Decoding the representations for depth perception in the visual brain

Yiran Chen

Integrated Program in Neuroscience

McGill University, Montreal, Canada

December 2024

A thesis submitted to McGill University in partial fulfilment of the requirement of the degree of

Doctor of Philosophy

Table of Contents

Abstract

Human visual system can perceive depth in real world using binocular cues and monocular cues (shading, texture, motion cues, etc.). It is unclear how depth information presented in real-world scenarios can be used by the visual system to facilitate the recognition of object and scenes, which is largely omitted by current computer vision models (such as HMAX and newer convolutional neural network models) that achieve object recognition from 2-D edge information. The aim of the studies presented in this thesis is to develop interpretable multivariate data analysis tools for neuroimaging data and to use these tools to understand how objects and scene are represented in the visual brain with the presence of the depth cues. Three different approaches were adopted to investigate the visual processing of depth. In the first study, I investigated the cortical representations of naturalistic movie clips in a binocular version versus a monocular version, using magnetoencephalography (MEG) to measure multivariate responses pattern among different frequency bands. I found that the cortical representations of movie scenes were similar across subjects and were correlated with the depth of the movie scenes in the gamma frequency bands but not in the lower frequency bands. As gamma band activity is thought to reflect important stimulus properties, the finding that depth is a strong predictor of gamma activity suggests that depth is an important visual feature, in addition to contrast, spatial frequency, etc. If depth *is* a prominent stimulus feature, then the visual cortex ought to broadly represent it, in a manner similar to contrast. In the second study, I used functional Magnetic Resonance Imaging (fMRI) to investigate the functional connectivity (FC) changes in 3D movie viewing versus 2D movie viewing as a gauge of cortical engagement. I found that shared FC networks across the subjects reflected the viewing condition better than subject-specific FC networks. Inter-subject FC in dorsal visual stream subnetworks were most effective for decoding 3D vs 2D viewing conditions, while ventral stream subnetworks were effective to a lesser degree. Thus, depth information is represented broadly in the visual system, and cortical networks are dynamically engaged to process

the depth information in scenes. While the first two studies emphasized the central importance of depth representation, it is not clear whether scene and object percepts have multiple, separate depth-cue specific representations or depth-cue invariant representations, or both. In the third study, I investigated the cortical representations of 3D objects evoked by different depth cues using event-related multi-channel MEG decoding and using well-controlled renderings of objects using single depth cue. I discovered that the object information in the MEG data can be decoded and generalized across different depth cues but at different times. Depth cue specific object information emerged before the transferable generalization, supporting a two-stage theory that the cue-specific object information is processed separately before the depth-cue invariant object representation is formed. Across the three studies, I found that depth cue information is central in driving cortical signatures of scene and object processing, namely by stimulus-dependent engagement of gamma band oscillations and dynamic functional networks. Given the finding of the third study—that pure depth-cue stimuli always resulted in invariant and generalizable representations—and the fact that natural scenes, such as those used in the first two studies, always include multiple depth cues, the most parsimonious explanation is that not only depth is a fundamental stimulus attribute that is broadly represented across the cortex, but that object and scene representations are inherently depth-cue invariant and therefore implicitly three-dimensional. This stands at odds with cortical models of object recognition based on feed-forward convolutions and complex 2D feature representations.

Résumé

Le système visuel humain est capable de percevoir la profondeur dans l'environnement réel grâce à des indices binoculaires et monoculaires tels que les ombres, les textures et les indices de mouvement. La manière dont les informations de profondeur présentes dans des scénarios réels sont utilisées par le système visuel pour faciliter la reconnaissance d'objets et de scènes reste cependant peu claire. Cette dimension est largement négligée dans les modèles de vision par ordinateur actuels, tels que HMAX et les nouveaux modèles de réseaux neuronaux convolution els, qui dérivent leurs algorithmes de reconnaissance d'objets d'images 2D. L'objectif des études présentées dans cette thèse est de comprendre comment les objets et les scènes sont représentés dans le cerveau visuel lorsque des indices de profondeur sont forts. Trois approches différentes ont été adoptées pour étudier le traitement visuel de la profondeur.

Dans la première étude, nous avons examiné les représentations corticales de clips de films naturalistes en versions binoculaires et monoculaires, utilisant la magnétoencéphalographie (MEG) pour mesurer les patrons de réponses multivariées dans différentes bandes de fréquence. Nos résultats ont montré que les représentations corticales des scènes de films étaient similaires entre les sujets et étaient corrélées avec la profondeur des scènes dans les bandes de fréquence gamma, mais pas dans les bandes de fréquence plus basses. Étant donné que l'activité de la bande gamma est considérée comme reflétant des propriétés importantes des stimuli, la constatation que la profondeur est un fort prédicteur de l'activité gamma suggère que la profondeur est une caractéristique visuelle importante, au même titre que le contraste ou la fréquence spatiale.

Dans la deuxième étude, nous avons utilisé l'imagerie par résonance magnétique fonctionnelle (IRMf) pour étudier les changements de connectivité fonctionnelle (CF) lors de la vision de films 3D versus la vision de films 2D, comme indicateur de l'engagement cortical. Nos résultats ont montré que les réseaux de CF partagés entre les sujets reflétaient mieux la condition de vision que les réseaux de CF

spécifiques aux sujets. La CF inter-sujets dans les sous-réseaux des voies visuelles dorsales était la plus efficace pour décoder les conditions de vision 3D vs 2D, tandis que les sous-réseaux des voies ventrales étaient moins efficaces. Ces résultats indiquent que l'information de profondeur est largement représentée dans le système visuel et que les réseaux corticaux sont dynamiquement engagés pour traiter cette information dans les scènes.

Dans la troisième étude, nous avons étudié les représentations corticales d'objets 3D évoquées par différents indices de profondeur, utilisant une méthode de décodage MEG multi-canal liée à des événements et un rendu bien contrôlé d'objets à indice de profondeur unique. Nos résultats ont montré que l'information d'objet dans les données MEG pouvait être décodée et généralisée entre différents indices de profondeur, mais à des moments différents. Les informations d'objet spécifiques aux indices de profondeur ont émergé avant la généralisation transférable, soutenant une théorie à deux étapes selon laquelle les informations d'objet spécifiques aux indices sont traitées séparément avant que la représentation d'objet invariante aux indices de profondeur ne soit formée.

Dans l'ensemble de ces études, nous avons constaté que l'information d'indice de profondeur joue un rôle central dans la conduction des signatures corticales du traitement de scènes et d'objets, notamment par l'engagement dépendant du stimulus des oscillations de bande gamma et des réseaux fonctionnels dynamiques. Compte tenu des résultats de la troisième étude - qui montrent que des stimuli d'indice de profondeur pur donnent toujours des représentations invariantes et généralisables - et du fait que les scènes naturelles, telles que celles utilisées dans les deux premières études, incluent toujours plusieurs indices de profondeur, nous pouvons conclure que la profondeur est non seulement un attribut fondamental des stimuli largement représenté dans tout le cortex, mais que les représentations d'objets et de scènes sont inhéremment invariantes aux indices de profondeur et donc implicitement tridimensionnelles. Cette conclusion est en contradiction avec les modèles corticales de

reconnaissance d'objets basées sur des convolutions d'avant en arrière et des représentations complexes de caractéristiques 2D.

Contribute to Original Knowledge

The contribution to original knowledge from this thesis is the demonstration that depth cues play a central role in driving cortical representations of scene and object in the human visual system. Specifically, the studies revealed that:

- Cortical representations of movie scenes are related to the gamma frequency bands, suggesting that depth is a vital feature in movie scene recognition in the human visual system.
- Shared functional connectivity networks across subjects reflect the viewing condition better than subject-specific networks, and the dorsal visual stream subnetworks are particularly effective for decoding 3D vs 2D viewing conditions, indicating that depth information is broadly represented and dynamically processed in the visual system.
- Cortical representations of 3D objects evoked by different depth cues can be decoded and generalized across cues, supporting a two-stage theory of object processing where cue-specific information is processed separately before forming a depth-cue invariant representation.

Overall, these findings challenge cortical models of object recognition based on feed-forward convolutions and complex 2D feature representations and suggest that object and scene representations are inherently depth-cue invariant and implicitly three-dimensional.

Acknowledgment

I sincerely thank my supervisor Dr. Reza Farivar for the strong support and guidance during my work in the lab. I also thank my advisory committee members Dr. Christopher Pack and Dr. Sylvain Baillet for the helpful comments and feedbacks on my progress during my PhD. I am honored to meet and work with the amazing coworkers, including Angela Zhang, Tatiana Ruiz, Sebastien Proulx, Hassan Akhavein, Luiza Passos Volpi, Giovana dos Santos Cover, Laurie Goulet, Shael Brown, William Mathieu, Roland Pilgram, and Haneieh Molaei. Special thanks to Sebastien Proulx, Angela Zhang, and Hassan Akhavein for the comments in data analysis. I would like to thank the IPN program for the administrative support. I thank my wife Han Hao for family support.

Contribution of Authors

Under the expert guidance of my supervisor, Dr. Reza Farivar, I designed the experimental framework, conducted the data analysis, and authored the introduction and discussion to the thesis.

For Chapter 2, Angela Zhang and I jointly gathered the data. Dr. Reza Farivar and I formulated the hypothesis and collaborated on the experimental design. I was responsible for analyzing the data and writing the manuscript.

For Chapter 3, Angela Zhang and I jointly gathered the data. Dr. Reza Farivar and I formulated the hypothesis and collaborated on the experimental design. I was responsible for analyzing the data and writing the manuscript.

For Chapter 4, Armita Dehmoobadsharifabadi and Hassan Akhvein contributed to the subject recruitment and data collection. Dr. Reza Farivar and I collaboratively formulated the hypothesis and devised the experimental design. I conducted the data analysis and authored the manuscript.

Chapter 1: Introduction

In this thesis, I investigate the representations of depth perception in the human brain by using different neuroimaging techniques. Neuroimaging, particularly functional MRI (fMRI) and magnetoencephalography (MEG), provide extremely rich data in time and space, which result in massive, multidimensional data with particular challenges pertaining to accurate anatomical registration and effective group inference. To tackle these challenges, tools must be developed to analyze, abstract, and present the data in a way with simplicity. In visual neuroscience, this goal is to derive interpretable models to convey the ideas of perceptual representations on real scenes and their contents, which are visual objects. In this introduction, I address the two challenges that need to be considered: (1) Improvements in the spatial and temporal resolution of neuroimaging techniques necessitate increased complexity of multivariate methods to draw valid inferences and to test models of cortical function and (2) the anatomical and functional space of the neuroimaging data cannot be perfectly aligned across multiple individuals, resulting in neuroimaging analysis that can have significant dependence on individual anatomy, which in turn may bring inaccuracies to group analyses.

Multivariate analysis can be used here to tackle both challenges (Feilong et al 2018, Guntupalli et al 2016). Firstly, multivariate analysis can be used to concentrate on specific neural activity patterns, reducing the dimensionality of neuroimaging data. Secondly, multivariate analysis, especially representational similarity analysis (Kriegeskorte & Diedrichsen 2019, Kriegeskorte et al 2008a), can be used to create interpretable abstraction of single-subject data, allowing for testing of cross-subject prototypical effects (Chen et al 2020b, Finn et al 2020, Nastase et al 2019). For these reasons, this thesis utilizes and extends multiple multivariate methods to the analysis of complex neuroimaging data to answer fundamental questions in visual neuroscience.

A crucial topic in visual neuroscience pertains to the representation of objects and scenes (Bar 2004, DiCarlo & Cox 2007, Oliva & Torralba 2007). While early models of object recognition emphasised 3-D representations (Biederman 1987, Hummel & Biederman 1992), the current emphasis has shifted to 2-D representations (HMAX (Riesenhuber & Poggio 1999); Updated HMAX (Serre et al 2007)), with success of convolutional neural networks supporting the utility of 2-D edge information for complex object representation (DiCarlo & Cox 2007, Kheradpisheh et al 2016, Kriegeskorte 2015, Riesenhuber & Poggio 2000, Wen et al 2018). But interestingly, we can recognize complex objects complete devoid of 2-D edge information as long as 3-D information is present (Akhavein et al 2018, Dehmoobadsharifabadi & Farivar 2016, Farivar 2009, Farivar et al 2009) suggesting depth information can play a crucial role in object understanding and recognition.

There is evidence for dissociable cortical representations of individual depth cues, including structure-from-motion (Peuskens et al 2004), disparity(Brooks 2017, Goncalves et al 2015, Verhoef et al 2016), shading (Norman & Wiesemann 2007), texture (Ichihara et al 2007), etc. Depth perception is important in human perception of objects and scenes in both naturalistic settings and lab-created environments. In the following chapters 2-4, I pursued three studies that investigate depth perception using multiple neuroimaging techniques to understand the following questions: 1) Does depth induce similarity across subjects and if so, how? and 2) How do different visual depth cues integrate to form coherent object representations?

1.1 Measuring object representations with non-invasive neuroimaging

There exist approximately 100 billion neurons in the cerebral cortex and the number of connections that are formed by the neuronal synapses in the brain are estimated to be 125 trillion (Herculano-Houzel 2009). Undoubtedly, the quantity of neurons and the connections in the brain are too great to be measured individually (von Bartheld et al 2016). Non-invasive neuroimaging techniques instead focus on the mesoscale to infer the local brain response. These methods greatly reduce the number of brain

regions needed to be measured for mapping the whole brain, but even measurements this mesoscale produce too much spatial data to be readily comprehensible. Combined with temporal data at different scales (seconds or milliseconds), this tremendous data flow imposes challenges for analysis. In the following sections, challenges, traditions, and developments of neuroimaging approaches will be discussed for tackling these analysis problems.

Neuroimaging data can be analyzed to demonstrate spatio-temporal linkage and to relate these to perceptual representations in the brain (Cichy et al 2014, Kriegeskorte & Diedrichsen 2019, Kriegeskorte et al 2008a). The goal of analysis (and neuroimaging analysis) is to derive valid, interpretable inferences from the data under some specific assumptions. Under each method's set of assumptions, inferences can then be made to identify the relationship between inputs—in our case, the visual scene—and the outputs—in our case, spatial and/or temporal patterns of brain activity (Cohen et al 2017).

One common approach is to consider neural activity patterns as a vector of quantities that are measured at various brain locations at a given time. Neural activity patterns can be measured e.g., by fMRI, represented by voxel BOLD patterns, or for MEG, magnetic sensors pattern data could be acquired. Note that, depending on the measurement method used, a neural activity pattern represents different physical entities (either BOLD signal changes in case of fMRI (Logothetis 2003), or magnetic potentials acquired by the MEG sensors in the case of MEG (da Silva 2010). Such vectors can be taken to represent brain activity such that they jointly relate to how the brain is processing or representing the inputs (in the case of the studies in this thesis, visual inputs), all else being equal. However, *inferring* the relationship between neural measurement representations and the stimulus input, or even the relationship between different measures of the same input, necessitates multivariate analysis that allows for some degree of abstraction (Kriegeskorte & Diedrichsen 2019). In the following sections, the principles underlying two popular neuroimaging measurement tools are discussed, followed by a treatment of multivariate analysis tools and the ways such results have been interpreted.

1.1.1 Functional magnetic resonance imaging

Functional MRI (fMRI) measures the blood-oxygenation-level-dependent (BOLD) image contrast, which is a combination of blood oxygenation, blood volume, and blood flow (Logothetis 2003, Logothetis et al 2001) BOLD changes are understood to occur due to local brain tissue hemodynamics and metabolic modulations, which are related to neural activity (Logothetis 2001, 2008). BOLD signals largely reflect the synaptic transmission of neurons, which are related to local circuits (Jueptner & Weiller 1995). The shape of the hemodynamic response can vary from region to region as well, highlighting the extent to which it may represent very local vascular and neural circuits (Handwerker et al 2004). Typical fMRI datasets are represented as four-dimensional volume series of image pixel intensities over time, and can be normalized to represent percent signal changes. The size of the fMRI voxels in the 3-D volume are usually about 2-3 mm^3, producing a sufficient spatial resolution for identifying spatial patterns that may contain information—for example, information that could be used to decode the preferred orientation (Kamitani & Tong 2005) or even inferring the structure of the input image (Nishimoto et al 2011, Rakhimberdina et al 2021). It is fascinating that such information can be recovered from voxels, given that even one cubic millimeter of macaque primary visual cortex contains already approximately 120,000 neurons (Palackal et al., 1993).

In a typical visual cognition study, BOLD signals in each fMRI dataset are analyzed using a general linear model (GLM; Handwerker et al 2004, Monti 2011, Poline & Brett 2012) which estimates the beta to a timeseries model consisting of the stimulus onsets with the hemodynamic response model. GLM allows for setting up of simple and complex contrasts between conditions, and the resulting statistical parametric maps can be thresholded to identify brain regions that are linearly related to the conditions—they exhibit a difference in the relative magnitude of the BOLD signal as estimated using the difference of the beta of the GLM models for each condition. In this sense, the GLM approach allows

one to search brain areas that differentiate between two or more conditions in mean amplitude. Alternatively, the GLM analysis can be conducted on a selected set of voxels or brain regions of interests (ROIs)—this is done by pooling all the voxels in each brain location to determine condition-induced amplitude variations in the BOLD signal in different ROIs and to increase sensitivity to detecting differences.

While there are multivariate GLM methods, the conventional approach to fMRI analysis is univariate—each voxel's magnitude over time is modelled as a sum of linear terms. Despite being widely employed, the univariate approach suffers limitations in relation to the linear assumptions of the BOLD amplitude and independence of voxels (Nimon 2012). The independence of voxels assumes that the covariance of the multi-voxel data possesses no information regarding the hemodynamic response, so that the model will bare no information on the interactions between voxels. The univariate model also assumes that the error term, like all the linear regression models, is independent and identically distributed (i.i.d.) as Gaussian variables for each of the voxels or ROIs (Monti 2011, Penny et al 2011, Worsley & Friston 1995).

While the multiple comparison problem is universal to any type of analysis that seeks to detect differences, the problem is of particular concern with univariate GLM analysis because each voxel is treated independently. Conservative approaches to controlling the rate of false positives, such as Bonferroni correction (Logan et al 2008, Logan & Rowe 2004) can be applied but are excessively contraining given the fact that one test is done at each voxel. A popular alternative is the Family-Wise Error correction based on Keith Worsley's research in the 1990s (Worsley et al 1996) which uses concepts from Guassian Random Field Theory (RFT) to generate a threshold that is more liberal than Bonferroni methods but principled so as to avoid false positives. RFT utilizes a spatial smoothing element to take into account the smoothness of the activation map in the brain, and the null hypothesis for RFT's statistical inference is that the signal at each voxel is a random variable with a Gaussian

distribution with zero mean and a known variance. In RFT, the goal is to identify clusters of contiguous voxels that show a significant deviation from the null hypothesis. The p-value for significance is assigned to each cluster rather than individual voxels, and the size of the cluster is taken into account in the detection of significance. Worsley and colleagues (Worsley et al 1996) have shown that the use of RFT can lead to a more accurate identification of significant brain regions than traditional voxel-based analysis methods. The RF method is therefore based on an intuitive assumption that voxels that are closer to each other should be similar in activation (i.e., perform a similar function), thus utilizing the spatial arrangement or spatial *patterns* (in this case, correlated activity) for statistical inference and shed light on the idea that the spatial patterns are relevant to understanding brain function. However, at its core, it is still a linear univariate analysis technique, with the caveats that the spatial variations in amplitude of neighbouring voxels are largely ignored and a cluster's correlated activity helps it reach detection threshold using RFT methods. As discussed below, multivariate analysis overcomes this important shortcoming of univariate methods and can allow for exploring the relationship between cortical spatial patterns of activity and experimental conditions.

1.1.2 Magnetoencephalography

Magnetoencephalography (MEG) is an alternate brain imaging technique that affords very high temporal resolution, but poor spatial resolution as compared to fMRI. In MEG systems, magnetometers measure the magnetic fields that are generated by the post-synaptic ionic currents (Baillet et al 2011, Hamalainen 1991, Hamalainen & Ilmoniemi 1994). Large, synchronized voltage-dependent activation gates on the dendritic neuronal membranes generate magnetic field changes that MEG sensors can measure, and this can be used to infer brain activity. It is worth noting that action potentials in the cortex generates trans-membrane electric currents in opposite directions, which associate magnetic fields that canceled out each other, thus will not be detected by MEG (Fred et al 2022, Hamalainen

1991). MEG is sometimes perceived as a sister electrophysiological technique to electroencephalography (EEG), but the signal sensitivity of MEG is superior to EEG's as the magnetic field signals are largely unaffected by tissue conductivity from scalp, skull and cerebrospinal fluid (CSF), in contrast to the electric field signals that would be distorted by tissue conductivity (Singh 2014). MEG also possesses better spatial resolution than EEG. However, magnetic fields signal decay as the distance increases, which leads to lower signal to noise ratio when detecting deep brain signals using MEG (Andersen et al 2020, Salmelin & Baillet 2009, Singh 2014). MEG datasets exhibit high temporal resolution—in the range of milliseconds— well-suited for measuring fast and dynamic brain activities. Furthermore, MEG is a whole brain coverage measurement tool, allowing the simultaneous measuring the activity of distant brain areas (Baillet 2017, Brookes et al 2011, Hillebrand et al 2012).

Using inverse modeling (Baillet & Garnero 1997, Friston et al 2008, Mosher et al 1999), MEG data can be used to infer the localization of brain source activations in brain mapping research. MEG generated time-resolved brain source maps are more accurate compared to attempts made from EEG-generated brain source maps, though they still suffer from inevitable distortions due to head shape inaccuracies, interaction with the skull bone, and the blood vessels that generate electric currents and magnetic fields (Baillet 2017, Vorwerk et al 2014). Previous research has shown that EEG signals can be used to separate the dipoles in a realistic skull phantoms with 7-8 mm accuracy, while the location error of MEG signals can reach 3mm (Leahy et al 1998). Brain regions that produce tangential current flows to the sensor orientation can yield MEG signals that are substantially stronger than those produced from radial current flows, thus providing the strongest component of the event-related MEG signals recorded (Goldenholz et al 2009, Hämäläinen et al 1993). In summary, MEG combined with source localization is a valuable methodology in neuroscience as it provides an alternate means of measuring population activity, allowing for high temporal resolutions over the entire brain non-invasively (Baillet 2017, Darvas

et al 2004, Schoffelen & Gross 2009). Similar to fMRI analysis, source localization maps can be analyzed using GLM statistical mapping methods mentioned above (i.e., univariate analysis).

1.1.3 Challenges of aligning functional brain maps in multi-subject datasets

The goal of the vast majority of neuroimaging studies is group inference—to test a hypothesis or explore a pattern across a large group of subjects so as to be able to claim generalizability to the population at large. By analyzing the datasets from multiple subjects, this generalization validity is increased, due to the fact that the noise in each subject's brain map is uncorrelated to the others and averaging them could therefore increase the signal-to-noise ratio. To aggregate maps across subjects for group-level inference, one needs to align the subjects somehow, allowing the derivation of a common spatial template—the data carried in the individual anatomical space must be aligned, registered, or projected to a common coordinate space (Ardekani et al 2004, Cox 1996, Cox & Hyde 1997, Tahmasebi et al 2009). These steps are currently performed through alignment algorithms which rely on anatomical landmarks of individual subjects. For instance, a three-dimensional affine transformation or a more advanced non-linear warping technique has been used to register the subject's sulci and gyri in addition to brain volume (Ardekani et al 2004, Cox 1996). Surface-based registration has improved resolution compared to volumetric registration, but the data are still wrapped based on anatomical landmarks (Hagler et al 2006, Klein et al 2010).

Anatomical similarity does not guarantee functional similarity—in other words, alignment based on anatomical features or landmarks does not ensure that the functional patterns elicited by a given patch of cortex across two or more participants will also be similar (Guntupalli et al 2016, Tahmasebi et al 2009). However, anatomical registration drastically improves group-wise inference and it is now part of the conventional processing pipeline (Ardekani et al 2004). Due to functional topography differing across subjects, the quality of multiple subject registrations may remain as one of the greatest

challenges in brain imaging studies, although increases in the resolution of fMRI images is improving our ability to align subjects possibly even without an anatomical image (Dohmatob et al 2018).

Consequently, one process commonly observed in the fMRI data analysis pipeline is to spatially smooth the data in order to make the functional topography spatially coarse (Mikl et al 2008). On a coarse level, the mismatch between functional topography and the anatomical registration is reduced and can be ignored but blurring/smoothing the functional data may not be appropriate when the cortical system under study exhibits fine spatial structures for accurate measurement.

Besides the fact that functional areas are not (at a fine resolution) consistently located relative to anatomical landmarks, functional cortical areas also have spatial structures that are much finer than anatomical landmarks such as gyri or sulci. Anatomically-defined brain areas typically extend beyond 1 cm^2 on the cortical surface (Haxby et al 2011), while topographic brain maps may exist at much finer scales—for example, in the retinotopic map of V1, the functional representations for a location in the visual field are distinct from nearby locations in the visual field and are only millimeters apart from each other (Cheng et al 2001, Wandell & Winawer 2011, Yacoub et al 2007). Similar organization of fine topographical maps can be observed in the auditory cortex (Saenz & Langers 2014, Schonwiesner & Zatorre 2009), somatosensory cortex (Stringer et al 2011), and other cortical areas (Dale et al 2000). In the ventral temporal cortex, noticeable patterns of activity in distinct areas are elicited by images of objects, suggesting that these patterns are reflecting object features (Bracci et al., 2017; Contini et al., 2017). In particular, the fine spatial patterns of activity can potentially carry important features regarding representations that the anatomical models fail to capture at their coarse level. Such findings pose further challenges to the usual processing steps of blurring the data or registering the multi-subject dataset (Sabuncu et al., 2010).

Certain cross-subject registration algorithms can be implemented to overcome some of the difficulties of registering multi-subject data and to increase the fine-scale information that carries

functional motifs. One method is to register multi-subject datasets by utilizing functional data within individual subjects as guidance (Guntupalli et al 2016, Haxby et al 2011). Various studies have been conducted by aligning cortical functional maps across subjects using functional connectivity (FC)—a within-subject similarity measurement over regions of the cortex (Bryan R. Conroy 2009). Other approaches have aligned functional cortical topography across subjects through neural activity patterns within a searchlight—a vicinity of cortical locations related to the one of interest (Guntupalli et al., 2016; Sabuncu et al., 2010). However, this analysis involves potentially "warping" the data, and runs the risk of adding one layer of complexity to the analysis and interpretation of the data.

1.2 Assessing cortical representation with multivariate analysis

Neuroimaging techniques such as fMRI and EEG measure the activity of groups of neurons, rather than individual neurons and this can make it difficult to use a single model such as univariate analysis to describe the complex and varied response patterns of these neuronal populations to different stimuli or experimental conditions. In the case of visual perception, this can lead to the activation of multiple visual columns in primary visual cortex (V1) within a single voxel typical 3×3×3mm fMRI voxel, which can generate conflicting responses to visual stimuli. This is because different columns in V1 are specialized to process different aspects of visual information, such as orientation, spatial frequency, or possibly colour (Carandini 2012, Purves D 2001). Several studies have investigated the spatial resolution of fMRI in V1 and found that even at high field strengths and with advanced imaging techniques, the spatial resolution of fMRI in V1 is limited to hundreds of micrometers or larger (Olman et al., 2012, Kay et al., 2013, Kok et al., 2016). This means that fMRI voxels typically contain neurons from multiple visual columns in V1.

This problem is not restricted to V1. Different groups of neurons may exhibit different response patterns depending on their location, functional specialization, and interactions with other brain regions. Therefore, developing analytic models that can accurately capture the complexity and

variability of brain activity across different groups of neurons remains a major challenge in neuroscience research. Multivariate analysis, instead of univariate analysis, is a suitable method when searching for reproducible activity patterns over experimental conditions which may activate complex response patterns in the brain, overcoming the inclusive and complexity problem in the nature of neuroimaging signals. In a review article, Poldrack et al. (2011) argued that the assumption of linearity is often invalid in the analysis of brain data, and non-linear multivariate methods are more suitable for analyzing complex patterns of brain activity. Park et al. (2013) demonstrated that the functional connectivity of brain regions varies across individuals, and the brain organization is highly individual-specific, which makes it challenging to achieve accurate registration across different subjects.

At the group level, multivariate analysis has also been shown to be able to make inter-subject inferences. Chen et al. (2016) demonstrated that multivariate analysis techniques, such as support vector machine (SVM) and principal component analysis (PCA), can identify reliable and meaningful patterns of brain activity across subjects, even in the presence of inter-individual differences. Varoquaux and Craddock (2013) reviewed "connectivity-based parcellation" that allows for the identification of functional regions within the brain based on patterns of connectivity, rather than anatomical landmarks. This method showed that multivariate pattern analysis (MVPA) can reveal reproducible patterns of brain activity across subjects, even when univariate approaches fail to do so (Kriegeskorte & Bandettini 2007, Kriegeskorte et al 2006).

Multivariate analysis in brain imaging here refers to the supervised machine-learning classification problem in which classifiers attempt to identify the spatial patterns that differ between the experimental conditions (Haxby et al 2001, Kriegeskorte et al 2006). The algorithm determines a decision function predicated on features in a dataset and subsequently predicts the experimental conditions from which that dataset will be collected. A feature is a machine-learning term that refers to

variables or attributes. Individual features may be treated as one dimension in a high-dimensional space, and the dataset could be one data point in that high-dimensional space (Mahmoudi et al., 2012).

To obtain the decision function, the entire data must be categorized into two subsets: a training set, and a testing set. Using the training set, a classifier is trained, and the predictions are then formulated for the labels of the testing sets. During the training process, machine learning tools are used to manipulate the weights that are assigned to each of the features in the dataset that contribute to the decision function. Training continues until either a number of epochs have passed, or a target level of accuracy has been achieved.

The classifier is not limited to linear models in the feature space, like linear discriminant analysis (LDA). Other categorical classification methods, like support vector machines (SVM), are also commonly used. A linear classifier model can be perceived as the activity pattern of the feature space, which is easy to understand; non-linear classifiers like neural networks are often considered to lack direct plausible biological interpretations. However, neuroimaging studies of the human ventral visual stream demonstrate that representations of later stages in the hierarchy sometimes mimic the higher order layers in the neural network (Dobs et al 2022, Guclu & van Gerven 2015, Nonaka et al 2021). In summary, features in multivariate analyses in neural imaging are considered interactive, differing from a univariate approach that often views responding units to be independent, and fails to analyze the relationship between features. The feature interactions during multivariate analysis are related to patterns of activities in the given cortical areas, or even related to representational transformation across the hierarchy of cortical visual processing.

1.2.1 Approaches to multivariate analysis in neural imaging

1.2.1.1 FMRI multi voxel pattern analysis

Multi-voxel pattern analysis (MVPA) involves searching for reliable spatial BOLD patterns that can be differentiated depending upon the experimental conditions. As classifiers (typically linear classifiers) are being trained to probe for the connections between experimental conditions and fMRI spatial patterns, MVPA is considered a data-driven supervised learning problem (Haxby et al 2014, Mahmoudi et al 2012, Oosterhof et al 2016). During MVPA, various fMRI voxels are defined as "variables" or "features", which are commonly used in machine learning. Data of each trial in the experiment serves as an "example" in this context. The data is then categorized into training and testing set to provide for independent validation. The classifier on the training set serves to maximize a decision function, in which weights will be allocated to each feature to determine the labels in the training set. This weight vector denotes the relative contribution to the decision for each feature. In the case of multiple classes, the analysis often results in multiple two-class problems. Subsequently, the classifier is evaluated via the held-back testing set. Using the proportion of correct labels that the classifier returns, an accuracy value is then computed (Haynes & Rees 2005, Kamitani & Tong 2005).

Without cross-validation, the data can be categorized into training and testing sets through multiple ways, therefore generating multiple estimates of classification accuracy. Among said partitions, there is a maximum accuracy. To overcome the possible biases that the classifier may be overfitting during the split, cross-validation is often introduced to efficiently evaluate the classifier performance (Lemm et al., 2011). A commonly employed method for formulating the classifier is as a linear support-vector machine (SVM), in which the decision is concluded by a boundary of hyperplanes that can separate the labels of examples in the multi-dimensional space. Benefits of the SVM model are the high performance and the applicability to large high-dimensional datasets (Peltier et al 2009, Song et al

2011). A linear SVM prevails in a multitude of MVPA fMRI studies due to its interpretability in the decision boundaries (Guggenmos et al., 2018; Peltier et al., 2009).

1.2.1.2 MEG sensor decoding analysis

Due to the limited temporal resolution of BOLD signals, the application of MVPA to fMRI data reveals little regarding the temporal dynamics of mental representations. However, by employing either EEG, MEG, or intracranial recording methods, condition-related information can be extracted from the temporal dynamics. Various behavioral tasks, including motor tasks (Waldert et al., 2008), auditory tasks (King et al., 2013), conceptual tasks and semantic information (Chan et al., 2011; Nishida & Nishimoto 2018), and music genres (Schaefer et al., 2011), can all be decoded from MEG or EEG recordings.

The temporal mental representation of dynamic information can be obtained from training multiple classifiers at each time point on the condition labels from the experiment (King & Dehaene 2014; Marti et al., 2015). Using a time-varying dynamic decoding approach, two kinds of analyses can be performed. The first form of analysis is to test the decoding performance at each specific time point using the same data from that time point as testing sets. The result is a decoding accuracy over time, which reflects how the information flow changes over time while the cortex is processing the information (Carlson et al 2013, Isik et al 2014, King & Dehaene 2014).

The second approach is generalization over time—after training the classifiers at each time point, the testing sets are determined from each of the other time points on the recordings. The result is a generalization of the temporal map of the decoding performance, showing that the neural code identified once can recur at a different time point. Using the map of temporal generalization, one can determine not only the timepoint that condition labels become dissociated from one another, but also the timepoint at which this classification pattern becomes general, as in “generally valid” (King & Dehaene 2014). Additionally, possible generalizations over conditions allow interpretations of how different conditions induce reorganization of cortical processes (King et al., 2016). The MEG and EEG

decoding analyses are typically performed on the raw sensor data level, without a reverse projection to a brain source level, thus saving from interpretations of the data (Cichy & Pantazis 2017, Guggenmos et al 2018). Experimental conditions can also be decoded from a time-frequency domain, in which both time and frequency information can be assessed.

1.2.1.3 Representational similarity analysis (RSA)

MVPA allows the activity patterns of the brain to be interpreted as mental representations. To further characterize the mental representation, the activity patterns can be compared between each pair of experimental conditions, forming a representational dissimilarity matrix (RDM) (Kaneshiro et al 2015, Kriegeskorte et al 2008a). Each element in the RDM is the quantified distinctiveness of that specific condition pair. In RDM, correlation distance is a commonly used to quantify the distinctiveness measure, although Euclidean distance and cross-validated Euclidean distance are also utilized in analysis of fMRI data. For MEG data, the Mahalanobis distance is typically employed to reduce bias, because it represents the distance from a datapoint to designated distribution. (Allefeld & Haynes 2014, Kaneshiro et al., 2015).

RDM matrices can serve as signatures of specific representations and can be directly compared between brains and models—in other words, each model can be used to form an idealized RDM, and the distance between the model RDM and the measured RDM can be compared. This abstraction broadens the view of neuroimaging research by changing how data is interpreted from the comparison of activity patterns to the comparison of representational patterns (Anderson et al., 2016; Fischer-Baum et al., 2017; Kriegeskorte & Diedrichsen 2019).

To produce a direct comparison between the representations, RDMs can be compared by using 2nd order methods like rank-correlation. This method of comparison is referred to as representational similarity analysis (RSA) (Cichy et al 2016a, Grootswagers et al 2017, Kriegeskorte & Diedrichsen 2019).

In psychological terms, analysis of similarity structures of representations is a second-order isomorphism (Edelman 1998), meaning that this measurement is one abstract level beyond the stimulus-induced activity patterns, which can be deemed as a first-level isomorphism stimulus-response representation (Oliva & Torralba 2007, Riesenhuber & Poggio 1999, Tsunoda et al 2001). Employing RSA allows a potential merger of combinations of neuroscience branches through a second-level isomorphism (Guggenmos et al 2018, Kriegeskorte & Diedrichsen 2019).

RSA has at least four noticeable advantages. The first advantage is the provision of quantification tools for mental representations (Kriegeskorte & Diedrichsen 2019, Kriegeskorte et al 2008a). Evaluations and comparisons can be conducted by using said metric on different brain regions (Guntupalli et al 2016), different time points (King & Dehaene 2014), and even different frequencies of oscillatory brain waves (Xie et al 2020). To illustrate, it is possible to use RSA to provide answers to the following questions: which brain area best explains the conceptual model of the nature of the stimulus (Guntupalli et al 2016)? And, to what extent do two different brain areas process information similarly (Cichy et al 2016b, Kim et al 2017)? The second advantage of RSA is that quantitative comparisons can be observed between regions (Guntupalli et al 2016), subjects (Chen et al 2020b), modalities (Cichy et al 2016b), or even species (Kriegeskorte et al 2008b). The dimensions that the activity patterns are predicated on can differ among datasets. By comparing the abstracted RDMs, the dimensional mismatch can be circumvented (Kriegeskorte & Diedrichsen 2019). The third advantage is that brain-activity measures can be utilized to observe relationships with behavioral performance (Jeong & Xu 2016). And the fourth advantage is that the RSA allows comparison of abstract conceptual models by broadening the stimulus design, incorporating multiple complex conditions, and addressing multiple questions simultaneously (Bokeria et al 2021).

Combining fMRI and MEG to investigate the temporal and spatial dynamics of cortical processes is one of RSA's most exciting qualities (Cichy et al 2014, Cichy et al 2016b). Achieving accurate mapping in

both high temporal resolution and high spatial resolution in neuroimaging research is a difficult task. For current non-invasive neuroimaging approaches, fMRI offers high spatial resolution, often at the level of cubic millimeters, but suffers a low temporal resolution on the level between hundreds of milliseconds or seconds. In contrast, the temporal resolution of MEG and EEG is on the level of milliseconds, but the spatial resolution is inferior compared to fMRI. Combining fMRI, MEG, and EEG to achieve simultaneous high temporal and spatial resolution is challenging yet necessary to comprehend information propagation in the brain when completing complex tasks. In said case, RSA may function as a useful tool—it can be assumed that the neural representation of fMRI data and MEG data are similar in the same cortical processes (Guggenmos et al 2018). Under this assumption, the local neural representations of fMRI that matches with MEG's can be observed via the searchlight technique, allowing the spatiotemporal resolved neural activation pattern to be re-established (Cichy et al., 2014, Cichy et al., 2016).

1.2.1.4 Alternative multivariate analysis in non-invasive neuroimaging

Other than MVPA, several multivariate analysis methods have been established to understand the nature of the neural code and to solve the imperfect alignment of the anatomical and functional maps, of which a select few will be discussed here. The first approach that aims to solve the mismatch is hyper-alignment and the shared response model (Chen. et al 2015). In the shared response model, the high-dimensional fMRI data is often projected into a low-dimensional component space. For instance, in a shared response model, the subjects are exposed to a series of task sequences, which elicits different brain processes (for example, visual, auditory, etc.). Each of the identified shared responses maximizes the shared variance with involving brain states. Through the shared response model, cross-subject decoding can be established (Chen. et al 2015, Cohen et al 2017).

The second approach is the extraction of spatial priors from the fMRI data based on an understanding of how the brain organizes cognitive functions spatially. A particular assumption is that the brain's response to a specific stimulus is sparse and smooth (De Brecht & Yamagishi 2012). The sparse prior assumption is that only a small subset of the brain's locations is responsive to the stimuli. Regarding the smooth response, using prior knowledge, it is assumed that if one location on the brain is reacting to the stimulus, the adjacent locations are also likely to respond. Such models can be combined with Bayesian models to statistically map the brain across multiple subjects (Manning et al., 2014).

A third approach is to extract patterns of connections between brain locations using correlation of activity (Gonzalez-Castillo & Bandettini 2018, Guntupalli et al 2018). A functional connectivity approach can capture information that MVPA fails to locally represent. An example of attainable information is attention, in which certain brain regions are modulated through a top-down influence. Functional connectivity has been demonstrated to reveal diffused brain functions in supporting cognitive tasks, including attention (Al-Aidroos et al., 2012; Regev et al., 2019), adaptation (Cole et al., 2013), and other cognitive tasks.

1.2.2 Decoding mental representations using activity pattern

The concept of decoding is not to directly model the brain's information processing, but rather to "crack the neural code" (Cox & Savoy 2003, Haynes & Rees 2006, Tong & Pratte 2012). Decoding is intended to provide insight into the mystery of the brain's representation and has proved to be advantageous in numerous ways. One advantage is that it can be employed in brain mapping (Haxby et al 2014, Haynes & Rees 2006, Nishida & Nishimoto 2018, Zhang et al 2020). Individual brain areas can be investigated separately and thus, decoding works well with localized approaches such as the searchlight technique (Kriegeskorte et al 2006). Additionally, the fine-grained multivariate information can be

employed in the decoding models, significantly enhancing the resolution of multivariate analysis (Guntupalli et al 2016).

However, decoding analysis is limited in its interpretability. The decoding models employed for categorial label predictions are classifiers and cannot be seen as brain models *per se* when processing information (Haynes & Rees 2006, Kriegeskorte & Diedrichsen 2019, Kriegeskorte & Douglas 2019, Kriegeskorte & Wei 2021). With regards to investigating visual perception, decoding models are fundamentally the inverse of the brain's processes, meaning that decoding models are trying to predict the input of the system rather than output. Decoding models often rely on linear models, which is not the accurate reflection of possible non-linear processes of the brain (Allefeld & Haynes 2014, Haxby et al 2014). In summary, decoding is a useful tool for evaluating if information is present in the region, especially in multivariate analysis.

1.2.3 Information encoded in network connections

A unique perspective on how cortical activity organizes in health and disease is gained by investigating the *functional connectome*. In recent years, resting state networks, as measured by fMRI, have been exceptionally useful in the classification of numerous mental disorders, including autism (Abraham et al., 2017), major depression (Rosa et al., 2015)(Zeng et al 2014), and mild cognitive impairment (Bai et al 2009). Importantly, fMRI connectivity networks can also be utilized to decode mental states (Jiang et al., 2015), visual attention (Di & Biswal 2020; Sun et al., 2021), and auditory stimuli features (Zhang et al., 2020). An abundance of studies suggests that connectivity networks can provide informative signatures of mental states and cognitive processes.

1.3 Depth perception in object recognition

The majority of animals, including humans, possess horizontally separated eyes, causing them to acquire slightly different images between their eyes, which is referred to as binocular disparity (Barlow

et al 1967, Parker 2007, Tittle & Perotti 1997). As a result of binocular disparity, the visual system develops a computational model to calculate depth information from the differences of images acquired by the two eyes. Binocular disparity defined depth information is processed in absolute disparity and relative disparity (Backus et al 2001, Umeda et al 2007). The former describes the difference in angle subtended on the left and right retina of an object in space and gives an estimate of the depth of that object to the observer. Relative disparity is the comparative depth between two objects in space and arises when there are two or more depth planes present differences between the two eyes are calculated. Disparity tuned cells being located along the visual hierarchy starting from V1 appear to support computation of absolute disparity (Barlow et al., 1967; Pettigrew et al., 1968; Cottereau et al 2012). Numerous studies have addressed this issue, and the parietal cortex and dorsal visual pathway, along with the middle temporal area (MT) and area V3A, have been identified to be involved in processing depth from the relative disparity in various stimuli (Parker 2007; Tsao et al., 2003b).

In addition to binocular disparity, various monocular perceptual cues of depth are also present in our day-to-day life. Perspective cues, shading cues, texture cues, and motion cues all occur in the daily lives of humans. The human brain can decipher depth information from shading conditions. Objects and scenes exposed to a light environment generate a 2-D image in the retina that contains certain luminance contrast patterns predicated on the basic topology of light in that space. The human brain actively solves the inverse problem through the depth of objects and scenes based on the luminance contrast pattern of shadow and light, texture gradients, binocular disparity, and other depth cues. The countering of the inverse problem using shadow and light is believed to occur in the ventral visual pathway (Li & Pizlo 2011).

Renaissance painters are examples of humans that utilized light and shade to convey depth information (Brooks 2017). Incorporating the size and shape of the image captured by the eyes, the human brain also deciphers depth from perceptual and texture cues. The retinal image captured by the eyes appears to be larger when the surface is closer, and smaller otherwise. If the surface of an object possesses a uniform texture, such as noise dots patterns, the human visual system can infer the surface tilting angles via the distortions of the texture elements. Subsequently, the texture gradient information is utilized to infer a 3-D structure in the space (Todd & Thaler 2010). Furthermore, the human brain infers the 3-D spatial depth information by motion cues. Moving objects that are closer to the observer appear to be moving faster in the retina images—if an object rotates in depth, the surfaces that are closer to the observer also appear to be moving faster (Wallach & O'Connell 1953). This differential speed of different surface segments allows the generation of a depth cue that humans can use to recognize objects (Farivar et al., 2009). Despite evidence suggesting that ventral regions are involved in motion defined depth (Li et al., 2013), the most common theory is that these motion cues are processed in the dorsal stream of the visual system (Bisley & Pasternak 2000).

In the natural world, object and scenery typically include multiple depth cues inseparably, meaning that the visual system must combine information from multiple depth cues in order to assess the scene. Research has shown that humans can recognize complex 3-D objects regardless of depth cues (Dehmoobadsharifabadi & Farivar 2016; Farivar 2009). Predicated on the traditional belief that different depth cues are processed in different visual pathways, it has been hypothesized that there is a neural population that integrates dorsal and ventral visual pathway depth information. However, the method by which the visual system combines information from multiple visual pathways to achieve an integral perception of depth is still not fully understood (Dovencioglu et al 2013, Georgieva et al 2009). The aforementioned hypothesis significantly demonstrates how the different depth cues are processed

quasi-independently, and then integrated into a single neuronal population containing all the information for 3-D structures. This hypothesis assumes cortical representations of objects to be invariant of depth cues.

Solving such problems can be accomplished through multivariate analysis. Multivariate analysis can be utilized to mitigate the anatomical and functional mismatches across multiple subjects; the subject-specific multivariate patterns can be used to estimate the effects of stereopsis and can account for the individuality of subjects (Chen et al 2020a). Furthermore, multivariate analysis employs patterns and network analysis that is able to capture the information from diffuse and distributed signals across the visual system, which matches the characteristics of depth perception processes (Henderson et al 2019). Depth perception is facilitated by the integration of multiple depth cues, meaning that the process may require coordination of multiple cortical visual areas. Hence, a multivariate method allows enhanced understanding (Kriegeskorte & Diedrichsen 2019).

1.4 Naturalistic stimuli: A way of measuring cortical response with real-world experience

When experiencing the world, the human brain encounters dynamic, crowded, and interactive stimuli. Naturalistic stimuli, such as movies, spoken stories, or music, have become exceedingly popular in cognitive neuroscience in order to help resolve the uncertainty of whether the brain functions in complex real-life scenarios as it does in simple tasks and lab-created abstract stimuli (Finn et al 2020, Vanderwal et al 2017, Vanderwal et al 2015).

1.4.1 Traditional stimuli lack dynamics

Abstract stimuli adhere to relatively simple parameters in the design and deliver strictly controlled inputs to the brain. The traditional designs of abstract stimuli tightly control the variables involved in the

different conditions, focusing the fundamental insight on the targeted brain locations in order to understand the brain's processes. However, as the lack of complexity of experiments typically fail to resemble real-world scenarios, debate often arises regarding the biological plausibility of lab-created stimuli. Abstract images of faces therefore lack dynamic interactive components, which involves the functions of various integrative multimodal brain regions, such as the superior temporal sulcus (Calvert et al., 2001). Therefore, the use of naturalistic stimuli that incorporates dynamic and rich stimuli like movies, narrative stories, music, etc., to mimic real-life experiences has been increasingly employed (Nishimoto et al 2011, Sonkusare et al 2019).

1.4.2 The emergence of naturalistic stimuli in visual neuroscience

Although naturalistic stimuli also operate in a laboratory setting, they are providing superseding approximations of daily experiences to the participants, as opposed to traditional synthetic stimuli (Hasson et al., 2010; Sonkusare et al., 2019). The human brain is ecologically tuned to naturalistic stimuli. To illustrate, the naturalistic motion of face stimuli has been shown to enhance the cortical response to static facial images (Schultz & Pilz, 2009). Naturalistic stimuli require active engagement from the participants in the experiments. For instance, movie viewing requires attention, comprehension, and integration of information(Hasson et al 2008). Additionally, emotion, social interactions, language, multimodal perception, or some inferences of latent intentions are also involved in a multitude of comprehensive experiments. The complex tasks naturally involve high levels of engagement of the neural system, although it can be argued that the discrepancies in the level of engagement may be more significant if the participants possess diverse cultural and personal backgrounds (Tononi et al., 1996).

1.4.3 Methodological developments in analysis of naturalistic stimulation studies

Traditional fMRI analysis techniques, including GLM-based parametric block or trial designs, are uncommon when analyzing naturalistic stimuli. GLM requires the cortical BOLD response to the experimental conditions to be defined from trigger events, which are then used to construct the regressors in the GLM equations. In movie-viewing or other passive paradigms, GLM is hindered by the difficulties of constructing regressors from the day-to-day activities. Furthermore, the robustness of the temporal events reconstructed from the natural stimuli features are low (Ben-Yakov et al., 2012). Within-subject cortical responses to the same features in experiments is inconsistent and often various for naturalistic stimuli. The lack of control over stimulus features hampers the ability of traditional GLM based analysis of naturalistic stimuli (Bartels & Zeki 2004, Hasson et al 2004).

In response to traditional approaches like GLM failing to address the challenges of naturalistic paradigms, novel data-driven neuroimaging data analytic methods have begun to emerge. A particularly successful method is the inter-subject correlation (ISC) approach established by Hasson et al. (Hasson et al., 2008a; Hasson et al., 2004). ISC is a between-subjects measurement that does not require any prior model of the stimuli, meaning it is highly suitable for naturalistic stimuli. In regards to the experimental paradigms, ISC utilizes the mean pairwise voxel Pearson's correlation coefficient for temporal cortical responses, thus ISC is often believed to possess reliability across all subjects. Reliable cortical responses across subjects have been replicated in many studies that employ either fMRI or MEG in movie viewing (Chang et al., 2015; Hasson et al., 2010; Lankinen et al., 2014). Measuring the reliability of cortical responses in naturalistic experiment settings with variable time windows could help reveal the cortical hierarchical structure. While brain areas higher in the cortical hierarchy integrate over a longer temporal window, primary sensory regions can capture rapidly changing stimulus features in a shorter temporal receptor field (Murray et al., 2014).

ISC is a univariate fMRI analysis method that relies on the assumption of a linear response model. The linear response model assumes that anatomical cortical locations that have been defined by the cross-subject alignment of the grey matter also share a linear response model with the stimuli, at least in some regions. This assumption is true in some visual areas that rank lower in the cortical visual hierarchy (Hasson et al 2010, Hasson et al 2004). Particularly, some brain areas, including multimodal areas, can exhibit high variability across subjects but robust functional responses (Kauppi et al., 2017). The method of analyzing neuroimaging data from naturalistic stimuli can be expanded by multivariate analyses that employ multivoxel pattern responses. Pattern classification models can also be established to decode information relating to the naturalistic stimuli labels (Betti et al 2013, Isik et al 2017, Mandelkow et al 2017, Nishida & Nishimoto 2018). In this latter case, decoding performance represents the information contained in the cortical patterns.

1.4.4 Challenges of using naturalistic stimuli

In the past two decades, the use of naturalistic stimuli in cognitive neuroscience has increased substantially (Hasson et al 2008, Vanderwal et al 2017). Although natural paradigms have expanded views surrounding traditional task-based classical paradigms, there are still numerous issues. One limitation is that the naturalistic experimental settings are only approximations of everyday life activities. Despite being far more realistic and engaging compared to traditional task-based procedures, natural paradigms are still lab-created and lab-operating and are subject to directions originating from the experimental procedure. Researchers have shown that movies with strong narrative structures are linked to higher ISC rather than random uninteresting movies (Hasson et al., 2008b), which is likely due to the well-controlled directions of attention and gaze caused by the visual content. Eye tracking has also been used in various ISC experiments to ensure the level of commonality between subjects (Lu et al., 2016; Mandelkow et al., 2017), but free-viewing risks increasing the inter-subject variability. Given

that naturalistic stimuli are also associated with comparable levels of engagement and attention to different components across subjects, naturalistic viewing combined with some control over viewing (i.e., fixation) will serve to combine the best of both worlds—increased comparability across participants through the experimental control of fixation, and rich, naturalistic stimulation through real (as opposed to synthetic) stimuli. In the following chapters (Chapters 2 & 3), a fixation-viewing method was employed to ensure the comparability and reliability of visual stimulation across subjects, allowing us to focus on the visual impact of movie viewing.

1.5 Objectives and rationales

The main objectives of the present thesis are (1) to establish novel multivariate data analysis techniques suitable for understanding the representation of objects from depth cues and (2) to investigate the link between depth perception and object recognition using multi-subject datasets. The rationale and the objectives of the three following chapters that were employed in the present thesis are discussed in detail in the following sections.

In Chapter 2, the primary focus is gamma-activity-related representational similarity and its modulation by stereopsis during naturalistic movie viewing. The rationale of the study is that the gamma oscillatory band activity patterns were previously not found to be similar across subjects in a naturalistic setting, despite the fact that gamma band activity plays an important role in depth perception and object recognition in the visual system. A new inter-subject representational similarity method is needed here to bridge the gap between the widely known role of gamma activity in the visual system and the lack of knowledge of gamma oscillatory bands at the level of subject prototypicality. To establish this inter-subject representational similarity, we collected MEG datasets of 24 subjects watching naturalistic video clips from a documentary movie *Under the Sea*. To investigate which oscillatory frequency bands account for the depth-related object representations in the brain, the traditional ISC

method was compared with the novel inter-subject representational similarity method. By utilizing both ISC and the inter-subject representational similarity method, the commonality of subjects during the movie viewing was investigated in 6 different oscillatory frequency bands. The similarity of movie scene representations provides a quantitative measurement to the prototypicality of objects and scenes among the subjects and oscillatory frequency bands. Furthermore, representations of the movie scenes were developed from the depth information of the movie and the frequency bands related to the depth information of the movie were measured.

In Chapter 3, the main focus is on cortical network changes that occur during 3-D movie viewing. Binocular disparity can be related to visual cortical networks (Gaebler et al., 2014), but the specific cortical network structures have not been identified. Dorsal visual networks, which have been shown to be involved in relative disparity processing in static stimulus settings (Cottereau et al., 2014; Roe et al., 2007; Roy et al., 1992), have not been linked to connectivity network changes in viewing conditions (i.e., dichoptic 3-D vs. monoptic 2-D). To understand how dorsal visual network connectivity behaves in naturalistic 3-D movie viewings, the network engagement must be measured through multi-subject datasets using an inter-subject method to reduce the subject-specific activity profiles. In Chapter 3, we use a novel approach, known as inter-subject functional connectivity, which eliminates the subject-specific temporal profiles in time-series fMRI datasets. We compared results from this method to those obtained with traditional FC method for 3-D vs. 2-D viewing. A decoding framework from MVPA was adopted, but rather than utilizing voxel patterns to decode the stimulus conditions, connectivity networks and its sub-networks were used to better illustrate the relatedness of visual networks to the movie-viewing condition. Subsequently, the decoding performance and a standardized distance between 3-D and 2-D conditions of the sub-networks were measured. It was concluded that the connection within the dorsal visual pathway plays an important role in 3-D movie viewing and that disparity induces different visual networks in naturalistic viewing.

Chapter 4 attempts to understand the mechanisms of cortical integration of monocular depth cues in 3-D object recognition. Although the integration of the depth cues has been demonstrated in object recognition (known as depth-cue-invariant 3-D object recognition) (Akhavein et al 2018), the exact timing of different depth cue cortical processing is unknown. The timing of depth-cue-specific and depth-cue-invariant representations is essential to infer the role of depth cue integration in 3-D object recognition, since depth-cue-specific 3-D object recognition could infer a separate and auxiliary object recognition process, which may be complementary to the current theories of object recognition which are focused on a hierarchy of 2-D representation. MEG data were used in Chapter 4 because of the high temporal resolution and potential for spatial inference. In this study, the subjects viewed three sets of objects separately defined by unique depth cues—shading, texture, and 3-D structure-from-motion—while correlates of brain activity were recorded via MEG. Depth cue invariant object recognition was investigated using the multivariate decoding generalization method—by utilizing training data from one time point and testing data from other time points, the temporal dynamics of depth-cue-specific representation were revealed. Training the decoding models using the data from one depth cue and testing the model using the data from another depth cue revealed the dynamics of depth-cue-invariant object representation. The development of said tools provided a means to better understand how depth cues are processed and integrated in 3-D object recognition.

Chapter 2: Natural scene representations in the gamma band are prototypical across subjects

Previous studies have suggested that the depth information can be in object recognition. However, it is not clear how depth processing can be measured in the cortex in a naturalistic movie viewing condition, which mimics the natural world. It is crucial to establish the prototypicality among the subjects under these natural conditions. In this paper, I used 3D movies as the stimuli to provide a visually striking presentation of the underwater sea creature scenes. I used MEG to measure the representational prototypicality of the natural scenes in different frequency bands. Counterintuitive to the previous findings that the lower frequency bands responses are similar across the subjects, I demonstrated for the first time that gamma bands are similar across the subjects in the representation of the movie scenes. Further, my evidence also showed that this representational similarity in the gamma oscillatory bands is closely related to the depth contrast from the movie scenes but not luminance contrast. These findings showed that movie scene representations are closely related to the depth information in the scenes.

Title: Natural scene representations in the gamma band are prototypical across subjects

Abbreviated title: Gamma band representations are prototypical

Author names and affiliation: Yiran Chen[1,2], Reza Farivar[1,2]
[1]: McGill Vision Research, McGill University and [2]: Research Institute of the McGill University Health Centre, Montreal, QC, Canada

Corresponding author and coordinates:
Reza Farivar, PhD
Montreal General Hospital, Room L7-213,
1650 Cedar Ave,
Montreal, QC, Canada, H3G 1A4
reza.farivar@mcgill.ca

Conflict of Interest: The authors declare no competing financial interests.

Acknowledgements: The authors are grateful to Yassin Nazzar, Sylvain Baillet, Christopher Pack, Angela Zhang, Sébastien Proulx, and Tatiana Ruiz for comments on the analysis and manuscript. The study was funded by an NSERC Discovery grant to RF and start-up funds from the Research Institute of the McGill University Health Centre.

Abstract

Prototypical brain responses describe similarity in neural representations between subjects in response to a natural stimulus. During natural movie viewing, for example, inter-subject correlation (ISC) measured by fMRI is high in visual areas (Hasson et al 2004). But the electrophysiological basis for this fMRI ISC has been controversial. Previous reports have only found ISC in low frequency bands—below 12 Hz (Chang et al 2015). These findings stand in contrast to reports that gamma band oscillations—30 to 90Hz—are highly stimulus-driven in visual cortex (Perry et al 2015). To resolve this discrepancy, we carried out both ISC estimation and a novel inter-subject representational correlation analysis across six frequency bands extracted from MEG data of 24 subjects who each viewed four 5-minute clips of an underwater documentary. Region-of-interest-based and vertex-based temporal ISC estimates confirmed that low-frequency bands are significantly synchronized in visual areas and that gamma band has low temporal correlation. We also found the representational geometry of movie scenes were related to structural statistics from the stimuli. Crucially, our results show that the gamma band oscillations also reflect prototypical brain response in scene representations formed in response to naturalistic stimuli as revealed by inter-subject representational correlation.

Introduction

Shared sensory stimuli result in similar brain responses across different individuals. Inter-subject correlation (ISC) describes a level of commonality that can be captured in the synchronized fMRI activity of a group of subjects (Hasson et al 2004), even under conditions of naturalistic viewing—subjects exhibit temporally correlated hemodynamic brain activity in corresponding brain areas when watching the same movie (Hasson et al 2010, Haufe et al 2018). Shared psychological perspectives, including engagement and attention have been shown to related to ISC in both fMRI and EEG literatures (Dmochowski et al 2012, Lahnakoski et al 2014, Poulsen et al 2017). Shared components can also be linearly modelled in relation to the stimulus features including color, contrast and stereopsis (Bartels &

Zeki 2004, Gaebler et al 2014, Hasson et al 2008, Malinen & Hari 2011), and these can be used to create models for "brain reading" (Naselaris et al 2009).

In response to stimuli, cortical regions exhibit electrophysiological activity in different frequency bands, including gamma bands (30 – 100 Hz). Gamma bands have been shown to be directly tuned to various features of synthetic stimuli, including orientation, size, luminance and spatial frequencies (Hadjipapas et al 2007, Hermes et al 2015a, Muller et al 1996, Perry et al 2015, Porcaro et al 2011). Even higher-level processing of objects and natural scenes have been shown to be related to gamma band power (Brunet et al 2015, Martinovic et al 2008). Interestingly, gamma bands power changes in MEG recordings can be used to predict visual features of synthetic stimuli (Duncan et al 2010). Gamma band power is elicited by natural images as well, but with lower magnitude compared to synthetic stimuli such as Gabor gratings (Kayser et al 2003, Mukamel et al 2005, Muthukumaraswamy et al 2010). The link between stimulus features and gamma band power suggests that gamma band power may be directly related to BOLD activity measured in fMRI (Mukamel et al 2005).

Despite its prevalence and its implication for perception, no study has found evidence of ISC in gamma band power in MEG movie viewing datasets, which would suggest that the gamma band power is perhaps person-specific (i.e., not prototypical). Using MEG, Lankinen et al. (2014) found significant ISC in MEG canonical components below 5 Hz when subjects watched a 15-minute silent film "*At Land*". Bridwell et al. (2015) also found significant ISC in the low-frequency 0-4 Hz range when subjects watched identical movie clips using EEG. A recent movie study using EEG and ECoG found some similarity amongst a few ECoG subjects in the gamma range when subjects were watching a 325-s commercial audio movie (Haufe et al 2018), but a more general demonstration of a link between this putatively stimulus-driven band of oscillatory activity and naturalistic stimulus features has not yet been made.

ISC is a concept that uses Pearson's correlation on the temporal signal (Hasson 2004, Hasson 2008a, Chang 2015). Using ISC or other techniques that operate on the temporal profile (i.e. canonical

correlation), one can derive estimates of temporal similarity in sensor or cortical source responses across multiple individuals. An alternative multivariate approach, such as representational similarity analysis (RSA; Kriegeskorte, 2008) can be used to derive similarity metric that are not necessarily temporal and that are not captured by ISC. Such multivariate approaches are powerful and can be used to hyper-align the response of different subjects and improve ISC (Gola et al 2013, Guntupalli et al 2016, Haxby et al 2014). RSA has been applied to MEG/EEG data in studies of object recognition and visual pattern recognition (Cichy et al 2016a, Cichy et al 2014, Cichy et al 2016b, Kaneshiro et al 2015, Wardle et al 2016). Thus RSA may capture shared prototypical components that ISC may fail to detect.

We therefore sought to compare ISC and inter-subject representational correlation—a novel method that we used to combine ISC and representational similarity analysis—in capturing group commonalities across oscillatory bands to assess inter-subject similarity during naturalistic stimulation in all oscillatory bands, including high-frequency. In addition, we measured the relatedness between representational similarity from movie image statistics and representational similarity from gamma band activity. Using this novel approach with rich naturalistic stimuli allowed us to capture robust representational similarities between subjects in the high-frequency gamma bands.

Materials and Methods

Participants

Twenty-four healthy young participants (14 female, 10 male, average age 27.04 years, standard deviation of age 6.80 years, ranging from 18 to 44 years of age) with normal vision were recruited for the movie experiments. Written consent from participants was obtained for permission of data analysis and publication. All procedures were approved by the Research Ethics Board of the Montreal Neurological Institute.

Stimuli

Two clips of the IMAX movie *Under the Sea* (Hall 2009) were shown, each running for a duration of 5 minutes. Subjects viewed the same clips both monoscopically (Clips 1 and 3) and stereoscopically (Clips 2 and 4), for a total of four clips. The four clips were not repeated. The IMAX movie was copied in HD format to the hard drive using *DVD Decrypter for Blu-Ray*. The video file included two separate streams, one for each eye, stored as AVI files. The AVI files were imported into Matlab where video content was selected for the two final clips. To facilitate stereoscopic fusion for that viewing condition, the scenes included a frame consisting of striped black-and-white lines. Each movie clip consisted of a sequence of different scenes where each scene showed different undersea settings and creatures. A white fixation cross was shown in the center of the screen and participants were asked to fix their gaze on the center of the screen during movie viewing. This constrained viewing condition maximized the similarity of visual input to subjects and reduced the attentional modulation, (Ki et al 2016). Finally, a flickering pattern was encoded at the bottom-right side of the screen to allow the playback timing to be picked up with a photodiode. The resulting videos were re-encoded with an Xvid HD codec into separate AVI files.

Display and playback

We used *Stereoscopic Player* (3dtv.at, Linz, Austria) to present the videos. Videos were played from a desktop PC with an Intel i7 processor, 8G of RAM and a Nvidia GeForce card. The PC was attached to an LG 3-D LED monitor (D2342PY, height: 13.5”) at a viewing distance of 150 cm away from the subject. The refresh rate of the monitor was 60 Hz. The size of the image was ±5.7° vertical and ±9.5° horizontal the frame rate of the movie was 24 fps. The 3-D monitor used line-interlacing of separable circular-polarized signals to encode a dichoptic stimulus (for Clips 2 and 4). The stimulus was viewed by the subjects using a set of circular-polarized glasses. The four clips were presented to the subjects in a pre-

determined random order, which also differed from subject to subject. The recording room was completely dark aside from the stimulus screen. An infrared camera was present to monitor the room.

MEG acquisition

MEG signal was acquired using a 275-channel CTF/VSM instrument with superconducting quantum interference device (SQUID)-based axial gradiometers. Magnetic brain activities were recorded at a sampling rate of 2400 Hz in addition to electrophysiological signals: two electrodes were placed on the chest of the subject to measure ECG (cardiac activity); two electrodes were placed adjacent to the left eye, one above and one below, to record EOG (blink activity); and two electrodes were placed on both sides of the eyes to record EOG (saccade activity). A ground electrode was placed on the left shoulder. Head position tracking coils were placed on the subject and their locations were recorded using a Fastrak Polhemus device. Additional scalp points on the subjects' head were recorded to capture the subjects' head-shape for co-registration with the T1 anatomical scan. A photodiode placed on the bottom-right of the screen captured precise video frame timing of the videos on an additional analog channel along with the MEG.

Pre-processing and brain source activity estimation

FreeSurfer was used to reconstruct the 3D surface from the anatomical MRI. *Brainstorm 3.1* (Tadel et al 2011a) was used to preprocess the MEG signal and compute the inverse projection. The reconstructed surfaces were registered using the images of the head position tracking coils during head digitization. The recordings were band-pass filtered at 0.5-300 Hz and a notch filter was applied to remove 60 Hz harmonics power line contamination. Problematic sensors were identified according to the power spectrum density (Welch filter) of the recordings and were eliminated from later analysis. We also did an empty room recording with the LCD monitor in the recording room, resembling the experiment condition. We found 24 Hz and harmonics power peaks, in the control data, likely due to the framerate of the movie. We therefore incorporated notch filters at 24 Hz and harmonics to all our data.

The power spectrum of the subject-absent recordings and the subject-present recordings after applying the notch filter could be seen in supplementary figure 1. Data were then down-sampled to 300 Hz to reduce the computational cost of subsequent analysis. The sampling frequency is chosen because there is no evidence that cortical activity is related to signals over 150 Hz, the 300Hz is the Nyquist frequency of this sampling rate.

Signal-Space Projection (SSP) was used to remove the artifacts generated by eye-blinks and heartbeats based on EOG and ECG activity recorded during the experiment. Forty signal components were derived using independent component analysis (ICA) method in Brainstorm (Tadel et al 2011a). We also carried out ICA on the subject-absent experiment and identified one component with the topography matching the screen artifact that we had anticipated. The artifact topography was plotted in the supplementary figure 2. This artifact component was then removed from the signal.

Unconstrained brain source activities at 15,002 vertices were estimated according to the minimum norm method (Hamalainen & Ilmoniemi 1994). Empty room recordings were used for estimation of the noise covariance matrix. The source data were projected onto the Colin27 anatomy template (Holmes et al 1998) based on a nearest neighbor approach.

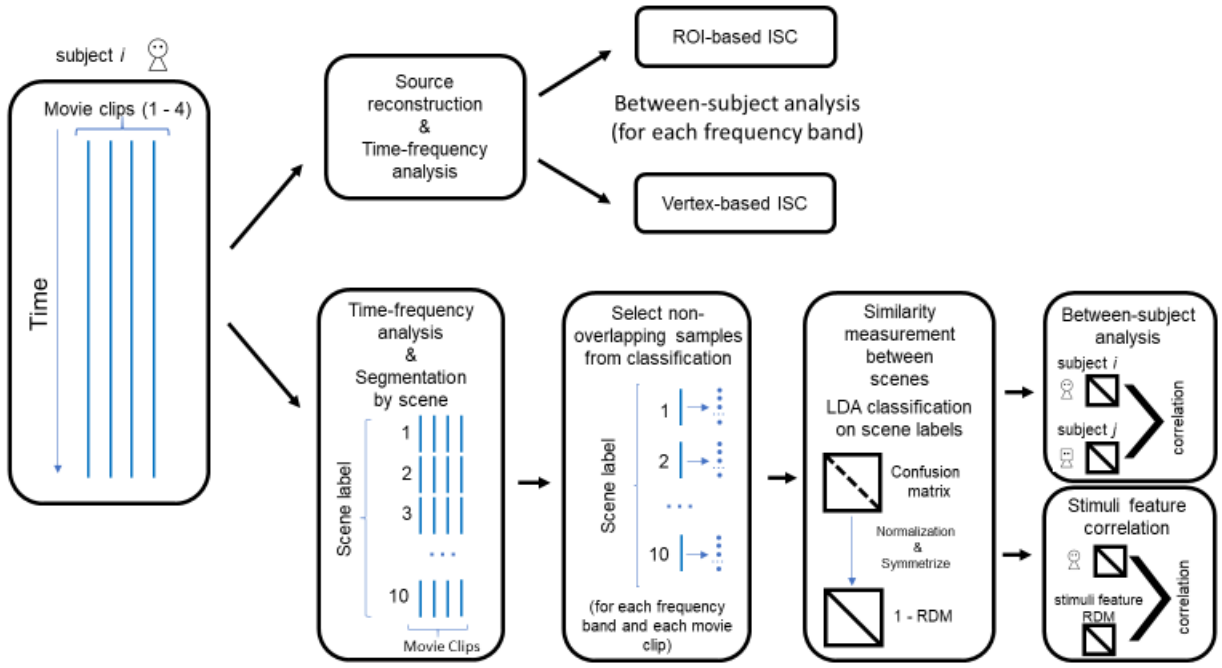


Figure 1. Our analysis consisted of the following two pipelines: In the first branch, we carried out MEG source data reconstruction and time-frequency analysis to generate time-dependent, frequency-specific power changes. We then performed ROI-based ISC and vertex based ISC in 6 frequency bands. In the second branch, we carried out the same time-frequency analysis on the MEG sensor data, and then segmented the time-series into 10 different scenes. We then further broke up the time-frequency data into 10 samples per scene and carried out LDA classification for scene labels. Confusion matrices were generated and transformed into representational dissimilarity matrices (RDM). Inter-subject representational correlation was calculated to measure consistency of movie scene geometry in the subject group. Also, Pearson correlations were performed between subject's RDM and the dissimilarity matrix generated from the movie image statistics.

Overview of the analysis pipeline

Figure 1 describes the analysis pipeline. Our data consisted of MEG responses to four 5-minutes movie clips for each subject. We carried out two analyses: For the first analysis (temporal ISC), we performed source reconstruction to project the data onto the vertices of the mesh describing the

cortical surface and carried out time-frequency analysis to decompose the time-series into 6 different frequency bands. We then calculated temporal ISC for these time-dependent frequency power changes both on various visual ROI and on the whole cortical surface. For the second analysis, we investigated representational similarity for movie scenes between subjects. We performed time-frequency analysis on the MEG sensor data. We then segmented the time-series data into matrices that could be used to train a multi-class classifier to discriminate between different movie scenes. Linear discriminant analysis (LDA) classifiers were trained to predict the scene labels from the test dataset and constructed scene similarity matrices from the LDA classification results. The subject-specific representational geometry was used in two different ways: First, we compared the scene RDM derived from the MEG results to the RDM derived using image statistics of the stimulus (i.e., movie structural statistics). Second, we explored inter-subject representational similarity by calculating the correlation of RDM for each subject pair.

Hilbert transform and band-pass filter

Bandpass filtering responses followed by Hilbert transformation provide an aggregate measure of time-dependent power in a specified frequency band. This approach boosts signal by pooling across frequencies within a band and provides an aggregate, time-dependent frequency response profile. To analyze the activity in both temporal domain and frequency domain, Hilbert transformation and bandpass filtering were performed in six different oscillatory frequency bands (Muller et al 1996). Because individual response peak differences exist for each frequency band, we performed a Hilbert transform on each frequency band, implemented in MEG/EEG data processing software, *Brainstrom (Tadel et al 2011a)*. We chose the six different bands as delta (2-4 Hz), theta (5-7 Hz), alpha (8-12 Hz), beta (15-29 Hz), gamma1 (30-59 Hz), and gamma2 (61-90 Hz) (Baillet 2017, Cohen 2014). The frequency boundaries were non-overlapping integers.

ISC for different frequency band estimation on vertices and ROIs

To compute ISC for each frequency band at every vertex, pairwise Pearson correlation for all subject pairs was computed for each frequency band and for each movie clip. Fisher z-transformation of the correlation coefficients was performed before we averaged the ISC value for 4 movie clips. These ISC values represent similarities between subjects on time-dependent power changes for each of the frequency bands. To determine the level of significance, a non-parametric test was used (described below) to transform the mean pairwise correlation value to a z-score. Bonferroni corrected critical z-score was used as a reference.

ROI-based analysis allows us to make inferences about specific visual areas while increasing the signal-to-noise ratio. In our study, we calculated ISCs in 25 different visual ROIs, which were described in a previous study (Wang et al 2015). To capture common features for each ROI, we used single value decomposition to extract the first spatial components using single value decomposition in each ROI instead of average all the source location in each ROI, as this has been demonstrated to be superior for remaining sensitive to small sources in large ROIs (Chang et al 2015). The first spatial component always retained over 40% of the total explained variance.

Linear regression analysis was employed to investigate the relationship between cortical hierarchy (as determined by the response delay) and ISC for a given ROI. The visual stimulus processing delay data were obtained from Experiment 2 of one MEG-fMRI fusion study using the same probabilistic atlas (Cichy et al 2016b, Wang et al 2015). To better estimate the change in ISC along the hierarchy of visual areas in each frequency bands, we used linear regression to estimate the relation between mean ISC scores and previously-reported cortical latencies (Cichy et al 2016b). Coefficients of determination and significance values were reported for the linear regression results in each frequency band.

MEG sensor data classification on movie scenes

We tested whether movie scene labels can be predicted from the sensor pattern in each frequency band. We first divided the MEG response in each frequency band to identical scene segments. Ten distinct scenes—wherein each contained a distinct undersea creature or group of creatures within a certain underwater environment—were identified and labeled, each of which lasted at least 6.5 seconds, with the first 500ms of each scene excluded from analysis to prevent any lingering brain activity from the previous one. Within each scene segment of the data, ten non-overlapping 500ms samples were drawn with random separation time between each sample of at least 50ms. With this criterion, only 10 samples could be selected from the shortest scene. This separation time ensured independence between each sample. The samples were averaged through time to improve signal-to-noise ratio. The size for all the data in each movie clip was Nsubject × Nsensor × Nband × Nscene ×Nsamples, where Nsubject was the number of subjects, which was 24; Nsensor is the number of sensors we used for multivariate-classifier; Nband was the number of frequency bands, which was 6; Nscene was 10, as we chose 10 distinct scene in each movie clip, Nsample was 10, as we extracted 10 different samples from each scene.

Movie scene label number classification was performed on each subject using LDA—a computationally inexpensive classification method when compared to alternatives such as Support Vector Machine (SVM); LDA also showed better performance when classifying neuroimaging data for natural scenes (Kaneshiro et al 2015). Unbiased estimates were acquired by leave-one-out cross-validation. A confusion matrix (CM) was generated whereby each element—in coordinates (*i*, *j*)—represented the proportion of samples where the LDA model classified a sample with label number *i* as label number *j*. The diagonal of the CM reflects the correct predictions and were used to compute the classification accuracy. We calculated the average classification accuracy for each movie clip, each subject and each band. We averaged the result for 4 movie clips for this analysis. We also carried out the

same analysis for the two subject-absent control recordings, where all the conditions are the same, except that there were no subjects—this served as a negative control.

We carried out two-sample t-tests (two-tailed) for each frequency band between the subject present condition and the subject absent condition. The degrees of freedom was Nsubject + Ncontrol -2 = 24, where Nsubject was the number of subjects, and Ncontrol was the number of control session we had.

We also estimated the contribution of each sensor to the classification performance to map the spatial pattern of the derived representations. We performed the classification analysis on each of the MEG sensors, averaged by subjects and movie clips, and then plot the classification accuracy on a MEG sensor topography, for a coarse visualization of sensor location of classification accuracies distribution. This process was repeated for each frequency band.

Inter-subject representational correlation

To assess the neural representation of scenes, we estimated the representational similarity between different scenes. The distance between different scenes can be shown in a symmetric matrix with zeros in the diagonal, termed representational dissimilarity matrix (RDM). To obtain the RDM, we used the method of (Kaneshiro et al 2015), whereby we first divided each entry of the CM by the diagonal value of the respective row, to generate normalized CM, $nCM_{ij} = CM_{ij}/CM_{ii}$. Then, we calculated the geometric mean of the matrix and its transpose matrix.

$$S = \sqrt{nCM \times nCM^{T}}$$

Here S denotes the similarity matrix. We symmetrized nCM because we considered the distance from representation of scene *i* to scene *j* was equal to the distance from the representation of scene *j* to scene *i*. The RDM was equal to the lower triangle elements from matrix (1- S). To visualize the grouping relations in a given example RDM, we used the unweighted pair grouping method with averaging (UPGMA) to generate a group-averaged similarity matrix, showing the hierarchical structure of the

representation. An example dendrogram describing scene representational structure in video clip one and hierarchy was constructed from the results of UPGMA.

Inter-subject representational correlation was explored by determining between-subject RDM correlation. It was defined as the mean pair-wise Pearson correlation of the subjects' RDM. We calculated inter-subject representational correlation for each movie clip and each frequency band. Significance levels were determined by one-sample t-tests to test the hypothesis that the mean Fisher-transformed inter-subject representational correlation values were significantly different from 0. The degrees of freedom for this t-test is Nclip -1 = 3, where Nclip was the number of different movie clips. We did this analysis on the data for each frequency.

Scene structural feature correlation

To estimate the structural similarity between scenes, we evenly divided the images of the video from all frames into 25 rectangle sections. For each section, we calculated several frame-wise image metrics, including the mean luminance, Root mean square (RMS) contrast, mean depth, and RMS depth contrast for each section (four metrics in total). We then calculated scene similarity as the Pearson correlation of the scene statistic within the 25 sections of one scene to the same section of another scene in the same clip, repeating this for every pair of scenes in the same clip. Pearson correlation of these feature similarity vector with the subject's similarity matrix generated from MEG sensor classification data was carried out. After combining the correlation value for 4 movie clips, significance levels were determined based on one-sample t-test. The degrees of freedom of this test are Nsubject – 1 = 23.

All stimulus features (scene image statistics) were estimated from the pixels of the movie frames. For mean luminance, we simply took the mean pixel intensity in each section, while for RMS contrast, we converted the RGB frame to grayscale using the National Television System Committee (NTSC) weights, and then calculated the RMS contrast for each section. A depth map was estimated from the

stereo pairs of the clip at each frame using local stereo matching with a guided filter (Asmaa Hosni, 2011). We took the mean of the disparity map in each section as the mean depth, and the RMS of the disparity map as RMS depth (i.e., depth contrast). We subsequently calculated the structural similarity matrices using these parameters.

Non-parametric tests for statistical inference

To robustly estimate statistical significance of the derived ISC, we used non-parametric permutation methods using the Amplitude-Adjusted Fourier Transform (AAFT) to estimate the null-distribution of the mean pairwise correlation by permutations from the phase-scramble time series (Chang et al 2015, Chen et al 2016). The estimated ISC were then z-scored with respect to the permutated null distributions.

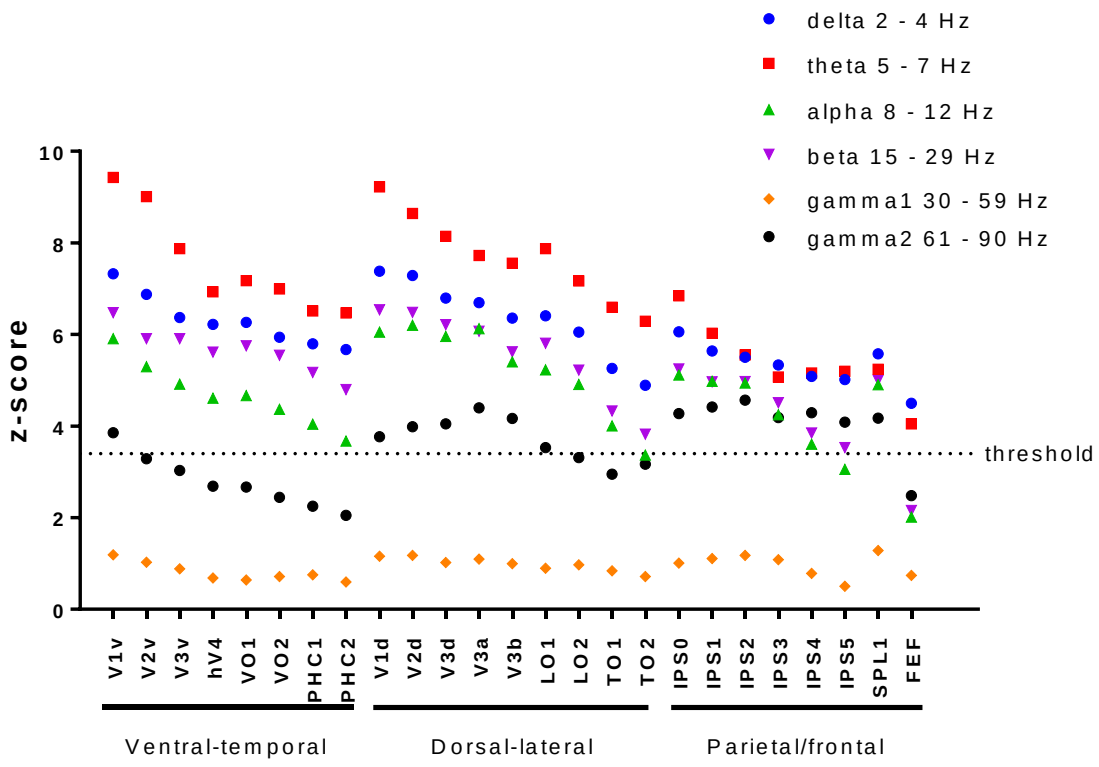


Figure 2. ISC of six different frequency bands were estimated from temporal MEG source data in 25 different ROIs obtained from a probabilistic atlas (Wang et al 2015). Each point depicts the average Z-score from responses to four movie clips. Z-scores were estimated from permuted distributions using temporal phase-scrambling to generate random time course signals (see Material and Methods for stimuli and non-parametric tests for statistical inference). Significance level of these Z-scores were calulated based on Z-test for standard normal distribution. Bonferroni-corrected threshold (z = 3.40, p = 0.00033) is depicted by the dotted line.

Results

ISC in different visual areas

With 25 visual areas across six bands, we set the critical *p* value for each test to 0.0003 (critical z-score of 3.4). We observed significant ISC for naturalistic movie viewing in several frequency bands and probabilistic ROIs. Consistent with previous research (Chang et al 2015), theta band demonstrated the highest ISC in most visual areas (5-7Hz; $p < 0.0003$ for all visual areas; Figure 2). The same was observed for the delta band (2-4Hz; $p < 0.0003$ for all visual areas). Alpha (8-12Hz; $p< 0.0003$ for all but TO2, IPS5 and FEF), and beta bands (15-29Hz; $p < 0.0003$ for all but FEF) also demonstrated ISCs greater than chance in visual ROIs. Gamma bands, however, were generally less significant, with the low-gamma band (30-59Hz) failing to show significant ISC in any area whereas the high-gamma band (60-90Hz) showed low ISCs in visual areas. Combining the ROIs, the overall averaged mean pair-wise correlation for the different frequency bands were 0.0137 ± 0.0021 for delta band, 0.0148 ± 0.0034 for theta band, 0.0090 ± 0.0022 for alpha band, $0.0032 \pm 7.22 \times 10^{-4}$ for the beta band, $4.18 \times 10^{-4} \pm 2.40 \times 10^{-4}$ for the gamma1 band, and $0.0034 \pm 7.11 \times 10^{-4}$ for the gamma2 band. We observed that ISCs consistently decreased with increasing visual area hierarchy across frequency. This decrease was found in both ventral visual pathway and dorsal visual pathways.

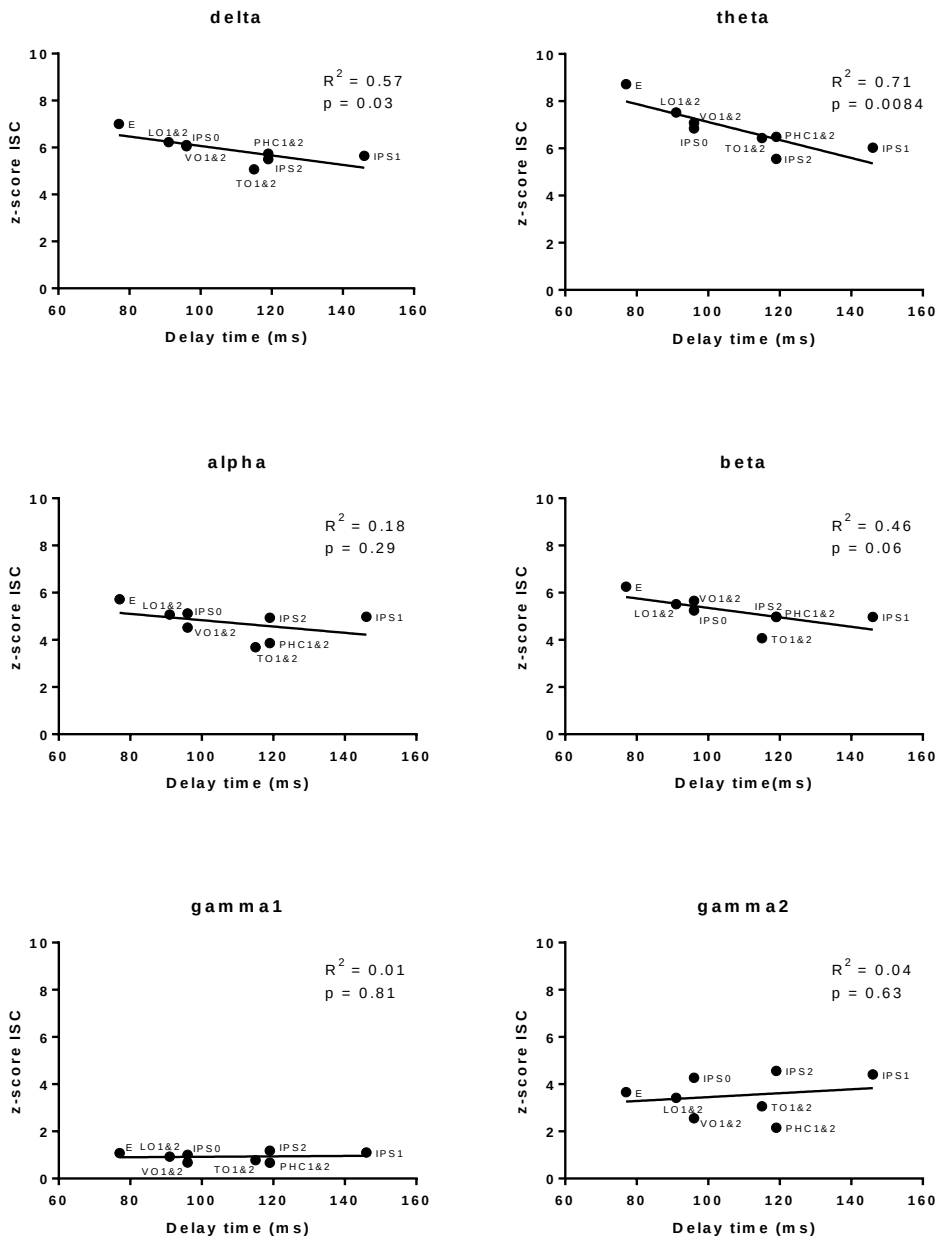


Figure 3. ISC of the ROI were related to the event-related latency of the visual area. Response latency of the ROI were obtained from an event-related object recognition task with natural backgrounds in a previous study (Cichy et al 2016b). In each of the different frequency bands, eight different areas were used to fit the linear relationship. The latency of the early visual areas is grouped, as done in Cichy et al. (2016), and depicted by the letter "E" in the figure. This point represents all ventral and dorsal V1, V2 and V3 areas.

We fitted linear models comparing ISCs and response delays in the visual areas. We used the reported delay time data for the visual areas from a recent MEG–fMRI fusion study (Cichy et al 2016b). Figure 3 shows the visual areas that were used to fit the linear relationships between delay and ISCs across frequency bands. The linear relationship between delay times and ISCs was significant in the delta ($F = 7.96$, $df = (1, 6)$, $p = 0.0303$) and theta ($F = 14.89$, $df = (1,6)$, $p = 0.0084$) bands; other bands showed no significant linear relationship.

Vertex-wise ISC

We performed vertex-wise ISC analysis and generated statistical maps of z-scores in six different frequency bands for visualization on the brain surface (see Figure 4). Significant ISCs ($z > 4.50$, Bonferroni corrected) were predominantly localized in the occipital visual areas. ISCs were highest in theta band over these occipital areas—in concordance with ROI-based analysis reported above. ISCs in low frequency bands—delta, alpha, beta—were also significant in the same regions, although at lower magnitudes. None of the gamma bands resulted in significant vertex-wise univariate ISCs.

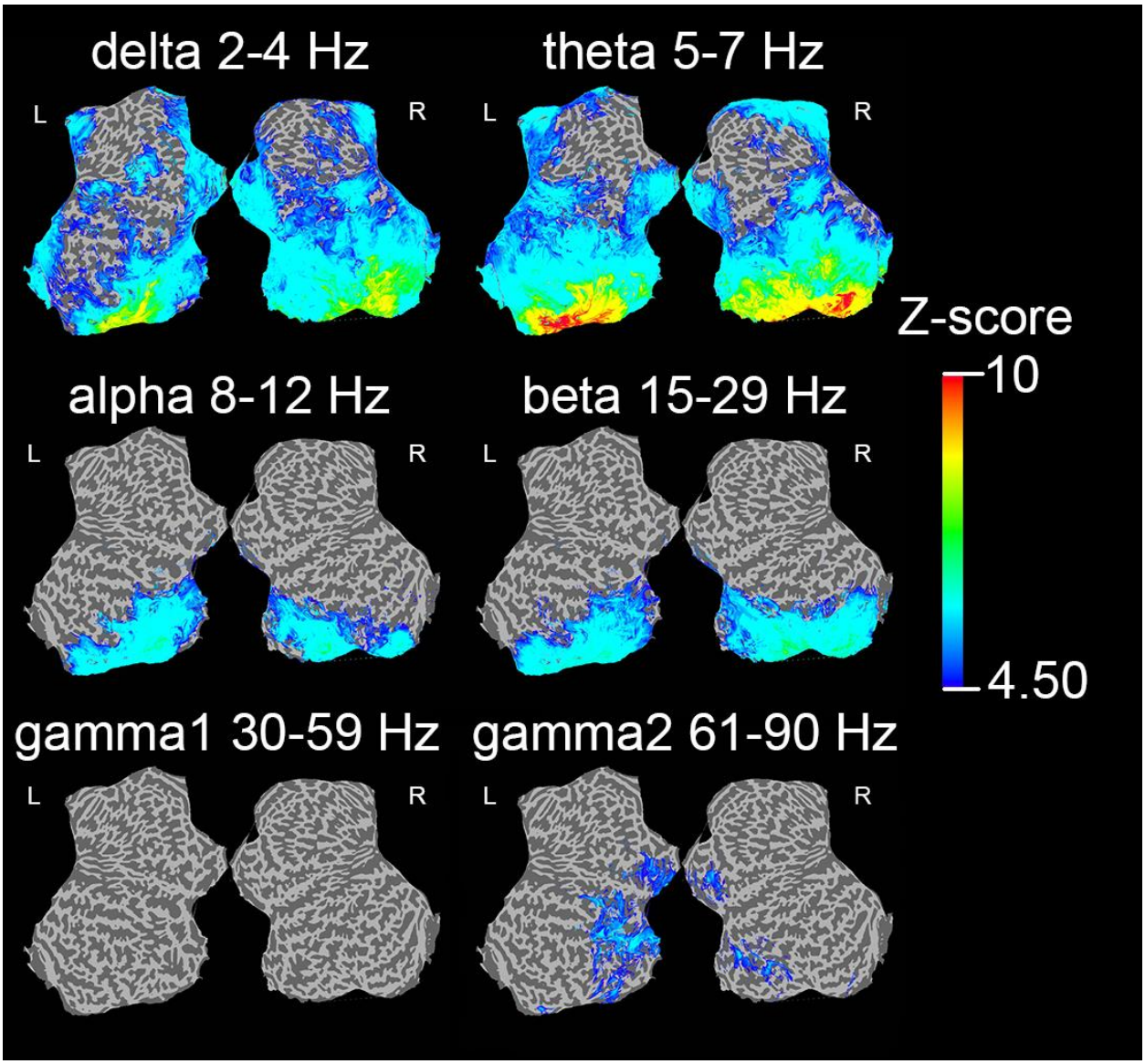


Figure 4. ISC was performed at each vertex of the cortical surface. A color scale for the z-score shows the range from Z = 4.50 (Bonferroni-corrected z-score threshold) to Z = 10. L, left hemisphere. R, right hemisphere.

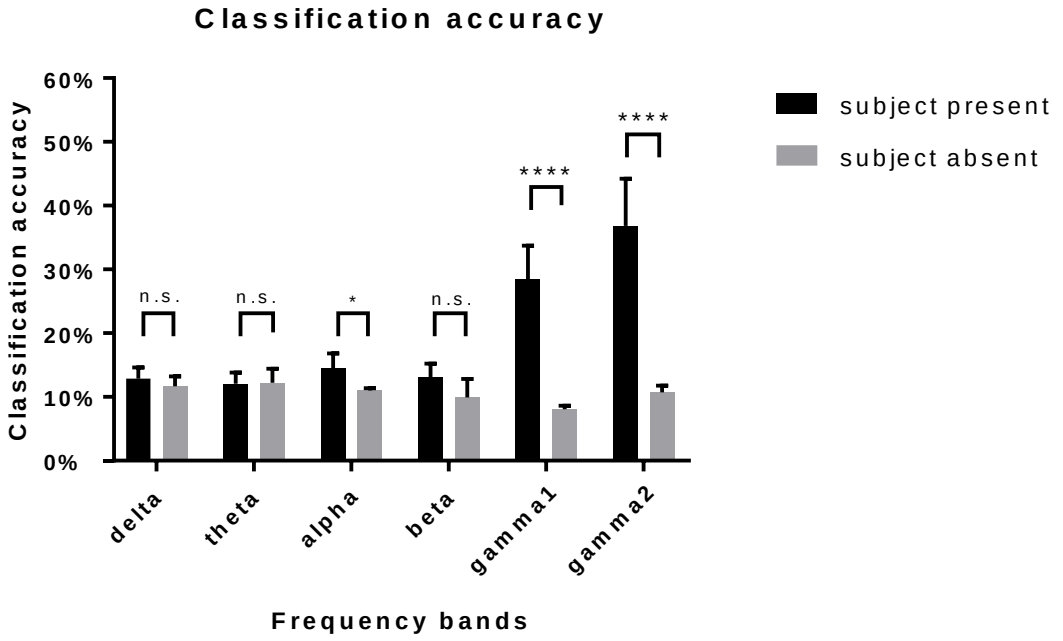


Figure 5. The spatial pattern of different frequency band responses measured by MEG were used to classify individual movie scenes for each subject. Error bars for the back bars depict standard deviation across 24 subjects for the subject present group. Error bars for grey bars represent standard deviation between two subject-absent session. Significant levels were determined by t-test between the two groups. *: $p < 0.05$. ****: $p < 0.0001$.

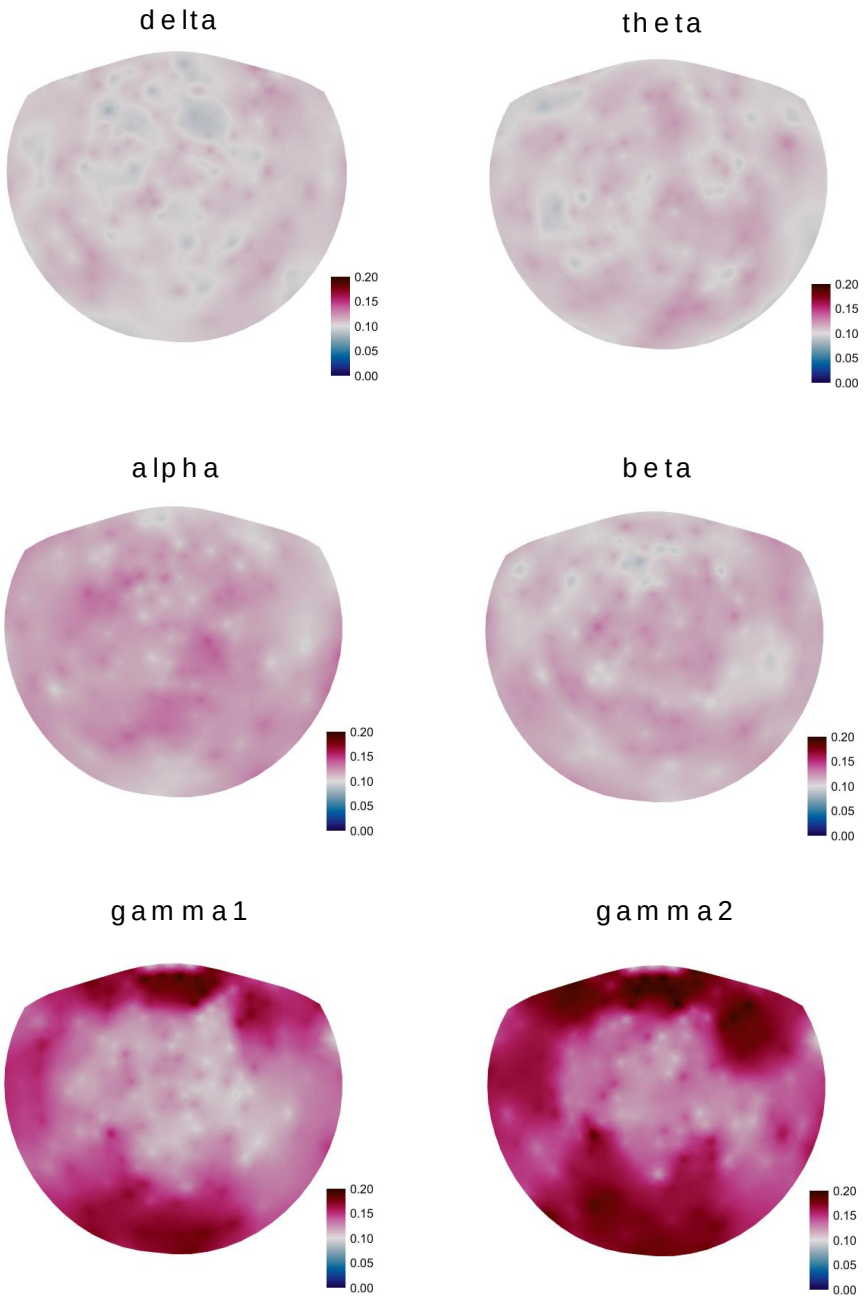


Figure 6. Single sensor classifications were performed on each subject and the results were then averaged. The color scale depicts classification accuracy results (proportion of correct classification). Noted that subjects were not in the same head position in a MEG scanner, these average results could only be used as a coarse reference.

Movie scene classification accuracies in different frequency bands

We tested the hypothesis that gamma bands can be used to classify movie scene labels from MEG sensor data. Out LDA classifier were able to achieve higher classification accuracies in subject-present group than subject-absent control group in different bands, when performing two-sample t-test (Figure 5, t_{alpha}= 2.07, p_{alpha}= 0.0495, t_{gamma1} = 5.43, p_{gamma1} = 1.4 × 10^{-5}, t_{gamma2} = 4.89, p_{gamma2} = 5.45 × 10^{-5}, all df = 24). Both gamma bands exhibited high classification accuracy in the subject present group (mean accuracy, 28.5% and 36.8%, respectively). Alpha band showed some significance between the subject-present and subject-absent control group. Other bands showed no significant differences between subject-present group and subject-absent control group. The subject-absent groups had classification performance close to 1/10, the theoretical chance level, while gamma bands showed much higher classification performance than chance level.

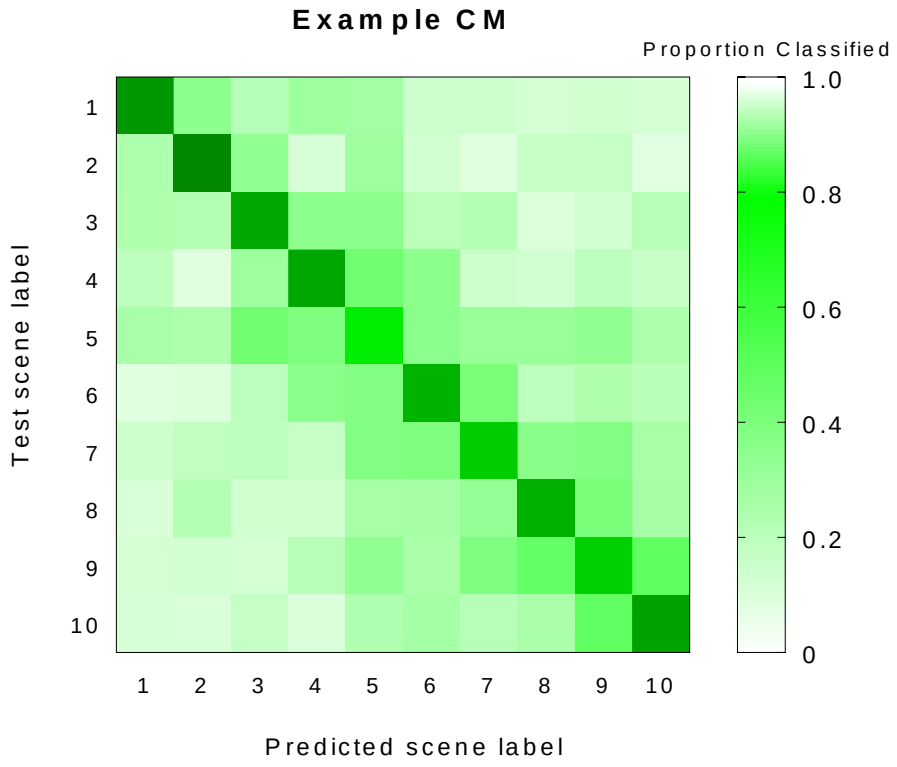
We performed classification on the data for each single sensor to determine the spatial distribution of informative sensors. The proportions of correct classifications were then visualized on a topographical map to show the physical distribution of informative sensors (see Figure 6). Both gamma1 and gamma2 bands showed higher correct classification proportions on some sensors, located on back of the head and front of the head.

Movie scene classification and scene representation geometry

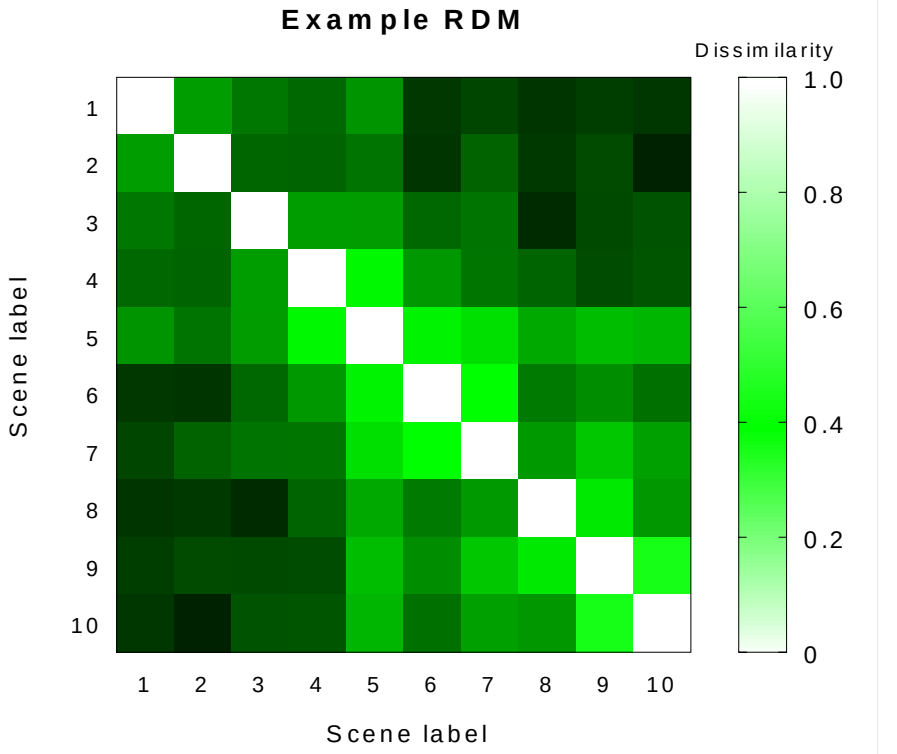
In order to estimate the representational geometry of different movie scenes, we transformed CM from classification results to RDM. The procedure was described in detail in the Methods. We show here a group-averaged CM derived from gamma2 frequency band activity in response to Clip 1 (Figure 7A). The diagonal values reflected proportion of correct classifications. Figure 7B showed an RDM generated from Figure 7A. This reflects the gamma2 representational geometry for movie scenes in Clip1 measured by the group. A dendrogram was used to visualize the representational geometry (Figure, 7C). The

scenes with lower dissimilarity on the dendrogram showed similar objects or similar backgrounds. These results implied that scene representational similarity bear object and/or scene information.

A)


Example CM
Proportion Classified
1.0
0.8
0.6
0.4
0.2
0
Test scene label
1
2
3
4
5
6
7
8
9
10
Predicted scene label

B)


Example RDM
Dissimilarity
1.0
0.8
0.6
0.4
0.2
0
Scene label
1
2
3
4
5
6
7
8
9
10
Scene label

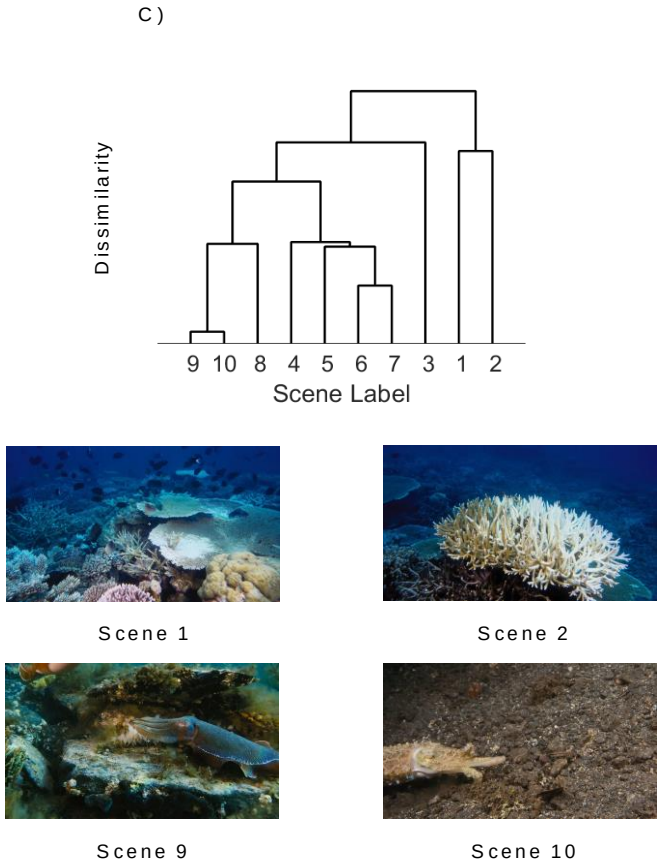


Figure 7. The 10-class LDA classification generated confusion matrices for each frequency band that were then used to compute the representational geometry for different movie scenes. (A) A sample confusion matrix generated from the classifier, trained for video Clip 1, gamma 2 frequency band, and then tested with leave-one-out cross validation. Each entry *(i, j)* in the matrix represents the proportion of test samples that the classifier predicted to be scene label *j*, while the true label of the test data was scene *i*. The color bar depicted the proportion classified. The CM shown here was an average from all subjects. (B) A sample dissimilarity matrix generated from the confusion matrix in (A), which reflected group averaged representational dissimilarity. The color bar depicts the dissimilarity between different scenes. (C) Dendrogram showing the hierarchical structure of the scene representation derived from the RDM produced in (B), with example screen captures from scenes 1, 2, 9 and 10 in this video clip shown to illustrate the movie scene similarity derived from this MEG band. The results suggest that the scene similarity metrics derived

from MEG signals can be interpretable, in that scenes grouped together share similar backgrounds and/or foreground objects.

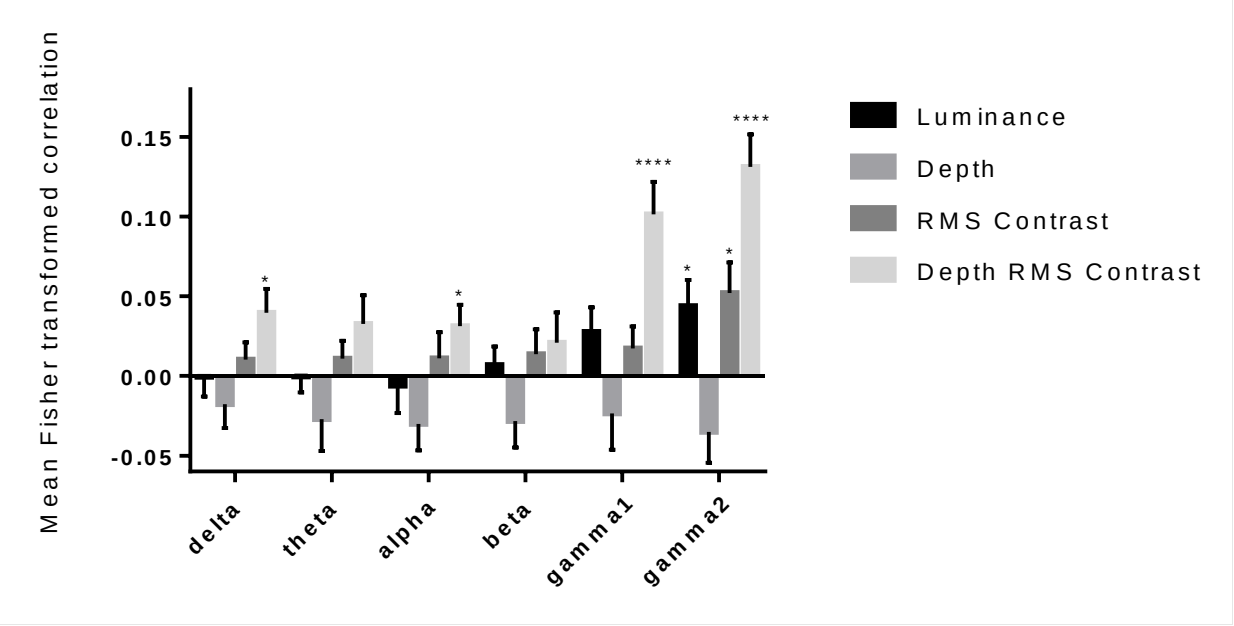


Figure 8. MEG-derived RDMS were correlated with movie scene dissimilarity metrics derived from scene structural image statistics. Luminance, depth, RMS contrast and depth RMS contrast stimuli feature conditions were plotted. Error bars represent SEM. (*= $p < 0.05$; **** = $p < 0.0001$). Depth RMS contrast structure correlated most strongly with movie scene dissimilarity derived from MEG signals, with the magnitude of the effect strongest in the gamma bands.

Scene Structural feature correlation.

We evaluated the relationship of each frequency band by correlating the representational dissimilarity matrix (RDM) of the sensors with the scene structural dissimilarity using image and depth statistics (Figure 8). Marginally significant correlations were observed for depth RMS contrast in the delta band, depth RMS contrast in the alpha band, luminance in the gamma2 band, and contrast in the gamma2 band. The t-statistics and the associated *p*-value with these conditions were $t_{dRMS\text{-}contrast,\ delta} = 2.64$, $p_{dRMS\text{-}contrast,\ delta} = 0.0145$, $t_{dRMS\text{-}contrast,\ alpha} = 2.32$, $p_{dRMS\text{-}contrast,\ alpha} = 0.0291$, $t_{luminance,\ gamma2} = 2.68$, $p_{luminance,\ gamma2} = 0.0133$, $t_{contrast,\ gamma2} = 2.72$, $p_{contrast,\ gamma2} = 0.0123$, all df = 23. Strong significant relations were observed for depth RMS contrast in the gamma1 and gamma2 bands. For this analysis, we combined the correlation values for different movie clips to maximize the group effect. The *t*-statistics and the associated p-values were $t_{dRMS\text{-}contrast,\ gamma1} = 5.00$, $p_{dRMS\text{-}contrast,\ gamma1} = 4.65 \times 10^{-5}$, $t_{dRMS\text{-}contrast,\ gamma2} = 6.48$, $p_{dRMS\text{-}contrast,\ gamma2} = 1.23 \times 10^{-6}$, all $df = 23$. This supported the view that the high gamma bands may be intimately related to complex scene structure that relate to the objects and/or background of the scene and its layout.

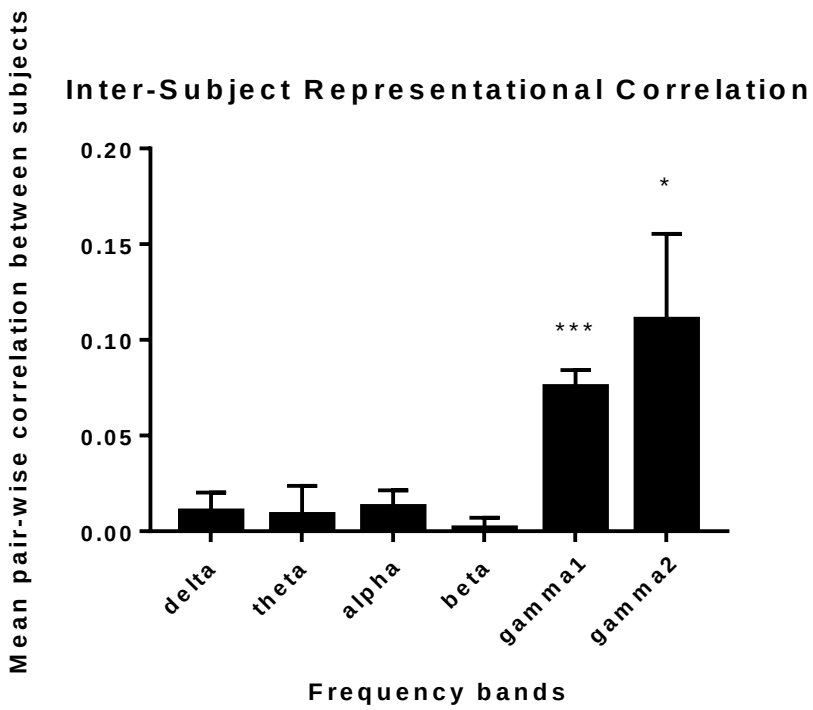


Figure 9. Inter-subject representational correlations, defined by mean pair-wise Pearson correlation of RDMs between subjects, indicated the similarity of subjects in movie scene representations. The error bars show standard deviation of inter-subject representational correlation values between the four movie clips. The stars depict significant level of inter-subject representational correlation. For gamma1, p = 0.00046. For gamma2, p = 0.016.

Inter-subject representational correlation

We only observed inter-subject similarity of scene representations in a few bands. Inter subject representational correlation was carried out by calculating the mean-pairwise correlation of RDM from all subject pairs. (Figure 9). Low frequency oscillatory bands—delta, theta, alpha and beta—all failed to show significant inter-subject representational correlation (all $t < 2.98$, all $p > 0.058$). Both gamma bands, however, demonstrated above chance inter-subject representational correlation ($t_{gamma1} = 16.84$, $p_{gamma1} = 4.56 \times 10^{-4}$; $t_{gamma2} = 4.89$, $p_{gamma2} = 0.016$, $df = 3$). These results showed that there exists strong inter-subject consistency in the representational structure for movie scenes in the gamma frequency range.

Discussion

By utilizing a multivariate metric for estimating inter-subject similarity, we were able to demonstrate for the first time that natural movie scene representations as measured by MEG are comparable across subjects in the gamma frequency range. This is in stark contrast to previous reports that had used temporal pattern similarity across MEG bands and only observed this in low-frequency bands (Bridwell et al 2015, Chang et al 2015, Lankinen et al 2014). With our dataset, we were able to largely replicate those previous findings using time-dependent frequency power changes. Additionally, we were able to demonstrate that the multivariate classification performance of movie scenes from the MEG could be used to derive an intuitive and interpretable movie scene similarity visualization. Our results further lend support to the notion that gamma band oscillations are important in stimulus representations (Hermes et al 2015a, Hermes et al 2015b, Tan et al 2016), and extend this notion to show that the pattern of gamma band scene representations is related to structural movie scene statistics, suggesting an important role for gamma band in depth perception.

ISCs in low frequency MEG brain response

Inter-subject correlation in the temporal domain has been used as a tool for probing similarity of brain responses to naturalistic stimuli, in various brain imaging studies (Chang et al 2015, Hasson et al 2008, Hasson et al 2010). In this study, we measured temporal ISC in various frequency power changes of the brain using MEG. We analyzed ISC in 25 different vision-related ROIs based on a probabilistic atlas (Wang et al 2015), and measured vertex-wise whole brain ISC. Low frequency bands (delta, theta, alpha, beta) showed significant levels of ISC, indicating high similarity of temporal brain responses between subjects during movie viewing. The mean pair-wise correlation coefficient in the lower frequency bands in this study is low but significant, which was consistent with previous studies, e.g. (Bridwell et al 2015). Visual response delay (Cichy et al 2016b) of a certain ROI was related to ISC score. Significant linear relationships between visual response delay and ISC indicated that ISC decreased along the visual hierarchy. We did not observed significance in ISC between monostotic viewing condition and stereoscopic viewing condition (data not shown). This suggests that ISC measurement lack sufficiency to detect the difference between the viewing conditions in our experiment.

The naturalistic stimuli used in this study were non-emotional and did not include any narrative structure. To mitigate the concern raised by a previous study that non-emotional movie induces inconsistent eye movements amongst subjects, we used a central fixation in all the viewing conditions—this has been shown to increase ISC (Ki et al 2016). We noticed elevations in the gamma rand bands in the MEG data (Supplementary Figure 1), which could be signs that our stimuli elicit strong responses in the gamma range.

Classification of movie scenes from MEG sensor patterns

A number of studies have shown MEG or EEG data can be used to classify stimuli category in event-related designs (Cichy et al 2014, Kaneshiro et al 2015). Here, we extend the idea of stimulus category classification from evoked responses to time intervals during a movie scene. We obtained significant

accuracy compared to the subject absent condition, when decoding movie scene labels using a 10-category LDA classification technique. The classification accuracies were the highest in the subject-present condition in gamma bands, indicate that reliable and consistent information exists during a movie scene and that it can be detected with MEG and classified using multi-class techniques such as LDA.

Scene classification patterns were meaningful, in that we were able to derive scene similarity estimates that were intuitive and interpretable. We visualized the movie scene structure in the representational space using a dendrogram (Figure 7C). Intuitively, the dendrogram representation suggests movie scenes were classified similarly based on the scene video content (objects and/or background). While preliminary, this result suggests that the representational structure estimated from the multivariate MEG classification was related to perceptual qualities of the scene. This was further supported by the fact that most occipital sensors exhibited high classification performance individually for gamma bands, meaning that the visual scene—as opposed to top-down components or possible semantic/narrative content—was the key driver of the observed patterns.

Movie scene classification performance across frequency bands

We found that only gamma band patterns contribute to robust scene classification. Gamma oscillatory activity is implicated as a core mechanism of cortical computation, but it is still debatable that in response to natural scene presentation, whether gamma amplitude is related to the presence of a stimulus or to the features of the stimuli scene (Brunet et al 2015, Hermes et al 2015a). Our results support the idea that gamma amplitude is related to natural scene features, because we found that gamma band patterns could be used to classify movie scenes. In this, we extended the notion of evoked and induced gamma band activity to multivariate gamma patterns over sensors. These findings provide a novel perspective in the debate of whether gamma band activity is tuned to stimulus features.

Low-frequency activity (below 30 Hz, except alpha band) showed no significant classification accuracy when compared to subject absent control data. Further, they showed no significant results in inter-subject representational correlation. Furthermore, we found that the movie scene statistics, especially depth RMS contrast, were significantly related to the gamma band RDMs. These results suggest that low-frequency activity patterns are not suitable for multivariate pattern analysis, while gamma band patterns are related to visual perception. While we did observe significant movie scene classification based on alpha band activity, the pattern of classification was not generalizable across subjects—in other words, the way alpha band patterns represented movie scenes appeared to be more subject-specific. Alpha band activity was previously reported to be related to visual attention selection and control (Gola et al 2013, Mathewson et al 2011, Sauseng et al 2005), which might be related to transient alert changes between movie scenes, or the linear temporal trend in the frequency powers during the presentation of the stimuli.

Inter-subject representational similarity in naturalistic movie viewing

We extended ISC from temporal correlation of MEG responses to inter-subject representational correlation, using representational similarity analysis based on a large corpus of previous reports (Guntupalli et al 2016, Haxby et al 2014, Kriegeskorte et al 2008a). Our results show that representational patterns of scenes were correlated and similar between subjects in only the gamma frequency range, but not in low-frequency bands. Inter-subject representational correlation likely reflects common perceived qualities of the naturalistic stimuli and thus is useful for identifying commonality between subjects in response to naturalistic stimuli in future studies. One recent fMRI study showed that inter-subject representational similarity could be identified in fMRI datasets (Chen et al 2020b). Our finding provides possible explanations that gamma bands could be the target of fMRI-measured inter-subject representational correlation.

Our use of inter-subject representational correlation based on representational similarity in individual frequency bands offers a complementary approach to estimation of prototypical responses from MEG/EEG signals. Firstly inter-subject representational correlation is anatomically independent and data from MEG sensors can be used for classification and estimation of inter-subject representational correlation, avoiding the information loss from the noise added to brain source projections (Baillet 2017). Inter-subject representational correlation does not require between-subject anatomical registration, avoiding a critical concern in previous ISC approaches where anatomical and functional registrations were not in agreement (Guntupalli et al 2016). Secondly, inter-subject representational correlation ISC reflects information about the movie scenes, as we have shown here—the high gamma band representation is similar to that of the depth structure in these scenes. This measurement may provide supplementary information beyond existing knowledge when investigating brain response to various visual stimulus features.

In interpreting our results, we considered situations that may have led to better classification by gamma band activity beyond relatedness to visual features. One of the possibilities was that in any time-frequency analysis, higher frequency signals would have better temporal resolution, but less frequency resolution. In our case of our classification procedure, the 500ms samples we drew from the movie scenes could have contain more information in higher frequency bands than lower ones. Future work on movie scene classification could include various sample length for different frequency bands, based on their relative temporal resolution. Another possibility is that the strong gamma inter-subject representational correlation in the MEG data were related to involuntary eye movement (Yuval-Greenberg et al 2008), which was unpreventable in long-term fixation movie experiment. We had incorporated SSP routines in our pre-processing step to remove saccade related artefacts. Micro-saccades and fixational eye movements most certainly will have been present even after pre-processing, but it is not clear that such movements are modulated by scene statistics, so it is not likely that they

contributed to the patterns reported here. Future work involving high spatial and temporal resolution eye recordings during MEG would help to explore this possibility.

Because of the inherent difficulties in simulating cortical oscillators, it is challenging to test the sensitivity of our approach—constructing the positive control is not so straightforward. The sensitivity of the approach is certainly limited by the sensitivity of the MEG sensors and their low spatial resolution, which obfuscate spatially-adjacent oscillators. Careful modelling of cortical oscillators and validation of such models may yield effective simulations of MEG responses. Once available, such models could allow for effective characterization of the sensitivity of different analytic methods.

In comparison to other ISC approaches where inter-subject similarity was estimated from temporal signatures, our naturalistic movie presentation is less affected by extraneous factors such as attentional, emotional or semantic aspects that are typically engaged by cinematic viewing (Hasson et al 2008). The drawback of inter-subject representational correlation is the lack of spatial information when only the sensor data were used. Very few successful MEG source classifications had been reported (Kozunov et al 2017). Future work on projection classifications on MEG source data may endow us with the capability to overcome spatial interpretation of inter-subject representational correlation and derive representation-related prototypical responses from the fast MEG/EEG signals.

Reference

Baillet S. 2017. Magnetoencephalography for brain electrophysiology and imaging. Nat Neurosci 20: 327-39

Bartels A, Zeki S. 2004. Functional brain mapping during free viewing of natural scenes. Hum Brain Mapp 21: 75-85

Bridwell DA, Roth C, Gupta CN, Calhoun VD. 2015. Cortical Response Similarities Predict which Audiovisual Clips Individuals Viewed, but Are Unrelated to Clip Preference. PLoS One 10: e0128833

Brunet N, Bosman CA, Roberts M, Oostenveld R, Womelsdorf T, et al. 2015. Visual cortical gamma-band activity during free viewing of natural images. Cereb Cortex 25: 918-26

Chang WT, Jaaskelainen IP, Belliveau JW, Huang S, Hung AY, et al. 2015. Combined MEG and EEG show reliable patterns of electromagnetic brain activity during natural viewing. Neuroimage 114: 49-56

Chen G, Shin YW, Taylor PA, Glen DR, Reynolds RC, et al. 2016. Untangling the relatedness among correlations, part I: Nonparametric approaches to inter-subject correlation analysis at the group level. Neuroimage 142: 248-59

Chen PA, Jolly E, Cheong JH, Chang LJ. 2020. Intersubject representational similarity analysis reveals individual variations in affective experience when watching erotic movies. Neuroimage: 116851

Cichy RM, Pantazis D, Oliva A. 2014. Resolving human object recognition in space and time. Nat Neurosci 17: 455-62

Cichy RM, Pantazis D, Oliva A. 2016. Similarity-Based Fusion of MEG and fMRI Reveals Spatio-Temporal Dynamics in Human Cortex During Visual Object Recognition. Cereb Cortex 26: 3563-79

Cohen XM. 2014. Analyzing Neural Time Series Data: Theory and Practice. pp. 33. The MIT Press.

Dmochowski JP, Sajda P, Dias J, Parra LC. 2012. Correlated components of ongoing EEG point to emotionally laden attention - a possible marker of engagement? Front Hum Neurosci 6: 112

Duncan KK, Hadjipapas A, Li S, Kourtzi Z, Bagshaw A, Barnes G. 2010. Identifying spatially overlapping local cortical networks with MEG. Hum Brain Mapp 31: 1003-16

Gaebler M, Biessmann F, Lamke JP, Muller KR, Walter H, Hetzer S. 2014. Stereoscopic depth increases intersubject correlations of brain networks. Neuroimage 100: 427-34

Gola M, Magnuski M, Szumska I, Wrobel A. 2013. EEG beta band activity is related to attention and attentional deficits in the visual performance of elderly subjects. Int J Psychophysiol 89: 334-41

Guntupalli JS, Hanke M, Halchenko YO, Connolly AC, Ramadge PJ, Haxby JV. 2016. A Model of Representational Spaces in Human Cortex. Cereb Cortex 26: 2919-34

Hadjipapas A, Adjamian P, Swettenham JB, Holliday IE, Barnes GR. 2007. Stimuli of varying spatial scale induce gamma activity with distinct temporal characteristics in human visual cortex. Neuroimage 35: 518-30

Hall H. 2009. Under the sea.

Hamalainen MS, Ilmoniemi RJ. 1994. Interpreting magnetic fields of the brain: minimum norm estimates. Med Biol Eng Comput 32: 35-42

Hasson U, Landesman O, Knappmeyer B, Vallines I, Rubin N, Heeger DJ. 2008. Neurocinematics: the neuroscience of film. Projections 2: 1 – 26

Hasson U, Malach R, Heeger DJ. 2010. Reliability of cortical activity during natural stimulation. Trends Cogn Sci 14: 40-8

Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R. 2004. Intersubject synchronization of cortical activity during natural vision. Science 303: 1634-40

Haufe S, DeGuzman P, Henin S, Arcaro M, Honey CJ, et al. 2018. Elucidating relations between fMRI, ECoG, and EEG through a common natural stimulus. Neuroimage 179: 79-91

Haxby JV, Connolly AC, Guntupalli JS. 2014. Decoding neural representational spaces using multivariate pattern analysis. Annu Rev Neurosci 37: 435-56

Hermes D, Miller KJ, Wandell BA, Winawer J. 2015a. Gamma oscillations in visual cortex: the stimulus matters. Trends Cogn Sci 19: 57-8

Hermes D, Miller KJ, Wandell BA, Winawer J. 2015b. Stimulus Dependence of Gamma Oscillations in Human Visual Cortex. Cereb Cortex 25: 2951-9

Holmes CJ, Hoge R, Collins L, Woods R, Toga AW, Evans AC. 1998. Enhancement of MR images using registration for signal averaging. J Comput Assist Tomogr 22: 324-33

Kaneshiro B, Perreau Guimaraes M, Kim HS, Norcia AM, Suppes P. 2015. A Representational Similarity Analysis of the Dynamics of Object Processing Using Single-Trial EEG Classification. PLoS One 10: e0135697

Kayser C, Salazar RF, Konig P. 2003. Responses to natural scenes in cat V1. J Neurophysiol 90: 1910-20

Ki JJ, Kelly SP, Parra LC. 2016. Attention Strongly Modulates Reliability of Neural Responses to Naturalistic Narrative Stimuli. J Neurosci 36: 3092-101

Kozunov V, Nikolaeva A, Stroganova TA. 2017. Categorization for Faces and Tools-Two Classes of Objects Shaped by Different Experience-Differs in Processing Timing, Brain Areas Involved, and Repetition Effects. Front Hum Neurosci 11: 650

Kriegeskorte N, Mur M, Bandettini P. 2008. Representational similarity analysis - connecting the branches of systems neuroscience. Front Syst Neurosci 2: 4

Lahnakoski JM, Glerean E, Jaaskelainen IP, Hyona J, Hari R, et al. 2014. Synchronous brain activity across individuals underlies shared psychological perspectives. Neuroimage 100: 316-24

Lankinen K, Saari J, Hari R, Koskinen M. 2014. Intersubject consistency of cortical MEG signals during movie viewing. Neuroimage 92: 217-24

Malinen S, Hari R. 2011. Data-based functional template for sorting independent components of fMRI activity. Neurosci Res 71: 369-76

Martinovic J, Gruber T, Hantsch A, Muller MM. 2008. Induced gamma-band activity is related to the time point of object identification. Brain Res 1198: 93-106

Mathewson KE, Lleras A, Beck DM, Fabiani M, Ro T, Gratton G. 2011. Pulsed out of awareness: EEG alpha oscillations represent a pulsed-inhibition of ongoing cortical processing. Front Psychol 2: 99

Mukamel R, Gelbard H, Arieli A, Hasson U, Fried I, Malach R. 2005. Coupling between neuronal firing, field potentials, and FMRI in human auditory cortex. Science 309: 951-4

Muller MM, Bosch J, Elbert T, Kreiter A, Sosa MV, et al. 1996. Visually induced gamma-band responses in human electroencephalographic activity--a link to animal studies. Exp Brain Res 112: 96-102

Muthukumaraswamy SD, Singh KD, Swettenham JB, Jones DK. 2010. Visual gamma oscillations and evoked responses: variability, repeatability and structural MRI correlates. Neuroimage 49: 3349-57

Naselaris T, Prenger RJ, Kay KN, Oliver M, Gallant JL. 2009. Bayesian reconstruction of natural images from human brain activity. Neuron 63: 902-15

Perry G, Randle JM, Koelewijn L, Routley BC, Singh KD. 2015. Linear tuning of gamma amplitude and frequency to luminance contrast: evidence from a continuous mapping paradigm. PLoS One 10: e0124798

Porcaro C, Ostwald D, Hadjipapas A, Barnes GR, Bagshaw AP. 2011. The relationship between the visual evoked potential and the gamma band investigated by blind and semi-blind methods. Neuroimage 56: 1059-71

Poulsen AT, Kamronn S, Dmochowski J, Parra LC, Hansen LK. 2017. EEG in the classroom: Synchronised neural recordings during video presentation. Sci Rep 7: 43916

Sauseng P, Klimesch W, Stadler W, Schabus M, Doppelmayr M, et al. 2005. A shift of visual spatial attention is selectively associated with human EEG alpha activity. Eur J Neurosci 22: 2917-26

Tadel F, Baillet S, Mosher JC, Pantazis D, Leahy RM. 2011. Brainstorm: a user-friendly application for MEG/EEG analysis. Comput Intell Neurosci 2011: 879716

Tan HM, Gross J, Uhlhaas PJ. 2016. MEG sensor and source measures of visually induced gamma-band oscillations are highly reliable. Neuroimage 137: 34-44

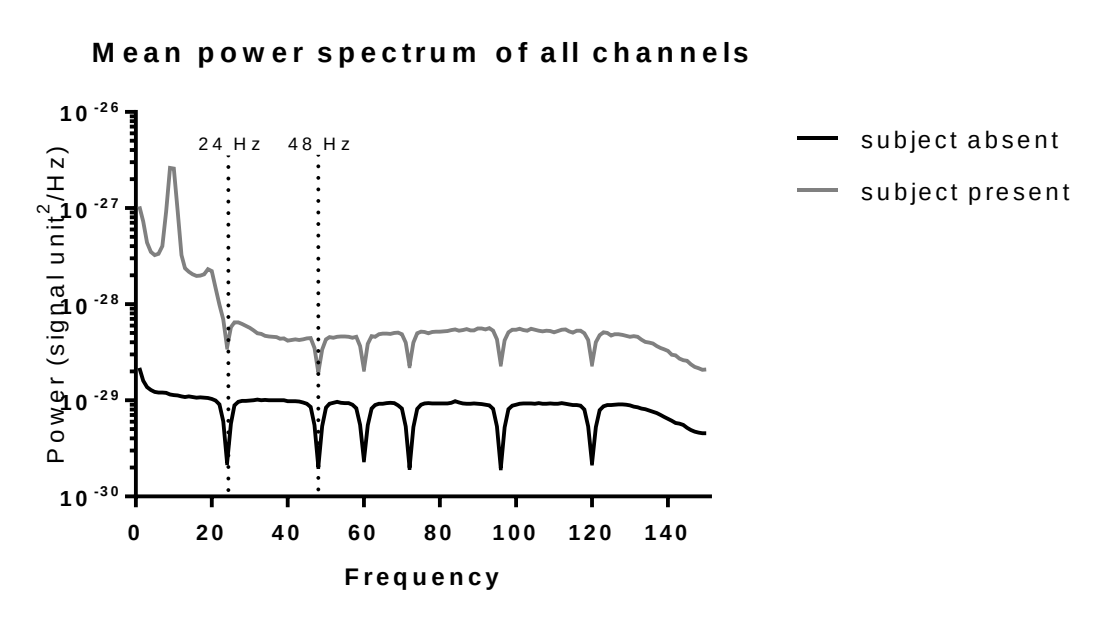
Wang L, Mruczek RE, Arcaro MJ, Kastner S. 2015. Probabilistic Maps of Visual Topography in Human Cortex. Cereb Cortex 25: 3911-31

Yuval-Greenberg S, Tomer O, Keren AS, Nelken I, Deouell LY. 2008. Transient induced gamma-band response in EEG as a manifestation of miniature saccades. Neuron 58: 429-41

Supplementary

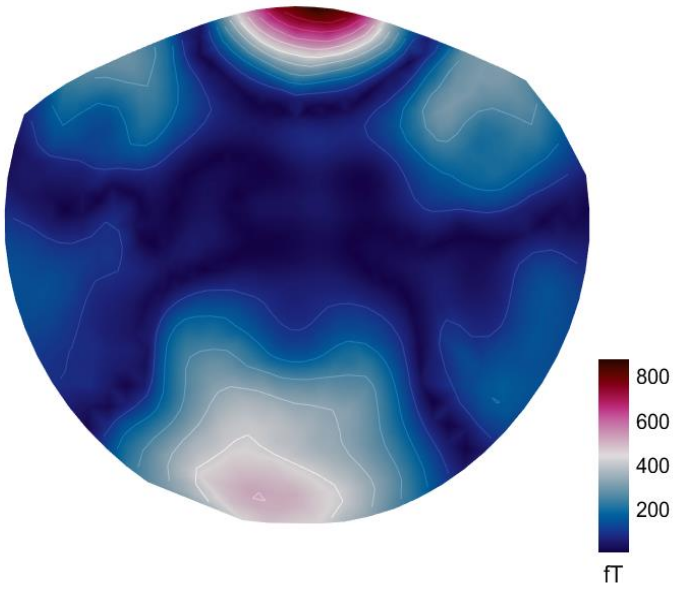
Supplementary figure 1.

Power spectrum of MEG recoding from subject absent experiment and an example from subject present data. Notch filters of power line noise (60 Hz) and its harmonics were applied to the data. Notch filters of screen frame rate (24 Hz) and its harmonics were applied to the data.



Supplementary figure 2.

Screen artifact topography from the subject absent recording.



Chapter 3: 3D movie viewing changes brain visual network across subjects

3D movie viewing changes brain visual network across subjects

Abstract

Visual cortical networks have been characterized as a hierarchical structural and stable, while also dynamically adapted to the visual stimuli (Cole et al 2013, Vanderwal et al 2017). It has been shown that stereopsis induced network responses in the visual brain when the subjects were watching naturalistic movies (Gaebler et al 2014). It is crucial to understand the network changes induced by stereopsis at the individual level. An effective way of eliminating subject-specific noise in the connectivity networks was inter-subject functional connectivity (ISFC), which focused on measuring cross-subject connections between brain areas rather than within-subject connections in traditional functional connectivity (FC) methods. Networks that engaged with the stimuli presented can be detected using ISFC methods compared to traditional FC (Cohen et al 2017, Di & Biswal 2020, Simony et al 2016). In this study, we adopted a naturalist 3D movie-viewing paradigm to assess the ISFC changes in the cortical visual areas. We identified connections in the dorsal visual stream that responded strongly in the 3D movie-viewing conditions. We conclude that the 3D movie-viewing engaged different network responses rather than increase the robustness of the existing visual networks.

Introduction

Connectivity networks

Functional connectivity (FC) networks have been shown to adapt and change in response to external stimuli (Cole et al 2014, Gonzalez-Castillo & Bandettini 2018, Li et al 2015, Mennes et al 2013, Vanderwal et al 2017, Wen et al 2018). However, the structural organization of the networks is stable across individuals and test sessions in network classifications (Zuo et al 2010). This is especially true in

the visual areas of the brain, where a hierarchical structural organization has been established using FC (Fiser et al 2004, Golland et al 2007, Moeller et al 2009). Even a complex visual stimulus has shown not to significantly alter this basic intrinsic structural organization (Vanderwal et al 2015). In the current study, we investigated how stable cortical networks respond to a naturalistic movie that adds an essential component—stereoscopic depth. We also compared FC to a new method—inter-subject functional connectivity (Simony et al 2016)—which has been proposed to reduce subject-specific effects from cortical connectivity estimates.

Depth perception and 3D movie viewing

Horizontal separation of the eyes results in slightly different image inputs between the two eyes, and this binocular disparity, along with other perceptual cues, provided bases for cortical computations of depth (for a review, see Parker, 2007). Classical studies using simple synthetic stimuli have shown that both dorsal and ventral pathways are involved in depth perception (Preston et al 2008).

The dorsal visual stream has been shown to be important in coding relative disparity, which is the key to depth perception (Backus et al 2001). Many neural imaging studies have found that along with many other regions, V3a, a dorsal visual stream region, is essential in relative disparity (Anzai & DeAngelis 2010, Cottereau et al 2012, Goncalves et al 2015, Henderson et al 2019). However, to understand how the brain processes stereoscopic depth in everyday life, naturalistic stimuli like 3D movies need to be used rather than simple, synthetic stimuli. 3D movie viewing has been shown to increase inter-subject synchrony across cortical visual areas (Gaebler et al 2014), which, at a minimum, implies that the depth processing in realistic stimuli different from that of simple synthetic stimuli. However, it is not clear what underlies the increased inter-subject synchrony—is it simply due to a reduction of noise brought by greater certainty of depth afforded by the addition of stereoscopic cues, or is it that disparity induces an additional, functionally distinct cortical network?

ISFC as a novel approach

Inter-subject functional correlation analysis (ISFC; Simony et al 2016) is a tool to investigate the stimulus-dependent FC across subjects. By combining the advantages of inter-subject correlation (ISC)—the mean correlation of the time-series response of pairs of subjects—and FC analysis, this new approach mitigates the subject-specific noise (e.g. physiological noise, scanner noise, etc.) in the FC. One study has shown that ISFC improves the characterization of dynamic correlation patterns responding to external stimuli, by revealing that strong ISFC connections can be found in both auditory networks and language networks when the subjects were listening to an intact story, while strong ISFC connections can only be found in auditory networks when the subjects were listening to word scrambled stories (Cohen et al 2017, Simony et al 2016). This phenomenon was not discovered using traditional FC methods. It is concluded that because ISFC eliminates subject-specific noise, it effectively increases the signal-to-noise ratio for the detection of stimulus-dependent inter-regional correlations.

In the current study, we utilized ISFC to detect disparity-induced variations in cortical networks. By using a natural, narrative-free, and visually stimulating movie that was acquired stereoscopically, we were able to assess functional connectivity changes that are driven by binocular disparity. Our analysis approach also increased our sensitivity to detect changes in FC, and we were able to then test whether binocular disparity simply increases the robustness of existing networks, or whether it engages an entirely different set of networks.

Method and material

Procedure and Stimuli

Fifty-one healthy subjects with normal or corrected to normal vision were recruited for this study (average age 25.3). Two different movie clips from “Under the Sea 3-D: IMAX”(Hall 2009) were shown while the subjects were scanned in a functional magnetic resonance imaging (fMRI) scanner. Each movie clip was approximately five minutes, and was presented twice, once in 2-D, and once in 3-D. The

subjects were wearing polarized 3-D glasses during the entire experiment. All subjects verbally reported perceiving the 3-D session of both movie clips. The movie clips were cast on a 3-D LCD BOLD screen, and a mirror reflected the movie content to subjects' eyes. The viewing distance was 1.7m, and the visual angle for the screen was 17° (width) by 9.4° (height). The playback was controlled by MATLAB code, using a stereoscopic player (http://www.3dtv.at). A dot was presented on the center of the screen, and subjects were asked to keep their eyes fixated upon it. No audio was played in the process of movie presentation.

Data acquisition and stimuli

The scan was performed using a 3T Siemens TIM Trio scanner, with the isotropic voxel size of 3mm. The time of repetition (TR) was 2.0s. We chose a field of view for the fMRI scan focusing on the back of the head (For detailed scanner parameters, see the previous published work from our lab, Zhang and Farivar, 2021).

Pre-processing

All preprocessing steps were performed using AFNI and SUMA (Cox 1996, Saad & Reynolds 2012). The preprocessing steps included slice-timing correction, distortion correction, motion correction. We also used ANATICOR model (Jo et al 2010) to detrend the data and reduce the influence of physiological noise. The pre-processed data for each subject was then registered to the subject's T1 anatomical image. All the transformation parameter files from each of the pre-processing steps were combined to a single set of transformation parameters and were stored in a single warping file. This method reduced the number of unnecessary interpolations in the warping process. We chose to perform the analysis on a surface model, which avoided non-linear volumetric registration between subjects, and showed better generalization between subjects (Saad & Reynolds 2012). We used AFNI's 3dVol2Surf function to project the volumetric data to the standard 60 pial surface. To analyze the structure connections between visual areas, we adopted a probabilistic atlas to define regions of interest (ROI; Wang et al 2015)—in total, all

22 visual ROIs from the atlas were selected. We averaged the time series of all vertices in each of the ROI to generate one timeseries for each ROI.

Inter-subject correlation (ISC)

Inter-subject correlation (ISC) was performed by averaging the Fisher transformed pair-wised Pearson correlation coefficient for *each node* on the cortical surface—not ROI (Hasson et al 2010). We compared the ISC between two viewing conditions (2-D vs. 3-D) using paired t-test. We used Bonferroni correction for multiple comparisons (Genovese et al 2002). The significance level alpha was set to 0.05.

Functional connectivity between visual areas

To investigate the FC changes in visual areas, we constructed one correlation matrix for the selected 22 different visual ROIs for each subject for each hemisphere and one correlation matrix for the interhemispheric connections. To do this, the BOLD time series were averaged within each visual area. Pearson correlation matrix were constructed. We then averaged the Fisher transformed correlation matrix for all subjects. The statistical significance of the differences between the two viewing conditions was determined by using paired t-tests for each element. The details of this statistical analysis are described in the next paragraph.

Inter-subject functional correlation (ISFC)

To assess the stimuli specific functional connectivity between visual areas while minimizing subject-specific effects and maximizing signal-to-noise, we performed inter-subject functional correlation (ISFC) (Simony et al 2016). We measured both ROI-based ISFC as well as seed-based ISFC. For ROI-based ISFC, we calculated Pearson's correlation between subjects and across different ROIs. We compared ROI-based ISFC between conditions and determined the significance level using paired t-tests in both hemispheres and intra-hemispheric connections. We used false discovery rate (FDR) to correct for multiple comparisons and set the significance level $q = 0.05$.

For seed based ISFC, we calculated Pearson's correlation between one subject's seed region data to other subjects' data from all vertices, excluding vertices that are beyond the field of view in our data acquisition. We used V1 and V3a as seed regions because they had been previously reported to being related to disparity perception (Goncalves et al 2015, Henderson et al 2019, Li et al 2017). The correlation coefficients for each pair were then Fisher-transformed prior to averaging across all pairs. The significance levels were determined by a paired t-test (p=0.0001).

Decoding using FC and ISFC networks and sub-networks

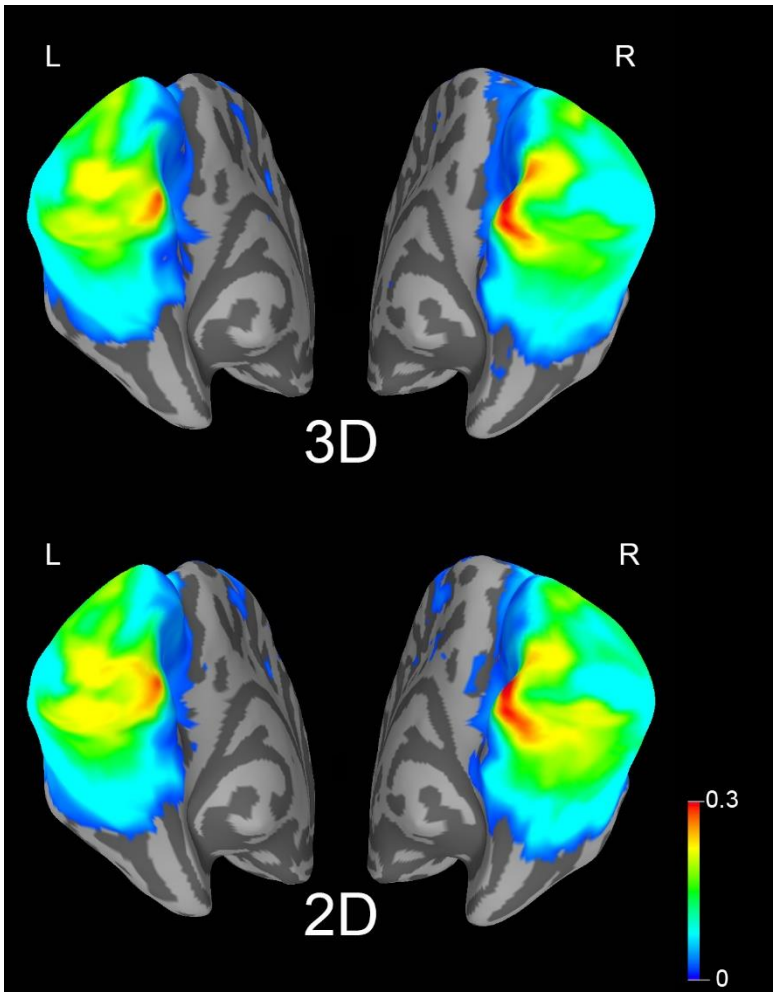
To investigate whether the network could be used to identify the viewing condition (2D vs 3D), we trained support vector machines (SVM) to classify the movie viewing conditions using the network estimates. The network metrics were symmetrized and vectorized before the decoding analysis. For FC networks, the SVMs were trained and tested by using leave-one-subject-out cross-validation. For ISFC networks, the classifications were performed for each subject, using the connectivity matrices between the chosen subject and all other subjects. We further confirmed the results by using a randomly generated time series as a true negative control (Gorgen et al 2018).

Results

ISC in 3D movie viewing

ISC was calculated on the cortical surface atlas in both 3D viewing condition and 2D viewing condition (Fig. 1A). The maximum value of vertex-wised ISC was 0.3174 in 3D viewing and 0.3494 in 2D viewing. High ISC values were showing in the primary visual regions, whereas the lower ISC values can be observed in other parts of the occipital cortex. This pattern is consistent in both 2D movie viewing and 3D movie viewing. Although no differences have been found in each of the vertices between the two conditions in the two movie clips (data not shown), the pattern of the ISC values deviated from the diagonal lines on the scatter plot, where vertex-wised was plotted between 3D and 2D viewing condition in both movie clips (Fig. 1B). On average between the two clips, 75.98% of the non-zero vertex-wised

ISCs were higher in 3D condition (below the diagonal $y = x$ black line in Fig. 1B), and 24.02% were higher in 2D condition (above the diagonal $y = x$ black line in Fig. 1B).


L
R
3D
L
R
2D
0.3
0

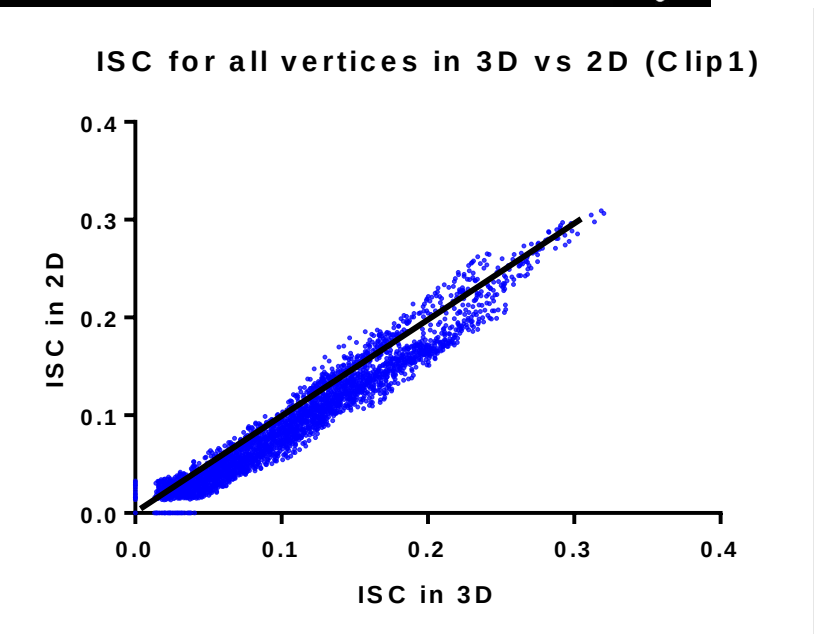
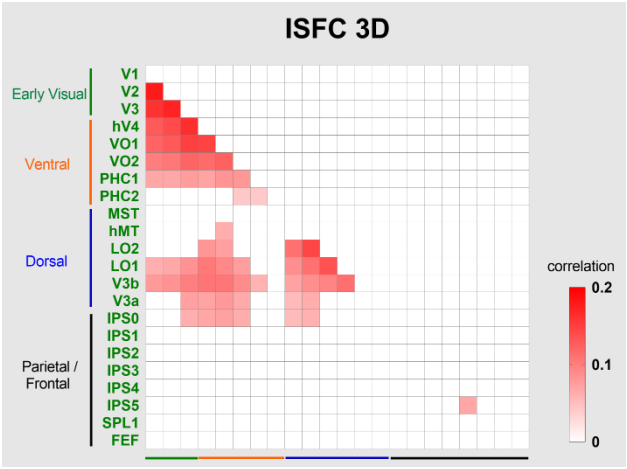

ISC for all vertices in 3D vs 2D (Clip1)
ISC in 2D
ISC in 3D
0.0
0.1
0.2
0.3
0.4

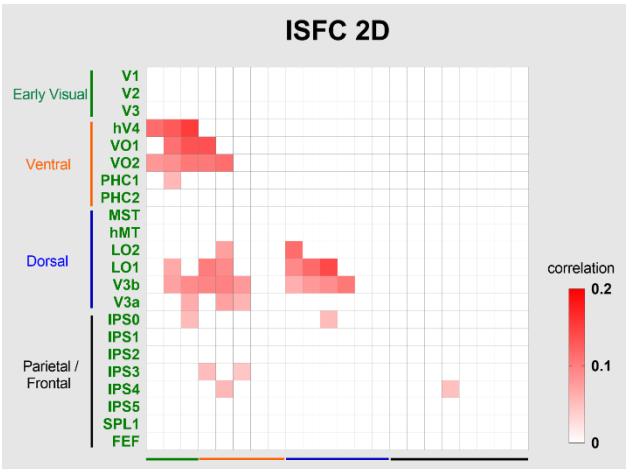
Figure 1. Inter-subject correlation in 3D movie viewing versus 2D movie viewing. A), Cortical voxel map of ISC in 3D movie viewing and 2D movie viewing. ISC is shown as mean pairwise Pearson's correlation coefficient for all subject pairs. A posterior view is chosen to clearly present the visual cortical regions. L, left hemisphere. R, right hemisphere. B), Scatter plot for all the vertices ISC, showing comparisons of 3D movie versus 2D movie. The result of movie clip 1 were shown. A black line in the diagonal is a line function $y=x$, representing no change between the two conditions.

Inter-subject functional connectivity of the visual networks

ISFC were measured in both 3D movie-viewing and 2D movie viewing conditions in all 22 visual regions, which were defined in a probabilistic atlas (Wang et al 2015). Significant connections in the networks were shown using an FDR error rate control of q = 0.05 (Fig. 2). The ISFC results showed that in both viewing conditions, detectable connections within dorsal visual areas and connections between early visual and dorsal areas were present. The number of detectable significant connections were 63 in 3D conditions, 40 in 2D conditions. Among the detectable connections, 34 of them were present in both 3D and 2D conditions, which was 54% in 3D condition, and 85% in 2D condition.

A)


ISFC 3D
Early Visual
Ventral
Dorsal
Parietal / Frontal
V1
V2
V3
hV4
VO1
VO2
PHC1
PHC2
MST
hMT
LO2
LO1
V3b
V3a
IPS0
IPS1
IPS2
IPS3
IPS4
IPS5
SPL1
FEF
correlation
0.2
0.1
0


ISFC 2D
Early Visual
Ventral
Dorsal
Parietal / Frontal
V1
V2
V3
hV4
VO1
VO2
PHC1
PHC2
MST
hMT
LO2
LO1
V3b
V3a
IPS0
IPS1
IPS2
IPS3
IPS4
IPS5
SPL1
FEF
correlation
0.2
0.1
0

B)

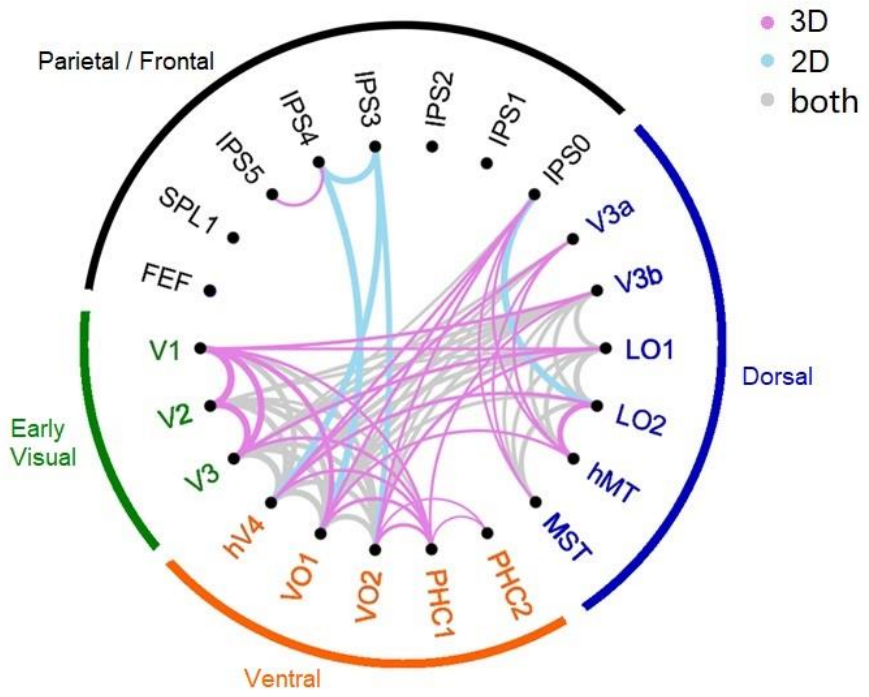
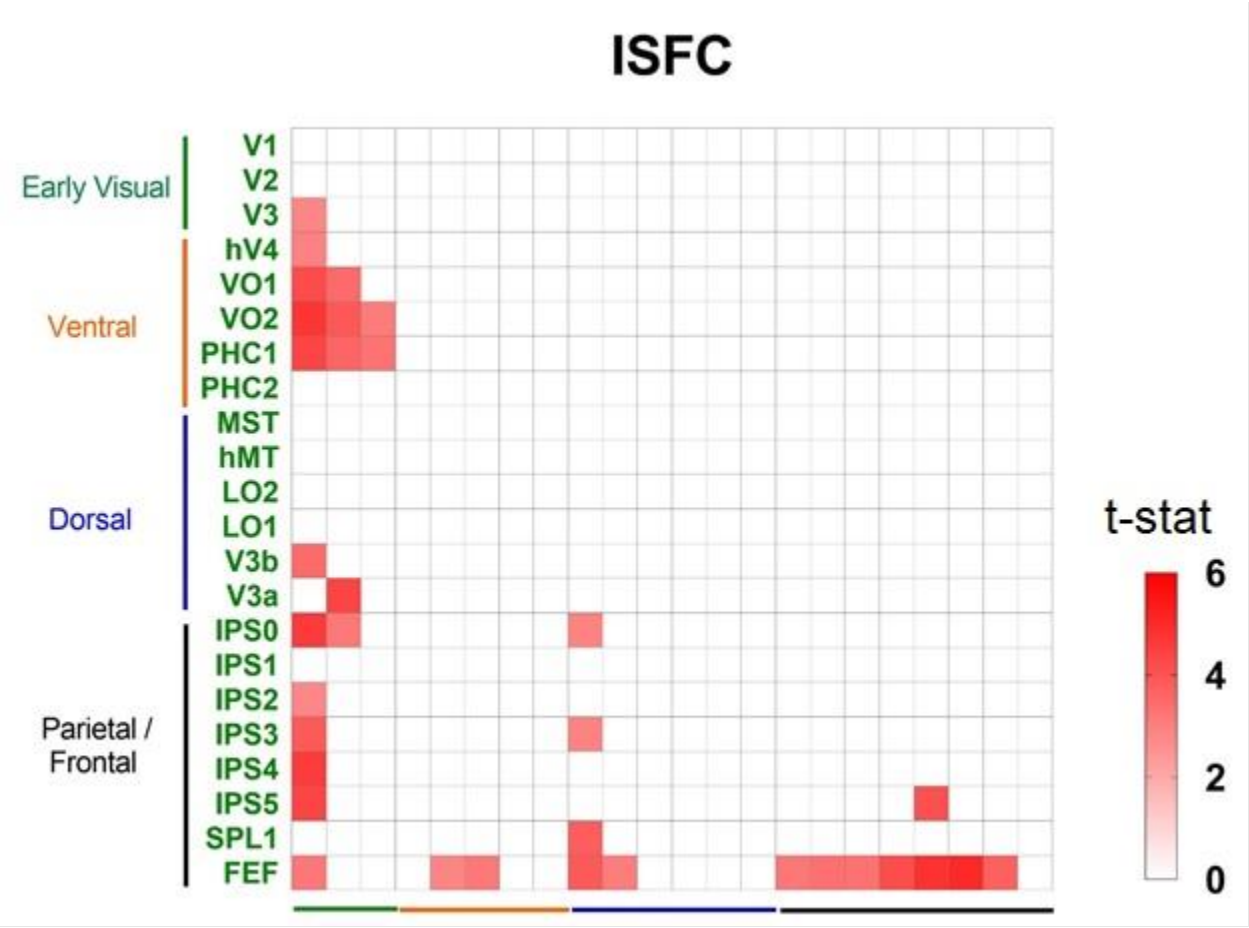


Figure 2. ISFC in 3D and 2D movie viewing. A) ISFC in 3D and 2D movie viewing. Correlation coefficients were tested for significance using an FDR error rate correction with q = 0.05. B) Detectable significant connections in 3D condition, 2D condition, and both.

Subject specific and across-subject network changes in 3D movie viewing

Functional connectivity (FC) and inter-subject functional connectivity (ISFC) changes between 3D and 2D conditions were calculated in 22 visual regions for 2 different movie clips (for FC, see supplementary Fig. S1. For ISFC, Fig. 3). Paired T-test was performed between the two viewing conditions for each connectivity network connection, under an FDR error rate control of q = 0.05. The FC network showed that 7.5% of its connection's strength had significant changes between viewing conditions. The ISFC network showed that 27.43% of its connection's strength had significant changes between viewing conditions, this result showed the ISFC networks were strengthened in 3D movie stimuli.

A)


ISFC
Early Visual
Ventral
Dorsal
Parietal / Frontal
V1
V2
V3
hV4
VO1
VO2
PHC1
PHC2
MST
hMT
LO2
LO1
V3b
V3a
IPS0
IPS1
IPS2
IPS3
IPS4
IPS5
SPL1
FEF
t-stat
6
4
2
0

B)

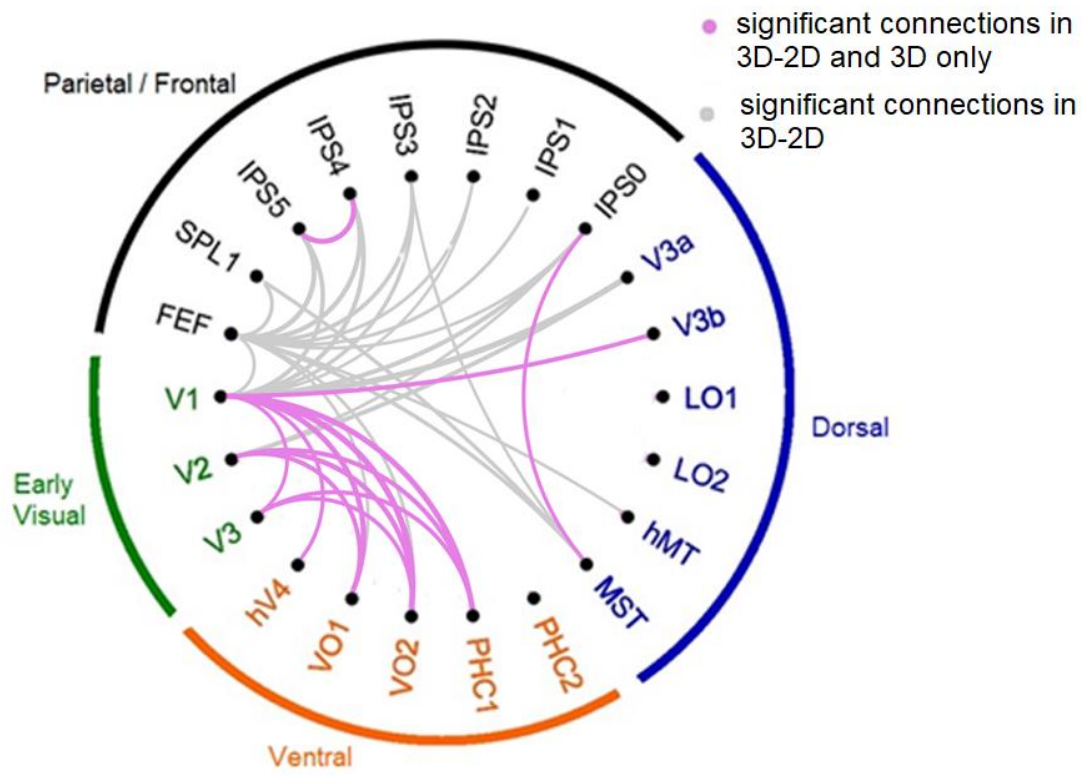


Figure 3. ISFC changes in 3D movie viewing. A) Paired student t-statistics showing connections with increased strength in 3D movie viewing compared to 2D movie viewing. B) ISFC connections strength increase. Pink, significant connections that were also present in 3D only ISFC networks. Gray, significant connections that were only present in the 3D vs. 2D paired test.

Classification of condition label using the connectivity network

FC from the within left hemisphere connections, within right hemisphere connections, and the inter-hemispheric connections demonstrated its weak ability to classify between 3D and 2D condition labels in an SVM algorithm (Fig. 4). The classification accuracy was 0.5392, 0.5931, and 0.5637, respectively. ISFC from each of the subject could be used to classify condition labels much more accurately. Both of within-hemispheric ISFC network and inter-hemispheric ISFC network showed classification accuracy above 0.85 (Fig. 4).

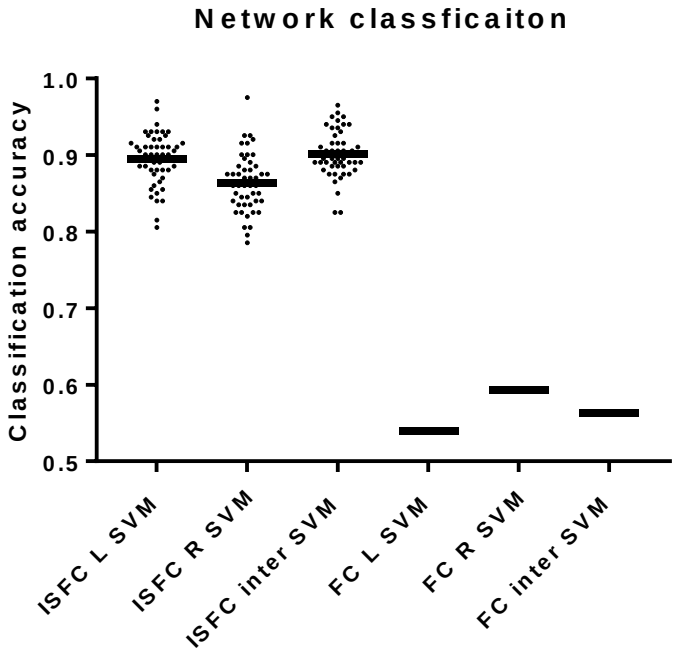
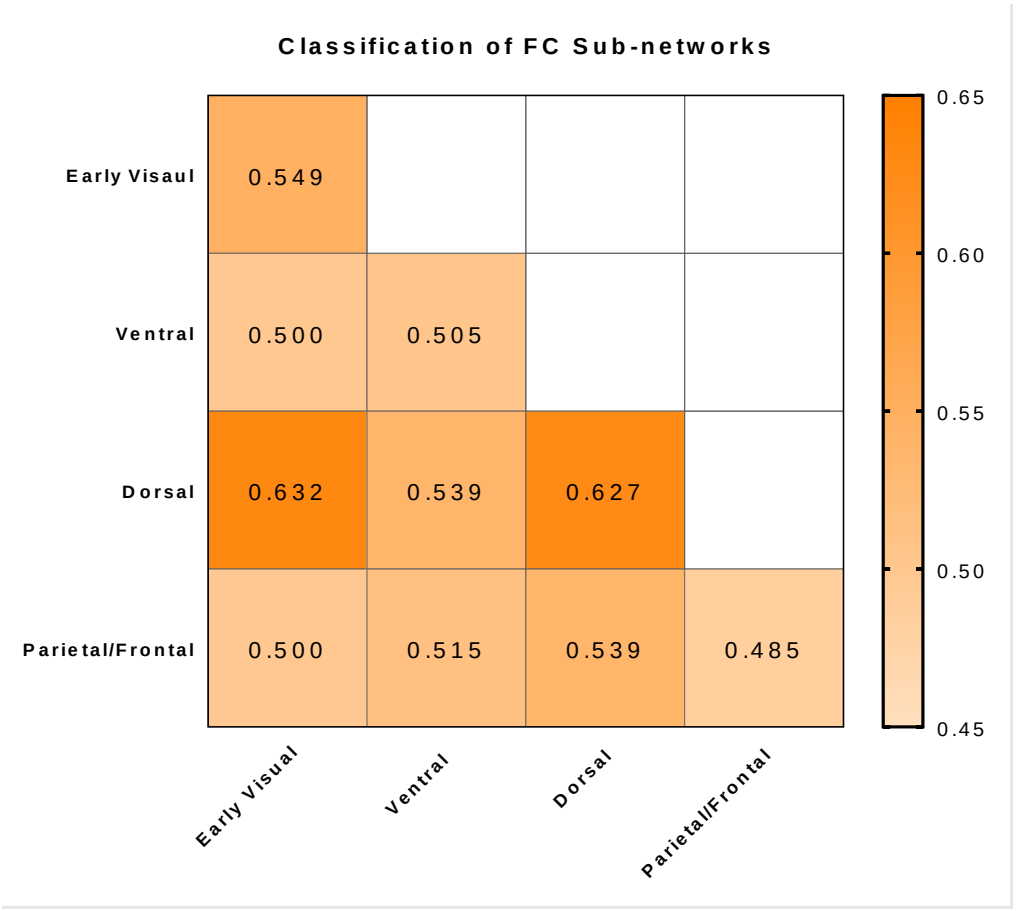


Figure 4. classification of movie labels based on FC and ISFC networks. L, left hemisphere connections. R, right hemisphere connections. Inter, interhemispheric connections. Classifications for ISFC networks were done in each subject (with subject pairs as repeats) and the accuracies were averaged. Black dots show the classification accuracies data points.

Classification using connectivity sub-networks

The FC and ISFC sub-networks between different visual streams could be used to classify the 2D vs. 3D viewing condition. The sub-networks in FC showed higher classification accuracy in dorsal visual areas to early visual areas connections, and dorsal visual areas to dorsal visual areas connections, with other connections close to 0.5 random chance level (Fig. 5A). All sub-networks in ISFC showed a higher classification accuracy than chance level. The sub-networks in ISFC that contained dorsal visual areas showed higher classification accuracy than other connections (Fig. 5B), which is consistent in the pattern showing in the FC results. The sub-networks in ISFC that contained ventral visual areas also demonstrated higher classification accuracy.

A)


Classification of FC Sub-networks
Early Visaul
Ventral
Dorsal
Parietal/Frontal
0.549
0.500
0.505
0.632
0.539
0.627
0.500
0.515
0.539
0.485
Early Visual
Ventral
Dorsal
Parietal/Frontal
0.65
0.60
0.55
0.50
0.45

B)

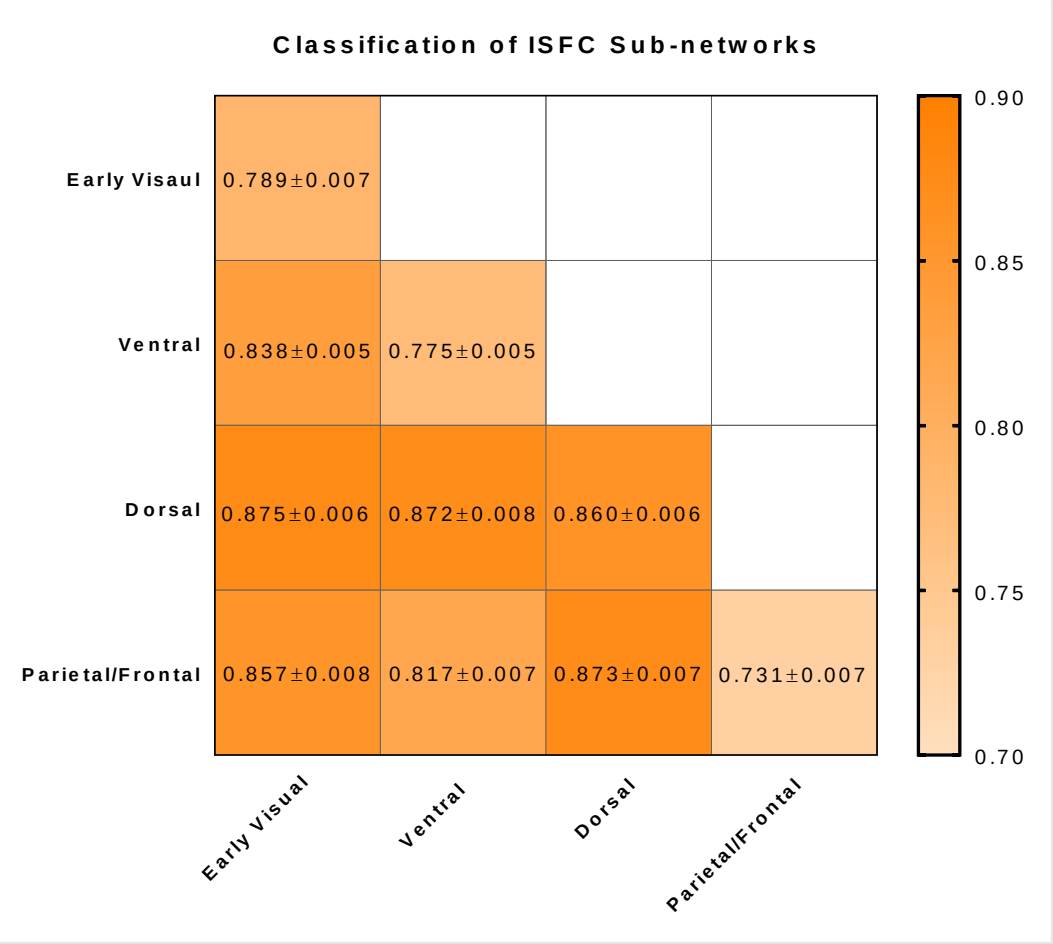


Figure 5. Classification of movie labels using sub-networks. A), Classification results for FC sub-networks. Early, ventral, dorsal, and parietal/frontal areas were picked based on the atlas (Wang et al 2015). The number in each column and each row showed the classification result of a sub-network in which the connections are between the column label and the row label. B), same as A), but using ISFC sub-networks. ISFC networks for each subjects contains multiple measures for every subject pair, thus can be used for classification. The numbers were the average classification results for all the subjects with the standard error of the mean.

Discussion

ISC as a tool to detect cortical response changes in 3D movie viewing

In this study, we performed a ISC procedure to measure the ISC changes in 3D movie viewing, compared to 2D movie viewing. We observed a significant ISC in the visual regions, but the changes

between the viewing conditions were not detected. These results suggested that the depth induced in 3D movies viewing would not result in detectable changes in individual vertex-wised results in naturalistic conditions. However, our data also showed that ISC can be different in conditions in terms of its pattern and distribution. This suggest that depth induced changes could be detected by other approaches.

Network changes induced by 3D movie viewing

Previous works have shown that stereopsis could induce changes of activities in networks of spatially dependent cortical regions (Gaebler et al 2014). In our data, we detected changes in connectivity between visual areas in 3D movie viewing condition in both of our movie clips. The ISFC networks showed more robust changes in its connections in 3D movie viewing. These findings were consistent with the previous finding that the intrinsic and noise signals were uncorrelated across subjects, and they were represented in a large part in FC networks, thus ISFC could be used to capture more between-subject commonalities (Simony et al 2016). This approach is especially valuable in 3D movie research, where the contents of the two movies (2D movie and 3D movie) are the same and will elicit similar cortical responses in the visual areas.

Decoding analysis using FC and ISFC networks

FC networks has been used recently to characterize different mental states. Our data supported the idea that FC networks could be used to identify movie viewing conditions. ISFC networks showed higher decoding accuracies, because within-subject decoding can be performed, in which the condition specific information could be preserved. Overall, we argue that the ISFC networks enhance the ability to detect condition-specific changes compared to FC networks in naturalistic stimuli settings.

Sub-networks relate to depth perception

We primarily focused on visual related areas, which were a sub-network for a whole brain FC. Our data suggested that decoding can be used as a powerful tool to identify sub-network connections

induced by binocular disparity induced in 3D movie viewing. Dorsal stream of the visual system, including V3a, has been characterized to be related to depth perception. This argument is supported by our data, which showed sub-networks that include dorsal visual areas had high classification accuracies. Specifically, we found that dorsal-to-dorsal and dorsal-to-early-visual connections to be most prominent connections. Interestingly, we also detect dorsal-ventral connections to be changed in 3D movie viewing when using ISFC but not FC networks.

Limitations and future directions

One of the limitations of the current studies is the number of movie clips been used. Future works that uses more diversifies movies are indicated. Our networks analysis was limited in the visual system. Future studies featuring whole-brain analysis are suggested. The tool we proposed in this study could also be used for future studies in clinical settings.

Reference

Anzai A, DeAngelis GC. 2010. Neural computations underlying depth perception. Curr Opin Neurobiol 20: 367-75

Backus BT, Fleet DJ, Parker AJ, Heeger DJ. 2001. Human cortical activity correlates with stereoscopic depth perception. J Neurophysiol 86: 2054-68

Cohen JD, Daw N, Engelhardt B, Hasson U, Li K, et al. 2017. Computational approaches to fMRI analysis. Nat Neurosci 20: 304-13

Cole MW, Reynolds JR, Power JD, Repovs G, Anticevic A, Braver TS. 2013. Multi-task connectivity reveals flexible hubs for adaptive task control. Nat Neurosci 16: 1348-55

Cottereau BR, McKee SP, Norcia AM. 2012. Bridging the gap: global disparity processing in the human visual cortex. J Neurophysiol 107: 2421-9

Cox RW. 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29: 162-73

Di X, Biswal BB. 2020. Intersubject consistent dynamic connectivity during natural vision revealed by functional MRI. Neuroimage: 116698

Gaebler M, Biessmann F, Lamke JP, Muller KR, Walter H, Hetzer S. 2014. Stereoscopic depth increases intersubject correlations of brain networks. Neuroimage 100: 427-34

Genovese CR, Lazar NA, Nichols T. 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage 15: 870-8

Goncalves NR, Ban H, Sanchez-Panchuelo RM, Francis ST, Schluppeck D, Welchman AE. 2015. 7 tesla FMRI reveals systematic functional organization for binocular disparity in dorsal visual cortex. J Neurosci 35: 3056-72

Hall H. 2009. Under the sea.

Henderson M, Vo V, Chunharas C, Sprague T, Serences J. 2019. Multivariate Analysis of BOLD Activation Patterns Recovers Graded Depth Representations in Human Visual and Parietal Cortex. eNeuro 6

Preston TJ, Li S, Kourtzi Z, Welchman AE. 2008. Multivoxel pattern selectivity for perceptually relevant binocular disparities in the human brain. J Neurosci 28: 11315-27

Saad ZS, Reynolds RC. 2012. Suma. Neuroimage 62: 768-73

Simony E, Honey CJ, Chen J, Lositsky O, Yeshurun Y, et al. 2016. Dynamic reconfiguration of the default mode network during narrative comprehension. Nat Commun 7: 12141

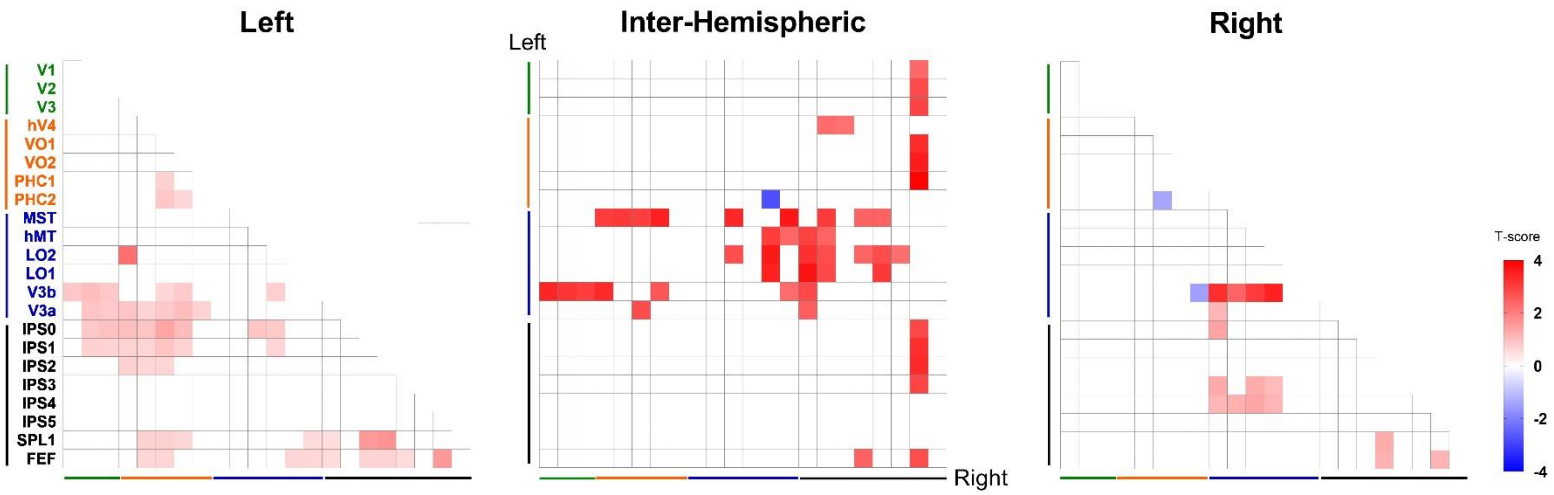
Vanderwal T, Eilbott J, Finn ES, Craddock RC, Turnbull A, Castellanos FX. 2017. Individual differences in functional connectivity during naturalistic viewing conditions. Neuroimage 157: 521-30

Wang L, Mruczek RE, Arcaro MJ, Kastner S. 2015. Probabilistic Maps of Visual Topography in Human Cortex. Cereb Cortex 25: 3911-31

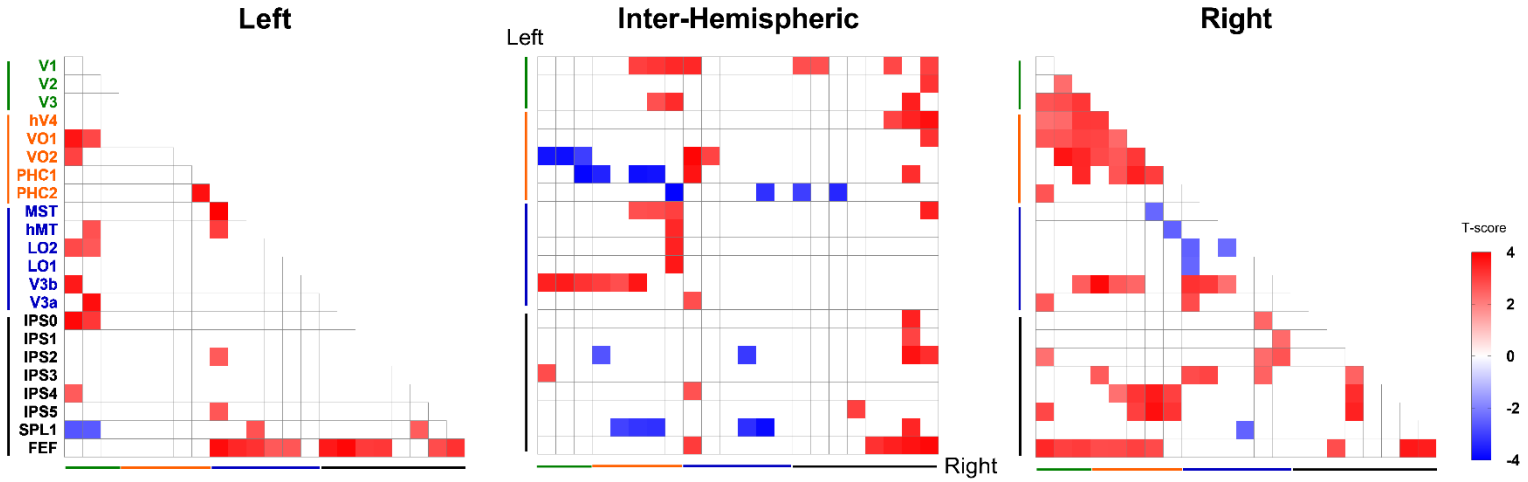
Supplementary

Supplementary Figure 1, FC and ISFC changes in each hemisphere and interhemispheric connections

A) FC changes in 3D vs 2D movie viewing



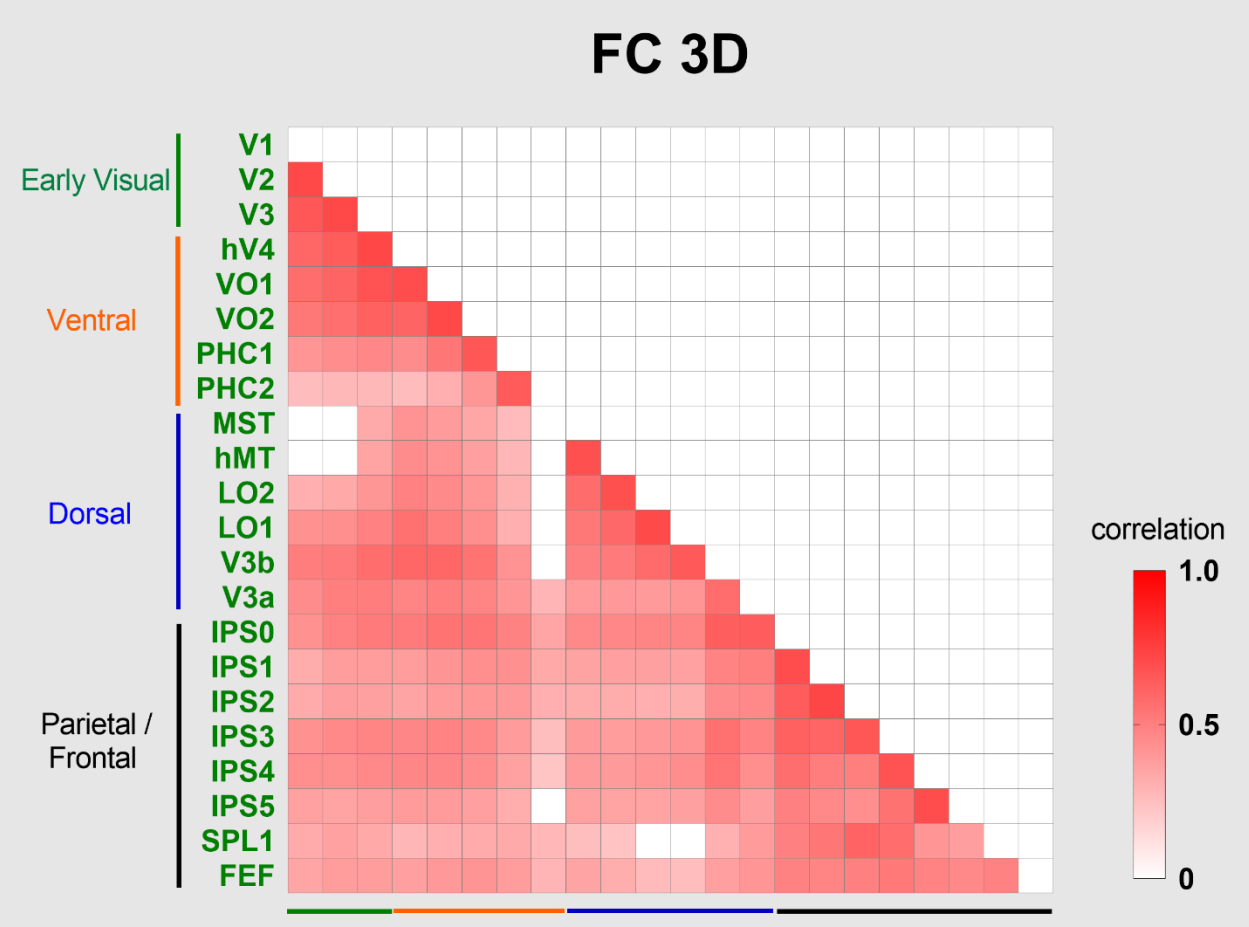
B) ISFC changes in 3D vs 2D movie viewing

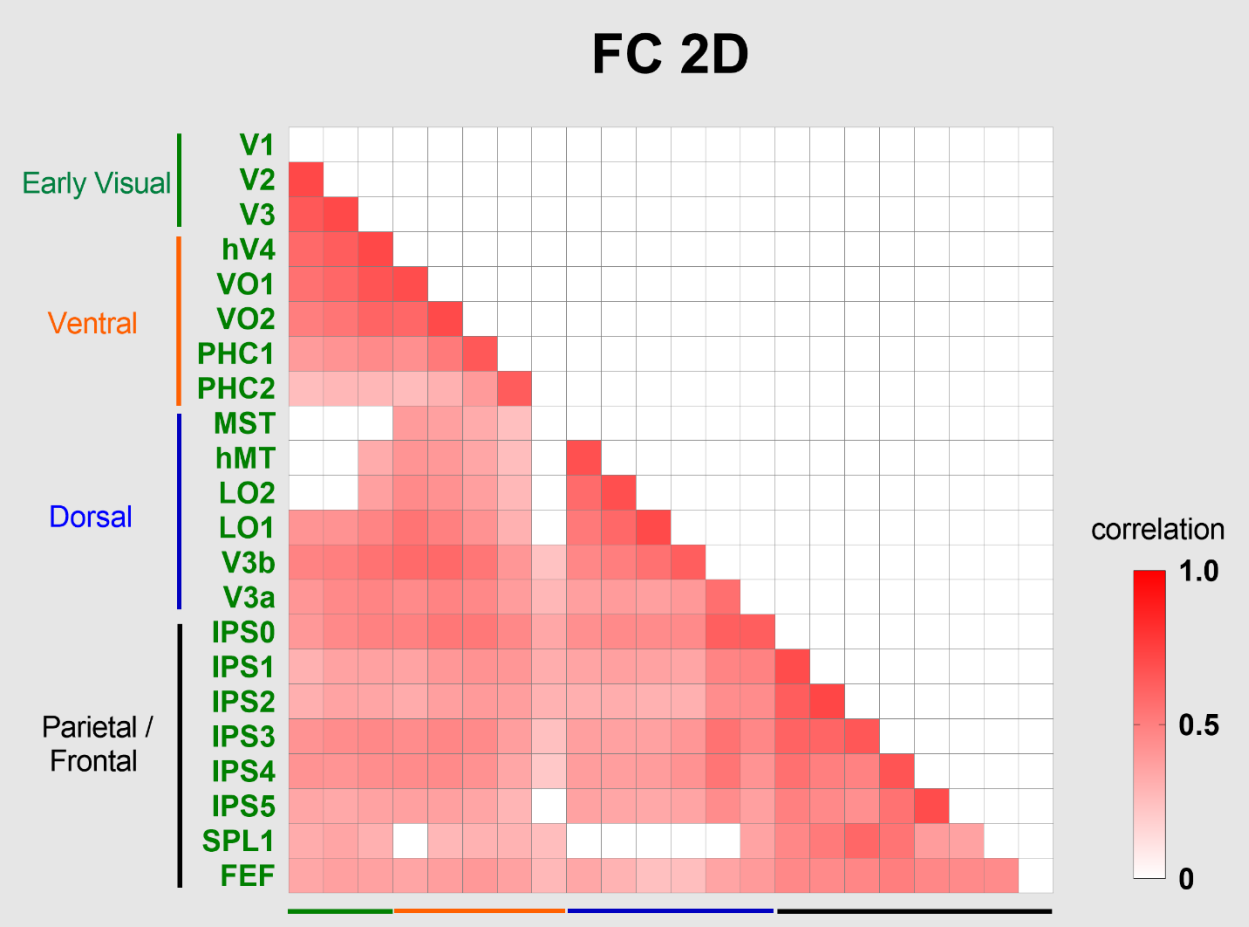


Functional connectivity changes in 3D movie viewing. A), functional connectivity changes between movie viewing in 3D versus 2D. Two blue boxes show the 22 visual areas in left and right hemisphere. A yellow box showed interhemispheric connectivity. The color showed t-statistics in a paired t-test for each connection.

The matrices only show significant connections ($q > 0.05$). B) inter-subject functional connectivity changes between movie viewing in 3D versus 2D.

Supplementary Figure 2. Functional connectivity in 3D and 2D movie viewing


FC 3D
Early Visual
Ventral
Dorsal
Parietal / Frontal
V1
V2
V3
hV4
VO1
VO2
PHC1
PHC2
MST
hMT
LO2
LO1
V3b
V3a
IPS0
IPS1
IPS2
IPS3
IPS4
IPS5
SPL1
FEF
correlation
1.0
0.5
0


FC 2D
Early Visual
Ventral
Dorsal
Parietal / Frontal
V1
V2
V3
hV4
VO1
VO2
PHC1
PHC2
MST
hMT
LO2
LO1
V3b
V3a
IPS0
IPS1
IPS2
IPS3
IPS4
IPS5
SPL1
FEF
correlation
1.0
0.5
0

Chapter 4: The dynamics of depth cue invariance in 3-D object recognition

In the previous chapters, I presented how scene recognition is related to depth perception in a dynamic naturalistic setting. However, the perception of depth also includes monocular cues. If the depth perception contributes to the recognition of object, the monocular cues should be involved in the process of object information. Thus, it is crucial to understand how the integration of different depth cue information is related 3D object perception. In the following chapter, I evaluated the dynamics of invariance object recognition using different objects including faces. I investigated the possibility of single depth cue object information existence and timing, and the integration of individual depth cue information process. I concluded that the information was transferable between depth cues after the initial stage in which the depth cues are processed separately in the cortex.

The dynamics of depth cue invariance in 3-D face recognition

Abstract

In the real world, human recognize faces by integration of information from different depth cues, including shading, texture, and structure from motion (SFM), to form stereopsis perception (Dehmoobadsharifabadi & Farivar 2016). Different depth cue information is processed in both ventral and dorsal visual pathways, and later integrated to form a mental representation of face. Understanding how and when brain integrates information from different depth cues are critical for deciphering stereopsis face recognition processes. Two competing hypothesis arises: 1) ventral and dorsal visual information from different visual cues are integrated in the 'face area' and represented in the same neural population in one stage; 2) different depth cues represent different face representations in different neural populations in the 'face area', and later, combined and integrated to form face perception (two stage hypothesis). To test which hypothesis is true, we used MEG to measure the precise timing of face recognition. Face and control objects were presented to the subjects using three different cues (shaded, texture and SFM). Multivariate pattern analysis, which involved a machine learning process to classify which one of the input stimuli is associated with the MEG recording data, was used to decode the category of stimuli in every time points. This technique showed that the precise timing of mental representation can be revealed by decoding accuracy changes over time. We could also test if the decoding model we trained from one of the models could accurately predict the input from other depth cues, to estimate the information been shared between depth cues. These results provided evidence about the existence and timing of depth-cue invariance face perception.

Introduction

3D objects can be recognized by the visual system across different depth cues, including shading, texture, and structure from motion (SFM). Previous research has shown that the visual system processes different depth cues in different visual streams. The shading and texture cues are static cues, which are processed in the ventral stream (Merigan 2000). The SFM cues are suggested to be processed in the dorsal visual stream (Kamitani & Tong 2006, Zeki 1991). In the nature, 3D object recognition problems include multiple depth cues, and the visual system is combining them to form a single percept. The 3D objects recognition is considered an integration process of these different processes (Dovencioglu et al 2013, Farivar 2009, Konen & Kastner 2008). In a previous study of our group, we illustrated that showing the stimuli in shading could result in an adaptation effect in the cortical responses when subjects were shown objects from other depth cues (texture and SFM), suggesting that the same object showed in different depth cues could invoke same neural population activities, which is the sign of integration of depth cue information from dorsal and ventral stream. Two different hypotheses emerge around how this integration happens. The first is that the object information from each depth cue is encoded in the same group of neural population. This is the integration hypothesis. Those competing hypothesis describe that the information about 3D objects presented in different depth cues are separated in multiple independent neural populations. Only in the later stage when these neural signals for the object are combined into a single representation. This is the independence hypothesis (Akhavein 2017, Dovencioglu et al 2013). In this paper, we investigate how object recognition in individual depth cues are integrated regarding to this depth cue invariance problem by using time-resolved MEG multivariate pattern analysis.

MEG multi-variate pattern analysis (MVPA) has shown to be effective in providing detailed temporal dynamics of neural representations. MVPA technique originates from fMRI research, where decoders are trained to classify the stimulus condition labels by machine learning algorithms, and then

to assess whether the stimulus information is present in the neural data. This technique has proven to be useful in many applications including electrophysiological recordings (Meyers et al 2008) , EEG data (Ratcliff et al 2009), and MEG data (Wardle et al 2016). Recent development of combining MEG and MVPA enables the method of fitting classifiers at each of the time samples. Using this method, the precise timing of whether the neural signal represents object information can be determined (Isik et al 2014, King & Dehaene 2014). Further, an important aspect of this technique allows each of the classifiers being trained at time *t* and been tested at another time point t'. This temporal generalization can reveal how information propagate dynamically, with the onset time and duration of information flow quantified (Marti et al 2015). Moreover, the generalization can also be applied to different experimental conditions. Here, we applied time resolved generalization to the 3D object recognition depth cue invariance MEG data. We identified successful decoding in the object recognition from shading, texture, and SFM depth cue conditions. We then tested if the generalization can be found across differed depth cues. In this way, we found that the depth-cue-specific object information can be observed before the generalization, supporting the hypothesis that the depth information for objects is characterized independently before integration.

Methods

Participants

Twelve healthy participants (8 female, 4 male) with normal vision were recruited for the experiments. Written consent forms for publication from all participants were obtained. All procedures were approved by the Research Ethics Board of the Montreal Neurological Institute.

Stimuli

In each trial, two stimulus epochs, termed S1 and S2, were presented in succession with a short inter-stimulus interval (ISI). The S1 stimulus consisted of a face, a chair, or a control face (cFace) defined

by one of the three depth cues (shading, texture, and SFM1). The S2 stimulus was always a shaded face whose identity varied on each trial, but the data of the S2 stimuli are not analyzed in this study. In 10 percent of the trials, S2 was a target shaded face that the subjects had to detect. The purpose of the task was to increase the subject's attention. The data from the target-present trials were excluded from all analysis. The S1 epoch lasted 1133ms (68 refreshes of the screen refreshing at 60Hz) and the duration of S2 and ISI were set to 200ms (12 refreshes) and 100ms (6 refreshes) respectively. There was a random pause of 1200ms ±200 between trials (ITI) and a red fixation point was presented at the center of the screen during the entire experimental run. Each run consisted of 10 pairs of conditions (3 depth cues * 3 object category and 1 target-present condition which was removed from analysis), and each condition was repeated 20 times (Figure 1). There were 5 recording sessions each lasting about 10 minutes.

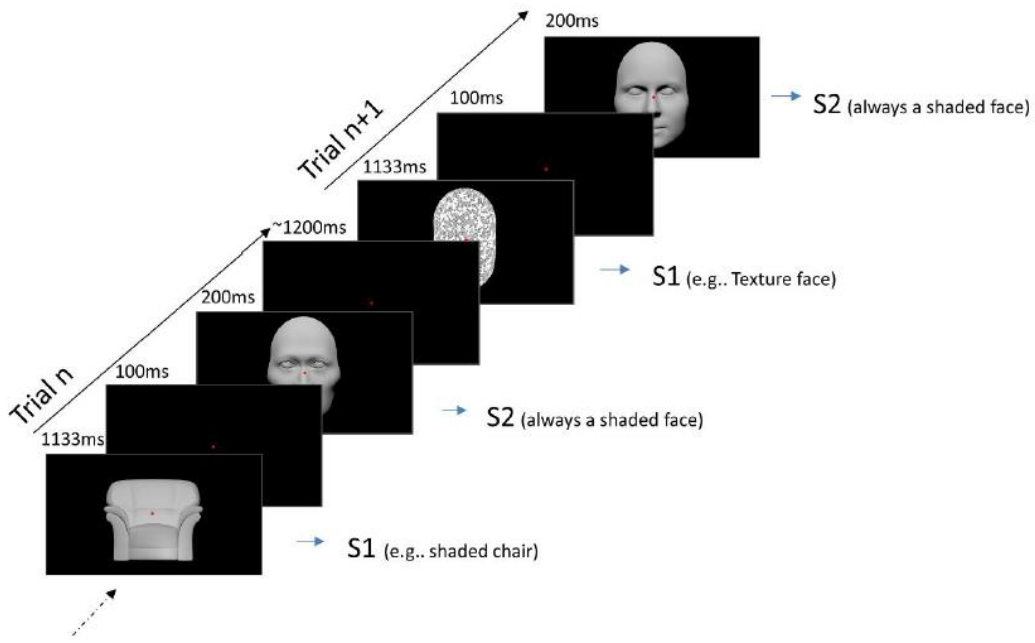


Figure 1. The experimental design. The experiment procedure consists of a presentation of S1 stimuli and S2 stimuli (data not used in this study) in each trial, with a randomized black screen inter-trial interval. The S1 stimuli were randomly selected from three difference object class (face, chair, and cFace) and three experiment conditions (shaded, texture, and SFM).

The stimuli depicted three object categories: faces, chairs and cFaces. Forty-one synthetic 3-D facial surfaces (20× S1, 20 × S2 and one target) were generated with random identities using FaceGen Modeller 3.5 Software. Chairs were obtained from online open-access libraries in .3ds format and later modified in Autodesk 3ds Studio Max 2013. Twenty chairs containing smooth surfaces were selectively chosen for this study to facilitate the perception of changes in depth. Twenty cFaces were generated by removing the internal facial features of the face but keeping the contour of the face intact using Autodesk 3ds Studio Max—the internal features of the face were removed from the facial surface and the final surfaces were smoothed. All object surfaces were later rendered to generate isolated depth cues as described below and following our previous work (Dehmoobadsharifabadi & Farivar, 2016; Farivar et al., 2009).

Shading: All the textures were removed from the surface and lights with 45° angle from the horizon was projected onto the surface of the object. The frontal view of the face was rendered in orthographic projection to avoid perspective information, to make this consistent across all conditions.

Texture stimuli: Dot textures were added to the facial surface. 100% self-luminance was applied to remove shading and shadows and the final surfaces were rendered with orthographic projections to remove perspective. This process resulted in 3-D surfaces that were defined solely by texture gradients.

SFM: Twenty-four thousand dots were projected onto the surface of the objects in spatially uniform random positions. The object rotated 0.5 degrees each frame. The movement range in depth (around the vertical axis) was from -4 to +4 degrees. For each frame, the dots were projected back to the 2-D image plane. The 2-D dot density was calculated, with which were reshuffled between high-

density and low-density regions to ensure uniform dot density distribution throughout the presentation. On average, 200 dots were shuffled on each frame to compensate for local density changes caused by the 0.5° change in depth. To increase spatial sampling of the object and improve the quality of the SFM stimuli, we generated four independently estimated dot positions over time and overlaid them for each stimulus. This resulted in a variety of dot luminance intensities but more spatial samples than simple white-on-black.

All stimuli were presented on an LG 3-D LED Widescreen monitor (D2342PY, height: 13.5”) placed approximately 1.7 meters away from the subject’s head in the MEG room which covered 9.6° visual angle.

The used stimuli were the same as in a previous study from our group (Akhavein et al 2018).

MEG data acquisition

MEG data were recorded using a 275 channel CTF/VSM instrument with superconducting quantum interference device (SQUID) based axial gradiometers. Magnetic brain activity was digitized at the sampling rate of 2400 Hz and using an anti-aliasing low pass filter of 600Hz for the time of recording. Additional EEG signals were recorded using bipolar derivations. Two electrodes were placed on the torso for ECG cardiac activity. Two electrodes were placed one above and one below the left eye to capture EOG blink activity. Two electrodes were placed one on either side of the eyes to capture EOG saccade activity. In addition, a ground electrode was placed on the left shoulder. Head position tracking coils were placed on the subject and their locations were digitized using a Fastrak Polhemus device. Additional scalp points were also recorded to capture the subject's head-shape for co-registration with the T1 anatomical scan. A photodiode was placed on the back side of the screen to record stimulus triggers for precise timing of visual presentation and recorded via the MEG system.

Preprocessing

We used Brainstorm 3.1 (Tadel et al 2011b) for preprocessing of the MEG signal. The recordings were band-pass filtered at 0.8-200Hz, and a notch filter was applied to remove 60Hz power line contamination. Bad channels were identified based on power spectrum density (Welch filter) from the recordings and eliminated from all further analysis. Signal-Space Projection (SSP) was used to remove the artifacts generated by eye-blinks and heartbeats based on EOG and ECG activity recorded during the experiment. Standard independent component analysis (ICA) with 60 components using the infomax algorithm was used to detect the artifacts generated by the noise of the LCD screen and the corresponding components were removed from the signals. Bad segments and trials with noisy recordings due to movements or large alpha wave amplitudes were eliminated from further analysis. All trials were separated in a time window of -0.2s to 1s (with 0s been the time of stimulus onset). Each condition contained ~100 trials.

Hilbert transform and band-pass filter

Bandpass filtering followed by Hilbert transformation were used to generate time-dependent frequency power in several specified frequency bands. Six different bands were used. These bands are delta (2-4 Hz), theta (5-7 Hz), alpha (8-12 Hz), beta (15-29 Hz), gamma1 (30-59 Hz), and gamma2 (61-90 Hz) (Baillet 2017, Cohen 2014). This approach will increase the signal-to-noise ratio within a band by provides an concentrated time-dependent frequency response profile(Muller et al 1996). We used MEG/EEG data processing software, *Brainstrom* (Tadel et al 2011a). The frequency boundaries were non-overlapping integers. Different time-dependent frequency power responses were used to dynamic decoding analysis using the same methods below.

MEG sensor decoding and onset of significance

To improve the signal to noise ratio, we randomly grouped four trials from each condition from the same run and calculated the mean of those four trials to form a new pseudo trial. We randomized the

grouping patterns 20 times to prevent bias. Further, we used a sliding window approach using a window length of 10ms to average the time points in the MEG data to improve the signal to noise ratio.

For each of the two different conditions, we extracted the MEG sensor data for each time point in the pseudo trials. Linear discriminant analysis (LDA) using MATLAB was carried out for each time point and leave-one-out cross-validation was performed for all pseudo trials. Classification accuracy was determined by the portion of the pseudo trials whose object labels (face, chair, or cFace) were successfully classified by the LDA model. We reported the average of classification accuracy across time, with the SEM determined by the subject-wise variation.

To determine the onset of significance time, we calculate the t-statistics of decoding accuracy at each time point by using the mean and standard deviation from cross-validation folds for each subject. We set the significant p-value to be 0.05, and used multiple comparison Bonferroni method (Genovese et al 2002), with the number of comparison equals to the number of time points. To avoid outliers, the first time point with another consecutive significant time point is determined to be the onset of significance time. The peak of significance levels was determined by the largest t-statistics in the time window.

Temporal and cross-cue generalization

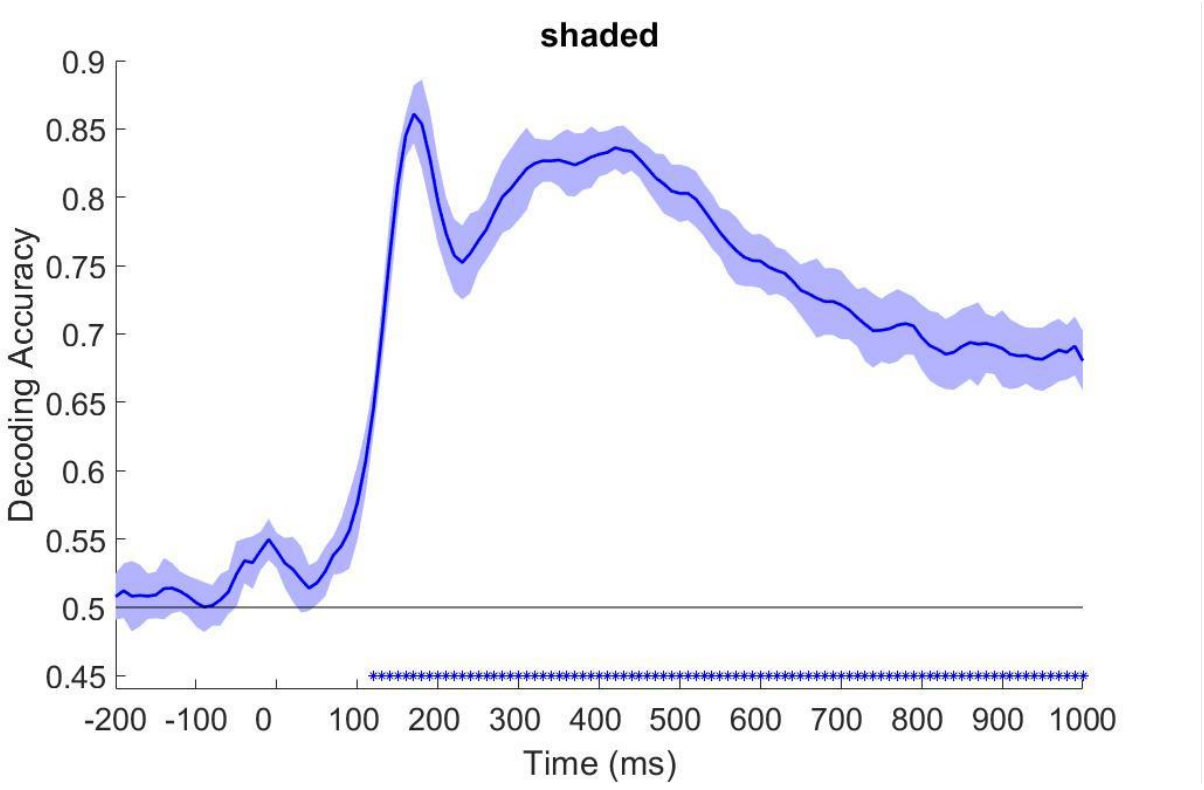
To generalize the 3D object representation, we used the MEG data to test temporal and cross-cue generalization. To investigate the temporal generalization, we trained the LDA model based on pseudo trials from one time point and tested the object type label using other time points from the same pseudo trial. We cross-validated the data using leave-one-pseudo-trial-out method. The cross-cue generalization was carried out using the same method, with the decoding model been trained based on data acquired from one 3-D cue the same model and tested using the data acquired from another 3-D cue.

Results

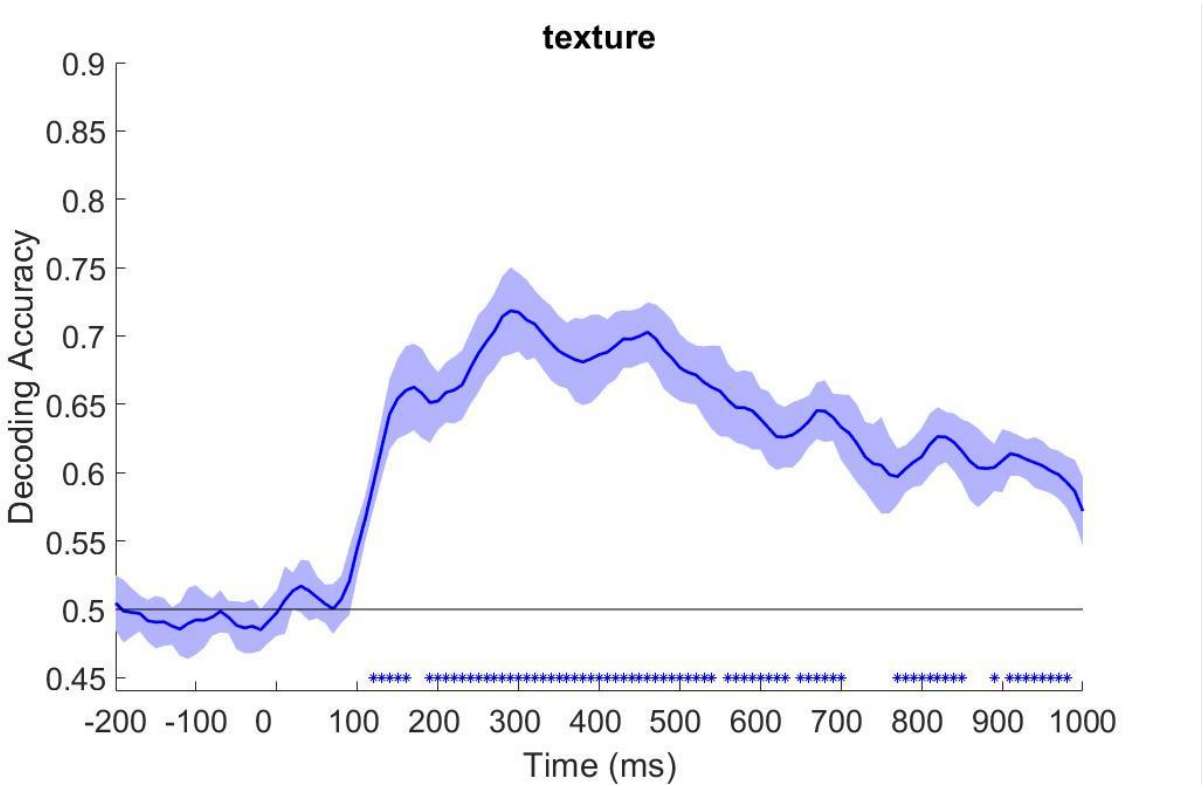
Decoding accuracy

We calculated the decoding accuracy between object types based on linear classifiers. The analysis was done for each time window in 10 ms sliding time window. Figure 2A shows the decoding accuracy between face and chair in three different depth cues. The decoding accuracy was averaged across all subjects, and the shaded area depicted SEM from 12 subjects' data. Shaded face vs. shaded chair showed the highest decoding accuracy, with a peak value of around 85% at 170 ms. The texture cue and SFM cue showed lower level and longer onset time of decoding accuracy. Significance level in each condition is Bonferroni corrected, and the significant decoding time point is marked with a star.

A)



B)


texture
Decoding Accuracy
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
-200 -100 0 100 200 300 400 500 600 700 800 900 1000
Time (ms)

C)

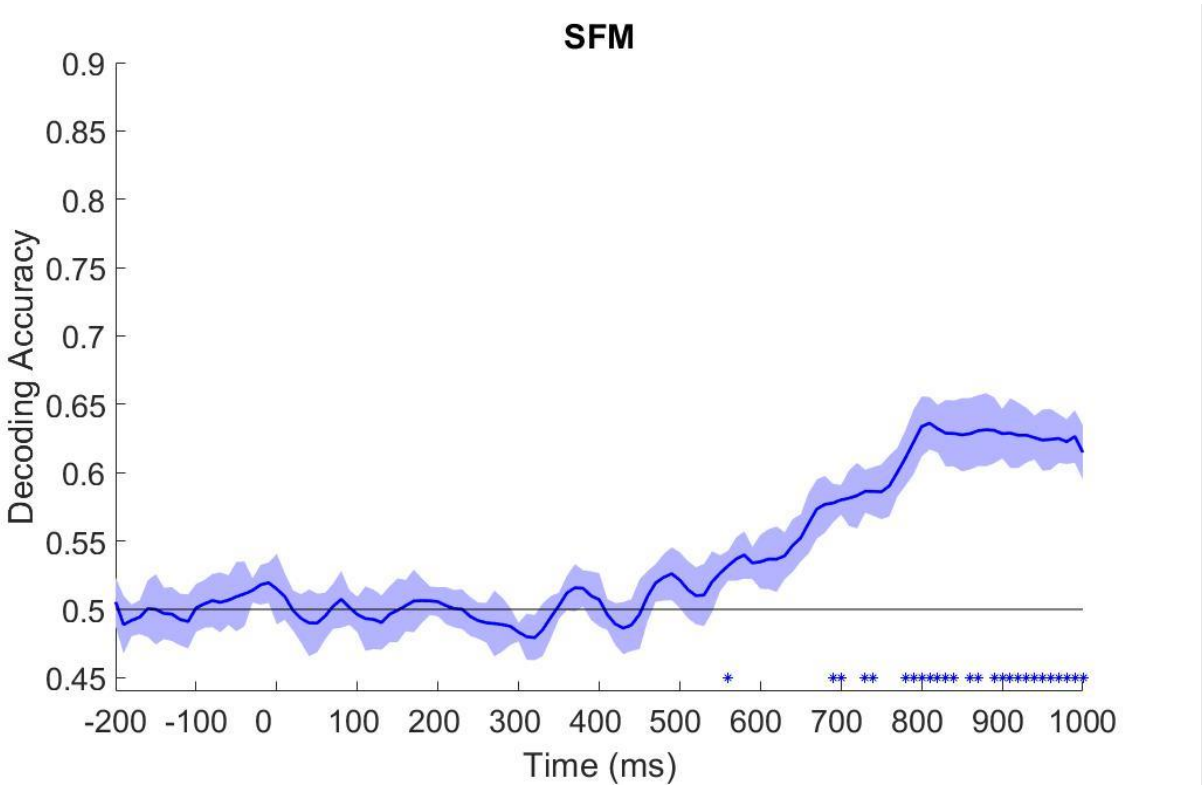


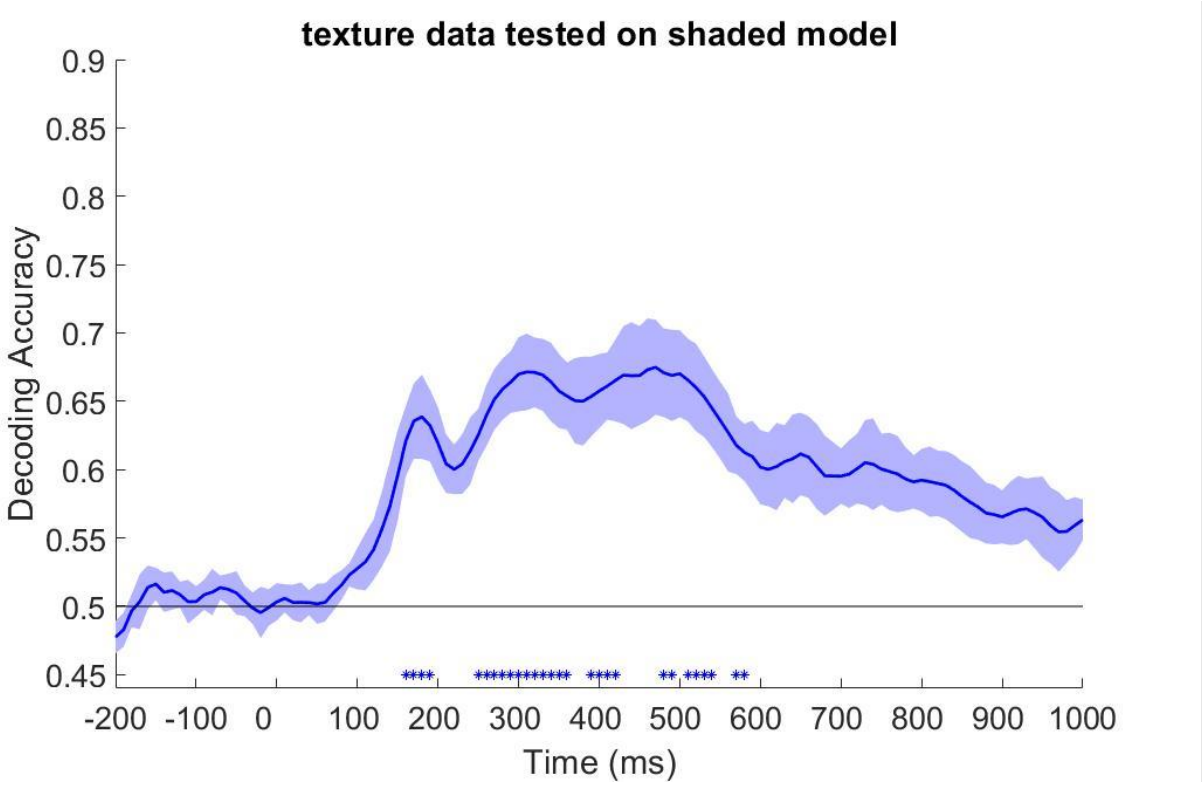
Figure 2. Time resolved decoding of depth-cue-specific object recognition (face vs. chair). 0s on the time axis depict stimuli onset. A 0.5 solid black line on the decoding accuracy axis shows the chance level. A solid blue line shows the average time resolved decoding accuracies across the subjects. Shaded blue area shows the standard errors of the mean. Blue stars below the 0.5 black line showed the significant time point, with $p < 0.05$. A), shaded stimuli. B), texture stimuli. C), structure from motion (SFM) stimuli.

Generalization across depth cues

To investigate the object information preserved in the decoding pattern from one 3D cue to another 3D cue, we performed time-resolved decoding generalization analysis. We generalized the decoding model by using the data from one 3D cue to train the decoder and test the performance on independent testing data from another 3D cue at the same time points for each subject. The training-testing pairs we show in the Fig. 3 were shade-texture and shade-SFM decoder performances on face vs. chair categorization tasks. Results from other training-testing pairs are presented in the supplementary

figures (Fig X). The results show that in the shade-texture model, the decoding performance is significant across all the subjects at around 100-200 ms after the stimulus is shown, and then rise to a peak at around 300-500 ms. The significant decoding information faded away after 600 ms. In shade-SFM models, the decoding performance started to show significance at around 700-800ms.

A)



B)

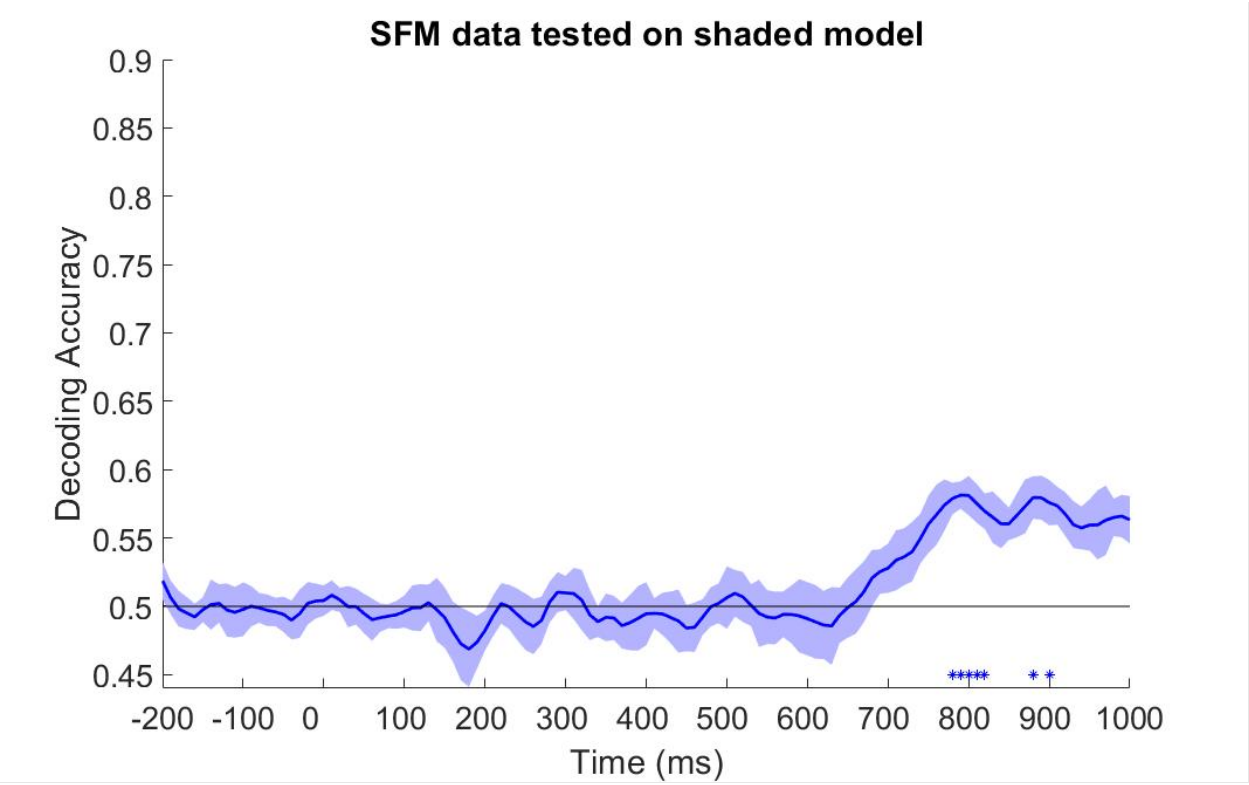


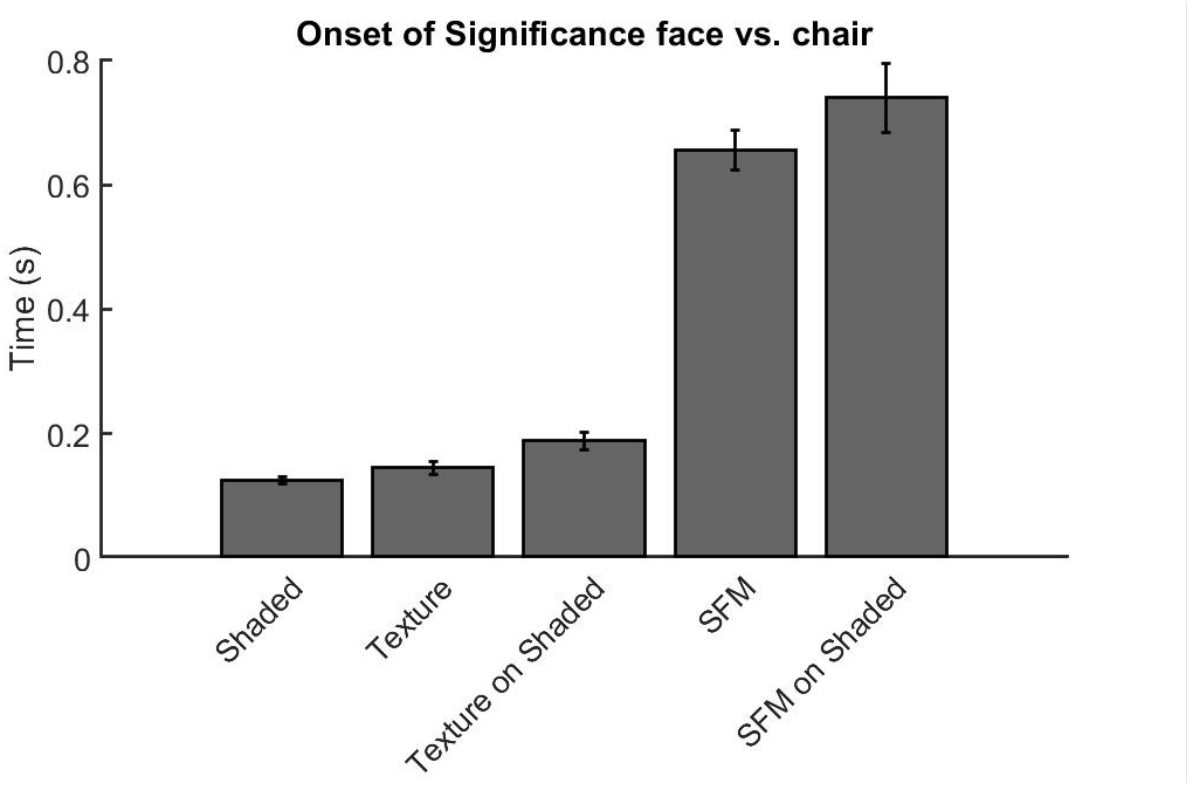
Figure 3. Cross-cue time-resolved decoding (face vs chair). The models were trained based on the shaded data and tested on A) texture data, and B) SFM data. 0s on the time axis depict stimuli onset. A 0.5 solid black line on the decoding accuracy axis shows the chance level. A solid blue line shows the average time resolved decoding accuracies across the subjects. Shaded blue area shows the standard errors of the mean. Blue stars below the 0.5 black line showed the significant time point, with $p < 0.05$.

Onset of significance and Peak decoding accuracy time

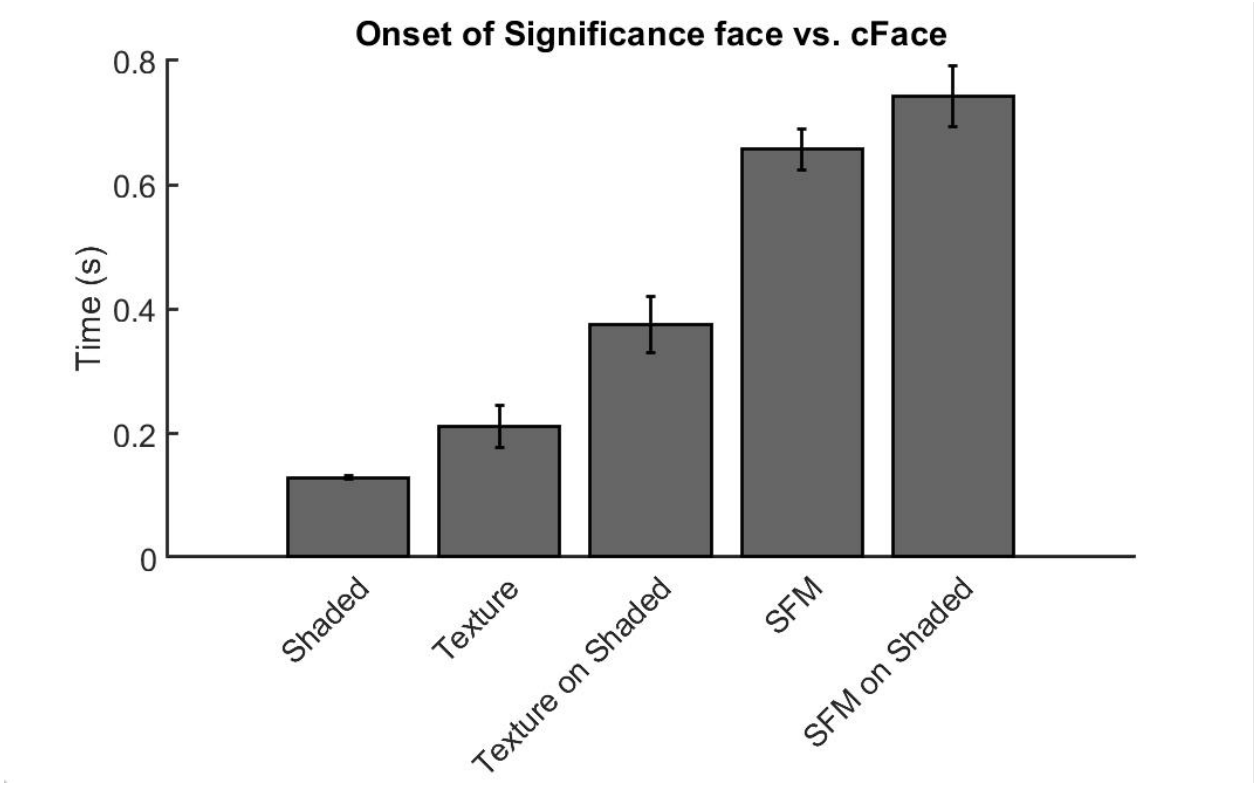
The onset of significance times for the decoder to successfully predict the labels in the testing data were calculated in each depth cue condition and in cross-depth-cue generalization decoding models (Figure 4A and B). The shaded depth cue stimulus and the texture depth cue stimulus showed significant decoding onset after ~100ms to ~200ms, which is consistent with previous findings using the face image vs. object images (Wardle et al 2016). The SFM depth cue stimulus showed later time values for onset of significance, indicating that the object recognition tasks were harder for the participants in our

experiment conditions. Interestingly, the cross-depth cue models (trained on shaded tested on texture or SFM) showed significant later time values for the onset of significant decoding accuracy than its single cue counterparts (trained and tested on the same cue, either texture or SFM). The peak decoding accuracy times were also shown in each depth cue conditions and cross-depth-cue generalization models. The difference peak time between the cross-depth-cue models and the single cue counterparts are more similar. These results suggest that the object information is processed in each condition significant earlier than generalized information, but at the maximum decoding performance time, the model contains information that can be generalized across depth cues.

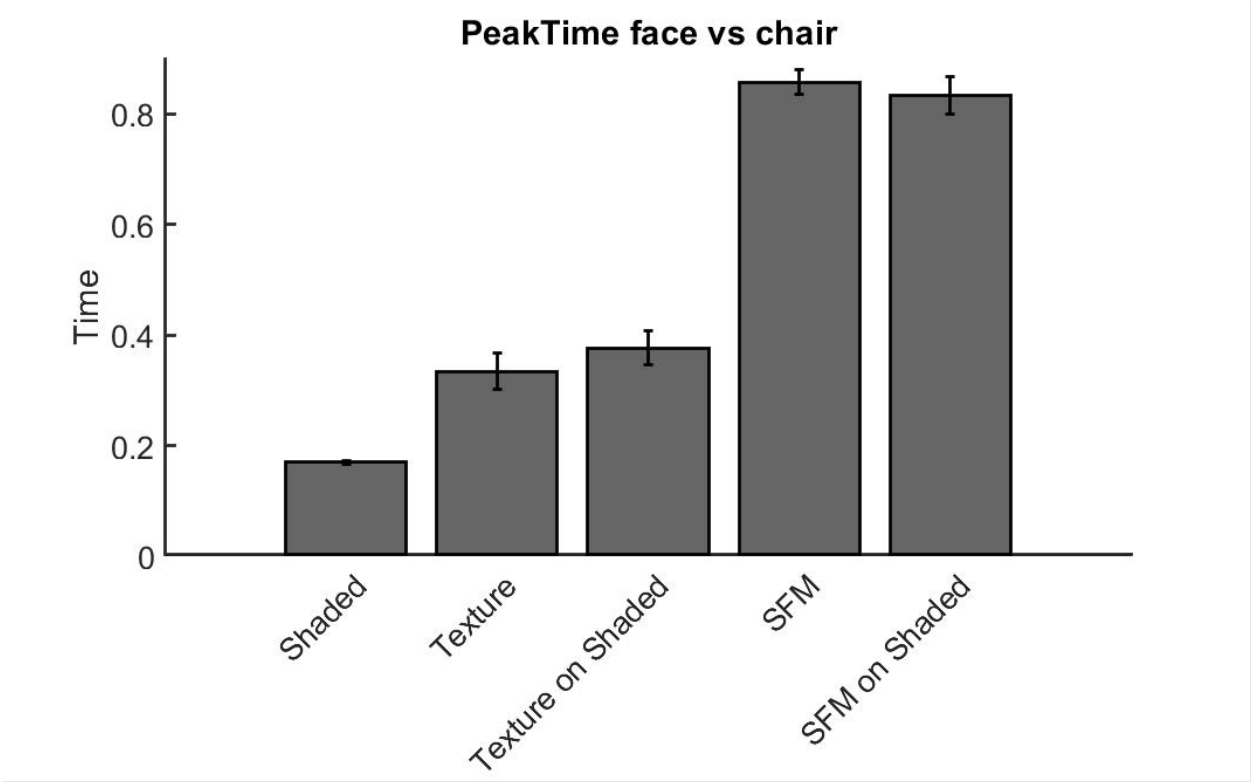
A)



B)


Onset of Significance face vs. cFace
Time (s)
0.8
0.6
0.4
0.2
0
Shaded
Texture
Texture on Shaded
SFM
SFM on Shaded

C)


PeakTime face vs chair
Time
0.8
0.6
0.4
0.2
0
Shaded
Texture
Texture on Shaded
SFM
SFM on Shaded

D)

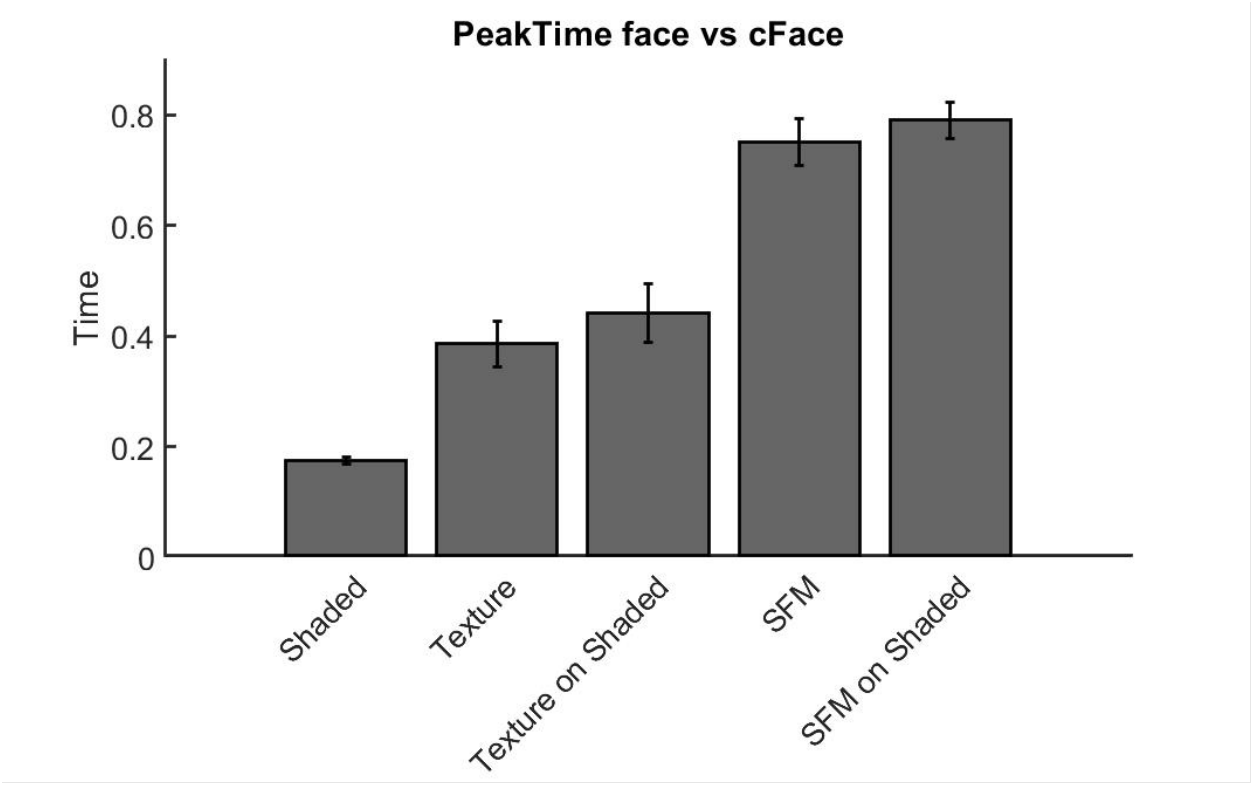


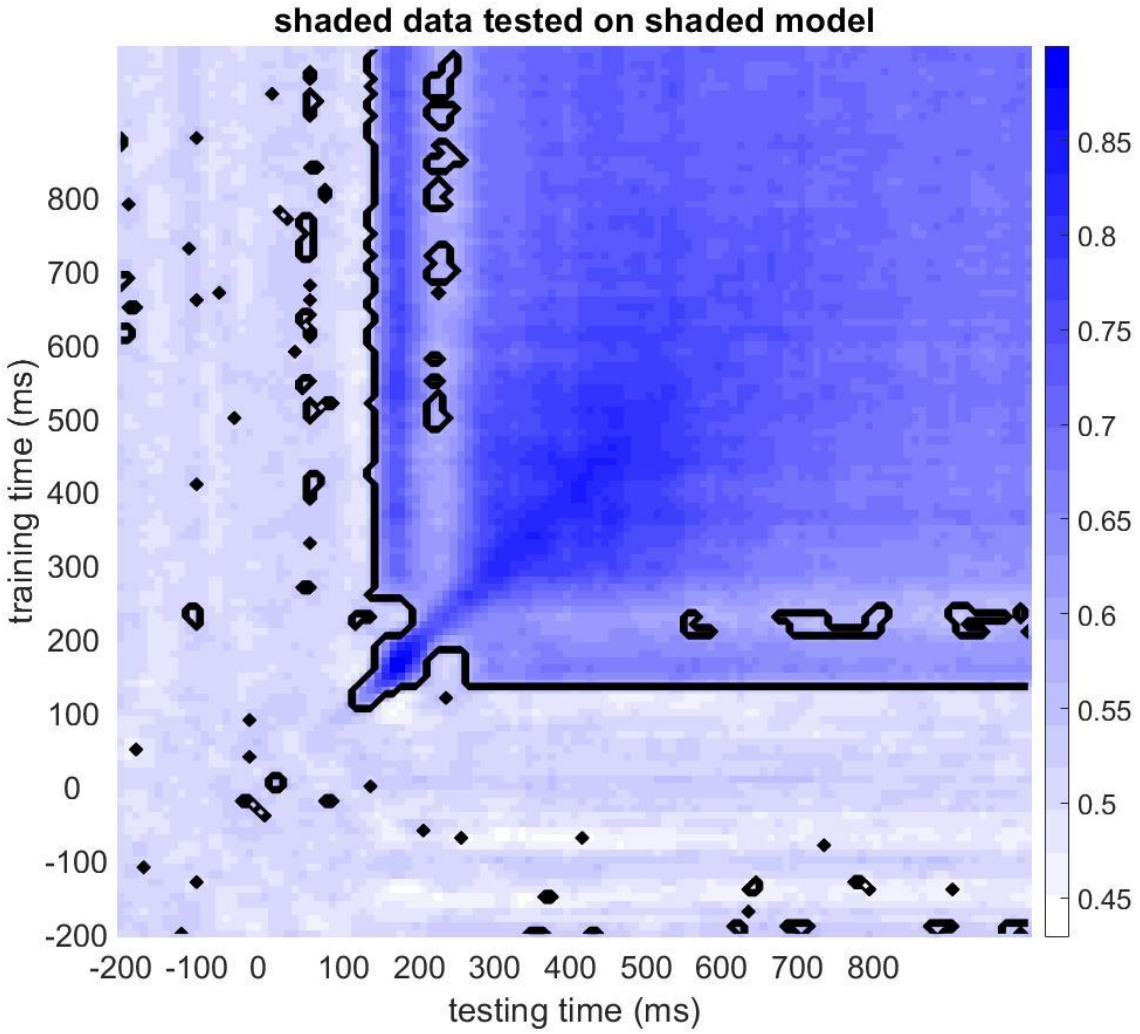
Figure 4. The onset of significance and the peak significance time in decoding accuracy. The onset of significance and peak significance times were calculated for each subject, and then averaged. The significance level is determined by decoding accuracy for the chance level, which is 0.5. Shaded: the decoders are trained on shaded data and tested on shaded data. Texture: the decoders are trained on texture data and tested on texture data. Texture on Shaded: the decoders are trained on shaded data and tested on texture data. SFM: the decoders are trained on SFM data and tested on SFM data. SFM on Shaded: the decoders are trained on shaded data and tested on SFM data.

Generalization across time

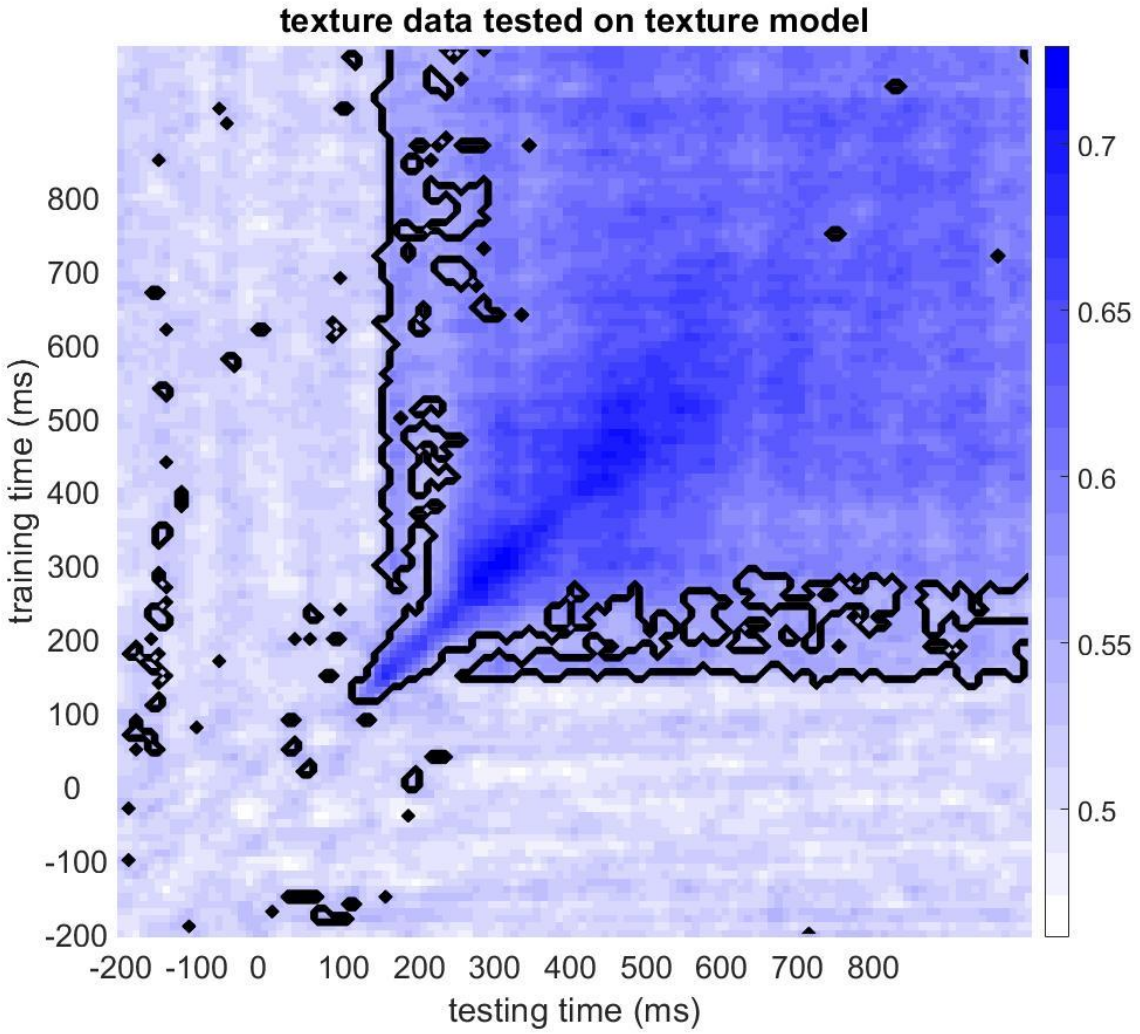
To investigate whether the information about the object is sustained in the system, we tested the temporal generalization for each of the three cues. This process involved training the linear classification model using the MEG sensor data in at one time point and tested on all other time points. For shaded face vs. chair classification, we observed that after the initial peak of high decoding accuracy, there

exists a second peak at around 300ms after the stimulus onset (figure 5). The information then sustained in the MEG sensor pattern. For the texture cue and SFM cue, the initial peak of high decoding accuracy around 170ms did not exist. For texture cue, the information sustained after around 200ms. For the SFM cues, the information sustained after around 750ms.

A)


shaded data tested on shaded model
training time (ms)
testing time (ms)
800
700
600
500
400
300
200
100
0
-100
-200
-200 -100 0 100 200 300 400 500 600 700 800
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45

B)


texture data tested on texture model
training time (ms)
testing time (ms)
800
700
600
500
400
300
200
100
0
-100
-200
-200 -100 0 100 200 300 400 500 600 700 800
0.7
0.65
0.6
0.55
0.5

C)

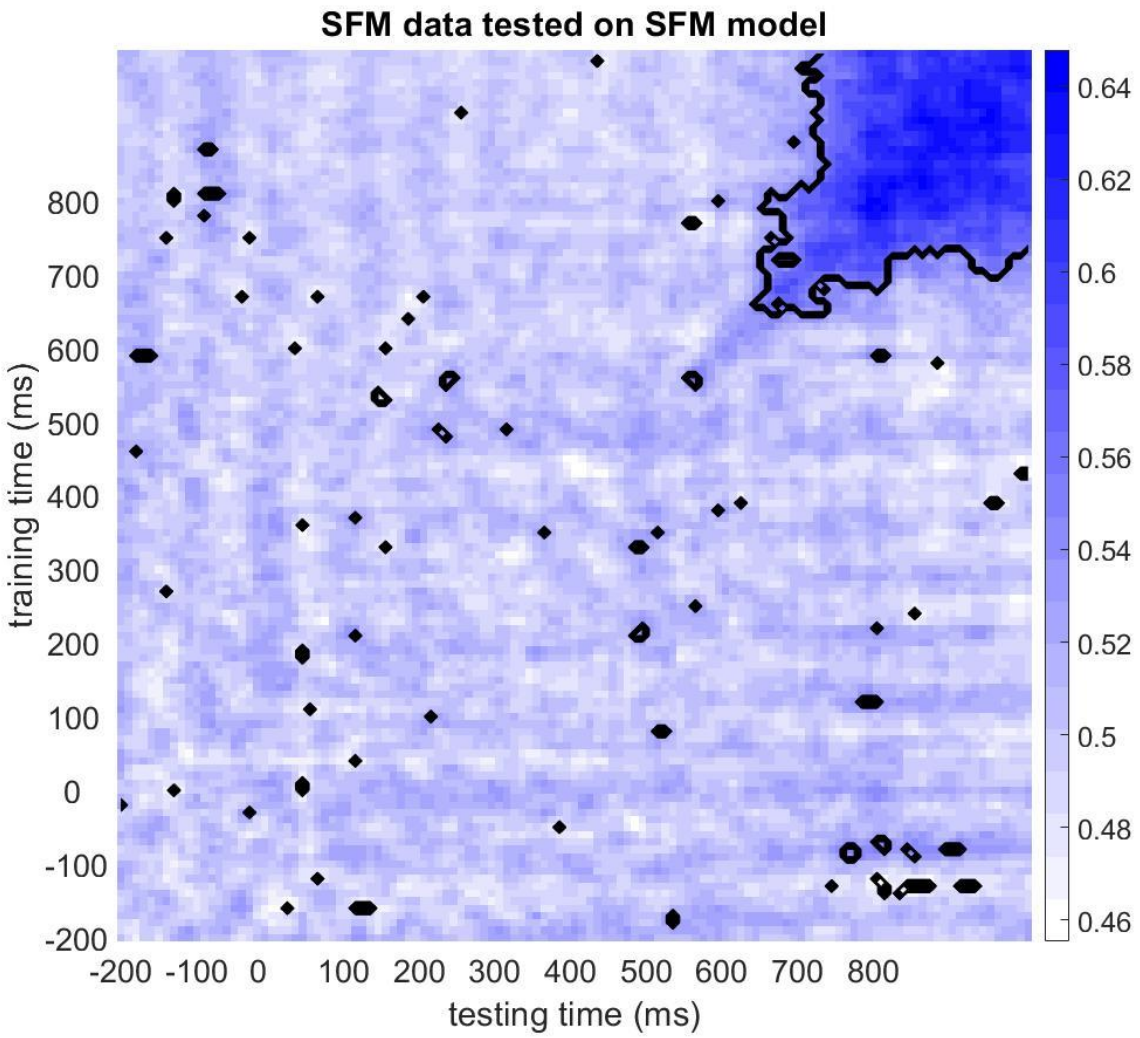
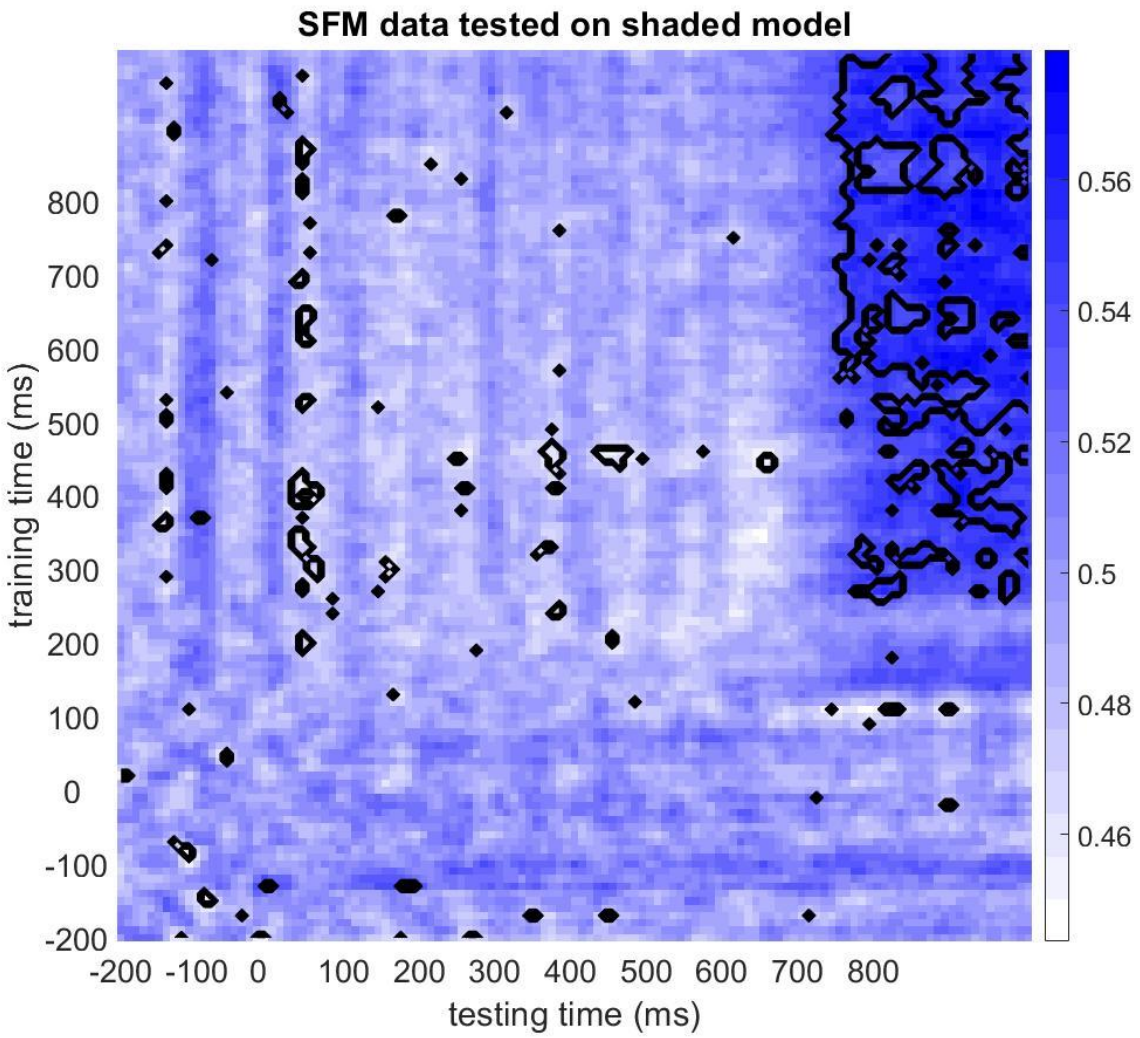


Figure 5. Temporal generalization of decoding models in face vs chair classification. Each point in the graph depicts a decoding model that was trained on one time point showing on the y axis and tested using the test data from another time point showing on the x axis. A color scale showed the classification accuracies. The black contour lines show the significant decoding models, with the significant level $p<0.05$. A), Temporal generalization for shaded stimuli. B), temporal generalization for texture stimuli. C), temporal generalization for SFM stimuli.

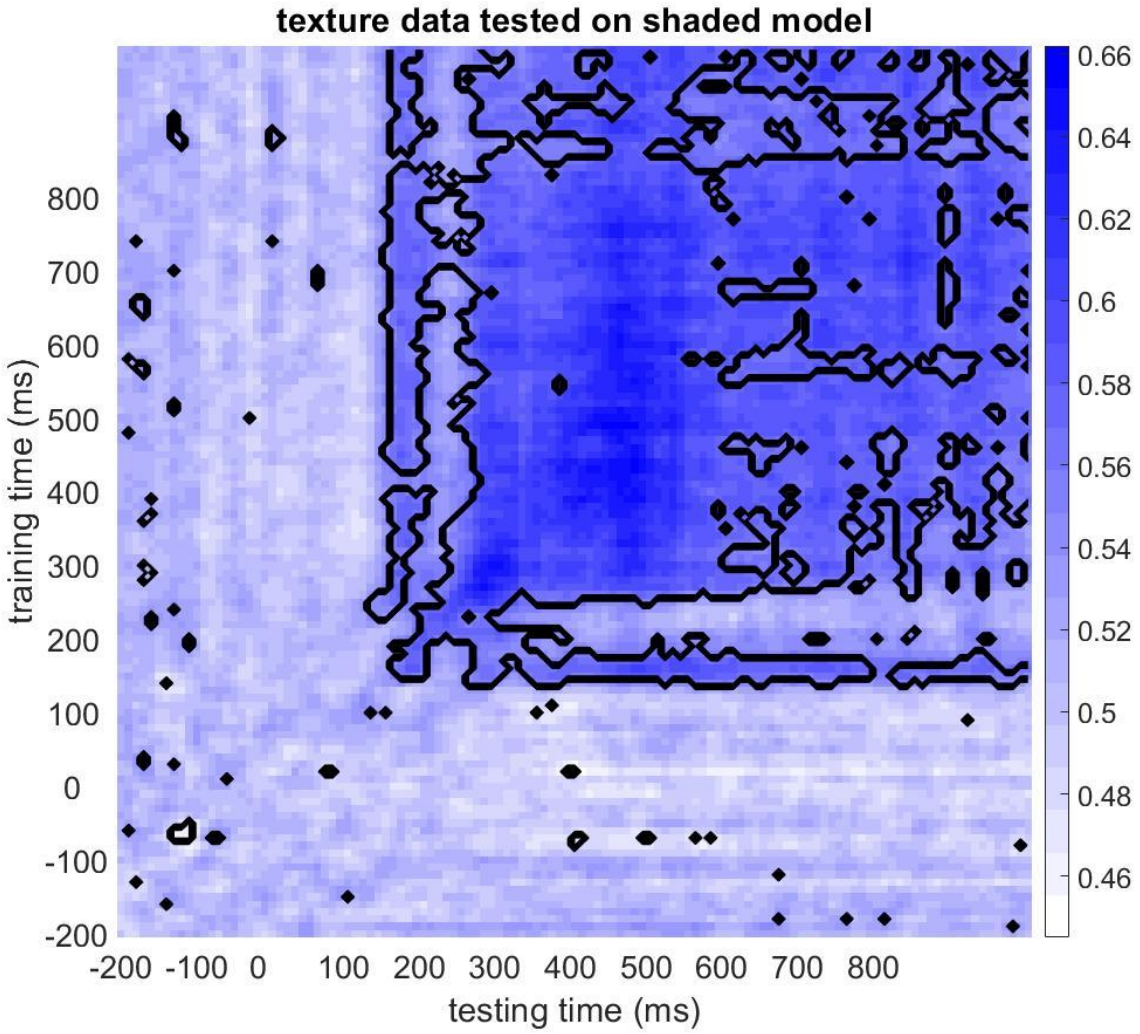
Generalization across both time and depth cue

To further address how invariance decoding can be observed, we carried out the generalization of decoding models in both across time and across depth cues. High classification accuracies can be seen from a rectangle that contains off-diagonal elements in the matrix of SFM data tested on shaded model (Fig. 6A), indicating that the patterns for MEG sensors were similar in the respective time window (750 ms – 1000 ms for SFM, 250 – 1000 ms for shaded). Similar rectangle could be observed from the results from texture data tested on shaded mode, and SFM data tested on texture model. The rectangle that contained most of the significant elements were 250 ms – 1000 ms for shaded, 250 ms – 600 ms for texture in texture data tested on shaded model (Fig. 6B). The rectangle that contained most of the significant elements were 250 ms – 600 ms for texture, 750 ms – 1000 ms for SFM for SFM data tested on texture model (Fig. 6C). These results showed that MEG sensor map patterns shared object related decoding information in the time window mentioned above, reflecting depth cue invariance object information sustaining.

A)


SFM data tested on shaded model
training time (ms)
testing time (ms)
800
700
600
500
400
300
200
100
0
-100
-200
-200 -100 0 100 200 300 400 500 600 700 800
0.56
0.54
0.52
0.5
0.48
0.46

B)


texture data tested on shaded model
training time (ms)
testing time (ms)
800
700
600
500
400
300
200
100
0
-100
-200
-200 -100 0 100 200 300 400 500 600 700 800
0.66
0.64
0.62
0.6
0.58
0.56
0.54
0.52
0.5
0.48
0.46

C)

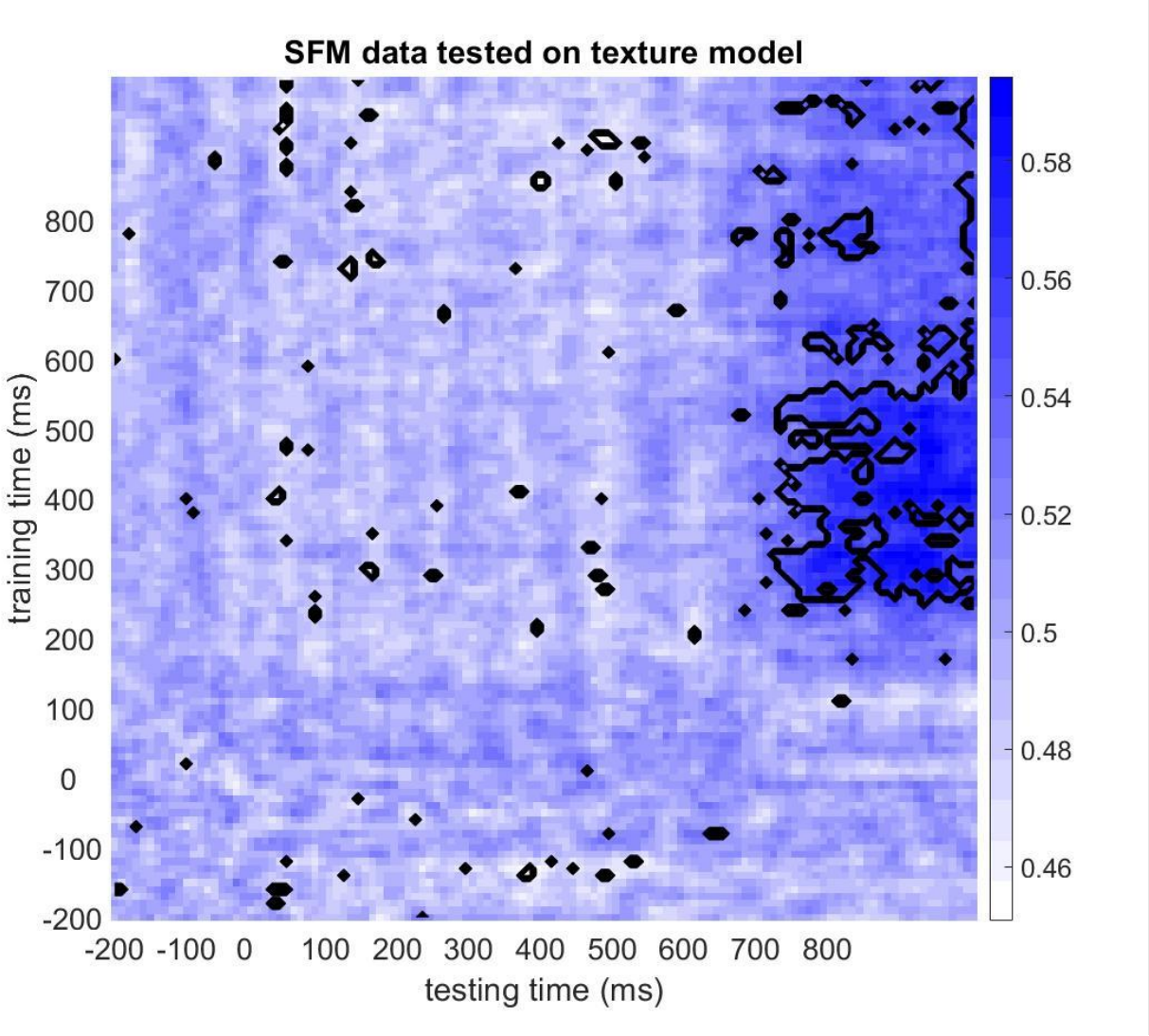


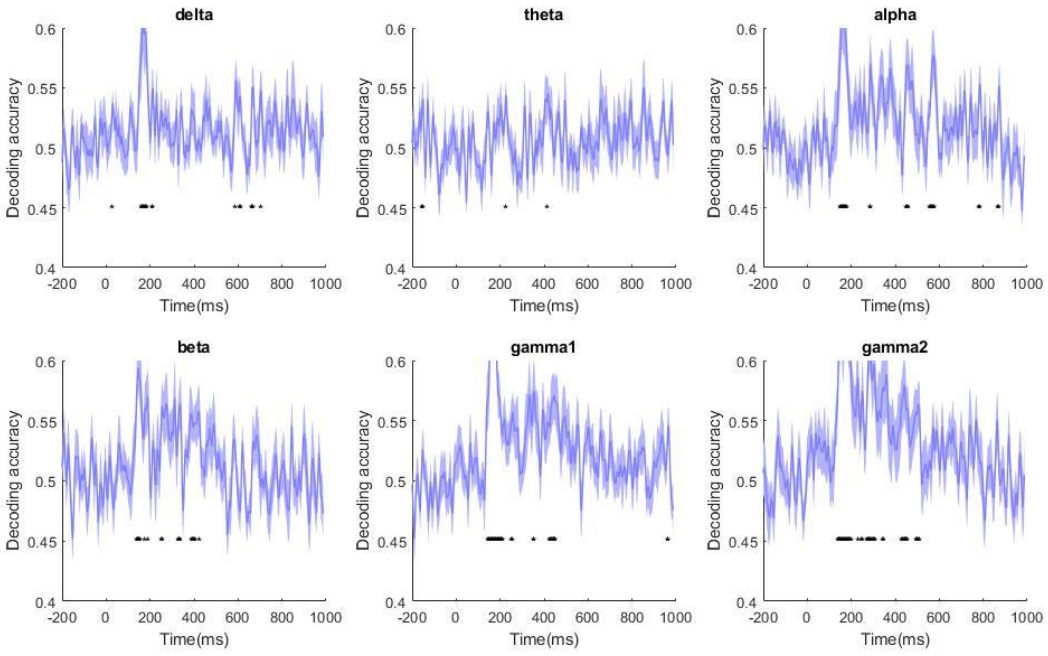
Figure 6. Generalization across both time and depth cue. Each point in the graph depicts a decoding model that was trained on one time point showing on the y axis and tested using the test data from another time point showing on the x axis. A color scale showed the classification accuracies. The black contour lines show the significant decoding models, with the significant level $p<0.05$. A), the classification models were trained on shaded data, and the testing data were from SFM. B), the classification models were trained on shaded data, and the testing data were from texture. C), the classification models were trained on texture data, and the texting data were from SFM.

MEG decoding using different frequency band power

We performed time-resolved decoding analysis for each depth cue for different frequency bands (delta, 2-4 Hz; theta, 5-7Hz; alpha, 8-12 Hz; beta, 15-29 Hz; gamma1, 30-59 Hz; gamma2, 61-90Hz). The results showed that gamma bands had high decoding accuracies compared to other frequency bands in all three conditions. The decoding accuracies levels were not stable compared to MEG channel data decoding.

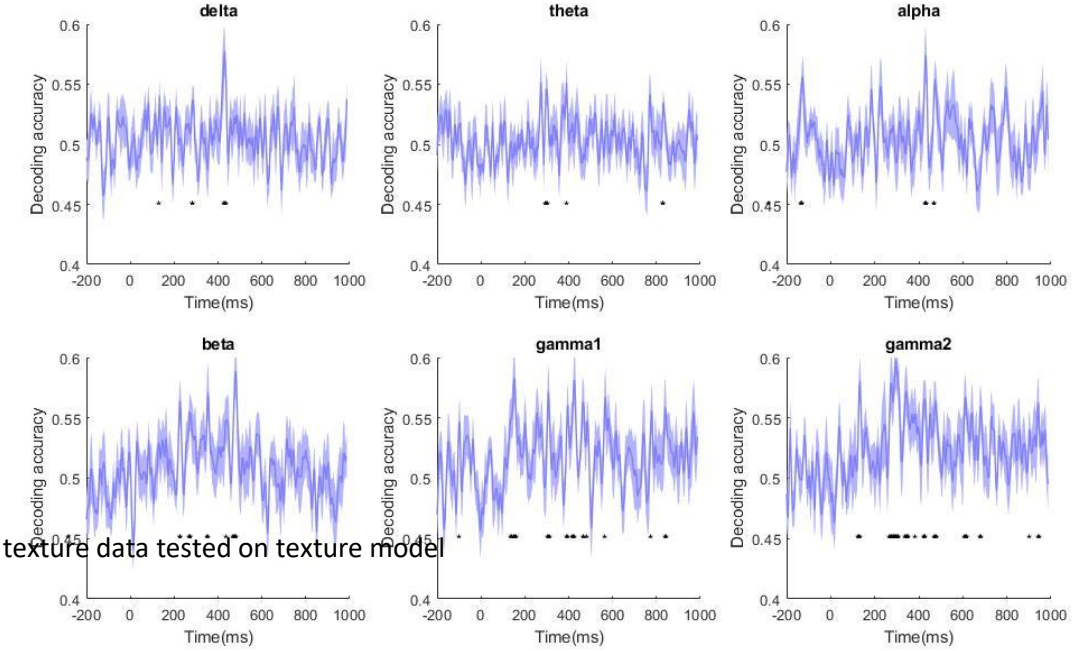
A)

Sh



B)

Texture



C)

SFM

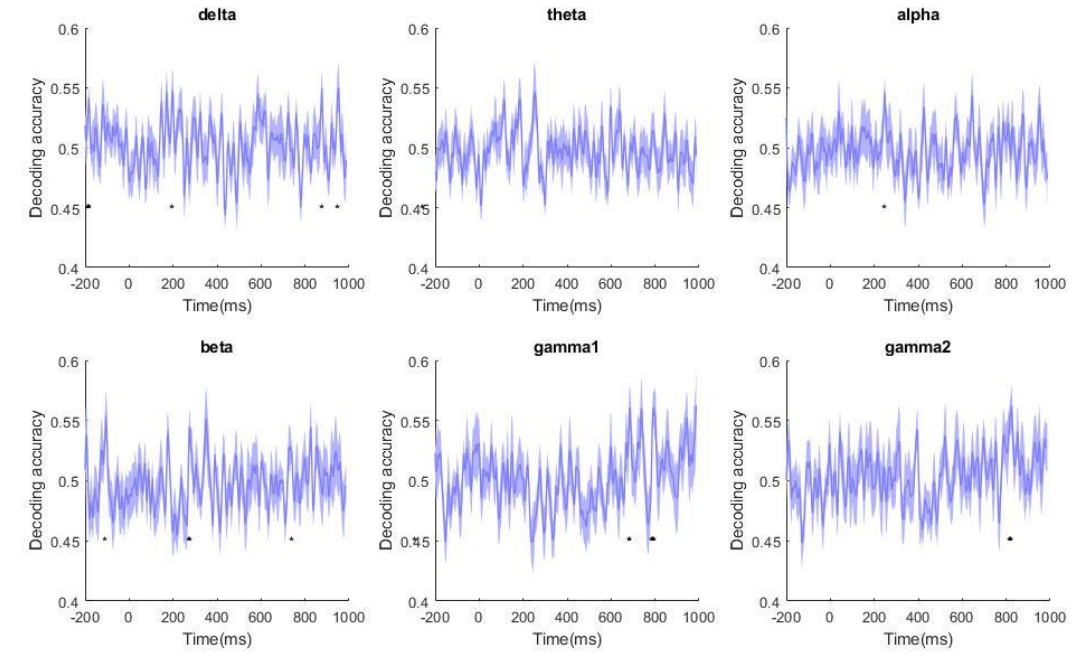


Figure 7. Decoding performance using different frequency bands. The time resolved decoding accuracies are

plotted for six different frequency bands (delta, 2-4 Hz; theta, 5-7Hz; alpha, 8-12 Hz; beta, 15-29 Hz; gamma1, 30-59 Hz; gamma2, 61-90Hz). 0s on the time axis depict stimuli onset. A solid blue line shows the average time resolved decoding accuracies across the subjects. Shaded blue area shows the standard errors of the mean. Blue stars below the 0.5 black line showed the significant time point, with $p < 0.05$.

Discussion

Depth cue invariance information can be verified using MEG MVPA decoding algorithms

Although depth cue invariance of objects has been observed behaviorally, it is not clear how this invariant process is represented in the visual system. Here we addressed this issue by using time resolved MEG decoding method to classify the object information from the data that acquired when the subjects were watching the objects using three types of depth cues (shading, texture, and SFM). We verified that the object information can be decoded from the temporal flow in all three depth cue conditions, supporting the idea that even a single depth cue is enough for the visual system to recognize objects. We were able to detect the object information from shading and texture objects as early as 120 ms and 200ms, respectively, which is consistent with previous findings (Isik et al 2014, Wardle et al 2016). The object detection form SFM stimuli showed large latency, at around 650ms. This long latency seeing in out data is due to the fact that the stimuli is hard to detect. The 3D objects defined by motion cues require depth information to be gathered over several video frames. This accumulative information is also likely to incur longer processing time in the visual system. We were able to see the gradual increase of decoding accuracy over time in SFM objects in the temporal generalization results, indicating this information is processed continuously for motion depth cue (Figure 5C). Stable object representations can be reached after the sufficient accumulation of object perception.

Supporting of the independent hypothesis: timing of cross-cue generalization

We were able to detect the 3D depth cue invariant object information when we performed the cross-depth cue decoding. The onset of significant invariant object information always come significant later than the cue-specific object information, while the peak decoding time has no significant difference between the invariant group and the cue-specific group. These results showed that the initial object identity signal from the 3D cues could be separated from the invariance signals like other invariance object information (position, size, viewpoint, etc.). Our Depth-cue specific object decoding existed indicating that the object recognition using 3D cues undergoes a two-stage process, in which each depth cue specific object information is processed independently, before integration happens and the intact object information is formed.

The dynamics of invariance decoding results showed a rectangle area in which the most significant decoding accuracies could be observed for each training-testing condition pairs (Fig. 6). The rectangle showed shared MEG sensor cap pattern between depth cue conditions, hence the depth cue invariant object information sustaining time. In our experimental conditions, this time is different for three different depth cues (250 ms – 1000 ms for shaded, 250 ms – 600 ms for texture, and 750 ms – 1000 ms for SFM), reflecting the difficulties in the vision tasks to discern the objects. Sustaining times for each condition were similar to onset of significant values, which were calculated from the diagonal values from generalization across both time and depth cues results. For SFM data tested on texture data models, using both generalization of time and depth cues were more accurate, than onset of significance, since the significant rectangles were local off-diagonal (Fig. 6C). Overall, these results showed compelling evidence that the visual system might take different latencies in processing different depth cues, especially the object recognition task is challenging.

Frequency responses: gamma response is related to cross-cue generalization

Objects and natural scenes have been shown to be related to gamma band power (Brunet et al 2015, Martinovic et al 2008). Here we have demonstrated that the 3D objects recognition is directly related to gamma band power, but not the lower frequency bands. Gamma band power may be a critical neural mechanism for 3D object recognition, potentially playing a key role in how the brain processes and represents information about the shape, texture, motion structure, and interactive relationships of objects in the environment. This could have implications for understanding the neural basis of visual perception. The object information in the gamma band could be used to help to refine or improve existing models of object recognition. By highlighting the importance of gamma band activity, more accurate and detailed models may be built for deciphering how the brain recognizes and represents objects. The gamma band could be established as a target oscillatory frequency for developing brain machine interface therapeutics regarding to object recognition abilities in individuals with perceptual deficits or neurological disorders. The roll of gamma band representations could be an important step forward in our understanding of the neural mechanisms underlying visual perception, and may have important implications for a wide range of fields, including neuroscience, psychology, and computer vision.

Reference

Akhavein H. 2017. Depth cue invariant object representations in the visual cortex. PhD Thesis

Akhavein H, Dehmoobadsharifabadi A, Farivar R. 2018. Magnetoencephalography adaptation reveals depth-cue-invariant object representations in the visual cortex. J Vis 18: 6

Baillet S. 2017. Magnetoencephalography for brain electrophysiology and imaging. Nat Neurosci 20: 327-39

Brunet N, Bosman CA, Roberts M, Oostenveld R, Womelsdorf T, et al. 2015. Visual cortical gamma-band activity during free viewing of natural images. Cereb Cortex 25: 918-26

Cohen XM. 2014. Analyzing Neural Time Series Data: Theory and Practice. pp. 33. The MIT Press.

Dehmoobadsharifabadi A, Farivar R. 2016. Are face representations depth cue invariant? J Vis 16: 6

Dovencioglu D, Ban H, Schofield AJ, Welchman AE. 2013. Perceptual integration for qualitatively different 3-D cues in the human brain. J Cogn Neurosci 25: 1527-41

Farivar R. 2009. Dorsal-ventral integration in object recognition. Brain Res Rev 61: 144-53

Genovese CR, Lazar NA, Nichols T. 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage 15: 870-8

Isik L, Meyers EM, Leibo JZ, Poggio T. 2014. The dynamics of invariant object recognition in the human visual system. J Neurophysiol 111: 91-102

Kamitani Y, Tong F. 2006. Decoding seen and attended motion directions from activity in the human visual cortex. Curr Biol 16: 1096-102

King JR, Dehaene S. 2014. Characterizing the dynamics of mental representations: the temporal generalization method. Trends Cogn Sci 18: 203-10

Konen CS, Kastner S. 2008. Two hierarchically organized neural systems for object information in human visual cortex. Nat Neurosci 11: 224-31

Marti S, King JR, Dehaene S. 2015. Time-Resolved Decoding of Two Processing Chains during Dual-Task Interference. Neuron 88: 1297-307

Martinovic J, Gruber T, Hantsch A, Muller MM. 2008. Induced gamma-band activity is related to the time point of object identification. Brain Res 1198: 93-106

Merigan WH. 2000. Cortical area V4 is critical for certain texture discriminations, but this effect is not dependent on attention. Vis Neurosci 17: 949-58

Meyers EM, Freedman DJ, Kreiman G, Miller EK, Poggio T. 2008. Dynamic population coding of category information in inferior temporal and prefrontal cortex. J Neurophysiol 100: 1407-19

Muller MM, Bosch J, Elbert T, Kreiter A, Sosa MV, et al. 1996. Visually induced gamma-band responses in human electroencephalographic activity--a link to animal studies. Exp Brain Res 112: 96-102

Ratcliff R, Philiastides MG, Sajda P. 2009. Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. Proc Natl Acad Sci U S A 106: 6539-44

Tadel F, Baillet S, Mosher JC, Pantazis D, Leahy RM. 2011a. Brainstorm: A User-Friendly Application for MEG/EEG Analysis. Computational Intelligence and Neuroscience 2011: 879716

Tadel F, Baillet S, Mosher JC, Pantazis D, Leahy RM. 2011b. Brainstorm: a user-friendly application for MEG/EEG analysis. Comput Intell Neurosci 2011: 879716

Wardle SG, Kriegeskorte N, Grootswagers T, Khaligh-Razavi SM, Carlson TA. 2016. Perceptual similarity of visual patterns predicts dynamic neural activation patterns measured with MEG. Neuroimage 132: 59-70

Zeki S. 1991. Cerebral akinetopsia (visual motion blindness). A review. Brain 114 (Pt 2): 811-24

Chapter 5: Discussion and Conclusion

Summary of findings

The representations of 3D objects and scenes were encoded in the human visual system in a diffused and distributed way. Previous studies in our lab demonstrated that humans could recognize objects from various monocular and binocular depth cues, and these perceptions can cause cross-depth-cue adaptation effects (Akhavein 2017, Dehmoobadsharifabadi & Farivar 2016), indicating an integration of depth cue processing. The concept that depth cues could and should be an important part of object and scene recognition inspired the chapters in this thesis. In this thesis, I investigated how the recognition of objects and scenes involved depth perception in three different ways. In the first approach, we used MEG to measure the prototypical representations of movie scenes across subjects. Our results revealed that the depth information of movie scenes was present in the gamma frequency bands but not in the lower frequency bands. In contrast to most inter-subject correlation studies in MEG and EEG indicating that the lower frequency bands were temporally similar across the subjects (Chang et al 2015, Chen et al 2020b, Lankinen et al 2014), we showed that multivariate classification could be used to demonstrate representations of movie scenes in an intuitive and interpretable way. Our results further showed that the gamma band patterns were directly related to the depth contrast of the movie scene in the stimuli, supporting the idea that the gamma bands reflect the visual stimulus representations (Hermes et al 2015a, Hermes et al 2015b).

I combined the traditional ISC approach and the multivariate representational similarity analysis technique to tackle the problem that the naturalistic stimuli neuroimaging data that were otherwise difficult to analyze. The methods for this work were inspired by a number of studies that showed MEG or EEG data could be used to classify stimulus labels in event-related designs (Cichy et al 2014, Kaneshiro et al 2015). To expand this classification method to naturalistic stimuli, I took advantage of the fact that

the movie clips were naturally segregated by scenes. Within each scene, the content of the movie remained relatively stable. I used each movie scene as a stimulus label to train and test linear classifiers. The gamma bands stood out to be the most effective oscillatory bands for detecting movie scene labels in the testing datasets. This reliable classification allowed us to establish the movie scene structures in the representational space, which was also intuitive and interpretable by dendrograms.

We took a step further to show that the representational structure of movie scenes is comparable between subjects and can be used to relate to movie scene statistics, especially depth contrast. These results showed evidence that depth contrast in the movie scenes were robustly related to the cortical representations of the scenes in gamma bands, and this relatedness was prototypical across subjects.

One of the concerns remained in Chapter 2 of this thesis was that the measurements of the representational structures were not direct measurements of cortical responses, that the outcomes are statistics (correlation coefficients) based on statistics (classification accuracy), hence secondary statistics. Concern could be raised that intrinsic errors in the data could have been amplified should the number of statistical procedures increases. I solved this problem by setting up negative control groups to ensure the specificity of the positive results. I used the phase randomization method to remove the relatedness of the data to set up the negative control (Chang et al 2015). This ensured that the relationships we identified in the data were not artifacts.

I found that gamma-related representations in the movie dataset were driven by depth contrast information in the stimuli. The fact that the depth contrast information was present prototypically across the subjects was important and novel. I argue that in rich and dynamic scenes (like the movies used in Chapter 2 and 3) the depth information is vital, because it provides information, both for the boundary of objects in the scenes, but also the texture of the objects. Out result in Chapter 2 provided strong support that the depth information and the scene representations (gamma bands) were closely bonded, if not inseparable.

In Chapter 3, I investigated how visual networks respond to binocular disparity in the same 3-D vs 2-D movie-viewing paradigm. I used fMRI to measure connectivity networks in the visual areas and found significant changes between the viewing conditions in ISFC networks, but not the FC networks. This finding supports the idea that similarity in functional connectivity networks can be an effective measurement of shared visual experience. In addition, my results showed that the ISFC networks could be used to decode the movie viewing condition labels, but the FC networks could not. We argue that by using inter-subject functional connectivity, the temporal subject-specific responses are removed from the dataset, and the shared visual representations can be robustly estimated.

I speculate that it is challenging to detect the response of the FC networks using traditional analysis, because the FC network encapsulates both intrinsic structural connections and stimulus-induced connections within each subject. For visual networks, the intrinsic structural connections prove to be far stronger (i.e., result in larger inter-areal BOLD correlation), so that we cannot detect the stereopsis-induced network response by using classifiers. It is consistent with the traditional view that the intrinsic structures of connectivity are stable and only small task-related modulations can take place. This finding shared similarities to ISC changes induced by stereopsis that these changes existed but were hard to detect using the standard measures—the ISFC method combined the benefits of FC and ISC and showed a different response to stimuli-induced pattern. ISFC networks showed a reliably distinguishable ability between 3D movie and 2D movie stimuli, which contrasted with the relatively non-selective FC network.

Furthermore, subnetworks from the dorsal visual stream were found to be strongly effective in carrying stereo-depth information. To summarize, we conclude that depth perception can play a role in the scene representations at the cortical connectivity level. Our findings in this Chapter were supplements for the previous chapter, where the same naturalistic movie stimuli were used. Although different neuroimaging tools were used in these two studies, I speculate that gamma representations

may have been related to dorsal visual networks at the inter-subject level. This could be a valuable direction for future works.

The ISFC approach showed that the correlation structure of the visual networks can undergo reliable changes when stereopsis is introduced. My results in this project reflect a similar duality between "signal correlation" and "noise correlation" in electrophysiological research (Ruff & Cohen 2014). Here in this analogy, the "signal correlation" can be viewed as the stimulus-related connectivity response, which is a group effect. The "noise correlation" on the other hand, represents the stable connectivity networks across the brain regions within a subject. The "noise correlation" generates from the slow waves, non-specific activations, and default mode networks that are not stimulus-related. The ISFC method offers a related concept of "signal correlation" at the BOLD connectivity level across subjects, separating from FC stable networks, which greatly improves the efficiency of detecting the stimuli effect.

In Chapter 4, I investigated how processing of depth cue defined 3D objects (face, chair, and face contour) evolves over the time immediately following stimulus onset to yield depth-cue invariant representations. We found that the decoding model constructed from one depth cue could be used to decipher information from other depth cues, supporting the idea that object recognition is depth-cue invariant. Further, we found that the timing of cross-depth-cue decoding was time sensitive, namely that depth cue invariant decoding came after the cue-specific decoding for each depth cue. These findings supported the idea that depth cue specific information exist before the depth cue invariant 3D object percepts formed in the visual system. My findings build up the evidence that 3D object recognition is a multi-staged process. The two-stage theory is favorable here that the depth cue related object information was processed in a separate stage, possibly locally, before the integration stage.

The dynamic decoding algorithms with moving windows of 10 ms were used to evaluate the object information contained in any time points without compromising the signal-to-noise ratio (Isik et al 2014,

King & Dehaene 2014). Thus, 10 millisecond temporal resolutions can be achieved, which is extremely short and allows for very fine-grained evaluation of cortical responses to stimuli. With this technique, we measured the exact timing of object information flow for each participant and discovered that the time needed for the onset of non-chance discrimination and peak discrimination times are very different between different depth cues, indicating that random perceptual latencies exist in different visual pathways. This temporal generalization technique can also be used to show the progression of the object recognition process. I demonstrated that the object information in the brain undergoes both an accumulation process and a sustaining process. Different depth cues showed different latencies of these two processes. Only in the sustain process did the object information became invariant. The dynamic MEG sensor decoding method can be improved in the future by increasing its spatial specificity. This can be complemented by using feature selection techniques to extract the components with higher decoding performances, including canonical correlations (Gaebler et al 2014). Another potential improvement for the study is to incorporate more object categories, thus, to expand the scope of investigation.

The future of using multivariate analysis in neuroimaging data

Multivariate analysis focuses on the aspects of the representations in the brain and uses abstract mathematical constructs to refine the neuroimaging data into a simpler, estimable forms. These refined forms of abstractions can be interpreted as representations of mental activities in the brain. One of the benefit of this abstraction is that the approach is based on data features rather than anatomical features, which greatly overcomes the issue that the anatomical localization and the functional brain map are not perfectly aligned (Guntupalli et al 2016). Bypassing this unsound anatomical registration step, group analysis can be done more accurately. For example, in Chapter 2 of this thesis, we performed representational similarity analysis that using the MEG channels as features, and the group

analysis were done based on the representational features (Chen et al 2020). Using multivariate analysis to measure the representations over comes the drawbacks that the basic measuring units of the neuroimaging techniques contains complex neural activity sources. The reason for this is that the multivariate analysis is often data driven and the information can be more efficiently extracted using the machine learning decoding algorithm. Beyond that, multivariate analysis can be used on comparison between different datasets (Cichy et al 2014), different measuring techniques (Cichy et al 2016), or even different species (Kriegeskorte et al 2008b, Mantini et al 2012). Multivariate analysis can be used as a bridging tool to connect different branches of neuroscience.

With this development in the multivariate analysis in the neuroimaging techniques, advanced machine learning algorithms that specifically developed for neuroimaging data may be required in the future. Although the current multivariate analysis is mainly focused on linear algorithms such as LDA or linear SVM, the fine-tuning analysis tools should be explored to unveil the non-linear relationships between the data features, which may reflect the complex interactions across distant brain areas.

Multivariate analysis of depth perception

In this thesis, I applied multivariate tools to contribute to the diversity of analysis, including representational similarity using MEG sensor in the second chapter (Chen & Farivar 2020), inter-subject functional connectivity in the third chapter, and the MEG sensor decoding the third chapter. Through my novel applications to these problems, I removed phenomena in the data that were without functional relevance—components which could reflect random temporal variations across the participants, random anatomical differences among the participants, or random neural noises. The selected methods were appropriate to assess the neural codes that carry the information of how the brain perceives the world via the visual depth. I showed that the depth perception is related to a diverse

brain network rather than a single region. Multivariate analysis fits the challenges to tackle this diffused pattern of cortical activities by summarizing the information that conveyed by all the nodes combined.

Challenges of multivariate analysis

Multivariate analysis will likely continue to be a major arsenal in deciphering the process of depth perception or functional mental activities, but these tools are not without their own unique challenges. One of the challenges is the requirement of data capacity. Multivariate analysis is often performed using data driven approach, which means that it relies on large amounts of data to identify patterns and relationships among variables. This can be challenging because collecting and processing large amounts of data can be time-consuming and costly. Additionally, it requires careful attention to data quality, as any errors or inconsistencies can compromise the accuracy of the analysis. Another challenge for the multivariate analysis is the interpretability. More specifically, data analysis that generates statistical inferences, patterns, similarity measurements, which does not have a one-to-one corresponding activation map to the anatomical structure. These statistical measures may not even directly relate to any physiological or neuronal activities, like action potentials or hemodynamic changes. These challenges make the understanding of multivariate decoding analysis a difficult task. A third challenge in multivariate analysis is the relatively intolerance to error. This is because that multivariate analysis, utilizing fine grained data features, is more precise and accurate compared to traditional univariate analysis. Multivariate analysis often involves multiple layers of statistical operations (e.g., correlations, distances, or classifications). Measurement errors can be amplified through the layers of statistical estimations. This can happen because each statistical operation adds its own sources of variation and potential error to the data. At each step of the data analysis, measurement errors in the variables can be amplified. Increase the data capacity could be the solution to this problem. To minimize the impact of measurement errors in multivariate analysis, it is important to carefully measure and validate the

variables, and to use appropriate statistical techniques that are robust to measurement errors. One of the potential ways for this is using negative controls in the experimental design for random shuffled data (Gorgen et al 2018). It is not uncommon for a permutation test to be performed to achieve accurate statistics compared to assuming normal distributions (Chang et al 2015).

3D representation of object and scenes

Objects and scenes in the real world are presented to our visual system with depth. However, contemporary object recognition models are focused on the complex configuration of 2D edges of objects and the responsive modules in the ventral visual pathways (DiCarlo & Cox 2007). I showed in the second chapter that the representations of naturalistic scenes are related to the representation of depth. I demonstrated that the stereopsis in a naturalistic condition can induce changes in the networks of the cortical visual connections, especially the dorsal visual pathway, in the third chapter. I also demonstrated that the invariance of depth cues in 3D object recognition, indicating that robust object representation can arise from unique depth cues. These findings suggest that the object recognition process should include the depth perception elements, which the current popular models often failed to appreciate (Akhavein et al 2018, Farivar 2009). Computer vision models for object recognition are developed based on 2D feature models. To improve the validity of these computer vision models in the real world that involves depth cues like shading, textures, and motions, depth calculations need to be considered. New models that describe the objects and scenes in the form of 3D presentation needs to be established with depth perception adding to the equation.

Reference

Akhavein H. 2017. Depth cue invariant object representations in the visual cortex. *PhD Thesis*

Akhavein H, Dehmoobadsharifabadi A, Farivar R. 2018. Magnetoencephalography adaptation reveals depth-cue-invariant object representations in the visual cortex. *J Vis* 18: 6

Allefeld C, Haynes JD. 2014. Searchlight-based multi-voxel pattern analysis of fMRI by cross-validated MANOVA. *Neuroimage* 89: 345-57

Andersen LM, Jerbi K, Dalal SS. 2020. Can EEG and MEG detect signals from the human cerebellum? *Neuroimage* 215: 116817

Anzai A, DeAngelis GC. 2010. Neural computations underlying depth perception. *Curr Opin Neurobiol* 20: 367-75

Ardekani BA, Bachman AH, Strother SC, Fujibayashi Y, Yonekura Y. 2004. Impact of inter-subject image registration on group analysis of fMRI data. *International Congress Series* 1265: 49-59

Backus BT, Fleet DJ, Parker AJ, Heeger DJ. 2001. Human cortical activity correlates with stereoscopic depth perception. *J Neurophysiol* 86: 2054-68

Bai F, Watson DR, Yu H, Shi Y, Yuan Y, Zhang Z. 2009. Abnormal resting-state functional connectivity of posterior cingulate cortex in amnestic type mild cognitive impairment. *Brain Res* 1302: 167-74

Baillet S. 2017. Magnetoencephalography for brain electrophysiology and imaging. *Nat Neurosci* 20: 327-39

Baillet S, Friston K, Oostenveld R. 2011. Academic software applications for electromagnetic brain mapping using MEG and EEG. *Comput Intell Neurosci* 2011: 972050

Baillet S, Garnero L. 1997. A Bayesian approach to introducing anatomo-functional priors in the EEG/MEG inverse problem. *IEEE Trans Biomed Eng* 44: 374-85

Bar M. 2004. Visual objects in context. *Nat Rev Neurosci* 5: 617-29

Barlow HB, Blakemore C, Pettigrew JD. 1967. The neural mechanism of binocular depth discrimination. *J Physiol* 193: 327-42

Bartels A, Zeki S. 2004. Functional brain mapping during free viewing of natural scenes. *Hum Brain Mapp* 21: 75-85

Betti V, Della Penna S, de Pasquale F, Mantini D, Marzetti L, et al. 2013. Natural scenes viewing alters the dynamics of functional connectivity in the human brain. *Neuron* 79: 782-97

Biederman I. 1987. Recognition-by-components: a theory of human image understanding. *Psychol Rev* 94: 115-47

Bokeria L, Henson RN, Mok RM. 2021. Map-Like Representations of an Abstract Conceptual Space in the Human Brain. *Front Hum Neurosci* 15: 620056

Bridwell DA, Roth C, Gupta CN, Calhoun VD. 2015. Cortical Response Similarities Predict which Audiovisual Clips Individuals Viewed, but Are Unrelated to Clip Preference. *PLoS One* 10: e0128833

Brookes MJ, Woolrich M, Luckhoo H, Price D, Hale JR, et al. 2011. Investigating the electrophysiological basis of resting state networks using magnetoencephalography. *Proc Natl Acad Sci U S A* 108: 16783-8

Brooks KR. 2017. Depth Perception and the History of Three-Dimensional Art: Who Produced the First Stereoscopic Images? *Iperception* 8: 2041669516680114

Brunet N, Bosman CA, Roberts M, Oostenveld R, Womelsdorf T, et al. 2015. Visual cortical gamma-band activity during free viewing of natural images. *Cereb Cortex* 25: 918-26

Carandini M. 2012. Area V1. *Scholarpedia* 7

Carlson T, Tovar DA, Alink A, Kriegeskorte N. 2013. Representational dynamics of object vision: the first 1000 ms. *J Vis* 13

Chang WT, Jaaskelainen IP, Belliveau JW, Huang S, Hung AY, et al. 2015. Combined MEG and EEG show reliable patterns of electromagnetic brain activity during natural viewing. *Neuroimage* 114: 49-56

Chen G, Shin YW, Taylor PA, Glen DR, Reynolds RC, et al. 2016. Untangling the relatedness among correlations, part I: Nonparametric approaches to inter-subject correlation analysis at the group level. *Neuroimage* 142: 248-59

Chen N, Chen Z, Fang F. 2020a. Functional specialization in human dorsal pathway for stereoscopic depth processing. *Exp Brain Res* 238: 2581-88

Chen PA, Jolly E, Cheong JH, Chang LJ. 2020b. Intersubject representational similarity analysis reveals individual variations in affective experience when watching erotic movies. *Neuroimage*: 116851

Chen Y, Farivar R. 2020. Natural scene representations in the gamma band are prototypical across subjects. *Neuroimage* 221: 117010

Chen. P-H, Chen. J, Yeshurun. Y, Hasson. U, Haxby. JV, Ramadge. PJ. 2015. A Reduced-Dimension fMRI Shared Response Model. *Advances in Neural Information Processing Systems 28 (NIPS 2015)*

Cheng K, Waggoner RA, Tanaka K. 2001. Human ocular dominance columns as revealed by high-field functional magnetic resonance imaging. *Neuron* 32: 359-74

Cichy RM, Khosla A, Pantazis D, Oliva A. 2016a. Dynamics of scene representations in the human brain revealed by magnetoencephalography and deep neural networks. *Neuroimage*

Cichy RM, Pantazis D. 2017. Multivariate pattern analysis of MEG and EEG: A comparison of representational structure in time and space. *Neuroimage* 158: 441-54

Cichy RM, Pantazis D, Oliva A. 2014. Resolving human object recognition in space and time. *Nat Neurosci* 17: 455-62

Cichy RM, Pantazis D, Oliva A. 2016b. Similarity-Based Fusion of MEG and fMRI Reveals Spatio-Temporal Dynamics in Human Cortex During Visual Object Recognition. *Cereb Cortex* 26: 3563-79

Cohen JD, Daw N, Engelhardt B, Hasson U, Li K, et al. 2017. Computational approaches to fMRI analysis. *Nat Neurosci* 20: 304-13

Cohen XM. 2014. *Analyzing Neural Time Series Data: Theory and Practice*. pp. 33. The MIT Press.

Cole MW, Reynolds JR, Power JD, Repovs G, Anticevic A, Braver TS. 2013. Multi-task connectivity reveals flexible hubs for adaptive task control. *Nat Neurosci* 16: 1348-55

Cottereau BR, McKee SP, Norcia AM. 2012. Bridging the gap: global disparity processing in the human visual cortex. *J Neurophysiol* 107: 2421-9

Cox RW. 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29: 162-73

Cox RW, Hyde JS. 1997. Software tools for analysis and visualization of fMRI data. *NMR Biomed* 10: 171-8

da Silva FHL. 2010. 1Electrophysiological Basis of MEG Signals In *MEG: An Introduction to Methods*, ed. P Hansen, M Kringelbach, R Salmelin, pp. 0: Oxford University Press

Dale AM, Liu AK, Fischl BR, Buckner RL, Belliveau JW, et al. 2000. Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron* 26: 55-67

Darvas F, Pantazis D, Kucukaltun-Yildirim E, Leahy RM. 2004. Mapping human brain function with MEG and EEG: methods and validation. *Neuroimage* 23 Suppl 1: S289-99

Dehmoobadsharifabadi A, Farivar R. 2016. Are face representations depth cue invariant? *J Vis* 16: 6

Di X, Biswal BB. 2020. Intersubject consistent dynamic connectivity during natural vision revealed by functional MRI. *Neuroimage*: 116698

DiCarlo JJ, Cox DD. 2007. Untangling invariant object recognition. *Trends Cogn Sci* 11: 333-41

Dmochowski JP, Sajda P, Dias J, Parra LC. 2012. Correlated components of ongoing EEG point to emotionally laden attention - a possible marker of engagement? *Front Hum Neurosci* 6: 112

Dobs K, Martinez J, Kell AJE, Kanwisher N. 2022. Brain-like functional specialization emerges spontaneously in deep neural networks. *Sci Adv* 8: eabl8913

Dohmatob E, Varoquaux G, Thirion B. 2018. Inter-subject Registration of Functional Images: Do We Need Anatomical Images? *Front Neurosci* 12: 64

Dovencioglu D, Ban H, Schofield AJ, Welchman AE. 2013. Perceptual integration for qualitatively different 3-D cues in the human brain. *J Cogn Neurosci* 25: 1527-41

Duncan KK, Hadjipapas A, Li S, Kourtzi Z, Bagshaw A, Barnes G. 2010. Identifying spatially overlapping local cortical networks with MEG. *Hum Brain Mapp* 31: 1003-16

Edelman S. 1998. Representation is representation of similarities. *Behav Brain Sci* 21: 449-67; discussion 67-98

Farivar R. 2009. Dorsal-ventral integration in object recognition. *Brain Res Rev* 61: 144-53

Farivar R, Blanke O, Chaudhuri A. 2009. Dorsal-ventral integration in the recognition of motion-defined unfamiliar faces. *J Neurosci* 29: 5336-42

Feilong M, Nastase SA, Guntupalli JS, Haxby JV. 2018. Reliable individual differences in fine-grained cortical functional architecture. *Neuroimage* 183: 375-86

Finn ES, Glerean E, Khojandi AY, Nielson D, Molfese PJ, et al. 2020. Idiosynchrony: From shared responses to individual differences during naturalistic neuroimaging. *Neuroimage* 215: 116828

Fred AL, Kumar SN, Kumar Haridhas A, Ghosh S, Purushothaman Bhuvana H, et al. 2022. A Brief Introduction to Magnetoencephalography (MEG) and Its Clinical Applications. *Brain Sci* 12

Friston K, Harrison L, Daunizeau J, Kiebel S, Phillips C, et al. 2008. Multiple sparse priors for the M/EEG inverse problem. *Neuroimage* 39: 1104-20

Gaebler M, Biessmann F, Lamke JP, Muller KR, Walter H, Hetzer S. 2014. Stereoscopic depth increases intersubject correlations of brain networks. *Neuroimage* 100: 427-34

Genovese CR, Lazar NA, Nichols T. 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 15: 870-8

Georgieva S, Peeters R, Kolster H, Todd JT, Orban GA. 2009. The processing of three-dimensional shape from disparity in the human brain. *J Neurosci* 29: 727-42

Gola M, Magnuski M, Szumska I, Wrobel A. 2013. EEG beta band activity is related to attention and attentional deficits in the visual performance of elderly subjects. *Int J Psychophysiol* 89: 334-41

Goldenholz DM, Ahlfors SP, Hamalainen MS, Sharon D, Ishitobi M, et al. 2009. Mapping the signal-to-noise-ratios of cortical sources in magnetoencephalography and electroencephalography. *Hum Brain Mapp* 30: 1077-86

Goncalves NR, Ban H, Sanchez-Panchuelo RM, Francis ST, Schluppeck D, Welchman AE. 2015. 7 tesla FMRI reveals systematic functional organization for binocular disparity in dorsal visual cortex. *J Neurosci* 35: 3056-72

Gonzalez-Castillo J, Bandettini PA. 2018. Task-based dynamic functional connectivity: Recent findings and open questions. *Neuroimage* 180: 526-33

Grootswagers T, Wardle SG, Carlson TA. 2017. Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. *J Cogn Neurosci* 29: 677-97

Guclu U, van Gerven MA. 2015. Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. *J Neurosci* 35: 10005-14

Guggenmos M, Sterzer P, Cichy RM. 2018. Multivariate pattern analysis for MEG: A comparison of dissimilarity measures. *Neuroimage* 173: 434-47

Guntupalli JS, Feilong M, Haxby JV. 2018. A computational model of shared fine-scale structure in the human connectome. *PLoS Comput Biol* 14: e1006120

Guntupalli JS, Hanke M, Halchenko YO, Connolly AC, Ramadge PJ, Haxby JV. 2016. A Model of Representational Spaces in Human Cortex. *Cereb Cortex* 26: 2919-34

Hadjipapas A, Adjamian P, Swettenham JB, Holliday IE, Barnes GR. 2007. Stimuli of varying spatial scale induce gamma activity with distinct temporal characteristics in human visual cortex. *Neuroimage* 35: 518-30

Hagler DJ, Jr., Saygin AP, Sereno MI. 2006. Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data. *Neuroimage* 33: 1093-103

Hall H. 2009. Under the sea.

Hämäläinen M, Hari R, Ilmoniemi RJ, Knuutila J, Lounasmaa OV. 1993. Magnetoencephalography---theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics* 65: 413-97

Hamalainen MS. 1991. Basic principles of magnetoencephalography. *Acta Radiol Suppl* 377: 58-62

Hamalainen MS, Ilmoniemi RJ. 1994. Interpreting magnetic fields of the brain: minimum norm estimates. *Med Biol Eng Comput* 32: 35-42

Handwerker DA, Ollinger JM, D'Esposito M. 2004. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage* 21: 1639-51

Hasson U, Landesman O, Knappmeyer B, Vallines I, Rubin N, Heeger DJ. 2008. Neurocinematics: the neuroscience of film. *Projections* 2: 1–26

Hasson U, Malach R, Heeger DJ. 2010. Reliability of cortical activity during natural stimulation. *Trends Cogn Sci* 14: 40-8

Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R. 2004. Intersubject synchronization of cortical activity during natural vision. *Science* 303: 1634-40

Haufe S, DeGuzman P, Henin S, Arcaro M, Honey CJ, et al. 2018. Elucidating relations between fMRI, ECoG, and EEG through a common natural stimulus. *Neuroimage* 179: 79-91

Haxby JV, Connolly AC, Guntupalli JS. 2014. Decoding neural representational spaces using multivariate pattern analysis. *Annu Rev Neurosci* 37: 435-56

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293: 2425-30

Haxby JV, Guntupalli JS, Connolly AC, Halchenko YO, Conroy BR, et al. 2011. A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* 72: 404-16

Haynes JD, Rees G. 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat Neurosci* 8: 686-91

Haynes JD, Rees G. 2006. Decoding mental states from brain activity in humans. *Nat Rev Neurosci* 7: 523-34

Henderson M, Vo V, Chunharas C, Sprague T, Serences J. 2019. Multivariate Analysis of BOLD Activation Patterns Recovers Graded Depth Representations in Human Visual and Parietal Cortex. *eNeuro* 6

Herculano-Houzel S. 2009. The human brain in numbers: a linearly scaled-up primate brain. *Front Hum Neurosci* 3: 31

Hermes D, Miller KJ, Wandell BA, Winawer J. 2015a. Gamma oscillations in visual cortex: the stimulus matters. *Trends Cogn Sci* 19: 57-8

Hermes D, Miller KJ, Wandell BA, Winawer J. 2015b. Stimulus Dependence of Gamma Oscillations in Human Visual Cortex. *Cereb Cortex* 25: 2951-9

Hillebrand A, Barnes GR, Bosboom JL, Berendse HW, Stam CJ. 2012. Frequency-dependent functional connectivity within resting-state networks: an atlas-based MEG beamformer solution. *Neuroimage* 59: 3909-21

Holmes CJ, Hoge R, Collins L, Woods R, Toga AW, Evans AC. 1998. Enhancement of MR images using registration for signal averaging. *J Comput Assist Tomogr* 22: 324-33

Hummel JE, Biederman I. 1992. Dynamic binding in a neural network for shape recognition. *Psychol Rev* 99: 480-517

Ichihara S, Kitagawa N, Akutsu H. 2007. Contrast and depth perception: effects of texture contrast and area contrast. *Perception* 36: 686-95

Isik L, Meyers EM, Leibo JZ, Poggio T. 2014. The dynamics of invariant object recognition in the human visual system. *J Neurophysiol* 111: 91-102

Isik L, Singer J, Madsen JR, Kanwisher N, Kreiman G. 2017. What is changing when: Decoding visual information in movies from human intracranial recordings. *Neuroimage*

Jeong SK, Xu Y. 2016. Behaviorally Relevant Abstract Object Identity Representation in the Human Parietal Cortex. *J Neurosci* 36: 1607-19

Kamitani Y, Tong F. 2005. Decoding the visual and subjective contents of the human brain. *Nature Neuroscience* 8: 679-85

Kamitani Y, Tong F. 2006. Decoding seen and attended motion directions from activity in the human visual cortex. *Curr Biol* 16: 1096-102

Kaneshiro B, Perreau Guimaraes M, Kim HS, Norcia AM, Suppes P. 2015. A Representational Similarity Analysis of the Dynamics of Object Processing Using Single-Trial EEG Classification. *PLoS One* 10: e0135697

Kayser C, Salazar RF, Konig P. 2003. Responses to natural scenes in cat V1. *J Neurophysiol* 90: 1910-20

Kheradpisheh SR, Ghodrati M, Ganjtabesh M, Masquelier T. 2016. Deep Networks Can Resemble Human Feed-forward Vision in Invariant Object Recognition. *Sci Rep* 6: 32672

Ki JJ, Kelly SP, Parra LC. 2016. Attention Strongly Modulates Reliability of Neural Responses to Naturalistic Narrative Stimuli. *J Neurosci* 36: 3092-101

Kim D, Kay K, Shulman GL, Corbetta M. 2017. A New Modular Brain Organization of the BOLD Signal during Natural Vision. *Cereb Cortex*: 1-17

King JR, Dehaene S. 2014. Characterizing the dynamics of mental representations: the temporal generalization method. *Trends Cogn Sci* 18: 203-10

Klein A, Ghosh SS, Avants B, Yeo BT, Fischl B, et al. 2010. Evaluation of volume-based and surface-based brain image registration methods. *Neuroimage* 51: 214-20

Konen CS, Kastner S. 2008. Two hierarchically organized neural systems for object information in human visual cortex. *Nat Neurosci* 11: 224-31

Kozunov V, Nikolaeva A, Stroganova TA. 2017. Categorization for Faces and Tools-Two Classes of Objects Shaped by Different Experience-Differs in Processing Timing, Brain Areas Involved, and Repetition Effects. *Front Hum Neurosci* 11: 650

Kriegeskorte N. 2015. Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing. *Annu Rev Vis Sci* 1: 417-46

Kriegeskorte N, Bandettini P. 2007. Analyzing for information, not activation, to exploit high-resolution fMRI. *Neuroimage* 38: 649-62

Kriegeskorte N, Diedrichsen J. 2019. Peeling the Onion of Brain Representations. *Annu Rev Neurosci* 42: 407-32

Kriegeskorte N, Douglas PK. 2019. Interpreting encoding and decoding models. *Curr Opin Neurobiol* 55: 167-79

Kriegeskorte N, Goebel R, Bandettini P. 2006. Information-based functional brain mapping. *Proc Natl Acad Sci U S A* 103: 3863-8

Kriegeskorte N, Mur M, Bandettini P. 2008a. Representational similarity analysis - connecting the branches of systems neuroscience. *Front Syst Neurosci* 2: 4

Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, et al. 2008b. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60: 1126-41

Kriegeskorte N, Wei XX. 2021. Neural tuning and representational geometry. *Nat Rev Neurosci* 22: 703-18

Lahnakoski JM, Glerean E, Jaaskelainen IP, Hyona J, Hari R, et al. 2014. Synchronous brain activity across individuals underlies shared psychological perspectives. *Neuroimage* 100: 316-24

Lankinen K, Saari J, Hari R, Koskinen M. 2014. Intersubject consistency of cortical MEG signals during movie viewing. *Neuroimage* 92: 217-24

Leahy RM, Mosher JC, Spencer ME, Huang MX, Lewine JD. 1998. A study of dipole localization accuracy for MEG and EEG using a human skull phantom. *Electroencephalogr Clin Neurophysiol* 107: 159-73

Logothetis NK. 2003. The underpinnings of the BOLD functional magnetic resonance imaging signal. *J Neurosci* 23: 3963-71

Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A. 2001. Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412: 150-7

Mahmoudi A, Takerkart S, Regragui F, Boussaoud D, Brovelli A. 2012. Multivoxel pattern analysis for FMRI data: a review. *Comput Math Methods Med* 2012: 961257

Malinen S, Hari R. 2011. Data-based functional template for sorting independent components of fMRI activity. *Neurosci Res* 71: 369-76

Mandelkow H, de Zwart JA, Duyn JH. 2017. Effects of spatial fMRI resolution on the classification of naturalistic movies. *Neuroimage* 162: 45-55

Marti S, King JR, Dehaene S. 2015. Time-Resolved Decoding of Two Processing Chains during Dual-Task Interference. *Neuron* 88: 1297-307

Martinovic J, Gruber T, Hantsch A, Muller MM. 2008. Induced gamma-band activity is related to the time point of object identification. *Brain Res* 1198: 93-106

Mathewson KE, Lleras A, Beck DM, Fabiani M, Ro T, Gratton G. 2011. Pulsed out of awareness: EEG alpha oscillations represent a pulsed-inhibition of ongoing cortical processing. *Front Psychol* 2: 99

Merigan WH. 2000. Cortical area V4 is critical for certain texture discriminations, but this effect is not dependent on attention. *Vis Neurosci* 17: 949-58

Meyers EM, Freedman DJ, Kreiman G, Miller EK, Poggio T. 2008. Dynamic population coding of category information in inferior temporal and prefrontal cortex. *J Neurophysiol* 100: 1407-19

Mikl M, Marecek R, Hlustik P, Pavlicova M, Drastich A, et al. 2008. Effects of spatial smoothing on fMRI group inferences. *Magn Reson Imaging* 26: 490-503

Monti MM. 2011. Statistical Analysis of fMRI Time-Series: A Critical Review of the GLM Approach. *Front Hum Neurosci* 5: 28

Mosher JC, Leahy RM, Lewis PS. 1999. EEG and MEG: forward solutions for inverse methods. *IEEE Trans Biomed Eng* 46: 245-59

Mukamel R, Gelbard H, Arieli A, Hasson U, Fried I, Malach R. 2005. Coupling between neuronal firing, field potentials, and FMRI in human auditory cortex. *Science* 309: 951-4

Muller MM, Bosch J, Elbert T, Kreiter A, Sosa MV, et al. 1996. Visually induced gamma-band responses in human electroencephalographic activity--a link to animal studies. *Exp Brain Res* 112: 96-102

Muthukumaraswamy SD, Singh KD, Swettenham JB, Jones DK. 2010. Visual gamma oscillations and evoked responses: variability, repeatability and structural MRI correlates. *Neuroimage* 49: 3349-57

Naselaris T, Prenger RJ, Kay KN, Oliver M, Gallant JL. 2009. Bayesian reconstruction of natural images from human brain activity. *Neuron* 63: 902-15

Nastase SA, Gazzola V, Hasson U, Keysers C. 2019. Measuring shared responses across subjects using intersubject correlation. *Soc Cogn Affect Neurosci* 14: 667-85

Nishida S, Nishimoto S. 2018. Decoding naturalistic experiences from human brain activity via distributed representations of words. *Neuroimage* 180: 232-42

Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu B, Gallant JL. 2011. Reconstructing visual experiences from brain activity evoked by natural movies. *Curr Biol* 21: 1641-6

Nonaka S, Majima K, Aoki SC, Kamitani Y. 2021. Brain hierarchy score: Which deep neural networks are hierarchically brain-like? *iScience* 24: 103013

Norman JF, Wiesemann EY. 2007. Aging and the perception of local surface orientation from optical patterns of shading and specular highlights. *Percept Psychophys* 69: 23-31

Oliva A, Torralba A. 2007. The role of context in object recognition. *Trends Cogn Sci* 11: 520-7

Oosterhof NN, Connolly AC, Haxby JV. 2016. CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. *Front Neuroinform* 10: 27

Parker AJ. 2007. Binocular depth perception and the cerebral cortex. *Nat Rev Neurosci* 8: 379-91

Peltier SJ, Lisinski JM, Noll DC, LaConte SM. 2009. Support vector machine classification of complex fMRI data. *Annu Int Conf IEEE Eng Med Biol Soc* 2009: 5381-4

Penny WD, Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE. 2011. *Statistical parametric mapping: the analysis of functional brain images*. Elsevier.

Perry G, Randle JM, Koelewijn L, Routley BC, Singh KD. 2015. Linear tuning of gamma amplitude and frequency to luminance contrast: evidence from a continuous mapping paradigm. *PLoS One* 10: e0124798

Peuskens H, Claeys KG, Todd JT, Norman JF, Van Hecke P, Orban GA. 2004. Attention to 3-D shape, 3-D motion, and texture in 3-D structure from motion displays. *J Cogn Neurosci* 16: 665-82

Porcaro C, Ostwald D, Hadjipapas A, Barnes GR, Bagshaw AP. 2011. The relationship between the visual evoked potential and the gamma band investigated by blind and semi-blind methods. *Neuroimage* 56: 1059-71

Poulsen AT, Kamronn S, Dmochowski J, Parra LC, Hansen LK. 2017. EEG in the classroom: Synchronised neural recordings during video presentation. *Sci Rep* 7: 43916

Preston TJ, Li S, Kourtzi Z, Welchman AE. 2008. Multivoxel pattern selectivity for perceptually relevant binocular disparities in the human brain. *J Neurosci* 28: 11315-27

Purves D AG, Fitzpatrick D. 2001. The Columnar Organization of the Striate Cortex In *Neuroscience*

Rakhimberdina Z, Jodelet Q, Liu X, Murata T. 2021. Natural Image Reconstruction From fMRI Using Deep Learning: A Survey. *Front Neurosci* 15: 795488

Ratcliff R, Philiastides MG, Sajda P. 2009. Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. *Proc Natl Acad Sci U S A* 106: 6539-44

Riesenhuber M, Poggio T. 1999. Hierarchical models of object recognition in cortex. *Nat Neurosci* 2: 1019-25

Riesenhuber M, Poggio T. 2000. Models of object recognition. *Nat Neurosci* 3 Suppl: 1199-204

Ruff DA, Cohen MR. 2014. Attention can either increase or decrease spike count correlations in visual cortex. *Nat Neurosci* 17: 1591-7

Saad ZS, Reynolds RC. 2012. Suma. *Neuroimage* 62: 768-73

Saenz M, Langers DR. 2014. Tonotopic mapping of human auditory cortex. *Hear Res* 307: 42-52

Salmelin R, Baillet S. 2009. Electromagnetic brain imaging. *Hum Brain Mapp* 30: 1753-7

Sauseng P, Klimesch W, Stadler W, Schabus M, Doppelmayr M, et al. 2005. A shift of visual spatial attention is selectively associated with human EEG alpha activity. *Eur J Neurosci* 22: 2917-26

Schoffelen JM, Gross J. 2009. Source connectivity analysis with MEG and EEG. *Hum Brain Mapp* 30: 1857-65

Schonwiesner M, Zatorre RJ. 2009. Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. *Proc Natl Acad Sci U S A* 106: 14611-6

Serre T, Wolf L, Bileschi S, Riesenhuber M, Poggio T. 2007. Robust object recognition with cortex-like mechanisms. *IEEE Trans Pattern Anal Mach Intell* 29: 411-26

Simony E, Honey CJ, Chen J, Lositsky O, Yeshurun Y, et al. 2016. Dynamic reconfiguration of the default mode network during narrative comprehension. *Nat Commun* 7: 12141

Singh SP. 2014. Magnetoencephalography: Basic principles. *Ann Indian Acad Neurol* 17: S107-12

Song S, Zhan Z, Long Z, Zhang J, Yao L. 2011. Comparative study of SVM methods combined with voxel selection for object category classification on fMRI data. *PLoS One* 6: e17191

Sonkusare S, Breakspear M, Guo C. 2019. Naturalistic Stimuli in Neuroscience: Critically Acclaimed. *Trends Cogn Sci* 23: 699-714

Stringer EA, Chen LM, Friedman RM, Gatenby C, Gore JC. 2011. Differentiation of somatosensory cortices by high-resolution fMRI at 7 T. *Neuroimage* 54: 1012-20

Tadel F, Baillet S, Mosher JC, Pantazis D, Leahy RM. 2011a. Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput Intell Neurosci* 2011: 879716

Tadel F, Baillet S, Mosher JC, Pantazis D, Leahy RM. 2011b. Brainstorm: A User-Friendly Application for MEG/EEG Analysis. *Computational Intelligence and Neuroscience* 2011: 879716

Tahmasebi AM, Abolmaesumi P, Zheng ZZ, Munhall KG, Johnsrude IS. 2009. Reducing inter-subject anatomical variation: effect of normalization method on sensitivity of functional magnetic resonance imaging data analysis in auditory cortex and the superior temporal region. *Neuroimage* 47: 1522-31

Tan HM, Gross J, Uhlhaas PJ. 2016. MEG sensor and source measures of visually induced gamma-band oscillations are highly reliable. *Neuroimage* 137: 34-44

Tittle JS, Perotti VJ. 1997. The perception of shape and curvedness from binocular stereopsis and structure from motion. *Percept Psychophys* 59: 1167-79

Tsunoda K, Yamane Y, Nishizaki M, Tanifuji M. 2001. Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nat Neurosci* 4: 832-8

Umeda K, Tanabe S, Fujita I. 2007. Representation of stereoscopic depth based on relative disparity in macaque area V4. *J Neurophysiol* 98: 241-52

Vanderwal T, Eilbott J, Finn ES, Craddock RC, Turnbull A, Castellanos FX. 2017. Individual differences in functional connectivity during naturalistic viewing conditions. *Neuroimage* 157: 521-30

Vanderwal T, Kelly C, Eilbott J, Mayes LC, Castellanos FX. 2015. Inscapes: A movie paradigm to improve compliance in functional magnetic resonance imaging. *Neuroimage* 122: 222-32

Verhoef BE, Vogels R, Janssen P. 2016. Binocular depth processing in the ventral visual pathway. *Philos Trans R Soc Lond B Biol Sci* 371

von Bartheld CS, Bahney J, Herculano-Houzel S. 2016. The search for true numbers of neurons and glial cells in the human brain: A review of 150 years of cell counting. *J Comp Neurol* 524: 3865-95

Vorwerk J, Cho JH, Rampp S, Hamer H, Knosche TR, Wolters CH. 2014. A guideline for head volume conductor modeling in EEG and MEG. *Neuroimage* 100: 590-607

Wandell BA, Winawer J. 2011. Imaging retinotopic maps in the human brain. *Vision Res* 51: 718-37

Wang L, Mruczek RE, Arcaro MJ, Kastner S. 2015. Probabilistic Maps of Visual Topography in Human Cortex. *Cereb Cortex* 25: 3911-31

Wardle SG, Kriegeskorte N, Grootswagers T, Khaligh-Razavi SM, Carlson TA. 2016. Perceptual similarity of visual patterns predicts dynamic neural activation patterns measured with MEG. *Neuroimage* 132: 59-70

Wen H, Shi J, Chen W, Liu Z. 2018. Deep Residual Network Predicts Cortical Representation and Organization of Visual Features for Rapid Categorization. *Sci Rep* 8: 3752

Worsley KJ, Friston KJ. 1995. Analysis of fMRI time-series revisited--again. *Neuroimage* 2: 173-81

Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC. 1996. A unified statistical approach for determining significant signals in images of cerebral activation. *Hum Brain Mapp* 4: 58-73

Xie S, Kaiser D, Cichy RM. 2020. Visual Imagery and Perception Share Neural Representations in the Alpha Frequency Band. *Curr Biol* 30: 3062

Yacoub E, Shmuel A, Logothetis N, Ugurbil K. 2007. Robust detection of ocular dominance columns in humans using Hahn Spin Echo BOLD functional MRI at 7 Tesla. *Neuroimage* 37: 1161-77

Yuval-Greenberg S, Tomer O, Keren AS, Nelken I, Deouell LY. 2008. Transient induced gamma-band response in EEG as a manifestation of miniature saccades. *Neuron* 58: 429-41

Zeki S. 1991. Cerebral akinetopsia (visual motion blindness). A review. *Brain* 114 (Pt 2): 811-24

Zeng LL, Shen H, Liu L, Hu D. 2014. Unsupervised classification of major depression using functional connectivity MRI. *Hum Brain Mapp* 35: 1630-41

Zhang J, Zhang G, Li X, Wang P, Wang B, Liu B. 2020. Decoding sound categories based on whole-brain functional connectivity patterns. *Brain Imaging Behav* 14: 100-09