COMENIUS UNIVERSITY IN BRATISLAVA

FACULTY OF MATHEMATICS, PHYSICS, AND INFORMATICS

PRACTICAL APPROACHES TO ERP ANALYSIS IN THE CONTEXT OF VISUAL SHORT-TERM MEMORY

Master's thesis

2016

BC. MILAN MITKA

COMENIUS UNIVERSITY IN BRATISLAVA

FACULTY OF MATHEMATICS, PHYSICS, AND INFORMATICS

PRACTICAL APPROACHES TO ERP ANALYSIS IN THE CONTEXT OF VISUAL SHORT-TERM MEMORY

Master's thesis

Course:	Cognitive Science (master's degree, full-time)
Study programme:	2503 Cognitive Science
Department:	FMFI.KAI – Department of Applied Informatics
Supervisor:	Igor Riečanský, MUDr., PhD.

BRATISLAVA 2016 BC. MILAN MITKA

UNIVERZITA KOMENSKÉHO V BRATISLAVE

FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

PRAKTICKÉ PRÍSTUPY K ANALÝZE ERP V KONTEXTE ZRAKOVEJ KRÁTKODOBEJ PAMÄTI

Diplomová práca

Študijný program:	kognitívna veda
Študijný odbor:	2503 kognitívna veda
Katedra:	FMFI.KAI – Katedra aplikovanej informatiky
Vedúci práce:	MUDr. Igor Riečanský, PhD.

BRATISLAVA 2016 BC. MILAN MITKA

Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname:	Bc. Milan Mitka
Study programme:	Cognitive Science (Single degree study, master II. deg., full time form)
Field of Study:	Cognitive Science
Type of Thesis:	Diploma Thesis
Language of Thesis:	English
Secondary language:	Slovak

Title: Practical approaches to ERP analysis in the context of visual short-term memory

Aim:
1. Get familiar with the concept of visual short-term memory (VSTM) and the underlying cognitive processes, namely information encoding, active maintanance, inhibition of interfering information and retrieval, and investigate their contribution to individual differences in VSTM capacity.
2. Focus on the use of electroencephalography and related methods to study the underlying neural processes.
3. Provide a description of the available methodological approaches to analysis of event-related potentials (ERP).

Literature:
Riečanský I., Tomova L., Katina S., Bauer H., Fischmeister F. & Lamm C. (2013). Visual image retention does not contribute to modulation of event-related potentials by mental rotation. Brain and Cognition, 83(2), 163-170.
Drew T., McCollough A. & Vogel E. (2006). Event-Related Potential Measures of Visual Working Memory. Clinical EEG and Neurosci., 37(4), 286-291.

Annotation: The exploration of brain processes of VSTM using ERP can provide new insights into the interindividual differences in VSTM and identify contribution of different VSTM component processes to overall storage accuracy. The summary of analytical approaches to ERP will provide a useful guidance for further studies linking cognitive and neuronal processes.

Supervisor:	MUDr. Igor Riečanský, PhD.
Department:	FMFI.KAI - Department of Applied Informatics
Head of department:	prof. Ing. Igor Farkaš, Dr.
Assigned:	02.11.2014
Approved:	02.11.2014

prof. Ing. Igor Farkaš, Dr.
Guarantor of Study Programme

Student

Supervisor

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Milan Mitka
Študijný program: kognitívna veda (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: kognitívna veda
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: anglický
Sekundárny jazyk: slovenský

Názov: Practical approaches to ERP analysis in the context of visual short-term memory
Praktické prístupy k analýze ERP v kontexte zrakovej krátkodobej pamäte

Cieľ: 1. Oboznámte sa s konceptom zrakovej krátkodobej pamäte (VSTM) a zúčastnenými kognitívnymi procesmi, zahŕňajúcimi kódovanie informácií, aktívne držanie, inhibíciu interferujúcich informácií a vybavovanie, a preskúmajte ich vplyv na individuálne rozdiely v kapacite VSTM.
2. Zamerajte sa na využitie elektroencefalografie a príbuzných metód na štúdium neurálnych procesov tvoriacich ich podklad.
3. Poskytnite popis dostupných metodických prístupov k analýze na udalosť viazaných potenciálov (ERP).

Literatúra: Riečanský I., Tomova L., Katina S., Bauer H., Fischmeister F. & Lamm C. (2013). Visual image retention does not contribute to modulation of event-related potentials by mental rotation. Brain and Cognition, 83(2), 163-170.
Drew T., McCollough A. & Vogel E. (2006). Event-Related Potential Measures of Visual Working Memory. Clinical EEG and Neurosci., 37(4), 286-291.

Anotácia: Skúmanie mozgových procesov VSTM s použitím ERP môže priniesť nové poznatky objasňujúce interindividuálne rozdiely vo VSTM a identifikovať mieru účasti rôznych čiastkových procesov VSTM na celkovej presnosti uskladnenia informácií. Zhrnutie prístupov k analýze ERP poskytne užitočný návod pre budúce štúdie spájajúce kognitívne a neurálne procesy.

Vedúci: MUDr. Igor Riečanský, PhD.
Katedra: FMFI.KAI - Katedra aplikovanej informatiky
Vedúci katedry: prof. Ing. Igor Farkaš, Dr.

Dátum zadania: 02.11.2014

Dátum schválenia: 02.11.2014

prof. Ing. Igor Farkaš, Dr.
garant študijného programu

.....
študent

.....
vedúci práce

ACKNOWLEDGEMENTS

I would like to express my utmost gratitude to Dr. Igor Riečanský for his professional guidance, overall kindness, and exceptional patience, as well as for the data that were used in the practical part of this work.

I also thank my mum for her general support and the developers of the free software that I used.

COLOPHON

This document was created with the use of LaTeX2e and BibLaTeX, using document class *memoir*, *fontspec* package, and typeset by LuaLaTeX.
The text is set in Minion Pro, Myriad Pro, and `Fira Code`.

ABSTRACT

This thesis provides an overview of practical approaches to the analysis of even-related potentials (ERP) and demonstrate the use of such methods in the context of visual short-term memory (VSTM).

Inter-individual differences in humans performing short-term memory tasks can be attributed to different factors including different stages of the memory process, namely encoding, maintenance, resistance to interference, and recall.

We analyse which of these stages is the strongest predictor of performance measured by response accuracy using data from a short-term memory (STM) task in which 28 subjects shortly observed a visual target stimulus which was then covered with a mask for 3 seconds and had to be retained in STM. They were then shown a probe stimulus and had to judge if it was rotated clockwise or counter-clockwise with respect to the target, and indicate this by a key press.

First, we introduce the theory in a methodical fashion, starting with a global overview of memory (history, models, experiments), later focusing on the specific type of memory that we set out to explore, followed by a general description of electroencephalography (EEG) and ERPs, including recent findings and questions that are still unresolved, such as the origin of ERPs. At the end of chapter one, we cover time-domain analysis and statistical methods developed to make reliable inferences based on electrophysiological data, which contain many observations.

Second, we present the methodology and results of our exploratory analysis. We begin with a brief description of data collection and pre-processing and then explain the sensor-level analysis from start to finish.

The main finding is a consistent area of strong correlation ($r > 0.5$, $p < 0.01$) between the response accuracy and the amplitude of the respective ERPs in the time window between 376 and 552 ms after the onset of the target stimulus, i.e. during the late encoding stage and extending shortly into the period of maintenance, over the medial part of the frontal lobe.

Finally, we attempt to perform a source-level analysis in order to obtain an approximation of the neural sources of signals collected outside of the head. The limitations of our procedure are discussed at length in the final chapter.

Keywords: *visual short-term memory (VSTM), event-related potentials (ERP), EEG signal processing, human electrophysiology*

ABSTRAKT

Táto práca poskytuje prehľad praktických prístupov k analýze na udalosť viazaných potenciálov (ERP) a demonštruje využitie týchto metód v kontexte zrakovej krátkodobej pamäti.

Interindividuálne rozdiely v ľudskom výkone počas vykonávania úloh zameraných na krátkodobú pamäť môže byť pripísané rôznym faktorom vrátane rôznych štádií pamäťového procesu (kódovanie, držanie, odolnosť voči interferencii, vybavovanie).

V rámci práce analyzujeme, ktoré štádium je najsilnejším prediktorom výkonu vo forme presnosti odpovede s využitím dát z úlohy zameranej na krátkodobú pamäť, ktorej bol 28 subjektom zbežne prezentovaný a následne na tri sekundy maskou prekrytý zrakový podnet. Ich úlohou bolo držať ho v krátkodobej pamäti a po prezentácii ďalšieho podnetu stlačením tlačidla posúdiť, či bol tento voči pôvodnému otočený v smere alebo proti smeru hodinových ručičiek.

Ako prvé uvádzame logicky zoradenú teóriu od globálneho prehľadu pamäti (história, modely, experimenty), neskôr za zameriavame na špecifický typ pamäti, ktorý sme sa rozhodli skúmať, nasledovaný všeobecným popisom elektroencefalografie (EEG) a ERP. Na konci prvej kapitoly sa venujeme analýze v časovej doméne a štatistickým metódam určeným na vytváranie spoľahlivých úsudkov založených na elektrofyziologických dátach, ktoré obsahujú mnoho pozorovaní.

Ďalej prezentujeme metodiku a výsledky našej exploračnej analýzy. Najskôr uvádzame základný popis metód zberu dát a predspracovania, potom popisujeme kompletnú analýzu na úrovni senzorov.

Hlavným zistením je konzistentná oblasť silnej korelácie ($r > 0.5$, $p < 0.01$) medzi presnosťou odpovede a amplitúdou zodpovedajúcich ERP v časovom okne medzi 376 a 552 ms po prezentácii prvého podnetu, teda v neskorom štádiu kódovania a mierne zasahujúca do intervalu držania, nad stredovou časťou čelového laloka.

Nakoniec sa pokúšame o vykonanie analýzy na úrovni zdrojov za účelom získania aproximácie neurálnych zdrojov signálov zaznamenaných na povrchu hlavy. Limitácie nášho postupu sú popísané v poslednej kapitole.

Kľúčové slová: *zraková krátkodobá pamäť (VSTM), na udalosť viazané potenciály (ERP), spracovanie signálu EEG, humánna elektrofyziológia*

Contents

1 Introduction and theory **1**

1.1 Memory 1

1.1.1 History 1

1.1.2 Short-term memory 3

1.1.3 Visual short-term memory 7

1.2 Electroencephalography 11

1.2.1 Event-related potentials 13

1.3 Time-domain analysis 16

1.3.1 Overview 16

1.3.2 Statistics 18

1.3.3 Non-parametric permutation testing 21

2 Methodology and results **23**

2.1 Data collection and pre-processing 23

2.2 Sensor-level analysis 25

2.2.1 Tool overview 25

2.2.2 Importing the data 26

2.2.3 Time-locked averaging 28

2.2.4 Plotting at the sensor level 30

2.2.5 Statistical design 32

2.2.6 Correcting for multiple comparisons 37

2.3 Source-level analysis 41

2.3.1 Head model 43

2.3.2 Forward solution 45

2.3.3 Inverse solution 45

2.3.4 Statistical inference 47

3 Discussion and conclusion **49**

Bibliography **52**

CHAPTER 1

Introduction and theory

1.1 MEMORY

Memory is the process of encoding, storing, and retrieving information indispensable to human existence. We need it to perform a virtually endless array of functions ranging from the most basic ones like motor control or recognising simple shapes or syllables, to complex cognitive processing such as learning new skills. Regardless of the substrate, this process consists of three distinct stages, namely encoding (the reception of information and their transformation to a form that can be stored and used for further processing), storage (the protection and maintenance of encoded information for the required amount of time), and retrieval, i.e. the process of accessing the storage space and accessing the required information on cue for additional processing.

1.1.1 HISTORY

To give a brief overview of the historical context, Baddeley, Eysenck, and Anderson (2009) provide a comprehensive account of the main milestones: Researchers have been trying to describe this process, differentiate between various apparent sub-types of memory, and devise a comprehensive framework for its empirical study since the 1930s, based mainly on the

work of Hull (1943; as cited in Baddeley et al., 2009) who used rigorous and explicit modelling to construct a general theory of learning, and Tolman (1948; as cited in Baddeley et al., 2009) who thought of learning as forming cognitive representations of the outside world through active exploration. In the 1960s, Herman Ebbinghaus was the first to demonstrate the use of experimental methods to study memory in humans and thereby shifted the interest from learning in itself to how new information interacted with what was already known.

At roughly the same time, another approach to memory developed in Britain based on Bartlett's (1932; as cited in Baddeley et al., 2009) book *Remembering*. He emphasised the role of meaning and explained errors in the memory process in terms of individuals' cultural assumptions about the world, proposing that these depend on internal representations which he termed *schemas*, however without an empirical way of studying them. A possible answer to the problem came with the development of computers and the idea of representing theories as models and using computers for their development (Craik, 1943; as cited in Baddeley et al., 2009). This started a new approach to psychology based on the computer metaphor and incited young scientists to apply these findings to practical problems originating in the wartime period, concluding with Neisser's (1967; as cited in Baddeley et al., 2009) book and formal birth of cognitive psychology. Human memory was henceforth regarded with the digital computer analogy in mind, comprising one or more storage systems with the three basic functions (encoding, storage, retrieval) that can interact and directly influence the overall quality of remembering information.

Baddeley et al. (2009) continue with an account of how the cognitive approach to psychology initiated the shift from the assumption of a single memory system of stimulus-response associations towards multiple systems, each adapted to perform a different function. The formation of memories was organised as a pipeline beginning with a stream of data from environment that is processed by a series of sensory memory systems that provide an interface between perception and memory, then temporarily stored in short-term memory (STM), and finally passed on to long-term memory (LTM). A version of this model developed by Atkin-

son and Shiffrin (1968; as cited in Baddeley et al., 2009) was particularly influential and its amended version is still being taught today.

According to some authors, memory should be regarded as a process rather than a static storage framework because of similarities between different memory tasks that may imply common processes and hence a single memory system (Nairne, 2002; Neath & Surprenant, 2003; as cited in Baddeley et al., 2009). Baddeley et al. (2009) respond by suggesting that similarly to an analysis of the brain that requires both static anatomical and dynamic physiological knowledge, we need to regard memory in terms of structures such as stores and the processes that operate on them, and that the common features should not encourage us to ignore the differences. We assume separate systems for sensory, short-term, and long-term content, but there is a lot of evidence suggesting that the flow of information is bidirectional (the contents of our long-term memory can also affect the sensory input, which is known as top-down processing) and more complicated than it was assumed originally.

1.1.2 SHORT-TERM MEMORY

We continue with a more elaborate account of short-term memory which will—as the title of this thesis implies—be divided into subsystems further in this chapter. For now, we turn back to Baddeley et al. (2009) and their excellent overview of this topic, starting with the difference between STM and working memory as these terms tend to be confused and are often used interchangeably. The former is used to refer to *"performance on a particular type of task, one involving the simple retention of small amounts of information, tested either immediately or after a short delay"* and the memory system or systems responsible for STM form part of the working memory system, which is the term used for *"a system that not only temporarily stores information but also manipulates it so as to allow people to perform such complex activities as reasoning, learning, and comprehension"* (Baddeley et al., 2009, p. 19). To put it briefly, short-term memory is the theory-free capacity to store information over a brief interval while working memory also entails the capacity to manipulate

information within the famous model developed and later extended by Baddeley and Hitch (1974). It is assumed that working memory provides a temporary workspace which is necessary to perform complex cognitive tasks, described in a number of theoretical approaches based on studies of attention (Cowan, 2001; as cited in Baddeley et al., 2009), individual differences in performance on complex tasks (Miyake et al., 2000; Engle & Kane, 2004; as cited in Baddeley et al., 2009), or neurophysiological considerations (Goldman-Rakic, 1996; as cited in Baddeley et al., 2009). For the purposes of this work, we only focus on short-term memory itself.

The first empirical assessment of STM appears to have been done by John Jacobs (1887; as cited in Baddeley et al., 2009) who devised a simple test in which the participant heard a sequence of numbers and repeated them back. The longest successfully reproduced sequence was termed the digit span and later included in tests of intellect although this basic version (unlike the working memory span, which can predict a range of cognitive skills) does not correlate highly with general intelligence. As such, the digit span is a measure that reflects STM (Baddeley et al., 2009). For most people, digit span is limited to six or seven digits but the dispersion is rather high, ranging from about four to ten or even more. Since it requires the person to remember what the items are as well as the order in which they were presented, one is expected to perform better on tasks in which the items are already known, such as familiar digits. Unfamiliar digits or words need to be remembered in addition to their order and as such are considerably harder to recall. There is also the question of how order is remembered. According to (Baddeley et al., 2009), one might think that each item is linked to the next, thereby creating a chain of items, however experimental evidence suggests that this is not the case since, although there is a decline in performance following a mistake, it does not collapse as chaining would predict. Let us now move from sequences of numbers to letters. We get much better performance with strings of letters that can be divided into pronounceable word-like subgroups, also called chunks, than random strings that do not follow known grammatical rules. Therefore, it was suggested that the capacity of STM is limited by the number of chunks rather than items that need to be recalled, for example syllables

instead of letters (Miller, 1956; as cited in Baddeley et al., 2009). This effect can also be induced by rhythm (e.g. making pauses after groups of numbers) whereby patterns consistent with long-term memory habits result in the same effect which seems to work best for groups of three items (Wickelgren, 1964; Ryan, 1969; as cited in Baddeley et al., 2009).

Another important observation was made by Conrad (1964, as cited in Baddeley et al., 2009) when he noticed that errors in visually presented consonants were more likely to be similar in sound than shape of the letter. This led to the discovery that memory for strings of consonants similar in sound is substantially worse than for sequences comprising of sounds that are unique. The interpretation for this effect was that the respective STM store depends on a rapidly fading acoustic code which had to operate with fewer distinguishing features (Conrad & Hull, 1964; as cited in Baddeley et al., 2009).

Peterson and Peterson (1959; as cited in Baddeley et al., 2009) developed a technique to investigate the effect of distraction on short-term retention in which participants were given a consonant triplet and asked to count down backwards in threes from a given number, after which they were asked to recall the triplet with various delay. They found that the percentage of correctly recalled triplets declined with time which was later confirmed by Murdock (1960; as cited in Baddeley et al., 2009) who found an equivalent effect for triplets of words. One possible explanation was that the numbers were interfering with the letters stored in memory, however McGeogh & MacDonald (1931; in Baddeley et al., 2009) found that *"such interference depends on the similarity between the remembered and the interfering material, and that for [long-term memory] at least, numbers do not interfere with letters*", which is why the Petersons *"suggested that their results reflected the rapid fading of a short-term memory trace"* (Baddeley et al., 2009, p. 23), a conclusion consistent with that of Brown (1958; as cited in Baddeley et al., 2009) but contrary to the commonly held view of memory as a unitary system in which forgetting was the result of interference.

This was later challenged by Keppel and Underwood (1962; as cited in Baddeley et al., 2009) who noticed that rapid forgetting only happened

after the first few trials of the experiment and that the first triplet showed little to no forgetting, suggesting that the interference came from earlier triplets rather than the numbers. The experiment was therefore adjusted to use triplets of words in which each item belonged to the same semantic category (e.g. birds or colours), which was changed after five trials. The performance on this task showed periodical behaviour, i.e. a steady decline over the first five trials, recovery with the introduction of another category, and another steady decline until the category was changed again (Loess, 1968; as cited in Baddeley et al., 2009).

Another experimental paradigm that became popular and theoretically controversial during 1960s was free recall, a task in which the participants are given a list of items to remember and subsequently recall in any order they wish. Postman and Phillips (1965; as cited in Baddeley et al., 2009) conducted an experiment in which 10, 20, or 30 words were presented to be recalled immediately or after a filled delay of 15 seconds, concluding that: *"(1) the likelihood of recalling an individual item is less for longer lists, although the total number of items recalled is likely to increase; (2) all lists showed a tendency for the first few items to be somewhat better recalled, the so-called primacy effect; (3) regardless of list length, if recall is immediate then the last few items are very well recalled, the recency effect; and (4) this effect is eliminated by a brief delay filled by some activity such as counting"* (Postman & Phillips, 1965; as cited in Baddeley et al., 2009, p. 24). According to the same authors, the primacy effect likely depends mainly on long-term memory at least partly due to the tendency to rehearse the first few items at the beginning of the task and throughout it (Tan & Ward, 2000; as cited in Baddeley et al., 2009). The variables that are known to influence overall performance at the beginning and throughout the middle of the task include presentation rate (slower is better), word frequency (familiar is easier), imageability of the words (visualisable is better), age of the participant (young adults are better than children or the elderly), or physiological state (drugs impair performance), but they have no comparable impact on the recency effect itself, and while Postman and Phillips (1965; as cited in Baddeley et al., 2009) offered an interpretation based on the interference theory, it failed to account for the

recency. Glanzer (1972; as cited in Baddeley et al., 2009) offered a popular interpretation that this effect simply reflected a temporary short-term store with different characteristics than the long-term store which was responsible for performance on earlier items, however this was later challenged by demonstrating that such effects can occur under conditions in which no short-term trace should be present (e.g. Bjork and Whitten, 1974; as cited in Baddeley et al., 2009). Similar recency effect have also been shown over much longer intervals, for example when rugby players were asked to recall which teams they had played that season (Baddeley & Hitch, 1977; as cited in Baddeley et al., 2009), or remembering parking locations (Pinto & Baddeley, 1991; as cited in Baddeley et al., 2009). This suggests that recency is not limited to a single type of memory system but reflects a retrieval strategy exploiting the fact that most recent events are more available for recollection, therefore the most plausible interpretation appears to be in terms of retrieval and with increasing delay, the discrimination becomes more difficult (Baddeley & Hitch, 1977; Brown, Neath, & Chater, 2007; as cited in Baddeley et al., 2009).

1.1.3 VISUAL SHORT-TERM MEMORY

Let us now move on to the specific type of memory system that we chose to explore in our work. Visual short-term memory (VSTM) is the capacity for holding a small amount of visual information in a readily available state for a short period of time, generally in the order of seconds, and limited to about two to five objects at any given moment. This ability has been shown to vary substantially across individuals and should be distinguished from working memory, a theory-specific system that not only temporarily stores but also manipulates information (including different modalities) to allow the performance of complex tasks such as learning. In addition to this distinction, it is generally regarded as the *"what?"* subcomponent of visuo-spatial STM which stores information related to the nature of objects (colour, shape) as opposed to the spatial *"where?"* subcomponent which stores their position in space (Mammarella, Pazzaglia, & Cornoldi, 2008; Klauer & Zhao, 2004).

Posner and Konick (1966; as cited in Baddeley et al., 2009) devised an experiment to investigate the attributes of the spatial *"where?"* memory in which they asked the participants to remember where along a line a stimulus had occurred and recall the position after a filled an unfilled delay. They found that retention was good after an unfilled delay but when the participants had to process digits, the performance decreased with the complexity of the task. Since the digit processing was not visual or spatial in nature, they concluded that it interfered with capacity to rehearse or retain the initial stimulus, which was later confirmed by Dale (1973; as cited in Baddeley et al., 2009) using a similar approach. Irwin and Andrews (1996; as cited in Baddeley et al., 2009) focused on the object *"what?"* memory, asking their participants to remember an array of differently coloured letters and then recall the letter, colour, or both, at a given location indicated by an asterisk. The performance was practically equal regardless of which feature or their combination had to be recalled and highly accurate up to four items. However, the use of letters raised questions about the possible involvement of verbal processing, which is why Vogel, Woodman, and Luck (2001; as cited in Baddeley et al., 2009) conducted a series of studies in which the stimuli were changed to bars of varying width, orientation, and other features that rendered verbal coding in the brief presentation interval useless, even if articulatory suppression was used to actively prevent it. They confirmed the same limit of four objects and found that the number of features had little effect on performance and even very complicated objects were encoded effectively (see also Awh, Barton, & Vogel, 2007). It was concluded that *"the system was limited on the basis of the number of objects but that objects could vary in complexity without affecting performance"* (Vogel et al., 2001; as cited in Baddeley et al., 2009, p. 34; see also Luria, Sessa, Gotler, Jolicoeur, & Dell'Acqua, 2010). Woodman and Luck (2004; as cited in Baddeley et al., 2009) combined the memory task with visual search, requiring the participants to scan an array for visually specified targets in the interval between presentation and test of sets of coloured shapes, and found that memory for spatial locations was impaired but object memory was not. Similarly, Oh and Kim (2004; as cited in Baddeley et al., 2009) found that

the need to remember a set of objects had no impact on the participants' visual search rate whereas a spatial retention task slowed them down. The summary of these findings is that people can remember up to four objects virtually regardless of the number of features they are composed of, and the information remain stored without declining over a period of a few seconds and regardless of interpolated activity, while memory for spatial location is more prone to damage by other spatial processing (Baddeley et al., 2009).

While the two memory systems work together, the distinction between spatial and object memory can be emphasised with the use of special types of tasks designed to isolate one of them: *"A classic spatial task is the block tapping test in which the participant is faced with an array of nine blocks […] the experimenter taps a number of blocks in sequence and the participant attempts to imitate this, with the length of sequence increasing until performance breaks down […] this is known as Corsi span, after the Canadian neuropsychologist who invented it, and is typically around five blocks, usually about two items below digit span"* (Baddeley et al., 2009, p. 34). On the other hand, *"visual span can be measured using a series of matrix patterns in which half the cells are filled and half left blank […] the participant is shown a pattern and asked to reproduce it by marking the filled cells in an empty matrix; testing starts with a simple 2 × 2 pattern, and the number of cells in the matrix is gradually increased to a point at which performance breaks down, usually around the point at which the matrix reaches around 16 cells"* (Baddeley et al., 2009, 34–35). This distinction is supported by studies that use a potentially interfering activity between presentation and test, which results in deterioration of normal performance (Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999; as cited in Baddeley et al., 2009).

Neurophysiological studies of visual STM investigating the sources of inter-individual differences and the associated predictors revealed a range of new information (Drew, McCollough, & Vogel, 2006). Vogel and Machizawa (2004) established an electrophysiological index of storage capacity limitations using event-related potentials by taking advantage of lateralised activity that reflects the encoding and maintenance

of items in visual memory. They used a task consisting of a memory array that contained a fixation cross in the middle and same number of coloured squares for each hemifield. The participants were cued to remember the colours of objects in either hemifield during a brief presentation period followed by a retention interval and then had to indicate whether the test array was the same or one of the objects had a different colour. By comparing the contralateral and ipsilateral event-related potentials, they found a considerable drop in contralateral electrodes 200 ms after the stimulus onset over posterior parietal and lateral occipital electrodes. In order to rule out executive processes, increased effort, or arousal, they removed non-specific bilateral activity by constructing difference waves (contralateral minus ipsilateral activity) and manipulating the number of squares per each hemifield. They found that the amplitude was highly sensitive to the number of objects and considerably lower when the answers were incorrect, i.e. the accuracy was low. They also found no significant differences between numbers of objects exceeding the individual VSTM capacity, concluding that this *contralateral delay activity (CDA)* is indeed an index of active representations in . The more objects are maintained, the higher the CDA amplitude becomes.

Murray, Nobre, and Stokes (2011) investigated the effect of preparatory spatial attention on VSTM encoding and performance while recording EEG activity of participants performing a task which manipulated the spatial distribution of attention at encoding, the memory load, and the relative difficulty of the change discrimination at the memory probe, and found that individual differences in preparatory brain activity indicated by ERP markers of anticipatory spatial attention, namely attention directing negativity (ADAN) and late directing attention positivity (LDAP), predicted cue-related changes in recall accuracy. Their conclusions suggest that attention affects the probability that an item enters VSTM and is successfully maintained.

Vogel, Woodman, and Luck (2006) used a change-detection task to measure the time course of consolidation of objects in visual working memory, which was theorised to be very slow, however they found that it can be as fast as 50 ms per item.

Most importantly, at least for the purposes of our work since our practical demonstration of approaches to ERP analysis attempts to replicate their results using different techniques, Riečanský, Tomova, Fischmeister, Bauer, and Lamm (2011) were looking at the dynamics of human brain activity related to VSTM which relies on cognitive processes such as encoding, retention (or maintenance), resistance to interference, and recall, but the contribution of these processes to individual differences in VSTM capacity remains unknown. In order to investigate the precision of storage for a single object, they recorded ERPs in a delayed orientation discrimination task (this is described in section 2.1), computed an individual orientation discrimination threshold which determined and represented individual VSTM capacity (the lower the threshold, the higher the VSTM capacity), and used temporal-spatial PCA to identify the underlying brain sources. This method revealed networks of synchronously active brain sources related to the underlying low-level processes; one of them—active during the encoding phase and shortly after the mask was displayed (the beginning of the maintenance period)—was significantly positively associated with the orientation discrimination threshold and localised to the midcingulate cortex. They concluded that the capacity of VSTM mostly relies on the efficiency of information encoding and that the *"involvement of the midcingulate cortex suggests that encoding efficiency is reflected by the engagement of the executive attention system"* (Riečanský et al., 2011; Smith, 1999).

1.2 ELECTROENCEPHALOGRAPHY

Electroencephalography (EEG) is an electrophysiological and typically non-invasive method of recording fluctuations in electric potential generated by ionic current flow within neuronal populations, mostly consisting of cortical pyramidal neurons, which are spatially aligned and fire synchronously. It is a direct measure of brain activity with great temporal and decent spatial resolution that has been widely used in research and clinical applications since 1924 when the first human EEG recording was

obtained by Hans Berger (Niedermeyer & Silva, 2004).

The temporal resolution of EEG depends on the sampling rate during acquisition and is usually between hundreds and a few thousands samples per second. For most analyses, the appropriate temporal resolution would be between 250 and 1000 Hz. As for temporal precision, i.e. the certainty of the measurement at each time point, it depends on the subsequent analysis, being the highest for raw data and lowered by filtering which leads to temporal leakage because each time point becomes a weighted average of temporally surrounding activity in the respective band. Temporal accuracy—the relationship between the timings of the obtained signal and the biophysical events that generate it—is extremely high since the electrical signal conveys information practically instantaneously (Cohen, 2014).

The spatial resolution is determined by the number of electrodes which is in turn usually selected depending on the analyses one needs to perform. For example, brain source localisation techniques can benefit from an increased number of electrodes but the increase in accuracy is not linear. The spatial precision is relatively low but can be improved using various techniques that will be discussed in the next chapter. The spatial precision of source localisation can be high, however it is difficult to obtain due to many factors—such as anatomically precise forward models—affecting the solution. Finally, the spatial accuracy is low because activity recorded from a given electrode does not represent only the signal from neurons directly below it but practically from the whole body, though weak and distant sources are of course less prominent (Cohen, 2014).

Cohen (2014) also states that it is useful to differentiate three spatial scales: a) microscopic, which refers to areas of less than a few cubic millimetres, comprised of neural columns, neurons, synapses, and other structures similar in size, whose dynamics are most likely invisible to EEG, b) mesoscopic, i.e. patches of cortex between several cubic millimetres and centimetres, whose dynamics can be resolved with EEG, albeit with at least 64 electrodes, and c) macroscopic, meaning larger regions of cortex that span many cubic centimetres and can be measured with fewer electrodes.

1.2.1 EVENT-RELATED POTENTIALS

The potential of recording EEG was later expanded by discovering that brain activity evoked by specific stimuli can be measured using a technique called event-related potential (ERP). These are electrical potentials generated by the brain that are related to specific internal or external events obtained by averaging over many trials with respect to a given event to clean the signal of noise (Luck, 2012). This technique results in a number of characteristic waveforms associated with various aspects of cognitive processing, including those related to memory and being as specific as e.g. reflecting the individual capacity of VSTM (Vogel & Machizawa, 2004).

According to Cohen (2014), there are four main advantages of ERPs. First of all, they are simple and fast to compute and require few assumptions or parameters concerning the nature of the investigated effect, thereby making them ideal in cases where one has no preconceptions about the electrophysiological dynamics that are involved. Second, their high temporal precision and accuracy make them better for careful exploration of latencies than for example methods that require time-frequency decomposition which involves temporal smoothing. Third, it is a very well-documented method commonly used in cognitive neuroscience, thereby providing one with a good basis for the development of new hypotheses and techniques, whereas there is reportedly less published research on time-frequency electrophysiological characteristics of cognitive processes. Fourth, they allow for quick and easy quality check of single-subject data, i.e. one can look for specific spatial-temporal progression of ERP activity to confirm that the obtained data are consistent with previous findings.

Cohen (2014) also lists two main limitations. The first one is that there are many kinds of dynamics in EEG data that are not represented in ERPs (such as time-locked but not phase-locked activity) which therefore reveal little information compared to time-frequency representations, and the second that they provide limited opportunities for identifying links between results and physiological mechanisms since neurophysiological mechanisms that produce ERPs are less understood than those which produce oscillations. Examples of such neural phenomena that have been

linked to brain function and cognition are interregional synchronisation and cross-frequency coupling which are recognised as fundamental mechanisms underlying neural computation and interregional communication, whereas this is not possible with ERPs and our hypotheses can effectively not be tested.

Recent simulations suggest that ERPs may be produced by complex additive and nonlinear effects (David, Kilner, & Friston, 2006; as cited in Cohen, 2014) or rapid changes in frequencies (Burgess, 2012; as cited in Cohen, 2014), however there is not as much in vitro evidence as in the case of oscillation mechanisms. Cohen (2014) describes a few different models of how ERPs can emerge from ongoing or oscillatory activity—

- *additive,* according to which ERP reflects a signal elicited by an external stimulus or an internal event and is added to ongoing background oscillations; since the oscillations are not related to the stimulus, they are attenuated in trial averaging; the model assumes a distinction between neurophysiological events that produce oscillations and those that produce ERPs;

- *phase reset,* which proposes that they *"result at least partially from a sudden alignment of the phases of ongoing oscillations"* (Makeig et al., 2002; as cited in Cohen, 2014, p. 57), which means that *"when a stimulus appears, the ongoing oscillation at a particular frequency band is reset to a specific phase value, which may reflect a return to a specific neural network configuration"* (Cohen, 2014, p. 57), however it is unlikely to account for later "cognitive" ERP components (Fell et al., 2004; as cited in Cohen, 2014).

- *amplitude asymmetry or baseline shift,* according to which outward-going currents may be less detectable from the scalp (Mazaheri & Jensen, 2008; as cited in Cohen, 2014), which would *"produce an asymmetry in the oscillations measured by scalp EEG electrodes such that peaks and troughs are not equally distributed"* and might also *"produce a baseline shift, which would also effectively produce asymmetries between peaks and troughs of oscillations"* (Nikulin,

> Linkenkaer-Hansen, Nolte, & Curio, 2010; as cited in Cohen, 2014, p. 57); *"changes in overall power could thus produce asymmetries in ongoing oscillations, which, when averaged over trials, might appear as a slow ERP"* (de Munck & Bijma, 2010; Jensen, van Dijk, & Mazaheri, 2010; Nikulin et al., 2007; as cited in Cohen, 2014, p. 57).

Further empirical evidence and understanding of their neurophysiological mechanisms is needed before a single model can be selected, however it may be the case that different ERP components have different neural origins, thereby limiting the possibility of a unified explanation (Krieg et al., 2011; Yeung, Bogacz, Holroyd, Nieuwenhuis, & Cohen, 2007; Burgess, 2012; as cited in Cohen, 2014).

Another question about the nature of electrical fields measured using EEG that we should pose is whether they are causally involved in cognitive processes. Unfortunately, Cohen (2014) notes that the current evidence is insufficient to draw this conclusion reliably, however several lines of evidence suggest their causal involvement in neural computation and information transfer. One example comes from in vitro studies on the relationship between local field potential oscillations and synaptic events such as long-term potentiation in the hippocampus which is thought to allow memory formation through Hebbian learning and occurs predominantly at specific phases of theta-band oscillations (Axmacher, Mormann, Fernández, Elger, & Fell, 2006; as cited in Cohen, 2014). Another one comes from studies which show that *"the timing of many but not all neurons is constrained by the local field potential, such that neurons are more likely to emit an action potential during some phases of the local field potential oscillation"* (Cohen, 2014, p. 58), which has led to theories of phase coding based on the synchronisation between action potential timing and field potential phase (Lisman & Otmakhova, 2001; Yamaguchi et al., 2007; as cited in Cohen, 2014). Other theories focus on the interregional oscillatory synchronisation which may be crucial for perceptual and cognitive processes through phase synchronisation between spatially disparate neural networks, allowing the transfer of information (Akam & Kullmann, 2012; Fries, 2005; Singer, 1993; as cited in Cohen, 2014).

What would happen if it turned out that electrical fields were not causally involved in cognition and that all these theories were therefore majorly flawed—would it be the end of cognitive neurophysiology? According to Cohen (2014), not at all. The use of field potential oscillations to study brain organisation would be still valid, much like the blood oxygenation-level-dependent (BOLD) signal measured by fMRI is thought to be an important index of brain function even though it is not widely believed that it is a causal mechanism of neural information processing. One can therefore safely assume that regardless of causality, the results of ERP studies can still provide useful and valid insights into brain function.

1.3 TIME-DOMAIN ANALYSIS

1.3.1 OVERVIEW

The computation of ERPs is straightforward. Each trial contains signal which is similar across trials, and noise which is randomly distributed around zero, therefore aligning the time-domain EEG signal to a given event based on our hypothesis and averaging across many trials cancels out noise and leaves us with the signal known as ERP. The resulting average is considerably smaller in magnitude because all non-phase-locked activity (mostly frequencies above 15 Hz) is subtracted out during averaging which acts as a low-pass filter (Cohen, 2014).

If one intends to make inferences about cognitive processes, there is a range of issues and analytical approaches that need to be considered before reaching a reliable conclusion, and as the primary objective of this work is to provide a comprehensive overview with a practical demonstration, let us now move on to some specifics of ERP analysis.

According to Cohen (2014), while filtering ERPs is not always necessary (particularly if one has many trials to work with and is mainly interested in late components with greater temporal extension), it is a common practice in order to minimise residual high-frequency fluctuations that allows for better peak-based component quantification be reducing the risk of mistaking random spikes for signal. In addition to averaging over

trials, filtering can be done by averaging across subjects which leads to further temporal smoothing as the timing of brief neural events is not exactly the same between individuals, or by using digital filters, however one should be aware that poorly designed filters can introduce ripples (also known as ringing artefacts) that may subsequently be mistaken for oscillations—this can be avoided by constructing filters with gentle transition zones. Apart from ripples, some filter settings such as the use of forward-only or causal filters could cause systematic biases in ERP components (Acunzo, MacKenzie, & van Rossum, 2012; Rousselet, 2012; as cited in Cohen, 2014). These pitfalls are discussed in detail in chapter 14 of Cohen's (2014) book where he also provides quantitative and qualitative measures of proper filter construction. For these reasons, one should always consider whether and to what extent is it necessary to apply additional filtering beyond averaging, otherwise the results could end up contaminated with unnecessary steps.

In order to visually inspect and confirm the timing of task event representations, one could use either a butterfly plot or topographical variance plots, e.g. *global field power*. The former is useful for detecting bad or noisy electrodes by showing the ERP from all electrodes referenced to the global average overlaid in the same figure so that it is easy to detect high variance. The latter is the standard deviation of activity over all electrodes which increases as different brain regions become more active and is also best detected when using the average reference (Murray, Brunet, & Michel, 2008; as cited in Cohen, 2014).

The spatial distribution of EEG recordings is often visualised in the form of topographical maps which provide a comprehensive and widely used standard for data exploration and reporting. They are created by plotting the electrodes arranged according to their layout on a two- or three-dimensional surface and interpolating the values over the surface, using colour coding to visualise the spatial distribution. The advantage of two-dimensional plots is that they show data from all electrodes simultaneously while three-dimensional plots require multiple views, though they are a bit easier to interpret. Whichever projection is used, they provide excellent and rapid data inspection possibilities and allow one to confirm

the timing of task events or detect bad or noisy electrodes (Cohen, 2014).

Finally, we arrive at ERP images which are two-dimensional (time × trial) representations of EEG data from a single electrode before averaging stacked vertically and colour coded to show changes in amplitude as changes in colour. They can be used to reveal trials with large amplitudes that are likely to contain artefacts or to link trial-varying task parameters or behaviours to the signal by sorting the trials according to values of the aligning event, e.g. reaction time.

1.3.2 STATISTICS

In addition to visual inspection, one must of course turn to statistics in order to test hypotheses and draw robust inferences based on quantitative approaches. On the other hand, it is necessary to select the right tools and pay close attention to qualitative patterns rather than rely on statistical thresholds and base all reporting on the rather infamous p-values. The following methods are commonly used in cognitive electrophysiology and their description in this chapter is based on Cohen's (2014) excellent overview. For more detailed information about the methods or examples, the reader is advised to read chapters 32–36 of his book.

The most basic consideration of every study in the field of cognitive electrophysiology is the unit of data for statistical analysis. A within-subjects design would be looking for differences between trials and is useful in cases where there is not enough subjects such as rare patient groups or many trials from several subjects, whereas a between-subject design would be looking for differences between individuals or groups of individuals, and is the more commonly used in cognitive electrophysiology. Within-subjects statistics (also called level-1) can take advantage of the larger number of trials where testing between subjects would be too unreliable—this is generally when the number of subjects is below eight (Cohen, 2014). Another potential use of within-subjects statistics in cognitive electrophysiology is to support claims about the robustness of an effect which can be considered more robust if it turns out to be statistically significant within 90% of subjects as opposed to an effect which is only

significant at the group level. The same approach may also be necessary for analyses whose raw values are either uninterpretable or incomparable across subjects, such as phase-amplitude cross-frequency coupling where the within-subject values must essentially be converted to a metric that can be used at the group-level (Cohen, 2014).

The more commonly used approach is to average data from all trials for each subject and then perform between-subjects statistics (also called level-2) which may operate with as few as ten or twenty points per test as opposed to hundreds of points for within-subjects analyses. Note that "*it is important to realise that within-subjects and group-level analyses have different goals and different interpretations; within subjects analyses provide information regarding the cross-trial variability of an effect relative to the magnitude of the effect [but] they provide no information regarding the generalisability of the effect to other subjects [whereas] group-level analyses […] provide information regarding the consistency of the direction of the effect across the group of subjects, and provide little information regarding the within-subject variability*" (Cohen, 2014, 451–452). One could therefore easily miss effects that are only statistically significant at subject-level or misinterpret a general effect due to high cross-trial variability.

As for statistical significance, we need to at least briefly address the question of p-values which are somewhat frustrating to many people due to how those who do not understand statistics misrepresent them and how the five per cent value is practically set in stone. (Cohen, 2014) concurs that p-values are entirely arbitrary and despite the accepted significance threshold, there is little difference in terms of effect sizes between $p = 0.051$ and $p = 0.049$, however $p = 0.05$ still remains the accepted threshold and one should avoid questionable methods just to fit the criterion. Whether the statistical threshold is appropriate for a given analysis also depends on its goals. Hypothesis-driven research with a small number of statistical tests motivated by theories or previous results may be considered reliable under the five per cent threshold whereas exploratory data-driven analyses that use thousands or more tests should go as low as 0.1 % or be used in combination with multiple-comparisons corrections, which is often necessary with ERP data that may consist of multiple dimensions, for

example subject × electrode × time. This correction can be performed in a number of ways depending on the nature and objectives of the intended analysis.

One of them—the Bonferroni correction (Weisstein, n.d.), which is commonly used in psychology—involves dividing the p-value by the number of statistical comparisons, however it assumes that the data are not autocorrelated and corrects for the number of tests instead of the amount of information available in those tests, which renders it insensitive and therefore likely inappropriate for cognitive electrophysiology data in which case nonparametric permutation testing (Nichols & Holmes, 2001) should be used (Cohen, 2014). Another disadvantage of p-values, this time specifically derived through parametric statistical procedures, is that they are based on the number of data points, so with enough data points, practically any small effect can be shown as statistically significant and vice versa (Cohen, 2014), which is perhaps why even psychology journals are starting to outright ban p-values (Woolston, 2015) and statisticians refuse to use them where they feel that they do not make any sense (Bates, 2006).

There are two main types of statistics based on the assumptions about the distribution from which the data and their parameters were drawn—*parametric*, which assume a known distribution (usually Gaussian) and statistical parameters such as p-values are computed based on an theoretical distribution, and *nonparametric*, which make no such assumption and instead are evaluated using permutation testing which is covered later. The p-values are then *"derived not from theoretical distributions but, rather, from distributions that are created from the data, by creating situations within the dataset that could arise if the null hypothesis were true"* (Cohen, 2014, p. 456). The selection of appropriate tools therefore depends upon characteristics of the data and one's hypotheses, or lack thereof. For data that are not normally distributed, nonparametric tests are more appropriate; if one takes the exploratory approach and needs to correct for multiple comparisons, again nonparametric tests might be beneficial; mixed- or random-effects designs based on hypotheses about main effects and interactions, one should choose parametric tests; if these are appropriate but the data is not normally distributed, one can use transformations to sat-

isfy this requirement, for example baseline correction for time-frequency power. Overall, Cohen (2014) advocates the use of standard parametric statistics whenever possible for hypothesis testing and nonparametric permutation testing when there are no such hypotheses or parametric tests would be inappropriate.

1.3.3 NON-PARAMETRIC PERMUTATION TESTING

Permutation testing can be used to assess statistical significance of M/EEG data (Maris & Oostenveld, 2007) without relying on assumptions about the underlying distributions and their parameters, and while providing correction for multiple comparisons. Instead of comparing the test statistic such as the t-value against its theoretical distribution under the null hypothesis and computing the probability (p-value) of obtaining a t-value (or other test statistic) at least as large as the observed one, no such assumptions are made for nonparametric permutation testing and *"the distribution is created from the data you have […] by observing what the test statistic would be if the null hypothesis were true […] for example by iteratively shuffling the condition labels over trials (for within-subject analyses) or over subjects (for group-level analyses) and recomputing the test statistic"* (Cohen, 2014, p. 460). If one has an EEG dataset with two conditions A and B, and predicts that a measure of activity will be greater in A compared to B, the hypothesis is evaluated by performing a t-test between the two conditions, where the null-hypothesis would be that there is no difference. Therefore, if trials were randomly exchanged between conditions, the expected test statistic value would be zero and if it was not, it would be attributed to a sampling error or outliers in one of the conditions. The same principle applies for two continuous variables such as when testing the statistical significance of a correlation coefficient. *"In this case, creating a test statistic value under the null hypothesis involves swapping data points in terms of their value of variable B; because the independent variable has been randomly shuffled, a correlation coefficient of zero would be expected, and any nonzero correlation coefficient would be attributed to sampling error or outliers […] when those procedures are repeated hundreds or thousands*

of times, the null hypothesis values over all iterations create a distribution of test statistic values observed under the null hypothesis" (Cohen, 2014, p. 462).

If the observed test statistic happens to be well within the boundaries of the null-hypothesis test statistic values, the null hypothesis can not be rejected, otherwise, if it is distant enough, the effect is considered statistically significant.

The number of iterations required to obtain robust results depends on the number of trials and conditions and in practice falls between several hundred to a few thousand iterations. The more iterations (permutations) are computed, the more reliable the result becomes, but of course at the expense of computation time. It is always better to choose more for the sake of greater confidence—for most applications, 1000 iterations should suffice but if the data are especially noisy or from just a few trials, more is recommended (Cohen, 2014).

Statistical significance or p-value associated with the observed test statistic based on the distribution of test values expected under the null hypothesis can be computed in two ways: *a)* by simply counting the number of null hypothesis test values which are more extreme than the observed test value and dividing them by the total number of tests, or *b)* by converting the observed statistic to standard deviation unit of the null hypothesis distribution (Z-value) and evaluating its position on a Gaussian probability density. Since the data are randomly reshuffled, the p-value changes every time the null hypothesis distribution is recomputed, but the changes should be minimal if the number of iterations is sufficient. If the fluctuation is high and around a p-value critical for reporting, Cohen (2014) recommends a *"meta-permutation test"*, i.e. running a test with more than 1000 iterations about 20 times and averaging the p-values together.

CHAPTER 2

Methodology and results

2.1 DATA COLLECTION AND PRE-PROCESSING

In order to demonstrate the described practical approaches to the analysis of event-related potentials and to provide a comprehensive guide, we decided to utilise the data from a short-term memory delayed object discrimination task collected by Riečanský et al. (2013) in which 32 healthy adult volunteers were presented with a target stimulus which had to be maintained in STM during a delay interval and then used to assess whether a subsequently presented probe stimulus was rotated with respect to the target, the capital letter 'E', presented for 520 ms in an oblique orientation (45°, 135°, 270°, or 315°) randomly varied across trials (Fig. 2.1).

The target presentation time and target stimulus were chosen based on pilot tests which had revealed variance in discrimination thresholds between characters, and compromised performance with shorter presentation times. The target was replaced by a random-dot mask (to prevent the formation of an afterimage) which was presented for the delay interval of 3000 ms during which the subject had to keep the exact visual image of the target stimulus in memory. Following the delay interval, the probe stimulus was presented. This was the same character rotated by 2°, 4°, 6°, or 8° with respect to the target either clockwise or counter-clockwise, and remained on the screen until a response was given using keys that

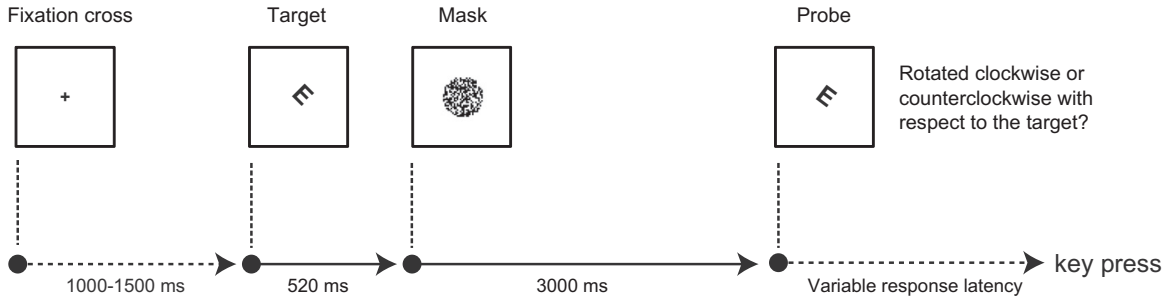

Figure 2.1: A schematic depiction of the delayed object discrimination task. The first 520 ms after the stimulus onset are considered the encoding stage, the following 3000 ms the maintenance stage, and the variable time span until a key press represents the retrieval stage. Figure used with permission from Riečanský et al. (2013).

were counterbalanced across subjects. The task was to determine in which direction the orientation of the probe deviated from that of the target. In total, 160 trials were presented, i.e. 40 for each target-to-probe offset (20 for clockwise and 20 for counter-clockwise deviation).

The EEG signal was recorded from 61 equidistant scalp sites using sintered Ag-AgCl electrodes mounted on an elestic cap (EASYCAP GmbH, Herrsching, Germany), referenced to a non-cephalic sternovertebral reference derivation (Stephenson & Gibbs, 1951). The electrodes were connected to a 64-channel DC-amplifier (Ing. Zickler GmbH, Pfaffstätten, Austria) and their impedances were kept below 3 kΩ. The signal was analog-filtered in the range of 0–1000 Hz, sampled at 3000 Hz, and digitally down-sampled to 250 Hz resolution. Further details describing the recording procedure can be found in Riečanský et al. (2013, p. 165).

The behavioural data from trials in which the reaction time was shorter than 300 ms and longer than 3000 ms were discarded. The percentage of correct responses for each subject and target-to-probe angular disparity (2°, 4°, 6°, and 8°) was calculated. Four subjects were excluded because their results either did not exceed the 75 % cut-off value for random responding even at the largest disparity, or showed no evident increase in accuracy as a function of angular disparity (Riečanský et al., 2013).

The EEG data were originally processed using the EEGLAB toolbox

(Delorme & Makeig, 2004) and MATLAB (The MathWorks, Massachussets). The signal was digitally filtered in the range of 0.1–80 Hz and carefully inspected for artefacts. The portions of data containing coarse artefacts were removed and then independent component analysis (Delorme & Makeig, 2004) was performed. Components separating artefactual signals such as eye-movements or blinks were identified based on activity time course, topography and spectrum, and eliminated from the data (Jung et al., 2000). Each epoch was then baseline-corrected to the mean activity within 300 ms preceding target onset and error trials were eliminated. ERPs were then calculated by truncating individual epochs at the time of the onset of the probe stimulus and averaging across all trials with respect to the onset of the target stimulus.

2.2 SENSOR-LEVEL ANALYSIS

2.2.1 TOOL OVERVIEW

The pre-processed data were taken for further analysis using FieldTrip (Oostenveld, Fries, Maris, & Schoffelen, 2011) and R (R Development Core Team, 2012) in multiple steps which are discussed in the next sections in chronological order. Although we are using specific tools to demonstrate the practical approach, the same methods or similar alternatives are available in other toolboxes and libraries for different environments.

FieldTrip (Oostenveld et al., 2011) is an open source software package for the analysis of EEG, MEG, and other electrophysiological data, that has been in development since 2003, currently at the Donders Institute for Brain, Cognition and Behaviour of the Radboud University Nijmegen, the Netherlands, together with other collaborators. It is implemented in MATLAB and has a relatively large user base—according to Google Scholar, the reference paper currently has 1273 citations and in 2010 there were approximately 650 subscribers to the e-mail discussion list. The toolbox provides algorithms for data preprocessing, event-related field/response analysis, parametric and nonparametric spectral analysis, forward and inverse source modelling, connectivity analysis, classification,

real-time data processing, and statistical inference. It offers no graphical user interface and the user interacts with high-level functions organised in scripts or entered to the command line. A script consists of a sequence of functions which perform distinct parts of the analysis pipeline and provide a comprehensive overview compliant with reproducible research guidelines.

High-level functions always take one or more parameters, the first one being a configuration structure containing options related to the function, detailing its expected behaviour, optionally followed by one or more data structures that are to be processed. The output is another MATLAB structure that includes a configuration field with those settings in order to provide the option of backtracking the steps that were taken to get to that point.

FieldTrip makes use of native structure arrays to define its own data structures for different types of data, for example segmented sensor-level time domain data would contain different fields than their time-frequency representation. For more technical details, consult the reference paper (Oostenveld et al., 2011) or the website.

FieldTrip can be downloaded from various locations listed at http://www.fieldtriptoolbox.org/download. Upon completion, the folder needs to be added to the MATLAB path and initiated with the command ft_defaults which provides access to all the available functions from the environment.

2.2.2 IMPORTING THE DATA

The first task ahead is to import our data into a format that FieldTrip understands. It provides import functions for many popular formats listed at http://www.fieldtriptoolbox.org/reading_data including the popular EEGLAB toolbox which was used to pre-processed our data. In this case, however, we will use raw ERP data and create the data structure from the scratch in order demonstrate how universal it is.

The raw data are ERP organised in a three-dimensional array of 64 channels × 1075 time points × 28 participants called `erp`. These are ac-

companied by another 1×1075 matrix containing values respective to the time points (`timesOrient`) in milliseconds and a 28×1 matrix with the overall accuracy scores for each participant (`accuracy`). When these are loaded into MATLAB, we have two options described at the FieldTrip website ("FAQ: How can I import my own data format", 2015): we can either extend FieldTrip with a custom algorithm to handle the import procedure which is especially handy when you are working with lots of data in the same format, in which case one needs to alter the reading functions, or circumvent these by reformatting the data within MATLAB into a compatible data structure that needs to a regular structure that contains the following fields:

`struct.label` a cell-array of electrode labels (channels × 1)

`struct.fsample` the sampling frequency in Hz (numeric)

`struct.trial` a cell-array of channels × samples matrices, one for each trial (1 × trials)

`struct.time` a cell-array of 1 × samples time axes for each trial (1 × trials)

`struct.trialinfo` an optional field that can be used for miscellaneous trial-specific information such as condition numbers or reaction times

For a continuous recording, one can either use a different data structure with trial definitions described in the documentation, or use the function `ft_redefinetrial` to split it into individual trials as above. Since our trials (in fact representing individual participants) were already segmented, we did not need to do this and neither did we use the field trialinfo.

The first step was to create the structure using

```
data = struct('label', {cell(61, 1)}, 'fsample', {0.25}, 'trial', {cell
    (1, 28)}, 'time', {cell(1, 28)});
```

to reflect the Easycap M10 layout of 61 channels, the sampling frequency 250 Hz divided by 1000 (because our time labels were in milliseconds), the number of trials (28) representing time-locked averaged data from participants, and the time labels for each participants individually (28).

Then we needed to populate it, starting with the electrode labels. FieldTrip provides many 2D template files located in `fieldtrip/template/layout` including what was seemingly our M10 layout, however the electrode labels and the order of positions were different and later found to be specific to the Brain Research Lab of the Department for Psychology at the University of Vienna where the data were collected. If the required layout is not included with FieldTrip or available online, as in our case, there are guidelines for constructing a custom one ("Specifying the channel layout for plotting", 2015).

Another option is to take a 3D template from `fieldtrip/template/electrode`, where again was an M10 file which defined the position of electrodes by two angles on a sphere, albeit with different labels and ordering. Fortunately, we were provided with a file in the same format and after changing the header to be the same as the one in the template (FieldTrip needs this to correctly recognise the format), we could use the function `ft_read_sens` which in this case assumed a given sphere diameter to construct a 3D electrode layout with the correct labels, and copy them into our data structure:

```
elec = ft_read_sens('easycap-M10-BRL.txt');
data.label = elec.label;
```

Next, we had to import the time labels (-300, -296, …, 3992, 3996) from `timesOrient`

```
data.time(1:28) = {timesOrient};
```

and the ERP data themselves (the first three rows of `erp` were not actual channels and hence were omitted)

```
for i = 1:28
  data.trial(i) = {erp(4:64, :, i)};
end
```

This concluded the creation of our minimal data structure and left us only with the structure itself and the accuracy matrix.

2.2.3 TIME-LOCKED AVERAGING

We then used the function ft_timelockanalysis to compute the grand average over all participants and produce a data structure that is fully

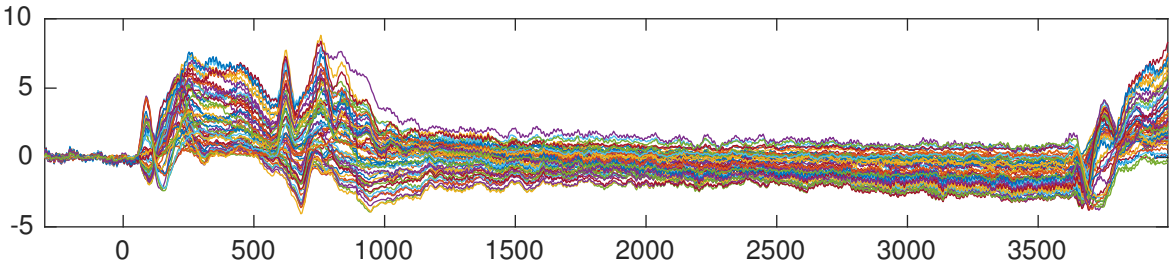

Figure 2.2: ERP data from all 61 channels plotted over the whole time period between −300 and 4000 ms.

compatible with all functions that provide methods for further analysis of time-locked data. This function allows one to select a subset of channels or trials, compute the covariance matrix which is used for source reconstruction, and remove or keep the original trials depending on whether they are needed. In this case, the covariance window was chosen as everything before the zero time point where there was no activity of interest and the reasons for this will be described later when dealing with source reconstruction.

```
cfg = [];
cfg.channel = 'all';
cfg.trials = 'all';
cfg.keeptrials = 'yes';
cfg.covariance = 'yes';
cfg.covariancewindow = [-inf 1000];
tlck = ft_timelockanalysis(cfg, data);
```

The resulting data structure included additional fields, most importantly `avg` (61×1075) that contained the grand average which could be visualised simply using

```
plot(tlck.time, tlck.avg)
```

(Fig. 2.2) or a range of FieldTrip plotting functions which provide interactive interface and other useful features. This brought us to the next logical step: to look at the data and confirm that the import process and averaging were done correctly.

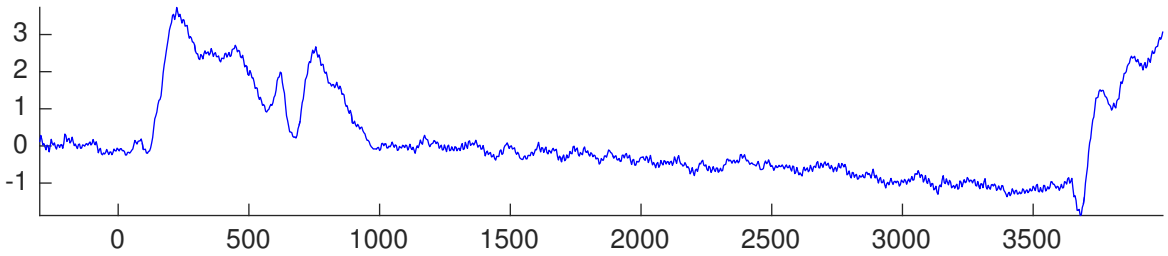

Figure 2.3: Grand average of ERP data from all 61 channels plotted over the whole time period between −300 and 4000 ms.

2.2.4 PLOTTING AT THE SENSOR LEVEL

FieldTrip provides three high-level functions for plotting 2D data at the sensor level, namely:

ft_singleplotER to plot ERPs of a single channel or the average over multiple channels, with the option to perform a baseline correction and select the desired channels and trials (Fig. 2.3).

```
cfg = [];
cfg.channel = 'all'; % or {'chan', 'chan', …}
cfg.trials = 'all'; % or [trial, trial …]
cfg.baseline = [-inf 0]; % prestim period
cfg.xlim = 'maxmin'; % or [max min]
cfg.ylim = 'maxmin'; % or [max min]
ft_singleplotER(cfg, tlck);
```

ft_multiplotER to plot ERPs arranged according to their location specified in the layout, including the additional functionality of the previous function. This is useful to visualise the signal at each electrode over time (Fig. 2.4).

```
cfg = [];
cfg.layout = 'easycap-M10-BRL.txt';
cfg.trials = 'all'; % or [trial, trial …]
cfg.showlabels = 'yes';
cfg.fontsize = 12;
cfg.hlim = 'maxmin'; % or [max min]
cfg.vlim = 'maxmin'; % or [max min]
ft_multiplotER(cfg, tlck);
```

ft_topoplotER to plot the topographic distribution of ERPs over the scalp in a two-dimensional space. The user can control a huge range of parameters that are described in the documentation (help ft_topoploter),

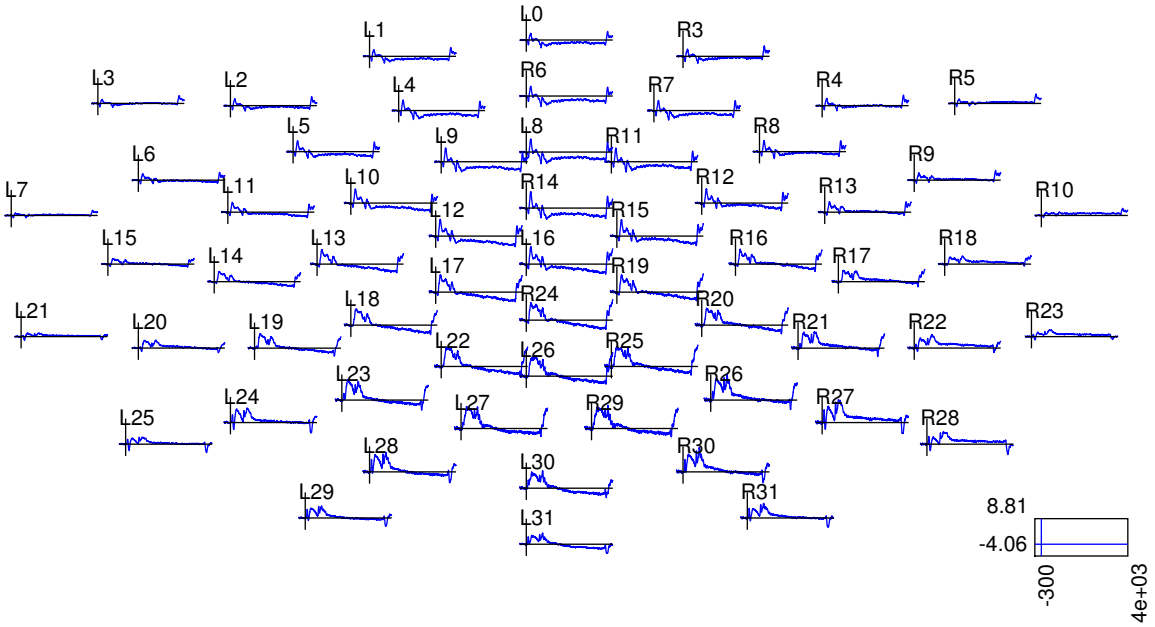

Figure 2.4: The result of the function `ft_multiplotER`. Data from each channel are plotted at its location according to the layout.

including colours, limits, channel markers, methods of interpolation and plotting styles, resolution, shading, etc. One also has the ability to highlight specific channels for publication purposes (Fig. 2.5).

```
cfg = [];
cfg.layout = 'easycap-M10-BRL.txt';
cfg.trials = 'all';
cfg.xlim = 'maxmin'; % time
cfg.zlim = 'maxmin'; % amplitude
cfg.colorbar = 'yes';
cfg.colormap = 'jet';
cfg.marker = 'labels';
cfg.markersize = 5;
cfg.markerfontsize = 12;
cfg.style = 'both'; % also straight, contour, fill
cfg.gridscale = 500; % resolution, default 67
ft_topoplotER(cfg, tlck);
```

The same function can also provide multiple plots representing different time windows in a single figure by altering the `cfg.xlim` parameter. As an example, to get the topographic distributions between 300 and 500 ms in 20 ms intervals, one would write:

```
cfg = [];
```

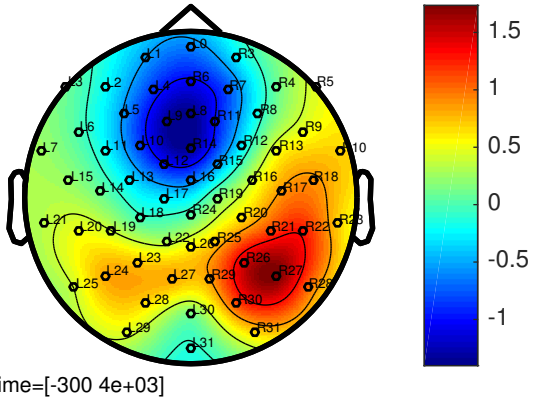

Figure 2.5: The result of the function `ft_topoplotER`. Data from a given time interval are averaged, interpolated between the channel locations, and extrapolated in order to be projected inside a circle representing the head.

```
cfg.layout = 'easycap-M10-BRL.txt';
cfg.xlim = [300:20:500]; % time
cfg.zlim = [-1 7]; % default different for each plot
cfg.colormap = 'jet';
cfg.comment = 'xlim';
cfg.commentpos = 'title';
ft_topoplotER(cfg, tlck);
colorbar; % one is enough
```

Once we confirmed that the data appeared as expected and there were no errors, we could move on to another stage of our analysis at the sensor level.

2.2.5 STATISTICAL DESIGN

Since we effectively had a single condition (one ERP per participant coupled with a single scalar value representing the overall accuracy on the task) and wanted to determine which portion of the signal could be deemed the best predictor of individual accuracy (and by extension the capacity of one's visual short-term memory), thereby performing an exploratory analysis without a priori hypotheses, we chose to conduct a correlation test with nonparametric permutation testing to determine statistical significance for reasons discussed in section 1.3.2. In general terms, this is applicable to any statistical testing of the relation between a neurobio-

logical and a behavioural variable, and is outlined in the FieldTrip user documentation [web reference here]. Neurobiological signals are typically considered the dependent variable whereas the independent variable can be either the experimental condition (task instructions, stimulus type, etc.) or a behavioural variable such as the response accuracy or reaction time. Since the latter is not experimentally controlled, the dependent vs independent role is assigned by convention—in FieldTrip, the one with the least number of dimensions is considered independent. One must also make a distinction between categorical (e.g. left vs. right hemifield) and quantitative (e.g. response time or accuracy) variables because they are both associated with different test statistics, as are designs with a single observation for every unit of observation (between-UO) vs multiple conditions that are to be compared (within-UO). In single-subject studies, these units are considered trials whereas in multi-subject studies, they are the subjects. FieldTrip then provides specific test statistics for every combination of these, as well as other more complex options.

In our case, we had a quantitative independent variable and a between-UO design so according to the documentation [cite documentation], we needed to use the independent samples regression T-statistic implemented in the function ft_statfun_indepsamplesregrT. This only depends on the Pearson correlation between the dependent and independent variable, therefore we could also use ft_statfun_correlationT which expresses the above T-statistic as a function of this correlation.

Since we did not have a hypothesis to test and were doing an exploratory analysis on data from many subjects, channels, and time points, a permutation correlation test was our method of choice. As outlined in [chapter n], the null hypothesis that the computed probability distribution of the dependent variable is identical for all possible values of the independent variable, i.e. that they are statistically independent, is tested by randomly permuting the values of the independent variable. In a between-UO design, this is done by permuting across the units of observations (in our case subjects), and in a within-UO design across the conditions. It is also important to point out that the three test statistics for quantitative independent variables implemented in FieldTrip are “only

sensitive to deviations from statistical independence that can be captured as a linear relation between the dependent and the independent variable [although] new statistics [optimised for particular non-linear deviations] can be formulated" [cite this].

FieldTrip contains three basic statistical functions according to the type of data that is to be tested, namely ft_timelockstatistics for time-locked data (the output from ft_timelockanalysis, in our case stored in the variable tlck), and ft_freqstatistics and ft_sourcestatistics for time-frequency and reconstructed source space data respectively. These high-level functions take a number of parameters that specify their exact behaviour. In order to compute significance probabilities and critical values of a non-parametric permutation correlation test as described earlier in this section, we used ft_timelockstatistics with the following configuration:

```
cfg = [];
cfg.statistic = 'ft_statfun_correlationT';
cfg.method = 'montecarlo';
cfg.numrandomization = 5000;
cfg.design = accuracy;
cfg.ivar = 1;
cfg.alpha = 0.001;
cfg.tail = 0;
cfg.correcttail = 'alpha';
stat = ft_timelockstatistics(cfg, tlck);
```

Note that some of these parameters are not available in the documentation for the ft_timelockstatistics function but rather in the ft_statfun_correlationT or other ft_statfun_xxx functions that implement individual test statistics, and ft_statistics_montecarlo which performs a non-parametric test by calculating Monte-Carlo estimates of the significance probabilities and/or critical values from the permutation distribution. The meaning of the parameters specified in the configuration structure is as follows:

`cfg.statistic` a string specifying which test statistic is to be used

`cfg.method` a string specifying the method for calculating the significance probability and/or critical value; our option calculates Monte-Carlo estimates

`cfg.numrandomization` the number of randomisations

cfg.design a matrix containing the independent variable, in our case the 28 × 1 accuracy matrix

cfg.ivar the number of independent variables

cfg.alpha the critical alpha value for rejecting the null-hypothesis per tail (default 0.05)

cfg.tail 0 for a two-sided test; -1 or 1 for a one-sided test, left or right respectively

cfg.correcttail a string specifying whether to correct p-values or alpha-values when doing a two-sided test

The output from this function is a structure with fields containing the results of the statistical analysis. Some of them are described here:

stat.prob the corrected p-values (channels × time)

stat.cirange the respective confidence interval range (channels × time)

stat.mask a binary matrix where ones denote channel-time points for which the test statistic is lower than the critical value given the set alpha, i.e. the result is statistically significant, and zeroes vice versa (channels × time)

stat.stat the test statistic (channels × time)

stat.rho the correlation coefficient (channels × time)

These results can be visualised using either standard MATLAB plotting functions or the FieldTrip ones specified in section 2.2.4. Since the FieldTrip plotting functions have no option for element-wise multiplication of the parameters with the mask, it is useful to compute three additional fields for prob, stat, and rho weighted by mask so that one is able to plot the masked data:

```
stat.maskprob = stat.prob .* stat.mask;
stat.maskstat = stat.stat .* stat.mask;
stat.maskrho = stat.rho .* stat.mask;
```

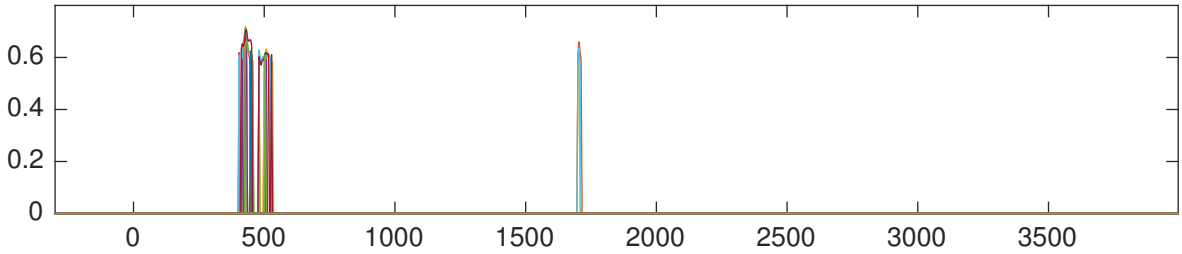

Figure 2.6: The correlation coefficient for every time point and channel showing only statistically significant ($p < 0.001$) results.

This allows one to easily plot the correlation coefficient for all electrodes as a function of time only where the correlation is statistically significant given the selected alpha value (Fig. **??**):

```
plot(stat.time, stat.maskrho);
xlim([stat.time(1) stat.time(end)]);
```

The function `ft_singleplotER` is not particularly useful for this since it plots either a single electrode or the mean of multiple electrodes, however `ft_multiplotER` can provide a useful plot of any of these values specified as a function of time arranged according to the electrode layout. Unfortunately, the layout is not transferred from `tlck` to `stat` so we had to use the option `cfg.layout` again to select the one we used earlier:

```
cfg = [];
cfg.showlabels = 'yes';
cfg.fontsize = 12;
cfg.layout = 'easycap-M10-BRL.txt';
cfg.parameter = 'maskrho'; % maskstat, stat, ...
ft_multiplotER(cfg, stat);
```

We could then use the interactive plots to select a number of channels where the statistically significant correlation was high and explore their mean as well as the topographical map, in this case showing the correlation coefficient.

There was only one considerable peak approximately between 420 and 540 ms on electrodes R11, R12, R14, R15, L10, and L12, where the mean correlation coefficient was reaching as high as 0.7 and was consistently high throughout the period (Fig. 2.7).

To visualise the results in multiple topographical plots between 420 and

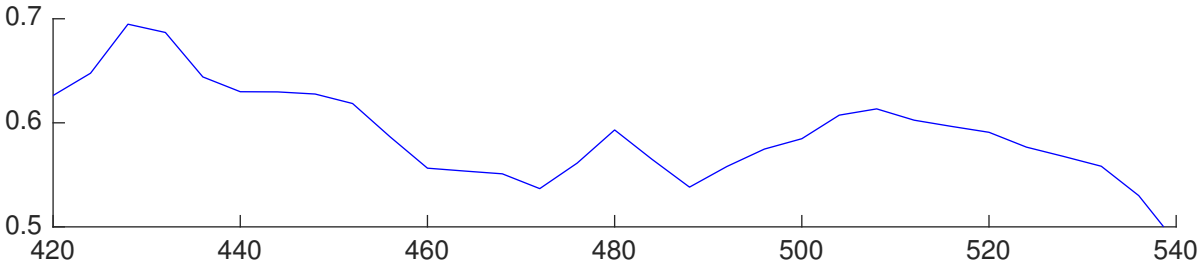

Figure 2.7: The mean of correlation coefficient from electrodes R11, R12, R14, R15, L10, and L12, between 420 and 540 ms, showing only statistically significant ($p < 0.001$) results.

540 ms averaged over 20 ms intervals, we used the function `ft_topoplotER` with the following settings:

```
cfg = [];
cfg.xlim = [420:20:540];
cfg.zlim = [0 0.7]; % amplitude
cfg.colorbar = 'yes';
cfg.colormap = 'jet';
cfg.style = 'both'; % also straight, contour, fill
cfg.gridscale = 200; % resolution, default 67
cfg.layout = 'easycap-M10-BRL.txt';
cfg.parameter = 'maskrho';
ft_topoplotER(cfg, stat);
```

These results indicate that there is a consistent spatio-temporal window of strong correlation ($r > 0.5$, $p < 0.001$) between task performance (response accuracy) and the amplitude of the event-related potentials approximately between 420 and 540 ms after the onset of the target stimulus, peaking at around 430 ms, i.e. during the late encoding stage, while the electrodes yielding high correlation are situated over anterior scalp.

2.2.6 CORRECTING FOR MULTIPLE COMPARISONS

In order to correct for multiple comparisons, FieldTrip implements various methods but the two most sensitive for neurophysiological data are the 'maximum statistic' which, instead of building up the null distribution separately for each electrode × time pair, selects the maximum test statistic at each permutation and uses it to create the null distribution, and the 'clustering' approach which is based on the fact that effects at neighbouring

electrodes and time-points are highly correlated—this is called a cluster-based permutation test (Maris & Oostenveld, 2007).

Maximum statistic correction

Based on the uncorrected results, we decided to narrow the temporal scope of our analysis by testing only between 400 and 600 ms (51 time points) in order to further reduce the number of statistical tests and increase the sensitivity of the multiple comparisons correction. First, we tried the maximum statistic correction at a five per cent alpha level; the other settings were left unchanged:

```
% ...
cfg.correctm = 'max'; % correction method
cfg.alpha = 0.05;
cfg.latency = [400 600]; % temporal window
stat_max = ft_timelockstatistics(cfg, tlck);
```

These results were more focused and revealed a peak of strong correlation ($r > 0.5$, $p_{corr} < 0.05$) between 420 and 440 ms over electrodes R11, R12, R14, R15, L8, L9, and L10, and a wider peak between 424 and 452 ms over electrode L12 where the correlation was consistently just below $r = 0.7$ for 20 ms. This allowed us to reach practically the same conclusion as before, only with greater confidence.

Cluster-based correction

Finally, we decided to try the cluster-based multiple comparisons correction which should be the most sensitive option for this type of data. This method requires the user to define neighbouring sensors which can be done in FieldTrip using the function ft_prepare_neighbours. Again, the authors of FieldTrip provide templates for the most commonly used layouts which are located in fieldtrip/template/neighbours, however our Easycap M10 was missing so we had to create it ourselves using distance calculation:

```
cfg = [];
cfg.method = 'distance'; % ''triangulation, ''template
cfg.neighbourdist = 46;
neighbours = ft_prepare_neighbours(cfg, tlck);
```

We found that the alternative triangulation method of calculating the neighbouring sensors was creating too many neighbours for most of the electrodes according to the diagram provided by the manufacturer which we wanted to reproduce, therefore we used the distance method and tweaked the maximum distance between electrodes (`cfg.neighbourdist`) to match the diagram as closely as possible (we later confirmed that the few remaining differences had little impact on the results, however these could be resolved manually).

The output is a 1×61 structure containing fields `label` and `neighblabel` where every label is simply coupled with a cell array of neighbouring sensor labels. The user can either explore this structure or use the function `ft_neighbourplot` to produce a plot for visual inspection:

```
cfg = [];
cfg.neigbours = neighbours;
cfg.method = 'distance'; % distance
cfg.neighbourdist = 46;
cfg.elec = tlck.elec;
ft_neighbourplot(cfg);
```

At last, we ran the cluster-based permutation correlation test with the following settings:

```
cfg = [];
cfg.statistic = 'ft_statfun_correlationT';
cfg.method = 'montecarlo';
cfg.correctm = 'cluster';
cfg.clusteralpha = 0.05;
cfg.clustertail = 0;
cfg.clusterstatistic = 'maxsum';
cfg.neighbours = neighbours;
cfg.minnbchan = 2;
cfg.tail = 0;
cfg.alpha = 0.01;
cfg.correcttail = 'alpha';
cfg.design = accuracy;
cfg.numrandomization = 5000;
cfg.latency = [400 600]; % temporal window
stat_cluster = ft_timelockstatistics(cfg, tlck);
```

The meaning of the new parameters specified in the configuration structure is as follows:

`cfg.clusteralpha` the critical values used for thresholding the sample-specific test statistic to decide whether it should be considered

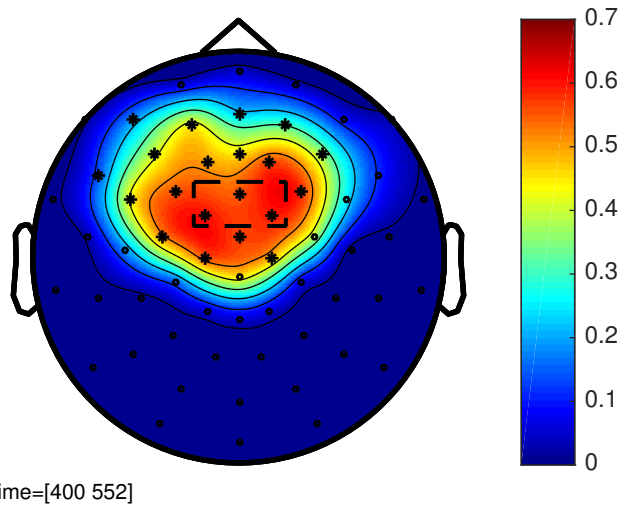

Figure 2.8: The topographical variance plot of correlation coefficient values between 400 and 552 ms corrected for multiple comparisons using the cluster-based correction. Electrodes R14, R15, and L14 yielding high correlation (mean $r > 0.7$) between 400 and 420 ms are highlighted.

a member of a larger cluster of samples; it does not affect the false alarm rate at the cluster level

`cfg.clustertail` the same as cfg.tail, only at the cluster level

`cfg.clusterstatistic` the test statistic evaluated under the permutation distribution, in this case the largest of the cluster level statistics

`cfg.neighbours` the structure containing the definitions of neighbouring sensors

`cfg.minnbchan` the minimum number of neighbouring channels that is required for a selected sample to be included in the clustering algorithm

Using the uncorrected exploratory analysis to select a temporal window of interest between 400 and 600 ms after the stimulus onset, the cluster-based permutation correlation test revealed a significant level of correlation ($r > 0.5$, $p < 0.01$) between task performance (response accuracy) and the amplitude of the event-related potentials. In this latency

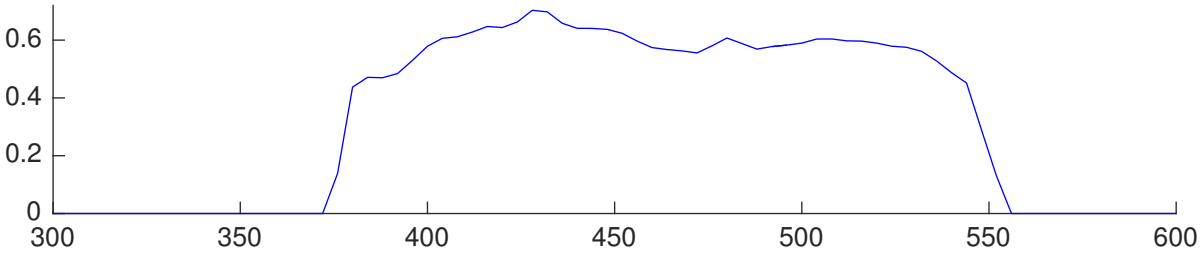

Figure 2.9: The mean correlation coefficient from electrodes R14, R15, and L14, corrected for multiple comparisons, showing high significant correlation ($r > 0.5$, $p < 0.01$) between 376 and 552 ms.

range, the correlation was most pronounced over anterior scalp, mainly under electrodes R14, R15, and L12 between 420 and 440 ms (mean $r > 0.7$), i.e. during the late encoding stage (Fig. 2.8). Since we noticed that the mean correlation coefficient was already high at 400 ms, we decided to extend the lower temporal boundary to 300 ms post-hoc and perform the test again, revealing a significant cluster of strong correlation between 376 and 552 ms ($r > 0.5$, $p < 0.01$, Fig. 2.9).

These results can be visualised for each time point using the function ft_clusterplot which highlights the significant sensors at a given critical alpha level, however we found the previous plotting functions more useful:

```
cfg = [];
cfg.alpha = 0.01;
cfg.layout = 'easycap-M10-BRL.txt';
cfg.comment = 'xlim';
cfg.highlight = 'on';
ft_clusterplot(cfg, stat_cluster);
```

2.3 SOURCE-LEVEL ANALYSIS

The first thing we need to state at the beginning of this section is that while we are quite confident about the sensor-level analysis, the documentation of source-level analyses in FieldTrip is not only sparse, but sometimes contradicts itself (mainly example scripts and tutorials), and despite our best effort, the following should be taken as an approximate guide leading to an approximate result rather than *the* way to do it, although the result

appears to be rather sensible. We were unsure whether to include it at all but since the result appears to be in line with previous conclusions and it a lot of energy to achieve it, we decided to provide a description that can at least be taken as a starting point and help the potential reader deal with undocumented or obscured steps. More about this lack of confidence and the reasons behind it will be discussed in the next chapter.

FieldTrip provides a range of algorithms for M/EEG source reconstruction which are listed in the `ft_sourceanalysis` documentation. We chose two methods appropriate for evoked data in the time domain which do not need any prior assumptions about the number of active sources: *minimum norm estimation* (MNE; Dale et al., 2000) and *linearly constrained minimum variance* beamforming (LCMV; Van Veen, van Drongelen, Yuchtman, & Suzuki, 1997). Since we could not get the MNE source estimation output to work with the functions responsible for statistical analysis that we needed due to what appeared to be a plethora of errors resulting from the lack of dimensionality definitions in the data structure containing the estimations for each participant, we later resorted to LCMV alone which did not have this issue. Our main source of guidance was provided by a tutorial from the 2014 NatMEG workshop in Stockholm, Sweden ("Beamforming evoked fields and potentials in combined MEG/EEG data", 2014), but other bits and pieces from all over the FieldTrip website and the e-mail discussion list (full of many, but mostly too specific questions) were used as well—unfortunately, we can not provide references to all of them.

Our goal was to identify the anatomical correlate associated with the evoked potential that we found to be strongly correlated with individual task performance (accuracy). According to Van Veen et al. (1997), a known activity pattern in the brain can be used to calculate the electromagnetic field generated on the outside of the brain in a unique fashion, which is the so-called *forward solution* that depends on various attributes of the model related to geometry and electromagnetic properties of tissues. Some models are simple spheres, others are based on anatomical data from imaging techniques such as magnetic resonance imaging (MRI) and are composed of multiple compartments with different properties.

However, the real problem is the *inverse* one. Given the electromagnetic field sampled at different sensor locations around the brain, we need to determine the underlying activity which does not have a unique solution, hence additional constraints must be introduced.

LCMV is based on spatial filtering, which is a technique designed to pass electrical activity from a specified location while attenuating activity from different locations. *"The power at the output of a spatial filter is an estimate of the neural power originating within the spatial passband of the filter; a map of neural power as a function of location is obtained by designing multiple spatial filters, each with a different passband, and depicting output power as a function of passband location,"* Van Veen et al. (1997, p. 867). The inverse filter is based on minimising the source power at a given location subject to *unit-gain constraint*, i.e. something the author does not pretend to understand in detail. It reportedly uses the temporal variation in signal in order to compute a covariance matrix which is then used to obtain a three-dimensional spatial distribution of the power of the sources which can be subjected to statistical analysis ("Beamforming evoked fields and potentials in combined MEG/EEG data", 2014).

2.3.1 HEAD MODEL

Ideally, one should use subject-specific MRI and recorded electrode positions to construct a model of the head (a volume conduction model), and the source model which is either a three-dimensional grid or a cortical sheet representing the source points. Since these were unavailable to us, we had to resort to templates.

For the volume conduction model, we decided to use one based on the *boundary element method (BEM)* developed by Oostendorp and van Oosterom (1989), provided by FieldTrip at `fieldtrip/headmodel/standard_bem.mat`. The model describes the flow of currents through the tissue while assuming realistic information about the interface between the skin, skull, and brain surfaces. Once the file was in our working directory, we invoked the following commands to load and plot the model as a mesh:

```
1 load standard_bem % later renamed to 'vol' for convenience
```

```
ft_plot_vol(vol, 'facecolor', 'none'); alpha 0.5;
```

Next, we needed to align the electrodes with the head model. This can be done automatically using fiducial landmarks—a set of at least three points common to the head and electrode space—however since our electrode layout did not include these, we had to do a manual approximation of the correct position using the function `ft_electroderealign` which provides an interactive plot that can be used to rotate, scale, or translate the position of the electrodes in order to achieve a snug fit (they are later projected on the outermost surface of the head model so there is presumably little need for absolute perfection):

```
cfg           = [];
cfg.method    = 'interactive';
cfg.elec      = tlck.elec;
cfg.headshape = vol.bnd(1); % outermost layer of the BEM
tlck.elec  = ft_electroderealign(cfg);
```

In order to confirm our adjustment, we used the following functions to plot the three layers of the head model with the new electrode locations overlaid:

```
ft_plot_mesh(vol.bnd(1), 'facecolor',[0.2 0.2 0.2], 'facealpha', 0.2, '
    edgealpha', 0.05);
ft_plot_mesh(vol.bnd(2),'edgecolor','none','facealpha',0.4);
ft_plot_mesh(vol.bnd(3),'edgecolor','none','facecolor',[0.4 0.6 0.4]);
ft_plot_sens(tlck.elec, 'style', 'sk');
```

More information about this process and the construction of the BEM from an anatomic MRI can be found online ("Creating a BEM volume conduction model of the head for source-reconstruction of EEG data", n.d.).

At this point, we also had to change the channel labels to numbers due to the peculiar way FieldTrip uses to determine the type of the sensors and even more peculiar errors associated with this issue:

```
for i = 1:61;
    tlck.elec.label{i} = num2str(i);
    tlck.label{i} = num2str(i);
end; clear i;
```

2.3.2 FORWARD SOLUTION

We experimented with different source models, initially starting with a cortical sheet template consisting of 8196 vertices located at `fieldtrip/template/sourcemodel/cortex_8196.surf.gii` but finally ending up with a regular grid with a 1 cm resolution like in "Beamforming evoked fields and potentials in combined MEG/EEG data" (2014), providing us with 1909 possible dipole locations inside the source compartment of the BEM. Increasing the resolution to 0.5 cm resulted in over 15,000 dipole locations and subsequently considerably longer computation times and greater number of statistical tests, however to no observable advantage, therefore we decided to keep it low in line with the tutorial. The grid along with the forward solution (a so-called lead field) were computed using the function `ft_prepare_leadfield` based on the electrode positions and the head model:

```
cfg = [];
cfg.elec = tlck.elec; % electrode locations
cfg.vol = vol; % head model
cfg.grid.resolution = 1;
cfg.grid.unit = 'cm';
grid = ft_prepare_leadfield(cfg);
```

2.3.3 INVERSE SOLUTION

The inverse solution is handled by the function `ft_sourceanalysis` which takes parameters such as the desired method, the forward solution computed at the preceding stage, the head model, or how to calculate the spatial filters (single trials vs their average). This is where we started to face major issues. We wanted to localise the activity between 376 and 552 ms which was the temporal region of interest based on our cluster-based permutation correlation analysis at the sensor level, but we could find no comprehensive guidance regarding the specification of such a time window. As mentioned, the covariance matrix computed at the very first step by `ft_timelockanalysis` is the key to all beamforming techniques, however we were unable to find reliable information about how to choose the covariance window.

"Beamforming evoked fields and potentials in combined MEG/EEG data" (2014) split the data into two groups of equal length around the stimulus onset, then compute the covariance matrix and calculate the spatial filters based on the entire interval, and then just apply those filters to the pre-stimulus and post-stimulus data respectively. At last, they use a cluster-based test and dependent samples T-statistic to obtain the contrast. Unfortunately, substituting it with the correlation T-statistic and adjusting the other parameters accordingly did not yield meaningful results.

On the other hand, "Localizing sources using beamformer techniques" (n.d.) used only one covariance matrix computed over the entire interval, however it was very short and they did not provide guidance regarding statistical analysis. The description of figure two by Ramirez (2008) further suggests that this is a valid practice: "*The data covariance matrix (used in methods b-d) [method c was LCMV] was computed using all of the data.*"

According to Woolrich, Hunt, Groves, and Barnes (2011, p. 1466) who discuss the calculation of covariance matrices for spatial filtering, "*when the noise levels are high, or when there is only a small amount of data available, the data covariance matrix is estimated poorly and the signal-to-noise ration (SNR) of the beamformer output degrades,*" which led us to believe that the covariance window should encompass the pre-stimulus baseline period as well as the post-stimulus range of interest, therefore we chose to extend it over the period of −300 to 520 ms (see the first snippet of code in subsection 2.2.3) to include both the pre-stimulus period and the encoding phase, which we hypothesised to be the primary correlate of task performance based on the sensor-level analysis. The inverse solution was computed for each participant separately using the function `ft_sourceanalysis`:

```
cfg = [];
cfg.method = 'lcmv';
cfg.grid = grid;
cfg.headmodel = vol;
cfg.rawtrial = 'yes';
cfg.keeptrial = 'yes';
source = ft_sourceanalysis(cfg, tlck);
```

2.3.4 STATISTICAL INFERENCE

The resulting `source` structure included an estimate of the neural power for each voxel within the brain and for every participant, therefore the task performance (accuracy) was tested in a between-subject design for every voxel using a cluster-based permutation correlation test, i.e. much like in subsection 2.2.6 except with many more 'sensors' and without the additional dimension representing time:

```
cfg = [];
cfg.statistic = 'ft_statfun_correlationT';
cfg.method = 'montecarlo';
cfg.numrandomization = 5000;
cfg.correctm = 'cluster';
cfg.clusteralpha = 0.05;
cfg.clusterstatistic = 'maxsum';
cfg.design = accuracy;
cfg.parameter = 'pow';
cfg.ivar = 1;
cfg.alpha = 0.001;
cfg.tail = 0;
cfg.clustertail = 0;
cfg.correcttail = 'alpha';
stat = ft_sourcestatistics(cfg, source);
```

The explanation of these settings can be found in subsection 2.2.6. There was no need to define the neighbouring sensors since the algorithm automatically detected and used the connectivity of voxels. We found twelve positive and zero negative clusters, one of them statistically significant ($p < 0.05$) and including voxels that were highly correlated with accuracy in the delayed orientation discrimination task ($r > 0.6$). Unfortunately, we were unable to use automatic labelling but we estimate that the correlation coefficient peaked in the right premotor cortex and supplementary motor area (Fig. 2.10).

In order to obtain the figure (2.10), we had to find the corresponding MRI file which happened to be located at `fieldtrip/template/anatomy/single_subj_T1.nii`, re-slice it, and interpolate the statistical map onto the voxels using `ft_sourceinterpolate`, however since the mask that is used to select only the statistically significant voxels did not work properly, we first had to create it manually and plot the figure:

```
stat.idx = find(stat.prob < 0.05);
```

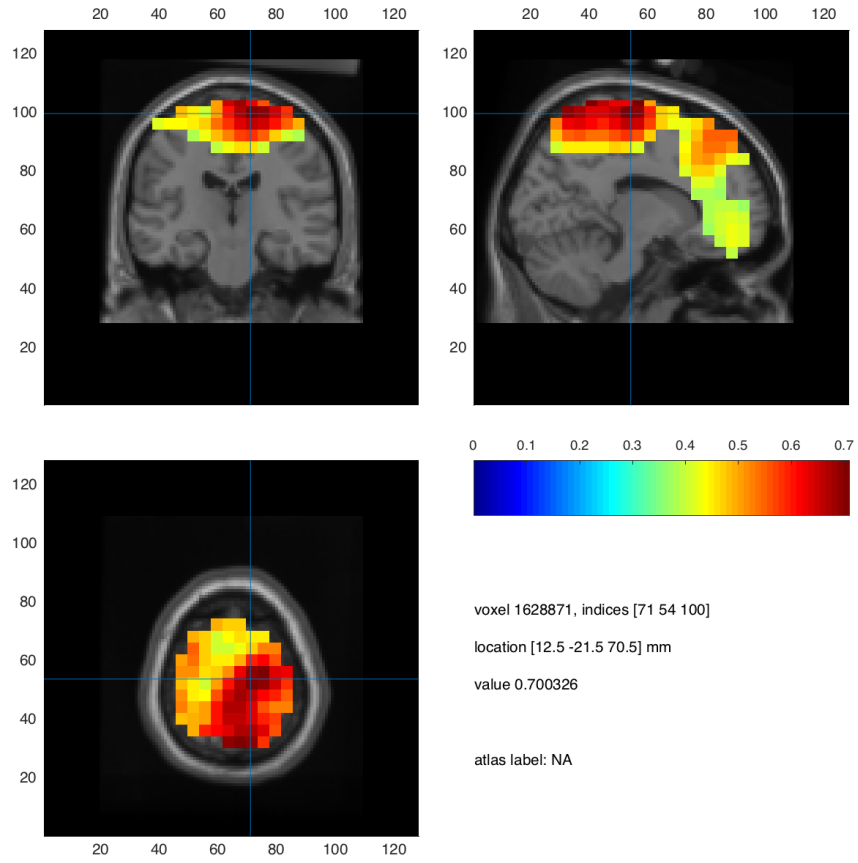

Figure 2.10: The significant ($p < 0.05$) cluster of high correlation, including voxels that were highly correlated with accuracy in the delayed orientation discrimination task ($r > 0.6$).

```
stat.mask(stat.idx) = true;
mri = ft_read_mri('single_subj_T1.nii');
mri = ft_volumereslice([], mri);
cfg = [];
cfg.parameter = 'all';
cfg.interpmethod = 'nearest';
cfg.downsample = 2;
statplot = ft_sourceinterpolate(cfg, stat, mri);

cfg = [];
cfg.method        = 'ortho'; % ortho, surface, slice
cfg.funparameter  = 'rho';
cfg.maskparameter = 'mask';
cfg.funcolorlim = 'zeromax';
ft_sourceplot(cfg, statplot);
```

CHAPTER 3

Discussion and conclusion

The main purpose of this work was to describe approaches to data analysis in electrophysiological research using event-related potentials and demonstrate their usage. We presented what we hope to be a comprehensive account of important phenomenons such as visual short-term memory, as well as the tools that can be used to answer associated questions about their nature.

In the first chapter, we provide rich, yet concise introduction covering the history of memory research and its advancements, as well as relevant time-domain analyses. Considering the emphasis our study programme places on multidisciplinary approaches, we attempted to reflect this aspect throughout the work by mixing knowledge from different fields (cognitive neuroscience, electrophysiology, statistics, psychology). We cover the hierarchy of memory processes leading to VSTM itself, accompanied by some of the recent advancements that are relevant to this work—especially Riečanský et al. (2011) whose results we attempted and mostly succeeded to replicate using different approaches throughout chapter two, which is provides a practical step-by-step demonstration of how one could approach such questions.

Chapter two begins with a description of one of the free toolboxes that are widely used in order to speed up the analysis, and we consider this toolbox-specific approach better than a general description, since it can be

used as a guide for potential newcomers such as the author himself a few months ago. Although the developers already provide several tutorials and experienced users may well use them to infer the rest, we could not find one similar to this (i.e. using similar methodology and being complete from start to finish) and often needed to advance by trial and error. We describe the process of importing data that were already pre-processed using other tools, time-locked averaging, various plotting routines, describe several forms of common statistical design, perform non-parametric permutation correlation testing, and also illustrate and compare appropriate methods for multiple comparisons correction.

The objective of our exploratory analysis was to investigate which processing stage or stages are the strongest predictors of performance measured by response accuracy in a visual short-term memory task. Our analysis revealed results similar to Riečanský et al. (2011) and therefore provided further evidence that the capacity of visual short-term memory relies predominantly on the efficiency of information encoding. We also found that the signal mostly associated with this activity was measured over the medial part of the frontal lobe.

Apart from the analysis at the sensor level, we made an attempt at source-level analysis using spatial filtering, however as we note multiple times throughout section 2.3, we can not report the results with confidence. Nevertheless, the section provides at least a starting point along with references to non-trivial issues that must be considered and that we suspect are sometimes ignored due to how little reliable information we were able to find that would guide us. We were unable to obtain the same result of source reconstruction (pointing to the midcingulate cortex) that was reached by Riečanský et al. (2011) or correspond with other findings (Fuster, 1993), which suggests that ours was likely inaccurate, possibly due to lower signal-to-noise ratio or simply improper methodology.

Other limitations of this work are as follows: Although we do not question the collection and pre-processing of the data, we recognise that working with a semi-finished product introduces some uncontrollable risks, however we consider these irrelevant, which also extends to possible errors in third-party tools that were used to conduct the analysis. We also

have great confidence in the sensor-level analysis. The parameters such as alpha thresholds or the number of permutations were set according to the referenced literature (e.g. Cohen, 2014) and there should be no errors as a result of improper settings. The statistical design was based on the data and the methods were also chosen accordingly. The fact that our results largely support the conclusions mentioned in the previous paragraphs, which were reached by people far more experienced than the author of this work, adds another level of confidence.

This can, however, not be said about the source-level analysis. Regardless of the quality of our methodology, we were working with templates and defaults. For example, the positions of electrodes were imported in a spherical distribution and the adjustment was performed manually rather than using fiducial points (which were not available). The covariance matrix for the purposes of LCMV beamforming where it performs an essential part was chosen without reliable guidelines and might severely impair the results. More work needs to be done in order to fully understand this method and perform a reliable analysis, however this is not necessarily our *fault* since we had no issues with well-documented procedures.

Overall, the practical approaches to ERP analysis that are presented in this work are by all means not exhaustive, yet we are confident that they demonstrate a range of essential methods with comprehensive implementation that are used by researchers worldwide every day, however not always in the correct way. While the author does not consider himself an expert in this field and his work is still lacking confidence, as demonstrated, it appears that even senior researchers tend to make methodological mistakes, which was partly the main motivations for this work. We found this work enriching and hope that the potential readers will consider the output useful.

Bibliography

Acunzo, D. J., MacKenzie, G., & van Rossum, M. C. W. (2012). Systematic biases in early ERP and ERF components as a result of high-pass filtering. *Journal of Neuroscience Methods*, *209*(1), 212–218. doi:10.1016/j.jneumeth.2012.06.011

Akam, T. E. & Kullmann, D. M. (2012). Efficient "communication through coherence" requires oscillations structured to minimize interference between signals. *PLoS computational biology*, *8*(11), e1002760. doi:10.1371/journal.pcbi.1002760

Atkinson, R. C. & Shiffrin, R. M. (1968). Human Memory: A Proposed System and its Control Processes. *Psychology of Learning and Motivation - Advances in Research and Theory*, *2*(100), 89–195. doi:10.1016/S0079-7421(08)60422-3

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual Working Memory a Fixed Number of Represents Items Regardless of Complexity. *18*(7), 622–628.

Axmacher, N., Mormann, F., Fernández, G., Elger, C. E., & Fell, J. (2006). Memory formation by neuronal synchronization. doi:10.1016/j.brainresrev.2006.01.007

Baddeley, A. D., Eysenck, M. W., & Anderson, M. C. (2009). *Memory*. East Sussex: Psychology Press.

Baddeley, A. D. & Hitch, G. (1974). Working memory. doi:10.1016/S0079-7421(08)60452-1

Baddeley, A. D. & Hitch, G. (1977). Recency re-examined. *Attention and performance VI*.

Bartlett, F. C. (1932). Remembering : A Study in Experimental and Social Psychology. *Cambridge, Social Psychology*, 1–11. doi:10.1111/j.2044-8279.1933.tb02913.x

Bates, D. (2006). [R] lmer, p-values and all that. Retrieved 2016-05-25, from https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html

Bjork, R. A. & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology*, *6*(2), 173–189. doi:10.1016/0010-0285(74)90009-7

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*(3), 539–576. doi:10.1037/0033-295X.114.3.539

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, *10*(907173425), 12–21. doi:10.1080/17470215808416249

Burgess, A. P. (2012). Towards a unified understanding of event-related changes in the EEG: the firefly model of synchronization through cross-frequency phase modulation. *PloS one*, *7*(9), e45630. doi:10.1371/journal.pone.0045630

Cohen, M. (2014). Analyzing Neural Time Series Data. *Theory and Practice*, 600. doi:10.1017/CBO9781107415324.004. arXiv: arXiv:1011.1669v3

Conrad, R. (1964). Acoustic Confusions In Immediate Memory. doi:10.1111/j.2044-8295.1964.tb00899.x

Conrad, R. & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology*, *55*(4), 429–432. doi:10.1111/j.2044-8295.1964.tb00928.x

Cowan, N. (2001). Metatheory of storage capacity limits. *Behavioral and Brain Sciences*, *24*(01), 154–176.

Craik, K. J. W. (1943). *The Nature of Explanation*. CUP Archive.

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., & Halgren, E. (2000). Dynamic Statistical Parametric Mapping: Combining fMRI and MEG for High-Resolution Imaging of Cortical Activity. *Neuron*, *26*(1), 55–67. doi:10.1016/S0896-6273(00)81138-1

Dale, H. C. A. (1973). Short term memory for visual information. *British Journal of Psychology*, *64*(1), 1–8.

David, O., Kilner, J. M., & Friston, K. J. (2006). Mechanisms of evoked and induced responses in MEG/EEG. *NeuroImage*, *31*(4), 1580–91. doi:10.1016/j.neuroimage.2006.02.034

de Munck, J. C. & Bijma, F. (2010). How are evoked responses generated? The need for a unified mathematical framework. *Clinical neurophysiology: of-*

ficial journal of the International Federation of Clinical Neurophysiology, *121*(2), 127–9. doi:10.1016/j.clinph.2009.10.002

Della Sala, S., Gray, C., Baddeley, A., Allamano, N., & Wilson, L. (1999). Pattern span: A tool for unwelding visuo-spatial memory. doi:10.1016/S0028-3932(98)00159-6

Drew, T. W., McCollough, a. W., & Vogel, E. K. (2006). Event-Related Potential Measures of Visual Working Memory. *Clinical EEG and Neuroscience*, *37*(4), 286–291. doi:10.1177/155005940603700405

Engle, R. W. & Kane, M. (2004). Executive Attention, Working Memory Capacity, and a Two-Factor Theory of Cognitive Control. *The Psychology of Learning and Motivation: Advances in Research and Theory*, 145–199. doi:10.1016/S0079-7421(03)44005-X

Fell, J., Dietl, T., Grunwald, T., Kurthen, M., Klaver, P., Trautner, P., … Fernández, G. (2004). Neural bases of cognitive ERPs: more than phase reset. *Journal of cognitive neuroscience*, *16*, 1595–1604. doi:10.1162/0898929042568514

Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. doi:10.1016/j.tics.2005.08.011

Fuster, J. M. (1993). Frontal lobes. *Curr Opin Neurobiol*, *3*(MARCH), 160–165. doi:10.1126/science.283.5408.1657

Glanzer, M. (1972). Storage mechanisms in recall. *Psychology of learning and motivation*.

Goldman-Rakic, P. S. (1996). The prefrontal landscape: implications of functional architecture for understanding human mentation and the central executive. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *351*(1346), 1445–53. doi:10.1098/rstb.1996.0129

Hull, C. L. (1943). *Principles of Behavior: An Introduction to Behavior Theory*.

Irwin, D. E. & Andrews, R. V. (1996). Integration and accumulation of information across saccadic eye movements.

Jacobs, J. (1887). Experiments on "prehension". *Mind*.

Jensen, O., van Dijk, H., & Mazaheri, A. (2010). Amplitude asymmetry as a mechanism for the generation of slow evoked responses. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, *121*(7), 1148–9, author reply 1149–50. doi:10.1016/j.clinph.2010.01.037

Kappenman, E. S. & Luck, S. J. (2016). Techniques and Methods Best Practices for Event-Related Potential Research in Clinical Populations. *Biologi-*

cal Psychiatry: Cognitive Neuroscience and Neuroimaging, *1*(2), 110–115. doi:10.1016/j.bpsc.2015.11.007

Keppel, G. & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, *1*, 153–161. doi:10.1016/S0022-5371(62)80023-1

Klauer, K. C. & Zhao, Z. (2004). Double dissociations in visual and spatial short-term memory. *Journal of experimental psychology. General*, *133*(3), 355–381. doi:10.1037/0096-3445.133.3.355

Krieg, J., Trébuchon-Da Fonseca, A., Martínez-Montes, E., Marquis, P., Liégeois-Chauvel, C., & Bénar, C.-G. (2011). A comparison of methods for assessing alpha phase resetting in electrophysiology, with application to intracerebral EEG in visual areas. *NeuroImage*, *55*(1), 67–86. doi:10.1016/j.neuroimage.2010.11.058

Lisman, J. E. & Otmakhova, N. A. (2001). Storage, recall, and novelty detection of sequences by the hippocampus: elaborating on the SOCRATIC model to account for normal and aberrant effects of dopamine. *Hippocampus*, *11*(5), 551–68. doi:10.1002/hipo.1071

Loess, H. (1968). Short-term memory and item similarity. *Journal of Verbal Learning and Verbal Behavior*, *7*(1), 87–92. doi:10.1016/S0022-5371(68)80169-0

Luck, S. J. (2012). Event-Related Potentials. In *Apa handbook of research methods in psychology* (Vol. 1, pp. 1–18).

Luria, R., Sessa, P., Gotler, A., Jolicoeur, P., & Dell'Acqua, R. (2010). Visual short-term memory capacity for simple and complex objects. *Journal of cognitive neuroscience*, *22*(3), 496–512. doi:10.1162/jocn.2009.21214

Makeig, S., Westerfield, M., Jung, T. P., Enghoff, S., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2002). Dynamic brain sources of visual evoked responses. *Science (New York, N.Y.)* *295*(5555), 690–694. doi:10.1126/science.1066168

Mammarella, I. C., Pazzaglia, F., & Cornoldi, C. (2008). Evidence for different components in children's visuospatial working memory. *British Journal of Developmental Psychology*, *26*, 337–355. doi:10.1348/026151007X236061

Maris, E. & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. doi:10.1016/j.jneumeth.2007.03.024

Mazaheri, A. & Jensen, O. (2008). Asymmetric amplitude modulations of brain oscillations generate slow evoked responses. *The Journal of neuroscience :*

the official journal of the Society for Neuroscience, *28*(31), 7781–7. doi:10.1523/JNEUROSCI.1631-08.2008

Miller, G. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review*, *101*(2), 343–352. doi:10.1037/h0043158

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, a. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. *Cognitive psychology*, *41*(1), 49–100. doi:10.1006/cogp.1999.0734

Murdock, B. J. (1960). The distinctiveness of stimuli. *Psychological review*.

Murray, A. M., Nobre, A. C., & Stokes, M. G. (2011). Markers of preparatory attention predict visual short-term memory performance. *Neuropsychologia*, *49*(6), 1458–1465. doi:10.1016/j.neuropsychologia.2011.02.016

Murray, M. M., Brunet, D., & Michel, C. M. (2008). Topographic ERP analyses: A step-by-step tutorial review. doi:10.1007/s10548-008-0054-5

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology*, *53*, 53–81. doi:10.1146/annurev.psych.53.100901.135131

Neath, I. & Surprenant, A. (2003). *Human Memory*.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

Nichols, T. E. & Holmes, A. P. (2001). Nonparametric Permutation Tests for Functional Neuroimaging Experiments: A Primer with examples. *Human Brain Mapping*, *15*(1), 1–25. doi:10.1002/hbm.1058

Niedermeyer, E. & Silva, F. H. L. D. (2004). *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*.

Nikulin, V. V., Linkenkaer-Hansen, K., Nolte, G., Lemm, S., Müller, K. R., Ilmoniemi, R. J., & Curio, G. (2007). A novel mechanism for evoked responses in the human brain. *European Journal of Neuroscience*, *25*(10), 3146–3154. doi:10.1111/j.1460-9568.2007.05553.x

Nikulin, V. V., Linkenkaer-Hansen, K., Nolte, G., & Curio, G. (2010). Non-zero mean and asymmetry of neuronal oscillations have different implications for evoked responses. *Clinical Neurophysiology*, *121*(2), 186–193. doi:10.1016/j.clinph.2009.09.028

Oh, S.-H. & Kim, M.-S. (2004). The role of spatial working memory in visual search efficiency. *Psychonomic Bulletin & Review*, *11*(2), 275–281. doi:10.3758/BF03196570

Oostendorp, T. F. & van Oosterom, A. (1989). Source parameter estimation in inhomogeneous volume conductors of arbitrary shape. *Biomedical Engineering, IEEE Transactions on*, *36*(3), 382–391. doi:10.1109/10.19859

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*. doi:10.1155/2011/156869. arXiv: 156869

Creating a BEM volume conduction model of the head for source-reconstruction of EEG data. (n.d.). Retrieved 2016-05-30, from http://www.fieldtriptoolbox.org/tutorial/headmodel_eeg_bem

Localizing sources using beamformer techniques. (n.d.). Retrieved 2016-05-30, from http://www.fieldtriptoolbox.org/tutorial/beamformer_lcmv

Beamforming evoked fields and potentials in combined MEG/EEG data. (2014). Retrieved 2016-05-29, from http://www.fieldtriptoolbox.org/tutorial/aarhus/beamformingerf

FAQ: How can I import my own data format. (2015). Retrieved 2016-05-29, from http://www.fieldtriptoolbox.org/faq/how_can_i_import_my_own_dataformat

Specifying the channel layout for plotting. (2015). Retrieved 2016-05-29, from http://www.fieldtriptoolbox.org/tutorial/layout

Peterson, L. R. & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of experimental psychology. General*, *58*(3), 193–8. doi:10.1037/h0049234

Pinto, A. D. C. & Baddeley, A. D. (1991). Where did you park your car? Analysis of a naturalistic long-term recency effect. doi:10.1080/09541449108406231

Posner, M. I. & Konick, A. F. (1966). Short-term retention of visual and kinesthetic information. *Organizational Behavior and Human Performance*, *1*(1), 71–86. doi:10.1016/0030-5073(66)90006-7

Postman, L. & Phillips, L. W. (1965). Short-term temporal changes in free recall. *The Quarterly Journal of Experimental Psychology*, *17*(2), 132–138. doi:10.1080/17470216508416422

Ramirez, R. (2008). Source localization. *Scholarpedia*, *3*(11), 1733. doi:10.4249/scholarpedia.1733

Riečanský, I., Tomova, L., Fischmeister, F. P., Bauer, H., & Lamm, C. (2011). Dynamics of human brain activity related to visual short-term memory. In

Activitas nervosa superior rediviva: the official journal of the cians collegium internationale activitatis nervosae superioris (Vol. 53, *3*, p. 146).

Riečanský, I., Tomova, L., Katina, S., Bauer, H., Fischmeister, F. P. S., & Lamm, C. (2013). Visual image retention does not contribute to modulation of event-related potentials by mental rotation. *Brain and Cognition*, *83*(2), 163–170. doi:10.1016/j.bandc.2013.07.011

Rousselet, G. A. (2012). Does filtering preclude us from studying ERP time-courses? doi:10.3389/fpsyg.2012.00131

Ryan, J. (1969). Grouping and short-term memory: Different means and patterns of grouping. *The Quarterly Journal of Experimental Psychology*, *21*(2), 137–147. doi:10.1080/14640746908400206

Singer, W. (1993). Synchronization of cortical activity and its putative role in information processing and learning. *Annual Review of Physiology*, *55*, 349–374. doi:10.1146/annurev.ph.55.030193.002025

Smith, E. E. (1999). Storage and Executive Processes in the Frontal Lobes. *Science*, *283*(5408), 1657–1661. doi:10.1126/science.283.5408.1657

Stephenson, W. & Gibbs, F. (1951). A balanced non-cephalic reference electrode. *Electroencephalography and Clinical Neurophysiology*, *3*(2), 237–240. doi:10.1016/0013-4694(51)90017-X

Tan, L. & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal Of Experimental Psychology-Learning Memory And Cognition*, *26*(6), 1589–1625. doi:10.1037/0278-7393.26.6.1589

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological review*, *55*(4), 189–208. doi:10.1037/h0061626

Van Veen, B., van Drongelen, W., Yuchtman, M., & Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical engineering*, *44*(9), 867–880. doi:10.1109/10.623056. arXiv: 10.623056 [10.1109]

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology Human Perception and Performance*, *27*(1), 92–114. doi:10.1037//0096-1523.27.1.92

Vogel, E. K. & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*(6984), 748–51. doi:10.1038/nature02447

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of experimental psychology. Human perception and performance*, *32*(6), 1436–51. doi:10.1037/0096-1523.32.6.1436

Weisstein, E. W. (n.d.). Bonferroni Correction. Wolfram Research, Inc. Retrieved 2016-05-25, from http://mathworld.wolfram.com/BonferroniCorrection.html

Wickelgren, W. A. (1964). Size of Rehearsal Group and Short-Term Memory. *Journal of experimental psychology*, *68*(4), 413–9. doi:10.1037/h0043584

Woodman, G. F. & Luck, S. J. (2004). Visual search is slowed when visuospatial working memory is occupied. *Psychonomic Bulletin & Review*, *11*(2), 269–274. doi:10.3758/BF03196569

Woolrich, M., Hunt, L., Groves, A., & Barnes, G. (2011). MEG beamforming using Bayesian PCA for adaptive data covariance matrix regularization. *NeuroImage*, *57*(4), 1466–1479. doi:10.1016/j.neuroimage.2011.04.041

Woolston, C. (2015). Psychology journal bans P values. *Nature*, *519*(7541), 9–9. doi:10.1038/519009f

Yamaguchi, Y., Sato, N., Wagatsuma, H., Wu, Z., Molter, C., & Aota, Y. (2007). A unified view of theta-phase coding in the entorhinal-hippocampal system. *Current opinion in neurobiology*, *17*(2), 197–204. doi:10.1016/j.conb.2007.03.007

Yeung, N., Bogacz, R., Holroyd, C. B., Nieuwenhuis, S., & Cohen, J. D. (2007). Theta phase resetting and the error-related negativity. *Psychophysiology*, *44*(1), 39–49. doi:10.1111/j.1469-8986.2006.00482.x