Revealing Representational Content with Pattern-Information fMRI – an Introductory Guide

Marieke Mur[1,2*], Peter A. Bandettini[1,3], Nikolaus Kriegeskorte[1]

[1]Section on Functional Imaging Methods, Laboratory of Brain and Cognition,
National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA

[2]Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands

[3]Functional Magnetic Resonance Imaging Facility, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA

* Corresponding author. Email: marieke@mail.nih.gov

ABSTRACT

Conventional statistical analysis methods for functional magnetic resonance imaging (fMRI) data have been very successful in detecting brain regions that are involved in specific mental activities. However, these conventional methods do not look *into* the regions, in that they do not consider the multivoxel patterns of activity within a brain region. These patterns of activity reflect neuronal population activity, which is thought to represent mental content. The content of regional representations can be investigated by pattern-information analysis, which targets the information carried by a region's multivariate response pattern. This tutorial introduction motivates pattern-information analysis, explains its underlying assumptions, introduces the most widespread methods in an intuitive way, and outlines the basic sequence of analysis steps.

INTRODUCTION

Conventional statistical analysis of functional magnetic resonance imaging (fMRI) data focuses on finding macroscopic brain regions that are "involved" in specific mental activities (Friston et al., 1994; Friston et al., 1995; Worsley and Friston, 1995). In order to find and characterize brain regions that become activated as a whole, data is spatially smoothed and activity is averaged across voxels within a region of interest (ROI). These analysis steps increase sensitivity to spatially extended activations, but result in loss of sensitivity to fine-grained spatial-pattern information. In recent years, there has been a growing interest in going beyond *activation* assessment and analyzing fMRI data for the *information* carried by fine-grained patterns of activity within each functional region (Norman et al., 2006; Haynes and Rees, 2006; Kriegeskorte and Bandettini, 2007a). The goal of this tutorial paper is to motivate the use of pattern-information analysis and to provide a step-by-step introduction on how to implement this method.

A region's *involvement* in task processing versus its *representational content*

Conventional analysis focuses on regions that become activated as a whole during the performance of a specific task. This motivates spatial smoothing of the data and averaging of activity across an ROI. Since this approach focuses on "activations" in the sense of blobs consisting of multiple voxels all showing effects in the same direction, we refer to it as activation-based analysis. Activation-based analysis is highly sensitive to extended activations, which are usually interpreted to indicate the "involvement" of the region in a specific mental function. However, activation-based analysis disregards information carried by patterns of activity within an ROI. These patterns are treated as noise in the conventional approach to imaging analysis, even though they can reflect the population activity that constitutes the internal representation of each functional region (for a striking example, see Kamitani and Tong, 2005). The blood-oxygen-level-dependent (BOLD) fMRI signal provides an indirect and complex summary of underlying neural activity and is affected by noise (Boynton et al., 1996; Logothetis, 2008). As a consequence, interpretation of the BOLD fMRI signal in terms of underlying neural activity requires some caution. Nevertheless, BOLD fMRI signal changes may be interpreted as representing changes in underlying neural activity (Logothetis et al., 2001).

Similarly, the precise fine-grained activity patterns measured by fMRI do not *directly* reflect neural activity patterns. However, a change of signal (patterns) across conditions can be interpreted as a change of neural population activity. The same spatial-average activation could result from different patterns encoding different representational content (see Fig. 1). These different patterns will go undetected by activation-based analysis. A recent study using pattern-information analysis showed that perceptually discriminable speech sounds are represented by different patterns of activity in right auditory cortex; however, these speech sounds elicited similar spatial-average activations (Raizada et al., 2008). This example illustrates how pattern-information analysis allows us to look into predefined regions and read out and analyze information carried by local population codes (see Table 1, Fig. 1).

The use of pattern-information analysis is not restricted to investigating functional regions defined by activation-based analysis. It can also be used to investigate patterns of activity across more widely distributed sets of voxels (e.g. Haxby et al., 2001; Carlson et al., 2003) or to *define* functional regions by mapping the whole volume for effects using a multivariate searchlight ("information-based brain mapping", Kriegeskorte et al., 2006, 2007). The change that activation-based analysis is sensitive to – all voxels changing their activity in *the same direction* – can be viewed as a special case of the changes that pattern-information analysis can detect: any change of the pattern, including spatial-mean activity changes as well as pattern changes where the spatial-mean is unaffected. This general sensitivity makes pattern-information analysis a powerful statistical tool. With many successful applications in neuroimaging, the approach has gained momentum in recent years (e.g. Haxby et al., 2001; Carlson et al., 2003; Cox and Savoy, 2003; Friston et al., 2008; Hanson et al., 2004; Kamitani and Tong, 2005; Haynes and Rees, 2005; Haynes et al., 2007; Kriegeskorte et al., 2007; Kriegeskorte et al., 2008 (in press); Mourao-Miranda et al., 2005; Mitchell et al., 2008; O'Toole et al., 2005; Pereira et al., 2008 (submitted); Raizada et al., 2008 (submitted)). Note that related multivariate methods as well as prediction frameworks have been explored before in neuroimaging analysis (Strother et al., 2002; Worsley et al., 1997), but with different conceptual goals.

Pattern-information fMRI is fundamentally limited by the amount of information about the neural population codes that can be provided by fMRI. Voxel resolution is one such limitation, thus motivating the use of high-resolution fMRI in conjunction with pattern-information analysis (Kriegeskorte and Bandettini, 2007a; Kriegeskorte et al., 2007). A technique that also targets the representational content of functional regions and that is not limited by voxel resolution is fMRI adaptation (Grill-Spector and Malach, 2001). This approach can potentially resolve sub-voxel representations by inferring neural selectivity from fMRI adaptation responses. However, the interpretation of positive findings ("release from adaptation") in terms of neural population selectivity relies on assumptions that have been questioned by recent experimental results (Tolias et al., 2005; Sawamura et al., 2006; Krekelberg et al., 2006). These results showed that release from adaptation does not necessarily reflect selectivity of the underlying neural population as measured by classical electrophysiological methods. Other explanations, e.g. attentional effects or carry-over of effects from connected regions (Tolias et al., 2005; Krekelberg et al., 2006), can account for release from adaptation as well. While the fMRI adaptation paradigm compares activation between pairs of either different or repeated stimuli and then *infers* single-stimulus selectivity from these activation differences, pattern-information fMRI follows the simpler logic of contrasting experimental conditions directly to determine if there is an effect on the dependent variable: the activity pattern within an ROI. Although its sensitivity is limited by the measurement technique of fMRI, a positive result, i.e. statistically distinct activity patterns, provides strong evidence for a difference between the underlying neural activity patterns in the region. It has recently been shown that it is possible to combine pattern-information fMRI and fMRI adaptation in a single experiment and simultaneously estimate activity patterns and adaptation effects (Aguirre, 2007).

PATTERN-INFORMATION ANALYSIS: STEP-BY-STEP

This section describes the basic steps of pattern-information analysis. First, we describe how to test for a multivariate activity-pattern difference. A significant pattern difference implies that the condition can be decoded (with some accuracy above chance level) from the activity patterns. In other words, it implies pattern-information about the experimental condition. We also discuss what study design and imaging parameters to use in pattern-information fMRI. Then, we provide a step-by-step description of the methods for extracting patterns of activity from fMRI data and for analyzing these patterns. These steps are summarized in Figure 3.

Testing a multivariate pattern difference

There is a wide variety of multivariate methods that can be used for pattern-information analysis. All these methods aim to determine whether the patterns of activity associated with different conditions are statistically discriminable (i.e. significantly different). As in conventional analysis, every activity pattern we estimate from the data results from a combination of true effects and noise. Noise is always present and will make every pattern unique (just as in a univariate t-test there is always a small difference between the estimates of the two means to be compared, even if the null hypothesis is true). We need to determine whether the patterns associated with, say, condition A and condition B, are more different than expected under the null hypothesis of equal activity patterns in both conditions. Under the null hypothesis, any differences between the pattern estimates would be produced by noise alone.

Univariate data is usually analyzed using a t-test or analysis of variance (ANOVA). For multivariate data, the equivalent method would be a multivariate analysis of (co)variance (MANOVA). However, this method assumes that the distribution of the residuals is multinormal, an assumption that might not hold for fMRI data. This is one reason why most of the cited studies approach pattern analysis as a classification problem: If we can classify the experimental conditions (which elicit the representational states we are interested in) on the basis of the activity patterns better than chance, this indicates that the response pattern carries information about the experimental conditions. This approach has been referred to as “brain reading” (Cox and Savoy, 2003) or “decoding”.

Linear decoding: the most widespread and successful pattern-information analysis in neuroimaging so far

Multivoxel patterns of activity can be viewed as points in multidimensional space (with as many dimensions as voxels). One way to classify an activity pattern is to assign it to the condition which centroid (multivariate mean) it is closest to in multivariate space. This method is referred to as minimum-distance classification (Fig. 2a). Multivariate distance between centroid and to-be-classified pattern can be measured by Euclidean distance (the length of a straight line that connects two points) or correlation distance (1-correlation; e.g. Haxby et al., 2001). Implicit to this minimum-distance classification is a linear decision boundary (i.e. a hyperplane) in multidimensional space (Fig. 2a). Classifiers that use a linear decision boundary are referred to as *linear* or *hyperplane* classifiers.

The three most widespread methods in pattern-information fMRI (Fig. 2) are the minimum-distance classifier (e.g. Haxby et al., 2001), the linear support vector machine (linear SVM; e.g. Cox and Savoy, 2003) and Fisher linear discriminant analysis (FLDA; e.g. Carlson et al., 2003). These three methods are very similar in that they are all linear (i.e. hyperplane) classifier methods.[1] While each of them will have optimal sensitivity under slightly different circumstances, they tend to perform somewhat similarly on fMRI data and there is no strong evidence to date suggesting a general superiority of any one of them in this context (but see Ku et al., 2008; Mourao-Miranda et al., 2005).

Equivalent to placing a decision hyperplane in multivariate space, we can compute a linear combination (weighting) of the voxel responses and apply a decision threshold. This is equivalent because, geometrically, computing a linear combination corresponds to orthogonally projecting each activity pattern (point in multivariate space) onto a linear discriminant dimension (a line in multivariate space). The decision hyperplane is the

[1] Non-linear classification algorithms have also been used for pattern-information analysis (e.g. Cox and Savoy, 2003; LaConte et al., 2005). These algorithms can capture more complicated class boundaries than linear classifiers. However, non-linear classification methods are more prone to overfit the data than linear classification methods. Overfitting is a problem in fMRI because the number of repetitions in an fMRI study is typically not very large in relation to the number of voxels in the ROI. Overfitting leads to lower generalization performance (i.e. lower accuracy on the test data set) and a decrease in power for detecting linear pattern effects (step 5).

hyperplane orthogonal to the discriminant dimension and passing through the decision threshold (Fig. 2).

The three methods, then, differ only in how the voxel weights are determined. Each aims at achieving optimal classification performance in a slightly different way (for details see Fig. 2 and step 4, below). One intuitive method would be to weight each voxel by how well its activity discriminates two conditions, for example by using its t-value for the contrast between these two conditions (A-B). This means that a voxel responding more to condition A than B (positive t-value) will be given a positive weight, and a voxel responding more to condition B than A (negative t-value) will be given a negative weight. A voxel that responds similarly to A and B will be given a weight close to zero. The methods for voxel weighting shown in Figure 2b-c are mathematically more complex, but conceptually similar to using contrast t-values as voxel weights.

Study design

Both event-related and block designs can be used in combination with pattern-information analysis. The following points should be considered in choosing the design of your study. Block designs yield a higher functional contrast-to-noise ratio than event-related designs. This holds both for constant inter-stimulus-interval (ISI) event-related designs (Bandettini and Cox, 2000) and jittered rapid event-related designs (Birn et al., 2002). This means that block designs will yield better estimates of the average response pattern (i.e. the centroid) for a small number of conditions. Therefore, they are preferred if you would like to discriminate a small number of conditions that are expected to elicit relatively different response patterns (e.g. Haxby et al., 2001). However, if you would like to make more fine-grained condition discriminations, especially in combination with a large number of conditions, an event-related design becomes an appropriate choice. Event-related designs can handle larger condition sets and yield more independent data points than block designs and will therefore give a better estimate of the shape of each condition's multivariate response distribution. This additional information can benefit classification methods in making fine-grained discriminations between conditions.

Imaging parameters

Imaging parameters will largely follow from the experimental question and design. If information on a fine spatial scale is of interest, high-resolution fMRI (Kriegeskorte et al., 2007) might be the best choice. However, the tradeoff between the functional-contrast-to-noise ratio and the resolution has to be carefully considered (Kriegeskorte and Bandettini, 2007a). A voxel size of about 2-mm in each dimensions appears to be a reasonable compromise at 3 Tesla. It is important to note, however, that most pattern-information analyses so far have utilized lower-resolution fMRI data (see Haxby et al., 2001; Kamitani and Tong, 2005; Haynes and Rees, 2005), indicating that larger-scale patterns – even if dominated by vascular effects – can contain a considerable amount of information even about quite fine-grained neuronal patterns (consider Kamitani and Tong 2005).

STEP 1: Data splitting and preprocessing

Before analysis, the data should be split into an independent training and test set to ensure unbiased testing results. The training data set should be used for voxel selection (step 3) and classifier training (step 4). Both these steps involve voxel weighting, either binary (voxel selection) or continuous (classifier training). Voxel weighting can bias testing results if performed on the same data and therefore it is crucial to use an independent data set for classifier testing (step 5). To make sure the data are independent, the two sets should be based on different runs (i.e. even and odd runs) that use independent stimulus sequences. One option is to split the data into two halves. However, the training data set is generally chosen to be larger than the test set in order to obtain stable voxel weights. Efficient use of the data can be achieved by cross-validation: divide the data into a number of independent subsets (e.g. five or ten, each a run in your experiment), use all but one subset as training data and use the left out subset as test data; then repeat this procedure until each subset has been used as test data once. Performance on the different subsets is combined to obtain overall classifier performance. Preprocessing should be performed separately for training and test data sets. If you use cross-validation, this means that each subset should be preprocessed separately. Preprocessing steps are the same as in activation-based analysis (i.e. slice-scan-time correction, motion correction, trend removal), except that the data should not be spatially smoothed to preserve fine-grained pattern information.

STEP 2: Extracting the single-subject patterns

Single-subject patterns are extracted by univariate analysis at each voxel using the general linear model (GLM) (Friston et al., 1994, 1995; Worsley and Friston, 1995). Each condition or each example belonging to a condition (if estimating the shape of the response distribution) is a predictor in the model. This part of the analysis is identical to activation-based analysis and will yield a beta-value for each predictor and voxel. The beta-values for one predictor across voxels form the pattern of activity for a specific condition (e.g. Haxby et al., 2001). For block designs or slow event-related designs, where BOLD responses to different conditions do not overlap, it is possible to stay closer to the data and use temporally averaged normalized signal intensity values as patterns of activity (e.g. Kamitani and Tong, 2005). Pattern extraction yields a set of training patterns and a set of test patterns. The patterns should not be averaged across subjects to preserve fine-grained subject-specific information. This implies that analysis is performed in native subject space.

STEP 3: Selecting the voxels

Once activity values are computed, the next step is to decide which voxels to include for pattern-information analysis. These voxels are selected using the training data set or another data set independent from the test set (e.g. anatomical data or functional data from a separate block-localizer experiment). One option would be to analyze the patterns of activity in a specific ROI. If defined by activation-based analysis, ROIs will be spatially contiguous sets of voxels, but they do not have to be. For example, to investigate object-category discrimination, the most visually responsive voxels in object-selective cortex could be selected for subsequent analysis, irrespective of whether these voxels are adjacent or not. A computationally more demanding option would be to analyze the pattern of activity across all brain voxels. This would increase informational content, but it will also add noise. In addition, most classification algorithms (especially FLDA, see Cox and Savoy, 2003) show a decrease in performance if the number of voxels heavily outnumbers the number of training patterns. Possible solutions include selecting fewer voxels and transforming the original voxel space into a lower dimensional space (dimensionality reduction). Voxels can also be selected using information-based brain mapping (Kriegeskorte et al., 2006, 2007). This can be seen as the multivariate equivalent of univariate statistical parametric mapping (SPM) (Friston et al., 1995).

STEP 4: Training the classifier

To investigate whether a region's pattern of activity discriminates two conditions, we first use the training data set to determine a set of weights (one for each voxel) that linearly combines the voxel responses in such a way as to maximize the difference between the two conditions (classifier training). Beyond the minimum-distance classifier described above, the two other frequently used linear methods are Fisher linear discriminant analysis (Fig. 2b; e.g. Carlson et al., 2003; Haynes and Rees, 2005; Kriegeskorte et al., 2007) and linear support vector machines (Fig. 2c; e.g. Cox and Savoy, 2003; Kamitani and Tong, 2005). These methods will perform optimally (thus providing maximum sensitivity to pattern information) under different assumptions about the distribution of the response patterns (Fig. 2). In neuroimaging, however, results are often similar. This suggests that choosing any of them is acceptable; if more than one method is used all results need to be reported (picking the significant result among different approaches would require correction for multiple comparisons).

FLDA computes the weights by maximizing the ratio of between-condition and within-condition variance. It assumes multinormality of the residuals and homoscedasticity (distributions are the same across conditions). Note that, in contrast to MANOVA, the specificity of FLDA is not dependent on the assumption of multinormality of the residuals because classification algorithms use independent data sets for training and testing. Strong violations of multinormality will affect sensitivity, but not specificity, so a test of pattern information is valid. Linear SVM does not assume multinormality or homoscedasticity, but it does assume linear separability. It searches for the linear boundary that has the maximum margin (maximum separation from the nearest response patterns) (Fig. 2c). The response patterns on the margins are referred to as the "support vectors", because they "support" the margins and define the decision hyperplane.

Most classifiers can also be trained on data from multi-condition experiments (Pereira et al., 2008). However, multi-class discriminations are often approached as a combination of multiple two-class discriminations. This approach is motivated by the fact that two-class discriminations are generally of neuroscientific interest, even if an experiment contains more than two conditions. For a detailed overview on using linear classification algorithms, and their mathematical descriptions, see Pereira et al. (2008). In addition, several pattern analysis toolboxes are listed in the reference section of this paper.

STEP 5: Testing the classifier

The weights computed during training are set to yield optimal classification performance on the training data set. To test whether good classification performance generalizes (i.e. is not based largely on noise present in the training data set), the weights are applied to an independent test data set. Performance of the classifier on the test data set can be measured by percent correct classification (accuracy). The null hypothesis is that the classifier would perform at chance level. To test whether classification accuracy is significantly better than chance, we can use a chi-square test (or a Monte-Carlo method in case of few observations). If the statistical test shows a significant result, this indicates that the region's response contains information about the experimental conditions. Another way to test the classifier is to perform a univariate t-test on the projected test patterns (Kriegeskorte et al., 2007). As described above, projection (voxel weighting) converts the activity patterns into one-dimensional values. These values can then be analyzed by a conventional univariate t-test. Similar to a classification accuracy that is significantly better than chance, a significant t-value for the difference between the two conditions would indicate that the region's response contains information about the experimental conditions.

CONCLUSION

Pattern-information analysis investigates the representational content of a region by analyzing the information carried by a region's pattern of activity. This information would not be detected by classical activation analysis and can significantly contribute to our understanding of neural representations of mental content. Pattern-information analysis invites high-resolution fMRI and typically does not involve spatial smoothing or subject-averaging to prevent loss of fine-grained pattern information. The most widespread method is linear decoding, which analyzes a region's activity pattern by means of a weighted sum of the single-voxel responses, with the weights chosen to maximally discriminate response patterns associated with different conditions. Statistical inference is performed on a data set independent of that used for ROI definition and voxel weighting to prevent statistical circularity.

The conceptual appeal of pattern-information fMRI is that it allows us to look into the regions and investigate their representational content. Recent neuroscientific successes in the domain of sensation and perception suggest that higher-order cognitive functions in the domain of social and cognitive neuroscience might also benefit from the pattern-information approach. .

REFERENCES

Aguirre, G.K. (2007). Continuous carry-over designs for fMRI. *Neuroimage 35*, 1480-1494.

Bandettini, P.A., Cox, R.W. (2000). Event-related fMRI contrast when using constant interstimulus interval: Theory and experiment. *Magnetic Resonance in Medicine, 43*, 540-548.

Birn, R.M., Cox, R.W., Bandettini, P.A. (2002). Detection versus estimation in event-related fMRI: Choosing the optimal stimulus timing. *Neuroimage, 15*, 252-264.

Boynton, G.M., Engel, S.A., Glover, G.H., Heeger, D.J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *The Journal of Neuroscience 16*, 4207-4221.

Carlson, T.A., Schrater, P., He, S. (2003). Patterns of activity in the categorical representations of objects. *Journal of Cognitive Neuroscience, 15*, 704-717.

Cox, D.D., Savoy, R.L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage 19*, 261-270.

Friston, K.J., Jezzard, P., Turner, R. (1994). Analysis of functional MRI time-series. *Human Brain Mapping 1*, 153-171.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.-P., Frith, C.D., Frackowiak, R.S.J. (1995). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping 2*, 189-210.

Friston, K.J., Holmes, A.P., Poline, J.-B., Grasby, P.J., Williams, S.C.R., Frackowiak, R.S.J., Turner, R. (1995). Analysis of fMRI time-series revisited. *Neuroimage 2*, 45-53.

Friston, K., Chu, C., Mourao-Miranda, J., Hulme, O., Rees, G., Penny, W., Ashburner, J. (2008). Bayesian decoding of brain images. *Neuroimage, 39*, 181-205.

Grill-Spector, K., Malach, R. (2001). fMRI-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychologica, 107*, 293-321.

Haxby, J.V., Gobbini, M.I., Fury, M., Ishai, A., Schouten, J.L., Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*, 2425-2430.

Haynes, J.-D., Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience, 8*, 686-691.

Haynes, J.-D., Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience, 7,* 523-534.

Haynes, J.-D., Sakai, K., Rees, G. Gilbert, S., Frith, C., Passingham, R.E. (2007). Reading hidden intentions in the human brain. *Current Biology, 17*, 323-328.

Kamitani, Y., Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience, 8,* 679-685.

Krekelberg, B., Boynton, G.M., Van Wezel, R.J.A. (2006). Adaptation: from single cells to BOLD signals. *Trends in Neurosciences, 29*, 250-256.

Kriegeskorte, N., Goebel, R., Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences USA, 103*, 3863-3868.

Kriegeskorte, N., Formisano, E., Sorger, B., Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences USA, 104*, 20600-20605.

Kriegeskorte, N., Bandettini, P. (2007a). Analyzing for information, not activation, to exploit high-resolution fMRI. *Neuroimage, 38*, 649-662.

Kriegeskorte, N., Bandettini, P. (2007b). Combining the tools: Activation- and information-based fMRI analysis. *Neuroimage 38*, 666-668.

Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., Bandettini, P.A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, in press.

Ku, Sp., Gretton, A., Macke, J., Logothetis, N. K. (2008). Comparison of pattern recognition methods in classifying high-resolution BOLD signals obtained at high magnetic field in monkeys. *Magnetic Resonance Imaging, 26*, 1007-1014.

LaConte, S., Strother, S., Cherkassky, V., Anderson, J., Hu, X. (2005). Support vector machines for temporal classification of block design fMRI data. *Neuroimage, 26*, 317-329.

Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature, 412*, 150-157.

Logothetis, N.K. (2008). What we can do and what we cannot do with fMRI. *Nature Reviews, 453*, 869-878.

Mourao-Miranda, J., Bokde, A.L.W., Born, C., Hampel, H., Stetter, M. (2005). Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on fMRI data. *Neuroimage, 28*, 980-995.

Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences, 10*, 424-430.

Pereira, F., Mitchell, T., Botvinick, M. (2008). Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage* (submitted).

Raizada, R.D.S., Tsao, F.M., Liu, H.M., Kuhl, P.K. (). Quantifying the adequacy of neural representations for a cross-language phonetic discrimination task: prediction of individual differences. *Cerebral Cortex*, submitted.

Sawamura H, Orban GA, Vogels R (2006) Selectivity of neuronal adaptation does not match response selectivity: a single-cell study of the fMRI adaptation paradigm. *Neuron*, *49*, 307-318.

Strother, S.C., Anderson, J., Hansen, L.K., Kjems, U., Kustra, R., Sidtis, J., Frutiger, S., Muley, S., LaConte, S., Rottenberg, D. (2002). The quantitative evaluation of functional neuroimaging experiments: The NPAIRS data analysis framework. *Neuroimage, 15*, 747-771.

Tolias AS, Keliris GA, Smirknakis SM, Logothetis NK (2005) Neurons in macaque area V4 acquire directional tuning after adaptation to motion stimuli. Nat Neurosci 8: 591-593.

Worsley, K.J., Friston, K.J. (1995). Analysis of fMRI time-series revisited – again. *Neuroimage 2*, 173-181.

Worsley, K.J., Poline, J.-B., Friston, K.J., Evans, A.C. (1997). Characterizing the response of PET and fMRI data using multivariate linear models. *Neuroimage 6*, 305-319.

Pattern-information analysis toolboxes

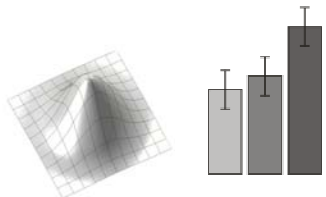
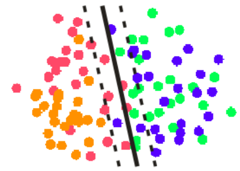
AFNI 3dsvm plug-in
(http://www.cpu.bcm.edu/laconte/3dsvm.html)

Princeton MVPA toolbox
(http://www.csbmb.princeton.edu/mvpa/)

PyMVPA toolbox
(http://pkg-exppsy.alioth.debian.org/pymvpa/)

LIBSVM toolbox
(http://www.csie.ntu.edu.tw/~cjlin/libsvm)

Table 1 Overview of activation-based and pattern-information analysis

	Activation-based analysis 	**Pattern-information analysis** 
Goal of the analysis	Investigating the *involvement* of regions in a specific mental activity	Investigating the *representational content* of regions
Experimental contrast	Difference between mental activity *including* component of interest and mental activity *excluding* component of interest	Difference between representation of object 1 and representation of object 2
Analytical comparison	Compare spatial-average activation across conditions	Compare patterns of activity across conditions
Spatial resolution	Benefits of high-resolution imaging will be limited if data are smoothed	Fine-grained spatial information provided by high-resolution imaging is used effectively
Statistical methods	• Spatial smoothing • Combine single-voxel signals by smoothing and averaging activity within ROI • Univariate analysis • Group analysis in common stereotactic space	• No spatial smoothing • Combine single-voxel signals by computing multivariate statistics • Multivariate analysis (typically linear discriminant analysis) • Single-subject analysis in native subject space

Images in this table were adapted with permission from Kriegeskorte and Bandettini (2007b).

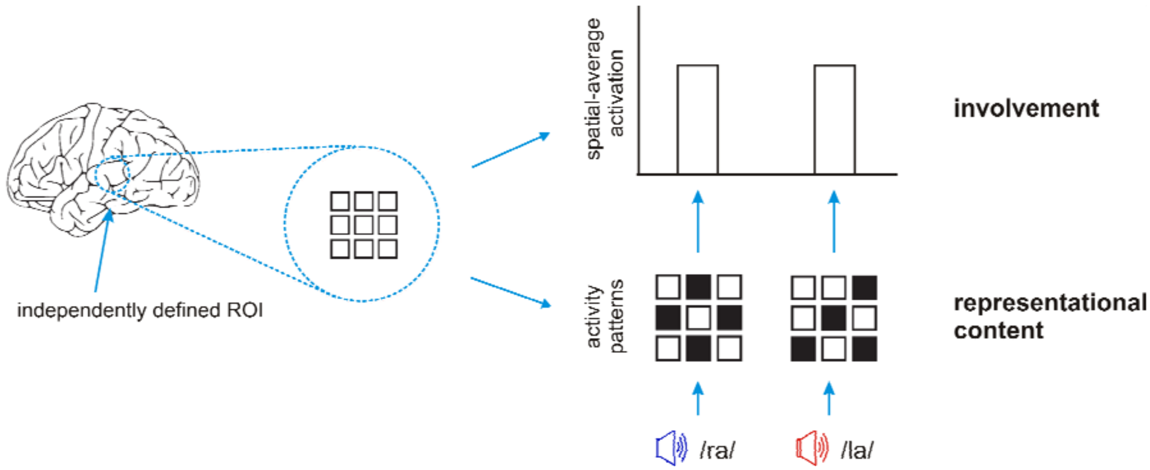


Figure 1 Activation indicates "involvement", pattern-information indicates "representational content". A specific region of interest (ROI) can show the same spatial-average activation resulting from different patterns encoding different representational content. This figure shows a hypothetical ROI consisting of 9 voxels. The ROI's multivoxel pattern of activity is different for /ra/ than /la/ speech sounds, but these different patterns result in the same spatial-average activation. This difference will go undetected by classical activation-based analysis. Pattern-information analysis can be used to show that an ROI's multivoxel activity pattern differs significantly across conditions, i.e. that the region contains information about the experimental conditions. Differences in multivoxel patterns across conditions can be interpreted as reflecting differences in underlying neuronal population activity. This figure has been adapted with permission from Raizada et al. (2008).

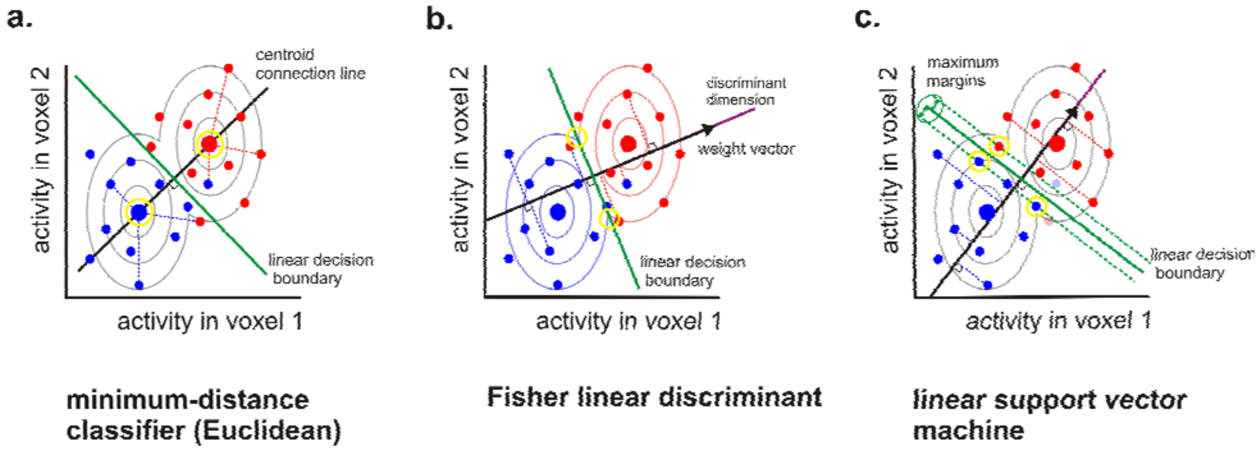


Figure 2 Linear classification methods all define a linear decision boundary, but the boundary is placed slightly differently. This is shown for a given set of hypothetical activity patterns. The blue dots represent activity patterns for one experimental condition (e.g. the speech sound /ra/), the red dots represent activity patterns for a second condition (e.g. the speech sound /la/). For simplicity, the displayed activity patterns are based on activity of only two voxels. Nevertheless, the classification methods generalize to higher-dimensional voxel spaces. The ellipses in the background of each panel are iso-probability-density contours describing the bivariate normal distribution of the activity patterns for each condition. The yellow circles indicate the geometrical features that define the linear decision boundary (green) for each classifier. **(a)** Minimum-distance classifier. This classifier first determines the centroids of the two multivariate distributions (large dots). Each activity pattern is then classified to the centroid that it is closest to in multivariate space (using Euclidean distance or 1- correlation across voxels (Haxby et al., 2001) as a multivariate distance measure, as shown by the dotted lines). This implies a linear decision boundary (i.e. a hyperplane) orthogonal to the centroid connection line, equally dividing the distance between the two centroids. **(b)** Fisher linear discriminant analysis (FLDA). Response patterns are projected onto a linear discriminant dimension by weighting each voxel's activity in order to maximize the ratio of between-condition and within-condition variance. The voxel weights define a weight vector that points in the direction of the linear discriminant dimension. The patterns (i.e. the data points) are orthogonally projected onto the discriminant dimension and a threshold is

used for classification. This implies a linear decision boundary (i.e. a hyperplane) orthogonal to the linear discriminant dimension. **(c)** Linear support vector machine (SVM). Same description as FLDA, except for the way the voxel weights are computed. The voxel weights computed by linear SVM are set to yield a linear decision boundary that maximizes the margin (i.e. the distance of the nearest data point to the decision boundary). To make this intuitive, we can imagine starting with a decision boundary that perfectly classifies the training set, then widening the margin equally on both sides while adjusting the angle and position of the decision boundary, until the margin cannot be widened anymore without including one of the training data points. The response patterns closest to the decision boundary (points in yellow circles) then define the margins and the decision boundary halfway in-between the margins. These points are therefore called "support vectors". In order to handle overlapping distributions, SVM algorithms are typically set to allow for a few misclassifications on the training set (see the two transparent points in our hypothetical example).

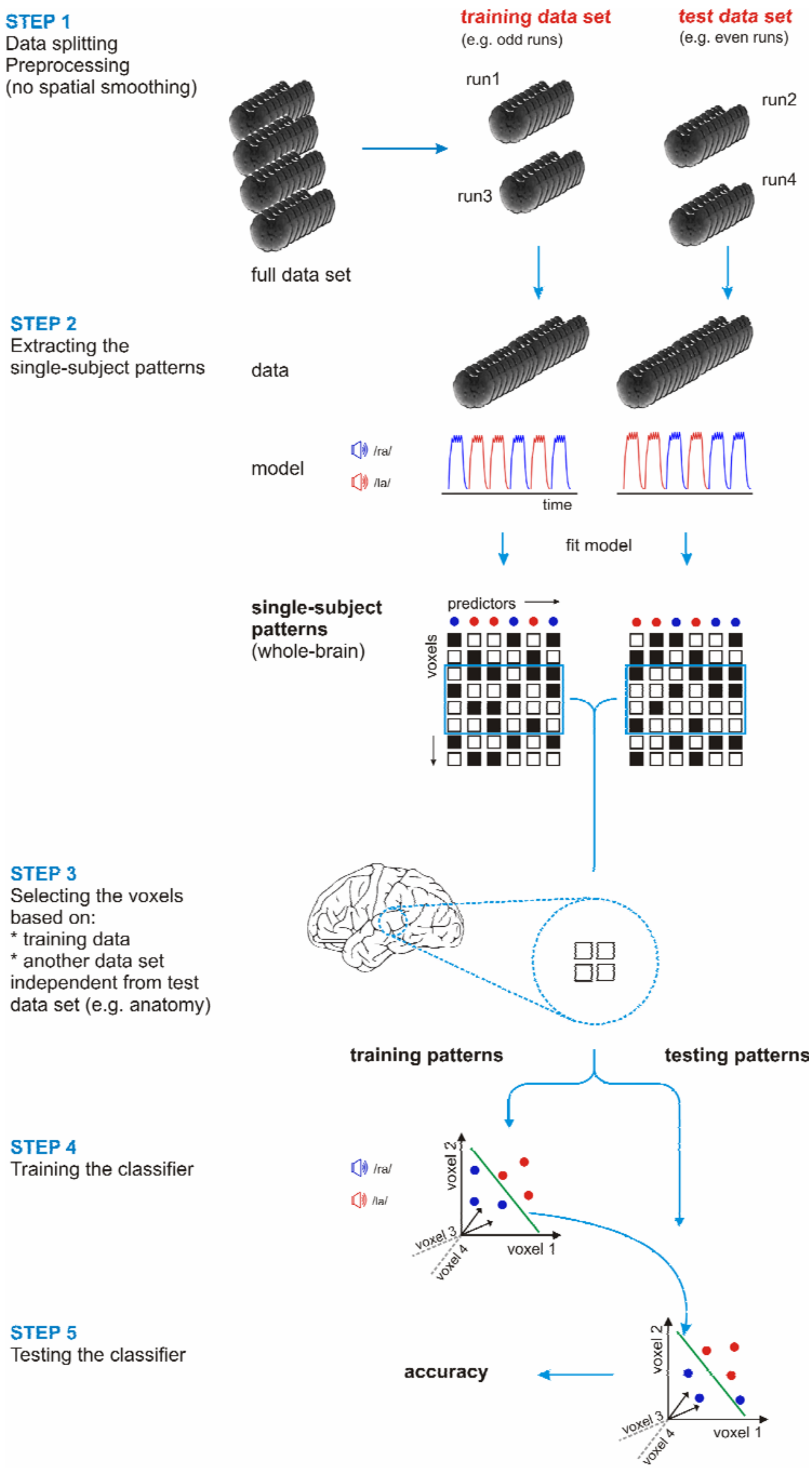

STEP 1
Data splitting
Preprocessing
(no spatial smoothing)
training data set
(e.g. odd runs)
test data set
(e.g. even runs)
run1
run2
run3
run4
full data set
STEP 2
Extracting the
single-subject patterns
data
model
/ra/
/la/
time
fit model
single-subject
patterns
(whole-brain)
predictors
voxels
STEP 3
Selecting the voxels
based on:
* training data
* another data set
independent from test
data set (e.g. anatomy)
training patterns
testing patterns
STEP 4
Training the classifier
/ra/
/la/
voxel 2
voxel 1
voxel 3
voxel 4
STEP 5
Testing the classifier
accuracy
voxel 2
voxel 1
voxel 3
voxel 4

Figure 3 Pattern-information analysis: step by step. Schematic illustration of the five steps of pattern-information analysis as described in the text. First, data are split into a training and a test data set and preprocessed separately. Then, single-subject patterns of activity are extracted from the data using univariate analysis (GLM) at each voxel. This results in whole-brain activity patterns consisting of beta-estimates. Black boxes indicate activated voxels; white boxes indicate non-activated voxels. Note that activity levels are continuous in analysis and only stated as binary here for simplicity. There will be as many patterns as there are predictors (conditions) in the model. Pattern-extraction is done separately for the training and test data set. The third step consists of selecting voxels for pattern-information analysis. This can be done based on anatomy, function or both. For simplicity, the shown example region consists of four voxels only. Voxel selection should be based on the training data set or another data set that is independent from the test data set in order to prevent biased testing results. This also applies to the fourth analysis step: voxel weighting (or computation of the condition centroids for a minimum-distance classifier) should be performed on the training data set to prevent biased testing results. Voxels are weighted in order to maximize discriminability of the patterns belonging to the two conditions. The voxel weights computed in step 4 can then be tested on the test data set in step 5. If the weights capture true differences between the two conditions, good performance (classification accuracy) on the training data set will generalize to the test data set. Performance significantly better than chance indicates that the ROI contains information about the experimental conditions, i.e. the representational content of the region differs across conditions. The image for step 3 has been adapted with permission from Raizada et al. (2008).