Chapter 6

Neuroinformatics

Ricardo Vigário, Miguel Almeida, Nicolau Gonçalves, Nima Reyhani, Jarkko Ylipaavalniemi, Elina Karp, Jayaprakash Rajasekharan, Jyri Soppela, Janne Nikkilä, Eerika Savia, Samuel Kaski, 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, the fields of application include often the analysis and modelling of neuronal behaviour, as well as the efficient handling and mining of scientific databases. The group aims at proposing algorithmic and methodological solutions for the analysis of elements and networks of functional brain activity, addressing several forms of communication mechanisms. Motivation and application areas include the understanding of ongoing brain activity and the neuronal responses to complex natural stimulation.

From a methodological viewpoint, the neuroinformatics group has studied properties of source separation methods, such as their reliability and extensions to subspaces. We have also assessed the suitability of such methods to the analysis of electrophysiological recordings (EEG and MEG), and functional magnetic resonance images (fMRI). We proposed also methods for the study of phase synchrony within the central nervous system, and between this and the peripheral nervous system. We have also developed methods for the analysis of neural responses of natural stimulation, based on a novel approach of capturing statistical dependencies between brain activity and the stimulus itself.

In addition to the analysis of fMRI recordings from natural stimulation, we have been also involved in the analysis of single trial event-related MEG data. Albeit its significantly higher temporal resolution, the signal-to-noise ratios are typically very poor, and averaging across hundreds of stimuli is often required. We currently search as well for efficient tissue segmentation of structural MRI.

Although not a natural research topic in neuroinformatics, we have also been involved in the study of phonocardiographic signals, to detect the sources and severities of cardiac murmurs in infants. The signal processing is a challenging one, and a successful application should have a great societal impact.

In addition to these ongoing but stable research topics, we have made a pilot in document mining. The goal is to extract, in a semi-automatic manner, functional information from neuroscience journals, hence reducing the dependence on curator intervention.

Research reported in this section has been carried out in collaboration with experts in neuroscience and cardiology. In the following, we highlight some of the results attained in the reported years.

References

[1] Ylipaavalniemi, J., E. Savia, R. Hari, R. Vigário, and S. Kaski. Towards True Brain Correlates: Dependencies Between Stimuli and Independent fMRI Sources. *Neuroimage 48*, 176–185, 2009.

[2] Almeida, M. and R. Vigário. Source separation of phase-locked subspaces. In *Proc. 8th Int. Conf. on Independent Component Analysis and Blind Signal Separation (ICA'2009)*, Paraty, Brazil, pp. 203–210, 2009.

[3] Karp, E., L. Parkkonen, and R. Vigário. Denoising single trial event related magnetoencephalographic recordings. In *Proc. 8th Int. Conf. on Independent Component Analysis and Blind Signal Separation (ICA'2009)*, Paraty, Brazil, pp. 203–210, 2009.

[4] Ylipaavalniemi, J., and J. Soppela. Arabica: Robust ICA in a Pipeline. In *Proc. 8th Int. Conf. on Independent Component Analysis and Blind Signal Separation (ICA'2009)*, Paraty, Brazil, pp. 379–386, 2009.

[5] Gonçalves, N., Nikkilä, J., and R. Vigário. Partial clustering for tissue segmentation in MRI. To appear in *Proc. 15th Int. Conf. on Neuro- Information Processing (ICONIP'2008)*, Auckland, New Zealand, 2008.

[6] Gonçalves, N., and R. Vigário. Semi-automatic approach for brain tissue segmentation using MRI. In *1st INCF Congress of Neuroinformatics: Databasing and Modeling the Brain (Neuroinformatics 2008)*, Stockholm, Sweden, pp. 106, 2008.

[7] Vigário, R. and E. Oja. BSS and ICA in Neuroinformatics: From current practices to open challenges. *IEEE Rev. Biomed. Eng. 1*, 50–61, 2008.

[8] Ylipaavalniemi, J., and R. Vigário. Analyzing Consistency of Independent Components — An fMRI Illustration. *Neuroimage 39*, 169–180, 2008.

[9] Ylipaavalniemi, J., and R. Vigário. Matching complex activation patterns with features of natural stimuli. In *1st INCF Congress of Neuroinformatics: Databasing and Modeling the Brain (Neuroinformatics 2008)*, Stockholm, Sweden, pp. 71, 2008.

6.2 Complex neural responses to complex stimuli

Natural stimuli are increasingly used in 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 have 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 either canonical correlation analysis (a journal article in NeuroImage) or its distribution-free variant Nonparametric Dependent Component Analysis (DeCA, a conference article in ICASSP'09). Our method looks for combinations of stimulus features and the corresponding combinations of functional patterns.

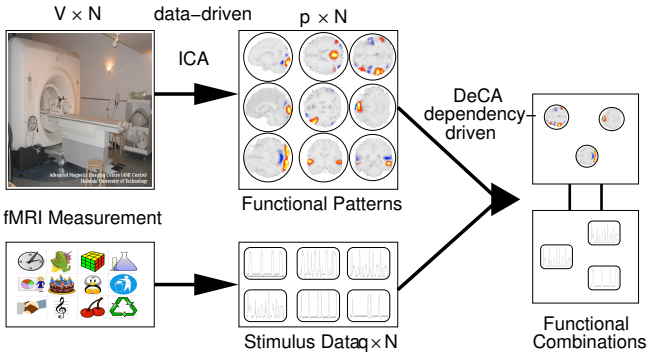

Figure 6.1: Sketch of the framework. (V = # of voxels, N = # of measurement time points, p = # of reliable ICA components, q = # of stimulus time courses).

This two-step approach has been used to analyze measurements from a fMRI study during multi-modal stimulation, in collaboration with Riitta Hari and co-workers. It seems promising to analyze data using natural stimulation.

6.3 Phase synchrony

Interest in phase synchronization phenomena has a long history, when studying the interaction of complex, natural or artificial, dynamic systems. Although not completely adopted, synchronization 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, which requires knowledge of the instantaneous phase of an observed signal.

During the reported years, we extended the set of algorithmic tools for the identification of phase synchronous phenomena. Our earlier methods dealt with the extraction of sources phase-locked to a reference signal and the clustering of a population of oscillators into synchronous sub-populations. We proposed now a method for the extraction of phase-locked subspaces, following an approach akin to the underlying considerations in independent component analysis.

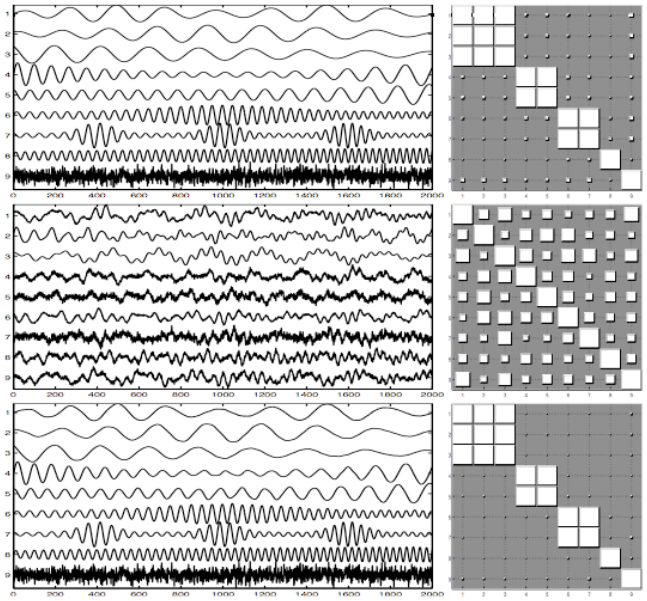

Figure 6.2: Results of one run of IPA: original sources (top left) and PLFs between them (top right); mixed signals (middle left) and PLFs between them (middle right); ex- tracted sources (bottom left) and PLFs between them (bottom right). Results obtained for ? = 0.2, after manually compensating for permutation, scaling and sign of the extracted sources. The Amari Performance Index was 0.06.

6.4 Single trial event related studies

Functional brain mapping is often performed by analysing neuronal responses evoked by external stimulation. Assuming constant brain responses to repeated identical stimuli, averaging across trials is usually applied to improve typically poor signal-to-noise ratios. However, since wave shape and latency vary from trial to trial, information is lost when averaging.

To mitigate this problem, and enable the identification of inter-trial signal variations, we proposed a method to correct the trial-to-trial jitters, in a visually evoked MEG study. The approach was based on a template-based denoising source separation framework. The results were physiologically plausible and presented a clear improvement compared to the classical averaging. We are currently searching as well for a competing approach to estimate, in addition to the jitters, variations in amplitude for each trial's evoked signal.

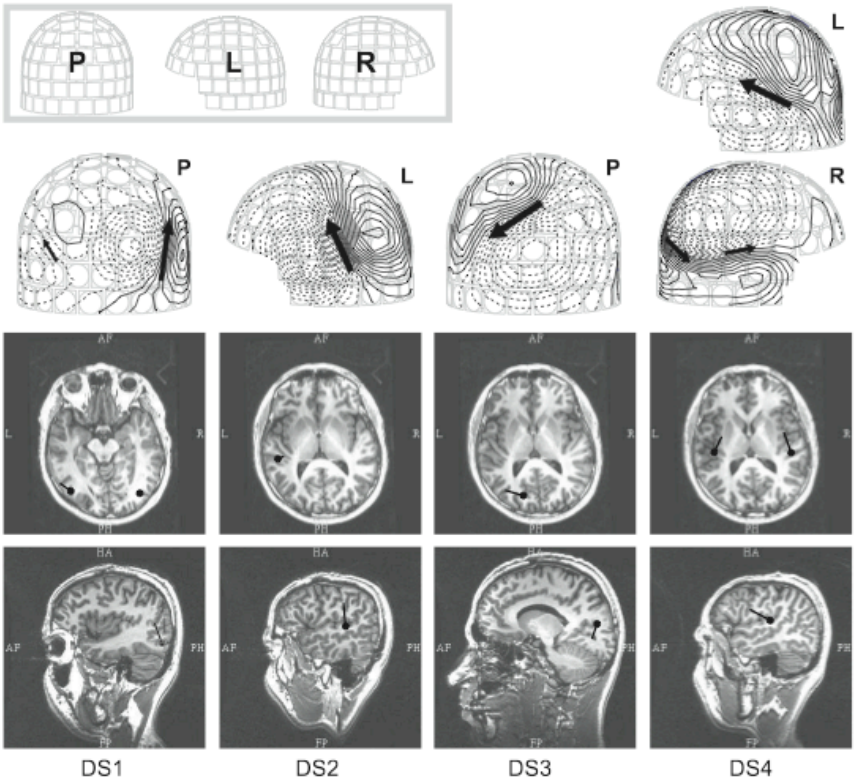

Figure 6.3: Fieldpatterns associated with the denoised sources (DS) and the locations of corresponding equivalent current dipoles. P stands for posterior, R forright and L for left view.

6.5 Tissue segmentation in MRI

MRI is a widely used clinical imaging technique, able to distinguish between soft tissues with an exceptional spatial resolution. In most clinical applications, several settings are used, depending on the targeted tissue, leading to a multi-spectral image set. Automatic follow-up of pathologies in the brain should make full use of the multi-spectra, and be capable of clear segmentation of each tissue.

Our approach is based on the discriminative clustering (DC) algorithm. Since DC is a supervised method, it requires labelled training data, which we produce through a boosting use of the self-organising maps. We achieved absolute classification results in par with the best methods currently in use. As a result of the clustering approach, partial volume information for each image voxel is available, and degenerative pathologies can be better assessed.

Figure 6.4: Classification result for each class of the brainweb data set. CSF, white and grey matter from left to right. The classification is shown overlaying a T2 sequence. Voxels in black correspond to the voxels that have most of their membership in that class.

Tissue	CSF	White	Gray	MS Lesion	Misclassification rate
normal-set	99.70%	97.05%	96.42%	–	3.11%
lesion-set	96.61%	98.90%	98.26%	34.38%	2.61%

Figure 6.5: Numerical results of the DC classification. The percentages shown correspond to the amount of voxels correctly classified.

6.6 Document mining

There is an ever increase in the number of scientific publications in many areas in general, and in neurosciences in particular. Hundreds of articles are published each month. When comparing the results one obtains with a given experimental setup and existing information in literature, one may validate, integrate or confront different opinions and theories. The compilation of such a vast amount of information is not only crucial, but currently also rather human-intensive.

With that in mind, we have conducted a pilot study on document mining of journal publications reporting results on fMRI experiments. We have focused on the image content of the articles. The rather positive preliminary results suggest that a more systematic use of the methodology, and its improvement may help as well reducing the amount of curating work required for the construction of functional databases.

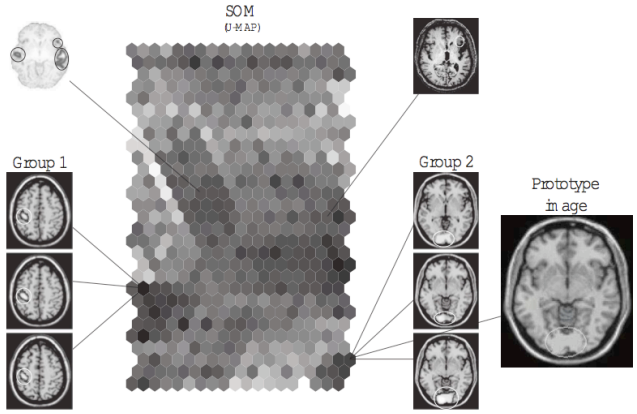

Figure 6.6: Self Organizing Map $\tilde{\text{N}}$ U-matrix trained with 16 dimensional feature vectors, from a set of 100 images extracted from 11 journal papers. Two distinct cluster regions are observed at the lower left and right sides of the map. The prototype image, depicted in the upper left corner fits the expected cluster.