Reconstructing and decoding imagined letters from early visual cortex using ultra-high field fMRI

Authors

Mario Senden[1,2*], Thomas Emmerling[1,2*], Rick van Hoof[1,2*], Martin Frost[1,2], and Rainer Goebel[1,2,3]

* These authors contributed equally to the paper

1) Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, 6201BC Maastricht, The Netherlands

2) Maastricht Brain Imaging Centre, Faculty of Psychology and Neuroscience, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands

3) Department of Neuroimaging and Neuromodeling, Netherlands Institute for Neuroscience, an Institute of the Royal Netherlands Academy of Arts and Sciences (KNAW), 1105BA Amsterdam, The Netherlands

Address for Correspondence

Mario Senden, Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Oxfordlaan 55, 6200 MD Maastricht, P.O. Box 616, The Netherlands, Phone number: +31 43 38 82071;

Email: mario.senden@maastrichtuniversity.nl

Abstract

Brain-computer interfaces offer a way to communicate for people with severe motor and speech disabilities. However, successful current letter speller implementations require perception-driven paradigms (EEG) or cognitively demanding tasks (fMRI, fNIRS) which are not directly linked to letters visualized in the mind's eye. A more natural, content-based, BCI speller system immediately decoding imagined letters from associated brain activity is desirable. In the current study, we take the first steps towards such a BCI and offer new insights into the neural underpinnings of visual mental imagery, a process which is considered one of the main sources of human cognitive complexity. We demonstrate for the first time the feasibility to reconstruct visual field images which carry recognizable content of imagined letter shapes. Using submillimeter resolution fMRI data of six participants, detailed population receptive field maps, and a denoising autoencoder, we were able to reconstruct the visual field during perception and imagery of four different letter shapes. We, furthermore, achieve greater-than-chance classification accuracy on the four letters in five out of six participants. Finally, we show that reconstructions can be recognizable on a trial-by-trial basis, paving the way for real-time BCI applications.

Keywords

Mental imagery, fMRI, Decoding, Reconstruction, Letters, Visual cortex, BCI

Introduction

Brain-computer interfaces (BCIs) hold the promise of restoring the ability to communicate for patients suffering from complete or partial loss of voluntary motor control (Wolpaw, Birbaumer, McFarland, Pfurtscheller, & Vaughan, 2002). Non-invasive BCIs using electro-encephalography (EEG) have been successfully employed for severely motor-impaired patients (e.g. motor neuron degenerative diseases and paralysis) exploiting stimulus evoked responses (Birbaumer et al., 1999; De Massari et al., 2013; Nijboer et al., 2008; Wolpaw et al., 2002). For some patients, however, EEG-based BCIs are not effective (Chaudhary, Xia, Silvoni, Cohen, & Birbaumer, 2017), particularly for patients with reduced or lost control of the eye muscle (completely locked-in state, CLIS). For patients that do not benefit from EEG BCIs, functional magnetic resonance imaging (fMRI) and functional near-infrared spectroscopy (fNIRS) provide potential hemodynamic BCIs. While fNIRS, like EEG, can be used at the bedside of a patient, only fMRI has been demonstrated until now to provide a robust hemodynamic letter speller BCI (Sorger, Reithler, Dahmen, & Goebel, 2012) where subjects engage in various mental tasks (e.g. mental spatial navigation, mental calculation, mental drawing or inner speech). So far, fMRI-based BCI communication systems have mostly focused on coding schemes arbitrarily mapping brain activity in response to diverse mental imagery tasks, and hence originating from distinct neural substrates, onto letters of the alphabet (Birbaumer et al., 1999; Sorger et al., 2012). As such, current BCI speller systems do not offer a meaningful connection between the intended letter and the specific content of mental imagery. This is demanding for users as it requires them to memorize the mapping in addition to performing imagery tasks. Therefore, a more natural, content-based, BCI speller system immediately decoding imagined (i.e. internally visualized) letters from their associated brain activity is desirable; especially for novice BCI speller users. In order for this to be feasible, activity in response to different items within the same domain, and thus originating from a single neural substrate, must be sufficiently discriminable to uniquely identify each item. For visual shapes, such as

letters, this is principally given by their unique spatial activation profile (voxel pattern; VP) resulting from the retinotopic organization of early visual cortex (Holmes, 1918; Sperry, 1963).

It is likely that these retinotopy-based, discriminable, spatial activation profiles are also present for visual mental imagery since it shares neural circuitry with perception in early visual cortex (Kosslyn & Thompson, 2003; Kosslyn, Thompson, & Ganis, 2006; Pearson, Naselaris, & Holmes, 2015). Indeed, several studies have shown that visual mental imagery activates cortical networks that are also activated during corresponding perceptual tasks (Ganis, Thompson, & Kosslyn, 2004; R Goebel, Khorram-Sefat, & Muckli, 1998; Ishai, Ungerleider, & Haxby, 2000; Kosslyn, Thompson, & Alpert, 1997; Mechelli, Price, Friston, & Ishai, 2004; O'Craven & Kanwisher, 2000). Additionally, applying different forms of machine learning approaches to functional magnetic resonance imaging (fMRI) data enabled the decoding of imagery content regarding visual mental imagery of orientations (Albers, Kok, Toni, Dijkerman, & de Lange, 2013; Harrison & Tong, 2009), motion (Emmerling, Zimmermann, Sorger, Frost, & Goebel, 2016), objects (Cichy, Heinzle, & Haynes, 2012; Lee, Kravitz, & Baker, 2012; Reddy, Tsuchiya, & Serre, 2010), shapes (Stokes, Saraiva, Rohenkohl, & Nobre, 2011; Stokes, Thompson, & Cusack, 2009), and scenes (Johnson & Johnson, 2014; Naselaris, Olman, Stansbury, Ugurbil, & Gallant, 2015). Finally, two previous studies (Klein, Dubois, Mangin, Kherif, & Flandin, 2004; Slotnick, Thompson, & Kosslyn, 2005) demonstrated functionally specific retinotopic activations during visual imagery.

This raises confidence that decoding of internally visualized letters is possible. Additionally, recent advancements in the reconstruction of perceived visual stimuli from fMRI data (Miyawaki, Uchida, Yamashita, Sato, & Morito, 2008; Schoenmakers, Barth, Heskes, & Gerven, 2013; Thirion et al., 2006) – i.e. a visualization of what participants saw based on their brain activations – pose the question of whether reconstruction rather than mere decoding is possible for mental imagery as well. Studies that reconstruct visual perception based on neuroimaging data leveraged the retinotopic organization of early visual areas and fit invertible encoding

models (cf. Sprague, Saproo, & Serences, 2015) to individual voxels in these areas. A particularly popular, and straightforwardly invertible, encoding model is the two-dimensional isotropic Gaussian model of population receptive fields (pRFs; Dumoulin & Wandell, 2008). Inversion of population receptive fields of a large number of voxels measured at high spatial resolution may thus not only be used to reconstruct from brain activation in response to perceived but also imagined shapes. Using an integrative approach combining ultra-high field fMRI, inverted encoding models (IEMs) based on pRFs, and machine learning, it is the aim of the present study to decode and reconstruct the content of visual mental imagery.

Materials and Methods

Participants

Six participants (2 female, age range = (21 - 49), mean age = 30.7) with normal or corrected-to-normal visual acuity took part in this study. All participants were experienced in undergoing high field fMRI experiments, gave written informed consent and were paid for participation. All procedures were conducted with approval from the local Ethical Committee of the Faculty of Psychology and Neuroscience at Maastricht University.

Stimuli and Tasks

Each participant completed three training sessions in order to practice the controlled imagery of visual letters prior to a single scanning session which comprised four experimental (imagery) runs of ~11 minutes and one control (perception) run of ~ 9 minutes as well as one pRF mapping run of ~16 minutes.

Training Session and Task

Training sessions lasted ca. 45 minutes and were scheduled one week prior to scanning. Before the first training session, participants filled in the Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973) and the Object- Spatial Imagery and Verbal Questionnaire (OSIVQ; Blazhenkova &

Kozhevnikov, 2009). These questionnaires measure the subjective clearness and vividness of imagined objects and cognitive styles during mental imagery, respectively. In each training trial, participants saw one of four white letters ('H', 'T', 'S', or 'C') enclosed in a white square guide box (8° by 8° visual angle) on grey background and a red fixation dot in the center of the screen (see figure 1). With the onset of visual stimulation, participants heard a pattern of three low tones (note C5) and one high tone (note G5) that lasted 1000 ms. This tone pattern was associated with the visually presented letter with specific patterns randomly assigned for each participant. After 3000 ms the letter started to fade out until it completely disappeared at 5000 ms after trial onset. The fixation dot then turned orange and participants were instructed to maintain a vivid image of the presented letter. After an 18 second imagery period, the fixation dot turned white and probing started. With an inter-probe-interval of 1500 ms (jittered by ±200 ms) three white probe dots appeared within the guide box. These dots were located within the letter shape or outside of the letter shape (however, always within the guide box). Participants were instructed to indicate by button press whether a probe was located inside or outside the imagined letter shape. Depending on the response, the fixation dot turned red (incorrect) or green (correct) before turning white again as soon as the next probe was shown. The positions of the probe dots were randomly chosen such that they had a minimum distance of 0.16° and a maximum distance of 0.32° of visual angle from the edges of the letter (and the guide box), both for inside and outside probes. This ensured similar task difficulty across trials. A resting phase of 3000 ms or 6000 ms followed the three probes. At the beginning of a training run all four letters were presented for 3000 ms each, alongside the associated tone pattern (reference phase). During one training run, each participant completed 16 pseudo-randomly presented trials. In each training session, participants completed two training runs during which reference letters were presented in each trial (described above) and two training runs without visual presentation (i.e. the tone pattern was the only cue for a letter). Participants were instructed to maintain central fixation throughout

the entire run. After the training session, participants verbally reported the imagery strategies they used.

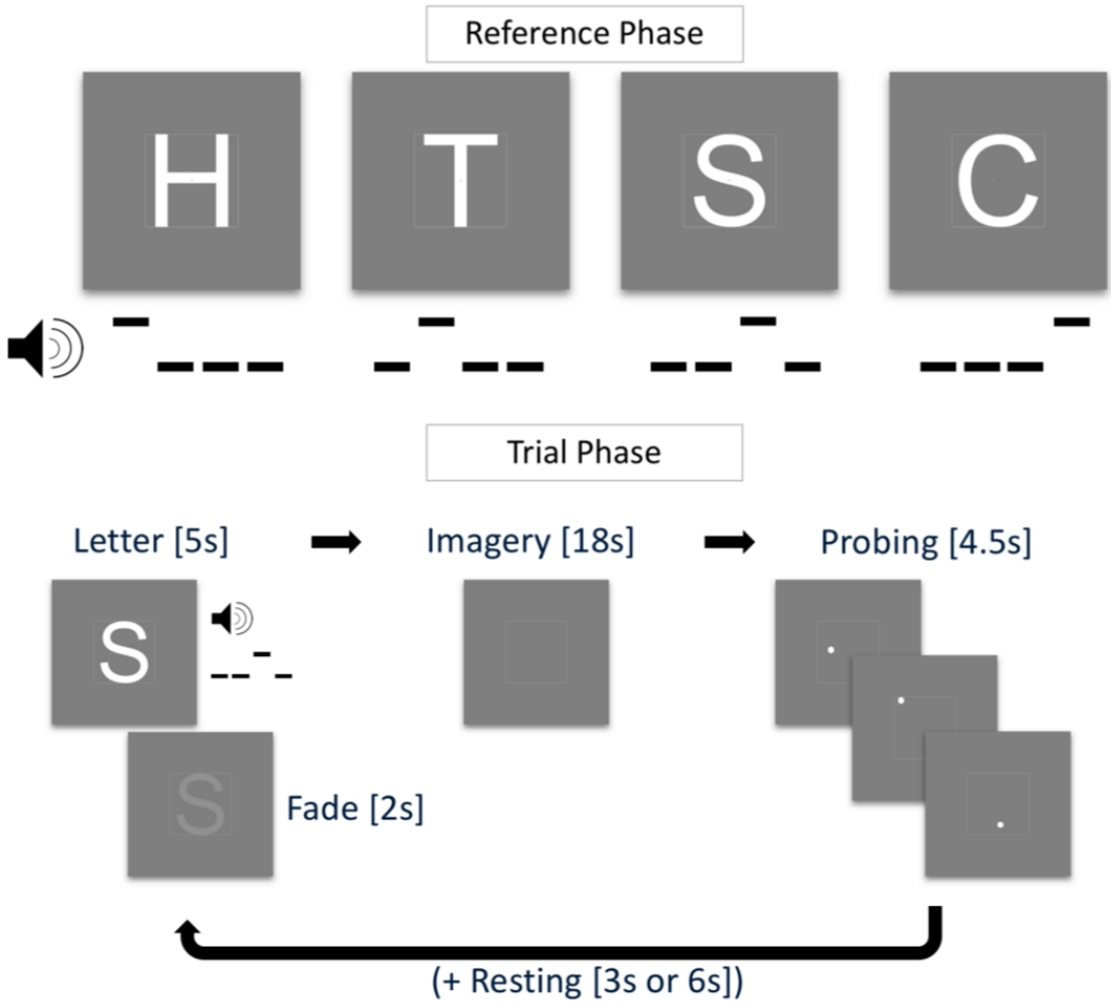


Figure 1: Training task. In the reference phase (top), four letters H, T, 'S' & 'C' were paired with a tone pattern. In the trial phase (bottom), the tone pattern was played and the letter shown for 5s (fading out after 3s) followed by an imagery period of 18s, a probing period of 4.5s, and a resting period of 3s or 6s

Imagery Runs

Imagery runs were similar to the training task with changes to the probing phase and the timing of the trial phase. After the reference phase in the beginning of each run, there was no visual stimulation other than the fixation dot and the guide box. Imagery phases started when participants heard the tone pattern and the fixation dot turned orange. Imagery phases lasted 6s. Participants were instructed to imagine the letter associated with the tone pattern as vividly and accurately as possible. The guide box aided the participant by acting as a reference for the physical dimensions

of the letter. The resting phases that followed each imagery phase lasted 9s or 12s. There was no probing phase in normal trials. In each experimental run, there were 32 normal trials and two additional catch trials which entailed a probing phase of four probes. There was no visual feedback for the responses in the probing phase (the fixation dot remained white). Data from the catch trials were not included in the analyses.

Perception Run

To measure brain activation patterns in visual areas during the perception of the letters used in the imagery runs we recorded one perception run during the scanning session. The four letters were visually presented using the same trial timing parameters as in the experimental runs. There was no reference nor probing phases. Letters were presented for the duration of the imagery phase (6s) and their shape was filled with a flickering checkerboard pattern (10 Hz). No tone patterns were played during the perception run. The recorded responses were also used to train a denoising auto-encoder (see below).

pRF mapping

A bar aperture (1.33° wide) revealing a flickering checkerboard pattern (10 Hz) was presented in four orientations. For each orientation the bar covered the entire screen in 12 discrete steps (each step lasting 3 seconds). Within each orientation the sequence of steps (and hence of the locations) was randomized (cf. Senden, Reithler, Gijsen, & Goebel, 2014). Each orientation was presented six times.

Stimulus Presentation

The bar stimulus used for pRF mapping was created using the open source stimulus presentation tool BrainStim (http://svengijsen.github.io/BrainStim/).Visual and auditory stimulation in the imagery and perception runs were controlled with PsychoPy (version 1.83.03; Peirce, 2007). Visual stimuli were projected on a frosted screen at the top end of the scanner table by means of an LCD projector (Panasonic, No PT- EZ57OEL; Newark, NJ, USA). Auditory stimulation was presented using MR-compatible insert earphones (Sensimetrics, Model S14; Malden,

MA, USA). Responses to the probes were recorded with MR-compatible button boxes (Current Designs, 8-button response device, HHSC-2x4-C; Philadelphia, USA).

Magnetic resonance imaging

We recorded anatomical and functional images with a Siemens Magnetom 7 Tesla scanner (Siemens; Erlangen, Germany) and a 32-channel head-coil (Nova Medical Inc.; Wilmington, MA, USA). Prior to functional scans, we used a T1-weighted magnetization prepared rapid acquisition gradient echo (3D-MP2RAGE; Marques et al., 2010) sequence [240 sagittal slices, matrix = 320 320, voxel size = 0.7 0.7 0.7 mm3, first inversion time TI1 = 900 ms, second inversion time TI2 = 2750 ms, echo time (TE) = 2.46 ms, repetition time (TR) = 5000 ms, first nominal flip angle = 5°, second nominal flip angle = 3°] to acquire anatomical data. For all functional runs we acquired high- resolution gradient echo (T2* weighted) echo-planar imaging (Moeller, Yacoub, & Olman, 2010) data (TE = 26 ms, TR = 3000 ms, generalized auto-calibrating partially parallel acquisitions (GRAPPA) factor = 3, multi-band factor = 2, nominal flip angle = 55°, number of slices = 82, matrix = 186 by 186, voxel size = 0.8 by 0.8 by 0.8 mm). The field-of-view covered occipital, parietal, and temporal areas. Additionally, before the first functional scan we recorded five functional volumes with opposed phase encoding directions to correct for EPI distortions that occur at higher field strengths (Andersson, Skare, & Ashburner, 2003).

Processing of (f)MRI data

We analyzed anatomical and functional images using BrainVoyager 20 (version 20.0; Brain Innovation; Maastricht, The Netherlands) and custom code in MATLAB (version 2015a; The Mathworks Inc.; Natick, MA, USA). We interpolated anatomical images to a nominal resolution of 0.8 mm isotropic to match the resolution of functional images. In the anatomical images, the grey/white matter boundary was detected and segmented using the advanced automatic segmentation tools of BrainVoyager 20 which are optimized for high-field MRI data. A region-growing approach analyzed local intensity histograms, corrected topological errors of the segmented grey/white matter border and finally reconstructed meshes of

the cortical surfaces (Rainer Goebel, Esposito, & Formisano, 2006; Kriegeskorte & Goebel, 2001). The functional images were corrected for motion artefacts using the 3D rigid body motion correction algorithm implemented in BrainVoyager 20 and all functional runs were aligned to the first volume of the first functional run. We corrected EPI distortions using a method similar to that described in Andersson, Skare, and Ashburner (Andersson et al., 2003) and the 'topup' tool implemented in FSL (S. Smith, Jenkinson, Woolrich, & Beckmann, 2004). The pairs of reversed phase encoding images recorded in the beginning of the scanning session were used to estimate the susceptibility-induced off-resonance field and correct the distortions in the remaining functional runs. After this correction, functional data were high-pass filtered using a general linear model (GLM) Fourier basis set of three cycles sine/cosine, respectively. This filtering included a linear trend removal. Finally, functional runs were co-registered and aligned to the anatomical scan using an affine transformation (9 parameters) and z-normalized to eliminate signal offsets and inter-run variance.

pRF Mapping and region-of-interest definition

For each subject, we fit location and size parameters of an isotropic Gaussian population receptive field model (Dumoulin & Wandell, 2008) by performing a grid search. In terms of pRF location, the visual field was split into a circular grid of 100 by 100 points whose density decays exponentially with eccentricity. Receptive field size exhibits a linear relationship with eccentricity with the exact slope depending on the visual area (Freeman & Simoncelli, 2011). For this reason, we explored slopes in the range from 0.1 to 1 (step = 0.1) as this effectively allows for exploration of a greater range of receptive field sizes (10 for each unique eccentricity value). Polar angle maps resulting from pRF fitting were projected onto inflated cortical surface reconstructions and used to define region-of-interests (ROIs) for bilateral visual areas V1, V2, and V3. The resulting surface patches from the left and right hemisphere were projected back into volume space (from -1 mm to +3 mm from the

segmented grey/white matter boundary). Volume ROIs were then defined for V1, V2, V3, and a combined ROI (V1V2V3).

Voxel patterns

All our analyses and reconstructions are based on letter-specific spatial activation profiles of voxel co-activations, which we refer to as voxel patterns (VPs). Voxel patterns within each ROI were obtained for both perceptual and imagery runs. First, for each run, single trial letter-specific VPs were obtained by averaging BOLD activations in the range from +2 until +3 volumes following trial onset and *z*-normalizing the result. This lead to a total of 32 (eight per letter) perceptual VPs and 128 (four runs each with eight trials per letter) imagery VPs. We furthermore computed perceptual and imagery grand-average VPs per letter by averaging over all single trial VPs (and runs in case of imagery) of a letter and *z*-normalizing the result. Imagery grand-average VPs were used in an encoding analysis (see below) while perceptual grand-average VPs were used for training a denoising autoencoder.

Encoding analysis

To validate the assumption that spatial activation profiles of visual mental imagery is similarly retinotopically organized as perception, we test whether voxel activations predicted from the encoding model (one isotropic Gaussian per voxel) and a physical (binary) stimulus corresponding to the imagined letter provides a significantly better fit with measured voxel activations than predictions from the remaining binary letter stimuli. Specifically, for each participant and ROI, we predicted voxel activations for each of the four letters based on pRF estimates and physical letter stimuli. Subsequently, we performed a mixed-model regression for the grand-average voxel activations of each imagined letter within each ROI with physical letter as fixed and participant as random factors, respectively. Finally, we performed a contrast analysis. For each imagined letter the contrast was always between the corresponding physical stimulus and all remaining physical stimuli. For example, when considering voxel activations for the imagined letter ‘H’, a weight of 3 was placed on activations predicted from the physical letter ‘H’ and a weight of

-1 was placed on activations predicted from each of the remaining three letters. Since we repeated the analysis for each imagined letter (4) and single region ROI (3), we performed a total of 12 tests and considered results significant at a corrected cutoff of $\alpha_c = 0.05/12 = 0.0042$.

Autoencoder

We trained a denoising autoencoder (Vincent, Larochelle, Bengio, & Manzagol, 2008) with a single hidden layer (see figure 2a) to reproduce grand-average perceptual VPs from noise corrupted versions per subject and ROI. Since the values of VPs follow a Gaussian distribution with a mean of zero and unit standard deviation, we opted for zero-mean additive Gaussian noise with a standard deviation $\sigma = 12$ for input corruption. The hidden layer consisted of rectified linear units (ReLUs) while output units activated linearly. We used mean squared distances to measure reconstruction loss and implemented the autoencoder in the TensorFlow library (Abadi et al., 2016) for Python (version 2.7, Python Software Foundation, https://www.python.org/). The autoencoder was trained using the Adam optimizer (Kingma & Ba, 2014) with a learning rate of 1×10^{-5} and a batch size of 100 for 2,000 iterations. In addition to the four grand-average perceptual VPs, we also included an equal amount of noise corrupted mean luminance images to additionally force reconstructions to zero, if the input contained no actual signal.

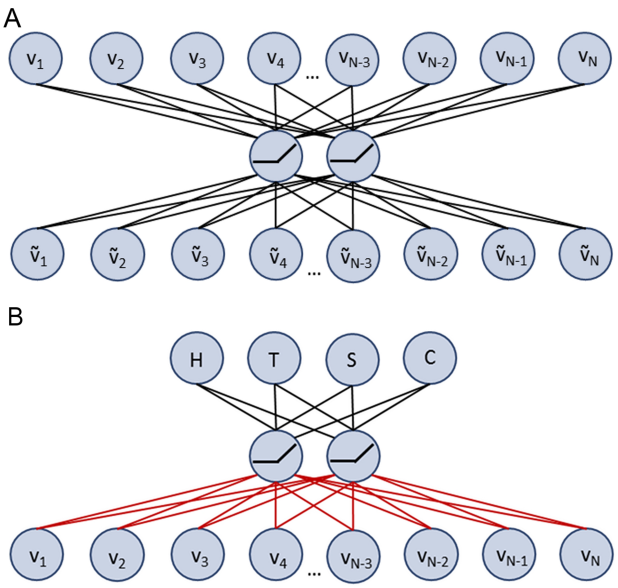


Figure 2: Network architectures. *Panel A)* shows a three-layer denoising autoencoder. The input and output layer consist of one unit per voxel while the number of units in the hidden layer is 10% of that in the input. Input is corrupted by additive Gaussian noise. Units in the hidden layer have a nonlinear activation function (ReLU) while output units activate linearly. Encoding weights (from input to hidden layer) and decoding weights (from hidden to output layer) are shared such that $\mathbf{W_d} = \mathbf{W_e}^T$. *Panel B)* shows a three-layer classification architecture. The output layer has one unit per letter and uses the softmax activation function. Input and hidden layers are as in the autoencoder. Red weights indicate that these weights have been trained previously in the autoencoder and remain fixed while training the classifier.

Classification

Once the autoencoder was trained, we replaced its output layer with a four-unit (one for each letter) softmax classifier (see figure 2b). Weights from the hidden to the classification layer as well as the biases of output units were then trained to classify single trial VPs in imagery runs using cross-entropy as a measure of loss. Note that pretrained weights from input to hidden layer as well as hidden unit biases remained fixed. This is conceptually similar to performing multinomial logistic regression on hidden layer representations. Imagery runs were split into training and testing datasets in a leave-one-run-out procedure such that the classifier was repeatedly trained on a total of 96 VPs (32 trials for each of three runs) and tested on the remaining 32 VPs. We again trained the network using the Adam optimizer. However, in this case the learning rate was

1×10^{-4}, the batch size equal to 96, and training lasted merely 250 iterations.

For each subject and ROI, we evaluated average classification accuracy across the four runs against a Null-distribution obtained from 1,000 permutations of the leave-one-run out procedure with randomly scrambled labels. We consider accuracy results significant if they exceed the 95th percentile of the Null distribution.

Reconstruction

For each subject and ROI, we reconstructed the visual field for the grand-average perceptual VP for each letter as well as from intra-trial averages of imagery VPs (i.e. averages over letter trials within a run). We obtained weights mapping the cortex to a visual field image (VFI) by inverting the mapping from visual field to cortex given by the population receptive fields:

$$\mathbf{W}_{VFI} = \left(\mathbf{W}_{pRF}\mathbf{D} \right)^{T}$$

where $\mathbf{W}_{pRF}$ is a v-by-p matrix (with v being the number of voxels and p the number of pixels) mapping a 150-by-150 pixel visual field to the cortex (i.e. $p = 22500$ pixels; after vectorizing the visual field) and $\mathbf{D}$ is a diagonal matrix of the inverse outdegree of each pixel in the visual field. The diagonal matrix $\mathbf{D}$ ensures that the sum total of weights impinging on each pixel is equal to one and corrects for overrepresentations of central regions in the visual field due to cortical magnification. Using these weights, VFIs could be obtained from VPs via a simple matrix multiplication

$$\mathbf{VFI} = \mathbf{W}_{VFI}\mathbf{VP}$$

.

For each letter, we assessed the quality of its reconstruction by calculating the correlation between its VFI and the corresponding binary letter

stimulus. This constitutes a first-level correlation metric of reconstruction quality. However, since the four letters bear different visual similarities with each other (e.g. 'S' and 'C' might resemble each other more closely than either resemble 'H'), we also define a second-level correlation metric. Specifically, we obtain one vector of all pairwise correlations between physical letter stimuli and a second vector of pairwise correlations between corresponding VFIs and compute correlations between these pairwise correlation vectors.

Results

Behavioral results

VVIQ and OSIVQ scores for each participant are shown in figure 3. The average score over participants for VVIQ was 4.07 (95% CI [3.71, 4.43]). For the object, spatial, and verbal sub-scales of OSIVQ, average scores were 2.88 (95% CI [2.48, 3.27]), 3.08 (95% CI [2.75, 3.41]), and 3.81 (95% CI [3.33, 4.29]), respectively. Participants reported that they tried to maintain the afterimage of the fading stimulus as a strategy to enforce vivid and accurate letter imagery. Furthermore, participants determined through button presses whether a probe was located inside or outside the letter shape while the letter was either visible or imagined. A repeated-measures ANOVA with task (visible or invisible runs) and time as within-subject factors revealed a statistically significant effect of time on probing accuracy ($F_{(2,10)} = 19.84$, $p \ll 0.001$), and no significant difference for task ($F_{(1,5)} = 1.10$, $p = .341$).

Table 1. Probing accuracy (averages over participants and time).

	T1	T2	T3
Visible	60.42 (95% CI [48.2, 72.64])	75.39 (95% CI [66.70, 84.08])	77.73 (95% CI [69.36, 86.10])
Invisible	62.02 (95% CI [44.57.36, 79.45])	73.18 (95% CI [65.98, 80.38])	81.57(95% CI [75.47, 87.67])

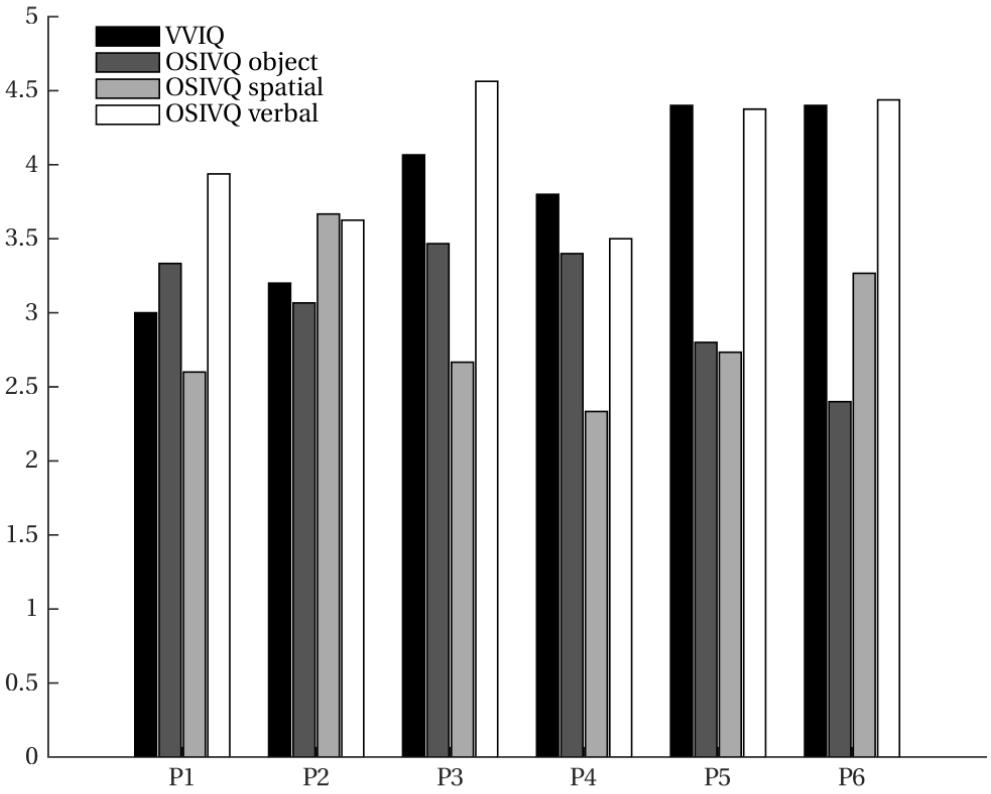


Figure 3: Vividness of visual imagery. Vividness of Visual Imagery Questionnaire (VVIQ) and Object-Spatial Imagery and Verbal Questionnaire (OSIVQ) scores (with the subscales for "object", "spatial", and "verbal" imagery styles) are shown for all participants.

Encoding

For each imagined letter (H, T, S, C) in each single-area ROI (V1, V2, V3), we investigated whether voxel activations can be predicted from a pRF encoding model and the corresponding physical stimulus. That is, for each imagined letter-ROI combination, we ran a mixed-model regression with observed imagery voxel activations (averaged within the time window +2 to +3 volumes following trial onset) as outcome and predicted voxel activations for each physical letter stimulus as predictors and participants as grouping variable. Since we were specifically interested in evaluating our assumption that imagery voxel activations are (just as their perceptual counterparts) retinotopically organized and that this is sufficient to

distinguish among different imagined letters, we performed contrast analyses between the physical letter corresponding to the imagery and all remaining letters (see Methods for details). Contrasts were significant after applying Bonferroni correction ($\alpha_c = 0.0042$) for each of the twelve letter-ROI combinations. Specifically, for V1, predictions based on the physical letter 'H' gave a better account of voxel activations observed for the imagery of letter 'H' than those based on every other physical letter ($t_{(2)}$ = 32.11, p = 0.0004). Similarly for 'T' ($t_{(2)}$ = 48.00, p = 0.0002), 'S' ($t_{(2)}$ = 14.10, p = 0.0025), and 'C' ($t_{(2)}$ = 29.84, p = 0.0006). In the same vain for V2, 'H' ($t_{(2)}$ = 25.21, p = 0.0008), 'T' ($t_{(2)}$ = 67.63, p = 0.0001), 'S' ($t_{(2)}$ = 19.64, p = 0.0013), and 'C' ($t_{(2)}$ = 47.48, p = 0.0002) and V3, 'H' ($t_{(2)}$ = 47.90, p = 0.0006), 'T' ($t_{(2)}$ = 27.60, p = 0.0007), 'S' ($t_{(2)}$ = 11.48, p = 0.0038), and 'C' ($t_{(2)}$ = 32.83, p = 0.0005). Figure 4 visualizes these results in the form of boxplots of first-level beta values (i.e. distribution over participants per physical letter) in each letter-ROI combination.

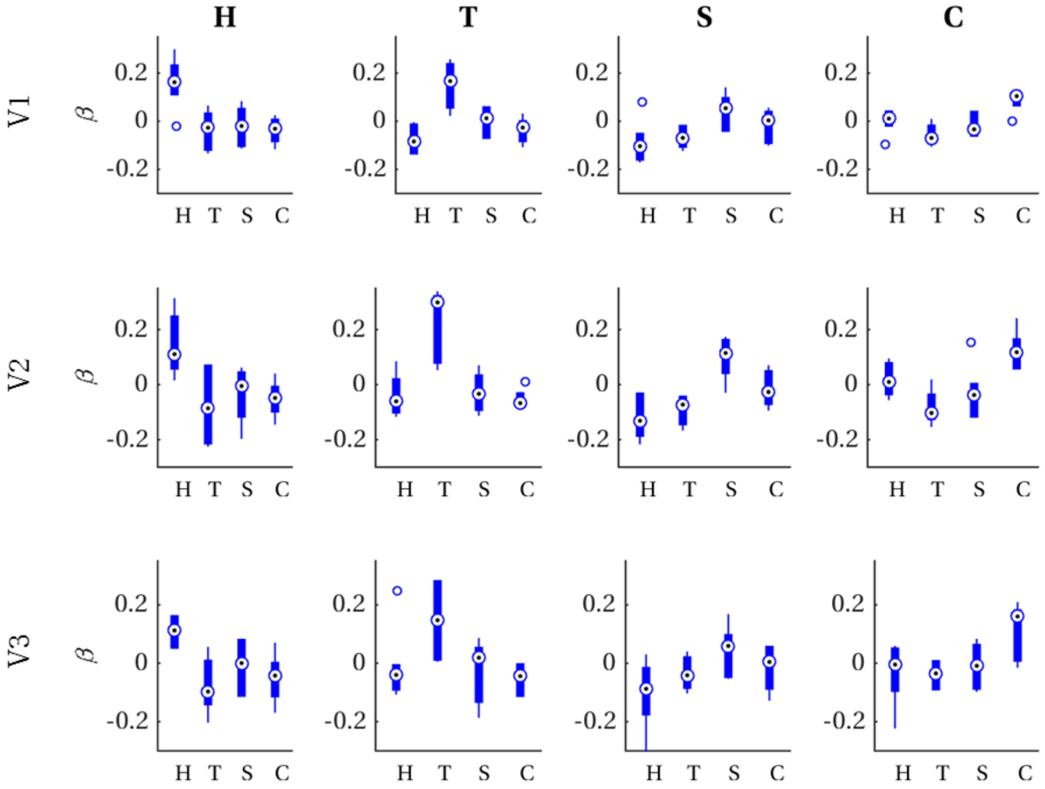


Figure 4: First-level beta distributions. Distribution of first-level beta values (across participants) for VPs predicted from each physical letter (x-axis) for all combinations of ROI (rows) and imagined letters (columns).

Classification

Having validated the assumption that voxel activations in response to visual mental imagery are at least in part given by voxels' population receptive fields and the shape of the imagined letters, we proceeded to construct a neural network classifier. The classifier consists of three layers with the output layer being a softmax classifier stacked onto the hidden layer of an autoencoder pretrained to denoise perceptual VPs (see Methods for details). We trained the classifier on imagery data using leave-one-run-out procedure; that is, we trained the classifier on three of the four imagery runs and tested classification accuracy on the left-out run. Figure 5 shows average classification accuracies per subject and ROI (including the combined ROI 'V1V2V3'). For five of the six participants, average classification accuracies exceeded theoretical chance levels (25% correct) as well as the 95th percentile of 1,000 permutation runs (randomly

scrambled labels) in all ROIs. For participant six, theoretical chance levels as well as the 95th percentile were (barely) exceeded for V2 only.

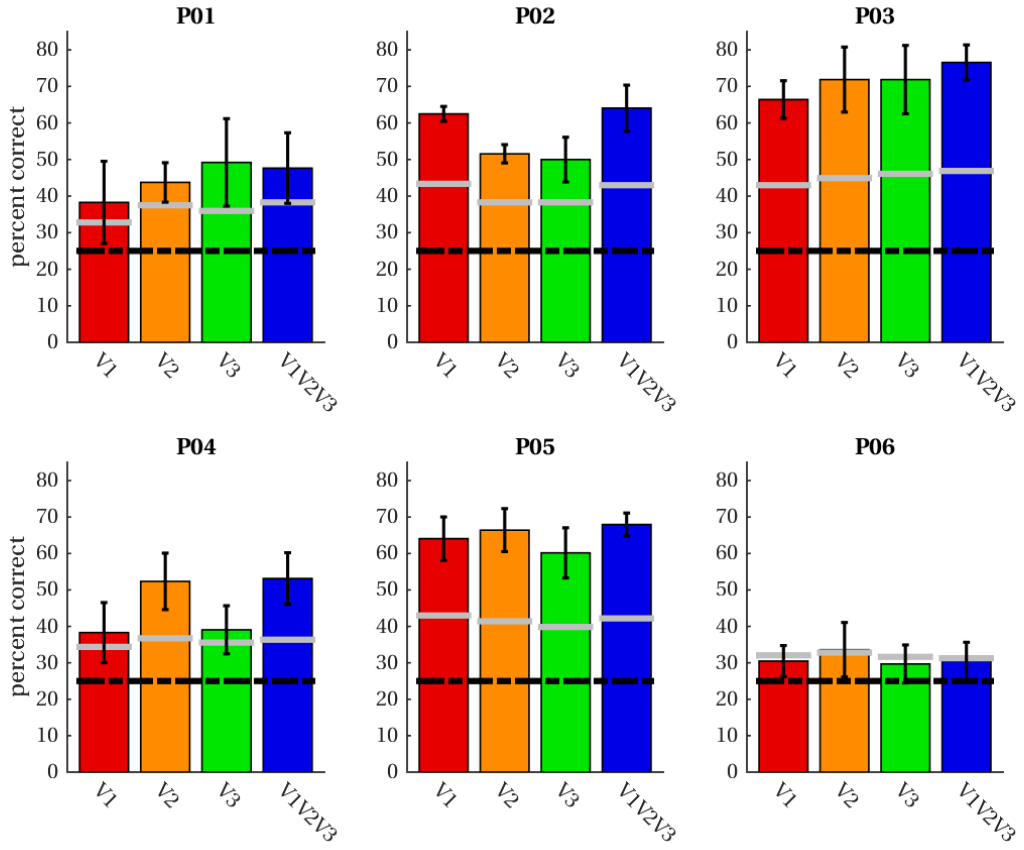


Figure 5: Classification accuracies. Average classification accuracies across four leave-one-out runs of imagery data are given for four ROIs in each participant. Classification was performed for letter-specific voxel patterns averaged in the range from +2 until +3 volumes after trial onset. The black dashed line indicates accuracies expected by chance; grey lines demarcate the 95th percentile of permutation classification accuracies.

As can be appreciated from these results as well as the figure, classification accuracies vary across ROIs as well as across subjects. Differences between subjects might be due to differences in their ability to imagine shapes accurately and vividly as measured by the VVIQ and OSIVQ questionnaires. Differences between ROIs might be due to differences with respect to their retinotopy (mostly receptive field sizes) or due to different numbers of voxels we included for analysis of each ROI. Only the former would be a true ROI effect. We investigate which factors account for the observed average accuracy by performing a mixed-model regression with questionnaire scores, ROI (using dummy coding, V1 =

reference), and number of selected voxels as predictors with number of voxels being grouped by ROI. We further performed stepwise model reduction to arrive at the most parsimonious account of our results. In the full model the VVIQ and the OSIVQ spatial and OSIVQ object scores were included but the OSIVQ verbal scores were not since those correlated highly with VVIQ scores (leading to collinearity) and are arguably the least relevant for mental imagery of visual shapes. To further prevent collinearity, we also only included single-area ROIs in this analysis and not the combined ROI. The most parsimonious model retains three significant predictors of average classification accuracy: number of voxels ($t_{(14)}$ = 5.37, $p \ll 0.001$), the object sub-score of OSIVQ ($t_{(14)}$ = 4.5712, $p <$ 0.001), and the spatial sub-score of OSIVQ ($t_{(14)}$ = 2.95, p = 0.011).

Reconstruction

After feeding the BOLD timeseries of each run through the denoising autoencoder, we computed clean grand-average imagery VPs from which we reconstructed VFIs (see figure 6). Average correlations between reconstructed imagery and physical letters are presented in table 2 (for comparison, table 3 shows correlations between reconstructed perception and physical letters). We performed a mixed-model regression with the VVIQ and the OSIVQ spatial and OSIVQ object scores, ROI (using dummy coding, V1 = reference), letter (dummy coding, 'H' = reference), and number of selected voxels (again grouped by ROI) as predictors to assess which factors account for the observed correlations (transformed to Fisher z-scores for statistical analysis). We again performed stepwise model reduction to arrive at the most parsimonious account of our results. The final model retained number of voxels ($t_{(62)}$ = 3.95, $p < 0.001$) and the OSIVQ object score $t_{(68)}$ = 6.03, $p \ll 0.001$) as significant quantitative predictors. Furthermore, both categorical predictors were significant. Specifically, letter 'T' ($t_{(68)}$ = 8.37, $p \ll 0.001$) presented with significantly improved correlation values over the reference letter 'H' whereas the letter 'S' ($t_{(68)}$ = -2.95, p = 0.004) presented with significantly decreased correlation values with respect to the reference. Finally, correlations were

significantly reduced for both V2 ($t_{(68)}$ = -3.38, p = 0.001) and V3 ($t_{(68)}$ = -2.80, p = 0.007) with respect to V1.

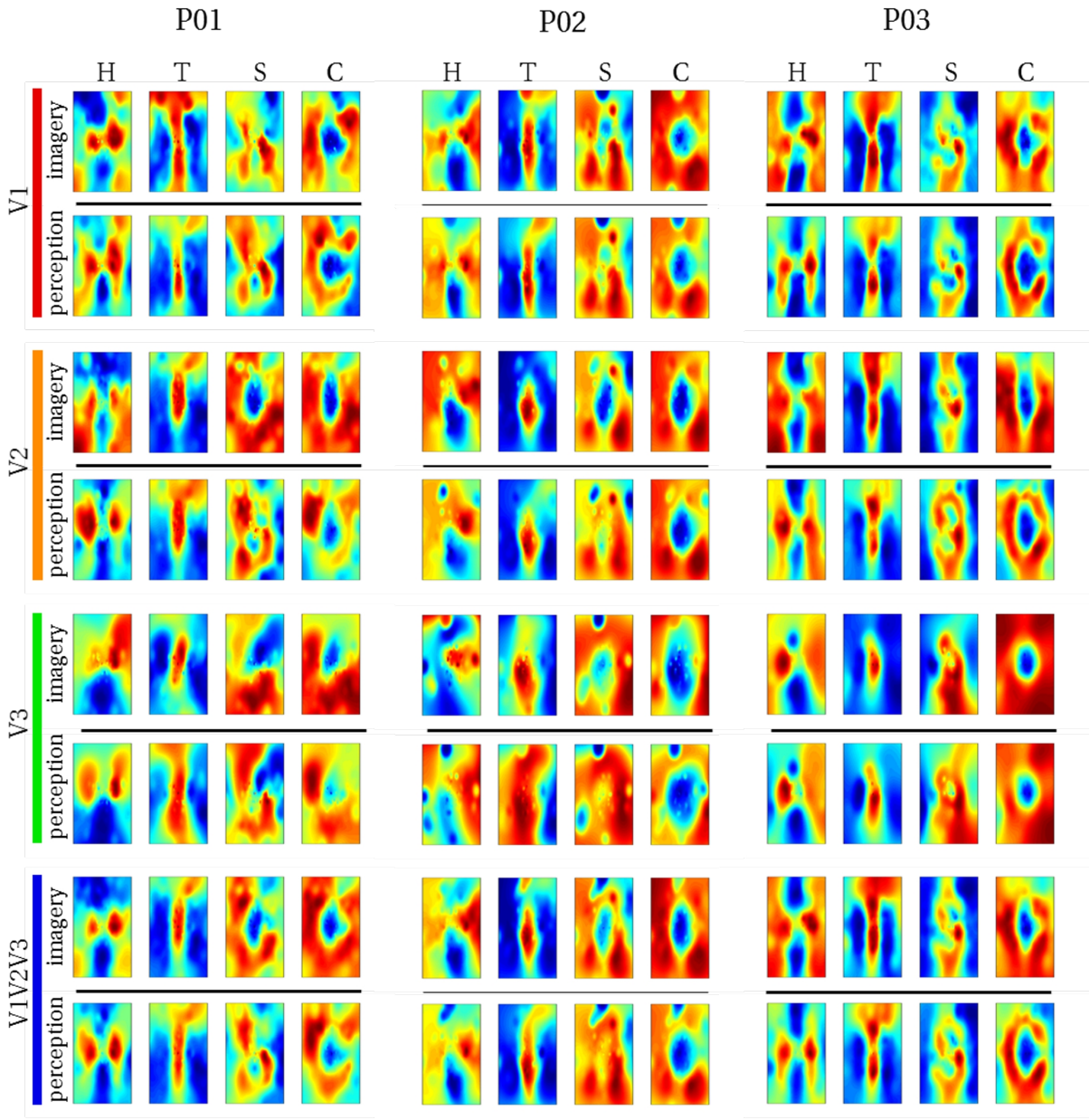


Figure 6: Reconstructed visual field images (participants 1-3). Reconstructed average VFIs are visualized for each ROI of participants one, two, and three. Reconstructions of the remaining three subjects are shown in figure 7. Perceptual voxel patterns were obtained from the raw BOLD time-series while imagery voxel patterns were obtained from cleaned BOLD time-series after feeding raw data through the autoencoder. For comparison, reconstructions of imagined letters without using the autoencoder can be found in supplementary figure 1.

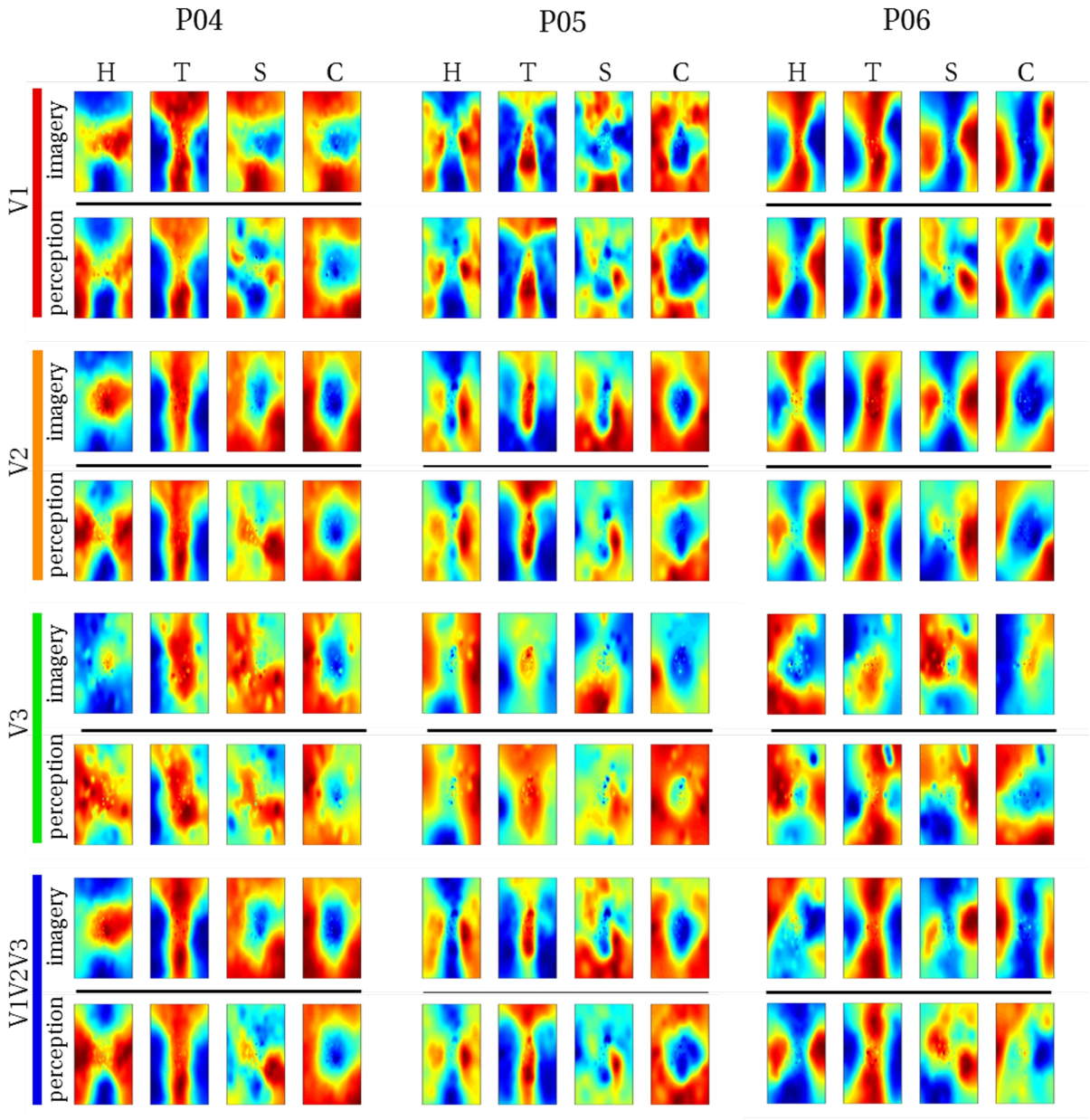


Figure 7: Reconstructed visual field images (participants 4-6). Reconstructed average VFIs are visualized for each ROI of participants four, five, and six. Reconstructions of the remaining three subjects are shown in figure 6. Perceptual voxel patterns were obtained from the raw BOLD time-series while imagery voxel patterns were obtained from cleaned BOLD time-series after feeding raw data through the autoencoder. For comparison, reconstructions of imagined letters without using the autoencoder can be found in supplementary figure 2.

Next, we examined the second-level correlation metric of reconstruction quality. Correlations between physical and reconstruction pairwise first-level correlation vectors were 0.65 (95% CI [0.39, 0.82], p = 0.080) for V1, 0.60 (95% CI [0.23, 0.82], p = 0.106) for V2, 0.44 (95% CI [0.10, 0.69], p = 0.189) for V3, and 0.77 (95% CI [0.58, 0.89], p = 0.037) for V1V2V3, respectively. None of these correlations were significant after Bonferroni

correction ($\alpha = 0.05/4 = 0.0125$). However, the correlation observed for V1V2V3 was significant at an uncorrected alpha level. Finally, we again performed a mixed regression to assess which factors account for the observed correlations (again transformed to Fisher z-scores). We included VVIQ and the OSIVQ spatial and OSIVQ object scores, ROI (using dummy coding, V1 = reference), and number of selected voxels (grouped by ROI) as predictors and performed stepwise model reduction. The only significant predictor remaining after this procedure was the VVIQ score ($t_{(16)}$ = -3.41, p = 0.004). Note that a lower VVIQ score corresponds to higher imagery vividness.

Table 2. First order correlations between reconstructed imagined letters and physical stimuli (averages over participants).

	H	T	S	C
V1	0.25 (95% CI [0.01, 0.46])	0.56 (95% CI [0.53, 0.59])	0.14 (95% CI [-0.01 0.29])	0.27 (95% CI [0.18, 0.35])
V2	0.16 (95% CI [-0.06, 0.36])	0.50 (95% CI [0.43, 0.56])	0.07 (95% CI [-0.05, 0.19])	0.27 (95% CI [0.23, 0.32])
V3	0.20 (95% CI [0.08, 0.31])	0.36 (95% CI [0.25, 0.45])	0.01 (95% CI [-0.11, 0.14])	0.19 (95% CI [0.04, 0.32])
V1V2V3	0.28 (95% CI [0.14, 0.41])	0.55 (95% CI [0.52, 0.58])	0.11 (95% CI [-0.04, 0.24])	0.25 (95% CI [0.17, 0.33])

Table 3. First order correlations between reconstructed perceived letters and physical stimuli (averages over participants).

	H	T	S	C
V1	0.37 (95% CI [0.34, 0.41])	0.57 (95% CI [0.54, 0.60])	0.26 (95% CI [0.16, 0.35])	0.30 (95% CI [0.22, 0.38])
V2	0.35 (95% CI [0.30, 0.40])	0.51 (95% CI [0.46, 0.57])	0.19 (95% CI [0.09, 0.28])	0.30 (95% CI [0.23, 0.36])
V3	0.24 (95% CI [0.19, 0.30])	0.38 (95% CI [0.29, 0.47])	0.06 (95% CI [-0.06, 0.18])	0.27 (95% CI [0.24, 0.29])
V1V2V3	0.39 (95% CI [0.35, 0.43])	0.55 (95% CI [0.51, 0.60])	0.22 (95% CI [0.12, 0.31])	0.30 (95% CI [0.24, 0.36])

In addition to reconstructing VFIs from grand average imagery VPs, it can be illuminating to examine reconstructions from single trial VPs. We focus here on participants three and five whose classification accuracy indicates that they were highly successful at imagery on a trial by trial basis. Obviously, these participants are not representative of the population at large but provide an indication of what is possible for people with a strong ability to imagine visual shapes. Figures 8 shows imagery VFIs for individual trials of each letter in a single run of participant three with denoising. For denoised data, mean correlation values across trials (and runs) were 0.39 (95% CI [0.32, 0.45]) for 'H', 0.55 (95% CI [0.46, 0.62]) for 'T', 0.10 (95% CI [0.04, 0.16]) for 'S', and 0.09 (95% CI [0.06, 0.12]) for 'C', respectively. As a comparison, mean correlations for data which has not been denoised were 0.19 (95% CI [0.15, 0.21]) for 'H', 0.33 (95% CI [0.28, 0.38]) for 'T', -0.02 (95% CI [-0.06, 0.02]) for 'S', and 0.02 (95% CI [-0.02, 0.06]) for 'C', respectively. Figure 9 shows imagery VFIs for individual trials of each letter in a single run of participant five. For denoised data, mean correlations were 0.28 (95% CI [0.20, 0.35]) for 'H', 0.53 (95% CI [0.43, 0.61]) for 'T', 0.08 (95% CI [0.00, 0.17]) for 'S', and 0.21 (95% CI [0.12, 0.31]) for 'C', respectively. For data which has not been denoised, mean correlations were 0.12 (95% CI [0.08, 0.15]) for 'H', 0.32 (95% CI [0.25, 0.37]) for 'T', 0.02 (95% CI [-0.01, 0.06]) for 'S', and 0.07 (95% CI [0.03, 0.10]) for 'C', respectively.

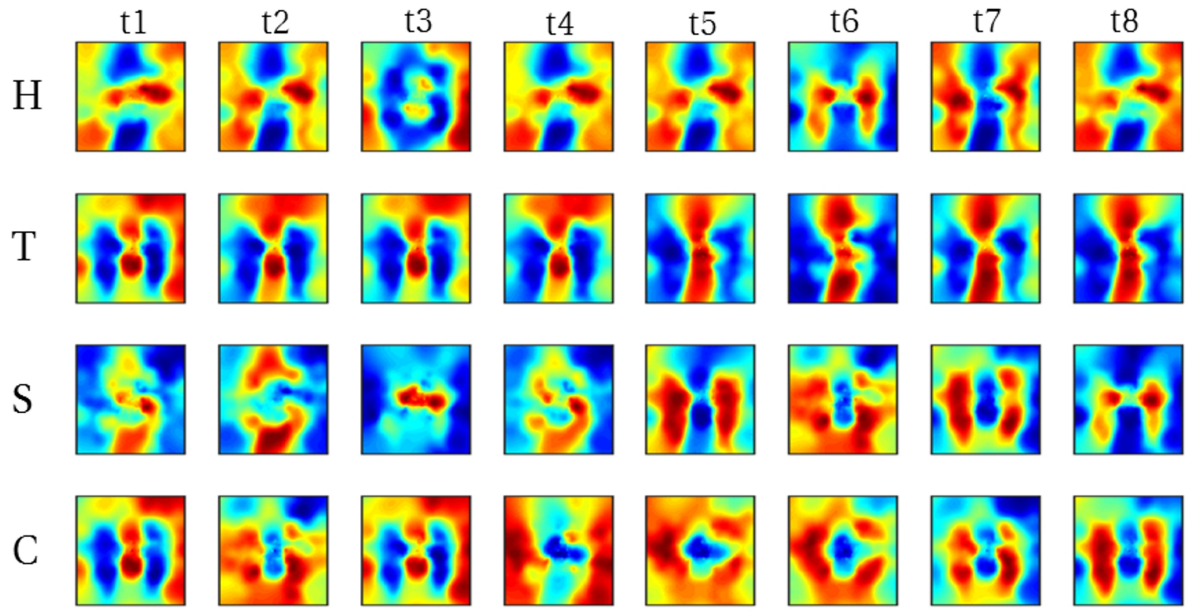


Figure 8: Reconstructed visual field images from denoised single trials in a single run of participant 3. Each run comprised of 8 trials (columns) per letter (rows). Recognizable reconstructions can be obtained for a number (though not all) individual trials. For comparison, reconstructions of imagined letters without using the autoencoder can be found in supplementary figure 3.

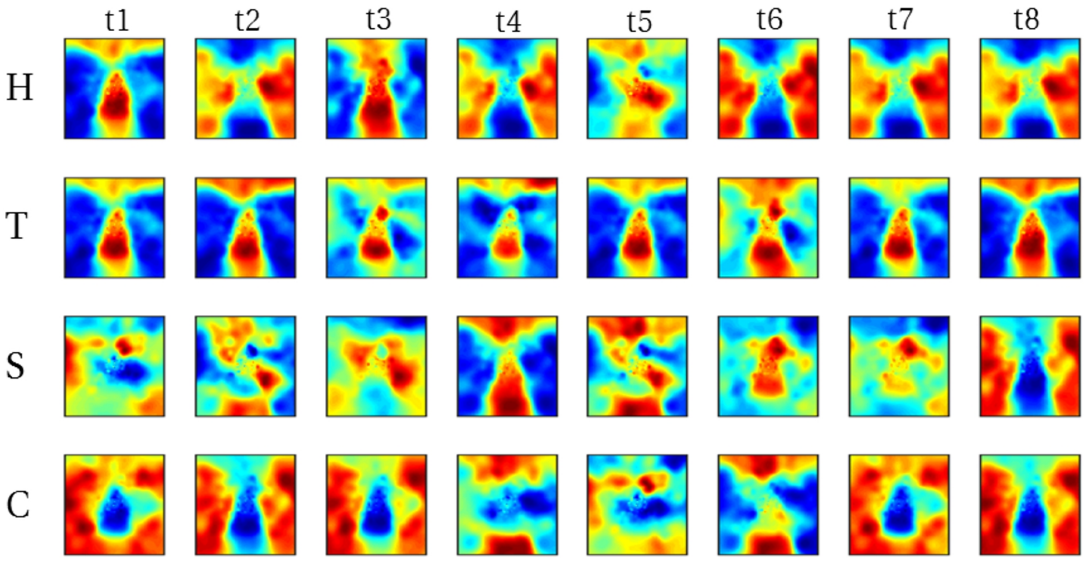


Figure 9: Reconstructed visual field images from denoised single trials in a single run of participant 5. Each run comprised of 8 trials (columns) per letter (rows). Recognizable reconstructions can be obtained for a number (though not all) individual trials.For comparison, reconstructions of imagined letters without using the autoencoder can be found in supplementary figure 4.

Discussion

We provide evidence that specific content of visual mental imagery in the shape of letters can not only be decoded but also reconstructed from 7 Tesla sub-millimeter voxel activity patterns. Our novel fMRI decoding approach employs inverted encoding models to project individual pRFs back into the visual field and machine learning tools to discriminate among four visually imagined letters using submillimeter fMRI images of early visual cortex. Importantly, our approach offers a more direct link to visual imagery content which is especially relevant for BCI letter speller applications.

Over training sessions all participants reached a high probing accuracy for both imagery and perception trials, showing that they could reliably indicate the location of the invisible letter shape in visual space. The ability to imagine the borders of the letter in absence of visual stimulation suggests participants were able to internally visualize the instructed letter. Next, we showed that voxel activations predicted by an pRF encoding model and a physical (binary) letter stimulus can account for observed voxel activations in response to mental imagery of the letter corresponding to the physical stimulus. Given that pRF mapping has shown to accurately predict fMRI responses to visual stimuli (Wandell & Winawer, 2015), our results suggest that intrinsic geometric organization of visual experiences are also maintained in visual mental imagery.

In five out of six participants, we were able to classify imagined letters with a high degree of accuracy from at least one region of interest (between 50% and 75% correct). Interestingly, classification accuracy varied not only across subjects but also across ROIs. Yet, ROIs do not constitute a significant predictor of classification accuracy. Rather, the number of voxels included for any given ROI determined classification accuracy. However, this does not imply that uncritically adding more voxels will lead to higher classification accuracies. We included only those voxels for which pRF mapping yielded a high fit. It is likely that classification accuracy benefits from a large number of voxels whose pRF

can be estimated to a high degree of precision (i.e. which show a strong spatially selective visual response) rather than a large number of voxels per se. Our analysis revealed two additional significant predictors, namely the OSIVQ object and OSIVQ spatial scores. This indicates that these scores can be useful for screening participants to ensure that only those who show a strong ability for vivid imagery need to undergo costly and physically exhausting fMRI measurements.

With respect to visual field reconstructions, we found significant overlap between reconstructed VFIs of imagery data with the physical stimulus. This was expected given our and previous findings that visual mental imagery exhibits retinotopic organization in early visual cortex (Albers, Kok, Toni, Dijkerman, & de Lange, 2013; Pearson et al., 2015; Slotnick et al., 2005). Our first-level correlation metric of reconstruction quality revealed that reconstructions based on V1 data showed the highest resemblance to reference images compared to V2 and V3. This finding is reasonable given that receptive fields in V1 are smaller than in the other regions (A. T. Smith, Singh, Williams, & Greenlee, 2001) allowing for resolving finer spatial detail. If quality of reconstruction indeed depends on the ability to resolve fine spatial detail, then one would also expect that stimuli exhibiting finer (coarser) spatial layouts would be harder (easier) to reconstruct. This fits with the observation that reconstruction quality of the letter ‘S’ was significantly reduced while that of letter ‘T’ was significantly improved with respect to that of letter ‘H’. As with classification accuracy the OSIVQ object score was a significant predictor of reconstruction quality given by the first-level correlation metric. In contrast to classification accuracy, the OSIVQ spatial score displayed no predictive value for first-level reconstruction quality. Furthermore, neither OSIVQ sub-score was a significant predictor of second-level reconstruction quality. Only the VVIQ was a significant predictor of this metric. This is interesting given that the VVIQ could not predict classification accuracy nor first-level reconstruction quality.

There are substantial differences between participants both in terms of reconstruction quality and classification accuracy rendering pre-screening

highly important. While it is not possible to achieve good reconstruction for everyone, we showed that for those who exhibit a strong ability for visual mental imagery, it is possible to obtain recognizable reconstructions even at the single trial level when using the denoising autoencoder. This offers the opportunity to provide real-time visual feedback to participants in the form of online reconstructions of their imagined letters. This feedback could serve as a visual aid for participants' imagery which might free sufficient mental resources to focus imagery on adjusting poorly reconstructed regions of the letter.

Overall, our letter classification and reconstruction approach could be particularly suitable for communication in cases where voluntary muscle movement is impaired (e.g. locked-in syndrome), and imagery of letters can be identified in matter of seconds in a more natural and direct way than current BCI letter speller implementations (Sorger et al., 2012). Nevertheless, questionnaires might not be sufficient tools for screening and both extensive training and feedback are desirable, especially when including all letters of the alphabet in future studies. Our work constitutes an important first step in the development of content-based BCI speller systems.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 7202070 (HBP SGA1) as well as under ERC-2010-AdG grant (269853).

Author Contributions

M.S. developed the decoding and reconstruction procedures. T.E. and R.G. designed the experiments. T.E. and R.H. performed experiments. M.S., T.E. and R.H analyzed the data. MS., T.E. and R.H wrote the manuscript with comments from other authors. M.F. and R.G. provided guidance and expertise.

Declaration of Interests

The authors declare no competing interests.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., … Brain, G. (2016). TensorFlow: A System for Large-Scale Machine Learning TensorFlow: A system for large-scale machine learning. In *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16)* (pp. 265–284). https://doi.org/10.1038/nn.3331

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & de Lange, F. P. (2013). *Shared Representations for Working Memory and Mental Imagery in Early Visual Cortex*. *Current Biology* (Vol. 23). https://doi.org/10.1016/j.cub.2013.05.065

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & de Lange, F. P. (2013). *Shared Representations for Working Memory and Mental Imagery in Early Visual Cortex*. *Current Biology* (Vol. 23). https://doi.org/10.1016/j.cub.2013.05.065

Andersson, J. L. R., Skare, S., & Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage*, *20*(2), 870–888. https://doi.org/10.1016/S1053-8119(03)00336-7

Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kübler, A., … Flor, H. (1999). A spelling device for the paralysed. *Nature*, *398*(6725), 297–298. https://doi.org/10.1038/18581

Blazhenkova, O., & Kozhevnikov, M. (2009). The new object-spatial-verbal cognitive style model: Theory and measurement. *Applied Cognitive Psychology*, *23*(5), 638–663. https://doi.org/10.1002/acp.1473

Chaudhary, U., Xia, B., Silvoni, S., Cohen, L. G., & Birbaumer, N. (2017). Brain–Computer Interface–Based Communication in the Completely Locked-In State. *PLOS Biology*, *15*(1), e1002593. https://doi.org/10.1371/journal.pbio.1002593

Cichy, R. M., Heinzle, J., & Haynes, J.-D. (2012). Imagery and Perception Share Cortical Representations of Content and Location. *Cerebral Cortex*, *22*(2), 372–380. https://doi.org/10.1093/cercor/bhr106

De Massari, D., Ruf, C. A., Furdea, A., Matuz, T., van der Heiden, L., Halder, S., … Birbaumer, N. (2013). Brain communication in the locked-in state. *Brain*, *136*(6), 1989–2000. https://doi.org/10.1093/brain/awt102

Dumoulin, S. O., & Wandell, B. A. A. (2008). Population receptive field estimates in human visual cortex. *Neuroimage*, *39*(2), 647–660. Retrieved from http://www.sciencedirect.com.ezproxy.ub.unimaas.nl/science/article/pii/S1053811907008269

Emmerling, T. C., Zimmermann, J., Sorger, B., Frost, M. A., & Goebel, R. (2016). Decoding the direction of imagined visual motion using 7T ultra-high field fMRI. *NeuroImage*, *125*, 61–73. https://doi.org/10.1016/j.neuroimage.2015.10.022

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream.

Nature Neuroscience, *14*(9), 1195–1201. https://doi.org/10.1038/nn.2889

Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2004). Brain areas underlying visual mental imagery and visual perception: an fMRI study. *Cognitive Brain Research*, *20*(2), 226–241. https://doi.org/10.1016/j.cogbrainres.2004.02.012

Goebel, R., Esposito, F., & Formisano, E. (2006). Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Human Brain Mapping*, *27*(5), 392–401. https://doi.org/10.1002/HBM.20249

Goebel, R., Khorram-Sefat, D., & Muckli, L. (1998). The constructive nature of vision: direct evidence from functional magnetic resonance imaging studies of apparent motion and motion imagery. *European Journal of*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1046/j.1460-9568.1998.00181.x/full

Harrison, S., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*. Retrieved from http://www.nature.com/nature/journal/v458/n7238/abs/nature07832.html

Holmes, G. (1918). Disturbances of vision by cerebral lesions. *The British Journal of Ophthalmology*, *2*(7), 353. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC513514/

Ishai, A., Ungerleider, L., & Haxby, J. (2000). Distributed neural systems for the generation of visual images. *Neuron*. Retrieved from http://www.sciencedirect.com/science/article/pii/S0896627300001689

Johnson, M. R., & Johnson, M. K. (2014). Decoding individual natural scene representations during perception and imagery. *Frontiers in Human Neuroscience*, *8*, 59. https://doi.org/10.3389/fnhum.2014.00059

Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. Retrieved from http://arxiv.org/abs/1412.6980

Klein, I., Dubois, J., Mangin, J., Kherif, F., & Flandin, G. (2004). Retinotopic organization of visual mental images as revealed by functional magnetic resonance imaging. *Cognitive Brain*. Retrieved from http://www.sciencedirect.com/science/article/pii/S0926641004001934

Kosslyn, S., & Thompson, W. (2003). When is early visual cortex activated during visual mental imagery? *Psychological Bulletin*. Retrieved from http://psycnet.apa.org/psycinfo/2003-99991-004

Kosslyn, S., Thompson, W., & Alpert, N. (1997). Neural systems shared by visual imagery and visual perception: A positron emission tomography study. *Neuroimage*. Retrieved from http://www.sciencedirect.com/science/article/pii/S1053811997902950

Kosslyn, S., Thompson, W., & Ganis, G. (2006). *The case for mental imagery*. Retrieved from https://books.google.nl/books?hl=en&lr=&id=igi-Z_w38CUC&oi=fnd&pg=PR7&dq=the+Case+for+Mental+Imagery+ko

sslyn&ots=JqEDTaDP81&sig=ti9bTcjHaZNyZMUTfsppNtgCuNc

Kriegeskorte, N., & Goebel, R. (2001). An efficient algorithm for topologically correct segmentation of the cortical sheet in anatomical MR volumes. *NeuroImage*, *14*(2), 329–346. Retrieved from http://www.sciencedirect.com/science/article/pii/S1053811901908316

Lee, S., Kravitz, D., & Baker, C. (2012). Disentangling visual imagery and perception of real-world objects. *Neuroimage*. Retrieved from http://www.sciencedirect.com/science/article/pii/S1053811911012195

Marques, J. P., Kober, T., Krueger, G., van der Zwaag, W., Van de Moortele, P.-F., & Gruetter, R. (2010). MP2RAGE, a self bias-field corrected sequence for improved segmentation and T 1-mapping at high field. *Neuroimage*, *49*(2), 1271–1281. Retrieved from http://www.sciencedirect.com/science/article/pii/S1053811909010738

Mechelli, A., Price, C., Friston, K., & Ishai, A. (2004). Where bottom-up meets top-down: neuronal interactions during perception and imagery. *Cerebral Cortex*. Retrieved from http://cercor.oxfordjournals.org/content/14/11/1256.short

Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M., & Morito, Y. (2008). Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron*. Retrieved from http://www.sciencedirect.com/science/article/pii/S0896627308009586

Moeller, S., Yacoub, E., & Olman, C. (2010). Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magnetic*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/mrm.22361/full

Naselaris, T., Olman, C. A., Stansbury, D. E., Ugurbil, K., & Gallant, J. L. (2015). A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *NeuroImage*, *105*, 215–228. https://doi.org/10.1016/j.neuroimage.2014.10.018

Nijboer, F., Sellers, E. W., Mellinger, J., Jordan, M. A., Matuz, T., Furdea, A., … Kübler, A. (2008). A P300-based brain–computer interface for people with amyotrophic lateral sclerosis. *Clinical Neurophysiology*, *119*(8), 1909–1916. https://doi.org/10.1016/j.clinph.2008.03.034

O'Craven, K., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of Cognitive Neuroscience*. Retrieved from http://www.mitpressjournals.org/doi/abs/10.1162/089892900051137549

Pearson, J., Naselaris, T., & Holmes, E. (2015). Mental imagery: functional mechanisms and clinical applications. *Trends in Cognitive*. Retrieved from http://www.sciencedirect.com/science/article/pii/S1364661315001801

Peirce, J. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods*. Retrieved from http://www.sciencedirect.com/science/article/pii/S0165027006005772

Reddy, L., Tsuchiya, N., & Serre, T. (2010). Reading the mind's eye:

decoding category information during mental imagery. *NeuroImage*. Retrieved from http://www.sciencedirect.com/science/article/pii/S1053811909012701

Schoenmakers, S., Barth, M., Heskes, T., & Gerven, M. van. (2013). Linear reconstruction of perceived images from human brain activity. *NeuroImage*. Retrieved from http://www.sciencedirect.com/science/article/pii/S1053811913007994

Senden, M., Reithler, J., Gijsen, S., & Goebel, R. (2014). Evaluating Population Receptive Field Estimation Frameworks in Terms of Robustness and Reproducibility. *PLoS ONE*, *9*(12), e114054. https://doi.org/10.1371/journal.pone.0114054

Slotnick, S., Thompson, W., & Kosslyn, S. (2005). Visual mental imagery induces retinotopically organized activation of early visual areas. *Cerebral Cortex*. Retrieved from http://cercor.oxfordjournals.org/content/15/10/1570.short

Smith, A. T., Singh, K. D., Williams, A. L., & Greenlee, M. W. (2001). Estimating Receptive Field Size from fMRI Data in Human Striate and Extrastriate Visual Cortex. *Cerebral Cortex*, *11*(12), 1182–1190. https://doi.org/10.1093/cercor/11.12.1182

Smith, S., Jenkinson, M., Woolrich, M., & Beckmann, C. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*. Retrieved from http://www.sciencedirect.com/science/article/pii/S1053811904003933

Sorger, B., Reithler, J., Dahmen, B., & Goebel, R. (2012). Report A Real-Time fMRI-Based Spelling Device Immediately Enabling Robust Motor-Independent Communication. *Current Biology*, *22*, 1333–1338. https://doi.org/10.1016/j.cub.2012.05.022

Sperry, R. W. (1963). Chemoaffinity in the orderly growth of nerve fiber patterns and connections. *Proceedings of the National Academy of Sciences*, *50*(4), 703–710. Retrieved from http://www.pnas.org/content/50/4/703.short

Sprague, T. C., Saproo, S., & Serences, J. T. (2015). Visual attention mitigates information loss in small- and large-scale neural codes. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2015.02.005

Stokes, M., Saraiva, A., Rohenkohl, G., & Nobre, A. (2011). Imagery for shapes activates position-invariant representations in human visual cortex. *Neuroimage*. Retrieved from http://www.sciencedirect.com/science/article/pii/S1053811911002485

Stokes, M., Thompson, R., & Cusack, R. (2009). Top-down activation of shape-specific population codes in visual cortex during mental imagery. *Journal of*. Retrieved from http://www.jneurosci.org/content/29/5/1565.short

Thirion, B., Duchesnay, E., Hubbard, E., Dubois, J., Poline, J.-B., Lebihan, D., & Dehaene, S. (2006). Inverse retinotopy: inferring the visual content of images from brain activation patterns. *Neuroimage*, *33*(4), 1104–1116. Retrieved from

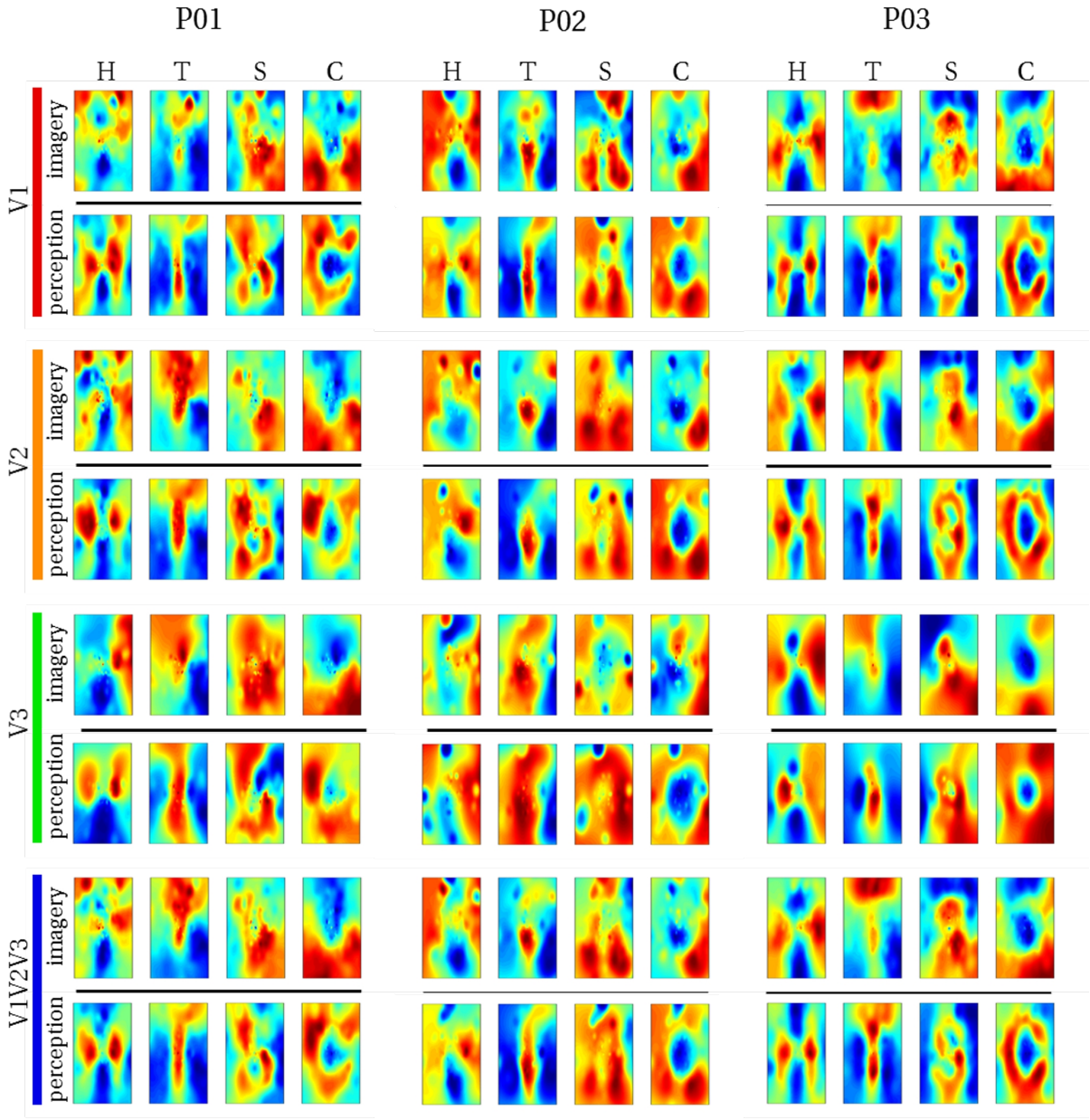
http://www.sciencedirect.com/science/article/pii/S1053811906007373

Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P.-A. (2008). Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning - ICML '08* (pp. 1096–1103). New York, New York, USA: ACM Press. https://doi.org/10.1145/1390156.1390294

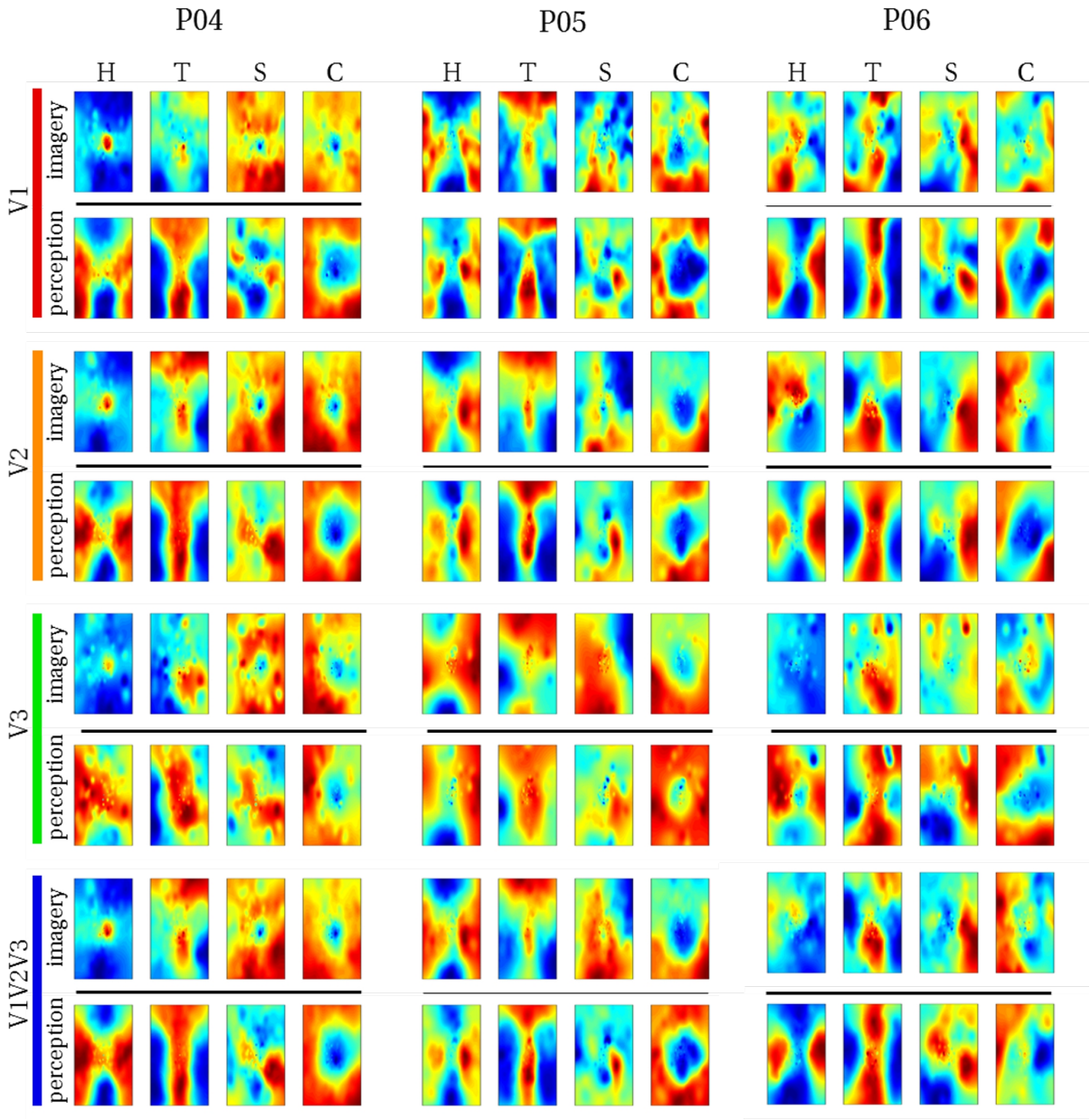
Wandell, B. A., & Winawer, J. (2015). Computational neuroimaging and population receptive fields. *Trends in Cognitive Sciences*. Retrieved from http://www.sciencedirect.com/science/article/pii/S1364661315000704

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain–computer interfaces for communication and control. *Clinical Neurophysiology*, *113*(6), 767–791. https://doi.org/10.1016/S1388-2457(02)00057-3

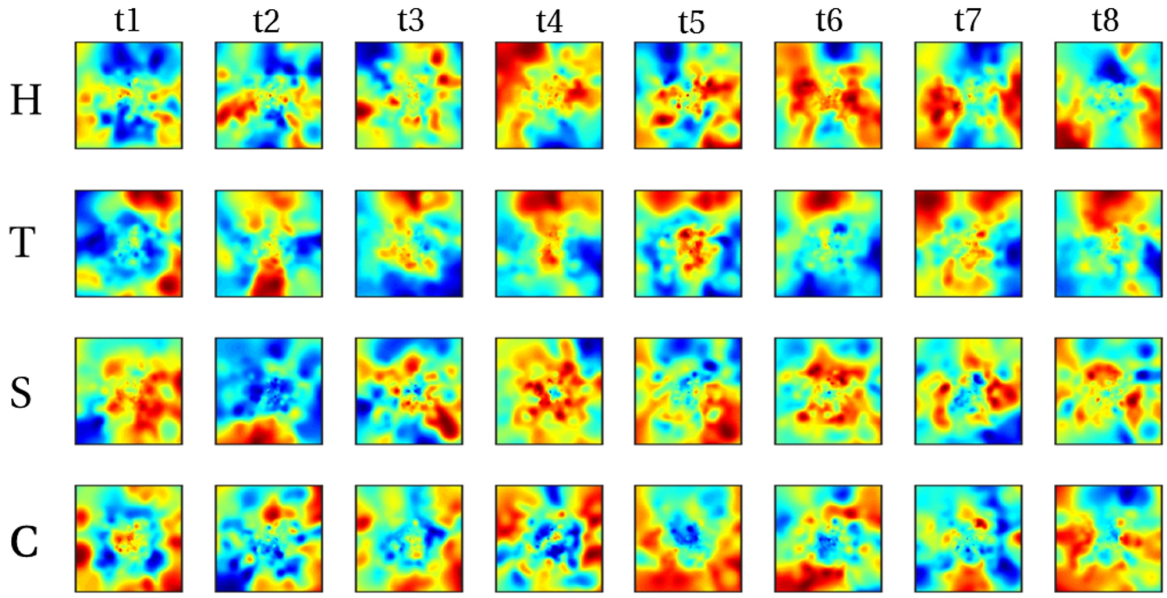
Supplementary Figures



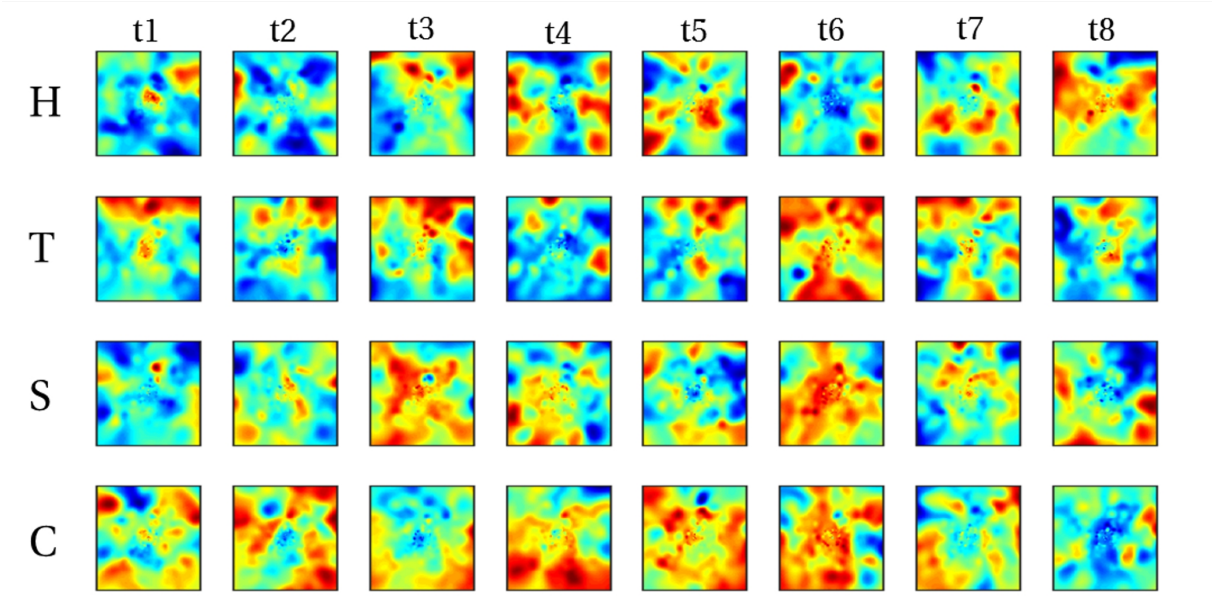
Supplementary Figure 1: Reconstructed visual field images (participants 1-3). Reconstructed average VFIs are visualized for each ROI of participants one, two, and three. Reconstructions of the remaining three subjects are shown in supplementary figure 2. Perceptual and imagery voxel patterns were obtained from the raw BOLD time-series.



Supplementary Figure 2: Reconstructed visual field images (participants 4-6).
Reconstructed average VFIs are visualized for each ROI of participants four, five, and six. Reconstructions of the remaining three subjects are shown in supplementary figure 1. Perceptual and imagery voxel patterns were obtained from the raw BOLD time-series.



Supplementary Figure 3: Reconstructed visual field images from single trials in a single run of participant 3. Each run comprised of 8 trials (columns) per letter (rows).



Supplementary Figure 4: Reconstructed visual field images from single trials in a single run of participant 5. Each run comprised of 8 trials (columns) per letter (rows).