Towards semantic visual decoding of naturalistic movies with high-density diffuse optical tomography

Wiete Fehner[1], Morgan Fogarty[1], Aahana Bajracharya[1], Zachary E. Markow[1], Dana Wilhelm[1], Jason Trobaugh[1], Alexander G. Huth[2], Joseph P Culver[1]*
[1]Washington University in St. Louis, St. Louis, Missouri 63130, USA, [2]The University of Texas at Austin, Austin, 78712, TX, USA
**Presenting Author email address: f.wiete@wustl.edu*

connect with me

Background

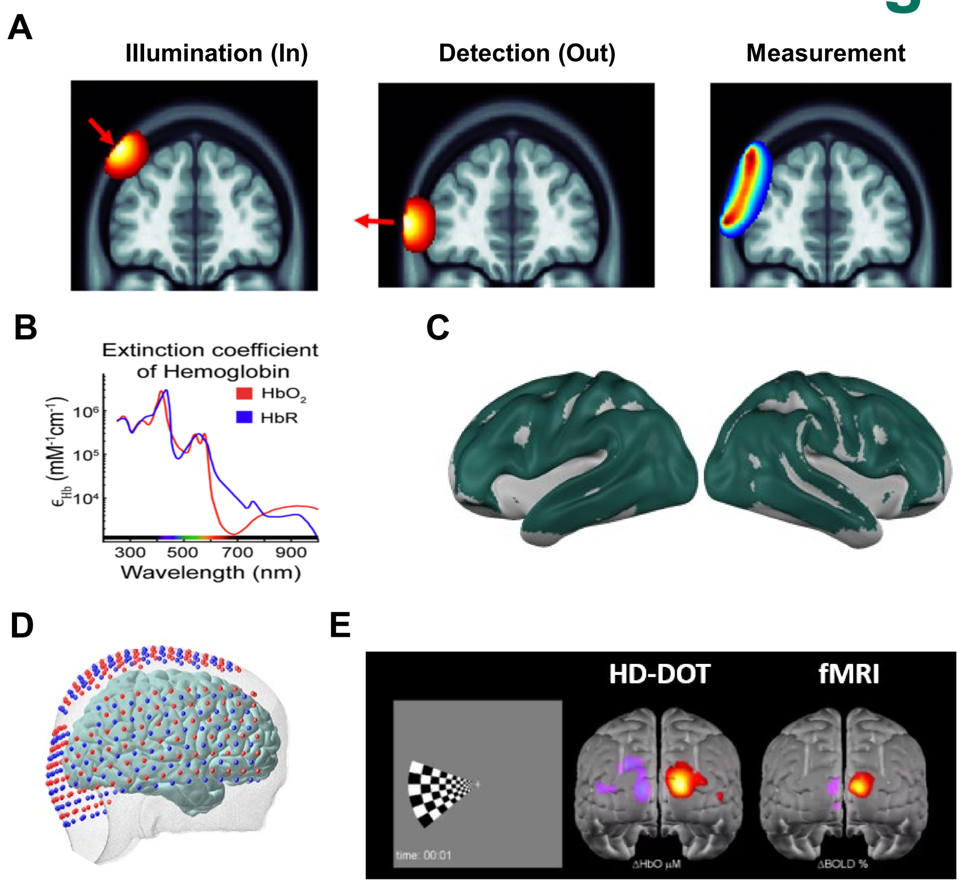


Figure 1. High-Density Diffuse Optical Tomography. (A) Sensitivity profile of measurement channel. **(B)** Extinction coefficients of hemoglobin [4]. **(C-D)** VHD-DOT imaging field of view and position of sources and detectors. **(E)** HD-DOT allows to map visual processing like fMRI [2].

- Optical imaging systems, such as functional near-infrared spectroscopy (fNIRS) and diffuse optical tomography (DOT) measure the changes in blood flow of the brain by analyzing the light absorption in the near-infrared optical window between a source and a detector.
- Diffuse optical tomography (DOT) methods reconstruct spatially overlapping multi-distance source-detector measurement channels into 3-D maps [4].
- The density of the sources and detectors impacts image quality by increasing the field of view (FOV) and spatial resolution.

HD-DOT allows cortical mapping like fMRI.

Motivation

Aphasia affects >100k individuals per year in the US

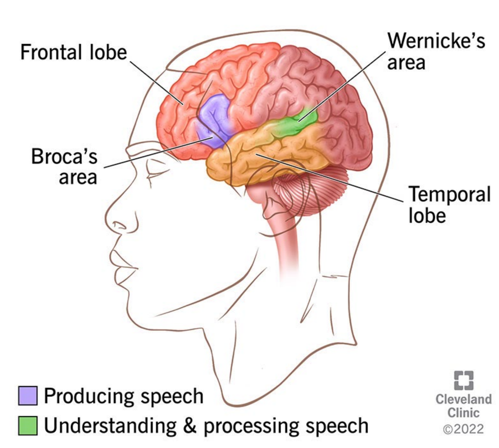


Figure 2. Aphasia. Figure from [5].

- **Expressive Aphasia** (Broca's Aphasia).
  - Trouble with speaking or writing but not with understanding.
- **Receptive Aphasia** (Wernicke's Aphasia).
  - Inability to receive and understand what is being said.

- Language is encoded in the cortex through language-specific areas and semantic representations.
- Visual semantic maps can potentially help identify alternative brain pathways that remain intact after a stroke (visual semantics).

Personalized language mapping (encoding) & BCI for augmented communication (decoding) could be useful.

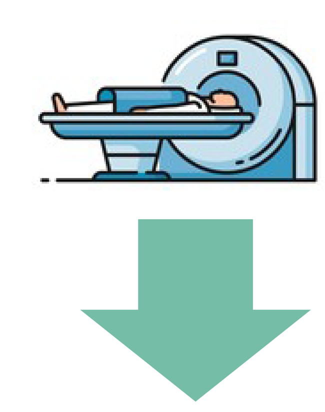


Figure 3. HD-DOT in Naturalistic Settings. Fiber-based HD-DOT (HD-DOT) system (bottom left) and newer wearable HD-DOT (WHD-DOT) systems (bottom right) allow for more naturalistic imaging environment compared to MRI.

Logistics of fMRI are NOT suitable for naturalistic environments.

Experimental Design

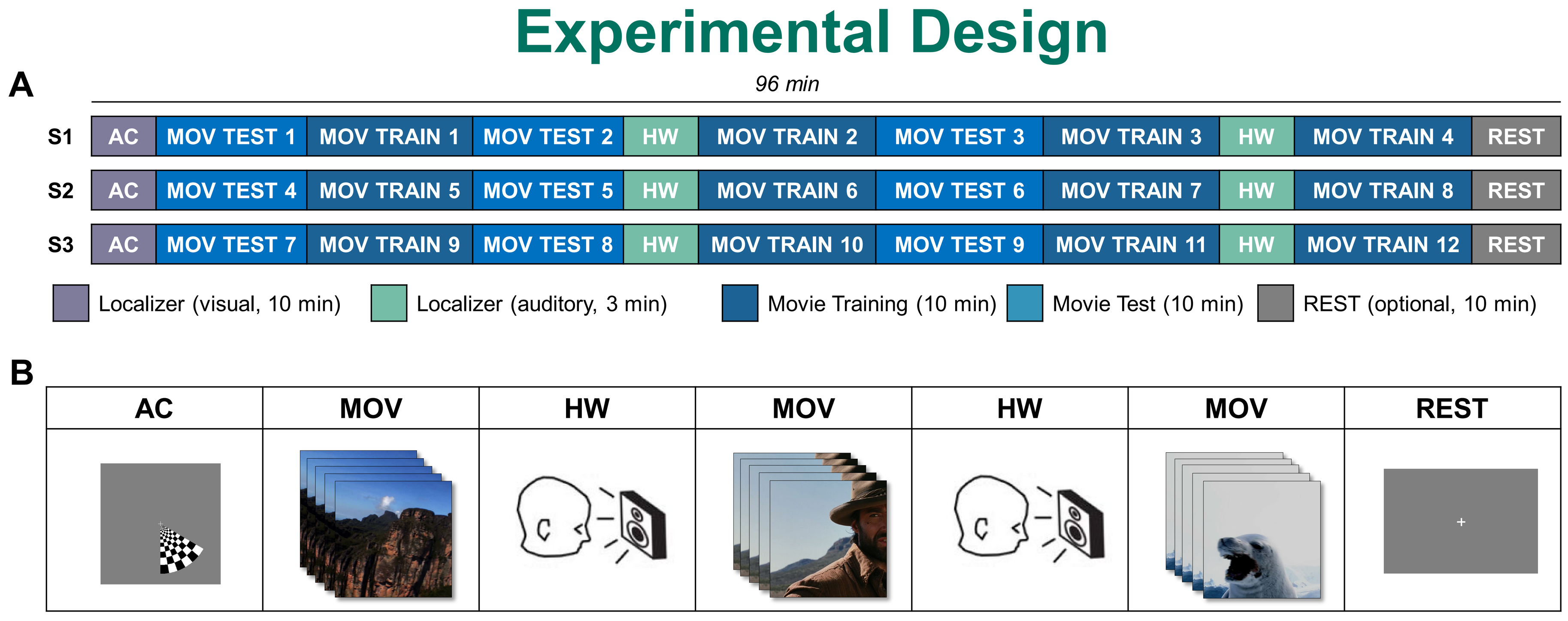
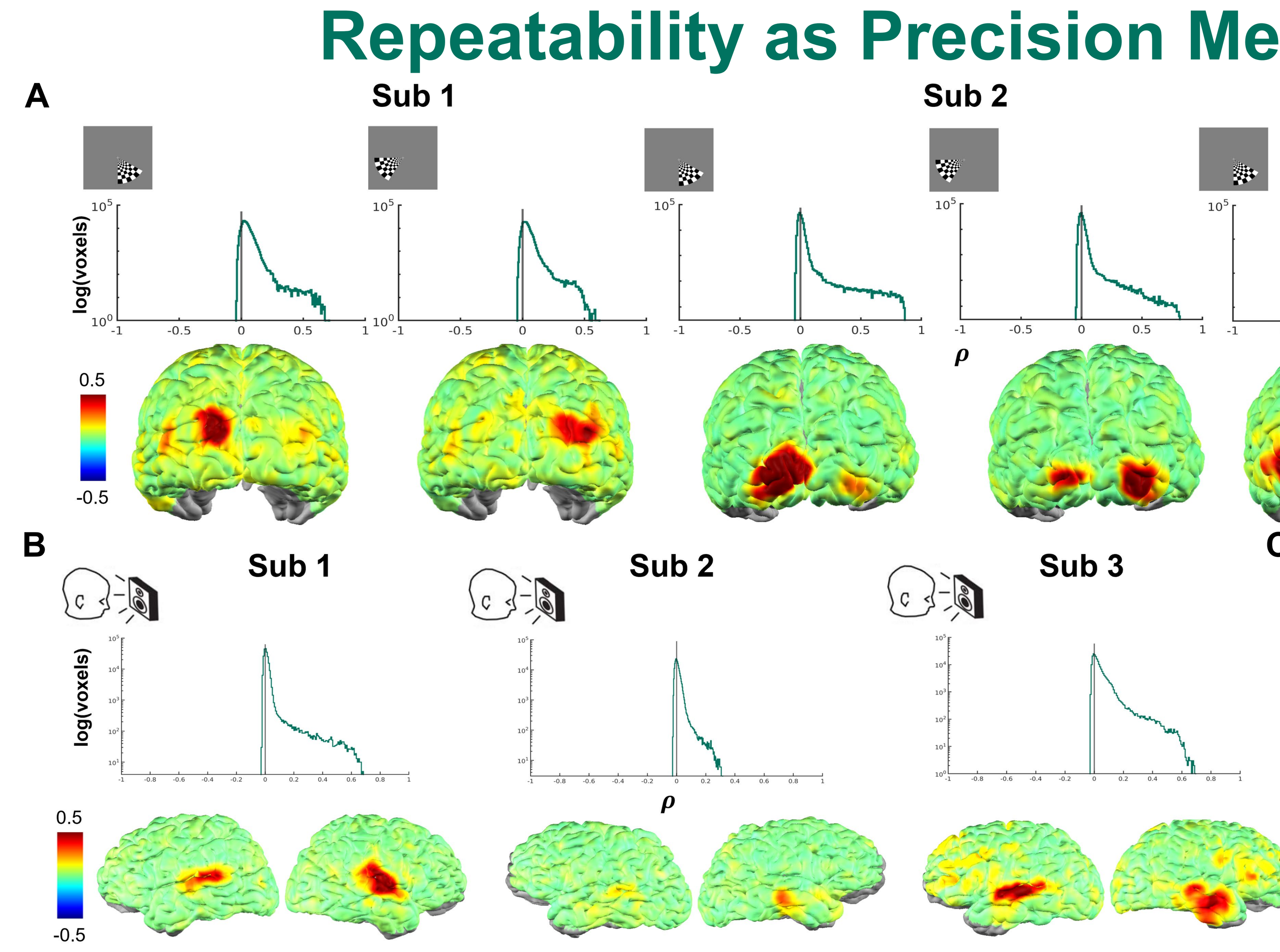


Figure 4. Experimental Design. (A) Overview of the experimental setup, including three scanning sessions with auditory (HW: hearing word) and visual (AC: alternating checkerboard) localizers to assess consistency across sessions. **(B)** Movie clips were validated in prior fMRI studies [3] and labeled using semantic categories from WordNet. Example of stimuli for task runs.

This gives 120 min of unique training data and 90 minutes of testing data (N=3). The testing data consist of 9 unique minutes.

Repeatability as Precision Measure



C

$$\frac{1}{\binom{m}{2}}\sum_{j=1}^{\binom{m}{2}} corr(BlockComb_{j(1)}, BlockComb_{j(2)})$$

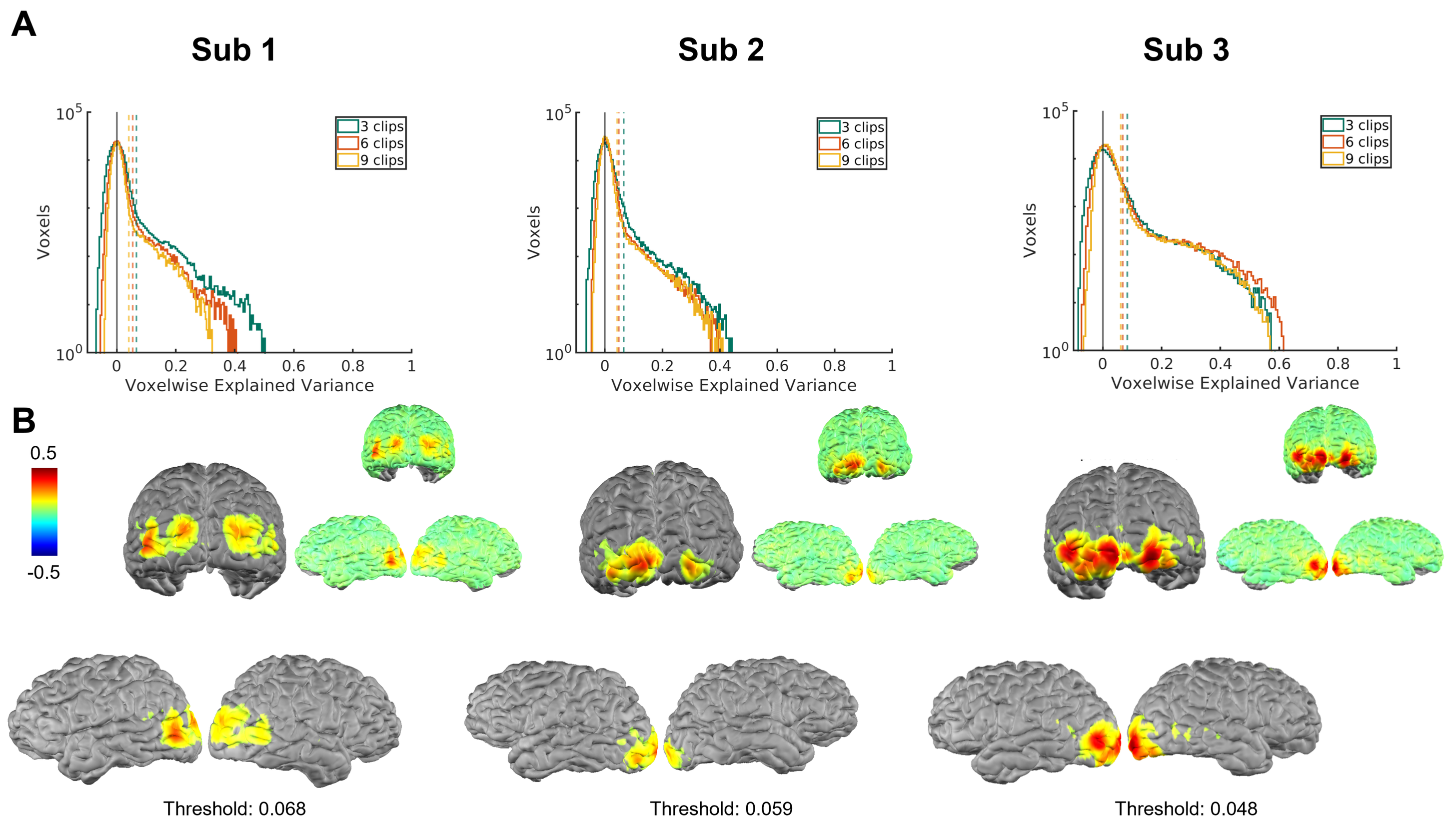
where

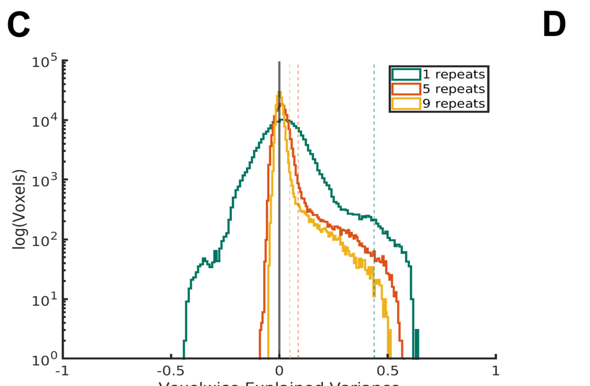
$$corr(\boldsymbol{x},\boldsymbol{y}) = \frac{\sum_i^n(\boldsymbol{x}_i-\bar{\boldsymbol{x}})(\boldsymbol{y}_i-\bar{\boldsymbol{y}})}{\sqrt{\sum_i^n(\boldsymbol{x}_i-\bar{\boldsymbol{x}})^2\sum_i^n(\boldsymbol{y}_i-\bar{\boldsymbol{y}})^2}}$$

- *m* is the total number of task blocks
- $BlockComb_{j(k)}$ is the k-th block in the j-th combination of blocks for comparison
- x and y are measurements from two task blocks
- x_i and y_i are the i-th measurement
- $\bar{x}$ and $\bar{y}$ are the means of the vectors
- n is the number of measurements

Figure 5. Repeatability Maps for Localizer Tasks. (A-B) Repeatability maps and distributions for visual and auditory localizers. **(C)** Equation used to calculate repeatability.

Explainable Variance of Test Movies





D

$$\boldsymbol{Y} \in R^{n\,x\,m\,x\,k}$$

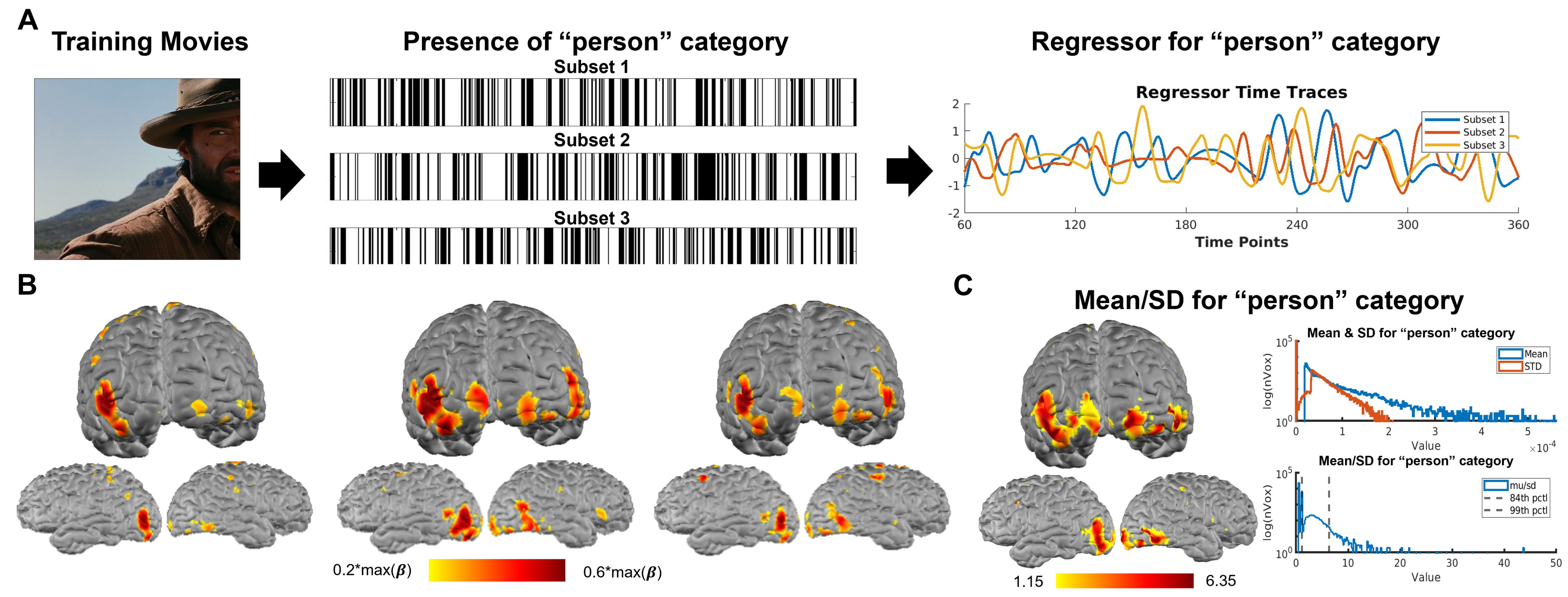
where $\boldsymbol{Y}$ is our dataset, *n* is the number of repetitions *m* is the number of time points, and *k* is the number of voxels.

- $\boldsymbol{y}_i, i = 1 \dots N$ measured signal in a voxel for each of the N repetitions
- $\bar{\boldsymbol{y}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{y}_i$ average brain response across repetitions
- $\boldsymbol{y}_i - \bar{\boldsymbol{y}}$ residual for each repeat

$$EV = \frac{1}{N}\sum_{i=1}^{N}Var(\boldsymbol{y}_i) - \frac{N}{N-1}\sum_{i=1}^{N}Var(\boldsymbol{y}_i - \bar{\boldsymbol{y}})$$

Figure 6. Explained Variance Analysis for Test Movie Clips. (A) Histograms of EV for 3, 6, and all 9 clips. **(B)** Brain maps using abs(min(EV)) as the threshold. **(C)** Effect of repetition on EV for Sub 3. **(D)** Equation for EV. The maps demonstrate high repeatability in visual areas, with the highest EV in these regions, indicating the maximum prediction accuracy achievable by the encoding model.

Semantic Mapping



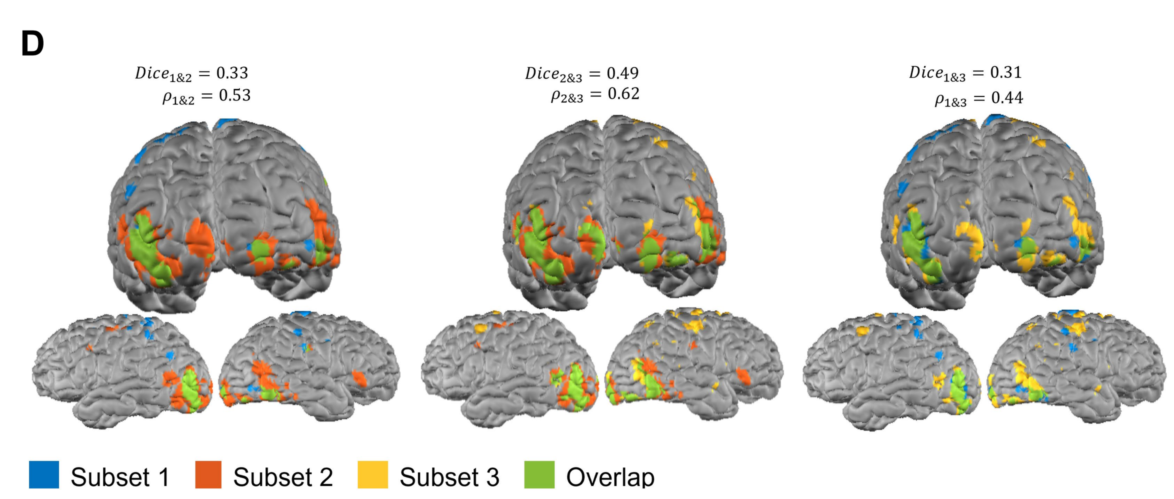
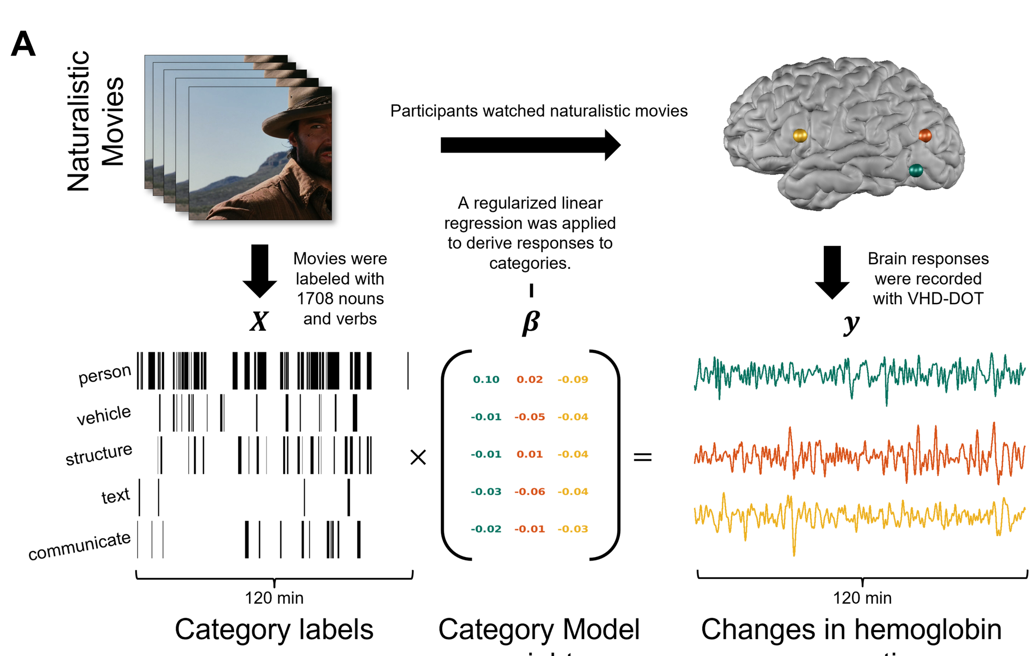


Figure 7. Semantic Mapping of the "Person" Category (in Sub 3). (A) A "person" category regressor was created for three subsets of the training data. **(B)** A GLM was applied to generate semantic maps. **(C)** Mean maps divided by their standard deviation highlight statistically meaningful areas. **(D)** Similarity measures demonstrate reproducibility of these maps.

VHD-DOT allows reproduceable mapping of semantic categories.

Voxelwise Encoding Model

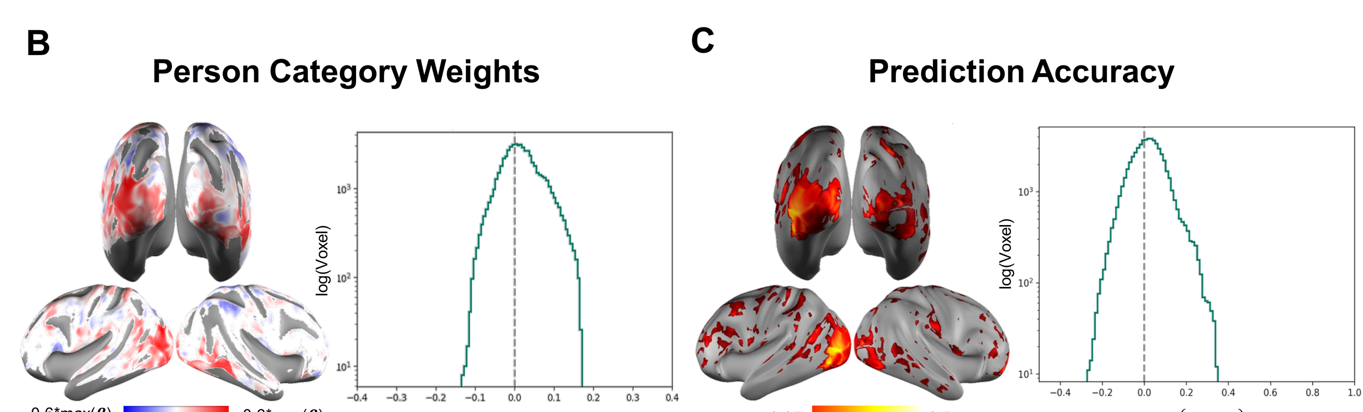


Ridge Regression

$$X\beta = y$$

$$\beta = argmin_{\beta}\{\|X\beta - y\|^2 + \alpha\|\beta\|^2\}$$

$$\beta = (X^TX + \alpha I)^{-1}X^Ty$$



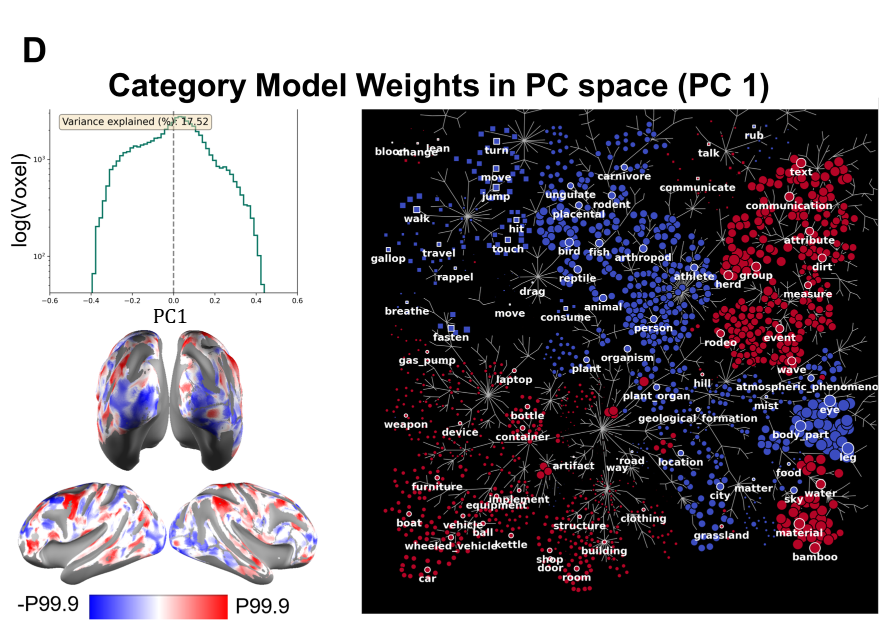


Figure 8. Voxelwise Encoding (in Sub 3). (A) Schematic of the voxelwise encoding model. **(B)** Extracted category weights for the "person" category. **(C)** Model prediction accuracy. **(D)** Complex semantic maps derived from Principal Component Analysis (PCA) on the category model weight matrix, visualized for PC 1 with corresponding categories.

Future Directions & Conclusions

- We demonstrated the feasibility of VHD-DOT for visual semantic mapping.
- VHD-DOT proved reproducibility of semantic category mapping and feasibility for complex voxelwise encoding models for semantic mapping.
- This study lays a foundation for visual semantic mapping in naturalistic environments and applications in aphasia.
- Future directions include expanding the encoding model and building towards a decoding model.

References and Acknowledgements

[1] Bajracharya, A. et al (2023). Precision Functional Mapping of Cortical Activity Using High-Density Diffuse Optical Tomography (HD-DOT). https://doi.org/10.1364/boda.2023.jtu4b.15
[2] Eggebrecht AT, Ferradal SL, Robichaux-Viehoever A, Hassanpour MS, Dehghani JH, Snyder AZ, Hershey T, Culver JP. Mapping distributed brain function and networks with diffuse optical tomography. Nat Photonics 2014 Jun; 8(6):448-454
[3] Huth, A. G. et al (2012). A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories across the Human Brain. Neuron, 76(6), 1210–1224. https://doi.org/10.1016/j.neuron.2012.10.014
[4] Wheelock, M. D., Culver, J. P., & Eggebrecht, A. T. (2019). High-density diffuse optical tomography for imaging human brain function. *Review of Scientific Instruments*, 90(5). https://doi.org/10.1063/1.5086809
[5] Image courtesy of Cleveland Clinic. © 2022 Cleveland Clinic.

This work was funded by USA NIH grants U01EB027005, R01NS109487, T32EB014855. Drs. Joseph Culver and Jason Trobaugh have a financial ownership interest in EsperImage LLC and may financially benefit if the company is successful in marketing its products that are related to this research. The material on this poster does not necessarily reflect the views of these funding sources.