Supplementary Information for

Laminar Specific fMRI Reveals Directed Interactions in Distributed Networks During Language Processing

Daniel Sharoh, Tim van Mourik, Lauren J. Bains, Katrien Segaert, Kirsten Weber, Peter Hagoort, David G. Norris

Correspondence to Peter Hagoort
E-mail: peter.hagoort@mpi.nl

This PDF file includes:

Supplementary text
Figs. S1 to S5
Captions for Databases S1 to S2
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Databases S1 to S2

Supporting Information Text

Methods and materials

Experimental design.

Task paradigm. Twenty-four native Dutch subjects (13 female, 18-30 years of age, 21 right handed) performed a single word reading task which presented words, pseudo-words and false-font items as task conditions. Two subjects were ultimately excluded from analysis owing to computer failure and signal dropout, leaving 22 data sets for analysis. Data for one subject was unrecoverable following computer failure during image reconstruction. The second subject experienced signal dropout in the left occipitotemporal sulcus and so was excluded from all analysis. Subjects had normal or corrected to normal vision and were screened for reading impairment. Left handed participants were included because regions of interest were determined through functional localization. Language function was observed to be left lateralized in all participants. Informed consent for all experimental procedures was obtained in accordance with the procedures of ethical approval of the Donders Centre for Cognitive Neuroimaging and the Erwin L. Hahn Institute.

Initially, the experiment was intended to use a 3×2 task design of Lexicality (words, pseudo-words, false-font items) $\times$ Length (short, long) with 120 items for each level, and where 'short' and 'long' designated the length of the words in terms of number of syllables. The length manipulation was intended to vary the bottom-up signal contribution to the LOTS. It was determined through piloting that the length manipulation was ineffective and was not analyzed as part of this study. The number of participants was chosen on the basis of previous work using similar acquisition techniques (1). Our task thus included three relevant conditions, two of which (words, pseudo-words) were conditions of interest for our analysis, and one of which (false-font items) was used to localize the functional Region Of Interest (fROI) for analysis. Figure S5 visualizes mini-blocks from each condition.

Item creation. Word items were selected from a list of high frequency, concrete Dutch nouns taken from the Celex database (2). Words were selected to maximize frequency, minimize the standard deviation of word frequency, and minimize standard deviation of these values for short and long items.

Pseudo-words were generated using Wuggy (3). Pseudo-word generation was constrained on the basis of phonemic neighborhood density, consonant/vowel structure and the number of characters of the word items. The word and pseudo-word stimuli, and the parameters used in Wuggy to generate the pseudo-word stimuli can be found in supplementary materials (StimulusList.xlsx).

False-font items were created by rendering the word items in the false font. The false font (4) was designed to preserve the low level features of familiar orthographically legal characters, but to be visually different from letter shapes. These items are included in supplementary materials (FalseFontItems.pdf). Collapsing across the length manipulation, there were 240 items of each stimulus type. Sample items can be seen in the main text.

Stimulus presentation. Items were presented during fMRI measurements taken over 12 runs. Runs were delimited by breaks in data acquisition. Twenty items of each stimulus type were presented per run, with 60 items presented in total per run.

Individual stimuli were visually presented for 800ms in the center of the display. One item was presented per trial. Items were rendered in white on a black background, as shown in the main text.

Presentation onset was jittered around the 3960 ms TR based on the design optimization calculations obtained using optseq (5). A black screen was presented for the remaining 3160 ms of each trial. Stimuli were presented minimally with a 200 ms delay from the trial onset and offset resulting in a minimum ISI of 400ms and a maximum ISI of 5920 ms. In practice, ISIs fell between these two extremes. A fixation cross was presented for a full trial (3960 ms) at the beginning of each run and following each mini-block.

Stimuli were presented in 5 item mini-blocks in which all 5 items were of the same condition type. Each mini-block was followed by a fixation cross presented for the duration of one trial. On three pseudo-random occasions per run, a question mark was presented that instructed the participant to indicate via button-box whether the previous mini-block contained existing Dutch words. Button-box responses were not analyzed and were considered only to ensure participant compliance. Prior to the experiment, subjects were briefed on the type of items they were to see and instructed to silently read the items on the screen.

The experiment was performed using Presentation ® software (Version 16.1, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). Two versions of the experiment were created with roughly half of the participants assigned to each version. The versions differed in block and item order. The different experiment versions were intended to capture latent, unintended effects inherent in presentation order or other version specific properties.

Task design model. The task design matrix included condition regressors, temporal and spatial dispersion derivatives, physiologic regressors, motion regressors produced by using SPM version 12 (6), drift terms, frequency filters, outlier censors, and constant terms modeling the mean signal per run. Outlier time points were determined using *3dToutcount* in AFNI (7). We considered a voxel to be an outlier if the probability of the distance of its intensity value from the trend exceeded $p = 0.001$ as defined by its location within a Gaussian probability distribution. Time points were excluded from analysis if 2% or more of the voxels at that time point were categorized as outliers.

Stimulus onsets were modeled as instantaneous events with zero duration and convolved with the canonical hemodynamic response function. Condition regressors were created separately for word, pseudo-word and false-font items; and for the long

and short items within each condition. In total six condition types were modeled, but only three were analyzed after collapsing across length.

Acquisition.

Functional acquisition. Near whole brain, submillimeter (0.943, 0.900 slice direction) resolution T2*-weighted GE-BOLD data were acquired using a GRAPPA accelerated (acceleration factor 8 × 1) 3D-EPI acquisition protocol (8) with CAIPI shift kz = 0, ky = 4 (9, 10); effective TE = 20ms, TR = 44ms, effective TR = 3960ms, BW = 1044Hz/Px, FoV = 215mm × 215mm × 215mm with 112 phase encode steps in the slice direction (100.8mm), $\alpha = 13°$, partial Fourier factor=6/8 in both slice and phase-encoding directions. The first phase encoding gradient was applied in the posterior to anterior direction. An axial slab was collected in each subject and positioned to include the occipitotemporal sulcus. The 10cm slab was sufficient to allow complete brain coverage in several subjects and near complete coverage in the remaining subjects. Data were acquired on a Siemens Magnetom 7 Tesla scanner (Siemens Healthineers, Erlangen, Germany) with a 32-channel head coil (Nova Medical, Wilmington, USA) at The Erwin L. Hahn Institute in Essen, Germany. Functional data consisted per subject of 12 3D-EPI data sets of 77 volumes each, although some sessions were incomplete owing to time constraints or other difficulties. No session contained fewer than 10 functional data sets.

Anatomy acquisition. Two anatomic images were acquired in each subject using the MP2RAGE (11) acquisition protocol (voxel resolution = 0.75mm × 0.75mm × 0.75mm, TR = 6000ms, TE = 3.06ms, $T1_1$ = 800ms, $T1_2$ = 2700ms, $\alpha_1 = 4°$, $\alpha_2 = 5°$, BW = 240Hz/Px, FoV = 240mm × 240mm with 192 slices (144mm)) and a T1-weighted inversion recovery EPI (IR-EPI) protocol based on the parameters used in the functional acquisition protocol. To create the T1 contrast, the following parameters were modified from the functional acquisition: $\alpha = 90°$, T1 = 800ms, TR = 200ms, TE = 20ms. Example images from each type of anatomic acquisition can be seen in figure S1.

It was necessary to increase the number of phase encode steps in the slice direction from 112 to 160 and to expand the FoV in the slice direction to ensure fully overlapping coverage with the functional data. The IR-EPI images were used for image registration as they are known to provide high tissue contrast while preserving the geometric distortions of the functional images (12). High accuracy, cross-modal registration is challenging, particularly with high resolution acquisitions known to exaggerate geometric distortions. Performing co-registration taking the IR-EPI as the source image mitigated the challenges caused by these distortions.

Field maps were also acquired in some subjects for potential use in distortion correction, though these were not used. The IR-EPI provided sufficient contrast in the native functional space to facilitate high quality registration without the need for distortion correction.

Choice of acquisition method for laminar resolution imaging. lfMRI relies on the same neurovascular coupling mechanisms exploited in standard BOLD imaging. In recent years there has been increased support for the concept that the neurovascular coupling occurs at a sufficiently fine scale to make lfMRI feasible (13). However, the requirement for submillimeter resolution has led to considerable discussion as to the best MR-contrast for interrogating the hemodynamic response.

The standard gradient echo BOLD contrast is highly sensitive to functional activation, but is known to have a considerable contribution from vessels downstream from the site of activation. The less commonly used spin-echo BOLD sequence only acquires data from a subset of the contrast mechanisms that contribute to gradient echo BOLD, but is believed to have a superior intrinsic spatial localization at high static magnetic field strengths (14–16). In addition, contrasts based on cerebral blood flow and volume (CBF, CBV) should also be considered. It is technically far easier to test and compare these contrasts in animal models, and historically such experiments largely preceded human lfMRI (17–24). The conclusion drawn from these was that CBV was consistently found to have the superior characteristics in terms of spatial resolution, and gradient echo BOLD the poorest. Spin-echo BOLD and CBF are somewhere between these two extremes. This hierarchy may be explained in terms of the current view that blood volume changes occur in the arterioles and capillaries (19, 25–27), and hence CBV contrast should not be a downstream contrast as is BOLD.

The first laminar fMRI studies in humans are comparatively recent (28–30), and utilized gradient echo BOLD contrast. Since then, the VASO technique for measuring CBV noninvasively (21) has been further developed for application for laminar fMRI at high static magnetic field strengths (31–33), and a number of spin echo (34–36) and combined spin-echo and gradient echo studies (37, 38) have been performed. Laminar CBF has to date not been published for human studies. Our rationale for selecting gradient echo BOLD for the current study was based primarily on its exclusive ability to acquire high spatial resolution data from large volumes within an acceptable acquisition time. The two main alternatives – CBV and spin-echo – are currently techniques that are restricted in their volume coverage and suffer from comparatively long acquisition times (33, 36).

The whole-brain gPPI results we report suggest, however, that GE-BOLD may be capable of more refined spatial localization than previously believed. As discussed in the main text, the ability of the gPPI to account for the task effects likely enhanced our ability to localize signal variance unique to individual depth bins. Simulations from Markuerkiaga et al. (39) based on reported depth dependent responses in visual cortex to identified a depth dependent peak to tail response ratio of at least 5:1 in all cortical depths at 7T, which would reduce the detectability of unique variance downstream from its source. The gPPI results suggest that this ratio may be conservative, or perhaps influenced by task properties. Our reading experiment presented stimuli at a high frequency relative to presentation rates discussed in Markuerkiaga et al. (39), which should have produced relatively higher frequency task signal. The vasculature attributed to downstream BOLD effects consists of post capillary vessels draining into larger vessels, whereby differences in the vessel length and flow velocity will act to reduce the coherence of

the signal leading the vasculature bed to act as a low-pass filter. High frequency task signal components would therefore be expected to undergo greater attenuation than lower frequency components, and would experience a larger depth dependent peak-to-tail response ratio. In light of our results, it seems clear that unique variance related to each bin was well localized, and that signal contamination was isolated to the main task effects where it could be removed during the gPPI analysis.

Image registration. High quality registration is critical to laminar fMRI. Given the complications inherent in the registration of submillimeter data, different combinations of tools were necessary to achieve accurate registrations in different participants. The criteria for success were constant, however, across all participants and of an entirely anatomical basis. Alignment quality was determined by visual inspection of brain edges and the left occipitotemporal sulcus. The registration procedure is described in this section.

Motion correction. Image realignment was performed using *spm_realign* from SPM 12 (6), with parameter values set to achieve the highest quality registration. During this step, a mean functional image was computed to be used as the base image in cross-modal registration.

Skull removal. Skull removal was performed on functional and anatomic data prior to cross-modal image registration. Different skull removal procedures were used depending on image modality. The FreeSurfer (40) watershed function was applied to the IR-EPI data sets, sometimes following a first pass B1 bias field correction (discussed in *B1 Correction*). Nearly all processed brains required manual intervention to remove voxels containing unwanted skull or tissue, or to reintroduce voxels removed in error. Skull removal was performed on all MP2RAGE images in the same manner.

Skull removal was performed on the mean functional images produced during realignment. In this procedure, we manually edited the result of AFNI's *3dAutomask* program. *3dAutomask* is typically used to remove the skull in images with poor tissue contrast, such as with T2*-weighted images. Parameters for this program were optimized on a per subject basis, and all results were manually edited to ensure that only voxels containing skull were removed. We found these results to be adequate on the basis of visual inspection following manual intervention, where 'adequate' describes results which did not contain residual skull or exclude voxels containing brain-matter. *3dAutomask* parameters were iteratively optimized until a result was obtained which reasonably limited the necessary manual intervention.

B1 correction. B1 correction on the IR-EPI data was unsuccessful on 5 data sets using the standard tools available in the FreeSurfer suite, resulting in failed skull removal and inaccurate segmentations. We were able reduce B1 inhomogeneity by applying an additional B1 correction before applying FreeSurfer tools. Our approach was to calculate a first-pass transform for the mean functional and IR-EPI images (with the skull) and apply the transformation to the mean functional image. We then utilized the B1 bias captured in the mean functional image to correct the IR-EPI anatomic images. Following the initial coregistration, the mean functional image was smoothed and voxel-wise intensity scaled between $0.3v$ and $0.9v$ of its intensity value v to prevent extreme values from unduly influencing bias correction. The IR-EPI was then divided by the scaled image, and the result was taken as the corrected image. The corrected image could then be coregistered to the original mean functional image and used as the input dataset for the standard FreeSurfer processing pipeline. We observed a marked improvement in both the coregistration results and the results of the FreeSurfer segmentation and surface generation after performing this correction (figure S2). Computation time was drastically reduced as well, in some cases up to 15 hours.

Coregistration. Within-subject coregistration was performed using the skull-removed mean functional image and the skull-removed IR-EPI image. Using this image set mitigated registration error owing to image distortion typically observed in EPI acquisitions. High quality coregistration was crucial to the laminar analysis featured in this experiment, as the accurate definition of tissue boundaries in functional space follows only from a highly accurate coregistration of the structural and functional images. Note also that the transformation computed in this step was applied to the structural image to avoid introducing interpolation errors.

Several coregistration tools were used to calculate optimal image alignment. For a given subject, multiple transforms were calculated and visually inspected. The best alignment as determined by visual inspection was taken for further analysis. Volume coregistration was performed using FreeSurfer's robust, outlier insensitive registration cost function as implemented in *mri_robust_register*. If the resulting transformation resulted in poor registration, we then used the NMI cost function implemented in *mri_robust_register* and finally the NMI cost function implemented in AFNI's *3dAllineate*. If necessary, manual improvements were applied to the best transformations generated by these tools. Registration quality was assessed by visual inspection of alignment along the left occipitotemporal sulcus and brain edges.

In 11 subjects, failure to reconstruct surfaces from the IR-EPI image made it necessary to perform surface reconstruction on MP2RAGE data, and therefore to bring the MP2RAGE surfaces into register with the functional data.. Following the initial coregistration described above for the IR-EPI images, the IR-EPI images were generally in good alignment with the task data and could serve as the source image for this purpose. In this case, MP2RAGE surfaces were aligned to the IR-EPI volumes using FreeSurfer's boundary based registration program. If the IR-EPI volume was not in good alignment with the functional data, an initial alignment between the MP2RAGE and functional data was first computed using the tools described in the previous paragraph before performing the boundary based registration. The boundary based registration procedure is described below in a dedicated section.

Normalization for group analysis. Functional data for each subject were mapped into MNI space for use in the whole-brain gPPI group analysis. The skull-removed MP2RAGE image was first brought into alignment with the skull-removed mean functional image. After inspecting the quality of the registration, the functionally aligned MP2RAGE images were aligned to the MNI128 template available in standard FreeSurfer installations. This transformation was concatenated with that obtained from the inverted matrix from the initial coregistration and applied to the motion corrected functional data. The result of this spatial normalization was MNI mapped functional data for each subject.

Tissue segmentation and surface generation. Tissue segmentation was performed in FreeSurfer using the skull-removed IR-EPI image. Failures to properly reconstruct subject surfaces were addressed by inserting control points, applying additional normalization as described in a previous section, or disabling the correction of defects in surface topology if they did not occur in experiment critical regions. The IR-EPI images commonly included artifacts in noncritical locations that would result in discontiguities in the surface and unsuccessful surface generation. As these defects did not often occur near LOTS, it was possible to generate accurate surfaces even after bypassing correction. If surface reconstruction failed following these interventions, the MP2RAGE dataset was used in place of the IR-EPI, and additional registration steps were applied (discussed below).

Boundary based registration. Surfaces reconstructed from the IR-EPI image did not require additional alignment to the functional data beyond resampling the FreeSurfer generated surfaces from "conformed space" to functional space. "Conformed space," native to FreeSurfer, is a 1mm isotropic 256^3 grid in the RAS coordinate system.

Surfaces reconstructed from MP2RAGE images underwent an additional registration step using *bbregister*, FreeSurfer's boundary based registration (BBR) tool. The goal of this procedure was to produce surfaces in register with the functional data. As the IR-EPI and MP2RAGE data sets were generally well aligned from the coregistration procedure described previously, the main purpose of the BBR was to find a solution accommodating the distortions affecting surface placement along the fROI. Using a boundary based cost function, IR-EPI images that were unable to be used for surface generation were aligned with the boundaries generated from the MP2RAGE images. The inverse of this transformation was then applied to the surfaces to align the boundaries to the IR-EPI image. If the alignment generated by the boundary based registration procedure was found to be inaccurate, simple solutions to improve accuracy involved optimizing the registration for the fROI through a weighting mask, manual intervention, or improving the alignment of the two images prior to the boundary based registration. Failing a simple solution, we also computed a nonlinear boundary based registration (41). In this approach, the registration algorithm recursively divided and aligned surface segments to increase registration accuracy.

The importance of highly accurate image alignment in laminar resolution imaging cannot be overstated. In the present work, registration inaccuracies in excess of 1mm had the potential to displace entire bins, leading to meaningless results. Great care was taken to ensure accurate registrations and alignment of the surfaces with the functional images. As in the other registration procedures, registration quality was assessed only through visual inspection of key anatomy.

Equivolume contouring. The gray matter volume of each subject was partitioned into equivolume bins using the OpenFmri (https://github.com/TimVanMourik/OpenFmriAnalysis) implementation of the equivolume contouring approach described in Waehnert et al. (42). The equivolume method increases the likelihood that the histological profile of each bin is consistent throughout the given region.

We partitioned the gray matter volumes into 3 bins: the smallest number of bins which allowed for the dissociation of the deep, middle and superficial contributions to the overall BOLD signal. For the purpose of the spatial GLM, it was necessary to include two additional non-cortical bins representing white matter and CSF volumes respectively. The inclusion of these additional bins was due to partial volume effects caused by voxels extending outside the cortical strip. Voxels observed within these boundaries were assigned a value representing the fractional volume observed within a particular set of boundaries.

The ultimate output of this procedure was a 4D dataset whose first 3 dimensions represented spatial coordinates and whose 4th dimension represented the different bins. Incrementing over the 4th dimension indices gave the fractional volume of each voxel found in that particular bin. This volume is referred to as the layer-volume distribution.

In our partition scheme, the volume subsuming the six histological layers was partitioned into three bins. We argue that the histologically coarse bins were sufficient to dissociate top-down and bottom-up signal contributions. At the mesoarchitectural level of lfMRI, it is not practical to measure individual histological layers. One common approach to this challenge has been to consider a simplified model of layer interactions which merges supragranular (layers I,II,III), granular (layer IV) and infragranular (layers V,VI) histological layers into three logical layers based on shared connection tendencies (1, 28, 37, 43, 44). This model is based in large part on patterns of laminar connectivity discovered in Rockland & Pandya (45) when exploring the link between anatomy and functional hierarchy, and schematized in the Felleman and Van Essen hierarchy (46). The simplified laminar model has proven valuable when constrained by functional data, and has informed efforts in lfMRI.

Physiologic noise removal. Cardiac and respiration data were collected concurrently with the functional data using a pulse oximeter and pneumatic belt. Physiologic regressor estimation up to the 6th (cardiac) and 8th (respiration) order was performed using a modified version of the PhysIO toolbox of the TAPAS suite (47). These minor modifications were necessary to account for unique log file formats produced by the equipment at the scan site. Regressors were then included in the design as nuisance regressors. Regressor quality was assessed with a partial F-test.

Non-laminar analysis. The task design was fitted using the generalized least squares regression implemented in the OpenFmri analysis suite. First level T-statistics were calculated in MATLAB version R2014B (The Mathworks Inc.) for condition and contrast effects in each subject individually. Several versions of the first level analysis were performed to calculate parameter fits in both native space and MNI space, and with different levels of spatial smoothing applied. This was necessary owing to requirements of the different analyses reported in this work.

Native space data with were spatially smoothed with a 1mm and 4mm Gaussian kernel. Data smoothed with the 4mm kernel were mapped to MNI space and used in the whole-brain gPPI analysis. Data smoothed with the 1mm kernel were analyzed in the first level GLM used to identify the fROIs. Following fROI identification, the fROIs identified in each subject were resampled to native resolution and used as an inclusive mask of the original resolution, native space functional data. The voxels included in this mask then underwent depth-dependent signal extraction.

Procedure to define functional region of interest. Anatomically, the region of interest was located proximal to the fundus of the occipitotemporal sulcus. It was functionally defined as a cluster of voxels which responded to visually presented words and pseudo-words, but preferentially to pseudo-words. In addition, this region is known to express reduced BOLD amplitude to false-font items compared to items composed of orthographically legal characters (48).

The region was defined in each subject through a series of masking operations implemented with AFNI's voxel-wise dataset calculator *3dcalc*.These operations were performed on the t-statistics from the 1mm smoothed, native space analysis. First, voxels were removed if they did not reach threshold in both the word and pseudo-word conditions. This was defined as a t-statistic where $1 \leq t \leq 3$. T was initially set to $t = 2$ and was increased or reduced if the number of surviving voxels fell outside of the desired range (see below). We then removed all voxels with a larger t-statistic for the false-font condition than for either the word or pseudo-word conditions. Finally, voxels were excluded if the difference between t-statistics of words and pseudo-words was larger than the original activation threshold. Clusters were considered for inclusion if they were 1) located within the extent of the occipitotemporal sulcus, if 2) cluster size was between 100 and 400 voxels, if 3) 30-50% of the voxels responded preferentially to the word condition over the pseudo-word condition, and if 4) the total response was comparable between the word and pseudo-word preferred voxels when considering the proportion of voxels preferring each condition. These criteria were selected to isolate a functional region which is known to respond preferentially to both words and pseudo-words compared to false-font items, prefer pseudo-words to words, and contain a mixture of individual voxels which prefer each condition. The fROI selection procedure was biased by design toward pseudo-word activation because stronger pseudo-word activation is a functional feature of the region (48).

Further considerations were made with respect to the importance of cluster contiguity. Given the high spatial resolution of our data, it was possible to distinguish populations of active voxels spanning the CSF boundary bridging the occipitotemporal sulcus. Following the removal of the voxels located in CSF in some subjects, formerly contiguous clusters became distinguishable. We determined that the most reasonable approach was to include formerly contiguous voxels in the laminar analysis. Given that this region is often functionally defined and generally identified near the fundus of the occipitotemporal sulcus, partial volume effects have almost certainly influenced fMRI measurements at standard resolutions. The decision to exclude populations of voxels stranded on either side of the chasm would have proven arbitrary in that non-laminar studies investigating this region typically lack the resolution to distinguish fusiform and inferior temporal populations. We concluded that allowing for discontiguities in the left OTS fROI more faithfully adhered to the literature definition of the region than an ad hoc justification for voxel removal. Native space functional ROIs for all participants can be seen in figures S3 and S4.

Depth dependent signal extraction. The fROI produced through the procedure described above was used to mask the layer-volume distribution. This resulted in a layer-volume distribution specific to the fROI. By treating this distribution as a design matrix such that rows were voxels and columns were bins, it was possible to regress it against the signal observed in the fROI for each time point in the experiment (49, 50).

Fitting the voxel-volume distribution to each time point in the experiment yielded the relative contribution of each bin to the overall signal at each time point, thereby representing a depth dependent time-series for each depth bin. These were treated similarly to voxel time-courses and used to fit the task design.

The task design model was then fitted to the extracted depth dependent time courses. Percent signal change was calculated as a division between β-weights assigned to each condition and the average weight assigned to the constant terms. The percent signal change values were then analyzed at the group level in an ANOVA and subsequent two-tailed, paired t-test comparing the responses to words and pseudo-words. T-statistics and ANOVA results were determined to be significant at $p = 0.01$. These analyses were implemented in MATLAB.

Statistical analysis.

Intraregional gPPI. The gPPI analysis is a generalized version of the PPI analysis. In the generalized form, the analysis is designed to span the full experiment space (51). In gPPI analysis, the first level model is extended by including the time-course of a seed region in addition to interaction terms of the seed region with each task condition regressor.

As the goal of this analysis is to observe the effect of the interaction between the task and the neuronal response of a seed region, a deconvolution is typically applied to the seed time-course before computing the interaction term. Given the problems associated with deconvolution (52) and the novel nature of this work, we omitted deconvolution from our gPPI analysis. The depth dependent hemodynamic response function (HRF) is not well understood. In the absence of this knowledge, the task

regressors used in this experiment were created using the canonical HRF. Deconvolution based on the canonical HRF would have therefore exacerbated errors in modeling introduced by the initial convolution. Furthermore, the design of this study sequentially presented five items of each condition type, essentially in 20 second blocks. O'Reilly et al. (52) have shown that omitting deconvolution is not expected to greatly affect the outcome in block designs such as that used in the present study. We therefore considered it more prudent to omit rather than include deconvolution in this analysis. There is no known precedent for laminar specific gPPI. The gPPI design was created by adding seed region time-courses and interaction terms to the original design. Interaction terms were calculated as the product of the detrended depth dependent time-series and binary condition vectors (1 when a condition response was expected, 0 when it was not) derived from the task regressors. A time point was included in the interaction term if the task regressor diverged from 0 by 0.0001.

Different models were created for each inter-regional analysis to assess the interaction between two depth- bins while alternating seed/target assignment. We did not model the third remaining bin. Six models were created in total, each containing six interaction terms (each of the six conditions multiplied by the seed-region time-course), the seed-region itself, and the full design as discussed previously. Group effects were assessed using AFNI's *3dANOVA3*. Paired two-tailed t-statistics were computed on the word and pseudo-word condition contrast. Results were deemed significant at $p = 0.01$.

Bin to whole brain gPPI. In a separate gPPI analysis, we modeled the task-dependent effect of the deep and middle bins on the whole brain. The superficial bin was excluded from the whole brain gPPI for several reasons. The LOTS is hypothesized to connect to left temporal cortex through either primarily bottom-up or top-down configurations (48). To distinguish among these and thus explore the ability of lfMRI to distinguish between top-down from bottom-up network arrangements, it was necessary only to include the predicted top-down and bottom-up bins associated with word reading. Voxels within the superficial bin are also susceptible to partial volume artifacts due to vessels on the pial surface, which could possibly affect the analysis. The inclusion of the superficial bin would have required the gPPI model to include ten additional regression terms which would have been collinear with the experimentally interesting deep and middle bin terms, and so this was not considered further in the interests of a parsimonious data analysis.

The analysis was performed on the MNI normalized data with a 4mm Gaussian smoothing kernel applied. Alignment quality was assessed partially on the basis of alignment accuracy of the middle temporal gyrus. Individual subject t-statistic maps were used in conjunction with subject anatomy and the MNI template used for normalization to determine the alignment quality of task critical regions. Inaccuracies in subject registrations were addressed with a manually created, secondary transformation containing small translations intended to improve task critical region alignment without introducing large global inaccuracies. As the experimental question related to the depth dependent connectivity to regions that respond to the word/pseudo-word contrast, the use of the non-laminar first level maps to facilitate alignment was independent of the gPPI analysis. Group results were assessed using AFNI's *3dANOVA3*, as in the previous section.

The parameters of the spatial AutoCorrelation Function (ACF) representing the smoothness of the data were computed using AFNI's *3dFWHMx* on the residual time-series of first level analysis. The ACF parameters were used by the AFNI program *3dClustSim* to compute the likelihood of random clusters given the ACF parameters 0.5815, 2.9134 and 8.3001, in a volume of the dimensions $63 \times 82 \times 55$ with 2mm isotropic voxels. The dimensions of the volume used for permutation testing were determined with a group level functional data mask. Clusters were deemed significant at $p_{\text{uncorr}} = 0.001$, $\alpha = 0.05$.

The task-independent connectivity from the deep and middle bins was assessed using the deep and middle bin time-courses included in the gPPI model. The final results were visualized using the rendering plugin in AFNI. In addition to subjects 2 and 19, subject 4 was excluded from this analysis. We were unable to successfully bring subject 4 into MNI space. Large inaccuracies in the registration resulted in the exclusion of this subject from the whole brain analysis. This subject was included in all analyses performed in native space.

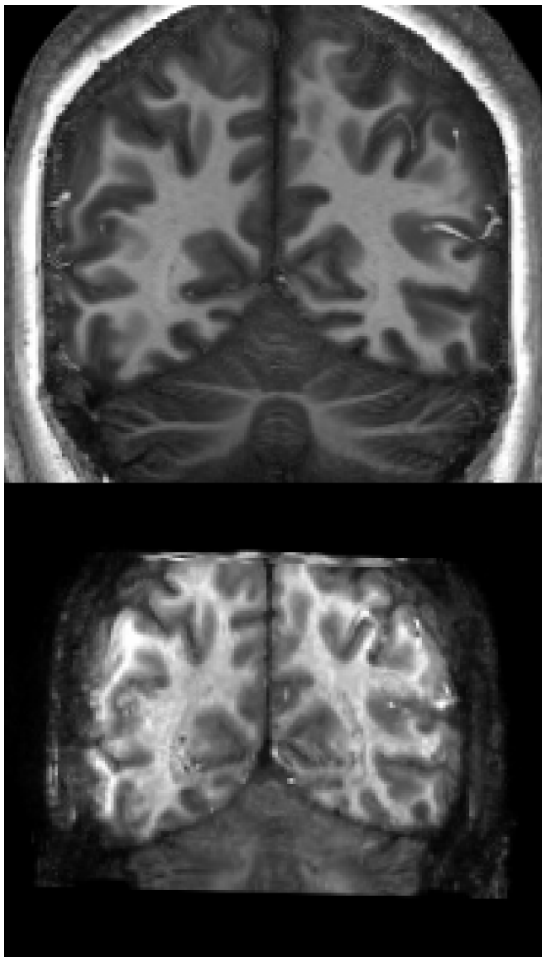

Fig. S1. Different anatomy acquisitions used in this experiment. MP2RAGE (top), IR-EPI (bottom). The left hemisphere is shown on the right side of the images.

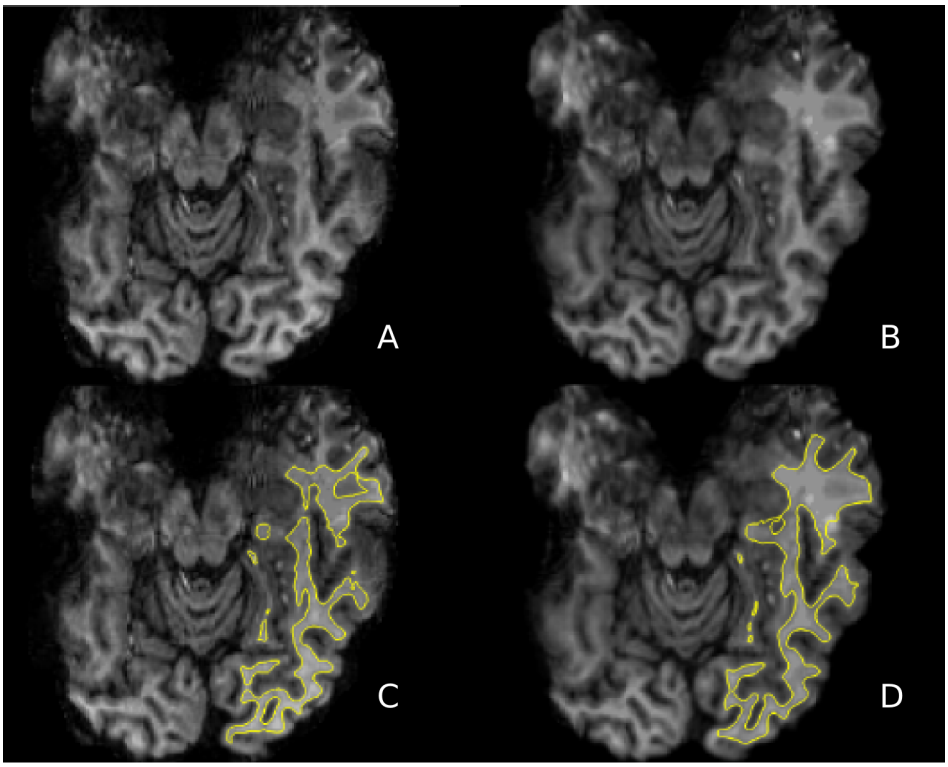

Fig. S2. Comparison of FreeSurfer white matter surface generation on the left hemisphere before and after supplemental B1 correction. (A) uncorrected; (B) corrected; (C) surfaces generated from A; (D) surfaces generated from B. The left hemisphere is shown on the right side of the image.

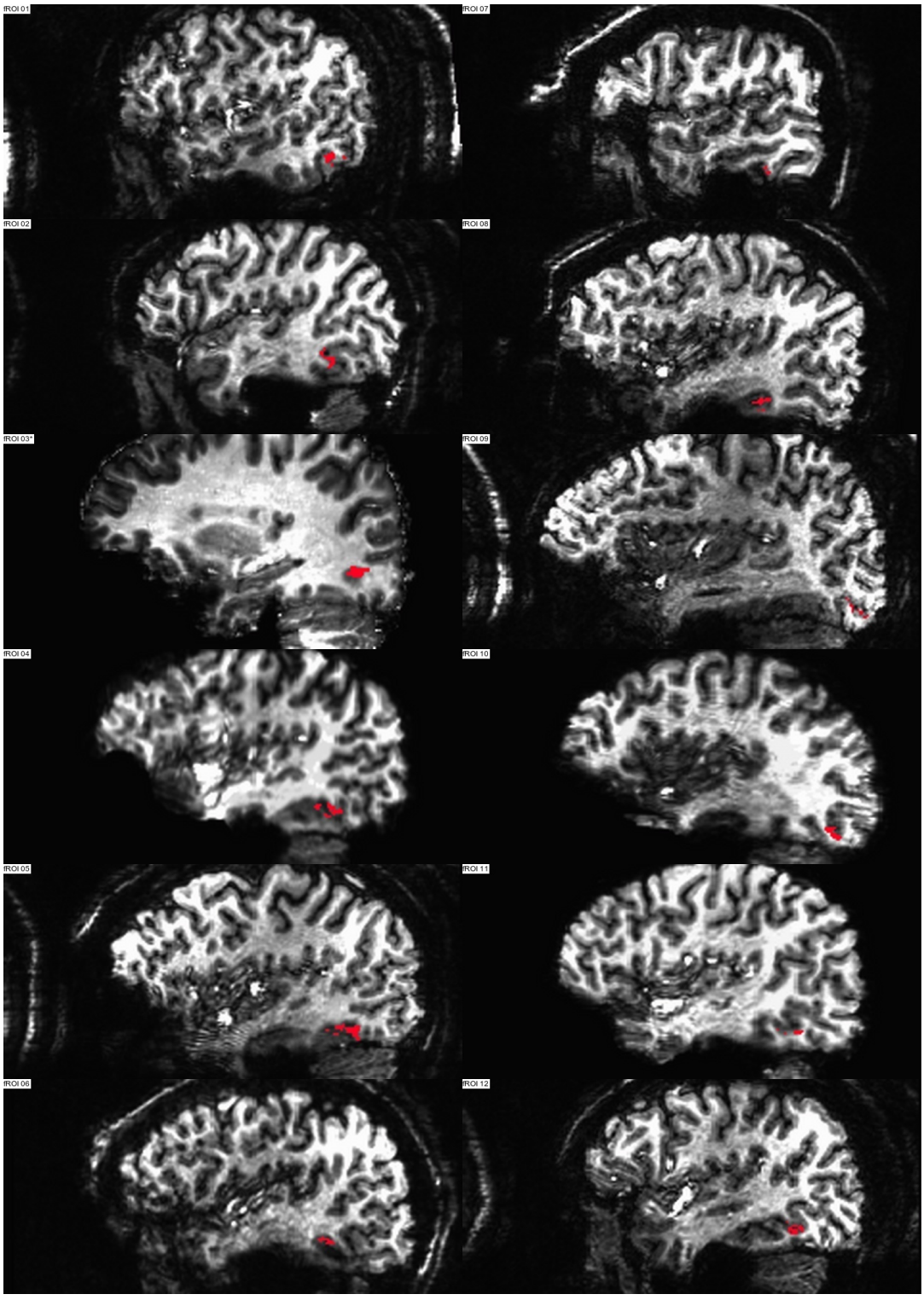

Fig. S3. Native space fROIs 1-12
∗Subject 03 excluded from whole-brain gPPI analysis

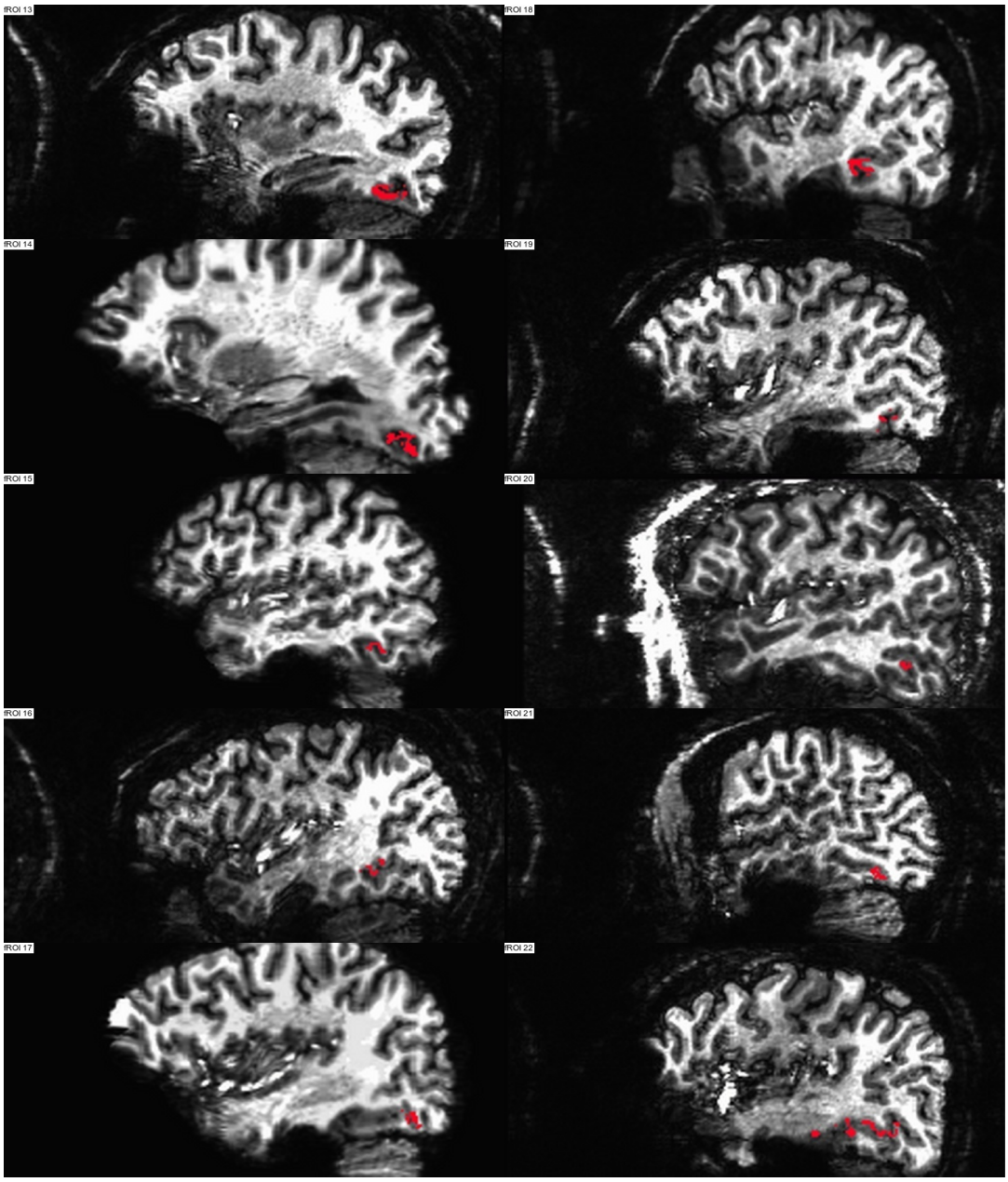

Fig. S4. Native pace fROIs 13-22

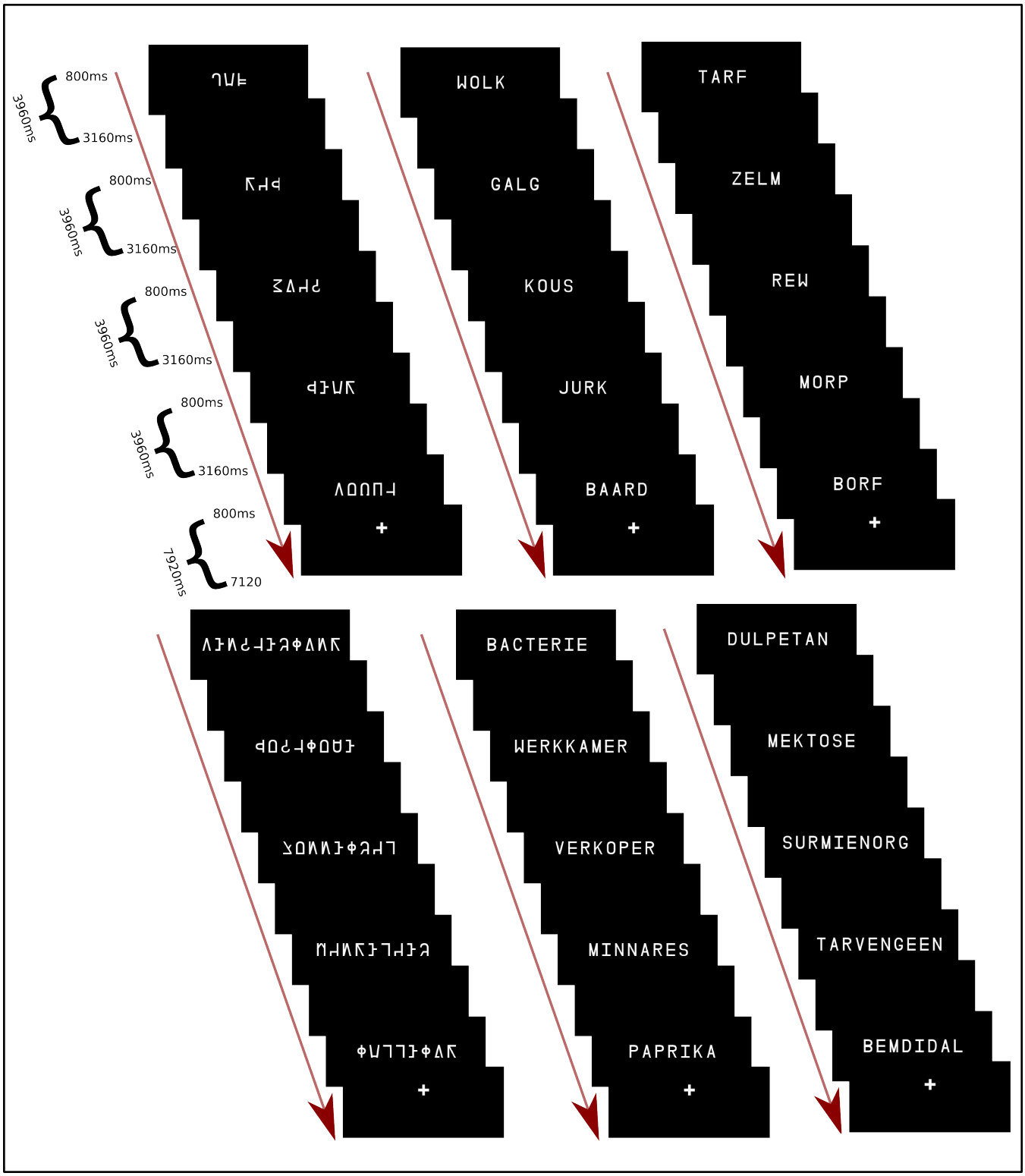

Fig. S5. Sample stimuli showing mini-blocks from each condition. Stimuli were presented in 5 trial mini-blocks with the items from the same condition presented in the trials within a block. The duration of each trial was 3960 ms. Each item was presented for 800 ms within the 3960 ms according to parameters obtained using optseq2. A black screen was presented for the remaining 3160 ms in each trial. A fixation was presented for the duration of one trial (3960 ms) at the beginning of each run and following each mini-block. From left to right: false fonts, words, pseudo-words. One syllable items are shown in the top portion and three syllable items are shown on the bottom.

Additional data table S1 (StimulusList.xlsx)
word and pseudo-word items

Additional data table S2 (FalseFontItems.pdf)
false-font items

References

1. Kok P, Bains LJ, van Mourik T, Norris DG, de Lange FP (2016) Selective activation of the deep layers of the human primary visual cortex by top-down feedback. *Current Biology* 26(3):371–6.
2. Baayen RH, Piepenbrock R, Gulikers L (1995) The celex lexical database (release 2).
3. Keuleers E, Brysbaert M (2010) Wuggy: A multilingual pseudoword generator. *Behavior Research Methods* 42(3):627–633.
4. Cohen L, et al. (2002) Language-specific tuning of visual cortex? functional properties of the visual word form area. *Brain* 125(Pt 5):1054–69.
5. Dale AM (1999) Optimal experimental design for event-related fmri. *Human Brain Mapping* 8(2-3):109–114.
6. Penny WD, Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE (2011) *Statistical parametric mapping: the analysis of functional brain images*. (Elsevier).
7. Cox RW (1996) Afni: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research* 29(3):162–173.
8. Poser BA, Koopmans PJ, Witzel T, Wald LL, Barth M (2010) Three dimensional echo-planar imaging at 7 tesla. *Neuroimage* 51(1):261–6.
9. Breuer FA, et al. (2005) Controlled aliasing in parallel imaging results in higher acceleration (caipirinha) for multi-slice imaging. *Magnetic Resonance in Medicine* 53(3):684–691.
10. Setsompop K, et al. (2012) Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magnetic Resonance in Medicine* 67(5):1210–1224.
11. Marques JP, et al. (2010) Mp2rage, a self bias-field corrected sequence for improved segmentation and t1-mapping at high field. *Neuroimage* 49(2):1271–1281.
12. Dumoulin SO, Fracasso A, van der Zwaag W, Siero JC, Petridou N (2018) Ultra-high field mri: advancing systems neuroscience towards mesoscopic human brain function. *Neuroimage* 168:345–357.
13. O'Herron P, et al. (2016) Neural correlates of single-vessel haemodynamic responses in vivo. *Nature* 534(7607):378.
14. Yacoub E, et al. (2003) Spin-echo fmri in humans using high spatial resolutions and high magnetic fields. *Magnetic Resonance in Medicine* 49(4):655–64.
15. Lee SP, Silva AC, Uğurbil K, Kim SG (1999) Diffusion-weighted spin-echo fmri at 9.4 t: microvascular/tissue contribution to bold signal changes. *Magnetic Resonance in Medicine* 42(5):919–28.
16. Duong TQ, et al. (2003) Microvascular bold contribution at 4 and 7 t in the human brain: gradient-echo and spin-echo fmri with suppression of blood effects. *Magnetic Resonance in Medicine* 49(6):1019–27.
17. Goense J, Logothetis NK (2006) Laminar specificity in monkey v1 using high-resolution se-fmri. *Magnetic Resonance Imaging* 24(4):381 – 392.
18. Harel N, Lin J, Moeller S, Uğurbil K, Yacoub E (2006) Combined imaging-histological study of cortical laminar specificity of fmri signals. *Neuroimage* 29(3):879–87.
19. Kim T, Hendrich KS, Masamoto K, Kim SG (2007) Arterial versus total blood volume changes during neural activity-induced cerebral blood flow change: implication for bold fmri. *Journal of Cerebral Bloodflow Metabolism* 27(6):1235–47.
20. Kim T, Kim SG (2010) Cortical layer-dependent arterial blood volume changes: improved spatial specificity relative to bold fmri. *Neuroimage* 49(2):1340–9.
21. Lu H, et al. (2004) Spatial correlations of laminar bold and cbv responses to rat whisker stimulation with neuronal activity localized by fos expression. *Magnetic Resonance in Medicine* 52(5):1060–8.
22. Silva AC, Koretsky AP (2002) Laminar specificity of functional mri onset times during somatosensory stimulation in rat. *Proceedings of the National Academy of Sciences U.S.A.* 99(23):15182–7.
23. Smirnakis SM, et al. (2007) Spatial specificity of bold versus cerebral blood volume fmri for mapping cortical organization. *Journal of Cerebral Bloodflow Metabolism* 27(6):1248–61.
24. Zappe AC, Pfeuffer J, Merkle H, Logothetis NK, Goense JB (2008) The effect of labeling parameters on perfusion-based fmri in nonhuman primates. *Journal of Cerebral Bloodflow Metabolism* 28(3):640–52.
25. Behzadi Y, Liu TT (2005) An arteriolar compliance model of the cerebral blood flow response to neural stimulus. *Neuroimage* 25(4):1100–11.
26. Devor A, et al. (2007) Suppressed neuronal activity and concurrent arteriolar vasoconstriction may explain negative blood oxygenation level-dependent signal. *Journal of Neuroscience.* 27(16):4452–9.
27. Hillman EM, et al. (2007) Depth-resolved optical imaging and microscopy of vascular compartment dynamics during somatosensory stimulation. *Neuroimage* 35(1):89–104.
28. Koopmans PJ, Barth M, Norris DG (2010) Layer-specific bold activation in human v1. *Human Brain Mapping* 31(9):1297–1304.

29. Ress D, Glover GH, Liu J, Wandell B (2007) Laminar profiles of functional activity in the human brain. *Neuroimage* 34(1):74–84.
30. Polimeni JR, Fischl B, Greve DN, Wald LL (2010) Laminar analysis of 7t bold using an imposed spatial activation pattern in human v1. *Neuroimage* 52(4):1334–46.
31. Huber L, et al. (2014) Investigation of the neurovascular coupling in positive and negative bold responses in human brain at 7 t. *Neuroimage* 97:349–62.
32. Huber L, et al. (2015) Cortical lamina-dependent blood volume changes in human brain at 7 t. *Neuroimage* 107:23–33.
33. Huber L, et al. (2017) High-resolution cbv-fmri allows mapping of laminar activity and connectivity of cortical input and output in human m1. *Neuron* 96(6):1253–1263.e7.
34. Olman CA, et al. (2012) Layer-specific fmri reflects different neuronal computations at different depths in human v1. *PLoS One* 7(3):e32536.
35. De Martino F, et al. (2013) Cortical depth dependent functional responses in humans at 7t: improved specificity with 3d grase. *PLoS One* 8(3):e60514.
36. Kemper VG, De Martino F, Yacoub E, Goebel R (2016) Variable flip angle 3d-grase for high resolution fmri at 7 tesla. *Magnetic Resonance in Medicine* 76(3):897–904.
37. Muckli L, et al. (2015) Contextual feedback to superficial layers of v1. *Current Biology* 25(20):2690 – 2695.
38. Moerel M, et al. (2018) Sensitivity and specificity considerations for fmri encoding, decoding, and mapping of auditory cortex at ultra-high field. *Neuroimage* 164:18–31.
39. Markuerkiaga I, Barth M, Norris DG (2016) A cortical vascular model for examining the specificity of the laminar bold signal. *Neuroimage* 132:491–8.
40. Fischl B (2012) Freesurfer. *NeuroImage* 62(2):774 – 781.
41. van Mourik T, Koopmans PJ, Norris DG (2018) Improved cortical boundary registration for locally distorted fmri scans.
42. Waehnert M, et al. (2014) Anatomically motivated modeling of cortical laminae. *NeuroImage* 93, Part 2:210 – 220.
43. De Martino F, et al. (2015) Frequency preference and attention effects across cortical depths in the human primary auditory cortex. *Proceedings of the National Academy of Sciences U.S.A.* 112(52):16036–41.
44. Lawrence S, et al. (2018) Laminar organization of working memory signals in human visual cortex. *Current Biology* 28(21):3435 – 3440.e4.
45. Rockland KS, Pandya DN (1979) Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Research* 179(1):3 – 20.
46. Felleman DJ, Van Essen D (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex* 1(1):1–47.
47. Kasper L, et al. (2017) The physio toolbox for modeling physiological noise in fmri data. *Journal of Neuroscience Methods* 276:56–72.
48. Price CJ (2012) A review and synthesis of the first 20 years of pet and fmri studies of heard speech, spoken language and reading. *Neuroimage* 62(2):816–47.
49. van Mourik T, van der Eerden JP, Bazin PL, Norris DG (2019) Laminar signal extraction over extended cortical areas by means of a spatial glm. *PLoS One* 14(3):e0212493.
50. Polimeni JR, Greve DN, Fischl B, Wald LL (2010) Depth-resolved laminar analysis of resting-state fluctuation amplitude in high-resolution 7t fmri. Proceedings: 18th Annual Meeting International Society for Magnetic Resonance in Medicine, Stockholm, p. 1168.
51. McLaren DG, Ries ML, Xu G, Johnson SC (2012) A generalized form of context-dependent psychophysiological interactions (gppi): A comparison to standard approaches. *NeuroImage* 61(4):1277 – 1286.
52. O'Reilly JX, Woolrich MW, Behrens TE, Smith SM, Johansen-Berg H (2012) Tools of the trade: psychophysiological interactions and functional connectivity. *Social Cognitive and Affective Neuroscience* 7(5):604–609.