

TEACHING STATISTICAL ANALYSIS OF fMRI DATA

Ian Lai¹, Randy Gollub^{2,3,4}, Richard Hoge³, Douglas Greve³, Mark Vangel³,
Russ Poldrack⁵, Julie E. Greenberg^{4,6}

¹ Department of Electrical Engineering and Computer Science, MIT

² Department of Psychiatry, Massachusetts General Hospital

³ MGH/MIT/HMS Martinos Center for Biomedical Imaging

⁴ Harvard-MIT Division of Health Sciences and Technology

⁵ Department of Psychology, UCLA

⁶ Research Laboratory of Electronics, MIT

Abstract

Functional magnetic resonance imaging (fMRI) represents a new and important topic in biomedical engineering. Statistical analysis of fMRI data is typically performed using free or commercial software packages that do not facilitate learning about the underlying assumptions and analysis methods; these shortcomings can lead to misinterpretation of the fMRI data and spurious results. We are developing an instructional module for learning the fundamentals of statistical analysis of fMRI data. The goal is to provide a tool for learning about the steps and assumptions underlying standard fMRI data analysis so that students and researchers can develop insights required to use existing analysis methods in an informed fashion and adapt them to their own purposes. The module includes a simulation of fMRI data analysis that provides students with opportunities for hands-on exploration of the key concepts using phantom data as well as sample human fMRI data. The simulation allows students to control relevant parameters and observe intermediate results for each step in the analysis stream (spatial smoothing, motion correction, statistical model parameter selection). It is accompanied by a tutorial that directs students as they use the simulation. The tutorial guides students through the individual processing steps, considering multiple cycles of fMRI data analysis and prompting them to make direct comparisons, with emphasis on how processing choices affect the ultimate interpretation of the fMRI data.

I. Introduction

While magnetic resonance imaging (MRI) was introduced for clinical use in the 1970s, functional magnetic resonance imaging (fMRI) was discovered in the early 1990s¹⁻⁴. This relatively new research tool has found widespread use in a variety of applications at the intersection of biomedical engineering and neuroscience, for example, mapping the boundaries between functional regions of the brain, identifying tumor margins prior to surgery, and investigating the pathology underlying diseases such as schizophrenia. fMRI detects activity in the brain by taking advantage of the change in magnetic properties of the blood surrounding

neuronal activation, which produces a blood oxygen level dependent (BOLD) signal that can be detected by an MRI scanner. In a typical fMRI experiment, the magnitude of the BOLD signal is quite small, on the order of 0.1% to 5.0% change in signal level, and is easily obscured by noise. Due to this inherent limitation, multiple measurements are required, and researchers must rely on statistical analyses to detect significant response^{5,6}.

Although the field of fMRI research is still relatively young, it has experienced explosive growth; in the years 1999-2001 more than 900 abstracts were submitted to the International Conferences on the Functional Mapping of the Human Brain⁷. While some investigators in the field have expertise concerning the details of fMRI data analysis, others simply apply free or commercial software packages to their data. These packages often include a multitude of parameters that can be optimized by an fMRI expert, but are rarely adjusted by others⁸. Since the packages have preset defaults that may not be appropriate for all situations, investigators lacking a proper understanding of fMRI data analysis may draw false conclusions from their data.

Students and researchers wanting to learn more about fMRI data analysis have limited resources available to them. The most widely available resources are two recently published textbooks^{9,10} and documentation that accompanies particular software packages¹¹⁻¹⁵. There are also a number of courses and workshops on fMRI, for example¹⁶⁻²⁹, but these represent a limited resource that is not available to all researchers in the field.

This paper describes an on-line, interactive module for learning the fundamentals of fMRI data analysis. The goal is to provide a tool for learning about the steps and assumptions underlying standard fMRI data analysis so that students and researchers can develop insights required to use existing software packages in an informed fashion and adapt them to their own purposes. The focus is on the fundamental processing steps and parameters commonly used in fMRI data analysis. Topics such as fMRI experimental design, the physics of fMRI data acquisition, and advanced methods for data processing and analysis are beyond the scope of this module; when possible, the module provides links to external sites with additional materials on these topics.

The intended audience includes advanced undergraduate and graduate students, as well as investigators who wish to use fMRI in their research but are not familiar with the methods and techniques of fMRI data analysis. This module is intended to be a standalone source for learning about fMRI data analysis, although it may also be a useful adjunct to existing courses and workshops.

II. Instructional module

The module includes two main components, a simulation of fMRI data analysis and a tutorial that guides students through a series of key concepts as they use the simulation.

A. Simulation of fMRI data analysis

The simulation provides students with opportunities for hands-on exploration of the individual processing steps in fMRI data analysis. They can control the relevant parameters and observe

intermediate results for each step in the processing stream. This allows the user to see how the parameters at each processing step affect the overall outcome of the analysis.

The user interface for the simulation consists of a panel of menus and buttons, shown in Figure 1. The panel divides the menus and buttons into seven frames. One frame controls selection of the input data, and four frames (Motion Correction, Spatial Filtering, Statistical Model, and Inference) correspond to steps in the normal flow of the processing. The remaining frames provide functions to compute statistical properties and to control which results are displayed; these functions facilitate examination of the data but do not affect its analysis.

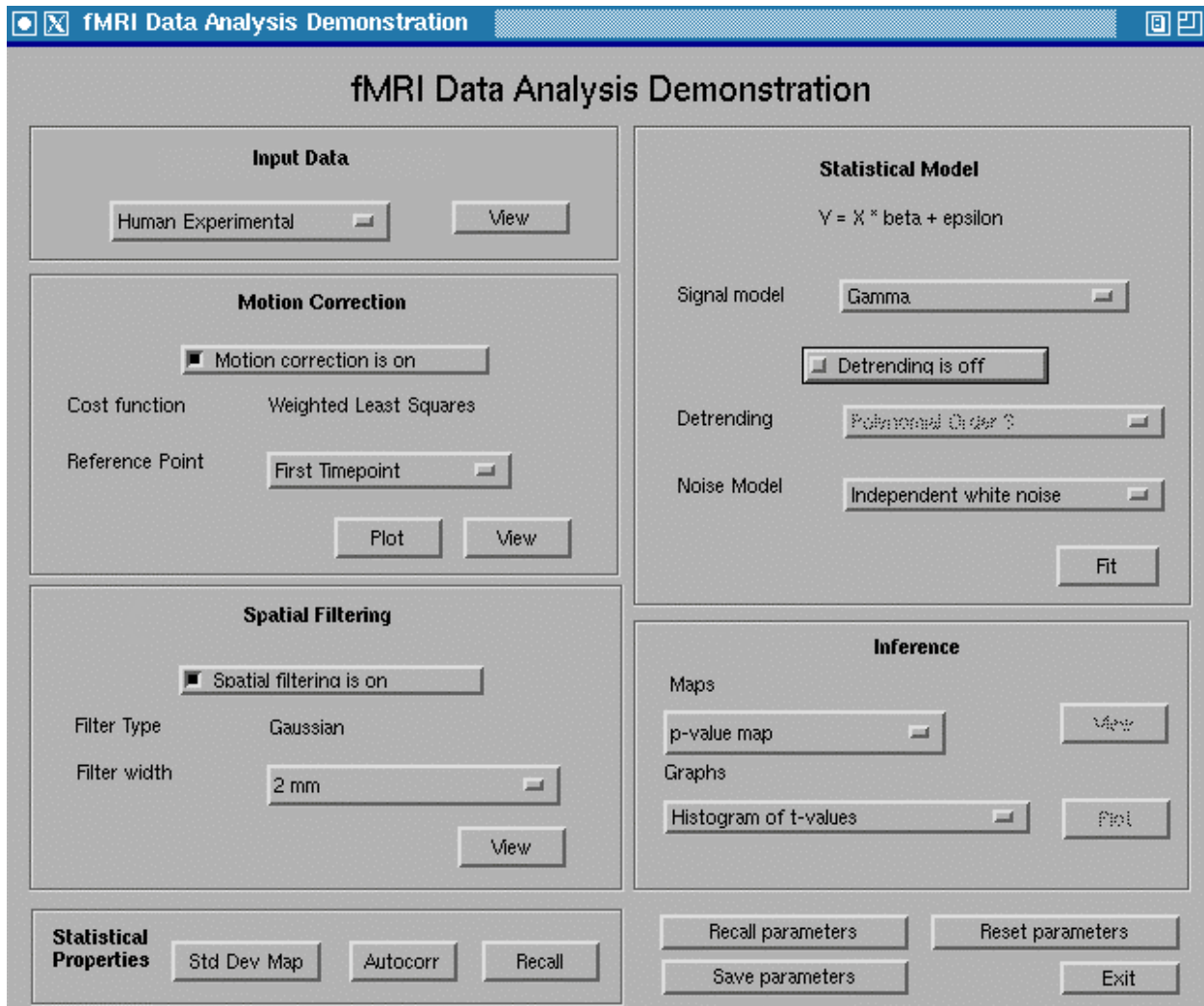


Figure 1: The simulation's main control panel. The user can select a data set, enable or disable processing steps, and vary parameters for processing and statistical analysis. The user can also select from a variety of displays showing: the data's statistical properties, intermediate results of the processing, and results of the statistical analysis.

Data selection: Currently, the user can select from phantom (a jug of paramagnetic salt solution in distilled water) data, resting human fMRI data, and experimental human fMRI data. For the resting data the human subject was not exposed to any stimuli, while for the experimental data

the same human subject viewed a visual stimulus; the stimulus was a flashing checkerboard pattern displayed intermittently to produce a time-varying response. Future versions may permit the user to upload their own fMRI data. These are four-dimensional data sets consisting of time series of three-dimensional brain volume images. The user may view the four-dimensional data sets as described below.

Motion correction: The user can enable or disable motion correction. When motion correction is enabled, it uses a weighted least squares cost function. A menu of cost functions may be added in the future. The user can select the reference for motion correction to be the brain volume at the first time point or the middle time point of the scan. The user may plot the estimated rotation and translation over time necessary to align the brain volumes. The user may also view the four-dimensional data after application of motion correction.

Spatial filtering: The user can enable or disable spatial filtering. When spatial filtering is enabled, it uses a Gaussian filter. A menu of filter shapes may be added in the future. The user can select the filter width for spatial filtering to be 2, 4, 6, or 8 mm. The user may view the four-dimensional data after application of spatial filtering.

Statistical model: The data analysis is based on the general linear model (GLM). The user can choose from two signal models. One signal model uses prior knowledge about the delay and dispersion known to occur in the hemodynamic coupling between the neuronal response and the measured fMRI signal change. This model assumes that the hemodynamic response to be detected has the shape of the weighted difference between two gamma distributions; one fits the initial rise and the other fits the subsequent undershoot. The second signal model assumes that the response is finite but otherwise unknown and attempts to determine the signal shape empirically. The user can choose the noise model to be independent white noise or first-order autoregressive noise.

Within the statistical model, the user can also enable or disable a detrending function. When detrending is enabled, it fits a polynomial function to the time course for each voxel and adjusts for that component when applying the signal and noise models. The user can select a polynomial of order 0, 1, 2, 3 4, or 5.

Inference: The inference block provides a variety of ways to examine the results of applying the statistical model. The user can select one of two activations maps corresponding to the p -values and the t -statistics. The user can view the activation maps as three-dimensional brain volumes. The user can also choose to display plots showing a number of intermediate variables and results: a histogram of values of the t -statistics; the design matrix used in the GLM as determined by the parameter choices in the *statistical model* processing block; the signal model determined by the GLM at each voxel; the false-positive rate; and a comparison of the raw data and the fitted model data for a particular voxel.

Statistical Properties: The statistical properties block allows the user to view maps of the temporal autocorrelations and standard deviations.

An additional set of user controls facilitates comparison between different processing parameters. The *Save Parameters* button causes the current set of parameters to be saved for future reference; the *Recall Parameters* button produces a dialog box listing all previously saved parameter sets. When a set of parameters is recalled, it is automatically applied and the user views the data corresponding to the chosen parameters. A simplified version of the *Recall* function is used with the display-only Autocorrelation and Standard Deviation Map buttons in the *Statistical Properties* block. In this case, the parameters and displays viewed are saved automatically and the *Recall* button produces a list of all the previously viewed plots. This special case allows the user to compare statistical properties of the data more easily.

The simulation utilizes a locally developed software package known as Dview for navigating three- and four-dimensional brain image data (Figures 2 and 3). Dview displays a run of fMRI data, showing standard transverse, sagittal, and coronal views of the brain corresponding to the point in space occupied by the cursor; the user may click on any of the three views to move the cursor and change the views. For four-dimensional data sets, another display frame shows the time-course data for the pixel corresponding to the cursor.

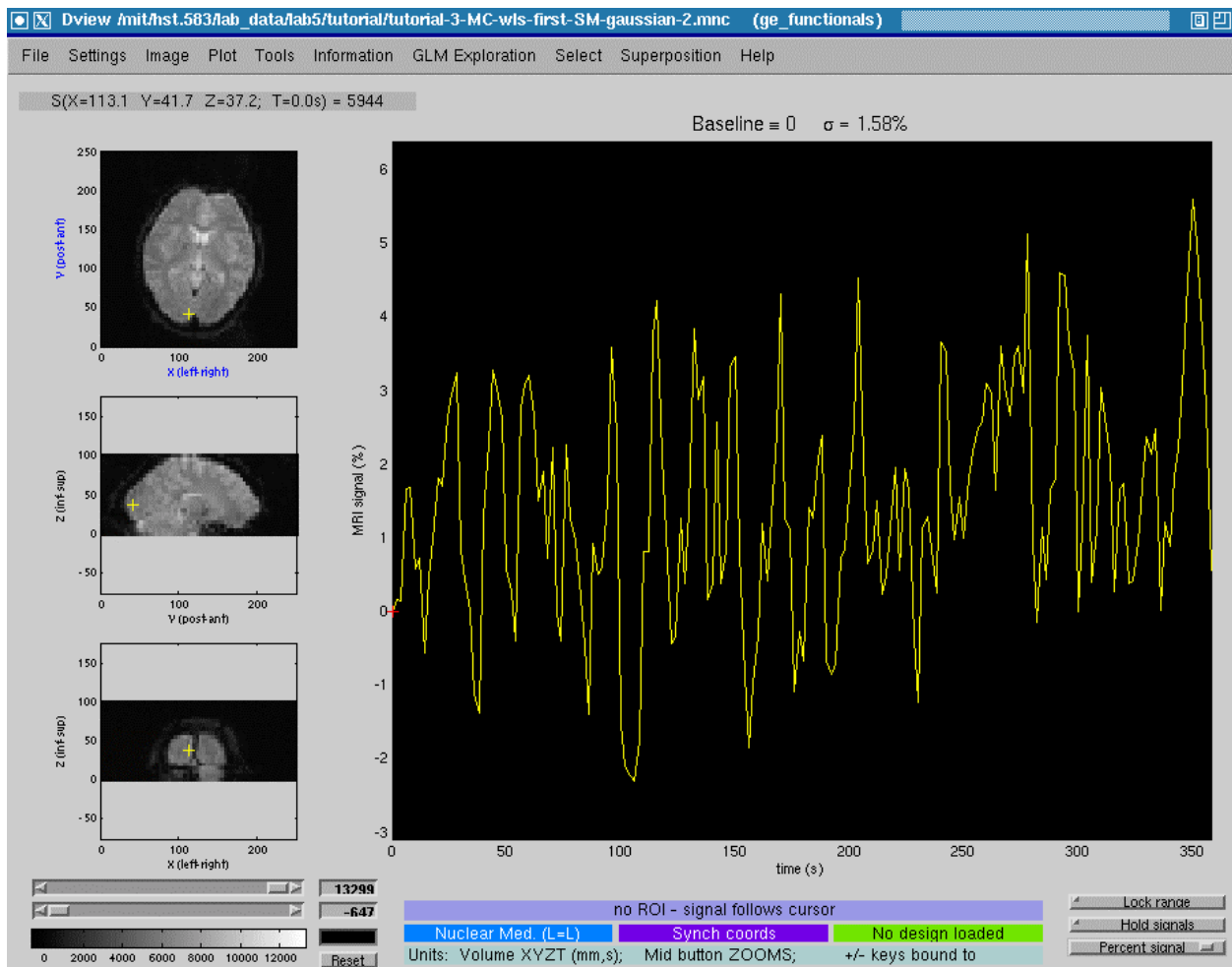


Figure 2: Dview display of the magnetic resonance signal as a function of time. The time-course plotted on the right corresponds to the single voxel indicated by the cursor in the three cross-sectional images on the left. The user can display different voxels by clicking to move the cursor.

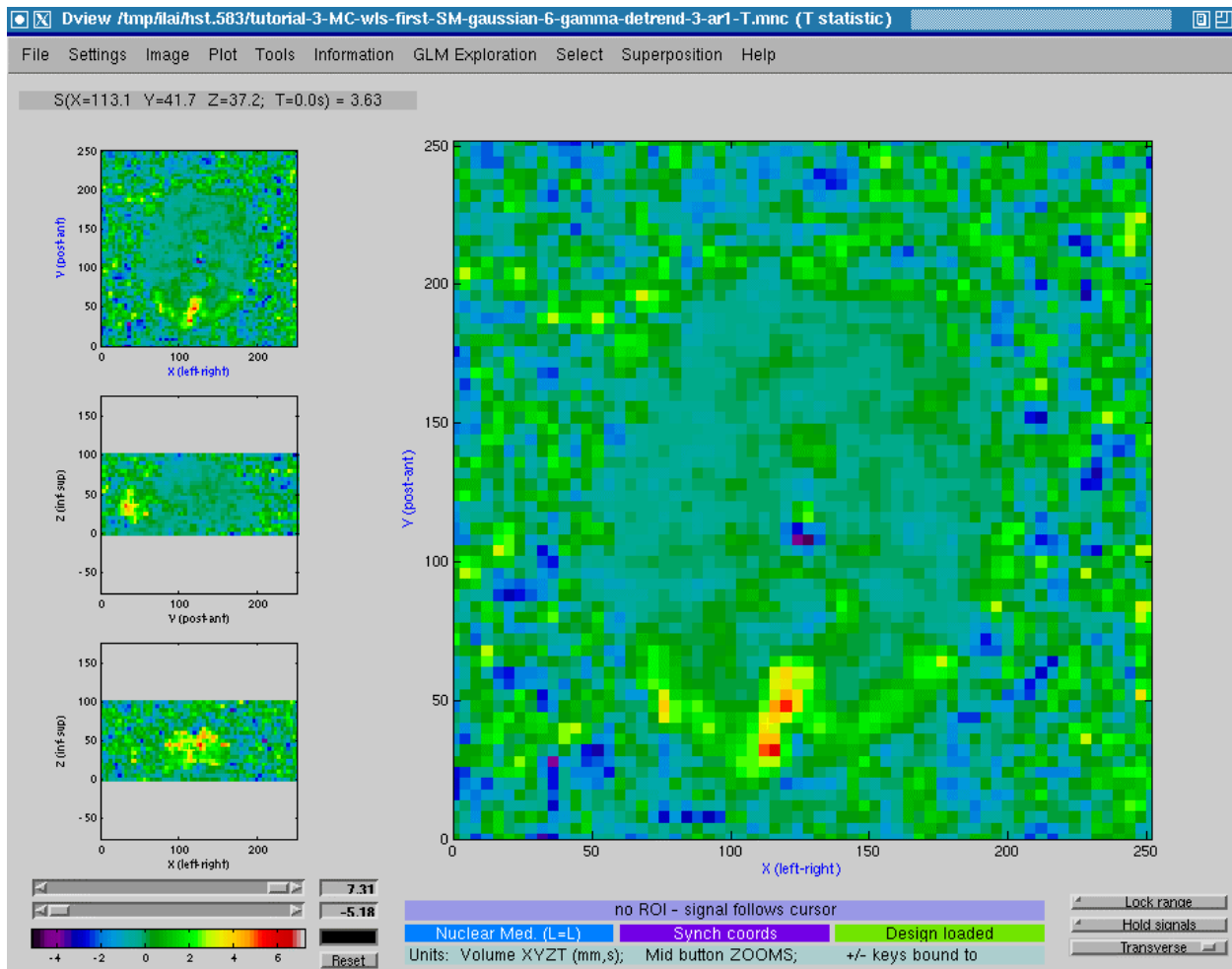


Figure 3: Dview display of an activation map of the brain for data from a visual stimulus experiment. The user can navigate the three-dimensional data set by clicking to move the cursor in the three cross-sectional images on the left. Any one of the three cross sections can be displayed as a larger image on the right. The user can also change the threshold by clicking the color bar below the cross sections.

The simulation and Dview are both implemented in MATLAB[®]³⁰. Due to the computational complexity of the motion correction and spatial filtering processing, some results are pre-computed for all data sets and parameter choices and then stored for future display.

B. Tutorial

The goal of the tutorial is to guide students as they use the simulation to develop an understanding of the key concepts in fMRI data analysis. The use of simulations in educational settings is most effective when students are working toward a clear goal, yet the assigned tasks are not too narrowly defined.³¹

The tutorial directs students to examine the individual processing steps for different fMRI data sets and to make direct comparisons, with emphasis on how processing choices affect the

ultimate interpretation of the data. The tutorial provides three types of material: step-by-step instructions for using the simulation, background information about the individual steps in the processing, and questions that probe the students' understanding of the material.

Four main learning objectives are stated explicitly in the tutorial. They are:

- Understanding temporal and spatial correlation in fMRI data;
- Understanding how to construct a statistical model for fMRI data;
- Identifying sources of noise and how they affect fMRI signals; and
- Understanding the effects of motion correction and spatial filtering on the outcome of statistical analysis of fMRI data.

As an example, Figure 4 shows an excerpt of the tutorial text related to the first learning objective.

III. Preliminary use of module

This instructional module was piloted during Fall term 2002 in a course offered by the Harvard-MIT Division of Health Sciences and Technology, *HST.583: Functional Magnetic Resonance Imaging: Data Acquisition and Analysis*.¹⁸ In the context of the course, which included lectures on experimental design and statistical analysis and assigned readings, students spent two one-and-a-half hour lab sessions using the simulation and working through the tutorial. Students in the class liked the lab exercises in general, and informal feedback solicited on this module in particular was generally positive. Some students expressed a desire for additional guidance in the tutorial.

The instructor's impression was that students who attended the four lectures specifically on fMRI data analysis and completed the assigned readings got the most benefit from the module. Individuals who came to lab without adequate preparation did not learn as much from the simulation and tutorial.

IV. Ongoing work

Our ongoing work on this instructional module includes two main components: expanding the scope of the tutorial and making the module more widely accessible.

The version of the tutorial piloted during Fall term 2002 was intended to accompany, and depended on, four lectures on fMRI data analysis. Student feedback indicated a desire for more guidance from the tutorial itself, and this coincides with our goal of developing an instructional module that stands independently. In order to accomplish this, we are generating additional content for the tutorial so that it can be used, in conjunction with the simulation, to learn about fMRI data analysis without need for attending lectures or additional reading.

To make the module widely accessible to students, educators and researchers, we are currently transforming the simulation into an application that can be run over the World Wide Web. The current implementation of the simulation is limited by its dependence on local installation of

1. Temporal Autocorrelation

1. In the **Input Data** box, click on the drop-down menu and select the **Human Experimental** data set.
2. Click **Autocorr** in the **Autocorr** box. Dview will now show the temporal autocorrelation for the voxel under the yellow crosshair cursor.
3. Move the cursor by clicking on another part of the brain in any of the transverse, sagittal, or coronal views. You should see the autocorrelation change depending on which voxel you have selected.
4. Now examine the autocorrelation in air, gray matter, white matter, and CSF. For gray matter, pick three locations among the frontal, parietal, and temporal cortices, and one in the basal ganglia.

Question 3. (a) Plot the autocorrelation graph for each of the seven tissue regions that you examined and label each. Describe what you see; is it what you expected? (b) Given what you know about the experimental paradigm, find an area in the brain that has paradigm-related signal changes. Plot this separately and label with the correct anatomical region. How does the autocorrelation here differ from other regions?

5. Select the **Phantom** data set in the **Input Data** box, and click **Autocorr** to show the autocorrelation. Move around different areas of the phantom and the surrounding air.

Question 4. Plot the autocorrelation graph for 2 locations of your choice in the phantom. What does the autocorrelation look like in the phantom? How does it vary within the phantom? Is it what you expected?

Detrending is a way of removing signal drift and other low-frequency noise, which may result from slow head movements, gradient heating, etc. Here we look at how the use of detrending impacts the temporal autocorrelation.

1. In the **Input Data** box, again select the **Human Experimental** data set. Turn off **Spatial Filtering** if it is on.
2. Turn on **Detrending** and set the **polynomial order** to 3.
3. Now examine the autocorrelation in the same tissues (CSF, 4 gray matter areas, white matter, air) that you used above by clicking on **Autocorr** and moving the crosshair to the appropriate regions.

Question 5. What effect does detrending have on temporal autocorrelation? Is it the same in all 7 regions? Explain why the effects occurred.

Aside: Polynomial Detrending. In the demo, we remove trends in the data which are polynomial in nature. The order of the polynomial trend removed can be varied to yield different results. For instance, removing 0th-order polynomials simply subtracts the mean from the signal, and removing 1st order polynomials does that and also removes any linear trend. As a thought exercise, consider the effects of removing higher-order polynomial trends. Why wouldn't it be a good idea to perform detrending with, say, 85th-order polynomials?

Figure 4: Excerpt from tutorial on fMRI data analysis

MATLAB®, as well as the size of the data sets. Work is currently underway to transition the simulation to a Web-accessible application using a conventional client/server model. The server stores the data sets and performs the majority of the computations, while the client performs user interface functions.

We intend to make the Web-accessible version of the module available to the general public via the Biomedical Informatics Research Network (BIRN).³² BIRN is a consortium of university and hospital research organizations concerned with neuroimaging data acquisition and analysis. The consortium's initial projects are all focused on structural and functional neuroimaging data. They include a Web site to support biomedical education on topics related to BIRN's research efforts, an ideal place for dissemination of this module.

We plan to use this module in conjunction with an fMRI statistics workshop that will be offered during the summer of 2003 at the MGH/MIT/HMS Martinos Center for Biomedical Imaging. This two-week long workshop will meet for two-hour sessions three times per week. Participants will attend a series of lectures as well as lab sessions that utilize this module. We will collect specific feedback on various aspects of the simulation and tutorial from the participants. This input will inform future revisions of the module.

V. Summary

This paper describes an instructional module on the topic of fMRI data analysis. The goal is to provide students and researchers with a tool to explore the processing steps that underlie the methods of data analysis in this relatively new and rapidly growing field of biomedical engineering. The module includes an interactive simulation of fMRI data analysis, which allows the user to explore the individual processing steps by controlling relevant parameters and observing intermediate results for each step in the processing stream. The simulation is accompanied by a tutorial that guides students as they use the simulation, providing step-by-step instructions for using the simulation, background information about the individual steps in the processing, and questions that probe the students' understanding of the material. Initial student reaction to the module was favorable. Ongoing work is aimed at expanding the scope of the tutorial, making the simulation Web-accessible and evaluating the module in a workshop to be delivered during the summer of 2003.

Bibliography

1. Ogawa, S. and Lee, T.M. "Magnetic resonance imaging of blood vessels at high fields: in vivo and in vitro measurements and image simulation," *Magn Reson Med* **16**(1) pp. 9-18 (1990).
2. Belliveau, J.W., Kennedy, D.N., Jr, McKinstry, R.C., Buchbinder, B.R., Weisskoff, R.M., Cohen, M.S., Vevea, J.M., Brady, T.J., and Rosen, B.R. "Functional mapping of the human visual cortex by magnetic resonance imaging," *Science* **254**(5032) pp. 716-9 (1991).
3. Ogawa S., Tank, D.W., Menon, R., Ellermann, J.M., Kim, S.G., Merkle, H., and Ugurbil, K. "Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging," *Proc Natl Acad Sci USA* **89**(13) pp. 5951-5 (1992).

4. Kwong, K.K., Belliveau, J.W., Chesler, D.A., Goldberg, I.E., Weisskoff, R.M., Poncelet, B.P., Kennedy, D.N., Hoppel, B.E., Cohen, M.S., Turner, R., Cheng, H.M., Brady, T.J., and Rosen, B.R. "Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation," *Proc Natl Acad Sci USA* **89**(12) pp. 5675-9 (1992).
5. Bandettini, P.A., Jesmanowicz, A., Wong, E.C., and Hyde, J.S. "Processing strategies for time-course data sets in functional MRI of the human brain," *Magn Reson Med* **30**(2) pp. 161-73 (1993).
6. Friston, K.J., Jezzard, P., and Turner, R. "Analysis of Functional MRI Time-Series," *Human Brain Mapping* **1** pp. 153-171 (1994).
7. Savoy, R.L. "History and future directions of human brain mapping and functional neuroimaging," *Acta Psychologica*, **107**, pp. 9-42 (2001).
8. Gold, S., Christian, B., Arndt, S., Zeien, G., Cizadlo, T., Johnson, D.L., Flaum, M., and Andreasen, N.C., "Functional MRI Statistical Software Packages: A Comparative Analysis," *Human Brain Mapping*, **6**(2), pp. 73-84 (1998).
9. Jezzard, P., Matthews, P.M., and Smith, S.M., Eds. *Functional MRI: An Introduction to Methods*. Oxford, UK: Oxford University Press (2001).
10. Buxton, R. Ed. *Introduction to Functional Magnetic Resonance Imaging: Principles and Techniques*. Cambridge, UK: Cambridge University Press (2001).
11. <lbc.nimh.nih.gov/fidap/fidap.html>, accessed January 13, 2003.
12. <afni.nimh.nih.gov/afni>, accessed January 13, 2003.
13. <www.mrc-cbu.cam.ac.uk/Imaging/spm_theory.htm>, accessed January 13, 2003.
14. <www.fmrib.ox.ac.uk/fsl/feat4>, accessed January 13, 2003.
15. <defiant.ssc.uwo.ca/Jody_web/fmri4dummies.htm>, accessed January 13, 2003.
16. <www.sph.umich.edu/~nichols/fMRIcourse>, accessed January 13, 2003.
17. <airto.loni.ucla.edu/BMCweb/CourseWork/M285/>, accessed January 13, 2003.
18. <web.mit.edu/hst.583/www/>, accessed January 13, 2003.
19. <fmri.neuro.brown.edu/BN293/BN293>, accessed January 13, 2003.
20. <www.fmri.org/course.htm>, accessed January 13, 2003.
21. <www.klab.caltech.edu/~geraint/mri/descrip.html>, accessed January 13, 2003.
22. <www.fmrib.ox.ac.uk/internal/medx/course/>, accessed January 13, 2003.
23. <www.neco.ucl.ac.be/mp/journees_IRMf.html>, accessed January 13, 2003.
24. <www.firc.mcw.edu/course>, accessed January 13, 2003.
25. <www.nmr.mgh.harvard.edu/fmriivfp>, accessed January 13, 2003.
26. <www.fmri.nl/General%20stuff/id45.htm>, accessed January 13, 2003.

27. <futu.hut.fi/fmri.html>, accessed January 13, 2003.
28. <www.hbm2001.ucl.org.uk/register/course10.html>, accessed January 13, 2003.
29. <www.ismrm.org/02/morningcat2('02).htm>, accessed January 13, 2003.
30. <www.mathworks.com>, accessed January 13, 2003.
31. Laurillard, D., "Learning through collaborative computer simulations," *British Journal of Educational Technology*, **23**(3) pp. 164-171 (1992).
32. <www.nbirn.net>, accessed January 13, 2003.

Acknowledgements

This work was supported by the NSF Engineering Research Centers award number EEC-9876363, by the MGH/MIT/HMS Martinos Center for Biomedical Imaging, and by the MIND Institute.

IAN LAI is a graduate student at the Massachusetts Institute of Technology. He expects to receive the S.B. and M.Eng. degrees in Electrical Engineering and Computer Science in June 2003.

RANDY L. GOLLUB, M.D., Ph.D. is an Assistant Professor of Psychiatry at Harvard Medical School, the Assistant Director of Psychiatric Neuroimaging at Massachusetts General Hospital and a Member of the Faculty of the Harvard-MIT Division of Health Sciences and Technology (HST) at HMS. She has been actively involved in neuroscience research for the past 25 years and is currently pursuing studies of acupuncture analgesia using state of the art functional neuroimaging methodologies. She is the course director for an HST graduate course entitled fMRI: Data Acquisition and Analysis.

RICHARD HOGE, Ph.D., is an Instructor at the Martinos Center for Biomedical Imaging and an Assistant Physicist at the Massachusetts General Hospital. His research activities focus on the use and development of functional MRI (fMRI) technology for the study of brain activation physiology. He has a particular interest in understanding the mechanisms of image contrast in blood oxygenation level-dependent (BOLD) fMRI.

DOUGLAS N. GREVE, PhD, is director of fMRI Post-Processing at the NMR Center at Massachusetts General Hospital. He has written much of the fMRI analysis software used at the NMR Center. His research interests include optimization of event-related fMRI, modeling of the event-related BOLD response, and spatio-temporal analysis of fMRI. He is also a lecturer for fMRI: Data Acquisition and Analysis, an HST graduate course.

MARK VANGEL, PhD, is Statistician at the NMR Center at Massachusetts General Hospital, and at the MIT General Clinical Research Center. He has approximately 20 years of experience in applying statistics to a wide range of problems in science and engineering. His research interests include linear models, applied Bayesian statistics, calibration, interlaboratory comparisons, and statistical inverse problems.

RUSSELL A. POLDRACK, Ph.D. is Assistant Professor in the Department of Psychology and Brain Research Institute at the University of California, Los Angeles. His research uses functional magnetic resonance imaging to investigate the neural basis of learning and memory. He has been involved in the teaching of fMRI methods at MGH, MIT, and UCLA.

JULIE E. GREENBERG, Ph.D. is a Lecturer in the Harvard-MIT Division of Health Sciences and Technology and a Research Scientist in the Research Laboratory of Electronics at the Massachusetts Institute of Technology. Her research interests include signal processing for hearing aids and the use of technology in bioengineering education. She is the course director for an HST course on Biomedical Signal and Image Processing. Dr. Greenberg is a member of the ASEE and the IEEE.