Applying FSL to the FIAC data: model-based and model-free analysis of voice and sentence repetition priming

FMRIB Technical Report TR06CB1
(A related paper has been accepted for publication in Human Brain Mapping)

C.F. Beckmann, M. Jenkinson, M.W. Woolrich, T.E.J. Behrens, D.E.F. Flitney, J.T. Devlin and S.M. Smith

Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB),
Department of Clinical Neurology, University of Oxford, John Radcliffe Hospital,
Headley Way, Headington, Oxford, UK
Corresponding author is Christian F. Beckmann: beckmann@fmrib.ox.ac.uk

Abstract

This paper presents results obtained from applying various tools from FSL (FMRIB Software Library) to data from the repetition priming experiment used for the HBM'05 Functional Image Analysis Contest. We present analyses from the model-based General Linear Model (GLM) tool (FEAT) and from the model-free independent component analysis tool (MELODIC). We also discuss the application of tools for the correction of image distortions prior to the statistical analysis and the utility of recent advances in FMRI time series modelling and inference such as the use of optimal constrained HRF basis function modelling and mixture modelling inference. The combination of HRF and mixture modeling, in particular, revealed that both sentence content and speaker voice priming effects occurred bilaterally along the length of the STS. These results suggest that both are processed in a single underlying system without any significant asymmetries for content vs. voice processing.

1 Introduction

The techniques available for the interrogation and analysis of neuroimaging data have a large influence in determining the flexibility, sensitivity and scope of neuroimaging experiments.

The Functional Image Analysis Contest (FIAC), organised as part of the 11th Annual Meeting of the Organization for Human Brain Mapping (Toronto, 2005), was designed to investigate the interactions between the neuroscientific interpretation of results and the particular type of data analysis. In this paper we present analysis results based on the tools available in the FMRIB Software Library (FSL).

The aim of the original FMRI experiment was to investigate cortical regions sensitive to auditory sentence comprehension as well as the voice of the speaker using a repetition priming paradigm. The experiment manipulated the sentence content (repeated vs. difference sentences) and the speaker's identity (same vs. different speaker) in a 2 x 2 factorial design; participants passively listened to auditory sentences in four conditions (see Dehaene-Lambertz et al., 2006, for details):

1. Different sentences, different speakers (DStDSp). This condition served as the baseline for evaluating priming effects.
2. Same sentence, different speakers (SStDSp) was used to evaluate repetition priming effects due to repeated sentences.
3. Different sentence, same speakers (DStSSp) was used to evaluate repetition priming effects due to hearing the same voice across sentences.
4. Same sentence, same speaker (SStSSp). This condition represents the interaction between sentence content and voice.

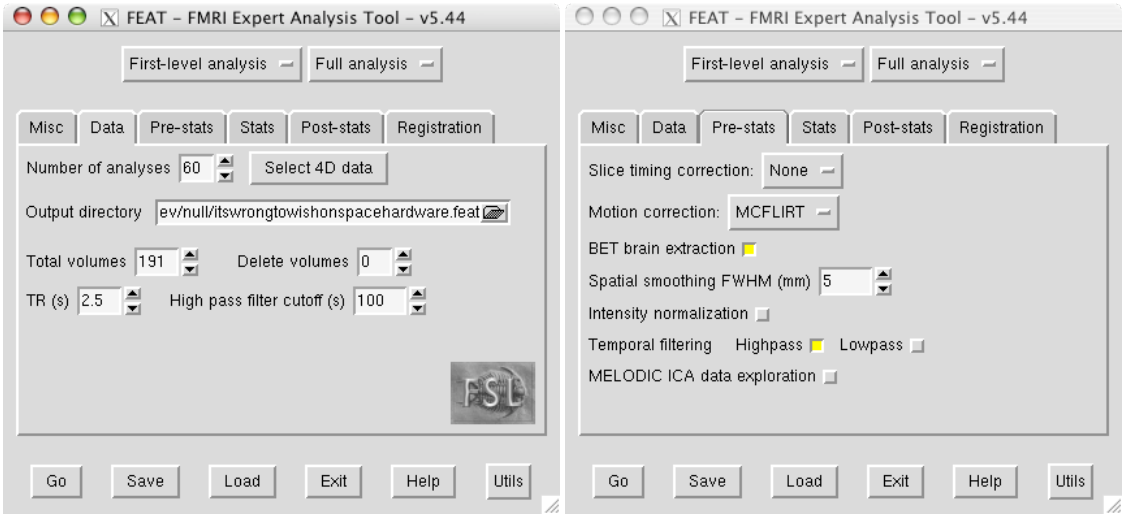


Figure 1: Setup of pre-statistical processing steps in the FEAT GUI

The experiment was run four times for each subject, twice using a blocked design and twice using a pseudo-randomised event-related design. In the blocked design, each 20s block of sentences was separated by a 9 sec block of silent rest. There was no comparable "resting" condition in the event-related design. Previous studies have shown that at a neural level, repetition priming tends to manifest as a reduction in activation in auditory association areas (Dehaene et al., 2001; Vuilleumier et al., 2002), possibly due to habituation of neuronal responses (Desimone, 1996). Consequently, by manipulating the relation between sentences, one can investigate sensitivity to sentence content and the spoken voice.

1.1 FMRIB Software Library - Overview

FMRIB's Software Library (FSL) is a comprehensive library of image analysis and statistical tools for FMRI, MRI and DTI brain imaging data, freely available as both source and binary distributions for various computing platforms. Most of the tools can be run both from the command line and as GUIs ("point-and-click" graphical user interfaces). As well as analysis tools, FSL also includes an intuitive yet powerful 3D/4D image display tool, FSLView, including multiple orthogonal or lightbox views, 3D rendering, time series display, image editing and histogram viewing. For more detail on the FSL analysis tools, see www.fmrib.ox.ac.uk/fsl and Smith et al., 2004 (also see table 1 for a list of FSL tools).

2 Model-Based FMRI Analysis using FEAT

The *FMRI Expert Analysis Tool* (FEAT) is an advanced GLM-based FMRI analysis tool with a straightforward but powerful GUI, carrying out data preprocessing (including slice timing correction and motion correction); first-level GLM time-series analysis with prewhitening; registration to subject-specific structural images and standard space; and mixed-effects group analysis using Bayesian estimation techniques.

Complete analysis for a single simple FMRI experiment can often be set up in less than 1 minute, whilst a more complex experiment typically need take no longer than 5 minutes to set up. Multiple experiments having the same design can be analysed with a single GUI setup. A complete setup can be saved to file, for easy reloading, amendment and re-running later, or to be used in script-based analysis of multiple experiments. The FEAT program produces a web page analysis report (including activation overlay images, activation cluster tables, time-course plots of data vs model, registration overlay images and an "Analysis Methods" paragraph describing the exact analysis carried out, including references and parameter value settings).

2.1 Data preprocessing

All 60 FIAC functional data sets (block and event-related) were pre-processed using the following analysis steps: motion correction using MCFLIRT (Jenkinson et al., 2002); non-brain removal using BET (Smith, 2002); spa-

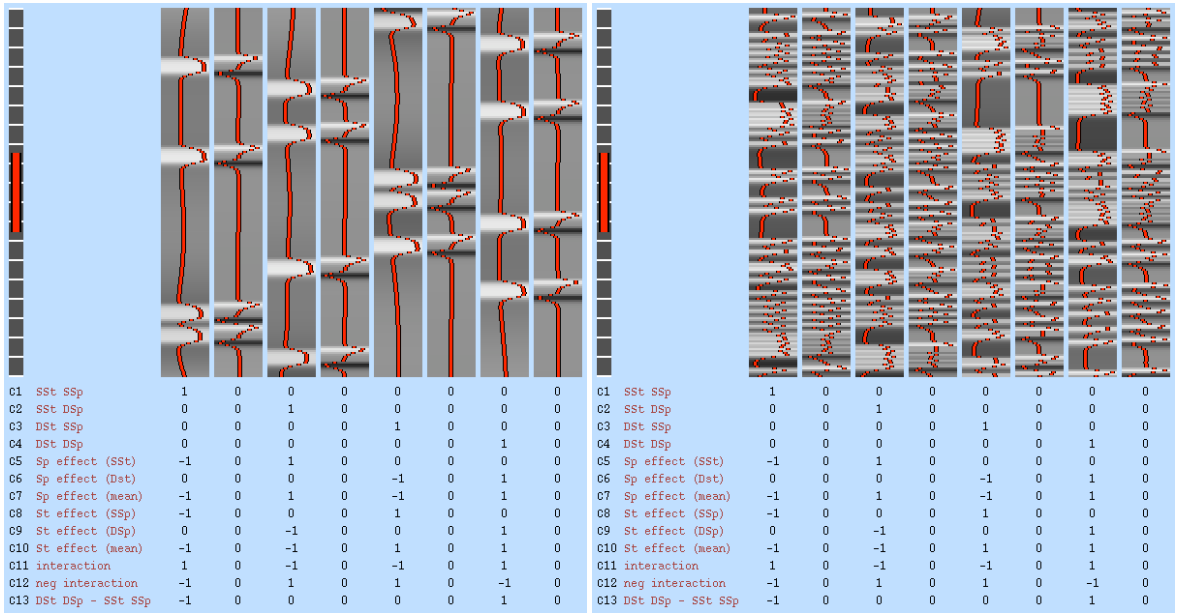


Figure 2: First-level design matrices used for the model-based (GLM) analysis. The columns of the design matrices show graphical representation of the GLM regressors. For the FIAC data there are 4 regressors modelling the BOLD changes during the 4 experimental conditions, together with temporal derivatives (each regressor is immediately followed by its temporal derivative), which are used to account for mis-specification of the haemodynamic lag. The design matrix plots also show the width of the temporal high-pass filter (far left; any low-frequency fluctuation with periodicity larger than the indicated red bar is removed from the data). At the bottom are the contrast vectors used for testing different primary or differential BOLD changes.

tial smoothing using a Gaussian kernel of full-width-at-half-maximum (FWHM) of $5mm$; mean-based intensity normalisation of all volumes by the same factor; high-pass temporal filtering (Gaussian-weighted least-squares straight line fitting, sigma=$50s$). All this pre-processing is easily set up in the "Pre-stats" part of the FEAT GUI (fig. 1).

2.2 First-level General Linear Modelling

The data from the FIAC block and event-related experiments were supplied with paradigm timing files. These were easily converted into the timing file format required by FEAT, with one timing file per explanatory variable (EV) - a column in the design matrix. In total 8 EVs were used for each design matrix, the four different conditions together with a temporal derivative per primary regressor, to account for potential mis-specification of the haemodynamic delay. Figure 2 shows example first-level design matrices generated by FEAT for one block-design data set and one event-related data set.

2.2.1 FILM parameter estimation

Time-series statistical analysis was carried out using FILM (FMRIB's Improved Linear Model), which corrects each voxel's timeseries for temporal smoothness. Unless this smoothness is correctly accounted for, the time-series analysis is at best inefficient (in terms of sensitivity to true activation) and at worst statistically invalid (in terms of inflated type-1 error due to incorrect degrees-of-freedom, DOF). To achieve this, FILM uses voxel-wise autocorrelation estimation involving temporal high-pass preprocessing (to remove the worst of the large-scale, non-stationary components and low frequency noise), estimation and regularisation of the auto-correlation coefficients in the frequency domain, followed by spatial regularisation of the estimates using non-linear smoothing, within tissue type. This approach has been demonstrated to have close to zero bias at probability levels as low as 0.00001 while obtaining close to optimally efficient estimation of the model parameters, giving greater sensitivity to activation detection (Woolrich et al., 2001).

FMRIB :	Functional Magnetic Resonace Imaging of the Brain Centre, Dept. of Clinical Neurology, University of Oxford, Oxford, UK (see http://www.fmrib.ox.ac.uk).
FSL :	*FMRIB's Software Library*; a collection of tools for the analysis of neuroimaging data.
BET :	*Brain Extraction Tool*; segments brain from non-brain in structural and functional data, and models skull and scalp surfaces.
PRELUDE :	*Phase Region Expanding Labeller for Unwrapping Discrete Estimates*; performs 3D phase unwrapping of fieldmap images.
FUGUE :	*FMRIBs Utility for Geometrically Unwarping EPIs*; performs unwarping of an EPI image based on unwrapped fieldmap data.
FLIRT :	*FMRIB's Linear Image Registration Tool*; fully automated robust and accurate tool for linear (affine) intra- and inter-modal brain image registration.
MCFLIRT :	*Motion Correction using FMRIB's Linear Image Registration Tool*; intra-modal motion correction tool.
FEAT :	*FMRI Expert Analysis Tool*; model-based FMRI analysis tool with a powerful GUI. FEAT provides a single consistent interface for setting up all aspect of pre-processing, linear modelling, registration and inference.
FILM :	*FMRIB's Improved Linear Model*; tool for estimating first-level GLMs using pre-ehitening.
FLAME :	*FMRIB's Local Analysis of Mixed Effects*; tool for estimating higher-level mixed-effects (between session/subject) GLMs using Bayesian estimation techniques.
FLOBS :	*FMRIB's Linear Optimised Basis Sets*; program for generating optimal basis sets for use in HRF convolution in FMRI linear modelling such as in FEAT.
MELODIC :	*Multivariate Exploratory Linear and Optimised Decomposition into Independent Components*; ICA-based model-free analysis of 4D Data.

Table 1: List of FSL tools.

2.2.2 Contrasts and T/Z-statistics

Each regressor (column) in the design matrix generates a parameter estimate as a result of the model-fitting. Questions (such as "where is the response to the different-speaker condition greater than the response to same-speaker") are asked by defining "contrasts" of parameter estimates (COPEs) which specify how to take linear combinations of the parameter estimates. The resulting COPEs can be then tested for statistical significance (against the null-hypothesis of zero effect size) by converting into T and/or F statistics, taking the estimated error in the fitting into account.

For the FIAC data, the primary questions of interest are:

- Primary areas of activation under the 4 different conditions (SStSSp, SStDSp, DStSSp, DStDSp; contrasts C1-C4).
- The effect of the speaker (under same sentence condition, different sentence condition and averaged over these; contrasts C5-C7). Of particular interest here is contrast C6, the direct comparison of the DStSSp and DStDSp conditions (voice repetition priming).
- The effect of sentence (under same speaker condition, different speaker condition and averaged; contrasts C8-C10). Of particular interest here is contrast C9, the direct comparison of the SStDSp and DStDSp conditions (sentence repetition priming).
- Positive (C11) or negative (C12) sentence-speaker interaction and the overall maximum effect size of the repetition suppression (*cognitive interaction*: BOLD amplitude decrease during SStSSp when compared to DStDSp; contrast C13).

2.2.3 Registration and correction for geometric EPI distortions

FEAT carries out pre-processing and timeseries statistics before estimating the alignment transformation between the low-resolution FMRI data and standard (e.g. MNI152) space. By default FEAT achieves this using FLIRT (Jenkinson et al., 2002) to first align an example low-resolution FMRI image to the same subject's high-resolution T1-weighted structural image, and then registering the structural image to a standard space average. In Smith et al., 2005, we showed the added value of this two-stage approach in terms of decreased registration-induced between-subject variance, compared with registering directly from FMRI to standard space.

In addition to the linear transforms applied in this way, it is possible to unwarp FMRI EPI (echo-planar imaging) data prior to registration to the structural image. Distortion and dropout of EPI-based functional images is

a particular problem for high field (3T and higher) MR scanners such as the one used for the acquisition of the FIAC data. Such artefacts are particularly noticeable in the inferior temporal and frontal lobes and limit the use of standard FMRI or diffusion imaging techniques in these areas (Jezzard and Balaban, 1995; Devlin et al., 2000).

In addition, the distortions can also lead to global errors in registration, causing errors in the spatial localisation of activations from any brain region, including those where there is little or no distortion present. If one acquires B0 field-map images one can use these to estimate and correct the EPI distortion.

In the case of the FIAC data the EPI images were acquired using an echo time of $35ms$ which is likely to result in substantial distortions at 3T. However, the FIAC data includes raw fieldmap images, which we used to undistort the FMRI data before running the linear alignments. The fieldmap images were first phase-unwrapped using PRELUDE (Phase Region Expanding Labeller for Unwrapping Discrete Estimates, Jenkinson, 2003). This corrects for the fact that MR phase measurements are wrapped within the range 0-2π. A corrected fieldmap was calculated from the difference of two unwrapped phase images. This fieldmap was then registered to the FMRI EPI data (using 6 DOF), the unwarping was applied to an example EPI image using FUGUE (FMRIB's Utility for Geometrically Unwarping EPIs), and the undistorted EPI image was registered to the subject's structural (using 7 DOF). In the registration steps, the signal loss (resulting from through-slice field gradients) was calculated and used as a cost function mask to exclude voxels in the image where the signal loss was severe (Jenkinson, 2004).

All of the above was carried out using command-line programs supplied with FSL, and is inserted as a "correction" to the FLIRT-estimated registrations after the first-level FEAT analyses have completed, before running cross-subject FEAT analyses.

2.3 Higher-level GLM

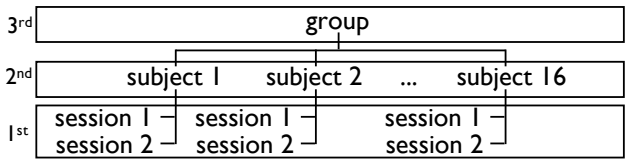


Figure 3: Hierarchical multi-level GLM approach for mixed-effects analysis of the FIAC data: at the first level, each individual session's data is analysed using first-level designs described in figure 2. At the intermediate level each subject's different sessions are combined into mixed-effects subject-specific means. At the third (top) level these subject means are then used to calculate mixed-effects group mean estimates.

In the present study the challenge is to address questions about activation effects in generic populations of subjects, i.e. analyse the data in such a way as to allow for hypothesis tests at the population level. In order to be able to generate results that accurately extend to the wider population, we need to account for the fact that these limited samples from the population are random quantities with associated random effects variances on top of the within-session fixed-effects variance contribution (Holmes and Friston, 1998; Beckmann et al., 2003a).

In FEAT, such *mixed-effects* group statistics are generated using a hierarchical summary statistics approach where the different levels of the hierarchy are separate GLMs. That is, the within-session first level, the within-subject-cross-session second level and the cross-subject third level (Beckmann et al., 2003a; Woolrich et al., 2004a). Note that in many studies, such as those with a single session per subject, only a two-level approach is needed.

A practical fully Bayesian inference technique is used for higher-level analyses in FEAT, referred to as FLAME (FMRIB's Local Analysis of Mixed Effects, Woolrich et al., 2004a). FLAME is a two-stage process: a fast approach using maximum a posterior estimates and a slower, more accurate approach using Markov Chain Monte Carlo (MCMC). By taking into account the correct lower-level summary statistics (estimates of the mean, variance and degrees-of-freedom), a substantial improvement in higher-level activation estimation accuracy can be obtained (see Woolrich et al., 2004a, for details).

For the FIAC analyses, each subject's different sessions are combined into a subject-specific mean at second level. These means are all estimated as part of one single second-level design (as opposed to a separate second-level design for each subject) because the number of sessions per subject is too small to obtain a robust estimate of the between-session variance separately for each subject; hence the session variance estimation is pooled across subjects to improve accuracy (at the expense of assuming similar between-session variance between subjects). At the third (top) level these subject means are then used to calculate mixed-effects group mean estimates. The higher-

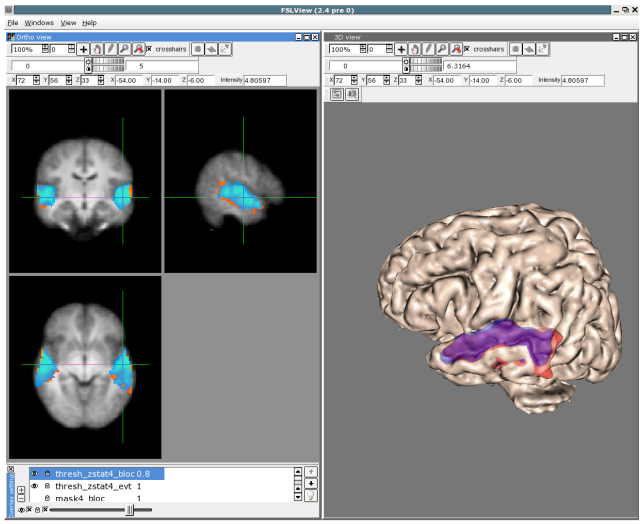


Figure 4: Cross-subject mean (contrast 4: different sentences different speakers). Results are shown for the block-design data (blue) and the event-related design (red). All 2D image slices are shown left-right reversed. 3D rendering carried out using FSLView.

level analyses are run separately for each contrast specified in the first-level designs. Figure 3 gives a graphical representation of the multi-level hierarchical GLM approach used for the model-based analysis of the FIAC data.

2.4 Model-based Group Mean results

The group mean mixed-effects Z (Gaussianised T/F, Jenkinson and Woolrich, 2000) statistic images were thresholded using clusters determined by $Z > 2.3$ and a (corrected) cluster significance threshold of $P = 0.05$ assuming a Gaussian random field for the Z statistics (Worsley et al., 1992).

Figure 4 shows the cross-subject mean for the different-sentence-different-speaker condition (contrast 4), averaged over all 16 subjects (blue:block-design; red: event-related design). In the context of this experiment this condition is typically referred to as the "baseline" condition. This condition generates the highest level of activation, i.e. largest estimated BOLD amplitude modulation, and repetition priming in other conditions is expected to cause suppression of the activation level relative to this baseline. Significant BOLD activation was observed bilaterally in Heschl's gyrus (HG), planum temporale, planum polare, lateral superior temporal gyrus, superior temporal sulcus (STS), and middle temporal gyrus. These regions encompass primary and secondary auditory cortices as well as extensive auditory association cortex and correspond to previous studies reporting auditory sentence processing (Cardillo et al., 2004; Crinion et al., 2003). There were no substantial differences in the pattern of activation between the event-related and blocked versions. For the differential contrasts of interest, table 2 lists all significant clusters and the location of the maximum statistics in MNI152 space.

2.4.1 Priming effects due to speaker repetition

Compared to the sentence effect, repetition priming due to the repetition of the speaker is expected to elicit a smaller reduction in BOLD amplitude. Indeed, there was no significant main effect of voice repetition (contrast 6: DStSSp $<$ DStDSp) in either the event-related or blocked versions when using the default null-hypothesis testing approach with Gaussian-random field theory. A significant cluster was seen for the simple main effect of voice repetition (contrast 5: SStSSp $<$ SStDSp) in the left STS (see figure 5; event-related data only).

2.4.2 Priming effects due to sentence repetition

BOLD amplitude decreases due to the repetition of sentences are explicitly tested by contrasts 8-10. Figure 6 shows the primary sentence effect of interest, the priming effect of different speakers presenting the same sentence (contrast 9: SStDSp $<$ DStDSp). BOLD amplitude decreases due to the repetition of sentences were only found in the left anterior STS and did not differ between the event-related and blocked designs.

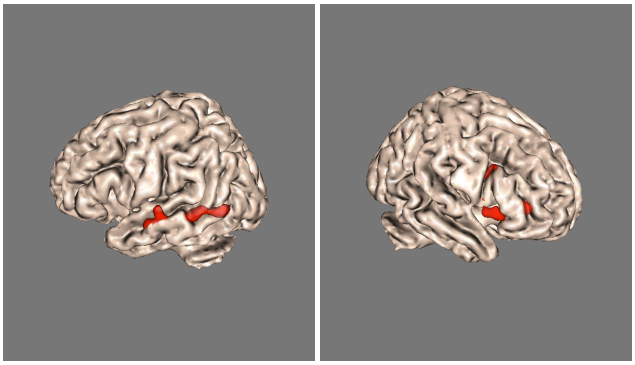


Figure 5: Repetition effect of speakers (contrast 5: SStSSp $<$ SStDSp) estimated from the event-related data sets (red) when sentence content is not changed between different speakers. No supra-thresholded voxels were found in the block-design data.

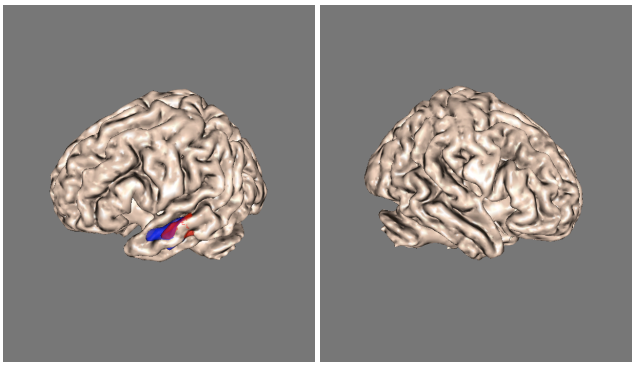


Figure 6: Repetition effect of sentences (contrast 9: SStDSp $<$ DStDSp) estimated from the block-design data (blue) and the event-related data sets (red). No supra-thresholded voxels were found in the right hemisphere.

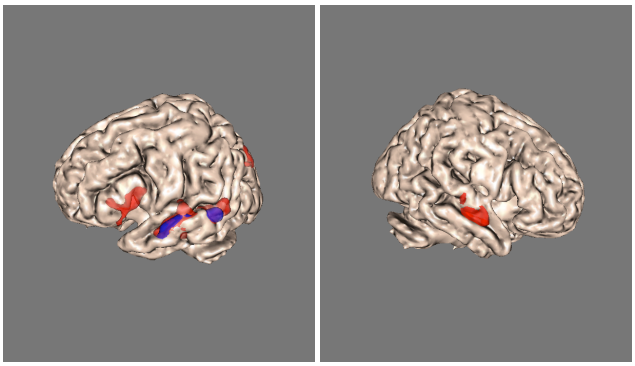


Figure 7: Effect of the cognitive interaction: maximum effect of repetition surpression (contrast 13: SStSSp $<$ DStDSp).

It is interesting to note that the cluster thresholding is hiding a fairly strong (but spatially smaller) BOLD amplitude change in a more posterior area of left STS, another in right anterior STS, and a slightly weaker priming effect in Broca's area. As a result, the GRF thresholded activation pattern of contrast 13 (figure 8) appears far more extensive than just sentence priming. This, however, is likely to be just a thresholding artefact in that the extra statistical power due to voice priming is increasing average Z-statistics within a magnitude and they are then spreading into each other, passing the spatial threshold, and appearing as novel priming effects, even though they really aren't.

2.4.3 Interaction effects between sentence content and voice

The simple cognitive interaction effect (i.e. overall effect of repetition priming) is shown in figure 7 (contrast 13: SStSSp $<$ DStDSp). Unlike the previous contrasts (e.g. contrast 4), there were more differences between the event-related and blocked versions. Priming was more extensive in the event-related design with significant reductions in BOLD signal seen in the left hemisphere along the length of the left superior temporal sulcus (STS), in lateral HG, all subdivisions of ventrolateral prefrontal cortex (pars opercularis, pars triangularis, and pars orbitalis), and posterior intra-parietal sulcus (IPS). In the right hemisphere, only lateral HG and anterior STS showed neural priming effects. On the other hand, in the blocked version the only significant priming was found in the left STS. Part of this apparent difference is an artefact of cluster thresholding, as the blocked version led to significant priming effects in pars opercularis in the left hemisphere and the anterior STS in the right hemisphere. Nonetheless, several regions showed priming effects in the event-related, but not blocked, versions and these included lateral HG, pars triangularis, pars orbitalis, and the posterior IPS in the left hemisphere. Presumably these are related to the well-known strategic processing differences associated with "blocking" priming trials (Neely, 1991; Mummery et al., 1999). Significant negative interaction (contrast 12) was found in right Broca's area and left posterior STS/MTG. No significant activation is found when testing for positive interactions between the sentence effect and the speaker effect.

2.5 Improved inference using Mixture Modelling

Commonly in FMRI, null hypothesis testing is used on a contrast image to label voxels, or clusters of voxels, as being "active" if they reject the null hypothesis at a given false positive rate (FPR). This depends on knowing the distribution for the relevant statistic (eġ, regression parameter, t-statistic or pseudo-t) under the null hypothesis.

An alternative approach for inference is to use mixture modelling on the statistic of interest (Everitt and Bullmore, 1999). This involves fitting a mixture of distributions to the histogram of the statistic of interest; in our case we use a Gaussian for the central non-activation part of the data, and a gamma for the activation and possibly another for "deactivation" (Beckmann et al., 2003b; Woolrich et al., 2005). By using mixture modelling, instead of asking the question "Is the activation zero or not?", we ask the question "Is the activation bigger than the overall background noise?". Adaptability in modelling the non-activating ("null") part of the distribution can also help to protect against violations of the modelling assumptions, such as poorly modelled noise structure. For example, the presence of large-scale signal fluctuations which are not removed by filtering can result in a significant shift of the assumed background Gaussian distribution associated with non-activated voxels. Furthermore, mixture

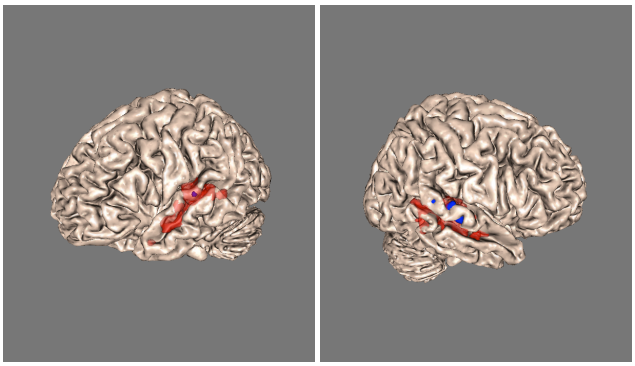


Figure 8: Estimated speaker effect when changing sentence content (contrast 6: DStSSp $<$ DStDSp) from the event-related data. No significant difference was found when null-hypothesis testing with Gaussian field theory was used for inference. When using alternative hypothesis testing via mixture modelling ($p > 0.5$), however , some difference was found, shown here in blue. Even more significant voxels were reported when, in addition, the HRF modelling was enhanced through the use of constrained optimal linear basis sets (shown in red).

modelling provides more inference flexibility compared with null hypothesis testing. One can either control FPR (or TPR–true positive rate) by using the "non-activating" or "activating" distributions respectively. Controlling the TPR may be of real importance when using FMRI for pre-surgical planning.

For the FIAC data, figure 8 (blue) shows the result of thresholding the GLM spatial Z-statistic image of contrast 6 (DStSSp $<$ DStDSp) with a Gaussian-Gamma mixture model. Unlike GRF thresholding, which failed to show any activation, the mixture model approach identifies small areas in lateral HG bilaterally, consistent with these regions showing a preference for human voice processing (Belin et al., 2000), although a different voice-repetition study identified a lateralised priming effect more anteriorally in the right STS (Belin and Zatorre, 2003).

2.6 Improved signal modelling using basis sets

The fact that contrast 6 (DStSSp<DStDSp), testing for the primary speaker effect of interest, did not show any significant activation clusters when tested against the null hypothesis of zero effect size potentially is influenced by a vast number of modelling inadequacies. In particular, the standard GLM methodology described above used a canonical model for the assumed haemodynamic response and does not allow for any variation in the shape of the response at different voxel locations (though small variations in the delay are taken into account by adding the temporal derivative).

In order to investigate the impact of the fixed HRF assumption we used FLOBS (FMRIB's Linear Optimal Basis Set), a tool based around the idea of generating optimal basis sets for use in HRF convolution in FMRI linear modelling (Woolrich et al., 2004b). FLOBS allows the specification of sensible ranges for various HRF-controlling parameters (delays and heights for the different parts of the HRF convolution kernel), generates lots of example HRFs where each timing/height parameter is randomly sampled from the range specified, and then uses PCA to generate a basis set that optimally "spans the space" of the generated samples.

The resulting basis functions are then used in the GLM to obtain constrained HRF fits. This is achieved by re-projecting the basis set onto the original set of samples in order to "learn" priors on the expected means and covariances of the individual basis functions. Using these priors when fitting the OBS model to the data means that the basis set is prevented from creating implausible HRF shapes. Constraining the shape of the estimated HRFs means that the noise is less "randomly fit" by the OBS model, giving better separation between the null part of the final statistics map and the activation part, iė; better activation modelling power. Final basis functions approximately correspond to the canonical HRF (first), its temporal derivative (second) and its dispersion derivative (third).

Accurate HRF modelling is much more important in event-related analyses than with block designs. In order to test whether constrained basis sets could improve the FIAC analysis, we re-ran the event-related analyses using 3 optimal basis functions. In order to estimate group mean effect using a summary of *power* (rather than shape) for each condition for each subject, we generated an F-statistic (across the 3 basis functions) for each contrast of interest at first-level, converted this into a pseudo-Z, averaged across sessions, and then carried out OLS mixed effects analysis at the group level.

Location	Cluster size (mm^3)	stat. significance	max. stats	Coordinates[†]		
				x	y	z
Effects of speaker repetition restricted to same sentence condition (fig. 5, conntrast C5)						
right Broca's area	7120	$p < 0.0001$	$Z = 3.23$	46	20	24
left STS/MTG	6336	$p < 0.0003$	$Z = 3.44$	-54	-58	6
Main effects of sentence repetition (fig. 6, contrast C9)						
left STS/MTG (middle portion)	3592	$p < 0.007$	$Z = 3.69$	-62	-14	-12
Interaction effects between sentence content and voice (negative interaction C12)						
right Broca's area	3136	$p < 0.02$	$Z = 3.33$	44	8	26
left posterior STS/MTG	2920	$p < 0.021$	$Z = 3.62$	-60	-66	6
Maximum effect of repetition surpression (fig. 7, contrast C13)						
left STS/MTG	12904	$p < 10^{-6}$	$Z = 4.35$	-58	-16	-8
Broca's area	4776	$p < 0.003$	$Z = 4.13$	-58	16	18
pos. intrapariatal sulcus	4528	$p < 0.004$	$Z = 3.19$	-32	-64	42
right ant. STS/MTG	3848	$p < 0.011$	$Z = 3.54$	62	0	-12
Main effects of speaker repetition (fig. 8, FLOBS & mixture modelling on contrasts C6)						
bilat. STG extending onto	7544	$p > 0.95^{\ddagger}$	$F = 4.23$	-64	-24	2
lat. Heschl's Gyrus	6584	$p > 0.95^{\ddagger}$	$F = 3.65$	68	-20	4

Table 2: FMRI activation cluster inferred from the model-based GLM analysis.
[†]coordinates are reported in *mm* from the anterior commissure (AC) in MNI152 standard space.
[‡]alternative hypothesis test based on non-spatial Gaussian/Gamma mixture modelling of the F-statistics image.

The constrained basis-set fitting approach results in statistic images where the null distribution at the highest level cannot be assumed. Therefore we carried out inference using mixture modelling, thresholding at the alternative hypothesis test threshold of $p > 0.5$. The inference was limited to F-statistic values which also show a positive fit to the first basis function (corresponding to the canonical HRF). This is in order to re-introduce directionality information which is not represented in the F-statistics. Final post-thresholded activation patterns thus represent potential changes in size and shape. The results can be see in figure 8 (red). Compared to mixture-modelling results obtained from the original GLM Z-statistic images, the fit to a flexible HRF basis set model results in large areas of activation seen in the posterior, middle and anterior regions of the STS, bilaterally.

While generally it is the case that 'larger' activations do not imply improved sensitivity, it is important to note that the improvements presented above were obtained using techniques which carefully control the probability of false-positive detection. Consequently, these results are validly comparable and not attributable to an increase in false-positive rate. Additional activation is only found in plausible auditory association areas, unlikely to be due to a regional unspecific increase in false detections.

3 Model-free data analysis

The GLM analysis approach described above explicitly tests for differences in the BOLD amplitude modulation under the different conditions. Differences in shape are typically not tested for, though the inclusion of the temporal derivatives (see figure 2) or more flexible time series modelling (eġuse of optimal basis sets) allow to test for significant differences in the haemodynamic delay and/or differences in the shape of the response between conditions. Though these types of analysis add to the flexibility of the GLM approach, model-based analysis is always fundamentally limited by the restrictions of the model chosen.

Exploratory data analysis approaches such as principal component analysis or independent component analysis decompose 2D data matrices (where 3D space is unwrapped into a single 1D vector of voxels, hence the original

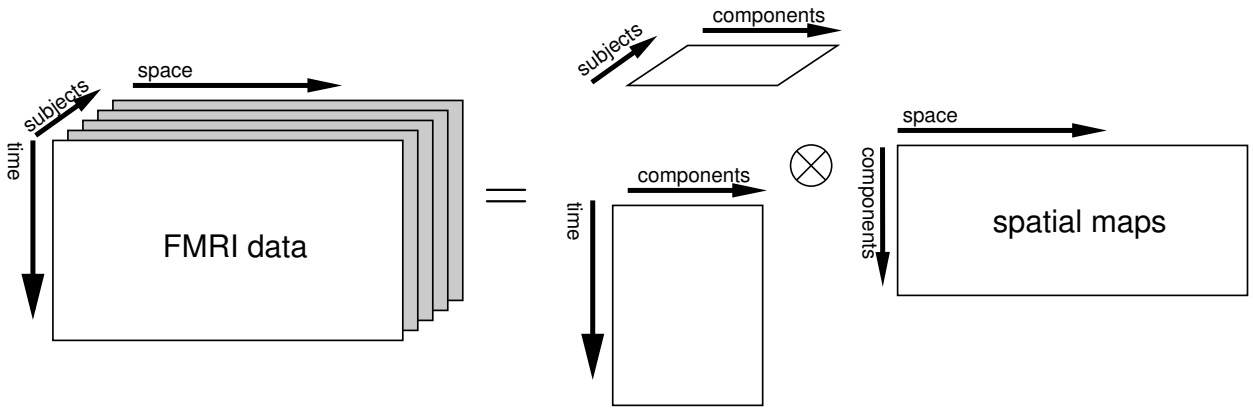


Figure 9: Schematic illustration of the model-free multi-subject Tensor-ICA approach: the data is represented as a time×space×subject block of data which is then decomposed as the outer product of matrices describing the different signal components in the spatial, temporal and subject domains.

4D FMRI data set is represented as a time×voxels 2D matrix) into a set of pairs of time courses and associated spatial maps. Each timecourse plus spatial map describes a different component in the data. An advantage of such approaches is the ability to detect unknown, yet structured spatiotemporal processes in neuroimaging data (McKeown et al., 1998; Beckmann and Smith, 2004a) and the additional modelling flexibility potentially results in an increased sensitivity in the detection of signals of interest.

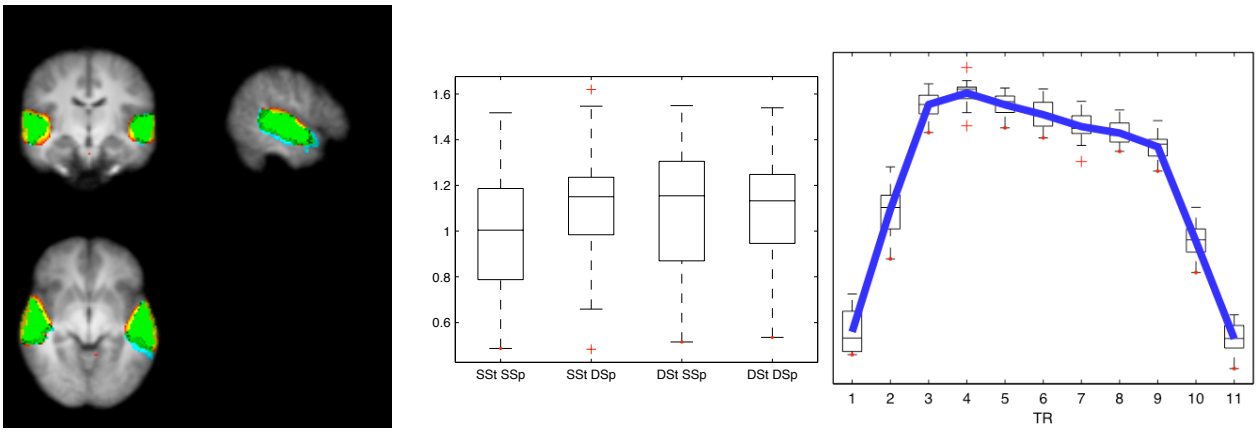
3.1 Tensor-ICA for the analysis of group FMRI data

The application of standard ICA to FMRI data was originally limited to the analysis of a single session's dataset at a time. In order to generalise ICA to the investigation of multiple sessions, subjects or subject groups, we have extended the standard 2D ICA methodology to higher dimensions (Tensor-ICA,see Beckmann and Smith, 2004b for a detailed desciption of the model and a comparison with other approaches to exploratory multi-subject analysis). In the Tensor-ICA approach, the group FMRI data is simultaneously decomposed into a set of components which now characterise the spatial, temporal *and* subject domains. Figure 9 shows a schematic illustration of this multi-dimensional decomposition technique. Note that spatial maps obtained from such a data-driven decomposition are explicitly optimised to violate a null-Hypothesis of random regression coefficients and therefore cannot be tested for significant activation based on simple null-hypothesis tests. Instead, the approach taken here assigns voxel-wise probabilities using an alternative hypothesis testing approach based on Gaussian/Gamma mixture modelling (see Beckmann and Smith, 2004a, for details).

On activation data the tensor-ICA approach has been demonstrated to extract plausible activation maps, time courses and estimates of the signal variation across the population (Beckmann and Smith, 2004b). Furthermore, it provides a richer description of the data than model-based techniques, showing additional processes of interest such as secondary activation patterns or artefacts.

3.2 Mean effects tensor-ICA

For the FIAC data, this approach is primarily used to help investigate the temporal signal characteristics in the auditory system for the block-design data. In particular we are interested in further investigating the question of whether there is significant within-block repetition suppression, iė; if there is any evidence for a BOLD amplitude decrease within blocks under the different conditions. For the tensor-ICA analysis, the pre-processed FMRI data for each subject was split into 4 different data sets, one for each stimulation condition, each containing 44 volumes. For each condition these were then combined across subjects before being fed into tensor-ICA, resulting in a set of time courses, spatial maps and subject/session modes. Each decomposition showed exactly one strongly stimulus-related effect where the spatial map shows significant voxels in primary and secondary auditory cortices and the associated time courses show clear block structure of the 4 repeats. Figure 10 summarises the results of the tensor-decompositions. The spatial map shows the T-ICA estimated map (red-yellow) together with the original FEAT GLM map (blue) for the DStDSp condition (C4; overlapping areas are rendered in green). The box-plots show



(i) across space (DStDSp) (ii) across subjects (iii) across time

Figure 10: Results from the Tensor-ICA decompositions of the block-design data. For the spatial map DStDSp, red/yellow=tICA, blue=GLM, green=both. The tICA spatial maps were thresholded by transforming the spatial raw IC estimates into Z-statistic maps (dividing raw IC estimates by the voxel-wise residual standard error of the decomposition) and thresholding these values at a posterior probability level of 'activation' ($p > 0.5$) based on a Gaussian/Gamma mixture model fitted to the distribution of spatial Z values.

the estimated response sizes of the 30 sessions for each of the four different conditions. There appears to be weak evidence for an overall reduction of BOLD modulation under the SStSSp condition, though not statistically significant. The timecourse plot shows the "peri-block response" as estimated over the four conditions and all four repeats (box-plot at every TR within the block show the distribution of the 16 estimated values around the median). While the estimated BOLD amplitude decreases within the block there is no evidence of a change in this response curve in the different conditions, i.e. no evidence of a systematic difference in shape over and above differences in magnitude.

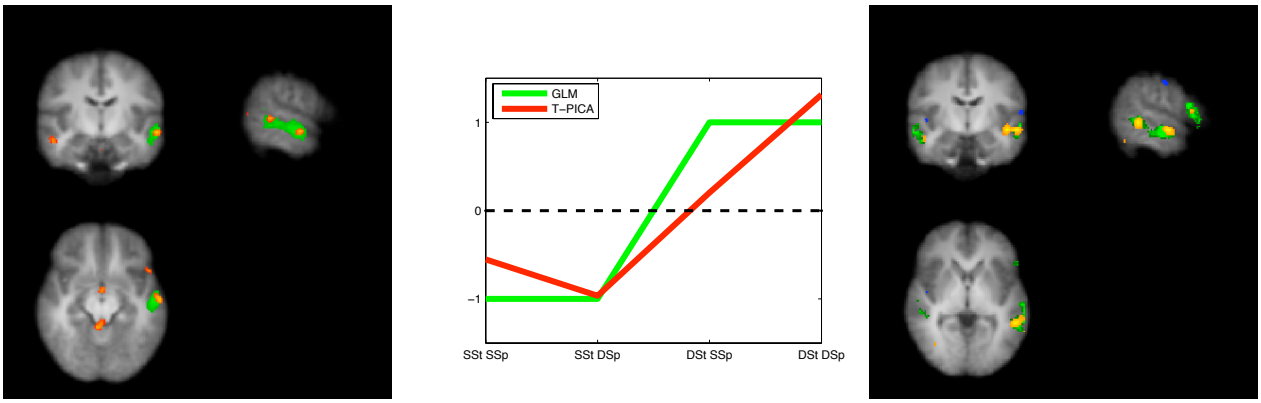
3.2.1 Between-condition analysis

The specific questions of voice and/or sentence repetition suppression, which within the GLM framework are being investigated by means of differential contrasts, can be approached in two different ways. Firstly, similar to the GLM, we can compare the estimated Tensor-ICA spatial maps directly by calculating contrast maps, i.e. by directly generating difference-maps from the 4 primary activation maps, obtained as the primary activation component in each of the 4 Tensor-ICA decompositions. As an example, figure 11(i) shows the contrast of the primary Tensor-ICA map calculated from the SStSSp condition and the primary map calculated from the data obtained during the DStDSp condition. Also shown is the GLM Z-statistic image for contrast 13 (green).

As an alternative to calculating the contrasts explicitly as difference images between primary tensor-ICA maps, we can instead use an exploratory data analysis at the second level and run standard ICA on the 4 primary tensor-ICA maps (concatenated into a $4 \times N$ matrix) in order to identify maximally non-Gaussian source processes and associated "exploratory contrasts". Figure 11(ii) shows one of the resulting images. The associated contrast vector (red line) is mostly similar to the contrast 10 vector (SStSSp+SStDSp $<$ DStSSp+DStDSp).

3.3 Additional tensor-ICA group maps

Out of the 19 maps which the tensor-ICA estimated from the multisubject data (the number of components was estimated from the Eigenspectrum of the data covariance matrix using the Laplace approximation to the model order for a probabilistic PCA model, Beckmann and Smith, 2004a), a few spatial maps show non-stimulus-related yet spatially structured BOLD signal fluctuations. These spatial patterns can also be identified in resting FMRI data (Beckmann et al., 2005) and various researchers have suggested that these signal variations are of neuronal origin and correspond to functional *resting-state networks* (RSNs). Examples of these patterns from the decompositions of the FIAC data are shown in figure 12(i). Regardless of their underlying cause, these patterns are a major source of non-modelled noise in FMRI. In particular, the visuospatial system (*default-mode network*, Gusnard and Raichle, 2001; blue) shows strong negative temporal correlation with experimental stimulation. Figure 12(ii) shows



(i) (ii)

Figure 11: "Contrasts" of primary tensor-ICA mean effects maps: (i) contrast 13 (SStSSp $<$ DStDSp) calculated explicitly from the 2 primary Tensor-ICA maps (red-yellow) together with contrast 13 from the GLM analysis (green); (ii) "exploratory contrast" calculated by decomposing the 4 primary auditory maps from the initial Tensor-ICA decompositions using standard ICA. One of the resulting columns of the mixing matrix shows strong correlation with the standard GLM contrast 10 and the spatial map (red-yellow) is shown on top of GLM mixed-effects results for contrast 10.

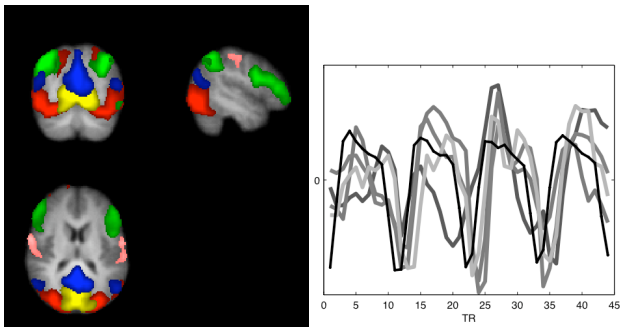


Figure 12: (i): Additional group maps, depicting structured BOLD fluctuations in functionally relevant cortical areas, such as medial visual cortex (yellow), lateral visual cortical areas (red), visual stream (green), motor (pink) and visuo-spatial attention (blue), often termed the *defaut-mode network* (Gusnard and Raichle, 2001). (ii): estimated temporal responses within the default-mode network for the 4 different conditions (grey, shown inverted) together with the temporal response for the primary activation pattern (black; average over all 4 conditions).

the estimated temporal characteristics corresponding to this estimated default mode network (inverted for display here); different grey-scales show the response as estimated within the 4 different conditions. The solid black line shows the mean estimated response in the primary auditory system. In all four data sets, the estimated temporal response of the 'default-mode' de-activation is strongly (negatively) correlated with the response of the primary auditory system (Pearson correlation $r > 0.5$).

4 Discussion

This paper has presented results obtained from applying various tools from FSL (FMRIB Software Library) to data from the repetition priming experiment used for the HBM'05 Functional Image Analysis Contest. We first discussed a standard GLM-based analysis using the FEAT model-based tool; this showed expected temporal lobe activation in the main conditions (Cardillo et al., 2004; Crinion et al., 2003), and also the reductions in activation upon repetition priming (Desimone, 1996; Dehaene et al., 2001; Vuilleumier et al., 2002). We then showed the improvement in sensitivity for detecting priming effects (for example, with respect to the speaker effect) when using recent more advanced analysis tools for modelling and inference (constrained optimal HRF basis functions and alternative hypothesis testing via mixture modelling). Note that these improvements were obtained using techniques which control the probability of false-positive detection, so that these results are validly comparable and not attributable to an increase in false-positive rate. Note also that these improvements are regionally specific to relevant areas. Finally, we applied model-free analysis to the FIAC data using tensor-ICA, which generated various results not obtainable by model-based analyses; we showed the spatial, temporal and cross-subject modes in the main experimental conditions, proposed two different approaches for investigating specific differential contrasts in such datasets, and showed several resting state networks found consistently across subjects.

One of the main goals of this competition was to comprehensively analyse a particular data set in order to highlight how different analysis techniques can identify findings beyond those seen in a basic GLM analysis. Thus it is of considerable interest that we do not replicate the asymmetries observed by Dehaene-Lambertz and colleagues (Dehaene-Lambertz et al., 2006) given the same data. That is, for repeated sentences by different speakers, we observed a reliable neural priming effect along the length of the STS, bilaterally. In addition, by modelling HRF variability and using using mixture modelling for inference testing, we demonstrated a similar voice-repetition neural priming effect which also ran the length of the STS, bilaterally. In other words, both sentence content and speaker's voice led to extensive reductions in auditory association areas in *both* hemispheres. Qualitatively, at least, there appears to be no difference in the response of the left and right STS to content vs. voices. There may, however, be a quantitative difference – at least for sentence content – as the repetition effects were larger on the left than right hemisphere. No such laterality difference was apparent for voice repetition. The effects of voice were smaller than those of sentence and therefore required more sensitive statistical techniques to identify. Nonetheless, these results call into question the notion that sentence content is primarily left lateralized while speaker identity is right lateralized or even limited to more anterior regions of the STS (Dehaene-Lambertz et al., 2006; Belin et al., 2000). Instead, the results suggest that both content and voice modulate a single underlying sentence processing system that includes anterior, middle, and posterior parts of the STS and MTG, bilaterally.

5 Acknowledgements

We are grateful for financial support from the UK MRC, EPSRC and BBSRC research councils and the Wellcome Trust. We also acknowledge vital collaborations with the many individuals listed at www.fmrib.ox.ac.uk/fsl/fsl/contributors.html and our other colleagues at FMRIB.

References

Beckmann, C., De Luca, M., Devlin, J., and Smith, S. (2005). Investigations into resting-state connectivity using independent component analysis. *Philosophical Transactions of the Royal Society*, 360(1457):1001–1013.

Beckmann, C., Jenkinson, M., and Smith, S. (2003a). General multi-level linear modelling for group analysis in FMRI. *NeuroImage*, 20:1052–1063.

Beckmann, C. and Smith, S. (2004a). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. on Medical Imaging*, 23(2):137–152.

Beckmann, C. and Smith, S. (2004b). Tensorial extensions to independent component analysis for multi-subject/session FMRI data. In *Tenth Int. Conf. on Functional Mapping of the Human Brain*.

Beckmann, C., Woolrich, M., and Smith, S. (2003b). Gaussian / Gamma mixture modelling of ICA/GLM spatial maps. In *Ninth Int. Conf. on Functional Mapping of the Human Brain*.

Belin, P. and Zatorre, R. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport*, 2105–2109.

Belin, P., Zatorre, R., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, 403(6767):309–312.

Cardillo, E., Aydelott, J., Matthews, P., and Devlin, J. (2004). Left inferior prefrontal cortex activity reflects inhibitory rather than facilitatory priming. *J. Cog. Neurosci.*, 16(9):1552–1561.

Crinion, J., Lambon-Ralph, M., Warburton, E., Howard, D., and Wise, R. (2003). Temporal lobe regions engaged during normal speech comprehension. *Brain*, 126(5):1192–1201.

Dehaene, S., Naccache, L., Cohen, L., Bihan, D., Mangin, J., Poline, J., and Riviere, D. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nat Neurosci.*, 4(7):678–680.

Dehaene-Lambertz, G., Dehaene, S., Anton, J., Champagne, A., Ciuciu, P., Dehaene, G., Denghien, I., Jobert, A., LeBihan, D., Sigman, M., Pallier, C., and Poline, J. (2006). Functional segregation of cortical language areas by sentence repetition. *HBM*.

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proc. Natl. Acad. Sci.*, 93(24):13494–13494.

Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., Wilson, J., Moss, H., Matthews, P. M., and Tyler, L. (2000). Susceptibility-Induced Loss of Signal: Comparing PET and fMRI on a Semantic task. *NeuroImage*, 11(6):589–600.

Everitt, B. and Bullmore, E. (1999). Mixture model mapping of brain activation in functional magnetic resonance images. *Human Brain Mapping*, 7:1–14.

Gusnard, D. and Raichle, M. (2001). Searching for a baseline: Functional imaging and the resting human brain. *Nature Review, Neuroscience*, 2:685–92.

Holmes, A. and Friston, K. (1998). Generalisability, random effects & population inference. In *Fourth Int. Conf. on Functional Mapping of the Human Brain, NeuroImage*, volume 7, page S754.

Jenkinson, M. (2003). A fast, automated, n-dimensional phase unwrapping algorithm. *Magnetic Resonance in Medicine*, 49(1):193–197.

Jenkinson, M. (2004). Improving the registration of B0-disorted EPI images using calculated cost function weights. In *Tenth Int. Conf. on Functional Mapping of the Human Brain*.

Jenkinson, M., Bannister, P., Brady, J., and Smith, S. (2002). Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17(2):825–841.

Jenkinson, M. and Woolrich, M. (2000). Asymptotic T to Z and F to Z statistic transformations. Internal Technical Report TR00MJ1, Oxford Centre for Functional Magnetic Resonance Imaging of the Brain, Department of Clinical Neurology, Oxford University, Oxford, UK.

Jezzard, P. and Balaban, R. (1995). Correction for geometric distortion in echo planar images from B0 field variations. *Magnetic Resonance in Medicine*, 34:65–73.

McKeown, M. J., Makeig, S., Brown, G. G., Jung, T. P., Kindermann, S. S., Bell, A. J., and Sejnowski, T. J. (1998). Analysis of fMRI data by blind separation into independent spatial components. *Human Brain Mapping*, 6(3):160–88.

Mummery, C., Shallice, T., and Price, C. (1999). Dual-process model in semantic priming: A functional imaging perspective. *Neuroimage*, 9(5):516–525.

Neely, J. (1991). Semantic priming in visual word recognition: A selective review of the current theories and findings. In Besner, D. and Humphries, G., editors, *Basic processes in reading: Visual work recognition*. Lawrence Erlbaum.

Smith, S. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, 17(3):143–155.

Smith, S., Beckmann, C., Ramnani, N., Woolrich, M., Bannister, P., Jenkinson, M., Matthews, P., and McGonigle, D. (2005). Variability in FMRI: A re-examination of intersession differences. *Human Brain Mapping*, 24:248–257.

Smith, S., Jenkinson, M., Woolrich, M., Beckmann, C., Behrens, T., Johansen-Berg, H., Bannister, P., De Luca, M., Drobnjak, I., Flitney, D., Niazy, R., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J., and Matthews, P. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23(S1):208–219.

Vuilleumier, P., Henson, R., Driver, J., and Dolan, R. (2002). Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nat Neurosci*, 5(5):491–499.

Woolrich, M., Behrens, T., Beckmann, C., Jenkinson, M., and Smith, S. (2004a). Multi-level linear modelling for FMRI group analysis using Bayesian inference. *NeuroImage*, 21(4):1732–1747.

Woolrich, M., Behrens, T., Beckmann, C., and Smith, S. (2005). Mixture models with adaptive spatial regularisation for segmentation with an application to FMRI data. *IEEE Trans. on Medical Imaging*, 24(1):1–11.

Woolrich, M., Behrens, T., and Smith, S. (2004b). Constrained linear basis sets for HRF modelling using Variational Bayes. *NeuroImage*, 21(4):1748–1761.

Woolrich, M., Ripley, B., Brady, J., and Smith, S. (2001). Temporal autocorrelation in univariate linear modelling of FMRI data. *NeuroImage*, 14(6):1370–1386.

Worsley, K., Evans, A., Marrett, S., and Neelin, P. (1992). A three-dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism*, 12:900–918.