Extracting Hidden Neural Representations In Human Visual Perception

A Thesis submitted for the completion of
requirements for the degree of

Bachelor of Science
(Research)

by

Atharv Suryawanshi
Undergraduate Programme
Indian Institute of Science

Under the supervision of

Prof. Dr. Martin Hebart
Max Planck Institute for Human Cognitive and Brain Sciences

Prof. Chandni Usha
Indian Institute of Science

Dedicated to my younger self

Declaration

This is to confirm that the bachelor's thesis titled
"Extracting Hidden Neural Representations in Human Visual Perception", submitted by **Atharv Suryawanshi** (Sr. No. 11-01-00-10-91-21-1-20582) to the Indian Institute of Science, Bangalore, in partial fulfillment of the requirements for the award of a Bachelor's degree in Physics (Research), is a record of the genuine work carried out by him under my supervision and guidance during the academic year 2024–25.

Attested by
Prof. Dr. Martin Hebart (Primary Guide)
Computational Cognitive Neuroscience and Quantitative Psychiatry
Justus Liebig University Giessen, Germany
Vision and Computational Cognition Group
Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Date: 09-04-2025 Signature:

Dr. Chandni Usha (Department Guide)
Department of Instrumentation and Applied Physics
Indian Institute of Science
Bangalore - 560012, India

Date: 09-04-2025 Signature:

Declaration by the Candidate

I Atharv Suryawanshi, certify that –

- The work presented in this report was carried out by me under the supervision of my advisor.
- It has not been submitted elsewhere for the award of any degree or diploma.
- I have adhered to the ethical guidelines of the Institute.
- All sources and contributions have been duly acknowledged and cited.

Date: 9th April 2025
Place: IISc, Bangalore

Signature:

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my parents, whose unwavering support, encouragement, and love have been the steady foundation throughout my academic journey. Without them, none of this would have been possible.

I am especially thankful to Professor Dr. Martin Hebart for his mentorship, guidance, and insightful feedback throughout this project. His sharp eye for detail and thought-provoking discussions have helped shape this thesis into what it is. I would also like to acknowledge the wonderful members of the Hebart lab - Tonghe, Sander, Lenny, Malin, Juhi, and Fernanda, each of whom brought a unique perspective to the table. Their suggestions, questions, and willingness to discuss ideas made for an intellectually rich and collaborative environment. I am grateful to have learned from them.

I also owe my thanks to Dr. Narayanan, Dr. Dhawale, Dr. Rougier, and Dr. Leblois, whose teaching and research kindled my interest in computational neuroscience. The curiosity you sparked and the tools you shared laid the foundation for much of what I explored in this thesis.

To my friends - Sourabh, Sanjith, Megha, Anusha, Divija, Gaurav, Aditi, Akshank, Oviya, and Anmol thank you for your constant presence, emotional support, and the well-timed distractions that helped me stay sane. You kept my morale high when the only thing high was my folder count of intermediate plots.

A special mention goes to the UGCL server, which worked tirelessly (and with much less complaint than I did) to run my experiments efficiently and on time. Your GPUs will always have my respect.

And finally, to Pepper, the cat who often sat next to me during long nights of writing - your ability to nap through my existential crises and purr through my debugging made everything feel a little lighter. Thank you for reminding me to take breaks, stretch, and occasionally chase a sock for no reason.

Abstract

The brain encodes and stores information about the external world through neural representations, forming the foundation of perception and cognition. Objects, as core elements of our visual environment, have traditionally been studied through the lens of object recognition. While classical theories emphasize categorical recognition, newer perspectives propose a multidimensional representational space that captures object properties, behavioural relevance, and contextual associations beyond simple identification. In this study, we take a hypothesis-neutral, data-driven approach to uncover hidden neural representations of everyday objects. Using magnetoencephalography (MEG) recordings from multiple participants, we analyze latent dimensions underlying visual processing. Using the THINGS dataset, a systematically sampled collection of diverse, concrete, and nameable object concepts, we identify distinct and highly structured neural representations within individuals. However, these representations exhibit slight variability across participants, suggesting individual differences in how the brain encodes object concepts. Our findings contribute to a more detailed understanding of neural representations, moving beyond object recognition to a richer model of how the brain organizes and interprets visual information. This work offers insights into the underlying principles of neural coding and the variability of object representations across individuals.

Contents

Chapter 1

Introduction

1.1 Organization

Neural representation refers to the structured encoding of information within the brain through a pattern of neuronal activity. This includes sensory stimuli, motor commands, cognitive processes, memory, emotions, and abstract concepts. Instead of storing information explicitly, the brain represents it through a distributed activity across populations of neurons. A specific population activity pattern corresponds to a distinct perceptual, motor, or cognitive state. These representations are shaped by the brain's intrinsic organization and experience-dependent plasticity, varying across individuals. Traditional hypothesis-driven approaches have primarily focused on predefined categories, limiting our understanding of how category selectivity emerges in our brain and its overall functional organization. To address this, we take a data-driven approach to extract how the brain represents object categories and how they vary with time. Rather than relying on a predefined hypothesis, we use matrix decomposition methods to identify naturally occurring latent neural dimensions, offering a more unbiased view of the brain's representational structure. Using a non-negative least squares approach, we track how these representations vary in importance over time during visual processing, revealing how distinct neural dimensions emerge and contribute to perception.

1.2 Background

1.2.1 Principles of Visual Recognition

The ability to perceive and interpret visual information is one of the most complex and essential functions of the brain. For years, researchers have considered object recognition to be the most important goal of the visual system, with the ventral visual pathway regarded as the major component of object perception, also called the 'what' pathway. This contrasts with the dorsal pathway, also called the 'how' pathway, which specialises in action-related tasks [1]. Using this study as a foundation, object vision research has explored mechanisms to understand how the ventral pathway solves the problem of object recognition. However, vision is not merely about capturing static images; it involves dynamic interaction between sensory input and cognitive processes. At its core, visual processing involves extracting meaningful representations from incoming sensory data within the context of a human's rich behavioural and cognitive repertoire. The visual system recognises objects, assigns semantic meaning, and integrates information across different spatial and temporal scales. This process is crucial for interacting with the environment, as it enables the identification of objects regardless of variations in lighting, perspective, or occlusion. A central challenge to this process is the problem of generalisation, i.e. how the brain constructs robust representations that extend beyond the immediate physical input.

- consider a 100 x 100 pixel black-and-white image representing a single digit. The raw visual input comprises of $2^{10,000}$ possible configuration. To solve this recognition task, the brain would have to store an explicit dictionary of all possible mappings, which is infeasible due to the immense memory requirements of a comparatively simple task.
- Instead, our brain solves this problem by extracting meaningful features from the image rather than memorising individual instances. It uses hierarchical processing. In the visual system, lower-level neurons detect simple elements like edges and textures, while higher-level neurons integrate these features into increasingly abstract representations. This enables the brain to recognise objects despite variations in lighting, orientation, and occlusion, allowing for flexible and efficient generalisation.

Understanding these mechanisms provides valuable insights into both human cognition and the development of computational models inspired by biological vision.

1.2.2 Visual Feature Processing in the brain

As briefly highlighted in the previous section, the process of visual perception is grounded in feature analysis, where the brain extracts and processes different visual attributes to form coherent representations. This task is carried out effortlessly to recognize a variety of objects in just a few milliseconds with exceptional precision despite the massive size of input coming through the eyes. It is a fascinating phenomenon attracting the attention of neuroscientists.

The human visual system processes information hierarchically, with increasing complexity at successive stages of the ventral visual pathway. At the earliest stage, in the retina, photoreceptors (rods and cones) detect light intensity and wavelength, transmitting raw visual information to the lateral geniculate nucleus (LGN) of the thalamus. The LGN refines contrast and spatial information before forwarding it to the primary visual cortex (V1), where neurons are tuned to simple features such as oriented edges [2], [3], spatial frequency [4], and direction of motion [5].

As visual information progresses through the mid-level processing stages in V2 and V4, neuronal responses become more complex, encoding attributes like contours, textures, and curvature [6], [7]. V4, in particular, plays a key role in integrating colour, shape, and texture, with neurons responding to curvature and shape fragments [8]. This processing enables the system to maintain representations that are invariant to minor transformations such as position, size, or viewpoint.

Further along the ventral stream, in the inferior temporal (IT) cortex, neurons become highly selective for complex object features and entire object categories. Early electrophysiological studies found that IT neurons in monkeys respond to specific stimuli, such as faces and hands [9]. Subsequent work demonstrated that IT neurons encode specific shape features in a sparse but highly selective manner [10] and that these representations remain stable despite variations in appearance [11]. A longstanding question in visual neuroscience is how object representations are organized within IT. Early studies suggested that neurons selective for different object categories were intermingled with little spatial organization [12]. However, later work revealed that IT exhibits domain-specific clustering, with distinct patches responding to faces, bodies, places, and objects [13],[14]. Representational similarity analysis (RSA) further demonstrated that object categories are not only clustered but also share structured representational relationships [15].

Summarizing this, the ventral visual pathway transforms raw sensory input into meaningful object representations through a progressive buildup of complexity. Beginning with edge detection in V1, feature integration in V4, and categorical selectivity in IT, this hierarchical system enables the rapid and invariant recognition of objects in our environment. Building on this understanding of hierarchical processing, the next question arises: How exactly are these object representations encoded at the neural level?

1.2.3 Encoding of Representations

A central question in understanding neural representations is how they are stored and accessed in the brain. Two dominant theories have emerged: local and distributed representations. In the localist framework, specific neurons or small clusters encode highly specific concepts or objects, an idea epitomized by the hypothetical ”grandmother cell”

[16]. This view gained empirical support from studies [17], which identified neurons in the human medial temporal lobe that responded selectively to particular individuals (e.g., Jennifer Aniston) or landmarks, suggesting sparse and highly selective coding. In contrast, evidence for distributed representations challenges this view. Here, object identity emerges from coordinated activity across large populations of neurons, aligning with principles of population coding, i.e. no single neuron uniquely represents a concept, but many contribute to multiple representations [18]. Functional MRI studies of the ventral temporal cortex reveal that even seemingly category-selective regions (e.g., for faces or places) encode information about other categories within distributed activity patterns [15]. Multi-electrode recordings in non-human primates also show that object identity can be reliably decoded from population responses in the inferior temporal cortex [11], reinforcing the high-dimensional and distributed nature of neural coding. Despite these distinctions, growing evidence suggests that the brain may adopt a hybrid coding strategy. Some neurons exhibit high selectivity (akin to localist coding), while others engage in distributed encoding depending on the context, task demands, or familiarity [19], [20]. This hybrid architecture provides both the robustness and specificity necessary to represent the immense diversity of stimuli encountered in natural vision.

1.2.4 Data-Driven Analysis for Hidden Representations

In understanding the function of the brain, a data-driven approach can reveal hidden representations that are often overlooked when working within the confines of pre-existing theories. Unlike hypothesis-driven methods, which begin with specific hypotheses and seek to confirm or refute them, data-driven methods allow important insights and patterns to emerge naturally from large-scale datasets, making them perfect for exploring the complexity of visual processing in an unbiased manner.

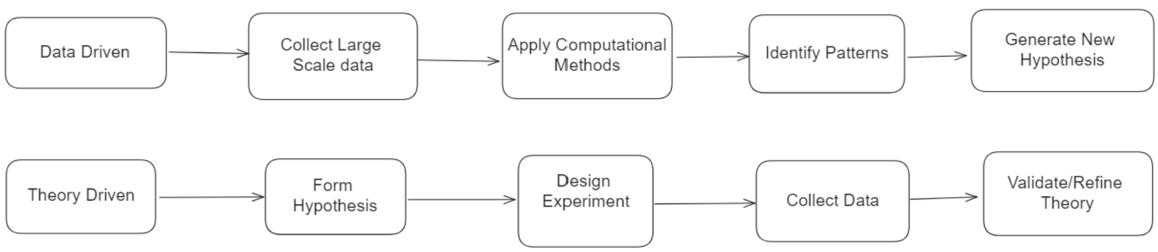


Figure 1.1: Structural diagram of pipelines: theory-driven and data-driven methods

This exploratory technique is not restricted by the researcher's prior assumptions, biases or imagination. For example, in the ventral visual pathway, data-driven analysis has found categorical representations that were not discovered by the classical models. This reveals special patches for object categories that have never been theorized. By focusing on the patterns of neural activity, these methods offer more comprehensive and potentially surprising insights into cortical function, highlighting the richness and diversity of representations involved in visual perception. Before we look at how brain data is collected and analyzed, we must first understand the essential physics behind how neurons fire and communicate.

1.2.5 Physics Behind Neural Firing

Neurons are the fundamental computational units of the brain, responsible for processing and transmitting information via electrical and chemical signals. Their function relies on well-defined physical principles governing electrical potential, ionic movement, and electromagnetic field generation.

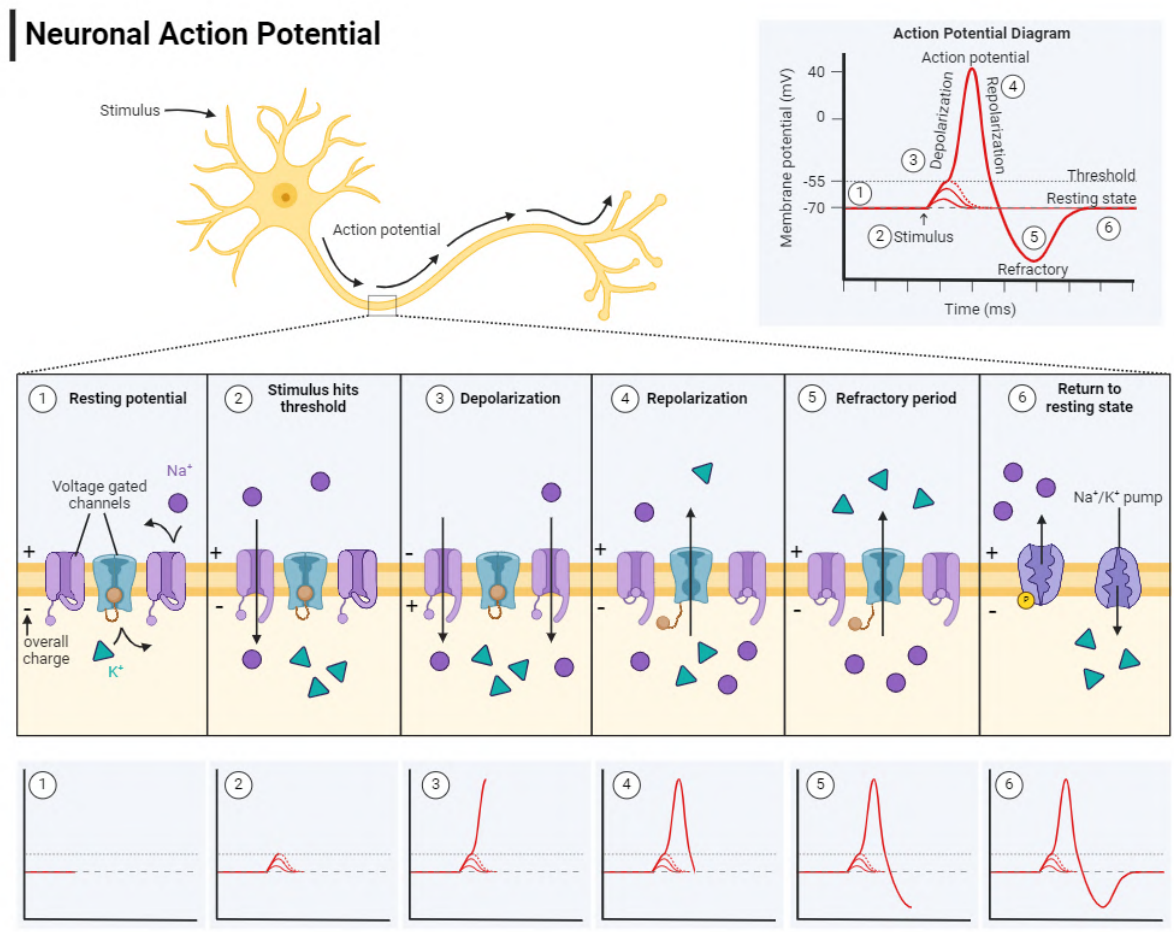


Figure 1.2: A: Structure of a neuron and propagation of action potential. B. Ion channel configurations for sodium and potassium transport. C. Phases of the action potential cycle. *[Image source: Biorender: Sarah Faber]*

Each neuron maintains a resting membrane potential of approximately -70 mV, meaning the inside of the cell is negatively charged relative to the outside. This potential is maintained by the selective permeability of the neuronal membrane and active ion transport mechanisms, primarily the sodium-potassium (Na^+/K^+) pump and leak channels. The pump actively moves three Na^+ ions out and two K^+ ions in, creating both a concentration gradient and an electrochemical gradient across the membrane.

When a neuron is sufficiently stimulated by external input (e.g., from sensory receptors or other neurons), its membrane potential undergoes a rapid change called an action potential. This electrical signal propagates along the axon as the depolarization wave activates neighbouring voltage-gated channels, transmitting the action potential down the neuron's length.

1.2.6 Generation of EM Fields in the Brain

The biophysical basis of electric and magnetic field generation in the brain arises from the ionic currents associated with neuronal activity. When a neuron is activated, voltage-gated ion channels open, leading to the flow of ions across the cell membrane. This ionic movement generates local intracellular and extracellular currents. While individual action potentials are brief and spatially confined, the synchronous activity of large populations of neurons, mainly cortical pyramidal cells aligned in parallel, can result in macroscopic field potentials.

Due to their geometry and orientation perpendicular to the cortical surface, these neurons produce net dipolar currents that summate across space and time. The resulting current flow through the extracellular medium produces electric fields, which can be detected by electrodes placed on the scalp, as in electroencephalography (EEG). The spatial configuration and temporal coherence of underlying neural populations govern the strength and distribution of these electric fields.

According to Maxwell's equations, the movement of electric charge also generates magnetic fields. Although these fields are orders of magnitude weaker than the Earth's magnetic field, the coordinated activity of thousands of neurons can produce a detectable signal outside the head. These magnetic fields propagate through the skull and scalp with minimal distortion and are measurable using superconducting quantum interference devices (SQUIDs), which are used in magnetoencephalography (MEG). Unlike electric fields, which are significantly influenced by the conductivity of intervening tissues, magnetic fields retain a relatively direct correspondence to the underlying neural currents.

Thus, both electric and magnetic fields generated by neuronal activity provide a window into the dynamic processes of the brain and form the foundation for several noninvasive neuroimaging techniques.

1.2.7 Neural Encoding

The brain encodes information using spatial and temporal mechanisms to ensure efficient sensory processing, motor control, and cognition. Spatial encoding relies on population coding, where information is distributed across groups of neurons rather than individual units, enhancing robustness and redundancy. This is evident in sensory systems, where neurons in the visual cortex respond to edges, colours, and motion, and in motor control, movement direction and force are encoded collectively. Neural activity is also spatially organized into topographic maps, such as retinotopic maps in vision and somatotopic maps in touch perception. Temporal encoding, on the other hand, conveys information through the timing of neural activity. Rate coding reflects stimulus intensity, synchronization of neuronal spikes supports attention and movement planning, and phase coding, as seen in hippocampal theta oscillations, aids spatial navigation and memory formation. These encoding strategies allow the brain to integrate and interpret complex information in real-time.

1.2.8 Measuring Brain Activity

Neural activity can be measured through different physiological signals, primarily electrical, magnetic, and hemodynamic responses. Each method offers distinct advantages and limitations, making them suitable for different research applications.

Electromagnetic vs. Hemodynamic Signals

Electromagnetic signals capture the electrical activity of neurons in real-time. When neurons fire, they generate electric fields, which can be recorded using electroencephalography (EEG). These electrical signals also induce weak magnetic fields, which can be measured with magnetoencephalography (MEG). Both EEG and MEG provide high temporal resolution and can detect neural changes on the millisecond scale, making them ideal for studying rapid cognitive processes such as visual perception. However, their spatial resolution is more limited due to challenges in accurately localizing the sources of these signals within the brain.

In contrast, hemodynamic signals, such as those measured by functional magnetic resonance imaging (fMRI), rely on changes in blood oxygenation levels (BOLD response). When a brain region becomes active, local blood vessels dilate to increase oxygen supply, leading to measurable changes in blood flow. This method provides excellent spatial resolution, allowing researchers to pinpoint which brain regions are involved in a given task. However, blood oxygenation changes occur on a slow timescale, with a delay of several seconds relative to neural activity. This makes fMRI unsuitable for studying fast processes like visual perception, where neural responses occur within milliseconds.

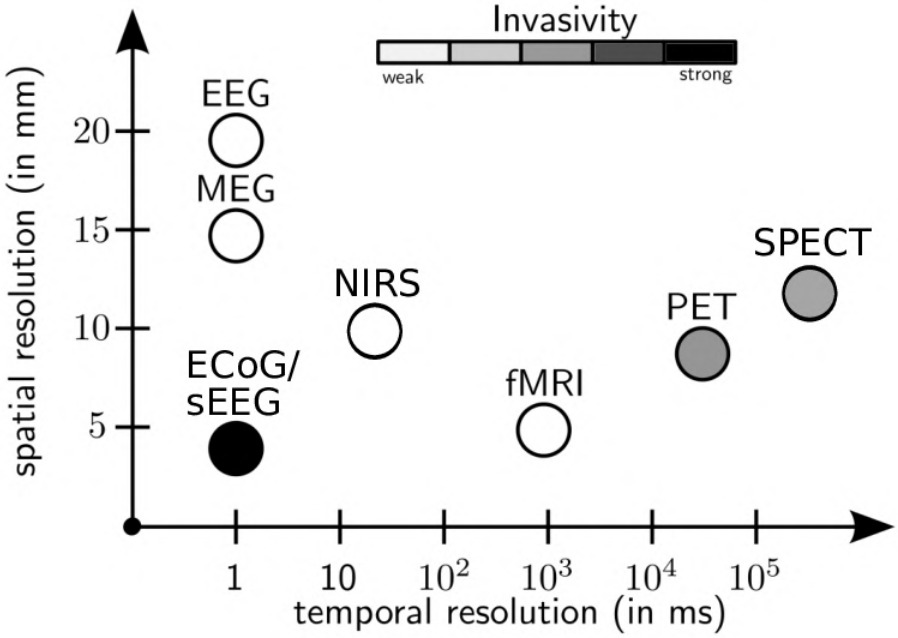


Figure 1.3: Comparison of spatial and temporal resolution of different brain imaging techniques *[Source: Sebastian Hitziger 2015]*

Why Electromagnetic Methods for Visual Tasks?

Since visual tasks evolve in a matter of milliseconds, capturing transient neural responses requires a method with high temporal resolution. Hemodynamic imaging is too slow to track the precise timing of neural events, whereas EEG and MEG allow us to measure neural responses in real time. MEG, in particular, offers an advantage over EEG in spatial localization due to its ability to measure magnetic fields with minimal distortion from the scalp and skull.

Comparing Different EM Methods

- MEG (Magnetoencephalography): Measures magnetic fields generated by neuronal activity. Magnetic fields pass through the scalp and skull without distortion, allowing for more accurate source localization. However, MEG requires expensive, magnetically shielded rooms and superconducting sensors, making it less accessible.
- EEG (Electroencephalography): Measures electrical potentials on the scalp. EEG is more affordable and widely available but suffers from signal distortion due to the conductivity of the skull and scalp, making source localization more challenging.

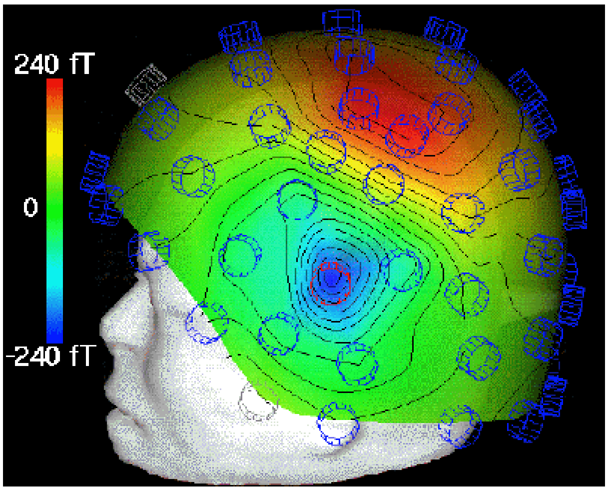


Figure 1.4: Magnetoencephalography (MEG) sensor array for detecting sources of electrical activity within the brain. The contours correspond to the magnitude of the field induced by the source. *[Source: CTF Systems Inc.]*

1.2.9 Magnetoencephalography

How Neurons Generate Magnetic Fields

Magnetoencephalography (MEG) measures the magnetic fields produced by neuronal activity. These fields originate from the movement of ions such as Na, K, Cl, and Ca^2 across neuronal membranes during action potentials and synaptic transmission. The resulting electrical currents can be classified into:

- Primary Currents: Postsynaptic currents flow along the dendrites of pyramidal neurons in the cerebral cortex when they receive input. These currents generate small but measurable magnetic fields.
- Secondary Currents: The primary currents induce weaker return currents in the surrounding tissue, further influencing the brain's overall electromagnetic activity.

As described by Maxwell's equations, the movement of these electric currents produces magnetic fields that extend beyond the scalp. However, because neural activity occurs in a complex three-dimensional structure, only specific current orientations produce detectable external fields. MEG is particularly sensitive to tangential currents in the sulci of the cerebral cortex, where pyramidal neurons align in parallel.
The net effect of synchronised activity in large neuronal populations produces a macroscopic magnetic field that MEG can record with high temporal precision.

Strength of Magnetic Fields in MEG

Neuronal magnetic fields are remarkably weak, measuring only 10–100 femtoteslas (fT) (1 fT = 10^{1} T), several orders of magnitude smaller than the magnetic fields encountered in everyday environments. For comparison, the Earth's magnetic field is approximately 50 microteslas (µT) (1 µT = 10 T), making it nearly a billion times stronger than the signals generated by neural activity. Additionally, environmental noise from power lines, electronic devices, and urban infrastructure produces magnetic fields far exceeding the strength of neuronal signals. Because MEG relies on detecting these minuscule brain-generated fields, recordings must occur in magnetically shielded rooms to eliminate external interference and enhance signal clarity.

How MEG Measures Brain Activity

To detect the brain's extremely weak magnetic fields, MEG relies on Superconducting Quantum Interference Devices (SQUIDs), highly sensitive magnetometers that measure minute changes in magnetic flux. SQUID sensors operate using superconducting loops that detect even the slightest fluctuations in magnetic fields. To maintain superconductivity and ensure optimal sensitivity, these devices function at cryogenic temperatures of approximately 269°C, cooled by liquid helium. A typical MEG system consists of 200–300 SQUID sensors arranged around the head to capture signals from multiple brain regions. While this dense sensor array enables detailed mapping of neural activity, precise source localisation requires computational techniques. Since MEG measures the magnetic fields generated by neuronal currents rather than direct neural firing, source localisation algorithms such as beamforming and minimum-norm estimation are employed to infer the precise origin of brain activity. Additionally, because neural signals are orders of magnitude weaker than environmental magnetic noise, MEG recordings take place in magnetically shielded rooms, which minimises external interference and improves signal clarity.

Advantages of MEG

MEG is a non-invasive, high-temporal-resolution method ideal for investigating real-time brain activity. Unlike hemodynamic techniques such as fMRI, which have slow sampling rates, MEG can detect neural dynamics at the millisecond scale, making it particularly useful for studying cognitive and perceptual processes. Additionally, because magnetic fields pass through biological tissue with minimal distortion, MEG offers better source localisation than EEG. These advantages make MEG a crucial tool for exploring neural mechanisms underlying cognition, perception, and neurological disorders.

Methods of Analyzing MEG

In magnetoencephalography (MEG) analysis, source localisation and matrix factorisation methods are two widely used approaches for extracting neural signals. While both aim to infer the underlying brain activity from external recordings, they differ in their conceptual frameworks and the nature of the outputs they provide. Source localisation methods attempt to solve the inverse problem by estimating where the measured magnetic fields originate in the brain. Techniques such as Minimum Norm Estimation (MNE), beamforming, and dipole fitting are commonly employed, making specific assumptions about neural sources' spatial configuration and strength. These methods offer high spatial resolution but are sensitive to modelling assumptions and noise characteristics. In contrast, matrix factorisation methods like Independent Component Analysis (ICA) and Non-negative Matrix Factorization (NMF) take a data-driven approach, decomposing the recorded MEG signals into statistical components without relying on anatomical priors. These techniques are beneficial for identifying functional patterns or separating noise and artefacts from meaningful brain signals. However, they do not provide direct localisation of brain activity. For this study, we adopt matrix factorisation methods, allowing us to explore dominant patterns in the data without imposing strong a priori constraints on source structure.

1.2.10 Latent dimensions

Many complex systems generate high-dimensional data, but a lower-dimensional structure often governs their behaviour. These fundamental underlying variables, known as latent dimensions, capture the essential patterns that drive the system's dynamics. An analogy comes from classical physics: Consider a double pendulum, where two arms move in a seemingly chaotic fashion. Although its motion appears highly complex, it is ultimately

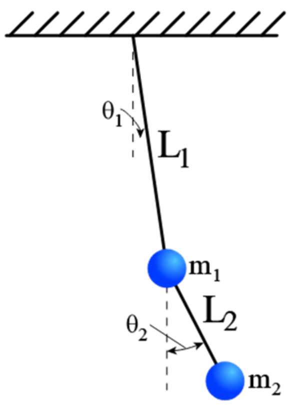


Figure 1.5: Double Pendulum

determined by a few independent variables, such as the angles and angular velocities of the two arms. Similarly, while thousands of neurons fire simultaneously in neural activity, their activity is not entirely independent. Instead, neurons tend to fire in coordinated patterns, reducing the actual degrees of freedom in the system. This suggests that high-dimensional neural data, such as MEG recordings, may have a lower-dimensional intrinsic structure that reflects the brain's underlying computational principles.

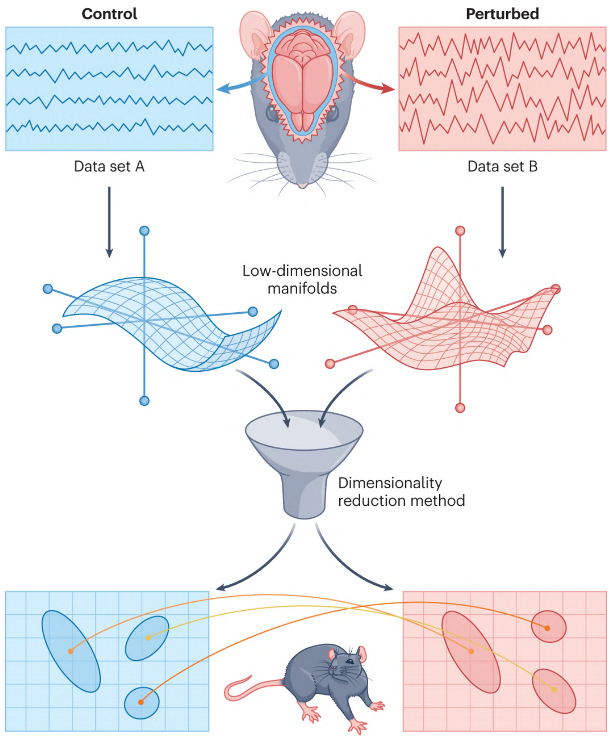


Figure 1.6: Schematic illustrating how dimensionality reduction reveals changes in neural dynamics between control and perturbed brain states in a rodent model.

To extract these latent dimensions, we turn to factorization and dimensionality reduction methods, which help simplify complex datasets by identifying meaningful co-activation patterns. Principal Component Analysis (PCA) identifies orthogonal axes that explain the most variance in the data, providing an efficient representation of neural activity. Independent Component Analysis (ICA) separates statistically independent sources, often helpful in isolating neural signals from artefacts. Non-negative Matrix Factorization (NMF) decomposes data into additive components, making it particularly suited for extracting interpretable patterns of neural activity. Additionally, clustering methods and dictionary learning provide alternative approaches to discovering meaningful low-dimensional representations. By applying these techniques to MEG data, we aim to uncover structured patterns in neural dynamics, offering insight into how different object categories are represented over time.

1.3 Previous Literature

Non-negative Matrix Factorization (NMF) has gained significant attention in recent years as a powerful technique for dimensionality reduction, offering advantages over traditional methods like Principal Component Analysis (PCA) by incorporating a non-negativity constraint [21]. This constraint leads to a parts-based representation of the data, meaning that components learned through NMF tend to represent additive and interpretable parts of the whole. One key strength of NMF lies in its ability to enforce sparseness in the resulting components. This sparseness helps isolate distinct features or categories in the data, ensuring that different components are more equally distributed and specialised, in contrast to PCA, where a few principal components often capture most of the data variance, leading to less interpretable features [21].

A comparative study of NMF and PCA applied to facial expression datasets highlights this distinction clearly. While PCA captures broad variance patterns, the learned features are distributed and difficult to localise. In contrast, NMF isolates distinct facial parts, such as eyes, mouths and other facial features across components. Additionally, the study demonstrated that NMF significantly outperformed PCA in terms of recognition accuracy [22].

Figure 1.7: NMF basic images components on face dataset *[Source: Zhao 2018]*

Interestingly, initialising NMF with PCA-derived matrices did not diminish its superior recognition capabilities, further suggesting that the non-negativity and sparsity constraints in NMF play a critical role in yielding more meaningful and discriminative features. Together, these results support the growing use of NMF in domains where interpretability and parts-based decomposition are desirable, particularly in high-dimensional datasets such as images or neural recordings [22].

The growing interest in understanding high-dimensional neural representations of object categories has led to the development and application of large-scale, multimodal datasets such as the THINGS database [23]. The study by Hebart et al [23], [24], utilised this dataset by combining both MEG and fMRI modalities to analyse the time-resolved neural dynamics and spatial activation patterns underlying object perception [24]. Pairwise decoding analyses performed on the MEG data revealed that component-wise representations of different object categories, such as animate and inanimate, could be reliably distinguished in time, underscoring the discriminability and informativeness of the evoked neural signals [24].

Building on this, Teichmann et al. proposed a behavior-guided approach to analysing object representations in the brain using the same THINGS dataset [23]. By extracting multidimensional object properties from millions of behavioural similarity judgments, they modelled single-trial MEG responses to over 26,000 images. Their data-driven analysis revealed that each behaviorally derived object dimension is reflected in the MEG signal. The components fall into two general temporal profiles: early visual dimensions peaking around 125 ms and later conceptual dimensions peaking closer to 300 ms. Early effects were consistent across participants, whereas late effects varied, suggesting that conceptual features may be more individually tuned [25].

Similar insights come from a separate fMRI study by Khosla et al., which used Non-negative Matrix Factorization (NMF) to find latent neural components responsive to object categories in the ventral visual pathway [26]. In contrast to conventional voxel-wise hypothesis-driven approaches, this method recovered distinct, spatially distributed components that included known category selectivity (e.g., faces, bodies, words), as well as a previously uncharacterised component selectively responsive to food images. Interestingly, this food-selective component could not be explained by lower-level features such as colour or texture, indicating the presence of high-level categorical tuning.

Together, these studies demonstrate the potential of data-driven methods like NMF, paired with rich behavioural datasets, to decompose and characterise the neural architecture underlying complex object representations across time and space.

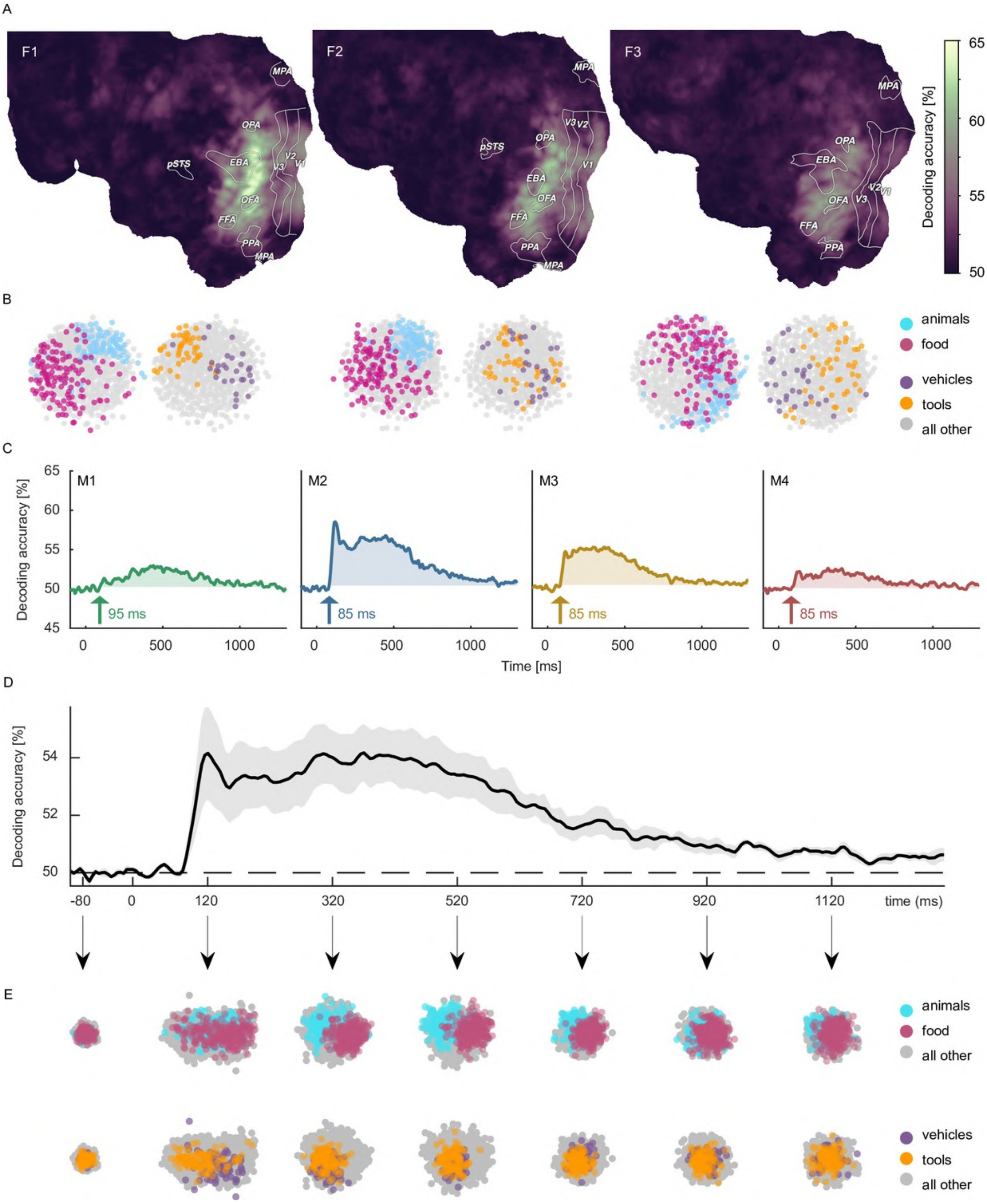


Figure 1.8: A: Pairwise decoding accuracy of fMRI data across 3 participants visualised on the cortical surface in the THINGS dataset B: fMRI response across channels with segregation between categories C: Pairwise decoding accuracy for MEG data across 4 participants D: Group average of the subject-wise MEG decoding accuracy E: Emergence of separate category clusters with time highlighting differential responses *[Source: Hebart 2023]*

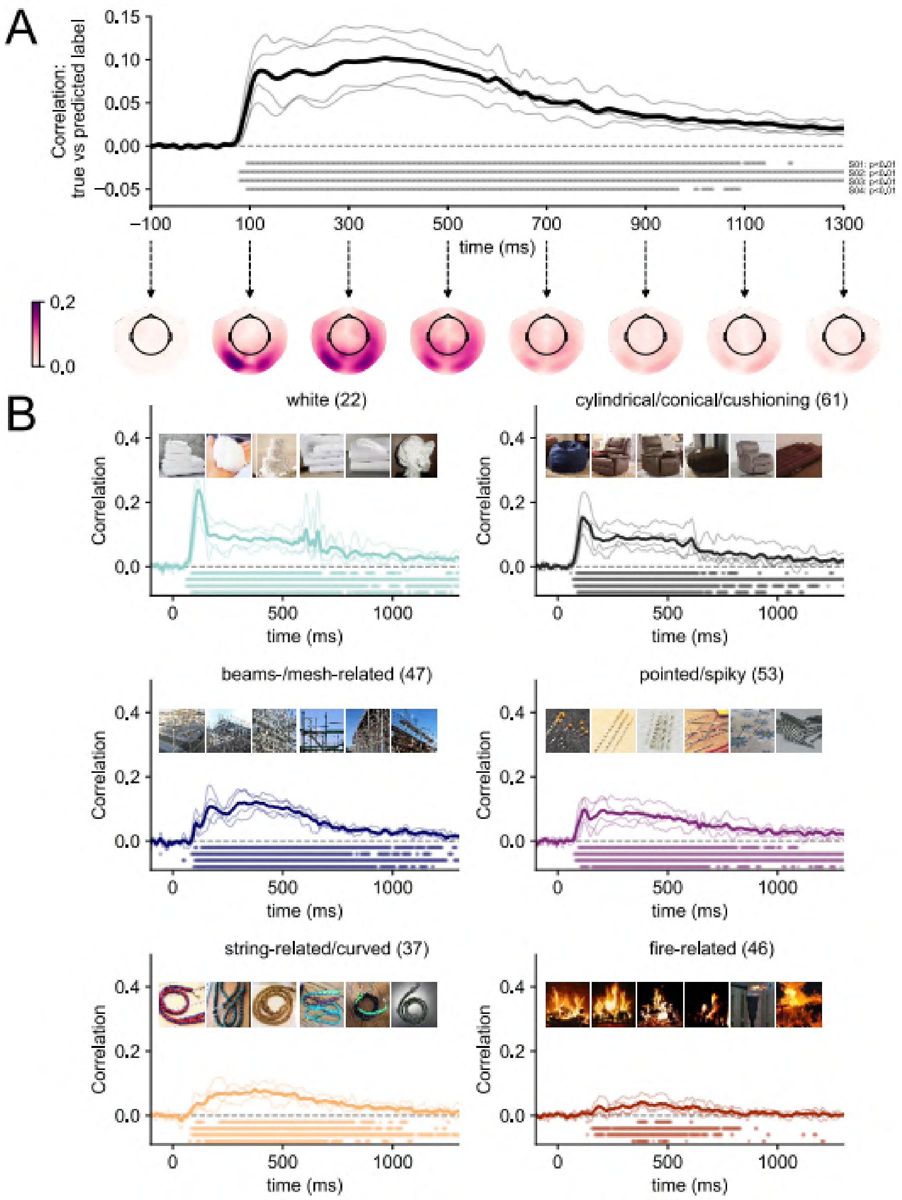


Figure 1.9: A: Correlation between the predicted and true behavioural embeddings across all dimensions over time. B: Example time dynamics for the 6 dimensions
[Source: Teichnann 2024]

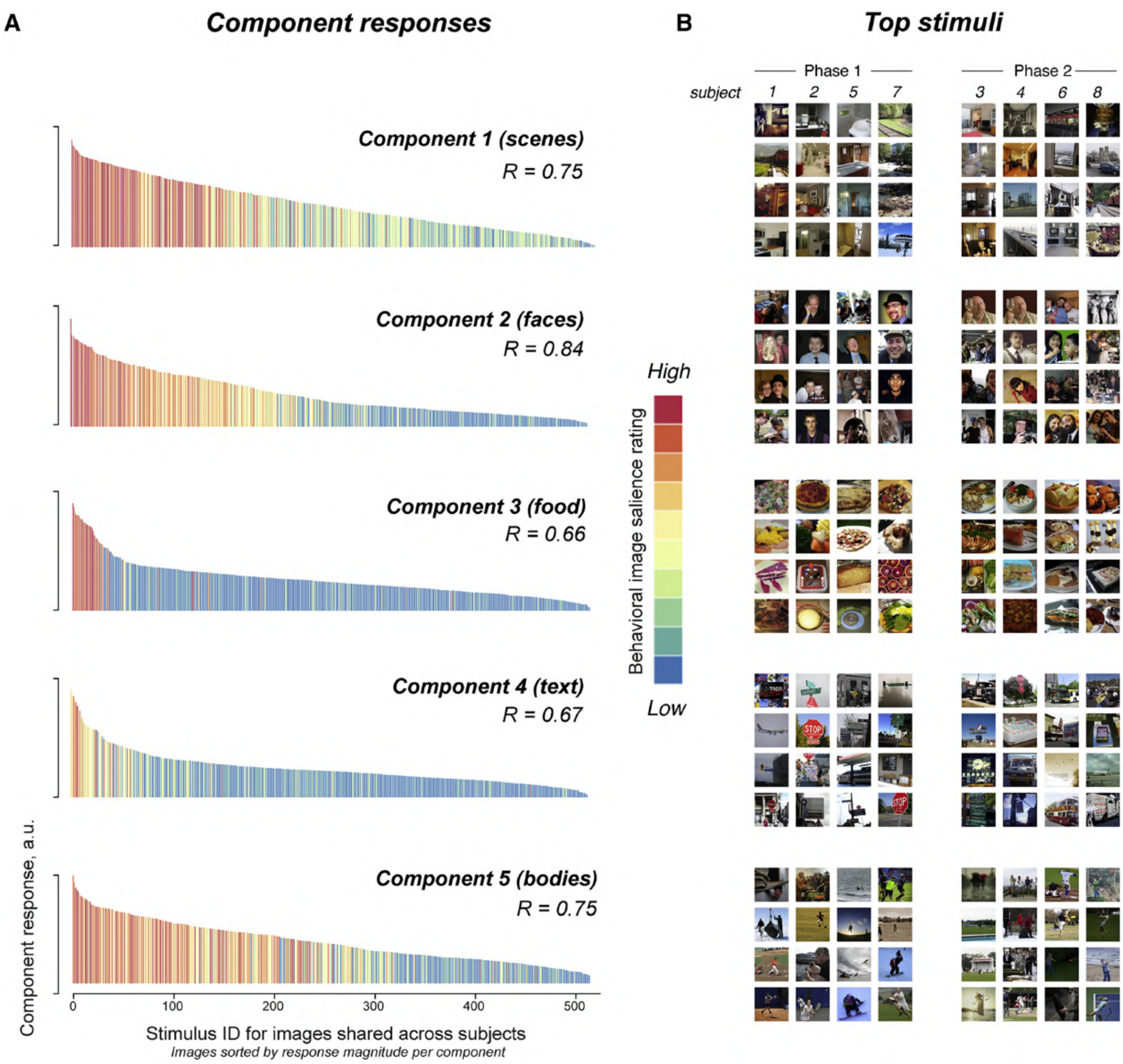


Figure 1.10: A: Responses of components of each stimulus image B: Examples images which load highest on each component *[Source: Khosla 2022]*

Chapter 2

Methods

2.1 Dataset

2.1.1 Ethics

The dataset used in this work is openly available online from Hebart et al [23], [24]. This research was conducted in accordance with ethical guidelines and regulatory requirements and was approved by the NIH Institutional Review Board under protocol 93-M-0170 (NCT00001360). All participants provided informed consent for both participation and data sharing and received financial compensation for their involvement. The MEG study included four healthy adult volunteers (2 female and 2 male; mean age at study onset: 23.25 years). The sample size (n = 4) was determined in advance as a balance between data quality and the substantial effort required for data collection. All participants were right-handed, had normal or corrected-to-normal visual acuity, and had prior experience in experiments requiring sustained visual fixation. Participants were screened for suitability and availability prior to inclusion in the study [23].

2.1.2 THINGS

The THINGS dataset is a large-scale, multimodal neuroimaging and behavioural resource designed to investigate object representations in the human brain. It consists of 1,854 diverse object categories, each represented by 12 distinct naturalistic images, totalling 22,448 unique images. The dataset includes magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and extensive behavioural similarity judgments, enabling comprehensive analyses across modalities.

One of the key advantages of THINGS is its wide usage across multiple studies, facilitating cross-comparison and reproducibility. In particular, the availability of the

THINGS-similarity matrix quantifying pairwise behavioural similarity across all object categories offers a robust ground for model validation. Previous research has applied encoding and decoding models (e.g., regression-based approaches) on MEG and fMRI data to uncover the representational structure of object concepts, yielding compelling results (Previous Literature). Applying Non-negative Matrix Factorization (NMF) to this dataset represents a novel approach to uncovering the latent dimensions of object representations, offering interpretability and potential alignment with behavioural and neural similarity structures.

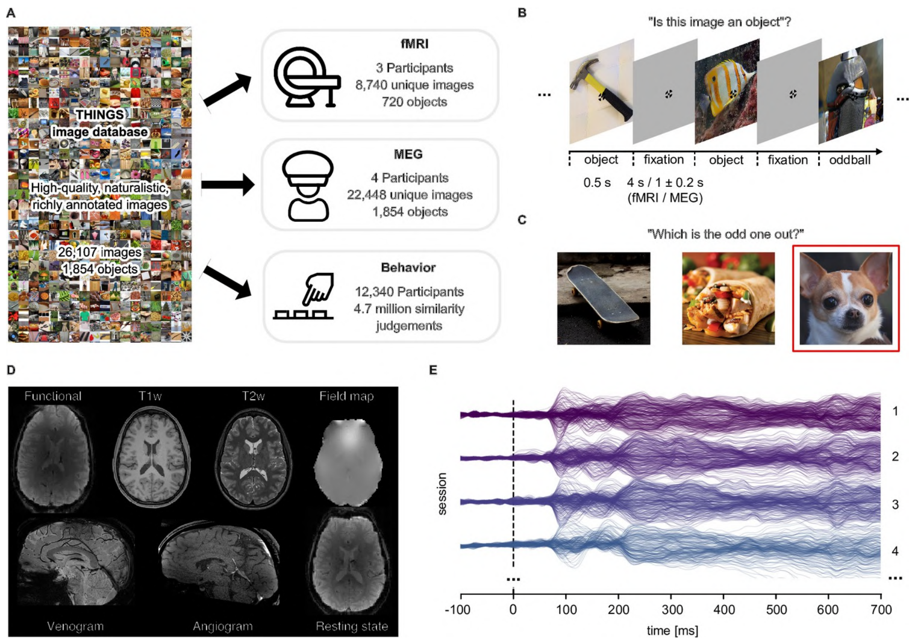


Figure 2.1: THINGS Dataset comprising of MEG, fMRI and behavioural responses to large samples of object images. *[Source: Hebart 2023]*

In the MEG portion of the dataset, participants were shown the 22,448 object images over 12 experimental sessions. Each image was presented briefly, and participants performed an oddball detection task to ensure sustained attention. Specifically, a subset of 200 images was shown repeatedly across sessions to estimate noise ceilings and serve as a test set for model evaluation.

Each MEG participant completed 12 sessions, with each session consisting of 10 runs of approximately 5 minutes. Within each run, 185–186 object images were presented, along with 20 test images and 20 catch (oddball) images. Stimuli were displayed for 500 ms, followed by a variable fixation interval of 1000 ± 200 ms (stimulus onset asynchrony: 1500 ± 200 ms). This temporal jitter was designed to reduce alpha-band synchronization effects related to trial onset. Overall, each run contained 225–226 trials, resulting in 2,254 trials per session and a total of 27,048 trials per participant.

2.1.3 MEG Data Acquisition

MEG data was recorded using a CTF 275-channel system equipped with radial first-order gradiometer/SQUID sensors. Recordings were conducted inside a magnetically shielded room to minimize environmental noise. Data were sampled at 1,200 Hz, with third-order gradient balancing applied online to suppress background interference further. Participants remained seated throughout the recording sessions to ensure signal stability. Due to malfunction in three MEG channels (MLF25, MRF43, and MRO13), the final dataset consisted of 272 functional MEG channels.

2.1.4 Preprocessing

Preprocessing was performed using the MNE-Python toolbox. The raw MEG signals were bandpass-filtered between 0.1 and 40 Hz to remove slow drifts and high-frequency noise. Data were then epoched around stimulus onset, and baseline correction was applied by subtracting the mean and dividing by the standard deviation within a 100 ms pre-stimulus baseline period. The epoched data were downsampled to 200 Hz to reduce computational complexity in downstream analyses. As a result, the final dataset is organized as:

$$1854 \text{ categories} \times 12 \text{ images} \times 271 \text{ channels} \times 281 \text{ timepoints}$$

2.1.5 Postprocessing

Prior to analysis, preprocessing steps were applied to standardize and clean the MEG data. Each channel was mean-centered to zero, and extreme outliers, defined as exceeding five standard deviations from the mean, were removed to mitigate artifacts. Following this, the data were shifted such that all values were strictly positive, ensuring numerical stability for subsequent analytical methods.

To enhance the signal quality for each object concept, MEG responses were averaged across all images belonging to a single category. This category-wise averaging increased the signal-to-noise ratio, allowing for a more robust representation of neural activity associated with each object concept.

Given the focus on visual processing, 39 occipital MEG sensors were selected for analysis, isolating neural responses primarily driven by the visual system. This sensor selection ensured that the extracted signals were predominantly influenced by early and mid-level visual processing regions. Further, the Non Negative Matrix Factorization requires a 2D input. There are two common ways to transform the data:

1. $(n_{\text{channels}},\ n_{\text{samples}} \times n_{\text{timepoints}})$: This method treats, each channel like a separate source of signal focusing on temporal patterns across all trials and timepoints for each channel. NMF extracts temporal components that are common across sensors. It is useful for isolating time-evolving data structures. But it does not preserve trial identity.

2. $(n_{\text{samples}},\ n_{\text{channels}} \times n_{\text{timepoints}})$: This method has a trial centric view. It Flattens out channels and timepoint activity into a single dimension by concatenation. NMF identifies spatiotemporal components that capture recurring trial wise patterns across sensors. It is useful for studying trial-by-trial variability while preserving spatial and temporal structure as the feature dimension.

In our scenario, we decide to transform our data such that the timepoints across all channels are concatenated across a single dimension. Fig: Thus our data has the shape:

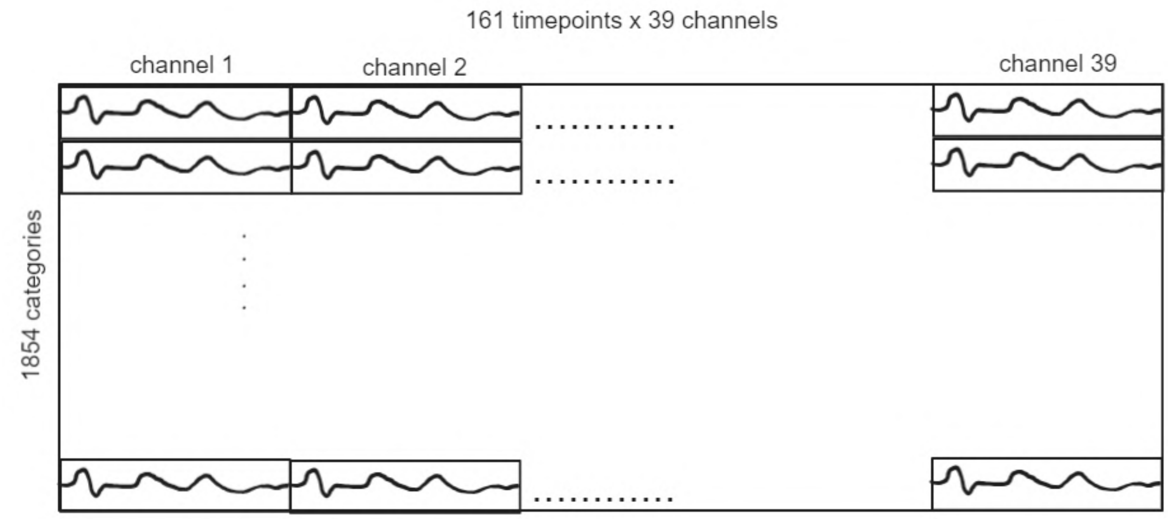


Figure 2.2:

$$1854 \text{ categories} \times (39 * 161) \text{ spatial-temporal dimension}$$

2.2 Non-Negative Matrix Factorization

To reveal latent structures in neural representations, we apply Non-Negative Matrix Factorization (NMF), a dimensionality reduction technique that decomposes the data into a set of interpretable, non-negative components. NMF allows us to express the recorded MEG data in terms of a smaller number of underlying features, facilitating the discovery of structured neural representations across object categories and time.

2.2.1 Mathematical Formulation

Given an input data matrix X, NMF seeks to approximate it as a product of two lower-dimensional matrices, W and H:

$$X \sim W \times H$$

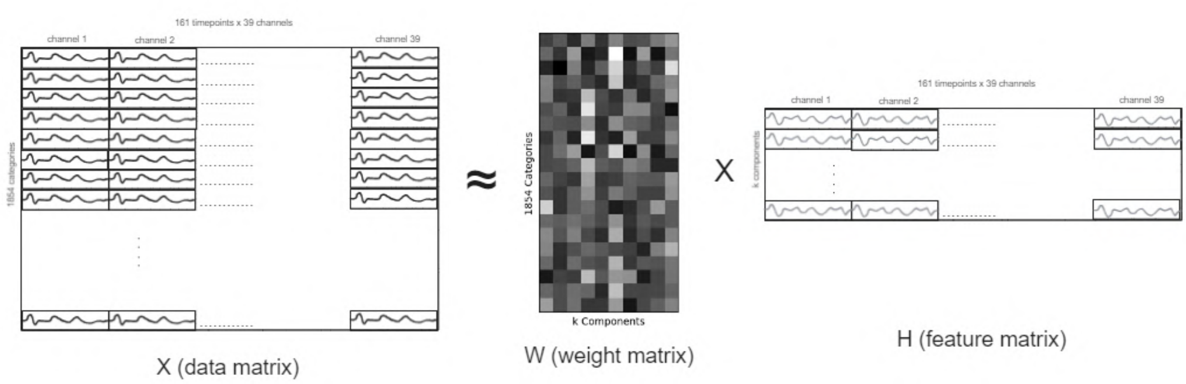


Figure 2.3: Data matrix decomopsition under NMF

where:

- X is the original data matrix with dimensions (categories, timepoints x channels), representing the MEG response for each object category across time and spatial channels.
- W is the basis matrix (categories , components) that captures how strongly each object category is associated with the learned components.
- H is the coefficient matrix (components , timepoints x channels) that describes the temporal and spatial features of each component.

Such that $\min || X - WH ||^2$ is satisfied.

2.2.2 Interpretation of W and H Matrices

- H (Feature Matrix): Represents the discovered latent features, capturing shared temporal-spatial patterns across object categories. Each row in H corresponds to a specific component and describes how it varies across time and MEG channels.
- W (Weight Matrix): Assigns category-specific weights to these features, determining how strongly each object category expresses a given component. Higher weights indicate a stronger contribution of a component to a particular category.

2.2.3 Optimum Number of Dimensions

Selecting the optimal number of components, denoted as k, is a crucial step in Non-negative Matrix Factorization (NMF). This value determines the level of abstraction and the granularity with which the original data is represented. A smaller k yields a more compressed, generalized representation, potentially capturing higher-level structure at the cost of fine details. Conversely, a larger k can preserve more of the original information but risks overfitting and reduced interpretability.

There is no universally correct choice for k; it often depends on the balance between reconstruction accuracy, computational efficiency, and interpretability. Several methods exist for determining the optimal k, including:

- Reconstruction Error Minimization: Choosing k that minimizes the Frobenius norm between the original matrix and its reconstruction.
- Cross-Validation: Partitioning the data into training and validation sets and selecting the k that best generalizes.
- Information-Theoretic Criteria, such as the Bayesian Information Criterion (BIC).

2.2.4 Bayesian Information Criterion (BIC)

The BIC is an objective metric that balances model fit with model complexity. It is defined as:

$$\mathrm{BIC} = k \log(n) - 2 \log(L)$$

- k is the number of free parameters in the model (in NMF, this includes both the entries in W and H),
- n is the number of data points,
- L is the likelihood of the data given the model.

Since NMF is typically posed as a matrix factorization problem rather than a probabilistic one, we approximate the log-likelihood $\log(L)$ by assuming a Gaussian noise model for reconstruction error. That is:

$$\log(L) \propto -\frac{1}{2\sigma^2} \|X - WH\|_F^2$$

where X is the original data matrix and $\|\cdot\|_F$ is the Frobenius norm. Thus, minimizing BIC favors models that fit the data well (low reconstruction error) while penalizing models with excessive complexity (large k).

By computing the BIC across a range of k values, we select the dimensionality at which this trade-off is optimized. This approach allows for a principled, data-driven determination of how many components best summarize the underlying structure in the data without overfitting.

2.3 Consensus Approach

The Consensus NMF approach is used to improve the robustness and reliability of matrix factorization by using multiple runs of Non-Negative Matrix Factorization (NMF). Instead of relying on a single decomposition, this method performs multiple NMF replicates and then finds a stable, consensus-based factorization by clustering across solutions. This helps mitigate variability that arises from different initializations of NMF and ensures that the extracted components are more consistent and interpretable.

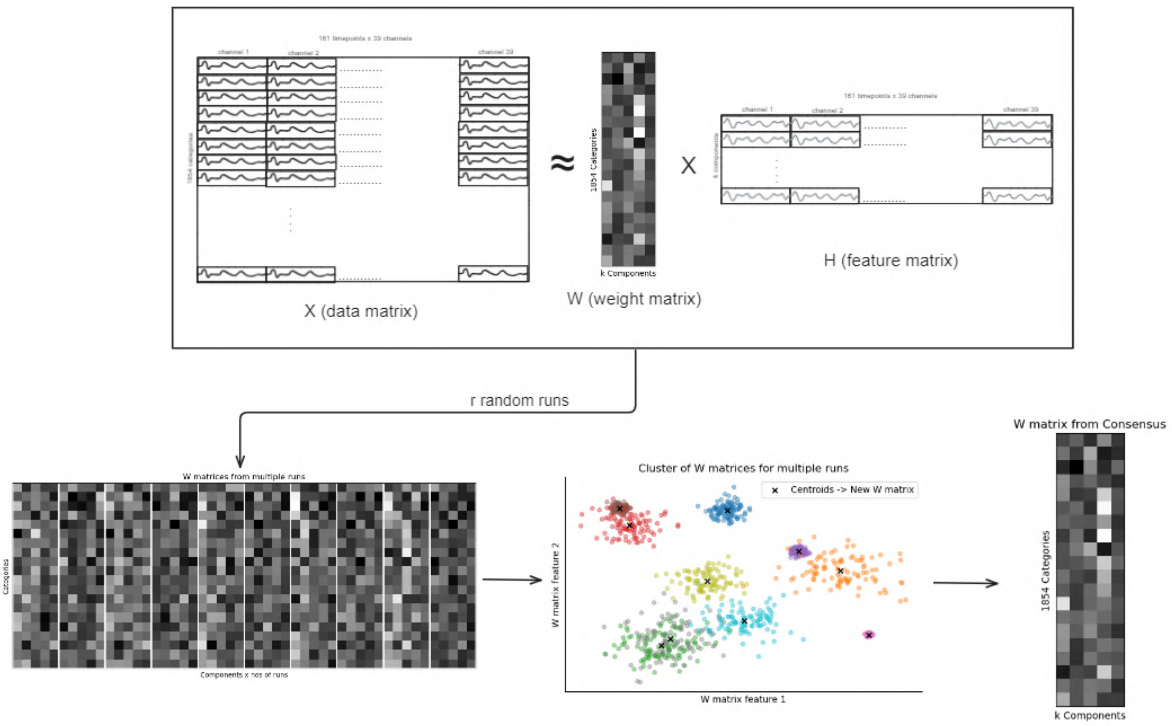


Figure 2.4: Multiple runs of NMF algorithm produce many W matrices. K means cluster is applied on them to find the most stable components

Step 1: Multiple NMF Replicates

Since NMF can produce slightly different results depending on initialization, we first run the algorithm multiple times (n replicates) on the same dataset. Each run results in different factorized matrices (W and H), capturing slightly different decompositions. A probabilistic inference approach replaces standard NMF if a Bayesian NMF is used.

Step 2: Normalization and Aggregation

To compare across replicates, we normalize the resulting W matrices, ensuring they are on the same scale. These matrices are then aggregated into a single dataset, forming an extended version of W, where all replications are stacked together. This creates a larger pool of candidate components from which stable patterns can be extracted.

Step 3: Clustering to Find Consensus Components

Instead of selecting components arbitrarily, we apply clustering (e.g., K-Means) to identify similar patterns across the multiple NMF runs. By grouping components that appear consistently across different solutions, we form more robust and reproducible features. The consensus representation of each cluster is then computed using median values, ensuring that the final W matrix is representative of the most stable patterns.

Step 4: Estimating the Final H Matrix

Once a consensus W matrix is established, we solve for H using a linear regression approach. The goal is to find the best coefficients that reconstruct the original data X given the consensus-based W matrix. This ensures that the final H matrix aligns well with the input data while preserving the structure found in the consensus components.

This approach improves the stability and reliability of extracted components using NMF by clustering across multiple factorizations. This is especially useful for noisy, high-dimensional data like MEG, ensuring more consistent and interpretable neural representations.

2.3.1 Quantifying goodness of cluster components

In order to measure how good a cluster is, we use two quantities: the number of components inside a cluster and the mean of pairwise correlation of each component inside that cluster. It is defined as:

$$\text{MPC} = \sum_{i=0}^{N} \sum_{j \neq i}^{N} \text{corr-coeff}(c_i, c_j)$$

Clusters with a high Mean Pairwise Correlation (MPC) indicate high consistency between the components inside that cluster, making it a good cluster. Additionally, if the size of the cluster N is significant, then we can classify it as a stable cluster.

2.4 Time Dynamics of Each Component

2.4.1 Non-Negative Least Squares Regression

Non-Negative Least Squares (NNLS) is an optimization problem where the goal is to find a vector x that minimizes the squared error between Ax and a target vector b, subject to the constraint that all elements of x are non-negative. Mathematically, it is formulated as:

$$\text{minimize}_{x \geq 0} \quad \|Ax - b\|_2^2$$

where:

- A is a real (m, n) known data matrix
- b is real observation column vector of size m
- x is the solution vector with the constraint that $x_i \geq 0$ for all values of i

We use this method to calculate the time dynamics of each component with time.

Chapter 3

Results

3.1 Preprocessed Data

Preprocessed Magnetoencephalographic (MEG) recordings for all 4 participants include 1854 categories, 12 images per category, 271 channels, and 281 time points.

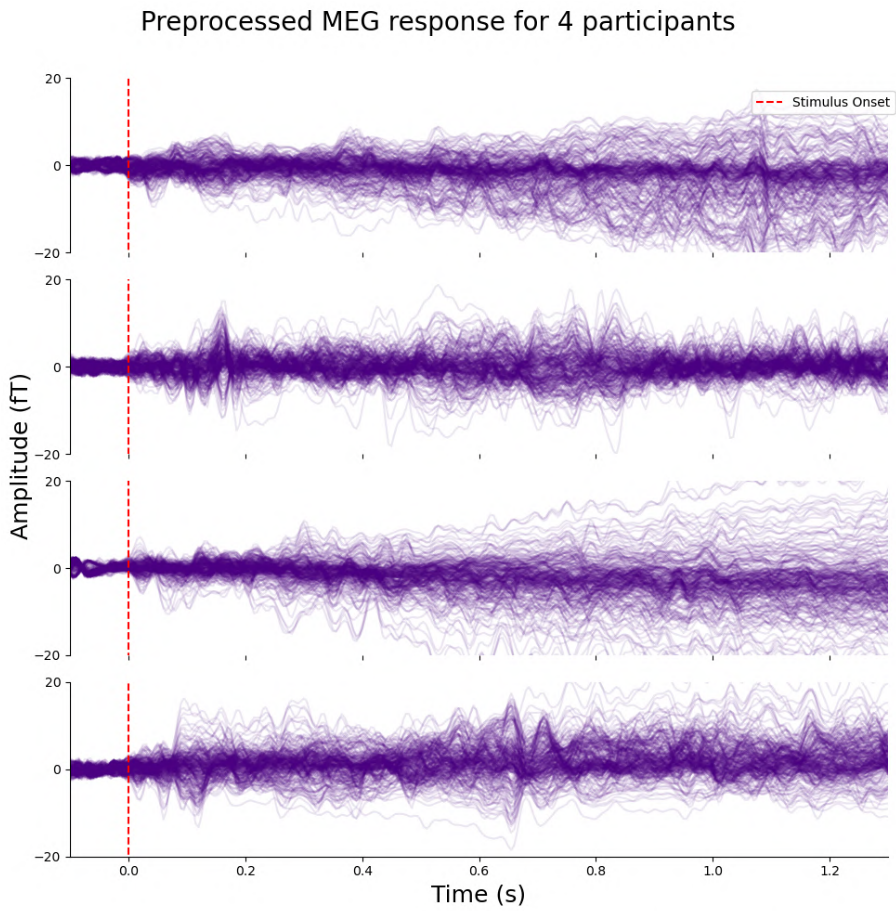


Figure 3.1: Preprocessed data for a category in all 4 participants

3.2 Post Processed Data

Post processing involves choosing MEG data from occipital channels and a time window of 0 to 800 ms. Additionally, the data is averaged over all the objects in a given object category. See the Methods section for additional details.

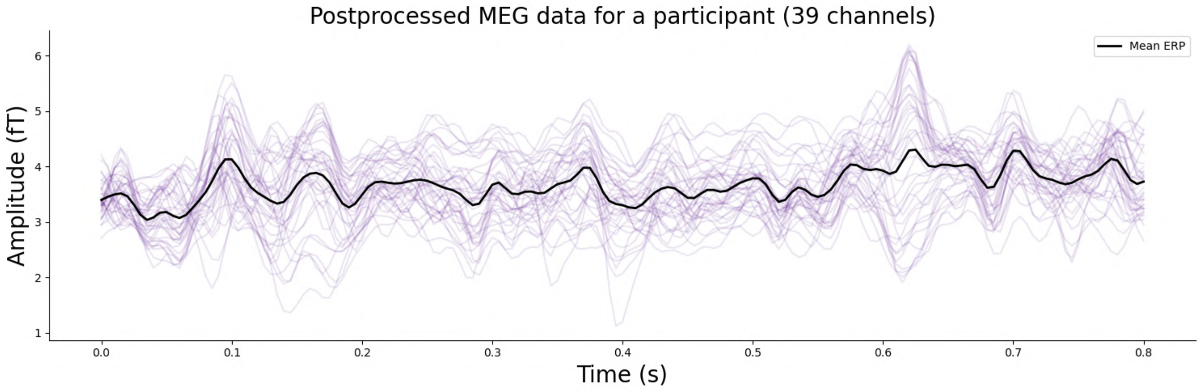


Figure 3.2: Post processed data for participant 1 for a single object category

3.3 Optimum Rank: Number of components

Using the Bayesian Information Criterion, we find the optimum number of components to maximize categories separated without overfitting and redundancy in components. We

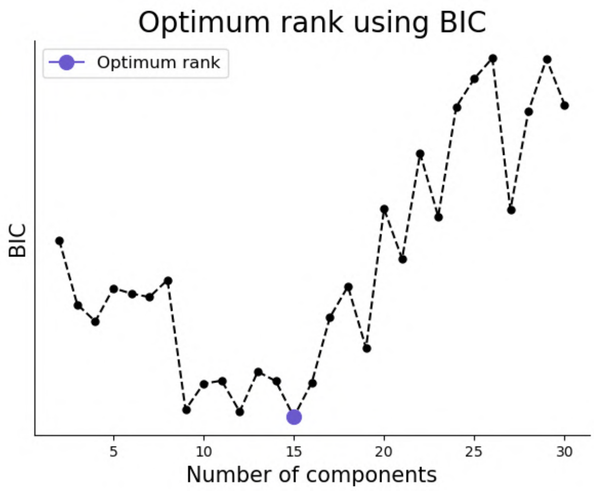


Figure 3.3: BIC criterion vs Number of Components to yield optimum value number of components

obtain that 15 is the optimum number of components to achieve the best results.

3.4 Examining the components

3.4.1 Goodness of a component

Using the Consensus approach on NMF, we obtain k clusters mapped to k components in the reconstructed W matrix. To identify how well a component is, we observe two quantities: Mean Pairwise Correlation and Size of the Cluster, i.e. number of pre-consensus components in a cluster.

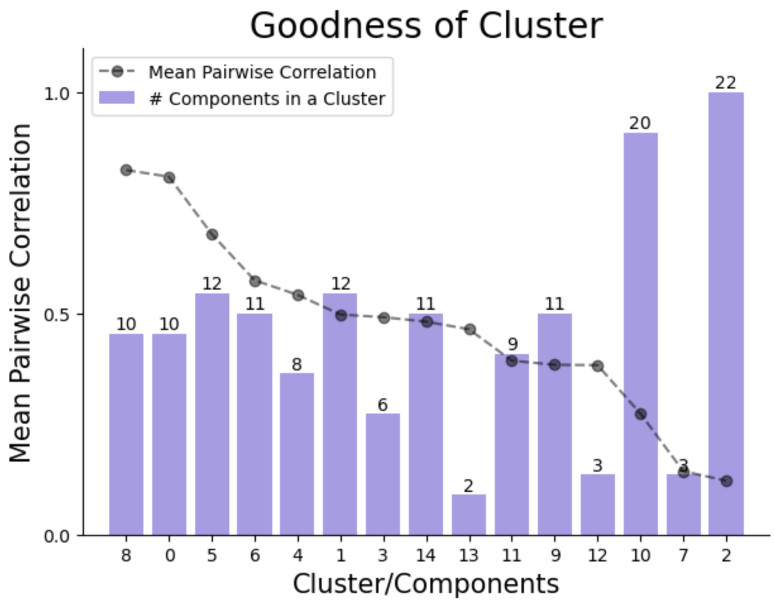


Figure 3.4: Quantification of the goodness of components for Participant 1

3.4.2 Representative Object Categories

The W matrix obtained through the Consensus approach reflects the strength with which each object category is associated with each component. To interpret how object categories are differentiated across components, we analyze the categories that exhibit the highest weights for a given component, i.e., the categories most strongly linked to the corresponding neural pattern. It is important to note that since MEG signals are averaged across all images within a category, the representative images shown illustrate the overall category as viewed by the participant rather than a specific image viewed by the participant.

To observe how components vary with MPC and cluster size, we first look at the components with the highest MPC and a decent cluster size.

Figure 3.5: Object categories loading best on component 8

Component 8

In this component, the highest-loading images predominantly feature straight lines and exhibit prominent edge-like patterns, suggesting that the component is sensitive to geometric structure and edge detection. These include parallel arrangements like multiple screwdrivers, vertical patterns like bamboo trees, grid-like layouts such as bathroom tiles and waffles, and distinct edge features in objects like dice and wrapped candies.

Component 0

Figure 3.6: Object categories loading best on component 0

In this component, the highest-loading images predominantly feature white objects, highlighting a contrast-driven representation. Additionally, the component appears sensitive to geometric features such as rectangles and corners. Representative images include white screens, notebook pages, tables, plates, and tiled surfaces, i.e. objects that emphasize both brightness and structural edges.

Now, let's look at the components that do not have the best MPC.

Component 13

This component has a mean pairwise correlation (MPC) of approximately 0.5 and a cluster size of 2, indicating that it captures two distinct groups of images with low inter-category correlation. One group primarily includes nature-related images such as trees, bushes, animals, and broccoli. The other group is characterized by strong checkered

Figure 3.7: Object categories loading best on component 13

or grid-like patterns featuring objects like chessboards, dice, nets, and blankets. This suggests that the component is sensitive to organic textures and structured geometric patterns.

Component 2

Figure 3.8: Object categories loading best on component 2

This component exhibits a very low mean pairwise correlation (MPC) of 0.2 and has the highest cluster size of 22. The high cluster size suggests that many individual components over all the runs converge onto this pattern. A significant proportion of the high-loading images depict animals, insects, and reptiles, while a subset also features objects with box-like or square structures. The presence of these two qualitatively different groups with limited shared features leads to the overall low MPC. Moreover, the large cluster size itself may naturally lead to lower average correlations due to the increased diversity across contributing categories.

Component 7

Lastly, we examine Component 7, which displays both a very low mean pairwise correlation (MPC) and a small cluster size. The component captures a highly diverse set of images, some featuring parallel lines, others rich in colour, a few with multiple small spherical objects, and even a few animal images. This diversity indicates that the component does not represent a coherent or dominant visual feature shared across categories. The small cluster size further suggests that this component is not a consistently emerging

Figure 3.9: Object categories loading best on component 7

pattern across different runs, reinforcing the idea that this compo reflects a rare or weakly defined representational structure.

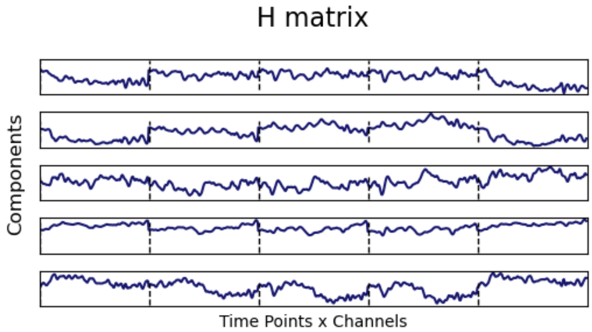
By examining cluster size, mean pairwise correlations (MPC), and the object categories that load onto each component, we can draw insights into the robustness and consistency of the components. Components with high MPC and moderate cluster sizes tend to be robust and well-defined, indicating that they consistently capture a specific pattern across trials. Components with moderate MPC but low cluster size are less frequently observed but show strong internal coherence, suggesting that when they do occur, they represent a distinct and reliable pattern. In contrast, components with high cluster size but low MPC appear frequently across runs, indicating stability, but their low correlation suggests they capture mixed or less pure representations, possibly combining multiple object categories. Finally, components with both low MPC and low cluster size are neither consistent nor commonly observed and likely represent noisy or weak patterns.

3.5 Time dynamics of these components

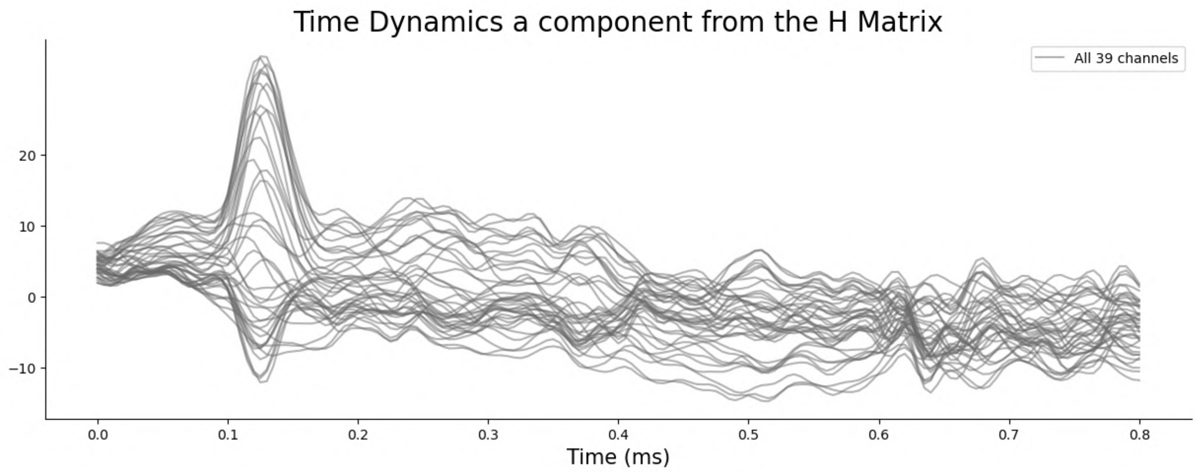
Now that we have looked at how different object categories load on components and what property each component represents, we now turn to the temporal aspect of neural processing. Specifically, we explore how the importance of each component evolves over time, providing insight into when certain object features are emphasized during visual perception in the brain.

3.5.1 The H Matrix

The H matrix represents the temporal dynamics of each component, essentially describing how strongly each component is activated over time. In the context of neural data, this reflects how the underlying neural representations captured by the W matrix evolve during visual processing. Each row in the H matrix corresponds to a component, and each column corresponds to a time point, allowing us to trace the contribution of specific features across the time course of the MEG recording.



To visualize the H matrix properly, we separate the concatenated time series of different channels and plot them to visualize the time dynamics of the components. ![](attachments/Pasted



This figure illustrates the temporal dynamics of the H matrix components. However, there are a few issues to note. Firstly, we observe negative values in the component activations, which is theoretically inconsistent since the H matrix, derived from a non-negative factorization, is expected to have only non-negative values. Secondly, the component magnitudes fluctuate in a relative manner rather than reflecting absolute importance, making the interpretation of ”negative importance” misleading. This suggests a need for normalization or reevaluation of the decomposition constraints to better align with the non-negative nature of the model.

3.5.2 Using NNLS to represent time

To estimate the time-varying contribution of each component, we perform Non-Negative Least Squares (NNLS) regression. This method computes the weights of each component across all time points and channels, ensuring that the resulting time dynamics are interpretable and non-negative (see Methods for details).

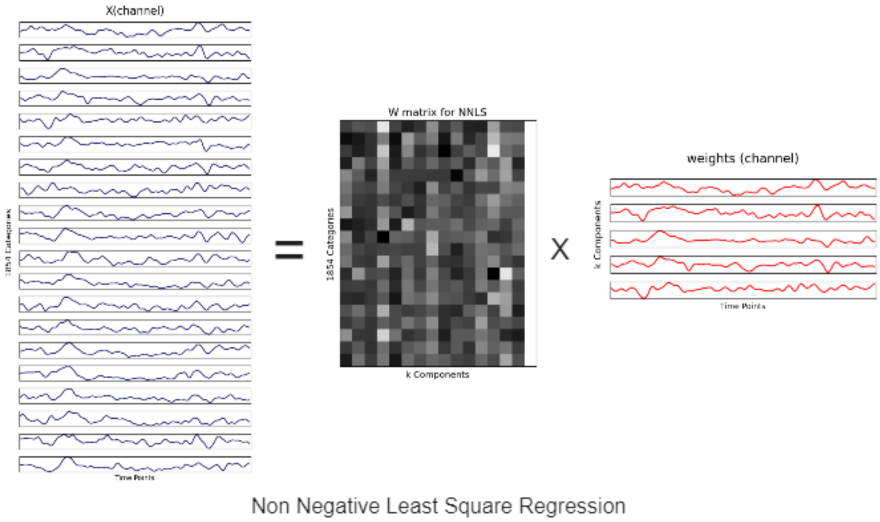


Figure 3.10: Enter Caption

To enforce positivity more robustly, we append a constant column vector to the component matrix, effectively acting as a bias term (β_0) in the regression. This ensures that the fitted time courses remain in the non-negative domain and allows the model to account for any baseline shifts across time. Here are the NNLS-derived weights for a single

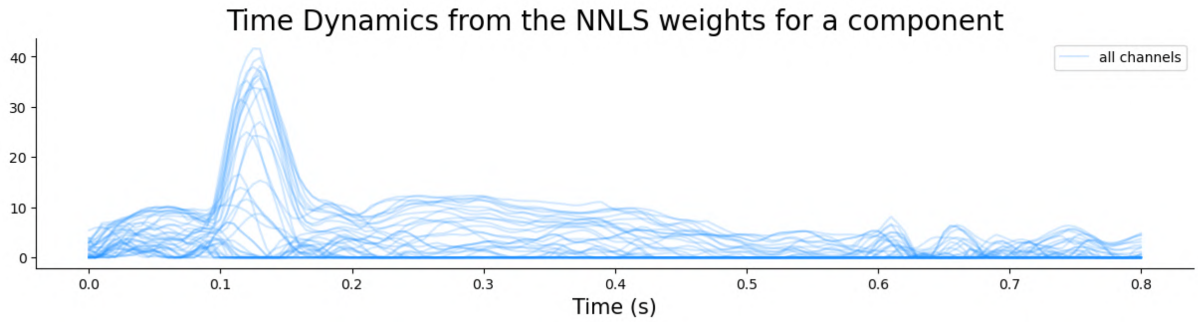


Figure 3.11: Enter Caption

component across all channels. As expected, the weights are strictly positive, reflecting the non-negative constraint of the regression. They indicate how strongly each channel contributes to the selected component. Now, we observe how the NNLS weights change vary across channels. We plot the correlation coefficient of NNLS weights across channels for all components. We observe the following: The correlation across channels for a single component varies significantly from values close to 1 down to around 0.3. As a result, simply averaging the signals across all channels would lead to a loss of important spatial information. To preserve this diversity, we again apply NMF on the NNLS-derived component activity across channels. We then examine whether a small number of latent

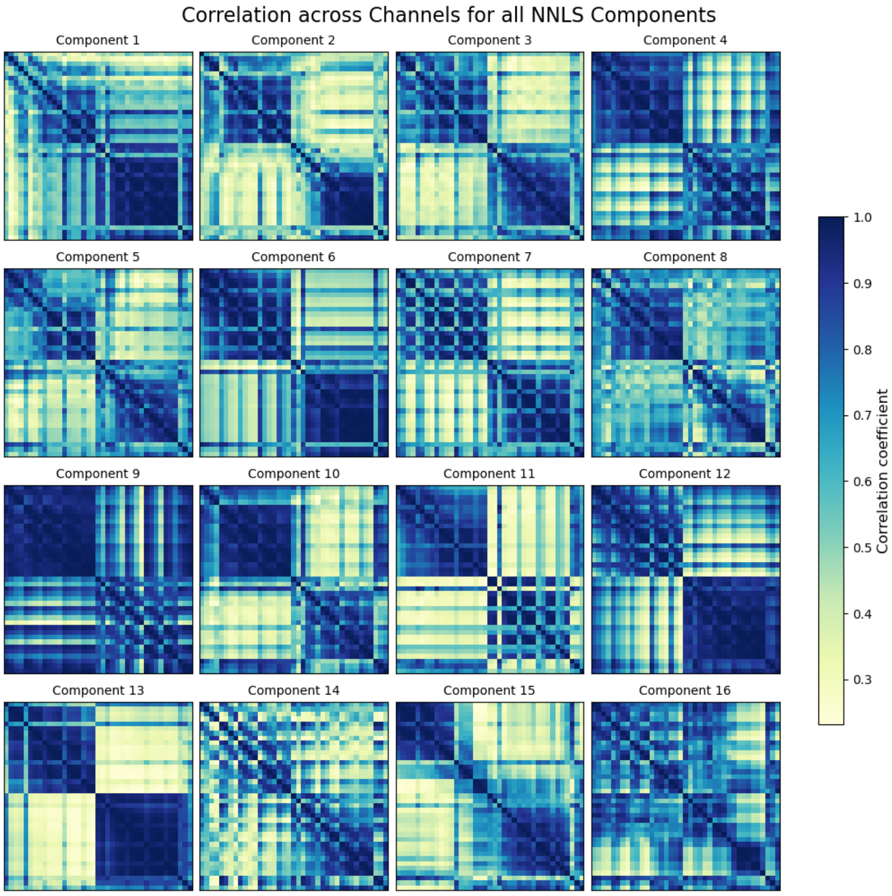


Figure 3.12: Enter Caption

patterns can explain most of the variance, ideally approaching an explained variance close to 1. We find that applying NMF to the NNLS-derived component activity across channels captures approximately 91% of the variance. This indicates that the summarized component retains most of the original spatial structure, validating the effectiveness of using NMF for dimensionality reduction in this context.

The function which plots NNLS summarise components and figures

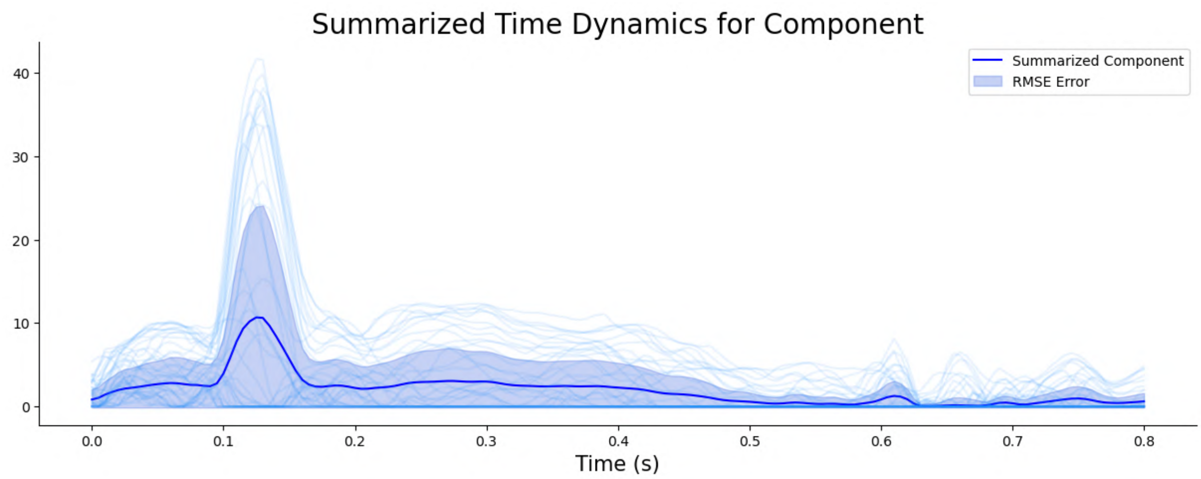


Figure 3.13: Summarizing time dynamics in a component over channels using NMF.

3.6 Components for other participants

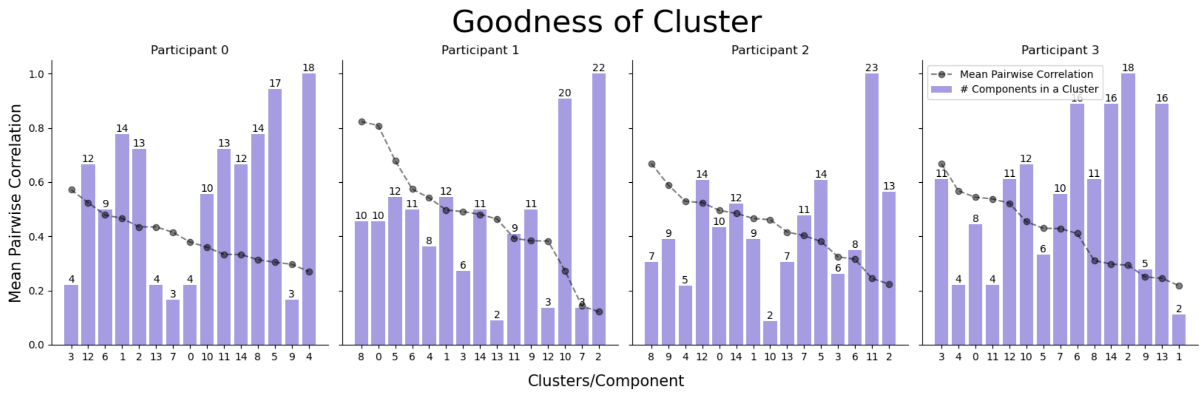


Figure 3.14: Goodness of Clusters over all 4 participants

To identify common patterns across participants, we compare their components derived from the W matrix. Specifically, we construct a correlation matrix between components across participants and select the component pairs with the highest correlation, without replacement. Figure 3.15 displays the correlation between components of participant 1 and participant 2. Figure 3.16 shows that component 0 of participant 1 is highly selective for white and rectangular objects, with mild selectivity for body parts. In contrast, component 0 of participant 2 exhibits strong selectivity for body parts. The temporal dynamics also differ: participant 1 shows an early peak around 0.12 s, whereas participant 2's peak develops more gradually over time. Figure 3.17 shows that both components in participant 1 show a very high selectivity for rectangles, corners, boxes. This is also categorized by an early peak in their time dynamics followed by a smaller, gradual peak later. Figure 3.18 shows that both components exhibit strong selectivity toward images containing multiple colors, often arranged in a checkered or mosaic-like pattern formed by fine-grained color textures.

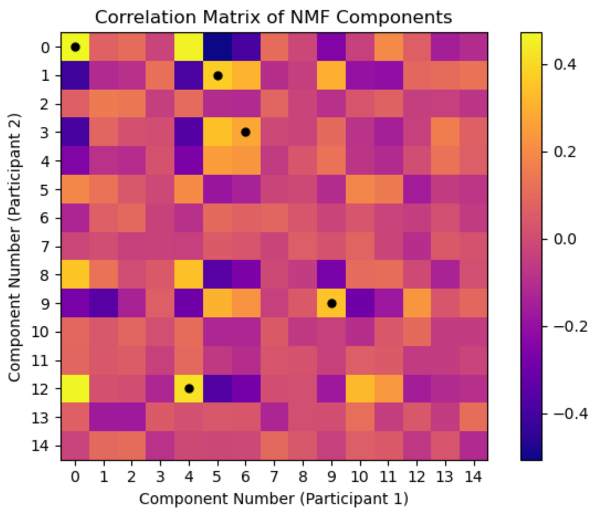


Figure 3.15: Comparison of components across different participants

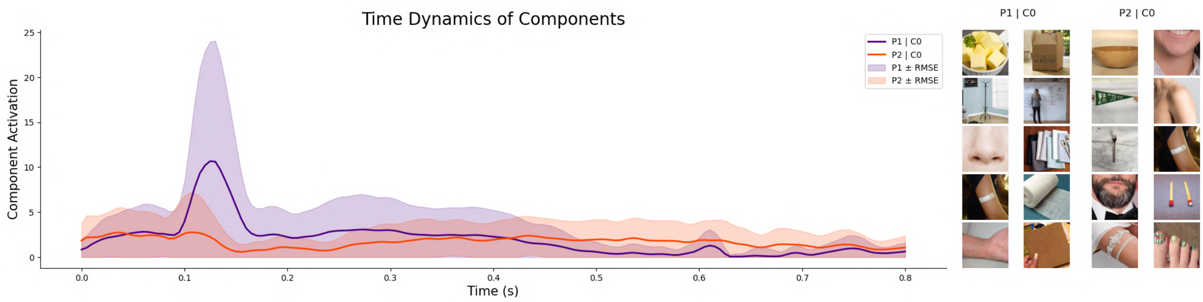


Figure 3.16: Participant 1 component 0 and participant 2 component 0

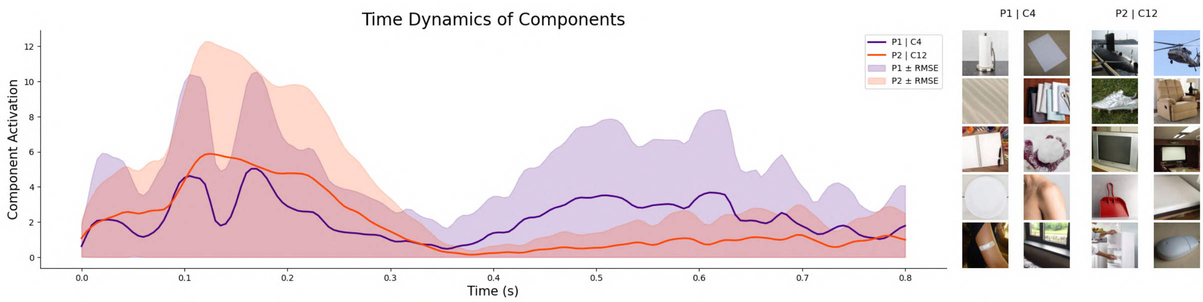


Figure 3.17: Participant 1 component 4 and participant 2 component 12

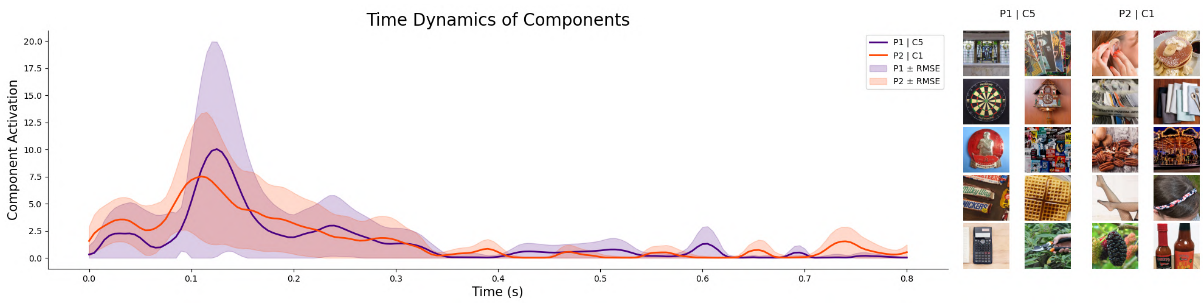


Figure 3.18: Participant 1 component 5 and participant 2 component 1

In Figure A.1, we observe a comparable pattern between participants 1 and 3. Both components exhibit the strongest selectivity for images dominated by white, rectangular shapes and sharp corners. This is followed by images featuring multicolored, mosaic-like patterns. Interestingly, the third most correlated component again shows selectivity for white, rectangular images, suggesting a recurring preference for this visual structure.

The temporal dynamics of the first two component pairs display clear early peaks. In contrast, the third pair presents a more complex temporal profile—an initial early peak followed by a gradual rise. Notably, participant 3's component in this pair exhibits high-frequency oscillations, which are unlikely to reflect genuine visual processing and may instead indicate a recording artifact.

Similarly, Figure A.2 reveals patterns consistent with previous comparisons. The first component pair is strongly selective for body parts in both participants. The second pair shows a preference for geometric shapes, such as rectangles and corners, while the third highlights mosaic-like, multicolored textures. All three component pairs exhibit early peaks in their temporal dynamics, suggesting rapid visual processing. Additionally, the second pair shows a secondary, delayed peak of smaller amplitude.

3.6.1 Components with Predominantly Late Temporal Peaks

In contrast to the early-peaking components discussed in the previous section, some participants exhibited components with delayed or sustained temporal responses, peaking after 200 ms. These components were often selective for categories that engage higher-level or conceptual visual processing, such as "body-/people-related" or "food-related" images.

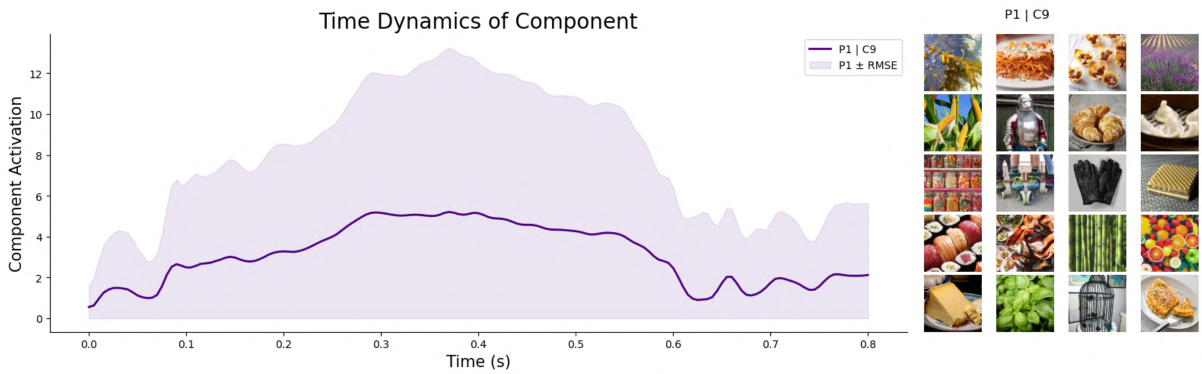


Figure 3.19: Temporal dynamics of a component with delayed peak, showing selectivity for repetitive patterns and food-related images.

In Figure 3.19, we observe a component that loads selectively on both repetitive patterns and food-related images. While the former may reflect low-level visual properties, the latter involves higher-level conceptual associations. Accordingly, the temporal profile

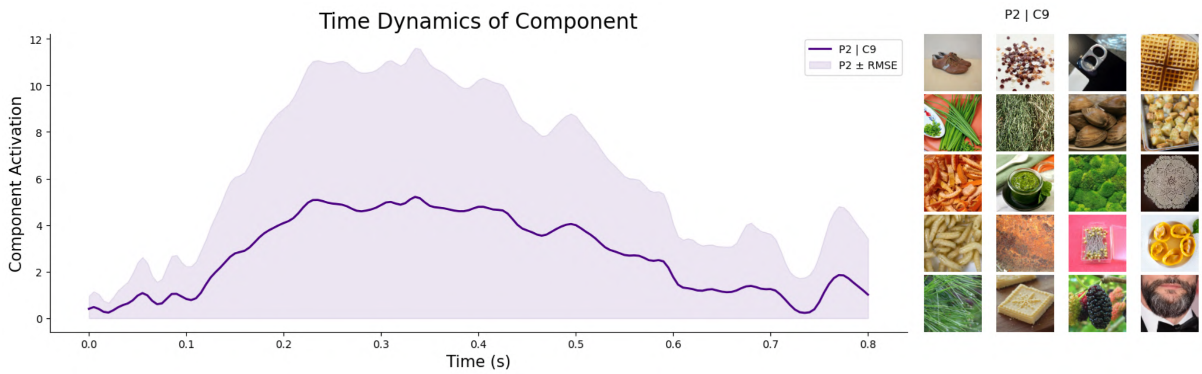


Figure 3.20: Component exhibiting strong selectivity for food-related images with a clear delayed peak at approximately 350 ms.

shows a small early peak at around 100 ms, followed by a gradual and more pronounced peak near 400 ms.

A clearer example is shown in Figure 3.20, where the component is highly selective for food items. The temporal dynamics are marked by a smooth, delayed peak around 350 ms, with minimal early response. This profile is consistent with slower, possibly more integrative processing associated with semantically rich or behaviorally relevant content.

3.6.2 Components with Both Early and Late Peaks

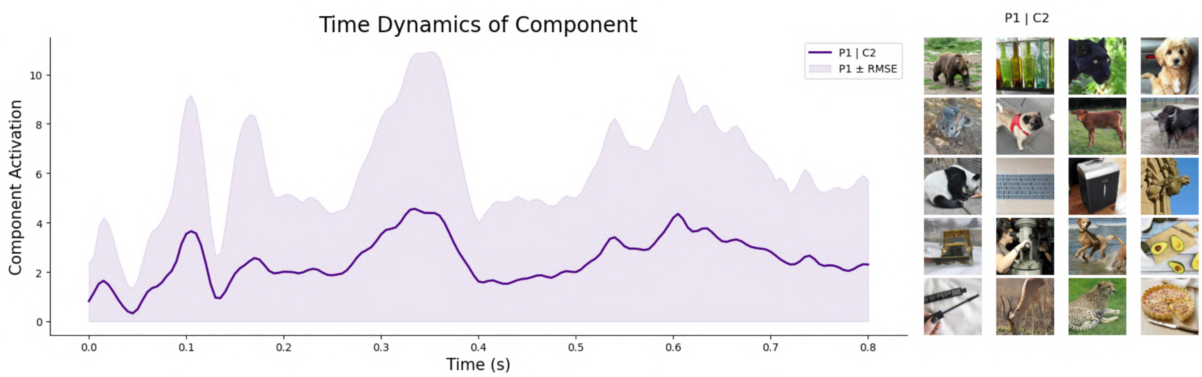


Figure 3.21: Component with selectivity for both low-level visual features (straight lines) and higher-level categories (animals), exhibiting multiple temporal peaks.

In Figure 3.21, we observe a component that loads highly on both straight lines and animal-related images. This dual selectivity manifests as two distinct peaks in the temporal profile: an early peak around 100 ms, likely reflecting sensitivity to low-level features, and a later peak near 350 ms, potentially corresponding to semantic or categorical processing related to animals. Interestingly, there is an additional peak around 600 ms, which may either reflect a recording artifact or point to further processing of yet another conceptual feature.

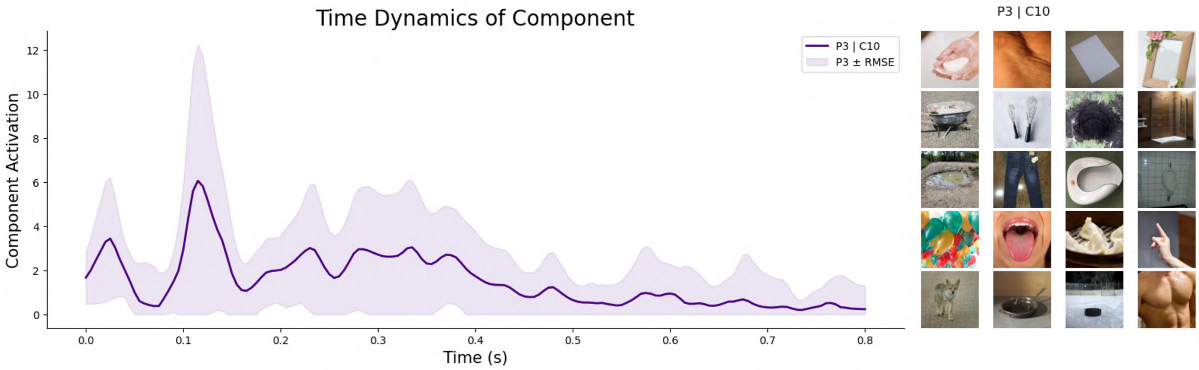


Figure 3.22: Component showing early and late peaks, with selectivity for both geometric shapes and body parts.

A similar temporal pattern is observed in Figure 3.22, where the component is selectively tuned to white rectangles and various body parts. The temporal dynamics reveal a sharp early peak around 120 ms, followed by a smaller, gradual rise at approximately 350 ms. These examples suggest that components with mixed selectivity often display composite temporal profiles—early peaks for consistent visual features, and later peaks for higher-level, potentially participant-specific semantic categories.

Chapter 4

Discussion

4.1 Summary of Results

Our analysis demonstrated that selecting approximately 15 components for Non-negative Matrix Factorization (NMF) provided an optimal balance between level of detail and interpretability. This number effectively minimized overlap between components, ensuring that each component captured distinct representational features. However, some very low-level features, such as rectangles, lines, or white objects against dark backgrounds, were occasionally split across multiple components, likely due to their ubiquitous presence in the MEG dataset.

The consensus approach proved to be highly effective in ensuring the stability and interpretability of the components. Clustering components across multiple NMF runs allowed for grouping similar components, and the median of these clusters provided a robust representation of each component. Additionally, we used two metrics to quantify the reliability and prominence of each component:

- **Mean Pairwise Correlation (MPC)** across components within a cluster, which served as an indicator of internal consistency,
- **Cluster Size** reflected the frequency of the component's occurrence across different runs.

These metrics allowed us to draw several conclusions about the quality of the components:

By analyzing the components using the Mean Pairwise Correlation (MPC) and cluster size, we observed several distinct patterns that provided insights into the stability and structure of the components. Components with high MPC and moderate to large cluster sizes were found to be highly consistent and selective, typically representing a single object category. These components exhibited strong internal consistency, meaning

that the same object category was consistently represented across different runs, making them robust and interpretable.
In contrast, components with moderate MPC and cluster sizes often captured two related object categories. Although these components were still relatively consistent, they tended to represent categories that were somewhat related to each other, which explained the moderate correlation between them. For example, a component might represent both geometric shapes and simple textures. These components were stable but more diverse in their representation.
Components with low MPC but large cluster sizes indicated the co-occurrence of two highly stable but distinct object categories, such as animals and geometric patterns (e.g., straight lines). While these components had a large cluster size due to the frequent appearance of these categories together, the low MPC reflected the lack of correlation between the two categories, which were dissimilar in nature. As a result, the cluster size was large, but the internal consistency (MPC) was low.
Finally, components with both low MPC and small cluster sizes were characterized by the presence of multiple, unrelated object categories. These components were less consistent, with their representations varying across runs, and the small cluster size suggested that they did not consistently emerge as a dominant pattern. This indicated that these components might represent weakly defined or poorly structured object categories, which were less stable and less likely to provide reliable information about the neural representations across different participants or runs.

In terms of temporal dynamics, the H matrix representing component activations over time did not reflect the expected absolute importance of each component. Instead, the activations fluctuated relatively, and the presence of negative values from the baseline was inconsistent with the non-negative constraints of the model. These fluctuations suggested that the components did not have a stable baseline and that one component's activation often inversely corresponded to another's. This necessitated a re-evaluation of the decomposition and normalization process.

To address this issue, we used Non-Negative Least Squares (NNLS) regression to obtain time courses for each component. We incorporated a bias term (β_0) to remove relative effects and establish a consistent baseline for all components. The time dynamics derived from this approach were more stable and interpretable. Although some components exhibited moderate reconstruction errors (RMSE), the overall temporal profiles were consistent across participants and runs.

Additionally, we analyzed components across participants by correlating the consensus W matrix for each participant. This allowed us to identify recurring patterns in the neural representations, such as the selective representation of rectangles, white corners, body parts, and mosaic textures. Notably, body parts, despite being a higher-level object category, consistently emerged across participants, suggesting a robust and early neural representation of this category.

The temporal dynamics of the components revealed distinct processing timelines. Early peaks around 100–120 ms were commonly observed for low-level features, including rectangles, straight lines, and mosaic patterns. In contrast, high-level categories, such as food, showed later peaks in the 300–400 ms range, indicating a delayed processing phase for these more complex object categories.

Finally, components with mixed selectivity, i.e., those representing low and high-level features, often exhibited bimodal time dynamics. Specifically, such components displayed one early peak (120 ms) followed by another at 350 ms, indicative of a two-stage visual and conceptual processing process. This pattern was evident in components that were selective for animals and straight lines or body parts and geometric shapes, supporting the notion of a temporally structured progression from perceptual to conceptual processing.

4.2 Nature of Visual Representations Over Time

Our results provide strong evidence that object representations in the brain evolve over time, with early peaks corresponding to low-level visual features and later peaks corresponding to higher-level, abstract object categories. The early peaks, typically around 100–120 ms, correspond to the processing of low-level visual features such as shapes, lines, and basic textures. These features are essential for object recognition at the perceptual stage, where the brain begins to break down the raw sensory input into meaningful components.

In contrast, the later peaks, observed around 350–400 ms, seem to correspond to higher-level categories, such as food, animals, or people. This temporal shift suggests a hierarchical processing structure in the brain, where initial sensory-driven processing of basic features is followed by more abstract, conceptual processing of objects. This aligns with the broader theory of visual processing in the brain, where perceptual features form the building blocks for higher-level cognitive representations. Our unsupervised approach, using Non-negative Matrix Factorization (NMF) and time-resolved analysis, reproduces the findings of Teichmann et al. (2024) [25] by revealing distinct temporal patterns in object representation. Teichmann et al. used a trained behavioural decoding model to find similar temporal dynamics in object recognition. However, our approach is unique because it does not rely on behavioural input or labels. Instead, it reveals the underlying structure purely from the neural data, supporting the idea that object representations evolve over time, from consistent perceptual processing in early stages to more abstract, conceptually-driven processing in later stages.

This evolution from low-level to high-level representation reinforces the notion that object representations are dynamic and change over time, with an initial sensory-driven stage that gradually transforms into more abstract and generalized cognitive representations. The hierarchy or cascade of these representations suggests a structured process where different brain regions or networks are activated at different time points, depending on the complexity of the object being processed.

4.3 Stability of components across participants

The presence of shared components across subjects, such as the consistent activation patterns associated with white rectangles or mosaic-like textures, suggests that the brain may rely on canonical mechanisms for organizing and compressing visual information. These stable components likely reflect fundamental neural sources or functional modules engaged during the early stages of visual processing. Specifically, areas such as V1 and V2 are well-established in the literature as critical for early, low-level visual feature processing, such as orientation, contrast, and texture [2], [3]. These areas are thought to operate in a feedforward manner, extracting simple features that form the basis of more complex visual perception.

Our findings align with this theory, suggesting that low-level visual features such as shapes, lines, or textures are processed in a relatively uniform manner across participants. This consistency supports the notion that neural representations of these basic features are robust and may be governed by standard neural mechanisms that are primarily invariant across individuals. The stability of these low-level components across subjects indicates that the brain has a reliable and organized architecture for processing fundamental visual information, consistent with findings from early visual areas like V1 and V2. On the other hand, higher-level visual representations, such as object categories and complex textures, appear to vary more across individuals. This variability could reflect individual differences in experience, attention, or the organization of higher-level processing areas (e.g., lateral occipital complex or fusiform gyrus) that process more abstract visual information. This divergence between low-level and high-level components suggests that while the brain has stable, canonical processes for early visual processing, higher-order representations are more flexible and may depend on top-down influences or personal experiences, as seen in higher-order visual areas [27].

Thus, the consistency of low-level components across participants, paired with the variability in higher-level object representations, highlights the hierarchical and modular nature of visual processing. Our use of unsupervised learning methods, like Non-negative Matrix Factorization (NMF), extracts these stable and flexible features, supporting the idea that MEG captures these intrinsic patterns of brain activity, which are critical for visual object recognition.

4.4 Why Representations differ across subjects

Variability in subject representations can be attributed to several factors, including both fundamental differences in neural processing and methodological limitations.

One possibility is that individuals process the same visual input differently based on their experiences, attention, or cognitive strategies. Variations in attentional focus, prior knowledge, or cultural background can lead to divergent interpretations of complex visual stimuli, particularly in higher-level object categories, where representations are more abstract and influenced by personal context. For example, categories such as food, animals, or people may be processed differently across subjects, with distinct patterns of neural activation shaped by factors like familiarity, expertise, or emotional associations. As a result, representations of these higher-level categories are more likely to vary significantly between individuals.

Another source of variability arises from methodological issues inherent in the use of MEG and unsupervised learning techniques like Non-negative Matrix Factorization (NMF). For instance, the noise present in MEG data, which can vary across participants and conditions, may introduce discrepancies in the extraction of components, resulting in slight differences in the components identified across subjects. Additionally, NMF operates under the assumption of a parts-based representation, which may lead to misalignment in the ordering of components or difficulties in interpreting mixed components. Unknown rotation and scaling of components in the NMF space can also contribute to apparent variability, causing components to appear different across subjects even when they reflect similar underlying representations. These technical factors create the illusion of variability in the data, even when shared neural processing is at play.

4.5 Limitations

While our analysis provides valuable insights into the representation of visual objects in the brain, it is important to acknowledge the limitations of our approach.

First, we did not employ decoding or supervised models, which means we cannot formally assess the generalizability or predictive power of our findings, limiting our ability to test how well these components might generalize to new, unseen stimuli or behavioural tasks.

Moreover, the use of Non-negative Matrix Factorization (NMF) introduces certain limitations, particularly in that it assumes a parts-based linear combination of components,

which may not fully capture the complexity and non-linearity of brain dynamics.
Additionally, our analysis did not consider how neural representations might vary spatially, particularly concerning different brain regions or channels. We only included data from 39 occipital channels, though a total of 271 channels were available. This narrow spatial focus could limit our understanding of how object representations are distributed across the entire brain.
Furthermore, our use of Non-negative Least Squares (NNLS) assumes linearity in the temporal domain, neglecting the possibility of feedback loops or recurrence processes that could be integral to neural processing. Finally, some of the components we identified were challenging to label or exhibited ambiguous selectivity. This suggests that while the components derived from NMF are interpretable, they may not fully reflect the richness of underlying neural activity, and further refinement may be necessary to improve their clarity and interpretability.

4.6 Future Directions

Looking forward, there are several promising avenues for extending this work. One significant step would be to test the behavioural or semantic relevance of the components we identified and linking them more directly to human perception or cognitive judgments. This could be done by pairing our unsupervised approach with decoding techniques, allowing us to assess the predictive power and generalizability of individual components.

Another direction would be to develop a quantitative metric or algorithm that evaluates the "quality" of a component based on which object categories load onto it. Such a tool could assign a consistency or selectivity score, enabling systematic comparison across components, subjects, or methods. Similarly, employing Bayesian NMF combined with a consensus approach could yield more stable and interpretable components, as Bayesian formulations offer probabilistic guarantees and tend to be more robust to noise.

Comparing NMF with alternative matrix decomposition or dimensionality reduction techniques such as ICA, PCA, or autoencoders could help determine the best method for extracting interpretable representations from MEG data. Another key extension would be to explore the spatial dynamics of neural representations: how components manifest across different sensor locations and how early and late representations map onto specific brain regions. This could be facilitated by incorporating source localization techniques.

Lastly, a long-term goal would be to build a unified model of time-evolving representations that accounts for shared patterns across individuals and meaningful individual differences. Such a framework could provide broader insights into the neural basis of object recognition while also acknowledging the diversity of human cognition.

Appendix A

A.1 Algorithms and Formulae

A.1.1 Multiplicative Update Rules for NMF

The goal of Non-negative Matrix Factorization (NMF) is to approximate a non-negative data matrix $V \in \mathbb{R}^{m \times n}$ as the product of two lower-rank non-negative matrices $W \in \mathbb{R}^{m \times k}$ and $H \in \mathbb{R}^{k \times n}$, such that:

$$V \approx WH, \quad \text{with } W \geq 0,\ H \geq 0.$$

This factorization is achieved by minimizing the Frobenius norm of the reconstruction error:

$$\mathcal{L}(W, H) = \frac{1}{2}\|V - WH\|_F^2.$$

Lee and Seung (2001) [28] introduced multiplicative update rules that ensure both non-negativity and convergence. These updates take the form:

$$H \leftarrow H \cdot \frac{W^T V}{W^T W H}, \quad W \leftarrow W \cdot \frac{V H^T}{W H H^T},$$

where all operations are element-wise. These update rules are derived from a gradient descent framework, but instead of subtracting the gradient, each parameter is updated by a positive scaling factor. This guarantees that all entries in W and H remain non-negative throughout training.

The intuition behind the multiplicative form is that it improves components contributing positively to the reconstruction while suppressing less useful elements. Because NMF only allows additive combinations (due to non-negativity), it encourages a parts-based representation—capturing local, interpretable features rather than holistic ones.

A.1.2 K-means Clustering Algorithm

K-means is an unsupervised clustering algorithm that partitions n data points into K clusters, such that each point belongs to the cluster with the nearest mean (centroid). The algorithm seeks to minimize the total intra-cluster variance.

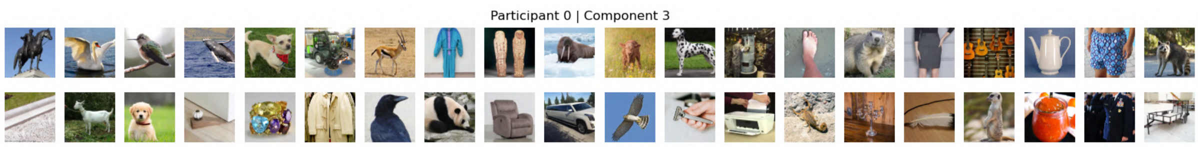
Algorithm Steps:

1. **Initialization:** Randomly select K data points as initial cluster centroids $\{\mu_1, \ldots, \mu_K\}$.
2. **Assignment Step:** Assign each data point $x_i \in \mathbb{R}^d$ to the cluster whose centroid is nearest:
$$C_k = \left\{x_i : \|x_i - \mu_k\|^2 \leq \|x_i - \mu_j\|^2, \forall j = 1, \ldots, K\right\}$$
3. **Update Step:** Recompute the centroid of each cluster as the mean of the data points assigned to it:
$$\mu_k = \frac{1}{|C_k|} \sum_{x_i \in C_k} x_i$$
4. **Repeat Steps 2–3** until the centroids converge (i.e., changes are below a threshold or zero).

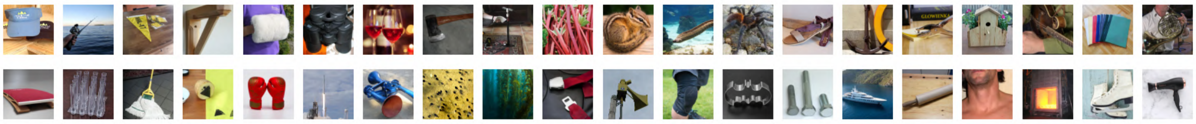
K-means is efficient and straightforward, but its results can be sensitive to the initial choice of centroids. Techniques such as *k-means++* provide improved initialization strategies to enhance clustering stability and performance.

A.2 Additional Results

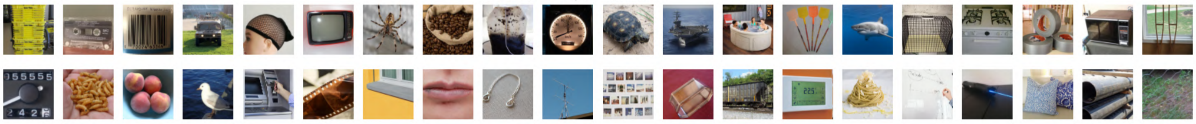
A.2.1 Top image categories on each components



Participant 0 | Component 6

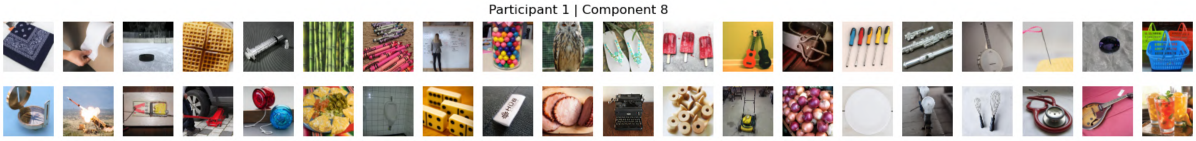


Participant 0 | Component 1



Participant 0 | Component 2

Participant 1 | Component 8

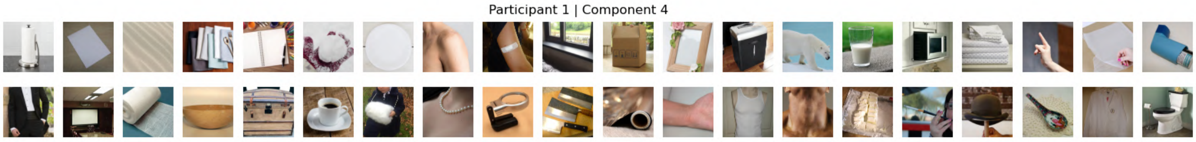


Participant 1 | Component 0

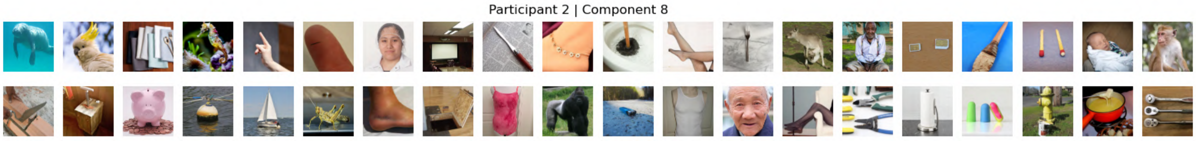
Participant 1 | Component 5

Participant 1 | Component 6

Participant 1 | Component 4

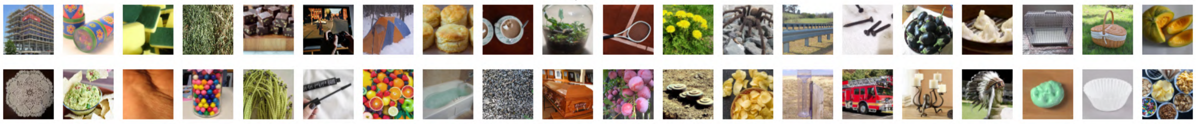


Participant 2 | Component 8



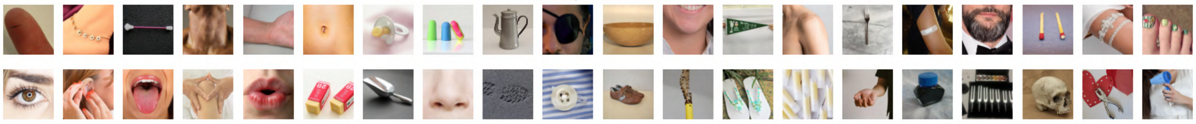
Participant 2 | Component 9

Participant 2 | Component 4



Participant 2 | Component 12

Participant 2 | Component 0

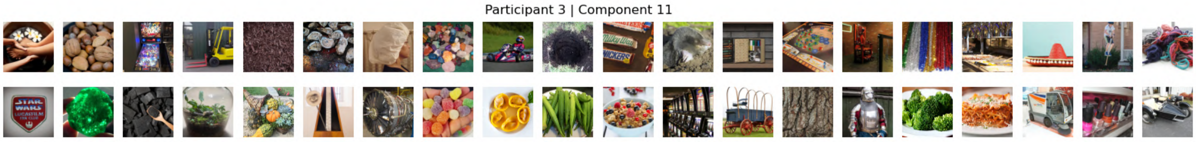


Participant 3 | Component 3

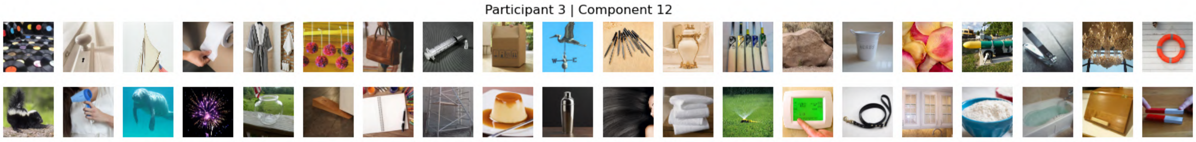
Participant 3 | Component 4

Participant 3 | Component 0

Participant 3 | Component 11



Participant 3 | Component 12



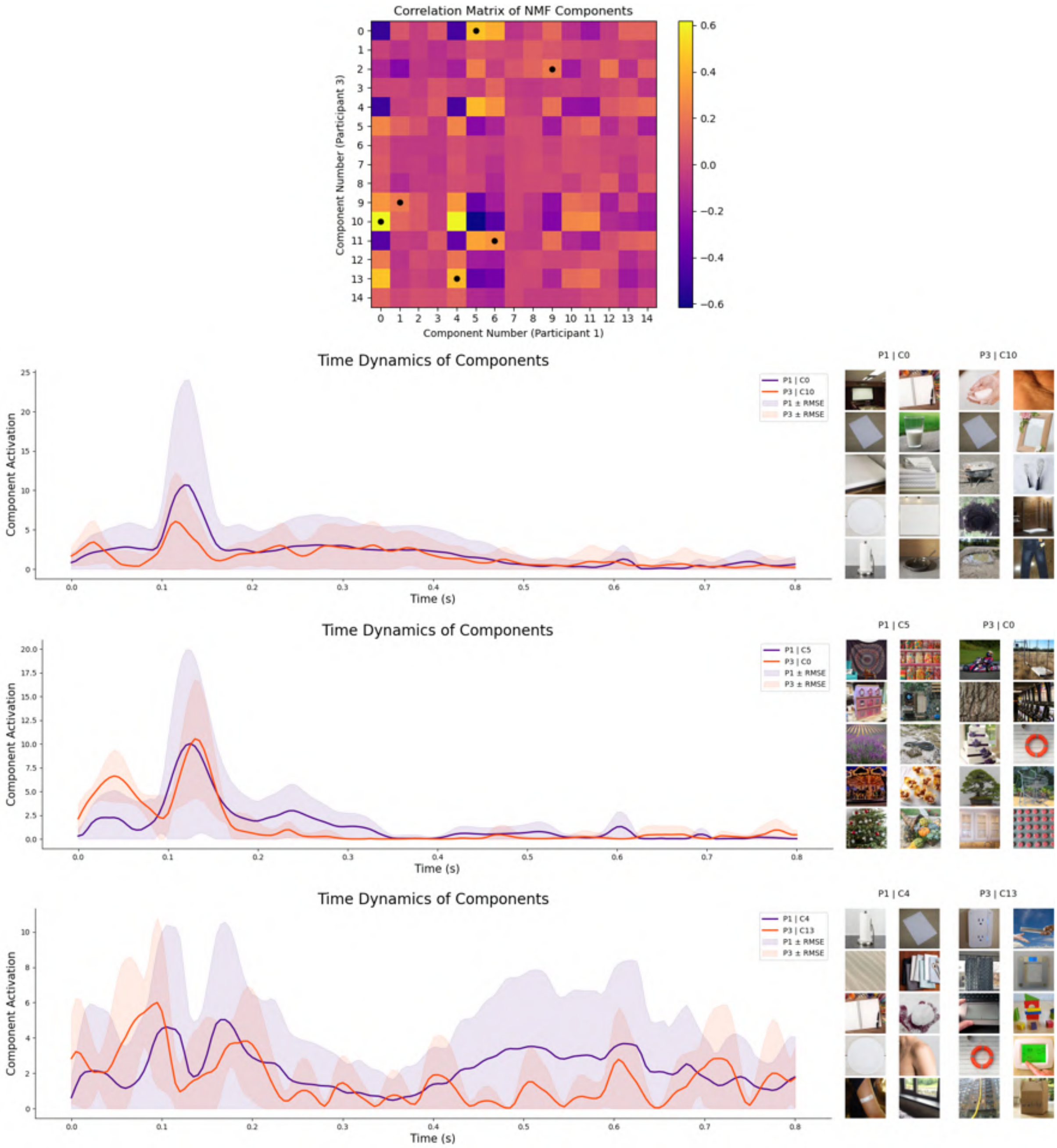


Figure A.1: Time dynamics of highest correlated components in Participant 1 and 3

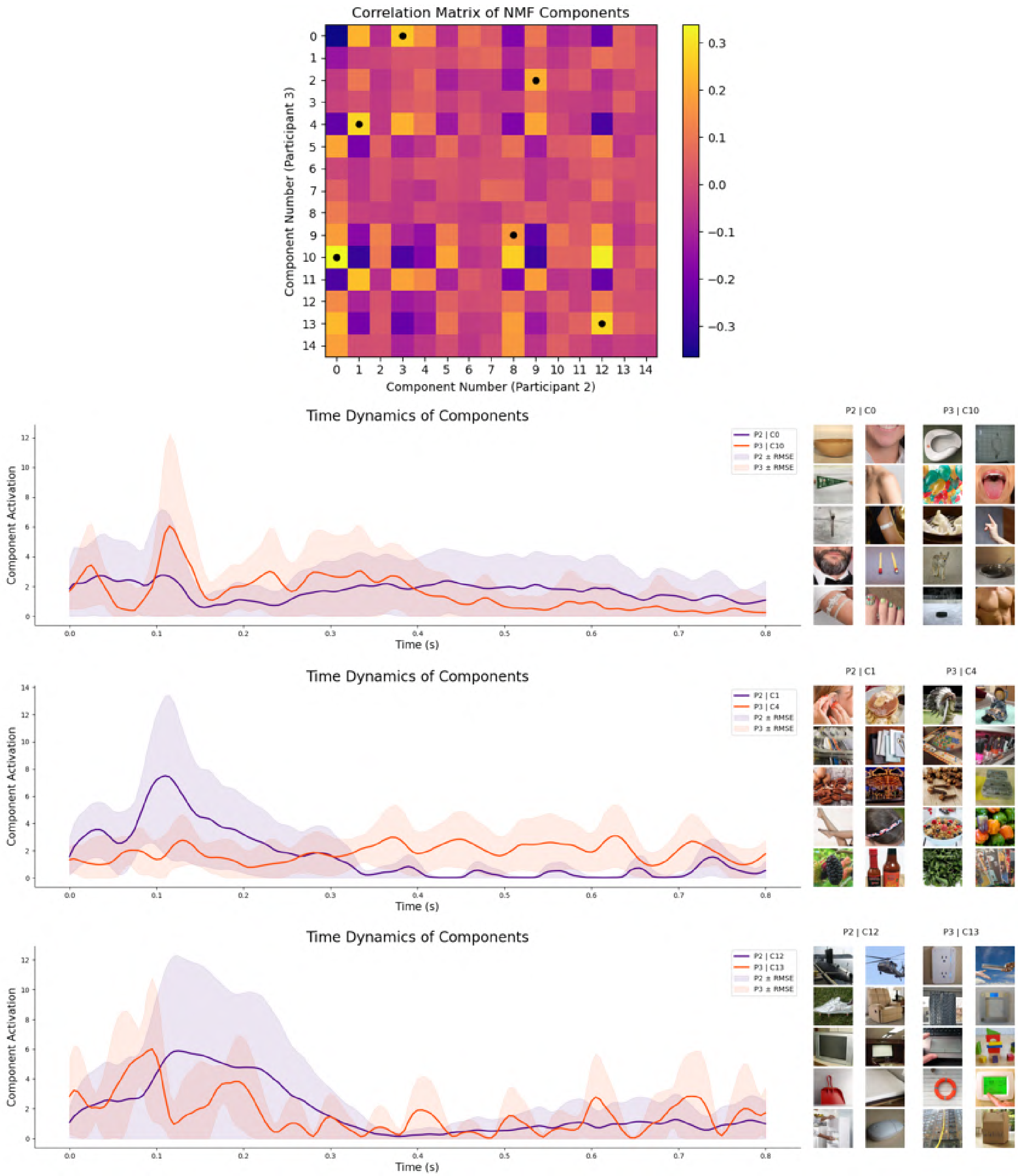


Figure A.2: Time dynamics of highest correlated components in Participant 2 and 3

Bibliography

[1] Melvyn A. Goodale and A.David Milner. "Separate visual pathways for perception and action". In: *Trends in Neurosciences* 15.1 (Jan. 1992), pp. 20–25. ISSN: 0166-2236. DOI: 10.1016/0166-2236(92)90344-8. URL: http://dx.doi.org/10.1016/0166-2236(92)90344-8.

[2] D. H. Hubel and T. N. Wiesel. "Receptive fields of single neurones in the cat's striate cortex". In: *The Journal of Physiology* 148.3 (Oct. 1959), pp. 574–591. ISSN: 1469-7793. DOI: 10.1113/jphysiol.1959.sp006308. URL: http://dx.doi.org/10.1113/jphysiol.1959.sp006308.

[3] D. H. Hubel and T. N. Wiesel. "Receptive fields and functional architecture of monkey striate cortex". In: *The Journal of Physiology* 195.1 (Mar. 1968), pp. 215–243. ISSN: 1469-7793. DOI: 10.1113/jphysiol.1968.sp008455. URL: http://dx.doi.org/10.1113/jphysiol.1968.sp008455.

[4] Russell L. De Valois, Duane G. Albrecht, and Lisa G. Thorell. "Spatial frequency selectivity of cells in macaque visual cortex". In: *Vision Research* 22.5 (Jan. 1982), pp. 545–559. ISSN: 0042-6989. DOI: 10.1016/0042-6989(82)90113-4. URL: http://dx.doi.org/10.1016/0042-6989(82)90113-4.

[5] J Movshon et al. "The analysis of moving visual patterns". English (US). In: *Pattern recognition mechanisms*. Ed. by C. Chagas, R. Gattass, and C. Gross. Pontificiae Academiae Scientiarum Scripta Varia. Vatican Press, 1985, pp. 117–151.

[6] R Desimone et al. "Stimulus-selective properties of inferior temporal neurons in the macaque". In: *The Journal of Neuroscience* 4.8 (Aug. 1984), pp. 2051–2062. ISSN: 1529-2401. DOI: 10.1523/jneurosci.04-08-02051.1984. URL: http://dx.doi.org/10.1523/JNEUROSCI.04-08-02051.1984.

[7] Jay Hegdé and David C. Van Essen. "Selectivity for Complex Shapes in Primate Visual Area V2". In: *The Journal of Neuroscience* 20.5 (Mar. 2000), RC61–RC61. ISSN: 1529-2401. DOI: 10.1523/jneurosci.20-05-j0001.2000. URL: http://dx.doi.org/10.1523/JNEUROSCI.20-05-j0001.2000.

[8] Anitha Pasupathy and Charles E. Connor. "Population coding of shape in area V4". In: *Nature Neuroscience* 5.12 (Nov. 2002), pp. 1332–1338. ISSN: 1546-1726. DOI: 10.1038/972. URL: http://dx.doi.org/10.1038/972.

[9] C G Gross, C E Rocha-Miranda, and D B Bender. "Visual properties of neurons in inferotemporal cortex of the Macaque." In: *Journal of Neurophysiology* 35.1 (Jan. 1972), pp. 96–111. ISSN: 1522-1598. DOI: 10.1152/jn.1972.35.1.96. URL: http://dx.doi.org/10.1152/jn.1972.35.1.96.

[10] Gang Wang, Keiji Tanaka, and Manabu Tanifuji. "Optical Imaging of Functional Organization in the Monkey Inferotemporal Cortex". In: *Science* 272.5268 (June 1996), pp. 1665–1668. ISSN: 1095-9203. DOI: 10.1126/science.272.5268.1665. URL: http://dx.doi.org/10.1126/science.272.5268.1665.

[11] Chou P. Hung et al. "Fast Readout of Object Identity from Macaque Inferior Temporal Cortex". In: *Science* 310.5749 (Nov. 2005), pp. 863–866. ISSN: 1095-9203. DOI: 10.1126/science.1117593. URL: http://dx.doi.org/10.1126/science.1117593.

[12] Edmund T. Rolls, Gordon C. Baylis, and Michael E. Hasselmo. "The responses of neurons in the cortex in the superior temporal sulcus of the monkey to band-pass spatial frequency filtered faces". In: *Vision Research* 27.3 (Jan. 1987), pp. 311–326. ISSN: 0042-6989. DOI: 10.1016/0042-6989(87)90081-2. URL: http://dx.doi.org/10.1016/0042-6989(87)90081-2.

[13] Nancy Kanwisher, Josh McDermott, and Marvin M. Chun. "The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception". In: *The Journal of Neuroscience* 17.11 (June 1997), pp. 4302–4311. ISSN: 1529-2401. DOI: 10.1523/jneurosci.17-11-04302.1997. URL: http://dx.doi.org/10.1523/JNEUROSCI.17-11-04302.1997.

[14] Doris Y. Tsao et al. "A Cortical Region Consisting Entirely of Face-Selective Cells". In: *Science* 311.5761 (Feb. 2006), pp. 670–674. ISSN: 1095-9203. DOI: 10.1126/science.1119983. URL: http://dx.doi.org/10.1126/science.1119983.

[15] Nikolaus Kriegeskorte. "Representational similarity analysis – connecting the branches of systems neuroscience". In: *Frontiers in Systems Neuroscience* (2008). ISSN: 1662-5137. DOI: 10.3389/neuro.06.004.2008. URL: http://dx.doi.org/10.3389/neuro.06.004.2008.

[16] H B Barlow. "Single Units and Sensation: A Neuron Doctrine for Perceptual Psychology?" In: *Perception* 1.4 (Dec. 1972), pp. 371–394. ISSN: 1468-4233. DOI: 10.1068/p010371. URL: http://dx.doi.org/10.1068/p010371.

[17] R. Quian Quiroga et al. “Invariant visual representation by single neurons in the human brain”. In: *Nature* 435.7045 (June 2005), pp. 1102–1107. ISSN: 1476-4687. DOI: 10.1038/nature03687. URL: http://dx.doi.org/10.1038/nature03687.

[18] James V. Haxby et al. “Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex”. In: *Science* 293.5539 (Sept. 2001), pp. 2425–2430. ISSN: 1095-9203. DOI: 10.1126/science.1063736. URL: http://dx.doi.org/10.1126/science.1063736.

[19] B OLSHAUSEN and D FIELD. “Sparse coding of sensory inputs”. In: *Current Opinion in Neurobiology* 14.4 (Aug. 2004), pp. 481–487. ISSN: 0959-4388. DOI: 10.1016/j.conb.2004.07.007. URL: http://dx.doi.org/10.1016/j.conb.2004.07.007.

[20] Sidney R. Lehky et al. “Statistics of visual responses in primate inferotemporal cortex to object stimuli”. In: *Journal of Neurophysiology* 106.3 (Sept. 2011), pp. 1097–1117. ISSN: 1522-1598. DOI: 10.1152/jn.00990.2010. URL: http://dx.doi.org/10.1152/jn.00990.2010.

[21] Yu-Xiong Wang and Yu-Jin Zhang. “Nonnegative Matrix Factorization: A Comprehensive Review”. In: *IEEE Transactions on Knowledge and Data Engineering* 25.6 (2013), pp. 1336–1353. DOI: 10.1109/TKDE.2012.51.

[22] Lihong Zhao, Guibin Zhuang, and Xinhe Xu. “Facial expression recognition based on PCA and NMF”. In: *2008 7th World Congress on Intelligent Control and Automation* (2008), pp. 6826–6829. URL: https://api.semanticscholar.org/CorpusID:6881693.

[23] Martin N. Hebart et al. “THINGS: A database of 1, 854 object concepts and more than 26, 000 naturalistic object images”. In: *PLOS ONE* 14.10 (Oct. 2019). Ed. by Fabian A. Soto, e0223792. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0223792. URL: http://dx.doi.org/10.1371/journal.pone.0223792.

[24] Martin N Hebart et al. “THINGS-data, a multimodal collection of large-scale datasets for investigating object representations in human brain and behavior”. In: *eLife* 12 (Feb. 2023). Ed. by Morgan Barense et al., e82580. ISSN: 2050-084X. DOI: 10.7554/eLife.82580. URL: https://doi.org/10.7554/eLife.82580.

[25] Lina Teichmann, Martin N. Hebart, and Chris I. Baker. “Dynamic representation of multidimensional object properties in the human brain”. In: *bioRxiv* (2025). DOI: 10.1101/2023.09.08.556679. eprint: https://www.biorxiv.org/content/early/2025/02/28/2023.09.08.556679.full.pdf. URL: https://www.biorxiv.org/content/early/2025/02/28/2023.09.08.556679.

[26] Meenakshi Khosla, N. Apurva Ratan Murty, and Nancy Kanwisher. "A highly selective response to food in human visual cortex revealed by hypothesis-free voxel decomposition". In: *Current Biology* 32.19 (2022), 4159–4171.e9. ISSN: 0960-9822. DOI: https://doi.org/10.1016/j.cub.2022.08.009. URL: https://www.sciencedirect.com/science/article/pii/S0960982222012866.

[27] Kalanit Grill-Spector, Zoe Kourtzi, and Nancy Kanwisher. "The lateral occipital complex and its role in object recognition". In: *Vision Research* 41.10 (2001), pp. 1409–1422. ISSN: 0042-6989. DOI: https://doi.org/10.1016/S0042-6989(01)00073-6. URL: https://www.sciencedirect.com/science/article/pii/S0042698901000736.

[28] Daniel Lee and H. Sebastian Seung. "Algorithms for Non-negative Matrix Factorization". In: *Advances in Neural Information Processing Systems*. Ed. by T. Leen, T. Dietterich, and V. Tresp. Vol. 13. MIT Press, 2000. URL: https://proceedings.neurips.cc/paper_files/paper/2000/file/f9d1152547c0bde01830b7e8bd60024c-Paper.pdf.