Chapter 6

Neuroinformatics

Ricardo Vigário, Jaakko Särelä, Sergey Borisov, Astrid Pietilä, Jan-Hendrik Schleimer, Jarkko Ylipaavalniemi, Alexander Ilin, Samuel Kaski, Eerika Savia, Erkki Oja

6.1 Introduction

Neuroinformatics has been defined as *the combination of neuroscience and information sciences to develop and apply advanced tools and approaches essential for a major advancement in understanding the structure and function of the brain.* Aside from the development of new tools, often the fields of application include the analysis and modelling of neuronal behaviour, as well as the efficient handling and mining of scientific databases. With the current configuration, the group aims at proposing algorithmic and methodological solutions for the analysis of elements and networks of functional brain activity, studying several kinds of communication mechanisms. These are to be applied in the understanding of ongoing brain activity, as well as responses to natural stimulation.

From a methodological viewpoint, the neuroinformatics group has been involved in studying certain properties of ICA, such as its reliability and applicability to the analysis of electrophysiological brain data (namely electroencephalograms, EEGs and magnetoencephalograms, MEG), as well as to functional magnetic resonance images (fMRI). Within the study of ICA reliability, subspace effects have been made evident, and their potential in functional network interpretability suggested (see Sec. 6.2).

Two explorative studies into functional brain networks have been started. One made explicit use of complex stimulation in fMRI, resulting in the detection of several networks of functional activity with clear interpretability (see Sec. 6.3). Another targeted phase synchrony, which is expected to play a central role in the communication within the central nervous system, as well as between this and the peripheral nervous system (see Sec. 6.4).

Several other topics have been researched in the field of biomedical signal processing, which are not thoroughly reported here. In particular, the denoising source separation framework (DSS) introduced earlier in the laboratory of computer and information science, has been used in the study of phonocardiographic signals, as well as in the investigation of different possible origins for high- and low-amplitude alpha-activity in EEG. We have as well studied measurement fMRI artefacts using a reliable ICA approach with a standard spherical phantom. All of these topics are collected in Sec. 6.5. The application of our methods to tissue segmentation in magnetic resonance imaging (MRI), to the detection of brain lesions will appear in the report of next biennial period.

Research reported in this section has been carried out in collaboration with experts in neuroscience and cardiology.

References

[1] Schleimer, J.-H., and R. Vigário. Clustering limit cycle oscillators by spectral analysis of the synchronisation matrix with an additional phase sensitive rotation. In *Proc. 17th Int. Conf. on Artificial Neural Networks (ICANN'2007)*, Porto, Portugal, pp. 944–953, 2007.

[2] Schleimer, J.-H., and R. Vigário. Order in Complex Systems of Nonlinear Oscillators: Phase Locked Subspaces. In *Proc. of 15th European Symposium on Artificial Neural Networks (ESANN'07)*, Bruges, Belgium, pp. 13–18, 2007.

[3] Ylipaavalniemi, J., E. Savia, R. Vigário, and S. Kaski. Functional elements and networks in fMRI. In *Proc. of 15th European Symposium on Artificial Neural Networks (ESANN'07)*, Bruges, Belgium, pp. 561–566, 2007.

[4] Ylipaavalniemi, J., and R. Vigário. Subspaces of Spatially Varying Independent Components in fMRI. In *Proc. 7th Int. Conf. on Independent Component Analysis and Blind Signal Separation (ICA'2007)*, London, England, pp. 665–672, 2007.

[5] Schleimer, J.-H., and R. Vigário. Reference-based extraction of phase synchronous components. In *Proc. 16th Int. Conf. on Artificial Neural Networks (ICANN'2006)*, Athens, Greece, pp. 230–238, 2006.

[6] Borisov, S., A. Ilin, R. Vigário, and E. Oja. Comparison of BSS methods for the detection of α-activity components in EEG. In *Proc. 6th Int. Conf. on Independent Component Analysis and Blind Signal Separation (ICA'2006)*, Charleston, South Carolina, USA, pp. 430–437, 2006.

[7] Pesonen, M., M. Laine, R. Vigário, and C. Krause. Brain oscillatory EEG responses reflect auditory memory functions. *abstr. 13th World Congress of Psychophysiology, International Organization of Psychophysiology (IOP'2006)*, Istanbul, Turkey, 2006.

[8] Pietilä, A., M. El-Segaier, R. Vigário, and E. Pesonen. Blind Source Separation of Cardiac Murmurs from Heart Recordings. In *Proc. 6th Int. Conf. on Independent Component Analysis and Blind Signal Separation (ICA'2006)*, Charleston, South Carolina, USA, pp. 470–477, 2006.

[9] Ylipaavalniemi, J., S. Mattila, A. Tarkiainen, and R. Vigário. Brains and Phantoms: An ICA Study of fMRI. In *Proc. 6th Int. Conf. on Independent Component Analysis and Blind Signal Separation (ICA'2006)*, Charleston, South Carolina, USA, pp. 503–510, 2006.

6.2 Reliable ICA and subspaces

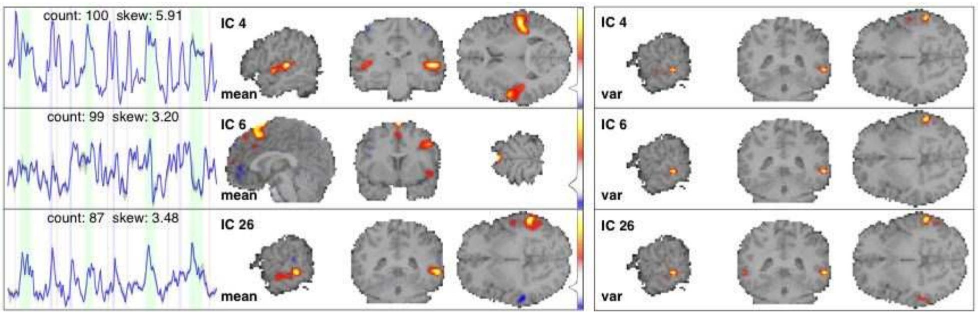
In contrast to traditional hypothesis-driven methods, independent component analysis (ICA) is commonly used in functional magnetic resonance imaging (fMRI) studies to identify, in a blind manner, spatially independent elements of functional brain activity. Particularly, in studies using multi-modal stimuli or natural environments, where the brain responses are poorly predictable, and their individual elements may not be directly relatable to the given stimuli.

In earlier reported work, we have analysed the consistency of ICA estimates, by focusing on the spatial variability of the components. The optimization landscape of ICA is defined by structure of the data, noise, as well as the objective function used. The landscape can form elongated or branched valleys, containing many strong points, instead of singular local optima. Multiple runs of the ICA algorithm with varying random initial conditions and re-sampling allows to characterise the optimisation landscape and the robustness of the estimates.

Previous studies have analyzed the consistency of independent components, and suggested that some components can have a characteristic variability. The goal was to provide additional insight into the components, that is not possible to attain with single run approaches. Complex valleys can also be considered as separate subspaces, where statistical independence is not necessarily the best objective for decomposition.

We have now proposed a novel method for reliably identifying subspaces of functionally related independent components. We also proposed two approaches to further refine the decomposition into functionally meaningful components. One refinement method uses clustering, to distinguish the internal structure of the subspace. Another method is based on finding the coordinate system inside the subspace that maximally correlates with the temporal dynamics of the stimulation. The directions are found with canonical correlation analysis (CCA).

A study of subspaces was conducted on multi-modal fMRI recordings, including several forms of auditory stimulation. In the following figure, we can see a set of components, strongly related to auditory stimulation. Each component is consistent, appearing in all or most of the 100 runs. The mixing variability is also minimal. However, the spatial variance reveals a coincident location of variability, shared by all components. The variability links the components into a three dimensional subspace, even though ICA has consistently identified directions within the subspace.



(a) (b)

Figure 6.1: Tested approaches for alpha extraction from simulated data.

Within the study, we postulate that, based on spatial variance, components can be roughly divided into 3 classes: individual and consistent components, with distributed variance due to noise; consistent members of a subspace, with focal variance coincident with the variance of the other members; and inconsistent subspaces, with variances coincident with their own mean. Such subspaces can provide information on networks of related activity in a purely data-driven manner. Criteria to disambiguate each subspace will then be of crucial relevance.

6.3 Towards brain correlates of natural stimuli

Natural stimuli are increasingly used in functional magnetic resonance imaging (fMRI) studies to imitate real-life situations. Consequently, challenges are created for novel analysis methods, including new machine learning tools. With natural stimuli it is no longer feasible to assume single features of the experimental design alone to account for the brain activity. Instead, relevant combinations of rich-enough stimulus features could explain the more complex activation patterns.

We proposed a novel two-step approach, where independent component analysis is first used to identify spatially independent brain processes, which we refer to as functional patterns. As the second step, temporal dependencies between stimuli and functional patterns are detected using dependency exploration methods. Our proposed framework looks for combinations of stimulus features and the corresponding combinations of functional patterns.

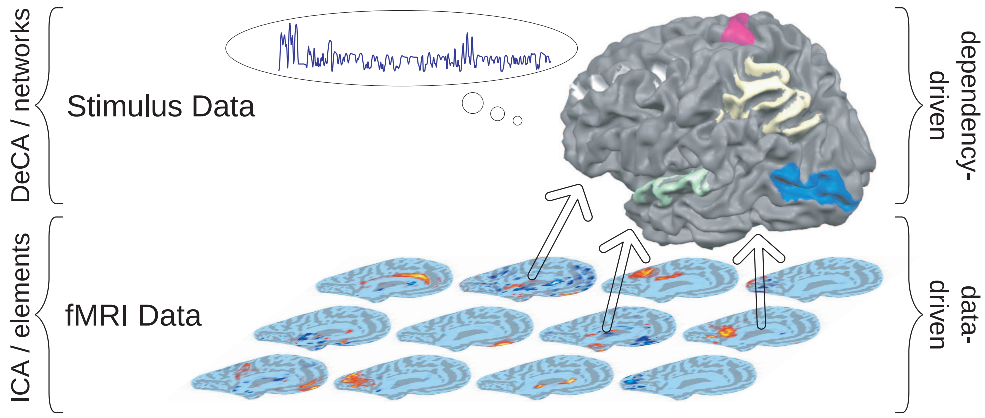
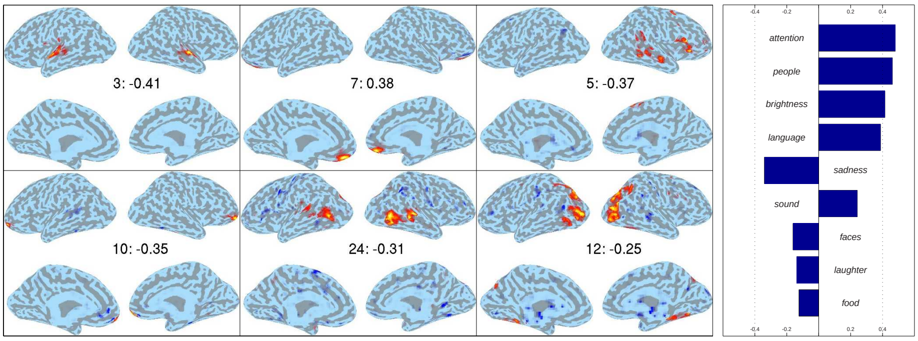


Figure 6.2: The proposed framework: elements of functional brain activity emerge from the data via ICA. Functional networks are revealed by DeCA, based on covariation between the elements and task goals, encoded as features.

This two-step approach was tested on fMRI recordings of brain responses to natural stimuli, consisting of a movie with 20 minutes duration. Rather subjective features were extracted from the movie, including labels such as *"attention"*, *"sadness"*, *"people"* or *"laughter"*.

As an illustrative example, we can look into a network comprising brain areas that individually correspond to, *e.g.*, auditory (IC3), visual (IC12), and multi-modal integration (IC24). This suggests that the functional role of the whole network is related to combining information from many sensory inputs. Indeed, the four highest scoring features of the dependent component are *attention*, *people*, *brightness* and *language*.

The found networks seem plausible, considering the limited and very subjective nature of the available stimulus features. Some elements were a part of several networks, with different functional contribution to each networks common task. A more controlled study has been carried out since, to verify the results and to further develop the approach. These will be reported in the next biennial report.



(a) (b)

Figure 6.3: (a) The 6 ICs corresponding to the highest loadings in the first dependent component. The loading value is shown in the middle of each square. (b) Respective loadings of the stimulus features.

6.4 Synchrony exploration

Interest in phase synchronisation phenomena has a long history, when studying the interaction of complex, natural or artificial, dynamic systems. Although not completely adopted, synchronisation was attributed a role in the interplay between different parts of the central nervous system as well as across central and peripheral nervous systems. Such phenomena can be quantified by the phase locking factor (PLF), which requires knowledge of the instantaneous phase of an observed signal.

Linear sources separation methods treat scenarios in which measurements do not represent direct observations of the dynamics, but rather superpositions of underlying latent processes. Such a mixing process can cause spuriously high PLF's between the measurements, and camouflage the phase locking to a provided reference signal. Essentially, synchronisation is either caused by a common input or by interactions between neurons.

Reference-based approach

The PLF between a linear projection of the data and a reference can be maximised as an optimisation criterion, revealing the most synchronous source component present in the data, with its corresponding amplitude. This is possible despite the amplitude distributions being Gaussian, or the signals being statistically dependent, common assumptions in blind sources separation techniques without a-priori knowledge, e.g. in form of a reference signal.

We first addressed this reference-based problem, and proposed a new algorithm capable of extracting sources phase-locked with a reference. In the following illustration one can see the efficiency of such a method. The sources, depicted on the right frame, were chosen so that neither high-order statistics methods, e.g., FastICA, nor methods based on temporal decorrelation, e.g., SOBI would perform the desired source estimation.

We tested this approach on MEG recordings, with a 306-sensor Vectorview neuromagnetometer, together with left and right hand EMG's. The subject was instructed to simultaneously keep isometric contraction in left and right hand muscles, using a special squeezing device. We then used the right hand EMG as a reference for the phase exploration into the MEG recordings. The results achieved agreed with early studies performed in the same recordings.

We also addressed the “internal neuronal synchronisation” problem, where no clear reference is available, proposing to cluster a population of oscillators into segregated sub-

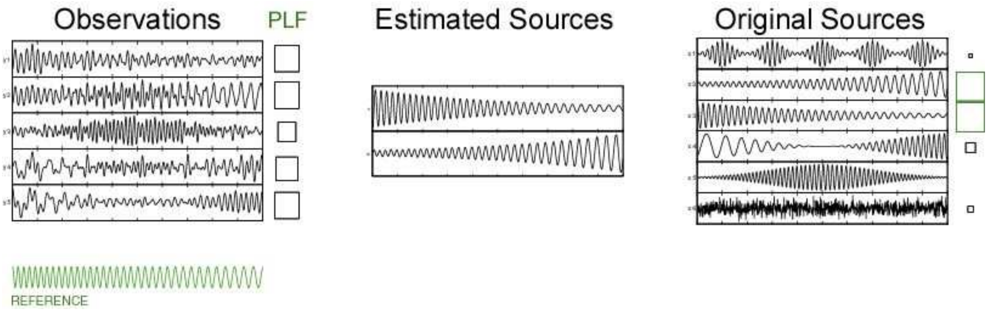
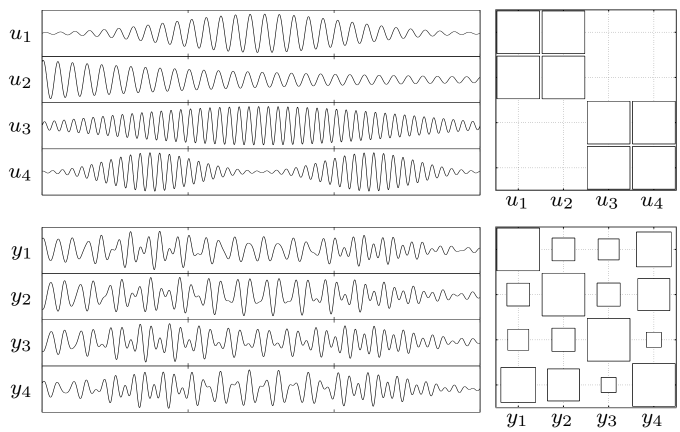


Figure 6.4: Six signals, of which only two are locked in phase.



(a) Oscillators with distinct amplitudes, grouped in synchronous subspaces. PLFs are depicted as a Hinton diagram. ($P = 0.08$)

(b) Mixtures with an overall spuriously phase locking. ($P = 3.66$)

populations, exhibiting high internal interactions. Approaches to solve this problem have often assumed different frequencies for the various sub-populations, usually neglecting phase information. These assumptions pose a restriction to the analysis of the dynamic world of natural systems, where communication can be unrelated to the natural frequency of the constituent oscillators.

Our solution makes explicit use of phase information, extracted from known models of physical interactions. The approach relies on a post-rotation of the eigenvectors of the synchronisation matrix.

With simulations, we show the effect of the post-rotation, in the estimation of underlying sources for which their frequency has been drawn from a global distribution. In neurobiological terms, this means that the neuron's system parameters, which determine its natural frequency, do not depend on the synaptic connections it has formed. In such formulation, frequency can not be used to identify the sources anymore. Phase clustering is then crucial for the task.

We have also proposed a method to reveal phase-locked subspaces, based on a concept of order in complex systems of nonlinear oscillators. Any order parameter quantifying the degree of organisation in a physical system can be studied in connection to source extraction algorithms. Independent component analysis, by minimising the mutual information of the sources, falls into that line of thought, since it can be interpreted as searching components with low complexity. Complexity pursuit, a modification minimising Kolmogorov complexity, is a further example.

Using such concept of order, we designed an algorithm capable of revealing subspaces of oscillators such that: oscillators of the same subspace are completely phase locked; whereas between subspaces there is no Phase locking. The following illustrations exemplify the algorithm's performance. Estimated sources coincide with the true ones.

6.5 Overview of other topics

Within the duration of this biennial report, several research topics have been addressed with a more prospective view. Some will be the subject of more thorough development in further reports, whereas other will stay as simple case studies. This section reviews four of those.

Extraction of alpha activity

Following earlier work on the characterisation of low- and high-amplitude alpha brain activity, we tested a two-step blind source separation approach for the extraction of such rhythms from ongoing EEG. The method comprised a denoising stage, performed by DSS, followed by either a high-order statistical independent component analysis source estimation, FastICA, or a temporal decorrelation one, TDSEP.

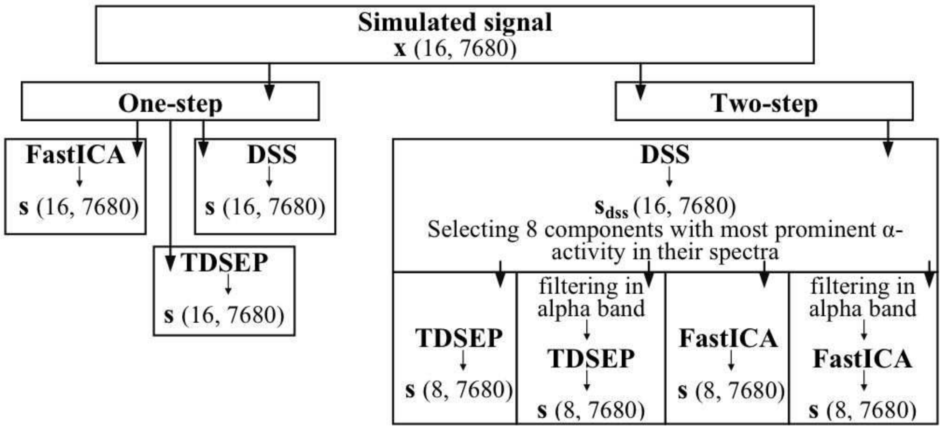


Figure 6.5: Tested approaches for alpha extraction from simulated data.

The main findings are that denoising has, as expected, a positive effect in rendering the subsequent source separation algorithms more efficient. In addition, we observed that high-order statistics ICA was more adequate in such separation than TDSEP, in spite of the latter being particularly suited for dealing with temporally structured sources. A targeting α-filter, placed between the denoising and the TDSEP modules, resulted in good estimates, rendering the combination rather efficient. Such filtering seems to not affect significantly FastICA.

Artefact removal in ERD/ERS study

Still within the rhythmic activity of the brain, we participated in a study of brain oscillatory EEG responses to auditory memory functions. The analysis concentrated on event related de-synchronisation and synchronisation (ERD and ERS, respectively), in the theta and alpha frequency ranges for ERS and also in beta for ERD.

The outcomes of that study suggested that theta frequency ERS responses may be associated with working memory functions, whereas alpha ERD/ERS responses robustly dissociate between auditory memory encoding and recognition.

ICA showed to be crucial in denoising the raw ongoing EEG, prior to wavelet processing. Several subjects displayed considerable artefacts that rendered most of the event-related responses virtually unusable.

BSS of cardiac murmurs

A significant percentage of young children present cardiac murmurs. However, only one percent of them are caused by a congenital heart defect; others are physiological. An automated system for an initial recording and analysis of the cardiac sounds could enable the primary care physicians to make the initial diagnosis and thus decrease the workload of the specialised health care system. independent component analysis source estimation, FastICA, or a temporal decorrelation one, TDSEP.

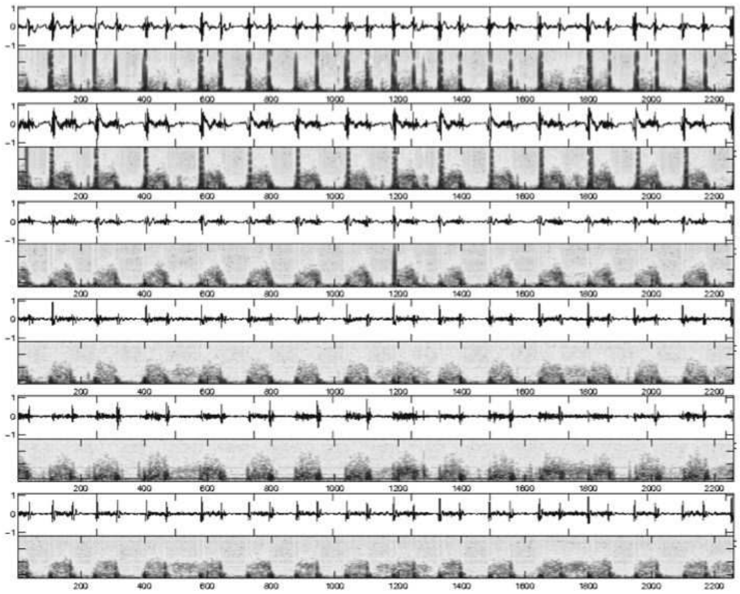


Figure 6.6: Six-channel PCG recordings from a patient, together with their spectrograms. The S1 and S2 are clearly visible in the first spectrogram as periodic pairs of vertical bars covering all the frequencies. Murmurs are visible in the systole, between the S1 and S2, present on all six recordings.

The first step to such analysis is the identification of the different components of the cardiac cycle, with particular emphasis to the separation of the murmurs. We have proposed a new methodological framework to address this issue, combining ICA and DSS. independent component analysis source estimation, FastICA, or a temporal decorrelation one, TDSEP.

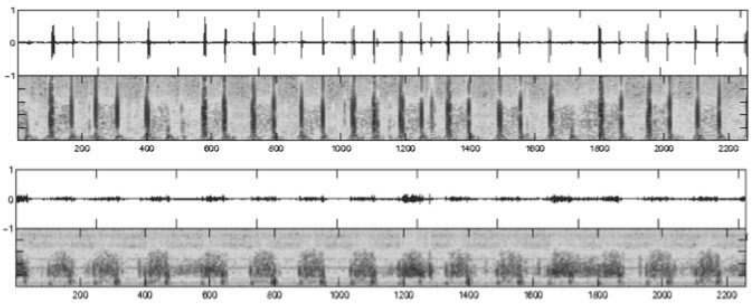


Figure 6.7: Heart sounds S1 and S2, clearly isolated from all other signals. In the second frame are uncontaminated murmurs.

Using such approach, we have been able to isolate rather efficiently the murmurs, as well as heart sounds S1 and S2 and artefacts such as voices recorded during the measurements.

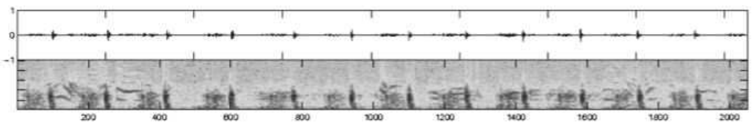


Figure 6.8: Speech artefacts present in PCG recordings. Formant structures are clearly visible.

With the aforementioned results, the collaboration with the Lund University Hospital, Sweden, has been strengthened, and further research outcomes are expected in the next reports.

Phantom study in fMRI

Phantom measurements are routinely used for verifying and calibrating the quality of MRI machinery. However, data-driven analysis of phantom fMRI data has been largely overlooked, possibly due to the lack of a method for assessing the reliability of the solutions. We have now used a reliable ICA approach to such analysis, and revealed evidence for possible misinterpretations in ICA studies with real subjects.

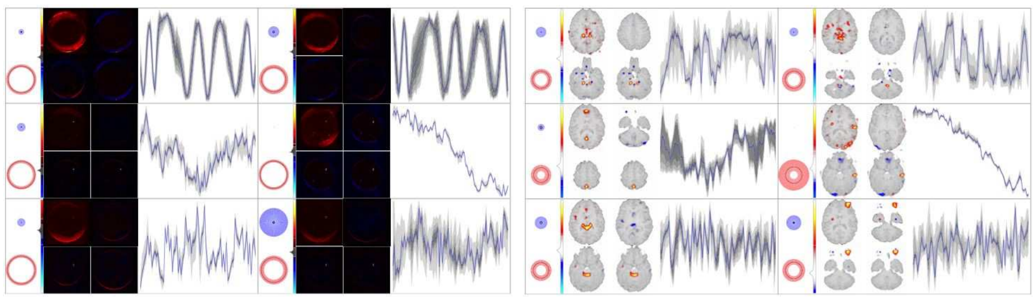


Figure 6.9: Reliable independent components, extracted from fMRI of a spherical phantom (a), and a real subject (b). Corresponding temporal 'activation' patterns are shown on the right of each estimate.

Several independent components found on a real subject presented a temporal structure that follows clearly that of phantoms. We speculate that methods other than ICA can also suffer from a similar kind of misinterpretation. We therefore suggest the need for a better understanding of the artificial, scanner- or environmentally-induced artefacts, prior to the automatic analysis of any fMRI recording. A comparison between real brain ICA and phantom-based decompositions may help in the validation of the estimated components.