



**Politecnico
di Torino**

POLITECNICO DI TORINO

Biomedical Engineering

A.a. 2025/2026

Graduation Session 03 2026

SUBJECT-INDEPENDENT SEMANTIC
DECODING FROM EEG: COMPARING
INTERPRETABLE FEATURE-BASED,
MACHINE-LEARNING, AND
DEEP-LEARNING APPROACHES

Supervisors:

Luca Mesin
Hossein Ahmadi

Candidate:

Vittorio Fonte

Abstract

Semantic decoding aims to identify, directly from brain activity, which concept a person is perceiving or imagining, without any motor or behavioural response. Although electroencephalography (EEG) is the most widely adopted recording modality in brain–computer interfacing, its use for semantic decoding remains an emerging line of research because of its practical advantages: non-invasiveness, low cost, portability, and high temporal resolution. Realising this potential, however, requires overcoming two limitations that affect most decoding work: poor cross-subject generalisation, since within-subject and pooled evaluation reward subject-specific idiosyncrasies, and limited interpretability, since the field is increasingly dominated by opaque deep-learning models. This work addresses these limitations on the multimodal semantic EEG dataset of Wilson et al., applying a strict Leave-One-Subject-Out (LOSO) protocol to all three classification tasks: perception versus imagination, sensory modality, and semantic concept. For the latter two, this is the first evaluation under such a protocol. The central contribution is a novel, interpretable pipeline that couples an anatomical regional decomposition with region-wise Common Spatial Patterns (CSP) and a parameter-free decision rule. It is benchmarked against two comparators: a Riemannian-geometric method mapping band-specific covariance matrices to the tangent space, and a topographic deep-learning approach classifying scalp-voltage map sequences with a convolutional network. Performance is assessed with balanced accuracy and non-parametric statistics. The work establishes subject-independent benchmarks for these tasks and proposes interpretable feature-based decoding as a complementary direction alongside established machine-learning and deep-learning approaches.

Table of Contents

1	Introduction	1
1.1	Motivation	1
1.2	Semantic decoding in the EEG	2
1.3	Research Gap	3
1.4	Original Contribution	5
1.5	Research Questions	6
1.6	Thesis Overview	6
2	Background & State of The Art	8
2.1	Brain Computer Interface	8
2.1.1	Neuro-imaging with Brain Computer Interface	10
2.1.2	Why EEG for semantic decoding?	16
2.2	Electroencephalography Features (EEG)	17
2.2.1	EEG Artifacts	19
2.2.2	Event Related Potentials	21
2.2.3	EEG frequency bands	22
2.3	Spectral and Time–Frequency Analysis	24
2.4	Machine Learning on the EEG	27
2.5	Deep Learning on the EEG	29
3	Materials & Methods	33
3.1	Data Set	33
3.2	Pipeline 1	39
3.3	Pipeline 2	44
3.4	Pipeline 3	48
3.5	Experimentation Setups	51
3.5.1	Tools	51
3.5.2	Preprocessing	52
3.5.3	Evaluation metrics	54
3.5.4	Statistical analysis	55
3.5.5	Experimental Design	56

4	Results	61
4.1	Pipeline 1	62
4.1.1	Imagination VS Perception	62
4.1.2	Sensory Modality	64
4.1.3	Semantic Concept	66
4.2	Pipeline 2	68
4.2.1	Imagination VS Perception	68
4.2.2	Sensory Modality	70
4.2.3	Semantic Concept	72
4.3	Pipeline 3	74
4.3.1	Imagination VS Perception	74
4.3.2	Sensory modality	78
4.3.3	Semantic Concept	81
4.4	Inter- and Intra- Pipeline Comparisons	86
4.4.1	Comparison Across the Three Main Pipelines	86
4.4.2	Pipeline 1: Comparison of the Three Decision Rules	91
4.4.3	Pipeline 2: Comparison of the Two Classifiers	92
5	Discussion	98
5.1	Reuslts discussion	98
5.2	Research Questions	101
5.2.1	RQ1	101
5.2.2	RQ2	102
5.2.3	RQ3	103
6	Conclusion	106
6.1	Summary of Findings and Contributions	106
6.2	Limitation & Future perspective	108
6.3	Concluding remarks	109
A	Appendix	111
A.1	Pipeline 1 materials	111
A.2	Pipeline 2 materials	116
A.3	Pipeline 3 materials	120
B	Appendix	124
B.1	Pipeline 1 Only Energy	124
B.1.1	Imagination VS Perception	124
B.1.2	Sensory Modality	126
B.1.3	Semantic Concept	128
B.2	Pipeline 1 Only Attention	130
B.2.1	Sensory Modality	130

B.2.2	Semantic Concept	132
C	Appendix	134
C.1	Pipeline 2 Linear SVM	134
C.1.1	Imagination VS Perception	134
C.1.2	Sensory Modality	136
C.1.3	Semantic Concept	138
D	Appendix	140
D.1	Pipeline 1 Friedman test results	140
E	Appendix	141
	Bibliography	144

Chapter 1

Introduction

1.1 Motivation

Brain-Computer Interfaces (BCIs) are systems that record neural activity, process it, and translate it into output signals, such as device commands or communication outputs, without relying on the brain's normal neuromuscular output pathways[1]. They are particularly relevant for clinical rehabilitation, as they offer a pathway to restore functions affected by severe neuromuscular disorders. While early BCI paradigms predominantly focused on decoding low-level motor commands, such as intended limb movements, and on restoring motor function in paralysed patients, a growing body of research has begun to extend this framework toward higher-level neural processes[2, 3, 4, 5]. It is precisely within this neurocognitive branch of BCI research that semantic decoding is situated: the challenge of identifying, directly from brain activity, which semantic concept a subject is perceiving or imagining, without requiring any motor action or behavioural response. Achieving robust and generalisable semantic decoding would represent a fundamental step toward BCIs capable of direct conceptual communication, a goal with profound implications for assistive technology, cognitive neuroscience, and our understanding of how the brain encodes meaning. This perspective is also connected to recent work on brain-to-brain communication, where EEG-based representations are treated not only as classification inputs but also as information-bearing signals whose robustness, security, and semantic content become central design requirements[6, 7]. Among the available methods to record neural activity, electroencephalography (EEG) stands as the most widely adopted in BCI research[8], owing to several practical advantages detailed in Section 2.1.2, making it attractive for real-world deployment while simultaneously posing fundamental challenges for semantic decoding, including low signal-to-noise ratio, high temporal non-stationarity, susceptibility to muscular and ocular artifacts, and inherently limited spatial resolution[9].

1.2 Semantic decoding in the EEG

The term “neural decoding” describes the process of predicting information about a stimulus from recordings of brain activity. Semantic decoding applies this specifically to meaning, inferring which semantic concept (e.g., “dog”, “cat”) a person is perceiving or thinking about from their neural signals. The difficulty lies in the fact that a concept is the meaning of a thing independent of how it looks or sounds[5]; it therefore targets the conceptual content of thought rather than its low-level perceptual features. Demonstrating that such conceptual content is accessible from scalp-level EEG recordings, however, first required empirical evidence that electrical brain responses are sensitive to higher-order cognitive processes. This evidence emerged progressively through Event-Related Potential (ERP) research: the P300 showed that scalp potentials are modulated by higher-order cognitive variables such as stimulus uncertainty[10], while the N400 demonstrated specific sensitivity to semantic incongruity[11], establishing that electrophysiological responses recorded at the scalp can track meaning and providing an empirical foundation for subsequent decoding approaches. Building upon this electrophysiological evidence, Suppes et al.[12] demonstrated, in 1997, the first systematic classification of word identity from brain signals. Electrical and magnetic brain activity was recorded from seven subjects across three experimental conditions, namely auditory comprehension, internal (imagined) speech, and normal speech, while subjects processed a closed set of seven target words. Classification was achieved by constructing a per-word prototype for each condition via trial averaging, applying subject-specific optimal predictive filtering in the frequency domain, and subsequently identifying the target word through least-squares matching between each filtered test sample and the corresponding prototype set in the time domain. The work of Suppes et al. cannot yet be characterised as a semantic classifier, that is, a system capable of recognising a word on the basis of its meaning. As the authors themselves state, semantic aspects were not yet analysed; the approach instead operates as a waveform-matching procedure that exploits discriminative temporal and spectral features distinguishing the electrophysiological signatures of individual words from one another[12, 13]. The transition to properly semantic decoding, meaning the classification of a stimulus on the basis of its meaning rather than its raw waveform, emerged with Simanova et al.[14] and Murphy et al.[15], who showed that multivariate pattern analysis (MVPA) of single-trial EEG could recover broad semantic categories (such as animals versus tools) well above chance. Both studies also provided evidence that the decoded patterns reflected conceptual rather than perceptual content: Simanova et al. decoded category from pictures as well as from spoken and written words, while Murphy et al. exploited single-trial predictions to systematically rule out a range of low-level visual, lexical, and task-related confounds. The growing evidence that EEG signals encode conceptual rather than

merely perceptual content prompted a broader shift toward semantic decoding as a research focus in its own right; the field progressively moved toward large-scale open datasets, most notably THINGS-EEG[16] and THINGS-EEG2[17], which provided concept-level neural responses to thousands of visual object stimuli, yet these resources remained confined to visual perception paradigms. To address the lack of open datasets capturing semantic representation across multiple sensory modalities and across both perception and imagination tasks, Wilson et al.[18] introduced the dataset adopted in the present work.

1.3 The Research Gap in Semantic EEG Decoding

Despite significant advances in multivariate pattern analysis and deep learning for neural decoding, a critical bottleneck persists across EEG-based decoding, and semantic decoding is no exception: the absence of robust cross-subject generalisation. Inherent inter-subject variability in brain anatomy, cortical folding, and individual neural response patterns introduces a severe distribution shift between training and unseen subjects; the predominant within-subject evaluation paradigm renders existing decoding frameworks inherently subject-specific, thereby limiting their scalability and practical applicability in real-world deployment[19, 20, 21]. This limitation is further compounded by the heterogeneous and frequently optimistic evaluation protocols adopted in the literature, a shortcoming to which the dataset introduced by Wilson et al.[18] is not immune. Comprising only 12 subjects, this dataset constrains both the statistical power of cross-subject evaluation and the volume of available training data, thereby incentivising the adoption of data-maximising protocols. Across the classification studies that employ it, three evaluation regimes predominate. In **within-subject k -fold** validation[18, 22, 23, 24], the trials of each individual are partitioned into k equally sized folds and the procedure is repeated until each fold has served once as the test set, with training and evaluation performed entirely within that subject’s data. The most common choice is $k = 5$, yielding an 80/20 training test ratio[18, 23, 24], although other values of k have also been explored[22]. Wilson et al.[18] further repeat the five-fold procedure over fifty independent random shuffles to stabilise the accuracy estimate. Although this protocol permits assessment of decodability within a given participant, it is uninformative with respect to the transferability of the resulting decoder to previously unseen subjects: the features it learns may reflect subject-specific signal idiosyncrasies rather than stimulus-driven representational structure, and thus carry no guarantee of generalising beyond the individual on whom the model was trained. In **pooled cross-subject validation**[25, 26, 27], the trials of all subjects are aggregated prior to partitioning, with no separation enforced at the

subject level. This regime encompasses three implementation variants. The first is a single hold-out split[26], in which each subject’s trials are divided into fixed training and test portions before pooling, such that the final training and test sets each contain samples from all subjects in a fixed proportion. The second is a pure k -fold scheme[25], in which the aggregated data are partitioned into k equally sized folds and the procedure is repeated until each fold has served once as the test set, ensuring that every trial participates in evaluation. The third is a hybrid hold-out and k -fold scheme[27], in which a fixed portion of the pooled data is reserved as a permanent test set, while k -fold cross-validation is applied exclusively to the remaining data to rotate the validation set across training iterations; in this variant, the test set never changes across folds. Regardless of the specific implementation, trials from the same subject are distributed across both training and test partitions; the model is therefore exposed to subject-specific signal idiosyncrasies during both training and evaluation, inadvertently rewarding subject recognition rather than semantic decoding and producing accuracy estimates that do not reflect genuine generalisation. In **Leave-One-Subject-Out (LOSO)** validation[28, 29], by contrast, all subjects but one contribute to training, with the held-out individual serving as the sole test case; the procedure is repeated until each subject has occupied the test role exactly once. By never exposing the model to data from the test individual during training, this scheme eliminates the within-subject sources of leakage and provides a direct estimate of cross-subject generalisation. Yet the reliability of this estimate depends on the number of subjects available: with small cohorts, the variance across held-out folds can be substantial, and any demographic or physiological biases present in the subject pool are inherited by the evaluation. Nevertheless, the present work deliberately adopts LOSO as its evaluation protocol. Despite the severe statistical penalty imposed by the small sample size of 12 subjects[18], this choice is preferred precisely because it tests whether the learned decision relies on invariant semantic structure rather than on subject-specific signal characteristics or anatomical idiosyncrasies. The existing subject-independent work on this dataset [28, 29] together with recent task-independent semantic feature learning approaches[30], has mainly focused on the perception–imagination distinction and on representation-level structure. The discrimination among sensory modalities (auditory, orthographic, pictorial) and the classification of semantic concepts (flower, guitar, penguin) remain unexamined under this regime. This gap is addressed in the present study. A further gap concerns the interpretability of the analytical frameworks employed in semantic EEG decoding. The recent literature on this task is increasingly dominated by deep learning architectures, notably Convolutional Neural Networks and Transformer-based models, which optimise classification performance through end-to-end learned feature representations[31, 32]. Despite competitive empirical results, these architectures are inherently opaque: as Sturm et al. [33] observe, deep neural networks are “notorious black boxes” whose

internal representations resist direct neuroscientific interpretation without post-hoc analysis tools that are themselves approximate and architecture dependent. Systematic reviews of deep learning for EEG classification confirm that achieving interpretability in these models is structurally difficult and often entails a trade-off with predictive performance[31, 34]. This opacity is particularly problematic when the scientific goal extends beyond decoding accuracy, as is explicitly the case for the Wilson et al.[18] dataset, whose stated objectives encompass not only classification performance but an understanding of the neural mechanisms underlying perception, imagination, and the processing of semantic content across sensory modalities. Handcrafted, spatially interpretable approaches that could provide direct insight into these mechanisms remain largely underexplored in this context. The present work therefore proposes a framework grounded in regional signal decomposition and Common Spatial Pattern (CSP) filtering[35, 36], which yields features with direct anatomical and physiological interpretability. This approach deliberately prioritises transparency in the decoding process over the maximisation of absolute classification performance, with the aim of providing a principled, interpretable basis for analysing semantic brain representations, one that is complementary to, rather than in competition with, deep learning approaches.

1.4 Original Contribution

This thesis proposes an initial investigation into three classification problems on the multimodal semantic EEG dataset of Wilson et al.[18]: the discrimination of perception from imagination, the identification of the sensory modality through which a stimulus was presented, and the recognition of the semantic concept. To this end, one novel handcrafted pipeline is proposed and evaluated alongside two comparative frameworks, all assessed under a strict subject-independent regime. The first is a novel handcrafted pipeline that attempts to address the interpretability gap identified above through anatomically motivated regional brain decomposition and Common Spatial Pattern (CSP) filtering. The second is a Riemannian-geometric machine learning approach, an established framework in EEG signal processing[9], in which single-trial covariance matrices are mapped to a tangent space and subsequently fed into conventional machine learning classifiers. This comparative design follows the broader observation that EEG decoding performance depends strongly on the chosen representational form, as shown in recent work comparing handcrafted, spectral, connectivity, entropy-based, and deep representations across BCI paradigms[37]. The third is a deep learning approach based on topographic decoding[28], in which EEG epochs are converted into sequences of scalp-voltage maps and submitted to a Convolutional Neural Network (CNN). The novel handcrafted pipeline is described in detail in what follows. The choice of this approach

is motivated by the known spatial organisation of semantic representations in the brain. Semantic representations in the human brain are distributed across temporal, parietal, and prefrontal cortical regions[38], and this distributed spatial organisation suggests that decomposition methods sensitive to the regional structure of the EEG signal may offer a relevant analytical lens for semantic decoding. The Common Spatial Patterns algorithm, which extracts spatial filters that maximise the variance contrast between conditions, represents a natural candidate. Rather than adopting the conventional CSP pipeline, in which component log-variances are submitted to a supervised classifier, the proposed framework derives two interpretable scalar measures directly from the CSP decomposition: the variance of the class-specific CSP component, reflecting the signal energy along the most discriminant spatial direction, and a decorrelation measure, serving as a per-trial indicator of the structural reliability of the decomposition within each region. Classification is performed through a parameter-free decision rule.

1.5 Research Questions

In response to the Research Gap, this work evaluates three decoding pipelines on the three classification tasks described, namely perception versus imagination, sensory modality discrimination, and semantic concept recognition, under a strict Leave-One-Subject-Out (LOSO) protocol. The overarching aim is twofold: to establish a set of benchmark results for these tasks under a subject-independent evaluation regime, and to answer the following three research questions.

- RQ1. Does the variance of the class-specific CSP component constitute a discriminative feature for subject-independent semantic EEG decoding, and does the proposed decorrelation measure carry additional discriminative value beyond uniform regional energy aggregation?
- RQ2. Can Riemannian-geometric covariance features, an established representation in EEG signal processing, sustain classification performance across the three semantic decoding tasks under a strict LOSO regime?
- RQ3. Given that topographic decoding has already demonstrated strong results on this dataset for the pictorial perception–imagination distinction [28], does this approach generalise effectively to the remaining tasks, namely sensory modality discrimination and semantic concept recognition?

1.6 Thesis Overview

The remainder of this thesis is organised as follows.

- Chapter 2 reviews the background and state of the art necessary to contextualise the present work. It introduces Brain-Computer Interface systems and the neuroimaging modalities available, motivates the choice of EEG for semantic decoding, and surveys the relevant EEG features, including Event-Related Potentials and frequency-band representations, before reviewing established machine learning and deep learning approaches to EEG classification.
- Chapter 3 describes the materials and methods employed. It presents the dataset and preprocessing pipeline, defines the three classification tasks and the experimental design adopted, and details the three decoding pipelines evaluated: the novel CSP-based handcrafted approach (Pipeline 1), the Riemannian-geometric framework (Pipeline 2), and the topographic map CNN approach (Pipeline 3).
- Chapter 4 reports the classification results obtained by each pipeline on the three tasks, namely perception versus imagination, sensory modality discrimination, and semantic concept recognition, under the Leave-One-Subject-Out evaluation protocol. For each task it compares performance within intra- and inter- pipelines.
- Chapter 5 discusses the results obtained by the three pipelines on each of the three tasks, namely perception versus imagination, sensory modality discrimination, and semantic concept recognition, including the intra- and inter-pipeline comparisons. It then addresses the three research questions in turn, discussing each findings and limitations.
- Chapter 6 concludes the thesis. It summarises the main contributions and the answers to the research questions, and outlines the limitations of the present work together with directions for future research.

Chapter 2

Background & State of The Art

2.1 Brain Computer Interface

The brain is the organ responsible for receiving sensory information, processing it, and generating appropriate responses in the human body. The principal functional unit of the Brain is the neuron, a cell composed of a cell body (soma) and branching extensions called dendrites and axons (Figure 2.1). Neurons communicate through electrochemical signals: at the synapse, the junction between the axon terminal of the transmitting neuron and the dendrite of the receiving neuron, communication is chemical, mediated by neurotransmitters, while signal propagation along the axon is electrical, in the form of action potentials. The neural signals flowing through the input–processing–output pathway can, thanks to devices called Brain-Computer Interfaces, be recorded, decoded, and used to directly drive external devices, bypassing the natural neuromuscular output. Several methods exist to record brain activity; the most common approaches measure the electrical and magnetic activity generated by neuronal populations. Other techniques instead capture metabolic changes associated with neural activation, such as increases in oxygen consumption or cerebral blood flow, the haemodynamic response exploited by methods such as fMRI and fNIRS.

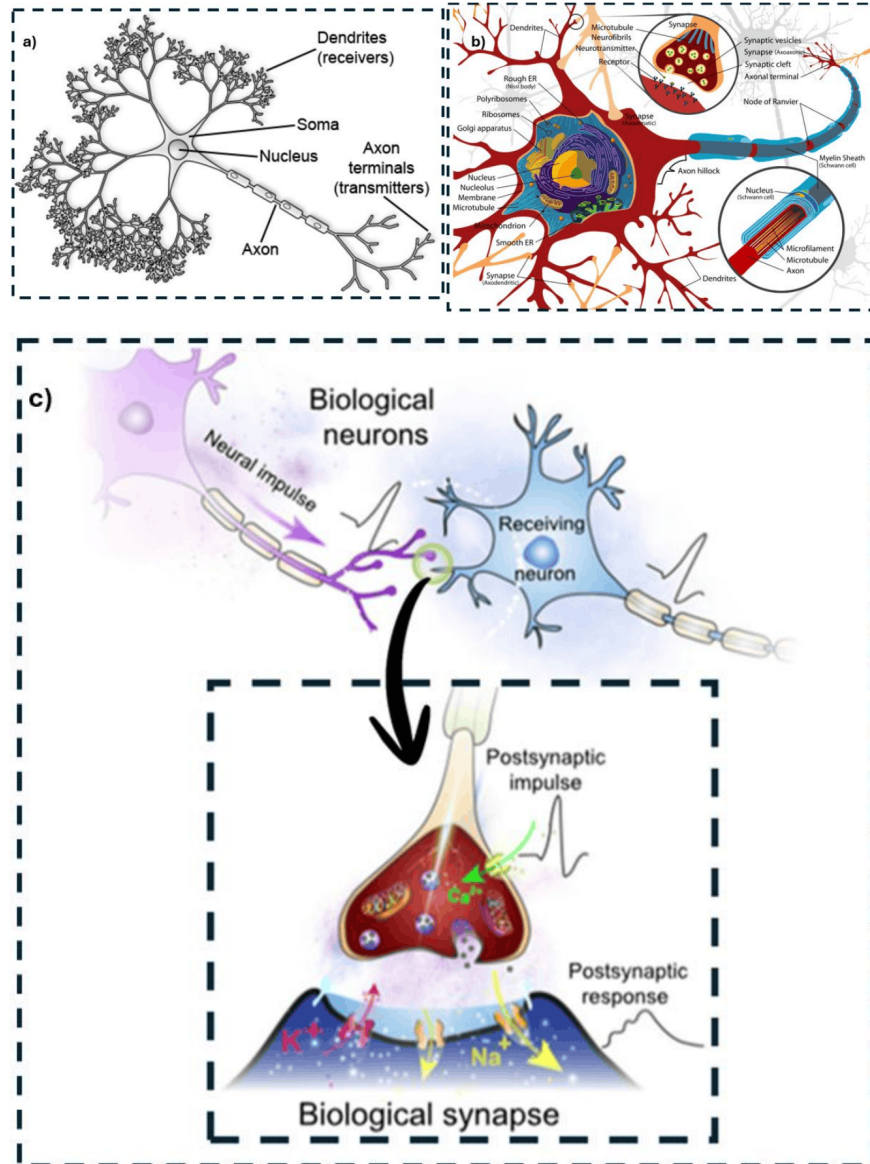


Figure 2.1: Structural and functional organization of the biological neuron. (a) Schematic representation of a neuron. (Source: N. Rougier, Wikimedia Commons). (b) Detailed anatomical view of the neuron, highlighting subcellular structures. (Source: LadyofHats, Wikimedia Commons). (c) Illustration of synaptic transmission between two biological neurons: a neural impulse travels along the axon toward the synapse, where neurotransmitters are released into the synaptic cleft, triggering ion fluxes that generate a postsynaptic impulse and response in the receiving neuron. Adapted from [39]

2.1.1 Neuro-imaging with Brain Computer Interface

Brain-Computer Interfaces are systems that enable direct communication between the brain and external devices, bypassing conventional neuromuscular output channels entirely[1]. This fundamental property makes BCIs particularly valuable for people affected by severe motor impairments, such as amyotrophic lateral sclerosis (ALS), locked-in syndrome, and spinal cord injury, conditions in which the natural pathways for communication and motor control are partially or completely compromised. EEG-based BCI and EEG pattern-recognition methods have also been explored in broader clinical contexts, including consciousness evaluation and coma-outcome prediction, where feature-engineering and machine-learning pipelines are used to extract clinically informative structure from neural signals[40, 41]. The first system reminiscent of a Brain-Computer Interface can be traced back to the performance of the American composer Alvin Lucier, held at the Rose Art Museum, Brandeis University, on May 5th, 1965. In his piece *Music for Solo Performer*, EEG electrodes were placed on Lucier's own scalp, allowing his alpha wave activity (Section 2.2.3) to directly drive percussive instruments distributed around the performance space[42]. Over a decade later, the term "Brain-Computer Interface" was formally introduced by Jacques Vidal in his landmark paper "Toward Direct Brain-Computer Communication"[43]. In a subsequent study (1977), Vidal provided the first practical demonstration of a BCI system, using non-invasive Visual Evoked Potentials (VEP) to control the movement of a cursor-like object through a graphical maze displayed on a computer screen[44]. Since Vidal's pioneering work, the field of BCI has expanded considerably, extending well beyond clinical rehabilitation to broader research domains. BCI technology is now a rapidly evolving field with the potential to influence research, clinical practice, and cognitive neuroscience alike[45, 46]. In recent years, a growing body of research has begun investigating higher-level neurocognitive processes through BCI frameworks, moving beyond simple motor commands to the decoding of perception, attention, and conceptual thought[5]. It is precisely within this neurocognitive branch of BCI research that semantic decoding is situated. BCI systems can be broadly categorised into two main groups according to their signal acquisition approach:

- Non invasive systems, which record brain activity from outside the skull, without any surgical procedure. The most common are:
 - functional Magnetic Resonance (fMRI)
 - functional Near Infrared Spectroscopy (fNIRS)
 - Magnetoencephalography (MEG)
 - Electroencephalography (EEG)

- Invasive systems, which require surgical implantation of electrodes directly on or within the cortex. The most common are:
 - Electrocorticography (ECoG)
 - Micro-electrode

fMRI. Magnetic Resonance Imaging (MRI) is a non-invasive medical imaging technique based on the physical principle of Nuclear Magnetic Resonance[47, 48]. When biological tissue is placed within a strong external magnetic field, the hydrogen protons present in tissues, primarily within water molecules, are selectively excited and emit a detectable radio-frequency signal, which is spatially encoded and reconstructed into detailed anatomical images[49]. Conventional MRI, however, captures only static anatomical structure and cannot resolve dynamic changes in neural activity. The oxygenated and deoxygenated forms of haemoglobin exhibit different magnetic susceptibility properties, and this difference can be exploited to extend MRI beyond static imaging[50]. When neurons become active, their elevated metabolic demand triggers a local increase in cerebral blood flow, delivering oxygenated blood in excess of immediate consumption. This haemodynamic response locally increases the relative concentration of oxygenated haemoglobin, whose distinct magnetic properties produce a detectable change in the MRI signal, through a contrast mechanism termed Blood-Oxygen-Level-Dependent (BOLD)[50]. Functional MRI (fMRI) exploits this contrast mechanism to map brain activity non-invasively, and has since become one of the most widely adopted tools in cognitive neuroscience. Figure 2.2 illustrates a representative fMRI-based BCI paradigm, highlighting the high spatial resolution of the BOLD activation maps. Despite its broad adoption, fMRI presents some fundamental limitations. Most critically, the temporal resolution of fMRI is fundamentally constrained by the slowness of the haemodynamic response, which peaks several seconds after neuronal firing, making conventional fMRI inadequate to capture the millisecond-scale dynamics of cognitive processing[51, 52]. A further limitation concerns its portability: fMRI is inherently confined to fixed laboratory infrastructures, as the requirement for a strong static magnetic field and the considerable size and cost of the equipment currently represent a significant barrier to its deployment in portable or naturalistic settings[53].

fNIRS. In 1977 it was observed that near-infrared light, in the wavelength range of approximately 650–950 nm, is capable of penetrating through bone and soft tissue[55]. As a consequence the concept of the optical window emerged from this observation: a spectral region in which biological tissue exhibits low optical absorption, enabling light to penetrate up to several centimetres of depth. Thanks to this observation, functional Near-Infrared Spectroscopy (fNIRS) was developed

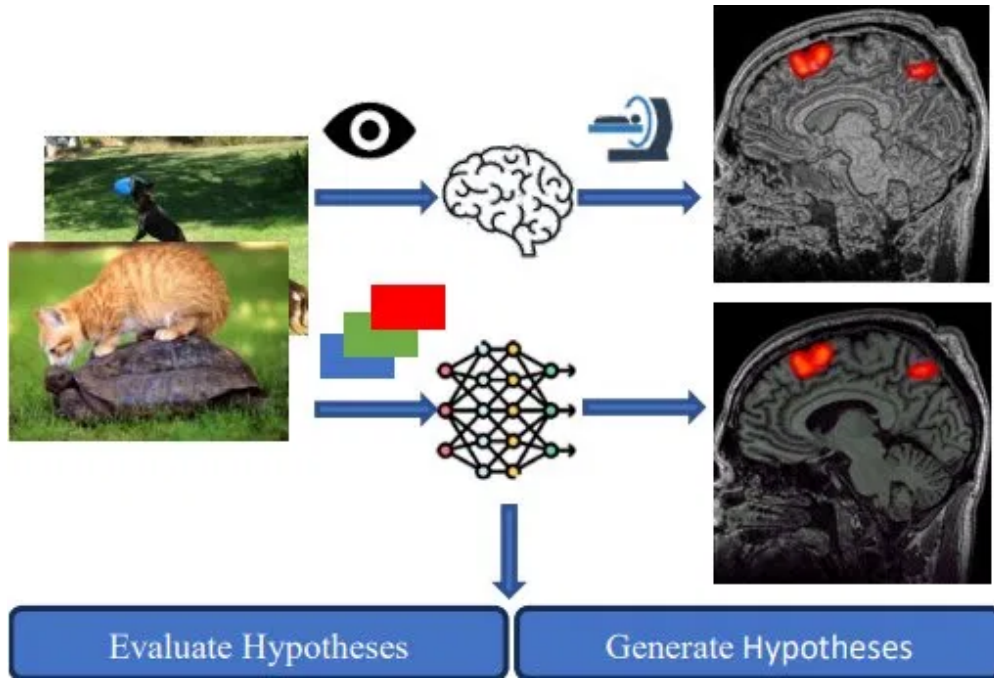


Figure 2.2: Example of an fMRI-based BCI paradigm for visual cortex encoding. A transformer-based neural network receives natural images as input and predicts the corresponding BOLD activation maps in the visual cortex, paralleling the human brain’s response to the same stimuli. Adapted from [54]

as a non-invasive neuroimaging technique. In 1993, four independent research groups simultaneously demonstrated for the first time that fNIRS could detect haemodynamic changes associated with cortical activation in humans[56]. fNIRS measures the haemodynamic response associated with neural activity by detecting changes in the concentration of oxyhaemoglobin (HbO) and deoxyhaemoglobin (HbR) in the cerebral cortex. This is made possible by the distinct absorption spectra of the two forms of haemoglobin in the near-infrared range: by shining NIR light through the scalp and measuring the returning signal, concentration changes can be quantified via the modified Beer-Lambert Law[57]. Figure 2.3 shows an example of the fNIRS used to understand the brain activity. Despite its advantages in terms of portability, tolerance to motion artifacts, applicability in naturalistic settings, fNIRS presents two fundamental limitations. First, it is limited to cortical surface measurements due to restricted depth penetration, being confined to superficial cortical regions, and has lower spatial resolution compared to fMRI. Second, like fMRI, its temporal resolution is constrained by the slowness of the haemodynamic response, having the problem to caught the millisecond-scale dynamics of neural activity[58].

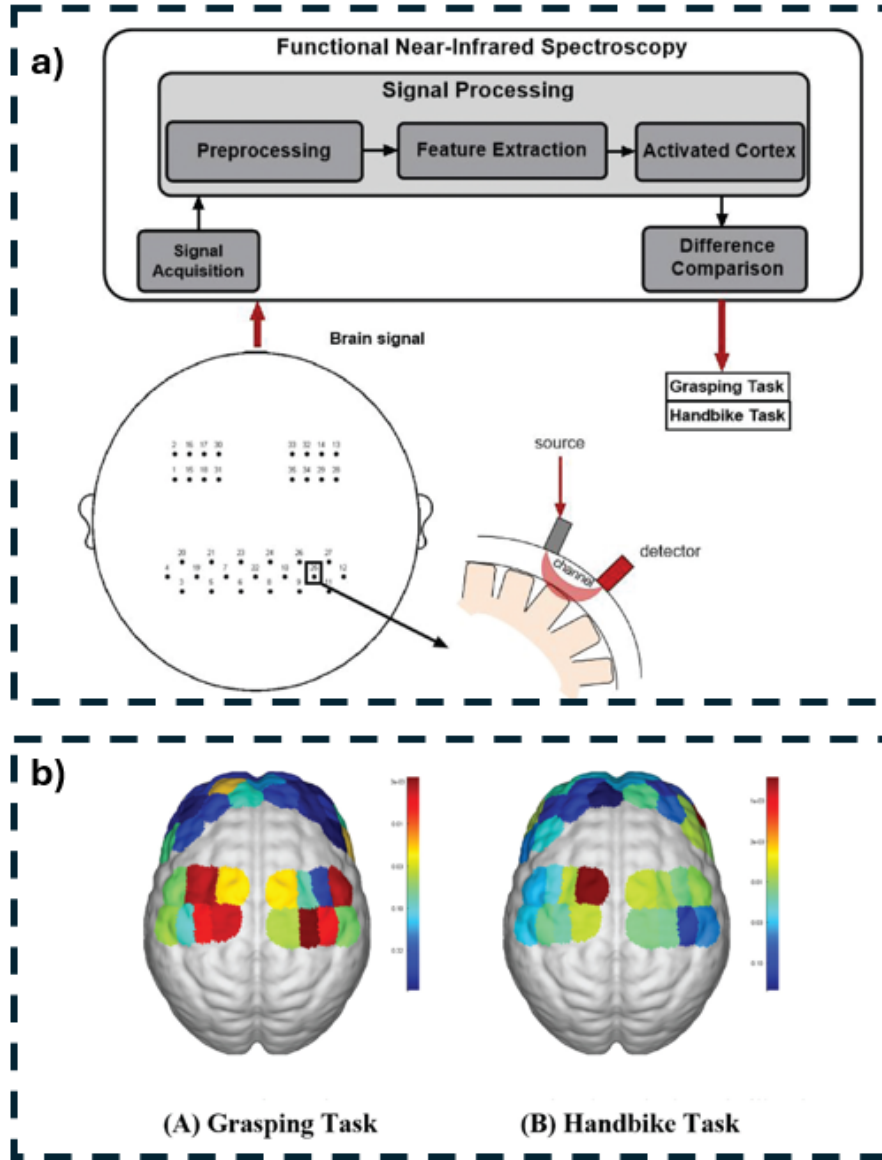


Figure 2.3: Example of fNIRS-BCI paradigm. (a) fNIRS acquisition and processing pipeline to distinguish a grasping from a handbike task. (b) Cortical activation maps rendered on a 3D brain model for the two motor tasks; the color scale encodes the amplitude of the haemodynamic response, from low (blue) to high (red). Adapted from [59]

EEG. The electrical activity of the brain arises from ionic currents flowing across neuronal membranes, driven by electrochemical gradients. The synchronised activity of large neuronal populations generates an electric potential field

that propagates through the surrounding tissue and can be recorded by surface electrodes placed on the scalp. This non-invasive recording technique is known as Electroencephalography (EEG). The first description of electrical activity from the brain cortex dates back to Richard Caton in 1875[60], while the first human EEG recording was performed by Hans Berger in 1924 and published in 1929[61]. Today, EEG is the most widely used signal acquisition technology for non-invasive BCIs[8, 62] due to several advantages. Most critically, it directly reflects neural electrical activity with millisecond-scale temporal resolution, making it uniquely suited to capture the rapid dynamics of cognitive processing. Furthermore, EEG systems are non-invasive, relatively low-cost, and portable, enabling their deployment in both clinical environments and naturalistic experimental settings[63]. On the other hand, EEG presents some limitations. Its spatial resolution, compared to haemodynamic brain imaging techniques such as fMRI, is inherently poor, on the order of 5–9 cm at the scalp for standard electrode configurations, due to the attenuation and distortion of electrical signals as they propagate through the skull and surrounding tissues. Furthermore, EEG recordings are susceptible to artifacts from muscle activity, eye movements, and other physiological sources, which can corrupt the signal. Finally, EEG captures primarily cortical surface activity and provides limited information about deep subcortical sources[63].

MEG. The ionic currents responsible for the brain’s electrical activity generate magnetic fields, a direct consequence of Ampère’s law and Maxwell’s equations. Magnetoencephalography (MEG) detects these fields, with intensities on the order of femtotesla, by measuring the magnetic flux through superconducting pickup coils (SQUIDs) placed on the scalp. The first detection of such fields was achieved by David Cohen in 1968 using a copper induction coil[64], and subsequently refined in 1972 with the introduction of SQUID-based sensors[65]. The progressive development of multi-channel SQUID arrays through the 1980s eventually led to the first whole-head MEG system in the early 1990s, marking the emergence of MEG as a practical neuroimaging tool[66]. Unlike electric fields, magnetic fields are only marginally attenuated by biological tissue, granting MEG a spatial resolution superior to that of EEG (Figure 2.4) while preserving the same millisecond-scale temporal resolution. These advantages, however, come at a considerable practical cost. SQUID sensors must be cooled to cryogenic temperatures with liquid helium, and, given the femtotesla-scale magnitude of the signals, the recordings must be acquired inside a magnetically shielded room in order to suppress ambient magnetic interference. As a result, MEG systems are bulky, expensive, and entirely non-portable, which restricts their use to specialised laboratory facilities and limits their accessibility relative to EEG, Figure 2.4(f). More recently, optically pumped magnetometers (OPMs) have emerged as a promising alternative that operates at room temperature and enables wearable, motion-tolerant MEG, although it

remains an active area of instrumentation research[67].

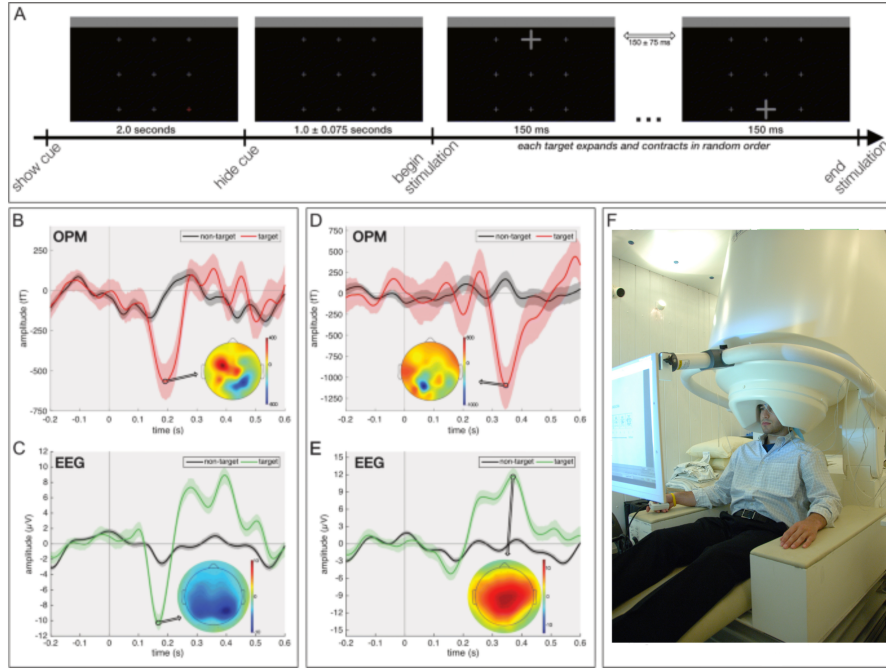


Figure 2.4: Example of MEG application. (a) Experimental protocol: a target cross is cued within a 3×3 grid and subjects mentally count its expansions during sequential stimulation. (b - c) Event-related responses (M/N200 component) recorded via OPM-MEG and scalp-EEG respectively; both modalities detect the event, but OPM-MEG exhibits a more localised scalp topography, reflecting its superior spatial resolution. (d-e) The same observation holds for the cognitive response (M/P300 component). Adapted from[68]. (f) Example of a traditional SQUID-based MEG system, illustrating the practical limitations of the technology: the bulky helmet-shaped sensor array and the requirement for a magnetically shielded room make the system inherently non-portable and confined to specialised laboratory facilities. (Source: National Institute of Mental Health, NIH, Wikimedia Commons)

Invasive methods: ECoG and Micro-electrode. A further family of neuroimaging methods also relies on measuring electrical potentials, but differs from EEG in sensor placement: the electrodes are surgically implanted in direct contact with the cortical tissue, bypassing the skull and scalp and positioning them closer to the neural sources. This proximity translates into substantially higher spatial resolution, signal-to-noise ratio, and accessible frequency range (most notably the high-gamma band)[69]. What distinguishes them from one another is essentially

their spatial scale and the depth at which the electrodes are implanted, Figure 2.5(a). Among these, ECoG and Micro-electrode arrays are the most relevant modalities in the context of semantic decoding research. Electrocorticography (ECoG) records electric potentials by means of electrode grids or strips placed directly on the cortical surface, either beneath the dura (subdural) or above it (epidural), Figure 2.5(b). Originally developed for presurgical mapping in refractory epilepsy, ECoG has since become a primary tool for speech and semantic BCI research, most notably demonstrated by Moses et al.[70], who achieved real-time decoding of intended speech with 94% accuracy in a paralysed patient. While ECoG electrodes rest on the cortical surface, micro-electrode arrays penetrate directly into the cortical tissue, enabling recordings from individual neurons or small neuronal populations at varying depths. This deeper implantation allows access to neural signals that surface recordings cannot capture, including both local field potentials and single-unit spiking activity. Among the most widely adopted devices of this class is the Utah Intracortical Electrode Array (UIEA)[71], which became the standard platform for invasive BCI research. While micro-electrode arrays provide the highest spatial resolution among all recording modalities, both ECoG and micro-electrode arrays require neurosurgery, which confines their application almost exclusively to clinical settings.

2.1.2 Why EEG for semantic decoding?

For semantic decoding, the choice of recording modality trades the fidelity of the neural signal against practical scalability. The best fidelity belongs to invasive recordings, whose direct cortical contact has enabled, for example, real-time speech decoding in a paralysed patient[70]; but this rests on neurosurgery, which is costly and risky, and restricts these methods to clinical populations, making them unsuitable for the large cohorts of healthy participants that a scalable system requires. The non-invasive methods are limited in different ways: fMRI has itself driven remarkable semantic results, up to the reconstruction of continuous language[4], yet its slow haemodynamic response limits temporal resolution and confines it to fixed, costly infrastructure, while MEG, though spatially superior, depends on cryogenic sensors and magnetic shielding that leave it equally non-portable. EEG alone combines the properties that matter here: direct, millisecond-resolution recording suited to the fast dynamics of semantic processing, together with non-invasiveness, low cost, portability, and a mature ecosystem of hardware and open datasets. Its one serious weakness, poor spatial resolution, is less limiting than it appears, since semantic representations are not focally localised but distributed across temporal, parietal, and prefrontal cortical regions[38], a spread-out organisation that EEG’s coarse spatial sampling captures reasonably well. EEG is therefore adopted throughout this thesis.

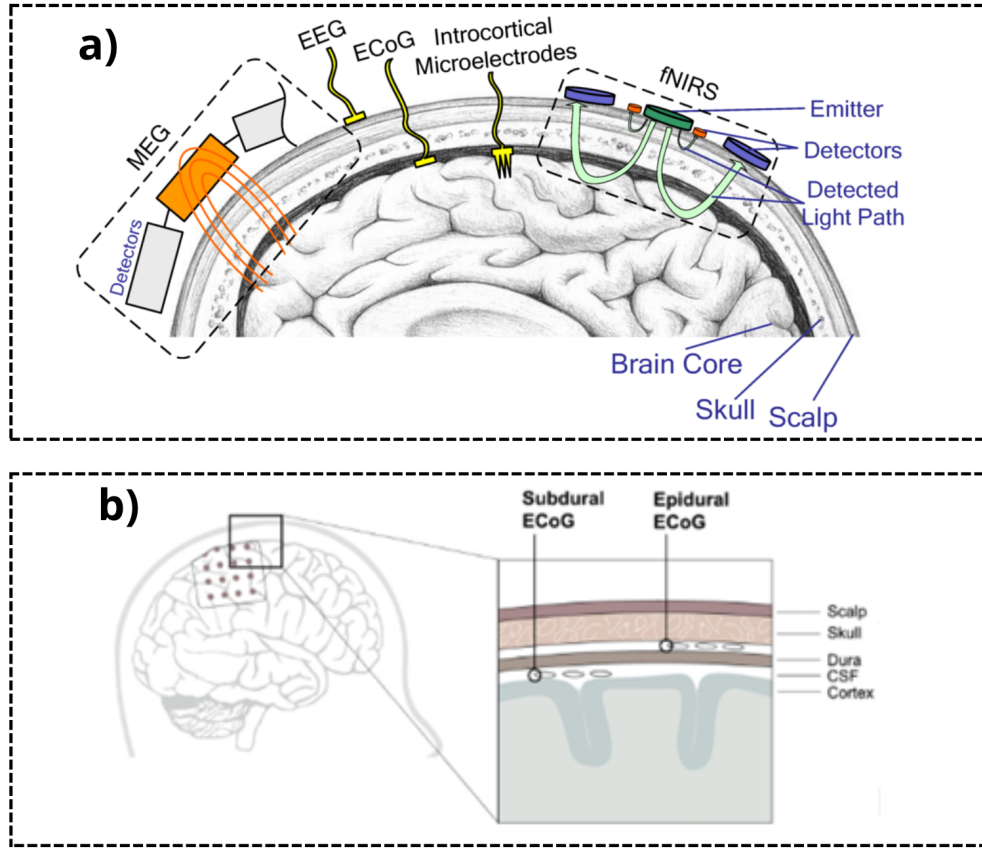


Figure 2.5: Scalp-reference position of neuroimaging devices. a) ECoG electrodes are placed directly on the cortical surface beneath the skull, eliminating the signal distortion introduced by bone and scalp. Intracortical microelectrodes penetrate the cortical tissue, enabling recordings at the level of individual neurons or small neuronal populations. The progressive increase in spatial proximity to neural sources comes at the cost of increasing surgical invasiveness. (Source: Peksa and Mamchur, Wikimedia Commons). b) Placement of subdural and epidural ECoG electrode grids relative to the cortical layers (scalp, skull, dura, CSF and cortex). Adapted from[72]

2.2 Electroencephalography Features (EEG)

The scalp EEG records the electrical potentials produced by the underlying cortical neurons. These do not arise from action potentials, which are too brief and too localised to summate at a distance, but from the slower postsynaptic potentials of cortical pyramidal neurons[69]. Aligned in parallel and oriented perpendicular to the cortical surface, these neurons act as current dipoles whose fields add constructively

when the population is active in synchrony and cancel when it is not, so that a measurable signal reaches the scalp only from large assemblies (tens of thousands to millions of neurons) acting in concert. The EEG thus reflects the synchronised activity of cortical populations rather than that of single neurons. On its way to the electrodes this field crosses the cerebrospinal fluid, skull, and scalp; the skull, a poor conductor, attenuates and spatially smears it, the volume-conduction effect behind EEG's limited spatial resolution (Section 2.1.1). The potentials reaching the scalp are correspondingly small, a few tens of microvolts, with each electrode recording the blurred sum of many sources. In practice the signal is recorded by scalp electrodes (commonly Ag/AgCl with conductive gel), each typically measured against a common reference and then amplified, filtered, and digitised at a few hundred to a thousand hertz. To ensure comparability across subjects and studies, electrode positions follow the International 10–20 system[73](Figure 2.6), which locates them at proportional fractions of the distances between skull landmarks (the nasion, inion, and preauricular points); its higher-density extensions, the 10–10 and 10–5 systems, apply the same principle to accommodate more electrodes.

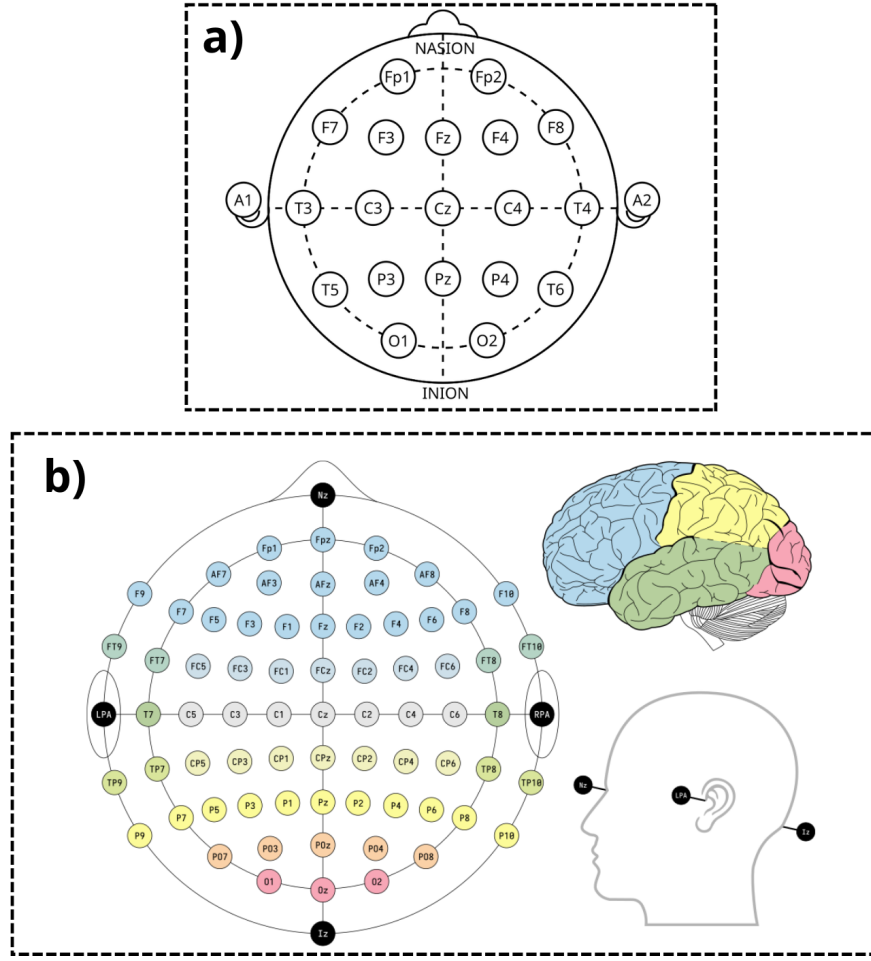


Figure 2.6: Example of recording systems. a) The standard 10–20 recording system. (Source: Wikimedia Commons) b) The 10–10 recording system, in which the electrode colours indicate the anatomical brain region (cerebral lobe) each electrode overlies. (Source: Laurens R. Krol, Wikimedia Commons)

2.2.1 EEG Artifacts

Not every fluctuation recorded at the scalp originates from the brain. Artifacts are components of the EEG that arise from non-cerebral sources and superimpose themselves on the genuine cortical activity, and their relevance can hardly be overstated: their amplitude is frequently far larger than that of the neural signal of interest, so that they may either bury the underlying activity or, more insidiously, imitate it. This is a particularly acute problem for a technique that, like EEG, already operates at low amplitude and with an unfavourable signal-to-noise ratio. Figure 2.7 illustrates how dramatically such contaminations can dominate a

recording, often masking the few tens of microvolts of the cortical signal. Artifacts are conventionally grouped into two broad classes according to their origin: physiological artifacts, generated by the subject's own body, and non-physiological (or technical) artifacts, which arise from the recording equipment and the surrounding environment. Physiological artifacts stem from biological sources other than cortical activity. The most prominent are the ocular artifacts produced by eye movements and blinks: because the eyeball behaves as an electrical dipole, a blink generates a large, low-frequency deflection that is maximal over the frontal electrodes and attenuates towards the back of the head (Figure 2.7.(1)). Muscular artifacts (EMG) arise from the contraction of the facial, jaw, neck, and scalp muscles and manifest as high-frequency, broadband bursts (Figure 2.7.(3)); because their spectrum overlaps that of the higher EEG rhythms, they are especially detrimental to the analysis of those bands. Cardiac activity contributes a further class of artifacts, either as the rhythmic electrical signature of the heartbeat (ECG) or as a pulse artifact when an electrode happens to lie over a blood vessel. Slower contaminations also occur, most notably the baseline drifts induced by perspiration and the modulation introduced by respiration. Non-physiological artifacts, by contrast, originate in the measurement system and its environment. Power-line interference, at 50 Hz in Europe and 60 Hz in North America, is the most ubiquitous example, coupled into the recording from the surrounding electrical mains. Electrode-related artifacts arise from poor or unstable electrode-skin contact, high electrode impedance, and the displacement of electrodes or cables; they typically appear either as abrupt transients, such as the electrode pop visible in Figure 2.7.(4), or as slow drifts. Broader environmental electromagnetic interference, together with movement of the participant or of the recording leads, completes this category. Because of their magnitude and ubiquity, the identification and suppression of artifacts is an essential stage of any EEG-processing pipeline. Two complementary strategies are commonly adopted. The first is artifact rejection, in which the contaminated segments of the recording are simply discarded; the second is artifact correction, in which the artifactual contribution is estimated and removed while the underlying neural signal is preserved. Correction is frequently achieved through blind source separation techniques, most notably Independent Component Analysis (ICA), which decomposes the multichannel recording into maximally independent components and allows those identified as artifactual, such as the ocular or muscular ones, to be discarded before the signal is reconstructed[74]; a complementary role is played by frequency-selective filtering, including the notch filters used to suppress power-line interference. The preprocessing pipeline adopted in the present work is described in detail in Chapter 3.5.2.

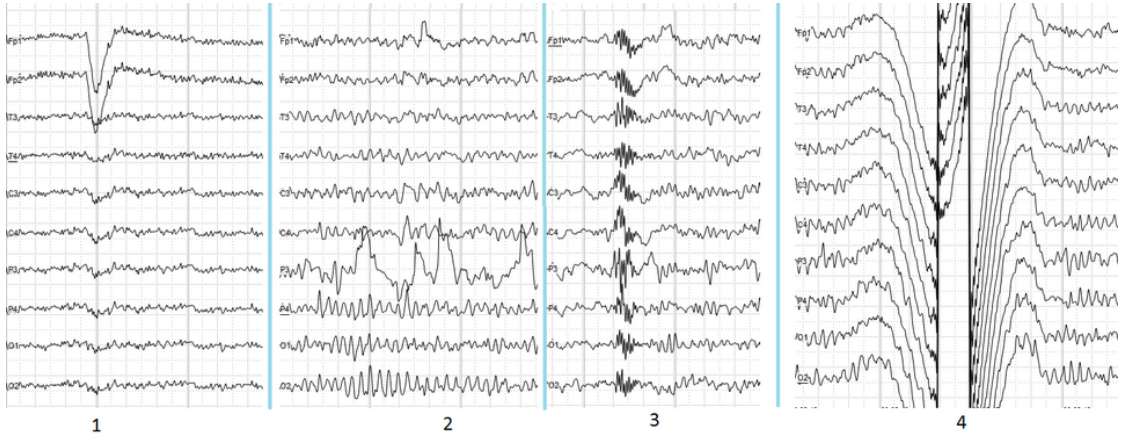


Figure 2.7: Representative examples of EEG artifacts. Four multichannel segments (10–20 montage: Fp1, Fp2, T3, T4, C3, C4, P3, P4, O1, O2) in which non-cerebral contaminations dominate the trace. (1) Ocular artifact: a large, low-frequency deflection produced by an eye blink, maximal over the frontal electrodes (Fp1, Fp2) and attenuating posteriorly. (2) Movement artifact: large-amplitude, slow rhythmic activity consistent with motion of the participant or of the recording leads, here most pronounced over the central and posterior channels. (3) Muscular (EMG) artifact: high-frequency, broadband bursts arising from muscle contraction, whose spectrum overlaps the higher EEG rhythms. (4) Electrode/movement artifact: synchronous high-amplitude slow deflections across all channels, accompanied by an abrupt transient (electrode pop). In every case the artifact amplitude far exceeds that of the genuine cortical signal. (Source: Andrii Cherninskyi, Wikimedia Commons)

2.2.2 Event Related Potentials

An Event-Related Potential (ERP) is the electrophysiological response of the brain that is time-locked to a specific event, be it a sensory stimulus, a cognitive operation, or a motor act. Because this response is small in comparison with the ongoing background EEG, it cannot generally be observed in a single trial; it is instead recovered by averaging the EEG epochs recorded around many repetitions of the same event. Under the assumption that the event-related response is constant across repetitions while the background activity is random with respect to the event, averaging causes the uncorrelated background to cancel progressively and the time-locked response to emerge, with the signal-to-noise ratio improving in proportion to the square root of the number of averaged trials. The resulting waveform is composed of a sequence of positive and negative deflections, termed components (Figure 2.8), which are characterised by their polarity, their latency relative to the eliciting event, and their distribution across the scalp. By convention, components

are labelled by a letter denoting their polarity, P for positive and N for negative, followed by their approximate latency in milliseconds or by an ordinal index (for example P1, N1, P2). The earliest components, occurring within roughly the first hundred milliseconds, are largely exogenous: they reflect the sensory and perceptual processing of the stimulus and depend chiefly on its physical properties. Later components are predominantly endogenous: they reflect higher-order cognitive processing and are comparatively insensitive to the low-level characteristics of the stimulus. Two endogenous components are of particular significance for the present work, since both demonstrate that scalp-level potentials are modulated by cognitive and semantic variables (Section 1.2). The P300 (or P3) is a positive deflection arising approximately 300 ms after stimulus onset, classically elicited by an infrequent or task-relevant stimulus embedded among frequent ones and linked to attention and to the resolution of stimulus uncertainty[10]. The N400, a negative deflection peaking at around 400 ms over centro-parietal regions, is specifically sensitive to the processing of meaning: it is enhanced when a stimulus is semantically incongruent with its context, and constitutes the canonical electrophysiological index of semantic processing[11]. The existence of such components provides the empirical foundation upon which the semantic decoding pursued in this thesis rests.

2.2.3 EEG frequency bands

Beyond its time-domain description, the EEG can also be characterised in the frequency domain, where its oscillatory activity is conventionally partitioned into a small number of frequency bands, each associated with particular physiological and cognitive states (Figure 2.9). Although the exact boundaries between these bands vary slightly across the literature, the principal bands, in order of increasing frequency, are the following. The delta band (δ , approximately 0.5–4 Hz) comprises the slowest and highest-amplitude rhythms and predominates during deep, slow-wave sleep; its presence in the waking adult is generally regarded as pathological. The theta band (θ , 4–8 Hz) emerges during drowsiness and light sleep and has additionally been linked to memory processes and spatial navigation. The alpha band (α , 8–13 Hz) characterises relaxed wakefulness with the eyes closed; it is most pronounced over posterior (occipital) regions and is attenuated upon opening the eyes or engaging in mental effort, a phenomenon known as alpha blocking, or desynchronisation. The beta band (β , 13–30 Hz) is associated with active, alert cognition and with motor control, the sensorimotor rhythms it contains being modulated during the planning and execution of movement. Finally, the gamma band (γ , above 30 Hz) has been related to higher cognitive functions such as attention, perceptual binding, and conscious perception; owing to its low amplitude and high frequency, it is especially vulnerable to contamination

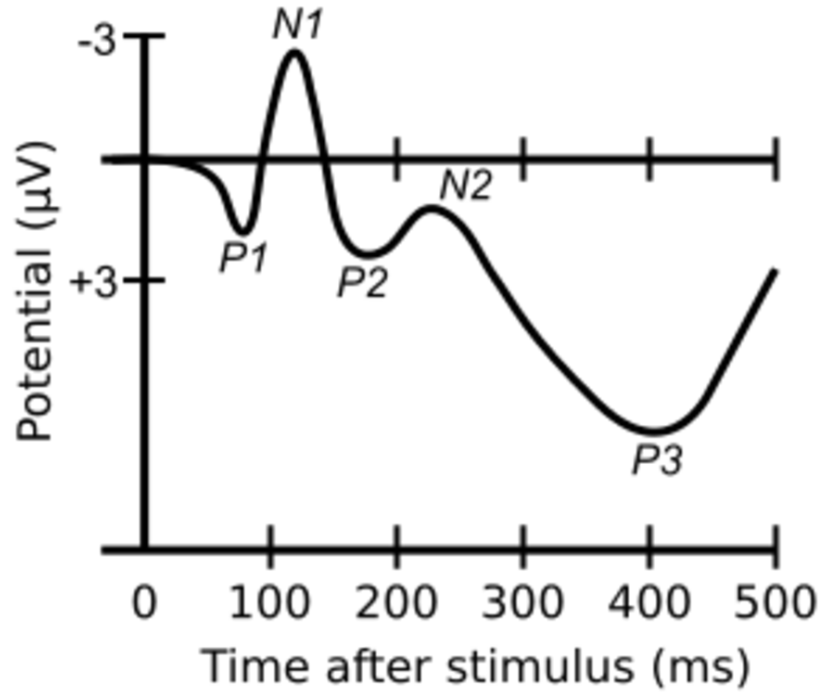


Figure 2.8: Schematic representation of an averaged ERP waveform and its components. The event-related potential is displayed as voltage plotted against time from stimulus onset (0 ms). The waveform unfolds as a stereotyped sequence of positive (P) and negative (N) deflections, each defined by its polarity and its peak latency. The early components (within roughly the first 100 ms) are largely exogenous and reflect the sensory and perceptual processing of the stimulus, whereas the later endogenous components, most notably the P300 and the N400, index higher-order cognitive and semantic processing. (Source: Wikimedia Commons)

by muscular artifacts (Section 2.2.1). The distribution of signal power across these bands constitutes one of the most widely used families of features in EEG classification, since different cognitive and perceptual processes are reflected in different portions of the spectrum[9]. This spectral perspective is directly relevant to the present work, as several of the decoding pipelines described in Section 3.5.2 operate on band-limited representations of the signal. In semantic EEG analysis, band-limited descriptors have also been investigated through power- and entropy-based approaches, including studies relating PSD features to perception, imagination, and imagery vividness, as well as entropy measures for semantic concept differentiation in alpha and beta rhythms[75, 76].

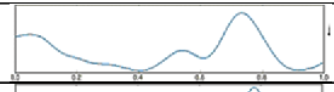
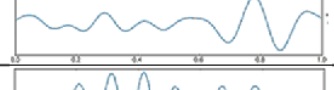
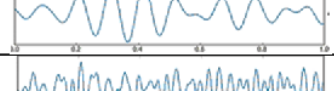
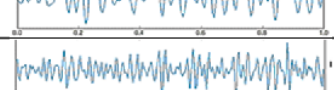
Waves	Frequency bands (Hz)	Behaviour Trait	Signal Waveform
Delta	0.3 – 4	Deep sleep	
Theta	4 – 8	Deep Meditation	
Alpha	8 – 13	Eyes closed, awake	
Beta	13 – 30	Eyes opened, thinking	
Gamma	30 and above	Unifying consciousness	

Figure 2.9: Canonical EEG frequency bands. For each band, the table reports the approximate frequency range, a representative behavioural or cognitive state, and an illustrative signal waveform. Adapted from[77].

2.3 Spectral and Time–Frequency Analysis

The spectral description of the EEG introduced in Section 2.2.3 presupposes a means of estimating how the power of the signal is distributed across frequency. This estimate is not unique: it depends on the analytical method adopted, and the choice of method is far from a technical detail, since it determines the very quantities (band powers, time–frequency maps, spectral coefficients) that are subsequently passed to the classifier as features. Three families of techniques are of particular relevance to the present work and, more broadly, dominate the EEG-processing literature. The first comprises the non-parametric methods founded on the Fourier transform, which estimate the spectrum directly from the data; the second extends these to the time–frequency plane, relaxing the assumption of stationarity through wavelet decompositions; and the third replaces the direct estimate with a parametric model of the signal. The remainder of this section examines each in turn.

Fourier-based Spectral Estimation. The Fourier transform expresses a signal as a superposition of sinusoids, resolving it into the frequencies of which it is composed. For a signal acquired in discrete time over a finite interval, as the digitised EEG necessarily is, the relevant operator is the Discrete Fourier Transform (DFT), whose efficient evaluation through the Fast Fourier Transform (FFT) algorithm underpins virtually all practical spectral analysis[78]. The DFT of an

N -sample epoch $x[n]$ is

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi kn/N}, \quad k = 0, \dots, N-1, \quad (2.1)$$

and the natural estimate of the power distribution that follows from it is the *periodogram*, the squared magnitude of the transform normalised by the record length,

$$\hat{S}_x(f_k) = \frac{1}{N} |X[k]|^2. \quad (2.2)$$

The periodogram, however, is a statistically poor estimator of the Power Spectral Density (PSD): its variance does not diminish as the record lengthens, so that the estimate remains erratic regardless of the quantity of data available, and the abrupt truncation of the signal at the edges of the analysis window introduces *spectral leakage*, whereby power from one frequency is smeared across neighbouring bins[79]. The method of Welch[80] addresses both deficiencies and has become the de facto standard for EEG band-power estimation. The epoch is divided into K shorter, overlapping segments; each segment is multiplied by a tapering window, typically a Hann window, that smoothly attenuates its edges and thereby suppresses leakage; a periodogram is computed for every segment; and the resulting estimates are averaged:

$$\hat{S}_x^W(f) = \frac{1}{K} \sum_{i=1}^K \hat{S}_x^{(i)}(f). \quad (2.3)$$

The averaging in (2.3) reduces the variance of the estimate by roughly a factor of K , yielding the smooth and stable spectra on which frequency-band features rest, at the cost of a coarser frequency resolution dictated by the shortened segment length, a trade-off characteristic of non-parametric estimation. Underlying the entire construction is the assumption that the signal is at least weakly stationary over the analysis window, an idealisation that the EEG, with its evolving and event-locked dynamics, only approximately satisfies.

Time–Frequency Analysis and the Morlet Wavelet. The Fourier-based estimate returns a single spectrum for the entire epoch and, in doing so, discards all information about *when* a given oscillation occurred. For a stationary signal this loss is immaterial, but the EEG is intrinsically non-stationary: its spectral content evolves over time and, in the event-related paradigms central to this thesis, the very quantities of interest are transient modulations of power time-locked to a stimulus. A representation that resolves the signal jointly in time and frequency is therefore required. The Short-Time Fourier Transform (STFT), which applies the Fourier transform within a sliding window, constitutes the first such remedy, but it imposes a single fixed window, and hence a single, uniform resolution, across

the whole time–frequency plane, a rigidity that the uncertainty principle renders fundamentally limiting: a narrow window localises events in time but blurs them in frequency, and a wide window does the converse[81]. The Continuous Wavelet Transform (CWT) overcomes this limitation by replacing the fixed window with a family of scaled and translated copies of a single prototype function, the *mother wavelet*. The transform correlates the signal $x(t)$ with these dilated and shifted wavelets,

$$W_x(s, \tau) = \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} x(t) \psi^*\left(\frac{t - \tau}{s}\right) dt, \quad (2.4)$$

where s is the scale, inversely related to frequency, and τ the temporal shift. Because the wavelets are compressed at small scales and dilated at large ones, the CWT affords fine temporal resolution at high frequencies and fine frequency resolution at low frequencies, a multi-resolution behaviour naturally suited to the multi-scale organisation of neural oscillations. Among the available mother wavelets, the complex Morlet wavelet is by far the most widely adopted for the time–frequency analysis of the EEG[82]. It consists of a complex exponential modulated by a Gaussian envelope,

$$\psi(t) = \pi^{-1/4} e^{j\omega_0 t} e^{-t^2/2}, \quad (2.5)$$

in which ω_0 governs the number of oscillatory cycles within the envelope and thereby sets the trade-off between temporal and spectral precision. The Morlet wavelet owes its popularity to two properties: its Gaussian envelope attains the theoretical lower bound on the joint time–frequency uncertainty, so that no other wavelet localises an event more sharply; and, being complex, it yields at each point of the plane both an instantaneous amplitude and an instantaneous phase. The squared modulus of the resulting coefficients constitutes a time-resolved estimate of power, commonly referred to as a scalogram, from which event-related changes in band power can be quantified directly, while the phase supports the analysis of inter-trial consistency and of coupling between cortical regions.

Parametric Spectral Estimation. The methods considered so far estimate the spectrum directly from the observations, without positing any structure for the process that generated them. Parametric estimation proceeds from the opposite premise: the signal is assumed to be the output of a linear system excited by white noise, and the problem of estimating the spectrum is reduced to that of identifying the parameters of this system. For the EEG the autoregressive (AR) model is the customary choice, in which each sample is expressed as a linear combination of its p predecessors together with a white-noise innovation $e[n]$,

$$x[n] = - \sum_{k=1}^p a_k x[n - k] + e[n]. \quad (2.6)$$

Once the coefficients $\{a_k\}$ and the innovation variance σ^2 have been determined, the PSD follows in closed form as

$$\hat{S}_x(f) = \frac{\sigma^2}{|1 + \sum_{k=1}^p a_k e^{-j2\pi f k}|^2}, \quad (2.7)$$

a continuous function of frequency that, unlike the periodogram, is not tied to a discrete grid of bins. The coefficients themselves may be obtained in several ways: the Yule–Walker method solves the system of normal equations relating them to the estimated autocorrelation sequence, whereas the Burg method estimates them by minimising the forward and backward prediction errors jointly, without recourse to the autocorrelation, an approach that proves more accurate on the short records typical of single EEG epochs[83]. The principal advantage of the parametric approach is precisely this capacity to yield smooth, high-resolution spectra from comparatively brief segments, free of the leakage that afflicts the Fourier estimate; its principal liability is the need to select the model order p , since too low a value over-smooths the spectrum and obscures genuine peaks, while too high a value introduces spurious detail by fitting the noise[79]. Taken together, these three families of methods furnish the frequency-domain and time–frequency representations from which the features of the following section are derived. The band powers obtained by integrating a Welch spectrum over the canonical bands, the time-resolved power yielded by a Morlet decomposition, and the smooth spectra of an autoregressive model are not competing answers to a single question but complementary descriptions, each emphasising a different facet of the signal, and the feature-extraction stage of the machine-learning pipeline examined next is built upon representations of exactly this kind.

2.4 Machine Learning on the EEG

Machine learning encompasses the family of computational methods that infer predictive models directly from data, in place of explicitly programmed decision rules[84]. In the supervised paradigm most pertinent to brain–computer interfacing, a model is exposed to a collection of labelled examples (pairs of input observations and their corresponding target classes) and is optimised to recover a mapping that generalises from these examples to previously unseen data. Cast in these terms, neural decoding becomes a problem of statistical pattern recognition: given a representation extracted from a segment of brain activity, a classifier is trained to predict the associated stimulus, task, or mental state. Conventionally, this process is decomposed into two largely separable stages: *feature extraction*, in which the raw signal is transformed into a compact and discriminative representation, and *classification*, in which a decision function is fitted over that representation[9].

The transfer of this pipeline to the EEG is complicated by the characteristic properties of the signal. EEG recordings are high-dimensional, comprising tens to hundreds of channels sampled at hundreds of hertz, whereas the number of trials obtainable from a single participant is comparatively limited, a regime in which classifiers readily overfit and the curse of dimensionality becomes a dominant concern[9]. The signal is further marked by a low signal-to-noise ratio, pronounced non-stationarity within and across recording sessions, and sensitivity to ocular and muscular artifacts. For these reasons, classical EEG decoding seldom operates on raw voltages, relying instead on a carefully engineered feature-extraction stage that compresses each trial into a low-dimensional and discriminative representation, simultaneously isolating the information relevant to the task and keeping the number of inputs commensurate with the limited number of available trials. The features employed for EEG classification are conventionally grouped by the domain in which they are computed. Time-domain descriptors capture the morphology of the signal directly, through amplitude samples, statistical moments, or the peak amplitudes and latencies of event-related potentials; frequency-domain descriptors, of which band power and the power spectral density are the canonical instances, summarise the energy distributed across the physiologically defined oscillatory bands (Section 2.2.3); and time-frequency descriptors, typically derived from wavelet decompositions, preserve the temporal evolution of these spectral components[9]. Irrespective of the representation adopted, the resulting feature vector is frequently high-dimensional, and an unsupervised dimensionality-reduction step is often interposed before classification in order to discard redundant or noise-dominated directions. Principal Component Analysis (PCA) is the most widespread such technique: it re-expresses the data along mutually orthogonal axes of decreasing variance, so that the representation may be truncated to its leading components, thereby decorrelating the features and curbing the propensity to overfit before they are passed to a downstream classifier[84]. A complementary and especially influential line of work exploits the spatial structure of the multichannel recording. Common Spatial Patterns (CSP) is a supervised spatial-filtering technique that learns linear combinations of channels maximising the variance of the projected signal for one class while minimising it for the other; the log-variances of the resulting components then constitute a compact and highly discriminative feature set, an approach first devised for motor imagery and subsequently adopted across BCI paradigms[35, 36]. More recently, representations grounded in Riemannian geometry have established themselves as a powerful alternative that dispenses with explicit spatial filtering. Each trial is summarised by its spatial covariance matrix, which jointly encodes the variance of, and the correlation between, the recording channels. Because covariance matrices are symmetric and positive-definite, they do not populate a flat Euclidean space but a curved manifold whose intrinsic geometry must be respected; operating directly on this manifold, for example

by projecting each matrix onto the tangent space at the geometric mean of the training set, where ordinary vector-space classifiers become applicable, yields representations that are notably robust to the linear distortions and inter-session shifts affecting the EEG, and that have proven accurate, computationally simple, and comparatively free of tuning parameters across a range of BCI tasks[85, 86]. Over features derived in any of these ways, the decision itself is generally entrusted to a shallow classifier; linear discriminant analysis and support vector machines remain the default choices on account of their robustness in the small-sample, high-dimensional EEG setting, with logistic regression and ensemble methods such as random forests also in common use[9]. Ensemble-learning strategies have recently been explored in motor-imagery EEG classification, including weighted and stacked adaptive ensemble models and correlation-optimised stacking frameworks designed to improve robustness across datasets and classifiers[87, 88]. This handcrafted-feature paradigm underpinned the earliest evidence that conceptual content is recoverable from scalp EEG: through multivariate pattern analysis, that is, the application of a supervised classifier to spatially or temporally distributed single-trial features, Simanova et al.[14] and Murphy et al.[15] discriminated broad semantic categories above chance, establishing machine learning as a viable instrument for probing the neural representation of meaning. On the multimodal dataset adopted in the present work[18], traditional machine learning likewise furnishes the baseline against which subsequent approaches are assessed. As part of their technical validation, Wilson et al.[18] applied a logistic-regression classifier to the perception–imagination distinction within each sensory modality, obtaining higher accuracy for the visual conditions and substantially lower accuracy for the auditory condition; Khanday et al.[26] subsequently examined a broader set of shallow classifiers (random forests, support vector classifiers, and gradient-boosted trees) on the same data, finding that, while they attained moderate accuracy on the perception–imagination task, they overfitted markedly and, on the three-way classification of semantic concept, none surpassed chance level. Taken together, these findings illustrate both the feasibility and the persistent difficulty of applying classical machine learning to semantic EEG decoding, and set the stage for the deep-learning approaches examined in the following section.

2.5 Deep Learning on the EEG

Whereas the classical pipeline of the preceding section rests on a deliberate separation between a handcrafted feature-extraction stage and a downstream classifier, deep learning dissolves this distinction: a hierarchy of trainable non-linear transformations is optimised end-to-end, so that the representation most useful for the task is discovered directly from the data rather than prescribed in advance

by the analyst. Successive layers compose increasingly abstract descriptors, from local patterns in the raw signal to global, task-relevant abstractions, and the parameters of every layer are adjusted jointly by gradient descent under a single classification objective. Propelled by the growing availability of data and of parallel computing hardware, this paradigm has come to dominate pattern recognition across domains, and EEG decoding is no exception[31]. The transfer of deep learning to the EEG is, however, far from straightforward. The very properties that complicate classical decoding (high dimensionality, low signal-to-noise ratio, pronounced non-stationarity, and a scarcity of trials relative to the number of free parameters) are, if anything, more acute for models comprising millions of weights, which overfit readily in this small-sample regime[9, 31]. The response has been the design of architectures whose inductive biases are tailored to the structure of the signal. Convolutional Neural Networks (CNNs) predominate: by sliding learnable filters along the temporal and spatial axes, they exploit the same channel-and-time organisation that motivates spatial filtering, while sharing parameters across the recording to economise on data. In parallel, recent EEG representation-learning studies have explored topology-preserving and task-independent deep architectures, including convolutional autoencoders, recurrent or Transformer-based temporal modules, and latent-space diagnostics for semantic EEG analysis[89, 30]. The compact EEGNet architecture[90], which factorises spatial and temporal filtering through depthwise and separable convolutions under a minimal parameter budget, and the deeper Deep ConvNet and the shallower Shallow ConvNet of Schirrneister et al.[32] have become de facto baselines across BCI paradigms. Recurrent networks and Long Short-Term Memory units have been employed to model the temporal evolution of the signal, and more recent work has turned to attention mechanisms and Transformer architectures in order to capture long-range dependencies[31]. A parallel line of research confronts the scarcity of labelled trials directly, through self-supervised and contrastive objectives that learn useful representations from unlabelled or weakly labelled data prior to fine-tuning on the decoding task. These gains in predictive accuracy come at a recognised cost. End-to-end learned representations are difficult to interpret: as already observed in Section 1.3, deep networks are routinely described as opaque “black boxes” whose internal computations resist direct neuroscientific reading[33], and systematic reviews report that interpretability and predictive performance frequently stand in tension[31, 34]. This opacity is consequential whenever the analytical objective extends beyond classification accuracy to an understanding of the neural representations themselves, which is precisely the concern that motivates the interpretable, spatially grounded approach pursued in the present work. Within semantic decoding specifically, and on the multimodal dataset of Wilson et al.[18] adopted here, deep learning constitutes the bulk of the recent literature. Among these studies, the CNN of Khanday et al.[26], supplied jointly with raw epochs and wavelet-derived spectral descriptors,

separated perception from imagination at an accuracy comfortably above that of the shallow classifiers evaluated alongside it; on the three-way classification of semantic concept, however, the same architecture did not exceed chance. A markedly higher accuracy on the perception–imagination contrast was subsequently reported by Huang et al.[22], using a spatio-temporal capsule network with self-correlation routing applied across the three modalities. A distinct and influential strand, developed by Tong and Chen, departs from the prevailing single-process focus by treating the documented similarity between perception and imagination as an explicit modelling principle. Their end-to-end PI Net[25] couples a dynamic convolution operating jointly in the spatial, temporal and frequency domains with a gated linear-attention unit; embedding all conditions within a single framework, it attains high accuracy on the binary perception–imagination task but much lower accuracy on a demanding eighteen-way classification spanning task, modality and concept, while its ablation analysis localises the most informative content to the delta band and the parietal region. MMPI Net[23] recasts the problem as one of multimodal fusion, pairing perception and imagination epochs of common semantics through a cross-attention module with shared parameters and decoding the fused representation with a Kolmogorov–Arnold Network; relative to a single-process baseline it raises semantic accuracy on the visual conditions. SSIDNet[24] pursues the same intuition through a parameter-sharing branch dedicated to the semantic information common to the two cognitive processes and a capsule network that fuses it with process-specific features, lifting accuracy on the visual pictorial conditions. A related direction has recently considered semantic-aware decoding of covert inner speech, where multimodal overt supervision from EEG, EMG, and audio is used to support subject-held-out decoding of inner-speech commands from non-invasive EEG[91]. A complementary contribution by Tran et al.[27] applies contrastive learning to the perception data alone, forming positive and negative pairs across the visual and auditory modalities and reporting that the resulting cross-sensory representation improves object classification over a convolutional baseline, while confirming that visual stimuli elicit more discriminable responses than auditory ones. Two features of this body of work bear directly on the present study. First, with the partial exception of the eighteen-way analysis of Tong and Chen[25], the deep-learning literature on this dataset concentrates overwhelmingly on the binary perception–imagination distinction; the discrimination of sensory modality and, above all, the classification of semantic concept receive markedly less attention, and where the latter has been attempted it has proved considerably harder. Second, as detailed in Section 1.3, these models are almost without exception evaluated within subject or on pooled data, so that their reported accuracies do not speak to cross-subject generalisation. The sole departure is the topographic approach of Ahmadi and Mesin[28], which renders each epoch as a sequence of scalp-voltage maps decoded by a convolutional network and is assessed under a

strict leave-one-subject-out protocol, representing the first such evaluation reported on the dataset and the basis for the third pipeline examined in this thesis. This subject-independent line has most recently been extended by the same group: Ahmadi, Impagnatiello and Mesin[29] learn, in place of the explicit scalp-voltage maps, a task-independent hierarchical autoencoder (a topology-preserving convolutional stage feeding a Transformer) whose unsupervised embeddings, decoded by lightweight per-modality heads under leave-one-subject-out, attain high accuracy with low across-subject variance on the perception–imagination distinction across the three modalities. Their emphasis, however, falls on interpretability rather than accuracy: through a correlation-based pruning of the latent space they derive two representation-level diagnostics, a Semantic Sensitivity Index and a cross-modal overlap measure, which they relate to self-reported imagery vividness (VVIQ and BAIS), finding that imagination occupies a more compact, less redundant subset than perception in several participants and that a substantial modality-invariant core spans the three modalities, even where the direct accuracy–vividness correlations remain weak. Taken together, the predictive strength of deep learning, its characteristic opacity, and the narrowness of both the tasks addressed and the evaluation regimes adopted delineate the space within which the present work is situated: an interpretable, spatially grounded analysis of all three classification tasks, conducted throughout under subject-independent evaluation.

Chapter 3

Materials & Methods

3.1 Data Set

The experiments conducted in this thesis rely on the open-source EEG dataset released by Wilson et al.[18] which, as declared by the authors, constitutes the first publicly available dataset to capture semantic representation across several sensory modalities and for both perception and imagination within a single participant sample. The released data are distributed through the OpenNeuro repository and organised according to the Brain Imaging Data Structure (BIDS) specification, with both the raw recordings and an author-provided pre-processed version made available.

Participants. Twelve participants took part in the study, and the majority of them were students at the University of Bath. The study was approved by the institution’s Psychology Research Ethics Committee, and every participant gave informed consent both to take part and to the subsequent public release of their data. Nine of the twelve participants completed a single recording session, while three returned for a second, so that the dataset contains fifteen sessions in total; each session lasted approximately two hours. The inclusion criteria were initially restricted to participants with normal or corrected vision and hearing, with epilepsy as an exclusion criterion; they were subsequently extended to include participants with sensory impairments. One participant with a visual and hearing impairment is consequently included in the final sample and is flagged as such in the repository.

Semantic concepts. The categories to be decoded are three concrete concepts: *flower*, *penguin* and *guitar*. These were selected to be maximally distinct in meaning while remaining matched on the low-level attributes most liable to confound a semantic decoder. Semantic separation was assessed by embedding each word

in a Word2Vec vector space, namely a model that represents words as high-dimensional vectors such that semantically similar words lie close together[18], confirming that the three concepts were mutually distant within it; the resulting configuration was then visualised in two dimensions by means of t-distributed stochastic neighbour embedding (t-SNE), a dimensionality reduction technique that projects high-dimensional data into a two-dimensional space preserving local structure. Two further considerations governed the selection. First, the three words were constrained to share the same number of syllables (two each): in the auditory paradigm this control is essential, since holding syllable count constant ensures that classification is driven by meaning rather than by word length. Second, deliberately uncommon objects were preferred over everyday ones, both to avoid the unpredictable semantic associations that familiar items accrue through daily use, and to avoid categories, such as faces, in which perceptual expertise can elicit spatially clustered, modular cortical responses that would limit the generality of the findings.

Experimental paradigms and stimuli. The experiment is built from six paradigm variations, obtained by crossing two cognitive tasks (perception and imagination) with three sensory modalities through which each concept is presented: visual pictorial, visual orthographic, and auditory comprehension. Within every modality the participant first perceived a stimulus and subsequently imagined it. In the *visual pictorial* modality the stimuli are coloured images of the three objects, rendered at 1200×1200 pixels on a black background and deliberately stripped of any surrounding scene so as not to introduce extraneous semantic associations or to inflate the difficulty of the imagination task; each concept is depicted at three levels of visual complexity: simple, intermediate and naturalistic. For *flower* and *guitar*, eight simple and nine intermediate and nine naturalistic exemplars are provided, while *penguin* has nine exemplars at each level. In the *visual orthographic* modality the stimuli are the written word forms of the three concepts, presented on a 1200×1200 pixel white background, with thirty exemplars per concept spanning five font colours (black, blue, red, green and purple) and six font styles. In the *auditory comprehension* modality the stimuli are two-second audio recordings of the three words, produced by speakers who took no part in the EEG sessions and denoised in Audacity, each word recorded in a normal, low or high pitch; in the corresponding imagination condition participants were instructed to re-imagine the word in the speaker’s voice rather than in their own inner voice. This pronounced within-category variability (multiple exemplars, levels of complexity, colours, fonts and voices) is the central design principle of the dataset: by ensuring that stimuli belonging to the same concept differ substantially in their low-level sensory properties, the decoder is forced to rely on semantic content rather than on superficial perceptual features. Examples of the pictorial and orthographic stimuli

are shown in Figure 3.1.

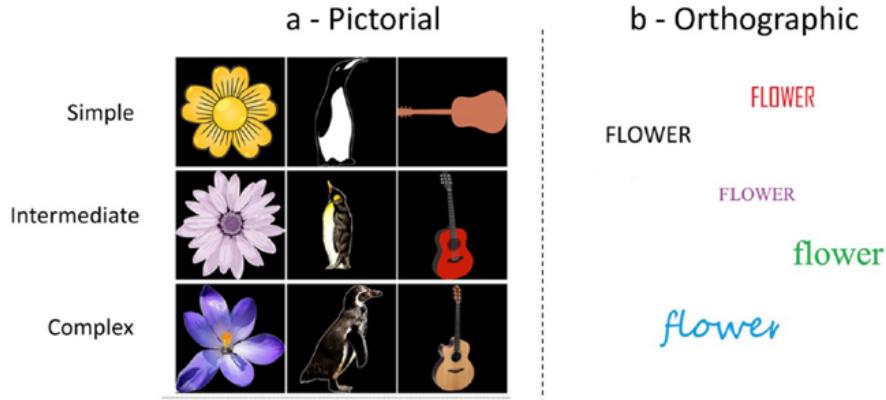


Figure 3.1: Examples of visual stimuli used in the experiment. (a) Pictorial stimuli at three levels of complexity: simple, intermediate and naturalistic, for *flower*, *penguin* and *guitar*. (b) Orthographic stimuli varying in colour and font style. Adapted from[18]

Trial structure and epoch duration. Stimuli were presented with PsychoPy (v2021.2.3)[18]. The onset of each stimulus was time-stamped via a Lab Streaming Layer (LSL) network, which synchronised the presentation computer with the EEG acquisition software. Trials were organised into blocks, each presenting every concept five times per modality; the full protocol consisted of ten blocks, separated by participant-paced breaks, as illustrated in Figure 3.2. Due to fatigue, most participants did not complete all ten blocks, resulting in an imbalance in the number of trials across participants: the realised counts range from a minimum of 64 to a maximum of 150 trials per condition. Epoch duration depends on the condition: imagination epochs last four seconds throughout, while perception epochs last two seconds for the auditory modality and three seconds for the two visual modalities, as summarised in Table3.1.

Table 3.1: Duration of the analysis window for each task–modality condition, following the original acquisition protocol[18].

Task	Sensory modality	Duration
Perception	Auditory	2 s
Perception	Orthographic	3 s
Perception	Pictorial	3 s
Imagination	Auditory	4 s
Imagination	Orthographic	4 s
Imagination	Pictorial	4 s

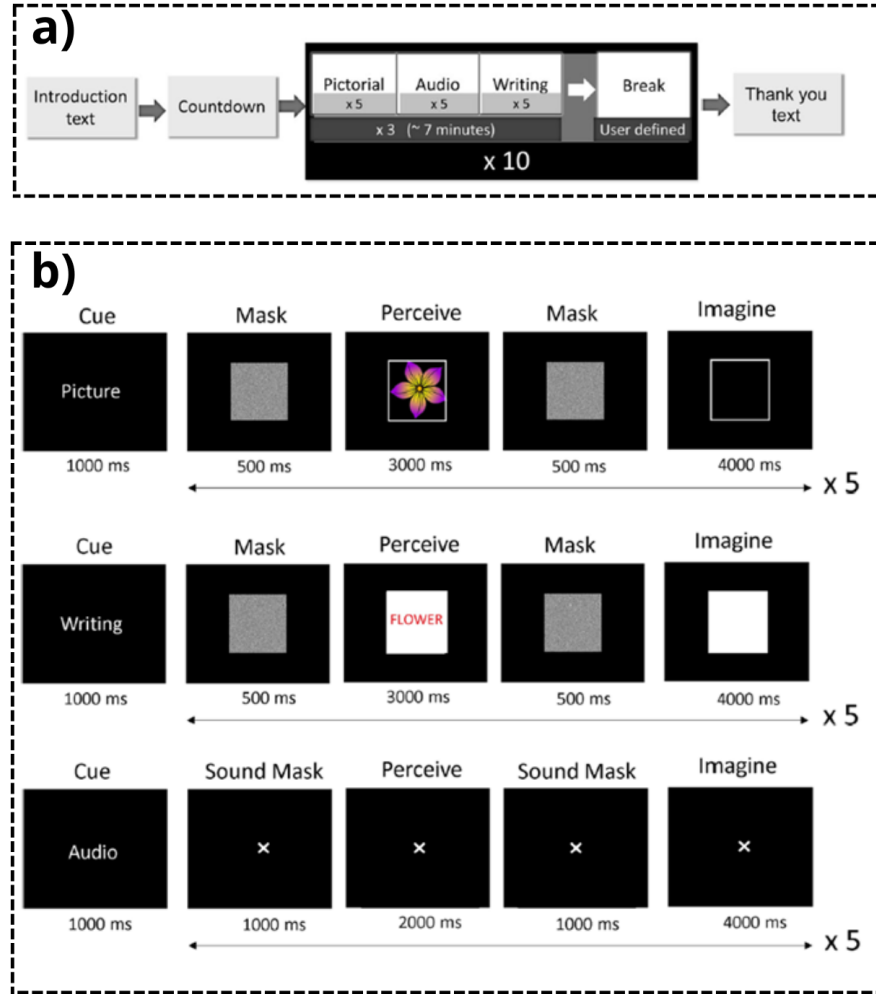


Figure 3.2: Schematic representation of the experimental protocol. (a) Overview of the session structure. (b) Temporal structure of individual trials for each modality. In the pictorial and orthographic conditions, each trial consists of a cue (1000 ms), a visual mask (500 ms), a perception phase (3000 ms), a second mask (500 ms) and an imagination phase (4000 ms). In the auditory condition, visual masks are replaced by sound masks (1000 ms each) and the perception phase lasts 2000 ms. Each sequence is repeated five times per block. Adapted from [18].

Imagery vividness. Because an individual’s capacity for mental imagery is known to modulate the neural correlates of imagination, and may therefore influence both the decodability of imagined content and the strategy best suited to decoding it, the subjective vividness of each participant’s imagery was assessed before the recording through two established self-report instruments. The Vividness

of Visual Imagery Questionnaire (VVIQ), scored from one to five, measures the subjective clarity of an imagined visual scene, while the vividness subscale of the Bucknell Auditory Imagery Scale (BAIS-V), scored from one to seven, measures the corresponding clarity of an imagined sound. The mean scores were 3.75 (SD 0.55) for the VVIQ and 4.76 (SD 0.85) for the BAIS-V, and the two measures were significantly correlated (Spearman’s $\rho = 0.79$, $p = 0.007$); the questionnaires were not completed by two participants. It should be noted that, although these metrics index an individual’s general imagery ability, they are more likely to bear on the decodability of low-level sensory content than on that of semantic representation as such. This interpretation is consistent with previous EEG analyses of the same semantic dataset, where PSD-based features and VVIQ/BAIS measures were examined to study the relation between perception, imagination, and subjective vividness[75], and with representation-level analyses suggesting that vividness may be more informative when related to latent-space geometry than to accuracy alone[29]

Acquisition and data format. The EEG was recorded with a 128-channel ANT Neuro eego Mylab system providing 124 scalp electrodes, mounted in a gel-based waveguard cap whose active shielding suppresses 50/60 Hz environmental interference; the signals were digitised at 1024 Hz with 24-bit resolution. Electrode positions follow the five-percent (10–5) system, a high-density extension of the international 10–20 montage; the recording was referenced to CPz, with the ground placed on the left mastoid, and electrode impedances were kept below 50 k Ω . For each participant and session the repository provides two versions of the data. The first is the continuous raw recording, converted from the native ANT Neuro `.cnt/.evt` files into a format compatible with the MNE-Python ecosystem and aligned to the BIDS event convention. The second is a pre-processed version produced by the authors’ own pipeline. Every event carries a descriptive label encoding the task, the modality, the concept, the stimulus variant and the exemplar identity. For instance, a label may refer to the imagination of a high-pitched spoken version of the word *flower*, or to the perception of a naturalistic image of a flower. This labelling allows trials to be selected flexibly according to the requirements of each classification task.

3.2 Pipeline 1

Pipeline1 is the novel handcrafted approach whose rationale was set out in Section 1.4. Its design responds to two concerns. The first is the distributed nature of semantic representation: conceptual content is encoded across widely distributed cortical networks, a spatial organisation to which a whole-scalp regional decomposition is naturally suited. The second is interpretability: instead of feeding the signal into a black-box classifier, the pipeline summarises each scalp region by two scalar quantities, namely an energy and a decorrelation measure, and combines them through a parameter-free decision rule with no learned parameters beyond the CSP spatial filters themselves. The pipeline combines an anatomical regional partition of the electrode montage with region-wise CSP filtering, extracting from each region a regional energy and a decorrelation measure, which are then aggregated into a class decision.

Common Spatial Patterns. Common Spatial Patterns (CSP) is a supervised spatial-filtering technique that, given two classes, derives linear combinations of the recording channels whose variance is maximal for one class and minimal for the other [35, 36]. For a multichannel segment $\mathbf{X}$ (channels $\times$ time), the trace-normalised spatial covariance is $\mathbf{X}\mathbf{X}^\top / \text{tr}(\mathbf{X}\mathbf{X}^\top)$; averaging it over the trials of each class produces the two class covariances $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$, whose sum is the composite covariance:

$$\mathbf{\Sigma} = \mathbf{\Sigma}_1 + \mathbf{\Sigma}_2. \quad (3.1)$$

The composite is whitened through its eigendecomposition $\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$, which defines the whitening matrix $\mathbf{P} = \mathbf{\Lambda}^{-1/2}\mathbf{U}^\top$ such that $\mathbf{P}\mathbf{\Sigma}\mathbf{P}^\top = \mathbf{I}$; this requires $\mathbf{\Lambda} \succ \mathbf{0}$, i.e. $\mathbf{\Sigma}$ must be full rank, a condition ensured in practice by regularisation or dimensionality reduction when necessary. Applying $\mathbf{P}$ to the two class covariances:

$$\mathbf{S}_1 = \mathbf{P}\mathbf{\Sigma}_1\mathbf{P}^\top, \quad \mathbf{S}_2 = \mathbf{P}\mathbf{\Sigma}_2\mathbf{P}^\top, \quad (3.2)$$

yields two matrices that, since $\mathbf{S}_1 + \mathbf{S}_2 = \mathbf{I}$, share the same eigenvectors and have eigenvalues summing to one. With $\mathbf{S}_1$ symmetric, $\mathbf{B}$ is orthogonal; diagonalising $\mathbf{S}_1 = \mathbf{B}\mathbf{\Psi}\mathbf{B}^\top$ therefore diagonalises $\mathbf{S}_2 = \mathbf{B}(\mathbf{I} - \mathbf{\Psi})\mathbf{B}^\top$ as well. Along each common direction the variance of the first class is ψ and that of the second $1 - \psi$: eigenvectors with ψ near one are maximally discriminative for the first class and those with ψ near zero for the second. The columns of $\mathbf{B}$ are ordered by decreasing ψ , so that the most discriminative directions for each class appear at the two extremes. The spatial filters are the rows of:

$$\mathbf{W} = \mathbf{B}^\top \mathbf{P}. \quad (3.3)$$

Projecting the segment onto them yields $\mathbf{Z} = \mathbf{W}\mathbf{X}$, which concentrates the class-discriminative variance in the first and last filtered components.

Notation and regional decomposition. A single trial is represented as a preprocessed multichannel epoch $\mathbf{X} \in \mathbb{R}^{C \times T}$, where $C = 124$ is the number of scalp electrodes and T the number of time samples; the length depends on the task and modality, the analysis window ranging from two to four seconds according to Table 3.1. Of the 124 scalp electrodes, the four midline electrodes (Fz , FCz , Cz , Pz) are excluded, as they lie on the central line between the lateralised left and right regions and cannot be assigned to either side; the remaining 120 are partitioned into $R = 10$ anatomical regions of interest, each of cardinality $C_r = 12$, so that $R \times C_r = 120$. The two bilateral regions, frontal-polar and occipital, are not lateralised and therefore retain the midline electrodes that fall within them. The regions span the principal functional areas implicated in semantic processing: frontal-polar, frontal-temporal, frontal-central, parieto-central, parieto-temporal, and occipital/posterior-parietal. The complete electrode-to-region mapping is reported in Table 3.2. Each trial is thus reorganised from $\mathbf{X} \in \mathbb{R}^{124 \times T}$ into $\mathbf{X}' \in \mathbb{R}^{10 \times 12 \times T}$.

Region-wise Common Spatial Patterns. This procedure, Common Spatial Patterns, is applied independently within each of the $R = 10$ regions. Since CSP is inherently binary, the three-class task is reduced to three binary problems in a One-vs-Rest scheme: for every class $c \in \{1, \dots, K\}$, with $K = 3$, and every region r , the covariance Σ_1 is estimated from the trials of class c and Σ_2 from those of the remaining classes pooled together, and a filter bank $\mathbf{W}_c^{(r)} \in \mathbb{R}^{m \times C_r}$ is obtained, $m = 2$ filters being retained and $C_r = 12$ channels per region. The two retained filters are the two extremes of the CSP ordering, one maximising the variance of class c and the other that of the remaining classes. Projecting a regional epoch onto these filters gives:

$$\mathbf{Z}_c^{(r)} = \mathbf{W}_c^{(r)} \mathbf{X}^{(r)} \in \mathbb{R}^{2 \times T}, \quad (3.4)$$

whose rows $\mathbf{z}_{c,1}^{(r)}$ and $\mathbf{z}_{c,2}^{(r)}$ are the two CSP components; by the ordering convention of the decomposition, the first, $\mathbf{z}_{c,1}^{(r)}$, is the component of maximal variance for class c and thus identifies the spatial direction along which class c is most strongly expressed within region r , while the second, $\mathbf{z}_{c,2}^{(r)}$, is the direction of maximal variance for the remaining classes.

Regional attention as a decorrelation measure. For each region and class, the first feature measures how distinctly the two retained components separate. The construction draws on the inner-product similarity at the core of the attention mechanism, introduced by Bahdanau et al.[92] and later central to Transformer architectures[93], where the inner product between two representations scores their similarity and weights their contribution. The regional attention transfers this idea to the two retained CSP components, each treated as a vector in $\mathbb{R}^T$, and is defined

Table 3.2: Partition of the 120 retained scalp electrodes into ten anatomical regions of interest (ROIs). Each ROI contains exactly 12 electrodes ($C_r = 12 \forall r$).

ROI	Region	Electrodes
1	Left Frontal-Temporal	F5, F7, FT7, FT9, FTT7h, FTT9h, FC5, T7, FFT7h, C5, AF7, M1
2	Frontal Polar-Prefrontal	Fp1, Fpz, Fp2, AFp3h, AFp4h, AF3, AFz, AF4, AFF1, AFF2, AFF5h, AFF6h
3	Right Frontal-Temporal	F6, F8, FT8, FT10, FTT8h, FTT10h, FC6, T8, FFT8h, C6, AF8, M2
4	Left Frontal-Central	F3, F1, FC1, FC3, FFC1h, FFC3h, FFC5h, FCC1h, FCC3h, FCC5h, C1, C3
5	Right Frontal-Central	F4, F2, FC2, FC4, FFC2h, FFC4h, FFC6h, FCC2h, FCC4h, FCC6h, C2, C4
6	Left Parieto-Central	P1, P3, P5, CP1, CP3, CP5, CCP1h, CCP3h, CCP5h, CPP1h, CPP3h, CPP5h
7	Occipital-Posterior Parietal	O1, O2, OI1h, OI2h, POz, PO3, PO4, PPO1, PPO2, POO3h, POO4h, Iz
8	Right Parieto-Central	P4, P2, P6, CP2, CP4, CP6, CCP2h, CCP4h, CCP6h, CPP2h, CPP4h, CPP6h
9	Left Parieto-Temporal	P7, P9, PO7, PO9, PPO5h, PPO9h, POO9h, TP7, TPP7h, TPP9h, TTP7h, I1
10	Right Parieto-Temporal	P8, P10, PO8, PO10, PPO6h, PPO10h, POO10h, TP8, TPP8h, TPP10h, TTP8h, I2

as the complement of their absolute cosine similarity:

$$\alpha_c^{(r)} = 1 - \frac{|\langle \mathbf{z}_{c,1}^{(r)}, \mathbf{z}_{c,2}^{(r)} \rangle|}{\|\mathbf{z}_{c,1}^{(r)}\| \|\mathbf{z}_{c,2}^{(r)}\|} \in [0,1]. \quad (3.5)$$

Since the z-scoring applied during preprocessing centres each channel to zero mean, and any linear combination of zero-mean channels is itself zero-mean, the two CSP components have zero temporal mean; the cosine similarity then coincides with the Pearson correlation between the two time courses:

$$\rho(\mathbf{a}, \mathbf{b}) = \frac{\sum_t (a_t - \bar{a})(b_t - \bar{b})}{\sqrt{\sum_t (a_t - \bar{a})^2} \cdot \sqrt{\sum_t (b_t - \bar{b})^2}} \quad (\text{Pearson correlation}) \quad (3.6)$$

$$\rho(\mathbf{a}, \mathbf{b}) = \frac{\sum_t a_t \cdot b_t}{\sqrt{\sum_t a_t^2} \cdot \sqrt{\sum_t b_t^2}} = \frac{\langle \mathbf{a}, \mathbf{b} \rangle}{\|\mathbf{a}\| \cdot \|\mathbf{b}\|} = \cos \theta \quad (\text{if } \bar{a} = \bar{b} = 0) \quad (3.7)$$

so that $\alpha_c^{(r)}$ is equivalently a decorrelation index:

$$\alpha_c^{(r)} = 1 - |\cos \theta| = 1 - \left| \rho(\mathbf{z}_{c,1}^{(r)}, \mathbf{z}_{c,2}^{(r)}) \right| \quad (3.8)$$

$\alpha_c^{(r)}$ is close to unity when the two CSP components are orthogonal, meaning the decomposition found two well-separated directions, and close to zero when they are collinear, meaning the decomposition is unreliable in that region; the hypothesis is that regions where the decomposition is more reliable are also more informative for classification. Evaluated over all classes and regions, $\alpha_c^{(r)}$ constitutes, for each trial, the attention map $\boldsymbol{\alpha} \in \mathbb{R}^{K \times R}$.

Regional energy. The second feature is the variance of the first CSP component, computed directly rather than in its logarithmic form as in the standard CSP pipeline:

$$\varepsilon_c^{(r)} = \text{Var}(\mathbf{z}_{c,1}^{(r)}) = \frac{1}{T-1} \sum_{t=1}^T \left(z_{c,1}^{(r)}(t) - \bar{z}_{c,1}^{(r)} \right)^2. \quad (3.9)$$

Because the first filter is constructed so as to maximise the variance of class c , a trial genuinely belonging to that class is expected to produce a large $\varepsilon_c^{(r)}$ along the corresponding direction, whereas trials of the remaining classes should excite it comparatively little. The energies, computed over all classes and regions, form the per-trial energy map $\boldsymbol{\varepsilon} \in \mathbb{R}^{K \times R}$. Since the Decision rules rely on a simple sum of regional energies, and the combined rule multiplies each energy by the attention $\alpha_c^{(r)}$, the choice of the energy representation directly affects the aggregation. The log-variance, commonly adopted in the standard CSP pipeline, is negative whenever the variance is below unity and diverges towards $-\infty$ as the variance approaches zero; a single region with low energy can therefore dominate the sum and override the contribution of all the others. This is particularly problematic in the combined rule, where the multiplicative weighting $\alpha_c^{(r)} \varepsilon_c^{(r)}$ is only meaningful for non-negative energies: a negative energy in a reliable region (high attention) yields a large negative contribution, inverting the intended role of the weighting. The raw variance, being non-negative, avoids both issues and is therefore adopted here.

Decision rules. Since both the attention map $\boldsymbol{\alpha}$ and the energy map $\boldsymbol{\varepsilon}$ reduce each epoch to a single scalar per region and class, the variable length T of the analysis window across conditions introduces no dimensional inhomogeneity into the representation. Classification proceeds by aggregating the regional maps into a per-class score and selecting the maximal-scoring class, with no intervening

learned parameter. Unlike the conventional CSP pipeline, in which the component log-variances are submitted to a supervised classifier, here the decision relies solely on a parameter-free aggregation of the regional maps. For the two three-class tasks (sensory modality and semantic concept), three rules are considered. The first, based on the attention map, assigns the trial to the class with the highest attention summed over all regions, under the hypothesis that the true class produces the most reliable, most decorrelated regional decompositions.

$$\hat{y}_\alpha = \arg \max_{c \in \{1, \dots, K\}} \sum_{r=1}^R \alpha_c^{(r)}. \quad (3.10)$$

Second, the energy rule returns the class whose class-specific direction carries the greatest cumulative power.

$$\hat{y}_\varepsilon = \arg \max_{c \in \{1, \dots, K\}} \sum_{r=1}^R \varepsilon_c^{(r)}. \quad (3.11)$$

Third, the combined rule weights each regional energy by its attention before pooling, so that regions exhibiting a more reliable decomposition contribute more strongly to the energy-based vote.

$$\hat{y}_{\alpha\varepsilon} = \arg \max_{c \in \{1, \dots, K\}} \sum_{r=1}^R \alpha_c^{(r)} \varepsilon_c^{(r)}. \quad (3.12)$$

In this last formulation the attention operates as a per-region, per-trial reliability gate modulating the discriminative energy, the two features being combined multiplicatively rather than treated as independent votes. None of the three rules introduces a free parameter or a trained decision function; the regional CSP filter banks $\{\mathbf{W}_c^{(r)}\}$ are the only quantities learned from data.

Binary case: perception versus imagination. The One-vs-Rest construction above is required only for the two three-class tasks. For the binary perception–imagination distinction a single CSP per region suffices, since its two retained components are tied directly to the two classes without recourse to separate decompositions: by construction the first, $\mathbf{z}_1^{(r)}$, maximises the variance of imagination and the second, $\mathbf{z}_2^{(r)}$, that of perception. The two components therefore yield two class-specific regional energies:

$$\varepsilon_I^{(r)} = \text{Var}(\mathbf{z}_1^{(r)}), \quad \varepsilon_P^{(r)} = \text{Var}(\mathbf{z}_2^{(r)}), \quad (3.13)$$

and the decision is taken on their signed contrast rather than by an $\arg \max$. Aggregating the per-region energy difference across the scalp:

$$\Delta\varepsilon = \sum_{r=1}^R (\varepsilon_I^{(r)} - \varepsilon_P^{(r)}), \quad (3.14)$$

the trial is assigned to imagination when $\Delta\varepsilon \geq 0$ and to perception otherwise, the sign of $\Delta\varepsilon$ indicating which class carries the greater cumulative power. The attention here reduces to a single per-region scalar, $\alpha^{(r)} = 1 - |\cos(\mathbf{z}_1^{(r)}, \mathbf{z}_2^{(r)})|$, common to both classes, and is incorporated by weighting each regional contrast before summation:

$$\Delta\varepsilon_\alpha = \sum_{r=1}^R \alpha^{(r)} (\varepsilon_I^{(r)} - \varepsilon_P^{(r)}), \quad (3.15)$$

with the same sign-based assignment. Both binary rules fix the decision threshold at zero rather than tuning it, so that they too remain free of any learned parameter beyond the CSP filters; the attention-only rule has no binary counterpart, as the attention is symmetric with respect to the two classes and therefore carries no discriminative information on its own.

Training and evaluation. The preprocessing pipeline of Wilson et al.[18] is adopted without modification; in particular, no task-specific band-pass is applied, so that the CSP operates on the full preprocessed band (high-pass filtered at 2 Hz, with power-line notches at 50, 100 and 150 Hz). Standard CSP practice restricts the signal to the narrow band carrying the discriminative rhythm, since the variance-driven spatial filters are otherwise dominated by the highest-power components; this band is well established for sensorimotor paradigms (the mu and beta rhythms) but not for multimodal semantic perception and imagination, whose spectral signature is not yet consolidated. Fixing a band a priori would risk discarding discriminative information, so the full band is retained here as a deliberately assumption-light baseline, with data-driven spectral selection (for instance a filter-bank extension) left as a natural refinement. Prior to feature extraction, each epoch is z-scored along the time axis independently for each channel, so that all channels have zero mean and unit variance. The regional CSP filter banks are then estimated only on the training partition (the $S - 1$ subjects retained at each Leave-One-Subject-Out fold) and are then applied, unaltered, to the epochs of the held-out subject. As the attention, the energy and the three decision rules are entirely parameter-free, no additional quantity is fitted: the spatial filters constitute the sole learned component, and the pipeline generalises to an unseen subject through interpretable spatial projections alone.

3.3 Pipeline 2

Pipeline 2 constitutes the machine-learning comparator of this study. Where Pipeline 1 decomposes the scalp into anatomical regions and derives handcrafted scalar descriptors, Pipeline 2 applies no explicit spatial filtering; instead it characterises each trial through the spatial covariance matrix of its multichannel signal.

These covariances live on the curved manifold of symmetric positive-definite (SPD) matrices rather than in a flat Euclidean space. Covariances are estimated independently within each of the five physiological frequency bands. Since distinct cognitive and perceptual processes are reflected in different portions of the EEG spectrum (Section 2.2.3), computing a separate covariance matrix per band allows the pipeline to capture the spatial correlation structure among channels that is specific to that band; concatenating the five resulting representations then provides the classifier with a joint spectral and spatial characterisation of each trial. Each covariance is mapped onto the tangent space at the Riemannian mean of the training set, which serves as the reference point for the alignment, producing a Euclidean feature vector suitable for conventional classifiers. The pipeline addresses RQ2, namely whether Riemannian covariance features can sustain classification performance across the three classification tasks (perception versus imagination, sensory modality discrimination, and semantic concept recognition) under a strict subject-independent regime, while also serving as a comparator for the novel handcrafted approach of Pipeline 1.

Notation and band decomposition. A single trial is a preprocessed multi-channel epoch $X \in \mathbb{R}^{C \times T}$, with $C = 124$ scalp electrodes and T time samples (see Table 3.1). Unlike Pipeline 1, no electrode is discarded. The epoch is first decomposed into five band-limited versions by zero-phase band-pass filtering. For each band $b \in \{\delta, \theta, \alpha, \beta, \gamma\}$, with passband as defined in Section 3.5.5, the signal is filtered to obtain $X_b \in \mathbb{R}^{C \times T}$. A zero-phase, forward-backward filter is used: the prototype filter is applied to the signal in both temporal directions so that the two phase responses cancel and the band-limited signal is left free of phase distortion, preserving the temporal structure of the waveform. Each trial is thereby re-expressed as the collection $\{X_b\}_{b=1}^B$ of $B = 5$ band-limited epochs. It should be noted that, because the preprocessing applies a high-pass at 2 Hz (Section 3.5.2), the delta passband is in practice confined to roughly 2–4 Hz, its portion below 2 Hz having been removed. This is a deliberate trade-off: the 2 Hz cut-off affords a more effective suppression of low-frequency drifts and a better-conditioned independent component decomposition, while the information lost is of limited relevance to the present task, delta being the slowest cortical rhythm and associated chiefly with deep slow-wave sleep rather than with waking cognition (Section 2.2.3).

Spatial covariance estimation. Within each band the trial is summarised by its $C \times C$ spatial covariance matrix. Rather than the plain sample covariance, the Ledoit-Wolf shrinkage estimator is adopted [94], which regularises the sample covariance $S_b = \frac{1}{T-1} X_b X_b^\top$ toward a scaled identity matrix:

$$\hat{\Sigma}_b = (1 - \rho) S_b + \rho \mu_b I, \quad (3.16)$$

where $\mu_b = \text{tr}(S_b)/C$ is the mean eigenvalue of S_b and $\rho \in [0,1]$ is the shrinkage coefficient estimated analytically from the data[94]. This regularisation pulls the eigenvalue spectrum of S_b toward a uniform distribution, improving the numerical conditioning of the estimate and guaranteeing symmetric positive-definiteness regardless of the ratio T/C , as required by the matrix logarithm in (3.19).

Tangent-space mapping. Since the SPD manifold $\mathcal{M}$ is curved, Euclidean distances do not respect the geometry of the manifold; geodesic distances (see Figure 3.3), measured along the manifold itself, must be used instead to define a geometrically consistent reference point. Once that reference is established, each covariance is projected onto the flat tangent plane erected at it, yielding a Euclidean vector on which ordinary classifiers can operate[85, 86]. The metric adopted for measuring geodesic distances is the affine-invariant Riemannian metric, so named because it satisfies $\delta_R(APA^\top, AQA^\top) = \delta_R(P, Q)$ for any invertible matrix A : the distance between two covariances is unchanged by any invertible linear transformation applied to both, a property that reflects the fact that EEG signals are themselves a linear mixture of underlying neural sources. Under this metric the geodesic distance between two SPD matrices P and Q is:

$$\delta_R(P, Q) = \left\| \log(P^{-1/2}QP^{-1/2}) \right\|_F = \left(\sum_{i=1}^C \log^2 \lambda_i \right)^{1/2}, \quad (3.17)$$

with $\log(\cdot)$ the matrix logarithm and λ_i the eigenvalues of $P^{-1}Q$. This metric defines a geometrically consistent notion of centre of mass on $\mathcal{M}$: the Riemannian (Fréchet) mean of the band-specific training covariances is the SPD matrix $\bar{\Sigma}_b$ that minimises the total squared geodesic distance to the training set, where the minimisation ranges over all candidate matrices G on the manifold:

$$\bar{\Sigma}_b = \arg \min_{G \in \mathcal{M}} \sum_{n=1}^{N_{\text{tr}}} \delta_R^2(G, \Sigma_b^{(n)}), \quad (3.18)$$

which has no closed form and is computed iteratively[85]. Each covariance is then whitened by this reference and projected onto the tangent plane at $\bar{\Sigma}_b$ through the matrix logarithm, the resulting symmetric matrix being vectorised by retaining only its upper-triangular entries:

$$s_b^{(n)} = \text{upper} \left(\log(\bar{\Sigma}_b^{-1/2} \Sigma_b^{(n)} \bar{\Sigma}_b^{-1/2}) \right) \in \mathbb{R}^{C(C+1)/2}, \quad (3.19)$$

where $\text{upper}(\cdot)$ collects the $C(C+1)/2$ upper-triangular coefficients of its argument, the off-diagonal ones weighted by $\sqrt{2}$ so that the Euclidean norm of $s_b^{(n)}$ equals the Riemannian norm of the corresponding tangent vector. The mapping linearises the

manifold locally around the training mean, so that distances and inner products between nearby covariances are well approximated by ordinary vector operations on their tangent-space coordinates. For $C = 124$ each tangent vector has dimension $C(C + 1)/2 = 7750$. The $B = 5$ band-specific tangent vectors are concatenated into the final per-trial representation of dimension $5 \times 7750 = 38\,750$.

$$f^{(n)} = \left[s_{\delta}^{(n)} ; s_{\theta}^{(n)} ; s_{\alpha}^{(n)} ; s_{\beta}^{(n)} ; s_{\gamma}^{(n)} \right] \in \mathbb{R}^{B \cdot C(C+1)/2}, \quad (3.20)$$

A separate tangent space, with its own reference mean $\bar{\Sigma}_b$, is built per band, the coordinates being stacked rather than mixed across bands.

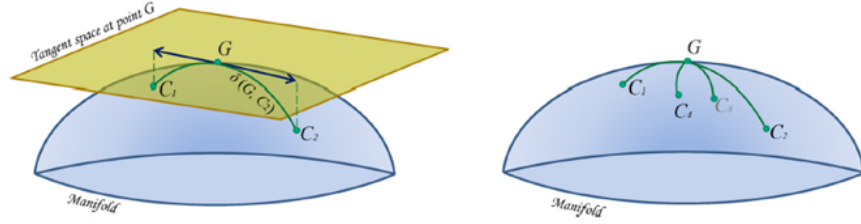


Figure 3.3: Geometric interpretation of the Riemannian manifold. Each EEG trial covariance matrix is represented as a point on the manifold. *Left:* the tangent plane erected at the reference point G . The shortest path connecting two points C_1 and C_2 along the manifold surface is called the geodesic; its length defines the Riemannian distance $\delta_R(C_1, C_2)$ of (3.17). Geodesics passing through G are mapped onto straight lines in the tangent plane, so that the local geometry of the manifold is well approximated by ordinary Euclidean operations in a neighbourhood of G . *Right:* G is the Riemannian (Fréchet) mean of the four points C_1, C_2, C_3, C_4 , defined as the manifold point minimising the total sum of squared geodesic distances to all four points, as formalised in (3.18); it serves as the geometrically consistent reference around which the tangent plane is constructed. Adapted from[9].

Classification. The 38 750-dimensional tangent vectors are submitted to two standard supervised classifiers, trained and evaluated separately. The first is a linear support vector machine (SVM), which fits the maximum-margin separating hyperplane under a soft-margin (hinge-loss) objective. The second is a random forest, an ensemble of decision trees each grown on a bootstrap resample of the training trials and on a random subset of features at every split, the prediction being taken by majority vote; this non-linear estimator captures higher-order interactions among the tangent coordinates and tolerates their high dimensionality and redundancy comparatively well[84]. The binary perception–imagination task is handled by both classifiers directly; for the three-class sensory-modality and

semantic-concept tasks the random forest is inherently multiclass, while the SVM is extended through a one-vs-one decomposition. The resulting 38 750-dimensional representation is high-dimensional relative to the number of available training trials; this disproportion may have adversely affected classification performance, and dimensionality reduction techniques such as PCA on the tangent vectors were not explored in this work, representing a natural direction for future investigation.

Training and evaluation. In keeping with the subject-independent protocol adopted throughout, every quantity estimated from data (the per-band Riemannian reference means $\bar{\Sigma}_b$, any feature-scaling statistics, and the classifier parameters) is fitted exclusively on the training partition of each Leave-One-Subject-Out fold, that is the $S - 1$ retained subjects, and is then applied unchanged to the held-out subject. In particular, each band’s tangent-space reference is computed from the training covariances alone and reused to project those of the test subject, so that no information from the held-out individual can enter either the geometric mean or the decision rule; the resulting accuracy thus measures genuine cross-subject generalisation.

3.4 Pipeline 3

Pipeline 3 constitutes the deep-learning comparator of this study. It adapts the topographic decoding framework of Ahmadi and Mesin[28], in which each EEG epoch is re-expressed as a sequence of scalp-voltage maps, or *topomaps*, and submitted to a Convolutional Neural Network (CNN). Whereas Pipeline 1 provides an interpretable, spatially grounded representation and Pipeline 2 an established covariance based one, Pipeline 3 contributes an end-to-end learned alternative that operates directly on the spatial topography of the signal and requires no handcrafted feature-extraction stage. Pipeline 3 thus directly addresses RQ3, which asks whether topographic decoding, already shown to separate pictorial perception from imagination with high accuracy[28], generalises to the remaining classification tasks (sensory modality discrimination, semantic concept recognition, and the perception–imagination distinction in the auditory and orthographic modalities). The present implementation, however, is not a reproduction of the original work[28]: it differs from theirs both in the generation of the topographic maps and in the convolutional architecture used to classify them. These differences are deliberate. The aim in adopting this pipeline is not to replicate their setup or to benchmark the present results against those they report, but to provide a complementary realisation of the topographic approach within the subject-independent framework of this work, evaluated alongside Pipelines 1 and 2 so as to give a more complete account of how the three method families behave across the classification tasks

considered.

Representative epoch. For each subject and each class c of the task under consideration, the N_c standardised epochs assigned to that class (Section Preprocessing) are averaged sample-by-sample into a single representative epoch

$$\bar{X}_c = \frac{1}{N_c} \sum_{i=1}^{N_c} X_c^{(i)} \in \mathbb{R}^{C \times T}, \quad (3.21)$$

where $C = 124$ is the number of scalp electrodes, $X_c^{(i)}$ the i -th standardised epoch of class c , and T the number of time samples of the analysis window, which depends on the task and modality and ranges from two to four seconds (Table 3.1); the topographic maps are subsequently derived from $\bar{X}_c$ rather than from the individual trials. Working from an averaged epoch suits the topographic representation, whose scalp-voltage maps are most informative when computed from a stable signal rather than from a single noisy trial: in this respect it follows the same logic as the Event-Related Potential (Section 2.2.2), the averaging attenuating the activity uncorrelated with stimulus onset while preserving the stimulus-locked response. It is also expedient in practice, since each epoch is expanded into a dense sequence of topographic images and rendering and training on the maps of every individual trial would be prohibitively costly (given the computational resources used in this work), whereas a single representative epoch per class keeps the number of generated frames tractable.

Topographic map generation. Each representative epoch is converted into a temporally ordered sequence of topographic maps. At successive intervals of 10 ms along the epoch, the instantaneous multichannel voltage is interpolated from the 124 electrode positions onto a regular two-dimensional grid spanning the scalp. Each map is rendered as an RGB image in which a diverging colour scale encodes the polarity and magnitude of the scalp potential. Whereas the original work renders these maps at a resolution of $150 \times 150 \times 3$, the present implementation reduces them to $32 \times 32 \times 3$, again for the computational reasons. The sampling step is held fixed at 10 ms across all tasks and modalities, a further departure from the original work, where the interval was adapted to each condition (15 ms for pictorial perception and 20 ms for pictorial imagination) so that every epoch yielded exactly two hundred maps and the two classes were balanced[28]. A uniform step preserves a common temporal resolution throughout, but makes the number of frames depend on the epoch duration: four hundred for a four-second imagination epoch and three hundred for a three-second visual perception epoch, for example (Table 3.1). The variability therefore affects the number of frames per epoch and not the size of the images, which remain $32 \times 32 \times 3$ regardless of condition. Every

frame inherits the class label of the epoch from which it originates and constitutes a single classification instance, in keeping with the frame-as-sample paradigm of [28].

Convolutional neural network. The topomap frames are classified by a compact convolutional network. It is deliberately simpler than the squeeze-and-excitation (SE) architecture of the original work[28], dispensing with its SE blocks and pooling layers in favour of depthwise and strided convolutions. Applied to the $32 \times 32 \times 3$ input, the network is a short stack: a standard convolution (2×2 kernel, stride two, no bias), two depthwise convolutions (one filter per channel; the first strided, the second at unit stride), a flattening step, and a fully connected ReLU layer feeding the task-specific output. Each convolution is followed by batch normalisation and dropout, and the dense layer by dropout; the two strided convolutions provide all spatial downsampling, with no pooling. This recurring regularisation limits overfitting in the small-sample EEG setting (Section 2.5). The capacity of the network is matched to the task. For the three-way classifications (sensory modality and semantic concept), the initial convolution uses 128 filters, the fully connected layer comprises 150 units, and the output is a three-unit softmax. For the binary perception–imagination distinction the structure is identical but lighter, the initial convolution using 64 filters and the fully connected layer 100 units, with a single sigmoid unit at the output; this reduced capacity reflects the lower complexity of a two-class decision.

Training and evaluation. In keeping with the subject-independent protocol adopted throughout this work, the network is trained and evaluated under Leave-One-Subject-Out (LOSO) cross-validation: at each fold the frames of a single subject are withheld for testing while those of the remaining subjects constitute the training set, the procedure being repeated until every subject has served once as the held-out test case. The training frames are randomly shuffled before batching, so that each mini-batch intermixes frames originating from different subjects, classes, and time instants; this prevents the optimisation from being influenced by the temporal ordering of the maps within an epoch or by the grouping of frames by subject. Performance is reported using the metrics defined in Section 3.5.3. The remaining training hyperparameters (learning rate, batch size, and regularisation) are reported in Section 3.5.5.

3.5 Experimentation Setups

The experimental work of this thesis addresses the three classification problems introduced in Section 1.4: the discrimination of perception from imagination, the identification of the sensory modality through which a stimulus is presented, and the recognition of the semantic concept. Each problem is evaluated with all three pipelines (the handcrafted CSP-based Pipeline 1, the Riemannian-geometric Pipeline 2 and the topographic CNN Pipeline 3) under the Leave-One-Subject-Out protocol motivated in Section 1.3. Each problem is decomposed into condition-specific tasks:

- **Perception versus imagination:** three binary tasks, one within each sensory modality (auditory, orthographic, pictorial);
- **Sensory modality:** (auditory versus orthographic versus pictorial); two three-class tasks, one within perception and one within imagination;
- **Semantic concept:** (flower versus penguin versus guitar); six three-class tasks, one within each task–modality condition.

The experimental setup adopted to address these tasks is described in what follows.

Cross-validation protocol. All three pipelines are evaluated under a leave-one-out cross-validation applied at the level of the individual recording session: each of the fifteen sessions is held out in turn as the test set while the remaining sessions form the training set, and every quantity estimated from data is fitted on the training partition alone. For the nine participants who completed a single session this coincides exactly with a leave-one-subject-out scheme; for the three who completed two, one session may remain in the training set while the other is held out, so that strict subject-independence holds for nine of the twelve participants. Each session is therefore treated as an independent unit, following the only prior work, to evaluate this dataset under a leave-one-subject-out regime [28, 29]; the choice is stated explicitly here, and its practical impact is assessed empirically in Chapter 4.

3.5.1 Tools

Hardware. All experiments were carried out on a single Dell Inspiron 7580 laptop, equipped with an Intel Core i7-8565U processor (four cores and eight threads at 1.80 GHz), an NVIDIA GeForce MX150 graphics card with 2 GB of dedicated memory, and an integrated Intel UHD Graphics 620 unit. This is a modest computational platform, and its constraints (the limited GPU memory in particular) motivated several of the implementation choices of Pipeline 3, namely

the reduced resolution of the topographic maps and the use of a single representative epoch per class.

Software. The entire study was implemented in Python (v3.12.3). The EEG recordings were handled with MNE-Python (v1.9.0), which was used to load the BIDS-formatted data, to apply the preprocessing pipeline and to extract the epochs. Numerical computation and linear algebra across the three pipelines relied on NumPy (v1.26.4) and SciPy (v1.14.0), the latter also providing the standardisation routine of Section 3.5.2. The covariance estimation, Riemannian averaging and tangent-space projection of Pipeline 2 were performed with PyRiemann (v0.9), and its support-vector-machine and random-forest classifiers were taken from scikit-learn (v1.6.1); the convolutional network of Pipeline 3 was built and trained with TensorFlow (v2.17.0). All code was developed in Visual Studio Code (v1.124.2).

3.5.2 Preprocessing

All three pipelines operate on an identically preprocessed signal: the preprocessing procedure released by Wilson et al.[18] is adopted without modification, both to preserve comparability with the original technical validation and because the dataset is distributed together with an author-provided preprocessed version generated by exactly this procedure. The pipeline is implemented within the MNE-Python ecosystem and is applied to the continuous raw recording of each participant and session, before any epoching or feature extraction. It comprises five stages: bad-channel handling, re-referencing, filtering, artefact removal by independent component analysis, and epoching.

Bad-channel detection and interpolation. Defective channels were identified by combining manual and automatic inspection. Channels judged faulty through visual examination of the raw traces were logged for each participant and session, and were complemented by the automatic detection of the PyPrep `PrepPipeline`, which flags channels that fail to correlate with the remaining electrodes, that exhibit abnormally low or high amplitude, that carry an excessive amount of high-frequency noise, or that return a flat signal. The channels so identified were reconstructed by spherical-spline interpolation from the surrounding electrodes, restoring the complete 124-channel montage.

Re-referencing. The signals were acquired with respect to CPz. Because re-referencing modifies the statistical trend of the whole recording, it was recomputed at several points of the pipeline rather than once. A common-average reference was applied both before and after bad-channel interpolation, so that the interpolation operated on a consistently referenced montage, and a third common-average step

was performed after filtering in order to suppress residual low-frequency drift. Common-average referencing expresses each electrode relative to the mean across all electrodes, an appropriate choice for the high-density montage employed here.

Filtering. Two filtering operations were carried out. A notch filter removed the 50 Hz power-line interference characteristic of the United Kingdom mains, together with its harmonics at 100 and 150 Hz. A high-pass filter with a cut-off at 2 Hz then attenuated the slow drifts produced by head movement, scalp perspiration and cable motion; this cut-off is also the value recommended for obtaining well-conditioned independent component decompositions. No low-pass filter was applied, leaving the upper portion of the spectrum intact, so that the high-frequency activity later exploited by the band-limited representation of Pipeline 2 is preserved.

Artefact removal. Ocular and muscular artefacts were suppressed by independent component analysis. The FastICA algorithm was fitted to the continuous, filtered recording rather than to the epoched data, fifty components being extracted. Eye-related components were located automatically by relating them to electrooculographic events estimated from the frontopolar electrodes Fp1 and Fp2, and the offending components were projected out during reconstruction. The signal recovered after this rejection retains all 124 original channels, so that the full montage is available to every downstream pipeline.

Epoching. The continuous recording was finally segmented into single-trial epochs by means of the event labels marking the onset of each stimulus. Since the analysis window differs across conditions, each epoch was cut to the duration prescribed for its task and modality, ranging from two seconds for auditory perception to four seconds for every imagination condition, in accordance with Table 3.1; only the intervals corresponding to the perception and imagination phases were retained. These epochs constitute the common input from which Pipelines 1, 2 and 3 derive their respective representations.

Channel-wise standardisation. A further standardisation step, common to the three pipelines, is then applied to these epochs before any pipeline-specific processing. Each epoch $\mathbf{X} \in \mathbb{R}^{C \times T}$ is standardised by z-scoring along its temporal axis, so that every channel of every trial is centred to zero mean and scaled to unit variance across the T samples of its analysis window, independently of all other channels and trials; with the epochs stacked into a single array of shape $N \times C \times T$, this corresponds to reducing along the last axis. The operation is performed with the `scipy.stats.zscore` function of the SciPy library. Equalising the per-channel scale in this manner prevents a small number of high-amplitude

electrodes from dominating the representations built by the three pipelines, and it is this step that secures the zero temporal mean assumed in the derivation of the regional decorrelation measure of Pipeline 1. The resulting standardised epochs are the actual input to the feature-extraction stages of Pipelines 1, 2 and 3.

3.5.3 Evaluation metrics

All three pipelines are assessed with the same metrics, so that their results can be compared on a common footing. The primary one is the balanced accuracy, the mean of the per-class recalls,

$$\text{BA} = \frac{1}{K} \sum_{c=1}^K \frac{\text{TP}_c}{\text{TP}_c + \text{FN}_c}, \quad (3.22)$$

where K is the number of classes and TP_c and FN_c are the true positives and false negatives of class c . By weighting every class equally regardless of its frequency, the balanced accuracy is insensitive to class imbalance and cannot be inflated by a bias towards the majority class. This property is essential for Pipeline 3, whose frame-level dataset is unbalanced across conditions of different duration (Section 3.4); adopting the same metric for Pipelines 1 and 2, where it coincides with the ordinary accuracy whenever the classes are balanced, ensures that all three pipelines are judged by an identical criterion and remain directly comparable. The structure of the errors is examined further through the confusion matrix of each task and through the per-class precision, recall and F1-score, which reveal whether performance is uniform across the classes or concentrated on a subset of them. For the convolutional network of Pipeline 3, the training and validation loss and accuracy are additionally tracked across epochs and reported as curves, in order to monitor convergence and detect overfitting. For visualisation purposes the loss and accuracy were logged over the full 100 epochs and plotted on the same range for every task, giving all curves the same epoch axis and keeping the gradual onset of overfitting visible. This logging is decoupled from the early-stopping criterion, which governs the weights actually used at test time: in most folds the best validation performance, and hence the restored weights, was reached within the first 30–40 epochs, well before the end of training. The later epochs are therefore retained only to expose the diverging trend of the training and validation curves, not because they contribute to the reported performance. These loss and accuracy curves are specific to the iterative training of the network and have no counterpart in Pipeline 1, which is parameter-free, or in the classifiers of Pipeline 2.

3.5.4 Statistical analysis

All analyses are carried out on the balanced accuracies of Section 3.5.3, one value per held-out fold. Following the Cross-validation protocol, each of the fifteen recording sessions is treated as an independent unit, so that every task is evaluated on fifteen observations per model. Given the small sample and the absence of a normality assumption, non-parametric tests are used throughout, with the significance level fixed at $\alpha = 0.05$. Because many tests are performed, p -values are corrected for multiple comparisons with the Benjamini–Hochberg procedure, which controls the false discovery rate (FDR); the correction is applied separately within each family of tests addressing the same question, rather than pooling all p -values into a single correction. The first question is whether each model performs above chance. For every task, the fifteen balanced accuracies of each model are compared with the theoretical chance level ($1/K$, that is 0.333 for the three-class tasks and 0.5 for the binary ones) by a one-sample Wilcoxon signed-rank test, the null hypothesis being that their median equals chance and the alternative one-sided (greater). This test is applied to every model evaluated in this work (the three decision rules of Pipeline 1, namely energy alone, attention alone and their combination, both classifiers of Pipeline 2, namely random forest and linear support-vector machine, and the convolutional network of Pipeline 3), and not only to the configurations later retained for the inter-pipeline comparison. The resulting p -values are corrected with Benjamini–Hochberg separately within each of the three classification problems (perception versus imagination, sensory-modality recognition and semantic-concept recognition), so that each problem forms its own family. Within each problem, all models and all conditions are corrected together as a single family, rather than model by model. The attention-only rule, being undefined for the binary perception–imagination distinction, enters only the two three-class problems. In the few degenerate cases in which the signed-rank test is undefined because a classifier attains exactly the chance level on every fold (all differences from chance being null), the corresponding p -value is set to one (no evidence against chance) and is retained in the family so as not to alter its size. The second question is whether the three pipelines differ from one another, each represented by its principal configuration: the combined energy–attention rule for Pipeline 1, the random forest for Pipeline 2 and the convolutional network for Pipeline 3. For each condition the fifteen per-session accuracies form a complete block design, one value per session and pipeline, to which a Friedman test is applied as an omnibus check. The Friedman test acts only as a gate: where it is significant, the three pipeline pairs are compared post-hoc with Wilcoxon signed-rank tests, the magnitude and direction of each difference being summarised by the median per-session difference in balanced accuracy. The Friedman p -value itself is not carried into the multiple-comparison correction, which is applied to the post-hoc

p -values alone, these forming a single family across all conditions of all tasks. The contribution of the attention-based weighting is isolated by comparing the three decision rules of Pipeline 1 among themselves, condition by condition. On the eight three-class conditions the three rules (energy-only, attention-only and their combination) are compared with a Friedman omnibus test; as no omnibus test reached significance in any condition, no post-hoc comparison was warranted and only the Friedman p -values are reported, without further correction. For the binary perception–imagination distinction the attention-only rule is undefined, so the combined and energy-only rules alone are compared, per sensory modality, by a paired Wilcoxon signed-rank test; given the small number of comparisons, raw p -values are reported without multiple-comparison correction, each summarised by the median per-session difference in balanced accuracy. All tests are computed with the `scipy.stats` module (Wilcoxon signed-rank and Friedman tests), the Benjamini–Hochberg correction being implemented following the standard step-up procedure.

3.5.5 Experimental Design

Pipeline 1. Within Pipeline 1 the only components exposing configurable parameters are the spatial-covariance estimator and the region-wise CSP; everything downstream, the regional attention, the energy and the three decision rules, is parameter-free, so that the CSP spatial filters remain the sole quantities learned from data. The configuration of these two components is reported in Table 3.3.

Pipeline 2. The configurable parameters of Pipeline 2 concern four stages: the band-pass filter that decomposes each epoch into the five physiological rhythms, the spatial-covariance estimator, the Riemannian alignment (the tangent-space mapping and its reference mean), and the two supervised classifiers, a linear support vector machine and a random forest, applied to the resulting tangent vectors. To keep the groups distinct they are reported separately: the band-pass filter and the frequency bands in Table 3.4, the covariance and alignment in Table 3.5, and the classifier settings in Table 3.6.

Table 3.3: Configuration of Pipeline 1. Settings of the two data-driven components, both from PyRiemann (v0.9): the spatial-covariance estimator that feeds the CSP, and the region-wise CSP. The CSP spatial filters are the sole quantities learned from data.

Component	Parameter	Value / meaning
Covariance estimator	<code>estimator</code>	Ledoit–Wolf shrinkage (<code>lwf</code>) [†]
Region-wise CSP	<code>nfilter</code>	2 — two retained filters (the two extremes of the CSP ordering)
	<code>metric</code>	<code>euclid</code> — arithmetic (Euclidean) mean of class covariances
	<code>ajd_method</code>	<code>ajd_pham</code> — not invoked (binary One-vs-Rest sub-problems)
	<code>log</code>	<code>True</code> — bypassed: <code>csp.filters_</code> (W) is used directly

Table 3.4: Band-pass filter and frequency bands of Pipeline 2. Each standardised epoch is decomposed into five band-limited versions by zero-phase band-pass filtering. The design cut-offs are the frequencies passed to the filter; for a Chebyshev type II they mark the -20 dB stop-band edges and were deliberately widened with respect to the target physiological band, so that the realised passband coincides with it.

Filter parameter	Value
Implementation	MNE <code>filter_data</code> (<code>method='iir'</code>)
Filter type	Chebyshev type II (<code>cheby2</code>)
Order	4 (effective 8 after the forward-backward pass)
Stop-band attenuation	20 dB (<code>rs</code>)
Phase	zero-phase, forward-backward (<code>filtfilt</code>)

Band	Symbol	Design cut-offs (Hz)	Target band (Hz)
Delta	δ	0.75–5	1–4*
Theta	θ	3.5–9	4–8
Alpha	α	7.5–13	8–12
Beta	β	10–35	12–30
Gamma	γ	23–107	30–100

* For delta the realised lower edge is set by the 2 Hz preprocessing high-pass rather than by the design cut-off, so the effective band is ≈ 2 –4 Hz, consistent with the band decomposition described earlier.

Table 3.5: Covariance estimation and Riemannian alignment in Pipeline 2.

Both components are from PyRiemann (v0.9). The per-band reference mean is the only quantity estimated from the training data: it is fixed at training time and reused, unchanged, to project the held-out subject.

Stage	Parameter	Value
Covariance estimator	<code>estimator</code>	Ledoit–Wolf shrinkage (<code>lwf</code>)
Tangent-space metric	<code>metric</code>	Affine-invariant Riemannian (<code>riemann</code>)
Reference point	—	Riemannian (Fréchet) mean of the training covariances
Covariate-shift update	<code>tsupdate</code>	<code>False</code> (reference fixed from training)

Table 3.6: Classifier settings in Pipeline 2. The tangent vectors are fed to two supervised classifiers, trained and evaluated separately (scikit-learn v1.6.1). Parameters not listed are left at the scikit-learn defaults.

Classifier	Parameter	Value
Linear SVM (SVC)	<code>kernel</code>	linear
	<code>C</code>	1.0
	<code>multiclass strategy</code>	one-vs-one
Random forest	<code>n_estimators</code>	1500
	<code>max_depth</code>	10
	<code>criterion</code>	gini
	<code>max_features</code>	sqrt
	<code>bootstrap</code>	True
	<code>random_state</code>	42

Table 3.7: Training configuration of Pipeline 3. Optimisation and regularisation settings of the convolutional network (TensorFlow v2.17.0); the network architecture is described in Section 3.4. All settings are shared across tasks, except the loss, which matches the output layer.

Hyperparameter	Value
Optimiser	Adam (learning rate 10^{-3})
Batch size	64
Epochs (max.)	100
Early stopping	patience 20, validation accuracy, best weights restored
Loss (binary task)	binary cross-entropy
Loss (3-class tasks)	sparse categorical cross-entropy
Dropout rate	0.3 (convolutional layers), 0.5 (dense layer)

Chapter 4

Results

All three pipelines are assessed with the same metric, the balanced accuracy of Section 3.5.3, computed per session under the Leave-One-Subject-Out protocol (Section 3.5). The chapter is organised as follows. For each of the three principal configuration pipelines, and for each classification task, the confusion matrix is reported together with a summary table giving the mean balanced accuracy ($\pm$ standard deviation) and the p -values (both raw and after Benjamini–Hochberg correction) of the one-sided Wilcoxon signed-rank test against chance. The principal configuration is: the combined energy–attention rule for Pipeline 1 (since this is the decision rule around which the proposed methodology was designed around), the random forest for Pipeline 2 (since the linear SVM failed to perform above chance) and the Convolutional Neural Network for Pipeline 3 (which admits a single configuration). The per-subject balanced accuracies and the per-class metrics (precision, recall and F1-score) are collected in Appendix A. The three pipelines are then compared: first, a boxplot of the per-session balanced accuracies (from table reported in Appendix A) provides a visual summary of the classification level attained by each method; this is followed by the tables reporting, for every pipeline pair, the median per-session difference in balanced accuracy (Δ BA) together with the corresponding raw and corrected p -values. Finally, the configurations not retained as principal (the remaining decision rules of Pipeline 1 in Appendix B and the linear SVM of Pipeline 2 in Appendix C) are reported separately, in a more compact form. For each of them the per-subject and mean balanced accuracies (with the Wilcoxon test against chance) and the per-class metrics are given; no confusion matrix is included, as these configurations do not perform above chance and the per-class recalls already suffice to reveal any prediction bias or class collapse, so a full matrix would take up space without adding much.

4.1 Pipeline 1

4.1.1 Imagination VS Perception

Under Pipeline 1 the Perception–Imagination distinction does not exceed chance in any sensory modality: none of the one-sided Wilcoxon signed-rank tests against chance survives Benjamini–Hochberg correction (Table 4.1). The confusion matrices (Figure 4.1) are consistent with this: predictions are biased toward perception in the auditory modality and toward imagination in the pictorial one, largely irrespective of the true label, while in the orthographic case trials are spread across both classes without a dominant diagonal. This asymmetry is reflected in the per-class recalls (Table A.2).

Table 4.1: Balanced accuracy (mean $\pm$ std over 15 folds) for the perception–imagination task, by sensory modality, with the one-sided Wilcoxon signed-rank test against chance (0.5): raw p -value and Benjamini–Hochberg corrected p -value. Significance is assessed on the corrected value only: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; n.s. = not significant.

Modality	Balanced accuracy	p (raw)	p (BH)
Audio	0.508 ± 0.032	0.143	0.248 n.s.
Orthographic	0.523 ± 0.090	0.165	0.248 n.s.
Pictorial	0.523 ± 0.071	0.216	0.270 n.s.

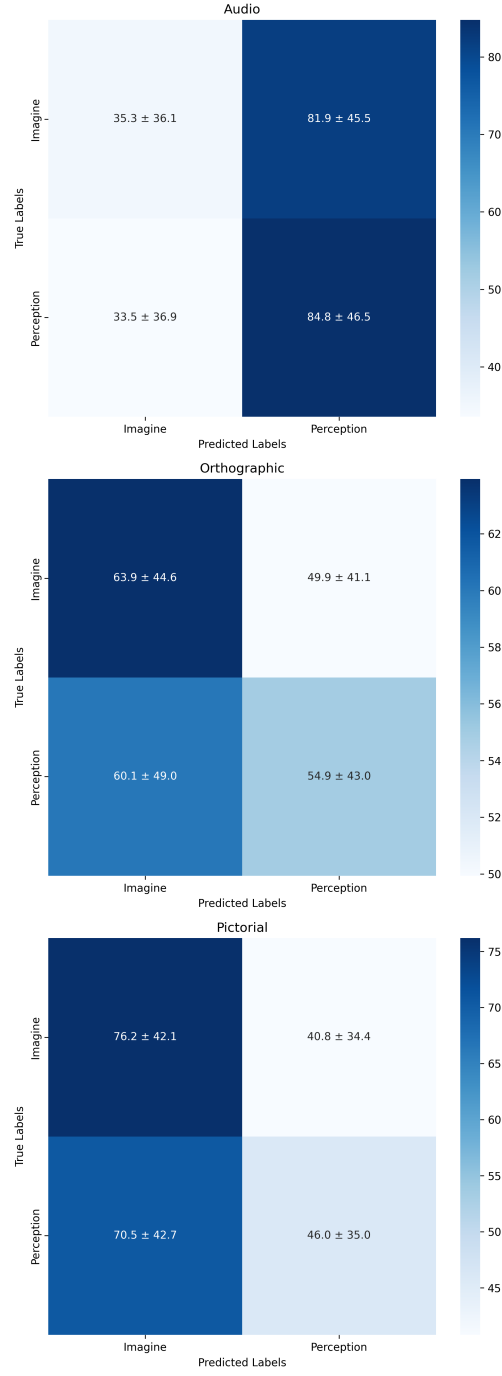


Figure 4.1: Confusion matrices for the Perception-Imagination task under Pipeline 1. Each cell reports the absolute number of trials (mean $\pm$ std over the 15 Leave-One-Subject-Out folds).

4.1.2 Sensory Modality

The Pipeline 1 also in the sensory modality cannot be discriminated above chance, neither in imagination nor in perception: neither one-sided Wilcoxon signed-rank test against chance survives Benjamini–Hochberg correction (Table 4.2). The confusion matrices (Figure 4.2) reveal a class-dependent prediction bias rather than genuine discrimination: in the imagination condition predictions concentrate on the pictorial and audio classes while the orthographic class is almost never predicted, whereas in the perception condition trials are assigned predominantly to the audio class largely irrespective of their true modality. This is reflected in the per-class recalls (Table A.4), where orthographic is effectively ignored in imagination and audio dominates in perception.

Table 4.2: Balanced accuracy (mean $\pm$ std over 15 folds) for the sensory-modality discrimination task, by task, with the one-sided Wilcoxon signed-rank test against chance (0.333): raw p -value and Benjamini–Hochberg corrected p -value. Significance is assessed on the corrected value only: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; n.s. = not significant.

Task	Balanced accuracy	p (raw)	p (BH)
Imagine	0.329 ± 0.023	0.820	1.000 n.s.
Perception	0.332 ± 0.019	0.681	1.000 n.s.

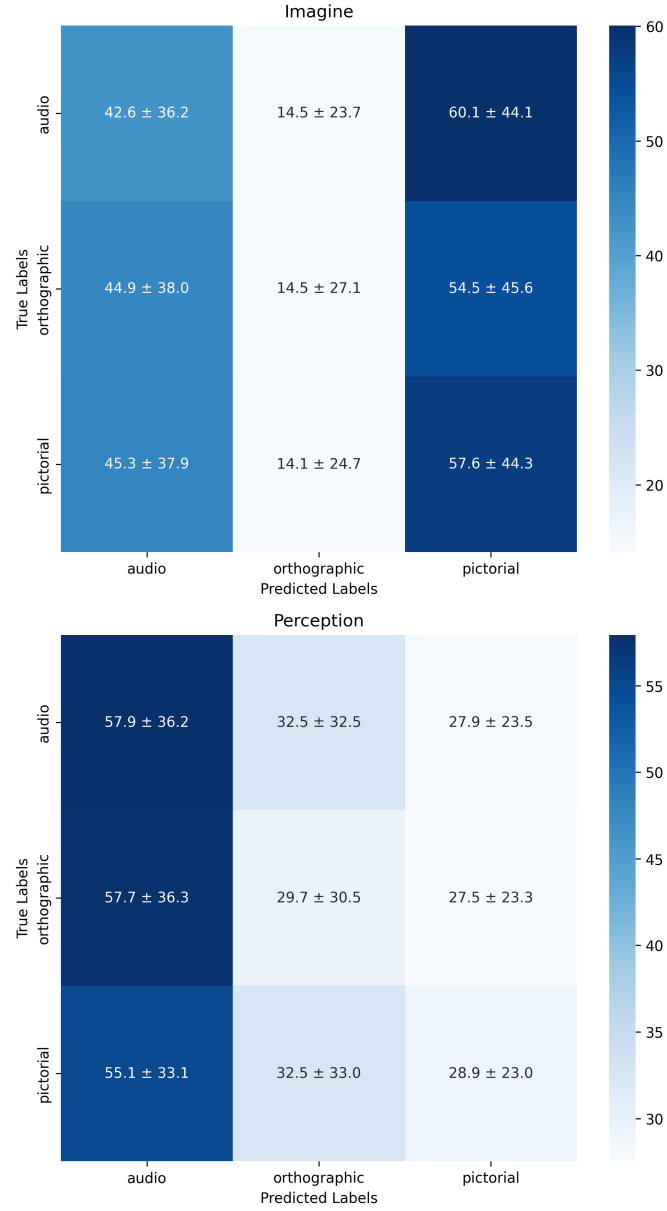


Figure 4.2: Confusion matrices for the Sensory Modality task under Pipeline 1. Each cell reports the absolute number of trials (mean $\pm$ std over the 15 Leave-One-Subject-Out folds).

4.1.3 Semantic Concept

Under Pipeline 1 the semantic concept cannot be recognised above chance in any of the six task–modality conditions: no one-sided Wilcoxon signed-rank test against chance survives Benjamini–Hochberg correction (Table 4.3). As in the previous two tasks, the confusion matrices (Figure 4.3) display a condition-dependent prediction bias rather than a diagonal structure: in each condition one concept is preferentially predicted while another is almost never selected, and the favoured concept changes from one condition to the next. The per-class recalls (Table A.6) confirm this instability—no class is consistently recovered across conditions, and the concept most reliably retrieved differs between them (for instance guitar under imagine–audio and penguin under perception–audio).

Table 4.3: Balanced accuracy (mean $\pm$ std over 15 folds) for the semantic-concept recognition task, by task–modality condition, with the one-sided Wilcoxon signed-rank test against chance (0.333): raw p -value and Benjamini–Hochberg corrected p -value. Significance is assessed on the corrected value only: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; n.s. = not significant.

Modality	Stimulus	Balanced accuracy	p (raw)	p (BH)
Imagine	Audio	0.332 ± 0.039	0.619	0.857 n.s.
Imagine	Orthographic	0.344 ± 0.028	0.076	0.227 n.s.
Imagine	Pictorial	0.315 ± 0.037	0.947	0.997 n.s.
Perception	Audio	0.331 ± 0.021	0.616	0.857 n.s.
Perception	Orthographic	0.357 ± 0.045	0.053	0.214 n.s.
Perception	Pictorial	0.349 ± 0.030	0.068	0.222 n.s.

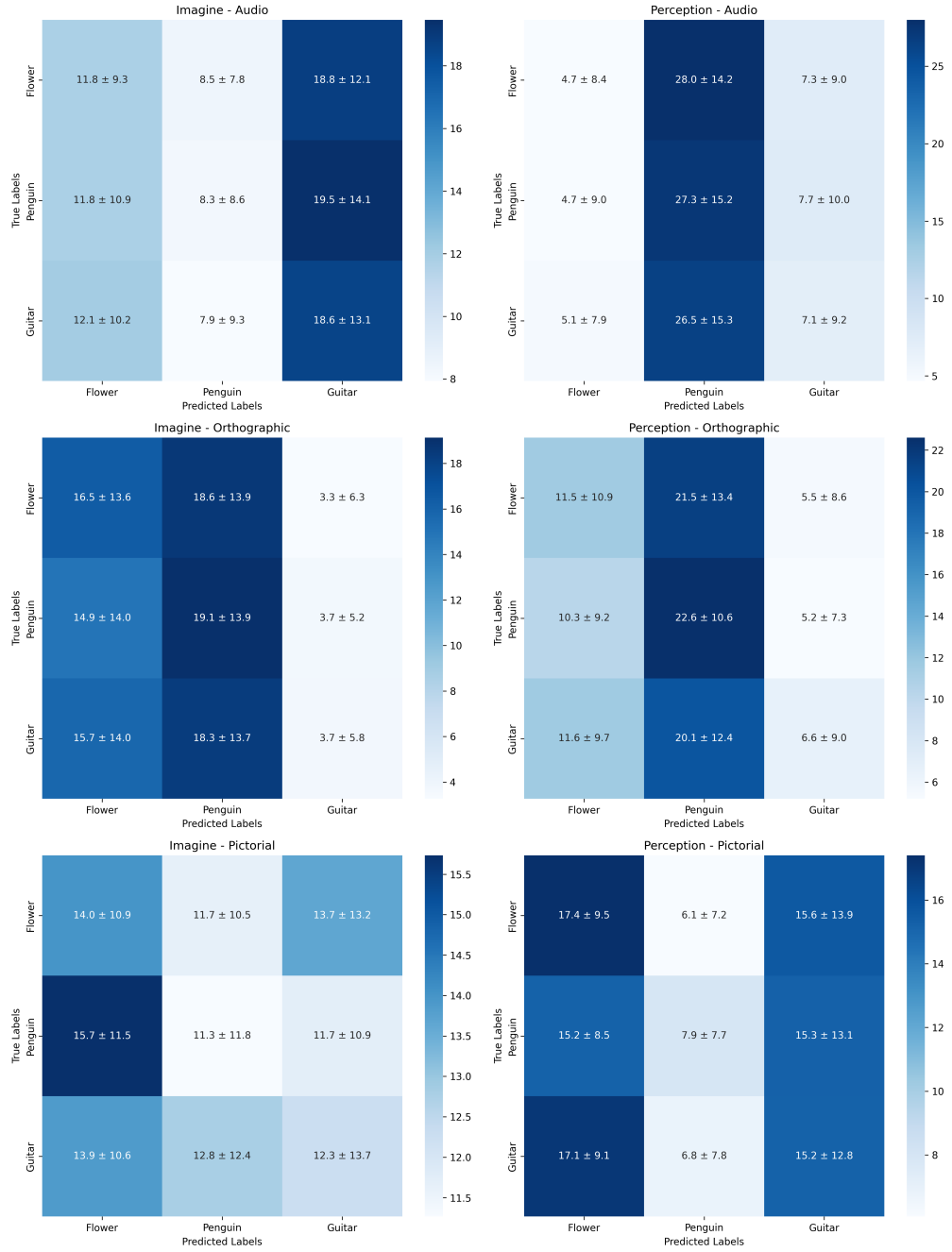


Figure 4.3: Confusion matrices for the Semantic Concept task under Pipeline 1. Each cell reports the absolute number of trials (mean $\pm$ std over the 15 Leave-One-Subject-Out folds).

4.2 Pipeline 2

4.2.1 Imagination VS Perception

Under Pipeline 2 the Perception–Imagination distinction is decoded well above chance in every sensory modality, and all three one-sided Wilcoxon signed-rank tests against chance remain significant after Benjamini–Hochberg correction (Table 4.4). Performance is highest for auditory stimuli and somewhat lower, though still clearly above chance, for the orthographic and pictorial modalities. Consistently, the confusion matrices (Figure 4.4) exhibit a marked diagonal structure: in the auditory modality the two classes are almost perfectly separated, while in the orthographic and pictorial ones the imagination class is recovered very reliably and most of the residual error falls on perception trials misclassified as imagination. The per-class metrics (Table A.8) reflect this asymmetry, with near-perfect recall for imagination and a lower, more variable recall for perception outside the auditory modality.

Table 4.4: Balanced accuracy (mean $\pm$ std over 15 folds) for the perception–imagination task, by sensory modality, with the one-sided Wilcoxon signed-rank test against chance (0.5): raw p -value and Benjamini–Hochberg corrected p -value. Significance is assessed on the corrected value only: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; n.s. = not significant.

Modality	Balanced accuracy	p (raw)	p (BH)
Audio	0.937 ± 0.125	< 0.001	< 0.001 ***
Orthographic	0.789 ± 0.170	< 0.001	< 0.001 ***
Pictorial	0.784 ± 0.182	< 0.001	0.001 **

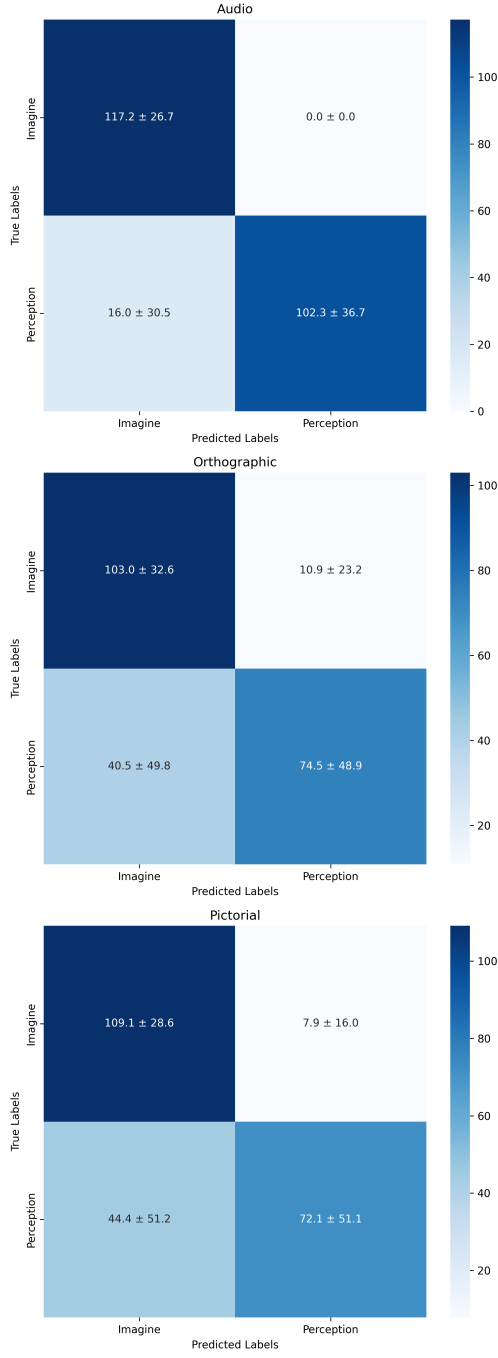


Figure 4.4: Confusion matrices for the Perception-Imagination task under Pipeline 2. Each cell reports the absolute number of trials (mean $\pm$ std over the 15 Leave-One-Subject-Out folds).

4.2.2 Sensory Modality

Under Pipeline 2 the sensory modality is discriminated above chance in both conditions, both one-sided Wilcoxon signed-rank tests against chance surviving Benjamini–Hochberg correction (Table 4.5); the two conditions are, however, markedly different in magnitude. In perception the balanced accuracy reflects genuine discrimination, whereas in imagination it lies only marginally above chance. The confusion matrices (Figure 4.5) clarify this contrast: in the perception condition the auditory class is recovered very reliably while the orthographic and pictorial classes are largely confused with each other, so that the above-chance performance is driven almost entirely by the separability of auditory stimuli. In the imagination condition no class is properly recovered, the orthographic class is almost never predicted and trials are assigned mainly to the audio and pictorial classes, so that the marginal significance reflects a prediction bias rather than genuine modality discrimination. The per-class metrics (Table A.10) confirm this reading, with high auditory recall in perception and a near-zero orthographic recall in imagination.

Table 4.5: Balanced accuracy (mean $\pm$ std over 15 folds) for the sensory-modality discrimination task, by task, with the one-sided Wilcoxon signed-rank test against chance (0.333): raw p -value and Benjamini–Hochberg corrected p -value. Significance is assessed on the corrected value only: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; n.s. = not significant.

Task	Balanced accuracy	p (raw)	p (BH)
Imagine	0.348 ± 0.019	0.009	0.027 *
Perception	0.567 ± 0.089	< 0.001	< 0.001 ***

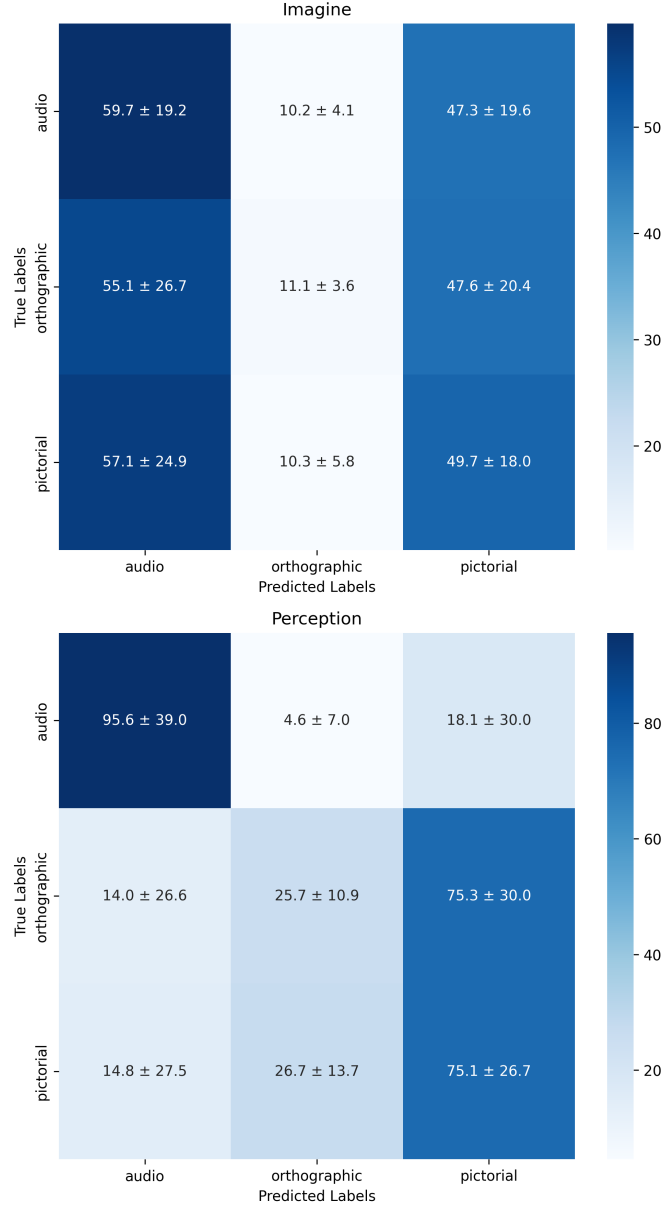


Figure 4.5: Confusion matrices for the Sensory Modality task under Pipeline 2. Each cell reports the absolute number of trials (mean $\pm$ std over the 15 Leave-One-Subject-Out folds).

4.2.3 Semantic Concept

Under Pipeline 2 the semantic concept cannot be recognised above chance in any of the six task–modality conditions: no one-sided Wilcoxon signed-rank test against chance survives Benjamini–Hochberg correction (Table 4.6). As already observed for Pipeline 1, the confusion matrices (Figure 4.6) show a condition-dependent prediction bias rather than a diagonal structure, one concept being preferentially predicted while another is seldom selected, with the favoured concept changing across conditions. The per-class metrics (Table A.12) confirm the absence of genuine recognition, no class being consistently recovered across conditions. The covariance-based representation that proved effective for the perception–imagination distinction therefore carries no discriminative information for the semantic concept under a subject-independent regime.

Table 4.6: Balanced accuracy (mean $\pm$ std over 15 folds) for the semantic-concept recognition task, by task–modality condition, with the one-sided Wilcoxon signed-rank test against chance (0.333): raw p -value and Benjamini–Hochberg corrected p -value. Significance is assessed on the corrected value only: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; n.s. = not significant.

Modality	Stimulus	Balanced accuracy	p (raw)	p (BH)
Imagine	Audio	0.331 ± 0.035	0.700	0.857 n.s.
Imagine	Orthographic	0.339 ± 0.046	0.339	0.679 n.s.
Imagine	Pictorial	0.307 ± 0.031	0.997	0.997 n.s.
Perception	Audio	0.327 ± 0.064	0.640	0.857 n.s.
Perception	Orthographic	0.327 ± 0.028	0.719	0.857 n.s.
Perception	Pictorial	0.332 ± 0.036	0.555	0.857 n.s.

Statistical Comparison Against Chance Level

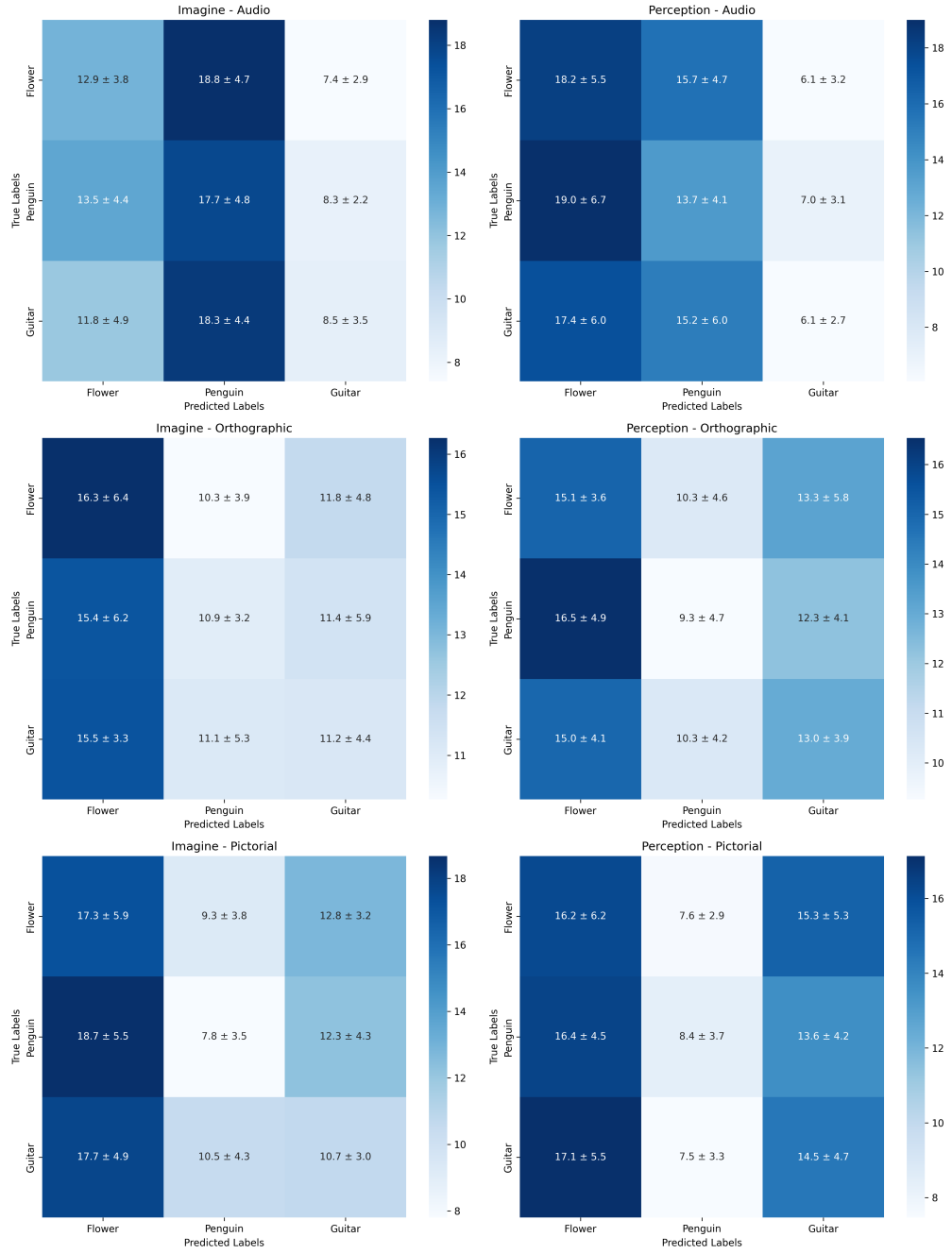


Figure 4.6: Confusion matrices for the Semantic Concept task under Pipeline 2. Each cell reports the absolute number of trials (mean $\pm$ std over the 15 Leave-One-Subject-Out folds).

4.3 Pipeline 3

4.3.1 Imagination VS Perception

Training & Validation Curves The training dynamics are consistent across the three sensory modalities (Figure 4.7). In every case the training loss decreases smoothly and the training accuracy rises steadily towards high values. The validation curves, by contrast, follow a non-monotonic trend: the validation loss decreases over the first few epochs, reaches a minimum and then drifts slowly upwards, while the validation accuracy mirrors this behaviour, rising early to a peak before gradually declining. This early minimum in the validation loss marks the point of best generalisation (the one recovered by the early-stopping criterion) after which the widening gap between training and validation is the expected signature of overfitting under the subject-independent protocol: the network fits the training subjects well but transfers only partially to the held-out one. The wide shaded band, spanning the $\pm$ std across the 15 Leave-One-Subject-Out folds, indicate that this transfer varies considerably from subject to subject, so the trend, though clearly present, should be read as qualitative. The validation accuracy nonetheless stabilises above chance, anticipating the above-chance balanced accuracies reported below.

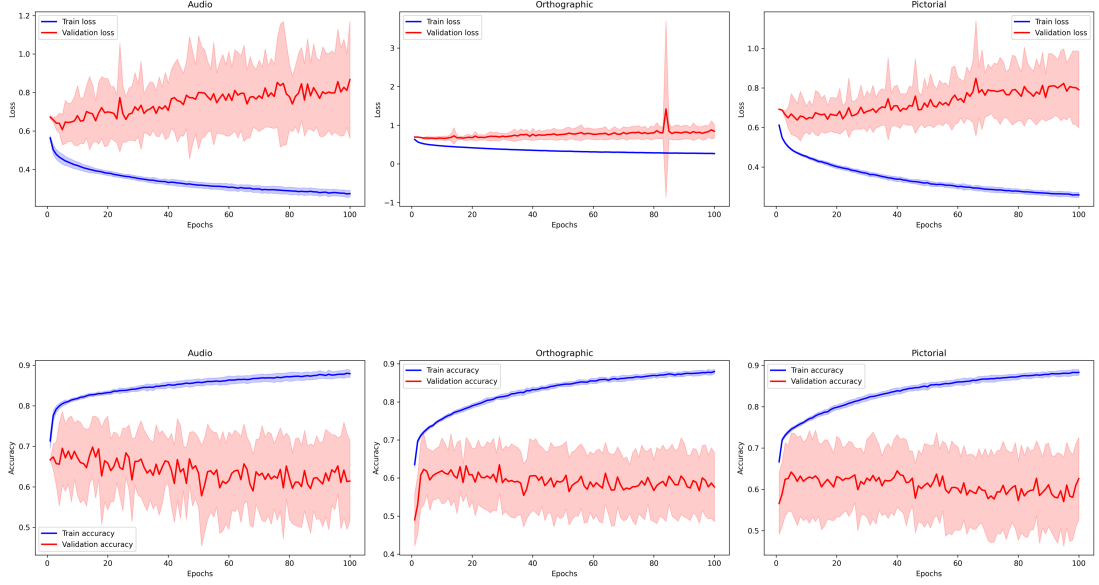


Figure 4.7: Training and validation curves for the Perception–Imagination task under Pipeline 3, by sensory modality (columns: audio, orthographic, pictorial). Top row: loss; bottom row: accuracy. Solid lines are the mean over the 15 Leave-One-Subject-Out folds and shaded bands the corresponding $\pm$ std; training curves in blue, validation curves in red.

Statistical Comparison Against Chance Level Under Pipeline 3 the Perception+Imagination distinction is decoded above chance in every sensory modality, all three one-sided Wilcoxon signed-rank tests against chance surviving Benjamini–Hochberg correction (Table 4.7); performance is comparable across the three modalities. The confusion matrices (Figure 4.8) show a clear diagonal but a pronounced asymmetry between the two classes: the imagination class is recovered very reliably, whereas a substantial share of perception trials is misclassified as imagination. The per-class metrics (Table A.14) reflect this imbalance, with high and stable recall for imagination and a markedly lower, more variable recall for perception across all modalities.

Table 4.7: Balanced accuracy (mean $\pm$ std over 15 folds) for the perception–imagination task, by sensory modality, with the one-sided Wilcoxon signed-rank test against chance (0.5): raw p -value and Benjamini–Hochberg corrected p -value. Significance is assessed on the corrected value only: $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; n.s. = not significant.

Modality	Balanced accuracy	p (raw)	p (BH)
Audio	0.690 ± 0.095	< 0.001	0.002 **
Orthographic	0.666 ± 0.058	< 0.001	< 0.001 ***
Pictorial	0.701 ± 0.054	< 0.001	< 0.001 ***

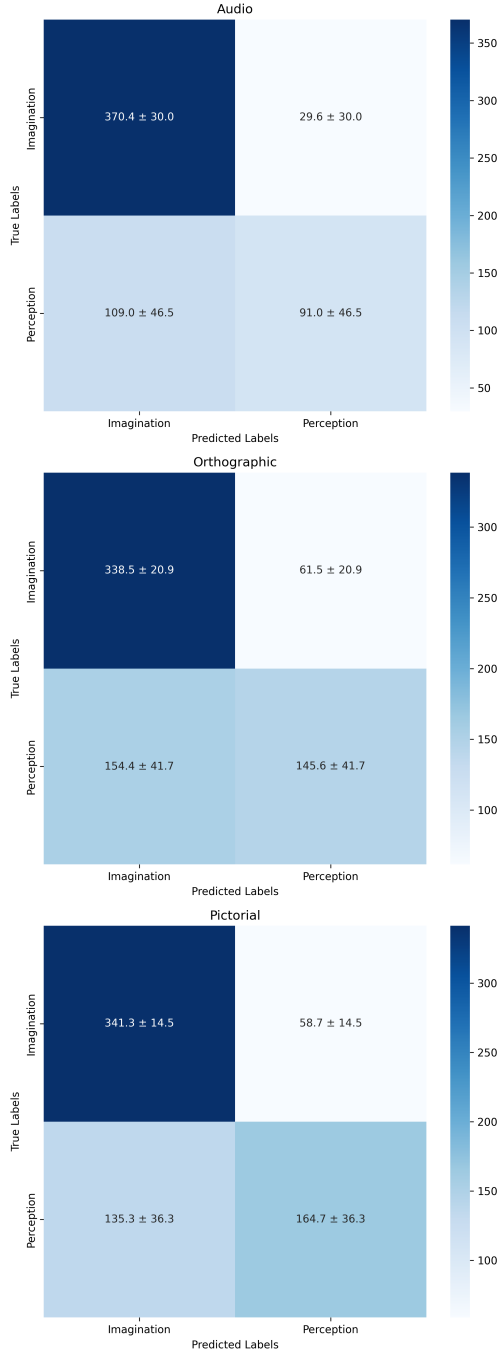


Figure 4.8: Confusion matrices for the Perception–Imagination task under Pipeline 3. Each cell reports the absolute number of trials (mean $\pm$ std over the 15 Leave-One-Subject-Out folds).

4.3.2 Sensory modality

Training & Validation Curves The training dynamics for the Sensory Modality task (Figure 4.9) again show the training accuracy rising steadily towards high values while the validation curves plateau at a far lower level, but the two conditions differ appreciably. In the imagination condition the validation accuracy remains essentially flat, only marginally above chance, and the validation loss rises from the very first epochs, indicating that little transferable structure is learned. In the perception condition the validation accuracy stabilises at a visibly higher level, pointing to a modest but genuine degree of cross-subject generalisation, consistent with the higher balanced accuracy reported below. In both conditions the validation loss is markedly noisier than for the Perception–Imagination task, with isolated spikes and wide ($\pm$) std bands across the 15 Leave-One-Subject-Out folds, so the gap between the two conditions, though clear, should be read as qualitative. It should be noted that the averaged validation curves understate the actual performance of the pipeline: read on their own they suggest a model hovering around chance, yet the balanced accuracies and confusion matrices (Table 4.8, Figure 4.10) place both conditions above chance. This discrepancy is an averaging artefact. The reported balanced accuracy is computed from the best-epoch (early-stopped) weights of each fold, whereas the curves average the validation metric epoch by epoch across the 15 folds; since each held-out subject attains its peak at a different epoch, these per-subject improvements do not align in time and largely cancel out when averaged, flattening the mean curve even though most folds individually rise above chance. The per-subject balanced accuracies in Table A.15 are consistent with this reading, being predominantly above chance yet widely dispersed across subjects.

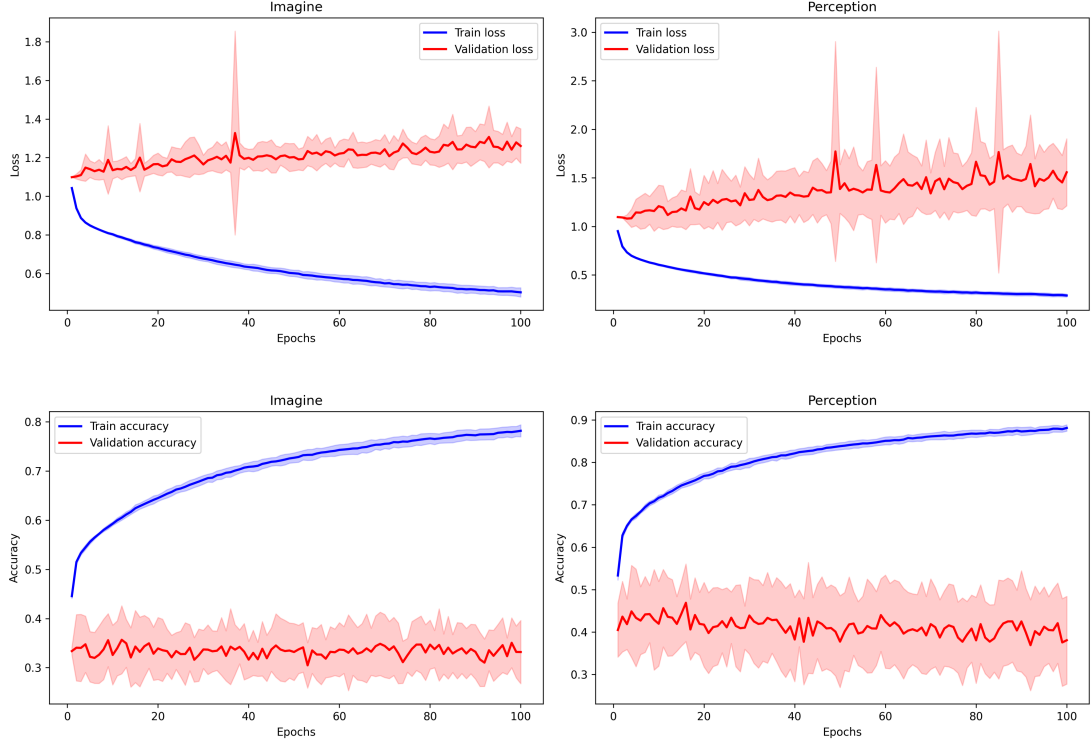


Figure 4.9: Training and validation curves for the Sensory Modality discrimination task under Pipeline 3, for the imagination (left) and perception (right) conditions. Top row: loss; bottom row: accuracy. Solid lines are the mean over the 15 Leave-One-Subject-Out folds and shaded bands the corresponding $\pm$ std; training curves in blue, validation curves in red.

Statistical Comparison Against Chance Level Under Pipeline 3 the Sensory Modality is discriminated above chance in both conditions, both one-sided Wilcoxon signed-rank tests against chance surviving Benjamini–Hochberg correction (Table 4.8), with perception somewhat higher than imagination. The confusion matrices (Figure 4.10) show a clear diagonal in both conditions: unlike Pipeline 2, where discrimination rested almost entirely on the auditory class, here all three modalities are recovered to a comparable extent, each true class being predicted most often as itself. The per-class metrics (Table A.16) confirm this more uniform behaviour, with recalls broadly balanced across the three classes rather than concentrated on a single one; in the perception condition the orthographic and pictorial classes are recovered most reliably, whereas the auditory class is the least well identified.

Table 4.8: Balanced accuracy (mean $\pm$ std over 15 folds) for the Sensory Modality discrimination task, by task, with the one-sided Wilcoxon signed-rank test against chance (0.333): raw p -value and Benjamini–Hochberg corrected p -value. Significance is assessed on the corrected value only: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; n.s. = not significant.

Task	Balanced accuracy	p (raw)	p (BH)
Imagine	0.471 ± 0.038	< 0.001	< 0.001 ***
Perception	0.538 ± 0.094	< 0.001	< 0.001 ***

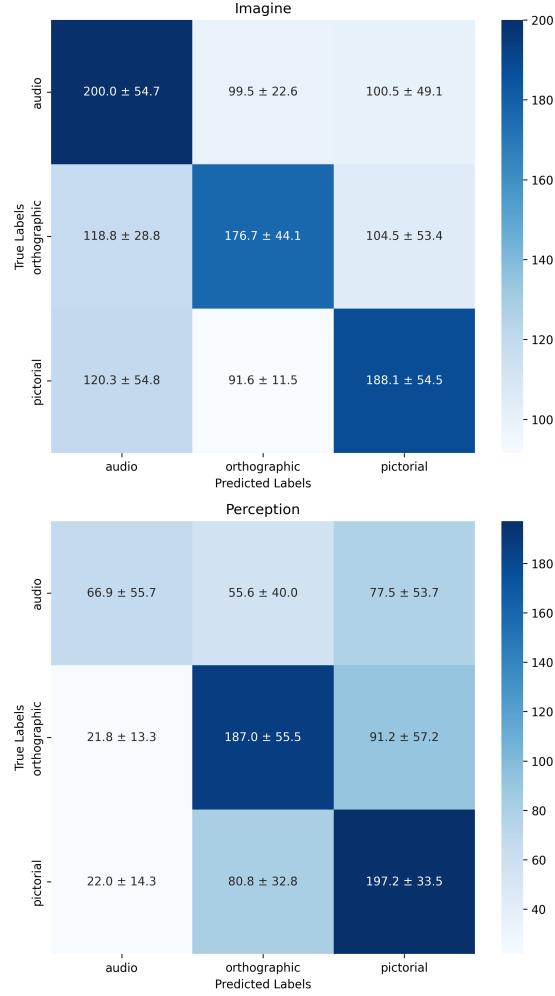


Figure 4.10: Confusion matrices for the Sensory Modality task under Pipeline 3. Each cell reports the absolute number of trials (mean $\pm$ std over the 15 Leave-One-Subject-Out folds).

4.3.3 Semantic Concept

Training & Validation Curves The training dynamics for the Semantic Concept task are shown separately for loss (Figure 4.11) and accuracy (Figure 4.12), across the six task–modality conditions. The picture is uniform and marks the most pronounced overfitting of the three tasks: in every condition the training loss decreases smoothly and the training accuracy climbs towards high values, whereas the validation loss rises from the early epochs and the validation accuracy remains essentially flat, close to chance, throughout training. The shaded $\pm$ std bands across the 15 Leave-One-Subject-Out folds are particularly wide, especially in the

perception conditions, indicating that the little transferable structure the network captures varies strongly from subject to subject. As with the sensory-modality task, the averaged validation curves understate the actual performance: the balanced accuracies reported in Table 4.9 place all six conditions above chance, a level that the epoch-by-epoch mean masks because each held-out subject reaches its best epoch at a different point and these peaks cancel out under averaging.

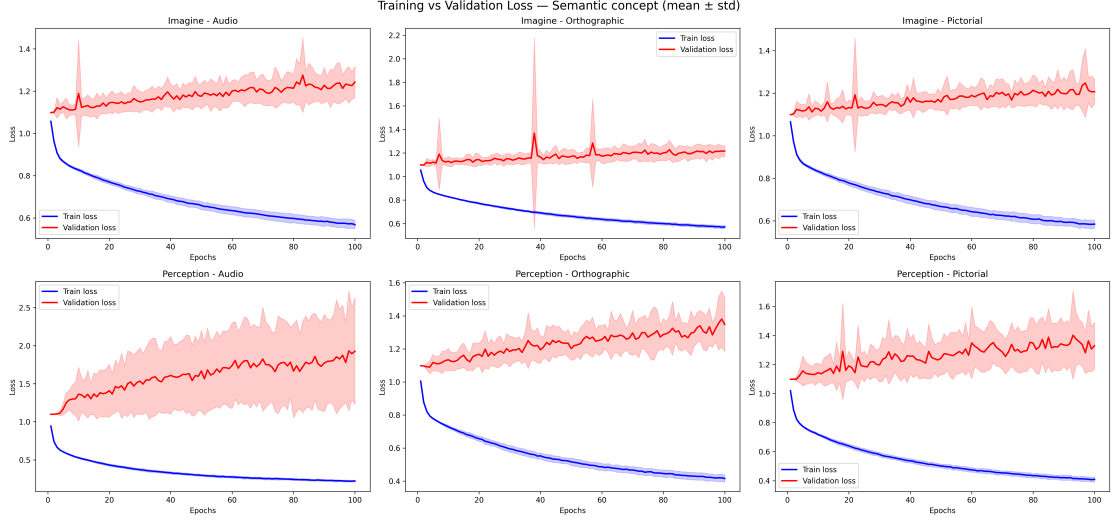


Figure 4.11: Training and validation loss for the Semantic Concept recognition task under Pipeline 3, for each task–modality condition (top row: imagination; bottom row: perception $\times$ audio, orthographic, pictorial). Solid lines are the mean over the 15 Leave-One-Subject-Out folds and shaded bands the corresponding $\pm$ std; training loss in blue, validation loss in red.

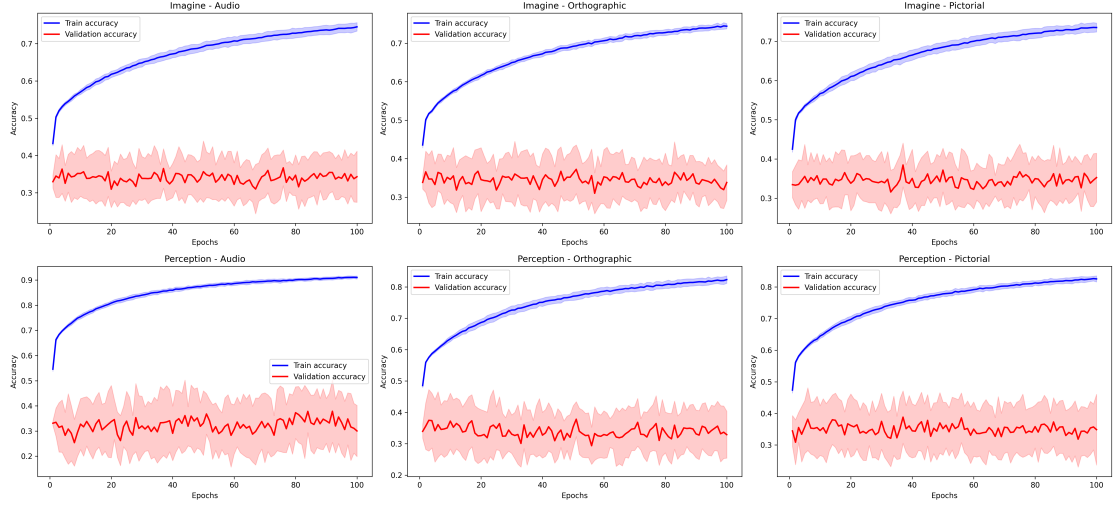


Figure 4.12: Training and validation accuracy for the Semantic Concept recognition task under Pipeline 3, for each task–modality condition (top row: imagination; bottom row: perception $\times$ audio, orthographic, pictorial). Solid lines are the mean over the 15 Leave-One-Subject-Out folds and shaded bands the corresponding $\pm$ std; training accuracy in blue, validation accuracy in red.

Statistical Comparison Against Chance Level Under Pipeline 3 the semantic concept is recognised above chance in all six task–modality conditions, every one-sided Wilcoxon signed-rank test against chance surviving Benjamini–Hochberg correction (Table 4.9). This is the only pipeline to exceed chance on the semantic-concept task, where both Pipeline 1 and Pipeline 2 remained at chance level in every condition. The confusion matrices (Figure 4.13) display a discernible diagonal throughout: in each condition every true concept is predicted most often as itself, and the per-class metrics (Table A.18) show recalls broadly balanced across flower, penguin and guitar rather than concentrated on a single concept. The margins above chance are nonetheless modest and, as reflected in the wide per-fold standard deviations, vary appreciably across held-out subjects.

Table 4.9: Balanced accuracy (mean $\pm$ std over 15 folds) for the semantic-concept recognition task, by task–modality condition, with the one-sided Wilcoxon signed-rank test against chance (0.333): raw p -value and Benjamini–Hochberg corrected p -value. Significance is assessed on the corrected value only: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; n.s. = not significant.

Modality	Stimulus	Balanced accuracy	p (raw)	p (BH)
Imagine	Audio	0.460 ± 0.040	< 0.001	< 0.001 ***
Imagine	Orthographic	0.477 ± 0.038	< 0.001	< 0.001 ***
Imagine	Pictorial	0.464 ± 0.034	< 0.001	< 0.001 ***
Perception	Audio	0.524 ± 0.114	< 0.001	< 0.001 ***
Perception	Orthographic	0.492 ± 0.039	< 0.001	< 0.001 ***
Perception	Pictorial	0.492 ± 0.059	< 0.001	< 0.001 ***

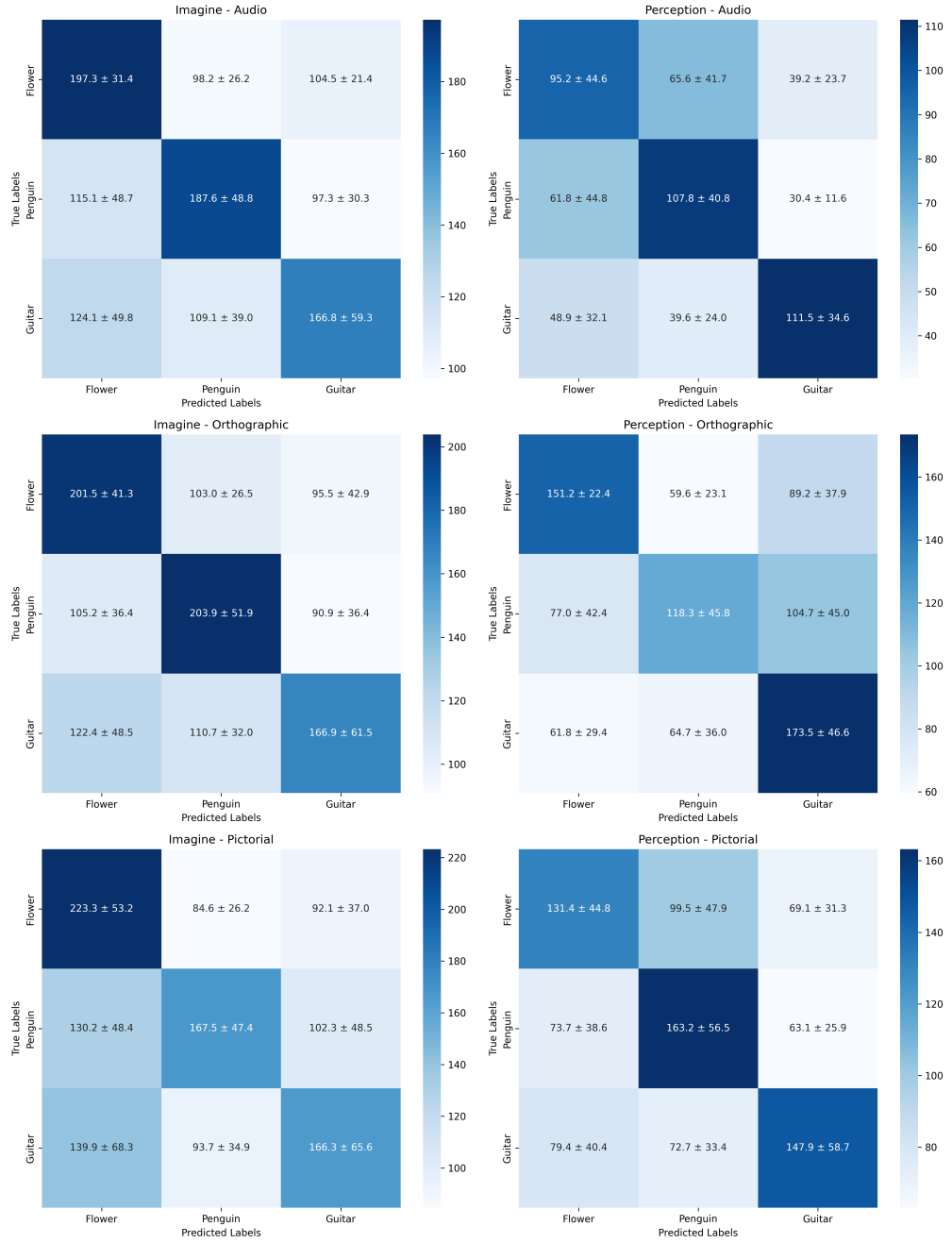


Figure 4.13: Confusion matrices for the Semantic Concept task under Pipeline 3. Each cell reports the absolute number of trials (mean $\pm$ std over the 15 Leave-One-Subject-Out folds).

4.4 Inter- and Intra- Pipeline Comparisons

This section compares the pipelines from two complementary angles. The inter-pipeline analysis contrasts the three principal configurations (the combined Energy–Attention rule for Pipeline 1, the Random Forest for Pipeline 2 and the Convolutional Network for Pipeline 3) on each of the three tasks. The intra-pipeline analysis then examines the remaining configurations of Pipelines 1 and 2 (the alternative decision rules and the linear SVM), to establish whether they differ from the retained ones. For the three principal configurations, which are the only ones to yield above-chance performance, a per-session boxplot is reported alongside the statistical comparison, so as to describe the dispersion of the balanced accuracies across subjects. For the remaining configurations no boxplot is shown, since they do not perform above chance; these are assessed by the statistical analysis alone (a Friedman omnibus test, where applicable, gating pairwise Wilcoxon signed-rank post-hoc tests corrected with Benjamini–Hochberg) to determine whether the choice of decision rule or classifier has any measurable effect.

4.4.1 Comparison Across the Three Main Pipelines

Perception VS Imagination The three pipelines are first compared on the Perception–Imagination task. Figure 4.14 summarises the per-session balanced accuracies of the three principal configurations, and Table 4.10 reports the corresponding statistical comparison: for each modality a Friedman omnibus test gates the three pairwise Wilcoxon signed-rank post-hoc tests, each summarised by the median per-session difference in balanced accuracy (ΔBA). The Friedman test is significant in all three modalities (all $p < 0.001$), so the pipelines are not equivalent. The post-hoc comparisons establish a consistent ordering, $P2 > P3 > P1$: Pipeline 2 significantly outperforms both Pipeline 1 and Pipeline 3 in every modality, and Pipeline 3 in turn significantly outperforms Pipeline 1. The advantage of Pipeline 2 over Pipeline 3 is largest for auditory stimuli and narrows for the visual modalities, becoming non-significant in the pictorial condition after Benjamini–Hochberg correction; the ordering itself, however, holds throughout. The boxplots are consistent with this picture: Pipeline 1 sits astride the chance line in every modality, whereas Pipeline 2 and Pipeline 3 lie clearly above it, Pipeline 2 attaining the highest medians (nearly perfect for audio) at the cost of a visibly wider spread across sessions than the more compact Pipeline 3.

Table 4.10: Inter-pipeline comparison for the Imagination–Perception task. For each condition, a Friedman omnibus test (uncorrected, used only as a gate) precedes pairwise Wilcoxon signed-rank post-hoc tests, summarised by the median per-session difference in balanced accuracy (ΔBA , first minus second pipeline). Post-hoc p -values are reported raw and after Benjamini–Hochberg correction. Significance on the corrected value: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; n.s. = not significant.

Condition	Friedman p	Comparison	Median ΔBA	p (raw)	p (BH)
Audio	< 0.001	P1–P2	−0.481	< 0.001	< 0.001 ***
		P1–P3	−0.220	< 0.001	< 0.001 ***
		P2–P3	+0.259	< 0.001	< 0.001 ***
Orthographic	< 0.001	P1–P2	−0.339	0.001	0.002 **
		P1–P3	−0.133	< 0.001	< 0.001 ***
		P2–P3	+0.201	0.018	0.024 *
Pictorial	< 0.001	P1–P2	−0.343	0.002	0.002 **
		P1–P3	−0.188	< 0.001	< 0.001 ***
		P2–P3	+0.159	0.229	0.280 n.s.

Sensory Modality The three pipelines are next compared on the Sensory Modality task. Figure 4.15 summarises the per-session balanced accuracies of the three principal configurations, and Table 4.11 reports the corresponding statistical comparison, the Friedman omnibus gating the pairwise Wilcoxon signed-rank post-hoc tests as before. The Friedman test is significant in both conditions (all $p < 0.001$). The post-hoc pattern, however, differs between them. In the imagination condition Pipeline 3 significantly outperforms both other pipelines, while Pipeline 2 exceeds Pipeline 1 only marginally, so that the ordering is $P3 > P2 \gtrsim P1$, all three lying close to chance. In the perception condition both Pipeline 2 and Pipeline 3 significantly outperform Pipeline 1, but they do not differ significantly from each other (median $\Delta BA +0.032$, n.s. after correction), so that the two comparators are statistically on par and clearly above Pipeline 1. The boxplots are consistent with this: Pipeline 1 remains centred on the chance line in both conditions, whereas in perception Pipeline 2 and Pipeline 3 rise well above it with comparable medians, though Pipeline 3 is again the more consistent across sessions, its values being less dispersed than those of Pipeline 2.

Table 4.11: Inter-pipeline comparison for the sensory modality task. For each condition, a Friedman omnibus test (uncorrected, used only as a gate) precedes pairwise Wilcoxon signed-rank post-hoc tests, summarised by the median per-session difference in balanced accuracy (ΔBA , first minus second pipeline). Post-hoc p -values are reported raw and after Benjamini–Hochberg correction. Significance on the corrected value: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; n.s. = not significant.

Condition	Friedman p	Comparison	Median ΔBA	p (raw)	p (BH)
Imagine	< 0.001	P1–P2	−0.021	0.015	0.021 *
		P1–P3	−0.141	< 0.001	< 0.001 ***
		P2–P3	−0.108	< 0.001	< 0.001 ***
Perception	< 0.001	P1–P2	−0.270	< 0.001	< 0.001 ***
		P1–P3	−0.185	< 0.001	< 0.001 ***
		P2–P3	+0.032	0.561	0.639 n.s.

Semantic Concept The three pipelines are finally compared on the Semantic Concept task. Figure 4.16 summarises the per-session balanced accuracies of the three principal configurations across the six task–modality conditions, and Table 4.12 reports the corresponding statistical comparison, the Friedman omnibus gating the pairwise Wilcoxon signed-rank post-hoc tests as before. The Friedman test is significant in all six conditions (all $p < 0.001$): Pipeline 3 significantly outperforms both Pipeline 1 and Pipeline 2 in every condition, whereas Pipeline 1 and Pipeline 2 never differ significantly from each other (median ΔBA close to zero, n.s. after correction in all six conditions). The ordering is therefore $\text{P3} > \text{P1} \approx \text{P2}$ uniformly across tasks and modalities. The boxplots make the separation plain: Pipeline 1 and Pipeline 2 overlap almost entirely and sit on the chance line, while Pipeline 3 lies clearly above it in every condition. This is the sharpest inter-pipeline dissociation of the three tasks: the semantic concept is recovered only by the topographic convolutional network, the handcrafted and covariance-based representations remaining at chance regardless of the cognitive task or sensory modality.

Table 4.12: Inter-pipeline comparison for the semantic concept task. For each condition, a Friedman omnibus test (uncorrected, used only as a gate) precedes pairwise Wilcoxon signed-rank post-hoc tests, summarised by the median per-session difference in balanced accuracy (ΔBA , first minus second pipeline). Post-hoc p -values are reported raw and after Benjamini–Hochberg correction. Significance on the corrected value: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; n.s. = not significant.

Condition	Friedman p	Comparison	Median ΔBA	p (raw)	p (BH)
Imagine-Audio	< 0.001	P1–P2	+0.000	0.975	0.978 n.s.
		P1–P3	−0.144	< 0.001	< 0.001 ***
		P2–P3	−0.143	< 0.001	< 0.001 ***
Imagine-Orthographic	< 0.001	P1–P2	+0.020	0.978	0.978 n.s.
		P1–P3	−0.139	< 0.001	< 0.001 ***
		P2–P3	−0.126	< 0.001	< 0.001 ***
Imagine-Pictorial	< 0.001	P1–P2	+0.018	0.599	0.659 n.s.
		P1–P3	−0.152	< 0.001	< 0.001 ***
		P2–P3	−0.152	< 0.001	< 0.001 ***
Perception-Audio	< 0.001	P1–P2	−0.012	0.934	0.978 n.s.
		P1–P3	−0.206	< 0.001	< 0.001 ***
		P2–P3	−0.187	< 0.001	< 0.001 ***
Perception-Orthographic	< 0.001	P1–P2	+0.032	0.107	0.136 n.s.
		P1–P3	−0.119	< 0.001	< 0.001 ***
		P2–P3	−0.174	< 0.001	< 0.001 ***
Perception-Pictorial	< 0.001	P1–P2	+0.017	0.277	0.326 n.s.
		P1–P3	−0.133	< 0.001	< 0.001 ***
		P2–P3	−0.164	< 0.001	< 0.001 ***

4.4.2 Pipeline 1: Comparison of the Three Decision Rules

The three decision rules of Pipeline 1 (combined Energy–Attention, Energy-Only and Attention-Only) are compared to assess whether the attention weighting adds discriminative value beyond regional energy alone. The energy-only rule (Appendix B.1) and the attention-only rule (Appendix B.2) behave essentially like the combined rule, remaining at chance on the sensory-modality and semantic-concept tasks, with the sole exception of the energy-only rule, which marginally exceeds chance for auditory stimuli in the perception–imagination task. This qualitative equivalence is confirmed statistically. For the binary Perception–Imagination task (where the attention-only rule is undefined) the combined and energy-only rules are compared by a paired Wilcoxon signed-rank test (Table 4.13); for the two three-class tasks the three rules are compared by a Friedman omnibus test per condition, reported in Appendix D. The rules are statistically indistinguishable almost everywhere: no Friedman test reaches significance in either the Sensory Modality or the Semantic Concept task, and in the Perception–Imagination task the combined and energy-only rules differ significantly only for auditory stimuli (median ΔBA -0.004 , $p = 0.046$; Table 4.13). The negative sign is notable: the attention weighting, in the only case where it has a measurable effect, slightly lowers the balanced accuracy rather than raising it. The paired comparison between the two rules is reported without multiple-comparison correction, given the small number of tests here.

Table 4.13: Pipeline 1 ablation for the Perception–Imagination task: combined energy–attention rule versus energy-only rule, per sensory modality. Each row reports the median per-session difference in balanced accuracy (ΔBA , combined minus energy-only) and the raw p -value of the paired Wilcoxon signed-rank test. Significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; n.s. = not significant.

Condition	Median ΔBA	p (raw)
Audio	-0.004	0.046 *
Orthographic	$+0.007$	0.277 n.s.
Pictorial	-0.006	0.229 n.s.

4.4.3 Pipeline 2: Comparison of the Two Classifiers

The two classifiers applied to the Pipeline 2 tangent-space features, the random forest and the linear SVM, are compared per condition by a paired Wilcoxon signed-rank test, summarised by the median per-session difference in balanced accuracy (ΔBA , random forest minus linear SVM). The random forest is the principal configuration; the full classification results of the linear SVM are reported in Appendix C.

Imagination VS Perception On the Perception–Imagination task the linear SVM is fully degenerate. Its balanced accuracy is exactly the chance level in every session and modality (0.500 ± 0.000 ; Table C.2), with no variance across the 15 Leave-One-Subject-Out folds (Table C.1). The per-class metrics (Table C.3) reveal the cause: in the auditory and orthographic modalities the imagination class is never predicted (recall 0.000) while every trial is assigned to perception (recall 1.000), and in the pictorial modality the same collapse occurs in the opposite direction, onto imagination. The classifier therefore predicts a single constant class per fold, which for a balanced two-class problem yields a balanced accuracy of exactly 0.5 irrespective of the true labels. The comparison with the random forest is reported in Table 4.14. The random forest significantly outperforms the linear SVM in all three modalities after Benjamini–Hochberg correction, by a wide margin: the median per-session difference reaches +0.496 balanced-accuracy points for auditory stimuli and exceeds +0.37 for the two visual modalities. This gap does not reflect a close contest between two working classifiers but the collapse of the linear SVM onto a constant prediction, against which the random forest recovers a genuinely above-chance decoding of the perception–imagination distinction.

Table 4.14: Comparison of the two Pipeline 2 classifiers (random forest vs linear SVM) for the Imagination–Perception task. For each condition a paired Wilcoxon signed-rank test is reported, summarised by the median per-session difference in balanced accuracy (ΔBA , random forest minus linear SVM). p -values are given raw and after Benjamini–Hochberg correction. Significance on the corrected value: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; n.s. = not significant.

Condition	Median ΔBA	p (raw)	p (BH)
Audio	+0.496	< 0.001	< 0.001 ***
Orthographic	+0.375	< 0.001	< 0.001 ***
Pictorial	+0.381	< 0.001	0.003 **

Sensory modality On the Sensory Modality task the linear SVM is again degenerate. Its balanced accuracy equals the chance level exactly in both conditions

(0.333 ± 0.000 ; Table C.5), with no variance across the 15 Leave-One-Subject-Out folds (Table C.4). The per-class metrics (Table C.6) show that the orthographic class is never predicted (recall 0.000) in either condition, while trials are distributed over the audio and pictorial classes; in perception the audio class in particular absorbs almost every trial (recall 0.933). The classifier thus fails to recover the three-way structure, collapsing onto a subset of the classes, which for this balanced three-class problem again pins the balanced accuracy at chance. The comparison with the random forest is reported in Table 4.15. The random forest significantly outperforms the linear SVM in both conditions after Benjamini–Hochberg correction, but the margin is markedly asymmetric: it is large for perception (median $\Delta\text{BA} +0.268$) and small, though still significant, for imagination ($+0.016$). This mirrors the behaviour of the random forest itself, which achieved genuine above-chance discrimination only in the perception condition: where the random forest has little signal to exploit, as in imagination, the gap over the collapsed SVM narrows accordingly, whereas in perception the gap reflects the random forest’s clearly above-chance performance against a constant-subset prediction.

Table 4.15: Comparison of the two Pipeline 2 classifiers (random forest vs linear SVM) for the sensory modality task. For each condition a paired Wilcoxon signed-rank test is reported, summarised by the median per-session difference in balanced accuracy (ΔBA , random forest minus linear SVM). p -values are given raw and after Benjamini–Hochberg correction. Significance on the corrected value: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; n.s. = not significant.

Condition	Median ΔBA	p (raw)	p (BH)
Imagine	+0.016	0.018	0.033 *
Perception	+0.268	< 0.001	< 0.001 ***

Semantic Concept On the Semantic Concept task the linear SVM no longer collapses. Unlike the previous two tasks, the per-subject balanced accuracies vary genuinely across folds (Table C.7) and the per-class recalls are distributed fairly evenly over the three concepts, with no class systematically ignored (Table C.9). The classifier is therefore operating rather than degenerate, but it remains at chance in every condition after Benjamini–Hochberg correction (Table C.8); the only exception is the perception–pictorial condition (0.378 ± 0.055 , $p_{\text{BH}} = 0.032$), where it marginally exceeds chance. Its behaviour thus matches that of the random forest on this task, both classifiers failing to recover the semantic concept under the subject-independent regime. The comparison with the random forest is reported in Table 4.16. Consistent with both classifiers sitting at chance, the two are statistically indistinguishable in five of the six conditions, the median per-session

differences being negligible in magnitude. The sole significant difference, after Benjamini–Hochberg correction, is the perception–pictorial condition (median ΔBA -0.048), where the linear SVM slightly outperforms the random forest; this is the same condition in which the SVM marginally exceeded chance, and given its small size and isolated occurrence it carries no practical weight. On this task, neither of the two recovers the semantic concept from the tangent-space representation.

Table 4.16: Comparison of the two Pipeline 2 classifiers (random forest vs linear SVM) for the semantic concept task. For each condition a paired Wilcoxon signed-rank test is reported, summarised by the median per-session difference in balanced accuracy (ΔBA , random forest minus linear SVM). p -values are given raw and after Benjamini–Hochberg correction. Significance on the corrected value: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; n.s. = not significant.

Condition	Median ΔBA	p (raw)	p (BH)
Imagine-Audio	+0.004	0.762	0.838 n.s.
Imagine-Orthographic	−0.004	0.277	0.381 n.s.
Imagine-Pictorial	−0.011	0.073	0.115 n.s.
Perception-Audio	+0.021	0.934	0.934 n.s.
Perception-Orthographic	−0.014	0.489	0.597 n.s.
Perception-Pictorial	−0.048	0.007	0.015 *

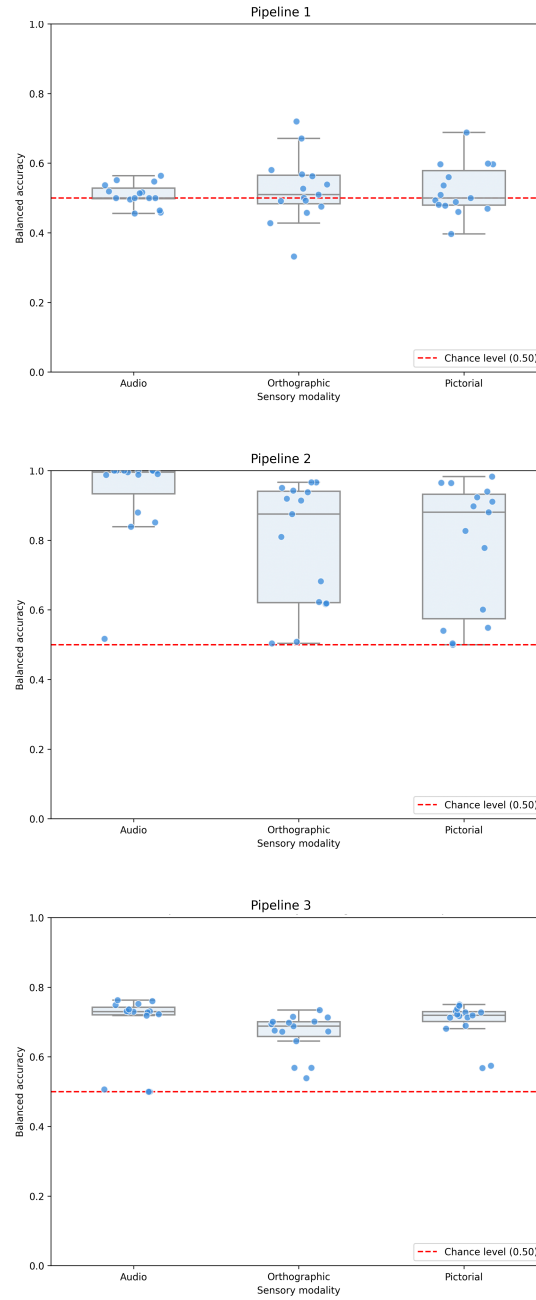


Figure 4.14: Per-session balanced accuracy for the perception–imagination task, by sensory modality, for the three principal pipelines (combined energy–attention rule for Pipeline 1, random forest for Pipeline 2, convolutional network for Pipeline 3). Boxplots summarise the distribution over the 15 Leave-One-Subject-Out folds (median, interquartile range and whiskers); the superimposed strip plot shows the individual per-session values (jittered horizontally). The dashed line marks the chance level (0.5).

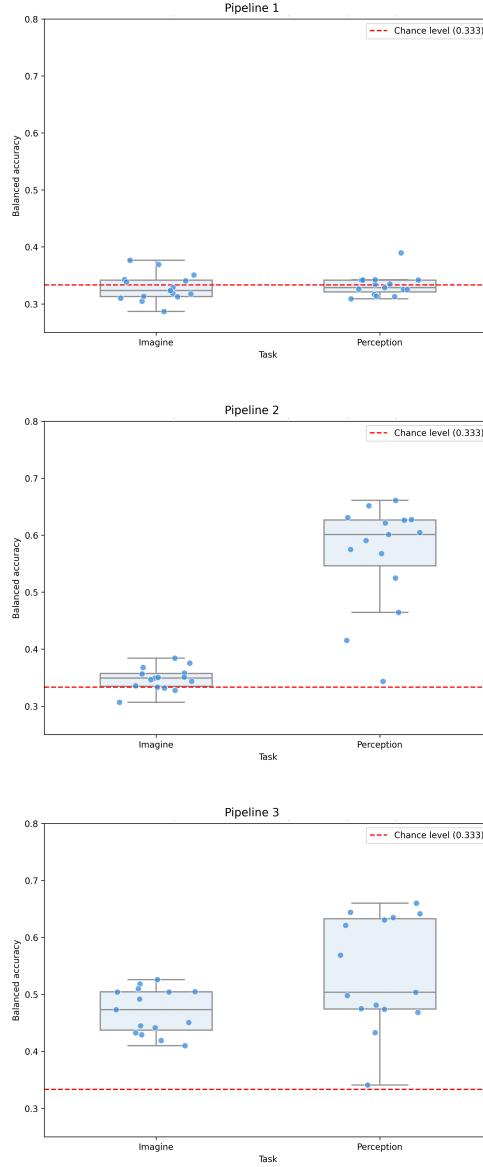


Figure 4.15: Per-session balanced accuracy for the Sensory Modality discrimination task, by task condition (imagination, perception), for the three principal pipelines (combined Energy–Attention rule for Pipeline 1, Random Forest for Pipeline 2, Convolutional Network for Pipeline 3). Boxplots summarise the distribution over the 15 Leave-One-Subject-Out folds (median, interquartile range and whiskers); the superimposed strip plot shows the individual per-session values (jittered horizontally). The dashed line marks the chance level (0.333).

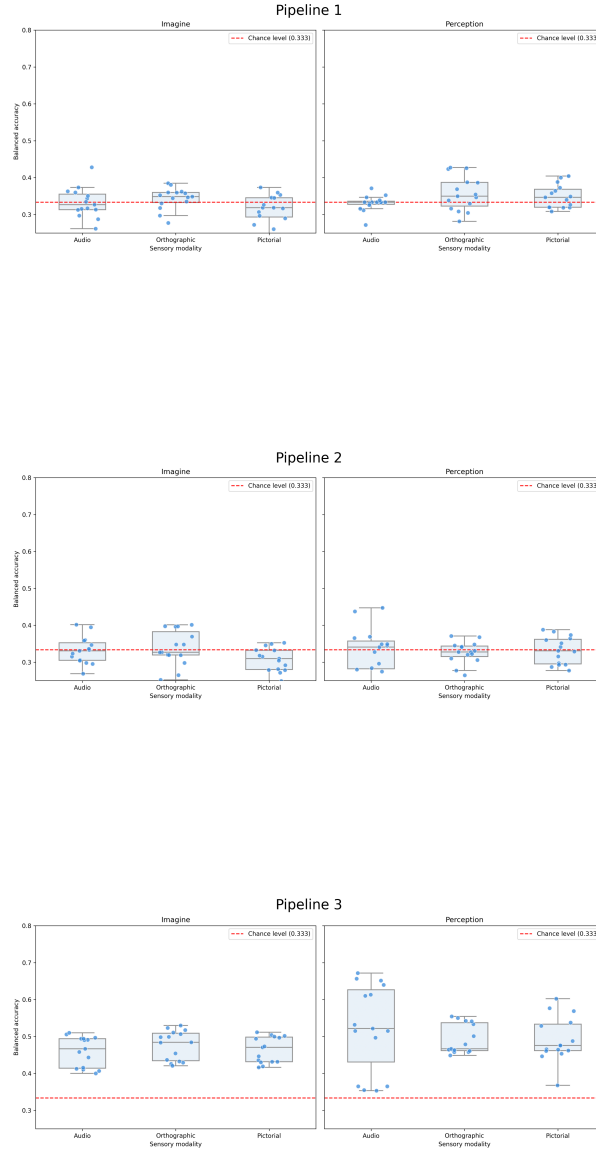


Figure 4.16: Per-session balanced accuracy for the Semantic Concept recognition task, by task–modality condition, for the three principal pipelines (combined energy–attention rule for Pipeline 1, random forest for Pipeline 2, convolutional network for Pipeline 3). Each panel shows the imagination and perception conditions across the three sensory modalities. Boxplots summarise the distribution over the 15 Leave-One-Subject-Out folds (median, interquartile range and whiskers); the superimposed strip plot shows the individual per-session values (jittered horizontally). The dashed line marks the chance level (0.333).

Chapter 5

Discussion

The three pipelines were evaluated on three decoding problems of increasing semantic depth (the Perception–Imagination distinction, the Sensory Modality of the stimulus, and its Semantic Concept) under a strictly subject-independent Leave-One-Subject-Out protocol, with balanced accuracy tested against chance by a one-sided Wilcoxon signed-rank test and Benjamini–Hochberg correction. Read together, the results draw a consistent picture: performance was governed jointly by the cognitive task and by the representation each pipeline extracts, and no single method dominated across all three problems.

Imagination VS Perception report On the Perception–Imagination task, the two data-driven pipelines decoded the distinction well above chance in every sensory modality, whereas the proposed Energy–Attention rule (Pipeline 1) never exceeded chance (Tables 4.1). The covariance-based random forest (Pipeline 2) was the strongest decoder, reaching 0.94 balanced accuracy for auditory stimuli and remaining clearly above chance for the orthographic and pictorial modalities (≈ 0.79); the Convolutional Network (Pipeline 3) was also robustly above chance and more uniform across modalities (≈ 0.67 – 0.70). The inter-pipeline comparison confirmed the ordering $P2 > P3 > P1$ in every modality, the advantage of Pipeline 2 over Pipeline 3 being largest for audio and no longer significant for the pictorial modality (Table 4.10). Both above-chance pipelines exhibited the same class asymmetry: near-perfect recall for imagination but substantially lower recall for perception, with most errors being perception trials misclassified as imagination.

5.1 Results discussion

Sensory Modality report In perception both data-driven pipelines discriminated the three modalities above chance and on a par with each other ($P2 \approx P3 >$

P1), whereas in imagination only the Convolutional Neural Network attained a genuine, if modest, above-chance level, the other pipelines sitting close to chance (Tables 4.2, 4.5 and 4.8). The two pipelines also differed in the structure of their success: Pipeline 2’s above-chance perception performance rested largely on the separability of the auditory class, the orthographic and pictorial classes being substantially confused, whereas Pipeline 3 recovered all three modalities more evenly.

Semantic Concept Report The semantic-concept task was the only one on which the three pipelines diverged completely. Neither the Energy–Attention rule nor the covariance-based random forest exceeded chance in any of the six task–modality conditions, while the Convolutional Neural Network exceeded chance in all six (Tables 4.3, 4.6 and 4.9). Pipeline 3 was therefore the only method to recover semantic information under the subject-independent regime, significantly outperforming both other pipelines in every condition ($P3 > P1 \approx P2$), with Pipelines 1 and 2 statistically indistinguishable from each other and from chance throughout (Table 4.12, Figure 4.16). The margins above chance were consistent but modest and, as the wide per-subject dispersion indicates, varied appreciably across held-out subjects.

Intra- Pipeline Comparisons Report Two secondary analyses are done. Within Pipeline 1, the attention weighting added no discriminative value over regional energy alone: the three decision rules were statistically indistinguishable on the modality and concept tasks, and in the only condition where the two rules differed significantly (auditory Perception–Imagination), adding the attention term slightly reduced balanced accuracy instead of improving it; as an isolated significant result at an uncorrected p close to 0.05, however, this difference should not be over-interpreted (Table 4.13, Appendix D). Within Pipeline 2, the choice of classifier was decisive: on the two easier tasks the linear SVM collapsed onto a constant (or near-constant) prediction, pinning balanced accuracy at exactly chance, so that the random forest’s clear advantage on these two tasks reflects the SVM’s degeneracy rather than a close contest (Tables 4.14 and 4.15). On the Semantic-Concept task the SVM operated without collapsing but, like the random forest, remained at chance in all but one of the six conditions; the sole exception (Perception–Pictorial, $p_{BH} = 0.032$) marginally exceeded chance, but as an isolated significant result among six comparisons it carries little weight and does not warrant a substantive interpretation (Tables 4.16 and C.8).

Summary of Principal Results Overall, three findings stand out and structure the discussion that follows. First, the Perception–Imagination distinction is the most readily decodable of the three targets, and is best captured by second-order

(covariance) features. Second, decoding performance degrades systematically as the target becomes more semantic and, for imagined content, more abstract. Third, only the convolutional network generalises across subjects to the semantic level: the covariance-based representation, despite its strength on the binary task, carries no semantic information under a subject-independent regime.

5.2 Research Questions

5.2.1 RQ1

Does the variance of the class-specific CSP component constitute a discriminative feature for subject-independent semantic EEG decoding, and does the proposed decorrelation measure carry additional discriminative value beyond uniform regional energy aggregation? Within the parameter-free decision rules adopted, the variance of the class-specific CSP component did not prove a discriminative feature under the subject-independent regime. Pipeline 1 failed to exceed chance on all three tasks (perception versus imagination, sensory modality and semantic concept) in every condition. The energy-only rule, which isolates this feature most directly, cleared chance in a single, marginal case (auditory perception–imagination) and remained at chance everywhere else. Aggregated across the anatomical regions by a simple sum, the class-specific CSP variance therefore does not carry, on the present evidence, information that transfers across subjects for any of the three problems. The second part of the question concerns whether the decorrelation measure proposed here adds value beyond uniform energy aggregation. It does not. Across the eight three-class conditions the three decision rules (energy-only, attention-only and their combination) are statistically indistinguishable, no Friedman omnibus test reaching significance; in the binary Perception–Imagination task the combined and energy-only rules differ significantly in a single modality (auditory), where the attention weighting slightly *lowers* the balanced accuracy rather than raising it (median $\Delta\text{BA} = -0.004$; Table 4.13). The hypothesis motivating the measure, “that regions in which the CSP decomposition is more reliable, that is, more decorrelated, are also more informative for classification”, is thus not supported: reweighting the regional energies by their decorrelation yielded no gain, and in the only case with a measurable effect a small loss. Two qualifications temper the strength of this negative conclusion. First, it is specific to the configuration tested. In keeping with the deliberately assumption-light design of Section 3.2, the CSP operated on the full preprocessed band and combined the regional descriptors through fixed, unweighted decision rules; a data-driven spectral selection, such as a filter-bank extension, or a trained aggregation could in principle recover discriminative structure that the present setup leaves untouched. The result therefore establishes that the class-specific CSP variance does not transfer across subjects under these minimal choices, not that it is intrinsically uninformative. Second, the statistical power available here is limited: with fifteen per-session folds drawn from twelve participants the tests are underpowered, and a non-significant difference cannot be read as proof of exact equivalence, absence of evidence not being evidence of absence. The appropriate reading is thus the cautious one: the decorrelation measure conferred no benefit in this study, and what little measurable

effect it had pointed in the adverse direction, but the evidence is not sufficient to rule out that a differently configured or better-powered analysis might yet extract value from it.

5.2.2 RQ2

Can Riemannian-geometric covariance features, an established representation in EEG signal processing, sustain classification performance across the three semantic decoding tasks under a strict LOSO regime?

Only in part. The tangent-space covariance representation sustained above-chance classification of the perception–imagination distinction across all three sensory modalities, but it did not extend to the remaining problems in comparable measure. Among the sensory-modality and semantic-concept conditions it exceeded chance genuinely only in the perception sensory-modality task, where the discrimination rested almost entirely on the separability of the auditory class; the imagination sensory-modality condition cleared chance only marginally and, on the confusion matrix, as a prediction bias rather than as true three-way separation, while the semantic concept remained at chance in every condition. The representation thus captures the cognitive-state contrast robustly, the coarse modality structure only partially and only in perception, and the semantic content not at all under the subject-independent regime. A methodological point is relevant to how these results should be read. The analysis window differs by condition (Table 3.1): imagination epochs last four seconds throughout, whereas perception epochs last two seconds for the auditory modality and three for the two visual ones. This asymmetry was retained deliberately rather than equalised. The onset of an imagined stimulus is not reliably time-locked to the cue: a participant may begin to form the mental image only after a variable delay, so that the discriminative activity need not fall within the first two or three seconds, and this latency is liable to grow with fatigue as the protocol is repeated. Truncating the imagination epochs to a fixed shorter window would therefore risk discarding precisely the interval in which the imagined representation becomes available. Because a covariance matrix has the same $C \times C$ dimension irrespective of the number of samples, epochs of unequal length pose no dimensional obstacle to Pipeline 2; what does depend on length is the stability of the estimate, a longer epoch yielding a better-conditioned covariance. The four-second imagination epochs are thus summarised by a covariance estimated from more samples than the shorter perception epochs, which offers one plausible contributor to the class asymmetry noted in Section 5, the imagination class, being estimated more stably, is the more reliably recovered of the two. To probe this length dependence directly, a control experiment was conducted on the Perception–Imagination task alone, in which the imagination epochs were

truncated to the duration of the corresponding perception epochs before entering Pipeline 2 (results reported in Appendix E). Its motivation was a specific tension in the results. In the intra-subject validation of Wilson et al. [18] and in the subject-independent study of Ahmadi et al. [29], the auditory modality was reported to be, on average, among the least accurately decoded; yet at native epoch lengths it was here the most accurately decoded of the three, raising the question of whether its apparent superiority reflected a genuine cognitive-state difference or the pronounced length asymmetry of the auditory condition, four seconds against two. Once the epochs were equalised the ordering changed markedly (Table E.2): the auditory modality fell to the lowest mean accuracy of the three, remaining above chance but by a reduced margin, whereas the two visual modalities stayed above 0.60 and well above chance; for the pictorial modality the decrease relative to the native-length configuration did not reach statistical significance (Table E.2). Two non-mutually-exclusive factors are consistent with this pattern. The first is the length dependence of the covariance estimate itself: equalising the epochs removes a systematic difference in estimation stability between the two classes, and the resulting drop indicates that part of the native-length separability, most acutely for audio, was tied to that difference rather than to cognitive state. The second is temporal: that the collapse is sharpest for audio, whose window was cut from four seconds to two, is consistent with much of the discriminative imagination activity arriving only after the first two seconds, in line with the delayed and variable onset of imagery discussed above; the visual modalities, truncated only to three seconds, retain more of that interval and are correspondingly less affected. Read together, these findings qualify the auditory result of Section 5: its near-perfect native-length accuracy should be regarded as substantially inflated by epoch length and treated with caution. They also, however, reinforce the conclusion for the visual modalities, which continue to separate perception from imagination well above chance even once the length confound is removed, indicating a cognitive-state signal that is not merely an artefact of window duration. Given the small number of participants, the relative contribution of the two factors cannot be established with certainty, and the control experiment is best read as evidence that epoch length is a non-negligible determinant of the covariance-based results.

5.2.3 RQ3

Given that topographic decoding has already demonstrated strong results on this dataset for the pictorial perception–imagination distinction[28], does this approach generalise effectively to the remaining tasks, namely sensory modality discrimination and semantic concept recognition? In a qualified sense, yes: of the three pipelines, the topographic one is the only representation that generalises to both remaining tasks. It exceeds chance in both

Sensory Modality conditions and, uniquely, in all six Semantic Concept conditions, the two handcrafted and covariance-based pipelines remaining at chance on the concept throughout. It is also the more consistent of the working pipelines across held-out subjects, its per-session accuracies being less dispersed than those of the covariance-based comparator. Within the space of methods examined here, topographic decoding is therefore the most promising strand for the harder tasks. The qualification matters, however: the margins above chance are modest, most acutely for the semantic concept, and vary appreciably from subject to subject. The approach generalises in the sense of clearing chance where the others do not (at least as implemented in the present work); this does not exclude that modifications to the pipeline might improve upon these margins. The reason this generalisation is better read as promising than as marginal is not one of accuracy alone, but of what the averaged representation implies about the underlying signal. As detailed in Section 3.4, Pipeline 3 does not operate on single trials but on a representative epoch obtained, for each subject and each class, by averaging that subject’s trials sample-by-sample, in direct analogy with the Event-Related Potential (Section 2.2.2). Such averaging improves the signal-to-noise ratio by a factor of roughly $\sqrt{N}$, but only under the condition that a stimulus-locked component is common to the averaged trials. The averaged per-subject representations are classifiable above chance. This would be unlikely if the averaged trials shared no stimulus-locked component, since in that case the average would tend towards zero. Their classifiability therefore suggests that such a common component exists within each subject; that a network trained on the representations of some subjects transfers to the held-out one suggests, further, that this component is sufficiently consistent *across* individuals to be recognised in a previously unseen subject. Read this way, the above-chance performance suggests that the stimuli elicit a spatially structured, stimulus-related pattern that is not merely subject-specific but repeatable across subjects, however weakly. This reading also anticipates a natural objection, namely that the averaging is a device contrived to ease the task and inflate the results. The evidence does not support that interpretation. Had the averaging trivialised the problem, one would expect strong accuracies to follow; instead the results remained modest in every condition, most notably for the semantic concept. The averaging thus appears to reveal a real but weak repeatable structure, rather than to create separability that is not there. Two caveats nonetheless bound the strength of the conclusion. First, the inference rests on twelve participants: whether the repeatable pattern would persist in a wholly different and larger cohort cannot be established from the present data, and the modest margins leave little room for optimism about robustness. Second, the representation is learned end-to-end, with the interpretive limitations discussed in Section 1.3: the network’s above-chance behaviour indicates that some transferable structure is captured, but not what that structure is, so the “repeatable pattern” remains an inference from the decoding outcome rather than

a localised electrophysiological finding. It cannot be excluded, in particular, that part of what the network exploits is a task-irrelevant regularity that happens to recur across subjects, inflating the reported accuracy without reflecting genuine semantic structure.

Chapter 6

Conclusion

This thesis has investigated the decoding of semantic representations from scalp EEG under a strictly subject-independent regime, addressing three classification problems of increasing semantic depth (the discrimination of perception from imagination, the identification of the sensory modality through which a stimulus is presented, and the recognition of the semantic concept) on the multimodal dataset of Wilson et al. Three decoding pipelines were developed and evaluated under a common Leave-One-Subject-Out protocol: a novel handcrafted approach grounded in an anatomical regional decomposition and region-wise Common Spatial Pattern filtering (Pipeline 1); a Riemannian-geometric framework operating on band-specific spatial covariances mapped to a tangent space (Pipeline 2); and a topographic approach in which each epoch is rendered as a sequence of scalp-voltage maps and classified by a Convolutional Neural Network (Pipeline 3). The work was motivated by two gaps in the literature surrounding this dataset: the near-exclusive reliance on opaque, end-to-end models, and the almost complete absence of subject-independent evaluation, most prior results having been reported within subject or on pooled data.

6.1 Summary of Findings and Contributions

The principal contribution of this work is less a single high-performing decoder than a coherent set of benchmark results, established under a rigorous evaluation regime across all three tasks, together with an interpretation of what those results reveal about the accessibility of semantic content at the scalp. Three principal findings can be grouped as follows. First, the Perception–Imagination distinction is the most readily decodable of the three targets. On the balanced accuracies alone, the Riemannian pipeline appears the representation best suited to separating the perceived from the imagined state, attaining the highest scores in every sensory

modality; the topographic network did so as well, whereas the handcrafted pipeline did not. This apparent primacy, however, may not come from the discriminative power of the methodology itself. The Riemannian pipeline involves several stages, and its advantage may be affected at more than one of them; One plausible interpretation is that the asymmetry in epoch length may affect the covariance estimates, and the auditory stimuli, by the design of the experimental protocol, are the most exposed to this asymmetry. The pipeline is therefore better seen as a promising route to the cognitive-state contrast than as clearly the strongest of the three. Second, decoding performance degrades systematically as the target becomes more semantic and, for imagined content, more abstract, the sensory modality being recoverable only partially and the semantic concept only marginally. Third, of the three method families examined, only the topographic convolutional network showed a weak but consistent cross-subject generalisation to the semantic level, recovering the concept above chance in every condition in which both the covariance-based and the handcrafted representations remained at chance throughout. For the answers to the research questions, no clear-cut yes-or-no emerged. For **RQ1**, in fact, within the deliberately assumption-light and parameter-free configuration adopted, the variance of the class-specific CSP component did not constitute a feature that transfers across subjects for any of the three tasks, and the proposed regional decorrelation measure conferred no discriminative value beyond a uniform aggregation of regional energy. In the single condition in which its effect was measurable, it slightly lowered accuracy rather than raising it. This negative result is itself informative: it delimits what a minimally-configured, interpretable regional CSP can and cannot achieve on this problem. It should also be noted that the proposed decorrelation measure was assessed solely in terms of classification accuracy. This does not rule out that it may hold some useful or neurophysiologically meaningful role, which was not investigated here as it fell outside the main scope of this work. For **RQ2**, the tangent-space covariance representation sustained classification only in part: robustly for the cognitive-state contrast, partially and only in perception for the sensory modality, where discrimination rested largely on the separability of the auditory class, and not at all for the semantic concept. The Riemannian pipeline comprises several stages, so it is hard to pinpoint where its advantage originates. A control experiment on different epoch lengths suggests that the temporal asymmetry may play a role, the near-perfect auditory result on the perception–imagination task dropping once the epochs were equalised. This need not be the only cause: shortening the imagined epochs may also remove genuinely discriminative information, since for imagined stimuli the onset is not guaranteed to coincide with the protocol cue and much of the stimulus-related activity may appear only after the first two seconds. For the visual stimuli this drop was less marked, since only one second was removed and any late stimulus-related activity is better preserved. Within the experimental setup adopted in this work, the Riemannian

pipeline appears a promising route to the imagination versus perception distinction. For **RQ3**, topographic decoding was, in a qualified sense, the only approach to generalise to the harder tasks; the margins above chance were nonetheless modest, most acutely for the semantic concept, and, having been obtained from epochs averaged per subject and per class, they indicate a stimulus-locked spatial pattern that is weak but repeatable across individuals rather than merely subject-specific.

6.2 Limitation & Future perspective

Limitations Several limitations affect these conclusions and should be kept in mind when reading them. The most pervasive is the size of the cohort: with twelve participants and fifteen sessions the statistical tests are underpowered, so that a non-significant difference cannot be taken as evidence of equivalence, and the modest, widely dispersed margins of the topographic pipeline on the semantic-concept task leave its robustness uncertain. The most pervasive is the size of the cohort: with twelve participants and fifteen sessions the statistical tests are underpowered, so that a non-significant difference cannot be taken as evidence of equivalence, and the modest, widely dispersed margins of the topographic pipeline on the semantic-concept task leave its robustness uncertain. The negative result of Pipeline 1 is, moreover, specific to the configuration tested, namely a full-band CSP combined through fixed, unweighted decision rules, and establishes that the class-specific CSP variance does not transfer under these minimal choices, not that it is intrinsically uninformative. The two comparators are likewise somewhat weakened realisations of their respective method families: the tangent-space representation of Pipeline 2 is very high-dimensional relative to the number of available trials, a disproportion left unaddressed by any dimensionality reduction, while the convolutional network of Pipeline 3 is deliberately small and simple and operates on low-resolution maps, both choices dictated by the modest computational platform, and does not reproduce the architecture from which it is adapted. A more fundamental qualification concerns the object on which Pipeline 3 operates: unlike the other two pipelines, which classify single trials, it is trained and evaluated on representative epochs formed by averaging each subject’s trials within a class. This averaging gives a denoising advantage that the other pipelines do not have and, in relying on the class labels to construct those representative epochs, addresses a comparatively easier version of the problem; its above-chance semantic performance therefore attests to a repeatable averaged structure, and not to semantic decoding at the level of the individual trial.

Future directions These limitations map directly onto the natural directions for future work. For the handcrafted pipeline, the assumption-light choices that

most plausibly suppressed its performance could be relaxed: a data-driven spectral selection, such as a filter-bank extension restricting the CSP to the bands carrying the discriminative rhythm, together with a trained rather than parameter-free aggregation of the regional descriptors, might recover structure that the present configuration leaves untouched, while preserving the anatomical interpretability that motivated the approach. For the Riemannian pipeline, an unsupervised dimensionality reduction of the tangent vectors, principal component analysis being the obvious candidate, would address the imbalance between feature count and sample size, and represents a natural next step. Further directions include analysing the covariance of each individual band, since taking all bands together may have penalised the most informative ones on the more specific tasks, as well as exploring other approaches such as different covariance estimation methods and different types of Riemannian alignment. A further direction is to segment the signals into one-second windows, which would yield epochs of uniform duration but, on the other hand, produce an imbalanced dataset and therefore require balanced accuracy for any evaluation. The topographic pipeline offers the widest scope for extension, and here the very simplicity of the present network is the most immediate lever. Treating each topographic map as an independent frame discards the temporal ordering of the sequence; a three-dimensional convolutional network, operating on the topomap sequence as a single spatio-temporal volume, would instead exploit that ordering directly, and coupling such a stage with an attention-based or Transformer module could capture the longer-range temporal dependencies across frames that a shallow, frame-wise network cannot. Rendering the maps at full resolution and, more ambitiously, extending the topographic approach from averaged to single-trial epochs would further improve both the fidelity and the interpretation of the result. Underlying all of these, replication on a larger and independent cohort remains the essential prerequisite: it is the only means of establishing whether the weak, repeatable semantic structure identified here persists beyond the twelve participants on which it currently rests.

6.3 Concluding remarks

The results of this work have shown that good performance can be achieved even in a LOSO setting without necessarily using end-to-end models: the Riemannian pipeline, working on covariance features, proved promising for distinguishing the cognitive state of perception from that of imagination, within the limitations imposed by the size of this kind of dataset, even though the same cannot be said for the handcrafted model proposed here. As the target moves from this cognitive-state contrast towards more semantic distinctions, however, decoding becomes progressively harder, and the conceptual level itself is recoverable only weakly, and

only through a learned topographic representation whose success relies on averaging the signal across trials, which removes the trial-by-trial variability. Indeed, given the intra-class variability built into the dataset, obtaining excellent results even on the averaged epochs remains an open challenge. The modest results obtained at the conceptual level under the LOSO evaluation protocol thus provide a benchmark against which future, more powerful approaches may be measured. Even when such performance is reached, however, there is no guarantee that it reflects a genuine shared feature, as it may simply come from noise that has propagated through the data in a systematic and repeatable way. For this reason, the author suggests pursuing interpretable handcrafted models, applied to the averaged epochs as well, in order to look for repeatable patterns of genuine neurophysiological meaning.

Appendix A

Appendix

A.1 Pipeline 1 materials

Table A.1: Per-subject balanced accuracy for the perception–imagination task, by sensory modality (one value per Leave-One-Subject-Out fold).

Subject	Audio	Orthographic	Pictorial
3_3	0.458	0.491	0.688
8_3	0.537	0.500	0.493
10_1	0.496	0.720	0.489
11_1	0.500	0.492	0.396
12_1	0.564	0.671	0.597
12_2	0.516	0.568	0.500
13_1	0.500	0.510	0.470
14_1	0.513	0.476	0.597
14_2	0.519	0.457	0.481
15_1	0.455	0.428	0.536
15_2	0.551	0.527	0.560
16_1	0.464	0.581	0.478
17_1	0.500	0.332	0.460
18_1	0.547	0.562	0.509
19_1	0.500	0.538	0.599

Table A.2: Per-class precision, recall and F1 (mean $\pm$ std over 15 folds) for the perception–imagination task, by sensory modality.

Modality	Class	P	R	F1
Audio	Imagine	0.398 ± 0.322	0.321 ± 0.329	0.294 ± 0.256
	Perception	0.538 ± 0.127	0.695 ± 0.329	0.545 ± 0.157
Orthographic	Imagine	0.553 ± 0.144	0.557 ± 0.332	0.481 ± 0.211
	Perception	0.492 ± 0.168	0.490 ± 0.366	0.429 ± 0.252
Pictorial	Imagine	0.519 ± 0.061	0.641 ± 0.296	0.534 ± 0.182
	Perception	0.460 ± 0.244	0.406 ± 0.297	0.396 ± 0.235

Table A.3: Per-subject balanced accuracy for the sensory-modality discrimination task, by task (one value per Leave-One-Subject-Out fold).

Subject	Imagine	Perception
3_3	0.343	0.390
8_3	0.318	0.335
10_1	0.313	0.326
11_1	0.377	0.328
12_1	0.305	0.309
12_2	0.313	0.325
13_1	0.329	0.341
14_1	0.310	0.334
14_2	0.351	0.342
15_1	0.369	0.313
15_2	0.319	0.325
16_1	0.324	0.342
17_1	0.287	0.317
18_1	0.338	0.342
19_1	0.341	0.314

Table A.4: Per-class precision, recall and F1 (mean $\pm$ std over 15 folds) for the sensory-modality discrimination task.

Task	Class	P	R	F1
Imagine	Audio	0.308 ± 0.058	0.388 ± 0.316	0.290 ± 0.162
	Orthographic	0.244 ± 0.164	0.120 ± 0.189	0.124 ± 0.144
	Pictorial	0.356 ± 0.137	0.479 ± 0.318	0.344 ± 0.145
Perception	Audio	0.314 ± 0.067	0.486 ± 0.295	0.349 ± 0.156
	Orthographic	0.298 ± 0.104	0.270 ± 0.271	0.216 ± 0.169
	Pictorial	0.341 ± 0.078	0.241 ± 0.166	0.253 ± 0.123

Table A.5: Per-subject balanced accuracy for the semantic-concept recognition task, by task–modality condition (one value per Leave-One-Subject-Out fold).

Subject	Imagine			Perception		
	Audio	Orthographic	Pictorial	Audio	Orthographic	Pictorial
3_3	0.288	0.362	0.261	0.312	0.426	0.364
8_3	0.360	0.360	0.317	0.333	0.330	0.320
10_1	0.316	0.297	0.319	0.333	0.282	0.400
11_1	0.363	0.278	0.307	0.333	0.308	0.349
12_1	0.344	0.360	0.353	0.333	0.423	0.358
12_2	0.297	0.385	0.239	0.353	0.386	0.319
13_1	0.317	0.347	0.346	0.333	0.347	0.320
14_1	0.335	0.330	0.290	0.272	0.316	0.319
14_2	0.313	0.318	0.345	0.316	0.339	0.308
15_1	0.262	0.358	0.360	0.330	0.388	0.326
15_2	0.350	0.353	0.297	0.347	0.427	0.404
16_1	0.313	0.344	0.318	0.326	0.354	0.347
17_1	0.327	0.335	0.373	0.333	0.305	0.340
18_1	0.373	0.381	0.272	0.339	0.350	0.388
19_1	0.428	0.349	0.326	0.371	0.369	0.373

Table A.6: Per-class precision, recall and F1 (mean $\pm$ std over 15 folds) for the semantic-concept recognition task, by task–modality condition.

Modality	Stimulus	Class	P	R	F1
Imagine	Audio	Flower	0.370 ± 0.138	0.312 ± 0.223	0.281 ± 0.131
		Penguin	0.335 ± 0.257	0.203 ± 0.189	0.205 ± 0.149
		Guitar	0.373 ± 0.195	0.482 ± 0.298	0.340 ± 0.160
Imagine	Orthographic	Flower	0.430 ± 0.167	0.424 ± 0.310	0.326 ± 0.163
		Penguin	0.341 ± 0.070	0.511 ± 0.316	0.360 ± 0.140
		Guitar	0.277 ± 0.288	0.096 ± 0.154	0.107 ± 0.142
Imagine	Pictorial	Flower	0.269 ± 0.126	0.378 ± 0.288	0.286 ± 0.166
		Penguin	0.266 ± 0.146	0.268 ± 0.250	0.228 ± 0.161
		Guitar	0.232 ± 0.146	0.299 ± 0.325	0.227 ± 0.197
Perception	Audio	Flower	0.138 ± 0.170	0.113 ± 0.180	0.113 ± 0.148
		Penguin	0.329 ± 0.029	0.694 ± 0.335	0.413 ± 0.127
		Guitar	0.173 ± 0.165	0.187 ± 0.233	0.162 ± 0.180
Perception	Orthographic	Flower	0.294 ± 0.155	0.297 ± 0.295	0.245 ± 0.187
		Penguin	0.386 ± 0.090	0.598 ± 0.239	0.430 ± 0.081
		Guitar	0.295 ± 0.267	0.175 ± 0.228	0.178 ± 0.170
Perception	Pictorial	Flower	0.345 ± 0.044	0.482 ± 0.287	0.364 ± 0.141
		Penguin	0.365 ± 0.193	0.200 ± 0.182	0.216 ± 0.157
		Guitar	0.378 ± 0.180	0.365 ± 0.259	0.306 ± 0.126

A.2 Pipeline 2 materials

Table A.7: Per-subject balanced accuracy for the perception–imagination task, by sensory modality (one value per Leave-One-Subject-Out fold).

Subject	Audio	Orthographic	Pictorial
3_3	0.988	0.810	0.827
8_3	1.000	0.966	0.923
10_1	0.851	0.616	0.540
11_1	1.000	0.875	0.898
12_1	0.517	0.508	0.500
12_2	0.996	0.938	0.881
13_1	1.000	0.966	0.983
14_1	1.000	0.914	0.940
14_2	1.000	0.682	0.778
15_1	1.000	0.942	0.965
15_2	0.879	0.619	0.601
16_1	0.988	0.919	0.911
17_1	0.839	0.503	0.503
18_1	1.000	0.950	0.964
19_1	0.990	0.623	0.549

Table A.8: Per-class precision, recall and F1 (mean $\pm$ std over 15 folds) for the perception–imagination task, by sensory modality.

Modality	Class	P	R	F1
Audio	Imagine	0.916 ± 0.141	1.000 ± 0.000	0.950 ± 0.090
	Perception	1.000 ± 0.000	0.873 ± 0.251	0.902 ± 0.233
Orthographic	Imagine	0.805 ± 0.201	0.909 ± 0.162	0.820 ± 0.133
	Perception	0.930 ± 0.102	0.669 ± 0.382	0.683 ± 0.333
Pictorial	Imagine	0.793 ± 0.202	0.935 ± 0.115	0.832 ± 0.117
	Perception	0.880 ± 0.249	0.634 ± 0.406	0.649 ± 0.374

Table A.9: Per-subject balanced accuracy for the sensory-modality discrimination task, by task (one value per Leave-One-Subject-Out fold).

Subject	Imagine	Perception
3_3	0.384	0.591
8_3	0.357	0.605
10_1	0.333	0.525
11_1	0.358	0.601
12_1	0.349	0.343
12_2	0.328	0.661
13_1	0.307	0.631
14_1	0.350	0.622
14_2	0.347	0.415
15_1	0.376	0.627
15_2	0.344	0.575
16_1	0.368	0.626
17_1	0.336	0.465
18_1	0.332	0.652
19_1	0.351	0.568

Table A.10: Per-class precision, recall and F1 (mean $\pm$ std over 15 folds) for the sensory-modality discrimination task.

Task	Class	P	R	F1
Imagine	Audio	0.359 ± 0.032	0.512 ± 0.123	0.413 ± 0.038
	Orthographic	0.371 ± 0.085	0.103 ± 0.035	0.155 ± 0.041
	Pictorial	0.351 ± 0.037	0.429 ± 0.126	0.373 ± 0.057
Perception	Audio	0.859 ± 0.162	0.819 ± 0.276	0.779 ± 0.222
	Orthographic	0.471 ± 0.124	0.234 ± 0.098	0.297 ± 0.102
	Pictorial	0.468 ± 0.061	0.649 ± 0.182	0.523 ± 0.076

Table A.11: Per-subject balanced accuracy for the semantic-concept recognition task, by task–modality condition (one value per Leave-One-Subject-Out fold).

Subject	Imagine			Perception		
	Audio	Orthographic	Pictorial	Audio	Orthographic	Pictorial
3_3	0.333	0.397	0.332	0.447	0.345	0.329
8_3	0.360	0.327	0.281	0.296	0.342	0.293
10_1	0.306	0.369	0.333	0.345	0.331	0.315
11_1	0.298	0.252	0.350	0.284	0.264	0.277
12_1	0.336	0.319	0.249	0.280	0.277	0.341
12_2	0.315	0.265	0.271	0.369	0.327	0.351
13_1	0.331	0.326	0.310	0.341	0.310	0.364
14_1	0.304	0.396	0.353	0.328	0.348	0.383
14_2	0.269	0.298	0.292	0.275	0.341	0.297
15_1	0.323	0.397	0.304	0.366	0.320	0.360
15_2	0.395	0.320	0.279	0.220	0.328	0.388
16_1	0.358	0.348	0.346	0.349	0.322	0.287
17_1	0.347	0.401	0.315	0.349	0.306	0.373
18_1	0.295	0.348	0.318	0.437	0.368	0.293
19_1	0.402	0.320	0.278	0.222	0.371	0.331

Table A.12: Per-class precision, recall and F1 (mean $\pm$ std over 15 folds) for the semantic-concept recognition task, by task–modality condition.

Modality	Stimulus	Class	P	R	F1
Imagine	Audio	Flower	0.344 ± 0.056	0.328 ± 0.060	0.333 ± 0.047
		Penguin	0.320 ± 0.045	0.448 ± 0.060	0.373 ± 0.049
		Guitar	0.346 ± 0.099	0.219 ± 0.069	0.266 ± 0.077
Imagine	Orthographic	Flower	0.335 ± 0.063	0.424 ± 0.136	0.371 ± 0.092
		Penguin	0.351 ± 0.076	0.304 ± 0.081	0.319 ± 0.063
		Guitar	0.328 ± 0.070	0.289 ± 0.084	0.301 ± 0.056
Imagine	Pictorial	Flower	0.316 ± 0.034	0.434 ± 0.101	0.362 ± 0.054
		Penguin	0.281 ± 0.077	0.205 ± 0.092	0.231 ± 0.081
		Guitar	0.300 ± 0.048	0.284 ± 0.083	0.288 ± 0.061
Perception	Audio	Flower	0.339 ± 0.069	0.459 ± 0.121	0.387 ± 0.084
		Penguin	0.311 ± 0.066	0.354 ± 0.111	0.327 ± 0.077
		Guitar	0.332 ± 0.117	0.168 ± 0.090	0.215 ± 0.094
Perception	Orthographic	Flower	0.324 ± 0.029	0.405 ± 0.107	0.356 ± 0.059
		Penguin	0.302 ± 0.068	0.232 ± 0.083	0.257 ± 0.072
		Guitar	0.342 ± 0.062	0.343 ± 0.065	0.339 ± 0.057
Perception	Pictorial	Flower	0.323 ± 0.051	0.411 ± 0.110	0.358 ± 0.066
		Penguin	0.351 ± 0.082	0.215 ± 0.071	0.261 ± 0.069
		Guitar	0.335 ± 0.056	0.371 ± 0.083	0.348 ± 0.059

A.3 Pipeline 3 materials

Table A.13: Per-subject balanced accuracy for the perception–imagination task, by sensory modality (one value per Leave-One-Subject-Out fold).

Subject	Audio	Orthographic	Pictorial
3_3	0.752	0.672	0.728
8_3	0.730	0.569	0.681
10_1	0.731	0.715	0.731
11_1	0.730	0.568	0.717
12_1	0.731	0.694	0.712
12_2	0.736	0.700	0.722
13_1	0.749	0.734	0.574
14_1	0.728	0.713	0.713
14_2	0.500	0.539	0.690
15_1	0.719	0.688	0.750
15_2	0.722	0.701	0.738
16_1	0.500	0.673	0.728
17_1	0.762	0.698	0.747
18_1	0.506	0.645	0.568
19_1	0.760	0.675	0.719

Table A.14: Per-class precision, recall and F1 (mean $\pm$ std over 15 folds) for the perception–imagination task, by sensory modality.

Modality	Class	P	R	F1
Audio	Imagination	0.780 ± 0.057	0.926 ± 0.075	0.843 ± 0.033
	Perception	0.682 ± 0.292	0.455 ± 0.232	0.521 ± 0.256
Orthographic	Imagination	0.692 ± 0.047	0.846 ± 0.052	0.759 ± 0.028
	Perception	0.699 ± 0.069	0.485 ± 0.139	0.562 ± 0.121
Pictorial	Imagination	0.720 ± 0.046	0.853 ± 0.036	0.780 ± 0.027
	Perception	0.733 ± 0.060	0.549 ± 0.121	0.622 ± 0.101

Table A.15: Per-subject balanced accuracy for the sensory-modality discrimination task, by task (one value per Leave-One-Subject-Out fold).

Subject	Imagine	Perception
3_3	0.429	0.481
8_3	0.518	0.660
10_1	0.442	0.475
11_1	0.451	0.569
12_1	0.419	0.631
12_2	0.410	0.635
13_1	0.504	0.504
14_1	0.504	0.433
14_2	0.492	0.468
15_1	0.510	0.498
15_2	0.526	0.642
16_1	0.473	0.341
17_1	0.505	0.644
18_1	0.433	0.474
19_1	0.445	0.621

Table A.16: Per-class precision, recall and F1 (mean $\pm$ std over 15 folds) for the sensory-modality discrimination task.

Task	Class	P	R	F1
Imagine	Audio	0.455 ± 0.050	0.500 ± 0.137	0.468 ± 0.084
	Orthographic	0.474 ± 0.069	0.442 ± 0.110	0.454 ± 0.090
	Pictorial	0.484 ± 0.077	0.470 ± 0.136	0.461 ± 0.112
Perception	Audio	0.484 ± 0.221	0.334 ± 0.279	0.367 ± 0.269
	Orthographic	0.579 ± 0.091	0.623 ± 0.185	0.583 ± 0.140
	Pictorial	0.567 ± 0.100	0.657 ± 0.112	0.596 ± 0.050

Table A.17: Per-subject balanced accuracy for the semantic-concept recognition task, by task–modality condition (one value per Leave-One-Subject-Out fold).

Subject	Imagine			Perception		
	Audio	Orthographic	Pictorial	Audio	Orthographic	Pictorial
3_3	0.400	0.523	0.436	0.610	0.464	0.464
8_3	0.506	0.499	0.447	0.353	0.449	0.529
10_1	0.411	0.454	0.430	0.355	0.554	0.466
11_1	0.458	0.518	0.502	0.365	0.462	0.453
12_1	0.510	0.429	0.471	0.365	0.467	0.602
12_2	0.416	0.530	0.473	0.532	0.501	0.569
13_1	0.491	0.511	0.419	0.652	0.533	0.462
14_1	0.413	0.432	0.432	0.515	0.541	0.368
14_2	0.490	0.498	0.497	0.522	0.458	0.461
15_1	0.467	0.507	0.417	0.672	0.479	0.447
15_2	0.494	0.426	0.512	0.613	0.458	0.538
16_1	0.494	0.437	0.494	0.657	0.461	0.488
17_1	0.497	0.421	0.504	0.640	0.542	0.577
18_1	0.443	0.484	0.432	0.497	0.463	0.477
19_1	0.407	0.483	0.500	0.515	0.550	0.476

Table A.18: Per-class precision, recall and F1 (mean $\pm$ std over 15 folds) for the semantic-concept recognition task, by task–modality condition.

Modality	Stimulus	Class	P	R	F1
Imagine	Audio	Flower	0.464 ± 0.056	0.493 ± 0.078	0.471 ± 0.034
		Penguin	0.475 ± 0.059	0.469 ± 0.122	0.465 ± 0.081
		Guitar	0.438 ± 0.069	0.417 ± 0.148	0.421 ± 0.115
Imagine	Orthographic	Flower	0.475 ± 0.063	0.504 ± 0.103	0.482 ± 0.065
		Penguin	0.485 ± 0.049	0.510 ± 0.130	0.490 ± 0.088
		Guitar	0.464 ± 0.069	0.417 ± 0.154	0.425 ± 0.126
Imagine	Pictorial	Flower	0.462 ± 0.052	0.558 ± 0.133	0.496 ± 0.050
		Penguin	0.488 ± 0.088	0.419 ± 0.119	0.442 ± 0.089
		Guitar	0.426 ± 0.129	0.416 ± 0.164	0.414 ± 0.143
Perception	Audio	Flower	0.466 ± 0.206	0.476 ± 0.223	0.464 ± 0.203
		Penguin	0.501 ± 0.169	0.539 ± 0.204	0.514 ± 0.182
		Guitar	0.608 ± 0.140	0.557 ± 0.173	0.575 ± 0.155
Perception	Orthographic	Flower	0.539 ± 0.095	0.504 ± 0.075	0.512 ± 0.051
		Penguin	0.491 ± 0.117	0.394 ± 0.153	0.423 ± 0.120
		Guitar	0.477 ± 0.063	0.578 ± 0.155	0.512 ± 0.085
Perception	Pictorial	Flower	0.456 ± 0.108	0.438 ± 0.149	0.440 ± 0.125
		Penguin	0.485 ± 0.069	0.544 ± 0.188	0.495 ± 0.125
		Guitar	0.491 ± 0.166	0.493 ± 0.196	0.487 ± 0.171

Appendix B

Appendix

B.1 Pipeline 1 Only Energy

B.1.1 Imagination VS Perception

Table B.1: Per-subject balanced accuracy for the perception–imagination task, energy rule, by sensory modality (one value per Leave-One-Subject-Out fold).

Subject	Audio	Orthographic	Pictorial
3_3	0.527	0.501	0.759
8_3	0.520	0.493	0.423
10_1	0.500	0.686	0.407
11_1	0.500	0.500	0.422
12_1	0.650	0.653	0.643
12_2	0.537	0.568	0.523
13_1	0.503	0.513	0.470
14_1	0.524	0.429	0.600
14_2	0.509	0.484	0.498
15_1	0.443	0.370	0.535
15_2	0.585	0.516	0.545
16_1	0.483	0.576	0.478
17_1	0.503	0.312	0.470
18_1	0.574	0.552	0.532
19_1	0.500	0.606	0.604

Table B.2: Balanced accuracy (mean $\pm$ std over 15 folds) for the perception–imagination task, by sensory modality, with the one-sided Wilcoxon signed-rank test against chance (0.5): raw p -value and Benjamini–Hochberg corrected p -value. Significance is assessed on the corrected value only: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; n.s. = not significant.

Modality	Balanced accuracy	p (raw)	p (BH)
Audio	0.524 ± 0.047	0.021	0.044 *
Orthographic	0.517 ± 0.095	0.198	0.270 n.s.
Pictorial	0.527 ± 0.091	0.165	0.248 n.s.

Table B.3: Per-class precision, recall and F1 (mean $\pm$ std over 15 folds) for the Perception–Imagination task, energy rule, by sensory modality.

Modality	Class	P	R	F1
Audio	Imagine	0.543 ± 0.336	0.327 ± 0.325	0.310 ± 0.263
	Perception	0.551 ± 0.126	0.721 ± 0.308	0.570 ± 0.137
Orthographic	Imagine	0.572 ± 0.166	0.528 ± 0.324	0.467 ± 0.197
	Perception	0.462 ± 0.173	0.507 ± 0.386	0.428 ± 0.264
Pictorial	Imagine	0.547 ± 0.102	0.647 ± 0.273	0.545 ± 0.167
	Perception	0.467 ± 0.231	0.407 ± 0.307	0.402 ± 0.240

B.1.2 Sensory Modality

Table B.4: Per-subject balanced accuracy for the Sensory Modality discrimination task, energy rule, by task (one value per Leave-One-Subject-Out fold).

Subject	Imagine	Perception
3_3	0.419	0.365
8_3	0.331	0.330
10_1	0.324	0.348
11_1	0.318	0.325
12_1	0.315	0.313
12_2	0.291	0.304
13_1	0.331	0.311
14_1	0.329	0.334
14_2	0.362	0.283
15_1	0.329	0.315
15_2	0.291	0.338
16_1	0.321	0.371
17_1	0.319	0.371
18_1	0.324	0.333
19_1	0.366	0.317

Table B.5: Balanced accuracy (mean $\pm$ std over 15 folds) for the Sensory Modality discrimination task, by task, with the one-sided Wilcoxon signed-rank test against chance (0.333): raw p -value and Benjamini–Hochberg corrected p -value. Significance is assessed on the corrected value only: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; n.s. = not significant.

Task	Balanced accuracy	p (raw)	p (BH)
Imagine	0.331 ± 0.030	0.896	1.000 n.s.
Perception	0.331 ± 0.024	0.681	1.000 n.s.

Table B.6: Per-class precision, recall and F1 (mean $\pm$ std over 15 folds) for the Sensory Modality discrimination task, energy rule.

Task	Class	P	R	F1
Imagine	Audio	0.330 ± 0.149	0.306 ± 0.265	0.265 ± 0.143
	Orthographic	0.206 ± 0.172	0.173 ± 0.252	0.146 ± 0.178
	Pictorial	0.317 ± 0.059	0.515 ± 0.331	0.349 ± 0.159
Perception	Audio	0.326 ± 0.094	0.488 ± 0.324	0.339 ± 0.168
	Orthographic	0.298 ± 0.215	0.270 ± 0.298	0.206 ± 0.173
	Pictorial	0.276 ± 0.143	0.234 ± 0.234	0.224 ± 0.160

B.1.3 Semantic Concept

Table B.7: Per-subject balanced accuracy for the Semantic Concept recognition task, energy rule, by task–modality condition (one value per Leave-One-Subject-Out fold).

Subject	Imagine			Perception		
	Audio	Orthographic	Pictorial	Audio	Orthographic	Pictorial
3_3	0.324	0.365	0.291	0.333	0.326	0.344
8_3	0.340	0.333	0.357	0.333	0.325	0.313
10_1	0.351	0.337	0.304	0.333	0.343	0.388
11_1	0.343	0.323	0.309	0.347	0.300	0.336
12_1	0.309	0.343	0.333	0.333	0.349	0.275
12_2	0.288	0.405	0.241	0.319	0.345	0.370
13_1	0.328	0.347	0.359	0.327	0.340	0.261
14_1	0.327	0.302	0.310	0.308	0.360	0.362
14_2	0.347	0.345	0.309	0.303	0.317	0.310
15_1	0.271	0.302	0.358	0.320	0.326	0.307
15_2	0.305	0.365	0.339	0.332	0.322	0.366
16_1	0.363	0.362	0.330	0.318	0.372	0.362
17_1	0.300	0.344	0.360	0.333	0.326	0.306
18_1	0.356	0.317	0.236	0.296	0.350	0.363
19_1	0.355	0.294	0.298	0.345	0.348	0.425

Table B.8: Balanced accuracy (mean $\pm$ std over 15 folds) for the Semantic Concept recognition task, by task–modality condition, with the one-sided Wilcoxon signed-rank test against chance (0.333): raw p -value and Benjamini–Hochberg corrected p -value. Significance is assessed on the corrected value only: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; n.s. = not significant.

Modality	Stimulus	Balanced accuracy	p (raw)	p (BH)
Imagine	Audio	0.327 ± 0.026	0.738	0.857 n.s.
Imagine	Orthographic	0.339 ± 0.028	0.255	0.574 n.s.
Imagine	Pictorial	0.316 ± 0.038	0.924	0.997 n.s.
Perception	Audio	0.325 ± 0.014	0.982	0.997 n.s.
Perception	Orthographic	0.337 ± 0.018	0.227	0.545 n.s.
Perception	Pictorial	0.339 ± 0.042	0.211	0.542 n.s.

Table B.9: Per-class precision, recall and F1 (mean $\pm$ std over 15 folds) for the Semantic Concept recognition task, energy rule, by task-modality condition.

Modality	Stimulus	Class	P	R	F1
Imagine	Audio	Flower	0.425 ± 0.262	0.286 ± 0.254	0.248 ± 0.157
		Penguin	0.228 ± 0.166	0.181 ± 0.186	0.181 ± 0.154
		Guitar	0.342 ± 0.143	0.515 ± 0.335	0.344 ± 0.149
Imagine	Orthographic	Flower	0.351 ± 0.127	0.438 ± 0.300	0.334 ± 0.152
		Penguin	0.317 ± 0.048	0.470 ± 0.294	0.343 ± 0.129
		Guitar	0.280 ± 0.317	0.110 ± 0.198	0.112 ± 0.147
Imagine	Pictorial	Flower	0.357 ± 0.207	0.363 ± 0.284	0.282 ± 0.141
		Penguin	0.290 ± 0.141	0.312 ± 0.290	0.250 ± 0.160
		Guitar	0.305 ± 0.205	0.272 ± 0.309	0.211 ± 0.185
Perception	Audio	Flower	0.112 ± 0.141	0.095 ± 0.167	0.092 ± 0.131
		Penguin	0.314 ± 0.050	0.705 ± 0.343	0.409 ± 0.141
		Guitar	0.241 ± 0.257	0.176 ± 0.223	0.159 ± 0.167
Perception	Orthographic	Flower	0.273 ± 0.152	0.259 ± 0.220	0.231 ± 0.157
		Penguin	0.343 ± 0.033	0.695 ± 0.255	0.441 ± 0.084
		Guitar	0.242 ± 0.264	0.056 ± 0.061	0.084 ± 0.083
Perception	Pictorial	Flower	0.327 ± 0.114	0.479 ± 0.296	0.359 ± 0.165
		Penguin	0.270 ± 0.180	0.171 ± 0.176	0.181 ± 0.152
		Guitar	0.326 ± 0.059	0.368 ± 0.263	0.311 ± 0.109

B.2 Pipeline 1 Only Attention

B.2.1 Sensory Modality

Table B.10: Per-subject balanced accuracy for the Sensory Modality discrimination task, attention rule, by task (one value per Leave-One-Subject-Out fold).

Subject	Imagine	Perception
3_3	0.322	0.338
8_3	0.378	0.344
10_1	0.314	0.392
11_1	0.305	0.300
12_1	0.329	0.332
12_2	0.311	0.307
13_1	0.283	0.355
14_1	0.303	0.350
14_2	0.343	0.332
15_1	0.385	0.353
15_2	0.375	0.342
16_1	0.341	0.316
17_1	0.335	0.292
18_1	0.330	0.356
19_1	0.357	0.327

Table B.11: Balanced accuracy (mean $\pm$ std over 15 folds) for the Sensory Modality discrimination task, by task, with the one-sided Wilcoxon signed-rank test against chance (0.333): raw p -value and Benjamini–Hochberg corrected p -value. Significance is assessed on the corrected value only: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; n.s. = not significant.

Task	Balanced accuracy	p (raw)	p (BH)
Imagine	0.334 ± 0.029	0.533	1.000 n.s.
Perception	0.336 ± 0.025	0.381	0.914 n.s.

Table B.12: Per-class precision, recall and F1 (mean $\pm$ std over 15 folds) for the Sensory Modality discrimination task, attention rule.

Task	Class	P	R	F1
Imagine	Audio	0.352 ± 0.088	0.340 ± 0.214	0.306 ± 0.125
	Orthographic	0.307 ± 0.119	0.376 ± 0.295	0.288 ± 0.166
	Pictorial	0.318 ± 0.087	0.286 ± 0.213	0.267 ± 0.155
Perception	Audio	0.322 ± 0.158	0.302 ± 0.268	0.257 ± 0.172
	Orthographic	0.312 ± 0.145	0.288 ± 0.304	0.231 ± 0.168
	Pictorial	0.356 ± 0.189	0.417 ± 0.320	0.314 ± 0.170

B.2.2 Semantic Concept

Table B.13: Per-subject balanced accuracy for the Semantic Concept recognition task, attention rule, by task–modality condition (one value per Leave-One-Subject-Out fold).

Subject	Imagine			Perception		
	Audio	Orthographic	Pictorial	Audio	Orthographic	Pictorial
3_3	0.287	0.330	0.344	0.351	0.360	0.288
8_3	0.300	0.340	0.301	0.277	0.280	0.307
10_1	0.334	0.339	0.355	0.383	0.338	0.376
11_1	0.287	0.221	0.377	0.368	0.354	0.357
12_1	0.328	0.334	0.387	0.390	0.396	0.305
12_2	0.324	0.223	0.342	0.337	0.370	0.317
13_1	0.312	0.354	0.248	0.374	0.321	0.375
14_1	0.311	0.343	0.283	0.292	0.339	0.390
14_2	0.344	0.270	0.327	0.343	0.336	0.322
15_1	0.371	0.396	0.428	0.310	0.289	0.380
15_2	0.300	0.324	0.340	0.258	0.312	0.354
16_1	0.363	0.360	0.303	0.302	0.329	0.362
17_1	0.300	0.361	0.370	0.373	0.339	0.372
18_1	0.333	0.268	0.288	0.350	0.423	0.433
19_1	0.368	0.271	0.365	0.337	0.351	0.312

Table B.14: Balanced accuracy (mean $\pm$ std over 15 folds) for the Semantic Concept recognition task, by task–modality condition, with the one-sided Wilcoxon signed-rank test against chance (0.333): raw p -value and Benjamini–Hochberg corrected p -value. Significance is assessed on the corrected value only: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; n.s. = not significant.

Modality	Stimulus	Balanced accuracy	p (raw)	p (BH)
Imagine	Audio	0.324 ± 0.027	0.864	0.972 n.s.
Imagine	Orthographic	0.316 ± 0.051	0.719	0.857 n.s.
Imagine	Pictorial	0.337 ± 0.045	0.339	0.679 n.s.
Perception	Audio	0.336 ± 0.039	0.360	0.682 n.s.
Perception	Orthographic	0.342 ± 0.036	0.165	0.457 n.s.
Perception	Pictorial	0.350 ± 0.039	0.060	0.217 n.s.

Table B.15: Per-class precision, recall and F1 (mean $\pm$ std over 15 folds) for the Semantic Concept recognition task, attention rule, by task–modality condition.

Modality	Stimulus	Class	P	R	F1
Imagine	Audio	Flower	0.291 ± 0.145	0.368 ± 0.241	0.293 ± 0.148
		Penguin	0.297 ± 0.105	0.334 ± 0.257	0.283 ± 0.143
		Guitar	0.335 ± 0.134	0.271 ± 0.246	0.245 ± 0.144
Imagine	Orthographic	Flower	0.304 ± 0.087	0.335 ± 0.292	0.267 ± 0.159
		Penguin	0.351 ± 0.213	0.359 ± 0.292	0.287 ± 0.166
		Guitar	0.269 ± 0.132	0.253 ± 0.248	0.218 ± 0.149
Imagine	Pictorial	Flower	0.301 ± 0.132	0.421 ± 0.262	0.329 ± 0.154
		Penguin	0.380 ± 0.097	0.230 ± 0.175	0.248 ± 0.104
		Guitar	0.335 ± 0.153	0.360 ± 0.266	0.298 ± 0.147
Perception	Audio	Flower	0.308 ± 0.116	0.250 ± 0.196	0.256 ± 0.140
		Penguin	0.329 ± 0.151	0.260 ± 0.203	0.257 ± 0.136
		Guitar	0.308 ± 0.063	0.499 ± 0.242	0.361 ± 0.130
Perception	Orthographic	Flower	0.425 ± 0.177	0.362 ± 0.224	0.320 ± 0.139
		Penguin	0.242 ± 0.168	0.265 ± 0.275	0.212 ± 0.189
		Guitar	0.342 ± 0.072	0.400 ± 0.264	0.328 ± 0.141
Perception	Pictorial	Flower	0.348 ± 0.045	0.396 ± 0.267	0.334 ± 0.135
		Penguin	0.312 ± 0.138	0.307 ± 0.211	0.286 ± 0.140
		Guitar	0.309 ± 0.116	0.347 ± 0.245	0.300 ± 0.159

Appendix C

Appendix

C.1 Pipeline 2 Linear SVM

C.1.1 Imagination VS Perception

Table C.1: Per-subject balanced accuracy for the Perception–Imagination task, by sensory modality (one value per Leave-One-Subject-Out fold).

Subject	Audio	Orthographic	Pictorial
3_3	0.500	0.500	0.500
8_3	0.500	0.500	0.500
10_1	0.500	0.500	0.500
11_1	0.500	0.500	0.500
12_1	0.500	0.500	0.500
12_2	0.500	0.500	0.500
13_1	0.500	0.500	0.500
14_1	0.500	0.500	0.500
14_2	0.500	0.500	0.500
15_1	0.500	0.500	0.500
15_2	0.500	0.500	0.500
16_1	0.500	0.500	0.500
17_1	0.500	0.500	0.500
18_1	0.500	0.500	0.500
19_1	0.500	0.500	0.500

Table C.2: Balanced accuracy (mean $\pm$ std over 15 folds) for the Perception–Imagination task, by sensory modality, with the one-sided Wilcoxon signed-rank test against chance (0.5): raw p -value and Benjamini–Hochberg corrected p -value. Significance is assessed on the corrected value only: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; n.s. = not significant.

Modality	Balanced accuracy	p (raw)	p (BH)
Audio	0.500 ± 0.000	1.000	1.000 n.s.
Orthographic	0.500 ± 0.000	1.000	1.000 n.s.
Pictorial	0.500 ± 0.000	1.000	1.000 n.s.

Table C.3: Per-class precision, recall and F1 (mean $\pm$ std over 15 folds) for the Perception–Imagination task, by sensory modality.

Modality	Class	P	R	F1
Audio	Imagine	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
	Perception	0.502 ± 0.009	1.000 ± 0.000	0.669 ± 0.008
Orthographic	Imagine	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
	Perception	0.504 ± 0.011	1.000 ± 0.000	0.670 ± 0.009
Pictorial	Imagine	0.466 ± 0.125	0.933 ± 0.249	0.622 ± 0.167
	Perception	0.032 ± 0.121	0.067 ± 0.249	0.044 ± 0.163

C.1.2 Sensory Modality

Table C.4: Per-subject balanced accuracy for the Sensory Modality discrimination task, by task (one value per Leave-One-Subject-Out fold).

Subject	Imagine	Perception
3_3	0.333	0.333
8_3	0.333	0.333
10_1	0.333	0.333
11_1	0.333	0.333
12_1	0.333	0.333
12_2	0.333	0.333
13_1	0.333	0.333
14_1	0.333	0.333
14_2	0.333	0.333
15_1	0.333	0.333
15_2	0.333	0.333
16_1	0.333	0.333
17_1	0.333	0.333
18_1	0.333	0.333
19_1	0.333	0.333

Table C.5: Balanced accuracy (mean $\pm$ std over 15 folds) for the Sensory Modality discrimination task, by task, with the one-sided Wilcoxon signed-rank test against chance (0.333): raw p -value and Benjamini–Hochberg corrected p -value. Significance is assessed on the corrected value only: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; n.s. = not significant.

Task	Balanced accuracy	p (raw)	p (BH)
Imagine	0.333 ± 0.000	1.000	1.000 n.s.
Perception	0.333 ± 0.000	1.000	1.000 n.s.

Table C.6: Per-class precision, recall and F1 (mean $\pm$ std over 15 folds) for the Sensory Modality discrimination task.

Task	Class	P	R	F1
Imagine	Audio	0.221 ± 0.157	0.667 ± 0.471	0.332 ± 0.235
	Orthographic	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
	Pictorial	0.106 ± 0.150	0.333 ± 0.471	0.161 ± 0.228
Perception	Audio	0.313 ± 0.084	0.933 ± 0.249	0.468 ± 0.126
	Orthographic	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
	Pictorial	0.019 ± 0.073	0.067 ± 0.249	0.030 ± 0.113

C.1.3 Semantic Concept

Table C.7: Per-subject balanced accuracy for the Semantic Concept recognition task, by task–modality condition (one value per Leave-One-Subject-Out fold).

Subject	Imagine			Perception		
	Audio	Orthographic	Pictorial	Audio	Orthographic	Pictorial
3_3	0.329	0.382	0.270	0.242	0.314	0.338
8_3	0.333	0.320	0.369	0.324	0.329	0.340
10_1	0.264	0.374	0.319	0.323	0.318	0.383
11_1	0.267	0.272	0.352	0.261	0.344	0.430
12_1	0.390	0.402	0.255	0.311	0.322	0.412
12_2	0.360	0.367	0.315	0.339	0.386	0.400
13_1	0.356	0.320	0.331	0.351	0.342	0.327
14_1	0.339	0.373	0.353	0.279	0.384	0.390
14_2	0.342	0.331	0.348	0.345	0.316	0.331
15_1	0.347	0.449	0.324	0.344	0.391	0.470
15_2	0.322	0.380	0.290	0.321	0.341	0.453
16_1	0.205	0.383	0.319	0.310	0.358	0.380
17_1	0.273	0.347	0.383	0.416	0.266	0.352
18_1	0.383	0.332	0.330	0.369	0.324	0.248
19_1	0.323	0.301	0.355	0.347	0.262	0.416

Table C.8: Balanced accuracy (mean $\pm$ std over 15 folds) for the Semantic Concept recognition task, by task–modality condition, with the one-sided Wilcoxon signed-rank test against chance (0.333): raw p -value and Benjamini–Hochberg corrected p -value. Significance is assessed on the corrected value only: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; n.s. = not significant.

Modality	Stimulus	Balanced accuracy	p (raw)	p (BH)
Imagine	Audio	0.322 ± 0.048	0.681	0.857 n.s.
Imagine	Orthographic	0.355 ± 0.043	0.047	0.213 n.s.
Imagine	Pictorial	0.327 ± 0.034	0.619	0.857 n.s.
Perception	Audio	0.325 ± 0.042	0.719	0.857 n.s.
Perception	Orthographic	0.333 ± 0.037	0.598	0.857 n.s.
Perception	Pictorial	0.378 ± 0.055	0.006	0.032 *

Table C.9: Per-class precision, recall and F1 (mean $\pm$ std over 15 folds) for the Semantic Concept recognition task, by task–modality condition.

Modality	Stimulus	Class	P	R	F1
Imagine	Audio	Flower	0.327 ± 0.082	0.329 ± 0.070	0.327 ± 0.073
		Penguin	0.319 ± 0.053	0.322 ± 0.069	0.319 ± 0.058
		Guitar	0.321 ± 0.070	0.316 ± 0.086	0.316 ± 0.070
Imagine	Orthographic	Flower	0.388 ± 0.057	0.397 ± 0.063	0.390 ± 0.053
		Penguin	0.335 ± 0.071	0.339 ± 0.086	0.333 ± 0.066
		Guitar	0.346 ± 0.058	0.331 ± 0.065	0.336 ± 0.055
Imagine	Pictorial	Flower	0.329 ± 0.068	0.342 ± 0.075	0.334 ± 0.070
		Penguin	0.336 ± 0.082	0.292 ± 0.051	0.311 ± 0.059
		Guitar	0.322 ± 0.052	0.348 ± 0.060	0.333 ± 0.053
Perception	Audio	Flower	0.317 ± 0.046	0.313 ± 0.052	0.314 ± 0.047
		Penguin	0.343 ± 0.071	0.343 ± 0.076	0.342 ± 0.071
		Guitar	0.318 ± 0.075	0.320 ± 0.084	0.317 ± 0.076
Perception	Orthographic	Flower	0.334 ± 0.070	0.352 ± 0.078	0.342 ± 0.072
		Penguin	0.340 ± 0.049	0.330 ± 0.062	0.335 ± 0.054
		Guitar	0.325 ± 0.067	0.317 ± 0.061	0.320 ± 0.063
Perception	Pictorial	Flower	0.344 ± 0.066	0.379 ± 0.078	0.361 ± 0.071
		Penguin	0.380 ± 0.066	0.360 ± 0.081	0.369 ± 0.071
		Guitar	0.413 ± 0.070	0.395 ± 0.067	0.403 ± 0.065

Appendix D

Appendix

D.1 Pipeline 1 Friedman test results

Table D.1: Pipeline 1 ablation for the sensory modality task. The three decision rules (combined energy–attention, energy-only, attention-only) are compared per condition with a Friedman omnibus test. As no omnibus test reached significance, no post-hoc comparison was warranted and only the Friedman p -value is reported. Significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; n.s. = not significant.

Condition	Friedman p
Imagine	0.819 n.s.
Perception	0.247 n.s.

Table D.2: Pipeline 1 ablation for the semantic concept task. The three decision rules (combined energy–attention, energy-only, attention-only) are compared per condition with a Friedman omnibus test. As no omnibus test reached significance, no post-hoc comparison was warranted and only the Friedman p -value is reported. Significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; n.s. = not significant.

Condition	Friedman p
Imagine-Audio	0.193 n.s.
Imagine-Orthographic	0.549 n.s.
Imagine-Pictorial	0.282 n.s.
Perception-Audio	0.395 n.s.
Perception-Orthographic	0.207 n.s.
Perception-Pictorial	0.449 n.s.

Appendix E

Appendix

Table E.1: Per-subject balanced accuracy of Pipeline 2 on the Perception–Imagination task, by sensory modality, with the imagination epochs truncated to the same duration as the perception epochs(one value per Leave-One-Subject-Out fold).

Subject	Audio	Orthographic	Pictorial
3_3	0.484	0.500	0.505
8_3	0.497	0.689	0.659
10_1	0.531	0.570	0.704
11_1	0.502	0.542	0.500
12_1	0.517	0.500	0.500
12_2	0.510	0.505	0.523
13_1	0.512	0.825	0.817
14_1	0.597	0.612	0.858
14_2	0.569	0.799	0.862
15_1	0.487	0.506	0.647
15_2	0.537	0.652	0.767
16_1	0.545	0.661	0.711
17_1	0.509	0.521	0.520
18_1	0.498	0.500	0.500
19_1	0.631	0.694	0.893

Table E.2: Balanced accuracy (mean $\pm$ std over 15 folds) of Pipeline 2 on the Perception–Imagination task, by sensory modality, with the imagination epochs truncated to the same duration as the perception epochs, evaluated by the one-sided Wilcoxon signed-rank test against chance (0.5). The last two columns compare it with the alternative preprocessing by a paired Wilcoxon signed-rank test, summarised by the median per-session difference in balanced accuracy (Δ BA). p -values are uncorrected. Significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; n.s. = not significant.

Modality	Balanced accuracy	p (vs chance)	Median Δ BA	p (comparison)
Audio	0.528 ± 0.041	0.009 **	−0.443	< 0.001 ***
Orthographic	0.605 ± 0.107	0.001 **	−0.258	0.010 *
Pictorial	0.665 ± 0.145	0.001 **	−0.166	0.084 n.s.

Bibliography

- [1] Jonathan R. Wolpaw, Niels Birbaumer, Dennis J. McFarland, Gert Pfurtscheller, and Theresa M. Vaughan. «Brain–computer interfaces for communication and control». In: *Clinical Neurophysiology* 113.6 (2002), pp. 767–791. DOI: 10.1016/S1388-2457(02)00057-3 (cit. on pp. 1, 10).
- [2] Maryam M. Shanechi. «Brain–machine interfaces from motor to mood». In: *Nature Neuroscience* 22 (Oct. 2019), pp. 1554–1564. DOI: 10.1038/s41593-019-0488-y (cit. on p. 1).
- [3] Alexander G. Huth, Wendy A. de Heer, Thomas L. Griffiths, Frédéric E. Theunissen, and Jack L. Gallant. «Natural speech reveals the semantic maps that tile human cerebral cortex». In: *Nature* 532.7600 (Apr. 2016), pp. 453–458. DOI: 10.1038/nature17637 (cit. on p. 1).
- [4] Jerry Tang, Amanda LeBel, Shailee Jain, and Alexander G. Huth. «Semantic reconstruction of continuous language from non-invasive brain recordings». In: *Nature Neuroscience* 26 (2023), pp. 858–866. DOI: 10.1038/s41593-023-01304-9 (cit. on pp. 1, 16).
- [5] Milan Rybář and Ian Daly. «Neural decoding of semantic concepts: a systematic literature review». In: *Journal of Neural Engineering* 19.2 (2022), p. 021002. DOI: 10.1088/1741-2552/ac619a (cit. on pp. 1, 2, 10).
- [6] Hossein Ahmadi, Ali Kuhestani, and Luca Mesin. «Adversarial Neural Network Training for Secure and Robust Brain-to-Brain Communication». In: *IEEE Access* 12 (2024), pp. 39450–39469. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2024.3376657 (cit. on p. 1).
- [7] Hossein Ahmadi, Ali Kuhestani, Mohammadreza Keshavarzi, and Luca Mesin. «Securing Brain-to-Brain Communication Channels Using Adversarial Training on SSVEP EEG». In: *IEEE Access* 13 (2025), pp. 14358–14378. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2025.3528770 (cit. on p. 1).

- [8] Roberto Portillo-Lara, Bogachan Tahirbegi, Christopher A. R. Chapman, Josef A. Goding, and Rylie A. Green. «Mind the gap: State-of-the-art technologies and applications for EEG-based brain-computer interfaces». In: *APL Bioengineering* 5.3 (July 2021), p. 031507. DOI: 10.1063/5.0047237 (cit. on pp. 1, 14).
- [9] Fabien Lotte, Laurent Bougrain, Andrzej Cichocki, Maureen Clerc, Marco Congedo, Alain Rakotomamonjy, and Florian Yger. «A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update». In: *Journal of Neural Engineering* 15.3 (June 2018), p. 031005. DOI: 10.1088/1741-2552/aab2f2 (cit. on pp. 1, 5, 23, 27–30, 47).
- [10] Samuel Sutton, Margery Braren, Joseph Zubin, and E. Roy John. «Evoked-potential correlates of stimulus uncertainty». In: *Science* 150.3700 (Nov. 1965), pp. 1187–1188. DOI: 10.1126/science.150.3700.1187 (cit. on pp. 2, 22).
- [11] Marta Kutas and Steven A. Hillyard. «Reading senseless sentences: brain potentials reflect semantic incongruity». In: *Science* 207.4427 (Jan. 1980), pp. 203–205. DOI: 10.1126/science.7350657 (cit. on pp. 2, 22).
- [12] Patrick Suppes, Zhong-Lin Lu, and Bing Han. «Brain wave recognition of words». In: *Proceedings of the National Academy of Sciences of the United States of America* 94.26 (Dec. 1997), pp. 14965–14969. DOI: 10.1073/pnas.94.26.14965 (cit. on p. 2).
- [13] Patrick Suppes, Bing Han, and Zhong-Lin Lu. «Brain-wave recognition of sentences». In: *Proceedings of the National Academy of Sciences of the United States of America* 95.26 (1998), pp. 15861–15866. DOI: 10.1073/pnas.95.26.15861 (cit. on p. 2).
- [14] Irina Simanova, Marcel van Gerven, Robert Oostenveld, and Peter Hagoort. «Identifying Object Categories from Event-Related EEG: Toward Decoding of Conceptual Representations». In: *PLoS ONE* 5.12 (2010), e14465. DOI: 10.1371/journal.pone.0014465 (cit. on pp. 2, 29).
- [15] Brian Murphy, Massimo Poesio, Francesca Bovolo, Lorenzo Bruzzone, Michele Dalponte, and Heba Lakany. «EEG decoding of semantic category reveals distributed representations for single concepts». In: *Brain and Language* 117.1 (Apr. 2011), pp. 12–22. DOI: 10.1016/j.bandl.2010.09.013 (cit. on pp. 2, 29).
- [16] Tijl Grootswagers, Ivy Zhou, Amanda K. Robinson, et al. «Human EEG recordings for 1,854 concepts presented in rapid serial visual presentation streams». In: *Scientific Data* 9 (2022), p. 3. DOI: 10.1038/s41597-021-01102-7 (cit. on p. 3).

- [17] Andrew T. Gifford, Kshitij Dwivedi, Gemma Roig, and Radoslaw M. Cichy. «A large and rich EEG dataset for modeling human visual object recognition». In: *NeuroImage* 264 (Dec. 2022), p. 119754. DOI: 10.1016/j.neuroimage.2022.119754 (cit. on p. 3).
- [18] Holly Wilson, Mohammad Golbabaee, Michael J. Proulx, et al. «EEG-based BCI Dataset of Semantic Concepts for Imagination and Perception Tasks». In: *Scientific Data* 10 (2023), p. 386. DOI: 10.1038/s41597-023-02287-9 (cit. on pp. 3–5, 29, 30, 33–37, 44, 52, 103).
- [19] Simanto Saha and Mathias Baumert. «Intra- and Inter-subject Variability in EEG-Based Sensorimotor Brain Computer Interface: A Review». In: *Frontiers in Computational Neuroscience* 13 (Jan. 2020), p. 87. DOI: 10.3389/fncom.2019.00087 (cit. on p. 3).
- [20] Taida Li, Yujun Yan, Fei Dou, Wenzhan Song, and Xiang Zhang. «Cross-subject generalization for EEG decoding: a survey of deep learning methods». In: *Progress in Biomedical Engineering* 8.2 (May 2026), p. 022013. DOI: 10.1088/2516-1091/ae65f0 (cit. on p. 3).
- [21] Shuai Huang et al. «NEED: Cross-Subject and Cross-Task Generalization for Video and Image Reconstruction from EEG Signals». In: *Proceedings of the Thirty-ninth Annual Conference on Neural Information Processing Systems*. 2026. URL: <https://openreview.net/forum?id=L3aEdxJMh1> (cit. on p. 3).
- [22] Jian Huang, Yuhu Chang, Wei Li, Jinze Tong, and Shengguang Du. «A Spatio-Temporal Capsule Neural Network with Self-Correlation Routing for EEG Decoding of Semantic Concepts of Imagination and Perception Tasks». In: *Sensors* 24 (2024), p. 5988. DOI: 10.3390/s24185988 (cit. on pp. 3, 31).
- [23] Jinze Tong and Wanzhong Chen. «MMPI Net: A Novel Multimodal Model Considering the Similarities Between Perception and Imagination for Image Evoked EEG Decoding». In: *IEEE Journal of Biomedical and Health Informatics* 29.8 (2025), pp. 5549–5560. DOI: 10.1109/JBHI.2025.3554664 (cit. on pp. 3, 31).
- [24] Jinze Tong and Wanzhong Chen. «Electroencephalogram decoding driven by shared semantic information for perception and imagination cognitive processes». In: *Biomedical Signal Processing and Control* 116 (2026), p. 109586. DOI: 10.1016/j.bspc.2026.109586 (cit. on pp. 3, 31).
- [25] Jinze Tong and Wanzhong Chen. «PI Net: An end-to-end semantic decoding model for EEG signals in perception and imagination tasks». In: *Digital Signal Processing* 164 (2025), p. 105250. DOI: 10.1016/j.dsp.2025.105250 (cit. on pp. 3, 4, 31).

- [26] Owais Mujtaba Khanday, Marc Ouellet, Jose L. Pérez-Córdoba, Asma Hasan Sbaih, Laura Miccoli, and Jose A. Gonzalez-Lopez. «Decoding the Mind: Neural Differences and Semantic Representation in Perception and Imagination Across Modalities». In: *Proceedings of IberSPEECH 2024*. Aveiro, Portugal, Nov. 2024, pp. 126–130. DOI: 10.21437/IberSPEECH.2024-26 (cit. on pp. 3, 4, 29, 30).
- [27] Xuan-The Tran, Quoc-Toan Nguyen, Linh Le, Thomas Do, and Chin-Teng Lin. «EEG-Based Contrastive Learning Models For Object Perception Using Multisensory Image-Audio Stimuli». In: *Proceedings of the 1st International Workshop on Brain-Computer Interfaces (BCI) for Multimedia Understanding*. BCIMM '24. New York, NY, USA: Association for Computing Machinery, 2024, pp. 39–47. DOI: 10.1145/3688862.3689116 (cit. on pp. 3, 4, 31).
- [28] H. Ahmadi and L. Mesin. «Decoding Visual Imagination and Perception from EEG via Topomap Sequences». In: *Proceedings of the 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Copenhagen, Denmark, 2025, pp. 1–7. DOI: 10.1109/EMBC58623.2025.11251641 (cit. on pp. 4–6, 31, 48–51, 103).
- [29] H. Ahmadi, M. Impagnatiello, and L. Mesin. «Semantic Latent Geometry Reveals Imagination–Perception Structure in EEG». In: *Applied Sciences* 16 (2026), p. 661. DOI: 10.3390/app16020661 (cit. on pp. 4, 32, 38, 51, 103).
- [30] Hossein Ahmadi and Luca Mesin. «Universal semantic feature extraction from EEG signals: a task-independent framework». In: *Journal of Neural Engineering* 22.3 (2025), p. 036003. ISSN: 1741-2560. DOI: 10.1088/1741-2552/add08f (cit. on pp. 4, 30).
- [31] Alexander Craik, Yongtian He, and Jose L. Contreras-Vidal. «Deep learning for electroencephalogram (EEG) classification tasks: a review». In: *Journal of Neural Engineering* 16.3 (June 2019), p. 031001. DOI: 10.1088/1741-2552/ab0ab5 (cit. on pp. 4, 5, 30).
- [32] Robin T. Schirrmeister, Jost T. Springenberg, Lukas D. J. Fiederer, Martin Glasstetter, Katharina Eggersperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. «Deep learning with convolutional neural networks for EEG decoding and visualization». In: *Human Brain Mapping* 38.11 (Nov. 2017), pp. 5391–5420. DOI: 10.1002/hbm.23730 (cit. on pp. 4, 30).
- [33] Irene Sturm, Sebastian Lapuschkin, Wojciech Samek, and Klaus-Robert Müller. «Interpretable deep neural networks for single-trial EEG classification». In: *Journal of Neuroscience Methods* 274 (2016), pp. 141–145. DOI: 10.1016/j.jneumeth.2016.10.008 (cit. on pp. 4, 30).

- [34] Param Rajpura, Hubert Cecotti, and Yogesh Kumar Meena. «Explainable artificial intelligence approaches for brain–computer interfaces: a review and design space». In: *Journal of Neural Engineering* 21.4 (Aug. 2024), p. 041003. DOI: 10.1088/1741-2552/ad6593 (cit. on pp. 5, 30).
- [35] Herbert Ramoser, Johannes Müller-Gerking, and Gert Pfurtscheller. «Optimal spatial filtering of single trial EEG during imagined hand movement». In: *IEEE Transactions on Rehabilitation Engineering* 8.4 (Dec. 2000), pp. 441–446. DOI: 10.1109/86.895946 (cit. on pp. 5, 28, 39).
- [36] Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, and Klaus-Robert Müller. «Optimizing Spatial Filters for Robust EEG Single-Trial Analysis». In: *IEEE Signal Processing Magazine* 25.1 (2008), pp. 41–56. DOI: 10.1109/MSP.2008.4408441 (cit. on pp. 5, 28, 39).
- [37] Mohammadjavad Asgari. «EEG Representational Form and Decoding Performance Across Motor Imagery, P300, and SSVEP Paradigms». Relatori: Luca Mesin, Hossein Ahmadi. Laurea magistrale in Data Science and Engineering. Politecnico di Torino, 2026, p. 85. URL: <https://webthesis.biblio.polito.it/id/eprint/38759> (cit. on p. 5).
- [38] Jeffrey R. Binder, Rutvik H. Desai, William W. Graves, and Lisa L. Conant. «Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies». In: *Cerebral Cortex* 19.12 (2009), pp. 2767–2796. DOI: 10.1093/cercor/bhp055 (cit. on pp. 6, 16).
- [39] Huanhuan Wei et al. «Mimicking efferent nerves using a graphdiyne-based artificial synapse with multiple ion diffusion dynamics». In: *Nature Communications* 12 (2021), p. 1068. DOI: 10.1038/s41467-021-21319-9 (cit. on p. 9).
- [40] Hossein Ahmadi, Paolo Costa, and Luca Mesin. *A Novel Hierarchical Binary Classification for Coma Outcome Prediction Using EEG, CNN, and Traditional ML Approaches*. Preprint. Nov. 2024. DOI: 10.36227/techrxiv.173220750.05940724/v1 (cit. on p. 10).
- [41] Francesco Serracca. «Elaborazione del segnale EEG di pazienti in coma». Relatore: Luca Mesin. Laurea magistrale in Ingegneria Biomedica. Politecnico di Torino, 2025, p. 103. URL: <https://webthesis.biblio.polito.it/id/eprint/32784> (cit. on p. 10).
- [42] Volker Straebel and Wilm Thoben. «Alvin Lucier’s Music for Solo Performer: Experimental music beyond sonification». In: *Organised Sound* 19.1 (2014), pp. 17–29. DOI: 10.1017/S135577181300037X (cit. on p. 10).
- [43] Jacques J. Vidal. «Toward Direct Brain-Computer Communication». In: *Annual Review of Biophysics and Bioengineering* 2 (1973), pp. 157–180. DOI: 10.1146/annurev.bb.02.060173.001105 (cit. on p. 10).

- [44] Jacques J. Vidal. «Real-Time Detection of Brain Events in EEG». In: *Proceedings of the IEEE* 65.5 (1977), pp. 633–641. DOI: 10.1109/PROC.1977.10542 (cit. on p. 10).
- [45] Reza Abiri, Soheil Borhani, Eric W. Sellers, Yang Jiang, and Xiaopeng Zhao. «A comprehensive review of EEG-based brain–computer interface paradigms». In: *Journal of Neural Engineering* 16.1 (2019), p. 011001. DOI: 10.1088/1741-2552/aaf12e (cit. on p. 10).
- [46] Thorsten O. Zander and Christian Kothe. «Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general». In: *Journal of Neural Engineering* 8.2 (2011), p. 025005. DOI: 10.1088/1741-2560/8/2/025005 (cit. on p. 10).
- [47] Felix Bloch. «Nuclear Induction». In: *Physical Review* 70.7–8 (1946), pp. 460–474. DOI: 10.1103/PhysRev.70.460 (cit. on p. 11).
- [48] Edward M. Purcell, Henry C. Torrey, and Robert V. Pound. «Resonance Absorption by Nuclear Magnetic Moments in a Solid». In: *Physical Review* 69.1–2 (1946), pp. 37–38. DOI: 10.1103/PhysRev.69.37 (cit. on p. 11).
- [49] Paul C. Lauterbur. «Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance». In: *Nature* 242 (1973), pp. 190–191. DOI: 10.1038/242190a0 (cit. on p. 11).
- [50] Seiji Ogawa, Tso-Ming Lee, Alan R. Kay, and David W. Tank. «Brain magnetic resonance imaging with contrast dependent on blood oxygenation». In: *Proceedings of the National Academy of Sciences* 87.24 (1990), pp. 9868–9872. DOI: 10.1073/pnas.87.24.9868 (cit. on p. 11).
- [51] Nikos K. Logothetis, Jon Pauls, Mark Augath, Torsten Trinath, and Axel Oeltermann. «Neurophysiological investigation of the basis of the fMRI signal». In: *Nature* 412.6843 (July 2001), pp. 150–157. DOI: 10.1038/35084005 (cit. on p. 11).
- [52] Seong-Gi Kim, Wolfgang Richter, and Kâmil Uğurbil. «Limitations of temporal resolution in functional MRI». In: *Magnetic Resonance in Medicine* 37.4 (Apr. 1997), pp. 631–636. DOI: 10.1002/mrm.1910370427 (cit. on p. 11).
- [53] Vanessa Scarapicchia, Cassandra Brown, Chantel Mayo, and Jodie R. Gawryluk. «Functional Magnetic Resonance Imaging and Functional Near-Infrared Spectroscopy: Insights from Combined Recording Studies». In: *Frontiers in Human Neuroscience* 11 (2017), p. 419. DOI: 10.3389/fnhum.2017.00419 (cit. on p. 11).

- [54] Ruixing Liang et al. «Unidirectional Brain-Computer Interface: Artificial Neural Network Encoding Natural Images to fMRI Response in the Visual Cortex». In: *ICASSP 2024 – IEEE International Conference on Acoustics, Speech and Signal Processing*. 2024. DOI: 10.1109/ICASSP48485.2024.10446366 (cit. on p. 12).
- [55] Frans F. Jöbsis. «Noninvasive, Infrared Monitoring of Cerebral and Myocardial Oxygen Sufficiency and Circulatory Parameters». In: *Science* 198.4323 (1977), pp. 1264–1267. DOI: 10.1126/science.929199 (cit. on p. 11).
- [56] Marco Ferrari and Valentina Quaresima. «A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application». In: *NeuroImage* 63.2 (Nov. 2012), pp. 921–935. DOI: 10.1016/j.neuroimage.2012.03.049 (cit. on p. 12).
- [57] David T. Delpy, Mark Cope, Pieter van der Zee, Simon Arridge, Susan Wray, and John Wyatt. «Estimation of optical pathlength through tissue from direct time of flight measurement». In: *Physics in Medicine and Biology* 33.12 (Dec. 1988), pp. 1433–1442. DOI: 10.1088/0031-9155/33/12/008 (cit. on p. 12).
- [58] Paola Pinti, Ilias Tachtsidis, Antonia Hamilton, Joy Hirsch, Clarisse Aichelburg, Sam Gilbert, and Paul W. Burgess. «The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience». In: *Annals of the New York Academy of Sciences* 1464.1 (Mar. 2020), pp. 5–29. DOI: 10.1111/nyas.13948 (cit. on p. 12).
- [59] Ning Zhang et al. «Recruitment of the Central Nervous System in Different Hand Tasks in Patients with Hand Dysfunction after Stroke Based on Functional Near-Infrared Spectroscopy: An Exploratory Study». In: *Journal of Rehabilitation Medicine* 58 (2026), jrm44712. DOI: 10.2340/jrm.v58.44712. URL: <https://doi.org/10.2340/jrm.v58.44712> (cit. on p. 13).
- [60] Paolo M. Rossini, Jonathan Cole, Walter Paulus, Ulf Ziemann, and Robert Chen. «1924–2024: First centennial of EEG». In: *Clinical Neurophysiology* 170 (Feb. 2025). Epub 2024 Dec 19, pp. 132–135. DOI: 10.1016/j.clinph.2024.11.021 (cit. on p. 14).
- [61] Hans Berger. «Über das Elektrenkephalogramm des Menschen». In: *Archiv für Psychiatrie und Nervenkrankheiten* 87 (1929), pp. 527–570. DOI: 10.1007/BF01797193 (cit. on p. 14).
- [62] Xiu-Yun Liu et al. «Recent applications of EEG-based brain-computer-interface in the medical field». In: *Military Medicine Research* 12.1 (Mar. 2025), p. 14. DOI: 10.1186/s40779-025-00598-z (cit. on p. 14).
- [63] Fernando Lopes da Silva. «EEG and MEG: Relevance to Neuroscience». In: *Neuron* 80.5 (Dec. 2013), pp. 1112–1128. DOI: 10.1016/j.neuron.2013.10.017 (cit. on p. 14).

- [64] David Cohen. «Magnetoencephalography: Evidence of magnetic fields produced by alpha-rhythm currents». In: *Science* 161.3843 (1968), pp. 784–786. DOI: 10.1126/science.161.3843.784 (cit. on p. 14).
- [65] David Cohen. «Magnetoencephalography: Detection of the brain’s electrical activity with a superconducting magnetometer». In: *Science* 175.4022 (1972), pp. 664–666. DOI: 10.1126/science.175.4022.664 (cit. on p. 14).
- [66] Matti Hämäläinen, Riitta Hari, Risto J. Ilmoniemi, Jukka Knuutila, and Olli V. Lounasmaa. «Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain». In: *Reviews of Modern Physics* 65.2 (1993), pp. 413–497. DOI: 10.1103/RevModPhys.65.413 (cit. on p. 14).
- [67] Elena Boto, Niall Holmes, James Leggett, Gillian Roberts, Vishal Shah, Sofie S. Meyer, Leonardo Duque Muñoz, Karen J. Mullinger, Tim M. Tierney, et al. «Moving magnetoencephalography towards real-world applications with a wearable system». In: *Nature* 555.7698 (2018), pp. 657–661. DOI: 10.1038/nature26147 (cit. on p. 15).
- [68] Benjamin Wittevrongel et al. «Practical real-time MEG-based neural interfacing with optically pumped magnetometers». In: *BMC Biology* 19 (2021), p. 158. DOI: 10.1186/s12915-021-01073-6 (cit. on p. 15).
- [69] György Buzsáki, Costas A. Anastassiou, and Christof Koch. «The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes». In: *Nature Reviews Neuroscience* 13.6 (May 2012), pp. 407–420. DOI: 10.1038/nrn3241 (cit. on pp. 15, 17).
- [70] David A. Moses et al. «Neuroprosthesis for Decoding Speech in a Paralyzed Person with Anarthria». In: *New England Journal of Medicine* 385.3 (2021), pp. 217–227. DOI: 10.1056/NEJMoA2027540 (cit. on p. 16).
- [71] Edwin M. Maynard, Chris T. Nordhausen, and Richard A. Normann. «The Utah Intracortical Electrode Array: a recording structure for potential brain-computer interfaces». In: *Electroencephalography and Clinical Neurophysiology* 102.3 (1997), pp. 228–239. DOI: 10.1016/S0013-4694(96)95176-0 (cit. on p. 16).
- [72] Mariana P. Branco, Simon H. Geukes, Erik J. Aarnoutse, Nick F. Ramsey, and Mariska J. Vansteensel. «Nine decades of electrocorticography: A comparison between epidural and subdural recordings». In: *European Journal of Neuroscience* 57.8 (2023), pp. 1260–1288. DOI: 10.1111/ejn.15941 (cit. on p. 17).

- [73] George H. Klem, Hans O. Lüders, Herbert H. Jasper, and Christian Elger. «The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology». In: *Electroencephalography and Clinical Neurophysiology. Supplement* 52 (1999), pp. 3–6 (cit. on p. 18).
- [74] Tzyy-Ping Jung, Scott Makeig, Colin Humphries, Te-Won Lee, Martin J. McKeown, Vicente Iragui, and Terrence J. Sejnowski. «Removing electroencephalographic artifacts by blind source separation». In: *Psychophysiology* 37.2 (Mar. 2000), pp. 163–178 (cit. on p. 20).
- [75] Martina Impagnatiello. «Extraction of Semantic Features from EEG Signals: Correlation with Perception, Imagination, and Vividness». Relatore: Luca Mesin. Laurea magistrale in Ingegneria Biomedica. Politecnico di Torino, 2025, p. 68. URL: <https://webthesis.biblio.polito.it/id/eprint/34882> (cit. on pp. 23, 38).
- [76] Francesco Alessandro. «Entropy-Based Semantic Feature Extraction from EEG Signals». Relatori: Luca Mesin, Hossein Ahmadi. Laurea magistrale in Ingegneria Biomedica. Politecnico di Torino, 2024, p. 94. URL: <https://webthesis.biblio.polito.it/id/eprint/32130> (cit. on p. 23).
- [77] Hooi Nee Oon, A. Saidatul, and Z. Ibrahim. «Analysis on Non-Linear Features of Electroencephalogram (EEG) Signal for Neuromarketing Application». In: *Proceedings of the International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA)*. 2018, pp. 1–8 (cit. on p. 24).
- [78] James W. Cooley and John W. Tukey. «An Algorithm for the Machine Calculation of Complex Fourier Series». In: *Mathematics of Computation* 19.90 (1965), pp. 297–301. DOI: 10.1090/S0025-5718-1965-0178586-1 (cit. on p. 24).
- [79] Petre Stoica and Randolph L. Moses. *Spectral Analysis of Signals*. Upper Saddle River, NJ: Prentice Hall, 2005 (cit. on pp. 25, 27).
- [80] Peter D. Welch. «The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms». In: *IEEE Transactions on Audio and Electroacoustics* 15.2 (1967), pp. 70–73. DOI: 10.1109/TAU.1967.1161901 (cit. on p. 25).
- [81] Stéphane Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way*. 3rd ed. Burlington, MA: Academic Press, 2009. DOI: 10.1016/B978-0-12-374370-1.X0001-8 (cit. on p. 26).
- [82] Catherine Tallon-Baudry and Olivier Bertrand. «Oscillatory Gamma Activity in Humans and Its Role in Object Representation». In: *Trends in Cognitive Sciences* 3.4 (1999), pp. 151–162. DOI: 10.1016/S1364-6613(99)01299-1 (cit. on p. 26).

- [83] Steven M. Kay. *Modern Spectral Estimation: Theory and Application*. Englewood Cliffs, NJ: Prentice Hall, 1988 (cit. on p. 27).
- [84] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. New York: Springer, 2006. ISBN: 978-0-387-31073-2 (cit. on pp. 27, 28, 47).
- [85] Alexandre Barachant, Stéphane Bonnet, Marco Congedo, and Christian Jutten. «Multiclass Brain–Computer Interface Classification by Riemannian Geometry». In: *IEEE Transactions on Biomedical Engineering* 59.4 (2012), pp. 920–928. DOI: 10.1109/TBME.2011.2172210 (cit. on pp. 29, 46).
- [86] Marco Congedo, Alexandre Barachant, and Rajendra Bhatia. «Riemannian Geometry for EEG-Based Brain–Computer Interfaces; a Primer and a Review». In: *Brain-Computer Interfaces* 4.3 (2017), pp. 155–174. DOI: 10.1080/2326263X.2017.1297192 (cit. on pp. 29, 46).
- [87] Hossein Ahmadi and Luca Mesin. «Enhancing MI EEG Signal Classification With a Novel Weighted and Stacked Adaptive Integrated Ensemble Model: A Multi-Dataset Approach». In: *IEEE Access* 12 (2024), pp. 103626–103646. URL: <https://api.semanticscholar.org/CorpusID:271550829> (cit. on p. 29).
- [88] Hossein Ahmadi and Luca Mesin. «Enhancing Motor Imagery Electroencephalography Classification with a Correlation-Optimized Weighted Stacking Ensemble Model». In: *Electronics* 13.6 (2024). ISSN: 2079-9292. DOI: 10.3390/electronics13061033. URL: <https://www.mdpi.com/2079-9292/13/6/1033> (cit. on p. 29).
- [89] Vito Di Leo. «Unsupervised Learning for Extracting Latent Semantic Representations from EEG Data». Relatori: Luca Mesin, Hossein Ahmadi. Laurea magistrale in Ingegneria Biomedica. Politecnico di Torino, 2026, p. 58. URL: <https://webthesis.biblio.polito.it/id/eprint/38394> (cit. on p. 30).
- [90] Vernon J. Lawhern, Amelia J. Solon, Nicholas R. Waytowich, Stephen M. Gordon, Chou P. Hung, and Brent J. Lance. «EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces». In: *Journal of Neural Engineering* 15.5 (2018), p. 056013. DOI: 10.1088/1741-2552/aace8c (cit. on p. 30).
- [91] Hossein Ahmadi, Eduardo Santamaria Vazquez, Luca Mesin, and Roberto Hornero. *Semantic-aware Decoding of Covert Inner Speech: A Multimodal EEG–EMG–Audio Framework*. Available at SSRN. 2025. DOI: 10.2139/ssrn.6874239. URL: <https://ssrn.com/abstract=6874239> (cit. on p. 31).
- [92] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. *Neural Machine Translation by Jointly Learning to Align and Translate*. 2015. arXiv: 1409.0473 [cs.CL] (cit. on p. 40).

- [93] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. «Attention Is All You Need». In: *Advances in Neural Information Processing Systems*. Vol. 30. 2017. DOI: 10.48550/arXiv.1706.03762 (cit. on p. 40).
- [94] Olivier Ledoit and Michael Wolf. «A well-conditioned estimator for large-dimensional covariance matrices». In: *Journal of Multivariate Analysis* 88.2 (2004), pp. 365–411. DOI: 10.1016/S0047-259X(03)00096-4 (cit. on pp. 45, 46).

Acknowledgements

My first thanks go to my family, who supported me enormously during these past difficult years. I would also like to thank all my friends, and among them I want to give special thanks to the guys from Ixem Labs, who gave me tremendous emotional support: Giovanni, Elena, Fabio, Irene, Ussama, Nicholas, and Daniele. In addition, I thank Professor Mesin and Hossein for allowing me to work on this topic. Last but not least, I would like to thank Dr. Francesca Tardio and Dr. Francesco Cuniberti, who gave me enormous personal support over this past year, helping me through a difficult time.