Computational Intelligence & Neuroscience

EEG/MEG Signal Processing

Guest Editors: Andrzej Cichocki and Saied Sanei

EEG/MEG Signal Processing

EEG/MEG Signal Processing

Guest Editors: Andrzej Cichocki and Saied Sanei

Contents

The P300 as a Marker of Waning Attention and Error Propensity, Avijit Datta, Rhodri Cusack, Kari Hawkins, Joost Heutink, Chris Rorden, Ian H. Robertson, and Tom Manly
Volume 2007, Article ID 93968, 9 pages

Hindawi Publishing Corporation
Computational Intelligence and Neuroscience
Volume 2007, Article ID 97026, 2 pages
doi:10.1155/2007/97026

Editorial
EEG/MEG Signal Processing

A. Cichocki[1] and S. Sanei[2]

[1] *Laboratory for Advanced Brain Signal Processing, RIKEN Brain Science Institute, Saitama 351-0198, Japan*
[2] *Centre of Digital Signal Processing, Cardiff University, Cardiff CF24 3AA, Wales, UK*

Correspondence should be addressed to S. Sanei, saneis@cf.ac.uk

Received 13 November 2007; Accepted 13 November 2007

Since its invention by the Hans Berger of the electroencephalography (EEG) in 1929, it was a strong scientific curiosity in analysis of human brain activity. In fact, the electroencephalography (EEG) and magnetoencephalography (MEG) have developed into one of the most important and widely used quantitative diagnostic tools in analysis of brain signals and patterns. EEG and MEG potentially contain a rich source of information related to functional, physiological, and pathological status of the brain. In particularly, they are essential for the identification of mental disorders and brain rhythms extremely useful for the diagnosis and monitoring of brain activity and offer not only the functional but also pathological, physiological, and metabolic changes within the brain and perhaps other parts in the body.

Recording and analysis of the EEG and MEG now involve a considerable amount of signal processing for S/N enhancement, feature detection, source localization, automated classification, compression, hidden information extraction, and dynamic modeling. These involve a variety of innovative signal processing methods, including adaptive techniques, time-frequency and time-scale procedures, artificial neural networks and fuzzy logic, higher-order statistics and nonlinear schemes, fractals, hierarchical trees, Bayesian approaches, and parametric modeling. This special issue contributes to the current status of EEG and MEG signal processing and analysis, with particular regard to recent innovations. It reports some promising achievements by academic and commercial research institutions and individuals, and provides an insight into future developments within this exciting and challenging area of functional brain imaging.

Noninvasive functional brain imaging has become an important tool used by neurophysiologists, cognitive psychologists, cognitive scientists, and other researchers interested in brain function. In the last five decades the technology of noninvasive functional imaging has flowered, and researchers today can choose from EEG, MEG, PET, SPECT, MRI, NIRS, and fMRI. Each method has its own strengths and weaknesses. Development of signal processing tools mitigates the problems and alleviates some of the weaknesses.

This issue includes the following contributions which cover a wide range of signal processing techniques for analysis, understanding, and recognition of EEG/MEG information.

The first paper, "Canonical source reconstruction for MEG" by J. Mattout et al., describes a new, simple but efficient solution to the problem of reconstructing electromagnetic sources into a canonical or standard anatomical space. Electromagnetic lead fields are computed using the warped mesh, in conjunction with a spherical head-model (which does not rely on individual anatomy). The ensuing forward model is inverted using an empirical Bayesian scheme that was described previously in several publications. This enables the pooling of data from multiple subjects and the reporting of results in stereotactic coordinates. Furthermore, it allows the graceful fusion of fMRI and MEG data within the same anatomical framework.

The second paper, "A subspace method for dynamical estimation of evoked potentials" by S. Georgiadis et al., describes method for single-channel trial-to-trial EP characteristics estimation. Prior information about phase-locked properties of the EPs is assessed by means of estimated signal subspace and eigenvalue decomposition. Then for those situations that dynamic fluctuations from stimulus-to-stimulus could be expected, prior information can be exploited by means of state-space modeling and recursive Bayesian mean square estimation methods (Kalman filtering and smoothing). The authors demonstrate that a few dominant eigenvectors of the data correlation matrix are able to model

trend-like changes of some component of the EPs, and that Kalman smoother algorithm is to be preferred in terms of better tracking capabilities and mean square error reduction. They also demonstrate the effect of strong artifacts, particularly eye blinks, on the quality of the signal subspace and EP estimates by means of independent component analysis (ICA) applied as a prepossessing step to the multichannel measurements.

The third paper, "Inferring functional brain states using temporal evolution of regularized classifiers," by A. Zhdanov et al., proposes a framework for functional brain state inference problem that utilizes the temporal information present in the brain signals. This application suggests that the relation between the regularization parameters and the temporal profile of the classifier helps improving the classifier accuracy.

In the fourth paper, "Removing ocular movement artefacts by a joint smoothened subspace estimator," by R. Robert Phlypo et al., a joint smoothened subspace estimator calculates the low- and high-order statistic information subject to the constraint that the resulting estimated ocular movement artifact source is smooth in time domain. This results in combination of blind source separation with different order statistics. The results have been compared to those of well-known blind source separation methods and have shown the capability of the system in mitigating the ocular artefacts automatically.

The fifth contribution, "A framework to support automated classification and labeling of brain electromagnetic patterns," by G. A. Frishkoff et al., focuses on patterns in averaged EEG (ERP) data to define high-level rules and concepts for ERP components and to design an automated data processing system that implements these rules. This is with a broader objective of designing an oncology-based system to support cross laboratory, cross paradigm, and cross modal integration of brain functional data.

The next paper, "Statistical modeling and analysis of laser-evoked potentials of electrocorticogram recordings from awake humans," by Z. Chen et al., provides a comprehensive analysis of electrocorticogram recorded using invasive laser stimulation. Both averaging and single trial laser-evoked potentials (LEP) have been considered. Then the LEPs have been extracted from both types of trials, and the variations in power, amplitude, and latency have been studied using probabilistic modeling, factor analysis, independent component analysis, wavelet domain, and quantitative and qualitative analyses.

The seventh paper"A Novel constrained topographic independent component analysis for separation of epileptic seizure signals," by Min Jing and Saeid Sanei, addresses a constrained source separation method which exploits the correlation among the nearby brain sources as well as characteristics of the seizure signals in space and frequency domains to highlight the sources of interest. In this method the space-frequency characteristics of the data is utilized as the constraint term in the update equation of the topographic ICA system. The results clearly show that the synchronously generated seizure sources are grouped together.

The next paper, "Clustering approach to long term spatio-temporal interactions in epileptic electroencephalograph," by A. Hegde et al., attempts to identify the spatio-temporal interactions of an epileptic brain using an existing nonlinear dependency measure based on a clustering approach. The mutual interactions have been analyzed using an index measure based on a self-organizing map (SOM) network. The results report a long-term structural connectivity related to various seizure states. In addition, the authors have aimed at developing engineering tools to determine spatiotemporal groupings in a multivariate epileptic brain.

The ninth paper, "Automatic seizure detection based on time-frequency analysis and artificial neural networks," By A. T. Tzallas et al., uses an artificial neural network system for detection of epileptic seizures from a set of features estimated from time-frequency domain EEG data.

Next paper, "Canonical decomposition of ictal scalp EEG and accurate source localisation: principles and simulation study," by M. De Vos et al., uses a dipole-based method for localization of epileptic seizure sources. In this method a canonical decomposition procedure extracts the seizure source by a three-way model assumption.

The eleventh paper, "The implicit function as squashing time model a novel parallel nonlinear EEG analysis technique distinguishing mild cognitive impairment and Alzheimer's disease subjects with high degree of accuracy," by M. Buscema et al., introduces an ANN-based method in which the MCI and AD can be classified based on the spatial information content of the restino EEGs. In this procedure the ANNs do not use EEGs as the input; rather, the inputs for the classification are the weights of the connections within the ANN to generate the recorded EEG data. The introduced TWIST system selects the best features.

The last paper, "The P300 as a marker of waning attention and error propensity," By Avijit Kumar Datta, Rhodri Cusack, Kari Hawkins, Joost Heutink, Christopher Rorden, Ian Robertson, and Tom Manly, studies and examines the variation of P300 ERP with respect to the error in responding to the stimuli. During the course of this research it has been found that errors are associated with significant reduction in the amplitude of preceding P300, and the fluctuations in P300 amplitude across the task formed a reliable associate of individual error propensity, supporting its use as a marker of our sustained control over action.

ACKNOWLEDGMENTS

The guest editors are extremely grateful to all the reviewers who took time and consideration to assess the submitted manuscripts. Their diligence has contributed greatly to ensuring that final papers have conformed to the high standards expected in this publication. The guest editors of this special issue are much indebted to their authors and reviewers, who put a tremendous amount of effort and dedication to make this issue a reality.

A. Cichocki
S. Sanei

Hindawi Publishing Corporation
Computational Intelligence and Neuroscience
Volume 2007, Article ID 67613, 10 pages
doi:10.1155/2007/67613

Research Article

Canonical Source Reconstruction for MEG

Jérémie Mattout,[1] Richard N. Henson,[2] and Karl J. Friston[3]

[1] *INSERM U821, Dynamique Cérébrale et Cognition, Lyon, France*
[2] *MRC Cognition and Brain Sciences Unit, Cambridge CB2 7EF, UK*
[3] *The Wellcome Trust Centre for Neuroimaging, University College, London WC1N 3BG, UK*

Correspondence should be addressed to Jérémie Mattout, jeremiemattout@yahoo.fr

Received 11 January 2007; Revised 24 April 2007; Accepted 27 May 2007

Recommended by Saied Sanei

We describe a simple and efficient solution to the problem of reconstructing electromagnetic sources into a canonical or standard anatomical space. Its simplicity rests upon incorporating subject-specific anatomy into the forward model in a way that eschews the need for cortical surface extraction. The forward model starts with a canonical cortical mesh, defined in a standard stereotactic space. The mesh is warped, in a nonlinear fashion, to match the subject's anatomy. This warping is the inverse of the transformation derived from spatial normalization of the subject's structural MRI image, using fully automated procedures that have been established for other imaging modalities. Electromagnetic lead fields are computed using the warped mesh, in conjunction with a spherical head model (which does not rely on individual anatomy). The ensuing forward model is inverted using an empirical Bayesian scheme that we have described previously in several publications. Critically, because anatomical information enters the forward model, there is no need to spatially normalize the reconstructed source activity. In other words, each source, comprising the mesh, has a predetermined and unique anatomical attribution within standard stereotactic space. This enables the pooling of data from multiple subjects and the reporting of results in stereotactic coordinates. Furthermore, it allows the graceful fusion of fMRI and MEG data within the same anatomical framework.

1. INTRODUCTION

Source reconstruction in neuroimaging, particularly PET and fMRI, is usually into a standard anatomical space (e.g., that defined by the Atlas of [1]). Reconstruction into a canonical space facilitates the formal or informal meta-analysis of findings in imaging neuroscience and provides a useful framework within which to define structure-function relationships. In PET and fMRI the construction of spatially normalized images comprises two distinct steps. First, the raw data are reconstructed into images of source activity within the subject's own anatomical space. Second, these data are then spatially normalized into a standard space using a template matching approach (e.g., [2]). For EEG and MEG, however, source reconstruction and spatial or anatomical normalization cannot be separated because the reconstruction depends upon the spatial configuration of sources.

The central idea, upon which this work is based, is to include anatomical variability in a forward model that links MEG responses to canonical sources. Specifically, the anatomical differences between a particular subject and a canonical subject (who conforms to the standard space) enter the forward model. Note that these differences are expressed in both cortical anatomy and in the geometrical and physical properties of other tissues (e.g., skull and scalp), through which electromagnetic fields propagate to the sensors. However, we restrict ourselves here to the effect of inter-subject variability in cortical anatomy, given that for MEG, spherical conductor models, which need not incorporate subject specific information about the head, generally provide a sufficiently good approximation compared with more realistic head models such as those using boundary element methods (BEM) see [3, 4]. In contradistinction, the inverse solution is highly sensitive to the source location and orientation, when defined by the cortical anatomy [5]. The nice thing about the approach used here is that spatial normalization becomes an implicit part of the inverse solution. In this paper, we describe how this can be implemented using fully automated procedures that are already in routine use and are freely available as academic software (see Software note).

The basic idea is to formulate a forward or generative model of how a specific subject's MEG data were caused and then invert this model using standard Bayesian techniques. We start with a canonical subject whose anatomy conforms to a predefined space; the MNI-space based upon the Talairach and Tournoux system [1]. This is the same space as used by the SPM software and, more generally, by the neuroimaging community when reporting fMRI and PET results. A canonical mesh is defined within this space, coding the position and orientation of dipolar sources. Warping the mesh to match the subject's anatomy creates a subject specific model. After warping, subject specific forward fields (i.e., a gain matrix) are computed using standard electromagnetic forward modelling procedures. In this paper, we use a single-sphere head model, fit to the template scalp surface. The resulting forward model has two components an anatomical component that displaces and reorientates the dipoles into subject specific anatomy and an electromagnetic component that projects the source activity to measurement space (i.e., channels). Reconstruction of the canonical sources corresponds to the inversion of this forward model, given some data. The conditional estimates of source activity can then be treated within a canonical space. In other words, the source activity is associated with the original mesh (before warping).

There are several advantages of the approach described in this paper. The primary advantage is that it allows for anatomically informed source reconstruction into a standard space that facilitates inter-subject pooling and standardized reporting of results. The second main advantage is that it does not entail the accurate extraction of a subject specific cortical surface. This means that the spatial constraints can be based upon any anatomical information, irrespective of whether its quality would support cortical surface extraction or not. Another advantage is that the estimation can proceed even in the absence of a subject's MRI. In this instance, the reconstruction assumes that the subject's anatomy was, in fact, canonical. A final advantage, which will be pursued in a subsequent paper, is that conditional uncertainty about the subject's anatomy can be handled gracefully during Bayesian inversion. It is worth noting that the two key methodologies, namely, estimating the mapping from canonical to subject specific anatomical space and Bayesian inversion of MEG forward models, are fully established and in routine use. Furthermore, because they are fully automated and deterministic, there is no need for human intervention, which renders the procedure totally reproducible.

The aim of this paper is first to motivate and to describe the operational details of a fully automated canonical source reconstruction. Second, we demonstrate, quantitatively, the performance of this inverse-normalized canonical mesh approach in comparison with (i) reconstructions based upon the subject's native mesh and (ii) the canonical mesh without any spatial transformation. In a later paper, we will use canonical reconstructions in a hierarchical model of multi-subject responses measured with EEG and MEG. This paper is restricted to the analysis of single subjects.

This paper is organized as follows. In Section 1 we review the theoretical aspects of the procedure. This entails a brief review of our Bayesian approach to conventional forward models. We then consider spatial normalization. Finally, we see how these two components are integrated to enable canonical source reconstruction. The second section is an empirical demonstration of the utility of the approach. Because the estimation scheme is Bayesian, we can use Bayesian model comparison to evaluate different models. This comparison rests on the log evidence or likelihood of the data given a particular model (having integrated out any dependencies on the model's parameters or hyperparameters). Put simply, we can quantify the likelihood of any given data set given one model, relative to another. Here, we compare three sorts of models: first, a baseline model where the electromagnetic model was based upon a canonical mesh without spatial transformation. The second model, used to explain the same data, incorporated anatomically informed spatial transformations of the canonical mesh. We also evaluated a gold-standard model where the cortical mesh was obtained from a cortical surface extraction, using the subject's MRI data. We hoped to show that including the spatial transformation in the reconstructions would yield a greater log evidence than for the baseline model, and that this log evidence was not significantly less than for the gold-standard model based upon the subject's cortical surface.

2. THEORY

2.1. Bayesian source reconstruction

In a series of papers [6, 7] we have described a Bayesian approach to inverting forward models for EEG and MEG. These forward models start with a subject specific cortical dipole mesh or three-dimensional grid, referred to as the subject's source space. This, in conjunction with the position of the sensors, is used to compute a Gain matrix L in the usual way, under quasistatic Maxwellian assumptions. The inversion of the ensuing electromagnetic forward model uses a hierarchical linear observation model and conforms to parametric empirical Bayes (PEB) using restricted maximum likelihood (ReML). The Bayesian aspect accommodates the regularization required for ill-posed inverse problems. The empirical aspect allows us to identify the ReML estimators of hyperparameters λ controlling multiple noise and prior covariance components, $Q_i^{(1)}$ and $Q_i^{(2)}$, respectively. The key advantage of this approach is that it can accommodate multiple priors in a principled and efficient way. Its efficiency stems from the fact that the ReML scheme estimates covariance components in low-dimensional sensor space, as opposed to high-dimensional source space.

The objective function used by this scheme is equivalent to the ReML objective function, which, as shown in [8], is identical to the (negative) variational free energy

$$F = \langle \ln p(y \mid j, \lambda) + p(j \mid \lambda) - \ln q \rangle_q, \tag{1}$$

where y is the data, j are the source activities, and $q(j)$ is their conditional or posterior density. Under Gaussian assumptions, when F is maximized; $q(j) = p(j \mid y, \lambda)$, and the (negative) free energy becomes the log likelihood of the

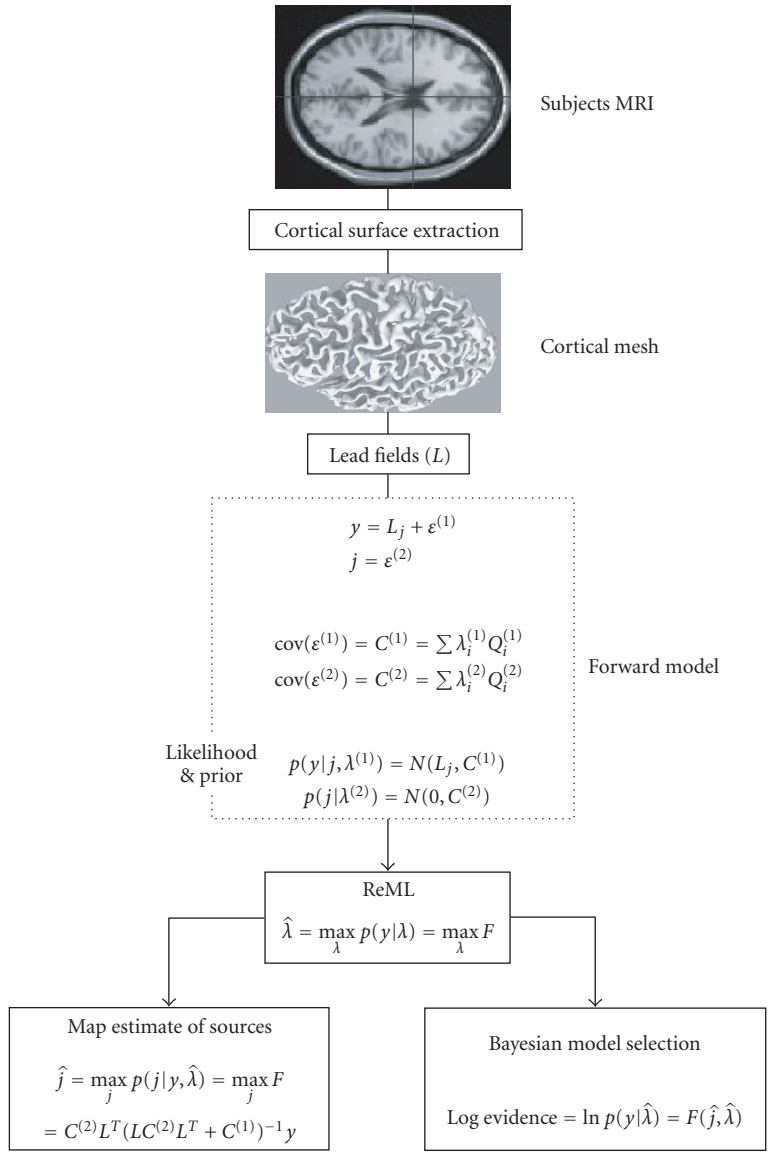


Figure 1: Bayesian inversion scheme.

model or its log evidence $F \to \ln p(y \mid \lambda)$ [9]. We have shown how the log evidence can be used to compare and adjudicate among different models comprising different prior covariance components or different source configurations [7]. We use exactly the same approach below, to compare three different sorts of anatomical source models, each with slightly different configurations of a cortical mesh subtending the lead fields. Figure 1 provides a schematic that summarizes this Bayesian inversion scheme.

2.2. Spatial normalization

Spatial normalization is a term that refers to the warping or mapping of a subject specific image into a standard anatomical space. It is used routinely in fMRI and PET to enable inter-subject pooling. The parameters θ_i that define the transformation $x^{(0)} \to x^{(n)}$ are identified using a Bayesian scheme that incorporates constraints on the smoothness of the transformation [2]. $x_i^{(n)}$ represents the position of the ith control point after n iterations. In brief, the warping is parameterized in terms of spatial basis functions (in SPM, we use a discrete cosine set). These encode the change in position effected by each transform parameter $\partial x/\partial \theta_i$. The coefficients of these basis functions maximize their conditional probability (i.e., maximize the likelihood and prior density). The likelihood is computed using a forward model, which mixes several canonical templates and then warps them to predict the observed image. The mismatch between the warped mixture of templates and the observed image constitutes a prediction error. Under Gaussian assumptions this error gives the likelihood of the observed image, given the mixing and warping parameters. Rough transformations are penalized by appropriate shrinkage priors on the coefficients, formulated

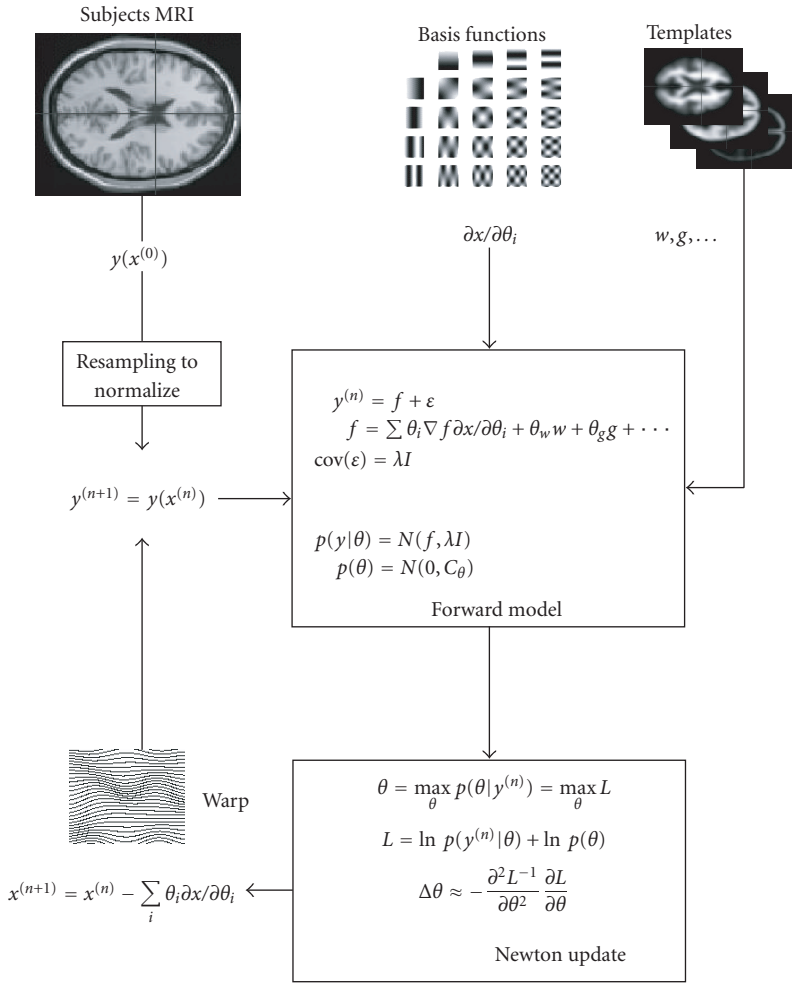


Figure 2: Spatial normalization scheme.

in terms of their covariance. The parameters are computed using a Newton method. The inverse of the template warping is applied to the image and the process iterated until convergence and the image is spatially normalized (see Figure 2 for a schematic).

Once the normalizing transformation has been identified, given some structural image it is usually applied to spatially normalize the subject's functional time series so that analysis can proceed in standard space. A full description of the assumptions and procedures entailed by spatial normalization can be found in a series of papers [10, 11]. Here, we do not use the spatial transformation to normalize reconstructed sources but to spatially *unnormalize* a canonical mesh to inform the forward model about how that subject's electromagnetic signals were generated. This simply involves applying the inverse spatial transformation $x^{(n)} \rightarrow x^{(0)}$ to the locations of the canonical mesh dipoles.

2.3. Canonical source reconstruction

Canonical source reconstruction is identical to our Bayesian source reconstruction (see Figure 1) with the addition of an anatomical component to the forward model. This component is the spatial transformation of a canonical cortical mesh to match the subject's anatomy using the inverse of the spatially normalising transformation (see Figure 3). After transformation, the subject specific mesh is used in the usual way to create an electromagnetic forward model that is inverted as described above. The evidence for this model that comprises both the anatomical and electromagnetic components can then be used to compare different models.

In the next section, we apply the above theory to both synthetic and real MEG data. Our primary goal is to ascertain the relative likelihoods of the different models considered. However, we also take the opportunity to demonstrate the procedure and provide a worked example of its application.

3. MODEL COMPARISON

3.1. Anatomical models

In what follows, we use the following acronyms for the meshes used by the models, which differ only in their anatomical information.

(i) SCS (subject's cortical surface) refers to the mesh obtained from cortical surface extraction, using the subject's structural MRI. This constitutes our gold standard in the sense it makes the least anatomical assumptions. The meshes were obtained using the BrainVISA[1] software [12, 13]. A "fine" mesh was used to generate the synthetic MEG data, while a "coarse" mesh was used to reconstruct the cortical activity, comprising 7204 and 4004 vertices, respectively.

(ii) CCS (canonical cortical surface) refers to a subject specific canonical mesh obtained by applying an inverse spatial transformation to a template mesh in canonical space (the TCS). The transformation is derived by normalising the subject MRI as described in Section 2.2.

(iii) TCS (template cortical surface) refers to the (untransformed) template mesh in canonical space. This model would be used typically when no structural MRI of the subject is available.

To build the TCS, a cortical mesh of a neurotypical male was extracted from his structural MRI, using BrainVISA. This furnished a high-density mesh, with a uniform discrete coverage of the grey/white matter interface. This mesh corresponds to the TCS currently available in the latest release of the SPM software package (see Software note). Here we use the TCS mesh downsampled to 4004 vertices, to match the SCS for source reconstruction. For any given mesh, each vertex location corresponds to a dipole position, whose orientation is fixed perpendicular to the surface. Note that our forward models, based on high-density meshes, could be replaced with low-density meshes with free dipole orientations to compensate for the loss of degrees of freedom implicit in reducing dipole number.

The single subject we considered here was a healthy female volunteer who participated in an MEG study of face perception. We chose a female to deliberately maximize the differences between subject and template anatomy.[2] This enabled us to assess the effectiveness of the warping procedure, under a substantial anatomical distance between SCS and TCS. Furthermore, it induced a greater difference between the warped (CCS) and unwarped (TCS) cortical surfaces, whose influence on the ensuing reconstruction could be observed. Clearly, we anticipate formal and anecdotal replications of the analyses presented in this paper that will allow its conclusions to be generalized to the population of normal subjects.

The two anatomical models for this subject (SCS and CCS) as well as the template mesh (TCS) were compared in the context of simulations and real experiment. In all cases, the sensor locations were registered to source space and the gain matrix was computed using a single sphere-head model [14], fit to the template scalp mesh. The latter was obtained with BrainVISA and used to get the best fitting sphere to be used in the forward computation. As a consequence, the head model was common to each anatomical model and based on the template geometry. We are thus in the position to compare the models, based on their representation of the cortical anatomy only. Bayesian inversion of the ensuing forward model assumed independent channel noise and simple minimum norm priors (i.e., $Q_i^{(1)}$ and $Q_i^{(2)}$ were identity matrices). This corresponds to the classical minimum norm solution, although the relative weight of the likelihood and prior are optimized using ReML as opposed to the conventional L-curve heuristic. ReML has been shown to provide optimal hyperparameter estimates [6, 7], when compared to alternative schemes. Furthermore, this Bayesian inversion enables us to use log-normal hyperpriors on the hyperparameters and enforce a positive contribution of each variance component [9].

Although the log evidence reflects both goodness of fit and model complexity [15], the complexity term for each model was exactly the same. This is because the only difference between the models was in the location of the dipoles encoded by the cortical meshes. In short, the three models compared here match perfectly in terms of complexity and number of free parameters (degrees of freedom).

3.2. Analyses of real data

The MEG dataset came from the female subject, who participated in a multimodal study on face perception (for description of paradigm see [16]). The subject made symmetry judgments on faces and scrambled faces. The MEG data were acquired on a 151-channel CTF Omega system at the Wellcome Trust Laboratory for MEG Studies, Aston University, England. The epochs (80 face trials, collapsing across familiar and unfamiliar faces, and 84 scrambled trials) were baseline corrected from −100 milliseconds to 0 millisecond, averaged over trials and bandpass filtered (between 1 and 30 Hz). The subject's T1-weighted MRI was obtained at a resolution of $1 \times 1 \times 1\,\text{mm}^3$. The subject's head shape was digitized with a 3D Polhemus Isotrak and was used to coregister the MEG sensor locations to anatomical space using a rigid-body (six-parameter) affine transformation. Figure 4 shows the three meshes SCS, CCS, and TCS defining the three models.

3.3. Results for real data

The two types of event-related fields (faces and scrambled) were subtracted to isolate a face-specific effect occurring around 170 milliseconds after stimulation ("M170"). Figure 5 shows the MEG setup and the M170 component elicited. Average responses, over a time window from 150 to 190 milliseconds, were estimated using the three models described above. The resulting log evidences are shown in Table 1. Figure 6 shows the corresponding maps of peak responses (conditional expectations of source activity at the time bin containing the maximum response).

Although slightly different, the estimated responses all show very similar activation patterns, namely in inferior occipital gyri (mostly right) and bilateral orbitofrontal poles.

[1] http://www.brainvisa.info [12].

[2] By maximizing anatomical differences, we refer to cortical size and shape. The gender difference ensures a global difference in size. Moreover, differences in the shape and location of sulci are clearly visible (see Figure 4).

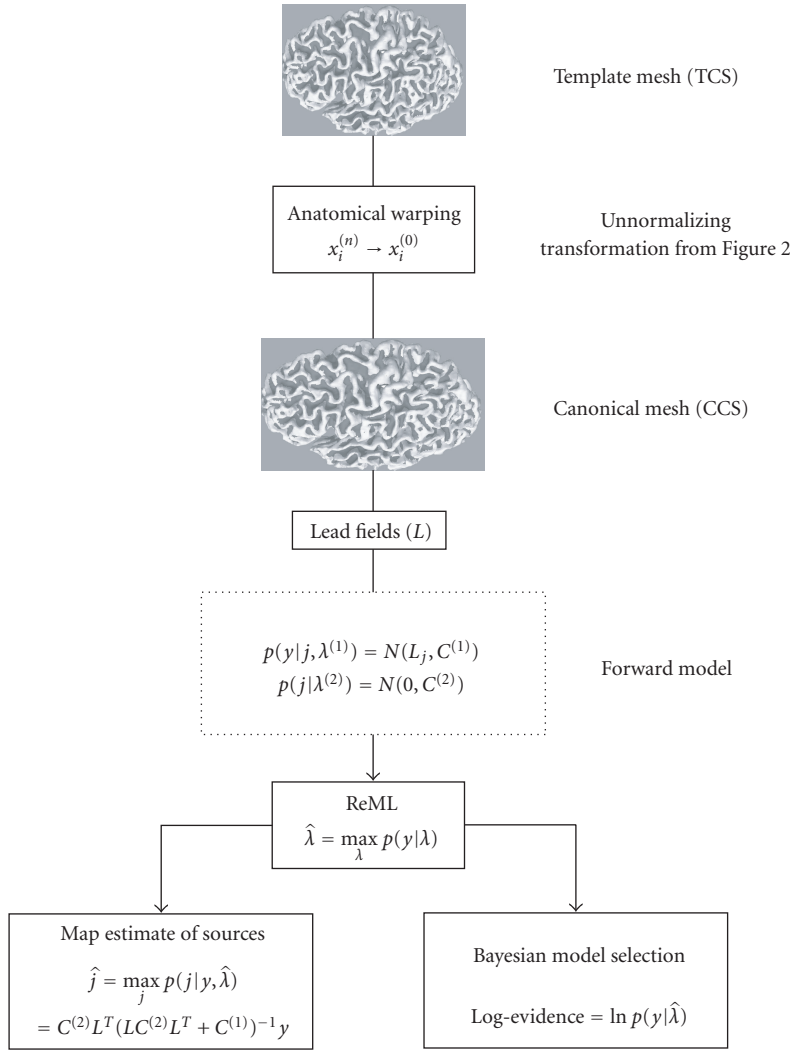
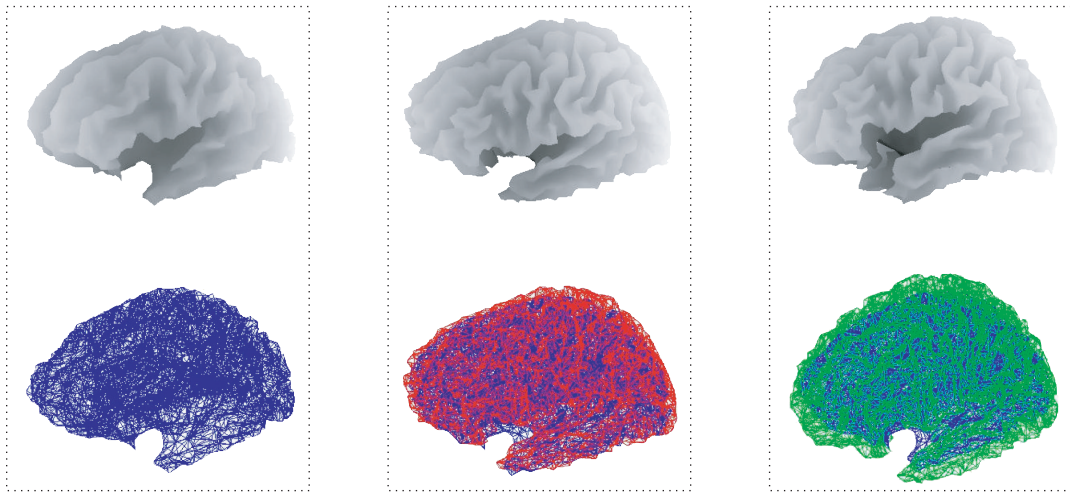
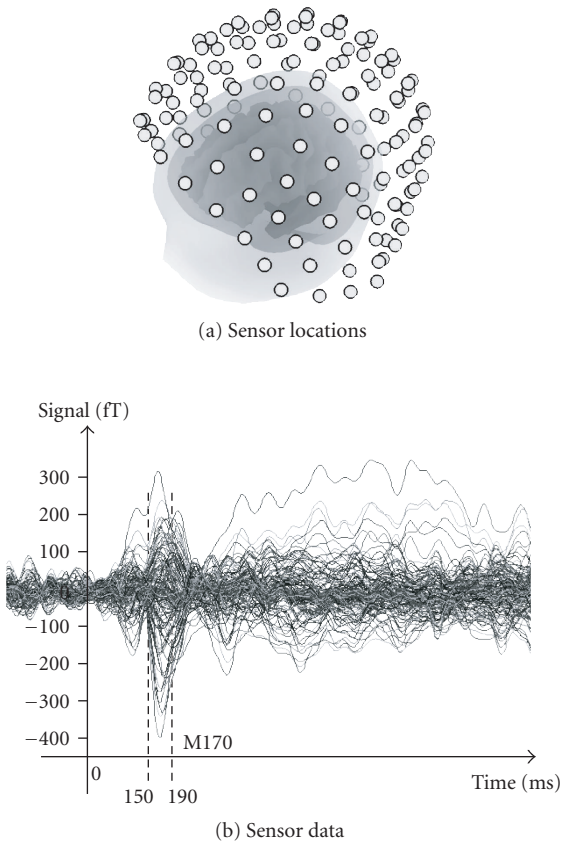


Figure 3: Overview of canonical source reconstruction.



(a) SCS (b) CCS (c) TCS

Figure 4: Surface rendering (upper row) and meshes (lower row) encoding the three cortical models: SCS (a), CCS (b), and TCS (c). CCS (red) and TCS (green) meshes are superimposed on the SCS mesh (blue).



(a) Sensor locations

(b) Sensor data

Figure 5: (a) Sensor locations coregistered with the subject's MRI-derived meshes of the cortical, skull, and scalp surfaces; (b) sensor data for the difference between faces and scrambled event related fields.

Table 1: Log evidences obtained using the real MEG dataset for the three anatomical models.

	SCS	CCS	TCS
Log evidence	14084	14072	14058

It should be noted that our simple minimum-norm solution has favored superficial activity (a well-known property of minimum norm solutions); analyses of the same data using more realistic models (with multiple sensor and source covariance components) place the maximum response more ventrally in both the fusiform and orbitofrontal regions [17]. However, we used the simplest model because this is the most established and our focus here is on *differences* in the reconstructed activity.

The log evidences for the three models are relatively close. One can assess the differences (log ratios or Bayes factors) using the semantics proposed by Kaas and Raftery by analogy with classical inference [15, 18]. In this context, a Bayes factor of twenty means that the data are twenty times more likely to have been generated by one model relative to another (cf., of P-value of .05). A Bayes factor of twenty corresponds to a difference in log evidence of about three, which is the typical threshold one would use to declare that one model was

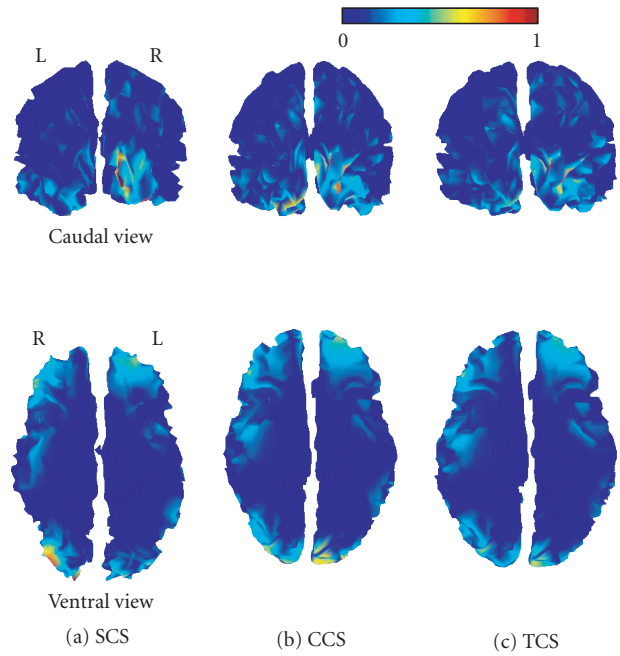


Figure 6: Caudal (upper row) and ventral (lower row) views of the cortical source energy estimated at the peak of the M170 for each of the three anatomical models: SCS (a), CCS (b), and TCS (c). Maps have been normalized to their maximum.

better than another. Given that the differences among the log evidences for our models were about twelve, there is strong evidence that SCS is better than CCS and that CCS is better than TCS. However, one cannot generalize from a single illustrative example. In Section 3.4, we present an extensive simulation study to assess quantitatively and statistically the difference between the three models.

3.4. Synthetic data

MEG data were simulated using the fine SCS mesh and the MEG setup described in Section 3.2 (see Figures 4(a) and 5, resp.). A hundred independent simulations were preformed, each using a single-extended source. For each simulation, the active source comprised a cluster of dipoles. Each cluster was constructed by selecting a random dipole and its nearest mesh neighbors, up to second order (including the nearest neighbors of the nearest neighbors). The cluster size was 7 ± 3 dipoles. Since the dipoles are spread uniformly over the cortical surface, this random dipole selection ensures that all brain regions were represented equally, over simulations. The activity of each source was modelled (over 321 time bins) with two gamma functions, whose parameters were selected randomly, subject to the constraint that the simulated activity reached a peak within time window modelled. Finally, after projection to sensor space, white Gaussian noise was added (SNR $=$ 8 dB) (see Figure 7 for an example of simulated data).

The three models were inverted for each of the hundred simulated datasets. Since we know the true cortical activity,

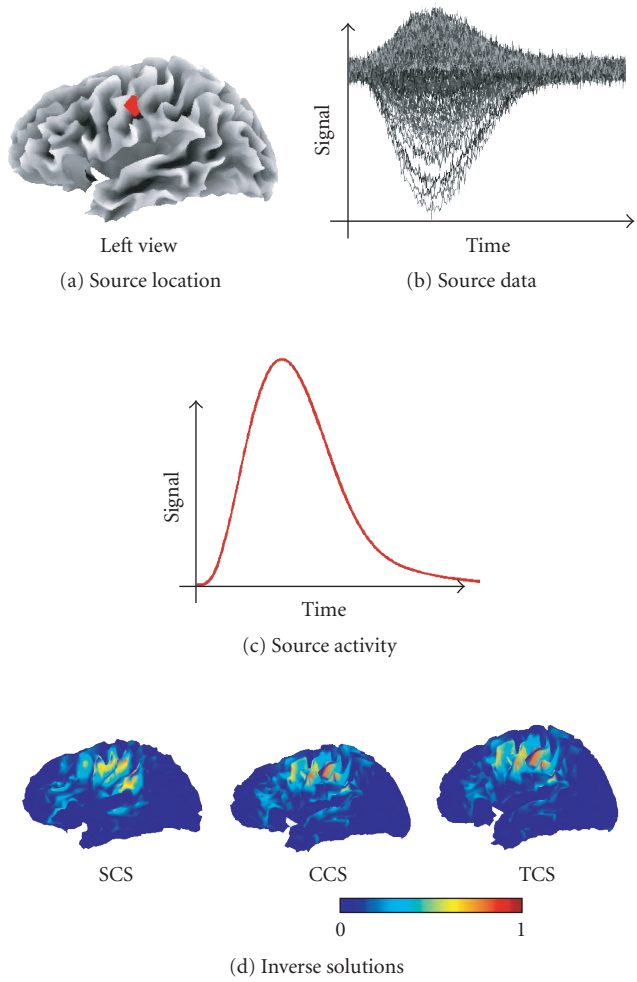


FIGURE 7: Example of a synthetic MEG dataset and its associated inverse solutions. Each map has been normalized to its maximum.

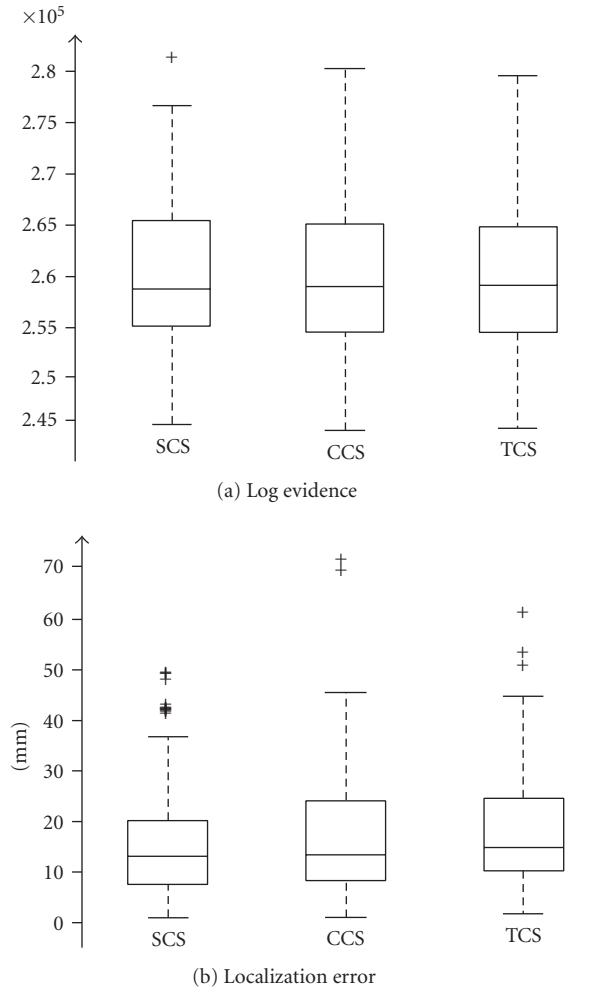


FIGURE 8: Whisker plots of the log evidences and LE values obtained with synthetic MEG data (similar to the example shown in Figure 7) for each of the three anatomical models (SCS, CCS, and TCS).

we supplemented our model comparison using the log evidence with the localization error (LE). LE is the distance between the true source and the dipole exhibiting the maximum estimated energy. This comparative metric complements the log evidence and speaks to the performance of the inversion in terms of the deployment of reconstructed activity, which is an important consideration in multisubject studies. To calculate LE for the SCS-(resp., CCS and TCS) based solution, we used the dipole on the coarse SCS (resp., CCS and TCS) which was closest to the truly activated source on the fine SCS.

3.5. Simulation results

Figure 7 shows an example of synthetic data and the three solutions obtained for each mesh. Figure 8 shows the distributions (whisker plots) of the log evidence and LE over all simulations, for each of the three cortical models. The variance of the log evidences over source configurations is large. It is worth emphasizing here that a given log evidence has no meaning in itself. It only becomes meaningful when compared to the log evidence of another model applied to the same data.

The means of the log evidences, over models, show the same tendency as in the real-data example. Furthermore, the one-way within-dataset ANOVA on the log evidences was significant ($F = 7.81$; $P < .0005^{***}$). Specifically, multiple comparisons with Bonferroni correction show that the only significant differences are between TCS and the two other models; suggesting that there is no demonstrable difference in the performance of the Bayesian inversion of the SCS and CCS models. Similarly, the one-way within-dataset ANOVA on the localization errors proved significant ($F = 15.25$; $P < .0005^{***}$). Again, multiple comparisons with Bonferroni correction show that the only significant pairwise differences are between TCS and the two other models.

To summarize, the localizations based on the reference mesh (SCS) are significantly better than the ones based on the anatomically uninformed template mesh (TCS). Critically, when we transform the template mesh into the subjects anatomical space (CCS) there is no significant difference in

localization error. Note that these results are obtained despite the fact that the SCS model should have been the best; since the synthetic data were generated using the similar, but with higher resolution, SCS model.

4. CONCLUSION

In this paper we have described a simple solution to the problem of reconstructing electromagnetic sources in a canonical anatomical space. Its simplicity rests on embedding subject specific anatomy into an extended forward model in a way that circumvents the need for cortical surface extraction. The forward model starts with a canonical cortical mesh, defined in a standard stereotactic space. The mesh is then warped into the subject's anatomical space. A conventional electromagnetic forward model is computed using the resulting warped mesh. The ensuing forward model is inverted using an established Bayesian scheme. Critically, the canonical mesh is warped using the inverse of the transformation used in conventional spatial normalization. This means that subject specific anatomy, encoded by the spatial transformation, can be derived from the subject's structural image using fully automated spatial normalization procedures that do not rely on high resolution or contrast.

The contribution of this work is twofold: first, conceptually we have formulated the problem of inter-subject anatomical variability as an explicit part of the forward model. This entails the notion of a canonical subject, whose cortical mesh is transformed anatomically to produce subject specific mesh. This places important constraints on individual meshes that enter the forward model; critically there must exist a diffeomorphic anatomical mapping between any subject and the canonical subject. We can exploit this constraint by always starting with the canonical mesh and warping it to match each subject. This has several fundamental advantages. First, it eschews the problems of cortical surface extraction from an individual's MRI; second it uses all the anatomical information in the MRI to construct a subject specific forward model (this information is not just confined to the cortical surface but includes all the information used in spatial normalization). Third, it ensures the cortical mesh is topologically valid (because it is derived under the diffeomorphism constraint). Finally, it enforces a standard solution space that facilitates inter-subject averaging and reporting. These standard spaces have proved very useful in fMRI.

The second contribution is the use of Bayesian model comparison, based on the model evidence or marginal likelihood to compare competing forward models. This enabled us to show that the models based on canonical meshes were at least as good as those based on individual cortical surface extraction. This provided a quantitative and principled way to explore model space and assess advances in model specification, of the sort addressed here.

We used Bayesian model comparison and localization error to evaluate the advantage of anatomically informed models (CCS) and to establish their construct validity in relation to conventional forward models based on cortical surface extraction (SCS). Importantly, our results do not show any systematic difference between the SCS and CCS models. This supports the idea that CCS is a sufficiently anatomically informed model to furnish a reasonable solution to the inverse problem. In other words, MEG data do not contain enough information about the fine-scale spatial configuration of sources to distinguish between the two models. Furthermore, TCS was significantly different from the other two models. This suggests that SCS or CCS should be used when possible. However, in the absence of structural MRI for any given subject, TCS remains a reasonable approximation, provided that it can be appropriately coregistered with the MEG data. The latter issue is crucial and will be addressed in a subsequent paper on optimizing use of template meshes, using only spatial information about sensor space (i.e., fiducials and head-shape data). Note finally, that we have focused on MEG and the use of a spherical head model. Although this approach could generalize to EEG in a straightforward way, we have not evaluated it yet in that context. This will require a careful analysis, due to the sensitivity of EEG to the geometry (and conductivity) of head tissues. This geometry is also subject specific and has been ignored here, because it is less important for MEG. However, a more realistic subject specific head model could be derived using the same approach used for the cortical mesh. This would entail using more realistic spheres or a boundary element model based on a canonical subject and warping it as described above. Again, Bayesian model comparison would enable us to assess the quantitative effect of realistic head tissue modeling.

To conclude, we have focussed on demonstrating the validity of the CCS model. This anatomically informed model has the twofold advantage of eschewing the need for cortical extraction and affording a one-to-one mapping with the canonical cortical surface. The latter is important for pooling results over subjects and reporting single subject or group localizations in the same stereotactic space. It also enables us to consider a full hierarchical model for multisubject analysis: namely, a unified inference scheme for group averages, instead of the conventional two-stage procedures (e.g., [17]). This includes, for example, incorporation of spatial priors on the MEG/EEG inverse solution based on normalized fMRI results from a group of subjects. This will be the focus of future work.

Software note

The algorithms described in this paper are available within SPM5 and can be downloaded from http://www.fil.ion.ucl.ac.uk/spm. It is worth emphasizing that the canonical cortical surface, given any subject's MRI, can be obtained automatically and robustly using the well-established spatial normalization schemes described in Section 2.2. SPM5 uses a unified forward model for anatomical deformations that includes tissue classification and inhomogeneity correction.

ACKNOWLEDGMENTS

The authors would like to thank Marcia Bennett for helping to prepare this manuscript. The Wellcome Trust and Medical Research Council funded this work. Jérémie Mattout is funded by the *Foundation pour la Recherché Médicale* (FRM). The authors are also grateful to Aude Entringer and Franck Morotti for helpful suggestions.

REFERENCES

[1] J. Talairach and P. Tournoux, *Co-Planar Stereotaxic Atlas of the Human Brain*, Thieme, New York, NY, USA, 1988.

[2] J. Ashburner and K. J. Friston, "Nonlinear spatial normalization using basis functions," *Human Brain Mapping*, vol. 7, no. 4, pp. 254–266, 1999.

[3] M. Fuchs, R. Drenckhan, H. A. Wischmann, and M. Wagner, "An improved boundary element method for realistic volume-conductor modeling," *IEEE Transactions on Biomededical Engineering*, vol. 45, no. 8, pp. 980–997, 1998.

[4] R. M. Leahy, J. C. Mosher, M. E. Spencer, M. X. Huang, and J. D. Lewine, "A study of dipole localization accuracy for MEG and EEG using a human skull phantom," *Electroencephalography and Clinical Neurophysiology*, vol. 107, no. 2, pp. 159–173, 1998.

[5] J. C. Mosher, M. E. Spencer, R. M. Leahy, and P. S. Lewis, "Error bounds for EEG and MEG dipole source localization," *Electroencephalography and Clinical Neurophysiology*, vol. 86, no. 5, pp. 303–321, 1993.

[6] C. Phillips, J. Mattout, M. D. Rugg, P. Maquet, and K. J. Friston, "An empirical Bayesian solution to the source reconstruction problem in EEG," *NeuroImage*, vol. 24, no. 4, pp. 997–1011, 2005.

[7] J. Mattout, C. Phillips, W. D. Penny, M. D. Rugg, and K. J. Friston, "MEG source localization under multiple constraints: an extended Bayesian framework," *NeuroImage*, vol. 30, no. 3, pp. 753–767, 2006.

[8] K. J. Friston, W. D. Penny, C. Phillips, S. Kiebel, G. Hinton, and J. Ashburner, "Classical and Bayesian inference in neuroimaging: theory," *NeuroImage*, vol. 16, no. 2, pp. 465–483, 2002.

[9] K. J. Friston, J. Mattout, N. Trujillo-Barreto, J. Ashburner, and W. D. Penny, "Variational free energy and the Laplace approximation," *NeuroImage*, vol. 34, no. 1, pp. 220–234, 2007.

[10] K. J. Friston, C. D. Frith, P. F. Liddle, and R. S. J. Frackowiak, "Plastic transformation of PET images," *Journal of Computer Assisted Tomography*, vol. 15, no. 4, pp. 634–639, 1991.

[11] J. Ashburner and K. J. Friston, "Unified segmentation," *NeuroImage*, vol. 26, no. 3, pp. 839–851, 2005.

[12] Y. Cointepas, J.-F. Mangin, L. Garnero, J.-P. Poline, and H. Benali, "BrainVISA: software platform for visualization and analysis of multi-modality brain data," in *Proceedings of the 7th Annual Meeting of the Organization for Human Brain Mapping (HBM '01)*, p. S98, Brighton, UK, June 2001.

[13] J.-F. Mangin, "From 3D magnetic resonance images to structural representations of the cortex topography using topology preserving deformations," *Journal of Mathematical Imaging and Vision*, vol. 5, no. 4, pp. 297–318, 1995.

[14] J. Sarvas, "Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem," *Physics in Medicine and Biology*, vol. 32, no. 1, pp. 11–22, 1987.

[15] W. D. Penny, K. E. Stephan, A. Mechelli, and K. J. Friston, "Comparing dynamic causal models," *NeuroImage*, vol. 22, no. 3, pp. 1157–1172, 2004.

[16] R. N. Henson, Y. Goshen-Gottstein, T. Ganel, L. J. Otten, A. Quayle, and M. D. Rugg, "Electrophysiological and haemodynamic correlates of face perception, recognition and priming," *Cerebral Cortex*, vol. 13, no. 7, pp. 793–805, 2003.

[17] R. N. Henson, J. Mattout, K. D. Singh, G. R. Barnes, A. Hillebrand, and K. J. Friston, "Population-level inferences for distributed MEG source localisation under multiple constraints: application to face-evoked fields," to appear in *Human Brain Mapping*.

[18] R. E. Kass and A. E. Raftery, "Bayes factors," *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 773–795, 1995.

Hindawi Publishing Corporation
Computational Intelligence and Neuroscience
Volume 2007, Article ID 61916, 11 pages
doi:10.1155/2007/61916

Research Article

A Subspace Method for Dynamical Estimation of Evoked Potentials

Stefanos D. Georgiadis, Perttu O. Ranta-aho, Mika P. Tarvainen, and Pasi A. Karjalainen

Department of Physics, University of Kuopio, P.O. Box 1627, 70211 Kuopio, Finland

Correspondence should be addressed to Stefanos D. Georgiadis, stefanos.georgiadis@uku.fi

Received 16 February 2007; Revised 7 June 2007; Accepted 18 September 2007

Recommended by Saied Sanei

It is a challenge in evoked potential (EP) analysis to incorporate prior physiological knowledge for estimation. In this paper, we address the problem of single-channel trial-to-trial EP characteristics estimation. Prior information about phase-locked properties of the EPs is assesed by means of estimated signal subspace and eigenvalue decomposition. Then for those situations that dynamic fluctuations from stimulus-to-stimulus could be expected, prior information can be exploited by means of state-space modeling and recursive Bayesian mean square estimation methods (Kalman filtering and smoothing). We demonstrate that a few dominant eigenvectors of the data correlation matrix are able to model trend-like changes of some component of the EPs, and that Kalman smoother algorithm is to be preferred in terms of better tracking capabilities and mean square error reduction. We also demonstrate the effect of strong artifacts, particularly eye blinks, on the quality of the signal subspace and EP estimates by means of independent component analysis applied as a prepossessing step on the multichannel measurements.

1. INTRODUCTION

Evoked potentials (EPs) and ongoing brain activity oscillations, obtained by scalp electroencephalogram (EEG) recordings, have been linked with various cognitive processes and provide means for studying cerebral brain function [1]. An EP is usually considered to be a wave or complex elicited by and time-locked to a physiological or nonphysiological stimulation or event. EPs are buried into background brain activity, and nonneural activity like muscle noise. Since many parallel mental processes may occur simultaneously in the brain, it is difficult to observe and determine an evoked potential on a single-trial base. Therefore, the simplest way to investigate EPs is to use ensemble averages of time-locked EEG epochs obtained by repeated stimulation. It is well known that this signal enhancement implies a loss of information related to trial-to-trial variability, and nonstationary features of event-related phenomena.

The generation mechanism of evoked responses is not precisely known in many situations. EPs are assumed to be generated either separately of ongoing brain activity, or through stimulus-induced reorganization of ongoing activity. For example, it might be possible that during the performance of an auditory oddball discrimination task, the brain activity is being restructured while attention is focused on the target stimulus [2]. Phase synchronization of ongoing brain activity is one possible mechanism for the generation of event-related responses. That is, following the onset of a sensory stimulus, the phase distribution of ongoing activity changes from uniform to one which is centered around a specific phase [3]. Moreover, several studies have concluded that averaged EPs are not separate from ongoing cortical processes, but rather, are generated by phase synchronization and partial phase resetting of ongoing activity [4, 5]. However, phase coherence over trials observed with common signal decomposition methods (e.g., wavelets) can result both from a phase-coherent state of ongoing rhythms and from the presence of a phase-coherent event related potential, which is additive to ongoing EEG [6]. Furthermore, stochastic changes in amplitude and latency of different components of the EPs are able to explain significant part of intertrial variability of the measurements [6–9].

Several methods have been proposed for EP estimation and denoising; see, for example, [10–13]. In general, most of the methods for single-trial EP analysis aim to decompose the measurements into relevant components or to explain

the data through some parameters. The parametrization gives the necessary means to investigate, for example, the changes that the stimulus causes to the ongoing EEG signal, or that the repetition of the test causes to the responses. Most of the methods are based on an explicit model or on some specific assumptions for the EPs. Every decomposition then involves at least two main considerations. On the one hand, if the resulting estimates follow too closely the measurements, it is possible that some features of the data are still going to be hidden by phenomena unrelated to the stimulation. On the other hand, if the estimates do not follow the measurements, some features may have been neglected. Usually a balance between these considerations is made and care is given to the correct interpretation of a parametrization that is able to reveal specific features of the experiment.

The performance and applicability of every single-trial estimation method depends on the prior information used and the statistical properties of the EP signals. Here, we focus on the case that some parameters of the EPs change dynamically from stimulus to stimulus. This situation could be a trend-like change of the amplitude or latency of some phase-locked component of the EPs. Although, for example, the above-mentioned methods [10–13] could be used to estimate such changes, they do not take into account in the estimation procedure this trend-like variability.

The most obvious way to handle time variations between single-trial measurements is subaveraging of the measurements in groups. Subaveraging could give optimal estimators if the EPs are assumed to be invariant within the subaveraged groups. A better approach is to use moving window or exponentially weighted average filters; see, for example, [14, 15]. Other adaptive methods have also been proposed for EP estimation, especially for brain stem potential tracking, for example, [16]. The statistical properties of some average filters and different recursive estimation methods for EP estimation have been discussed through Kalman filtering in [17]. Some smoothing methods have also been proposed for modeling trial-to-trial variability in EPs (e.g., [18]).

An elegant way to describe trial-to-trial variations in EPs can be given through state-space models. State-space modeling for single-trial dynamical estimation considers the EP as a vector-valued random process with stochastic fluctuations from stimulus to stimulus [17]. Then, past and future realizations contain information of relevance to be used in the estimation procedure. Recursive estimates for the states, that are optimal in the mean square sense, are given by Kalman filter and smoother algorithms. Of importance is also the parametrization of the problem and the selection of an observation model for the measurements. For example, in [16, 17] generic observation models were used based on shifted Gaussian-shaped smooth functions. While other generic observation models could also be considered, when all the measurements are available, data-based observation models can be used.

In this paper, we extend the method presented in [17] to the use of Kalman smoother algorithm. We demonstrate that for batch processing the use of the smoother algorithm is preferable. Fixed-interval smoothing improves the tracking performance of EP characteristics and reduces greater the noise. In parallel, we propose a novel method for state-space modeling of EPs. The method is based on the eigenvalue decomposition of the ensemble data correlation matrix. A few dominant eigenvectors form a signal subspace that can be used for single-trial estimation. Subspace-based methods have already been proposed for EP estimation, for example, in [12, 19]. However, these approaches do not take into account in the estimation procedure the situation that some characteristics of the EPs change dynamically from stimulus to stimulus. In this paper, we demonstrate that such a signal subspace can be used to model dynamic changes present in EP measurements.

The approach is demonstrated with simulated and real measurements obtained by an auditory EP experiment. Finally, we investigate the effect of strong artifacts on the quality of the estimates by means of independent component analysis (ICA), which is applied as a prepossessing step on the multichannel measurements.

2. METHODS

The sampled potential (from channel l) relative to the successive stimulus or trial t can be denoted with a column vector of length M:

$$z_t = \begin{pmatrix} z_t(1) \\ z_t(2) \\ \vdots \\ z_t(M) \end{pmatrix}, \quad t = 1, \ldots, T, \tag{1}$$

where T is the total number of trials.

2.1. Linear estimation and additive noise model

A widely used model for EP estimation is the additive noise model. The observations are then assumed to be of the form

$$z_t = s_t + v_t. \tag{2}$$

The vector s_t corresponds to the part of the activity that is related to the stimulation, and the rest of the activity v_t is usually assumed to be independent of the stimulus and the EP. Single-trial EPs can be further modeled as a linear combination of some preselected basis vectors. Then, the observation model takes the form

$$z_t = H_t \theta_t + v_t, \tag{3}$$

where H_t is the observation matrix, which contains the basis vectors $\psi_{t,1}, \ldots, \psi_{t,k}$ of length M in its columns, and θ_t is a parameter vector of length k. The estimated EPs $\hat{s}_t$ can then be obtained by using the estimated parameters $\hat{\theta}_t$ as follows:

$$\hat{s}_t = H_t \hat{\theta}_t. \tag{4}$$

By treating both θ_t and v_t as random, the estimator $\hat{\theta}_t$ that minimizes the mean square Bayes cost $B_{\mathrm{MS}} = E\{\| \theta_t - \hat{\theta}_t \|^2\}$ is given by the conditional mean [20]

$$\hat{\theta}_t = E\{\theta_t \mid z_t\} \tag{5}$$

of the posterior distribution

$$\begin{aligned} p(\theta_t \mid z_t) &\propto p(z_t \mid \theta_t)\, p(\theta_t) \\ &\propto p_{v_t}(z_t - H_t\theta_t \mid \theta_t)\, p(\theta_t). \end{aligned} \tag{6}$$

By taking into account the linear observation model, and that θ_t and v_t are assumed uncorrelated, that is, $C_{\theta_t, v_t} = 0$. the linear conditional mean estimator takes the form [20]

$$\hat{\theta}_t = \left(H_t^T C_{v_t}^{-1} H_t + C_{\theta_t}^{-1}\right)^{-1} \left(H_t^T C_{v_t}^{-1} z_t + C_{\theta_t}^{-1} \eta_{\theta_t}\right), \tag{7}$$

where C_{θ_t} and η_{θ_t} are, respectively, the covariance and the mean of θ_t. C_{v_t} is the covariance of the zero mean measurement noise, and $(\cdot)^T$ denotes transpose. The estimator is optimal in the mean square sense among all possible estimators, not only linear, if θ_t and v_t are Gaussian. In Bayesian estimation this is also called the maximum a posteriori estimator (MAP), and C_{θ_t} and η_{θ_t} represent prior information about the parameters θ_t. If they are not available, we can assume $C_{\theta_t}^{-1} = 0$ corresponding to infinite prior variance for the parameters. In this case, the estimator reduces to the ordinary minimum variance Gauss-Markov estimator, which treats the parameters as nonrandom. If we assume that the errors are independent with equal variances $C_{v_t} = \sigma_{v_t}^2 I$. the estimator is identical to the ordinary least squares estimator

$$\hat{\theta}_t = \left(H_t^T H_t\right)^{-1} H_t^T z_t. \tag{8}$$

2.2. State-space modeling of EPs

Estimators of the form (7) can be used to model time-varying characteristics of EPs, for example, in terms of amplitude and latency estimates of some characteristic peak of the signals. However, such estimators do not take into account situations that some dynamical behavior is expected from stimulus to stimulus. A mathematical plausible way to incorporate prior information for estimation about time-varying phenomena is given through state-space modeling.

The measurement vectors z_t can be considered as realizations of a stochastic vector process, that depends on some unobserved parameters θ_t (state vector) through the model (3). The parameters θ_t are the quantities that we are primarily interested in, and their form depends on the parametrization of the estimation problem. In order to model the time evolution of the hidden process θ_t, a linear first-order Markov model can be used, that is,

$$\theta_t = F_t\theta_{t-1} + \omega_t, \tag{9}$$

with some initial distribution for θ_0. Equations (3) and (9) form a linear state-space model, where F_t and H_t are preselected matrices. Other important assumptions for the model are

(i) for every $i \neq j$, the observation noise vectors v_i, v_j as well as the state noise vectors ω_i, ω_j are mutually independent and also mutually independent of the initial state θ_0,

(ii) the vectors ω_i, v_j are mutually independent for all i, j.

For the white noise sequences ω_t and v_t, we can also assume $E\{\omega_t\} = 0$ and $E\{v_t\} = 0$ for every t, but the covariances C_{ω_t}, C_{v_t} can still be time-varying.

2.3. Kalman filter and smoother algorithms

The Kalman filtering problem is related to the determination of the mean square estimator $\hat{\theta}_t$ for the state θ_t given the observations $z_1, \ldots, z_t$. This is equal to the conditional mean

$$\hat{\theta}_t = E\{\theta_t \mid z_1, \ldots, z_t\} = E\{\theta_t \mid Z_t\}, \tag{10}$$

that relates to the density [20]

$$p(\theta_t \mid Z_t) \propto p(z_t \mid \theta_t)\, p(\theta_t \mid Z_{t-1}), \tag{11}$$

where

$$p(\theta_t \mid Z_{t-1}) = \int p(\theta_t \mid \theta_{t-1})\, p(\theta_{t-1} \mid Z_{t-1})\, d\theta_{t-1}. \tag{12}$$

The optimal linear mean square estimator can then be obtained recursively by restricting to a linear conditional mean, or by assuming v_t and ω_t to be Gaussian [20]. The recursive estimator can be written as

$$\hat{\theta}_t = \left(H_t^T C_{v_t}^{-1} H_t + C_{\tilde{\theta}_{t|t-1}}^{-1}\right)^{-1} \left(H_t^T C_{v_t}^{-1} z_t + C_{\tilde{\theta}_{t|t-1}}^{-1} \hat{\theta}_{t|t-1}\right), \tag{13}$$

where $\hat{\theta}_{t|t-1}$ is the prediction of θ_t based on $\hat{\theta}_{t-1}$ and $\hat{\theta}_{t-1} = E\{\theta_{t-1} \mid z_{t-1}, \ldots, z_1\}$ is the optimal MS estimate at time $t-1$. Clearly this is of the form (7), which is the Bayesian MAP estimator using the last available estimate as prior information. After adding the initializations, Kalman filter algorithm can be written as follows.

(i) Initialization:

$$\begin{aligned} C_{\tilde{\theta}_0} &= C_{\theta_0}, \\ \hat{\theta}_0 &= E\{\theta_0\}. \end{aligned} \tag{14}$$

(ii) Prediction step:

$$\begin{aligned} \hat{\theta}_{t|t-1} &= F_t \hat{\theta}_{t-1}, \\ C_{\tilde{\theta}_{t|t-1}} &= F_t C_{\tilde{\theta}_{t-1}} F_t^T + C_{\omega_t}. \end{aligned} \tag{15}$$

(iii) Filtering step:

$$\begin{aligned} K_t &= C_{\tilde{\theta}_{t|t-1}} H_t^T \left(H_t C_{\tilde{\theta}_{t|t-1}} H_t^T + C_{v_t}\right)^{-1}, \\ \hat{\theta}_t &= \hat{\theta}_{t|t-1} + K_t\left(z_t - H_t\hat{\theta}_{t|t-1}\right), \\ C_{\tilde{\theta}_t} &= \left(I - K_t H_t\right) C_{\tilde{\theta}_{t|t-1}}, \end{aligned} \tag{16}$$

for $t = 1, \ldots, T$. The matrix K_t is called the Kalman gain matrix.

If all the measurements are available, that is, $z_t, t = 1, \ldots, T$, then the fixed interval smoothing problem can be considered, that is,

$$\hat{\theta}_t^s = E\{\theta_t \mid z_1, \ldots, z_T\} = E\{\theta_t \mid Z_T\}, \tag{17}$$

that relates to the density [21]

$$p(\theta_t \mid Z_T) = p(\theta_t \mid Z_t) \int \frac{p(\theta_{t+1} \mid \theta_t)\, p(\theta_{t+1} \mid Z_T)}{p(\theta_{t+1} \mid Z_t)}\, d\theta_{t+1}. \tag{18}$$

The last form suggests again a recursive estimation procedure for the determination of the conditional density. It is thus possible to compute filtered and prediction distributions in a forward (filtering) recursion, and then execute a backward recursion with each smoothed distribution $p(\theta_t \mid Z_T)$ relying upon the quantities calculated in the forward run and the previous (in reverse time) smoothed distributions $p(\theta_{t+1} \mid Z_T)$. This property enables the formulation of the forward-backward method for the smoothing problem [22], which gives the smoother estimates as corrections of the filter estimates. So for the linear or Gaussian case the smoothing problem is complete through the backward recursion.

Smoothing:

$$\begin{aligned} A_t &= C_{\tilde{\theta}_t} F_{t+1}^T C_{\tilde{\theta}_{t+1|t}}^{-}, \\ \widehat{\theta}_t^s &= \widehat{\theta}_t + A_t\left(\widehat{\theta}_{t+1}^s - \widehat{\theta}_{t+1|t}\right), \\ C_{\tilde{\theta}_t^s} &= C_{\tilde{\theta}_t} + A_t\left(C_{\tilde{\theta}_{t+1}^s} - C_{\tilde{\theta}_{t+1|t}}\right)A_t^T, \end{aligned} \tag{19}$$

for $t = T-1, T-2, \ldots, 1$. For the initialization of the backward recursion the filter estimates can be used, that is, $\widehat{\theta}_T^s = \widehat{\theta}_T$.

2.4. Signal and noise subspaces

Singular value decomposition (SVD) has many theoretical and practical applications in signal processing and identification problems [23]. In relatively high signal-to-noise ratio conditions (SNR), SVD of a data matrix can divide measurements into signal and noise subspaces. Alternatively, it can also be understood in terms of principal component regression (PCR) as a combined method for signal enhancement and optimal model dimension reduction [24]. The subspace method has been used to enhance stimulus phase-locked activity in different studies (e.g., [19]).

The available data matrix $Z = [z_1, \ldots, z_T] \in \mathbb{R}^{M\times T}$, which has as columns the EEG sampled epochs relative to the stimulation, can be decomposed as

$$Z = U\Sigma V^T, \tag{20}$$

where $U \in \mathbb{R}^{M\times M}$ satisfies $U^TU = I$, $V \in \mathbb{R}^{T\times T}$ satisfies $V^TV = I$, and $\Sigma \in \mathbb{R}^{M\times T}$ is a pseudodiagonal matrix with nonnegative diagonal elements σ_i such that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(M,T)} \geq 0$. If $M \leq T$, then Σ has the form $\Sigma = [\Sigma_1, 0]$, where $\Sigma_1 = \operatorname{diag}(\sigma_1, \ldots, \sigma_M)$ and 0 is a zero matrix. If $M > T$, then Σ has the form $\Sigma = \left[\begin{smallmatrix}\Sigma_1\\ 0\end{smallmatrix}\right]$, where $\Sigma_1 = \operatorname{diag}(\sigma_1, \ldots, \sigma_T)$. Only r singular values are nonzero, where $r = \operatorname{rank}(Z)$.

For the additive noise model and relatively small noise the following decomposition can be considered:

$$Z = [U_s, U_v]\begin{bmatrix}\Sigma_s & 0\\ 0 & \Sigma_v\end{bmatrix}[V_s, V_v]^T. \tag{21}$$

The matrix Σ_s contains the k largest singular values and U_s the respective left singular vectors associated mainly with the signals s_t. Thus the matrices (U_s, Σ_s, V_s) represent a signal subspace, and (U_v, Σ_v, V_v) represent primarily the noise subspace.

From the SVD of the matrix $Z = U\Sigma V^T$ we also have

$$ZZ^T = U\Sigma_1^2U^T. \tag{22}$$

This means that the left singular vectors of Z are the eigenvectors of the matrix ZZ^T, or the eigenvectors of the data correlation matrix

$$\widehat{R} = \frac{1}{M}ZZ^T. \tag{23}$$

If we denote with H_s the matrix with columns the k dominant eigenvectors, then the ordinary least squares estimator for the parameters θ_t becomes

$$\widehat{\theta}_t = \left(H_s^TH_s\right)^{-1}H_s^Tz_t = H_s^Tz_t. \tag{24}$$

Estimates for the EPs can then be obtained from (4). Quantitatively, the first basis vector is the best mean-square fit of a single waveform to the entire set of epochs. Thus, the first eigenvector is similar to the mean of the epochs, and the corresponding parameters or principal component $\widehat{\theta}_t(1)$ $(t = 1, 2, \ldots, T)$ reveal the contribution of the eigenvector to each epoch. The rest of the dominant eigenvectors model primarily amplitude differences between individual EP peak components, and latency variations from trial to trial. Therefore, since this basis contains prior information about phase-locked characteristics of the EP signals, we consider the following state-space model for dynamical estimation:

$$\begin{aligned} \theta_t &= \theta_{t-1} + \omega_t, \\ z_t &= H_s\theta_t + \upsilon_t, \end{aligned} \tag{25}$$

with the selections $F_t = I$, $t = 1, \ldots, T$, that is, a random walk model, and $H_t = H_s$ for all t. Estimates for the parameters can then be obtained by Kalman filter and smoother algorithms for different selections of state and observation noise covariance matrices. Thus, the applicability of the proposed method relates on the quality of the signal subspace in low signal-to-noise ratio conditions, as well as on the assumption of hidden dynamical behavior from trial-to-trial.

2.5. Artifact correction by ICA

Individual EEG channels measure superimposed activity generated simultaneously by various brain sources. The behavior of the sources is stochastic and generally nonstationary. In addition, artifact sources, such as eye blinks, can distort statistical properties of the signals and increase complexity. For the problem of blind source separation (BSS) of the multichannel EEG measurements, target is to recover unobserved brain generated initial source signals by using only the available sensor data and some statistical properties assumed for the sources [25, 26].

Fundamentally, the basic problem that BSS attempts to solve assumes a set of L measured data points $x_n = (x_n(1), \ldots, x_n(l), \ldots, x_n(L))^T$ at time instant n $(n = 1, \ldots, N)$

to be a linear combination of m unknown sources $y_n = (y_n(1), \ldots, y_n(m))^T$, that is

$$x_n = Ay_n + v_n. \tag{26}$$

For EEG measurements, L is the number of available channels, and the measurements can be summarized in a matrix X having the vectors x_n in its columns and different channel recordings in each row. A time-invariant mixing matrix A is the common approach for ICA and BSS of EEG, for example, in event-related studies [3]. This model can be interpreted as the fixed biophysical structure of the brain itself whilst the sources distributed within this structure change their intensity over time [25].

A general formulation for BSS without any assumptions (prior information) about the nature of the data, noise, or mixing system will leave the problem of EEG separation intractable. Therefore, some basic assumptions are needed. For example, the goal of ICA is to recover independent sources given only sensor observations that are unknown linear mixtures of unobserved independent source signals [27, 28].

The assumption of physiological independence of the sources can be quite obvious in some situations, for example, when used in artifact rejection separating brain signals from ocular artifacts. Note that the ICA model considers the signals as independent and identically distributed, and requires non-Gaussian sources. Thus, by ignoring time structure, the estimation is based solely on investigating structure across the sensors as estimated by the sample distribution of the measurements, and an embedded density parametrization (differentiating at least between sub-Gaussian and super-Gaussian sources). Therefore, the model might not be able to separate every kind of sources (e.g., stationary Gaussian random processes). However, in many situations predominant artifacts show a highly kurtotic sample distribution that enables estimation.

ICA methods carry ambiguities about the ordering and the overall amplitude and sign of the estimated sources. The rows of the data matrix X are the EEG channel recordings and are decomposed as $X = AY$, where Y has in its rows the independent components. The mixing matrix A contains the spatial information of the sources obtained at the sensors. Therefore, the columns of A are the spatial distributions of the estimated sources, which are normalized to unit variance. For example, eye movements and eye blinks project mainly to frontal sites. An artifact source can be eliminated and removed from the measurements by backprojection.

3. RESULTS

In this section, we present the performance of Kalman filter and smoother algorithms on tracking dynamic variations, and estimating single-trial EPs in a simulated and a real data set. In parallel, we investigate the performance of the method when the signal subspace is enhanced by rejecting eye-related artifacts with the use of ICA.

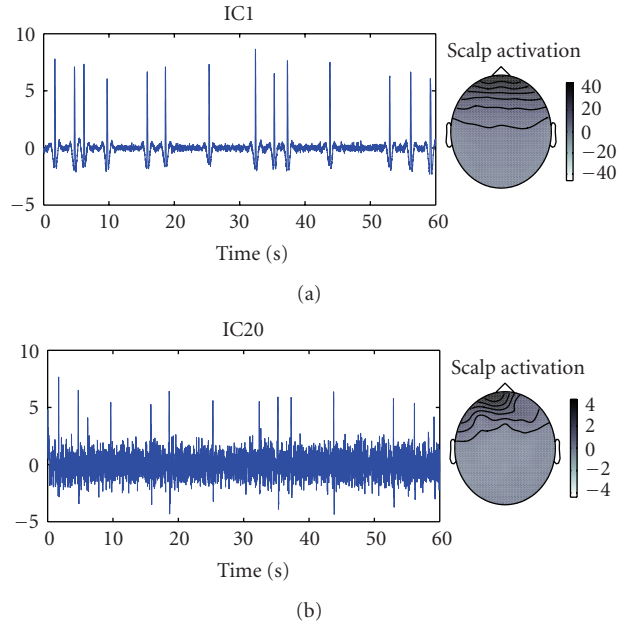


Figure 1: Blink-related components estimated with ICA. Time activations (left) and scalp activations (right). The left plots correspond to the first minute of the measurement set.

3.1. Measurements and artifact removal

EEG measurements were obtained from a standard oddball paradigm with auditory stimulation (1 subject, 60 EEG channels, reference: ears). In the recording, 569 auditory stimuli were presented with an interstimulus interval of 1 second. Eighty-five percent of the stimuli were the standard tones at 800 Hz. Fifteen percent were the deviant tones at 560 Hz. The deviant tones were randomly presented. The subject was sitting in a chair, and was asked to press a button every time he heard the deviant target tone. The sampling rate of the measurements was 500 Hz.

Reduction in noise for EEG signals can be done with linear filtering without altering the basic ICA model [27]. If we further assume less sources than sensors and that the sensor noise is relatively small, then principal component analysis (PCA) on the data covariance matrix and dimension reduction can be used to reduce the noise and to prevent overlearning [27]. For the analysis, the data were digitally filtered in the range (1–35 Hz). All the measurement set (about 10 minutes) was used for the estimation of the separating matrix. The dimension of the data was reduced with PCA to 31, by keeping eigenvectors associated with eigenvalues larger than 1, resulting in more than 99% of explained variance. The FastICA algorithm in parallel form [27] was used for the estimation of independent components.

By visual inspection of the estimated components and scalp activations two components showed to be related to eye activity. The blink components are presented in Figure 1. On the left, the time activations corresponding to the first minute of the recordings are presented, and on the right the spatial distributions. Furthermore, these components did not show any significant correlation with the two types of

stimuli (standard and target). Correlation with stimulation time was investigated by computing EP image plots for every estimated component. The component-based EP image plots are not shown here, but such images are also used in the next section (Figures 3 and 5). EP image plots are constructed by color-coding potential variations occurring in single-trial epoch vectors (e.g., [3]). The thin color-coded horizontal bars, each representing a single-trial, are, for example, stacked row-by-row according to data collection time (data epochs sampled relative to successive stimulus or trial t) producing an EP image.

Note that PCA-based dimension reduction is a rather subjective approach for the determination of the number of brain source signals in EEG measurements [25]. Some relatively weak brain sources, as measured at the sensors, may be eliminated. Additionally, some estimated independent components may remain the mixture of more than one source signals. However, by computing different EP image plots we did not observe any significant loss of phase-locked EP activity. Furthermore, filtering and dimension reduction provided good estimates for the blink components and fast convergence for the FastICA algorithm. Therefore, the performance was considered satisfactory for ocular artifact removal, and for the demonstration needs of the proposed subspace method for dynamical estimation of single-channel single-trial EPs.

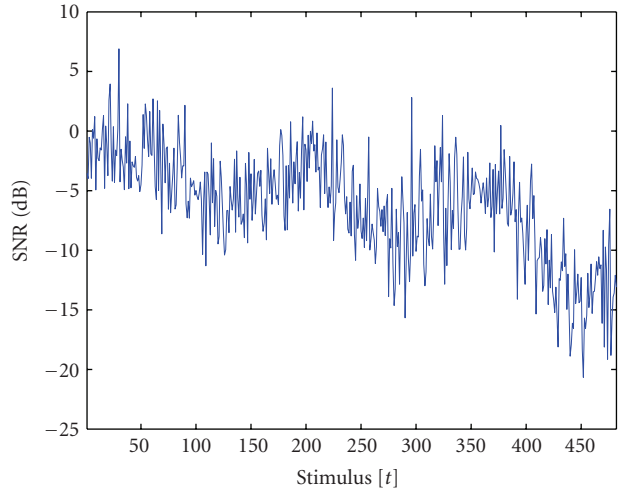


Figure 2: Time-varying SNR(dB) for the simulated second peak as a function of the stimulus number (trial) t, that is, $\mathrm{SNR}_t = 10\log_{10}\sum_i s_t^2(i)/\sum_i v_t^2(i)$, $t = 1,\ldots,T$, where s_t are the simulated noise-free single-trial EPs and v_t prestimulus EEG epochs sampled relative to the standard tone from channel CZ after ocular artifact removal with ICA. The sums were considered in a smaller interval around 200 milliseconds covering only the second peak (see also Figure 3).

3.2. Single-trial estimation

Real EEG data were used as background EEG activity, or noise, in the simulations. From the recordings, we used only the channel CZ, after preprocessing and artifact removal by ICA. Only ocular artifacts were considered. As background activity for the simulations, we sampled prestimulus EEG epochs from −500 milliseconds to 0 millisecond relative to the standard stimulus onset. Simulated EPs were constructed according to the additive noise model by superimposing upon the selected real EEG epochs linear combinations of 2 Gaussian-shaped functions. In order to be consistent to the real measurements (standard tones and N100/P200 complex), each pseudoreal EP vector has two Gaussian peaks: a negative after 100 milliseconds and a positive after 200 milliseconds. Trial-to-trial sinusoidal variations for the amplitude and latency of the second peak were generated. Random variations were also added to the amplitudes, latencies, and widths of both simulated peaks.

The estimated time-varying SNR with respect only to the second peak can be seen in Figure 2 as a function of the stimulus number t. Therefore, the important assumption in the simulations is the trend-like behavior in low signal-to-noise ratio conditions. By construction the simulated EPs have trend-like trial-to-trial characteristics. This can be observed in Figure 3 (left) and the EP image plots. In the same figure (bottom, left), they are also presented the 10 dominant eigenvectors of the data correlation matrix obtained before and after EEG addition.

It must be noted that the aim in the creation of the simulations was that the average of the simulated EPs is close to the average of the real measurements (standard tones and N100/P200 complex at channel CZ). The average of the real measurements has a negative peak around 110 milliseconds (N100) with amplitude about $-4\,\mu V$, and a positive peak (P200) around 230 milliseconds with amplitude about $5\,\mu V$. Then the simulations were created as follows. For the first peak random variability in a small range in amplitude and latency was simulated that gives ensemble average with peak amplitude about $-4\,\mu V$ at the required latency. For the second peak dynamic variability was created with range of about $10\,\mu V$ (2–$12\,\mu V$) in amplitude and about 45 milliseconds in latency, such that the average has peak amplitude about $6\,\mu V$ and similar latency to the real measurements. Then prestimuli EEG was added. In that respect, SNR conditions were not directly considered, but instead a reasonable range for the time-varying behavior was assumed that can produce similar average with the real measurements.

For estimation the state-space model (25) was selected. For the covariances we used $C_{\omega_t} = \sigma_\omega^2 I$ and $C_{v_t} = \sigma_v^2 I$ for every stimulus t. Then the selection of the last variance term is not essential since only the ratio $\sigma_v^2/\sigma_\omega^2$ has effect on the estimates. Then the choice $C_{v_t} = I$ can be made and care should be given to the selection of σ_ω^2. In general, if it is tuned too small, fast fluctuations of EPs are going to be lost, and if it is selected too big the estimates have too much variance and they will tend to be similar to the ordinary least squares or principal component regression solution. The selection can be based on experience and visual inspection of the estimates as a balance between preserving expected dynamic variability and greater noise reduction.

In order to identify an optimal value for the variance term σ_ω^2 for the simulations we calculated root mean square

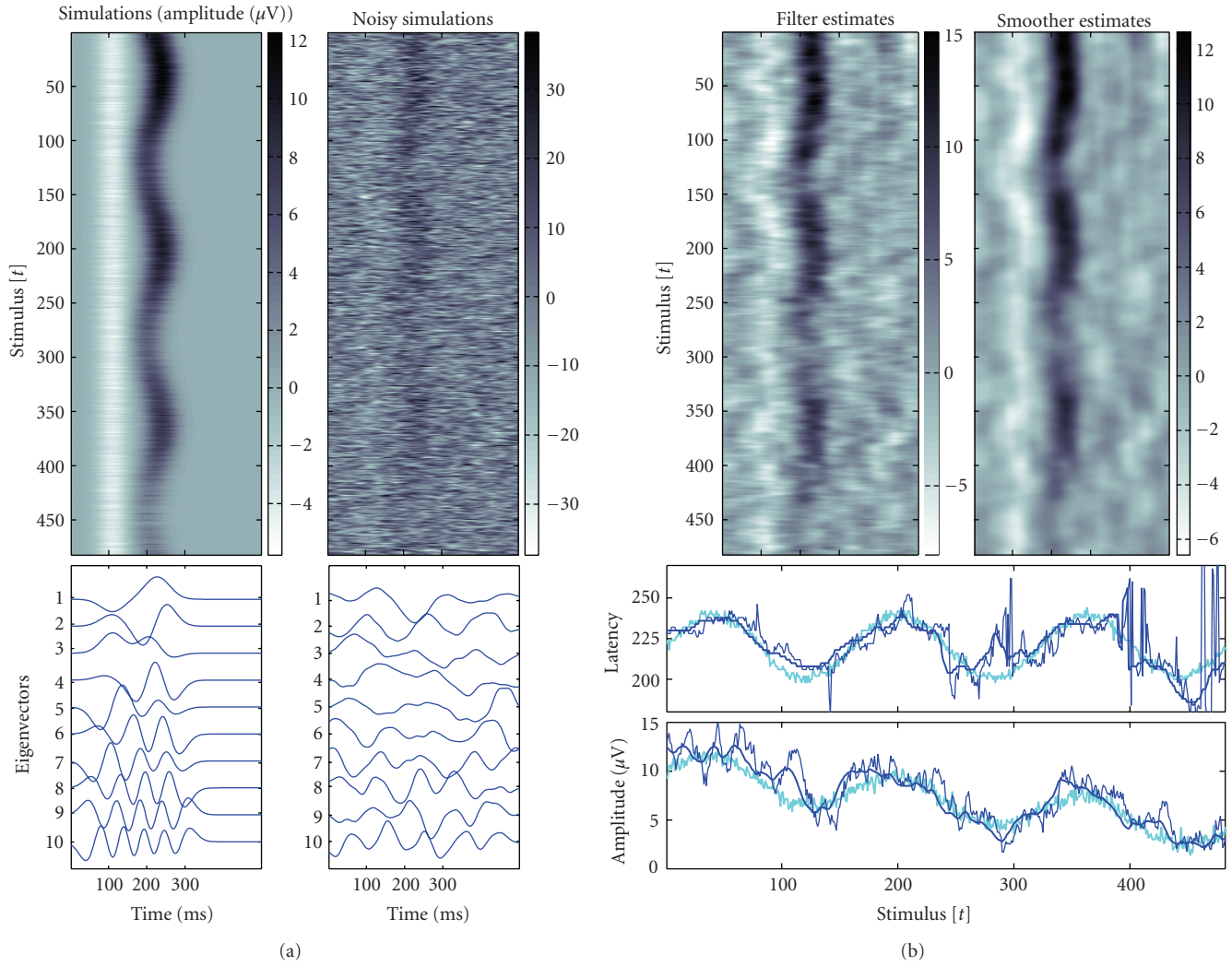


FIGURE 3: Simulations resembling the N100/P200 auditory complex and obtained estimates. For background noise prestimuli EEG samples relative to standard tones from channel CZ after ocular artifacts removal were used. Left: simulations (Gaussian functions) and noisy simulations, single-trials as image plots (up), and the respective 10 dominant eigenvectors of the data correlation matrix (bottom). The EP images represent stimulus locked stacked epochs (row-by-row). The color-maps describe the amplitude level in μV, y-axis represents successive stimulus or trial t, and the x-axis represents within a trial latency variation. Right: single-trial estimates as image plots with Kalman filter and smoother (up) and estimated variability of the second positive peak (bottom). Simulated amplitude and latency trends (light bold), estimates based on Kalman filter (dark thin) and based on fixed-interval Kalman smoother (dark bold). For estimation the selection $\sigma_\omega^2 = 10^{-2}$ was used.

errors (RMSEs) between the estimates based on the noisy data and the noiseless simulated EPs. The RMSEs were computed with respect to the second peak only over a smaller time interval around 200 milliseconds. For initialization of the algorithms we used half the data set by filtering backwards in time. The last estimates were used for initializing the forward run. Finally, the last state estimate of the Kalman filter forward run was used to initialize the backward smoothing procedure.

Means of RMSEs over all single-trials for different values of state noise variance parameter and for different dimensions of the observation matrix are presented in Figure 4 as contour plots for Kalman filter (top) and smoother (middle). In all the cases, Kalman smoother provides smaller error than the filter. This is to be expected, since all the measurements are included in the estimation procedure. The reduction of the error during backward smoothing is due to greater noise cancellation, as well as better tracking of the dynamic fluctuations. Optimal values of σ_ω^2 for all the selected observation matrices are between 10^{-3} and 10^{-2}. By considering the contour plots and by inspection of the estimates, around 10 eigenvectors are enough for tracking the dynamic fluctuations. Single-trial estimates for that dimension ($k = 10$) and with the selection $\sigma_\omega^2 = 10^{-2}$ are presented in Figure 3 as image plots for Kalman filter and smoother. In the right (bottom) of the same figure they are presented estimates for the single-trial latency and amplitude of the second peak as a function of the stimulus number or trial t.

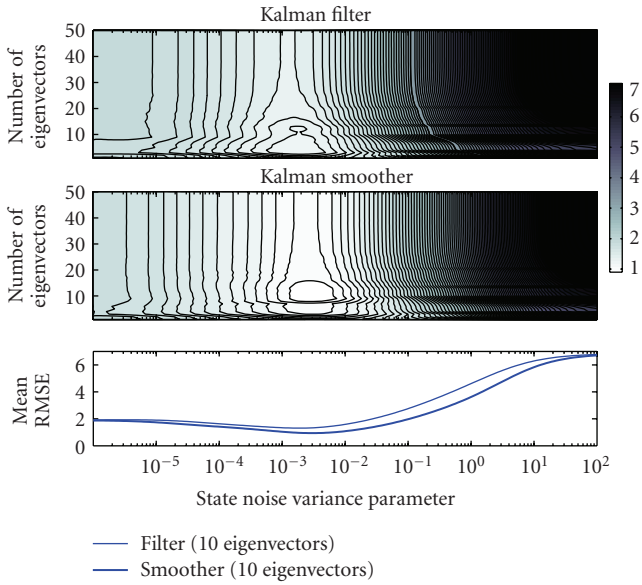


Figure 4: Means of RMSEs for different values of the state noise variance parameter σ_ω^2 and different number of dominant eigenvectors included in the observation model. Contour plots of the means for Kalman filter (top) and smoother (middle). Means when 10 eigenvector are included in the observation model (bottom). In all plots the x-axis is in logarithmic scale.

State-space representation and a few dominant eigenvectors obtained from the ensemble data correlation matrix are able to model the amplitude and latency changes. Bayesian recursive mean square estimation is able to reveal the hidden dynamic variability under unfavorable signal-to-noise ratio conditions. Clearly, Kalman smoother tracks better the dynamic changes and reduces greater the noise.

For the real measurements we considered epochs 0–500 milliseconds after the presentation of the standard tones from channel CZ before and after eye artifact removal. For the two data sets we selected 10 eigenvectors of the data correlation matrix for estimation. The strong blink contributions clearly affect the eigenvectors and the signal subspace, especially after the first half of the measurements, see Figure 5. This can also be seen by observing the first two eigenvectors that reflect mainly blink artifacts. However, since the blinks occur random enough, recursive mean square estimation is largely reducing their contribution. This can be observed in Figure 5 from the estimates, which are obtained with Kalman smoother with the same choices $\sigma_\omega^2 = 10^{-2}$ and $k = 10$ for both data sets. The estimated dynamic variability of the second peak (P200) in terms of amplitudes and latencies is presented in the left (bottom) of the same figure.

Some representative individual single-trial estimates are presented in Figure 6 for the simulations (left) and real EP measurements (right). The estimates for the simulations and the real EP measurements (standard tones and N100/P200 complex) are based on the artifact corrected EEG and Kalman smoother algorithm. The identification of peak potentials from raw measurements can be misleading even in simple simulations (e.g., stimulus number $t = 50$, left). The proposed method produced accurate estimates for the simulations even in very low SNR conditions (e.g., stimulus number $t = 450$, left). This is because we assumed a trend-like variability. The evaluation of the estimates for the real EPs is naturally more difficult. For example, clear N100 and P200 peaks are obtained for stimuli 50 and 250 (right). Though, the identification of peaks is not trivial for stimulus 450 (right). However, it must be noted that the proposed method does not make assumptions for the number of peaks and their exact form. This information is obtained from the estimated signal subspace and the included eigenvectors.

In summary, the proposed approach for single-trial dynamical estimation of EPs consists of the following steps. (1) Band-pass filter the selected EEG channels. This has as an effect on the improvement of the quality of the signal subspace. For example, it can reduce high-frequency components, and therefore, it can provide smoother eigenvectors and estimates. (2) Enhance the quality of the signal subspace. If the EEG epochs contain strong artifact contributions, such as eye blinks, an artifact correction method can be applied, for example, ICA. (3) Estimate the data correlation matrix and compute eigenvectors. In the simplest case, a basic artifact correction method based on thresholding of potential values and excluding very noisy single-trial epochs can be applied prior to the computation of the correlation matrix. (4) Select a few dominant eigenvectors to form the observation model for estimation. The estimated signal subspace must be able to model latency changes for different phase-locked EP components. (5) Estimate EP characteristics with Kalman smoother algorithm. The smoothing parameter can be selected by visual inspection of the estimates (EP image plots), and by considering the expected trial-to-trial variability of individual peaks.

4. DISCUSSION AND CONCLUSION

We presented a new dynamical estimation method for single-trial EP estimation based on a state-space representation for the trial-to-trial evolution of EP characteristics. The method uses the eigenvalue decomposition of the data correlation matrix for the identification of the state-space model. This is an extension of the method presented in [17], where a generic observation model was used. A few dominant eigenvectors obtained from the ensemble measurements incorporate prior information about shape characteristics and within trials correlations of individual EP peaks. This approach takes also into account individual subject characteristics for estimation. Therefore, the method is applicable for different types of EP experiments as long as dynamical behavior from trial-to-trial could be expected. For a Gaussian basis selection like in [17], someone has to select the number of basis vectors and their width. This is not always trivially easy, since a given wave shape may perform in a different way for every individual peak. Therefore, a benefit of SVD is the rather easy selection of observation model that can take into account shape information about different peaks and individual subject characteristics. However, for very weak EPs a generic observation model may have better performance.

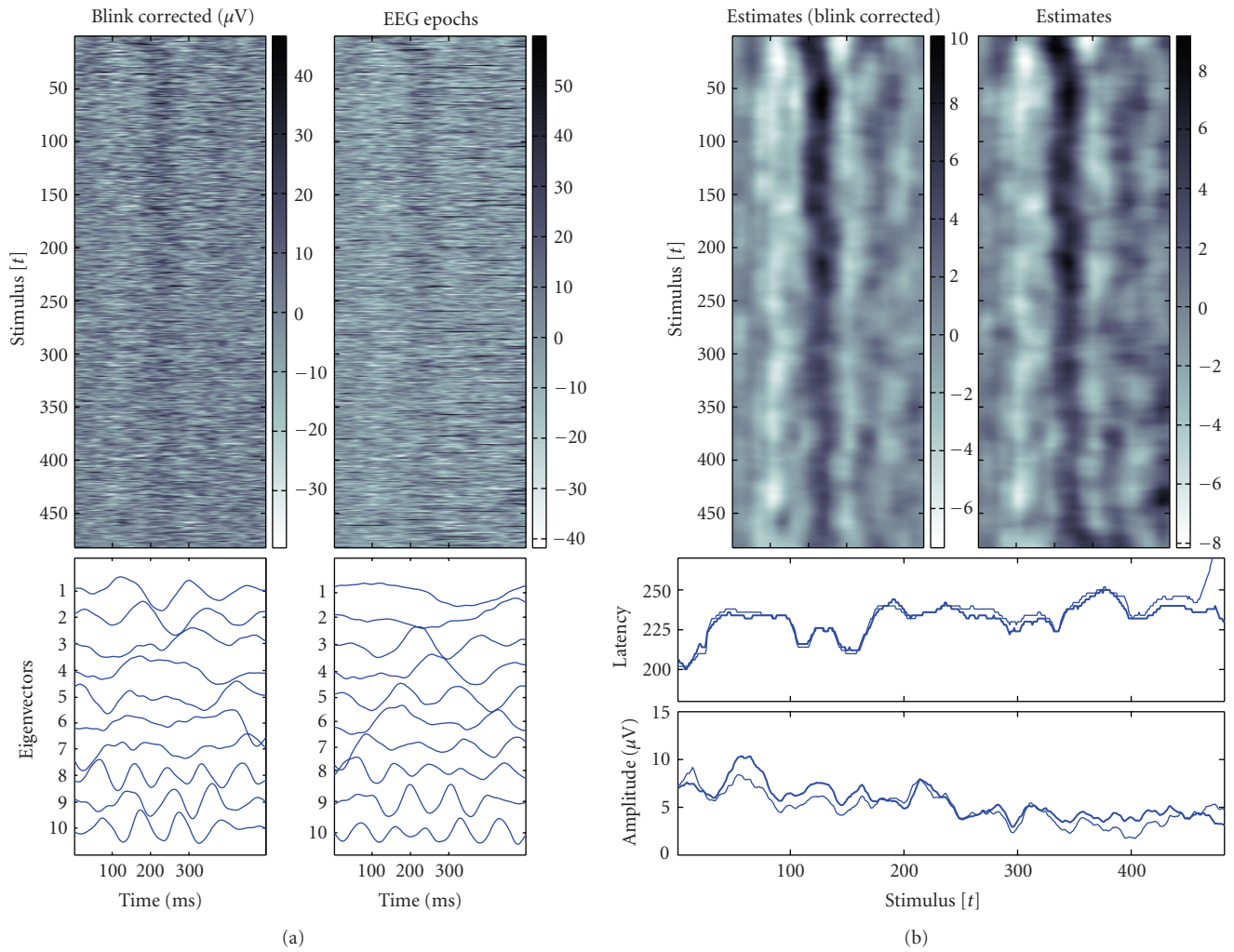


Figure 5: N100/P200 auditory complex, measurements from channel CZ. EEG epochs relative to the standard tones (0–500 milliseconds after auditory stimulation), and obtained estimates. Left: EEG epochs as image plots after and before blink correction (up) and the respective 10 dominant eigenvectors of the data correlation matrix (bottom). The EP images represent stimulus locked stacked epochs (row-by-row). The color-maps describe the amplitude level in μV, y-axis represent successive stimulus or trial t, and the x-axis within a trial latency variation. Right: single-trial estimates as image plots with Kalman smoother (up) based on artifact corrected measurements and original measurements respectively, amplitude and latency estimates of the P200 peak (bottom) based on original measurements (thin) and artifact corrected measurements (bold). For estimation the selection $\sigma_\omega^2 = 10^{-2}$ was used.

Estimates for the state parameters are obtained with Kalman filter and fixed-interval smoother algorithms. Both share the optimality of Bayesian recursive mean square estimation. The fixed-interval smoothing method estimates better the hidden dynamic changes and reduces greater the noise. Therefore, it should be preferred when all the measurements are available. The same behavior can be shown when other observation models are considered, for example, generic basis vectors as in [17]. Therefore, the present paper introduces the use of Kalman smoother algorithm for dynamical estimation of EPs. The use of the filter is appropriate for online estimation. However, compromises between better tracking capabilities and almost online estimation can be searched in terms of fixed-lag smoothing methods [29].

For the demonstration of the methods we used measurements from an auditory experiment (oddball paradigm). Since the aim was to investigate to performance of the methods when strong artifacts exist, we only considered the standard tone measurements and not the deviant and the P300 target response. For this data set the blink artifacts were more prominent for the standard tones. In addition, the estimates of latency and amplitude of the P200 peak (slower and smaller responses towards the end of the measurements) just show that even in ordinary experiments some dynamic behavior from stimulus to stimulus could be expected. However, the method should be addressed to the study of more specific experimental settings. The investigation of latency or amplitude estimates could, for example, be used to study

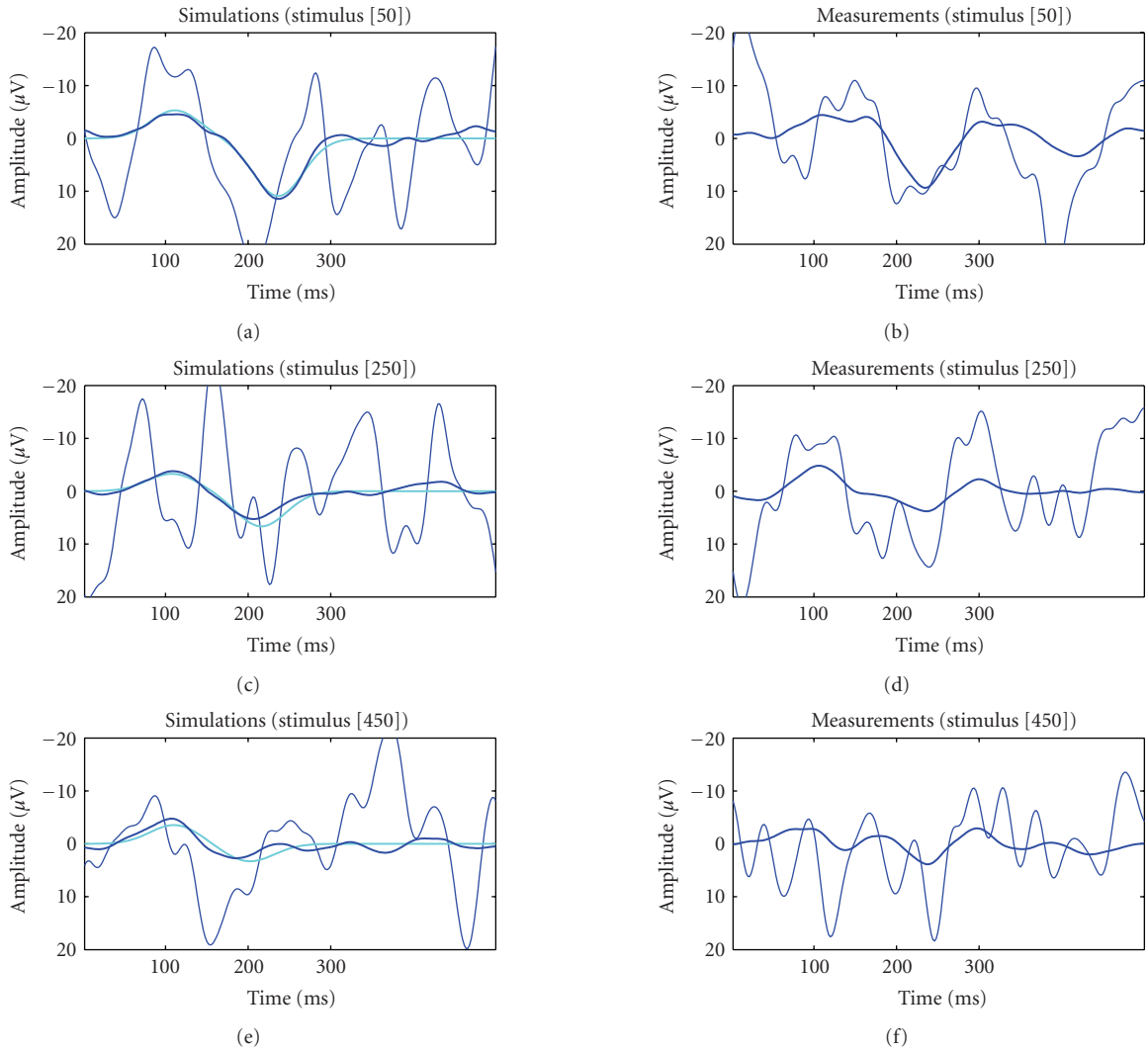


Figure 6: Representative single-trial estimates based on Kalman smoother algorithm. Estimates for the simulations (left) and for the real measurements (right) (standard tones and N100/P200 complex after ocular artifact correction by ICA). Measurements and noisy simulations (dark thin), noise-free simulations (light bold), and estimates (dark bold).

possible habituation effects due to repetition of stimuli, or to study cognitive changes due to time-varying task difficulty or extra distraction. Latency or amplitude changes of peak potentials can also be used to track changes caused by sedative drugs during anesthesia.

EP measurements are usually made with multiple electrodes providing spatial information for the experiment. This information can be used at least to remove artifacts from the signals. We showed by means of ICA that even when the signal subspace is distorted from characteristic artifacts the method is still able to track changes in EP peak components. This is because in the filtering or smoothing procedure phenomena uncorrelated from trial to trial are largely eliminated. In fact, this is exactly the main advantage of dynamical estimation for single-trial EP analysis. However, accurate artifact removal or further elimination of undesirable brain generated components can enable better quality for the signal subspace and individual channel measurements. Extensions to multichannel measurements could be searched by applying the method to each channel separately. Then the variable signal-to-noise ratio conditions from channel to channel should be considered. Another approach could be to direct introduce spatial information in the state-space model. Such multichannel extensions could be investigated for further development of the method. Finally, the signal subspace method can be extended to multichannel measurements. Then it could, for example, be combined with BSS methods.

REFERENCES

[1] E. Niedermeyer and F. Lopes da Silva, *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, Williams and Wilkins, Baltimore, Md, USA, 1999.

[2] J. Intriligator and J. Polich, "On the relationship between background EEG and the P300 event-related potential," *Biological Psychology*, vol. 37, no. 3, pp. 207–218, 1994.
[3] S. Makeig, S. Debener, J. Onton, and A. Delorme, "Mining event-related brain dynamics," *Trends in Cognitive Sciences*, vol. 8, no. 5, pp. 204–210, 2004.
[4] S. Makeig, M. Westerfield, T.-P. Jung, et al., "Dynamic brain sources of visual evoked responses," *Science*, vol. 295, no. 5555, pp. 690–694, 2002.
[5] B. H. Jansen, G. Agarwal, A. Hegde, and N. N. Boutros, "Phase synchronization of the ongoing EEG and auditory EP generation," *Clinical Neurophysiology*, vol. 114, no. 1, pp. 79–85, 2003.
[6] V. Mäkinen, H. Tiitinen, and P. May, "Auditory event-related responses are generated independently of ongoing brain activity," *NeuroImage*, vol. 24, no. 4, pp. 961–968, 2005.
[7] W. A. Truccolo, M. Ding, K. H. Knuth, R. Nakamura, and S. L. Bressler, "Trial-to-trial variability of cortical evoked responses: implications for the analysis of functional connectivity," *Clinical Neurophysiology*, vol. 113, no. 2, pp. 206–226, 2002.
[8] K. H. Knuth, A. S. Shah, W. A. Truccolo, M. Ding, S. L. Bressler, and C. E. Schroeder, "Differentially variable component analysis: identifying multiple evoked components using trial-to-trial variability," *Journal of Neurophysiology*, vol. 95, no. 5, pp. 3257–3276, 2006.
[9] A. Holm, P. O. Ranta-aho, M. Sallinen, P. A. Karjalainen, and K. Müller, "Relationship of P300 single-trial responses with reaction time and preceding stimulus sequence," *International Journal of Psychophysiology*, vol. 61, no. 2, pp. 244–252, 2006.
[10] S. Cerutti, V. Bersani, A. Carrara, and D. Liberati, "Analysis of visual evoked potentials through Wiener filtering applied to a small number of sweeps," *Journal of Biomedical Engineering*, vol. 9, no. 1, pp. 3–12, 1987.
[11] M. von Spreckelsen and B. Bromm, "Estimation of single-evoked cerebral potentials by means of parametric modeling and Kalman filtering," *IEEE Transactions on Biomedical Engineering*, vol. 35, no. 9, pp. 691–700, 1988.
[12] P. A. Karjalainen, J. P. Kaipio, A. S. Koistinen, and M. Vauhkonen, "Subspace regularization method for the single-trial estimation of evoked potentials," *IEEE Transactions on Biomedical Engineering*, vol. 46, no. 7, pp. 849–860, 1999.
[13] R. Quian Quiroga and H. Garcia, "Single-trial event-related potentials with wavelet denoising," *Clinical Neurophysiology*, vol. 114, no. 2, pp. 376–390, 2003.
[14] N. V. Thakor, C. A. Vaz, R. W. McPherson, and D. F. Hanley, "Adoptive Fourier series modeling of time-varying evoked potentials: study of human somatosensory evoked response to etomidate anesthetic," *Electroencephalography and Clinical Neurophysiology*, vol. 80, no. 2, pp. 108–118, 1991.
[15] C. Doncarli, L. Goerig, and P. Guiheneuc, "Adaptive smoothing of evoked potentials," *Signal Processing*, vol. 28, no. 1, pp. 63–76, 1992.
[16] W. Qiu, C. Chang, W. Liu, et al., "Real-time data-reusing adaptive learning of a radial basis function network for tracking evoked potentials," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 2, pp. 226–237, 2006.
[17] S. D. Georgiadis, P. O. Ranta-aho, M. P. Tarvainen, and P. A. Karjalainen, "Single-trial dynamical estimation of event-related potentials: a Kalman filter-based approach," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 8, pp. 1397–1406, 2005.
[18] B. I. Turetsky, J. Raz, and G. Fein, "Estimation of trial-to-trial variation in evoked potential signals by smoothing across trials," *Psychophysiology*, vol. 26, no. 6, pp. 700–712, 1989.
[19] A. Cichocki, R. R. Gharieb, and T. Hoya, "Efficient extraction of evoked potentials by combination of Wiener filtering and subspace methods," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '01)*, vol. 5, pp. 3117–3120, Salt Lake, Utah, USA, May 2001.
[20] H. W. Sorenson, *Parameter Estimation: Principles and Problems*, Marcel Dekker, New York, NY, USA, 1980.
[21] M. Askar and H. Derin, "A recursive algorithm for the bayes solution of the smoothing problem," *IEEE Transactions on Automatic Control*, vol. 26, no. 2, pp. 558–561, 1981.
[22] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA Journal*, vol. 3, pp. 1445–1450, 1965.
[23] G. H. Golub and C. F. van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, Md, USA, 1989.
[24] I. Jolliffe, *Principal Component Analysis*, Springer, New York, NY, USA, 1986.
[25] C. J. James and C. W. Hesse, "Independent component analysis for biomedical signals," *Physiological Measurement*, vol. 26, no. 1, pp. R15–R39, 2005.
[26] A. Cichocki, "Blind signal processing methods for analyzing multichannel brain signals," *International Journal of Bioelectromagnetism*, vol. 6, no. 1, 2004.
[27] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, New York, NY, USA, 2001.
[28] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing, Learning Algorithms and Applications*, John Wiley & Sons, New York, NY, USA, 2002.
[29] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, Springer, New York, NY, USA, 2005.

Hindawi Publishing Corporation
Computational Intelligence and Neuroscience
Volume 2007, Article ID 52609, 8 pages
doi:10.1155/2007/52609

Research Article

Inferring Functional Brain States Using Temporal Evolution of Regularized Classifiers

Andrey Zhdanov,[1,2] Talma Hendler,[1,3] Leslie Ungerleider,[4] and Nathan Intrator[2]

[1] *Functional Brain Imaging Unit, Tel Aviv Sourasky Medical Center, 6 Weizmann Street, Tel Aviv 64239, Israel*
[2] *The School of Computer Science, Tel Aviv University, P.O. Box 39040, Tel Aviv 69978, Israel*
[3] *Psychology Department and Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel*
[4] *Laboratory of Brain and Cognition, National Institute of Mental Health (NIMH), National Institute of Health (NIH), Bethesda, MD 20892-1366, USA*

Correspondence should be addressed to Andrey Zhdanov, zhdanova@post.tau.ac.il

Received 18 February 2007; Accepted 16 July 2007

Recommended by Saied Sanei

We present a framework for inferring functional brain state from electrophysiological (MEG or EEG) brain signals. Our approach is adapted to the needs of functional brain imaging rather than EEG-based brain-computer interface (BCI). This choice leads to a different set of requirements, in particular to the demand for more robust inference methods and more sophisticated model validation techniques. We approach the problem from a machine learning perspective, by constructing a classifier from a set of labeled signal examples. We propose a framework that focuses on temporal evolution of regularized classifiers, with cross-validation for optimal regularization parameter at each time frame. We demonstrate the inference obtained by this method on MEG data recorded from 10 subjects in a simple visual classification experiment, and provide comparison to the classical nonregularized approach.

1. INTRODUCTION

Historically, the goal of inferring person's functional state from brain signals on a single-trial basis was most extensively pursued in the field of EEG-based brain-computer interface (BCI) design [1, 2]. EEG-based BCI systems attempt to distinguish among a small number of consciously controllable mental states from accompanying EEG signals, using the response potential evoked by the stimulus [3, 4]. This approach is often based on machine learning principle using a set of labeled examples to construct a (usually linear) classifier. First BCI experiments utilized a single-trial ERP setup in which subject was presented with stimuli in a controlled fashion and communicated his or her decision by changing mental state (e.g., focus of attention) [3]. Another approach to BCI design attempts to infer subject's mental state exclusively from EEG signals without relying on pacing cues [5–7]. Typically, this free-paced BCIs would split ongoing EEG activity into short (usually less than 1 second) intervals and examine each interval independently in search of EEG patterns, characteristic of one of the predefined mental states.

A wide variety of different algorithms utilizing different features of EEG signal were proposed over the last three decades. The simplest ones like the one described in [8] rely on subjects learning to control their cortical potentials at certain electrode locations, thus reducing the classification algorithm to simple thresholding. More complex algorithms use spatial [9] or spatio-temporal [5–7, 10, 11] features of the EEG signal in conjunction with some classification techniques. Typically, these algorithms treat either raw EEG data or energy of some predefined frequency bands (such as motor-related μ and β rhythms) as features. Those features are then fed into some classifiers to produce the final classification. Most BCIs use a variation of a linear classifier such as regularized fisher linear discriminant analysis (LDA) [5], common spatial patterns [9], or support vector machines (SVM) [12]. Some attempts are also made to address the problem with nonlinear classifiers such as artificial

neural networks [11]. An extended discussion on comparative merits of linear and non-linear methods can be found in [13].

One type of EEG signal features particularly widely used in BCI is the amount of energy in a certain frequency band. Large neuronal populations are capable of generating large-scale synchronized oscillatory electrical activity observable by EEG. As a general rule, the frequency of such oscillatory activity is believed to decrease as the number of neuronal assemblies forming the network increases [14]. This activity is transient and can be evoked (event-related synchronization, ERS) or suppressed (event-related desynchronization, ERD) by various experimental events such as stimulus presentation. Two particular frequency bands—the Rolandic μ rhythm (7–13 Hz) and the central β rhythm (above 13 Hz)—are particularly useful for BCI design as they are amenable to conscious control by means of motor imagery (see [15, 16]). More extensive discussion of the ERS/ERD phenomenon can be found in [4].

Current BCI systems are capable of achieving typical classification accuracies in the range of 80–95% for a two-outcome classification trial (one exception is a report in [17] of 100% classification accuracy over 160 trials).

Recently, application of mental state inference techniques to brain research received a lot of attention from the fMRI community [18–21]. While it has been a valuable tool in investigation of endogenously triggered changes of mental states such as bistable perceptual phenomena, it suffers from low temporal resolution. Unlike fMRI, electrophysiological measurements (EEG and MEG) provide a rich source of temporal information; therefore, it is expected that the analysis of the temporal evolution of these signals can be used for fine temporal mental state inference. While mental state inference from EEG signals has been researched extensively in the BCI context, there is little investigation into EEG- and MEG-based inference as a functional neuroimaging research technique.

To be useful outside the BCI domain, inference techniques need to satisfy a set of requirements that differs significantly from the requirements of the BCI design.

(1) The choice of functional states that need to be distinguished is often outside the experimenter's control.

(2) The subject is not trained to improve the inference accuracy.

(3) The inference techniques need to be applicable to modalities other than EEG. In particular, inferring functional states from MEG or fMRI signals raises two major problems: (a) the dimension of input data is much higher than that of EEG and (b) due to technical and cost limitations, the amount of available data is much smaller.

(4) The inference method attempts to provide a physiologically meaningful interpretation of the inference criteria.

(5) Unlike with BCI, the experimenter has greater control over the experimental environment, making scenarios that require relatively complicated setups (for example, single-trial evoked response potentials (ERPs) experiments) much more attractive.

Figure 1: Examples of the stimulus category presented to the subjects.

These differences require a more high-dimensional and robust classifiers than those used for BCI. In addition, the scarcity of data for MEG and fMRI modalities means that more advanced model validation techniques (such as cross-validation, bootstrapping, etc.) are needed.

In this work, we describe a framework for inference of the temporal evolution of functional states. We formulate the inference problem as that of discriminating between two classes of signals time locked to experimental events. Central concepts of the proposed framework are the temporal evolution of regularized linear classifiers constructed from instantaneous signal values and their relation to the regularization parameter. We investigate the behavior of these quantities on MEG dataset from a simple classification experiment that involves switches between two stimulus categories. We construct a classifier by choosing the combination of time-point and regularization parameter that jointly minimize estimated misclassification rate and analyze the classifier's performance.

2. MEG EXPERIMENTAL SETUP

The MEG experiment was performed on 10 healthy volunteer subjects at the Lab of Brain and Cognition, National Institute for Mental Health (NIMH), Bethesda, Maryland. The study was approved by the Institutional Review Board committee of the NIMH. During the experiment, MEG signals were recorded while subjects were presented with images from two different categories—faces and houses. The images of faces were taken from the Ekman and Friesen [22] and KDEF [23] databases and were composed of 4–6 female or male particulars exhibiting fearful or neutral facial expression (for an example of a particular, see Figure 1). The images were presented in twelve (subjects TE and ZK) or eight (the remaining 8 subjects) 40-second-long epochs separated by 10-second rest intervals of a grey screen with fixation. During each epoch, the subject was presented only with images of faces and houses (no blanks, fixation screens, etc. were used), with the stimulus switching between face and house

Table 1: Number of training samples for each subject.

Subject	CT	ER	FB	JMB	JMM	MC	MKN	SH	TE	ZK
No. of switches from house to face	42	39	47	48	74	65	80	55	57	72
No. of switches from face to house	39	36	46	44	68	61	76	56	53	66

at irregular intervals—approximately every several seconds. The numbers of switches for each subject are summarized in Table 1.

Throughout the experiment, the subjects were requested to fixate at a black point in the center of the screen and report the stimulus category switches by pressing the button corresponding to the category that appeared (i.e., face or house) with the right hand. The MEG experiment used in our study served as a control condition in a larger emotional binocular rivalry experiment.

2.1. Data acquisition and preprocessing

MEG signals were recorded using 275-sensor whole-head CTF-275 system by VSM MedTech Ltd. Coquitlam, Canada. Because of a failure of one of the sensors, only 274 channels were recorded. All the sensors were 2nd-order axial gradiometers. The data was sampled at 600 Hz.

For computational efficiency reasons, the MEG signals were downsampled to 60 Hz. Then they were segmented into intervals of $[-0.33\ 1]$ seconds or $[-20\ 60]$ samples around the stimulus switch. Next, each interval was baseline corrected by subtracting the average of the first 20 samples from each sample in the interval. In this manner for each subject, we obtained several dozens of signals, each containing 274 (number of channels) $*$ 81 (number of time slices) values. Each of the signals was associated with class label "face" if it was recorded while stimulus switched from house to face and with class label "house" otherwise.

3. FISHER LDA-BASED FRAMEWORK FOR FUNCTIONAL BRAIN STATE INFERENCE

In a classical Fisher LDA setup, one is given two sets of scalars, $X = \{x_1, x_2, \ldots, x_n\}$ and $Y = \{y_1, y_2, \ldots, y_m\}$, and the Fisher separation measure is given by

$$d(x, y) = \frac{|\mu_x - \mu_y|}{\sqrt{\sigma_x^2 + \sigma_y^2}}, \tag{1}$$

where μ_x and μ_y are means and σ_x and σ_y are standard deviations of the two sets. The separation measure quantifies the "distinctiveness" of the two sets and can be thought of as signal-to-noise ratio of the associated classification problem.

For two sets of k-dimensional column vectors (representing labeled samples of two classes), $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ and $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m\}$, the direction $\mathbf{p}_f$ in the k-dimensional space that maximizes the Fisher separation between the projections of $\mathbf{X}$ and $\mathbf{Y}$,

$$\mathbf{p}_f = \arg\max_{\mathbf{p}} d(\mathbf{p}^t\mathbf{X}, \mathbf{p}^t\mathbf{Y}), \tag{2}$$

is given by

$$\mathbf{p}_f = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_x - \boldsymbol{\mu}_y), \tag{3}$$

where $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_y$ is the sum of covariance matrices for $\mathbf{X}$ and $\mathbf{Y}$ and $\boldsymbol{\mu}_x$, $\boldsymbol{\mu}_y$—vector means of $\mathbf{X}$ and $\mathbf{Y}$ (see [24] for details). The inversion of $\boldsymbol{\Sigma}$ is problematic when the dimensionality of $\boldsymbol{\Sigma}$ is high and the number of observations is small. In that case, $\boldsymbol{\Sigma}$ is singular or close to singular, due to dimensions where the variance is zero or very small, and the inversion leads to large errors in the estimation of correct values even for dimensions where the variance is large.

Below, we extend this approach to temporal signals and address the singularity of the covariance matrix.

Following the MEG data preprocessing, we obtain a set of labeled signals, each signal being a matrix of 274 channels sampled at 81 consecutive time points (timeslices). Our main goal is to develop a method for inferring correct label from the signal matrix.

We assume a time-point-wise correspondence among the signals (the assumption is partially justified by the fact that the segmentation is timelocked to the stimulus). This assumption implies entrywise correspondence of the signal matrices, allowing us to treat each signal as a point in a $274 * 81$-dimensional feature space. Thus, we can formulate our inference problem as a high-dimensional pattern classification problem.

Such high-dimensional classification problem poses 2 challenges:

(1) feature selection—selecting a small subset of the $274 * 81$-dimensional feature set that is most informative of the signal label.
(2) classifier construction—building robust classifier from the selected feature subset.

3.1. Feature selection

There are many possible strategies for the feature selection step. In this study, we employed a very simple strategy of selecting the set of 274 MEG sensor readings from a single most predictive time-point as a feature set for the classifier construction step (i.e., selecting the most predictive column from the 274 by 81 feature matrix). This reduces the dimension of the data from $274 * 81$ to 274. We evaluate the predictiveness of each timepoint by evaluating the performance of the resulting classifier using 100-fold cross-validation on all the data available.

3.2. Classifier construction

Once a set of 274 features is selected, one needs to construct a classifier for 274-dimensional vectors using a set of several

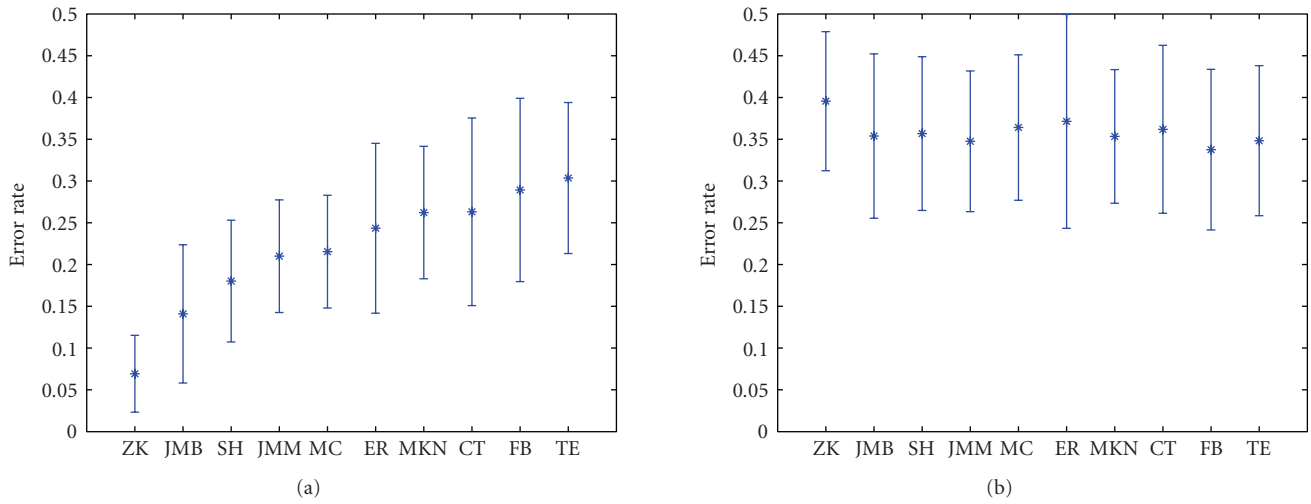


FIGURE 2: (a) Classifier error rates for all 10 subjects; regularization parameter and the input time slice were selected to minimize the classification error using 100-fold cross-validation. (b) Control results obtained using the same algorithm on data with randomly scrambled target labels; both plots show average error estimated using 100-fold cross-validation; error bars denote 1-std-wide margin around the estimate.

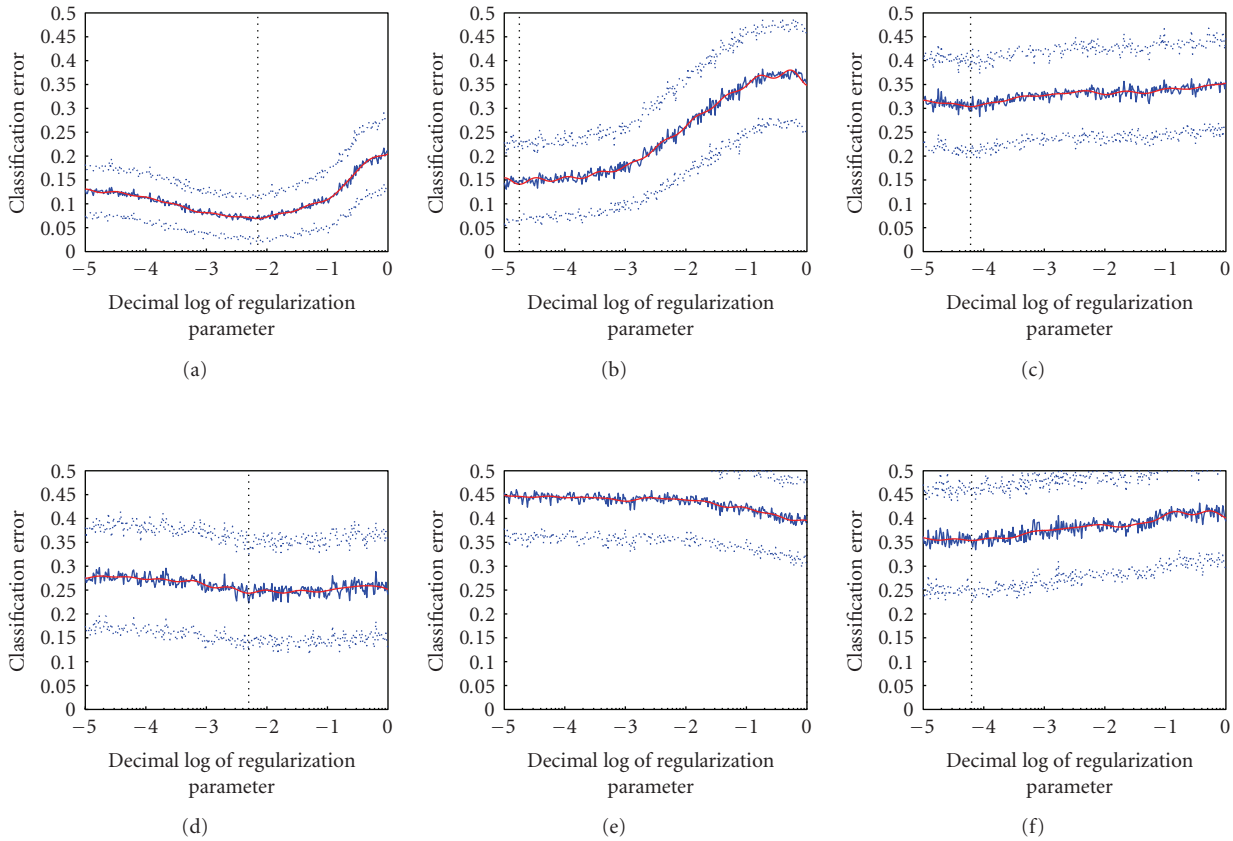


FIGURE 3: Prediction error at the best time slice versus log of regularization parameter. (a), (b) predictable subjects—ZK and JMB. (c), (d) unpredictable subjects—TE and ER. (e), (f) control experiments, in which category labels for subjects ZK and JMB were randomly scrambled before constructing the classifier. Classifier's prediction error was estimated using 100-fold cross-validation on 20% of the data. Dotted lines denote 1-std-wide margins of the estimate. The dotted vertical line marks the global minimum of the smoothed error estimate (smooth red line).

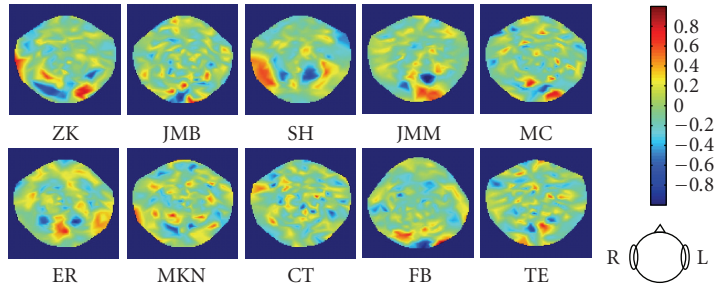


Figure 4: MEG sensor weight maps for the 10 subjects. Each map corresponds to the time slice and the regularization value that yield lowest prediction error estimate for the given subject. The maps are presented in the order of increasing classifier error (from left to right and from top to bottom).

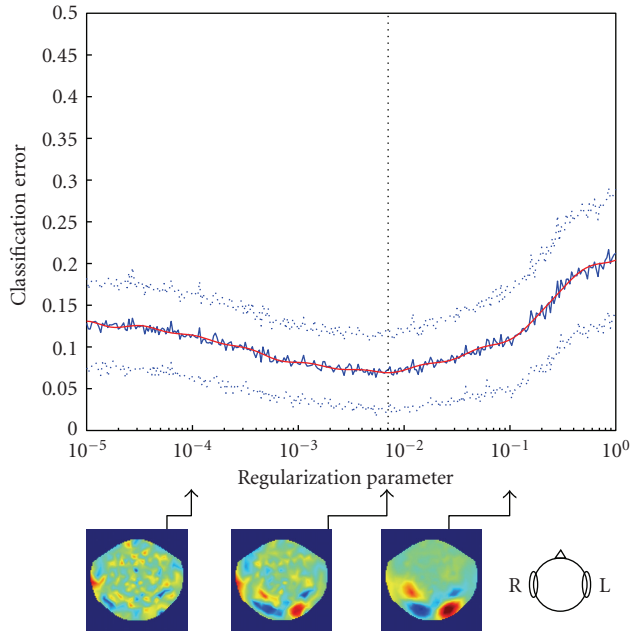


Figure 5: Error rate as a function of regularization parameter for subject ZK. Solid blue line denotes the average error rate over 100-fold cross-validation, dotted lines mark 1-std-wide margin; the vertical line marks the minimum of the smoothed error rate (red line). Three plots below show the distribution of sensor weights corresponding to different values of the regularization parameter.

dozens of labeled examples. We construct the classifier by computing from the labeled examples the optimal projection direction $\mathbf{p}_f$ in the 274-dimensional space using regularized Fisher LDA (see above). A new sample $\mathbf{s}$ is classified by projecting it onto $\mathbf{p}_f$ and applying a simple nearest-neighbor rule: for two classes $\mathbf{X}$ (faces) and $\mathbf{Y}$ (houses), decide that $\mathbf{s}$ belongs to $\mathbf{X}$ if

$$|\mathbf{p}_f^t\mathbf{s} - \mathbf{p}_f^t\boldsymbol{\mu}_x| < |\mathbf{p}_f^t\mathbf{s} - \mathbf{p}_f^t\boldsymbol{\mu}_y| \tag{4}$$

and that $\mathbf{s}$ belongs to $\mathbf{Y}$ otherwise.

Regularization technique

We construct the classifier using Fisher LDA with slightly modified version of regularization described in [25]:

$$\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma} + \lambda e_{\max}\mathbf{I}, \tag{5}$$

where $e_{\max}$ is the largest eigenvalue of the covariance matrix. Normalizing the second term of (5) by $e_{\max}$ allows a heuristic estimation of the relation between λ and the condition number of $\boldsymbol{\Sigma}$. To illustrate this, let us assume that $\boldsymbol{\Sigma}$ is diagonal; in which case, its entries along the main diagonal are its eigenvalues. The condition number c of $\boldsymbol{\Sigma}^*$ is then given by

$$c = \frac{e_{\max} + \lambda e_{\max}}{e_{\min} + \lambda e_{\max}}, \tag{6}$$

where $e_{\min}$ is the lowest eigenvalue of $\boldsymbol{\Sigma}$. Since in our case the number of data samples is less than the data dimension, $\boldsymbol{\Sigma}$ is degenerate and has the lowest eigenvalue $e_{\min} = 0$. Substituting zero for $e_{\min}$ in (6) gives us the relation between λ and the condition number

$$c = \frac{1 + \lambda}{\lambda}. \tag{7}$$

While (7) holds strictly only if $\boldsymbol{\Sigma}$ is diagonal, it can be used for heuristic approximation of c as a function of λ for any degenerate covariance matrix.

3.3. Relationships between λ and time

We argue that relations among λ, timepoint index t, and the classifier accuracy (estimated, e.g., by cross-validation) provide a wealth of information on both statistical and biological aspects of the problem (see the results section). This information can be utilized to guide feature selection, and evaluate data quality and other tasks. The current version of the proposed mental state inference technique uses this information to perform a very simple optimization—it selects the combination of t and λ yielding the lowest prediction error estimate.

The final classification of each signal is performed by doing single timepoint classification using the values of t and λ that minimize the estimated error.

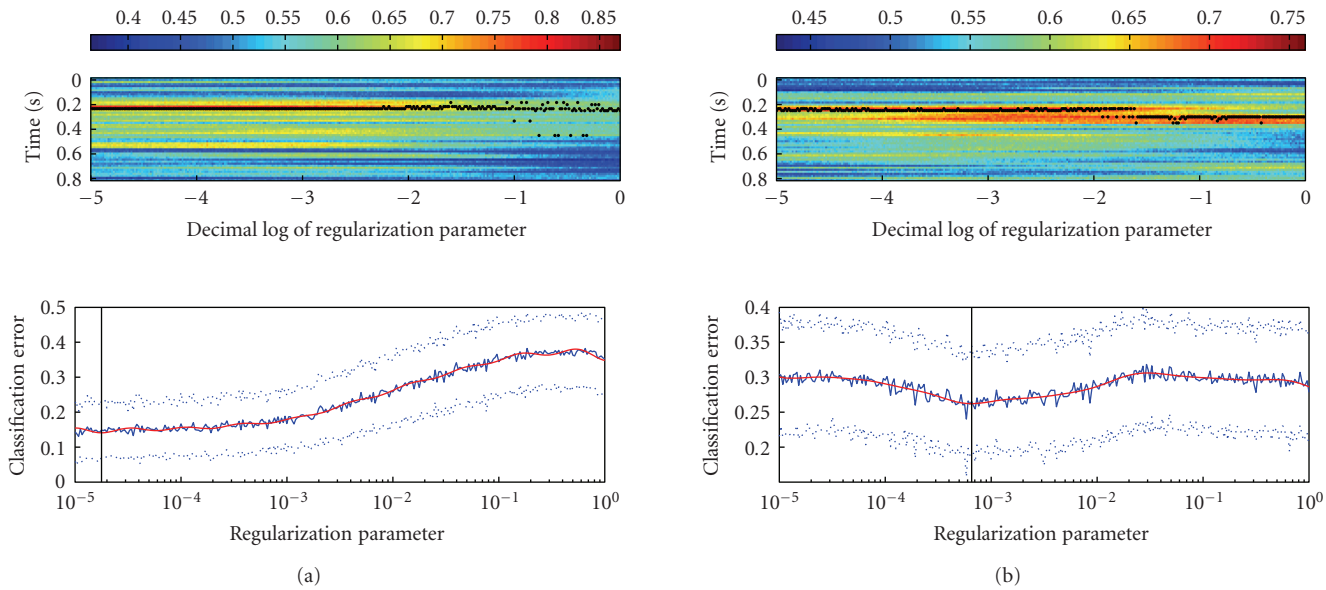


FIGURE 6: (a) Temporal stability of the best separating timeslice as a function of regularization parameter for subject JMB. The upper plot shows the accuracy of the classifier as a function of timeslice and regularization parameter. The accuracy is denoted by the color according to the colorbar above the plot. Timeslice yielding maximum accuracy for each value of the regularization parameter is marked by a black dot. The lower part of the plot shows the best (over all timeslices) error plotted against the regularization parameter using the same timescale as the upper part. (b) Same as (a) but for subject MKN.

3.4. Computational experiments

We estimated the classifier accuracy for each timeslice in the interval $[-0.33\ 1]$ seconds and each value of the regularization parameter $\lambda \in [10^{-5}, 1]$. According to (7), the lower limit of $\lambda = 10^{-5}$ yields regularized matrix $\mathbf{\Sigma}^*$ with condition number of order of magnitude 10^5, which is the largest value for which the computation of the inverse of $\mathbf{\Sigma}+\lambda e_{\max}\mathbf{I}$ is still numerically stable. Using the values from the lower part of the range corresponds to the fixed diagonal regularization proposed in [26]. 300 values of λ were sampled uniformly on the logarithmic scale (i.e., the ratio of the two successive samples was constant) from the interval $[10^{-5}\ 1]$.

For each timeslice and each value of λ, the classifier accuracy was estimated with 100-fold cross-validation using all the data available. In each iteration of the cross-validation, 80% of the data was used for training the classifier and 20% for testing.

4. RESULTS

4.1. Overall error rates

The lowest (over all timeslices and regularization parameter values) error rates achieved for each subject are summarized in Figure 2. Since minimizing the error over any free parameters biases, the error estimate downwards; we compare the estimated error to the estimate obtained by applying exactly the same algorithm to the data with randomly scrambled class labels (see Figure 2(b)). The difference between the mean error estimates is significant for all subjects ($P < 10^{-3}$ for all subjects, estimated using Student's t-test).

4.2. Relation between classifier error and regularization parameter

For a classification problem that uses regularization, one typically expects that the (estimated) classifier error as a function of regularization parameter exhibits a clear global minimum. In our case, the classification error when plotted against the regularization parameter clearly revealed such minimum in some subjects, while in others it remained completely flat (see Figure 3). Subjects that produced such flat plots also tended to achieve lower classification accuracy, which lead us to speculate the convexity of the plot might be indicative of the amount of noise in the data. One might think of the phenomenon in terms of a continuum of different signal-to-noise ratios: the more noise there is in the subject's data, the more similar it is to the random controls, both in terms of minimal achievable error and in terms of convexity of the plot.

4.3. Best separating weight maps

The set weights assigned to the MEG channels by the regularized Fisher LDA analysis can be interpreted as a weight map over the MEG helmet surface indicating the contribution of each point to the classification decision.

We examined the weight maps obtained for the combination of λ and timeslice that yield the lowest estimated prediction error. The maps display a prominent structure consisting of several small clusters of interleaved positive and negative weights (see Figure 4). As expected from animal single unit and fMRI human studies [27], this structure is fairly localized to occipitotemporal regions that might correspond to

a neural source in the fusiform gyrus. The structure seems to be more clearly exhibited in the predictable subjects. We also investigated the relation between the value of λ and the structure of corresponding weight maps. As one could have expected, increasing the regularization parameter causes the resulting optimal weight maps to become smoother (see Figure 5).

4.4. Spatiotemporal structure of the signal and its relation to the regularization parameter

Another item of particular interest is the temporal structure of the signal and its relation to the regularization parameter. We discovered that the stability of the best separating timeslice as a function of regularization and classifier performance as a function of regularization are closely related. The temporal location of the best separating timeslice tends to be more stable for the λ values that yield lower classification error (see Figure 6).

The figure also reveals that the most informative timeslices are located approximately 0.2 seconds after the stimulus switch. This finding is consistent with previous findings about the N170 wave—an increase in negative potential at the parietal parts of the scalp, approximately 0.17 seconds after stimulus presentation [28, 29]. One can also see that there are other timeslices in addition to those located at 0.2 seconds, that can potentially contribute to improved classification (e.g., the timeslices located near 0.32 and 0.5 seconds in Figure 6(b)).

4.5. Comparison to other classification techniques

Finally, we compared regularized Fisher LDA to two other more straightforward techniques: sensorwise difference of average signals for faces and houses and sensorwise difference normalized by sensorwise signal variance (see Figure 7). Note that each classifier attains best separation at a different time. Regularized Fisher linear discriminant differs from the other methods in 3 aspects: (1) it achieves much lower error rate: 14% against 37% and 39% for the other methods; (2) the global minimum of the error function is much more clearly localized in time; (3) the corresponding weights map shows a prominent pattern localized to the sensors located over occipital region of the brain.

4.6. Neuronal basis of the classification

The differential neuronal activity that allows distinguishing between the two types of stimulus switches can be attributed to the differences in visual processing of the stimulus, the differences in the planning and execution of the response motor task, or both. However, observations support the notion that differences in activity detected by the classifier are predominantly of the visual category processing nature. First, the classifier accuracy when plotted as a function of time peaks at about 200 milliseconds which is consistent with other findings regarding the N170 wave and its role in face processing [28, 29]. As expected from N170 distribution, weight maps resulting from the presented classification tend to assign higher importance to sensors located over the occipital and temporal lobes. Finally, behaviorally there was no significant difference between average reaction times for the two stimulus categories suggesting that for both stimulus classes the motor-related neuronal activity is similar.

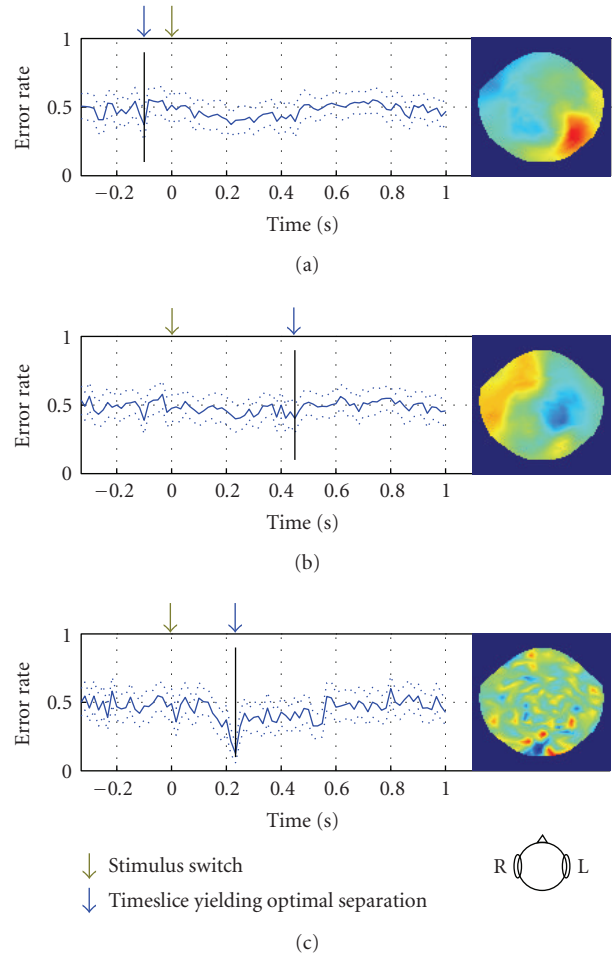


Figure 7: Comparison between different linear discrimination methods for subject JMB. (a) Using sensorwise difference of mean signals for two conditions as weights. (b) Same as (a) but the weight of each sensor is normalized by the variance of the signal at that sensor. (c) Regularized Fisher linear discriminant analysis. The plots depict error estimate of the classifier as a function of time slice of MEG signal to which it was applied. Dotted lines denote 1-std-wide margin around the estimate. The maps depict distribution of weights over the scalp (flattened helmet viewed from above) at the time slice that yields best separation (marked by blue arrow).

5. CONCLUSIONS

We have proposed a new framework for the functional brain state inference problem. The framework utilizes temporal information present in EEG and MEG signals and is particularly adapted to the needs of functional neuroimaging. Application of the framework to MEG data suggests that the relation between regularization parameter and temporal profile of the classifier reveals a lot of structure that can be utilized

for improving classification accuracy. This structure can be exploited to construct more accurate classifiers, for example, by fusing information across different combinations of regularization parameters and times. The proposed classification framework opens a new horizon for whole-brain functional imaging where combined temporal and spatial characteristics of brain signals can reveal the underlying physiological mechanism of an individual's functional state. It can further promote studies on internally driven mental events such as spontaneous switching in awareness, emerging of volition, and formulation of intention.

ACKNOWLEDGMENT

The first and the second authors provided equal contributions to the paper.

REFERENCES

[1] J. J. Vidal, "Toward direct brain-computer communication," *Annual Review of Biophysics and Bioengineering*, vol. 2, pp. 157–180, 1973.

[2] J. J. Vidal, "Real-time detection of brain events in EEG," *Proceedings of the IEEE*, vol. 65, no. 5, pp. 633–641, 1977.

[3] E. Donchin, K. M. Spencer, and R. Wijesinghe, "The mental prosthesis: assessing the speed of a P300-based brain-computer interface," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 174–179, 2000.

[4] G. Pfurtscheller and C. Neuper, "Motor imagery direct communication," *Proceedings of the IEEE*, vol. 89, no. 7, pp. 1123–1134, 2001.

[5] B. Blankertz, G. Dornhege, C. Schäfer, et al., "Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 127–131, 2003.

[6] J. D. R. Millán, F. Renkens, J. Mouriño, and W. Gerstner, "Noninvasive brain-actuated control of a mobile robot by human EEG," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1026–1033, 2004.

[7] J. R. Wolpaw and D. J. McFarland, "Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 51, pp. 17849–17854, 2004.

[8] N. Birbaumer, A. Kubler, N. Ghanayim, et al., "The thought translation device (TTD) for completely paralyzed patients," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 190–193, 2000.

[9] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441–446, 2000.

[10] S. Lemm, B. Blankertz, G. Curio, and K.-R. Müller, "Spatio-spectral filters for improving the classification of single trial EEG," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 9, pp. 1541–1548, 2005.

[11] B. O. Peters, G. Pfurtscheller, and H. Flyvbjerg, "Mining multi-channel EEG for its information content: an ANN-based method for a brain-computer interface," *Neural Networks*, vol. 11, no. 7-8, pp. 1429–1433, 1998.

[12] E. Arbabi, M. B. Shamsollahi, and R. Sameni, "Comparison between effective features used for the Bayesian and the SVM classifiers in BCI," in *Proceedings of the 27th Annual International Conference of the Engineering in Medicine and Biology Society (EMBS '05)*, vol. 7, pp. 5365–5368, Shanghai, China, September 2005.

[13] K.-R. Müller, C. W. Anderson, and G. E. Birch, "Linear and nonlinear methods for brain-computer interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 165–169, 2003.

[14] W. Singer, "Synchronization of cortical activity and its putative role in information processing and learning," *Annual Review of Physiology*, vol. 55, pp. 349–374, 1993.

[15] G. Pfurtscheller and A. Aranibar, "Event related cortical desynchronization detected by power measurements of scalp EEG," *Electroencephalography and Clinical Neurophysiology*, vol. 42, no. 6, pp. 817–826, 1977.

[16] G. C. Pfurtscheller and F. H. Lopes da Silva, "Functional meaning of event-related desynchronization (ERD) and synchronization (ERS)," in *Handbook of Electroencephalography and Clinical Neurophysiology*, G. C. Pfurtscheller and F. H. Lopes da Silva, Eds., vol. 6, pp. 51–65, Elsevier, Amsterdam, The Netherlands, 1999.

[17] G. Pfurtscheller, C. Guger, G. Müller, G. Krausz, and C. Neuper, "Brain oscillations control hand orthosis in a tetraplegic," *Neuroscience Letters*, vol. 292, no. 3, pp. 211–214, 2000.

[18] D. D. Cox and R. L. Savoy, "Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex," *NeuroImage*, vol. 19, no. 2, pp. 261–270, 2003.

[19] Y. Kamitani and F. Tong, "Decoding the visual and subjective contents of the human brain," *Nature Neuroscience*, vol. 8, no. 5, pp. 679–685, 2005.

[20] K. A. Norman, S. M. Polyn, G. J. Detre, and J. V. Haxby, "Beyond mind-reading: multi-voxel pattern analysis of fMRI data," *Trends in Cognitive Sciences*, vol. 10, no. 9, pp. 424–430, 2006.

[21] S. M. Polyn, V. S. Natu, J. D. Cohen, and K. A. Norman, "Category-specific cortical activity precedes retrieval during memory search," *Science*, vol. 310, no. 5756, pp. 1963–1966, 2005.

[22] P. Ekman and W. V. Friesen, *Pictures of Facial Affect*, Consulting Psychologists Press, Palo Alto, Calif, USA, 1976.

[23] D. Lundqvist, A. Flykt, and A. Ohman, *The Karolinska Directed Emotional Faces (KDEF)*, Department of Neurosciences, Karolinska Hospital, Stockholm, UK, 1998.

[24] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, pp. 179–188, 1936.

[25] J. H. Friedman, "Regularized discriminant analysis," *Journal of the American Statistical Association*, vol. 84, pp. 165–175, 1989.

[26] N. Efron and N. Intrator, "Multi-dimensional feature scoring for gene expression data," submitted.

[27] D. L. Sheinberg and N. K. Logothetis, "The role of temporal cortical areas in perceptual organization," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 7, pp. 3408–3413, 1997.

[28] S. Bentin, T. Allison, A. Puce, E. Perez, and G. McCarthy, "Electrophysiological studies of face perception in humans," *Journal of Cognitive Neuroscience*, vol. 8, no. 6, pp. 551–565, 1996.

[29] D. Carmel and S. Bentin, "Domain specificity versus expertise: factors influencing distinct processing of faces," *Cognition*, vol. 83, no. 1, pp. 1–29, 2002.

Hindawi Publishing Corporation
Computational Intelligence and Neuroscience
Volume 2007, Article ID 75079, 13 pages
doi:10.1155/2007/75079

Research Article

Removing Ocular Movement Artefacts by a Joint Smoothened Subspace Estimator

Ronald Phlypo,[1] Paul Boon,[2] Yves D'Asseler,[1] and Ignace Lemahieu[1]

[1] *The Medical Image and Signal Processing (MEDISIP) Group, ELIS Department, Faculty of Engineering Sciences (Firw), Ghent University, The Institute for Broadband Technology (IBBT), Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium*
[2] *Department of Neurology, The Laboratory for Clinical and Experimental Neurophysiology (LCEN), Ghent University Hospital 10K1, De Pintelaan 185, 9000 Ghent, Belgium*

Correspondence should be addressed to Ronald Phlypo, ronald.phlypo@ugent.be

Received 18 February 2007; Revised 25 May 2007; Accepted 21 August 2007

Recommended by Andrzej Cichocki

To cope with the severe masking of background cerebral activity in the electroencephalogram (EEG) by ocular movement artefacts, we present a method which combines lower-order, short-term and higher-order, long-term statistics. The joint smoothened subspace estimator (JSSE) calculates the joint information in both statistical models, subject to the constraint that the resulting estimated source should be sufficiently smooth in the time domain (i.e., has a large autocorrelation or self predictive power). It is shown that the JSSE is able to estimate a component from simulated data that is superior with respect to methodological artefact suppression to those of FastICA, SOBI, pSVD, or JADE/COM1 algorithms used for blind source separation (BSS). Interference and distortion suppression are of comparable order when compared with the above-mentioned methods. Results on patient data demonstrate that the method is able to suppress blinking and saccade artefacts in a fully automated way.

1. INTRODUCTION

Recording of cerebral activity by means of the electroencephalogram (EEG) is a widespread technique that is well embedded in today's healthcare environment. The potentials recorded at the patient's scalp are a direct reflection of cerebral activity patterns and thus may serve as an indication to neurological diseases such as epilepsy, encephalopathies and sleep disorders. Being a noninvasive technique with a high temporal resolution, it is also frequently used in experimental settings in neurophysiology and psychology, where responses to external stimuli are measured.

Although the first article mentioning EEG registration already dates from 1924 (Hans Berger), there still remain a lot of side effects, inherent to the recording, that are to be dealt with. The major issue to be tackled when preprocessing the EEG is the contamination of the signal by artefacts. The latter hamper the interpretation by physicians of the cerebral activity, since they are often many times larger in amplitude than the neuronal activity of interest. The most well-known interfering sources are power line noise, muscle activity, and ocular movements. The classical frequency bands of interest in the EEG are situated between half a Hertz (delta band lower limit) and approximately 35–40 Hz (gamma band upper limit), although studies are found where in upper limits of gamma activity in evoked potentials extend beyond the classical 35–40 Hz limits up to 80 Hz [1]. Even scalp effects up to 200 Hz have been recorded as reported in [2], tagged as oscillation in the high gamma frequency band (60–200 Hz) while test persons imagined they were singing. In general, though, power line noise is easy to deal with since the main spectral bands of interest are usually limited between 0 Hz and 35 Hz. Because in these cases there is no spectral overlap with the EEG bands of interest, a simple low-pass or Notch filter with cancellation at 50 or 60 Hz suffices for the elimination of this artefact. Muscle artefact suppression is harder to resolve since the frequencies are situated in the upper part of the EEG spectrum. Moreover, the activity of the muscles stems from fast changing polarization of different muscle fibers, displaying undeterministic (low autocorrelation) behaviour. From this point of view, De Clercq et al. [3] proposed to make use of the canonical correlation analysis method to reduce the influence of muscle activity in EEG recordings. For ocular movement artefact suppression

several solutions have already been put forward. Nevertheless, we will focus on the latter, showing where previous techniques fail and can thus be ameliorated and how to validate these studies based on objective measures derived from simulated data.

Whenever ocular movements are present in the EEG, the underlying cerebral activity cannot be interpreted by the physician, the experienced EEG technician, or the automated file processor. In the past, many solutions have been presented to suppress artefacts as much as possible. One of the earliest techniques was to request the patients or subjects to move their eyes as little as possible in order to obtain a nice and clean EEG recording. However, working with children and disabled people, which are still the main groups in clinical settings, seemed quasi infeasible. Besides, the current trend is shifting more and more to long-term monitoring, a set up in which it is infeasible to ask the subjects to avoid moving their eyes during recording. Also, when considering task related potential distributions through event related potentials, the creation of an additional task (staring) decreases the amplitude and the visibility of the cerebral response in the grand average related to the primary task, see, for example, [4–6]. Therefore, rejection of trials or segments contaminated with ocular artefacts has been put forward as an alternative solution. Although very commonly used in the past, this technique suffers from huge data losses since spontaneous blinks occur at a ratio of approximately 20 per minute [7]. For statistical tests based on long-term recorded data, this loss of data is highly unwanted, since the statistical results would suffer from the absence of the discarded data epochs. The same holds for examinations of seizure onsets in long-term recordings of epileptic patients, a period that may be heavily contaminated by muscular and ocular artefacts [8]. Moreover, for some ERP processing the blinking of the eyes are temporally highly correlated with the cerebral response of interest. Hence, this would result in an unwanted rejection of epochs of interest. [9].

For these reasons and because of increasing computational resources, the trend is shifting toward artefact correction methods. Although correction would be beneficial and preferable to rejection for the reasons mentioned above, it is claimed that rejection is still to be preferred over correction in EEG recordings of children [10]. Some of the more widely spread correction methods in research and clinical uses are the temporal and spectral regression techniques [11–13] and source separation or extraction methods [14, 15]. The regression methods, differentiated based on details in the implementation, all start from a set of reference signals and calculate the weighted contribution of each of those references at the recording sites or electrodes [12, 13, 16]. Although these methods have proven to be able to cope with different eye movement artefacts in the EEG, the major drawback is still the cross contamination between the reference electrooculogram (EOG) channels and the EEG channels. It cannot be guaranteed that the EOG is free from contributions of cerebral origin, because propagation takes place in the same way as the ocular potentials influence the scalp potentials or EEG. This can be simply explained by the electrical property of reciprocity. Therefore, it is required that the reference used to perform regression with is well chosen and appropriately preprocessed. Recent studies [16, 17] showed that there exist means to tackle this problem by using adapted versions of the EOG channels. However, the validity of these results in clinical data is questionable, since the findings are based on a construction model that is equivalent to the correction process, hence biasing the outcome toward the presented method.

Apart from these regression techniques, a lot of research has been done on blind source separation (BSS) models. The EEG, being narrowband potential measurements that are the resultant from current sources in the brain, can be described by the linear approximation of the Maxwell equations for volume conduction. This also implies that the measurements are an instantaneous reflection of the underlying activity, and thus no delays should be considered. The general linear and instantaneous mixture model is given by

$$x(t) = \mathbf{A}\mathbf{s}(t) + \eta(t), \tag{1}$$

where $x(t) \in \mathbb{R}^M$ are the measured data from m electrodes sampled at time instance t, $\mathbf{s}(t) \in \mathbb{R}^N$ are the n sources, at time instant t, $\mathbf{A} \in \mathbb{R}^{M\times N}$ is the linear mixing matrix, and $\eta(t)$ is the additive noise. The ith column of $\mathbf{A}$, $\mathbf{a}_i$, is a measure for the spreading of the activity of the ith source in $\mathbf{s}$, s_i, to the scalp electrodes, that is, the so-called source topography. Since $x(t)$ and $\mathbf{s}(t)$ can be seen as random samples sampled independently from a multivariate distribution, we omit the time index t throughout the subsequent work, with explicit usage only there where needed to interpret the variables or the equations they are involved in. The additive noise in (1), when not negligible, will be considered as one of the sources in the subsequent work, and thus the term $\eta(t)$ can and will be dropped.

In the midtwentieth century, the most commonly used model to solve for the estimated sources (and their corresponding topographies), given only the measurement data x, was that of principal component analysis (PCA) [18], a technique based on the well-known singular value or eigenvalue decomposition (SVD, resp., EVD, also known as the Karhunen-Loève transform in information theory). These methods try to estimate underlying sources, based on maximisation of variance in a decorrelation framework for the sources and their topographies. However, many researchers have pointed out that the constraint of orthogonality on the source topographies does not stroke with the reality of the physiological sources underlying the EEG. Indeed, there is no ground on which we should assume that physiological source topographies are mutually orthogonal. Nevertheless, some interesting results based on these assumptions have been published recently, for example, [16, 19, 20].

In the last decade, independent component analysis (ICA) has become a popular technique to decompose the EEG signal into cerebral and noncerebral source estimates. The extra assumption of maximal source independence has found a lot of support in the EEG research community. In contrast to PCA/SVD, the sources no longer need to have mutually uncorrelated topographies, meaning the topographies are no longer constrained to be orthogonal. To solve for the linear and instantaneous mixture model subjected to the

constraint of maximal independence, the source estimates $\hat{s}_i$ for the sources s_i in (1) can be obtained through maximisation or minimisation of an appropriate object function, called a contrast. There exist a lot of contrasts in literature that one could optimise for in order to obtain maximally independent sources. Among the most popular are those of Hyvärinen and Oja [21] (Kurtosis based with a nonlinearity in the updating function), Lee et al. [22] (maximum likelihood based), Belouchrani et al. [23] (joint diagonalisation over a specific set of matrices) and Comon [24] or Cardoso and Souloumiac [25] (both cumulant tensor based). The first application of ICA to solving the EEG problem came from Makeig et al. [14] where they attempted to separate the raw EEG signal into physiological sources. An extension thereof was given in [15]. The research presented in the latter has been based on patient data as well as simulations and shows some of the abilities of ICA in EEG applications. In spite of the use of PCA/SVD in the preprocessing step of ICA implementations, the so-called whitening, PCA/SVD itself is often regarded as an inferior BSS algorithm compared to ICA.

In this paper, we show that by careful selection of a parameter set we can use the benefits of SVD, namely, a limited number of samples needed to estimate the covariance matrices with a sufficient precision, to our advantage in EOG source interference suppression. Although the results obtained by a such decomposition suffer a lot from methodological artefacts. Therefore we introduce an additional estimation step carried out by ICA (we used JADE [23]) and merge both results through a joint smoothened subspace estimator. The method of Canonical Correlation Analysis (CCA) [26] is an excellent candidate method for the latter, since the linear combination between both subspaces is automatically calculated and prevents us from introducing prefixed weighting scalars. Moreover, the subspace estimator reduces the subspace to the smoothest components only. For objective validation we propose a dipole-based model with eye activity modelling based on the model described in [7].

For ocular artefact suppression, other methods, such as wavelet transforms [27], the use of neural networks [28], and advanced filtering techniques [29], have been proposed. However, they form a minority and the reports on their success (or failure) are not much discussed in literature. We like to inform the reader that the given list of methods is certainly not exhaustive, but the given background should suffice to demonstrate the weaknesses in the current methods and to support the strategy we opted for.

2. MATERIALS AND METHODS

2.1. Materials

The patient data were collected at the Laboratory for Clinical and experimental neurophysiology (LCEN), at the Ghent University Hospital (Ghent, Belgium). Data were used from ten patients. The EEGs were recorded using a Telefactor Beehive system at a sample rate of 200 Hz. Twenty-one electrodes were placed on the patients' heads according to the 10–20 international system, together with six electrodes covering the lower temporal regions. One patient, showing numerous eye movement artefacts, was chosen to display the results on patient data, while the others were used for resampling as discussed below in Section 2.2.

In the rest of the paper, we consider the EEG as a 27-channel recording as defined above. Nevertheless, the presented method is valid for all average reference-based recordings with a reasonable number of channels M, including frontal channels.

2.2. EEG simulation model

The use of patient data has an enormous drawback in that there exists no way to qualify the performance of the algorithms, except for subjective scoring by physicians or experienced EEG readers. The latter are often ambiguous and suffer from interscorer variability [30]. We therefore propose to use a simulation model that simulates EEG data, given certain patient specifications. The method consists of building a dataset based upon forward modelling of dipoles. The calculation of the electrical field created by randomly activated cortical dipoles in a three-layer spherical head model (brain tissue, skull, and scalp) gives rise to the potentials at the respective electrodes. Each dipole is chosen to have a radial orientation, since the cortical activation patterns are known to be perpendicular to the cortical layers, a result of the physical layout of the pyramidal cortical cells [31]. The latter are then filtered according to spectral statistics derived from the patient data. An additional ocular dipole is added following the model in [7]. To add a waveform mimicking an eye blink we simulate the electrical shortcut that is created by the closure of the eyelid by a Gaussian bell curve and the eyeball rotation by a rotating dipole. An example of the resulting potentials due to these eye movements are given in Figure 1.

The strength of the model lies in the separate modelling of the activity of interest $\mathbf{e}$, which is actually the artefact source, and the background activity $\mathbf{b}$. By consequence, any method that has as a goal to separate the activity of interest from the background EEG can be evaluated using this model in combination with appropriate measures. The most simple measure would be the direct comparison between the background $\mathbf{b}$ and its estimate $\hat{\mathbf{b}}$ returned by the method under investigation. The difference, which can also be expressed as a function of an estimate of the activity of interest $\hat{\mathbf{e}}$, may be given as the average sample distance asd:

$$\text{asd} = \frac{1}{M}\sum_{i=1}^{M} \mathrm{E}\left\{\sqrt{(\mathbf{b}-\hat{\mathbf{b}})^2}\right\} = \frac{1}{M}\sum_{i=1}^{M} \mathrm{E}\left\{\sqrt{(\mathbf{e}-\hat{\mathbf{e}})^2}\right\}, \quad (2)$$

where $\mathbf{b}$, $\mathbf{e}$, $\hat{\mathbf{b}}$, and $\hat{\mathbf{e}} \in \mathbb{R}^M$ are the samples of the background or the activity of interest used either by the simulation model or estimated by the algorithm under investigation, respectively. This measure, although attractive because of its simplicity, does not reveal any details about the source of error. Therefore we will turn to more sophisticated error measures to compare different methods and their performance, see 2.8. However, the asd measure will further on be used to estimate appropriate parameter settings for the pSVD method, as discussed in 3.1 and for evaluation of the algorithms under noise and relative scaling of b with respect to e.

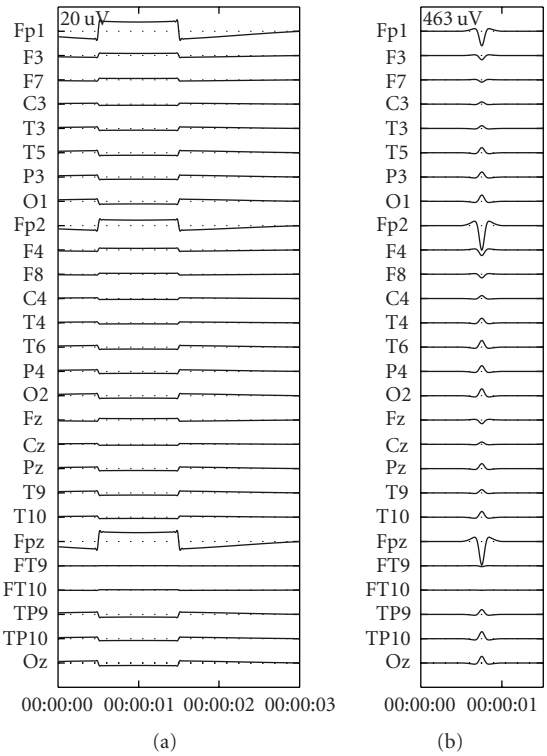


Figure 1: The potentials resulting from the simulated eye movements for horizontal saccade (a) and blinking (b).

To have a measure of performance in different scenarios, we define the SNR level as the mean ratio between the background signal and the eye movement at the samples where contamination occurs. Consider the set of all time instances t in the observed window $\mathcal{W}$, we can take a subset containing only artefact contaminated samples $t \in \mathcal{T} (\subset \mathcal{W})$. The definition of SNR is then given as follows:

$$\mathrm{SNR} = 10 \log_{10} \frac{\|\mathbf{e}_{t\in\mathcal{T}}\|_F^2}{\|\mathbf{b}_{t\in\mathcal{T}}\|_F^2}, \tag{3}$$

where $\|\cdot\|_F$ denotes the Frobenius norm of the observation matrix.

2.3. BSS

The model that is used in the linear BSS framework is based on a direct mixture model:

$$x = \mathbf{A}\mathbf{s}, \tag{4}$$

where the measurements $x \in \mathbb{R}^M$ are linear combinations of the sources $\mathbf{s} \in \mathbb{R}^M$ through a mixture $\mathbf{A} \in \mathbb{R}^{M\times M}$. We assume an equal number of sources with respect to the measurements. In the latter case, the inverse of $\mathbf{A}$ exists whenever $\mathbf{A}$ is full column rank. The aim of BSS algorithms is to find an estimate of $\mathbf{s}$, $\hat{\mathbf{s}}$, by estimating the unmixing matrix $\mathbf{W}$. In the ideal case the matrix $\mathbf{W}$ would be equal to $\mathbf{A}^{-1}$, and thus the source estimates $\hat{\mathbf{s}}$ would equal the original sources $\mathbf{s}$ in the mixture. However, in most cases the estimated sources are only an approximation to the real sources and it is up to the user to find out whether the approximation is sufficient for his application. These techniques are termed blind since they only use the available data x as prior information, although some authors suggest to call databased source separation techniques semi blind since one always has to start from some additional basic assumption(s) [32], see below in Sections 2.3.1 and 2.3.2. However, to avoid confusion with the field of communications, where semi blind is used for methods where some parts of the source signals are known *a priori*, we dissuade the use of the terminology in any circumstances where the latter is not the case.

2.3.1. Piecewise SVD

In a decomposition based on SVD, the additional basic assumption is the decorrelation or linear independence of the sources as well as of the topographies while pursuing maximal variance of the estimated sources. The temporal and spatial decorrelations are derived from the left and right correlation matrices of $\mathbf{X}$, where $\mathbf{X}$ is the stacking of all samples $x(t)$ in a rectangular window of size T. If we denote by t_0 the first sample index of the window, then $X = [x(t_0)x(t_0 + 1)\cdots x(t0 + T - 1)]$. The decomposition of the data is given as

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \tag{5}$$

with $\mathbf{X} \in \mathbb{R}^{M\times T}$ the measurements, $\mathbf{U}$ and $\mathbf{V}$ the eigenvectors of $\mathbf{X}\mathbf{X}^T$ and $\mathbf{X}^T\mathbf{X}$, respectively, and a matrix $\mathbf{\Sigma}$ containing the therewith associated singular values on its diagonal. The columns of $\mathbf{U}$ can be seen as the (mutually uncorrelated or orthogonal) source voltage distribution maps at the electrodes known as topographies. The columns of $\mathbf{V}$ are the (mutually uncorrelated or orthogonal) source activations and the ith singular value σ_i on the diagonal of $\mathbf{\Sigma}$ is a measure for the explained variance by the corresponding i-th source in the original measurement data $\mathbf{X}$. The sources are ordered according to their nondecreasing associated σ_i with increasing index i.

When using the SVD in our source separation model we will use it in a sliding window of $T = 32$ samples (160 milliseconds), moving with 8 samples per window position. The chosen windowing parameters are justified in 3.1. For each window d the SVD decomposition is calculated and the source topography $\mathbf{u}_1$ associated with the source $\mathbf{v}_1$ with maximal variance (σ_1) is checked upon Criterion 1.

Criterion 1. The signal $\mathbf{X}$ is deflated by a topography $\mathbf{u}_1(d)$ iff $\arg\max_j c_j = \mathrm{abs}(\mathbf{topo}_j^T \mathbf{u}_1(d))/(|\mathbf{topo}_j \| \mathbf{u}_1(d)|) \geq 0.6$.

In the above criterion, deflation of $\mathbf{X}$ is performed by setting the corresponding $\sigma_1(d)$ to zero in the reconstruction (cf. (5). The template library containing the vectors $\mathbf{topo}_j$ is given in Table 1. This library is build from vectorially transcribed versions of the descriptions of spatial maps associated to ocular activity as can be found in [33]. The topographies in the table are reduced to the affected electrodes only, unmentioned electrodes are set to zero in the reference spatial

Table 1: The template library composed of column vectors (here transposed and limited to their core information) as used in Criterion 1.

	Fp1	Fp2	Fpz	F7	F8	FT9	FT10	F3	F4	C3	C4
$\mathbf{topo}_1$	1	1	0.5	0	0	0	0	0	0	0	0
$\mathbf{topo}_2$	1	0	0	0	0	0	0	0	0	0	0
$\mathbf{topo}_3$	0	1	0	0	0	0	0	0	0	0	0
$\mathbf{topo}_4$	0	0	1	0	0	0	0	0	0	0	0
$\mathbf{topo}_5$	0	0	0	1	0	0	0	0	0	0	0
$\mathbf{topo}_6$	0	0	0	0	1	0	0	0	0	0	0
$\mathbf{topo}_7$	0	0	0	1	-1	1	-1	0	0	0	0
$\mathbf{topo}_8$	1	1	1	0	0	0	0	1	1	1	1

maps. The first four topographies in the library contain maps generally associated with blinks, while the last four describe the eye gazing and horizontal movements.

Criterion 1 is no more than thresholding the subspace correlation [34] between the first component in $\mathbf{U}(d)(\mathbf{u}_1(d))$ and the template library composed of $[\mathbf{topo}_1\mathbf{topo}_2\cdots\mathbf{topo}_8]$ as given in Table 1. The fuzziness included in the threshold (0.6) is due to the generality of the library and the mismatch between the true correlation and the estimated calculated correlation from an SVD based on 32 samples only. The short time windows are chosen as such as to cope with the nonstationarity of the EEG that is caused by the waxing and waning of sources in the background. The window of 32 samples or 160 milliseconds is a tradeoff between the oscillatory processes of 80–100 milliseconds [35] and a sufficient sample size for the SVD calculation, see Section 3.1 for more details. The reconstructed EEG with the locally deflated subspaces is then calculated for the first 8 samples of window d, given the results calculated for the windows $d-3\cdots d-1$, as the mean of these local reconstructions:

$$\mathbf{X}^D_{i\cdots i+7} = \frac{1}{4}\sum_{j=0}^{3}\mathbf{U}(d-j)\mathbf{\Sigma}^D(d-j)\mathbf{V}_{i+8j+1\cdots i+8(j+1)}(d-j), \tag{6}$$

where $\mathbf{\Sigma}^D$ denotes putting the first eigenvalue on the diagonal of Σ to zero when required so by the deflation criterion (Criterion 1). The i-th sample is the first sample of window d. The final result $\mathbf{X}^D$ is then the concatenation of all the subwindows d that are corrected as above in (6). From here on we will call this deflation method piecewise SVD (pSVD) [20] referring to the window per window deflation approach.

2.3.2. ICA

ICA algorithms try to find a decomposition based on the constraint of maximal statistical independence between the sources. In EEG, we may assume the model to be linear and instantaneous as given in (1) if the sources are stationary during the observation. Assume for a while that this constraint has been met. The maximal statistical independence of the sources is then a weaker constraint than the one used in the SVD, in the sense that there are no assumptions made about the topographies. It involves also an indeterminacy concerning permutation and scaling of the sources [24]. For this reason we cannot rely on the ordering of the components in the decomposition nor on the variances of the estimated sources for our selection criteria. Hence, we test all the component topographies of the decomposition against the decision rule in criterion 1. When testing all the topographies upon their subspace correlation with the template library we might find that multiple components obey the Criterion. In that case the data gets deflated by the subspace containing all these components. This can be done in the reconstruction of x (4) by setting all columns in $\mathbf{A}$ associated to the sources in s containing ocular activity (according to Criterion 1) to zero. The obtained cleaned dataset will subsequently be called $\mathbf{X}^D_{\mathrm{ICA}}$.

ICA decompositions were taken from windows of 2000 samples or 10 seconds at 200 Hz sampling rate. This is a sufficient tradeoff between the nonstationarity and the samples needed to obtain an appropriate decomposition [36]. We here use the algorithm of JADE/COM1 [24, 37] because of its stability and its statistically robust approach. The algorithm does not suffer from initialisation, nor from parametrisation issues.

2.4. Joint smoothened subspace estimation

In a last step, both the estimates of the ocular components, $\mathbf{X}^D_{\mathrm{SVD}}$ and $\mathbf{X}^D_{\mathrm{ICA}}$, are fed to a joint smoothened subspace estimator (JSSE). The algorithm that lends itself best to calculate the joint smoothened component(s) is the canonical correlation analysis (CCA). Let the subspaces estimated by the SVD and the ICA algorithm be $\mathcal{Y}$ and $\mathcal{Z}$, respectively. Since there is no linear component from the pSVD algorithm that can be estimated (the general mixing matrix is nonexisting, it is a chain of short time linear mixtures), we use the piecewise back projected versions of the components onto the original EEG subspace, that is $\mathbf{y} = x - x^D_{\mathrm{SVD}}$, to represent the subspace $\mathcal{Y}$. If we take as a basis for $\mathcal{Z}$, $\mathbf{z} = x - x^D_{\mathrm{ICA}}$, we can calculate a common, smoothened component for the subspaces $\mathcal{Y}$ and $\mathcal{Z}$ by calculating the CCA for the joint subspace. For this, we proceed as follows (see also [26, 34] for more details on the canonical correlation analysis and calculation of angles between subspaces).

(1) Take the QR decomposition of the joint signal space of the stacked matrix $\mathbf{P} = \binom{\mathbf{Y}}{\mathbf{Z}}$, where $\mathbf{Y} = [\mathbf{y}(t_0)\mathbf{y}(t_0+1)\cdots\mathbf{y}(t_0+T)]$ and $\mathbf{Z} = [\mathbf{z}(t_0)\mathbf{z}(t_0+1)\cdots\mathbf{z}(t_0+T)]$. The joint signal subspace can be found by taking the SVD of $\mathbf{P} = \mathbf{U}_s\mathbf{\Sigma}_s\mathbf{V}_s^T$ and truncating at a noise level of 1%, that is $\mathbf{P}^D = \mathbf{U}_s\mathbf{\Sigma}_s^D\mathbf{V}_s^T$, where $\mathbf{\Sigma}_s^D$ describes the signal subspace formed by retaining only the p highest eigen values which acumulate to 99% of the total energy and where the relative energy is calculated as $\sum_{i\in\mathrm{Set}}\sigma_i^2/\sum_{i=1}^{M}\sigma_i^2$. The QR decomposition of these sources results in $\mathbf{V}_s^T = \mathbf{Q}_0\mathbf{R}_0$. Repeating this for a time delayed version of $\mathbf{P}$ gives $\mathbf{V}_s^T(\tau) = \mathbf{Q}_\tau\mathbf{R}_\tau$, where both $\mathbf{Q}_0, \mathbf{Q}_\tau \in \mathbb{R}^{N\times N}$ are orthogonal and both $\mathbf{R}_0, \mathbf{R}_\tau \in \mathbb{R}^{N\times M}$ are quasi upperdiagonal. τ is taken one sample period.

(2) Calculate the SVD of $\mathbf{Q}_0^T\mathbf{Q}_\tau = \mathbf{U}_q\mathbf{\Sigma}_q\mathbf{V}_q^T$.

(3) The joint and smoothened component is then found by taking $\sigma_{1\cdots j}\mathbf{U}_{q,1\cdots j}^T\mathbf{R}_0$, where $\sigma_{1\cdots j}$ denote the first j entries in the diagonal of $\mathbf{\Sigma}$ put on the diagonal in a $j \times j$ matrix and $\mathbf{U}_{q,1\cdots j}$ are the first j columns of $\mathbf{U}_q$.

The number of components j that are to be retained depends on the angles between the estimated components and the two subspaces. The cosine of these angles are given in descending order on the diagonal of Σ. To describe the common subspace between both $\mathcal{Y}$ and $\mathcal{Z}$, it suffices to limit the number of components to the ones that are exceeding a subspace correlation of 0.9.

To find the source contribution in the original dataspace, we have to calculate back each of the estimated source contributions. The estimated source can be expressed as

$$\begin{aligned}\hat{\mathbf{s}} &= \mathbf{U}_{q,1}^T\mathbf{Q}_0^T \\ &= \mathbf{U}_{q,1}^T\left(\mathbf{V}_s^D\mathbf{R}_0^{-1}\right)^T \\ &= \mathbf{U}_{q,1}^T\left(\mathbf{R}_0^{-1}\right)^T\mathbf{V}_s^{D^T} \\ &= \mathbf{U}_{q,1}^T\left(\mathbf{R}_0^{-1}\right)^T\left(\mathbf{\Sigma}_s^D\right)^{-1}\mathbf{U}_s^T\mathbf{P}.\end{aligned} \tag{7}$$

From (7), it can be seen that the source can be expressed as a function of $\mathbf{Y}$ and $\mathbf{Z}$, with a mixing matrix that is equal to $\mathbf{U}_{q,1}^T\left(\mathbf{R}_0^{-1}\right)^T\left(\mathbf{\Sigma}_s^D\right)^{-1}\mathbf{U}_s^T$. The expression of $\mathbf{Y}$ as a function of the original data is piecewise linear which makes the resulting sources in $\hat{\mathbf{s}}$ a weighted sum of a piecewise linear mixture (pSVD) and a completely linear mixture (JADE) of the original data in x.

2.5. Reconstruction phase

To have an adequate reconstruction of our artefact free data, we need to deflate our original dataspace by the projection of both subspaces $\mathcal{Y}$ and $\mathcal{Z}$ projected on the common subspace as returned by JSSE. Since these subspaces are a mixture of stationary, linear mixing and a nonstationary (and thus temporal nonlinear) mixing, respectively, the mixing matrix should be evaluated piecewise (i.e., temporally nonstationary), see (6) and (7). However, the estimated component $\hat{\mathbf{s}}_i$ returned by the JSSE can be seen as being linearly mixed in the data, since it is itself already a combination of stationary and nonstationary source estimates ($\mathbf{y}$ and $\mathbf{z}$, and their respective stationary and nonstationary mixing matrix, as stated above). Hence, we might calculate its contribution as the least squares estimate between our original data x and a linear mixture $\mathbf{h}_i \in \mathbb{R}^M$ of the component $\hat{s}_i, i \in \{i, \forall i : |\Sigma_{ii}| \geq 0.9\}$. This is given by

$$\mathbf{h}_i = \arg\min_{\mathbf{h}_i} E\{||x - \mathbf{h}_i\hat{s}_i||_2\}. \tag{8}$$

To deflate the dataspace $\mathcal{X}$ (associated with x) by the subspace $\mathcal{H}$ spanned by the vectors $\mathbf{h}_i$, we use an iterative procedure, replacing x in (8) by the current estimate of the background activity $\hat{\mathbf{b}}(k)$ ($\hat{\mathbf{b}}(1)$ being x itself). Subsequently the projection of h_i onto the already calculated subspace $\mathcal{H}^p = \mathrm{span}[\mathbf{h}_1\mathbf{h}_2\cdots\mathbf{h}_{i-1}]$ is subtracted from $\mathbf{h}_i$ by using a Gram-Schmidt orthogonalisation procedure [34, pages 230–232]. The new estimate of the background estimate is then calculated as

$$\hat{\mathbf{b}}(k+1) = \left(\mathbf{I} - \mathbf{h}_i(\mathbf{h}_i^T\mathbf{h}_i)^{-1}\mathbf{h}_i^T\right)\hat{\mathbf{b}}(k), \tag{9}$$

where $\mathbf{I}$ is the identity matrix in $\mathbb{R}^{M\times M}$ and $(\mathbf{h}_i^T\mathbf{h}_i)^{-1}\mathbf{h}_i^T$ is the (left) Moore-Penrose pseudoinverse of $\mathbf{h}_i$.

2.6. Alternatives to the proposed method

As noted in the introduction, the concept of ICA can be approached in different ways, yet leading to the same objective of mutual information reduction or maximal mutual independence. For comparison we include three ICA algorithms based on different point of views on statistical independence, that is, FastICA, JADE/COM1, and SOBI.

2.6.1. FastICA

FastICA [21] is probably the most widely spread ICA method in various research communities. The popularity of FastICA can be explained mainly through its ease of use and the various possibilities to manipulate the objective, see [15, 38, 39] amongst others. Basically, the algorithm is supported by the general definition of statistical independence, saying that variables are independent if they are uncorrelated through every function. Furthermore, the method makes use of the optimal decorrelation function, namely the inverse cumulative density function of the source variables. Both assumptions are united in the decorrelation of the output of a fixed nonlinear function (e.g., tanh, x^4) of the prewhitened data. This has been shown to be similar to maximising the kurtosis of the estimated sources in case the nonlinear function approximates the inverse cumulative density function.

2.6.2. JADE or COM1

The JADE [25] and COM1 [24] algorithms are both based on the maximisation of the marginal source cumulants by minimising the cross-cumulants of fourth order, either by jointly diagonalising tensor slices (JADE), either by pairwise processing of the entry signals (COM1). The idea originates from the Edgeworth expansion of the density functions, providing a sufficient statistic when truncated at order four. Both algorithms return equal performance rates and differ mainly in computational complexity [24, 37].

2.6.3. SOBI

Relying solely on second order techniques, SOBI is an ICA algorithm using spatial as well as temporal information from the observed dataset[1]. The objective is to jointly decorrelate

[1] This is in contrast with all of the above algorithms which consider the observations as stochastic independent realisations and thus neglecting any possible temporal dependencies.

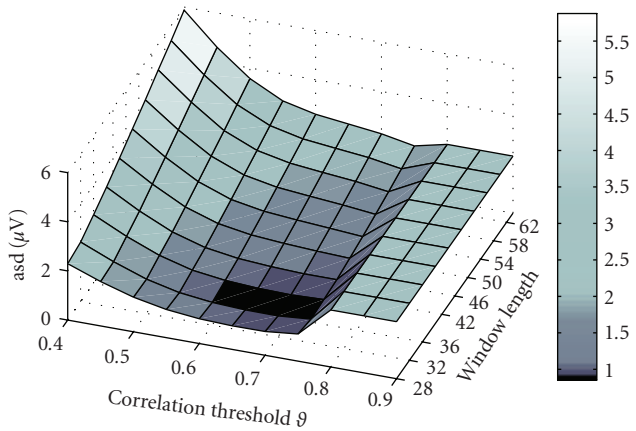


FIGURE 2: The asd as a function of the correlation threshold ϑ and of the window length T.

the data spatially and temporally, based on the information in the autocorrelation matrices of the data $\mathbf{R}_{\mathbf{X}(0)\mathbf{X}(\tau)}$. The input to the algorithm requires an additional set of time lags upon which SOBI will act. For this work, the set of time lags τ has been chosen as $T = \{\tau_i | 0 \leq \tau_i \leq 4 * T_s, \forall i \in \mathbb{Z}\}$, where T_s is the sampling period of the signal.

2.7. pSVD

Although not an ICA algorithm, we have a closer look at pSVD as it is one of the basic methods underlying JSSE. It is mainly used here to contrast the performance of pSVD outside, respectively, within the JSSE framework.

2.8. Evaluation measures

For a comparison, we evaluate the method and put its outcome next to that of the underlying basic techniques of pSVD and FastICA and the alternatives JADE and SOBI. Since simple measures such as asd do not suffice for a detailed error evaluation, we opt for performance measures as they are given in [40]. The measures take into account the source interferences, the methodological artefact and the total distortion. For clarity the definitions of the measures used here are repeated from [40] below.

Definition 1. Source to interference ratio is given as $SIR = 10 \log_{10}(\|\mathbf{s}_t\|_F^2 / \|\mathbf{e}_i\|_F^2)$.

Definition 2. Source to artefact ratio is given as $SAR = 10 \log_{10}((\|\mathbf{s}_t + \mathbf{e}_i\|_F^2) / \|\mathbf{e}_a\|_F^2)$.

Definition 3. Source to distortion ratio is given as $SDR = 10 \log_{10}(\|\mathbf{s}_t\|_F^2 / (\|\mathbf{e}_i + \mathbf{e}_a\|_F^2))$.

Where $\|\cdot\|_F$ denotes the Frobenius norm of its argument, s_t is the source estimate, and e_i and e_a are the interference and artefact error, respectively.

The advantages of this set of measures is that it splits up the error in the estimated source $\hat{\mathbf{s}}_j$ into a contribution that is related to the projection on the original source space of $\mathbf{s}_j$ ($\mathbf{s}_t$), a projection of $\hat{\mathbf{s}}_j$ on the subspace spanned by the vectors $\mathbf{s}_k$, $\forall k \neq j$ (interference $\mathbf{e}_i$) and an artefactual source that is the projection on the remaining subspace which cannot be explained by any of the above two projections (artefact $\mathbf{e}_a$). The latter is directly related to our methodologically introduced error or to numerical (round off) errors. The definitions of the above-mentioned measures resemble the familiar SNR definitions but are slightly altered to share mutually as little information as possible.

Note that we have omitted the noise term in all definitions, because we do not evaluate any noise perturbation studies. Noise perturbation studies of the pSVD algorithm can be found in [20] and for the JADE/COM1 algorithm in [24], amongst others.

3. RESULTS

3.1. Parameter settings

To have an optimal parameter set for pSVD we minimalise the asd on a group of 250 simulated datasets over a set of correlation parameters and window lengths. Since we want to keep the library as general as possible, no changes are made in the spatial reference maps $\mathbf{topo}_i$, $\forall i : 1 \leq i \leq 8$, restricting the tuning to the two parameters mentioned above. The correlation parameter ϑ and the window length T were varied independently, whereupon the minimum asd (mean over the 250 datasets) was found at a window length of $T = 32$ with a correlation threshold $\vartheta = 0.7$.

Figure 2 shows the mean values of asd as a function of T and ϑ. The minimum is reached at $T = 32$ and $\vartheta = 0.7$, respectively. Nevertheless, from hereon a threshold value of $\vartheta = 0.6$ is chosen in Criterion 1 to make the method as robust as possible to small changes in the data. It can be seen from Figure 2 that this small alteration in ϑ does not change a lot in the final asd value, but it will ensure a better performance in patient data where there is a higher effect of interfering background activity. Taking a value that is greater than 0.75 results in a value that is equal to no change, that is there is no component that will be identified as being close enough to the template library. From Figure 2 it is clear that the optimal window length is 32. Increasing the number of samples suffers from the orthogonality constraint and the stationarity assumptions that are made during this long-lasting window, decreasing the number of samples will result in an insufficient sample size for a robust estimation of the correlation.

3.2. Simulated data

We show the consecutive steps for the artefact reduction with JSSE on a simulated dataset with an SNR of -18 dB. Figure 3 shows the simulated dataset consisting of the background EEG and ocular artefacts. There are 4 blinks in the dataset with varying amplitude and varying topography (left and right eye blink). Figure 4 shows the results of JSSE acting on the dataset in Figure 3. In Figure 5, the intermediate estimated sources are displayed for JADE and pSVD together with the final estimate through their combination using JSSE.

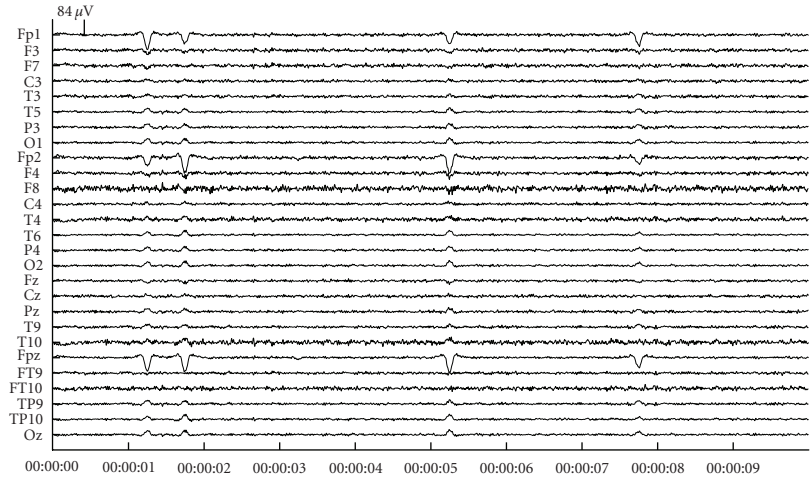


Figure 3: An example of a simulated dataset with blinks at an average of 5 dB above the background EEG level (SNR −5 dB).

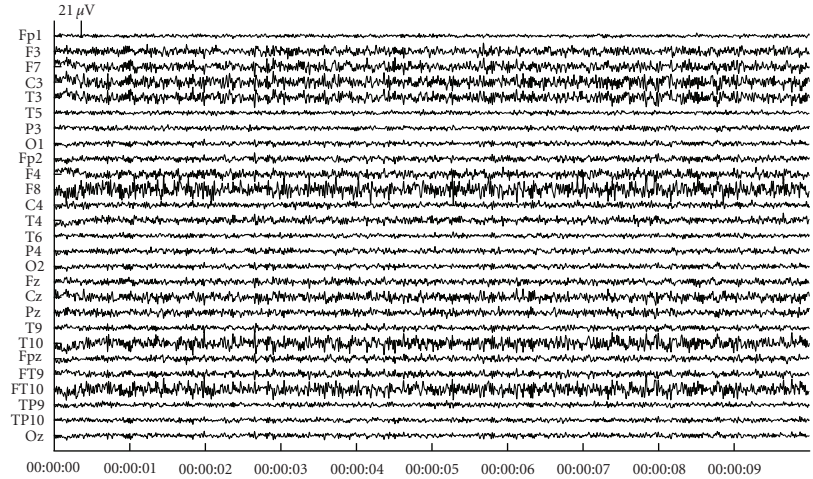


Figure 4: The cleaned simulation dataset.

Table 2: Results from 250 runs of simulations for which the mean was taken over SNR levels of −20 dB to 10 dB.

	SOBI	JADE	FastICA	pSVD	JSSE
SIR (dB)	21.71	14.93	17.74	26.06	17.07
SAR (dB)	10.58	15.51	12.28	18.98	39.51
SDR (dB)	8.50	9.36	10.16	17.97	14.80

In Table 2, we show the results from 250 trials on simulated datasets for an SNR (see (3)) range of −20 dB to 0 dB. We compare the combined subspace method JSSE to the underlying algorithms that provide the subspace estimation (JADE and pSVD) and the two proposed alternatives SOBI and FastICA. To see the behaviour of the algorithms as a function of the SNR values of the datasets, we set out SDR, SAR and SIR values against SNR in Figure 6.

3.3. Patient data

Figures 7 and 9 contain two snippets of patient datasets recorded at the Ghent University Hospital. Figure 7 contains clear blinking artefacts at seconds 1, 3, and 7, whereas Figure 9 contains clear saccades at seconds 1, 5, and 8. Both dataframes have been subjected to JSSE of which the obtained results can be seen in Figures 8 and 10, respectively. For clarity, the spectrum of JSSE that accompanies the results in Figure 7 (i.e., the values on the diagonal of Σ obtained at the second step of JSSE, see Section 2.4) are given in Figure 12 and a profile of the pSVD correction is given in Figure 11. The latter shows how many windows were deflated to reconstruct the current 8 samples.

Figure 13 shows two scalp maps, representing the weighing of a source estimate of JADE, respectively, JSSE onto the scalp electrodes (both components were taken to correspond to the same eye movement, i.e., a right eye blink). The scalp

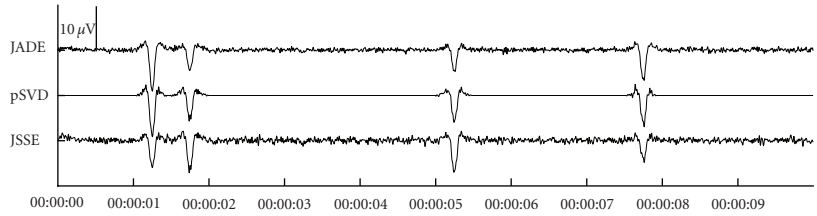


Figure 5: The first extracted source as estimated by JADE, pSVD, and JSSE. The ordening by JADE was done with descending kurtosis.

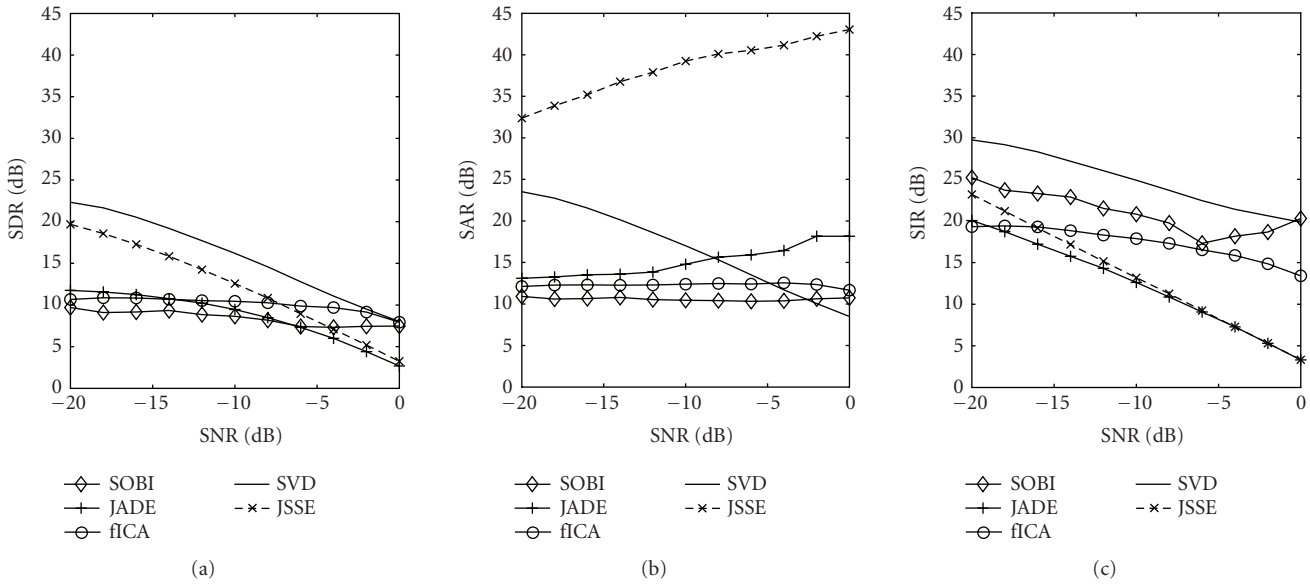


Figure 6: Values of SDR (a), SAR (b), and SIR (c) as a function of the SNR levels.

map associated to the JSSE component is given by the entry $\mathbf{h}_i$ in the topographical matrix H. Remember that a scalp map reflecting the activity of the eye movement estimated by pSVD cannot be given, since it includes nonstationarities which are inherent to the method of pSVD.

4. DISCUSSION

Combining two statistical estimation algorithms through the JSSE results in an ameliorated eye movement estimation from the EEG. The motivation to use short time statistics (pSVD) to cope with the nonstationarity of the cerebral activity is justified in the sense that it results in a minimisation of interference from other sources present in the EEG (reflected in maximal SIR), although it might introduce too much artefactual components caused by its windowing (SAR). Using the prior that in most cases eye movements are independent from the cerebral processes, the results of JADE show a quite good artefact suppression—although lower than that of pSVD—in the considered window (SAR) but are disappointing with respect to the interference suppression (SIR). The introduction of the joint and smoothened subspace estimation offers a solution hereto by augmenting the SAR through joining the advantages of both techniques. The results in Table 2 and Figure 6 show that the extracted component results in an interference and distortion suppression that are close to the pSVD results, while the enoying windowing artefact is suppressed outstandingly in its combination with JADE through the JSSE.

From Figure 6, it can be seen that the price to pay for a such amelioration in artefact suppression is an approximately constant 3 dB loss in SDR with respect to piecewise corrected EEG (pSVD). For the interference suppression, this even runs up to 17 dB (at an SNR level of 0 dB), although being acceptable at reasonable SNR levels (approx., 8 dB at −15 dB SNR). We thus have to give in on both SDR and SIR if a gain in SAR is of importance. From Figure 11 it can be seen that methodological artefacts can be introduced quite easily by the windowing that is inherent to the pSVD method. Since the first component of each local SVD decomposition does not necessarily the same spatial projection, the reconstruction introduces discontinuities related to the windowing. The reconstruction being influenced by a possible nonstationarity of the eye movement vector, nonstationarities in the background activity as well as the on/off switching caused by the binary decision process. The latter is directly reflected in the artefact error and by consequence in both the SDR and SAR values.

Experiments on patient EEG showed promising results concerning the suppression of blinks and saccades. Although

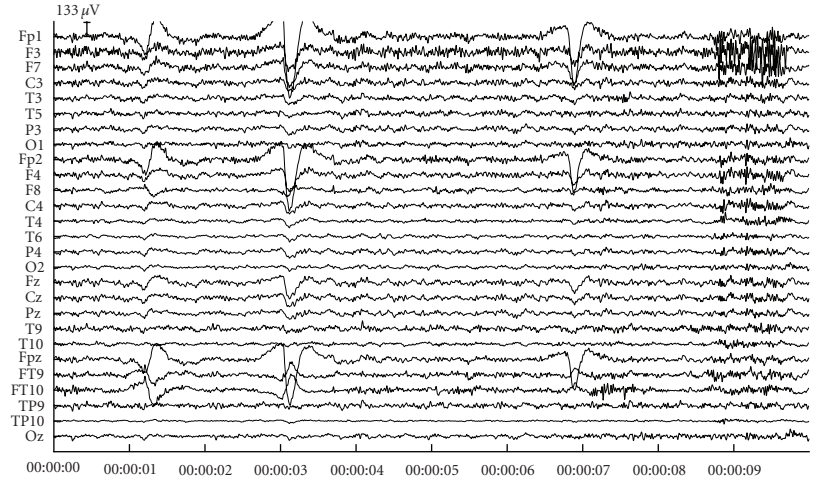


Figure 7: An example fragment of blinking artefacts. The blinks are clearly visible at seconds 1, 3, and 7.

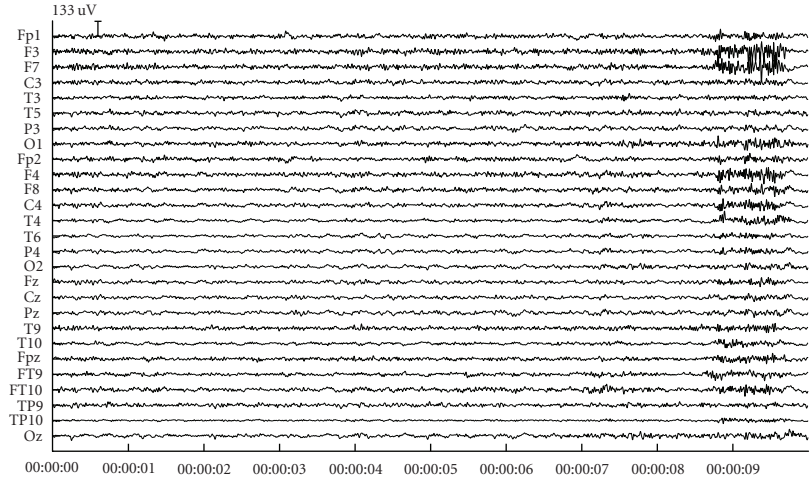


Figure 8: The results after having subjected the blink fragment of Figure 7 to JSSE.

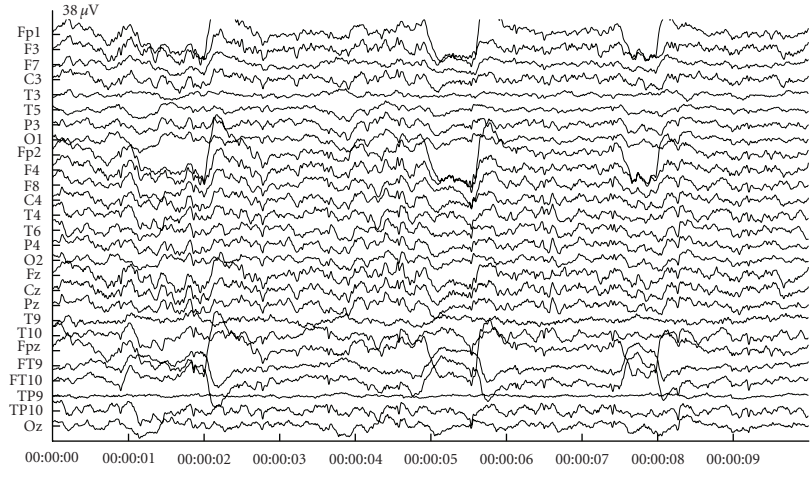


Figure 9: An example fragment of saccades. The saccades are clearly visible at seconds 1, 5, and 8.

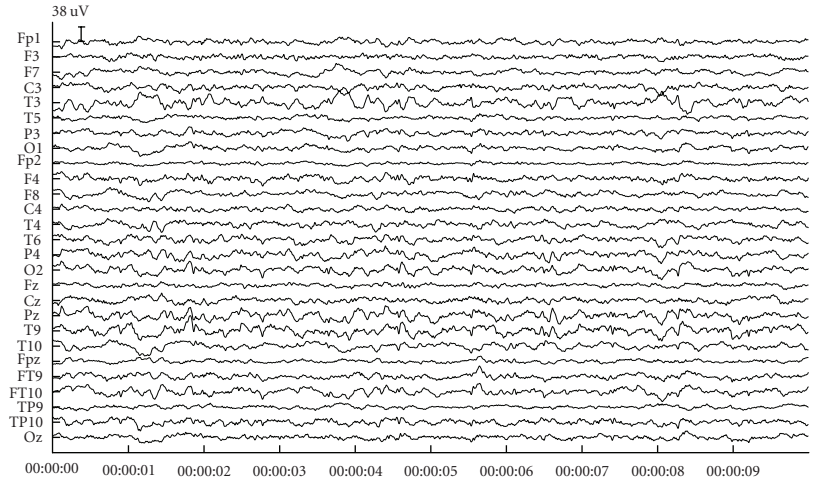


FIGURE 10: The results after having subjected the saccade fragment of Figure 9 to JSSE.

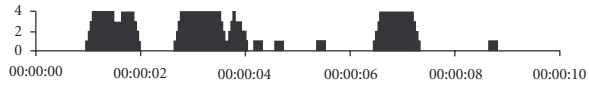


FIGURE 11: The pSVD profile associated with the blink fragment of Figure 7. The bars denote the number of deflations that occurred for each 8 sample window (with a maximum of 4 occurrences, see text).

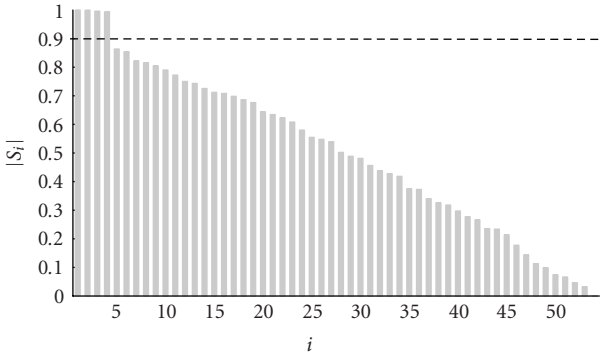


FIGURE 12: The spectrum of the subspace angles as obtained by the blink frame of Figure 7. The dashed horizontal line denotes the threshold level of 0.9.

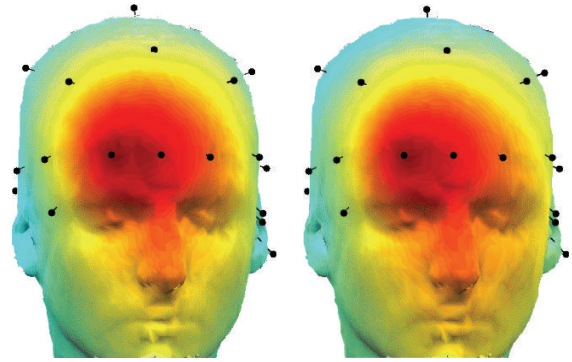


FIGURE 13: The topography of the first source of JADE (a) and an estimation of the JSSE topography, based on the reconstruction in (9) (b).

it is difficult, or even impossible, to show objective measures for evaluation, we observe that the estimated topographies are close to the topographies as estimated through JADE. An exemplar topography as in Figure 13 shows that the topography is spatially even more concentrated around the eye, pointing at a closer to dipole behaviour, which is in line with the model proposed in [7].

From Figure 12 it can be seen that the chosen threshold for the spectrum of the JSSE falls in the spectral gap. Although this is not always the case, the threshold at 0.9 offers a reliable reconstruction in the majority of the cases as proved by the simulation results and the two patient frames presented, where even on visual inspection it can be seen that the method leaves almost no traces at the time spots where it interacted on the EEG recording (see Figures 4, 8, and 10).

One could of course think of other subspace combining methods. The simplest form being the reconstruction of the EEG by taking a weighted sum of both partial reconstructions as they are given by JADE and pSVD, respectively. However, the disadvantage would be that the errors are added while JSSE has a more sophisticated error suppression with respect to the errors introduced by the two supporting methods (see Figure 6 and Table 2). Yet another combination method would be to take a threshold onto the profile provided by pSVD (as in Figure 11) and only consider the time instances of the ICA (in our case, this would be JADE) reconstruction that are labelled by this thresholding, leaving the remaining time instances untouched. Unfortunately, this does not resolve for the artefact suppression. On the contrary, the discontinuities will be more articulated if the threshold would be augmented (resulting in lower values of SAR), and the interference at the time instances considered will not exceed the performance of a regular ICA algorithm. Moreover,

it is implicitly assumed that the background activity would be stationary along the complete frame (i.e., 10 seconds in our case), which is quite in contrast with the findings in [35].

5. CONCLUSION

This study shows the importance of combining BSS techniques with different order statistics. It convincingly shows that merging short time signal characteristics (pSVD) with more global measures (JADE) into a joint smoothened subspace estimator (JSSE) provides acceptable to outstanding results compared to many of the commonly used standard ICA techniques. More specifically, it follows directly from our simulations that the proposed method is superior in artefact suppression, while it keeps up with the methods of FastICA, JADE, pSVD, and SOBI concerning the interference and distortion suppression, especially at low (highly negative) SNR values.

The proposed method has proven to be capable of suppressing ocular artefacts in the EEG in a fully automated way, relying on a set of patient-independent reference topographies as a prior.

ACKNOWLEDGMENTS

The authors kindly acknowledge the (FWO) *Fonds voor Wetenschappelijk Onderzoek, Vlaanderen/Scientific Research Funds, Flanders* to support this work through Grant G.0416 .04. the authors also would like to acknowledge the collaborations of the Laboratory for Clinical and Experimental Neurophysiology at the Ghent University Hospital for providing them with the patient data files for the study. A special thank goes to the reviewers who supported this work through their valuable comments.

REFERENCES

[1] A. Lutz, L. L. Greischar, N. B. Rawlings, M. Ricard, and R. J. Davidson, "Long-term meditators self-induce high-amplitude gamma synchrony during mental practice," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 46, pp. 16369–16373, 2004.

[2] A. Gunji, R. Ishii, W. Chau, R. Kakigi, and C. Pantev, "Rhythmic brain activities related to singing in humans," *NeuroImage*, vol. 34, no. 1, pp. 426–434, 2007.

[3] W. de Clercq, A. Vergult, B. Vanrumste, W. van Paesschen, and S. van Huffel, "Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 12, part 1, pp. 2583–2587, 2006.

[4] T. C. Weerts and P. J. Lang, "The effects of eye fixation and stimulus and response location on the contingent negative variation (CNV)," *Biological Psychology*, vol. 1, no. 1, pp. 1–19, 1973.

[5] R. Verleger, "The instruction to refrain from blinking affects auditory P3 and N1 amplitudes," *Electroencephalography and Clinical Neurophysiology*, vol. 78, no. 3, pp. 240–251, 1991.

[6] C. J. Ochoa and J. Polich, "P300 and blink instructions," *Clinical Neurophysiology*, vol. 111, no. 1, pp. 93–98, 2000.

[7] M. Iwasaki, C. Kellinghaus, A. V. Alexopoulos, et al., "Effects of eyelid closure, blinks, and eye movements on the electroencephalogram," *Clinical Neurophysiology*, vol. 116, no. 4, pp. 878–885, 2005.

[8] H. Hallez, A. Vergult, R. Phlypo, et al., "Muscle and eye movement artifact removal prior to EEG source localization," in *Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS '06)*, pp. 1002–1005, New York, NY, USA, August-September 2006.

[9] S. A. Hillyard and R. Galambos, "Eye movement artifact in the CNV," *Electroencephalography and Clinical Neurophysiology*, vol. 28, no. 2, pp. 173–182, 1970.

[10] R. J. Somsen and B. van Beek, "Ocular artifacts in children's EEG: selection is better than correction," *Biological Psychology*, vol. 48, no. 3, pp. 281–300, 1998.

[11] M. Fatourechi, A. Bashashati, R. K. Ward, and G. E. Birch, "EMG and EOG artifacts in brain computer interface systems: a survey," *Clinical Neurophysiology*, vol. 118, no. 3, pp. 480–494, 2007.

[12] R. J. Croft and R. J. Barry, "Removal of ocular artifact from the EEG: a review," *Neurophysiologie Clinique*, vol. 30, no. 1, pp. 5–19, 2000.

[13] A. Schlögl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller, "A fully automated correction method of EOG artifacts in EEG recordings," *Clinical Neurophysiology*, vol. 118, no. 1, pp. 98–104, 2007.

[14] S. Makeig, A. J. Bell, T.-P. Jung, and T. J. Sejnowski, "Independent component analysis of electroencephalographic data," in *Advances in Neural Information Processing Systems*, vol. 8, pp. 145–151, MIT Press, Cambridge, Mass, USA, 1996.

[15] R. N. Vigário, "Extraction of ocular artefacts from EEG using independent component analysis," *Electroencephalography and Clinical Neurophysiology*, vol. 103, no. 3, pp. 395–404, 1997.

[16] G. L. Wallstrom, R. E. Kass, A. Miller, J. F. Cohn, and N. A. Fox, "Automatic correction of ocular artifacts in the EEG: a comparison of regression-based and component-based methods," *International Journal of Psychophysiology*, vol. 53, no. 2, pp. 105–119, 2004.

[17] G. Wallstrom, R. Kass, A. Miller, J. Cohn, and N. Fox, "Correction of ocular artifacts in the EEG using Bayesian adaptive regression splines," in *Bayesian Statistics*, vol. 6, pp. 351–366, Springer, New York, NY, USA, 2002.

[18] O. G. Lins, T. W. Picton, P. Berg, and M. Scherg, "Ocular artifacts in recording EEGs and event-related potentials II: source dipoles and source components," *Brain Topography*, vol. 6, no. 1, pp. 65–78, 1993.

[19] S. Casarotto, A. M. Bianchi, S. Cerutti, and G. A. Chiarenza, "Principal component analysis for reduction of ocular artefacts in event-related potentials of normal and dyslexic children," *Clinical Neurophysiology*, vol. 115, no. 3, pp. 609–619, 2004.

[20] R. Phlypo, P. van Hese, H. Hallez, et al., "PSVD: a method for robust, real time eye movement artifact rejection from the EEG," in *Proceedings of the 3rd IET International Conference on Advances in Medical, Signal and Information Processing (MEDSIP '06)*, p. 4, Glasgow, Scotland, July 2006.

[21] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, no. 7, pp. 1483–1492, 1997.

[22] T.-W. Lee, M. Girolami, and T. J. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources," *Neural Computation*, vol. 11, no. 2, pp. 417–441, 1999.

[23] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444, 1997.

[24] P. Comon, "Independent component analysis. a new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[25] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proceedings F*, vol. 140, no. 6, pp. 362–370, 1993.

[26] M. Borga and H. Knutsson, "A canonical correlation approach to blind source separation," Tech. Rep. LiU-IMT-EX-0062, Department of Biomedical Engineering, Linköping University, Linköping, Sweden, 2001.

[27] V. Krishnaveni, S. Jayaraman, L. Anitha, and K. Ramadoss, "Removal of ocular artifacts from EEG using adaptive thresholding of wavelet coefficients," *Journal of Neural Engineering*, vol. 3, no. 4, pp. 338–346, 2006.

[28] A. Erfanian and B. Mahmoudi, "Real-time ocular artifact suppression using recurrent neural network for electroencephalogram based brain-computer interface," *Medical and Biological Engineering and Computing*, vol. 43, no. 2, pp. 296–305, 2005.

[29] S. Puthusserypady and T. Ratnarajah, "Robust adaptive techniques for minimization of EOG artefacts from EEG signals," *Signal Processing*, vol. 86, no. 9, pp. 2351–2363, 2006.

[30] R. Agarwal, T. Takeuchi, S. Laroche, and J. Gotman, "Detection of rapid-eye movements in sleep studies," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 8, pp. 1390–1396, 2005.

[31] H. Hallez, P. van Hese, B. Vanrumste, et al., "Dipole localization errors due to not incorporating compartments with anisotropic conductivities: simulation study in a spherical head model," *International Journal of Bioelectromagnetism*, vol. 7, no. 1, pp. 134–137, 2005.

[32] M. Babaie-Zadeh and C. Jutten, "Semi-blind approaches for source separation and independent component analysis," in *Proceedings of the 14th European Symposium on Artificial Neural Networks (ESANN '06)*, pp. 301–312, Bruges, Belgium, April 2006.

[33] J. M. Stern and J. Engel Jr., *Atlas of EEG Patterns*, Lippincott Williams & Wilkins, Philadelphia, Pa, USA, 2004.

[34] G. H. Golub and C. F. van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, Md, USA, 3rd edition, 1996.

[35] W. J. Freeman, "Origin, structure, and role of background EEG activity—part 2: analytic phase," *Clinical Neurophysiology*, vol. 115, no. 9, pp. 2089–2107, 2004.

[36] A. Hyvärinen, J. Särelä, and R. Vigário, "Bumps and spikes: artifacts generated by independent component analysis with insufficient sample size," in *Proceedings of the 1st International Workshop on Independent Component Analysis and Blind Signal Separation (ICA '99)*, pp. 425–429, Aussois, France, January 1999.

[37] J.-F. Cardoso, "High-order contrasts for independent component analysis," *Neural Computation*, vol. 11, no. 1, pp. 157–192, 1999.

[38] C. W. Hesse and C. J. James, "On semi-blind source separation using spatial constraints with applications in EEG analysis," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 12, part 1, pp. 2525–2534, 2006.

[39] W. Lu and J. C. Rajapakse, "ICA with reference," in *Proceedings of the 3rd International Conference on Independent Component Analysis and Blind Source Separation (ICA '01)*, pp. 120–125, San Diego, Calif, USA, December 2001.

[40] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 4, pp. 1462–1469, 2006.

Hindawi Publishing Corporation
Computational Intelligence and Neuroscience
Volume 2007, Article ID 14567, 13 pages
doi:10.1155/2007/14567

Research Article

A Framework to Support Automated Classification and Labeling of Brain Electromagnetic Patterns

Gwen A. Frishkoff,[1] Robert M. Frank,[2] Jiawei Rong,[3] Dejing Dou,[3] Joseph Dien,[4] and Laura K. Halderman[1]

[1] *Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA 15260, USA*
[2] *NeuroInformatics Center, University of Oregon, 1600 Millrace Drive, Eugene, OR 97403, USA*
[3] *Computer and Information Sciences, University of Oregon, Eugene, OR 97403, USA*
[4] *Department of Psychology, University of Kansas, 1415 Jayhawk Boulevard, Lawrence, KS 66045, USA*

Correspondence should be addressed to Gwen A. Frishkoff, gwenf@pitt.edu

Received 19 February 2007; Revised 28 July 2007; Accepted 7 October 2007

Recommended by Saied Sanei

This paper describes a framework for automated classification and labeling of patterns in electroencephalographic (EEG) and magnetoencephalographic (MEG) data. We describe recent progress on four goals: 1) specification of rules and concepts that capture expert knowledge of event-related potentials (ERP) patterns in visual word recognition; 2) implementation of rules in an automated data processing and labeling stream; 3) data mining techniques that lead to refinement of rules; and 4) iterative steps towards system evaluation and optimization. This process combines top-down, or knowledge-driven, methods with bottom-up, or data-driven, methods. As illustrated here, these methods are complementary and can lead to development of tools for pattern classification and labeling that are robust and conceptually transparent to researchers. The present application focuses on patterns in averaged EEG (ERP) data. We also describe efforts to extend our methods to represent patterns in MEG data, as well as EM patterns in source (anatomical) space. The broader aim of this work is to design an ontology-based system to support cross-laboratory, cross-paradigm, and cross-modal integration of brain functional data. Tools developed for this project are implemented in MATLAB and are freely available on request.

1. INTRODUCTION

The complexity of brain electromagnetic (EM) data has led to a variety of processes for EM pattern classification and labeling over the past several decades. The absence of a common framework may account for the dearth of statistical metaanalyses in this field. Such cross-lab, cross-paradigm reviews are critical for establishing basic findings in science. However, reviews in the EM literature tend to be informal, rather than statistical: it is difficult to generalize across datasets that are classified and labeled in different ways.

To address this problem, we have designed a framework to support automated classification and labeling of patterns in electroencephalographic (EEG) and magnetoencephalographic (MEG) data. In the present paper, we describe the framework architecture and present an application to averaged EEG (event-related potentials, or ERP) data collected in a visual word recognition paradigm. Results from this study illustrate the importance of combining top-down and bottom-up approaches. In addition, they suggest the need for ongoing system evaluation to diagnose potential sources of error in component analysis, classification, and labeling. We conclude by discussing alternative analysis pathways and ways to improve efficiency of implementation and testing of alternative methods. It is our hope that this framework can support increased collaboration and integration of ERP results across laboratories and across study paradigms.

1.1. Classification of ERPs

A standard technique for analysis of EEG data involves averaging across segments of data (*trials*), time-locking to stimulus or response *events*. The resulting measures are characterized by a sequence of positive and negative deflections distributed across time and space (scalp locations). In principle, activity that is not event-related will tend towards zero as the number of averaged trials increases. In this way, ERPs

provide increased signal-to-noise, and thus increased sensitivity, to functional (e.g., task) manipulations. Signal averaging assumes that the brain signals of interest are time-locked to (or "evoked by") the events of interest. As illustrated in recent work on induced (nontime-locked) versus evoked (time-locked) EEG activity, this assumption does not always hold ([1, 2]).

In the past several decades, researchers have described several dozen spatiotemporal ERP patterns (or *components*), which are thought to index a variety of neuropsychological processes. Some patterns are observed across a range of experimental contexts, reflecting domain-general processes, such as memory, decision-making, and attention. Other patterns are observed in response to specific types of stimuli, reflecting human expertise in domains such as mathematics, face recognition, and reading comprehension (for reviews see [3, 4]). Previous investigations of these patterns have demonstrated the effectiveness of ERP methods for addressing basic questions in nearly every area of psychology.

Given the success of this methodology, ERPs are likely to remain at the forefront of research in clinical and cognitive neuroscience, even as newer methods for EEG and MEG analyses are developed as alternatives to signal averaging (e.g., [1, 2, 5–7]).

At the same time, ERP methods face some important challenges. A key challenge is to identify standardized methods for measure generation, as well as objective and reliable methods for identification and labeling of ERP components. Traditionally, researchers have characterized ERP components in respect to both physiological (spatial, temporal) and functional criteria [8, 9]. Physiological criteria include latency and scalp distribution, or topography. For example, as illustrated in Figure 1, the visual "P100 component" is characterized by a positive deflection that peaks at ~100 milliseconds after onset of a visual stimulus (A) and is maximal over occipital electrodes, reflecting activity in visual cortex (B).

Despite general agreement on criteria for ERP component identification [9], in practice such patterns can be hard to identify, particularly in individual subjects. This difficulty is due in part to the superposition of patterns generated by multiple brain regions at each time point [10], leading to complex spatial patterns that reflect the mixing of underlying patterns. Given this complexity, ERP researchers have adopted a variety of solutions for scalp topographic analysis (e.g., [11, 12]). It can therefore be difficult to compare results from different studies, even when the same experimental stimuli and task are used.

Similarly, researchers use a variety of methods for describing temporal patterns in ERP data [13]. For example, early components, such as the P100, tend to be characterized by their peak latency, while the time course of later components, such as the N400 or P300, is typically captured by averaging over time "windows" (e.g., 300–500 milliseconds). The latency of other components, such as the N400, has been quantified in a variety of ways. Finally, there is variability in how functional information (e.g., subject-, stimulus-, or task-specific variables) is used in ERP pattern classification. Some patterns, such as the P100, are easily observed as large deflections in the raw ERP waveforms. Other patterns, such as the mismatch negativity are more reliably seen in difference measures, calculated by subtracting ERP amplitude in one condition from the ERP amplitude in a contrasting condition. This inconsistency may lead to confusion, particularly when the same label is used to refer to two different measures, as is often the case.

1.2. Outline of paper

In summary, the complexity of ERP data has led to multiple processes for measure generation and pattern classification that can vary considerably across different experiment paradigms and across research laboratories. Ultimately, this limits the ability both to replicate prior results and to generalize across findings to achieve high-level interpretations of ERP patterns.

In light of these challenges, the goal of this paper is to describe a framework for automated classification and labeling of ERP patterns. The framework presented here comprises both *top-down* (knowledge-driven) and *bottom-up* (data-driven) methods for ERP pattern analysis, classification, and labeling. Following, we describe this framework in detail (Section 2) and present an application to patterns in ERP data from a visual word processing paradigm (Section 3). Section 4 describes approaches to system evaluation. Section 5 describes data mining for refinement of expert-driven (top-down) methods. In Section 6, we draw some general conclusions and discuss extensions of our framework for representation of patterns in source space, and ontology development to support cross-paradigm, cross-laboratory, and cross-modal integration of results in EM research.

2. PATTERN CLASSIFICATION FRAMEWORK

As illustrated in Figure 2, our framework comprises five main processes.

(i) *Knowledge engineering*. Known ERP patterns are cataloged (1). High-level rules and concepts are described for each pattern (2).

(ii) *Pattern analysis* and *measure generation*. Analysis methods are selected and applied to ERP data (3). The goal is transformation of continuous spatiotemporal data into discrete patterns for labeling. Statistics are generated (4) to capture the rules and concepts identified in (2).

(iii) *Data mining*. Unsupervised clustering (7) and supervised learning (8) are used to explore how measures cluster, and how these clusters may be used to identify and label patterns using rules derived independently of expert knowledge.

(iv) *Operationalization and application of rules*. Rules are operationalized by combining metrics in (4) with prior knowledge (2). Data mining results (7-8) may be used to validate and refine the rules. Rules are applied to data, using an automated labeling process (6) detailed below.

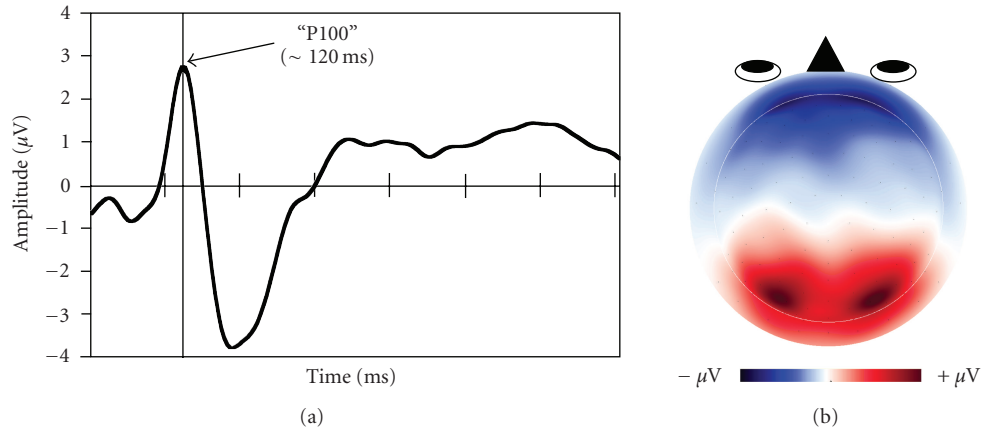


Figure 1: (a) Time course of P100 pattern, plotted at left occipital electrode, O1. Time is plotted on the x-axis (0–700 milliseconds); each vertical hash mark represents 100 milliseconds. Amplitude is plotted on the y-axis (scale, $\pm 4\,\mu$V). The dark vertical line marks the time of peak amplitude (~120 milliseconds). (b) Scalp topography of the P100 pattern, plotted at the time of peak amplitude. Red, positive. Blue, negative.

Following, we describe how these processes have been implemented in a series of MATLAB procedures. We then report results from the application of this process to data from a visual word processing experiment. Results are evaluated against a "gold standard" that consists of expert judgments regarding the presence or absence of patterns, and their prototypicality, for each of 144 observations (36 subjects ×4 experiment conditions).

2.1. Knowledge engineering (process 1, 2)

The goal of knowledge engineering is to identify concepts that have been documented for a particular research domain. Based on prior research on visual word processing we have tentatively identified eight spatiotemporal patterns that are commonly observed from ~100 to ~700 milliseconds after presentation of a visual word stimulus, including the P100, N100, late N1/N2b, N3, P1r, MFN, N400, and P300. Space limitations preclude a detailed discussion of each pattern (see reviews in [3, 4]). The left temporal N3 and medial frontal negativity (MFN) components are less well known, but have been described in several high-density ERP studies of visual word processing (e.g., [14–16]). The P1r [17] has also been referred to as a posterior P2 [18]. The late N1/N2b has variously been referred to as an N2, an N170, and a recognition potential (see [15] for discussion and references). It is not clear that the late N1/N2 represents a component that is functionally distinct from the N1 and N3, though it sometimes emerges in tPCA results as a distinct spatiotemporal pattern (e.g., see Section 3). These eight patterns reflect a working taxonomy of ERP in research on visual word processing between ~60–700 milliseconds. Application of the present framework to large numbers of datasets collected across a range of paradigms, and across different ERP research labs, would contribute to the refinement of this taxonomy.

A note of caution is in order, concerning the labels for scalp regions of interest (ROIs). By convention, areas of the scalp are associated with anatomical labels, such as "occipital," "parietal," "temporal," and "frontal" (see Table 1). It is well known, however, that a positive or negative deflection over a particular scalp ROI is not necessarily generated in cortex directly below the measured data. ERP patterns can reflect sources tangential to the scalp surface. In this instance, the positive and negative fields may be maximal over remote regions of the scalp, reflecting a dipolar scalp distribution (e.g., with a positive maximum over frontal scalp regions, and a negative maximum over temporal scalp regions). Thus, the ROI labels should not be interpreted as literal references to brain regions. The ROI clusters used in the present study are shown in Appendix A.

Table 1: Spatial and temporal concepts used to define the eight target patterns. Regions of interest (ROIs) are defined in Appendix A.

Pattern	Window	ROI
P100	60–150	occipital
N100	151–230	occipital
N2	231–300	post-temporal
P1r	250–400	parietal
N3	250–400	left anterior
MFN	250–450	frontal
N4	350–550	parietal
P300	401–700	parietal

2.2. Data summary

Prior to analysis, ERP data consist of complex waveforms (time series), measured at multiple electrode sites. To simplify analysis and interpretation of these data, a standard practice is to transform the ERPs into discrete patterns. Traditional methods for data summary include identification of peak latency within a specified time window ("peak picking") and computing the mean amplitude over a time window for each electrode ("windowed analysis"), or averaged over

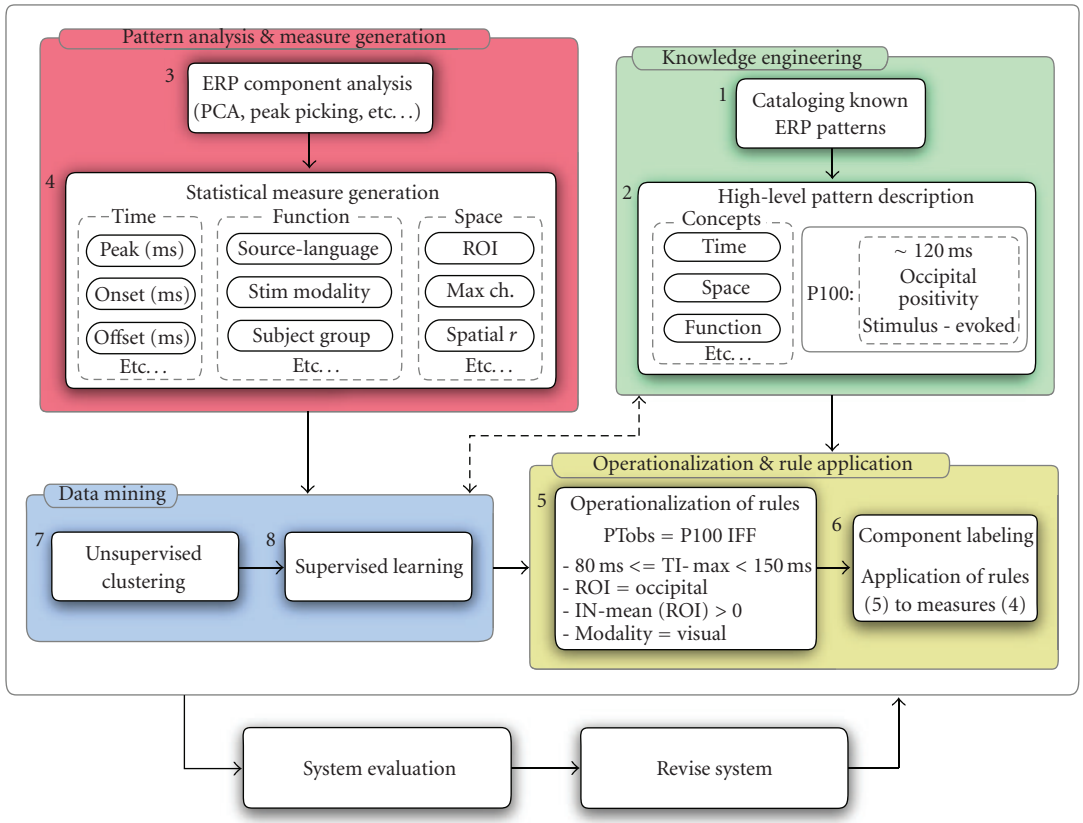


Figure 2: Pattern classification and labeling scheme. *Knowledge engineering* (processes 1, 2) includes "top-down" specification of ERP concepts and rules, formulated by domain experts. *Component analysis* and *measure generation* (processes 3, 4) yield summary metrics that are used for pattern classification and labeling. Implementation and operationalization of pattern rules (processes 5, 6) are detailed in Section 2. *Data mining* (processes 7, 8) includes "bottom-up" or data-driven methods for clustering and discovery of pattern rules (Section 5). *System evaluation* is detailed in Section 4.

electrode clusters (regions of interest—*ROIs*). An alternative method is principal components analysis (PCA), which decomposes the data into "latent" patterns, or *factors*. The following subsection describes this method in detail, and explains the utility of PCA for automated pattern classification.

2.2.1. Temporal PCA methods (process 3)

PCA belongs to a class of factor-analytic procedures, which use eigenvalue decomposition to extract linear combinations of variables (latent "factors") in such a way as to account for patterns of covariance in the data parsimoniously, that is, with the fewest factors. Mathematically, the goal of PCA is to take intercorrelated variables $(x_1, \ldots, x_n)$ and combine them such that the tranformed data, the "principal components" (PC), are linear combinations of x, weighted to maximize the amount of variance captured by each eigenvector (v_i):

$$\mathrm{PC}_1 = v_{11}x_1 + v_{12}x_2 + \cdots + v_{1n}x_n. \quad (1)$$

In this way, the original set of variables $(x_1, \ldots, x_n)$ is "projected" into a new data space, where the dimensions of this new space are captured by a small number of latent factors (the eigenvectors).

In ERP data, the variables $(x_1, \ldots, x_n)$ are the microvolt readings either at consecutive time points (temporal PCA) or at each electrode (spatial PCA). The major source of covariance isassumed to be the ERP components, characteristic features of the wave form that are spread across multiple time points and multiple electrodes. Ideally, each latent factor corresponds to a separate ERP component, providing a statistical decomposition of the brain electrical patterns that are superposed in the scalp-recorded data. To achieve this ideal factor-to-pattern mapping, the factors may be "rotated" so that the variance associated with the original variables (timepoints) is redistributed across the factors in such a way that maximizes "simple structure," that is, that achieves a simple and transparent mapping from variables to factors. (See [19] for a review of PCA and related factor-analytic methods for ERP data decomposition.)

In the present application, we used temporal PCA (tPCA) as implemented in the Dien PCA Toolbox [20]. In temporal PCA, the data are organized with the variables corresponding to time points and observations corresponding to the different waveforms in the dataset. The waveforms vary across subjects, electrodes, and experimental conditions. Thus, subject, spatial, and task variance are collectively responsible for covariance among the temporal variables. The data matrix

is then self-multiplied and mean-corrected to produce a covariance matrix. The covariance matrix is subjected to eigenvalue decomposition, and the resulting nonnoise factors are rotated using Promax to obtain a more transparent relationship between the PCA factors and the latent variables of interest (i.e., ERP components).

After transformation of the ERP data into factor space, the data are projected back into the original data space, by multiplying factor scores by factor loadings and by the standard deviation at each timepoint (see the appendix in [21]). In this way, it is possible to visualize and extract information about the strength of the pattern at each electrode, to determine the spatial distribution of the pattern for a given subject and experiment condition. Visualizing the spatial projection of each factor in this way is useful in interpreting tPCA results (e.g., see Figure 3(b)).

For our initial attempts to automate data description and classification, tPCA offered several advantages over traditional methods. First, tPCA is able to separate overlapping spatiotemporal patterns. Second, tPCA automatically extracts a discrete set of temporal patterns. Third, when implemented and graphed appropriately, tPCA results are easily interpreted with respect to previous findings, as illustrated below. tPCA is therefore easily incorporated in an automated process for ERP pattern extraction and classification. In the final section, we address some limitations of tPCA as a method of ERP pattern analysis.

2.2.2. Measure generation (process 4)

For each tPCA factor, we extracted 32 summary metrics that characterize spatial, temporal, and functional dimensions of the data. The full set of metrics, along with their definitions, is listed in Appendix C. Note that our expert-defined rules, which were used for the tPCA autolabeling process, mainly involved two metrics (see Section 2.2.3 for details): *In-mean (ROI)* and *TI-max*. In-mean (ROI) represents the amplitude over a region-of-interest (ROI), averaged over electrode clusters for each latent factor at the time of peak latency, after the factor has been projected back into channel space. TI-max is the peak latency and is measured on the factor loadings, which are sign-invariant.

Although these two metrics intuitively capture the spatial and temporal dimensions of the ERP data that are most salient to ERP researchers, our prior data mining results suggested that additional metrics might improve the tPCA autolabeling results [22, 23]. In particular, some failures in the autolabeling process (i.e., cases where the modal factor for a given pattern did not show a match to the rule in a given condition, for a given subject) were due to component overlap that remained even after tPCA. For example, in one of our four pilot datasets [23], the P100 pattern was partially captured by a factor corresponding to the N100. For some subjects, most of the P100 was in fact captured by this "N100 factor." The factor showed a slow negativity, beginning before the stimulus onset, and the P100 appeared as a positive going deflection that was superposed on this sustained negativity. However, because the rule specified that the mean amplitude over the occipital electrodes should be positive, the factor did not meet the P100 rule criteria.

To address this issue, we implemented *onset and offset metrics*. Each onset latency was estimated as the midpoint of four consecutive sliding windows in which corresponding t-tests (threshold, $P = .05$) indicated that the means of their respective windowed signals diverged significantly from a baseline value, typically zero. The subsequent offset was the temporal midpoint at which the four consecutive t-tests showed their windowed signal means returned to baseline. The procedure is implemented as described in [24].

Using the onset latency to determine a "baseline" (0-point or onset) for each pattern, we then computed *peak-to-baseline* and *baseline-to-peak* metrics to capture phasic deflections that could be confused with slow potentials. The baseline intensity was computed as the signal mean within an interval centered on component onset. We predicted that data mining results would incorporate these measures to yield improved accuracy in the labeling process.

In addition, we added metrics to capture variations in amplitude due to experimental variables. Four measures were computed: *Pseudo-Known* (difference in response to nonwords versus words), *RareMisses-RareHits* (difference in response to unknown rare words versus words that we correctly recognized), *RareHits-Known* (difference in response to rare versus low-frequency words), and *Pseudo-RareMisses* (difference in nonwords versus missed rare words). Because prior research has shown that semantic processing can affect the N2, N3, MFN, N4, and P3 patterns, we predicted that the data mining procedures would identify one or more of these metrics as important for pattern classification.

2.2.3. Rule operationalization (process 5)

Rules for each ERP pattern were formulated initially based on results from prior literature and were operationalized using metrics defined in Process 4 (Section 2.2.2). After application of the initial rules to test data, we evaluated the results against a "Gold Standard" (see Section 4 for details) and modified the pattern rules to improve accuracy. For example, after initial testing, the visual "P100" pattern (P100v) was defined as follows: for any n, FA_n = P100v *if and only if*

(i) 80 ms < TI-max (FA_n) ≤ 150 milliseconds,

(ii) In-mean(ROI) > 0,

(iii) EVENT (FA_n) = stimon,

(iv) MODALITY (EVENT) = visual,

where FA_n is defined as the nth tPCA factor, and P100v is the visual-evoked P100 ("v" stands for "visual"). TI-max is the time of peak amplitude, In-mean(ROI) is the mean amplitude over the region-of-interest (ROI), and ROI for P100v is specified as "occipital" (i.e., mean intensity over occipital electrodes). "Stimon" refers to stimulus onset, which is the event that is used for time-locking single trials to derive the ERP. "MODALITY" refers to the stimulus modality (e.g., visual, auditory, somatosensory, etc.). See Appendix B for a full listing of rule formulae.

These rules represent informed hypotheses, based on expert knowledge. As described below (Section 5), bottom-up methods can be used to refine these rules. Further, as the rules are applied to larger and more diverse sets of data, they are likely to undergo additional refinements (see Section 4.1).

2.2.4. Automated labeling (process 6)

For each condition, subject, and tPCA factor, we used MATLAB to compute temporal and spatial metrics on that factor's contribution to the scalp ERP. The values of the metrics specified in the expert defined rules were then compared to rule-specific thresholds that characterized specific ERP components. Thresholds were determined through expert definitions that were formulated and tested as described in Section 2.2.3). The results of the comparisons were recorded in a true/false table, and factors meeting all criteria were flagged as capturing the specified ERP component for that subject and condition. All data were automatically saved to Excel spreadsheets organized by rule, condition, and subject.

2.3. Data mining

As described in Section 2.1, ERP patterns are typically discovered through a "manual" process that involves visual inspection of spatiotemporal patterns and statistical analysis to determine how the patterns differ across experiment conditions. While this method can lead to consensus on the high-level rules and concepts that characterize ERP patterns in a given domain, operationalization of these rules and concepts is highly variable across research labs, as described in Section 1. Bottom-up (data-driven) methods can contribute to standardization of rules for classifying known patterns, and possibly to discovery of new patterns, as well. Here we describe two bottom-up methods, unsupervised learning (i.e., clustering) and supervised learning (i.e., decision tree classifiers).

2.3.1. Clustering (process 7)

In this study, we used the expectation-maximization (EM) algorithm for clustering [25], as implemented in WEKA [26]. EM is used to approximate distributions using mixture models. It is a procedure that iterates around the expectation (E) and maximization (M) steps. In the E-step for clustering, the algorithm calculates the posterior probability, h_{ij}, that a sample j belongs to a cluster C_i:

$$h_{ij} = P(C_i \mid D_j) = \frac{p(D_j \mid \theta_i)\pi_i}{\sum_{m=1}^{C} p(D_j \mid \theta_m)\pi_m}, \tag{2}$$

where π_i is the weight for the ith mixture component, D_j is the measurement, and θ_i is the set of parameters for each density functions. In the M-step, the EM algorithm searches for optimal parameters that maximize the sum of weighted log-likelihood probabilities. EM automatically selects the number of clusters by maximizing the logarithm of the likelihood of future data. Observations that belong to the same pattern type should ideally be assigned to a single cluster.

2.3.2. Classification (process 8)

We use a traditional classification technique, called a decision tree learner. Each internal node of a decision tree represents an attribute, and each leaf node represents a class label. We used J48 in WEKA, which is an implementation of C4.5 algorithm [27]. The input to the decision tree learner for the present study consisted of a pattern factor metrics vector of dimension 32, representing the 32 statistical metrics (Appendix C). Cluster labels were used as classification labels. The labeled data set was recursively partitioned into small subsets as the tree was being built. If the data instances in the same subset were assigned to the same label (class), the tree building process was terminated. We then derived If-Then rules from the resulting decision tree and compared them with expert-generated rules.

3. APPLICATION: VISUAL WORD PROCESSING

The ERP data for this study consisted of 144 observations (36 subjects $\times$4 experiment conditions) that were acquired in a lexical decision task (see [28] for details). Participants viewed word and pseudoword stimuli that were presented, one stimulus at a time, in the center of a computer monitor and made word/nonword judgments to each stimulus using their right index and middle fingers to depress the "1" and "2" keys on a keyboard ("yes" key counterbalanced across subjects). Stimuli consisted of 350 words and word-like stimuli, including low-frequency words that were familiar to subjects (based on pretesting) and rare words like "nutant" (which were unlikely to be known by participants). Letters were lower-case Geneva black, 26 dpi, presented foveally on a white screen. Words and nonwords were matched in mean length and orthographic neighborhood [29, 30].

3.1. ERP experiment data

ERP data were recorded using a 128-channel electrode array, with vertex recording reference [31]. Data were sampled at a rate of 250 per second and were amplified with a 0.01 Hz highpass filter (time constant ~10 seconds). The raw EEG was segmented into 1500 milliseconds epochs, starting 500 milliseconds before onset of the target word. There were four conditions of interest: correctly classified, low-frequency words (*Known*); correctly classified rare words (*RareHits*), rare words rated as nonwords (*RareMisses*); and correctly classified nonwords (*Pseudo*).

Segments were marked as bad if they contained ocular artifacts (EOG $> 70\,\mu$V), or if more than 20% of channels were bad on a given trial. The artifact-contaminated trials were excluded from further analysis.

Segmented data were averaged across trials (within subjects and within conditions) and digitally filtered with a 30-Hz lowpass filter. After further channel and subject exclusion,

bad (excluded) channels were interpolated. The data re-referenced to the average of the recording sites [32], using a polar average reference to correct for denser sampling over superior, as compared with inferior, scalp locations [33, 34]. Data were averaged across individual subjects, and the resulting "grand-averaged" ERPs were used for inspection of waveforms and topographic plots.

4. TPCA AUTOLABELING RESULTS

Temporal PCA (tPCA) was used to transform the ERP data into a set of latent temporal patterns (see Section 2.2.1 for details). We extracted the first 15 latent factors from each of the four datasets, accounting for approximately 80% of the total variance. These 15 tPCA factors were then subjected to a Promax rotation.

After the tPCA factors were projected back into the original data space (Section 2.2.1), we applied our expert-defined rules to determine the percentage of observations that matched each target pattern. Results are shown in Table 2.

We assigned labels to the first 10 factors based on the correspondence between the target patterns and the tPCA factors. Results were as follows: Factor 4 = P100, Factor 3 = N100, Factor = N2, Factor 7 = N3/P1r, Factor 2 = MFN/N4, and Factor 9 = P3. Figure 3 displays the time course and topography for these six *pattern factors*.

Note that many patterns showed splitting across two or more factors. This may reflect misallocation of pattern variance across the factors (i.e., inaccuracies in the tPCA decomposition), inaccuracies in rule definitions, or both. A complementary problem is seen in the case of factors 2, 7, and 10, which show matches to more than one target pattern. Again, this may reflect misallocation of variance. Alternatively, these results may suggest a need to refine our pattern descriptions, the rules that are used to identify pattern instances, or both. In either case, these findings point to the need for systematic evaluation of results. Diagnosing potential sources of error is the first step towards systematic improvements of methods.

4.1. Evaluation of top-down methods

In our framework, top-down methods for pattern classification are dependent on the accuracy of both the data summary methods and the expert-defined rules. In particular,

(1) data summary methods should yield discrete patterns that reflect different underlying neuropsychological processes, or "components;"

(2) rules that are applied to summary metrics should be implemented in a way that effectively discriminates between separate patterns.

Our initial efforts have led to encouraging classification results, as illustrated above. However, several findings suggest the need to consider possible misallocation of variance in the data summary process, and ways of optimizing pattern rules.

4.1.1. Diagnosing misallocation of variance

A well-known critique of PCA methods, including temporal PCA, is that inaccuracies in the decomposition can lead to misallocation of variance ([21, 35]). For example, in our results, the left temporal N3 and parietal P1r patterns were both assigned to a single factor (cf. [15] for similar results). Recent methods can achieve separation of patterns that have been confounded in an initial PCA (see [19] for a discussion). A more serious problem is that of the pattern splitting: well-known patterns like the P100 are expected to map to a single rule (factor). Indeed, this simple mapping was obtained in 3 or our 4 pilot datasets [23]. Splitting of the P100 across two factors therefore suggests a possible misallocation of variance in the tPCA. A future challenge will be to develop rigorous methods of diagnosing misallocation of variance in the decomposition of ERPs. In the final section, we consider alternatives to tPCA, which may address this issue.

4.1.2. Comparison with a "gold standard"

The validity of our tPCA autolabeling procedures was assessed by comparing autolabeling results with a "gold standard," which was developed through manual labeling of patterns. Two ERP analysts visually inspected the raw ERPs for each subject and each condition. For each target pattern, the analysts indicated whether the pattern was present, based on inspection of temporal data (waveforms, butterfly plots) and spatial data (topography at time of peak activity in pattern interval). Analysts also provided confidence ratings and rated the typicality of each pattern instance using a 3-point scale.

An initial set of ratings on 100 observations (25 subjects $\times$4 conditions) was collected. Raters met to discuss results and to calibrate procedures for subsequent ratings. Experts then proceeded to label another 116 ERP observations (4 observations were omitted due to a technical error in the data file). This set of labeled data constituted the "gold standard" for system evaluation.

Interrater reliability for test data was computed for two of the patterns (P100 and N100) using the Spearman-Brown prophecy coefficient [36]. Results are graphed in Table 3 ("*" = moderate reliability, "**" = high reliability).

For both patterns, the highest level of reliability was reflected in the typicality ratings. In addition, reliability was considerably higher for the P100 pattern. Inspection of the data revealed that the low reliability for N100 "presence" judgments was due to a systematic difference in use of categories: one rater consistently rated as "not present" cases where the other rater indicated the pattern was "present" but atypical ("1" on typicality scale).

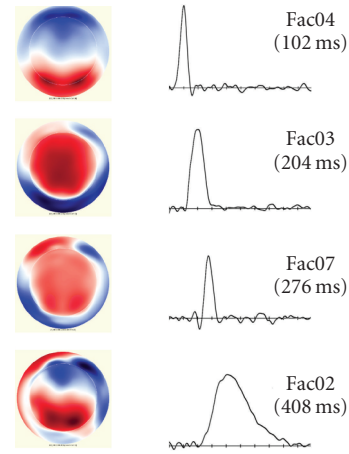
Accuracy of the autolabeling procedures was defined as the percentage of system labels that matched the gold-standard labels (%Agr; see Table 4). Across the eight patterns, the autolabeling results and expert ratings had an averaged Pearson r correlation of +.36. This leads to an effective interrater reliability of +.52 as measured by the Spearman-Brown formula. Note that while the %Agr was relatively high for the

Table 2: Percentage of ERP observations for each factor that matched expert-defined rule criteria.

	% Observations meeting pattern criteria							
Factor	P100	N100	N2	N3	P1r	MFN	N4	N3
Fac#01	—	—	—	—	—	—	—	—
Fac#02	—	—	—	—	—	**36.81**	**9.72**	59.72
Fac#03	—	**82.64**	—	—	—	—	—	—
Fac#04	**82.64**	—	—	—	—	—	—	—
Fac#05	—	—	—	—	—	—	—	—
Fac#06	—	—	—	—	—	—	—	—
Fac#07	—	—	69.44	**42.36**	**64.58**	22.92	—	—
Fac#08	34.72	—	—	—	—	—	—	—
Fac#09	—	—	—	—	—	—	—	**56.94**
Fac#10	—	51.39	**51.39**	—	—	—	—	—
Fac#11	—	—	—	47.92	25.69	34.03	35.42	—
Fac#12	—	—	—	—	—	—	—	—
Fac#13	—	—	—	59.03	62.50	40.97	—	—
Fac#14	—	—	—	—	—	—	—	—
Fac#15	—	—	—	—	—	—	—	9.72

Pattern	Fac01	Fac02	Fac03	Fac04	Fac05	Fac06	Fac07
P1	-	-	-	82.64	-	-	-
N1	-	-	82.64	-	-	-	-
N2	-	-	-	-	-	-	69.44
N3	-	-	-	-	-	-	42.36
P1r	-	-	-	-	-	-	64.58
MFN	-	36.81	-	-	-	-	22.92
N4	-	9.72	-	-	-	-	-
P3	-	59.72	-	-	-	-	-

(a)



(b) (c)

Figure 3: Autoclassification and labeling results. (a) Percentage of observations matching rule criteria for each pattern. (b) Topogragraphy and (c) time course of pattern factors.

Table 3: Interrater reliability (Spearman-Brown r).

	Presence	*Confidence*	*Typicality*
P100	.51*	.41*	.72**
N100	−.04	.35*	.45*

Table 4: Comparison of autolabeling with expert labels.

Pattern	Person r	Spearman-Brwon	%Agr
P100	0.60	0.75	0.90
N100	0.26	0.41	0.84
N2	0.12	0.21	0.53
N3	0.41	0.58	0.63
P1r	0.47	0.64	0.76
MFN	0.33	0.49	0.40
N4	0.37	0.54	0.81
P3	0.30	0.46	0.64

N100 (0.84), the Spearman-Brown coefficient was considerably lower (0.41), consistent with the lower interrater reliability observed between ERP analysts for this pattern.

5. DATA MINING RESULTS

Input to the data mining ("bottom-up") analyses consisted of 32 metrics for each factor, weighted across each of the 144 labeled observations (total $N = 4608$). Pattern labels for each observation were a combination of the autolabeling results (pattern present versus pattern absent for each

factor, for each observation), combined with typicality ratings, as follows. Observations that met the rule criteria ("pattern present" according to autolabeling procedures) and were rated as "typical" (rating > "1") were assigned to one category label. Observations that *either* failed to meet pattern criteria ("pattern absent") *or* were rated as atypical ("1" on rating scale), or both, were assigned to a second category. The combined labels were used to capitalize on the high reliability and greater sensitivity of the typicality + presence/absence ratings, as compared with the presence/absence labels by themselves.

For the EM procedures, we set the number of clusters to be 9 (8 patterns + nonpatterns). We then clustered the 144 observations derived from the pattern factors, based on the 32 metrics. As shown in Table 5, the assignment of observations to each of the 9 clusters largely agreed with the results from the top-down (autolabeling) procedures (compare Table 2).

Ideally, each cluster will correspond to a unique ERP pattern. However, as noted above, inaccuracies in either the data summary (tPCA) procedures, or the expert rules, or both, can lead to pattern splitting. Thus, it is not surprising that patterns in our clustering analysis were occasionally assigned to two or more clusters. For instance, the P100 pattern split into two clusters (clusters 4 and 5), consistent with the autolabeling results (Table 2).

Supervised learning (decision tree) methods were used to derive pattern rules, independently of expert judgments. According to the information gain rankings of the 32 attributes, *TI-max* and *In-mean(ROI)* were most important, consistent with our previous results [22]. These findings validate the use of these two metrics in expert-defined rules. Decision trees revealed the importance of additional spatial metrics, suggesting the need for finer-grained characterization of pattern topographies in our rule definitions. In addition, difference measures (*Pseudo-RareMisses* and *RareMisses-RareHits*) were highly ranked for certain patterns (the N2 and P300, resp.), suggesting that functional metrics may be useful for classification of certain target patterns.

6. CONCLUSION

The goal of this study was to define high-level rules and concepts for ERP components in a particular domain (visual word recognition) and to design, evaluate, and optimize an automated data processing and labeling stream that implements these rules and concepts. By combining rule definitions based on expert knowledge (top-down approach) with rule definitions that are generated through data mining (bottom-up approach), we predicted that our system would achieve higher accuracy than a system based on either approach in isolation. Results suggest that the combination of top-down and bottom-up methods is indeed synergistic: while domain knowledge was used effectively to constrain the number of clusters in the data mining, decision tree classifiers revealed the importance of additional metrics, including multiple measures of topography and, for certain patterns, functional metrics that correspond to experiment manipulations.

Ongoing work is focused on the following goals:

(i) refinement of procedures for expert labeling of patterns in the "raw" (untransformed) ERP data;
(ii) testing of alternative data summary and autolabeling methods;
(iii) modification of rules and concepts, based on integration of bottom-up and top-down classification methods.

6.1. Alternative data summary procedures

In the present study, we applied temporal PCA (tPCA) to decompose ERP data into discrete patterns that are input to our automated component classification and labeling process. PCA is a useful approach because it is automated, is data-driven, and has been validated and optimized for decomposition of event-related potentials [21]. At the same time, as illustrated here, PCA is prone to misallocation of variance across the latent factors. Further, differences in the time course of patterns across subjects and experiment conditions are a particular problem for tPCA methods: latency "jitter" can lead to mischaracterization of patterns [7].

For this reason, we are currently testing alternative approaches to ERP component analysis. One approach involves application of sequential (temporo-spatial) PCA. Temporo-spatial PCA is a refinement and extension of temporal PCA (see [12, 19] for details). The factor scores from the temporal PCA, which quantify the extent to which their respective latent factors are present in the ERP data, undergo a spatial PCA. The spatial PCA further decomposes the factor scores into a second tier of latent factors that capture correlations between channels across subjects and conditions. The latent factors from the two decompositions are then combined to yield a finer decomposition of the patterns of variance that are present in the ERP data.

6.1.1. Windowed analysis of ERPs

The second approach is to adopt the traditional methods of parsing ERP data into discrete temporal "windows" for analysis. By focusing on temporal windows corresponding to known ERP patterns, the algorithms we developed for extracting statistics from the tPCA factors can be extended to the raw ERP, with some modification. While the raw ERP is more complex, with overlapping temporo-spatial patterns, the autolabeling process applied to raw ERPs would correspond directly to the expert "gold standard" labeling procedure. Furthermore, it would not be subject to one weakness of tPCA, namely, that the time courses of the factor loadings are invariant across subjects and conditions.

6.1.2. Microstate analysis

We are also evaluating the use of microstate analysis, an approach to ERP pattern segmentation that was introduced by Lehmann and Skrandies [37]. Microstate analysis is a data parsing technique that partitions the ERP into windows based upon characteristics of its evolving topography.

Table 5: EM clustering results (NP: nonpatterns).

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
P100	0	0	0	0	**60**	**49**	0	0	0
N100	1	0	0	0	0	0	7	30	**77**
N2	**104**	0	0	0	17	0	0	3	8
N3	5	0	0	0	4	2	2	**40**	1
P1r	11	0	14	0	14	6	5	**51**	0
MFN	0	0	0	**56**	0	9	0	0	0
N4	0	0	0	**15**	0	1	0	0	0
P3	0	**113**	0	2	0	0	0	0	0
NP	26	28	22	197	39	16	**33**	64	20

Consecutive time slices, whose topographies are similar under a metric, such as global map similarity, are grouped together into a single microstate. This microstate in turn corresponds to a distinct distribution of neuronal activity. Microstate analysis may hold promise for separating ERP components that have minimal temporal overlap. Moreover, this method has been implemented as a fully automated process (see [38] for downloadable software and [39, 40] for discussion of automated segmentation using microstate analysis).

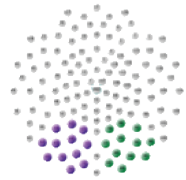
6.2. *Development of neural electromagnetic ontologies (NEMO)*

In previous work [22] we have described progress on the design of a domain ontology mining framework and its application to EEG data and patterns. This represents a first step in the development of Neural ElectroMagnetic Ontologies (NEMO). The tools that are developed for the NEMO project can be used to support data management and pattern analysis within individual research labs. Beyond this goal, ontology-based data sharing can support collaborative research that would advance the state of the art in EM brain imaging, by allowing for large-scale metaanalysis and high-level integration of patterns across experiments and imaging modalities. Given that researchers currently use different concepts to describe temporal and spatial data, ontology development will require us to develop a common framework to support spatial and temporal references.

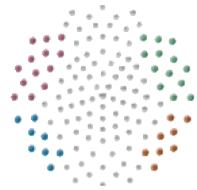
A practical goal for the NEMO project is to build a merged ERP-ERF ontology for the reading and language domain. This accomplishment would demonstrate the utility of ontology-based integration of averaged EEG and MEG measures, and make strong contributions to the advancement of multimodal neuroinformatics. To accomplish this goal, we have developed concurrent strategies for representation of ERP and ERF data in sensor space and in source (anatomical) space. To link to these ontology databases and to support integration of EM measures with results from other neuroimaging techniques, we are working to extend our pattern classification process to brain-based coordinate systems, through application of source analysis to dense-array EEG and whole-head MEG datasets.

APPENDICES

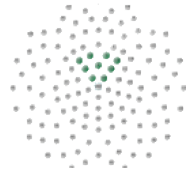
A. CHANNEL GROUPINGS FOR SPATIAL METRICS (REGIONS OF INTEREST—ROIS)



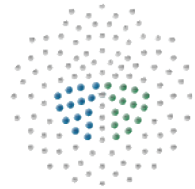
Left occipital	77, 78, 83, 84, 85, 86, 89, 90, 91, 92, 95, 96
Right occipital	59, 60, 64, 65, 66, 67, 69, 70, 71, 72, 74, 75



Left anterotemporal	27, 28, 33, 34, 35, 39, 40, 41, 44, 45, 46, 49, 128
Right anterotemporal	1, 2, 109, 110, 114, 115, 116, 117, 120, 121, 122, 123, 125
Left posterotemporal	50. 56, 57, 58, 63, 64 65, 69
Right posterotemporal	91, 96, 97, 100, 101, 102, 108



Medial frontal	5, 6, 7, 12, 13, 21 107, 113, 119



Left parietal	7, 31, 32, 37, 38, 42, 43, 48, 52, 53, 54, 60, 61, 67
Right parietal	78, 79, 80, 81, 86, 87, 88, 93, 94, 99, 104, 105, 106, 107

B. ERP PATTERN RULES HYPOTHESIZED FOR VISUAL WORD RECOGNITION

Rule #1 (pattern PT_1 = P100)

Let ROI = occipital (average of left and right occipital). For any n, $FA_n = PT_1$ *iff*

(i) 60 ms < TI-max (FA_n) ≤ 150 ms AND
(ii) |IN-mean(ROI) | ≥ .4 mV AND
(iii) IN-mean(ROI) > 0.

Table 6

	Metric	Description
Function	Pseudo-known	Difference in mean intensity over ROI at time of peak latency (Nonwords-Words)
	RareMisses-RareHits	Difference in mean intensity over ROI at time of peak latency (RareMisses-RareHits)
	RareHits-Known	Difference in mean intensity over ROI at time of peak latency (RareHits-Known)
	Pseudo-RareMisses	Difference in mean intensity over ROI at time of peak latency (Nonwords-RareMisses)
Intensity	IN-max	Maximum intensity (in microvolts) at time of peak latency
	IN-max to Baseline	Maximum intensity (in microvolts) at time of peak latency with respect to intensity at TI-begin
	IN-min	Maximum intensity (in microvolts) at time of peak latency
	IN-min to Baseline	Maximum intensity (in microvolts) at time of peak latency with respect to intensity at TI-begin
	SP-max	Channel associated with maximum intensity, IN-max
	SP-max ROI	Channel group (ROI) containing SP-max
	SP-min	Channel associated with manimum intensity, IN-min
	SP-min ROI	Channel group (ROI) containing SP-min
Space	IN-mean ROI	Mean intensity (in microvolts) at time of peak latency for a specified channel group
	IN-LOCC	Mean intensity (in microvolts) at time of peak latency for left occipital channel group
	IN-ROCC	Mean intensity (in microvolts) at time of peak latency for right occipital channel group
	IN-LPAR	Mean intensity (in microvolts) at time of peak latency for left parietal channel group
	IN-RPAR	Mean intensity (in microvolts) at time of peak latency for right parietal channel group
	IN-LPTEM	Mean intensity (in microvolts) at time of peak latency for left posterior temporal channel group
	IN-RPTEM	Mean intensity (in microvolts) at time of peak latency for right posterior temporal channel group
	IN-LATEM	Mean intensity (in microvolts) at time of peak latency for left anterior temporal channel group
	IN-RATEM	Mean intensity (in microvolts) at time of peak latency for right anterior temporal channel group
	IN-LORB	Mean intensity (in microvolts) at time of peak latency for left orbital channel group
	IN-RORB	Mean intensity (in microvolts) at time of peak latency for right orbital channel group
	IN-LFRON	Mean intensity (in microvolts) at time of peak latency for left frontal channel group
	IN-RFRON	Mean intensity (in microvolts) at time of peak latency for right frontal channel group
	SP-cor	Correlation between factor topography and topography of target pattern
Time	TI-max	Latency (in milliseconds) of maximum or minimum amplitude
	TI-begin	Onset (in milliseconds) of waveform excurstion containing peak intensity
	TI-end	Conclusion (in milliseconds) of waveform excurstion containing peak intensity
	TI-duration	Duration (in milliseconds) of pattern, equal to TI-begin minus TI-end

Rule #2 (pattern PT_2 = N100)

Let ROI = occipital (average of left and right occipital). For any n, $FA_n = PT_2$ *iff*

(i) 151 ms < TI-max (FA_n) ≤ 229 ms AND

(ii) |IN-mean(ROI)| ≥ .4 mV AND

(iii) IN-mean(ROI) < 0.

Rule #3 (pattern PT_3 = N2)

Let ROI = occipital-temporal (average of occipital, posterior temporal). For any n, $FA_n = PT_3$ *iff*

(i) 230 ms < TI-max (FA_n) ≤ 300 ms AND

(ii) |IN-mean(ROI)| ≥ .4 mV AND

(iii) IN-mean(ROI) < 0.

Rule #4 (pattern PT_4 = N3)

Let ROI = left anterior temporal. For any n, $FA_n = PT_4$ *iff*

(i) 250 ms < TI-max (FA_n) ≤ 400 ms AND

(ii) |IN-mean(ROI)| ≥ .4 mV AND

(iii) IN-mean(ROI) < 0.

Rule #5 (pattern PT_5 = P1r)

Let ROI = parietal temporal (average of left parietal, right parietal) For any n, $FA_n = PT_5$ *iff*

(i) 250 ms ≥ TI-max (FA_n) ≤ 400 ms AND

(ii) |IN-mean(ROI)| ≥ .4 mV AND

(iii) IN-mean(ROI) > 0.

Rule #6 (pattern PT_6 = MFN)

Let ROI = frontocentral (average of left frontocentral, right frontocentral) For any n, $FA_n = PT_6$ *iff*

(i) 250 ms < TI-max (FA_n) ≤ 450 ms AND
(ii) |IN-mean(ROI)| ≥ .4 mV AND
(iii) IN-mean(ROI) < 0.

Rule #7 (pattern PT_7 = N4)

Let ROI = parietal temporal (average of left parietal, right parietal) For any n, $FA_n = PT_7$ *iff*

(i) 350 ms < TI-max (FA_n) ≤ 550 ms AND
(ii) |IN-mean(ROI)| ≥ .4 mV AND
(iii) IN-mean(ROI) < 0.

Rule #8 (pattern PT_8 = P300)

Let ROI = parietal temporal (average of left parietal, right parietal) For any n, $FA_n = PT_8$ *iff*

(i) 401 ms ≥ TI-max (FA_n) ≤ 700 ms AND
(ii) |IN-mean(ROI)| ≥ .4 mV AND
(iii) IN-mean(ROI) > 0.

C. STATISTICAL METRICS

For statistical metrics see Table 6.

ACKNOWLEDGMENTS

This work was supported by a grant to the Oregon NeuroInformatics Center from the Brain, Biology, and Machine Initiative, Department of Defense, Telemedicine Advanced Technology Research Center (TATRC), DAMD17-01-1-0750, "Acquisition of the Oregon ICONIC Grid for Integrated Cognitive Neuroscience, Informatics, and Computation," NSF Major Research Instrumentation, NSF BCS-0321388.

REFERENCES

[1] W. Klimesch, "Memory processes, brain oscillations and EEG synchronization," *International Journal of Psychophysiology*, vol. 24, no. 1-2, pp. 61–100, 1996.
[2] T.-P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and T. J. Sejnowski, "Analysis and visualization of single-trial event-related potentials," *Human Brain Mapping*, vol. 14, no. 3, pp. 166–185, 2001.
[3] M. Fabiani, G. Gratton, and M. G. H. Coles, "Event-related brain poten-tials: methods, theory, and applications," in *Handbook of Psychophysiology*, J. Cacioppo, L. Tassinary, and G. Berntson, Eds., chapter 3, pp. 53–84, Cambridge University Press, New York, NY, USA, 2000.
[4] A. M. Proverbio and A. Zani, "Electromagnetic manifestations of mind and brain," in *The Cognitive Electrophysiology of Mind and Brain*, A. Zani and A. M. Proverbio, Eds., chapter 2, pp. 13–37, Academic Press, New York, NY, USA, 2002.
[5] T. Gasser, J. C. Schuller, and U. S. Gasser, "Correction of muscle artefacts in the EEG power spectrum," *Clinical Neurophysiology*, vol. 116, no. 9, pp. 2044–2050, 2005.
[6] O. Hauk, M. H. Davis, M. Ford, F. Pulvermuller, and W. D. Marslen-Wilson, "The time course of visual word recognition as revealed by linear regression analysis of ERP data," *NeuroImage*, vol. 30, no. 4, pp. 1383–1400, 2006.
[7] K. Spencer, "Averaging, detection, and classification of single-trials ERPs," in *Event-Related Potentials: A Methods Handbook*, T. Handy, Ed., pp. 209–228, MIT Press, Cambridge, Mass, USA, 2005.
[8] E. Donchin and E. Heffley, "Multivariate analysis of event-related potential data: a tutorial review," in *Multidisciplinary Perspectives in Event-Related Brain Potential Research*, D. Otto, Ed., pp. 555–572, U.S. Government Printing Office, Washington, DC, USA, 1978.
[9] T. W. Picton, S. Bentin, P. Berg, et al., "Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria," *Psychophysiology*, vol. 37, no. 2, pp. 127–152, 2000.
[10] P. L. Nunez, *Electric Fields of the Brain: The Neurophysics of EEG*, Oxford University Press, New York, NY, USA, 1981.
[11] G. Gratton, M. G. H. Coles, and E. Donchin, "A procedure for using multi-electrode information in the analysis of components of the event-related potential: Vector filter," *Psychophysiology*, vol. 26, no. 2, pp. 222–232, 1989.
[12] K. M. Spencer, J. Dien, and E. Donchin, "A componential analysis of the ERP elicited by novel events using a dense electrode array," *Psychophysiology*, vol. 36, no. 3, pp. 409–414, 1999.
[13] T. Handy, "Basic principles of ERP quantification," in *Event-Related Potentials: A Methods Handbook*, T. Handy, Ed., pp. 33–56, MIT Press, Cambridge, Mass, USA, 2005.
[14] A. C. Nobre and G. McCarthy, "Language-related ERPs: scalp distributions and modulation by word type and semantic priming," *Journal of Cognitive Neuroscience*, vol. 6, no. 3, pp. 233–255, 1994.
[15] J. Dien, G. A. Frishkoff, A. Cerbone, and D. M. Tucker, "Parametric analysis of event-related potentials in semantic comprehension: evidence for parallel brain mechanisms," *Cognitive Brain Research*, vol. 15, no. 2, pp. 137–153, 2003.
[16] G. A. Frishkoff, "Hemispheric differences in strong versus weak semantic priming: evidence from event-related brain potentials," *Brain and Language*, vol. 100, no. 1, pp. 23–43, 2007.
[17] P. E. Compton, P. Grossenbacher, M. I. Posner, and D. M. Tucker, "A cognitive-anatomical approach to attention in lexical access," *Journal of Cognitive Neuroscience*, vol. 3, no. 4, pp. 304–312, 1991.
[18] K. Huang, K. Itoh, S. Suwazono, and T. Nakada, "Electrophysiological correlates of grapheme-phoneme conversion," *Neuroscience Letters*, vol. 366, no. 3, pp. 254–258, 2004.
[19] J. Dien and G. A. Frishkoff, "Introduction to principal components analysis of event-related potentials," in *Event-Related Potentials: A Methods Handbook*, T. Handy, Ed., pp. 189–208, MIT Press, Cambridge, Mass, USA, 2005.
[20] J. Dien, "PCA toolbox (version 1.093)," Lawrence, Kan, USA, October 2004.
[21] J. Dien, "Addressing misallocation of variance in principal components analysis of event-related potentials," *Brain Topography*, vol. 11, no. 1, pp. 43–55, 1998.
[22] D. Dou, G. Frishkoff, J. Rong, R. M. Frank, A. Malony, and D. M. Tucker, "Development of NeuroElectroMagnetic Ontologies (NEMO): a framework for mining brainwave ontologies," in *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '07)*, pp. 270–279, San Jose, Calif, USA, August 2007.
[23] J. Rong, D. Dou, G. A. Frishkoff, R. M. Frank, A. Malony, and D. M. Tucker, "A semi-automatic framework for mining ERP

patterns," in *Proceedings of the 21st International Conference on Advanced Information Networking and Applications Workshops (AINAW '07)*, vol. 1, pp. 329–334, Niagara Falls, Canada, May 2007.

[24] A. Rodriguez-Fornells, B. M. Schmitt, M. Kutas, and T. F. Münte, "Electrophysiological estimates of the time course of semantic and phonological encoding during listening and naming," *Neuropsychologia*, vol. 40, no. 7, pp. 778–787, 2002.

[25] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[26] Weka 3, "Data Mining Software in Java," http://www.cs.waikato.ac.nz/ml/weka/.

[27] J. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann, San Mateo, Calif, USA, 1993.

[28] G. A. Frishkoff, C. Perfetti, and C. Westbury, "ERP measures of partial semantics knowledge: left temporal indices of skill differences and lexical quality," *Biological Psychology*, in revision.

[29] D. A. Medler and J. R. Binder, "MCWord: an online orthographic database of the English language," 2005, http://www.neuro.mcw.edu/mcword/.

[30] M. D. Wilson, "The MRC psycholin-guistic database: machine readable dictionary," *Behavioural Research Methods, Instruments and Computers*, vol. 20, no. 1, pp. 6–11, 1988.

[31] Electrical Geodesics (EGI), "Eugene, Oregon," http://www.egi.com/.

[32] J. Dien, "Issues in the application of the average reference: review, critiques, and recommendations," *Behavior Research Methods, Instruments, and Computers*, vol. 30, no. 1, pp. 34–43, 1998.

[33] M. Junghöfer, T. Elbert, D. M. Tucker, and C. Braun, "The polar average reference effect: a bias in estimating the head surface integral in EEG recording," *Clinical Neurophysiology*, vol. 110, no. 6, pp. 1149–1155, 1999.

[34] M. Junghöfer, T. Elbert, D. M. Tucker, and B. Rockstroh, "Statistical control of artifacts in dense array EEG/MEG studies," *Psychophysiology*, vol. 37, no. 4, pp. 523–532, 2000.

[35] G. McCarthy and C. C. Wood, "Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models," *Electroencephalography and Clinical Neurophysiology*, vol. 62, no. 3, pp. 203–208, 1985.

[36] R. Rosenthal and R. Rosnow, *Essentials of Behavioral Research: Methods and Data Analysis*, McGraw-Hill, New York, NY, USA, 2nd edition, 1991.

[37] D. Lehmann and W. Skrandies, "Spatial analysis of evoked potentials in man—a review," *Progress in Neurobiology*, vol. 23, no. 3, pp. 227–250, 1984.

[38] "Cartool software," Functional Brain Mapping Laboratory, Geneva, Switzerland, http://brainmapping.unige.ch/Cartool.htm.

[39] T. Koenig, K. Kochi, and D. Lehmann, "Event-related electric microstates of the brain differ between words with visual and abstract meaning," *Electroencephalography and Clinical Neurophysiology*, vol. 106, no. 6, pp. 535–546, 1998.

[40] T. Koenig and D. Lehmann, "Microstates in language-related brain potential maps show noun-verb differences," *Brain and Language*, vol. 53, no. 2, pp. 169–182, 1996.

Hindawi Publishing Corporation
Computational Intelligence and Neuroscience
Volume 2007, Article ID 10479, 24 pages
doi:10.1155/2007/10479

Research Article

Statistical Modeling and Analysis of Laser-Evoked Potentials of Electrocorticogram Recordings from Awake Humans

Zhe Chen,[1,2] Shinji Ohara,[3,4] Jianting Cao,[1,5] François Vialatte,[1] Fred A. Lenz,[3] and Andrzej Cichocki[1]

[1] *Laboratory for Advanced Brain Signal Processing, RIKEN Brain Science Institute, Wako-shi 351-0198, Saitama, Japan*
[2] *Neuroscience Statistics Research Lab, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA*
[3] *Department of Neurosurgery, Johns Hopkins Hospital, Baltimore, MD 21287-7274, USA*
[4] *Department of Neurosurgery, Kyoto Kizugawa Hospital 26-1, Nishi-Rokutan, Hirakawa Joyo 610-0101, Kyoto, Japan*
[5] *Department of Electronics and Information Engineering, Saitama Institute of Technology, Fukaya-shi 369-0293, Saitama, Japan*

Correspondence should be addressed to Zhe Chen, zhechen@neurostat.mgh.harvard.edu

Received 25 December 2006; Accepted 18 May 2007

Recommended by Saied Sanei

This article is devoted to statistical modeling and analysis of electrocorticogram (ECoG) signals induced by painful cutaneous laser stimuli, which were recorded from implanted electrodes in awake humans. Specifically, with statistical tools of factor analysis and independent component analysis, the pain-induced laser-evoked potentials (LEPs) were extracted and investigated under different controlled conditions. With the help of wavelet analysis, quantitative and qualitative analyses were conducted regarding the LEPs' attributes of power, amplitude, and latency, in both averaging and single-trial experiments. Statistical hypothesis tests were also applied in various experimental setups. Experimental results reported herein also confirm previous findings in the neurophysiology literature. In addition, single-trial analysis has also revealed many new observations that might be interesting to the neuroscientists or clinical neurophysiologists. These promising results show convincing validation that advanced signal processing and statistical analysis may open new avenues for future studies of such ECoG or other relevant biomedical recordings.

1. INTRODUCTION

Pain is an essential function for the organism to enable immediate awareness of actual or threatening injury for further adopting a self-protective behavior. Roughly speaking, pain is a complex and subjective experience in the brain; it involves sensory, affective, cognitive, and motivational components and is associated with autonomous activity, nocifensive reflexes and reactions. In clinical practice, neurophysiological evaluation of pain in humans has been an important subject of research in the last decade (Bromm and Lorenz [1]).

In the literature, there are many approaches for monitoring and measuring the pain-related brain activities, including electroencephalogram (EEG), magnetoencephalogram (MEG), and fMRI. In particular, the electrocorticogram (ECoG) records directly the cortical (electrical) activities from subdural electrode grids that are implanted in the human subjects for collecting information for surgical treatments of medically intractable epilepsy (i.e., patients in the hospital upon approval). As an invasive recording tool, ECoG offers some superior features that are unavailable for EEG or MEG recordings. Specifically, unlike EEG that measures the electrical potentials recorded from the scalp, ECoG directly records the potentials from the cortical surface, thereby having a higher signal-to-noise ratio (SNR) and higher spatial resolution (because of closer electrode spacing). Consequently, activities in beta or gamma bands are better recorded in ECoG due to less spatial summation and phase cancelation (or high-cut filter effect) than in scalp EEG recordings.

Since the energy of the infrared laser can be used to produce a brief thermal stimulus applied to the skin such as to selectively activate the skin nociceptor, the recordings of brain responses to short laser pulses (the so-called laser-evoked potentials, or LEPs) have increasingly become a useful method for evaluating the function of central nociceptive pathways. The roles of LEPs for detecting abnormalities in patients have been noted (García-Larrea et al. [2]). Generally, there are two or three major peaks in the pain-evoked LEPs, which may be generated in multiple regions. In the literature, most research efforts focused on two peaks of the

LEPs, the so-called N2 and P2, which correspond to the vertex negative-positive complex.[1] The timing when the peak of the LEP appears is referred to the latency of LEPs. Typically, N2 was found around 150–400 milliseconds, and P2 was found around 230–500 milliseconds, depending on the laser pulse duration and intensity, as well as the stimulus site or area (Bromm and Lorenz [1]). The difference in latency is essentially related to the response differences in peripheral conduction distance. Specifically, LEP reflects an integrative cortical response to the painful laser stimuli rather than a simple reaction of the sensory cortex to it; thus, in the healthy subject the amplitude of cortical LEPs correlates with the subjective sensation of pain, rather than with the physical stimulus intensity (García-Larrea et al. [2]). For instance, paying attention to the laser stimulus simultaneously increases the subjective pain sensation and the LEP amplitude, both of which decrease in turn when the subject is distracted from the stimulus (García-Larrea et al. [3]). In addition to the amplitude, the latencies of the LEPs are often important for the neurophysiological evaluation of pain (Bromm and Lorenz [1]). In a later section, we will analyze the amplitudes and latencies of LEP components N2 and P2 in detail. As suggested in the literature, the negative component (N2) seems to be induced mainly by the activation in the bilateral operculoinsular cortices and contralateral primary somatosensory cortex (SI) (e.g., Tarkka and Treede [4], Iannetti et al. [5]), and the positive component (P2) is mainly generated by the cingulate gyrus (e.g., Tarkka and Treede [4], Lenz et al. [6], Iannetti et al. [5]). However, it should also be noted that both N2 and P2 could be recorded and observed at multiple cortical regions simultaneously (e.g., Ohara et al. [7–9]); therefore, although there may be some evidence that one LEP is more related to a particular region than the other, a full understanding of their underlying mechanisms remains unclear.

In the previous studies (Ohara et al. [8]) of the ECoG recordings from the awake humans, it was found that attention to painful cutaneous laser stimuli enhances pain-related LEPs in cortical regions receiving nociceptive input, typically at multiple cortical sites (Ohara et al. [9]). Specifically, it was observed that at primary somatosensory (SI), parasylvian (PS), and medial frontal (MF: anterior cingulate and supplementary motor area) cortex areas, the amplitudes of the negative ($N2^*$) and positive ($P2^{**}$) LEP components[2] were enhanced by attention to (counting stimuli), in comparison with distraction from the stimuli (reading for comprehension). It was suggested therein that attention controls both early ($N2^*$) and late ($P2^{**}$) pain-related input to SI (and other) cortical regions, while the late positive deflections (that follow the $P2^{**}$ peak) are specifically related to attention. It was also reported in other independent EEG studies (e.g., Legrain et al. [10, 11]) that LEPs can be modulated by selective spatial attention. In [7], Ohara et al. observed that attention to painful stimuli leads to enhanced event-related desynchronization (ERD) in cortical regions receiving input from nociceptors, and the alpha ERD is more widespread and more intense during attention to the laser than distraction from the stimuli. This was also consistent with the observations from other studies using EEG or MEG recordings (Mouraux et al. [12], Ploner et al. [13]).

In recent years, many statistical tools, such as principal component analysis (PCA), independence component analysis (ICA), parallel factor analysis (PARAFAC), common spatial subspace decomposition (CSSD), statistical wavelet thresholding (SWT), and Kalman filtering, have been used for analyzing biological or biomedical data, including EEG, MEG, and fMRI (e.g., Lee et al. [14], Cao et al. [15, 16], Makeig et al. [17], Anemüller et al. [18], Miwakeichi et al. [19], Browne and Cutmore [20], Wang et al. [21], Galka et al. [22], Cichocki [23, 24]). The common goal of these mathematical tools is to discover the hidden components underlying the data and extract the markers for characterizing specific events (e.g., event-related potentials). In addition, combing ICA or other statistical tools with advanced time-frequency analysis methods has also been advocated in cognitive neuroscience and neuroimaging (e.g., Makeig et al. [25], Mørup et al. [26]).

In this paper, we conduct both quantitative and qualitative analyses of ECoG data induced by pain stimuli controlled by a laser pulse. The investigation is focused on two selected human subjects under several different controlled stimuli conditions: attention, distraction, as well as under different laser intensity levels. Statistical analysis was conducted for both averaging trials and single trials. The averaging-trial study attempts to find out the dominant and common components (especially LEPs) by averaging all trials (of one subject) under the same conditions. In contrast, the single-trial study aims to search for instantaneous brain waves and to analyze the corresponding LEP properties (such as the amplitude and latency). The signal-trial analysis is important because the spontaneous brain activities that are regarded as "noise" are often diminished by averaging. We believe that the results obtained from the single trials, if analyzed appropriately, often offer extra information that is unavailable in the averaging-trial study (e.g., Makeig et al. [25]).

To achieve our goal, we select proper processing procedures and mathematical tools, including factor analysis (FA) and ICA, to the experimental recordings. This builds on the assumption that within a short timescale the ECoG recordings are approximated by an instantaneous linear generative model that is corrupted by additive noise. The LEPs of interest and other instantaneous brain activities are assumed to be mutually independent. To blindly separate the sources of interest (i.e., LEPs), we first resort on a dimensionality reduction procedure followed by an efficient and robust ICA estimation method. In addition, with eigenvalue decomposition, an energy ratio threshold is defined to reject nonsignificant components, which are regarded as the interfering noise from the raw ECoG recordings. The values of these two

[1] This is in contrast to the earliest component N1, which is a lateralized, relatively small negative peak.

[2] In the previous studies, the authors were not sure if the ECoG-LEPs correspond to the scalp-recorded LEPs (N2 and P2), therefore they used the nomenclatures $N2^*$ and $P2^{**}$. Here, for simplicity, we use N2 and P2 for referring to these two LEPs. The LEPs can also be labeled by polarity and latency; when the latency is known, we also use N150 or P200 for the same reference purpose.

statistical analysis methods have been demonstrated in both averaging and single trials (e.g., Cao et al. [15, 16]). In terms of single-trial analysis, wavelet-based time-frequency analysis is also used to assist the quantitative analysis of Z-score transformed power across different frequency bands. Whilst these statistical methods are not new the contribution of this paper is to integrate these methods with careful computational procedures and present a systematic study of the ECoG recordings for their LEP characterizations, which might offer some insights for the neurophysiological or clinical practice. To our best knowledge, we are in the first position or for the first time, to employ the statistical ICA tools to pain-related ECoG recordings. We describe the computational modeling and analysis in detail and present some interpretations and discussions from our experimental results. On the one hand, we strive to relate the results to the reported neurophysiological observations in the literature; on the other hand, we also pinpoint several interesting findings and observations in our single-trial data analysis.

2. DATA

2.1. Recordings

To obtain the ECoG recordings, special grid electrodes were implanted on the cortical surface of the subjects (i.e., patients for surgical treatment of epilepsy). The grid consisted of platinum-iridium circular electrodes (2.3 mm diameter) with a center-to-center distance between electrodes of 1 cm (Ad-Tech, Racine, Wis, USA). The LEPs were recorded with the implanted grid electrodes over the SI, PS, and MF regions; see Figure 1 for an illustration. During recordings, the subjects wore goggles and reclined on a bed, quietly wakeful with eyes open. Painful heat stimulation was delivered to the contralateral hand dorsum (contralateral to the grid) by a Thulium YAG laser (Neurotest, Wavelight Inc., Starnberg, Germany). The duration of each pulse was 1 millisecond and the beam diameter was 6 mm. Laser energy level was determined to produce a painful sensation of 3-4/10 on a decimal scale (with 0 denoting no pain, and 10 denoting the most intense pain). The ECoG signals were recorded with sampling frequency 1000 Hz. The recordings were carried out at the Johns Hopkins Hospital between 1999 and 2003 (Ohara et al. [7–9]). The protocol was reviewed and approved annually by the Institutional Review Board of the Johns Hopkins Hospital and all subjects signed an informed consent for the studies.

2.2. Subjects

For the purpose of presentation clarification and due to space limit, we have chosen two human subjects in the current study. The statistics of the recording setup regarding the selected two subjects are listed in Table 1. Specifically, the first subject was a 21-year old woman with medically intractable seizures since age 10; her neurological examinations and brain magnetic resonance images (MRIs) were normal. Subdural electrode grids were planted over the frontal-central-parasylvian cortex (no. 1–64 channels) and the medial wall of the left hemisphere (no. 65–80 channels). The second subject was a 21-year old man with complex partial seizures since age 4, whose MRI showed a small cavernoma in the right parietal lobe (contralateral to the side of the implantation). The ECoG signals were recorded from the left fronto-parietal lobe (64 channels) and medial frontal lobe (16 channels). All the signals were recorded with reference to one intracranial electrode.

2.3. Experimental paradigm

There are two types of experimental protocols designed for subjects: *attention/distraction*, and *intensity*. In the attention condition, the subject was asked to count the number of painful stimuli and to report both that number and the average pain intensity after each run of laser pulses; in the distraction condition, the subject read a magazine article and answered questions about it after the run. In these two conditions, constant level of laser intensity was used for the subject, and 38 laser pulses were delivered with an interstimulus interval that was randomly varied between 50 and 10 seconds within each run. Additionally, in the intensity experiment, varying levels of laser stimuli were delivered to the subject, and the subject was asked to rate the subjective pain sensation according to the decimal scale.

2.4. Filtering

Upon loading the raw ECoG recordings to the computer, the data were amplified and band-pass filtered at 0.1–300 Hz (Astro-Med, Inc., West Warwick, RI, USA). Subsequently, we conducted a simple notch filtering procedure to filter out the AC components of power supply (60 Hz).

3. MATHEMATICAL MODELING AND ANALYSIS

3.1. Generative model

The experimental data are assumed to be generated by a probabilistic generative model that is described by two equations as follows:

$$\mathbf{x}_t = \boldsymbol{\mu} + \mathbf{B}\mathbf{z}_t + \boldsymbol{\epsilon}_t, \tag{1}$$

$$\mathbf{z}_t = \mathbf{A}\mathbf{s}_t, \tag{2}$$

where t denotes the time index. Equation (1) is essentially a factor analysis (FA) model, where $\mathbf{z}_t \in \mathbb{R}^n$ is the hidden variable called "factor," the $m \times n$ matrix $\mathbf{B}$ is called the "loading matrix," $\mathbf{x}_t \in \mathbb{R}^m$ denotes the observed multichannel signals measured in the electrodes, $\boldsymbol{\mu} \in \mathbb{R}^m$ denotes the constant mean vector that is often assumed to be zero, and $\boldsymbol{\epsilon}_t \in \mathbb{R}^m$ denotes the additive uncorrelated noise that corrupts the measurements. Equation (2) describes a linear mixture model that is related to the blind source separation (BSS) problem of our interest, where $\mathbf{s}_t \in \mathbb{R}^N$ denotes the independent source signals originated from the brain, $\mathbf{A}$ denotes a linear mixing matrix that roughly models the mixing process and the stationary propagation or scattering effect within a short timescale (say, 200 to 600 mesc); and the

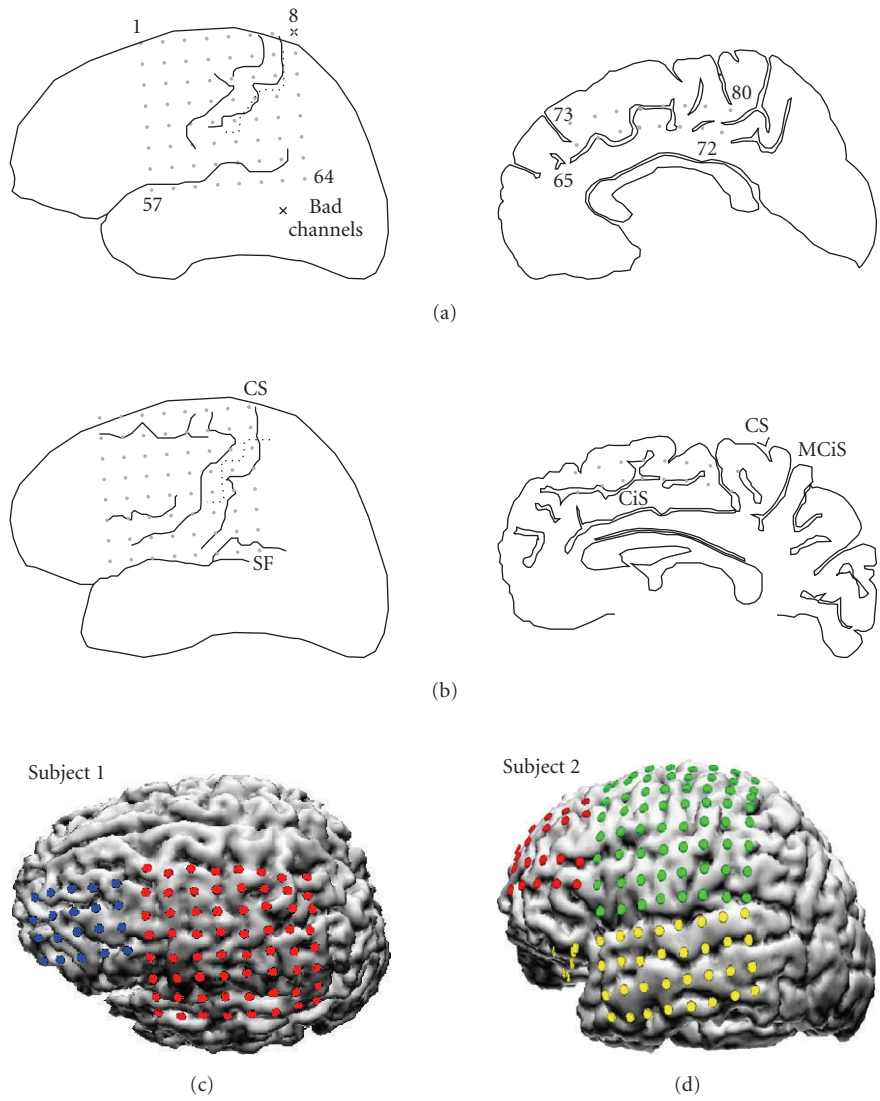


Figure 1: The implanted electrodes' layout; the somatosensory cortex that is associated with the sensation of the pain is located in the parietal lobe of the brain. (a) subject 1; (b) subject 2 (where CS and SF correspond to no. 8 and no. 64 channels, resp.); (c), (d) implanted grids imposed on the reconstructed 3D magnetic resonance images of two subjects. Note that the number of implanted grids shown on the 3D images is more than the number of the available channels shown in Table 1; because of the limitation in data acquisition, only a subset of the grids were selected (CS: central sulcus; SF: sylvian fissure; CiS: cingulate sulcus; MCiS: marginal CiS).

mixed signals consist of the hidden factor $\mathbf{z}_t$ obtained in (1). In the current setting of this paper, we assume $m > n = N$.

No doubt that the generative model described by (1) and (2) is somewhat oversimplified for the ECoG data. However, we believe that the instantaneous linear mixing model is rather reasonable at a short timescale and therefore can be used in the first step. In addition, we assume that matrices $\mathbf{A}$ and $\mathbf{B}$ are constant within the a short duration of measurements. Now, the statistical estimation problem is to infer the independent sources $\mathbf{s}_t$ given the observed $\mathbf{x}_t$. We will tackle this problem via these two statistical tools as described below. Notably, similar methodology has been applied to MEG or EEG recordings with successes in some other real-life recordings (e.g., Cao et al. [15, 16]).

3.2. Factor analysis

Without loss of generality, we assume that $\boldsymbol{\mu} = \mathbf{0}$, and the factor variables satisfy $\mathbb{E}[\mathbf{z}_t] = \mathbf{0}$ and $\mathbb{E}[\mathbf{z}_t\mathbf{z}_t^T] = \mathbf{C}_\mathbf{z}$, where $\mathbf{C}_\mathbf{z}$ is the covariance matrix; and the noise is Gaussian distributed with zero mean and covariance matrix $\boldsymbol{\Sigma}$, which we denote by $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$. In light of (1), we have

$$\begin{aligned} \mathbb{E}[\mathbf{x}_t] &= \mathbf{0}, \\ \mathbb{E}[\mathbf{x}_t\mathbf{x}_t^T] \equiv \mathbf{C}_x &= \mathbf{B}\mathbf{C}_z\mathbf{B}^T + \boldsymbol{\Sigma}. \end{aligned} \tag{3}$$

If $\mathbf{z}_t$ is Gaussian distributed, then $\mathbf{x}_t$ is also Gaussian distributed. If we further restrict that the factor $\mathbf{z}_t$ is whitened, then $\mathbf{C}_\mathbf{z} = \mathbf{I}$ (where $\mathbf{I}$ denotes the identity matrix); this

assumption is reasonable since we can always scale the loading factors $\mathbf{B}$ to satisfy the original model equation. Typically, $\dim(\mathbf{x}) > \dim(\mathbf{z})$, therefore FA is also a dimensionality-reduction method. A close examination of our experimental multielectrode recordings indicates that there are strong correlations between adjacent electrodes, which therefore justifies the necessity of dimensionality reduction.

From a probabilistic point of view, we can write $p(\mathbf{z}_t) = \mathcal{N}(\mathbf{0}, \mathbf{I})$, then $p(\mathbf{x}_t) = \mathcal{N}(\mathbf{0}, \mathbf{B}\mathbf{B}^T + \mathbf{\Sigma})$. Under the Gaussian assumption of the factor analyzer, the posterior probability $p(\mathbf{z}_t \mid \mathbf{x}_t)$ is also Gaussian, with mean and covariance, respectively, defined by

$$\mathbb{E}[\mathbf{z}_t \mid \mathbf{x}_t] = \left(\mathbf{B}^T\mathbf{\Sigma}^{-1}\mathbf{B} + \mathbf{I}\right)^{-1}\mathbf{B}^T\mathbf{\Sigma}^{-1}\mathbf{x}_t, \tag{4}$$

$$\mathrm{Cov}[\mathbf{z}_t \mid \mathbf{x}_t] = \left(\mathbf{B}^T\mathbf{\Sigma}^{-1}\mathbf{B} + \mathbf{I}\right)^{-1}. \tag{5}$$

Now the goal of FA is to estimate the unknown matrices $\mathbf{B}$ and $\mathbf{\Sigma}$, given the observed data $\{\mathbf{x}_t\}$. In the literature, two types of estimation procedures can be employed.

• *Maximum likelihood estimation*

By deriving the log likelihood function (see the appendix) with respect to the unknown variables, we can use iterative optimization procedures, such as the gradient ascent or expectation-maximization (EM) algorithm, to obtain the optimal solution. Upon obtaining the maximum likelihood estimates of $\mathbf{B}$ and $\mathbf{\Sigma}$, we can further calculate the hidden factor $\mathbf{z}_t$ by (4).

• *Least-squared estimation*

Given observed samples $\{\mathbf{x}_t\}_{t=1}^T$, we can calculate the sample covariance matrix (assuming zero mean) and conduct its eigenvalue decomposition (EVD) as follows:

$$\hat{\mathbf{C}}_x = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t^T = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T, \tag{6}$$

where $\mathbf{U}$ is the $m \times m$ orthogonal matrix that consists of eigenvectors as its column vectors, $\mathbf{\Lambda}$ is a diagonal matrix that consists of the diagonal entries as eigenvalues. Note that when the noise is zero or the noise is negligible and has a diagonal covariance matrix, then FA reduces to PCA as a special case. Upon PCA, we can empirically estimate the noise covariance. Let $\mathbf{U}_n$ denote an $m \times n$ matrix that consists of the first n dominant eigenvectors, then we can estimate the noise covariance by

$$\hat{\mathbf{\Sigma}} = \hat{\mathbf{C}}_x - \mathbf{U}_n\mathbf{\Lambda}_n\mathbf{U}_n^T, \tag{7}$$

and estimate the loading matrix by

$$\hat{\mathbf{B}} = \mathbf{U}_n\mathbf{\Lambda}_n^{1/2}. \tag{8}$$

Finally, the factor variable $\mathbf{z}_t$ is produced by a linear transformation:

$$\mathbf{z}_t = \mathbf{Q}\mathbf{x}_t, \tag{9}$$

where $\mathbf{Q} = (\hat{\mathbf{B}}^T\hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{B}})^{-1}\hat{\mathbf{B}}^T\hat{\mathbf{\Sigma}}^{-1}$. Note that in this case, the dimensionality of $\mathbf{z}_t$ can be determined by PCA with dimensionality reduction, whereas the remaining components are considered to be "significant" in terms of variance or energy contribution.

3.3. Independent component analysis

Upon performing the model reduction using FA, we further aim to apply the blind source separation (BSS) approach, using the tool of ICA (e.g., Cichocki and Amari [27]), to recover the hidden sources in (2). Roughly speaking, ICA is built upon the assumption that the hidden sources in $\mathbf{s}_t$ are mutually independent and subject to an instantaneous linear mixing.

There are many ICA/BSS algorithms available in the literature. To our interest, two kinds of batch (i.e., noniterative) ICA/BSS algorithms are considered.

Time-domain method

Specifically, we focus on the BSS algorithms based on generalized EVD of the time-delayed cross-correlation matrices or cumulant statistics, such as the SOBI (second-order blind identification) and JADE (joint approximate diagonalization of eigen-matrices) algorithms. These methods are fast and noniterative (thereby independent of the initial conditions). In our experiments, we have tried and compared the SOBI and JADE algorithms, and found that their results were qualitatively similar. However, JADE is more desirable and preferred since it incorporates higher-order statistics.

Time-frequency method

Specifically, the source separation criterion of this method is conducted in time-frequency domain based on joint diagonalization of the spatial time-frequency distribution (TFD). A representative example is the algorithm described by Févotte and Doncarli [28]. This method is more intuitively appealing (by taking into account of the information in both time and frequency) and has been demonstrated to be robust to noise.[3]

Notably, although the hidden factor $\mathbf{z}$ is whitened (with zero mean and unit variance), it is still likely that the mixing matrix is ill-conditioned, which thereby makes the estimation of its inverse (or Moore-Penrose pseudoinverse), the demixing matrix $\mathbf{W} = \mathbf{A}^{-1}$ (or $\mathbf{W} = \mathbf{A}^{\dagger} \equiv (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$), rather difficult, especially in single-trial experiments. One way to overcome this problem is to conduct a two-stage ICA procedure. The essence of the two-stage ICA is as follows: the role of the first-stage ICA is "rough tuning," which produces a guess (or poor estimate) of the ill-conditioned mixing matrix; and the final "fine tuning" job is accomplished by the second-stage ICA routine. The trick of such a two-stage

[3] The Matlab code is available at http://www-sigproc.eng.cam.ac.uk/~cf269.

Table 1: Summary of the experimental recordings of two human subjects.

Subject	Condition	Laser intensity	No. of electrodes	No. of runs	No. of trials at each run
1	Attention	720 mJ	89	3	38, 38, 38
1	Distraction	720 mJ	89	3	38, 38, 38
2	Attention	720 mJ	80	2	38, 38
2	Distraction	720 mJ	80	2	38, 38
2	Intensity	480 mJ	80	4	8, 12, 10, 10
2	Intensity	640 mJ	80	4	11, 11, 12, 10
2	Intensity	800 mJ	80	4	10, 10, 10, 10

ICA often helps to recover the hidden components in many ill-conditioned scenarios—if it is not the case, the second-stage ICA simply produce improved or similar results as in the first-stage ICA.

The significance of the (uncorrelated or independent) components is determined by their relative energy (or variance). Physiologically, we believe those sources that have relative great energy are more meaningful in terms of repeatability. In practice, selecting the number of principal components is done by EVD followed by a threshold selection. In our experiments, five to eight principal components were typically selected, which account for about 97–99% of the total energy. Specifically, let $\mathbf{\Lambda} = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ denote the diagonal matrix that contains the nondecreasing eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$, the number of significant components, k, is chosen according to the following criterion:

$$k = \arg\min_i L_i \quad \text{s.t. } L_i = \frac{\sum_{j=1}^{i} \lambda_j}{\sum_{j=1}^{n} \lambda_j} > \mathrm{Th}, \tag{10}$$

in which the threshold Th was empirically set as 0.97; the nonnegative eigenvalue indicates the relative significance of specific component in terms of its energy contribution.

3.4. Identification of interested source by deflation

Let $\mathbf{y}^{(1)} = \mathbf{W}^{(1)}\mathbf{z}$ and $\mathbf{y} = \mathbf{W}^{(2)}\mathbf{y}^{(1)}$ denote, respectively, the first- and second-stage ICA unmixing equations, where $\mathbf{W}^{(1)}$ and $\mathbf{W}^{(2)}$ denote the associated unmixing matrices; then the final unmixed signals, $\mathbf{y}_t$, can be estimated as

$$\mathbf{y}_t = \mathbf{W}\mathbf{z}_t = \mathbf{W}^{(2)}\mathbf{W}^{(1)}\mathbf{z}_t, \tag{11}$$

where $\mathbf{W} = \mathbf{W}^{(2)}\mathbf{W}^{(1)}$ denote the global (combined) unmixing matrix.[4] Notably, each column of $\mathbf{W}^{-1}$ contains the relative strengths of a source component at the individual scalp electrodes, which can be used to identify the interested source component.

Given the estimated $\mathbf{y}_t = [y_1(t), y_2(t), \ldots, y_n(t)]^T$, we can also reconstruct the *partial* hidden factor by projecting the ith component of $\mathbf{y}_t$, denoted by $y_i(t)$, backward onto the subspace[5]

$$\hat{\mathbf{z}}_t = \mathbf{W}^{\dagger}[0, \ldots, 0, y_i(t), 0, \ldots, 0]^T \equiv [\mathbf{W}^{\dagger}]_i y_i(t), \tag{12}$$

where $[\mathbf{W}^{\dagger}]_i$ denotes the ith column vector of the matrix $\mathbf{W}^{\dagger}$. Furthermore, we can reconstruct the specific source of interest in the observed data space (i.e., the scalp signals contributed merely to the ith source)

$$\hat{\mathbf{x}}_t = \mathbf{Q}^{\dagger}\hat{\mathbf{z}}_t = \mathbf{Q}^{\dagger}\mathbf{W}^{\dagger}[0, \ldots, 0, y_i(t), 0, \ldots, 0]^T. \tag{13}$$

By projecting $\hat{\mathbf{x}}_t$ to the original channels' positions (i.e., the 8×8 electrode layout), we essentially *identify* the source(s) of interest. It should be noted that the "source identification" here is only limited to the two-dimensional scalp surface, and does not refer to localization of the three-dimensional spatial position of the "voxel."

In addition, in order to evaluate the *relative* contribution of every electrode to the extracted independent component (especially for the LEP), we need to consider the *joint* effect of $\hat{\mathbf{x}}_t$ and $\mathbf{W}$. For this purpose, we may also calculate the *weighted* estimate of the sensor space $\hat{\mathbf{x}}_t$ as follows:

$$\tilde{\mathbf{x}}_t = \mathbf{w}_i^T \odot \hat{\mathbf{x}}_t = [w_{i1}\hat{x}_1(t), w_{i2}\hat{x}_2(t), \ldots, w_{in}\hat{x}_n(t)]^T, \tag{14}$$

where $\odot$ denotes the Hadamard (elementwise) product, $\mathbf{w}_i = [w_{i1}, w_{i2}, \ldots, w_{in}]$ denotes the ith row vector of the matrix $\mathbf{W}$, and $\hat{\mathbf{x}}_t = [\hat{x}_1(t), \hat{x}_2(t), \ldots, \hat{x}_n(t)]^T$ is the back-projected sensor space from the ith independent source via (13). As a distinction, we call the reconstructed $\hat{\mathbf{x}}_t$ in the sensory space as "unweighted map" and the reconstructed $\tilde{\mathbf{x}}_t$ in the sensory space as "weighted map." Notably, because of the degeneracy of $\mathbf{W}$, the "weighted map" is subject to the scaling and algebraic sign uncertainties.

3.5. Time-frequency analysis

In addition to analyzing temporal signals, we also resort on time-frequency analysis tools (such as the short-time Fourier transform, or Wigner-Ville distribution, and wavelet transform) to extract more information for quantitative compar-

[4] If only one-stage ICA is used, then $\mathbf{W}^{(1)} = \mathbf{I}$, $\mathbf{y}^{(1)} = \mathbf{z}$, and $\mathbf{W} = \mathbf{W}^{(2)}$.

[5] If the demixing matrix is square, then the matrix pseudoinverse $\mathbf{W}^{\dagger}$ will reduce to the matrix inverse $\mathbf{W}^{-1}$.

isons. Specifically, wavelet analysis is appealing and considered superior to the short-time Fourier transform for nonstationary signals, including EEG (e.g., Mallat [29]; Tallon-Baudry et al. [30], Düzel et al. [31], Mouraux et al. [12], Ohara et al. [7]). Here, we choose the continuous wavelet transform for our purpose because of its adaptive time-frequency analysis via multiscale decomposition. However, because of the *uncertainty principle*, in order to obtain a good frequency resolution, sufficient time samples are required. In the experiments, we will use the Wigner-Ville distribution for an illustration purpose, while in the quantitative analysis we will use the continuous wavelet transform.

For a temporal signal $x(t)$ (i.e., the raw recordings from one electrode channel), the power of its continuous-time wavelet transform is described by

$$X(t,\omega_0) = |x(t) * \psi(t,\omega_0)|^2, \tag{15}$$

where $*$ denotes convolution product between the signal and the mother wavelet function, and $\psi(t,\omega_0)$ is a complex-valued Morlet mother function:

$$\psi(t,\omega_0) = (\sigma^2\pi)^{-1/4} \exp\left(\frac{-t^2}{2\sigma^2}\right) \exp(j2\pi\omega_0 t), \tag{16}$$

where $j = \sqrt{-1}$, and σ is the bandwidth parameter. The width of the Morlet wavelet, defined by $2\pi\sigma\omega_0$, is set to 7 in our study.[6] The central frequency ω_0 ranges from 1 to 60 Hz in steps of 1 Hz. To analyze the specific temporal window of interest, we select a 100-millisecond prestimulus period and a 500-millisecond poststimulus period, with a total window length 600 milliseconds.

To compare the power change between the prestimulus and poststimulus periods, we need to introduce some "relative" measures to obtain a baseline for the poststimulus power. This is important because we are not interested in the "absolute" power statistic *per se*, but interested in the stimulus-induced relative power change. In the literature, the measure of event-related band power change (ERBP) was defined as (e.g., Ohara et al. [7])

$$\mathrm{ERBP}(t,\omega_0) = 10\log\left(\frac{X(t,\omega_0)}{m(\omega_0)}\right)(\mathrm{dB}), \tag{17}$$

where $m(\omega_0)$ denotes the *median* power envelope during the prestimulus period. Alternatively, we can use another measure, which we refer to as "Z-score transformed poststimulus power," by using the Z-score transformation (e.g., Browne and Cutmore [20]):

$$\widetilde{X}(t,\omega_0) = \frac{X(t,\omega_0) - \mu(\omega_0)}{\sigma(\omega_0)}, \tag{18}$$

[6] Generally, the greater the width parameter is, the better frequency resolution we can obtain; nevertheless, this is at the cost of sacrificing temporal resolution. The temporal resolution has a reciprocal relationship with respect to the frequency resolution.

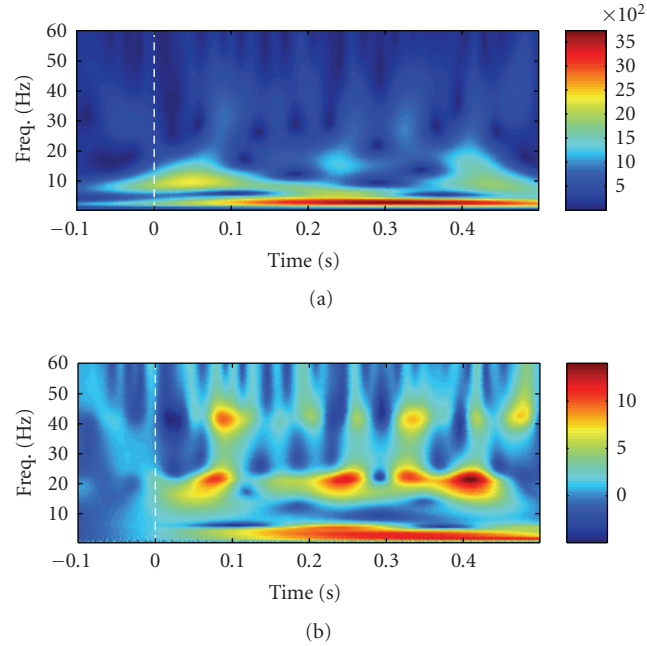


Figure 2: The original (upper panel, unit μV^2) versus Z-score transformed (bottom panel, unitless) wavelet scalogram of one selected channel in a single trial (subject 2, attention task, laser intensity 720 mJ). The white dash lines indicate the laser stimulus onset. As seen, the ERS and ERD are highlighted more clearly by the Z-score transformation given by (18).

where $\mu(\omega_0)$ and $\sigma(\omega_0)$ are, respectively, the *mean* and *standard deviation* of the power in a specific channel band (with center frequency ω_0) during the prestimulus period. The motivation of (18) is to introduce baseline power values across different frequency bands for the poststimulus power statistics, which are used for standardized comparisons. In doing so, the low-amplitude component at high frequency will be highlighted, which also makes the time-frequency atom in the *gamma* ($>$ 32 Hz) band more visible. See Figure 2 for an illustrative example. Note that the Z-score power value can be negative; the positive values indicate the *event-related synchronization* (ERS), and the negative values indicate the *event-related desynchronization* (ERD), both between the prestimulus and poststimulus periods. Hence, the Z-score transformation provides a clearer understanding of the time-frequency map (in terms of relative power change).

In some cases, the resulted Z-score transformed poststimulus power will be converted to a two-dimensional time-frequency distribution map, denoted by $E(t,\omega)$, and further normalized to unity such that $\iint E(t,\omega)d\omega\,dt = 1$, which we refer to as the normalized power. In doing so, each time-frequency atom can be interpreted by a nonnegative probability in the time-frequency plane.

4. COMPARATIVE EXPERIMENTS FOR AVERAGING TRIALS

We first apply the above described computational procedure and statistical tools for averaging trials, the signal-trial

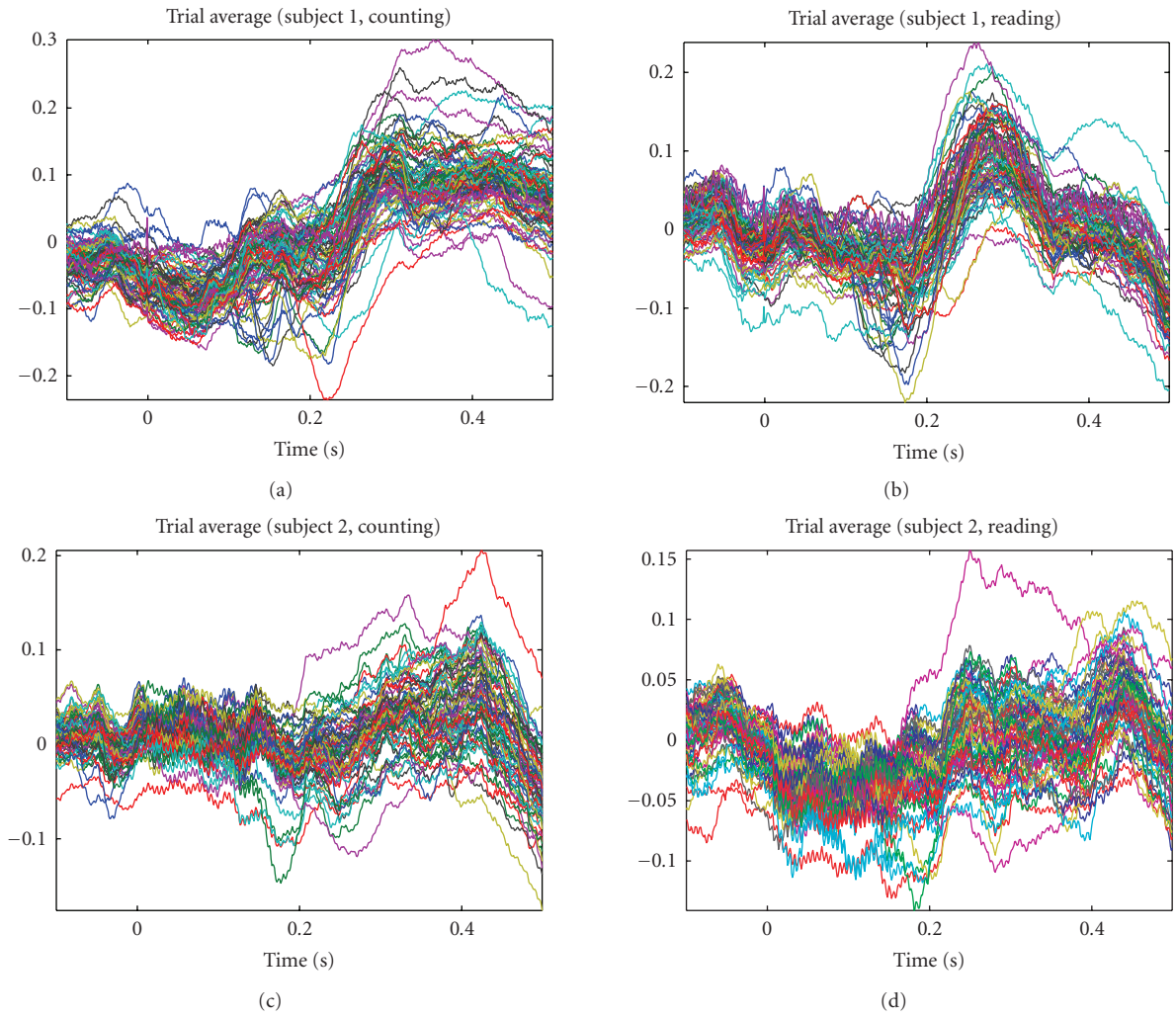


Figure 3: The averaging waveforms (arbitrary scaling) from averaging trials for both subjects in two tasks. As seen, the averaging evoked potentials are not clearly evident in these plots.

experiments will be treated later in more detail. The experimental results reported in this section will be illustrated for subject 1; two kinds of conditions, counting and reading, are considered.

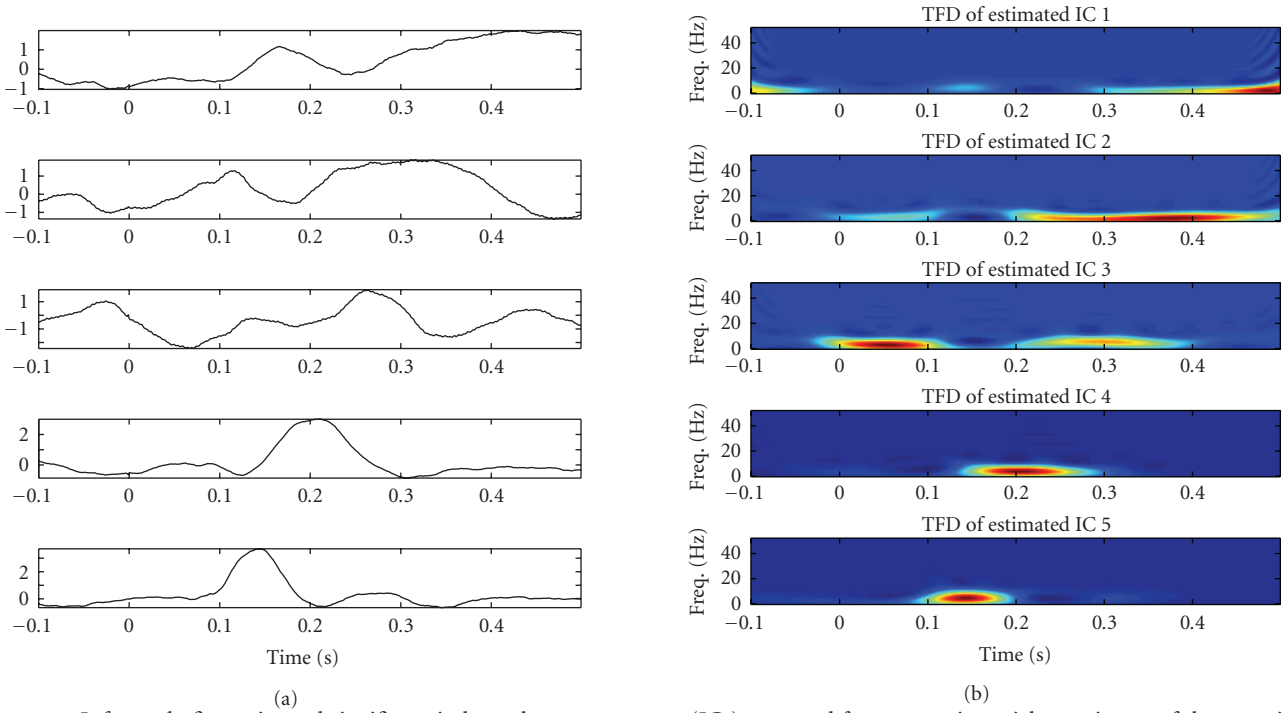
4.1. Extraction of laser-evoked potentials

First, we aim at extracting LEPs for the averaging-trial experiments. Specifically, according to the laser onset tag, the ECoG recordings (of all channels) were averaged upon the total number of trials at each run. By doing so, the effect of the visual or muscle artifacts may be greatly reduced. However, it is difficult to identify the LEPs from the averaging ECoG waveforms of all channels (see Figure 3). Not only the peaks of the LEPs are less evident, the averaging waveforms still suffer from noise and artifact corruption.

To overcome these issues, we then apply the statistical tools (FA and ICA) to further process these trial-averaging signals. In the averaging-trial experiments for subject 1, we selected five independent components for the purpose of extracting LEPs. These five independent components are considered to be "significant" because they contribute mostly to the averaged ECoG data in terms of variance or energy.[7] Due to the averaging/smoothing effect, one-stage ICA procedure (with the JADE algorithm) was found typically sufficient in the experiments.[8] The experimental results for the subject 1, in the time domain as well as in the time-frequency domain, are illustrated in Figures 4 and 5. As observed in the figures, we can extract typical peaks around 150 milliseconds and 200 milliseconds, which might correspond to the hypothetic N2 and P2 peaks (or N2* and P2**) of LEPs, which we also refer to as N150 and P200, respectively; the other components can be viewed as other significant independent spontaneous brain activities. These findings were confirmed in both attention (counting) and distraction (reading) conditions.

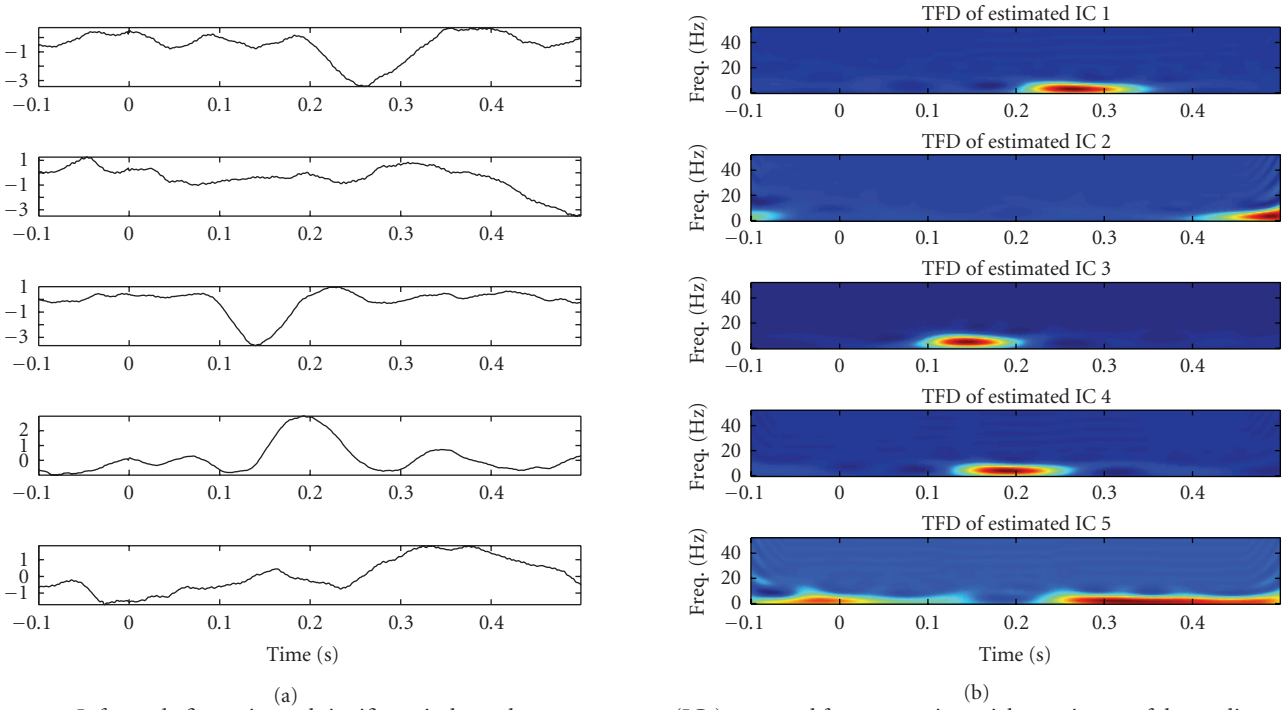
[7] Specifically, the two LEP components are more significant and relatively robust in that they remain approximately unchanged when we vary (increase or decrease) the selected number of components by 1 or 2.

[8] This is in contrast to the two-stage ICA procedure in single-trials; this is often witnessed by the observation that the outcomes of the second-stage ICA are not much different from the results of the first-stage ICA.



(a) (b)

FIGURE 4: *Left panels:* five estimated significant independent components (ICs) extracted from averaging-trial experiment of the counting task (attention situation) for subject 1. *Right panels:* the associated time-frequency distribution (TFD) maps.



(a) (b)

FIGURE 5: *Left panels:* five estimated significant independent components (ICs) extracted from averaging-trial experiment of the reading task (distraction situation) for subject 1. *Right panels:* the associated time-frequency distribution (TFD) maps.

Next, we conduct the task of LEP source identification. This is done by back-projecting the ith independent component (i.e., the estimated LEP) back to the observed sensor space. Specifically, the power contour maps of N2 (N150) and P2 (P200) under the attention and distraction conditions are illustrated in Figures 6 and 8, respectively. The results are qualitatively close (but not identical) to the previous study (Ohara et al. [7, 8]), in which the LEP peak was found over the interhemispheric (medial) surface. We also plot the *combined* contributions of the power contour map for N150 and

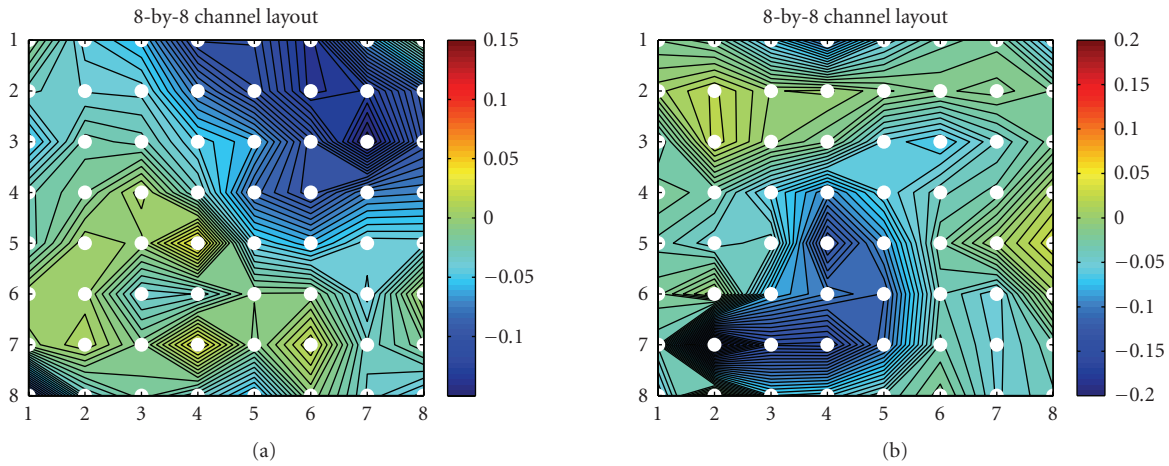


Figure 6: Source identification in the averaging trial of subject 1: the back-projected 8×8 (first 64 channels) scaled amplitude contour map of the LEP peak at N150 (the 5th independent source at 150 milliseconds, left panel) and P200 (the 4th independent source at 200 milliseconds, right panel) in the counting task (attention condition).

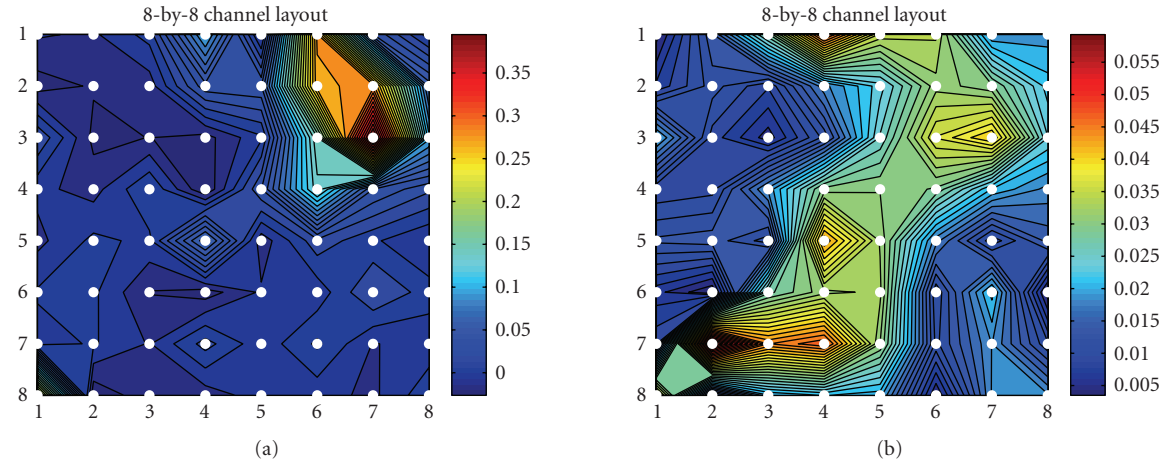


Figure 7: *Left panel:* the "weighted" map of LEP-N150 (compared to the "unweighted" map the left panel of Figure 5) from (14). *Right panel*: the back-projected 8-by-8 (first 64 channels) power (i.e., the absolute value of the amplitude) contour map of the two LEPs, N150 and P200, averaged between 120 milliseconds to 240 milliseconds (subject 1, attention condition).

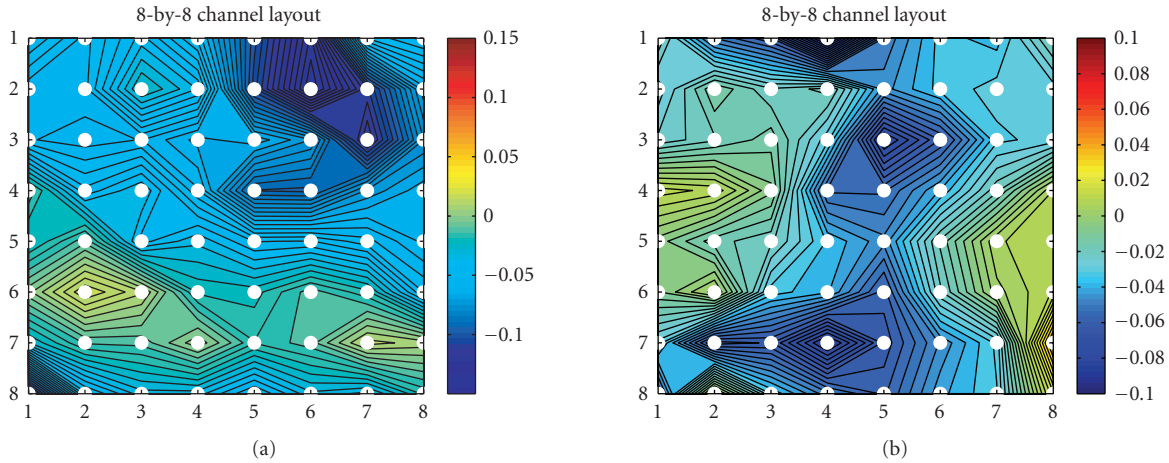


Figure 8: Source identification in the averaging trial of subject 1: the back-projected 8-by-8 (first 64 channels) scaled amplitude contour map of the LEP peak at N150 (the 3rd independent source at 150 milliseconds, left panel) and P200 (the 4th independent source at 200 milliseconds, right panel) in the reading task (distraction condition).

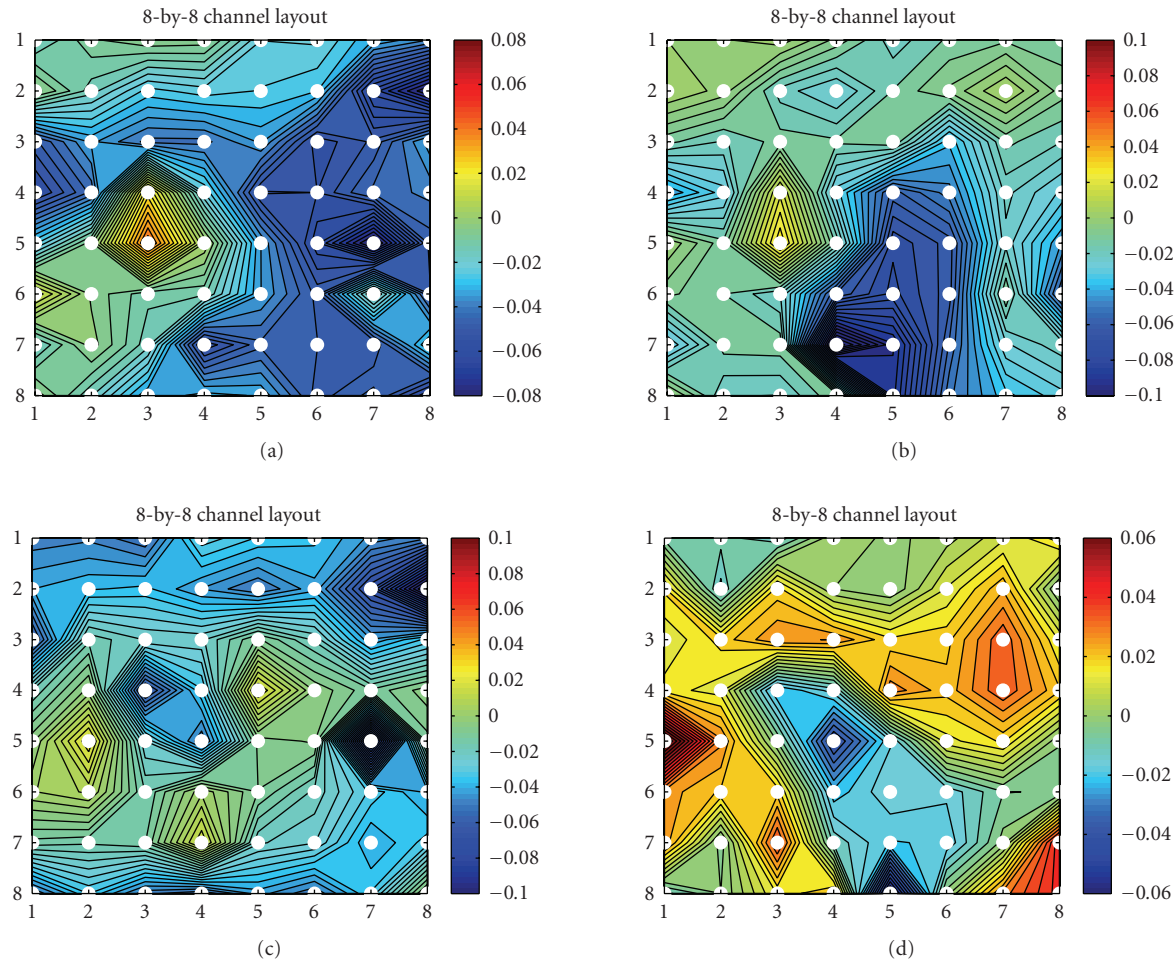


FIGURE 9: Source identification in the averaging trial of subject 2: the back-projected 8-by-8 (first 64 channels) scaled amplitude contour map of the LEP peak at N2 and P2 in the counting (top 2 panels) and reading (bottom 2 panels) tasks.

P200, averaged from 120 milliseconds to 240 milliseconds, as shown in Figure 7. As seen in the figure, in the LEP-N2 (i.e., N150), the greatest brain activities happen around the vertex (Cz)—the upper right corner of the 8×8 electrode layout (see Figures 1, 6, and 7), these observations were consistent with our early result (Ohara et al. [7, 8]), as well as other independent findings using EEG and fMRI with a similar setup (e.g., see Figure 1 of Iannetti et al. [5]). Similarly, we also obtained the LEPs' mappings for subject 2 (see Figure 9).

4.2. *Relative power*

We compare the time-frequency distribution (TFD) power between the prestimulus and poststimulus periods. The averaged total power (per channel) and the averaged power (per channel) of specific frequency bands, including *theta* (4–7.5 Hz), *alpha* (8–12 Hz), *beta* (12.5–32 Hz), and *gamma* (32–60 Hz), are all calculated. In Table 2, we summarize the statistics of two subjects under the attention (counting) and distraction (reading) conditions. The corresponding scatter plots of prestimulus and poststimulus power (of selected frequency bands) of all channels are shown in Figures 10 and 11.

From Table 2, several observations are noteworthy.

(i) The power in the poststimulus period is generally greater than that in the prestimulus period, which is obviously evidenced in terms of total power, θ and α power.

(ii) The θ power increase (or ERS) is relatively more pronounced in the attention condition than in the distraction condition.

(iii) The β power remains roughly the same level after the laser stimulus, regardless of the undertaken tasks.

(iv) The γ power is typically small in all conditions, with slightly greater value in the attention condition than in the distraction condition.

It is noteworthy that the above observations are consistent with the findings reported in neuroscience and neurophysiology (to name a few, Bromm and Lorenz [1], García-Larrea et al. [2], Ohara et al. [7]). Although the statistics summarized in Table 2 are calculated based on the averaging trials, statistical test (see the next subsection) on single trials also reveals statistical significance.

Table 2: The relative power comparisons between prestimulus period (100 milliseconds) and poststimulus period (500 milliseconds) in the averaging-trials. The statistics are averaged over the total number of channels (namely, divided by 89 and 80 for subjects 1 and 2, resp.) and the relative time period. The values are unitless, reflecting the ratio among the normalized energy of the time-frequency map.

	Subject 1 (counting)	Subject 1 (reading)	Subject 2 (counting)	Subject 2 (reading)
Ave. prestimulus total power	0.0667	0.0796	0.0748	0.0969
Ave. prestimulus θ power	0.0160	0.0265	0.0175	0.0333
Ave. prestimulus α power	0.0048	0.0105	0.0112	0.0102
Ave. prestimulus β power	0.0047	0.0078	0.0136	0.0087
Ave. prestimulus γ power	0.0021	0.0019	0.0026	0.0025
Ave. poststimulus total power	0.1867	0.1841	0.1850	0.1806
Ave. poststimulus θ power	0.0486	0.0617	0.0539	0.0502
Ave. poststimulus α power	0.0074	0.0124	0.0154	0.0160
Ave. poststimulus β power	0.0051	0.0073	0.0134	0.0104
Ave. poststimulus γ power	0.0024	0.0024	0.0033	0.0023

4.3. Statistical hypothesis testing

In order to evaluate the results of the averaging trials, we conduct some statistical hypothesis tests in order to confirm the "statistical meaning" of the extracted LEPs. This procedure is necessary because the result of the extracted LEPs in the averaging trials does not tell anything *in statistical sense* about each single trial; namely, we need to be sure if the results we are tempted to interpret are due to random effects from averaging, or due to the consistent causality in all or most of individual single trials.

Two popular hypothesis testing methods we consider here are the ANOVA (analysis of variance, or F-test) and Mann-Whitney test (or U-test). In our experiments, we first use the Mann-Whiteny test to calculate the so-called P-value. Second, we also apply a logarithm transformation of the raw samples in attempt to obtain the Gaussianity (i.e., the raw samples are lognormal distributed, as confirmed by the Shapiro-Wilk test), and then apply the ANOVA to calculate the P-values.[9]

To conduct the statistical tests, we apply the estimated unmixing matrix $\mathbf{W}$ from the averaged trial to each single trial; then we obtain the *surrogate* "single-trial LEP"[10] for individual single trials, for either LEP-N150 or LEP-P200. For a specific LEP component, we expect that there is a consistent and significant difference between the prestimulus and poststimulus periods in terms of their absolute values. In our case, the statistical test was conducted in the time-frequency domain. For instance, for the LEP-N150 (or LEP-P200), according to its time-frequency map, we empirically choose a window around the maximum power value (i.e., the magnitude) in the time-frequency map,[11] and further conducted the Mann-Whitney test for each extracted LEP component in all single trials, in which the comparison was done in the time-frequency domain. Specifically, we compared the *average mean* of the power value inside the time-frequency window centered around the maximum point (which in the time domain corresponds to the extracted LEP peak) with that of the prestimulus period (with the same time-frequency window size), both averaged across all frequency bins. Consequently, we may expect that the signal amplitude in the region of interest is significantly greater than that of the baseline. The statistical hypothesis testing results are summarized in Table 3, and the corresponding boxplots are shown in Figure 12. As seen in the table, the P-values of U-test are all smaller than .05, and three of them are much smaller than .01, consequentially, they are statistically significant. For the sake of completeness and sanity check, we also calculated the P-values that are not associated with the LEPs, (i.e., the other independent components extracted from the averaged-trials), we have consistently observed that their P-values are greater than .2 (around $.2 \sim .6$); hence, we can conclude that these non-LEP components obtained in the averaged trials are ascribed by the random effect that is not consistent in each single trial.

5. QUALITATIVE AND QUANTITATIVE ANALYSES OF SINGLE-TRIAL RECORDINGS

The averaging-trial experiments and statistical tests described above present an informative baseline and guideline for further single-trial experiments. As we mentioned earlier,

[9] It should be noteworthy that it is also possible to apply the logarithm transformation to the samples before the Mann-Whitney test; in this case, the P-values will remain unchanged, except that the standard deviation will become smaller after the logarithm transformation.

[10] We call it surrogate single-trial LEP because the LEP is not estimated directly from individual single trial alone; instead, its recovery arises from the unmixing matrix that is estimated based on averaging all single trials.

[11] Typically, the window of temporal axis is centered at 150 milliseconds (or 200 milliseconds) with width 30 milliseconds, and the window of frequency axis covers from 4.5 to 6.5 Hz (with the resolution of 0.25 Hz for each frequency bin).

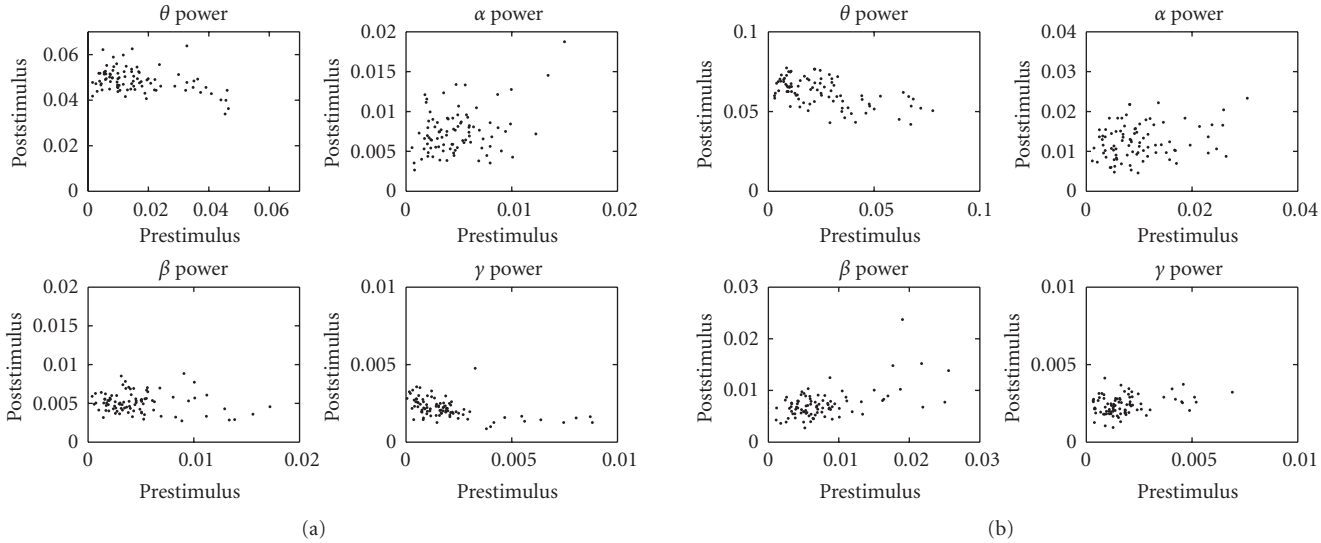


Figure 10: The scatter plots of prestimulus and poststimulus power comparisons in averaging trials for 89 channels (subject 1, left: counting task, right: reading task).

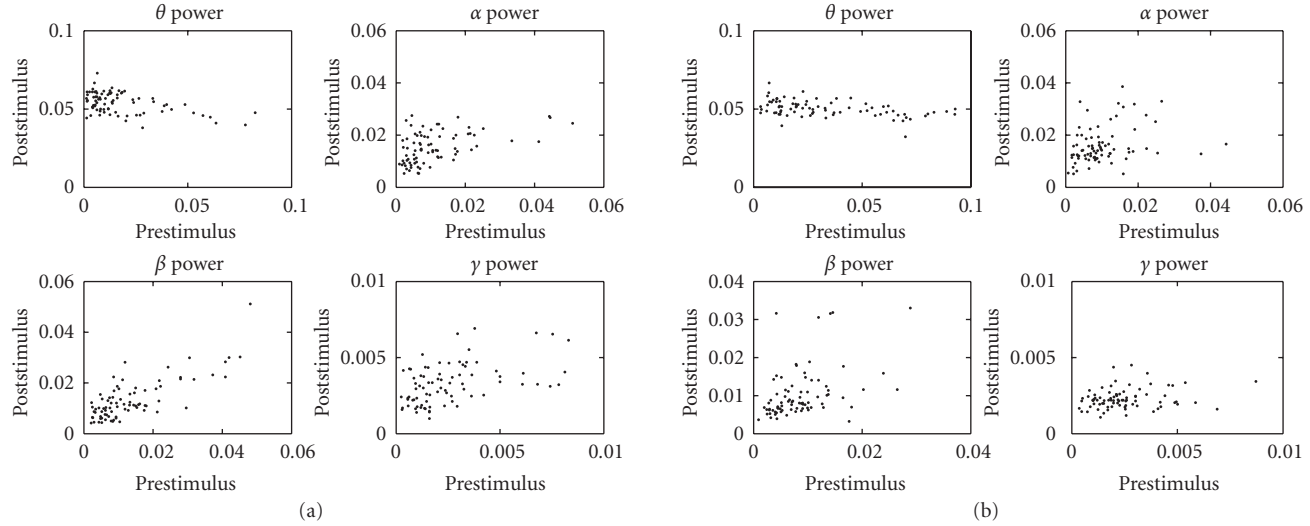


Figure 11: The scatter plots of prestimulus and poststimulus power comparisons in averaging trials for 80 channels (subject 2, left: counting task, right: reading task).

it is well known that by averaging the ECoG recordings, we might lose some valuable information due to cancelation. For this reason, single-trial experimental findings would be also interesting. Nevertheless, single-trial analysis is more challenging because of the random background activities and artifacts; hence, obtaining consistent yet interpretable results is quite difficult. To succeed, we may require additional care or more sophisticated processing. Table 4 lists the operation comparisons between the averaging and single-trial analyses at each stage of procedure.

Notably, in contrast to the averaging-trial experiments in which the artifact effects are greatly reduced, strong artifacts may exist in the single-trial experiments. In practice, artifacts (often with low-frequency components) are sometimes observed by visual inspection. In this case, we will be cautioned about using these "bad" channels. A simple solution is to discard them or average with their neighboring channels. Selection of bad channels is often assisted with the reference of averaging trials. For instance, channels with extremely high amplitude and low frequency are generally regarded as eye movement artifacts. Since the FA/ICA statistical methods

Table 3: Statistical hypothesis testing statistics of various extracted LEPs in averaging trials for subject 1. The Mann-Whitney U-test was applied to the "absolute value" of the raw samples, and the ANOVA F-test was applied to the logarithm transformation of the absolute value of the raw samples. The N/A implies that the samples are neither normally nor log-normally distributed and therefore cannot be used for ANOVA.

P-value	Counting (N150)	Counting (P200)	Reading (N150)	Reading (P200)
U-test	.0029	.0013	7×10^{-5}	.0269
F-test	.0003	N/A	6×10^{-5}	.0183

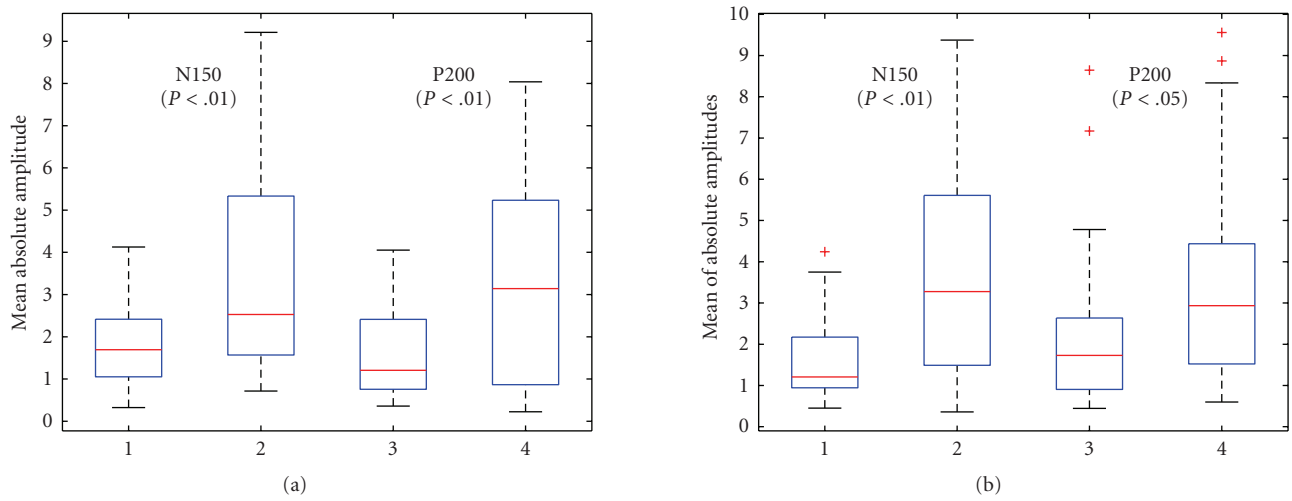


Figure 12: Boxplots of the absolute value of raw samples, together with their Mann-Whitney test P-values on the counting (left panel) and reading (right panel) tasks (subject 1, laser intensity 720 mJ).

Table 4: A comparison of main operations between the averaging and single-trial analyses.

Routine	Averaging trials	Single trials	Purpose
Averaging	Yes	No	Smoothing
FA+PCA	Yes	Yes	Noise and dimensionality reduction
first-stage ICA	Yes	Yes	Extracting independent sources
second-stage ICA	No	Optional	Fine tuning of the sources
Source identification	Yes	Yes	Locating the LEPs of interest
WVD	Optional	Optional	Visualization
Wavelet transform	Yes	Optional	Z-score transform
Statistical test	Optional	Optional	Testing hypothesis

described above can somehow reduce these effects, hence only those channels with obvious artifacts were removed in the experimental procedure.

In the sequel, we will conduct qualitative and quantitative comparisons of single-trial recordings for different measurements listed in Table 1.

5.1. Setup

In single-trial experiments, the number of independent components usually varies from trial to trial (for the purpose of extracting LEPs), and we typically choose the number between 5 and 8. This is because in individual single trials, some small-amplitude but potentially important components at high frequency may play a crucial role, which is also interesting to observe. For the same purpose, we will use the two-stage ICA procedure (JADE algorithm followed by TFD joint diagonalization) described earlier in Section 3.

Upon extracting the LEP of interest, we further identify the LEP localization in the sensor space and focus on the analysis on *one* specific channel (in contrast to the analysis of *all* channels in the averaging trials). Specifically, we will examine the single-trial recordings under attention and distraction conditions, as well as the statistics of the LEP attributes (latency and amplitude) with varying pain levels (i.e., given different laser intensities).

5.2. Single trials versus averaging trials

In the single-trial experiments, we apply the above-described procedure with the goal of extracting the LEPs under different conditions, and the results obtained in the averaging trials are considered to be the baselines for qualitative comparison.

Typically, not all single trials have good quality recordings compared with the averaging trials. Here we show a few successful examples that are capable of identifying the markers of the LEPs. Notably, in our experiments, it was observed that the LEP-N2 can be easily identified, while the LEP-P2 is more difficult to separate. See Figure 13 for two illustrations under different setting conditions.

In order to evaluate the variability between different single trials, we apply the estimated demixing matrix $\mathbf{W}$ obtained from averaging trials to all individual single trials, by which we obtain a set of LEP components for N150 and P200 (one pair for each single trial). Furthermore, we may use the available tools of the EEGLAB toolbox (http://www.sccn.ucsd.edu/eeglab; Delorme and Makeig

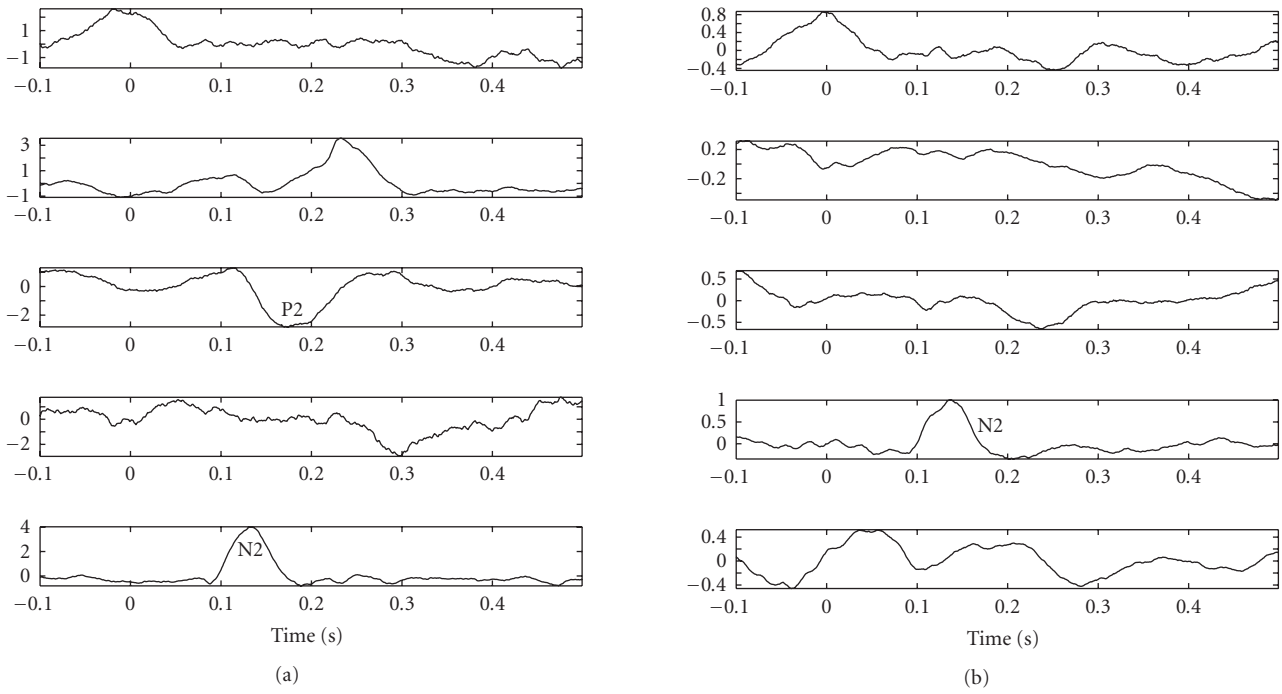


FIGURE 13: The extracted independent components (including LEPs) in signal-trials experiments for counting (left) and reading (right) tasks (subject 1, laser intensity 720 mJ).

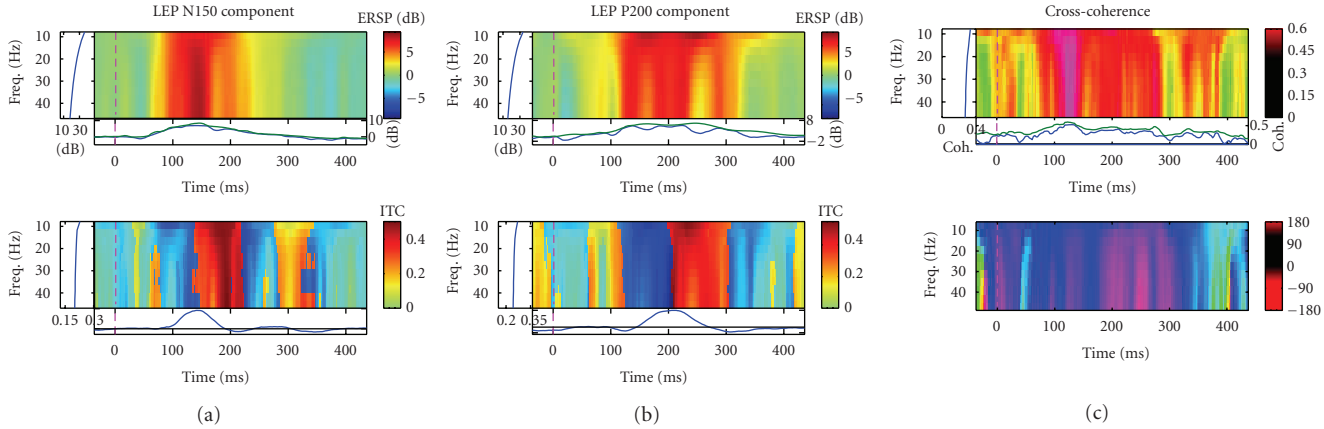
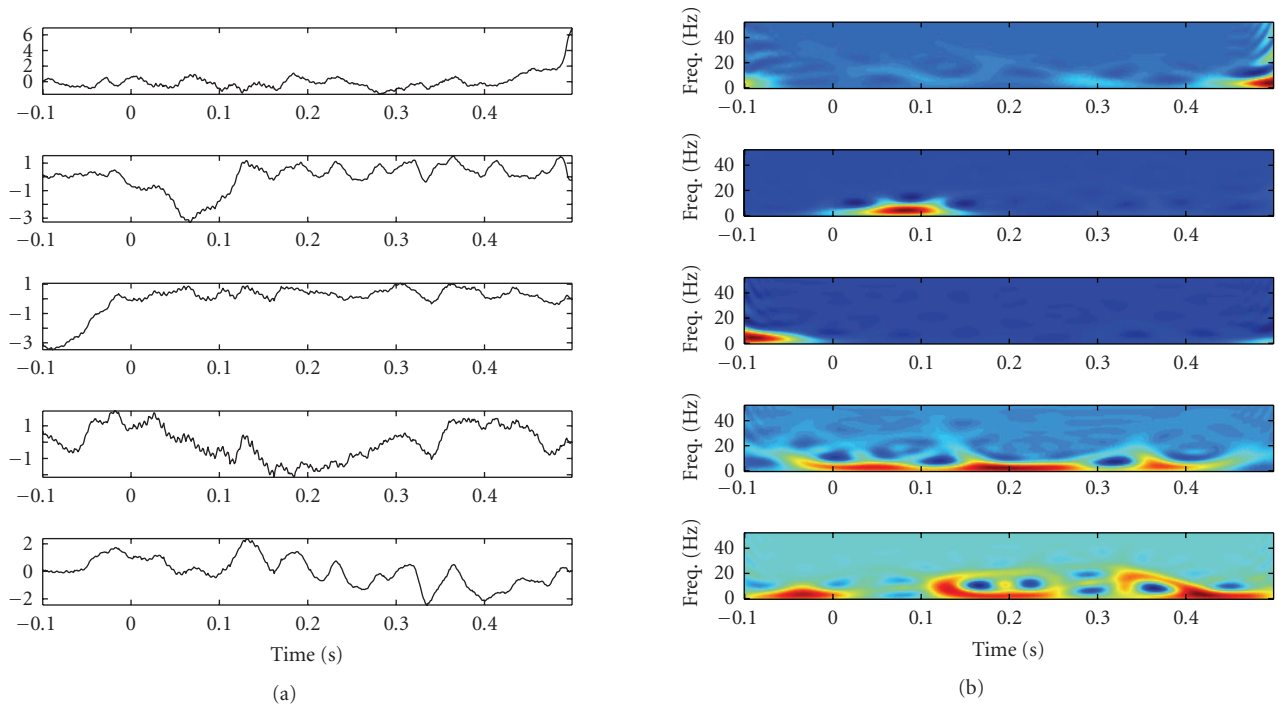


FIGURE 14: *Left and middle panels:* event-related (log) power spectral perturbation (ERSP, in dB, top row) and inter-trial coherence (ITC, from 0 to 1, bottom row) changes time locked to the LEP components in single trials (subject 1, attention task). *Right panel:* cross-coherence between LEP N150 and P200, with magnitude plot (from 0 to 1; top row) and phase plot (from −180 to 180 degree; bottom row).

[32], Delorme et al. [33], Makeig et al. [25]) to visualize the event-related spectral perturbation (ERSP) and the intertrial coherence (ITL) for the specific LEP components, as well as the cross-coherence between the independent LEP components. Specifically, the ERSP shows the spectral power change from prestimulus baseline (in dB) relative to the stimulus onset; and the ITL measures the consistency or reproducibility of the phase of stimulus-locked trial activity in the selected independent components. For instance, see Figure 14 for an illustrative example of two LEP components obtained from the attention task (recalling Figure 4). As seen in the figure, the cross-coherence magnitude (from 0 and 1) indicates the degree of synchronization between two independent LEP components, and the cross-coherence phase (from −180 to 180 degree) indicates that the LEP-N150 component is leading ahead of the LEP-P200 component.

5.3. Attention versus distraction

For subjects 1 and 2, consistent alpha waves were found among many (but not all) single trials in the reading task (i.e., distraction condition); whereas in the counting task



(a) (b)

Figure 15: *Left panels:* the 5 estimated sources extracted from a single-trial experiment of the reading task (subject 2). The 5th independent source contains typical alpha waves. *Right panels:* the corresponding time-frequency representation.

Table 5: Comparative statistics of the relative power of the normalized wavelet scalogram followed by Z-score transformation (for subject 2, no. 14 electrode) in single-trial analysis. The mean and standard deviation (mean ± SD) statistics are calculated by averaging the number of trials in each run.

Run	No. of trials	Intensity	θ power	α power	β power	γ power
1a	8	480 mJ	0.10 ± 0.76	0.24 ± 0.76	0.33 ± 0.70	0.27 ± 0.62
1a	11	640 mJ	0.22 ± 0.83	0.09 ± 0.48	0.03 ± 0.30	0.10 ± 0.49
1a	10	800 mJ	0.57 ± 2.92	0.09 ± 0.72	−0.03 ± 0.39	0.14 ± 0.67
1b	12	480 mJ	0.38 ± 1.63	0.19 ± 0.73	0.25 ± 0.44	0.22 ± 0.40
1b	11	640 mJ	0.58 ± 0.91	0.22 ± 0.57	0.08 ± 0.58	0.05 ± 0.48
1b	10	800 mJ	0.66 ± 1.23	0.59 ± 0.79	0.47 ± 1.39	0.41 ± 0.92
2a	10	480 mJ	0.29 ± 0.61	0.40 ± 0.57	0.53 ± 1.26	0.43 ± 0.66
2a	12	640 mJ	0.76 ± 1.62	0.35 ± 0.93	0.21 ± 0.66	0.11 ± 0.61
2a	10	800 mJ	0.77 ± 2.11	0.24 ± 0.69	0.12 ± 0.33	0.36 ± 0.51
2b	10	480 mJ	−0.16 ± 0.30	−0.04 ± 0.31	0.17 ± 0.28	0.31 ± 0.38
2b	10	640 mJ	0.30 ± 0.93	0.34 ± 0.85	0.28 ± 0.84	0.05 ± 0.44
2b	10	800 mJ	0.64 ± 1.56	0.21 ± 0.59	0.10 ± 0.27	0.20 ± 0.33

(i.e., attention condition), the significant alpha component was not observed in most of single trials. In some reading tasks, no obvious LEP was identified, while the dominant alpha waves can be observed. See Figure 15 for an illustration. In such cases, since there are no clear LEP peaks being observed, it remains an open question that whether this phenomenon is ascribed to "*habituation to the pain*" or "*loss of attention*." The reason that alpha rhythms appear frequently in the reading task might be due to the fact that the subject was in a relatively relaxed mood (especially compared with the counting task).

In addition, we also measure the coherence of signal-trial ECoG data under different conditions. In Figure 16, the coherency of alpha (8–12 Hz) and beta (12.5–32 Hz) bands between pairwise channels during the poststimulus period is illustrated. In order to visualize the coherency, putting all connections in one plot will be informative. Specifically, the complete 8-by-8 layout illustrates the first 64 electrodes' positions; at each electrode's position, we also plot aN 8-by-8 contour plot that represents the pairwise coherence between a specific electrode and the other electrodes, in which the specific electrode is marked by a relatively big filled

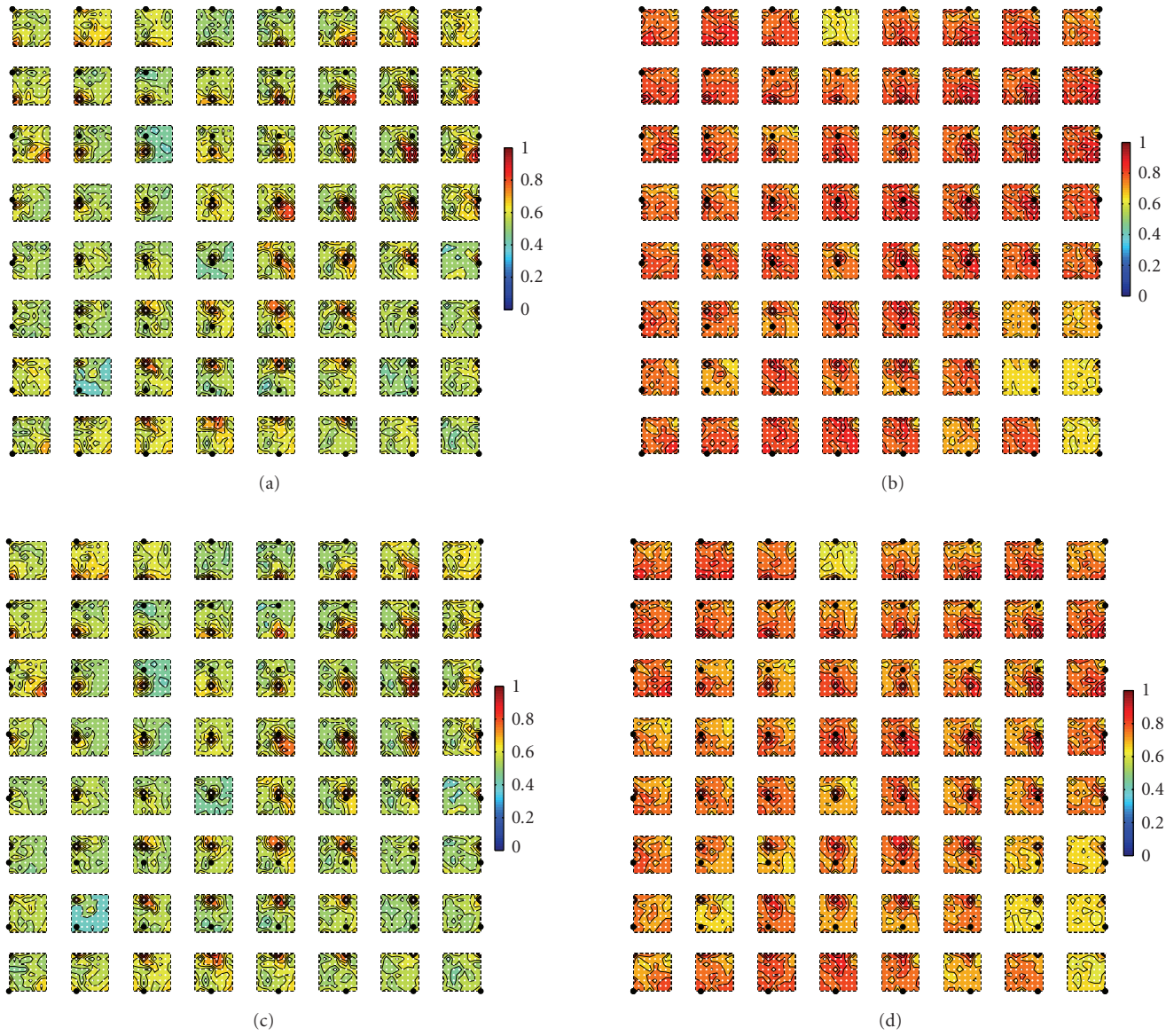


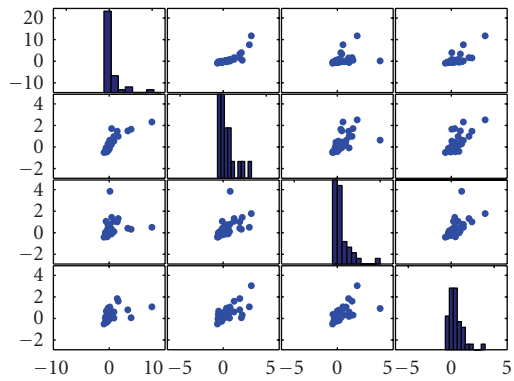
FIGURE 16: Pairwise coherence maps between the first 64 channels (subject 1, laser intensity 720 mJ) averaged over all single trials within a duration of 800 milliseconds in poststimulus period. (a) *alpha*-range coherence in the counting task; (b) *alpha*-range coherence in the reading task. (c) *beta*-range coherence in the counting task; (d) *beta*-range coherence in the reading task.

circle. As seen, typically there is strong coherence in the range of neighboring electrodes. Comparing Figure 16(a) with Figure 16(b), and Figure 16(c) with Figure 16(d), we can observe that there is stronger coherence in the alpha and beta bands in the distraction condition than in the attention condition.

5.4. LEP-component power versus laser intensity

For the same human subject in a series of single trials, it is expected that varying the level of stimuli (by changing the laser intensity), the *amplitude* and *latency* of the LEPs will consequently vary, so does the *power* of the LEP components in the time-frequency map. For this purpose of analyzing the power of LEP components at difference frequency bands, we have conducted quantitative and comparative analysis for subject 2 under varying controlled conditions.

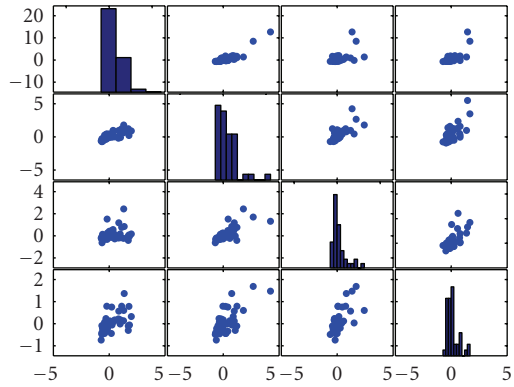
The power statistics are summarized in Table 5. It should be noted that the power values in Table 5 refer to the "Z-score transformed" poststimulus power according to (18), and all the values are averaged over the total number of trials in each run. The statistics are calculated for the first 64 channels including the one that has the highest power contribution (no. 14 channel) at each trial. The scatter plots of the Z-score transformed power for *theta, alpha, beta*, and *gamma* bands are shown in Figure 17. In the off-diagonal subplots, the scatter plots of cross-band power are shown; whereas the diagonal subplots show the histograms of the



$$\mathbf{C} = \begin{pmatrix} 1 & 0.909 & 0.591 & 0.837 \\ 0.909 & 1 & 0.734 & 0.843 \\ 0.591 & 0.734 & 1 & 0.772 \\ 0.837 & 0.843 & 0.772 & 1 \end{pmatrix}$$

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & .0303 & 0 \\ 0 & 0 & 0 & 0 \\ .0303 & .0003 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

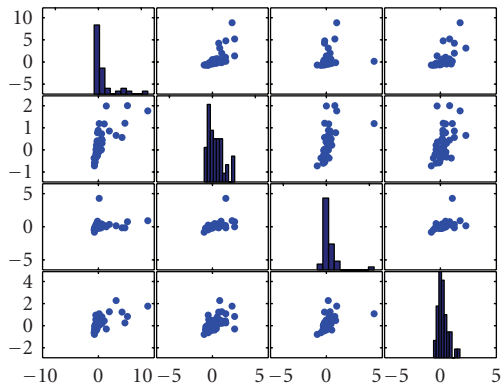
(a)



$$\mathbf{C} = \begin{pmatrix} 1 & 0.868 & 0.563 & 0.732 \\ 0.868 & 1 & 0.814 & 0.704 \\ 0.563 & 0.814 & 1 & 0.721 \\ 0.732 & 0.704 & 0.721 & 1 \end{pmatrix}$$

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & .0559 & .0003 \\ 0 & 0 & 0 & .0009 \\ .0559 & 0 & 0 & .0005 \\ .0003 & .0009 & .0005 & 0 \end{pmatrix}$$

(b)



$$\mathbf{C} = \begin{pmatrix} 1 & 0.841 & 0.866 & 0.801 \\ 0.841 & 1 & 0.691 & 0.694 \\ 0.866 & 0.691 & 1 & 0.660 \\ 0.801 & 0.694 & 0.660 & 1 \end{pmatrix}$$

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & .0269 & 0 \\ 0 & 0 & .0016 & .0014 \\ .0265 & .0016 & 0 & .0047 \\ 0 & .0014 & .0047 & 0 \end{pmatrix}$$

(c)

Figure 17: The scatter plots (using the MATLAB function "`plotmatrix`") of the Z-score transformed power for the *theta, alpha, beta*, and *gamma* bands: (a) 480 mJ, (b) 640 mJ, (c) 800 mJ, each based upon 40, 44, and 40 single trials, respectively. At each panel, the diagonal plots show the histograms of Z-score power of the associated frequency bands (from left to right, theta, alpha, beta, and gamma); the off-diagonal plots show the scatter plots of Z-score power across different frequency bands. Matrix **C** contains the correlation coefficients, and matrix **P** contains the associated P-values from the student's t-test.

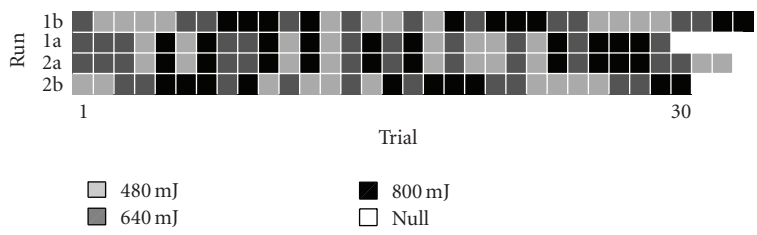


Figure 18: The graphical illustration of the laser intensity presentation orders at different runs (1a, 1b, 2a, and 2b). Note that the combined 62 trial sequences of "1a + 1b" and "2a + 2b" are of identical order.

power distribution in the relative frequency bands. In each subplot, the correlation coefficient between the power across different frequency bands is also calculated (stored in matrix **C**), as well as the associated P-values for the student's t-test (stored in matrix **P**). As seen, with different levels of laser intensities, the Z-score transformed power across different bands is correlated to certain degree: as the laser intensity increases, the degree of correlation at certain frequency bands (e.g., between *theta* and *alpha*) tends to decrease. A cut-off correlation coefficient of 0.7 was considered as a sign of significance. Each P-value indicates the probability of testing the hypothesis of no correlation, or the probability of getting a correlation as large as the observed value by random chance, when the true correlation is zero. If $P(i, j)$ is small (say, less than 0.05), then the correlation $C(i, j)$ is statistically significant.

From our data analysis, several observations are noteworthy.

(i) Compared to the prestimulus period, the power across different frequency bands in the poststimulus period mostly (or in majority) increases, as evidenced by the *positive* mean values of the Z-score transformed (relative) power, although their standard deviations are relative large.

(ii) In one specific run, the general trend is that the Z-score transformed θ power increases as the laser intensity increases; it seems that no general rule can be found for α, β, and γ power among our experiments.

(iii) In different runs (i.e., 1a, 1b, 2a, 2b), the mean power statistics with the same laser intensity often vary. This is not unreasonable because in each run the conditions of the subject may be different; in addition, the (random) order for presenting the laser stimuli is also different in each run (see Figure 18), their overall effects (say, e.g., between 480 → 640 → 800 and 640 → 800 → 480) would be certainly distinct. Such a "hysteresis" phenomenon is well known in psychology and psychophysics. In an effort to investigate this phenomenon, we take the 800 mJ intensity level as an example. According to Figure 18, the total numbers of 480 mJ, 640 mJ, and 800 mJ preceding 800 mJ are 10, 16, and 14, respectively. In order to compare their effects on the Z-score power, we calculate the means and standard deviations of different frequency bands under these three different conditions (namely, 480 → 800, 640 → 800, 800 → 800), and the results are shown in Figure 19. It is interesting to observe from the figure that their Z-score power statistics are quite different especially at the low-frequency (theta and alpha) bands. Generally, the Z-score power are highest for 480 → 800, followed by 640 → 800, and then lowest for 800 → 800—this is not surprising considering the sensation habituation effect. Statistical tests show at the theta and alpha bands, the pairwise comparison of Z-scored power among three conditions is statistically significant (ANOVA, $P < .01$).

5.5. LEP amplitude and latency versus laser intensity

Consistent with the previous studies (Ohara et al. [7, 8]), the peak amplitudes were measured from the baseline value, which was defined as the averaged value during the prestimulus period. Latencies were measured as the time of the peak amplitude (except for the artifact) for each component; and peak was regarded as significant when the peak amplitude was above the mean ± SD prestimulus level. However, in the previous studies, peak amplitudes and latencies were both measured from reproducible, averaged waveforms; here, we attempt to measure the latencies from single trials, while the amplitude will still be measured from averaging (over the trials at each run) because of its strong randomness; and the standard deviation of the amplitude estimate is calculated based on 4 independent runs among the recordings. In the meantime, we will focus the measurements on the first 64 electrodes (channels) for the primary somatosensory (SI) region, while the analyses for the parasylvian and medial frontal (MF) regions are ignored here. As observed in our experiments (Table 6), the averaged amplitudes of the LEPs (for both N2 and P2) increase as increasing laser intensity, except for one case of P2 under the 800 mJ condition; however, the mean statistic is also accompanied with a relatively large standard deviation, which reflects the random variations of measurement and/or subject conditions.

In our single-trial experiments, it was found that the latencies of the LEPs vary from trials to trials, evidenced by a large standard deviations (see Figure 20). In addition, by varying the laser intensity, the LEP-N2 and LEP-P2 also exhibit different attributes in terms of latency and amplitude. The corresponding statistics are summarized in Table 6 and Figure 20. Specifically, several observations are noteworthy.

(i) As seen in Table 6, the stronger is the laser intensity, the sooner the LEP appears; namely, the value of the LEP latency is smaller. See Figure 21 for two illustrative results.

(ii) When the laser intensity is small (e.g., 480 mJ), it is quite difficult to extract the LEP (either one or two) with the available ICA technique. This is partly because the LEP is so weak that it is overwhelmed in the background "noise" (brain activities). Indeed, it is even difficult to identify the peaks via visual inspection from the averaging recordings.

Generally, the amplitude of the LEP is a reflection of the sensation of the pain. Although it seems difficult to discover quantitative relationship between the intensity of the laser beam and the amplitude/latency of the LEPs, it is qualitatively clear that there exists correlation between them, especially when the intensity difference is large. This phenomenon might serve as a useful evaluation tool in the clinical practice.

To evaluate the statistical significance of the LEP peak amplitude and latency, we also conduct statistical tests based on their single-trial measurements. We first conduct a robust linear regression fit (using the MATLAB function "`robustfit`") between the laser intensity value (regression variable) and the measure of interest (amplitude or latency

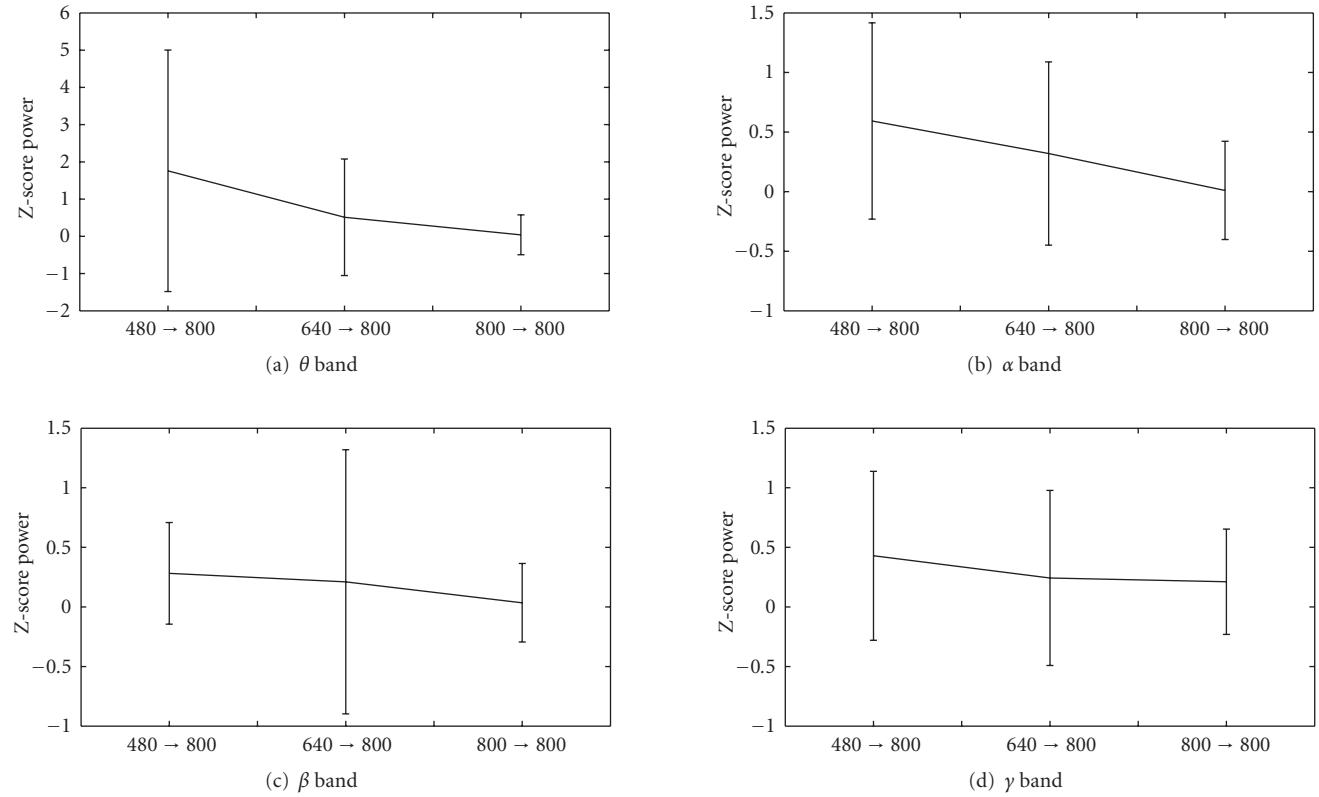


Figure 19: The Z-score power comparison of three different hysteresis effects. (a) *theta* band, (b) *alpha* band, (c) *beta* band, (d) *gamma* band.

Table 6: Comparative results of the estimated amplitudes and latencies of the LEPs (subject 2, under rating condition) from single and averaging trials. The last row indicate the selected number of single trials (by excluding some bad trials) used to evaluate the latencies.

	N2 (SI region)			P2 (SI region)		
Intensity (mJ)	480	640	800	480	640	800
Latency (milliseconds)	260	178	122	300	248	171
Amplitude (μV)	-121 ± 18	-125 ± 23	-150 ± 41	112 ± 31	126 ± 55	98 ± 20
No. of trials	28	36	34	22	26	26

of the LEP), and then obtain Pearson's correlation statistic r. Next, we calculate the t statistic as follows:

$$t = \frac{r\sqrt{\ell - 2}}{\sqrt{1 - r^2}}, \tag{19}$$

where ℓ denotes the number of regression sample pairs. From the t-statistic, we can further evaluate the statistical significance (i.e., P-value) from the t-table. In our case, we found the linear fit for LEP's latency is significant ($r = 0.87$, $P < .05$); however, the linear fit for LEP's amplitude is not significant.

5.6. Subjective sensation versus laser intensity

Finally, we follow the procedure of Ohara et al. [34] to analyze the relationship between the subjective sensation (in terms of pain rating) and the laser intensity. Specifically, the subject was asked to rate the pain level in decimal scale (0 no pain, 10 the most intense pain sensation). The mean and standard deviation statistics are calculated based on all single trials given three different laser intensities, as shown in Figure 22. Generally, it is seen that the average subjective pain sensation increases as the level of the laser intensity increases. Statistical tests show significant sensation differences between different levels of laser intensities (ANOVA, $P < .001$ between 480 mJ and 800 mJ; $P < .05$ between 640 mJ and 800 mJ). Moreover, we also evaluate the correlation between subjective sensation and LEP amplitude; however, no significant correlation was observed between the pain sensation rating and LEP amplitude for subject 2. We suspect this is partially due to the large variations among the subjective pain rating, even with the same laser intensity

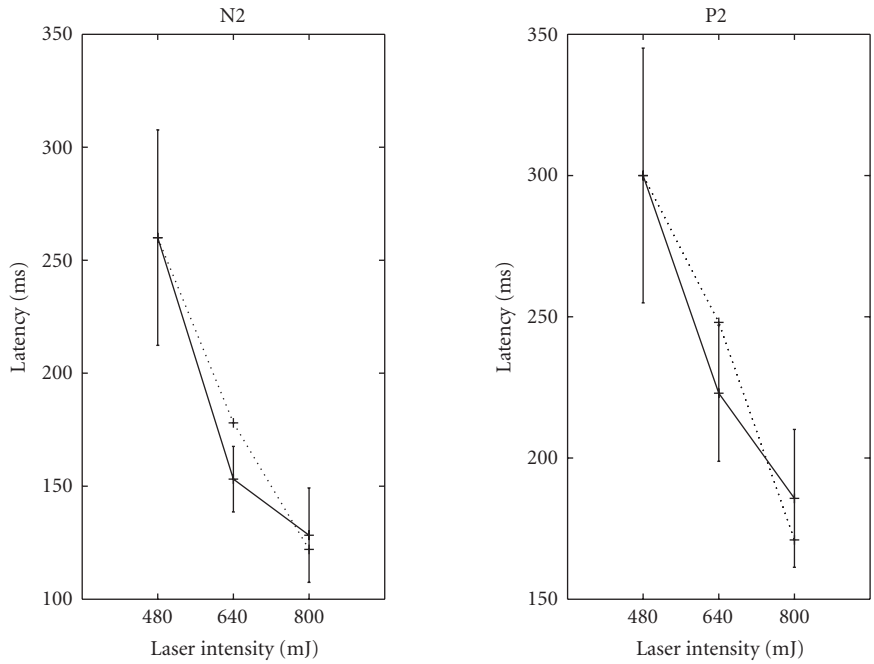


FIGURE 20: The error bars (mean ± SD) of the estimated latencies (from single trials) for LEP-N2 (left panel) and LEP-P2 (right panel) with varying laser intensities. The dotted lines indicate the estimated latencies from the averaging trials.

(specifically, the mean ± SD of the pain rating value for laser intensities 480 mJ, 640 mJ, and 800 mJ are 0.15 ± 0.70, 1.05 ± 1.49, and 2.10 ± 2.47, resp.). Although our data here seem to suggest that the subjective pain sensation and the objective LEP attribute observation might not be necessarily correlated, we should also be cautioned that the pain is a very complex sensation and is susceptible to many human factors and experimental conditions. Verification of any claim in this matter require more data and careful analysis.

6. DISCUSSION AND CONCLUSION

In this paper, we have used the statistical tools of FA/ICA for extracting and analyzing the LEPs. To our best knowledge, the statistical analysis and quantitative results reported here are among the new (if not the first) reports that apply sophisticated and systematic statistical analyses to the laser-induced pain data in the literature. In both averaging and single trials, we have demonstrated that the pain-evoked event potentials can be extracted and further analyzed with careful design of statistical procedure, and that the ICA/BSS approaches show a promising role in analyzing the multichannel ECoG data recorded from the awake human subjects. Our results here have also validated our previous findings in the early investigations and the reported neurophysiological observations in the literature. This is encouraging in that it justifies the merits of blind signal processing for neurobiological or physiological data analysis. The next challenge of this line of research is to extract *consistently* less-dominant (in terms of power) and potentially important pain-related components that are beyond the LEPs from single trials, which will be the subject of future study.

We have focused on one particular type of blind signal processing tool (namely, ICA) in this paper. However, we make no claim that the choice is unique or optimal. Indeed, we have been aware of the strengthes and weaknesses of the ICA during the experimental investigations (e.g., Makeig et al. [25]), although other improved ICA models, such as the spatially constrained ICA (Ille et al. [35], Hesse and James [36]) or the temporally constrained ICA (James and Gibson [37]), can be considered. It is also noteworthy to point out several other powerful blind signal processing tools and statistical algorithms, which might be valuable for the future investigation:

(i) nonnegative matrix factorization (NMF) (e.g., Lee and Seung [38]), which is an *approximate* matrix factorization method for nonnegative data (e.g., spectra, or time-frequency map). Unlike ICA, the independence assumption is relaxed or unnecessary in NMF, on the other hand, extra constraints (such as the smoothness or sparsity) can be imposed for this statistical model.[12]

(ii) parallel factor analysis (PARAFAC) (e.g., Bro [39]), which is a well-suited method for analyzing high-dimensional tensorial data; PARAFAC can be viewed

[12] For various implementation, see online resources and MATLAB toolboxes developed in our lab: http://www.bsp.brain.riken.jp/ICALAB/nmflab.html.

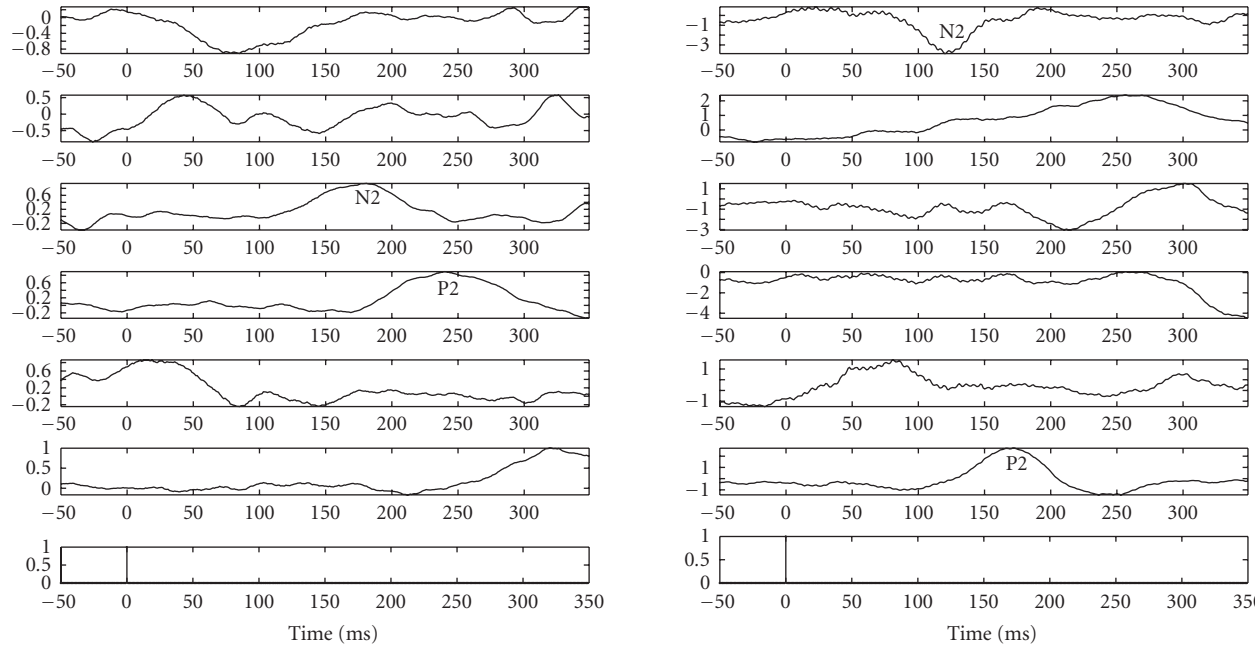


Figure 21: The 6 estimated sources extracted from averaged trials (subject 2) with different laser intensities (left panel: 640 mJ, right panel: 800 mJ).

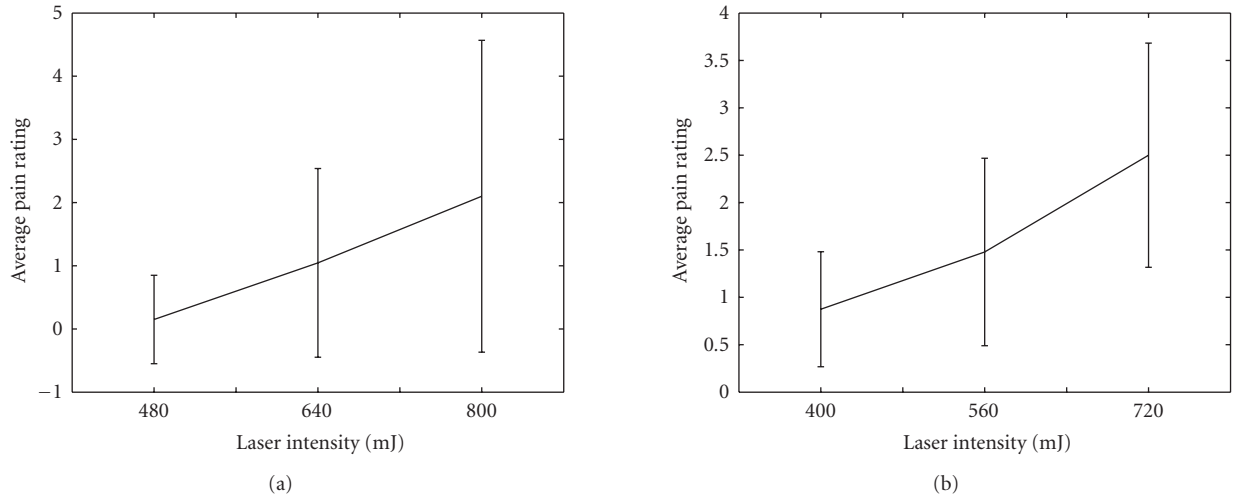


Figure 22: Pain sensation rating at different levels of laser intensity (left panel, averaged over all 124 single trials for subject 2). This is compared with another subject (right panel, averaged over 124 single trials, data from Ohara et al. [34]).

as a generalization of higher-order FA or high-dimensional NMF (if additional nonnegativity constraint is imposed).

(iii) common spatial subspace decomposition (CSSD) (Wang et al. [21]), which is a spatial filtering method for extracting signal components specific to one condition from multichannel electrode recordings given multiple task conditions. This kind of common spatial pattern algorithm may be used for evaluating the ECoG recordings under different task conditions; however, unlike the ICA method, it is a supervised algorithm that uses labeled data for classification.

In addition to the above-mentioned statistical tools, it would be also interesting to investigate the instantaneous brain activities and dynamics (Makeig et al. [25]), which may provide useful information of interactions inside the brain for specific patients with ECoG recordings. Finally, we believe what we reported here is only the first step towards a complete "statistical" understanding of the pain-evoked ECoG

data, the substantiation of our observations, claims, and conclusion made in this article would require more experimental verification of ECoG recordings in the future.

APPENDIX

MAXIMUM LIKELIHOOD ESTIMATION OF FACTOR ANALYSIS

Let us consider a general factor analysis (FA) model as follows:

$$\mathbf{x}_t = \boldsymbol{\mu} + \mathbf{B}\mathbf{z}_t + \boldsymbol{\epsilon}_t, \tag{A.1}$$

where $\mathbf{x}_t \in \mathbb{R}^m$ denotes the observed variable, $\boldsymbol{\mu}$ denotes the mean vector, $\mathbf{z}_t \in \mathbb{R}^n$ denotes the hidden variable called "factor," and $\mathbf{B}$ is an $m \times n$ "loading matrix." With the probabilistic assumptions that $\mathbf{z}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, and $\mathbb{E}[\mathbf{z}_t \boldsymbol{\epsilon}_t] = \mathbf{0}$, then we may derive that

$$\begin{aligned} \mathbb{E}[\mathbf{x}_t] &= \boldsymbol{\mu}, \qquad \mathbb{E}[\mathbf{x}_t \mid \mathbf{z}_t] = \boldsymbol{\mu} + \mathbf{B}\mathbf{z}_t, \\ \operatorname{Var}[\mathbf{x}_t] &= \mathbf{B}\mathbf{B}^T + \boldsymbol{\Sigma}, \qquad \operatorname{Cov}[\mathbf{x}_t, \mathbf{z}_t] = \mathbf{B}. \end{aligned} \tag{A.2}$$

Let $\boldsymbol{\theta} = (\boldsymbol{\mu}, \mathbf{B}, \boldsymbol{\Sigma})$ denote the unknown parameters, then the log likelihood of the FA model is written as

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) &= \sum_{t=1}^{T} \ln p(\mathbf{x}_t \mid \boldsymbol{\theta}) \\ &= -\frac{T}{2} \ln |\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{t=1}^{T} \mathbf{x}_t^T \mathbf{x}_t \\ &\quad - \frac{1}{2} \sum_{t=1}^{T} \left\{ (\mathbf{x}_t - \mathbf{B}\mathbf{z}_t)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_t - \mathbf{B}\mathbf{z}_t) \right\} \\ &= -\frac{T}{2} \ln |\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{t=1}^{T} \operatorname{tr}[\mathbf{x}_t \mathbf{x}_t^T] \\ &\quad - \frac{1}{2} \sum_{t=1}^{T} \operatorname{tr}[(\mathbf{x}_t - \mathbf{B}\mathbf{z}_t)(\mathbf{x}_t - \mathbf{B}\mathbf{z}_t)^T \boldsymbol{\Sigma}^{-1}], \end{aligned} \tag{A.3}$$

where $\operatorname{tr}[\cdot]$ denotes the trace operator, and $|\boldsymbol{\Sigma}|$ denotes the determinant of $\boldsymbol{\Sigma}$. Maximizing $\mathcal{L}(\boldsymbol{\theta})$ with respect to the unknown parameters yields the maximum likelihood estimate. An elegant solution can be obtained by using the iterative EM algorithm.

ACKNOWLEDGMENTS

The authors thank two anonymous reviewers for their critical and valuable comments. This project was done in collaboration between the Laboratory for Advanced Brain Signal Processing at RIKEN Brain Science Institute and the Department of Neurosurgery and Neurology at Johns Hopkins University. The experimental data reported in this article were recorded in the Johns Hopkins Hospital. We thank Liangyu Zhao for assistance during the early investigation. The ICA software—"ICALAB" based on MATLAB platform is available for download in our website http://www.bsp.brain.riken.jp/ICALAB.

REFERENCES

[1] B. Bromm and J. Lorenz, "Neurophysiological evaluation of pain," *Electroencephalography and Clinical Neurophysiology*, vol. 107, no. 4, pp. 227–253, 1998.

[2] L. García-Larrea, P. Convers, M. Magnin, et al., "Laser-evoked potential abnormalities in central pain patients: the influence of spontaneous and provoked pain," *Brain*, vol. 125, no. 12, pp. 2766–2781, 2002.

[3] L. García-Larrea, R. Peyron, B. Laurent, and F. Mauguière, "Association and dissociation between laser-evoked potentials and pain perception," *NeuroReport*, vol. 8, no. 17, pp. 3785–3789, 1997.

[4] I. M. Tarkka and R.-D. Treede, "Equivalent electrical source analysis of pain-related somatosensory evoked potentials elicited by a CO_2 laser," *Journal of Clinical Neurophysiology*, vol. 10, no. 4, pp. 513–519, 1993.

[5] G. D. Iannetti, R. K. Niazy, R. G. Wise, et al., "Simultaneous recording of laser-evoked brain potentials and continuous, high-field functional magnetic resonance imaging in humans," *NeuroImage*, vol. 28, no. 3, pp. 708–719, 2005.

[6] F. A. Lenz, M. Rios, A. Zirh, D. Chau, G. Krauss, and R. P. Lesser, "Painful stimuli evoke potentials recorded over the human anterior cingulate gyrus," *Journal of Neurophysiology*, vol. 79, no. 4, pp. 2231–2234, 1998.

[7] S. Ohara, N. E. Crone, N. Weiss, and F. A. Lenz, "Attention to a painful cutaneous laser stimulus modulates electrocorticographic event-related desynchronization in humans," *Clinical Neurophysiology*, vol. 115, no. 7, pp. 1641–1652, 2004.

[8] S. Ohara, N. E. Crone, N. Weiss, R.-D. Treede, and F. A. Lenz, "Cutaneous painful laser stimuli evoke responses recorded directly from primary somatosensory cortex in awake humans," *Journal of Neurophysiology*, vol. 91, no. 6, pp. 2734–2746, 2004.

[9] S. Ohara, N. E. Crone, N. Weiss, H. Vogel, R.-D. Treede, and F. A. Lenz, "Attention to pain is processed at multiple cortical sites in man," *Experimental Brain Research*, vol. 156, no. 4, pp. 513–517, 2004.

[10] V. Legrain, J.-M. Guérit, R. Bruyer, and L. Plaghki, "Attentional modulation of the nociceptive processing into the human brain: selective spatial attention, probability of stimulus occurrence, and target detection effects on laser evoked potentials," *Pain*, vol. 99, no. 1-2, pp. 21–39, 2002.

[11] V. Legrain, R. Bruyer, J.-M. Guérit, and L. Plaghki, "Nociceptive processing in the human brain of infrequent task-relevant and task-irrelevant noxious stimuli. A study with event-related potentials evoked by CO_2 laser radiant heat stimuli," *Pain*, vol. 103, no. 3, pp. 237–248, 2003.

[12] A. Mouraux, J. M. Guérit, and L. Plaghki, "Non-phase locked electroencephalogram (EEG) responses to CO_2 laser skin stimulations may reflect central interactions between A∂- and C-fibre afferent volleys," *Clinical Neurophysiology*, vol. 114, no. 4, pp. 710–722, 2003.

[13] M. Ploner, J. Gross, L. Timmermann, B. Pollok, and A. Schnitzler, "Pain suppresses spontaneous brain rhythms," *Cerebral Cortex*, vol. 16, no. 4, pp. 537–540, 2006.

[14] T.-W. Lee, M. Girolami, and T. J. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources," *Neural Computation*, vol. 11, no. 2, pp. 417–441, 1999.

[15] J. Cao, N. Murata, S.-I. Amari, A. Cichocki, and T. Takeda, "Independent component analysis for unaveraged single-trial MEG data decomposition and single-dipole source localization," *Neurocomputing*, vol. 49, no. 1–4, pp. 255–277, 2002.

[16] J. Cao, N. Murata, S.-I. Amari, A. Cichocki, and T. Takeda, "A robust approach to independent component analysis of signals with high-level noise measurements," *IEEE Transactions on Neural Networks*, vol. 14, no. 3, pp. 631–645, 2003.

[17] S. Makeig, M. Westerfield, T.-P. Jung, et al., "Dynamic brain sources of visual evoked responses," *Science*, vol. 295, no. 5555, pp. 690–694, 2002.

[18] J. Anemüller, T. J. Sejnowski, and S. Makeig, "Complex independent component analysis of frequency-domain electroencephalographic data," *Neural Networks*, vol. 16, no. 9, pp. 1311–1323, 2003.

[19] F. Miwakeichi, E. Martínez-Montes, P. A. Valdés-Sosa, N. Nishiyama, H. Mizuhara, and Y. Yamaguchi, "Decomposing EEG data into space-time-frequency components using parallel factor analysis," *NeuroImage*, vol. 22, no. 3, pp. 1035–1045, 2004.

[20] M. Browne and T. R. H. Cutmore, "Low-probability event-detection and separation via statistical wavelet thresholding: an application to psychophysiological denoising," *Clinical Neurophysiology*, vol. 113, no. 9, pp. 1403–1411, 2002.

[21] Y. Wang, P. Berg, and M. Scherg, "Common spatial subspace decomposition applied to analysis of brain responses under multiple task conditions: a simulation study," *Clinical Neurophysiology*, vol. 110, no. 4, pp. 604–614, 1999.

[22] A. Galka, O. Yamashita, T. Ozaki, R. Biscay, and P. Valdés-Sosa, "A solution to the dynamical inverse problem of EEG generation using spatiotemporal Kalman filtering," *NeuroImage*, vol. 23, no. 2, pp. 435–453, 2004.

[23] A. Cichocki, "Blind signal processing methods for analyzing multichannel brain signals," *International Journal of Bioelectromagtism*, vol. 6, no. 1, 2004, http://www.ijbem.org/.

[24] A. Cichocki, "Generalized component analysis and blind source separation methods for analyzing multichannel brain signals," in *Statistical and Process Models of Cognitive Aging*, M. J. Wenger and C. Schuster, Eds., Erlbaum, Mahwah, NJ, USA, 2006.

[25] S. Makeig, S. Debener, J. Onton, and A. Delorme, "Mining event-related brain dynamics," *Trends in Cognitive Sciences*, vol. 8, no. 5, pp. 204–210, 2004.

[26] M. Mørup, L. K. Hansen, C. S. Herrmann, J. Parnas, and S. M. Arnfred, "Parallel factor analysis as an exploratory tool for wavelet transformed event-related EEG," *NeuroImage*, vol. 29, no. 3, pp. 938–947, 2006.

[27] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*, John Wiley & Sons, New York, NY, USA, 2002.

[28] C. Févotte and C. Doncarli, "Two contributions to blind source separation using time-frequency distributions," *IEEE Signal Processing Letters*, vol. 11, no. 3, pp. 386–389, 2004.

[29] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, New York, NY, USA, 2nd edition, 1999.

[30] C. Tallon-Baudry, O. Bertrand, C. Delpuech, and J. Pernier, "Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human," *The Journal of Neuroscience*, vol. 16, no. 13, pp. 4240–4249, 1996.

[31] E. Düzel, R. Habib, B. Schott, et al., "A multivariate, spatiotemporal analysis of electromagnetic time-frequency data of recognition memory," *NeuroImage*, vol. 18, no. 2, pp. 185–197, 2003.

[32] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods*, vol. 134, no. 1, pp. 9–21, 2004.

[33] A. Delorme, S. Makeig, M. Fabre-Thorpe, and T. Sejnowski, "From single-trial EEG to brain area dynamics," *Neurocomputing*, vol. 44–46, pp. 1057–1064, 2002.

[34] S. Ohara, N. E. Crone, N. Weiss, R.-D. Treede, and F. A. Lenz, "Amplitudes of laser evoked potential recorded from primary somatosensory, parasylvian and medial frontal cortex are graded with stimulus intensity," *Pain*, vol. 110, no. 1-2, pp. 318–328, 2004.

[35] N. Ille, R. Beucker, and M. Scherg, "Spatially constrained independent component analysis for artifact correction in EEG and MEG," *NeuroImage*, vol. 13, no. 6, supplement 1, p. 159, 2001.

[36] C. W. Hesse and C. J. James, "The FastICA algorithm with spatial constraints," *IEEE Signal Processing Letters*, vol. 12, no. 11, pp. 792–795, 2005.

[37] C. J. James and O. J. Gibson, "Temporally constrained ICA: an application to artifact rejection in electromagnetic brain signal analysis," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 9, pp. 1108–1116, 2003.

[38] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[39] R. Bro, *Multi-way analysis in the food industry: models, algorithms and applications*, Ph.D. thesis, University of Amsterdam, Amsterdam, The Netherlands, 1998.

Hindawi Publishing Corporation
Computational Intelligence and Neuroscience
Volume 2007, Article ID 21315, 7 pages
doi:10.1155/2007/21315

Research Article

A Novel Constrained Topographic Independent Component Analysis for Separation of Epileptic Seizure Signals

Min Jing and Saeid Sanei

Centre of Digital Signal Processing, Cardiff University, Cardiff CF24 3AA, Wales, UK

Correspondence should be addressed to Min Jing, jingm@cf.ac.uk

Received 30 December 2006; Accepted 27 May 2007

Recommended by Andrzej Cichocki

Blind separation of the electroencephalogram signals (EEGs) using topographic independent component analysis (TICA) is an effective tool to group the geometrically nearby source signals. The TICA algorithm further improves the results if the desired signal sources have particular properties which can be exploited in the separation process as constraints. Here, the spatial-frequency information of the seizure signals is used to design a constrained TICA for the separation of epileptic seizure signal sources from the multichannel EEGs. The performance is compared with those from the TICA and other conventional ICA algorithms. The superiority of the new constrained TICA has been validated in terms of signal-to-interference ratio and correlation measurement.

1. INTRODUCTION

Epilepsy is the most common brain disorder only second to stroke, which affects nearly 60 million people in the world [1]. Many studies have been carried out from different aspects in order to explore the mechanisms of epileptogenesis and the possible solutions for anticipation and therapia [1–5]. Seizure detection has been under research for approximately three decades [6]. The most popular methods are based on time-frequency analysis [7] and artificial neural networks [8]. These methods do not exploit the multichannel electroencephalogram (EEG) information effectively.

Independent component analysis (ICA) has been increasingly applied to brain signal analysis for decomposition of multivariate EEGs to extract the desired sources. It has found a fruitful application in the analysis of multichannel EEGs [9] including epileptic seizure signals. The applications include the implementation of joint approximate diagonalization of eigenmatrices (JADE) and fastICA for seizure detection [10, 11], artifact rejection from epileptic intracranial EEGs by minimization of mutual information [12] and spatial filtering [13], and tracking of the epileptiform activity by incorporating the spatial constraint within the fastICA [14]. A novel approach proposed in [15, 16] applied an ICA approach to separate the seizure signals for prediction purpose and verified the predictability of epileptic seizure from the scalp EEGs. The main concept of this approach is to consider the seizures as independent components which are linearly and instantaneously combined together and with the noise and artifacts over the scalp. Subject to the mutual independency of the sources, the independent components can be separated by ICA algorithms and the seizure sources can be selected by postprocessing. The traditional nonlinear analysis methods can be applied to these seizure components for investigation of predictability. This approach can be further improved if a better performance of separation can be achieved. The objective of this work is to develop such method which can provide more plausible estimation of the seizure sources and eventually pave the way for the prediction of epileptic seizures from the scalp EEGs.

The conventional ICA model is built based on the statistical assumptions such that (1) the source signals are statistically independent; (2) the independent components must have nonGaussian distributions; (3) the number of independent components are less or equal to the number of input channels [17]. The ICA model has its own limitations. Apart from the scale ambiguity and the permutation problem, sometime the classic ICA cannot take all the prior physiological information into account and the results of separation cannot be interpreted physiologically. That is why

in real applications the ICA algorithms have been modified to incorporate the relevant additional information into the separation processing as constraints to enhance both efficiency and efficacy of the process.

Topographic ICA (TICA) proposed by Hyvärinen et al. [18] is a modified ICA model, which relaxes the assumption of statistical independency of the components, considering the components topographically closed to each other are not completely independent but have certain dependencies. The dependencies are used to define a topographic order between these components. This provides a very efficient method for separation of the multichannel EEG source signals. Generally, the EEG recordings reveal the sum of the action potentials of the neural cells, which are very complicated to be understood physiologically and mathematically. The dependencies between such sources cannot be simply cancelled out by some statistical assumptions. In this paper, we show how TICA works for the separation of the epileptic seizure EEGs, and how the performance can be improved by introducing novel spacial and frequency constraints in TICA. (In this paper, the constrained TICA is denoted as CTICA).

The paper is organized as follows. Section 2 describes the algorithm development. First, the basic TICA model and principles are explained. Then, the CTICA model is developed. Section 3 gives the experimental results obtained by applying the proposed methods to the epileptic seizure EEGs. The performance of CTICA and TICA is compared, and the superiority of CTICA is demonstrated by comparing with other commonly used ICA algorithms. The final section concludes the paper.

2. ALGORITHM DEVELOPMENT

2.1. Topographic ICA

The conventional noise-free ICA model can be expressed as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t), \tag{1}$$

where $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots, x_n(t)]^T$, $\mathbf{x} \in \Re^n$ is the vector of observed signals at time t, $(\cdot)^T$ denotes transpose operation, $\mathbf{s}(t) = [s_1(t), s_2(t), \ldots, s_m(t)]^T$ is the unknown independent source, $\mathbf{s} \in \Re^m$, $m \leq n$ for over-determined mixtures, and $\mathbf{A} \in \Re^{n\times m}$ is the mixing matrix. The estimated sources $\mathbf{y}(t) = [y_1(t), y_2(t), \ldots, y_m(t)]^T$ can be obtained by a separation matrix $\mathbf{W}$ through the inversion of the above mixing model,

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t), \tag{2}$$

where $\mathbf{W} = \mathbf{A}^\dagger$ is the pseudoinverse of the mixing matrix and $\mathbf{WA} = \mathbf{I}$. In the conventional ICA, the sources are assumed to be completely statistically independent, and the estimated signals have no particular order. But in most real applications, some sources may be more or less dependent on each other, such as the EEG sources which are fired from close locations within the cortex. In order to estimate the dependency of the independent components, Hyvärinen et al. proposed the TICA [18]. In TICA, the independency of the components has been relaxed, which means that the sources geometrically far from each other in topography are considered approximately independent and those close to each other are assumed to have certain dependencies. The dependency is defined as the higher-order correlation between the estimated sources, such as the correlation of the energies:

$$\operatorname{cov}(s_i^2, s_j^2) = E\{s_i^2 s_j^2\} - E\{s_i^2\}E\{s_j^2\} \neq 0, \tag{3}$$

where $\operatorname{cov}(\cdot)$ is the covariance of the two sources s_i and s_j, and $E\{\cdot\}$ is the expectation operator. Therefore, the estimated sources from the TICA are still uncorrelated, but their energies are not.

In the TICA model, the variances of estimated components are not constant, instead, they are generated by some high-order independent variables. These variables are mixed linearly in the topographic neighborhood, which are defined by a neighborhood function $h(i, j)$. Based on this model, the estimated components in the same neighborhood are energy-correlated. The approximation of the density of source s is given as [18]

$$\widetilde{p}(s) = \prod_k \exp\left(G\left(\sum_i h(i,k)s_i^2\right)\right), \tag{4}$$

where k is the index of the components within the same neighborhood. $G(\cdot)$ is the scalar function defined by incorporating certain nonlinearity. In this work, $G(\cdot)$ has been defined in [18]:

$$G(y) = -\alpha\sqrt{\epsilon + y}, \tag{5}$$

where α and ϵ are scalar constants.

The approximation of the log likelihood of this model is given in the following equation; more details of the derivation can be found in [18]:

$$\log\widetilde{L}(\mathbf{W}) = \sum_{t=1}^{N}\sum_{j=1}^{n} G\left(\sum_{i=1}^{n} h(i,j)\left(\mathbf{w}_i^T\mathbf{x}(t)\right)^2\right) + N\log\left(|\det\mathbf{W}|\right), \tag{6}$$

where $\mathbf{w}_i$ is the column vector of the unmixing matrix, N is the length of the data, and $h(i, j)$ is the neighborhood function, which can be defined as a monotonically decreasing function of some distance. The second term of the above equation can be ignored, since the unmixing matrix is constrained to be orthogonal and the determinant of an orthogonal matrix is one. Therefore, the estimation of the TICA model changes to choosing the optimal matrix $\mathbf{W}_{\text{opt}}$ that maximizes the above log-likelihood function. The estimation of maximization of the log likelihood of (6) can be found by

$$\frac{\partial}{\partial\mathbf{W}}\log\widetilde{L}(\mathbf{W})|_{\mathbf{W}=\mathbf{W}_{\text{opt}}} = 0. \tag{7}$$

The gradient is obtained as [18]

$$\nabla_{\mathbf{W}_k} = 2\sum_{t=1}^{N}\mathbf{x}(t)\left(\mathbf{w}_k^T\mathbf{x}(t)\right)\sum_{j=1}^{n}h(k,j)g\left(\sum_{i=1}^{n}h(i,j)\left(\mathbf{w}_i^T\mathbf{x}(t)\right)^2\right), \tag{8}$$

where $g(\cdot)$ is the derivative of the scalar function $G(\cdot)$.

2.2. Constrained topographic ICA

The estimated components from the TICA may be dependent if they fall into the same neighborhood, that is, the sources coming from the nearby location will be grouped together. However, the performance of TICA algorithm has certain limits. It may not be easy to get the sources grouped together unless the nearby sources are active at the same time. In [18], in order to obtain better visualization results, the experiment was designed to generate some typical high energy sources, such as biting teeth for 20 seconds. However, in most cases of real applications, the source signals may not be so significant, or there may be only one or two of active sources. Another factor is the number of input channels. It is obvious that the more input channels, the more information one can have and the better results can be achieved. This can be another limitation for the practical applications. However, the performance can be improved by introducing certain constraints into the algorithm.

Adding prior information, as a constraint, to classic ICA has been previously applied to EEG signal separation and analysis [15, 19–23]. The conventional ICA does not exploit the dependency of the sources, therefore, does not always provide the desired outputs. For EEGs, there is valuable prior knowledge which can help to separate the desired sources. In this study, we consider two constraints which are based on spatial and frequency information. Firstly, in the focal epileptic seizures, the location of the seizure sources, "epileptogenic zone," is often known as the prior information. Secondly, the seizure signals manifest themselves within certain frequency band. Based on the research findings from the clinicians and the neurologists, although the dominant frequency may vary for different types of seizures, the frequency band of the epileptic seizure onset is normally from 2.5 to 15.5 Hz. (Frequencies below 2.5 Hz are considered to be mainly due to eye-blinking artifacts) [24–26]. Therefore, the constraint can be determined based on both spatial and frequency domain information. The model of the constrained TICA problem can be expressed as

$$\max J_m(\mathbf{W}), \quad \text{s.t. } \min J_c(\mathbf{W}), \tag{9}$$

where $J_m(\mathbf{W})$ is the main cost function, which is based on TICA as shown in (6). $J_c(\mathbf{W})$ is the constraint which can be defined as minimizing the distance between the output and a reference signal:

$$J_c(\mathbf{W}) = \arg\min_{\mathbf{w}}\sum_{t=1}^{N}\left\|\mathbf{w}_i^T\mathbf{x}(t) - \mathbf{y}_r(t)\right\|_2^2, \tag{10}$$

where $\mathbf{y}_r$ is the reference signal defined based on the spatial and frequency constraints and $\|\cdot\|_2$ measures the Euclidean distance. The CTICA is then changed to an unconstrained function by using a Lagrange multiplier. Therefore, the overall cost function is written as

$$J(\mathbf{W},\Lambda) = J_m(\mathbf{W}) - \Lambda J_c(\mathbf{W}), \tag{11}$$

where $\Lambda = \text{diag}\{\Lambda_{ii}\}$, $i = 1,\ldots,m$, is a diagonal weight matrix formed by

$$\Lambda = p\cdot\text{diag}(\text{cor}(y_r, y_i)), \tag{12}$$

where p is an adjust constant, $\text{cor}(\cdot)$ is the correlation measurement, and y_i is the ith estimated source. Then, the update equation is obtained as

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mu(k)\left\{\frac{\partial J_m(\mathbf{W})}{\partial\mathbf{W}} + \Lambda\left(\mathbf{X}(\mathbf{W}\mathbf{X} - \mathbf{Y}_r)^T\right)\right\}, \tag{13}$$

where μ is the learning rate which is updated iteratively. $\mathbf{Y}_r$ is the matrix with the reference signal y_r in each row.

3. EXPERIMENT

The experiments consist of the application of the proposed CTICA algorithm to two patients with focal epileptic seizure. Generalized seizure was not considered in this work because the main purpose of this study was to investigate the predictability of epileptic seizure which is possible for only focal seizures. The epileptogenic zone was confirmed by the clinical experts as the prior information. Both patients' data contained epileptic seizure onset were truncated from the original long recording EEGs and were used in the experiments to validate the algorithm. The first experiment compared the performance of CTICA and TICA in terms of the signal-to-interference ratio (SIR). The second experiment provided the comparison of CTICA and three algorithms in terms of correlation measurement. Both experiments used topography to assist the visualization of the results.

In order to evaluate the performance, SIR was defined to be the averaged signal energy for the estimated source $y(t)$ from the direct source divided by the energy stemming from the other sources; higher value of SIR indicates a better performance:

$$\text{SIR} = \frac{(1/m)\sum_i^m\left|\mathbf{W}_{ii}^{-1}\right|^2\left\langle\left|\mathbf{y}_i\right|^2\right\rangle}{(1/m(m-1))\sum_{i\neq j}^m\sum_j^m\left|\mathbf{W}_{ij}^{-1}\right|^2\left\langle\left|\mathbf{y}_j\right|^2\right\rangle}, \tag{14}$$

where $\mathbf{W}_{ii}^{-1}$ includes the diagonal elements in the inverse of unmixing matrix, that is, the weights from source y_i to sensor x_i. The off-diagonal elements $\mathbf{W}_{ij}^{-1}$ provide the weights from the source y_j to the sensor x_i. It shows how the source y_j interferes the source y_i, since each column of the inverse of unmixing matrix indicates the distribution of each source in the mixtures.

The parameters used in the experiments were set up as follows. In (5), the scalar function $G(\cdot)$ parameters α and ϵ are chosen, respectively, as 1 and 0.005 refering to [18]. The adjust constant p in (12) was chosen between 6 to 10 based on the experiments performance. The initial value of learning rate μ in (13) was set to 0.1.

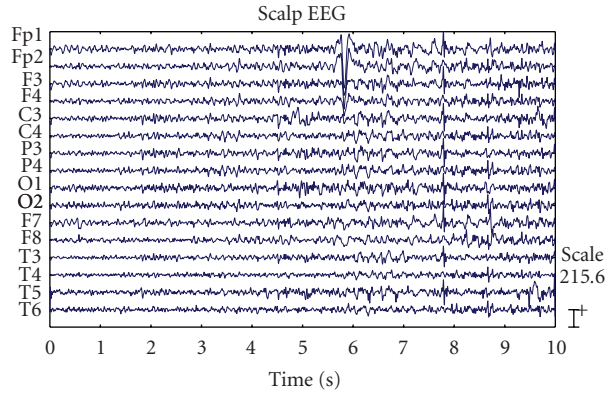


Figure 1: Multichannel EEG signals from an epilepsy patient including the seizure onset.

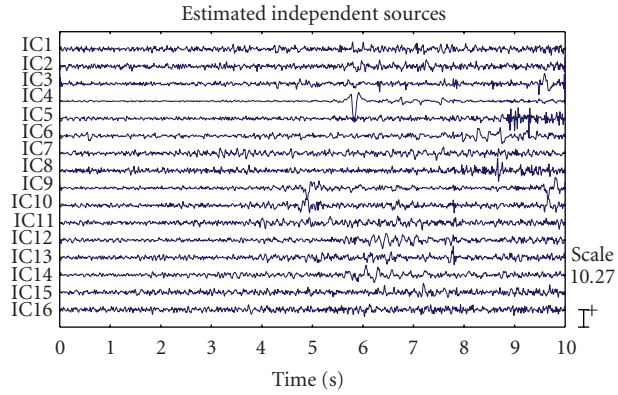


Figure 3: The EEG source signals estimated by TICA.

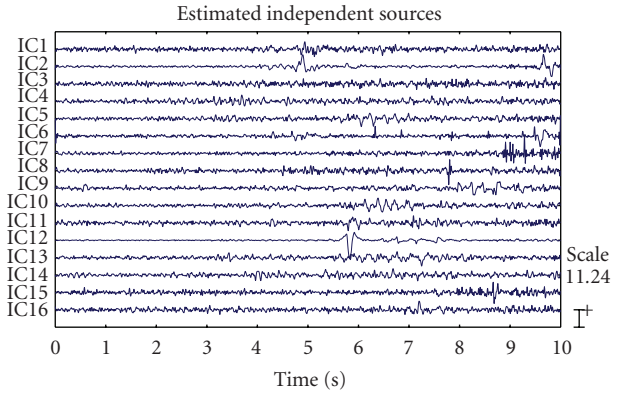


Figure 2: The EEG source signals estimated by TICA.

3.1. Experiment I

3.1.1. Data acquisition and the experiment setup

The multichannel EEGs with the frontal focal epileptic seizure were recorded using the standard silver cup electrodes applied according to the "Maudsley" electrode placement system, which is a modification of the extended 10–20 system [27]. This system provides a more extensive coverage of the lower part of the cerebral convexity, increasing the sensitivity for the recording from basal subtemporal structures. The 16 channels EEGs were sampled at 200 Hz and bandpass filtered in the frequency range of 0.3–70 Hz. The system input range was 2 mV and the data were digitized with a 12-bit analog-to-digital converter [15]. The signals were preprocessed by first removing the baseline to alleviate the effect of low frequency artifacts. Then, the EEGs were filtered by a 10th order Butterworth digital filter with a cut frequency of 45 Hz in order to eliminate the 50 Hz frequency component. The EEGs used in the following experiment were truncated from the original recordings to include the duration of 10 seconds with seizure onset as shown in Figure 1.

3.1.2. Reference

The reference signal was obtained by first averaging the special channels closed to the epileptogenic zone. In these experiments, F3, F4, F7, F8, C3, and C4 were selected. Then, 3–15 Hz bandpass filtering was undertaken to extract the information within the seizure frequency band. The final reference is a vector bounded within the designed spatial and frequency information of the seizure.

3.1.3. Neighborhood function

The neighborhood function indicates how the estimated sources are energy correlated with each other, which can be defined as a function of the width of the neighborhood. In this study, because of the limited number of input channels, the function was chosen as the simple one-dimensional form, such as $h(i, j) = 1$, if $|i - j| \leq m$, otherwise, $h(i, j) = 0$, where m is the width of the neighborhood. It can be noticed that the neighborhood function is symmetric as $h(i, j) = h(j, i)$.

3.1.4. Results

The separation results of TICA and CTICA are given in Figures 2 and 3. Figure 7 gives the convergence curve of CTICA. Both algorithms used the width of neighborhood $m = 1$. A simple detection rule based on the dominant frequency and respective estimated spectrum is applied to select the sources which have the significant ictal activities. The source with a maximum spectrum amplitude higher than a threshold and also with the dominant frequency in the seizure band, is taken as a seizure source. These sources are IC7, IC8, IC9, and IC10 in Figure 2, IC5, IC6, IC7, and IC8 in Figure 3. One can see that the high amplitude spike signals are separated from the other sources. Another distinct source related to the eye blink can be seen from two of the outputs, which is IC12 in Figure 2 and IC4 in Figure 3.

It may not be easy to decern the differences between the source candidates only by visual inspection of the time course of the sources, hence the topography was used to

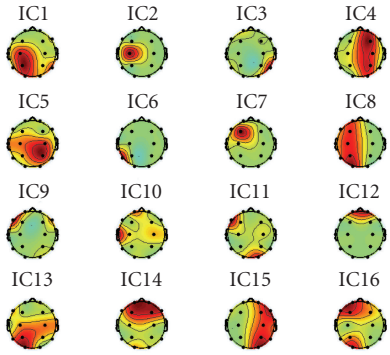


Figure 4: Topography of the estimated EEG sources from TICA.

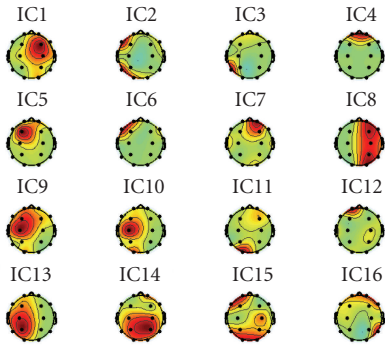


Figure 5: Topography of the estimated EEG sources from CTICA.

help visualization of the results. Figures 4 and 5 provide the topographies of the sources estimated, respectively, by TICA and CTICA. A topography can be obtained by back-projecting the estimated source onto the original signal space, that is, multiplying the column vector of the inverse of unmixing matrix by the corresponding estimated source. Topography reveals how the source signal contributes to each recordings, for example, one can notice that, in both sets of results, the distribution of eye blink (IC12 in Figure 4 and IC4 in Figure 5) appears on the area near the electrodes Fp1 and Fp2. It can be found that the four selected ICs are grouped together. The difference is, in Figure 5, the selected ICs (IC5, IC6, IC7, and IC8) from the CTICA are localized in the frontal region, but in Figure 4, the distribution of the corresponding sources (IC7, IC8, IC9, and IC10) by the TICA are rather dispersed. For instance, for IC10, the spatial distribution is highlighted in both frontal and temporal areas. A similar result can be noticed for IC11.

The performance of the algorithm was evaluated by the average of five trials for both TICA and CTICA. The SIR was calculated based on the definition given in (14). Figure 6 illustrates the separation performance (SIR) via the changes of the width of the neighborhood. It can be noticed that the SIR of TICA decreases with the increase of the neighborhood width. This is because the wider the neighborhood is, the more the source will be separated based on energy correlation. However, for the CTICA, due to the spatial and frequency constraints, the SIR slightly decreases at the beginning, then stays approximately at certain level. It shows that, generally, the CTICA has a better performance than the TICA. It also works better than the TICA when the width of the neighborhood increases.

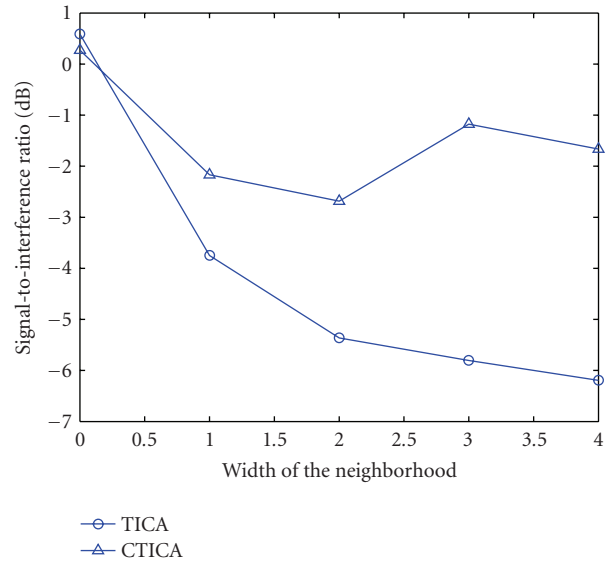


Figure 6: Performance comparison of TICA and CTICA.

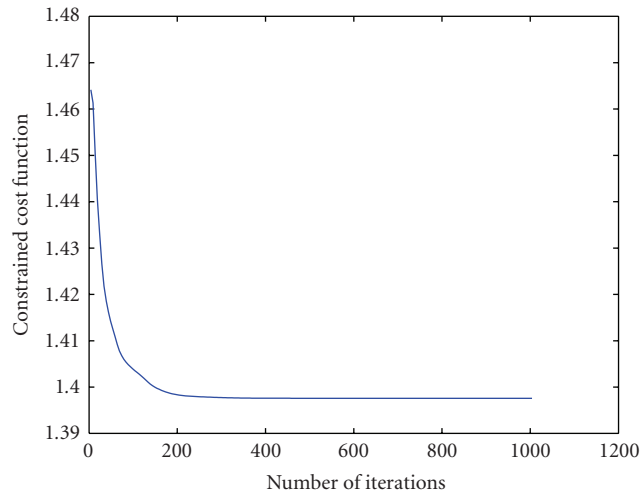


Figure 7: Algorithm convergence of CTICA.

3.2. Experiment II

3.2.1. Data acquisition and the experiment setup

In order to validate the performance of CTICA, in the second experiment, CTICA and other three popular ICA algorithms (JADE, SOBI, and Infomax) were applied to a patient with the right temporal seizure. The multichannel EEGs were obtained from a simultaneous EEG-fMRI recording system, in which the data were recorded during the fMRI scanning process. The fMRI scan period was 3 seconds and the scanner artifacts within EEGs were removed by the data provider.

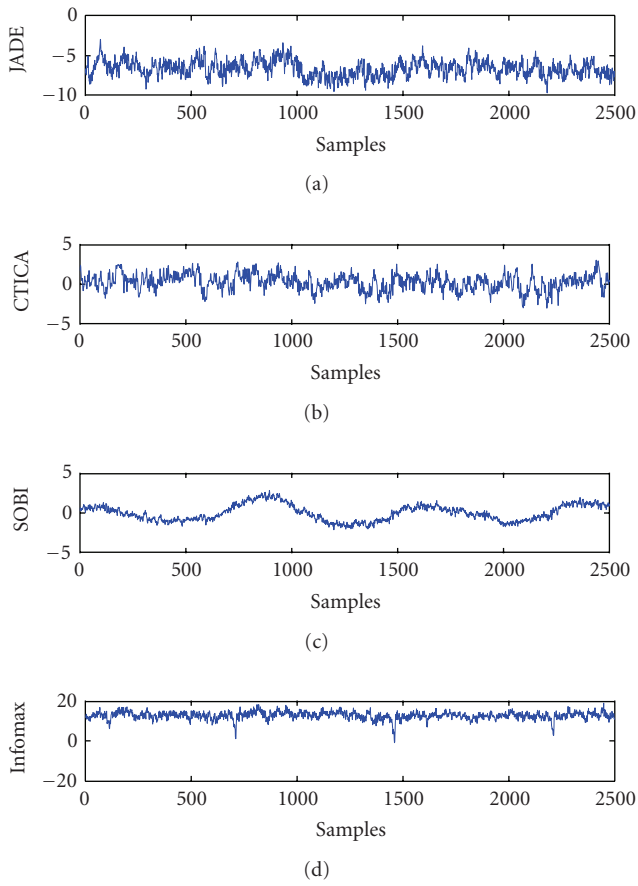


Figure 8: The separated sources from four ICA algorithms. The source which had the maximum correlation with the reference was selected from each algorithm.

The 64 channels EEGs were sampled at 250 Hz and filtered by a 10th order Butterworth low-pass digital filter with a cut frequency of 45 Hz. The data were then truncated with duration of 10 seconds for the separation. The reference signal was formed by averaging the signals from two electrodes T8 and P8. The width of the neighborhood function was $m = 1$. The rest of the parameters was set as in the first experiment.

3.2.2. Results

The performance of the four algorithms were compared in terms of correlation coefficient. For each algorithm, the source which had the maximum correlation with the reference was selected, are the correlation coefficient is shown in Table 1. It can be seen that the source obtained from CTICA has the maximum correlation with the reference and the source from Infomax has the minimum correlation.

The time course of the selected source is shown in Figure 8. It can be noticed that the source from Infomax has clear spikes with a period of 3 seconds, which is the same as the fMRI scan period. The spikes were most likely the fMRI scanner artifacts remained in the EEGs, and Infomax seemed not to separate these artifacts from the desired sources.

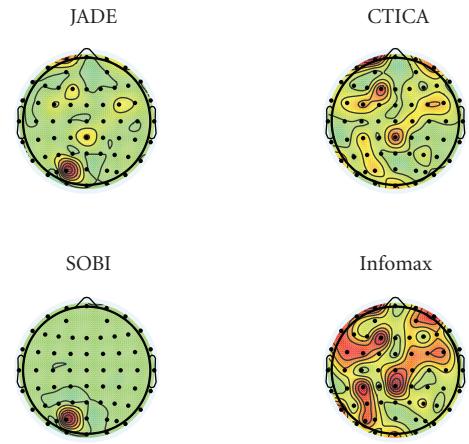


Figure 9: Topography of the selected sources from four algorithms.

Table 1: Correlation between reference and selected source.

JADE	CTICA	SOBI	Infomax
0.5510	0.6832	0.5142	0.3292

Figure 9 compares the topography of the sources selected from the four algorithms. It can be seen that although the topography does not highlight them at the area of interest (which can be due to the depth of the sources), the sources from JADE and CTICA have shown the distribution around the these regions (the right temporal area), and CTICA performs better than JADE. SOBI does not provide the promising result in the area of interest. Topography of the source from Infomax highlights a quite large area in the brain, which is typically caused by the scanning process. This also matches its source time course, in which the spikes were due to the scanner artifacts (as in Figure 8).

4. CONCLUSION

A novel constrained topographic ICA algorithm has been developed for separation of the epileptic seizure signals, which not only relaxes the independence assumption of nearby sources, but also further constrains the mixing model in spatial and frequency domains by using application-specific knowledges of epileptic seizures in the form of an averaged and band-limited reference signal. The CTICA algorithm achieves better performance than other ICA algorithms in terms of the SIR and correlation with the reference signal. This provides very promising results for further application of epileptic seizure analysis.

ACKNOWLEDGMENT

The authors wish to acknowledge Dr. Khalid Hamandi, the Consultant Neurologist from University Hospital of Wales for providing the data and valuable clinical advices.

REFERENCES

[1] L. D. Iasemidis, "Epileptic seizure prediction and control," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 5, pp. 549–558, 2003.

[2] F. H. L. da Silva, W. Blanes, S. N. Kalitzin, J. Parra, P. Suffczynski, and D. N. Velis, "Dynamical diseases of brain systems: different routes to epileptic seizures," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 5, pp. 540–548, 2003.

[3] M. L. van Quyen, J. Martinerie, V. Navarro, M. Baulac, and F. J. Varela, "Characterizing neurodynamic changes before seizures," *Journal of Clinical Neurophysiology*, vol. 18, no. 3, pp. 191–208, 2001.

[4] L. D. Iasemidis, D.-S. Shiau, W. Chaovalitwongse, et al., "Adaptive epileptic seizure prediction system," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 5, pp. 616–627, 2003.

[5] L. D. Iasemidis, J. C. Principe, and C. J. Sackellares, "Measurement and quantification of spatio-temporal dynamics of human epileptic seizures," in *Nonlinear Signal Processing in Medicine*, pp. 294–318, IEEE Press, Piscataway, NJ, USA, 1999.

[6] J. Gotman and P. Gloor, "Automatic recognition and quantification of interictal epileptic activity in the human scalp EEG," *Electroencephalography and Clinical Neurophysiology*, vol. 41, no. 5, pp. 513–529, 1976.

[7] P. J. Franaszczuk, G. K. Bergey, P. J. Durka, and H. M. Eisenberg, "Time-frequency analysis using the matching pursuit algorithm applied to seizures originating from the mesial temporal lobe," *Electroencephalography and Clinical Neurophysiology*, vol. 106, no. 6, pp. 513–521, 1998.

[8] C. Kurth, F. Gllliam, and B. J. Steinhoff, "EEG spike detection with a Kohonen feature map," *Annals of Biomedical Engineering*, vol. 28, no. 11, pp. 1362–1369, 2000.

[9] S. Sanei and J. Chambers, *EEG Signal Processing*, John Wiley & Sons, New York, NY, USA, 2007.

[10] S. Sanei, M. A. Latif, and L. Shoker, "Detection and localization of epileptic seizures using ICA," in *Proceedings of the 3rd Sensor Array and Multichannel Signal Processing Workshop (SAM '04)*, pp. 173–176, Sitges, Spain, July 2004.

[11] S. Faul, L. Marnane, G. Lightbody, G. Boylan, and S. Connolly, "A method for the blind separation of sources for use as the first stage of a neonatal seizure detection system," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, vol. 5, pp. 409–412, Philadelphia, Pa, USA, March 2005.

[12] H. Liu, K. E. Hild II, J. B. Gao, D. Erdogmus, J. C. Príncipe, and J. C. Sackellares, "Evaluation of a BSS algorithm for artifacts rejection in epileptic seizure detection," in *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS '04)*, vol. 1, pp. 91–94, San Francisco, Calif, USA, September 2004.

[13] N. Ille, P. Berg, and M. Scherg, "Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies," *Journal of Clinical Neurophysiology*, vol. 19, no. 2, pp. 113–124, 2002.

[14] C. W. Hesse and C. J. James, "Tracking epileptiform activity in the multichannel ictal EEG using spatially constrained independent component analysis," in *Proceedings of the 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS '05)*, pp. 2067–2070, Shanghai, China, September 2005.

[15] J. Corsini, L. Shoker, S. Sanei, and G. Alarcón, "Epileptic seizure predictability from scalp EEG incorporating constrained blind source separation," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 5, pp. 790–799, 2006.

[16] M. Jing, S. Sanei, J. Corsini, and G. Alarcón, "Incorporating BSS to epileptic seizure predictability measure from scalp EEG," in *Proceedings of the 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS '05)*, pp. 5950–5953, Shanghai, China, September 2005.

[17] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, New York, NY, USA, 2001.

[18] A. Hyvärinen, P. O. Hoyer, and M. Inki, "Topographic independent component analysis," *Neural Computation*, vol. 13, no. 7, pp. 1527–1558, 2001.

[19] W. Wang, M. G. Jafari, S. Sanei, and J. A. Chambers, "Blind separation of convolutive mixtures of cyclostationary signals," *International Journal of Adaptive Control and Signal Processing*, vol. 18, no. 3, pp. 279–298, 2004.

[20] M. A. Latif, S. Sanei, J. Chambers, and L. Shoker, "Localization of abnormal EEG sources using blind source separation partially constrained by the locations of known sources," *IEEE Signal Processing Letters*, vol. 13, no. 3, pp. 117–120, 2006.

[21] L. Spyrou, M. Jing, S. Sanei, and A. Sumich, "Separation and localisation of P300 sources and their subcomponents using constrained blind source separation," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, Article ID 82912, 10 pages, 2007.

[22] W. Wang, S. Sanei, and J. A. Chambers, "Penalty function-based joint diagonalization approach for convolutive blind separation of nonstationary sources," *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1654–1669, 2005.

[23] N. Ille, R. Beucker, and M. Scherg, "Spatially constrained independent component analysis for artifact correction in EEG and MEG," *NeuroImage*, vol. 13, no. 6, supplement 1, p. 159, 2001.

[24] G. Lantz, C. M. Michel, M. Seeck, O. Blanke, T. Landis, and I. Rosén, "Frequency domain EEG source localization of ictal epileptiform activity in patients with partial complex epilepsy of temporal lobe origin," *Clinical Neurophysiology*, vol. 110, no. 1, pp. 176–184, 1999.

[25] O. Blanke, G. Lantz, M. Seeck, et al., "Temporal and spatial determination of EEG-seizure onset in the frequency domain," *Clinical Neurophysiology*, vol. 111, no. 5, pp. 763–772, 2000.

[26] P. Wahlberg and G. Lantz, "Approximate time-variable coherence analysis of multichannel signals," *Multidimensional Systems and Signal Processing*, vol. 13, no. 3, pp. 237–264, 2002.

[27] J. L. Fernández Torre, G. Alarcón, C. D. Binnie, and C. E. Polkey, "Comparison of sphenoidal, foramen ovale and anterior temporal placements for detecting interictal epileptiform discharges in presurgical assessment for temporal lobe epilepsy," *Clinical Neurophysiology*, vol. 110, no. 5, pp. 895–904, 1999.

Hindawi Publishing Corporation
Computational Intelligence and Neuroscience
Volume 2007, Article ID 83416, 18 pages
doi:10.1155/2007/83416

Research Article

Clustering Approach to Quantify Long-Term Spatio-Temporal Interactions in Epileptic Intracranial Electroencephalography

Anant Hegde,[1] Deniz Erdogmus,[2] Deng S. Shiau,[3] Jose C. Principe,[1] and Chris J. Sackellares[3,4]

[1] *Computational NeuroEngineering Laboratory, Department of Electrical & Computer Engineering, University of Florida, Gainesville, FL 32611, USA*
[2] *Department of Computer Science and Electrical Engineering CSEE, OGI School of Science & Engineering, Oregon Health & Science University, Portland, Beaverton, OR 97006, USA*
[3] *Optima Neuroscience, Inc., Gainesville, FL 32601, USA*
[4] *Malcolm Randal VA Medical Center, Gainesville, FLa, FL 32608, USA*

Correspondence should be addressed to Anant Hegde, ashegde2004@gmail.com

Received 18 February 2007; Accepted 19 August 2007

Recommended by Saied Sanei

Abnormal dynamical coupling between brain structures is believed to be primarily responsible for the generation of epileptic seizures and their propagation. In this study, we attempt to identify the spatio-temporal interactions of an epileptic brain using a previously proposed nonlinear dependency measure. Using a clustering model, we determine the average spatial mappings in an epileptic brain at different stages of a complex partial seizure. Results involving 8 seizures from 2 epileptic patients suggest that there may be a fixed pattern associated with regional spatio-temporal dynamics during the interictal to pre-post-ictal transition.

1. INTRODUCTION

There is sufficient evidence to believe that the brain dynamics can be effectively modeled through complex nonlinear interactions. Application of nonlinear dynamical measures [1, 2] such as short-term Lyapunov exponents (STLmax) and correlation dimension on an epileptic brain have revealed that the complexity of the brain dynamics reduces significantly as a seizure is approaching. In other words, the temporal dynamics of the brain progresses from a "high-dimensional" nonconvergent (chaotic) state to a much smaller dimensional "chaotic" state.

Much of the analysis on temporal dynamics focuses on analyzing and characterizing the irregular behavior of the time signal of either intracranial or scalp EEG. However, it is important to realize that the brain is a multidimensional system with a large set of neuronal oscillators that are physically and functionally coupled together. Obviously, neurons communicate with each other through synaptic potentials resulting in microscopic action potential discharges. Abnormal neural population synchrony can also produce mesoscopic transient activity, clinically called sharp waves or spikes. Depending on the pathophysiological states, the nature of the spikes with respect to their frequency of occurrence, amplitude, and shape, can be very distinctive. Particularly in an epileptic brain, it would be natural to expect the distinctions between interictal, preictal, and ictal spikes could possibly be a consequence of the dynamical changes in spatio-temporal communications between various regions of the brain. Therefore, it is essential to unravel the functional connectivity of the neural networks and analyze how the structures change during seizure events.

Even though observations that the macroscopic EEG cannot be distinguished from linearly correlated noise [3], many nonlinear approaches have been able to extract interregional coupling information in a manner that would not have been possible by spectral approaches. Nonlinear dependencies between multiple signals have been studied in the last two decades, with the hope of enhancing the tool set provided by the linear methods. Unfortunately, they have faced some practical implementation problems such as sensitivity to noise, choice of parameters, and the high computational cost. Most of the state-space methods rely on finding the functional dependencies between two-time series based

on how their trajectories in the embedded phase space describe each other. Inspired by the similarity–index (SI) technique introduced by Arnhold et al. [4], we earlier proposed a self-organizing map (SOM)-based computationally efficient measure, SOM-SI [5, 6], to measure asymmetric dependencies between time sequences. Conceptually, the SI and the SOM-SI methods rely on the assumption that if there is a functional dependency between two signals, the neighboring points in the state space of one signal correspond to neighborhoods of their counterpart. The SOM-SI method maps the embedded data from signals onto a quantized output space through an SOM [7, 8] specialized on these signals, and utilizes the activation of SOM processing elements (PE) to infer about the influence directions between the signals. This approach reduces the computational complexity drastically by exploiting the accurate quantization properties of the SOM in representing the dynamics of the signal in the phase space. Our previous work [6] showed that the SOM-SI was capable of determining the temporal evolution of dependencies between various cortical sites, at different stages of temporal lobe epileptic seizures.

Epileptic seizures, in particular, are characterized by dynamic states (interictal, ictal, preictal, and postictal) that are known to possess both local and global spatio-temporal groupings. Channels associate and deassociate in time; however, depending on the psycho-physiological state of the brain, certain groups of channels might have a higher likelihood of sharing same channel connectivities, thus forging a long-term association. In epileptic intracranial EEG, identifying such state-dependent clusters may provide us with useful insights on the evolution of brain patterns during seizure states. In this study, we propose a spatio-temporal clustering model to qualitatively analyze the spatio-temporal groupings in multidimensional epileptic structures. Unlike in many other clustering approaches, where dynamical features extracted from the data are used as basis to determine groupings, our proposed clustering approach uses the dependencies among the original data recordings to do the same. Our approach, in short, essentially seeks to analyze the regional grouping of cortical sites at different stages of a seizure, based on their mutual interactions.

On a clinical perspective, this study intends to investigate spatio-temporal relationships across various regions of an epileptic brain to help determine the epileptic focus and the dynamical changes that lead to a seizure. In order to achieve this ultimately goal, it is necessary to develop appropriate signal processing tools that extract features to cluster different regions of the brain based on their functional dependencies. The highlight of this clustering measure is that it uses a similarity or a proximity matrix that is entirely data-dependent to determine regional dependencies. Our idea is two folds: (a) to propose a novel tool to determine clusters and present synthetic simulations and real data to support the validity and robustness of this measure, (b) to apply this measure on real-epileptic data and present a detailed clinical investigation on the outcome. The study was made on 8 complex partial seizures from 2 patients suffering from temporal lobe epilepsy. The conclusions of this paper are based on observations from these 8 seizures only.

The paper is organized as follows. We first present a brief review of SOM-SI in Section 2. Section 3 discusses the spectral-clustering approach and the proposed spatio-temporal cluster model. Data description is provided in Section 4 followed by clinical evaluation of the clustering approach on the epileptic EEG data, in Section 5. Section 6 discusses about potential directions for future study.

2. SIMILARITY INDEX (SI) MEASURE

(A) Original SI measure

Assume that X and Y are two time series generated by a system, which are embedded into two vector signals in time using delays. $N(X \mid Y)$ is defined as the average dependency of X on Y and it can be written as [5],

$$N(X \mid Y) = \frac{1}{N} \sum_{n=0}^{N-1} \frac{R^n(X) - R^n(X \mid Y)}{R^n(X)}, \tag{1}$$

where $R^n(X)$ is the average Euclidean distance between the state-vector of X^n and the remaining state-vectors in X. The Y-conditioned Euclidean distance $R^n(X \mid Y)$ measures the average Euclidean distance between X^n and the vectors in X whose corresponding time partners are the k-nearest neighbors of Y^n. This measure takes values in $[0, 1]$, where 0 implies no coupling and 1 implies perfect synchronization [4]. Average dependence of Y on X, $N(Y \mid X)$, is similarly computed. The difficulty with this approach is that at every time instant n, we must search for the k nearest neighbors of the current embedded signal vectors among all N sample vectors; this process requires $O(N^2)$ operations. This high complexity hinders real-time implementation and analysis. In addition, the measure depends heavily on the free parameters, namely, the number of nearest neighbors and the neighborhood size ε. The neighborhood size ε needs to be adjusted every time the dynamic range of the windowed data changes.

2.1. SOM-based similarity index (SOM-SI)

The self-organized-map- (SOM-) based SI algorithm [5] is fundamentally aimed at reducing the computational complexity of the SI technique. The central idea is to create a statistically quantized representation of the dynamical system using an SOM [7, 8]. An SOM is a neural-network in which spatial patterns from the input space are mapped onto an ordered output space consisting of a set of processing elements (PE). Thus each PE in the SOM, based on its location on the map, compactly models different features/dynamics of the input.

For best generalization, the map needs to be trained to represent all possible states of the system (or at least with as much variation as possible). As an example, if we were to measure the dependencies between EEG signals recorded from different regions of the brain, it is necessary to create an SOM that represents the dynamics of signals collected from all channels. The SOM can then be used as a prototype to represent any signal recorded from any spatial location on the brain, assuming that the SOM PEs have specialized in the dynamics from different regions.

One of the salient features of the SOM is topology preservation; that is, the neighboring PEs in the feature space correspond to neighboring states in the input data. In the application of SOM modeling to the similarity index concept, the topology preserving quality of the SOM will be of added advantage, because of the fact that the neighboring PEs in the feature space will now correspond to neighboring states in the input data.

Assume X and Y are two time series generated by a system, which are embedded into two vector signals in time-using delays. Define the activation region of a PE in the SOM as the set of all input vectors (the embedded signal vectors) for which the PE is the winner based on some distance metric (Euclidean in most cases). Let X_n be the set of time indices of input vectors x_j that are in the activation region of the winner PE corresponding to the input vector x_n at time n. Similarly define the set Y_n.

Then the procedure to estimate the directed SOM-SI between X and Y is as follows:

(1) Train an SOM using embedded vectors from both X and Y as the input.
(2) At time n, find W_n^x, the winner PE for vector x_n, and find W_n^y, the winner PE for vector y_n.
(3) To find $R^n(X)$, compute the average Euclidean distance between W_n^x and all the other winner PEs in the SOM. Similarly, compute $R^n(Y)$.
(4) Determine the sets X_n and Y_n for W_n^x and W_n^y, respectively.
(5) Determine the nearest PEs $W_{n,j}^y$ corresponding to vectors y_j, where $j \in X_n$. Determine the nearest PEs $W_{n,j}^x$ corresponding to vectors y_j, where $j \in Y_n$.
(6) Calculate $R^n(X \mid Y) = (1/q)\Sigma_{j=1}^q \| W_n^x - W_{n,j}^x \|$, where q is the number of elements in X_n. Calculate $R^n(Y \mid X) = (1/q)\Sigma_{j=1}^q || W_n^y - W_{n,j}^y ||$, where q is the number of elements of Y_n.
(7) Compute the ratios,

$$N^n(X \mid Y) = \left(R^n(X) - R^n(X \mid Y)\right)/R^n(X),$$
$$N^n(Y \mid X) = \left(R^n(Y) - R^n(Y \mid X)\right)/R^n(Y). \quad (2)$$

(8) Find *interdependencies* $N(X \mid Y)$ and $N(Y \mid X)$ as the average of $N^n(X \mid Y)$ and $N^n(Y \mid X)$ over all n.
(9) Compute the SOM-SI as the difference,

$$\chi = N(Y \mid X) - N(X \mid Y). \quad (3)$$

Positive values of χ indicate that influence of X on Y is more than the influence of Y on X, while negative values indicate the opposite. Higher magnitude of χ indicates a stronger coupling of the signals.

The computational savings of the SOM approach is an immediate consequence of the quantization of the input (signal) vector space. The nearest neighbor search involves $O(NM)$ operations as opposed to $O(N^2)$ in the original SI, where M is the number of PEs. Traditionally $M \ll N$, hence, SOM-SI offers a significantreduction in computations compared to original SI.

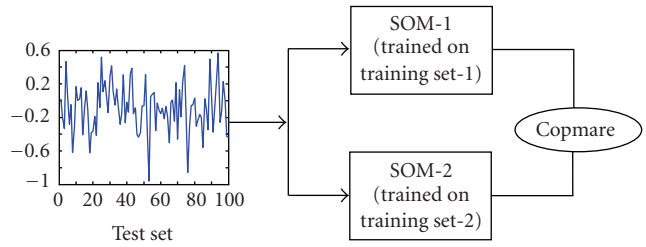


Figure 1: Experimental setup to compare SOM-Similarity Indices obtained from two (2) separate maps.

2.2. Testing the robustness of SOM-SI on multiple SOMs

To illustrate the accuracy of the SOM-based measure, we previously presented a few experimental simulations [5, 9] involving synthetically constructed linear and nonlinear interactive models. Results from each of them demonstrated the accuracy of our quantized measure, validated through statistical quantification with results from the original SI measure. For application on seizures especially, a 25×25 sized, 2-dimensional SOM grid was trained to embed all the dynamical states of an EEG attractor. SOM, being one of the most important elements of this improvised measure, one of the pre-requisites of this approach, is to ensure that: (a) for data modeling purposes, the training set captures the variance found in the dynamics of the ictal states from all the channels for a given patient and (b) the similarity indices computed using the SOM's processing elements are independent of the SOM and the corresponding training dataset. Put in other words, pair-wise similarity indices computed on two separate SOMs should be significantly close to each other if not equal.

While the previous test results [9] were a testimony to the former, the independence of the observed interactions through similarity indices to a given SOM needed to be tested before proceeding with extensive data analysis. From the multivariate EEG data samples of an epileptic patient, two separate training sets were constructed. One of the training sets (say training set-1) consisted of portions of data sampled from the interictal, ictal, preictal, and postictal states of seizures 1 and 2. The other training set (say training set-2) consisted of data portions picked around seizure 4 and 5. Using the same normalization procedures on both the sets and with the same set of training parameters as before, two separate SOMs (called as SOM-1 and SOM-2 for convenience) were trained. Post training, the SOM-similarity indices were obtained from pair-wise analysis of interdependence among channels chosen from the ROF and LOF regions of the brain, as illustrated in Figure 1.

Test data from three (3) recording sites in right orbitofrontal region (namely, ROF1, ROF2, and ROF3) and 3 sites from left orbitofrontal regions (namely LOF1, LOF2, and LOF3) were picked from intervals surrounding seizures 4, 5, 6, 7 and seizure 11, respectively. The similarity index profiles $\{N^1(X \mid Y)\}_t$ and $\{N^2(X \mid Y)\}_t$ obtained from computing the SOM-SI on large intervals (say time $t = 1, \ldots, T$)

of seizure data are quantitatively compared using the classical correlation coefficient and error-percentage as the comparison metrics. The error-percentage is computed as follows:

$$\{e\} = 100^* \left\{ \frac{N^1(X \mid Y)_t - N^2(X \mid Y)_t}{N^1(X \mid Y)_t} \right\}_{t=1}^{T}, \quad (4)$$

where $N(X \mid Y)$ is the normalized interdependency of X on Y. Note that the notations X and Y are used to denote the two channels of interest. Normalized error e quantifies the percentage difference between the interdependency values from SOM-2 and SOM-1, keeping interdependency value from SOM-1 as the reference. From the error population, the fraction of the absolute error values less than 20% and the fraction less than 10% are computed to determine the degree of dependence of the SOM-SI measure on the data used to train a SOM.

For illustration, the results from analyzing the interdependency of LOF3 on LOF4 on various seizures are shown in Figure 2. The histograms correspond to the error ensembles obtained from analyzing over long seizure intervals. Qualitatively, the superimposed traces in Figure 2 indicate the extent of agreement or disagreement between the SOM-SI profiles. Table 1 compiles a summary of the agreement between the SOM-SI profiles for about 13 hours of EEG data. A large fraction of errors less than 20%, supported by a high correlation coefficient between the two SOM-SI profiles, suggests that there was very little disparity between the SOM-SI profiles from SOM-1 and SOM-2. Besides, the high percentages also seem to suggest the EEG data dynamics might not vary drastically from one seizure to another, and therefore the two SOM models produced almost identical SI results. This finding consequently supports our original belief that a well-trained SOM and a well-picked training dataset is sufficient to carry out inter-dependency analysis on all the seizures of a patient.

Overall, pair-wise analyses of the interdependency among 6 channels (15 combinations) on 5 seizures of the epileptic patient were performed on SOM-1 and SOM-2. The average correlation coefficient and the error results between the SOM-SI profiles are shown in Table 2.

Results from Table 2 indicate that in around 80% of the times, the differences between the SOM-SI results are less than 20%. This is not surprising considering that the differences are measured in percentages (3), and therefore even small discrepancies in the case of small dependency values can appear magnified. In addition, we also speculate that the discrepancies could be the outcome of the two SOMs being trained in an identical fashion instead of being fine-tuned to obtain the lowest reconstruction error in each.

In general, if the SOMs can be designed to obtain the lowest reconstruction error, by iteratively choosing the best sets of parameters, a slight improvement in the performances can be easily achieved; but as it stands, a slight discrepancy can nevertheless be always expected although it may have very little impact in the overall scheme of analysis.

3. SPATIO-TEMPORAL CLUSTERING MODEL

Often time series structures collected from a multi-dimensional dynamical system share similar information that reflect system wide interactions or even synchronization abilities. By definition, the word similar could mean that the information shared among a set of channels are stronger than the information they share with other channels. Such spatial similarities could possibly be transient up to a few seconds or could even stretch to several minutes or hours. As we postulated earlier, dynamial similarities in spatio-temporal behavior could be one of the driving factors to trigger certain events in biological systems. From a clinical point of view, we believe that analyzing the temporal changes in channel similarities could reveal some interesting aspects about the epileptic brain.

Similarity-based time-series clustering [10, 11] is a well-researched topic in the area of dynamical graph theory. It is an extremely useful approach to characterize spatial groupings in time sequences. Similar time sequences are typically grouped based on their mutual interactions. In this study, using the SOM-SI as a computational tool to derive the distance/similarity/proximity matrix, we propose a clustering model to dynamically analyze the spatio-temporal groupings in mutivariate time sequences.

3.1. 1 Model for spatio-temporal clustering

In this section, we propose a clustering approach to extract information on spatio-temporal distribution of multivariate time measurements. A 3-fold approach, consisting of spatial-discretization of the data using spectral-clustering technique [12, 13], temporal quantification using Hamming distance, followed by application of another clustering technique, is presented in Figure 3. The rational will become apparent during the explanation.

Spectral clustering is one of the many clustering methods that use subspace decomposition on data-derived affinity matrix to achieve data-clustering. Using kernel methods, the data samples are projected onto a higher dimensional space where the discriminant analysis is much easier. Projecting the data onto a feature space results in tightly formed clusters such that the between cluster entropy is maximized and the within-cluster entropy is minimized. In our study, we apply the standard spectral clustering algorithm by Ng et al. [12] to spatially cluster the similarity indices obtained by the SOM-SI technique.

Pair-wise evaluation of SOM-SI measure on all the possible combinations (C_2^N, where N is assumed to be the number of channels) of a portion of a multivariate time series leads to $k = 2^*(C_2^N)$ similarity indices in $[0, 1]$. k is multiplied by 2 because of the asymmetric nature of the SOM-SI measure. If we imagine the time series as various inter-connected nodes in a multidimensional graph, the SOM-SI similarity indices represent the affinity or rather the weights of the connection between those nodes. Therefore, we can translate them into a square matrix of size $N \times N$, where N is the number of channels. Since the weighting is normalized between 0 and 1,

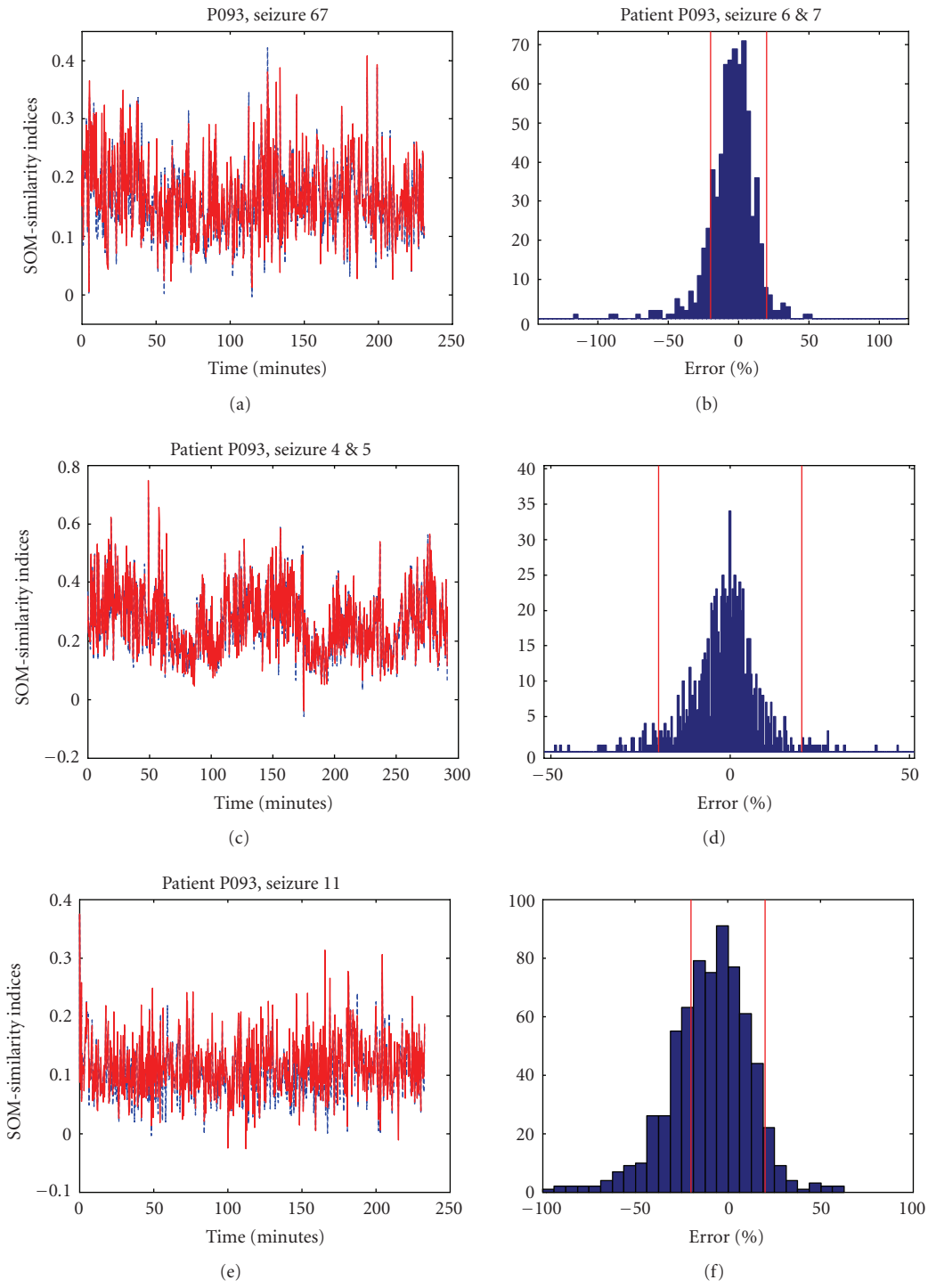


Figure 2: Comparing interdependencies between channels LOF3 and LOF4. Left: SOM-similarity profiles from the output of SOM-1 and SOM-2 are superimposed. Right: Histogram of the errors in %. Top: Seizure 4 and 5. Middle: Seizure 6 and 7. Bottom: Seizure 11.

Table 1: Quantitative comparisons between the SOM-SI profiles obtained from SOM-1 and SOM-2. LOF3 and LOF4 data was projected on each of the SOMs and then the SOM-SI measure was applied to analyze the dependency of LOF3 on LOF4.

Interdependency N(LOF3| LOF4)	Correlation Coefficient (%)	Fraction of error less than 20%	Fraction of error less than 10%
Seizure 6 and 7	95.74	0.8504	0.5597
Seizure 4 and 5	98.45	0.9234	0.7543
Seizure 11	91.59	0.6452	0.3614

Table 2: Summary of the comparisons between the SOM-SI profiles from SOM-1 and SOM-2. Each row represents the statistics (mean and variance) of pair-wise SOM-SI analyses of the epileptic EEG data from 6 channels (15 combinations).

	Correlation Coefficient (%)	Fraction of error less than 20%	Fraction of error less than 10%
Seizure 6 and 7	94.32 ± 2.85	0.79 ± 0.1	0.54 ± 0.12
Seizure 4 and 5	97.46 ± 1.08	0.91 ± 0.06	0.73 ± 0.12
Seizure 11	93.24 ± 2.06	0.71 ± 0.08	0.41 ± 0.07

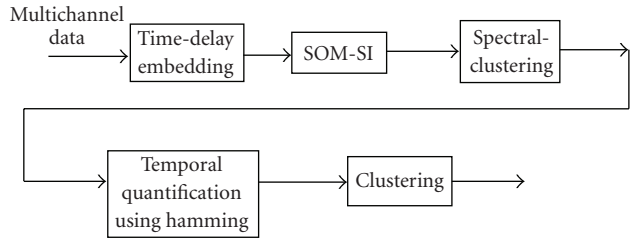


Figure 3: Block diagram to extract spatio-temporal groupings information in Multivariate EEG structures.

the diagonal elements, representing the affinity of a channel with itself, are coded as 1.

However, to be able to perform spectral-decomposition on an affinity matrix, Ng's algorithm [12] requires that the affinity matrix be square and symmetric in nature. This is because the eigen decomposition yields orthogonal column vectors (also called eigenvectors) only if the projection matrix is square-symmetric. The asymmetric matrix can be transformed to a symmetric matrix by adding it to its transpose and dividing each entry by 2. Following the eigen decomposition on the transformed affinity matrix, we have a set of labeled clusters representing the membership of the channels.

If the above procedure is repeated over consecutive time (T) windows (overlapping or nonoverlapping), channel groupings obtained on each time window ($t = 1 \cdots T$) can be arranged in a matrix (of dimension $N \times T$) as in (5).

$$\kappa_{\text{spect}} = \begin{bmatrix} 3 & 2 & 2 & . & . & . & . & 3 & 1 \\ 1 & 2 & 2 & . & . & . & . & 3 & 2 \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ 3 & 1 & 2 & . & . & . & . & 1 & 2 \end{bmatrix}. \tag{5}$$

To characterize the average clustering of the channels over a longer period of time, we propose another, albeit simple, hierarchical clustering approach that uses Hamming distance to derive the proximity matrix.

3.2. 2 Temporal quantification using hamming distance

We showed in the previous section that the multivariate time series can be grouped by using similarity-based clustering techniques such as spectral clustering. The spectrally clustered labels specify the groups of channels exhibiting high degree of within-cluster similarities and low degree of between-cluster similarities. Often in applications such as epileptic EEG analyses where associations last longer, it is important to identify channel groupings over a longer time-window.

State-dependent connections can be quantified by clustering rows of the κ_{spect} matrix that are similar with each other over a longer time interval, say T. In this context, we propose a simple statistic that computes the relative frequency of any two channels sharing the same labels/groupings to determine the degree of similarity. In other words, in a time window of length T, we check the average number of times when the two channels of interest, share the same cluster label.

In an algebraic context, the above operation is equivalent to computing pair-wise Hamming distance in a time window T. Similarity can be quantified by subtracting the Hamming distance from 1. That is, if d_{ij}^{ham} is the hamming-distance between channels "i" and "j," similarity in probabilistic terms can be obtained as

$$p_{ij}^{\text{sim}} = 1 - d_{ij}^{\text{ham}}. \tag{6}$$

Thus, computing the pair-wise similarity for all i and j combinations will result in a $\boldsymbol{P}$ matrix of size $N \times N$ (N is the number of channels). For convenience, we will call the matrix $\boldsymbol{P}$ the cluster-similarity matrix in all our future references.

Finally, hierarchical clustering on the cluster-similarity matrix $\boldsymbol{P}$ will yield information on the cluster groupings over a time T. In the context of EEG data, clustering will thus enable us to know the groups of channels that have similar behavioral structure in the brain, over a longer time frame.

4. EPILEPTIC EEG DATA DESCRIPTION

Intracranial EEG signals were recorded from the hippocampus, subtemporal and frontal cortex structures of epileptic patients having a history of complex-partial and secondary generalized seizures of temporal lobe focus, using bilaterally and surgically implanted electrodes (Figure 4). The clinical motivation for the location of the electrodes was mainly to identify focal area for presurgical evaluation. Using amplifiers with an input range of ±0.6 mv, the recorded signals were converted to a narrow-band using an antialiasing filter with a cutoff range between 0.1 Hz and 70 Hz. Using an analog-to-digital converter with 10-bit quantization precision, the narrow-band signals were sampled/digitized at 200 samples/sec. Measurements involved recording EEGs from multiple sensors (28 to 32, with common reference channels) and the recordings spanned over 6 continuous days. A total of 55 seizures, of temporal lobe onset were recorded from 5 patients, in the range of 6 to 18 seizures for each patient.

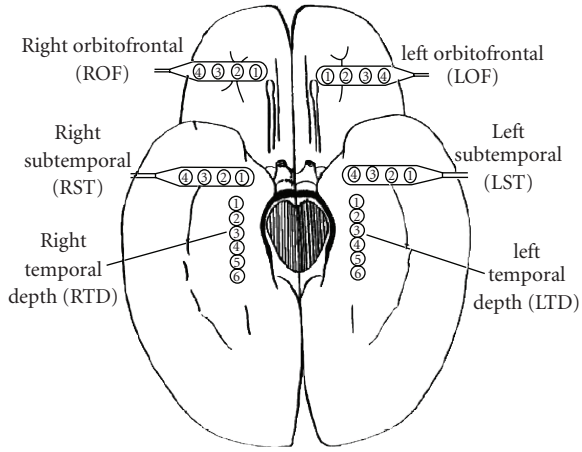


FIGURE 4: Diagram of the depth and subdural electrode montage in an epileptic brain. Electrode strips are placed over the left orbitofrontal (LOF), right orbitofrontal (ROF), left subtemporal (LST), right subtemporal cortex (RST). Depth electrodes are placed on the left temporal depth (LTD) and right temporal depth (RTD), to record hippocampus EEG activity.

The distinction of these patients from general patients with temporal lobe epilepsy is their seizures are medically refractory. In other words, these patients' seizures cannot be controlled by the currently available anticonvulsant medications. We note that all the patients had to undergo surgery as part of their treatment.

5. RESULTS

In the last section, we proposed a spatio-temporal model to extract groupings from long-term multivariate recordings. In this section, we will focus on the application of that model on the epileptic intracranial EEG time series. The first part of the section will describe the details on the application of the model and the second part will discuss the results of analyses on 8 seizures, from 2 patients. With respect to selecting seizures for our analysis, the underlying reasoning was to be able to understand the following:

(a) complex partial types of seizures;
(b) how the functional relationships among different cortical sites of the brain changed over time; and
(c) the temporal variability of functional relationships across successive seizures.

For (c), we selected pairs of seizures that were neither too close nor too distant in time to introduce coupling from previous seizure events or external effects of many other environmental variables. Therefore, we picked pairs of seizures that were between 60 minutes and 6 hours apart only. The minimum of 60 minutes was chosen so that the second seizure was not in the postictal region of the preceding seizure. Pairs of seizures more than 6 hours apart were treated as seizures in isolation and therefore were left out of the selection.

5.1. 1 Application on epileptic intracranial EEG data

The temporal changes in the spatial structure of an epileptic brain was analyzed on twenty four (24) representative channels recorded bilaterally from the orbitofrontal, temporal, and subtemporal regions on the brain. One of the fundamental requirements for analyzing the dynamics of a non linear system is to construct the state-space attractor from just a single recording of the time series. From previous studies that estimated intracranial EEG attractor size using correlation-dimension techniques [14, 15], the EEG state-space dimensionality using Taken's embedding theorem [16] was bounded between 3 and 10. In our intracranial EEG data, the embedding dimension (m) and the delay (τ) were chosen to be $m = 10$ and $\tau = 4$. The parameters were compatible with other studies [14, 15], performed on the same data. The following steps describe the procedure to track the spatio-temporal connectivity patterns in intracranial EEG data.

(1) The intracranial EEG attractors were reconstructed in the high dimensional state space. On nonoverlapping 10-second epochs, one set of pair-wise interdependence values among 24 channels are computed using the SOM-SI measure.
(2) The similarity indices, from every window, are translated into a symmetric similarity/affinity/proximity matrix. With the number of clusters (say n_1) specified apriori as discussed below, spectral clustering on the affinity matrix results in channels being labeled as one of the n_1 clusters.
(3) Steps 1 and 2 are repeated for all the successive windows, representing 10-second stationary segments. However, the overall ability of the channels to associate with each other over longer time duration needs to be quantified.

On T: 30-minute time segments (equal to 90, 10-second windows), pair-wise Hamming-distance based cluster-similarity matrix $\boldsymbol{P}$ is computed among all the channels. The matrix elements essentially index the probability of channels to group into the same cluster over a 30-minute time interval.

Spectral clustering or any other clustering algorithm on the cluster-similarity matrix $\boldsymbol{P}$ will result in final cluster memberships. The number of clusters is fixed to n_2 as specified below. For computing similarity indices in step 1, the epoch length of 10 seconds is chosen as a tradeoff between stationarity and sample-size requirements. Also note that the successive windows are 10 seconds apart (alternate 10-second windows) for reasons specific to computational feasibility.

We now describe step 2 in more details. The channel interdependencies obtained from SOM-SI represent the spatio-temporal correlation indices obtained by computing pair-wise similarity index among 24 channels. In spectral clustering jargon, the resulting matrix can be interpreted as an affinity matrix representing the pair-wise distances between 24 nodes. After spectral-clustering, we have a set of labeled clusters representing the membership of the channels [17]. Repeating this procedure on every 10-second window will yield a discrete-valued matrix κ_{spect} similar to (5).

Typically, the choice for the number of clusters n_1 in step 2 is conditioned on the significant eigenvalues. The dimensionality of the space will affect tremendously the computational complexity of the overall procedure. In our analysis, the sum of the first 3 eigenvalues typically ranged from 60% to 80% of the total variance, due to changes in seizure states. Considering this variability between epochs, and the fact that the number of clusters need to be the same for all epochs in order to be able to determine the overall grouping in channels (using cluster-similarity matrix $\boldsymbol{P}$), we fixed the number of clusters to $n_1 = 3$.

Experimental studies using nonlinear dynamics have shown [1, 2] that the quantitative descriptors of EEG exhibit seizure precursors in the form of interictal to preictal state transitions. The preictal transition time is not exactly known, however the literature [1, 2] suggests that it has a broad range of 5 minutes to 60 minutes before seizure. Therefore in step 5, as a tradeoff between state transition periods and time resolution, we choose a 30-minutes time window to characterize both the preictal and the postictal periods.

Patient P093

This patient had a history of complex partial seizures, localized in the mesial structures of the temporal lobe. Surgery revealed a lesion (mesial temporal sclerosis) in the right hippocampus (RTD electrodes) region. The set of 24 channels are listed below:

Channels 1 to 4: LTD3, LTD5, LTD7, LTD9,
Channels 5 to 8: RTD4, RTD6, RTD8, RTD10,
Channels 9 to 12: LST1, LST2, LST3, LST4,
Channels 13 to 16: RST1, RST2, RST3, RST4,
Channels 17 to 20: LOF1, LOF2, LOF3, LOF4,
Channels 21 to 24: ROF1, ROF2, ROF3, ROF4.

Before data analysis, a validation test was utilized to check whether application of different clustering algorithms on $\boldsymbol{P}$ would consistently result in same cluster memberships or not. For a given number of clusters n_2, it turned out that all the clustering algorithms including spectral clustering produced the same outputs. Therefore, we decided to choose the simple hierarchical clustering algorithm used in Matlab 6.5 owing to its graphical support.

Cluster-similarity matrices $\boldsymbol{P}$ indicating the probability that two channels share the same grouping in a 30-minute time segment are shown gray-scale coded in Figure 5. Preseizure analysis on 30-minute windows is shown for up to 3 hours. Similarly, the postseizure analysis is shown for the first 30 minutes. The ability of the left side channels to have a higher tendency to group together compared to the right hemisphere channels is quite noticeable from Figure 5. In addition, the orbitofrontal lobes seem like the only brain area to have a high probability of making a cross-hemisphere grouping. On the left hemisphere, the LST and the LTD channels are consistently seen to share the same clusters.

To confirm the observations from Figure 5, the hierarchical clustering algorithm was applied on each of those $\boldsymbol{P}$ matrices. Figure 6 graphically illustrates two instances of the clustering outputs through dendrograms. A dendrogram is strictly defined as a binary tree with a distinguished root that has all the data items at its leaves. Conventionally, all the leaves are shown at the same level of the drawing. The ordering of the leaves is arbitrary. The heights of the internal nodes are related to the metric information ($\boldsymbol{P}$ here) used to form the clustering. Using a threshold of 0.4 and the average-linkage technique to determine fusion levels, clustering was performed on a predefined number of clusters (n_2). For determining apriori the number of clusters n_2, several dendrograms were visually analyzed. There seemed to be at least 3 to 4 strong groupings among channels in most of the dendrograms. For consistency, therefore, we chose to fix the number of clusters n_2 to 3 for all the analyses.

Both dendrograms in Figure 6 clearly translate the spatial patterns observed in the corresponding $\boldsymbol{P}$ matrices of Figure 5. The top dendrogram in Figure 6 corresponds to the 2.5-to-3 hour time window (indicated by -5) in Figure 5. It is easy to see that the dendrogram considers the RTD and the RST as isolated clusters due to their weak between-cluster fusion level. Since the number of clusters n_2 is restricted to 3, all the remaining channels form a single large cluster. Similarly, the bottom dendrogram in Figure 6 corresponds to the P matrix indicated by -1 in Figure 5. In this case, the RST and the RTD channels group into one cluster; also well supported by a dark patch in Figure 5. This enables the LST/LTD channels and the LOF/ROF channels to group together as separate clusters.

The overall cluster configuration is listed in Table 3.

We summarize the spatial patterns at different time intervals of seizure 11 as follows.

(1) The LST and the LTD channels, in particular, exhibit a strong tendency to belong to the same group.

(2) The LOF and the ROF channels form a strong bilateral homologous connection, as seen from all the matrices in Figure 5.

(3) Relatively strong similarity can be seen between RTD and the RST channels.

(4) Common observation in all the matrices is the strong similarity between the left hemisphere channels as opposed to the right hemisphere channels. This is reflected in the ability of LOF channels to have a higher probability of sharing clusters with other left hemisphere channels, as seen in Figure 5.

(5) Interestingly, no temporal changes are seen in the spatial-patterns yet.

5.2. Statistical validation

The cluster configurations observed from analyzing 30-minute segments necessitates validation. Previously [9], we partially validated our model (up to the spectral clustering stage), using synthetically coupled multivariate time sequences (both nonlinear and linear). Simulations involving creation of dynamic graphs involve multidimensional time series that continuously change cluster memberships over time. Determining the average spatio-temporal groupings from a collection of multivariate time series is relatively easier to be demonstrated in linear coupling cases. However, nonlinear dynamic model constructions are extremely hard and mostly nontrivial. We therefore decided to pursue

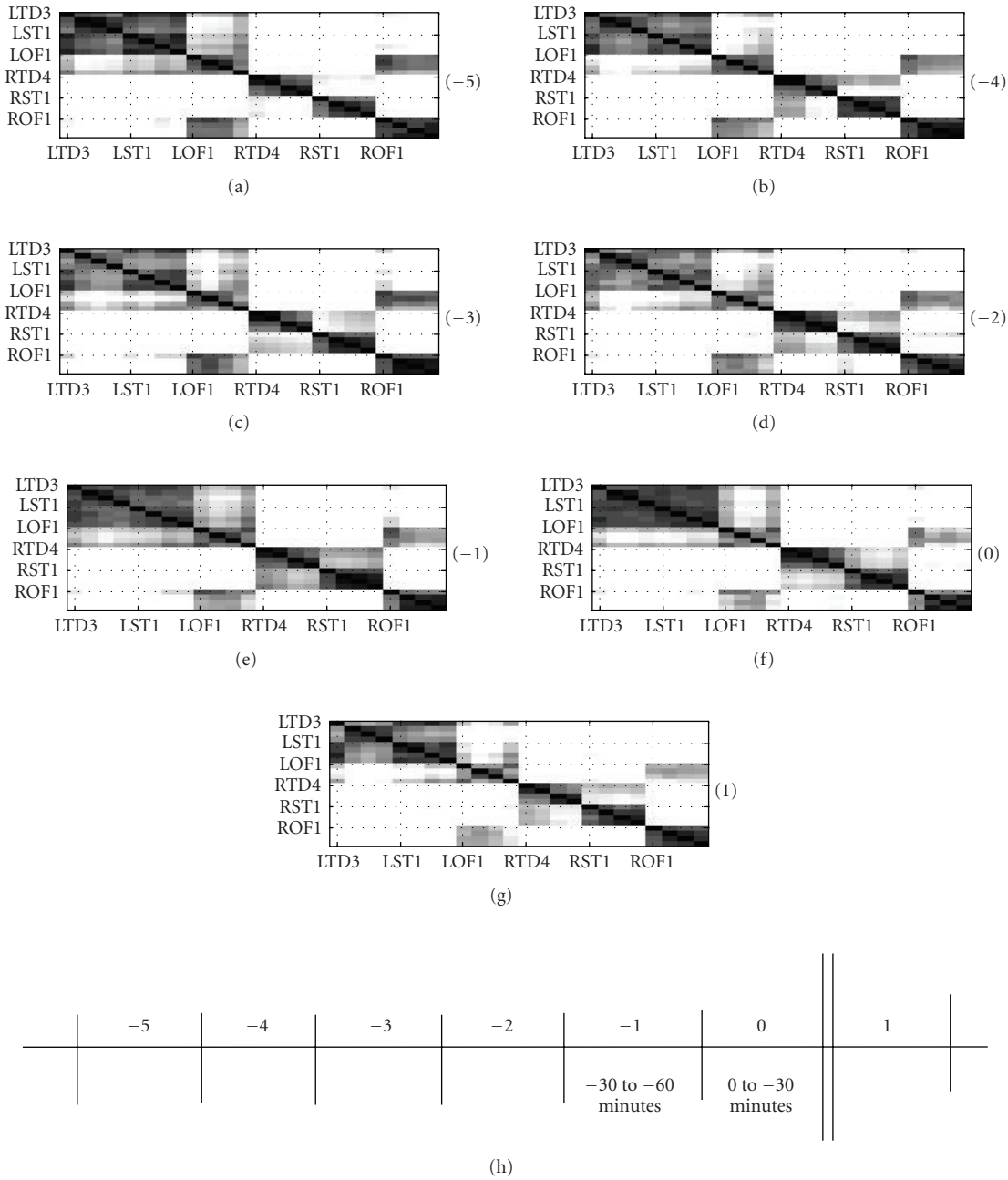


Figure 5: Seizure 11 of patient P093: Number in bracket indicates the 30-minute time interval when the cluster-similarity matrices were computed. The cluster-similarity matrices represent the probability that two channels share the same cluster label in a 30-minute time interval.

Table 3: Spatio-temporal groupings as obtained for seizure 11 of patient P093.

P093, Seizure 11	C_1	C_2	C_3
Preseizure, (2.5–3 hrs)	RTD	RST	LTD, LST, LOF, ROF
Preseizure, (2–2.5 hrs)	RTD, RST	LOF, ROF	LTD, LST
Preseizure, (1.5–2 hrs)	RTD, RST	LOF, ROF	LTD, LST
Preseizure, (1–1.5 hrs)	RTD, RST	LOF, ROF	LTD, LST
Preseizure, (30 mins–1 hr)	RTD, RST	LOF, ROF	LTD, LST
Preseizure, (0–30 mins)	RTD, RST	LOF, ROF	LTD, LST
Postseizure, (30 mins–1hr)	RTD, RST	LOF, ROF	LTD, LST

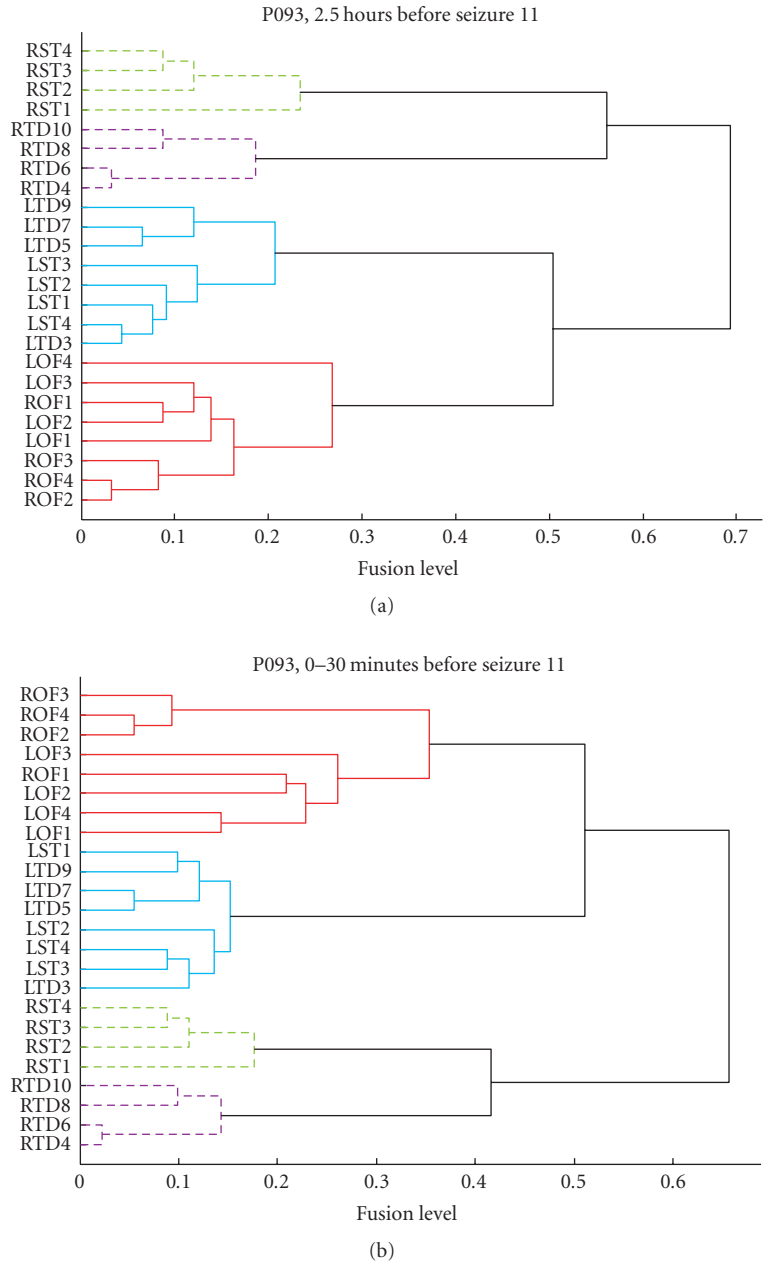


Figure 6: Dendrogram representation of the cluster results in Seizure 11, P093. TOP: Dendrogram corresponding to 2.5 hours before seizure. BOTTOM: Dendrogram corresponding to the 30-minute preseizure period.

a verification of the time-averaged cluster groupings on the intracranial EEG data, using the quasisurrogate analysis technique [18–20].

Recall that the cluster groupings obtained over 30-minute time segments involve two steps. First step consists of applying spectral clustering technique on the SOM-similarity indices (computed on 10-second intracranial EEG data segments). Then similar grouping patterns among channels are extracted by using hierarchical clustering approach on the cluster-similarity matrices $\boldsymbol{P}$. In order to validate this 2-step approach, we define our hypothesis as follows.

H_0: The average within-cluster channel interaction at each window (out of 91, 10-second windows) is not significantly different from the corresponding between-cluster channel interactions.

We propose to test this hypothesis on all the 3 (n_2) clusters separately, for every 10-second window within the 30-minute period. Within-cluster interaction is computed by averaging the pair-wise similarity indices for all the channels within a cluster. For between-cluster interaction, the pair-wise interactions among 3 channels picked randomly from each of the 3 clusters are computed. A between-cluster interaction statistic is formed by computing the average

Table 4: P093, Seizure 11: Over each 30-minute (91 samples total) window, number of times the within-cluster interaction is greater than between-cluster interaction, at 95% significance level.

P093, Sz 11	−5	−4	−3	−2	−1	0 (Sz)	1
C1	1	1	0.91	0.95	0.99	1	0.93
C2	0.82	0.89	0.96	0.91	0.89	0.85	0.98
C3	0.95	0.55	0.80	0.70	0.46	0.46	0.97

Table 5: Spatio-temporal groupings as obtained for seizures 4 and 5 of patient P093.

P093, Seizure 4 and 5	C_1	C_2	C_3
Preseizure 4, (30–60 mins)	RTD, RST	LOF, ROF	LTD, LST
Preseizure 4, (0–30 mins)	RTD, RST	LOF, ROF	LTD, LST
Postseizure 4, (0–30 mins)	RTD	LTD, LST, LOF, ROF	RST
Postseizure 4, (30 mins–1 hr)	RTD	LOF, ROF	LTD, LST, RST
Preseizure 5, (30 mins–1 hr)	RTD	LTD, LST, LOF, ROF	RST
Preseizure 5, (0–30 mins)	RTD	LTD, LST, LOF, ROF	RST
Postseizure 5, (30–1 hr)	RTD	LTD, LST, LOF, ROF	RST

interactions from random selection of 3 channels (one from each cluster) over a number of trials. We found that this statistic follows a quasinormal distribution, implying that the within-cluster interaction value can now be compared with the mean and the variance sample estimates of the between-cluster statistic. Mathematically, we construct the z-score as follows

$$Z_t^i = \frac{|\, C_{w_t}^i - \langle C_{b_t} \rangle \,|}{\sigma(C_b)} \quad t = 1, 2, \ldots, 90 \text{ and } i = 1, 2, 3, \qquad (7)$$

where $C_{w_t}^t$ is the within-cluster interaction at time "t", for cluster "i"; $\langle C_{b_t} \rangle$ is the mean and $\sigma(C_{bt})$ is the standard deviation of the between-cluster interaction at time "t"; Z_t^i reflects the z-score and is considered significant at the 95 percentile significance if $Z_t^i > 1.96$ (reject H_0). In Table 4, the bolded value in each cell represents the number of windows (out of 91) having significant z-score in the 30-minute period corresponding to Figure 5 (P093, Seizure 11). It is easy to observe that the null-hypothesis H_0 is rejected beyond doubt, validating the clustering results.

Seizures 4, 5, 6 and 7:

Spatio-temporal clustering analyses, similar to the one described on seizure 11 were performed on several other seizures, of the same patient P093. The cluster-similarity matrices $\boldsymbol{P}$ obtained from time intervals surrounding seizures 4 and 5 and 6 and 7 of patient P093 are shown in Figures 8 and 9, respectively. Channel groupings for the same are listed in tables 5 and 6, respectively. All the 4 seizures present very consistent groupings.

(1) Consistent to the observation in seizure 11, we observe the temporal depth and the subcortical regions of the left hemisphere are always grouped together.

(2) Once again, the association of ROF-LOF areas into the same cluster suggests a strong homologous connection between the orbitofrontal areas of the brain. This observation is also in agreement with those in seizure 11.

(3) The dendrograms once again presented 4 unambiguous clusters in the form of RST, RTD, LST/LTD, and LOF/ROF. The fusion levels, indicating the strength of connection between clusters, often turn out in favor of RTD and RST to be grouped separately. Owing to the fact that we have predefined the number of clusters to 3, the LST, LTD, LOF & ROF channels will consequently get grouped into one cluster.

(4) Once again, temporal changes are not very evident in the spatial patterns. However, observing Figures 8 and 9 and their corresponding dendrograms (not shown), the fusion levels and the topology of the connections change with time. These changes can only be quantified using statistical tests such as Mantel test statistics or the Double Permutation Statistics (DPS).

Patient P092

In this section, we present the summary results of the clustering analyses performed on patient P092 suffering from a lesion (mesial temporal sclerosis) in the medial temporal lobe structures of the right hemisphere. Channel configuration for the patient P092 is as follows:

Channels 1 to 4: LTD1, LTD3, LTD5, LTD7,
Channels 5 to 9: RTD2, RTD4, RTD6, RTD8, RTD12,
Channels 10 to 13: LST1, LST2, LST3, LST4,
Channels 14 to 17: RST1, RST2, RST3, RST4,
Channels 18 to 21: LOF1, LOF2, LOF3, LOF4,
Channels 22 to 24: ROF1, ROF2, ROF3.

Note that a separate 25×25-sized, 2-dimensional EEG-SOM grid was created to model the data dynamics of P092. Postspectral clustering analysis on 30-minute data segments led to some interesting observations.

Figure 10 shows the dendrograms created for seizure segments 2 hours prior to seizure 1 and 30 minutes preseizure, respectively. As before, the number of clusters (n_1) specified in the spectral-clustering step after SOM-SI block was fixed to 3. The fusion levels between most of the channel clusters

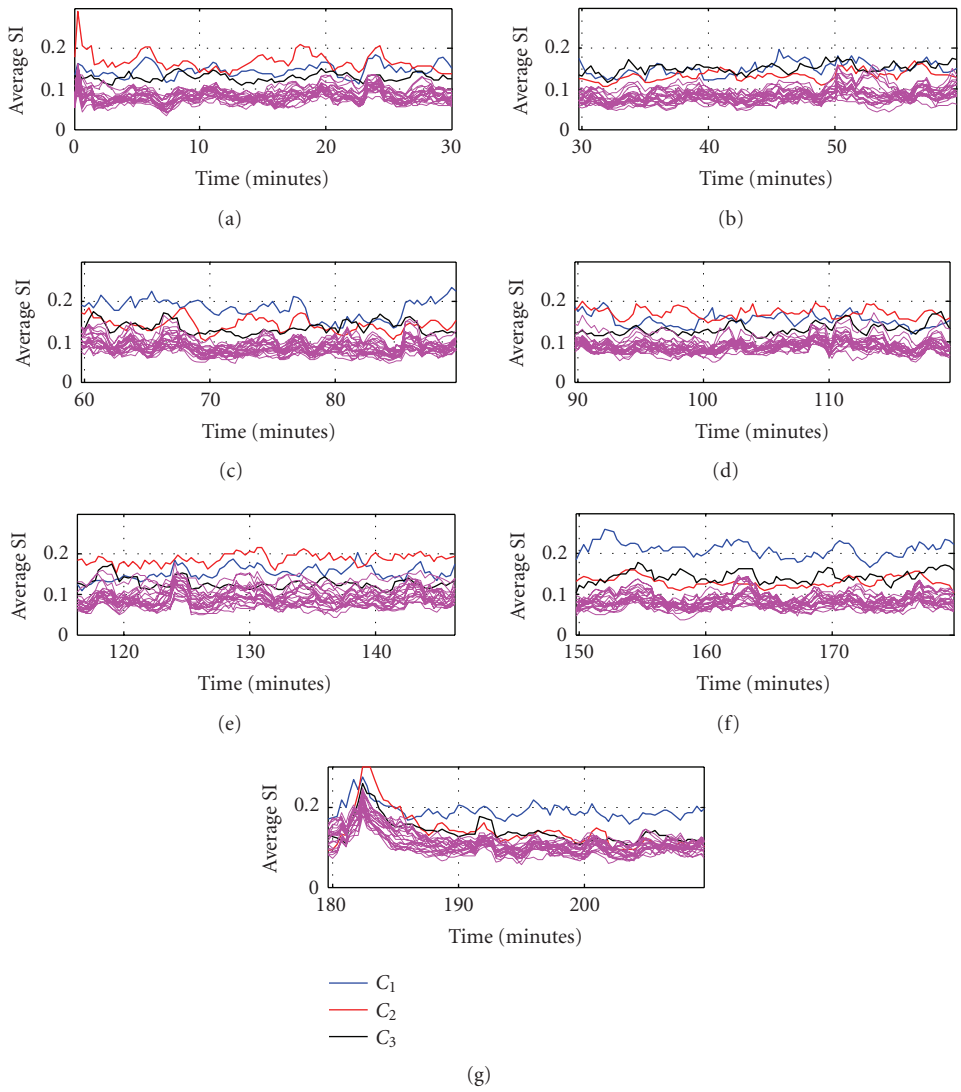


Figure 7: Statistical validation of the clustering results. In each panel, thick lines are used to represent the profiles of the three clusters in a 30-minute time interval. The thin lines are the surrogate profiles indicating between-cluster interactions. Cluster veracity can be visually verified by observing that amplitudes representing within-cluster interaction for cluster profiles are mostly higher that the amplitudes representing between-cluster interaction for surrogate profiles, at each time instance.

Table 6: Spatio-temporal groupings as obtained for seizure 6 and 7 of patient P093.

P093, Seizure 6 and 7	C_1	C_2	C_3
Postseizure 6, (0–30 mins)	RTD, RST	LTD, LST	LOF, ROF
Preseizure 7, (30 mins–1 hr)	RTD, RST	LTD, LST	LOF, ROF
Preseizure 7, (0–30 mins)	RTD	LTD, LST, LOF, ROF	RST
Postseizure 7, (0–30 mins)	RTD	LTD, LST, RST	LOF, ROF
Postseizure 7, (30 mins–1 hr)	RTD	LTD, LST, LOF, ROF	RST
Postseizure 7, (1 hr–1.5 hrs)	RTD	LTD, LST, LOF, ROF	RST

is greater than 0.4, indicating a lack of strong connectivity between regions.

For the second level of clustering, as before, let the number of clusters n_2 be fixed at 3. Cluster analysis on the 30 minutes segment 2 hours prior to seizure 1 (top dendrogram in Figure 10) results in the following groups of channels:

Cluster number 1: LTD and LST,
Cluster number 2: RTD and RST,
Cluster number 3: LOF and ROF.

Observe the cluster formed from LTD and LST channels, in the dendrogram. It is made up of two subclusters, a large and a small cluster. The small cluster consists of only

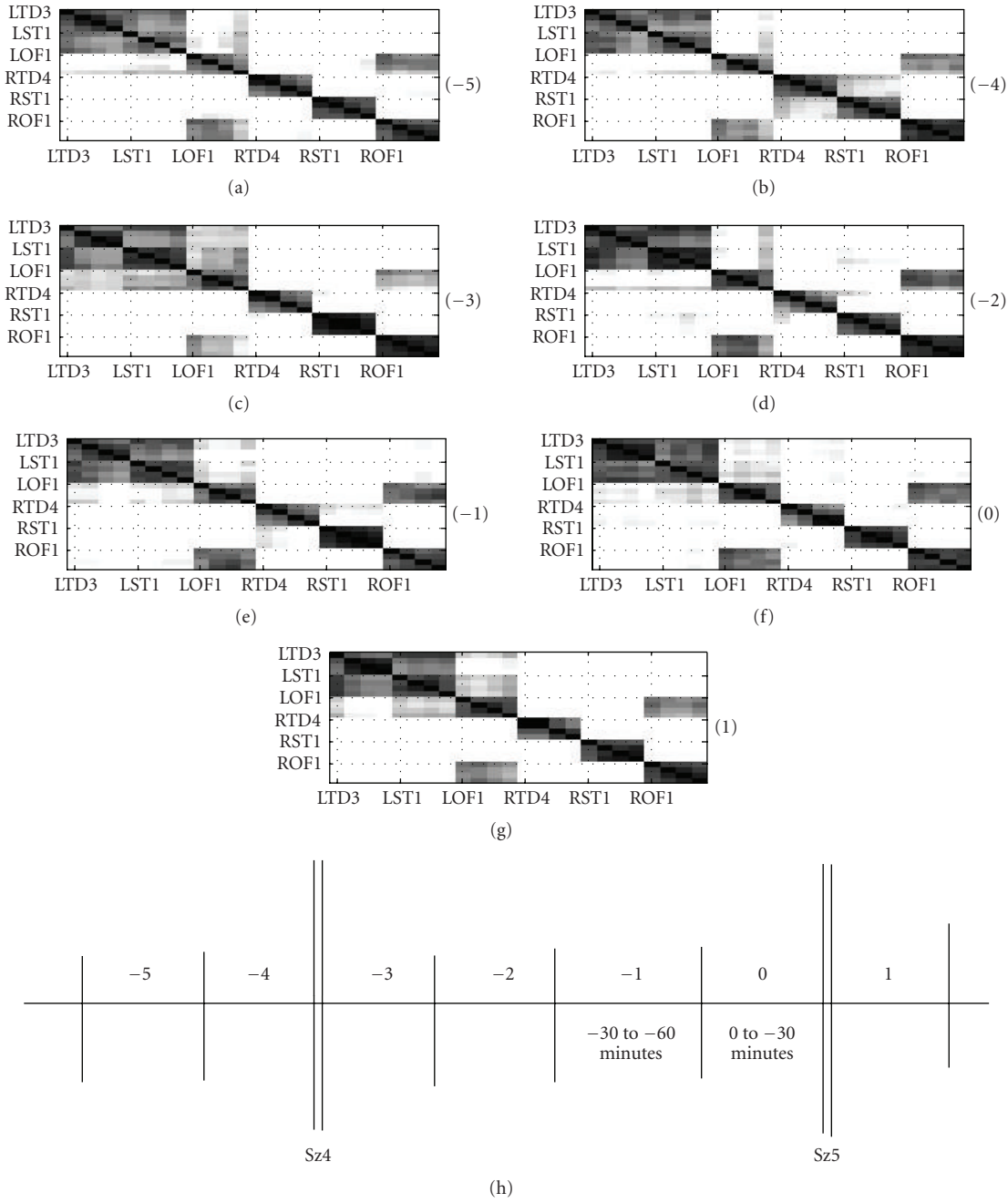


Figure 8: Seizures 4 and 5 of patient P093. Number in bracket indicates the 30-minute time interval when the cluster-similarity matrices were computed. The cluster-similarity matrices represent the probability that two channels share the same cluster label in a 30-minute time interval.

two channels, LTD (3 and 5) and fuses with the other subcluster at a very high fusion level (implying weak link). If n_2 was to be increased to 4, the clustering algorithm would classify this subcluster as an independent cluster. A detailed analysis on all seizures in P092 revealed a strong intrachannel correlation (or low fusion level) between channels LTD (3 and 5) and a weak interchannel correlation with the rest of the channels. Surrogate analysis also confirmed the imbalance by having very few rejections for the cluster consisting of LTD (3 and 5) channels. It is obvious that the average interaction (within-cluster interaction) of the largest cluster would be pulled down if there are subclusters that have a strong within-subcluster interaction, but a weak between-subcluster interaction. Consequently, the within-cluster interaction of the largest cluster can be expected to be as weak as or marginally better than the between-cluster interactions, leading to fewer rejections of the null hypothesis H_0.

This problem can possibly be overcome by increasing the number of clusters to 4 or more. However, for consistency,

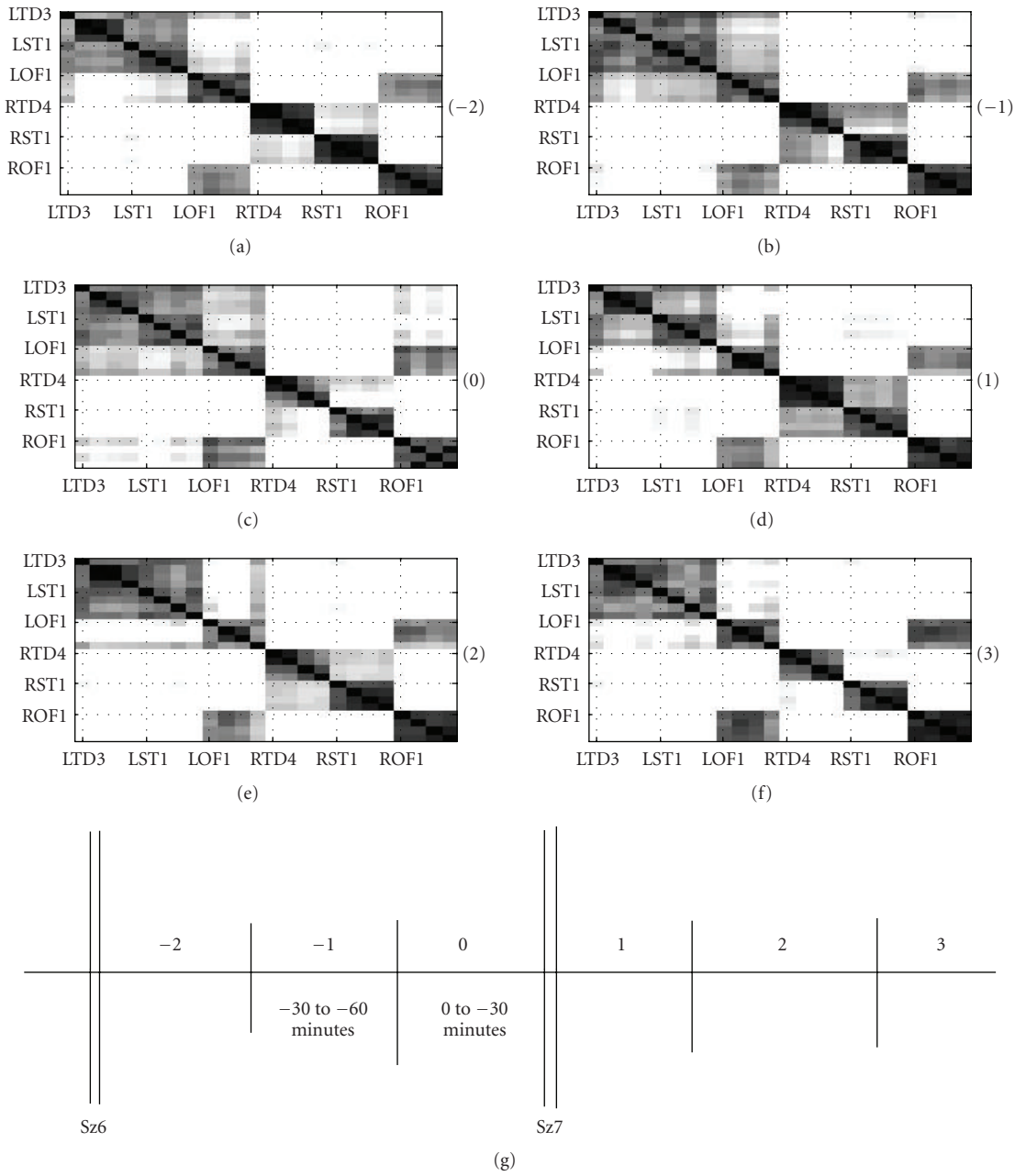


Figure 9: Seizures 6 and 7 of patient P093: Number in bracket indicates the 30-minute time interval when the cluster-similarity matrices were computed. The cluster-similarity matrices represent the probability that two channels share the same cluster label in a 30-minute time interval.

we let the number of clusters n_2 be fixed at 3 in the rest of the analyses.

Seizures 1, 3, and 4:

For illustration, the cluster-similarity matrices corresponding to seizure 1 is shown in Figure 11. Overall, the spatiotemporal clustering results for seizures 1, 3, and 4 are summarized in Tables 7 to 9.

From the cluster results of patient P092, we note the following:

(1) The nonfocal zone LTD has a strong coupling with the LST region. Correspondingly, strong affinity is observed between RTD and RST as well. These observations are consistent with the observations for P093. However, unlike in P093, we also see here that LTD connects and disconnects with several other channels, depending on the seizure state.

(2) As in P093, we observe an exclusively strong connection between ROF-LOF regions at all stages surrounding a seizure. There are few instances where the ROF breaks into a separate group. We do not have any explanation for this drift in ROF, at this point in time.

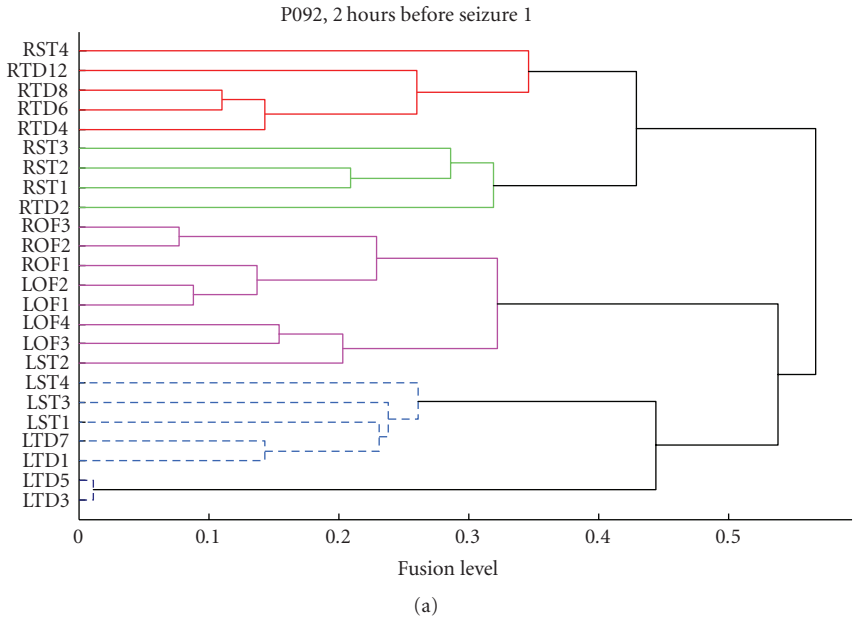


Figure 10: Dendrograms corresponding to P092, Seizure 1. Top: 2 hours before Seizure. Bottom: 30-minute preseizure.

Table 7: Spatio-temporal groupings as obtained for seizure 1 of Patient P092.

P092, Seizure 1	C_1	C_2	C_3
Preseizure, (1.5–2 hrs)	RTD, RST	LTD, LST (1, 3, 4)	LOF, ROF, LST (2)
Preseizure, (1–1.5 hrs)	RTD	LST, RST, LOF, ROF, LTD (1, 7)	LTD (3, 5)
Preseizure, (30 mins–1 hr)	RTD, RST	LTD, LST	LOF, ROF
Preseizure, (0–30 mins)	RTD, RST	LTD, LST	LOF, ROF
Postseizure, (0–30 mins)	RTD, RST	LTD, LST	LOF, ROF
Postseizure, (30–1 hr)	RTD	LTD, LST, LOF, RST	ROF

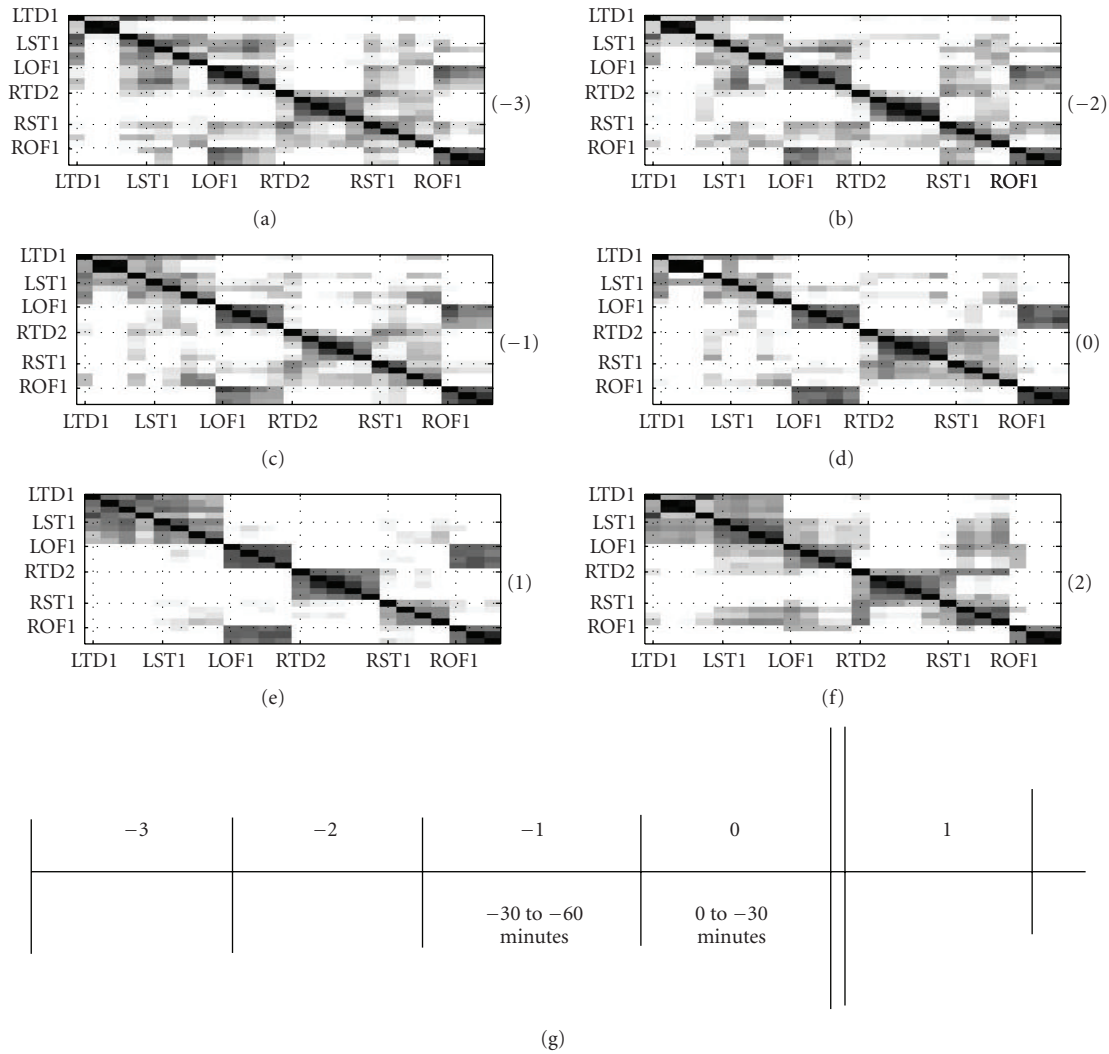


Figure 11: Seizure 1 of patient P092 Number in bracket indicates the 30-minute time interval when the cluster-similarity matrices were computed. The cluster-similarity matrices represent the probability that two channels share the same cluster label in a 30-minute time interval.

Table 8: Spatio-temporal groupings as obtained for seizure 3 of Patient P092.

P092, Seizure 3	C_1	C_2	C_3
Preseizure, (1.5–2 hrs)	RTD	LST, LTD, RST	LOF, ROF
Preseizure, (1–1.5 hrs)	RTD	LST, LTD, RST	LOF, ROF
Preseizure, (30 mins–1 hr)	RTD, RST	LST, LTD	LOF, ROF
Preseizure, (0–30 mins)	RTD, RST	LST, LTD	LOF, ROF
Postseizure, (0–30 mins)	RTD, RST	LST, LTD	LOF, ROF
Postseizure, (30–1 hr)	RTD, RST	LST, LTD	LOF, ROF

Table 9: Spatio-temporal groupings as obtained for seizure 4 of Patient P092.

P092, Seizure 4	C_1	C_2	C_3
Preseizure, (1.5–2 hrs)	RTD, RST	LST, LTD	LOF, ROF
Preseizure, (1–1.5 hrs)	RTD, RST	LST, LTD	LOF, ROF
Preseizure, (30 mins–1 hr)	RTD	LST, LTD, RST	LOF, ROF
Preseizure, (0–30 mins)	RTD, RST	LST, LTD	LOF, ROF
Postseizure, (0–30 mins)	RTD, RST	LST, LTD	LOF, ROF
Postseizure, (30–1 hr)	RTD, RST	LST, LTD	LOF, ROF

(3) Statistics from the surrogate analyses confirmed the veracity of the technique in most of the cases. As pointed out earlier, discrepancies occurred in a few instances for the clusters containing LTD (3, 5) channels.

Finally, we summarize the analysis on 2 patients and 8 complex partial seizures:

(1) Contrary to the accepted view that the seizure activity initiates in the focal zone followed by a gradual propagation to other regions, we observed that the spatial organization reflected by EEG activity exhibits either minimal or no progressive changes from the focal zone (RTD) to other zones (based on how it groups with other regions in the brain).

(2) Evidence show stronger ipsilateral connection between the LTD and LST zones compared to the connection strength between RTD-RST. Statistical analysis to check if a significant difference in intrahemisphere coupling strengths exists is needed.

(3) We also found evidence to show a strong cross-hemispheric activity by observing consistent groupings of the right and left orbitofrontal lobes at all seizure states.

(4) Patient P093 was seen to have qualitatively lesser spatio-temporal changes in its $\boldsymbol{P}$ matrices than P092 across the 30-minute analysis. It remains to be checked whether a significant change in the spatial organization before seizure is a pre-requisite to its initiation.

6. DISCUSSION

In this study, we applied the SOM-based similarity index measure to analyze the mutual interactions among critical areas of an epileptic brain. Based on the functional relationships, we analyzed long term structural connectivity's related to various seizure states by proposing a spatio-temporal clustering model. On analyzing 8 complex partial seizures from 2 patients suffering from temporal lobe epilepsy, we found that the orbitofrontal regions always exhibit a strong homologous connectivity while maintaining a low relationship with other regions. The left subtemporal and the lefttemporal depth regions (nonfocal hemisphere) were identified to have a strong ipsilateral connection, regardless of seizure states. Finally, we found that the epileptic focus, namely, the right hippocampus depth region, maintained a relatively strong connection with the right subtemporal region. Interestingly, the configuration of the groupings between different regions always remained the same, regardless of whether the patient was in an interictal, preictal, or postictal state although the inter-region connectivity strengths seemed to vary slightly across states.

So far, because of the data size, we were constrained to analyze only on 8 seizures from 2 patients. Future effort in this direction would be to apply the proposed approach on a larger set of seizures and more patients. In addition, since we analyzed only complex partial seizures, it would be worthwhile to check the cluster grouping in other types of seizures such as partial secondary generalized and subclinical seizures.

Recall from the results that certain channels were always grouped together regardless of the seizure states. This raises a question if this pattern is unique to an epileptic patient, and therefore be considered as a blueprint of seizures. While it is almost impossible to obtain intracranial EEG on normal subjects, one plausible way to answer this speculation would be to apply the proposed clustering approach on scalp EEG data from normal subjects and then analyze the differences in groupings with that of scalp EEG obtained from seizure patients.

One of our other main objectives in this study was to develop engineering tools to determine spatio-temporal groupings in a multivariate epileptic brain. We proposed a similarity-based clustering approach and used it to extract hidden structures from an epileptic brain. One of the obvious limitations with any clustering approach is determining the optimal number of clusters. Techniques to address cluster size have been researched, without much success. In eigenvector-based methods such as spectral clustering, cluster size can possibly be approximated to be equal to the number of eigenvectors corresponding to significant eigenvalues. In multiple datasets however, the optimal cluster size need not have to be the same across different datasets rendering cluster comparisons weak. In our approach, we analyzed a large number of data sets and empirically, fixed the cluster size to 3. This may not be an efficient or a systematic approach to tackle the problem. Theoretic efforts are needed to develop a mathematical criterion that allows us to determine a fixed cluster size, suitable to all groups of data. Besides, exploring tools better than clustering to unravel hidden patterns in multidimensional time sequences would be very beneficial.

REFERENCES

[1] L. D. Iasemidis, J. C. Principe, J. M. Czaplewski, R. L. Gilman, S. N. Roper, and J. C. Sackellares, "Spatiotemporal transition to epileptic seizures: a nonlinear dynamical analysis of scalp and intracranial EEG recordings," in *Spatiotemporal Models in Biological and Artificial Systems*, F. H. Lopes da Silva, J. C. Principe, and L. B. Almeida, Eds., pp. 81–89, IOS Press, Amsterdam, The Netherlands, 1997.

[2] L. D. Iasemidis, L. D. Olson, J. C. Sackellares, and R. S. Savit, "Time dependencies in the occurrences of epileptic seizures," *Epilepsy Research*, vol. 17, no. 1, pp. 81–94, 1994.

[3] K. J. Blinowska, R. Kuś, and M. Kamiński, "Granger causality and information flow in multivariate processes," *Physical Review E*, vol. 70, no. 5, Article ID 050902, 4 pages, 2004.

[4] J. Arnhold, P. Grassberger, K. Lehnertz, and C. E. Elger, "A robust method for detecting interdependences: application to intracranially recorded EEG," *Physica D*, vol. 134, no. 4, pp. 419–430, 1999.

[5] A. Hegde, D. Erdogmus, Y. N. Rao, J. C. Principe, and J. B. Gao, "SOM-based similarity index measure: quantifying interactions between multivariate structures," in *Proceedings of the 13th IEEE Workshop on Neural Networks for Signal Processing (NNSP '03)*, pp. 819–828, Toulouse, France, September 2003.

[6] A. Hegde, D. Erdogmus, and J. C. Principe, "Synchronization analysis of epileptic ECOG data using SOM-based SI measure," in *Proceedings of the 26th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC '04)*, vol. 2, pp. 952–955, San Francisco, Calif, USA, September 2004.

[7] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice-Hall, London, UK, 2nd edition, 1999.

[8] J. C. Principe, N. R. Euliano, and W. C. Lefebvre, *Neural and Adaptive Systems: Fundamentals through Simulations*, John Wiley & Sons, New York, NY, USA, 2000.

[9] A. Hegde, D. Erdogmus, D. S. Shiau, J. C. Principe, and C. J. Sackellares, "Quantifying spatio-temporal dependencies in epileptic ECOG," *Signal Processing*, vol. 85, no. 11, pp. 2082–2100, 2005, special issue on Neuronal Coordination in the Brain: A Signal Processing Perspective.

[10] J. A. García, J. Fdez-Valdivia, F. J. Cortijo, and R. Molina, "A dynamic approach for clustering data," *Signal Processing*, vol. 44, no. 2, pp. 181–196, 1995.

[11] E. Keogh, J. Lin, and W. Truppel, "Clustering of time series subsequences is meaningless: implications for previous and future research," in *Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM '03)*, pp. 115–122, Melbourne, Fla, USA, November 2003.

[12] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: analysis and an algorithm," in *Advances in Neural Information Processing Systems 14*, pp. 849–856, The MIT Press, London, UK, 2002.

[13] J. Malik, S. Belongie, T. Leung, and J. Shi, "Contour and texture analysis for image segmentation," *International Journal of Computer Vision*, vol. 43, no. 1, pp. 7–27, 2001.

[14] L. D. Iasemidis, K. E. Pappas, J. C. Principe, and J. C. Sackellares, "Spatiotemporal dynamics of human epileptic seizures," in *Proceedings of the 3rd Experimental Chaos Conference*, R. G. Harrison, L. Weiping, W. Ditto, L. Pecora, M. Spano, and S. Vohra, Eds., pp. 26–30, World Scientific, Singapore, August 1996.

[15] L. D. Iasemidis, P. Pardalos, J. C. Sackellares, and D. S. Shiau, "Quadratic binary programming and dynamical system approach to determine the predictability of epileptic seizures," *Journal of Combinatorial Optimization*, vol. 5, no. 1, pp. 9–26, 2001.

[16] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence*, D. A. Rand and L.-S. Young, Eds., vol. 898 of *Lecture Notes in Mathematics*, pp. 366–381, Springer, Berlin, Germany, 1981.

[17] A. Hegde, D. Erdogmus, and J. C. Principe, "Spatio-temporal clustering of epileptic ECOG," in *Proceedings of the 27th Annual International Conference of the Engineering in Medicine and Biology Society (EMBS '05)*, pp. 4199–4202, Shanghai, China, September 2005.

[18] D. Prichard and J. Theiler, "Generating surrogate data for time series with several simultaneously measured variables," *Physical Review Letters*, vol. 73, no. 7, pp. 951–954, 1994.

[19] T. Schreiber, "Measuring information transfer," *Physical Review Letters*, vol. 85, no. 2, pp. 461–464, 2000.

[20] T. Schreiber and A. Schmitz, "Improved surrogate data for nonlinearity tests," *Physical Review Letters*, vol. 77, no. 4, pp. 635–638, 1996.

Hindawi Publishing Corporation
Computational Intelligence and Neuroscience
Volume 2007, Article ID 80510, 13 pages
doi:10.1155/2007/80510

Research Article

Automatic Seizure Detection Based on Time-Frequency Analysis and Artificial Neural Networks

A. T. Tzallas,[1,2] M. G. Tsipouras,[2] and D. I. Fotiadis[2,3]

[1] *Department of Medical Physics, Medical School, University of Ioannina, GR 451 10 Ioannina, Greece*
[2] *Unit of Medical Technology and Intelligent Information Systems, Department of Computer Science, University of Ioannina, GR 451 10 Ioannina, Greece*
[3] *Biomedical Research Institute, Foundation for Research and Technology-Hellas (BRI-FORTH), University of Ioannina, GR 451 10 Ioannina, Greece*

Correspondence should be addressed to D. I. Fotiadis, fotiadis@cs.uoi.gr

Received 31 December 2006; Revised 16 July 2007; Accepted 7 October 2007

Recommended by Saied Sanei

The recording of seizures is of primary interest in the evaluation of epileptic patients. Seizure is the phenomenon of rhythmicity discharge from either a local area or the whole brain and the individual behavior usually lasts from seconds to minutes. Since seizures, in general, occur infrequently and unpredictably, automatic detection of seizures during long-term electroencephalograph (EEG) recordings is highly recommended. As EEG signals are nonstationary, the conventional methods of frequency analysis are not successful for diagnostic purposes. This paper presents a method of analysis of EEG signals, which is based on time-frequency analysis. Initially, selected segments of the EEG signals are analyzed using time-frequency methods and several features are extracted for each segment, representing the energy distribution in the time-frequency plane. Then, those features are used as an input in an artificial neural network (ANN), which provides the final classification of the EEG segments concerning the existence of seizures or not. We used a publicly available dataset in order to evaluate our method and the evaluation results are very promising indicating overall accuracy from 97.72% to 100%.

1. INTRODUCTION

Epilepsy is one of the most common neurological disorders with a prevalence of 0.6–0.8% of the world's population. Two-thirds of the patients achieve sufficient seizure control from anticonvulsive medication, and another 8–10% could benefit from resective surgery. For the remaining 25% of patients, no sufficient treatment is currently available [1]. The epilepsy is characterized by a sudden and recurrent malfunction of the brain, which is termed "seizure." Epileptic seizures reflect the clinical signs of an excessive and hypersynchronous activity of neurons in the brain. Depending on the extent of the involvement of other brain areas during the course of the seizure, epilepsies can be divided into two main classes. Generalized seizures involve almost the entire brain, while focal (or partial) seizures originate from a circumscribed region of the brain (epileptic focus) and remain restricted to this region. Epileptic seizures may be accompanied by impairment or loss of consciousness: psychic, autonomic or sensory symptoms, or motor phenomena [2, 3].

Traditionally, suspected seizures are evaluated using a routine electroencephalogram (EEG), which is typically a 20-minute recording of the patient's brain waves. Because a routine EEG is of short duration, it is unlikely that actual events are recorded. Routine EEGs may record interictal hallmarks of epilepsy, including spikes, sharp waves, or spike-and-wave complexes. However, diagnostic difficulties arise when a person has a suspected seizure, or a neurological event of unclear etiology, not obvious in the routine EEG. The current gold standard is the continuous EEG recording along with video monitoring of the patient, which usually requires inpatient admission. This is a costly endeavour, which is not always available. The patient is away from his environment and routine, which may be associated with factors that provoke the patient's events [4]. The introduction of portable recording systems (ambulatory EEG), however, has allowed

out-patient EEG recording to become more common. This has the advantage that patients are monitored in their normal environment without the reduction in seizure frequency usually occurring during in-patient sessions [4, 5].

Clinical neurophysiologists can then periodically review the EEG recordings and analyze the seizures that may have occurred during the monitoring session. However, reviewing a continuous EEG recording lasting several days can be a time-consuming process. In practice, the patient can indicate that a seizure occurs through the use of an alarm button, so that only the recording sections around the use of the button need to be analyzed. Unfortunately, in many cases, patients are not aware of the occurrence of their own seizures. An automated seizure detection system can thus be of great interest in identifying EEG sections that need to be reviewed. The main difficulty with it lies in the wide variety of EEG patterns that can characterize a seizure, such as "low-amplitude desynchronization, polyspike activity, rhythmic waves for a wide variety of frequencies and amplitudes, and spikes and waves" [6]. In extracranial recordings, EMG, movement, and eye blink artefacts often obscure seizures. Thus, from the pattern recognition point of view, the problem is extremely complex.

Research in automated seizure detection began in the 1970s and various algorithms addressing this problem [5–7] have been presented. Methods for automatic detection of seizures may rely on the identification of various patterns such as an increase in amplitude [8], sustained rhythmic activity [9, 10], or EEG flattening [11]. Several algorithms have been developed based on spectral [12–18] or wavelet features [19–23], amplitude relative to background activity [12, 24] and spatial context [24–27]. Chaotic features [28–31] such as correlation dimension [32, 33], Lyapunov exponents [34], and entropy [35] have also been proposed to characterize the EEG signal. These features can then be used to classify the EEG signal using statistical methods [28–30], nearest neighbour classifiers [36], decision trees [16], ANNs [21, 34], support vector machines (SVMs) [18, 37], or adaptive neurofuzzy inference systems [23, 35] in order to identify the occurrence of seizures. It is crucial for seizure detection systems to result in high sensitivity, even if this results in a large number of false detections. Such systems can then be used to reduce considerably the amount of data that need to be reviewed; neurophysiologists can then easily discard false detections.

In addition, to seizure detection systems, warning systems have also become increasingly valuable since detection of seizures at an early stage can warn the patient that a seizure is occurring. Also, they alert medical staff, and allow them to perform behavioral testing to further assess which specific functions may be impaired as a result of a seizure and help them in localizing the source of the seizure activity. Techniques used to forecast seizures include time-domain analysis [38], frequency-based methods [39], nonlinear dynamics and chaos [31, 40], methods of delays [41], and intelligent systems [42]. Advances in seizure prediction promise to give rise to implantable devices able to warn of impending seizures and to trigger therapy to prevent clinical epileptic attacks [2]. Treatments such as electrical stimulation of focal drug infusion could be given on demand and might eliminate side effects in some patients taking antiepileptic drugs.

Consequently, epileptic seizures give rise to changes in certain frequencies bands. Recent works have focused on the analysis of the δ (0.4–4 Hz), θ (4–8 Hz), α (8–12 Hz), β (12–30 Hz) rhythms, and their relation to epilepsy. An epileptic signal is nonstationary, having time-varying frequency components. Time-frequency (TF) representations combine both time and frequency information into a single representation and have proven to be powerful tools for the analysis of nonstationary signals [43], and have been used for neonatal seizure detection [44, 45].

In this work, we use TF analysis in order to extract several features from EEG segments, and subsequently use these features to classify the segments concerning epileptic seizures. The method is divided into three stages. Initially, TF analysis is performed for each EEG segment and its spectrum is acquired. Then, several features are extracted from it, measuring the fractional energy on specific TF windows. For this purpose, several partitions on the time axis and the frequency axis are tested. Finally, these features are used as inputs in an ANN, which provides the final classification according to the specified number of categories. A dataset of 500 EEG segments is used, while the method is evaluated for four different classification problems, each of them addressing a different interpretation of the medical problem and thus different selection of EEGs from the whole EEG segment dataset is required for each classification problem. TF analysis and feature extraction, reflecting the energy distribution over the TF plane, have been employed only for neonatal epileptic seizure detection and have not been previously applied in general epileptic seizure detection. In addition, no work addresses all four classification problems, which are directly related to the diagnosis provided by an expert. The obtained results indicate high accuracy compared to other existing approaches.

The rest of the paper is structured as follows. In Section 2, the dataset used in our work along with the employed methodology is described in detail. Then, the evaluation procedure and the obtained results are presented (Section 3), followed by an extensive discussion regarding them (Section 4). Finally, some concluding remarks are included in Section 5.

2. MATERIALS AND METHODS

The flowchart of the proposed method is shown in Figure 1. Below the dataset and its partitions used are briefly discussed and the three stages (time-frequency analysis, feature extraction, and classification) of the method are explained in detail.

2.1. Dataset

An EEG dataset, which is available online [46] and includes recordings for both healthy and epileptic subjects, is used. The dataset includes five subsets (denoted as Z, O, N, F, and S) each containing 100 single-channel EEG segments, each one having 23.6-second duration. The subsets Z and O have been acquired using surface EEG recordings of five healthy volunteers with eyes open and closed, respectively. Signals in

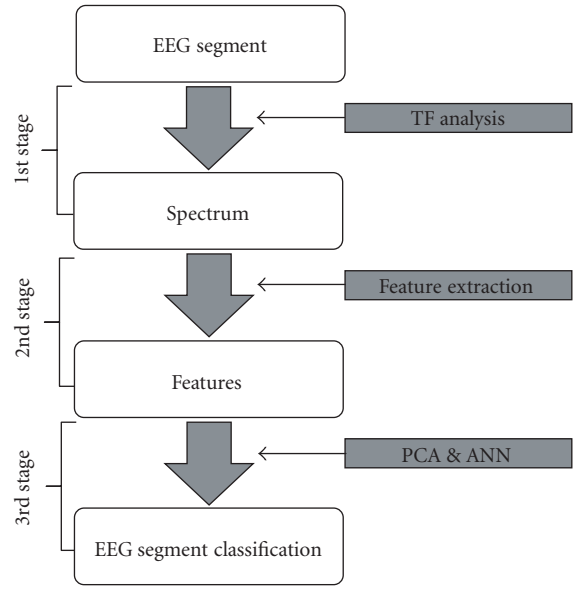


Figure 1: The flowchart of the proposed method.

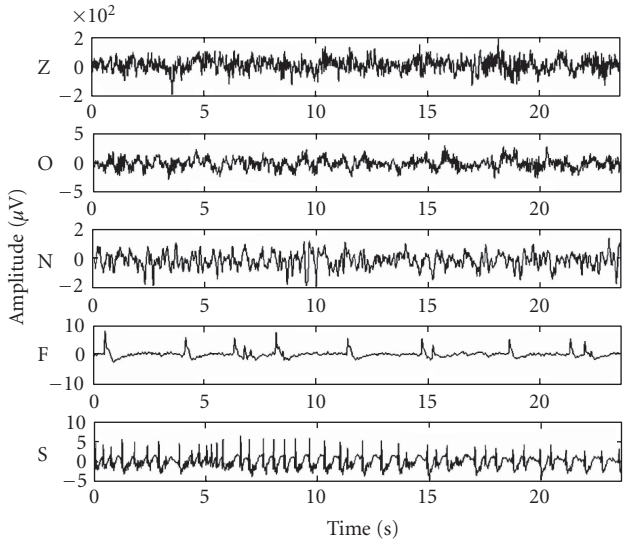


Figure 2: Exemplary EEG segments from each of the five subsets (Z, O, N, F, and S). From top to bottom: subset Z to subset S. The amplitudes of surface EEG recordings are typically in the order of some μV. For intracranial EEG recordings, the amplitudes range around 100 μV. For seizure activity, these voltages can exceed 1000 μV.

two sets have been measured in seizure-free intervals from five patients in the epileptogenic zone (set F) and from the hippocampal formation of the opposite hemisphere of the brain (set N). Finally, subset S contains seizure activity, selected from all recording sites exhibiting ictal activity. Subsets Z and O have been recorded extracranially, using standard electrode positioning (according to the international 10–20 system [47]), whereas subsets N, F, and S have been recorded intracranially. More specifically, depth electrodes are implanted symmetrically into the hippocampal formation. EEG segments of subsets N and F were taken from all contacts of the relevant depth electrode [46]. In addition, strip electrodes are implanted onto the lateral and basal regions (middle and bottom) of the neocortex. EEG segments of the subsets S were taken from contacts of all electrodes (depth and strip). All EEG signals were recorded with the same 128-channel amplifier system, using an average common reference. The data were digitized at 173.61 samples per second using 12 bit resolution and they have the spectral bandwidth of the acquisition system, which varies from 0.5 Hz to 85 Hz. Typical EEG segments (one from each category of the dataset) are shown in Figure 2.

In our analysis, we use the above-described dataset to create four different classification problems and then we tested our method with all of them.

(1) In the first, all the EEG segments from the dataset were used and they were classified into three different classes: Z and O types of EEG segments were combined to a single class, N and F types were also combined to a single class, and type S was the third class. This set is the one closest to real medical applications including three categories; normal (i.e., types Z and O), seizure-free (i.e., types N and F) and seizure (i.e., type S).

(2) In the second, again all the EEG segments from the dataset were used and they were classified into two different classes: Z, O, N, and F types are included in the first class and type S in the second class. This is also close to real medical applications, being slightly simpler than the previous, classifying the EEG segments into nonseizures and seizures.

(3) The third has similar classes with the first, that is, normal, seizure-free and seizure, but not all the EEG segments from the dataset were employed. The normal class includes only the Z-type EEG segments, the seizure-free class the F-type EEG segments, and the seizure class the S-type.

(4) The fourth has similar classes with the second, that is, normal and seizure, but again not all the EEG segments from the dataset were employed. The normal class includes only the Z-type EEG segments while the seizure class includes the S-type.

The above classification problems are shown in detail in Table 1.

2.2. Time-frequency analysis

In the proposed method, the smoothed pseudo-Wigner-Ville distribution (SPWVD) [48, 49] is applied to each EEG segment, defined as

$$\mathrm{SPWVD}_x(t,\omega) = \int_{-\infty}^{+\infty} h(\tau)\left(\int_{-\infty}^{+\infty} g(s-t)x\left(s+\frac{\tau}{2}\right)x^*\left(s-\frac{\tau}{2}\right)ds\right)e^{-j2\pi\omega\tau}d\tau, \quad (1)$$

where $x(\cdot)$ is the signal, t is the time, ω is the frequency, and $g(\cdot)$ and $h(\cdot)$ are time and frequency smoothing window functions, respectively. SPWVD can substantially suppress

Table 1: The classes and the corresponding number of EEG segments of the four classification problems.

Classification problem	Classes	Number of EEG segments
1	Normal (Z, O)	200
	Seizure-free (N, F)	200
	Seizure (S)	100
	Total	500
2	Nonseizure (Z, O, N, F)	400
	Seizure (S)	100
	Total	500
3	Normal (Z)	100
	Seizure-free (N)	100
	Seizure (S)	100
	Total	300
4	Normal (Z)	100
	Seizure (S)	100
	Total	200

Table 2: The frequency ranges (Hz) of four frequency subbands (4, 5, 7, and 13).

	Frequency subbands			
	4	5	7	13
Frequency ranges (Hz)	0–4	0–2.5	0–2	0–2
	4–8	2.5–5.5	2–4	2–4
	8–12	5.5–10.5	4–6.5	4–6
	12–40	10.5–21.5	6.5–9	6–8
	—	21.5–43.5	9–12	8–10
	—	—	12–25	10–12
	—	—	25–40	12–16
	—	—	—	16–20
	—	—	—	20–24
	—	—	—	24–28
	—	—	—	28–32
	—	—	—	32–36
	—	—	—	36–40

the cross terms, which is a major limitation of the time-frequency analysis. The time smoothing window was selected to be a Hamming 64-point length window, which was the same for all tests performed for evaluation. The length of the frequency smoothing window is not always the same; we have selected several different frequency resolutions (64, 128, 256, and 512 points length window), and we tested the method for all of them. Time-frequency (TF) analysis is used to calculate the spectrum of the signal. Figure 3 shows the spectrum of five EEG segments, one of each of the original dataset categories (Z, O, N, F, and S), using a 512-point length window.

2.3. Feature extraction

The spectrum of the signals, computed using TF analysis, is used to extract several features. To do that, a grid is used,

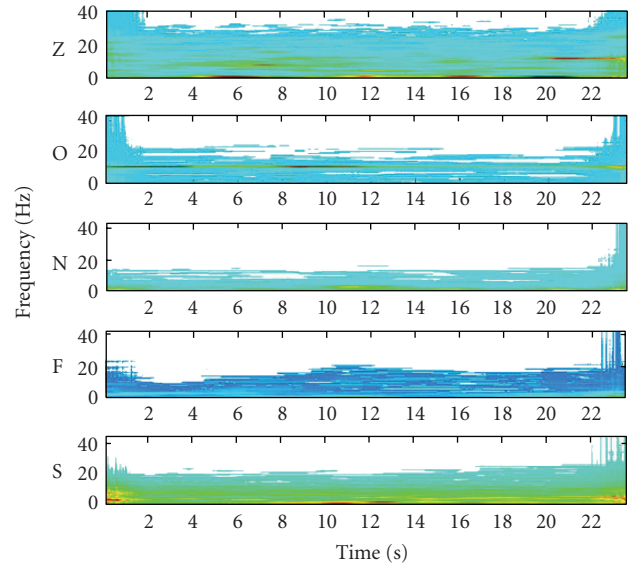


Figure 3: The obtained spectrum for five EEG segments, one for each of the original dataset categories (Z, O, N, F, and S).

based on a time and a frequency partition. In the time domain, two different partitions were used, having three and five equal-sized windows, respectively, while in the frequency domain, four different partitions were used, which divide the frequency domain in 4, 5, 7, and 13 subbands. These subbands, which are not always equal, are shown in Table 2 and they are created using medical knowledge about the EEG and the features that are expected to be found in certain frequency bands for the specific types of EEG segments included in the original dataset. All the combinations between these time and frequency partitions are used, in order to extract several sets of features. The result of the application of TF analysis in an EEG segment for different combinations of time windows and frequency subbands is shown in Figure 4.

Each feature, $f(i, j)$, is calculated as

$$f(i,j) = \int_{t_i} \int_{\omega_j} \mathrm{SPWVD}_x(t,\omega) d\omega\, dt, \tag{2}$$

where t_i is the ith time window and ω_j is the jth frequency band. Each feature represents the fractional energy of the signal in a specific frequency band and time window; thus the total feature set depicts the distribution of the signal's energy over the TF plane. Therefore, it is expected that each feature set carries sufficient information related to the nonstationary properties of the signal and thus, it can be useful for the classification process. The feature set initially is represented as an $N \times M$ matrix, where N is the number of time windows and M is the number of frequency subbands, and then it is reshaped into an $N \cdot M$ size vector. The length of the feature vector is not the same in all cases and it depends only on the time and frequency partitions. In all cases, an additional feature is used, which is the total energy of the signal. Thus, in each case the total number of features is $N \cdot M + 1$.

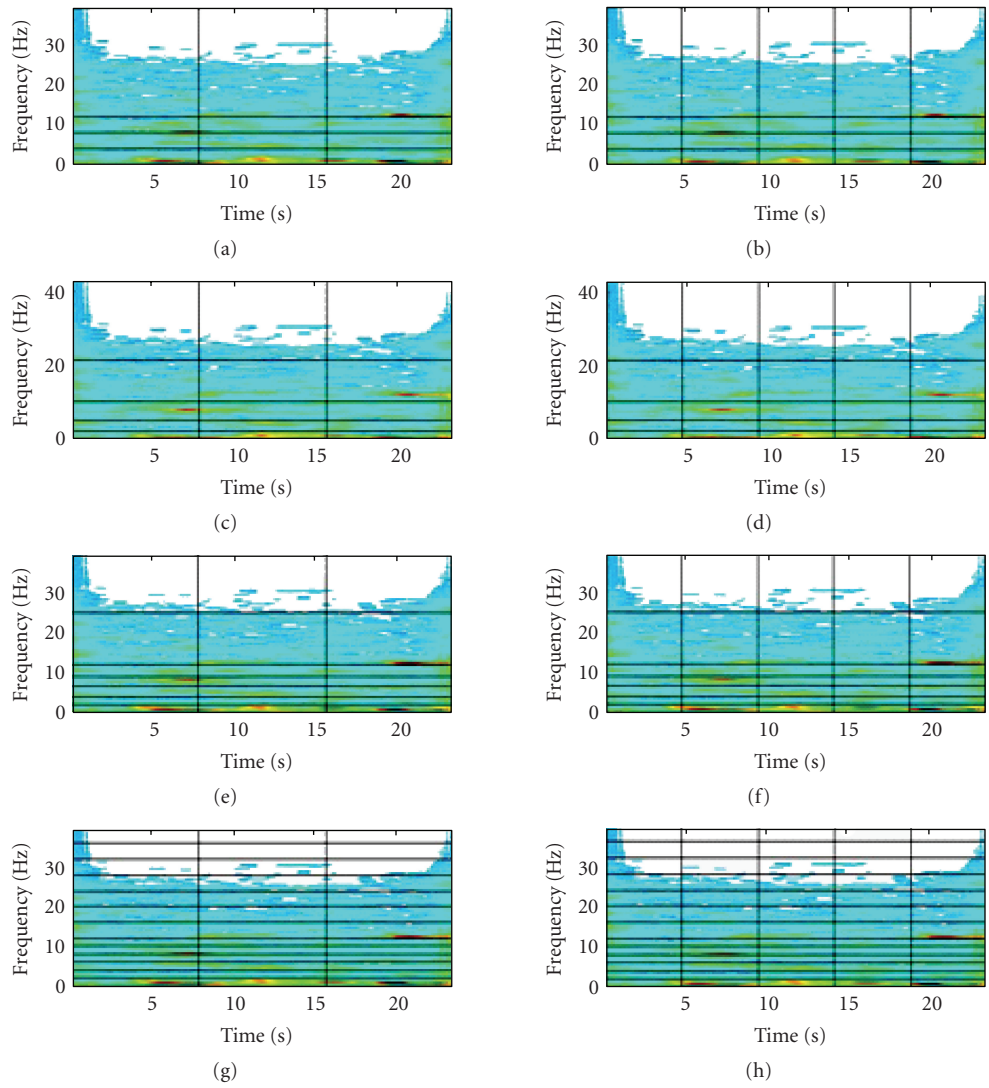


FIGURE 4: The spectrums obtained for various combinations of time and frequency partitions: (a) 3 time windows and 4 frequency subbands, (b) 5 time windows and 4 frequency subbands, (c) 3 time windows and 5 frequency subbands, (d) 5 time windows and 5 frequency subbands, (e) 3 time windows and 7 frequency subbands, (f) 5 time windows and 7 frequency subbands, (g) 3 time windows and 13 frequency subbands, and (h) 5 time windows and 13 frequency subbands.

2.4. Classification

The calculated features are fed into a feed-forward artificial neural network (ANN). To reduce the dimensionality of the input patterns, principal component analysis (PCA) is employed with the threshold set to 1%. The architecture of the neural network is different in each classification problem: N inputs (N is the number of features resulted from the PCA), one hidden layer with 4∗N neurons, and M outputs (M is the number of the classes), each of them being a real number in the interval [0, 1]. The units in the hidden layer are sigmoid units with hyperbolic tangent as activation function, while the outputs are linear. Half of the patterns of the dataset were randomly selected to be used for training, while the rest were used for testing. The network is trained using a standard backpropagation algorithm [50]. Ten different training-test sets were created for each classification problem and thus ten different neural networks were optimized. The final result is obtained as the average of their results.

3. RESULTS

The four classification problems, described above, are used to evaluate the proposed method. For each of them, all combinations between frequency resolutions (64, 128, 256, or 512), time windows (3 or 5), and frequency bands (4, 5, 7, or 13) were tested; totally 32 different combinations for each classification problem. For each problem, half of the EEG segments, randomly selected, were used for the training of the neural network, while the other half for testing.

The size of the confusion matrix depends on the classification problem: 3×3 for problems (1) and (3), 2×2 for

Table 3: Results for the first classification problem, in terms of sensitivity (Sens), specificity (Spec), and selectivity (Sel) in % values. Those are given for all TF resolutions (64, 128, 256, and 512), time windows (3 and 5), and frequency subbands (4, 5, 7, and 13).

Frequency subbands				4			5			7			13		
Classes				ZO	NF	S	ZO	NF	S	ZO	NF	S	ZO	NF	S
Frequency resolution 64	Time windows	3	Sens	98.90	94.20	97.00	97.90	96.80	95.60	95.80	93.20	93.80	96.10	97.80	90.80
			Spec	98.00	98.53	98.40	97.53	97.80	99.75	95.67	95.80	99.35	98.27	95.53	99.30
			Sel	97.06	97.72	93.81	96.36	96.70	98.96	93.65	93.67	97.30	97.37	93.59	97.01
		5	Sens	95.40	91.90	90.80	95.70	95.50	88.80	96.20	92.90	92.20	93.30	93.30	88.60
			Spec	95.53	95.07	98.40	96.20	94.40	99.85	96.27	96.27	98.20	95.33	93.87	98.55
			Sel	93.44	92.55	93.42	94.38	91.92	99.33	94.50	94.31	92.76	93.02	91.02	93.86
Frequency resolution 128	Time windows	3	Sens	99.20	95.50	97.40	97.90	95.20	91.20	99.60	96.90	94.60	96.80	93.20	87.40
			Spec	97.60	98.87	99.35	96.80	96.80	99.15	98.00	98.13	99.80	95.47	95.47	98.65
			Sel	96.50	98.25	97.40	95.33	95.20	96.41	97.08	97.19	99.16	93.44	93.20	94.18
		5	Sens	95.90	92.70	97.40	96.80	93.10	92.00	96.30	93.60	89.80	96.20	91.60	93.00
			Spec	96.73	96.47	98.75	95.67	96.40	98.90	95.33	96.07	98.85	94.93	97.60	97.75
			Sel	95.14	94.59	95.12	93.71	94.52	95.44	93.22	94.07	95.13	92.68	96.22	91.18
Frequency resolution 256	Time windows	3	Sens	96.60	95.70	98.00	98.20	90.80	87.20	98.00	96.00	96.20	96.50	98.00	93.00
			Spec	98.27	97.20	99.05	93.53	95.87	99.25	97.20	97.93	99.70	97.60	97.67	99.05
			Sel	97.38	95.80	96.27	91.01	93.61	96.67	95.89	96.87	98.77	96.40	96.55	96.07
		5	Sens	95.20	93.50	93.80	94.90	92.40	89.80	94.00	91.90	92.80	96.80	92.00	85.00
			Spec	96.80	95.40	98.65	95.27	94.13	99.05	95.53	94.73	98.45	92.47	97.33	98.30
			Sel	95.20	93.13	94.56	93.04	91.30	95.94	93.35	92.08	93.74	89.55	95.83	92.59
Frequency resolution 512	Time windows	3	Sens	97.50	95.00	93.80	98.50	97.30	91.60	97.30	95.70	96.40	**98.80**	**99.00**	**93.00**
			Spec	98.20	96.67	98.55	97.33	98.13	99.20	98.53	97.20	98.80	**98.20**	**98.20**	**99.85**
			Sel	97.31	95.00	94.18	96.10	97.20	96.62	97.79	95.80	95.26	**97.34**	**97.35**	**99.36**
		5	Sens	95.70	90.20	95.80	95.60	92.70	90.00	92.30	90.40	90.00	96.00	95.70	83.20
			Spec	95.47	96.27	98.10	95.27	95.93	98.25	93.60	93.27	98.70	95.67	94.13	99.30
			Sel	93.37	94.15	92.65	93.09	93.83	92.78	90.58	89.95	94.54	93.66	91.58	96.74

problems (2) and (4). Results for each class i are derived in terms of sensitivity (Sens), specificity (Spec), and selectivity (Sel):

$$\text{Sens}_i = \frac{\text{Number of patterns of class } i \text{ classified in class } i}{\text{Total number of patterns in class } i}, \quad (3)$$

$$\text{Spec}_i = \frac{\text{Number of patterns not in class } i \text{ classified not in class } i}{\text{Total number of patterns not in class } i}, \quad (4)$$

$$\text{Sel}_i = \frac{\text{Number of patterns of class } i \text{ classified in class } i}{\text{Total number of patterns classified in class } i}. \quad (5)$$

The results for the classification problems (1)–(4) are shown in Tables 3–6, respectively.

The accuracy (Acc), defined as

$$\text{Acc} = \text{Trace}(\text{cm}), \quad (6)$$

where cm is the confusion matrix, defined as

$$\text{cm}_{i,j} = \text{number of patterns belonging to class } i \text{ and classified to class } j, \quad (7)$$

is calculated for each confusion matrix. The computed accuracies, along with the standard deviations are presented in Table 7. Additionally, the initial number of features and the reduced number of features after the PCA application are presented. For each classification problem, overall results have been derived, that is, the maximum and minimum accuracies (for all combinations between frequency resolutions, time windows, and frequency subbands) as well as the average accuracy and the standard deviation. For the first classification problem, the best obtained accuracy is 97.72%, achieved for 512 frequency resolution, 3 time windows, and 13 frequency subbands. For the second classification problem, the best obtained accuracy is 97.73%, achieved for 512 frequency resolution, 3 time windows, and 5 frequency subbands. For the third classification problem, the best obtained accuracy is 99.28%, achieved for 128 frequency resolution, 3 time windows, and 4 frequency subbands. Finally, for the fourth classification problem, the best obtained accuracy is 100%, achieved in most of the cases; in 28 out of 32 different

Table 4: Results for the second classification problem, in terms of sensitivity (Sens), specificity (Spec), and selectivity (Sel) in % values. Those are given for all TF resolutions (64, 128, 256, and 512), time windows (3 and 5), and frequency subbands (4, 5, 7, and 13).

Frequency subbands					4		5		7		13	
Classes					ZONF	S	ZONF	S	ZONF	S	ZONF	S
Frequency resolution	64	Time windows	3	Sens	98.40	97.60	98.55	95.60	99.30	96.00	99.10	92.40
				Spec	97.60	98.40	95.60	98.55	96.00	99.30	92.40	99.10
				Sel	99.39	93.85	98.90	94.28	99.00	97.17	98.12	96.25
			5	Sens	97.70	97.00	99.35	93.20	98.70	91.80	98.65	91.80
				Spec	97.00	97.70	93.20	99.35	91.80	98.70	91.80	98.65
				Sel	99.24	91.34	98.32	97.29	97.97	94.64	97.96	94.44
Frequency resolution	128	Time windows	3	Sens	99.50	98.40	99.25	92.60	99.55	96.80	98.05	92.40
				Spec	98.40	99.50	92.60	99.25	96.80	99.55	92.40	98.05
				Sel	99.60	98.01	98.17	96.86	99.20	98.17	98.10	92.22
			5	Sens	99.50	97.80	99.05	92.80	98.95	93.20	98.10	94.80
				Spec	97.80	99.50	92.80	99.05	93.20	98.95	94.80	98.10
				Sel	99.45	98.00	98.22	96.07	98.31	95.69	98.69	92.58
Frequency resolution	256	Time windows	3	Sens	99.25	99.00	98.90	86.40	99.45	96.60	99.40	93.60
				Spec	99.00	99.25	86.40	98.90	96.60	99.45	93.60	99.40
				Sel	99.75	97.06	96.68	95.15	99.15	97.77	98.42	97.50
			5	Sens	98.55	96.20	99.00	94.40	98.70	94.20	97.15	92.60
				Spec	96.20	98.55	94.40	99.00	94.20	98.70	92.60	97.15
				Sel	99.05	94.31	98.61	95.93	98.55	94.77	98.13	89.04
Frequency resolution	512	Time windows	3	Sens	98.90	96.20	**99.05**	**94.20**	98.85	95.60	99.70	94.20
				Spec	96.20	98.90	**94.20**	**99.05**	95.60	98.85	94.20	99.70
				Sel	99.05	95.63	**98.56**	**96.12**	98.90	95.41	98.57	98.74
			5	Sens	98.35	95.00	98.65	92.60	98.75	93.40	98.85	89.20
				Spec	95.00	98.35	92.60	98.65	93.40	98.75	89.20	98.85
				Sel	98.74	93.50	98.16	94.49	98.36	94.92	97.34	95.10

evaluations of the fourth classification problem we obtained accuracy 100%.

For the first two classification problems, the obtained accuracies of the different evaluations varied significantly; almost 6.5% (max-min) for both of them, with average 95% and standard deviation 1.7%. For the third classification problem, the max-min difference is 3% and the average 97.94%, with 0.75% standard deviation. Finally, for the fourth classification problem, the max-min difference is 1.3% and the average 99.92%, with 0.26% standard deviation.

4. DISCUSSION

We have proposed an automated method for seizure detection in EEG recordings. The method is based on TF analysis of the EEG segments and extraction of several features from the spectrum of the signal. These features are fed into neural networks, which provide the final classification of the EEG segments. The method is evaluated using four different classification problems originated from the type of medical diagnosis, which can be obtained. The effect of different parameters of the method on the classification accuracy is examined. Those parameters are the frequency resolution of the TF analysis, the length of the time window, and the width of the frequency subbands used in the feature extraction. The different combinations among all the afore-mentioned parameters result in a large number of different experimental settings (32) for each classification problem (4) and 10 different realizations (selections of training/test datasets) for each of them—totally 1280 optimized and evaluated ANNs—and results are presented for all of them. This is considered an extensive validation procedure, which can sufficiently exploit the potentials of the proposed method.

In this method, the SPWVD has been employed for the TF analysis of the EEG signals. Other distributions have been also tried but the better results were obtained for SPWVD.

The frequency resolution, used in the TF analysis, does not greatly affect the accuracy of the proposed method; the average accuracies of all different combinations of time windows and frequency subbands, for the four classification problems, are 96.71%, 97.13%, 96.7%, and 96.87% for 64, 128, 256, and 512 points length windows, respectively. It is obvious that the use of 128 points length window slightly improves the results. On the other hand, the number of the time windows is important for the analysis; in the case of three

Table 5: Results for the third classification problem, in terms of sensitivity (Sens), specificity (Spec), and selectivity (Sel) in % values. Those are given for all TF resolutions (64, 128, 256, and 512), time windows (3 and 5), and frequency subbands (4, 5, 7, and 13).

Frequency subbands				4			5			7			13		
Classes				Z	F	S	Z	F	S	Z	F	S	Z	F	S
Frequency resolution 64	Time windows	3	Sens	99.00	93.80	96.60	97.80	95.20	98.20	97.00	87.80	97.80	97.40	98.00	93.20
			Spec	98.90	97.80	98.00	98.60	98.50	98.50	94.60	98.30	98.40	99.30	94.60	98.90
			Sel	97.83	95.52	96.02	97.22	96.95	97.04	89.98	96.27	96.83	98.54	90.07	97.69
		5	Sens	94.60	84.20	96.40	96.40	92.40	92.80	95.20	92.80	94.80	90.20	90.60	93.60
			Spec	94.50	96.20	96.90	95.80	95.30	99.70	97.40	95.50	98.50	96.40	92.80	98.00
			Sel	89.58	91.72	93.96	91.98	90.77	99.36	94.82	91.16	96.93	92.61	86.29	95.90
Frequency resolution 128	Time windows	3	Sens	**99.20**	**90.60**	**97.40**	99.40	93.40	92.80	99.40	98.60	93.00	97.40	95.80	93.40
			Spec	**96.50**	**98.50**	**98.60**	97.00	97.50	98.30	98.60	97.50	99.40	98.50	96.40	98.40
			Sel	**93.41**	**96.79**	**97.21**	94.31	94.92	96.47	97.26	95.17	98.73	97.01	93.01	96.69
		5	Sens	95.40	91.60	95.60	98.20	96.20	92.20	95.00	95.40	95.20	96.40	90.60	94.00
			Spec	97.30	95.60	98.40	97.80	96.00	99.50	98.30	96.00	98.50	95.40	97.80	97.30
			Sel	94.64	91.24	96.76	95.71	92.32	98.93	96.54	92.26	96.95	91.29	95.37	94.57
Frequency resolution 256	Time windows	3	Sens	92.00	92.80	98.80	99.20	96.20	92.60	96.40	96.00	94.00	98.20	97.80	95.20
			Spec	97.90	95.40	98.50	97.90	96.50	99.60	97.80	96.60	98.80	98.40	97.80	99.40
			Sel	95.63	90.98	97.05	95.94	93.22	99.14	95.63	93.39	97.51	96.84	95.69	98.76
		5	Sens	95.60	91.40	95.40	90.20	92.80	94.40	94.20	91.00	92.40	98.40	91.80	97.00
			Spec	97.40	95.70	98.10	97.10	93.00	98.60	95.70	94.50	98.60	96.60	98.20	98.80
			Sel	94.84	91.40	96.17	93.96	86.89	97.12	91.63	89.22	97.06	93.54	96.23	97.59
Frequency resolution 512	Time windows	3	Sens	99.80	95.20	96.20	99.60	97.40	96.20	94.20	94.20	94.60	99.40	96.40	93.80
			Spec	98.20	98.30	99.10	98.40	98.60	99.60	98.00	95.20	98.30	98.30	97.30	99.20
			Sel	96.52	96.55	98.16	96.89	97.21	99.18	95.93	90.75	96.53	96.69	94.70	98.32
		5	Sens	97.60	85.80	95.40	98.00	91.20	94.00	91.40	92.60	96.40	96.20	94.00	89.40
			Spec	94.70	96.60	98.10	97.50	96.50	97.60	97.90	94.50	97.80	96.50	94.20	99.10
			Sel	90.20	92.66	96.17	95.15	92.87	95.14	95.61	89.38	95.63	93.22	89.02	98.03

time windows, the average accuracy of all different combinations between the frequency resolutions and frequency subbands, for all four classification problems, is 97.52%, while the accuracy in the case of five time windows is 96.2%. This means that analyzing EEG segments of approximately 8-second length reveals more information for the epileptic seizures than having 5-second windows. Other statistical measurements lead to the same conclusion; in the case of three time windows, the minimum accuracy of all cases is 93.04% and the standard deviation 1.8%, while the accuracy for five time windows is 91.08% and the standard deviation 2.9%, respectively. Finally, concerning the number of frequency subbands, again the reported average accuracies for all combinations among the frequency resolutions and the time windows, for all classification problems, are 97.07%, 96.87%, 96.84%, and 96.62% for 4, 5, 7, and 13 frequency subbands, respectively. This gives indications that the separation in δ, θ, α, and β rhythms is the one that mostly detects the TF components that characterize the signal regarding epileptic seizures, compared to 5 and 7, which have been used in other methods [20, 22], and 13, which is defined in this work to examine if a frequency resolution with a large number of frequency subbands improves the classification accuracy. The results indicate that all selections for frequency subbands result in similar high-average accuracies—the difference between the best and worst age accuracy is 0.45%. This can be justified since they are generated either based on expert knowledge or have been previously proposed in the literature. Concerning the frequency subbands, the higher their number, is the lower (slightly) the average accuracy obtained.

To our knowledge, TF analysis and feature extraction, which reflect the energy over the TF plane, have been only applied in the analysis of neonatal EEG signals (and mainly for neonatal epileptic seizure detection) and not EEG signals in general. Moreover, the quality of the proposed method can be proved from the obtained results. The accuracy achieved by our method for the epileptic seizure detection is more than satisfactory and also its automated nature makes it suitable to be used in real clinical conditions. Besides the feasibility of a real-time implementation of the proposed method, the diagnosis can be made more accurate by increasing the number of parameters. A system that may be developed as a result of this study may provide feedback to the experts for classification of the EEG signals quickly and accurately by examining the EEG signal.

Table 6: Results for the fourth classification problem, in terms of sensitivity (Sens), specificity (Spec), and selectivity (Sel) in % values. Those are given for all TF resolutions (64, 128, 256, and 512), time windows (3 and 5), and frequency subbands (4, 5, 7, and 13).

Frequency subbands					4		5		7		13	
Classes					Z	S	Z	S	Z	S	Z	S
Frequency resolution	64	Time windows	3	Sens	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
				Spec	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
				Sel	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
			5	Sens	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
				Spec	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
				Sel	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Frequency resolution	128	Time windows	3	Sens	**100**	**100**	**100**	**100**	**100**	**100**	100	99.80
				Spec	**100**	**100**	**100**	**100**	**100**	**100**	99.80	100
				Sel	**100**	**100**	**100**	**100**	**100**	**100**	99.80	100
			5	Sens	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
				Spec	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
				Sel	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Frequency resolution	256	Time windows	3	Sens	**100**	**100**	99.80	97.60	**100**	**100**	100	98.80
				Spec	**100**	**100**	97.60	99.80	**100**	**100**	98.80	100
				Sel	**100**	**100**	97.65	99.80	**100**	**100**	98.81	100
			5	Sens	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
				Spec	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
				Sel	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Frequency resolution	512	Time windows	3	Sens	**100**	**100**	**100**	**100**	**100**	**100**	100	99.00
				Spec	**100**	**100**	**100**	**100**	**100**	**100**	99.00	100
				Sel	**100**	**100**	**100**	**100**	**100**	**100**	99.01	100
			5	Sens	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
				Spec	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
				Sel	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**

Table 8 presents a comparison between our method and other methods proposed in the literature. Only methods evaluated in the same dataset are included so that a comparison between the results is feasible. For the two classes' problem, using only the Z and S types of EEG segments, the results obtained from the evaluation of our method are the best presented for this dataset. The difference between our result and all other results proposed in the literature varies from 0.4% to 10%. The second two classes' problem that we used to evaluate our method also presents high-accuracy results (97.73%). It is worth to mention here that a method that discriminates EEGs into nonseizure and seizure is much closer to the expert needs.

Regarding the three classes' problem, the results obtained from our method are the best presented for this dataset, either using only the Z, F, and S types or all the available dataset. In the case of using the third problem to evaluate our method (i.e., only the Z, F, and S types), the difference between our results and all others' results varies from 2.5% to 13.4%. In the case of using the first classification problem to evaluate our method (i.e., the Z and O, F and N, S types), the difference between our results and all others' results ranges from 1% to 12%. The second case has also the advantage of being a more realistic classification, dividing the dataset to normal, seizure-free, and seizure EEGs, and thus being closer to clinical conditions.

Still, however, there are several other aspects either technical or medical which must be addressed. From the technical point of view, although we have examined the effect of various parameters (frequency resolution, number of time windows, and frequency bands), some other, like time-frequency distributions (e.g., reduced interference distributions), have not been explored. Furthermore, we mainly focused on the effects of the parameters related to frequency analysis, either for the calculation of the spectrum of the signal or for the frequency resolution for feature extraction. More detailed examination of the time resolution for feature extraction may also reveal important information regarding the seizure detection; this feature will be addressed in feature communications. From the medical point of view, the most important feature is that currently the method is used to characterize predetermined (with respect to their length) EEG segments. An important aspect is also the modification of the proposed method in order to be able to automatically detect highly suspicious segments (regardless of their length) into long time EEG recordings and classify them.

Table 7: Accuracy (%), standard deviation (in the parenthesis), and initial number of features/reduced number of features after PCA application, for all classification problems (1, 2, 3, and 4) reported, for all TF resolutions (64, 128, 256, and 512), time windows (3 and 5), and frequency subbands (4, 5, 7, and 13).

Frequency resolution	Time windows	Frequency subbands	Classification problem			
			1	2	3	4
64	3	4	96.64 (0.34) 13/3	96.47(0.45) 13/3	98.24 (0.33) 13/3	**100 (0) 13/3**
		5	97 (0.76) 16/3	97.07(0.78) 16/3	97.96 (0.61) 16/3	**100 (0) 16/3**
		7	94.36 (0.58) 22/5	94.2 (0.89) 22/5	98.64 (0.34) 22/5	**100 (0) 22/5**
		13	95.72 (0.71) 40/4	95.2 (1.25) 40/4	97.76 (0.33) 40/4	**100 (0) 40/4**
	5	4	93.08 (0.96) 21/4	91.73 (0.84) 21/4	97.56 (0.39) 21/4	**100 (0) 21/4**
		5	94.24 (0.54) 26/4	93.87 (1.08) 26/4	98.12 (0.6) 26/4	**100 (0) 26/4**
		7	94.08 (0.7) 36/4	94.27 (0.95) 36/4	97.32 (0.19) 36/4	**100 (0) 36/4**
		13	92.36 (0.81) 66/4	91.47 (0.82) 66/4	97.28 (0.37) 66/4	**100 (0) 66/4**
128	3	4	97.36 (0.34) 13/3	95.73 (0.47) 13/3	**99.28 (0.17) 13/3**	**100 (0) 13/3**
		5	95.48 (0.33) 16/3	95.2 (0.61) 16/3	97.92 (0.32) 16/3	**100 (0) 16/3**
		7	97.52 (0.25) 22/4	97 (0.47) 22/4	99 (0.34) 22/4	**100 (0) 22/4**
		13	93.48 (0.80) 40/5	95.53 (1.3) 40/5	96.92 (0.42) 40/5	99.9 (0.32) 40/5
	5	4	94.92 (0.71) 21/4	94.2 (1.41) 21/4	99.16 (0.35) 21/4	**100 (0)** 21/4
		5	94.36 (0.72) 26/4	95.53 (0.71) 26/4	97.8 (0.28) 26/4	**100 (0) 26/4**
		7	93.92 (1.1) 36/4	95.2 (0.93) 36/4	97.8 (0.39) 36/4	**100 (0) 36/4**
		13	93.72 (0.9) 66/5	93.67 (1.18) 66/5	97.44 (0.47) 66/5	**100 (0) 66/5**
256	3	4	96.52 (0.27) 13/3	94.53 (0.42) 13/3	99.2 (0) 13/3	**100 (0) 13/3**
		5	93.04 (0.78) 16/3	96 (0.7) 16/3	96.4 (0.53) 16/3	98.7 (0.82) 16/3
		7	96.84 (0.35) 22/5	95.47 (0.53) 22/5	98.88 (0.41) 22/5	**100 (0) 22/5**
		13	96.4 (0.9) 40/6	97.07 (0.84) 40/6	98.24 (0.39) 40/6	99.4 (0.52) 40/6
	5	4	94.24 (0.8) 21/4	94.13 (1.21) 21/4	98.08 (0.53) 21/4	**100 (0) 21/4**
		5	92.88 (0.53) 26/5	92.47 (1.18) 26/5	98.08 (0.49) 26/5	**100 (0) 26/5**
		7	92.92 (0.6) 36/5	92.53 (0.61) 36/5	97.8 (0.43) 36/5	**100 (0) 36/5**
		13	92.52 (0.71) 66/5	95.73 (0.84) 66/5	96.24 (0.63) 66/5	**100 (0) 66/5**
512	3	4	95.76 (0.28) 13/3	97.07 (0.72) 13/3	98.36 (0.4) 13/3	**100 (0) 13/3**
		5	96.64 (0.34) 16/4	**97.73 (1) 16/4**	98.08 (0.62) 16/4	**100 (0) 16/4**
		7	96.48 (0.59) 22/5	94.33 (0.85) 22/5	98.2 (0.28) 22/5	**100 (0) 22/5**
		13	**97.72 (0.38) 40/6**	96.53 (0.69) 40/6	98.6 (0.47) 40/6	99.5 (0.53) 40/6
	5	4	93.52 (0.67) 21/5	92.93 (0.9) 21/5	97.68 (0.41) 21/5	**100 (0) 21/5**
		5	93.32 (0.46) 26/5	94.4 (1.1) 26/5	97.44 (0.43) 26/5	**100 (0) 26/5**
		7	91.08 (1.18) 36/5	93.47 (0.88) 36/5	97.68 (0.45) 36/5	**100 (0) 36/5**
		13	93.32 (1.16) 66/5	93.2 (1.47) 66/5	96.92 (0.5) 66/5	**100 (0) 66/5**
		Total				
		Max	97.72	97.73	99.28	100
		Min	91.08	91.47	96.24	98.7
		Average	94.73	94.81	97.94	99.92
		SD	1.78	1.63	0.75	0.26

5. CONCLUSIONS

In this paper, we explored the ability of the TF analysis to classify EEG segments which contain epileptic seizures. We have extracted several time-frequency features and we examined the effect of the parameters entering the problem, that is, the frequency resolution of the time-frequency analysis and the number of time windows and frequency subbands used for feature extraction. Promising results have been reported after the evaluation of the proposed method in four different classification problems, derived from a well-known database. However, several types of artefacts have been removed from this database after visual inspection. This is a limitation of the evaluation of our method and thus further evaluation under real clinical conditions is required in order to fully exploit its potential. Another limitation is that in the current study high-frequency components (over 40 Hz) were not measured and thus taken under consideration; the

Table 8: A comparison of the results obtained by our method and others' methods (classification accuracy) for two and three categories classification problems.

Classes	Authors (year)	Method	Dataset	Accuracy
2	Nigam et al. [15] (2004)	Nonlinear preprocessing filter, diagnostic artificial neural network (LAMSTAR)	Z, S	97.2
	Srinivasan et al. [14] (2005)	Time & frequency domain features, recurrent neural network (RNN)	Z, S	99.6
	Kannathal et al. [42] (2005)	Entropy measures, adaptive neurofuzzy inference system (ANFIS)	Z, S	92.22
	Kannathal et al. [35] (2005)	Chaotic measures, surrogate data analysis	Z, S	~ 90
	Polat et al. [16] (2006)	Fast Fourier transform (FFT), decision tree (DT)	Z, S	98.72
	Subasi [22] (2007)	Discrete wavelet transform (DWT), mixture of expert model	Z, S	95
	This work (2007)	**Time frequency (TF) analysis, artificial neural network (ANN)**	**Z, S**	**100**
	This work (2007)	**Time frequency (TF) analysis, artificial neural network (ANN)**	**(Z, O, N, F), S**	**97.73**
3	Guler et al. [34] (2005)	Lyapunov exponents, recurrent neural network (RNN)	Z, F, S	96.79
	Sadati et al. [23] (2006)	Discrete wavelet transform (DWT), adaptive neural fuzzy network (ANFN)	Z, F, S	85.9
	This work (2007)	**Time frequency (TF) analysis, artificial neural network (ANN)**	**Z, F, S**	**99.28**
	This work (2007)	**Time frequency (TF) analysis, artificial neural network (ANN)**	**(Z, O), (N, F), S**	**97.72**

employment of high-frequency components, such as gamma activity, and their importance concerning epileptic seizure detection will be addressed in a future communication. Finally, several technical aspects can be further investigated, such as different techniques for feature reduction and alternative classification algorithms.

REFERENCES

[1] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lenhnertz, "Seizure prediction: the long and winding road," *Brain*, vol. 130, no. 2, pp. 313–333, 2006.

[2] B. Litt and J. Echauz, "Prediction of epileptic seizures," *Lancet Neurology*, vol. 1, no. 1, pp. 22–30, 2002.

[3] K. Lehnertz, F. Mormann, T. Kreuz, et al., "Seizure prediction by nonlinear EEG analysis," *IEEE Engineering in Medicine and Biology Magazine*, vol. 22, no. 1, pp. 57–63, 2003.

[4] E. Waterhouse, "New horizons in ambulatory electroencephalography," *IEEE Engineering in Medicine and Biology Magazine*, vol. 22, no. 3, pp. 74–80, 2003.

[5] N. McGrogan, *Neural network detection of epileptic seizures in the electroencephalogram*, Ph.D. thesis, Oxford University, Oxford, UK, February 1999.

[6] J. Gotman, "Automatic detection of seizures and spikes," *Journal of Clinical Neurophysiology*, vol. 16, no. 2, pp. 130–140, 1999.

[7] R. O. Sirne, S. I. Isaacson, and C. E. D'Attellis, "A data-reduction process for long-term EEGS: feature extraction through digital processing in a multiresolution framework," *IEEE Engineering in Medicine and Biology Magazine*, vol. 18, no. 1, pp. 56–61, 1999.

[8] P. F. Prior, R. S. M. Virden, and D. E. Maynard, "An EEG device for monitoring seizure discharges," *Epilepsia*, vol. 14, no. 4, pp. 367–372, 1973.

[9] J. Gotman, "Automatic recognition of epileptic seizures in the EEG," *Electroencephalography and Clinical Neurophysiology*, vol. 54, no. 5, pp. 530–540, 1982.

[10] W. R. S. Webber, R. P. Lesser, R. T. Richardson, and K. Wilson, "An approach to seizure detection using an artificial neural network (ANN)," *Electroencephalography and Clinical Neurophysiology*, vol. 98, no. 4, pp. 250–272, 1996.

[11] G. W. Harding, "An automated seizure monitoring system for patients with indwelling recording electrodes," *Electroencephalography and Clinical Neurophysiology*, vol. 86, no. 6, pp. 428–437, 1993.

[12] A. M. Murro, D. W. King, J. R. Smith, B. B. Gallagher, H. F. Flanigin, and K. Meador, "Computerized seizure detection of complex partial seizures," *Electroencephalography and Clinical Neurophysiology*, vol. 79, no. 4, pp. 330–333, 1991.

[13] A. J. Gabor, "Seizure detection using a self-organizing neural network: validation and comparison with other detection strategies," *Electroencephalography and Clinical Neurophysiology*, vol. 107, no. 1, pp. 27–32, 1998.

[14] V. Srinivasan, C. Eswaran, and A. N. Sriraam, "Artificial neural network based epileptic detection using time-domain and frequency-domain features," *Journal of Medical Systems*, vol. 29, no. 6, pp. 647–660, 2005.

[15] V. P. Nigam and D. Graupe, "A neural-network-based detection of epilepsy," *Neurological Research*, vol. 26, no. 1, pp. 55–60, 2004.

[16] K. Polat and S. Güneş, "Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast

Fourier transform," *Applied Mathematics and Computation*, vol. 187, no. 2, pp. 1017–1026, 2007.

[17] A. Alkan, E. Koklukaya, and A. Subasi, "Automatic seizure detection in EEG using logistic regression and artificial neural network," *Journal of Neuroscience Methods*, vol. 148, no. 2, pp. 167–176, 2005.

[18] B. Gonzalez-Vellon, S. Sanei, and J. A. Chambers, "Support vector machines for seizure detection," in *Proceedings of the 3rd IEEE International Symposium on Signal Processing and Information Technology (ISSPIT '03)*, pp. 126–129, Darmstadt, Germany, December 2003.

[19] S. Blanco, C. E. D'Attellis, S. I. Isaacson, O. A. Rosso, and R. O. Sirae, "Time-frequency analysis of electroencephalogram series. II. Gabor and wavelet transforms," *Physical Review E*, vol. 54, no. 6, pp. 6661–6672, 1996.

[20] H. Adeli, Z. Zhou, and N. Dadmehr, "Analysis of EEG records in an epileptic patient using wavelet transform," *Journal of Neuroscience Methods*, vol. 123, no. 1, pp. 69–87, 2003.

[21] A. Subasi, A. Alkan, E. Koklukaya, and M. K. Kiymik, "Wavelet neural network classification of EEG signals by using AR model with MLE preprocessing," *Neural Networks*, vol. 18, no. 7, pp. 985–997, 2005.

[22] A. Subasi, "EEG signal classification using wavelet feature extraction and a mixture of expert model," *Expert Systems with Applications*, vol. 32, no. 4, pp. 1084–1093, 2007.

[23] N. Sadati, H. R. Mohseni, and A. Maghsoudi, "Epileptic seizure detection using neural fuzzy networks," in *Proceedings of IEEE International Conference on Fuzzy Systems (FUZZY '06)*, pp. 596–600, Vancouver, BC, Canada, July 2006.

[24] A. A. Dingle, R. D. Jones, G. J. Carroll, and W. R. Fright, "A multistage system to detect epileptiform activity in the EEG," *IEEE Transactions on Biomedical Engineering*, vol. 40, no. 12, pp. 1260–1268, 1993.

[25] A. T. Tzallas, P. S. Karvelis, C. D. Katsis, D. I. Fotiadis, S. Giannopoulos, and S. Konitsiotis, "A method for classification of transient events in EEG recordings: application to epilepsy diagnosis," *Methods of Information in Medicine*, vol. 45, no. 6, pp. 610–621, 2006.

[26] F. I. M. Argoud, F. M. de Azevedo, J. M. Neto, and E. Grillo, "SADE^3: an effective system for automated detection of epileptiform events in long-term EEG based on context information," *Medical and Biological Engineering and Computing*, vol. 44, no. 6, pp. 459–470, 2006.

[27] B. L. K. Davey, W. R. Fright, G. J. Carroll, and R. D. Jones, "Expert system approach to detection of epileptiform activity in the EEG," *Medical and Biological Engineering and Computing*, vol. 27, no. 4, pp. 365–370, 1989.

[28] P. E. McSharry, T. He, L. A. Smith, and L. Tarassenko, "Linear and non-linear methods for automatic seizure detection in scalp electro-encephalogram recordings," *Medical and Biological Engineering and Computing*, vol. 40, no. 4, pp. 447–461, 2002.

[29] N. Paivinen, S. Lammi, A. Pitkanen, J. Nissinen, M. Penttonen, and T. Gronfors, "Epileptic seizure detection: a nonlinear viewpoint," *Computer Methods and Programs in Biomedicine*, vol. 79, no. 2, pp. 151–159, 2005.

[30] N. Kannathal, U. R. Acharya, C. M. Lim, and P. K. Sadasivan, "Characterization of EEG—a comparative study," *Computer Methods and Programs in Biomedicine*, vol. 80, no. 1, pp. 17–23, 2005.

[31] L. D. Iasemidis and J. C. Sackellares, "Chaos theory and epilepsy," *Neuroscientist*, vol. 2, no. 2, pp. 118–126, 1996.

[32] D. E. Lerner, "Monitoring changing dynamics with correlation integrals: case study of an epileptic seizure," *Physica D*, vol. 97, no. 4, pp. 563–576, 1996.

[33] K. Lehnertz and C. E. Elger, "Spatio-temporal dynamics of the primary epileptogenic area in temporal lobe epilepsy characterized by neuronal complexity loss," *Electroencephalography and Clinical Neurophysiology*, vol. 95, no. 2, pp. 108–117, 1995.

[34] N. F. Güler, E. D. Übeyli, and İ. Güler, "Recurrent neural networks employing Lyapunov exponents for EEG signals classification," *Expert Systems with Applications*, vol. 29, no. 3, pp. 506–514, 2005.

[35] N. Kannathal, M. L. Choo, U. R. Acharya, and P. K. Sadasivan, "Entropies for detection of epilepsy in EEG," *Computer Methods and Programs in Biomedicine*, vol. 80, no. 3, pp. 187–194, 2005.

[36] H. Qu and J. Gotman, "A patient-specific algorithm for the detection of seizure onset in long-term EEG monitoring: possible use as a warning device," *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 2, pp. 115–122, 1997.

[37] A. B. Gardner, A. M. Krieger, G. Vachtsevanos, and B. Litt, "One-class novelty detection for seizure analysis from intracranial EEG," *Journal of Machine Learning Research*, vol. 7, pp. 1025–1044, 2006.

[38] H. H. Lange, J. P. Lieb, J. Engel Jr., and P. H. Crandall, "Temporo-spatial patterns of pre-ictal spike activity in human temporal lobe epilepsy," *Electroencephalography and Clinical Neurophysiology*, vol. 56, no. 6, pp. 543–555, 1983.

[39] S. J. Schiff, D. Colella, G. M. Jacyna, et al., "Brain chirps: spectrographic signatures of epileptic seizures," *Clinical Neurophysiology*, vol. 111, no. 6, pp. 953–958, 2000.

[40] K. Lehnertz, R. G. Andrzejak, J. Arnhold, et al., "Nonlinear EEG analysis in epilepsy: its possible use for interictal focus localization, seizure anticipation, and prevention," *Journal of Clinical Neurophysiology*, vol. 18, no. 3, pp. 209–222, 2001.

[41] M. Le Van Quyen, J. Martinerie, V. Navarro, et al., "Anticipation of epileptic seizures from standard EEG recordings," *The Lancet*, vol. 357, no. 9251, pp. 183–188, 2001.

[42] A. B. Geva and D. H. Kerem, "Forecasting generalized epileptic seizures from the EEG signal by wavelet analysis and dynamic unsupervised fuzzy clustering," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 10, pp. 1205–1216, 1998.

[43] B. Boashash, M. Mesbah, and P. Golditz, "Time frequency detection of EEG abnormalities," in *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*, chapter 15, pp. 663–670, Elsevier, Oxford, UK, 2003.

[44] H. Hassanpour, M. Mesbah, and B. Boashash, "Time-frequency based newborn EEG seizure detection using low and high frequency signatures," *Physiological Measurement*, vol. 25, no. 4, pp. 935–944, 2004.

[45] B. Boashash and M. Mesbah, "Time-frequency methodology for newborn electroencephalographic seizure detection," in *Applications in Time-Frequency Signal Processing*, A. Papandreou-Suppappola, Ed., chapter 9, pp. 339–369, CRC Press, Boca Raton, Fla, USA, 2003.

[46] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state," *Physical Review E*, vol. 64, no. 6, Article ID 061907, 8 pages, 2001.

[47] H. H. Jasper, "Ten-twenty electrode system of the International Federation," *Electroencephalography and Clinical Neurophysiology*, vol. 10, pp. 371–375, 1958.

[48] R. L. Allen and D. W. Mills, *Signal Analysis: Time, Frequency, Scale, and Structure*, Wiley-IEEE Press, New York, NY, USA, 2004.

[49] F. Auger, P. Flandrin, P. Gonçalvès, and O. Lemoine, *Time-Frequency Toolbox*, Rice University, CNRS, France, 1996.

[50] C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford, UK, 1995.

Hindawi Publishing Corporation
Computational Intelligence and Neuroscience
Volume 2007, Article ID 58253, 10 pages
doi:10.1155/2007/58253

Research Article

Canonical Decomposition of Ictal Scalp EEG and Accurate Source Localisation: Principles and Simulation Study

Maarten De Vos,[1] Lieven De Lathauwer,[2] Bart Vanrumste,[1] Sabine Van Huffel,[1] and W. Van Paesschen[3]

[1] *ESAT-SISTA, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, 3001 Heverlee-Leuven, Belgium*
[2] *CNRS-ETIS, 6 Avenue du Ponceau BP 44, 95014 Cergy-Pontoise, France*
[3] *Department of Neurology, University Hospital Gasthuisberg, Katholieke Universiteit Leuven, Herestraat 49, 3000 Leuven, Belgium*

Correspondence should be addressed to Maarten De Vos, maarten.devos@esat.kuleuven.be

Received 16 February 2007; Revised 13 June 2007; Accepted 2 October 2007

Recommended by Andrzej Cichocki

Long-term electroencephalographic (EEG) recordings are important in the presurgical evaluation of refractory partial epilepsy for the delineation of the ictal onset zones. In this paper, we introduce a new concept for an automatic, fast, and objective localisation of the ictal onset zone in ictal EEG recordings. Canonical decomposition of ictal EEG decomposes the EEG in atoms. One or more atoms are related to the seizure activity. A single dipole was then fitted to model the potential distribution of each epileptic atom. In this study, we performed a simulation study in order to estimate the dipole localisation error. Ictal dipole localisation was very accurate, even at low signal-to-noise ratios, was not affected by seizure activity frequency or frequency changes, and was minimally affected by the waveform and depth of the ictal onset zone location. Ictal dipole localisation error using 21 electrodes was around 10.0 mm and improved more than tenfold in the range of 0.5–1.0 mm using 148 channels. In conclusion, our simulation study of canonical decomposition of ictal scalp EEG allowed a robust and accurate localisation of the ictal onset zone.

1. INTRODUCTION

Epilepsy is one of the most common, severe neurological diseases. People suffering from epilepsy, who are not helped by medication, can potentially benefit from epilepsy surgery [1]. In order to remove the epileptogenic region, a precise localisation of the epileptic focus is mandatory. One of the diagnostic tools to localize this region of seizure onset zone is recording of ictal scalp electroencephalogram (EEG) [2]. The EEG measures electric potential distributions at discrete recording sites on the scalp. These potential distributions are the direct consequence of internal electrical currents associated with the synchronous firing of neurons. EEG recordings have an excellent temporal resolution, but a rather poor spatial accuracy due to the limited number of recording sites and the shielding effect of the skull. Visual analysis of EEG recordings aims to determine which lobe or which electrodes are activated. A challenging problem in neuroscience is to estimate in a more objective and precise way the regions of the brain that are active, given only the measured potential distributions.

Estimating the electrical source in the brain from the scalp EEG is a difficult problem since an infinite number of internal electrical currents can generate the same potential distribution on the scalp. Several different approaches to solve this source localisation or inverse problem exist based on different assumptions [3, 4]. One assumption is that the surface potentials are generated by a dense set of dipolar sources distributed on the cortical surface. The most popular method from this "distributed source" family is Loreta [5]. In a second approach, which is the most common, a limited number of "equivalent dipoles" are assumed to generate the measured potential distribution [6]. Dipole modeling is a well-established technique for localising interictal spikes, see, for example, [7, 8] and references herein. Ictal EEG recordings have been subjected to dipole modeling much less often than interictal spikes. The seizure discharge is a very complex pattern. Mainly artifacts, such as electromyogram, movement, eye blinks, and eye movements artifacts, render modeling difficult [9]. Even visual analysis of seizure onset can be significantly improved by removing muscle artifacts [10]. Moreover, the low signal-to-noise ratio of the seizure

signal can render the correct localisation very diffuse. However, when source localisation of seizure onset would be possible, it can reduce the need for invasive intracranial EEG recordings. So far, the results of ictal EEG source localisation have been discouraging. One study reports that the used "inverse solution" [11] is not useful at all for localising seizure onsets [12]. Some studies were restricted to temporal lobe seizures [13, 14]. One reason to select temporal lobe seizures is that source analysis is most reliable during periods of relative signal stationarity in order to average repetitive ictal waveforms, which is more common in temporal than in extratemporal lobe seizures. Another reason for selecting only temporal lobe seizures is that extratemporal lobe seizures are much more frequently contaminated by severe artifacts. Two other studies were not restricted to temporal lobe seizures. Gotman [9] obtained reliable models for seizure onset in 6 out of 15 patients (40%) and Boon et al. [15] in 31 out of 100 patients (31%). In the latter study, the ictal EEG was filtered with a narrowband filter (1–14 Hz), while ictal seizure activity is known to consist of rhythmical waves with a frequency between 3 and 29 Hz [16]. Filtering should be avoided because these filters suppress all high-frequency activity, including electrical brain activity. Moreover, muscle artifacts filtered by a lowpass filter can resemble cerebral activity [17]. All these studies illustrate how difficult it is to reliably estimate ictal sources, and indicate that the current ictal scalp EEG source analysis tools can not be used for a reliable localisation of the ictal onset zone during presurgical evaluation. A recent study on source analysis developed a novel integrative approach to characterise the structure of seizures in the space, time, and frequency domains and showed some promising results [18].

The localising value of dipole modeling of ictal EEG can be improved by first removing artifacts and afterwards estimating the sources [19]. Another possibility is to decompose the measured EEG in a sum of individual contributions of distinct brain sources and localising the epilepsy-related source in order to estimate the epileptic focus. Space-time decomposition techniques like principal component analysis (PCA) and independent component analysis (ICA) of multichannel EEG can be used for artifact removal [20, 21] or for extracting activities of interest [22, 23]. However, in order to obtain a matrix decomposition like PCA and ICA, assumptions like orthogonality or independence—which are physically maybe irrelevant—have to be imposed. Recently, we have shown that a space-time-frequency decomposition of a three-way array containing wavelet-transformed EEG by the canonical decomposition (Candecomp), also known as parallel factor analysis (Parafac), reliably separated a seizure atom from the noise and background activity with a sensitivity of more than 90% [24]. This work was inspired by [25, 26]. The main advantage of this decomposition is that no extra assumptions have to be imposed. After the decomposition, the potential distribution over the electrodes of the epilectical activity was obtained, and displayed as a 2D image. Electrodes with large potential amplitudes could be considered as close to the focus. The aim of the present study was twofold. First, we wanted to investigate whether it was possible to localise the ictal onset zone in the head by applying dipole source localisation after canonical decomposition of ictal EEG recordings. Second, we wanted to investigate the accuracy of this localising method with realistic simulations under different conditions. We were especially interested (i) in the influence of the frequency of the seizure activity on the localisation, (ii) how the dipole localisation would be influenced by changes in frequency, and (iii) if the dipole estimation accuracy could be improved by increasing the number of electrodes.

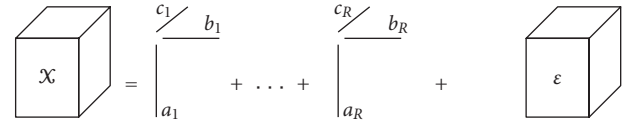


Figure 1: The Candecomp model with R components.

We start by revising the canonical decomposition of a higher-order array (Section 2.1). We then define how we constructed realistically simulated EEG (Section 2.2), assessed the accuracy of our method (Section 3) and finally discuss our results (Section 4).

2. MATERIALS AND METHODS

2.1. Method

In our application, a three-way data array $\mathcal{X}$ with dimensions (space, scale, time) is obtained by wavelet-transforming every channel of the original (or simulated) EEG matrix. The continuous wavelet transform C at scale a and time t of a signal $x(t)$ is defined as

$$C(a,t) = \int_{-\infty}^{\infty} x(t)\varphi^{\star}(a,t,\tau)d\tau \tag{1}$$

with $\phi^{\star}$ the chosen wavelet. Different real wavelets can be used. In this study, we used a biorthogonal wavelet with decomposition order 3. From the scale a of the wavelet, the frequency f of the signal can be estimated as

$$f \approx \frac{f_c}{(a\Delta t)} \tag{2}$$

with f_c the center frequency of the wavelet and Δt the sampling period.

The trilinear Candecomp [27–29] is a generalisation of the singular value decomposition (SVD) for higher orders. It is defined for a three-way array $\mathcal{X}(I \times J \times K)$ as

$$x_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr} + e_{ijk}, \tag{3}$$

where R is the number of components used in the Candecomp model and e_{ijk} are the residuals containing the unexplained variation. A pictorial representation of the Candecomp model is given in Figure 1. The Candecomp model is a trilinear model: fixing the parameters in two modes, x_{ijk} is expressed as a linear function of the remaining parameters. Another equivalent and useful expression of the same Candecomp model is given with the Khatri-Rao product $\odot$, defined as the column-wise Kronecker product [30].

Stack the elements of the tensor $\mathcal{X}^{I\times J\times K}$ in a matrix $\mathbf{X}^{IJ\times K}$ as

$$\mathbf{X}_{(i-1)J+j,k} = x_{ijk}. \tag{4}$$

Construct a matrix $\mathbf{E}$ in a similar way. Collect the elements a_{ir} in $\mathbf{A}$; b_{jr} in $\mathbf{B}$ and c_{kr} in $\mathbf{C}$. Then

$$\mathbf{X}^{IJ\times K} = \left(\mathbf{A}^{I\times R} \odot \mathbf{B}^{J\times R}\right)\left(\mathbf{C}^{K\times R}\right)^{T} + \mathbf{E}^{IJ\times K}. \tag{5}$$

Comparing the number of free parameters of a generic tensor and a Candecomp model, it can be seen that this model is very restricted. The advantage of this model is its uniqueness under mild conditions [31–33]:

$$k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}} \geqslant 2R + 2 \tag{6}$$

with $k_{\mathbf{M}}$ the k-rank of matrix $\mathbf{M}$. The k-rank of matrix $\mathbf{M}$ is defined as the maximal number r such that any set of r columns of $\mathbf{M}$ is linearly independent. For tensors of which one dimension is greater than the rank, another less restrictive condition has recently been derived in [34].

The canonical decomposition is usually computed by means of an alternating least-squares (ALS) algorithm [30]. This means that the least-squares cost function

$$f(A,B,C) = \left|\left|\mathcal{X} - \sum_{r=1}^{R} A_r \circ B_r \circ C_r\right|\right|^2 \tag{7}$$

is minimized by means of alternating updates of one of its matrix arguments, keeping the other two matrices fixed. Because the canonical decomposition is a multilinear decomposition, each update just amounts to solving a classical linear least-squares problem. The convergence may be local. To increase the probability that the global minimum is found, the algorithm is reinitialized a couple of times. Since the introduction of the ALS algorithm, other computational schemes have been proposed [34–37].

When Candecomp is used for seizure localisation, 2 seconds of EEG at the seizure onset is wavelet transformed. The obtained three-way array is decomposed with Candecomp with R atoms. Several techniques exist to determine the optimal number of atoms [30]. Corcondia was used to determine the optimal number of atoms R. After decomposition, each atom has a component in the space ($\mathbf{a}_i$), time ($\mathbf{b}_i$), and frequency domain ($\mathbf{c}_i$). The seizure atom(s) can be selected based on characteristic signatures in the different domains. At the ictal onset, seizure activity is recognised by rhythmical activity that is well localised in space and frequency. This was also described in [38]. Another possibility is to reconstruct the decomposed atoms in EEG settings by means of the inverse continuous wavelet transform (ICWT) [39]. We illustrate this approach with an example. EEG containing clear ictal activity in the right temporal lobe is given in Figure 2. The seizure starts at Second 3, and the EEG between Second 3 and 5 is wavelet transformed and decomposed with Candecomp (see Figure 3). Corcondia indicated that a decomposition in two atoms would be appropriate. The first atom is recognized as seizure atom. The frequency component peaks around 3 Hz and the time component is a rhythmical waveform that increases in amplitude. When this component is reconstructed in EEG settings, pure ictal activity can be seen (see Figure 4). Because the atoms in the canonical decomposition have a very simple, trilinear structure, we propose to fit only 1 dipole for every atom. We expect that, when a patient suffers from multifocal epilepsy, different atoms will be related to activity generated by the different dipoles.

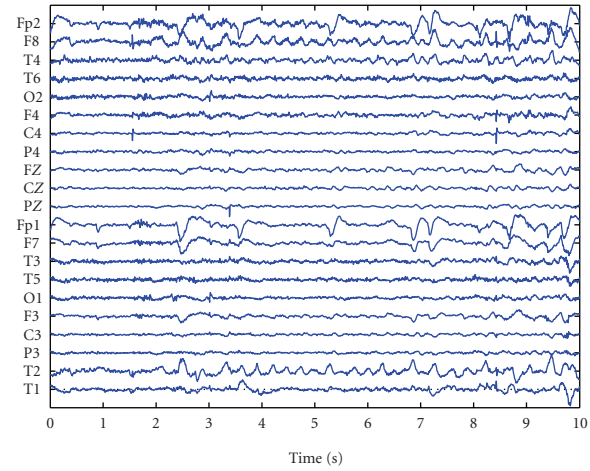


Figure 2: 10 seconds of EEG containing the start of a seizure.

Dipole estimation then determines the dipole's coordinates and orientation that best generate the given potential distribution in a least-squares sense. For computational simplicity, we used a spherical head model in this study.

2.2. Simulation

Consider a matrix $\mathbf{X}$ of dimension 500-by-21 representing a 21-channel EEG section of 2.0 seconds long. Each vector $\mathbf{x_s}$, $s = 1,\ldots,21$ of $\mathbf{X}$ contains the time course of an EEG channel:

$$\mathbf{X} = [\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_{21}}]^T. \tag{8}$$

In this simulation study $\mathbf{X}$ includes both seizure activity, and superimposed noise. Both signals are described as follows.

2.2.1. Synthetic seizure activity

The EEG of the ictal activity was generated using a fixed dipole in a three-shell spherical head model. The different time courses generated by the dipole are described below. The amplification factors at each electrode were computed by solving the forward problem for a dipole in a three-shell spherical head model consisting of a brain, a skull, and a scalp compartment [40]. Each compartment had a specific conductivity with a ratio equal to 1:1/16:1 for the brain, skull, and scalp compartment, respectively [41]. The brain and scalp conductivity was $3.3 \times 10^{-4}/\Omega$mm [42]. Radii of the outer boundary of the brain, skull, and scalp region equal to, respectively, 8 cm, 8.5 cm and 9.2 cm were used. Anumber of 21 electrodes were used: *Fp2, F8, T4, T6, O2, F4, C4, P4,*

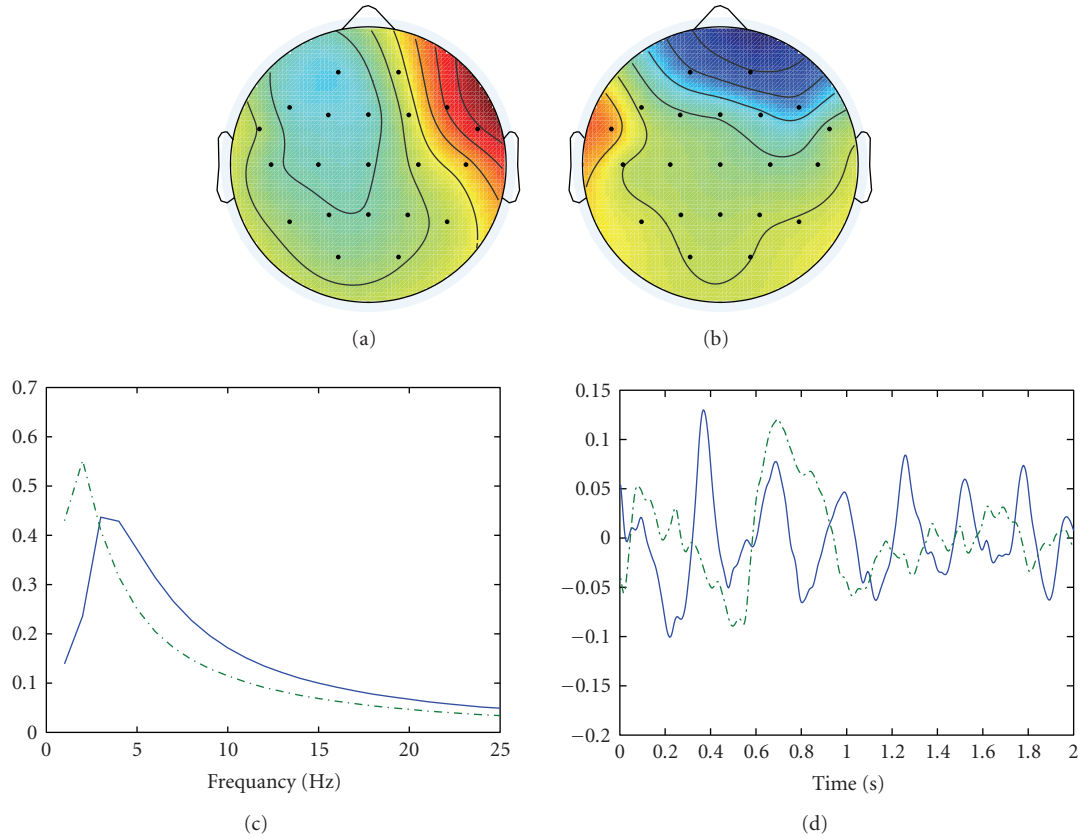


FIGURE 3: Seconds 3 to 5 of the seizure shown in Figure 2 are decomposed with the canonical decomposition with 2 atoms. (a) (b) the spatial potential distributions of the two atoms. (c) The frequency content of the atoms. (d) The time course of the atoms. First atom drawn in solid line correspond with (a). Dash-dotted line correspond with (b). First atom is seizure atom.

Fz, Cz, Pz, Fp1, F7, T3, T5, O1, F3, C3, and *P3* placed according to the 10–20 system for electrode placement [43] and additional electrodes T1 and T2 on the temporal region. The time course of the scalp potentials was stored in a 500-by-21 dimensional matrix **A**, representing 2 seconds of EEG with sample frequency of 250 Hz.

Unless otherwise stated, dipole coordinates x (left ear to right ear), y (posterior to anterior) and z (up, through the Cz electrode) were $[-0.5\ 0\ 0.1]$ and the dipole orientations d_x, d_y, and d_z were $[1\ 0\ 0]$.

The following seizure characteristics were simulated:

(A) Seizure activity in patients with mesial temporal lobe epilepsy (MTLE) is typically expressed by a 4 Hz sinusoidal waveform [44]. In a first simulation we estimated the dipole localisation error when seizure activity was represented by a 4 Hz sinusoid at different noise levels (see Figure 5(a)). We also investigated the influence of the specific waveform and estimated the localisation error when seizure activity was represented by a 4 Hz sawtooth, instead of a sinusoidal wave, at different noise levels.

(B) Ictal EEG activity can have a frequency in the delta, theta, alpha, or beta range. In a second simulation, therefore, we estimated the influence of the frequency of the seizure signal on ictal scalp EEG source localisation at a fixed noise level. We were particularly interested if the possible overlap in frequency content between faster ictal activity and seizure activity would bias the decomposition and thus the dipole estimate.

(C) Epileptic seizure activity can rapidly change in frequency. Ictal EEG activity is often characterized by low-voltage fast activity in the beta range which gradually slows down to alpha or theta frequencies with increasing amplitude. The canonical decomposition exploits frequency information during the decomposition. In order to test possible shortcomings of the canonical decomposition of ictal EEG, we wanted to estimate the accuracy when the model is violated. In a third simulation, we assessed the dipole localisation error when the frequency changed during the 2 seconds under investigation. This does not give a trilinear signal after wavelet transformation. We simulated a chirp that linearly changed in frequency from 8 Hz at the start to 4 Hz at the end of the considered 2 seconds. The signal also doubled in amplitude.

(D) In our previous study [24], two atoms were obtained after the decomposition of in vivo seizures and a distinction could be made between a seizure and a nonseizure atom. An interesting question is how well

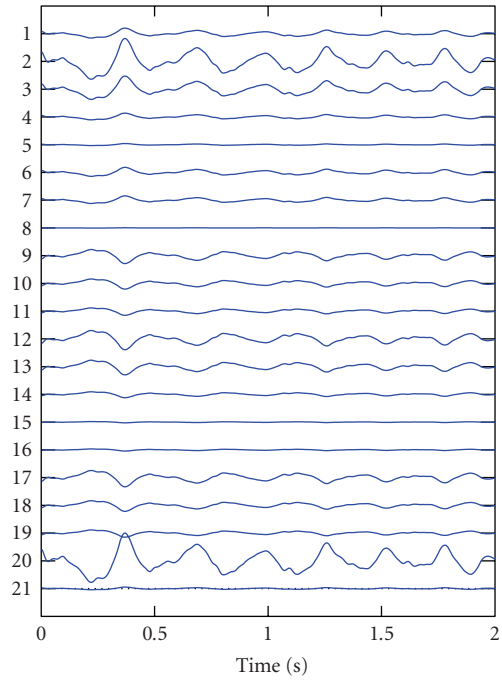


FIGURE 4: The seizure atom from Figure 3 is reconstructed in EEG coordinates after canonical decomposition.

different dipoles generating similar ictal signals will be distinguished from each other. Such activity can be measured in the case of multifocal epilepsy. In a fourth simulation, we considered two rhythmical sources firing at the same frequency separated from each other by about 1 cm: the second dipole had coordinates $[-0.4\ 0\ 0.1]$. These dipoles generated similar potential distributions at the scalp.

(E) In a fifth simulation, the influence of the dipole localisation was investigated. Deeper sources generate a weaker signal captured by the electrodes and are possibly less accurately separated from background EEG. We varied the z-coordinate of the dipole between 0 and 0.8. x and y were kept fixed at -0.5 and 0, respectively.

(F) 21-channel EEG does not have an optimal spatial resolution due to the low spatial sampling. In a last simulation, we investigated how much the dipole localisation error could be improved by using dense array EEG [45]. We used 148 electrodes, uniformly distributed over the realistic domain of the same spherical head model.

2.2.2. Noise

A 500-by-21 noise matrix **B** contained 2 seconds of awake background EEG activity, recorded with the same electrode configuration as in (A), from a normal subject. On this matrix **B**, muscle artifacts were superimposed. These muscle artifacts were separated from contaminated background activity using BSS-CCA [46]. For the last simulation with dense-array EEG, the noise was Gaussian, because no background EEG was available with this high number of electrodes.

2.2.3. The simulated signal

In the simulation study the noise matrix **B** is superimposed on the signal matrix **A** containing the epileptical activity:

$$\mathbf{X}(\lambda) = \mathbf{A} + \lambda \cdot \mathbf{B} \tag{9}$$

with $\lambda \in \mathbb{R}$. The root mean-squared (RMS) value of the signal is then equal to

$$\mathrm{RMS}(\mathbf{A}) = \sqrt{\frac{1}{S \cdot N} \sum_{s=1}^{S} \sum_{n=0}^{N-1} \left(\mathbf{A}(n,s)\right)^2} \tag{10}$$

with N the number of time samples; and the RMS value of the noise is equal to

$$\mathrm{RMS}(\lambda \cdot \mathbf{B}) = \sqrt{\frac{1}{S \cdot N} \sum_{s=1}^{S} \sum_{n=0}^{N-1} \left(\lambda \cdot \mathbf{B}(n,s)\right)^2}. \tag{11}$$

The signal-to-noise ratio (SNR) is then defined as follows:

$$\mathrm{SNR} = \frac{\mathrm{RMS}(\mathbf{A})}{\mathrm{RMS}(\lambda \cdot \mathbf{B})}. \tag{12}$$

Changing the parameter λ alters the noise level of our simulated signal.

3. RESULTS

Figure 6(a) shows the dipole localisation error in function of the SNR when one dipole was fitted on the potential distribution extracted with Candecomp. At an SNR of 0.4, the localisation error became smaller than 1 cm and at an SNR of 0.7, the error between the simulated and the fitted dipole was only 5 mm. At SNRs lower than 0.26, there was no atom that clearly corresponded to the seizure activity as can be seen by the large localisation error. At higher noise levels, one atom contained pure rhythmical activity as can be seen by the sudden improvement in dipole localisation error. Figure 6(b) shows the dipole fit error when a sawtooth was used to simulate ictal EEG. The error was slightly larger compared to the perfect sinusoidal signal, but still in the same range.

Figure 7 shows the dipole localisation error for different frequencies of the simulated epileptic signal at an SNR of 0.7 (see Figure 5(b)). From this figure, it can be seen that the accuracy of the separation of ictal EEG and the dipole fit does not depend on the frequency of the signal. At all frequencies, a dipole is fitted with an error smaller than 1 cm. This means that even when the frequency content of ictal activity overlaps with frequency content of muscle artifacts, a good separation is obtained.

Figure 8 shows the dipole localisation error in function of the SNR when the simulated epileptic signal changed in frequency and amplitude during the considered 2 seconds. The figure strongly resembles Figure 6(a). This means that, although the signal is not well localised in frequency, the decomposition still reliably detects the correct location. This does not mean that the seizure activity is fully separated into one atom. When we looked at the frequency component

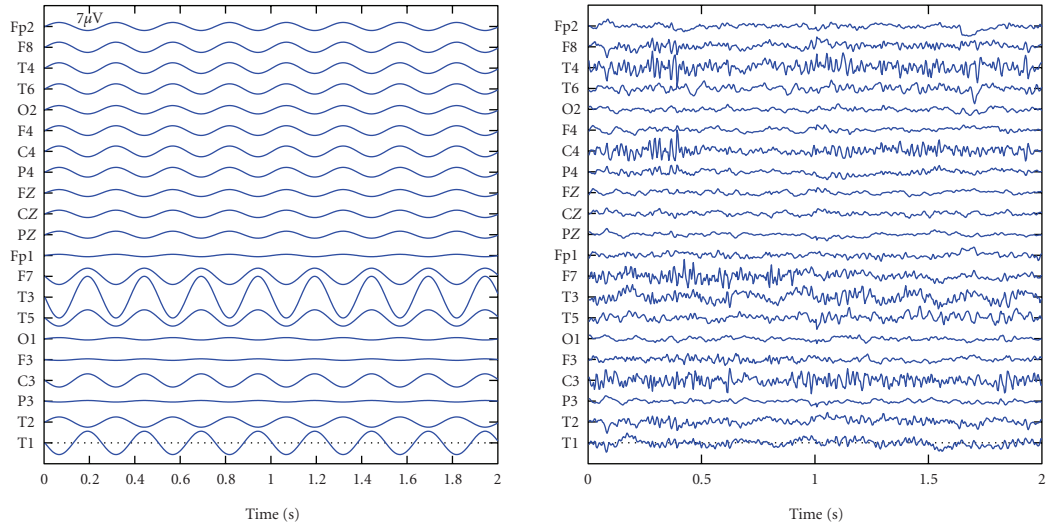


Figure 5: Simulated data. (a) The time course of the scalp potentials reflecting the 4 Hz epileptiform activity on each electrode. (b) The simulated data matrix for an SNR equal to 0.7.

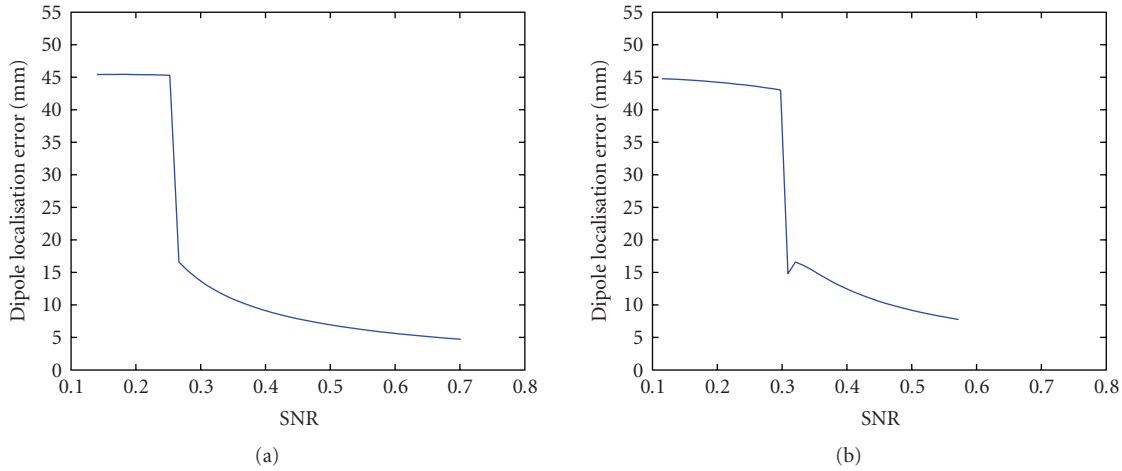


Figure 6: (a) The dipole localisation error in function of the noise level when a sinus waveform was used as epileptic signal. (b) Idem as (a) but a sharp wave was used as epileptic signal.

of the epileptic atom, this component had maximal values around 6 Hz, that is, the average of the start (8 Hz) and end frequency (4 Hz), while the frequency component in the first simulation peaked around 4 Hz. When the epileptic atom is reconstructed (see Figure 9), the change in frequency is not captured and the reconstruction is poor in the beginning and at the end. This is also reflected by a lower Candecomp fit percentage. In the first simulation, the fit percentage was about 75%, while in this simulation only 58% of the activity could be modeled. However, the best trilinear approximation captures a good localisation.

Figure 10 shows in (a) the simulated localisation of two close dipoles and in (b) the estimated localisation with the proposed method at an SNR of 0.7. The Corcondia [30] indicated that three atoms were the correct number of atoms for this simulated EEG. Two of them corresponded to the 2 dipolar foci. The localisation error was for both sources about 5 mm, which indicates that a reliable separation and localisation was obtained.

The dipole localisation error as a function of the position of the dipole is shown in Figure 11.

The last figure, Figure 12, shows the dipole estimation error when 148 electrodes are used to acquire the EEG. It can be seen that with a high spatial sampling, the estimation accuracy became about 1 mm.

4. DISCUSSION

In [24], we introduced an automatic, fast, and sensitive method for visualizing the ictal onset zone. The method was

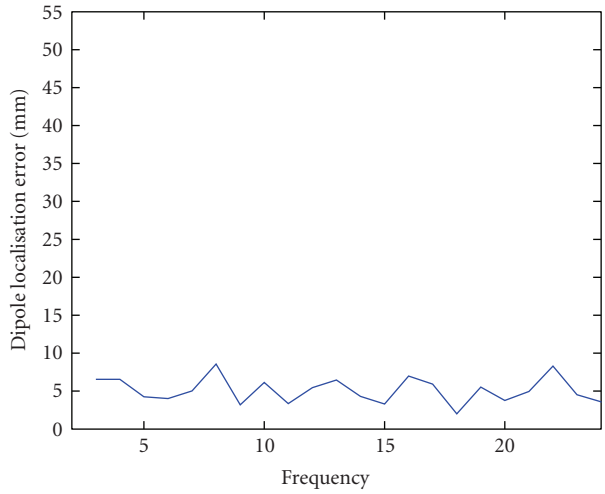


Figure 7: The dipole localisation error as a function of the seizure frequency.

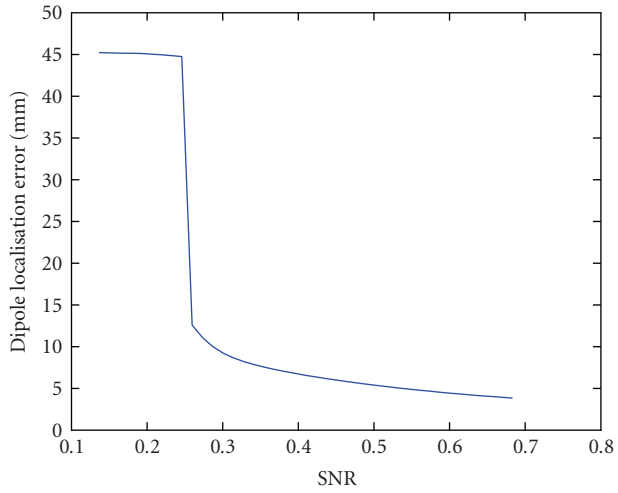


Figure 8: The dipole localisation error as a function of the noise level, when the seizure activity is changed in frequency during the time interval under investigation.

based on the multiway Candecomp of wavelet-transformed EEG in distinct "atoms." After the decomposition, one atom could be identified as the epileptical atom, and the spatial component of this atom revealed the focus. The method was also validated on a large number of in vivo seizures, and was not influenced by the presence of strong artifacts. However, in that study, the extracted localising information was limited to the 2D potential distribution of epileptic activity over the electrodes. In the present study, we looked at the 3D localisation in a spherical head, and investigated the localising accuracy of a dipolar source fitted to the extracted potential distribution.

It is known that an infinite number of internal electrical currents correspond with exactly the same potential distribution on the scalp. The discussion if dipolar sources are superior to distributed sources is beyond the scope of this study. We chose the dipolar source because it is most popular. It is known that the generator of ictal activity can be an extended area, and that a dipole situated in a certain region should be considered as the center of mass of a larger activated brain region [7]. In [25], source densities were computed after Candecomp.

We present here the framework for seizure onset localisation with Candecomp as preprocessing step for EEG source localisation. In fact, we focussed in the paper on seizure activity. However, the method can also be used to localise all origins of oscillatory activity. We have shown that in a spherical head model with realistically simulated EEG, our algorithm correctly localised the seizure-related atom with an accuracy of about 5 mm, even at SNR ratios that are lower than one encounters during real ictal recordings. SNRs below 1 mean that the signal contains more noise than signal (see, e.g., Figure 5). Although the shape of seizure activity will not be perfectly sinusoidal, we have shown that the exact shape of the seizure signal did not really influence the localisation accuracy. In a second simulation, we have shown that the localisation error does not depend on the frequency of the epileptic signal, and that overlapping frequency content of signal and noise, representing muscle artifacts, does not lower the reliability of the decomposition. The third simulation investigated a more challenging, but maybe more realistic situation in which the frequency of the seizure changed during the considered time interval. The resulting atom could not fully capture the exact frequency-varying signal, as indicated by a lower fit-percentage of Candecomp and the reconstructed epileptic signal. However, the best trilinear approximation still reliably localised the signal. We should emphasize that Candecomp is an interesting decomposition method due to its uniqueness properties. However, the trilinear decomposition in space-time-frequency components really restricts the activity that can be fully captured. When the frequency content changes in time at a fixed position, the exact signal will not be fully separated. However, the best trilinear approximation will separate a rhythmical signal at the correct location. A similar result is observed when a moving dipole was simulated. Moving activity cannot be captured with a trilinear model, but the best approximation will result in an "average" localisation. The fourth simulation showed that the localisation error is quite insensitive to dipole localisation. In [47], it was observed that dipoles closer located to the scalp, are slightly better estimated due the higher SNR associated with higher dipoles. However, in our simulation this effect is negligible. We investigated also the situation in which two dipolar sources generating the same signal were placed near each other. This simulates multifocal epilepsy. The Corcondia [30] indicated that three atoms were the correct number of atoms for this simulated EEG. Two of them corresponded to the 2 dipolar sources. This example illustrates the interesting uniqueness property of Candecomp [30] for EEG source localisation. When matrix decomposition techniques like SVD or independent component analysis (ICA) would have been used to decompose the EEG, only 1 rhythmical source would be extracted as the 2 simulated sources are not independent nor uncorrelated. It would then not be obvious to determine the correct number of dipoles. In our approach, Candecomp

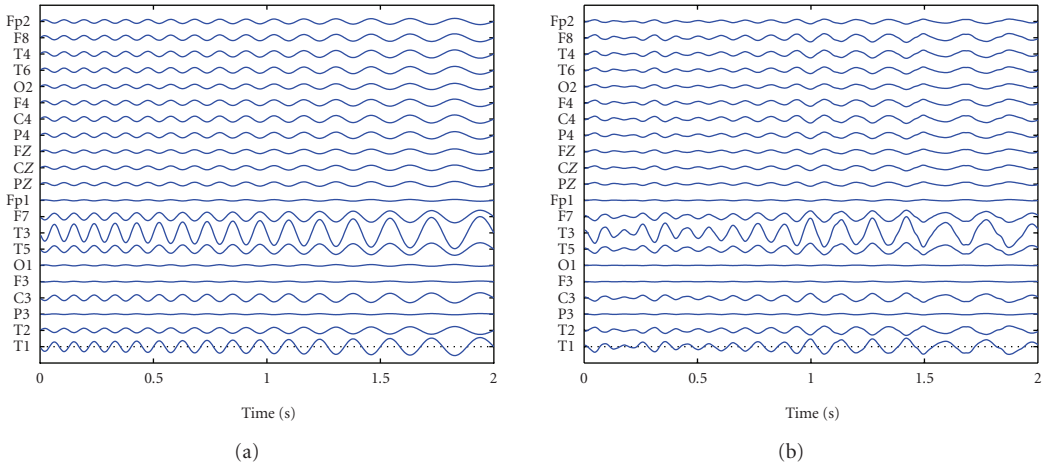


Figure 9: (a) The simulated frequency-modulated signal used in Figure 8 (b) The reconstructed atom.

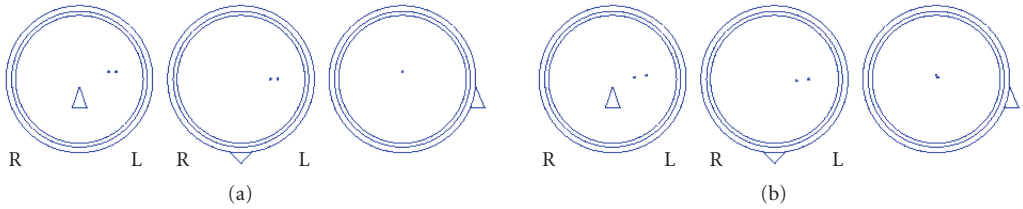


Figure 10: (a) The original dipole localisation of two simulated dipoles. (b) The dipole localisation when three atoms were estimated with Candecomp.

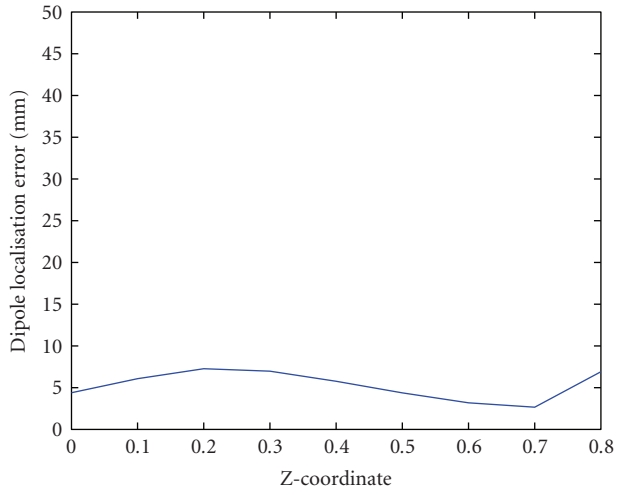


Figure 11: The dipole localisation error as a function of the z-coordinate of the dipole, in order to assess the influence of the depth of the dipole location.

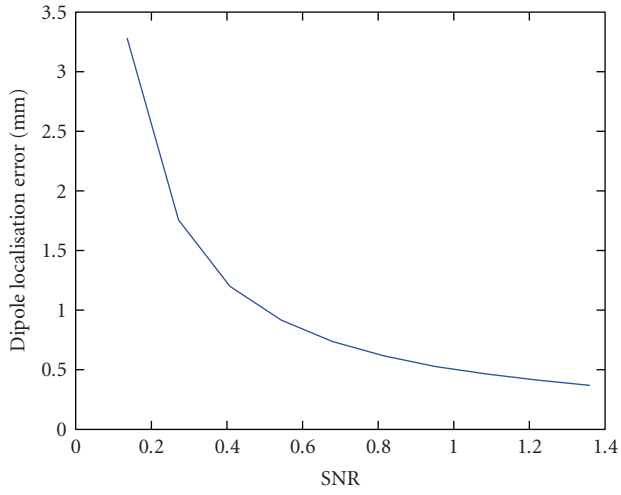


Figure 12: The dipole localisation error as a function of the noise level, when the EEG is recorded with 148 electrodes.

determines the optimal number of components and only 1 dipole will correspond to each atom. Tensor decomposition techniques offer clearly advantages over matrix decomposition techniques as preprocessing technique for EEG source localisation. The last simulation assessed the accuracy when more electrodes are used. It is known that dipole localisation based on 21 electrode measurements gives only an approximate indication of source localisation. However, using 148 electrodes can reduce the dipole estimation error to less than 1 mm at the same low SNR's. So we think it is worth to record the EEG with denser spatial sampling.

The current simulation study is the most reliable validation of our method. In the future, we plan to validate our method on in vivo seizures with a gold standard. This gold standard can be intracranial EEG, ictal SPECT, or the site of epilepsy surgery in patients who were rendered seizure free.

Comparing the estimated dipole localisation to other data, like ictal SPECT or MR-visible lesions, however, will be biased by the accuracy of the onset delineation with these diagnostic tools. We anticipate that the higher sensitivity and objectivity of our Candecomp method as compared with visual assessment of the ictal EEG's will improve and streamline the noninvasive presurgical evaluation of patients with refractory partial epilepsy.

ACKNOWLEDGMENTS

We would like to thank Guido Van Driel, Wim De Clercq, and Anneleen Vergult for helpful discussions on epilepsy and EEG. We would like to thank M. A. Kulesh for providing useful material and information on the computation of the inverse continuous wavelet transform. This research is funded by a PhD Grant of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). Research supported by Research Council KUL: GOA-AMBioRICS, CoE EF/05/006 Optimization in Engineering, IDO 05/010 EEG-fMRI; Flemish Government: FWO: projects, G.0407.02 (support vector machines), G.0360.05 (EEG, Epileptic), G.0519.06 (Noninvasive brain oxygenation), FWO-G.0321.06 (Tensors/Spectral Analysis), G.0341.07 (Data fusion), research communities (ICCoS, ANMMM); IWT: PhD Grants; Belgian Federal Science Policy Office IUAP P6/04 ("Dynamical systems, control and optimization," 2007–2011); EU: BIOPATTERN (FP6-2002-IST 508803), ETUMOUR (FP6-2002-LIFESCIHEALTH 503094), Healthagents (IST-2004-27214), FAST (FP6-MC-RTN-035801); ESA: Cardiovascular Control (Prodex-8 C90242).

REFERENCES

[1] J. Engel Jr., "Update on surgical treatment of the epilepsies: summary of the second international palm desert conference on the surgical treatment of the epilepsies (1992)," *Neurology*, vol. 43, no. 8, pp. 1612–1617, 1993.

[2] F. Rosenow and H. Lüders, "Presurgical evaluation of epilepsy," *Brain*, vol. 124, no. 9, pp. 1683–1700, 2001.

[3] S. Baillet, J. C. Mosher, and R. M. Leahy, "Electromagnetic brain mapping," *IEEE Signal Processing Magazine*, vol. 18, no. 6, pp. 14–30, 2001.

[4] C. Michel, M. Murray, G. Lantz, S. Gonzalez, L. Spinelli, and R. Grave de Peralta, "EEG source imaging," *Clinical Neurophysiology*, vol. 115, no. 10, pp. 2195–2222, 2004.

[5] R. Pascual-Marqui, C. Michel, and D. Lehmann, "Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain," *International Journal of Psychophysiology*, vol. 18, no. 1, pp. 49–65, 1994.

[6] M. Scherg and D. Von Cramon, "Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model," *Electroencephalography and Clinical Neurophysiology*, vol. 62, no. 1, pp. 32–44, 1985.

[7] I. Merlet and J. Gotman, "Reliability of dipole models of epileptic spikes," *Clinical Neurophysiology*, vol. 110, no. 6, pp. 1013–1028, 1999.

[8] K. Kobayashi, H. Yoshinaga, Y. Ohtsuka, and J. Gotman, "Dipole modeling of epileptic spikes can be accurate or misleading," *Epilepsia*, vol. 46, no. 3, pp. 397–408, 2005.

[9] J. Gotman, "Noninvasive methods for evaluating the localization and propagation of epileptic activity," *Epilepsia*, vol. 44, supplement 12, pp. 21–29, 2003.

[10] A. Vergult, W. De Clercq, A. Palmini, et al., "Improving the interpretation of ictal scalp EEG: BSS-CCA algorithm for muscle artifact removal," *Epilepsia*, vol. 48, no. 5, pp. 950–958, 2007.

[11] T. Krings, K. H. Chiappa, B. N. Cuffin, B. R. Buchbinder, and G. R. Cosgrove, "Accuracy of electroencephalographic dipole localization of epileptiform activities associated with focal brain lesions," *Annals of Neurology*, vol. 44, no. 1, pp. 76–86, 1998.

[12] T. B. J. Wiederin, K. H. Chiappa, T. Krings, et al., "The utility of dipole source analysis of seizure onsets in the localization of epileptogenic zones as assessed by postsurgical outcome," *Journal of Contemporary Neurology*, vol. 1A, pp. 2–11, 1999.

[13] S. Mine, H. Iwasa, Y. Kasagi, and A. Yamaura, "Ictal dipole source analysis based on a realistic scalp-skull-brain head model in localizing the epileptogenic zone," *Neuroscience Research*, vol. 51, no. 4, pp. 453–461, 2005.

[14] J. S. Ebersole, "Noninvasive localization of epileptogenic foci by EEG source modeling," *Epilepsia*, vol. 41, supplement 3, pp. S24–S33, 2000.

[15] P. Boon, M. D'Havé, B. Vanrumste, et al., "Ictal source localization in presurgical patients with refractory epilepsy," *Journal of Clinical Neurophysiology*, vol. 19, no. 5, pp. 461–468, 2002.

[16] J. Gotman, "Automatic recognition of epileptic seizures in the EEG," *Electroencephalography and Clinical Neurophysiology*, vol. 54, no. 5, pp. 530–540, 1982.

[17] D. W. Klass, "The continuing challenge of artifacts in the EEG," *American Journal of EEG Technology*, vol. 35, no. 4, pp. 239–269, 1995.

[18] L. Ding, G. A. Worrell, T. D. Lagerlund, and B. He, "Ictal source analysis: localization and imaging of causal interactions in humans," *NeuroImage*, vol. 34, no. 2, pp. 575–586, 2007.

[19] H. Hallez, A. Vergult, R. Phlypo, et al., "Muscle and eye movement artifact removal prior to EEG source localization," in *Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS '06)*, pp. 1002–1005, New York, NY, USA, August 2006.

[20] E. Urrestarazu, J. Iriarte, M. Alegre, M. Valencia, C. Viteri, and J. Artieda, "Independent component analysis removing artifacts in ictal recordings," *Epilepsia*, vol. 45, no. 9, pp. 1071–1078, 2004.

[21] P. LeVan, E. Urrestarazu, and J. Gotman, "A system for automatic artifact removal in ictal scalp EEG based on independent component analysis and Bayesian classification," *Clinical Neurophysiology*, vol. 117, no. 4, pp. 912–927, 2006.

[22] L. Zhukov, D. Weinstein, and C. Johnson, "Independent component analysis for EEG source localization," *IEEE Transactions on Engineering in Medicine and Biology Magazine*, vol. 19, no. 3, pp. 87–96, 2000.

[23] K. Kobayashi, I. Merlet, and J. Gotman, "Separation of spikes from background by independent component analysis with dipole modeling and comparison to intracranial recording," *Clinical Neurophysiology*, vol. 112, no. 3, pp. 405–413, 2001.

[24] M. De Vos, A. Vergult, L. De Lathauwer, et al., "Canonical decomposition of ictal scalp EEG reliably detects the seizure onset zone," *NeuroImage*, vol. 37, no. 3, pp. 844–854, 2007.

[25] F. Miwakeichi, E. Martínez-Montes, P. A. Valdés-Sosa, N. Nishiyama, H. Mizuhara, and Y. Yamaguchi, "Decomposing EEG data into space-time-frequency components using parallel factor analysis," *NeuroImage*, vol. 22, no. 3, pp. 1035–1045, 2004.

[26] E. Martínez-Montes, P. A. Valdés-Sosa, F. Miwakeichi, R. I. Goldman, and M. S. Cohen, "Concurrent EEG/fMRI analysis by multiway partial least squares," *NeuroImage*, vol. 22, no. 3, pp. 1023–1034, 2004.

[27] F. L. Hitchcock, "The expression of a tensor or a polyadic as a sum of products," *Journal of Mathematical Physics*, vol. 6, pp. 164–189.

[28] J. D. Carroll and J.-J. Chang, "Analysis of individual differences in multidimensional scaling via an *n*-way generalization of "Eckart-Young" decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, 1970.

[29] R. A. Harshman, "Foundations of the parafac procedure: models and conditions for an 'explanation' multi-modal factor analysis," UCLA Working Papers in Phonetics, 16, 1970.

[30] A. Smilde, R. Bro, and P. Geladi, *Multi-Way Analysis with Applications in the Chemical Sciences*, John Wiley & Sons, New York, NY, USA, 2004.

[31] J. B. Kruskal, "Three-way arrays: rank and uniqueness of trilinear decomposition with applications to arithmetic complexity and statistics," *Linear Algebra and Its Applications*, vol. 18, no. 2, pp. 95–138, 1977.

[32] N. D. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of *n*-way arrays," *Journal of Chemometrics*, vol. 14, no. 3, pp. 229–239, 2000.

[33] A. Stegeman and N. D. Sidiropoulos, "On Kruskal's uniqueness condition for the Candecomp/Parafac decomposition," *Linear Algebra and Its Applications*, vol. 420, no. 2-3, pp. 540–552, 2007.

[34] L. De Lathauwer, "A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization," *SIAM Journal on Matrix Analysis and Applications*, vol. 28, no. 3, pp. 642–666, 2006.

[35] P. Paatero, "The multilinear engine: a table-driven, least squares program for solving multilinear problems, including the *n*-way parallel factor analysis model," *Journal of Computational and Graphical Statistics*, vol. 8, no. 4, pp. 854–888, 1999.

[36] L. De Lathauwer, B. De Moor, and J. Vandewalle, "Computation of the canonical decomposition by means of a simultaneous generalized schur decomposition," *SIAM Journal on Matrix Analysis and Applications*, vol. 26, no. 2, pp. 295–327, 2004.

[37] S. A. Vorobyov, Y. Rong, N. D. Sidiropoulos, and A. B. Gershman, "Robust iterative fitting of multilinear models," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, part 1, pp. 2678–2689, 2005.

[38] E. Acar, C. Aykut-Bingol, H. Bingol, R. Bro, and B. Yener, "Multiway analysis of epilepsy tensors," *Bioinformatics*, vol. 23, no. 13, pp. i10–i18, 2007.

[39] M. A. Kulesh, M. S. Diallo, and M. Holschneider, "Wavelet analysis of ellipticity, dispersion, and dissipation properties of Rayleigh waves," *Acoustical Physics*, vol. 51, no. 4, pp. 425–434, 2005.

[40] Y. Salu, L. G. Cohen, D. Rose, S. Sato, C. Kufta, and M. Hallett, "An improved method for localizing electric brain dipoles," *IEEE Transactions on Biomedical Engineering*, vol. 37, no. 7, pp. 699–705, 1990.

[41] T. F. Oostendorp, J. Delbeke, and D. F. Stegeman, "The conductivity of the human skull: results of in vivo and in vitro measurements," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 11, pp. 1487–1492, 2000.

[42] B. N. Cuffin, D. Cohen, K. Yunokuchi, et al., "Tests of EEG localization accuracy using implanted sources in the human brain," *Annals of Neurology*, vol. 29, no. 2, pp. 132–138, 1991.

[43] M. R. Nuwer, G. Comi, R. Emerson, et al., "IFCN standards for digital recording of clinical EEG," *Electroencephalography and Clinical Neurophysiology*, vol. 106, no. 3, pp. 259–261, 1998.

[44] E. Niedermeyer, "Epilepstic seizure disorder," in *Electroencephalography: Basic Principels, Clinical Applications and Related Fields*, E. Niedermeyer and F. Lopes da Silva, Eds., chapter 27, Urban and Swarzenberg, Baltimore, Md, USA, 2nd edition, 1987.

[45] R. Oostenveld and P. Praamstra, "The five percent electrode system for high-resolution EEG and ERP measurements," *Clinical Neurophysiology*, vol. 112, no. 4, pp. 713–719, 2001.

[46] W. De Clercq, A. Vergult, B. Vanrumste, W. Van Paesschen, and S. Van Huffel, "Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 12, part 1, pp. 2583–2587, 2006.

[47] G. Van Hoey, B. Vanrumste, M. D'Havé, R. Van de Walle, I. Lemahieu, and P. Boon, "Influence of measurement noise and electrode mislocalisation on EEG dipole-source localisation," *Medical and Biological Engineering and Computing*, vol. 38, no. 3, pp. 287–296, 2000.

Hindawi Publishing Corporation
Computational Intelligence and Neuroscience
Volume 2007, Article ID 35021, 15 pages
doi:10.1155/2007/35021

Research Article

The Implicit Function as Squashing Time Model: A Novel Parallel Nonlinear EEG Analysis Technique Distinguishing Mild Cognitive Impairment and Alzheimer's Disease Subjects with High Degree of Accuracy

Massimo Buscema,[1] Massimiliano Capriotti,[1] Francesca Bergami,[1] Claudio Babiloni,[2,3,4] Paolo Rossini,[3,5,6] and Enzo Grossi[7]

[1] *Semeion Research Centre of Sciences of Communication, Via Sersale, 117, 00128 Rome, Italy*
[2] *Department of Human Physiology and Pharmacology, University of Rome La Sapienza, 00185 Rome, Italy*
[3] *Ospedale San Giovanni Calibita "Fatebenefratelli", Isola Tiberina, 00153 Rome, Italy*
[4] *Casa di cura San Raffaele Cassino (Frosinone), San Raffaele Pisana, Rome, Italy*
[5] *IRCCS Centro San Giovanni di Dio Fatebenefratelli, 25100 Brescia, Italy*
[6] *Department of Clinical Neurosciences, University of Rome Campus Biomedico, 00155 Rome, Italy*
[7] *Bracco SpA Medical Department, Via E. Folli, 50, 20134 Milan, Italy*

Correspondence should be addressed to Massimo Buscema, m.buscema@semeion.it

Received 19 December 2006; Revised 7 June 2007; Accepted 1 August 2007

Recommended by Saied Sanei

Objective. This paper presents the results obtained using a protocol based on special types of artificial neural networks (ANNs) assembled in a novel methodology able to compress the temporal sequence of electroencephalographic (EEG) data into spatial invariants for the automatic classification of mild cognitive impairment (MCI) and Alzheimer's disease (AD) subjects. With reference to the procedure reported in our previous study (2007), this protocol includes a new type of artificial organism, named TWIST. The working hypothesis was that compared to the results presented by the workgroup (2007); the new artificial organism TWIST could produce a better classification between AD and MCI. *Material and methods*. Resting eyes-closed EEG data were recorded in 180 AD patients and in 115 MCI subjects. The data inputs for the classification, instead of being the EEG data, were the weights of the connections within a nonlinear autoassociative ANN trained to generate the recorded data. The most relevant features were selected and coincidently the datasets were split in the two halves for the final binary classification (training and testing) performed by a supervised ANN. *Results*. The best results distinguishing between AD and MCI were equal to 94.10% and they are considerable better than the ones reported in our previous study (∼92%) (2007). *Conclusion*. The results confirm the working hypothesis that a correct automatic classification of MCI and AD subjects can be obtained by extracting spatial information content of the resting EEG voltage by ANNs and represent the basis for research aimed at integrating spatial and temporal information content of the EEG.

1. INTRODUCTION

The electroencephalogram (EEG), since its introduction, was considered the only methodology allowing a direct and online view of the "brain at work." At the same time, abnormalities of the "natural" aging of the brain have yet been noticed in different types of dementias. The introduction of different structural imaging technologies in the 1970's and 1980's (computed tomography and magnetic resonance imaging) and the good results in the study of brain function obtained with techniques dealing with regional metabolism, glucose and oxygen consumption, and blood flow (single-photon emission computed tomography, positron emission tomography, functional magnetic resonance imaging) during the following two decades closet the role of EEG in a secondary line, particularly in the evaluation of Alzheimer's dementia (AD) and related dementias.

Lately, EEG computerized analysis in aged people has been enriched by various modern techniques able to manage the large amount of information on time-frequency processes at single recording channels (wavelet, neural networks, etc.) and on spatial localization of these processes [2–10]. The results have encouraged the scientific community in exploring electromagnetic brain activity, which changes by aging and can greatly deteriorate, through the different stages of the various forms of dementias. The use of neural networks represents an alternative and very promising attempt to make EEG analysis suitable for clinical applications in aging—thanks to their ability in extracting specific and smooth characteristics from huge amounts of data. Computerized processing of a large quantity of numerical data in wakeful relaxed subjects ("resting" EEG) made easier the automatic classification of the EEG signals, providing promising results even using relatively simple linear classifiers such as logistic regression and discriminant analysis. Using global field power (i.e., the sum of the EEG spectral power across all electrodes) as an input, some authors reached an accurate differential diagnosis between AD and MCI subjects with accuraces of 84% and 78%, respectively[11, 12]. Using evaluation of spectral coherence between electrode pairs (i.e., a measure of the functional coupling) as an input to the classification, the correct classification reached 82% when comparing the AD and normal aged subjects [13, 14].

Spatial smoothness and temporal fluctuation of the EEG voltage are considered as measures of the synaptic impairment, along with the notion that cortical atrophy can affect the spatiotemporal pattern of neural synchronization generating the scalp EEG. These parameters have been used to successfully discriminate the respective distribution of probable AD and normal aged subjects [15]. The interesting new idea in that study [15] was the analysis of resting EEG potential distribution instant by instant rather than the extraction of a global index along periods of tens of seconds or more.

Table 1 summarizes the results of a higher preclassification rate with ANN's analysis than with standard linear techniques, such as multivariate discriminatory analysis or the nearest-neighbour analysis [16]. Some authors [17] developed a system consisting of recurrent neural nets processing spectral data in the EEG. They succeeded in classifying AD patients and non-AD patients with a sensitivity of 80% and a specificity of 100%. In other studies, classifiers based on ANNs, wavelets, and blind source separation (BSS) achieved promising results [18, 19]. In a study from the same workgroup of this paper, we used a sophisticated technique based on blind source separation and wavelet preprocessing developed by Vialatte et al. [18] and Cichocki et al. [20–22] recently, whose results appear to be the best in the field when compared to the literature. We named this method *BWB model* (blind source separation + wavelet + bumping modeling), [1]. The results obtained in the classifications tasks, comparing AD patients to MCI subjects, using the BWB model, ranged from 78.85% to 80.43% (mean = 79.48%).

The aim of this study is to assess the strength of a novel parallel nonlinear EEG analysis technique in the differential classification of MCI subjects and AD patients, with a high degree of accuracy, based on special types of artificial neural networks (ANNs) assembled in a novel methodology able to compress the temporal sequence of electroencephalographic (EEG) data into spatial invariants. The working hypothesis is that this new approach to EEG based on nonlinear ANNs-based methods can contribute to improving the reliance of the diagnostic phase in association with other clinical and instrumental procedures. Compared to the results already presented by the workgroup [1], the included new artificial organism TWIST could produce a better classification between AD and MCI.

2. MATERIAL AND METHODS

The IFAST method includes two phases.

(1) A squashing phase: an EEG track is compressed in order to project the *invariant patterns* of that track on the connections matrix of an autoassociated ANN. The EGG track/subject is now represented by a vector of weights, without any information about the target (AD or MCI).

(2) "TWIST" (training with input selection and testing) phase: a technique of data resampling based on the genetic algorithm GenD, developed at Semeion Research Center. The new dataset which is composed by the connections matrix (output of the squashing phase), plus the target assigned to each vector, is splitted into two sub samples, each one for five times with a similar probability density function, in order to train, test, and validate the ANN models.

2.1. The IFAST method

2.1.1. General philosophy

The core of this new methodology is that the ANNs do not classify subjects by directly using the EEG data as an input. Rather, the data inputs for the classification are the weights of the connections within a recirculation (nonsupervised) ANN trained to generate the recorded EEG data. These connection weights represent a model of the peculiar spatial features of the EEG patterns at the scalp surface. The classification, based on these weights, is performed by a standard supervised ANN.

This method, named IFAST (acronym for implicit function as squashing time), tries to understand the implicit function in a multivariate data series compressing the temporal sequence of data into spatial invariants and it is based on three general observations.

(1) Every multivariate sequence of signals coming from the same natural source is a complex asynchronous dynamic highly nonlinear system, in which each channel's behavior is understandable only in relation to all the others.

(2) Given a multivariate sequence of signals generated from the same source, the implicit function defining the above-mentioned asynchronous process is the conversion of that same process into a complex

Table 1: EEG automatic classification (* = severe AD ** = mild AD; S. no. = Sample; N. aged = normal aged; ANN = artificial neural networks; LDA = linear discriminant analysis; ACC = accuracy (%); SE = sensibility; SP = specificity).

Author year	S. no.	AD	N. aged	MCI	Length (s)	Classificators		ACC	SE	SP
						ANN	LDA			
Pritchard et al. (1994)	39	14	25		nd	x	x	85	nd	nd
Besthorn et al. (1997)	nd	nd	nd		nd	x	x	86.60		
Huang et al. [6, 11]	93	38	24	31	nd		x	81	84	78
Knott et al. (2001)	65	35	30		nd		x	75		
Petrosian et al. [17]	20	10	10		120	x		90	80	100
Cichocki et al. [20]	60		38	22	20		x	78.25	73	84
Melissant et al. [16]	36	15*	21		40	x		94	93	95
Melissant et al. [16]	38	28**	10		40	x		82	64	100

hypersurface, representing the interaction in time of all the channels' behavior.

(3) The 19 channels in the EEG represent a dynamic system characterized by asynchronous parallelism. The nonlinear implicit function that defines them as a whole represents a metapattern that translates into space (hypersurface) that the interactions among all the channels create in time.

The idea underlying the IFAST method resides in thinking that each patient's 19-channel EEG track can be synthesized by the connection parameters of an autoassociated nonlinear ANN trained on the same track's data.

There can be several topologies and learning algorithms for such ANNs; what is necessary is that the selected ANN be of the autoassociated type (i.e., the input vector is the target for the output vector) and that the transfer functions defining it be*non linear and differentiable* at any point.

Furthermore, it is required that all the processing made on every patient be carried out with the same type of ANN, and that the initial randomly generated weights have to be *the same* in every learning trial. This means that, for every EEG, every ANN has to have the same starting point, even if that starting point is random.

We have operated in two ways in order to verify this method's efficiency.

(1) Different experiments were implemented based on the same samples. By "experiment," we mean a complete application of the whole procedure to every track of the sample.

(2) The second way is using autoassociated ANNs with different topologies and algorithms on the entire sample in order to prove that any autoassociated ANN can carry out the task of *translating into the space domain* the whole EEG track through its connections.

2.1.2. The squashing phase

The first application phase of the IFAST method may be defined as "*squashing*." It consists in compressing an EEG track in order to project the *invariant patterns* of that track on the connections of an auto-associated ANN.

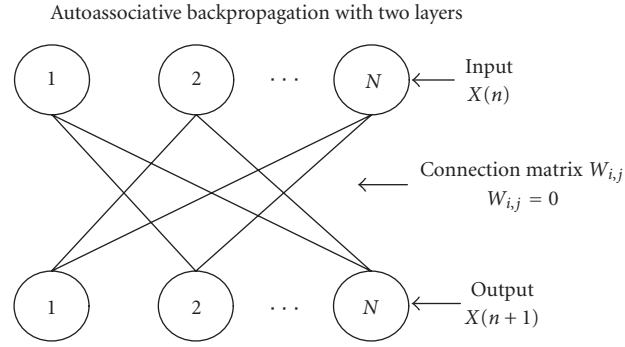


Figure 1: Autoassociative backpropagation ANN with $W_{j,j} = 0$, as the connections on the main diagonal are not present.

More formally

if

$F_i()$ = implicit function of the i-th EEG track

X_i = matrix of the values of the i-th EEG

$W^*_{i_{j,k}}$ = trained matrix of the connections of the i-th EEG ($*$ = objective of the squashing)

$W_{0_{j,k}}$ = random starting matrix, the same for all EEGs then in the case of a two-layered autoassociated ANN

$X_i = F_i(X_i, W^*_{i_{j,k}}, W_{0_{j,k}})$; con $W_{0_{j,j}} = 0$.

$W_{i_{j,j}} = 0$ means that every ith EEG track is processed by the two-layered autoassociated ANN in which $W_{j,j} = 0$, as the connections on the main diagonal are not present (see Figure 1).

It is possible to use different types of autoassociated ANNs to run this search for spatial invariants in every EEG.

(1) A backpropagation without a hidden unit layer and without connections on the main diagonal (for short, AutoBp):

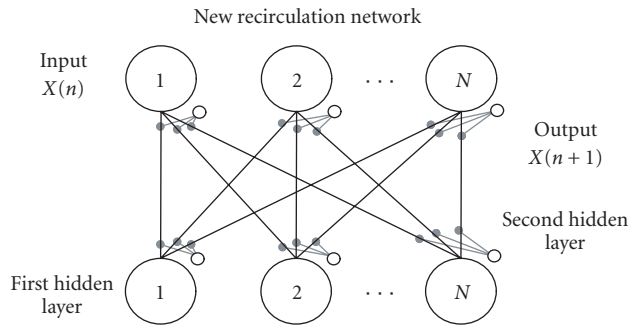


Figure 2: New recirculation network (NRC), with one connection matrix and four layers of nodes: one input layer, one output layer, and two layers of hidden nodes.

This is an ANN featuring an extremely simple learning algorithm:

$$\begin{aligned}
\text{Output}_i &= f\left(\sum_j^N \text{Input}_j \cdot W_{i,j} + \text{Bias}_i\right) \\
&= \frac{1}{1+e^{-\left(\sum_j^N \text{Input}_j \cdot W_{i,j} + \text{Bias}_i\right)}}, \qquad W_{i,i} = 0; \\
\delta_i &= (\text{Input}_i - \text{Output}_i) \cdot f'(\text{Output}_i) \\
&= (\text{Input}_i - \text{Output}_i) \cdot \text{Output}_i \cdot (1 - \text{Output}_i); \\
\Delta W_{i,j} &= \text{LCoef} \cdot \delta_i \cdot \text{Input}_j, \quad \text{LCoef} \in [0,1], \\
\Delta \text{Bias}_i &= \text{LCoef} \cdot \delta_i.
\end{aligned} \tag{1}$$

AutoBP is an ANN featuring $N^2 - N$ internode connections and N bias inside every exit node, for a total of N^2 adaptive weights. This algorithm works similarly to logistic regression and can be used to establish the dependency of variables from each others.

The advantage of AutoBP is due to its learning speed, in turn due to the simplicity of its topology and algorithm. Moreover, at the end of the learning phase, the connections between variables, being direct, have a clear conceptual meaning. Every connection indicates a relationship of faded excitement, inhibition, or indifference between every pair of channels in the EEG track of any patient.

The disadvantage of AutoBP is its limited convergence capacity, due to that same topological simplicity. That is to say, complex relationships between variables may be approximated or ignored (for details, see [23, 24]).

(2) New recirculation network (for short, NRC) is an original variation [25] of an ANN that has existed in the literature [26] and was not considered to be useful to the issue of autoassociating between variables.

The topology of the NRC which we designed includes only one connection matrix and four layers of nodes: one input layer, corresponding to the number of variables; one output layer whose target is the input vector; two layers of hidden nodes with the same cardinality independent from the cardinality of the input and output layers. The matrix between input-output nodes and hidden nodes is fully connected and in every learning cycle, it is modified both ways, according to the following equations:

$$\begin{aligned}
\text{Hidden1}_i &= f\left(\sum_j^N \text{Input}_j \cdot W_{i,j} + \text{BiasHidden}_i\right) \\
&= f\left(\text{Net}_i^{\text{Hidden1}}\right) = \frac{1}{1+e^{-\text{Net}_i^{H1}}}; \\
\text{Output}_j &= R \cdot \text{Input}_j + (1-R) \\
&\quad \cdot f\left(\sum_i^M \text{Hidden1}_i \cdot W_{j,i} + \text{BiasOutput}_j\right) \\
&= R \cdot \text{Input}_j + (1-R) \cdot f\left(\text{Net}_j^{\text{Output}}\right) \\
&= R \cdot \text{Input}_j + (1-R) \cdot \frac{1}{1+e^{-\text{Net}_j^{\text{Output}}}}; \\
R &\in [0,1] / {}^* \text{Projection Coefficient}^* / \\
\text{Hidden2}_i &= R \cdot \text{Hidden1}_i + (1-R) \\
&\quad \cdot f\left(\sum_j^N \text{Output}_j \cdot W_{i,j} + \text{BiasHidden}_i\right) \\
&= R \cdot \text{Hidden1}_i + (1-R) \cdot f\left(\text{Net}_i^{\text{Hidden2}}\right) \\
&= R \cdot \text{Hidden2}_i + (1-R) \cdot \frac{1}{1+e^{-\text{Net}_i^{\text{Hidden2}}}}; \\
\Delta W_{j,i} &= \text{LCoef} \cdot (\text{Input}_j - \text{Output}_j) \cdot \text{Hidden1}_i; \\
\Delta \text{BiasOutput}_j &= \text{LCoef} \cdot (\text{Input}_j - \text{Output}_j); \\
\text{LCoef} &\in [0,1] / {}^* \text{Learning Coefficient}^* / \\
\Delta W_{i \cdot i} &= \text{LCoef} \cdot (\text{Hidden1}_i - \text{Hidden2}_i) \cdot \text{Output}_j; \\
\Delta \text{BiasHidden}_i &= \text{LCoef} \cdot (\text{Hidden1}_i - \text{Hidden2}_i).
\end{aligned} \tag{2}$$

NRC then features N^2 internode adaptive connections and $2 \cdot N$ intranode adaptive connections (bias). The advantages of NRC are its excellent convergence ability on complex datasets and, as a result, an excellent ability to interpolate complex relations between variables.

The disadvantages mainly have to do with the vector codification that the hidden units run on the input vectors making the conceptual decoding of its trained connections difficult.

(3) Autoassociative multilayer perceptron (for short, AMLP) may be used with an auto-associative purpose (encoding)— thanks to its hidden units layer, that decomposes the input vector into main nonlinear components. The algorithm used to train the MLP is a typical backpropagation algorithm [27].

The MLP, with only one layer of hidden units, features two connection matrices and two intranode connection vectors (bias), according to the following definitions:

N = number of input variables
= number of output variables;
M = number of nodes in the hidden layer;

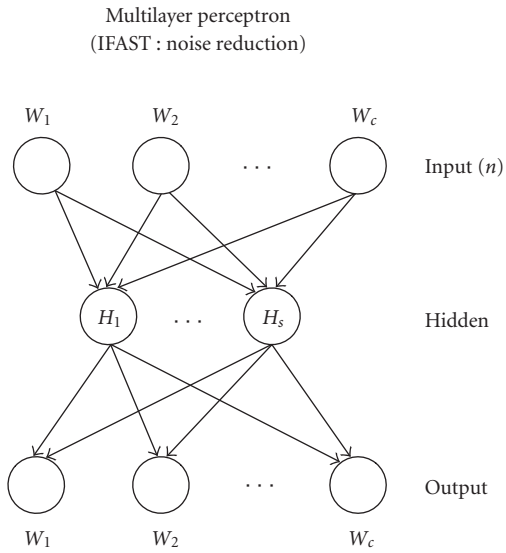


Figure 3: Multilayer perceptron; its hidden units layer decomposes the input vector into main nonlinear components.

C = total number of internode and intranode connections (bias);

$$C = 2 \cdot N \cdot M + N + M. \tag{3}$$

The advantages of MLP are its well-known flexibility and the strength of its backpropagation algorithm. Its disadvantages are the tendency to saturate the hidden nodes in the presence of nonstationary functions, and the vector codification (allocated) of the same hidden nodes.

(4) Elman's hidden recurrent [28] can be used for autoassociating purposes, again using the backpropagation algorithm (for short, autoassociative hidden recurrent AHR, see Figure 4). It was used in our experimentation as a variation for MLP with memory set to one step. It is not possible to call it a proper recurring ANN in this form, because the memory would have been limited to one record before. We used this variation only to give the ANN an input vector modulated at any cycle by the values of the previous input vector. Our purpose was not to codify the temporal dependence of the entrance signals, but rather to give the ANN a "smoother" and more mediated input sequence. The number of connections in the AHR BP is the same as an MLP with extended input, whose cardinality is equal to the number of hidden units:

$$C = 2 \cdot N \cdot M + N + M + M^2. \tag{4}$$

The software IFAST (developed in Borland C) [29] produces the squashing phase through the training operated by these four networks; in the "MetaTask" section the user can define the whole procedure by selecting

(i) the files that will be processed (in our case every complete EEG),

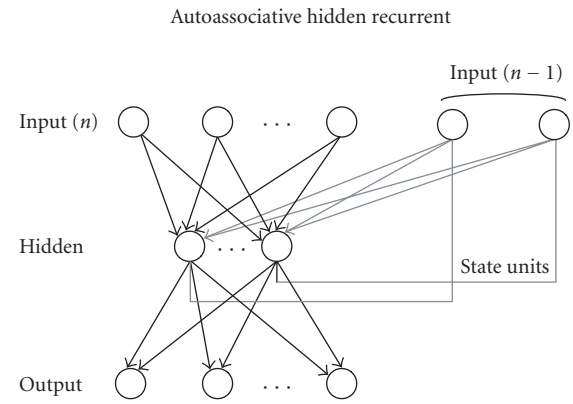


Figure 4: Elman's hidden recurrent ANN for auto-associating purposes using the backpropagation algorithm.

(ii) the type of network,
(iii) the sequence of the records for every file (generally random),
(iv) the number of epochs of training,
(v) a training stop criterion (number of epochs or minimum RMSE),
(vi) the number of hidden nodes of the autoassociated network, which determines the length of the output vector of the file processed
(vii) the number of matrices, depending on the type of the autoassociated network selected,
(viii) the learning coefficient and delta rate.

2.2. TWIST

From this phase, the procedure is completely different from the one described in our precedent work [1]. The choice of following a different methodology was due to the will of improving the classification results and removing causes of loss of information.

In the former study, the dataset coming from the squashing phase was compressed by another autoassociated ANN, in the attempt of eliminating the invariant pattern, codified from the previous ANN, relating to specific characteristic of the brain (anxiety level, background level, etc.) which is not useful for the classification, leaving the most significant ones unaltered. Then the new compressed datasets were split into two halves, (training and test) using T&T [30] evolutionary algorithm, for the final binary classification.

Rather in this work, the elimination of the noisiest features and the classification run parallel to each other. We will show that the new procedure has obtained better performances.

First of all, a new dataset called "Diagnostic DB" was created for easier understanding. The diagnostic gold standard has been established, for every patient, in a way that is completely independent of the clinical and instrumental examinations (magnetic resonance imaging, etc.) carried out by a group of experts whose diagnosis has been also reconfirmed in time.

The diagnoses have been divided into the following two classes, based on delineated inclusion criteria:

(a) elderly patients with "cognitive decline" (MCI);
(b) elderly patients with "probable Alzheimer" (AD);

We rewrote the last generated dataset, adding to every H_{n_s} vector the diagnostic class that an objective clinical examination had assigned to every patient. The H_{m_s} vectors represent the invariant traits s as defined by the squashing phase for every m-th subject EEG track, that is, the columns number of the connections matrix depending on the specific autoassociated network used.

Then the dataset is ready for the next step. This new phase is called TWIST [31] and includes the utilization of two systems T&T and IS [30], both based on a genetic algorithm, GenD, developed at Semeion Research Centre [32].

T&T systems are robust data resampling techniques able to arrange the source sample into subsamples, each one with a similar probability density function. In this way the data split into two or more subsamples in order to train, test, and validate the ANN models more effectively.

The IS system is an evolutionary system for feature selection based on a wrapper approach. While the filter approach looks at the inner properties of a dataset providing a selection that is independent of the classification algorithm to be used afterwards, in the wrapper approach various subsets of features are generated and evaluated using a specific classification model using its performances as a guidance to optimization of subsets.

The IS system reduces the amount of data while conserving the largest amount of information available in the dataset. The combined action of these two systems allows us to solve two frequent problems in managing artificial neural networks:

(1) the size and quality of the training and testing sets,
(2) the large number of variables which, apparently, seem to provide the largest possible amount of information. Some of the attributes may contain redundant information, which is included in other variables, or confused information (noise) or may not even contain any significant information at all and be completely irrelevant.

Genetic algorithms have been shown to be very effective as global search strategies when dealing with nonlinear and large problems.

The "*training and testing*" algorithm (T&T) is based on a population of n ANNs managed by an evolutionary system. In its simplest form, this algorithm reproduces several distribution models of the complete dataset D_Γ (one for every ANN of the population) in two subsets ($d_\Gamma^{[\mathrm{tr}]}$, the training set, and $d_\Gamma^{[\mathrm{ts}]}$, the testing set). During the learning process each ANN, according to its own data distribution model, is trained on the subsample $d_\Gamma^{[\mathrm{tr}]}$ and blind-validated on the subsample $d_\Gamma^{[\mathrm{ts}]}$.

The performance score reached by each ANN in the testing phase represents its "fitness" value (i.e., the individual probability of evolution). The genome of each "network individual" thus codifies a data distribution model with an associated validation strategy. The n data distribution models are combined according to their fitness criteria using an evolutionary algorithm. The selection of "network individuals" based on fitness determines the evolution of the population, that is, the progressive improvement of performance of each network until the optimal performance is reached, which is equivalent to the better division of the global dataset into subsets. The evolutionary algorithm mastering this process, named "genetic doping algorithm" (GenD for short), created at Semeion Research Centre, has similar characteristics to a genetic algorithm [33–37] but it is able to maintain an inner instability during the evolution, carrying out a natural increase of biodiversity and a continuous "evolution of the evolution" in the population.

The elaboration of T&T is articulated in two phases.

In a preliminary phase, an evaluation of the parameters of the fitness function that will be used on the global dataset is performed. The configuration of a standard backpropagation network that most "suits" the available dataset is determined: the number of layers and hidden units, some possible generalizations of the standard learning law, the fitness values of the population's individuals during evolution. The parameters thus determined define the configuration and the initialization of all the individual networks of the population and will then stay fixed in the following computational phase. The accuracy of the ANN performance with the testing set will be the fitness of that individual (i.e., of that hypothesis of distribution into two halves of the whole dataset).

In the computational phase, the system extracts from the global dataset the best training and testing sets. During this phase, the individual network of the population is running, according to the established configuration and the initialization parameters.

Parallel to T&T runs "*Input Selection*" (IS), an adaptive system, based on the same evolutionary algorithm GenD, consisting of a population of ANN, in which each one carries out a selection of the independent and relevant variables on the available database.

The elaboration of IS, as for T&T, is developed in two phases. In the preliminary phase, a standard backpropagation ANN is configured in order to avoid possible over fitting problems. In the computational phase, each individual network of the population, identified by the most relevant variables, is trained on the training set and tested on the testing set.

The evolution of the individual network of the population is based on the algorithm GenD. In the I.S. approach, the GenD genome is built by n binary values, where n is the cardinality of the original input space. Every gene indicates if an input variable is to be used or not during the evaluation of the population fitness. Through the evolutionary algorithm GenD, the different "hypotheses" of variable selection, generated by each ANN of the population, change over time, at each generation; this leads to the selection of the best combination of input variables. As in the T&T systems, the genetic operators crossover and mutation are applied on the ANNs population; the rates of occurrence for both operators

are self-determined by the system in an adaptive way at each generation.

When the evolutionary algorithm no longer improves its performance, the process stops, and the best selection of the input variables is employed on the testing subset.

The software based on TWIST phase algorithm (developed in *C*-Builder [31]) allows the configuration of the genetic algorithm GenD:

- the population (the number of individual networks),
- number of hidden nodes of the standard BP,
- number of epochs,
- the output function SoftMax,
- the cost function (classification rate in our case).

The generated outputs are the couple of files SetA and SetB (subsets of the initial db defined by the variables selected) that will be used in the validation protocol (see Section 2.3).

2.3. The validation protocol

The validation protocol is a fundamental procedure to verify the models' ability to generalize the results reached in the Testing phase of each model. The application of a fixed protocol measures the level of performance that a model can produce on data that are not present in the testing and/or training sample. We employed the so-called 5×2 cross-validation protocol (see Figure 6) [38]. This is a robust protocol that allows one to evaluate the allocation of classification errors. In this procedure, the study sample is randomly divided ten times into two subsamples, always different but containing a similar distribution of cases and controls.

The ANNs' good or excellent ability to diagnostically classify all patients in the sample from the results of the confusion matrices of these 10 independent experiments would indicate that the spatial invariants extracted and selected with our method truly relate to the functioning quality of the brains examined through their EEG.

2.4. Experimental setting

2.4.1. Subjects and diagnostic criteria

The population study included

(a) 180 AD patients (gender: 50 males/130 females; age: mean = 77 ± 6.78 SD, range from 54 to 91; MMSE: mean = 19.9, ± 4.89 SD, range from 5 to 30);
(b) 115 MCI subjects (gender: 49 males/66 females; age: mean = 76 ± 6.37 SD, range from 42 to 88; MMSE: mean = 25.2, ± 2.35 SD, range from 17.3 to 29).

The samples were matched for age, gender, and years of education. Part of the individual data sets was used for previous EEG studies [2–4]. In none of these studies we addressed the specific issue of the present study. Local institutional ethics committees approved the study. All experiments were performed with the informed and overt consent of each participant or caregiver.

The present inclusion and exclusion criteria for MCI were based on previous seminal studies [39–46] and designed for selecting elderly persons manifesting objective cognitive deficits, especially in the memory domain, who did not meet criteria for a diagnosis of dementia or AD, namely, with, (i) objective memory impairment on neuropsychological evaluation, as defined by performances ≥ 1.5 standard deviation below the mean value of age and education-matched controls for a test battery including memory rey list (immediate recall and delayed recall), Digit forward and Corsi forward tests; (ii) normal activities of daily living as documented by the patient's history and evidence of independent living; (iii) clinical dementia rating score of 0.5; (iv) geriatric depression scale scores < 13.

Exclusion criteria for MCI were: (i) mild AD, as diagnosed by the procedures described above; (ii) evidence of concomitant dementia such as frontotemporal, vascular dementia, reversible dementias (including pseudodepressive dementia), fluctuations in cognitive performance, and/or features of mixed dementias; (iii) evidence of concomitant extrapyramidal symptoms; (iv) clinical and indirect evidence of depression lower than 14 as revealed by GDS scores; (v) other psychiatric diseases, epilepsy, drug addiction, alcohol dependence, and use of psychoactive drugs including acetylcholinesterase inhibitors or other drugs enhancing brain cognitive functions; (vi) current or previous systemic diseases (including diabetes mellitus) or traumatic brain injuries.

Probable AD was diagnosed according to NINCDS-ADRDA criteria [47]. Patients underwent general medical, neurological, and psychiatric assessments and were also rated with a number of standardized diagnostic and severity instruments that included MMSE [48], clinical dementia rating scale [49], geriatric depression scale [50], Hachinski ischemic scale [51], and instrumental activities of daily living scale [52]. Neuroimaging diagnostic procedures (computed tomography or magnetic resonance imaging) and complete laboratory analyses were carried out to exclude other causes of progressive or reversible dementias, in order to have a homogenous probable AD patient sample. The exclusion criteria included, in particular, any evidence of (i) front temporal dementia diagnosed according to criteria of Lund and Manchester groups [53]; (ii) vascular dementia as diagnosed according to NINDS-AIREN criteria [54] and neuroimaging evaluation scores [55, 56]; (iii) extra pyramidal syndromes; (iv) reversible dementias (including pseudo dementia of depression); (v) Lewy body dementia according to the criteria by McKeith et al. [57]. It is important to note that benzodiazepines, antidepressant, and/or antihypertensive drugs were withdrawn for about 24 hours before the EEG recordings.

2.4.2. EEG recordings

EEG data were recorded in wake rest state (eyes-closed), usually during late morning hours from 19 electrodes positioned according to the international 10–20 system (i.e., Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2; 0.3–70 Hz filtering band passes). A specific reference electrode was not imposed to all recording units of this multi-centric study, since any further data

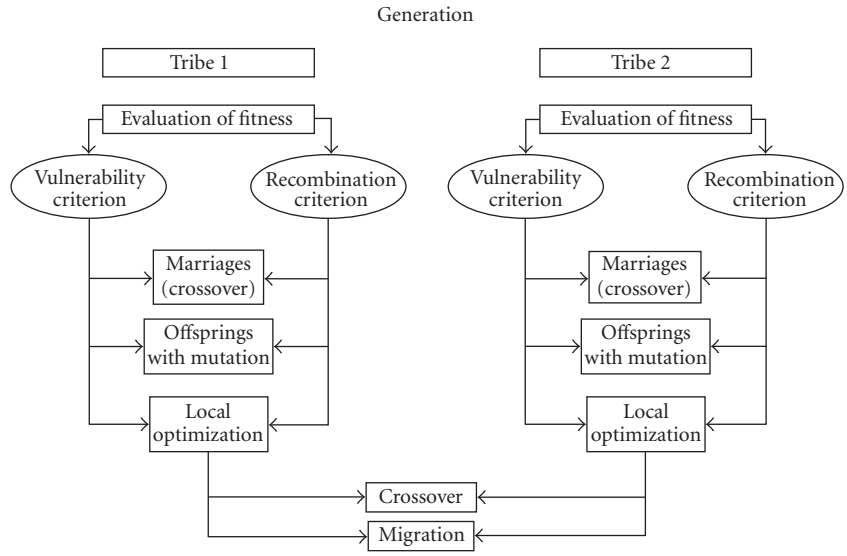


Figure 5: The structure and the operators of the evolutionary algorithm GenD.

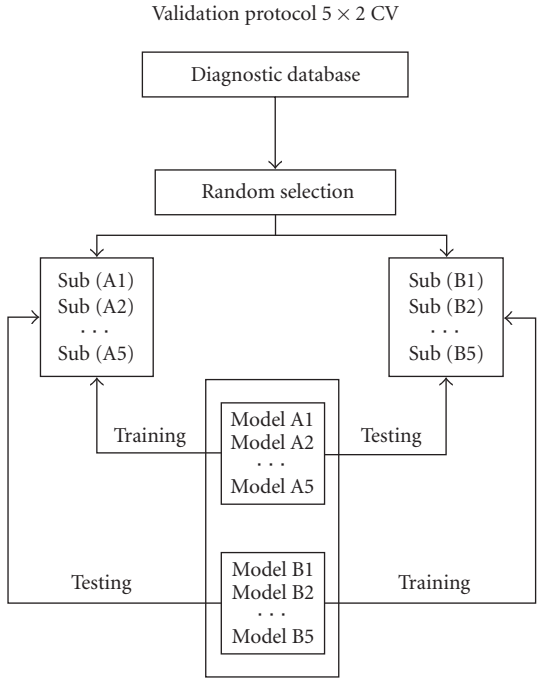


Figure 6: 5 × 2 validation protocol for the independent identification of the spatial invariants of EEGs.

analysis was carried out after EEG data were rereferenced to a common average reference. The horizontal and vertical electrooculogram was simultaneously recorded to monitor eye movements. An operator controlled, online, the subject and the EEG traces by alerting the subject any time there were signs of behavioural and/or EEG drowsiness in order to keep the level of vigilance constant. All data were digitized (5 minutes of EEG; 0.3–35 Hz band pass 128 Hz sampling rate).

The duration of the EEG recording (5 minutes) allowed the comparison of the present results with several previous AD studies using either EEG recording periods shorter than 5 minutes [58–62] or shorter than 1 minute [7, 8]. Longer resting EEG recordings in AD patients would have reduced data variability, but they would have increased the possibility of EEG "slowing" because of reduced vigilance and arousal.

EEG epochs with ocular, muscular, and other types of artefact were preliminarily identified by a computerized automatic procedure. Those manifesting sporadic blinking artefacts (less than 15% of the total) were corrected by an autoregressive method [63].

The performances of the software package on EOG-EEG-EMG data related to cognitive-motor tasks were evaluated with respect to the preliminary data analysis performed by two expert electroencephalographists (gold standard). Due to its extreme importance for multicentric EEG studies, we compared the performances of two representative "regression" methods for the EOG correction in time and frequency domains. The aim was the selection of the most suitable method in the perspective of a multicentric EEG study. The results showed an acceptable agreement of approximately 95% between the human and software behaviors, for the detection of vertical and horizontal EOG artifacts, the measurement of hand EMG responses for a cognitive-motor paradigm, the detection of involuntary mirror movements, and the detection of EEG artifacts. Furthermore, our results indicated a particular reliability of a "regression" EOG correction method operating in time domain (i.e., ordinary least squares). These results suggested the use of the software package for multicentric EEG studies.

Two independent experimenters—blind to the diagnosis— manually confirmed the EEG segments accepted for further analysis. A continuous segment of artefact-free EEG data lasting for 60 seconds was used for subsequent analyses for each subject.

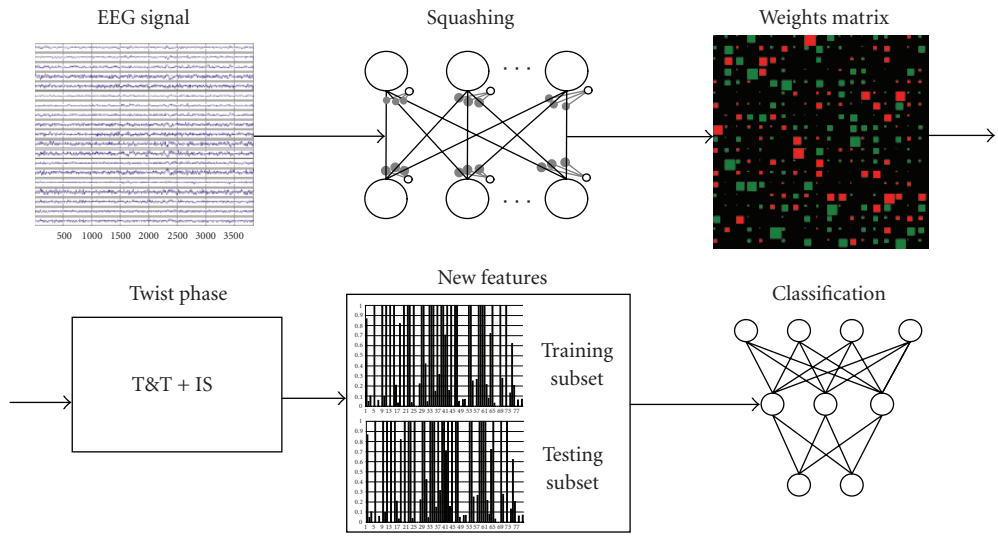


Figure 7: Procedure's scheme: from the squashing phase applied to EEG signal, the TWIST phase, to the final classification phase by ANNs.

2.4.3. *Preprocessing protocol*

The entire sample of 466 subjects was recorded at 128 Hz for 1 minute. The EEG track of each subject was represented by a matrix of 7680 sequential rows (time) and 19 columns (the 19 channels).

The *squashing phase* was implemented using the four autoassociative ANNs described [29]:

(a) an autoassociative BP with 2 layers (ABP);

(b) a new recirculation ANN (NRC);

(c) an autoassociative multilayer perceptron with 3 layers (AMLP);

(d) an autoassociative hidden recurrent (AHR).

Every autoassociative ANN independently processed every EEG of the total sample in order to assess the different capabilities of each ANN to extract the key information from the EEG tracks.

After this processing, each EEG track is squashed into the weights of every ANN resulting in 4 different and independent datasets (one for each ANN), whose records are the squashing of the original EEG tracks and whose variables are the trained weights of every ANN.

After TWIST processing, the most significant features for the classification were selected and at the same time the training set and the testing set with a similar function of probability distribution that provides the best results in the classification were defined.

The validation protocol 5x2CV was applied blindly to test the capabilities of a generic supervised ANN to correctly classify each record (the number of inputs depending on the number of variables selected by IS).

A supervised MLP was used for the classification task, without hidden units. In every experimentation, in fact, we were able to train perfectly the ANN in no more than 100 epochs (root mean square error (RMSE) < 0.0001). That means that in this last phase, we could have used also a linear classifier to reach up the same results.

3. RESULTS

The experimental design consisted in 10 different and independent processing for the classification AD versus MCI. Every experiment was conducted in a blind and independent manner in two directions: training with subsample A and blind testing with subsample B versus training with subsample B and blind testing with subsample A.

Table 3 shows the mean results summary for the classifications of AD versus MCI, compared to the results obtained in the experimentations reported in a previous study [1], based on a different protocol (without the TWIST phase).

Regarding the protocol IFAST-TWIST, the ABP and AHR achieved the best results comparing AD with MCI subjects (94.10% and 93.36%), but all the performances are considerably better than those obtained in the previous study.

Tables 4, 5, 6 and 7 show the details of the results obtained by each autoassociated ANN, where

SE = sensibility,

SP = specificity,

VP+ = positive predictive value,

VP− =negative predictive value,

LR+ = likelihood ratio for positive test results (benchmark value $\geq$ 2),

LR− = likelihood ratio for negative test results (benchmark value $\leq$ 0.2),

AUC = area under ROC curve (average ROC curve calculated by the threshold method),

Figures 8, 9, 10, and 11 show the respective average Roc curves.

Table 2: Autoassociative ANN types and parameters used during the processing.

ANN parameters type	AbP	NRC	AMLP	AHR
Number of inputs	19	19	19	19
Number of outputs	19	19	19	19
Number of state units	0	0	0	10
Number of hidden units	0	19	10	10
Number of weights	361	399	409	509
Number of epochs	200	200	200	200
Learning coefficient	0.1	0.1	0.1	0.1
Projection coefficient	Null	0.5	Null	Null

Table 3: Summary and comparison of AD results versus MCI.

Blind classification	AD versus MCI		
Type of input vector	Sensitivity	Specificity	Accuracy
ABP	90.73	97.46	94.1
NRC	89.27	93.32	91.29
AMLP	92.42	94.14	93.28
AHR	92.11	92.61	92.36

Table 4: Details of the ABP results.

	ABP results (%)									
ANN	SE	SP	A.MeanAcc.	W.MeanAcc.	Errors	VP+	VP−	LR+	LR−	AUC
FF_Bp(ab)	97.14	94.92	96.03	96.12	5	95.77	96.55	19.1	0.03	~ 0.98
FF_Bp(ba)	84.31	100	92.16	89.87	16	100	77.78	+ Inf	0.16	~ 0.928
Mean results	90.73	97.46	94.1	93	10.5	97.88	87.17	+ Inf	0.1	~ 0.948

* Average ROC curve calculated by the threshold method.

Table 5: Details of the NRC results.

	NRC results (%)									
ANN	SE	SP	A.MeanAcc.	W.MeanAcc.	Errors	VP+	VP−	LR+	LR−	AUC
FF_Bp(ab)	84.16	96.15	90.16	88.24	18	97.7	75.76	21.88	0.16	~ 0.898
FF_Bp(ba)	94.37	90.48	92.42	92.54	10	91.78	93.44	9.91	0.06	~ 0.932
Mean results	89.27	93.32	91.29	90.39	14	94.74	84.6	15.90	0.11	~ 0.926

Table 6: Details of the AMLP results.

	AMLP results (%)									
ANN	SE	SP	A.MeanAcc.	W.MeanAcc.	Errors	VP+	VP−	LR+	LR−	AUC
FF_Bp(ab)	93.26	92.19	92.72	92.81	6	94.32	90.77	11.94	0.07	~ 0.930
FF_Bp(ba)	91.57	96.08	93.82	93.28	7	97.44	87.5	23.35	0.09	~ 0.935
Mean results	92.42	94.14	93.28	93.05	6.5	95.88	89.14	17.65	0.08	~ .933

Table 7: Details of the AHR results.

	AHR results (%)									
ANN	SE	SP	A.MeanAcc.	W.MeanAcc.	Errors	VP+	VP−	LR+	LR−	AUC
FF_Bp(ab)	97.22	89.23	93.23	93.43	9	90.91	96.67	9.03	0.03	~ 0.940
FF_Bp(ba)	87	96	91.5	90	15	97.75	78.69	21.75	0.14	~ 0.904
Mean results	92.11	92.62	92.37	91.72	12	94.33	87.68	15.39	0.09	~ 0.926

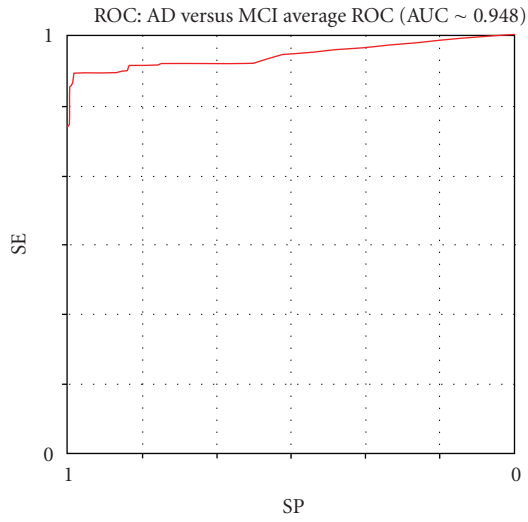


Figure 8: The average ROC curve of the ABP performance (threshold method).

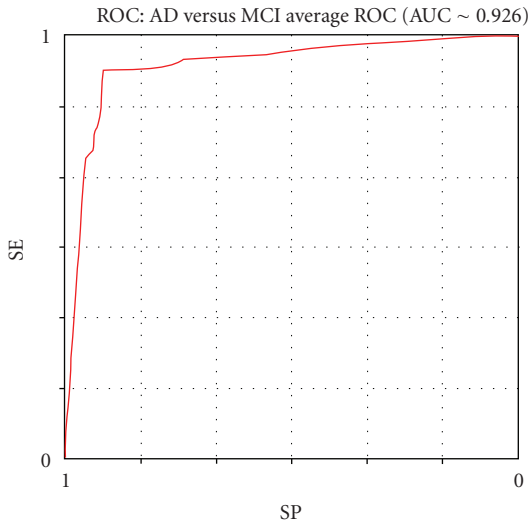


Figure 9: The average ROC curve of the NRC performance (threshold method).

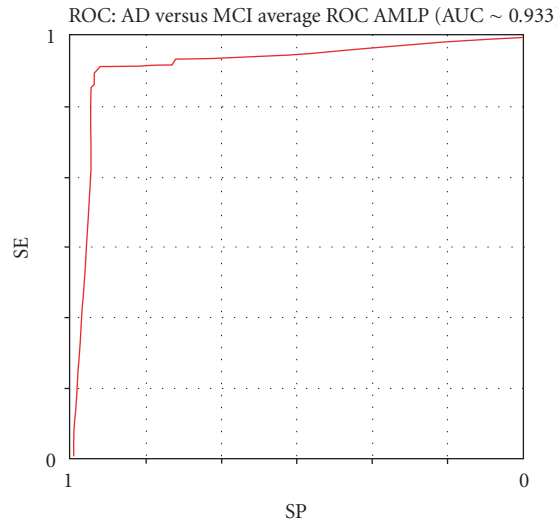


Figure 10: The average ROC curve of the AMLP performance (threshold method).

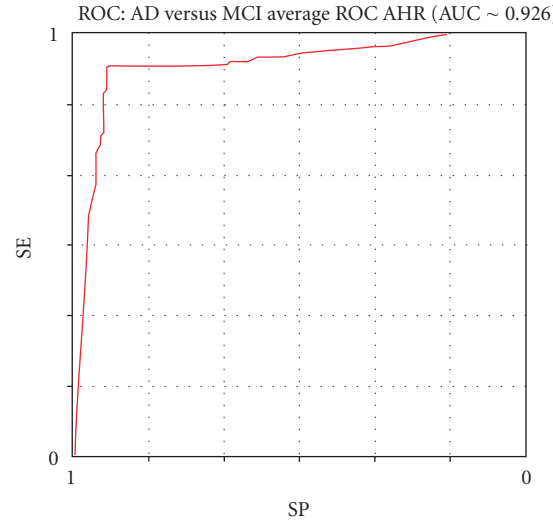


Figure 11: The average ROC curve of the AHR performance (threshold method).

4. DISCUSSION

Various types of nonreversible forms of dementias represent a major health problem in all those countries where the average life span is progressively increasing. There is a growing amount of scientific and clinical evidences that brain neural networks rearrange their connections and synapses to compensate neural loss due to neuro degeneration [64]. This process of plasticity maintains brain functions at an acceptable level before clear symptoms of dementia appear. The length of this presymptomatic period is currently unknown but, in the case of AD, often preceded by MCI, it lasts several years. Despite the lack of an effective treatment, able to block progression and/or to reverse the cognitive decline, it is generally agreed that early beginning of the available treatment (i.e., inhibitors of anticholinesterase drugs) provides the best results [65]. A significant advancement in the fight against dementias would be to have in our hands a non-invasive, easy-to-perform, and low-cost diagnostic tool capable of screening with a high rate of positive prognostication a large at-risk population sample (i.e., MCI, subjects with genetic defects and a family history of dementias or other risk factors). To test this issue, we performed automatic classification of MCI and AD subjects extracting with ANNs the spatial content of the EEG voltage. The results showed that the correct automatic classification rate reached 94.10% for AD versus MCI, better than the classification rate obtained with the more advanced currently available nonlinear techniques. These results confirm the working hypothesis that this EEG approach based on ANNs can contribute to improve the precision of

the diagnostic phase in association with other clinical and instrumental procedures.

The present results suggest that the present variant of IFAST procedure (TWIST) could be used for a large screening of MCI subjects under control, to detect the first signs of conversion to AD for triggering further clinical and instrumental evaluations crucial for an early diagnosis of AD (this is invaluable for the beginning of cholinergic therapies that are generally carried out only in overt AD patients due to gastro intestinal side effects). Indeed, the actual percentage of correct discrimination between MCI and probable AD is around 94%. This rate is clearly insufficient for the use of the IFAST procedure for a diagnosis, due to 6% of misclassifications. The present results prompt future studies on the predictive value of cortical EEG rhythms in the early discrimination of MCI subjects who will convert to AD. This interesting issue could be addressed by a proper longitudinal study. MCI subjects should be divided into "converted" and "stable" subgroups, according to final out-come as revealed by followup after about 5 years (i.e., the period needed for conversion of all MCI subjects fated to decline over time based on the mentioned literature). That study should demonstrate that the spatial EEG features at baseline measurement as revealed by the IFAST procedure might be discriminated between MCI converted and MCI stable subjects. Furthermore, baseline values of spatial EEG features in individual MCI subjects should be successfully used as an input by the IFAST procedure to predict the conversion to dementia. This intriguing research perspectives are the sign of the heuristic value of the present findings. However, apart from clinical perspectives, the present findings have an intrinsic value for clinical neurophysiology. They provided further functional data from a large aged population to support the idea that spatial features of EEG, as a reflection of the cortical neural synchronization, convey information content able to discriminate preclinical stage of dementia (MCI) from probable AD.

Furthermore, the evaluation of that diagnostic contribution may motivate future scientific studies probing its usefulness for prognosis and monitoring of AD across temporal domain.

Although EEG would fulfil up all the previous requirements, the way in which it is currently utilized does not guarantee its ability in the differential diagnosis of MCI, early AD, and healthy nonimpaired aged brains. The neurophysiologic community always had the perception that there is much more information about brain functioning embedded in the EEG signals than those actually extracted in a routine clinical context. The obvious consideration is that the generating sources of EEG signals (cortical postsynaptic currents at dendritic tree level) are the same ones as those attacked by the factors producing symptoms of dementia. The main problem is that usually in the signal-to-noise ratio the latter is largely overcoming the former.

This paper suggests that the reasons why the clinical use of EEG has been somewhat limited and disappointing with respect to early diagnosis of AD and identification of MCI—despite the progresses obtained in recent years—are due to the following, erring, general principles:

(A) identify and synthesizing the mathematical components of the signal coming from each *individual* recording site, considering the EEG channel as exploring only one, discrete brain area under the exploring electrode, and suming up all of them in attempt to reconstruct the general information;
(B) focusing on the time variations of the signal coming from each *individual* recording site,
(C) mainly employing linear analysis instruments.

The basic principle which is proposed in this work is very simple; *all* the signals from *all* the recording channels are analyzed together—and not individually—in both time and space. The reason for such an approach is quite simple; the instant value of the EEG in any recording channel depends, in fact, upon its previous and following values, and upon the previous and following values of *all* the other recording channels.

We believe that the EEG of each individual subject is defined by a specific *background signal model*, distributed in time and in the space of the recording channels (19 in our case). Such a model is a set of background invariant features able to specify the quality (i.e., cognitive level) of the brain activity, even in so a called resting condition. We all know that the brain never rests, even with closed eyes and if the subject is required to relax. The method that we have applied in this research context completely ignores the subject's contingent characteristics (age, cognitive status, emotions, etc.). It utilized a recurrent procedure which *squeezes* the significant signal and progressively selects the features useful for the classification.

5. CONCLUSIONS

We have tested the hypothesis that a correct automatic classification of MCI and AD subjects can be obtained extracting spatial information content of the resting EEG voltage by ANNs. The spatial content of the EEG voltage was extracted by a novel step-wise procedure. The core of this procedure was that the ANNs did not classify individuals using EEG data as an input; rather, the data inputs for the classification were the weights of the connections within an ANN trained to generate the recorded EEG data. These connection weights represented a useful model of the peculiar spatial features of the EEG patterns at scalp surface. Then the new system TWIST, based on a genetic algorithm, processed the weights to select the most relevant features and at the same time to create the best subset, training set, and testing set, for the classification. The results showed that the correct automatic classification rate reached 94.10% for AD versus MCI. The results obtained are superior to those obtained with the more advanced currently available nonlinear techniques. These results confirm the working hypothesis and represent the basis for research designed to integrate EEG-derived spatial and temporal information content using ANNs.

From methodological point of view, this research shows the need to analyze the 19 EEG channels of each person as a whole complex system, whose decomposition and/or linearization can involve the loss of many key information.

The present approach extends those of previous EEG studies applying advanced techniques (wavelet, neural networks, etc.) on the data of single recording channels; it also complements those of previous EEG studies in aged people, evaluating the spatial distributions of the EEG data instant by instant and the brain sources of these distributions [2–10].

With complex systems, it is not possible to establish a priori which information is relevant and which is not. Nonlinear autoassociative ANNs are a group of methods to extract from these systems the maximum of linear and nonlinear associations (features) able to explain their "strange" dynamics.

This research also documents the need to use different architectures and topologies of ANNs and evolutionary systems within complex procedures in order to optimize a specific medical target. This study's EEG analysis used

(1) different types of nonlinear autoassociative ANNs for squashing data;
(2) a new system, TWIST, based on a genetic algorithm, which manages supervised ANNs in order to select the most relevant features and to optimize the distribution of the data in training and testing sets;
(3) a set of supervised ANNs for the final patterns recognition task.

It is reasonable to conclude that ANNs and other adaptive systems should be used as cooperative adaptive agents within a structured project for complex, useful applications.

NOTE

IFAST is a european patent (application no. EP06115223.7—date of receipt 09.06.2006). The owner of the patent is Semeion Research Center of Sciences of Communication, Via Sersale 117, Rome 00128, Italy. The inventor is Massimo Buscema. For software implementation, see [53]. Dr. C. D. Percio (Associazione Fatebenefratelli per la Ricerca) organized the EEG data cleaning.

REFERENCES

[1] M. Buscema, P. Rossini, C. Babiloni, and E. Grossi, "The IFAST model, a novel parallel nonlinear EEG analysis technique, distinguishes mild cognitive impairment and Alzheimer's disease patients with high degree of accuracy," *Artificial Intelligence in Medicine*, vol. 40, no. 2, pp. 127–141, 2007.

[2] C. Babiloni, G. Binetti, E. Cassetta, et al., "Mapping distributed sources of cortical rhythms in mild Alzheimer's disease. A multicentric EEG study," *NeuroImage*, vol. 22, no. 1, pp. 57–67, 2004.

[3] C. Babiloni, G. Frisoni, M. Steriade, et al., "Frontal white matter volume and delta EEG sources negatively correlate in awake subjects with mild cognitive impairment and Alzheimer's disease," *Clinical Neurophysiology*, vol. 117, no. 5, pp. 1113–1129, 2006.

[4] C. Babiloni, L. Benussi, G. Binetti, et al., "Apolipoprotein E and alpha brain rhythms in mild cognitive impairment: a multicentric electroencephalogram study," *Annals of Neurology*, vol. 59, no. 2, pp. 323–334, 2006.

[5] N. Tsuno, M. Shigeta, K. Hyokid, P. L. Faber, and D. Lehmann, "Fluctuations of source locations of EEG activity during transition from alertness to sleep in Alzheimer's disease and vascular dementia," *Neuropsychobiology*, vol. 50, no. 3, pp. 267–272, 2004.

[6] C. Huang, L.-O. Wahlund, T. Dierks, P. Julin, B. Winblad, and V. Jelic, "Discrimination of Alzheimer's disease and mild cognitive impairment by equivalent EEG sources: a cross-sectional and longitudinal study," *Clinical Neurophysiology*, vol. 111, no. 11, pp. 1961–1967, 2000.

[7] T. Dierks, R. Ihl, L. Frölich, and K. Maurer, "Dementia of the Alzheimer type: effects on the spontaneous EEG described by dipole sources," *Psychiatry Research*, vol. 50, no. 3, pp. 151–162, 1993.

[8] T. Dierks, V. Jelic, R. D. Pascual-Marqui, et al., "Spatial pattern of cerebral glucose metabolism (PET) correlates with localization of intracerebral EEG-generators in Alzheimer's disease," *Clinical Neurophysiology*, vol. 111, no. 10, pp. 1817–1824, 2000.

[9] T. Dierks, L. Frölich, R. Ihl, and K. Maurer, "Correlation between cognitive brain function and electrical brain activity in dementia of Alzheimer type," *Journal of Neural Transmission*, vol. 99, no. 1–3, pp. 55–62, 1995.

[10] J. Hara, W. R. Shankle, and T. Musha, "Cortical atrophy in Alzheimer's disease unmasks electrically silent sulci and lowers EEG dipolarity," *IEEE Transactions on Biomedical Engineering*, vol. 46, no. 8, pp. 905–910, 1999.

[11] C. Huang, L.-O. Wahlund, T. Dierks, P. Julin, B. Winblad, and V. Jelic, "Discrimination of Alzheimer's disease and mild cognitive impairment by equivalent EEG sources: a cross-sectional and longitudinal study," *Clinical Neurophysiology*, vol. 111, no. 11, pp. 1961–1967, 2000.

[12] K. Bennys, G. Rondouin, C. Vergnes, and J. Touchon, "Diagnostic value of quantitative EEG in Alzheimer's disease," *Neurophysiologie Clinique*, vol. 31, no. 3, pp. 153–160, 2001.

[13] M. Nuwer, "Assessment of digital EEG, quantitative EEG, and EEG brain mapping: report of the American Academy of Neurology and the American Clinical Neurophysiology Society," *Neurology*, vol. 49, no. 1, pp. 277–292, 1997.

[14] G. Adler, S. Brassen, and A. Jajcevic, "EEG coherence in Alzheimer's dementia," *Journal of Neural Transmission*, vol. 110, no. 9, pp. 1051–1058, 2003.

[15] T. Musha, T. Asada, F. Yamashita, et al., "A new EEG method for estimating cortical neuronal impairment that is sensitive to early stage Alzheimer's disease," *Clinical Neurophysiology*, vol. 113, no. 7, pp. 1052–1058, 2002.

[16] C. Melissant, A. Ypma, E. E. E. Frietman, and C. J. Stam, "A method for detection of Alzheimer's disease using ICA-enhanced EEG measurements," *Artificial Intelligence in Medicine*, vol. 33, no. 3, pp. 209–222, 2005.

[17] A. A. Petrosian, D. V. Prokhorov, W. Lajara-Nanson, and R. B. Schiffer, "Recurrent neural network-based approach for early recognition of Alzheimer's disease in EEG," *Clinical Neurophysiology*, vol. 112, no. 8, pp. 1378–1387, 2001.

[18] F. Vialatte, A. Cichocki, G. Dreyfus, T. Musha, S. L. Shishkin, and R. Gervais, "Early detection of Alzheimer's disease by blind source separation, time frequency representation, and bump modeling of EEG signals," in *Proceedings of the 15th International Conference on Artificial Neural Networks: Biological Inspirations (ICANN '05)*, vol. 3696 of *Lecture Notes in Computer Science*, pp. 683–692, Springer, Warsaw, Poland, September 2005.

[19] J. Jeong, "EEG dynamics in patients with Alzheimer's disease," *Clinical Neurophysiology*, vol. 115, no. 7, pp. 1490–1505, 2004.

[20] A. Cichocki, S. L. Shishkin, T. Musha, Z. Leonowicz, T. Asada, and T. Kurachi, "EEG filtering based on blind source

separation (BSS) for early detection of Alzheimer's disease," *Clinical Neurophysiology*, vol. 116, no. 3, pp. 729–737, 2005.

[21] A. Cichocki, "Blind signal processing methods for analyzing multichannel brain signals," *International Journal of Bioelectromagtism*, vol. 6, no. 1, 2004.

[22] A. Cichocki and S.-I. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*, Wiley, New York, NY, USA, 2003.

[23] D. E. Rumelhart, P. Smolensky, J. L. McClelland, and G. E. Hinton, "Schemata and sequential thought processes in PDP models," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, J. L. McClelland and D. E. Rumelhart, Eds., vol. 2, pp. 7–57, The MIT Press, Cambridge, Mass, USA, 1986.

[24] M. Buscema, "Constraint satisfaction neural networks," *Substance Use & Misuse*, vol. 33, no. 2, pp. 389–408, 1998, special issue on artificial neural networks and complex social systems.

[25] M. Buscema, "Recirculation neural networks," *Substance Use & Misuse*, vol. 33, no. 2, pp. 383–388, 1998, special issue on artificial neural networks and complex social systems.

[26] G. E. Hinton and J. L. McClelland, "Learning representation by recirculation," in *Proceedings of IEEE Conference on Neural Information Processing Systems*, Denver, Colo, USA, November 1988.

[27] Y. Chauvin and D. E. Rumelhart, Eds., *Backpropagation: Theory, Architectures, and Applications*, Lawrence Erlbaum Associates, Hillsdale, NJ, USA, 1995.

[28] J. L. Elman, "Finding structure in time," *Cognitive Science*, vol. 14, no. 2, pp. 179–211, 1990.

[29] M. Buscema, "I FAST Software, Semeion Software #32," Rome, Italy, 2005.

[30] M. Buscema, E. Grossi, M. Intraligi, N. Garbagna, A. Andriulli, and M. Breda, "An optimized experimental protocol based on neuro-evolutionary algorithms: application to the classification of dyspeptic patients and to the prediction of the effectiveness of their treatment," *Artificial Intelligence in Medicine*, vol. 34, no. 3, pp. 279–305, 2005.

[31] M. Buscema, "TWIST Software, Semeion Software #32," Rome, Italy, 2005.

[32] M. Buscema, "Genetic doping algorithm (GenD): theory and applications," *Expert Systems*, vol. 21, no. 2, pp. 63–79, 2004.

[33] L. Davis, *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, New York, NY, USA, 1991.

[34] S. Harp, T. Samed, and A. Guha, "Designing application-specific neural networks using the genetic algorithm," in *Advances in Neural Information Processing Systems*, D. Touretzky, Ed., vol. 2, pp. 447–454, Morgan Kaufman, San Mateo, Calif, USA, 1990.

[35] M. Mitchell, *An Introduction to Genetic Algorithms*, The MIT Press, Cambridge, Mass, USA, 1996.

[36] D. Quagliarella, J. Periaux, C. Polani, and G. Winter, *Genetic Algorithms and Evolution Strategies in Engineering and Computer Science*, John Wiley & Sons, Chichester, UK, 1998.

[37] G. Rawling, *Foundations of Genetic Algorithms*, Morgan Kaufman, San Mateo, Calif, USA, 1991.

[38] T. G. Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms," *Neural Computation*, vol. 10, no. 7, pp. 1895–1923, 1998.

[39] E. H. Rubin, J. C. Morris, F. A. Grant, and T. Vendegna, "Very mild senile dementia of the Alzheimer type I. Clinical assessment," *Archives of Neurology*, vol. 46, no. 4, pp. 379–382, 1989.

[40] M. Albert, L. A. Smith, P. A. Scherr, J. O. Taylor, D. A. Evans, and H. H. Funkenstein, "Use of brief cognitive tests to identify individuals in the community with clinically diagnosed Alzheimer's disease," *International Journal of Neuroscience*, vol. 57, no. 3-4, pp. 167–178, 1991.

[41] C. Flicker, S. H. Ferris, and B. Reisberg, "Mild cognitive impairment in the elderly," *Neurology*, vol. 41, no. 7, pp. 1006–1009, 1991.

[42] M. Zaudig, "A new systematic method of measurement and diagnosis of "mild cognitive impairment" and dementia according to ICD-10 and DSM-III-R criteria," *International Psychogeriatrics*, vol. 4, supplement 2, pp. 203–219, 1992.

[43] D. P. Devanand, M. Folz, M. Gorlyn, J. R. Moeller, and Y. Stern, "Questionable dementia: clinical course and predictors of outcome," *Journal of the American Geriatrics Society*, vol. 45, no. 3, pp. 321–328, 1997.

[44] R. C. Petersen, G. E. Smith, R. J. Ivnik, et al., "Apolipoprotein E status as a predictor of the development of Alzheimer's disease in memory-impaired individuals," *Journal of the American Medical Association*, vol. 273, no. 16, pp. 1274–1278, 1995.

[45] R. C. Petersen, G. E. Smith, S. C. Waring, R. J. Ivnik, E. Kokmen, and E. G. Tangelos, "Aging, memory, and mild cognitive impairment," *International Psychogeriatrics*, vol. 9, no. 1, pp. 65–69, 1997.

[46] R. C. Petersen, R. Doody, A. Kurz, et al., "Current concepts in mild cognitive impairment," *Archives of Neurology*, vol. 58, no. 12, pp. 1985–1992, 2001.

[47] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M. Stadlan, "Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of department of health and human services task force on Alzheimer's disease," *Neurology*, vol. 34, pp. 939–944, 1984.

[48] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "Mini mental state: a practical method for grading the cognitive state of patients for the clinician," *Journal of Psychiatric Research*, vol. 12, no. 3, pp. 189–198, 1975.

[49] C. P. Hughes, L. Berg, W. L. Danziger, L. A. Coben, and R. L. Martin, "A new clinical scale for the staging of dementia," *The British Journal of Psychiatry*, vol. 140, pp. 566–572, 1982.

[50] J. A. Yesavage, T. L. Brink, T. L. Rose, et al., "Development and validation of a geriatric depression screening scale: a preliminary report," *Journal of Psychiatric Research*, vol. 17, no. 1, pp. 37–49, 1983.

[51] W. G. Rosen, R. D. Terry, P. A. Fuld, R. Katzman, and A. Peck, "Pathological verification of ischemic score in differentiation of dementias," *Annals of Neurology*, vol. 7, no. 5, pp. 486–488, 1980.

[52] M. P. Lawton and E. M. Brody, "Assessment of older people: self maintaining ad instrumental activities of daily living," *Gerontologist*, vol. 9, no. 3, pp. 179–186, 1969.

[53] A. Brun, B. Englund, L. Gustafson, et al., "Consensus on clinical and neuropathological criteria for fronto-temporal dementia," *Journal of Neurology, Neurosurgery and Psychiatry*, vol. 57, pp. 416–418, 1994.

[54] G. C. Roman, T. K. Tatemichi, T. Erkinjuntti, et al., "Vascular dementia: diagnostic criteria for research studies: report of the NINDS-AIREN international workshop," *Neurology*, vol. 43, no. 2, pp. 250–260, 1993.

[55] G. B. Frisoni, A. Beltramello, G. Binetti, et al., "Computed tomography in the detection of the vascular component in dementia," *Gerontology*, vol. 41, no. 2, pp. 121–128, 1995.

[56] S. Galluzzi, C. F. Sheu, O. Zanetti, and G. B. Frisoni, "Distinctive clinical features of mild cognitive impairment with subcortical cerebrovascular disease," *Dementia and Geriatric Cognitive Disorders*, vol. 19, no. 4, pp. 196–203, 2005.

[57] I. G. McKeith, D. Galasko, K. Kosaka, et al., "Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): report of the consortium on DLB international workshop," *Neurology*, vol. 47, no. 5, pp. 1113–1124, 1996.

[58] R. J. Buchan, K. Nagata, E. Yokoyama, et al., "Regional correlations between the EEG and oxygen metabolism in dementia of Alzheimer's type," *Electroencephalography and Clinical Neurophysiology*, vol. 103, no. 3, pp. 409–417, 1997.

[59] E. Pucci, N. Belardinelli, G. Cacchiò, M. Signorino, and F. Angeleri, "EEG power spectrum differences in early and late onset forms of Alzheimer's disease," *Clinical Neurophysiology*, vol. 110, no. 4, pp. 621–631, 1999.

[60] B. Szelies, R. Mielke, J. Kessler, and W.-D. Heiss, "EEG power changes are related to regional cerebral glucose metabolism in vascular dementia," *Clinical Neurophysiology*, vol. 110, no. 4, pp. 615–620, 1999.

[61] G. Rodriguez, P. Vitali, C. De Leo, F. De Carli, N. Girtler, and F. Nobili, "Quantitative EEG changes in Alzheimer patients during long-term donepezil therapy," *Neuropsychobiology*, vol. 46, no. 1, pp. 49–56, 2002.

[62] C. Babiloni, R. Ferri, D. V. Moretti, et al., "Abnormal frontoparietal coupling of brain rhythms in mild Alzheimer's disease: a multicentric EEG study," *European Journal of Neuroscience*, vol. 19, no. 9, pp. 2583–2590, 2004.

[63] D. V. Moretti, F. Babiloni, F. Carducci, et al., "Computerized processing of EEG-EOG-EMG artifacts for multi-centric studies in EEG oscillations and event-related potentials," *International Journal of Psychophysiology*, vol. 47, no. 3, pp. 199–216, 2003.

[64] Y. Stern, "Cognitive reserve and Alzheimer disease," *Alzheimer Disease and Associated Disorders*, vol. 20, no. 2, pp. 112–117, 2006.

[65] S. G. Gauthier, "Alzheimer's disease: the benefits of early treatment," *European Journal of Neurology*, vol. 12, no. 3, pp. 11–16, 2005.

Hindawi Publishing Corporation
Computational Intelligence and Neuroscience
Volume 2007, Article ID 93968, 9 pages
doi:10.1155/2007/93968

Research Article

The P300 as a Marker of Waning Attention and Error Propensity

Avijit Datta,[1,2] Rhodri Cusack,[1] Kari Hawkins,[1] Joost Heutink,[1] Chris Rorden,[1,3] Ian H. Robertson,[1,4] and Tom Manly[1]

[1] *Medical Research Council Cognition and Brain Sciences Unit, CB2 7EF Cambridge, UK*
[2] *Hull York Medical School, York Hospitals NHS Foundation Trust, York YO31 8HE, UK*
[3] *Arnold School of Public Health, University of South Carolina, Columbia, SC 29208, USA*
[4] *Department of Psychology and Institute of Neuroscience, Trinity College, Dublin 2, Ireland*

Correspondence should be addressed to Avijit Datta, avijit.datta@york.nhs.uk

Received 17 February 2007; Accepted 9 October 2007

Recommended by Saied Sanei

Action errors can occur when routine responses are triggered inappropriately by familiar cues. Here, EEG was recorded as volunteers performed a "go/no-go" task of long duration that occasionally and unexpectedly required them to withhold a frequent, routine response. EEG components locked to the onset of relevant go trials were sorted according to whether participants erroneously responded to immediately *subsequent* no-go trials or correctly withheld their responses. Errors were associated with a significant relative reduction in the amplitude of the preceding P300, that is, a judgement could be made bout whether a response-inhibition error was likely before it had actually occurred. Furthermore, fluctuations in P300 amplitude across the task formed a reliable associate of individual error propensity, supporting its use as a marker of sustained control over action.

1. INTRODUCTION

"Absent-minded" slips of action often result from the inappropriate production of an automatic or routine response [1]. Many of us will have repeatedly attempted to switch on light bulbs that we "know" need replacing, or automatically driven a familiar route when we were intending to go elsewhere. Although routine activities may be skilfully performed with little requirement for continuous control, there are occasions when such unsupervised actions can have serious consequences, from personal accidents to major disasters [2]. Moreover, the tendency to make such action errors significantly increases following traumatic brain injury, focal frontal lesions, and in some developmental disorders [3–12]. Here we examine whether time-locked EEG components may be sensitive to different states in which such errors are more or less likely to occur.

Slips of attention have been studied both in terms of predicting difficulties faced by clinical groups and in developing models of normal executive control over action. Norman and Shallice [13] and Shallice [10], for example, proposed an influential framework in which routine actions are controlled in a relatively automatic or stimulus-driven manner. Within this view, the expression of one behavioral sequence rather than another is governed by a competitive process determined by the strength of environmental triggers. Via such a system, apparently complex activities such as those involved in driving a car can be performed appropriately with little requirement for higher-level control. The second level of control, termed supervisory attention, is then proposed to modulate action selection if, for example, the most active behavioral sequence is inappropriate in relation to an overall goal. Such control is also experienced subjectively as effortful and conscious attention. More recently proposed frameworks draw similar distinctions. Dehaene and Naccache [14], for example, argue for a fronto-parietal circuit that acts as a "global workspace," regulating more routine processes and which is associated with conscious effort. One set of conditions under which supervisory control is argued to be crucial is that presented in sustained attention tasks. In such tasks, the environmental triggers for goal-related behavior are reduced to a minimum, either by making the task "boring,"

increasing the time over which a participant has to self-maintain a readiness to respond, and/or increasing the duration beyond a point of tedium [15–17]. The more successful a task is in reducing environmental support, the greater is its emphasis on the internal, or "endogenous," maintenance of the appropriate processing stance.

Robertson et al. [9] developed a simple paradigm designed to assess self-maintained attention to current action. In the sustained attention to response task (SART), participants' watch-as-single digits are presented on a computer screen at a regular, invariant rate. They are asked to press a single button for each digit as it appears. The rhythmic nature of this response, coupled with the lack of selection, was designed to rapidly establish a relatively automatic, task-driven response. Periodically and unpredictably, however, a "no-go" target is presented to which no response should be made. In order to maximize the chances of not making an error, it has been argued, participants must try and counter the tendency to lapse into routine responding and maintain a high degree of control over action throughout the task. This brief and reliable task has proved to be sensitive to the frequency of everyday action lapses in traumatically brain injured patients [9] and in neurologically healthy volunteers [18].

The electroencephalogram (EEG) signal reflects brain activity including that which is in response to a specific environmental event. Such event-related responses are often difficult to separate from other activity on a trial-by-trial basis. If time-locked signals to many identical events are averaged, however, the unrelated signal tends to cancel out and the event related potentials (ERPs) emerge. The electrophysiological correlates of performance on tasks, such as the SART, that emphasise alternation between responding and not responding (termed "go/no-go" tasks) have been extensively examined [19–23]. The emphasis in such studies has been on differential responses to the presentation of the no-go stimulus relative to the go trial. Mäntysalo [23], for example, found increased amplitude of a negative component (N200) and a positive component (P300) on no-go trials a feature subsequently interpreted by Kok [22], and by Eimer [19] as reflecting response-inhibition processes. Jackson et al. [20] also found that the P300 component to the visual stimulus was more rapidly suppressed during no-go trials. These studies place emphasis on what happens after a "no-go" trial is presented. The focus here is on what happens *before* a no-go trial is unexpectedly presented. If, as has been argued, the ability to control action on no-go trials is determined by a *pre-existing attentive* state (sustained attention during the task), then it may be possible to assess this independently of overt behavior using ERP measures. Our hypothesis was that correct go trials that precede a correctly withheld response in a no-go trial should show evidence of this heightened attentive control relative to go trials that precede an error. A conceptual advantage of this approach lies in the degree to which other factors that might influence the ERP are controlled. In each case, the comparison is between correct go trials that are identical in terms of the stimulus presented (go), the response made (press), the instructional set (do not press for no-gos), and the probability of a subsequent trial being a no-go signal (1/8). If reliable differences emerge between trials that precede an action slip and those that do not, this can be interpreted with some confidence as being related to the attentional state of the participant under which subsequent errors are more or less likely.

There were cogent reasons for us to focus on the P300 ERP component as a likely predictor of errors in the SART go/no-go tasks. The P300 is a positive wave occurring approximately in 300 milliseconds following stimulus presentation [24, 25]. In contrast to some earlier components within the ERP, the P300 has been argued to reflect higher-level processes that are sensitive to task context, such as attentive selection [24, 26]. Increased P300 amplitude has been reported when participants detect that they have made an error in go/no-go tasks [27, 28], which may be interpreted in terms of error detection or the consequent establishment of a more attentional stance in which subsequent error probability is reduced. Further, studies have shown that the P300 is significantly reduced in survivors of traumatic brain injury, a group who have particular difficulty in avoiding errors on the SART [29–31].

In the current study, a group of neurologically healthy volunteers performed multiple blocks of the SART task to establish whether variations in P300 amplitude were associated with action errors in the SART. For each participant, the 250 *no-go* trials from the 10 blocks of the SART were first indexed and sorted according to whether the participant had made a commission error, by incorrectly pressing the response key, or had correctly withheld the response. For each of these categories, the visual ERPs to go trials that immediately *preceded* these no-go trials were then averaged first for each participant and then for the group of 25 participants as a whole. From previous studies, we anticipated sufficiently high error rates in this group to allow a reasonable comparison between events prior to a correct no-go trial and prior to an action error.

Previous studies have shown that SART is relatively reliable in picking up enduring individual differences in error propensity. In addition to the hypothesis that relatively high or low P300 amplitude would be associated at a *within-subject* level with different subsequent error rates, we therefore further hypothesized that individual differences in the degree to which the P300 component was maintained across all of the go trials would be associated with individual differences in error rates.

For both analyses, there were advantages if gross individual differences in P300 amplitude (e.g., due to the quality of electrode contact, skull thickness, etc.) could be reduced. To this end, we expressed P300 in proportion to that of an earlier ERP component, the P200 (P200 : P300 ratio). The P200 should be subject to the same intersubject differences affecting absolute amplitude but, in being thought to reflect more perceptual aspects of the neural response, less likely to be modulated by current attentional engagement with the task. For this reasoning to be valid, it would be necessary to additionally demonstrate *in the current task* that variations in the P300 are related to subsequent error while variations in the P200 are not.

The hypotheses can therefore be summarized as follows.

(1) Having first grouped no-go trials according to whether or not an error occurred, the average amplitude of the P300 on pre*ceding go trials* will vary in relation to the outcome on those subsequent no-go trials. The earlier and more perceptual P200 will not.
(2) If so, this will allow us to reduce gross between-subject differences by expressing P300 amplitude relative to that of the P200 (P200 : P300). Averaged across the group, we then predict that the "normalized" P300 value will differ between go-trials preceding an error and those preceding a correct no-go trial.
(3) In addition to those go trials that immediately precede no-go trials, we would expect the degree to which the normalized P300 amplitude is maintained across the task as a whole to reflect error rates. Specifically, in a correlational analysis, the mean normalized P300 amplitude (P200 : P300) across *all* go trials will be associated with individual differences in error propensity.

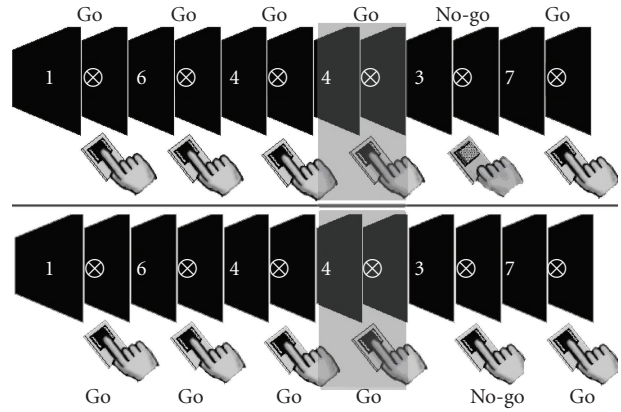


Figure 1: Selection of trials for the main comparison. Each figure represents the sequence of events in the SART where go trials are defined by any digit between 1 and 9 (except 3) and the no-go target by the 3. In each sequence, the participant is responding correctly to go trials. In the upper panel, the presentation of the target is followed by a correctly withheld response. In the lower panel, by an error, the correct go trials *prior to* these no-go signals (highlighted) form the basis of the comparison.

2. MATERIALS AND METHODS

2.1. Participants

Following ethical committee approval, 25 neurologically healthy right-handed volunteers (13 women and 12 men, age range 20–47) gave informed consent for their participation in the study.

2.2. Electrophysiological recording and averaging

EEG recordings were made from 3 midline sites (Fz, Cz, Pz) using silver/silver chloride electrodes (Grass). Four additional electrodes were applied for eye blink and movement monitoring, grounding, and reference. The electrodes were referenced to the right mastoid during recording. The horizontal electro-oculogram electrodes were referenced to each other. The EEG and EOG signals were amplified with a bandwidth of 0.05–100 Hz. The digitization rate for the analogue-to-digital conversion was 500 samples per second. Prior to averaging, artefact rejection was performed on the data to discard epochs in which amplifier saturation, eye movements, blinks or excessive muscle, or movement artefacts occurred. The same rejection criteria were used for all participants. In some cases, the rejection values for eye artefacts were individually adjusted, to correct for individual differences in amplitudes of artefacts and EEG. This procedure resulted in an average rejection of no more than 2% of the trials for each of the 25 subjects included in the analysis. Electromyogram signals (EMG) in the responding hand were monitored using bipolar silver/silver chloride electrodes from an index finger flexor (first dorsal interosseous muscle) and an index finger extensor (extensor indicis). EEG was amplified 20 000 fold, EMG 1000 fold, and EOG 2000 fold using AC coupled amplifiers (Biopac Systems Inc., Santa Barbara). Filtering was 10 Hz–5 KHz, 1–35 Hz, and 0.05 Hz–100 Hz for EMG, EEG, and EOG, respectively. Full-wave rectification of the EMG was performed digitally. All data was digitized at 500 Hz, indexed for go, no-go stimulus, correct and incorrect response, archived, and averaged offline using a purpose-written averaging program. For stimulus-locked averages, the P300 was defined as the maximum positive peak amplitude between 250–450 milliseconds after stimulus presentation. Latencies of peaks were clearly identifiable in each case.

2.3. Behavioral task: the sustained attention to response test

The task [32] was presented on a Dell Latitude laptop computer isolated from the mains supply. On each trial, a single digit (1–9) was selected at random and presented for 250 milliseconds, followed by a mask for 900 milliseconds, at the center of the 185 mm × 245 mm screen. Participants, who were at a comfortable viewing distance from the screen (around 40 cm), were asked to press a mouse button with the index finger of their preferred hand as quickly as possible after each digit presented, with the exception of 3, to which no response should be made. They were asked to press the button "as quickly but as accurately as possible" following the onset of the trial. The randomization meant that 25 no-go trials (3 seconds) appeared unpredictably amid 200 go trials (all digits other than 3) in each block. Each participant completed 10 blocks with the opportunity to rest from the task between each.

Testing took place in a quiet, darkened room that was free from distraction. The total testing session, including setting up and removing the recording electrodes, lasted for approximately 3 hours. Figure 1 illustrates the sequence of events in the task and the two types of go trial (defined by immediately subsequent no-go trial error) that inform the main ERP comparison of this study.

3. RESULTS

3.1. Performance on the task

The participants completed 10 blocks of the SART, comprising 2000 go trials and 250 (11.1%) randomly intermixed no-go targets. The participants correctly withheld their responses to 147.72 (59%) of the 250 no-go trials (SD 16.12) and made errors of commission on an average of 102.28 (41%) of these trials (SD 16.105). As is common, errors of *omission* (i.e., not pressing the response key on go trials) were very rare, occurring on an average of 0.55 of the 2000 go trials (0.061%, SD 0.17%).

3.2. ERPs to the visual stimulus prior to a correct no-go trial and prior to an action error

Previous behavioral studies with the SART suggest that, other than in severely brain injured individuals, correct responses on no-go trials are likely to outweigh errors of commission. As reported above, this was the case with the healthy participants tested here. Correct responses accounted for about 60% of the no-go trials with around 40% attracting action errors. This error rate is somewhat highly compared with previous studies and may be related to the presentation of 10 consecutive blocks, rather than the more conventional single block. This higher rate is, however, to our advantage in comparing pre-error and pre-correct go trials. With both categories yielding between 80 and 170 trials per person (pre-error mean = 102.28, SD 16.11, range 80–147, pre-correct mean = 147.72, SD 16.10, range 103–170), there are sufficient numbers for noise to tend towards zero in the averaged signals for each participant in both categories, with these values then being again averaged across the group. Any differences between the waveforms should *not* therefore be attributable to a disparity between the overall number of error and correct trials. However, the risk of this unlikely confound is further reduced by our focus on a single wave, the P300, as the component that should show a difference. If it is the P300 which is indeed different while other components are broadly equivalent, it is less likely that *differential* amounts of noise, which would be distributed across the signal, would have such a specific effect.

The Pz ERPs for go trials before an error and before a correct response suppression are presented in Figure 2 below. Figure 2(a) shows the pattern that preceded a correct no-go trial while Figure 2(b) shows the pattern preceding an action error. Each panel shows (from top to bottom) averaged rectified agonist and antagonist muscle electromyogram (EMG), averaged scalp electroencephalogram (EEG; at Pz), and averaged electrooculogram (EOG; used in controlling for eye movements).

In both Figures 2(a) and 2(b), a triphasic response is observed in Pz EEG with peaks of each wave occurring at similar latencies (270 milliseconds, 384 milliseconds, and 544 milliseconds in Figure 2(a), and 250 milliseconds, 388 milliseconds, and 542 milliseconds in Figure 2(b)). The amplitude and latency of the positivity between 200–300 milliseconds after the onset of the trial (P200) is strikingly similar to the two trial types. Given that these go trials are effectively identical, other than in what *subsequently* happens, it is perhaps not surprising that the early perceptual components of the ERPs are so similar. As discussed above, we therefore exploited this stable feature in order to allow a comparison of the P300 components that was relatively free from the influence of noise and inter-subject variables such as signal intensity. The amplitude of the P300 component was therefore expressed as a ratio of the P200 amplitude for each participant averaged across the two trial "types" illustrated in Figure 1.

Comparison of the P300/P200 ratio between trials preceding an action error (mean 0.92, SD 0.33, $n = 25$) and trials preceding correct withholding of the response (mean 1.28, SD 0.48) reveals a robust and statistically significant difference ($t(24) = 3.63$, $P < .001$). Moreover, Figure 3 shows that the median P300/P200 ratio prior to an action error falls below even the interquartile range of the P300/P200 prior to a correct no-go trial.

In both Figures 2(a) and 2(b), EOG traces are flat until 600 milliseconds after presentation of the visual stimulus. This lack of contamination of eye movement allows confident interpretation of EEG traces and P300/P200 ratios that we have obtained. Although similar responses were seen at Cz and Fz sites, these were less compelling in magnitude and did not reach statistical significance for this group size. For brevity we will therefore focus on the Pz results in subsequent analyses (see later for discussion).

In summary, in two groups of go trials which are identical other than in what occurs on the *subsequent* no-go trial, there appears to be a difference reflected in the P300 at Pz which is related to the probability of a subsequent error. When this amplitude is relatively low, errors are more likely. For reasons outlined in the introduction, therefore, it is tempting to argue that this component is reflecting some form of enhanced attention to the stimulus/task which makes errors less likely.

3.3. Individual differences in error propensity

In the previous section we considered only those go-trials that immediately preceded no-go trials. If it is the case, as the results suggest, that an increased P300 amplitude is associated with more attention and fewer errors, it might be expected that the mean value of this marker across the whole task could reflect an individual's capacity to maintain an attentive state and overall "resistance" to inhibition errors. To examine this, we examined the Pearson correlation between each participant's (averaged) normalized P300 amplitude across all of the 2000 go trials in the task and their overall commission error rates on no-go trials.

The relationship was statistically significant (Pearson's $r = -0.46$, $P < .05$), the lower the relative average amplitude of the P300, the more action lapses a particular participant was prone to make. This relationship is further illustrated by the division of the participants into "high" and "low" relative P300 groups based on a median split. As can be seen in Figure 4, 12 participants with low Pz P300/P200 ratio values (between 0.34 and 1.07) had a mean error rate of 47% (SD 15.4%) while 13 participants with high Pz P300/P200 ratio

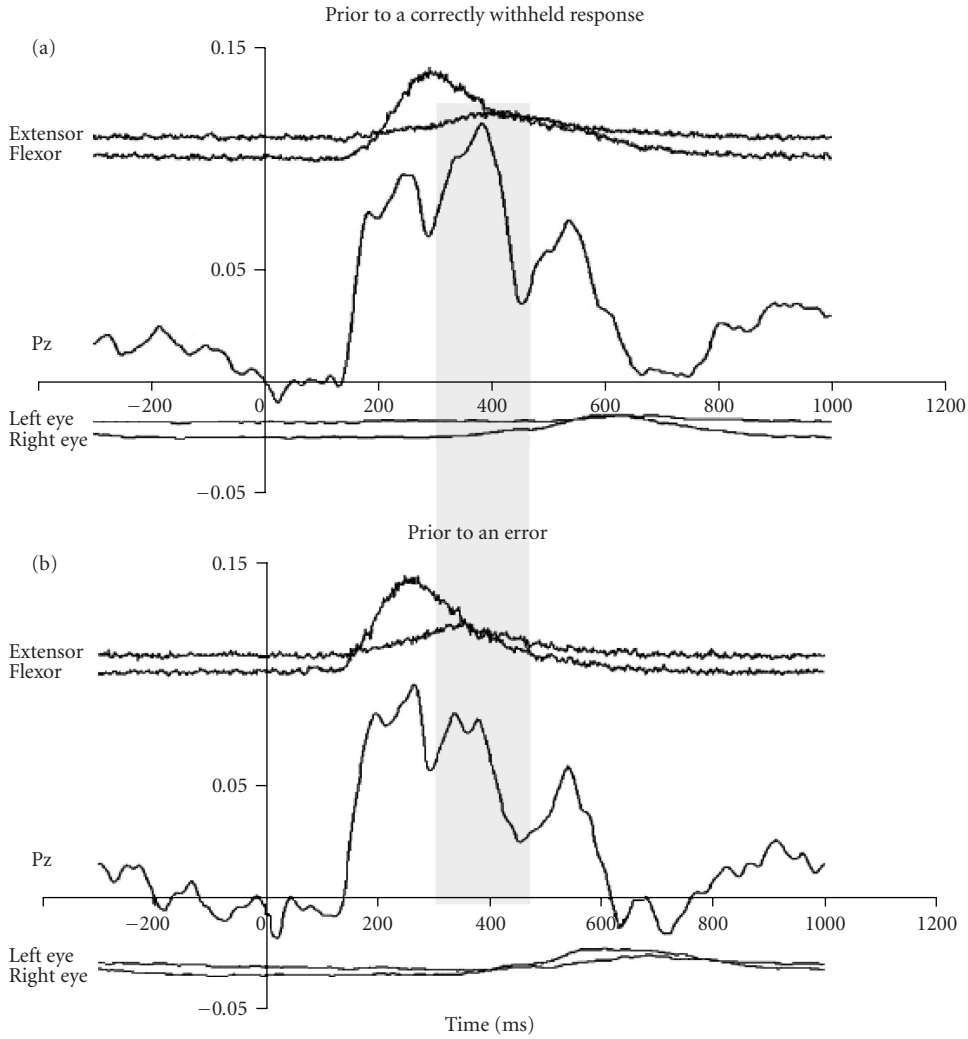


Figure 2: The difference between go trials preceding a correct or erroneous no-go trial. Each figure shows EEG (at Pz), finger muscle activity (extensor/flexor), and eye movements (left and right) averaged across all available relevant trials. The crucial difference between these go trials appears to be in the amplitude of the P300 ERP peak, highlighted in the grey band.

values (between 1.09 and 2.25) made significantly fewer errors (32.5% (SD 13.4%); $t(23) = 2.51$, $P = .02$).

3.4. Error detection and reaction time effects

So far we have seen that a reduced relative amplitude of the Pz P300 is associated with a higher probability of an error on a subsequent no-go trial and, over all of the go trials in the task, associated with increased no-go error propensity. We have so far interpreted this in terms of reflecting waning attention to the stimulus and task. However, before we can do that with confidence, there are a couple of potential confounds that should be addressed. These are "contamination" of our go trial ERPs with processes related to the detection of a *previous* error and the possibility that trials preceding errors had rather different reaction times to those preceding correct no-go trials.

The SART is a continuous task. If one has a high *overall* rate of errors on no-go trials, it is more probable that any given go trial might have occurred *after* a previous error, as well as possibly *preceding* other errors. If one made errors on 100% of no-go trials, for example, all but the first go trials may be considered to have "followed" an error. This is important because increases in the P300 have been associated with error detection, albeit that this is a feature that appears to be relatively short lived in the ERP trace [27, 28, 33]. It is possible, therefore, that the relationships so far reported between Pz P300 amplitude and subsequent error and overall error rates are mediated by error detection factors, although it should be noted that, were this the case, the direction of this relationship would be reversed (more errors = higher P300 amplitude). To examine this possibility, we compared the mean Pz P300/P200 ratio for those go trials that immediately preceded *no-go trials,* with the average for all go trials. Go trials that occur immediately before a no-go trial tend, by definition, to be as distant from a previous no-go trial as it is possible to be within the task and are therefore *less* likely to be influenced by error detection processes triggered by

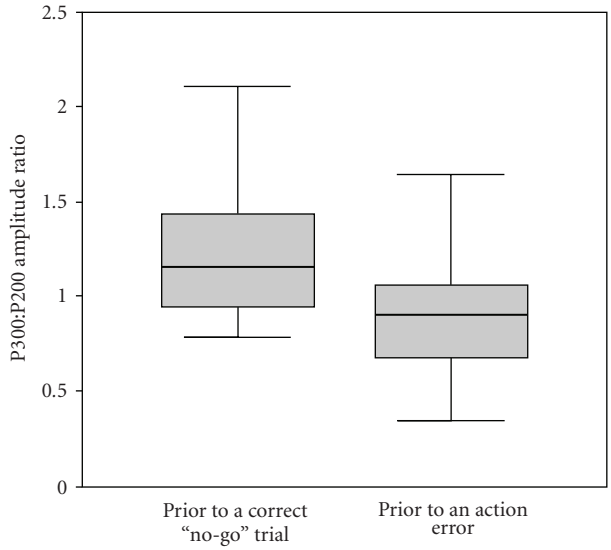


Figure 3: The difference between go trials preceding a correct or erroneous no-go trial. The figure shows a boxplot for the P300 : P200 amplitude ratio for the two "types" of go trial. Each shows the median (heavy line), interquartile range (shaded area), and total range for the 25 volunteers. Prior to a correct no-go trial, the median normalized P300 amplitude exceeds even the interquartile range of that seen in go-trials prior to an no-go error.

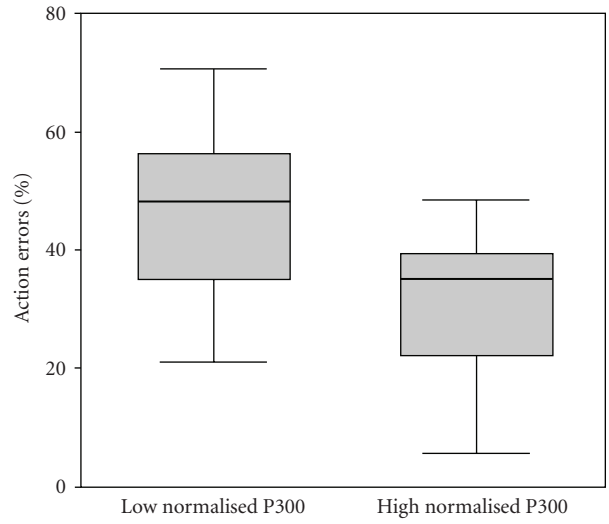


Figure 4: Propensity to error is associated with the maintenance of the P300 : P200 ratio across the task. The boxplot shows error frequencies for participants with high or low mean P300 values (defined by a median split of the total participant group), respectively.

a previous mishap. If our previously reported correlation was substantially due to error-detection processes, we would expect the relationship between the P300 amplitude in *these* trials and overall error rates to be reduced. In fact, if anything, it was enhanced (Pearson's $r = 0.57$, $P < .01$).

Finally, we investigated whether the predictive qualities of the normalized Pz P300 may be mediated by reaction time (RT) differences. The relationship between mean RT to go stimuli and error rates across subjects did not, however, reach statistical significance ($r = -0.281$, $P = .174$), meaning that, in this group, we could not predict errors on the basis of how fast individuals were responding. In addition, there was no relationship between mean RT of participants and mean amplitude of their normalized P300 response ($r = 0.1$, $P = .455$, $n = 25$), further suggesting that the relationship between individual error propensity and P300 amplitude was not mediated by response speed differences.

4. DISCUSSION

In this study, we asked participants to perform a simple go/no-go tasks in which no-go targets appeared infrequently and unpredictably within a random sequence. Previous research has suggested that this task is sensitive to everyday absentminded lapses in people with brain-injuries and healthy participants. Furthermore, there is evidence to suggest that success on no-go trials is related to how well people are able to maintain active attentive control over their responses, rather than allowing them to be "driven along" by the repetitive, regular pacing. The basis for this study was that, if there is a fluctuating state of attention allocation in which errors are sometimes more probable and sometimes less, we might be able to see this within fluctuating electrophysiological signal before the critical no-go target has even appeared.

The results were consistent with this view. From the random sequence of trials in the task, we first found go trials that happened to have occurred before a no-go trial. We then divided these into those that had been followed by a correct response suppression and those that had been followed by an error. The EEG was then averaged for each grouping, time-locked to trial onset. It is again important to stress that, from the participants' perspective, trials preceding no-go signals hold no special status, indeed, any given go trial is around 8 times more likely to be followed by another go trial than a no-go trial. The ERPs on go trials had a characteristic triphasic pattern. Given that the trials are perceptually indistinguishable, it was not surprising that the early perceptual response in the EEG was similar whether the go trial occurred before an error or a correct no-go trial. A substantial difference was, however, apparent in the P300. When its amplitude was relatively low, it was associated with increased errors on subsequent no-go trials. When its amplitude was relatively high, participants were more likely to succeed in withholding their responses on subsequent no-go trials. As might be expected from this finding, the degree to which the amplitude of the P300 was maintained across the task was associated with individual error propensity among the participants.

It seems, therefore, that the P300 formed an electrophysiological marker of *something* that is probabilistically associated with subsequent error. It is tempting to view this as a fluctuating "top-down" goal-directed signal which, if it could speak, would be saying things like "watch out, don't press on the no-go trial, don't get distracted, keep focusing on the task, and so forth." However, the averaging of the ERPs to the onset of each trial makes it less likely that we are directly sampling the intensity of such a signal. Instead, we are more probably detecting the *consequence* of that maintained stance

in the attention/decision making allocation to each digit. In the SART, the really important presented digit is the one nominated as the no-go target. The others just mean that the current trial is not a no-go target and, when no-go trials are rare, arguably this encourages a stance in which evaluation becomes rather scant (and in which commission errors are more likely). This would be reflected in the reduced P300 to each digit presentation. The influence of a maintained goal-directed stance would be to resist this and encourage more active trial-by-trial decision making about the response with reference to the digit. This would be reflected in increased digit-onset locked P300 amplitude. The results are therefore consistent with many previous studies associating the P300 with increased attention to a particular stimulus (e.g., [19–23, 32]). The novel feature or argument here is in the relation of this individual stimulus processing to some more generally maintained executive stance to the task. More simply, it might be expressed as If there is a good attention at trial n (high P300), then it is more likely that there will be good attention at trial $n + 1$—which will be particularly useful if it happens to be a no-go trial.

There are a number of confounds or different interpretations of these findings which we have attempted to address. The first is that the results are an artefact of the different number of pre-error and pre-correct trials delivered to us by the participants. The actual number of trials contributing to the averages was, however, relatively high (between 80 and 147 for pre-error trials and between 103 and 170 for pre-correct trials). The averaging process should, therefore, have had a reasonable opportunity to reduce the contribution of random noise to near zero levels, and therefore a *differential* contribution of noise to the comparison should be minimal. In addition, and assuming that noise would be temporally as well as randomly distributed, the inference is strengthened by our focus on the P300 and the lack of marked difference in other components within the ERPs. Finally, in this respect, we further minimized the risk by expressing the magnitude of the P300 as a ratio of the P200. This process should further cancel any noise difference (in that the P200 should be equally susceptible) as well as offering other advantages in terms of normalizing the response. A second concern was that the P300 association with error was mediated by *previous* error detection. The observations that error detection has generally been associated with an *increase* in P300 (rather than the *decrease* that we see here associated with more errors), and that the "error-signal" is a rather short-lived phenomenon [34–36] both suggest that this account is unlikely. Furthermore, by comparing the correlation with overall error rates in go trials that immediately preceded no-go trials (which are as "far as you can get" from a previous no-go trial and hence error) with go trials in general, we found the P300 magnitude was increased, not decreased, with remoteness from an error. Finally, we found no significant differences in reaction times that could account for the results.

Despite this, we still need to be cautious. The statistically significant differences and correlations that we report are all from the Pz region. Although they were broadly in a consistent direction, the differences at Cz and Fz were less impressive. However, the site of biggest ERP signal difference may not be obviously connected to the origins of that difference and it seems improbable that the prefrontal cortex is not in some way involved in the allocation and maintenance of attention [32, 37–40]. It is also true that a plethora of functional imaging and other results now suggest that parietal regions tend to be coactivated with those of the dorso- and ventro-lateral prefrontal cortex in tasks requiring effortful or conscious processing [14]. While the current study may have little to say about the location(s) of the sources of the observed ERP differences, other studies may be more useful guides. Robertson et al. [32] examined ERP correlates of go no-go task performance in head-injured and healthy participants. As with our study, they found no significant differences between no-go and visually identical go trials in the early perceptual components in the ERP (up to and including the 200 milliseconds bin) in either group. For the healthy participants, increased amplitude at P300 did differentiate the trial types and was interpreted by the authors as reflecting increased attention and/or the launching of an inhibitory signal to prevent a response. In this respect, the healthy participants showed a greater differential response to no-go trials than the patient group, which may be reflected in their relatively lower error rates. Interestingly, in terms of our current discussion, Roche et al. identified two components in their P300. The P300 was reported to be maximal at the frontal electrode site whilst the slightly later P300b was of greater magnitude and most apparent at the Pz site. It is possible to question whether the common coactivation of frontal and parietal regions in effortful tasks which is commonly seen in functional magnetic resonance imaging (fMRI) studies reflects the simultaneous engagement of a large distributed network or whether, for example, parietal activity may be a secondary consequence of frontal activation. The combination of the good temporal resolution of ERPs and the spatial resolution of fMRI may be necessary to further address this question.

Electrophysiological measures such as this provide one route out of the conceptual circularity inherent in some purely behavioral analyses. Errors on the SART have previously been attributed to the poor maintenance of attention with that poor maintenance being marked by the occurrence of the error. This is a reasonable but circular argument that requires additional measures such as the frequency of attention problems in everyday life, self-reports of "task unrelated thought" propensity, and the effect of cues to maintain attention, if it is to be sustained [41]. An alternative, though not mutually exclusive, account might emphasise response inhibition efficiency as contributing towards errors in the task. Following Logan et al. [42] we might therefore view success or failure on a no-go trial as depending upon the outcome of a race between the erroneous "go" response and an inhibitory signal launched at the start of the trial. We would know if the internal "stop!" signal was a good or poor competitor based on the number of errors made and we would explain the number of errors made based on the hypothetical speed of this signal. Again, independent measures of response inhibition from other tasks or from everyday life would be required to avoid circularity. The advantage of the electrophysiological approach used here is that we can see that there is some influence at work *before* the critical no-go trial has been

presented. Whether or not a race model is accurate or appropriate (and there are good reasons to believe it is both), the results suggest that there is *something* in place biasing the odds of that race before it has begun. This seems to chime with everyday experience of inhibitory failures. Returning to the light bulb example, if one enters the dark room thinking "Concentrate... habit tells you to switch on the light but you know that, in this case, it will not help!", then—with luck—the action error is less likely.

This issue is not trivial as there are, as discussed, many clinical groups said to suffer from inhibitory deficits. In addition to the possibility of tweaking the efficiency of inhibitory control, perhaps pharmacologically, the results suggest that other interventions could serve to reduce the consequences of inhibitory difficulty. These would include programs designed to help people recognize situations in which a more attentive stance might offset inhibitory slips, training in maintaining such a stance, and the use of cues to externally support such maintenance when necessary. These programs would have application in rehabilitation of neurological patients and also assist situations where prolonged vigilance is vital such as in industrial or military scenarios. Although tasks such as the SART may be somewhat artificial models of aspects of everyday situations, their value lies in allowing close, controlled analysis of cognitive failures and, therefore, in refining understanding and evaluating interventions. They are also, in their repetitive structured way, compatible with the averaging over multiple similar events necessary for ERP analysis. We conclude that identifying EEG markers, such as the P300, which appear to reflect a well maintained top-down stance to a task therefore has multiple potential benefits in predicting and preventing potentially catastrophic errors in civilian and military life.

ACKNOWLEDGMENTS

Support from the MRC is acknowledged (MRC programme U.1055.01.003.00001.01). Support from the NIH (NHLBI RO1 HL079937-01) to Dr. Datta is acknowledged.

REFERENCES

[1] J. Reason, "Lapses of attention in everyday life," in *Varieties of Attention*, R. Parasuraman and D. R. Davies, Eds., Academic Press, Orlando, Fla, USA, 1984.

[2] G. E. Larson, D. L. Alderton, M. Neideffer, and E. Underhill, "Further evidence on dimensionality and correlates of the cognitive failures questionnaire," *British Journal of Psychology*, vol. 88, no. 1, pp. 29–38, 1997.

[3] R. A. Barkley, "Is there an attention deficit in ADHD?" *The ADHD report*, pp. 3–11, 1995.

[4] R. A. Barkley, "Response inhibition in attention-deficit hyperactivity disorder," *Mental Retardation and Developmental Disabilities Research Reviews*, vol. 5, no. 3, pp. 177–184, 1999.

[5] P. W. Burgess and T. Shallice, "Bizarre responses, rule detection and frontal lobe lesions," *Cortex*, vol. 32, no. 2, pp. 241–259, 1996.

[6] P. W. Burgess and T. Shallice, "Response suppression, initiation and strategy use following frontal lobe lesions," *Neuropsychologia*, vol. 34, no. 4, pp. 263–272, 1996.

[7] F. Lhermitte, "Human autonomy and the frontal lobes—part II: patient behavior in complex and social situations: the *environmental dependency syndrome*," *Annals of Neurology*, vol. 19, no. 4, pp. 335–343, 1986.

[8] A. R. Luria, *Higher Cortical Functions in Man*, Tavistock, London, UK, 1966.

[9] I. H. Robertson, T. Manly, J. Andrade, B. T. Baddeley, and J. Yiend, "'Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects," *Neuropsychologia*, vol. 35, no. 6, pp. 747–758, 1997.

[10] T. Shallice, *From Neuropsychology to Mental Structure*, Cambridge University Press, Cambridge, UK, 1988.

[11] T. Shallice and P. W. Burgess, "Deficits in strategy application following frontal lobe damage in man," *Brain*, vol. 114, no. 2, pp. 727–741, 1991.

[12] D. T. Stuss and C. A. Gow, "'Frontal dysfunction' after traumatic brain injury," *Neuropsychiatry, Neuropsychology and Behavioral Neurology*, vol. 5, no. 4, pp. 272–282, 1992.

[13] D. A. Norman and T. Shallice, "Attention to Action: Willed and automatic control of behaviour," Centre for Human Information Processing, 1980.

[14] S. Dehaene and L. Naccache, "Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework," *Cognition*, vol. 79, no. 1-2, pp. 1–37, 2001.

[15] T. Manly and I. H. Robertson, "Sustained attention and the frontal lobes," in *Methodology of Frontal and Executive Function*, P. Rabbitt, Ed., pp. 135–150, Psychology Press, Hove, UK, 1997.

[16] D. T. Stuss, T. Shallice, M. P. Alexander, and T. W. Picton, "A multidisciplinary approach to anterior attentional functions," *Annals of the New York Academy of Sciences*, vol. 769, no. 1, pp. 191–211, 1995.

[17] A. J. Wilkins, T. Shallice, and R. McCarthy, "Frontal lesions and sustained attention," *Neuropsychologia*, vol. 25, no. 2, pp. 359–365, 1987.

[18] T. Manly, I. H. Robertson, M. Galloway, and K. Hawkins, "The absent mind: further investigations of sustained attention to response," *Neuropsychologia*, vol. 37, no. 6, pp. 661–670, 1999.

[19] M. Eimer, "Effects of attention and stimulus probability on ERPs in a Go/Nogo task," *Biological Psychology*, vol. 35, no. 2, pp. 123–138, 1993.

[20] S. R. Jackson, G. M. Jackson, and M. Roberts, "The selection and suppression of action: ERP correlates of executive control in humans," *NeuroReport*, vol. 10, no. 4, pp. 861–865, 1999.

[21] L. Karlin, M. J. Martz, and A. M. Mordkoff, "Motor performance and sensory-evoked potentials," *Electroencephalography and Clinical Neurophysiology*, vol. 28, no. 3, pp. 307–313, 1970.

[22] A. Kok, "Effects of degradation of visual stimuli on components of the event-related potential (ERP) in go/nogo reaction tasks," *Biological Psychology*, vol. 23, no. 1, pp. 21–38, 1986.

[23] S. Mäntysalo, "N2 and P3 of the ERP to Go and Nogo stimuli: a stimulus-response association and dissociation," *Electroencephalography and Clinical Neurophysiology. Supplement*, vol. 40, pp. 227–234, 1987.

[24] L. Karlin, M. J. Martz, E. Brauth, and A. M. Mordkoff, "Auditory evoked potentials, motor potentials and reaction time," *Electroencephalography and Clinical Neurophysiology*, vol. 31, no. 2, pp. 129–136, 1971.

[25] T. W. Picton, "The P300 wave of the human event-related potential," *Journal of Clinical Neurophysiology*, vol. 9, no. 4, pp. 456–479, 1992.

[26] W. Ritter, R. Simson, and H. G. Vaughan Jr., "Association cortex potentials and reaction time in auditory discrimination," *Electroencephalography and Clinical Neurophysiology*, vol. 33, no. 6, pp. 547–555, 1972.

[27] E. Donchin, G. Gratton, D. Dupree, and M. Coles, "After a rash action: latency and amplitude of the P300 following fast guesses," in *Neurophysiology and Psychophysiology: Experimental and Clinical Applications*, G. C. Galbraith, M. L. Kietzman, and E. Donchin, Eds., pp. 173–188, Lawrence Erlbaum, Hillsdale, NJ, USA, 1988.

[28] M. Falkenstein, N. A. Koshlykova, V. N. Kiroj, J. Hoormann, and J. Hohnsbein, "Late ERP components in visual and auditory Go/Nogo tasks," *Electroencephalography and Clinical Neurophysiology*, vol. 96, no. 1, pp. 36–43, 1995.

[29] K. B. Campbell, S. Houle, D. Lorrain, D. Deacon-Elliott, and G. Proulx, "Event-Related Potentials as an index of cognitive functioning in head-injured outpatients," *Electroencephalography and Clinical Neurophysiology*, vol. 38, pp. 486–488, 1986.

[30] M. D. Rugg, C. P. Cowan, M. E. Nagy, A. D. Milner, I. Jacobson, and D. N. Brooks, "Event related potentials from closed head injury patients in an auditory 'Oddball' task: evidence of dysfunction in stimulus categorisation," *Journal of Neurology Neurosurgery and Psychiatry*, vol. 51, no. 5, pp. 691–698, 1988.

[31] S. J. Segalowitz, J. Dywan, and A. Unsal, "Attentional factors in response time variability after traumatic brain injury: an ERP study," *Journal of the International Neuropsychological Society*, vol. 3, no. 2, pp. 95–107, 1997.

[32] I. H. Robertson, T. Ward, V. Ridgeway, and I. Nimmo-Smith, "The structure of normal human attention: the test of everyday attention," *Journal of the International Neuropsychological Society*, vol. 2, no. 6, pp. 525–534, 1996.

[33] M. Falkenstein, J. Hohnsbein, J. Hoormann, and L. Blanke, "Effects of crossmodal divided attention on late ERP components—II: error processing in choice reaction tasks," *Electroencephalography and Clinical Neurophysiology*, vol. 78, no. 6, pp. 447–455, 1991.

[34] W. J. Gehring and R. T. Knight, "Prefrontal-cingulate interactions in action monitoring," *Nature Neuroscience*, vol. 3, no. 5, pp. 516–520, 2000.

[35] S. Nieuwenhuis, K. R. Ridderinkhof, J. Blom, G. P. H. Band, and A. Kok, "Error-related brain potentials are differentially related to awareness of response errors: evidence from an antisaccade task," *Psychophysiology*, vol. 38, no. 5, pp. 752–760, 2001.

[36] M. Ullsperger, D. Y. von Cramon, and N. G. Müller, "Interactions of focal cortical lesions with error processing: evidence from event-related brain potentials," *Neuropsychology*, vol. 16, no. 4, pp. 548–561, 2002.

[37] J. V. Pardo, P. T. Fox, and M. E. Raichle, "Localization of a human system for sustained attention by positron emission tomography," *Nature*, vol. 349, no. 6304, pp. 61–64, 1991.

[38] R. M. Cohen, W. E. Semple, M. Gross, H. H. Holcomb, M. S. Dowling, and T. E. Nordahl, "Functional localization of sustained attention: comparison to sensory stimulation in the absence of instruction," *Neuropsychiatry, Neuropsychology and Behavioral Neurology*, vol. 1, no. 1, pp. 3–20, 1988.

[39] J. S. Lewin, L. Friedman, D. Wu, et al., "Cortical localization of human sustained attention: detection with functional MR using a visual vigilance paradigm," *Journal of Computer Assisted Tomography*, vol. 20, no. 5, pp. 695–701, 1996.

[40] T. Manly, A. M. Owen, L. McAvinue, et al., "Enhancing the sensitivity of a sustained attention task to frontal damage: convergent clinical and functional imaging evidence," *Neurocase*, vol. 9, no. 4, pp. 340–349, 2003.

[41] J. Smallwood, J. B. Davies, D. Heim, et al., "Subjective experience and the attentional lapse: task engagement and disengagement during sustained attention," *Consciousness and Cognition*, vol. 13, no. 4, pp. 657–690, 2004.

[42] G. D. Logan, R. J. Schachar, and R. Tannock, "Impulsivity and inhibitory control," *Psychological Science*, vol. 8, no. 1, pp. 60–64, 1997.