bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Title: Multi-Connection Pattern Analysis: Decoding the Representational Content of Neural Communication

Authors: Yuanning Li[1,2,3*] and Avniel Singh Ghuman[1,2,3]

Affiliations:
[1]Center for the Neural Basis of Cognition.
[2]Program in Neural Computation, Carnegie Mellon University and University of Pittsburgh.
[3]University of Pittsburgh, Department of Neurological Surgery.

***Correspondence to**:
Yuanning Li
S906 Scaife Hall
3550 Terrace Street
Pittsburgh, PA 15261
ynli@cmu.edu

Keywords: multi-connection pattern analysis, functional connectivity, multivariate analysis, decoding, representation similarity analysis, intracranial electroencephalography (iEEG), occipital face area, fusiform face area

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

It is unknown what information is represented in distributed brain circuit interactions because we lack methods for decoding the representational content of interregional neural communication. Here we present Multi-Connection Pattern Analysis (MCPA), which is designed to probe the nature of the representational space contained in the multivariate functional connectivity pattern between neural populations. MCPA works by learning mappings between the activity patterns of the populations separately for each condition, stimulus, or brain state in training data. These maps are used to predict the activity from one neural population based on the activity from the other population as a factor of the information being processed in test data. Successful MCPA-based decoding indicates the involvement of distributed computational processing and provides a framework for probing the representational structure of the interaction. Simulations demonstrate the efficacy of MCPA for decoding distributed information processing across a set of realistic circumstances and show that MCPA is insensitive to local information processing. Furthermore, applying MCPA to human intracranial electrophysiological data demonstrates that the interaction between occipital face area and fusiform face area contains information about individual faces. Representational analysis indicates that the OFA-FFA interaction codes face information differently than either the OFA or FFA individually. These results support the hypothesis that face individuation occurs not in a single region, but through interactive computation and distributed information representation across the face processing network. Thus, MCPA, can be used to assess the information processed the coupled activity of distributed, interacting neural circuits.

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Significance Statement

Information is represented in the brain by the coordinated activity of neurons both at the regional level and the level of large-scale, distributed networks. Multivariate methods from machine learning have advanced our understanding of the representational structure of local information coding, but the nature of distributed information representation remains unknown. Here we present a novel method that integrates multivariate connectivity analysis with machine learning classification techniques that can be used to decode the representational structure of neural interactions. This method is then used to provide a novel neuroscientific insight, that information about individual faces is distributed across at least two critical nodes of the face-processing network. Thus, this work provides a framework to assess the representational content of circuit-level processing.

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Since at least the seminal studies of Hubel and Wiesel (1), the computational role that neurons and neural populations play in processing has defined, and has been defined by, how they are tuned to represent information. The classical approach to address this question has been to determine how the activity recorded from different neurons or neural populations varies in response to parametric changes of the information being processed. Single unit studies have revealed tuning curves for neurons from different areas in the visual system responsive to features ranging from the orientation of a line, shapes, and even high level properties such as eye, nose, and mouth properties of the face (1-3) Multivariate methods, especially pattern classification methods from modern statistics and machine learning, such as multivariate pattern analysis (MVPA), have gained popularity in recent years and have been used to study neural population tuning and the information represented via population coding in neuroimaging and multiunit activity (4-11). These methods allow one to go beyond examining involvement in a particular neural process by probing the nature of the representational space contained in the pattern of population activity(12-14)

Neural populations do not act in isolation, rather the brain is highly interconnected and cognitive processes occur through the interaction of multiple populations. Indeed, many models of neural processing suggest that information is not represented solely in the activity of local neural populations, but rather at the level of recurrent interactions between regions (15-20). However previous studies only focused on the information representation within a specific population (3, 5, 11, 21-24), as no current multivariate methods allow one to directly assess what information is represented

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in the pattern of functional connections between distinct and interacting neural populations. Such a method would allow one to assess the content and organization of the information represented in the neural interaction. Thus, it remains unknown whether functional connections passively transfer information between encapsulated modules (25) or whether these interactions play an active computational role in processing. For example, the interaction could bind the information represented in distinct populations together (e.g. visual letter form and phonetic pronunciation) or allow for the computational transformations between distinct representations of a particular type of information (e.g. transforming a face image between viewpoint-dependent to viewpoint-independent representations in service of identity recognition).

Neural interactions are assessed based on functional and effective connectivity where the strength of the relationship between the activity from different populations is quantified as a factor of a particular neural process or disease state (26-28). Functional and effective connectivity can be determined by number of methods, such as ones based on correlation, causality, mutual information, etc. However, traditional methods for assessing neural connectivity do not allow one to go beyond determining if circuits are communicating in a particular cognitive state to probe the information represented in the interaction.

In this paper, we introduce a multivariate analysis method combining functional connectivity and pattern recognition analyses that we term Multi-Connection Pattern Analysis (MCPA). MCPA works by learning the discriminant information represented in the shared activity between distinct neural populations by combining multivariate correlational methods with pattern classification techniques from machine learning in a

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

novel way. The MCPA method consists of an integrated process of learning connectivity maps based on the pattern of coupled activity between two populations A and B and using these maps to classify the information representation in shared activity between A and B in test data. The rationale for MCPA is that if the activity in one area can be predicted based on the activity in the other area and the mapping that allows for this prediction is sensitive to the information being processed, then the areas are communicating with one another and the nature of the communication is sensitive to the type of information being processed. This is operationalized by learning a connectivity map by finding the mapping that maximizes the multivariate correlation between the activities of the two populations in each condition. This map can be thought of like the regression weights that transform the activity pattern in area A to the activity pattern in area B (properly termed "canonical coefficients" because a canonical correlation analysis [CCA] is used to learn the map). These maps are then used to generate the predictions as part of the classification algorithm. Specifically, a prediction of the activity pattern in one region is generated for each condition based on the activity pattern in the other region projected through each mapping. Single trial classification is achieved by comparing these predicted activity patterns with the true activity pattern (see Figure 1 for illustration). With this framework, single trial classification based on multivariate functional connectivity patterns is achieved allowing the nature of the representational space of the interaction to be probed.

To validate MCPA we first demonstrate the ability to detect the information represented in the functional connectivity patterns for data with realistic signal-to-noise ratios (SNR). We then present a number of control cases to show that classification of the

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

information contained in the functional connectivity between areas is independent of the local information contained within each area. Finally, we demonstrate the utility of MCPA by using it to examine the circuit-level representation for faces using intracranial electroencephalography (iEEG) data. Specifically, we show that the interaction between the occipital face area (OFA) and the fusiform face area (FFA), areas previously shown to be anatomically and functionally connected (29-31), represents information about individual faces. These results demonstrate that MCPA can be used to probe the nature of representational space resulting from information processing distributed across neural regions.

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

Overview

The MCPA method consists of learning phase and a test phase (as in machine learning, where a model is first learned, then tested). In the learning phase, the connectivity maps for each condition that characterize the pattern of shared activity between two populations is learned. In the test phase, these maps are used to generate predictions of the activity in one population based on the activity in the other population as a factor of condition and these predictions are tested against the true activity in the two populations. Similar to linear regression where one can generate a prediction for the single variable A given the single variable B based on the line that correlates A and B, MCPA employs a canonical correlation model (a generalization of multivariate linear regression) and produces a mapping model for each condition as a hyperplane that correlates multidimensional spaces A and B. Thus one can generate a prediction of the observation in multivariate space A given the observation in multivariate space B on a single trials basis. In this sense, MCPA is more analogous to a machine learning classifier combined with a multivariate extension of the psychophysiological index (32) rather than being analogous to correlation-based functional connectivity measures.

The general framework of MCPA is to learn the connectivity map between the populations for each task or stimulus condition separately based on training data. Specifically, given two neural populations (referred to as A and B), the neural activity of the two populations can be represented by feature vectors in multi-dimensional spaces (14). The actual physical meaning of the vectors would vary depending on modality, for example spike counts for a population of single unit recordings; time point features for

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

event-related potentials (ERP) or event-related fields; time-frequency features for electroencephalography, electrocorticography (ECoG) or magnetoencephalography; or single voxel blood-oxygen-level dependent responses for functional magnetic resonance imaging. A mapping between A and B is calculated based on any shared information between them for each condition on the training subset of the data. This mapping can be any kind of linear transformation, such as any combination of projections, scalings, rotations, reflections, shears, or squeezes.

These mappings are then tested as to their sensitivity to the differential information being processed between cognitive conditions by determining if the neural activity can be classified based on the mappings. Specifically, for each new test data trial, the maps are used to predict the neural activity in one area based on the activity in the other area and these predictions are compared to the true condition of the data. The trained information-mapping model that fits the data better is selected and the trial is classified into the corresponding condition. This allows one to test whether the mappings were sensitive to the differential information being represented in the neural interaction in the two conditions.

Mathematically, MCPA involves first calculating the cross-correlation matrix for each condition and reducing its dimensionality using the canonical correlation algorithm. Then multivariate prediction and classification are performed on the data in these reduced, cross-correlation spaces.

Connectivity Map

The first phase of MCPA is to build the connectivity map between populations. Here we followed a similar approach as the method proposed by Brookes et al. (33). The

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

neural signal in each population can be decomposed into two parts: the part that encodes shared information, and the part that encodes non-shared local information (including any measurement noise). We assume that the parts of the neural activities that represent the shared information in the two populations are linearly correlated (though, this can easily be extended by the introduction of a non-linear kernel). The model can be described as follows

$$\boldsymbol{C} \sim \mathcal{N}(0, \boldsymbol{I}_d), \min\{m_A, m_B\} \geq d \geq 1$$

$$\boldsymbol{A}|\boldsymbol{C} = \boldsymbol{W}_A\boldsymbol{C} + \boldsymbol{D}, \boldsymbol{D} \sim \mathcal{N}(\boldsymbol{\mu}_A, \boldsymbol{\Psi}_A), \boldsymbol{W}_A \in \mathbb{R}^{m_A \times d}, \boldsymbol{\Psi}_A \succcurlyeq 0$$

$$\boldsymbol{B}|\boldsymbol{C} = \boldsymbol{W}_B\boldsymbol{C} + \boldsymbol{E}, \boldsymbol{E} \sim \mathcal{N}(\boldsymbol{\mu}_B, \boldsymbol{\Psi}_B), \boldsymbol{W}_B \in \mathbb{R}^{m_B \times d}, \boldsymbol{\Psi}_B \succcurlyeq 0$$

where $\boldsymbol{C}$ is the common activity, $\boldsymbol{D}$ and $\boldsymbol{E}$ are local activities, m_A, m_B are the dimensionalities of activity vector in population A and B respectively. Without loss of generality, $\boldsymbol{\mu}_A = \boldsymbol{\mu}_B = 0$ can be assumed. The activity in population A can be decomposed into shared activity $\boldsymbol{W}_A\boldsymbol{C}$ and local activity $\boldsymbol{D}$, while activity in B can be decomposed into shared activity $\boldsymbol{W}_B\boldsymbol{C}$ and local activity $\boldsymbol{E}$. More importantly, the shared discriminant information only lies in the mapping matrix $\boldsymbol{W}_A$ and $\boldsymbol{W}_B$ since $\boldsymbol{C}$ always follows the standard multivariate normal distribution (though correlation measures that do not assume normally distributed data can also be applied with minor modifications to the calculation).

In statistics, canonical correlation analysis (CCA) can be used to fit such model and estimate the linear mappings (34, 35). In brief, let S be the covariance matrix

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{S}_{AA} & \boldsymbol{S}_{AB} \\ \boldsymbol{S}_{BA} & \boldsymbol{S}_{BB} \end{bmatrix} = \mathbb{E}\left[\begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{pmatrix} \begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{pmatrix}^T\right]$$

Therefore $\boldsymbol{W}_A$ and $\boldsymbol{W}_B$ can be estimated by solving the following eigen problem

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$\begin{cases} \boldsymbol{S}_{AA}^{-1}\boldsymbol{S}_{AB}\boldsymbol{S}_{BB}^{-1}\boldsymbol{S}_{BA}\widehat{\boldsymbol{U}}_A = \rho^2\widehat{\boldsymbol{U}}_A \\ \boldsymbol{S}_{BB}^{-1}\boldsymbol{S}_{BA}\boldsymbol{S}_{AA}^{-1}\boldsymbol{S}_{AB}\widehat{\boldsymbol{U}}_B = \rho^2\widehat{\boldsymbol{U}}_B \end{cases}$$

and we have

$$\begin{cases} \widehat{\boldsymbol{W}}_A = \boldsymbol{S}_{AA}\widehat{\boldsymbol{U}}_{Ad}M_1 \\ \widehat{\boldsymbol{W}}_B = \boldsymbol{S}_{BB}\widehat{\boldsymbol{U}}_{Bd}M_2 \end{cases}$$

where $\widehat{\boldsymbol{U}}_{Ad}$ and $\widehat{\boldsymbol{U}}_{Bd}$ are the first d columns of canonical directions $\widehat{\boldsymbol{U}}_A$ and $\widehat{\boldsymbol{U}}_B$, and M_1, $M_2 \in \mathbb{R}^{d\times d}$ are arbitrary matrices such that $\boldsymbol{M}_1\boldsymbol{M}_2^T = \boldsymbol{P}_d$, $\boldsymbol{P}_d$ is the diagonal matrix with the first d elements of $\boldsymbol{P} = \boldsymbol{U}_B^T\boldsymbol{S}_{BA}\boldsymbol{U}_A$.

With $\boldsymbol{W}_A$ and $\boldsymbol{W}_B$, the shared information $\boldsymbol{C}$ can be estimated using its posterior mean $\mathbb{E}(\boldsymbol{C}|\boldsymbol{A})$ and $\mathbb{E}(\boldsymbol{C}|\boldsymbol{B})$, where $\mathbb{E}(\boldsymbol{C}|\boldsymbol{A}) = \boldsymbol{M}_1^T\boldsymbol{U}_A^T\boldsymbol{A}$ and $\mathbb{E}(\boldsymbol{C}|\boldsymbol{B}) = \boldsymbol{M}_2^T\boldsymbol{U}_B^T\boldsymbol{B}$. Let $\boldsymbol{M}_1 = \boldsymbol{M}_2$ and equate $\mathbb{E}(\boldsymbol{C}|\boldsymbol{A})$ and $\mathbb{E}(\boldsymbol{C}|\boldsymbol{B})$, this shared information can be used as a relay to build the bidirectional mapping between A and B. Specifically, $\widehat{\boldsymbol{B}} = (\boldsymbol{M}_2^T\boldsymbol{U}_B^T)^\dagger\boldsymbol{M}_1^T\boldsymbol{U}_A^T\boldsymbol{A} = \boldsymbol{U}_B^{T\dagger}\boldsymbol{U}_A^T\boldsymbol{A} = \boldsymbol{R}\boldsymbol{A}$ and $\widehat{\boldsymbol{A}} = (\boldsymbol{M}_1^T\boldsymbol{U}_A^T)^\dagger\boldsymbol{M}_2^T\boldsymbol{U}_B^T\boldsymbol{B} = \boldsymbol{U}_A^{T\dagger}\boldsymbol{U}_B^T\boldsymbol{B} = \boldsymbol{R}^\dagger\boldsymbol{B}$, where $\boldsymbol{R} = \boldsymbol{U}_B^{T\dagger}\boldsymbol{U}_A^T\boldsymbol{A}$.

In the first step, the connectivity map is estimated for each condition separately. If we have n_1 trials in condition 1 and n_2 trials in condition 2 in the training set, the training data for the two conditions are represented in matrices as $\left[\boldsymbol{X}_A^{(1)}, \boldsymbol{X}_B^{(1)}\right]^T$ and $\left[\boldsymbol{X}_A^{(2)}, \boldsymbol{X}_B^{(2)}\right]^T$ respectively, where $\boldsymbol{X}_A^{(1)} \in \mathbb{R}^{m_A\times n_1}$, $\boldsymbol{X}_B^{(1)} \in \mathbb{R}^{m_B\times n_1}$ are the population activity for A and B under condition 1 respectively, and $\boldsymbol{X}_A^{(2)} \in \mathbb{R}^{m_A\times n_2}$, $\boldsymbol{X}_B^{(2)} \in \mathbb{R}^{m_B\times n_2}$ are the population activity for A and B under condition 2 respectively. The testing data vector is then represented as $[\boldsymbol{x}_A, \boldsymbol{x}_B]^T$, where $\boldsymbol{x}_{\square} \in \mathbb{R}^{m_A}$ and $\boldsymbol{x}_B \in \mathbb{R}^{m_B}$ are population activities in A and B respectively. Using CCA, the estimations of the mapping matrices with respect to different conditions are $\boldsymbol{R}^{(1)}$ and $\boldsymbol{R}^{(2)}$.

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To sum up, by building the connectivity map, a linear mapping function $\boldsymbol{R}$ is estimated from the data for each condition so that the activity of the two populations can be directly linked through bidirectional functional connectivity that captures only the shared information.

Classification

The second phase of MCPA is a pattern classifier that takes in the activity from one population and predicts the activity in a second population based on the learned connectivity maps conditioned upon the stimulus condition or cognitive state. The testing data is classified into the condition to which the corresponding model most accurately predicts the true activity in the second population.

The activity from one population is projected to another using the learned CCA model, i.e. $\boldsymbol{x}_B^{(i)} = \boldsymbol{U}_B^{(i)\dagger}\boldsymbol{U}_A^{(i)}\boldsymbol{x}_{\boldsymbol{A}}$. The predicted projections $\boldsymbol{x}_B^{(i)}$ are compared to the real observation $\boldsymbol{x}_{\boldsymbol{B}}$, and then the testing trial is labeled to the condition where the predicted and real data match most closely. Any similarity metric could be used for this comparison; here cosine similarity (correlation) is used. The mapping is bidirectional, so A can be projected to B and vice versa. In practice, the similarities from the two directions are averaged in order to find the condition that gives maximum average correlation coefficient. Specifically, $corr_B^{(i)} = \frac{\langle \boldsymbol{x}_B, \boldsymbol{x}_B^{condi} \rangle}{|\boldsymbol{x}_B||\boldsymbol{x}_B^{condi}|}$, $corr_A^{(i)} = \frac{\langle \boldsymbol{x}_A, \boldsymbol{x}_A^{condi} \rangle}{|\boldsymbol{x}_A||\boldsymbol{x}_A^{condi}|}$ and $predict\ label = \arg\max_i \left\{\frac{corr_A^{(i)} + corr_B^{(i)}}{2}\right\}$.

Simulated data

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To test the performance of MCPA, we simulated shared and local activity in two populations and tested the performance of MCPA on synthetic data as a factor of the number of dimensions in each population and signal-to-noise ratio (SNR; figure 2a). The shared activity for both conditions in population A was drawn independently from a *d*-dimensional normal distribution $\boldsymbol{Y}_A^{(i)} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}_d)$, for $i = 1,2$. The shared activity in population B under two different conditions were generated by rotating $\boldsymbol{Y}_A$ with different rotation matrices separately, $\boldsymbol{Y}_B^{(i)} = \boldsymbol{R}^{(i)} \boldsymbol{Y}_A^{(i)}$, where $\boldsymbol{R}^{(1)}$ and $\boldsymbol{R}^{(2)}$ were two *d*-by-*d* random rotation matrices corresponding to the information mapping under condition 1 and 2 respectively, and $\boldsymbol{R}^{(i)^{\boldsymbol{T}}} \boldsymbol{R}^{(i)} = \boldsymbol{I}_d$.

The two important parameters here are the dimensionality *d* and the variance σ^2. SNR was used to characterize the ratio between the variance of shared activity and variance of local activity, and the logarithmic decibel scale $\mathrm{SNR}_{dB} = 10 \log_{10}(\sigma^2)$ was used. To cover the wide range of possible data recorded from different brain regions and different measurement modalities, we tested the performance of MCPA with *d* ranging from 2 to 25 and SNR ranging from -20 dB to 20 dB (σ^2 ranged from 0.01 to 100). Note that each of the *d* dimensions contain independent information about the conditions though have the same SNR. Thus the overall SNR does not change, but the amount of pooled information does change with *d*. For each particular setup of parameters, the rotation matrices $\boldsymbol{R}^{(i)}$ were randomly generated first, then 200 trials were randomly sampled for each condition and evenly split into training set and testing set. MCPA was trained using the training set and tested on the testing set to estimate the corresponding true positive rate (TPR) and false positive rate (FPR) for the binary classification. The sensitivity index *d'* was then calculated as $d' = Z(TPR) - Z(FPR)$, where $Z(x)$ is the

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

inverse function of the cdf of standard normal distribution. This process was repeated 100 times and the mean and standard errors across these 100 simulations were calculated. Note that the only discriminant information about the two conditions is the pattern of interactions between the two populations, and neither of the two populations contains local discriminant information about the two conditions in its own activity. We further tested and confirmed this by trying to classify the local activity in populations A and B (see below). To avoid an infinity *d'* value, with 100 testing trials, the maximum and minimum for TPR or FRP were set to be 0.99 and 0.01, which made the maximum possible *d'* to be 4.65.

The MCPA method captures the pattern of correlation between neural activities from populations and is invariant to the discriminant information encoded in local covariance. To see this, we first take the simulation data described above and apply MVPA (naïve Bayes) to each of the two populations separately. Note that in each of the two populations, we set the two conditions to have the same mean and covariance. As a result there should be no local discriminant information within any of the two populations alone.

Control simulations

In addition to the MVPA control described above, we further evaluated the following three control experiments to demonstrate that MCPA is insensitive to the presence or change in the local information. In the first control experiment (no functional connectivity, no shared information, varying local information), we simulated the case where two populations are totally independent under both conditions, but there is local

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

discriminant information in each (figure 2b). Specifically, for condition 1, $\boldsymbol{X}_A^{(1)}, \boldsymbol{X}_B^{(1)}$ were drawn independently from the same distribution $\mathcal{N}(0, \boldsymbol{I}_d)$; for condition 2, $\boldsymbol{X}_A^{(2)}, \boldsymbol{X}_B^{(2)}$ were drawn independently from the same distribution $\mathcal{N}(0, \boldsymbol{I}_d)$. Then we changed the local variance in one of the conditions. For the features in population A and B under condition 1, we used $\boldsymbol{X}_A^{(1)'} = k\boldsymbol{X}_A^{(1)}$ and $\boldsymbol{X}_B^{(1)'} = k\boldsymbol{X}_B^{(1)}$, where k ranged from 1 to 9. Thus, in both populations, the variance of condition 1 was different from the variance of condition 2, and such difference would increase as k became larger. Therefore, there was no information shared between the two populations under either condition, but each of the population had discriminant information about the conditions encoded in the variance for any $k \neq 1$.

In the second control experiment (functional connectivity, constant shared information, varying local information), we introduced local discriminant information into population A without changing the amount of shared information between populations A and B (figure 2c). We fixed the dimensionality at 10 and SNR at 0 dB ($\sigma^2 = 1$) and kept the rotation matrices of different conditions different from each other. As a result, the amount of shared discriminant information represented in the patterns of interactions stayed constant. Then we changed the local variance in one of the conditions. For the features in population A under condition 1, we used $\boldsymbol{X}_A^{(1)'} = k\boldsymbol{X}_A^{(1)}$, where k ranged from 1 to 9. Thus, population A, the variance of condition 1 was different from the variance of condition 2, and such difference would increase as k became larger. According to our construction of MCPA, it should only pick up the discriminant information contained in the interactions and should be insensitive to the changes in local discriminant information from any of the two populations.

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In the third control experiment (functional connectivity, no shared information, varying local information), we eliminated the information represented in the pattern of interaction, but maintained the functional connectivity by keeping the correlation between populations invariant with regard to conditions. We introduced local discriminant information into the two populations to demonstrate that MCPA is insensitive to the presence of constantly correlated local information (figure 2d). We fixed the dimensionality at 10 and SNR at 0 dB ($\sigma^2 = 1$) and kept the rotation matrices constant for different conditions. As a result, the amount of shared discriminant information represented in the patterns of interactions was 0. Then we changed the local variance in one of the conditions. Then we changed the local variance in one of the conditions. For the features in population A and B under condition 1, we used $\boldsymbol{X}_A^{(1)'} = k\boldsymbol{X}_A^{(1)}$ and $\boldsymbol{X}_B^{(1)'} = k\boldsymbol{X}_B^{(1)}$, where k ranged from 1 to 9. Thus, in both populations, the variance of condition 1 was different from the variance of condition 2, and such difference would increase as k became larger. Notably, such local information was actually correlated through interactions between the populations. However, since the pattern of interaction did not vary as the condition changed, there was no discriminant information about the conditions represented in the interactions. According to our construction of MCPA, it should not pick up any discriminant information in this control case.

Examining OFA-FFA coding for individual faces using MCPA

Subject

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A human subject underwent surgical placement of iEEG depth electrodes (stereotactic electroencephalography) as standard of care for surgical epilepsy localization. The subject was male, age 56. There was no evidence of epileptic activity shown on the electrodes used in this study.

The experimental protocols were approved by the Institutional Review Board of the University of Pittsburgh. Written informed consent was obtained from the participant.

Stimuli

In Experiment 1, 180 images of faces (50% male), bodies (50% male), words, hammers, houses, and phase scrambled faces were used as a functional localizer. Each category contained 30 images. Phase scrambled faces were created in Matlab by taking the 2-dimensional spatial Fourier spectrum of each of the face images, extracting the phase, adding random phases, recombining the phase and amplitude, and taking the inverse 2-dimensional spatial Fourier spectrum. Each image was presented in pseudorandom order and repeated once in each session.

Faces in Experiment 2 were taken from the Karolinska Directed Emotional Faces stimulus set (18). Frontal views and 5 different facial expressions (happy, sad, angry, fearful, and neutral) from all 70 faces (50% male) in the database were used, which yielded a total of 350 face images, each presented once in random order during a session.

All stimuli were presented on an LCD computer screen placed approximately 2 meters from participants' heads.

Experimental paradigms

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In experiment 1, each image was presented for 900 ms with 900 ms inter-trial interval during which a fixation cross was presented at the center of the screen (~ 10˚ x 10˚of visual angle). At random, 25% of the time an image would be repeated. Participants were instructed to press a button on a button box when an image was repeated (1-back). Only the first presentations of repeated images were used in the analysis.

In experiment 2, each face was presented for 1500 ms with 500 ms inter-trial interval during which a fixation cross was presented at the center of the screen. Faces subtended approximately 5 degrees of visual angle in width. Subjects were instructed to report whether the face was male or female via button press on a button box.

Paradigms were programmed in Matlab™ using Psychtoolbox and custom written code.

Data preprocessing

The electrophysiological activity in OFA and FFA were recorded simultaneously using iEEG electrodes at 1000 Hz. They were subsequently bandpass filtered offline from 1-170 Hz using a fifth order Butterworth filter to remove slow and linear drift, the 180 Hz harmonic of the line noise, and high frequency noise. The 60 Hz line noise and the 120 Hz harmonic noise were removed using DFT filter. To reduce potential artifacts in the data, trials with maximum amplitude 5 standard deviations above the mean across the rest of the trials were eliminated. In addition, trials with a change of more than 25 μV between consecutive sampling points were eliminated. These criteria resulted in the elimination of less than 1% of trials.

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

As the last step of the data preprocessing, we extracted wavelet features using Morlet wavelets. The number of cycles of the wavelet was set to be 7. The entire epoch length of the data was 1500ms (-500 ~ 1000 ms relative to stimulus onset). To avoid numerical issues in MATLAB, the lowest frequency was set at 7 Hz. The wavelet features were estimated using FieldTrip™ toolbox. Finally, we took all the wavelet features at 7, 8, 9, …, 100 Hz at every 10 ms as features, which yielded a 94-dimensional feature vector at every time point. All the wavelets were normalized to the baseline by subtracting the mean value and divided by the standard deviation of the data from 350ms to 50ms before stimulus onset.

Electrode selection

Face sensitive electrodes were selected based on anatomical and functional considerations. Electrodes of interest were restricted to those that were located in or near the fusiform gyrus or inferior occipital cortex. In addition, MVPA was used to functionally select the electrodes that showed sensitivity to faces, comparing to other conditions in experiment 1. Specifically, electrodes were selected such that their peak 6-way classification *d'* score (see below for how this was calculated) exceeded 1 ($p < 0.001$ based on a permutation test, as described below) and the event related potential (ERP) for faces was larger than the ERP for the other non-face object categories.

There were 12 contacts on a depth electrode on the ventral temporal lobe extending along the anterior-posterior axis. Among all the contacts, only three (the 1st, 6th and 7th contacts, see figure 3a for the location of these contacts) satisfied the criterion described above (see Figure S1 for *d*' timecourses from all contacts on the depth

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

electrode). The first contact was near the mid-fusiform gyrus while the other two were near posterior end of the fusiform gyrus/anterior end of the inferior occipital cortex. Hence we used the data from the first electrode as FFA signal and the averaged data across the 6th and 7th electrodes as the OFA signal (see Figure S2 for averaged ERP data in the two areas). Notably, the post-operative structural MRI scan did not allow us to carefully distinguish the precise localization of the "OFA" electrodes and it may be that these electrodes are in fact in the posterior fusiform and properly labeled "FFA-1" according to the recent nomenclature introduced byWeiner et al. (36). However, considering OFA and FFA-1 are contiguous with one another and it has not been determined what, if any, functional distinction there is between the two, we use "OFA" for the label of the electrodes out of convenience.

MCPA Analysis

MCPA was applied to classify the OFA-FFA connectivity for each possible pair of faces (total of 2415 pairs). For each specific pair of faces, averaged wavelet features within a 50 ms time window were used as features in MCPA. Principal Component Analysis (PCA) was used to reduce the dimensionality from 94 to *P*, where *P* corresponds to the number of PCs that capture 95% of variation in the data, the typical value of *P* is around 7~8. Leave-one-trial-out cross-validation was used in order to estimate the classification accuracy. This procedure was repeated for all 2415 pairs and all time windows slid with 10 ms step between 0 and 600ms after stimulus onset. Similar to previous simulations, *d'* was used to quantify the performance of MCPA.

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Permutation test was used to determine the significance of the *d'* timecourse of MCPA (37). During each permutation, the condition labels of all the trials were randomly permuted and the same procedure as described above was used to calculate the timecourse of *d'* for each permutation. The permutation was repeated for a total of 200 times. The mean *d'* during 200-500 ms of each permutation was used as the test statistic and the null distribution of the test statistic was estimated using the histogram of the permutation test. The time window 200-500 ms was chosen based on the fact that the sensitivity of facial identity was only presented in OFA and FFA roughly 200 -500 ms after stimulus onset. (11)

MVPA Analysis

Similarly, MVPA was applied to classify the neural activity within OFA and FFA separately for each possible pair of faces (total of 2415 pairs). The same features extracted from OFA and FFA as described above were used in MVPA analysis. Naïve Bayes classifier was used as the linear classifier and leave-one-trial-out cross-validation was used in order to estimate the classification accuracy. This procedure was repeated for all 2415 pairs and all time windows slid with 10 ms step between 0 and 600 ms after stimulus onset. Similar to previous simulations, *d'* was used to quantify the performance of MVPA. Permutation test, as described above, was used to test for the significance of the classification accuracy.

Representational Similarity Analysis

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Based on the classification results, for each classification analysis (MCPA, MVPA on FFA, and MVPA on OFA), the dissimilarity matrix $\boldsymbol{M}$ was constructed such that the *j*th element in the *i*th row m_{ij} equals the dissimilarity (classification accuracy) between the face *i* and face *j* in the corresponding representational space defined by the analysis. The representation similarity between the representational space of the OFA-FFA interactions and the individual representational space of OFA and FFA is measured as the correlation coefficient (Pearson's linear correlation) between their corresponding dissimilarity matrices.

Results

Simulations

We used simulations to test and verify the performance and properties of MCPA on synthetic data. Specifically, synthetic data representing neural activity of two distinct populations and the information represented in the interaction between those populations were manipulated to construct different testing conditions. These tests included determining the sensitivity and specificity of MCPA with respect to SNR, demonstrating the insensitivity of MCPA to information only encoded locally, and showing that MCPA is specifically sensitive to the representational content of the interaction and not functional connectivity alone.

In the first simulation, we evaluated the ability of MCPA to detect information represented in the functional connectivity pattern when it was present as a factor of the SNR and the number of dimensions of the data. The mean and standard error of *d*' from 100 simulation runs for each particular setup (dimensionality and SNR) are shown in

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2a. The performance of the MCPA classifier increased when SNR or effective dimensionality increased. Classification accuracy saturated to the maximum when SNR and number of dimensions were high enough (SNR > 10 dB, dimensionality > 10). The performance of MCPA was significantly higher than chance ($p < 0.01$, permutation test) for SNRs above -5 dB ($\sigma^2 = 0.32$) for all cases where the dimensionality was higher than 2, when the pattern of the multivariate mapping between the activity was changed between conditions. It is notable that significant MCPA classification was seen despite there being no local information present in either of the two simulated populations ($p > 0.1$ for all SNRs and numbers of dimensions for MVPA on each individual activity pattern). Thus MCPA can decode the information represented in the neural interaction when the connectivity pattern has discriminant information, even when no local discriminant information is present in any of the local populations.

After confirming that MCPA is sensitive to information represented in neural communication even in the absence of local information under realistic circumstances, we ran two control simulations to demonstrate that MCPA is not sensitive to changes in local information.

The first control simulation was designed to confirm that when two unconnected populations both carry local discriminant information, MCPA would not be sensitive to that piece of information. As shown in Figure 2b, MCPA did not show any significant classification accuracy above chance ($d' = 0$) as k changed. On the other hand, the MVPA classifier that only took the data from local activity showed significant classification accuracy above chance level and the performance increased as local discriminant information increased.

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The second control simulation was designed to test if MCPA would be sensitive to changes in local discriminant information when there was constant information coded in neural communication. Local discriminant information was injected into the populations by varying the ratio of the standard deviation (k) between the two conditions.. When MVPA was applied to the local activity, increasing classification accuracy was seen as *k* became larger (figure 2c). This result confirmed that discriminant information was indeed encoded in the local activity in the simulation. On the other hand, the performance of MCPA did not change with the level of local discriminant information, demonstrating that MCPA is only sensitive to changes in information contained in neural interactions.

The final control simulation tested whether MCPA is simply sensitive to the presence of functional connectivity between two populations *per se* or is only sensitive to the whether the functional connectivity contains discriminant information. Specifically, are local discriminant information in two populations, and a correlation between their activity, sufficient for MCPA decoding? It should not be considering that MCPA requires that the pattern of the mapping between the populations to change as a factor of the information being processed (see figure 1). For example, the local activity in either or both populations could code for the information being processed, but the mapping between the activity in each region could be constant and insensitive to the changes in conditions, e.g. the CCA coefficients could be the same. This would be the case if each population was an informationally encapsulated module where information transfer occurs in the same way regardless of the stimulus being processed or cognitive state. In this case, one would not want to infer that distributed processing was taking place

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

because the nature of the interregional communication is not sensitive to the computation being performed (e.g. the information transfer is passive, rather than reflecting distributed computational processing) and all of the information processing is done locally in each population. The final control simulation was designed to assess whether MCPA is sensitive to the case where two populations communicate, but in a way that would not imply distributed computational processing. Specifically, neural activity in areas A and B were simulated local discrimination was possible in each population and the activity of the two populations was correlated, but the interaction between them was invariant to the information being processed. Figure 2d shows that in this case MCPA did not classify the activity above chance, despite significant correlation between the regions and significant local classification (MVPA). This control simulation demonstrates that indeed MCPA is only sensitive to the case where the mapping itself changes with respect to the information being processed, which is a test of the presence of distributed neural computation.

Taken together, the simulation results show that MCPA is sensitive to the discriminant information that is represented in the pattern of interactions between two populations, is insensitive to the discriminant information represented in unshared local activity, and the presence of functional connectivity and local information is not sufficient to infer distributed computational processing and produce significant MCPA decoding.

IEEG Data

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To assess its performance on real neural data, MCPA was applied on intracranial electroencephalography (iEEG) data recorded from OFA and FFA in one human epileptic patient during a visual cognition task (see Figure 3a for the electrode locations). MCPA was applied in the classification between each possible pair of faces. Previous studies on the timecourse of face individuation (11) have demonstrated that the 200-500 ms time window is critical for the processing of face individuation information. For MCPA, as shown in Figure 3, the classification accuracy was significantly above chance level across that time window (averaged $d' = 0.1491$, $p < 0.05$, permutation test). Using MVPA, classification accuracy was significantly above chance level across that time window in FFA (averaged $d' = 0.3604$, $p < 0.05$, permutation test), replicating previous reports (Ghuman et al. 2014), however classification accuracy did not reach chance level across that time window in OFA (averaged $d' = 0.1073$, $p > 0.1$, permutation test). On the other hand, in the early time window, which is 50 – 200 ms after stimulus onset, MCPA did not show significant classification accuracy (averaged $d' = 0.0711$, $p > 0.1$, permutation test).

As a control analysis, we took a contact outside of the fusiform gyrus that did not show face sensitivity and performed the same analysis between the control contact and the OFA and FFA contacts. As shown in Figure 3b, the averaged d' of MCPA between the control contact and both the OFA and FFA contacts was not significant above chance level ($d' = 0.0641$ for control & FFA, $d' = 0.0094$ for control & OFA, $p > 0.1$).

To assess whether the information represented in the OFA-FFA interaction reflected a distinct computational process or merely reflected the representation in either OFA or FFA, a representational similarity analysis was performed. Specifically, we calculated the representational similarity based on pairwise classification accuracy

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

between each pair of the 70 different faces using MVPA on the activity in OFA and FFA and MCPA on the interaction. No significant correlation was seen between the representational structure in OFA and the OFA-FFA interaction (Pearson's correlation coefficient $\rho = 0.0195$, $p > 0.1$) or between FFA and the OFA-FFA interaction (Pearson's correlation coefficient $\rho = 0.0176$, $p > 0.1$). Thus, despite both the FFA and the OFA-FFA interaction showing significant face decoding, little similarity between the representations reflected in those different aspects of the neural activity is seen. This suggests that the OFA-FFA interaction plays a computational role in face individuation that is distinct from the role that either the FFA or OFA play alone.

These results support the hypothesis individual level face information is represented in the OFA-FFA interaction pattern and that distributed computational processing across this circuit plays a critical role in face individuation.

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion

This paper presents a novel method to assess the information represented in the patterns of interactions between two neural populations. MCPA works by learning the mapping between the activity patterns from the populations from a training data set, and then classifying the neural communication pattern using these maps in a test data set. Simulated data demonstrated that MCPA was sensitive to information represented in neural interaction for realistic SNR ranges. Furthermore, MCPA is only sensitive to the discriminant information represented through different patterns of interactions irrespective of the information encoded in the local populations. Finally, we used this method to provide a novel neuroscientific insight: that the multivariate connectivity pattern between OFA and FFA represents information at the level of individual faces.

It is worth noting that significant discrimination within each population and significant functional connectivity between them is not sufficient to produce MCPA and indeed local classification within each population is not even necessary (Figures 2d and 2a respectively). MCPA requires the pattern of connectivity between the two populations to vary across the different conditions. As an example, if the two populations interact, but the interaction behaves like a passive filter, mapping the activity between the populations in a similar way in all conditions, MCPA would not be sensitive to the interaction because the mapping does not change (Figure 2d). Instead, MCPA is more akin to testing for adaptive filtering or distributed, interactive computation where the nature of the interaction changes depending on the information that is being processed. Recent studies demonstrate that neural populations in perceptual areas alter their response properties based on context, task demands, etc. (38). These modulations of response properties

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

suggest that lateral and long-distance interactions are adaptive and dynamic processes responsive to the type of information being processed and MCPA provides a platform for examining the role of interregional connectivity patterns in this adaptive process. Indeed, MCPA can be interpreted as testing whether distributed computational "work" is being done in the interaction between the two populations and the interaction does not just reflect a passive relay of information between two encapsulated modules (25).

In addition to allowing one to infer whether distributed computational work is being done in service of information processing, MCPA provides a platform for assessing its representational structure. Specifically, much as MVPA has been used in representational similarity analyses to measure the structure of the representational space at the level local neural populations (12, 13, 39), MCPA can be used to measure the structure of the representational space at the level of network interactions. Specifically, the representational geometry of the interaction can be mapped in terms of the similarity among the multivariate functional connectivity patterns corresponding to the brain states associated with varying input information. It is notable that this type of single-trial or single-stimulus representational similarity analysis is not possible by directly applying a classifier to functional connectivity features and requires learning the mapping between neural activity patterns, as described for MCPA (see next paragraph for further discussion of this distinction). The representational structure can be compared to behavioral measures of the structure to make brain-behavior inferences and assess what aspects of behavior a neural interaction contributes to. It can also be compared to models of the structure to test theoretical hypotheses regarding the computational role of the neural interaction (13, 39).

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

These two properties of MCPA, being able to assess distributed computational processing rather than just whether or not areas are communicating and being able to determine the representational structure of the information being processed, set MCPA apart from previously proposed multivariate functional connectivity methods. In these previous methods the functional connectivity calculation is performed separately from the classification calculation. Specifically, multivariate classification has been performed on a population of functional connectivity data (40-45) or classification is first performed in each region to extract the information and then the information represented in the two areas are correlated (13, 46). This separation of the connectivity and classification calculations precludes being able to assess distributed computational processes because these methods are sensitive to passive information exchange between encapsulated modules, as described above, and thus conflate passive and active information exchange. Critically, this separation also does not allow for single trial or single sample classification, as is required to perform the representational similarity analysis in a practical manner and decode how the information processed in the interaction is encoded and organized. As a concrete example, these previous methods would not be able to assess whether the interaction between OFA and FFA contain information about individual faces (Figure 3) and perform a representational similarity analysis.

Two examples of the types of studies that can be performed using MCPA help highlight the potential utility of this method. First, MCPA can be used to provide a strong test of the binding-by-synchrony hypothesis (47). This hypothesis asserts that information that results from computations arising from spatially distinct regions of the brain combine their information into a coherent representation through interregional synchronization.

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Thus far this hypothesis has primarily been tested by demonstrating increased interregional synchrony for conditions that show greater binding, such as showing greater synchrony when an image is perceived as a coherent Gestalt (48). However, increased synchrony could reflect a number of different effects, not all of which necessarily imply binding (49-51). MCPA provides a stronger test of the binding-by-synchrony hypothesis by allowing one to decode the representational content of the interaction. As a concrete example, one may be able to decode shape from certain regions of the brain (e.g. lateral occipital complex) and color from other regions (e.g. V4), but it is unclear how color and shape are bound into a coherent percept. One could certainly decode color and shape from pooled activity from V4 and the lateral occipital complex, but because these are distinct neural populations, the "AND" operation that binds color and shape still needs to be performed by the brain. If MCPA could be used to decode color and shape from the pattern of interactions between the regions, it would suggest that this "AND" operation is performed in the interregional synchrony pattern, consistent with the binding-by-synchrony hypothesis. If color and shape could not be decoded using MCPA, that would suggest a third region that acts as a convergence zone to bind these features that needs to be found, inconsistent with the binding-by-synchrony hypothesis.

A second potential use for MCPA is that a MCPA-based representational similarity analysis may help inform models of how representations are transformed between neural populations along a processing pathway (14). For example, when asked to verbally name a word being read, visual-orthographic representations must be transformed into phonological representations. This orthographic-phonological mapping implies a representational structure that is neither orthographic nor phonological, but

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

rather an intermediary between them. If a MCPA-based representational similarity analysis between regions responsible for orthographic and phonological computations were consistent with what would be predicted by an orthographic-phonological mapping, this would indicate that the mapping is performed through the interregional interactions between the areas. If not, it would suggest that another neural region should be found that is responsible for this mapping.

The specific instantiation of MCPA presented here treats connectivity as a bi-directional linear mapping between two populations. However, the MCPA framework could be easily generalized into more complicated cases. For example, instead of using correlation-based methods like CCA, other directed functional connectivity algorithms, such as Granger causality based on an autoregressive framework, could be used to examine directional interactions. Additionally, kernel methods, such as kernel CCA, could be applied to account for non-linear interactions. A more general framework would be to use non-parametric functional regression method to build a functional mapping between the two multidimensional spaces in the two populations. MCPA can also be expanded to look at network-level representation by implementing the multiset canonical correlation analysis, wherein the cross-correlation among multiple sets of activity patterns from different brain areas is calculated (52). MCPA could be used with a dual searchlight approach to examine whole brain, or multifeature communication (53). Also, MCPA could be optimized by regularizing the CCA algorithm to find the connectivity maps that uniquely describe, or at least best separate, the conditions of interest. Furthermore, both with and without these modification, the framework of MCPA may have a number of applications outside of assessing the representational content of

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

functional interactions in the brain, such as detecting the presence of distributed processing on a computer network, or examining genetic or proteomic interactions.

The current prevalent view is that face perception is mediated by a distributed network with multiple brain areas including the OFA and FFA. Structural and functional connectivity analysis for the core network has shown that FFA is strongly connected to OFA (29-31). While these results suggest the hypothesis that face individuation may involve the interaction between these populations (and likely other face processing regions), direct evidence for this hypothesis has been lacking. Our results here support the hypothesis that individual-level facial information is not only encoded by the activity within certain brain populations, but also represented through recurrent interactions between multiple populations at a network level. In addition, MCPA showed significant face individuation in approximately the 200 – 500 ms time window after stimulus onset, but did not show any significant face individuation in the early time window (50 – 200 ms after stimulus onset), which is consistent with a previous MVPA study based on iEEG recording from FFA only (11). This suggests that the face individuation process involves temporally synchronized, recurrent interactions between OFA and FFA and likely other nodes in the face-processing network. Representational similarity analysis showed that the structure of the representations of face information in OFA and FFA are different from the structure in the interaction between these areas (with the strong caveats that a non-significant correlation does not allow one to accept the null hypothesis and that these results require replication in more subjects). It is beyond the scope of the current work to assess whether this distinct representation reflects the binding of the OFA and FFA representations, the transformation of information between the representation in OFA and

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FFA, or some other computational process. More broadly, the MCPA results suggest that the computational work done in service of face individuation occurs not only on the local level, but also at the level of distributed brain circuits. Further studies with broader coverage, especially anterior temporal areas, are required to shed light on what other populations are involved in face individuation.

Conclusion

Previously, multivariate pattern analysis have been used to analyze either the information processing within a certain area and functional connectivity methods have been used to assess whether or not brain networks participate in a particular process. With MCPA, the two perspectives are merged into one method, which extends multivariate pattern analysis to enable the detailed examination information processing at the network level. Thus, the introduction of MCPA provides a platform for examining how computation is carried out through the interactions between different brain areas, allowing us to directly test hypotheses regarding circuit-level information processing.

Acknowledgements

We would like to thank the patient for participating in the iEEG experiments, and Dr. R. Mark Richardson, Michael Ward and the epilepsy monitoring unit staff, Cheryl Plummer,

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Gena Ghearing, and administration for their assistance and cooperation with our research. We thank Marc Coutanche and Julie Fiez for their insightful comments and feedback on this work. This work was supported by the National Institute on Drug abuse under award NIH R90DA023420 (to YL) and the National Institute of Mental Health under award NIH R01MH107797 (to ASG). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Reference:

1. Hubel DHW, Torsten N (1959) Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology* 148(3).
2. Desimone R, Albright TD, Gross CG, & Bruce C (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* 4(8):2051-2062.
3. Tsao DY, Freiwald WA, Tootell RB, & Livingstone MS (2006) A cortical region consisting entirely of face-selective cells. *Science* 311(5761):670-674.
4. Sugase Y, Yamane S, Ueno S, & Kawano K (1999) Global and fine information coded by single neurons in the temporal visual cortex. *Nature* 400(6747):869-873.
5. Haxby JV*, et al.* (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293(5539):2425-2430.
6. Cox DD & Savoy RL (2003) Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage* 19(2 Pt 1):261-270.
7. Kamitani Y & Tong F (2005) Decoding the visual and subjective contents of the human brain. *Nat Neurosci* 8(5):679-685.
8. Polyn SM, Natu VS, Cohen JD, & Norman KA (2005) Category-specific cortical activity precedes retrieval during memory search. *Science* 310(5756):1963-1966.

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

9. Haynes JD & Rees G (2006) Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience* 7(7):523-534.

10. Poldrack RA (2011) Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. *Neuron* 72(5):692-697.

11. Ghuman AS*, et al.* (2014) Dynamic encoding of face information in the human fusiform gyrus. *Nat Commun* 5:5672.

12. Edelman S, Grill-Spector K, Kushnir T, & Malach R (1998) Toward direct visualization of the internal shape representation space by fMRI. *Psychobiology* 26(4):309-321.

13. Kriegeskorte N & Kievit RA (2013) Representational geometry: integrating cognition, computation, and the brain. *Trends Cogn Sci* 17(8):401-412.

14. Haxby JV, Connolly AC, & Guntupalli JS (2014) Decoding Neural Representational Spaces Using Multivariate Pattern Analysis. *Annu Rev Neurosci* 37:435-456.

15. Fukushima K (1980) Neocognitron - a Self-Organizing Neural Network Model for a Mechanism of Pattern-Recognition Unaffected by Shift in Position. *Biol Cybern* 36(4):193-202.

16. Grossberg S (1980) How Does a Brain Build a Cognitive Code. *Psychol Rev* 87(1):1-51.

17. Mumford D (1992) On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biol Cybern* 66(3):241-251.

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

18. Lundqvist DF, Anders; Öhman, A (1998) The Karolinska directed emotional faces (KDEF). *CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet*:91-630.

19. Lee TS & Mumford D (2003) Hierarchical Bayesian inference in the visual cortex. *J Opt Soc Am A* 20(7):1434-1448.

20. Kveraga K, Ghuman AS, & Bar M (2007) Top-down predictions in the cognitive brain. *Brain Cogn* 65(2):145-168.

21. Freiwald WA, Tsao DY, & Livingstone MS (2009) A face feature space in the macaque temporal lobe. *Nat Neurosci* 12(9):1187-1196.

22. Kriegeskorte N, Formisano E, Sorger B, & Goebel R (2007) Individual faces elicit distinct response patterns in human anterior temporal cortex. *P Natl Acad Sci USA* 104(51):20600-20605.

23. Kay KN, Naselaris T, Prenger RJ, & Gallant JL (2008) Identifying natural images from human brain activity. *Nature* 452(7185):352-355.

24. Nestor A, Plaut DC, & Behrmann M (2011) Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *P Natl Acad Sci USA* 108(24):9998-10003.

25. Fodor JA (1983) *The modularity of mind : an essay on faculty psychology* (MIT Press, Cambridge, Mass.) p 145 p.

26. Perkel DH, Gerstein GL, & Moore GP (1967) Neuronal spike trains and stochastic point processes. II. Simultaneous spike trains. *Biophys J* 7(4):419-440.

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

27. Friston KJ (1994) Functional and effective connectivity in neuroimaging: a synthesis. *Human Brain Mapping* 2(1-2):56-78.

28. McIntosh AR, *et al.* (1994) Network analysis of cortical visual pathways mapped with PET. *J Neurosci* 14(2):655-666.

29. Ishai A (2008) Let's face it: it's a cortical network. *Neuroimage* 40(2):415-419.

30. Gschwind M, Pourtois G, Schwartz S, de Ville DV, & Vuilleumier P (2012) White-Matter Connectivity between Face-Responsive Regions in the Human Brain. *Cereb Cortex* 22(7):1564-1576.

31. Pyles JA, Verstynen TD, Schneider W, & Tarr MJ (2013) Explicating the face perception network with white matter connectivity. *PLoS One* 8(4):e61611.

32. Friston KJ, *et al.* (1997) Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage* 6(3):218-229.

33. Brookes MJ, *et al.* (2014) Measuring temporal, spectral and spatial changes in electrophysiological brain network connectivity. *Neuroimage* 91:282-299.

34. Borga M (1998) Learning multidimensional signal processing.

35. Bach FRJ, Michael I (2005) A probabilistic interpretation of canonical correlation analysis.

36. Weiner KS, Sayres R, Vinberg J, & Grill-Spector K (2010) fMRI-Adaptation and Category Selectivity in Human Ventral Temporal Cortex: Regional Differences Across Time Scales. *J Neurophysiol* 103(6):3349-3365.

37. Maris E & Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* 164(1):177-190.

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

38. Gilbert CD & Li W (2013) Top-down influences on visual processing. *Nat Rev Neurosci* 14(5):350-363.

39. Kriegeskorte N (2011) Pattern-information analysis: From stimulus decoding to computational-model testing. *Neuroimage* 56(2):411-421.

40. Richiardi J, Eryilmaz H, Schwartz S, Vuilleumier P, & Van De Ville D (2011) Decoding brain states from fMRI connectivity graphs. *Neuroimage* 56(2):616-626.

41. Shirer WR, Ryali S, Rykhlevskaia E, Menon V, & Greicius MD (2012) Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cereb Cortex* 22(1):158-165.

42. Turk-Browne NB (2013) Functional Interactions as Big Data in the Human Brain. *Science* 342(6158):580-584.

43. Finn ES*, et al.* (2015) Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat Neurosci* 18(11):1664-1671.

44. Wang YD, Cohen JD, Li K, & Turk-Browne NB (2015) Full correlation matrix analysis (FCMA): An unbiased method for task-related functional connectivity. *J Neurosci Meth* 251:108-119.

45. Rosenberg MD*, et al.* (2016) A neuromarker of sustained attention from whole-brain functional connectivity. *Nat Neurosci* 19(1):165-171.

46. Coutanche MN & Thompson-Schill SL (2013) Informational connectivity: identifying synchronized discriminability of multi-voxel patterns across the brain. *Front Hum Neurosci* 7:15.

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

47. Singer W (1999) Neuronal synchrony: a versatile code for the definition of relations? *Neuron* 24(1):49-65, 111-125.

48. Rodriguez E*, et al.* (1999) Perception's shadow: long-distance synchronization of human brain activity. *Nature* 397(6718):430-433.

49. Shadlen MN & Movshon JA (1999) Synchrony unbound: a critical evaluation of the temporal binding hypothesis. *Neuron* 24(1):67-77, 111-125.

50. Reynolds JH & Desimone R (1999) The role of neural mechanisms of attention in solving the binding problem. *Neuron* 24(1):19-29, 111-125.

51. Roelfsema PR, Lamme VA, & Spekreijse H (2004) Synchrony and covariation of firing rates in the primary visual cortex during contour grouping. *Nat Neurosci* 7(9):982-991.

52. Kettenri.Jr (1971) Canonical Analysis of Several Sets of Variables. *Biometrika* 58(3):433-&.

53. Kriegeskorte N, Goebel R, & Bandettini P (2006) Information-based functional brain mapping. *P Natl Acad Sci USA* 103(10):3863-3868.

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

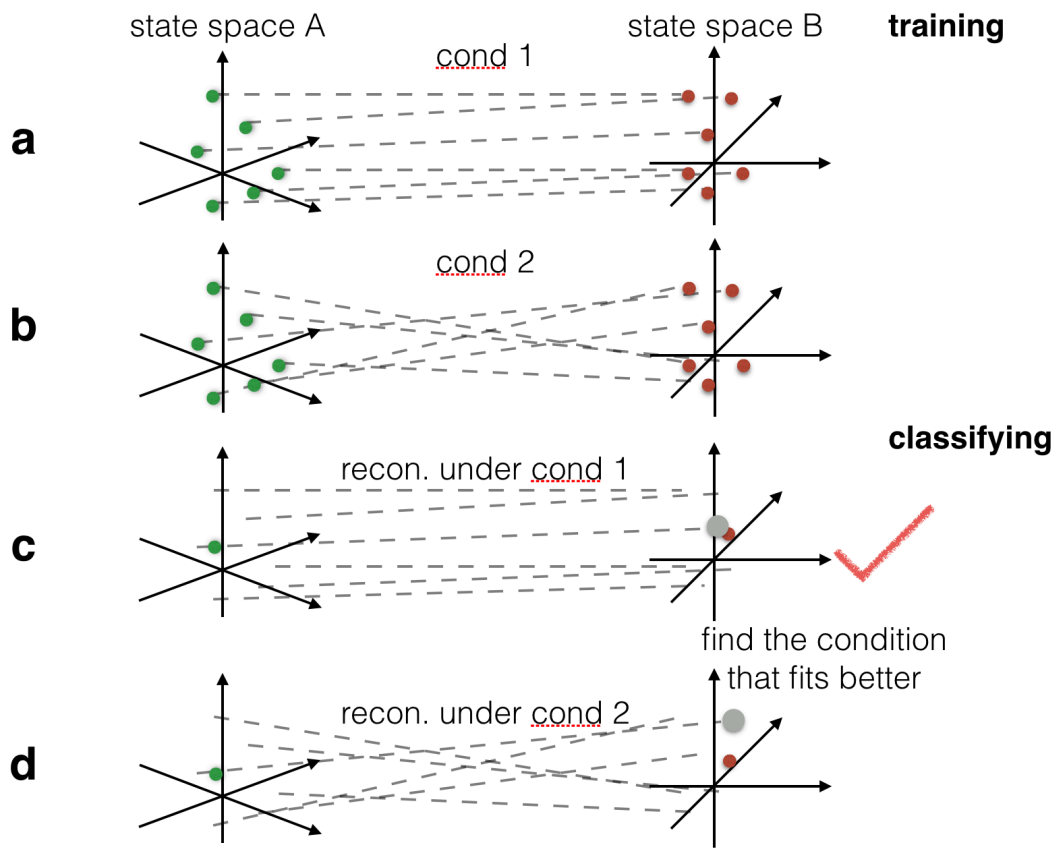

Figure 1 **Illustration of the connectivity map and classifier of MCPA.**

a) An illustration of functional information mapping between two populations under condition 1. The representational state spaces of the two populations are shown as A and B, each colored dot corresponds to an observed data point from the population. The functional information mapping is demonstrated as the dotted lines that project points from one space onto another. One observation of the two populations results in two dots that are linked by one dotted line.

b) An illustration of functional information mapping between two populations under condition 2.

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

c) An illustration of the predicted signal (shown as the grey dot) by mapping the observed neural activity from one population onto another using the mapping patterns learned from condition 1. The real signal in the second population is shown by red dot.

d) An illustration of the predicted signal (shown as the grey dot) by mapping the observed neural activity from one population onto another using the mapping patterns learned from condition 2. The real signal in the second population is shown by red dot. In this case, MCPA would classify the activity as arising from condition 1 because of the better match between the predicted and real signal.

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

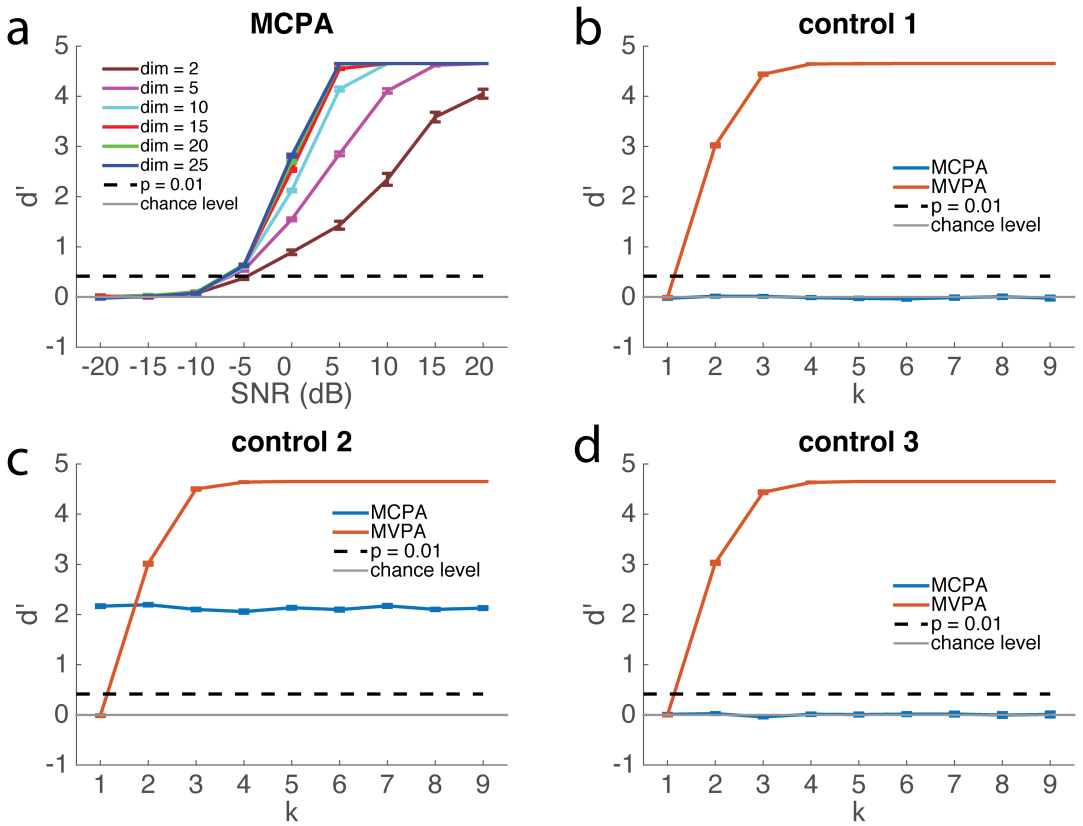

Figure 2. Synthetic data and control simulation experiments. The mean and standard error for 100 simulation runs are plotted. The horizontal gray line corresponds to chance level (*d*' = 0). The dashed line (*d*' = 0.42) corresponds to the chance threshold, p = 0.01, based on a permutation test. The maximum possible *d'* = 4.65 (equivalent to 99% accuracy because the d' for 100% accuracy is infinity).

a) The sensitivity of MCPA for connectivity between two populations as a factor of SNR and the number of effective dimensions in each population. MCPA was applied to synthetic data, where two conditions had different patterns of functional connectivity (measured by SNR and dimensionality). Performance of MCPA was significantly higher than chance level when SNR ≥ -5 dB and the number of dimensions ≥ 2. Performance of MCPA saturated to maximum when SNR > 5 dB and the number of dimensions > 10.

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

b) The insensitivity of MCPA when there is variable local discriminant information, but no circuit-level information (control case 1). MCPA and MVPA were applied to control case 1. The SNR was fixed at 0 dB and the number of dimensions is fixed at 10 for panels b, c, and d. *k* corresponds to the ratio of the standard deviations of the two conditions in panels b, c, and d.

c) The insensitivity of MCPA to changes in local discriminant information with fixed circuit-level information when there is both local and circuit-level information (control case 2).

d) The insensitivity of MCPA to variable local discriminant information when the circuit-level activity is correlated, but does not contain circuit-level information about what is being processed (control case 3).

bioRxiv preprint doi: https://doi.org/10.1101/046441; this version posted March 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

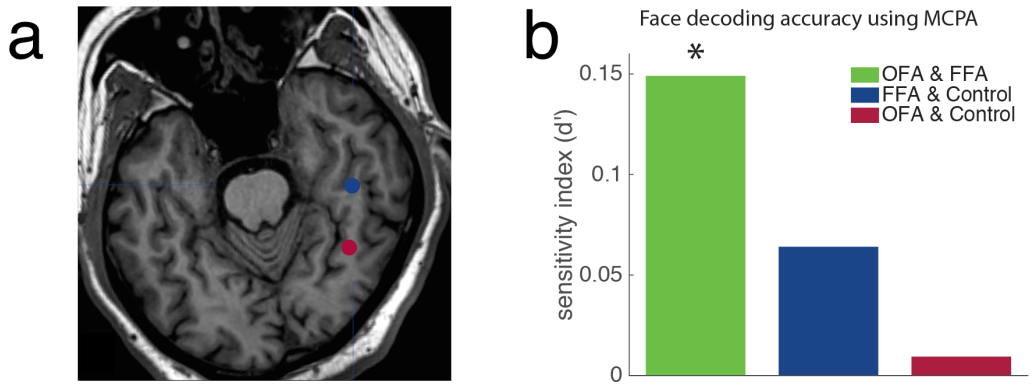

Figure 3. iEEG expriments and MCPA results.

a) Location of the electrodes of interest. The blue dot corresponds to the location of the FFA contact while the red dot corresponds to the location of the OFA contacts.

b) MCPA applied between (1) the OFA and FFA channels, (2) the FFA channel and the control channel, (3) the OFA channel and the control channel. The mean d' of pairwise face classification over all 2415 pair of faces across the 200-500 ms timewindow after stimulus onset is plotted. * $p < 0.05$, permutation test.