The goal of my research is to contribute to our understanding of the brain function. Specifically, I am interested in developing new analysis methodologies and software solutions in the domain of computational and cognitive neuroscience, as well as formalizing existing software. As an investigator in Neuroscience and Psychology, my background in computer science and engineering allows me to leverage my technical skills for cross-disciplinary knowledge transfer between computer and natural sciences.

Seeking answers to the question *How could we decode neural processes and their organization at different spatial and temporal scales given available instrumentation?* I have carried out and participated in the research projects on multimodal (e.g. EEG/ fMRI ) data analysis, large-scale decoding of the mental states, causal structure inference, and transformations of neural code representations across individuals. Through my involvement in neuroscience research I have observed that the idea of “open science” is often hindered by absence of convenient means to share data and difficulties with dissemination and deployment of developed methodologies. Therefore to streamline my own research and to help others with answering their research questions, I began a long-term collaboration with Michael Hanke (Magdeburg, Germany) in 2006 to develop PyMVPA and NeuroDebian, both are free and open-source (FOSS) platforms. In the course of those projects I became an active contributor to a variety of FOSS projects related to neuroimaging (e.g. Nipype, Psychopy, etc.) and statistical learning (e.g. scikit-learn, Shogun, ANTS, etc.). I have joined INCF “Standards for Datasharing” task force to contribute to the development of efficient means for data sharing in neuroimaging.

With all of these ongoing research and software development activities, I hope to improve reproducibility in neuroscience research by making existing tools, methodologies and data readily available in an integrated computing environment for any researcher, not only to facilitate new research directions but also to replicate published findings.

**PyMVPA**

*Python framework for analysis of neural data*

[http://www.pymvpa.org](http://www.pymvpa.org)

PyMVPA is a flexible and versatile Python platform for the analysis of neural data through employing recent advances in statistical learning methods. It affords the efficient use of normative machine learning methods on neuroscience data, while at the same time fostering the development of novel data analysis methods. PyMVPA provides a rich collection of classifiers and regressions, estimation methods of generalization, statistical significance and similarities. To provide a scalable rich environment we interface to the functionality developed in other FOSS projects, such as Shogun, scikit-learn, MDP, etc. in a uniform API allowing researchers to explore wider variety of methods than any single toolbox could make available alone.

We initiated the PyMVPA project in 2007 to facilitate our own research, and from the beginning have developed it as FOSS project accepting contributions from the community. Since inception PyMVPA has had over 20 releases, became equipped with a rich collection of unit- and doc-tests, and an extended tutorial for interested researchers to start using it in a matter of hours. PyMVPA project acquired a vibrant users community. It is inspiring to see PyMVPA is used not only by our colleagues at Dartmouth (PBS and Music Department, Dr. M. Casey) and Magdeburg, but also by various research groups around the globe. In the upcoming years I am planing to extend PyMVPA with methodologies for addressing the following research questions.

**Connectivity characterization of the functional units**

Based on multivariate auto-regressive modeling and the idea behind Granger causality I am developing analysis procedures to characterize basic functional units with the structure of neural information flow in their neighborhoods as well as their involvement in distant functional connections. Discovered distant connections and the structure of the local information flow allow to characterize each voxel in degree of its involvement in local computation vs. communication gateway to distant areas. Preliminary results on a sample dataset showed plausibility of the suggested methodology: voxels were found consistently differentiable in regard to the existence of distant connections across different conditions. I am going to explore the discovered local information flow, further validate this approach on data for complex natural stimuli, and extend it to non-linear mappings.

**Topologies of neural code representations**

The domain of fMRI data analysis is dominated by linear methods despite a variety of reported effects which invalidate basic linearity assumptions of the signal in various respects. In particular, in the domain of machine learning analysis of fMRI, linear classifiers are used nearly exclusively since voxel space is very high-dimensional and investigation of non-linear decision boundaries would require more intensive model-selection, and thus more data samples to obtain reliable models and avoid a selection bias. Recent work in our lab on the alignment of neural code representations across different individuals (codename...
hyperalignment \(^{\text{HGC}+11}\), implemented in PyMVPA) allows to overcome the problem of a limited samples size through carrying out the analyses across subjects. In the scope of ongoing in the lab study of social aspects of human facial perception I am going to explore the topological organizations of identity information in the brain, and to what major factors to attribute low performance of a classifier: absent signal of interest or just low SNR, high variability of patterns across the trials, or improper choice of the model (linear vs. continuous non-linear vs. combinatorial) to model the topology of the categories.

**NeuroDebian**

**Turnkey scientific platform for Neuroscience**

http://neuro.debian.net

It is fortunate that majority of the data analysis software projects in neuroimaging field follow free and open source (FOSS) development model. Unfortunately availability of the sources and limited set of binary builds alone neither provide an adequate media for wide dissemination of the new methodologies, nor allow for maintenance of a robust heterogeneous research computing environment. Without appropriate centralized distribution maintenance of such environment becomes increasingly tedious, thus disconnecting researchers from the recent developments in the field and limiting them in number of tools they might employ for their research. To overcome such a problem for myself, and to share benefits with the others, together with Michael Hanke in 2006 we founded NeuroDebian project.

NeuroDebian’s goal is to bring developers and users of software for neuroscience closer together, by providing a turnkey platform where research software is well tested, integrated and made conveniently available to users. It achieves this goal through the integration of neuroscience FOSS within a massive community-driven project — Debian — which delivers a robust, free and open universal operating system with possibly the largest archive of maintained software. At the moment NeuroDebian is used by thousands researchers “natively” or via virtualization, making it possible for any researcher not only to maintain up-to-date and secure computing environment adequate with scientific demands, but also to share complete computing environments, thus facilitating reproducibility and collaboration. Adhering to open standards developed in Debian community, I frequently contribute back to the projects regarding various aspects of FOSS development -- copyright and licensing conflicts, modularity of the design, and compatibility. Moreover to enrich Debian environment and to attract “fresh blood” to the project I am mentoring work of the contributors working on packaging of neuroscience-related projects for Debian.

Besides ongoing maintenance of already supported software products, we have a number of planned projects for NeuroDebian, which I am going to address

- Expansion of NeuroDebian into a wider range of research applications:
  - extend software coverage beyond fMRI/DTI-based neuroimaging to tools for intra/extra-cellular recording and modeling, EEG/MEG, and data management (e.g. NEURON, XNAT)
  - integrate essential Matlab™-based open-source software: e.g. PsychToolbox, EEGLAB; and, where feasible, assure compatibility with Octave
- Sustained availability of software and precise re-creation of complete research environments
  - employ Debian's existing software archive snapshotting framework to preserve and distribute all previous and current versions of supported by NeuroDebian software (currently in place, public interface is in works)
  - develop the utility to describe a particular analysis environment (with all versioned dependencies) to be able to reconstruct it at any later point in time - by anyone -- given access to the specification and to the aforementioned software archive snapshots
- Adequate acknowledgment of the research software

Extend aforementioned utility to gather bibliographic references for software and data resources used for any particular data analysis. It will provide users with ready-to-use bibliography thus adequately acknowledging the efforts of the research software developers

---

This PDF document embeds hyperlinks to external URLs providing additional information (e.g. information about user-base, community, corresponding projects). Hyperlinked text is colored in blue. If hyperlinks get removed by mathjobs portal, you can obtain the original from http://haxbylab.dartmouth.edu/publications/yoh-research_statement.pdf

**HH08** S.J. Hanson and Y.O. Halchenko “Brain reading using full brain support vector machines for object recognition: there is no ‘face’ identification area”. Neural Computation, 20, 2008


**PHH09** R.A. Poldrack, Y.O. Halchenko and S.J. Hanson “Decoding the Large-Scale Structure of Brain Function by Classifying Mental States Across Individuals.” Psychological Science, 20, 2009

