Receiver Operating Characteristics Curves in functional magnetic resonance imaging

Finn Årup Nielsen

Informatics and Mathematical Modelling
Technical University of Denmark

2003 September

Overview

ROC curves

Binomial mixture model

Superior temporal sulcus. Digitization, extraction, analysis

Clustering

Receiver operating characteristics (ROC)

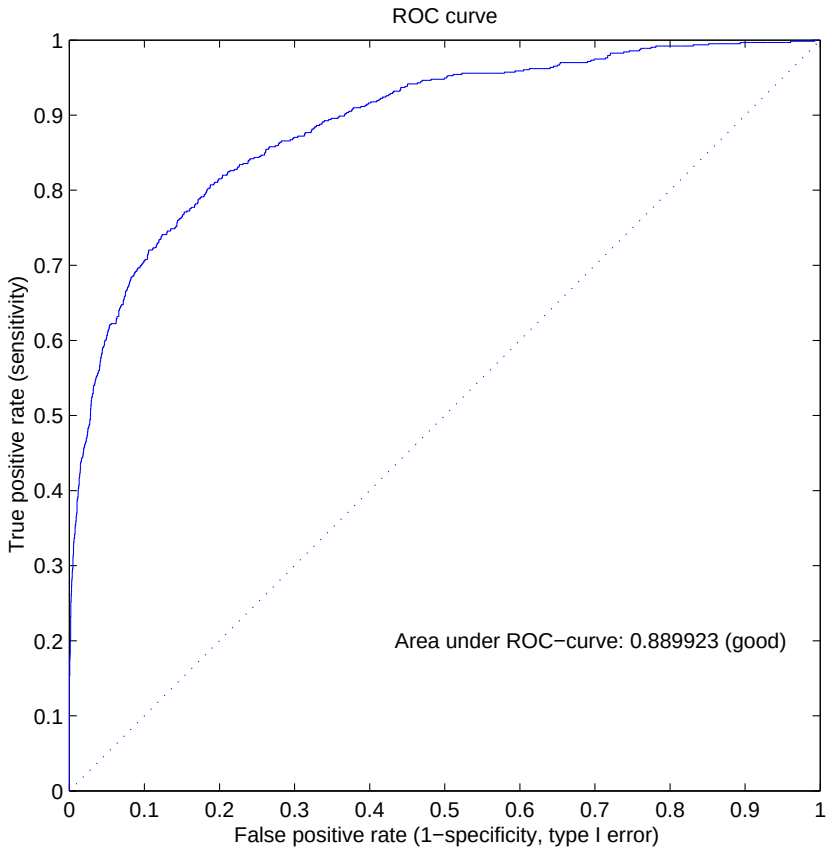


Figure 1: ROC curve with artificial data plotted with `brede_vol_plot_roc.m`.

ROC curve/analysis in fMRI (Constable et al., 1995).

Model comparison with null data and simulated response (Sorenson and Wang, 1996; Lange et al., 1999; Lange et al., 1998).

Requires the "ground truth"

ROC analysis with no "ground truth"

No ground truth but repeated experiments (Genovese et al., 1997; Genovese et al., 1996).

A mixture of two binomial distributions (Gelfand and Solomon, 1974)

$$p(\mathbf{n}|P_A, P_I, \lambda) \propto \sum_{m=0}^{M} n_m \ln\left[\lambda P_A^m (1-P_A)^{M-m} + (1-\lambda) P_I^m (1-P_I)^{M-m}\right],$$

where M is the number of replications, P_A and P_I are probabilities for a positive classification to be truly active and truely inactive, respectively.

Application for, e.g., evaluation of respiratory artifact correction techniques (Noll et al., 1996), comparison of Student t and Kolmogorov-Smirnov statistics (Genovese et al., 1997).

Small binomial mixture example

$M = 4$ replicated thresholded "volumes":

$$\underbrace{\left.\left.\begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}\right\}\right.}_{\text{Voxels } V = 12} \text{Replications } M = 4 \tag{1}$$

The sufficient statistics $\mathbf{n}$, with, e.g., the first element counting the number of voxels that are zero in all replications.

$$\mathbf{n} = [5, 0, 2, 2, 3]\,. \tag{2}$$

Estimation of the parameters P_A, P_I and λ

$$P_A = 0.78, \qquad P_I = 0, \qquad \lambda = 0.58 \tag{3}$$

ROC binomial mixture — Example

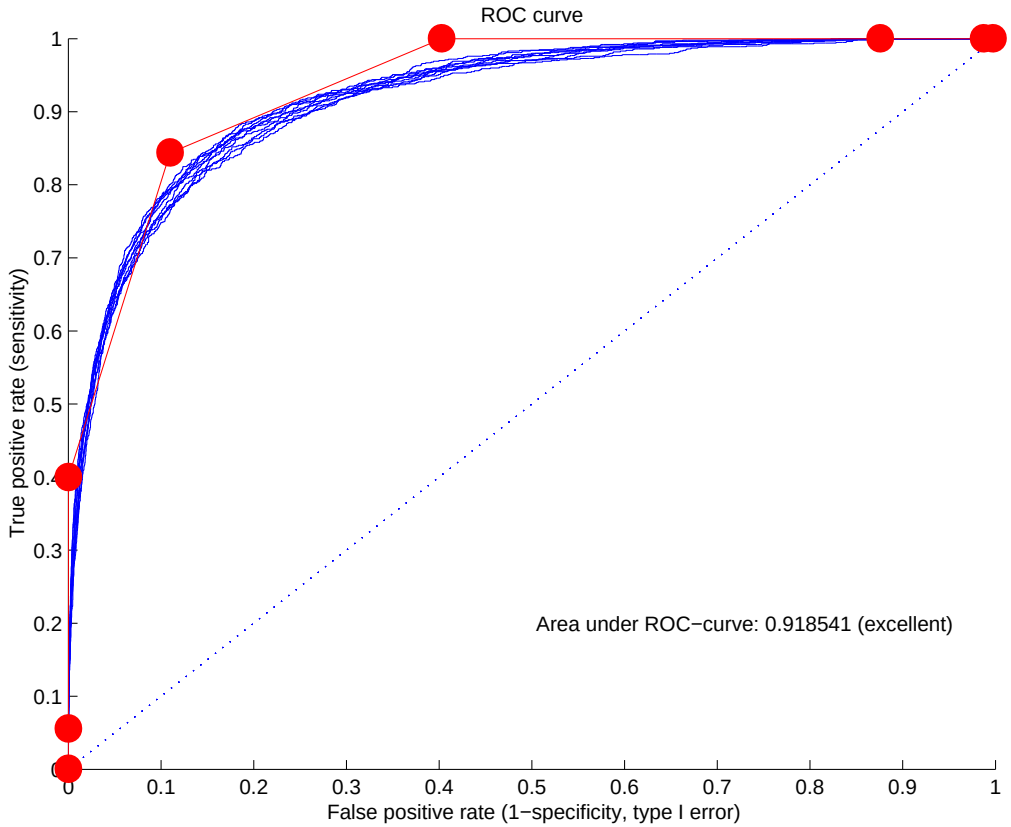


Figure 2: ROC curve with artificial data plotted.

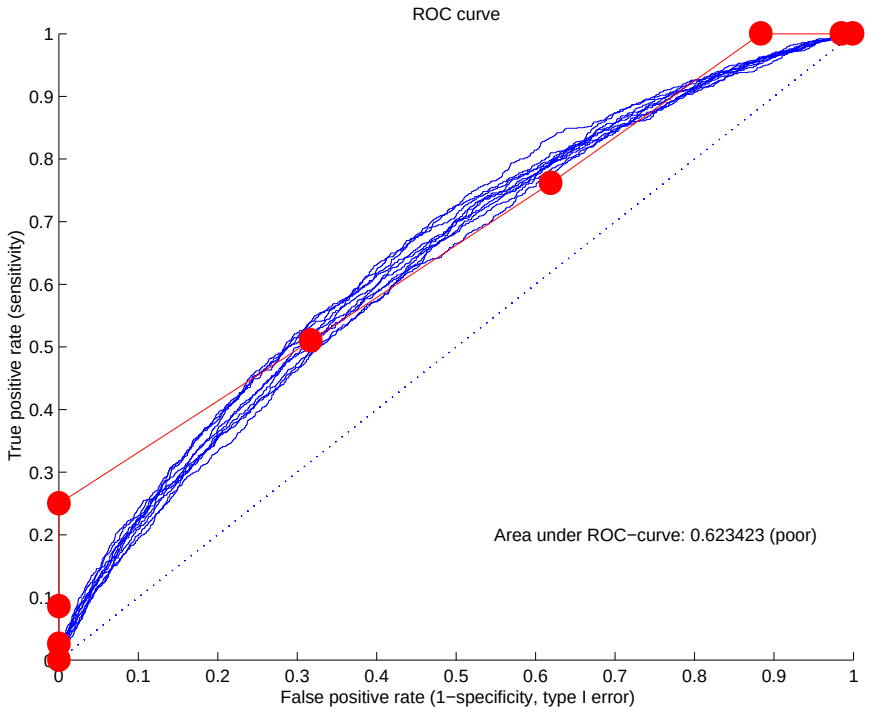
"Volume" with 1000 truly active, 9000 truly inactive "voxels".

Addition of independent Gaussian noise

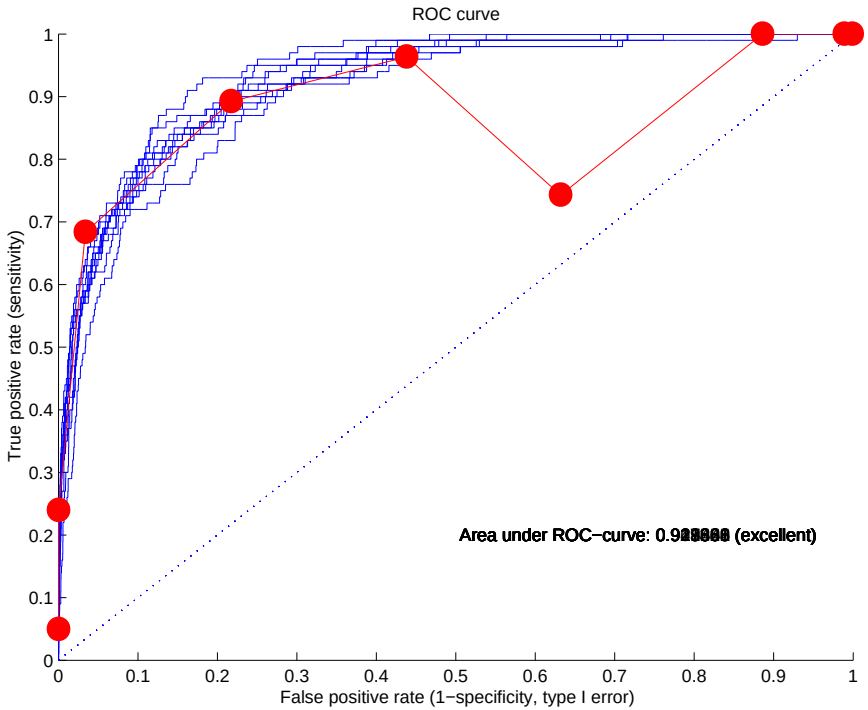
Estimation of P_A, P_I and λ from 10 replications at different thresholds.

Plotting of P_A (y-axis) and P_I (x-axis) in the ROC plot (the red dots).

ROC and binomial mixture — Example ...

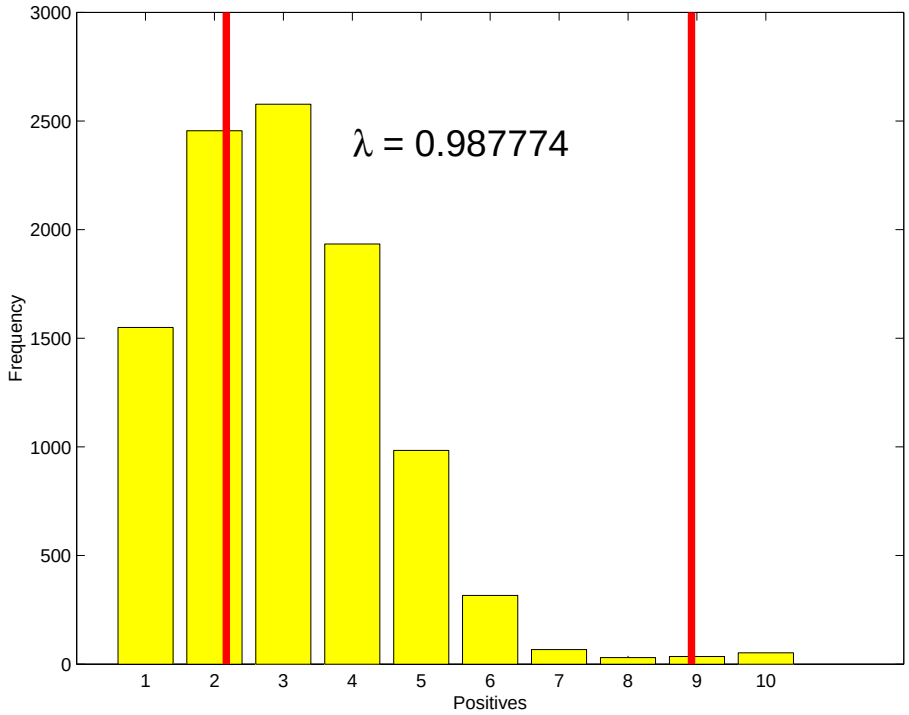


(a) More noise



(b) Fewer truly active: 100

ROC and binomial mixture — Example ...



Bar plot of $\mathbf{n}$.

Estimates of P_A, P_I and λ

The two modes of $\mathbf{n}$ is in this case modeled well.

In this case the mixing coefficient is also modeled well:

$$\lambda = 0.988 \approx 9900/10000 \qquad (4)$$

Binomial mixture for ROC

A noisy estimate of the ROC curve.

Might be wrong if there is an imbalance between the number of active and inactive voxels.

Assumption on spatial independence among voxels.

Bias/variance: The binomial mixture will only model the variance. If the methods are systematically wrong then this will not be accounted for.

Digitization

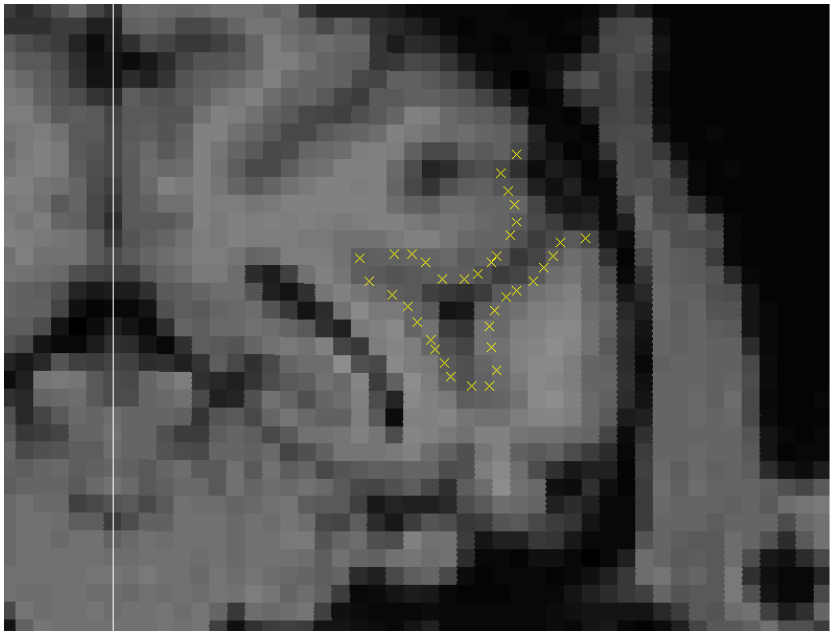


Figure 3: Digitization of points along STS at $y = -3$mm.

Digitization of superior temporal sulcus (STS) from coronal anatomical T1 (jerom_cmpr) images $y = 21$mm to $y = -8$mm.

Focus on specific interesting area

Global multivariate methods are influenced by noise and signal from other areas.

Superior temporal sulcus

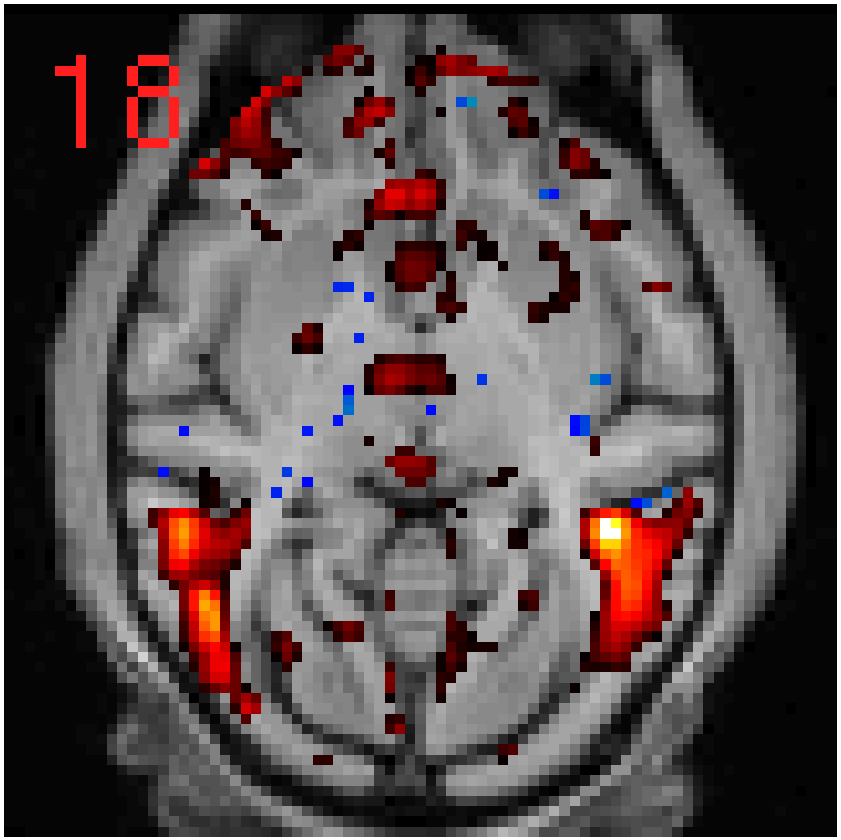
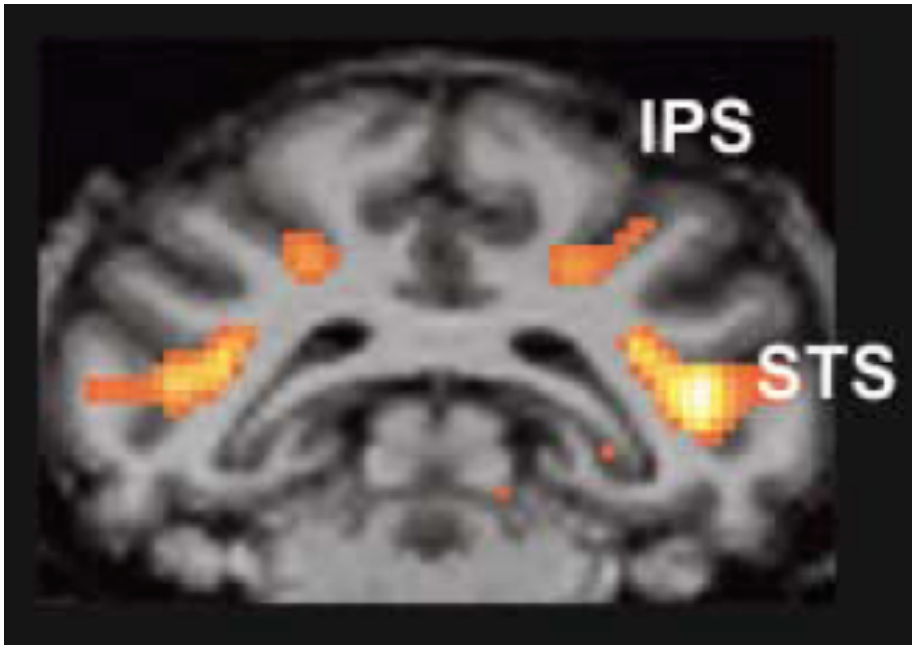


Figure 4: Z-score map for harmonic analysis threshold at $|z| > 2$ with `jerom_cmpr` as background. Jerom session 25, run 17.

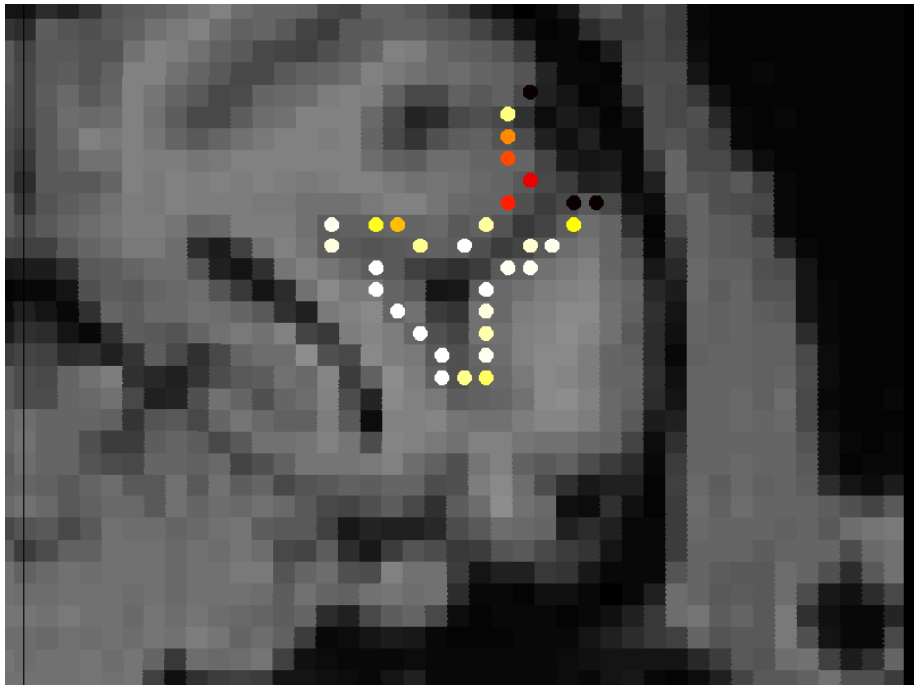
Superior temporal sulcus (STS) contains MT visual motion area.

Strong signal around $(18, -5, 18)$ and $(18, -4, 17)$ Jerom space in Jerom session 25.

Here analyzed with Fisher's G that compares the strongest periodogram component with the rest (Fisher, 1929; Fisher, 1940). Suggested for automated analysis in neuroimaging (Van Horn et al., 2002). No parameters in the model, Experiment should be periodic



(a) From Vanduffel et al., 2001, figure 2b, MION signal from Jerom at $y \approx -3$mm



(b) With digitization. Analyzed with Fisher's G. Hot color scale

ROC curves for Jerom data

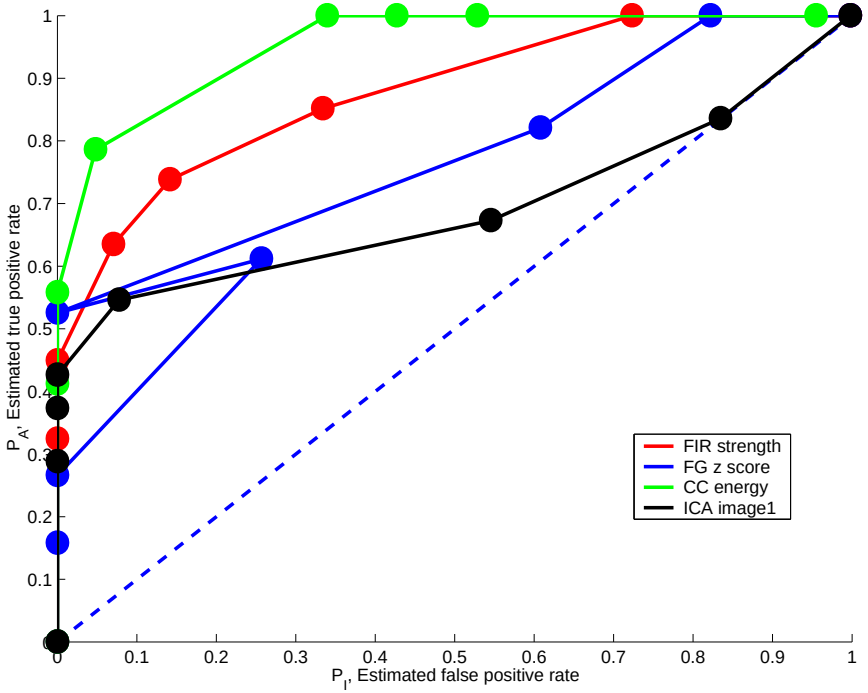


Figure 5: Estimated ROC curves.

Estimated ROC curves from different analyses on Jerom, session 25.

Nine replications: Type A runs.

Ranking: Cross-correlation $>$ FIR $>$ Fisher G $>$ ICA

Performance estimate of cross-correlation might be biased, e.g., it does not divide by the noise.

"Dependent" likelihood

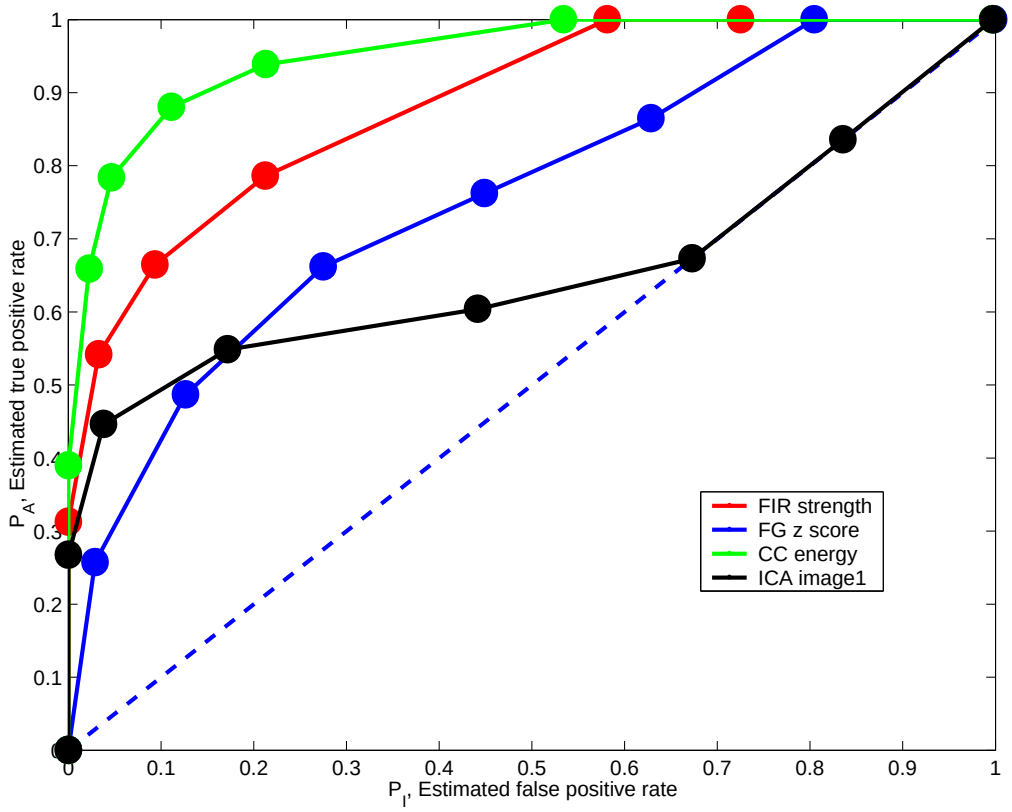
λ common for all k: Interpretable as the number of truly positives should not change across the ROC curve.

$$p(\mathbf{N}|P_{A,1}, P_{I,1}, \ldots, P_{A,K}, P_{I,K}, \lambda) \propto \sum_{k=1}^{K} \sum_{m=0}^{M} n_{k,m} \ln \Big[\lambda P_{A,k}^{m}(1-P_{A,k})^{M-m} + (1-\lambda)P_{I,k}^{m}(1-P_{I,k})^{M-m}\Big],$$

where K is the number of points, e.g., along the ROC curve.

Matrix with sufficient statistics $\mathbf{N}(K \times M+1)$

ROC curve with common λ



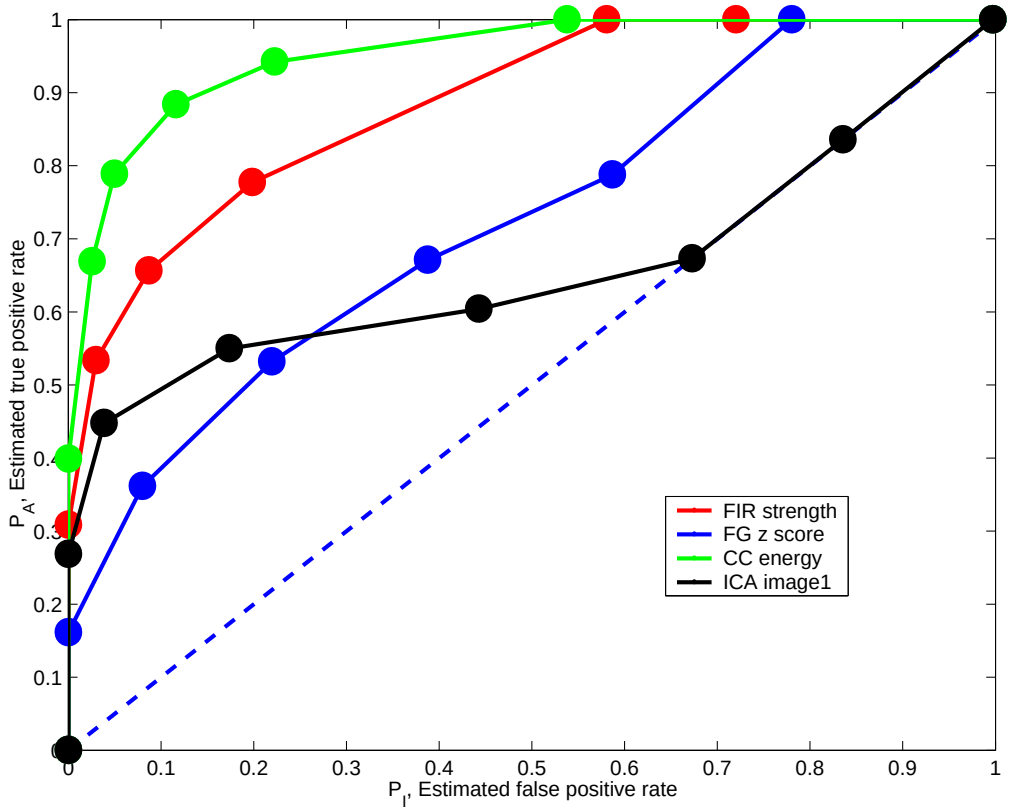
ROC curve with common λ true the ROC curve.

Smooth monotonous curve.

Still different λ for each analysis method: $1-\lambda$: 0.56, 0.50, 0.81, 0.60

Figure 6: ROC curve with common λ.

ROC curve with common λ



ROC curve with common λ true the ROC curve and across models

Common $\lambda = 0.43$.

Small differences.

Figure 7: ROC curve with common λ.

SVD of summary images

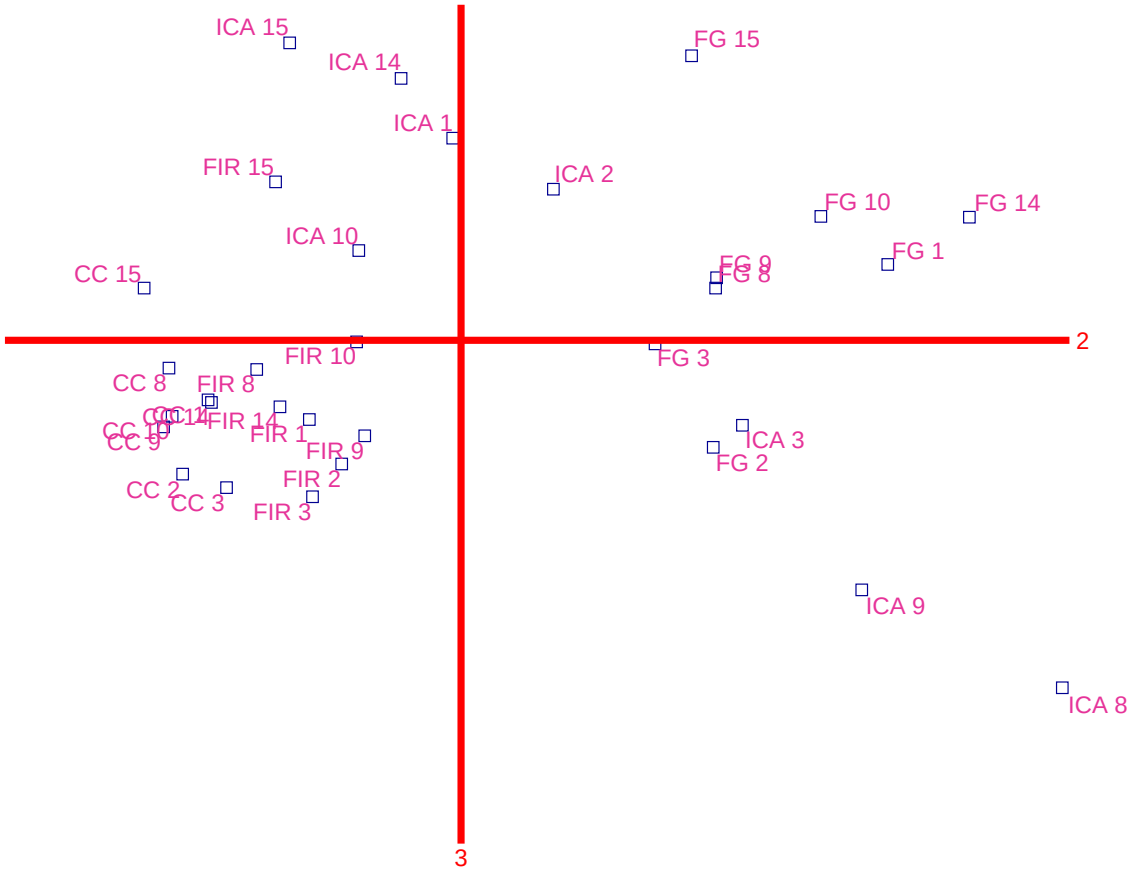


Figure 8: Second and third principal component.

Singular value decomposition (SVD) of summary images across runs and methods.

Histogram equalized to a Gaussian distribution.

ICA show the highest variance and cross-correlation the lowest, i.e., in accordance with the ROC curve.

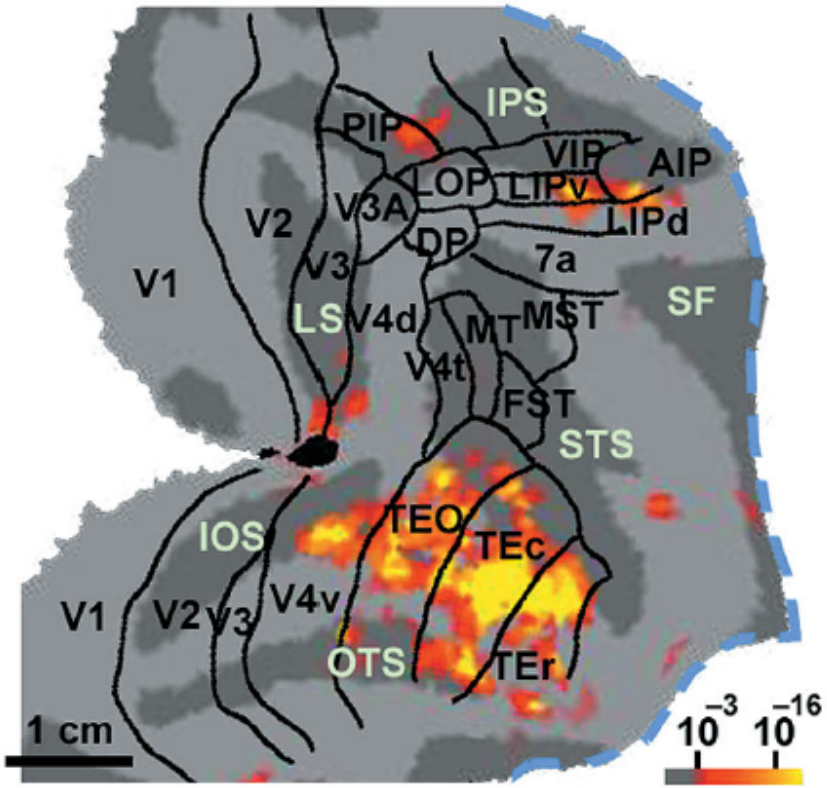
Clustering

K-means clustering (Goutte et al., 1999; Goutte et al., 2001; Balslev et al., 2002) implemented in Lyngby and Brede.

Unsupervised segmentation from functional data: Dynamic complex visual scene segmented (James Bond movie) with independent component analysis (Zeki et al., 2003).

Unsupervised segmentation from diffusion data: Thalamic nuclei with K-means (Wiegell et al., 2003).

Segmentation of STS

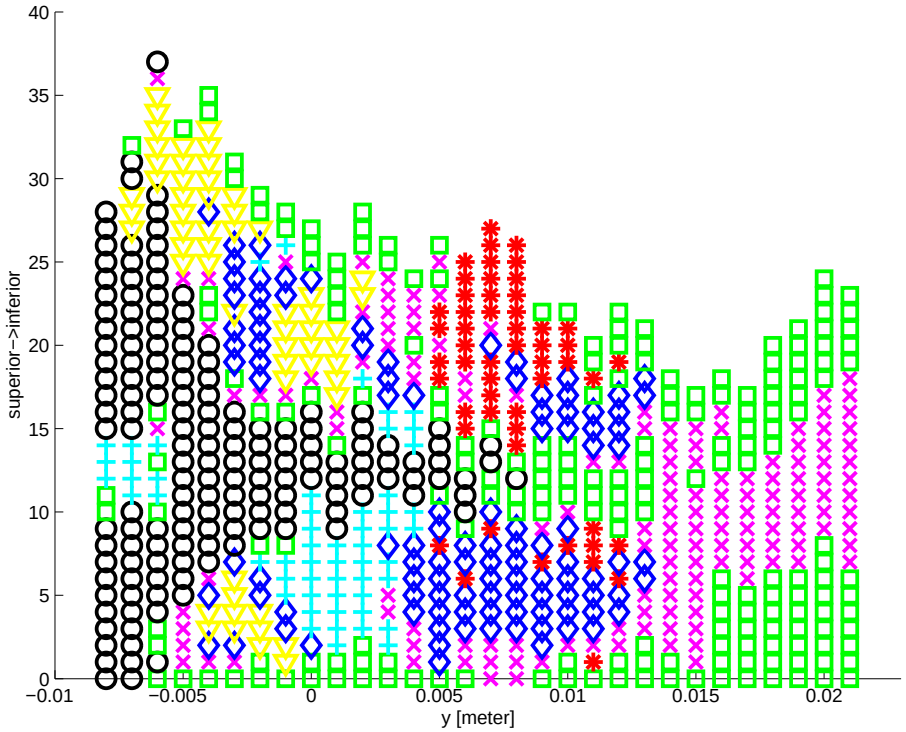


Functional areas within superior temporal sulcus (STS):

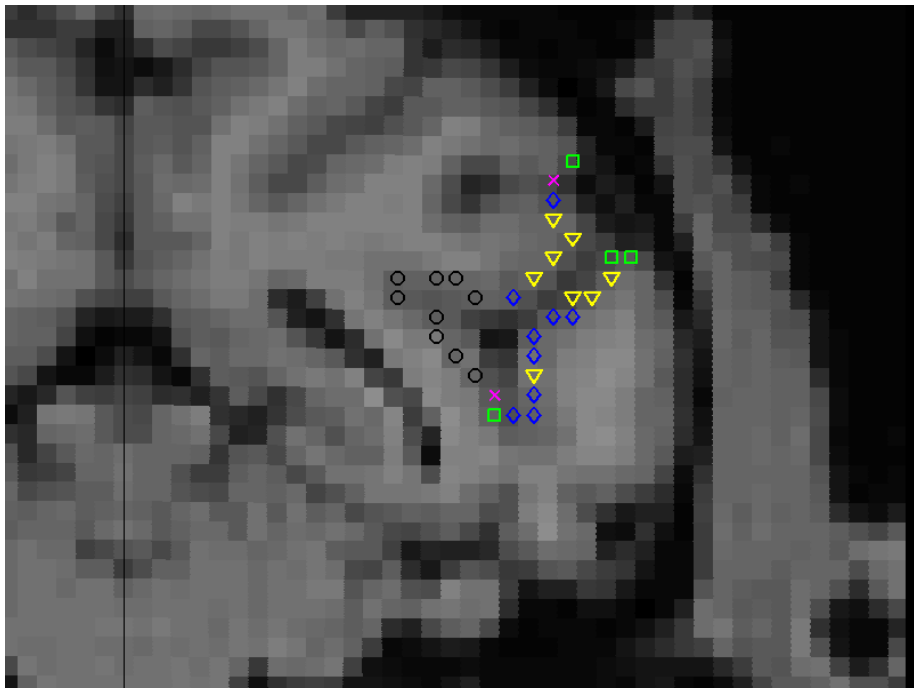
V4t, MT, MST, FST, TEO, TEc, TEr

Figure 9: Segmentation of STS. From (Tsao et al., 2003, figure 1a).

Segmentation of STS



(a) "Flatmap"



(b) $Y = 3$mm

Relations to workpackages

Workpackage 5: Comparative study of fMRI models.

Deliverable 5.1, Publication ROC evaluation consensus artificial data: Consensus models (Hansen et al., 2001).

Deliverable 5.2, Software ROC evaluation consensus artificial data: Functions implemented in Lyngby (Hansen et al., 1999) and Brede (Nielsen and Hansen, 2000): `lyngby_cons_main`, `brede_vol_plot_roc`, `brede_pde_binmix`

Deliverable 5.3, Publication ROC consensus 2DG data: No 2DG/fMRI data available.

Workpackage 4: Novel approaches to generation of activity maps

Deliverable 4.1, Publication on feature extraction: Feature space clustering (Goutte et al., 2001).

Deliverable 4.6, Software feature extraction & pattern recognition technique for activity maps: Lyngby and Brede extended with, e.g., meta clustering, independent component analysis, non-negative matrix factorization and Fisher's G.

Workpackage 3: Warping (intra- and intersubject)

Bibliography on Image Registration, a small list of methods and tools
http://www.imm.dtu.dk/~fn/bib/Nielsen2001BibImage/

Conclusion

ROC curves without ground truth possible but probably biased estimate.

Lyngby and Brede able to analyze data that is not necessarily a volume, e.g., might be interesting for functional segmentation of local areas based on fMRI.

References

Balslev, D., Nielsen, F. Å., Frutiger, S. A., Sidtis, J. J., Christiansen, T. B., Svarer, C., Strother, S. C., Rottenberg, D. A., Hansen, L. K., Paulson, O. B., and Law, I. (2002). Cluster analysis of activity-time series in motor learning. *Human Brain Mapping*, 15(3):135–145. http://www3.interscience.wiley.com/cgi-bin/abstract/89011762/. ISSN 1097-0193.

Constable, R. T., Skudlarski, P., and Gore, J. C. (1995). An ROC approach for evaluating functional brain MR imaging and postprocessing protocols. *Magnetic Resonance in Medicine*, 34(1):57–64. PMID: 7674899. ISSN 0740-3194.

Fisher, R. A. (1929). Tests of significance in harmonic analysis. *Proceedings of the Royal Society, A*, 125:54–59.

Fisher, R. A. (1940). On the similarity of the distributions found for the test of significance in harmonic analysis, and in Steven's problem in geometrical probability. *Annals of Eugenics*, 10:14–17.

Gelfand, A. E. and Solomon, H. (1974). Modeling jury verdict in the american legal system. *Journal of the American Statistical Association*, 69(345):32–37.

Genovese, C. R., Noll, D. C., and Eddy, W. F. (1996). Statistical estimation of test-rest reliability in fMRI. In *Proceedings of the International Society of Magnetic Resonance in Medicine, Fourth Scientific Meeting and Exhibition*, volume 1, page 345, Berkeley, California, USA. Society of Magnetic Resonance in Medicine. ISSN 1065-9889.

Genovese, C. R., Noll, D. C., and Eddy, W. F. (1997). Estimating test-retest reliability in functional MR imaging I: Statistical methodology. *Magnetic Resonance in Medicine*, 38:497–507. Presentation of a method for assessing the reliability of fMRI methods with the use on a binomial mixture model.

Goutte, C., Hansen, L. K., Liptrot, M. G., and Rostrup, E. (2001). Feature-space clustering for fMRI meta-analysis. *Human Brain Mapping*, 13(3):165–183. PMID: 11376501. http://www3.interscience.wiley.com/cgi-bin/abstract/82002382/START.

Goutte, C., Toft, P., Rostrup, E., Nielsen, F. Å., and Hansen, L. K. (1999). On clustering fMRI time series. *NeuroImage*, 9(3):298–310.

Hansen, L. K., Nielsen, F. Å., Strother, S. C., and Lange, N. (2001). Consensus inference in neuroimaging. *NeuroImage*, 13(6):1212–1218. PMID: 11352627. http://www.idealibrary.com/links/doi/10.1006/nimg.2000.0718.

Hansen, L. K., Nielsen, F. Å., Toft, P., Liptrot, M. G., Goutte, C., Strother, S. C., Lange, N., Gade, A., Rottenberg, D. A., and Paulson, O. B. (1999). "lyngby" — a modeler's Matlab toolbox for spatio-temporal analysis of functional neuroimages. In Rosen, B. R., Seitz, R. J., and Volkmann, J., editors, *Fifth International Conference on Functional Mapping of the Human Brain, NeuroImage*, volume 9, page S241. Academic Press. http://isp.imm.dtu.dk/publications/1999/hansen.hbm99.ps.gz. ISSN 1053–8119.

Lange, N., Hansen, L. K., Anderson, J. R., Nielsen, F. Å., Savoy, R., Kim, S.-G., and Strother, S. C. (1998). An empirical study of statistical model complexity in neuro-fMRI. *NeuroImage*, 7(4, part 2):S764.

Lange, N., Strother, S. C., Anderson, J. R., Nielsen, F. Å., Holmes, A. P., Kolenda, T., Savoy, R., and Hansen, L. K. (1999). Plurality and resemblance in fMRI data analysis. *NeuroImage*, 10(3):282–303. PMID: 10458943. DOI: 10.1006/nimg.1999.0472. http://www.sciencedirect.com/science/article/B6WNP-45FCP48-13/2/bd7e7f72099b83540609e24c627a2fc4.

Nielsen, F. Å. and Hansen, L. K. (2000). Experiences with Matlab and VRML in functional neuroimaging visualizations. In Klasky, S. and Thorpe, S., editors, *VDE2000 - Visualization Development Environments, Workshop Proceedings, Princeton, New Jersey, USA, April 27–28, 2000*, pages 76–81, Princeton, New Jersey. Princeton Plasma Physics Laboratory. http://www.imm.dtu.dk/pubdb/views/edoc_download.php/1231/pdf/imm1231.pdf. CiteSeer: http://citeseer.ist.psu.edu/309470.html.

Noll, D. C., Genovese, C. R., Vazquez, A. L., and Eddy, W. (1996). Evaluation of respiratory artifact correction techniques in fMRI using ROC analysis. In *Proceedings of the International Society of Magnetic Resonance in Medicine, Fourth Scientific Meeting and Exhibition*, volume 1, page 343, Berkeley, California, USA. Society of Magnetic Resonance in Medicine. ISSN 1065-9889.

References

Sorenson, J. A. and Wang, X. (1996). ROC methods for evaluation of fMRI techniques. *Magnetic Resonance in Medicine*, 36(5):737–744. PMID: 8916024. ISSN 0740-3194.

Tsao, D. Y., Greiwald, W. A., Knutsen, T. A., Mandeville, J. B., and Tootell, R. B. H. (2003). Faces and objects in macaque cerebral cortex. *Nature Neuroscience*, 6(9):989–995.

Van Horn, J. D., Woodward, J., Aslam, J., Grethe, J., and Gazzaniga, M. (2002). Statistical time course feature vectors for use in rapid assessment and clustering. *NeuroImage*, 16(2). http://www.academicpress.com/journals/hbm2002/15098.html. Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2–6, 2002, Sendai, Japan. Available on CD-Rom.

Wiegell, M. R., Tuch, D. S., Larsson, H. B., and Wedeen, V. J. (2003). Automatic segmentation of thalamic nuclei from diffusion tensor magnetic resonance imaging. *NeuroImage*, 19(2):391–401. PMID: 12814588.

Zeki, S., Perry, R. J., and Bartels, A. (2003). The processing of kinetic contours in the brain. *Cerebral Cortex*, 13(2):189–202. PMID: 12507950. WOBIB: 52. ISSN 1047-3211.