



Published in final edited form as:

*Neuroimage*. 2021 February 01; 226: 117516. doi:10.1016/j.neuroimage.2020.117516.

## Decoding visual information from high-density diffuse optical tomography neuroimaging data

Kalyan Tripathy<sup>a,\*</sup>, Zachary E. Markow<sup>a</sup>, Andrew K. Fishell<sup>a</sup>, Arefeh Sherafati<sup>a</sup>, Tracy M. Burns-Yocum<sup>a</sup>, Mariel L. Schroeder<sup>a</sup>, Alexandra M. Svoboda<sup>a</sup>, Adam T. Eggebrecht<sup>a</sup>, Mark A. Anastasio<sup>b</sup>, Bradley L. Schlaggar<sup>c</sup>, Joseph P. Culver<sup>a</sup>

<sup>a</sup>Department of Radiology, Washington University School of Medicine, St. Louis, Missouri, USA

<sup>b</sup>University of Illinois, Urbana-Champaign, Illinois, USA

<sup>c</sup>Kennedy Krieger Institute, Baltimore, Maryland, USA

### Abstract

**Background:** Neural decoding could be useful in many ways, from serving as a neuroscience research tool to providing a means of augmented communication for patients with neurological conditions. However, applications of decoding are currently constrained by the limitations of traditional neuroimaging modalities. Electroencephalography requires invasive neurosurgery, magnetic resonance imaging (MRI) is too cumbersome for uses like daily communication, and alternatives like functional near-infrared spectroscopy (fNIRS) offer poor image quality. High-density diffuse optical tomography (HD-DOT) is an emerging modality that uses denser optode arrays than fNIRS to combine logistical advantages of optical neuroimaging with enhanced image quality. Despite the resulting promise of HD-DOT for facilitating field applications of neuroimaging, decoding of brain activity as measured by HD-DOT has yet to be evaluated.

---

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

\*Corresponding author. Couch Biomedical Research Building, 4515 McKinley Avenue, 2<sup>nd</sup> Floor, St. Louis, MO 63110, USA. [tripathyk@wustl.edu](mailto:tripathyk@wustl.edu).

#### Credit Author Statement

Kalyan Tripathy: conceptualization, methodology, software, formal analysis, investigation, data curation, visualization, funding acquisition, writing (original draft, review and editing). Zachary Markow: methodology, software, validation, formal analysis, resources, writing (review and editing). Andrew Fishell: conceptualization, methodology, software, writing (review and editing). Arefeh Sherafati: investigation, data curation, resources, writing (review and editing). Tracy Burns-Yocum: investigation, data curation, writing (review and editing). Mariel Schroeder: investigation, data curation, writing (review and editing). Alexandra Svoboda: investigation, data curation, writing (review and editing). Adam Eggebrecht: software, resources, visualization, writing (review and editing). Mark Anastasio: conceptualization, writing (review and editing). Bradley Schlaggar: conceptualization, methodology, funding acquisition, writing (review and editing). Joseph Culver: conceptualization, methodology, formal analysis, resources, supervision, project administration, funding acquisition, writing (review and editing).

**Publisher's Disclaimer:** This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

#### Data And Code Availability Statement

To ensure transparency, facilitate reuse of our data, and encourage comparative analyses by other groups, our data will be made available upon request and can be obtained by contacting the corresponding author. Code for pre-processing and reconstructing HD-DOT data is available through Github ([https://github.com/WUSTL-ORL/NeuroDOT\\_Beta](https://github.com/WUSTL-ORL/NeuroDOT_Beta)), and additional code specific to decoding experiments can also be obtained by contacting the corresponding author.

**Declarations of interest:** none

**Objective:** To assess the feasibility and performance of decoding with HD-DOT in visual cortex.

**Methods and Results:** To establish the feasibility of decoding at the single-trial level with HD-DOT, a template matching strategy was used to decode visual stimulus position. A receiver operating characteristic (ROC) analysis was used to quantify the sensitivity, specificity, and reproducibility of binary visual decoding. Mean areas under the curve (AUCs) greater than 0.97 across 10 imaging sessions in a highly sampled participant were observed. ROC analyses of decoding across 5 participants established both reproducibility in multiple individuals and the feasibility of inter-individual decoding (mean AUCs > 0.7), although decoding performance varied between individuals. Phase-encoded checkerboard stimuli were used to assess more complex, non-binary decoding with HD-DOT. Across 3 highly sampled participants, the phase of a 60° wide checkerboard wedge rotating 10° per second through 360° was decoded with a within-participant error of 25.8±24.7°. Decoding between participants was also feasible based on permutation-based significance testing.

**Conclusions:** Visual stimulus information can be decoded accurately, reproducibly, and across a range of detail (for both binary and non-binary outcomes) at the single-trial level (without needing to block-average test data) using HD-DOT data. These results lay the foundation for future studies of more complex decoding with HD-DOT and applications in clinical populations.

## Keywords

Decoding; High-density diffuse optical tomography; Functional neuroimaging; Retinotopy

## 1. Introduction

While much cognitive neuroscience research has focused on mapping brain regions that are activated while participants perform tasks, recent work has emphasized the value of the reverse analysis known as decoding, i.e., deducing task information from recordings of brain activity (Haynes and Rees, 2006; Hebart and Baker, 2018). Potential applications of decoding range from building brain-computer interfaces that could restore movement or communication in paralyzed patients (Abdalmalak et al., 2017; Collinger et al., 2013) to recreating scenes from brain activity measured during visual experiences (Kay et al., 2008; Nishimoto et al., 2011; Wen et al., 2018), imagination (Albers et al., 2013; Horikawa and Kamitani, 2017), memory tasks (Harrison and Tong, 2009), or dreams (Horikawa et al., 2013). However, decoding applications are constrained by the limitations of established neuroimaging modalities. For instance, many decoding research studies so far have relied on electrocorticography (ECoG), which is invasive, or functional magnetic resonance imaging (fMRI), which is not conducive to imaging in certain populations and applications due to its logistics (Anumanchipalli et al., 2019; Haynes and Rees, 2006; Pinti et al., 2020). Functional near-infrared spectroscopy (fNIRS) addresses some of these issues as a noninvasive, portable, optical imaging method that supports imaging in an open environment (Pinti et al., 2020). However, the low channel count of traditional fNIRS limits its spatial resolution, coverage, and image quality, thereby constraining the complexity and precision of decoding (Abdalmalak et al., 2017; Emberson et al., 2017; White and Culver, 2010a). High-density diffuse optical tomography (HD-DOT) is an emerging optical neuroimaging modality that uses a dense array of light sources and detectors to capture thousands of overlapping

measurements, overcoming several limitations of traditional fNIRS and more closely matching fMRI in extensive brain mapping studies (Eggebrecht et al., 2014, 2012; Gregg et al., 2010; White and Culver, 2010a, 2010b; Zeff et al., 2007). However, decoding of brain activity as measured by HD-DOT has yet to be evaluated. Here, we examine the feasibility of decoding visual stimulus information from evoked brain activity measured with HD-DOT, assessing accuracy, reproducibility, and the level of detail obtainable.

Prior research using other neuroimaging modalities has made a compelling case for the clinical and neuroscientific utility of decoding (Brandman et al., 2018; Haynes and Rees, 2006; Hebart and Baker, 2018). For one, decoding could provide a means of augmented communication for patients who cannot speak or move effectively as a result of various neurological disorders including, but not limited to, stroke, neurodegenerative diseases, and developmental conditions like cerebral palsy. The complexity of decoding-based communication has so far ranged from obtaining yes-no responses using traditional fNIRS noninvasively (Abdalmalak et al., 2017) to generating comprehensible speech from ECoG arrays in epilepsy patients undergoing neurosurgical treatment (Anumanchipalli et al., 2019). Neuroimaging signals have also been decoded to drive motor prosthetics for amputees and paralyzed patients (Brandman et al., 2018). Success with decoding motor signals has ranged from noninvasive but coarse decoding of binary movement direction using fNIRS in healthy participants (Sitaram et al., 2007) to detailed control of robotic arms using intracortical electrodes in paraplegic patients (Collinger et al., 2013). Furthermore, cognitive neuroscience studies have decoded naturalistic images (Kay et al., 2008), movies (Nishimoto et al., 2011; Wen et al., 2018), speech (Correia et al., 2014, 2015), music (Hoefle et al., 2018), dreams (Horikawa et al., 2013), and semantic content (Huth et al., 2016) from fMRI data and provided a powerful window into human brain function (Hebart and Baker, 2018).

While promising, these applications of decoding are currently constrained by the limitations of mainstream neuroimaging modalities. The advanced decoding achieved using ECoG has remarkable performance, but implantation of the electrode arrays requires invasive neurosurgery and is therefore limited to small clinical populations (Anumanchipalli et al., 2019). Meanwhile, fMRI is noninvasive, but is still challenging for certain populations, such as young children, and contraindicated in others, such as patients with implanted metallic devices. In addition, it is not feasible for patients to regularly use cumbersome, technically demanding, and prohibitively expensive MRI scanners for applications such as longitudinal communication.

In contrast, fNIRS is much more portable and cost-effective, allowing widespread use and longitudinal imaging in more natural settings (Pinti et al., 2020). In addition, fNIRS is without major contraindications: the instrumentation is compatible with implants, and the open imaging environment is well suited to imaging children with a parent nearby for comfort. However, the sparse optode arrays and low channel counts used in traditional fNIRS compromise spatial resolution, coverage, and image quality, which are likely to limit the information available for accurate and detailed decoding (Fishell et al., 2020; White and Culver, 2010a).

HD-DOT utilizes a comparatively dense arrangement of light sources and detectors to collect ten-fold to a hundred-fold more measurements than traditional, sparse fNIRS. The numerous overlapping channels of HD-DOT enable higher spatial resolution, tomographic reconstruction, superficial signal regression, and a reduction in signal localization artifacts (Fishell et al., 2020; Gregg et al., 2010; White and Culver, 2010a). As a result, HD-DOT has been found to have a similarly high signal-to-noise ratio and high spatial concordance compared to fMRI in brain mapping studies using phase-encoded visual stimuli (Eggebrecht et al., 2012; White and Culver, 2010b), hierarchical language tasks (Eggebrecht et al., 2014; Hassanpour et al., 2015), and resting state functional connectivity analysis (Eggebrecht et al., 2014). By combining logistical advantages of optical neuroimaging with improved space-bandwidth product (i.e., the field-of-view divided by the point-spread-function, or roughly the number of independent voxels) and image quality, HD-DOT holds the potential to support detailed, accurate decoding in naturalistic environments and advance the clinical and neuroscience applications of decoding.

In the current study, we evaluate the feasibility and performance of visual decoding with HD-DOT, focusing on decoding the positions of checkerboard stimuli. The visual system is a particularly useful model for proof of principle, due to the elaborate, consistent, and well-characterized organization of visual features in neuroanatomical space (Hubel and Wiesel, 1962). Indeed, early fMRI decoding research also focused on visual decoding and gradually progressed from decoding fundamental visual features such as stimulus position (Thirion et al., 2006) and line orientation (Kamitani and Tong, 2005) to eventually performing intricate reconstructions of naturalistic images and movies (Kay et al., 2008; Nishimoto et al., 2011; Wen et al., 2018). Furthermore, prior work with HD-DOT has validated retinotopic mapping with HD-DOT against the gold standard of fMRI (Eggebrecht et al., 2012; Zeff et al., 2007), making retinotopic decoding a promising initial goal.

Herein, we first establish, in healthy adults, the feasibility of binary visual decoding with HD-DOT using a template matching strategy, and we evaluate its sensitivity, specificity, and accuracy across a range of thresholds using receiver operating characteristic (ROC) analysis. We then assess the reproducibility of this binary decoding across multiple imaging sessions and multiple participants. Finally, we extend the analysis of decoding performance to non-binary cases. We use phase-encoded retinotopic stimuli to evaluate the feasibility, accuracy, and replicability of 18-way and 36-way classifications of stimulus position across different parts of the visual field at individual time points without block-averaging test data. These studies reveal that HD-DOT can allow sensitive, specific, and reproducible retinotopic decoding, encouraging future studies of more complex decoding paradigms and applications in atypical populations.

## 2. Methods

### 2.1. HD-DOT imaging

This study aims to investigate the decoding performance of optical neuroimaging using a HD-DOT imaging array with an increased space-bandwidth product relative to traditional fNIRS. Imaging was performed using a previously characterized continuous-wave HD-DOT system that illuminates the back and sides of the head with 750nm and 850nm light through

a grid of 96 LED sources interlocked with 92 APD detectors, collectively yielding over 1200 usable source-detector measurements per wavelength. The weight of the fiber optics is supported by an extruded aluminum frame and two suspended wooden rings. Fiber tips contact the head via a custom-built cap that spaces the optodes 13 mm apart (hence with first-nearest through fourth-nearest source-detector separations of 1.3, 3.0, 3.9 and 4.7 cm) across the posterior and lateral surfaces of the head (Eggebrecht et al., 2014). During each cap fit, the participant's hair was first parted and tied if necessary to minimize obstruction of light transmission within the field-of-view. The participant was then asked to sit in a chair placed below the HD-DOT cap, hold the front straps of the cap, and comb the cap's optodes through their hair and up against the scalp surface. Anatomical markers such as the tragus and inion were used to guide cap positioning across sessions and participants. Specifically, the vertical position of the cap was adjusted such that the lowest row of optodes contacted the head at the level of the inion, and the horizontal distance between reference optodes and the tragus was measured on each side and the positioning of the cap adjusted accordingly to ensure symmetry. The straps of the cap were then tightened and fastened. Real-time data quality metrics such as light and noise levels were used to guide any further optimization of the cap fit, e.g. combing any occluded optodes through any obstructing hair to ensure uniform optode-to-scalp coupling and light levels across the field of view. Photographs of the imaging cap placement were taken from both sides, from both upper and lower viewing angles, and used to confirm cap placement during data processing.

## 2.2 Data processing

Following data acquisition, data were pre-processed, reconstructed, and subjected to spectroscopy, as summarized below and detailed in Eggebrecht et al. 2014.

**2.2.1 Pre-processing:** Raw light measurements were converted to log ratio time series using the temporal mean of each measurement as its relative baseline. Noisy channels with >7.5% variance across a run were excluded from further analysis, as this excessive variance was likely to reflect non-physiological nuisance signals such as head motion as opposed to hemodynamic changes associated with brain activity (Eggebrecht et al., 2014). High-pass filtering with a 0.02-Hz cutoff was performed to reduce long-term drift. Subsequently, superficial signal regression was performed by first averaging all first-nearest-neighbor measurements, which sample mostly the scalp, to use as an estimate of global systemic signals. This global superficial signal time trace was then subtracted out of every source-detector measurement time trace using linear regression. This approach has been shown to work in conjunction with HD-DOT to improve contrast-to-noise ratio, and the use of a single superficial signal regressor avoids both overfitting and removal of brain activation signals (Eggebrecht et al., 2014; Gregg et al., 2010; Zeff et al., 2007). Low-pass filtering (with a 0.5-Hz cutoff) removed residual pulse and other high-frequency noise, and data were then downsampled to 1 Hz (Eggebrecht et al., 2014).

**2.2.2 Light modeling and reconstruction:** Pre-processed data were reconstructed using an anatomically based light propagation model generated from the Montreal Neurological Institute (MNI) non-linear ICBM152 atlas (Ferradal et al., 2014; Fonov et al., 2011; Mazziotta et al., 2001), using Freesurfer for segmentation (Dale et al., 1999; Fischl et

al., 2001; Ségonne et al., 2004), NIRVIEW for mesh generation, finite element modeling with spring relaxation approaches for source and detector positioning, and the NIRFAST toolbox for determining a solution to the optical diffusion equation so as to model photon diffusion through the head (Dehghani et al., 2008, 2003; Eggebrecht et al., 2014). The resulting sensitivity matrix  $A$  is the linear transformation for each time point between  $x$ , the vector of absorption coefficients at 750nm and 850nm within the brain volume, and  $y$ , the vector of relative changes in measured light levels at each wavelength detected at the head surface, as per the linear Rytov approximation:

$$y = A x \quad (1)$$

This sensitivity matrix was inverted using Tikhonov regularization ( $\lambda_1=0.01$ ) and spatially variant regularization ( $\lambda_2 = 0.1$ ), following Eggebrecht et al. (2014). The wavelength-dependent absorption and scattering coefficients (units  $\text{mm}^{-1}$ ) for the five non-uniform tissue compartments were as follows: scalp ( $\mu_a, 750 = 0.017$ ;  $\mu_a, 850 = 0.019$ ;  $\mu_s, 750' = 0.74$ ;  $\mu_s, 850' = 0.64$ ), skull ( $\mu_a, 750 = 0.012$ ;  $\mu_a, 850 = 0.014$ ;  $\mu_s, 750' = 0.94$ ;  $\mu_s, 850' = 0.84$ ), cerebrospinal fluid ( $\mu_a, 750 = 0.004$ ;  $\mu_a, 850 = 0.004$ ;  $\mu_s, 750' = 0.3$ ;  $\mu_s, 850' = 0.3$ ), grey matter ( $\mu_a, 750 = 0.018$ ;  $\mu_a, 850 = 0.019$ ;  $\mu_s, 750' = 0.84$ ;  $\mu_s, 850' = 0.67$ ), and white matter ( $\mu_a, 750 = 0.017$ ;  $\mu_a, 850 = 0.021$ ;  $\mu_s, 750' = 1.19$ ;  $\mu_s, 850' = 1.01$ ) (Bevilacqua et al., 1999; Custo et al., 2006; Eggebrecht et al., 2012; Fishell et al., 2019; Strangman et al., 2002).

**2.2.3 Spectroscopy:** Relative changes in oxy- and deoxy- hemoglobin concentrations were calculated from the differential absorption image  $x$  at each time point through spectral decomposition:

$$\Delta C = E^{-1}x \quad (2)$$

where  $C$  is a vector of oxy- and deoxy-hemoglobin concentration changes across the brain volume, and  $E$  is a matrix of extinction coefficients of oxy- and deoxy- hemoglobin.

### 2.3. Participants

While functional domains of the brain are generally common across individuals, there are readily quantified individual-specific differences in functional localization even after anatomical spatial normalization (Gordon et al., 2017). Because of this high inter-individual variability, it has been common for fMRI decoding studies to focus on 2-3 highly sampled research participants in order to amass sufficient data per participant (Dumoulin and Wandell, 2008; Kay et al., 2008; Nishimoto et al., 2011). Here, we adopted a similar approach of collecting large amounts of data in a few participants to analyze decoding accuracy and reproducibility across multiple imaging sessions. We also collected smaller quantities of data across other participants to assess reproducibility across individuals. In total, 8 healthy adults participated in this study (age range 24-54 years, 7 female). In the first phase of the study (binary decoding), we collected extensive data in participant 1 across 10 imaging sessions, and evaluated reproducibility across  $n=5$  individuals (participants 1-5). In the second phase of the study (non-binary decoding), we used pilot data from participant 1 to optimize decoding parameters, then analyzed 18-way and 36-way decoding extensively in

participant 6, and finally evaluated reproducibility of the more complex 36-way decoding across  $n=3$  individuals (participants 6-8). The participants' demographic information and the data that they contributed to each experiment are detailed in Table 1. We intentionally selected participants whom we expected to provide high-quality data, based on prior studies, as we aimed to evaluate the feasibility and performance of decoding without the confounds of poor data quality.

Informed consent was obtained from all participants in accordance with the IRB protocol approved by the Human Research Protection Office at Washington University School of Medicine.

#### 2.4. Stimulus protocols

Participants were imaged with HD-DOT while they performed multiple runs of three different types of checkerboard viewing tasks, adapted from prior retinotopic mapping studies (Eggebrecht et al., 2012; White and Culver, 2010b, 2010a; Zeff et al., 2007). The Psychophysics Toolbox 3 package for MATLAB was used to display the visual stimuli (Brainard, 1997). In all tasks, a checkerboard grid pattern was displayed, consisting of a black and white grid that reverses at 8 Hz in time against a 50% gray background, changing over time as follows:

**Task 1 (binary decoding):** Left- and right-sided checkerboard wedge stimuli provided an intuitive starting point to begin studying decoding as previous studies show that these two stimulus conditions produce distinct, reproducible, well-defined activation maps (Zeff et al., 2007). Participants were asked to maintain central fixation through multiple rounds of viewing a flickering checkerboard wedge for 10 seconds interspersed with rest periods of viewing just a fixation cross for 24 seconds. Longer inter-stimulus rest periods of 48s and 72s were used during a subset of the task runs to check whether this had any effects on stimulus response maps and decoding performance, but none were seen indicating that 24 seconds was sufficient to separate stimulus presentations. The checkerboard wedges extended over a polar angle of  $70^\circ$  and a radial angle of  $2.5\text{-}10^\circ$ . During each block, the checkerboard stimulus was presented on either the lower left quadrant or the lower right quadrant of the screen. Over the course of each task run, the stimulus was presented an equal number of times on each side in a pseudorandom order. The number of repetitions ranged from 5-8 depending on the duration of the inter-stimulus interval; fewer repetitions were delivered when the inter-stimulus interval was extended beyond 24s to limit total run time.

**Task 2 (18-way decoding):** To create a more challenging decoding task with a greater number of targets, we presented participants with expanding/contracting checkerboard ring stimuli. Participants were asked to maintain central fixation while a flickering checkerboard ring expanded or contracted for 8 cycles through 18 positions on the screen at 2 seconds per position.

**Task 3 (36-way decoding):** For another evaluation of decoding with a greater number of targets, we also presented participants with rotating checkerboard wedge stimuli. Participants maintained central fixation while a flickering checkerboard wedge rotated

through 10 revolutions, either clockwise or counterclockwise. The wedge subtended a polar angle of  $60^\circ$ , a radial angle of  $2.5\text{--}10^\circ$ , and rotated  $10^\circ$  at a time through 36 positions spanning  $360^\circ$  on the screen at 1 second per position (i.e.,  $10^\circ$  per second, or 36 seconds per revolution).

## 2.5. Decoding by template matching

For an initial assessment of the feasibility of decoding HD-DOT, we wanted to use a simple classification algorithm and hence chose a template matching strategy. For every decoding attempt, distinct task runs were used for training and testing. The training data were always derived from a single task run that included 5 to 10 rounds of stimulus presentation, block-averaged to construct oxyhemoglobin signal maps for different stimulus conditions. These mean activation maps served as “templates” of brain activity corresponding to each stimulus response, in voxel space and spatially normalized to the MNI atlas as detailed in section 2.2. To create the templates at maximum signal-to-noise ratio, we used a data-driven, paradigm-specific time window to capture the peak signal. While the hemodynamic response is fixed, the signal measured is a convolution of the hemodynamic response function and the stimulus timing (Boynton et al., 1996), with the latter varying across tasks. Longer presentations of the stimulus in each position delay the peak response (Lindquist et al., 2009). For Task 1, in which stimulus presentations were most spread out, left-sided and right-sided template maps were constructed by block-averaging the signal across a 5-second time window starting 10 seconds after stimulus onset, as this coincided with the peak signal based on prior hemodynamic studies (Supplemental Figure 1A) (Gregg et al., 2010; Zeff et al., 2007). For Task 2, the stimulus only spent 2 seconds in each position at a time, so each template was constructed by averaging data across blocks from two time points, 7 and 8 seconds after the checkerboard had passed through the position under consideration. This 7-8 second time window was empirically optimized using pilot data from participant 1 by evaluating decoding performance for a range of time delays and selecting the time delay that minimized decoding error (Supplemental Figure 1B). For Task 3, the stimulus rotated at 1 second per position, so each template was constructed by block-averaging data from a single time point across the 10 revolutions of the stimulus. Here, the optimal time delay was found to be 6 seconds (Supplemental Figure 1C–1D).

**Decoding with 2 templates:** Initially, stimulus state was decoded at every time point to evaluate the feasibility of decoding at individual time points (Figure 1). This demonstrates the temporal quality of the decoding. To perform a statistical analysis of binary decoding performance (Figures 2–4), we restricted the data to clearly independent trials, reasoning that individual time points were not all independent due to the temporal blurring of the hemodynamic response. Each stimulus presentation and each inter-stimulus rest period in the test data were treated as separate trials. A trial response was defined as the oxyhemoglobin signal map during the single frame 12 seconds after the onset of the stimulus presentation or the inter-stimulus interval (to coincide with the middle of the time window used to construct the stimulus response templates), in the same MNI atlas voxel space as the templates. We then evaluated binary decoding performance on a trial-wise basis. Each trial response was compared to both the left-sided and the right-sided templates by calculating a spatial Pearson correlation coefficient  $r_{n,m}$  between the  $n^{\text{th}}$  template  $T_n$  and the



oxyhemoglobin response map for the  $m^{\text{th}}$  trial  $S_m$  as follows, where both  $T_n$  and  $S_m$  were zero-mean-centered (i.e.,  $\text{mean}(T_n)=0$ ,  $\text{mean}(S_m)=0$ ):

$$r_{n,m} = \frac{(T_n) \cdot (S_m)}{|T_n||S_m|} \quad (3)$$

These correlation coefficients were compared to an adjustable threshold to determine the decoding output ( $D_m$  for each trial  $m$ ). The threshold value was initially set ad hoc for Figure 1 and later optimized using ROC analysis as described in section 2.6 and Figure 2. If all values of  $r_{n,m}$  on the  $m^{\text{th}}$  trial fell below the threshold, the decoding output was set to  $D_m = 0$  (i.e., decoded as rest). If  $r_{n,m}$  rose above the threshold for only the  $n^{\text{th}}$  template, the decoding output was set to  $D_m = n$ . If  $r_{n,m}$  rose above the threshold for multiple templates in one trial, the decoding output was determined by the template with the maximum correlation value  $r_{n,m}$  during that trial:

$$D_m = \text{argmax}_n(r_{n,m}) \quad (4)$$

**Decoding with 18 and 36 templates:** The template matching approach was similar for the non-binary decoding experiments, except that decoding involved a larger number of templates (either 18 or 36) and a decoding output was computed for every time point in the test run. Specifically, a spatial Pearson correlation  $r_{n,t}$  was calculated between the  $n^{\text{th}}$  template  $T_n$  and the oxyhemoglobin response map  $S_t$  at the  $t^{\text{th}}$  time point for each of the 18 or 36 templates and every single time point in the test data:

$$r_{n,t} = \frac{(T_n) \cdot (S_t)}{|T_n||S_t|} \quad (5)$$

Here,  $T_n$  and  $S_t$  were again first zero-mean-centered (i.e.,  $\text{mean}(T_n)=0$ ,  $\text{mean}(S_t)=0$ ). The decoding output  $D_t$  at each time point was determined by the template number  $n$  that had the maximum correlation with the oxyhemoglobin signal map at time  $t$ :

$$D_t = \text{argmax}_n(r_{n,t}) \quad (6)$$

## 2.6. Evaluation of binary decoding performance

After calculating the full vector of decoding outputs ( $D_m$ ), we evaluated decoding performance by comparing  $D_m$  to the ground truth of the study design. ROC analysis was conducted in order to quantify the sensitivity and specificity of decoding across the full range of possible thresholds (Metz, 2006, 1978). At each threshold value,  $D_m$  was compared to the true stimulus state for each trial to classify every trial as a true or false positive or negative. For each threshold, the sensitivity, specificity and Youden  $J$  statistic (Youden, 1950) were then calculated as follows:

$$\text{sensitivity} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} \quad (7)$$

$$\text{specificity} = \frac{\text{true negatives}}{\text{true negatives} + \text{false positives}} \quad (8)$$

$$J = \text{sensitivity} + \text{specificity} - 1 \quad (9)$$

Sweeping the correlation threshold across the full range of possible values (from  $-1$  through  $0$  to  $+1$  in increments of  $0.01$ ) allowed us to plot ROC curves for each possible pairwise classification, i.e., one ROC curve for distinguishing left-versus right-sided checkerboard stimulus presentations, one for decoding left-sided and rest trials, and one for differentiating right-sided and rest trials (Metz, 2006, 1978). Areas under the curves were computed as a metric for decoding performance. The maximum value of the Youden  $J$  statistic was used to determine the optimal threshold value balancing sensitivity and specificity of decoding.

Reproducibility of binary decoding was assessed across 18 task runs in one highly sampled participant by changing which task run was used for template construction, pooling all the remaining runs as test data, and then repeating the decoding and ROC procedure for every possible template.

As an additional measure of reproducibility, ROC analysis was also repeated using only a single task run as test data (rather than pooling test data across runs), but for every possible pairing of template and test run. This approach was used to evaluate reproducibility both across sessions in one highly sampled participant as well as across multiple participants.

## 2.7. Evaluation of 18-way and 36-way decoding performance

For the 18-way and 36-way decoding experiments, decoding error  $E_t$  was calculated at every time point  $t$  in the test task run as the absolute value of the difference between the decoded stimulus position ( $D_t$ ) and the actual stimulus position ( $S_{t+I}$ ), after sliding the entire true stimulus position time course by a lag time  $I$  to allow for the delayed hemodynamic response and match the time window of template construction (i.e.,  $I = 8$  s for Task 2 and  $I = 6$  s for Task 3):

$$E_t = |D_t - S_{t+I}| \quad (10)$$

This error value was then averaged across all time points in the task run to calculate mean decoding error.

In order to assess whether the resulting mean decoding error was significantly different from chance-level performance, permutation tests were conducted. For example, to evaluate 36-way decoding performance across sessions within a participant, decoding and mean error calculation procedures were repeated for all 30 possible pairings of training and test data set across 6 task runs in one participant to generate a distribution of mean decoding error. The

template labels in the training data were then randomly shuffled 10 times for each of the 30 decoding attempts, and the same process of template matching and error calculation was repeated each time to generate a null distribution of decoding error values derived using 300 shuffled template sets. This null distribution was compared to the error distribution for true decoding attempts to evaluate the statistical significance of our decoding performance.

To further study decoding performance across different parts of the visual field, mean decoding error was also calculated for each possible position of the stimulus rather than simply averaging across all positions.

Finally, to evaluate the reproducibility of this detailed decoding both within and across individuals, a total of 3 participants performed the rotating checkerboard task 6 times each over the course of 2 imaging sessions involving 2 separate cap fits per participant. The decoding and error calculation process was repeated for every possible pairing of training and test data sets, and the resulting matrix of error values was used to evaluate reproducibility of decoding within and between sessions and participants. Permutation testing was again conducted to test statistical significance, comparing the observed mean decoding error to the error distribution resulting from decoding with 10 random permutations of the training data set for every decoding attempt.

## 2.8. Data and code sharing

To ensure transparency, facilitate reuse of our data, and encourage comparative analyses by other groups, our data will be made available upon request and can be obtained by contacting the corresponding author. Code for pre-processing and reconstructing HD-DOT data is available through Github ([https://github.com/WUSTL-ORL/NeuroDOT\\_Beta](https://github.com/WUSTL-ORL/NeuroDOT_Beta)), and additional code specific to decoding experiments can also be obtained by contacting the corresponding author.

## 3. Results

### 3.1. Feasibility of decoding visual stimulus position from HD-DOT data

We first assessed the feasibility of decoding HD-DOT data using a simple stimulus position classification problem. Participants were imaged with HD-DOT while they performed multiple runs of a block-design visual task, in which a flickering checkerboard wedge was presented to either the left or the right visual hemi-field interspersed with rest periods of no checkerboard stimulus (Zeff et al., 2007). We anticipated that these three stimulus conditions (left-sided checkerboard, right-sided checkerboard, and no checkerboard) would produce spatially separable HD-DOT activations. A template matching strategy was therefore used to decode checkerboard position. Using one <10-minute task run as training data, we mapped templates of the oxyhemoglobin response to the left- and right-sided checkerboard stimuli. In a second independent task run used as test data, the stimulus condition was decoded as either left-sided, right-sided, or rest at each time point, based on the spatial correlation between the HD-DOT signal map at that time and each of the templates (as explained in section 2.5). Time traces of the Pearson correlation with each template, the decoded

stimulus, and the true stimulus condition for a typical decoding attempt in one participant illustrate the general agreement between decoded and actual stimulus (Figure 1).

### 3.2. Quantitative assessment of binary decoding accuracy

In order to objectively quantify decoding performance, we converted our three-way decoding problem into three binary classifications (left-sided versus no checkerboard, right-sided versus no checkerboard, and left-sided versus right-sided stimulus), and performed ROC analysis for each case (Metz, 2006, 1978). Furthermore, to ensure that each test data point was temporally independent, we calculated decoding accuracy on a trial-wise basis rather than across every time point. Each stimulus presentation and inter-stimulus rest period in the test data was treated as an independent event or “trial”. To maintain a large sample of data points for a robust analysis amid this trial-wise evaluation of sparse block-design task data, test data were pooled across multiple task runs and imaging sessions where available. For example, one highly sampled participant performed the same checkerboard-viewing task 18 times over the course of 10 imaging sessions, so data from a single task run was used to construct templates, and the other 17 runs were pooled as test data. A plot of correlations for all of this participant’s trial responses to the left and right templates reveals a clustering based on stimulus condition (Figure 2A).

The thresholds separating different decoding outputs were swept across the full range of possible values, and true and false positive rates of decoding were calculated in each case to construct three ROC curves (Figure 2B). The large areas under all these curves (AUCs > 0.98) reflect the high sensitivity and specificity of decoding. The maximum value of the Youden  $J$  statistic (Youden, 1950) for each ROC curve was used to guide the selection of a data-driven optimal threshold value, and sensitivity, specificity, and overall accuracy were calculated at this threshold (Table 2).

We also evaluated the feasibility and performance of decoding the deoxyhemoglobin signal in place of the oxyhemoglobin signal (Supplemental Table 1).

### 3.3. Reproducibility of decoding across imaging sessions

One approach that fMRI studies have taken to assess reproducibility is to collect large quantities of data on individual participants across multiple task runs and imaging sessions and then compare the similarity of results between all possible pairings of the data sets (Gordon et al., 2017). Adapting this approach to our HD-DOT decoding study, we used data from the participant who performed the same checkerboard-viewing task 18 times over the course of 10 different imaging sessions, spanning more than 1 year. We repeated the ROC analysis described in section 3.2 17 more times, changing which of the 18 task runs was used for template construction each time and again pooling all the remaining data as test data, to produce 18 ROC curves for each of the 3 binary classification problems (Figure 3A). The consistently high AUC values (mean AUCs = 0.98 for left vs. rest, 0.99 for left vs. right, 0.99 for right vs. rest) reflect the high sensitivity and specificity of decoding. As an alternate approach to evaluating reproducibility within this data, we also repeated our decoding and ROC analysis using single task runs for template construction again but then using only single task runs as test data (rather than pooling test data across runs), for every possible pair

of template and test run (Figure 3B). While the number of test trials is lower here than in Figure 3A by design, we still find consistently high AUC values across each matrix of >300 possible pairings of template and test data (mean AUCs = 0.98 for left vs. rest, 0.99 for left vs. right, 0.99 for right vs. rest).

### 3.4. Reproducibility of binary decoding across participants and inter-subject decoding

We also sought to investigate how replicable our decoding was across multiple different participants, and hence took HD-DOT data from four additional healthy adults performing the same checkerboard-viewing task twice each. We conducted ROC analysis of decoding performance within all five participants as described in section 3.2, here using one task run in one participant as training data to decode the other task run in that participant (Figure 4A).

In addition, we conducted a reproducibility assay similar to that described in section 3.3 except across participants, evaluating the feasibility of inter-individual decoding. For this experiment we used one task run from one participant as training data to decode a second task run from each of the participants, constructing ROC curves to assess performance for each decoding attempt. This analysis was iterated to evaluate decoding for every possible combination of training and test participant (Figure 4B). Mean AUC values were 0.72 for left vs. rest, 0.82 for left vs. right, and 0.84 for right vs. rest, with a majority of the values ranging from 0.75 to 1.0 both along the matrices' diagonals and across off-diagonal elements. These results indicate that both within-subject decoding and between-subject decoding were effective across multiple participants. A one-sample Wilcoxon signed rank test was used to confirm that inter-participant decoding performance was significantly better than chance ( $p = 0.011$  for left vs. rest;  $p = 2.9 \times 10^{-4}$  for left vs. right;  $p = 2.5 \times 10^{-4}$  for right vs. rest). Lower AUC values were observed for some pairings of training and test data sets with participants 4 and 5. Raw data quality may be one contributor to this variance (Supplemental Figure 2).

### 3.5. Complex, non-binary decoding:

18- and 36- way classification of moving stimulus position One of the main advantages of HD-DOT over sparser fNIRS imaging systems is the improved spatial resolution and image quality afforded by the high-density arrays of light sources and detectors with their thousands of overlapping measurements (Eggebrecht et al., 2014; White and Culver, 2010a). In order to harness the spatial resolution of HD-DOT and assess the feasibility of more elaborate decoding than the binary classification established so far, we imaged participants while they observed several patterns of moving checkerboard stimuli. As there were 18 possible sizes for the expanding/contracting ring stimuli and 36 possible positions of the rotating checkerboard wedges, we constructed sets of 18 and 36 templates, respectively, for these two task paradigms by block-averaging single <7-minute task runs. We then used a template matching strategy to decode stimulus position at each time point in independent runs of the same task without any block-averaging of the test data. Graphs of Pearson correlation values for all the templates and plots of the actual and decoded stimulus positions are shown for a typical contracting ring task run (Figure 5A) and a typical rotating wedge task run (Figure 6A, Supplemental Movie 1).

The similarity between the actual and decoded stimulus traces indicates that decoding each time point was feasible even in these more complex 18-way and 36-way classification problems. After correcting for the hemodynamic time delay, visible in the 6-8 second lag between the actual and decoded stimulus traces (Figures 5A and 6A), the discrepancy between the actual and decoded stimulus positions was calculated for each time point and these values were averaged across the run as a quantitative measure of accuracy. For example, across all 30 of the 36-way rotating stimulus decoding attempts in one high-performing participant, the mean error ( $\pm$  standard deviation) in decoding stimulus position was  $18.0 \pm 17.4^\circ$  (compared to the mean error for decoding at chance which would have been  $90^\circ$ ). To assess whether decoding performance in this participant was significantly better than chance, we conducted a permutation test across all the data collected in this participant as described in Section 2.7. The separation between the true decoding error distribution and the null distribution illustrates the statistical significance of the results (Figure 5B,  $p < 0.0083$ , and Figure 6B,  $p < 0.0033$ ). Interestingly, breaking down the mean absolute error calculation for each stimulus position revealed no significant variation in decoding error with stimulus eccentricity (Figure 5C), but greater error when the checkerboard wedge rotated through the upper visual hemifield than the lower half of space (Figure 6C).

### 3.6. Reproducibility of non-binary HD-DOT decoding across participants

For the decoding run depicted in Figure 6A, both the training and test data were collected from the same participant and during the same imaging session. To further investigate the robustness of this more elaborate decoding, we imaged three different participants as they each viewed the rotating checkerboard stimulus sequence six times spread over the course of two imaging sessions involving two separate cap fits. We then repeated our decoding and error calculation analysis, training with one task run and testing with another independent run, for every possible pairing of training and test data set across all the participants and imaging sessions (Figure 7A–7B). Although performance varied between participants, we observed reproducible decoding across task runs and imaging sessions, with a mean within-participant decoding error ( $\pm$  standard deviation) of  $25.8 \pm 24.7^\circ$  relative to chance performance at  $90^\circ$  ( $p < 0.0011$ ). We also observed effective inter-participant decoding (i.e., when training and test data were from different individuals), with performance varying between sessions and participants, but mean inter-participant decoding error passing permutation testing for statistical significance ( $p < 0.0005$ ).

## 4. Discussion

To establish the feasibility of decoding with HD-DOT, we have used a well-validated visual stimulation protocol and have evaluated the accuracy and replicability of decoding stimulus position across multiple imaging sessions and research participants for a range of decoding complexity. Using a straightforward template matching strategy with training and test data from separate 5-10 minute-long task runs, we found that HD-DOT data can be used to accurately decode stimulus position at individual time points without needing to block-average test data. We performed ROC analysis to quantify the accuracy of binary visual decoding, obtaining mean AUC values  $> 0.97$  across 10 imaging sessions (including 18 task runs) in one highly sampled subject, and obtaining mean (intra- and inter-participant

decoding) AUC values of 0.7-0.85 across a group of 5 participants. More challenging 18-way and 36-way decoding experiments also showed strong decoding performance across multiple imaging sessions and participants. For these latter studies, the phase of a 60° wide checkerboard wedge rotating 10° per second through 360° was decoded with a mean error of 18.0±17.4° in our best participant across multiple task runs and cap fits. This decoding performance varied with data quality but remained significantly above chance based on permutation testing across imaging sessions in two additional participants. Inter-participant decoding, with training and test data taken from different individuals, was feasible for both the binary and the more complex decoding.

#### 4.1. Study design in relation to prior decoding research

Some of the most elaborate decoding research so far has used either ECoG or fMRI to sample brain activity and reconstruct things like intelligible speech and detailed visual experiences (Anumanchipalli et al., 2019; Nishimoto et al., 2011; Wen et al., 2018). However, ECoG is too invasive and fMRI too cumbersome for translation beyond clinical populations and laboratory studies into widespread use. While fNIRS and electroencephalography (EEG) overcome these challenges as noninvasive and portable imaging methods that have been used in previous decoding research, their low spatial resolution limits the space-bandwidth product available for decoding. On the one hand, decoding could have a meaningful impact even with a low bit rate; for instance, studies in patients with locked-in syndrome have shown the potential of fNIRS to enable patients with few other means of communication to respond yes or no to questions (Abdalmalak et al., 2017). However, by increasing the optical channel count and density relative to traditional fNIRS, HD-DOT combines logistical advantages of optical neuroimaging with image quality that is closer to that of fMRI, providing motivation to study the feasibility and performance of more detailed decoding with HD-DOT.

Retinotopic decoding was chosen as the focus of the current study for several reasons. Firstly, using an externally controlled stimulus, rather than studying internal thought or other higher-order cognitive functions, provided definite knowledge of ground truth and a means to manipulate it. This approach enabled a fairly simple experimental design to objectively quantify decoding accuracy. In addition, it had already been shown that the flickering checkerboard stimuli used here evoke repeatable and distinct brain activity patterns at different positions in the visual field that can be mapped by HD-DOT (Eggebrecht et al., 2012; White and Culver, 2010b; Zeff et al., 2007), making decoding the positions of these stimuli a reasonable goal. Finally, using the visual system as a model for initial proof of principle follows the wisdom of neuroscience literature in general, exemplified by Hubei and Wiesel's seminal plasticity studies (Hubei and Wiesel, 1965; Wiesel and Hubei, 1965a, 1965b, 1963), and follows the arc of the successful fMRI decoding literature in particular (Haynes and Rees, 2006; Kamitani and Tong, 2006, 2005; Thirion et al., 2006).

#### 4.2. Accuracy of binary visual decoding with HD-DOT

The feasibility of decoding with HD-DOT was evident in how closely our decoding results mirrored actual stimulus conditions. We employed ROC analysis to further evaluate decoding accuracy, objectively quantifying true and false positive rates across a range of

decoding threshold values (Metz, 2006, 1978). The consistently large area under the ROC curves highlights the high sensitivity, specificity, and accuracy of the single-trial level decoding presented (Figures 2B and 3, Table 2). Decoding the deoxyhemoglobin signal in place of oxyhemoglobin also yielded similarly high performance (Supplemental Table 1). While head-to-head comparison to previous fNIRS studies is difficult due to differences in study design, a study of binary audiovisual decoding in infants attained trial-wise decoding accuracy in the 55-70% range (Emberson et al., 2017). While our focus on simple stimuli and adult participants may have partly facilitated our higher decoding performance, another likely contributor to our gain in accuracy is the high channel count of HD-DOT (supporting ten-fold to a hundred-fold more measurements than most fNIRS arrays), which improves the space-bandwidth product of our data, increasing the amount of information that can be leveraged for decoding. These various contributing factors could be separated in future research.

#### 4.3. Reproducibility of binary visual decoding with HD-DOT across sessions and participants

Both the research and clinical applications of decoding hinge on its reliability, so we sought to assess the replicability of our results across multiple data sets. Recent fMRI studies have highlighted the advantages of conducting research with highly sampled individuals (Gordon et al., 2017; Laumann et al., 2015), so we chose to first investigate reproducibility in one participant across 18 task runs conducted over 10 different cap fits. Across combinations of training and test data, we obtained mean ROC AUC values  $>0.97$ , demonstrating the reproducibility of accurate binary visual decoding that can be achieved within a single participant, even when data is collected across multiple imaging sessions and cap fits over an extended period of time. This consistency of decoding across sessions illustrates the reproducibility of HD-DOT signals as well as the reliability of our cap fit procedure, which leverages anatomical landmarks, the structural integrity and flexibility of the imaging cap, and the use of feedback from real-time data quality visualizations (Eggebrecht et al., 2014). In addition, this observed reproducibility is encouraging for the continued pursuit of research and applications wherein a patient would rely on HD-DOT as a means of communication; it is technically feasible, for instance, that a HD-DOT decoder could be trained over one or more days, and then be used for decoding for months to years afterwards. This potential use case also further supports the study of highly sampled individuals.

Interestingly, there were some cases in which a run served as a consistently good test data set but not as reliable a training data set (e.g. run 6 for left vs. rest classification), and vice versa (e.g. run 16 was a consistent training data set but not always a good test data set for left vs. right decoding). These occasional asymmetries likely reflect differences in processing for training and test data. Templates were constructed from training data by block-averaging across a 5 second time window and across all presentations of a stimulus. Meanwhile, for test data, a single time point was used to assess decoding for each trial. Furthermore, inconsistencies in decoding performance for some specific pairings of data sets and not others may be a result of minor variations in the positioning of the cap between sessions. Future studies could investigate these potential sources of variance and attempt to further improve the consistency of decoding. Data was taken from four additional



participants and subjected to a similar analysis of decoding accuracy and reproducibility across individuals (Figure 4). Here, the within-participant ROC curves (Figure 4A) and the AUC values along the diagonal of the reproducibility matrices (Figure 4B) reflect high within-subject reliability in most of the participants, while off-diagonal matrix values illustrate that with HD-DOT data it is feasible to train a decoder on one individual and successfully infer the visual stimulus seen by another (Figure 4B). The inter-individual variability in decoding outcomes, with some participants yielding higher decoding accuracy than others, is likely partly related to variation in data quality between individuals associated with a combination of factors such as anatomical variability and participant compliance with regard to both maintaining central fixation and minimizing motion. Indeed, as illustrated in Supplemental Figure 2, participants with poor decoding performance had poor data quality to begin with, as assessed by signal-to-noise ratio and template laterality. This underscores the importance of monitoring data quality.

Nevertheless, the reproducibility of HD-DOT decoding across multiple sessions and individuals and the feasibility of inter-individual decoding suggest that it should be possible to pool training data across multiple imaging sessions and participants in future studies. While in this study we were able to train our decoder sufficiently with data from a single 5-10 minute task run, more complex decoding tasks will likely require significantly more data than could be acquired in one continuous session. For instance, prior visual decoding studies using fMRI data and more complex algorithms to reconstruct novel naturalistic stimuli outside of the training set collected 2-3 hours of training data in each of 3 participants to train their decoders without over-fitting and also block-averaged hours' worth of test data (Nishimoto et al., 2011; Wen et al., 2018). Furthermore, though subject-specific training data may often facilitate the most accurate decoding, the decoding performance we observed across sessions and individuals suggests that a decoder could be pre-trained with group data to enhance efficiency of training and performance of decoding in some cases. While inclusion of low quality training data would likely impair decoding performance, including high quality data from other individuals could even improve decoding; e.g. decoding left vs. right in participant 4 is even more effective using the high quality training data from participants 1 and 2 than the data from participant 4 themselves. The success of a pre-trained decoder could further be enhanced by anatomical and functional coregistration of group-level templates to a subject-specific space and by updating the decoder with additional subject-specific training data. Future studies can systematically assess the quantity of HD-DOT data required for an optimized pre-training approach and its effects on decoding performance.

#### **4.4. Performance of more spatially and temporally detailed retinotopic decoding with HD-DOT**

Finally, we reasoned that the increased space-bandwidth product afforded by our high-density imaging array may enable us to distinguish a larger number of targets than prior optical decoding studies that have mostly performed binary or 4-way decoding (Abdalmalak et al., 2020; Emberson et al., 2017; Hosseini et al., 2011; Luu and Chau, 2009; Sitaram et al., 2007). As a result, we increased the complexity of our decoding and performed the 18-way and 36-way classification experiments with the moving checkerboard ring and wedge

stimuli. We found that we were able to localize a 60° wide checkerboard wedge rotating 10° at a time through 360° with a mean within-participant error of  $18.0 \pm 17.4^\circ$  across all sessions in our highest performing participant (Figure 6B) and  $25.8 \pm 24.7^\circ$  across all participants. While this decoding performance is significantly better than chance (90°,  $p < 0.0033$ ), the error indicates that the decoder cannot independently resolve all of the 36 positions separated by 10°. A better estimate of the number of independent radial positions that could be decoded is the full cycle (360°) divided by the mean error, or approximately 20 positions in our highest performing participant and fewer for other participants (~14 positions on average).

We observed several intriguing patterns in our evaluation of 36-way decoding across multiple participants (Figure 7). Decoding performance within a participant appears to be consistent across runs and sessions, and is best in participant 1 and worst in participant 2. We interpret this as reflecting that the repeatability of precise retinotopic activations is greatest in participant 1 and lowest in participant 2, perhaps as a result of differences in participant compliance with central fixation and differences in data quality. Another possible contributing factor is anatomical variability between participants in the precise folding of visual cortex. For example, in some participants, responses to stimuli in all four quadrants of the visual field can be mapped by HD-DOT, but in others only responses to the lower visual field hemi-field can be captured with high signal-to-noise (White and Culver, 2010b). The fidelity of activations appears to be even more critical for test data than training data based on the asymmetries observed in the matrix. For instance, testing with participant 1 after training with participant 2 yields better results than both testing with participant 2 after training with participant 1 as well as using both training and test data from participant 2. These observations are consistent with the fact that template maps are created by averaging training data across blocks, which boosts signal-to-noise ratio, while every time point is evaluated separately in the test data, which makes consistency particularly important. However, additional factors evidently influence inter-participant decoding; for instance, decoding of test data from participant 1 is better using templates from participant 2 (who had the worst within-participant decoding performance) than using templates from participant 3. It is likely that inter-individual differences in anatomy and cap positioning contribute to such trends. For example, participants 1 and 2 may have had more similar head shapes, occipital cortex anatomy, and cap fits, such that retinotopic activations were more similar between them than to those seen in participant 3, potentially explaining why interparticipant decoding was more successful with participants 1 and 2.

In addition, the visual stimulation paradigms used allowed us to study how decoding performance varied across the visual field. It emerged that there was a discrepancy between decoding accuracy in the lower and upper hemifields with generally lower decoding error for stimuli in the lower hemifield (Figure 6C). This result is consistent with prior reports of variations in HD-DOT signal-to-noise ratio across different parts of visual cortex (White and Culver, 2010b). These observations likely stem from the anatomy of visual areas V1, V2, and V3, organized retinotopically around the calcarine fissure with the upper hemifield represented deeper below the cranial surface and hence less accessible to photonic measurements (DeYoe et al., 1996; Wandell et al., 2007). Structural and functional neuroanatomical variability between individuals likely explains why this effect of stimulus

position on decoding error was more pronounced in some individuals than others (Amunts et al., 2000; Dougherty et al., 2003; Dumoulin and Wandell, 2008; Iaria and Petrides, 2007; Kochunov et al., 2003; Thompson et al., 1996).

Nonetheless, the feasibility of this detailed decoding even with training and test data from different cap fits and individuals reflects the potential of HD-DOT for supporting applications of neural decoding. Furthermore, the success of this spatially detailed decoding encourages exploring other forms of more elaborate decoding, such as deciphering naturalistic stimulus information. In fact, some of the error recorded in the current decoding experiments may have stemmed from lapses in compliance with central fixation (as it is challenging to keep gaze fixated on a central crosshair while bright, distracting patterns flicker in the periphery), so a more naturalistic visual stimulus may even improve decoding performance. Decoding of more complex stimuli could also leverage signals from additional areas beyond early visual cortex to potentially distinguish a larger number of targets.

Aside from the increased spatial detail of decoding revealed by the 18- and 36-way classification experiments, the success of decoding stimulus location at individual time points without having to average across multiple trials is also a step towards developing the real-time decoding that would enable efficient brain-computer interfaces for clinical use. The current study aimed to evaluate the feasibility of decoding while using established methods for processing and reconstructing optical data, including steps that use the full run of data such as zero-phase high-pass filtering, superficial signal regression, and mean-subtraction. Future studies could reassess the performance of single-trial and single-time-point decoding using a modified processing pipeline that changes these steps to respect causal relations, only using prior data for any given time point. Such analysis would provide an assessment of pseudo-real-time decoding. Based on the accuracy and timing of this proposed pseudo-real-time decoding, subsequent studies could decode HD-DOT data in real time while participants are being imaged and deliver feedback to investigate closed-loop HD-DOT-based brain-computer interfaces.

#### 4.5. Limitations, potential solutions, and future directions

It is important to note that our study intentionally enriched for participants and data sets with high raw data quality, as evaluated by light levels, signal-to-noise, pulse signal, and head motion. Our rationale for this decision was multifactorial. Firstly, as we were evaluating a new method of analyzing HD-DOT data, it was important to ensure that the quality of the data itself was not a confounding factor undermining the study. Secondly, the myriad possible sources of noise (instrumentation, cap fit, head motion, participant attention, confounding physiology, etc.) render a reasonable analysis of these factors and their effects on decoding beyond the scope of the current study. Furthermore, data quality is a moving target, with newer systems and strategies mitigating the effects of head motion and improving signal-to-noise (Chitnis et al., 2016; Sherafati et al., 2020; Zhao and Cooper, 2017). Data quality is also expected to vary dramatically as decoding is applied to different tasks, domains, and populations. Therefore, additional research will be required to evaluate generalizability across a broader population and range of data quality. Future studies with larger numbers of participants would also be better powered to more systematically evaluate

the relationship between different components or measures of data quality and decoding performance. Overall, we anticipate average decoding performance might be poorer among a general population. However, on the upside, most of the potential sources of noise discussed are addressable. For instance, anatomical variability could be at least partly addressed through subject-specific head modeling (Eggebrecht et al., 2012). Participant compliance could be improved through training and feedback for both maintaining central fixation (Guzman-Martinez et al., 2009) and minimizing head motion (Greene et al., 2018; Yang et al., 2005). And given the fluid nature of data quality and the progression of the fNIRS research field towards ensuring higher quality data, it may be possible to further improve decoding performance through advances in hardware such as wireless systems with high signal-to-noise (Chitnis et al., 2016; Zhao and Cooper, 2017) and algorithms for rigorous monitoring and optimization of data quality (Sherafati et al., 2020). As a result, the field of optical neuroimaging will likely see improvement in the performance and scope of decoding in future research.

The current study was restricted to visual decoding, but it also provides a framework to begin exploring decoding of other modalities with HD-DOT. For example, other sensory systems and the motor system also organize information systematically in neuroanatomical space (Besle et al., 2019; Penfield and Boldrey, 1937), and these maps could be leveraged to decode motor imagery or auditory signals with HD-DOT to replicate and expand on prior work with fMRI and fNIRS (Correia et al., 2015; Hoefle et al., 2018; Sitaram et al., 2007). In these applications too, HD-DOT could combine the strengths of fNIRS, such as portability and a quiet scanning environment, with higher spatial resolution and image quality to potentially perform the more detailed decoding that fMRI has supported.

It is left for future studies to determine if the decoding established here using a simple task paradigm and normative adult population can be extended to more challenging tasks and populations, as have been explored with other imaging modalities. In particular, decoding naturalistic stimuli outside the decoder's training set, as previous fMRI studies have done, will require more sophisticated decoding strategies than the straightforward template matching approach used here. Future HD-DOT studies may hence explore decoding using stimulus feature encoding regression models or convolutional neural networks for classification (Kay et al., 2008; Nishimoto et al., 2011; Wen et al., 2018). Decoding more complex tasks will also rely on capturing more subtle signals distributed across broader areas of cortex than the robust signal localized to visual cortex that is evoked by checkerboard stimuli. However, HD-DOT has been shown to be capable of functional brain mapping with high signal-to-noise ratio during tasks such as verb generation, covert reading, and viewing multimodal naturalistic stimuli (Eggebrecht et al., 2014; Fishell et al., 2019), which is encouraging for decoding covert and complex signals pertaining to language and other domains beyond vision. Decoding in other populations, such as infants and patients with neurological disorders who have participated in prior fNIRS decoding studies (Abdalmalak et al., 2017; Emberson et al., 2017), will present additional data acquisition challenges such as increased levels of motion and the need for imaging at the bedside. However, previous HD-DOT brain mapping studies have already established the feasibility of imaging at the bedside in neonates and stroke patients (Culver et al., 2016; Ferradal et al., 2016; Liao et al., 2012), while newer lightweight (Bergonzi et al., 2018) or fiber-less (Zhao

and Cooper, 2017) designs will further increase portability and reduce motion artifacts facilitating decoding studies in such populations.

Finally, potential real-world applications of HD-DOT decoding such as brain-computer interfaces in neurological populations would build on the combination of advances in imaging hardware, real-time data processing methods, and decoding algorithms. Although motor prosthetic control and augmented communication commonly garner more attention, one of the earliest demonstrations of a brain-computer interface was based on decoding EEG responses to checkerboard visual stimuli (Vidal, 1977). In that study, participants guided a cursor through a maze by shifting their center of fixation relative to a checkerboard stimulus, which produced visual evoked potentials that were decoded and used to update the location of the cursor on the screen. A similar paradigm could be used to apply our retinotopic decoding in a visual HD-DOT-based brain-computer interface in healthy participants for proof of principle. This framework for real-time processing and feedback could then be combined with other parallel progress in HD-DOT decoding, such as motor and semantic decoding, and with hardware advances, such as increasingly wearable HD-DOT systems, to develop HD-DOT-driven prosthetics and communication for patients with motor disabilities.

There are indeed many steps between our study of explicit visual decoding and long-term goals such as brain-computer interfaces for neurological patients. However, the current validation of the feasibility, accuracy, and reproducibility of detailed, single-trial visual decoding with HD-DOT provides a solid foundation to build upon in future research.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

This work was funded by the National Institutes of Health [grant numbers U01EB027005, R01NS090874] awarded to J. P. C. and a McDonnell Foundation fellowship for cognitive, computational, and systems neuroscience awarded to K.T. The funding sources had no involvement in study design, data collection or analysis, writing, or the decision to submit this article for publication. In addition, the authors thank Lindsey Brier, Dillan Newbold, and Pooja Tripathy for helpful discussions regarding the study, as well as all study participants for generously lending their time to this research.

## Abbreviations:

|               |   |
|---------------|---|
| <b>HD-DOT</b> | high-density diffuse optical tomography |
| <b>fNIRS</b>  | functional near-infrared spectroscopy   |
| <b>fMRI</b>   | functional magnetic resonance imaging   |
| <b>ECoG</b>   | electrocorticography                    |
| <b>ROC</b>    | receiver operating characteristic       |
| <b>AUC</b>    | area under the curve                    |

## References

- Abdalmalak A, Milej D, Norton L, Debicki DB, Gofton T, Diop M, Owen AM, Lawrence K St., 2017. Single-session communication with a locked-in patient by functional near-infrared spectroscopy. *Neurophotonics* 4, 040501. 10.1117/1.nph.4.4.040501 [PubMed: 29296627]
- Abdalmalak A, Milej D, Yip LCM, Khan AR, Diop M, Owen AM, Lawrence K St., 2020. Assessing Time-Resolved fNIRS for Brain-Computer Interface Applications of Mental Communication. *Front. Neurosci* 14, 105. 10.3389/fnins.2020.00105 [PubMed: 32132894]
- Albers AM, Kok P, Toni I, Dijkerman HC, de Lange FP, 2013. Shared Representations for Working Memory and Mental Imagery in Early Visual Cortex. *Curr. Biol* 23, 1427–1431. 10.1016/j.cub.2013.05.065 [PubMed: 23871239]
- Amunts K, Malikovic A, Mohlberg H, Schormann T, Zilles K, 2000. Brodmann's Areas 17 and 18 Brought into Stereotaxic Space- Where and How Variable? *Neuroimage* 11,66–84. 10.1006/nimg.1999.0516 [PubMed: 10686118]
- Anumanchipalli GK, Chartier J, Chang EF, 2019. Speech synthesis from neural decoding of spoken sentences. *Nature* 568, 493–498. 10.1038/s41586-019-1119-1 [PubMed: 31019317]
- Bergonzi KM, Burns-Yocum TM, Bumstead JR, Buckley EM, Mannion PC, Tracy CH, Mennerick E, Ferradal SL, Deghani H, Eggebrecht AT, Culver JP, 2018. Lightweight sCMOS-based high-density diffuse optical tomography. *Neurophotonics* 5, 035006. 10.1117/1.NPh.5.3.035006 [PubMed: 30137925]
- Besle J, Mougín O, Sánchez-Panchuelo R-M, Lanting C, Gowland P, Bowtell R, Francis S, Krumbholz K, 2019. Is Human Auditory Cortex Organization Compatible With the Monkey Model? Contrary Evidence From Ultra-High-Field Functional and Structural MRI. *Cereb. Cortex* 29, 410–428. 10.1093/cercor/bhy267 [PubMed: 30357410]
- Bevilacqua F, Pigué D, Marquet P, Gross JD, Tromberg BJ, Depeursinge C, 1999. In vivo local determination of tissue optical properties: applications to human brain. *Appl. Opt* 38, 4939. 10.1364/ao.38.004939 [PubMed: 18323984]
- Boynton GM, Engel SA, Glover GH, Heeger DJ, 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. *J. Neurosci* 16, 4207–4221. 10.1523/jneurosci.16-13-042076.1996 [PubMed: 8753882]
- Brainard DH, 1997. The Psychophysics Toolbox. *Spat. Vis* 10, 433–436. [PubMed: 9176952]
- Brandman DM, Cash SS, Hochberg LR, 2018. Review: Human intracortical recording and neural decoding for brain computer interfaces. *IEEE Trans Neural Syst Rehabil Eng.* 25, 1687–1696. 10.1109/TNSRE.
- Chitnis D, Cooper RJ, Dempsey L, Powell S, Quaggia S, Highton D, Elwell C, Hebden JC, Everdell NL, 2016. Functional imaging of the human brain using a modular, fibre-less, high-density diffuse optical tomography system. *Biomed. Opt. Express* 7, 4275. 10.1364/boe.7.004275 [PubMed: 27867731]
- Collinger JL, Wodlinger B, Downey JE, Wang W, Tyler-kabara EC, Weber DJ, Mcmorland AJC, Velliste M, Boninger ML, Schwartz AB, 2013. High-performance neuroprosthetic control by an individual with tetraplegia. *Lancet* 381, 557–564. 10.1016/S0140-6736(12)61816-9 [PubMed: 23253623]
- Correia J, Formisano E, Valente G, Hausfeld L, Jansma B, Bonte M, 2014. Brain-Based Translation: fMRI Decoding of Spoken Words in Bilinguals Reveals Language-Independent Semantic Representations in Anterior Temporal Lobe. *J Neurosci* 34, 332–338. 10.1523/JNEUROSCI.1302-13.2014 [PubMed: 24381294]
- Correia JM, Jansma BMB, Bonte M, 2015. Decoding Articulatory Features from fMRI Responses in Dorsal Speech Regions. *J Neurosci* 35, 15015–15025. 10.1523/JNEUROSCI.0977-15.2015 [PubMed: 26558773]
- Culver JP, Bergonzi KM, Eggebrecht AE, Fishell AK, Lee J, 2016. Optical imaging of functional connectivity at the bedside. *Conf Proc IEEE Eng Med Biol Soc* 65–67. 10.1109/EMBC.2016.7590641

- Custo A, Wells WM, Barnett AH, Hillman EMC, Boas DA, 2006. Effective scattering coefficient of the cerebral spinal fluid in adult head models for diffuse optical imaging. *Appl. Opt* 45, 4747–4755. 10.1364/AO.45.004747 [PubMed: 16799690]
- Dale AM, Fischl B, Sereno MI, 1999. Cortical Surface-Based Analysis. *Neuroimage* 9, 179–194. 10.1006/nimg.1998.0395 [PubMed: 9931268]
- Dehghani H, Eames ME, Yalavarthy PK, Davis SC, Srinivasan S, Carpenter CM, Pogue BW, Paulsen KD, 2008. Near infrared optical tomography using NIRFAST: Algorithm for numerical model and image Reconstruction. *Commun. Numer. Methods Eng* 25, 711–732. 10.1002/cnm.1162 [PubMed: 20182646]
- Dehghani H, Pogue BW, Poplack SP, Paulsen KD, 2003. Multiwavelength three-dimensional near-infrared tomography of the breast: initial simulation, phantom, and clinical results. *Appl. Opt.* 42, 135–45. 10.1364/ao.42.000135 [PubMed: 12518832]
- DeYoe EA, Carman GJ, Bandettini P, Glickman S, Wieser J, Cox R, Miller D, Neitz J, 1996. Mapping striate and extrastriate visual areas in human cerebral cortex. *Proc. Natl. Acad. Sci. U. S. A* 93, 2382–2386. 10.1073/pnas.93.6.2382 [PubMed: 8637882]
- Dougherty RF, Koch VM, Brewer AA, Fischer B, Modersitzki J, Wandell BA, 2003. Visual field representations and locations of visual areas V1/2/3 in human visual cortex. *J. Vis* 3, 586–598. 10.1167/3.10.1 [PubMed: 14640882]
- Dumoulin SO, Wandell BA, 2008. Population receptive field estimates in human visual cortex. *Neuroimage* 39, 647–660. 10.1016/j.neuroimage.2007.09.034 [PubMed: 17977024]
- Eggebrecht AT, Ferradal SL, Robichaux-Viehoever A, Mahlega SS, Dehghani H, Snyder AZ, Hershey T, Culver JP, 2014. Mapping distributed brain function and networks with diffuse optical tomography. *Nat. Photonics* 8, 448–454. 10.1038/nphoton.2014.107 [PubMed: 25083161]
- Eggebrecht AT, White BR, Ferradal SL, Chen C, Zhan Y, Snyder AZ, Dehghani H, Culver JP, 2012. A quantitative spatial comparison of high-density diffuse optical tomography and fMRI cortical mapping. *Neuroimage* 61, 1120–1128. 10.1016/j.neuroimage.2012.01.124 [PubMed: 22330315]
- Emberson LL, Zinszer BD, Raizada RDS, Aslin RN, 2017. Decoding the infant mind: Multivariate pattern analysis (MVPA) using fNIRS. *PLoS One* 12, e0172500. 10.1371/journal.pone.0172500 [PubMed: 28426802]
- Ferradal SL, Eggebrecht AT, Hassanpour M, Snyder AZ, Culver JP, 2014. Atlas-based head modeling and spatial normalization for high-density diffuse optical tomography: In vivo validation against fMRI. *Neuroimage* 85, 117–126. 10.1016/j.neuroimage.2013.03.069 [PubMed: 23578579]
- Ferradal SL, Liao SM, Eggebrecht AT, Shimony JS, Inder TE, Culver JP, Smyser CD, 2016. Functional Imaging of the Developing Brain at the Bedside Using Diffuse Optical Tomography. *Cereb. Cortex* 26, 1558–1568. 10.1093/cercor/bhu320 [PubMed: 25595183]
- Fischl B, Liu A, Dale AM, 2001. Automated Manifold Surgery: Constructing Geometrically Accurate and Topologically Correct Models of the Human Cerebral Cortex. *IEEE Trans. Med. Imaging* 20, 70–80. 10.1109/42.906426 [PubMed: 11293693]
- Fishell AK, Arbeláez AM, Valdés CP, Burns-Yocum TM, Sherafati A, Richter EJ, Torres M, Eggebrecht AT, Smyser CD, Culver JP, 2020. Portable, field-based neuroimaging using high-density diffuse optical tomography. *Neuroimage* 215, 116541. 10.1016/j.neuroimage.2020.116541 [PubMed: 31987995]
- Fishell AK, Burns-Yocum TM, Bergonzi KM, Eggebrecht AT, Culver JP, 2019. Mapping brain function during naturalistic viewing using high-density diffuse optical tomography. *Sci. Rep* 9, 11115. 10.1038/s41598-019-45555-8 [PubMed: 31366956]
- Fonov V, Evans AC, Botteron K, Almlí CR, McKinstry RC, Collins DL, 2011. Unbiased average age-appropriate atlases for pediatric studies. *Neuroimage* 54, 313–327. 10.1016/j.neuroimage.2010.07.033 [PubMed: 20656036]
- Gordon EM, Laumann TO, Gilmore AW, Newbold DJ, Greene DJ, Berg JJ, Ortega M, Hoyt-Draze C, Gratton C, Sun H, Hampton JM, Coalson RS, Nguyen AL, McDermott KB, Shimony JS, Snyder AZ, Schlaggar BL, Petersen SE, Nelson SM, Dosenbach NUF, 2017. Precision functional mapping of individual human brains. *Neuron* 95, 791–807. 10.1016/j.neuron.2017.07.011 [PubMed: 28757305]

- Greene DJ, Koller JM, Hampton JM, Wesevich V, Van AN, Nguyen AL, Hoyt CR, McIntyre L, Earl EA, Klein RL, Shimony JS, Petersen SE, Schlaggar BL, Fair DA, Dosenbach NUF, 2018. Behavioral interventions for reducing head motion during MRI scans in children. *Neuroimage* 171, 234–245. 10.1016/j.neuroimage.2018.01.023 [PubMed: 29337280]
- Gregg NM, White BR, Zeff BW, Berger AJ, Culver JP, 2010. Brain specificity of diffuse optical imaging: improvements from superficial signal regression and tomography. *Front. Neuroenergetics* 2, 14. 10.3389/fnene.2010.00014 [PubMed: 20725524]
- Guzman-Martinez E, Leung P, Franconeri S, Grabowecky M, Suzuki S, 2009. Rapid eye-fixation training without eye tracking. *Psychon Bull Rev* 16, 491–496. 10.3758/PBR.16.3.491 [PubMed: 19451374]
- Harrison SA, Tong F, 2009. Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458, 632–635. 10.1038/nature07832 [PubMed: 19225460]
- Hassanpour MS, Eggebrecht AT, Culver JP, Peelle JE, 2015. Mapping cortical responses to speech using high-density diffuse optical tomography. *Neuroimage* 117, 319–326. 10.1016/j.neuroimage.2015.05.058 [PubMed: 26026816]
- Haynes J-D, Rees G, 2006. Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci* 7, 523–534. 10.1038/nrn1931 [PubMed: 16791142]
- Hebart MN, Baker CI, 2018. Deconstructing multivariate decoding for the study of brain function. *Neuroimage* 180, 4–18. 10.1016/j.neuroimage.2017.08.005 [PubMed: 28782682]
- Hoefle S, Engel A, Basilio R, Alluri V, Toiviainen P, Cagy M, Moll J, 2018. Identifying musical pieces from fMRI data using encoding and decoding models. *Sci. Rep* 8, 2266. 10.1038/S41598-018-20732-3 [PubMed: 29396524]
- Horikawa T, Kamitani Y, 2017. Generic decoding of seen and imagined objects using hierarchical visual features. *Nat. Commun* 8, 15037. 10.1038/ncomms15037 [PubMed: 28530228]
- Horikawa T, Tamaki M, Miyawaki Y, Kamitani Y, 2013. Neural Decoding of Visual Imagery During Sleep. *Science* (80-.). 340, 639–643. 10.1126/science.1234330
- Hosseini SMH, Mano Y, Rostami M, Takahashi M, Sugiara M, Kawashima R, 2011. Decoding what one likes or dislikes from single-trial fNIRS measurements. *Neuroreport* 22, 269–273. 10.1097/WNR.0b013e3283451f8f [PubMed: 21372746]
- Hubel DH, Wiesel TN, 1965. Binocular interaction reared in striate artificial cortex squint. *J. Neurophysiol* 28, 1041–1059. 10.1152/jn.1965.28.6.1041 [PubMed: 5883731]
- Hubel DH, Wiesel TN, 1962. Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex. *J Physiol* 160, 106–154. 10.1113/jphysiol.1962.sp006837 [PubMed: 14449617]
- Huth AG, Lee T, Nishimoto S, Bilenko NY, Vu AT, Gallant JL, 2016. Decoding the Semantic Content of Natural Movies from Human Brain Activity. *Front. Syst. Neurosci* 10, 81. 10.3389/fnsys.2016.00081 [PubMed: 27781035]
- Iaria G, Petrides M, 2007. Occipital Sulci of the Human Brain: Variability and Probability Maps. *J Comp Neurol* 501,243–259. 10.1002/cne [PubMed: 17226764]
- Kamitani Y, Tong F, 2006. Decoding Seen and Attended Motion Directions from Activity in the Human Visual Cortex. *Curr. Biol* 16, 1096–1102. 10.1016/j.cub.2006.04.003 [PubMed: 16753563]
- Kamitani Y, Tong F, 2005. Decoding the visual and subjective contents of the human brain. *Nat. Neurosci* 8, 679–685. 10.1038/nn1444 [PubMed: 15852014]
- Kay KN, Naselaris T, Prenger RJ, Gallant JL, 2008. Identifying natural images from human brain activity. *Nature* 452, 352–5. 10.1038/nature06713 [PubMed: 18322462]
- Kochunov P, Hasnain M, Lancaster J, Grabowski T, Fox P, 2003. Improvement in Variability of the Horizontal Meridian of the Primary Visual Area Following High-Resolution Spatial Normalization. *Hum. Brain Mapp* 18, 123–134. 10.1002/hbm.10080 [PubMed: 12518292]
- Laumann TO, Gordon EM, Adeyemo B, Snyder AZ, Joo SJ, Chen M-Y, Gilmore AW, Mcdermott KB, Nelson SM, Dosenbach NUF, Schlaggar BL, Mumford JA, Poldrack RA, Petersen SE, 2015. Functional system and areal organization of a highly sampled individual human brain. *Neuron* 87, 657–670. 10.1016/j.neuron.2015.06.037 [PubMed: 26212711]

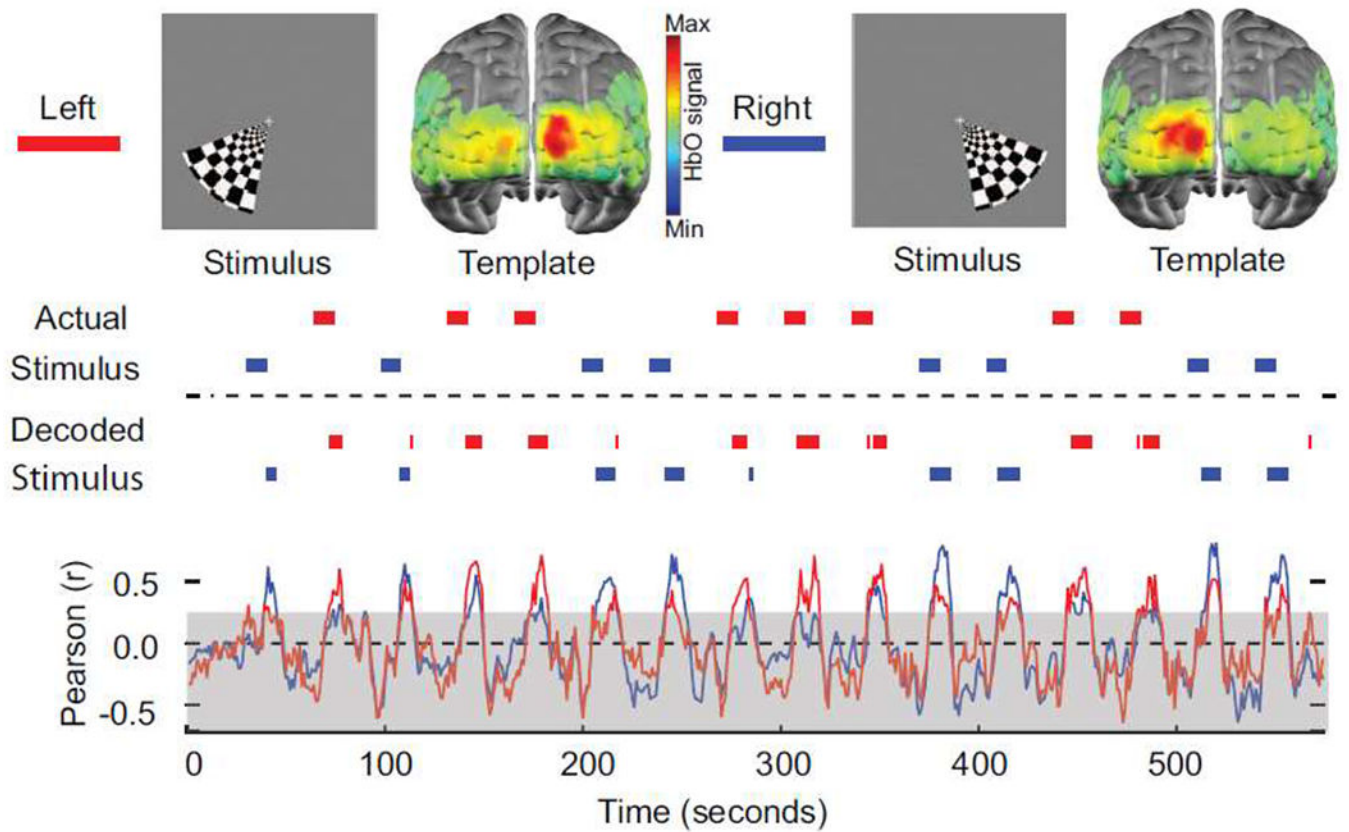


- Liao SM, Ferradal SL, White BR, Gregg N, Inder TE, Culver JP, 2012. High-density diffuse optical tomography of term infant visual cortex in the nursery. *J. Biomed. Opt* 17, 081414. 10.1117/1.JBO.17.8.081414 [PubMed: 23224175]
- Lindquist MA, Meng Loh J, Atlas LY, Wager TD, 2009. Modeling the hemodynamic response function in fMRI: efficiency, bias and mis-modeling. *Neuroimage* 45, S187–S198. 10.1016/j.neuroimage.2008.10.065 [PubMed: 19084070]
- Luu S, Chau T, 2009. Decoding subjective preference from single-trial near-infrared spectroscopy signals. *J Neural Eng* 6, 016003. 10.1088/1741-2560/6/1/016003 [PubMed: 19104138]
- Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, Woods R, Paus T, Simpson G, Pike B, Holmes C, Collins L, Thompson P, MacDonald D, Iacoboni M, Schormann T, Amunts K, Palomero-Gallagher N, Geyer S, Parsons L, Narr K, Kabani N, Le Goualher G, Boomsma D, Cannon T, Kawashima R, Mazoyer B, 2001. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philos. Trans. R. Soc. B Biol. Sci* 356, 1293–1322. 10.1098/rstb.2001.0915
- Metz CE, 2006. Receiver Operating Characteristic Analysis: A Tool for the Quantitative Evaluation of Observer Performance and Imaging Systems. *J Am Coll Radiol* 3, 413–422. 10.1016/j.jacr.2006.02.021 [PubMed: 17412096]
- Metz CE, 1978. Basic Principles of ROC Analysis. *Semin. Nucl. Med* 8, 283–98. 10.1016/S0001-2998(78)80014-2 [PubMed: 112681]
- Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu B, Gallant JL, 2011. Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies. *Curr. Biol* 21, 1641–1646. 10.1016/j.cub.2011.08.031 [PubMed: 21945275]
- Penfield W, Boldrey E, 1937. Somatic Motor And Sensory Representation In The Cerebral Cortex Of Man As Studied By Electrical Stimulation. *Brain* 60, 389–443. 10.1093/brain/60.4.389
- Pinti P, Tachtsidis I, Hamilton A, Hirsch J, Aichelburg C, Gilbert S, Burgess PW, 2020. The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Ann N Y Acad Sci*. 1464, 5–29. 10.1111/nyas.13948 [PubMed: 30085354]
- Ségonne F, Dale AM, Busa E, Glessner M, Salat D, Hahn HK, Fischl B, 2004. A hybrid approach to the skull stripping problem in MRI. *Neuroimage* 22, 1060–1075. 10.1016/j.neuroimage.2004.03.032 [PubMed: 15219578]
- Sherafati A, Snyder AZ, Eggebrecht AT, Bergonzi KM, Burns-Yocum TM, Lugar HM, Ferradal SL, Robichaux-Viehoever A, Smyser CD, Palanca BJ, Hershey T, Culver JP, 2020. Global motion detection and censoring in high-density diffuse optical tomography. *Hum. Brain Mapp* 4093–4112. 10.1002/hbm.25111 [PubMed: 32648643]
- Sitaram R, Zhang H, Guan C, Thulasidas M, Hoshi Y, Ishikawa A, Shimizu K, Birbaumer N, 2007. Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain–computer interface. *Neuroimage* 34, 1416–1427. 10.1016/j.neuroimage.2006.11.005 [PubMed: 17196832]
- Strangman G, Culver JP, Thompson JH, Boas DA, 2002. A Quantitative Comparison of Simultaneous BOLD fMRI and NIRS Recordings during Functional Brain Activation. *Neuroimage* 17, 719–731. 10.1006/nimg.2002.1227 [PubMed: 12377147]
- Thirion B, Duchesnay E, Hubbard E, Dubois J, Poline J, Lebihan D, Dehaene S, 2006. Inverse retinotopy: Inferring the visual content of images from brain activation patterns. *Neuroimage* 33, 1104–1116. 10.1016/j.neuroimage.2006.06.062 [PubMed: 17029988]
- Thompson PM, Schwartz C, Lin RT, Khan AA, Toga AW, 1996. Three-dimensional Statistical Analysis of Sulcal Variability in the Human Brain. *J Neurosci* 16, 4261–4274. 10.1523/JNEUROSCI.16-13-04261.1996 [PubMed: 8753887]
- Vidal JJ, 1977. Realtime Detection of Brain Events in EEG. *Proc. IEEE* 65, 633–641. 10.1109/PROC.1977.10542
- Wandell BA, Dumoulin SO, Brewer AA, 2007. Visual Field Maps in Human Cortex. *Neuron* 56, 366–383. 10.1016/j.neuron.2007.10.012 [PubMed: 17964252]
- Wen H, Shi J, Zhang Y, Lu K-H, Cao J, Liu Z, 2018. Neural Encoding and Decoding with Deep Learning for Dynamic Natural Vision. *Cereb. Cortex* 28, 4136–4160. 10.1093/cercor/bhx268 [PubMed: 29059288]

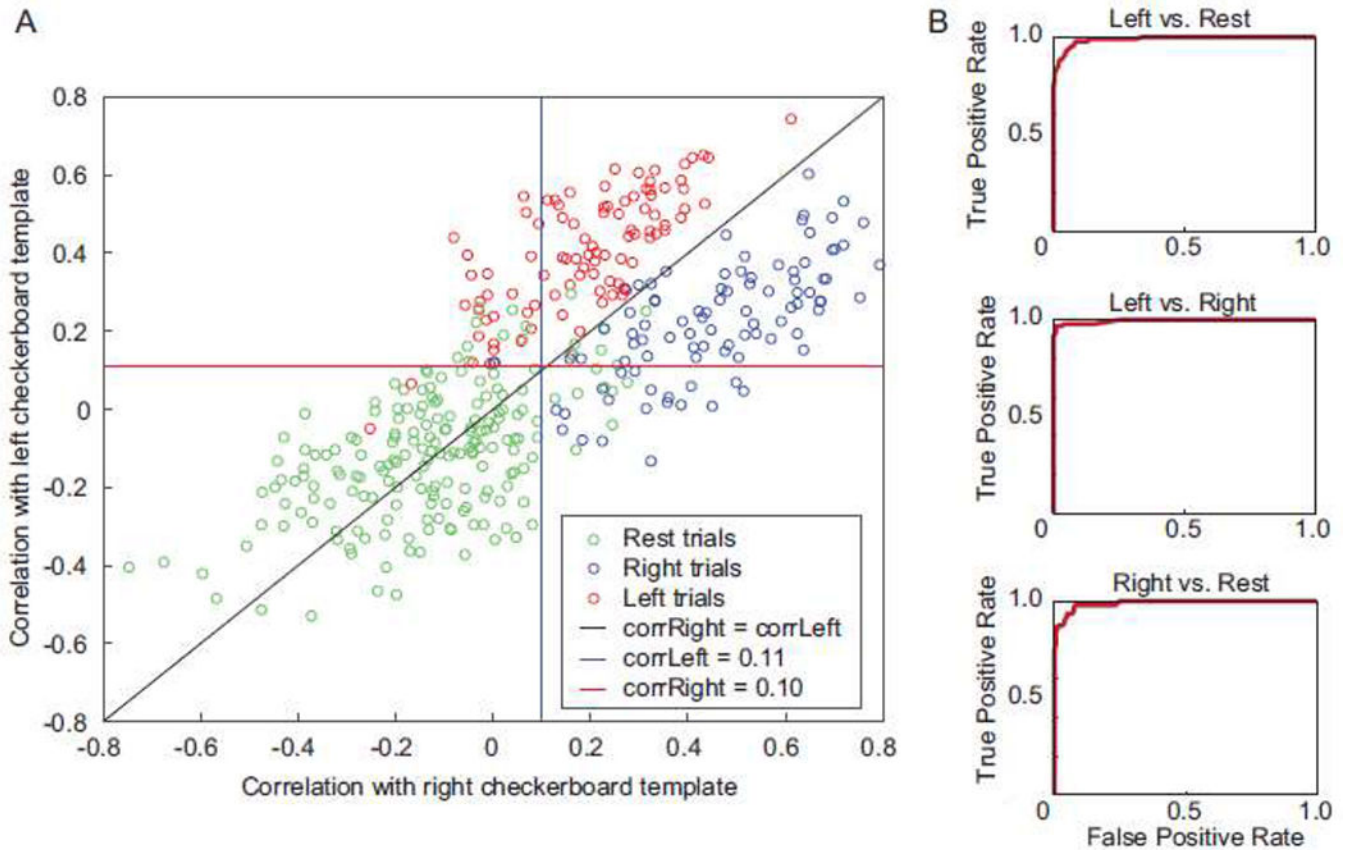
- White BR, Culver JP, 2010a. Quantitative evaluation of high-density diffuse optical tomography: in vivo resolution and mapping performance. *J. Biomed. Opt* 15, 026006. 10.1117/1.3368999 [PubMed: 20459251]
- White BR, Culver JP, 2010b. Phase-encoded retinotopy as an evaluation of diffuse optical neuroimaging. *Neuroimage* 49, 568–577. 10.1016/j.neuroimage.2009.07.023 [PubMed: 19631755]
- Wiesel TN, Hubel DH, 1965a. Comparison of the effects of unilateral and bilateral eye closure on cortical unit responses in kittens. *J Neurophysiol* 28, 1029–1040. 10.1152/jn.1965.28.6.1029 [PubMed: 5883730]
- Wiesel TN, Hubel DH, 1965b. Extent of recovery from the effects of visual deprivation in kittens. *J Neurophysiol* 28, 1060–1072. 10.1152/jn.1965.28.6.1060 [PubMed: 5883732]
- Wiesel TN, Hubel DH, 1963. Single-Cell Responses in Striate Cortex of Kittens Deprived of Vision in One Eye. *J Neurophysiol* 26, 1003–1017. 10.1152/jn.1963.26.6.1003 [PubMed: 14084161]
- Yang S, Ross TJ, Zhang Y, Stein EA, Yang Y, 2005. Head motion suppression using real-time feedback of motion information and its effects on task performance in fMRI. *Neuroimage* 27, 153–162. 10.1016/j.neuroimage.2005.02.050 [PubMed: 16023040]
- Youden WJ, 1950. Index for rating diagnostic tests. *Cancer* 3, 32–35. 10.1002/1097-0142(1950)3:1<32::aid-cnrcr2820030106>3.0.co;2-3 [PubMed: 15405679]
- Zeff BW, White BR, Dehghani H, Schlaggar BL, Culver JP, 2007. Retinotopic mapping of adult human visual cortex with high-density diffuse optical tomography. *Proc Natl Acad Sci U S A* 104, 12169–12174. 10.1073/pnas.0611266104 [PubMed: 17616584]
- Zhao H, Cooper RJ, 2017. Review of recent progress toward a fiberless, whole-scalp diffuse optical tomography system. *Neurophotonics* 5, 011012. 10.1117/1.NPh.5.1.011012 [PubMed: 28983490]

### Highlights

- HD-DOT can support binary visual decoding with high sensitivity and specificity
- HD-DOT can support detailed decoding of moving stimuli at individual time points
- HD-DOT decoding is replicable, although performance can vary across participants
- Inter-participant decoding is also feasible with HD-DOT data

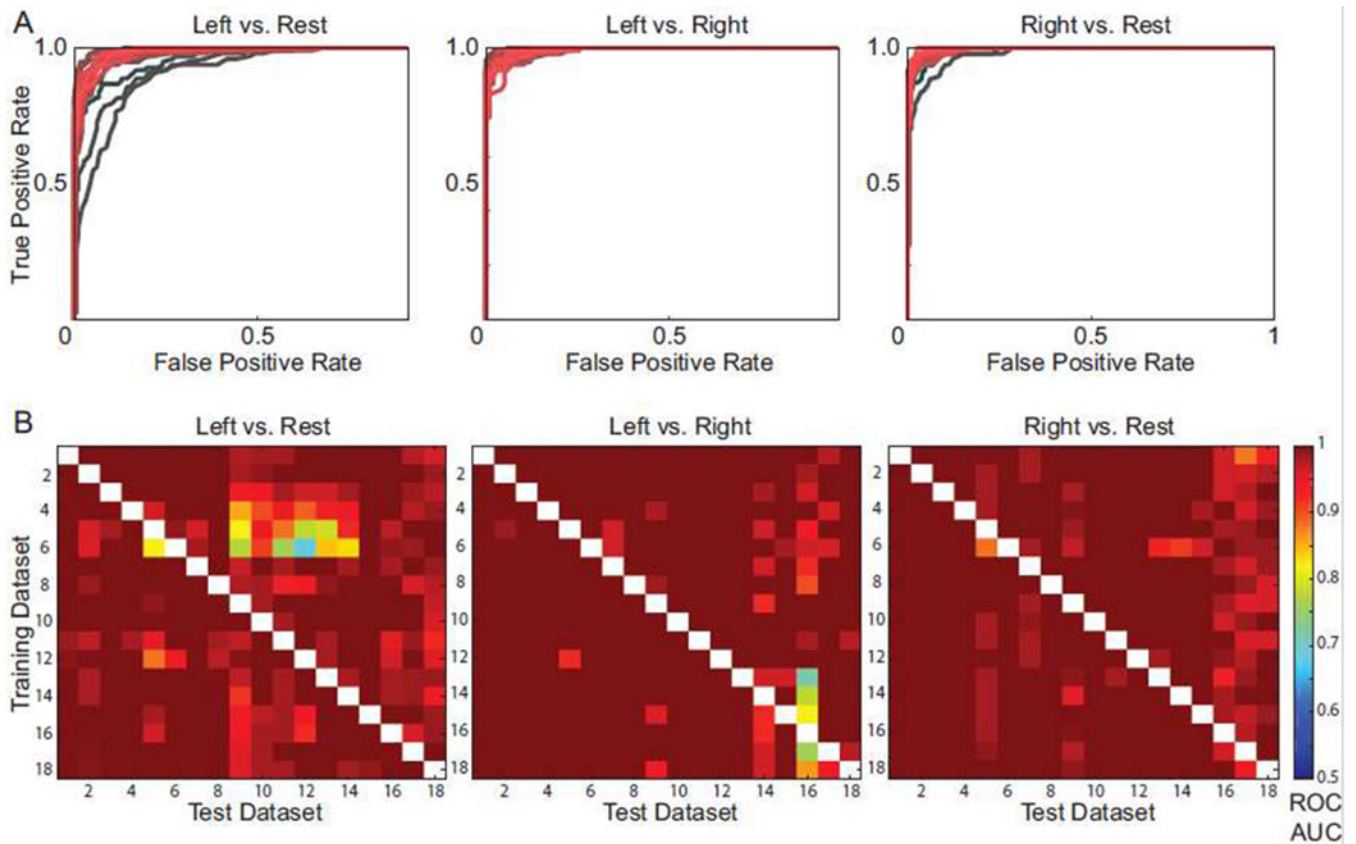


**Figure 1: Feasibility of visual decoding with HD-DOT data using a template matching strategy.** Results are illustrated for a typical decoding attempt in one participant. The participant viewed a checkerboard wedge flickering in either the left or the right visual hemifield, interspersed with rest periods, while being imaged by HD-DOT. Training data were block-averaged to define “templates” of expected brain activity for each stimulus condition (top panels). Spatial Pearson correlation coefficients were then computed between each template and the oxyhemoglobin signal map at each time point in the test data (bottom panel). The resulting correlation values were compared between templates and to a designated threshold value (here selected to be 0.25 on an ad hoc basis, but later optimized by ROC analysis in Figure 2) to determine the decoded stimulus condition at each time point, and this was compared to the actual stimulus state (middle panels).



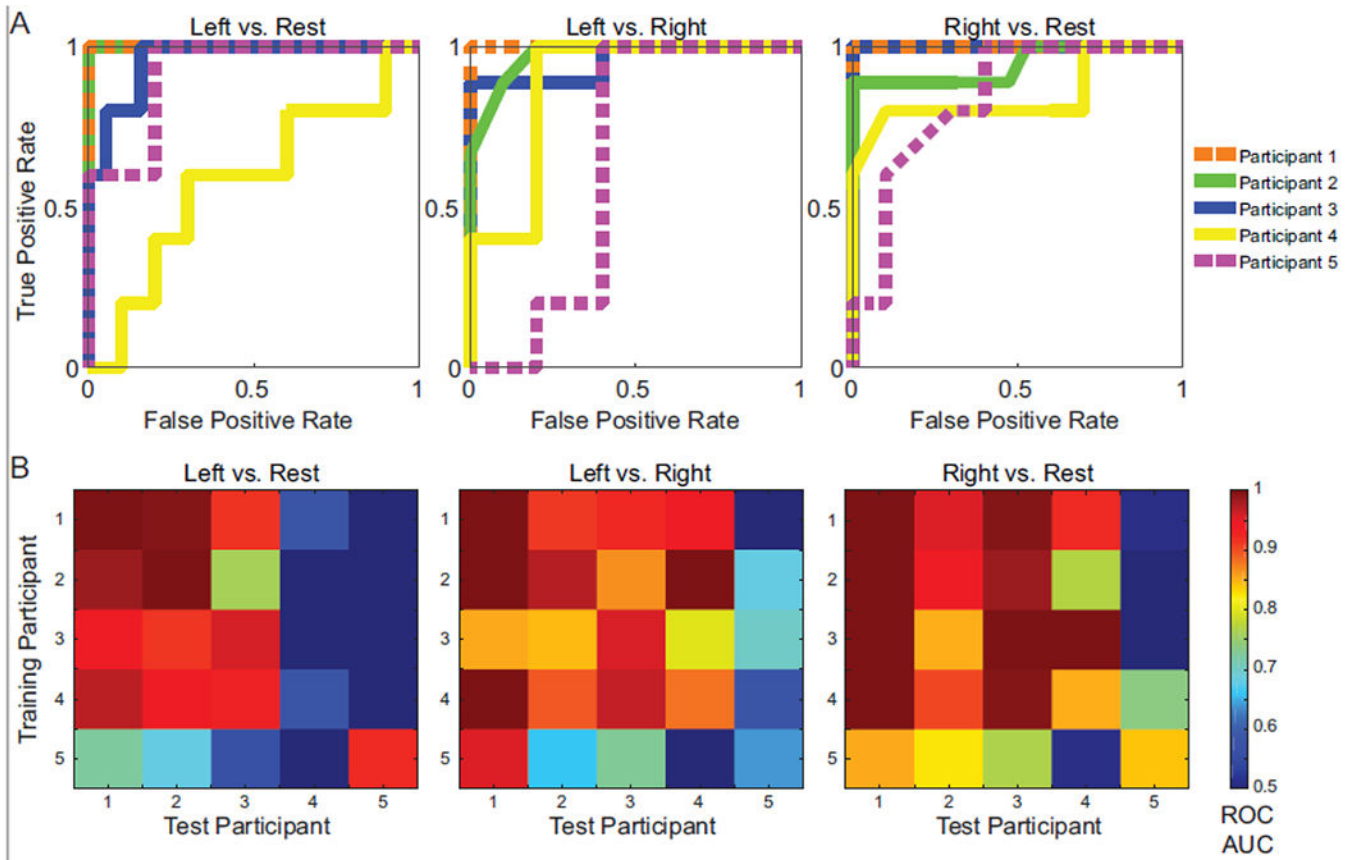
**Figure 2: Sensitivity and specificity of binary retinotopic decoding with HD-DOT data.**

(A) Taking data from 18 task runs in one participant, a single task run was used for construction of templates, and independent test trials were pooled across the remaining 17 runs. Plotting all trials by their correlations with each of the two templates reveals a clustering by trial type. Clusters can be separated by thresholds, which can be swept across the full range of possible values and optimized in a receiver operating characteristic (ROC) analysis. (B) ROC curves for the 3 possible binary classifications.



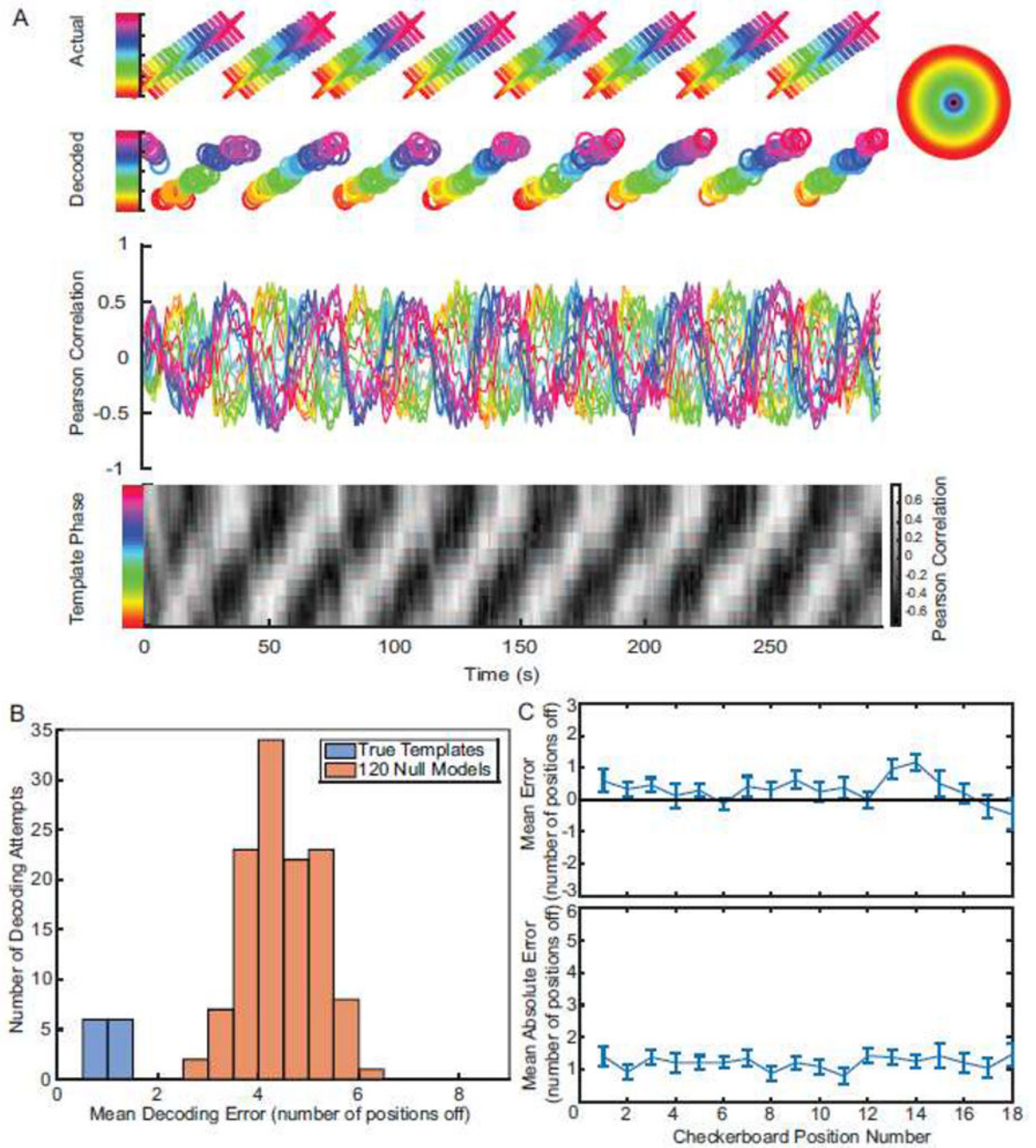
**Figure 3: Reproducibility of binary retinotopic decoding with HD-DOT across sessions in a highly sampled participant.**

(A) The data from the participant who performed the checkerboard viewing task 18 times were used to evaluate the reproducibility of HD-DOT decoding across imaging sessions. The single task run used as training data for template construction was changed 17 times, and each time all the remaining runs were pooled as test data to conduct an ROC analysis, producing 18 ROC curves for each of the three binary classification problems. Different shades between red and black were used to plot different ROC curves to make the individual curves more discernible. (B) ROC analysis was also conducted to evaluate decoding with every possible pairing of a single training task run and a single test task run (i.e., without pooling test data across multiple runs). Areas under the curve (AUCs) for all these ROC analyses are plotted in matrices for each of the three possible pairwise classifications, illustrating the reproducibility of accurate binary retinotopic decoding with HD-DOT.



**Figure 4: Reproducibility of binary HD-DOT decoding across n=5 participants.**

(A) Data was taken from 4 additional participants who performed the same checkerboard stimulus viewing task twice each. Decoding and ROC analysis were conducted for each participant using one task run for template construction and another task run as test data; resulting ROC curves are shown for all 5 participants. (B) Decoding and ROC analysis were also conducted using templates from one task run in one participant and test data from another run in any participant, across every possible pairing of training and test participants. AUC values along the diagonals of these matrices illustrate reproducibility and variability of within-participant decoding across multiple individuals, while off-diagonal values indicate the feasibility of inter-individual decoding.

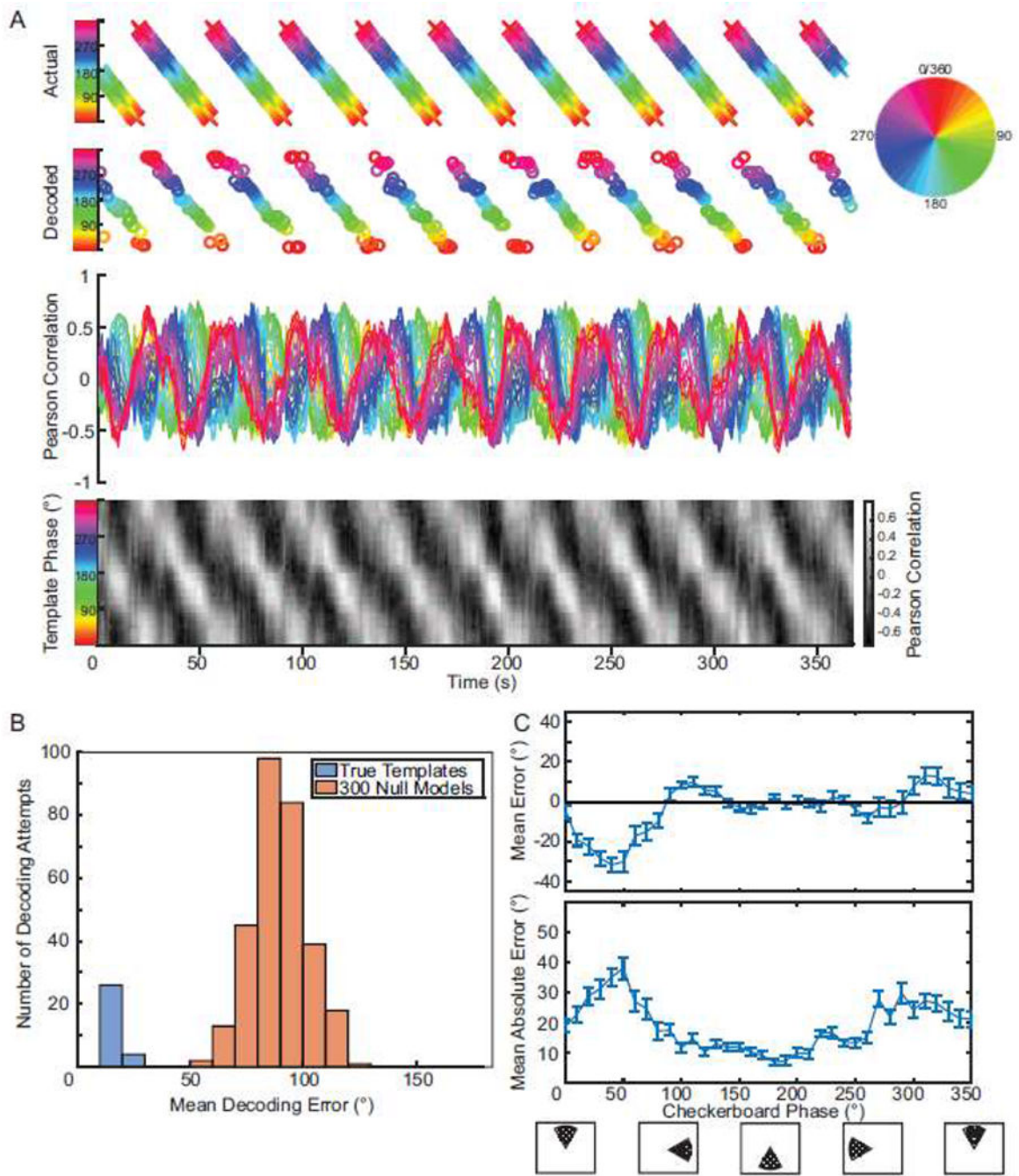


**Figure 5: 18-way classification of visual stimuli with varying eccentricity within a single participant.**

A participant was imaged using HD-DOT while watching a flickering checkerboard ring over 8 cycles of either periodic expansion or contraction through 18 concentric positions on a screen. A template matching strategy was again used to decode stimulus location at each time point in a test dataset, here using 18 template maps – 1 for each stimulus phase. (A) Each stimulus phase is assigned a color (as per the color wheel) and a position along the vertical axis in the plots of actual and decoded stimulus positions. Pearson correlation coefficients were calculated between each of the 18 templates and the oxyhemoglobin signal



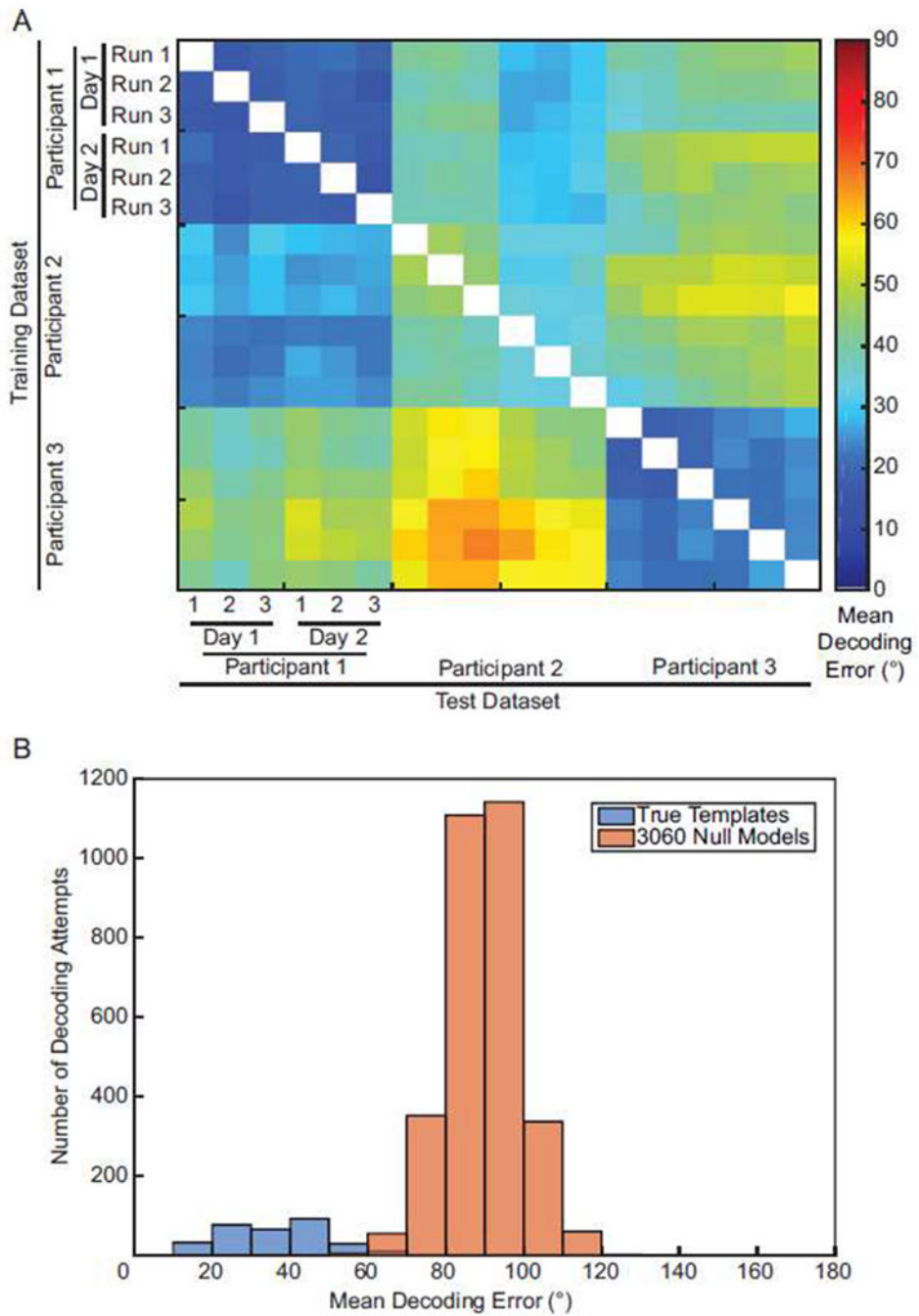
map at every time point in the test data, and are plotted on the two graphs at the bottom of this panel. The decoded stimulus corresponds to the template with the maximum correlation at each time point. (B) A permutation test was performed to evaluate the significance of this decoding using every one of the 12 possible pairings of the 4 task runs collected in this participant for training and testing. The mean decoding error obtained using true template sets was compared with a null distribution generated using 10 random permutations of the training data for every true decoding attempt ( $p < 0.0083$ ). (C) Mean decoding error across the 3 test runs for a single training run is plotted here as a function of true stimulus position, showing little variation in decoding performance with eccentricity.



**Figure 6: 36-way classification of rotating visual stimulus position within a single participant.**

A participant was imaged using HD-DOT while watching a flickering checkerboard wedge rotating 10 times through 36 positions over 36 seconds per revolution. A template matching strategy was used to decode stimulus location, here generating a set of 36 templates (one for each phase of the stimulus) from one training task run and decoding stimulus position at every time point in an independent test task run. (A) Each stimulus phase is assigned a color (as per the color wheel) and a position along the vertical axis in the plots of actual and decoded stimulus positions. Pearson correlation coefficients were calculated between each of

the 36 templates and the oxyhemoglobin signal map at every time point in the test data, and are plotted on the two graphs at the bottom of this panel. The decoded stimulus corresponds to the template with the maximum correlation at each time point. (B) A permutation test was used to evaluate the statistical significance of the decoding performance. Mean decoding error was calculated for every one of the 30 possible pairings of training and test data set across the 6 task runs performed by this participant. The resulting error distribution was compared with a null distribution generated using 10 random permutations of the training data for each true template set ( $p < 0.0033$ ). (C) Mean decoding error across the 5 test runs for a single training run is plotted as a function of true stimulus position, showing better decoding performance as the wedge rotates through the lower half of visual space.



**Figure 7: Reproducibility of complex, non-binary visual decoding with HD-DOT across n=3 participants.**

Three participants viewed the same rotating checkerboard stimulus paradigm six times across two imaging sessions each. Decoding performance was then evaluated using every possible pairing of template and test task run across participants. (A) The mean absolute error for each decoding attempt illustrates both the reproducibility and variability of within-subject and inter-individual decoding across the imaging sessions and participants. (B)

Permutation testing comparing decoding performance across all sessions and participants to a null distribution.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 1:**

Participant demographics and contributions to study

| Participant # | Age | Sex | Number of task runs contributed to study |                 |                 |
|---------------|-----|-----|--|-----------------|-----------------|
|               |     |     | Binary decoding                          | 18-way decoding | 36-way decoding |
| 1             | 28  | F   | 18                                       | 2               | 2               |
| 2             | 20  | F   | 2  | -               | -               |
| 3             | 50  | F   | 2  | -               | -               |
| 4             | 38  | M   | 2  | -               | -               |
| 5             | 54  | F   | 2  | -               | -               |
| 6             | 27  | F   | -  | 4               | 6               |
| 7             | 24  | F   | -  | -               | 6               |
| 8             | 31  | F   | -  | -               | 6               |

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2:**

Measures of decoding accuracy in a highly sampled participant from the ROC analyses in Figure 2B

| ROC            | AUC  | Sensitivity* | Specificity* | Accuracy* |
|----------------|------|--------------|--------------|-----------|
| Left vs. Rest  | 0.99 | 0.98         | 0.91         | 0.93      |
| Left vs. Right | 0.99 | 0.96         | 0.99         | 0.97      |
| Riaht vs. Rest | 0.99 | 0.98         | 0.91         | 0.93      |

\* Sensitivity, specificity, accuracy calculated at optimizec threshold based on Youden *J* statistic (Youden, 1950)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript