Published in final edited form as:
Neuroimage. 2009 August 15; 47(2): 745–755. doi:10.1016/j.neuroimage.2009.04.089.

Strategies for longitudinal neuroimaging studies of overt language production

Jed A. Meltzer[1], **Whitney A. Postman-Caucheteux**[2], **Joseph J. McArdle**[1], and **Allen R. Braun**[1]

1 *Language Section; Voice, Speech, and Language Branch, National Institute on Deafness and Other Communication Disorders, National Institutes of Health, 10 Center Drive, Building 10, Rm 5C410, Bethesda, MD 20892-1065*

2 *Department of Communication Sciences and Disorders, 110 Weiss Hall, 1701 N. 13th Street, Temple University, Philadelphia, PA 19122*

Abstract

Longitudinal fMRI studies of language production are of interest for evaluating recovery from post-stroke aphasia, but numerous methodological issues remain unresolved, particularly regarding strategies for evaluating single subjects at multiple timepoints. To address these issues, we studied of overt picture naming in eleven healthy subjects, scanned four times each at one-month intervals. To evaluate the natural variability present across repeated sessions, repeated scans were directly contrasted in a unified statistical framework on a per-voxel basis. The effect of stimulus familiarity was evaluated using explicitly overtrained pictures, novel pictures, and untrained pictures that were repeated across sessions. For untrained pictures, we found that activation declined across multiple sessions, equally for both novel and repeated stimuli. Thus, no repetition priming for individual stimuli at one-month intervals was found, but rather a general effect of task habituation was present. Using a set of overtrained pictures identical in each session, no decline was found, but activation was minimized and produced less consistent patterns across participants, as measured by intra-class correlation coefficients. Subtraction of a baseline task, in which subjects produced a stereotyped utterance to scrambled pictures, resulted in specific activations in the left inferior frontal gyrus and other language areas for untrained items, while overlearned stimuli relative to pseudo pictures activated only the fusiform gyrus and supplementary motor area. These findings indicate that longitudinal fMRI is an effective means of detecting changes in neural activation magnitude over time, as long as the effect of task habituation is taken into account.

INTRODUCTION

Neuroimaging studies of aphasia have sought to elucidate the changes in neural organization that underlie the recovery of function over time following brain damage, typically ischemic stroke. While early studies compared groups of stable aphasic patients with healthy controls (Karbe et al., 1998, Rosen et al., 2000, Blasi et al., 2002), recently there has been much interest in employing repeated measurements on the same individuals at multiple timepoints, in order to assess the effect of a specific therapy (Leger et al., 2002, Fridriksson et al., 2006, 2007) or

Corresponding author: Jed A. Meltzer, E-mail: jed.meltzer@aya.yale.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

to explore the natural timecourse of recovery in the first year after stroke (Saur et al., 2006, Cardebat et al., 2003, Cappa et al., 1997). Although numerous tasks have been shown to be effective at localizing cortical areas involved in language processing, overt picture naming is of particular interest as a language production task with practical applicability. Overt picture naming offers promise for recovery studies because nearly all kinds of aphasic syndromes include difficulties with naming (Goodglass, 1993), and it provides a clear behavioral measure of function that can even be assessed on a trial-by-trial basis in the scanner (Meinzer et al., 2006, Postman-Caucheteux et al., 2007, *submitted*).

Nonetheless, there are numerous methodological challenges to be overcome in order for longitudinal studies of picture naming to be effective in elucidating the mechanisms of aphasia recovery. Many of these are addressed in a recent review by Crosson et al. (2007). Some of the issues discussed therein deal with difficulties involved in imaging language production in general, while others relate to the reliability and stability of the signal across multiple sessions, which is an essential element of any longitudinal neuroimaging study. Consistency across sessions depends both on behavioral stability in patients and reproducibility of fMRI activation magnitudes in general (Kurland et al., 2006). In the present study, we focus on the latter issue of activation reproducibility, examining several factors relevant to the design and analysis of a longitudinal fMRI study of picture naming, using a cohort of eleven young, healthy subjects. Subjects were scanned on four occasions, approximately one month apart. The goals of this study were to identify the strategies that lead to the most consistent and reliable activation patterns within individuals, and also to establish a baseline of expected activations for overt picture naming and consistent trends for changes over multiple sessions in neurologically intact subjects. Only in the context of such a baseline can changes within individual patients with aphasia be interpreted. Patients with aphasia may exhibit variable performance across repeated sessions, due to both inherent variability in symptom severity and functional recovery over time. In contrast, healthy control subjects are expected to exhibit more consistent performance in picture naming, at ceiling level. Therefore, an evaluation of the amount of signal variability present in the healthy population over repeated scanning sessions is essential to interpreting changes in functional activation in aphasic patients as reflecting their behavioral status.

Another leading issue in investigations of aphasia recovery is that of practice. The most effective treatments of lexical retrieval deficits in people with aphasia are those that demonstrate expansion of patients' improvement on items trained in therapy to untrained items (e.g., Kiran & Thompson, 2003). Since generalization is one of the best indications of the efficacy of therapy and of real recovery (as opposed to practice), in functional naming tasks with patients with aphasia, pictures trained in therapy as well as untrained pictures are commonly incorporated as stimuli (e.g., Vitali et al. (2007), Meinzer et al. (2006)). This raises the important question of whether it is better to present the same stimuli in each session, in order to reduce inter-session variability, or to use novel stimuli in each session to produce optimally reliable activation patterns. Therefore, each imaging session in our study was divided. In the first portion, subjects named the same pictures in the same order every time, and were drilled on the same pictures before each session to make them maximally familiar. In the second portion, untrained pictures were used in each session. Additionally, within the untrained runs, half of the pictures were truly novel, and the other half were repeated across sessions (appearing once in each session), although the order was randomized and subjects were not informed of the repetition. This manipulation allowed us to evaluate the presence of any long-term priming or habituation effect for stimuli repeated over the course of 4 months, compared to any changes in activation magnitude that may be present for novel stimuli as well, the latter representing general processes of habituation to the picture naming task. Furthermore, we examined the use of a control condition in which subjects viewed scrambled “pseudo” objects and produced a stereotypical verbal response. This comparison was intended to subtract out low-level processes of vocal production and visual stimulation, thus isolating activity related to picture recognition and lexical access.

Besides evaluating the impact of these design factors on the nature and stability of activation patterns in overt picture naming, we also present a somewhat novel approach for the analysis of longitudinal neuroimaging data, featuring a direct statistical comparison of activation magnitude between multiple sessions. As the most relevant finding in a longitudinal study of aphasia recovery is changes between the sessions, our analytic method is optimized to detect evidence of such changes. In some studies, changes between multiple imaging sessions have been evaluated by visual comparison of statistical maps from each session independently, or by counting voxels within a given region that exceed a specific arbitrary statistical threshold. Unfortunately, such a strategy is notoriously unstable. Crosson et al. (2007) note that a voxel activated in one session is, on average, activated one out of three times in a second session, while counts of activated voxels within a region of interest (ROI) are somewhat more reliable. However, previous studies have compared voxel-counting approaches with measurements of activation magnitude (typically in the form of percent-signal change), and have found that activation magnitudes are much more consistent across sessions (Cohen and DuBois, 1999, Chee et al., 2003, Friedman et al., 2007, Kimberley et al., 2008). In this study, we present a method for producing a single statistical map representing the evidence for a significant change in activation magnitude across sessions. While a direct statistical comparison between multiple sessions is optimal, even this method depends on reliable measurement of activation magnitude within single sessions. Therefore, we compared the reliability of activation values within individual subjects obtained with the various conditions examined in this experiment, using the Intraclass Correlation Coefficient (ICC), a popular measure of reproducibility in neuroimaging studies.

In summary, we report here results of three different analyses related to the design of a longitudinal experiment of overt picture naming: 1) activation patterns observed for naming real pictures under overlearned and untrained conditions, and the subtraction of real vs. pseudo pictures under both conditions, 2) systematic effects of repeated scanning sessions, in the form of linear increases or decreases in activation magnitude, and 3) reliability of individual subjects' activation patterns for real naming and real vs. pseudo comparisons, calculated using the Intraclass Correlation Coefficient.

METHODS

Subjects

Eleven subjects (four females, age range 24–39) participated after providing written informed consent, and were financially compensated. All were right-handed, monolingual native speakers of North American English, with no history of neurological impairment. Each subject was scanned during four separate fMRI sessions, at approximately one month intervals, for a total of 44 individual data sets. Due to a scanner malfunction, data from five of these imaging sessions (all on the same day) were irretrievably corrupted. The missing data sets were from the fourth imaging session for two subjects, the third session for two more, and the second session for another subject. These missing data points impacted the longitudinal analysis strategies in minor ways, which will be noted below. It is important to note that although some imaging data points were unavailable, all subjects did complete all four scanning sessions, and so it is behaviorally valid to consider the final session as session 4 even if session 3 is missing, for example.

Task

An overt picture naming task was used throughout the experiment. Subjects were presented with black and white line drawing pictures and instructed to say the name of each object without delay. Over the four scanning sessions, subjects named a total of 504 objects (animals, foods, plants, body parts, clothing, appliances, furniture, musical instruments, and tools), selected from the International Picture Naming Project (‘IPNP’) database (Szekely et al., 2003, Bates et al., 2003, http://crl.ucsd.edu/~aszekely/ipnp). Additionally, a set of 320 “pseudo-objects” was constructed for a control condition. The pseudo-objects were pictures derived from digital distortions of line drawings of real objects, produced with Adobe Illustrator (Adobe Systems, San Jose, CA), such that they were not recognizable as real objects but maintained the same level of visual complexity. Subjects were instructed to utter a single stereotyped verbal response to each pseudo object, which was an arbitrarily chosen disyllabic CVCV non-word, “rado” [‘re^j-do^u] (stress on initial syllable). A fixation cross was displayed on the screen at all times except when the pictures appeared. Each picture appeared on the screen for 800 ms, projected against a black background using Presentation software (Neurobehavioral Systems, Albany, CA).

Each of the four experimental sessions consisted of 12 functional imaging runs. In each run, 18 real objects and 18 pseudo objects were presented. The inter-trial interval was jittered randomly in 1-second increments between 4 and 8 seconds, to allow for deconvolution of the BOLD response through multiple regression (Miezin et al., 2000). In order to reduce confusion between the real object and pseudo object conditions, the two conditions were grouped into blocks of 9 trials each, and a 20 second fixation period was inserted in between blocks to separate them and allow the hemodynamic response to return to baseline. The order of blocks (four per run) was randomized.

Besides the basic distinction between real and pseudo objects, real objects were grouped into three subcategories in order to investigate effects of stimulus novelty and familiarity in the context of repeated scans in a longitudinal experiment. 72 pictures were selected as an “overlearned” stimulus set, henceforth referred to as “OL.” These pictures were printed out on paper and assembled into a booklet. In a training session prior to the study and again immediately before each fMRI session, each subject was familiarized with the stimuli by being asked to name all of the pictures from the booklet aloud. When their responses did not match the dominant name of the target, they were trained on the dominant name. Thus, they were trained to utter the same name for a given OL stimulus at each scanning session. In all four fMRI scanning sessions over four months, the first four runs consisted of the 72 OL real pictures and 72 pseudo objects, in exactly the same order and timing for each session. In addition, subjects were familiarized with the task and the pseudo-object stimuli prior to the study, by practicing with the stimuli of these OL runs outside the scanner on a laptop computer.

After the four OL runs, the remaining eight runs consisted of untrained stimuli (henceforth “UT”). The real pictures in the UT set were randomly assigned to two categories, “novel” and “repeat.” In the first scanning session, all untrained pictures (n=144) were considered to be novel. In the three subsequent sessions, half of the UT real object pictures were novel, being presented for the first time. The other half of the real object pictures were repeated from the first session, such that during the fourth session, the subjects were naming these particular items (n=72) for the fourth time. Novel and repeated items were randomly interspersed during the real-object blocks of the UT runs. Each UT run (32 runs over four sessions) contained a unique arrangement of timings, novel pictures, and pseudo stimuli, and a randomized order of presentation for the repeated pictures. This arrangement allowed us to assess the potentially more subtle effect of repeating an individual stimulus that had been seen in a previous experimental session, as opposed to explicitly overtraining on a set of stimuli. Subjects were not informed about the repetition of certain stimuli across scans. Real pictures in the three categories (OL, UT novel, and UT repeat) were balanced across runs within a scanning session and across scanning sessions for semantic category, and the following variables based on norms from the IPNP: number of alternative names; percent name agreement; visual complexity; and length in syllables, frequency, and age of acquisition of the dominant target name.

FMRI acquisition

Blood oxygen level dependent imaging (BOLD) data were acquired on a 3-Tesla whole-body scanner (GE Signa; General Electric Medical Systems, Milwaukee, WI) using a standard quadrature head coil and a gradient-echo EPI sequence. The scan parameters were as follows: TR = 2000 ms, TE = 30 ms, flip-angle = 90°, 64×64 matrix, field of view 240 mm, 23 parallel axial slices covering the whole brain, 6 mm thickness. A high-resolution anatomical image (T1-weighted, axial MPRAGE, 0.86×0.86×1.2mm resolution) was also acquired. A total of twelve fMRI runs were conducted in each session, and the length of each run was 135 volumes (4min, 30sec), plus four initial volumes that were discarded to achieve steady-state magnetization. Foam padding was used to reduce subject movement. Subjects were instructed to name each picture with a minimal response (i.e. one word, except for occasional familiar compounds such as "vacuum cleaner.") Subjects were coached on producing unique responses with minimal head motion, while avoiding extraneous vocalizations.

FMRI processing

BOLD images were preprocessed in AFNI software (Cox, 1996), with standard steps, including brain extraction, motion correction, spatial smoothing (8mm FWHM), and voxelwise timecourse normalization to percent of mean signal level, thus scaling subsequent regression parameter estimates into percent signal change measures. Due to the increased risk of head motion involved in overt speech production, an automated procedure coded in Python was used to identify timepoints at which the EPI signal values were most likely to be affected by motion occurring within the time of volume acquisition, which cannot be corrected by the standard volumetric registration approach to motion correction. This procedure tagged all time points at which the first derivative of any of the six estimated motion parameters (3 translations and 3 rotations) went beyond two standard deviations of its values throughout each run. These timepoints, comprising 3% of the total, were excluded from subsequent regression analyses.

Hemodynamic response magnitudes for each condition (real and pseudo within the OL runs, and novel, repeat, and pseudo within the UT runs) were estimated by convolving the stimulus timings with a canonical gamma density function with default parameters and computing the regression coefficients with this function. In a subset of subjects, we also performed a full deconvolution of the hemodynamic response using a more flexible Finite Infinite Response modeling strategy (Miezin et al., 2000), in order to confirm that the canonical model was an adequate fit for the responses in this task. Visual inspection showed an excellent match for the canonical function, and so we elected to use it for subsequent analyses. The use of a single response function simplifies the computations for analyses across subjects and sessions; however, it may not be possible in the case of certain stroke patients, who may have altered hemodynamic responses, or highly delayed verbal reaction times (Bonakdarpour et al., 2007).

Statistical results (regression coefficients and t-values) from single-session analyses were spatially transformed into Talairach space using a single warp encompassing two computed transformations: a 6-parameter rigid transformation from the EPI base image (to which all other images were registered during motion correction) to the high-resolution T1, and then a fully non-linear grid-based deformation (Papademetris et al., 2004) to the standard AFNI reference brain, which is the "colin27" brain transformed to match the space of the Talairach atlas.

Coregistration and warping were done using the program BioImage Suite (http://bioimagesuite.org/).

For activation maps (main effects and across-session trends), the main statistical test took the form of a voxel-wise one-sample t-test applied to the regression contrast values obtained from each subject. To correct for multiple comparisons over all intracranial voxels, we used a voxel-wise threshold of p <.001 (2-tailed) in combination with a cluster size criterion (Forman et al., 1995) determined by Monte Carlo simulations using the AFNI program *Alphasim*. The chosen voxel-wise threshold, combined with the whole-brain search volume and smoothness of the data resulted in a cluster-size criterion of 62 voxels (496 μL), for a corrected family-wise error rate of p <.05.

Multi-session analysis

In most multi-subject fMRI studies, each subject is scanned once, and the activation magnitudes from each subject are submitted to a second-level statistical analysis. The present study is complicated by the presence of both multiple sessions and subjects. As usual, subjects are a random effect, sampled from a larger population. In contrast, sessions are a fixed effect, in that we are explicitly interested in the changes from session to session that may be present in the cohort. Therefore, we adapted a procedure for estimating both session effects and the main effects of conditions within each subject individually, and furthermore submitting these effects to second-level analysis across subjects. To do this, we combined the multiple imaging sessions for each subject into one regression analysis. This required spatial transformation of the preprocessed EPI time series into a common space across sessions. We computed a rigid transformation between each session’s EPI base image and its T1-weighted high-resolution anatomical image, and concatenated this transform with a rigid transformation between the T1 image for that session and that for the first session. Each EPI volume was then transformed into the space of the first-session anatomical image and interpolated to a 3mm isotropic resolution. To ensure that each voxel was adequately sampled in every session, we computed a binary mask for every in-brain EPI voxel, submitted the mask to the same spatial transformation, and limited the analysis only to voxels that were included in all four sessions.

In the regression analysis, trials occurring in each session were modeled as separate conditions. Thus, a multi-session regression for the OL runs yielded parameter estimates for real pictures in session 1, pseudo pictures in session 1, real pictures in session 2, etc. Thus, the total variance for within-subject statistical analyses was pooled across sessions. With this approach, changes in the amount of noise between sessions do not produce misleading changes in activation, as it is the activation magnitude that is compared, rather than a t-value. To compute main effects of conditions (real vs. baseline, pseudo vs. baseline, and real vs. pseudo), linear contrasts were used, resulting in an averaged magnitude over the four sessions, along with a single t-value reflecting the statistical likelihood of an activation, pooled over all sessions. The magnitude values were then spatially transformed to Talairach space and submitted to one-sample t-tests across subjects.

For analysis of changes across sessions, linear contrasts were used to compare, for example, real pictures in session 1 vs. real pictures in session 2. Additionally, a single linear contrast was used to test for a trend of increasing or decreasing activation across the four sessions. This contrast was achieved by multiplying the activation values from sessions 1 through 4 by the vector [−3 −1 1 3]. In the cases in which subjects completed four scanning sessions but one session’s data was excluded, the appropriate weights were still used, i.e., in the case of a missing third session, we used the weights [−3 −1 3], rather than [−1 0 1], as would be used if there were only three sessions performed.

Intra-class correlation analysis

Intra-class correlation (ICC) is a measure that originated in the field of psychometrics as a means of computing agreement between multiple raters (Fleiss, 1986), but has come to be widely used to evaluate the inter-session reproducibility of fMRI scans. In order to compute ICC, one first conducts an ANOVA with subject and session as random factors. From the ANOVA mean-squares values, the measure, which ranges from 0 to 1 is computed as follows:

$$\mathrm{ICC}=[\mathrm{BMS}-\mathrm{EMS}]/[\mathrm{BMS}+(k-1)\,\mathrm{EMS}]$$

where BMS is between-subject mean square, EMS is the error mean square, and k is the number of sessions (Fleiss, 1986). Note that systematic differences between subjects will tend to increase the ICC value, while differences between sessions, whether systematic or random, can be expected to decrease it. Therefore, the regions with the highest ICC may be those that are consistently activated in some subjects but not in others.

While the ICC in the form described above has been used as a measure of reliability in several multi-session fMRI studies, previous studies have varied greatly in the exact nature of the data that is entered into the equations. The kinds of measures that have been entered into ICC analyses include binary classifications of voxels as activated or not (Aron et al., 2006, Eaton et al., 2006), t-values over a certain threshold (Specht et al., 2003, Fernandez et al., 2003), sums of z-scores in suprathreshold voxels within an ROI (Wei et al., 2004), and signal change in super-threshold voxels within a region centered on an activation peak (Kong et al., 2007). Given this heterogeneity of methods, it is difficult to compare studies that purport to use the same ICC measure. Fortunately, some systematic comparisons have been performed with the ICC, and have unanimously indicated that measures of signal change magnitude are more reproducible than statistical values or crossing of an arbitrary threshold (Friedman et al, 2007, Kimberley et al., 2008), which agrees with previous studies that have examined reliability with measures other than the ICC (Chee et al., 2003, Cohen and DuBois, 1999). Therefore, in the present study, we elected to compute the ICC on the estimated signal change values from each subject and session, on a voxel-by-voxel basis. We compared the reliability of several different activation contrasts (see results).

RESULTS

Main effects of picture naming

Main effects reflect average activations across all four imaging sessions. Activation maps for naming real pictures under overlearned and untrained conditions are presented in figure 1A–B. These maps display regions in which a hemodynamic response to picture presentation is detected, relative to a fixation baseline. The activation patterns are quite similar, and cover extensive regions of the brain, including bilateral activations in motor and premotor cortex, supplementary motor area, visual cortex, auditory cortex, inferior parietal lobe, middle temporal gyrus, inferior frontal gyrus and insula, putamen and extensive portions of the thalamus. The widespread activation pattern is unsurprising, in that overt naming of pictures engages several neural systems, including areas specialized for visual, motor, auditory, attention, and linguistic processes. The subtraction of a control condition is required to highlight regions specifically involved in processes of linguistic interest, which include picture recognition, lexical access, response selection, phonological planning, and articulation. Therefore, we examined the main effect of naming real pictures vs. producing a stereotyped vocal response to pseudo pictures. The resulting activations for overlearned and untrained conditions are presented in figure 1C–D and table 1A–B. In the overlearned condition, significantly greater activation for real vs. pseudo picture naming was detected in only two regions, the left fusiform gyrus (LFusG) and a small cluster encompassing the dorsal anterior cingulate cortex (ACC) and supplementary motor area (SMA). In the untrained condition, however, activations are observed in the same two regions, but also in several additional regions, including a large portion of the left inferior frontal gyrus (LIFG), left middle temporal gyrus, bilateral inferior precentral gyrus, pre-SMA, and in subcortical regions including the putamen and thalamus.

Session effects

Within each individual subject, a linear contrast was computed reflecting the presence of a linear trend for increasing or decreasing activation over the four sessions. These contrasts were then submitted to voxel-wise random effects analysis. The resulting maps show regions in which activation tends to change over repeated testing in a consistent manner across subjects, but do not necessarily reflect the reliability of single-session measurements within individual subjects (for that, see the ICC analyses below). The use of a single linear trend contrast eliminates the problem of multiple comparisons involved in testing each pair of sessions against each other, and is sufficiently sensitive to reveal an overall pattern of increase or decrease even if the pattern is not perfectly linear. Trend analyses were done for real and pseudo picture naming in the overlearned category, while in the untrained category, separate trend contrasts were computed for novel real items and repeated real items, as half of the pictures repeated in all four sessions, while the other half were novel each time. All significant changes were negative. That is, activation magnitudes tended to decrease over subsequent sessions, but never increased. In the overlearned condition, activation for real object naming decreased in only one region, within the left precentral gyrus, while activation to pseudo objects decreased in a few small clusters within the right hemisphere only (table 2A–B). Across the untrained sessions, no significant changes were detected for pseudo object naming. The modest attenuation for pseudo objects occurring only in the overlearned runs may be attributable to the repetition of the same pseudo objects in the same order in each session, whereas the pseudo objects in the untrained runs were different. For real untrained objects, however, extensive decreases in activation magnitude occurred, for both novel and repeated items (figure 2A–B, table 2C–D). The resulting maps of linear decrease for novel and repeated items are similar, showing decreasing activation across sessions in bilateral motor cortex, bilateral inferior parietal cortex, supplementary motor area, and left premotor cortex. Signal decrease is also seen in the left inferior frontal gyrus and insula, but only in the repeated condition.

Using a paired t-test, we also tested for differences in the slope of activation changes across sessions, to answer the question of whether activation for repeated items declines more than activation for novel items, reflecting some sort of long-term priming effect over months. Despite the disparity in LIFG between the two maps, no significant clusters were found when testing over the whole brain for differences between trend slopes, indicating that the rates of decline for novel and repeated stimuli are statistically indistinguishable at the sensitivity level of whole-brain analysis, correcting for multiple comparisons on the cluster level. Similarly, we also tested for a difference between novel and repeated items *within* each session separately, but no significant differences were detected in any session at our specified threshold.

To illustrate these effects, we display in bar graph form the activation changes across multiple sessions in two regions, LIFG and LFusG. To produce these graphs, activation magnitudes were averaged across voxels in an 8mm spherical ROI centered at peak activation points for the main effect contrast Untrained Real vs. Untrained Pseudo. For the overlearned runs, signal in the LIFG is not consistently different for real and pseudo pictures (figure 2C). While some differences are present in certain sessions, the directionality of the differences is not consistent. In contrast, during untrained runs, real pictures consistently induce a much higher response than pseudo pictures, which do not induce much response at all (figure 2D). Furthermore, the magnitude of the response to real untrained pictures tends to decrease across the sessions, although the decline is similar for both novel and repeated items. In the fourth session alone, there seems to be a modest difference between novel and repeated items in this region, consistent with the significant linear decrease in LIFG seen in figure 2B. The direct comparison between novel and repeated responses in this ROI for the fourth session is not quite statistically significant [t(8) = 1.92, p=.09, 2-tailed], and the magnitude of this difference is obviously not nearly enough to meet thresholding criteria for whole-brain analysis. It can be seen from the bar graph that although the response to novel items seems to level off by the fourth session, there is still an overall trend of decreasing activation over the four sessions.

In LFusG, responses to real pictures were consistently greater than pseudo pictures, in both overlearned (figure 2E) and untrained runs (figure 2F). In this region, as in LIFG, there is no consistent effect of repeating individual untrained pictures, as changes in activation magnitude across sessions are approximately equal for novel and repeated pictures. This particular ROI, defined on the basis of the main effect for real vs. pseudo in untrained naming, was not part of a significant cluster for a decreasing linear activation trend across sessions. Nonetheless, inspection of the signal changes across sessions in this region, as in many others, shows that the signal change does tend to decrease across sessions, although the third session was a slight exception to the general trend. This pattern, along with the lack of a distinction between the novel and repeated items, indicates that general familiarity with the task and the fMRI environment may be a major factor in across-session changes.

Reproducibility: Intra-class correlation coefficients

In order to assess the reproducibility of activation patterns from individual subjects across multiple sessions, we applied the widely used ICC metric to percent signal change values, on a voxel-by-voxel basis. Due to the missing data points in this experiment (see Methods), this analysis was limited to eight subjects, for whom the first three imaging sessions were available. The purpose of this comparison was to see which contrast gave the most reproducible activation patterns within individual subjects, and also to see if activation was more reproducible in any particular brain region. We compared four different contrasts: naming of real overlearned pictures, naming of real untrained pictures, and the subtractions of real vs. pseudo naming under overlearned and untrained conditions. To evaluate the reproducibility of these contrasts throughout the brain in an unbiased fashion, we computed the ICC at each voxel, rather than selecting regions of interest. The resulting ICC maps were thresholded at an ICC value of 0.7, which is commonly considered to be a high level of reliability (Fleiss, 1986).

The thresholded maps are displayed in figure 3. Figure 3A displays the ICC map for naming overlearned real pictures. Relatively few voxels achieve the ICC level of 0.7 for this contrast. Next, the ICC map for naming untrained real pictures is displayed in figure 3B. The reliability of activation magnitudes for untrained pictures is much better than that for the overlearned pictures (figure 3A). This may seem surprising, as from the analyses of trends reported above, we know that activation for untrained pictures declines systematically across multiple sessions. This effect serves to increase the inter-session variance. Since the ICC is effectively a ratio of between-subject variance to between-session variance, the systematic activation decline is expected to decrease the observed ICC. Therefore, one might expect the ICC map for untrained pictures to indicate less reliability in the untrained condition. However, this is not the case. For untrained pictures, most voxels throughout the network of activated regions achieve the specified level of ICC, despite the known activation decline across sessions. High ICC values are seen across visual, auditory, and motor cortices, and also in bilateral IFG, middle temporal, and inferior parietal cortex. This indicates that untrained pictures in fact give a more consistent activation pattern within individual subjects, even though the magnitude of activation may decline with multiple testing sessions.

We also evaluated the reliability of the real vs. pseudo subtraction under overlearned and untrained conditions. In the overlearned condition, not a single voxel achieved the level of 0.7 (data not shown). In the untrained condition, the real vs. pseudo subtraction reached that level of ICC in relatively few voxels (figure 3C), but the one large region where high ICC was seen was an interesting one – the right IFG. Since ICC is increased by consistent differences between subjects, this indicates that a subset of the subjects tended to consistently activate RIFG more to real than to pseudo objects. In the random effects analysis across subjects, (figure 1D), this contrast activated the LIFG, indicating that all subjects tended to activate this region. The high ICC values seen in the right hemisphere homolog indicate that some subjects tend to activate this region as well in a consistent fashion, but they are only a subset of the group. This finding thus represents a confirmation of the well-known fact that healthy right-handed subjects vary in the degree to which their language functions are left-lateralized, with some having a reproducibly bilateral or even right-lateralized distribution (Chee et al., 1998, Knecht et al., 2001, 2003, Eaton et al., 2008). Nonetheless, a comparison of figures 3B and 3C indicates that the activation magnitude of naming real pictures alone (against a fixation baseline) is more reproducible than the subtraction of real naming vs. pseudo picture naming.

DISCUSSION

Background: Longitudinal neuroimaging of picture naming in aphasia

The purpose of this study was to evaluate the effectiveness of several methodological alternatives for a longitudinal study of picture naming in aphasia. Before discussing our findings and recommendations, we will briefly review the motivation for this study and current progress in the field. Picture naming is a useful task for assessing language function, and has been frequently studied in aphasic patients in order to assess the degree to which recovery of language functions can be attributed to the neuroplastic recruitment of brain regions that are not normally involved (Martin et al., 2005, Fridriksson et al., 2006, Leger et al., 2002). Most imaging studies of aphasia have conducted a single imaging session in each patient, choosing patients who are well past their stroke date, in order to ensure that the patient has reached a stable state. While confining studies to stable chronic patients reduces the confound of temporal variability within individuals, there may also be much to be gained from studying single subjects at multiple timepoints, particularly while they are showing improvement in the intervening time. While some studies have used longitudinal imaging sessions to evaluate the effect of a specific therapy on stable patients (Crosson et al., 2005, Leger et al., 2002), another very promising approach is to study patients in a fairly acute stage, in order to characterize the natural process of recovery that tends to be most pronounced in the first year post-stroke. One such study has been conducted (Saur et al., 2006), using a language comprehension paradigm with a plausibility judgment task. In that study, increased right hemisphere activation was seen in the subacute stage relative to acute and chronic stages, suggesting that RH activation may be a transient stage in recovery.

In order to evaluate the timecourse of language network reorganization, there is currently much interest in using overt picture naming as a production task. As methods for recording patient vocal responses during simultaneous EPI scanning are now widely available, it is possible to compare correct and incorrect responses within an individual, thus helping to distinguish between functional reorganization and “ineffective” or “maladaptive” activity. Thus far, a small number of studies have applied variations of this method. Postman-Caucheteux et al. (*submitted*) reported increased right (contralesional) inferior frontal activation for erroneous responses to novel pictures in three mildly impaired patients with conduction or anomic aphasia. In the recovery studies by Meinzer et al. (2006) and Vitali et al. (2007), more severely impaired stroke patients were scanned before and after treatment for anomia. The patient with Wernicke’s aphasia in the former investigation and a patient with phonological anomia in the latter, both showed RIFG activation associated with naming trained items, that increased with accuracy in tandem with left (perilesional) or left frontal activation. Given the temporal variability reported by Saur et al. (2006), the next logical step seems to be to evaluate patterns of activation related to performance at multiple timepoints, especially in patients who show spontaneous improvement in the first year poststroke. A longitudinal study of overt naming in recovering aphasics may be able to reveal the emergence of new language networks in the post-stroke brain, as changes in activation magnitude over multiple sessions (ideally more than two) may proceed in parallel with, and be correlated with, improvements in naming performance.

To prepare for such a study, it is necessary to evaluate the baseline pattern of activation and degree of variability to be seen in healthy controls undergoing repeated scans in a picture naming paradigm over several months. We have done so in the present study, while comparing different strategies for conducting repeated measurements of picture naming within individuals. We will now review the findings of this study, emphasizing the specific recommendations for a longitudinal study generated from the present data.

Use of a control condition, overlearned, and untrained materials

We compared the use of a simple contrast of naming real pictures, relative to an implicit fixation baseline, versus the subtraction of a control condition, i.e., naming real pictures minus “naming” pseudo pictures (via a stereotypical nonword response). We found that naming real pictures activated a large amount of tissue, including visual, auditory, and motor systems, a pattern which is not surprising given that visual stimulation, vocal response, and auditory stimulation (from the sound of one’s own voice) are essential components of the trial structure. The subtraction of a control condition was effective at producing activation maps specific to regions putatively involved in the higher cognitive aspects of picture naming, although the extent of the observed activation depended on whether the stimuli were overlearned or untrained. Using overlearned stimuli, we detected significantly greater activation for real picture naming relative to pseudo naming in only two regions: SMA/ACC and the left fusiform gyrus (LFusG). This finding implies that these two regions are critically involved in picture naming even when it is highly practiced and automatized.

Under untrained conditions, we observed extensive activation for real vs. pseudo picture naming in Left Inferior Frontal Gyrus (LIFG), a region commonly associated with speech production and also implicated in naming deficits (Hillis et al., 2006, Deleon et al., 2007). Our findings suggest a dissociation between LIFG and LFusG, in that LFusG is consistently activated (relative to the pseudo naming baseline) even when naming objects that are highly practiced, while LIFG is only engaged in the more effortful condition of naming untrained objects. It has been suggested that LIFG activations in picture naming may reflect selection from multiple alternatives, rather than processes of recognition and lexical access per se (Kan and Thompson-Schill, 2004). Our results are consistent with that interpretation, as there is very little selection involved when the pictures are overlearned and the subject has already decided what to call them. Nonetheless, the untrained condition is arguably more applicable to everyday language use, and thus it can be inferred from our results that LIFG does play an important role in word finding under normal conditions.

These findings suggest that LIFG is involved in lexical search and effortful word retrieval, but is not particularly engaged under conditions of low effort, when word retrieval is highly automatized. They are in good agreement with previous studies of language production, such as the PET study of Bookheimer et al., (2000), which demonstrated that Broca’s area was activated when reciting a memorized prose passage, but not by more automatic speech tasks such as repeating a phoneme sequence or the months of the year. Similarly, some studies of overt picture naming have shown no activation of LIFG. Etard et al. (2000), using PET, showed activation in LIFG for verb generation but not for naming the same pictures, while Rau et al.

(2007) conducted a longitudinal fMRI experiment of picture naming using three sessions with the same stimuli each time. Using a voxel-counting approach, Rau et al. showed very poor reproducibility for overt naming, but better reproducibility with a more complex task involving both naming and word generation. The authors note that priming effects may have reduced the activation in LIFG, which is confirmed by our finding that naming untrained pictures produces LIFG activations that are both larger and more reproducible than naming overlearned pictures.

In contrast to LIFG, the left fusiform gyrus was consistently activated in naming real pictures relative to pseudo pictures, regardless of how overlearned or automatized they were. This finding suggests that the left fusiform gyrus may be the single region most essential for picture naming performance. Although anomia is a common symptom in all forms of aphasia arising from various lesions (Goodglass and Wingfield, 1997), a particularly central role for the left fusiform gyrus is consistent with studies of hypoperfusion in acute stroke patients, which have demonstrated that reperfusion of this region is the single best predictor for improved naming performance (Hillis et al., 2006, Deleon et al., 2007). Models of naming commonly distinguish between initial stages of picture recognition and lexical access, associated with posterior regions of the temporal lobe, and a subsequent stage of phonological encoding, which is thought to involve frontal regions such as LIFG (Gordon, 1997, Levelt et al., 1999). The proposed serial order of these stages is in good accord with MEG findings that have detected responses in inferior temporal regions as preceding those in frontal regions (Salmelin et al., 1994, Levelt et al., 1998). Our findings suggest that while both LIFG and inferior temporal cortex are differentially involved in naming pictures relative to a pseudo picture baseline condition, the LIFG may play a role in lexical selection, but be less fundamental to the naming process once a target word has been selected, given that it can be strongly attenuated by overtraining on the stimuli. Analysis of aphasic naming errors has suggested that errors can arise at multiple stages of the naming process (Foygel and Dell, 2000, Schwartz et al., 2006), although thus far attempts to localize each stage to specific brain regions on the basis of lesions have had limited success. Nonetheless, fMRI studies of overt picture naming in aphasic populations may contribute greatly to dissociate between these stages by revealing brain activity differentially involved in the production of specific forms of errors, such as semantic and phonological paraphasias, perseverations, and neologisms.

To summarize our recommendations based on the main effects of this study (independent of multiple sessions), the use of untrained stimuli in a naming task would appear to give a more comprehensive picture of an individual’s language network than the use of trained (“overlearned”) stimuli. Furthermore, the subtraction of pseudo-object naming from real object naming is an effective means of isolating regions critically involved in word finding, as opposed to visual and motor processes.

Activation changes across sessions

We found that the activation magnitude for naming real, untrained pictures tends to decrease with each successive scanning session. Interestingly, the rate of decline in most regions, including SMA, LFusG, and LIFG, tended to be similar for both novel items and items that were repeated from scan to scan. In other words, we did not observe any significant long-term priming effect in hemodynamic activation magnitude as pictures were repeated at one-month intervals, in contrast to studies that have observed activation decrements for specific stimuli repeated at shorter intervals such as three days (Van Turennout et al., 2003). That activation tends to decline with repeated performance of the same task is not particularly surprising, as similar findings have been reported with, for example, working memory tasks (Milham et al., 2003, Jansma et al., 2001). The apparent lack of a difference in the practice effect between novel and repeated items indicates that the activation decline is most likely due to greater familiarity with the task of object naming, rather than priming on specific items. That is not to say, however, that individual item identity is a negligible factor. In the overlearned runs, there were no systematic trends for activation to naming real pictures over the four sessions. At first glance, this may seem to be a desirable aspect of using overlearned stimuli. However, our results indicate that the use of overlearned stimuli *stabilizes* the activation magnitude across sessions in most regions by *minimizing* it. The use of overlearned material produces a much more restricted activation pattern. Essentially, the activation for overlearned stimuli is already reduced as much as it can be, and so further repetitions do not make a difference. For this reason, overlearned stimuli are a less effective means to probe an individual's functional anatomy for word-finding. Therefore, we would recommend the use of novel materials in each imaging session in order to maximize the sensitivity of the fMRI technique.

Unfortunately, using all novel stimuli over several imaging sessions may result in a prohibitive number of stimuli being necessary, depending on how many sessions are desired. In this case, our findings indicate that the re-use of certain pictures does not drastically affect activation, as long as they are randomly interspersed with novel stimuli. In this way, a limited set of pictures can be stretched to cover more sessions. Furthermore, the repetition of some pictures allows for a very powerful comparison in patients who exhibit appreciable improvement in their anomia between sessions – one may compare the hemodynamic responses to the same pictures that elicit incorrect responses in earlier sessions but correct in later sessions (see e.g. Fridriksson et al., 2006).

The presence of a systematic decline in activation across sessions constrains the interpretation of sequential decreases that might be observed in longitudinal experiments with aphasic patients. A linear trend for decreasing activation may simply represent a task practice effect, as opposed to decreasing reliance on a particular brain region associated with network reorganization, given that such a decline is easily detected in healthy young control subjects upon repeated testing. On the other hand, we did not observe any systematic increases in activation in any area. Therefore, activation increases occurring over a timecourse of aphasia recovery are quite likely to reflect genuine reorganization and neural recruitment, rather than being artifacts of repeated task performance. Nonetheless, the behavioral variability in the performance of aphasic patients may help in interpreting any observed change, be it positive or negative. In this report, we have focused on systematic linear trends across sessions that are consistent across subjects. The predominant effect in this case appears to be a decline in activation associated with general task variability. In patients, one might expect the pattern of change to be more nonlinear, with the greatest changes in activation magnitude corresponding to the periods of maximum behavioral improvement, with comparatively small changes in activation magnitude expected when performance remains stable.

Reliability of individual activation patterns

Using the ICC as a metric of reliability, we evaluated four different contrasts, to determine which one gave the most reliable map of an individual's language network in picture naming. We found that the most reliable activation map was that of naming untrained real pictures, defined relative to an implicit fixation baseline. The reliability of naming overlearned pictures was much poorer, despite the fact that the systematic decline across sessions (which reduces reliability by increasing between-session variance) occurs only in the untrained runs. This reinforces our earlier statement that the use of overlearned stimuli stabilizes the activation magnitude by minimizing it, producing values that are closer to the noise level of measurement and hence less reliable. Therefore, we recommend that the activation magnitude for naming real pictures be used as the primary measure of interest for evaluating changes across sessions. Since this measure is relatively reliable, any observed changes are more likely to be attributable to genuine processes of neuronal plasticity.

Despite the fact that the subtraction of the pseudo condition gives a more specific picture of brain regions involved in lexical access as opposed to visual and motor processes, we do not recommend the use of this subtraction value as the measure of interest for studying between-session changes. The ICC values for the subtraction were lower than those for the activation magnitude of real naming alone. This is unsurprising, given that the variance in the subtraction is impacted by variance in both the real naming task and the pseudo naming task, and thus should be higher than the variance of either measure alone. Thus, if the goal is to create a map of an individual's language network in any given session, then the subtraction is the most appropriate measure. If, on the other hand, the goal is to create a map of regions that have significantly changed between sessions, then a direct contrast of real activation alone between sessions is more appropriate. Ideally, the regions highlighted in these analyses should overlap, if the observed plasticity is indeed related to improved picture naming abilities.

Conclusion

We have shown that longitudinal neuroimaging of overt picture naming has the power to reveal consistent patterns of activation in individual subjects as well as changes across sessions, using a statistically valid method of inter-session comparison. We have evaluated several approaches for the design and analysis of a longitudinal study using a control condition of pseudo object naming, thus generating an empirical basis for optimal design of aphasia recovery studies. Additionally, this study provides a baseline characterization of regions involved in overt picture naming under both overlearned and untrained conditions, thus dissociating regions that are necessarily involved in lexical access (even when well practiced) from those in which activation seems to be largely a function of cognitive effort required for naming. On the basis of these findings, we recommend against the exclusive use of overlearned materials, in favor of novel pictures, which produce a robust and reliable activation pattern in regions critical for lexical access. We also recommend the use of a control condition such as overt responses to pseudo-pictures, as this serves to better distinguish linguistic activation from low-level sensory and motor processes. Most importantly, for the evaluation of changes across sessions within individuals, we recommend that the activation magnitude from a control condition *not* be used, but rather that the activation magnitudes from the novel picture naming condition alone, acquired in multiple sessions, be directly contrasted with each other statistically.

Acknowledgments

This work was supported by the Intramural Research Program of the National Institute for Deafness and Other Communication Disorders, National Institutes of Health. We thank Vladimir Nechaev for preparation of the pseudo picture stimuli, and Carla Wettig for technical assistance.

References

Aron AR, Gluck MA, Poldrack RA. Long-term test-retest reliability of functional MRI in a classification learning task. Neuroimage 2006;29:1000–1006. [PubMed: 16139527]

Bates E, D'Amico S, Jacobsen T, Szekely A, Andonova E, Devescovi A, Herron D, Lu CC, Pechmann T, Pleh C, Wicha N, Federmeier K, Gerdjikova I, Gutierrez G, Hung D, Hsu J, Iyer G, Kohnert K, Mehotcheva T, Orozco-Figueroa A, Tzeng A, Tzeng O. Timed picture naming in seven languages. Psychon Bull Rev 2003;10:344–380. [PubMed: 12921412]

Blasi V, Young AC, Tansy AP, Petersen SE, Snyder AZ, Corbetta M. Word retrieval learning modulates right frontal cortex in patients with left frontal damage. Neuron 2002;36:159–170. [PubMed: 12367514]

Bonakdarpour B, Parrish TB, Thompson CK. Hemodynamic response function in patients with stroke-induced aphasia: implications for fMRI data analysis. Neuroimage 2007;36:322–331. [PubMed: 17467297]

Bookheimer SY, Zeffiro TA, Blaxton TA, Gaillard PW, Theodore WH. Activation of language cortex with automatic speech tasks. Neurology 2000;55:1151–1157. [PubMed: 11071493]

Cappa SF, Perani D, Grassi F, Bressi S, Alberoni M, Franceschi M, Bettinardi V, Todde S, Fazio F. A PET follow-up study of recovery after stroke in acute aphasics. Brain Lang 1997;56:55–67. [PubMed: 8994698]

Cardebat D, Demonet JF, De Boissezon X, Marie N, Marie RM, Lambert J, Baron JC, Puel M. Behavioral and neurofunctional changes over time in healthy and aphasic subjects: a PET Language Activation Study. Stroke 2003;34:2900–2906. [PubMed: 14615626]

Chee MW, Buckner RL, Savoy RL. Right hemisphere language in a neurologically normal dextral: a fMRI study. Neuroreport 1998;9:3499–3502. [PubMed: 9855306]

Chee MW, Lee HL, Soon CS, Westphal C, Venkatraman V. Reproducibility of the word frequency effect: comparison of signal change and voxel counting. Neuroimage 2003;18:468–482. [PubMed: 12595200]

Cohen MS, DuBois RM. Stability, repeatability, and the expression of signal magnitude in functional magnetic resonance imaging. J Magn Reson Imaging 1999;10:33–40. [PubMed: 10398975]

Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 1996;29:162–173. [PubMed: 8812068]

Crosson B, Fabrizio KS, Singletary F, Cato MA, Wierenga CE, Parkinson RB, Sherod ME, Moore AB, Ciampitti M, Holiway B, Leon S, Rodriguez A, Kendall DL, Levy IF, Rothi LJ. Treatment of naming in nonfluent aphasia through manipulation of intention and attention: A phase 1 comparison of two novel treatments. J Int Neuropsychol Soc 2007;13:582–594. [PubMed: 17521480]

Crosson B, Moore AB, Gopinath K, White KD, Wierenga CE, Gaiefsky ME, Fabrizio KS, Peck KK, Soltysik D, Milsted C, Briggs RW, Conway TW, Gonzalez Rothi LJ. Role of the right and left hemispheres in recovery of function during treatment of intention in aphasia. J Cogn Neurosci 2005;17:392–406. [PubMed: 15814000]

DeLeon J, Gottesman RF, Kleinman JT, Newhart M, Davis C, Heidler-Gary J, Lee A, Hillis AE. Neural regions essential for distinct cognitive processes underlying picture naming. Brain 2007;130:1408–1422. [PubMed: 17337482]

Eaton KP, Szaflarski JP, Altaye M, Ball AL, Kissela BM, Banks C, Holland SK. Reliability of fMRI for studies of language in post-stroke aphasia subjects. Neuroimage 2008;41:311–322. [PubMed: 18411061]

Etard O, Mellet E, Papathanassiou D, Benali K, Houde O, Mazoyer B, Tzourio-Mazoyer N. Picture naming without Broca's and Wernicke's area. Neuroreport 2000;11:617–622. [PubMed: 10718324]

Fernandez G, Specht K, Weis S, Tendolkar I, Reuber M, Fell J, Klaver P, Ruhlmann J, Reul J, Elger CE. Intrasubject reproducibility of presurgical language lateralization and mapping using fMRI. Neurology 2003;60:969–975. [PubMed: 12654961]

Fleiss, JL. The Design and Analysis of Clinical Experiments. Wiley; Hoboken, NJ: 1986.

Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. Magn Reson Med 1995;33:636–647. [PubMed: 7596267]

Foygel D, Dell GS. Models of impaired lexical Access in speech production. J Mem Lang 2000;43:182–216.

Fridriksson J, Morrow-Odom L, Moser D, Fridriksson A, Baylis G. Neural recruitment associated with anomia treatment in aphasia. Neuroimage 2006a;32:1403–1412. [PubMed: 16766207]

Fridriksson J, Morrow-Odom L, Moser D, Fridriksson A, Baylis G. Neural recruitment associated with anomia treatment in aphasia. Neuroimage. 2006b

Fridriksson J, Moser D, Bonilha L, Morrow-Odom KL, Shaw H, Fridriksson A, Baylis GC, Rorden C. Neural correlates of phonological and semantic-based anomia treatment in aphasia. Neuropsychologia 2007;45:1812–1822. [PubMed: 17292928]

Friedman L, Stern H, Brown GG, Mathalon DH, Turner J, Glover GH, Gollub RL, Lauriello J, Lim KO, Cannon T, Greve DN, Bockholt HJ, Belger A, Mueller B, Doty MJ, He J, Wells W, Smyth P, Pieper S, Kim S, Kubicki M, Vangel M, Potkin SG. Test-retest and between-site reliability in a multicenter fMRI study. Hum Brain Mapp 2008;29:958–972. [PubMed: 17636563]

Goodglass, H. Understanding Aphasia. Academic Press, Inc; San Diego, CA: 1993.

Goodglass, H.; Wingfield, A. Word-finding deficits in aphasia: Brain-behavior relations and clinical symptomatology. In: Goodglass, H.; Wingfield, A., editors. Anomia: Neuroanatomical and Cognitive Correlates. Academic Press; San Diego: 1997.

Gordon, B. Models of naming. In: Goodglass, H.; Wingfield, A., editors. Anomia: Neuroanatomical and cognitive correlates. Academic Press; San Diego: 1997.

Hillis AE, Kleinman JT, Newhart M, Heidler-Gary J, Gottesman R, Barker PB, Aldrich E, Llinas R, Wityk R, Chaudhry P. Restoring cerebral blood flow reveals neural regions critical for naming. J Neurosci 2006;26:8069–8073. [PubMed: 16885220]

Jansma JM, Ramsey NF, Slagter HA, Kahn RS. Functional anatomical correlates of controlled and automatic processing. J Cogn Neurosci 2001;13:730–743. [PubMed: 11564318]

Kan IP, Thompson-Schill SL. Effect of name agreement on prefrontal activity during overt and covert picture naming. Cogn Affect Behav Neurosci 2004;4:43–57. [PubMed: 15259888]

Karbe H, Thiel A, Weber-Luxenburger G, Herholz K, Kessler J, Heiss WD. Brain plasticity in poststroke aphasia: what is the contribution of the right hemisphere? Brain Lang 1998;64:215–230. [PubMed: 9710490]

Kimberley TJ, Birkholz DD, Hancock RA, VonBank SM, Werth TN. Reliability of fMRI during a continuous motor task: assessment of analysis techniques. J Neuroimaging 2008;18:18–27. [PubMed: 18190491]

Kiran S, Thompson CK. The role of semantic complexity in treatment of naming deficits: training semantic categories in fluent aphasia by controlling exemplar typicality. J Speech Lang Hear Res 2003;46:773–787. [PubMed: 12959459]

Knecht S, Drager B, Floel A, Lohmann H, Breitenstein C, Deppe M, Henningsen H, Ringelstein EB. Behavioural relevance of atypical language lateralization in healthy subjects. Brain 2001;124:1657–1665. [PubMed: 11459756]

Knecht S, Jansen A, Frank A, van Randenborgh J, Sommer J, Kanowski M, Heinze HJ. How atypical is atypical language dominance? Neuroimage 2003;18:917–927. [PubMed: 12725767]

Kong J, Gollub RL, Webb JM, Kong JT, Vangel MG, Kwong K. Test-retest study of fMRI signal change evoked by electroacupuncture stimulation. Neuroimage 2007;34:1171–1181. [PubMed: 17157035]

Kurland J, Naeser MA, Baker EH, Doron K, Martin PI, Seekins HE, Bogdan A, Renshaw P, Yurgelun-Todd D. Test-retest reliability of fMRI during nonverbal semantic decisions in moderate-severe nonfluent aphasia patients. Behav Neurol 2004;15:87–97. [PubMed: 15706052]

Leger A, Demonet JF, Ruff S, Aithamon B, Touyeras B, Puel M, Boulanouar K, Cardebat D. Neural substrates of spoken language rehabilitation in an aphasic patient: an fMRI study. Neuroimage 2002;17:174–183. [PubMed: 12482075]

Levelt WJ, Praamstra P, Meyer AS, Helenius P, Salmelin R. An MEG study of picture naming. J Cogn Neurosci 1998;10:553–567. [PubMed: 9802989]

Levelt WJ, Roelofs A, Meyer AS. A theory of lexical access in speech production. Behav Brain Sci 1999;22:1–38. [PubMed: 11301520]discussion 38–75

Martin PI, Naeser MA, Doron KW, Bogdan A, Baker EH, Kurland J, Renshaw P, Yurgelun-Todd D. Overt naming in aphasia studied with a functional MRI hemodynamic delay design. Neuroimage 2005;28:194–204. [PubMed: 16009568]

Meinzer M, Flaisch T, Obleser J, Assadollahi R, Djundja D, Barthel G, Rockstroh B. Brain regions essential for improved lexical access in an aged aphasic patient: a case report. BMC Neurol 2006;6:28. [PubMed: 16916464]

Miezin FM, Maccotta L, Ollinger JM, Petersen SE, Buckner RL. Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. Neuroimage 2000;11:735–759. [PubMed: 10860799]

Milham MP, Banich MT, Claus ED, Cohen NJ. Practice-related effects demonstrate complementary roles of anterior cingulate and prefrontal cortices in attentional control. Neuroimage 2003;18:483–493. [PubMed: 12595201]

Papademetris, X.; Jackowski, AP.; Schultz, RT.; Staib, LH.; Duncan, JS. Integrated intensity and point-feature non-rigid registration. In: Barillot, C.; Haynor, D.; Hellier, P., editors. Medical image computing and computer-assisted intervention. Springer; Saint_malo, France: 2004. p. 763-770.

Postman-Caucheteux WA, Birn R, Pursley R, Butman J, Solomon J, Picchioni D, McArdle JJ, Braun AR. Single-Trial fMRI shows contralesional activity linked to overt naming errors in chronic aphasic patients. J Cogn Neurosci.

Postman-Caucheteux WA, Hoffman S, Picchioni D, McArdle JJ, Birn R, Braun AR. Distinct activation patterns for accurate vs. inaccurate naming of actions and objects: An fMRI study with stroke patients with chronic aphasia. Brain and Language 2007;103:150–151.

Rau S, Fesl G, Bruhns P, Havel P, Braun B, Tonn JC, Ilmberger J. Reproducibility of activations in Broca area with two language tasks: a functional MR imaging study. AJNR Am J Neuroradiol 2007;28:1346–1353. [PubMed: 17698539]

Rosen HJ, Petersen SE, Linenweber MR, Snyder AZ, White DA, Chapman L, Dromerick AW, Fiez JA, Corbetta MD. Neural correlates of recovery from aphasia after damage to left inferior frontal cortex. Neurology 2000;55:1883–1894. [PubMed: 11134389]

Salmelin R, Hari R, Lounasmaa OV, Sams M. Dynamics of brain activation during picture naming. Nature 1994;368:463–465. [PubMed: 8133893]

Saur D, Lange R, Baumgaertner A, Schraknepper V, Willmes K, Rijntjes M, Weiller C. Dynamics of language reorganization after stroke. Brain 2006;129:1371–1384. [PubMed: 16638796]

Schwartz MF, Dell GS, Martin N, Gahl S, Sobel P. A case-series test of the interactive two-step model of lexical access: Evidence from picture naming. J Mem Lang 2006;54:228–264.

Specht K, Willmes K, Shah NJ, Jancke L. Assessment of reliability in functional imaging studies. J Magn Reson Imaging 2003;17:463–471. [PubMed: 12655586]

Szekely A, D'Amico S, Devescovi A, Federmeier K, Herron D, Iyer G, Jacobsen T, Arevalo AL, Vargha A, Bates E. Timed action and object naming. Cortex 2005;41:7–25. [PubMed: 15633703]

Szekely A, D'Amico S, Devescovi A, Federmeier K, Herron D, Iyer G, Jacobsen T, Bates E. Timed picture naming: extended norms and validation against previous studies. Behav Res Methods Instrum Comput 2003;35:621–633. [PubMed: 14748507]

van Turennout M, Bielamowicz L, Martin A. Modulation of neural activity during object naming: effects of time and practice. Cereb Cortex 2003;13:381–391. [PubMed: 12631567]

Vitali P, Abutalebi J, Tettamanti M, Danna M, Ansaldo AI, Perani D, Joanette Y, Cappa SF. Training-induced brain remapping in chronic aphasia: a pilot study. Neurorehabil Neural Repair 2007;21:152–160. [PubMed: 17312090]

Wei X, Yoo SS, Dickey CC, Zou KH, Guttmann CR, Panych LP. Functional MRI of auditory verbal working memory: long-term reproducibility analysis. Neuroimage 2004;21:1000–1008. [PubMed: 15006667]

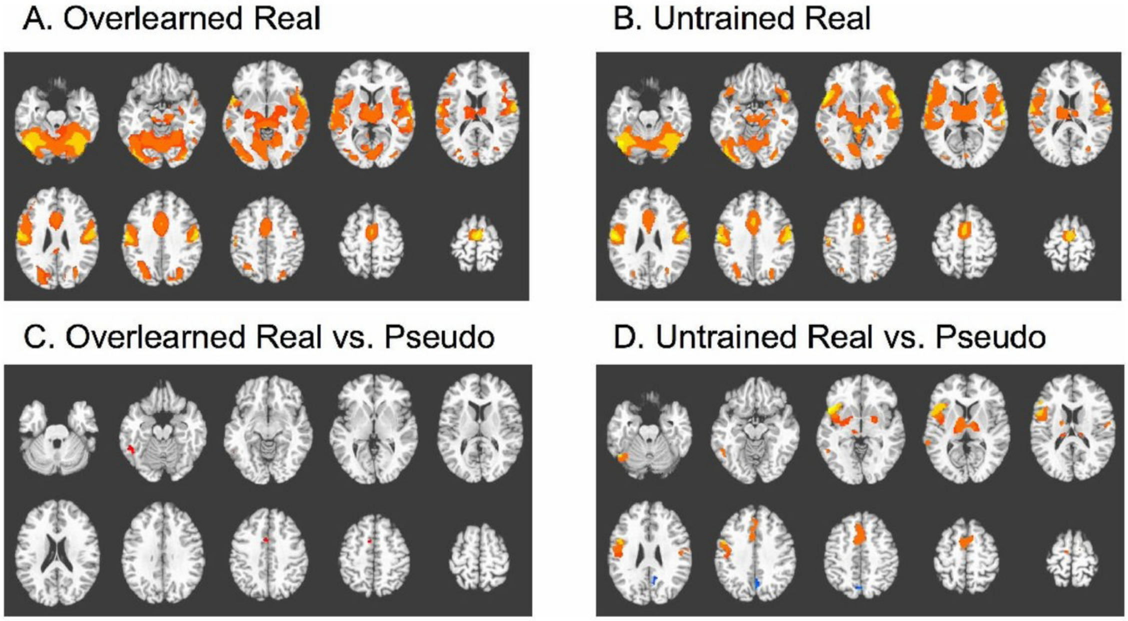

Figure 1. Main effects over all sessions

A: Regions showing a significant response to onset of real pictures during overlearned runs. B: Regions showing a significant response to onset of real pictures during untrained runs. C: Regions showing a significantly greater response to real pictures relative to pseudo pictures during overlearned runs (see Table 1A). D: Regions showing a significantly greater response to real pictures relative to pseudo pictures during untrained runs (see Table 1B). All montages shown depict 10 axial slices, ranging from z = −20 to z = +61, in increments of 9mm (except for figure 1C, which ranges from −25 to +56). Colors are scaled from red to yellow reflecting the relative magnitude of activation (or dark blue to light blue for negative activations). See Methods for thresholding information.

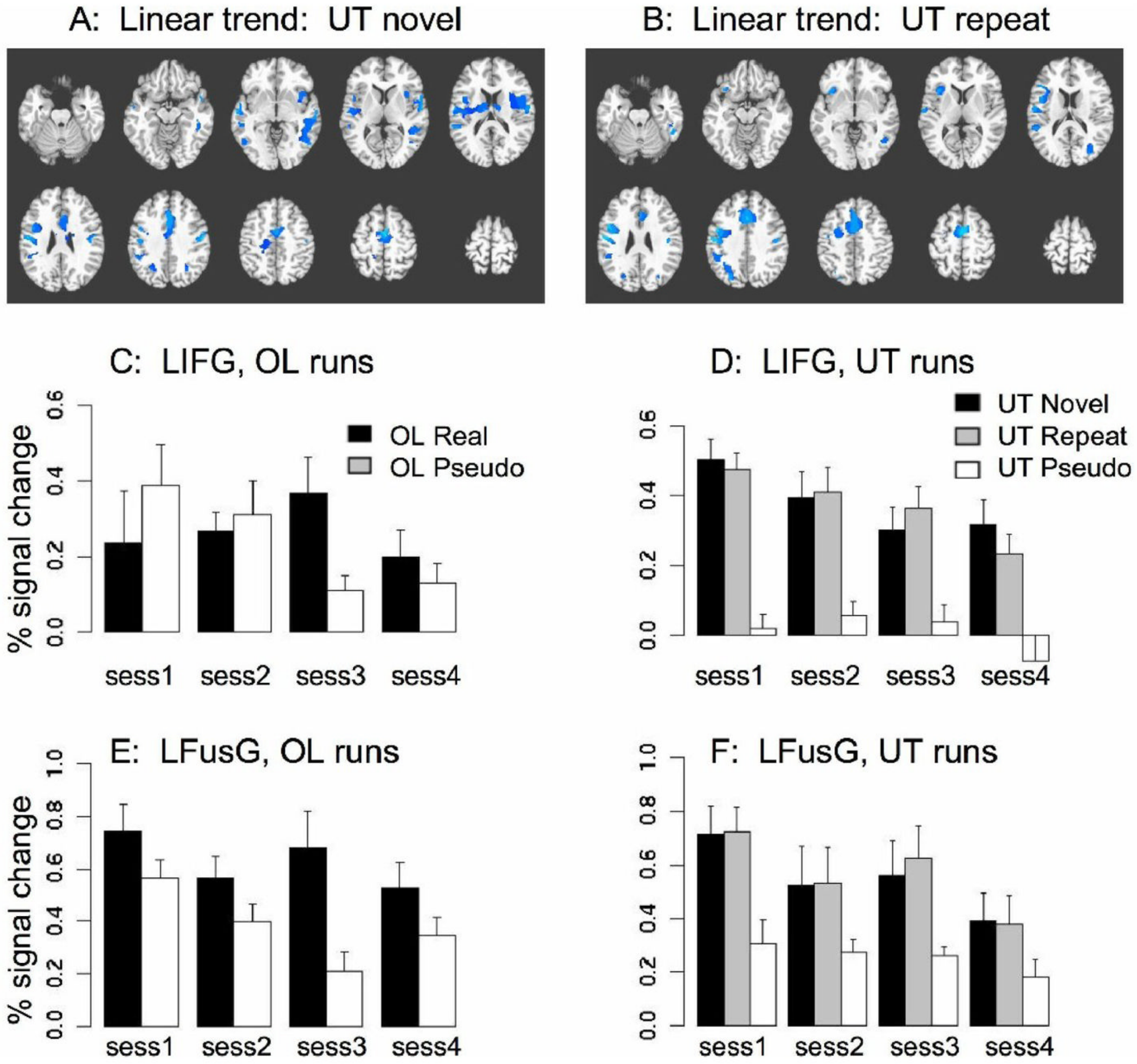

Figure 2. Trends of activation changes across sessions

A: Regions showing a significant negative trend in activation magnitude for novel items in untrained (UT) runs across the four sessions (see Methods). No positive trends were seen. Slice locations are identical to figure 1. B: Regions showing a significant decline in activation magnitude for the repeated set of pictures in untrained runs across the four sessions. The sets of regions showing a decline for novel and repeated items are similar but not identical; however, no voxels exhibited a significant difference in the rate of decline for novel vs. repeated pictures. (Note that in the first session, all pictures are effectively novel). C: Activation magnitude for overlearned (OL) real and pseudo pictures across the four sessions, in Left Inferior Frontal Gyrus. Values are percent signal change from a spherical ROI of radius 8mm, averaged over voxels and subjects, centered at Talairach coordinates −37, +13, 0. D: Activation for the LIFG ROI in untrained runs across four sessions, for novel real pictures, repeated real pictures, and pseudo pictures. Note that all pictures are actually novel in the first session – the repeated set are repeated during the subsequent sessions. E: Activation in the left fusiform gyrus (LFusG) for overlearned runs. ROI coordinates −45, −55, −18. F: Activation in the left fusiform gyrus for untrained runs.

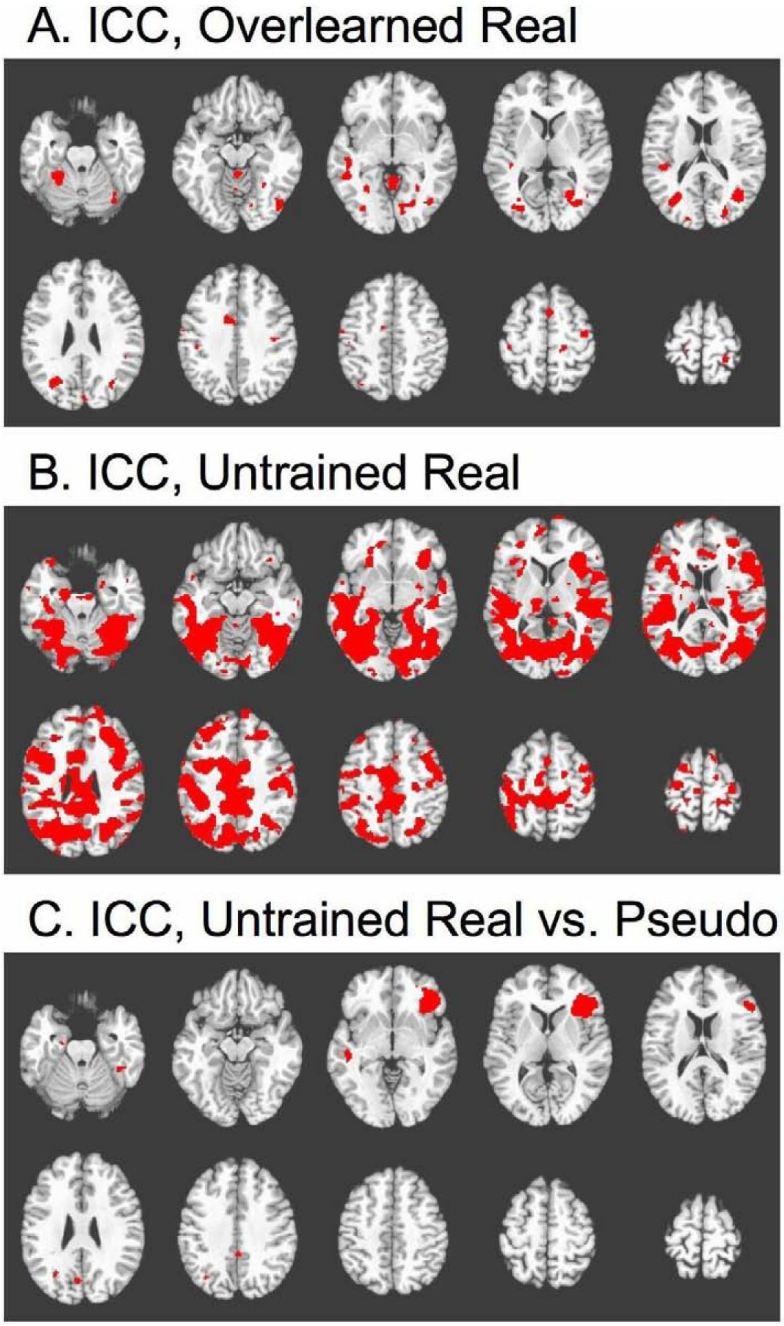

Figure 3. Reliability across sessions

All maps depict the Intraclass Correlation Coefficient (ICC) at each voxel (see Methods), binary thresholded at a value of 0.7, indicating high reliability of an individual subject's activation magnitude across multiple scanning sessions. Slice locations are identical to figure 1. A: ICC for responses to overlearned real pictures. B: ICC for untrained real pictures. C: ICC for the subtraction of untrained real vs. pseudo pictures.

Table 1

Clusters of activation in main effect comparisons. Areas are named by visual inspection of the extent of activation and the center of mass coordinates. BA indicates Brodmann areas. Volume is in voxels, which are 2×2×2mm. Coordinates represent cluster center of mass in Talairach atlas space.

area name	BA	Volume	x	y	z
A: overlearned real vs pseudo					
L Fusiform Gyr.	37	101	−49	−51	−13
Ant. Cingulate, SMA	6	83	−4	3	42
B: untrained real vs pseudo					
L Inf. Frontal Gyr., putamen	44,45,6	1955	−43	6	12
L SMA, pre-SMA	6	1176	−2	11	44
Thalamus		984	−2	−12	7
L Fusiform Gyr.	37	226	−44	−57	−17
R Post-central Gyr.	4,3	130	53	−9	15
R Cerebellum		113	32	−50	−30
L Superior/Middle Temporal Gyr.	22,21	67	−60	−36	5
* Precuneus	7	208	1	−66	36

* The activation in the precuneus was negative, i.e. real pictures induced a larger negative BOLD signal change than pseudo pictures.

Table 2

Clusters exhibiting a significant linear trend for activation changes across sessions (see Methods). Areas are named as outlined in Table 1.

area name	BA	Volume	x	y	z
A: trend for overlearned real					
L Pre-central Gyr.	4	107	−46	−12	40
B: trend for overlearned pseudo					
Ant. Cingulate, SMA	24,6	497	6	4	42
R Inf. Frontal Gyr.	44,45	167	51	9	17
R Mid. Temp. Gyr, Sup. Temp. Gyr.	22	129	53	−37	−1
R Mid. Frontal Gyr.	9	126	36	30	25
C: trend for untrained novel					
R Insula		1511	43	−4	16
L Insula		1224	−43	−13	15
Ant. Cingulate, SMA	24, 6	1117	2	3	40
R Mid. Temp. Gyr., Inf. Temp. Gyr.	22,37	973	47	−39	−1
L Inf. Frontal Gyr.	44	379	−40	3	21
L Mid. Temp. Gyr., Sup. Temp. Gyr.	22	247	−51	−37	15
L Ant. Cingulate	31	229	−17	−27	44
L Angular Gyr.	7,39	154	−25	−59	36
L Mid. Temp Gyr., Inf. Temp. Gyr.	22,37	84	−48	−57	−2
R Angular Gyr.	7,39	69	26	−58	35
D: trend for untrained repeat					
L Inf. Frontal Gyr., insula	44,45,47	1860	−40	1	22
Cingulate, SMA	24,6	1510	−1	11	41
L Intraparietal Sulcus	7	359	−26	−65	36
L Intraparietal Sulcus	13	295	−48	−40	26
R Mid. Occ. Gyr.	31	189	29	−71	18
R Mid. Temp. Gyr.	37	93	44	−58	0
R. Fusiform Gyr.	37	87	49	−41	−18
R Post-central Gyr.	4,3	80	44	−16	31

All clusters in this table are negative, as activation magnitude *decreased* across sessions.