

Towards EEG/fNIRS-based semantic brain-computer interfacing



Milan Rybář

School of Computer Science and Electronic Engineering

University of Essex

A thesis submitted for the degree of

Doctor of Philosophy

September 2022

Acknowledgements

I want to thank my supervisors Ian Daly and Riccardo Poli. I am grateful for their guidance, patience, friendly approach, and support during my PhD journey. This thesis would not have been possible without them.

It was an honor to be a part of the Brain-Computer Interfaces and Neural Engineering Laboratory which was full of amazing, welcoming, and friendly people. I would like to express my gratitude to all people from the lab, just to name a few, Saugat, Sebastian, Ana, Jacobo, Christoph, Tasos, and Rodrigo. I would also like to thank all other PhD candidates in the lab for keeping each other company in the shared office and supporting each another during the PhD journey and the pandemic, just to name a few, Eirini and Lena.

I would like to thank Zuzana; Sebastian, Nadine, and Loki; and Penelope for the walks that kept me sane during the pandemic. I would also like to thank Kateřina and Mark for sharing our PhD struggle and supporting each another.

I would also like to express my gratitude to my family who always supported me during my studies.

Abstract

Semantic neural decoding aims to identify which semantic concepts an individual is focused on at a given moment in time from recordings of their brain activity. This could be used by brain-computer interfaces (BCIs) for communication. These semantic BCIs have the potential to be highly intuitive by allowing direct communication of semantic concepts instead of spelling one character at a time, as is the case with current state-of-the-art BCI systems. This thesis explores the feasibility of semantic BCIs based on electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS).

We designed an experiment to differentiate between the semantic categories of animals and tools during a silent naming task (for the first time in fNIRS), and three novel and intuitive sensory-based imagery tasks using visual, auditory, and tactile perception. Participants were asked to visualize an object in their minds, imagine the sounds made by the object, and imagine the feeling of touching the object. We showed the possibility of semantic neural decoding in both neuroimaging modalities but with contrasting differences in comparison with other state-of-the-art research. Furthermore, we investigated the influence of cue presentation on EEG-based semantic decoding. We found that all EEG-based semantic decoding studies published to date could exploit neural activity recorded during the cue presentation period in their analyses. We showed that including the cue presentation period in the classification pipeline significantly increases classification accuracies. While this research area involves considerable challenges, this thesis made a step towards EEG/fNIRS-based semantic BCIs.

Table of contents

Acknowledgements	iii
Abstract	v
1 Introduction	1
1.1 Motivation	1
1.2 Research Objectives	4
1.3 Thesis Structure	5
2 A systematic literature review	7
2.1 Introduction	7
2.2 Literature review methods	9
2.2.1 Study selection	9
2.2.2 Definitions	11
2.2.2.1 Semantic concepts	11
2.2.2.2 Semantic encoding and decoding	13
2.3 Semantic decoding literature	14
2.3.1 Neuroimaging methods	15
2.3.2 Open datasets	18
2.3.3 Experimental design	19
2.3.3.1 Semantic concepts and categories	19

2.3.3.2	Mental tasks	21
2.3.3.3	Stimulus / cue	25
2.3.4	Feature extraction	26
2.3.5	Feature selection	29
2.3.6	Machine learning models	31
2.3.7	Measuring performance	32
2.3.8	Decoding information transfer rate	34
2.4	Key challenges	36
2.5	Current and future applications and directions	39
2.5.1	Tools for neuroscience	39
2.5.2	Clinical applications	40
2.5.3	Communication aids	42
2.5.4	Other applications and privacy concerns	43
2.6	Discussion	44
3	Experiment	47
3.1	Key choices	47
3.1.1	Neuroimaging methods	47
3.1.2	Mental tasks	48
3.1.3	Semantic categories	51
3.1.4	Stimuli	51
3.2	Mental tasks	52
3.3	Stimuli	54
3.4	Experimental design	56
3.5	Participants	57
3.6	Neuroimaging data	58
3.7	Order of tasks	60

4	Semantic decoding in EEG	63
4.1	Methods	63
4.1.1	EEG data preprocessing	63
4.1.2	Temporal analysis	65
4.1.3	Decomposition by PCA	66
4.1.4	Decomposition by CSP	67
4.1.4.1	An issue of decreased dimensionality of EEG data	67
4.1.4.2	Nested-cross-validation approach	68
4.1.4.3	Spatio-spectral CSP	68
4.1.4.4	Grid search	69
4.1.5	Time-frequency representation	69
4.1.6	Event-related (de)synchronization	70
4.2	Results	71
4.2.1	Temporal analysis	72
4.2.2	Decomposition by PCA	72
4.2.3	Decomposition by CSP	75
4.2.3.1	Nested-cross-validation approach	75
4.2.3.2	Spatio-spectral CSP	78
4.2.3.3	Grid search	78
4.2.4	Time-frequency representation	80
4.2.5	Event-related (de)synchronization	80
4.3	Discussion	87
4.3.1	Time variance	88
4.3.2	Image presentation confound	89
5	The effect of cue presentation on EEG-based semantic decoding	91
5.1	Introduction	91

5.2	Methods	94
5.2.1	Analysis in the temporal domain	94
5.2.2	Analysis 1 (single channel classification)	95
5.2.3	Analysis 2 (all channels classification)	95
5.2.4	Analysis 3 (selecting an optimal temporal window)	96
5.2.5	Analysis 4 (full period classification)	96
5.2.6	Analyses 5 and 6 (replication studies)	97
5.3	Results	98
5.3.1	Analysis 1 (single channel classification)	98
5.3.2	Analysis 2 (all channels classification)	100
5.3.3	Analysis 3 (selecting an optimal temporal window)	101
5.3.4	Analysis 4 (full period classification)	101
5.3.5	Analyses 5 and 6 (replication studies)	103
5.4	Discussion	105
6	Semantic decoding in fNIRS	109
6.1	Methods	109
6.1.1	fNIRS data preprocessing	109
6.1.1.1	Bad channels removal	109
6.1.1.2	Motion correction	110
6.1.1.3	Changes in oxy/deoxy-generated hemoglobin	110
6.1.1.4	Filtering	111
6.1.2	Single concept trial analysis	111
6.1.3	GLM approach	112
6.1.3.1	Design matrix	113
6.1.3.2	Correction for temporal correlations	115
6.1.3.3	Cross-validation	116

6.1.3.4	Extraction of evoked hemodynamic responses	116
6.1.4	Univariate classification	117
6.1.5	Multivariate classification with PCA features	118
6.1.5.1	Feature selection	118
6.2	Results	119
6.2.1	Univariate classification	119
6.2.2	Multivariate classification with PCA features	122
6.2.2.1	Feature selection	122
6.3	Discussion	126
7	Discussion	131
7.1	Mental tasks	133
7.2	Cue presentation confound	135
7.3	Time-variant approaches	136
7.4	Experiment shortcomings	137
7.5	Future research	139
8	Conclusions and future research	141
8.1	Contributions	141
8.2	Limitations and future work	143
	References	145
	Appendix A Functional near-infrared spectroscopy	167
A.1	Overview	167
A.2	Modified Beer-Lambert law	169
	Appendix B Potential pitfalls of widely used implementations of common spatial patterns	173

B.1	Introduction	174
B.2	Common spatial patterns	174
B.2.1	Geometric approach	176
B.2.2	Generalized eigenvalue problem	176
B.2.3	Covariance matrices without full rank	177
B.3	Methods	178
B.3.1	Evaluation	178
B.3.2	Implementations	179
B.3.3	EEG Toolboxes	180
B.4	Results	181
B.4.1	Geometric approach	181
B.4.2	Generalized eigenvalue problem	183
B.5	Conclusion	184
Appendix C Additional figures for Chapter 4		185
Appendix D Additional figures for Chapter 5		205

Chapter 1

Introduction

This chapter introduces the motivation, research objectives, and structure of this thesis.

1.1 Motivation

Semantic concepts are mental representations within our minds. They are a crucial part of our knowledge and understanding of the world and our thought processes [174].

Recent results in cognitive neuroscience, machine learning, and human language science have shown the possibility of semantic neural decoding. *Semantic neural decoding* aims to identify which specific semantic concepts an individual is focusing on, or thinking of, at a given moment in time from recordings of their brain activity [1, 221].

Work by several groups has shown that several pairs of semantic categories can be differentiated, for instance: animals and tools [240, 239, 120, 244, 113, 23, 45, 11, 161, 260], tools and buildings [244, 113, 23, 45, 11, 161], and animals and body parts [286, 244, 113, 24, 23, 45, 11, 161]. Recent work has also shown that it is possible to differentiate more than two categories at a time [256, 258, 23, 45, 52, 11].

This semantic neural decoding could be used by brain-computer interfaces (BCIs) for communication [84, 182]. BCIs provide an alternative pathway between a human brain and

external devices. BCIs are studied for a variety of applications ranging from helping people with disabilities [133] to gaming and the entertainment industry for healthy people [130].

BCIs have been successfully used as a technological solution to aid communication for people who experience difficulties communicating via other means [133, 94, 121]. For example, this is required in a severe clinical condition called *locked-in syndrome* (LIS) [110]. LIS is paralysis of all four limbs and the last cranial nerves without interfering with consciousness. In other words, patients with LIS are conscious, but trapped within their bodies, unable to move or communicate. LIS can result from a variety of clinical etiologies, such as amyotrophic lateral sclerosis (ALS) or spinal cord injury. Individuals with ALS experience progressive loss of motor function. LIS is divided into three varieties: classical, total, and incomplete [20]. Patients with classical LIS have intact eye movements or blinking. However, patients with total LIS have absence of all voluntary movements, including eye movements. Lastly, patients with incomplete LIS retain remnants of voluntary movement, such as a finger twitch. Apart from LIS, other neurological or head and neck injuries can result in anarthria which is the inability to speak, despite intact language comprehension and cognition. This differs from aphasia in which the formulation or understanding of language is affected but some speech is possible.

Unfortunately, current BCI communication speeds and accuracies are relatively low in comparison with other communication platforms [4, 245, 190]. The performance of BCIs is often measured by *information transfer rate* that determines the amount of information that is conveyed by a system's output within a set time [151, 277, 235]. Information transfer rate depends on the number of choices available to the user of the system, the accuracy of target detection by the system, and the average selection time taken by the system. The majority of current BCIs achieve average information transfer rates of 27 bits per minute [26, 283], while other communication aids can achieve considerably higher information transfer rates. For example, eye trackers can achieve average transfer rates of 41 bits per minute [265]. This makes BCIs for communication only really useful when other interfaces are not feasible [190].

According to surveys of potential BCI users [103, 102, 51, 95], a “good” BCI-based speller system for communication should have an accuracy of at least 90%, a speed of at least 15-20 letters per minute (corresponding to an information transfer rate of at least 65 bits per minute), and accidental exits from a standby mode no more than once every 2-4 hours.

Semantic BCIs based on semantic neural decoding could allow direct communication of semantic concepts. Semantic BCIs could have the potential to be highly intuitive but it is currently unknown if they might allow levels of accuracy and communication speed that meet the requirements of BCI users.

The main difference between current state-of-the-art BCI systems and our proposed semantic BCI system stems from the serial communication paradigm used in almost all current BCI systems, in which communication proceeds 1 character at a time, with the meaning of the concept transmitted becoming available only when a full word has been received. Contrary to this, semantic BCIs could achieve a form of parallel communication by directly identifying high-information semantic concepts. For example, instead of spelling out ‘I–A–M–H–U–N–G–R–Y’ one letter at a time, semantic BCI would allow directly identifying the semantic concept of ‘hunger’ allowing a faster and more natural communication.

While the most promising results to date have been reported using functional magnetic resonance imaging (fMRI) [160, 161, 238, 239] which is a whole-brain neuroimaging modality, semantic decoding using *neural signals recorded from the scalp*, such as electroencephalography (EEG) or functional near-infrared spectroscopy (fNIRS), is of particular interest for potential semantic BCIs. fMRI is a non-invasive neuroimaging technique which measures brain activity by detecting changes associated with cerebral blood flow. Its main drawbacks include application cost, non-portability, low temporal resolution, and the scanner environment which greatly restricts the types of cognitive tasks and abilities that can be investigated. EEG is a non-invasive technique which measures electrical activity from the surface of the scalp. The signal is quite noisy and attenuated because currents need to go through several layers of

non-neural tissue to reach electrodes. Thus, EEG has a lower spatial resolution of about 2 cm in comparison with other neuroimaging techniques but it offers a millisecond-range temporal resolution. Also, fNIRS is a non-invasive technique but it measures cortical brain activity up to about 2 cm deep through hemodynamic responses associated with neuron behavior similarly to fMRI. On the other hand, fNIRS is portable and cheaper than fMRI. This gives us a hope to substitute fMRI with fNIRS [226].

The joint recordings of EEG and fNIRS seem like an ideal combination for semantic BCIs. Both EEG and fNIRS are portable, relatively cheap, and provide better ecological validity in comparison with fMRI. EEG provides a good temporal resolution while its poor spatial resolution could be potentially improved by fNIRS, as these two techniques have complementary strengths.

1.2 Research Objectives

This thesis aims to explore the feasibility of semantic BCIs based on EEG and fNIRS. The main research objective is to identify or propose several mental tasks that would be suitable for semantic BCI paradigms. This will allow us to address the following research questions:

1. Does including neural activity recorded during a cue presentation period into decoding pipelines affect EEG-based semantic neural decoding in state-of-the-art?
2. Is it possible to differentiate between the semantic categories of animals and tools during selected mental tasks from EEG data?
3. Is it possible to differentiate between the semantic categories of animals and tools during a sequence of mental tasks from fNIRS data?

1.3 Thesis Structure

Chapter 2 presents a systematic literature review of semantic neural decoding from a range of neuroimaging modalities. It discusses specific neuroimaging methods, experimental designs, and machine learning pipelines that are employed to aid the decoding of semantic concepts. The chapter also discusses current challenges presented by this research area and discusses some possible emerging and speculative future directions for this research area.

Chapter 3 introduces an experiment used throughout this thesis to explore the feasibility of differentiation between the semantic categories of animals and tools in a silent naming task and three novel and intuitive sensory-based imagery tasks using visual, auditory, and tactile perception.

Chapter 4 explores the possibility of semantic neural decoding from the recorded EEG data.

Chapter 5 investigates the influence of cue presentation on semantic neural decoding in EEG.

Chapter 6 explores the possibility of semantic neural decoding from the recorded fNIRS data.

Chapter 8 provides some conclusions and makes suggestions for future research on semantic neural decoding.

Chapter 2

A systematic literature review

This chapter presents a systematic literature review of semantic neural decoding from neural activity recorded by any neuroimaging modality. This systematic literature review has been published in [\[221\]](#).

2.1 Introduction

Our experience of the world has long been regarded by some philosophers as an internal subjective experience that is individual to us [\[259, 33\]](#). We may have tasted the same apples, smelt the same roses, and heard the same bird song as our neighbors, but our individual mental states have long been thought to have a very distinct and subjective nature [\[259\]](#).

Many philosophers refer to this individual introspective experience as our ‘qualia’, our own introspectively accessible experience of the world [\[259, 33\]](#). It has long been considered, by some, to be impossible to know, with absolute certainty, how anyone else experiences the world.

While this may remain true, modern neuroscience is increasingly beginning to reveal how our brains respond to specific experiences within the world. We now know what specific patterns of activity occur in the brain as we eat an apple, or smell a rose, and, broadly speaking,

for many people the parts of the brain that become active during these experiences are similar [145, 262, 33].

Indeed a significant portion of modern neuroscience is focused on exactly how our conscious mental states as we experience the world (our ‘qualia’) relate to the activity in our brains [166]. This work has rapidly accelerated in recent years with the development of modern, non-invasive, neuroimaging tools that are capable of observing activity in our brains in real-time [91].

Techniques such as fMRI (developed in the 1990’s [184, 25]) and EEG (developed between the 1870’s to 1890’s [22], but much more recently coupled with powerful computer-driven statistical analysis techniques) have been combined with studies of neurological aetiologies to revolutionize our understanding of how semantic concepts are encoded in the brain. This new understanding has given rise to a further field of study, semantic decoding, defined as the decoding of semantic concepts from recordings of our brain activity.

Semantic decoding refers to a combination of hardware and software systems that may be employed to identify the specific semantic concept(s) an individual is focused on, or thinking of, from a recording of their brain activity [178]. It is a technique which opens the doors to a wide range of exciting possibilities and future applications.

This chapter reviews the current state of the art in semantic decoding methods and discusses current neuroimaging methods, experimental designs used in semantic decoding and how they may be combined with machine learning pipelines to reveal which specific semantic concepts an individual is focused on. It also discusses the current challenges in this research area, including how to effectively combine multi-modal neuroimaging techniques to more accurately decode semantic concepts and how to develop effective machine learning methods to deal with the typically large, non-stationary, noisy, multi-dimensional datasets involved in this work. Finally, this chapter discusses some current and future applications of this research area.

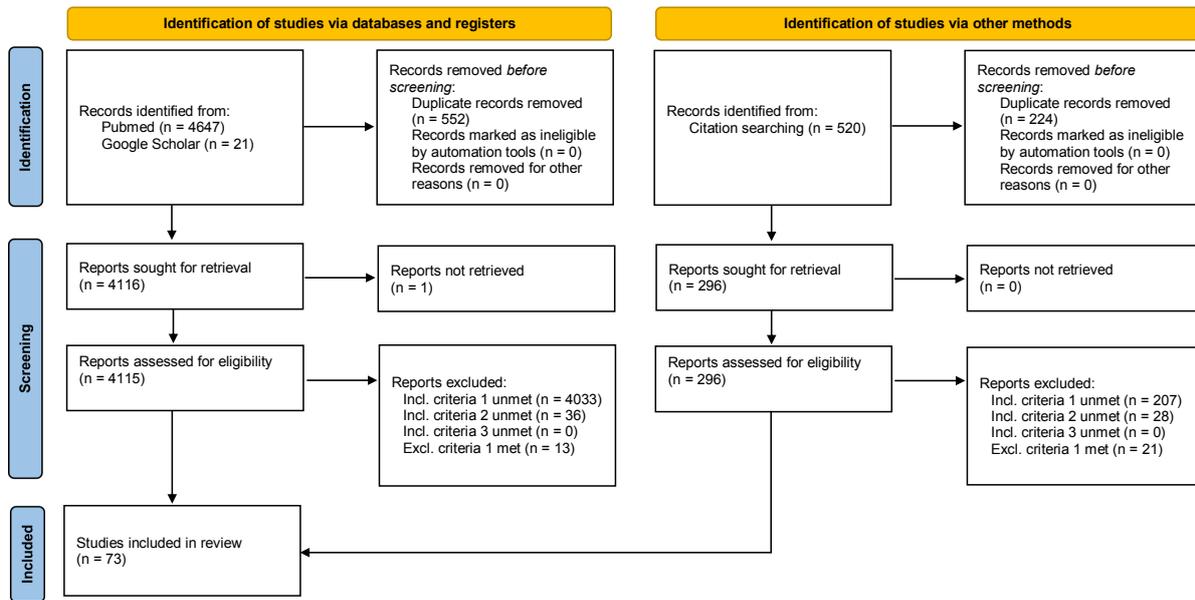


Fig. 2.1 Study selection flowchart.

2.2 Literature review methods

2.2.1 Study selection

To review the topic of semantic decoding, the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) guidelines [164] were followed. Figure 2.1 illustrates the process of study selection and the resulting number of identified studies. PubMed and Google Scholar databases were systematically searched to identify papers which report attempts at neural semantic decoding. The search was run in January 2022 and the search queries used for each database are listed in Table 2.1.¹

Duplicate results that arose from the four search queries were removed. The records were then screened for their relevance. Specifically, papers were included which described attempts to build and evaluate models that are able to decode the individual discrete semantic concepts an individual participant was focused on at a discrete moment of time from recordings of their

¹This search and the screening procedure were primarily performed by Dr Ian Daly for [221], which describes this systematic literature review of semantic neural decoding.

Table 2.1 Search queries used to identify articles relating to neural semantic decoding.

Database	Query
PubMed	((semantic AND decoding) OR (semantic AND prediction) OR (concept AND prediction) OR (concept AND decoding) OR (noun AND prediction) OR (noun AND decoding)) AND ("brain activity" OR neural)
PubMed	((semantic AND decoding) OR (semantic AND prediction) OR (concept AND prediction) OR (concept AND decoding) OR (noun AND prediction) OR (noun AND decoding)) AND (EEG OR electroencephalography OR electroencephalogram OR fMRI OR "functional magnetic resonance imaging" OR MEG OR "magnetoencephalogram" OR "magnetoencephalography" OR fNIRS OR "functional near infrared spectroscopy" OR ECoG OR "electrocortigraphy")
PubMed	((semantic AND decoding) OR (semantic AND prediction) OR (concept AND prediction) OR (concept AND decoding) OR (noun AND prediction) OR (noun AND decoding)) AND ('intracranial EEG' OR iEEG OR 'stereotactic EEG' OR sEEG OR 'invasive EEG' OR 'depth electrodes' OR 'implanted electrodes' OR 'human single-unit' OR 'human single neuron' OR 'concept cells')
Google Scholar	allintitle: (semantic AND decoding AND "brain activity") OR (semantic AND decoding AND neural) OR (semantic AND decoding) OR (semantic AND prediction AND "brain activity") OR (semantic AND prediction AND neural) OR (semantic AND prediction) OR (concept AND prediction AND "brain activity") OR (concept AND prediction AND neural) OR (concept AND prediction) OR (concept AND decoding AND "brain activity") OR (concept AND decoding AND neural) OR (concept AND decoding) OR (noun AND prediction AND "brain activity") OR (noun AND prediction AND neural) OR (noun AND prediction) OR (noun AND decoding AND "brain activity") OR (noun AND decoding AND neural) OR (noun AND decoding)

Table 2.2 Screening criteria for records returned by search queries.

Criteria	
Include	Report describes an attempt to develop and evaluate, on humans, a model capable of neural semantic decoding.
Include	Clear description of methods and results in terms of decoding accuracy / efficacy.
Include	Report published in a peer reviewed article (journal, conference, or peer-reviewed book chapter).
Exclude	Review, position, theory, and discussion articles.

neural activity. Consequently, the records were screened according to the criteria set out in Table 2.2.

To further identify additional articles not found by the initial search queries, each of the articles short-listed by applying the screening criteria in Table 2.2 were then inspected. Specifically, the reference list from each article was also screened according to the criteria in Table 2.2. This produced a final list of 73 articles which describe attempts to develop and evaluate neural semantic decoders.

2.2.2 Definitions

2.2.2.1 Semantic concepts

At the most basic level a concept is the idea of what something is or how it works and may be held in the mind or expressed in language. Semantics refers to the study of meaning. Thus, a semantic concept may be defined as the meaning of what something is or how it works. This may be distinguished from a perceptual concept, which defines how a concept is perceived (e.g., how it looks or sounds).

Within the field of neuroscience it has been known for some time that different neural systems exist for semantic processing of concepts and perceptual processing of those same concepts [271]. Specifically, early work by Elizabeth Warrington described how patients could

match perceptual features of objects without being able to match descriptions of the objects with their names.

More recently, the specific neuro-anatomical basis of these systems have been identified in detail by a series of neuroimaging studies as well as studies of individuals with neurological aetiologies that effect their ability to access semantic memory [191]. Specifically, semantic memory (the process of retrieving semantic information related to a concept) involves a distributed-plus-hub network in which a distributed network of brain regions selectively respond to modality specific features, while a central semantic hub acts to represent semantic similarity between concepts. There is considerable evidence locating this hub within the left hemisphere anterior temporal lobe [191].

A widely supported theory describing how semantic concepts are encoded in the brain is embodiment theory. This states that the meaning of a concept is situated within our experience of the world [154]. So for example, the concept of a tool is situated within our understanding of how tools are used (they are held in the hands, they are used to manipulate other objects, etc.). This may be contrasted with other approaches, which state that the meaning of a concept is grounded in abstract symbols or in a universal organizational system [174].

Neuroimaging support for both embodiment theory and the distributed-plus-hub model comes from fMRI studies, which report significant changes in blood flow within both brain regions responsible for percepto-motor circuits during processing of words related to perception of motion and the anterior temporal lobe. For example, processing of words related to tools has been shown to activate the sensori-motor cortex [209].

A review of embodiment theories by Meteyard and others [154] divides the current theories into four groups placed on a continuum from unembodied to strongly embodied. (1) Unembodied theories represent fully symbolic systems in which semantic information is truly symbolic. There is no role for sensory and motor information in semantic representation and thus it is completely independent from sensory-motor systems. Any interactions between semantic content

and sensory-motor systems are explained by an indirect route. (2) In secondary embodiment theories, semantic representations are amodal with a non-arbitrary relationship between semantic information and sensory-motor content. These theories posit a region for amodal semantic content plus modality specific regions which code experiential attributes. The above mentioned distributed-plus-hub model belongs to this group. (3) Weak embodiment theories state that semantic representations are at least partly constituted by sensory-motor information. There are distributed networks of areas which code integral modal information which is proximal to primary sensory and motor regions. (4) In strong embodiment theories, low level sensory and motor information is activated in primary cortical areas as part of semantic processing. They conclude that the fully symbolic, unembodied theories and strong embodiment theories are not supported. An additional important finding is the 'anterior shift' [154] in which the areas activated during the semantic processing are shifted anterior to those areas used in direct experience.

For the purposes of this chapter, we define a semantic concept as an idea of what something is or how it works that is independent of the perceptual features of the concept such as how it looks or how it sounds.

2.2.2.2 Semantic encoding and decoding

Semantic encoding may be broadly described as the study of how the brain encodes specific concepts. This includes studying how specific brain regions are involved in the encoding of concepts, as well as exploring how networks of brain regions work together to encode specific semantic concepts [257].

In general, semantic encoding and decoding may be realized by constructing *encoding and decoding models* [128, 131]. Semantic *encoding* models are a group of modeling techniques that seek to predict brain activity from stimuli, while semantic *decoding* models seek to predict the stimuli from neural activity [178].

Both types of model involve the development of signal processing and machine learning pipelines to relate distinct semantic categories to recordings of neural activity. Consequently, these models are frequently confused with one another in the literature [178]. Indeed, encoding and decoding models are often closely related to one another. Although, an encoding model is not a necessary prerequisite of a decoding model, it has two advantages over a decoding model. Firstly, it can in principle provide a complete description of the related encoding process, while a decoding model can provide only a partial description. Secondly, it can be transformed into an optimal decoding model, a process which is more difficult the other way around [178].

Encoding and decoding models are applicable to a wide range of questions in neuroscience. For example, decoding models have been developed to decode scenes from a TV show viewed by individuals [264], faces seen by individuals [139], pieces of music heard by participants [96], and the quantity of displayed objects [40].

In this chapter, we focus on decoding models applied to the problem of semantic decoding, identifying the single coherent semantic concept an individual is focused on at a given discrete epoch of time from recordings of their neural activity.

2.3 Semantic decoding literature

A wide range of different neuroimaging tools and methods have been employed by researchers seeking to decode semantic concepts from the brain.

Semantic decoding models seek to identify the discrete semantic concepts an individual is focused on at a given moment in time. Consequently, neural semantic decoding studies start with an experiment that is designed to cue participants to focus their attention on single semantic concepts for discrete periods of time. Neural activity is recorded while participants are cued to pay attention to a single concept. This recorded neural activity is then processed to remove signal noise and increase the signal to noise ratio of key discriminative features.

Table 2.3 Modalities used for developing neural semantic decoding models.

Modality	No.	References
EEG	5	[240, 173, 5, 54, 24]
EEG + MEG	3	[172, 44, 246]
EEG + ECoG + SEEG	1	[168]
fMRI	45	[2, 3, 239, 238, 19, 120, 198, 148, 18, 113, 87, 64, 278, 237, 55, 268, 242, 69, 82, 66, 10, 134, 72, 39, 144, 17, 256, 258, 116, 236, 42, 73, 23, 7, 45, 213, 11, 140, 8, 161, 269, 206, 89, 56, 181]
fMRI + MEG	1	[35]
fNIRS	1	[286]
MEG	6	[244, 208, 99, 241, 52, 78]
Macro electrodes (ECoG, SEEG)	9	[272, 141, 107, 175, 233, 159, 260, 255, 223]
Micro electrodes	2	[129, 212]

Semantic decoding models are then trained on these features and evaluated in terms of their decoding accuracy.

2.3.1 Neuroimaging methods

Table 2.3 lists the modalities used in neural semantic decoding studies.

The majority of decoding studies to date have used fMRI. This is due, in large part, to the superior spatial resolution provided by fMRI, which allows whole brain neuroimaging. However, the fMRI does have a number of disadvantages when it comes to studying brain activity related to semantic meaning. Specifically, as already mentioned fMRI has a particularly poor temporal resolution and is only able to detect and monitor changes in oxygenated blood flow (BOLD) that follow electrophysiological neural activity by 2–4 seconds [83]. Additionally, fMRI is extremely expensive, cumbersome, and requires participants to lie in a confined space in tightly controlled conditions for extended periods of time. Consequently, fMRI studies typically focus on small numbers of participants and are often only able to answer relatively straightforward questions [1].

In contrast, electrophysiological neuroimaging methods, such as EEG or magnetoencephalography (MEG), provide a direct recording of neural activity in mainly cortical neurons with a very high temporal resolution. This provides the potential to explore how semantic encoding patterns change over time [208] at the cost of a considerably poorer spatial resolution.

EEG has been explored as a tool for semantic decoding by a relatively small number of authors and has been demonstrated, in some circumstances, to be able to reveal activity related to processing of a subset of semantic concepts. For example in work by Murphy and colleagues [173] differences in EEG correlates of the concepts for ‘tools’ and ‘mammals’ were reported to allow a mean decoding accuracy of 72 %. Additionally, work by Simanova and colleagues [240] reported semantic decoding for the concepts of ‘animals’ and ‘tools’ with a mean accuracy of up to 79 %.

Two alternative neuroimaging techniques that provide direct recordings of electrophysiological neural activity with the same high temporal resolution as the EEG, while also affording a high spatial resolution and specificity, are macro intracranial electrodes (such as electrocorticogram (ECoG) and stereoelectroencephalography (SEEG)) and micro intracranial electrodes. Macro intracranial electrodes record neural activity from large groups of neurons via a grid of electrodes. This grid is either placed directly on the cortical surface under the skull, in the case of ECoG [158], or can be placed at a wide range of locations in the brain, in the case of SEEG [41]. On the other hand, micro intracranial electrodes allow activity to be recorded from individual neurons at any position in the brain. Consequently, both techniques provide signals with high spatial and temporal resolution that have high signal to noise ratios. However, this comes at the cost of coverage (ECoG and SEEG only cover a limited region of the brain and micro electrodes only allow recordings from a few dozen individual neurons) and with the added risk from the brain surgery that is necessary to implant the electrodes. A set of studies have demonstrated that recordings of ECoG signals, SEEG signals, and micro electrodes may be used for semantic decoding [272, 141, 107, 175, 260, 255, 129, 223, 212].

A recent work has also demonstrated that it is possible to differentiate semantic concepts from fNIRS [286]. fNIRS records levels of oxygenated and de-oxygenated hemoglobin in the cortex by shining an infra-red light through the skull and measuring how the reflected and refracted light changes with blood flow. It measures the same physiological process as fMRI, while allowing participants to sit or move more freely, which enables a wider range of experiment designs at the cost of lower spatial resolution and coverage.

Techniques that record electrophysiological brain activity, such as EEG, provide a direct measure of neural activity as it happens with very high time resolution, whereas blood flow based neuroimaging methods, such as fNIRS, are only able to provide indirect measures of neural activity via changes in the concentration of hemoglobin, a time delayed and spatially imprecise response to electrophysiological neural activity [1]. An additional consideration is that non-invasive technology, such as EEG and fNIRS, are relatively cheap and portable, potentially allowing their use in experiments that better capture everyday use of semantic concepts.

However, the considerably poorer spatial resolution of technologies such as EEG and fNIRS presents a significant challenge when compared to technologies that provide a higher resolution recording of brain activity such as fMRI, and this is reflected in the corresponding number of semantic decoding publications that make use of each technique. This is because different semantic concepts can be spatially encoded throughout the brain, including in sub-cortical regions [161] which can be observed by fMRI but, conventionally, are harder to measure with scalp based measurement technologies [119].

Indeed, work by Murphy and Poesio [172] suggests that the ability to identify semantic concepts from the EEG is closely related to the ease with which the associated neural activity may be identified from electrophysiological recordings of cortical brain activity (EEG and MEG). For example, the concepts of ‘tools’ and ‘mammals’ are differentiable from EEG data alone [173] and fMRI neuroimaging work by Pulvermüller and colleagues [209] has shown

these two concepts involve activations in the sensorimotor and parietal cortices, which are cortical regions observable via EEG. Conversely, other semantic concepts that are, perhaps, more complex in nature (e.g., such as specific foods or ‘hunger’) have been shown to involve sub-cortical brain areas, making them potentially considerably harder to identify via current non-invasive neuroimaging techniques [161].

2.3.2 Open datasets

A small proportion of the neuroimaging datasets that have been recorded during studies developing and evaluating neural semantic decoders have been made publicly available, allowing other research groups to re-use such datasets to develop and evaluate new methods. Table 2.4 lists publicly available datasets for developing and evaluating neural semantic decoders. It also shows which study originally recorded the dataset and other studies which have made use of the same dataset.

Note, a number of studies (such as [66, 8, 233, 255]) make use of datasets recorded in other studies but not made publicly available. This is typically because the studies were conducted within the same lab and re-used data that was available in the lab but not publicly available. A small number of studies make use of data that is described as being available on request, either to eligible researchers [120], or to all [278, 208]. However, these datasets are not published. Finally, one study by Carlson and colleagues [42] used data that is described in the study as publicly available. However, on further investigation the data was found to no longer be publicly available because the archive site was taken down due to lack of funding. These datasets are not included in Table 2.4.

Table 2.4 Publicly available datasets for developing and evaluating semantic decoding models.

Modality	Available at	Reported	Re-used in
ECoG	http://klab.tch.harvard.edu/resources/liuetal_timing3.html#sthash.BiYFH24Z.dpbs	[141]	[107]
ECoG	https://purl.stanford.edu/xd109qh3109	[159]	
Micro electrodes	https://github.com/rebrowski/abstractRepresentationsInMTL	[212]	
EEG	https://www.cs.cmu.edu/~tom/science2008/	[161]	[11]
fMRI	https://openneuro.org/datasets/ds000105/versions/00001	[89]	[87]
fMRI	https://datadryad.org/stash/dataset/doi:10.5061/dryad.vmcvdncpf	[242]	
fMRI	https://www.cs.cmu.edu/~tom/science2008/	[161]	[23, 11]

2.3.3 Experimental design

The experimental design is probably the most important set of decisions to make when attempting semantic decoding. Here, we review three crucial elements of experimental design: (1) semantic concepts and categories, (2) mental tasks, and (3) stimulus or cue presentation.

2.3.3.1 Semantic concepts and categories

The semantic categories that neural semantic decoders have been trained to differentiate vary from study to study. However, there are some groups of semantic categories that are frequently used to train and evaluate neural semantic decoders. Figure 2.2 illustrates which pairs of semantic categories neural decoders have been developed to differentiate between. Specifically, the figure shows a network on semantic categories where each node represents an individual category and each edge represents an attempt to build a decoder to differentiate those categories. The size of the nodes is proportional to the number of studies that report attempts to build decoders that recognize that category, while the widths of the edges between pairs of nodes are proportional to the number of studies that report attempts to differentiate those pairs of categories. Note that the positions of the concepts in the network diagram are arbitrary.

The most frequently differentiated semantic categories include animals and tools [240, 239, 120, 244, 113, 23, 45, 11, 161, 260], tools and buildings [244, 113, 23, 45, 11, 161], and animals and body parts [286, 244, 113, 24, 23, 45, 11, 161]. Several studies have also shown that it is possible to differentiate more than two semantic categories at a time [256, 258, 23, 45, 52, 11].

There is a relatively dense network of semantic categories that are frequently decoded, including tools, buildings, body parts, and animals. However, it is important to note that this may not necessarily indicate that these specific concepts are easier to decode than other concepts, as many authors simply opt to replicate and extend the work of other authors, particularly when selecting which categories to attempt to decode.

A set of studies focused on differentiating individual concepts within a single category [238, 161, 244, 148, 286, 120], for instance, physics concepts [148], sets of 180 words [198], and sets of 240 sentences [269]. These studies are not included in Figure 2.2 to avoid over-complicating it.

The selection of semantic categories and concepts is occasionally not clearly justified and only a few studies have focused on this problem in detail. For instance, Bauer and colleagues [19] used concepts based on a previous behavioral study that collected pairwise dissimilarity ratings. However, a small number of studies employed a data-driven strategy to generate the concepts or the semantic categories. For instance, Pereira and colleagues [198] partitioned a semantic vector space, which was used to encode individual concepts (see Section 2.3.4), by a clustering method and a core concept was selected from each region.

2.3.3.2 Mental tasks

A wide variety of different mental tasks have been used in semantic decoding studies to date. These are listed in Table 2.5. They all share the common goal of encouraging participants to hold a target concept in their minds, while at the same time many aim to also test participants focus during the experiment.

Table 2.5 Mental tasks used by the semantic decoding studies. Note, some studies employ two or more task types and, therefore, appear in two or more rows.

Task type	Specific task	References
Naming	Silent naming task	[172, 173, 242, 246, 56]
	Aloud naming task	[272, 246, 256, 175, 99]
Properties	Silent properties generation (think about a consistent pre-generated set of properties)	[238, 161, 3, 2, 148, 19, 11, 18, 113, 237, 258, 39, 23, 45]
	Think about sensory and motor properties of the concept	[272, 242]
	Think about taught features	[17]
Meaning	Think about characteristics of the concept	[236]
	Think freely about meaning of stimulus or evoked memories	[286]
	Contextual meaning reflection (think about the meaning of the concepts in the given context)	[198, 233, 255]
	Think about associated situation with the concept	[10, 8]
	Contextual meaning reflection (think about overall meaning of a sentence/phrase)	[269, 7, 278, 134, 116, 206]
	Read story then answer comprehension question	[66]
	Generate detailed mental images as similar as possible to pre-seen images	[213]
	Category / property recognition	
Out-of-category recognition	[240, 239, 54]	
In-category recognition	[55, 24]	
Yes/no questions	[244]	
Size judgment	[44]	
Orientation judgment	[159]	
Category specific judgment	[168, 129, 212]	
Answer whether it can be directly experienced with senses	[73, 72]	
Concept similarity judgment (scale 1–4)	[144]	
Semantic similarity of 2 words to a key word	[268]	
Semantic congruity judgment	[140]	
Name the color of the object or the background	[99]	
Silently name a word from a cued category with a cued initial letter	[241]	
Object recognition	Oddball task	[35, 208, 78, 236, 260]
	1-back task	[208, 141, 107, 52]
	1-back match task	[82, 89, 223]
	Delayed matching	[42, 181]
	Remember all six elements presented in a sequence	[5]
	Name an object that was closest to the one shown in the picture	[272]
Passive	Object identification + naming	[120]
	Passive task (see text)	[87, 69, 246, 64, 42]

In the “silent naming” task [172, 173, 242, 246, 56], participants are asked to silently name, in their minds, a semantic concept. An alternative, related task, is the “aloud naming” task [272, 246, 256, 175, 99] in which participants name the concept aloud. This task has the advantage that participant responses can be recorded but it also has the disadvantage that other processes (such as action planning and execution) and muscular artifacts may make the neuroimaging information worse.

Many studies [238, 161, 3, 2, 148, 19, 11, 18, 113, 237, 258, 39, 23, 45, 236] asked participants to think of the same properties of the semantic concept in each experimental trial. Each participant was asked to come up with a set of properties for each concept before the start of the experiment. Several studies [272, 242] restricted the properties to various sensory and motor properties. A study by Zinszer and colleagues [286] removed the constraint of generating the properties before the experiment and let participants think freely about the meaning of the given concept or any memory it evoked. Conversely, a study by Bauer and Just [17] asked participants to think about features of animals that they had been taught about thus far in the experiment. Pereira and colleagues [198] asked participants to think about the meaning of the concept in the given context (in a sentence, with an accompanying image, or with accompanying concepts). Additionally, a study by Reddy and colleagues [213] asked participants to vividly imagine detailed mental images as similar as possible to pre-seen images in the experiment. In a work by Anderson and colleagues [10, 8], participants were asked to imagine a situation that they individually associated with the concept.

Some related research focused on more complex concepts or scenarios described by sentences, typically one sentence was presented one word (or phrase) at a time and participants were asked to think about the meaning of each phrase as it appeared and then the overall meaning of the sentence [269, 7, 278, 134, 116]. For more information on this, see related research on encoding or decoding of episodic recollection and autobiographical memory [227, 88, 247, 228, 43, 34, 218, 46, 14], or procedural knowledge [149].

In several studies [240, 239, 54, 55, 24], participants were presented with target and non-target semantic categories and asked to respond upon the appearance of items from the non-target (or target) category, for instance, by pressing a mouse button. While, in some studies [208, 141, 107, 52], participants were asked to press a button if any image repeated itself consecutively (1-back task) to ensure that participants were paying attention. Studies by a few researchers [82, 89, 223] used a 1-back match task in which participants were asked to judge whether the category matched the category presented immediately before. Studies by Carlson and colleagues and by Niazi and colleagues [42, 181] used delay matching in which participants indicated which choice of stimulus matched the target stimuli presented previously. Additionally, a study by Alizadeh and colleagues [5] asked participants to remember all six elements presented in a sequence.

Other studies focused on other semantic aspects of the concepts. For example, in a study by Sudre and colleagues [244], participants answered semantic yes/no questions for concrete nouns (e.g., “Was it ever alive?”, “Can you pick it up?”). In a study by Chan and colleagues [44], participants responded based on a size judgment of the concept, whether it was smaller or larger than 0.3 meters in any dimension, while in a study by Miller and colleagues [159] the orientation of an image stimuli was used as a form of oddball task. Studies by Fernandino and colleagues [73, 72] asked participants whether the stimulus, either a word or a pseudoword, referred to something that can be experienced through the senses. In a study by Wei and colleagues [272], participants were instructed to name a concept that was closest to the one shown in the picture. Kivisaari and colleagues [120] provided three verbal clues for each concept (e.g., “has 4 legs”, “is found in the savannah”, “has a trunk”) and participants attempted to identify the concept. In a study by Simanova and colleagues [241], participants were asked to internally produce a word in the cued semantic category with the initial cued letter. Dehghani and colleagues [66] conducted a study in which participants were asked to read a story and then answer a comprehension question. In a study by Mahon and Caramazza [144], participants

were asked to judge the conceptual similarity of two objects on a scale from 1 to 4, while Wang and colleagues [268] asked participants to judge which of two words was most similar to a key word. In a study by Li and colleagues [140], participants were asked to silently judge semantic congruity of the presented stimuli with a cued category. Finally, in a study by Honari-Jahromi and colleagues [99] participants were asked to name the color of the object or the background (in images).

Several studies used passive tasks, for instance, passive viewing of images [42, 87], passive reading [69], and passive listening [246, 64]. It has been shown that the viewed object can be identified from the passive viewing of images [252, 117, 162, 179, 48, 192]. The same argument applies for instance for speech production and passive listening. For this reason, passive tasks alone may not be sufficient to allow semantic decoding. To mitigate this issue and ensure participants attention, several studies [35, 208, 78, 236, 260] included an oddball task in which participants were asked to respond, typically by pressing a button, when a different type of stimulus was presented.

2.3.3.3 Stimulus / cue

Table 2.6 lists stimulus modalities used to cue participants to focus on a particular semantic category.

The most common modality used is the visual image presentation modality, which 41 studies used. Stimuli included photographs (gray-scale or colored) or line drawings of the concepts the participants were instructed to focus on. In 12 studies written captions or spoken words were added to the images. Written words or text, presented all at once or each word one by one, were the second most used modality and were employed by 31 studies. Spoken words (or speech) and natural sounds were used less often and were employed in only 12 studies.

A concern when using cues to instruct participants to focus on particular categories is that the presentation of the stimulus may introduce potential perceptual mental processes that can

Table 2.6 Stimulus presentation modalities used by the semantic decoding studies.

Modality	References
Image	[240, 238, 237, 45, 239, 173, 172, 272, 87, 55, 24, 35, 208, 256, 99, 134, 42, 52, 140, 141, 107, 260, 89, 129, 181, 223, 159, 212]
Image + written caption	[2, 3, 244, 198, 17, 23, 11, 161, 206, 168]
Image + auditory (speech)	[3, 286]
Auditory (speech)	[240, 239, 44, 246, 54, 144, 213, 255, 82, 120]
Auditory (natural sounds)	[239, 64]
Written word	[240, 239, 19, 44, 148, 18, 113, 246, 237, 268, 242, 5, 10, 72, 39, 258, 236, 73, 241, 175, 8, 233]
Written text or phrases	[198, 278, 69, 66, 116, 7, 78, 269, 134]

act as confounds into the classification process. For instance, focusing on a concept while seeing its image raises the question of what is used for the differentiation between different concepts: the visual processing of the image (low-level perceptual features), the imagination of the concept, or some combination of brain activities related to both processes. Some studies explicitly analyzed the influence of some of these possible confounds. For instance, Murphy and colleagues [173] examined brightness, mean spatial frequency, and visual complexity of the stimuli images. However, the set of potential confounds and methods (for instance, how to measure image complexity) has not been comprehensively studied. An alternative method is to use only certain brain regions or networks in the analysis, typically excluding visual areas [198]. However, this approach is only feasible for neuroimaging techniques with good spatial resolution, such as fMRI, intracranial electrodes, or ECoG. The separation of the task and stimulus presentation can potentially avoid this issue, see also the related field of mental imagery [192, 152, 125, 123, 176, 135]

2.3.4 Feature extraction

Depending on the recording modality, a wide variety of features can be used for semantic decoding. In fMRI and fNIRS, signals are typically epoched from 4 up to 9 seconds after the

stimulus onset to account for the hemodynamic delay in event-related designs. EEG, MEG, and intracranial electrode recordings are traditionally analyzed in: (1) the temporal domain (e.g., ERP analysis), (2) the frequency domain to reveal the signal power distribution over frequencies, or (3) the time-frequency domain for varying spectral activities over time.

Apart from these traditional features, studies have started to use domain-specific multi-dimensional information in which each concept is encoded by “semantic features”. The two main approaches used can be categorized as attribute-based views and vector space models of semantics, see also a recent review [38].

In the attribute-based view, a concept can be encoded according to its semantic attributes or features. Each attribute is assigned a value or a set of values related to its probability, weight, or importance [216, 27, 58, 79, 220]. A study by Sudre and colleagues in MEG [244] used a semantic knowledge base consisting of 218 interpretable semantic attributes. This dataset was collected by asking 218 questions to a group of Amazon Mechanical Turk users about the semantic properties of 1000 concrete nouns [187, 244]. For example, some questions were related to size, shape, surface properties, context, and typical usage, with answers on a scale of 1 to 5, and then rescaled to a range of -1 to 1. In particular, they employed a two-stage classifier with a layer of intermediate semantic features between the input features and the class label. Fernandino and colleagues [73, 72] used a semantic model based on five semantic attributes directly related to sensory-motor processes: sound, color, shape, manipulability, and visual motion. Ratings for these attributes on a scale from 0 to 6 were collected for 900 words.

In another example, Anderson and colleagues [7] used an experiential attribute model with 65 attributes [27] that modeled semantic representation using people’s ratings of their association with different attributes of experience on a scale of 0 to 6. Collected attributes spanned sensory, motor, affective, spatial, temporal, causal, social, and abstract cognitive experiences. Lastly, a study by Wang [269] developed a set of 42 concept-level semantic features. These binary features included information from categories such as the perceptual and

affective characteristics of an entity (e.g., whether it was man-made, size, color, temperature, positive affective valence, and high affective arousal), animate beings (person, human-group, animal), and time and space properties (e.g., unenclosed setting, change of location). For example, the noun ‘judge’ was encoded with the following features: person, social norms, knowledge, and communication. The study used an encoding regression model to determine the mapping between 42 semantic features as well as 6 thematic role markers of phrases in sentences and neural activation patterns assessed with fMRI.

In vector space models of semantics, automated methods can be used to learn semantic features from the statistical properties of words and phrases in large text corpora [65, 136, 156, 196, 197]. Computational linguistics has shown that contextual information provides a good approximation to word meaning [157, 49, 71, 254]. Mitchell and colleagues [161] developed a model to learn predictive relationships between the statistics of word co-occurrences (with a set of 25 verbs in a large text corpus) and fMRI neural activation patterns (BOLD activation patterns). Zinszer and colleagues [286] used representational similarity-based neural decoding to test whether semantic information of words and pictures represented by textual co-occurrence frequency in large text corpora is encoded in fNIRS.

More recently, word2vec [155, 156] and GloVe [196] have become popular semantic spaces [16, 197]. In word2vec, semantic vector representations are learnt in a way that a word can be predicted given the average semantic vector of the other words in the context (e.g., 5 words before and 5 words after the word of interest). In GloVe, representations are created in a way that the dot product of two vectors equals the logarithm of the probability of the associated words co-occurring in text. For instance, Pereira and colleagues [198] used GloVe to decode individual word meanings in fMRI while participants were instructed to think about the meaning of a target word in the given context (either in a sentence, with an accompanying image or accompanying words). Djokic and colleagues [69] investigated processing of literal and metaphoric sentences in fMRI using GloVe, a visual model, and a compositional model.

While, Kivisaari and colleagues [120] used word2vec to decode a semantic concept in fMRI while participants read brief verbal descriptions of the target concept. Participants received clues about individual concepts in the form of three isolated semantic features, given as verbal descriptions. Dehghani and colleagues [66] used an extension of word2vec for paragraph vectors to decode specific stories participants were reading in fMRI. Honari-Jahromi and colleagues [99] used word2vec to investigate neural representations of words within phrases in MEG.

2.3.5 Feature selection

Nowadays, multivariate analyses methods, such as multivariate pattern analysis (MVPA) in fMRI literature, utilizing information from multiple channels (voxels in fMRI, electrodes in EEG, etc.) are dominant, while historically many studies used to apply univariate analyses methods to the semantic decoding problem. Feature selection methods are thus typically needed to decrease the number of features from inherently high-dimensional neuroimaging data. Furthermore, feature selection methods may be used to attempt to address inter-person differences in neural encoding.

A basic method is to restrict the neuroimaging data, for instance to certain channels, time points, or frequencies. For example, the analysis can be performed on anatomically defined regions of interest or performed iteratively on small local areas (searchlight analysis in fMRI literature). Many studies [161, 238, 148, 286, 269, 19, 120] attempt to select the most stable channels over presentations of concepts within a participant, while some studies [148, 269] have applied a two-level hierarchical factor analysis to select brain locations over multiple participants.

Figure 2.3 shows which regions of the brain are most commonly used in semantic neural decoding studies. Specifically, we coarse grain the brain regions into 8 regions: frontal, parietal, temporal, and occipital brain regions in the left and right hemispheres. We then report the

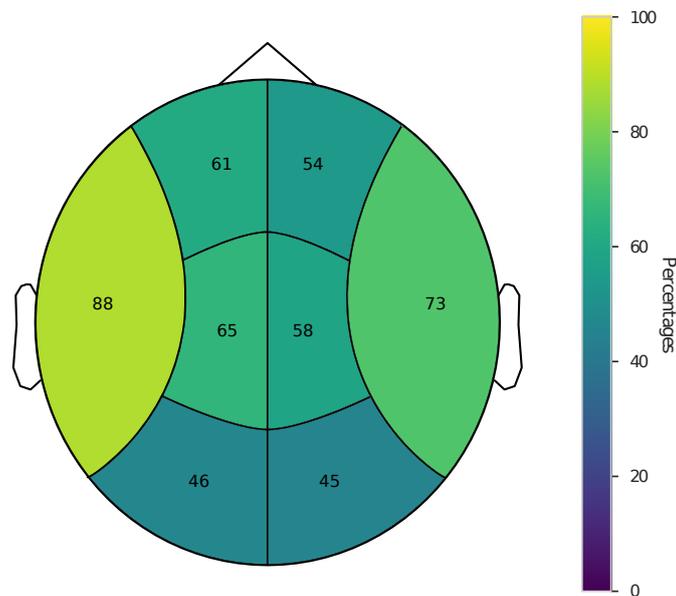


Fig. 2.3 Illustration of the most informative brain lobes for semantic decoding. Values represent percentages from the number of studies that reported this information, typically inspected in a post-analysis. Left and right frontal, parietal, temporal, and occipital lobes were chosen for a broad overview, which could be useful for a wide range of neuroimaging modalities including EEG and fNIRS decoding.

percentage of neural semantic decoding studies which make use of information from each region.

It can be seen that the left temporal lobe is most frequently used as the basis for extracting features for semantic decoding. This is not a surprising result as the left temporal lobe of the brain has been widely reported to be involved in conceptual naming [97] and, as we saw in Section 2.3.3.2, naming tasks are used in several studies, while, as we saw in Section 2.3.3.3, many studies use written or spoken concept names to present concepts to study participants. Furthermore, the anterior temporal lobe is well-known to be the hub, within the distributed-plus-hub model, for semantic memory retrieval in the brain [111, 191].

Statistical-based feature selection methods making use of the category labels can also be used. For instance, some studies [3, 2, 69] used channel selection based on ANOVA. Alternatively, supervised machine learning can be used to drive the channel selection.

Dimensionality reduction methods that project the data into a smaller subspace are popular. For example, principal component analysis (PCA) projects the data into a space with components that successively maximize the variance of the projected data, independent component analysis (ICA) decomposes the data into statistically independent components, and common spatial patterns (CSP) [169, 32] (used to aid binary classification) projects the data into a space that maximizes the signal variance for one class, while simultaneously minimizing the signal variance for the opposite class. These projections are then followed by a selection of only a certain number of dimensions, typically the ones that describe the most useful aspects of the data. It is important to note this decomposition can be spatial (over channels), temporal (over time), spectral (over frequencies), or any combination of these dimensions.

2.3.6 Machine learning models

Machine learning methods are used within semantic decoding to identify the specific semantic concept(s) an individual is focused on from a recording of their brain activity. Thus, the core aim of the machine learning part of the semantic decoding process is to categorize and classify recordings of brain activity into labels describing the associated semantic concepts.

Machine learning methods may, in general, be grouped into two distinct categories: unsupervised methods and supervised methods.

Unsupervised machine learning methods do not make use of any underlying category labels in order to process the data. Thus, they are best suited to aiding the categorization process by, for example, reducing the dimensionality of the feature space. However, they cannot, by themselves, be used to classify data [29].

Supervised machine learning methods, by contrast, make use of category labels in order to attempt to identify rules by which the data may be classified [29]. For example, supervised machine learning methods may be used to identify rulesets or thresholds to separate neural feature sets into their associated semantic category labels.

Table 2.7 Machine learning classifiers used by the semantic decoding studies. Note, some studies employ two or more classifiers and, therefore, appear in two or more rows.

Method	References
Support vector machine	[239, 173, 172, 19, 272, 44, 64, 242, 82, 54, 134, 236, 23, 45, 213, 140, 175, 141, 5, 17, 233, 255, 56, 212]
Logistic regression	[2, 3, 240, 18, 268, 241, 181]
Naive Bayes	[238, 272, 148, 113, 278, 237, 55, 24, 39, 258, 116, 23]
Regression	[120, 244, 66, 72, 198, 99, 73, 7, 78, 161, 269, 116, 260]
Linear discriminant analysis	[35, 208, 42, 52, 56, 129, 159]
K-nearest neighbors	[256, 23, 107, 246]
Neural network	[87, 10, 107, 206]
Correlation-based	[286, 69, 11, 144, 8, 89, 223]

Table 2.7 lists the machine learning classifiers used for semantic decoding. The most frequently used machine learning methods were support vector machines, naive Bayes classifiers, and regression based methods. Somewhat surprisingly, there have not been any semantic decoders to date that make use of deep learning methods such as convolutional neural networks or long short-term memory networks [137]. This is despite the rapid recent growth of the use of these methods in many related domains of neuroscience [57]. We anticipate that semantic decoding studies that use these advanced machine learning methods will begin to appear in the near future. We think the limiting factor is that the size of datasets gathered with humans is quite small in comparison with what is typically needed to train a deep ML model.

2.3.7 Measuring performance

The final step of any decoding pipeline is to evaluate the decoding performance. When only a few classes are being distinguished, standard machine learning evaluation methods can be used for binary or multi-class classification problems, such as classification accuracy, F1 score, Cohen's kappa, or preferably a confusion matrix.

With an increasing number of classes to distinguish, the above methods do not tell us the whole picture, for instance, the class may be incorrectly predicted but it would be the second

choice of a multi-class classifier or it may be semantically similar to the true class (if this makes sense in the application context). In these cases, several studies used rank accuracy [238, 19, 198, 148, 18, 113, 278, 72, 39, 17, 258, 116, 45, 161, 269], which is defined as the percentile rank of the correct class in the classifier's rank output. The list of predicted classes is rank-ordered from most to least likely and the normalized rank of a correct class in a sorted list is computed. Rank accuracy ranges from 0 to 1 and the chance level performance is 0.5.

Several studies used leave-two-out pairwise comparison [161, 244, 198, 120, 286, 66, 208, 99, 42, 11, 78, 8, 89]. This procedure leaves two samples s_1 and s_2 for testing during cross-validation. With two classes C_1 and C_2 , it compares two predicted classes and decides which order is a better match whether $(s_1 = C_1 \text{ and } s_2 = C_2)$ or $(s_1 = C_2 \text{ and } s_2 = C_1)$. The chance level performance is 0.5. For two samples and two classes, this is mathematically equivalent to the area under the receiver operating characteristic (ROC) curve measure. However, this metric makes comparisons between studies difficult unless more information is provided. Furthermore, performance measured this way is not appropriate for many real-world use-case scenarios where only two samples could be predicted and it does not consider the same class for the two samples. To make this issue more confusing, several studies incorrectly refer to this procedure as leave-two-out cross-validation. Whereas, from a machine learning perspective, leave-two-out cross-validation leaves two samples from the training and then classifies each sample separately to which class it belongs (from all possible classes). On a related note, a small number of studies only reported mean or individual pairwise accuracies from multi-class classification (e.g., from one-vs-one or one-vs-rest strategies) without trying to aggregate them together. Nevertheless, we must acknowledge that the main research focus of many studies, presented here, was on localization of brain regions involved in semantic decoding or encoding. Thus, not all reported performance metrics are useful when attempting to compare decoding accuracy between studies.

Overall, the selection of the evaluation metrics ultimately depends on the application scenario. We strongly suggest reporting everything necessary, such as confusion matrices, so that one can compute any other metric of interest, whenever this is feasible. Nevertheless, it is important to note that these metrics do not represent the whole picture of the approach used. This issue is similar to the issue of the information transfer rate (ITR) metric, which is a popular metric in BCI systems and measures the amount of information in bits that is conveyed by a system's output within a given time [151, 277, 235] (see Equation 2.2 in Section 2.3.8). Whereas, in real-case BCI scenarios, users' states, such as fatigue and perceived ease of use of the BCI must also be taken into consideration.

2.3.8 Decoding information transfer rate

We compared semantic decoding performance between studies. Due to differences in reporting metrics used in different studies it was not possible to compare performance of all the studies in this chapter. However, to make at least a partial comparison, we decided to use ITR to compare decoding performance. ITR incorporates the number of classes the semantic decoder is attempting to differentiate, the time taken to decode the concepts, and the reported decoding accuracy. It is defined, in [26], by

$$B = \log_2 C + p \log_2 p + (1 - p) \log_2 \left(\frac{1 - p}{C - 1} \right) \quad (2.1)$$

$$\text{ITR} = \frac{B}{T} \quad (2.2)$$

where C denotes the number of classes, p denotes the classification accuracy, and T denotes the time taken to make a selection in minutes.

This allows meaningful comparisons of decoding performance to be made between semantic decoding studies, even when different numbers of semantic categories and/or different time windows are employed. For comparison, consider the case where studies are compared in terms

of accuracy, or some similar metric such as the area under the ROC curve. Such a comparison is only meaningful when the number of classes and the time windows are the same across studies. For example, an accuracy of 50 % may be good when there are 4 different classes, but could be no better than random chance when there are only 2 classes.

Figure 2.4 shows ITRs in bits per minute for studies that reported decoding accuracies. Nevertheless, this information represents an optimistic view. To compute ITR, we ignored inter-stimulus intervals in experimental paradigms and instead only considered the end of the time window after the stimulus onset, that is the time window which was used for classification. It is important to note that all studies were conducted offline. In real-time semantic decoding applications, ITR would most probably be significantly lower.

As expected, neuroimaging techniques affected by a slow hemodynamic delay, such as fMRI, require longer times and thus they typically have lower ITRs (in a range from 0.02 to 9.08) in comparison with electrophysiological neuroimaging methods (with ITRs in a range from 0.09 to 149.83), even though they typically achieve greater accuracies. Indeed, if performance is measured without taking into account the length of the time window needed by the decoder for each of the neuroimaging modalities, fMRI would out-perform scalp based measures such as EEG. However, given the rapidity with which human thought can switch between semantic concepts we considered it appropriate to incorporate the length of the time window into our comparison of decoder performance.

It is worth noting that ITR is not a perfect metric for comparing semantic decoding studies as it does not take into account the semantic similarity of concepts. For example, pairs of concepts that are semantically unrelated to one another are likely to be much easier to decode than concepts that are closely related. This can be seen in Figure 2.2, which shows that categories that are quite distinct from one another, such as ‘animals’ and ‘tools’, are frequently used in semantic decoding studies, whereas more similar concepts, such as ‘celery’ and ‘carrots’, are rarely used. An ideal metric for measuring the performance of semantic decoders would

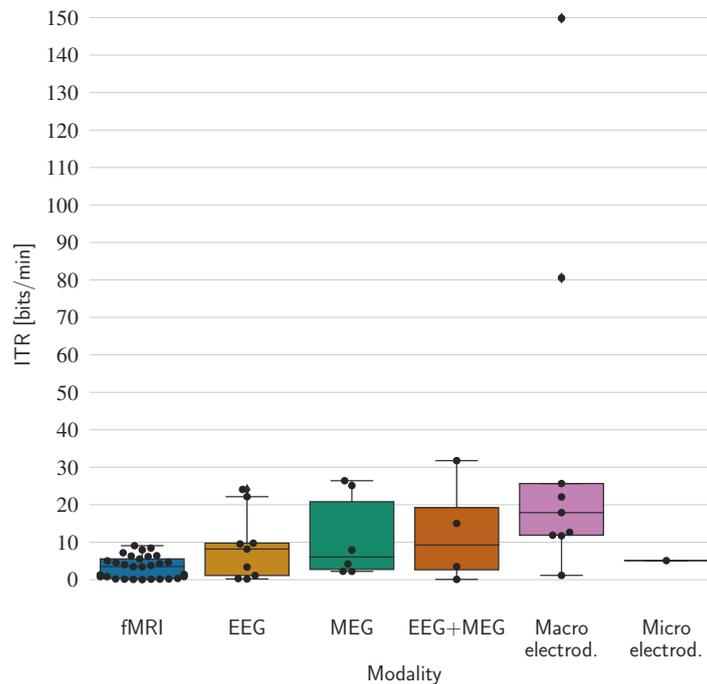


Fig. 2.4 Information transfer rate (ITR) in bits per minute for semantic decoding studies, for which it was possible to calculate ITR. Macro electrod. indicates macro intracranial electrodes such as ECoG or SEEG, while Micro electrod. indicates micro intracranial electrodes used for single unit recordings.

also incorporate some measure of the semantic similarity of the concepts that were decoded. However, as semantic similarity between concepts varies across languages, cultures, and even individuals, such a measure could prove challenging to develop and is beyond the scope of this review.

2.4 Key challenges

Semantic neural decoding has considerable potential to aid understanding of how concepts are held and processed in the brain. However, it is first necessary to overcome current gaps in our understanding of how the brain works. For example, more accurate characterization of activity patterns in terms of location, timing, and morphology has the potential to enable more accurate semantic neural decoding.

It is also necessary to improve current machine learning methods used to identify semantic concepts from neural data. This may include using joint recording methods, such as simultaneous EEG and fMRI to improve the accuracy of semantic decoding.

An additional challenge is identifying the most appropriate combinations of methods to differentiate specific sets of semantic concepts and determine which methods may be employed for particular applications. For example, fMRI may be used to differentiate a wide range of different semantic concepts, but is impractical for many possible applications of semantic decoding (for example, building a practical semantic communication device, see Section 2.5.3).

It is also important to note that the specific concepts that semantic decoders are able to differentiate currently depends largely on the neuroimaging methods employed. In general, I observed that decoders that used techniques with higher spatial resolution — such as fMRI or intracranial electrodes — were better able to decode concepts that are more semantically similar to one another than neuroimaging techniques with lower spatial resolution, such as EEG or fNIRS. Advances in signal processing techniques for the EEG and fNIRS may help to close this gap in future, but it is likely to remain the case that a higher spatial resolution is needed to more accurately decode more semantically similar concepts.

Finally, differences in inter-participant and inter-language neural encodings of semantic concepts represent a considerable challenge [2]. Ideally, one would wish to build a decoding model from one sub-group of individuals and be able to apply this with any new individual. However, neural signatures of semantic encoding vary considerably across individuals and even across experiments with the same individual [1]. There are a variety of reasons for this. In particular neuroanatomical differences between individuals mean that direct one-to-one mappings of neural encoding patterns for a given semantic concept between participants are not possible [1]. In addition, non-stationarity in neural representations of meaning results in differences in neural encoding patterns between experimental sessions with the same participant [90].

Some of these differences can be corrected for by pre-processing the recorded neural data. For example, fMRI recordings can be fit to common templates via a series of warping and translation steps to provide some degree of neuroanatomical homogeneity, at the cost of reduced spatial precision and resolution [77]. However, conceptual organization of semantic concepts differs between individuals as different people relate concepts to one another quite differently. For example, while one individual may relate the concept of ‘celery’ to the concept of ‘hunger’ another may not. These differences in conceptual organization result, according to embodiment theory, in differences in neuroanatomical localization of encoding patterns for concepts. Consequently, even with correct inter-person neuroanatomical alignment there may still be considerable differences in encoding patterns between individuals. Methods to address this include searching for signatures of semantic concepts within neural data [132] or joint feature ranking selection [2]. For example, joint feature ranking identifies signatures of concepts across different neuroanatomical structures and localization by searching for temporal dynamic modulations of neural activity that co-vary with presentations of specific semantic concepts.

An additional consideration is differences in neural encoding of semantic concepts by individuals who speak different languages. A semantic concept may be, to some extent, independent of language; the concept of ‘food’ (for example) is a universal one. However, the way specific concepts are encoded in our brain is determined by multiple factors including, but not limited to, mappings to other related concepts, and societal and cultural views of the concept. Moreover, the meaning of concepts can change over the life span [70].

Emerging evidence suggests a mixed picture, with some similarities in neural representations reported (e.g., [279, 285]). However it is not certain that these similarities will generalize well across all languages. Therefore, inter-participant / language differences (e.g., in neuroanatomical structure, as well as in structuring of neural encoding) need to be accounted for when attempting to understand semantic encoding or build semantic decoding models [1]. Meth-

ods have been developed to attempt to help overcome these differences, such as hyper-alignment analysis [90] or mutual similarity relationships [132, 210].

2.5 Current and future applications and directions

Semantic decoding allows the identification of the specific semantic concept(s) an individual is presented with or focused on at a given moment in time. This emerging field of research suggests many application areas.

2.5.1 Tools for neuroscience

Semantic decoding has the potential to provide a useful toolset to neuroscientists seeking to investigate how our brains store, relate, and process semantic concepts. For example, the multivariate pattern analysis method used in some semantic decoding studies has also been widely used to understand which brain regions are involved in representing semantic concepts [170]. Semantic decoding has also been used to build and test models of memory re-consolidation after receiving further, refining, information from input sentences [253].

A more specific example of this is the use of semantic decoding to explore neural representations of naturalistic stimulus complexity in the early visual and auditory cortices. A recent neuroscientific study by Güçlütürk and colleagues [86] used semantic decoding methods to identify how complex natural stimuli are encoded in these parts of the brain.

Other researchers have used tools developed for semantic decoding to explore how concepts at different ‘levels’ are encoded in our brains. For example, early work by Rosch and colleagues [217] defined a ‘basic level’ concept upon which other more complex concepts may be constructed. For example, Rosch defined a ‘bird’ as a basic level concept while more specific concepts (such as ‘robin’ or ‘crow’) exist at subordinate levels in this hierarchy.

This early conceptual framework has been shown, via the application of semantic decoding tools, to map to specific organizational structures for semantic encoding in the brain. For example, work by Bauer and Just [18] showed that ‘basic level’ concepts occupied more spatially distributed neural encoding patterns, while subordinate level concepts occupied less widely distributed, more concentrated brain areas.

This, in turn, relates to the distributed-plus-hub model of semantic memory retrieval in the brain. Under this model, visual, perceptual, and motor related features of individual concepts involve a distributed network of brain regions located within the brain regions responsible for the associated cognitive processes [191]. So for example, the concept of ‘tools’ is likely to be associated with motor-related cognitive processes and involves a distributed network that includes the motor cortex. This distributed network is then bound together in a central amodal hub, located within the anterior temporal lobe, which is responsible for relating semantic concepts to one another. So, for example, basic level and more complex concepts are related to one another in the anterior temporal lobe and semantic decoding studies can aid understanding of this process.

2.5.2 Clinical applications

The ability to accurately decode and classify concepts from recordings of brain activity has potential clinical applications in treating disease. An early review in this area suggested that many of the computational and neuroimaging techniques developed for semantic neural decoding could be employed to classify brain disorders such as schizophrenia and depression [36].

This approach was shown to be usable in the diagnosis of developmental dyscalculia in a small study with 13 individuals with dyscalculia and 36 control participants [167]. A time-resolved multivariate analysis method was used to analyze fMRI recorded from participants while they judged the correctness of multiplication results. The results showed detailed

differences between the groups, indicating that neural decoding techniques could be adopted for clinical diagnosis in future.

Following on from this early work, neural decoding techniques have also been applied to attempt to understand and treat aphasia [189]. Aphasia is a disorder of language that results from damage to the brain and causes deficits in the production and/or comprehension of speech. Pasley and Knight [189] suggested that neural decoding of semantic concepts could be used to understand how semantic encoding is affected by aphasia. Furthermore, they suggested that, during attempted treatment of aphasia, semantic decoding could be used to judge the effectiveness of the treatment. Treatments could then be adjusted according to this neural measure of their efficacy.

Semantic neural decoding has also been shown to be able to differentiate between individuals with schizophrenia and healthy controls [108]. Specifically, a multivariate state space model was used to analyze the representations of mental processes of individuals as they performed the Sternberg Item Recognition Paradigm [243]. Significant differences were found between controls and individuals with schizophrenia, suggesting a possible further clinical application.

More recently, neural decoding techniques have been shown, in two separate studies, to be able to differentiate between individuals with autism and control participants [114, 93].

Another recent exciting example of this is the suggestion that neural decoding of semantic concepts may be used as a potential test for Alzheimer's disease [9]. Alzheimer's disease is a progressive neuro-degenerative disease that leads to gradual loss of cognitive function and, in many cases, ultimately leads to death. One of the symptoms of Alzheimer's disease is a loss of semantic knowledge that begins years before the onset of dementia [186] and it has been suggested that this early loss of semantic knowledge could be used as an early test for Alzheimer's disease. Specifically, it was suggested in [9] that the semantic neural decoding methods developed in fMRI studies and extended to use with other neural imaging technologies, could be deployed as a test for Alzheimer's disease. However, there are considerable challenges

that first need to be overcome before this potential application can be realized. Specifically, the relationships between semantic knowledge decline and specific Alzheimer's disease pathologies needs to be more thoroughly investigated.

As a final example, neural decoding has also been demonstrated to allow identification of individuals who are engaged in suicidal ideation. Specifically, a fMRI study by Just and colleagues [115] was used to identify neural signatures related to the concepts of 'death', 'cruelty' and other concepts related to suicide in 17 suicidal ideators and 17 controls. Significant differences in neural encoding patterns for these concepts allowed differentiation of these groups with a 91 % accuracy, suggesting semantic decoding could potentially be used to identify individuals at risk of suicide.

2.5.3 Communication aids

The possibility to accurately decode the concept an individual is focused on also suggests an application as a communication aid; specifically, as a unique form of BCI.

Some work in the field of BCI is already moving in this direction. For example, the use of a single shot decoding to attempt to identify the concept an individual is focused on is one of the first attempts in BCI to deploy semantic decoding techniques as a communication paradigm [219, 244, 120, 150].

Related BCIs have been developed based on semantic relations. Geuze and colleagues [81] introduced a BCI based on EEG to determine which prime word a user had in mind. Users were presented with a probe word, the BCI detected whether the word is related to the prime word, and a new probe word was chosen from an association network. This process was repeated until a certain confidence threshold was met. An average decoding accuracy of 38 % was reported using 100 probes and 150 possible words. Additionally, Wenzel and colleagues [273] used a combination of EEG and eye gaze. Users looked for words belonging to a semantic category of interest from a stream of words on the screen. The online BCI detected whether the words

were subjectively relevant to the category. An average rank for the category of interest among the five categories was 1.62 after a hundred words had been read.

Some related research focuses on identification of cognitive concepts from neural signals in *cognitive* BCIs [6]. However, these cognitive BCIs make use of implanted electrodes (a technology which fundamentally limits their utility due to the inherent safety and ethical concerns entailed in such an approach), and are not based on semantic concepts, but rather the broader concept of cognitive states (which includes emotions, intention, executive function, motor commands, etc.) [6].

2.5.4 Other applications and privacy concerns

Finally, the ability to identify the specific semantic concept an individual is focused on, or thinking of, has numerous other potential applications that, to date, have only been briefly suggested in the literature.

One such application is the use of neural decoding in the field of “neuromarketing”. This field suggests the use of neuroscientific techniques to develop, refine, and test marketing strategies for commercial products, for example by measuring neural signature of affective (emotional) responses to particular products [13].

Semantic decoding methods may be used to identify which specific concepts an individual focuses on when shown advertising material. This could, in turn, be used to identify more effective advertising strategies.

However, applications such as this and other similar possible uses of semantic decoding suggest the need to consider the privacy and ethical issues raised by semantic decoding [91]. Specifically, neural decoding offers the possibility to decode and interpret a part of an individual's current mental state. This could, theoretically, be done without the permission of the individual, for example as a part of a criminal investigation.

The associated privacy and ethical issues are rarely considered in the majority of the literature on semantic neural decoding, perhaps because the technology is currently at a very early stage where such applications feel a long way off. However, one recent discussion paper [153] begins to consider these issues and develops an evaluation framework to consider issues of privacy and ethics in the field of neural decoding. We anticipate considerably more discussion on these issues as the field develops further.

2.6 Discussion

We systematically sought records of studies that attempted to develop semantic neural decoders. The search methodology included searches of PubMed records and Google Scholar and included all relevant peer reviewed articles that we could identify on these databases. However, no literature review can ever be completely comprehensive and we may have neglected to include some studies that describe semantic neural decoders, either because the title and abstract did not indicate that this was attempted in the study, or because we misunderstood the title and abstract and incorrectly excluded the paper. Thus, while this review considers the majority of semantic neural decoding studies it may not be fully comprehensive. Nevertheless, we are able to draw some key conclusions from our analysis of this literature.

Specifically, the majority of neural semantic decoders make use of the fMRI to record neural data, while a smaller number of studies use other methods such as EEG or MEG. The range of concepts that these decoders attempt to identify is relatively large but there is a core subset of concepts (such as animals and tools) that are very frequently decoded. Experimental designs vary considerably across studies with a wide range of different types of cues and experimental tasks used. On the other hand the range of machine learning methods used by semantic decoders is relatively modest, comprised largely of support vector machines and regression based methods.

The relationship between semantic encoding models and decoding models is not always consistently described in the literature. Indeed some studies confuse these two terms and present an encoding study as a decoding study or *vice versa*. We have endeavored to only include studies that present semantic decoding models. However, an important caveat is that some *encoding* models are constructed in such a way that adapting the model to achieve semantic *decoding* would be extremely trivial. Indeed, in some cases an encoding model is also, in effect, a decoding model because the predicted encoding maps the model identifies are explicitly linked to discrete semantic concepts. In such cases we have included the study in our review.

Understanding how our brains encode semantic concepts is an important goal in modern neuroscientific research and enables many new and exciting areas of research. Not least among these is the rapidly developing area of semantic decoding, the attempt to develop processing pipelines and decoding models to identify the specific semantic concept an individual is focused on from recordings of their brain activity.

We have identified several key methods employed to tackle the challenge of semantic decoding. Although there are many challenges inherent in developing and evaluating effective models, semantic decoding has the potential to identify, sometimes with quite high levels of accuracy, the specific concept an individual is focused on. This may, in future, enable a wide range of applications such as new clinical diagnostic tests or fast and accurate communication aids.

Chapter 3

Experiment

This chapter introduces an experiment which was designed to explore the feasibility of differentiating between the semantic categories of animals and tools in a silent naming task and three novel and intuitive sensory-based imagery tasks using visual, auditory, and tactile perception. Participants were asked to visualize an object in their minds, imagine the sounds made by the object, and imagine the feeling of touching the object. Recorded data via simultaneous EEG and fNIRS are analyzed in the following chapters.

3.1 Key choices

Chapter 2 reviewed studies exploring the possibility of semantic neural decoding, a technique which aims to identify specific semantic concepts an individual is focused on, or thinking of, at a given moment in time from their neural activity.

3.1.1 Neuroimaging methods

While the most promising results to date have been reported using fMRI (see Sections 2.3.1 and 2.3.8), semantic decoding using neural signals recorded from the scalp, such as EEG or fNIRS, is of particular interest for potential semantic BCIs. EEG and fNIRS are portable,

relatively cheap, and provide better ecological validity in comparison with fMRI. EEG provides a good temporal resolution but has a poor spatial resolution. This could potentially be improved by combining it with fNIRS, as these two techniques have complementary strengths. Thus, the joint recording of EEG and fNIRS was used, more information on this is in Section 3.6.

3.1.2 Mental tasks

Semantic decoding studies to date have used a variety of different mental tasks (see Table 2.5 in Section 2.3.3.2). However, not all these mental tasks would be suitable for BCI applications. An important difference with respect to research on semantic decoding is that, in semantic BCIs, users would freely choose and focus on a semantic concept of their choice (from a supported set of recognizable concepts). In other words, there would not be an external cue that is used to drive their particular choice. We selected four different mental tasks that could potentially be used for semantic BCIs.

First, the silent naming task, in which participants are asked to silently name in their minds a presented object on the screen, has been used in previous research in EEG [173, 172, 171]. However, the silent naming task has not been tested in fNIRS yet. Therefore, we test its feasibility for semantic decoding in fNIRS.

Many studies asked participants to think about the properties or meaning of the concept or to generate mental images of the concept (see task types for ‘Properties’, and ‘Meaning’ in Table 2.5). In general, many studies have shown that similar patterns of brain activity arise when perceiving and imagining objects [192, 48, 138, 213]. We follow their lead but we want to use a mental task based on pure imagination. This is analogous to motor imagery tasks in which participants imagine an action involving movement, for instance, moving a leg or an arm, or walking on a beach [180, 200, 225]. Thus, we move toward mental imagery tasks [152, 192, 125, 123] in which participants could freely use their mental imagery to think about concepts. Mental imagery is considered a fundamental ability of the human mind, which most

of us have [192, 230, 176]. We thus consider mental tasks based on mental imagery to be natural mental tasks to most of us. We propose three novel and intuitive sensory-based imagery tasks using visual, auditory, and tactile perception. We asked participants to visualize an object in their minds, imagine the sounds made by the object, and imagine the feeling of touching the object. There is more detail on this in Section 3.2. We considered these three modalities to be the most intuitive for mental imagery tasks that were also used in some forms in mental imagery research [152, 192, 125, 230, 199].

Although each imagery task instructs participants to use a different sensory modality, we do not expect participants to use only that modality. For instance, visual imagery will probably be used during the tactile imagery task. However, the focus here is not on the particular sensory modality but on whether the mental task instruction could be used for semantic neural decoding. Research suggests a shared neural network for mental imagery between different sensory modalities. For example, a review by McNorgan [152] investigating modality-specific imagery in the auditory, tactile, motor, gustatory, olfactory, and visual domains found a general imagery network across all modalities, while modality-specific imagery overlapped with corresponding somatosensory processing and motor execution areas. This shared network could then be used to investigate the feasibility of semantic neural decoding regardless of the modalities employed by participants.

In the silent naming task, participants might use ‘internal speech’ in addition to silently naming the concept. This behavior might also be present during the imagery tasks but probably to a lesser extent. The decoding algorithm could then exploit neural features related to internal speech instead of the concept. This would confound the decoding process and create false positives for semantic neural decoding. Neural correlates should be investigated for how much they are shared between the silent naming task and the imagery mental tasks for these possible confounds. However, successful speech decoding has primarily been reported in invasive

neuroimaging techniques with much less success in EEG [12, 207, 63] so this might not be an issue here.

Although not implemented in our experiment, in theory, all four mental tasks could be used without an external cue to drive participants' choice. The participants could voluntarily think about the concept they wish to communicate. This makes the mental tasks also suitable for use outside the lab settings. Without the need for external cues, BCI paradigms could be directly adapted for use outside the lab settings with relative ease, in contrast with the majority of the reviewed mental tasks from Section 2.3.3.2. However, it is unknown whether semantic BCIs would require any synchronization for the decoding process (to indicate the mental task start and intentional control) or self-paced BCIs would be possible.

The performance, and thus usability, of BCI systems is impacted by several interconnected factors, such as, users' engagement, motivation, attention, cognitive load, or fatigue [163, 109, 266, 118, 146]. Our idea is to use mental tasks that elicit a high degree of neural activity related to each concept while keeping the participants engaged. However, we did not perform any preliminary experiments to investigate the difficulty of the chosen mental tasks. If the mental tasks were too easy or too difficult, participants could be either bored, resulting in mind wandering, or anxious due to the task difficulty. In both cases, this would negatively affect the BCI performance. As mental imagery is considered to be innate to most of us, we did not consider these mental tasks to be difficult, providing enough time to perform them. We decided that 3 seconds is the optimal amount of time to perform mental imagery, see Section 3.4 for the experimental design. Nevertheless, imagery vividness and strength vary between individuals [194, 61]. Adapting the mental task length to participant psycho-physiological states could improve the system [163, 109]. To initially test the feasibility of the chosen mental tasks in the simplest scenario, we postpone these considerations for future research.

3.1.3 Semantic categories

To determine the feasibility of the proposed mental tasks, we attempt to differentiate between the semantic categories of animals and tools. This pair of semantic categories has been mostly used in existing semantic decoding studies (see Section 2.3.3.1). More on this is in Section 3.3.

3.1.4 Stimuli

Before using the mental tasks without any external cue in BCIs, the mental tasks must first be tested in an experimental design which cues participants for a particular semantic concept. We use a visual modality in the form of images to cue participants because this modality has been most frequently used in semantic decoding studies (see Section 2.3.3.3) and it is appropriate for all mental tasks.

In contrast with most semantic decoding studies, we attempt to ensure a separation between the cue presentation period and the mental task period. More on this is in Section 3.4. As discussed in Section 2.3.3.3, this separation avoids the problem of potential processing-related confounds in the classification process that is present in many studies. For instance, focusing on a concept while seeing its image raises the question of what is used for the differentiation between different concepts: the visual processing of the image (low-level perceptual features), the imagination of the concept, or some combination of brain activities related to both processes. This is a pertinent question as it has been shown that the viewed object can be identified from passive viewing of images [192, 48]. This study design mitigates this problem. This issue is further examined in Chapter 5. Additionally, a mask is used after the image presentation period to reduce visual persistence [214].

3.2 Mental tasks

Participants were presented with images of concepts from the two semantic categories of “animals” and “tools” to focus on. Each image presentation was followed by a set of four individual mental tasks: silent naming, visual, auditory, and tactile imagery. The order of mental tasks was randomized across blocks (see Section 3.4).

Silent naming task In the silent naming task, participants were asked to name the presented object in their minds, in their mother tongue (English), see Figure 3.1.

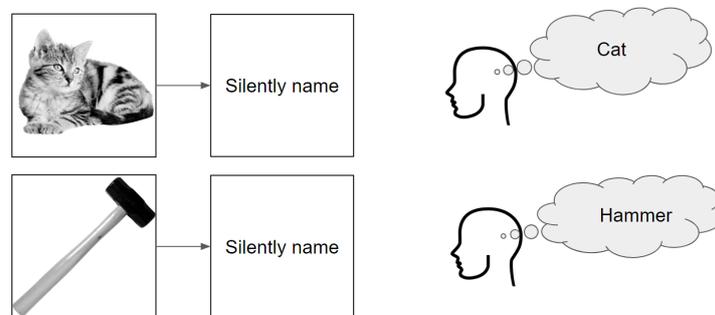


Fig. 3.1 Example of the silent naming task from a part of the material shown to participants.

Visual imagery task In the visual imagery task, participants were asked to visualize the presented object. They were instructed to try not to visualize the particular image they had seen but their own representation of the concept, see Figure 3.2.

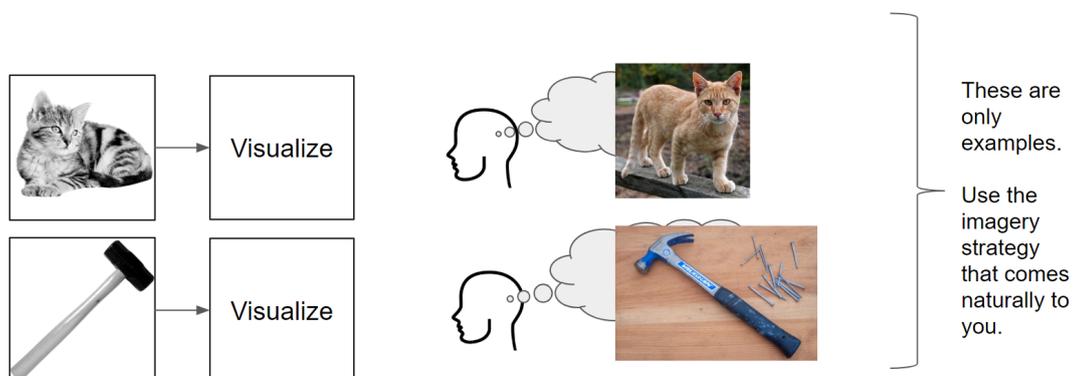


Fig. 3.2 Example of the visual imagery task from a part of the material shown to participants.

Auditory imagery task Similarly in the auditory imagery task, participants were asked to imagine sounds made by the presented object when they interact with it. For instance, the sounds made by an animal (such as the meowing of a cat) or the sounds produced when using a tool (such as the banging of a hammer), see Figure 3.3.

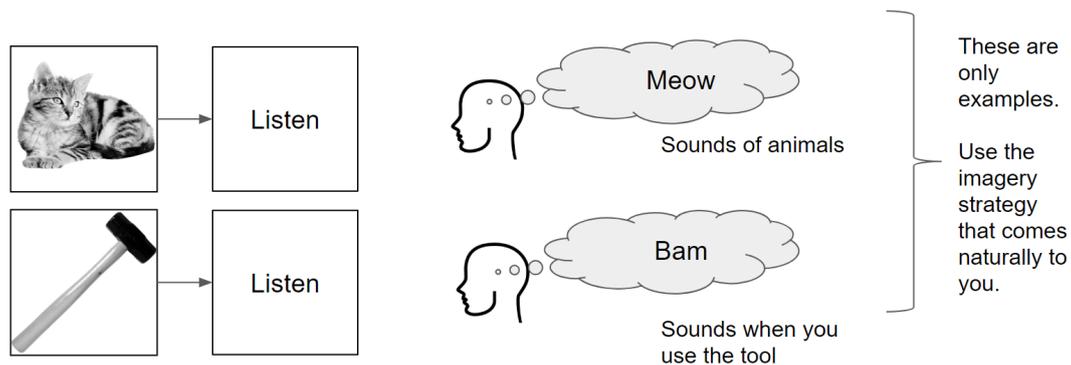


Fig. 3.3 Example of the auditory imagery task from a part of the material shown to participants.

Tactile imagery task Lastly, in the tactile imagery task, participants were asked to imagine the feeling of touching the presented object. For instance, when petting an animal or touching different parts of a tool, see Figure 3.4.

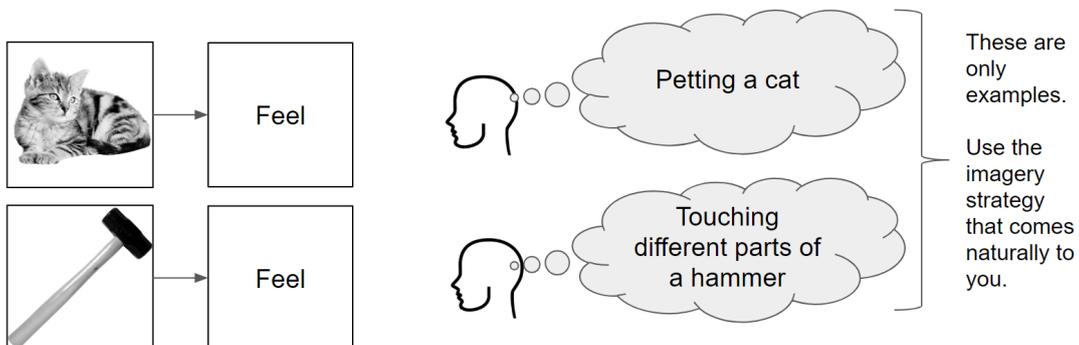


Fig. 3.4 Example of the tactile imagery task from a part of the material shown to participants. Participants were instructed to imagine the feeling of touching the presented object during the tactile imagery task, for instance, petting a cat or touching a hammer in this example.

Participants were presented with the above-mentioned descriptions of the mental tasks (including the examples reported) but instructed to use the imagery strategy that came most naturally to them. For all imagery tasks, they were instructed to be engaged for the whole

mental task duration (3 seconds). They were also instructed to try to avoid eye movements, facial muscle contractions, and head or body movements during the mental tasks.

Additionally, we did try not to influence the participants' interpretation of images. They were shown all images before the experiment. We named a particular image only if they could not recognize it. Otherwise, we let them use their own interpretations.

3.3 Stimuli

A set of 18 animals and 18 tools were selected. We selected as many concepts from previous studies [240, 173] as possible. We used concepts that are suitable for all mental tasks and recognizable by most people. For instance, many animals were excluded from the initial set as they were unsuitable for the auditory imagery task because we judged that a considerable number of people would not be able to recall, and then imagine, the sound such animals make. Images (with a license allowing non-commercial reuse with modifications) were sourced from the Internet. They were converted to gray-scale, cropped, resized to 400×400 pixels, and contrast stretched. In all images, the object was presented on a white background. We used photos instead of line drawings because we did not want to be constrained by their limited variety. The selected concepts are listed below. Figure 3.5 shows their corresponding images.

Animals: bear, cat, cock, cow, crab, crow, dog, donkey, duck, elephant, frog, lion, monkey, owl, pig, sheep, snake, and tiger.

Tools: axe, bottle-opener, broom, chain saw, computer keyboard, computer mouse, corkscrew, hammer, hand saw, Hoover, kettle, knife, microwave, pen, phone, scissors, shovel, and toothbrush.



Fig. 3.5 Images of concepts presented during the experiment.

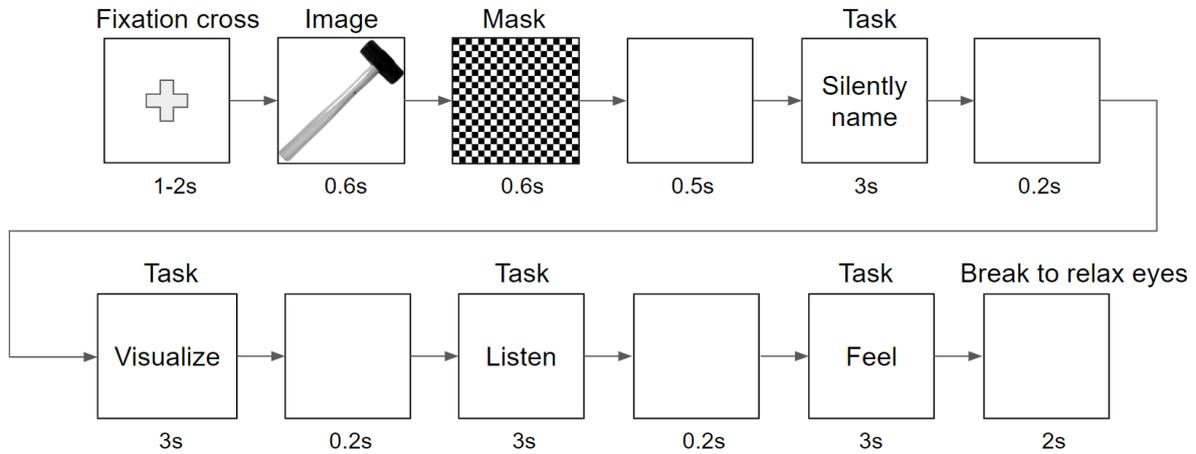


Fig. 3.6 Illustration of one concept trial. The order of mental tasks is randomized across blocks.

3.4 Experimental design

Figure 3.6 illustrates one concept trial. Each concept trial started with a black fixation cross on a white background for 1-2 seconds (uniformly distributed). The image of a concept was then presented for 0.6 seconds. A mask (the image of a checkerboard) followed for 0.6 seconds to reduce visual persistence and thus to get rid of potential effects of perceptual processing related neural activity from the image presentation after this mask presentation [214]. A blank white screen was shown for 0.5 seconds before a sequence of all four mental tasks. Each mental task lasted for 3 seconds and was separated from the following one by a blank white screen for 0.2 seconds. The type of the mental task was indicated by text presented on the screen for the whole mental task duration: “Silently name”, “Visualize”, “Listen”, or “Feel”. The last mental task was followed by a short break for 2 seconds indicated by a blank screen, which changed color over time from white to black and back. In total, one concept trial took 17.3–18.3 seconds depending on the duration of the fixation cross.

Each concept was presented five times, for a total of 90 trials per category (18 concepts, 5 repetitions each). The experiment was split into 15 blocks with 12 concepts per block (207.6–219.6 seconds). Blocks were separated by breaks of at least 30 seconds. There was also a longer break of at least 3 minutes in the middle of the experiment.

Table 3.1 Participants information.

Participant	Age	Sex	Handedness	fNIRS montage
1	26	female	right	frontal
2	32	female	right	frontal
3	57	male	right	frontal
4	47	female	right	frontal
5	23	female	right	frontal
6	21	female	right	frontal
7	29	female	right	temporal
8	50	female	right	temporal
9	27	female	right	temporal
10	33	female	right	temporal
11	28	male	right	temporal
12	20	male	right	temporal

The experiment started with two additional short blocks (86.5–91.5 seconds, containing a random subset of 5 concepts, each repeated two times) for familiarization with the experiment.

The order of concepts and mental tasks was pseudo-randomized with the following constraints. No concept was repeated twice in succession. All mental tasks in one block had the same order. Different blocks had different orders of mental tasks and no order in a given block was repeated in the following block.

3.5 Participants

Twelve right-handed native English speakers were recruited from the student and staff population of the University of Essex (3 males and 9 females, age range 20–57, mean 32.75, standard deviation 11.55; see Table 3.1 for more information). The recruitment search was limited to native English speakers to avoid a potential problem of differences in neural representations of semantic concepts by individuals who speak different languages [279, 285], which could be particularly important for the silent naming task (see Section 3.2). All participants had normal or corrected-to-normal vision. Participants received compensation of £16 for their time.

They all read, understood, and signed a consent form. The study was approved by the Ethics Committee of the University of Essex on 25th October 2018.

3.6 Neuroimaging data

EEG data were acquired with a BioSemi ActiveTwo system with 64 electrodes positioned according to the international 10-20 system, plus one electrode on each earlobe as references. Additionally, 2 electrodes placed on the left hand measured galvanic skin response and a respiration belt around the waist measured respiration. The sampling rate was 2048 Hz. fNIRS data were acquired with a NIRx NIRScoutXP continuous wave imaging system equipped with 4 light detectors, 8 light emitters (sources), and low-profile fNIRS optodes. Both electrodes and optodes were placed in a NIRx NIRScap for integrated fNIRS-EEG layouts. EEG and fNIRS signals were synchronized by sending a trigger at the onset of each event in the experiment from the computer running the experiment to both systems simultaneously via a parallel port and an active parallel port splitter box.

The brain regions selected for fNIRS channel placement were based on a review by Binder and colleagues [28] which analyzed 120 functional neuroimaging studies to locate the semantic system for semantic processing, which refers to the cognitive act of accessing stored knowledge about the world. Their analysis showed a left-lateralized network comprised of 7 regions: posterior inferior parietal lobe (angular gyrus and adjacent supramarginal gyrus), middle temporal gyrus (and posterior portions of the inferior temporal gyrus), fusiform and parahippocampal gyri, dorsomedial prefrontal cortex, inferior frontal gyrus (especially pars orbitalis), ventromedial prefrontal cortex, and posterior cingulate gyrus (and adjacent ventral precuneus). These findings are consistent with semantic decoding studies that were reviewed in Chapter 2. We focused on regions close to the scalp which could potentially be studied by fNIRS, namely left-lateralized: posterior inferior parietal lobe, middle temporal gyrus, and dorsomedial prefrontal cortex.

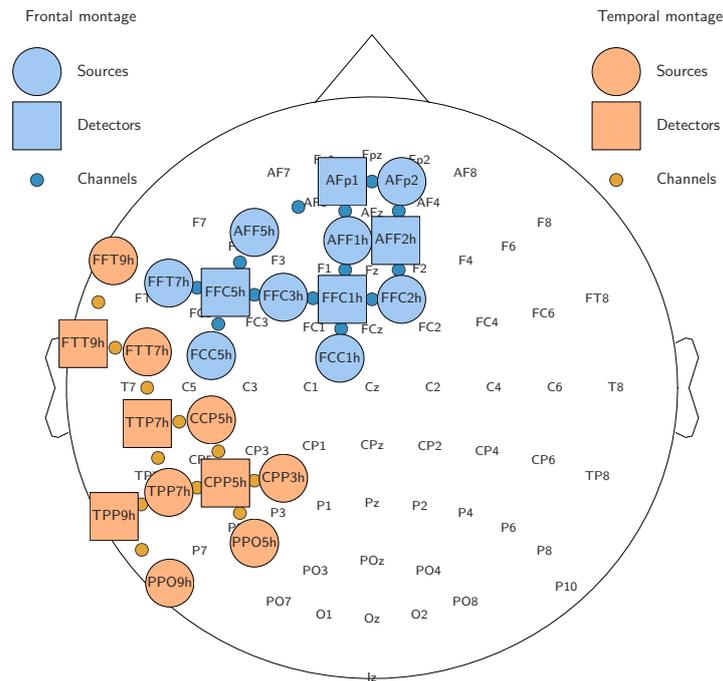


Fig. 3.7 The frontal and temporal montages used for fNIRS data acquisition with the joint EEG with 64 electrodes following the international 10-20 system. fNIRS sources (circles) and detectors (squares) positioned in the 10-5 system form channels (small circles) shown between them.

We designed two fNIRS montages because the lab's equipment did not include enough channels to record from all regions of interest. Figure 3.7 shows our two fNIRS montages. The first montage focused on the left temporal lobe and the posterior inferior parietal lobe. The montage was selected based on the fOLD toolbox [284], which allows for probe arrangement guided by brain regions of interest, with regions of interest in the left-lateralized inferior parietal lobe, angular gyrus, middle and inferior temporal gyrus. We selectively eliminated optodes with the lowest specificity for the regions of interest until the number of optodes satisfied the lab's equipment limitations. This montage used 4 detectors and only 7 sources creating 11 channels with a sampling rate of $f_s = 8.92$ Hz. The second montage focused on the left frontal cortex. The montage was selected using the same method based on the fOLD toolbox with regions of interest in Brodmann areas 9 and 46. This montage used 4 detectors and 8 sources creating 14 channels with a sampling rate of $f_s = 7.81$ Hz. The inter-optode separations were

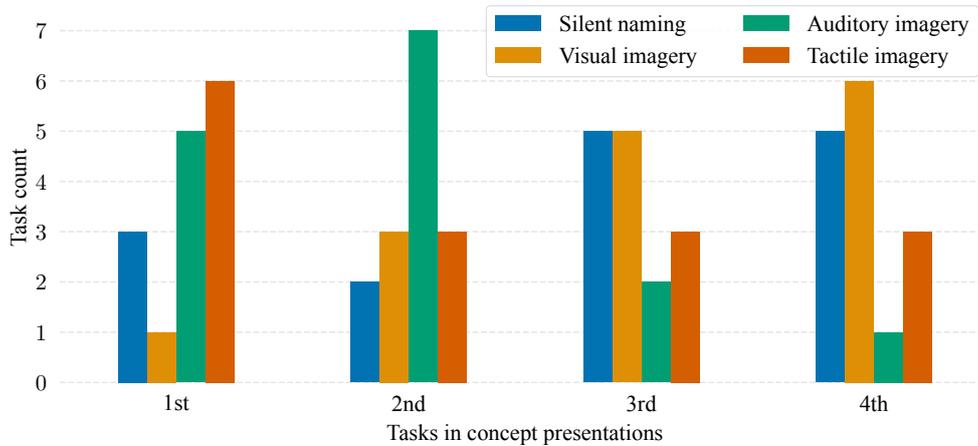


Fig. 3.8 Number of mental task appearances as the first, second, third, and fourth task in the shared order of mental tasks in participants 2 to 12.

approximately 3 cm. For the first six participants, the frontal montage was used and for the final six, the temporal montage was used (see Table 3.1).

3.7 Order of tasks

In the course of data analysis, we discovered a minor error in the experiment script that was used to present stimuli to participants. All participants except participant 1 had the same seed for the random number generator and they thus shared the same order of mental tasks and concepts. Figure 3.8 shows how many times mental tasks appeared as the first, second, third, and fourth mental task in blocks. The distribution is clearly not uniform. It was skewed to present the auditory and tactile imagery tasks more frequently as the first mental task and the auditory imagery task as the second mental task.

If the ease of semantic decoding depends on the mental task position during the sequence of four mental tasks, the randomization seed error might negatively affect the results. For example, if the first presented mental task, regardless of the type of the mental task, provided the most useful features to differentiate between the semantic categories, the tactile and auditory imagery tasks might have an unfair advantage. These mental tasks were presented more often than the

other two mental tasks, especially in comparison with the visual imagery task, resulting in providing more samples for the training process in the classification pipeline. Thus, they might be able to achieve higher classification accuracies.

We prepared a post-hoc analysis to investigate how much the results depend on different distributions of the order of task appearances. A particular analysis would be re-computed on subsets of blocks from the experiment by excluding individual blocks from the analysis to simulate different distributions. We used subsets of blocks with sizes from 7 to 14 blocks and only considered subsets in which all mental tasks appeared as the first, second, third, and fourth task. From these 4023 subsets, 100 were randomly selected. We would then check whether or not the original results would fall within the “middle” of the distribution. If yes, the results would be significant even if the order of the tasks had been correctly randomized across participants. In other words, the error we discovered in the experiment script would not adversely affect the results. However, as we will see in analyses of the EEG and fNIRS data in Chapters 4 and 6, the mean classification accuracies across participants were not statistically significant in any mental task. Thus, we did not perform this post-hoc analysis.

Ideally, we should inspect semantic decoding performance of each mental task for each task appearance position during the sequence of four mental task separately, for instance, classification accuracies when the silent naming task appears as the first task, then as the second task, and so on. This analysis would provide insights into the effects of the mental task position on semantic decoding while also providing answer whether or not this error negatively affected the results. Unfortunately, there is not enough data for each mental task at each position to perform this analysis.

If the mental task order is essential and affects the decoding performance, we expect to see some mental tasks achieve higher classification accuracies than other mental tasks. However, this is not the case, as we will see in our EEG and fNIRS analyses in Chapters 4 and 6. Mean classification accuracies were not statistically significant for any mental task in all exploratory

analyses. These results cannot be fully attributed to the mental task order. Thus, random seed error did not adversely affect the results.

Chapter 4

Semantic decoding in EEG

This chapter explores the possibility of semantic decoding with EEG in the experiment presented in Chapter 3.

4.1 Methods

We investigated the feasibility of differentiating between the semantic categories of animals and tools by a set of different analyses. All analyses presented in this chapter used a stratified 15-fold cross-validation to assess the decoding performance.

4.1.1 EEG data preprocessing

The EEG signals were referenced to the mean of electrodes placed on the left and right earlobe. Channels with bad signal quality were manually identified and removed ($7.5 \pm 7.3\%$ of channels, mean and standard deviation, see “Bad EEG channels” column in Table 4.1). The EEG data were high-pass filtered to remove slow drift artifacts by an IIR 4th-order Butterworth filter with 1 Hz cutoff (forward-backward filtering which creates an effective order of 8 with -6.02 dB at a 1 Hz cutoff, `sosfiltfilt` method from SciPy [263] using second-order sections to reduce numerical errors). Epochs representing concept trials (a presentation of a concept image with a

Table 4.1 Number of bad channels and bad concept trials in the EEG recording from each participant. EEG contains 64 channels in total. The experiment contains 180 concept trials in total.

Participant	Bad EEG channels	Bad EEG concept trials
1	5	17
2	16	16
3	4	18
4	12	37
5	3	21
6	1	30
7	5	12
8	excluded	excluded
9	2	29
10	1	24
11	0	15
12	4	29

sequence of all tasks, see Section 3.4) were extracted from the last second of the fixation cross, before the image presentation, until the end of the last task (14.3 seconds after the image onset). We inspected the concept trials for artifacts in any channel and marked bad concept trials that were then excluded from further analysis ($12.5 \pm 4.2\%$ of concept trials, see “Bad EEG concept trials” column in Table 4.1).

Artifacts from eye blinks were suppressed by independent component analysis (ICA) [112, 106]. Spatial ICA (FastICA from SciPy [263] with default settings but with a maximum number of iterations set to 1000 and with a number of components corresponding to a number of valid channels) was fitted on the valid concept trials for each participant. The ICA component explaining largest amount of the variance always represented eye blinks; we manually checked this. Finally, this component was removed from the high-pass filtered data.

Participant 8 was excluded from further analysis due to overall bad signal quality and artifacts in most concept trials.

4.1.2 Temporal analysis

We explored the possibility of differentiating between the semantic categories of animals and tools in the temporal domain using specific EEG frequency bands.

We selected the two most widely used frequency bands from previous EEG-based semantic decoding studies (see Table 5.1): 1–30 and 4–30 Hz. In both cases, the preprocessed EEG data were first high-pass filtered by an IIR 4th-order Butterworth filter (1 or 4 Hz cutoff with -6.02 dB at the cutoff, forward-backward filtering, `sosfiltfilt` method) and then convolved with a FIR low-pass filter (designed by the `firwin` method from SciPy [263], 0.661 s length, 30 Hz cutoff, 5 Hz bandwidth) [275]. Additionally, we used individual frequency bands of delta 1–4 Hz, theta 4–8 Hz, alpha 8–14 Hz, beta 14–20 Hz, and low gamma 20–30 Hz. This selection was based on [24]. For these frequency bands, the preprocessed data were filtered by a FIR band-pass filter (designed by the `firwin` method) with settings based on filtering recommendations from the MNE framework [85]. Concretely, settings (filter length, lower bandwidth, upper bandwidth) were (6759 samples (3.3 s), 1 Hz, 2 Hz) for 1–4 Hz, (3381 samples (1.651 s), 2 Hz, 2 Hz) for 4–8 Hz, (3381 samples (1.651 s), 2 Hz, 3.5 Hz) for 8–14 Hz, (1931 samples (0.943 s), 3.5 Hz, 5 Hz) for 14–20 Hz, and (1353 samples (0.661 s), 5 Hz, 7.5 Hz) for 20–30 Hz. In all cases, the filtered EEG data were then downsampled to 64 Hz. Finally, data were restructured into epochs from 0 to 3 seconds after the individual mental task onset (with the exception of data from the excluded concept trials).

Two approaches for the classification were investigated: a sliding window approach and the classification of the whole mental-task period. The sliding window approach was used to investigate the temporal evolution of semantic decoding. A temporal window was shifted in steps of half the window size in the mental task period. A classifier was separately trained in each position of the temporal window by using only time points from that window. Different lengths of the temporal window were investigated: 0.25, 0.5, and 1 second that correspond to 16, 32, and 64 time points at the sampling frequency of 64 Hz. In the classification of the whole

mental task period, the classifier could use information from all time points in the mental task period.

As a classifier, we chose a support vector machine (SVM) with a radial basis function kernel (from scikit-learn [195], default settings, $C = 1$). This is because SVM is the most used classifier in semantic decoding studies (see Table 2.7 in Section 2.3.6). We also tested other classifiers: logistic regression (LR) with L1 or L2 norm, and linear discriminant analysis (LDA). Even though there were several (minor) differences, the main message was the same and we will, thus, report only the results achieved with the SVM (unless it is specified otherwise).

The SVM could use information from all channels. Data in each temporal window or in the whole mental task period were normalized (z-scored) for each channel separately, based on training folds during cross-validation. A single channel classification was initially tested. However, we decided to only present all channels classification in this chapter to make the presentation clearer because the main message was the same for both approaches. (The results of single channel classification for analyses in this section are shown in the next chapter in Sections 5.3.1 and 5.3.4 for the frequency bands of 1–30 and 4–30 Hz.)

Naturally, using information from all channels might result in too many features for the classifier, which might cause overfitting. This problem might be solved by decomposition techniques which reduce the number of input features to a smaller set of highly informative ones. We investigated using the two decomposition techniques described in the following sections.

4.1.3 Decomposition by PCA

We used a spatial PCA to decrease data dimensionality in channel space. The preprocessed EEG data from Section 4.1.1 were filtered for the same selection of frequency bands and downsampled to 64 Hz as in Section 4.1.2. The filtered and downsampled data were projected into a smaller subspace by keeping only the first N PCA components. We tested a selection

of $N \in \{2, 5, 10, 15, 20, 25, 30\}$ components and also the number of components that explains $\{85\%, 90\%, 95\%, 98\%\}$ variance in the data. Similarly to the previous section, we investigated the possibility of semantic decoding using two approaches: sliding window and whole mental task period classification. In both approaches, the selected components were normalized (z-scored) separately.

4.1.4 Decomposition by CSP

We also used a CSP [169, 32, 211] algorithm to decrease data dimensionality in channel space. The CSP is a popular supervised decomposition method for EEG signal analysis to distinguish between two classes. This is in contrast with the PCA algorithm described in the previous section which is an unsupervised decomposition method. The CSP method finds spatial filters that maximize the signal variance for one class while simultaneously minimizing the signal variance for the opposite class. Section B.2 in Appendix B offers a more in depth description of the CSP algorithm and its possible implementations.

Similarly to previous sections, the preprocessed EEG data were filtered with the same selection of frequency bands and downsampled to 64 Hz. This transformed data were spatially filtered by CSP and a subset of N CSP components were extracted. The subset contained the $N/2$ CSP components with the largest eigenvalues and the $N/2$ components with the smallest eigenvalues. This is a typical approach in CSP applications, see Section B.2 in Appendix B. The log variance of the selected CSP components was extracted and used as an input for the SVM. We tested $N \in \{2, 4, 6, 8, 10\}$ CSP components. This approach was again tested with the sliding window and the whole mental task period classification.

4.1.4.1 An issue of decreased dimensionality of EEG data

The CSP algorithm assumes that covariance matrices of the EEG data have full rank. However, the ICA used for eye blinks suppression in the preprocessing step (see Section 4.1.1) removed

one IC and thus reduced the dimensionality of the EEG data. Apart from being mathematically incorrect, this issue leads to potential errors in the CSP decomposition and decreased classification accuracies [222]. Appendix B describes this issue in detail and shows one possible solution by using spatial PCA that is employed here. This analysis of the effectiveness of different CSP implementations has been published in [222].

In order to have full rank covariance matrices, a dimensionality reduction method, such as spatial PCA, can be used to project the EEG data into an appropriate space. In this case, the spatial PCA projects the EEG data into a space with a number of dimensions equal to the number of channels minus one. The number of dimensions was decreased by one because one IC was removed. This dimensionality reduction is employed before the CSP.

4.1.4.2 Nested-cross-validation approach

Additionally, we investigated the possibility of automatic selection of a number of CSP components by a nested-cross-validation approach. The number of CSP components was selected based on the maximal mean classification accuracy from the inner stratified 10-fold cross-validation.

4.1.4.3 Spatio-spectral CSP

In this section we explore whether or not the decoding performance improves when the classifier can use information from multiple frequency bands. We only considered the delta, theta, alpha, beta and gamma frequency bands for this approach. We used a spatio-spectral CSP, that is CSP filtering in the joint channel and frequency space. The number of dimensions of this space corresponds to the number of channels (up to 64 channels) multiplied by the number of frequency bands (5 frequency bands), that is up to 320 dimensions. We performed the PCA operation before the CSP to reduce the number of dimensions to the number of channels minus one multiplied by the number of frequency bands.

In this approach, we considered a selection of $N \in \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20\}$ CSP components. We then used a nested-cross-validation approach to automatically select the number of CSP components in the same manner as described before.

4.1.4.4 Grid search

The previously selected frequency bands may not always be appropriate for analysis using CSP for all participants. Several EEG-based semantic decoding studies first searched for participant-specific frequency bands [173, 172, 246]. We employed a grid search to test all time-frequency windows for each participant and mental task.

The mental task period (3 seconds) was split into a grid with a step of 0.2 seconds. All possible temporal windows on this grid were then considered: 0–0.2, 0–0.4, ..., 0–3, ..., 0.2–0.4, 0.2–0.6, ..., and 2.8–3 seconds. This resulted in 120 temporal windows in total.

Similarly, the frequency domain in the frequency range 0–50 Hz was split into a grid with 5 Hz steps. This resulted in 55 frequency windows in total. We filtered the data with a FIR band-pass filter (designed by the `firwin` method) with automatic filter settings based on the MNE `filter` method following the framework filtering recommendations. To avoid aliasing, the EEG signals were downsampled to 128 Hz instead of 64 Hz.

All combinations of $120 \times 55 = 6600$ time-frequency windows were tested. Additionally, classification in each time-frequency window was explored with more options than before. Namely, we tested $N \in \{4, 6, 8, 10, 12\}$ CSP components and different classifiers: SVM with $C \in \{0.01, 0.1, 1, 10, 100\}$, LR with L1 or L2 norm, and LDA.

4.1.5 Time-frequency representation

The epochs were transformed into a time-frequency representation by a wavelet decomposition using Morlet wavelets. The EEG signals were preprocessed as described in Section 4.1.1 with a high-pass filter with a cut-off frequency of 1 Hz and keeping the original sampling frequency

of 2048 Hz. Frequencies were extracted from 1 to 40 Hz with a step of 1 Hz. The number of cycles for each frequency was set to the half of this frequency. The temporal domain of time-frequency representation was then downsampled (decimated) to 128 Hz. To avoid edge artifacts, each epoch was extended by 0.5 seconds before and after the epoch period before the decomposition and cropped to the original period afterwards. We used the signal power of this time-frequency representation.

We used a spatio-spectral PCA to decrease data dimensionality in the joint channel and frequency space. The number of dimensions of this space corresponds to the number of channels (up to 64 channels) multiplied by the number of frequencies (40 frequencies), that is up to 2560 dimensions.

Signal powers were normalized (z-scored) for each channel and frequency separately. This data was then projected into a smaller subspace by keeping only the N PCA components that explain most of the data variance. We tested $N \in \{1, \dots, 100\}$.

We used a sliding window approach to inspect the decoding performance over time. In each temporal window, values of each component were averaged to further decrease the number of features. The number of features thus equals the number of components. These features were normalized (z-scored) before being passed to the SVM ($C = 1$) mentioned before.

We tested temporal window lengths of 125 ms (16 time points) and 250 ms (32 time points).

4.1.6 Event-related (de)synchronization

Event-related (de)synchronization (ERD/S) maps were explored for each mental task regardless of the semantic category. The preprocessed EEG data from Section 4.1.1 (removal of bad trials and suppression of eye blinks) were further transformed by the spatial Laplacian to reduce the negative impact of volume conduction (`compute_current_source_density` method from MNE). Time-frequency representations were computed by multitaper method (`tfr_multitaper` method from MNE) for the frequency range of 2–35 Hz. Epochs in time-

frequency representations were baseline corrected with the break period of 2 seconds to compute relative changes in percentages. Statistical significance was accessed by a cluster-based permutation test (`permutation_cluster_1samp_test` method from MNE, 10^5 permutations, $p < 0.05$).

4.2 Results

The classification accuracy of a single participant was tested for statistical significance by comparing it with that of a random classifier by using a one-sided Binomial test for a significance level of $p < 0.05$. The probability distribution of the random classifier for two categories (i.e., the semantic categories of animals and tools) can be described by the binomial distribution. The binomial distribution with parameters n and q describes the number of successes in a sequence of n independent experiments, each with a Boolean-valued outcome for success (with probability q) or failure (with probability $1 - q$). The semantic category of animals or tools for each epoch represents the Boolean-valued outcome and is independent from other epochs due to the random generation of the concept order. Thus, the binomial distribution can be used to describe the random classifier with n set to the number of epochs. Due to the removal of bad epochs during the preprocessing step, the categories of animals and tools were not balanced which is reflected by setting the parameter q to the ratio of samples between the two categories. A one-sided Binomial test is then used to measure the statistical difference of deviations from this theoretically expected distribution.

Significance for the mean classification accuracy across participants was calculated using a bootstrapping simulation (based on 10^6 simulations using a classifier which randomly choose a class). Significance borderlines were then 56.6% for $p = 0.05$, 59.33% for $p = 0.01$, and 62.25% for $p = 0.001$. Note that if the numbers of epochs were the same for all participant, this bootstrapping would be equivalent to a Binomial test.

4.2.1 Temporal analysis

Figure 4.1 shows the results for two representative participants obtained with the sliding temporal window of 0.5 second when the SVM ($C = 1$) can use information from all channels. The figure shows statistically significant classification accuracies ($p < 0.05$, one-sided Binomial test) for each mental task and frequency band. There was considerable variability in the temporal locations of significant classification accuracies between different participants for each mental task and frequency band, see Figure C.1 in Appendix C for the results from all participants. It was, therefore, not possible to select a temporal window across all participants that would allow discrimination between the semantic categories for any mental task in any frequency band. Due to this, it was not possible to compare which mental task and frequency band performs better for semantic decoding in the sliding window approach. This was also true for all other tested options of different window sizes and classifiers.

When the classifier can use information from the whole mental task period and all channels, mean classification accuracies across participants were not statistically significant in any frequency band and mental task. Due to this, it was not again possible to compare which mental task and frequency band performs better for semantic decoding.

4.2.2 Decomposition by PCA

Figure 4.2 shows the results for two representative participants for each mental task obtained with the sliding temporal window of 0.5 second when the SVM can use information from the first N PCA components. The results revealed variability in the temporal locations of significant classification accuracies between different participants for each mental task and frequency band, see Figures C.2, C.3, C.4, and C.5 in Appendix C for the results from all participants for each mental task. As a result, mean classification accuracies across participants were not statistically significant in any temporal window for all tested options, including different window sizes

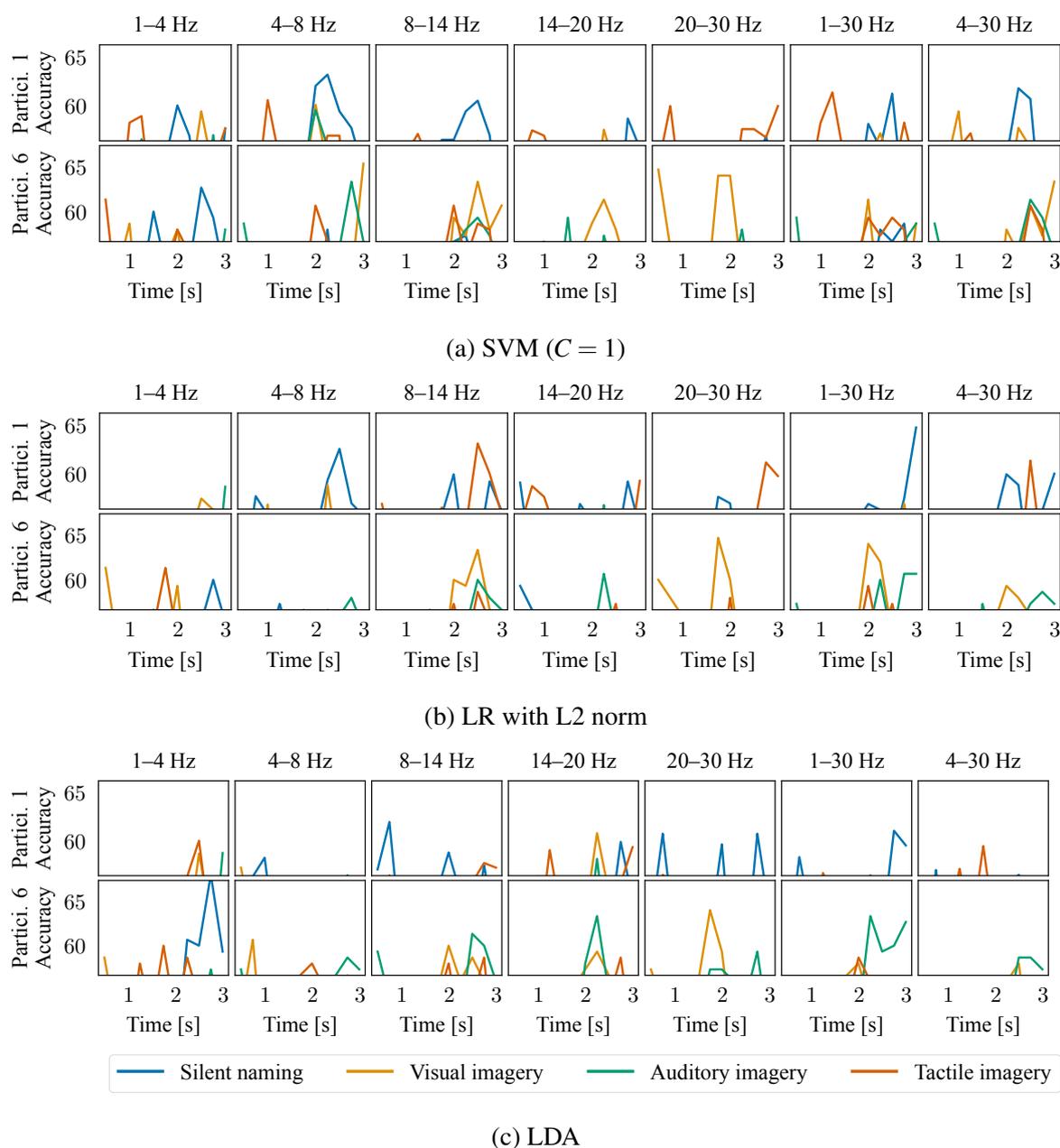


Fig. 4.1 Classification accuracies for two participants (1 and 6) obtained with the sliding temporal window of 0.5 second when the classifier can use information from all channels, see Section 4.1.2. Columns represent different frequency bands. The top row is participant 1 and the bottom row is participant 6. The Y axes are limited to statistically significant classification accuracies from $p = 0.05$ to $p = 10^{-5}$. Subfigures show the results achieved using different classifiers.

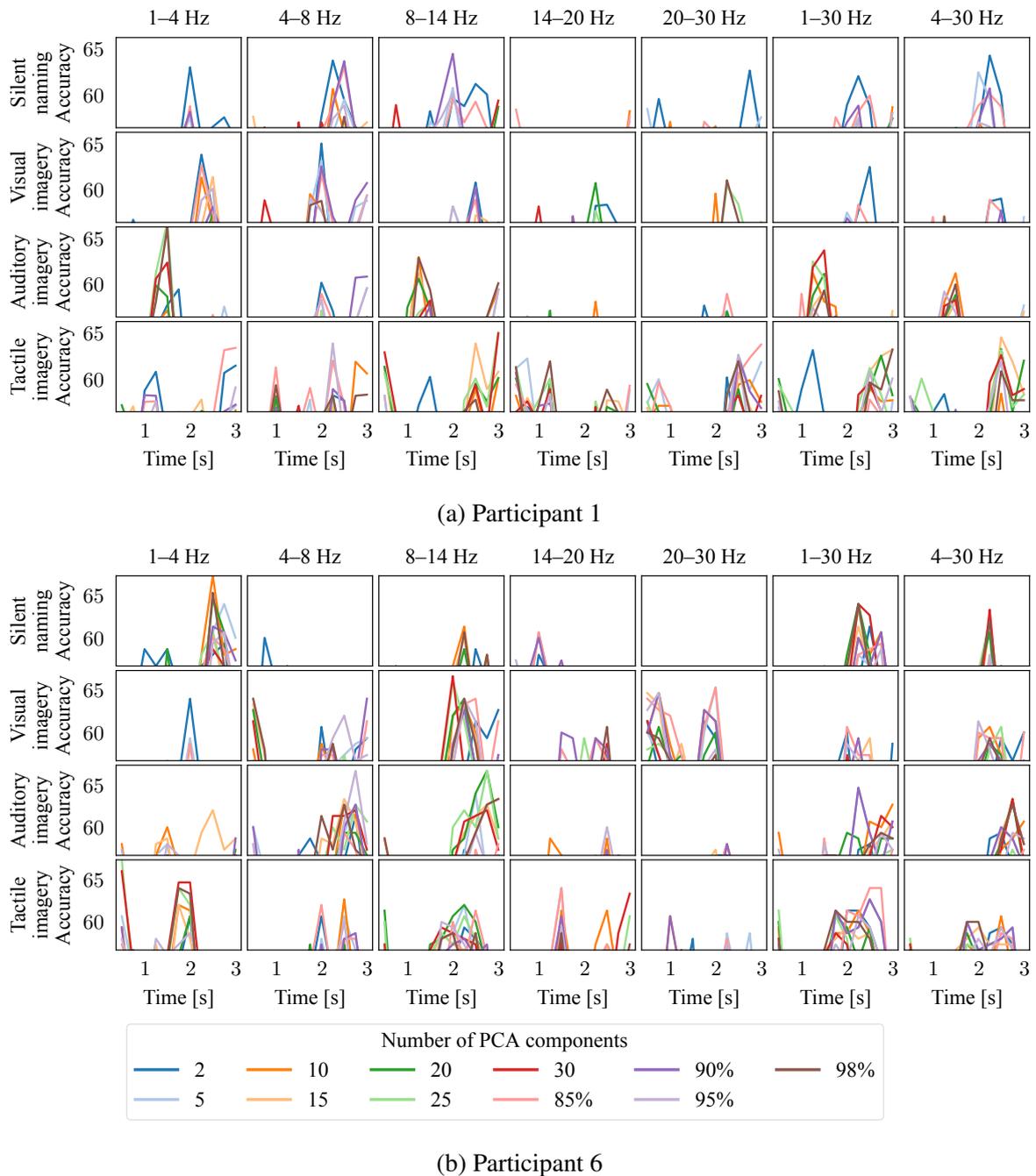


Fig. 4.2 Classification accuracies for two participants (1 and 6) obtained with the sliding temporal window of 0.5 second when the SVM ($C = 1$) can use information from the first N PCA components, see Section 4.1.3. Rows represent different mental tasks. Columns indicate different frequency bands. The Y axes are limited to statistically significant classification accuracies from $p = 0.05$ to $p = 10^{-5}$.

and classifiers. In the whole mental task classification, mean classification accuracies across participants were not statistically significant in any tested option.

4.2.3 Decomposition by CSP

Figure 4.3 shows the results for two representative participants for each mental task obtained with the sliding temporal window of 0.5 second when the SVM can use information from N CSP components. Once again, the results revealed variability in the temporal locations of significant classification accuracies between different participants for each mental task and frequency band, see Figures C.6, C.7, C.8, and C.9 in Appendix C for the results from all participants for each mental task. Mean classification accuracies across participants were not statistically significant in any temporal window for all tested options, including different window sizes and classifiers. In the whole mental task classification, mean classification accuracies across participants were not statistically significant in any tested option.

4.2.3.1 Nested-cross-validation approach

Figure 4.4 shows the results for two representative participants obtained with the sliding temporal window of 0.5 second and the SVM. The nested-cross-validation approach was used to select the number of CSP components which were used in the classification, see Section 4.1.4.2. There was variability in the temporal locations of significant classification accuracies between different participants for each mental task and frequency band, see Figure C.10 in Appendix C for the results from all participants. Mean classification accuracies across participants were not statistically significant in any temporal window for all tested options, including different window sizes and classifiers. In the whole mental task period classification, mean classification accuracies across participants were not statistically significant in any tested option.

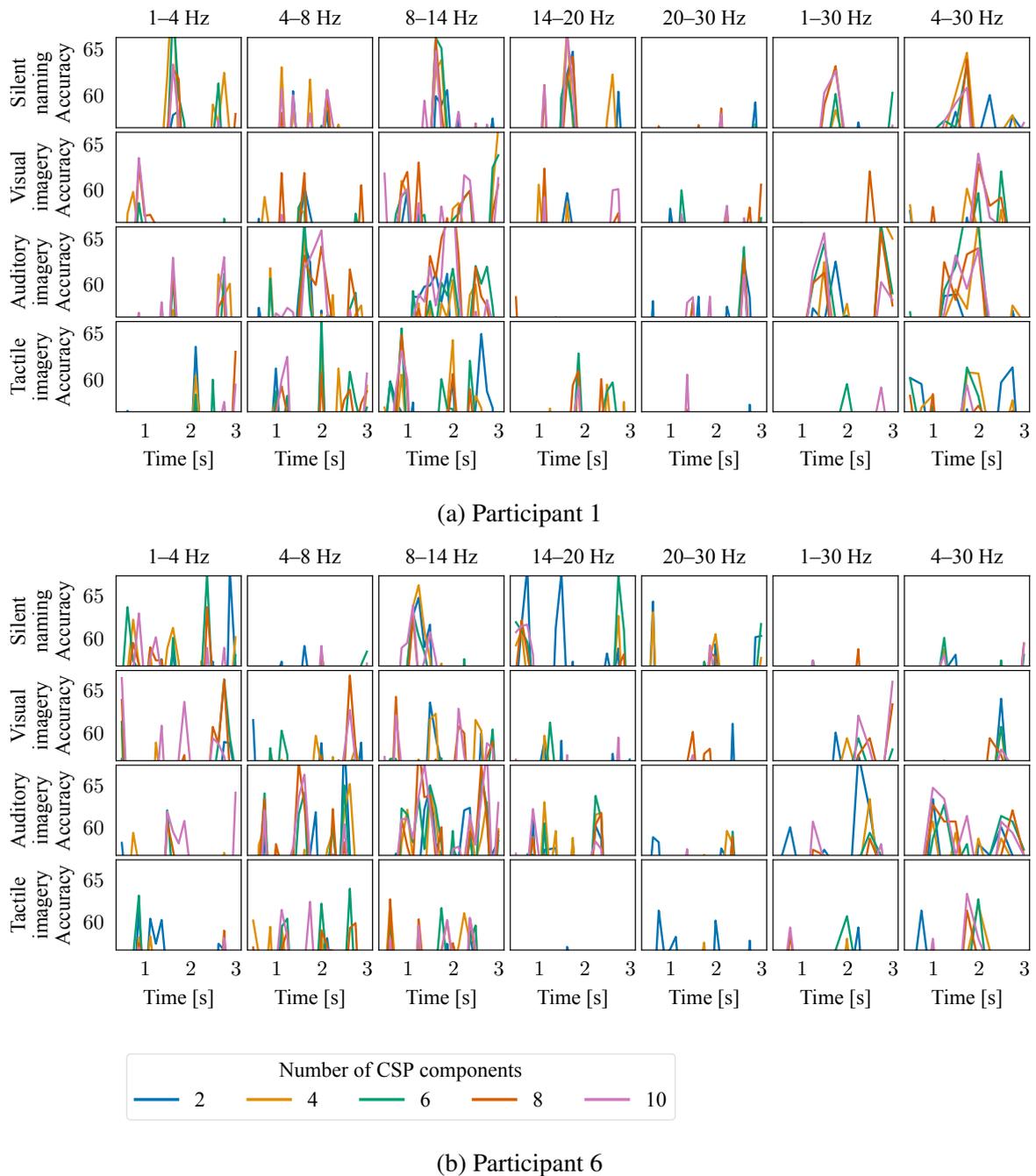


Fig. 4.3 Classification accuracies for two participants (1 and 6) obtained with the sliding temporal window of 0.5 second when the SVM ($C = 1$) can use information from N CSP components, see Section 4.1.4. Rows represent different mental task, while columns indicate different frequency bands. The Y axes are limited to statistically significant classification accuracies from $p = 0.05$ to $p = 10^{-5}$.

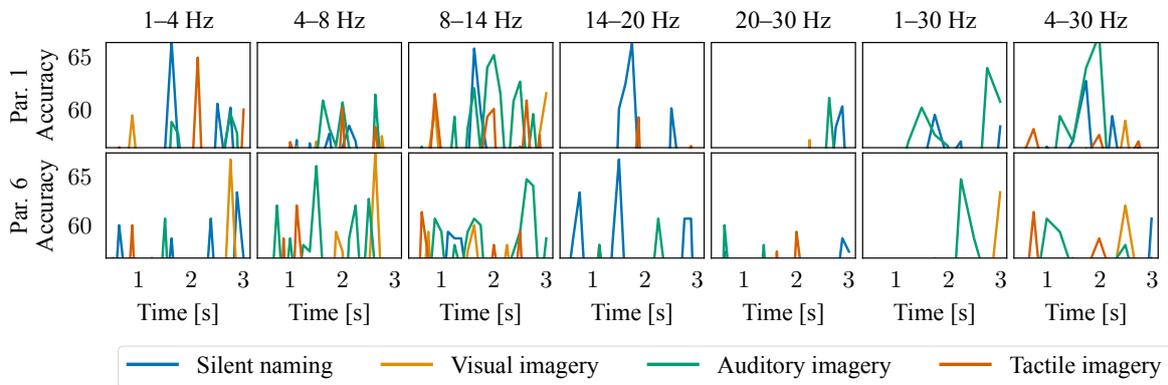


Fig. 4.4 Classification accuracies for two participants (1 and 6) obtained with the sliding temporal window of 0.5 second when the SVM ($C = 1$) can use information from the number of CSP components that was selected by the nested-cross-validation approach, see Section 4.1.4.2. Rows represent different participants, while columns indicate different frequency bands. The Y axes are limited to statistically significant classification accuracies from $p = 0.05$ to $p = 10^{-5}$.

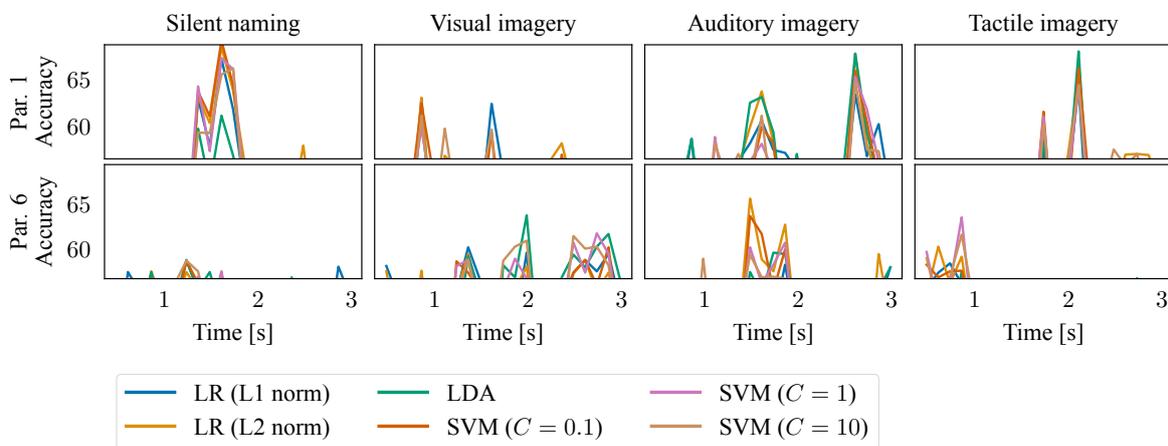


Fig. 4.5 Classification accuracies for two participants (1 and 6) obtained with the sliding temporal window of 0.5 second when the classifier can use information from the number of CSP components from the joint channel and frequency space that was selected by the nested-cross-validation approach, see Section 4.1.4.3. Rows represent different participants, while columns indicate different mental tasks. The Y axes are limited to statistically significant classification accuracies from $p = 0.05$ to $p = 10^{-6}$.

4.2.3.2 Spatio-spectral CSP

Figure 4.5 shows the results for two representative participants obtained with the sliding temporal window of 0.5 second. The nested-cross-validation approach was used to select the number of CSP components from the joint channel and frequency space, which were used in the classification, see Section 4.1.4.3. There was variability between different participants and mental tasks in the temporal locations where statistically significant classification accuracies were found, see Figure C.11 in Appendix C for the results from all participants. Mean classification accuracies across participants were not statistically significant in any temporal window for all tested options, including different window sizes and classifiers. In the whole mental task period classification, mean classification accuracies across participants were not statistically significant in any tested option.

4.2.3.3 Grid search

Due to the difficulty of visualizing high-dimensional space of all possible time-frequency windows, numbers of CSP components, and classifiers (see Section 4.1.4.4), let's first explore what would be the best achievable classification accuracies with the SVM, which was used in all previous analyses.

Figure 4.6 shows summarized results for frequency and temporal domains separately for two representative participants. For each temporal window shown in this figure, all temporal windows from the grid search that make use of this temporal range were considered across all tested options of different numbers of CSP components and frequency windows. A value of 90% quantile is then computed from their classification accuracies and shown in this figure. A similar procedure was applied to the frequency domain. The temporal and spectral locations where statistically significant accuracies were found varied between different participants and different mental tasks, see Figures C.12, C.13, C.14, and C.15 in Appendix C for the results from all participants. Overall, for the majority of participants and the majority of mental tasks

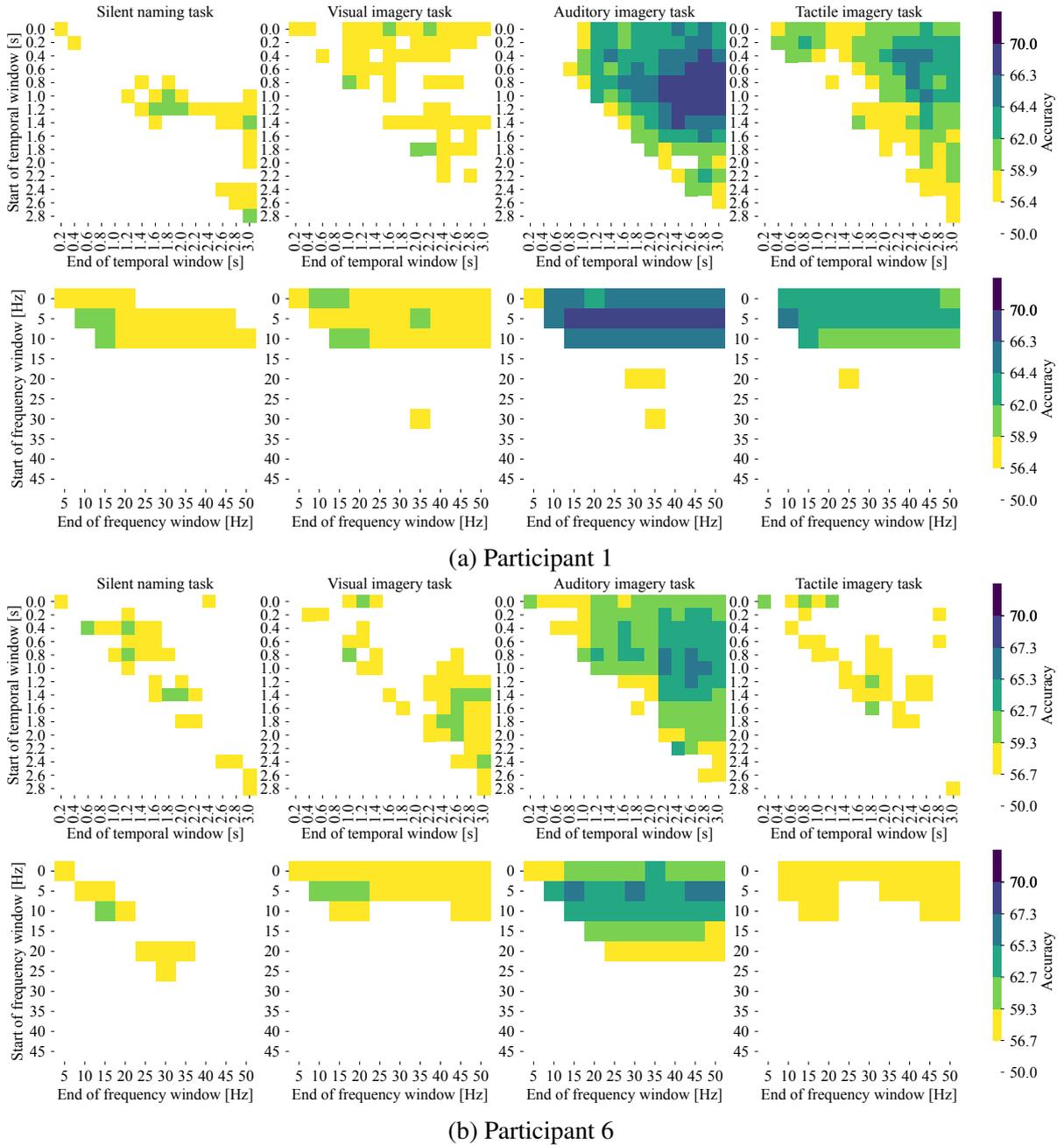


Fig. 4.6 Summary of the “best” achievable classification accuracies for two participants (1 and 6) from the grid search for the CSP analysis using the SVM ($C = 1$), see Section 4.1.4.4. Columns indicate different mental tasks. The top row represents temporal windows and the bottom row represents frequency windows. The Y axes show the start of the window and the X axes show the end of the window. For each temporal window shown in the top row, all temporal windows from the grid search that make use of this temporal range were considered across all tested options of different numbers of CSP components and frequency windows. A value of 90% quantile is then computed from their classification accuracies and shown in this figure. A similar procedure was applied to the frequency domain. Classification accuracies in colorbars represent significant borderlines for p -values of $p \in \{0.05, 0.01, 0.001, 0.0001, 0.00001\}$. White represents non-significant classification accuracies.

it was possible to find time-frequency regions that allow significant classification accuracies. However, these regions were not shared between different participants *with the exception of the auditory imagery task* in which the mean of maximal accuracies across all participants (bottom-right plot in Figure C.14) contained statistically significant regions. This was not present in any other mental task. The results when using other tested classifiers showed a similar variability in the spectral and temporal locations between different participants and mental tasks.

It is worth noting that we examined maximal accuracies across different options in this exploratory analysis. In a realistic classification analysis, accuracies would most probably be significantly lower.

4.2.4 Time-frequency representation

Both temporal window sizes of 125 and 250 ms had visually similar results in the sliding window approach when spatio-spectral PCA was employed to decrease the dimensionality of the time-frequency representation, see Section 4.1.5. We thus report here only the results for the temporal window size of 125 ms.

Figure 4.7 shows the results for two representative participants when the SVM can use information from up to 100 components from the spatio-spectral PCA of signal powers. The results revealed variability in temporal locations where statistical significant classification accuracies occurred between different participants for each mental task, see Figures C.16, C.17, C.18, and C.19 in Appendix C for the results from all participants. Mean classification accuracies across participants were not statistically significant in any mental task.

4.2.5 Event-related (de)synchronization

Figures 4.8, 4.9, 4.10, and 4.11 show ERD/S (ERD/ERS) maps for each mental task regardless of the semantic category for twenty representative electrodes, see Section 4.1.6. Overall, all

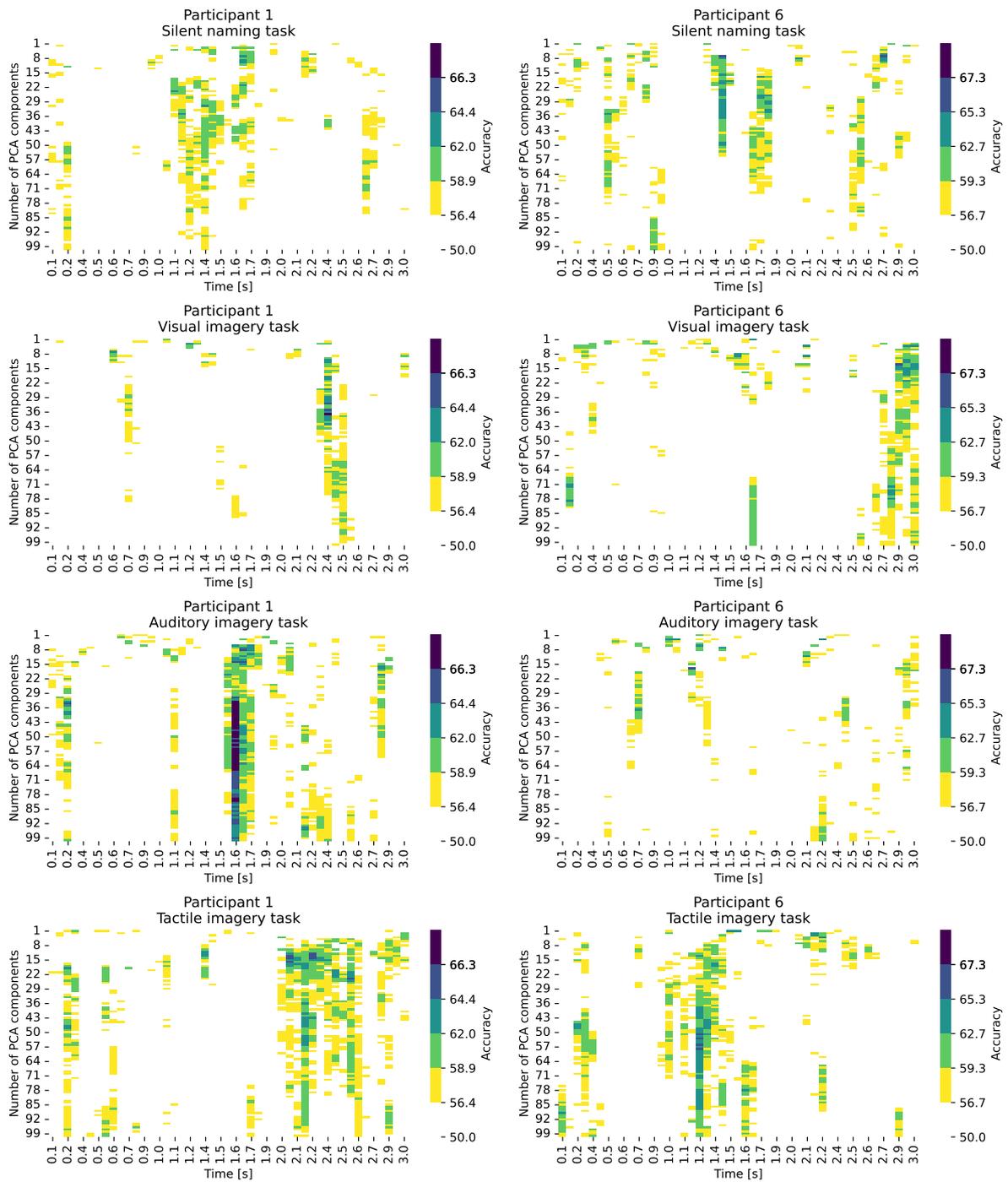


Fig. 4.7 Classification accuracies for two participants (1 and 6) with the sliding temporal window of 125 ms when the SVM can use information from up to N spatio-spectral PCA components from the joint channel and frequency space, see Section 4.1.5. The left column is participant 1 and the right column is participant 6. Rows represent different mental tasks. Classification accuracies in colorbars represent significant borderlines for p -values of $p \in \{0.05, 0.01, 0.001, 0.0001, 0.00001\}$. White represents non-significant classification accuracies.

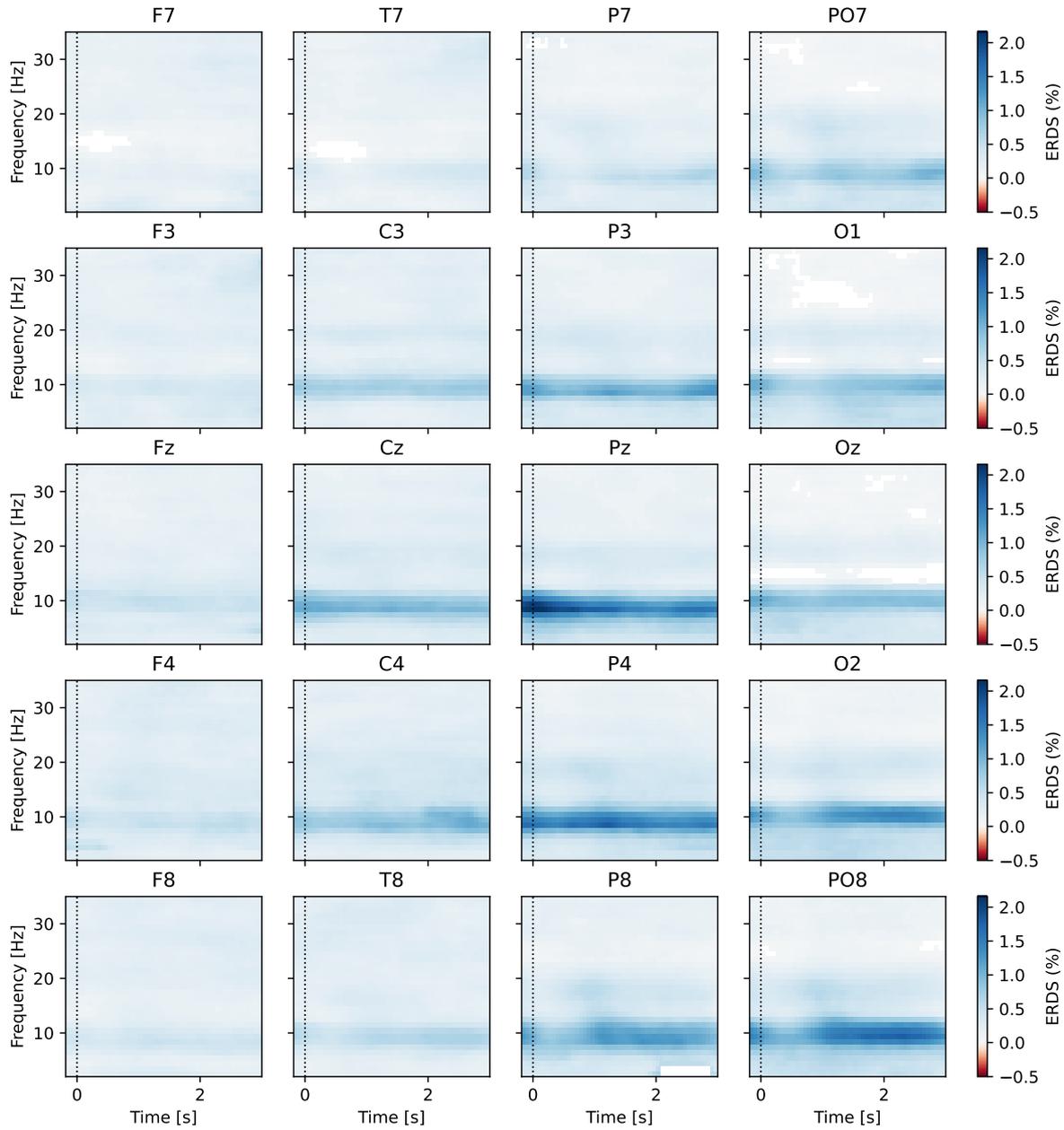


Fig. 4.8 ERD/S maps for the silent naming task of twenty representative electrodes. X axes are from 200 ms before to 3 seconds after the mental task onset. Y axes show the frequency range of 2–35 Hz. Only statistically significant values are shown (cluster-based permutation test), see Section 4.1.6.

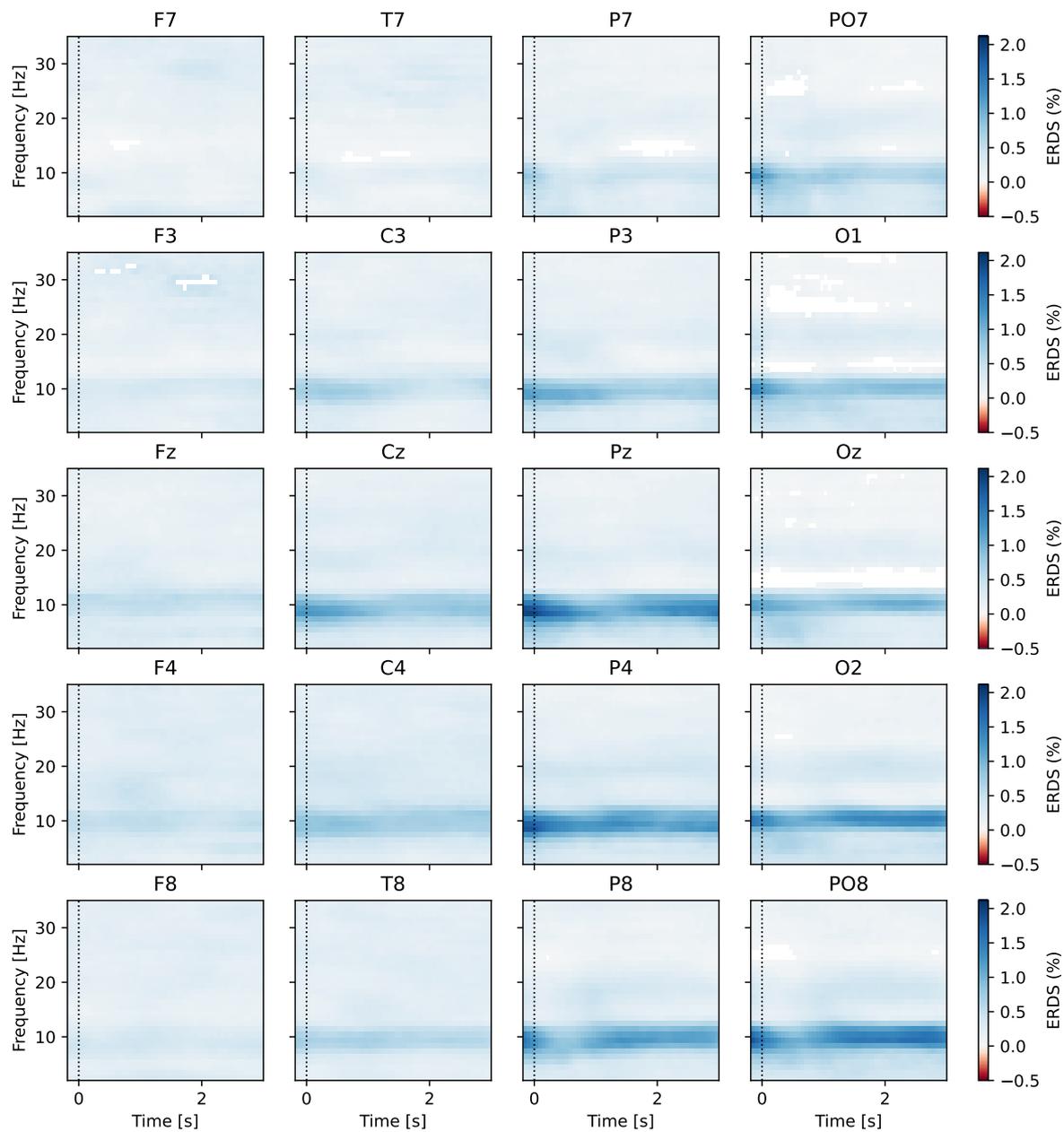


Fig. 4.9 ERD/S maps for the visual imagery task of twenty representative electrodes. X axes are from 200 ms before to 3 seconds after the mental task onset. Y axes show the frequency range of 2–35 Hz. Only statistically significant values are shown (cluster-based permutation test), see Section 4.1.6.

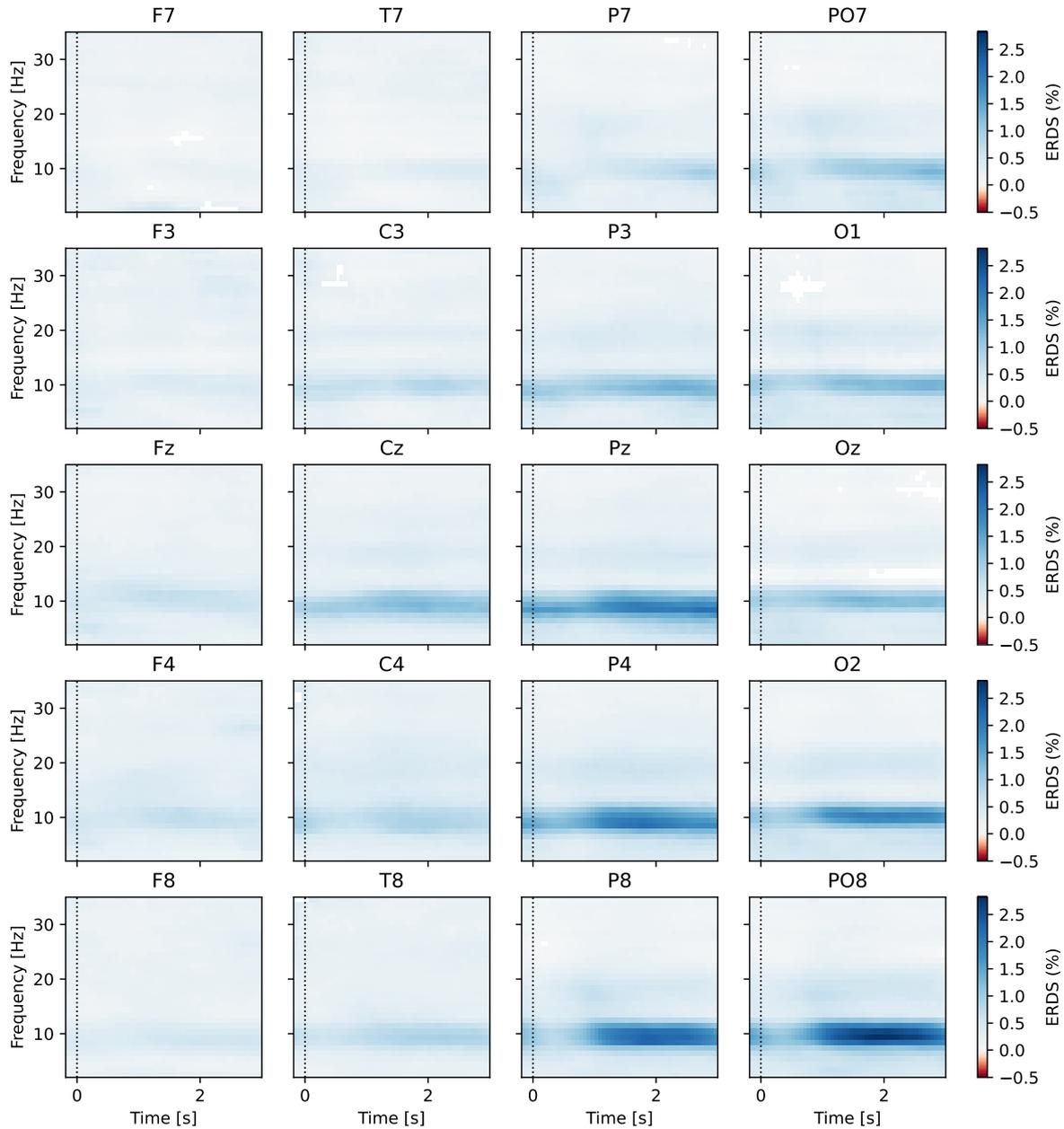


Fig. 4.10 ERD/S maps for the auditory imagery task of twenty representative electrodes. X axes are from 200 ms before to 3 seconds after the mental task onset. Y axes show the frequency range of 2–35 Hz. Only statistically significant values are shown (cluster-based permutation test), see Section 4.1.6.

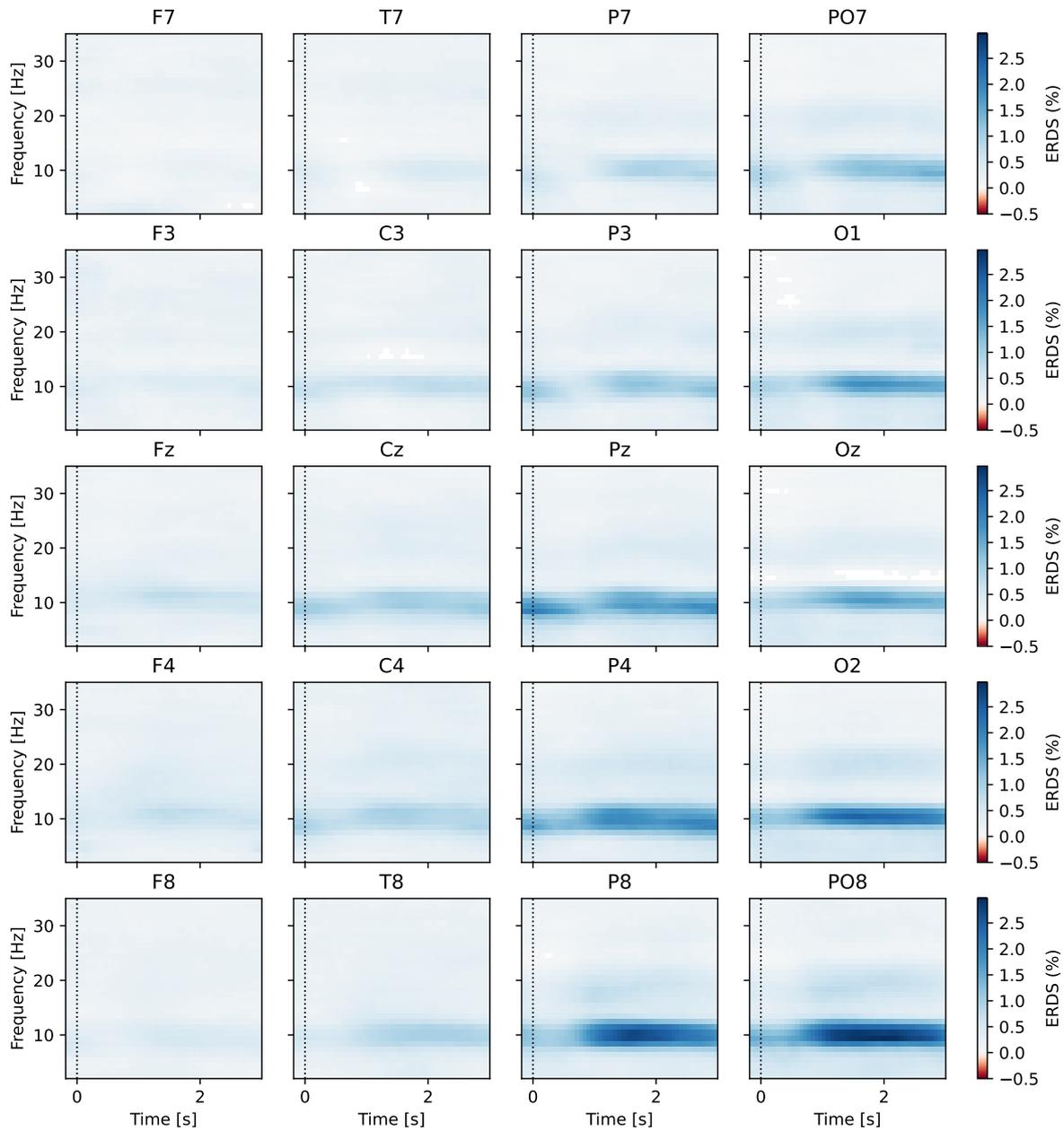


Fig. 4.11 ERD/S maps for the tactile imagery task of twenty representative electrodes. X axes are from 200 ms before to 3 seconds after the mental task onset. Y axes show the frequency range of 2–35 Hz. Only statistically significant values are shown (cluster-based permutation test), see Section 4.1.6.

four mental tasks had visually similar ERD/S maps with the strongest ERD/S activities in parietal and occipital electrodes and partially in central and temporal electrodes in the frequency band of 8–11 Hz.

The silent naming task had the strongest ERD/S in Pz, P2, and CPz with 2.16, 2.15, and 2.08% synchronization increase in comparison with the baseline, respectively. The visual imagery task had the strongest ERD/S in P2, CPz, and Pz with 2.12, 2.01, and 1.96%. The auditory imagery task had the strongest ERD/S in PO8 (2.82%), P2 (2.28%), P8 (2.28%), P4 (2.15%), P10 (2.14%), Pz (2.12%), PO4 (2.06%), and CPz (2.06%). The tactile imagery task had the strongest ERD/S in PO8 (2.97%), P8 (2.71%), PO4 (2.57%), O2 (2.40%), P10 (2.22%), POz (2.21%), P2 (2.11%), Pz (2.05%), and CPz (2.02%).

The strongest ERD/S were in the tactile and auditory imagery tasks both in PO8 (2.97 and 2.82% respectively) and P8 (2.71 and 2.28% respectively). The tactile and auditory imagery tasks had 9 and 8 electrodes with ERD/S above 2%, while the silent naming and visual imagery tasks had only 3 and 2 electrodes. Even though visually similar activities were located across both hemispheres, the strongest ERD/S were primarily right-lateralized.

While it is not possible to precisely locate corresponding brain regions from these results alone due to volume conduction, the strongest ERD/S electrodes were distributed across all modality-specific regions employed here. We did not expect a clear separation between modalities, as discussed in Section 3.1.2 for the imagery mental tasks. The mental task instructions gave freedom to perform the mental tasks in any way that was natural to each participant. Indeed, some participants mentioned that they used visual imagery while performing tactile and auditory imagery. These ERD/S maps are in line with embodiment theories, see Section 2.2.2.1. However, it is not possible to differentiate between secondary, weak, and strong embodiment theories from these EEG results. Therefore, we cannot answer whether or not primary cortical areas for sensory and motor information were activated to distinguish between strong and weak embodiment theories. Note that the embodiment theory review [154] rejected the validity of

strong embodiment outside conscious imagery. Similarly, it is unknown whether or not an amodal hub (or multiple hubs) is used due to visually similar activities across all mental tasks.

4.3 Discussion

Our results suggest the possibility of semantic decoding in EEG, but there is considerable variability in time and frequency locations of semantic information between different participants and mental tasks.

By using the sliding window approach to inspect the temporal evolution of semantic decoding, multivariate classification of all channels (see Section 4.2.1), spatial PCA features (see Section 4.2.2), spatial or spatio-spectral CSP features (see Section 4.2.3), and spatio-spectral PCA features of the signal power (see Section 4.2.4) revealed a variability in the temporal locations where statistically significant classification accuracies occurred between different participants, mental tasks, and frequency bands (where used). Due to this variability in temporal windows, it was not possible to select a single temporal window across all participants that would allow discrimination between the semantic categories for any mental tasks in any of the tested options. This variability was further highlighted by using a grid search over all possible time-frequency windows in the CSP analysis (see Section 4.2.3.3).

In the whole mental task period classification, while it was possible to differentiate between the semantic categories in a few participants, mean classification accuracies, for specific time-frequency windows, across participants were not statistically significant in any analyses. The only exception was the auditory imagery task in the CSP grid search analysis (see Figure C.14 in Section 4.2.3.3). However, this analysis presented the best achievable classification accuracies. The time-frequency windows identified by this analysis must be further assessed in future research.

Semantic decoding might thus be feasible in each mental task for only a subset of participants, with each participant using a different temporal window and sometimes also a different frequency window.

Our results are in contrast with the EEG-based semantic decoding studies identified in Chapter 2. These EEG studies were able to achieve information transfer rates of 0.21–24.09 bits per minute (see Section 2.3.8) with classification accuracies up to 79% for 2 classes [240]. We believe there are several possible interconnected reasons for this difference.

4.3.1 Time variance

EEG-based semantic decoding studies employed a variety of different mental tasks: out-of-category recognition [240, 54], in-category recognition [24], silent naming [173, 172, 246], size judgment [44], semantic judgment [168], memorization of all six elements presented in a sequence [5], and passive listening [246],

In comparison, the silent naming task was used in several studies [173, 172, 246]. For instance, Murphy and colleagues [173] were able to differentiate between the semantic categories of mammals and tools with a mean classification accuracy of 72%, while participants viewed images and performed the silent naming task. Their analysis used a temporal window of 95–360 ms (after the task onset) and a frequency window of 4.1–18.3 Hz to extract epochs that were then spatially filtered by CSP. Two CSP components (in log variance) were then classified by the SVM ($C = 1$). This analysis is similar to our CSP analysis in Section 4.2.3. In contrast, we did not find mean classification accuracies that were better than chance in the silent naming task.

On the other hand, no similar mental tasks to our proposed imagery tasks have been published in studies that made use of EEG. When other neuroimaging modalities were used, the most similar mental tasks were task types based on ‘Properties’, such as silent properties

generation, or ‘Meaning’, such as thinking freely about the meaning of a stimulus, as shown in Table 2.5.

Our experimental design was unfortunately quite complicated in retrospect. Due to the sequence of four different mental tasks, participants were required to quickly change between different mental tasks. With short gaps of only 200 ms between the mental task this may have caused a high degree of variability in the timing of when participants started performing the mental task between different trials. If this is the case, and the inter-trial variability was indeed high, it is not surprising that all tested analyses fail: they were all time-invariant analyses.

All tested analyses in this chapter and even all analyses from the EEG-based semantic decoding studies inherently assumed time-invariant neural activity (for each participant) or with only a slight shift in time. However, our results suggest that time-variant analyses might be required.

4.3.2 Image presentation confound

In contrast with most semantic decoding studies, we attempted to ensure separation of the image presentation period and the mental task period. As discussed in Section 2.3.3.3, this separation avoids the problem of potential processing-related confounds in the classification process that is present in many studies. For instance, focusing on a concept while seeing its image raises the question of what is used for the differentiation between different concepts: the visual processing of the image (low-level perceptual features), the imagination of the concept, or some combination of brain activities related to both processes. This is a pertinent question as it has been shown that the viewed object can be identified from passive viewing of images [192, 48]. Our experimental design mitigated this problem. We will examine this issue in detail in Chapter 5. We will show that all EEG-based semantic decoding studies published to date could exploit the cue presentation in their analyses. Furthermore, we will show that by

including the image presentation period into a classification process it will also be possible to achieve statistically significant mean classification accuracies across participants.

Chapter 5

The effect of cue presentation on EEG-based semantic decoding

Previous semantic decoding studies, that were reviewed in Chapter 2, make use of neural data that include both semantic and perceptual processing. For semantic BCIs, it is important to investigate what happens when these processes are separated in time. Therefore, we ensured a separation between the cue presentation and mental task period in our experiment presented in Chapter 3. This chapter investigates the consequences of this separation in contrast with previous EEG-based semantic decoding studies.

5.1 Introduction

Semantic decoding studies have shown the possibility of semantic neural decoding. Most of these semantic decoding studies use experiments that are designed to cue participants to focus their attention on instances of particular semantic concepts for short periods of time, for example by presenting visual or auditory stimuli. Neural activity is recorded while participants perform mental tasks involving the cued semantic concepts, for example, by silently naming

the instance of the concept. This recorded activity is then processed to attempt to differentiate between semantic concepts.

An important difference between semantic BCIs and the majority of semantic decoding studies is that in semantic BCIs, users would freely choose and focus on a semantic concept of their choice (from a supported set of recognizable concepts). In other words, in a semantic BCI, there would not be an external cue that is used to drive participant's particular choice.

Many semantic decoding studies employ experimental designs where the cue is present (e.g., as an image on screen) *while* participants perform mental tasks, other studies do not, but they differentiate between different semantic concepts using neural activity that was recorded during the cue presentation together with activity recorded during the mental task. In these case, semantic decoders might exploit perceptual processing related neural activity to increase their decoding performance.

Note that this is a general issue applicable to all neuroimaging modalities including fMRI with which the most promising results of semantic neural decoding have been obtained. However, we primary focus on EEG here. Table 5.1 shows a subset of semantic decoding studies from Chapter 2 that used EEG (alone or in combination with other neuroimaging modalities). If "Cue present" column is "Yes", there is no clear separation between the cue and the task. As indicated, semantic decoders in all of these EEG-based semantic decoding studies could exploit the cue presentation in one way or another.

In this chapter, we show that decoding of purely imagined concepts without any external cue, that is when semantic decoders can use only semantic activation caused by concept related mental tasks, may pose a more challenging problem with respect to the current state-of-the-art suggesting that semantic decoding is possible. There are at least two (interconnected) reasons for this: (1) an exploitation of perceptual processing by semantic decoders when the cue is presented, and (2) a lower degree of semantic activation caused only by the concept related

Table 5.1 Semantic decoding studies from Chapter 2 using EEG. The column ‘Pres. Mod.’ represents presentation modalities used: visual (V) (image), auditory (A) (spoken word), or orthographical (O) (written word). Multiple letters indicate that multiple modalities were tested separately, while V+O is an image with a corresponding written name. The column ‘Cue present’ refers to whether or not the cue is present in temporal window used by the semantic decoding model for classification.

Temporal features	Spectral features	Task	Pres. Mod.	Cue present	Ref.
0-700ms	1-30Hz	out-of-category recognition	VAO	Yes	[240]
95-360ms	4.1-18.3Hz	silent naming task	V	Yes	[173, 172]
200-700ms (category), 250-500ms (individual words)	1-30Hz	size judgment	AO	Yes	[44]
0-1.2s	optimized for each participant (0.5-55Hz)	passive listening, silent naming task	AO	Yes	[246]
0-1s	0.1-40Hz	remember all six elements presented in a sequence	O	Yes	[5]
0-1s	1-12Hz	out-of-category recognition	A	Yes	[54]
0-700ms	1-30Hz	in-category recognition	V	Yes	[24]
0-3500ms	2-100Hz	semantic judgment	V+O	Yes	[168]

mental tasks (when performing pure imagination or recollection of semantic concepts). We focus on the first issue here.

That this is an issue is further supported by several studies that explored experimental designs with different stimulus presentation modalities and reported significantly different decoding accuracies for each stimulus modality, even though it has been shown that semantic processing is independent of stimulus presentation modality [191, 271]. For instance, an EEG study by Simanova and colleagues [240] reported a higher mean classification accuracy when participants were cued by images in comparison with when they were cued by spoken or written words.

For these reasons, we argue that, when attempting to build and evaluate semantic decoding models with the focus towards semantic BCIs, the cue presentation should be separated from the mental task period. We designed an experiment, presented in Chapter 3, with this separation between the cue presentation period (i.e., an image presentation) and the mental task period. We will explore the effect of cue presentation on EEG-based semantic decoding by applying the same analyses during the cue presentation period and during different mental tasks attempted after the presentation of the cue. When using recorded neural activity only from the mental task period, semantic decoders cannot use perceptual processing from the cue presentation period, but there may still be some leftover from semantic activation caused by perceiving the cue (and, of course, there is semantic activation caused by the concept related mental tasks).

5.2 Methods

This chapter only focuses on EEG data from our experiment presented in Chapter 3. To investigate the effect of the cue presentation period on differentiating between the semantic categories of animals and tools, we present six analyses that were conducted on cue presentation periods in addition to the mental task periods. The EEG data were preprocessed as described in Section 4.1.1 in Chapter 4.

5.2.1 Analysis in the temporal domain

To explore the possibility of differentiating between the semantic categories of animals and tools, we used the most prevalent analysis approach from the prior EEG-based semantic decoding studies in Table 5.1 which is an analysis in the temporal domain.

This was conducted in the two most widely used frequency bands (see Table 5.1): 1–30 and 4–30 Hz. In both cases, the preprocessed EEG data were first high-pass filtered by an IIR 4th-order Butterworth filter (1 or 4 Hz cutoff, forward-backward filtering, `sosfiltfilt`

method) and then convolved with a FIR low-pass filter (designed by the `firwin` method from SciPy [263], 0.661 s length, 30 Hz cutoff, 5 Hz bandwidth) [275]. Lastly, the filtered data were downsampled (decimated) to 64 Hz and restructured into epochs (excluding data from the excluded concept trials).

To investigate the effect of the cue (image) presentation period on differentiating between the semantic categories, all the following analyses were also conducted on concept trials in addition to different mental task types.

5.2.2 Analysis 1 (single channel classification)

In order to investigate the temporal evolution of semantic decoding, a classifier was trained in a stratified 15-fold cross-validation over a sliding temporal window for each channel separately. The temporal window of 109.375 ms was used (7 time points at a sampling frequency of 64 Hz) with a step of half the window size. Data in each window were normalized (z-scored) for each channel separately, based on training folds during cross-validation.

We chose a SVM (from scikit-learn [195], default settings, $C = 1$) because it was the most used classifier in semantic decoding studies (see Table 2.7 in Section 2.3.6). We also tested other classifiers: linear regression with L1 or L2 norm, naive Bayes classifier, and linear discriminant analysis. As before, even though there were several (minor) differences, the main message is the same and thus we report only the results achieved with the SVM.

5.2.3 Analysis 2 (all channels classification)

Next, the same analysis as before was used with the only difference that all channels instead of a single channel were used in the sliding window approach detailed above. As before, each channel in a particular temporal window was separately normalized (z-scored).

In this approach, the number of features provided to the classifier was the size of the temporal window (7 time points) multiplied by the number of channels recorded from a

particular participant (from 48 to 64 channels, mean 59.18) resulting in up to $7 \times 64 = 448$ features. To avoid having more features than samples (and the curse of dimensionality), we also re-ran this analysis but instead of using all time points from each channel, only mean and standard deviation of each channel in that temporal window were used as features. This approach reduced the maximal number of features to $2 \times 64 = 128$. However, results from those two approaches were (qualitatively) similar so we report only the results using all time points.

5.2.4 Analysis 3 (selecting an optimal temporal window)

We tested identifying an optimal temporal window for classification using a nested-cross-validation approach for each channel separately. The analysis is the same as before with the only difference that instead of using a sliding temporal window, an optimal temporal window was selected (for each channel separately) by the maximal mean classification accuracy from all temporal windows using inner stratified 10-fold cross-validation (on the training folds from the outer cross-validation).

5.2.5 Analysis 4 (full period classification)

To take into consideration the full task period, all time points from a task period were used for classification, instead of using only time points from a particular temporal window as before. This was done using a single channel or all channels. Additionally, the mental task period was concatenated with the image presentation period to investigate the effect of cue presentation on classification accuracy. Note that the number of features for the classifier depends on a particular period length.

5.2.6 Analyses 5 and 6 (replication studies)

To further stress the need for separating the cue presentation period and the mental task period, we replicated analyses from two studies that used an image stimulus as a cue during the mental task period (from Table 5.1). This was done not to criticize these studies but to show the potential issue with attempting semantic category decoding using the neural data recorded during the cue presentation period.

In a study by Murphy and colleagues [173], participants were presented with images of mammals and tools. They were asked to silently name the shown object and to press a mouse button when finished (with a timeout of 3 seconds). The image was shown for the whole task duration. Their analysis used a temporal window of 95–360 ms (after the task (or image) onset) and a frequency window of 4.1–18.3 Hz to extract epochs that were then spatially filtered by common spatial patterns (CSP) [169, 32]. A subset of N CSP components (in log variance) were classified by an SVM ($C = 1$). They were able to achieve a mean classification accuracy of 72% (with $N = 2$).

We replicated their analysis in the following pipeline. The preprocessed EEG data were filtered and all epochs were extracted. We used a dimensionality reduction by principal component analysis (PCA) before the CSP to avoid the issue of decreased dimensionality of EEG signals after ICA for eye blinks suppression during the preprocessing phase [222], see Appendix B for more information. The number of dimensions was decreased by one because one IC was removed (see Section 4.1.1). Then this data were spatially filtered by CSP and a subset of N CSP components was extracted. The first half of N selected CSP components had the largest eigenvalues and the second half had the smallest eigenvalues, which is typical in CSP applications (see Section B.2 in Appendix B). The log variance of the selected CSP components was classified by an SVM ($C=1$). We tested $N \in \{2, 4, 6, 8, 10\}$.

In a study by Simanova and colleagues [240], participants were presented with either images (line drawings), written words, or spoken words of animals and tools. They were asked

to respond, by pressing a mouse button, upon appearance of items from the non-target task categories (clothing or vegetables). In the case of image stimulus, the image was shown for 300 ms followed by a blank screen for 1000–1200 ms. However, their analysis used a temporal window of 0–700 ms after the stimulus onset which includes the cue presentation period. The filtered signals, in the frequency band 1–30 Hz, from all channels were classified by logistic regression (LR) (Bayesian logistic regression with a multivariate Laplace prior). They achieved a mean classification accuracy of 79% when participants were cued by images.

We replicated their analysis in the following pipeline. The preprocessed EEG data were filtered and all epochs were extracted. Each channel was separately normalized (z-scored). A classifier used all channels for the classification. We tested two different classifiers. The first classifier was LR with L1 norm (from scikit-learn [195], default settings but with a maximum number of iterations of 10^5), which is similar to the approach of [240] (but without explicitly using a Bayesian approach). The second classifier was an SVM ($C=1$) because it was used in all other analyses. Additionally, both classifiers were also trained using nested-cross-validation (with inner stratified 10-fold cross-validation) to choose an appropriate parameter C in case of SVM and regularization strength in case of LR. Because our image presentation was only 600 ms, in comparison with their 700 ms, a temporal window of 600 ms was used for all classified periods except for the blank screen period before the first task which was only 500 ms. Note that this analysis is similar to Analysis 4, in Section 5.2.5, using all channels, but it uses only 600 ms after the mental task onset in comparison with the full 3 seconds in Analysis 4.

5.3 Results

5.3.1 Analysis 1 (single channel classification)

Analysis 1 revealed considerable variability in the temporal locations of significant classification accuracies ($p < 0.05$, one-sided Binomial test) between different participants for each mental

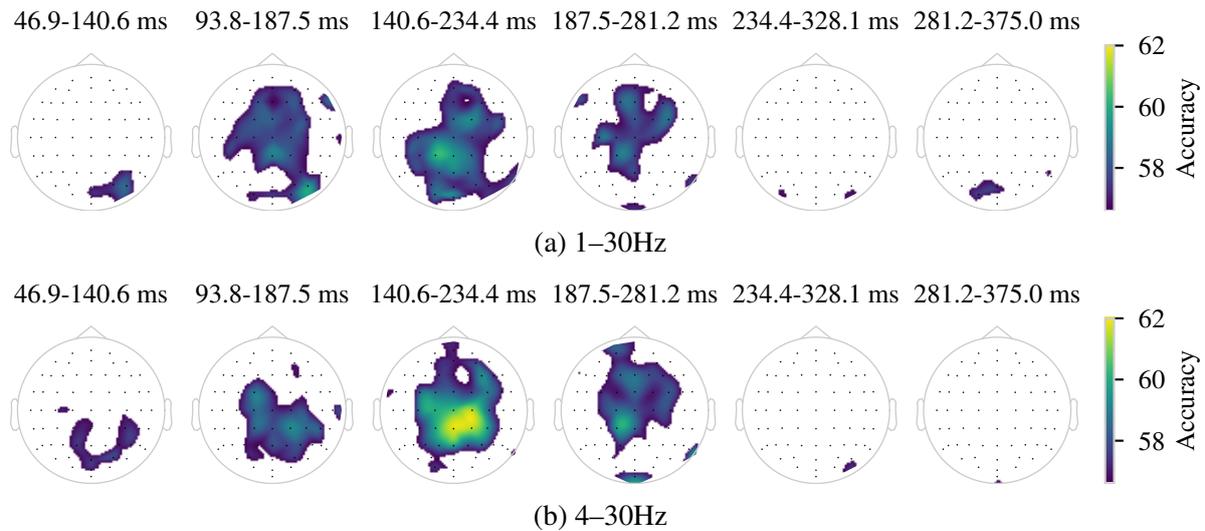


Fig. 5.1 Mean classification accuracies during the image presentation period for each channel (Analysis 1). Times represent the start and the end of each temporal window after the image onset. Scalp maps indicate performance above the significance borderline ($p = 0.05$, 56.6%). White represents non-significant classification accuracies.

task. Due to this variability, it was not possible to select a temporal window across all participants that would allow discrimination between the semantic categories for any mental task. Figures D.1 and D.2 in Appendix D show this variability across all participants.

On the other hand, it was possible to differentiate between the semantic categories during the image presentation period across all participants. Figure 5.1 shows scalp maps indicating for which channels mean classification accuracies were statistically significant during the image presentation period. Figures D.3 and D.4 in Appendix D show the variability across all participants. The peak of the mean classification accuracies across all participants was in the temporal window of 140.6–234.3 ms after the image onset in both frequency bands, see Figure D.5 in Appendix D for variability in classification accuracies across participants in this temporal window.

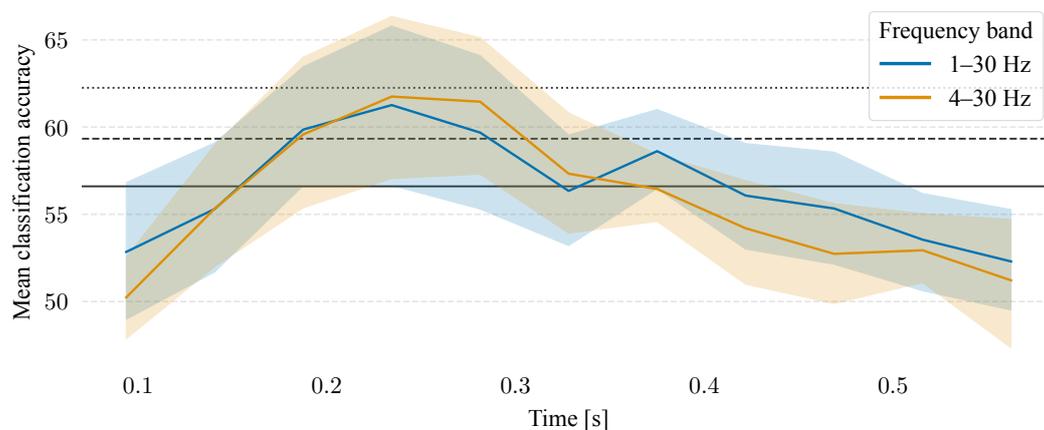


Fig. 5.2 Mean classification accuracies during the image presentation period when the classifier can use information from all channels (Analysis 2). Classification accuracies are shown with mean and 95% confidence interval. Horizontal lines represent significance borderlines for $p = 0.05$ (56.6%, solid), $p = 0.01$ (59.33%, dashed), and $p = 0.001$ (62.25%, dotted).

5.3.2 Analysis 2 (all channels classification)

Similarly to the results of the single channel classification in Analysis 1, there was variability across participants in the temporal locations of significant classification accuracies in the mental tasks in comparison with the temporal locations in the image presentation period. Due to this variability, it was not possible to select a temporal window across all participants that would allow discrimination between the semantic categories for any mental task.

On the other hand, it was possible to differentiate between the semantic categories during the image presentation period across all participants. Figure 5.2 shows mean classification accuracies during the image presentation period. The peak mean classification accuracy was 61%, again, in the temporal window of 140.6–234.3 ms after the image onset in both frequency bands. It was not possible to select a temporal window across all participants to discriminate between the semantic categories for any mental task, or for other periods within the trial (mask presentation or blank screen). Figure D.6 in Appendix D shows the variability across all participants.

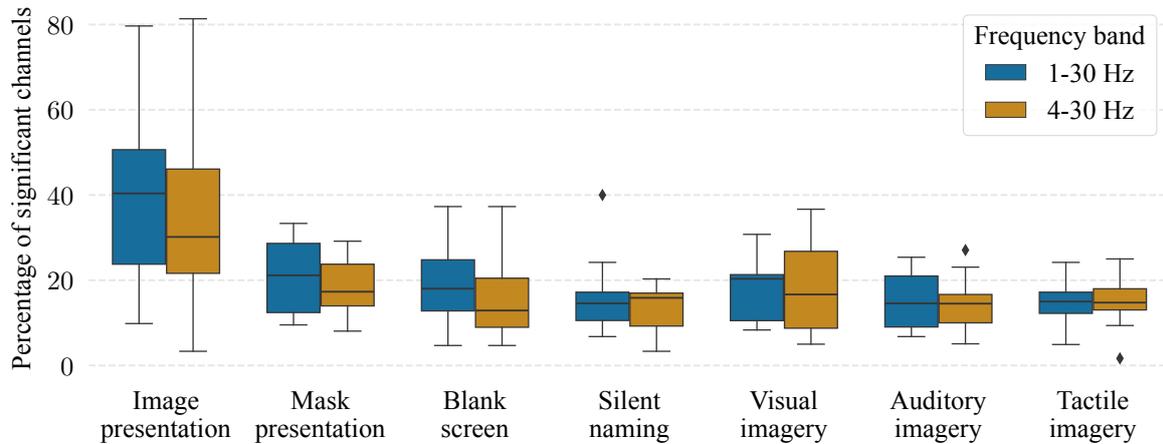


Fig. 5.3 Percentages of channels with significant classification accuracies ($p < 0.05$) for each participant using the selected optimal temporal windows during the nested-cross-validation (Analysis 3).

5.3.3 Analysis 3 (selecting an optimal temporal window)

The feasibility of selecting an optimal temporal window for each channel separately was assessed by a nested-cross-validation approach. Figure 5.3 shows the percentage of the EEG channels with significant classification accuracies ($p < 0.05$). Significantly more channels were found to contain information that allowed significant differentiation of the semantic categories during the image presentation period than during any of the other tasks ($p < 0.01$ except the visual and tactile imagery tasks in the 4–30 Hz frequency band with $p < 0.05$, one-sided Wilcoxon signed-rank test). Figures D.7 and D.8 in Appendix D show scalp maps indicating for which channels classification accuracies were statistically significant.

5.3.4 Analysis 4 (full period classification)

In a single channel classification, Figure 5.4 shows the percentage of the EEG channels with significant classification accuracies ($p < 0.05$). Markedly more channels had significant classification accuracies during the image presentation than during any of the other tasks ($p < 0.01$, one-sided Wilcoxon signed-rank test). The concatenation of the mental-task and

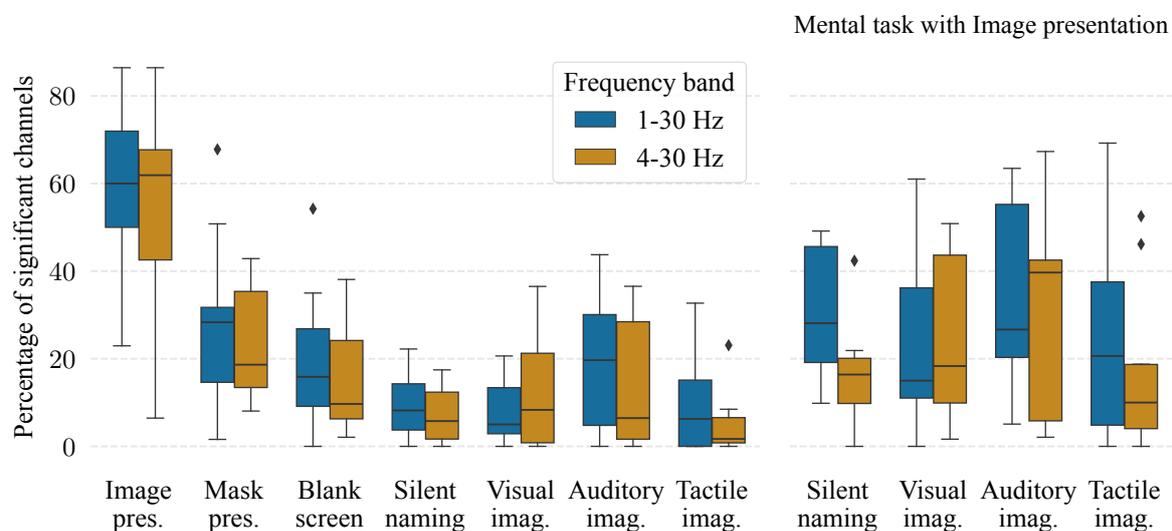


Fig. 5.4 Percentages of channels with significant classification accuracies ($p < 0.05$) for each participant using all time points from the corresponding period(s) (Analysis 4).

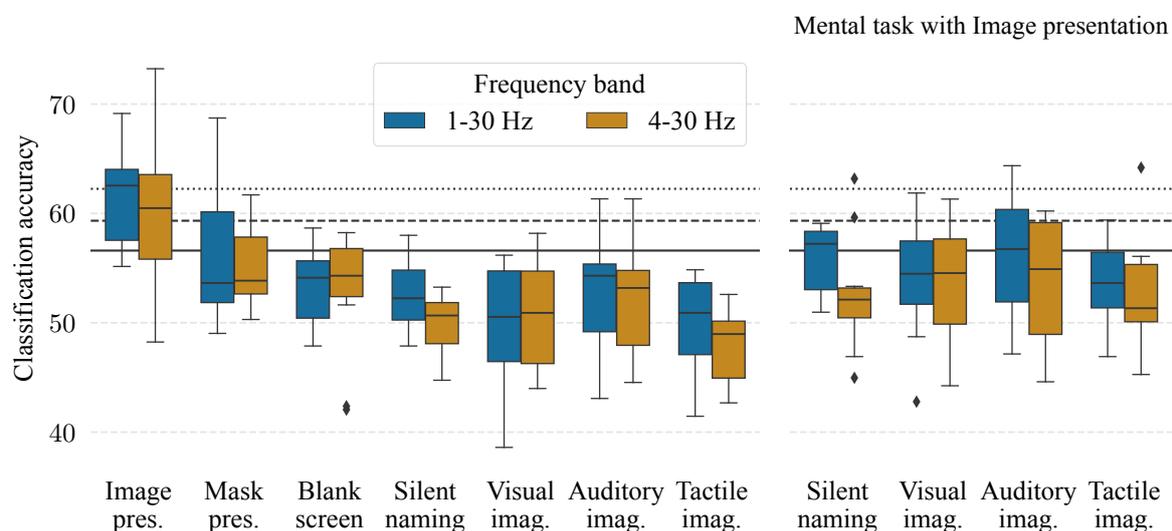


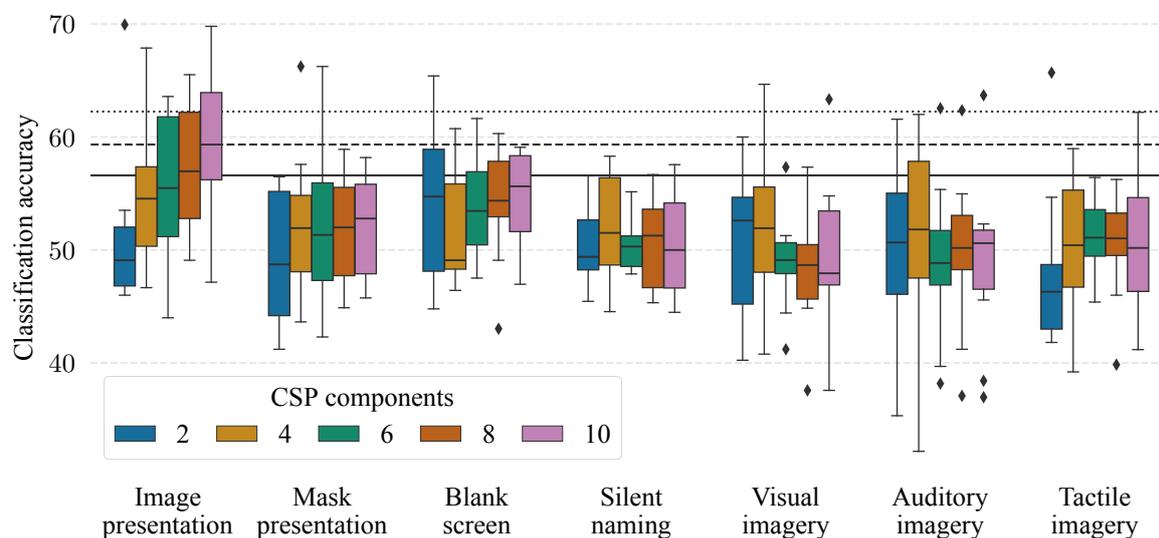
Fig. 5.5 Classification accuracies of all channels and all time points from corresponding period(s) (Analysis 4). Horizontal lines represent significance borderlines for $p = 0.05$ (56.6%, solid), $p = 0.01$ (59.33%, dashed), and $p = 0.001$ (62.25%, dotted).

the image-presentation periods increased the number of channels with significant classification accuracies in comparison with the mental-task period alone ($p < 0.01$ except the silent naming task in 4–30 Hz with $p < 0.05$). On the other hand, the concatenation of the mental-task with either mask-presentation or blank-screen period did not significantly increase the number of channels with significant accuracies.

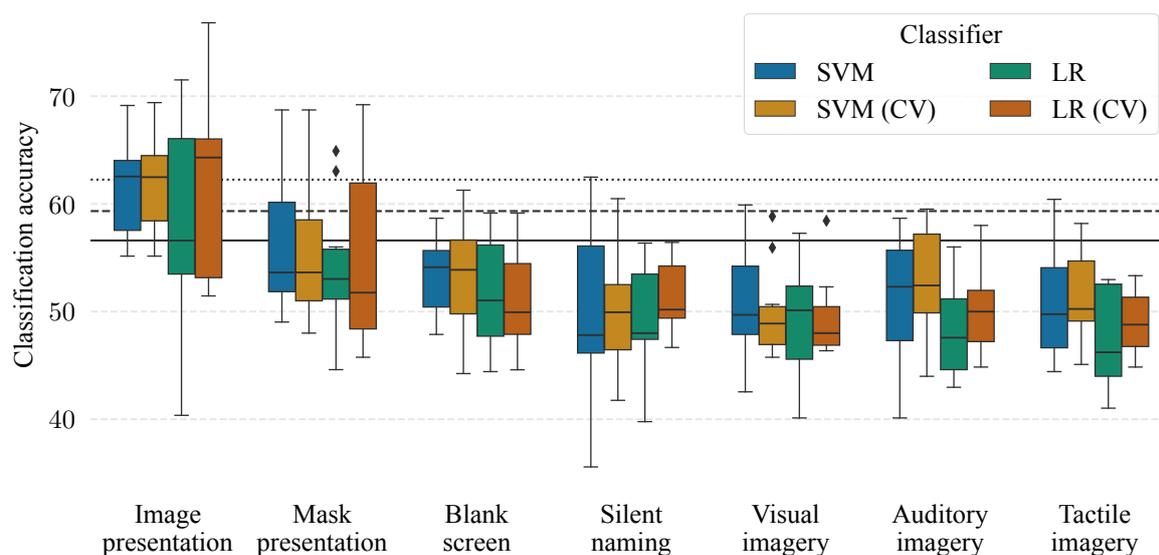
Figure 5.5 shows classification accuracies when the classifier can use information from all channels. In the frequency band of 1–30 Hz, it was possible to significantly differentiate between the semantic categories during the image presentation period in 9 out of 11 participants. While, for instance, the silent naming task had significant classification accuracies only in 1 participant but in 7 participants when it was concatenated with the image presentation period. Similarly, in the frequency band 1–40 Hz, significant semantic decoding was possible in 8 participants during the image presentation period. Overall, classification accuracies for the image presentation period were significantly greater than in any other task ($p < 0.001$, one-sided Wilcoxon signed-rank test). The concatenation of the mental-task and the image-presentation period significantly increased classification accuracies in comparison with the mental-task period alone ($p < 0.01$ except the silent naming and auditory imagery task in 4–30 Hz with $p < 0.05$).

5.3.5 Analyses 5 and 6 (replication studies)

Figure 5.6(a) shows the results of Analysis 5 replicating the analysis from a study by Murphy and colleagues [173], which used CSP. It was possible to significantly differentiate between the semantic categories during the image presentation period in 7 out of 11 participants with 10 CSP components, in 6 participants with 8 components, in 5 participants with 6 or 4 components, and in none with 2 components. Classification accuracies were significantly greater in the image presentation period in comparison with the mental tasks with 10 ($p < 10^{-6}$, one-sided Wilcoxon signed-rank test), 8 ($p < 0.001$), and 6 ($p < 0.01$) CSP components, while with 4



(a) Analysis 5 using CSP.



(b) Analysis 6 using all channels and a 600 ms period. CV indicates using nested-cross-validation to choose a classifier parameter value.

Fig. 5.6 Classification accuracies from the studies replication. Horizontal lines represent significance borderlines for $p = 0.05$ (56.6%, solid), $p = 0.01$ (59.33%, dashed), and $p = 0.001$ (62.25%, dotted).

CSP components for all mental tasks except the visual imagery task ($p < 0.05$), and only for the tactile imagery task with 2 CSP components ($p < 0.05$).

Figure 5.6(b) shows the results of Analysis 6 based on a study by Simanova and colleagues [240], which uses all channels and a 600 ms period. It was possible to significantly differentiate between the semantic categories during the image presentation period in 9 out of 11 participants with the SVM ($C = 1$ and C set by the nested-cross-validation), in 5 participants with LR, and in 6 participants with LR with regularization strength chosen by the nested-cross-validation. Classification accuracies for the image presentation period were significantly greater in comparison with the mental tasks for both SVMs ($p < 10^{-7}$) and both LRs ($p < 10^{-6}$ and $p < 10^{-7}$ with the nested-cross-validation).

Overall, both analyses revealed the possibility of differentiating between the semantic categories from only the image presentation period.

5.4 Discussion

Our results indicate a potential issue when the semantic decoder is allowed to use neural activity recorded in the cue presentation period. We used six analyses on the mental tasks and the beginnings of concept trials, which include the image presentation, mask presentation, and blank screen periods.

The results of the temporal evolution of classification accuracies by using a single channel in Analysis 1 (see Section 5.3.1) or all channels in Analysis 2 (see Section 5.3.2) suggest that semantic decoding might be feasible in each mental task for only a subset of participants, with each participant utilizing a different temporal window. On the other hand, neural activity recorded during the image presentation period allowed for significant differentiation between the semantic categories, especially in the temporal window of 140–234 ms after the image onset *across all participants*.

To account for the variability in temporal locations of significant accuracies from Analysis 1 in the mental tasks, the results of selecting an appropriate temporal window in Analysis 3 (see Section 5.3.3) suggest that the semantic decoding in the mental tasks and image presentation period might be feasible with an appropriate subset of channels. However, there were markedly *more channels that allowed significant differentiation of the semantic categories during the image presentation period than in any other tasks.*

The presence of this issue was also corroborated by Analysis 4 (see Section 5.3.4), which used all time points in a particular task period. Markedly more channels allowed significant differentiation between the semantic categories from the image presentation period than any other task. Furthermore, classification of neural activity from the mental task with the addition of neural activity from the image presentation period markedly increased the number of channels that allowed significant differentiation. This is in spite of possible overfitting because concatenated neural activity is longer and thus presents more features to the classifier in comparison with only the image presentation period. Additionally, the classifier using all channels allowed *semantic decoding in most participants for the image presentation period.*

Finally, replicated analyses from two studies (see Section 5.3.5) also allowed significant discrimination between the semantic categories for the image presentation period but with much less success for the mental task periods of the trials.

Overall, it was easier to differentiate between the semantic categories of animals and tools during the image presentation period than during the mental tasks (by all of the analyses presented in this chapter). Semantic decoding may thus be more challenging when the classifier is not allowed to exploit the perceptual processing related neural activity from the cue presentation period.

If the image presentation period had not been separated from the mental task period or if neural activity recorded during the image presentation period had been included with neural activity recorded during the mental task period, we could have claimed the possibility

of semantic decoding with our presented analyses. Mean classification accuracies across participants for the image presentation period were up to $61.6 \pm 4.9\%$ (mean \pm std) in Analysis 4, up to $59.5 \pm 6.1\%$ in Analysis 5, and up to $61.8 \pm 4.9\%$ in Analysis 6. While, mean classification accuracies for any of the mental tasks were below 56.6% in Analysis 4, 5, and 6 and were significantly lower in comparison with the image presentation classification accuracies.

Only a few semantic decoding studies, reviewed in Chapter 2, attempted semantic decoding during a free recall paradigm with minimal cue support [239, 168, 206]. During free recall, participants are often instructed to recall semantically related items consecutively. This is the most similar paradigm to our experimental design. For instance, a fMRI study [239] presented participants with semantic categories of animals and tools in four modalities: spoken words, written words, images, and natural sounds. Participants were asked to respond upon appearance to out-of-category exemplars. Additionally, the experiment ended with two free recall blocks in which participants were presented with a category name for 2 seconds and then instructed to recall all the entities from the cued category, seen during the experiment, in a 40 second period. A classifier was trained on the data from actual stimulus presentation and then tested on the recall data. Mean classification accuracies across 14 participants were significant only when the classifier was trained on a combination of all modalities (67%) and partly when trained on a natural sounds modality (65%, $p = 0.01$, but below the p -value threshold for statistical significance used by the study). In general, studies that examined neural activity during free recall showed the rise and fall of the category-specific neural activity with different effects in different neuroimaging modalities [168, 206, 234, 80, 142, 147, 205].

This effect of cue presentation on semantic decoding is also likely to be present in state of the art results from semantic decoding studies using neuroimaging modalities affected by a slow hemodynamic delay, such as fMRI and fNIRS. The effect of perceptual processing induced by the cue presentation on the recorded neural activity might be *present for a longer period*

due to the hemodynamic delay. This recorded neural activity is then more easily exploited by a classifier for differentiating between the semantic concepts.

A possible solution to partly avoid this issue, when the experimental design uses the cue presentation, is to exclude brain areas that are involved in perceptual processing from the analysis, for example, visual areas that are used for visual processing. However, while this approach is applicable to neuroimaging modalities with high degrees of spatial specificity, such as fMRI or ECoG, it is not very viable in neuroimaging modalities with low spatial resolution, such as EEG.

Chapter 6

Semantic decoding in fNIRS

This chapter investigates whether or not semantic decoding is possible solely with the fNIRS data recorded from the experiment presented in Chapter 3.

6.1 Methods

6.1.1 fNIRS data preprocessing

6.1.1.1 Bad channels removal

Bad channels were identified and removed based on the scalp coupling index (SCI) [204]. The main visible artifact in the fNIRS signal is due to the cardiac cycle, which is approximately 1 Hz. Its presence is related to intracranial physiological parameters and indicates a good contact between the optical probe and the scalp [251, 250].

To compute an SCI, raw fNIRS data were band-pass filtered between 0.5 and 2.5 Hz (4th-order Butterworth filter, forward-backward filtering by the `sosfiltfilt` method from SciPy [263]) and z-scored. For each channel, the SCI was computed as the Pearson correlation coefficient between the fNIRS signals of the two wavelengths, 785 and 830 nm, defining the channel. In-phase and counter-phase identical waveforms have an SCI of 1 and -1 , respectively,

Table 6.1 Number of bad channels in fNIRS recordings for each participant. The frontal montage contains 14 channels in total, while the temporal montage contains 11 channels in total.

Participant	Montage	Bad fNIRS channels
1	frontal	2
2	frontal	2
3	frontal	1
4	frontal	7
5	frontal	excluded
6	frontal	5
7	temporal	3
8	temporal	1
9	temporal	6
10	temporal	1
11	temporal	0
12	temporal	excluded

whereas uncorrelated signals have an SCI of 0. Channels with a mean SCI across all blocks below 0.75 were removed (i.e., $-1 \leq \text{SCI} < 0.75$). This threshold was suggested in [204].

Two participants (participants 5 and 12) were excluded from further fNIRS analysis because they had all channels removed. Table 6.1 shows the number of excluded channels for each participant.

6.1.1.2 Motion correction

Motion artifacts caused by relative motion between fNIRS optical fibers and the scalp were corrected with a wavelet transform [165, 53, 37] using the `hmrMotionCorrectWavelet` method from the Homer2 software package [105] with an interquartile range of 1.5, as suggested in the function's documentation.

6.1.1.3 Changes in oxy/deoxy-generated hemoglobin

Motion-corrected fNIRS signals with a wavelength of 785 and 830 nm were converted to changes in concentration of oxygenated and deoxygenated hemoglobin (in units M (molar

concentration)) using the modified Beer-Lambert law [68] (see Section A.2 in Appendix A for more information) with differential path length factors (unitless), $DPF^{785} = 7.25$ and $DPF^{830} = 6.38$, and molar absorption coefficients (in $M^{-1} cm^{-1}$) $\alpha_{HbO_2}^{785} = 1798.643$, $\alpha_{Hb}^{785} = 2295.285$, $\alpha_{HbO_2}^{830} = 2321.424$ and $\alpha_{Hb}^{830} = 1791.734$.

6.1.1.4 Filtering

The extracted changes in oxygenated and deoxygenated hemoglobin were filtered between 0.01 and 0.7 Hz by two filters. This choice was based on recommendations from the literature [201, 177, 231]. First, an IIR high-pass 9th-order Butterworth filter with 0.01 Hz cutoff reduced very low-frequency oscillations which are below 0.04 Hz (forward-backward filtering which creates an effective order of 18, `sosfiltfilt` method from SciPy [263] using second-order sections to reduce numerical errors). Then, a FIR low-pass filter with 0.7 Hz cutoff designed by the `firwin` method from SciPy [263] reduced the influence from heart rate which is around 1 Hz. The filter length was $\lfloor 42f_s \rfloor$ (374 for the temporal montage, 328 for the frontal montage) because 21 seconds was the shortest length after the last block before the end of the experiment from all participants. The usage of an IIR filter followed by a FIR filter was based on practical experience and suggested in [275].

6.1.2 Single concept trial analysis

Chapter 5 described the influence of image presentation on semantic decoding in EEG. This issue may be even worse in fNIRS. The effect of the image presentation period is present in fNIRS signals for a long time due to the hemodynamic delay. We thus started with an exploration of the whole concept trial period, which starts from the image presentation onset and includes the sequence of four mental tasks. In effect, this approach disregards different mental task types.

We explored a single concept trial analysis of two types of signals. First, the preprocessed data were used. We will refer to this data as ‘no-GLM’. Second, the preprocessed fNIRS signals were modeled via a general linear model [77, 47] to remove noise and influences from previous concept trials. We will refer to this data as ‘GLM’. The GLM approach will be described in detail in Section 6.1.3. Epochs of concept trials were then extracted and further processed in the same way for both approaches.

Epochs were extracted from 1 second before the image presentation (i.e., the last 1 second of the fixation cross) until 20 seconds after the image presentation. The last mental task ends 14.3 seconds after the image onset. So, this 20 seconds period thus contains 8.7 seconds of the hemodynamic response of the last mental task.

Epochs were further preprocessed before any analysis by using: (1) linear detrending by subtracting each epoch’s least squares fit and (2) baseline correction by subtracting the mean of the 1 second period before the image presentation. All the following analyses were tested: (1) no detrending and no baseline correction, (2) detrending but no baseline correction, (3) no detrending but baseline correction, and (4) detrending and baseline correction. Although, there were differences between each setting, the overall message was the same for all of them. Only one setting is thus reported here to simplify the presentation in which only the detrending (option (2) above) is used.

Finally, data were downsampled by a factor of 4 that is from 7.81 to 1.9525 Hz for the frontal montage and from 8.92 to 2.23 Hz for the temporal montage.

6.1.3 GLM approach

Each fNIRS channel (i.e., changes of concentration in oxygenated or deoxygenated hemoglobin) was modeled by a general linear model (GLM) [77, 47] to reduce noise and influences from preceding mental tasks. This is important due to the short gap (200 ms) between mental tasks.

Since fNIRS records brain activity in a similar way to fMRI, the GLM approach adopted for fMRI [77] can be applied to fNIRS [232, 202, 122, 280].

A single fNIRS channel can be modeled by the GLM as follows:

$$y = X\beta + \varepsilon \quad (6.1)$$

$$\varepsilon \sim \mathcal{N}(0, \sigma^2 I_N) \quad (6.2)$$

where $y \in \mathbb{R}^N$ represents N time samples of the fNIRS signal, $X \in \mathbb{R}^{N \times P}$ is a design matrix, P is the number of parameters, $\beta \in \mathbb{R}^P$ are unknown parameters, ε are normally distributed error terms with zero mean and variance σ^2 , and $I_N \in \mathbb{R}^{N \times N}$ is the identity matrix. The ordinary least squares estimates of β are given by

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (6.3)$$

6.1.3.1 Design matrix

The design matrix X models the experiment from the beginning of the first block until the end of the last block. Each row represents a different time step in the experiment. Columns represent regressors modeling the experiment. The first 90 regressors are the result of the convolution of 10 conditions modeled in the experiment and a set of 9 temporal basis functions.

We modeled the following 10 conditions: four conditions represent the execution of 4 mental tasks for the animal category, four conditions represent the execution of 4 mental tasks for the tools category, one condition represents the presentation of an image, and one condition represents the mask. An element of a particular condition is 1 if the condition was taking place at the time represented by the element. All other elements are 0. Figure 6.1 illustrates this process. Conditions are represented as long sequences on the left. One convolved condition is shown in the middle of the figure. All convolved conditions are placed in the columns of the design matrix X in the right of the figure.

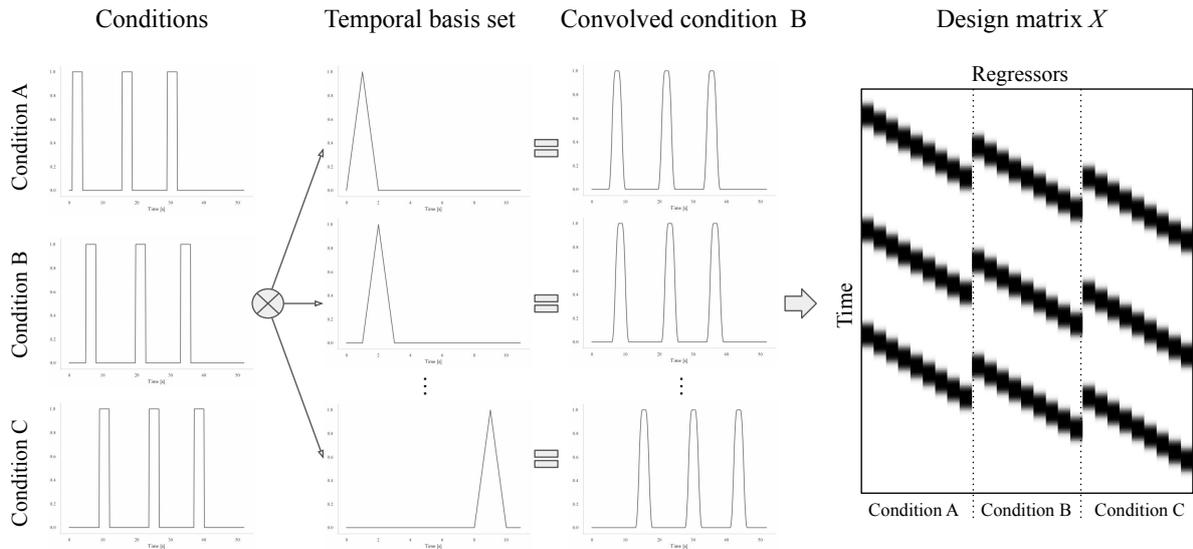


Fig. 6.1 Illustration of the process of creating the design matrix X with three conditions. Condition B (e.g., visual imagery task) is separately convolved by each function from the temporal basis set. The results are regressors placed in the corresponding columns of X .

We used a temporal basis set based on the B-spline functions (see Figure 6.1). The first function is non-zero from 0 to 2 seconds with a peak at 1 second. The following function is shifted by 1 second and so on. This basis set is flexible enough to model different hemodynamic responses, with different timing between oxygenated and deoxygenated hemoglobin [104]. The set contains 9 functions with the last peak at the 9th second. So, the overall maximum duration for any function expressible in this basis is 10 seconds. The number of functions was chosen based on the canonical hemodynamic response function to include its most informative segment.

All conditions and temporal basis functions were sampled at a higher temporal resolution ($16f_s$) and then downsampled (decimated) back to f_s after the convolution, similarly to the default setting in the fMRI SPM toolbox implementation (`defaults.stats.fmri.t = 16`) [77]. The last 2 regressors are one column representing the respiration signal and one column of ones representing an offset. The model contains $P = 92$ parameters to be estimated. The respiration signal was filtered between 0.01 and 0.7 Hz by the same filters used for the fNIRS signals during the preprocessing step and resampled to the sample rate of the fNIRS signal, f_s .

6.1.3.2 Correction for temporal correlations

The error term ε_t at time point t in the fNIRS signal is correlated with its temporal neighbors due to several slow physiological processes, such as respiration, heartbeat, and blood pressure changes. This temporal correlation has to be modeled. Most of the fNIRS-GLM analysis studies have used pre-whitening based on an autoregressive model (AR) of the error terms [202, 98, 15]. A first-order autoregressive model (AR(1)) of the error terms is

$$\varepsilon_t = a_1 \varepsilon_{t-1} + v_t \quad (6.4)$$

$$v_t \sim \mathcal{N}(0, \sigma_w^2) \quad (6.5)$$

where the value of ε_t at the current time t depends on the value from the previous time $t - 1$ and the normally distributed error term v_t with zero mean and variance σ_w^2 . First, the ordinary least squares estimates are calculated, as per Equation 6.3. The parameter a_1 is estimated by fitting AR(1) to the resulting residuals. Then, the Cochrane-Orcutt method [50] is used to reduce temporal correlation. This is defined as

$$y_t - a_1 y_{t-1} = X_t \beta_w - a_1 X_{t-1} \beta_w + \varepsilon_t - a_1 \varepsilon_{t-1} \quad (6.6)$$

which can be rewritten as

$$y_t^W = X_t^W \beta_w + v_t, \quad (6.7)$$

where $y_t^W = y_t - a_1 y_{t-1}$ and $X_t^W = X_t - a_1 X_{t-1}$ represent the pre-whitened data and design matrix respectively, and $\beta_w \in \mathbb{R}^P$ are new unknown parameters. We will use the superscript W to denote the pre-whitened data (y^W) and design matrix (X^W). Then, the transformed GLM with reduced temporal correlation is given by

$$y^W = X^W \beta_w + v, \quad (6.8)$$

and the ordinary least squares estimates of β_w are given by

$$\hat{\beta}_w = (X^{WT} X^W)^{-1} X^{WT} y^W . \quad (6.9)$$

6.1.3.3 Cross-validation

The GLM approach for a single fNIRS channel was adapted for the classification of evoked hemodynamic responses to differentiate between the semantic categories of animals and tools in a 15-block-wise cross-validation analysis in which one block is used for testing, while other blocks were used for training.

In one step of the cross-validation, let X_{train} be a modified version of the full design matrix X where all rows outside the training folds are set to zero. Data blocks were extended 5 seconds beyond the end of the experimental blocks to compensate for the hemodynamic delay and before possible artifacts in the fNIRS signal at the beginning of the breaks. Similarly, let y_{train} be the fNIRS signal zeroed outside the training folds. Keeping the original dimensions (but zeroing rows) makes things easier for overall manipulations, fitting AR(1) and pre-whitening without any time discontinuities. The terms a_1 and $\hat{\beta}_w$ are then estimated from X_{train} and y_{train} .

6.1.3.4 Extraction of evoked hemodynamic responses

The next step is to extract evoked hemodynamic responses. All four mental tasks in one concept trial share the same underlying hidden variable about the semantic category. Thus, it is not possible to extract the evoked hemodynamic response for an individual mental task. Instead, the evoked hemodynamic response must be extracted for the sequence of four mental tasks.

Let T be one such sequence of four mental tasks in one concept trial for which we want to extract the evoked hemodynamic response. The estimated GLM is used to remove influence from preceding trials and thus to isolate the evoked hemodynamic response for T . Let X_T be another modification of the design matrix X with the difference that, before the convolution,

the elements of the corresponding conditions for T are set to 0, instead of 1, when T is taking place. In other words, the design matrix X_T is excluding T from the design matrix X as if T did not take place in the experiment. The evoked hemodynamic response r for T , plus the normally distributed error, is then computed via

$$r = y^W - X_T^W \hat{\beta}_w. \quad (6.10)$$

In order to use the same analysis pipeline as for the non-GLM approach, the extracted hemodynamic responses were de-whitened. This approach brings the signal back to the original fNIRS space. To undo the effect of pre-whitening from Equation 6.7, the de-whitened response r^D is estimated from r , which is in the whitened space, as

$$r_t^D = r_t + a_1 r_{t-1} + a_1^2 r_{t-2}^D \quad (6.11)$$

with starting conditions $r_0^D = 0$ and $r_1^D = 0$.

6.1.4 Univariate classification

In order to investigate the temporal evolution of semantic decoding in a single channel, the sliding window approach was used. A temporal window was shifted in steps of half the window size in the concept trials, which start from the image presentation onset until 20 seconds after the image presentation onset. The window size was 4 samples, which correspond to about 2 seconds (2.04 seconds for the frontal montage and 1.79 seconds for the temporal montage).

The data in each temporal window and each channel were classified separately in 15-block-wise cross-validation. In each temporal window, the data were normalized (z-scored) and classified by a SVM ($C = 1$) with a radial basis function kernel.

6.1.5 Multivariate classification with PCA features

An approach based on spatial PCA was employed to decrease data dimensionality in the channel space while allowing the classifier to use information from all channels. The same sliding temporal window approach presented above was used.

In each temporal window, each channel was normalized (z-scored) separately. Spatial PCA then projected the data onto a smaller subspace by keeping only the $N \in \{1, 2, \dots, 10\}$ PCA components which explain most of the data variance. The selected components were normalized (z-scored) separately before being passed as features to the SVM mentioned above.

6.1.5.1 Feature selection

The sliding window approach was adapted for automatic feature selection by nested cross-validation. In an outer 15-block-wise cross-validation, each channel was normalized (z-scored) separately, as before, and spatial PCA was used to preserve N components.

The sliding temporal window was then used in the same configuration as before. Classification accuracies were computed in an inner stratified 10-fold cross-validation for each temporal window and each component separately. A maximum number of W temporal windows which had the highest accuracies were then selected. Only windows achieving significant accuracies ($p < 0.05$, more on this below) were considered.

The selected windows were then used for the test fold in the outer cross-validation. Concretely, the data in each window were normalized (z-scored) separately and the concatenation of all these windows were features for the same SVM mentioned above. We tested $W \in \{1, \dots, 20\}$ that correspond from 4 to 80 features and $N \in \{1, \dots, 10\}$ PCA components.

6.2 Results

Given the number of epochs in the experiment and the balance in the two classes, a classifier can be said to have a classification accuracy that is statistically significantly above chance level when its accuracy is above 56.11% for a significance level of $p < 0.05$ (one-sided Binomial test ($n = 180, q = 0.5$)), 58.89% for a significance level of $p < 0.01$, and 61.67% for a significance level of $p < 0.001$. This is the same statistical test that was discussed in Section 4.2 for the EEG analysis. Briefly, the Binomial distribution describes the probability distribution of a random classifier for two classes and the Binomial test is then used to assess the deviation from this distribution. Significance for a mean classification accuracy across participants was again calculated using the Binomial test and validated by a bootstrapping simulation (based on 10^6 simulations using a classifier that randomly choose a class).

6.2.1 Univariate classification

Figure 6.2 shows classification accuracies for the no-GLM and GLM data obtained with the sliding window approach. Statistically significant accuracies ($p < 0.05$) are shown for temporal windows starting from the image onset. Note that differences in temporal locations of classification accuracies may appear between the no-GLM and GLM approaches due to pre- and de-whitening in the GLM approach.

Each participant had at least one channel that allowed significant differentiation between the semantic categories of animals and tools. Even though this analysis disregarded the individual mental task types, temporal locations where significant accuracies occurred were found across the whole concept trial presentation period. Note that significant classification accuracies that occurred before 6–9 seconds may still be partially attributed to the image presentation due to the effect of the hemodynamic delay.

Figure 6.3 shows hemodynamic responses from participant 1 for the no-GLM approach. The generality of the temporal basis set adopted allows the timing of the main peak, if present,

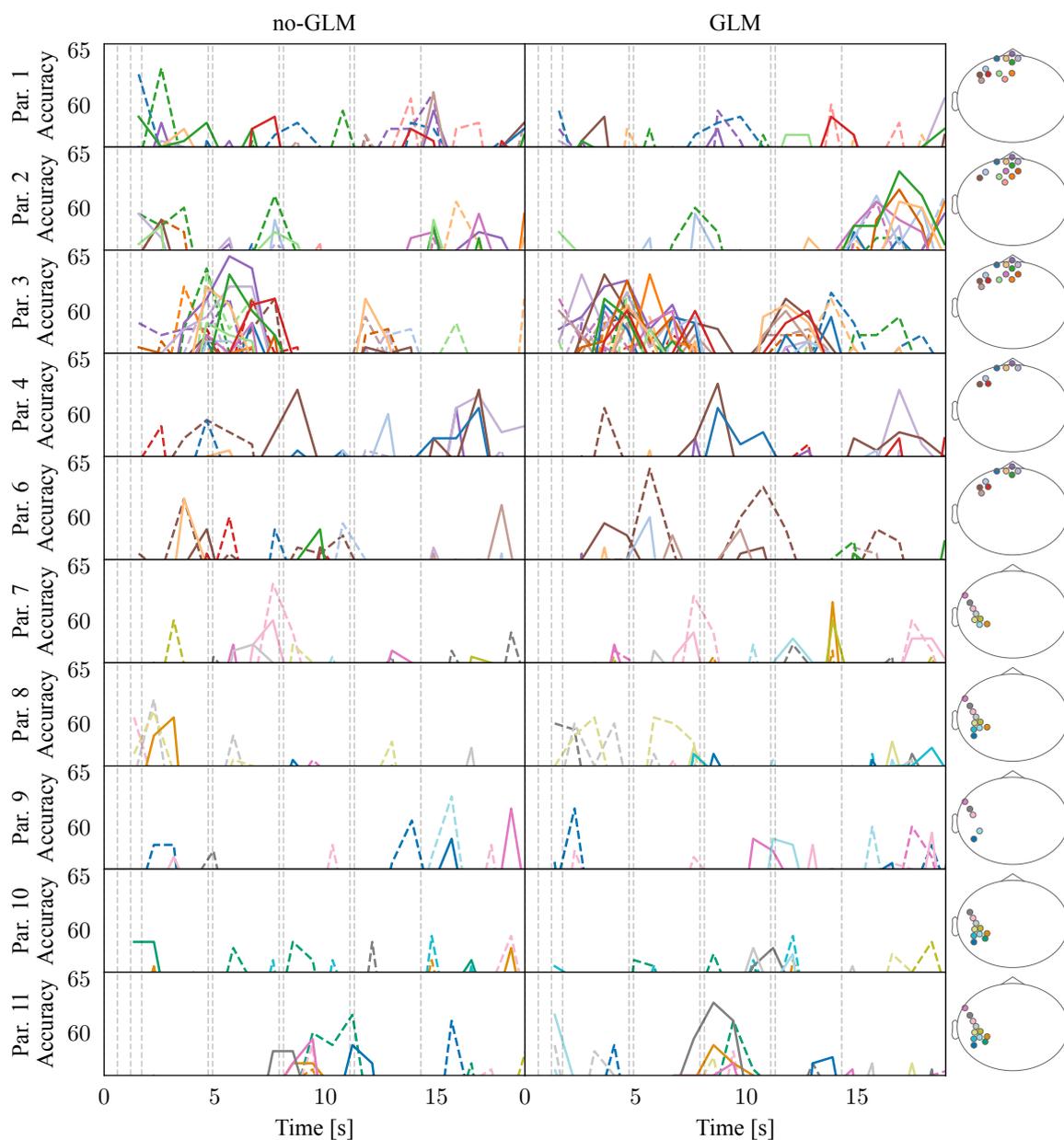
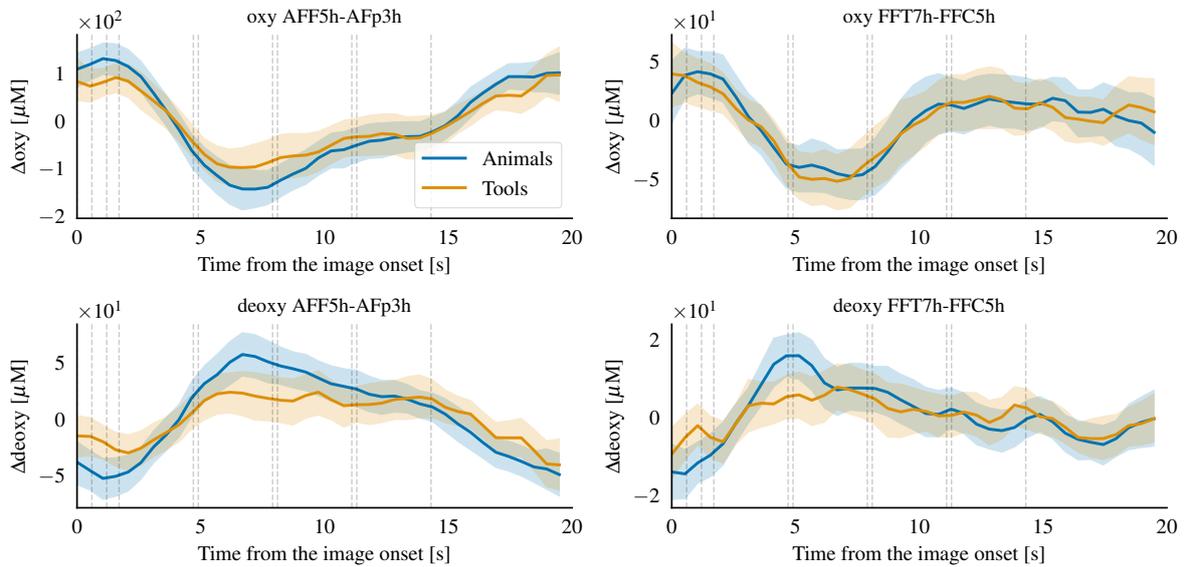
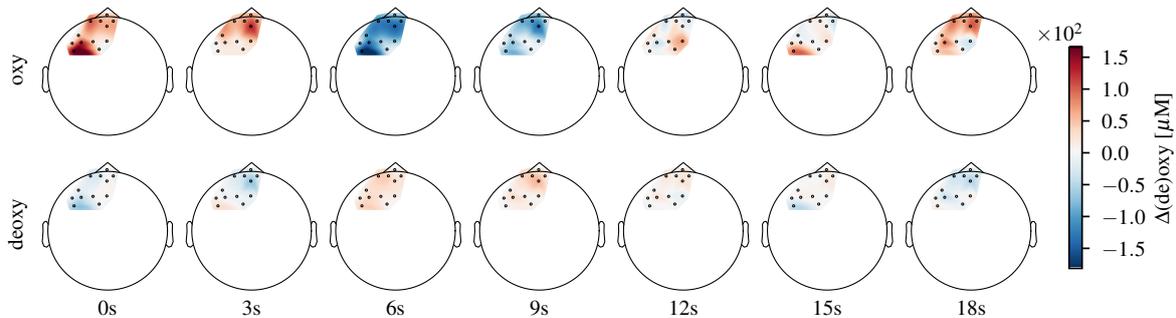


Fig. 6.2 Single channel classification accuracies obtained with the sliding temporal window starting from the image onset and the SVM ($C = 1$), see Section 6.1.4. Rows represent different participants, while columns indicate whether or not the GLM approach was used. The Y axes are limited to statistically significant classification accuracies from $p = 0.05$ to $p = 0.00001$. Solid lines indicate oxygenated channels, while dashed lines indicate deoxygenated channels. Vertical dashed lines represent events in concept trials.



(a) Hemodynamic responses for two channels: AFF5h-AFp3h (the left column) and FFT7h-FFC5h (the right column). Hemodynamic responses are shown with mean and 95% confidence interval. Changes in oxygenated hemoglobin are in the top row and changes in deoxygenated hemoglobin are in the bottom row.



(b) Mean hemodynamic responses for each channel starting from the image onset. Changes in oxygenated hemoglobin are in the top row and changes in deoxygenated hemoglobin are in the bottom row.

Fig. 6.3 Hemodynamic responses from participant 1 that were extracted using all blocks from the no-GLM data.

to vary for each channel type. Hemodynamic responses follow the expected relationship between changes in oxygenated and deoxygenated hemoglobin.

6.2.2 Multivariate classification with PCA features

Figures 6.4 and 6.5 show classification accuracies for the no-GLM and GLM data, respectively. Spatial PCA was used on only oxygenated, only deoxygenated, or both channel types together. This was done to inspect possible different temporal locations where statistically significant accuracies occur between oxygenated and deoxygenated channels.

The results from no-GLM and GLM data mostly shared similar temporal locations of significant classification accuracies. However, in participant 3, significant accuracies after 8 seconds in the GLM data were not prominent in the no-GLM data.

Some temporal locations where statistical significant accuracies occurred for either oxygenated or deoxygenated channels were not statistically significant when the PCA was used on both channel types together. For instance, participant 9 had a peak with deoxygenated channels under 5 seconds when using many different numbers of components. This peak was not present when spatial PCA was used together with the oxygenated channels.

6.2.2.1 Feature selection

Mean classification accuracies across participants for either frontal or temporal montages were not statistically significant when using only oxygenated, only deoxygenated, or both channel types in either the GLM or no-GLM approach.

Figures 6.6 and 6.7 shows how many participants had significant accuracies in the no-GLM and GLM approaches, respectively. It is difficult to statistically compare the results between different montages, channel types, or the no-GLM and GLM approaches when there are only 5 participants for the comparison and mean classification accuracies were not statistically significant in all tested options.

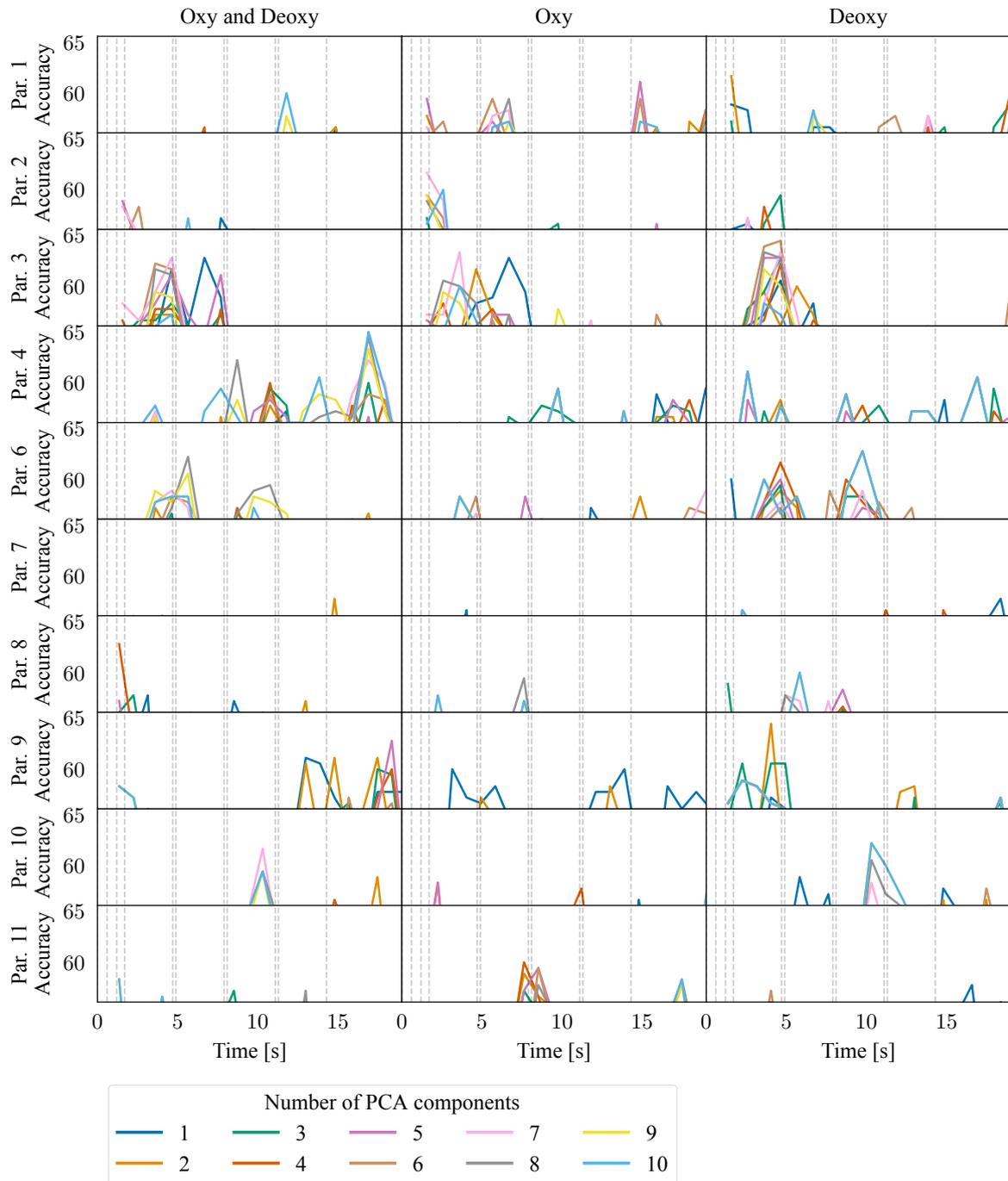


Fig. 6.4 Classification accuracies for the no-GLM approach obtained with the sliding temporal window starting from the image onset when the SVM ($C = 1$) can use information from the N PCA components, see Section 6.1.5. Rows represent different participants, while columns indicate channel types used. The Y axes are limited to statistically significant classification accuracies from $p = 0.05$ to $p = 0.00001$. Vertical dashed lines represent events in concept trials.

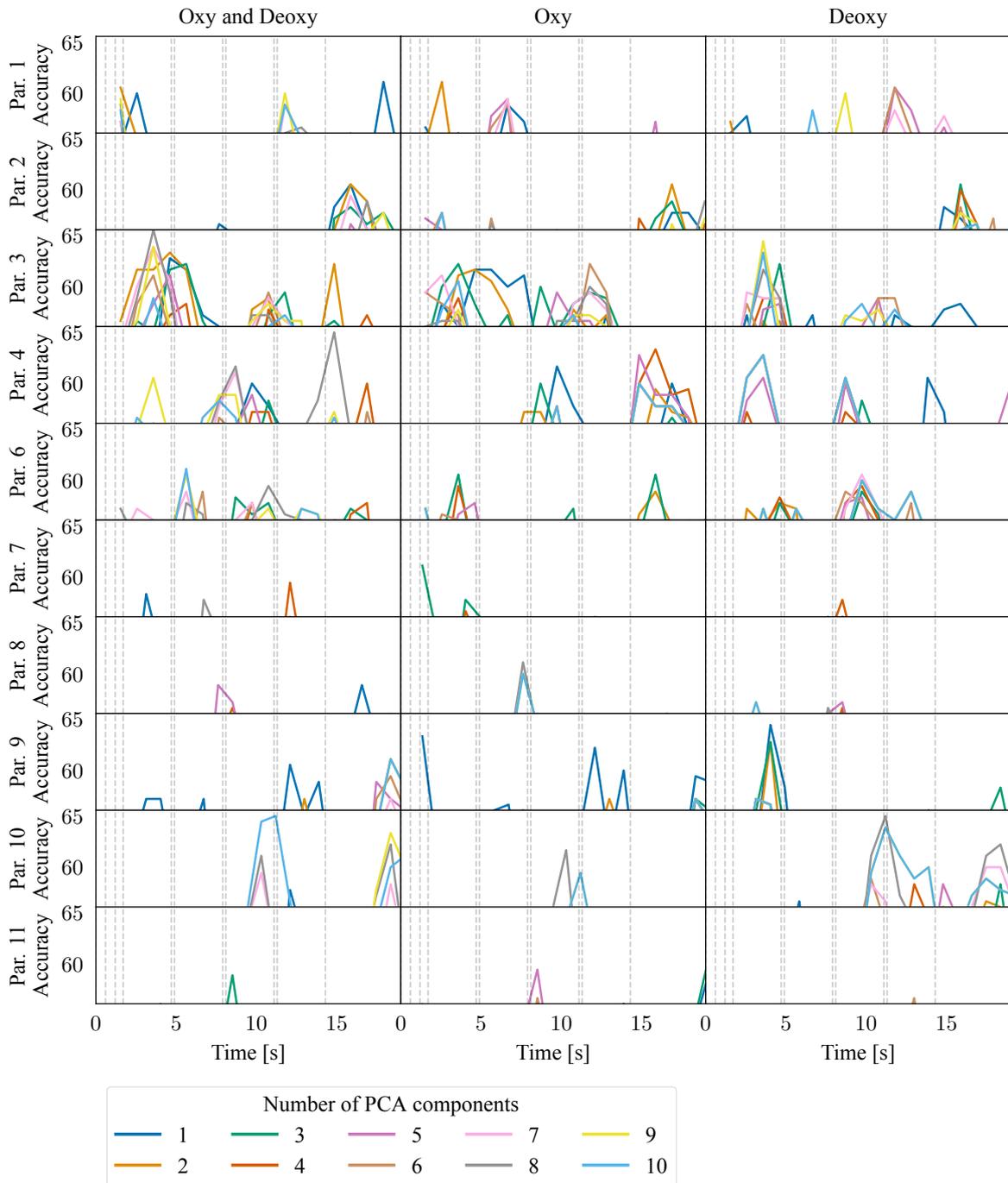


Fig. 6.5 Classification accuracies for the GLM approach obtained with the sliding temporal window starting from the image onset when the SVM ($C = 1$) can use information from the N PCA components, see Section 6.1.5. Rows represent different participants, while columns indicate channel types used. The Y axes are limited to statistically significant classification accuracies from $p = 0.05$ to $p = 0.00001$. Vertical dashed lines represent events in concept trials.

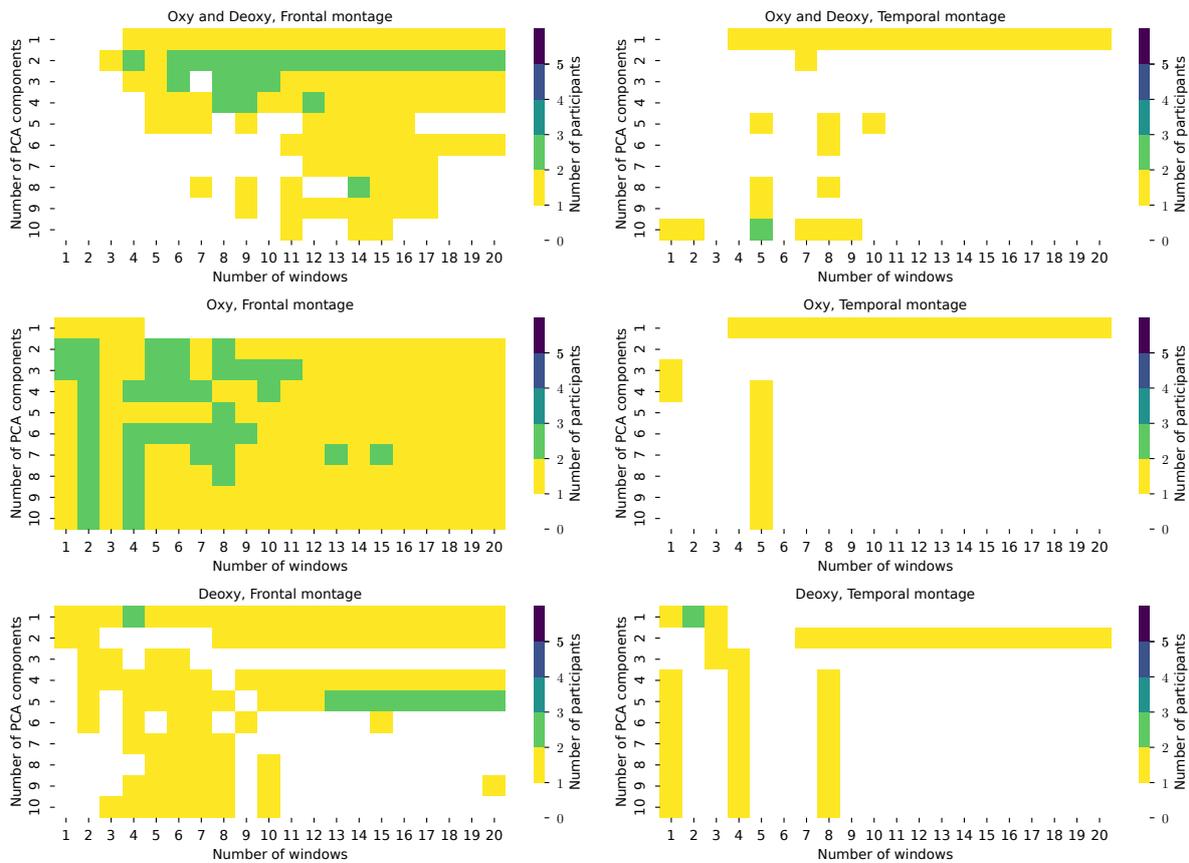


Fig. 6.6 Number of participants with significant classification accuracies for the no-GLM approach with feature selection achieved via the nested-cross-validation approach, see Section 6.1.5.1.

The frontal montage had more participants with significant classification accuracies. This should not be surprising because participants with the frontal montage had more time windows with statistically significant classification accuracies in comparison with participants with the temporal montage in the previous analysis without this time window selection, see Figures 6.4 and 6.5. The nested-cross-validation could then select these significant time windows resulting in more participants with significant classification accuracies. Similar observations can be made about differences between different channel types, and the no-GLM and GLM approaches based again on the analysis without this time window selection. However, it is unknown whether or not these observations would generalize with more participants because it was not possible to evaluate these observations with statistical tests.

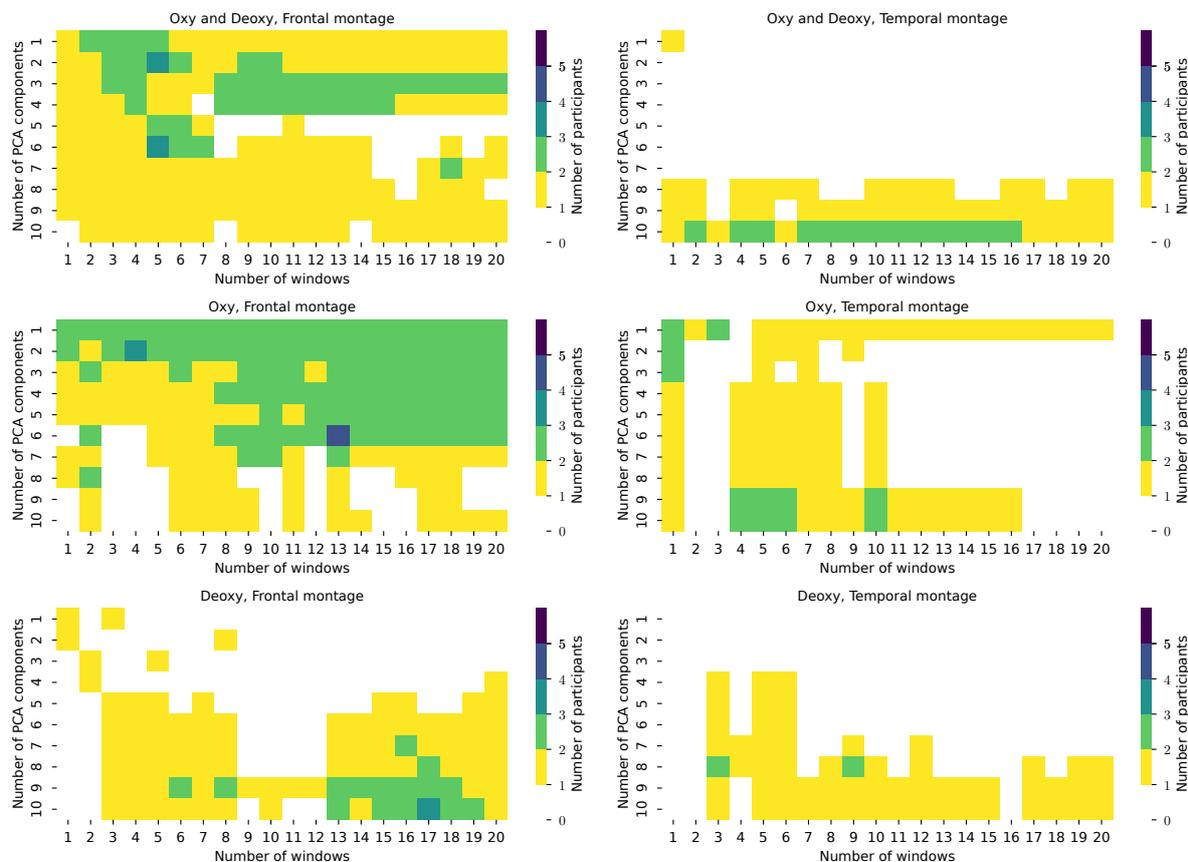


Fig. 6.7 Number of participants with significant classification accuracies for the GLM approach with feature selection achieved via the nested-cross-validation approach, see Section 6.1.5.1.

6.3 Discussion

Our results suggest the possibility of semantic decoding in fNIRS. Nevertheless, more future research is required to give us a conclusive answer as to whether or not semantic decoding is possible solely with fNIRS.

We used two analyses with the GLM and no-GLM data and a sliding window approach to inspect the temporal evolution of semantic decoding in the whole concept trial period. Univariate classification (see Section 6.2.1) and multivariate classification with PCA features (see Section 6.2.2) were able to achieve statistically significant classification accuracies for the majority of participants.

The temporal locations of semantic information varied between participants. These differences in temporal locations might be related to the variability in timing of when participants started performing the mental task between different trials, which was discussed in the chapter on EEG analysis (see Section 4.3). Due to the sequence of four different mental tasks, participants were required to quickly change between different mental tasks. With short gaps of only 200 ms between the mental tasks, this may have caused a high degree of variability in when participants started performing the mental task between different trials. However, this may be less of an issue for fNIRS signals than for EEG signals due to the hemodynamic delay, which acts as a low-pass filter. Additionally, the presented analyses disregarded mental task types and if the neural correlates of semantic decoding are not shared between mental tasks (i.e., do not generalize between mental tasks), this might have caused an additional difficulty for a classifier.

These results were achieved on the whole concept trial presentation. Significant classification accuracies that occurred before 6–9 seconds may still be partially attributed to the image presentation due to the effect of the hemodynamic delay. Chapter 5 described the influence of image presentation on semantic decoding in EEG. This issue may be even worse in fNIRS. The effect of the image presentation period is present in fNIRS signals for a long time due to the hemodynamic delay. Future research is required to further disentangle the influence of image presentation and mental task on the evoked hemodynamic response.

While our experimental design of the sequence of four different mental tasks and short gaps of only 200 ms between mental tasks is sufficient for EEG recordings, it might not have been the best decision for fNIRS recordings in terms of conclusiveness of the results. Due to the short gap between mental tasks, a classifier could exploit information from preceding and following mental task(s). While the GLM approach can model different mental tasks, it is not possible to remove the influence of preceding mental tasks in the same concept trial because all mental tasks share the same underlying information about the semantic category. This issue

should be partly mitigated by the random order in which the mental tasks were presented across different blocks.

There has been only one other fNIRS-based semantic neural decoding study to date. A study by Zinszer and colleagues [286] discriminated between semantic categories of animals and body parts while participants focused on audiovisual stimuli (photographs with a simultaneous auditory presentation of the object names) and thought about the meaning of that stimulus or any memory it evoked. Each stimulus was presented for 3 seconds and followed by an interstimulus interval of 6–9 seconds composed of fireworks and a short musical clip. Mean accuracies were 66%. Data were epoched from 6.5 to 9.0 seconds after the stimulus onset. I was not able to achieve similar mean classification accuracies. However, we were able to achieve significant classification accuracies for several participants in the range 60–65%. Additionally, the results achieved from ours and their study may still be affected by the cue presentation period.

In future experiments to further investigate the neural correlates of semantic decoding in fNIRS, we would modify our experimental design in the following ways. First, we would use only a single mental task after the image presentation. Second, a longer interstimulus interval should be employed, such as a canonical 6–9 seconds interval from fMRI research, to properly identify evoked hemodynamic responses. Lastly, to suppress the influence of image presentation, a longer interval between the image presentation and the mental task could be also used.

Additionally, future experiments should use additional short-distance detectors to detect and directly remove extracerebral signals close to the source [224, 281]. Components of fNIRS signals that are not driven by neuronal activity (i.e., neurovascular coupling) can act as possible confounds. These components include extracerebral hemodynamics and intracerebral hemodynamics caused by mental task-related systemic activity. They act as confounds as they may mimic the presence of, or attenuate, the neuronally induced hemodynamic response

[231, 248, 249]. In the GLM approach, the inclusion of the respiration signal as a regressor in the GLM and autoregressive model reduced the influence of only some of these components. However, we also tested using two preprocessing methods for fNIRS signals before the model estimation to further mitigate this issue. The results from both methods had the same message to the one presented in this chapter so they were not included in this chapter for simplicity. The first method was based on the PCA to remove global interference by removing an eigenvector with the highest eigenvalue [282, 76]. The second method was the common average reference [21]. Both methods were applied to the oxygenated and deoxygenated channels separately.

Overall, our results showed the exciting possibility of semantic decoding in fNIRS. However, we discovered that more research is required to give us a conclusive answer as to whether or not semantic decoding, without associated perceptual processing, is truly possible with fNIRS. Decoding of fNIRS signals entail additional challenges due to the slow hemodynamic delay in comparison with EEG signals. For this reason, future experiments must be designed with a special care to try to disentangle the effects of cue presentation and mental task on the evoked hemodynamic responses. This is an especially important prerequisite for semantic BCIs in which the cue presentation would not be present.

Chapter 7

Discussion

It is still an open question how semantic information, in general, is encoded, retrieved, and processed in the brain. While many theories have been proposed over the years and some theories have been rejected by a growing body of evidence, for instance, symbolic representation [193, 192, 154], there is still much we do not know.

This thesis focused on a decoding approach instead of trying to progress our understanding of the semantic system. Decoding approaches attempt to identify the current mental state of an individual from recordings of their neural activity by developing neural decoding models. These decoding models have been developed to decode a variety of activities, including, but not limited to, visual information [86, 101], affective states [127], visual imagery during sleep [100], story meaning [66], and music [62]. This thesis started testing the feasibility of semantic neural decoding for BCIs for communication.

The current research suggests an exciting possibility of semantic neural decoding, see Chapter 2. However, this suggestion comes from a variety of different neuroimaging studies, which differ in their research questions and, thus, mental tasks and experimental designs.

It is unknown whether it is possible to translate these results reported in the literature to BCI applications. The majority of the studies used fMRI. However, neuroimaging techniques that provide better ecological validity and allow use outside the lab are particularly interesting

for BCI applications. For these reasons, this thesis used EEG and fNIRS, two modalities which can be adapted for use outside the lab with relative ease. It is unclear whether or not the results, reported in the literature, based on BOLD signals, with a low temporal resolution and affected by the slow hemodynamic response, would translate to EEG, and other electrophysiological neuroimaging methods with a faster temporal resolution. On the other hand, EEG and fNIRS offer a lower spatial resolution in comparison with fMRI, and fNIRS can measure only cortical areas.

We selected four mental tasks that satisfy the requirements for semantic BCIs. These mental tasks could be used without any external cue. This property is in contrast with the majority of mental tasks, reviewed in Chapter 2, which would be difficult or near impossible to use outside their experimental designs. We designed an experiment to test the feasibility of differentiating between the semantic categories of animals and tools.

Even though neuroimaging studies suggested the possibility of semantic neural decoding, our mental tasks and experimental design significantly differed. We thus explored the recorded data with many different analyses to investigate whether or not the recorded neural activity contains any discriminative features that can be used for semantic neural decoding. However, no exploratory research can ever be completely comprehensive and we may have neglected to include some methods and approaches. This thesis presented exploratory analyses starting from the most simple features and methods to more complex ones. We explored more methods than presented in this thesis, but we omitted them if they did not provide any new insights or have the same message. For example, Behroozi and colleagues [24] discriminated semantic categories from the phase of the EEG signals. They reported that significant accuracies were achieved in a low-frequency band (1–4 Hz) using the phase but not the power of the EEG signal. On the other hand, the signal power provided significant accuracies in a high-frequency band (20–30 Hz). These accuracies were higher than when the phase was used. We also tested using the phase information alone and in combination with the signal power (as in Section 4.1.5).

However, the results were similar to using only the power of the EEG signals and the overall message was the same. We thus did not report these analyses.

7.1 Mental tasks

ERD/S maps revealed visually similar activities across all four mental tasks in Section 4.2.5 with the strongest ERD/S in parietal and occipital electrodes and partially in central and temporal electrodes. We believe that it is likely that visual imagery was included in the auditory and tactile imagery tasks. Thus, it is expected to see a large spatial overlap in those three mental imagery tasks.

The strongest ERD/S were in the tactile and auditory imagery tasks. However, this behavior might be related to the random seed error, which will be further discussed in Section 7.4. The tactile and auditory imagery tasks were most frequently presented as the first tasks during the sequence of four mental tasks in all but one participant, see Section 3.7. If this is indeed the case, the first mental task was the most important for semantic neural decoding.

For the silent naming tasks, Murphy and colleagues [173] reported that a wide range of occipital, parietal and frontal areas played a role in separating the semantic categories of mammals and tools in EEG. Soto and colleagues [242] reported significant decoding of semantic categories in fMRI in the middle temporal gyrus, anterior and inferior temporal, inferior parietal lobe, and prefrontal regions. Though prefrontal regions are typically thought to be involved in semantic control rather than in representing semantic knowledge [267, 203, 274].

No similar mental tasks to the proposed imagery tasks have been published in semantic decoding studies using EEG. When other neuroimaging modalities were used, the most similar mental tasks were task types based on ‘Properties’, such as silent properties generation, or ‘Meaning’, such as thinking freely about the meaning of a stimulus, as shown in Table 2.5. Nevertheless, our mental tasks should rather be compared with the results of mental imagery research [192, 152, 125, 123, 176, 135].

A review by McNorgan [152] found a general imagery network shared between different modalities. The review surveyed auditory, tactile, motor, gustatory, olfactory, and visual imagery in fMRI and PET studies. Shared activations between modalities (regardless of task) were found in bilateral dorsal parietal, left inferior frontal, and anterior insula regions. Modality-specific imagery for most modalities was also associated with activations in corresponding sensorimotor regions (but not necessarily in primary sensorimotor areas), primarily left-lateralized.

Specifically, auditory imagery did not activate the primary auditory cortex but it did activate the bilateral secondary auditory cortex (auditory associative areas) and the bilateral inferior frontal cortex [152].

Mental imagery research typically separates tactile and motor imagery [152]. However, there was no direct modality separation in the tactile imagery task employed here. Consider the examples given to participants: petting an animal or touching different parts of a tool, from Section 3.2. Both examples contain instructions for both motor and tactile imagery. According to the review [152], tactile imagery activates the left-lateralized primary sensorimotor cortex (primary somatosensory cortex) but it does not activate the primary motor cortex. Motor imagery activates the premotor cortex and primary somatosensory cortex but it does not activate the primary motor cortex.

The current evidence suggests that the early visual cortex is involved in visual imagery when the imagery task requires high-fidelity representations [124, 152]. Though visual imagery may activate the early visual cortex, different aspects of visual input are facilitated by upstream visual areas specialized, for instance, for color (neurons in area V4), motion (neurons in area V5/MT), and form perception (the lateral occipital complex) [152].

Whether or not the primary areas are activated during the mental tasks is an important discriminating factor between weak and strong embodiment theories [154]. However, the embodiment theory review [154] rejected the validity of strong embodiment outside conscious imagery.

A presentation of a stimulus image causes semantic priming. This form of visual priming could enhance performance of the visual imagery task in comparison with using a non-visual stimulus [229, 192, 123]. This prior semantic processing can also enhance subsequent picture naming performance [92]. However, we are not able to validate these potential issues due to none statistically significant mean classification accuracies.

Only a few semantic decoding studies, reviewed in Chapter 2, attempted semantic decoding during a free recall paradigm with minimal cue support [239, 168, 206]. During free recall, participants are often instructed to recall semantically related items consecutively. This is the most similar paradigm to the experimental design used here. For instance, an fMRI study [239] presented participants with semantic categories of animals and tools in four modalities: spoken words, written words, images, and natural sounds. Participants were asked to respond upon appearance to out-of-category exemplars. Additionally, the experiment ended with two free recall blocks in which participants were presented with a category name for 2 seconds and then instructed to recall all the entities from the cued category seen during the experiment in a 40 second period. A classifier was trained on the data from the actual stimulus presentation and then tested on the recall data. Mean classification accuracies across 14 participants were significant only when the classifier was trained on a combination of all modalities (67%) and partly when trained on a natural sounds modality (65%, $p = 0.01$, but below the p -value threshold for statistical significance used by the study). In general, studies that examined neural activity during free recall showed the rise and fall of the category-specific neural activity with different effects in different neuroimaging modalities [168, 206, 234, 80, 142, 147, 205].

7.2 Cue presentation confound

We showed that it is possible to differentiate between the semantic categories of animals and tools during the image presentation period from EEG in Chapter 5. We argued that this image presentation period should not be used in the decoding process to avoid confounds from: (1)

perceptual processing from the image presentation period and (2) semantic activation caused by perceiving the image. Semantic BCIs would not use an external cue so we were primarily interested in semantic activation caused by the concept-related mental tasks.

Although semantic neural decoding was possible during the image presentation, it cannot be fully contributed to perceptual processing from perceiving the image. The temporal evolution of single channel classification showed that significant mean classification accuracies moved from occipital (visual) areas to parietal, central, and frontal areas during the image presentation period, see Figure 5.1. The highest mean classification accuracies were located in parietal, central, and frontal areas in a 140.6–234.3 ms temporal window after the image onset. However, spatial locations of significant classification accuracies in this temporal window varied in individual participants, including occipital areas, see Figure D.5.

The image presentation informed participants about the concept that will be used in the following mental tasks. This was our effort to suppress the effect of external cues for semantic BCIs. However, this implementation does not necessary reflect the same neural activity that we would expect to be present in semantic BCIs. In semantic BCIs, users could start thinking about a particular concept of their own choice. On the other hand, participants had information about the concept in their minds before the onset of mental tasks from the image presentation period. In other words, our experimental design misses the initial semantic activation in mental tasks because this is only present during the image presentation.

7.3 Time-variant approaches

Our exploratory analyses showed a considerable variability in temporal and spectral locations of statistically significant classification accuracies across participants and mental tasks. Based on these results and due to the contrasting decoding difficulties in comparison with state-of-the-art research reported in the literature, we think that (non-phase-locked) time-variant approaches

are required. Examples of these methods are, for instance, recurrent neural networks such as long short-term memory networks [137, 57].

All presented exploratory analyses were, in effect, time-variant analyses to a certain degree. In other words, the discriminative features used by the classifier across different trials should be temporally located close to each another. However, this conclusion could be affected by the experiment design issues, which will be discussed in Section 7.4. Consequently, the task start time variance across trials could be responsible for this behavior.

7.4 Experiment shortcomings

The experiment had two shortcomings: the sequence of four mental tasks and the random seed error. The experimental design introduced the concept presentation followed by the sequence of four mental tasks. However, this design choice negatively affected the results' significance and conclusiveness for several reasons.

First, the sequence of four mental tasks required participants to keep switching between mental tasks. Each mental task lasted for 3 seconds and they were separated by only 200 ms. As we already discussed, this quick mental task switching with a short gap of 200 ms might have caused a high trial variability in mental task start times. In effect, time-invariant analyses would be inappropriate for this data. This issue could also explain the high variability in temporal locations of significant classification accuracies in all analyses.

Second, neural activity from the preceding mental task might still be present in the EEG signals due to the 200 ms gap between mental tasks. This could explain the visually similar ERD/S maps across all mental tasks.

Third, the mental task position in the sequence might be important for the performance of semantic neural decoding. For instance, the first mental task may be the easiest to decode. However, this might be related to the first issue of mental task start time variance. The first mental task neural activity might be more time-locked in regards to its starting time in

comparison with other mental tasks. In effect, its features would provide more discriminative information for semantic neural decoding in comparison with the following mental tasks.

Finally, the sequence was inappropriate for fNIRS analyses. It was not possible to make any claims regarding any particular mental task for two reasons. The 200 ms between mental tasks was too short, considering the slow hemodynamic responses and mental tasks in one sequence shared a variable about the category information. This proved to be ineffective for the GLM analysis.

The second shortcoming was the random seed error. The experiment had a fixed random seed for all but one participant, as discussed in Section 3.7. The effected participants shared the same order of mental tasks and concepts. However, the main issue was the shared order of mental tasks. Due to this issue, the mental task presentation at each position was not uniformly distributed across participants. For instance, the auditory and tactile imagery tasks were presented more frequently as the first mental task in comparison with the silent naming and visual imagery tasks. The strongest ERD/S were also in the tactile and auditory imagery tasks. The first mental task could have been the most important for semantic neural decoding. This might be for several reasons. As the first mental task, the neural activity was more “time-locked” to the mental task onset, while the temporal variability across trials could be higher for the following tasks because of the switching between mental tasks. The first mental task might have also produced neural activity with the most discriminative features for the decoding process. The degree of semantic neural activity might be lower in the following tasks in comparison with the first task. This lower degree of neural correlates might make the extraction of discriminate features more difficult.

If the ease of semantic decoding depends on the mental task position, this random seed error might have negatively affected the results. Then, we would expect to see that some mental tasks achieve higher classification accuracies than other mental tasks. However, this was not the case because mean classification accuracies were not statistically significant in any analyses.

We believe these results cannot be fully attributed to the mental task order. Thus, this random seed error did not adversely affect the results. Nevertheless, it is still possible that it affected the results but its influence was hidden by the previously mentioned experimental design issues resulting in insignificant classification accuracies.

7.5 Future research

The results from our exploratory analyses suggested the feasibility of semantic neural decoding for all four mental tasks. However, an additional experiment is required to conclusively confirm the results due to the discussed experimental design issues. Unfortunately, it was not possible to do another experiment with an improved experimental design due to the COVID-19 pandemic. Thus, this thesis could not compare mental tasks and make conclusive suggestions about their future potential for semantic BCIs.

A new experiment should fix the following issues in our experimental design. First, the sequence of four mental tasks should not be used. Instead, only a single mental task should be present after the concept presentation. This mental task would then be used in all trials in a single block (to avoid switching between different mental tasks). This change would test whether or not the time-variant analysis is indeed needed to get rid of the task order position confound and to remove the shared category variable for the fNIRS-GLM analysis. Second, testing the feasibility of four mental tasks was too much for a single experiment. A subset of mental tasks should be used instead. Based on the thesis results, the most promising mental tasks seem to be the auditory and tactile imagery tasks. The auditory imagery tasks had statistically significant mean classification accuracies in grid search for CSP analysis in Section 4.2.3.3. Both of these mental imagery tasks are also likely to contain visual imagery, which can help with semantic neural decoding. Third, gaps between trials should be longer for the fNIRS analysis to compensate for the hemodynamic delay. Fourth, electrooculography (EOG) recordings should be employed to suppress eye blinks and movements. This would

remove the need for eye blink suppression by ICA employed in this thesis for EEG. Fifth, an explicit baseline period should be introduced for fNIRS analysis. Finally, whole-brain fNIRS recordings would remove the need to use two different montages. Semantic neural decoding utilizes many different brain areas. However, the use of two separate montages limited our view only to a few brain areas. The whole-brain fNIRS activity would allow the inspection of the temporal evolution of semantic neural decoding across multiple bilateral brain regions.

Additional considerations for any BCI systems are user psycho-physiological states. The BCI system performance is impacted, for instance, by user's engagement, motivation, attention, cognitive load, or fatigue [163, 109, 266, 118, 146]. While "true" users of semantic BCIs for communication are likely to have a strong internal motivation during the mental tasks to communicate their desires to the outside world, we do not know the levels of engagement and motivation of participants in our experiment. Future research should consider increasing the participants' engagement, for instance, by providing feedback, gamification, or adapting to their psycho-physiological states to keep them in flow [143, 60, 59, 215, 163].

Chapter 8

Conclusions and future research

This chapter summarizes the main contributions of this thesis, discusses the limitations of this work, and suggests possible future research towards semantic BCIs.

8.1 Contributions

The current research suggests an exciting possibility of semantic neural decoding on which semantic BCIs could be built. Our systematic literature review in Chapter 2 showed a large variety of semantic concepts, mental tasks, experimental designs, and machine learning pipelines that have been employed to achieve semantic neural decoding. However, not all these mental tasks, experimental designs, and machine learning pipelines would be suitable for BCI scenarios. I needed to step back and first identify and test appropriate mental tasks for semantic BCIs.

In our experiment, presented in Chapter 3, we used one previously used mental task (a silent naming task) and proposed three novel and intuitive sensory-based imagery tasks using visual, auditory, and tactile perception. These mental tasks were then tested to differentiate between the semantic categories of animals and tools.

Our results suggested, and thus confirmed, the possibility of semantic decoding in EEG in Chapter 4 and in fNIRS in Chapter 6. However, we found contrasting differences in comparison

with other state-of-the-art research from the literature that suggested this possibility of semantic neural decoding. We discussed several interconnected reasons for this difference: the effect of cue presentation period on semantic decoding, the need of time-variant analysis, and our experimental design.

In contrast with most semantic decoding studies, we attempted to ensure the separation of the cue presentation period and the mental task period. We investigated this potential issue for EEG-based semantic decoding in Chapter 5. We found that all EEG-based semantic decoding studies published to date could exploit neural activity recorded during the cue presentation period in their analyses. When excluding the image presentation period, it was not possible to achieve similar levels of classification accuracies in comparison with these studies. However, this changed by including the image presentation period into the classification pipeline.

To summarize, for the main research objective, four mental tasks were identified that would be suitable for semantic BCI paradigms: the silent naming task and three novel imagery tasks using visual, auditory, and tactile perception. All these mental tasks would not require external cues in semantic BCI paradigms. The main contributions of this thesis, in regards to the research questions from Section 1.2, are the following:

1. This thesis identified an issue that affects many state-of-the-art semantic decoding studies. It is possible to differentiate between the semantic categories of animals and tools during the image presentation period in EEG. Including neural activity recorded during the cue presentation period into decoding pipelines negatively affects the conclusiveness of EEG-based semantic neural decoding.
2. It is possible to differentiate between the semantic categories of animals and tools during the four mental tasks from the EEG data. This thesis identified the need for a non-phase-locked and time-variant decoding analysis.
3. It is possible to differentiate between the semantic categories of animals and tools during the sequence of four mental tasks from the fNIRS data.

8.2 Limitations and future work

More research is required to investigate the effect of cue presentation period on semantic neural decoding. This is an especially important prerequisite for semantic BCIs in which the cue presentation would not be present. In addition to EEG, this effect of the cue presentation period might also influence state-of-the-art results from semantic decoding studies using neuroimaging modalities affected by the slow hemodynamic delay, such as fMRI and fNIRS.

Our EEG results in Chapter 4 revealed considerable variability in the time and frequency locations of semantic information between different participants and mental tasks. This suggests that time-variant analyses might be required. Alternatively, these differences in temporal locations could be caused by our experimental design. Due to the sequence of four mental tasks and short gaps of 200 ms between the mental tasks, participants were required to quickly switch between different mental tasks. This may have caused a high degree of variability when participants started performing the mental task between different trials. Future research is necessary to disentangle this issue.

Our fNIRS results in Chapter 6 were unfortunately influenced by not the most appropriate experimental design for fNIRS. Due to the slow hemodynamic response, fNIRS signals were affected by the above mentioned image presentation period and the sequence of four mental tasks with 200 ms gaps between them. While our results suggest the possibility of semantic neural decoding, future research with different experimental designs is needed to gain a better understanding of the evoked hemodynamic responses and to give us a conclusive answer as to whether or not fNIRS-based semantic decoding is truly achievable.

In this thesis, EEG and fNIRS signals were analyzed separately. However, a combined analysis of EEG and fNIRS could further improve decoding accuracy as these two techniques have complementary strengths.

It is important to note that this study and the majority of semantic decoding studies have been conducted offline. It is currently unknown whether similar results could be achieved in an online experimental paradigm for BCI applications.

Further research is needed to investigate which modality of mental imagery is the most suitable for semantic neural decoding and whether a single modality or a sequence of modalities performs better. Multimodal (multisensory) mental imagery could be especially promising [176, 152, 135]. In informal talks with participants during and after the experiment, some participants mentioned that they were unable to distinguish between different mental tasks. For instance, imagining touching an object included visualizing the object itself. This was not a problem here because the goal was to choose the mental tasks that elicit clear neural correlates for differentiation between the semantic categories of animals and tools. These questions are ultimately interconnected with imagery mental strategies and mental task instructions given to participants. Nevertheless, imagery vividness and strength vary between individuals [194, 61]. In potential BCI applications, the mental imagery strategy should be specifically selected for each individual according to their abilities and needs. Additionally, future research should investigate to what extent semantic decoding performance generalizes across different mental tasks.

This thesis was based on an exciting line of research that suggests the possibility of semantic neural decoding. In addition to semantic BCIs used for communication, this research has the potential to allow for universal communication based on cross-language neural decoding. While this research area involves considerable challenges (such as one-to-one mappings of neural encoding patterns for a given semantic concepts, neuroanatomical differences between participants, changes in concept meaning over the life span, and societal and cultural views of the concept), this thesis made a step towards EEG/fNIRS-based semantic BCIs.

References

- [1] Akama, H. and Murphy, B. (2017). Emerging methods for conceptual modelling in neuroimaging. *Behaviormetrika*, 44(1):117–133.
- [2] Akama, H., Murphy, B., Lei, M. M., and Poesio, M. (2014). Cross-participant modelling based on joint or disjoint feature selection: an fMRI conceptual decoding study. *Appl. Informatics*, 1(1):1.
- [3] Akama, H., Murphy, B., Na, L., Shimizu, Y., and Poesio, M. (2012). Decoding semantics across fMRI sessions with different stimulus modalities: a practical MVPA study. *Front. Neuroinform.*, 6:24.
- [4] Albilali, E., Aboalsamh, H., and Al-Wabil, A. (2013). Comparing brain-computer interaction and eye tracking as input modalities: An exploratory study. In *2013 Int. Conf. Curr. Trends Inf. Technol.*, pages 232–236. IEEE.
- [5] Alizadeh, S., Jamalabadi, H., Schönauer, M., Leibold, C., and Gais, S. (2017). Decoding cognitive concepts from neuroimaging data using multivariate pattern analysis. *Neuroimage*, 159:449–458.
- [6] Andersen, R. A., Hwang, E. J., and Mulliken, G. H. (2010). Cognitive Neural Prosthetics. *Annu. Rev. Psychol.*, 61(1):169–190.
- [7] Anderson, A. J., Binder, J. R., Fernandino, L., Humphries, C. J., Conant, L. L., Aguilar, M., Wang, X., Doko, D., and Raizada, R. D. S. (2017a). Predicting Neural Activity Patterns Associated with Sentences Using a Neurobiologically Motivated Model of Semantic Representation. *Cereb. Cortex*, 27(9):4379–4395.
- [8] Anderson, A. J., Kiela, D., Clark, S., and Poesio, M. (2017b). Visually Grounded and Textual Semantic Models Differentially Decode Brain Activity Associated with Concrete and Abstract Nouns. *Trans. Assoc. Comput. Linguist.*, 5:17–30.
- [9] Anderson, A. J. and Lin, F. (2019). How pattern information analyses of semantic brain activity elicited in language comprehension could contribute to the early identification of Alzheimer’s Disease. *NeuroImage Clin.*, 22:101788.
- [10] Anderson, A. J., Murphy, B., and Poesio, M. (2014). Discriminating Taxonomic Categories and Domains in Mental Simulations of Concepts of Varying Concreteness. *J. Cogn. Neurosci.*, 26(3):658–681.
- [11] Anderson, A. J., Zinszer, B. D., and Raizada, R. D. (2016). Representational similarity encoding for fMRI: Pattern-based synthesis to predict brain activity using stimulus-model-similarities. *Neuroimage*, 128:44–53.

- [12] Anumanchipalli, G. K., Chartier, J., and Chang, E. F. (2019). Speech synthesis from neural decoding of spoken sentences. *Nature*, 568(7753):493–498.
- [13] Ariely, D. and Berns, G. S. (2010). Neuromarketing: the hope and hype of neuroimaging in business. *Nat. Rev. Neurosci.*, 11(4):284–292.
- [14] Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., and Norman, K. A. (2017). Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron*, 95(3):709–721.e5.
- [15] Barker, J. W., Aarabi, A., and Huppert, T. J. (2013). Autoregressive model based algorithm for correcting motion and serially correlated errors in fNIRS. *Biomed. Opt. Express*, 4(8):1366–1379.
- [16] Baroni, M., Dinu, G., and Kruszewski, G. (2014). Don’t count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proc. 52nd Annu. Meet. Assoc. Comput. Linguist. (Volume 1 Long Pap.)*, volume 1, pages 238–247, Stroudsburg, PA, USA. Association for Computational Linguistics.
- [17] Bauer, A. J. and Just, M. A. (2015). Monitoring the growth of the neural representations of new animal concepts. *Hum. Brain Mapp.*, 36(8):3213–3226.
- [18] Bauer, A. J. and Just, M. A. (2017). A brain-based account of “basic-level” concepts. *Neuroimage*, 161:196–205.
- [19] Bauer, A. J. and Just, M. A. (2019). Brain reading and behavioral methods provide complementary perspectives on the representation of concepts. *Neuroimage*, 186:794–805.
- [20] Bauer, G., Gerstenbrand, F., and Rumpl, E. (1979). Varieties of the locked-in syndrome. *J. Neurol.*, 221(2):77–91.
- [21] Bauernfeind, G., Wriessnegger, S. C., Daly, I., and Müller-Putz, G. R. (2014). Separating heart and brain: on the reduction of physiological noise from multichannel functional near-infrared spectroscopy (fNIRS) signals. *J. Neural Eng.*, 11(5):056010.
- [22] BE, S. and ES, G. (1998). Timeline of the history of EEG and associated fields. *Electroencephalogr. Clin. Neurophysiol.*, 106(2):173–176.
- [23] Behroozi, M. and Daliri, M. R. (2014). Predicting brain states associated with object categories from fMRI data. *J. Integr. Neurosci.*, 13(04):645–667.
- [24] Behroozi, M., Daliri, M. R., and Shekarchi, B. (2016). EEG phase patterns reflect the representation of semantic categories of objects. *Med. Biol. Eng. Comput.*, 54(1):205–221.
- [25] Belliveau, J. W., Kennedy, D. N., McKinstry, R. C., Buchbinder, B. R., Weisskoff, R. M., Cohen, M. S., Vevea, J. M., Brady, T. J., and Rosen, B. R. (1991). Functional Mapping of the Human Visual Cortex by Magnetic Resonance Imaging. *Science (80-.)*, 254(5032):716–719.

- [26] Billinger, M., Daly, I., Kaiser, V., Jin, J., Allison, B. Z., Müller-Putz, G. R., and Brunner, C. (2012). Is It Significant? Guidelines for Reporting BCI Performance. In Allison, B. Z., Dunne, S., Leeb, R., Del R. Millán, J., and Nijholt, A., editors, *Toward Pract. Brain-Computer Interfaces Bridg. Gap from Res. to Real-World Appl.*, pages 333–354. Springer Berlin Heidelberg, Berlin, Heidelberg.
- [27] Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M., and Desai, R. H. (2016). Toward a brain-based componential semantic representation. *Cogn. Neuropsychol.*, 33(3-4):130–174.
- [28] Binder, J. R., Desai, R. H., Graves, W. W., and Conant, L. L. (2009). Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. *Cereb. Cortex*, 19(12):2767–2796.
- [29] Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- [30] Blankertz, B., Acqualagna, L., Dähne, S., Haufe, S., Schultze-Kraft, M., Sturm, I., Ušćumlic, M., Wenzel, M. A., Curio, G., and Müller, K.-R. (2016). The Berlin Brain-Computer Interface: Progress Beyond Communication and Control. *Front. Neurosci.*, 10:530.
- [31] Blankertz, B., Müller, K., Krusienski, D., Schalk, G., Wolpaw, J., Schlogl, A., Pfurtscheller, G., Millan, J., Schroder, M., and Birbaumer, N. (2006). The BCI Competition III: Validating Alternative Approaches to Actual BCI Problems. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 14(2):153–159.
- [32] Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., and Müller, K.-r. (2008). Optimizing Spatial filters for Robust EEG Single-Trial Analysis. *IEEE Signal Process. Mag.*, 25(1):41–56.
- [33] Blumenfeld, H. (2016). Neuroanatomical Basis of Consciousness. In *Neurol. Consciousness*, pages 3–29. Elsevier.
- [34] Bonnici, H. M., Chadwick, M. J., Lutti, A., Hassabis, D., Weiskopf, N., and Maguire, E. A. (2012). Detecting Representations of Recent and Remote Autobiographical Memories in vmPFC and Hippocampus. *J. Neurosci.*, 32(47):16982–16991.
- [35] Brandman, T. and Peelen, M. V. (2017). Interaction between Scene and Object Processing Revealed by Human fMRI and MEG Decoding. *J. Neurosci.*, 37(32):7700–7710.
- [36] Bray, S., Chang, C., and Hoeft, F. (2009). Applications of multivariate pattern classification analyses in developmental neuroimaging of healthy and clinical populations. *Front. Hum. Neurosci.*, 3:32.
- [37] Brigadoi, S., Ceccherini, L., Cutini, S., Scarpa, F., Scatturin, P., Selb, J., Gagnon, L., Boas, D. A., and Cooper, R. J. (2014). Motion artifacts in functional near-infrared spectroscopy: A comparison of motion correction techniques applied to real cognitive data. *Neuroimage*, 85:181–191.

- [38] Bruffaerts, R., De Deyne, S., Meersmans, K., Liuzzi, A. G., Storms, G., and Vandenberghe, R. (2019). Redefining the resolution of semantic knowledge in the brain: Advances made by the introduction of models of semantics in neuroimaging. *Neurosci. Biobehav. Rev.*, 103:3–13.
- [39] Buchweitz, A., Shinkareva, S. V., Mason, R. A., Mitchell, T. M., and Just, M. A. (2012). Identifying bilingual semantic neural representations across languages. *Brain Lang.*, 120(3):282–289.
- [40] Bulthé, J., De Smedt, B., and Op de Beeck, H. (2014). Format-dependent representations of symbolic and non-symbolic numbers in the human cortex as revealed by multi-voxel pattern analyses. *Neuroimage*, 87:311–322.
- [41] Cardinale, F., Casaceli, G., Raneri, F., Miller, J., and Lo Russo, G. (2016). Implantation of Stereoelectroencephalography Electrodes. *J. Clin. Neurophysiol.*, 33(6):490–502.
- [42] Carlson, T. A., Schrater, P., and He, S. (2003). Patterns of Activity in the Categorical Representations of Objects. *J. Cogn. Neurosci.*, 15(5):704–717.
- [43] Chadwick, M. J., Hassabis, D., Weiskopf, N., and Maguire, E. A. (2010). Decoding Individual Episodic Memory Traces in the Human Hippocampus. *Curr. Biol.*, 20(6):544–547.
- [44] Chan, A. M., Halgren, E., Marinkovic, K., and Cash, S. S. (2011). Decoding word and category-specific spatiotemporal representations from MEG and EEG. *Neuroimage*, 54(4):3028–3039.
- [45] Chang, K.-m. K., Mitchell, T., and Just, M. A. (2011). Quantitative modeling of the neural representation of objects: How semantic feature norms can account for fMRI activation. *Neuroimage*, 56(2):716–727.
- [46] Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., and Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nat. Neurosci.*, 20(1):115–125.
- [47] Christensen, R. (2011). *Plane Answers to Complex Questions*. Springer Texts in Statistics. Springer New York, New York, NY.
- [48] Cichy, R. M., Heinzle, J., and Haynes, J.-D. (2012). Imagery and Perception Share Cortical Representations of Content and Location. *Cereb. Cortex*, 22(2):372–380.
- [49] Clark, S. (2015). Vector Space Models of Lexical Meaning. In *Handb. Contemp. Semant. Theory*, pages 493–522. John Wiley & Sons, Ltd, Chichester, UK.
- [50] Cochrane, D. and Orcutt, G. H. (1949). Application of Least Squares Regression to Relationships Containing Auto-Correlated Error Terms. *J. Am. Stat. Assoc.*, 44(245):32–61.
- [51] Collinger, J. L., Boninger, M. L., Bruns, T. M., Curley, K., Wang, W., and Weber, D. J. (2013). Functional priorities, assistive technology, and brain-computer interfaces after spinal cord injury. *J. Rehabil. Res. Dev.*, 50(2):145.
- [52] Contini, E. W., Goddard, E., and Wardle, S. G. (2021). Reaction times predict dynamic brain representations measured with MEG for only some object categorisation tasks. *Neuropsychologia*, 151:107687.

- [53] Cooper, R. J., Selb, J., Gagnon, L., Phillip, D., Schytz, H. W., Iversen, H. K., Ashina, M., and Boas, D. A. (2012). A Systematic Comparison of Motion Artifact Correction Techniques for Functional Near-Infrared Spectroscopy. *Front. Neurosci.*, 6(OCT):147.
- [54] Correia, J. M., Jansma, B., Hausfeld, L., Kikkert, S., and Bonte, M. (2015). EEG decoding of spoken words in bilingual listeners: from words to language invariant semantic-conceptual representations. *Front. Psychol.*, 6(FEB):71.
- [55] Coutanche, M. N. and Thompson-Schill, S. L. (2015). Creating Concepts from Converging Features in Human Cortex. *Cereb. Cortex*, 25(9):2584–2593.
- [56] Cox, D. D. and Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) “brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage*, 19(2):261–270.
- [57] Craik, A., He, Y., and Contreras-Vidal, J. L. (2019). Deep learning for electroencephalogram (EEG) classification tasks: a review. *J. Neural Eng.*, 16(3):031001.
- [58] Cree, G. S. and McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *J. Exp. Psychol. Gen.*, 132(2):163–201.
- [59] Croom, A. M. (2015). Music practice and participation for psychological well-being: A review of how music influences positive emotion, engagement, relationships, meaning, and accomplishment. *Music. Sci.*, 19(1):44–64.
- [60] Csikszentmihalyi, M. (2000). *Beyond boredom and anxiety*. Jossey-Bass, San Francisco, CA, US.
- [61] Cui, X., Jeter, C. B., Yang, D., Montague, P. R., and Eagleman, D. M. (2007). Vividness of mental imagery: Individual variability can be measured objectively. *Vision Res.*, 47(4):474–478.
- [62] Daly, I. (2023). Neural decoding of music from the EEG. *Sci. Rep.*, 13(1):624.
- [63] Dash, D., Ferrari, P., and Wang, J. (2020). Decoding Imagined and Spoken Phrases From Non-invasive Neural (MEG) Signals. *Front. Neurosci.*, 14:290.
- [64] De Martino, F., Valente, G., Staeren, N., Ashburner, J., Goebel, R., and Formisano, E. (2008). Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *Neuroimage*, 43(1):44–58.
- [65] Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990). Indexing by latent semantic analysis. *J. Am. Soc. Inf. Sci.*, 41(6):391–407.
- [66] Dehghani, M., Boghrati, R., Man, K., Hoover, J., Gimbel, S. I., Vaswani, A., Zevin, J. D., Immordino-Yang, M. H., Gordon, A. S., Damasio, A., and Kaplan, J. T. (2017). Decoding the neural representation of story meanings across languages. *Hum. Brain Mapp.*, 38(12):6096–6106.
- [67] Delorme, A. and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods*, 134(1):9–21.

- [68] Delpy, D. T., Cope, M., van der Zee, P., Arridge, S., Wray, S., and Wyatt, J. (1988). Estimation of optical pathlength through tissue from direct time of flight measurement. *Phys. Med. Biol.*, 33(12):1433–1442.
- [69] Djokic, V. G., Maillard, J., Bulat, L., and Shutova, E. (2020). Decoding Brain Activity Associated with Literal and Metaphoric Sentence Comprehension Using Distributional Semantic Models. *Trans. Assoc. Comput. Linguist.*, 8:231–246.
- [70] Dubossarsky, H., De Deyne, S., and Hills, T. T. (2017). Quantifying the structure of free association networks across the life span. *Dev. Psychol.*, 53(8):1560–1570.
- [71] Erk, K. (2012). Vector Space Models of Word Meaning and Phrase Meaning: A Survey. *Lang. Linguist. Compass*, 6(10):635–653.
- [72] Fernandino, L., Humphries, C. J., Conant, L. L., Seidenberg, M. S., and Binder, J. R. (2016). Heteromodal Cortical Areas Encode Sensory-Motor Features of Word Meaning. *J. Neurosci.*, 36(38):9763–9769.
- [73] Fernandino, L., Humphries, C. J., Seidenberg, M. S., Gross, W. L., Conant, L. L., and Binder, J. R. (2015). Predicting brain activation patterns associated with individual lexical concepts based on five sensory-motor attributes. *Neuropsychologia*, 76:17–26.
- [74] Ferrante, A., Gavriel, C., and Faisal, A. (2015). Data-efficient hand motor imagery decoding in EEG-BCI by using Morlet wavelets & Common Spatial Pattern algorithms. In *2015 7th Int. IEEE/EMBS Conf. Neural Eng.*, pages 948–951. IEEE.
- [75] Fitzgibbon, S. P., Powers, D. M. W., Pope, K. J., and Clark, C. R. (2007). Removal of EEG Noise and Artifact Using Blind Source Separation. *J. Clin. Neurophysiol.*, 24(3):232–243.
- [76] Franceschini, M. A., Joseph, D. K., Huppert, T. J., Diamond, S. G., and Boas, D. A. (2006). Diffuse optical imaging of the whole head. *J. Biomed. Opt.*, 11(5):054007.
- [77] Friston, K., Ashburner, J., Kiebel, S., and Nichols, T. (2007). *Statistical Parametric Mapping*. Elsevier.
- [78] Fyshe, A., Sudre, G., Wehbe, L., Rafidi, N., and Mitchell, T. M. (2019). The lexical semantics of adjective–noun phrases in the human brain. *Hum. Brain Mapp.*, 40(15):4457–4469.
- [79] Garrard, P., Lambon Ralph, M. A., Hodges, J. R., and Patterson, K. (2001). Prototypicality, distinctiveness, and intercorrelation: Analyses of the semantic attributes of living and nonliving concepts. *Cogn. Neuropsychol.*, 18(2):125–174.
- [80] Gelbard-Sagiv, H., Mukamel, R., Harel, M., Malach, R., and Fried, I. (2008). Internally Generated Reactivation of Single Neurons in Human Hippocampus During Free Recall. *Science (80-.)*, 322(5898):96–101.
- [81] Geuze, J., Farquhar, J., and Desain, P. (2014). Towards a Communication Brain Computer Interface Based on Semantic Relations. *PLoS One*, 9(2):e87511.
- [82] Ghio, M., Vaghi, M. M. S., Perani, D., and Tettamanti, M. (2016). Decoding the neural representation of fine-grained conceptual categories. *Neuroimage*, 132:93–103.

- [83] Glover, G. H. (2011). Overview of Functional Magnetic Resonance Imaging. *Neurosurg. Clin. N. Am.*, 22(2):133–139.
- [84] Graimann, B., Allison, B., and Pfurtscheller, G. (2009). Brain–Computer Interfaces: A Gentle Introduction. In *Brain-Computer Interfaces Revolutionizing Human-Computer Interact.*, pages 1–27. Springer Berlin Heidelberg, Berlin, Heidelberg.
- [85] Gramfort, A., Luessi, M., Larson, E., Engemann, D., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., and Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Front. Neurosci.*, 7:267.
- [86] Güçlütürk, Y., Güçlü, U., van Gerven, M., and van Lier, R. (2018). Representations of naturalistic stimulus complexity in early and associative visual and auditory cortices. *Sci. Rep.*, 8(1):3439.
- [87] Hanson, S. J., Matsuka, T., and Haxby, J. V. (2004). Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a “face” area? *Neuroimage*, 23(1):156–166.
- [88] Hassabis, D., Kumaran, D., and Maguire, E. A. (2007). Using Imagination to Understand the Neural Basis of Episodic Memory. *J. Neurosci.*, 27(52):14365–14374.
- [89] Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., and Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science (80-.)*, 293(5539):2425–2430.
- [90] Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., Hanke, M., and Ramadge, P. J. (2011). A Common, High-Dimensional Model of the Representational Space in Human Ventral Temporal Cortex. *Neuron*, 72(2):404–416.
- [91] Haynes, J.-D. and Rees, G. (2006). Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.*, 7(7):523–534.
- [92] Heath, S., McMahon, K., Nickels, L., Angwin, A., MacDonald, A., van Hees, S., Johnson, K., and Copland, D. (2012). Priming Picture Naming with a Semantic Task: An fMRI Investigation. *PLoS One*, 7(3):e32809.
- [93] Heinsfeld, A. S., Franco, A. R., Craddock, R. C., Buchweitz, A., and Meneguzzi, F. (2018). Identification of autism spectrum disorder using deep learning and the ABIDE dataset. *NeuroImage Clin.*, 17:16–23.
- [94] Hill, N. J., Ricci, E., Haider, S., McCane, L. M., Heckman, S., Wolpaw, J. R., and Vaughan, T. M. (2014). A practical, intuitive brain–computer interface for communicating ‘yes’ or ‘no’ by listening. *J. Neural Eng.*, 11(3):035003.
- [95] Hochberg, L. R. and Anderson, K. D. (2012). BCI Users and Their Needs. In *Brain–Computer Interfaces Princ. Pract.*, pages 317–324. Oxford University Press.
- [96] Hoefle, S., Engel, A., Basilio, R., Alluri, V., Toivainen, P., Cagy, M., and Moll, J. (2018). Identifying musical pieces from fMRI data using encoding and decoding models. *Sci. Rep.*, 8(1):2266.

- [97] Hoffman, P. and Lambon Ralph, M. A. (2018). From percept to concept in the ventral temporal lobes: Graded hemispheric specialisation based on stimulus and task. *Cortex*, 101:107–118.
- [98] Hofmann, M. J., Herrmann, M. J., Dan, I., Obrig, H., Conrad, M., Kuchinke, L., Jacobs, A. M., and Fallgatter, A. J. (2008). Differential activation of frontal and parietal regions during visual word recognition: An optical topography study. *Neuroimage*, 40(3):1340–1349.
- [99] Honari-Jahromi, M., Chouinard, B., Blanco-Elorrieta, E., Pylkkänen, L., and Fyshe, A. (2021). Neural representation of words within phrases: Temporal evolution of color-adjectives and object-nouns during simple composition. *PLoS One*, 16(3):e0242754.
- [100] Horikawa, T., Tamaki, M., Miyawaki, Y., and Kamitani, Y. (2013). Neural Decoding of Visual Imagery During Sleep. *Science (80-.)*, 340(6132):639–642.
- [101] Huang, W., Yan, H., Wang, C., Li, J., Yang, X., Li, L., Zuo, Z., Zhang, J., and Chen, H. (2020). Long short-term memory-based neural decoding of object categories evoked by natural images. *Hum. Brain Mapp.*, 41(15):4442–4453.
- [102] Huggins, J. E., Moinuddin, A. A., Chiodo, A. E., and Wren, P. A. (2015). What Would Brain-Computer Interface Users Want: Opinions and Priorities of Potential Users With Spinal Cord Injury. *Arch. Phys. Med. Rehabil.*, 96(3):S38–S45.e5.
- [103] Huggins, J. E., Wren, P. A., and Gruis, K. L. (2011). What would brain-computer interface users want? Opinions and priorities of potential users with amyotrophic lateral sclerosis. *Amyotroph. Lateral Scler.*, 12(5):318–324.
- [104] Huppert, T., Hoge, R., Diamond, S., Franceschini, M., and Boas, D. (2006). A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *Neuroimage*, 29(2):368–382.
- [105] Huppert, T. J., Diamond, S. G., Franceschini, M. A., and Boas, D. A. (2009). HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Appl. Opt.*, 48(10):D280–D298.
- [106] Islam, M. K., Rastegarnia, A., and Yang, Z. (2016). Methods for artifact detection and removal from scalp EEG: A review. *Neurophysiol. Clin. Neurophysiol.*, 46(4-5):287–305.
- [107] Jahromy, F. Z. and Daliri, M. R. (2017). Semantic category-based decoding of human brain activity using a Gabor-based model by estimating intracranial field potential range in temporal cortex. *J. Integr. Neurosci.*, 16(4):419–428.
- [108] Janoos, F., Brown, G., Mórocz, I. Á., and Wells, W. M. (2013). State-Space Analysis of Working Memory in Schizophrenia: An FBIRN Study. *Psychometrika*, 78(2):279–307.
- [109] Jeunet, C., N’Kaoua, B., Subramanian, S., Hachet, M., and Lotte, F. (2015). Predicting Mental Imagery-Based BCI Performance from Personality, Cognitive Profile and Neurophysiological Patterns. *PLoS One*, 10(12):e0143962.
- [110] Johansson, V., Soekadar, S. R., and Clausen, J. (2017). Locked Out. *Cambridge Q. Healthc. Ethics*, 26(4):555–576.

- [111] Jung, J., Williams, S. R., Sanaei Nezhad, F., and Lambon Ralph, M. A. (2017). GABA concentrations in the anterior temporal lobe predict human semantic processing. *Sci. Rep.*, 7(1):15748.
- [112] Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., McKeown, M. J., Iragui, V., and Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37(2):163–178.
- [113] Just, M. A., Cherkassky, V. L., Aryal, S., and Mitchell, T. M. (2010). A Neurosemantic Theory of Concrete Noun Representation Based on the Underlying Brain Codes. *PLoS One*, 5(1):e8622.
- [114] Just, M. A., Cherkassky, V. L., Buchweitz, A., Keller, T. A., and Mitchell, T. M. (2014). Identifying Autism from Neural Representations of Social Interactions: Neurocognitive Markers of Autism. *PLoS One*, 9(12):e113879.
- [115] Just, M. A., Pan, L., Cherkassky, V. L., McMakin, D. L., Cha, C., Nock, M. K., and Brent, D. (2017a). Machine learning of neural representations of suicide and emotion concepts identifies suicidal youth. *Nat. Hum. Behav.*, 1(12):911–919.
- [116] Just, M. A., Wang, J., and Cherkassky, V. L. (2017b). Neural representations of the concepts in simple sentences: Concept activation prediction and context effects. *Neuroimage*, 157:511–520.
- [117] Kay, K. N., Naselaris, T., Prenger, R. J., and Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, 452(7185):352–355.
- [118] Keller, J. M. (1987). Development and use of the ARCS model of instructional design. *J. Instr. Dev.*, 10(3):2–10.
- [119] Kirschstein, T. and Köhling, R. (2009). What is the Source of the EEG? *Clin. EEG Neurosci.*, 40(3):146–149.
- [120] Kivisaari, S. L., van Vliet, M., Hultén, A., Lindh-Knuutila, T., Faisal, A., and Salmelin, R. (2019). Reconstructing meaning from bits of information. *Nat. Commun.*, 10(1):927.
- [121] Kleih, S. C., Herweg, A., Kaufmann, T., Staiger-Sälzer, P., Gerstner, N., and Kübler, A. (2015). The WIN-speller: a new intuitive auditory brain-computer interface spelling application. *Front. Neurosci.*, 9:346.
- [122] Koh, P. H., Glaser, D. E., Flandin, G., Kiebel, S., Butterworth, B., Maki, A., Delpy, D. T., and Elwell, C. E. (2007). Functional optical signal analysis: a software tool for near-infrared spectroscopy data processing incorporating statistical parametric mapping. *J. Biomed. Opt.*, 12(6):064010.
- [123] Kosslyn, S. M., Ganis, G., and Thompson, W. L. (2001). Neural foundations of imagery. *Nat. Rev. Neurosci.*, 2(9):635–642.
- [124] Kosslyn, S. M. and Thompson, W. L. (2003). When is early visual cortex activated during visual mental imagery? *Psychol. Bull.*, 129(5):723–746.

- [125] Kosslyn, S. M., Thompson, W. L., and Ganis, G. (2006). *The Case for Mental Imagery*. Oxford University Press.
- [126] Kothe, C. A. and Makeig, S. (2013). BCILAB: a platform for brain–computer interface development. *J. Neural Eng.*, 10(5):056014.
- [127] Kragel, P. A., Knodt, A. R., Hariri, A. R., and LaBar, K. S. (2016). Decoding Spontaneous Emotional States in the Human Brain. *PLOS Biol.*, 14(9):e2000106.
- [128] Kragel, P. A., Koban, L., Barrett, L. F., and Wager, T. D. (2018). Representation, Pattern Information, and Brain Signatures: From Neurons to Neuroimaging. *Neuron*, 99(2):257–273.
- [129] Kraskov, A., Quiroga, R. Q., Reddy, L., Fried, I., and Koch, C. (2007). Local Field Potentials and Spikes in the Human Medial Temporal Lobe are Selective to Image Category. *J. Cogn. Neurosci.*, 19(3):479–492.
- [130] Krepki, R., Blankertz, B., Curio, G., and Müller, K.-R. (2007). The Berlin Brain-Computer Interface (BBCI) – towards a new communication channel for online control in gaming applications. *Multimed. Tools Appl.*, 33(1):73–90.
- [131] Kriegeskorte, N. and Douglas, P. K. (2018). Interpreting encoding and decoding models. *Curr. Opin. Neurobiol.*, 55:167–179.
- [132] Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.*, 2:4.
- [133] Kübler, A., Furdea, A., Halder, S., Hammer, E. M., Nijboer, F., and Kotchoubey, B. (2009). A Brain-Computer Interface Controlled Auditory Event-Related Potential (P300) Spelling System for Locked-In Patients. *Ann. N. Y. Acad. Sci.*, 1157(1):90–100.
- [134] Kumar, M., Federmeier, K. D., Fei-Fei, L., and Beck, D. M. (2017). Evidence for similar patterns of neural activity elicited by picture- and word-based representations of natural scenes. *Neuroimage*, 155:422–436.
- [135] Lacey, S. and Sathian, K. (2011). Multisensory object representation: Insights from studies of vision and touch. In *Prog. Brain Res.*, volume 191, pages 165–176. Elsevier.
- [136] Landauer, T. K. and Dumais, S. T. (1997). A solution to Plato’s problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol. Rev.*, 104(2):211–240.
- [137] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.
- [138] Lee, S.-H., Kravitz, D. J., and Baker, C. I. (2012). Disentangling visual imagery and perception of real-world objects. *Neuroimage*, 59(4):4064–4073.
- [139] Li, Y., Richardson, R. M., and Ghuman, A. S. (2017). Multi-Connection Pattern Analysis: Decoding the representational content of neural communication. *Neuroimage*, 162:32–44.
- [140] Li, Y., Wang, G., Long, J., Yu, Z., Huang, B., Li, X., Yu, T., Liang, C., Li, Z., and Sun, P. (2011). Reproducibility and Discriminability of Brain Patterns of Semantic Categories Enhanced by Congruent Audiovisual Stimuli. *PLoS One*, 6(6):e20801.

- [141] Liu, H., Agam, Y., Madsen, J. R., and Kreiman, G. (2009). Timing, Timing, Timing: Fast Decoding of Object Information from Intracranial Field Potentials in Human Visual Cortex. *Neuron*, 62(2):281–290.
- [142] Long, N. M., Oztekin, I., and Badre, D. (2010). Separable Prefrontal Cortex Contributions to Free Recall. *J. Neurosci.*, 30(33):10967–10976.
- [143] Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D., and Munafò, M. R. (2016). Gamification of Cognitive Assessment and Cognitive Training: A Systematic Review of Applications and Efficacy. *JMIR Serious Games*, 4(2):e11.
- [144] Mahon, B. and Caramazza, A. (2010). Judging semantic similarity: an event-related fMRI study with auditory word stimuli. *Neuroscience*, 169(1):279–286.
- [145] Mahon, B. Z. and Caramazza, A. (2009). Concepts and Categories: A Cognitive Neuropsychological Perspective. *Annu. Rev. Psychol.*, 60(1):27–51.
- [146] Malone, T. W. and Lepper, M. R. (2021). Making Learning Fun : A Taxonomy of Intrinsic Motivations for Learning. 3:223–254.
- [147] Manning, J. R., Polyn, S. M., Baltuch, G. H., Litt, B., and Kahana, M. J. (2011). Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. *Proc. Natl. Acad. Sci.*, 108(31):12893–12897.
- [148] Mason, R. A. and Just, M. A. (2016). Neural Representations of Physics Concepts. *Psychol. Sci.*, 27(6):904–913.
- [149] Mason, R. A. and Just, M. A. (2020). Neural Representations of Procedural Knowledge. *Psychol. Sci.*, 31(6):729–740.
- [150] McCartney, B., Martinez-del Rincon, J., Devereux, B., and Murphy, B. (2019). A zero-shot learning approach to the development of brain-computer interfaces for image retrieval. *PLoS One*, 14(9):e0214342.
- [151] McFarland, D. J., Sarnacki, W. A., and Wolpaw, J. R. (2003). Brain-computer interface (BCI) operation: optimizing information transfer rates. *Biol. Psychol.*, 63(3):237–251.
- [152] McNorgan, C. (2012). A meta-analytic review of multisensory imagery identifies the neural correlates of modality-specific and modality-general imagery. *Front. Hum. Neurosci.*, 6(September):285.
- [153] Mecacci, G. and Haselager, P. (2019). Identifying Criteria for the Evaluation of the Implications of Brain Reading for Mental Privacy. *Sci. Eng. Ethics*, 25(2):443–461.
- [154] Meteyard, L., Cuadrado, S. R., Bahrami, B., and Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, 48(7):788–804.
- [155] Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient Estimation of Word Representations in Vector Space. *1st Int. Conf. Learn. Represent. ICLR 2013 - Work. Track Proc.*

- [156] Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. In *Adv. Neural Inf. Process. Syst.*
- [157] Miller, G. A. and Charles, W. G. (1991). Contextual correlates of semantic similarity. *Lang. Cogn. Process.*, 6(1):1–28.
- [158] Miller, K. J., Hermes, D., and Staff, N. P. (2020). The current state of electrocorticography-based brain–computer interfaces. *Neurosurg. Focus*, 49(1):E2.
- [159] Miller, K. J., Schalk, G., Hermes, D., Ojemann, J. G., and Rao, R. P. N. (2016). Spontaneous Decoding of the Timing and Content of Human Object Perception from Cortical Surface Recordings Reveals Complementary Information in the Event-Related Potential and Broadband Spectral Change. *PLOS Comput. Biol.*, 12(1):e1004660.
- [160] Mitchell, T. M., Hutchinson, R., Niculescu, R. S., Pereira, F., Wang, X., Just, M., and Newman, S. (2004). Learning to Decode Cognitive States from Brain Images. *Mach. Learn.*, 57(1):145–175.
- [161] Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K.-M., Malave, V. L., Mason, R. A., and Just, M. A. (2008). Predicting Human Brain Activity Associated with the Meanings of Nouns. *Science (80-.)*, 320(5880):1191–1195.
- [162] Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M.-a., Morito, Y., Tanabe, H. C., Sadato, N., and Kamitani, Y. (2008). Visual Image Reconstruction from Human Brain Activity using a Combination of Multiscale Local Image Decoders. *Neuron*, 60(5):915–929.
- [163] Mladenović, J., Frey, J., Bonnet-Save, M., Mattout, J., and Lotte, F. (2017). The Impact of Flow in an EEG-based Brain Computer Interface.
- [164] Moher, D., Liberati, A., Tetzlaff, J., and Altman, D. G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med.*, 6(7):e1000097.
- [165] Molavi, B. and Dumont, G. A. (2012). Wavelet-based motion artifact removal for functional near-infrared spectroscopy. *Physiol. Meas.*, 33(2):259–270.
- [166] Mormann, F. and Koch, C. (2007). Neural correlates of consciousness. *Scholarpedia*, 2(12):1740.
- [167] Mórocz, I. Á., Janoos, F., van Gelderen, P., Manor, D., Karni, A., Breznitz, Z., von Aster, M., Kushnir, T., and Shalev, R. (2012). Time-resolved and spatio-temporal analysis of complex cognitive processes and their role in disorders like developmental dyscalculia. *Int. J. Imaging Syst. Technol.*, 22(1):81–96.
- [168] Morton, N. W., Kahana, M. J., Rosenberg, E. A., Baltuch, G. H., Litt, B., Sharan, A. D., Sperling, M. R., and Polyn, S. M. (2013). Category-Specific Neural Oscillations Predict Recall Organization During Memory Search. *Cereb. Cortex*, 23(10):2407–2422.
- [169] Müller-Gerking, J., Pfurtscheller, G., and Flyvbjerg, H. (1999). Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clin. Neurophysiol.*, 110(5):787–798.

- [170] Mur, M., Bandettini, P. A., and Kriegeskorte, N. (2009). Revealing representational content with pattern-information fMRI—an introductory guide. *Soc. Cogn. Affect. Neurosci.*, 4(1):101–109.
- [171] Murphy, B., Baroni, M., and Poesio, M. (2009). EEG responds to conceptual stimuli and corpus semantics. In *Proc. 2009 Conf. Empir. Methods Nat. Lang. Process. Vol. 2 - EMNLP '09*, volume 2, pages 619–627, Morristown, NJ, USA. Association for Computational Linguistics.
- [172] Murphy, B. and Poesio, M. (2010). Detecting Semantic Category in Simultaneous EEG/MEG Recordings. In *Proc. NAACL HLT 2010 First Work. Comput. Neurolinguistics*, CN '10, pages 36–44, USA. Association for Computational Linguistics.
- [173] Murphy, B., Poesio, M., Bovolo, F., Bruzzone, L., Dalponte, M., and Lakany, H. (2011). EEG decoding of semantic category reveals distributed representations for single concepts. *Brain Lang.*, 117(1):12–22.
- [174] Murphy, G. L. G. L. (2002). *The big book of concepts*. MIT Press.
- [175] Na, Y., Choi, I., Jang, D. P., Kang, J. K., and Woo, J. (2019). Semantic-hierarchical model improves classification of spoken-word evoked electrocorticography. *J. Neurosci. Methods*, 311:253–258.
- [176] Nanay, B. (2018). Multimodal mental imagery. *Cortex*, 105:125–134.
- [177] Naseer, N. and Hong, K.-S. (2015). fNIRS-based brain-computer interfaces: a review. *Front. Hum. Neurosci.*, 9(JAN):3.
- [178] Naselaris, T., Kay, K. N., Nishimoto, S., and Gallant, J. L. (2011). Encoding and decoding in fMRI. *Neuroimage*, 56(2):400–410.
- [179] Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., and Gallant, J. L. (2009). Bayesian Reconstruction of Natural Images from Human Brain Activity. *Neuron*, 63(6):902–915.
- [180] Neuper, C., Müller-Putz, G. R., Scherer, R., and Pfurtscheller, G. (2006). Motor imagery and EEG-based control of spelling devices and neuroprostheses. In *Prog. Brain Res.*, volume 159, pages 393–409. Elsevier.
- [181] Niazi, A. M., van den Broek, P. L. C., Klanke, S., Barth, M., Poel, M., Desain, P., and van Gerven, M. A. J. (2014). Online decoding of object-based attention using real-time fMRI. *Eur. J. Neurosci.*, 39(2):319–329.
- [182] Nicolas-Alonso, L. F. and Gomez-Gil, J. (2012). Brain Computer Interfaces, a Review. *Sensors*, 12(2):1211–1279.
- [183] Niedermeyer, E. (1999). The Normal EEG of the Waking Adult. *Electroencephalogr. Basic Princ. Clin. Appl. Relat. Fields*.
- [184] Ogawa, S., Lee, T.-M., Nayak, A. S., and Glynn, P. (1990). Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magn. Reson. Med.*, 14(1):68–78.

- [185] Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Comput. Intell. Neurosci.*, 2011:1–9.
- [186] Pakhomov, S. V., Hemmy, L. S., and Lim, K. O. (2012). Automated semantic indices related to cognitive function and rate of cognitive decline. *Neuropsychologia*, 50(9):2165–2175.
- [187] Palatucci, M., Pomerleau, D., Hinton, G. E., and Mitchell, T. M. (2009). Zero-shot Learning with Semantic Output Codes. In Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C., and Culotta, A., editors, *Adv. Neural Inf. Process. Syst.*, volume 22. Curran Associates, Inc.
- [188] Parra, L. C., Spence, C. D., Gerson, A. D., and Sajda, P. (2005). Recipes for the linear analysis of EEG. *Neuroimage*, 28(2):326–341.
- [189] Pasley, B. N. and Knight, R. T. (2013). Decoding Speech for Understanding and Treating Aphasia. In *Prog. Brain Res.*, volume 207, pages 435–456. Elsevier.
- [190] Pasqualotto, E., Matuz, T., Federici, S., Ruf, C. A., Bartl, M., Olivetti Belardinelli, M., Birbaumer, N., and Halder, S. (2015). Usability and Workload of Access Technology for People With Severe Motor Impairment. *Neurorehabil. Neural Repair*, 29(10):950–957.
- [191] Patterson, K., Nestor, P. J., and Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci.*, 8(12):976–987.
- [192] Pearson, J. (2019). The human imagination: the cognitive neuroscience of visual mental imagery. *Nat. Rev. Neurosci.*, 20(10):624–634.
- [193] Pearson, J. and Kosslyn, S. M. (2015). The heterogeneity of mental representation: Ending the imagery debate. *Proc. Natl. Acad. Sci.*, 112(33):10089–10092.
- [194] Pearson, J., Naselaris, T., Holmes, E. A., and Kosslyn, S. M. (2015). Mental Imagery: Functional Mechanisms and Clinical Applications. *Trends Cogn. Sci.*, 19(10):590–602.
- [195] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.*, 12(85):2825–2830.
- [196] Pennington, J., Socher, R., and Manning, C. (2014). Glove: Global Vectors for Word Representation. In *Proc. 2014 Conf. Empir. Methods Nat. Lang. Process.*, pages 1532–1543, Stroudsburg, PA, USA. Association for Computational Linguistics.
- [197] Pereira, F., Gershman, S., Ritter, S., and Botvinick, M. (2016). A comparative evaluation of off-the-shelf distributed semantic representations for modelling behavioural data. *Cogn. Neuropsychol.*, 33(3-4):175–190.
- [198] Pereira, F., Lou, B., Pritchett, B., Ritter, S., Gershman, S. J., Kanwisher, N., Botvinick, M., and Fedorenko, E. (2018). Toward a universal decoder of linguistic meaning from brain activation. *Nat. Commun.*, 9(1):963.

- [199] Pfeiderer, B. (2002). Visualization of Auditory Habituation by fMRI. *Neuroimage*, 17(4):1705–1710.
- [200] Pfurtscheller, G. and Neuper, C. (2001). Motor imagery and direct brain-computer communication. *Proc. IEEE*, 89(7):1123–1134.
- [201] Pinti, P., Scholkmann, F., Hamilton, A., Burgess, P., and Tachtsidis, I. (2019). Current Status and Issues Regarding Pre-processing of fNIRS Neuroimaging Data: An Investigation of Diverse Signal Filtering Methods Within a General Linear Model Framework. *Front. Hum. Neurosci.*, 12:505.
- [202] Plichta, M., Heinzel, S., Ehlis, A.-C., Pauli, P., and Fallgatter, A. (2007). Model-based analysis of rapid event-related functional near-infrared spectroscopy (NIRS) data: A parametric validation study. *Neuroimage*, 35(2):625–634.
- [203] Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., and Gabrieli, J. D. (1999). Functional Specialization for Semantic and Phonological Processing in the Left Inferior Prefrontal Cortex. *Neuroimage*, 10(1):15–35.
- [204] Pollonini, L., Olds, C., Abaya, H., Bortfeld, H., Beauchamp, M. S., and Oghalai, J. S. (2014). Auditory cortex activation to natural speech and simulated cochlear implant speech measured with functional near-infrared spectroscopy. *Hear. Res.*, 309:84–93.
- [205] Polyn, S. M., Kragel, J. E., Morton, N. W., McCluey, J. D., and Cohen, Z. D. (2012). The neural dynamics of task context in free recall. *Neuropsychologia*, 50(4):447–457.
- [206] Polyn, S. M., Natu, V. S., Cohen, J. D., and Norman, K. A. (2005). Category-Specific Cortical Activity Precedes Retrieval During Memory Search. *Science (80-.)*, 310(5756):1963–1966.
- [207] Proix, T., Delgado Saa, J., Christen, A., Martin, S., Pasley, B. N., Knight, R. T., Tian, X., Poeppel, D., Doyle, W. K., Devinsky, O., Arnal, L. H., Mégevand, P., and Giraud, A.-L. (2022). Imagined speech can be decoded from low- and cross-frequency intracranial EEG features. *Nat. Commun.*, 13(1):48.
- [208] Proklova, D., Kaiser, D., and Peelen, M. V. (2019). MEG sensor patterns reflect perceptual but not categorical similarity of animate and inanimate objects. *Neuroimage*, 193:167–177.
- [209] Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nat. Rev. Neurosci.*, 6(7):576–582.
- [210] Raizada, R. D. S. and Connolly, A. C. (2012). What Makes Different People’s Representations Alike: Neural Similarity Space Solves the Problem of Across-subject fMRI Decoding. *J. Cogn. Neurosci.*, 24(4):868–877.
- [211] Ramoser, H., Muller-Gerking, J., and Pfurtscheller, G. (2000). Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans. Rehabil. Eng.*, 8(4):441–446.
- [212] Reber, T. P., Bausch, M., Mackay, S., Boström, J., Elger, C. E., and Mormann, F. (2019). Representation of abstract semantic knowledge in populations of human single neurons in the medial temporal lobe. *PLOS Biol.*, 17(6):e3000290.

- [213] Reddy, L., Tsuchiya, N., and Serre, T. (2010). Reading the mind's eye: Decoding category information during mental imagery. *Neuroimage*, 50(2):818–825.
- [214] Rolls, E. T. and Tovee, M. J. (1994). Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proc. R. Soc. London. Ser. B Biol. Sci.*, 257(1348):9–15.
- [215] Ron-Angevin, R. and Díaz-Estrella, A. (2009). Brain–computer interface: Changes in performance using virtual reality techniques. *Neurosci. Lett.*, 449(2):123–127.
- [216] Rosch, E. (1978). Principles of categorization. *Princ. Categ.*, pages 27–48.
- [217] Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., and Boyes-Braem, P. (1976). Basic objects in natural categories. *Cogn. Psychol.*, 8(3):382–439.
- [218] Rugg, M. D. and Vilberg, K. L. (2013). Brain networks underlying episodic memory retrieval. *Curr. Opin. Neurobiol.*, 23(2):255–260.
- [219] Rupp, K., Roos, M., Milsap, G., Caceres, C., Ratto, C., Chevillet, M., Crone, N. E., and Wolmetz, M. (2017). Semantic attributes are encoded in human electrocorticographic signals during visual object recognition. *Neuroimage*, 148:318–329.
- [220] Ruts, W., De Deyne, S., Ameel, E., Vanpaemel, W., Verbeemen, T., and Storms, G. (2004). Dutch norm data for 13 semantic categories and 338 exemplars. *Behav. Res. Methods, Instruments, Comput.*, 36(3):506–515.
- [221] Rybář, M. and Daly, I. (2022). Neural decoding of semantic concepts: a systematic literature review. *J. Neural Eng.*, 19(2):021002.
- [222] Rybář, M., Daly, I., and Poli, R. (2020). Potential pitfalls of widely used implementations of common spatial patterns. In *2020 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pages 196–199.
- [223] Sabra, Z., Bonilha, L., and Naselaris, T. (2020). Spectral Encoding of Seen and Attended Object Categories in the Human Brain. *J. Neurosci.*, 40(2):327–342.
- [224] Sato, T., Nambu, I., Takeda, K., Aihara, T., Yamashita, O., Isogaya, Y., Inoue, Y., Otaka, Y., Wada, Y., Kawato, M., Sato, M.-a., and Osu, R. (2016). Reduction of global interference of scalp-hemodynamics in functional near-infrared spectroscopy using short distance probes. *Neuroimage*, 141:120–132.
- [225] Savaki, H. E. and Raos, V. (2019). Action perception and motor imagery: Mental practice of action. *Prog. Neurobiol.*, 175:107–125.
- [226] Scarapicchia, V., Brown, C., Mayo, C., and Gawryluk, J. R. (2017). Functional Magnetic Resonance Imaging and Functional Near-Infrared Spectroscopy: Insights from Combined Recording Studies. *Front. Hum. Neurosci.*, 11:419.
- [227] Schacter, D. L., Addis, D. R., and Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. *Nat. Rev. Neurosci.*, 8(9):657–661.
- [228] Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., and Szpunar, K. K. (2012). The Future of Memory: Remembering, Imagining, and the Brain. *Neuron*, 76(4):677–694.

- [229] Schacter, D. L., Dobbins, I. G., and Schnyer, D. M. (2004). Specificity of priming: a cognitive neuroscience perspective. *Nat. Rev. Neurosci.*, 5(11):853–862.
- [230] Schmidt, T. T., Ostwald, D., and Blankenburg, F. (2014). Imaging tactile imagery: Changes in brain connectivity support perceptual grounding of mental images in primary sensory cortices. *Neuroimage*, 98:216–224.
- [231] Scholkmann, F., Kleiser, S., Metz, A. J., Zimmermann, R., Mata Pavia, J., Wolf, U., and Wolf, M. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *Neuroimage*, 85:6–27.
- [232] Schroeter, M. L., Bücheler, M. M., Müller, K., Uludağ, K., Obrig, H., Lohmann, G., Tittgemeyer, M., Villringer, A., and von Cramon, D. (2004). Towards a standard analysis for functional near-infrared imaging. *Neuroimage*, 21(1):283–290.
- [233] Schrouff, J., Mourão-Miranda, J., Phillips, C., and Parvizi, J. (2016). Decoding intracranial EEG data with multiple kernel learning method. *J. Neurosci. Methods*, 261:19–28.
- [234] Sederberg, P. B., Schulze-Bonhage, A., Madsen, J. R., Bromfield, E. B., Litt, B., Brandt, A., and Kahana, M. J. (2007). Gamma Oscillations Distinguish True From False Memories. *Psychol. Sci.*, 18(11):927–932.
- [235] Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell Syst. Tech. J.*, 27(3):379–423.
- [236] Sheikh, U. A., Carreiras, M., and Soto, D. (2021). Neurocognitive mechanisms supporting the generalization of concepts across languages. *Neuropsychologia*, 153:107740.
- [237] Shinkareva, S. V., Malave, V. L., Mason, R. A., Mitchell, T. M., and Just, M. A. (2011). Commonality of neural representations of words and pictures. *Neuroimage*, 54(3):2418–2425.
- [238] Shinkareva, S. V., Mason, R. A., Malave, V. L., Wang, W., Mitchell, T. M., and Just, M. A. (2008). Using fMRI Brain Activation to Identify Cognitive States Associated with Perception of Tools and Dwellings. *PLoS One*, 3(1):e1394.
- [239] Simanova, I., Hagoort, P., Oostenveld, R., and van Gerven, M. A. J. (2014). Modality-Independent Decoding of Semantic Information from the Human Brain. *Cereb. Cortex*, 24(2):426–434.
- [240] Simanova, I., van Gerven, M., Oostenveld, R., and Hagoort, P. (2010). Identifying Object Categories from Event-Related EEG: Toward Decoding of Conceptual Representations. *PLoS One*, 5(12):e14465.
- [241] Simanova, I., van Gerven, M. A. J., Oostenveld, R., and Hagoort, P. (2015). Predicting the Semantic Category of Internally Generated Words from Neuromagnetic Recordings. *J. Cogn. Neurosci.*, 27(1):35–45.
- [242] Soto, D., Sheikh, U. A., Mei, N., and Santana, R. (2020). Decoding and encoding models reveal the role of mental simulation in the brain representation of meaning. *R. Soc. Open Sci.*, 7(5):192043.

- [243] Sternberg, S. (1966). High-Speed Scanning in Human Memory. *Science (80-.)*, 153(3736):652–654.
- [244] Sudre, G., Pomerleau, D., Palatucci, M., Wehbe, L., Fyshe, A., Salmelin, R., and Mitchell, T. (2012). Tracking neural coding of perceptual and semantic features of concrete nouns. *Neuroimage*, 62(1):451–463.
- [245] Suefusa, K. and Tanaka, T. (2017). A comparison study of visually stimulated brain–computer and eye-tracking interfaces. *J. Neural Eng.*, 14(3):036009.
- [246] Suppes, P., Lu, Z.-L., and Han, B. (1997). Brain wave recognition of words. *Proc. Natl. Acad. Sci.*, 94(26):14965–14969.
- [247] Szpunar, K. K., Watson, J. M., and McDermott, K. B. (2007). Neural substrates of envisioning the future. *Proc. Natl. Acad. Sci.*, 104(2):642–647.
- [248] Tachtsidis, I. and Scholkmann, F. (2016). False positives and false negatives in functional near-infrared spectroscopy: issues, challenges, and the way forward. *Neurophotonics*, 3(3):031405.
- [249] Takahashi, T., Takikawa, Y., Kawagoe, R., Shibuya, S., Iwano, T., and Kitazawa, S. (2011). Influence of skin blood flow on near-infrared spectroscopy signals measured on the forehead during a verbal fluency task. *Neuroimage*, 57(3):991–1002.
- [250] Themelis, G., D’Arceuil, H., Diamond, S. G., Thaker, S., Huppert, T. J., Boas, D. A., and Franceschini, M. A. (2007). Near-infrared spectroscopy measurement of the pulsatile component of cerebral blood flow and volume from arterial oscillations. *J. Biomed. Opt.*, 12(1):014033.
- [251] Themelis, G., Selb, J., Thaker, S., Stott, J. J., Custo, A., Boas, D., and Franceschini, M. A. (2004). Depth of arterial oscillation resolved with NIRS time and frequency domain. In *Biomed. Top. Meet.*, page WF2, Washington, D.C. OSA.
- [252] Thirion, B., Duchesnay, E., Hubbard, E., Dubois, J., Poline, J.-B., Lebihan, D., and Dehaene, S. (2006). Inverse retinotopy: Inferring the visual content of images from brain activation patterns. *Neuroimage*, 33(4):1104–1116.
- [253] Tu, K., Cooper, D. G., and Siegelmann, H. T. (2009). Memory reconsolidation for natural language processing. *Cogn. Neurodyn.*, 3(4):365–372.
- [254] Turney, P. D. and Pantel, P. (2010). From Frequency to Meaning: Vector Space Models of Semantics. *J. Artif. Intell. Res.*, 37:141–188.
- [255] van de Nieuwenhuijzen, M., Axmacher, N., Fell, J., Oehr, C., Jensen, O., and van Gerven, M. (2016). Decoding of task-relevant and task-irrelevant intracranial EEG representations. *Neuroimage*, 137:132–139.
- [256] Van de Putte, E., De Baene, W., Brass, M., and Duyck, W. (2017). Neural overlap of L1 and L2 semantic representations in speech: A decoding approach. *Neuroimage*, 162:106–116.

- [257] Van Uden, C. E., Nastase, S. A., Connolly, A. C., Feilong, M., Hansen, I., Gobbini, M. I., and Haxby, J. V. (2018). Modeling Semantic Encoding in a Common Neural Representational Space. *Front. Neurosci.*, 12(JUL):437.
- [258] Vargas, R. and Just, M. A. (2020). Neural Representations of Abstract Concepts: Identifying Underlying Neurosemantic Dimensions. *Cereb. Cortex*, 30(4):2157–2166.
- [259] Velmans, M. (2009). *Understanding Consciousness*. Routledge.
- [260] Vidal, J. R., Ossandón, T., Jerbi, K., Dalal, S. S., Minotti, L., Ryvlin, P., Kahane, P., and Lachaux, J.-P. (2010). Category-Specific Visual Responses: An Intracranial Study Comparing Gamma, Beta, Alpha, and ERP Response Selectivity. *Front. Hum. Neurosci.*, 4.
- [261] Vidaurre, C., Sander, T. H., and Schlögl, A. (2011). BioSig: The Free and Open Source Software Library for Biomedical Signal Processing. *Comput. Intell. Neurosci.*, 2011:1–12.
- [262] Vigliocco, G., Vinson, D. P., Druks, J., Barber, H., and Cappa, S. F. (2011). Nouns and verbs in the brain: A review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neurosci. Biobehav. Rev.*, 35(3):407–426.
- [263] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., and van Mulbregt, P. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods*, 17(3):261–272.
- [264] Vodrahalli, K., Chen, P.-H., Liang, Y., Baldassano, C., Chen, J., Yong, E., Honey, C., Hasson, U., Ramadge, P., Norman, K. A., and Arora, S. (2018). Mapping between fMRI responses to movies and their natural language annotations. *Neuroimage*, 180:223–231.
- [265] Volosyak, I. and Hochschule Rhein-Waal (2016). *EEG-based brain-computer interfaces for healthcare applications*.
- [266] Vygotsky, L. and Cole, M. (1978). *Mind in society: Development of higher psychological processes*.
- [267] Wagner, A. D., Paré-Blagoev, E., Clark, J., and Poldrack, R. A. (2001). Recovering Meaning: Left Prefrontal Cortex Guides Controlled Semantic Retrieval. *Neuron*, 31(2):329–338.
- [268] Wang, J., Baucom, L. B., and Shinkareva, S. V. (2013). Decoding abstract and concrete concept representations based on single-trial fMRI data. *Hum. Brain Mapp.*, 34(5):1133–1147.
- [269] Wang, J., Cherkassky, V. L., and Just, M. A. (2017). Predicting the brain activation pattern associated with the propositional content of a sentence: Modeling neural representations of events and states. *Hum. Brain Mapp.*, 38(10):4865–4881.
- [270] Wang, Y. (2013). Classifying Single-Trial EEG during Motor Imagery with a Small Training Set. *ArXiv*, abs/1306.3.

- [271] Warrington, E. K. and Shallice, T. (1984). Category specific semantic impairments. *Brain*, 107(3):829–853.
- [272] Wei Wang, Degenhart, A. D., Sudre, G. P., Pomerleau, D. A., and Tyler-Kabara, E. C. (2011). Decoding semantic information from human electrocorticographic (ECoG) signals. In *2011 Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pages 6294–6298. IEEE.
- [273] Wenzel, M. A., Bogojeski, M., and Blankertz, B. (2017). Real-time inference of word relevance from electroencephalogram and eye gaze. *J. Neural Eng.*, 14(5):056007.
- [274] Whitney, C., Kirk, M., O’Sullivan, J., Lambon Ralph, M. A., and Jefferies, E. (2011). The Neural Organization of Semantic Control: TMS Evidence for a Distributed Network in Left Inferior Frontal and Posterior Middle Temporal Gyrus. *Cereb. Cortex*, 21(5):1066–1075.
- [275] Widmann, A., Schröger, E., and Maess, B. (2015). Digital filter design for electrophysiological data – a practical approach. *J. Neurosci. Methods*, 250:34–46.
- [276] Winkler, I., Haufe, S., and Tangermann, M. (2011). Automatic Classification of Artifactual ICA-Components for Artifact Removal in EEG Signals. *Behav. Brain Funct.*, 7(1):30.
- [277] Wolpaw, J., Birbaumer, N., Heetderks, W., McFarland, D., Peckham, P., Schalk, G., Donchin, E., Quatrano, L., Robinson, C., and Vaughan, T. (2000). Brain-computer interface technology: a review of the first international meeting. *IEEE Trans. Rehabil. Eng.*, 8(2):164–173.
- [278] Yang, Y., Wang, J., Bailer, C., Cherkassky, V., and Just, M. A. (2017a). Commonalities and differences in the neural representations of English, Portuguese, and Mandarin sentences: When knowledge of the brain-language mappings for two languages is better than one. *Brain Lang.*, 175:77–85.
- [279] Yang, Y., Wang, J., Bailer, C., Cherkassky, V., and Just, M. A. (2017b). Commonality of neural representations of sentences across languages: Predicting brain activation during Portuguese sentence comprehension using an English-based model of brain function. *Neuroimage*, 146:658–666.
- [280] Ye, J. C., Tak, S., Jang, K. E., Jung, J., and Jang, J. (2009). NIRS-SPM: Statistical parametric mapping for near-infrared spectroscopy. *Neuroimage*, 44(2):428–447.
- [281] Yücel, M. A., Selb, J., Aasted, C. M., Petkov, M. P., Becerra, L., Borsook, D., and Boas, D. A. (2015). Short separation regression improves statistical significance and better localizes the hemodynamic response obtained by near-infrared spectroscopy for tasks with differing autonomic responses. *Neurophotonics*, 2(3):035005.
- [282] Zhang, Y., Brooks, D. H., Franceschini, M. A., and Boas, D. A. (2005). Eigenvector-based spatial filtering for reduction of physiological interference in diffuse optical imaging. *J. Biomed. Opt.*, 10(1):011014.
- [283] Zhou, S., Jin, J., Daly, I., Wang, X., and Cichocki, A. (2016). Optimizing the Face Paradigm of BCI System by Modified Mismatch Negative Paradigm. *Front. Neurosci.*, 10:444.

-
- [284] Zimeo Morais, G. A., Balardin, J. B., and Sato, J. R. (2018). fNIRS Optodes' Location Decider (fOLD): a toolbox for probe arrangement guided by brain regions-of-interest. *Sci. Rep.*, 8(1):3341.
- [285] Zinszer, B. D., Anderson, A. J., Kang, O., Wheatley, T., and Raizada, R. D. S. (2016). Semantic Structural Alignment of Neural Representational Spaces Enables Translation between English and Chinese Words. *J. Cogn. Neurosci.*, 28(11):1749–1759.
- [286] Zinszer, B. D., Bayet, L., Emberson, L. L., Raizada, R. D. S., and Aslin, R. N. (2017). Decoding semantic representations from functional near-infrared spectroscopy signals. *Neurophotonics*, 5(1):011003.

Appendix A

Functional near-infrared spectroscopy

This appendix briefly overviews principles behind near-infrared spectroscopy and describes the Modified Beer-Lambert law (MBLL). The MBLL converts measured light intensities into changes in oxygenated and deoxygenated hemoglobin. These changes are then used as indirect measurements of underlying cortical neural activity.

A.1 Overview

Near-infrared spectroscopy (NIRS) is a non-invasive neuroimaging method that takes advantage of an optical window in the near-infrared (NIR) spectrum between 650–950 nm [231]. The NIR light can penetrate up to several centimeters of tissue of the head through skin and skull where it may interact with tissue chromophores (absorbing compound), such as hemoglobin.

The dominant chromophores in human tissue in the visible and near-infrared light are oxygenated and deoxygenated hemoglobin and water. Light below 650 nm is strongly absorbed by hemoglobin and above 950 nm too strongly by water. In this optical window, oxygenated and deoxygenated hemoglobin have different minima and maxima in their absorption spectra. Other substances have higher absorption coefficients but they are present only at relatively low concentrations. Oxygenated and deoxygenated hemoglobin can thus be considered as the

main chromophores. We denote oxygenated hemoglobin as HbO_2 despite different notations in literature (HbO_2 , O_2Hb , HbO) and deoxygenated hemoglobin as Hb despite different notations in literature (Hb , HHb , HbR).

NIR measurements of the brain are done by placing optodes (optical sensor devices) on the scalp. A measurement channel is formed by an optical emitter (source) and receiver (detector). The emitter sends the NIR light to the head. The tissue of the human head is highly scattering. Only a small fraction of the emitted light reached the surface of the head at a specific detector position. The light entering at a source location and exiting the head at a detector position samples a diffuse volume between these positions. The depth and shape of the probing volume is a complex function of the source-detector distance, local tissue optical properties, and other aspects of measurement geometry. Over time, changes in the detected light occur due to changes in optical absorption by the underlying tissue (i.e., changes in tissue chromophore concentrations).

Continuous wave NIRS (cwNIRS) is solely based on a light intensity measurement. It uses a stable light source to send a continuous NIR light into the tissue. The intensity of detected light is used to determine the amount of optical absorption. Other forms of NIRS techniques, such as frequency domain or time domain methods, additionally measure the time that light needs to travel through the tissue, for more information see a recent review [231]. They can be used for absolute measurement of chromophore concentration. The disadvantage of cwNIRS systems are that they cannot fully determine the optical properties of the tissue. Therefore, concentrations of tissue chromophore cannot be determined absolutely. However, with a few reasonable assumptions it is possible to quantify changes in their concentration. On the other hand, cwNIRS systems have a higher temporal resolution, lower cost, can be miniaturized and be wireless. We will focus on cwNIRS from now on because our lab have equipment of this type.

A.2 Modified Beer-Lambert law

The Beer-Lambert law states that attenuation in the light intensity is proportional to the concentration of chromophores in a non-absorbing medium and the path length of the photons. However, biological tissue is highly scattering medium. The MBLL [68] extends the Beer-Lambert law by introducing an additive term to account for scattering loss and a term for the change in the optical path length. Let I be the detected light intensity (in units V) and I_0 the incident light intensity (i.e., intensity of emitted light) of the wavelength λ . The MBLL describes the optical density OD^λ (sometimes called attenuation A , unitless) as

$$OD^\lambda(t) = -\log_{10} \left(\frac{I^\lambda(t)}{I_0^\lambda(t)} \right) = \sum_i \varepsilon_i^\lambda c_i(t) DPF^\lambda d + G^\lambda. \quad (\text{A.1})$$

The index i denotes all investigated chromophores. c_i are chromophore concentrations (in units M (molar concentration), or as $\text{mol} \cdot \text{L}^{-1}$). ε_i^λ are (molar) extinction coefficients (in units $\text{M}^{-1} \text{cm}^{-1}$; not to be mistaken with the molar absorption coefficient $\alpha = \ln(10)\varepsilon$ or the absorption coefficient $\mu_a = \alpha c$ in cm^{-1} when using the natural logarithm instead of base 10 logarithm in Equation A.1). DPF^λ is the differential path length factor (unitless) to account for the increased path length due to the scattering. d is the source-detector separation (in cm). G^λ is a geometric factor incorporating the scattering loss (unitless). Assuming that G^λ is time-invariant, it can be neglected when determining the change in optical density

$$\Delta OD^\lambda(\Delta t) = OD^\lambda(t_1) - OD^\lambda(t_0) \quad (\text{A.2})$$

for a time point t_1 against an initial time point t_0 . Additionally, we assume that the emitter intensity I_0 is constant and therefore this term cancels out. The differential form of the MBLL (dMBLL) is

$$\Delta OD^\lambda(\Delta t) = -\log_{10} \left(\frac{I^\lambda(t_1)}{I^\lambda(t_0)} \right) = \sum_i \varepsilon_i^\lambda \Delta c_i DPF^\lambda d, \quad (\text{A.3})$$

where

$$\Delta c_i = c_i(t_1) - c_i(t_0) \quad (\text{A.4})$$

is the temporal change in chromophore concentration. Given that oxygenated and deoxygenated hemoglobin are dominant chromophores, we get from Equation A.3

$$\Delta OD^\lambda(\Delta t) = DPF^\lambda d \left(\epsilon_{\text{Hb}}^\lambda \Delta[\text{Hb}] + \epsilon_{\text{HbO}_2}^\lambda \Delta[\text{HbO}_2] \right). \quad (\text{A.5})$$

To obtain $\Delta[\text{Hb}]$ and $\Delta[\text{HbO}_2]$, Equation A.5 is evaluated at two or more wavelengths. The resulting matrix formula for two wavelengths λ_1 and λ_2 is

$$\begin{pmatrix} \Delta[\text{Hb}] \\ \Delta[\text{HbO}_2] \end{pmatrix} = \begin{pmatrix} \epsilon_{\text{Hb}}^{\lambda_1} & \epsilon_{\text{HbO}_2}^{\lambda_1} \\ \epsilon_{\text{Hb}}^{\lambda_2} & \epsilon_{\text{HbO}_2}^{\lambda_2} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\Delta OD^{\lambda_1}}{DPF^{\lambda_1} d} \\ \frac{\Delta OD^{\lambda_2}}{DPF^{\lambda_2} d} \end{pmatrix} \quad (\text{A.6})$$

with the exact solution:

$$\begin{aligned} \Delta[\text{Hb}] &= \frac{\epsilon_{\text{HbO}_2}^{\lambda_1} \frac{\Delta OD^{\lambda_2}}{DPF^{\lambda_2} d} - \epsilon_{\text{HbO}_2}^{\lambda_2} \frac{\Delta OD^{\lambda_1}}{DPF^{\lambda_1} d}}{\epsilon_{\text{HbO}_2}^{\lambda_1} \epsilon_{\text{Hb}}^{\lambda_2} - \epsilon_{\text{HbO}_2}^{\lambda_2} \epsilon_{\text{Hb}}^{\lambda_1}}, \\ \Delta[\text{HbO}_2] &= \frac{\epsilon_{\text{Hb}}^{\lambda_1} \frac{\Delta OD^{\lambda_2}}{DPF^{\lambda_2} d} - \epsilon_{\text{Hb}}^{\lambda_2} \frac{\Delta OD^{\lambda_1}}{DPF^{\lambda_1} d}}{\epsilon_{\text{Hb}}^{\lambda_1} \epsilon_{\text{HbO}_2}^{\lambda_2} - \epsilon_{\text{Hb}}^{\lambda_2} \epsilon_{\text{HbO}_2}^{\lambda_1}}. \end{aligned} \quad (\text{A.7})$$

The dMBLL is based on two assumptions: a homogeneous change in chromophore concentrations in homogeneous medium and constant scattering loss. The first assumption does not hold true for measurements of the head. The head is not a homogeneous medium. However, this is not a problem because the inhomogeneity remains constant and is mostly covered by the constant G , which cancels out. On the other hand, the concentration change in oxygenated and deoxygenated hemoglobin is not homogeneous, that is it occurs only in the brain and not in other tissues such as skin and skull. This leads to an error in quantification. The MBLL

strongly underestimates the size of changes in oxygenated and deoxygenated hemoglobin. In principle, this error can be corrected by taking partial differential path lengths into account. This can be achieved by incorporating anatomical volumetric data from fMRI or computerized tomography to define the spatial boundaries of the skull and scalp. Additionally, DPF was found to be age, gender, and wavelength dependent and varies between participants.

Appendix B

Potential pitfalls of widely used implementations of common spatial patterns

During the course of my PhD research, we have uncovered serious flaws in handling EEG signals with a decreased rank in implementations of the CSP algorithm. The CSP algorithm assumes covariance matrices of the signal to have full rank. However, preprocessing techniques, such as artifact removal using independent component analysis, may decrease the rank of the signal, leading to potential errors in the CSP decomposition. We inspect what could go wrong when CSP implementations do not take this into consideration on a binary motor imagery classification task. We review CSP implementations in open-source toolboxes for EEG signal analysis (FieldTrip, BBCI Toolbox, BioSig, EEGLAB, BCILAB, and MNE). We show that the unprotected implementations decreased mean classification accuracy by up to 32%, with spatial filters resulting in complex numbers, for which corresponding spatial patterns do not have a clear interpretation. We encourage researchers to check their implementations and analysis pipelines.

This appendix has been published in [\[222\]](#).

B.1 Introduction

The CSP algorithm [169, 32, 211] is a popular supervised decomposition method for EEG signal analysis, which is used to distinguish between two classes (conditions). It finds spatial filters that maximize the signal variance for one class, while simultaneously minimizing the signal variance for the opposite class. Derivations of CSP assume that two covariance matrices of two classes have full rank. However, preprocessing techniques may decrease the rank of the signal. This issue is highlighted using independent component analysis (ICA) for artifact removal [112, 106]. This appendix focuses on ICA but this message applies for any technique with this feature.

A general solution has been shown in the literature [188]. Nevertheless, several studies using ICA for artifact removal followed by CSP in the original channel space struggle with this issue and often walk around the problem. For instance, work reported in [276] computed CSP on independent components (ICs) to avoid the issue, while work in [74] used a preprocessing transformation to obtain a full rank matrix. On the other hand, many studies use the same pipeline without mentioning the issue. This issue is highlighted here because in many cases it is unclear how other authors have solved this issue.

This appendix shows what could happen when CSP implementations do not take this issue into consideration on a binary motor imagery classification task of EEG trials from the BCI competition III dataset IVa [31]. Additionally, we review implementations in open-source toolboxes for EEG signal analysis.

B.2 Common spatial patterns

Here, we review the CSP algorithm and its two main implementation approaches. We assume that the EEG is already band-pass filtered and centered. Let $X_i \in \mathbb{R}^{C \times T}$ be the EEG signal of trial i where C is the number of channels and T is the number of samples per trial. We compute

the spatial covariance $R_1 \in \mathbb{R}^{C \times C}$ by averaging over trials of class 1:

$$R_1 = \frac{1}{|\mathcal{S}_1|} \sum_{i \in \mathcal{S}_1} \frac{X_i X_i^T}{\text{trace}(X_i X_i^T)} \quad (\text{B.1})$$

where \mathcal{S}_1 is the set of indices corresponding to trials belonging to class 1, $|\mathcal{S}_1|$ denotes the size of the set \mathcal{S}_1 , and trace is the trace of a matrix. The spatial covariance R_2 matrix is calculated equivalently for class 2. In the following derivations of CSP, we assume that R_1 and R_2 have full rank (i.e., $\text{rank}(R_1) = \text{rank}(R_2) = C$).

The goal of CSP is to find a decomposition matrix $W \in \mathbb{R}^{C \times C}$ that projects the signal $x(t) \in \mathbb{R}^C$ in the original channel space to $x_{CSP}(t) \in \mathbb{R}^C$ as follows:

$$x_{CSP}(t) = W^T x(t) \quad (\text{B.2})$$

with the following properties:

$$W^T R_1 W = D_1 \quad (\text{B.3})$$

$$W^T R_2 W = D_2 \quad (\text{B.4})$$

and scaling such that

$$D_1 + D_2 = I_C \quad (\text{B.5})$$

where $I_C \in \mathbb{R}^{C \times C}$ is the identity matrix. In other words, R_1 and R_2 share the same eigenvectors and the sum of the corresponding eigenvalues is always 1. The eigenvector with the largest eigenvalue for class 1 has the smallest eigenvalue for class 2 and vice-versa. Columns of W are spatial filters while columns of a matrix $A = (W^T)^{-1}$ represent spatial patterns.

B.2.1 Geometric approach

We factorize the composite spatial covariance $R_1 + R_2$ as

$$R_1 + R_2 = EFE^T \quad (\text{B.6})$$

where E is the orthogonal matrix of eigenvectors (in columns) and F is the diagonal matrix of their corresponding eigenvalues. We define the whitening transformation matrix U as

$$U = F^{-1/2}E^T \quad (\text{B.7})$$

and whiten matrix R_1

$$S_1 = UR_1U^T. \quad (\text{B.8})$$

We factorize matrix S_1 as

$$S_1 = PD_1P^T \quad (\text{B.9})$$

where P is the orthogonal matrix of eigenvectors and D_1 is the diagonal matrix of their corresponding eigenvalues. We define the decomposition matrix W^T as

$$W^T = P^T U. \quad (\text{B.10})$$

Then this W satisfies (B.3) and also (B.4) using (B.5).

B.2.2 Generalized eigenvalue problem

We can directly solve W by getting W^T from (B.5) [188] and by inserting this into (B.3) we get

$$R_1 W = D_1(R_1 + R_2)W, \quad (\text{B.11})$$

which is an equation of the generalized eigenvalue problem.

B.2.3 Covariance matrices without full rank

If the covariance matrices R_1 and R_2 do not have full rank, the above CSP derivations do not hold. Putting aside mathematical incorrectness, what could go wrong in their direct implementations?

In the geometric approach, the first eigendecomposition in (B.6) may have some zero eigenvalues. In the case of using ICA for artifact removal, the number of zero eigenvalues equals the number of removed ICs. The whitening transformation U in (B.7) is undefined due to division by zero. We can remove dimensions with zero eigenvalues at this point and the rest would work. This is similar to dimensionality reduction by principal component analysis (PCA) before CSP.

In the generalized eigenvalue problem approach, the generalized eigendecomposition in (B.11) may have a complex solution. The complex spatial filters and their corresponding complex spatial patterns do not have a clear interpretation.

In both cases, the EEG should be first projected into a space with the number of dimensions equal to the rank of the EEG before CSP decomposition. We use PCA here, see [188] for a general solution and [32] for the difference of the CSP solution on spatially filtered data. The covariance matrices will have full rank in this space. Note that, to compute spatial patterns on the original EEG channels, we must first multiply W^T with the PCA transformation matrix before the inversion.

B.3 Methods

B.3.1 Evaluation

We evaluated CSP implementations on a binary motor imagery classification task using ICA for artifact removal in a preprocessing step. We chose this pipeline because it is commonly used and the complex number problem is hidden by the classifier. We used public dataset IVa from the BCI competition III [31]. The single-trial EEG signals were recorded from five healthy participants during imagination of right hand and right foot movement without feedback (140 trials per class), see [31] for more details.

First, EEG signals were preprocessed by removing artifacts. EEG signals were FIR band-pass filtered between 1–40 Hz to remove slow drifts in the signal and high-frequency noise. For each participant, ICA (FastICA) was trained on all windows 0–4.5 s after the task onset. Artifactual components were identified by thresholding peak amplitudes of the EEG time series [183, 75]. A scalp projection of each IC was thresholded to $\pm 100 \mu\text{V}$ and peak-to-peak differences between maximum and minimum amplitudes in each window and channel were thresholded to $60 \mu\text{V}$, with ICs exceeding any criterion marked for removal (35, 32, 37, 63, and 23 ICs were removed from a total of 118 for the 5 participants).

We adopted the winning solution from the BCI competition using CSP [270]. EEG signals were FIR band-pass filtered between 12–14 Hz and trials were extracted from 0.5–4.5 s after the task onset. Additionally, dimensionality reduction by PCA on all trials was used if it was required by a particular CSP implementation.

For each CSP implementation, We trained a classifier using stratified 10-fold cross-validation and measured classification accuracy from all test folds. CSP was trained and even numbers of CSP components, columns of W , from 2 to 20 were selected. Components were ordered by sorting their corresponding eigenvalues in ascending order and the first $k/2$ and last $k/2$ components were selected for a desired k components. We choose 20 as a maximum

because we have not seen any usage of more CSP components reported. The winning solution from the BCI competition, which was adopted here, used 2 CSP components. The classifier features were the log-variance of the selected components [169, 32, 211]. The LDA and SVM (with the radial basis function kernel and the regularization parameter $C = 1$) were used as classifiers.

B.3.2 Implementations

In Python, there are two eigendecomposition methods, for the geometric approach, *eig* and *eigh* in the numpy and scipy packages. Implementations in both packages have the same behavior. The *eigh* is a specialized method for a real symmetric matrix, which always produces a real solution, while *eig* is for a general matrix. Nevertheless, *eigh* does not check this assumption and only uses the upper or lower triangular part. Similarly, two generalized eigendecomposition methods *eig* and *eigh* are implemented in the scipy package with the same logic as above. The *eigh* method raises an exception when the second matrix, $R_1 + R_2$ in (B.11) here, is not definite positive. We will refer to the used method in parentheses. In Matlab, there is only one method *eig* for everything.

We tested CSP implementations in Python 3.6 and Matlab R2018b Update 4, both 64-bit, with all the above permutations. The classification pipeline was implemented in Python and the only difference was a particular CSP implementation that ran in its required environment, directly Python or Matlab in a subprocess.¹ In the geometric approach, the same eigendecomposition method is used for both eigendecompositions in (B.6) and (B.9), and with and without removing dimensions during the whitening step with eigenvalues smaller than 10^{-14} after (B.6). All these possibilities are used with and without dimensionality reduction by PCA before CSP. We refer to the CSP implementation with the correct mathematical background ‘protected’ CSP and otherwise refer to CSP as ‘unprotected’.

¹Source code for the classification pipeline with all CSP implementations is available at https://github.com/milan-rybar/csp_evaluation.

B.3.3 EEG Toolboxes

We review CSP implementations in popular open-source toolboxes for the EEG analysis in Matlab and Python.

FieldTrip

FieldTrip (v. 20191025, October 2019) [185] for Matlab implements a geometric approach but uses singular value decomposition instead of eigenvalue decomposition. During the whitening step, it removes dimensions with eigenvalues of absolute value smaller than 10^{-14} .

BBCI Toolbox

BBCI Toolbox (commit a30ce0bc8d, March 2019) [30] for Matlab implements a geometric approach. During the whitening step, it keeps dimensions with eigenvalues larger than the fraction 10^{-10} of the largest eigenvalue. It shows a warning of this dimensionality reduction when it is applied.

BioSig

BioSig (v. 3.6.0, April 2019) [261] for Matlab implements both approaches. Both of them are unprotected. The implementation always returns only four spatial filters, two per class, thus we excluded it from the evaluation.

EEGLAB, BCILAB

EEGLab [67] for Matlab has 2 plugins with CSP implementations. CSP plugin (v. 1.1) implements a geometric approach and BCILAB (v. 1.1) [126] solves a generalized eigenvalue problem. Both of them are unprotected. Both implementations are difficult to adapt for a programmatic evaluation on given trials outside their desired processing pipelines without taking the code outside, thus we excluded them from the evaluation.

MNE

MNE (v. 0.17.2, April 2019) [85] for Python solves a generalized eigenvalue problem. It uses the *eigh* method for real symmetric matrices, which raises an exception when the covariance matrix $R_1 + R_2$ in (B.11) is not definite positive.

B.4 Results

We mainly compare our CSP implementations in Python and Matlab because we can guarantee that they differ only in the inspected part. Other implementations may differ in their definition of the covariance matrix in (B.1) and selection of CSP components. For instance, MNE implements different selection criterion based on [32]. Their results are provided mainly as a reference. Two methods are compared by their differences in classification accuracies on each test fold from cross-validation (10) and each participant (5). Both the LDA and SVM classifiers gave similar results, thus only results for LDA are reported here as a simpler classifier. Figure B.1 shows mean classification accuracies for all tested CSP implementations.

B.4.1 Geometric approach

Unprotected Python (*eigh*) encountered division by zero in (B.7) due to zero eigenvalues. On the other hand, unprotected Python (*eig*) is a little bit tricky. The *eig* function has a complex solution in (B.6) and it depends on the particular data how “close” the complex eigenvalues are to zero, as to whether it raises division by zero or not in (B.7).

Reducing dimensionality during the whitening step (i.e, removing eigenvectors with zero eigenvalues after (B.6)) or dimensionality reduction by PCA before the CSP (or both together) had equivalent results on the classification pipeline for any classifier, Python (*eig*, *eigh*) and Matlab implementation. We will group their results together and call them protected regardless what dimensionality reduction is used. However, Python (*eig*) with dimensionality reduction

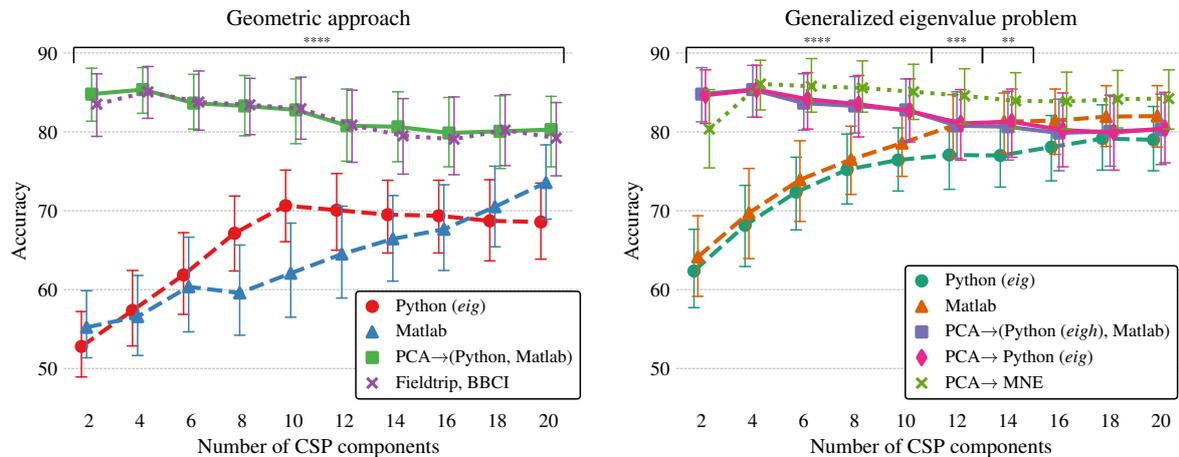


Fig. B.1 Mean classification accuracy over all test folds from cross-validation (10) and all participants (5) for different CSP implementations. Unprotected CSP implementations are shown with dashed lines, protected with solid lines, and external implementations with dotted lines. Error bars are 95% confidence intervals. Points for a particular number of CSP components are offset on the X axis for better visibility. Intervals of numbers of CSP components with asterisks represent statistically significant differences between the CSP implementations for the particular number of CSP components, as assessed by Friedman test, where **** is $p < 0.0001$, *** is $p < 0.001$, and ** is $p < 0.01$. ‘PCA→’ denotes dimensionality reduction by PCA before the CSP algorithm.

during the whitening step (without PCA) returned W and eigenvalues expressed in complex numbers but with zero imaginary parts. Matlab did not have this issue. Additionally, these methods also gave equivalent results to protected Python (*eigh*) and protected Matlab in the generalized eigenvalue problem approach.

The difference in classification accuracy between unprotected Python (*eig*) and protected Python (*eig*, *eigh*) or Matlab, which were equal as described above, was statistically significant when using any number of CSP components ($p < 0.0001$, one-sided Wilcoxon signed-rank test with Pratt modification for zero-differences). The difference decreased from 32 ± 2.1 (mean \pm standard error) for 2 components to 11.7 ± 1.3 for 20 components. The difference between unprotected Matlab and protected Matlab was statistically significant for 2 to 16 components ($p < 0.0001$), and 18 components ($p < 0.01$). The difference decreased from 29.5 ± 2.2 for 2

components to 9.5 ± 2.7 for 18 components. Both unprotected Python (*eig*) and Matlab had complex solutions, thus the resulting W and eigenvectors were complex.

B.4.2 Generalized eigenvalue problem

Unprotected Python (*eigh*) raised an exception because the covariance matrix $R_1 + R_2$ is not definite positive.

The difference in classification accuracy between unprotected Python (*eig*), protected Python (*eig*), and protected Python (*eigh*) was statistically significant for 2 to 12 components ($p < 0.0001$, Friedman test), and for 14 components ($p < 0.01$). The difference in classification accuracy between unprotected Python (*eig*) and protected Python (*eigh*) or Matlab, which were equal to protected CSP in geometric approach as described above, was statistically significant for 2 to 10 components ($p < 0.0001$, one-sided Wilcoxon test) and for 12 to 14 components ($p < 0.01$). The difference decreased from 22.4 ± 2.2 for 2 components to 3.6 ± 1.4 for 14 components. Similarly, the difference between unprotected Python (*eig*) and protected Python (*eig*) was statistically significant for 2 to 10 components ($p < 0.0001$), for 12 components ($p < 0.001$), and for 14 components ($p < 0.01$). The difference decreased from 22.2 ± 2.2 for 2 components to 4.2 ± 1.3 for 14 components. The difference between unprotected Matlab and protected Matlab was statistically significant for 2 to 6 components ($p < 0.0001$), 8 components ($p < 0.01$), and less for 10 components ($p < 0.05$). The difference decreased from 20.6 ± 2.4 for 2 components to 6.7 ± 2.1 for 8 components.

Unprotected Python (*eig*) and unprotected Matlab had a complex solution, thus the resulting W and eigenvectors were complex. Protected Python with *eig* and *eigh* did not have equivalent results, but they were not statistically different for any number of CSP components (two-sided Wilcoxon test). Surprisingly, protected Python (*eig*) had complex eigenvalues with zero imaginary parts but a W with real numbers (before components selection). Further inspection

showed that the difference in their eigenvalues D_1 was always less than 10^{-12} but their W were different, not just with flipped signs or reverse ordering.

B.5 Conclusion

We showed that unprotected CSP implementations in Python and Matlab can significantly decrease accuracy on a binary motor imagery classification task. Our results suggest that the less CSP components that are used the higher the decrease in classification accuracy between protected and unprotected CSP implementations. In Python, we strongly recommend using the *eigh* method in both CSP implementation approaches. We encourage researchers to check their implementations.

Appendix C

Additional figures for Chapter 4

This appendix complements the results from Chapter [4](#).

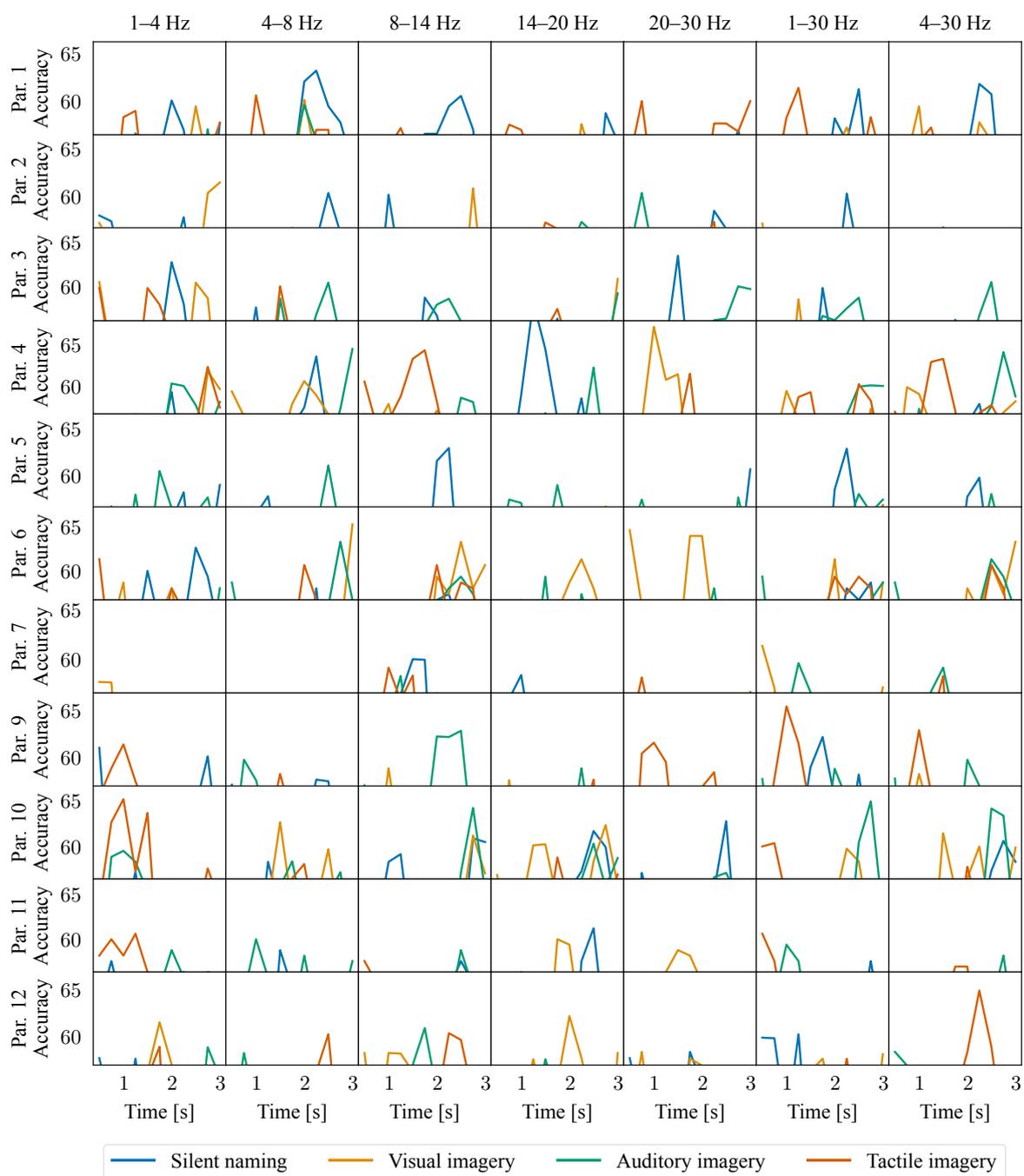


Fig. C.1 Classification accuracies obtained with the sliding temporal window of 0.5 second when the SVM ($C = 1$) can use information from all channels, see Section 4.1.2. Rows represent different participants, while columns indicate different frequency bands. The Y axes are limited to statistically significant classification accuracies from $p = 0.05$ to $p = 10^{-5}$.

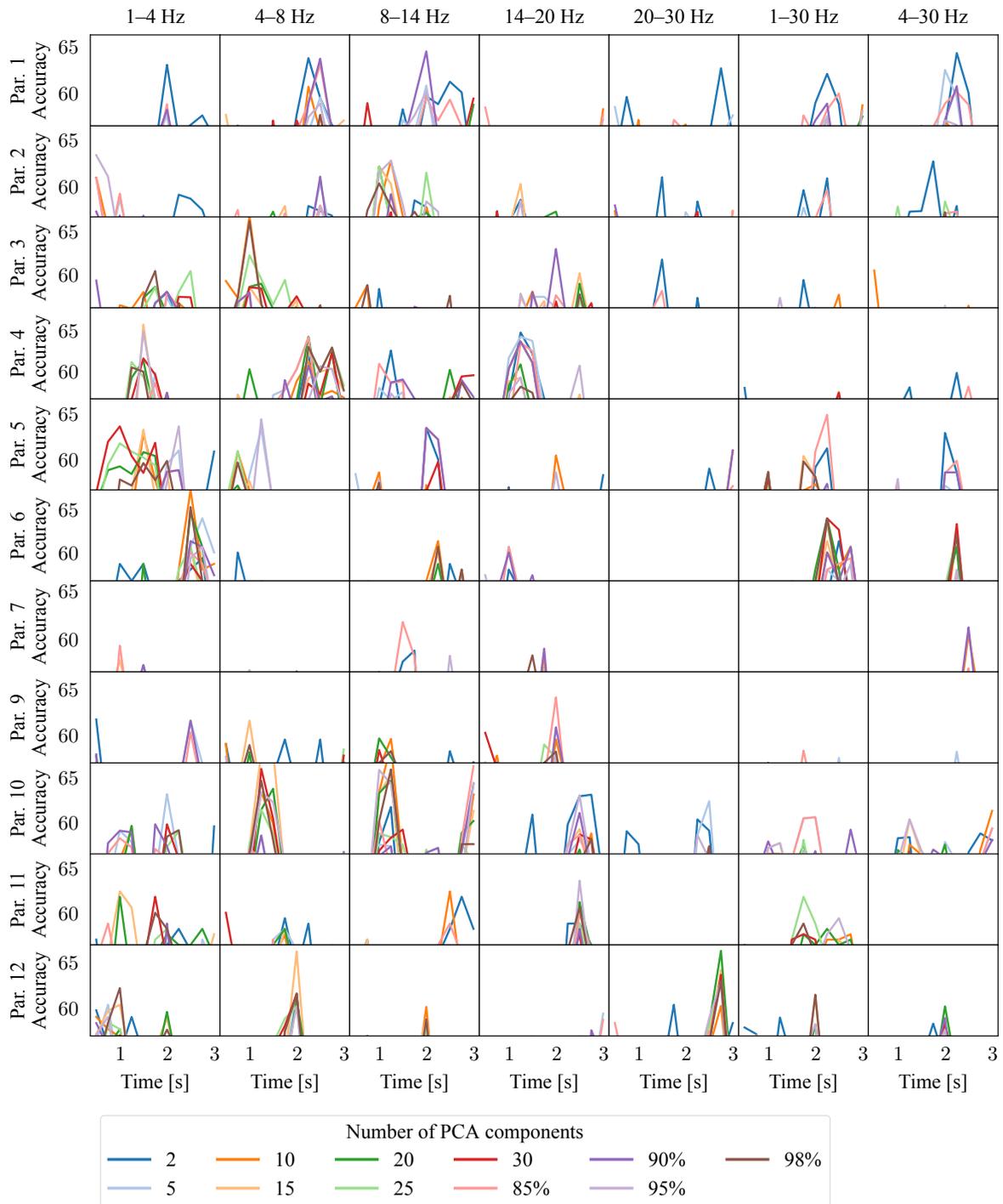


Fig. C.2 Classification accuracies for the silent naming task obtained with the sliding temporal window of 0.5 second when the SVM ($C = 1$) can use information from the first N PCA components, see Section 4.1.3. Rows represent different participants, while columns indicate different frequency bands. The Y axes are limited to statistically significant classification accuracies from $p = 0.05$ to $p = 10^{-5}$.

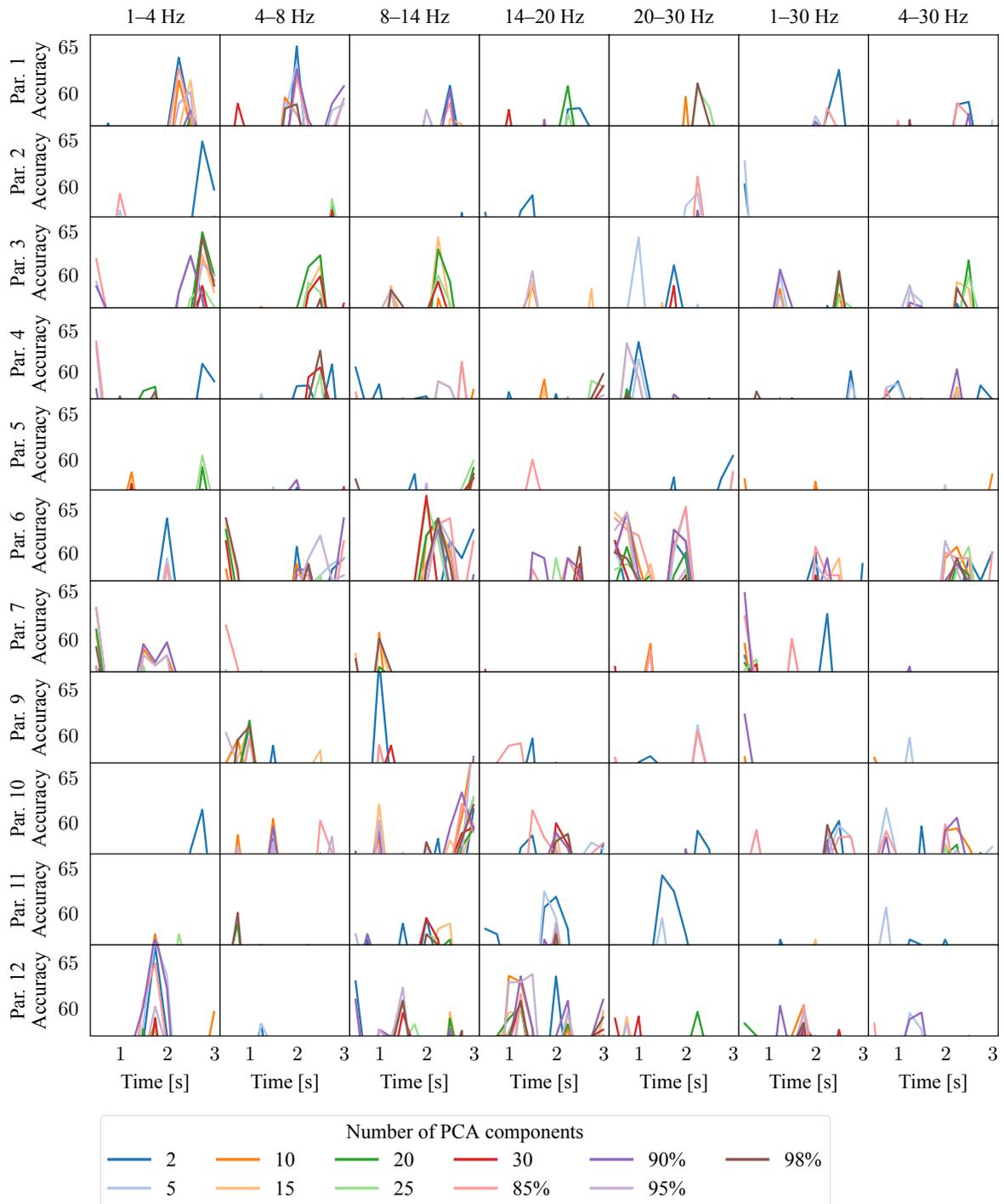


Fig. C.3 Classification accuracies for the visual imagery task obtained with the sliding temporal window of 0.5 second when the SVM ($C = 1$) can use information from the first N PCA components, see Section 4.1.3. Rows represent different participants, while columns indicate different frequency bands. The Y axes are limited to statistically significant classification accuracies from $p = 0.05$ to $p = 10^{-5}$.

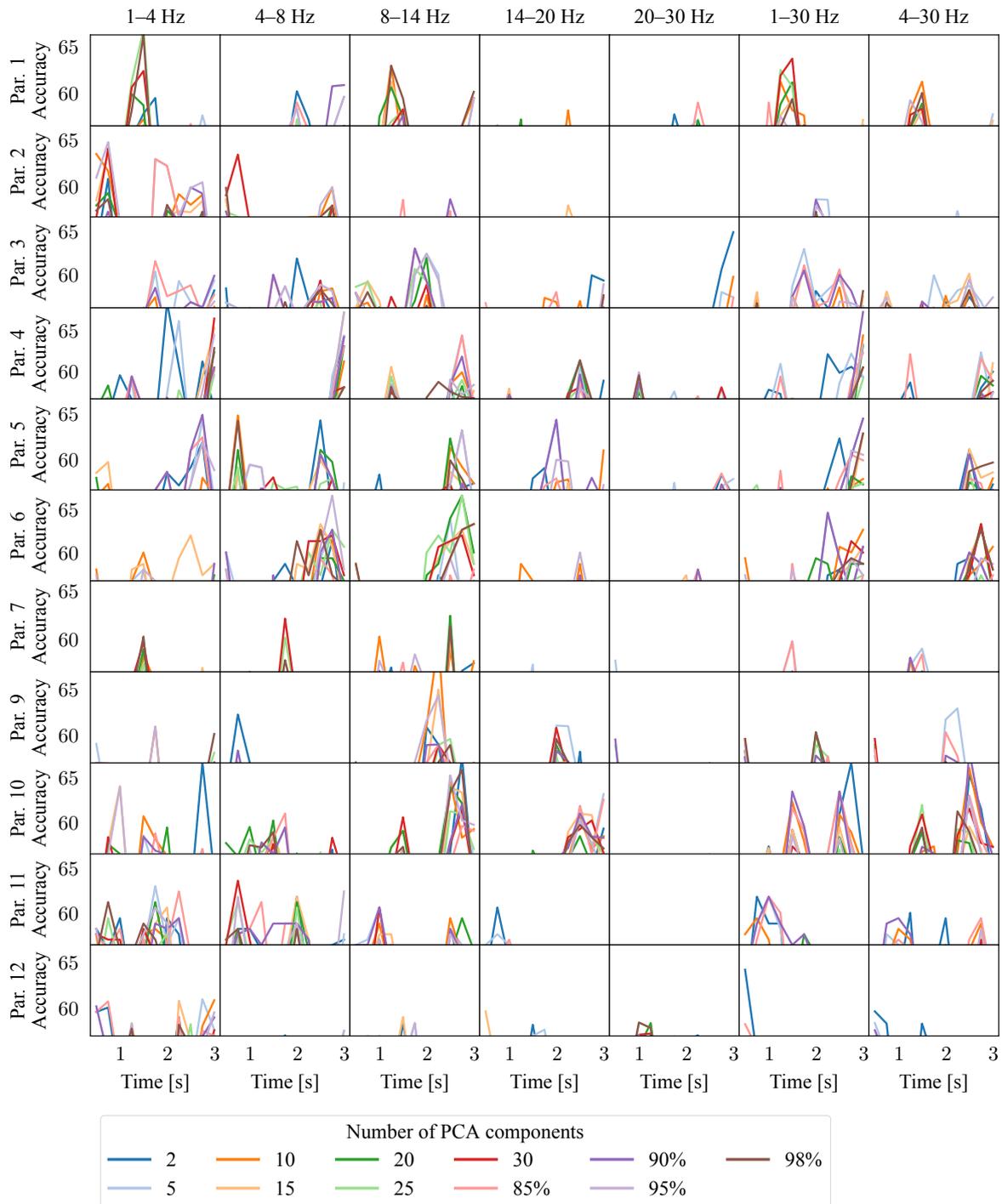


Fig. C.4 Classification accuracies for the auditory imagery task obtained with the sliding temporal window of 0.5 second when the SVM ($C = 1$) can use information from the first N PCA components, see Section 4.1.3. Rows represent different participants, while columns indicate different frequency bands. The Y axes are limited to statistically significant classification accuracies from $p = 0.05$ to $p = 10^{-5}$.

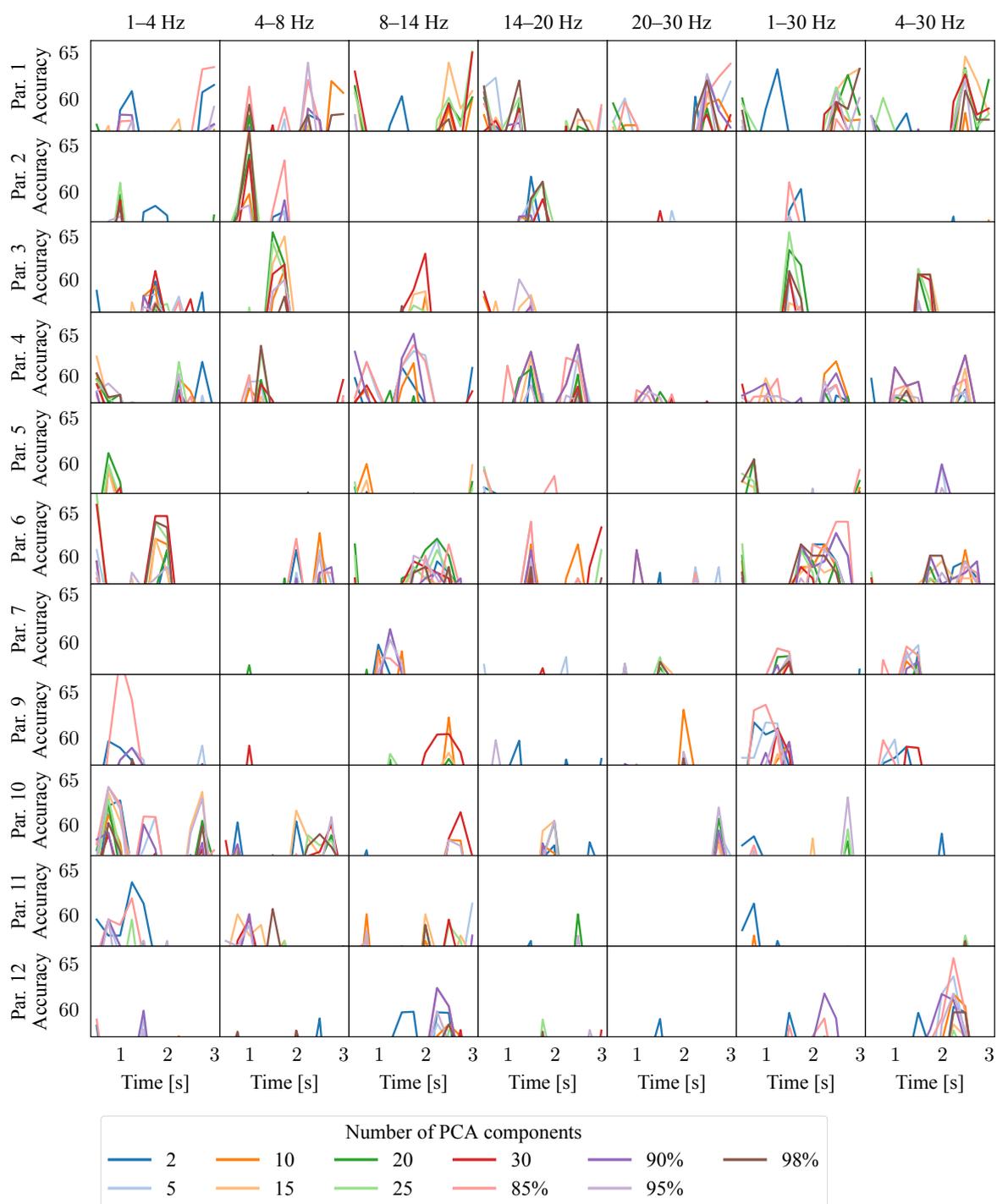


Fig. C.5 Classification accuracies for the tactile imagery task obtained with the sliding temporal window of 0.5 second when the SVM ($C = 1$) can use information from the first N PCA components, see Section 4.1.3. Rows represent different participants, while columns indicate different frequency bands. The Y axes are limited to statistically significant classification accuracies from $p = 0.05$ to $p = 10^{-5}$.

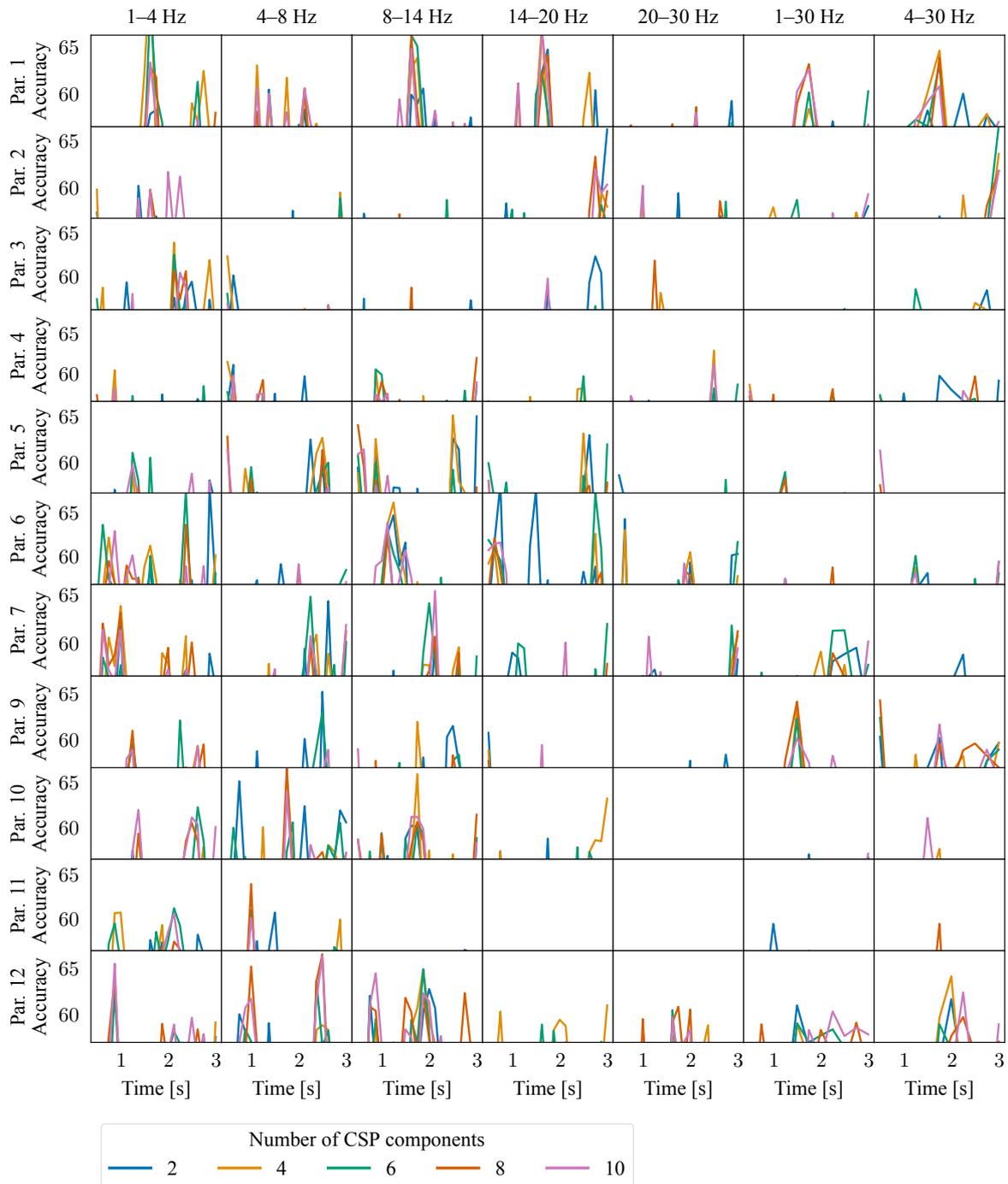


Fig. C.6 Classification accuracies for the silent naming task obtained with the sliding temporal window of 0.5 second when the SVM ($C = 1$) can use information from N CSP components, see Section 4.1.4. Rows represent different participants, while columns indicate different frequency bands. The Y axes are limited to statistically significant classification accuracies from $p = 0.05$ to $p = 10^{-5}$.

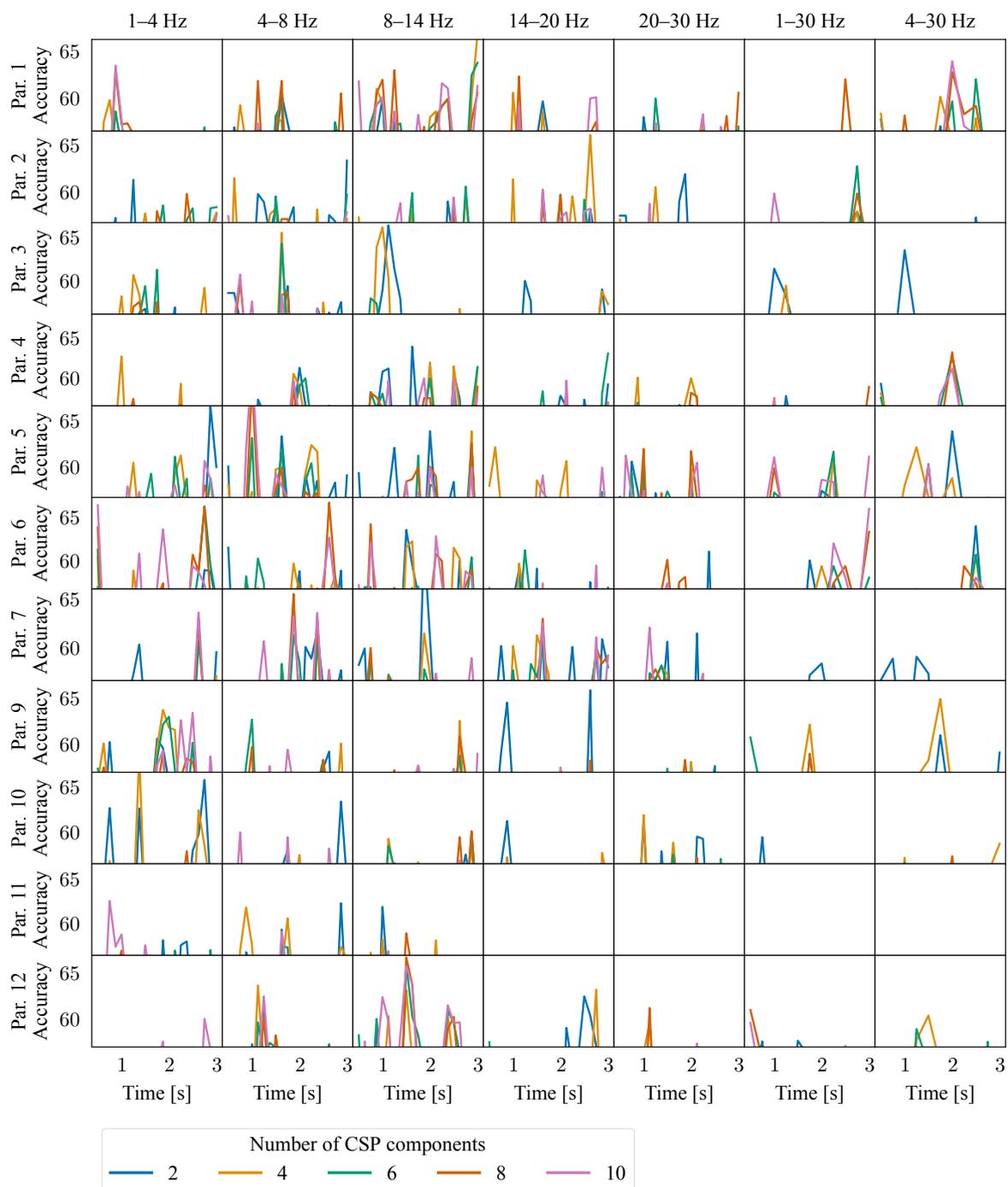


Fig. C.7 Classification accuracies for the visual imagery task obtained with the sliding temporal window of 0.5 second when the SVM ($C = 1$) can use information from N CSP components, see Section 4.1.4. Rows represent different participants, while columns indicate different frequency bands. The Y axes are limited to statistically significant classification accuracies from $p = 0.05$ to $p = 10^{-5}$.

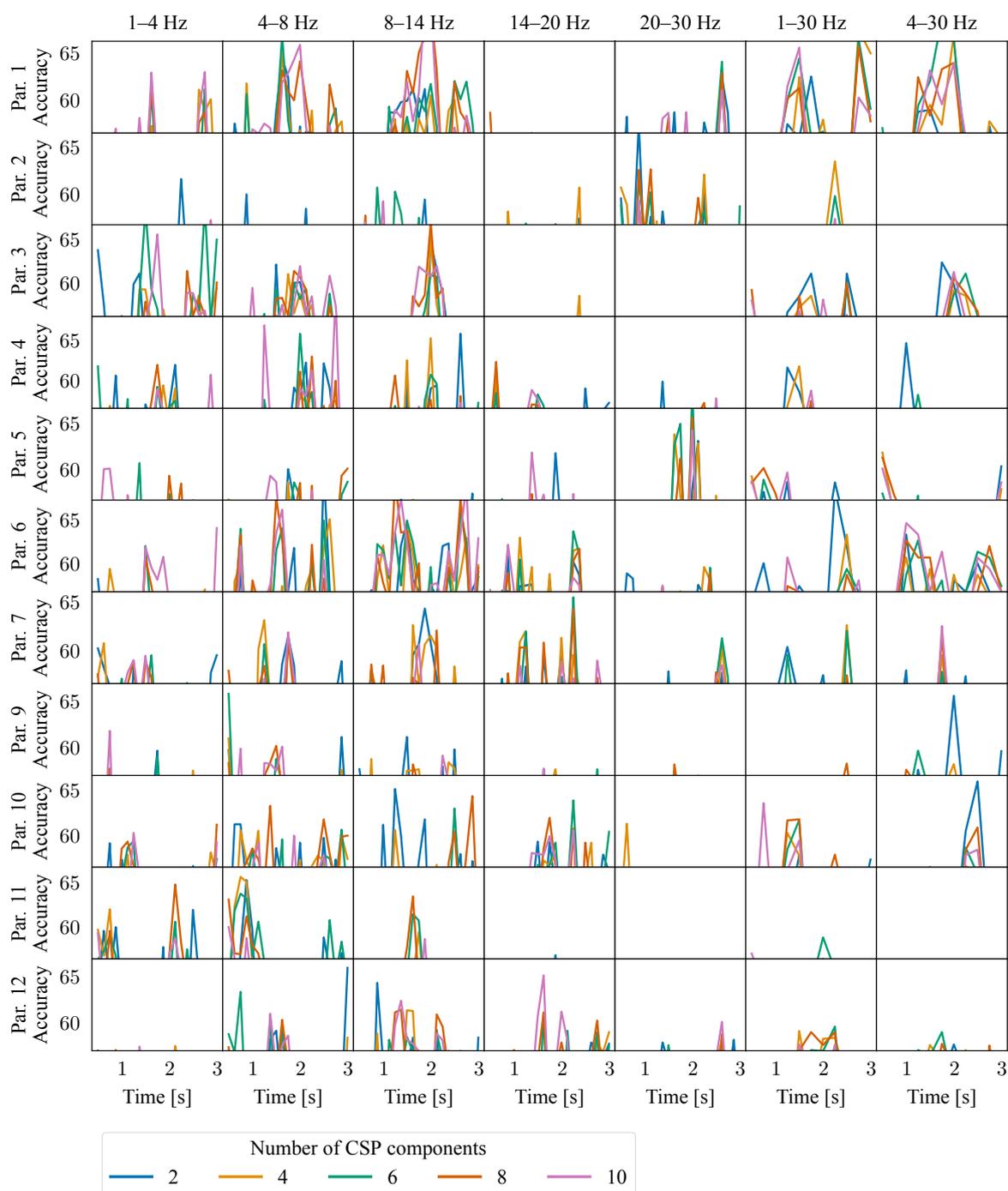


Fig. C.8 Classification accuracies for the auditory imagery task obtained with the sliding temporal window of 0.5 second when the SVM ($C = 1$) can use information from N CSP components, see Section 4.1.4. Rows represent different participants, while columns indicate different frequency bands. The Y axes are limited to statistically significant classification accuracies from $p = 0.05$ to $p = 10^{-5}$.

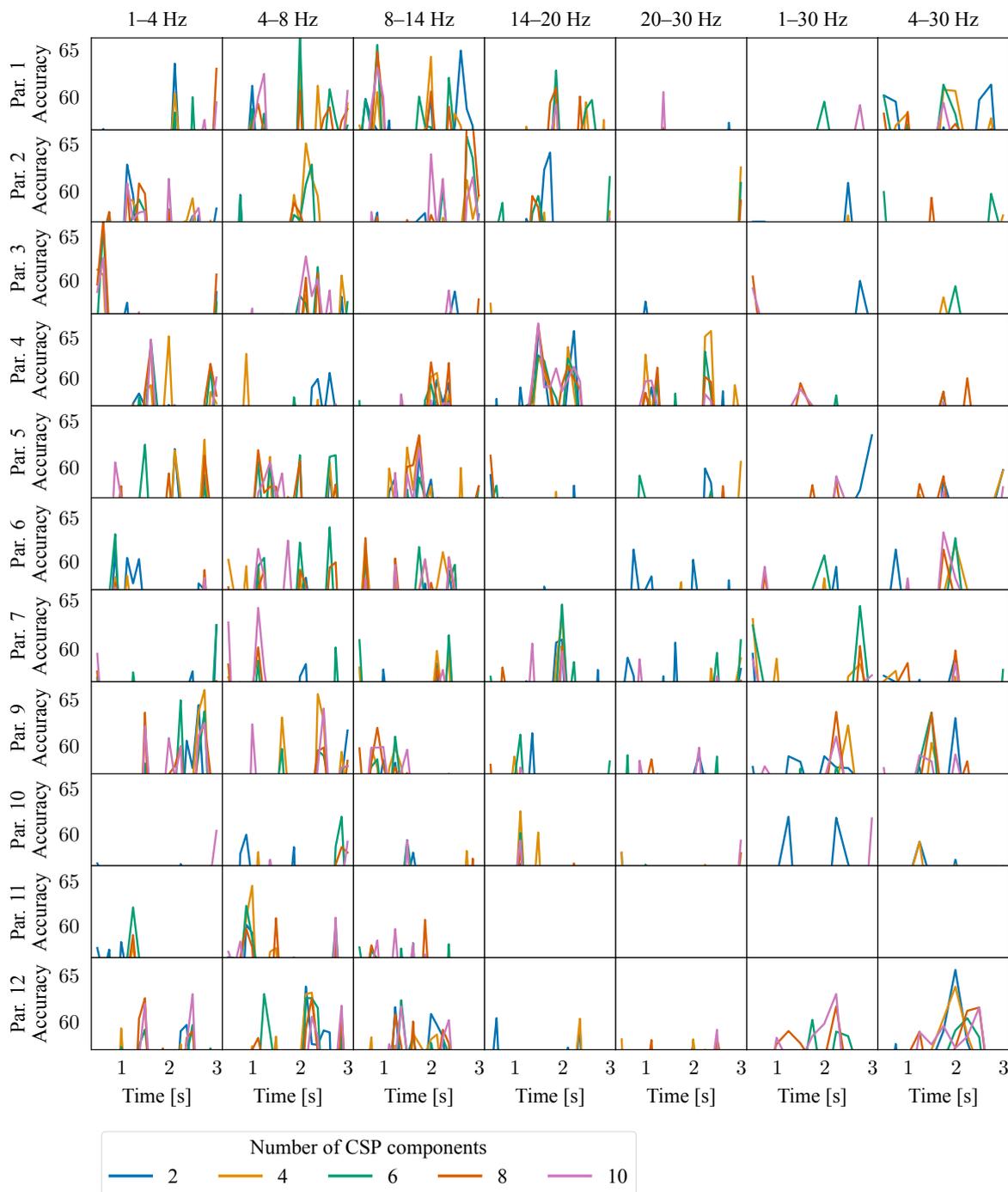


Fig. C.9 Classification accuracies for the tactile imagery task obtained with the sliding temporal window of 0.5 second when the SVM ($C = 1$) can use information from N CSP components, see Section 4.1.4. Rows represent different participants, while columns indicate different frequency bands. The Y axes are limited to statistically significant classification accuracies from $p = 0.05$ to $p = 10^{-5}$.

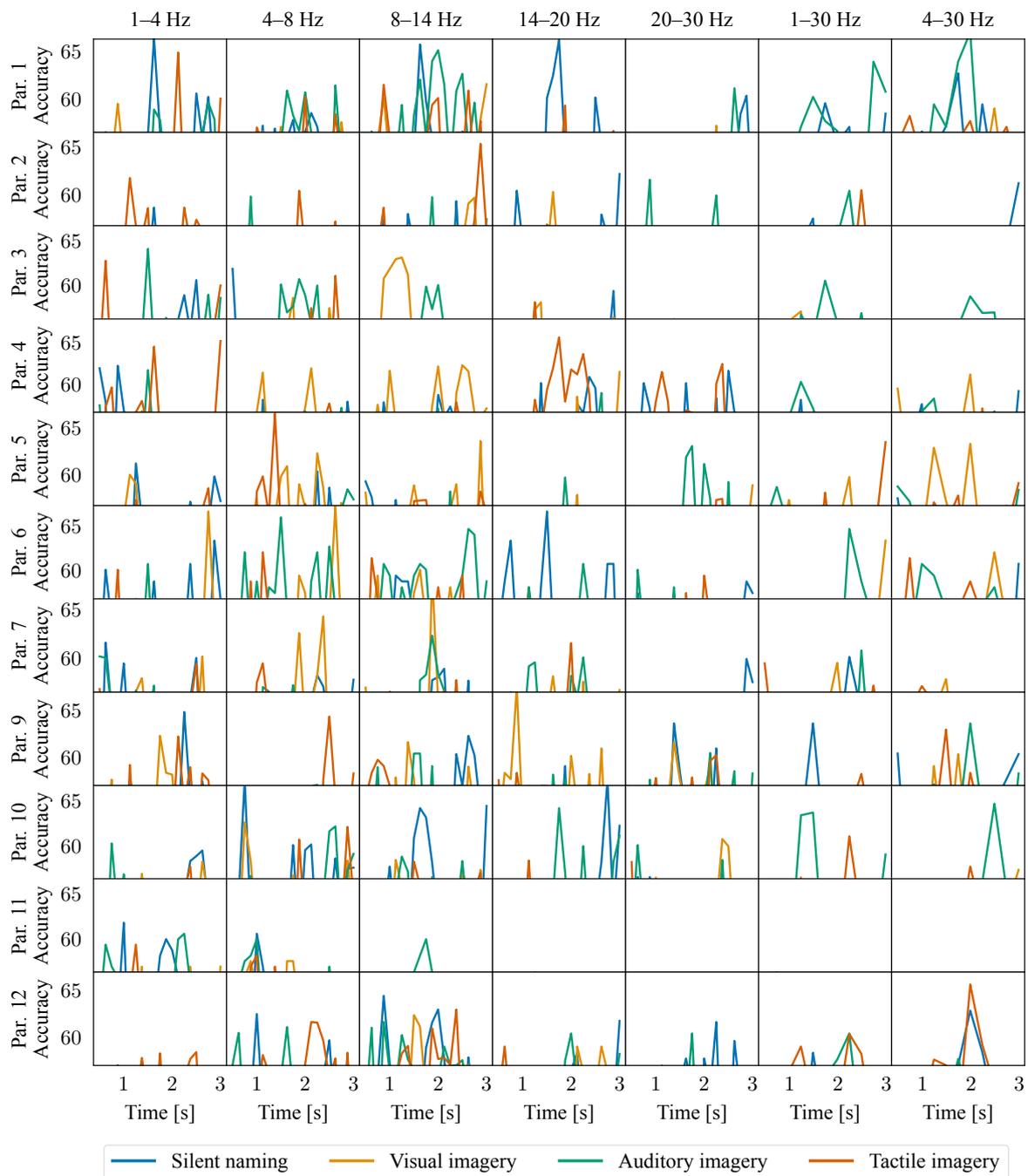


Fig. C.10 Classification accuracies obtained with the sliding temporal window of 0.5 second when the SVM ($C = 1$) can use information from the number of CSP components that was selected by the nested-cross-validation approach, see Section 4.1.4.2. Rows represent different participants, while columns indicate different frequency bands. The Y axes are limited to statistically significant classification accuracies from $p = 0.05$ to $p = 10^{-5}$.

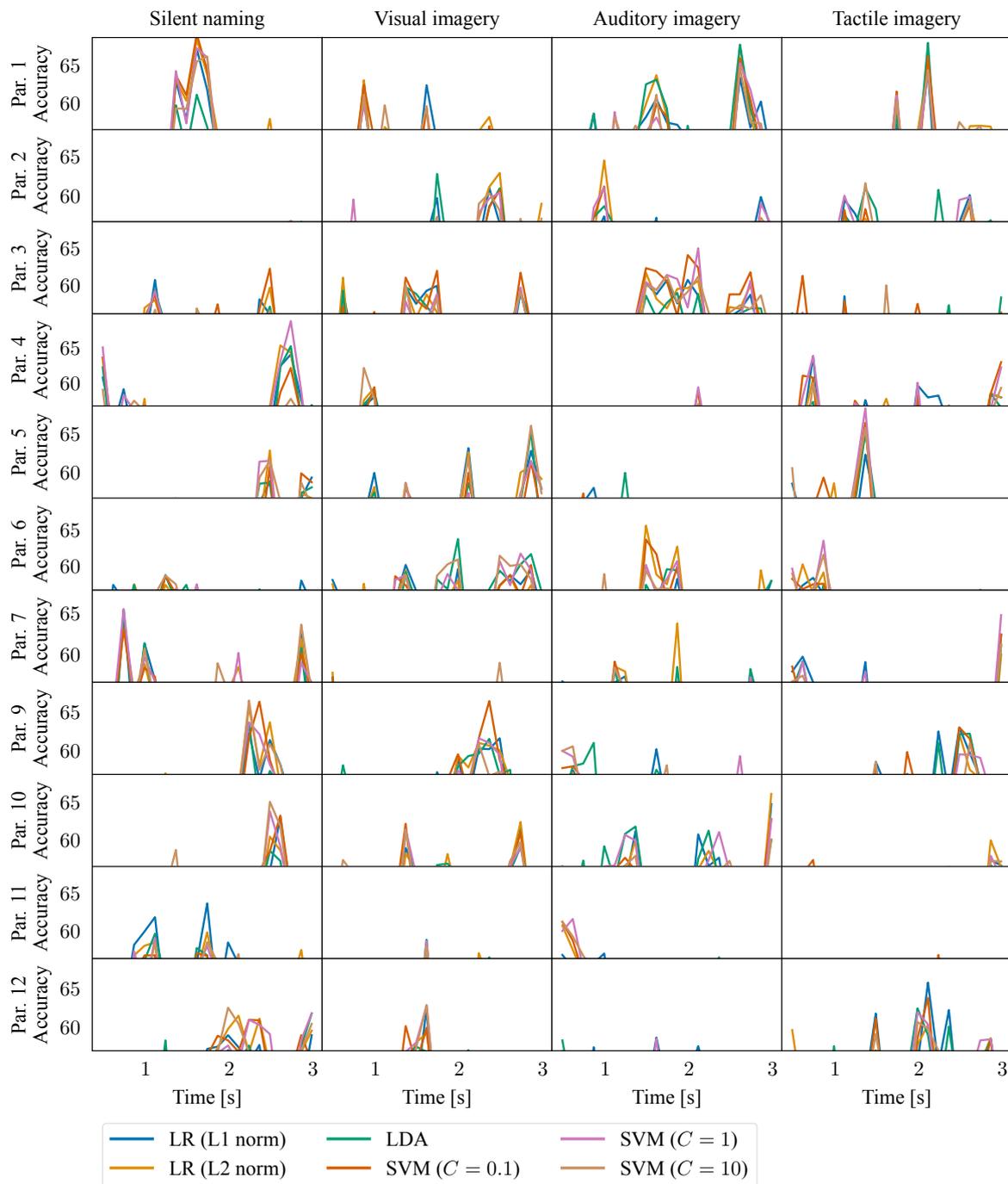


Fig. C.11 Classification accuracies obtained with the sliding temporal window of 0.5 second when the classifier can use information from the number of CSP components from the joint channel and frequency space that was selected by the nested-cross-validation approach, see Section 4.1.4.3. Rows represent different participants, while columns indicate different mental tasks. The Y axes are limited to statistically significant classification accuracies from $p = 0.05$ to $p = 10^{-6}$.

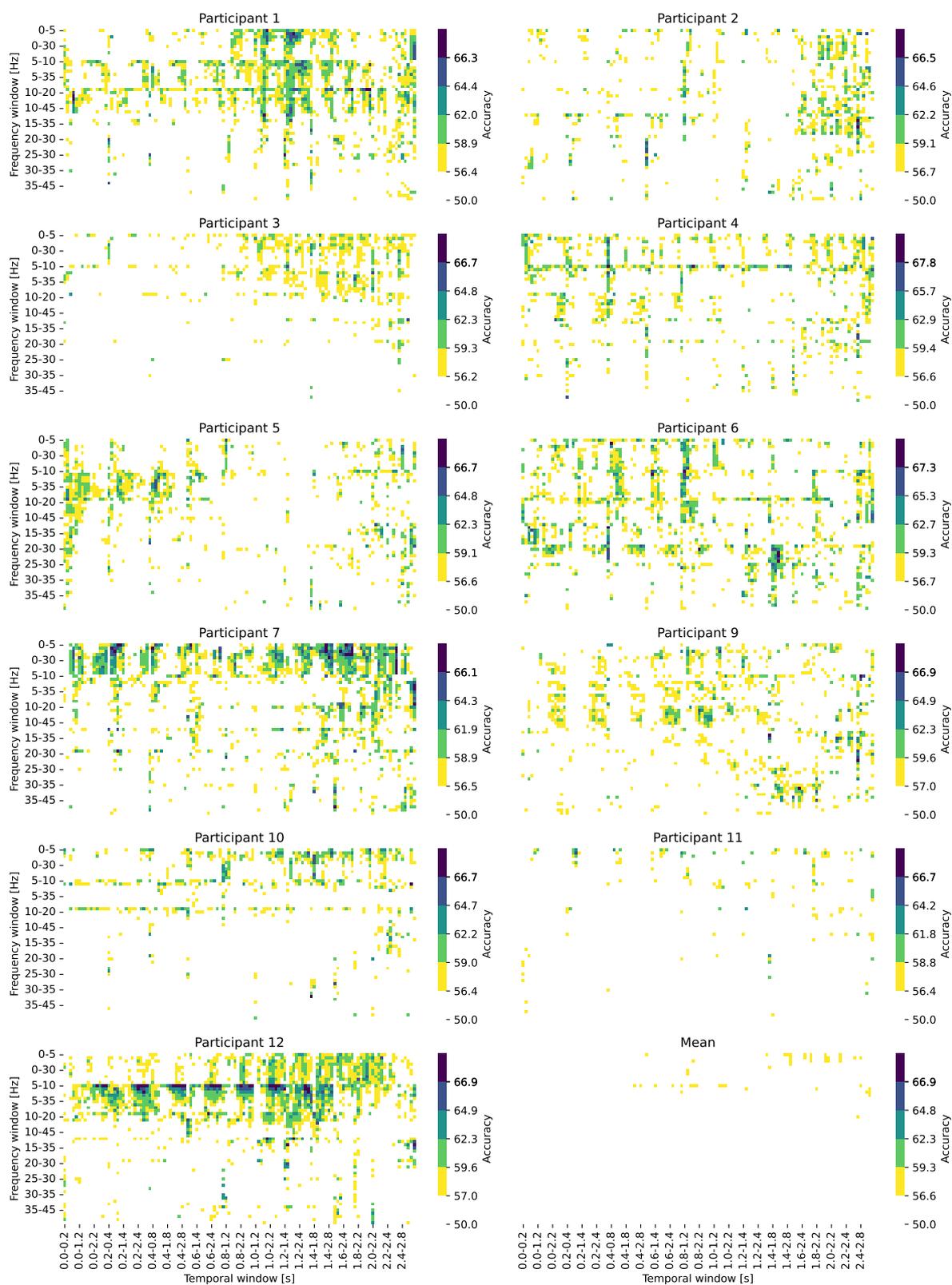


Fig. C.12 Maximal classification accuracies for the silent naming task for each time-frequency window from all tested N CSP components with the SVM ($C = 1$), see Section 4.1.4.4. Classification accuracies in colorbars represent significant borderlines for p -values of $p \in \{0.05, 0.01, 0.001, 0.0001, 0.00001\}$. White represents non-significant classification accuracies. Not all labels for X and Y axis are shown for simplicity.

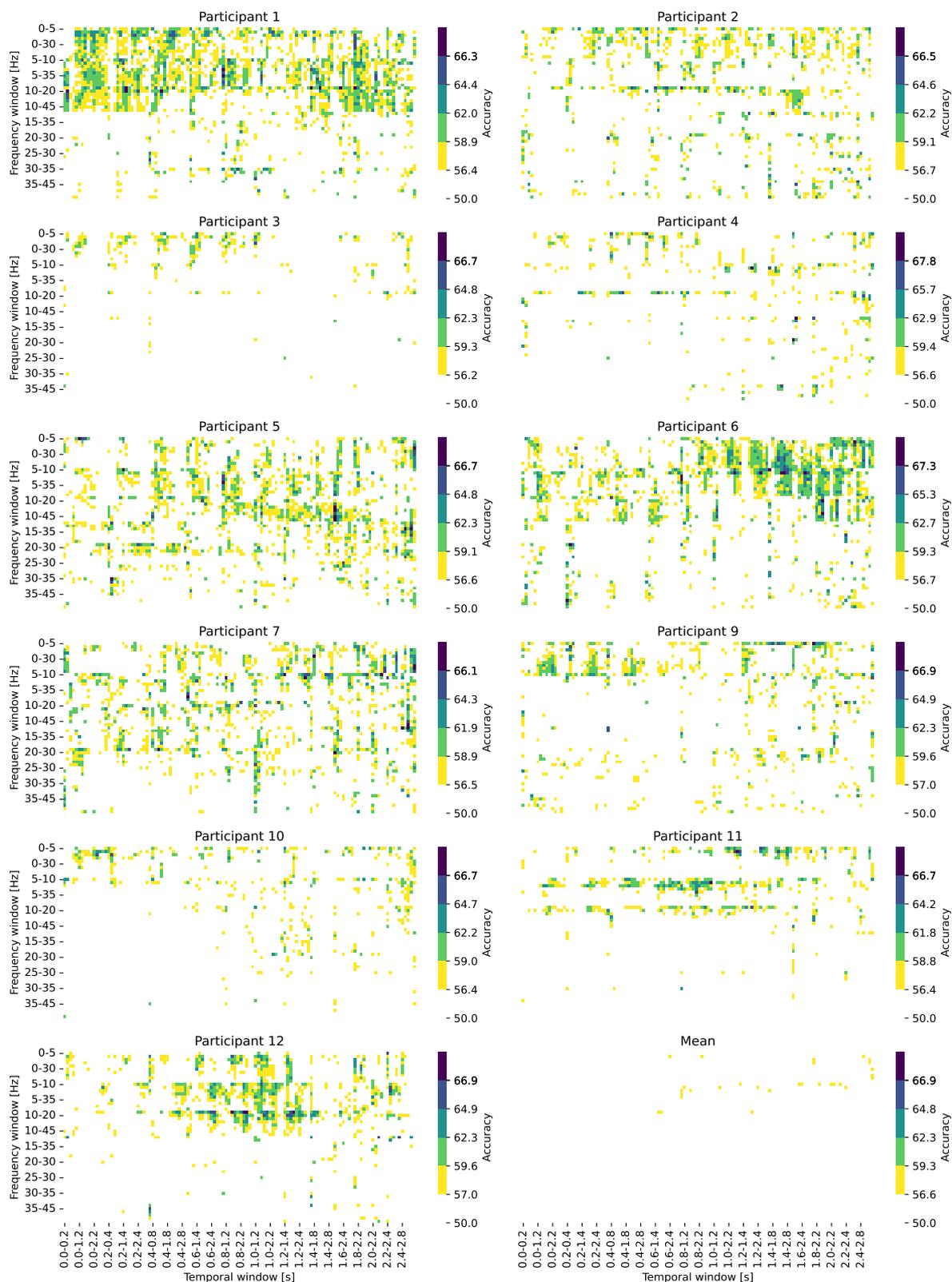


Fig. C.13 Maximal classification accuracies for the visual imagery task for each time-frequency window from all tested N CSP components with the SVM ($C = 1$), see Section 4.1.4.4. Classification accuracies in colorbars represent significant borderlines for p -values of $p \in \{0.05, 0.01, 0.001, 0.0001, 0.00001\}$. White represents non-significant classification accuracies. Not all labels for X and Y axis are shown for simplicity.

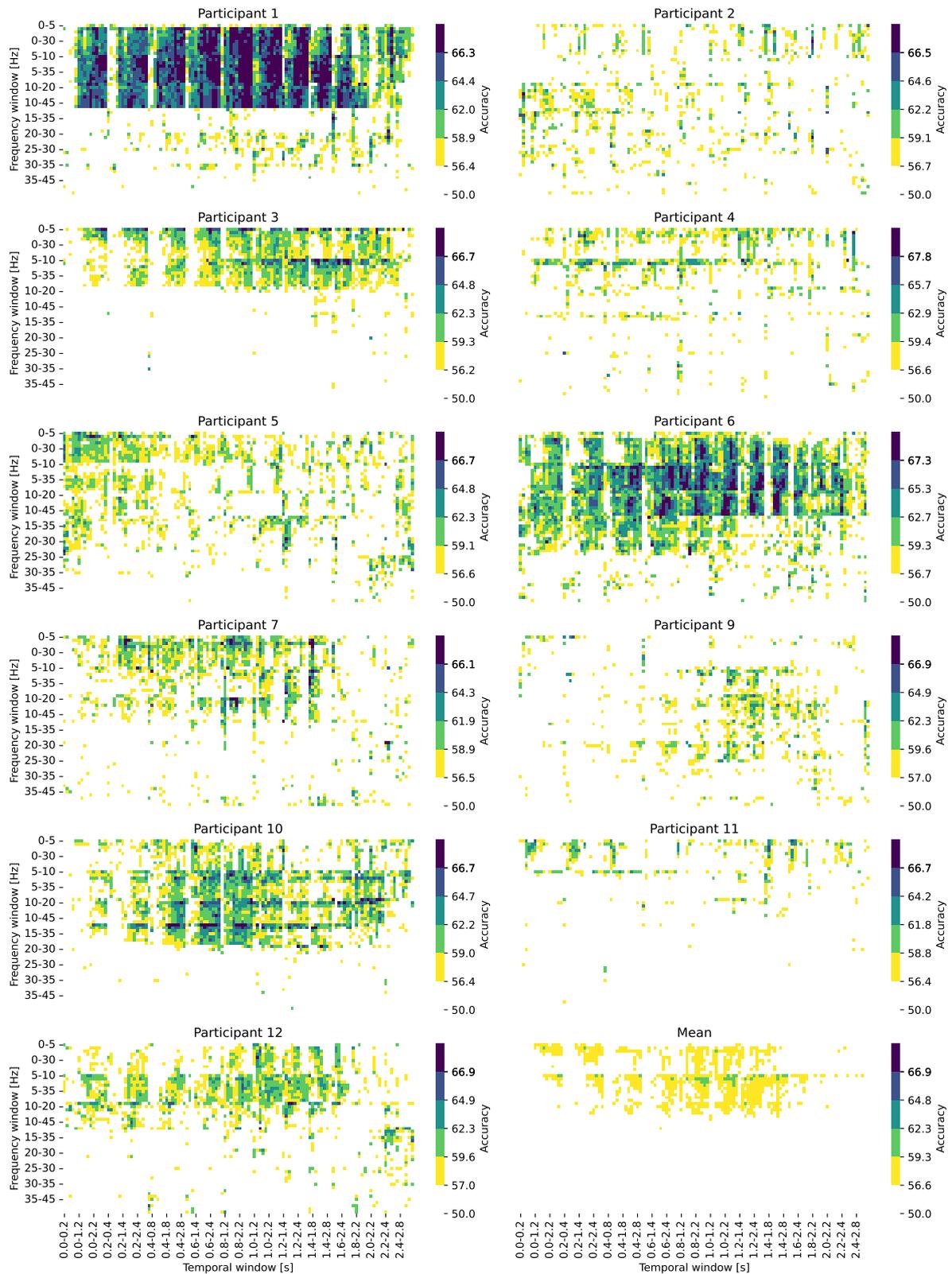


Fig. C.14 Maximal classification accuracies for the auditory imagery task for each time-frequency window from all tested N CSP components with the SVM ($C = 1$), see Section 4.1.4.4. Classification accuracies in colorbars represent significant borderlines for p -values of $p \in \{0.05, 0.01, 0.001, 0.0001, 0.00001\}$. White represents non-significant classification accuracies. Not all labels for X and Y axis are shown for simplicity.

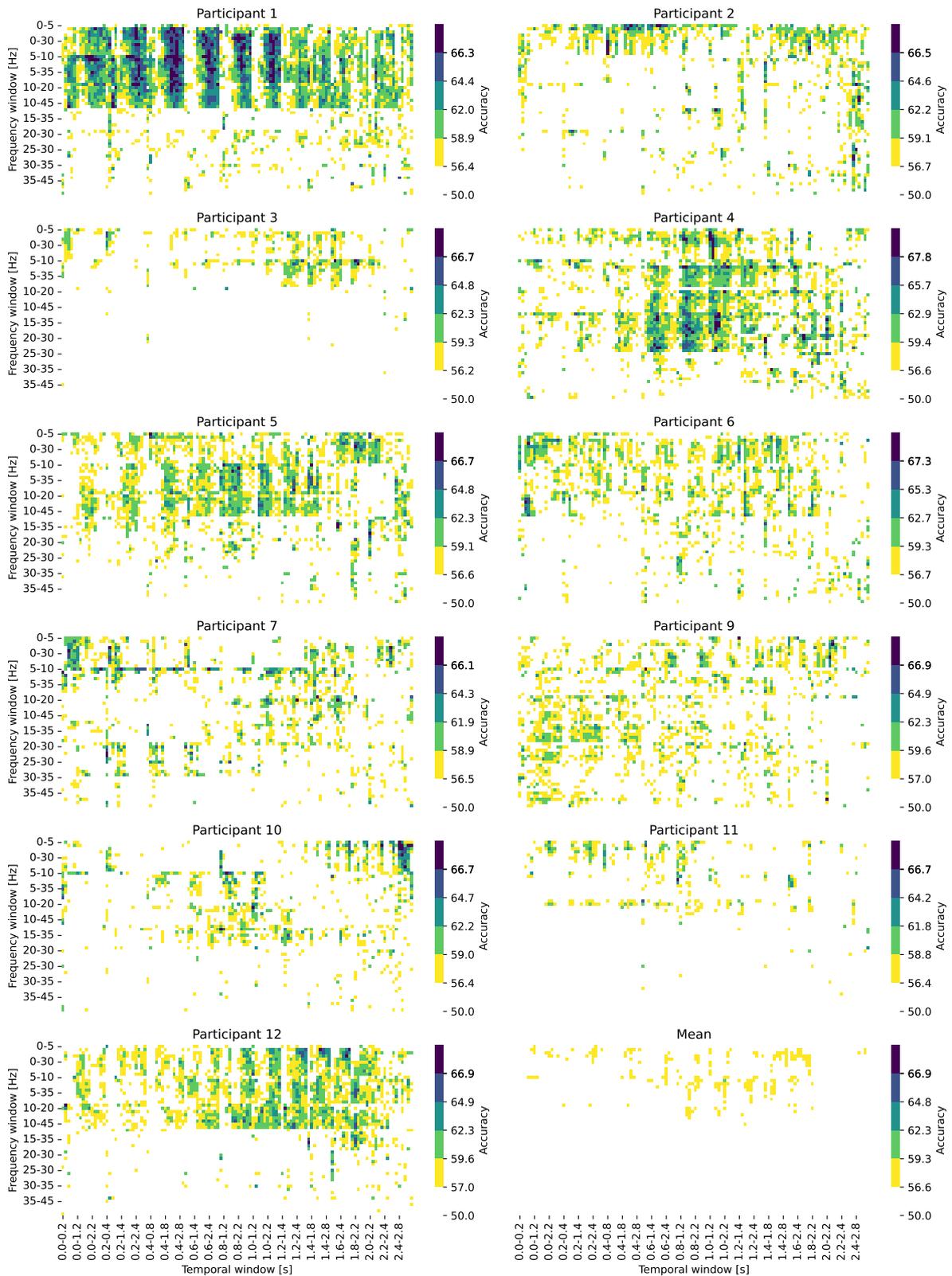


Fig. C.15 Maximal classification accuracies for the tactile imagery task for each time-frequency window from all tested N CSP components with the SVM ($C = 1$), see Section 4.1.4.4. Classification accuracies in colorbars represent significant borderlines for p -values of $p \in \{0.05, 0.01, 0.001, 0.0001, 0.00001\}$. White represents non-significant classification accuracies. Not all labels for X and Y axis are shown for simplicity.

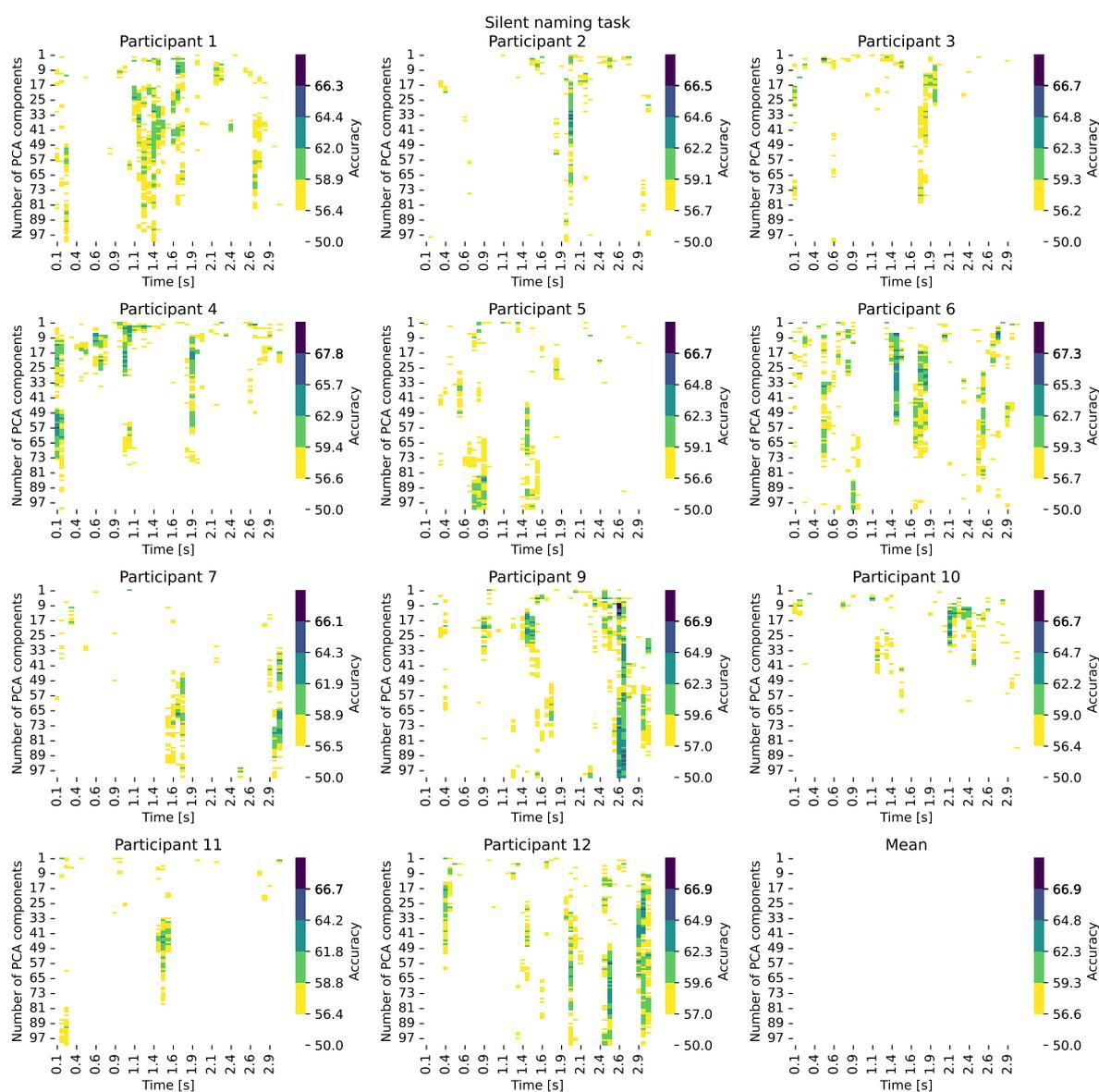


Fig. C.16 Classification accuracies for the silent naming task with the sliding temporal window of 125 ms when the SVM can use information from up to N spatio-spectral PCA components from the joint channel and frequency space, see Section 4.1.5. Classification accuracies in color-bars represent significant borderlines for p -values of $p \in \{0.05, 0.01, 0.001, 0.0001, 0.00001\}$. White represents non-significant classification accuracies.

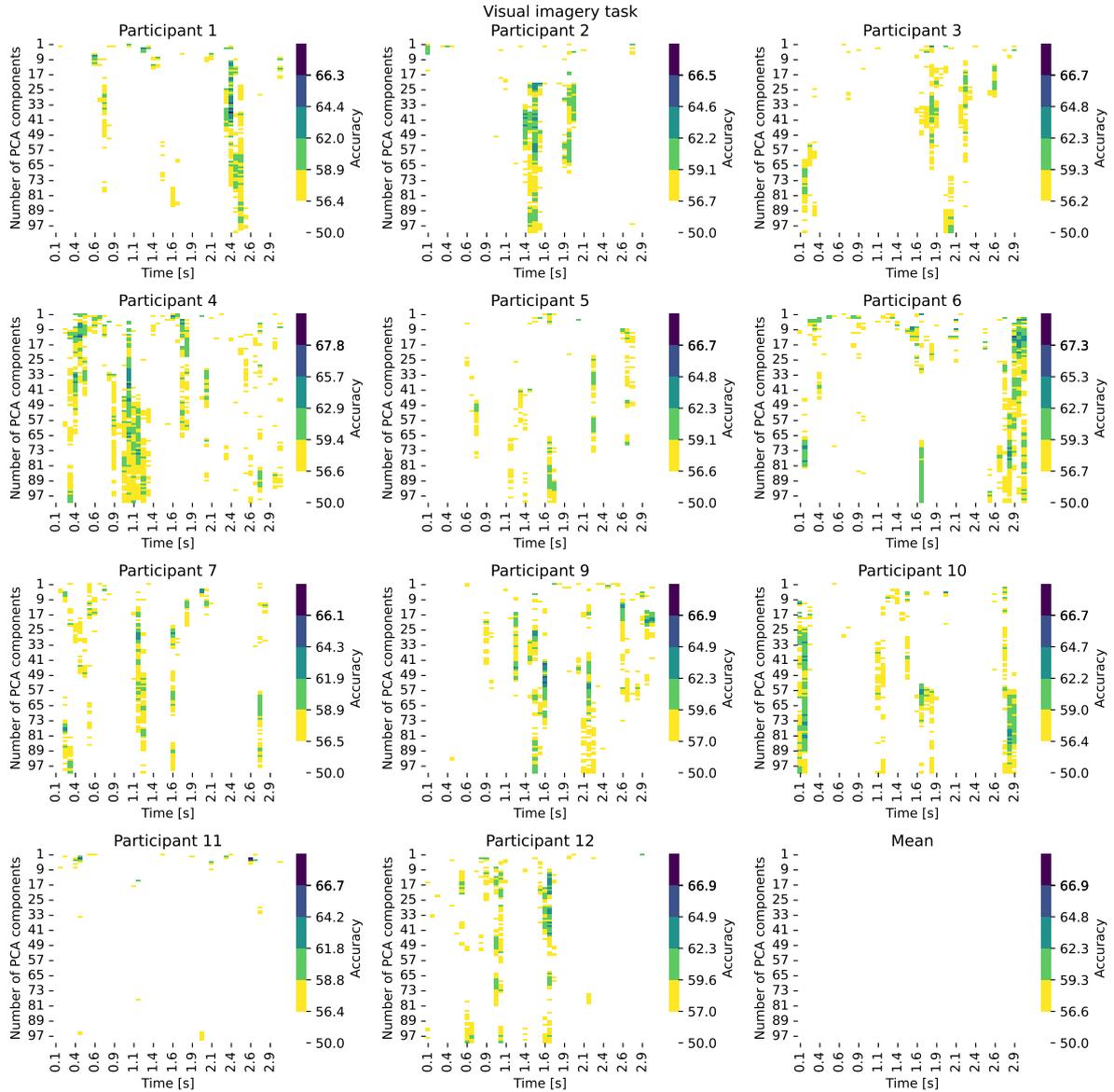


Fig. C.17 Classification accuracies for the visual imagery task with the sliding temporal window of 125 ms when the SVM can use information from up to N spatio-spectral PCA components from the joint channel and frequency space, see Section 4.1.5. Classification accuracies in color-bars represent significant borderlines for p -values of $p \in \{0.05, 0.01, 0.001, 0.0001, 0.00001\}$. White represents non-significant classification accuracies.

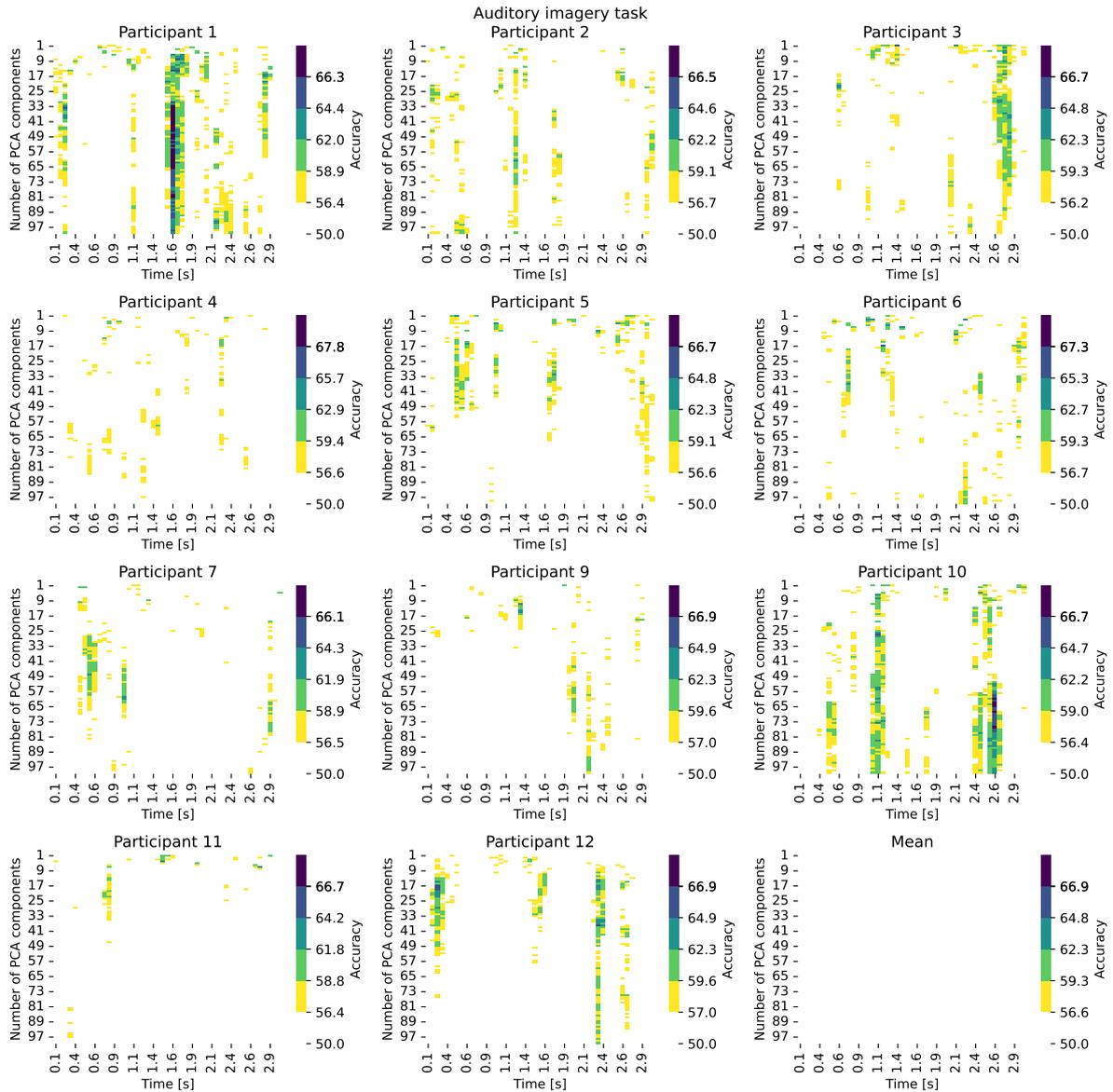


Fig. C.18 Classification accuracies for the auditory imagery task with the sliding temporal window of 125 ms when the SVM can use information from up to N spatio-spectral PCA components from the joint channel and frequency space, see Section 4.1.5. Classification accuracies in colorbars represent significant borderlines for p -values of $p \in \{0.05, 0.01, 0.001, 0.0001, 0.00001\}$. White represents non-significant classification accuracies.

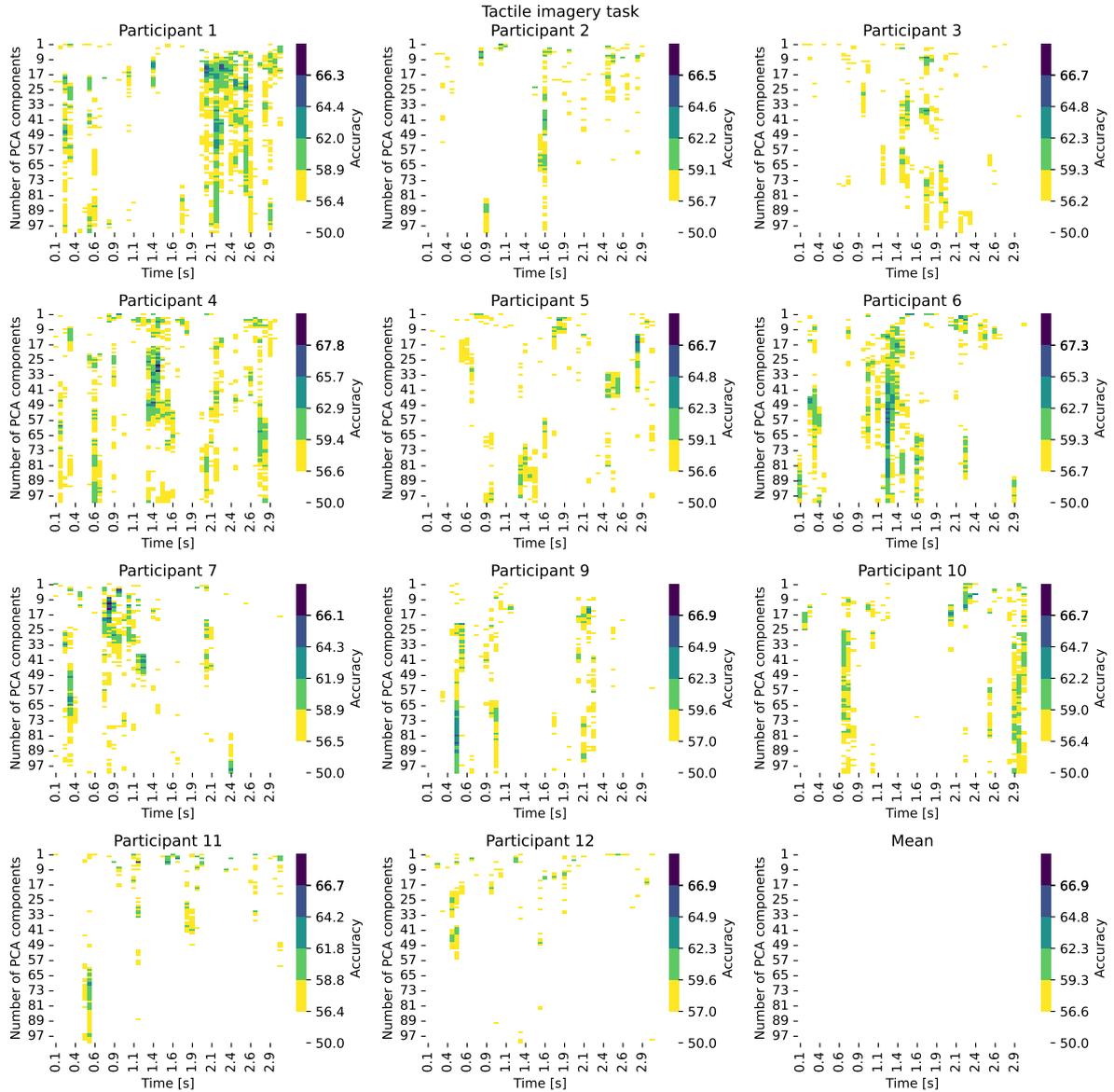


Fig. C.19 Classification accuracies for the tactile imagery task with the sliding temporal window of 125 ms when the SVM can use information from up to N spatio-spectral PCA components from the joint channel and frequency space, see Section 4.1.5. Classification accuracies in color-bars represent significant borderlines for p -values of $p \in \{0.05, 0.01, 0.001, 0.0001, 0.00001\}$. White represents non-significant classification accuracies.

Appendix D

Additional figures for Chapter 5

This appendix complements the results from [Chapter 5](#).

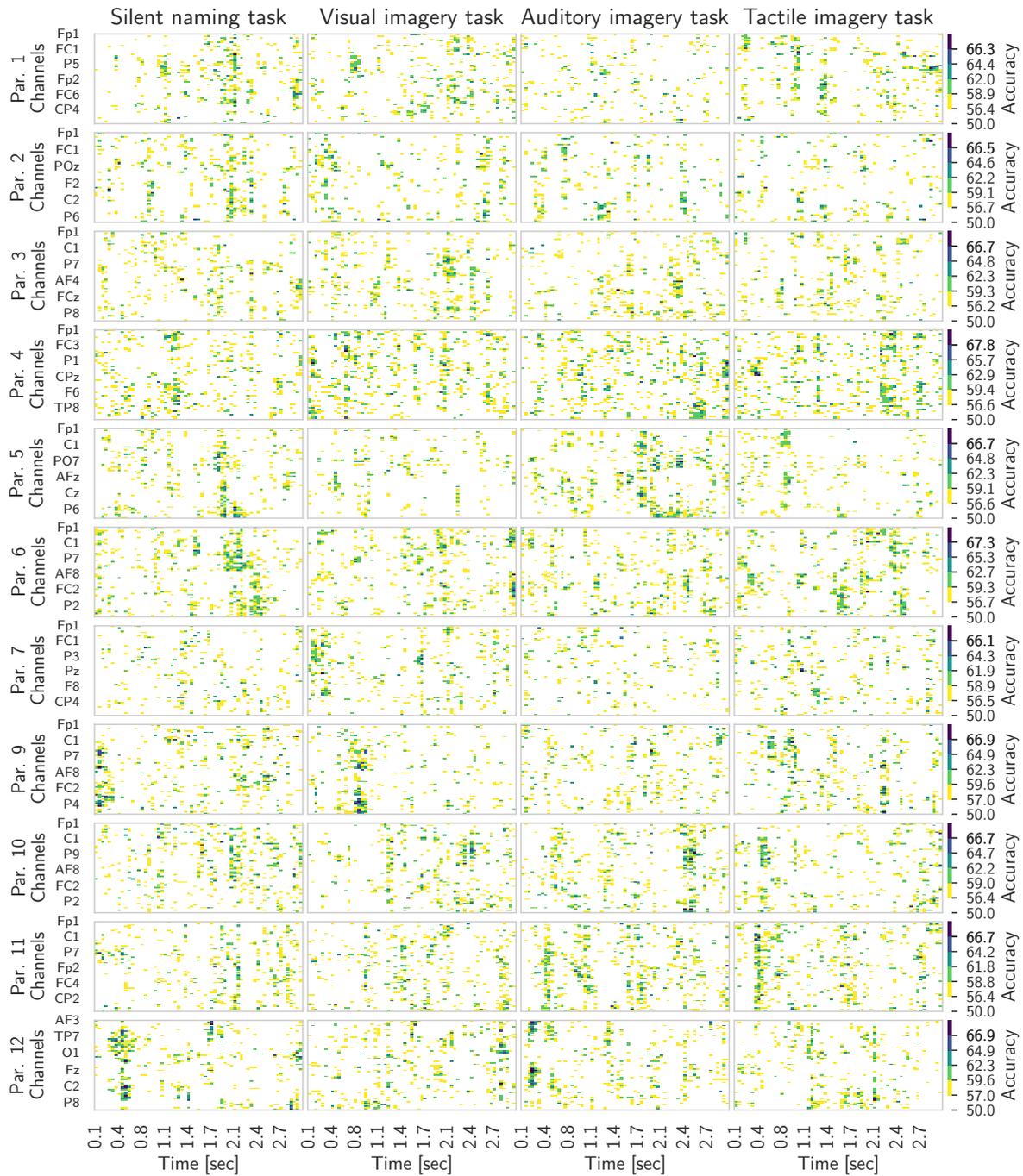


Fig. D.1 Classification accuracies for the 1–30 Hz frequency band for single channel classification over a sliding temporal window (Analysis 1). Rows represent different participants. Columns are different mental tasks. In each plot, the Y axis represents EEG channels but the Y label does not include all channel names to improve readability. Classification accuracies in colorbars represent significant borderlines for p -values of $p \in \{0.05, 0.01, 0.001, 0.0001, 0.00001\}$ for each participant. Note that accuracy borderlines differ between different participants due to the different number of epochs. White represents non-significant classification accuracies.

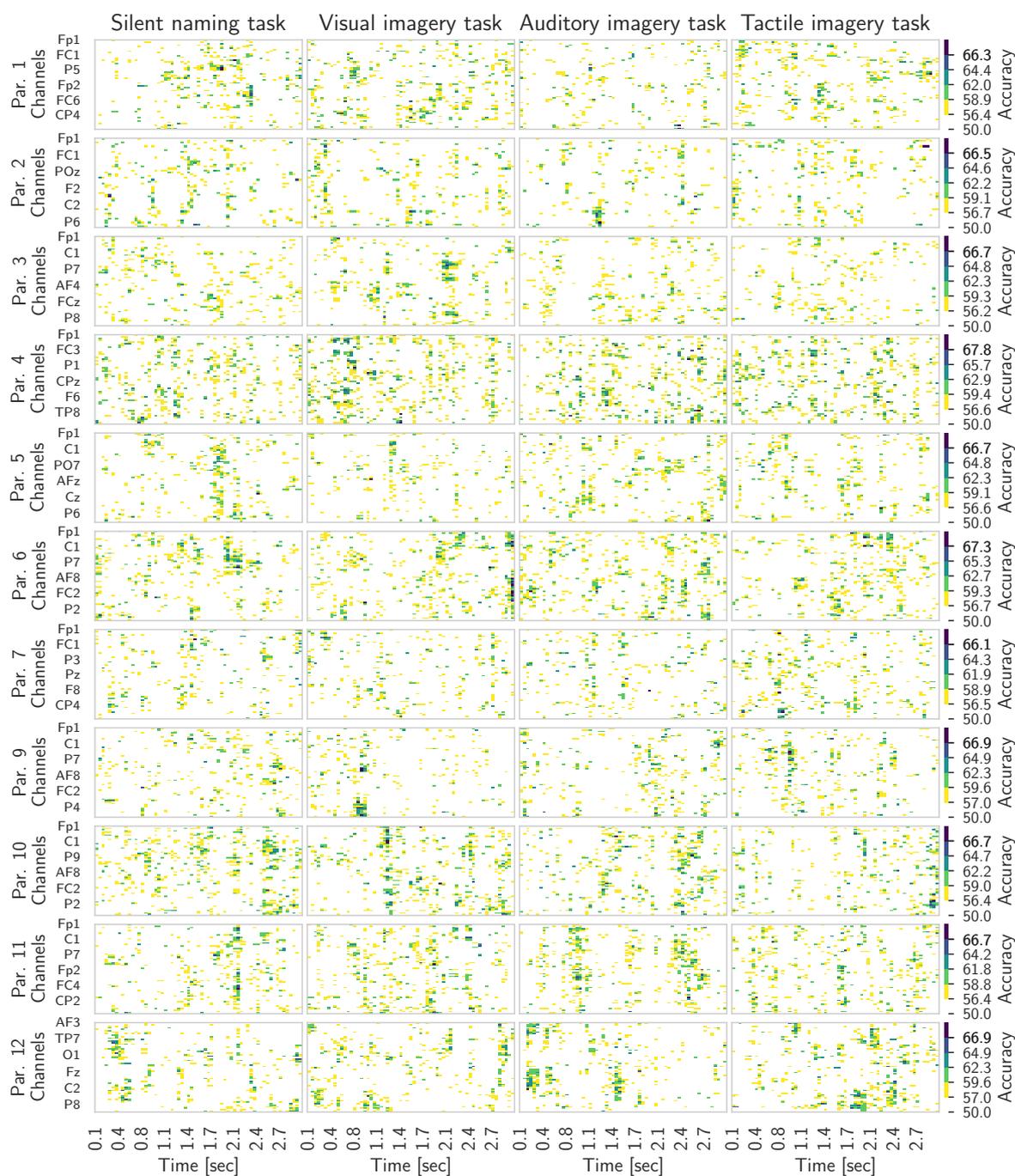


Fig. D.2 Classification accuracies for the 4–30 Hz frequency band for single channel classification over a sliding temporal window (Analysis 1). Rows represent different participants. Columns are different mental tasks. In each plot, the Y axis represents EEG channels but the Y label does not include all channel names to improve readability. Classification accuracies in colorbars represent significant borderlines for p -values of $p \in \{0.05, 0.01, 0.001, 0.0001, 0.00001\}$ for each participant. Note that accuracy borderlines differ between different participants due to the different number of epochs. White represents non-significant classification accuracies.

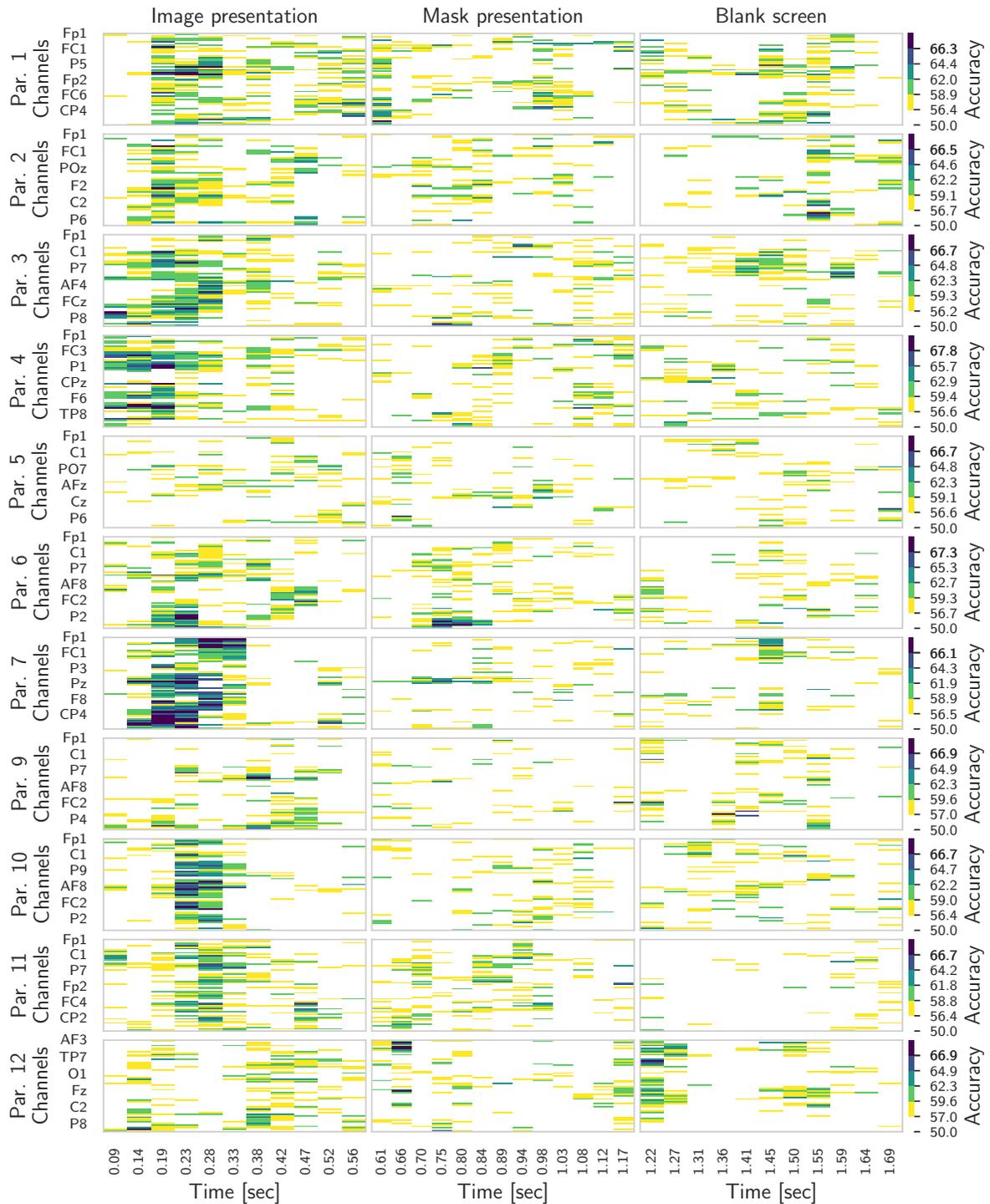


Fig. D.3 Classification accuracies for the 1–30 Hz frequency band for single channel classification over a sliding temporal window (Analysis 1) for the beginning of the concept trials (image presentation, mask, and blank screen). Rows represent different participants. In each plot, the Y axis represents EEG channels but the Y label does not include all channel names to improve readability. Classification accuracies in colorbars represent significant borderlines for p -values of $p \in \{0.05, 0.01, 0.001, 0.0001, 0.00001\}$ for each participant. Note that accuracy borderlines differ between different participants due to the different number of epochs. White represents non-significant classification accuracies.

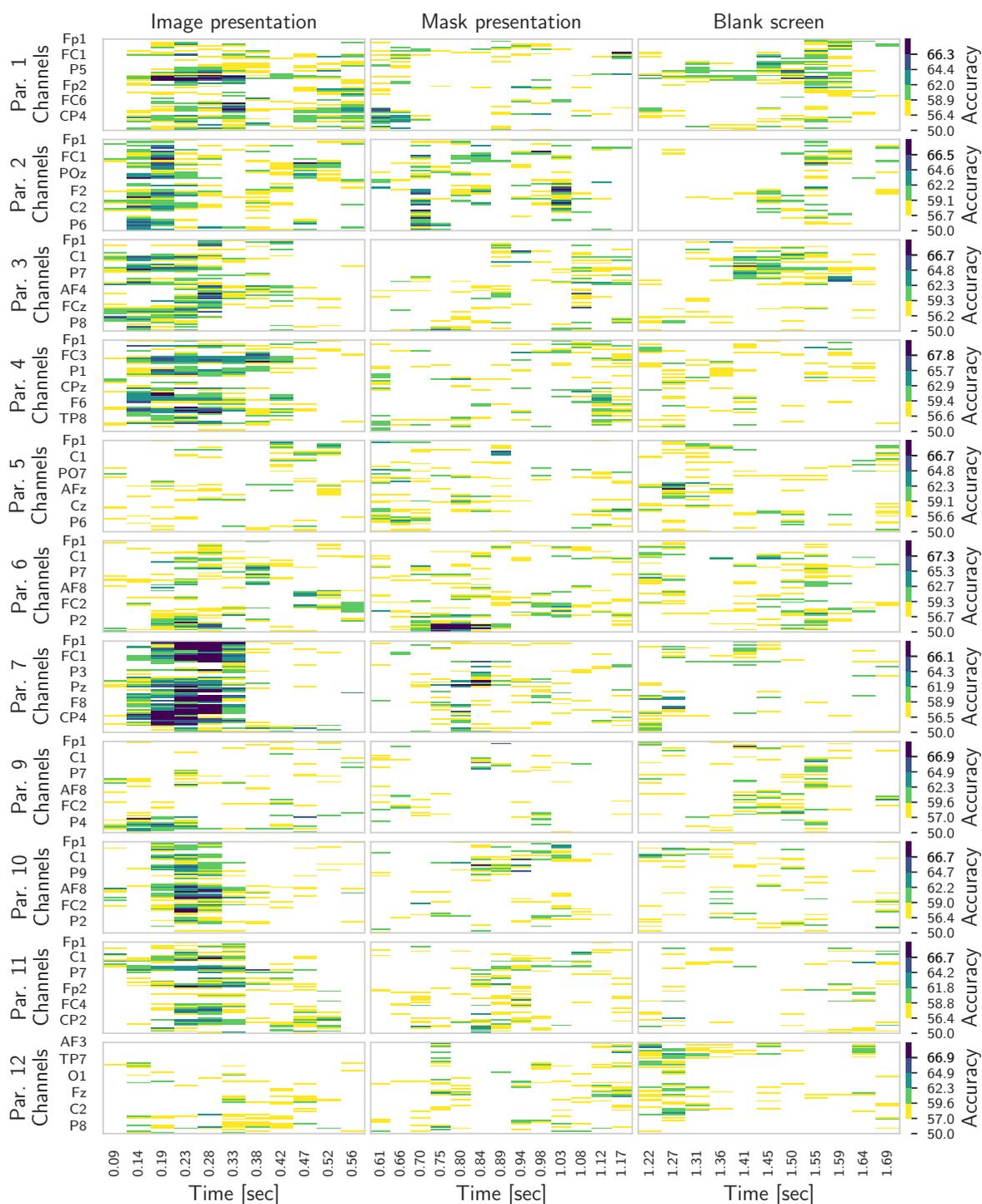


Fig. D.4 Classification accuracies for the 4–30 Hz frequency band for single channel classification over a sliding temporal window (Analysis 1) for the beginning of the concept trials (image presentation, mask, and blank screen). Rows represent different participants. In each plot, the Y axis represents EEG channels but the Y label does not include all channel names to improve readability. Classification accuracies in colorbars represent significant borderlines for p -values of $p \in \{0.05, 0.01, 0.001, 0.0001, 0.00001\}$ for each participant. Note that accuracy borderlines differ between different participants due to the different number of epochs. White represents non-significant classification accuracies.

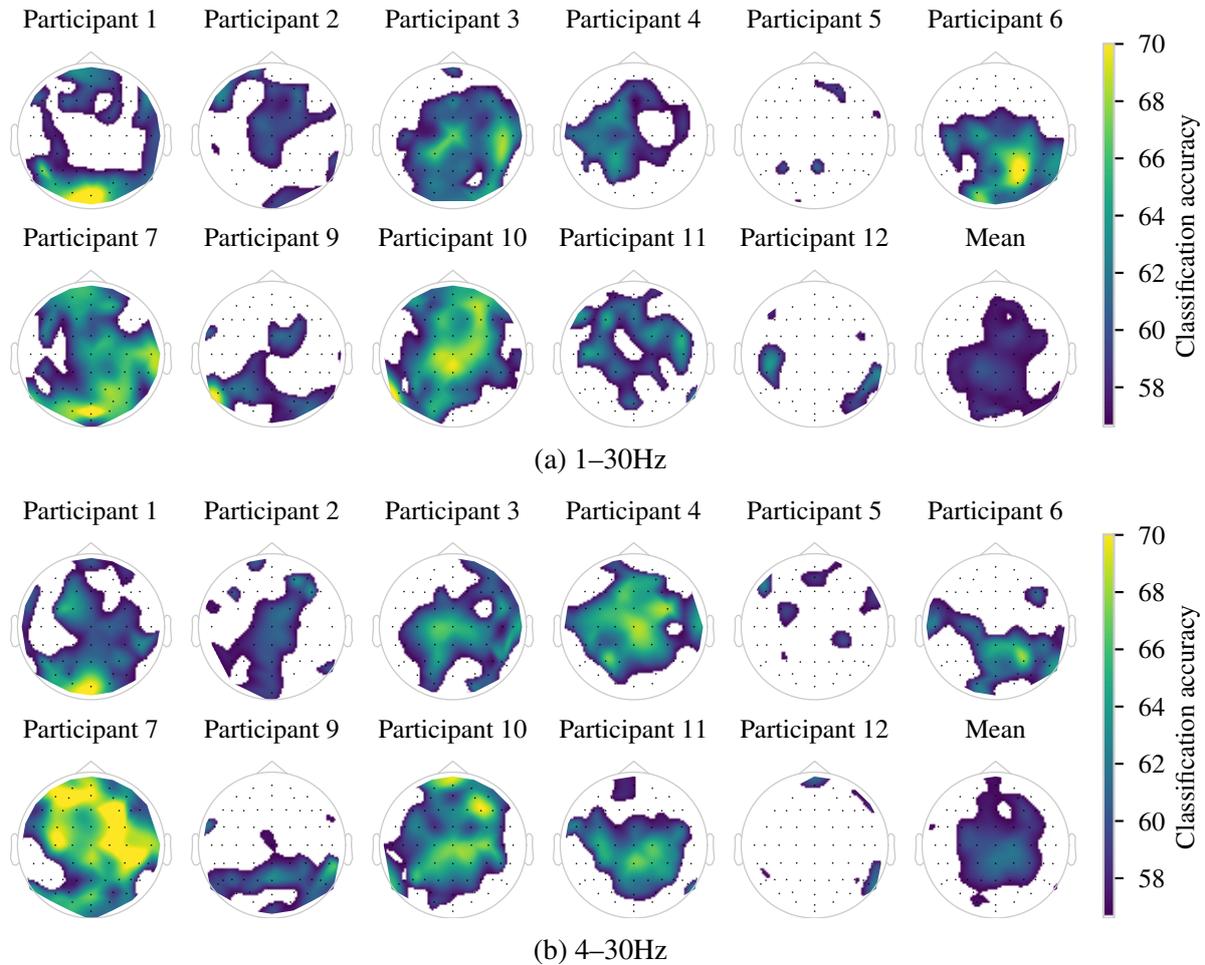


Fig. D.5 Classification accuracies during the image presentation period from each channel in the temporal window of 140.6–234.3 ms (i.e., the sliding temporal window from Analysis 1) after the image onset. This temporal window had the highest classification accuracy when averaged over all channels and all participants in both frequency bands. Scalp maps indicate performance above the significance borderline ($p = 0.05$, 56.6%). White represents non-significant classification accuracies.

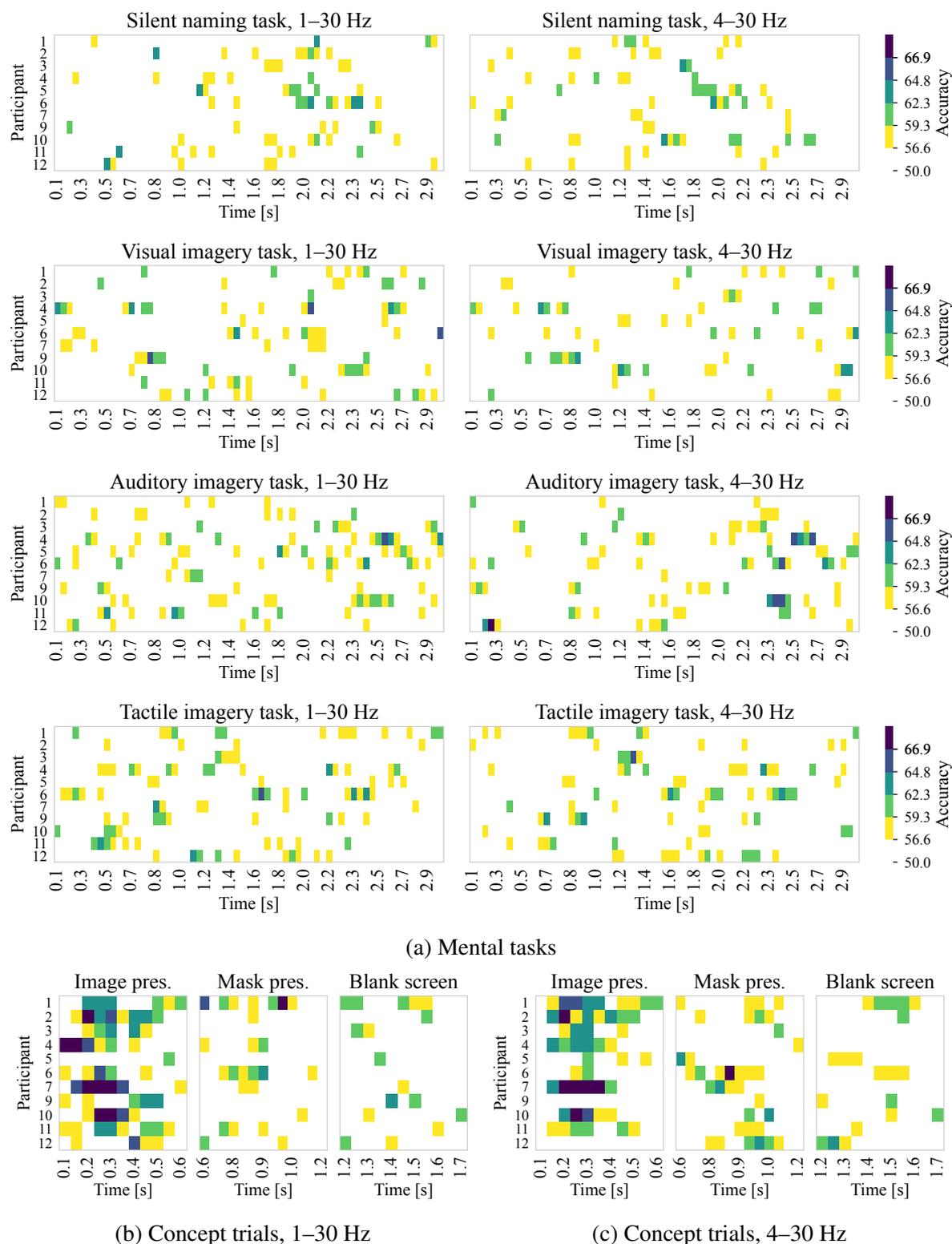


Fig. D.6 Classification accuracies for all channels over a sliding temporal window (Analysis 2) for the mental tasks and for the beginning of the concept trials (image presentation, mask, and blank screen). Classification accuracies in colorbars represent significant borderlines for $p = 0.05$ (56.6%), $p = 0.01$ (59.3%), $p = 0.001$ (62.3%), $p = 0.0001$ (64.8%), and $p = 0.00001$ (66.9%). White represents non-significant classification accuracies.

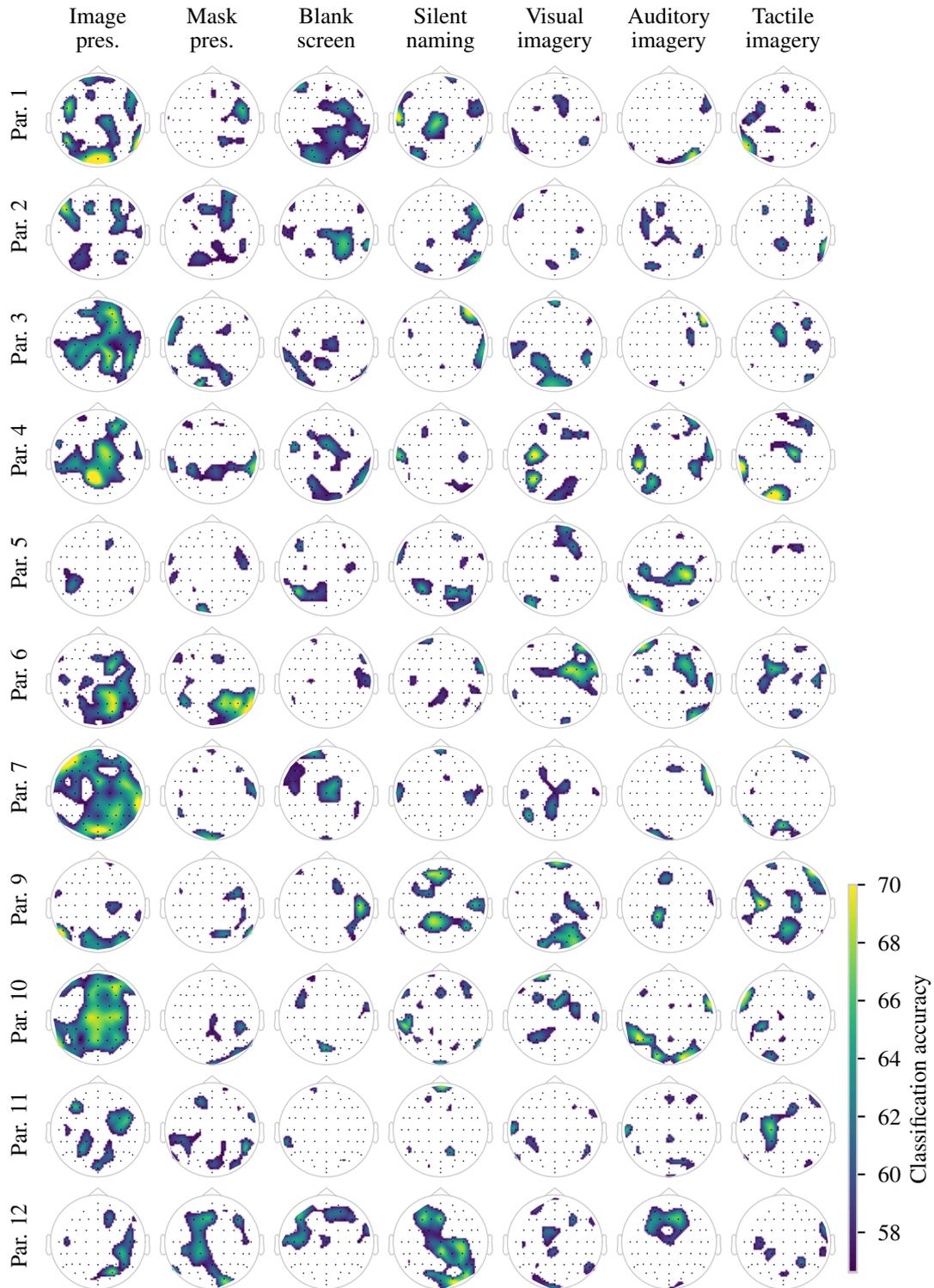


Fig. D.7 Classification accuracies for the 1–30 Hz frequency band using the nested-cross-validation approach to select an optimal temporal window (Analysis 3). Columns represent different periods. Rows correspond to different participants. Scalp maps indicate performance above the significance borderline ($p = 0.05$, 56.6%). White represents non-significant classification accuracies.

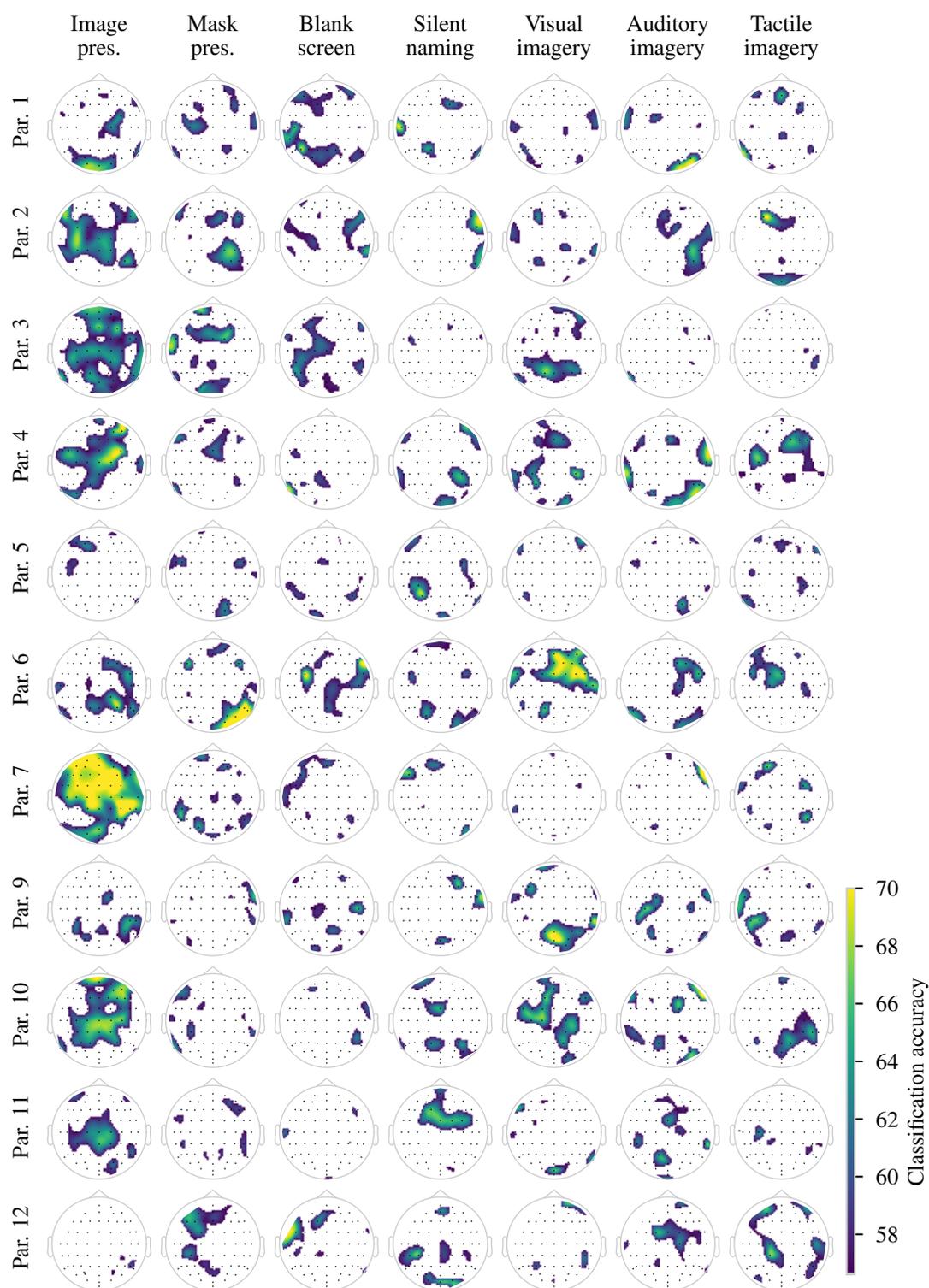


Fig. D.8 Classification accuracies for the 4–30 Hz frequency band using the nested-cross-validation approach to select an optimal temporal window (Analysis 3). Columns represent different periods. Rows correspond to different participants. Scalp maps indicate performance above the significance borderline ($p = 0.05$, 56.6%). White represents non-significant classification accuracies.