Towards Semantic Encoding of Visual Content in Movies via 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

Background

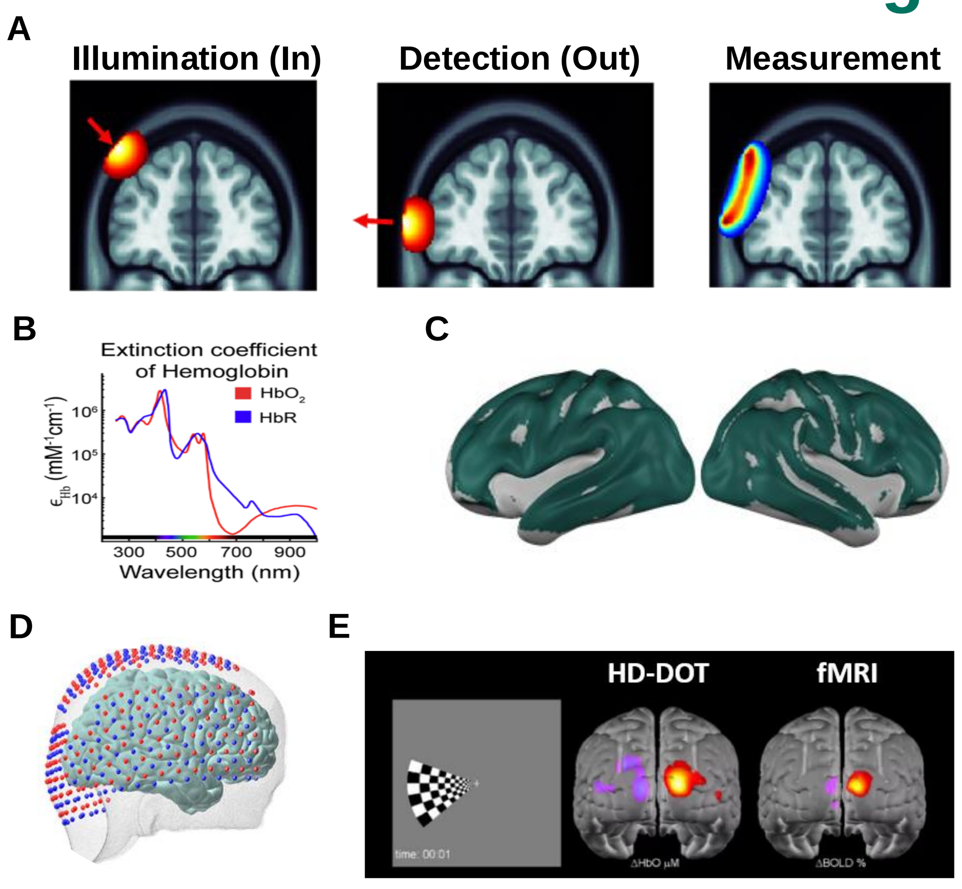


Figure 1. High-Density Diffuse Optical Tomography. (A) Sensitivity profile of measurement channel. **(B)** Extinction coefficients of hemoglobin [4]. **(C-D)** VHD-DOT imaging field of view and position of sources (N=255) and detectors (N=252). **(E)** HD-DOT allows to map visual processing like fMRI [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 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

- **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**).

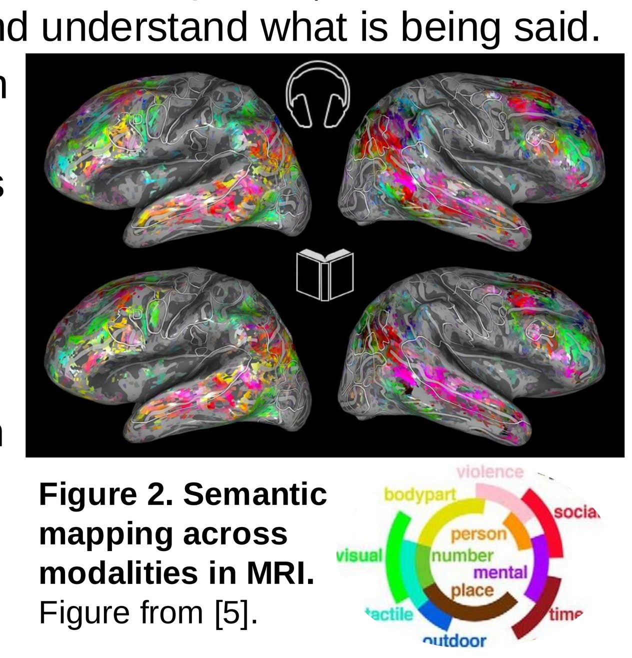


Figure 2. Semantic mapping across modalities in MRI. Figure from [5].

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

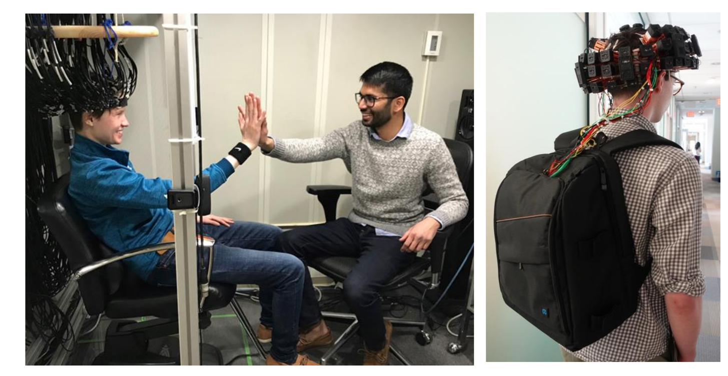


Figure 3. HD-DOT in Naturalistic Settings. Fiber-based HD-DOT (HD-DOT) system (bottom left) and newer wearable HD-DOT (WHD-DOT) systems (bottom right) allow for more naturalistic imaging environment compared to MRI.

Logistics of fMRI are NOT suitable for naturalistic environments.

Experimental Design & Repeatability

A (96 min)

S1	AC	MOV TEST 1	MOV TRAIN 1	MOV TEST 2	HW	MOV TRAIN 2	MOV TEST 3	MOV TRAIN 3	HW	MOV TRAIN 4	REST
S2	AC	MOV TEST 4	MOV TRAIN 5	MOV TEST 5	HW	MOV TRAIN 6	MOV TEST 6	MOV TRAIN 7	HW	MOV TRAIN 8	REST
S3	AC	MOV TEST 7	MOV TRAIN 9	MOV TEST 8	HW	MOV TRAIN 10	MOV TEST 9	MOV TRAIN 11	HW	MOV TRAIN 12	REST

Localizer (visual, 10 min) | Localizer (auditory, 3 min) | Movie Training (10 min) | Movie Test (10 min) | REST (optional, 10 min)

Training Data = 120 min, Testing Data = 90 min, N = 4

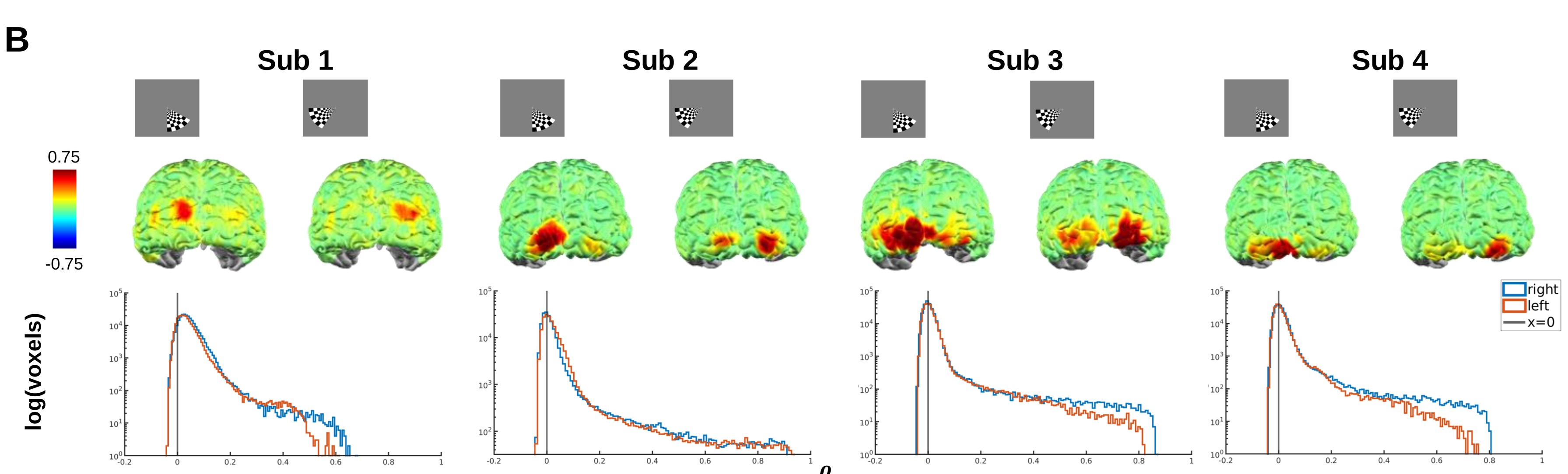
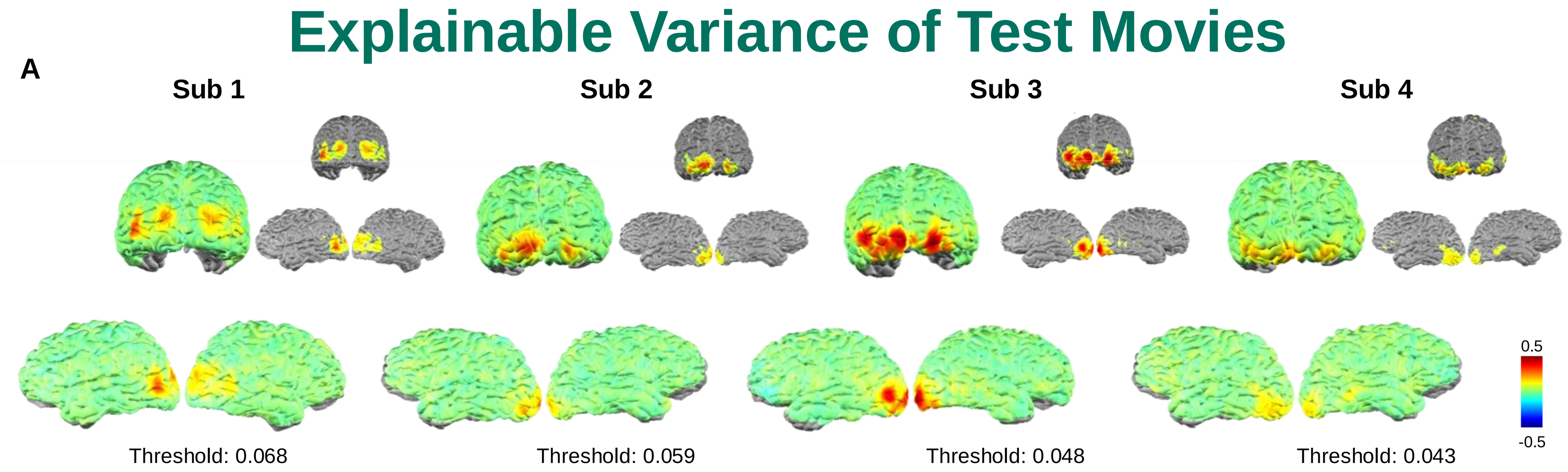
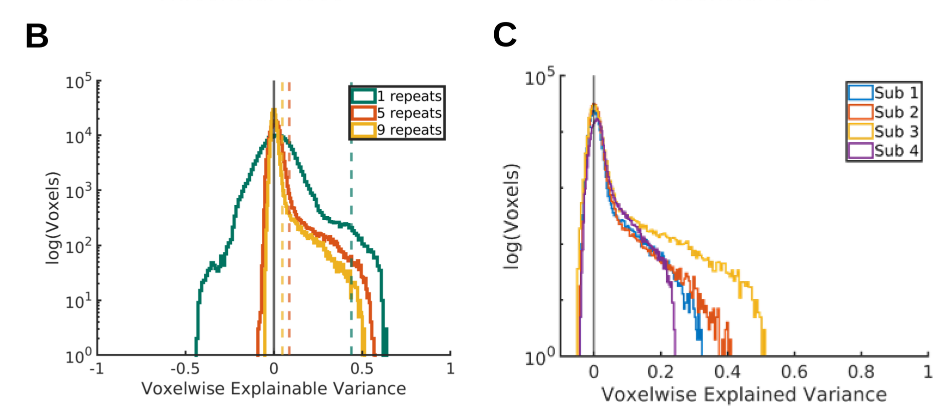


Figure 4. Experimental Design. (A) Three scanning sessions with auditory (HW: hearing word) and visual (AC: alternating checkerboard) localizers were used to assess consistency across sessions [1]. Movie clips were validated in prior fMRI studies [3] and labeled using semantic categories from WordNet. **(B)** Repeatability for visual localizers tasks.

Explainable Variance of Test Movies





D

$$Y \in R^{n \times m \times k}$$

where Y is our dataset, n is the number of repetitions m is the number of time points, and k is the number of voxels.

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

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

Figure 5. Explainable Variance Analysis for Test Movie Clips. (A) Brain maps using abs(min(EV)) as the threshold. **(B)** Effect of repetition on EV for Sub 3. **(C)** Histogram for EV for all subjects. **(D)** The equation for EV. The maps demonstrate high repeatability (EV) in visual areas, indicating the maximum prediction accuracy achievable by the encoding model.

Semantic Mapping

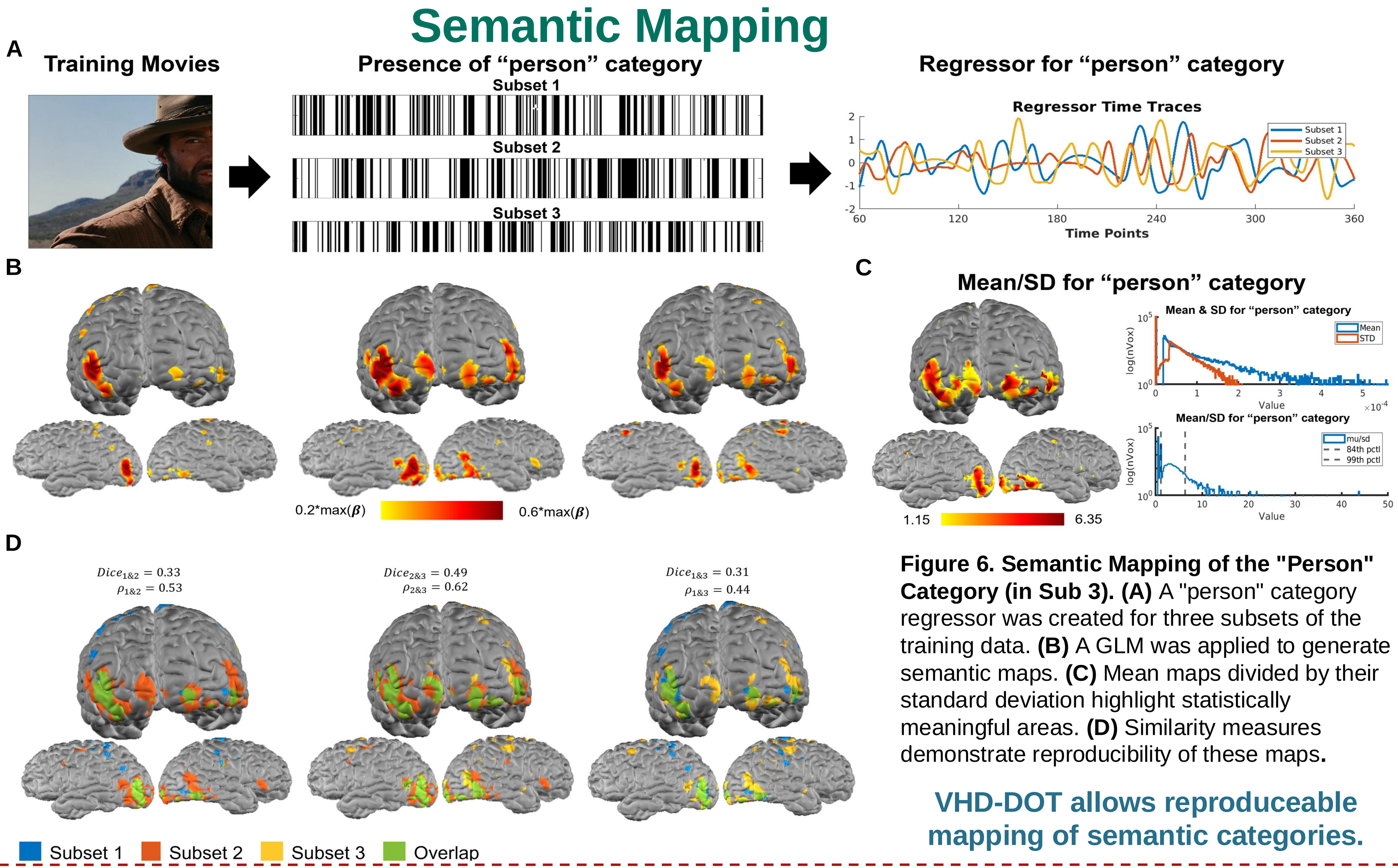
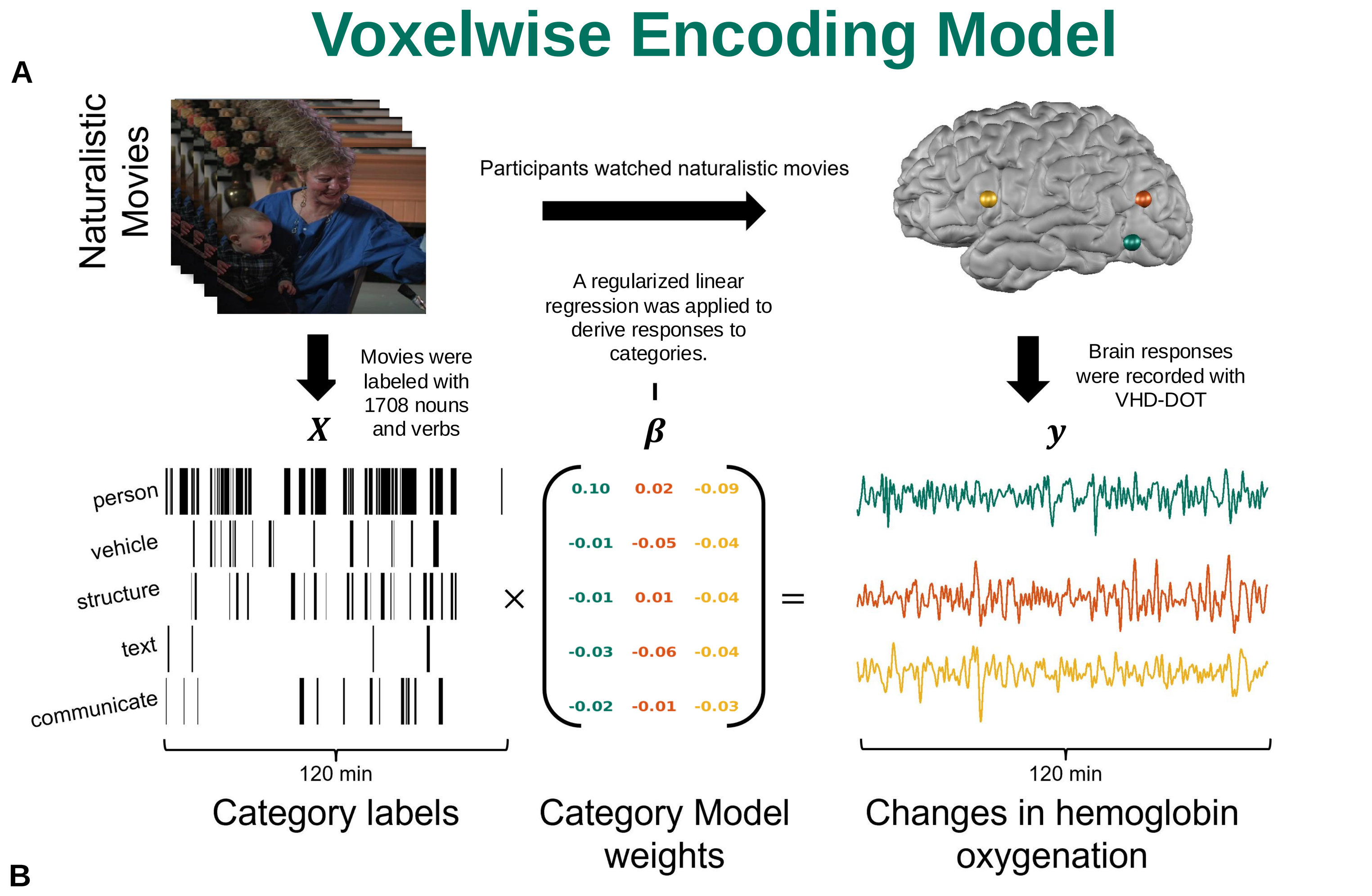


Figure 6. Semantic Mapping of the "Person" Category (in Sub 3). (A) A "person" category regressor was created for three subsets of the training data. **(B)** A GLM was applied to generate semantic maps. **(C)** Mean maps divided by their standard deviation highlight statistically meaningful areas. **(D)** Similarity measures demonstrate reproducibility of these maps.

VHD-DOT allows reproduceable mapping of semantic categories.

Voxelwise Encoding Model



B Ridge Regression

$$X\beta = y$$

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

$$\beta = \left(X^T X + \alpha I\right)^{-1} X^T y$$

Figure 7. Voxelwise Encoding Model. (A) Model and **(B)** corresponding equation. Ridge regression with bootstrapping is applied to derive category model weights.

Visual Semantic Encoding

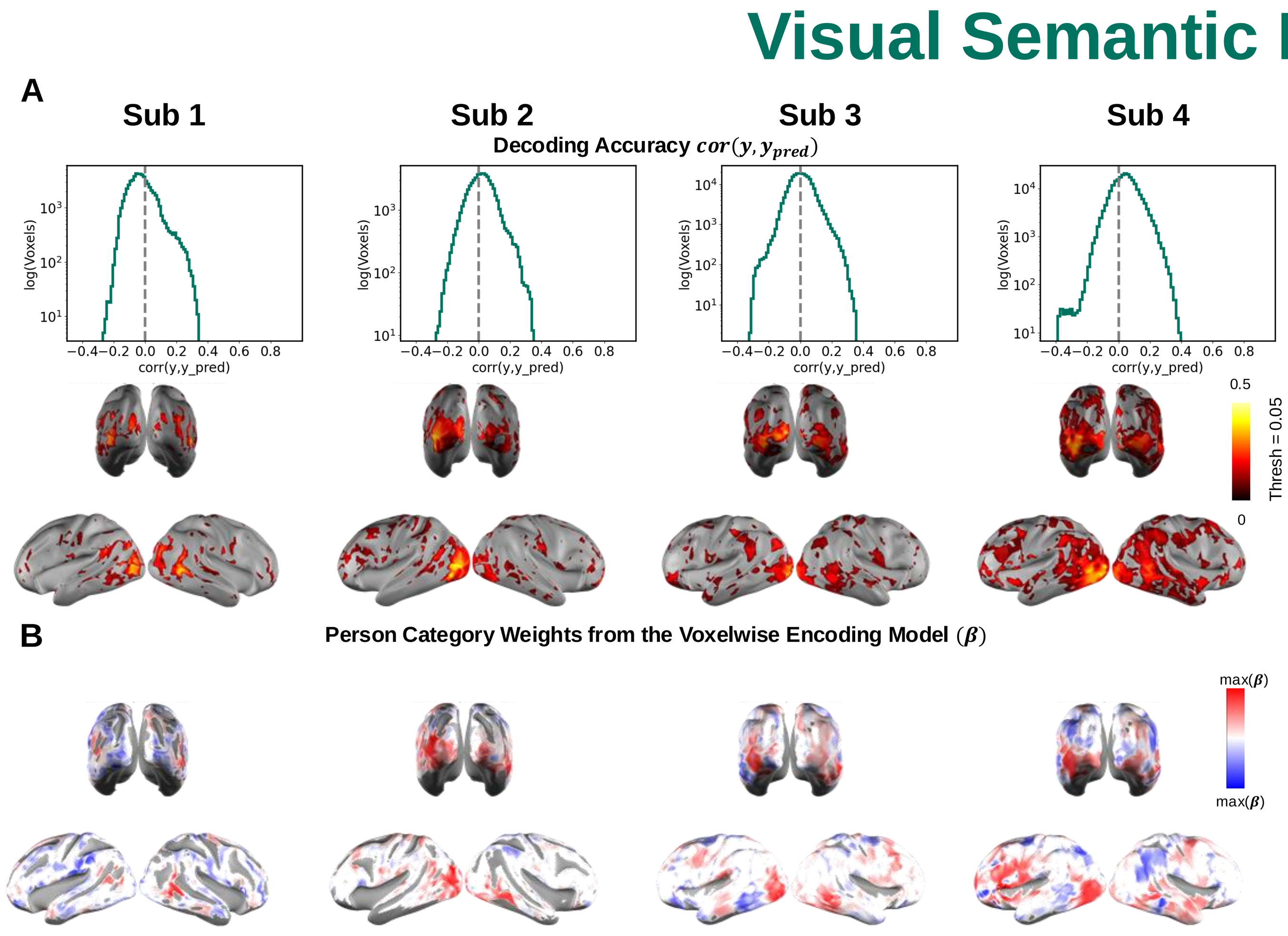
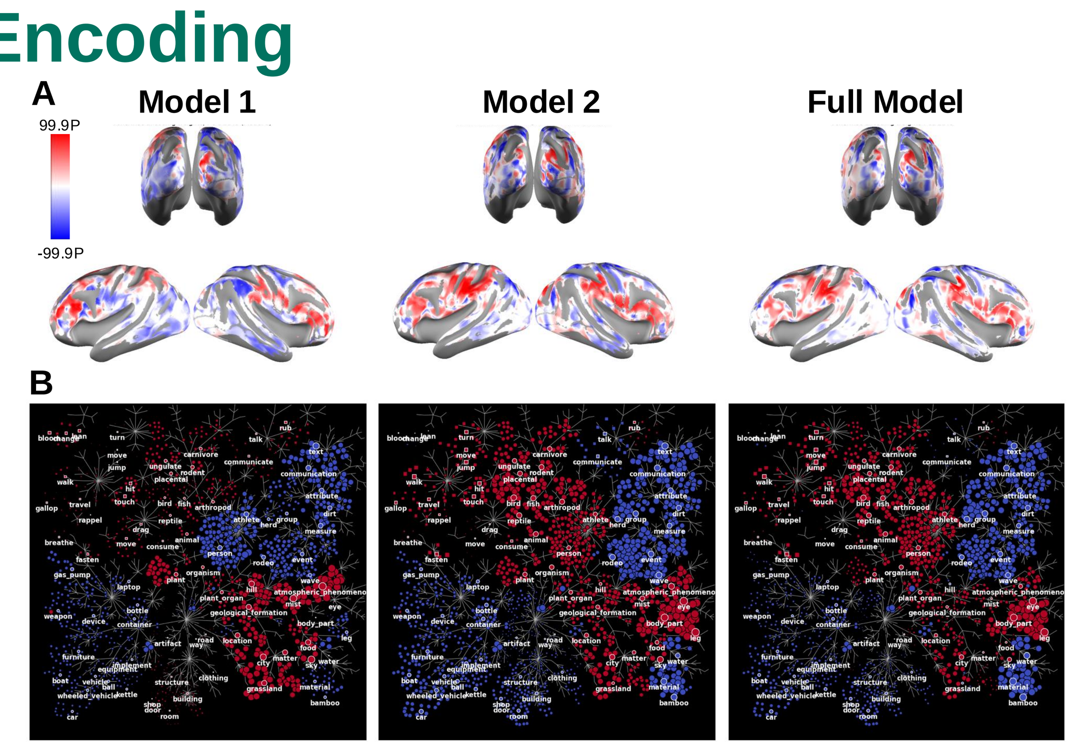


Figure 8. Visual Semantic Encoding. Here the encoding model from Figure 7 was applied. **(A)** Model prediction accuracy. **(B)** Extracted category weights for the "person" category.



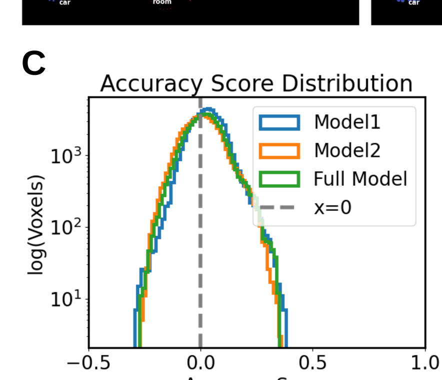


Figure 9. Test-Retest of Visual Semantic Encoding Model in Sub 2. Training data was split in half, and performance was compared to the model using all training data. Category model weights in PC space. **(A)** Maps and **(B)** corresponding wordnet graphs. **(C)** Prediction Accuracy of the encoding model.

Future Directions & Conclusions

- We demonstrated the feasibility of VHD-DOT for visual semantic mapping (encoding).
- VHD-DOT showed the reproducibility of semantic category mapping and the feasibility of 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).

[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.

[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).

[5] Figure from https://neurosciencenews.com/reading-brain-map-14740/ and Deniz, F., et al (2019). The Representation of Semantic Information Across Human Cerebral Cortex During Listening Versus Reading Is Invariant to Stimulus Modality. *Journal of Neuroscience*, 39(39), 7722–7736.

This work was funded by USA NIH grants U01EB027005, R01NS090874, R01EB03491902, 5T32EB014855. 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.