block's Probe and Irrelevant Critical stimuli were paired, so that the same celebrity (Probe) face and unknown (Irrelevant) face were presented within the same block – this will enable us to make direct comparisons between these paired-conditions.

Finally, subjects were told to keep their eyes fixed at the centre of the screen during the presentation of the RSVP stream (lasting 2.5 seconds), and to avoid movement or blinking. Also, they were informed that the Target image will appear pseudo-randomly, so they should not expect it in every trial, however, subjects were naïve to the presence of famous celebrity faces (i.e. Probes).

In this experiment, out of a total of 75 trials for each Critical condition, the number of trials that remained after artefact rejection, per condition, ranged between 59 and 73, and none of the subjects were excluded from the analysis due to removal of artefact trials:

Target (M = 68.91, SD = 5.64);
Probe (M = 71.72, SD = 3.53);
Irrelevant (M = 71.57, SD = 4.21).

4.3.4 – Experiment's Target Questions

As explained in Chapter 3, at the end of each RSVP stream, the subject was required to answer two question (see section 3.2.3), using a dedicated keypad, which was placed under the subject's right or left hand (whichever hand the subject preferred to use). The first question related to the finishing-item, which required the subject to select either key '1' or '2', and the second question related to Target-recognition, which could be answered using either key '4' or '5'.

Before starting the experiment, the subject was shown the Target image – this would be the same image for all subjects – which was chosen from the Distractor (i.e. unknown) database, and therefore, not familiar to the subject. Even so, the subject was asked, in the beginning, if they had ever seen, or could recognise, the Target face (none of our subjects had ever seen the Target face). As this is a task-based experiment, the subject was instructed to look only for that Target image, in each of the RSVP streams, and to expect a recognition question: "Did you see the Target face?", at the end of each trial (noting that this recognition question followed the finishing-item question). If the Target image was seen, the subject was instructed to answer 'Yes' (using '4' key), or 'No' if it was not perceived (using '5' key). If the Target was present, a 'Yes' (i.e. correct) answer would be a "HIT", and a 'No' (i.e. incorrect) answer would be a "MISS". Conversely, if the Target was absent, a 'Yes' (i.e. incorrect) answer would be a False-positive (FP), and a 'No' would be a correct rejection (see Table 4.1).

Out of 75 times that each subject was randomly presented with the Target face, the average Hit rate for the group was 81.4% (M = 61.07, SD = 7.89), and out of the remaining 150 other trials in which the Target was not

Target	HIT (75)	FP (150)
Subject 1:	75	11
	100.0%	7.3%
Subject 2:	64	4
	85.3%	2.7%
Subject 3:	57	9
	76.0%	6.0%
Subject 4:	70	13
	93.3%	8.7%
Subject 5:	56	11
	74.7%	7.3%
Subject 6:	66	23
	88.0%	15.3%
Subject 7:	46	12
	61.3%	8.0%
Subject 8:	66	7
	88.0%	4.7%
Subject 9:	66	5
	88.0%	3.3%
Subject 10:	50	7
	66.7%	4.7%
Subject 11:	61	9
	81.3%	6.0%
Subject 12:	58	15
	77.3%	10.0%
Subject 13:	65	23
	86.7%	15.3%
Subject 14:	55	28
	73.3%	18.7%
Mean:	61.07	*12.64*
	81.4%	8.4%
SD:	7.89	7.26

Table 4.1 – Subjects' HIT count (i.e. number of times that the subject correctly reported seeing the task-relevant Target face, in 75 trials), and False-Positive (FP) count (i.e. reported seeing the Target when it was not there, in the other 150 trials).

Group HIT rate of 61.1 (81.4%) and FP rate of 12.6 (8.4%), with corresponding MISS rate of 13.9 (18.6%) and correct rejection of 137.4 (91.6%), result in a response sensitivity measure of d' = 2.28.

presented, the False-Positive rate was 8.4% (M = 12.64, SD = 7.26). The resulting sensitivity measure (d' = 2.283) was within our tolerance range, and no subjects were excluded due to low sensitivity or high bias.

4.3.5 – Experiment's Probe/Irrelevant Questions

At the end of each block, the subject was given an additional recognition test, in the form of memory questions, to determine if the Probe or Irrelevant images were perceived/recognised (over and above the Target image). This end-of-block memory test consisted of four questions, appearing randomly, where each question accompanied an image that may or may not have been included in that block. Two questions related to the presence of the paired Probe and Irrelevant faces that were included in that block, and the other two questions related to a random Probe face and a random Irrelevant face that were not included in that block of the experiment. Whereas the former two questions (about the Probe/Irrelevant faces that were presented) would gauge the subject's ability to perceive faces that were included in that block, the latter two questions assess the subject's engagement with the tests (i.e. were subjects guessing the presence of salient faces?).

The online response to each of these four recognition/memory tests were handled in two parts: firstly, what is the subject's confidence rating of how often each of the 4 faces were presented (i.e. the Probe/Irrelevant that were present, and the Probe/Irrelevant that was absent), and secondly, a confidence rating of how well the subject knew each of the four faces, prior to the experiment. The responses to both of these confidence ratings used a scale of 1 to 5, where 1 is "Never", 2 is "Once or twice", 3 is "Few times", 4 is "Many times" and 5 is "A lot". Note, for the purposes of statistical comparison, 1 out of 5 (i.e. Never) is equivalent to 0% and 5 out of 5 (i.e. A lot) is equivalent to 100%. Thus, 2 out 5 = 25%, 3 out of 5 = 50% and 4 out of 5 = 75% (see Appendix A.1 for the full set of results).

4.3.5.1 – Overall Probe/Irrelevant recognition

The five Probe (celebrity) faces that were included in the experiment were reported to have been seen 60% of the time (Mean confidence rating of 3.4 out of 5), and subjects reported a high (pre-experimental) familiarity of 88.2% with these celebrity faces (4.5 out of 5). When comparing this to the (absent) Probe faces that were not included in the experiment, subjects reported a similar high (pre-experimental) familiarity of 83.6% (4.3 out of 5), and only reported seeing these 'absent' celebrities 19.6% of the time (1.8 out of 5), which is less than one-third of the celebrities that were included in the experiment.

The five Irrelevant (unknown) faces that were included in the experiment were reported to have been seen 10% of the time (1.4 out of 5), and, similarly, the absent Irrelevant faces that were not included in the experiment were reported to have been seen, at an average of 4.3% of the time (1.2 out of 5). Finally, subjects reported an imperceptible (pre-experimental) familiarity of 0% with all the Irrelevant/distractor faces (1.0 out of 5).

As we were comparing Probe faces with Irrelevant faces, it was encouraging to discover that Probes were reported 60% of the time ($M = 3.4$; $SD = 0.8771$), which was six times more than Irrelevants that were reported 10% of the time ($M = 1.4$; $SD = 0.532$). Note that both conditions (Probes and Irrelevants) were, in fact, presented an equal number of times. The mean confidence rating of the main comparison conditions, for all subjects, reveals a highly significant difference between the Probe (celebrity) faces and the Irrelevant (unknown) faces, using pair-wise comparison ($M = 2$, $SD = 0.8629$), $t(13) = 8.6722$, $p < 0.0001$, $d = 2.7572$).

4.3.5.2 – By-item Probe recognition

As the same five Probe (celebrity) faces were shown to all 14 subjects, we were able to draw 'by-item' comparisons between the Probes (see Figure 4.2, below). The least detected celebrity face was block-1's Jolie (53.6%), even though, this Probe was the second most recognised celebrity (94.6%). However, the first block possesses two

disadvantages: firstly, the subject is unaware of the possibility of a *celebrity* face within the first block, whereas, s/he is likely to infer the presence of more celebrity faces, in future blocks; secondly, as a result of a training effect, the greater the exposure to RSVP streams, the more likely it is that the subject will perceive the salient Probes in future blocks (e.g. the most detected celebrity face was block-4's Bieber (66.1%), even though, he was the least recognised celebrity (73.2%)). Note that similar improvements in detection/recognition of Irrelevant faces was not evident.

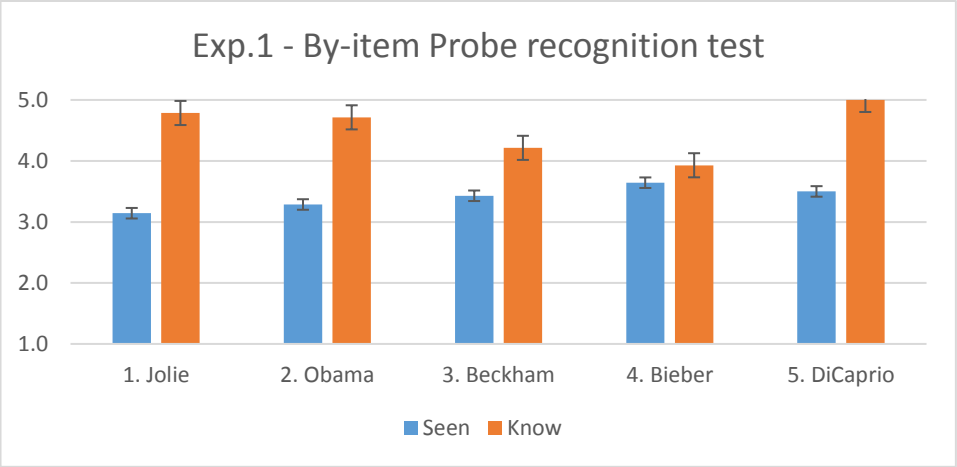

*Figure 4.2 – Experiment 1's by-item Probe (celebrity face) recognition tests: "**Seen**" rates (i.e. confidence rating of having detected the Probe) and "**Know**" rates (i.e. how well the subject recognises the Probe). On average, 54% of subjects had seen Jolie (rating = 3.1, SE:0.29), 57% had seen Obama (rating = 3.3; SE: 0.4), 61% had seen Beckham (rating = 3.4; SE: 0.37), 66% had seen Bieber (rating = 3.6; SE: 0.31), and 63% had seen DiCaprio (rating = 3.5; SE: 0.25). One-way ANOVA on the 'seen' ratings for the five celebrities confirms that there is no statistically significant difference between the means (p = 0.8477). As expected, subjects' familiarity (i.e. 'Know' ratings) with all five presented celebrities was very high (see Appendix A.1 for more detail).*

The behavioural data (i.e. all the above online responses to recognition questions) provided a useful indicator of the perceptual state of the subjects' mind, however, the primary aim of our research was to use the EEG data to detect the breakthrough of Probe (celebrity) faces, which could be differentially perceived and processed, as compared to Irrelevant (unfamiliar) faces. Therefore, the rest of this chapter will focus on the analysis of the EEG data, in the Time and Frequency domains.

4.4 Data Analyses

4.4.1 – Summary of Analysis

Although we were interested in the EEG data across all the midline electrodes (Pz, Cz and Fz), in-line with (Kaufmann, Schulz, Grünzinger, & Kübler, 2011), we expect the strongest brain responses to familiar faces, to be recorded at the Pz electrode. In this section, we will start by making a basic comparison of midline electrode grand-ERPs, and justify the use of detrending techniques, before focusing on the Pz electrode, reporting Time and Frequency domain analyses (at group and subject level), and, finally, reporting the same analyses at Fz and Cz.

4.4.1.1 – Pz Electrode

At group-level, the grand average ERPs of all three critical stimuli (i.e. the Target, Irrelevant and Probe conditions), at the Pz electrode site, revealed a clear difference between the conditions (see Figure 4.3, below). The **Target** condition was task-relevant, so it elicited a large classical P3, which was as expected because subjects were instructed to detect the Target face, throughout the experiment. The **Irrelevant** condition, which consisted of an unknown face (paired with each Probe, and repeated randomly, as many times as the Probe), did not present any feature/pattern of interest (other than an SSVEP); this was as expected, since non-salient stimuli were unlikely to breakthrough into conscious awareness, due to the high presentation rate of the RSVP streams. Finally, the **Probe** condition elicited a continuous oscillatory pattern, within a 300 to 600ms time frame (observed frequency of approx. 3-4 Hz). Although, we hypothesised a large difference between the Probe (celebrity face) and the Irrelevant (unknown face) conditions, and predicted a smaller difference between the Probe and Target (task-relevant face), the oscillatory nature of the Probe pattern, which has been recorded for the first time in an RSVP-based study of faces, on the fringe of awareness, is highly significant, and requires greater analysis in the time and frequency domains.

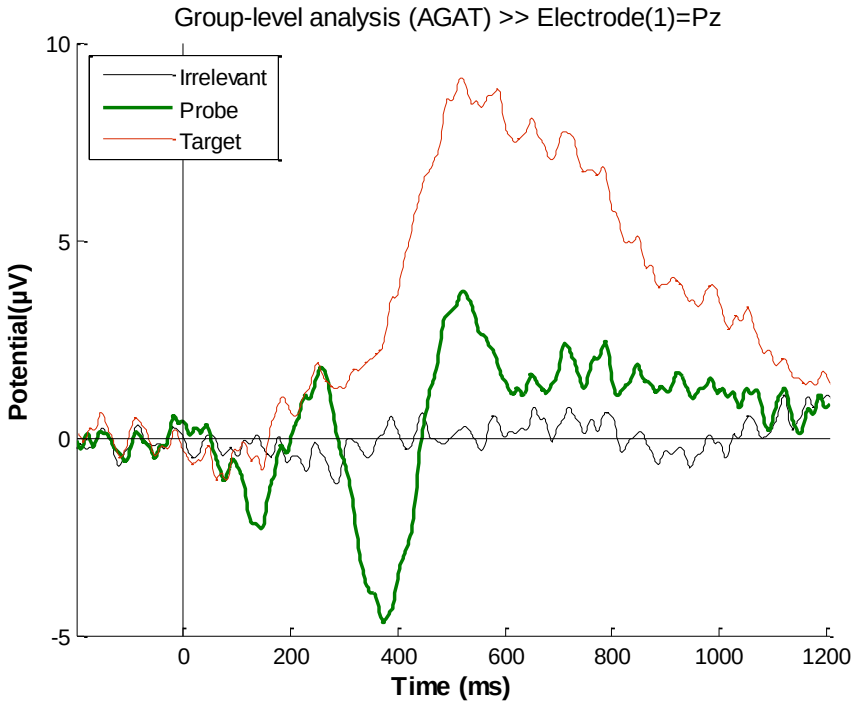

*Figure 4.3 – Grand average ERPs elicited by the three critical stimuli (Irrelevant, Probe and Target conditions), at **Pz** electrode, showing a P3 pattern for the **Target** (in red, peaking at +9μV), an oscillatory pattern for the **Probe** (in green, with an observed frequency of approx. 3-4 Hz), and no distinct pattern for the **Irrelevant** (in black, with SSVEP hovering around +/-1μV). Target was the stimulus that the subject was instructed to look for, whereas, they were not informed of the presence of the Probe (celebrity face). And yet, the oscillatory pattern for the Probe suggests a significant difference with the Irrelevants (unknown faces), which were presented as many times as the Probe.*

By collating and stacking all the trials (i.e. every single trial for all subjects in the group) for the Target condition, we observed a prevailing positivity, from 400ms onwards, for most trials, at the Pz channel. This channel-oriented representation of the trials was confirmed by the aggregated ERPs (see left plot of Figure 4.4), and the spatial dispersion of resultant waveform was depicted by the ERP scalp topographies (see right plot of Figure 4.4), which confirmed the Target condition's dominant positive wave, peaking at around 500ms.

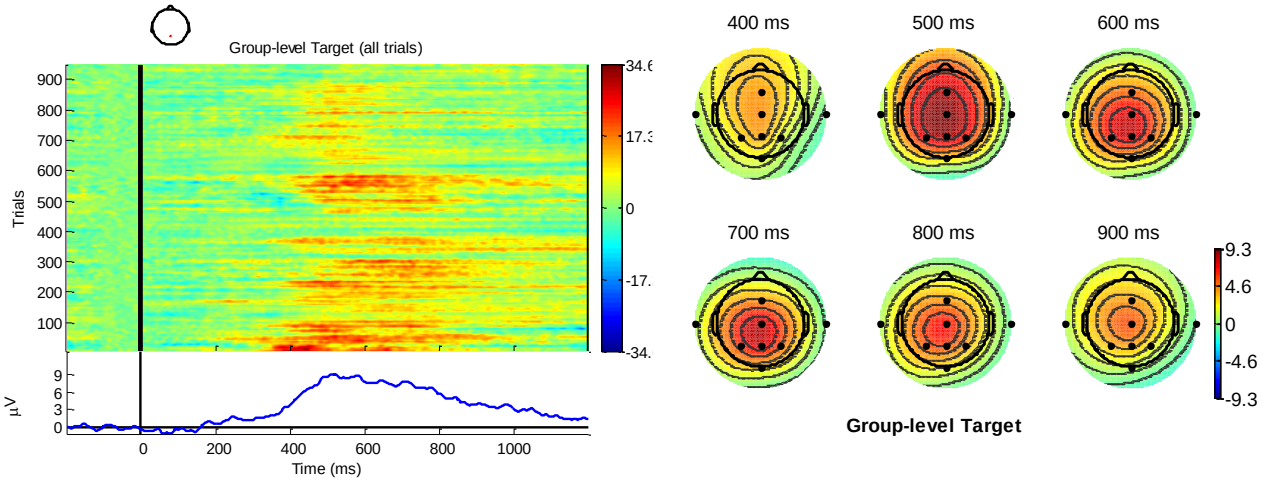

Figure 4.4 – Group-level view of all (956) Target trials, in order of appearance over time, at Pz (left plot), and the corresponding scalp topography of the ERPs (right plot), showing a prevailing positivity, peaking at around 500ms, with the electrical field moving posteriorly through time. Having used a limited number of (8) electrodes in this experiment, it must be noted that MATLAB employs an interpolatory algorithm to represent the full scalp pattern. Therefore, estimated electric potential values are used at scalp locations between the actual recording sites, and the presented scalp topographies carry considerable uncertainty, especially in respect of laterality of effects, since we have few electrodes beyond the central line. Note that the scalp map scale ranges from -9.3 to +9.3 μV.

Similarly, we stacked all the trials in the Probe condition, for all subjects at Pz, and observed the oscillatory waveform, with its peak negativity at around 350ms, and its peak positivity at around 500ms. In addition to stacked trials and their aggregated ERPs, we were able to observe this pattern in the ERP scalp topography (see Figure 4.5).

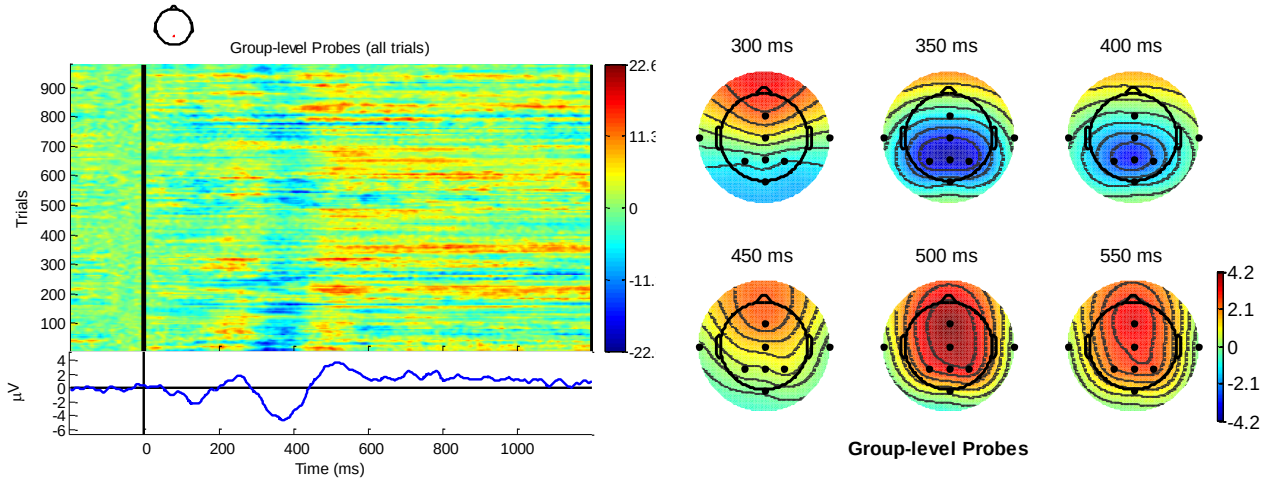

Figure 4.5 – Group-level view of all (971) Probe trials, over time, at Pz (left plot) and interpolated scalp topography of the ERPs (right plot, which must be treated with caution, due to the small number of electrodes), showing an oscillatory pattern, with lowest negativity at 350ms, and highest positivity at 500ms. Note that the scalp map scale ranges from -4.2 to +4.2 μV, which is lower than the scale for Target (see Figure 4.4).

As for the Irrelevant condition, other than the SSVEP, we observed relative inactivity, which supports our hypothesis that unknown faces, presented at a rapid rate, will not breakthrough into conscious awareness. This observation is evident in the stacked trials/ERP, and the relatively unchanging pattern in the ERP scalp topography (see Figure 4.6).

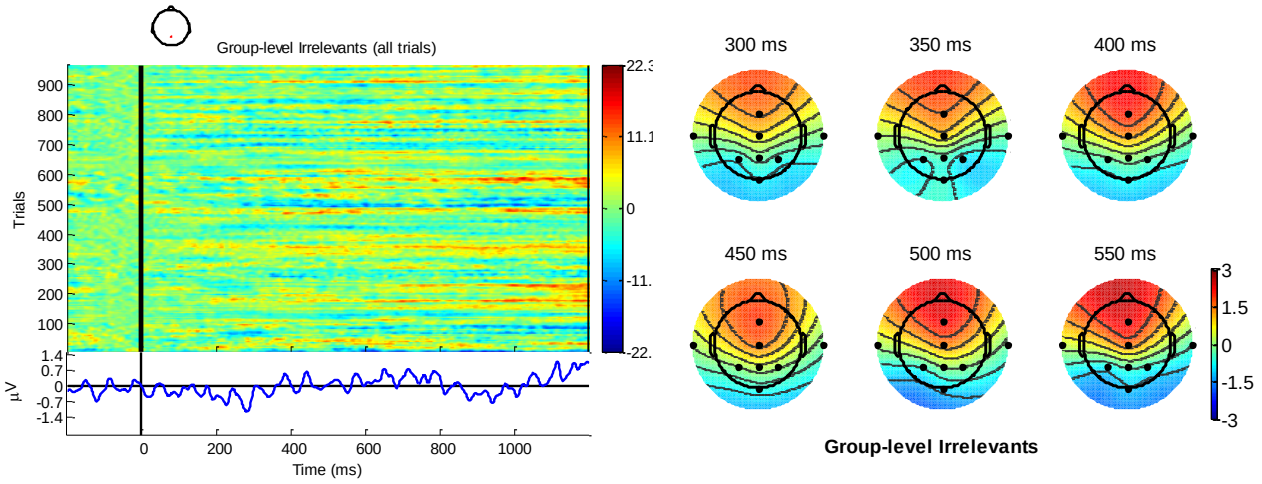

Figure 4.6 – Group-level view of all (963) Irrelevant trials, over time, at Pz (left side) and the interpolated scalp map representation of the ERPs (right plot, which must be treated with caution, due to the small number of electrodes), showing an unvarying pattern (albeit, SSVEP may be present). Note that the scalp map scale ranges from -3 to +3 μV, which is lower than the scale for Probe, (see Figure 4.5).

Ultimately, the main comparison was between the Probe and Irrelevant conditions, and our statistical tests showed a highly significant difference between them. Having aggregated all Probe and Irrelevant trials for all subjects, we employed the AGAT method, for orthogonal contrast time window placement (i.e. to independently find the most extreme 100ms mean amplitude interval) for the lowest negativity (N400f) and the highest positivity (P600f) components.

The null hypothesis (H0) was that there is no difference between the Probe and Irrelevant patterns, for the group. Our experimental hypothesis is that H0 can be rejected, at the group-level. As detailed in section 3.3.3.2 *(Group-level (AGAT) window placement*), a paired *t*-test of the mean amplitudes of Probe N400f/P600f and Irrelevant

N400f/P600f was used, across all participants, to calculate the group's *p*-values (compared to a critical alpha level of 0.05), and to determine the probability that the observed pattern could have arisen if the null hypothesis were true. This is a reliable way to determine the group's familiarity with the Probe faces.

Within the a-priori N400f time-frame (i.e. 300ms to 500ms), the AGAT orthogonal contrast method independently identified the 100ms time window, at 322ms to 422ms (M = -4.674, SD = 2.556), and our statistical tests produced a highly significant difference between the Probe (M = -3.6935) and Irrelevant (M = 0.06786), with a *p*-value of $p < 0.0001$. Similarly, within the a-priori P600f time-frame (i.e. 300ms to 900ms), the AGAT orthogonal contrast method independently identified the 100ms time window, at 479ms to 578ms (M = 1.5418, SD = 0.55908), and our statistical tests produced a highly significant difference between the Probe (M = 3.0133) and Irrelevant (M = 0.070288), with a *p*-value of $p = 0.0001$ (see Figure 4.7, below).

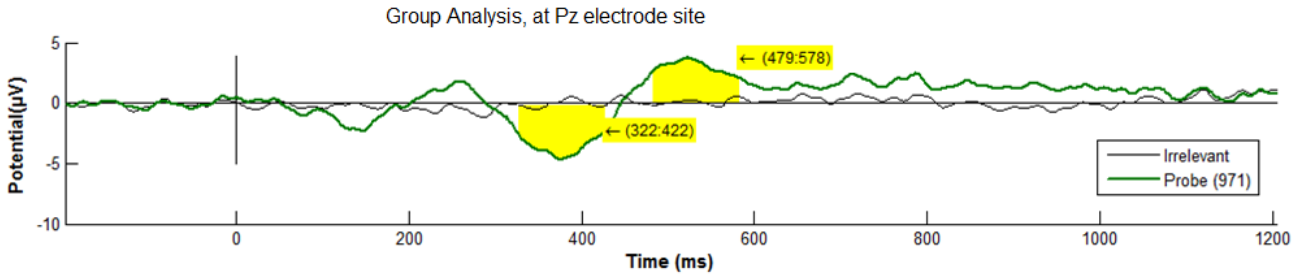

Figure 4.7 – Grand average ERPs elicited by Probe and Irrelevant (i.e. the main comparison conditions) at Pz. AGAT windows for N400f and P600f components are highlighted in yellow, and both p-values are highly significant (p < 0.001).

4.4.1.2 – Other Electrodes (Fz and Cz)

Both Probe and Irrelevant conditions present similar patterns at the other two midline electrode sites (i.e. Cz and Fz). However, unlike Pz, both Cz and Fz suffer from a slow drift in the signal that appears to skew the data, and interferes with our analyses. As shown in the following grand average ERPs at Cz and Fz electrodes (see Figure 4.8, below), a consistent drift existed in all three conditions (i.e. Prove, Irrelevant and Target), albeit, the drift for Cz was not as bad as Fz.

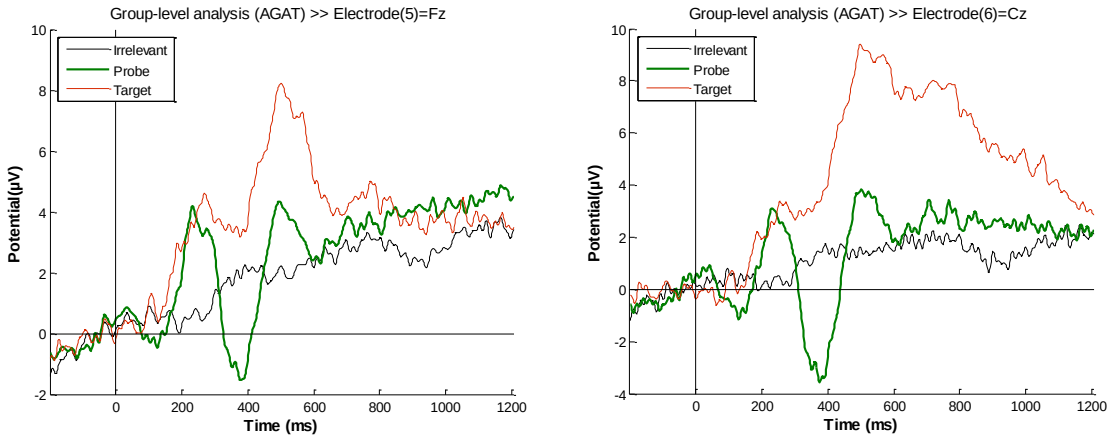

*Figure 4.8 – Grand average ERPs elicited by Irrelevant, Probe and Target conditions, at **Fz** (left plot) and **Cz** (right plot) electrodes, with similar patterns to Pz, but both suffer from a late drift in the EEG signal.*

Before continuing the analysis of the ERP data (at group, and then at subject level), we decided to review the traditional methods that have been employed to handle drift (i.e. high-pass filtering), with the knowledge that in this experiment, different subjects and/or electrodes have experienced varying degrees of drift.

4.4.2 – Traditional Handling of Drift

The common reason for the *drift* may relate to weak EEG brain signals that struggle to compete with various sources of noise (e.g. sweat on the scalp/skin, which degrades the signal, over time). In previous experiments, the standard method for dealing with any drift in the EEG data was to employ increasing high-pass filtering, by exceeding the standard cut-off rate of 0.5 Hz (e.g. increasing the high-pass filter to 1.0 Hz). However, filtering strategies may introduce temporal distortions in the signal, especially for low frequency P3 components that we are studying, where the amplitude of the P3 starts to reduce as the frequency cut-off increases. Furthermore, the practice of throttling the high-pass filter, in accordance with the perceived level of drift, was considered by us to be arbitrary. Therefore, we sought a robust and independent method

for mitigating the drift, wherever necessary, without introducing temporal artefacts and/or making value judgements. This was achieved through a *Detrending* technique, which was considered to be robust and safe because it is independently applied to all conditions (see section 4.4.2.3 – *Application of Detrending* – below).

4.4.2.1 – Common Causes of ERP Drift

No matter how carefully we setup our EEG recordings, prepare our electrode contacts, and instruct our participants to avoid physical movements, experiments are not immune to the presence of noise and artefacts. Therefore, the detection and removal of artefacts is an important part of our ERP analysis. To improve the signal-to-noise ratio, we eliminated physiological noise (e.g. eye blinks or heartbeats), and environmental sources (e.g. mains power line). Thereafter, we relied on well-established filtering techniques, namely low-pass, high-pass and notch filtering of the recorded data. Whereas low-pass and notch filtering parameters can be standardised for all subjects/electrode (i.e. 45Hz and 7 to 9Hz, respectively), high-pass filtering is considered to be a useful tool for handling drift. Therefore, a subject/channel's level of drift can influence the need to raise high-pass filtering, until the observed drift has been resolved. The optimal high-pass filter for each subject/electrode would be applied post-hoc, according to the experienced observations of the researcher/experimenter.

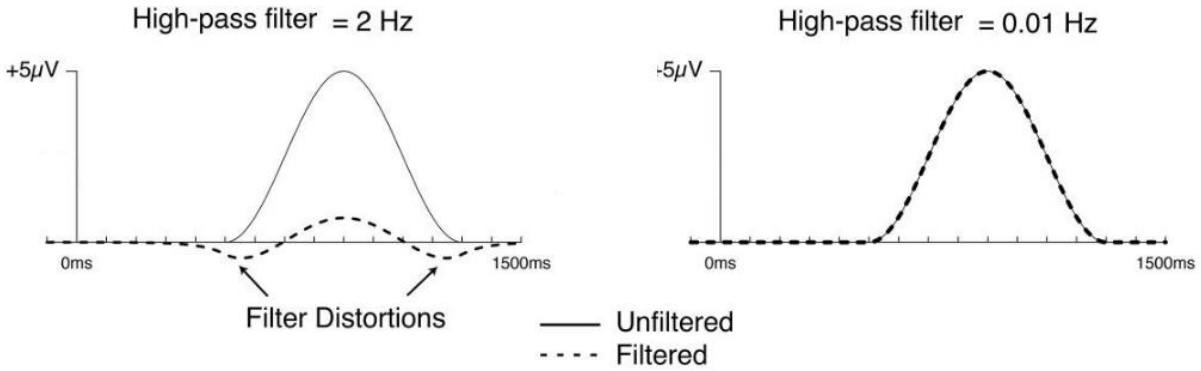

Figure 4.9 – Effects of excessive filtering, on simulated P600-like ERP component. Left plot shows the distortion of the original signal (solid 'Unfiltered' wave), when an excessive high-pass filter (2 Hz) has been applied (dotted 'Filtered' wave is distorted). The right plot shows no apparent distortion at minimal high-pass filter (0.01 Hz). Reproduced from (Tanner, 2015).

Traditionally, researchers have relied on post-hoc high-pass filtering to reduce the drift (i.e. by reducing the amplitude of slow ERP components), but recent studies (Tanner, 2015) have acknowledged that such interventions may introduce artefactual peaks that lead to incorrect conclusions (see a simulated example in Figure 4.9, above).

4.4.2.2 – Post-hoc High-pass Filtering

As discovered earlier (see Figure 4.8, above), at the Fz electrode site, the grand-average ERPs exhibited a prominent drift. An experienced experimenter would observe this anomaly and elect to mitigate the drift by increasing the high-pass filter, in order to reduce the amplitude of slow ERP components, and negate the drift for both Probe and Irrelevant.

As can be seen in Figure 4.10, without increasing the high-pass filter (left plot), the orthogonal time window placement, using the AGAT method, would incorrectly find a window towards the top-end of the search window range (i.e. 300 to 900ms), whereas, by increasing the high-pass filter to 1.0 Hz (right plot), the amplitude of the ERP has been reduced, resulting in the correct placement of the window over the *true* component of interest (P600f). As a result, the significance of our statistical analysis would improve, from $p = 0.095$ (AGAT win = 752:852ms, $M = 3.3903$, $SD = 0.44458$), to a highly significant $p = 0.002$ (AGAT win = 465:564ms, $M = 1.2256$, $SD = 0.54176$).

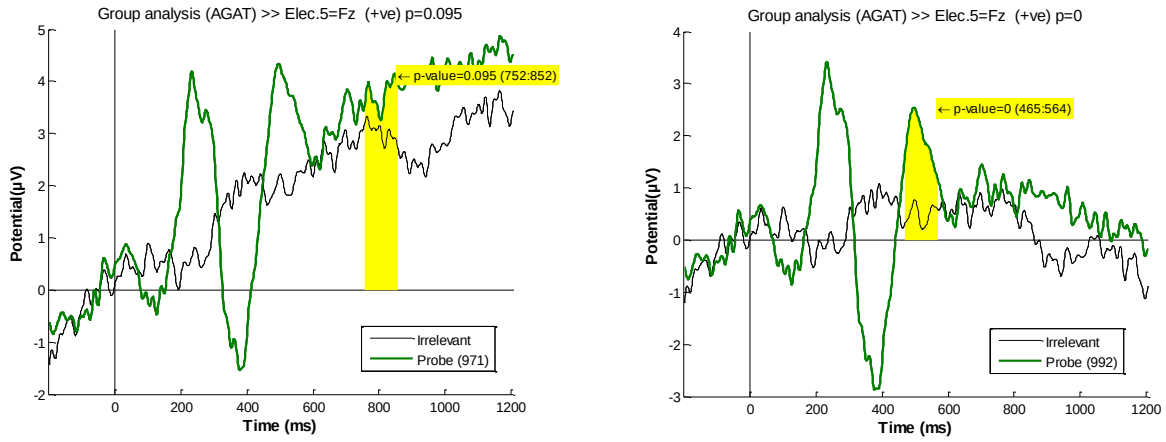

*Figure 4.10 – Comparative result of grand-average ERPs elicited by Probe and Irrelevant at **Fz** electrode, with the default 0.5 Hz high-pass filter (left-side plot, showing a late drift), and the effect of doubling the high-pass filter to 1.0 Hz (right-side plot), to reduce the amplitude of the slow ERP component, and mitigate the drift. Although, the increased high-pass filter was effective, we had to be very circumspect in applying excessive high-pass filtering, which could introduce pronounced N400/P600 effects, resulting in false conclusions (Tanner, 2015).*

As demonstrated above, the standard method for dealing with drift in EEG data is to increase the high-pass filter above the standard cut-off rate of 0.5 Hz (e.g. to 1.0 Hz for Fz electrode, as demonstrated in Figure 4.10). However, such filtering strategies were left to the experience of the experimenter, knowing that excessive use of high-pass filtering could introduce temporal distortions in the signal (as simulated in Figure 4.9). As previously discussed, fundamentally, the practice of throttling the high-pass filter, in accordance with the perceived level of drift, was considered by us to be arbitrary – especially as it could be different for different subjects within the same experiment (i.e. no consistency in its application, across the group).

Therefore, we were incentivised to find a robust and independent means of mitigating the drift, without introducing temporal artefacts and/or making value judgements. This was achieved through a *Detrending* technique, which was considered to be robust and safe because it can be independently applied to all subjects. By fixing our high-pass filter at the lowest default rate (i.e. 0.5 Hz), we utilised the detrending technique, to remove linear drifts from the data, for every epoch (Craston, Wyble, Chennu, & Bowman, 2009), and focus our analysis on the fluctuations, rather than the systematic increase or decrease in the artefact influenced data (e.g. sensor drift). Whereas the post-hoc application of high-pass filtering – which could be applied, anywhere from 0.5 Hz to 2 Hz – was traditionally a judgement call by the experimenter, our standard application of detrending, to all subjects, is independent of the observer.

4.4.2.3 – Application of Detrending

To establish a precedent for applying a standard Detrending technique to all EEG data (i.e. without the need to throttle the high-pass filter), we began by demonstrating the adverse effects of drift on our analyses: we employed the AGAT method, for orthogonal contrast time window placement (i.e. to independently find the highest 100ms mean amplitude interval, using combined Probe and Irrelevant trials) for the highest positivity (P600f), and then performed a statistical test, both before and after

detrending. As can be seen in the following ERP analysis (left plot of Figure 4.11), without detrending, the slow drift deceives the AGAT method into placing the time window towards the latter end of the ERP waveform (i.e. where there is no apparent component of interest), however, detrending safely removes the drift from both conditions (right plot of Figure 4.11), and places the window over the *true* component of interest (P600f). As a result, the statistical analyses, performed before-and-after, would improve the original *p*-value of $p = 0.095$ (AGAT win = 752:852ms, $M = 3.3903$, $SD = 0.44458$), to a highly significant $p = 0.002$ (AGAT win = 463:563ms, $M = 0.8726$, $SD = 0.52776$).

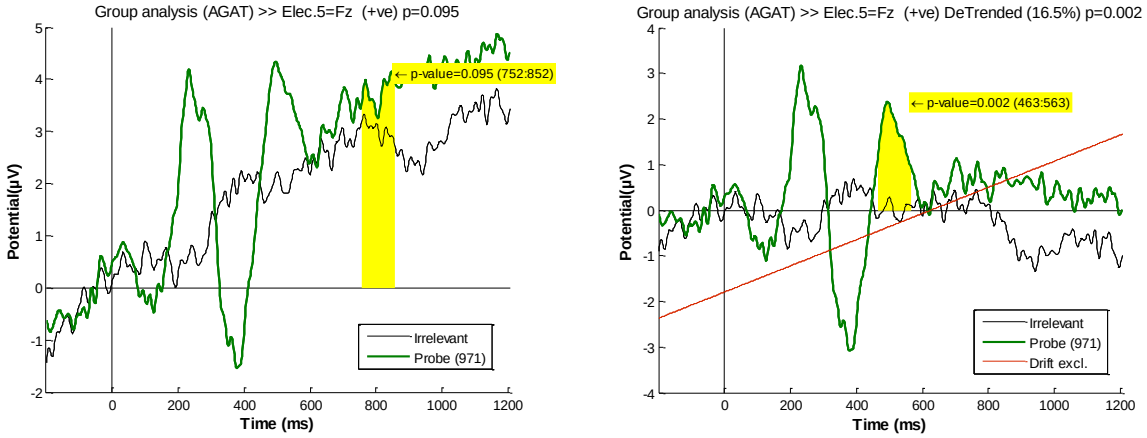

*Figure 4.11 – Grand average ERPs elicited by Probe and Irrelevant, at **Fz** electrode (left plot), presented a drift, which meant that the AGAT orthogonal contrast method of window selection 'overshot' the P600f component of interest (p = 0.095). However, once both conditions were detrended (right plot), the AGAT method was successful at independently finding the P600f component (p = 0.002). Note that the drift (shown in red) was found by calculating the combined Probe/Irrelevant trend away from the x-axis time domain - producing the linear-trend-line, at 16.5% to the vertical - before subtracting the 'drift' from every trial of both conditions.*

Similarly, the adverse effects of a slow drift on our analysis of the Cz electrode was demonstrated (see Figure 4.12, below). Even though the angle of the late-drift for Cz (10.2% from the horizontal) is not as sharp/prominent as the drift at Fz (16.5% from the horizontal), the adverse effect on our analysis is also damaging. Once again, we employed the AGAT method, for the highest positivity (i.e. to find P600f time window), both before and after detrending, and then performed statistical tests. As can be seen in the following ERP analysis (left plot of Figure 4.12), without detrending, the slow drift deceives the AGAT method into placing the window towards the latter end of the ERP

(i.e. where there is no apparent component of interest), whereas, detrending safely removes the drift from both conditions (right plot of Figure 4.12), and places the *true* window over the component of interest (P600f). As a result, the statistical analysis would improve the original *p*-value of $p = 0.031$ (AGAT win = 689:789, $M = 2.4569$, $SD = 0.54252$), to a highly significant $p < 0.0001$ (AGAT win = 467:566, $M = 1.1664$, $SD = 0.59001$).

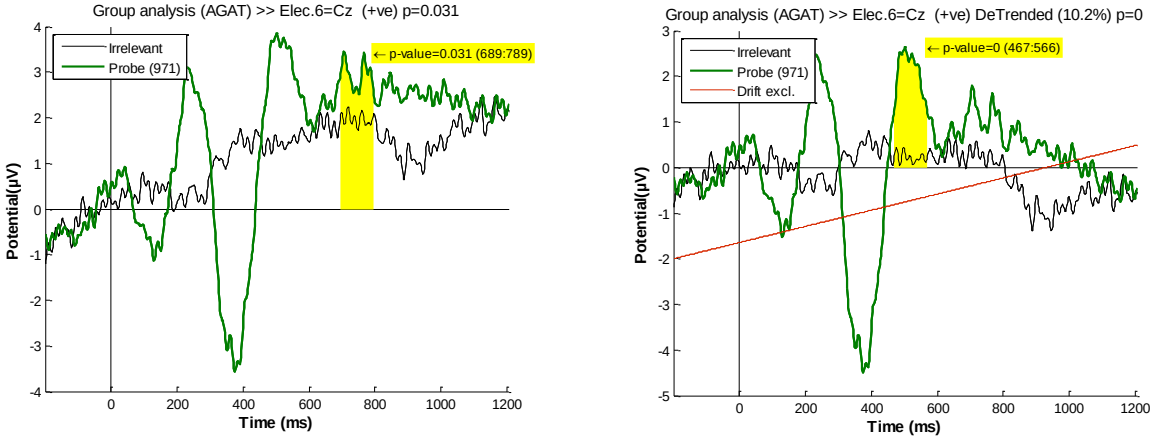

*Figure 4.12 – Grand average ERPs elicited by Probe and Irrelevant, at **Cz** electrode (left plot), presented a drift, which means that the AGAT orthogonal contrast method of window selection 'overshot' the P600f component of interest (albeit, finding a late component that appears after P600f, which was significant; p = 0.031). However, once both conditions were detrended (right plot), the AGAT method was successful at independently finding the highly significant P600f component (p < 0.001). Note that the drift (shown in red) was found by calculating the combined Probe/Irrelevant trend away from the x-axis time domain - producing the linear-trend-line, at 10.2% to the vertical - before subtracting the 'drift' from every trial of both conditions.*

Note that, in the above Cz electrode comparison (see Figure 4.12), whilst the pre-detrending *p*-value is significant ($p = 0.031$), the orthogonal window placement is clearly not optimised (689 to 789ms, using the AGAT method), and the component of interest (P600f) has not been correctly identified. However, once a robust detrending technique has been applied, the window placement is correctly identified (467 to 566ms), and the *p*-value becomes highly significant ($p < 0.0001$).

Finally, it must be noted that baseline correction was done after detrending, otherwise, we could be artificially lowering the late-components and increasing the early-ones (i.e. tilting the data, end-down and start-up).

4.4.3 – Group-level Analysis, at Pz

With an a-priori choice of focusing our statistical analyses at the Pz electrode (in-line with (Kaufmann, Schulz, Grünzinger, & Kübler, 2011)), and the justification of applying an independent detrending technique to mitigate the inevitable drift, we performed statistical tests (i.e. paired *t*-test of the mean amplitudes of Probe N400f/P600f and Irrelevant N400f/P600f, across all subjects), using a critical-alpha level of 0.05. As explained above (see section 3.3.3.2 – *Group-level (AGAT) window placement*), aggregating all Probe and Irrelevant trials for all subjects, and employing the AGAT method, for orthogonal contrast time window placement, would enable us to independently identify the highest 100ms mean amplitude interval for the lowest negativity (N400f) and the highest positivity (P600f). Despite the fact that the Pz electrode was the least affected by a drift in the EEG signal, we have demonstrated the benefits of using a safe and robust detrending technique, which has become a part of our standard procedure for preparing the EEG data for statistical analysis. Hereafter, all experiments will benefit from our detrending procedure, at trial level and before baseline correction.

4.4.3.1 – Group N400f

Within the a-priori N400f time-frame (i.e. 300ms to 500ms), the AGAT method independently identified an orthogonal contrast 100ms time window, at 322 to 422ms (see Figure 4.13), and our statistical tests produced a highly significant difference between the Probe (M = -4.333, SD = 2.0506) and Irrelevant (M = -0.546, SD = 1.1683), at Pz electrode site: (M = -3.787, SD = 2.3394), $t(13) = -6.057$, $p < 0.0001$, $d' = -2.269$.

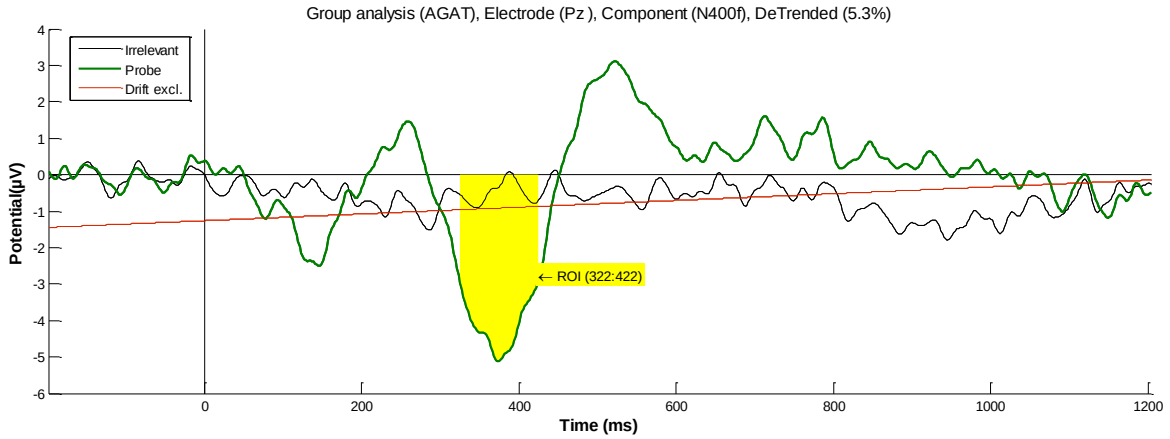

*Figure 4.13 – Grand average ERPs elicited by Probe and Irrelevant (i.e. the main comparison conditions) at **Pz**, showing an oscillatory pattern for the Probe condition (in green), which does not exist for the Irrelevant condition (in black). The linear Drift has been excluded (in red, at 5.3% to the vertical), with a detrending method. Even though subjects were not informed of the presence of the Probe (famous celebrity face), statistical tests show a highly significant difference between Probe and Irrelevant, for **N400f** component (t(13) = -6.057, p < 0.0001, d' = -2.269).*

4.4.3.2 – Group P600f

Similarly, within the a-priori P600f time-frame (i.e. 300ms to 900ms), the AGAT method independently identified an orthogonal contrast 100ms time window, at 479ms to 578ms (see Figure 4.14), and our statistical tests produced a highly significant difference between the Probe (M = 2.3426, SD = 1.761) and Irrelevant (M = -0.6063, SD = 1.4475), at the Pz electrode site (M = 2.9489, SD = 2.0213), $t(13) = 5.4587$, $p = 0.0001$, $d' = 1.829$).

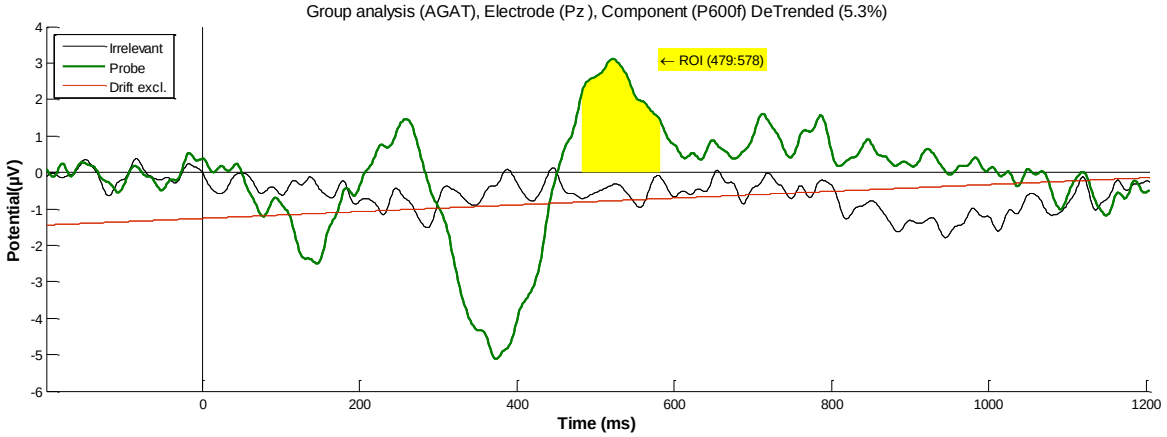

*Figure 4.14 – Grand average ERPs elicited by Probe and Irrelevant (i.e. the main comparison conditions) at **Pz**, showing a highly significant difference between them for **P600f** component (t(13) = 5.4587, p = 0.0001, d' = 1.8295).*

4.4.3.3 – By-item (block) Analysis

As explained in section 4.3.5.2 (*By-item Probe recognition*), the trials for each Celebrity were combined to calculate the by-item group significance, between the Probe/Irrelevant conditions (i.e. each Probe/famous face against its paired Irrelevant/unknown face). Although this additional (by-item) analysis would not inform the key enquiry, which is to make comparisons between the Probe and Irrelevant conditions at subject-level, we were interested in the by-item effect, as it would reveal the group-level effect of detecting different celebrity faces, and allow for a general comparison with the behavioural/recognition tests.

All celebrity faces exhibited similar Probe waveforms, with high significance at the lowest negativity (N400f), whilst statistical results of the highest positivity (P600f) appear to confirm the subject's recognition results (see section 4.3.5.2 – *By-item Probe recognition*), in which the first block was not perceived by all subjects. Studies (Sutton, Braren, Zubin, & John, 1965) have compared the P600 with the P3 – notably, the P3b, which relates to the *oddball* paradigm – as both components are characterised by similar latencies and topographical distributions, over the centro-parietal scalp region (albeit, the P600f often peaks at later latencies). Hence, the naïve state of subjects' exposure to the first block (who were not informed that celebrity faces may appear in the experiment) could explain the lowest recognition results for Jolie, whereas, it was more likely that subjects would infer the presence of celebrity faces in future blocks (see Figure 4.15, for a comparison between the first and last blocks). Therefore, the P600f (*P3b oddball*) effect in the first block (Jolie; $p = 0.275$) could not match subsequent blocks (e.g. DiCaprio; $p < 0.001$), despite Jolie being the second most recognised celebrity.

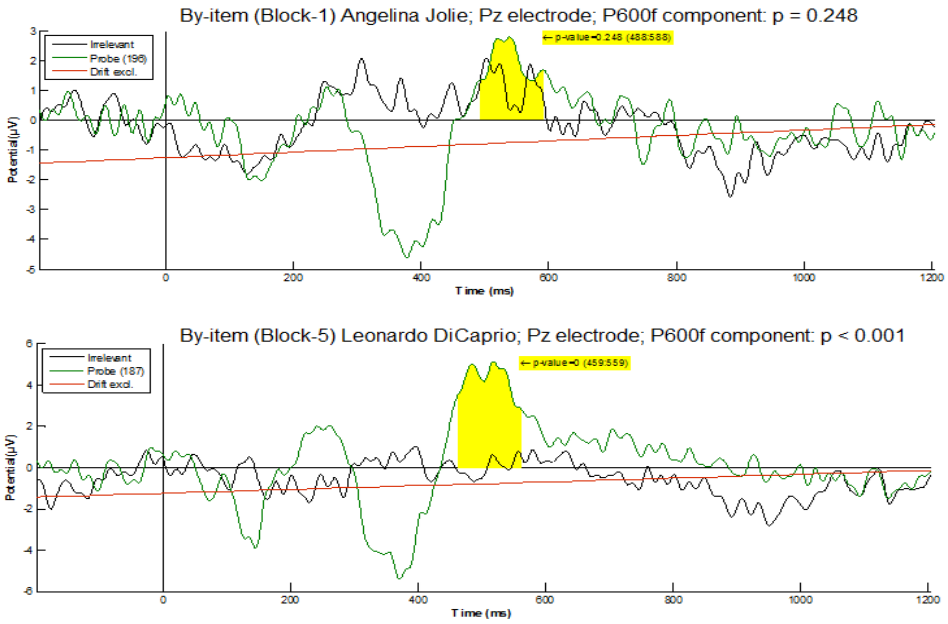

Figure 4.15 – By-item analysis of the first (Jolie, top-plot) and last (DiCaprio, bottom-plot) blocks, at Pz electrode, showing similar oscillatory pattern for the Probe, even though, the P600f differs in significance (p = 0.248 for Jolie and p <0.001 for DiCaprio). Note that the Probe's maximum P600f voltage for DiCaprio (5.5µV) is almost double that of Jolie (2.75µV), whereas, minimum N400f is similar for both (approx. -4.75µV).

4.4.4 – Subject-level Analysis

Having established that our goal was to statistically analyse the data at the Pz electrode site only (Kaufmann, Schulz, Grünzinger, & Kübler, 2011), and the fact that the main comparison was between the Probe and Irrelevant conditions (Bowman, et al., 2013), we performed statistical analyses of the ERP data, to determine whether the elicited response by the Probe (i.e. celebrity face) was significantly different from that elicited by the Irrelevant (i.e. the unknown face), on a subject-level basis. As outlined in section 3.3.3 (*Time Domain (ERP) Analysis*), subject-level analysis is based on analysing each experimental participant separately, to determine whether there was a significant difference for that subject alone – did the subject's brain response reveal a differential perception & processing of the celebrity face, as compared to the unknown face? The null hypothesis (H0) was that there is no difference between the subject's Probe and Irrelevant patterns. Our experimental hypothesis is that H0 can be rejected.

Having used the aERPt method to independently find the time window for each component of interest, a randomisation (i.e. Monte Carlo Permutation) test was used to define a *p*-value for each subject. Then, a null hypothesis distribution was generated in order to calculate the individual's *p*-value. This *p*-value would determine the probability that the observed pattern could have arisen if the null hypothesis were true. This is a reliable way to assess each subject's pattern individually, and to determine that subject's familiarity with the Probe. Whether the subject reported to have seen the Probes in the relevant block of the experiment, or not, we theorised that the results of our statistical analysis would infer their conscious and/or unconscious (i.e. sub/liminal) detection of celebrity faces – in a Concealed Information Test, this could infer the guilt of the subject.

4.4.4.1 – Synopsis of results

As shown in table 4.2 (below), subject-level statistical tests of Pz electrode's N400f component resulted in 10 of 14 subjects (71%) achieving critical-significance (at alpha level $p < 0.05$, shown in green), between the Probe and Irrelevant conditions. Furthermore, statistical tests of P600f component resulted in 7 of 14 subjects (50%) with *p*-values below our critical-significance. After combining each subject's *p*-values of the N400f and P600f components (as described in section 3.3.4 – *Combined probability test (Fisher's)*), all 14 subjects (100%) achieved Fisher combined levels at a minimal-significance (i.e. an alpha level of $p < 0.1$), as used in most of Farwell and Rosenfeld's deception detection studies (Farwell & Donchin, 1991); (Rosenfeld I. P., 2008). Out of these, 10 of 14 subjects (71.4%) achieved critical-significance (alpha level $p < 0.05$), which is our preferred alpha level, in all experiments.

SUBJECT	N400F	P600F	FISHER
1	*< 0.0001*	0.033	*<0.0001*
2	*< 0.0001*	0.062	*<0.0001*
3	*< 0.0001*	0.012	*<0.0001*
4	0.014	0.093	0.015
5	0.106	0.031	0.028
6	*0.001*	0.246	0.011
7	0.113	*< 0.0001*	*0.001*
8	0.045	0.394	0.094*
9	0.037	0.37	0.083*
10	0.049	0.002	0.011
11	0.44	0.028	0.076*
12	0.05	0.322	0.093*
13	0.169	*< 0.0001*	*<0.0001*
14	*< 0.0001*	0.24	*0.001*

*Table 4.2 – Subject-level analysis, at **Pz** electrode, for N400f and P600f components, and their **Fisher** combining probability. Note that minimal-significance (at alpha level p < 0.1) is shown in blue (with an astrix), and critical-significance (at alpha level p < 0.05) is shown in green. All 14 subjects (100%) achieved Fisher combined levels at minimal-significance (i.e. an alpha level of 0.1), and 10 of 14 subjects (71.4%) achieved critical-significance (alpha level 0.05).*

4.4.4.2 – Individual's N400f, by-item and by-subject

At the Pz electrode site, we began by exploring the presence of the N400f component within each of the five items of every subject (i.e. 5 experimental blocks for 14 subjects, equalling 70 item-blocks). Having independently searched for each component's 100ms aERPt time window (i.e. highest *negative* deflection, within the

a-priori search area that spans from the time range of 300ms to 500ms), we performed permutation tests for each individual block (see Appendix A.2 for more detail).

Consequently, five 'by-item' *p*-values were obtained for each subject's block (i.e. one for each celebrity), resulting in significant difference between the Probe and Irrelevant conditions for 20 of 70 blocks (28.6%). Despite the unfavourable number of significant *p*-values (i.e. only 20 of 70), it was noted that the Signal to Noise Ratio (SNR) of block-level analysis is low, due to the relatively low number of trials per item/block. However, by combining subject's trials (i.e. up to 75 trials per condition, to gain a safe and representative SNR), we were able to perform statistical tests on each subject (as shown in Table 4.2), resulting in a significant difference (at alpha level $p < 0.05$) between the Probe and Irrelevant conditions for 10 of 14 subjects (71.4%), of which 6 subjects have highly-significant *p*-values ($p < 0.001$).

As highlighted in Figure 4.16 (below), all subjects' Probes elicited a clear negative deflection within the N400f time-frame (300ms to 500ms), however, relative to the Irrelevant (i.e. the condition of comparison), subject 11's Probe does not possess a dominant negativity. Thus, the True Observed Value for subject 11 (i.e. N400f Probe -minus- N400f Irrelevant) was very small, resulting in the largest *p*-value ($p = 0.44$) of all 14 subjects.

Similarly, subjects 5, 7 and 13 were slightly above our critical-alpha level (i.e. *p*-values 0.106, 0.113 and 0.169, respectively), but interestingly, their positivity, depicting a P600f component was prominent and significant. The average window placement for all 14 subjects' N400f component was at 322 to 422ms.

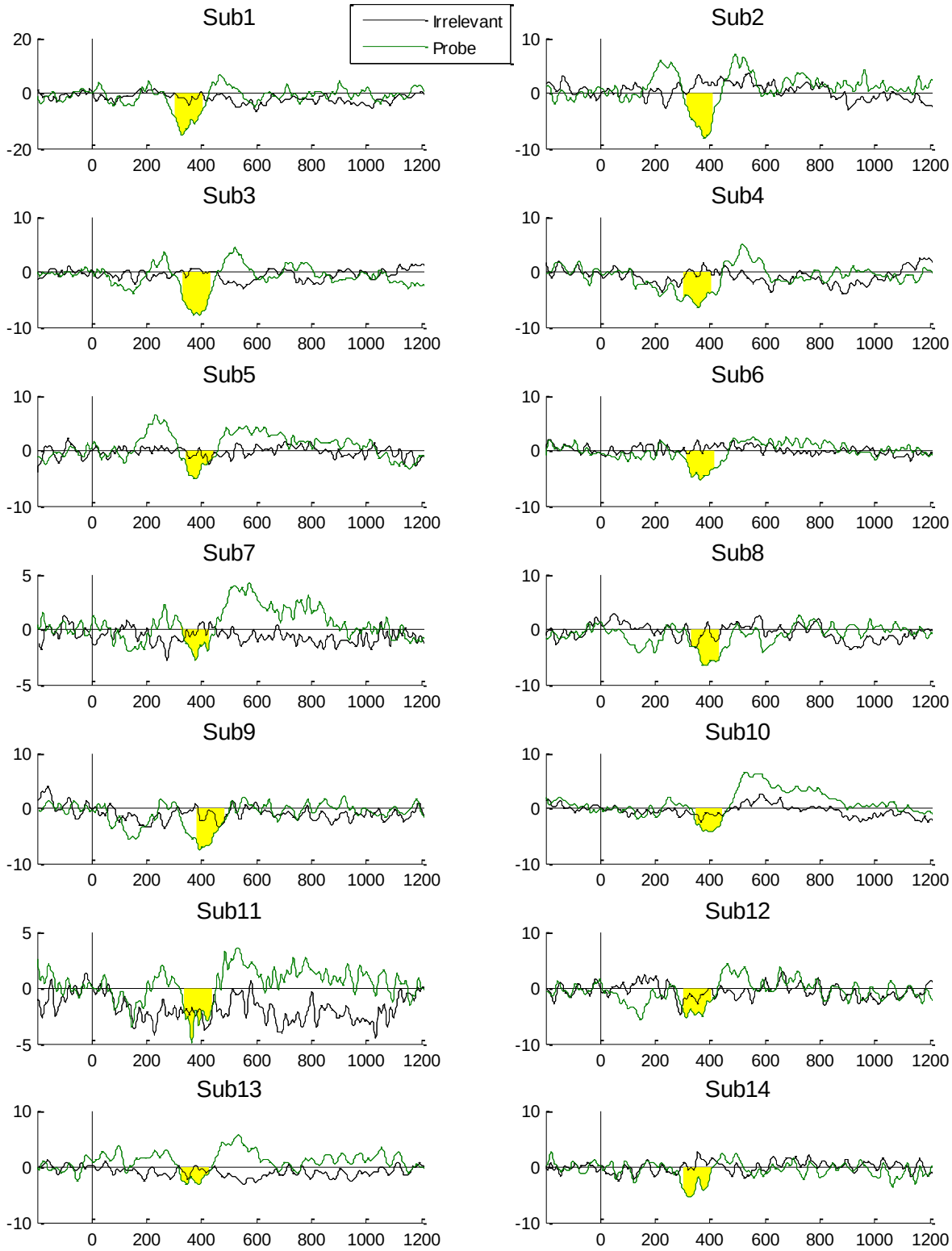

*Figure 4.16 – Subject-level Probe (in green) and Irrelevant (in black) ERPs, at the **Pz** electrode site (x-axis represents Time in miliseconds, and y-axis represents Potential in microvolts). Each ERP shows the orthogonally identified highest positive 100ms time window (yellow highlight) for **N400f** (using the aERPt method), where 10 of 14 subjects (71.4%) show a significant difference between the Probe and Irrelevant conditions. Almost all ERPs display a distinct Probe waveform, which is not present in the Irrevelant waveform.*

4.4.4.3 – Individual's P600f, by-item and by-subject

The same process and statistical tests that were used for N400f were repeated for P600f, albeit, the 100ms aERPt defined time window was the highest *positive* deflection, within the a-priori search area (i.e. 300ms to 900ms). As a result, *p*-values obtained for each subject's by-item arrangement (i.e. one *p*-value for each celebrity) showed a significant difference (see Appendix A.2 for more detail) between the Probe and Irrelevant conditions for 11 of 70 blocks (15.7%). As explained above (see section 4.4.4.2 – *Individual's N400f*), such unfavourable results related to the low SNR of block-level analysis (i.e. low number of trials per item/block), therefore, the accumulated trials for each subject was used to test the significance of each participant, which resulted in a significant difference between the Probe and Irrelevant conditions for 7 of 14 subjects (50%), at an alpha level $p < 0.05$. However, at an alpha level of 0.1 (often used as the level of significance in P3-based deception detection studies), 9 of 14 subjects (64.3%) showed a significant difference between Probe and Irrelevant, within the P600f component time window (see Table 4.2, above).

As highlighted in Figure 4.17 (below), all but three subjects' Probes elicited a clear positive deflection, within the P600f time-frame (300ms to 900ms). Relative to the Irrelevant (i.e. the condition of comparison), subjects 6, 8 and 9 possess Probes without a clear positivity. Thus, their True Observed Value (i.e. Probe -minus- Irrelevant) was very small, resulting in large *p*-values.

Additionally, subject 12's *p*-value was large ($p = 0.322$), even though, the Probe condition possesses a clear/high positivity. This was as a result of the independently searched aERPt method finding a later-than-ideal time window (i.e. 637 to 737ms, instead of 446 to 546ms, which is a better fit for its P600f). Had the correct P600f been selected, our statistical tests would have showed a highly significant difference between the Probe and Irrelevant, for subject 12.

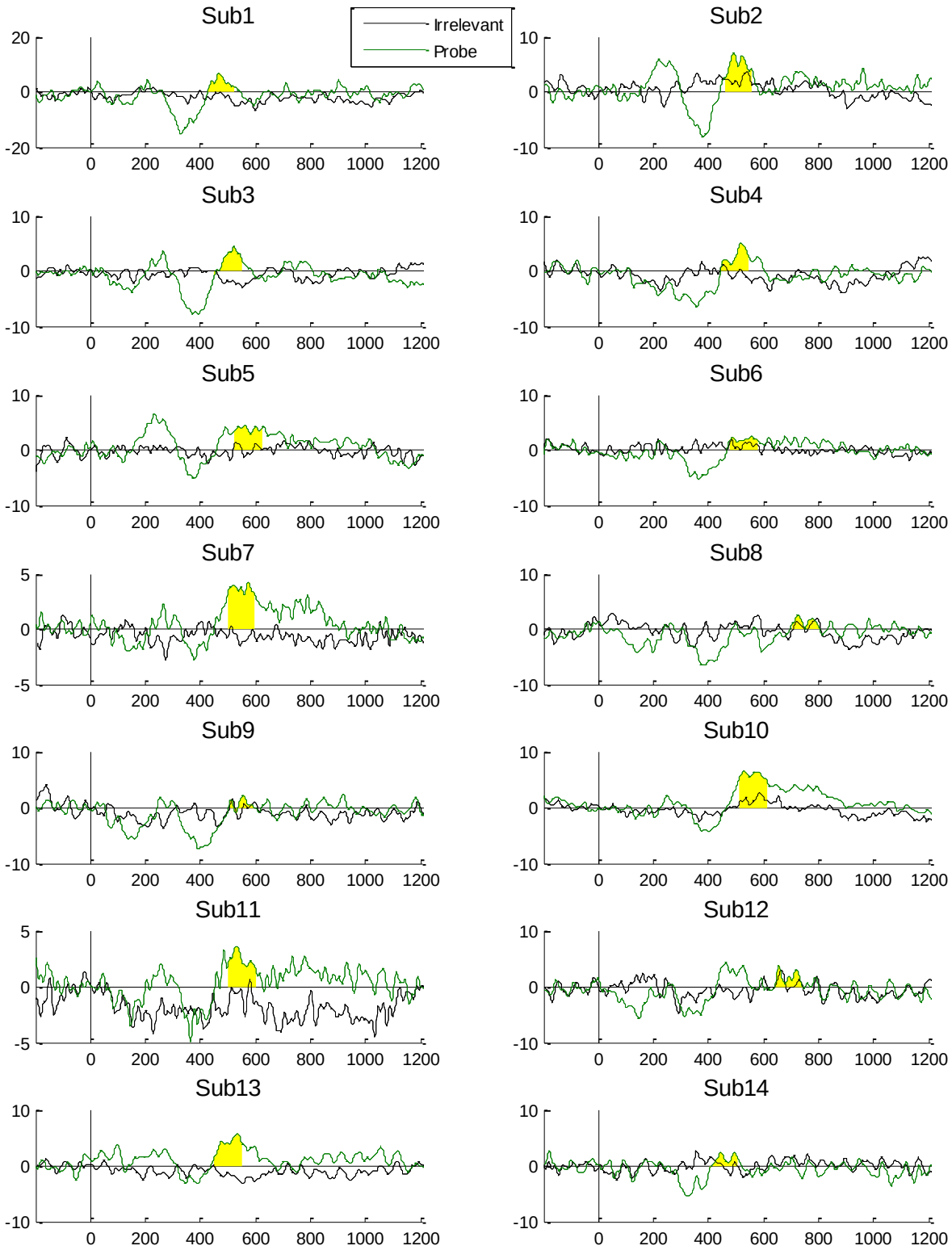

*Figure 4.17 – Subject-level Probe (in green) and Irrelevant (in black) ERPs, at the **Pz** electrode site (x-axis represents Time in miliseconds, and y-axis represents Potential in microvolts). Each ERP shows the orthogonally identified highest positive 100ms time window (yellow highlight) for **P600f** (using the aERPt method), where 7 of 14 subjects (50%) show a significant difference between the Probe and Irrelevant conditions.*

4.4.4.4 – Fisher Combining of N400f/P600f

Finally, we were able to combine the *p*-values of the N400f and the P600f components, for each subject, into a single *p*-value, by employing the Fisher combining procedure (see Figure 4.2, above). All 14 subjects (100%) achieved Fisher combined levels at a minimal-significance (i.e. an alpha level of $p < 0.1$), as used in most of Farwell and Rosenfeld's deception detection studies (Farwell & Donchin, 1991); (Rosenfeld I. P., 2008). Out of these, 10 of 14 subjects (71.4%) achieved critical-significance levels (i.e. an alpha level of $p < 0.05$), and 6 of them (42.9%) were highly-significant (i.e. an alpha level of $p < 0.001$).

As explained in section 3.3.4 (*Combined probability test (Fisher's)*), the Fisher combining procedure was able to generate a significant gain, by improving the average significance over both N400f and P600f conditions. According to simulations in (Bowman, et al., 2013), data points with *p*-values that go down, compared to the average, benefit from the application of Fisher's combining procedure, whilst those for which *p*-values go up suffer. To counter the criticism that simply multiplying the *p*-values of the two components (P400f and N600f) can inflate the false-positive rate, it must be noted that a further randomisation procedure is performed, at the level of Fisher values, as recommended by the aforementioned study.

As all subjects possessed at least one significant component (i.e. some subjects achieved significance with only N400f or P600f), their combined Fisher value was still within a minimal-significance, at an alpha level of $p < 0.1$ (see Figure 4.2, above).

4.4.5 – Time Frequency Analysis (TFA)

As outlined in section 3.3.5 (*Frequency Domain Analysis*), to analyse the power and coherence of the EEG data, we have employed two Time Frequency transforms: Even-Related Spectral Perturbation (ERSP) and Inter-Trial Coherence (ITC), using EEGLAB's toolbox (Delorme & Makeig, 2004). Whereas ERSP reflects the extent to which the signal power changes in relation to a specific time point (i.e. the baselining

window before stimulus-onset) at different frequencies in a signal, ITC reflects the phase consistency (or synchronisation) between the trials, at every time point and frequency range. ERSP/ITC changes in coherence enables us to measure and assess the multi-cycle oscillations that we had observed in the ERPs. The output is a colour-map of time (x-axis) against frequency (y-axis), where the time values before stimulus-onset (i.e. the appearance of the Probe or Irrelevant conditions) were considered to be the baseline (i.e. baseline was set at -100ms to 0ms). The power in the signal is represented by the colour, which can be green (i.e. statistically no change in the power), red (i.e. increase in power) or blue (decrease in power). A colour-bar (i.e. the key to the colour values) on the side of the plot defines the positive/negative values of the signal (in dB).

4.4.5.1 – TFA analysis framework

As outlined in section 3.3.5.1 (*Time Frequency Window Placement*), the group-level critical time window, for measuring ERSP/ITC, was placed based on the AGAT of power/coherence. As seen in Figure 4.18 (below), ERSP and ITC results of the AGAT of the Probe and Irrelevant conditions are combined together, across all 14 subject at Pz, with a large power increase around 200 to 550ms time-window (post-stimulus), mainly at the low frequency range of 0 to 10 Hz. It must be noted that the RSVP presentation rate (i.e. 133ms SOA) would evoke a Steady State Visual Evoked Potential (SSVEP), at a frequency of approximately 8 Hz (e.g. rhythmic pulses at 133, 266, 399, 532, 665, 798, 931, 1,064ms, and so on). Therefore, we applied a notch filter, between 7 and 9 Hz, during the initial processing/epoching of the EEG data (see section *3.3.2 – EEG data*). Consequently, we can justify fixing the upper boundary of our analysis at 7 Hz, and the lowest boundary is fixed by our standard high-pass filter, on 0.5 Hz. In addition to the fixed-boundary analysis window (0.5 to 7 Hz), we also performed the full ERSP/ITC analyses on the full frequency range (0.5 to 45 Hz), to assess the power/coherence changes at higher frequencies.

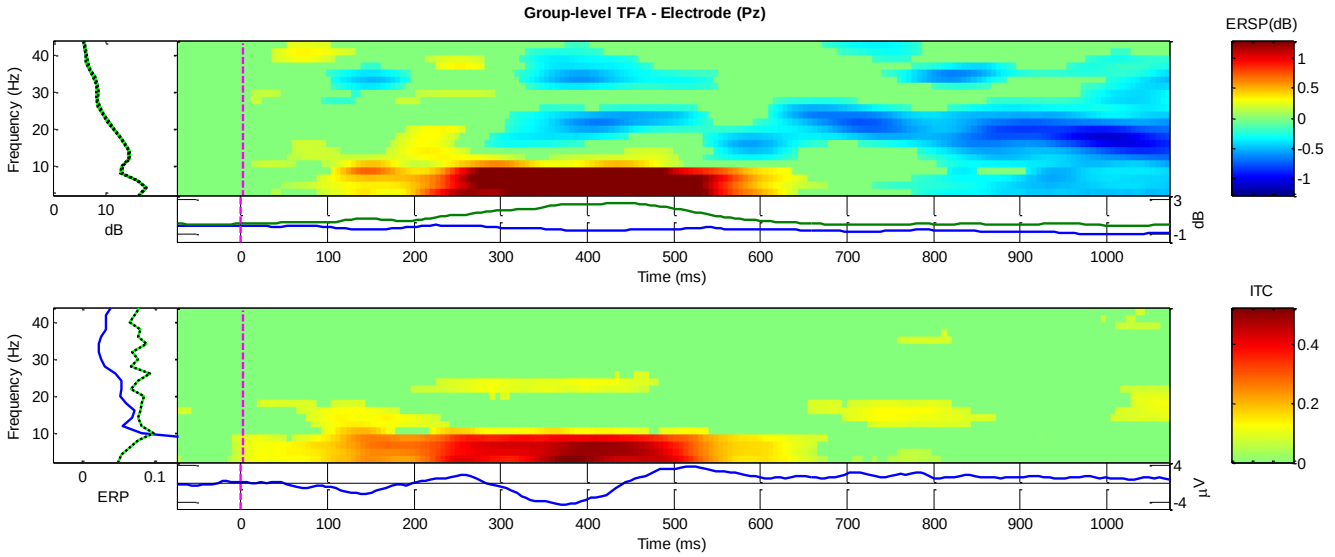

*Figure 4.18 – Group-level Time Frequency plots, at the Pz electrode, across full frequency range (i.e. **0.5 to 45 Hz**), using the combined Probe and Irrelevant conditions. The top plot relates to ERSP, and the bottom plot relates to ITC. The independent window selection (using AGAT method) for ERSP produced a Region of Interest (ROI) at 230:330ms, and 357:457ms for ITC. Increases in power/choherence have been mostly concentrated in the 0.5 to 10 Hz frequency range, and are strongest in the ROI time frames (SSVEP has been filtered out, by applying a 7:9Hz notch filter), confirming the suitability of further analysing the fixed boundary (i.e. 0.5 to 7Hz).*

4.4.5.2 – Group-level TFA

As explained in section *3.3.5.1 (Time Frequency Statistical Test)*, ERSP/ITC statistical tests were performed to compare the power and coherence changes between the two critical conditions: Probe and Irrelevant. To compare these conditions, two measures were obtained for each subject, and a two-tailed paired *t*-test was used to calculate the group-level significance. We performed independently measured statistical analyses, by obtaining an orthogonal contrast time window, using the group-level AGAT method (i.e. an aggregated grand average of trials for both power and coherence). The independent window selection, using the AGAT method, produced a 100ms Region of Interest (ROI) at 230 to 330ms for ERSP, and an ROI at 357 to 457ms for ITC. As can be seen in the grand-Probe versus grand-Irrelevant ERSP/ITC comparisons (see Figure 4.19, below), increases in power/coherence are predominantly evident in the grand-Probe condition, which suggests detection of the celebrity face (ERSP > 2.5dB, and ITC > 0.4). However, the grand-Irrelevant condition lacks any significant

power/coherence fluctuations, within the same time window, which implies little-to-no conscious or sub/liminal detection of the unknown face.

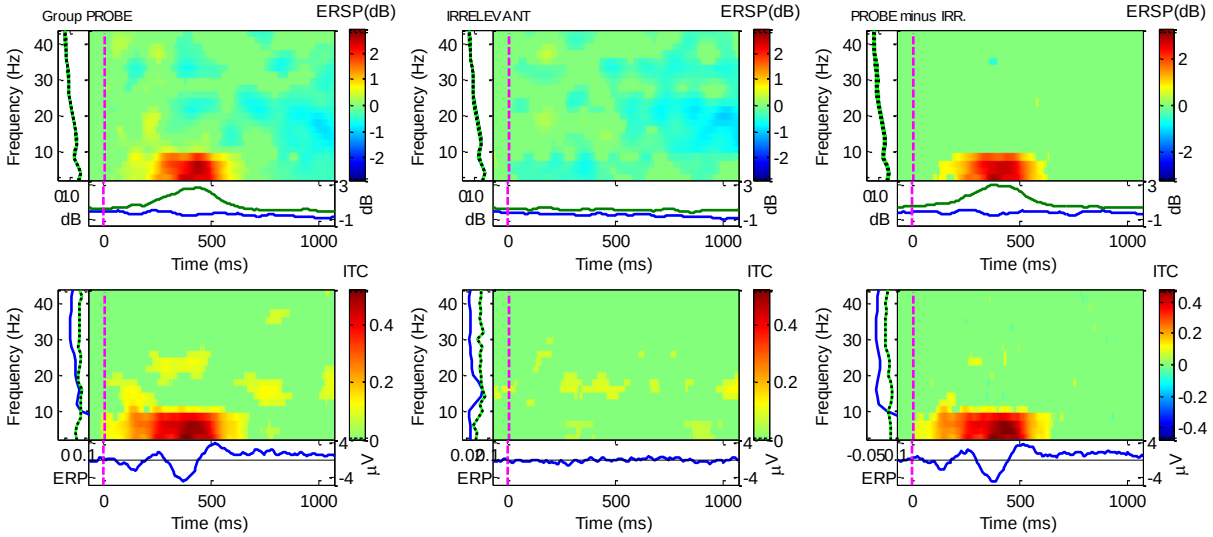

*Figure 4.19 – Group-level Time Frequency Analysis, at Pz electrode, for the difference between critical stimuli (Probe and Irrelevant), across full frequency range (**0.5 to 45 Hz**). Top row of 3 plots relates to ERSP, and the bottom row's 3 plots relates to ITC. The first column of ERSP/ITC plots show the power/coherence changes in the grand-Probe condition, and the second column shows the same for the grand-Irrelevant condition. The third column is the difference between grand-Probe and grand-Irrelevant (i.e. Probe minus Irrelevant), which confirms an increase in power and coherence at group-level for the grand-Probe only. The colour-bar (on the right-side of the plots) identifies the colour values at each frequency and time point – increase in power/coherence is in red, and decrease is in blue, whilst green indicates no significant change (i.e. all p-values > 0.01).*

Over the full frequency-range (i.e. 0.5 to 45 Hz), the group-level analysis at Pz electrode for ERSP revealed a highly significant result (see Figure 4.19, above), confirming a difference between Probe and Irrelevant conditions ($t(13) = 3.3723$, $p = 0.005$, $d = 1.2680$). For the group-level ITC over the same (maximum) frequency range, our statistical tests also confirmed high significance ($t(13) = 9.2154$, $p < 0.0001$, $d = 2.592$).

However, focusing on the narrower frequency-band (i.e. 0.5 to 7 Hz), the group-level analysis at Pz electrode for ERSP revealed a highly significant result, at the AGAT defined window 352 to 452ms (see Figure 4.10, below), confirming a difference between Probe and Irrelevant conditions ($t(13) = 6.6688$, $p < 0.0001$, $d = 2.5506$). For group-level ITC over the same (narrower) frequency range, our statistical tests also confirmed a highly significant result at the AGAT defined window 357 to 457ms: ($t(13) = 13.7146$, $p < 0.0001$, $d = 4.9355$).

The ERSP and ITC plots for the two conditions that were being compared in this study (i.e. grand-Probe and grand-Irrelevant) are shown in two columns – the top row represents power fluctuations (ERSP), and the bottom row represents coherence (ITC). For each of these two power and coherence analyses, the grand-Probe condition is shown on the left column, with the grand-Irrelevant in the middle, and the difference between these two conditions shown on the right column (i.e. Probe minus Irrelevant). At the Pz electrode, the difference plots (right column), for both ERSP and ITC, shows significant increases in power and coherence, over the maximal frequency range (see Figure 4.19, above), as well as the narrower frequency band (see Figure 4.20, below).

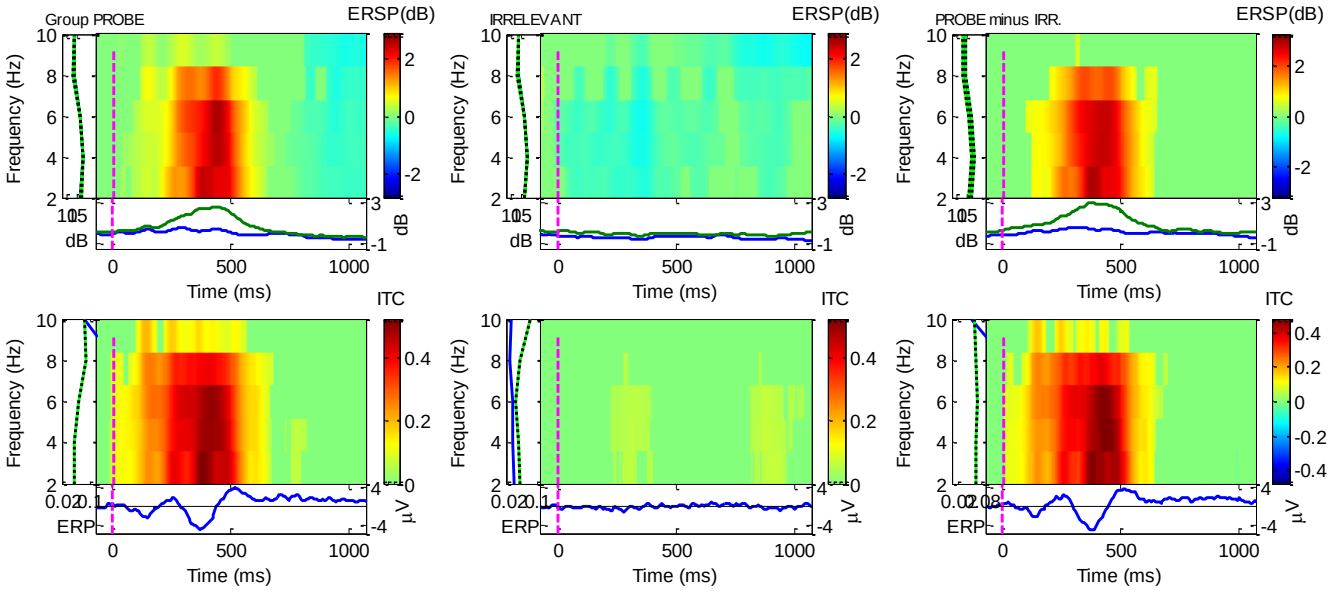

*Figure 4.20 – Group-level Time Frequency Analysis, at Pz electrode, for the difference between critical stimuli (Probe and Irrelevant), at the narrower frequency band (**0.5 to 7 Hz**). Top row of 3 plots relates to ERSP, and the bottom row's 3 plots relate to ITC. The first column of ERSP/ITC plots show the power/coherence changes in the grand-Probe condition, and the second column shows the same for the grand-Irrelevant condition. The third column is the difference between grand-Probe and grand-Irrelevant (i.e. Probe minus Irrelevant), which confirms an increase in power/coherence at group-level (especially, in 0.5 to 7 Hz frequency band), for the grand-Probe only. Note that at each frequency and time point, increases in power/coherence are in red; decreases in blue, and green indicates no significant change.*

4.4.5.3 – Subject-level TFA

Per subject statistical analysis – in the form of a randomisation test on the combined Probe and Irrelevant conditions – confirmed the high significance of the increase in the Probe's power (ERSP) and coherence (ITC), as compared to the Irrelevant. Statistical tests of the *narrower* frequency band (0.5 to 7 Hz), resulted in two

independently measured time windows (ERSP average window: 350 to 450ms, and ITC average window: 345 to 445ms) and *p*-values that revealed a difference between the Probe and Irrelevant conditions (see table 4.3, below). For ERSP (average $p = 0.001$), all 14 subjects' *p*-values (100%) were significant – nine out of fourteen subjects' *p*-values (64%) for ERSP were highly significant ($p < 0.001$), and the other five were significantly below the critical alpha level of 0.05 (highest ERSP $p < 0.007$ belonged to subject 4).

Likewise, all ITC *p*-values (average $p < 0.0001$) were significant – thirteen out of fourteen subjects' *p*-values (93%) for ITC were highly significant ($p < 0.0001$), and one was significantly below the critical alpha level of 0.05 (highest ITC $p = 0.002$, which belonged to subject 8). As would be expected, a Fisher combining procedure, which was applied to ERSP and ITC to produce a single/joint *p*-value, resulted in a highly significant *p*-values (100%) for all subjects (average $p < 0.001$), at the this narrow frequency band (0.5 to 7 Hz).

Subject no.	*p*-values ERSP	aERPt win.	*p*-values ITC	ITC win.
1	< 0.0001	242	< 0.0001	324
2	< 0.0001	393	< 0.0001	352
3	< 0.0001	369	< 0.0001	313
4	0.007	443	< 0.0001	404
5	< 0.0001	283	< 0.0001	289
6	0.002	404	< 0.0001	381
7	0.005	422	< 0.0001	295
8	< 0.0001	357	0.002	369
9	< 0.0001	324	< 0.0001	334
10	< 0.0001	416	< 0.0001	387
11	0.003	289	< 0.0001	271
12	< 0.0001	363	< 0.0001	369
13	< 0.0001	363	< 0.0001	375
14	0.001	225	< 0.0001	363

*Table 4.3 – Subject-level Time Frequency analysis of power (ERSP) and coherence (ITC), at **Pz electrode**, using the narrower frequency range (**0.5 to 7 Hz**). For each subject, an orthogonal contrast time window was employed (using the aERPt method), and p-values were obtained for ERSP and ITC, by comparing the Probe and Irrelevant conditions, using a randomisation statistical test. At an alpha level 0.05, all ERSP p-values (100%) were significant (average p < 0.001), and all ITC p-values (100%) were significant (average p < 0.001).*

Even though we have justified the reason why the upper boundary of our analysis was fixed at 7 Hz (i.e. due to SSVEP waveform, which required a notch-filter on 7 to 9 Hz), we confirmed that per-subject statistical analysis of the maximum frequency range (0.5 to 45 Hz), resulted in *p*-values that revealed a difference between the Probe and Irrelevant conditions (see Table 4.4, below). For ERSP (average $p = 0.027$), eleven out of 14 subjects' *p*-values (79%) were significant, at 0.05 alpha level. Although, the other three *p*-values were very slightly above our 0.05 alpha level (highest ERSP $p = 0.093$, which belonged to subject 4), they were still within a minimal-alpha level of 0.1, which is often used as the level of significance in P3-based deception detection studies. As for ITC (average $p = 0.003$), all *p*-values (100%) were significant – seven out of fourteen subjects' *p*-values (50%) for ERSP were highly significant ($p < 0.0001$), and the other seven were significantly below the critical alpha level of 0.05.

Subject no.	*p*-values ERSP	aERPt win.	*p*-values ITC	ITC win.
1	0.006	197	< 0.0001	209
2	0.001	404	< 0.0001	375
3	< 0.0001	434	< 0.0001	219
4	0.093	443	0.003	398
5	0.002	283	< 0.0001	219
6	0.019	248	< 0.0001	416
7	0.086	209	0.004	289
8	0.035	543	0.025	439
9	< 0.0001	230	0.005	150
10	0.003	439	< 0.0001	422
11	0.044	588	< 0.0001	289
12	0.002	410	0.001	369
13	0.054	23	0.002	334
14	0.036	721	0.003	254

*Table 4.4 – Subject-level Time Frequency analysis of power (ERSP) and coherence (ITC), at **Pz electrode**, using the maximal frequency range (**0.5 to 45 Hz**). For each subject, an orthogonal contrast time window was employed (using the aERPt method), and p-values were obtained for ERSP and ITC, by comparing the Probe and Irrelevant conditions, using a randomisation statistical test. At an alpha level 0.05, 11 of 14 ERSP p-values (79%) were significant (average p-value = 0.027), and all ITC p-values (100%) were significant (average p = 0.003).*

4.4.6 – Other midline electrode sites

All the above Time and Frequency domain analyses focused on the Pz electrode, but we were also interested in the other two midline electrodes (Cz and Fz), to confirm that, in-line with (Kaufmann, Schulz, Grünzinger, & Kübler, 2011), the strongest brain responses to familiar faces are recorded at Pz. The following analogous Time domain analyses of Fz and Cz, aim to find out if N400f/P600f evoked by the Probe was significantly different from that evoked by the Irrelevant.

4.4.6.1 – Fz electrode

At the group-level, the grand average ERPs of the two critical stimuli (i.e. the Probe and Irrelevant conditions), at the Fz electrode site, revealed a clear difference between the conditions (see Figure 4.21, below). The Irrelevant condition, which consisted of an unknown/distractor face (paired with the Probe, and repeated randomly, as many times as the Probe), did not present any feature/pattern of interest (other than the SSVEP). This was as expected, because non-salient information is unlikely to breakthrough into conscious awareness, due to the high presentation rate of the RSVP streams. However, the Probe condition elicited a continuous oscillatory pattern, within a 150ms to 600ms time frame, with a frequency of approximately 4 Hz. This waveform at Fz is very similar to the oscillatory waveform at Pz (see Figure 4.7, above), and it confirms the prediction of a large difference between the Probe (celebrity face) and the Irrelevant (unknown face) conditions, at all midline electrodes. At this Fz electrode site, two orthogonal contrast time windows, for the lowest negative (N400f) and highest positive (P600f) components were independently found (using the AGAT method), at 322 to 422ms and 469 to 563ms, respectively.

Statistical analyses – in the form of a paired *t*-test of the mean amplitudes of Probe and Irrelevant, across all participants – were employed to find the group level significance of both components. Our statistical tests of N400f produced a highly

significant difference between the Probe (M = -2.2829, SD = 0.5804) and Irrelevant (M = 0.4262, SD = 0.1387) , at Fz electrode site: $t(13)$ = -4.756, $p < 0.0001$, d' = -1.768.

Similarly, our statistical tests of the P600f component produced a significant difference between the Probe (M = 1.7439, SD = 0.4378) and Irrelevant (M = 0.0013, SD = 0.1590), at Fz electrode site: $t(13)$ = 3.2376, $p < 0.006$, d' = 1.317.

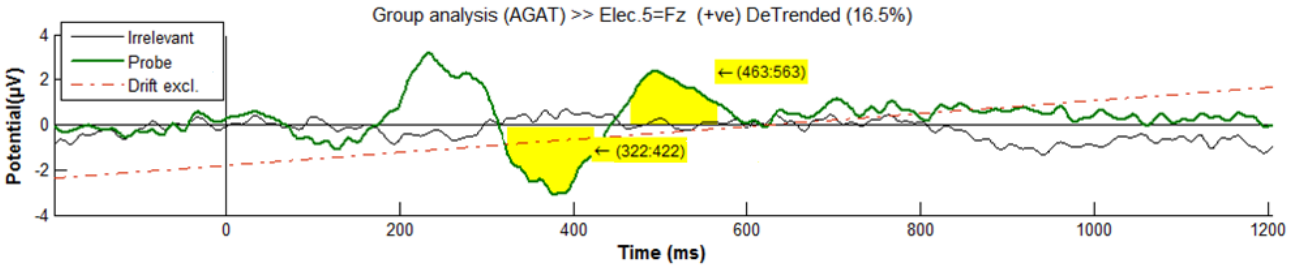

*Figure 4.21 – Grand average ERPs elicited by the Probe and Irrelevant, at the **Fz** electrode, showing an oscillatory pattern for the Probe condition (in green), Irrelevant (in black) and the Drift (in dotted-red), with a detrended drift at 16.5% to the vertical. Even though subjects were not informed of the presence of the Probe (celebrity face), statistical tests show a highly significant difference between Probe and Irrelevant, for **N400f** ($t(13) = -4.756$, $p < 0.0001$, $d' = -1.768$), and **P600f** ($t(13) = 3.2376$, $p < 0.006$, $d' = 1.317$).*

At the Fz electrode, subject-level statistical tests (i.e. Monte Carlo permutation) on the N400f component confirmed that 6 of 14 subjects (43%) showed critical-significance (0.05 alpha level) between Probe and Irrelevant. Statistical tests on the P600f component confirmed that 4 of 14 subjects (29%) showed a significant difference between Probe and Irrelevant. Finally, we were able to combine the p-values of the N400f and P600f components, for each subject, into a single p-value, by employing the Fisher combining procedure.

The following table summarises our subject-level results (see Table 4.5), at N400f and P600f components, as well as, the Fisher combined levels, whereby, 9 of 14 subjects (64%) achieved the critical significance (0.05 alpha level), for the Fz electrode. In terms of the number of subjects achieving significance, all three categories (i.e. N400f, P600f and Fisher) failed to match the equivalent results at Pz, agreeing with studies (Kaufmann, Schulz, Grünzinger, & Kübler, 2011) that report stronger brain responses (to familiar faces) at Pz.

SUBJECT	N400F	P600F	FISHER
1	0.018	0.121	0.026
2	0.234	0.014	0.031
3	0	0.672	0.011
4	0.485	0.366	0.496
5	0.087	0.041	0.029
6	0.002	0.073	0.009
7	0.238	0.001	0.01
8	0.046	0.723	0.166
9	0.09	0.808	0.266
10	0.047	0.175	0.048
11	0.065	0.43	0.145
12	0.312	0.1	0.148
13	0.875	0	0.006
14	0.001	0.6	0.012
AVERAGE:	0.179	0.295	0.100

*Table 4.5 – Subject-level analysis, at **Fz** electrode, for N400f and P600f components, and their Fisher combining probability. Statistical tests on N400f resulted in 6 of 14 subjects (43%) with significant p-values (0.05 alpha level), and statistical tests on P600f resulted in 4 of 14 subjects (29%) with significant p-values. The Fisher combining of N400f and P600f components resulted in 9 of 14 subjects (64%) being significant. All Fz categories failed to show a stronger brain response when compared to equivalent results at Pz.*

4.4.6.2 – Cz electrode

The same group-level analysis that was carried out at Fz (see section *4.4.6.1*), was performed at Cz, revealing a difference between the Probe and Irrelevant conditions (see Figure 4.22, below). Once again, the Irrelevant condition did not present any feature/pattern of interest (other than SSVEP), and the Probe condition elicited a continuous oscillatory pattern, within a 200ms to 600ms time frame, with a frequency of approximately 4 Hz. This waveform at Cz is very similar to the oscillatory waveforms at Pz and Fz (see Figures 4.7 and 4.21, respectively), and it confirms the prediction of a large difference between the Probe (celebrity face) and the Irrelevant (unknown face) conditions, at all midline electrodes. At this Cz electrode site, two orthogonal contrast

time windows, for the lowest negative (N400f) and highest positive (P600f) components were independently found (using the AGAT method), at 322 to 422ms and 467 to 567ms, respectively.

Statistical analyses, in the form of a paired *t*-test, were employed to find the group level significance of both components. Our statistical tests of N400f produced a highly significant difference between the Probe (M = -3.5505, SD = 0.6727) and Irrelevant (M = 0.4019, SD = 0.2066), at Cz electrode site: $t(13) = -6.1377$, $p < 0.0001$, $d' = -2.38$.

Similarly, our statistical tests of the P600f component produced a highly significant difference between the Probe (M = 2.1141, SD = 0.4446) and Irrelevant (M = 0.2187, SD = 0.1045), at Cz electrode site: $t(13) = 3.4301$, $p < 0.005$, $d' = 1.48$.

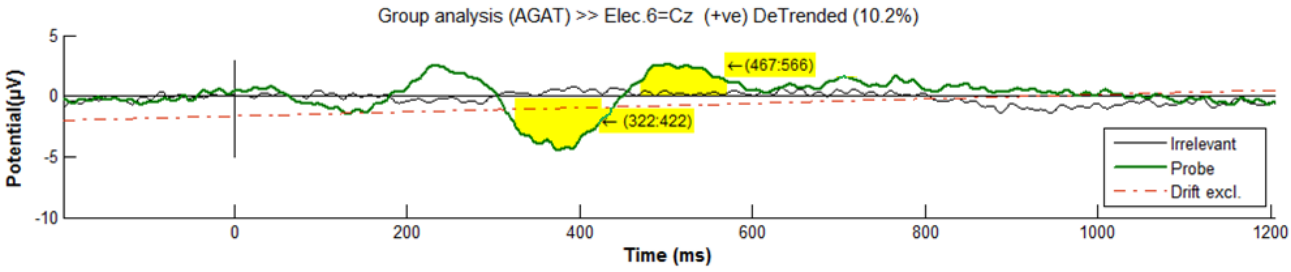

*Figure 4.22 – Grand average ERPs elicited by the Probe and Irrelevant, at the **Cz** electrode, showing an oscillatory pattern for the Probe condition (in green), Irrelevant (in black) and the Drift (in dotted-red), with a detrended drift at 10.2% to the vertical. Even though subjects were not informed of the presence of the Probe (celebrity face), statistical tests show a highly significant difference between Probe and Irrelevant, for **N400f** component ($t(13) = -6.1377$, $p < 0.0001$, $d' = -2.38$), and **P600f** ($t(13) = 3.4301$, $p < 0.005$, $d' = 1.48$).*

At the Cz electrode, subject-level statistical tests (i.e. Monte Carlo permutation) on the N400f component confirmed that 10 of 14 subjects (71%) showed critical-significance between Probe and Irrelevant (see Table 4.6). Statistical tests on the P600f component confirmed that 4 of 14 subjects (29%) showed a significant difference between Probe and Irrelevant. Finally, we were able to combine the *p*-values of the N400f and P600f components, for each subject, into a single *p*-value, by employing the Fisher combining procedure.

The following table summarises our combined Fisher subject-level results (see Table 4.6), whereby, 10 of 14 subjects (71%) achieved the critical-significance, for the Cz electrode. In terms of the number of subjects achieving significance at all three categories (i.e. N400f, P600f and Fisher), results at Pz beat Cz (albeit, Cz managed to beat Fz); once again, agreeing with studies (Kaufmann, Schulz, Grünzinger, & Kübler, 2011) that report stronger brain responses (to familiar faces) at Pz.

SUBJECT	N400F	P600F	FISHER
1	0.007	0.039	0.006
2	0.004	0.023	0.005
3	0	0.423	0.012
4	0.039	0.262	0.067
5	0.027	0.05	0.016
6	0.002	0.076	0.006
7	0.09	0.064	0.037
8	0.017	0.389	0.049
9	0.006	0.815	0.049
10	0.008	0.096	0.015
11	0.295	0.147	0.19
12	0.096	0.538	0.222
13	0.601	0.001	0.209
14	0.003	0.429	0.017
AVERAGE:	0.085	0.239	0.064

*Table 4.6 – Subject-level analysis, at **Cz** electrode, for N400f and P600f components, and their Fisher combining probability. Statistical tests on N400f resulted in 10 of 14 subjects (71%) with significant p-values (0.05 alpha level), and statistical tests on P600f resulted in 4 of 14 subjects (29%) with significant p-values. The Fisher combining of N400f and P600f components resulted in 10 of 14 subjects (71%) being significant. Whilst an improvement on Fz, Cz failed to show a stronger brain response when compared to equivalent results at Pz.*

Finally, we have demonstrated that all three midline electrodes (Pz, Fz and Cz) have exhibited similar oscillatory waveforms, and that statistical tests showed significant difference between the two conditions, Probe and Irrelevant. Although our choice to focus on the Pz electrode was a priori (in-line with (Kaufmann, Schulz, Grünzinger, & Kübler, 2011)), we found that the strongest brain responses to familiar faces, was indeed recorded at the Pz electrode.

4.5 Discussion

The primary aim of this experiment was to investigate whether faces can be used in an RSVP sub/liminal search paradigm (Bowman, et al., 2013). More specifically, in the first experiment of its kind, our objective was to examine whether famous faces can breakthrough into conscious awareness, and that the breakthrough event can be detected by EEG, on a per individual basis. We intended to achieve this through statistical analyses of the ERP data (in the Time domain) and single-trial data (in the Frequency domain), to determine whether the evoked response by the Probe (celebrity) faces were significantly different from that evoked by the Irrelevant (unknown) faces. The null hypothesis (H0) was that there is no difference between the Probe and Irrelevant conditions. Our experimental hypothesis is that there is a difference.

With the key comparison between Probe faces and Irrelevant faces, a significant difference was observed between the ERPs, at all three mid-line electrodes (Pz, Fz and Cz). In particular, an enhanced negative deflection, followed by a positivity, over the time frame 300ms to 600ms (for the Probe condition only) was identified, which is in-line with previous studies that investigated familiarity effects using faces (Bentin & Deouell, 2000); (Eimer, 2000); (Touryan, 2011). Albeit, the studies referenced here used a much higher Stimulus Onset Asynchrony (SOA) of 350 to 500ms, whereas, we have employed the RSVP technique (using 133ms SOA), which presents the images at a very fast rate, in order to allow the brain to process the salient stimuli only, on the fringe of human awareness. As such, we have successfully demonstrated that a comparison between unknown and familiar faces can be made, in an RSVP paradigm, using traditional ERP analysis in the time domain, as well as, the novel Time Frequency Analysis (TFA) of the oscillatory activity, in the frequency domain.

Therefore, our experiment's statistical test results demonstrated the viability of using faces in the RSVP paradigm, in order to infer recognition of familiar faces. Even though subjects were not informed that familiar faces may appear in the RSVP streams, our statistical tests confirmed the breakthrough events. Having been instructed to only look for the Target face (i.e. an unknown face that subjects were trained to recognise), the inclusion of Probe faces (i.e. celebrities – not associated with the explicit task) was meant to examine the subject's ability to perceive intrinsically salient faces. However, we acknowledge that once our subjects perceived a familiar (Probe) face, they were likely to look for more of these (familiar) faces. Even so, statistically testing the brain

responses by comparing Probes and Irrelevants, in the Time, as well as, the Frequency domains, enabled us to confirm the Probes' significance over the Irrelevants, at group and subject levels.

Finally, we have standardised the use of detrending to independently remove any drift in EEG data, and to avoid the legacy practice of post/ad-hoc increasing of the high-pass filter, which may adversely affect the low frequency P3 component and/or introduce waveform distortions. Furthermore, we have relied on the Fisher combining procedure to aggregate multiple probabilities, and to produce a single *p*-value across dimensions (e.g. N400f/P600f components).

4.5.1 – Time Domain

At the Pz electrode, subject-level statistical analyses of ERPs confirmed that, having found the orthogonal contrast window for the N400f component (average time window: 322 to 422ms), a total of 10 of 14 subjects (71.4%) had *p*-values below our critical-significance (alpha level 0.05), of which 6 subjects have highly significant *p*-values (0.01 or below), revealing a difference between the Probe and Irrelevant conditions.

As for P600f – with an independently identified region of interest, at an average time window of 479 to 579ms – a total of 7 of 14 subjects (50%) had *p*-values below our critical-significance (alpha level 0.05). Incidentally, at a *minimal-significance* (i.e. an alpha level of $p < 0.1$), as used in most of Farwell and Rosenfeld's deception detection studies (Farwell & Donchin, 1991); (Rosenfeld I. P., 2008), we could report 9 of 14 subjects (64.3%) with significant difference between Probe and Irrelevant, within the P600f component time window.

Using the Fisher combining procedure, we aggregated the *p*-values of the N400f and P600f components, for each subject, and showed that 10 of 14 subjects (71.4%) achieved critical-significance (alpha level 0.05), of which, 6 subjects (42.9%) were highly significant (alpha level 0.01). Furthermore, at the *minimal-significance* level (alpha level 0.1), all 14 subjects (100%) achieved Fisher combined levels of significance.

The results of our statistical analyses, within the Time Domain, provide evidence that the celebrity faces (Probe conditions) were differentially perceived and processed

by all subjects' brains, as compared to the unknown faces (Irrelevant conditions). Even though both conditions were treated equally, our experimental findings show major differences between the Probe and Irrelevant, which was as a result of the former stimuli reaching conscious awareness and generating pronounced electrical responses (as seen in the Probe ERPs), whilst the latter was not sufficiently perceived to encode into working memory, in order to generate a distinct electrical response.

4.5.2 – Frequency Domain

At the Pz electrode, subject-level statistical analyses of Time Frequency (across the *entire* frequency range, 0.5 to 45 Hz), using the independently measured time window for ERSP, confirmed that 11 out of 14 subjects' *p*-values (79%) were significant (at the 0.05 alpha level). Further, the same statistical tests on the *narrower* frequency band (0.5 to 7 Hz, attainable due to SSVEP waveform, which required a notch-filter of 7 to 9 Hz) showed that all 14 subjects' *p*-values (100%) were significant – nine out of fourteen subjects' *p*-values (64%) for ERSP were highly significant ($p < 0.001$). As for ITC, subject-level statistical analyses of Time Frequency (across the *entire* frequency range: 0.5 to 45 Hz, as well as, the *narrower* frequency band: 0.5 to 10 Hz) showed that all subjects' *p*-values (100%) were significant.

The results of our statistical analyses, within the Frequency Domain, provide additional evidence that the Probe (celebrity) faces were differentially perceived and processed by all subjects' brains, as compared to Irrelevant (unknown) faces. The large increases in power (ERSP) and coherence (ITC), which were observed and statistically confirmed in the Probe condition only, demonstrate that such changes in power and phase-locking coherence could have contributed to the generation of the components N400f/P600f, which were elicited within similar time windows of the same condition's Probe ERPs. This finding supports the hypothesis that oscillatory activity, in the frequency domain, is related to the ERP component, in the time domain (Makeig, Debener, Onton, & Delorme, 2004b) (Fuentemilla, 2008).

4.5.3 – Conclusion

This chapter's experimental findings confirm our first hypothesis that having substituted faces in place of word/numbers/letters, within an RSVP paradigm (i.e. in place of the previous critical stimuli that were used in similar EEG experiments), we were able to detect the group-level breakthrough of highly familiar/famous faces into consciousness. Furthermore, we agree that such breakthrough would be encoded in brain signals (Bowman, et al., 2013), and would generate ERP components/effects that would differ between the Probes (celebrity faces) and the Irrelevants (unknown faces). Through the effective use of our statistical analyses, in the time domain (using ERPs), as well as, the frequency domain (using single-trials), and the introduction of our standard new statistical testing techniques (e.g. detrending and independent window placement), we have successfully differentiated between the two conditions, and have since published our findings for this experiment (Alsufyani, et al., 2019).

Our second hypothesis was that in addition to the breakthrough of Probe faces at group-level, we can also use ERPs to detect the breakthrough events on an individual basis. With the aid of the Fisher combining method, our statistical tests, in the Time and Frequency domains, confirm the presence of large differences in brain responses for the Probe and Irrelevant conditions – having treated them equally in the experiment – at subject-level. Therefore, we infer that this approach can be used to determine whether a subject has high familiarity of a well known individual (e.g. a celebrity).

Our third and final hypothesis was that the strongest brain responses to the familiar (Probe) faces are recorded at the Pz electrode site. Having carried out the same statistical tests on all midline electrodes (Fz, Cz and Pz), we can confirm that whilst all three sites exhibited similar oscillatory waveforms for the Probe, the strongest brain responses to familiar faces was, indeed, recorded at the Pz electrode, in-line with other/similar studies (Kaufmann, Schulz, Grünzinger, & Kübler, 2011).

4.5.4 – Future work

This chapter's experiment was the first step in demonstrating that faces can be employed in RSVP-based fringe-P3 studies, and that highly familiar faces can breakthrough into conscious awareness, on an individual (subject-level) basis. The

results suggest that we can apply our findings to the differentiation of deceivers and non-deceivers, in the application of crime compatriots, whereby, a suspect's familiarity with a criminal/terrorist can be established using faces.

However, we recognised that famous faces of celebrities – who are often, rich, successful, arguably good looking and probably admired/hated – are highly recognisable, and, therefore, may have a greater impact on the breakthrough effect that is experienced by our subjects. In recognition of this potential incongruity, our next experiment (see Chapter 5) will substitute famous faces with familiar faces (i.e. real-life acquaintances), in the form of University lecturers, who have had long-term and close relationships with participants (i.e. their students).

Additionally, we shall improve the design of the next experiment, to mitigate the unintentional revealing of the nature of our experiment, whereby, in the previous (celebrity faces) experiment, after the presentation of the first critical item (i.e. at the end of the first block), the subject was asked if s/he recognised two (comparative) celebrity faces – this recognition question may have revealed to the subject, the fact that the experiment contained more celebrity faces, in the ensuing blocks (i.e. even if the subject had not perceived the celebrity face in the first block, the recognition question could give the game away). Therefore, we will move the recognition test, which used to take place at the end of each block, to the end of the experiment (i.e. after all blocks of the experiment have been completed). As a result, the next experiment will take us one step closer to using faces in RSVP-based EEG tests for deception detection applications of compatriots.

Chapter 5:

EEG study 2 – Recognition of Concealed Lecturer Faces

5.1 Introduction

So far, we have successfully established that famous/celebrity faces can breakthrough into conscious awareness, using an RSVP subliminal search paradigm, and that our statistical tests can differentiate between the Probe (celebrity) and Irrelevant (unknown) faces, at group and subject levels (see chapter 4). The objective of the current chapter was to demonstrate that we can substitute the highly evocative faces of famous celebrities with familiar faces that are personally known to the participants, in the form of the University's lecturers. Furthermore, we aimed to use the same standard analysis methods, established in the first experiment, to differentiate between the Probe (familiar University of Kent lecturer) and Irrelevant (unknown Christchurch University lecturers), at group and subject levels. Ultimately, this chapter's aim was to act as a bridge between the proof of concept (i.e. the first experiment) and the working prototype (i.e. the third-and-final experiment), which could be employed as a scientifically robust solution for deception detection applications of compatriots, using faces in RSVP-based EEG tests.

Once again, by referencing the standard design and analysis methods, described in Chapter 3, we will begin by outlining the familiar/lecturer faces experiment, and then summarise the group-level analysis. Next, we will describe our in-depth group and subject level analyses, in the Time (ERP) and Frequency (ERSP/ITC) domains. Finally, we will discuss the results and draw conclusions to our hypotheses, based on the evidence gathered.

5.2 Experiment's Hypotheses

Taking one step closer to exploring the suitability of the RSVP paradigm, to infer the recognition of familiar/compatriot's faces, in real-life EEG-based deception

detection tests, we substituted the first experiment's famous *celebrity* faces with familiar *lecturer* faces, in order to test the following hypotheses, experimentally:

i) Familiar faces of University lecturers that are personally known to the subject can be used in the RSVP paradigm, instead of the highly recognisable celebrity faces, and that the fringe-P3 method can be employed to detect the group-level breakthrough of Probe (familiar lecturer) faces, which are differentially perceived and processed, as compared to Irrelevant (unfamiliar lecturer) faces;

ii) In addition to the group-level breakthrough of Probe (lecturer) faces, we can detect the breakthrough events on an individual basis, even though, only the Target was task-relevant and subjects were not made aware of the appearance of the Probe conditions;

iii) In accordance with the first experiment, the strongest brain responses to the lecturer (Probe) faces are recorded at the Pz electrode site.

5.3 Design of the second Experiment

5.3.1 – Experiment's Participants

Fourteen participants were tested and none were excluded. Out of 14 subjects, 12 were male (86%) and 2 female (14%). The ages of the subjects ranged from 22 to 37 (M = 27.5 years, SD = 3.94); 12 of them were right-handed (86%), and 2 were left-handed (14%). Because this experiment's Probe stimuli consisted of University of Kent's lecturer faces, from the School of Computing, we needed subjects who had a close working relationship with their lecturers. Therefore, we asked the lecturers to covertly nominate PhD students only, so that we could be assured of a long-term relationship/familiarity between subjects and their lecturers' faces. Also, subjects were chosen on the basis of never having been included in a similar EEG/RSVP experiment, and at the end of each experiment, participants were instructed to avoid discussing the experiment with their colleagues, in order to avoid any priming of future participants. The duration of each experiment was (approx.) 1 hour and 45 minutes, and each subject was paid £10 (ten pounds) for their time.

5.3.2 – Experiment's Stimuli

As outlined in Chapter 3 (see section 3.2.2), the stimuli were split into two groups: **Distractors** (i.e. 524 unknown faces) and **Critical** images. The Critical group was further split into 3 categories: **Target** image (a single face that became task-relevant), **Irrelevant** images (unknown faces of Lecturers from another University) and **Probe** images (familiar Lecturer faces who are well known to the subject). Having photographed a large portion of the lecturers in the University of Kent's School of Computing (23 images in total), we were able to assign three Lecturers (as Probes) to each subject, knowing that they were highly familiar with one another (as confirmed by the Lecturers and/or the subject's colleagues). Additionally, each subject was randomly assigned three unknown lecturers (as Irrelevants) from a different University (i.e. Canterbury Christ Church University), whose photographs were taken with the same camera (9 images in total), and treated in the same manner as all the Probe images (for detailed explanation of the standards/methods used to take the photographs and edit them, please refer to section 3.2.2 – *Stimuli*).

5.3.3 – Experiment's design

As outlined in Chapter 3 (see section 3.2.3), each RSVP stream's 18 faces included a single Critical stimulus and 17 Distractors (with an SOA of 133ms). The Critical stimuli in each RSVP stream could either be a Probe (i.e. one of 3 familiar Lecturer faces), or an Irrelevant (i.e. one of 3 unknown Lecturer faces), or the Target (i.e. a single face that is task-relevant).

In total, Probes, Irrelevants and Targets were presented an equal number of times, and (in a statistical sense) in the same position in streams. As explained in the Introduction, the primary change between this (the second) experiment and the previous (the first) experiment was that the Probes became *Lecturer* images, instead of famous *Celebrity* images (noting that the Irrelevants were also changed from unknown

Distractor images to unknown Lecturer images). However, we also implemented a change to the design of the experiment: to improve the signal-to-noise ratio of unique Probe/Irrelevant Critical stimuli, we reduced the number of Probes/Irrelevants from five (celebrities) to three (lecturers), whilst maintaining the total number of trials for each subject (i.e. 225 trials in total, so that the number of trials per Critical condition would also remain the same between the two experiments). Therefore, in this (the second) experiment, each Probe was repeated 25 times, resulting in 75 Probe-trials (i.e. 25 times for each of the 3 Probes), and each Irrelevant was also repeated 25 times, resulting in 75 Irrelevant-trials (i.e. 25 times for each of the 3 Irrelevants). The single Target was, therefore, repeated 75 times, to equal the number of times that the other two Critical Stimuli category were included in RSVP streams. The resultant 225 RSVP trials were divided into 3 blocks, each block comprising 75 trials (i.e. 25 Probe trials, 25 Irrelevant trials, and 25 Target trials), and the order of the three Critical stimuli were randomised within the blocks. However, each block's Probe and Irrelevant Critical stimuli were paired, so that the same known lecturer (Probe) face and unknown lecturer (Irrelevant) face were presented within the same block – this will enable us to make direct comparisons between these paired-conditions.

Just as in the first experiment, subjects were told to keep their eyes fixed at the centre of the screen during the presentation of the RSVP stream (lasting approx. 2.5s), and to avoid movement or blinking. Also, they were informed that the Target image will appear pseudo-randomly, so they should not expect it in every trial, however, subjects were naïve to the presence of familiar lecturer faces (i.e. Probes).

In this experiment, out of a total of 75 trials for each Critical condition, the number of trials that remained after artefact rejection, per condition, ranged between 59 and 75, and none of the subjects were excluded from the analysis due to removal of artefact trials (e.g. excessive eye blinks):

Target (M = 70.5, SD = 4.26);
Probe (M = 71.93, SD = 3.27);
Irrelevant (M = 72.64, SD = 2.17).

5.3.4 – Experiment's Target Questions

As explained in Chapter 3, at the end of each RSVP stream, the subject was required to answer two question (see section 3.2.3), using a dedicated keypad, which was placed under the subject's right/left hand (whichever hand the subject preferred to use). The first question related to the finishing-item, which required the subject to select either key '1' or '2', and the second question related to Target-recognition, which could be answered using either key '4' or '5'.

Before starting the experiment, the subject was shown the Target image – this would be the same image for all subjects – which was chosen from the Distractor (i.e. unknown) database, and therefore, not familiar to the subject. Even so, the subject was asked, in the beginning, if they had ever seen, or could recognise, the Target face (none of our subjects had ever seen the Target face). As this is a task-based experiment, the subject was instructed to look only for that Target image, in each of the RSVP streams, and to expect a recognition question: "Did you see the Target face?", at the end of each trial (noting that this recognition question followed the finishing-item question). If the Target image was seen, the subject was instructed to answer 'Yes' (using '4' key), or 'No' if it was not perceived (using '5' key). If the Target was present, a 'Yes' (i.e. correct) answer would be a "HIT", and a 'No' (i.e. incorrect) answer would be a "MISS". Conversely, if the Target was absent, a 'Yes' (i.e. incorrect) answer would be a False-Positive (FP), and a 'No' would be a correct rejection (see Table 5.1).

Target	HIT (75)	FP (150)
Subject 1:	70	3
	93.3%	2.0%
Subject 2:	74	12
	98.7%	8.0%
Subject 3:	40	2
	53.3%	1.3%
Subject 4:	74	5
	98.7%	3.3%
Subject 5:	68	3
	90.7%	2.0%
Subject 6:	32	6
	42.7%	4.0%
Subject 7:	47	1
	62.7%	0.7%
Subject 8:	58	2
	77.3%	1.3%
Subject 9:	73	4
	97.3%	2.7%
Subject 10:	30	6
	40.0%	4.0%
Subject 11:	34	15
	45.3%	10.0%
Subject 12:	62	6
	82.7%	4.0%
Subject 13:	31	0
	41.3%	0.0%
Subject 14:	69	4
	92.0%	2.7%
Mean:	54.43	*4.93*
	72.6%	3.3%
SD:	17.87	4.12

Table 5.1 – Subjects' HIT count (i.e. number of times that the subject correctly reported seeing the task-relevant Target face, in 75 trials), and False-Positive (FP) count (i.e. reported seeing the Target when it was not there, in the other 150 trials) are shown.

Group HIT rate of 54.43 (72.6%) and FP rate of 4.93 (3.3%), with corresponding MISS rate of 20.57 (27.4%) and correct rejection of 145.07 (96.7%), result in a response sensitivity measure of d' = 2.49.

Out of 75 times that each subject was randomly presented with the Target face, the average Hit rate for the group was 72.6% (M = 54.43, SD = 17.87), and out of the

remaining 150 other trials in which the Target was not presented, the False-Positive rate was 3.3% ($M = 4.93$, $SD = 4.12$). The resulting sensitivity measure ($d' = 2.494$) was within our tolerance range, and no subjects were excluded due to low sensitivity or high bias.

5.3.5 – Experiment's Probe/Irrelevant Questions

Another change that we introduced, to improve the design of the experiment, was to move the end of block familiarity question to the end of the experiment, in order to remove the unintentional revealing of the Probes (lecturers), during the experiment. As we have noted earlier, in the previous (celebrity faces) experiment, the subject was asked if s/he recognised two (comparative) celebrity faces, at the end of each block; this recognition question may have revealed the fact that the experiment contained more celebrity (Probe) faces, in the remaining blocks. Thus, by moving the end of *block* familiarity question to the end of the *experiment*, we can avoid the unintentional revealing of the lecturer (Probe) faces, which is especially important when the subject had not perceived the lecturer face in any of the previous blocks.

Therefore, at the end of the experiment (rather than end of block), the subject was given a recognition test, in the form of memory questions, to determine if the three Probes and/or the three Irrelevants were perceived/recognised. This memory test consisted of 12 questions, appearing randomly, where each question accompanied an image that may or may not have been included in the experiment's three blocks. Six questions related to the presence of the paired Probe and Irrelevant Lecturer faces that were included in the three blocks (i.e. one pair of Probe/Irrelevant, per block), and the other six questions related to random Probe and Irrelevant (Lecturer) faces that were not included in the experiment. Whereas the former six questions (about the Probe/Irrelevant faces that were presented) would gauge the subject's ability to perceive faces that were included in the experiment, the latter six questions assess the subject's engagement with the tests (i.e. were subjects guessing the presence of salient faces?).

As before, the response to each of these 12 recognition/memory tests were handled in two parts: firstly, what is the subject's confidence rating of how often each face was presented (i.e. the Probe/Irrelevant that were present, and the Probe/Irrelevant that were absent), and secondly, a confidence rating of how well the subject knew each of the 12 faces, prior to the experiment. The responses to both of these confidence ratings used a scale of 1 to 5, where 1 is "Never", 2 is "Once or twice", 3 is "Few times", 4 is "Many times" and 5 is "A lot". Note, for the purposes of statistical comparison, 1 out of 5 (i.e. Never) is equivalent to 0% and 5 out of 5 (i.e. A lot) is equivalent to 100%. Thus, 2 out 5 = 25%, 3 out of 5 = 50% and 4 out of 5 = 75% (see Appendix B.1 for the full results).

5.3.5.1 – Overall Probe/Irrelevant recognition

The three Probe (familiar-lecturer) faces that were included in the experiment were reported to have been seen 33.9% of the time (Mean confidence rating of 2.4 out of 5), and subjects reported a very high (pre-experimental) familiarity of 94.6% with these Lecturer faces (4.8 out of 5). When comparing this to the (absent) Probe faces that were not included in the experiment, subjects reported a similar high (pre-experimental) familiarity of 79.2% (4.2 out of 5), and only reported seeing these 'absent' familiar-lecturers 3.6% of the time (1.1 out of 5), which is approximately one-ninth of the Lecturers that were, indeed, included in the experiment.

The three Irrelevant (*unknown*-lecturer) faces that were included in the experiment were reported to have been seen 4.8% of the time (1.2 out of 5), and, similarly, the absent Irrelevant faces that were not included in the experiment were reported to have been seen, at an average of 2.4% of the time (1.1 out of 5). Finally, subjects reported an imperceptible (pre-experimental) familiarity of 0% with all the Irrelevant/*unknown*-lecturer faces (1.0 out of 5).

As we were comparing Probe faces with Irrelevant faces, it was encouraging to discover that Probes were reported 33.9% of the time ($M = 2.4$; $SD = 1.2504$), which was seven times more than Irrelevants that were reported 4.8% of the time ($M = 1.2$;

SD = 0.428). Note that both conditions (Probes and Irrelevants) were, in fact, presented an equal number of times. The mean confidence rating of the main comparison conditions, for all subjects, reveals a highly significant difference between the Probe (known-lecturer) faces and the Irrelevant (unknown-lecturer) faces, using pair-wise comparison ($M = 1.1714$, $SD = 1.5122$), $t(13) = 2.898$, $p = 0.0125$, $d = 1.2545$).

5.3.5.2 – By-item Probe recognition

Unlike the previous (first) experiment, which used the same five Probe (celebrity) faces for all 14 subjects, the current (second) experiment was required to match subjects to their most familiar lecturers (i.e. the Probes with the highest familiarity). Therefore, the three Probe (familiar-lecturer) faces that were chosen for each subject could be different, and a '*by-item*' comparison (e.g. all 25 trials in the first block of all 14 subjects) would not show the response to an individual lecturer (i.e. the first block of the experiment's 14 subjects may consist of up to 14 different lecturers). However, we carried out *by-item* comparisons between the 3 blocks of all subjects, in order to quantify the group-level response to the Probe recognition/memory tests (see Figure 5.1, below).

The most detected Probe (familiar-lecturer) face was block-1's lecturer (50%), and the least detected was block-3's lecturer (21.4%), bearing in mind that different lecturers may be involved in each block. This result is in-line with the design of the experiment, whereby we ordered the 3 Probes, for each subject, by their predicted (pre-experimental) familiarity (e.g. subjects' supervisors were selected as the first block's Probe). The effect of prioritising the order of the familiar Probes appears to have overcome any recognition advantages that may be gained as a result of a training effect (i.e. the greater the exposure to RSVP streams, the more likely it is that the subject will perceive the salient Probes, in future blocks), as a one-way ANOVA confirms that there is no statistically significant difference between them ($F(2, 39) = 1.9031$, $p = 0.1627$).

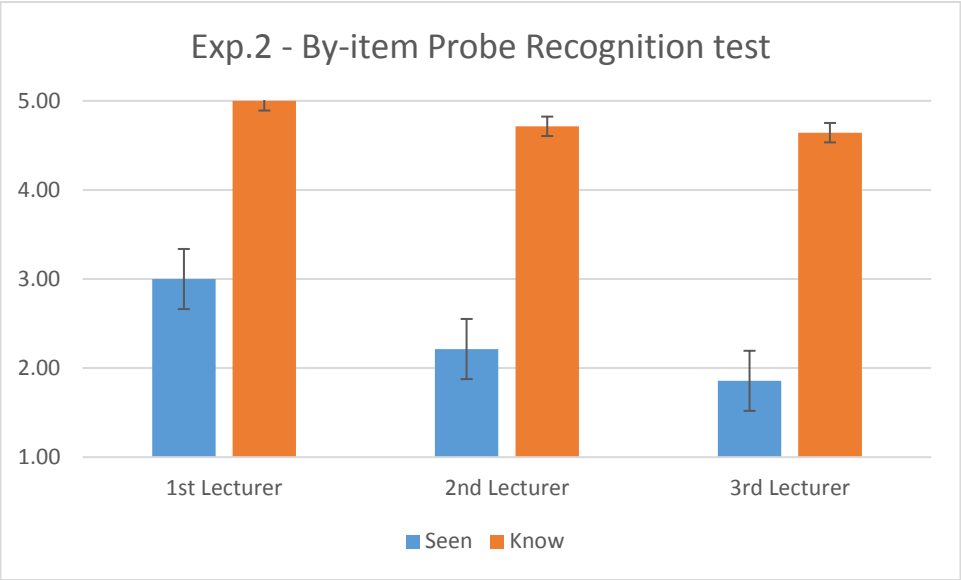

Figure 5.1 –Experiment 2's by-item Probe (familiar-lecturer face) recognition tests: "Seen" rates (i.e. confidence rating of having detected the Probe) and "Know" rates (i.e. how well the subject recognises the Probe). On average, 50% of subjects had seen the first-lecturer (rating = 3.0, SE:0.5), 30% had seen the second-lecturer (rating = 2.2; SE: 0.39), and 21% had seen the third-lecturer (rating = 1.9; SE: 0.36). One-way ANOVA on the 'seen' ratings for the 3 lecturers confirms that there is no statistically significant difference between the means (p = 0.1627). As expected, subjects' familiarity (i.e. 'Know' ratings) with all 3 presented lecturers was very high (see Appendix B.1 for full details).

As with the first experiment, the behavioural data (i.e. all the above online responses to recognition questions) provided a useful indicator of the perceptual state of the subjects' mind, however, the primary aim of our research was to use the EEG data to detect the breakthrough of Probe (familiar lecturer) faces, which could be differentially perceived and processed, as compared to Irrelevant (unknown lecturer) faces. Therefore, the rest of this chapter will focus on the analysis of the EEG data, in the Time and Frequency domains.

5.4 Data Analyses

5.4.1 – Summary of Analysis

As in the first (celebrity faces) experiment, we were interested in the EEG data across all the midline electrodes (Pz, Cz and Fz), but in-line with (Kaufmann, Schulz, Grünzinger, & Kübler, 2011), we expect the strongest brain responses to familiar faces, to be recorded at the Pz electrode. Therefore, we will start by focusing on the Pz electrode, reporting Time and Frequency domain analyses (at group and subject level), before reporting the same analyses at Fz and Cz.

5.4.1.1 – Pz Electrode

At group-level, the grand average ERPs of all three critical stimuli (i.e. the Target, Irrelevant and Probe conditions), at the Pz electrode site, revealed a clear difference between the conditions (see Figure 5.2, below). The Target condition was task-relevant, so it elicited a large classical P3, which was as expected because subjects were instructed to detect the Target face, throughout the experiment. The Irrelevant condition, which consisted of an unknown face (paired with each Probe, and repeated randomly, as many times as the Probe), did not present a similar pattern to the Probe or Target; this was as expected, because non-salient stimuli were unlikely to breakthrough into conscious awareness, due to the high presentation rate of the RSVP streams. In a similar fashion to the first (celebrity faces) experiment, the Probe condition elicited a continuous oscillatory pattern, within a 280 to 620ms time frame (and observed frequency of approximately 3-4 Hz).

In the first (celebrity faces) experiment, we confirmed our hypothesis that there is a large difference between the Probe (celebrity face) and the Irrelevant (unknown face) conditions, and in the current (lecturer faces) experiment, we expected a similar effect – this was confirmed in the group-level ERPs (see Figure 5.2). Furthermore, the first experiment revealed an oscillatory pattern for the Probe (celebrity) faces, which was recorded for the first time in an RSVP-based study of faces, and we predicted a

similar Probe pattern in the current (lecturer faces) experiment – this was also confirmed in the group-level ERPs, albeit, the oscillatory pattern lacked the same N400f peak voltages of the first experiment, which may affect the significance of the main comparison conditions, between the Probe and Irrelevant.

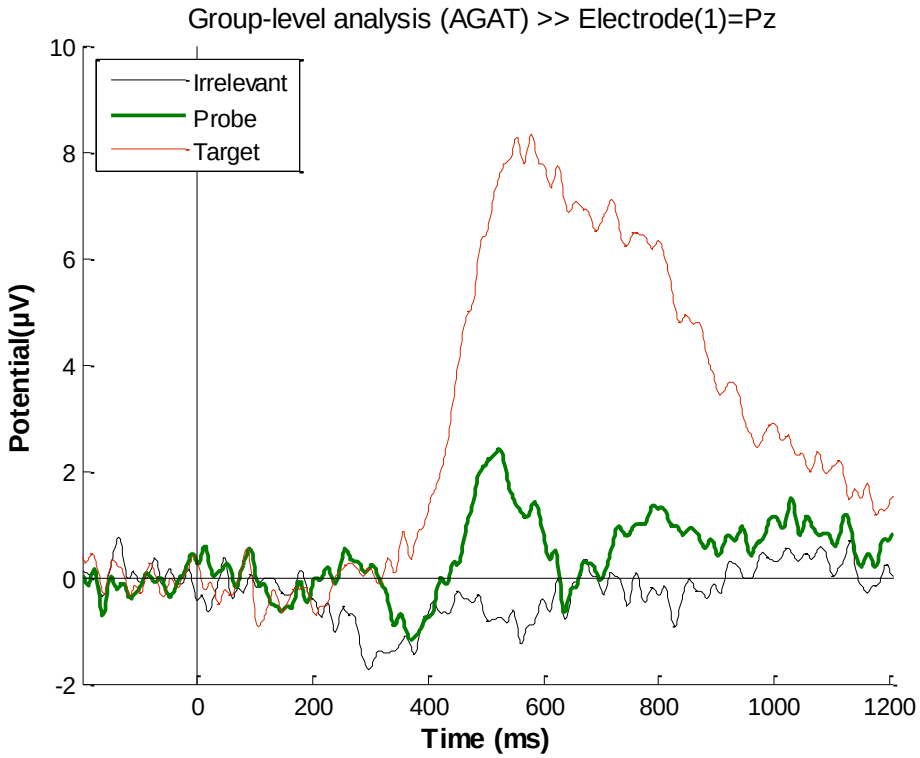

*Figure 5.2 – Grand average ERPs elicited by the three critical stimuli (Irrelevant, Probe and Target conditions), at Pz electrode, showing a P3 pattern for the **Target** (in red, peaking at +8.5μV), an oscillatory pattern for the **Probe** (in green, with an observed frequency of approx. 3-4 Hz),), and a different pattern for the **Irrelevant** (in black, containing an interesting negative deflection, peaking at 300ms) that is distinct from the Probe and Traget. As before, Target was the stimulus that the subject was instructed to look for, whereas, they were not informed of the presence of the Probe (familiar lecturer face). The oscillatory pattern for the Probe suggests a significant difference with the Irrelevants (unknown lecturer faces), which were presented as many times as the Probe. Whilst the Probe's oscillatory pattern, and the peak positivity around 500ms (i.e. P600f) is similar to the first (celebrity faces) experiment, the peak negativity around 400ms (i.e. N400f) is considerably smaller in amplitude.*

By collating and stacking all the trials (i.e. every single trial for all subjects in the group) for the Target condition, we observed a prevailing positivity, from 400ms onwards, for most trials, at the Pz channel. This channel-oriented representation of the trials was confirmed by the aggregated ERPs (see left plot of Figure 5.3), and the spatial

dispersion of resultant waveform was depicted by the ERP scalp topographies (see right plot of Figure 5.3), which confirmed the Target condition's dominant positive wave, peaking at around 550ms.

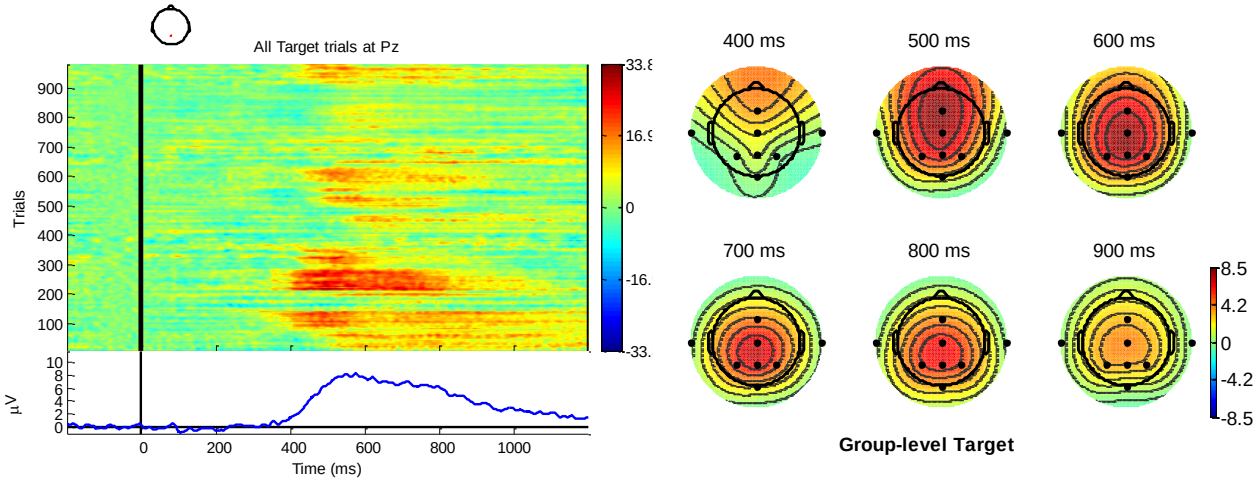

Figure 5.3 – Group-level view of all (987) Target trials, in order of appearance over time, at Pz (left plot), and the corresponding scalp topography of the ERPs (right plot), showing a prevailing positivity, peaking at around 550ms, with the electrical field moving posteriorly through time. Having used a limited number of (8) electrodes in this experiment, it must be noted that MATLAB employs an interpolatory algorithm to represent the full scalp pattern. Therefore, estimated electric potential values are used at scalp locations between the actual recording sites, and the presented scalp topographies carry considerable uncertainty, especially in respect of laterality of effects, since we have few electrodes beyond the central line. Note that the scalp map scale ranges from -8.5 to +8.5 μV.

Next, we stacked all the trials in the Probe condition, for all subjects at Pz, and observed the oscillatory waveform, with its peak negativity at around 350ms, and its peak positivity at around 500ms. However, in comparison with the first (celebrity faces) experiment, the current (lecturer faces) experiment's Probe possessed a weaker N400f feature, which was an unexpected outcome that requires further study. The Pz channel ERP image and the interpolatory scalp topography of the ERPs can be seen below (see Figure 5.4).

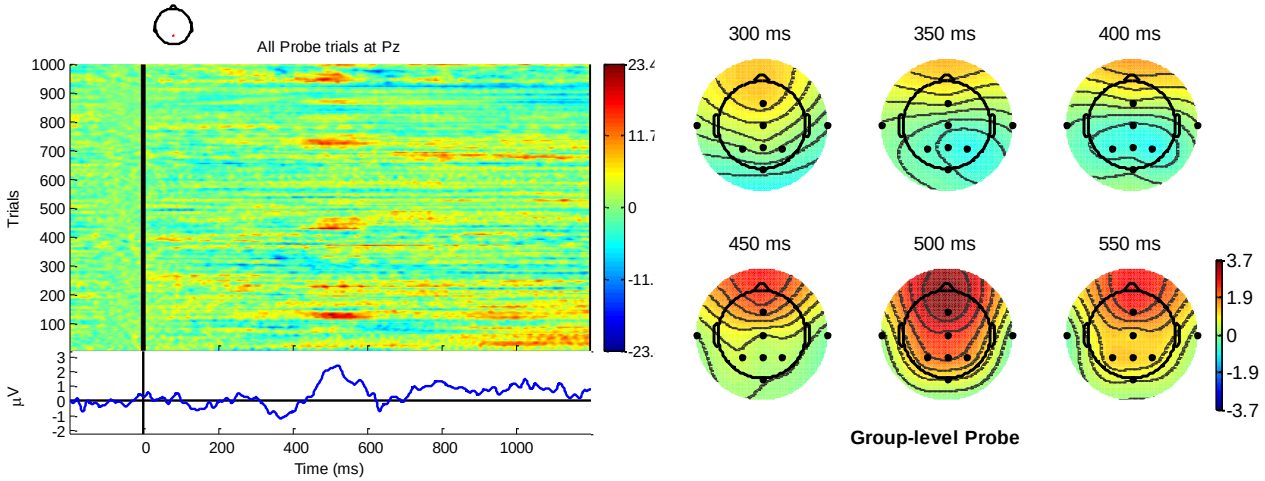

Figure 5.4 – Group-level view of all (1007) Probe trials, over time, at Pz (left plot) and interpolated scalp topography of the ERPs (right plot, which must be treated with caution, due to the small number of electrodes), showing a similar oscillatory pattern to the previous (celebrity faces) experiment (albeit, the N400f effect is considerably weaker). Lowest negativity can be observed at 350ms, and highest positivity at 500ms. Note that the scalp map scale ranges from -3.7 to +3.7 μV, which is lower than the scale for Target (see Figure 5.3).

As shown in Figure 5.5 (below), the Irrelevant condition did not show a similar oscillatory pattern that was observed in the Probe condition, supporting our hypothesis that unknown lecturer faces, presented at a rapid rate, will not breakthrough into conscious awareness. However, a small negativity (peaking at 300ms) was observed for the first time (i.e. it was not present in the previous *celebrity faces* experiment).

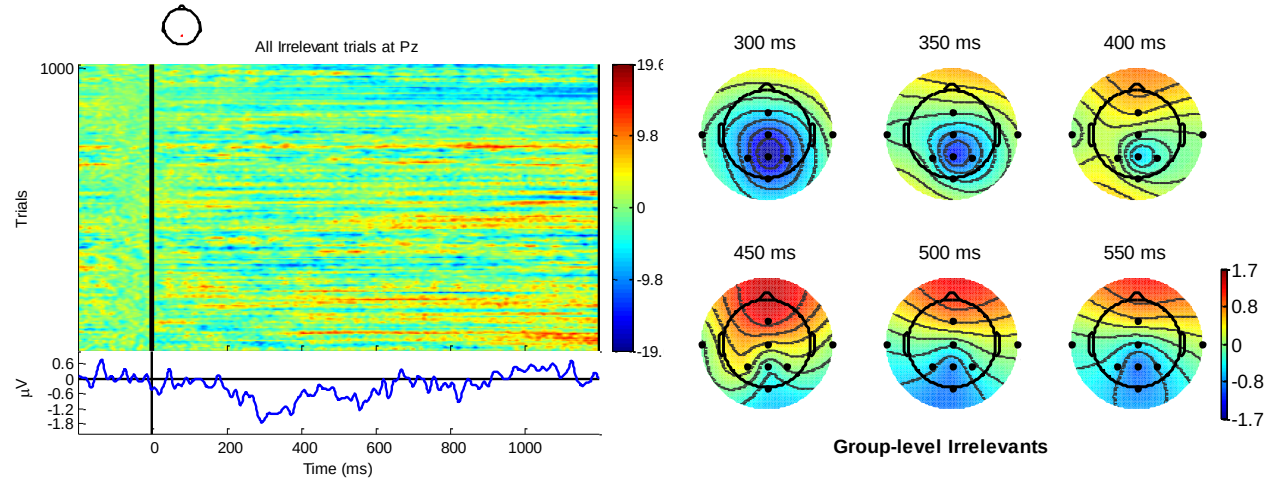

Figure 5.5 – Group-level view of all (1017) Irrelevant trials, over time at Pz (left side), and the interpolated scalp map representation of the ERPs (right plot, which must be treated with caution, due to the small number of electrodes), showing a distinctly different pattern when compared to the Probe (albeit, the small negativity, peaking at 300ms, was not expected). Note that the scalp map scale ranges from -1.7 to +1.7 μV, which is lower than the scale for Probe (see Figure 5.4).

As with the previous experiment, the main comparison was between the Probe and Irrelevant conditions, and our statistical tests showed a highly significant difference between them. Having aggregated all Probe and Irrelevant trials for all subjects, we employed the AGAT method, for orthogonal contrast time window placement (i.e. to independently find the most extreme 100ms mean amplitude interval) for highest positivity (P600f) components. Although the lowest negativity (N400f) was also analysed (i.e. to make comparisons with the previous experiment), the non-typical negativity in the Irrelevant condition – peaking at approximately 300ms (see Figure 5.6, below), and overlapping the Probe – meant that the results of our statistical tests on N400f were not significant (see section 5.4.2).

The null hypothesis (H0) was that there is no difference between the Probe and Irrelevant patterns, for the group. Our experimental hypothesis is that H0 can be rejected, at the group-level. As detailed in section 3.3.3.2 *(Group-level (AGAT) window placement*), a paired *t*-test of the mean amplitudes of Probe P600f/N400f and Irrelevant P600f/N400f was used, across all participants, to calculate the group's *p*-values (compared to a critical alpha level of 0.05), and to determine the probability that the observed pattern could have arisen if the null hypothesis were true. This is a reliable way to determine the group's familiarity with the Probe (lecturer) faces. Note that having justified the use of an independent detrending techniques (see section 4.4.2.3 – *Application of Detrending*), all the following statistical analyses will incorporate this method of removing any unwanted drift in the signal. Furthermore, detrending will always take place before baseline correction.

5.4.2 – Group-level Analysis, at Pz

For group-level analysis, the AGAT orthogonal contrast method enabled us to perform statistical tests, using a critical-alpha level of 0.05. As explained earlier (see section 3.3.3.2 (*Group-level (AGAT) window placement*), aggregating all Probe and

Irrelevant trials for all subjects, and employing the AGAT method, for orthogonal contrast time window placement, would enable us to independently identify the highest 100ms mean amplitude interval for the highest positivity (P600f).

Within the a-priori P600f time-frame (i.e. 300ms to 900ms), the AGAT method independently identified an orthogonal contrast 100ms time window, at 453 to 553 (see Figure 5.6), and our statistical tests produced a highly significant difference between the Probe (M = 2.0086, SD = 2.0809) and Irrelevant (M = -0.4146, SD = 1.3911), at the Pz electrode site (M = 2.4232, SD = 2.7311), $t(13) = 3.3198$, $p = 0.0055$, $d' = 1.3691$.

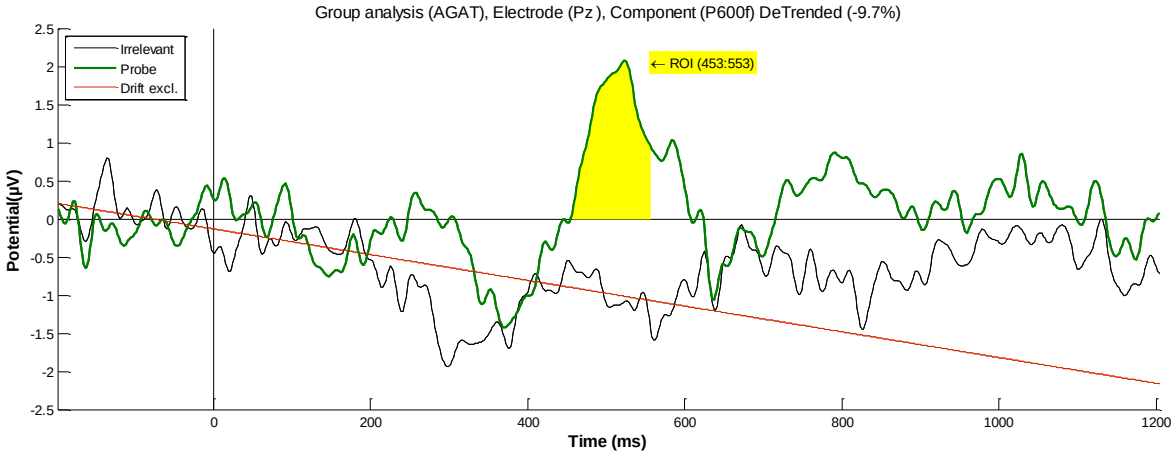

*Figure 5.6 – Grand average ERPs elicited by Probe and Irrelevant (i.e. the main comparison conditions) at **Pz**, showing an oscillatory pattern for the Probe condition (in green), which does not exist for the Irrelevant condition (in black). The linear Drift has been excluded (in red, at -9.7% to the vertical), with a detrending method. Even though subjects were not informed of the presence of the Probe (familiar lecturer face), statistical tests show a highly significant difference between Probe and Irrelevant, for **P600f** ($t(13) = 3.3198$, $p = 0.0055$, $d' = 1.3691$). However, the same statistical test on N400f was not significant ($t(13) = -1.9035$, $p = 0.0794$, $d' = 0.4532$).*

Unlike the previous (celebrity faces) experiment, the same statistical tests on the N400f component did not result in a significant difference between the two conditions of comparison: (M = -0.57, SD = 1.1205), $t(13) = -1.9035$, $p = 0.0794$, $d' = 0.4532$. Furthermore, subject-level statistical tests (i.e. Monte Carlo permutation) on the N400f component confirmed that none of 14 subjects (0%) showed critical-significance (0.05 alpha level) between Probe and Irrelevant (Mean p-value = 0.5744, SD = 0.2634). As can be seen in Figure 5.6 (above), the unusual negativity in the Irrelevant condition (peaking at approximately 300ms, and overlapping the Probe condition's N400f

component) meant that the results of our statistical tests on N400f were not significant. As a result, we chose to only focus on the analysis of the P600f component.

5.4.3 – Subject-level Analysis

We have established that our goal was to statistically analyse the data at the Pz electrode site (Kaufmann, Schulz, Grünzinger, & Kübler, 2011), and that the main comparison was between the Probe and Irrelevant conditions (Bowman, et al., 2013), at individual/subject level. Thus, we performed statistical analyses of the ERP data, to determine whether the elicited response by the Probe (i.e. familiar lecturer face) was significantly different from that elicited by the Irrelevant (i.e. unknown lecturer face), on a subject-level basis. As outlined in section 3.3.3 (*Time Domain (ERP) Analysis*), subject-level analysis is based on analysing each experimental participant separately, to determine whether there was a significant difference for that subject alone – did the subject's brain response reveal a differential perception and processing of the familiar lecturer faces, as compared to the unknown lecturer faces? The null hypothesis (H0) was that there is no difference between the subject's Probe and Irrelevant patterns. Our experimental hypothesis is that H0 can be rejected.

Having used the aERPt method to independently find the time window for each component of interest, a randomisation (i.e. Monte Carlo Permutation) test was used to define a *p*-value for each subject. Then, a null hypothesis distribution was generated in order to calculate the individual's *p*-value. This *p*-value would determine the probability that the observed pattern could have arisen if the null hypothesis were true. This is a reliable way to assess each subject's pattern individually, and to determine that subject's familiarity with the Probe. As with the first (celebrity faces) experiment, whether the subject reported to have seen the Probes in the relevant block of the experiment or not, we theorised that the results of our statistical analysis would infer their conscious and/or unconscious (i.e. sub/liminal) detection of familiar lecturer faces – in a Concealed Information Test, this could infer the guilt of the subject and/or knowledge of a compatriot.

5.4.3.1 – Synopsis of results

As shown in table 5.2, subject-level statistical tests of Pz electrode's P600f component, resulted in 8 of 14 subjects (57.1%) achieving critical-significance (at alpha level $p < 0.05$, shown in green), between the Probe and Irrelevant conditions. Note that this is an improvement on the previous (celebrity faces) experiment's results, by one subject, since that experiment's subject-level analysis, for P600f, achieved 50% significance (7 of 14 subjects).

Subject	Probe (*M*)	Irrelevant (*M*)	*p*-value
1	-0.8440	-2.4415	0.1660
2	3.4664	0.1121	0.0360
3	0.7447	1.5237	0.6510
4	2.0746	1.7694	0.4350
5	0.5265	-1.9373	0.0480
6	1.9794	0.7962	0.3280
7	4.2431	-0.1308	0.0020
8	-1.1190	0.5188	0.8580
9	1.4428	-1.2087	0.0080
10	2.0194	-1.3143	0.0270
11	4.6768	-1.5233	< 0.0001
12	0.1841	0.5206	0.6010
13	2.6896	-0.1072	0.0340
14	6.0366	-2.3814	< 0.0001

*Table 5.2 – Subject-level analysis, at **Pz** electrode, for the P600f component, showing the mean amplitude values of the Probe and Irrelevant conditions, from the same 100ms time window, which was independently found using the aERPt method. Note that critical-significance (at alpha level p<0.05) is shown in green. Out of 14 subjects, eight (57.1%) achieved critical-significance.*

5.4.3.2 – Individual's P600f, by-item and by-subject

At the Pz electrode site, we began by exploring the presence of the P600f component, within each of the three items of every subject (i.e. 3 experimental blocks for 14 subjects, equalling 42 item-blocks). Having independently searched for the P600f component's 100ms aERPt time window (i.e. highest positive deflection, within the a-priori search area that spans from the time range of 300ms to 900ms), we performed permutation tests for each individual block (see Appendix B.2 for full details). Consequently, three 'by-item' *p*-values were obtained for each subject (i.e. one for each block's familiar lecturer), resulting in significant difference between the Probe and Irrelevant conditions for 11 of 42 blocks (26.2%), which is higher than the previous (celebrity faces) experiment (i.e. the first experiment's *by-item* subject analysis, for P600f, which achieved 15.7% significance: 11 of 70 blocks). Note that part of this improvement may be attributed to the increase in the number of per-item trials (i.e. there were only 15-trials per block, in the first experiment, versus, 25-trials per block, in the current experiment), which can improve the Signal to Noise Ratio (SNR).

Next, by combining subject's trials (i.e. up to 75 trials per condition, to improve SNR), we were able to perform statistical tests on each subject (see the results in Table 5.2, above), resulting in a significant difference between the Probe and Irrelevant conditions for 8 of 14 subjects (57.1%). As explained above, the P600f significance for this experiment was higher than the previous (celebrity faces) experiment, however, we could not use the N400f component (and the Fisher combining method) to enhance the significance, at the subject-level.

As shown in Figure 5.7 (below), the majority of subjects' Probes elicited a clear positive deflection, within the aERPt identified highest positive 100ms time window (highlighted in yellow), of the P600f time-frame (300 to 900ms). However, relative to the Irrelevant (i.e. the condition of comparison), the Probe for five subjects (nos. 1, 3, 4, 6 and 12) failed to show a significant positivity for P600f, and one subject (no. 8) failed because the Probe ERP was consistently below the Irrelevant ERP – an uncharacteristic reversal, which may be related to the reversal of polarities (e.g. population of neurons pointing in the opposite direction).

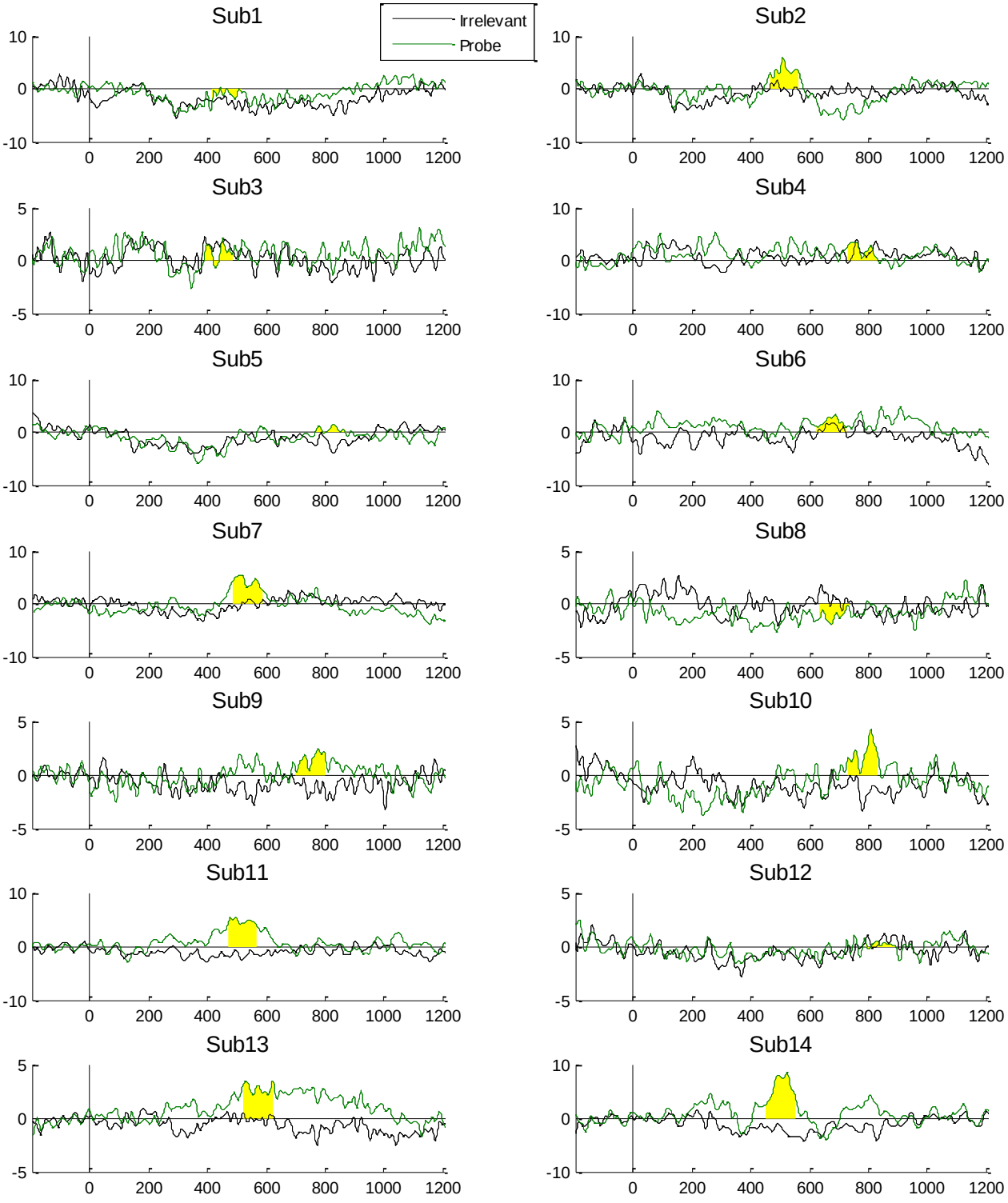

*Figure 5.7 – Subject-level Probe (in green) and Irrelevant (in black) ERPs, at the **Pz** electrode site (x-axis represents Time in miliseconds, and y-axis represents Potential in microvolts). Each ERP shows the orthogonally identified highest positive 100ms time window (yellow highlight) for **P600f** (using the aERPt method), where 8 of 14 subjects (57.1%) show a significant difference between the Probe and Irrelevant conditions.*

5.4.4 – Time Frequency Analysis (TFA)

As outlined in section 3.3.5 (*Frequency Domain Analysis*), and as used in the previous experiment, to analyse the power and coherence of the EEG data, we have employed two Time Frequency transforms: Even-Related Spectral Perturbation (ERSP) and Inter-Trial Coherence (ITC), using EEGLAB's toolbox (Delorme & Makeig, 2004). Whereas ERSP reflects the extent to which the signal power changes in relation to a specific time point (i.e. the baselining window before stimulus-onset) at different frequencies in a signal, ITC reflects the phase consistency (or synchronisation) between the trials, at every time point and frequency range. ERSP/ITC changes in coherence enables us to measure and assess the multi-cycle oscillations that we had observed in the ERPs.

Just as we had done in the previous experiment, we applied a notch filter, between 7 and 9 Hz, during the initial processing/epoching of the EEG data, in order to filter out any Steady State Visual Evoked Potential (SSVEP); as explained in section 3.3.2, the experiment's RSVP presentation rate (i.e. 133ms SOA) would evoke an SSVEP, at a frequency of approximately 8 Hz. Thus, as long as there are no significant power increases at higher frequencies, we could fix the upper boundary of our analysis at 7 Hz (also noting that the lowest boundary is fixed by our standard high-pass filter, on 0.5 Hz). Even so, in addition to the fixed-boundary analysis window (0.5 to 7 Hz), we also performed the full ERSP/ITC analyses on the full frequency range (0.5 to 45 Hz), to assess the power/coherence changes at higher frequencies.

5.4.4.1 – Group-level TFA

As outlined in section 3.3.5.1 (*Time Frequency Window Placement*), the group-level critical time window, for measuring ERSP/ITC, was placed based on the AGAT of power/coherence. As seen in Figure 5.8 (below), ERSP and ITC results of the AGAT of the Probe and Irrelevant conditions are combined together, across all 14 subject at Pz, with a large power increase around 300 to 650ms time-window (post-stimulus), mainly, at the low frequency range.

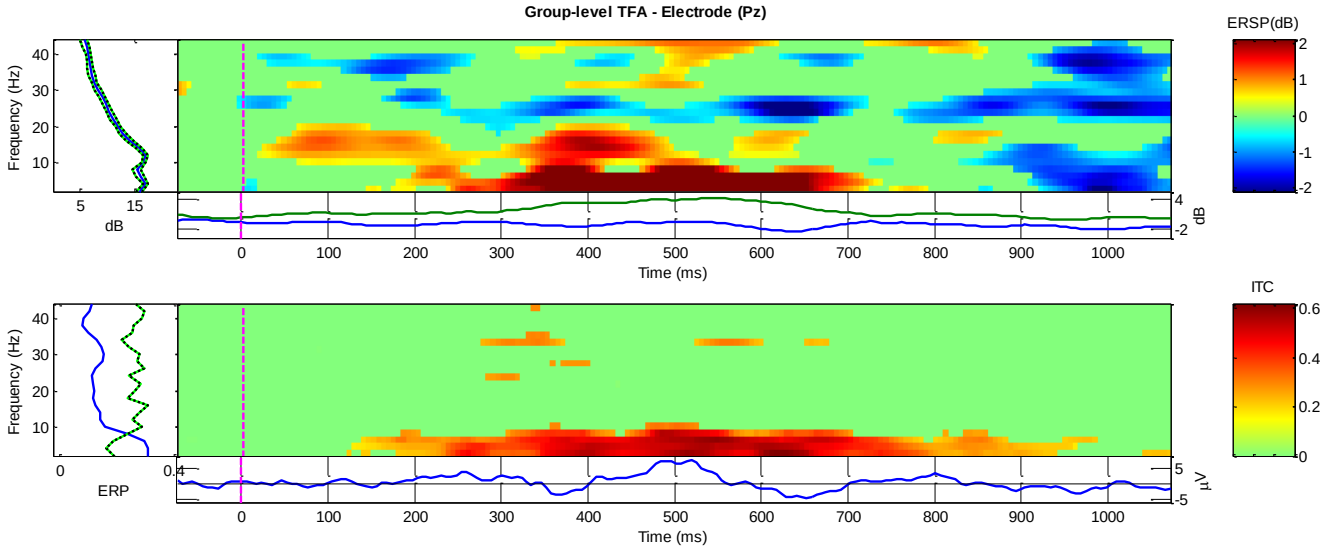

Figure 5.8 – Group-level Time Frequency plots, at the Pz electrode, using the combined Probe and Irrelevant conditions. The top plot relates to ERSP, and the bottom plot relates to ITC. The independent window selection (using AGAT method) for ERSP produced a Region of Interest (ROI) at 461:561ms, and earlier ROI for ITC, at 334:434ms. Increases in power/choherence have been mostly concentrated in the 0.5 to 10 Hz frequency range, and are strongest in the ROI time frames (SSVEP has been filtered out, by applying a 7:9 Hz notch filter).

As explained in section 3.3.5.1 (*Time Frequency Statistical Test)*, ERSP/ITC statistical tests were performed to compare the power and coherence changes between the two critical conditions: Probe and Irrelevant. To compare these conditions at the group-level, two measures were obtained for each subject, and a two-tailed paired *t*-test was used to calculate the significance for ERSP and ITC. We performed independently measured statistical analyses, by obtaining an orthogonal contrast time window, using the group-level AGAT method. As can be seen in the grand-Probe versus grand-Irrelevant ERSP/ITC comparisons (see figure 5.9, below), increases in power/coherence are predominantly evident in the grand-Probe condition, which suggests detection of the familiar lecturer face (ERSP > 4dB, and ITC > 0.4). However, the grand-Irrelevant condition lacks any significant power/coherence fluctuations, within the same time window, which implies little-to-no conscious or sub/liminal detection of the unknown lecturer face. Furthermore, nearly all the power/coherence fluctuations are occurring in the lower bandwidth (i.e. 0 to 7Hz).

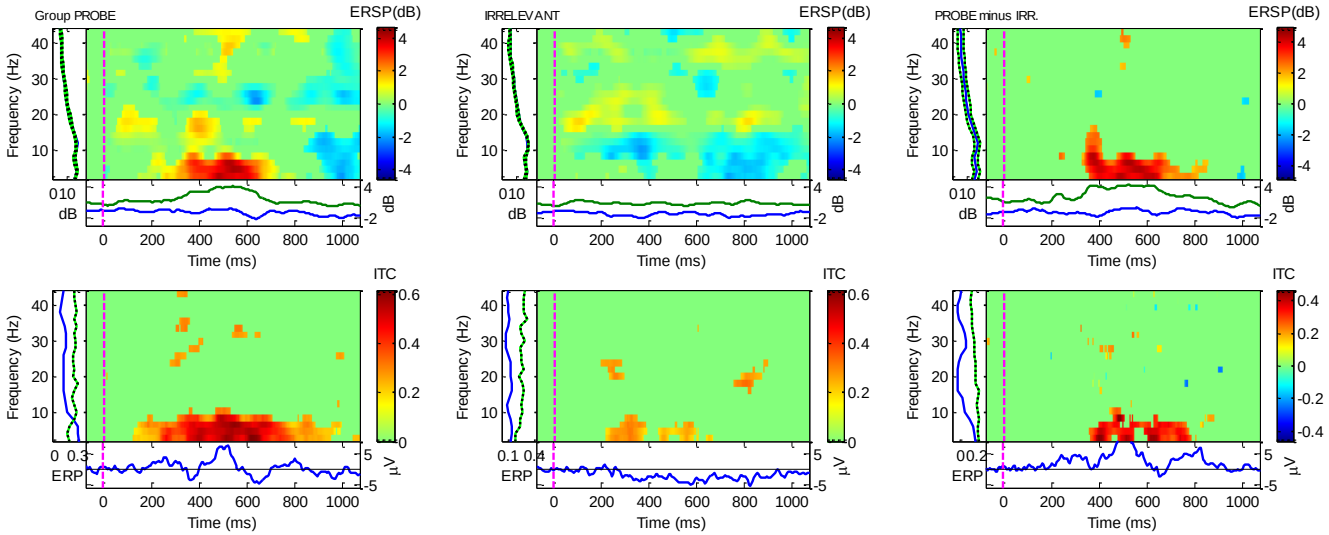

*Figure 5.9 – Group-level Time Frequency Analysis, at Pz electrode, for the difference between critical stimuli (Probe and Irrelevant), across the full frequency range (**0.5 to 45 Hz**)), showing ERSP (top row) and ITC (bottom row). The first column of ERSP/ITC plots show the power/coherence changes in the grand-Probe condition, and the second column shows the same for the grand-Irrelevant condition. The third column is the difference between grand-Probe and grand-Irrelevant (i.e. Probe minus Irrelevant), which confirms group-level increases in power/coherence for the grand-Probe only: ERSP (t(13) = 2.9737, p = 0.0108, d = 1.0417), and ITC (t(13) = 1.8064, p = 0.0941, d = 0.7997). Note that at each frequency and time point, increases in power/coherence are in red; decreases in blue, and green indicates no significant change.*

Over the full frequency-range (i.e. 0.5 to 45 Hz), the group-level analysis at Pz electrode for ERSP revealed a significant result (*alpha 0.05*), at the AGAT defined window 461 to 561ms (see Figure 5.9), confirming a difference between the Probe and Irrelevant conditions ($t(13) = 2.9737$, $p = 0.0108$, $d = 1.0417$). For the group-level ITC over the same/maximum frequency range, our statistical tests confirmed no significance, at the AGAT defined window 334 to 434ms: ($t(13) = 1.8064$, $p = 0.0941$, $d = 0.7997$).

However, focusing on the narrower frequency-band (i.e. 0.5 to 7 Hz), the group-level analysis at the Pz electrode for ERSP revealed a highly significant result (*alpha 0.05*), at the AGAT defined window 484 to 584ms (see Figure 5.10, below), confirming a difference between Probe and Irrelevant conditions ($t(13) = 3.4769$, $p = 0.0041$, $d = 1.2649$). For group-level ITC over the same (narrower) frequency range, our statistical tests also confirmed a significant result (*alpha 0.05*) at the AGAT defined window 588 to 688: ($t(13) = 2.322$, $p = 0.0371$, $d = 0.7442$).

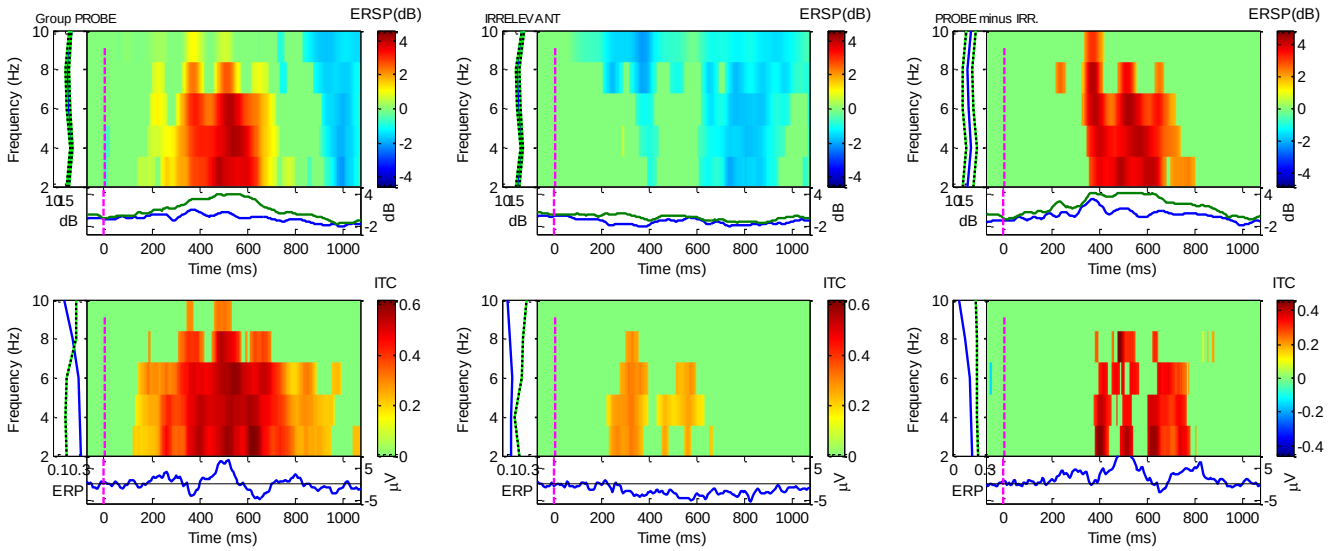

*Figure 5.10 – Group-level Time Frequency Analysis, at Pz electrode, for the difference between critical stimuli (Probe and Irrelevant), at the narrower frequency band (**0.5 to 7 Hz**), showing ERSP (top row) and ITC (bottom row). The first column of ERSP/ITC plots show the power/coherence changes in the grand-Probe condition, and the second column shows the same for the grand-Irrelevant condition. The third column is the difference between grand-Probe and grand-Irrelevant (i.e. Probe minus Irrelevant), which confirms group-level increases in power/coherence for the grand-Probe only: ERSP (t(13) = 3.4769, p = 0.0041, d = 1.2649), and ITC (t(13) = 2.322, p = 0.0371, d = 0.7442). Note that at each frequency and time point, increases in power/coherence are in red; decreases in blue, and green indicates no significant change.*

5.4.4.2 – Subject-level TFA

Per subject statistical analysis – in the form of a randomisation test on the combined Probe and Irrelevant conditions – confirmed the high significance of the increase in the Probe's power (ERSP) and coherence (ITC), as compared to the Irrelevant. Statistical tests of the *narrower* frequency band (0.5 to 7 Hz), resulted in two independently measured time windows (ERSP average window: 496 to 569, and ITC average window: 413 to 513) and *p*-values that revealed a difference between the Probe and Irrelevant conditions (see table 5.3, below). For ERSP, 10 out of 14 subjects' *p*-values (71.4%) were significant. Similarly, for ITC, 9 out of 14 subjects' *p*-values (64.3%) were significant, confirming the difference between the Probe and Irrelevant conditions.

Subject	ERSP *p*-values	aERPt win.	ITC *p*-values	ITC win.
1	0.6733	703	0.0107	369
2	<0.0001	508	<0.0001	525
3	0.04	461	0.2493	162
4	<0.0001	404	0.22	709
5	0.008	410	0.0147	520
6	0.0147	531	0.6213	260
7	<0.0001	514	0.0293	428
8	0.2147	795	0.244	53
9	0.2907	473	0.0187	674
10	0.264	307	0.043	629
11	<0.0001	490	<0.0001	484
12	0.0253	559	0.284	53
13	0.0133	455	0.04	473
14	<0.0001	334	<0.0001	449

*Table 5.3 – Subject-level Time Frequency analysis of power (ERSP) and coherence (ITC), at **Pz electrode**, using the narrower frequency range (**0.5 to 7 Hz**). For each subject, an orthogonal contrast time window was employed (using the aERPt method), and p-values were obtained for ERSP and ITC, by comparing the Probe and Irrelevant conditions, using a randomisation statistical test. At an alpha level 0.05, 10 of 14 ERSP p-values (71.4%) were significant, and 9 of 14 ITC p-values (64.3%) were significant.*

According to the above Frequency domain analysis (see Table 5.3), with the exception of subject 1, statistical test results of ITC appears to be closely correlated to the Time domain's statistical tests of the ERP data (see Table 5.2), at subject-level, since both analyses produce significant *p*-values for subjects 2, 5, 7, 9, 10, 11, 13 and 14.

Finally, even though we have justified the reason why the upper boundary of our analysis was fixed at 7 Hz (i.e. due to SSVEP waveform, which required a notch-filter on 7 to 9 Hz), we confirmed that per-subject statistical analysis of the maximum frequency range (0.5 to 45 Hz), resulted in *p*-values that revealed a difference between the Probe and Irrelevant conditions (see Table 5.4, below). For ERSP, 7 out of 14 subjects' *p*-values (50%) were significant, at 0.05 alpha level. As for ITC, 9 out of 14 subjects' *p*-values (64.3%) were significant.

Subject	ERSP *p*-values	aERPt win.	ITC *p*-values	ITC win.
1	0.35	455	0.002	357
2	0.002	484	0.001	381
3	0.056	738	0.093	191
4	0.004	422	0.04	53
5	0.006	307	0.136	410
6	0.038	53	0.025	795
7	0.001	525	0.012	455
8	0.157	756	0.166	490
9	0.391	59	0.002	635
10	0.183	600	0.002	53
11	0.105	289	0.092	145
12	0.06	600	0.332	53
13	0.043	53	0.05	496
14	< 0.001	461	< 0.001	334

*Table 5.4 – Subject-level Time Frequency analysis of power (ERSP) and coherence (ITC), at **Pz electrode**, using the maximal frequency range (**0.5 to 45 Hz**). For each subject, an orthogonal contrast time window was employed (using the aERPt method), and p-values were obtained for ERSP and ITC, by comparing the Probe and Irrelevant conditions, using a randomisation statistical test. At an alpha level 0.05, 7 of 14 ERSP p-values (50%) were significant, and 9 of 14 ITC p-values (64.3%) were significant.*

5.4.5 – Other midline electrode sites

All the above Time and Frequency domain analyses focused on the Pz electrode, but we were also interested in the other two midline electrodes (Cz and Fz), to confirm that, in-line with (Kaufmann, Schulz, Grünzinger, & Kübler, 2011), the strongest brain responses to familiar faces are recorded at Pz. The following analogous Time domain analyses of Fz and Cz, aim to find out if the P600f evoked by the Probe was significantly different from that evoked by the Irrelevant.

5.4.5.1 – Fz electrode

At the group-level, the grand-average ERPs of the two critical stimuli (i.e. the Probe and Irrelevant conditions), at the Fz electrode site, revealed a clear difference

between the conditions (see Figure 5.11, below). The Irrelevant condition, which consisted of an unknown lecturer face (paired with the Probe, and repeated randomly, as many times as the Probe), did not present any feature/pattern of interest. As explained earlier, this was as expected, since non-salient information is unlikely to breakthrough into conscious awareness, due to the high presentation rate of the RSVP streams. However, the Probe condition elicited a continuous oscillatory pattern, within a 200 to 620ms time frame. This waveform at Fz is similar to the oscillatory waveform at Pz (see Figure 5.6, above), and it confirms the prediction of a large difference between the Probe (familiar lecturer face) and the Irrelevant (unknown lecturer face) conditions, at all midline electrodes. At this Fz electrode site, an orthogonal contrast time window, for the highest positive (P600f) component, was independently found (using the AGAT method), at 434 to 533ms.

Statistical analyses at Fz – in the form of a paired *t*-test of the mean amplitudes of Probe and Irrelevant, across all participants – was employed to find the group level significance of the P600f component. Our statistical tests produced a significant difference between the Probe ($M = 2.0696$, $SD = 2.4934$) and Irrelevant ($M = -0.0349$, $SD = 1.7156$), at Fz electrode site: ($M = 2.1045$, $SD = 2.3754$), $t(13) = 3.3151$, $p = 0.0056$, $d' = 0.9834$.

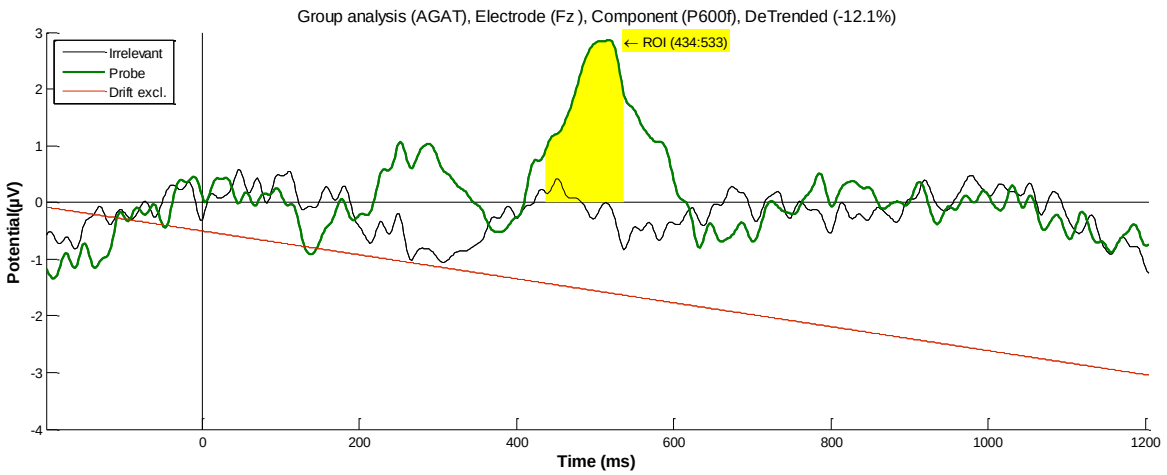

*Figure 5.11 – Grand average ERPs elicited by the Probe and Irrelevant, at the **Fz** electrode, showing an oscillatory pattern for the Probe condition (in green), which does not exist for the Irrelevant condition (in black). The linear Drift has been excluded (in red, at -12.1% to the vertical) with a detrending method. Even though subjects were not informed of the presence of the Probe (familiar lecturer face), statistical tests show a significant difference between Probe and Irrelevant, for **P600f** (t(13) = 3.3151, p = 0.0056, d' = 0.9834).*

At the Fz electrode, subject-level statistical tests (i.e. Monte Carlo permutation) on the P600f component confirmed that 6 of 14 subjects (42.9%) showed critical-significance (0.05 alpha level) between Probe and Irrelevant. The following table summarises our subject-level results (see Table 5.5), confirming that, in terms of the number subjects achieving significance, results at Pz (i.e. 8 of 14; 57.1% – see Table 5.2), beat Fz, agreeing with studies (Kaufmann, Schulz, Grünzinger, & Kübler, 2011) that report stronger brain responses (to familiar faces) at Pz.

Subject	Probe (*M*)	Irrelevant (*M*)	*p*-value
1	-2.2370	-1.9522	0.5900
2	4.5957	-1.0226	0.0070
3	2.7125	2.4847	0.4630
4	5.8049	1.4776	0.0120
5	3.9095	0.2096	0.0200
6	2.1781	1.2152	0.3290
7	6.2001	-0.9125	< 0.0001
8	-0.8330	1.4425	0.9090
9	0.9747	-0.8974	0.0790
10	1.3183	0.0575	0.2880
11	4.3378	0.7975	0.0160
12	0.6016	0.6258	0.5330
13	5.5821	1.8344	0.0160
14	1.8365	-0.8767	0.0610

*Table 5.5 – Subject-level analysis, at **Fz** electrode, for the P600f component, showing the mean amplitude values of the Probe and Irrelevant conditions, from the same 100ms time window, which was independently found using the aERPt method. Statistical tests on P600f resulted in 6 of 14 subjects (42.9%) being significant, which is not as high as equivalent results at Pz (i.e. 8 of 14: 57.1%), confirming that Fz failed to show a stronger brain response when compared to equivalent results at Pz.*

5.4.5.2 – Cz electrode

The same group-level analysis that was carried out at Fz (see section 5.4.5.1, above), was performed at Cz, revealing a difference between the Probe and Irrelevant

conditions (see Figure 5.12, below). Once again, the Irrelevant condition did not present a similar pattern to the Probe (or the Target), and the Probe condition elicited a continuous oscillatory pattern, within a 220 to 720ms time frame. This waveform at Cz is very similar to the oscillatory waveforms at Pz and Fz (see Figures 5.6 and 5.11, respectively), and it confirms the prediction of a large difference between the Probe (familiar lecturer face) and the Irrelevant (unknown lecturer face) conditions, at all midline electrodes. At this Cz electrode site, an orthogonal contrast time window, for the highest positive (P600f) component, was independently found (using the AGAT method), at 438 to 537ms.

Statistical analyses at Cz, in the form of a paired *t*-test, were employed to find the group level significance of the P600f component. Our statistical tests produced a significant difference between the Probe ($M = 1.2271$, $SD = 2.0023$) and Irrelevant ($M = -0.4156$, $SD = 1.3367$), at Cz electrode site: ($M = 1.6427$, $SD = 2.1683$), $t(13) = 2.8346$, $p = 0.0141$, $d' = 0.96494$.

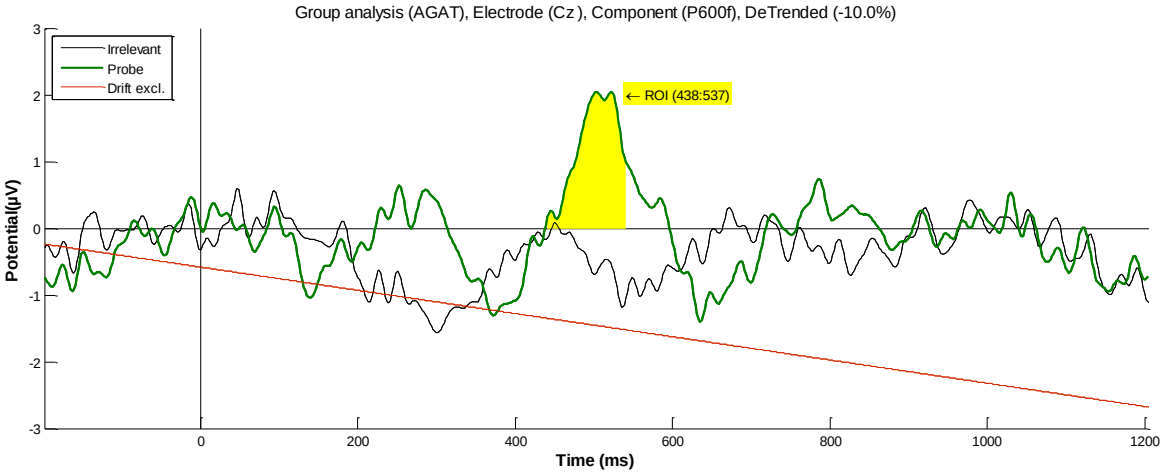

*Figure 5.12 – Grand average ERPs elicited by the Probe and Irrelevant, at the **Cz** electrode, showing an oscillatory pattern for the Probe condition (in green), which does not exist for the Irrelevant condition (in black). The linear Drift has been excluded (in red, at -10% to the vertical) with a detrending method. Even though subjects were not informed of the presence of the Probe (familiar lecturer face), statistical tests show a significant difference between Probe and Irrelevant, for **P600f** (t(13) = 2.8346, p = 0.0141, d' = 0.9649).*

At the Cz electrode, subject-level statistical tests (i.e. Monte Carlo permutation) on the P600f component confirmed that 4 of 14 subjects (28.6%) showed critical-

significance (0.05 alpha level) between Probe and Irrelevant. The following table summarises our subject-level results (see Table 5.6), confirming that, in terms of the number subjects achieving significance, results at Pz (i.e. 8 of 14; 57.1% – see Table 5.2) beat Cz; once again, agreeing with studies that report stronger brain responses (to familiar faces) at the Pz electrode site.

Subject	Probe (*M*)	Irrelevant (*M*)	*p*-value
1	-1.1798	-2.9437	0.1870
2	3.4261	-0.1361	0.0270
3	1.6282	1.6906	0.5040
4	0.6658	1.8322	0.7260
5	0.3553	-0.2447	0.3510
6	2.9873	0.8060	0.2100
7	4.4588	-0.2104	0.0010
8	-0.8165	1.1491	0.8790
9	0.9685	-1.0082	0.0620
10	1.2720	-0.2308	0.1990
11	4.2569	-0.0883	0.0020
12	0.4085	0.7474	0.6020
13	3.5615	0.7885	0.0610
14	3.4331	-1.8483	0.0020

*Table 5.6 – Subject-level analysis, at **Cz** electrode, for the P600f component, showing the mean amplitude values of the Probe and Irrelevant conditions, from the same 100ms time window, which was independently found using the aERPt method. Statistical tests on P600f resulted in 4 of 14 subjects (28.6%) being significant, which is not as high as equivalent results at Pz (i.e. 8 of 14: 57.1%).*

Finally, we have demonstrated that all three midline electrodes (Pz, Fz and Cz) have exhibited similar oscillatory waveforms, and that statistical tests showed significant difference between the two conditions, Probe and Irrelevant. Although our choice to focus on the Pz electrode was a priori (in-line with (Kaufmann, Schulz, Grünzinger, & Kübler, 2011)), we found that the strongest brain responses to familiar lecturer faces, was indeed recorded at the Pz electrode. Note that this finding is in accordance with the results of the first (celebrity faces) experiment.

5.5 Discussion

The primary aim of the second experiment was to investigate whether faces that are personally known to an individual (in the form of familiar lecturer faces) can breakthrough into conscious awareness (and that the breakthrough event can be detected by EEG, on a group and subject level basis), as successfully as the first experiment's famous celebrity faces. Once again, we intended to achieve this through statistical analyses of the ERP data (in the Time domain) and single-trial data (in the Frequency domain), to determine whether the evoked response by the Probe (familiar lecturer) faces were significantly different from that evoked by the Irrelevant (unknown lecturer) faces. The null hypothesis (H0) was that there is no difference between the Probe and Irrelevant conditions. Our experimental hypothesis is that there is a difference.

As the key comparison was between Probe faces and Irrelevant faces, a significant difference was observed between the ERPs, at all three mid-line electrodes (Pz, Fz and Cz). However, unlike the first experiment's prominent oscillatory negative-and-positive deflections for the Probe (i.e. N400f, followed by P600f), this experiment's negative deflection (N400f) was muted, but its enhanced positivity (P600f) was equivalent (over a similar time frame of 300ms to 600ms). Thus, even though there was evidence of a similar oscillatory pattern, it was acknowledged that the second experiment's ERPs are slightly different to the first experiment.

Having closely mirrored the first experiment's design and analysis standards, the single change to the current/second experiment's critical stimuli (i.e. replacement of the *celebrity* faces with *lecturer* faces) enabled us to make direct comparisons and draw conclusions, in light of evidence gathered. These findings were essential to bridge the gap between the previous/first experiment's proof of concept, and the next/third experiment's functional prototye, which could advance future applications of deception detection tests, using faces in RSVP-based EEG tests. Even though the first (celebrity faces) experiment established the viability of using faces to infer recognition in the RSVP paradigm, the next logical step was to investigate whether faces that are personally known to an individual would have a similar effect. After all, the application of deception detection tests are dependent on the personal relationship/familiarity of the subject under investigation, rather than the non-partisan knowledge of a famous person.

In keeping with the previous experiment, subjects were not informed that familiar faces may appear in the RSVP streams, and yet, our statistical tests confirmed the breakthrough events. Having been instructed to only look for the Target face, the inclusion of Probe faces (i.e. University lecturers who were personally familiar to the subject, but not associated with the explicit task) was meant to examine the subject's ability to perceive intrinsically salient faces. Statistically testing the brain responses by comparing Probes and Irrelevants, in the Time, as well as, the Frequency domains, enabled us to confirm the Probes' significance over the Irrelevants (i.e. unknown faces), at group and subject levels.

5.5.1 – Time Domain

At the Pz electrode, group-level analysis of ERPs confirmed the significance of the difference between the grand-Probe and grand-Irrelevant ($p = 0.0055$), and subject-level statistical analyses of ERPs confirmed that, having found the orthogonal contrast window for the P600f component, a total of 8 of 14 subjects (57.1%) had *p*-values below our critical-significance (alpha level 0.05), revealing a difference between the Probe and Irrelevant conditions. We noted that the subject results at Pz, for this experiment's P600f component was one subject more than the previous (celebrity faces) experiment, which achieved only 50% (7 of 14 subjects) significance.

The results of our statistical analyses, within the Time Domain, provide evidence that the personally familiar lecturer faces (Probe conditions) were differentially perceived and processed by most subjects' brains, as compared to the unknown lecturer faces (Irrelevant conditions). Even though both conditions were treated equally, our experimental findings show major differences between the Probe and Irrelevant, which was as a result of the former stimuli reaching conscious awareness and generating pronounced electrical responses (as seen in the Probe ERPs), whilst the latter was not sufficiently perceived to encode into working memory, in order to generate a distinct electrical response that resembled the Probe (or, for that matter, the Target). And yet, there was an interesting new electrical response (i.e. Irrelevant's negative deflection, peaking at 300ms), which may reflect subliminal awareness of repetition.

Finally, just as we had done in the first experiment, in the second experiment, we confirmed that the brain response to the Probe, at group-level ($p = 0.0055$) and at subject level (8 of 14), achieved more subjects at the Pz electrode site, when compared to Fz ($p = 0.0056$, and 6 of 14, respectively) and Cz ($p = 0.0141$, and 4 of 14, respectively).

5.5.2 – Frequency Domain

At the Pz electrode, group-level analysis of Time Frequency, across the narrower frequency band (0.5 to 7 Hz), confirmed the significance of the difference between the grand-Probe and grand-Irrelevant for ERSP ($p < 0.0041$) and for ITC ($p = 0.0371$). Subject-level statistical analyses of Time Frequency, across the same frequency band (0.5 to 7 Hz), using the independently measured time window for ERSP, confirmed that 10 out of 14 subjects' *p*-values (71.4%) were significant. As for ITC, subject-level statistical analyses of Time Frequency showed that 9 out of 14 subjects' *p*-values (64.3%) were significant, confirming the difference between the Probe and Irrelevant conditions.

The results of our statistical analyses, within the Frequency Domain, provide additional evidence that the Probe (familiar lecturer) faces were differentially perceived and processed by most subjects' brains, as compared to Irrelevant (unknown lecturer) faces. Albeit, the large increases in power (ERSP) and coherence (ITC), which were observed and statistically confirmed in the Probe condition only, were not as significant as the first (celebrity faces) experiment. Even so, they demonstrate that such changes in power and phase-locking coherence could have contributed to the generation of the P600f component, which was elicited within similar time windows of the same condition's Probe ERPs. Once again, this finding supports the hypothesis that oscillatory activity, in the frequency domain, is related to the ERP component, in the time domain (Makeig, Debener, Onton, & Delorme, 2004b) (Fuentemilla, 2008).

5.5.3 – Conclusion

This chapter's experimental findings confirm our first hypothesis that having substituted *lecturer* faces, who are personally known to the subjects, in place of the highly recognisable *celebrity* faces, we were able to detect the group-level breakthrough of familiar faces into consciousness. In a similar manner to the previous (celebrity faces) experiment, we agree that such breakthrough would be encoded in brain signals (Bowman, et al., 2013), and would generate ERP components/effects that would differ between the Probes (familiar lecturer faces) and the Irrelevants (unknown lecturer faces). Through the effective use of our statistical analyses, in the time domain (using ERPs), as well as, the frequency domain (using single-trials), we have successfully differentiated between the Probe and Irrelevant conditions, at the group-level, in all 3 mid-line electrodes (Pz, Cz and Fz).

Our second hypothesis was that in addition to the breakthrough of Probe faces at group-level, we can also use Time and Frequency domain analyses to detect the breakthrough events at subject-level. Whilst the Probe and Irrelevant conditions were treated equally in the experiment, we used each subject's ERPs (in the time domain, at Pz) to confirm the presence of large differences in brain responses between the conditions for 8 of 14 subjects. Furthermore, using the subject's single-trials (in the frequency domain, at Pz), we confirmed a difference between the Probe/Irrelevant conditions for 10 of 14 (for ERSP) and 9 of 14 (for ITC) subjects.

Whilst acknowledging that this experiment's subject-level statistical results are not as conclusive as the first (celebrity faces) experiment's statistical results, it must be noted that the first experiment's subjects would have been previously exposed to the images of the famous *celebrity* faces (i.e. published photographs, which were frequently in the public eye, and assuredly seen by all subjects, on many occasions and over a far longer period of time). Whereas, the second experiment's *lecturer* faces, whilst being personally known to the subjects, their photographs were seen for the first time, in the fast moving RSVP stream of images. In fact, the behavioural/recognition tests for both experiments support this premise, as subjects in the previous/first experiment reported seeing the emotive (celebrity) Probes with an average confidence rating of 3.4 out of 5 (60%), whereas, subjects in this/second experiment reported seeing the intrinsic (lecturer) Probes with an average confidence rating of only 2.4 out of 5 (33.9%). With

that in mind, we infer that this approach can still be used to determine whether a subject has high familiarity of real-life acquaintances.

Our third and final hypothesis was that the strongest brain responses to the familiar (Probe) faces are recorded at the Pz electrode site. Having carried out the same statistical tests on all mid-line elctrodes (Fz, Cz and Pz), we can confirm that whilst all three sites exhibited similar oscillatory waveforms for the Probe, the strongest brain responses to familiar faces was, indeed, recorded at the Pz electrode, which is in-line with our previous (celebrity faces) experiment and similar studies (Kaufmann, Schulz, Grünzinger, & Kübler, 2011).

5.5.4 – Future work

This chapter's experiment extended our earlier work, which, together, demonstrated that both highly evocative faces (i.e. first experiment's *celebrities*) and personally familiar faces (i.e. this experiment's *lecturers*) can be employed in RSVP-based fringe-P3 studies, and that highly familiar faces can breakthrough into conscious awareness, on an individual (subject-level) basis. The results suggest that we can apply our findings to the differentiation of deceivers and non-deceivers, in the application of crime compatriots, whereby, a suspect's familiarity with a criminal/terrorist can be established using faces.

However, taking into account the muted effect of the Probe oscillations, in the current (lecturer faces) experiment, when compared to the first (celebrity) faces experiment, we acknowledged that improvements in the signal-to-noise ratio would be instrumental in achieving significant findings in our next-and-final experiment. Therefore, before planning the next experiment (see chapter 7), we hypothesised that we can introduce changes to the next study, in order to produce improvements in statistical power (i.e. to reduce the risk of Type II errors), without the inflation of Type I errors (see chapter 6). With the aid of ground-truth data simulations, our aim was to improve the design of the study, as well as, the analysis methods, to make our research empirically relevant to the application of real-life deception detection of compatriots.

In addition to the benefits gained from the changes that will be justified in our methodological exploration (see chapter 6), the design of the next study (see Chapter 7) will take one step closer to the real-world scenario of concealed information tests, whereby, the hidden nature of the Probe critical stimuli will be revealed to the subjects, at the start of the experiment. Having demonstrated that the breakthrough of familiar faces can be detected even when subjects were not expecting to see celebrity-or-lecturer faces – because the presence of the Probe was concealed and the Target was the only task-relevant objective in the first two experiments – we aim to inform the subjects that, in addition to the Target (which will remain task-relevant), they may see a familiar lecturer face, in the RSVP streams. Naturally, we will not inform the subjects which lecture faces may appear in the experiment (or show them any photographs), and will only instruct them to focus on detecting the Target. This arrangement is closer to the real-life application of a deception detection test, in which the perpetrator is fully aware of the purpose of the experiment (i.e. to find out if s/he is familiar with an accomplice). As a result, the next/final experiment will take us one step closer to using faces in RSVP-based EEG tests for deception detection applications of compatriots.

Finally, in addition to improving the design and analysis methods of the next experiment, we aim to increase the potential for recording electrocortical processes, in our final study. Thus, instead of the current practice of applying 8 electrodes only – mainly because of precedent and our interest in mid-line electrodes only – we aim to increase our data capture points to 32 electrodes, in order to improve the estimated electric potential values used at scalp locations, between the actual recording sites. Whilst acknowledging that our next/final experiment may not take full advantage of all the data from the extra 24 electrodes (albeit, it will benefit the scalp topography of the ERPs, as shown in Figures 7.3, 7.4 and 7.5), the practice of capturing data from 32 electrodes will become a standard – along with detrending and AGAT/aERPt window placement – which will benefit all future experiments, leading to further opportunities, using new signal processing techniques (e.g. Independent Component Analysis).

Chapter 6:
Methodological Explorations to Improve Statistical Power

6.1 Introduction

In the first (celebrity faces) experiment, we established that famous faces can breakthrough into conscious awareness, using an RSVP subliminal search paradigm, and in the second (concealed lecturer faces) experiment, we demonstrated that we can substitute the highly evocative famous celebrity faces with familiar faces that are personally known to the participants (e.g. the University's lecturers). Furthermore, in both experiments, we showed that our statistical tests can differentiate between the Probe (familiar celebrity or lecturer) and Irrelevant (unknown) faces, at group and subject levels (see chapters 4 and 5). But, before we attempt to apply our findings to the next/final experiment, we decided to explore the potential and viability of using online statistical tests, during the experiment, to enable us to focus our data collection efforts on the critical stimulus that shows the highest significance.

In this chapter, we considered a new experimental design and technique that can rank the subject's familiarity with multiple critical stimuli (i.e. Probes), and provide online inference/feedback to inform the experiment, in real-time. As long as we can demonstrate that the proposed technique is not susceptible to fishing criticism, we theorise that we can improve the statistical power and signal-to-noise ratio (SNR) – *noting that, in this context, the SNR is considered to improve with the increase in the number of trials* – which would result in a higher subject-level significance. Even allowing for the well-known psychometric principle that an increase in the number of trials results in higher reliability, we propose that our two-part experimental design may result in a higher validity. To investigate the viability of different techniques, we propose the use of ground-truth data simulations, to evaluate the safety and benefits of different methods that can identify the salient Probe (i.e. the critical stimulus with the highest significance), in the middle of the experiment, and to re-use the independently identified stimulus in the rest of the experiment. This will improve statistical power (i.e. reduce the risk of Type II errors), without the inflation of Type I errors. This chapter's methodological explorations will inform the design of the next-and-final experiment

(see chapter 7), and the proposed new analysis method will advance future applications of deception detection of crime compatriots (i.e. to establish a suspect's familiarity with a criminal or terrorist, using faces).

In summary, notwithstanding the success of the previous two (*celebrity* and *lecturer* faces) experiments, we are constantly striving to improve the statistical power of detecting an effect and improving the SNR, whereby, improvements in the design could benefit our statistical tests, in the Time and Frequency domains. Furthermore, in addition to our ultimate goal of showing the significance of the difference between familiar faces (i.e. Probes) and unknown faces (i.e. Irrelevants), at subject-level, we predict further benefits in isolating each subject's significance down to a single Probe (i.e. *by-item*). To date, our RSVP-based experimental design was set-up to demonstrate the significance of a subject's familiarity with all presented Probes (e.g. five different Probes in the first experiment, or three different Probes in the second experiment). Thus, as formulated, it did not enable us to infer the significance of the subject's familiarity with an individual Probe (i.e. subject-level significance does not reflect *by-item* significance). Conversely, the lack of significance at subject-level, does not necessarily mean that there was no effect at an individual Probe (or *by-item*) level. Ultimately, the suitability of our new experimental design would depend on the exploration of the dangers of false-positive conclusions (i.e. rejection of a true null hypothesis, or Type I error), and the non-rejection of a false null hypothesis (i.e. type II error, or a false negative conclusion).

6.2 Exploration Hypotheses

The above scientific enquiries have incentivised us to consider alternative methods of designing the next experiment and performing statistical tests, to advance our deception detection techniques, and to make them more suitable for real-world applications. As a result, we are proposing a two-part experiment: Part I will be used to rank the subject's familiarity with three different Probes, and act as a feedback loop, so that Part II can focus on the Probe with the highest significance, only. With the aid of ground-truth data simulations, we intend to demonstrate that the improved new experimental design does not inflate type I errors, and reduces the risk of type II errors.

Because the previous (lecturer faces) experiment contained three blocks, we chose to retain the same number of blocks in Part I of the new experimental design, affording us options to make comparisons between the previous (concealed lecturer faces) experiment and the next (revealed lecturer faces) experiment. Thus, Part I of the latter will be compatible with the design of the former. However, the new experimental design will possess a Part II, which contains two further blocks, where both blocks will re-use the same stimulus that was shown to be statistically significant, from Part I. As a consequence, the following four different methods of statistically analysing the new two-part experiment are being considered. Note that the fourth (*Decider*) method, has been shown for reference purposes only, as *Decider* is the online technique that is used, after the completion of Part I, in order to find the most significant stimulus for Part II:

A) ***Abandoned*** method, where Part I is discarded/abandoned because it acts as the qualifier (using the *Decider* method) for Part II of the experiment. Therefore, the Abandon method contains the last two blocks of the experiment, from Part II only. We hypothesise that this method is safe, but due to having the least number of trials, the reduction in the statistical power may inflate Type II (false-negative) errors.

B) ***Biased*** method, where the chosen block in Part I (i.e. the one with the lowest *p*-value, whose Probe/Irrelevant pair will be re-used in Part II) is cherry-picked, to join with both blocks in Part II. Thus, Biased contains three blocks: one of the blocks from Part I, plus the fourth and fifth blocks, which comprises Part II. Despite an improved SNR (compared to method A), we hypothesise that this method is unsafe, due to the inflated possibility of Type I (false-positive) errors.

C) ***Combined*** method, where all five blocks (i.e. three in Part I and two in Part II) are joined together. We hypothesise that this method is safe and yields the highest SNR.

D) ***Decider*** method, where the three blocks in Part I are used to infer the paired Probe/Irrelevant conditions for Part II. As this method replicates the design of the previous (familiar lecturer faces) experiment, we hypothesise that it remains safe.

6.3 Methodological Explorations

6.3.1 – Overview

We began the exploration by demonstrating the ideal procedure for producing simulated/synthetic EEG data (see section 6.4), which included two key components: random noise and human EEG simulations, especially as, in classical terms, EEGs are considered to be the superposition of evoked signals onto background noise (i.e. human artifacts, task irrelevant brain activity and environmental noise). After generating a large number of EEG trials, in the usual way, we used an averaging process to increase the signal-to-noise ratio and reveal the evoked activity, in the form of Event Related Potentials (ERPs). Background noise was generated using the well-publicised Bogacz noise method (Yeung, Bogacz, Holroyd, & Cohen, 2004), which contained auto-correlation statistics consistent with real human EEG. This was used to scale the amplitudes of sinusoids, in generation of EEG noise, for two conditions (Probe and Irrelevant), where, under the null hypothesis, there is, in a statistical sense, no difference between the conditions (i.e. the null hypothesis is true).

In section 6.5 of this chapter, we tackle the probability of Type I errors (false positives), by ensuring that both Probe and Irrelevant conditions are the same, in a statistical sense. Through extensive iterations of simulated data, adjusted to represent similar conditions within blocks (i.e. Probe and Irrelevant conditions differed by the probability of random noise only), we were able to demonstrate rates of rejecting a true null hypothesis. This was achieved by exploring both scenarios in which all blocks used the same human-EEG signal (from our 2013 Names experiment: (Bowman, et al., 2013)), and where each block used one of three different subject-ERP signals (from our 2016 lecturer faces experiment: see chapter 5).

In section 6.6 of this chapter, we tackled the probability of Type II errors (false negatives), and focused on the new experimental design's probability of affirming a false null hypothesis. As we have noted earlier, our new experimental design allows for a two-part experiment, in which Part I (containing three blocks of unique Probe/Irrelevant conditions) determines the paired conditions that will be re-used in Part II (containing two more blocks of the same paired Probe/Irrelevant conditions that were discovered to be the most significant, in Part I). To increase the probability that the first

block in Part I achieves significance (in a statistical sense), we artificially increased the power/amplitude of its Probe condition, thus, promoting that block's paired Probe/Irrelevant conditions, which would be re-used in both blocks of Part II. Aside from this procedure, which identifies the most significant stimuli from Part I (i.e. the '*Decider*' method that analyses the first three blocks), this experimental design will afford us three unique methods, in which we can statistically test subjects' EEG data:

A) *Abandoned*, where we only use the data in Part II, by discarding Part I;

B) *Biased*, where we only use the significant block in Part I, plus all of Part II;

C) *Combined*, where we use all the data in Part I, plus everything in Part II.

To test our hypotheses, we intend to demonstrate, in this chapter, the statistical bias in method B (i.e. fishing for the best result of Part I's data), which increases the risk of a false-positive conclusion, whilst the redundancy in method A (i.e. abandoning Part I, and only using Part II's data) may increase the risk of a false-negative conclusion. Therefore, we predict the suitability of method C (i.e. combining all the data from Parts I and II), as the Goldilocks option, where we have safely increased the SNR and the statistical power, whilst limiting type I errors.

6.3.2 – Type I and II errors

In empirical research, our first objective is always to generate a hypothesis (based on a good research question), which may be tested critically by experimentation. As we cannot prove our hypothesis, our Null hypothesis may state that there is no difference between two conditions, thus, using statistical tests, falsification of this initial hypothesis enables us to reject the Null, in favour of the alternative hypothesis. Furthermore, using statistical tests, it is possible to estimate the probability that an observed difference between the two conditions could be due to chance alone. There are two possible error outcomes, known as Type I and Type II errors (see Figure 6.1). Type I error relates to rejecting the Null hypothesis when it is actually true (also known as *false positive*), and Type II error relates to failing to reject the Null when there was indeed a difference between the conditions (also known as *false negative*).

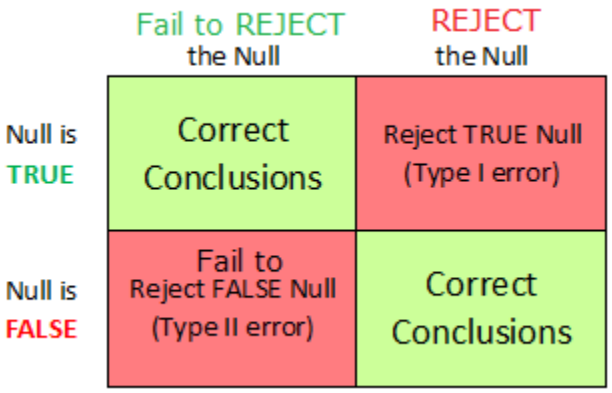

Figure 6.1 – The correct conclusion from a statistical analysis can be that the null hypothesis is true (i.e. fail to reject the null), or that it is false (i.e. reject the null), however, it is also possible to reject a true null hypothesis (i.e. Type I error), or fail to reject a false null hypothesis (i.e. Type II error).

In well-functioning classical statistical inference, the probability of committing type I error is determined by the alpha level (i.e. the level of statistical significance, or reasonable doubt). In all of our experiments, the a-priori alpha has been defined as 0.05, which means that if the test is sound, there is a 5% chance that the Null hypothesis is incorrectly rejected. The rest of this chapter will chronicle our investigations into false conclusions (i.e. detection of type I and II errors) that can be drawn from the results of our new experimental design and analysis methods.

6.4 Noise Generator

The generation of simulated (or synthetic) data is performed at the trial level, by generating noise according to the human frequency spectrum. Following the published works of the Bogacz group (Yeung, Bogacz, Holroyd, & Cohen, 2004), we generated simulated noise trials with the characteristics of the power spectrum of human EEG, and then scaled this noise with a constant (see Figure 6.2). The ideal scaling parameter for the random noise generator was chosen experimentally (see section 6.4.1), to ensure that the noise signal would not overwhelm the human EEG signal.

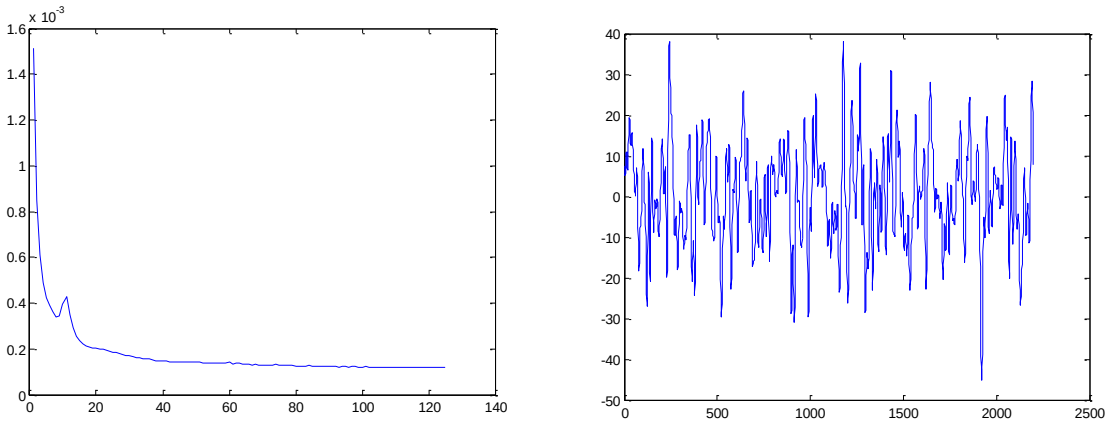

Figure 6.2 – Our noise generator combined signal and noise components, by loading Rafal Bogacz's "meanpower" (left plot), which uses an algorithm to scale its frequency range from 0 to 120 Hz, on the x-axis, along with a pre-defined amplitude-scaling parameter (as justified in section 6.4.1), to produce random noise with the power spectrum of human EEG. This method was used to scale the amplitudes of sinusoids, to generate a single EEG trial (right plot), contaning random noise.

Next, the resultant simulated noise trial was added to an actual human EEG signal (i.e. an a priori EEG signal that, in this case, was the grand-ERP of our 2013 Names experiment (Bowman, et al., 2013)), which contained (amongst others) a P300 component (see Figure 6.3). Finally, we baseline corrected the resultant waveform, by subtracting the mean of its first 100ms. Having generated the first condition (called Probe), we repeated the above process for the second condition (called Irrelevant).

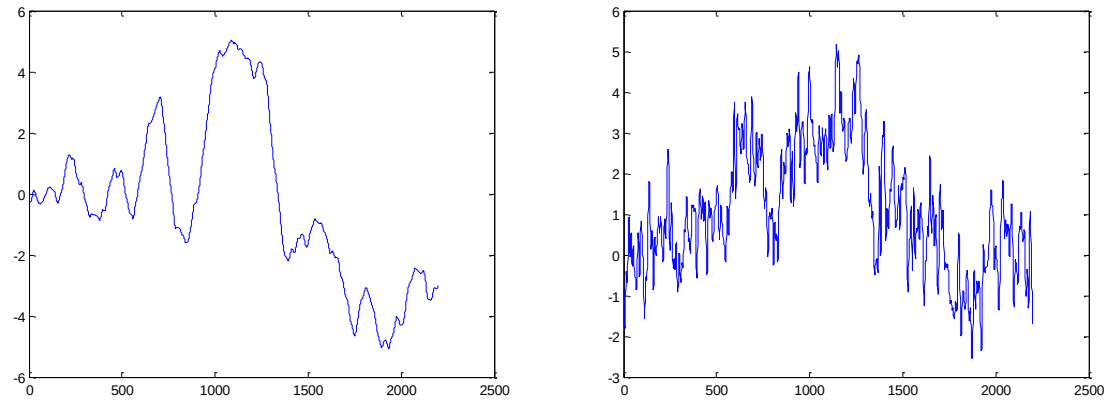

Figure 6.3 – Having loaded a human EEG signal (i.e. the grand-ERP from our 2013 Names experiment, left-plot), which contained (amongst others) a P300 component, we added the ERP signal to the simulated noise waveform (described in Figure 6.2). This formed a signal and noise condition that contained an artificial condition containing pure random noise, with the power spectrum of human EEG (right plot). Finally, we baseline corrected the resultant waveform, before repeating the process for subsequent trials.

6.4.1 – Ideal Noise amplitude scaling

To justify a standard amplitude scaling parameter (hereafter referred to as ASP) for random noise generation, we experimented with different settings, and achieved an intensity value that would neither overwhelm the human EEG-signal, nor have an unnoticeable impact. To that end, we performed 5,000 iterations of our new experiment, by setting ASP of random *Noise* at intervals of 10, 20, 30, 40 and 50, whilst increasing the size difference between the first Probe and its paired Irrelevant (i.e. using a *Multiplier*, to force a difference between the first block's Probe & Irrelevant conditions).

Note that the paired Probe/Irrelevant conditions for the remaining two blocks of Part I were the same (i.e. the only difference between the conditions in blocks 2 and 3 was random noise), and the most significant pair of conditions were re-used in Part II. Our hypothesis is that as the significance of the first block's difference between its Probe and Irrelevant conditions increases, the likelihood of inferring block-1's paired conditions for blocks 4 and 5 will also rise, leading to an increase in the number of significant *p*-values in Part II. With the aid of a surface plot, we demonstrated the distribution of the number of significant *p*-values, at different *Noise* ASP levels, versus *Multiplier* size-differences, across 5 blocks of the experiment (see Figure 6.4).

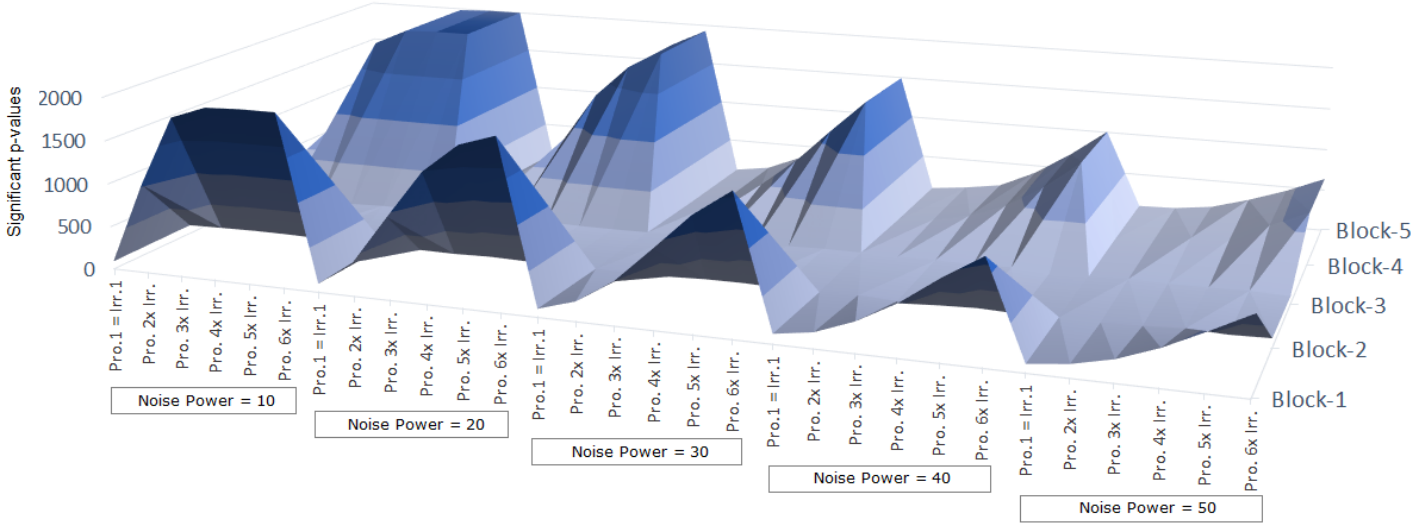

Figure 6.4 – Distribution of p-values for 5,000 iterations of our new two-part experiment, at 5 different Noise-Power ASP levels (10, 20, 30, 40 and 50), in relation to a Multiplier that forces a difference between the first block's Probe and its paired Irrelevant (e.g. Probe size increased by 6x that of its paired Irrelevant). As a result, we observed that adding the lowest Noise ASP level (i.e. 10) to the human EEG signal will not be enough to introduce sufficient randomness to the Probe/Irrelevant conditions because the slightest difference between them (e.g. when Probe size is increased by two times that of its paired Irrelevant) results in high significance. Conversely, the higher Noise Power levels (e.g. 50) swamps the human EEG signal, requiring excessively large differences between the paired Probe/Irrelevant conditions, to achieve any significance. Therefore, we chose 20, as our standard Noise ASP level.

The results of the above exploration confirmed that at the lowest Noise ASP level of 10, the human EEG signal was largely unaffected by the addition of random noise because the slightest difference between the Probe and Irrelevant conditions showed a high statistical significance for block-1 (as well as, its inferred blocks 4 and 5). We consider this set-up (i.e. ASP of 10) to be too low for our data simulation purposes, as the human EEG signal overpowers the random noise, even with the slightest difference between the Probe and Irrelevant conditions.

At the other end of the scale (i.e. the highest Noise ASP level of 50), the above results confirmed that the addition of random noise swamps the human EEG signal, such that relatively large differences between the Probe and Irrelevant signals (e.g. Probe being six times larger than the Irrelevant) would only result in marginal levels of significance. As a compromise, we chose the ideal Noise ASP level of 20 (as the standard setting for all simulations, in this chapter), with the knowledge that it would enable us to explore observed/statistical associations, by only regulating the size difference between the Probe and Irrelevant conditions (i.e. changing the power of the resultant time series, by adjusting the *Multiplier* factor of the first block's Probe signal).

6.5 Type I error investigations

Having devised a procedure for simulating our EEG data (see section 6.4 – *Noise Generator*), we were able to investigate the extreme variability in the data, using statistical tests, by analysing the probability that an observed difference between the two conditions could be due to chance alone. In this section, our aim is to test the probability of Type I errors (false positives) by ensuring that both Probe and Irrelevant conditions are, in a statistical sense, the same, resulting in a true null hypothesis, where type I error relates to rejecting the null hypothesis when it is actually true. To fully test false positive conclusions, we utilised our new online technique for inferring the significance of Part I of the experiment (hereafter, called *Best Block)*, where the most significant block in Part I of the experiment informs Part II.

After generating 50 simulated trials for each condition (i.e. 50 Probes and 50 Irrelevants), we averaged each condition's 50 trials, resulting in a single ERP for the Probe and a single ERP for the Irrelevant (see Figure 6.5). By combining these two conditions, we were able to search for the Region of Interest (ROI), using the aERPt method (i.e. by independently searching for the highest positive mean amplitude in a 100ms time window, throughout the waveform). Next, the mean of each condition (Probe and Irrelevant) was taken at the ROI window, and then they were subtracted to define the *true* Observed Difference of Means (ODoM).

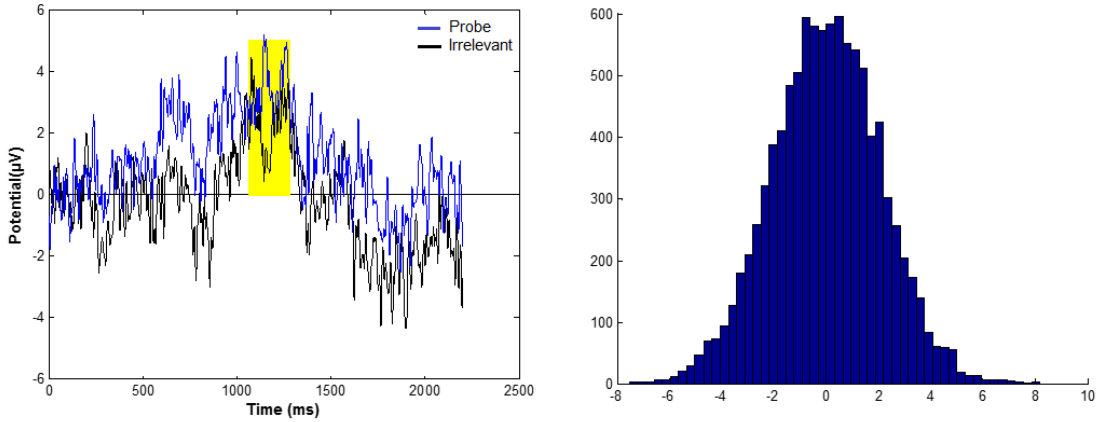

Figure 6.5 – Averaging 50 trials, for each condition, resulted in two ERPs (left plot): the Probe (in blue) and the Irrelevant (in black) conditions. Combining both conditions enabled us to search for a Region of Interest (ROI), using the aERPt method (i.e. highest positive mean amplitude window, which is highlighted in yellow), and subtracting the means of the two conditions, at the ROI, defined the true Observed Difference of Means (ODoM). To demonstrate that the null holds for the true-ODoM, we repeated the above process 10,000 times, and plotted a distribution of 10,000 true-ODoM iterations (right plot), showing a Normal (Gaussian) distribution, centered at zero.

To statistically test our null hypothesis that there is no difference between the two conditions, a randomisation (i.e. Monte Carlo permutation) test was used to produce a null distribution and a *p*-value: we started by combining all Probe and Irrelevant trials and shuffled them into a single bin. Next, we split them into two surrogate groups and used the original ROI window to find the *surrogate* ODoM. Having recorded this surrogate-ODoM, we repeated the resampling test, in order to obtain 1,000 surrogate-ODoM results. Finally, we compared the distribution of the 1,000 resampling iterations

(i.e. surrogate-ODoM values) with the single true-ODoM, and thus obtained a single *p*-value for the first block of the experiment (see left plot of Figure 6.6).

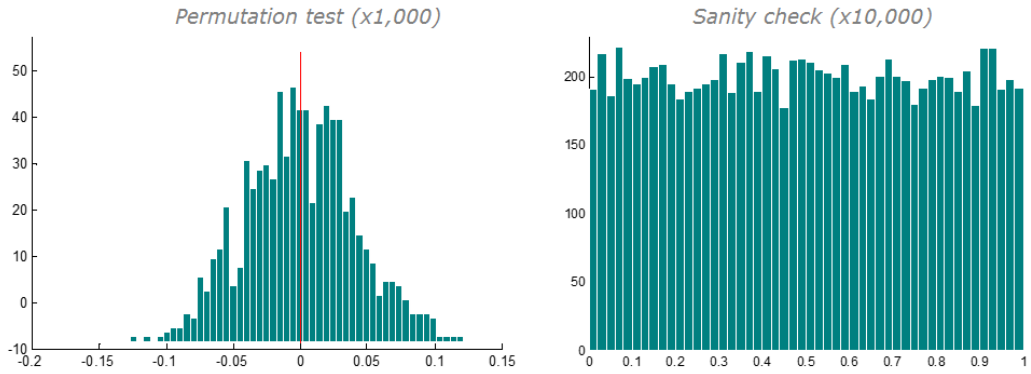

Figure 6.6 – To show that our analysis was well-behaved (i.e. did not inflate type I error rate), we used a randomisation test, by combining and shuffling all the trials of both conditions, and splitting them into two surrogate groups (i.e. surrogate-Probe and surrogate-Irrelevant). Then, using the same ROI window to find each group's highest positive mean value, we subtracted the two results to obtain a surrogate-ODoM. Repeating this process 1,000 times produced a distribution of surrogate-ODoM results, and comparing the distribution with the original true-ODoM resulted in a single p-value (p = 0.499, left plot). Note that, as is most likely, the p-value is not significant because both conditions are a product of random noise plus the same EEG signal. Therefore, to confirm lack of bias, we performed a sanity check, by repeating the entire procedure 10,000 times, and each time obtaining a new p-value (right plot). Plotting the distribution of all 10,000 p-values revealed a Uniform distribution because all iterations employed random noise plus the same human EEG signal, for each of their Probe and Irrelevant conditions. This is the full validation to show that, in a statistical sense, our procedure does not inflate false positive rates.

To demonstrate that our analysis was not biased (i.e. the data simulation procedure did not inflate false positive rates), we repeated the above process 10,000 times, and plotted the distribution of 10,000 *p*-values (see right plot of Figure 6.6), which revealed a Uniform distribution. This is the full validation that our data generation procedure and the method of statistically testing the results is well-behaved, and that it does not inflate type I error rates.

However, for the sake of completeness, we will also demonstrate that our new experimental design, which uses the results of Part I to infer the paired conditions for Part II, will not inflate false positive conclusions (vis-à-vis the *Decider* method), as long as the same human EEG signal (taken from our 2013 Names experiment) is used for all

blocks of Part I (see section 6.5.1). Additionally, we will introduce three different signals for the first three blocks of the experiment (i.e. each of the first three blocks will be assigned a different human ERP signal, taken from three different subjects in our 2016 lecturer faces experiment; see chapter 5), and demonstrate that both Part I and Part II of the new experimental design will remain well-behaved (see section 6.5.2), as long as each block of Part I uses its unique EEG signal for both of its paired Probe/Irrelevant conditions (i.e. the only difference between them will be the random noise data).

6.5.1 – Same EEG signal for all blocks of Part I

By applying the aforementioned noise generator process, which utilises a single human EEG signal (see Figure 6.3), to our new experimental design, we performed randomisation tests to obtain three *p*-values for the three blocks in Part I of the experiment. Of course, we have already demonstrated that, statistically speaking, Part I of the experiment remains well-behaved (i.e. none of the blocks inflate the false positive rate), but we wish to demonstrate that Part II, which will be influenced by the results of Part I, will not suffer any bias. Consequently, the lowest *p*-value from Part I advanced the *primary* block's paired-conditions (i.e. the Probe and Irrelevant from the block that was most significant in Part I of the experiment), which would then be re-used in block-4 and block-5, of Part II. The resultant five blocks defined the '*Best Block*' technique for the first pseudo-subject, and by repeating the whole process multiple times (e.g. for a total of 5,000 pseudo-subjects), we would be able to confirm that our experimental design will not, in a statistical sense, inflate type I error rates, in Part II.

As a result of the above process, each pseudo-subject possessed five blocks; three in Part I and two in Part II. The latter two blocks (i.e. block-4 and block-5) were inferred/influenced by the outcome of Part I, and, thus, there could be an inflation of false-positive rates. However, due to the fact that all Probes and Irrelevants in Part I were employing the same human ERP-signal (i.e. the only difference between them was random noise), our Null Hypothesis was that there would also be no difference between the two conditions in Part II. As expected, a distribution of 5,000 iterations of the experiment resulted in a uniform distribution (see Figure 6.7), and there was no

significant difference between Part I and Part II (i.e. the 2 blocks that inherited the best parameters from Part One) of 5,000 pseudo-subjects.

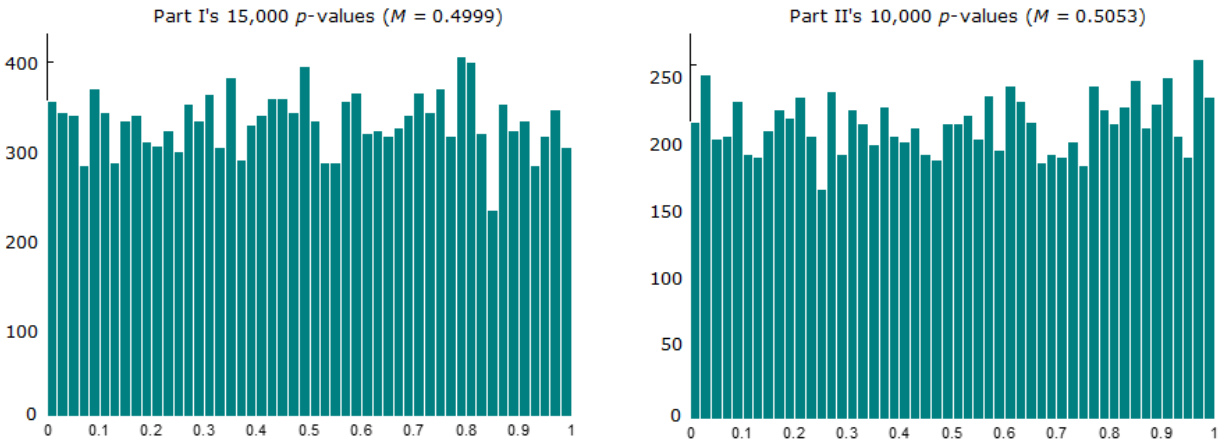

Figure 6.7 – In our 'Best Block' technique, statistical test results for Part I (i.e. left-plot's 15,000 p-values for the first three blocks) were similar to the results in Part II (i.e. right-plot's 10,000 p-values for the latter two blocks), even though, the two most significant paired conditions from Part I were independently selected and re-used in Part II. The uniform distributions for both parts of our new experimental design confirms the lack of bias.

Whilst the above confirmed that there would be no bias in Part II, our new experimental design requires confirmation that aggregating all the trials of each iteration of the experiment (i.e. in the form of the '*Combined'* method) will not inflate the type I error rates. Once again, we obtained three *p*-values for the three blocks in Part I of the experiment, and re-used the lowest *p*-value's paired-conditions for both the fourth and fifth blocks, of Part II. As before, the resultant five blocks would become the first pseudo-subject, and by repeating the whole process multiple times (e.g. for 5,000 pseudo-subjects), we were able to confirm that our preferred experimental design will not, in a statistical sense, inflate type I error rates.

In summary, even though both blocks in Part II re-used the same parameters from Part I, we demonstrated that there was no bias, and that a distribution of 5,000 iterations of the experiment, using the proposed *Combined* method, resulted in a uniform distribution (see Figure 6.8). This was the absolute validation that our experimental design does not inflate false positive rates.

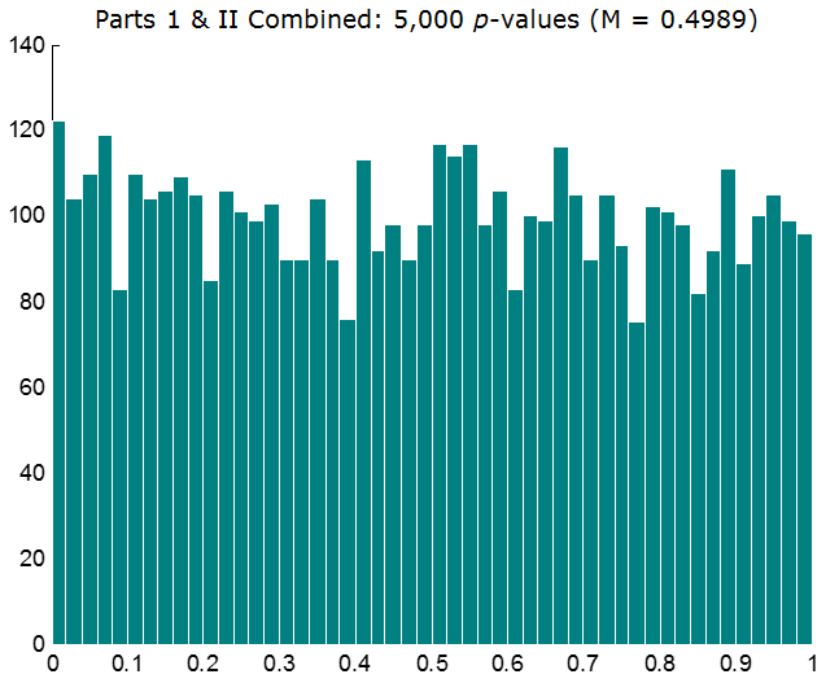

Figure 6.8 – In 'Best Block' technique, the most significant paired conditions from Part I were re-used for both blocks of Part II, and all trials for the experiment were combined to obtain a single p-value, using statistical tests. Dirstribution of 5,000 p-values resulted in a uniform distribution, confirming that the new experimental design, with its Combined analysis method, is well-bahaved.

6.5.2 – Unique ERP signals, for each block of Part I

The above demonstration confirmed that there could be no bias when all blocks in Part I utilise the same simulated human ERP-signal. However, we chose to investigate the rates of the probabilities of type I errors, when different EEG-signals were being used for each block in Part I. Thus, we selected three different subjects' ERP signals, from our 2016 familiar lecturer faces experiment (see Figure 6.9), to ensure that our simulated data generation procedure would produce three distinctly different ERPs, for each block of Part I. Using the same noise generator procedures (described in section 6.4), we substituted the single human EEG-signal (shown in Figure 6.3) with one of the unique subject-ERPs, so that each block's Probe and Irrelevant conditions, in Part I of the experiment, would be different. This arrangement would enable us to test the inflation of false-positive rates, and reveal that without a significant difference between each block's Probe and Irrelevant conditions, there would still be no bias.

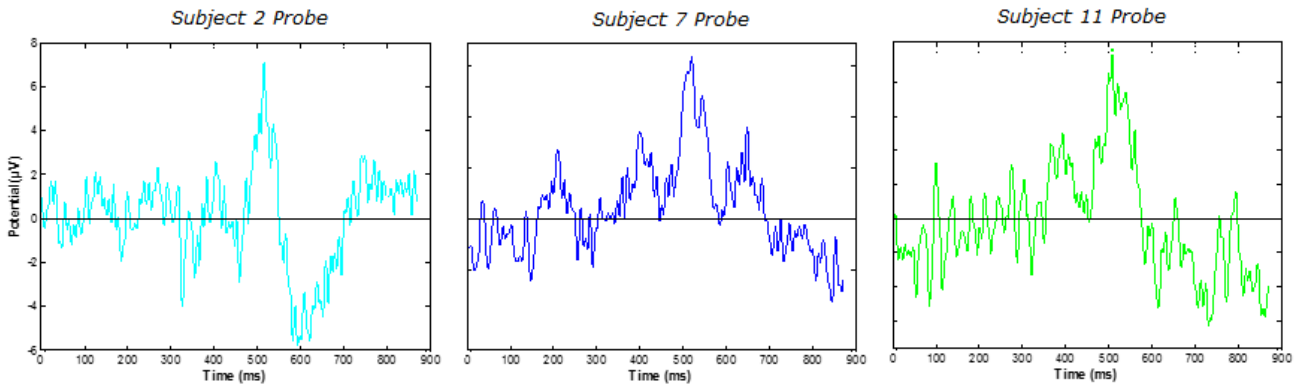

Figure 6.9 – Instead of the single simulated human-EEG signal, which had been used for all blocks of the previous data simulations (see section 6.5.1), we used three different subject-ERPs, from chapter 5's familiar lecturer faces experiment (i.e. Probe ERPs for subjects 2, 7 and 11). Each subject-ERP was randomly assigned (without replacement), to the two/paired Probe and Irrelevant conditions, for the three blocks of Part I. Therefore, with the exception of random noise, we expect no statistical difference between the paired conditions within the same block, but, of course, there could be significant differences between the three blocks, of Part I.

As a result of randomly assigning each of the three subject-ERPs to our data simulation process, we guaranteed differences between the three blocks of Part I. Thus, within the same block, the difference between the Probe and Irrelevant conditions was only influenced by the presence of random noise, albeit, each block in Part I would differ, due to its unique subject-ERP characteristics. Using statistical tests to compare the difference between Probes and Irrelevants, we obtained *p*-values for the three blocks in Part I, and selected the most significant block's paired Probe/Irrelevant conditions, for blocks 4 and 5, of Part II. Having repeated the above process 5,000 times, we wanted to examine which of the three analysis methods (i.e. Abandoned, Biased or Combined) might inflate false-positive conclusions.

Our hypothesis was that even if we use different EEG-signals for each block of Part I (i.e. using one of the three subject-ERPs from a recent experiment), as long as each block's paired Probe and Irrelevant conditions use the same subject-ERP, the statistical difference between the conditions of comparison will not be significant. However, whilst the Abandoned and Combined methods are well-behaved (i.e. neither inflates type I error rates, though, we will later demonstrate, in section 6.6, that the Combined method benefits from a higher statistical power), the Biased method is fallible, since it inflates the possibility of type I errors, as shown in the distribution of 2,000 iterations of the new experimental design (see Figure 6.10).

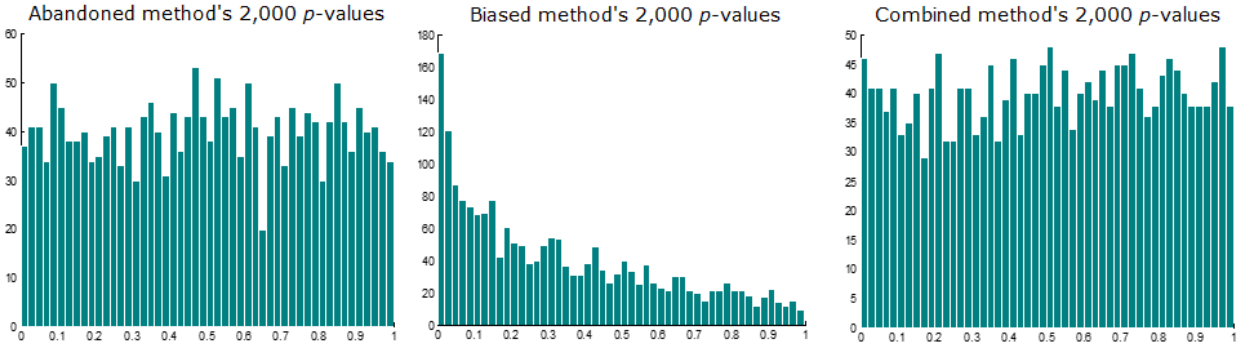

Figure 6.10 – Distribution of 2,000 iterations of the new experimental design, using the three different analysis methods: Abandoned (left plot), Biased (middle plot) and Combined (right plot). Applying three different subject-ERPs for the three blocks of Part I (and matching each block's Probe and Irrelevant conditions), enabled us to identify the block with the highest significance, so that its pair of conditions could be re-used for both blocks of Part II (i.e. using the 'Best Block' technique). Then, we combined the trials, according to the three analysis methods (i.e. Abandoned, Biased and Combined), and performed statistical tests to show a single p-value for each method. Consequently, a distribution of 2,000 iterations of the above process confirmed a uniform distribution for the Abandoned (left-plot) and Combined (right-plot) methods, but the Biased method inflated the type I errors (middle plot), as the best block in Part I was cherry-picked and combined with Part II (i.e. Biased is akin to fishing for results).

6.6 Type II error investigations

Next, we addressed the inflation of type II errors, which relates to failing to reject a False null (i.e. when there was indeed a difference between the conditions). Our hypothesis is that if the *Multiplier* factor (introduced in section 6.4.1) of the first block's Probe is increased whilst all the other Probes and Irrelevants in Part I remain the same, it is more likely that there will be a difference between the paired Probe & Irrelevant conditions of block-1. Thus, the artificially manipulated block-1 is (statistically speaking) more likely to be advanced to Part II of the experiment, in the majority of circumstances, where it can influence block-4 and block-5 of Part II. Note that we will be using unique subject-ERP signals, for each of the three blocks of Part I, as described in section 6.5.2.

In this section, our aim is to reduce the probability of false-negative conclusions, and to demonstrate that our new experimental design's *Combined* method, which did not inflate false positive rates (as shown in section 6.5), has the lowest type II error rates (i.e. the highest statistical power), when compared to the alternative analysis method: *Abandoned* (bearing in mind that *Biased* has already been excluded in section 6.5.2).

6.6.1 – Three Analysis Methods

As we have noted earlier, the new experimental design allows for a two-part experiment, where Part II (containing two blocks) is influenced by the significance of Part I (containing three blocks). Subsequently, the resultant five blocks of the experiment can be analysed using three unique methods:

A) Abandoned, where we will only use the data in Part II, by discarding all the data in Part I (i.e. as if Part II is a new experiment);

B) Biased, where we will only use the data of the most significant block of Part I, plus all the data in Part II (i.e. fishing for best results only);

C) Combined, where we will use all the data in Part 1, plus all the data in Part 2 (i.e. the entire two-part experiment).

It is our intention to demonstrate that the Abandoned (A) method is safe and without bias, as it is similar to running a new/unrelated experiment, however, this method is wasteful because the data in Part I is being discarded. Furthermore, method A will reduce the Signal-to-Noise Ratio (SNR), partly because the data in Part II is a small sub-set of all the available data (i.e. Part II contains less than half of the data in the whole experiment), thus, unless the number of Trials are significantly increased, method A will be at an SNR disadvantage. As we have already shown in section 6.5.2, the Biased (B) method is not safe and will inflate the false positive rates. In fact, method B can be likened to fishing for results by favouring the most significant blocks, thus, inflating type I error rates. Finally, we conclude that the Combined (C) method is both safe (compared to method B) and benefits from an improved SNR (compared to method A). However, the scientific question is whether we can justify method C and obtain few type II errors? In other words, does method C benefit from a higher statistical power?

To answer this question and to explore the limitations of the analysis methods, we chose to artificially manipulate the parameters that can influence the SNR, by varying three attributes: a) the '*Noise'* amplitude scaling parameter; b) the '*Multiplier'* difference between the two conditions; c) the number of '*Trials'* in each block.

6.6.2 – Manipulating the Noise

The first exploration of false-negative rates involved the manipulation of the noise Amplitude Scaling Parameter (ASP), which, incidentally, is the same parameter that was used to define the standard Noise ASP level, in section 6.4.1. As such, we applied different *Noise* ASP settings, using intervals of 10, 20, 30, 40 and 50, whilst simultaneously increasing the *Multiplier*'s amplitude difference between the first Probe and its paired Irrelevant (i.e. forcing a difference between the first block's Probe and Irrelevant conditions). So long as the number of *Trials* remained the same, each combination of the *Noise*-and-*Multiplier* levels could be repeated 5,000 times – using the new experimental design, where Part II of each iteration would be influenced by the significance of Part I – and the results could be processed using the Abandoned and Combined analysis methods (see Figure 6.11).

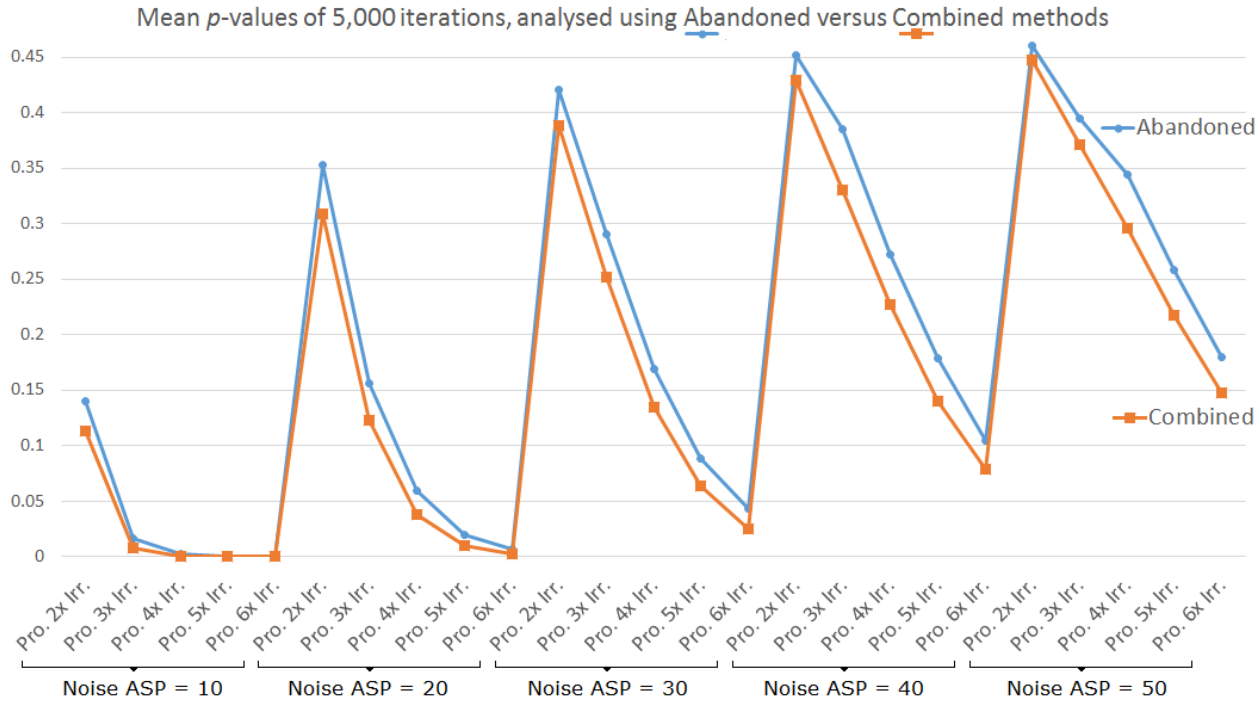

Figure 6.11 – Mean p-values of 5,000 iteratins of the new experimental design, for 25 different combinations of Noise ASP (10, 20, 30, 40 and 50) and Probe Miltiplier (2, 3, 4, 5 and 6 times that of the first block's Irrelevant). At every level, the Combined method (consisting of all five blocks of Part I and II) benefits from a higher statistical power when compared to the Abandoned method (consisting of two blocks in Part II, only), confirming that the improved SNR of the Combined method results in the lowest type II error rates.

Having explored 25 different combinations of *Noise* ASP (10, 20, 30, 40 and 50) and Probe *Miltiplier* (2, 3, 4, 5 and 6 times that of the first block's Irrelevant) settings, the mean *p*-value of 5,000 iterations, using the Abandoned and Combined methods of analysis, confirmed that the Combined method wins over the Abandoned method (see Figure 6.11). The improved statistical power of the Combined method can be attributed to the larger number of trials (i.e. utilising 5 blocks, as opposed to Abaondoned method's 2 blocks), resulting in improved SNR and lower type II error rates.

6.6.3 – Manipulating the Probe Multiplier

The second exploration takes a finer look at the Probe *Multiplier*, by maintaining a fixed *Noise* ASP of 20 (i.e. the default value that was justified in section 6.4.1), and by limiting the number of simulated *Trials* to 50 per iteration (split equally between 5 blocks). Thus, by progressively increasing the *Multiplier* of the first block's subject-ERP signal, to artificially generate a difference between block-1's Probe and Irrelevant conditions, we were able to explore a localised effect of the statistical power between the Abandoned and Combined methods (see Figure 6.12).

Using six *Multiplier* levels – where the first block's Probe would become 1.1, 1.5, 2, 3, 4, 5 and 6 times the amplitude of its paired Irrelevant – we simulated our new experimental design, whereby, the most significant block's paired conditions of Part I were re-used in both blocks of Part II. Next, we combined the trials of Part II for the Abandoned method (i.e. blocks 4 and 5 only), and separately combined all the trials of Parts I & II for the Combined method (i.e. blocks 1 to 5), before performing statistical tests to obtain two *p*-values; one for the Abandoned method, and the other for the Combined method. Finally, we repeated the above process 5,000 times, at each level of the Multiplier (i.e. 30,000 pseudo-experiments in total), and compared the statistical power of the mean *p*-values of the two analysis methods (Figure 6.12).

Multiplier	Abandoned	Combined
1.1 x	0.4852	0.4797
1.5 x	0.4506	0.4190
2 x	0.3823	0.3300
3 x	0.2424	0.1670
4 x	0.1308	0.0755
5 x	0.0712	**0.0298**

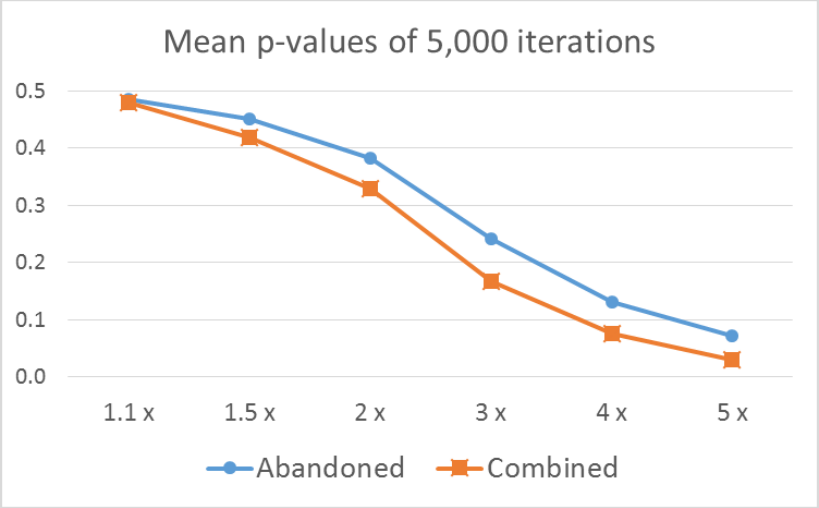

Figure 6.12 – Investigations into changes in Probe Multiplier, where the number of Trials for each block remained the same, and the Multiplier of the first block's Probe was progressively increased, from 1.1 to 5 times that of its paired Irrelevant. Using the new two-part experimental design, statistical tests produced a p-value for the Abandoned and the Combined methods. To demonstrate the difference between these two analysis methods, each of the six multiplier levels were subjected to 5,000 iterations, and the mean p-values (see the table, on the left side, and the plot on the right side) shows that Combined always wins over Abandoned, due to its higher statistical power.

Results of the above investigations into changes in Probe *Multiplier* level provided additional support that the Combined method wins over the Abandoned method, due to its higher statistical power, confirming the improved SNR of the Combined method (i.e. lower type II error rate).

6.6.4 – Manipulating the number of Trials

Whilst it is feasible to explore the theoretical effects of increasing the number of *Trials* in our experiments, it must be noted that there will always be a practical limit to the size and duration of a real-life experiment. In addition to the cost/convenience of extending the duration of an experiment well beyond two hours, a subject's effectiveness (e.g. ability to remain focused) can be a limiting factor. In past experiments, we had elected to limit the total number of trials to 225 (e.g. 3 blocks consisting of 75 trials per block), which would last approximately 1.5 hours. However, our new experimental design dictates that our next experiment will be extended to 375 trials (i.e. 5 blocks of 75 trails), which could last approximately 2 hours.

In this final exploration, we focused on the relationship between the number of *Trials* and the number of *Blocks*. To simulate a representative comparison between them, we chose to limit the *maximum* number of trials in the experiment to 180 because it enabled us to apply nine factor-pairs for our data simulations (i.e. trail/block combinations that are divisible by 180 are: 45/4, 36/5, 30/6, 20/9, 18/10, 15/12, 12/15, 10/18, and 5/36). This time, as well as fixing the *Noise* ASP (i.e. 20, as justified in section 6.4.1), we also limited the Probe *Multiplier* (i.e. 2 times greater than its paired Irrelevant, as investigated in section 6.6.3), so that we could explore the consequences of changing the relationship between the number of *Trials* (from 45 to 5 per block) and the number of *Blocks* (from 4 to 36 blocks per experiment). Note that the trial/block ratio must maintain the same total number of trials for each pseudo-experiment.

This simulation will be able to investigate an interesting set-up, whereby, the number of blocks in Part I of the experiment is always fixed (i.e. Part I has three blocks), but the number of blocks in Part II will increase from 1 to 33 (i.e. the total number of blocks in the experiment will start from 4 and reach a maximum of 36).

Since the total number of trials in the experiment will always remain the same, it must be noted that as the number of blocks in Part II increases, we are effectively sacrificing trials in Part I (i.e. the number of trials in the first 3 blocks will decrease), in order to boost the number of trials that will be required for the additional blocks, in Part II. Therefore, as the significance of Part I of the experiment will independently mandate the paired Probe/Irrelevant conditions that will be inferred to Part II (i.e. using the *Best Block* technique, where the most significant pair will be re-used in all the blocks of Part II), we hypothesise that the reduction in the number of trials-per-block will reduce the statistical power of inferring the *Best Block*, and result in reduced significance. To simulate the above set-up, we repeated the process 5,000 times, at each of the nine trial/block settings (i.e. 45,000 pseudo-experiments in total), and compared the statistical power of the mean *p*-values of the two analysis methods (see Table 6.1).

Trials	Blocks	Total trials in each Mehod:			Mean p-values of 5,000 iterations:			Natural Log:
per block	*per exp.*	A	B	C	Abandoned	Biased	Combined	ln(A) - ln(C)
45	4	45	135	180	0.1958	0.0741	0.1479	**0.2806**
36	5	72	144	180	0.1574	0.0829	0.1268	**0.2162**
30	6	90	150	180	0.1662	0.1072	0.1371	**0.1925**
20	9	120	160	180	0.1872	0.1441	0.155	**0.1888**
18	10	126	162	180	0.2043	0.1621	0.1716	**0.1744**
15	12	135	165	180	0.2144	0.1785	0.1854	**0.1453**
12	15	144	168	180	0.211	0.1848	0.1925	**0.0918**
10	18	150	170	180	0.2315	0.2058	0.2083	**0.1056**
5	36	165	175	180	0.2604	0.2464	0.2483	**0.0476**

Table 6.1 – Results of investigations into the comparative changes between the number of Trials and Blocks: by limiting the maximum number of trials in the experiment to 180, we were able to investigate the effects of changing the relationship between the number of Trails and Blocks (first two columns). Note that, as a result of this set-up, the three analysis methods will contain different number of trials (see columns A, B and C, under the heading "Total trials in each Method"), thus, as the number of blocks increase, the difference between the total number of trials in each method declines. However, as the number of Trials-per-block decreases, the statistical power of inferring the most significant pair of conditions (from Part I to Part II) reduces, resulting in higher p-values (as indicated by the colour-shading of the mean p-values). By combining the trials, in accordance to the three analysis methods, and using statistical tests to obtain a p-value, we were able to repeat this process 5,000 times, and show that the Combined method always wins over the Abandoned method (note that the difference of the natural log of the two methods, shown in the last column, confirms this finding).

Whilst remembering that the number of blocks in Part I is always three, we note that, as the number of trials-per-block decreases, the consequence of this set-up is that the number of blocks will increase, since the total number of trials for the experiment must remain the same (i.e. 180 in total). Thus, as we hypothesised, the reduction in the number of trials in each block of Part I will reduce the statistical power of inferring the correct pair of conditions to Part II, as shown in the results (see Table 6.1), where the mean *p*-values of all three methods begin to drop, but rise back up, as the number of trials-per-block decreases.

Furthermore, as shown in Table 6.1, the total number of trials for the Combined method is always the same (180), whereas, the number of trials in the other two analysis methods will be different, since Abandoned and Biased retain a subset of the total number of blocks. As we are only interested in the Abandoned and Combined methods, we must note that as the number of trials-per-block decrease, the difference between the total number of trials in the Abandoned method approaches that of the Combined

method. Whilst we hypothesised that the Combined method would always win over the Abandoned method, we acknowledged that as the difference in the number of trials decreases, the statistical power of the Abandoned method would close in on the Combined method (albeit, it will never match). This was confirmed by the results of our data simulations (see Table 6.1), and further highlighted in Figure 6.13 (below), which plots the difference between these two methods. In addition to showing the improved statistical power of the Combined method, the mean significance was shown to be at its greatest, at an optimum number of blocks (i.e. 5 blocks), which has been an influencing factor in the design of our next (Revealed Lecturer Faces) experiment (see Chapter 7).

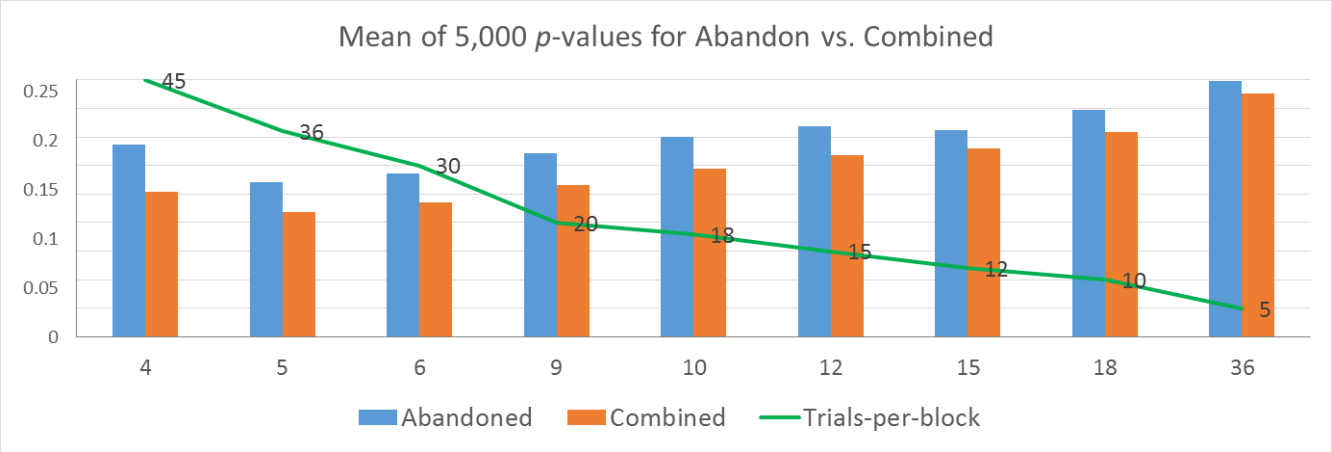

Figure 6.13 – According to the results in table 6.1, the mean p-values of 5,000 iterations of the experiment (y-axis marks p-values), using the Abandoned and Combined methods, confirmed that the Combined method wins over the Abandoned method, at all combinations of trial/block settings (x-axis marks the number of blocks, and the green line marks the number of trials-per-block). Note that the difference between the mean of 5,000 p-values for the two analysis methods progresssively reduces, as the number of blocks increase. Furthermore, the mean significance of both methods are shown at their greatest (i.e. lowest p-values) when the number of blocks equals 5, indicating an optimum level, which will be used in our next (Revealed Lecturer Faces) experiment.

Finally, by taking the natural log of each method (Abandoned versus Combined) and subtracting them from one another, we can confirm that – within the simulated range: 4-blocks of 45-trials to 36-blocks of 5-trials – the significance of the mean of 5,000 iterations, will always favour the Combined method of analysis (see Figure 6.14). Thus, the improved statistical power of the Combined method, over the Abandoned method, can be attributed to the larger number of trials (e.g. at their closest margin, 36-blocks of 5-trials, the Abandoned method possesses 165 trials, versus the Combined method's 180 trials), resulting in improved SNR and lower type II error rates.

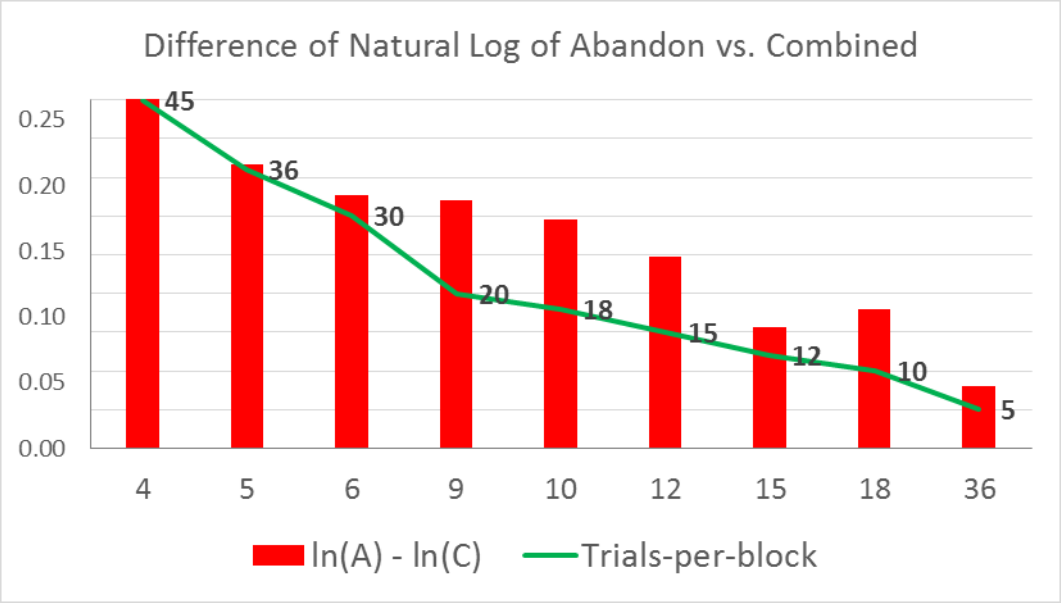

Figure 6.14 – According to the results in table 6.1, the difference between the natural log of the Abandoned and Combined methods confirmed that the significance of the mean of 5,000 iterations will always favour the Combined method of analysis, due to its inherent lower type II error rates. Thus, we have chosen the Combined method of analysis for our new (Revealed Lecturer Faces) experiment (see Chapter 7).

6.7 Discussion

This chapter's methodological explorations were intended to advance future applications of EEG-based deception detection studies of crime compatriots (i.e. to establish a suspect's familiarity with a criminal/terrorist, using faces). We facilitated our ground-truth explorations by using a noise generator (see section 6.4), which produced random noise with the power spectrum of human EEG, in which noise and signal components could be combined to generate epochs of simulated EEG.

Having formed our theories and verified our hypotheses, we expected our findings to inform the proposed new design of our next/final EEG experiment (see Chapter 7), which would adopt a novel technique of mid-experiment inference, whereby the most significant critical stimulus, from the first-part of the experiment, would be independently selected and re-used in the second part. To that end, this chapter's exploration of improvements in statistical power (i.e. to reduce the risk of Type II errors), without the inflation of Type I errors, focused on the suitability of three different methods for analysing our two-part experiments:

A) *Abandoned*: using two blocks of Part II only, by abandoning Part I;

B) *Biased*: using two blocks of Part II, plus the most significant block of Part I;

C) *Combined*: using all five blocks of Part I and Part II of the experiment.

To investigate the rates of type I errors (see section 6.5), we hypothesised that even if we combine random noise with different EEG signals (e.g. unique subject-ERPs for each block of Part I), as long as each block's Probe and Irrelevant conditions are derived from the same EEG-signal, statistically speaking, there will be no difference between them. After showing the uniform distribution of statistical tests carried out on the difference between the Probe/Irrelevant conditions (i.e. no inflation of type I errors), we introduced the new experimental design, consisting of two-parts, where the blocks in Part II are influenced by the significance of the blocks in Part I. Furthermore, we replaced the single EEG signal, which contributed to the data simulations, with three unique subject-ERPs, and performed thousands of pseudo-experiment iterations. Then, we aggregated the trials, in accordance with three new methods of analysis: *Abandoned*, *Biased* and *Combined*. Plotting distributions of statistical tests confirmed that whilst the Abandoned and Combined methods are well-behaved (i.e. neither inflated type I error rates), the Biased method inflated type I errors, and should be rejected.

To investigate the rates of type II errors (see section 6.6), we hypothesised that if the difference between one of the block's conditions is artificially enhanced (e.g. the Probe Multiplier of block-1 is increased by two times that of its paired Irrelevant), the expected *p*-values of all three analysis methods (i.e. Abandoned, Biased and Combined) will improve. However, we were no longer interested in method B (Biased), which proved to be unsafe because of its fishing characteristics that could inflate false-positive rates. Thus, we explored the false-negative rates of method A (Abandoned) versus method C (Combined). The former was considered to be safe because it focuses on Part II alone (i.e. similar to running a new/unrelated experiment, which has two blocks only), but as it suffers from lower SNR (i.e. there are fewer trials in method A), it could lead to higher false-negative rates. Using statistical tests of our data simulations, we showed that the same flaw was not present for the Combined (C) method, confirming that method C has the highest statistical power of the two.

In conclusion, even though the new experimental design consists of two parts, where Part II's parameters are independently influenced by Part I's results, our data simulations confirmed that as long as we use the Combined method of analysis, we can be assured that our statistical test results are well-behaved (compared to Biased), and possess the highest statistical power (compared to Abandoned). Therefore, we recommend the new experimental design, in all future RSVP-based experiments (as outlined in section 6.2), along with the Combined method of analysis, for statistically testing RSVP-based EEG data, at group and subject levels.

Chapter 7:
EEG study 3: Infer-and-Combine of Revealed Lecturer Faces

7.1 Introduction

Through the evidence gathered in the first two experiments, we have successfully established that famous faces of celebrities, as well as, familiar faces of lecturers can breakthrough into conscious awareness (using an RSVP subliminal search paradigm), and that our statistical tests can differentiate between the Probe (familiar celebrity or lecturer) and Irrelevant (unknown) faces, at group and subject levels (see chapters 4 and 5). Further, having investigated methods for improving statistical power, without inflating false-positive rates, we have established an experimental design (see chapter 6) that could promote the use of faces in RSVP-based EEG tests for deception detection applications. As a result, the objective of this chapter was to demonstrate that we can bring together all of our findings, to improve the detection at subject-level, and to demonstrate that our methods (i.e. experimental design and statistical analysis methods) can be used in real-world concealed information tests.

In this (the final) experiment, we continued to utilise the personally known faces of the University's lecturers (as Probes), and the unknown faces of lecturers from another University (as Irrelevants), in order to differentiate between the Probe and Irrelevant conditions, at group and subject levels. As explained in chapter 6, we used Part I of the experiment to independently select one of the three Probes and re-use the chosen Probe/Irrelevant pair in Part II of the experiment. This selection process was conducted using online statistical tests to infer the Probe (i.e. familiar lecturer face) that achieved the highest significance. The resultant five blocks (i.e. three in Part I and two in Part II) were then used to perform group and subject level statistical tests.

Aside from the latest improvements in the design/analysis techniques, the key change in this experiment related to the instruction given to subjects, at the start of the experiment; whereas, the presence of the Probes (i.e. familiar celebrity or lecturer faces) was *concealed* from subjects in the previous two experiments, in this (the third/final)

experiment, we revealed the possibility that familiar lecturers may be presented in the RSVP streams. However, subjects were still instructed to look for the Target (which remained task-relevant), and they were not told which lecturers might be included. By revealing the presence of the Probes, we have staged a real-life scenario, in which the subject is patently aware of the examiner's ultimate goal – in essence, a perpetrator who is being questioned about a crime, will naturally assume that the purpose of being shown a series of faces is to ascertain his/her familiarity with an accomplice. Therefore, we consider this (*revealed lecturer faces*) experiment to be a workable solution for deception detection applications of compatriots, using faces in RSVP-based EEG tests.

As before, by referencing the standard design and analysis methods, described in Chapter 3, we will begin by outlining the revealed lecturer faces experiment, and then summarise the group-level analysis. Next, we will describe our in-depth group and subject level analyses, in the Time (ERP) and Frequency (ERSP/ITC) domains. Finally, we will discuss the results and draw conclusions to our hypotheses, based on the evidence gathered.

7.2 Experiment's Hypotheses

In our final exploration of the suitability of the RSVP paradigm to infer the recognition of familiar/compatriot's faces, in real-life EEG-based deception detection tests, we revealed to the subject, the presence of the Probe (familiar lecturer faces), in order to test the following hypotheses, experimentally:

i) Revealing the presence of familiar lecturer faces that are personally known to the subject, instead of the previous practice of concealed inclusion of such lecturer faces, will result in a similar group-level breakthrough of Probe (familiar lecturer) faces, which are differentially perceived and processed, as compared to Irrelevant (unfamiliar lecturer) faces;

ii) In the first experiment of its kind, the application of the new design (which involves online qualification of the Probe), and the use of the new analysis method (which increases the SNR, without inflating the false positive/negative

rates) can improve the breakthrough and detection of Probe (lecturer) faces, on an individual basis, even though, only the Target was task-relevant.

iii) In accordance with the first and second experiments, the strongest brain responses to the lecturer (Probe) faces are recorded at the Pz electrode site.

7.3 Design of the third Experiment

7.3.1 – Experiment's Participants

Fifteen participants were tested, but one was excluded, as explained below, leaving fourteen subjects for our study. Out of 14 subjects, 13 were male (93%) and 1 female (7%). The ages of the subjects ranged from 22 to 28 (M = 23.6 years, SD = 1.83); 13 of them were right-handed (93%), and one was left-handed (7%). Subject no. 5 was excluded (and replaced with an additional subject, no. 15) because the participant could not easily recognise familiar faces – even though the participant reported no neurological conditions or cognitive disorders, the end-of-experiment recognition questions (see 7.3.5 – *Experiment's Probe/Irrelevant Questions*) highlighted a surprising lack of familiarity with lecturer faces that were well known to the subject. Upon further enquiry, the subject confirmed having difficulty remembering faces (akin to prosopagnosia), therefore, we did not process this subject's data, and added another subject (no. 15), to bring the total number of participants up to the intended 14 subjects.

Following on from the second (concealed lecturer faces) experiment, this experiment's Probe stimuli would also employ the University of Kent's lecturer faces (from the School of Computing), who had a close working relationship with their lecturers. As before, we asked the lecturers to covertly nominate PhD students only, so that we could be assured of a long-term relationship/familiarity between subjects and their lecturers' faces. Also, subjects were chosen on the basis of never having been included in a similar EEG/RSVP experiment, and at the end of each experiment, participants were instructed to avoid discussing the experiment with their colleagues, in order to avoid any priming of future participants. The duration of each experiment was over 2 hours, and each subject was paid £15 (fifteen pounds) for their time (note that this was £5 more than the previous two experiments, due to the extra duration).

7.3.2 – Experiment's Stimuli

As outlined in Chapter 3 (see section 3.2.2), the stimuli were split into two groups: **Distractors** (i.e. 524 unknown faces) and **Critical** images. The Critical group was further split into 3 categories: **Target** image (a single face that became task-relevant), **Irrelevant** images (unknown faces of Lecturers from another University) and **Probe** images (familiar Lecturer faces who are well known to the subject). Just as we had done for the second (concealed lecturer faces) experiment, we assigned three Lecturers (as Probes) to each subject, knowing that they were highly familiar with one another (as confirmed by the Lecturers and/or the subject's colleagues). Additionally, each subject was randomly assigned three unknown lecturers (as Irrelevants) from a different University (i.e. Canterbury Christ Church University), whose photographs were taken with the same camera, and treated in the same manner as all the Probe images (for detailed explanation of the standards/methods used to take the photographs and edit them, please refer to section 3.2.2 – *Stimuli*).

7.3.3 – Experiment's design

As described in section 3.2.3, each RSVP stream's 18 faces included a single Critical stimulus and 17 Distractors (with an SOA of 133ms). The Critical stimuli in each RSVP stream could either be a Probe (i.e. one of 3 familiar lecturer faces), or an Irrelevant (i.e. one of 3 unknown lecturer faces), or the Target (i.e. a single face that is task-relevant).

In total, Probes, Irrelevants and Targets were presented an equal number of times, and (in a statistical sense) in the same position in streams. As explained in the Introduction, the primary change between this (the third) experiment and the previous (the second) experiment was that the existence of Probes (i.e. familiar lecturer faces) were revealed, thus, in this experiment, we instructed subjects to expect seeing familiar lecturer faces. Although, of course, we did not name any particular lecturer or present their face, in advance of the experiment. Irrelevants remained the same (i.e. unknown Lecturer images) and the Target (i.e. an unknown Distractor face) continued to be task-

relevant. Additionally, according to the design improvements outlined in our ground-truth data simulations (see chapter 6), we split the experiment into two parts: Part I contained three blocks (i.e. a similar set-up to the previous, concealed lecturer faces, experiment), and after a short break, which enabled us to perform an online statistical test, Part II continued with a further two blocks. During the break (lasting approx. 5 minutes), the subject rested and we automatically processed Part I's data, to reveal the Probe with the highest significance. The results of this online test would determine the Probe/Irrelevant pair that could be used in Part II of the experiment.

7.3.3.1 – Part I of the new design

Part I of the experiment would replicate the entire design of the previous (concealed lecturer faces) experiment, which contained 3 Probes, 3 Irrelevants and a single Target. Each Probe was repeated 25 times, resulting in 75 Probe-trials (i.e. 25 times for each of the 3 Probes), and each Irrelevant was also repeated 25 times, resulting in 75 Irrelevant-trials (i.e. 25 times for each of the 3 Irrelevants). The single Target was, therefore, repeated 75 times, to equal the number of times that the other two Critical Stimuli category were included in RSVP streams. The resultant 225 RSVP trials were divided into 3 blocks, each block comprising 75 trials (i.e. 25 Probe trials, 25 Irrelevant trials, and 25 Target trials), and the order of the three Critical stimuli were randomised within the blocks. However, each block's Probe and Irrelevant Critical stimuli were paired, so that the same known lecturer (Probe) faces and unknown lecturer (Irrelevant) faces were presented within the same block – as with previous experiments, this will enable us to make direct comparisons between these paired-conditions.

As before, subjects were told to keep their eyes fixed at the centre of the screen during the presentation of the RSVP stream (lasting approx. 2.5 seconds), and to avoid movement or blinking. Finally, they were instructed to look for the Target image (i.e. the same task-relevant instructions as the previous two experiments), and informed that the Target image will appear pseudo-randomly, so they should not expect it in every trial, however, in this experiment, subjects were informed that they may see familiar lecturer faces (i.e. Probes that are well known to them).

7.3.3.2 – Part II of the new design

Having completed Part I of the experiment and performed an online statistical test on the difference between the Probe and Irrelevant conditions (i.e. a randomisation test on the mean amplitude measures of the highest positive 100ms window, as defined by the aERPt method), the Probe with the lowest *p*-value was chosen as the single Probe (plus its Irrelevant-pair and the Target, which remained the same) for Part II. Hence, the experiment continued, once subjects were given the same instructions as Part I, albeit, they were informed that the second part consisted of two blocks only (instead of three blocks in Part I). Consequently, in Part II, the inferred Probe was repeated 25 times, resulting in 50 Probe-trials (i.e. 25 times in each of the 2 blocks), and its paired Irrelevant was also repeated 25 times, resulting in 50 Irrelevant-trials (i.e. 25 times in each of the 2 blocks). The single Target was, therefore, repeated 50 times, to equal the number of times that the other two Critical Stimuli category were included in RSVP streams. Finally, the resultant 150 RSVP trials were divided into 2 blocks, each block comprising 75 trials (i.e. 25 Probe trials, 25 Irrelevant trials, and 25 Target trials), and the order of the Critical stimuli were randomised within the blocks.

7.3.3.3 – Combining Parts I and II

In Part I of this experiment, out of a total of 75 trials for each Critical condition, the number of trials that remained after artefact rejection, per condition, ranged between 54 and 75: *Target (M = 71.29, SD = 6.13); Probe (M = 71.57, SD = 4.55); Irrelevant (M = 72, SD = 5.46).* In Part II, out of a total of 50 trials for each Critical condition, the number of trials that remained ranged between 33 and 50: *Target (M = 47.29, SD = 4.68); Probe (M = 48.5, SD = 2.71); Irrelevant (M = 48.07, SD = 2.59).* Having combined Parts I and II, out of a total of 125 trials for each Critical condition (i.e. 75 trials in Part I and 50 trials in Part II), the number of trials that remained after artefact rejection, per condition, ranged between 87 and 125, and none of the subjects were excluded from the analysis due to removal of artefact trials:

Target (M = 118.57, SD = 10.63);
Probe (M = 120.07, SD = 7.08);
Irrelevant (M = 120.07, SD = 7.79).

7.3.4 – Experiment's Target Questions

As explained in Chapter 3, at the end of each RSVP stream, the subject was required to answer two question (see section 3.2.3), using a dedicated keypad, which was placed under the subject's right/left hand (whichever hand the subject preferred to use). The first question related to the finishing-item, and the second question related to Target-recognition. Both questions required the subject to select either key '1' or '2' (note that in previous experiments, the second question could be answered using keys 4-or-5). The reason for using the same key combination for both questions was because we wanted to avoid the previously observed head movements, which (occasionally) occurred when the subject adjusted his/her fingers from 1 & 2 to 4 & 5 keys.

Before starting the experiment, the subject was shown the Target image – this would be the same image for all subjects – which was chosen from the Distractor (i.e. unknown) database, and, therefore, not familiar to the subject. Even so, the subject was asked, in the beginning, if they had ever seen, or could recognise, the Target face (none of our subjects had ever seen the Target face). As this is a task-based experiment, the subject was instructed to look only for that Target image, in each of the RSVP streams, and to expect a recognition question: "Did you see the Target face?", at the end of each trial (noting that this recognition question followed the finishing-item question). If the Target image was seen, the subject was instructed to answer 'Yes' (using '1' key), or 'No' if it was not perceived (using '2' key). If the Target was present, a 'Yes' (i.e. correct) answer would be a "HIT", and a 'No' (i.e. incorrect) answer would be a "MISS". Conversely, if the Target was absent, a 'Yes'

Target	HIT (125)	FP (250)
Subject 1:	121	12
	96.8%	4.8%
Subject 2:	115	6
	92.0%	2.4%
Subject 3:	119	5
	95.2%	2.0%
Subject 4:	99	25
	79.2%	10.0%
Subject 6:	98	12
	78.4%	4.8%
Subject 7:	122	6
	97.6%	2.4%
Subject 8:	119	24
	95.2%	9.6%
Subject 9:	113	21
	90.4%	8.4%
Subject 10:	104	7
	83.2%	2.8%
Subject 11:	92	11
	73.6%	4.4%
Subject 12:	103	9
	82.4%	3.6%
Subject 13:	84	10
	67.2%	4.0%
Subject 14:	107	8
	85.6%	3.2%
Subject 15:	117	19
	93.6%	7.6%
Mean:	108.07	*12.50*
	86.5%	5.0%
SD:	11.82	6.88

Table 7.1 – Subjects' HIT count (i.e. number of times that the subject correctly reported seeing the task-relevant Target face, in 125 trials), and False-Positive (FP) count (i.e. reported seeing the Target when it was not there, in the other 250 trials) are shown.

Group HIT rate of 108.07 (86.5%) and FP rate of 12.5 (5%), with corresponding MISS rate of 16.93 (13.5%) and correct rejection rate of 237.5 (95%), result in a response sensitivity measure of d' = 2.75.

(i.e. incorrect) answer would be a False-Positive (FP), and a 'No' would be a correct rejection (see Table 7.1).

Out of 125 times that each subject was randomly presented with the Target face, the average Hit rate for the group was 86.5% (M = 108.07, SD = 11.82), and out of the remaining 250 other trials in which the Target was not presented, the False-Positive rate was 5% (M = 12.5, SD = 6.88). The resulting sensitivity measure (d' = 2.7461) was within our tolerance range, and no subjects were excluded due to low sensitivity or high bias.

7.3.5 – Experiment's Probe/Irrelevant Questions

Just as we had done in the previous (concealed lecturer faces) experiment, the familiarity questions were asked at the end of the experiment (i.e. after both Parts I and II were completed). Even though we had advised subjects that the RSVP streams may contain familiar lecturer faces (i.e. the key change between this and the previous experiment), we did not ask the familiarity questions at the end of Part I (i.e. during the short break) because the subject may not have perceived any of the lecturer faces (i.e. Probes) in the first three blocks of the experiment. Therefore, by asking the familiarity questions at the end of the experiment, we could avoid the unintentional revealing of which lecturer faces had been included in Part I's three blocks, and, thus, may be repeated in Part II's remaining two blocks.

So, at the end of the experiment, the subject was given a recognition test, in the form of memory questions, to determine if the 3 Probes and/or the 3 Irrelevants were perceived/recognised. Note that even though there were five blocks in the entire experiment, one pair of Probe/Irrelevant conditions that were used in Part I of the experiment would be independently selected (using online statistical tests, during the break), and re-used in both blocks of Part II. Thus, the total number of Probes and Irrelevants for the entire experiment remained the same as the previous (concealed lecturer faces) experiment (i.e. 3 Probes and 3 Irrelevants). However, it must also be

noted that the chosen Probe/Irrelevant pair that were inferred and used in Part II would have been presented three times more than the other two Probe/Irrelevant pairs that were only used in Part I of the experiment. As demonstrated in chapter 6, our new statistical analysis method of combining all blocks will ensure that the false-positive and false-negative rates are not inflated.

The end of experiment memory-test consisted of 12 questions, appearing randomly, where each question accompanied an image that may or may not have been included in the experiment's five blocks. Six questions related to the presence of the paired Probe (familiar lecturer) and Irrelevant (unknown lecturer) faces that were included in the experiment, and the other six questions related to random Probe and Irrelevant faces that were not included in the experiment. Whereas the former six questions (about the Probe/Irrelevant faces that were presented) would gauge the subject's ability to perceive faces that were included in the experiment, the latter six questions assess the subject's engagement with the tests (i.e. were subjects guessing the presence of salient faces?).

As with the previous two experiments, the response to each of these 12 recognition/memory tests were handled in two parts: firstly, what is the subject's confidence rating of how often each face was presented (i.e. the Probe/Irrelevant that was present, and the Probe/Irrelevant that was absent), and secondly, a confidence rating of how well the subject knew each of the 12 faces, prior to the experiment. The responses to both of these confidence ratings used a scale of 1 to 5, where 1 is "Never", 2 is "Once or twice", 3 is "Few times", 4 is "Many times" and 5 is "A lot". Note, for the purposes of statistical comparison, 1 out of 5 (i.e. Never) is equivalent to 0% and 5 out of 5 (i.e. A lot) is equivalent to 100%. Thus, 2 out 5 = 25%, 3 out of 5 = 50% and 4 out of 5 = 75% (see Appendix C.1 for full results).

7.3.5.1 – Overall Probe/Irrelevant recognition

The three Probe (familiar-lecturer) faces that were included in the experiment were reported to have been seen 50% of the time (Mean confidence rating of 3 out of 5), and subjects reported a high (pre-experimental) familiarity of 79.2% with these Lecturer

faces (4.2 out of 5). When comparing this to the (absent) Probe faces that were not included in the experiment, subjects reported a similar high (pre-experimental) familiarity of 83.3% (4.3 out of 5), and only reported seeing these 'absent' familiar-lecturers 8.3% of the time (1.3 out of 5), which is almost one-tenth of the Lecturers that were actually included in the experiment.

The three Irrelevant (unknown-lecturer) faces that were included in the experiment were reported to have been seen 8.9% of the time (1.4 out of 5), and the absent Irrelevant faces that were not included in the experiment were reported to have been seen, almost the same rate of 6% of the time (1.2 out of 5). Finally, subjects reported an imperceptible (pre-experimental) familiarity of 0% with all the Irrelevant/unknown-lecturer faces (1.0 out of 5).

As we were comparing Probe (familiar lecturer) faces with Irrelevant (unknown lecturer) faces, it was encouraging to discover that Probes were reported 50% of the time ($M = 3.0$; $SD = 0.6918$), which was nearly six times more than Irrelevants that were reported 8.9% of the time ($M = 1.4$; $SD = 0.4022$). Note that both conditions (Probes and Irrelevants) were, in fact, presented an equal number of times. The mean confidence rating of the main comparison conditions, for all subjects, reveals a highly significant difference between the Probe (familiar lecturer) faces and the Irrelevant (unknown lecturer) faces, using pair-wise comparison ($M = 1.6571$, $SD = 0.8582$), $t(13) = 7.2251$, $p < 0.0001$, $d = 2.9518$).

7.3.5.2 – By-item Probe recognition

As with the previous (concealed lecturer faces) experiment, we were required to match subjects to their most familiar lecturers (i.e. the Probes with the highest familiarity), therefore, in this (revealed lecturer faces) experiment, also, the three Probe (familiar-lecturer) faces that were chosen for each subject could be different. As a result, a 'by-item' comparison would not show the response to the same lecturers (i.e. if we were to consider the first block of all 14 subjects, we may find that 14 different lecturers were chosen as Probes). However, we carried out by-item comparisons between the 3 lecturers that were assigned to all subjects (noting that Part II of the experiment re-used

the same Probe/Irrelevant pairing that achieved the highest statistical significant in Part I), in order to quantify the group-level response to the Probe recognition/memory tests (see Figure 7.1, below).

Bearing in mind that different lecturers may be involved in each block, recognition results for the three lecturers (57.1%, 46.4% and 46.4% respectively), in the form of a one-way ANOVA, confirms that there is no statistically significant difference between them ($F(2, 39) = 0.3205$, $p = 0.7276$). Note that one of the Probes (and its paired Irrelevant) appeared more times than the other two because it was re-used in Part II of the experiment (see Appendix C.2 for a clarification of which Probe was re-used in Part II, for each subject).

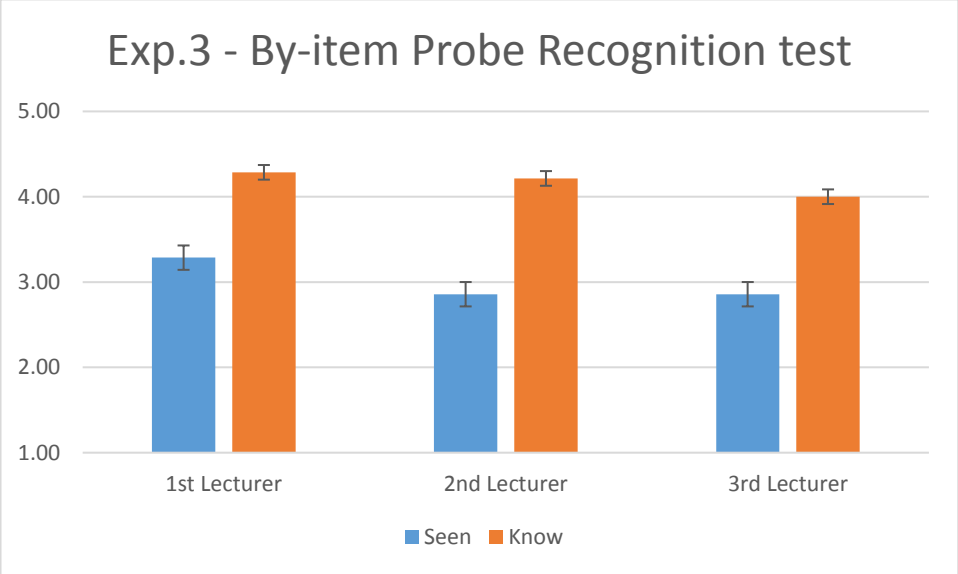

Figure 7.1 – Experiment 3's by-item Probe (familiar-lecturer face) recognition tests: "Seen" rates (i.e. confidence rating of having detected the Probe) and "Know" rates (i.e. how well the subject recognises the Probe). On average, 57.1% of subjects had seen the first-lecturer (rating = 3.3, SE:0.5), 46.4% had seen the second-lecturer (rating = 2.9; SE: 0.4), and 46.4% had seen the third-lecturer (rating = 2.9; SE: 0.4). One-way ANOVA on the 'seen' ratings for the three lecturers confirms that there is no statistically significant difference between the means (p = 0.7276). As expected, subjects' familiarity (i.e. 'Know' ratings) with all three presented lecturers was very high (see Appendix C.1 for more detail).

As with the previous two experiments, the behavioural data (i.e. all the above online responses to recognition questions) provided a useful indicator of the perceptual state of the subjects' mind, however, the primary aim of our research was to use the EEG data to detect the breakthrough of Probe (familiar lecturer) faces, which could be

differentially perceived and processed, as compared to Irrelevant (unknown lecturer) faces. Therefore, the rest of this chapter will focus on the analysis of the EEG data, in the Time and Frequency domains.

7.4 Data Analyses

7.4.1 – Summary of Analysis

In keeping with the previous two experiments, we remained interested in the EEG data across all the midline electrodes (Pz, Cz and Fz), but in-line with (Kaufmann, Schulz, Grünzinger, & Kübler, 2011) and our own findings, we expect the strongest brain responses to familiar faces, to be recorded at the Pz electrode. Once again, we will start by focusing on the Pz electrode's Time and Frequency domain analyses (at group & subject level), before reporting the same analyses at Fz and Cz.

7.4.1.1 – Pz Electrode

At group-level, the grand average ERPs of all three critical stimuli (i.e. the Target, Irrelevant and Probe conditions), at the Pz electrode site, revealed a clear difference between the conditions (see Figure 7.2, below). The Target condition was task-relevant, so it elicited a large classical P3, which was as expected because subjects were instructed to detect the Target face. The Irrelevant condition, which consisted of an unknown face (paired with each Probe, and repeated randomly, as many times as the Probe), did not present a similar pattern to the Probe or Target; this was as expected because non-salient stimuli were unlikely to breakthrough into conscious awareness, due to the high presentation rate of the RSVP streams. Just like the previous two experiments, the Probe condition elicited a continuous oscillatory pattern, within a 280 to 620ms time frame (and observed frequency of approximately 3-4 Hz).

In the previous (*concealed* lecturer faces) experiment, we confirmed our hypothesis that there is a large difference between the Probe (familiar lecturer face) and

the Irrelevant (unknown lecturer face) conditions, and in the current (*revealed* lecturer faces) experiment, we expected a similar effect – this was confirmed in the group-level ERPs (Figure 7.2), and an oscillatory pattern for the Probe (lecturer) faces was observed. Interestingly, this oscillatory wave was a closer match to the pattern that we observed in the first (*celebrity* faces) experiment, albeit, the peak negativity of the N400f component was, in fact, similar to the second (*concealed* lecturer faces) experiment (noting that both of these experiments' N400f components appeared to be half the size of the first experiment's N400f).

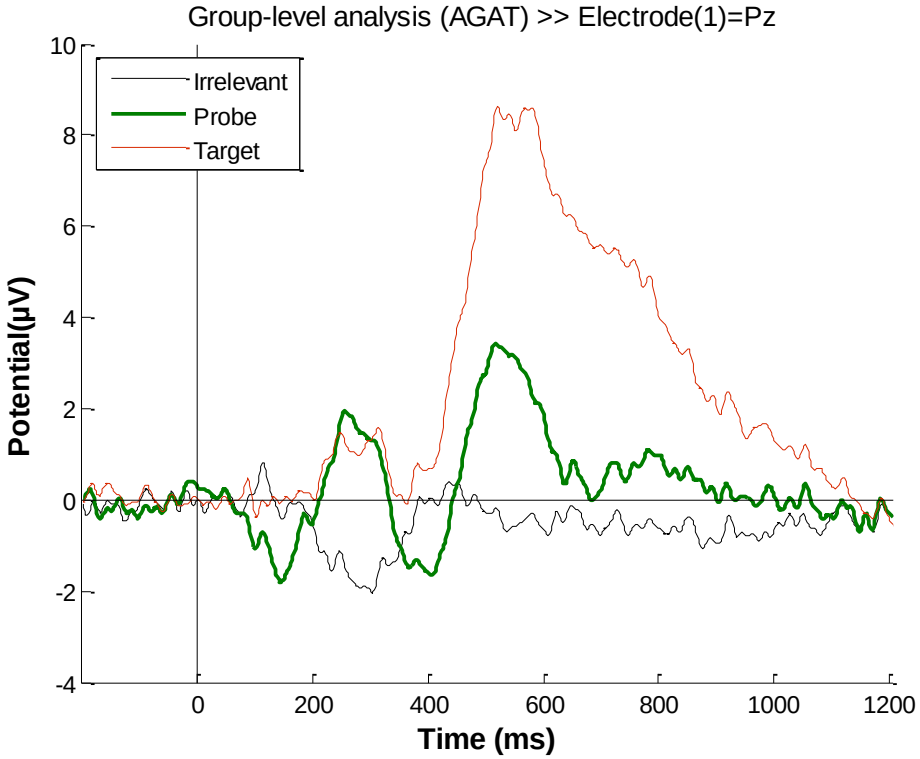

*Figure 7.2 – Grand average ERPs elicited by the three critical stimuli (Irrelevant, Probe and Target conditions), at Pz electrode, once again, showing a P3 pattern for the **Target** (in red, peaking at +8.5μV), an oscillatory pattern for the **Probe** (in green, with an observed frequency of approx. 3-4 Hz), and a different pattern for the **Irrelevant** (in black, containing an interesting negative deflection, peaking at 300ms) that is distinct from the Probe and Traget. Target was the stimulus that the subject was instructed to look for, whilst subjects knew that familiar faces of lecturers (Probe) may also be present. The oscillatory pattern for the Probe suggests a significant difference with the Irrelevants (unknown lecturer faces), which were presented as many times as the Probe. Whilst the Probe's oscillatory pattern, and the peak positivity around 500ms (i.e. P600f) is similar to the first (celebrity faces) experiment, the Probe's peak negativity around 400ms (i.e. N400f) is considerably smaller in amplitude.*

In keeping with the previous two experiments, by collating and stacking all Target trials, for all subjects in the group, we observed a prevailing positivity, from 400ms onwards, for most trials, at the Pz channel. This channel-oriented representation of the trials was confirmed by the aggregated ERPs (see left plot of Figure 7.3), and the spatial dispersion of resultant waveform was depicted by the ERP scalp topographies (see right plot of Figure 7.3), which confirmed the Target condition's dominant positive wave, peaking at around 550ms.

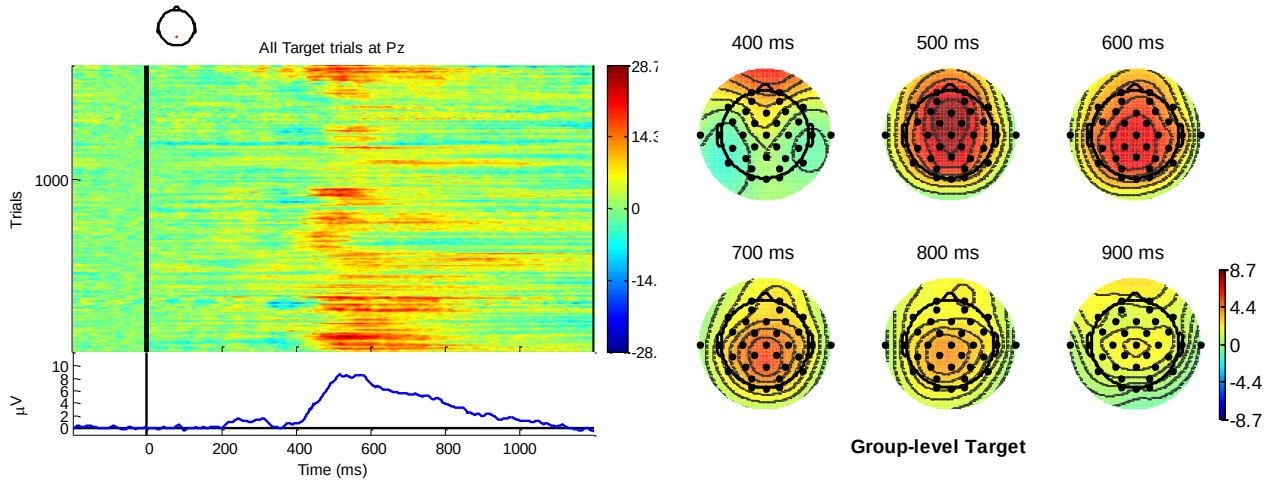

Figure 7.3 – Group-level view of all (1660) Target trials, in order of appearance over time, at Pz (left plot), and the corresponding scalp topography of the ERPs (right plot), showing a prevailing positivity, peaking at around 550ms, with the electrical field moving posteriorly through time. Unlike the previous two experiments that only used 8 electrodes, this experiment used 32 electrodes to increase the coverage. Even so, it must be noted that MATLAB employs an interpolatory algorithm to represent the full scalp pattern, therefore, estimated electric potential values are used at scalp locations between the actual recording sites. Note that the scalp map scale ranges from -8.7 to +8.7 μV.

Next, we stacked all the trials in the Probe condition, for all subjects at Pz, and observed the oscillatory waveform, with its peak negativity at around 400ms, and its peak positivity at around 500ms. However, in comparison with the first (celebrity faces) experiment, the current experiment's Probe possessed a weaker N400f feature, albeit, it was closer to the second (concealed lecturer faces) experiment. The Pz channel ERP image and the interpolatory scalp topography of the ERPs can be seen in Figure 7.4 (below).

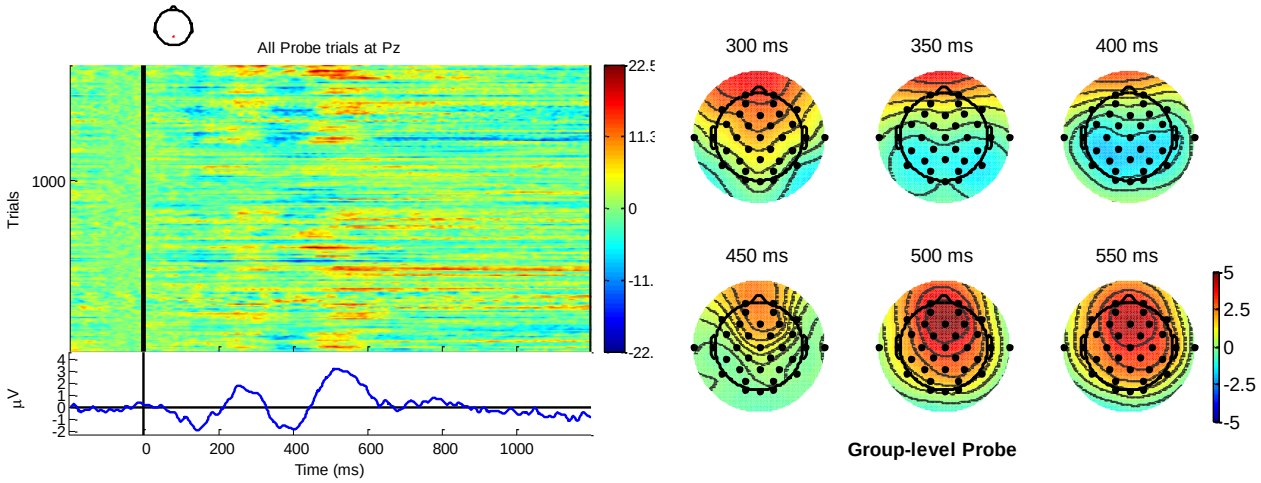

Figure 7.4 – Group-level view of all (1681) Probe trials, over time, at Pz (left plot) and interpolated scalp topography of the ERPs (right plot), showing a similar oscillatory pattern to the previous two experiments. The N400f effect is weaker than the first (celebrity faces) experiment, but stronger than the second (concealed lecturer faces) experiment. Lowest negativity can be observed at 400ms, and highest positivity at 550ms, with an oscillatory switching from frontal positivity (300ms) to posterior negativity (400ms). Note that the scalp map scale ranges from -5 to +5 μV, which is lower than the scale for the Target (see Figure 7.3).

As shown in Figure 7.5 (below), the Irrelevant condition did not show a similar oscillatory pattern to that observed in the Probe or Target conditions, supporting our hypothesis that unknown lecturer faces, presented at a rapid rate, will not breakthrough into conscious awareness. Albeit, the unusual posterior negativity (300ms), followed by a frontal positivity (450ms) may indicate a covert response to repetition.

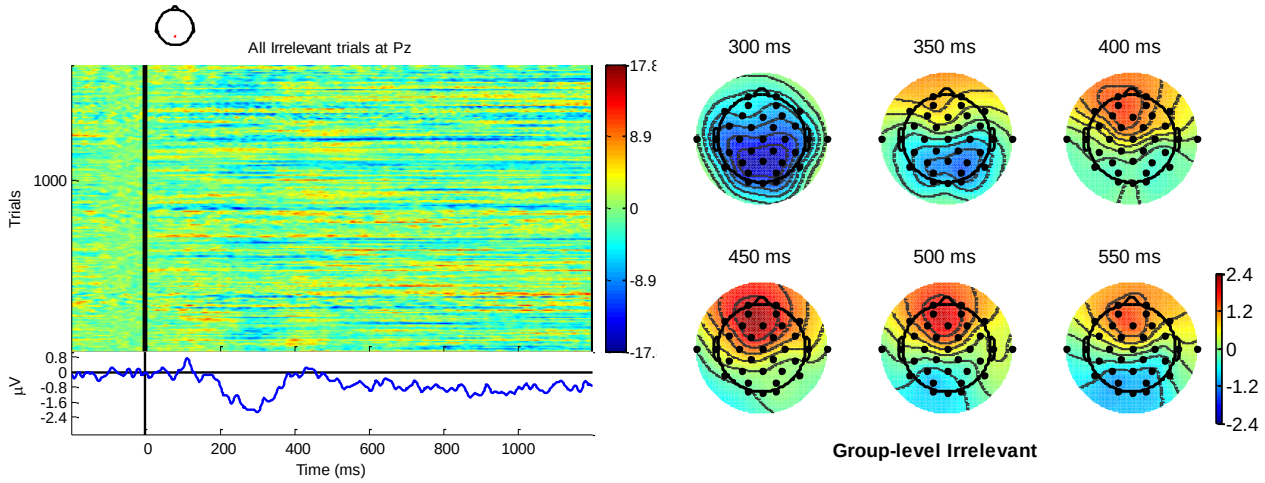

Figure 7.5 – Group-level view of all (1687) Irrelevant trials, over time at Pz (left side), and the interpolated scalp map representation of the ERPs (right plot), showing a posterior negativity (300ms) and a later frontal positivity (450ms). Note that the scalp map scale ranges from -2.4 to +2.4 μV, which is lower than the scale for the Probe condition (see Figure 7.4).

Even though the oscillatory pattern for the Probe (and the Target) was different to the Irrelevant, the existence of the Irrelevant's posterior negativity (peaking at 300ms) presents an interesting finding, since participants did not report seeing them (refer to section 7.3.5.1, where recognition results for Irrelevants that were included in the experiment was 8.9%, which was only slightly higher than the Irrelevants that were not included in the experiment, at 6%). Could this discovery – which was observed for the first time, in the previous (concealed lecturer faces) experiment – be related to subliminal registering (i.e. a covert response or threshold awareness) of a repetition by the brain, or some other incongruity? *We shall expand on these ideas in section 7.5*.

As with the previous two experiments, the main comparison was between the Probe and Irrelevant conditions, and our statistical tests showed a highly significant difference between them. Having aggregated all Probe and Irrelevant trials for all subjects, we employed the AGAT method, for orthogonal contrast time window placement (i.e. to independently find the most extreme 100ms mean amplitude interval) for highest positivity (P600f) components. Similarly, we used the AGAT method to independently find the lowest 100ms mean amplitude, within the N400f time-frame, but due to the non-typical negativity in the Irrelevant condition (peaking at 300ms, and partly overlapping the Probe, as seen in Figure 7.6, below), the results of our statistical tests on N400f were not significant. Whilst the oscillatory pattern for this experiment's Probe showed a strong similarity to the Probe condition in the previous experiments, we noted that, with the exception of the weaker N400f effect, it was a closer match to the first (celebrity faces) experiment. Conversely, we observed that this experiment's unexpected early negativity for the Irrelevant resembled a similar effect in the previous lecturer faces experiment (*for theories on these differences, see section 7.5.4*).

The null hypothesis (H0) was that there is no difference between the Probe and Irrelevant patterns, for the group. Our experimental hypothesis is that H0 can be rejected, at the group-level. As detailed in section 3.3.3.2 (*Group-level (AGAT) window placement*), a paired *t*-test of the mean amplitudes of Probe P600f/N400f and Irrelevant P600f/N400f was used, across all participants, to calculate the group's *p*-values (compared to a critical alpha level of 0.05), and to determine the probability that the observed pattern could have arisen if the null hypothesis were true. This is a reliable way to determine the group's familiarity with the Probe (lecturer) faces. Note that

having justified the use of an independent detrending technique (see section 4.4.2.3), all the following statistical analyses will incorporate this method of removing any unwanted drift in the signal. Furthermore, detrending will always take place before baseline correction.

7.4.2 – Group-level Analysis, at Pz

For group-level analysis, the AGAT orthogonal contrast method enabled us to perform statistical tests, using a critical-alpha level of 0.05. As explained earlier (see section 3.3.3.2), aggregating all Probe and Irrelevant trials for all subjects, and employing the AGAT method, for orthogonal contrast time window placement, would enable us to independently identify the highest 100ms mean amplitude interval for the highest positivity (P600f), as shown in Figure 7.6 (highlighted in yellow).

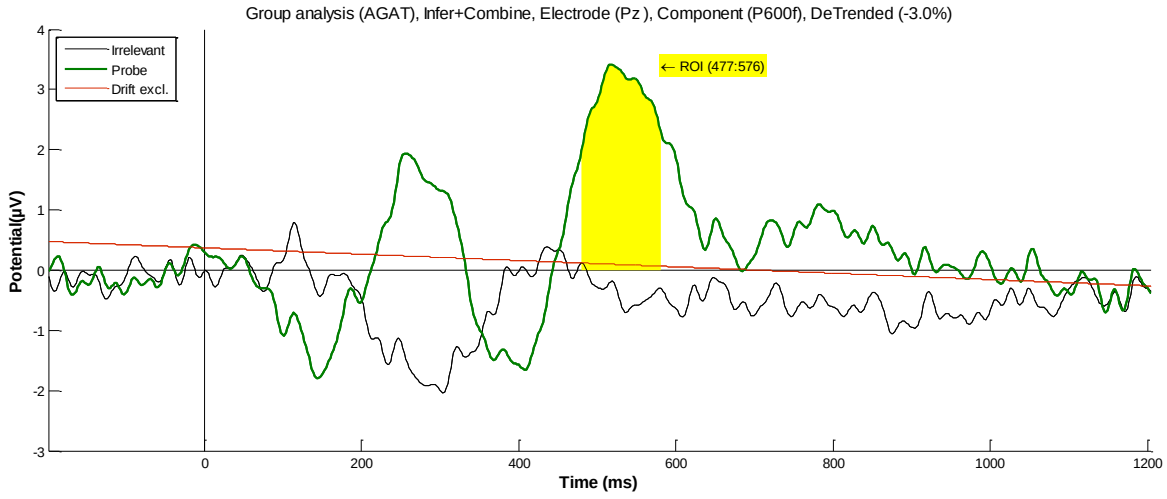

*Figure 7.6 – Grand average ERPs elicited by Probe and Irrelevant (i.e. the main comparison conditions) at **Pz**, showing an oscillatory pattern for the Probe condition (in green), which does not exist for the Irrelevant condition (in black). The linear Drift has been excluded (in red, at -3% to the vertical), with a detrending method. Even though subjects were not informed of the presence of the Probe (familiar lecturer face), statistical tests show a highly significant difference between Probe and Irrelevant, for **P600f** (t(13) = 4.6121, p = 0.0004, d' = 1.8923). However, the same statistical test on N400f was not significant (t(13) = 1.0474, p = 0.314, d' = -0.388).*

Within the a-priori P600f time-frame (i.e. 300ms to 900ms), the AGAT method independently identified an orthogonal contrast 100ms time window, at 477 to 576ms (see Figure 7.6), and our statistical tests produced a highly significant difference between the Probe (M = 2.9653, SD = 2.3258) and Irrelevant (M = -0.3899, SD = 0.9371), at the Pz electrode site (M = 3.3552, SD = 2.7219), $t(13) = 4.6121$, $p = 0.0004$, $d' = 1.8923$.

As explained earlier, the non-typical negativity in the Irrelevant condition (partly overlapping the Probe and peaking at 300ms), which was similar to the Irrelevant in the previous (concealed lecturer faces) experiment, meant that statistical tests, at group-level, on the N400f component did not result in a significant difference between the two conditions of comparison: (M = -0.5951, SD = 2.126), $t(13) = 1.0474$, $p = 0.314$, $d' = -0.3882$. Furthermore, subject-level statistical tests (i.e. Monte Carlo permutation) on the N400f component confirmed that only 5 of 14 subjects (35.7%) showed critical-significance (subjects 1, 8, 9, 12 and 14) between Probe and Irrelevant (Mean p-value = 0.4126, SD = 0.4107). As a result, the following section will mainly focus on the analysis of the P600f component.

7.4.3 – Subject-level Analysis

As before, our goal was to statistically analyse the data at the Pz electrode site (Kaufmann, Schulz, Grünzinger, & Kübler, 2011), and the main comparison was between the Probe and Irrelevant conditions (Bowman, et al., 2013), at subject-level. Statistical analyses of the ERP data will determine whether the elicited response by the Probe (i.e. familiar lecturer face) was significantly different from that elicited by the Irrelevant (i.e. unknown lecturer face). As outlined in section 3.3.3 (*Time Domain (ERP) Analysis*), subject-level analysis is based on analysing each experimental participant separately, to determine whether there was a significant difference (i.e. did the subject's brain response reveal a differential perception and processing of the lecturer faces, as compared to the unknown faces?). The null hypothesis (H0) was that there is no difference between the subject's Probe and Irrelevant patterns. Our experimental hypothesis is that H0 can be rejected.

Having used the aERPt method to independently find the time window for each component of interest, a randomisation (i.e. Monte Carlo Permutation) test was used to define a *p*-value for each subject. Then, a null hypothesis distribution was generated in order to calculate the individual's *p*-value. This *p*-value would determine the probability that the observed pattern could have arisen if the null hypothesis were true. This is a reliable way to assess each subject's pattern individually, and to determine that subject's familiarity with the Probe. As with the first (celebrity faces) and second (concealed lecturer faces) experiments, we theorised that the results of our statistical analysis would infer the subject's conscious and/or unconscious (i.e. sub/liminal) detection of familiar lecturer faces.

As outlined in the Introduction of this chapter (*and detailed in Chapter 6*), an added dimension in the design of this experiment was the method by which we used Part I to qualify the three Probes (i.e. with the aid of online statistical tests, we identified which Probe achieved the highest significance in Part I), and then re-used the chosen Probe/Irrelevant pair in Part II of the experiment. As demonstrated in Chapter 6, combining Parts I and II would enable us to raise the SNR, whilst ensuring that the false-positive and false-negative rates are not inflated. Having justified the use of the Combined (C) analysis method, we performed three additional analyses – which we named: Abandoned (A), Biased (B) and Decider (D) methods – for comparison purposes (see section 7.4.3.3 – *Alternative Methods of Analysis*).

7.4.3.1 – Synopsis of results

As shown in table 7.2, subject-level statistical tests of Pz electrode's P600f component, resulted in 11 of 14 subjects (78.6%) achieving critical-significance (alpha level 0.05), between the Probe and Irrelevant conditions. Note that this is an improvement on both of the previous experiments' results (to wit: the celebrity faces experiments achieved 7 of 14 subjects, and the concelaed lecturer faces achieved 8 of 14 subjects, on the P600f contrast).

Subject	Probe (*M*)	Irrelevant (*M*)	*p*-value
1	5.7219	0.0468	< 0.0001
2	4.1088	0.7413	< 0.0001
3	1.2010	-1.5738	0.047
4	3.5180	0.0985	0.002
6	1.8596	0.5753	0.338
7	5.0795	0.1546	< 0.0001
8	2.9328	-0.3709	0.001
9	-1.1494	0.4338	0.885
10	2.4216	-0.8157	0.004
11	0.3570	0.9624	0.728
12	5.0769	-0.2530	< 0.0001
13	4.7381	-0.5363	< 0.0001
14	3.2380	-1.2884	< 0.0001
15	7.4658	-1.0403	< 0.0001

*Table 7.2 – Subject-level analysis, at **Pz** electrode, for the P600f component, showing the mean amplitude values of the Probe and Irrelevant conditions, from the same 100ms time window, which was independently found using the aERPt method (significant p-values shown in green). Out of 14 subjects, eleven (78.6%) achieved critical-significance, which is the highest result in all 3 experiments, for the P600f.*

7.4.3.2 – Individual's P600f, by-item and by-subject

At the Pz electrode site, we began by exploring the presence of the P600f component within each of the three items of every subject (i.e. five experimental blocks for 14 subjects, equalling 70 item-blocks). Having independently searched for the P600f component's 100ms aERPt time window (i.e. highest positive deflection, within the a-priori search area that spans from the time range of 300ms to 900ms), we performed permutation tests for each individual block (see Appendix C.2 for full details).

Consequently, five 'by-item' *p*-values were obtained for each subject (i.e. one for each block's familiar lecturer, noting that the last 2 blocks re-used a familiar lecturer that was also used in an earlier block), resulting in significant difference between the Probe and Irrelevant conditions for 33 of 70 blocks (47.1%), which is higher than both of the previous two experiments (to wit: the celebrity faces experiment's by-item subject analysis achieved 15.7% significance, and the concealed lecturer faces experiment achieved 26.2%). Note that, as the number of trials-per-block remained the same as the previous experiment, one reason for the improvement in *by-item* results could be due to the revelatory instruction (given at the start of the experiment), in which subjects were informed that familiar lecturers may be included in the RSVP streams.

Whilst block-level/by-item results were interesting, our main enquiry was the significance at subject-level, for the P600f component. Therefore, so we combined each of the three conditions' trials and performed statistical tests on every subject (see Table 7.2, above), resulting in a significant difference between the Probe and Irrelevant conditions for 11 of 14 subjects (78.6%), which is more than the previous two experiments. However, we could not use the N400f component (& the Fisher combining method) to enhance the significance, at the subject-level.

As shown in Figure 7.7, nearly all subjects' Probes elicited a clear positive deflection, within the aERPt identified highest positive 100ms time window (highlighted in yellow), of the P600f time-frame (300 to 900ms). However, relative to the Irrelevant (i.e. the condition of comparison), the Probe for two subjects (nos. 6 and 11) failed to show a significant positivity for P600f, and one subject (no. 9) failed because the independently searched aERPt method found an earlier-than-ideal time window (i.e. 520 to 620ms, instead of a better fit at 791 to 891); had the correct P600f been selected for subject 9, our statistical tests would have showed a highly significant difference between that subject's Probe and Irrelevant.

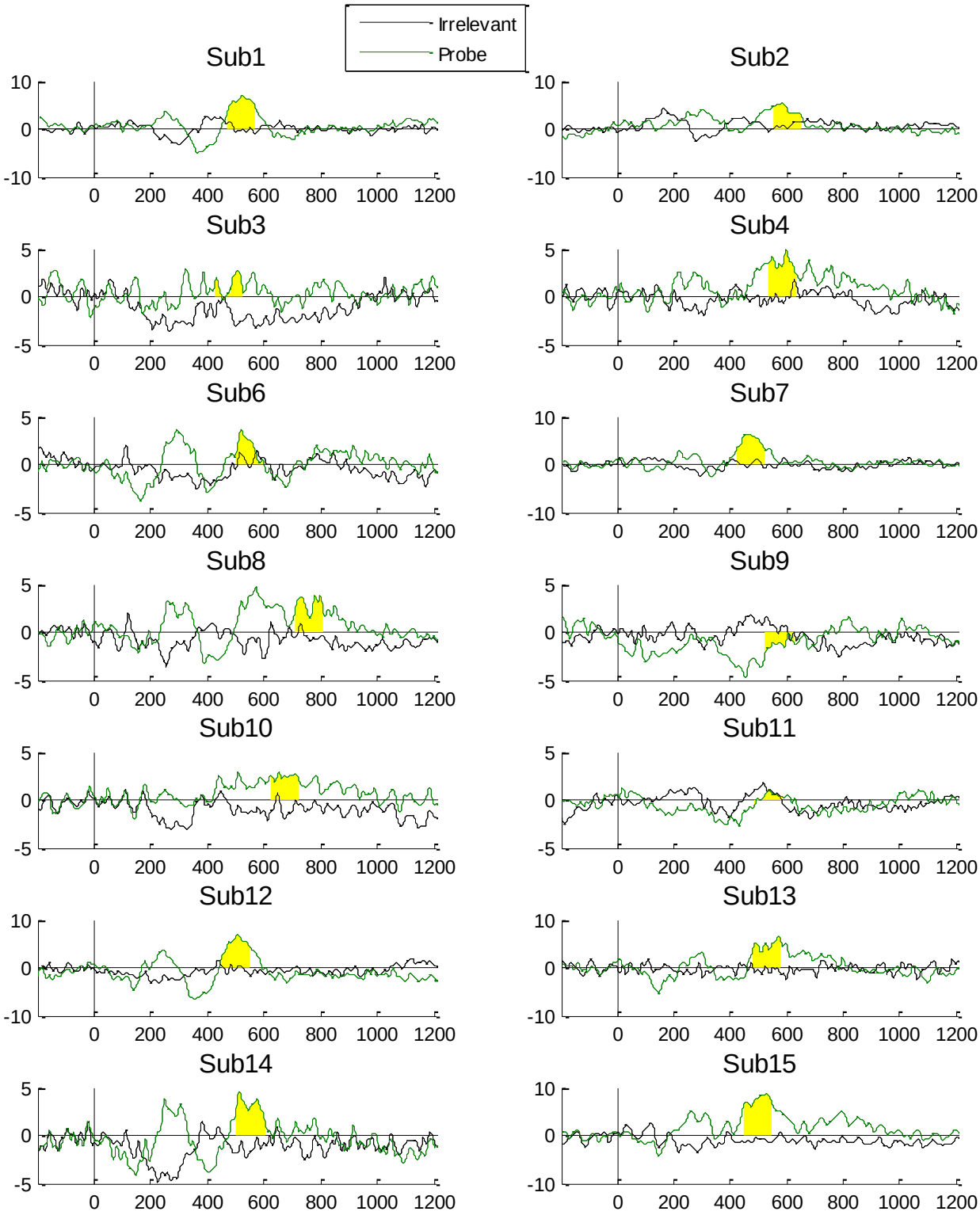

*Figure 7.7 –Subject-level Probe (in green) and Irrelevant (in black) ERPs, at the **Pz** electrode site (x-axis represents Time in miliseconds, and y-axis represents Potential in microvolts). Each ERP shows the orthogonally identified highest positive 100ms time window (yellow highlight) for **P600f** (using the aERPt method), where 11 of 14 subjects (78.6%) show a significant difference between the Probe and Irrelevant conditions which is the highest result in all three experiments, of this thesis.*

7.4.3.3 – Alternative Methods of Analysis

The conclusions of our ground-truth data simulations (see chapter 6) justified the adoption of the current experiment's two-part design, and the use of the *Combined* method of analysis. However, the application of the three alternative methods that were rejected (*Abandoned*, *Biased*, and *Decider*) may be of scientific interest to the reader. Therefore, after a brief introduction of these methods of analysis, we will apply them to the current experiment's data, as a means of comparison and contrast:

A) **Abandoned** method, where Part I is discarded because it acts as the decider for Part II only. Therefore, the Abandon method contains the last two blocks, which are in Part II of the experiment. In Chapter 6, we demonstrated that this method is safe, but due to having the least number of trials, the reduction in the statistical power may inflate type II (false-negative) errors.

B) **Biased** method, where the chosen block in Part I (i.e. the one with the lowest *p*-value, whose Probe/Irrelevant pair is re-used in Part II) joins both blocks in Part II. Therefore, the Biased method contains three blocks: one from Part I, plus the fourth and fifth blocks, which comprise Part II. Despite an improved SNR (compared to method A), this is an unsafe method, due to the inflated possibility of Type I (false-positive) errors.

C) **Combined** method, where all five blocks (3 in Part I and 2 in Part II) are joined together. This is our preferred method, which has the highest SNR (as it pertains to the largest number of trials) and is safe.

D) **Decider** method, where Part II is discarded and only the three blocks in Part I are used. This method is safe, but its SNR is similar to method B (albeit, better than method A). As we have noted earlier, the primary use of method D was to independently determine the paired conditions for Part II of each experiment, however, its secondary use was to make direct comparisons with the previous (Concealed Lecturer Faces) experiment, and to determine whether revealing the presence of lecturer faces would increase statistical power.

Having demonstrated that Method C (Combined) is the preferred technique, as it possesses the highest number of trials and is safe from Type I errors (see Chapter 6 for justification), we performed the same standard statistical tests on the other three methods, mainly, for comparison purposes (see table 7.3). As previously shown in table 7.2, using method C, subject-level statistical tests of Pz electrode's P600f component resulted in 11 of 14 subjects (78.6%) achieving critical-significance between the Probe and Irrelevant conditions. As predicted, the low SNR for methods A (Abandoned) and D (Decider) resulted in fewer significant subjects: 10 of 14 (71.4%) for the former and 9 of 14 (64.3%) for the latter. Whilst both methods A and D are safe (albeit, their SNR is low, due to fewer trials), we demonstrated that method B (Biased) raises the Type I error rate, which (unsurprisingly) resulted in the highest number of significant subjects: 13 of 14 (92.9%), with only subject 11 failing to achieve significance.

Subject	A. Abandoned	B. Biased	C. **Combined**	D. Decider
1	< 0.0001	< 0.0001	< 0.0001	< 0.0001
2	< 0.0001	< 0.0001	< 0.0001	0.235
3	0.657	0.012	0.047	0.008
4	0.004	< 0.0001	0.002	0.088
6	0.244	0.029	0.338	0.202
7	0.002	< 0.0001	< 0.0001	< 0.0001
8	< 0.0001	0.001	0.001	0.030
9	0.818	0.043	0.885	0.968
10	< 0.0001	< 0.0001	0.004	0.026
11	0.996	0.633	0.728	0.122
12	0.005	< 0.0001	< 0.0001	< 0.0001
13	< 0.0001	< 0.0001	< 0.0001	0.037
14	< 0.0001	< 0.0001	< 0.0001	0.046
15	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Significance	*10 of* 14	*13 of* 14	***11 of* 14**	*9 of* 14

Table 7.3 – Subject-level analysis (at Pz electrode, for the P600f component), showing p-values for 4 different methods that could be used to analyse the experiment (significant results are shown in green). In the Abandoned method (A) 10 of 14 (71.4%) achieved critical-significance; in the Biased method (B), 13 of 14 (92.9%); in the Combined method (C), 11 of 14 subjects (78.6%); and in the Decider method (D), 9 of 14 (64.3%).Whilst our preferred method C is safe and benefits from a high SNR, method B is note safe (i.e. it raises type I errors), and method A may inflate type II errors. However, methods A and D are both safe, but they possess lower SNR (albeit, despite method A having a lower SNR than method D, more subjects were shown to be significant, due to the inference of using the 'best block', in Part II of the experiment).

Whilst the number of subjects that were shown to be significant, using our method of choice (i.e. Combined: 11 of 14), was one higher than the Abandoned method (10 of 14), statistical tests showed that there was no significant difference between them, at the Pz electrode site (Difference = 7.2%) $X^2(1, N = 14) = 0.187$, $p = 0.6657$, CI_{95} [-23.92%, 36.72%].

Despite the Abandoned method (i.e. using two blocks in Part II only) having a lower SNR than the Decider method (i.e. using three blocks in Part I only), one more subject was shown to be significant using the Abandoned method, due to the inference of using the 'best block', in Part II of the experiment (i.e. the paired Probe/Irrelevant that were re-used in Part II were inferred by selecting the most significant condition in Part I). However, comparisons between Part I of the experiment (i.e. Abandoned: 10 of 14) and Part II (i.e. Decider: 9 of 14) did not show a significant difference between them either (Difference = 7.1%) $X^2(1, N = 14) = 0.156$, $p = 0.6929$, CI_{95} [-25.38%, 37.69%].

Finally, comparisons between subject-level results of the previous (*concealed* lecturer) experiment and Part I of the current (*revealed* lecturer) experiment were noteworthy because the only difference between the two experiments was the explicit instruction, given at the start (i.e. subjects were informed that familiar lecturer faces may appear in the latter, but they were naïve in the former). Despite an improvement in subject-level significance of the current experiment's Part I (i.e. 10 of 14 versus 8 of 14), statistical tests showed that there was no significant difference between them (Difference = 14.3%) $X^2(1, N = 14) = 0.601$, $p = 0.4382$, CI_{95} [-19.49%, 44.08%].

7.4.4 – Time Frequency Analysis (TFA)

As outlined in section 3.3.5 (*Frequency Domain Analysis*), to analyse the power and coherence of the EEG data, we have employed two Time Frequency transforms: Event-Related Spectral Perturbation (ERSP) and Inter-Trial Coherence (ITC), using EEGLAB's toolbox (Delorme & Makeig, 2004). Whereas ERSP reflects the extent to which the signal power changes in relation to a specific time period (i.e. the baselining window before stimulus-onset) at different frequencies in a signal, ITC reflects the phase consistency (or synchronisation) between the trials, at every time point and frequency range. ERSP/ITC changes in coherence enable us to measure and assess the multi-cycle oscillations that we had observed in the ERPs.

In-line with the previous two experiments, we applied a notch filter, between 7 and 9 Hz, during the initial processing/epoching of the EEG data, in order to filter out any Steady State Visual Evoked Potential (i.e. to remove the SSVEP, which results from the experiment's RSVP presentation rate, as explained in section 3.3.2 – *EEG data*). This would justify our focus on the fixed-boundary analysis window of 0.5 to 7Hz, assuming that there are no significant power increases at higher frequencies. However, in addition to the fixed-boundary analysis window (0.5 to 7 Hz), we also performed the full ERSP/ITC analyses on the full frequency range (0.5 to 45 Hz), to assess the power/coherence changes at all frequencies.

7.4.4.1 – Group-level TFA

As outlined in section 3.3.5.1 (*Time Frequency Window Placement*), the group-level critical time window, for measuring ERSP/ITC, was placed based on the AGAT of power/coherence. As seen in Figure 7.8 (below), ERSP and ITC results of the AGAT of the Probe and Irrelevant conditions are combined together, across all 14 subjects at Pz, with a large power increase, around 300 to 650ms time-window (post-stimulus), at a low frequency range. However, an interesting pattern can also be observed in the 10 to 20 Hz frequency range, which was not present in the previous two experiments.

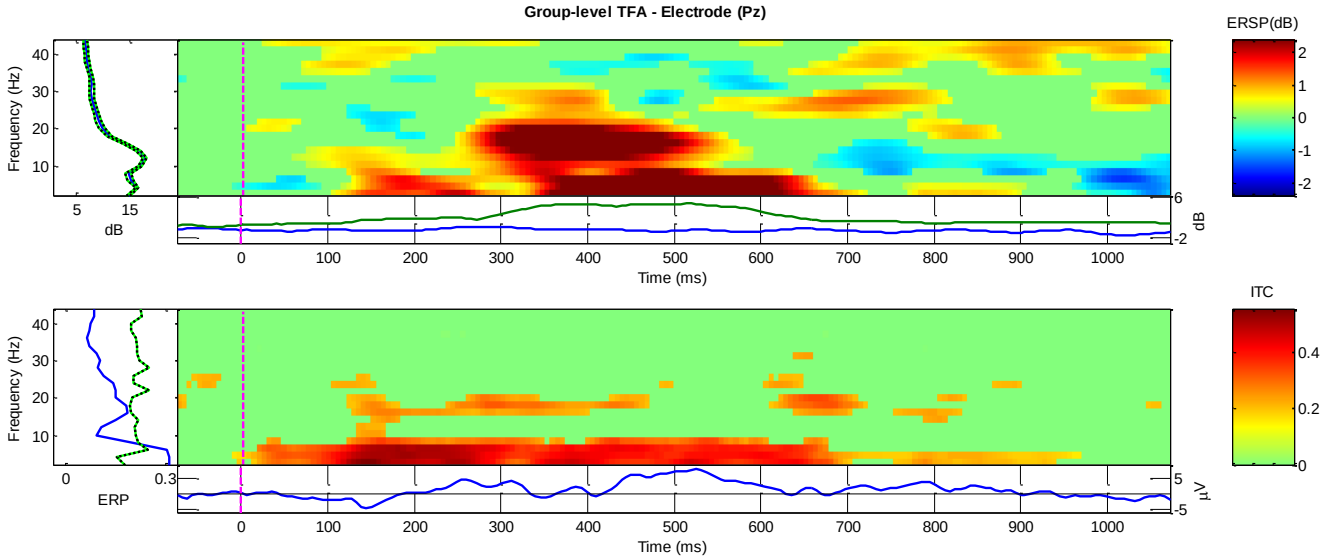

Figure 7.8 – Group-level Time Frequency plots, at the Pz electrode, using the combined Probe and Irrelevant conditions. The top plot shows the ERSP mean power spectrum (with its low/high envelope, directly below), and the bottom plot shows the ITC significance, when the EEG phase, at a given time and frequency, in single trials becomes locked across trials. Evoked increases in power/choherence have been concentrated in the 0.5 to 10 Hz frequency range, and an alpha/low-beta pattern can be seen, in the 10 to 20 Hz frequency range, which could not be seen in our previous two faces experiments. Note that SSVEP has been filtered out, by applying a 7:9 Hz notch filter.

As explained in section 3.3.5.1 (*Time Frequency Statistical Test*), ERSP/ITC statistical tests were performed to compare the power and coherence changes between the Probe & Irrelevant conditions; at the group-level, two measures were obtained for each subject, and a two-tailed paired *t*-test was used to calculate the significance for ERSP and ITC. As can be seen in the grand-Probe versus grand-Irrelevant ERSP/ITC comparisons (see Figure 7.9), increases in power/coherence are predominantly evident in the grand-Probe condition, which suggests detection of the familiar lecturer face (ERSP > 5dB, and ITC > 0.4). Whilst the grand-Irrelevant condition lacks a similar power/coherence fluctuations, within the same time window, its ITC plot shows a phase reset (between 100 – 300ms) without a corresponding ERSP power increase, which may be related to the unusual ERP negativity (peaking at 300ms) that we reported in figures 7.2 and 7.4 (i.e. a potential covert response to the repetition of the Irrelevant). These results are similar to the previous two experiments, but not all the power/coherence fluctuations are occurring in the lower band (i.e. 0.5 to 7Hz), since in this experiment, the Probe shows a power increase without phase reset (between 300 – 500ms), in the form of an alpha/low-beta pattern that can be seen, in the 10 to 20 Hz frequency range.

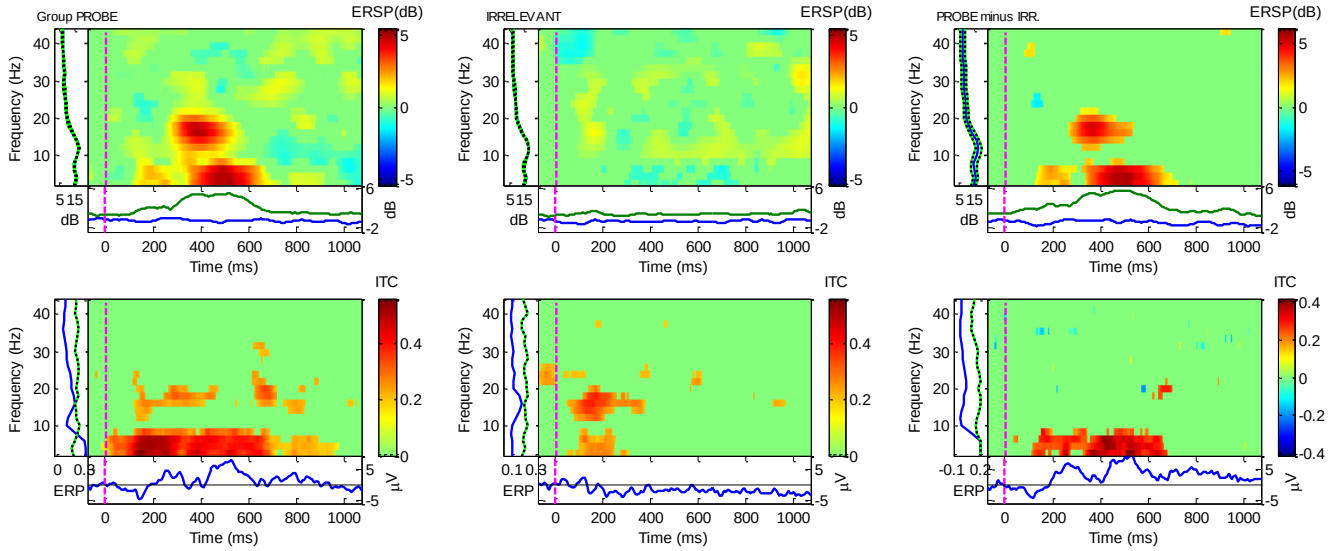

*Figure 7.9 – Group-level Time Frequency Analysis, at Pz electrode, for the difference between critical stimuli (Probe and Irrelevant), across the full frequency range (**0.5 to 45 Hz**), showing ERSP (top row) and ITC (bottom row). The first column of ERSP/ITC plots show the power/coherence changes in the grand-Probe condition, and the second column shows the same for the grand-Irrelevant condition. The third column is the difference between grand-Probe and grand-Irrelevant (i.e. Probe minus Irrelevant), which confirms significant group-level increases in power/coherence for the grand-Probe only: ERSP (t(13) = 3.9433, p = 0.0017, d = 1.4888), and ITC (t(13) = 2.3126, p = 0.0378, d = 1.0315). Note that at each frequency and time point, increases in power/coherence are in red; decreases in power/coherence are in blue, and green indicates no significant change in power/coherence. Whilst the majority of the difference (see the third column) is shown in the 0.5 to 10 Hz frequency range, an evoked response can be observed at alpha/low-beta (i.e. 10 to 20Hz, between 300 to 500ms), in the form of a power increase without phase reset. This evoked response can also be observed in the Probe condition (see the first column), but, interestinglty, in the Irrelevant condition (see the middle column) the opposite effect can be observed, between 100 to 300ms, where a phase reset does not exhibit a power increase.*

Over the full frequency-range (0.5 to 45 Hz), the group-level ERSP analysis at Pz electrode revealed a significant result, at the AGAT defined window 334 to 434ms (see Figure 7.9, above), confirming a difference between the Probe and Irrelevant conditions ($t(13) = 3.9433$, $p = 0.0017$, $d = 1.4888$). For the group-level ITC over the same (maximum) frequency range, our statistical tests confirmed a significance, at the AGAT defined window 428 to 528ms: ($t(13) = 2.3126$, $p = 0.0378$, $d = 1.0315$).

On the narrower frequency-band (0.5 to 7 Hz), the group-level ERSP analysis at the Pz electrode revealed a highly significant result, at the AGAT defined window 434 to 534ms (see Figure 7.10, below), confirming a difference between Probe and Irrelevant conditions ($t(13) = 5.9302$, $p < 0.0001$, $d = 2.2688$). For group-level ITC over

the same (narrower) frequency range, our statistical tests also confirmed a highly significant result at the AGAT defined window 381 to 481: ($t(13) = 5.5128$, $p = 0.0001$, $d = 2.1298$).

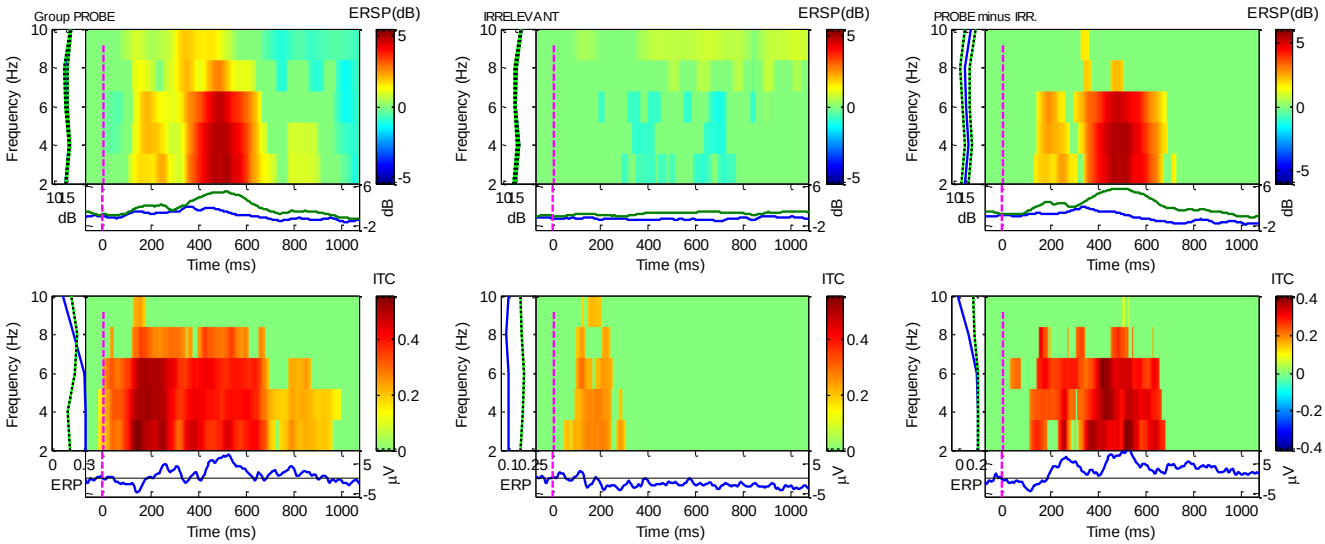

*Figure 7.10 – Group-level Time Frequency Analysis, at Pz electrode, for the difference between critical stimuli (Probe and Irrelevant), at the narrower frequency band (**0.5 to 7 Hz**), showing ERSP (top row) and ITC (bottom row). The first column of ERSP/ITC plots show the power/coherence changes in the grand-Probe condition, and the second column shows the same for the grand-Irrelevant condition. The third column is the difference between grand-Probe and grand-Irrelevant, which confirms group-level increases in power/coherence for the grand-Probe only: ERSP (t(13) = 5.9302, p < 0.0001, d = 2.2688), and ITC (t(13) = 5.5128, p = 0. 0001, d = 2.1298). Note that at each frequency and time point, increases in power/coherence are in red; decreases in blue, and green indicates no significant change.*

7.4.4.2 – Subject-level TFA

Per subject statistical analysis (i.e. a randomisation test on the combined Probe and Irrelevant conditions) confirmed the high significance of the increase in the Probe's power (ERSP) and coherence (ITC), as compared to the Irrelevant. Statistical tests of the *narrower* frequency band (0.5 to 7 Hz), resulted in two independently measured time windows and *p*-values that revealed a significant difference between the conditions (see table 7.4, below). For ERSP, 13 out of 14 subjects' *p*-values (92.9%) were significant, and for ITC, 12 out of 14 subjects' *p*-values (85.7%) were significant, confirming the difference between the Probe and Irrelevant conditions.

Subject	*ERSP p*-values	aERPt (ms)	*ITC p*-values	ITC (ms)
1	<0.0001	439	<0.0001	549
2	0.001	461	<0.0001	582
3	0.008	230	0.019	605
4	0.001	439	0.012	434
6	<0.0001	162	<0.0001	203
7	<0.0001	439	<0.0001	439
8	<0.0001	410	<0.0001	393
9	<0.0001	404	0.105	289
10	0.113	203	0.23	697
11	0.013	795	0.018	59
12	<0.0001	422	<0.0001	375
13	<0.0001	404	<0.0001	53
14	<0.0001	398	<0.0001	422
15	<0.0001	434	<0.0001	381

*Table 7.4 – Subject-level Time Frequency analysis of power (ERSP) and coherence (ITC), at Pz electrode, using the narrower frequency range (**0.5 to 7 Hz**). For each subject, an orthogonal contrast time window was employed (using the aERPt method), and p-values were obtained for ERSP and ITC, by comparing the Probe and Irrelevant conditions, using a randomisation statistical test. At an alpha level 0.05, 13 of 14 ERSP p-values (92.9%) were significant, and 12 of 14 ITC p-values (85.7%) were significant.*

Finally. even though we have justified the reason why the upper boundary of our analysis was fixed at 7 Hz (i.e. due to SSVEP waveform, which required a notch-filter on 7 to 9 Hz), we confirmed that per-subject statistical analysis of the maximum frequency range (0.5 to 45 Hz), resulted in *p*-values that revealed a difference between the Probe and Irrelevant conditions (see Table 7.5, below). For ERSP, 11 out of 14 subjects' *p*-values (78.6%) were significant. As for ITC, 12 out of 14 subjects' *p*-values (85.7%) were significant.

Subject	*ERSP p*-values	aERPt (ms)	*ITC p*-values	ITC (ms)
1	<0.0001	543	0.006	434
2	0.003	428	0.002	594
3	0.082	230	<0.0001	664
4	0.006	508	0.008	652
6	<0.0001	318	<0.0001	301
7	0.007	242	<0.0001	334
8	0.035	416	0.001	313
9	<0.0001	461	0.219	428
10	0.21	352	0.357	756
11	0.059	779	0.005	53
12	<0.0001	369	<0.0001	94
13	0.004	398	0.002	104
14	0.006	213	0.001	186
15	<0.0001	334	<0.0001	600

*Table 7.5 – Subject-level Time Frequency analysis of power (ERSP) and coherence (ITC), at Pz electrode, using the maximal frequency range (**0.5 to 45 Hz**). For each subject, an orthogonal contrast time window was employed (using the aERPt method), and p-values were obtained for ERSP and ITC, by comparing the Probe and Irrelevant conditions, using a randomisation statistical test. At an alpha level 0.05, 11 of 14 ERSP p-values (78.6%) were significant, and 12 of 14 ITC p-values (85.7%) were significant. These results are similar to those from the narrower frequency range (see Table 7.4), albeit, by focusing on 0.5 to 7 Hz, we observed a higher ERSP significance (i.e. 13 of 14 instead of 11 of 14), but ITC significance was the same.*

7.4.5 – Other midline electrode sites

In addition to the above analyses on the Pz electrode, we were also interested in the other two midline electrodes (Cz and Fz), to confirm that, in-line with (Kaufmann, Schulz, Grünzinger, & Kübler, 2011), the strongest brain responses to familiar faces are recorded at Pz. The following analogous Time domain analyses of Fz and Cz, aim to find out if the P600f evoked by the Probe was significantly different from that evoked by the Irrelevant.

7.4.5.1 – Fz electrode

At the group-level, the grand average ERPs of the two critical stimuli (i.e. the Probe and Irrelevant conditions), at the Fz electrode site, revealed a clear difference between the conditions (Figure 7.11, below). The Irrelevant condition, which consisted of an unknown lecturer face (paired with the Probe, and repeated randomly, as many times), did not present a similar pattern to the Probe, or the Target. This was as expected because non-salient information is unlikely to breakthrough into conscious awareness, due to the high presentation rate of the RSVP streams. However, the Probe condition elicited a continuous oscillatory pattern, within a 200 to 650ms time frame. This waveform at Fz is similar to the oscillatory waveform at Pz (Figure 7.6, above), and it confirms the prediction of a large difference between the Probe (familiar lecturer face) and the Irrelevant (unknown lecturer face) conditions, at all midline electrodes. At this Fz electrode site, an orthogonal contrast time window, for the highest positive (P600f) component, was independently found (using the AGAT method), at 469 to 568ms. Statistical analyses at Fz – in the form of a paired *t*-test of the mean amplitudes of Probe and Irrelevant, across all participants – was used to find the group-level significance of the P600f component. Our statistical tests produced a significant difference between the Probe ($M = 4.4937$, $SD = 2.6181$) and Irrelevant ($M = 1.4758$, $SD = 1.3568$), at Fz electrode site: ($M = 3.0179$, $SD = 3.3575$), $t(13) = 3.361$, $p = 0.0051$, $d' = 1.4473$.

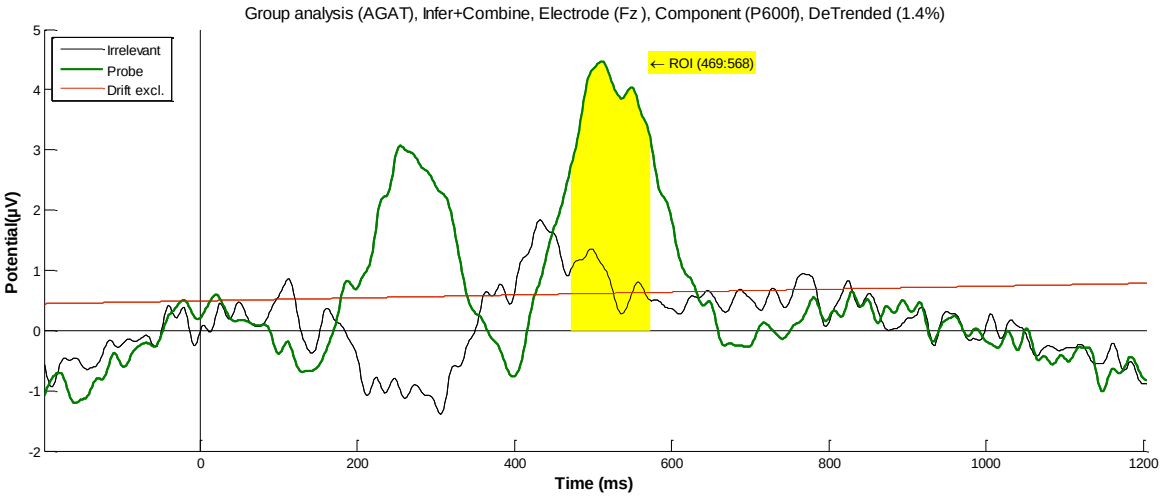

*Figure 7.11 – Grand average ERPs elicited by the Probe and Irrelevant, at the **Fz** electrode, showing an oscillatory pattern for the Probe condition (in green), which does not exist for the Irrelevant condition (in black). The linear Drift has been excluded (in red, at 1.4% to the vertical) with a detrending method. Even though subjects were informed of the presence of the Probe (familiar lecturer face), they were not told which lecturers were included. Statistical tests show a significant difference between Probe and Irrelevant, for P600f (t(13) = 3.361, p = 0.0051, d' = 1.4473).*

At the Fz electrode, subject-level statistical tests (i.e. Monte Carlo permutation) on the P600f component confirmed that 8 of 14 subjects (57.1%) showed critical-significance (0.05 alpha level) between Probe and Irrelevant (see Table 7.6). The 21.5% increase in the number of subjects that were shown to be significant at Pz (to wit: 11 of 14; 78.6% – see Table 7.2) imply a stronger brain response, to familiar faces, when compared to Fz (agreeing with studies (Kaufmann, Schulz, Grünzinger, & Kübler, 2011)), but statistical tests comparing the two electrode sites' *p*-values cannot confirm a significant difference: (Difference = 21.5%) $X^2(1, N = 14) = 1.430$, $p = 0.2317$, CI_{95} [-12.35%, 49.68%].

Subject	Probe (*M*)	Irrelevant (*M*)	*p*-value
1	9.7492	2.6109	<0.0001
2	4.4991	-0.5418	<0.0001
3	4.2241	5.2069	0.7220
4	3.6224	2.4896	0.2380
6	3.1290	2.5343	0.3680
7	5.5893	2.1383	0.0010
8	5.0339	0.6045	<0.0001
9	0.6038	2.5037	0.9050
10	1.8931	2.3348	0.6520
11	2.9475	3.4087	0.6130
12	8.3205	1.1370	<0.0001
13	4.5776	0.5088	0.0020
14	5.4466	2.5851	0.0450
15	8.9254	-1.3192	<0.0001

*Table 7.6 – Subject-level analysis, at **Fz** electrode, for the P600f component, showing the mean amplitude values of the Probe and Irrelevant conditions, from the same 100ms time window, which was independently found using the aERPt method. Statistical tests on P600f resulted in 8 of 14 subjects (57.1%) being significant, which is not as many as equivalent results at Pz (i.e. 11 of 14; 78.6%).*

7.4.5.2 – Cz electrode

The same group-level analysis that was carried out at Fz (see section 7.4.5.1), was performed at Cz, revealing a difference between the Probe and Irrelevant conditions (see Figure 7.12, below). Once again, the Irrelevant condition did not present a similar pattern to the Probe (or the Target), and the Probe condition elicited a continuous oscillatory pattern, within a 180ms to 700ms time frame. This waveform, at Cz, is very similar to the oscillatory waveforms at Pz and Fz (see Figures 7.6 and 7.11, respectively), and it confirms the prediction of a large difference between the Probe (familiar lecturer face) and the Irrelevant (unknown lecturer face) conditions, at all midline electrodes. At this Cz electrode site, an orthogonal contrast time window, for the highest positive (P600f) component, was independently found (using the AGAT method), at 471 to 570ms. Statistical analyses at Cz, in the form of a paired *t*-test, were employed to find the group level significance of the P600f component. Our statistical tests produced a significant difference between the Probe ($M = 3.6665$, $SD = 2.2793$) and Irrelevant ($M = 0.4912$, $SD = 1.189$), at Cz electrode site:

$(M = 3.1753, SD = 2.9436), t(13) = 4.0361, p = 0.0014, d' = 1.7468.$

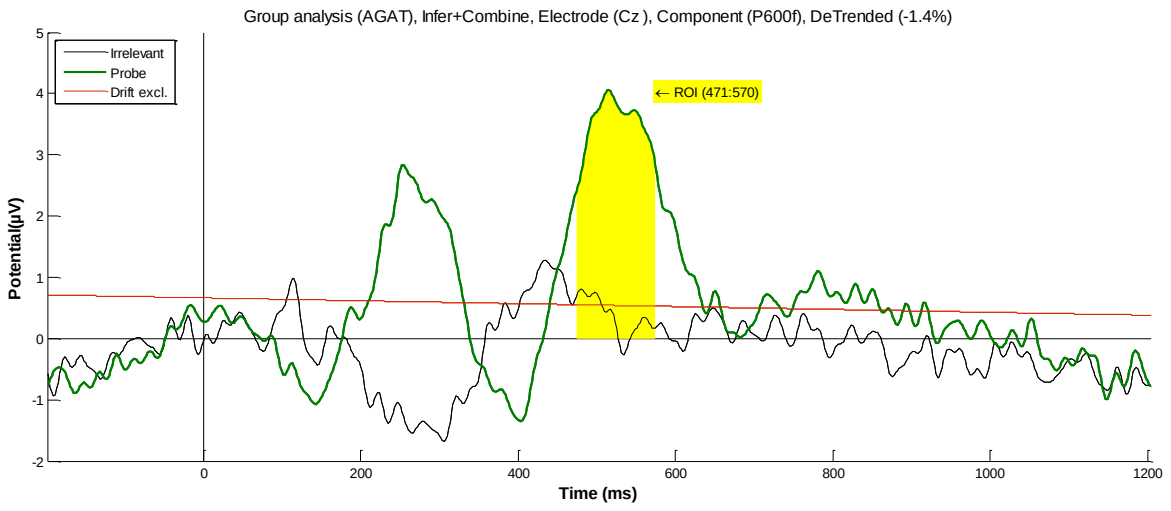

*Figure 7.12 – Grand average ERPs elicited by the Probe and Irrelevant, at the **Cz** electrode, showing an oscillatory pattern for the Probe condition (in green), which does not exist for the Irrelevant condition (in black). The linear Drift has been excluded (in red, at -1.4% to the vertical) with a detrending method. Even though subjects were informed of the presence of the Probe (familiar lecturer face), statistical tests show a significant difference between Probe and Irrelevant, for P600f (t(13) = 4.0361, p = 0.0014, d' = 1.7468).*

At the Cz electrode, subject-level statistical tests (i.e. Monte Carlo permutation) on the P600f component confirmed that 10 of 14 subjects (71.4%) showed critical-significance (0.05 alpha level) between Probe and Irrelevant (see Table 7.7). Despite a 7% increase in the number of subjects that were shown to be significant at Pz (to wit: 11 of 14; 78.6% – see Table 7.2), statistical tests comparing the two electrode sites' *p*-values cannot confirm a significant difference:

(Difference = 7.2%) $X^2(1, N = 14) = 0.187$, $p = 0.6657$, CI_{95} [-23.92%, 36.72%].

Subject	Probe (*M*)	Irrelevant (*M*)	*p*-value
1	7.3404	1.0204	<0.0001
2	4.1877	-0.8849	<0.0001
3	5.2249	5.0955	0.4640
4	3.1444	0.8431	0.0420
6	2.5051	1.5781	0.2860
7	4.9398	1.0250	<0.0001
8	3.6431	0.4196	0.0010
9	0.3401	0.9694	0.6790
10	2.5201	0.2825	0.0250
11	1.3796	1.6686	0.5940
12	5.6589	0.5160	<0.0001
13	4.3934	-0.3496	0.0020
14	4.6006	0.9850	0.0060
15	7.7840	-1.8335	<0.0001

*Table 7.7 – Subject-level analysis, at **Cz** electrode, for the P600f component, showing the mean amplitude values of the Probe and Irrelevant conditions, from the same 100ms time window, which was independently found using the aERPt method. Statistical tests on P600f resulted in 10 of 14 subjects (71.4%) being significant, which is not as many as equivalent results at Pz (i.e. 11 of 14; 78.6%).*

Finally, we have demonstrated that all three midline electrodes (Pz, Fz and Cz) have exhibited similar oscillatory waveforms, and that statistical tests showed significant difference between the two conditions, Probe and Irrelevant. Although our choice to focus on the Pz electrode was a priori (in-line with (Kaufmann, Schulz, Grünzinger, & Kübler, 2011)), we found some evidence that the strongest brain responses to familiar lecturer faces, was indeed recorded at the Pz electrode. Note that this finding is in accordance with the results of the first (celebrity faces) experiment, as well as, the second (concealed lecturer faces) experiment.

7.5 DISCUSSION

The ultimate aim of the third-and-final experiment was to investigate whether the new design (i.e. experiment's Part I informing Part II) and the additional instruction (i.e. revealing the presence of lecturer faces) can improve the breakthrough of familiar faces, into conscious awareness, using the RSVP subliminal search paradigm. Additionally, we investigated whether the new statistical analysis method (i.e. the 'Combined' method, as described in Chapter 6) can improve statistical power (i.e. increase SNR, since there are more trials), and deliver improved detection of the breakthrough event, at group and subject levels. This would be achieved through statistical analyses of the ERP data (in the Time domain) and single-trial ERSP/ITC data (in the Frequency domain), to determine whether the evoked response by the Probe/familiar faces were significantly different from that evoked by the Irrelevant/unknown faces. These results would take our findings in the first experiment (i.e. can celberity faces be used to infer recognition, using the RSVP paradigm?), and our findings in the second experiment (i.e. will the recognition of personally familiar faces achieve a similar breakthrough?), to our ultimate goal of developing a scientifically robust framework, in the form of our third experiment's functional prototype, which could advance future applications of deception detection tests, using faces in RSVP-based EEG tests.

In the current experiment, informing the subject of the presence of familiar (Probe) faces – without giving away any particulars about the Probes – was considered to be a natural progression towards real-life application of our RSVP-based deception detection test, as subjects would naturally assume the raison d'être of the experiment, as soon as, they are presented with faces. Additionally, in this experiment, we have introduced a new two-part experimental framework, which enables the examiner to shortlist the subject's familiarity with multiple Probes (e.g. compatriot faces) in Part I, and then focus the investigation on the most significant Probe (e.g. the partner in crime) in Part II. This is an improvement to previous experiments, in which subject-level significance would infer a combined/general familiarity with multiple Probes (e.g. up to five celebrity faces in the first experiment, and up to three lecturer faces in the second

experiment), rather than a framework that can focus on the subject's probe-level (by-item) significance.

In a similart fashion to the previous two experiments, we observed an oscillatory pattern for the Probe (lecturer) faces, which was not present for the Irrelevant (unknown) faces. As the key comparison was between Probe faces and Irrelevant faces, a significant difference was observed between the ERPs, at all three mid-line electrodes (Pz, Fz and Cz). In this experiment, the Probe's oscillatory wave was a closer match to the pattern that we observed in the first (celebrity faces) experiment, albeit, the peak negativity of the N400f component was, in fact, similar to the second (concealed lecturer faces) experiment. However, both (concealed and revealed) lecturer faces experiments' N400f components appeared to be muted, when compared to the celebrity faces experiment.

Furthermore, we observed that this experiment's unexpected early negativity for the Irrelevant (around 300ms) resembled a similar effect in the concealed lecturer faces experiment, albeit, no such deflection was present in the first (celebrity faces) experiment. As subjects do not report seeing the Irrelevants, we wondered if this effect could be related to subliminal registering (i.e. a covert response or threshold awareness) of a repetition by the brain, or some other (yet to be explained) incongruity? Whilst further experiments need to be run to investigate the reasons for the above two differences in the Probe/Irrelevant conditions, we will propose our conclusions (see section 7.5.4 – *Future Work*), after describing the statistical test results in the Time and Frequency domains.

7.5.1 – Time Domain

At the Pz electrode, group-level analysis of ERPs confirmed the significance of the difference between the grand-Probe and grand-Irrelevant ($p = 0.0004$), and subject-level statistical analyses of ERPs confirmed that, having found the orthogonal contrast window for the P600f component, a total of 11 of 14 subjects (78.6%) had *p*-values below our critical-significance (alpha level 0.05), revealing a highly significant difference between the Probe and Irrelevant conditions. In terms of the number of

subjects, this result was an improvement on the previous two experiments' results, for P600f at Pz, where the first (celebrity faces) experiment's subject-level analysis achieved 50% (7 of 14 subjects) significance, and the second (concealed lecturer faces) experiment's subject-level analysis achieved 57.1% (8 of 14 subjects) significance. However, statistical tests comparing their results did not show a significant difference between the first-and-third experiments: (Difference = 28.6%) $X^2(1, N = 14) = 2.405$, $p = 0.1209$, CI_{95} [-6.37%, 55.62%], or, indeed, the second-and-third experiments: (Difference = 21.5%) $X^2(1, N = 14) = 0.430$, $p = 0.2317$, CI_{95} [-12.35%, 49.68%].

The results of our statistical analyses, within the Time Domain, provide evidence that the personally familiar lecturer faces (Probe conditions) were differentially perceived and processed by nearly all subjects' brains, as compared to the unknown lecturer faces (Irrelevant conditions). Even though both conditions were treated equally, our experimental findings show major differences between the Probe and Irrelevant, which was as a result of the former stimuli reaching conscious awareness and generating pronounced electrical responses (as seen in the Probe ERPs), whilst the latter was not sufficiently perceived to encode into working memory, and generate a distinct electrical response that resembled the Probe (or the Target). And yet, there was an interesting new electrical response (i.e. Irrelevant's negative deflection, peaking at 300ms), which may reflect subliminal awareness of repetition.

7.5.2 – Frequency Domain

At the Pz electrode, group-level analysis of Time Frequency, across the narrower frequency band (0.5 to 7 Hz), confirmed the significance of the difference between the grand-Probe and grand-Irrelevant for ERSP ($p < 0.0001$) and for ITC ($p = 0.0001$). Subject-level statistical analyses of Time Frequency, across the same frequency band (0.5 to 7 Hz), using the independently measured time window for ERSP, confirmed that 13 out of 14 subjects' *p*-values (92.9%) were significant. As for ITC, subject-level statistical analyses of Time Frequency showed that 12 out of 14 subjects' *p*-values (85.7%) were significant, confirming the difference between the Probe and Irrelevant conditions.

The results of our statistical analyses, within the Frequency Domain, provide additional evidence that the Probe (familiar lecturer) faces were differentially perceived and processed by most subjects' brains, as compared to the Irrelevant (unknown lecturer) faces. Whereas, the previous (concealed lecturer faces) experiment's increases in power (ERSP) and coherence (ITC), in the Probe condition, were not as significant as the first (celebrity faces) experiment, the current (revealed lecturer faces) experiment established much larger increases in power/coherence (i.e. similar to the celebrity faces experiment). Thus, demonstrating that such changes in power and phase coherence could have contributed to the generation of the P600f component, which was elicited within similar time windows of the same condition's Probe ERPs. Once again, this finding supports the hypothesis that oscillatory activity, in the frequency domain, is related to the ERP component, in the time domain (Makeig, Debener, Onton, & Delorme, 2004b) (Fuentemilla, 2008).

7.5.3 – Conclusion

This chapter's experimental findings confirm our first hypothesis that having revealed the presence of familiar lecturer faces that are personally known to the subject (instead of the previous two experiments' concealed inclusion), we were able to detect the group-level breakthrough of familiar faces into consciousness. Just as we did in the previous two experiments, we agree that such breakthrough would be encoded in brain signals (Bowman, et al., 2013), and would generate ERP components/effects that would differ between the Probes (familiar lecturer faces) and the Irrelevants (unknown lecturer faces). Once again, through the effective use of our statistical analyses, in the time domain (using ERPs), as well as, the frequency domain (using single-trials), we have successfully differentiated between the Probe and Irrelevant conditions, at the group-level, in all 3 mid-line electrodes (Pz, Cz and Fz).

Our second hypothesis was that the new design (which involves online qualification of the Probe), and the use of the new 'Combined' analysis method (which increases the SNR, without inflating the false positive rate) can improve the breakthrough and detection of Probe (lecturer) faces, on an individual basis, even though, only the Target was task-relevant. Whilst the Probe and Irrelevant conditions

were treated equally in the experiment, we used each subject's ERPs (in the time domain) to confirm the presence of large differences in brain responses between the conditions for 11 of 14 subjects (i.e. 3 more subjects than the previous experiment). Furthermore, using the subject's single-trials (in the frequency domain), we confirmed a difference between the Probe/Irrelevant conditions in 13 of 14 (for ERSP) and 12 of 14 (for ITC) subjects, (i.e. 6 more subjects for ERSP, and 3 more subjects for ITC, in comparison with the previous experiment).

All the above results confirm that the approach we have proposed in the current experiment – precisely: a) revealing the presence of the Probes; b) inferring the Probe with the highest significance through online tests; c) using the 'Combined' method to statistically test the difference between the Probe and Irrelevant – can improve detection rates and can lead to applications in deception detection, to determine whether a subject has high familiarity of real-life acquaintances.

Our third hypothesis was that the strongest brain responses to the familiar (Probe) faces are recorded at the Pz electrode site. Having carried out the same statistical tests on all mid-line elctrodes (Fz, Cz and Pz), we can confirm that all three sites exhibited similar oscillatory waveforms for the Probe, and statistical tests comparing the three electrode site's *p*-values could not confirm a significant difference, however, the number of subjects whose results were significant was higher at Pz (11 of 14), beating the results at Fz (8 of 14) and at Cz (10 of 14).

7.5.4 – Future Work

This chapter's experiment concluded our work on RSVP-based fringe-P3 studies, suggesting that our latest design and analysis methods can be applied to deception detection applications, in order to determine whether a subject has high familiarity of real-life acquaintances. In the first experiment of its kind, we have demonstrated that, in addition to the evocative celebrity faces, personally familiar faces can also breakthrough into conscious awareness, on an individual basis, even when the stimuli are not task-relevant. Furthermore, we have revealed that through the application of our latest framework (i.e. the two-part experiment that can infer highest familiarity)

and improved statistical tests in the Time and Frequency domains (e.g. the Combined method, which benefits from a high SNR and avoids inflation in Type I errors), we can expect highly significant results, at subject-level.

Whilst the above finding are very promising, there are three areas of enquiry that would benefit from future studies: the first two are related to the differences between the first (celebrity faces) experiment and the latter two (concealed and revealed lecturer faces) experiments; namely, the weakness in the N400f component for the Probe (familiar lecturer) faces, and the unexpected negativity in the Irrelevant (unknown lecturer) faces. The third area of enquiry relates to a new territory in RSVP-based fringe-P3 studies, whereby we could extend subject-level significance to a more specific Probe-level (by-item) significance, thus, demonstrating the relationship between a subject and a single acquaintance (rather than a subject and multiple Probes, which is the currently accepted practice). We will now expand on these three areas of enquiry:

7.5.4.1 – Differences in the N400f component

We have already pointed out that the oscillatory pattern for the current (revealed lecturer faces) experiment's Probe shows a strong similarity to the Probe condition in the previous two (celebrity faces and concealed lecturer faces) experiments. However, the celebrity faces experiment experienced a more extreme N400f effect (approx. -5μV), when compared with both lecturer faces experiments (approx. -1.5μV).

We hypothesie that this differnce may be related to the nature of the Probe stimuli, which changed from highly evocative faces of famous celebrities – with vivid associations to beauty, wealth, power, etc. – to the more mundane faces of familiar lecturers that subjects have real-life/personal dealings with. Furthermore, subjects would have previously been exposed to the images of the famous celebrity faces – since we used highly publicised photographs, which were frequently in the public eye, and assuredly seen by all subjects, on many occasions and over a far longer period of time – whereas, the lecturer faces' images were seen for the first time (i.e. in the format that we had procurred for our experiments). Thus, this anomily may require further investigation into the role that fame plays on the human psyche.

7.5.4.2 – Differences in the Irrelevant condition

As explained earlier, the unexpected early negativity for the Irrelevant condition, in the current (revealed lecturer faces) experiment was similar to the previous (concealed lecturer faces) experiment, but the first (celebrity faces) experiment did not show the same negativity (peaking at approx. 300ms). It is noteworthy that participants did not report seeing the Irrelevant (unknown lecturer) faces, so we have formed two theories, which we would like to explore in future work. The first is that this posterior negativity may be related to subliminal registering (i.e. a covert response or threshold awareness) of a repetition by the brain, and the second is that it relates to an incongruity, between the Irrelevant and filler/distractor images, which was not present in the first (celebrity faces) experiment. More specifically, the Irrelevant images in the first (celebrity faces) experiment were chosen randomly from the Distractor database, but the Irrelevant images in the latter two (concealed and revealed lecturer faces) experiments did not come from the Distractor database - they were unknown lecturer faces from a different University.

To be precise, the first experiment's Irrelevants and Fillers were from the same Distractor database, whereas, both lecturer faces experiments only used the Distractor database as Fillers, preferring to use *bespoke* Irrelevants (i.e. unknown lecturers), which were treated in the same manner as the Probes (vis-à-vis *bespoke* photography and editing). Thus, the unknown Irrelevant faces that were used in the latter two experiments may have been, somehow, differently perceived to the unknown Irrelevant/distractor faces that were used in the first experiment.

Within each RSVP trial, even though the bespoke Irrelevants (unknown lecturer faces) were unlikely to breakthrough into conscious awareness (as evidenced by the behavioural/memory tests of the latter two experiments), the subjects' brain, nonetheless, may have perceived an aesthetic difference between the 17 library-photos that were used as fillers, and the single bespoke-photo of the Irrelevant that we had procured for the unknown lecturers category (i.e. using the same camera, technique and editing methods that we used for the Probe photos of the familiar lecturers). Whilst this potential disparity could not be helped (i.e. we did not have sufficient time or resources

to professionally capture 500+ photos, as fillers/distractors), we expect the real-life application of our deception detection test to ensure that all Distractor images are captured and edited, in a similar manner to the Probe and Irrelevant images.

7.5.4.3 – Viability of Probe-level significance

Whilst acknowledging that our new 'Combined' analysis method can improve the breakthrough and detection of Probe faces, on a subject-level basis (by increasing the SNR, without inflating the false positive rates), we hypothesise improvements in our technique, which would enable the examiner to pinpoint the familiarity of a subject to a single Probe. This could be a new territory in RSVP-based fringe-P3 studies, whereby we can extend subject-level significance to a more specific Probe-level significance, which can reveal the relationship between a subject and a single acquaintance (rather than a subject and multiple Probes, which is the currently accepted practice).

Therefore, we hypothesise that if we increase Part II's number of blocks from two to three, or more (i.e. independently select the same Probe/Irrelevant pair that was inferred from Part I, and use them in three or more blocks of Part II), the increase in the number of trials/SNR may confer a higher significance, at subject-level. Additionally, we can discard Part I (i.e. the Decider method, which informs Part II), and perform statistical tests on Part II alone (according to the Abandoned method), in the knowledge that the *by-item* significace of the Probe, when compared to its paired-Irrelevant, would show the subject's familiarity to a single Probe/face (rather than multiple Probes, as expected in the Combined method). Of course, the downside of such a strategy is that we require more time to complete each experiment, which may not be practicable.

In conclusion, we propose that the above three hypotheses and areas of enquiries should be explored in future work, in order to refine and improve our current deception detection framework. Additionally, the standard practice of capturing data using 32 electrodes (as achieved in the current experiment, only) should lend itself to the application of new signal processing techniques, like Independent Component Analysis.

PART III

–

DISCUSION

Chapter 8:
Conclusions and Future Work

In the final chapter of this thesis, we will look back at the central hypotheses that were previously outlined in chapter 1, and draw conclusions in section 8.1, by summarising our research findings (i.e. from chapters 4 to 7). Next, we will contemplate the direction of our thesis in section 8.2, and its contributions to the field of cognitive neuroscience. Finally, in section 8.3, we will propose future work, and suggest scientific questions that could advance the research presented in this thesis.

8.1 – Conclusions

The central hypothesis of this thesis (as outlined in section 1.2) was that a new category of critical stimuli, in the form of human faces, can be introduced to the ERP-based RSVP paradigm, for the first time, and that the fringe-P3 method could be successfully used to detect intrinsic salience of familiar faces. Furthermore, even when there was no task associated with the stimuli, familiar faces will differentially break through into conscious awareness, such that we can detect the breakthrough events in EEG, using statistical tests, in Time and Frequency domains.

8.1.1 – Celebrity Faces

To begin the examination of our central hypotheses, **chapter 4** investigated the sensitivity of the ERP-based RSVP paradigm, to infer recognition of *celebrity* faces, and used statistical tests to differentiate between known (Probe) and unknown (Irrelevant) faces, at group and subject levels. This was achieved whilst a Target face was task-relevant, and participants were unaware of the inclusion of the celebrity (Probe) faces. In this chapter, we introduced the use of detrending, to independently remove any drift in EEG data, and to avoid the legacy practice of post/ad-hoc increasing of the high-pass

filter, which may adversely affect the low frequency P3 component and/or introduce waveform distortions.

Within the Time (ERP) domain, the results of our statistical analyses – at the Pz electrode site, 7 of 14 subjects (50%) for P600f, and 10 of 14 subjects (71.4%) for N400f had *p*-values below our critical-significance – provided evidence that the celebrity faces were differentially perceived and processed, as compared to the unknown faces. Even though both conditions were treated equally, our experimental findings show major differences between the Probe and Irrelevant, which was as a result of the former stimuli generating pronounced electrical responses (as seen in the Probe ERPs), whilst the latter did not generate a distinct electrical response.

Within the Frequency Domain, the results of our statistical analyses – across the 0.5 to 7 Hz frquency band, all 14 subjects (100%) for ERSP and ITC showed significant *p*-values – provided additional evidence that the celebrity faces were differentially perceived and processed by all subjects' brains, as compared to unknown faces. The large increases in power (ERSP) and coherence (ITC), which were observed and statistically confirmed in the Probe condition only, suggest that such changes in power and phase coherence could have contributed to the generation of the N400f/P600f components, which were elicited within similar time windows of the same condition's Probe ERPs.

8.1.2 – Lecturer Faces

Having provided evidence that famous/celebrity faces can breakthrough into conscious awareness, using an RSVP subliminal search paradigm, we substituted the highly evocative faces of famous celebrities with familiar faces that were personally known to the participants, in the form of the University's *lecturers*. Subsequently, in **chapter 5**, we demonstrated that we can differentiate between the Probe (familiar University of Kent lecturer) and Irrelevant (unknown lecturers from another university) faces, at group and subject levels, using statistical analyses in the Time and Frequency domains.

Within the Time (ERP) domain, the results of our statistical analyses – at the Pz electrode site, 8 of 14 subjects (57.1%) for P600f, had significant *p*-values – provide evidence that the personally familiar lecturer faces were differentially perceived and processed by most subjects' brains, as compared to the unknown lecturer faces (i.e. one more than celebrity faces experiment).

Within the Frequency Domain, the results of our statistical analyses – across the 0.5 to 7 Hz frquency band, 10 of 14 subjects (71.4) for ERSP and 9 of 14 subjects (64.3%) for ITC showed significant *p*-values – provided additional evidence that the familiar lecturer faces were differentially perceived and processed by most subjects' brains, as compared to unknown faces. As before, the large increases in power (ERSP) and coherence (ITC), which were observed and statistically confirmed in the Probe condition only, demonstrated that such changes in power and phase coherence could have contributed to the generation of the P600f component, which was elicited within similar time windows of the same condition's Probe ERPs.

8.1.3 – Data Simulations

The results of the above two studies suggested that we could apply our findings to the differentiation of deceivers and non-deceivers, in the application of crime compatriots, whereby, a suspect's familiarity with a criminal/terrorist can be established using faces. However, we hypothesised that we could further improve the statistical power of detecting an effect and enhancing the signal-to-noise ratio (SNR). Thus, we used ground-truth data simulations, in **chapter 6**, to explore the viability of using online statistical tests, to focus experimental data collection efforts, on the critical stimulus with the highest significance, in order to improve statistical power (i.e. reduce the risk of Type II errors), without the inflation of Type I errors.

As a result of these methodological explorations, we were able to justify the design of a novel two-part experiment, in which Part I of the experiment would independently select the critical stimuli for Part II, using online statistical tests to infer the familiar face that achieves the highest significance. Furthermore, we demonstrated

that, out of three different methods of statistically analysing the data (i.e. Abandoned, Biased and Combined), the Combined method of analysis ensures that our statistical results are well-behaved (when compared to Biased), and possess the highest statistical power (when compared to Abandoned).

8.1.4 – Revealed Lecturer Faces

In our final experiment, we introduced a key change to the instructions given to the subject (at the start of the experiment), in order to modify the *covert* nature of presenting familiar faces (which was the modus operandi of the first two experiments), into an *overt* familiarity study, in **chapter 7**. Thus, by revealing the possibility of participants encountering faces that are personally known to them (without telling them who these familiar faces could be), our third-and-final experiment simulated a real-life scenario, in which subjects/perpetrators would deduce that the purpose of being shown a series of faces is to ascertain their familiarity with an accomplice. This experiment demonstrated the closest workable solution for deception detection applications, using faces in RSVP-based EEG tests.

Additionally, in this experiment, we have introduced a new two-part experimental framework (as formulated in chapter 6), which enables the examiner to shortlist the subject's familiarity with multiple Probes (e.g. compatriot faces) in Part I, and then focus the investigation on the most significant Probe (e.g. the partner in crime) in Part II. As noted earlier, this is an improvement to previous experiments, in which subject-level significance would infer a combined/general familiarity with multiple Probes (e.g. up to 5 celebrity faces in the first experiment), whereas, the new framework (as used in the final *revealed lecturer faces* experiment) could reveal the subject's probe-level (by-item) significance. As a result, we propose that this framework will be empirically relevant to the application of real-life deception detection of compatriots.

Within the Time (ERP) domain, the results of our statistical analyses – using the Combined method (as justified in chapter 6) at the Pz electrode site, 11 of 14 subjects (78.6%) for P600f, had significant *p*-values – provide evidence that the personally

familiar lecturer faces were differentially perceived and processed by most subjects' brains, as compared to the unknown lecturer faces (i.e. four more than celebrity faces experiment, and three more than covert lecturer faces experiment).

Within the Frequency Domain, the results of our statistical analyses – across the 0.5 to 7 Hz frquency band, 13 of 14 subjects (92.9) for ERSP and 12 of 14 subjects (85.7%) for ITC showed significant *p*-values – provided additional evidence that the celebrity faces were differentially perceived and processed by most subjects' brains, as compared to unknown faces. Once again, the large increases in power (ERSP) and coherence (ITC), which were observed and statistically confirmed in the Probe condition only, demonstrated that such changes in power and phase coherence could have contributed to the generation of the P600f component, which was elicited within similar time windows of the same condition's Probe ERPs.

In conclusion, the findings of our research provided evidence that familiar faces are differentially perceived and processed by participants' brains, as compared to novel (unfamiliar) faces. Therefore, we propose our final experiment to be a workable solution for deception detection applications of crime compatriots (e.g. accomplices), using faces in RSVP-based EEG tests.

8.1.5 – ERP Comparisons

By fixing the limits of the time/potential axes, we were able to plot and compare the Probe/Irrelevant ERPs (at Pz) for all experiments in this thesis (see Figure 8.1). All three experiments' ERPs showed a similar oscillatory pattern for the Probe, and their peak positivity for P600f (at 500ms) was highly significant. However, the peak negativity for N400f (at 400ms) was more extreme in amplitude in the first (celebrity faces) experiment. Indeed, the largest Probe oscillatory amplitudes were found in the celebrity faces experiment (+3.7 to -3.8 μV), whose pattern was closer to the third (*revealed* lecturer faces) experiment (+3.4 to -1.9 μV), and the second (*covert* lecturer faces) experiment showed the smallest amplitude effects of all (+2.6 to -1.1 μV).

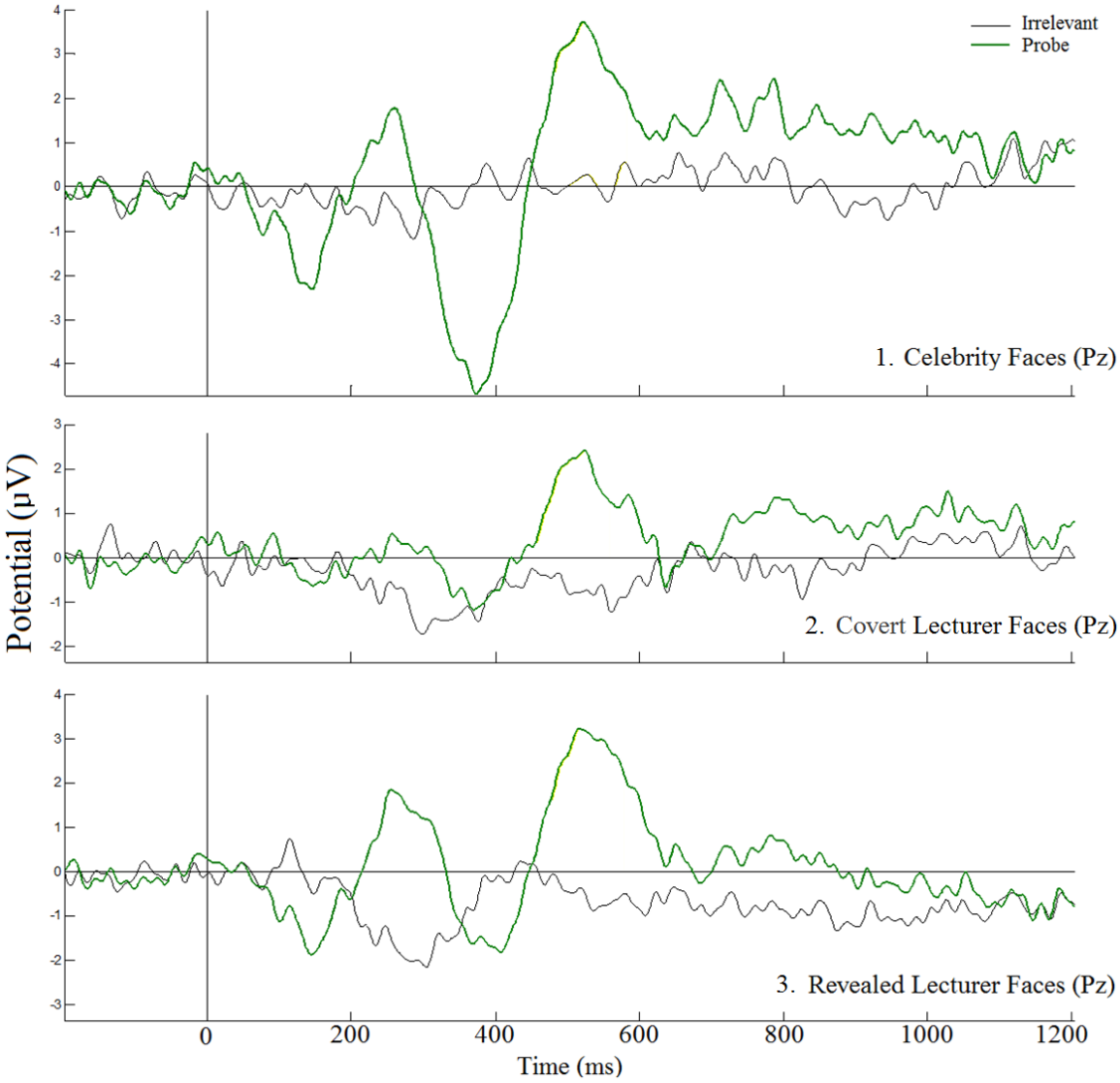

Figure 8.1 – Grand average ERPs elicited by the two critical stimuli that are being compared (Probe/Irrelevant), at Pz electrode site, for all three experiments, using the same axis limits. All three Probe conditions elicited a continuous oscillatory pattern, and a different pattern for the Irrelevant, which was presented as many times as the Probe. Whilst the Probe's oscillatory pattern, and the peak positivity for P600f (around 500ms), is similar for all three experiments, the peak negativity for N400f (around 400ms) is more extreme in amplitude in the first (celebrity faces) experiment. Furthermore, the early-negativity (around 300ms) in the Irrelevant condition of the second and third experiments, was unexpected, and will require further investigation.

As for the comparison of the Irrelevant conditions, the early-negativity (around 300ms) for the second and third experiments, was not present in the first experiment (see Figure 8.1). Furthermore, due to this unexpected negativity in the Irrelevant condition, which overlaps the lowest negativity in the Probe condition (around 400ms), the results of our group-level statistical tests (between the Probe and Irrelevant) for the

N400f component were not significant. Note that the first experiment did not elicit the same negative deflection for the Irrelevant condition. As subjects do not report seeing the Irrelevants, we wonder if this effect could be related to subliminal registering (i.e. a covert response or threshold awareness) of a repetition by the brain, or some other difference that has yet to be investigated.

8.1.6 – General Limitations

Having declared that the focus of our research has been to establish a proof of principle, and that the use of University students, as a selective sample of participants, is inherent in the vast majority of similar psychology/perception studies, we would like to acknowledge associated limitations of our research. Namely, it could be suggested that a key difference between our experiments, and those conducted in a real-world scenario, is that we expected our participants to cooperate with the instructions, whereas, the suspect/criminal participant may have a vested interest in confounding the test results. Therefore, future studies must apply our proof of principle to a wider group/class of participants, in order to demonstrate the applicability of our methods and findings.

Additionally, it could be argued that the modest sample size (i.e. 14 subjects in each experiment) would decrease statistical power, limiting our capacity to compare performance across our three experiments. Thus, future experiments that intend to compare different groups (e.g. groups of criminals vs. groups of innocent suspects) should include a larger sample size. However, it should be noted that the group-level effect, of each experiment, in this thesis, is highly significant ($p < 0.001$) and reliable. As for the reliability of subject-level significance, the sample size is reassuringly high (between 225 and 375 trials for each subject).

Furthermore, the lack of an unknowing (innocent) control group presents another limitation in our proof of principle research, which must be addressed in future studies. As a point of comparison, the findings of an earlier RSVP study, which was conducted by our group (Bowman, et al., 2014), included the results of an Innocents group, in its ‘Names’ experiment. Additionally, we must acknowledge potential

limitations in our EEG deception detection tests that can affect its use in real-world scenarios, due to neurological disorders (e.g. prosopagnosia), which prohibit its application on all human subjects.

Having highlighted the general limitations, it would be noteworthy to outline two potential shortfalls that have been addressed in our experiments, which may provide room for further improvement, in future studies:

1. We know that subjects can affect the results if they avoid (or fail to observe) images that are presented in the middle of the screen (e.g. by shifting their focus to the edge of the screen, or crossing their eyes). To mitigate this limitation, we monitor each subject's lack of neural response to visual stimulation (i.e. the absence of SSVEP, as described in section 3.3.2). Additionally, we measure the Target image Hit/Miss rates, in order to identify low sensitivity or high bias (see section 4.3.4). To improve upon these safeguards, we could also incorporate an eye-gaze system, which would monitor subjects' pupils, in real-time, to reveal attentional focus and cognitive strategies.

2) We expect a difference between the Probe (known face) and Irrelevant (unknown face), but in order to reduce opportunities for countermeasures, the stream's presentation rate must be high enough to only permit the Probe images to break through into conscious awareness. Indeed, Probes may not be perceived if the presentation speed is too high, and Irrelevants may breakthrough if the speed is too low. This can be considered as a limitation of the Fringe-P3 method, which relies on the breakthrough effect, so we record subjects' recognition of Probe/Irrelevant images, using end-of-block memory questions (see section 4.3.5), and compare them with questions that relate to known and unknown images that were not presented in the experiment. The former reveal's each subject's ability to perceive the presented faces, and the latter gauges the subject's engagement with the tests (i.e. were subjects guessing the presence of salient faces). Whilst we have adopted a fixed presentation rate (i.e. a-priori SOA of 133ms), future studies can improve the breakthrough difference between the Probe and Irrelevant, by running pre-experiment training session (i.e. using the 'staircase' procedure), which assesses each subject's ability to perceive a Target face, at the highest presentation rate possible.

8.2 – Direction and Contribution

Following on from David Hubel and Torsten Wiesel's discovery of neurons in the visual cortex that selectively encode whether a line is vertical, horizontal, or diagonal (Hubel & Wiesel, 1959), Rodrigo Quian Quiroga investigated neurons in the brain, and discovered that a single neuron can be highly selective (Quiroga, Reddy, Kreiman, Koch, & Fried, 2005). The so called "Jennifer Aniston" neuron would fire to different images (or written name) of the actress. Despite the findings that individual neurons have somewhat specialised roles, it can be argued that each individual brain is different, and that it is constantly changing and adapting, as a result of brain plasticity.

In chapter 2 (see section 2.2.5 - *Brain as an intelligent machine*), we argued that the role of our neurons is to reduce uncertainty, by merging information into a unified whole. The amount of information that produces the unified whole is defined by the 'Integrated Information theory' (Tononi, 2008), where "each moment of awareness is a fusion of information from all of our senses". The human retina alone, is responsible for transmitting millions of bits of information (*per second*) to the brain, but only a few bits of unified information may break into consciousness (Nørretranders, 1998). Even then, the resultant unified information that reaches our Conceptual Short Term Memory (CSTM), may be lost or superseded (Shelvin, 2017), if it is not attended or perceived in time. This concept was a key tenet upon which the RSVP-based fringe-P3 method was successfully developed (see section 2.4.2 – *Fringe-P3 in Concealed Information Tests*).

The RSVP technique enabled us to present information at a very high speed, whilst observing the brain's electrical signals using an EEG. In RSVP, since the subject's brain is searching for salient stimuli, at a very high presentation rate, it is possible to detect an electrophysiological marker (e.g. the P3 component), which indicates when the salient stimulus is detected. Being more robust in the hands of deceivers, who may want to confound the test using countermeasures, the *fringe-P3* method is considered to be "a novel deception detection system" (Bowman, et al., 2013).

Having introduced a new dimension, in the form of celebrity faces, to the existing numbers/words/names based deception detection studies (Bowman, Filetti,

Alsufyani, Janssen, & Su, 2014), we successfully demonstrated the use of statistical tests, to differentiate between celebrity and unknown faces (see chapter 4). Next, we revised our methods, to explore the sensitivity of our ERP-based RSVP paradigm to infer recognition of broadly familiar lecturer faces (see chapters 5). Then, we used ground-truth data simulations to demonstrate the benefits of employing online statistical tests, to focus data collection efforts on the critical stimulus with the highest significance, in order to improve statistical power (i.e. reduce the risk of Type II errors), without the inflation of Type I errors (see chapter 6). Finally, we successfully applied the new analysis methods and the two-part experimental design – where Part II's parameters are influenced by Part I's results, using online statistical tests – to the third-and-final experiment (see chapter 7).

Whereas the first two experiments kept subjects naïve (i.e. they were not given prior information about the inclusion of celebrity/lecturer images), the third experiment was a logical progression from the earlier two, whereby, the key design change was that the subject was informed that (undisclosed) lecturer faces could be present in the experiment. However, like the previous two experiments, the task was to look for and report the target face only (i.e. an unknown face that subjects were instructed to detect).

As a result of the improvements to the experimental design and the new analysis methods, we consider the contributions from this thesis to be an important addition to the real-life application of RSVP-based EEG deception detection; especially as, the accused may not know (or be able to read) the name of the person that is being investigated. Indeed, our published findings of the first experiment (Alsufyani, et al., 2019) provide evidence that famous faces are differentially perceived and processed by participants' brains, as compared to novel (unfamiliar) faces. Therefore, we suggest that our final experiment could be a workable solution for deception detection applications of crime compatriots (e.g. accomplices), using faces in RSVP-based EEG tests. In essence, our investigation into familiar face detection and recognition opens viable possibilities for applications in deception detection – more specifically, to reveal crime compatriots/accomplices – as well as, face-related Brain Computer Interface (BCI) solutions.

8.3 – Future Work

This thesis introduced image-based stimuli to the existing fringe-P3 deception detection studies, by utilising RSVP-based EEG, to infer recognition of familiar faces, and to successfully differentiate between known and unknown faces, using statistical tests in the Time and Frequency domains. Our findings provide evidence that familiar faces are differentially perceived and processed by participants' brains, as compared to novel (unfamiliar) faces, however, we recognise the potential of methodological limitations, as well as, the prospect of scientific advances, which could be pursued to advance our research.

In particular, we would like to propose four areas of enquiry that would benefit from future studies: 1) integrate EEG with other technologies/techniques, to improve detection and counter countermeasure tactics; 2) investigate the relatively high amplitude of the Probe condition's N400f component, in the first (celebrity faces) experiment, as compared to the latter two (covert and revealed lecturer faces) experiments; 3) study the unexpected negativity in the Irrelevant condition (around 300ms), for the latter two (covert and revealed lecturer faces) experiments, which was not apparent in the first (celebrity faces) experiment; and 4) pursue the prospect of a new territory in RSVP-based fringe-P3 studies, whereby we could extend subject-level significance to a more specific Probe-level (by-item) significance, thus, demonstrating the relationship between a subject and a single acquaintance (rather than a subject and multiple Probes, which is currently our accepted practice).

8.3.1 – Integration with other technologies

Although the fringe-P3 method has been shown to be more robust in the hands of deceivers who may want to confound the test using countermeasures, further study into countering countermeasures – possibly, with the aid of eye-gaze systems – would

be highly valuable. Indeed, the prospect of integrating our framework with eye-gaze equipment, in order to explore the relationship between EEG's ERP components and pupil dilation, or micro-saccades, could reveal new findings and provide useful counterbalancing of stimuli (e.g. ruling out influences of low-level visual features).

8.3.2 – Differences in the N400f component

In chapter 7, we pointed out that the oscillatory pattern for both lecturer faces experiments (i.e. covert and revealed lecturer faces) showed a relatively low amplitude N400f component (approx. -1.5μV), when compared with the celebrity faces experiment (approx. -5μV). We hypothesise that this difference may be related to the nature of the Probe stimuli, which is the same in the two lecturer faces experiments (i.e. both use familiar lecturer faces), but different in the celebrity faces experiment.

We suggest that the highly evocative faces of famous *celebrities* – with vivid associations to beauty, wealth and power – may produce additional emotional/memory processes in the brain, which are lacking in the relatively mundane faces of familiar *lecturers*. Furthermore, subjects would have previously been exposed to the images of the famous celebrity faces (because we used highly publicised photographs, which were frequently in the public eye, and assuredly seen by all subjects, on many occasions and over a far longer period of time), whereas, the lecturer faces' images were seen for the first time (i.e. in the format that we had procurred for our experiments). Therefore, a new study could explore such differences, and further investigate the effect of presenting more personally evocative images, like close friends and family.

8.3.3 – Differences in the Irrelevant condition

As we have noted earlier, the unexpected early negativity for the Irrelevant condition (peaking at approx. 300ms), in both lecturer faces experiments (i.e. covert and revealed lecturer faces) is a marked contrast to the celebrity faces experiment, which did not show a similar negativity. It is noteworthy that participants did not report seeing the Irrelevant (unknown lecturer) faces, so we have formed two theories, which we would

like to explore in future work. The first is that this posterior negativity may be related to subliminal registering (i.e. a covert response or threshold awareness) of a repetition by the brain, and the second is that it may relate to an incongruity, between the Irrelevant and filler/distractor images, which was not present in the first (celebrity faces) experiment. More specifically, the Irrelevant images in the first (celebrity faces) experiment were chosen randomly from the Distractor database, but the Irrelevant images in the latter two (lecturer faces) experiments did not come from the Distractor database – they were (unknown) lecturer faces from a different University.

Thus, the unknown Irrelevant/lecturer faces that were used in the latter two experiments may have been, somehow, differently perceived to the unknown Irrelevant/distractor faces that were used in the first experiment. To mitigate this potential disparity, we propose that a similar study is performed, by replacing all 500+ distractor images with those that are more compatible with the critical stimuli (e.g. each distractor image could be photographed and edited, in the same manner/class that the Probe and Irrelevant images are processed).

8.3.4 – Viability of Probe-level significance

Whilst acknowledging that our new 'Combined' analysis method (proposed/justified in chapter 6, and put into practice in chapter 7) can improve the breakthrough and detection of Probe faces, on a subject-level basis, we hypothesise improvements in our technique, which would enable the examiner to pinpoint the familiarity of a subject to a single Probe. This could be a new territory in RSVP-based fringe-P3 studies, whereby we can extend subject-level significance to a more specific Probe-level (by-item) significance, thus, revealing the relationship between a subject and a single acquaintance (rather than a subject and multiple Probes, which is the currently accepted practice). Therefore, a new study could maintain Part I of the new experimental design (i.e. three blocks that infer the Probe/Irrelevant conditions for Part II), but increase Part II of the experiment, from two blocks to five (or more), in order to improve SNR. Then, we could discard Part I (i.e. ignore the Decider portion of the data, which will only be used to inform Part II's parameters), and perform statistical tests on Part II alone (i.e. adopt the Abandoned method), in the knowledge that the significance

of the Probe, when compared to its paired-Irrelevant, would show the subject's familiarity to a single familiar face. However, we acknowledge that increasing the number of blocks that repeat the *same* paired Prove/Irrelevant conditions may produce an unwanted/unexpected brain response, as the Irrelevant may eventually breakthrough, due to the number of repetitions. Additionally, it must be noted that the extra time required to perform a longer experiment may not be practicable.

In closing, we propose that the above four areas of enquiries should be explored, in future work, to advance the research presented in this thesis. Additionally, the standard practice of capturing EEG data using 32 electrodes (as achieved in the final experiment only) should provide additional opportunities to use alternative signal processing techniques, like Independent Component Analysis (ICA), which has yet to be studied using the RSVP-based fringe-P3 approach. Finally, we submit that our methods, techniques and findings strongly support the potential of utilising our research in forensic applications (e.g. knowledge of compatriots and accomplices).

PART IV

–

APPENDIX

Appendix A

Table A.1 *– The first (Celebrity Faces) experiment's behavioral results, as reported by 14 subjects. Each subject reported a confidence rating (1 to 5) of seeing ('Seen' heading in green) and knowing ('Know' heading in green) the five Probes (i.e. celebrity faces) and five Irrelevants (i.e. unknown faces) that were presented in the experiment.*

Additionally, subjects reported confidence ratings for Probes and Irrelevants that were not presented in the experiment (i.e. red 'Seen' and red 'Know' headings, on right-side).

The responses to both of these confidence ratings used a scale of 1 to 5, where 1 is "Never", 2 is "Once or twice", 3 is "Few times", 4 is "Many times" and 5 is "A lot".

As we were comparing celebrity faces with unknown faces (i.e. the Probe and Irrelevant conditions, in green), it was encouraging to discover that Probes were reported, on average, 60% of the time (M = 3.4; SD = 0.8771), which was six times more than Irrelevants that were reported 10% of the time (M = 1.4; SD = 0.532).

Note that both conditions (Probes and Irrelevants) were, in fact, presented an equal number of times. The mean confidence rating of the main comparison conditions, for all subjects, reveals a highly significant difference between the Probe (celebrity) faces and the Irrelevant (unknown) faces, using pair-wise comparison (M = 2, SD = 0.8629), t(13) = 8.6722, p < 0.0001, d = 2.7572).

(More detail in section 4.3.5)

	5	Probe		Probe Seen	Irrelevant		Irr. Seen	~Probe		~Pro. Seen	~Irrelevant		~Irr. Seen
	Block	Seen	Know		Seen	Know		Seen	Know		Seen	Know	
Subject 1	1	4	5		4	1		1	5		1	1	
	2	4	5		3	1		1	4		1	1	
	3	4	5	3.8	3	1	3.0	1	5	1.0	1	1	1.0
	4	4	4		3	1		1	4		1	1	
	5	3	5		2	1		1	5		1	1	
Subject 2	1	4	5		1	1		1	5		1	1	
	2	3	5		1	1		1	3		1	1	
	3	5	5	4.0	1	1	1.8	1	5	1.0	1	1	1.0
	4	5	4		1	1		1	5		1	1	
	5	3	5		5	1		1	4		1	1	
Subject 3	1	5	5		1	1		3	5		1	1	
	2	4	5		1	1		3	3		1	1	
	3	4	4	4.0	2	1	1.4	3	5	2.8	1	1	1.0
	4	4	5		1	1		3	4		1	1	
	5	3	5		2	1		2	5		1	1	
Subject 4	1	2	5		1	1		3	2		1	1	
	2	5	5		1	1		2	5		1	1	
	3	4	4	4.2	1	1	1.0	1	5	1.6	2	1	1.2
	4	5	5		1	1		1	4		1	1	
	5	5	5		1	1		1	5		1	1	
Subject 5	1	4	4		2	1		2	4		1	1	
	2	2	5		1	1		1	4		1	1	
	3	3	4	3.2	2	1	1.6	1	4	1.2	2	1	1.2
	4	3	3		2	1		1	5		1	1	
	5	4	5		1	1		1	5		1	1	
Subject 6	1	3	5		1	1		1	3		1	1	
	2	1	5		1	1		1	2		1	1	
	3	1	4	1.8	1	1	1.0	2	4	1.4	1	1	1.0
	4	2	5		1	1		1	5		1	1	
	5	2	5		1	1		2	5		1	1	
Subject 7	1	1	5		3	1		3	5		2	1	
	2	4	5		1	1		2	5		3	1	
	3	4	5	3.2	1	1	1.4	2	4	2.2	1	1	2.0
	4	3	4		1	1		2	5		1	1	
	5	4	5		1	1		2	2		3	1	
Subject 8	1	3	5		1	1		2	2		1	1	
	2	2	4		1	1		1	5		2	1	
	3	1	4	2.2	1	1	1.0	2	4	2.0	1	1	1.2
	4	2	4		1	1		2	5		1	1	
	5	3	5		1	1		3	5		1	1	
Subject 9	1	4	5		2	1		3	5		1	1	
	2	5	4		2	1		3	5		2	1	
	3	4	5	4.2	1	1	1.6	2	3	2.6	2	1	1.6
	4	4	4		2	1		2	4		1	1	
	5	4	5		1	1		3	5		2	1	
Subject 10	1	3	5		1	1		2	5		1	1	
	2	3	3		1	1		2	4		1	1	
	3	4	5	3.4	1	1	1.2	1	3	1.6	1	1	1.0
	4	3	3		2	1		2	5		1	1	
	5	4	5		1	1		1	5		1	1	
Subject 11	1	2	5		1	1		3	5		1	1	
	2	5	5		1	1		2	5		2	1	
	3	4	3	3.8	2	1	1.4	1	4	2.0	1	1	1.2
	4	4	3		2	1		2	5		1	1	
	5	4	5		1	1		2	5		1	1	
Subject 12	1	2	5		1	1		3	2		1	1	
	2	5	5		1	1		2	5		1	1	
	3	5	5	4.4	1	1	1.0	2	5	1.8	1	1	1.0
	4	5	4		1	1		1	4		1	1	
	5	5	5		1	1		1	5		1	1	
Subject 13	1	4	5		2	1		4	5		1	1	
	2	2	5		1	1		2	4		1	1	
	3	4	4	3.6	1	1	1.2	2	3	2.4	1	1	1.0
	4	5	4		1	1		2	4		1	1	
	5	3	5		1	1		2	5		1	1	
Subject 14	1	3	3		1	1		1	4		1	1	
	2	1	5		1	1		1	5		1	1	
	3	1	2	1.8	1	1	1.0	2	4	1.4	1	1	1.0
	4	2	3		1	1		1	5		1	1	
	5	2	5		1	1		2	5		1	1	
14	*70*	3.4	*4.5*	3.4	1.4	*1.0*	1.4	1.8	*4.3*	1.8	1.2	*1.0*	1.2
	Mean (%)	60.0%	88.2%	60.0%	10.0%	0.0%	10.0%	19.6%	83.6%	19.6%	4.3%	0.0%	4.3%
	Std. Dev.		0.8771			0.5320			0.5736			0.2920	

Pz	Block	Block-level N400f
Subject 1	1	0.227
	2	0.291
	3	0.165
	4	0.004
	5	0.031
Subject 2	1	0.012
	2	0.201
	3	0.101
	4	0.067
	5	0.083
Subject 3	1	0.014
	2	0.939
	3	0.032
	4	0.004
	5	0.002
Subject 4	1	0.342
	2	0.297
	3	0.09
	4	0.068
	5	0.046
Subject 5	1	0.484
	2	0.146
	3	0.258
	4	0.276
	5	0.232
Subject 6	1	0.113
	2	0.077
	3	0.059
	4	0.008
	5	0.316
Subject 7	1	0.095
	2	0.261
	3	0.708
	4	0.418
	5	0.009

Pz	Block	Block-level N400f
Subject 8	1	0.026
	2	0.583
	3	0.539
	4	0.134
	5	0.264
Subject 9	1	0.515
	2	0.039
	3	0.003
	4	0.017
	5	0.613
Subject 10	1	0.258
	2	0.78
	3	0.33
	4	0.345
	5	0
Subject 11	1	0.956
	2	0.898
	3	0.002
	4	0.57
	5	357
Subject 12	1	0.202
	2	0.117
	3	0.089
	4	0.174
	5	0.643
Subject 13	1	0.148
	2	0.859
	3	0.487
	4	0.045
	5	0.383
Subject 14	1	0.022
	2	0.507
	3	0.273
	4	0.004
	5	0.031
		20 of 70

Pz	Block	Block-level P600f
Subject 1	1	0.415
	2	0.03
	3	0.885
	4	0.724
	5	0.385
Subject 2	1	0.585
	2	0.54
	3	0.231
	4	0.206
	5	0.14
Subject 3	1	0.351
	2	0.236
	3	0.05
	4	0.061
	5	0.421
Subject 4	1	0.863
	2	0.076
	3	0.05
	4	0.092
	5	0.193
Subject 5	1	0.217
	2	0.84
	3	0.504
	4	0.216
	5	0.03
Subject 6	1	0.25
	2	0.3
	3	0.252
	4	0.595
	5	0.541
Subject 7	1	0.735
	2	0.008
	3	0.182
	4	0.401
	5	0

Pz	Block	Block-level P600f
Subject 8	1	0.768
	2	0.306
	3	0.312
	4	0.891
	5	0.796
Subject 9	1	0.149
	2	0.842
	3	0.79
	4	0.696
	5	0.2
Subject 10	1	0.271
	2	0.039
	3	0.093
	4	0.005
	5	0.194
Subject 11	1	0.311
	2	0.24
	3	0.701
	4	0.098
	5	0.093
Subject 12	1	0.724
	2	0.644
	3	0.383
	4	0.138
	5	0.001
Subject 13	1	0.072
	2	0.22
	3	0.07
	4	0.153
	5	0.01
Subject 14	1	0.872
	2	0.687
	3	0.037
	4	0.815
	5	0.81
		11 of 70

Table A.2 *– The first (Celebrity Faces) experiment's results, for 14 subjects. Statistical tests at block-level (i.e. blocks 1 to 5) of Pz electrode's N400f component is shown in the left table (20 of 70), and for P600f component is shown in the right table (11 of 70).*

As shown in section 4.4.4, subject-level statistical tests of Pz electrode's N400f component, resulted in 10 of 14 subjects (71%) achieving critical-significance (at alpha level $p < 0.05$), between the Probe and Irrelevant conditions. Furthermore, statistical tests of P600f component resulted in 7 of 14 subjects (50%) with p-values below our critical-significance. After combining each subject's p-values of the N400f and P600f components (as described in '3.3.4 – Combined probability test (Fisher's)'), all 14 subjects (100%) achieved Fisher combined levels at a minimal-significance (i.e. an alpha level of $p < 0.1$), as used in most of Farwell and Rosenfeld's deception detection studies (Farwell & Donchin, 1991); (Rosenfeld I. P., 2008). Out of these, 10 of 14 subjects (71.4%) achieved critical-significance level (i.e. $p < 0.05$, which is our preferred alpha level).

Appendix B

Table B.1 *– The second (Lecturer Faces) experiment's behavioral results, as reported by 14 subjects. Each subject reported a confidence rating (1 to 5) of seeing ('Seen' heading in green) and knowing ('Know' heading in green) the five Probes (i.e. celebrity faces) and five Irrelevants (i.e. unknown faces) that were presented in the experiment.*

Additionally, subjects reported confidence ratings for Probes and Irrelevants that were not presented in the experiment (i.e. red 'Seen' and red 'Know' headings, on right-side).

The responses to both of these confidence ratings used a scale of 1 to 5, where 1 is "Never", 2 is "Once or twice", 3 is "Few times", 4 is "Many times" and 5 is "A lot".

As we were comparing lecturer faces with unknown faces (i.e. the Probe and Irrelevant conditions, in green), it was encouraging to discover that Probes were were reported 33.9% of the time (M = 2.4; SD = 1.2504), which was seven times more than Irrelevants that were reported 4.8% of the time (M = 1.2; SD = 0.428).

Note that both conditions (Probes and Irrelevants) were, in fact, presented an equal number of times. The mean confidence rating of the main comparison conditions, for all subjects, reveals a highly significant difference between the Probe (known-lecturer) faces and the Irrelevant (unknown-lecturer) faces, using pair-wise comparison (M = 1.1714, SD = 1.5122), t(13) = 2.898, p = 0.0125, d = 1.2545).

(More detail in section 5.3.5)

	3	Probe		Probe Seen	Irrelevant		Irr. Seen	~Probe		~Pro. Seen	~Irrelevant		~Irr. Seen
	Block	Seen	Know		Seen	Know		Seen	Know		Seen	Know	
Subject 1	1	○ 1	● 5		◑ 2	○ 1		○ 1	● 5		○ 1	○ 1	
	2	○ 1	● 5		○ 1	○ 1		○ 1	○ 1		○ 1	○ 1	
	3	○ 1	● 5	1.0	○ 1	○ 1	1.3	○ 1	◑ 3	1.0	○ 1	○ 1	1.0
Subject 2	1	● 5	● 5		○ 1	○ 1		○ 1	● 5		○ 1	○ 1	
	2	◔ 2	● 5		○ 1	○ 1		○ 1	● 5		○ 1	○ 1	
	3	● 5	● 5	4.0	○ 1	○ 1	1.0	○ 1	● 5	1.0	○ 1	○ 1	1.0
Subject 3	1	○ 1	● 5		○ 1	○ 1		○ 1	● 5		○ 1	○ 1	
	2	○ 1	● 5		○ 1	○ 1		○ 1	◑ 3		○ 1	○ 1	
	3	◑ 3	● 5	1.7	○ 1	○ 1	1.0	◔ 2	● 5	1.3	○ 1	○ 1	1.0
Subject 4	1	● 5	● 5		○ 1	○ 1		○ 1	● 5		○ 1	○ 1	
	2	◕ 4	● 5		○ 1	○ 1		○ 1	● 5		○ 1	○ 1	
	3	◑ 3	◕ 4	4.0	○ 1	○ 1	1.0	◔ 2	○ 1	1.3	○ 1	○ 1	1.0
Subject 5	1	○ 1	● 5		○ 1	○ 1		○ 1	◕ 4		○ 1	○ 1	
	2	◕ 4	● 5		○ 1	○ 1		○ 1	◕ 4		○ 1	○ 1	
	3	○ 1	● 5	2.0	○ 1	○ 1	1.0	○ 1	◕ 4	1.0	○ 1	○ 1	1.0
Subject 6	1	◕ 4	● 5		○ 1	○ 1		○ 1	◑ 3		○ 1	○ 1	
	2	○ 1	● 5		○ 1	○ 1		○ 1	◔ 2		○ 1	○ 1	
	3	○ 1	● 5	2.0	○ 1	○ 1	1.0	○ 1	◕ 4	1.0	○ 1	○ 1	1.0
Subject 7	1	◑ 3	● 5		○ 1	○ 1		○ 1	● 5		○ 1	○ 1	
	2	◔ 2	● 5		○ 1	○ 1		○ 1	● 5		○ 1	○ 1	
	3	◔ 2	● 5	2.3	○ 1	○ 1	1.0	○ 1	● 5	1.0	● 2	○ 1	1.3
Subject 8	1	◕ 4	● 5		○ 1	○ 1		○ 1	◔ 2		○ 1	○ 1	
	2	○ 1	◕ 4		○ 1	○ 1		○ 1	● 5		○ 1	○ 1	
	3	○ 1	◕ 4	2.0	○ 1	○ 1	1.0	○ 1	◕ 4	1.0	○ 1	○ 1	1.0
Subject 9	1	○ 1	● 5		◑ 2	○ 1		○ 1	● 5		● 2	○ 1	
	2	○ 1	◕ 4		● 3	○ 1		○ 1	● 5		○ 1	○ 1	
	3	○ 1	● 5	1.0	◑ 2	○ 1	2.3	○ 1	● 5	1.0	○ 1	○ 1	1.3
Subject 10	1	○ 1	● 5		○ 1	○ 1		○ 1	● 5		● 2	○ 1	
	2	○ 1	◑ 3		○ 1	○ 1		○ 1	◕ 4		○ 1	○ 1	
	3	○ 1	● 5	1.0	○ 1	○ 1	1.0	◔ 2	◑ 3	1.3	● 2	○ 1	1.7
Subject 11	1	● 5	● 5		○ 1	○ 1		◔ 2	● 5		○ 1	○ 1	
	2	◑ 3	● 5		○ 1	○ 1		○ 1	● 5		○ 1	○ 1	
	3	○ 1	◕ 4	3.0	○ 1	○ 1	1.0	○ 1	● 5	1.3	○ 1	○ 1	1.0
Subject 12	1	○ 1	● 5		○ 1	○ 1		◑ 3	◔ 2		○ 1	○ 1	
	2	○ 1	● 5		◑ 2	○ 1		○ 1	● 5		○ 1	○ 1	
	3	○ 1	● 5	1.0	● 3	○ 1	2.0	○ 1	● 5	1.7	○ 1	○ 1	1.0
Subject 13	1	● 5	● 5		○ 1	○ 1		○ 1	● 5		○ 1	○ 1	
	2	◕ 4	● 5		○ 1	○ 1		○ 1	◕ 4		○ 1	○ 1	
	3	○ 1	◕ 4	3.3	○ 1	○ 1	1.0	○ 1	◑ 3	1.0	○ 1	○ 1	1.0
Subject 14	1	● 5	● 5		○ 1	○ 1		○ 1	● 5		○ 1	○ 1	
	2	● 5	● 5		○ 1	○ 1		○ 1	● 5		○ 1	○ 1	
	3	◕ 4	◕ 4	4.7	○ 1	○ 1	1.0	○ 1	◕ 4	1.0	○ 1	○ 1	1.0
	14	42	2.4	4.8	2.4	1.2	1.0	1.2	1.1	4.2	1.1	1.1	1.0	1.1
	Mean (%)		33.9%	94.6%	33.9%	4.8%	0.0%	4.8%	3.6%	79.2%	3.6%	2.4%	0.0%	2.4%
	Std. Dev.				1.2504			0.4280			0.2154			0.2037

Table B.2 *– The second (Lecturer Faces) experiment's results, for 14 subjects. Statistical tests at block-level (i.e. blocks 1 to 3) of Pz electrode's P600f component is shown (11 of 42).*

To achieve these results, we began by exploring the presence of the P600f component within each of the three items of every subject (i.e. 3 experimental blocks for 14 subjects, equalling 42 item-blocks). Having independently searched for the P600f component's 100ms aERPt time window (i.e. highest positive deflection, within the a-priori search area that spans from the time range of 300ms to 900ms), we performed permutation tests for each individual block. Consequently, three 'by-item' p-values were obtained for each subject (i.e. one for each block's familiar lecturer), resulting in significant difference between the Probe and Irrelevant conditions for 11 of 42 blocks (26.2%).

As shown in section 5.4.3, subject-level statistical tests of Pz electrode's P600f component, resulted in 11 of 14 subjects (78.6%) achieving critical-significance (alpha level 0.05), between the Probe and Irrelevant conditions. Note that this is an improvement on both of the previous experiments' results (to wit: the celebrity faces experiments achieved 7 of 14 subjects, and the concelaed lecturer faces achieved 8 of 14 subjects, on the P600f contrast).

Pz	Block	Block-level P600f
Subject 1	1	0.383
	2	0.063
	3	0.227
Subject 2	1	0.375
	2	0.898
	3	0
Subject 3	1	0.354
	2	0.711
	3	0.29
Subject 4	1	0.036
	2	0.288
	3	0.82
Subject 5	1	0.71
	2	0.064
	3	0.544
Subject 6	1	0.16
	2	0.673
	3	0.093
Subject 7	1	0.007
	2	0.051
	3	0.012

Pz	Block	Block-level P600f
Subject 8	1	0.857
	2	0.159
	3	0.522
Subject 9	1	0.025
	2	0.063
	3	0.179
Subject 10	1	0.888
	2	0.003
	3	0.511
Subject 11	1	0
	2	0.017
	3	0.424
Subject 12	1	0.454
	2	0.606
	3	0.366
Subject 13	1	0
	2	0.14
	3	0.5
Subject 14	1	0
	2	0.083
	3	0.005
		11 of 42

Appendix C

Table C.1 *– The third (Revealed Lecturer Faces) experiment's behavioral results, as reported by 14 subjects. Each subject reported a confidence rating (1 to 5) of seeing ('Seen' heading in green) and knowing ('Know' heading in green) the five Probes (i.e. celebrity faces) and five Irrelevants (i.e. unknown faces) that were presented in the experiment.*

Additionally, subjects reported confidence ratings for Probes and Irrelevants that were not presented in the experiment (i.e. red 'Seen' and 'Know' headings).

The responses to both of these confidence ratings used a scale of 1 to 5, where 1 is "Never", 2 is "Once or twice", 3 is "Few times", 4 is "Many times" and 5 is "A lot".

As we were comparing lecturer faces with unknown faces (i.e. the Probe and Irrelevant conditions, in green), it was encouraging to discover that Probes were were reported 50% of the time (M = 3.0; SD = 0.6918), which was nearly six times more than Irrelevants that were reported 8.9% of the time (M = 1.4; SD = 0.4022).

Note that both conditions (Probes and Irrelevants) were, in fact, presented an equal number of times. The mean confidence rating of the main comparison conditions, for all subjects, reveals a highly significant difference between the Probe (familiar lecturer) faces and the Irrelevant (unknown lecturer) faces, using pair-wise comparison (M = 1.6571, SD = 0.8582), t(13) = 7.2251, p < 0.0001, d = 2.9518).

(More detail in section 7.3.5)

	5	Probe		Probe Seen	Irrelevant		Irr. Seen	~Probe		~Pro. Seen	~Irrelevant		~Irr. Seen
	Block	Seen	Know		Seen	Know		Seen	Know		Seen	Know	
Subject 1:	1	3	5		1	1		1	5		1	1	
	2	5	5		1	1		1	5		1	1	
	3	2	4	3.3	2	1	1.3	2	5	1.3	1	1	1.0
	4												
	5												
Subject 2:	1	5	5		2	1		1	5		1	1	
	2	4	5		1	1		3	3		1	1	
	3	3	4	4.0	2	1	1.7	2	5	2.0	1	1	1.0
	4												
	5												
Subject 3:	1	1	5		1	1		1	5		1	1	
	2	4	5		1	1		2	4		1	1	
	3	2	5	2.3	1	1	1.0	2	4	1.7	1	1	1.0
	4												
	5												
Subject 4:	1	5	4		1	1		1	4		1	1	
	2	2	3		2	1		1	4		1	1	
	3	1	3	2.7	1	1	1.3	1	4	1.0	1	1	1.0
	4												
	5												
Subject 6:	1	5	5		1	1		1	5		1	1	
	2	2	4		1	1		2	4		2	1	
	3	1	4	2.7	1	1	1.0	2	5	1.7	1	1	1.3
	4												
	5												
Subject 7:	1	5	5		1	1		3	5		1	1	
	2	3	4		1	1		3	5		3	1	
	3	1	2	3.0	2	1	1.3	1	5	2.3	1	1	1.7
	4												
	5												
Subject 8:	1	5	5		1	1		1	4		1	1	
	2	1	4		2	1		1	3		3	1	
	3	5	5	3.7	1	1	1.3	1	3	1.0	1	1	1.7
	4												
	5												
Subject 9:	1	1	4		1	1		1	4		1	1	
	2	4	4		1	1		1	3		1	1	
	3	5	4	3.3	1	1	1.0	1	4	1.0	1	1	1.0
	4												
	5												
Subject 10:	1	1	3		1	1		1	5		1	1	
	2	1	3		2	1		1	4		1	1	
	3	3	5	1.7	1	1	1.3	1	4	1.0	1	1	1.0
	4												
	5												
Subject 11:	1	2	4		2	1		1	5		1	1	
	2	1	3		3	1		2	1		1	1	
	3	3	5	2.0	2	1	2.3	1	5	1.3	1	1	1.0
	4												
	5												
Subject 12:	1	4	5		1	1		1	5		1	1	
	2	5	5		1	1		1	5		1	1	
	3	3	5	4.0	1	1	1.0	1	5	1.0	2	1	1.3
	4												
	5												
Subject 13:	1	3	3		1	1		1	5		3	1	
	2	1	4		3	1		1	4		3	1	
	3	5	5	3.0	2	1	2.0	1	4	1.0	1	1	2.3
	4												
	5												
Subject 14:	1	5	5		2	1		1	5		1	1	
	2	4	5		1	1		2	4		1	1	
	3	1	2	3.3	1	1	1.3	1	5	1.3	1	1	1.0
	4												
	5												
Subject 15:	1	1	2		1	1		1	5		1	1	
	2	3	5		1	1		1	4		1	1	
	3	5	3	3.0	1	1	1.0	1	4	1.0	1	1	1.0
	4												
	5												
14	42	3.0	4.2	3.0	1.4	1.0	1.4	1.3	4.3	1.3	1.2	1.0	1.2
Mean (%)		50.0%	79.2%	50.0%	8.9%	0.0%	8.9%	8.3%	83.3%	8.3%	6.0%	0.0%	6.0%
Std. Dev.				0.6918			0.4022			0.4336			0.4015

Pz		Block-level		Subject-level
	Block	P600f	METHODS used:	P600f
Subject 1:	1	0.197	**Subject 1:**	
	2	0	d=Blocks 1+2+3:	0
	3	0.952	a=Blocks 4+5:	0
	4	0.002	b=Blocks 2+4+5:	0
	5	0.008	c=Blocks 1+2+3+4+5:	0
Subject 2:	1	0.212	**Subject 2:**	
	2	0.496	d=Blocks 1+2+3:	0.235
	3	0.966	a=Blocks 4+5:	0
	4	0	b=Blocks 1+4+5:	0
	5	0.004	c=Blocks 1+2+3+4+5:	0
Subject 3:	1	0.031	**Subject 3:**	
	2	0.011	d=Blocks 1+2+3:	0.008
	3	0.013	a=Blocks 4+5:	0.657
	4	0.343	b=Blocks 3+4+5:	0.012
	5	0.809	c=Blocks 1+2+3+4+5:	0.047
Subject 4:	1	0.009	**Subject 4:**	
	2	0.937	d=Blocks 1+2+3:	0.088
	3	0.674	a=Blocks 4+5:	0.004
	4	0.05	b=Blocks 1+4+5:	0
	5	0.047	c=Blocks 1+2+3+4+5:	0.002
Subject 6:	1	0.009	**Subject 6:**	
	2	0.7	d=Blocks 1+2+3:	0.202
	3	0.562	a=Blocks 4+5:	0.244
	4	0.123	b=Blocks 1+4+5:	0.029
	5	0.499	c=Blocks 1+2+3+4+5:	0.338
Subject 7:	1	0	**Subject 7:**	
	2	0.196	d=Blocks 1+2+3:	0
	3	0.182	a=Blocks 4+5:	0.002
	4	0.042	b=Blocks 1+4+5:	0
	5	0.01	c=Blocks 1+2+3+4+5:	0
Subject 8:	1	0.024	**Subject 8:**	
	2	0.616	d=Blocks 1+2+3:	0.03
	3	0.035	a=Blocks 4+5:	0
	4	0.046	b=Blocks 1+4+5:	0.001
	5	0.045	c=Blocks 1+2+3+4+5:	0.001

Pz		Block-level		Subject-level
	Block	P600f	METHODS used:	P600f
Subject 9:	1	0.331	**Subject 9:**	
	2	0.629	d=Blocks 1+2+3:	0.968
	3	0.047	a=Blocks 4+5:	0.818
	4	0.281	b=Blocks 3+4+5:	0.043
	5	0.739	c=Blocks 1+2+3+4+5:	0.885
Subject 10:	1	0.068	**Subject 10:**	
	2	0.265	d=Blocks 1+2+3:	0.026
	3	0.059	a=Blocks 4+5:	0
	4	0.045	b=Blocks 1+4+5:	0
	5	0.004	c=Blocks 1+2+3+4+5:	0.004
Subject 11:	1	0.065	**Subject 11:**	
	2	0.148	d=Blocks 1+2+3:	0.122
	3	0.773	a=Blocks 4+5:	0.996
	4	0.981	b=Blocks 1+4+5:	0.633
	5	0.976	c=Blocks 1+2+3+4+5:	0.728
Subject 12:	1	0.001	**Subject 12:**	
	2	0	d=Blocks 1+2+3:	0
	3	0.706	a=Blocks 4+5:	0.005
	4	0.01	b=Blocks 2+4+5:	0
	5	0.114	c=Blocks 1+2+3+4+5:	0
Subject 13:	1	0.294	**Subject 13:**	
	2	0.321	d=Blocks 1+2+3:	0.037
	3	0.117	a=Blocks 4+5:	0
	4	0.069	b=Blocks 3+4+5:	0
	5	0.002	c=Blocks 1+2+3+4+5:	0
Subject 14:	1	0.005	**Subject 14:**	
	2	0.59	d=Blocks 1+2+3:	0.046
	3	0.99	a=Blocks 4+5:	0
	4	0	b=Blocks 1+4+5:	0
	5	0.043	c=Blocks 1+2+3+4+5:	0
Subject 15:	1	0.305	**Subject 15:**	
	2	0.007	d=Blocks 1+2+3:	0
	3	0	a=Blocks 4+5:	0
	4	0	b=Blocks 3+4+5:	0
	5	0	c=Blocks 1+2+3+4+5:	0
	14	70	**33 of 70**	*4 Methods per Subject:* **43 of 56**

Table C.2 *– The third (Revealed Lecturer Faces) experiment's results, for 14 subjects. Statistical tests at block-level (i.e. blocks 1 to 5) of Pz electrode's P600f component (33 of 70), as well as, the method-specific results (i.e. d=Decider, a = Abandoned, b = Biased and c = Combined).*

Note that the critical stimulus in Part I (i.e. from blocks 1, 2 or 3) that showed the highest significance (highlighted in yellow) using online statistical tests, was promoted to Part II, so that its paired Probe/Irrelevant conditions were re-used in blocks 4 and 5.

As shown in section 7.4.3, subject-level statistical tests of Pz electrode's P600f component, resulted in 11 of 14 subjects (78.6%) achieving critical-significance (alpha level 0.05), between the Probe and Irrelevant conditions. Note that this is an improvement on both of the previous experiments' results (to wit: the celebrity faces experiments achieved 7 of 14 subjects, and the concelaed lecturer faces achieved 8 of 14 subjects, on the P600f contrast).

Appendix D

Glossary of common terms

AGAT – Aggregated Grand Average of Trials is a data-driven (safe) window selection for Group-level analysis (as justified by Brooks, et al., 2017). For more information refer to section 3.3.3.2.

aERPt – The aggregated ERP of all Trials is a data-driven (safe) window selection for subject-level analysis (as justified by Brooks, et al., 2017). For more information refer to section 3.3.3.1.

Alpha rhythm - Rhythm at 8–13 Hz inclusive occurring during wakefulness over the posterior regions of the head, generally with maximum amplitudes over the occipital areas.

Artifact - (1) A physiological potential difference due to an extracerebral source present in EEG recordings, such as eye blinks and movements, electrocardiogram (ECG) or muscle contractions (EMG). (2) A modification of the EEG caused by extracerebral factors, such as instrumental distortion or malfunction, movement of the patient, or ambient electrical noise.

Asynchrony - The noncoherent occurrence of EEG activities over regions on the same or opposite sides (hemispheres) of the head. For example, two similar waveforms occurring at separate electrodes or channels, but not simultaneously due to a time lag between the channels.

Beta rhythm - Any EEG rhythm between 14 and 30 Hz (wave duration 33–72 ms). Most characteristically recorded over the fronto-central regions of the head during wakefulness.

CIT – Concealed Information Test detects a person's guilty knowledge of a crime, unlike the traditional polygraph Comparison Question Test that assesses deception to direct, accusatory questions.

Electrocorticography (ECoG) - Technique of recording electrical activity of the brain by means of electrodes applied over or implanted into the cerebral cortex.

Electrode - A conducting device applied over or inserted in a region of the scalp or brain.

Electroencephalography (EEG) - An electrophysiological monitoring method to record electrical activity of the brain. EEG is typically non-invasive, with the electrodes placed along the scalp, although invasive electrodes are sometimes used, as in electrocorticography (a.k.a. intracranial EEG).

Epoch - EEG segment with a defined duration. Duration of epochs is determined arbitrarily but should be specified.

ERSP – Event Related Spectral Perturbations, as used in single Trial (Frequency domain) analysis, measures power changes of EEG data, reflecting synchronisation of phase across trials.

Event-related potential (ERP) – Time-locked and averaged EEG activity (or ERP) helps capture neural activity related to both sensory and cognitive processes.

Frequency - Number of complete cycles of repetitive waves in one second. Measured in cycles per second (c/s) or Hertz (Hz).

Finge-P3 – A countermeasure resistant method (developed by Bowman, et. al., 2013) that uses RSVP to present stimuli at the fringe of human awareness, as used in our CIT experiments.

High frequency filter (or low pass filter)- A circuit that reduces the sensitivity of the EEG signals to relatively high frequencies (for example, above 45 Hz).

ITC – Inter-Trial Coherence, as used in single Trial (Frequency domain) analysis, measures coherence changes of EEG data, which reflects the synchronisation of phase across trials.

Low frequency filter (high pass filter) - A circuit that reduces the sensitivity of the EEG signal to relatively low frequencies (for example, below 0.5 Hz).

Magnetoencephalography (MEG) - Recording of magnetic fields generated from the cortical neurons.

Notch filter - A filter that selectively attenuates a very narrow frequency band, thus producing a sharp notch in the frequency response of an EEG signal (for example, between 7 and 10 Hz).

Phase - Time or polarity relationships between a point on a wave displayed in a derivation and the identical point on the same wave recorded simultaneously in another derivation.

Potential - Electrical activity (waveforms) generated by the nervous system.

Potential field - Amplitude distribution of the negative and positive potentials of an EEG signal at the surface of the head, or cerebral cortex or in the depth of the brain, measured at a given instant in time.

RSVP – Rapid Serial Visual Presentation is a scientific method for studying the timing of vision. In RSVP, a sequence of stimuli are shown (at a rapid rate) to an observer, at one location in their visual field.

SNR – Signal to Noise ratio is the ratio of the power in the signal of interest to the total noise power over the signal's bandwidth. Any noise present outside of the signal bandwidth can be filtered out without removing any signal power, so it can be ignored.

Ten-twenty (10–20) system - System of standardised scalp electrode placement recommended by the International Federation of Clinical Neurophysiology. According to this system, the placement of electrodes is determined by measuring the head from 4 external landmarks and taking 10 or 20 percentiles of these measurements.

Voltage - The difference in electric potential between two points (units: volts).

Wave - Any change of the potential difference between pairs of electrodes in EEG recording, which may arise in the brain (an EEG wave) or outside of it (i.e., extracerebral potential).

Bibliography

Abootalebi, V., Moradi, M., & Khalilzadeh, M. (2006). A comparison of methods for ERP assessment in a P300-based GKT. International Journal of Psychophysiologu. 62:309-320.

Adelson, R. (2004). The polygraph in doubt. APA Monitor, volume 35, no. 7, page 71.

Alonso-Nanclares, L., Gonzalez-Soriano, J., Rodriguez, J. R., & DeFelipe, J. (2008). Gender differences in human cortical synaptic density. Proceedings of the National Academy of Sciences, 105(38), 14615-14619.

Alsufyani, A., Hajilou, O., Zoumpoulaki, A., Filetti, M., Solomon, J., Gibson, SJ., & Bowman, H. (2019). Breakthrough Percepts of Famous Faces. Psychophysiology, 56 (1). e13279. ISSN 0048-5772. (doi:10.1111/psyp.13279).

APA. (2004). The Truth About Lie Detectors (aka Polygraph Tests). apa.org. American Psychological Association. Retrieved from https://www.apa.org/research/action/polygraph.

Atwood, H., & MacKay, W. (1989). Essentials of neurophysiology, B.C. Decker, Hamilton, Canada. ISBN 101556640552.

Azevedo, F., Carvalho, L., Grinberg, L., Farfel, J., Ferretti, R., Leite, R., & Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. Journal of Comparative Neurology, 513(5), 532-541. doi:10.1002/cne.21974.

Ben-Shakhar, G., & Elaad, E. (2003). The Validity of Psychophysiological Detection of Information With the Guilty Knowledge Test: A Meta-Analytic Review. The Journal of applied psychology. 88. 131-51. 10.1037/0021-9010.88.1.131.

Bentin, S., & Deouell, LY. (2000). Structural encoding and identification in face processing: ERP evidence for separate mechanisms. Cognitive Neuropsychology, 17, 35-54.

Bentin, S., Truett, A., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. Journal of Cognitive Neuroscience, 8, 551-565.

Berger, H. (1929). Über das elektroencephalogramm des menschen. Psychiatry, 87.

Bindemann, M., Burton, A. M., Leuthold, H., & Schweinberger, S. R. (2008). Brain potential correlates of face recognition: geometric distortions and the N250r brain response to stimulus repetitions. Psychophysiology, 45(4), 535e544.

Bond, CF., & DePaulo, BM. (2006). Accuracy of Deception Judgments. Pers Soc Psychol. vol. 10 no. 3 214-234.

Bowman, H., & Wyble, B. (2007). The simultaneous type, serial token model of temporal attention and working memory. Psychological review 114:38–70.

Bowman, H., Filetti, M., Alsufyani, A., Janssen, & Su, L. (2014). Countering countermeasures: detecting identity lies by detecting conscious breakthrough. PLoS One 9(3):e90595.

Bowman, H., Filetti, M., Janssen, D., Su, L., Alsufyani, A., & Wyble, B. (2013). Subliminal Salience Search Illustrated: EEG Identity and Deception Detection on the Fringe of Awareness. 10.1371/journal.pone.0054258.

Bresadola, M. (1998). Medicine and science in the life of Luigi Galvani. Brain Research Bulletin 46 (5): 367–380.

Bressler, S., & Ding, M. (2006). Event-Related Potentials. Wiley encyclopedia of biomedical. https://doi.org/10.1002/9780471740360.ebs0455.

Brodmann, K. (1909). Vergleichende lokalisationslehre der grobhirnrinde. Barth, Leipzig.

Brooks, J., Zoumpoulaki, A., & Bowman, H. (2017). Data-driven region-of-interest selection without inflating Type I error rate: Safe data-driven ROI selection. Psychophysiology. 54. 100-113. 10.1111/psyp.12682.

Bruce, V., & Young, A. (1986). Understanding face recognition. British Journal of Psychology, 77, 305-327.

Carey, BY., & Diamon, R. (1977). From piecemeal to configurational representations of faces. Science, 195(4275), 312-314.

Caton, R. (1875). The Electric Currents of the Brain. The Electric Currents of the Brain.

Chun, M., & Potter, M. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. Journal of Experimental Psychology: Human Perception and Performance. 21:109–127.

Clark, VP., Fan, S., & Hillyard, SA. (1995). Identification of early visually evoked potential generators by retinotopic and topographic analyses. Human Brain Mapping. 1995;2:170–187.

Craston, P., Wyble, B., Chennu, S., & Bowman, H. (2009). The attentional blink reveals serial working memory encoding: Evidence from virtual and human event-related potentials. Journal of Cognitive Neuroscience. 2009;21(3):550–566. doi: 10.1162/jocn.2009.21036.

Craston, Wyble & Bowman, H. (2006). An EEG study of masking effects in RSVP. Journal of Vision, 6(6), 1016-1016. doi:10.1167/6.6.1016.

Crevits, L. (1982). On and Off Contribution to the Combined Occipital On-Off Response to a Pattern Stimulus. Ophthalmologica 1982;184:169-173. doi: 10.1159/000309201.

Curan, T., & Hancock, J. (2007). The FN400 indexes familiarity-based recognition of faces. Neuroimage. Jun 2007; 36(2): 464–471.

Davis, P. (1939). Effects of acoustic stimuli on the waking human brain. Journal of Neurophysiology. 1939;2:494–499.

Sun, D., Chan, CC., & Lee, TM. (2012). Identification and Classification of Facial Familiarity in Directed Lying: An ERP Study; PLOS 10.1371/journal.pone.0031250.

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

Dien, J., & Santuzzi, AM. (2004). Application of repeated measures ANOVA to high-density ERP datasets: A review and tutorial. In: Handy, T. (Ed.), Event-Related Potentials: A Methods Handbook, Cambridge, Mass: MIT Press.

Dietrich, A., Hu, X., & Rosenfeld, J. (2014). The effects of sweep numbers per average and protocol type on the accuracy of the p300-based concealed information test. Applied psychophysiology and biofeedback - DOI:10.1007/s10484-014-9244-y.

Donchin, E. (1969). Average Evoked Potentials, Methods, Results, and Evaluations. U.S. Government Printing Office; Washington, D.C.

Donchin, E., Spencer, K.M., & Wijesinghe, R. (2000). The mental prosthesis: Assessing the speed of a P300-based brain-computer interface. Ieee Transactions on Rehabilitation Engineering, 8, 174-179.

Duncan-Johnson, CC. & Kopell, BS. (1981). The Stroop effect: Brain potentials localize the source of interference. Science, 214, 938–940.

Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity; Electroencephalogr Clin Neurophysiol. 1996 Sep; 99(3):225-34.

Eimer, M. (2000). Event-related brain potentials distinguish processing stages involved in face perception and recognition. Clinical Neurophysiology, 111, 694-705. 10.1016/S1388-2457(99)00285-0.

Ellis, A., Young, A., & Flude, B. (1990). Repetition priming and face processing: priming occurs within the system that responds to the identity of a face. The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 42(3), 495e512.

Fabiani, M., Karis, D., Coles, M. G., & Donchin, E. (1983). P300 and recall in an incidental memory paradigm. Psychophysiology, 20(4), 439–439.1010.

Fantz, R. (1963). Pattern vision in newborn infants. Science, 140, 296-297.

Farroni, T. C. (2002). Eye contact detection in humans from birth. Proceedings of the National Academy of Sciences (USA), 99, 9602-9605.

Farwell, L. & Donchin, E. (1991). The truth will out: Interrogative polygraphy (‘‘lie detection’’) with event-related brain potentials. Psychophysiology, 28, 531–547.

Farwell, L., & Donchin, E. (1986). The “brain detector”: P300 in the detection of deception. Psychophysiology. 23:434.

Fisher, R. (1932). Statistical Methods for Research Workers. Oliver and Boyd, Edinburgh. https://doi.org/10.1002/qj.49708235130.

Ford, JM., Mathalon, DH., Marsh, L., Faustman, WO., Harris, D., Hoff, AL., Beal, M., & Pfefferbaum, A. (1999). P300 amplitude is related to clinical state in severely and moderately ill patients with schizophrenia. Biological Psychiatry, 46, 94–101.

Freeman, F. (1994). Galen's ideas on neurological function. Department of Neurology, Vanderbilt University, Nashville, Tennessee 37203, USA.

Freunberger, RK. (2007). Visual P2 component is related to theta phase-locking. Neuroscience Letters, 426, 181-186.

Friston, K., Fletcher, P., Josephs, O., Holmes, A., Rugg, M., & Turner, R. (1998). Event-related fMRI: characterizing differential responses. Neuroimage. DOI: 10.1006/nimg.1997.0306.

Friston, K., Rotshtein, P., Geng, J., Sterzer, P., & Henson, R. (2006). A critique of functional localisers. Neuroimage. 30(4):1077-87.

Fuentemilla, L. (2008). Theta EEG oscillatory activity and auditory change detection. Brain Research, 1220, 93-101.

Gaspar, P., Ruiz, S., Zamorano, F., Altayó, M., Pérez, C., Bosman, C., & Aboitiz, F. (2011). P300 amplitude is insensitive to working memory load in schizophrenia. BMC Psychiatry. 11:29. doi: 10.1186/1471-244X-11-29.

Gauthier, I., & Tarr, M. (1997). Becoming a "Greeble" expert: Exploring mechanisms for face recognition. Vision Research, 37(12), 1673-1682.

Gold, J., Mundy, P., & Tjan, B. (2012). The perception of a face is no more than the sum of its parts". Psychological Science 23 (4): 427–434.

Goshen-Gottstein, Y., & Ganel, Y. (2000). Repetition priming for familiar and unfamiliar faces in a sex-judgment task: evidence for a common route for the processing of sex and identity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 26(5), 1198e1214.

Gosling, A., & Eimer, M. (2011). An event-related brain potential study of explicit face recognition. Neuropsychologia, 49(9), 2736e2745.

Grüter, T., Grüter, M., & Carbon, CC. (2008). Neural and genetic foundations of face recognition and prosopagnosia. J Neuropsychol 2 (1): 79–97. doi:10.1348/174866407X231001.

Guillery, R., & Sherman, S. (2002). Thalamic Relay Functions and Their Role in Corticocortical Communication: generalizations from the visual system. VOLUME 33, ISSUE 2, P163-175.

Hayasaka, S. (2004). ombining voxel intensity and cluster extent with a permutation test framework. doi:10.1016/j.neuroimage.2004.04.035.

Heinze, HJ., Mangun, GR., Burchert, W., Hinrichs, H., Scholz, M., Münte, T., Gos, A., Scherg, M., Johannes, S., & Hundeshagen, H. (1994). Combined spatial and temporal imaging of brain activity during visual selective attention in humans. Nature. 1994 Dec 8; 372(6506):543-6.

Hodgkin, A., & Huxley, A. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. J Physiol. 1952 Aug 28; 117(4): 500–544.

Honts, CR., & Kircher, JC. (1984). Mental and physical countermeasures reduce the accuracy of polygraph tests. Journal of Applied Psychology, 79(2), 252–259.

Hubel, D., & Wiesel, T. (1959). Receptive fields of single neurones in the cat's striate cortex. J Physiol. 148(3): 574–591.

Ichisugi, Y. (2007). A Cerebral Cortex Model that Self-Organizes Conditional Probability Tables and Executes Belief Propagation. International Joint Conference on Neural Networks, Orlando, FL, 2007, pp. 178-183.

Immerwahr, H. (1985). Herodotus, in The Cambridge History of Classical Greek Literature: Greek Literature, P. Easterling and B. Knox (eds), Cambridge University Press. ISBN:9781139054874.

Ishai, A., Underleider, L., Martin, A., Schouten, J., & Haxby, J. (1999). Distributed representation of objects in the human. Neurobiology. Vol. 96, pp. 9379–9384.

Jasper, H. (1958). The ten-twenty electrode system of the International Federation. Electroencephalography and Clinical Neurophysiology, 10, 371-375.

Johnson, M., & Morton, J. (1991). Biology and cognitive development. The case of face recognition. Oxford: Blackwell. ISBN: 978-0-631-17454-7.

Johnson, M., & Rosenfeld, J. (1992). Oddball-evoked P300-based method of deception detection in the laboratory II: Utilization of non-selective activation of relevant knowledge. International Journal of Psychophysiology, 12(3), 289–306.

Kanwisher, NG. (1990). Repetition Blindness: Levels of Processing. Journal of Experimental Psychology, Vol. 16, No. 1, 30-47.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 361(1476), 2109–2128. doi:10.1098/rstb.2006.1934.

Kaufmann, T., Schulz, S., Grünzinger, C., & Kübler, A. (2011). Flashing characters with famous faces improves ERP-based brain-computer interface performance. Journal of neural engineering. 8. 056016. 10.1088/1741-2560/8/5/056016.

Kilner, J. (2013). Bias in a common EEG and MEG statistical analysis and how to avoid it. Clinical Neurophysiology, Volume 124, Issue 10, Pages 2062–2063 .

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. Psychophysiology, 38(03), 557-577.

Kozel, F., Padgett, T., & George, M. (2004). A Replication Study of the Neural Correlates of Deception. Behavioral Neuroscience, 118(4): 852-56.

Kriegeskorte, N., Simmons, WK., Bellgowan, PS., & Baker, Cl. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. Nature neuroscience, 12(5), 535–540. doi:10.1038/nn.2303.

Kutas, M., & Hillyard, S. A. (1980). "Reading senseless sentences: Brain potentials reflect semantic incongruity". Science 207: 203–208. doi:10.1126/science.7350657.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time; Science 197(4305):792-5.

Labkovsky, E., & Rosenfeld, J. (2012). The P300-Based, Complex Trial Protocol for Concealed Information Detection Resists Any Number of Sequential Countermeasures Against Up to Five Irrelevant Stimuli. 37: 1. https://doi.org/10.1007/s10484-011-9171-0. Appl Psychophysiol Biofeedback.

Larson, JA. (1932). Lying and its detection. Lying and its detection: A study of deception and deception tests (p. 99). Chicago, IL: University of Chicago Press.

Larson, JA. (1923). The Cardio-Pneumo-Psychogram in Deception. Journal of Experimental Psychology, 6(6), 420.

Lawrence, D. (1971). Two studies of visual search for word targetswith controlled rates of presentation. Perception and Psychophysics 10:85-89.

Lefebvre, CM. (2007). Determining eyewitness identification accuracy using event-related brain potentials (ERPs). Psychophysiology, 44 6, 894-904.

Luck, S. (2005). An introduction to the event-related potential technique. MIT Press. ISBN: 9780262525855.

Luck, S. (2014). An Introduction to the Event-Related Potential Technique (2nd Ed.). Cambridge. ISBN: 9780262525855.

Luck, S., Hillyard, S., Mouloua, M., & Hawkins, H. (1996). Mechanisms of visual-spatial attention: resource allocation or uncertainty reduction? J Exp Psychol Hum Percept Perform. DOI: 10.1037//0096-1523.22.3.725.

Luck, Vogel & Shapiro. (1996). Word meanings can be accessed but not reported during the attentional blink. Nature, 383(6601), 616-618.

Lukács, G., Weiss, B., Dalos, VD., Kilencz, T., Tudja, S., Csifcsák, G, (2016). The first independent study on the complex trial protocol version of the P300-based concealed information test: Corroboration of previous findings and highlights on vulnerabilities. International Journal of Psychophysiology, Volume 110, 56-65.

Lykken, D. (1984). Polygraphic interrogation. Nature, 307, 681-684.

Lykken, D. (1959). The GSR in the detection of guilt. Journal of Applied Psychology, 43, 385-388.

Makeig, SD. (2004a). Electroencephalographic brain dynamics following manually responded visual targets. PLoS Biology, 2(6), e176. doi:10.1371/journal.pbio.0020176.

Makeig, S., Debener, S., Onton, J., & Delorme, A. (2004b). Mining event-related brain dynamics. Trends in cognitive sciences, 8(5), 204-210.

Malmivuo, J., Suihko, V., & Eskola, H. (1997). Sensitivity distributions of EEG and MEG measurements. IEEE Transactions on Biomedical Engineering. DOI: 10.1109/10.554766.

Mangun, GR., & Hillyard, SA. (1995). Mechanisms and models of selective attention. In M. D. Rugg & M. G. H. Coles (Eds.), Electrophysiology of mind: Event-related brain potentials and cognition (pp. 40–85). New York: Oxford University Press.

Markowitsch, H. J., & Staniloiu, S. (2012). Amnesic disorders. Lancet 380, 1229–1240. doi:10.1016/S0140-6736(11)61304-4.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. Nature. 264 (5588): 746–8. doi:10.1038/264746a0. PMID 1012311.

Meijer, E. S. (2009). The Contribution of Mere Recognition to the P300 Effect in a Concealed Information Test. Appl Psychophysiol Biofeedback (2009) 34: 221.

Meixner, J., Haynes, A., Winograd, M., Brown, J., & Rosenfeld, J. (2009). Assigned versus random, countermeasure-like responses in the p300 based complex trial protocol for detection of deception: task demand effects. Sep;34(3):209-20.

Melloni, L., Van-Leeuwen, S., Alink, A., & Mueller, N. (2012). Interaction between Bottom-up Saliency and Top-down Control: How Saliency Maps Are Created in the Human Brain. Cerebral cortex (New York, N.Y. : 1991). 22. 10.1093/cercor/bhr384.

Merriam, E., & Colby, C. (2005). Active vision in parietal and extrastriate cortex. Neuroscientist. 11:484–493.

Minear, M., & Park, DC. (2004). A lifespan database of adult facial stimuli. Behavior research methods, instruments, & computers : a journal of the Psychonomic Society, Inc, 36: 630-3.

Morriss-Kay, G. (2010). The evolution of human artistic creativity. Department of Physiology, Anatomy and Genetics, Oxford, UK. J Anat. 2010 Feb;216(2):158-76.

Naatanen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. Psychophysiology, 24, 375–425.

Nasr, S., & Tootell, R. B. (2012). Role of fusiform and anterior temporal cortical areas in facial recognition. Neuroimage 63, 1743–1753. doi:10.1016/j.neuroimage.2012.08.031.

Niedermeyer, E., & Lopes da Silva, F. (1993). Electroencephalography: Basic principles, clinical applications and related fields, 3rd edition, Lippincott, Williams & Wilkins, Philadelphia. ISBN: 0781751268 .

Nørretranders, T. (1998). The User Illusion: Cutting Consciousness Down to Size. Viking, New York. ISBN-10: 0140230122.

Olivi, E. (2011). Coupling of numerical methods for the forward problem in Magneto- and Electro-Encephalography. Page 11; HAL Id: tel-00838707.

Olson, I., & Marshuetz, C. (2005). Facial Attractiveness Is Appraised in a Glance. Emotion 5, 498–502. PMID: 16366753 DOI: 10.1037/1528-3542.5.4.498.

Osterhout, L., & Holcomb, PJ. (1992). Event related brain potentials elicited by syntactic anomaly. Journal of Memory and Language. 1992;31:785–786.

O'Toole, A. (2005). Psychological and neural perspectives on human face recognition: in Handbook of Face Recognition, eds S. Z. Li, and A. K. Jain (New York: Springer Science + Business Media), 349–369. ISBN: 978-0-387-40595-7.

Palermo, R., & Rhodes, G. (2002). Are you always on my mind? A review of how face perception and attention interact. Neuropsychologia, 45(1), 75e92.

Patton, L. (2018). Hermann von Helmholtz: Philosophy of Mind in the Nineteenth Century. The Stanford Encyclopedia of Philosophy. 9781138243965.

Pearce, J. (2016). Greek medicine: a new look, Brain, Volume 139, Issue 8. Pages 2322–2325.

Peterson, NN., Schroeder, CE., & Arezzo, JC. (1995). Neural generators of early cortical somatosensory evoked potentials in the awake monkey. Electroencephalogrphy and Clinical Neurophysiology. 1995;96:248–260.

Pierce, L., Scott, L., Boddington, S., Droucker, D., Curran, T., & Tanaka, J. (2011). The N250 Brain Potential to Personally Familiar and Newly Learned Faces and Objects. Front Hum Neurosci. 2011; 5: 111.

Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: An integrative review. Biological Psychology 41:103–146.

Potter, M. (1975). Meaning in visual search. Science. *187(4180):965-6.* DOI: 10.1126/science.1145183.

Pouget, A., Beck, JM., Ma, WJ., & Latham, PE. (2013). Probabilistic brains: knowns and unknowns. Nature neuroscience, 16(9), 1170–1178. doi:10.1038/nn.3495.

Quiroga, R., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. Nature. 435(7045):1102-7.

Radlo, S., Janelle, C., Barba, D., & Frehlich, S. (2001). Perceptual decision making for baseball pitch recognition: using P300 latency and amplitude to index attentional processing. Res Q Exerc Sport. 72(1):22-31.

Raymond, JE., Shapiro, KL., & Arnell, KM. (1992). Temporary Suppression of Visual Processing in an Rsvp Task - an Attentional Blink. Journal of Experimental Psychology-Human Perception and Performance, 18, 849-860.

Rohrbaugh, J., Donchin, E., & Eriksen, C. (1974). Decision making and the P300 component of the cortical evoked response. Perception & Psychophysics 15:368–374.

Rosenfeld, IP. (2008). The Complex Trial Protocol (CTP): A new, countermeasure-resistant, accurate, P300-based method for detection of concealed information. Psychophysiology, 45(6), 906-919. doi:10.1111/j.1469-8986.2008.00708.x.

Rosenfeld, J. P., Angell, A., Johnson, M., & Qian, J. (1991). An ERP-based, control-question lie detector analog: Algorithms for discriminating effects within individuals' average waveforms. Psychophysiology, 38, 319–335.

Rosenfeld, P., Soskins, M., Bosch, G., & Ryan, A. (2004). Simple, effective countermeasures to P300-based tests of detection of concealed information. Psychophysiology. 41(2):205-19.

Rosipal, TT. (2003). Kernel PLS-SVC for linear and nonlinear classification. In Proc. 20th Int. Conf. Machine Learning (ICML-2003), Washington, D.C., day Month 2003, Place of Publication: Publisher, pp.640-7.

Rossion B. (2000). The N170 occipito-temporal component is delayed and enhanced t oinverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. Neuroreport2000;11:69–74.

Saxe, L. (1991). Lying: Thoughts of an applied social psychologist. American Psychologist, 46, 409-415.

Saxe, L., & Ben-Shakhar, G. (1999). Admissibility of polygraph tests: The application of scientific standards post-Daubert. Psychology, Public Policy and the Law, 5(1): 203-23.

Schweinberger S. (2004). N250r: a face-selective brain response to stimulus repetitions. DOI: 10.1097/01.wnr.0000131675.00319.42.

Schweinberger, SR., Pickering, EC., Jentzsh, I., Burton, AM., & Kaufmann, JM. (2002). Event-related brain potential evidence for a response of inferior. Cogn Brain Res; 14: 398 409.

Shelvin, H. (2017). Conceptual short-term memory: a missing part of the mind? Journal of Consciousness Studies 24(7-8):163-88.

Shepherd, G. (1988). Neurobiology (2nd Edition). Oxford University Press, New York. ISBN: 9780195088434.

Silva, A. M. (2016). Always on My Mind? Recognition of Attractive Faces May Not Depend on Attention. Frontiers in psychology, 7, 53. doi:10.3389/fpsyg.2016.00053.

Skirry, J. (2014). René Descartes: The Mind-Body Distinction, The Internet Encyclopedia of Philosophy (IEP) (ISSN 2161-0002).

Spencer, & Eriksen. (1969). Rate of information processing in visual perception: some results and methodological considerations. Journal of Experimental Psychology, 79(2p2), 1.

Stout, D. T. (2008). Neural correlates of Early Stone Age toolmaking: technology, language and cognition in human evolution. Philos Trans R Soc Lond B Biol Sci. Jun 12, 2008; 363(1499): 1939–1949.

Sugiura, M., Mano, Y., Sasaki, A., & Sadato, N. (2011). Beyond the memory mechanism: person-selective and nonselective processes in recognition of personally familiar faces. J. Cogn. Neurosci. 23, 699–715. doi:10.1162/jocn.2010.21469.

Sutton, S., Braren, M., Zubin, J., & John, E. (1965). Evoked-potential correlates of stimulus uncertainty. Science, 150(3700), 1187-1188.

Synnott, J., Dietzel, D., & Ioannou, M. (2015). A review of the polygraph: history, methodology and current status, Crime Psychology Review, 1:1, 59-83.

Tanaka, J., & Simonyi, D. (2016). The “parts and wholes” of face recognition: a review of the literature. Q J Exp Psychol (Hove). 69(10): 1876–1889.

Tanner, D., Morgan-Shoer, K., & Luck, SJ. (2015). How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. Psychophysiology. 52. 997-1009. 10.1111/psyp.12437.

Taylor, J., Shehzad, S., & McCarthy, G. (2016). Electrophysiological correlates of face-evoked person knowledge. Biol Psychol. 118: 136–146.

Tononi, G. (2008). Consciousness as Integrated Information: a Provisional Manifesto. Marine Biological Laboratory. vol. 215 no. 3 216-242.

Touryan, J. (2011). Real-Time Measurement of Face Recognition in Rapid Serial Visual Presentation. Front Psychol. 2011; 2: 42.

Trenner, M. S., Jentzsch, I., & Sommer, W. (2004). Face repetition effects in direct and indirect tasks: An event-related brain potentials study. Brain research. Cognitive brain research. 21. 388-400.

Trovillo, P. (1939). A History of Lie Detection, Journal of Criminal Law: Criminology and Police Science, 29, 848-881.

Van der Stelt, O., Van der Molen, M., Boudewijn Gunning, W., & Kok, A. (2001). Neuroelectrical signs of selective attention to color in boys with attention-deficit hyperactivity disorder. Cognitive Brain Research, 12, 245–264.

Van Vugt, M., Sederberg, P., & Kahana, M. (2007). Comparison of spectral analysis methods for characterizing brain oscillations. Journal of Neuroscience Methods 162 49–63.

Verleger, R. (1988). Event-related potentials and cognition: A critique of the context updating hypothesis and an alternative interpretation of P3. Behavioral and Brain Sciences, 11(03), 343-356.

Vogel, E. & Luck, S. (2000). he visual N1 component as an index of a discrimination process. Psychophysiology. 2000 Mar; 37(2):190-203.

Vogel, E., & Machizawa, M. (2004). Neural activity predicts individual differences in visual working memory capacity. Nature; 428(6984):748-51.

Walter, W. C. (1964). Contingent Negative Variation: An electric sign of sensorimoto association and expectancy in the Human brain. Nature 203():380-4.

Wang, Y., & Jung, T. (2010). Visual stimulus design for high-rate SSVEP BCI. Electronics Letters, 46, 1057-U27.

Werner, N.-S., Kühnel, S., & Markowitsch, H. (2013). The Neuroscience of Face Processing and Identification in Eyewitnesses and Offenders ; DOI=10.3389/fnbeh.2013.00189.

Willis, J. & Todorov, A. (2006). First Impressions: Making Up Your Mind After a 100-Ms Exposure to a Face. Psychological Science 17, 592–598.

Wyble, B., Bowman, H., & Nieuwenstein, M. (2009). The attentional blink provides episodic distinctiveness: sparing at a cost. J Exp Psychol Hum Percept Perform. 35(3):787-807. doi: 10.1037/a0013902.

Yeung, N., Bogacz, R., Holroyd, CB., & Cohen, JD. (2004). Detection of synchronized oscillations in the electroencephalogram: an evaluation of methods. 2004.Psychophysiology, 41(6):822-32.

Zimmermann, F., & Eimer, M. (2013). Face learning & emergence of view-independent face recognition: an event-related brain potential study. Neuropsychologia. 51(7):1320-9.

Zimmermann, F., & Eimer, M. (2014). The activation of visual memory for facial identity is task-dependent: Evidence from human electrophysiology. ResearchGate, DOI: 10.1016/j.cortex.2014.02.008.

Zumsteg, D., & Wieser, H. G. (2005). Presurgical evaluation: current role of invasive EEG. Epilepsia, 41(s3), S55-S60.