Biomed Tech 2013; 58 (Suppl. 1) © 2013 by Walter de Gruyter · Berlin · Boston. DOI 10.1515/bmt-2013-4160

HIGH-LEVEL FEATURE EXTRACTION FROM ELECTROPHYSIOLOGICAL BRAIN SIGNALS IN THE TIME-FREQUENCY DOMAIN

René Heideklang and Galina Ivanova

Institut für Informatik, Humboldt-Universität zu Berlin, Deutschland

heidekla@informatik.hu-berlin.de

Abstract: *In this study, an application of modeling smooth patterns of signal power in the time-frequency domain is presented. This novel technique called SnaGe was recently developed and is applied to real EEG data here. In particular, its benefits for single trial analysis are highlighted. Models of varying degrees of freedom are fitted to the chosen data set, emphasizing the trade-off between goodness of fit and interpretability.*

Keywords: *Time-frequency analysis, Feature extraction, Model fitting, Signal analysis*

Introduction

The electroencephalogram (EEG) is the longest-studied noninvasive technique to measure brain signals. In the last decades, elaborate signal processing tools have been developed to automatically extract relevant information from the data. This task is quite challenging, owing to the complexity of the human brain. Indeed, each EEG measurement represents merely a snapshot of brain signals, but which are in fact varying: Even in a fixed experimental setting, the obtained signals exhibit considerable intra-individual and inter-individual biological variability. Yet, standard analysis techniques, such as time-frequency decompositions, do not appropriately account for this effect. For example, while they allow to study the signals at different time/frequency points, such precise information about regions of interest is often not available, e.g. because of the inherent variability. Usually, information from many neighboring time-frequency locations is therefore averaged, or the issue is simply ignored and the locations are treated as being independent. Both strategies, however, are not satisfactory: Either precise localization is lost, or the full potential of the data is not used. Thus, there is a need for adaptive signal representations which explicitly quantify the data *patterns* generated by the underlying biological processes.

In a previous paper, the authors published such an abstract representation called *Smooth Natural Gaussian Extension* (SnaGe) model [1]. Similarly to the well-known bivariate Gaussian function, this parametric model can be fitted to power distributions in the time-frequency domain, which are interpreted as images. Thereby obtained model parameters form high-level descriptors (*features*) about the data patterns. As these patterns may be too complex to be captured by a Gaussian peak, SnaGe is designed to represent whole paths of prominent signal power in the time-frequency domain. This path is essentially a smooth interpolating curve, and a surface is defined by exponential decay w.r.t. increasing distance from the path.

In this study, the previously developed and published technique is applied to a data set of real EEG measurements. The promising ability to analyze single trial recordings is demonstrated. To this end, an induced oscillation in the alpha band is investigated and the obtained high-level signal characteristics are presented.

Methods

An exemplary data analysis by means of the SnaGe method is carried out. The data set in this study stems from a motor experiment. Participants were asked to press a button with their index finger in reaction to a visual stimulus. EEG was measured at a sampling rate of 500 Hz from 128 channels according to the 10-20 system. Linked mastoids formed the reference channel. Response-locked trials were extracted, and after standard pre-processing, the signals are transformed to the time-frequency domain by the Short-Time Fourier Transform. The focus is on the investigation of induced neural activity in a region of interest around the alpha band (8–13 Hz), during the first four seconds after the response. A relevant channel localized near the motor and sensory-motor cortical area was selected for this demonstration. An example for a single trial signal is plotted in Fig. 1. A prominent pattern of event-related synchronization of neuronal activity is apparent in the visualization. In order to analyze the oscillation in the alpha band, different bivariate SnaGe models are fitted to the dynamic power spectrum image.

One fundamental decision to be made is the model order. While low-order models are robust to small perturbations of the signal, they might not be able to represent increasingly complex patterns. Higher-order models, on the other hand, are more prone to over-fitting, but offer greater flexibility. Therefore, in this study SnaGe models of order 5,6,7 and 8 are fitted. Also, as proposed in [1], the strategy of iteratively refining models of low order to yield more and more complex pattern representations can be useful for robust parameter extraction. Thus, in another experiment, an initial 5th order model is refined multiple times, which results in an additional 8th order representation. Lastly, the effects of taking into account a-priori knowledge about the data are studied. Because it is known that the neural excitement displays as a horizontal path of increased signal power in the time-frequency representation, paths going backward in time should be punished. This behavior can be prevented by adding a penalty term to the optimized cost function.

Results

Six SnaGe models are computed, whose predictions are shown in Fig. 1. To objectively compare these results, the goodness of fit is computed as the root mean squared error (RMSE) between the preprocessed time-frequency data and the models' predicted images, see Table 1.

All six models are able to approximate the pattern in the data, which is reflected by the overall low RMSE values. In general, higher-order models produce better matches, especially concerning values of signal power. Although the overall time-frequency distribution of signal power is almost always represented well, there are significant differences regarding the course of the modeled path (represented by white lines in Fig. 1). In particular, the 7th-order model and the refined model without regularization both exhibit loops in the SnaGe path. While such results impair interpretability as a path of instantaneous frequency, it is also clear that the obtained models are optimal in terms of goodness of fit. Among the other solutions which are free of loops, the 8th-order model is most accurate, followed by the refined regularized 8-th order result.

After meaningful parameters are extracted, they can form the basis for further analyses. As a simple demonstration, the mean frequency value along the modeled path is computed here. To this end, the most accurate loop-free model's curve is densely sampled, resulting in a mean value of 10.4 Hz. These observations are discussed in detail in the following section.

Table 1: Goodness of fit (gof) according to root mean squared error.

experiment	**gof**
order 5	67.18
order 6	59.20
order 7	**56.84**
order 8	56.94
order 8 (refined)	**55.59**
order 8 (refined, regularized)	57.59

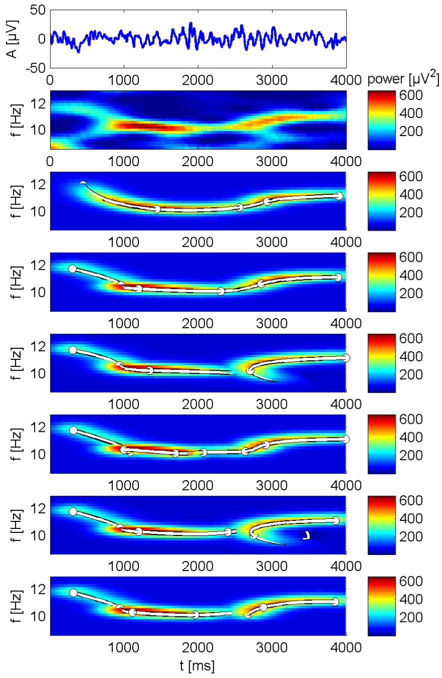

Figure 1: EEG single trial (top), time-frequency representation (2nd row) and several fitted SnaGe models (3rd to 8th row: order 5,6,7,8,8(refined),8(refined and regularized).

Discussion

The most accurate solutions were obtained by models of higher order, owing to the pattern's complexity. However, the more flexible models tend do develop loops in their paths, which shows that there usually is a trade-off between interpretability and goodness of fit.

The SnaGe is built on several model assumptions, most importantly it is designed for mono-component signals. This assumption can often be justified by filtering or masking other components, as it was done here by the application of a region of interest. However, the target pattern will never be ideal in this sense, for example there seems to be a peak off the main path of activation at 3000 ms / 9 Hz. Loops are the only means to compensate for such violated model assumptions and may be sources of over-fitting. In fact, this shows the optimization algorithm's strength to travel long distances in parameter search space to obtain very different, but in fact more optimal (w.r.t. the cost function, i.e. RMSE) solutions.

It was shown that the SnaGe model is able to accurately extract high-level features in the form of a time-frequency-power path of neural oscillatory activity. The stated trade-off can be controlled by adding penalty terms to the cost function. SnaGe is an appropriate tool for single trial analysis, since it is able to automatically adapt to the distribution of signal power and thus takes into account the biological variability. This suggests the method's potential as a measure of event related synchronization / desynchronization (ERS / ERD). Also, derived high-level features such as mean pattern frequency may be useful, for instance for the objective discrimination of groups of subjects in applied studies.

Bibliography

[1] R. Heideklang and G. Ivanova, "A novel flexible model for the extraction of features from brain signals in the time-frequency domain," *International Journal of Biomedical Imaging*, vol. 2013, pp. 1–12, 2013.