# Combining ESOMs Trained on a Hierarchy of Feature Subsets for Single-Trial Decoding of LFP Responses in Monkey Area V4

Nikolay V. Manyakov[1], Jonas Poelmans[2],
Rufin Vogels[1], and Marc M. Van Hulle[1]

[1] Laboratory for Neuro- and Psychofysiology, K.U.Leuven,
Herestraat 49, bus 1021, 3000 Leuven, Belgium
[2] Faculty of Business and Economics, K.U.Leuven,
Naamsestraat 69, 3000 Leuven, Belgium
{NikolayV.Manyakov,Rufin.Vogels,Marc.VanHulle}@med.kuleuven.be,
Jonas.Poelmans@econ.kuleuven.be

**Abstract.** We develop and combine topographic maps trained on different combinations of feature subsets for visualizing and classifying event-related responses recorded with a multi-electrode array chronically implanted in the visual cortical area V4 of a rhesus monkey. The monkey was trained, during consecutive training sessions, in a classical conditioning paradigm in which one stimulus was consistently paired with a fluid reward and another stimulus not. We opted for features from three categories: time-frequency analysis, phase synchronization between electrodes, and propagating waves in the array. The Emergent Self Organizing Map (ESOM) was used to explore the feasibility of single-trial decoding. Since the effective dimensionality of the feature space is rather high, a series of ESOMs was trained on features selected from different combinations of the three feature categories. For each trained ESOM, a classifier was developed, and classifiers of different ESOMs were combined so as to maximize the single-trial decoding performance.

## 1 Introduction

An event-related potential (ERP) is any stereotyped electrophysiological response to a stimulus. The last few years have witnessed an increasing interest in detecting ERPs, due to the development of brain-computer interfaces (BCIs). The problem of ERP detection requires a compromise between classification accuracy and response time, pushing the signal-to-noise ratio requirements to its limits. Indeed, in the ideal case, we would like to be able to detect an event based on a single ERP brain response, thus, without averaging. However, in practice, this is not very likely, since the ERP is a small signal in the ongoing brain activity. As a result, multiple presentations of the targeted stimulus are needed, and the corresponding time-locked responses averaged.

In this paper, we will deal with event-related local field potentials (LFPs) that represent the extracellular current flow due to the summed postsynaptic

L. Rutkowski et al. (Eds.): ICAISC 2010, Part II, LNAI 6114, pp. 548–555, 2010.

potentials of local cell groups [1]. The LFPs were recorded with a 96 electrode array implanted in the visual cortical area V4 of a rhesus monkey. The monkey was shown two stimuli, one for which he received a fluid reward, and another not (classical conditioning paradigm). It can be seen that the variability in the single LFP responses is large, and the difference between the responses to the two stimuli is not visible. But if we compute the averaged ERPs, we observe a clear difference in amplitude. As was shown in [2], the difference in the averaged ERP for the rewarded and unrewarded stimuli grows as a result of training, due to brain plasticity, from being identical in the beginning to markedly different at the end. But the question whether it is possible to detect differences in single trial recordings (thus, without any time-locked averaging) remains open. This question is important because it opens the way to perform real-time BCI. In this paper, we will concentrate on the possibility to detect and visualize differences in single trial LFP recordings using the Emergent Self Organizing Map (ESOM) [4].

The ESOM is a topographic map that differs from the traditional Self Organizing Map (SOM) [3] in that a very large number of neurons (at least a few thousands) is used [4]. According to [5], "emergence is the ability of a system to produce a phenomenon on a new, higher level." In order to achieve emergence, the existence and cooperation of a large number of elementary processes is necessary, so that large numbers of neurons can represent data clusters individually, which facilitates their detection. The ESOM is argued to be especially useful for visualizing sparse, high-dimensional datasets, yielding an intuitive overview of its structure [6]. In this paper we use, for the first time, the ESOM for visualizing and decoding intracranial recordings.

## 2 Experiment

We briefly summarize the experimental set-up; for more information, we refer to [2]. A rhesus monkey was implanted with a Utah array into the prelunate gyrus (area V4). The array measures $4 \times 4$ mm, and consists of $10 \times 10$ electrodes (4 of them are not connected). The local field potential signals were obtained by filtering the recorded signals between 0.3-250 Hz. The stimuli are displayed in Fig. 1. During training, a different sinusoidal noise background, that filled the display, was presented every 500 ms. At random intervals, a sinusoidal grating

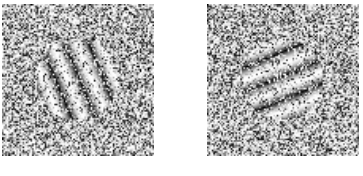

**Fig. 1.** Examples of rewarded (left panel) and unrewarded (right panel) stimuli. The sinusoidal grating has an orientation $157.5^o$ (rewarded) or $67.5^o$ (unrewarded), a spatial frequency of 2 c/deg, a diameter $4^o$ visual angle, and is superimposed on a sinusoidal noise background (S/N ratio=20%).

was presented for 500 ms. The reward was provided 400 ms after presentation of the grating pattern and, thus, partially overlapped with the grating presentation.

For our analysis, we looked only into the period of 300 ms after stimuli onset (thus, 100 ms before reward), to avoid the influence of the reward, and concentrate on ERP signals. Also, only those recordings of the rewarded and unrewarded stimuli were considered that were preceded by 3 background images, and that were observed by the monkey without failing to fixate a small dot on the screen. In this way, on average, 262 rewarded and 282 unrewarded trials per day were retained for further analysis.

## 3 Emergent Self Organizing Map

An Emergent Self Organizing Map (ESOM) map is composed of a set of formal (artificial) neurons $I$, arranged in a lattice or map. A neuron $i \in I$ is a tuple $(\mathbf{w}_i, \mathbf{x}_i)$ consisting of a weight vector $\mathbf{w}_i \in \mathbb{R}^n$ and a position $\mathbf{x}_i \in \mathbb{R}^2$ in the lattice. The input space $F$ is a metric subspace of $\mathbb{R}^n$. The training set $F^l = \{\mathbf{f}_1, ..., \mathbf{f}_N\}$ with $\mathbf{f}_1, ..., \mathbf{f}_N \in \mathbb{R}^n$ consists of input samples presented during ESOM training. We perform online training in which the best match for an input vector is searched for, and the corresponding weight vectors, and also those of its neighboring neurons of the map, are updated immediately.

When an input vector $\mathbf{f}_k$ is supplied to the training algorithm, the weight of neuron $i$ is modified as follows: $\Delta\mathbf{w}_i = \eta \cdot h(i, bm_k, R) \cdot (\mathbf{f}_k - \mathbf{w}_i)$, with $\eta \in [0, 1]$, and $bm_k$ the best-matching neuron of an input vector $\mathbf{f}_k$, *i.e.*, $bm_k = arg\min_{i \in I} \|\mathbf{f}_k - \mathbf{w}_i\|$, where $\|\mathbf{f}_k - \mathbf{w}_i\|$ is a distance in input space, according to some metric, between the input vector $\mathbf{f}_k$ and the weight vector $\mathbf{w}_i$ of the $i$th neuron; $h(i, bm_k, R)$ is a function that scales the update of neuron $i$ according to its position in a neighborhood with radius $R$ (according to some metric in the map) of the best-matching neuron $bm_k$. The ESOM training and visualization is done with the publicly available Databionic ESOM tool [7].

## 4 Features

The objective of feature selection is three-fold: 1) to improve the prediction performance of the classifier, 2) to provide classifiers with low computational complexitythat are also more cost-effective, and 3) to yield a better understanding of the underlying process that generated the data [8]. We considered the next feature subsets:

### 4.1 Time-Frequency Analysis

For the time time-frequency analysis, we used the continuous wavelet transformation (CWT) of the LFP signal $s(t)$, defined as:

$$W(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(t) \cdot \overline{\psi\left(\frac{t-b}{a}\right)}\, dt \tag{1}$$

where $b$ is a time shift, $a$ is a scale factor and $\psi$ is the predefined mother wavelet (we use the Daubechies wavelet db7) with zero mean, $\int_{-\infty}^{\infty} \psi(t)\, dt = 0$.

### 4.2 Phase Synchrony

As a second feature, we used an index of phase synchrony between the recorded LFPs, which was shown to be useful for decoding LFP recordings [9]. The phase from the prefiltered LFP signal $s(t)$ (in our case filtered below 30 Hz) is estimated according to $\varphi_s(t) = arctan\left(\frac{H(s(t))}{s(t)}\right)$, where $H(s(t))$ is the Hilbert transform of signal $s(t)$. As an index of bivariate phase synchrony between signals $x(t)$ and $y(t)$, we use the mean phase coherence, estimated as $\gamma_{x,y} = \sqrt{|\langle e^{i(\varphi_x(t)-\varphi_y(t))}\rangle|}$, where $\langle\cdot\rangle$ denotes the time average, and which measures how the relative phase difference is distributed over the unit circle. The result is in the interval $[0, 1]$, where 1 correspond to perfect synchrony.

### 4.3 Propagating Waves

As a third set of features, we took the directions and velocities of waves propagating in the array, since the presence of such phenomena was shown in [9,10]. In order to detect propagating waves, we used a phase-based estimate of the optical flow [11], in which case the electrode array should be regarded as a $10 \times 10$ pixel image, with each electrode corresponding to a pixel. By coupling the grey level of each pixel to the LFP of the corresponding electrode, the array becomes a frame in an image sequence (movie). To estimate the optical flow, the array's LFP recordings become a movie in the form $s(x, y, t)$, where $x$ and $y$ refer to the pixel's position and $t$ refers to time. The previously derived phases (see 4.2) are now referred to as $\varphi(x, y, t)$. The velocity of the coherent activity in the array is defined as the velocity of the lines of constant phase [11]. This velocity $\mathbf{v} = (dx/dt, dy/dt)$ is computed by taking the total derivative of $\varphi(x, y, t) = C$ with respect to time $d\varphi/dt = \bigtriangledown\varphi \cdot \mathbf{v} + \partial\varphi/\partial t$. The velocity direction, which is perpendicular to the lines of constant phase, is $-\bigtriangledown\varphi$.

## 5 Results

We considered, for our analysis, all electrodes for the phase synchrony and the wave features, but for the wavelet features only those of two electrodes (#18 and #80) that are representative for two groups of electrodes with similar LFP responses. As time-frequency features, we divide the wavelet scalograms (for 300 ms analyzed recordings, we applied a transformation for the scale factor $a$ ranging from 1 to 300) into $20 \times 20$ squares to avoid that too many redundant features would be taken. Within each square, the best separable features, based on the $t$-statistic, were chosen. Then, from the resulting 225 features, only the 15 best features (with smallest $p$-values) are retained. Hence, for the two electrodes considered, we have 15 best separable features for each electrode.

For the phase synchrony, we calculate the level of synchrony between all electrode pairs within non-overlapping windows of 20 ms (so, for each electrode pair, we have $300/20 = 15$ synchrony features). These features are then sorted

in terms of the separability between the two stimuli using the $t$-statistic. Only the 15 best phase synchrony features are retained.

As wave propagation features, the direction and magnitude of the optical flow in each electrode, for each moment after stimulus onset, are calculated. Among these features, also the 15 best separating ones are retained.

In total, we have 60 features for decoding. These features are developed for every training day and used for training separate ESOMs, not only for visualization purposes, but also for decoding the responses to the two stimuli. To test the decoding performance, the ESOM trained on a given day was used for predicting the decoding performance for the next day.

For days 1–36 of the monkey's training, we used the 2D ESOM with a toroid architecture sized $50 \times 82$ formal neurons, given all 60 features. ESOMs were trained for 20 presentations of the training set. As a distance metric, we used the Euclidian distance, both in data space and in lattice space. For every training sample, the weights of the winning neuron, and of the neurons in its neighborhood, were updated. The neighborhood function was a gaussian kernel. The neighborhood radius was linearly decreased during training from 24 to 1, and the learning rate from 0.5 to 0.1. Weights in the network were initialized by sampling the normal distribution. As a preprocessing, we transferred all features into $Z$-scores.

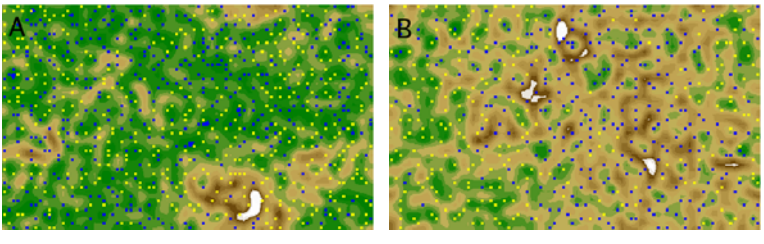


**Fig. 2.** ESOMs for monkey training day 1 (A) and day 36 (B) using all 60 features. The responses to the rewarded stimulus are shown as blue pixels, and those to the unrewarded one as yellow pixels. The background shows, for each neuron, the average distance of its weight vector to those of its lattice neighbors, normalized over all average distances in the lattice (U-matrix) (green: smallest distance; white largest distance).

In Fig. 2 A,B, the ESOMs for the first and last days of the monkey's training are shown. It can be seen how clusters for two separate groups were formed as a result of monkey training (note that no labeled information was used in training the ESOMs). As a measure of how well the clusters are formed, we opted for the standard deviation (in terms of distances) of the points in each class in a map $X$. Note that Euclidian distances were calculated by taking into account the toroidal structure. We plotted the mentioned measure as a function of the training days, for each stimulus [results not shown]. The result is that the standard deviation decreases over time, for both the rewarded and the unrewarded stimuli, which indicates that the clusters become more condensed.

We also calculated the average Euclidean distance between two consecutive responses to the same stimulus in the original feature space $F$ [results not shown]. We obtained that the average distance decreases over time for the rewarded stimulus, but not for the unrewarded one. This means that the responses to the consecutive rewarded stimulus vary less than to the unrewarded stimulus.

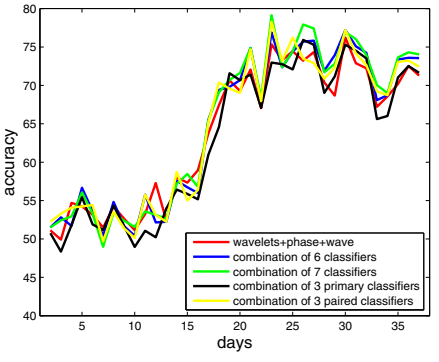


**Fig. 3.** Prediction accuracy of responses based on the map, trained in the previous day, for different combinations of classifiers

We also computed the prediction performance of the developed ESOMs, given all 60 features. For this, we used the weighted $k$NN classifier:

$$a(u; X^l, k) = arg \max_{y \in Y} \sum_{i=1}^{k} [y_{i,u} = y] \omega_i, \tag{2}$$

where $u$ is the coordinate of the best-matching neuron in the ESOM for each new item we want to classify; $X^l$ is the set of coordinates of the best-matching neurons for training examples $F^l$; $Y = \{-1, 1\}$ is the set of class labels (rewarded and unrewarded classes); $\omega_i = q^i$ $(q = 0.9)$ is a weight given to the $i$th closest labeled sample in the training set $X^l$; $y_{i,u}$ is the label of the $i$th closest sample with respect to $u$; $[a = b]$ is 1 when $a = b$ and 0 elsewhere. The parameter $k$ was determined using leave-one-out cross-validation on the training set.

Using the ESOM constructed for the previous day, we predict the stimulus type (rewarded or unrewarded) that corresponds to the response recorded on the current day. The results are shown in Fig. 3 (red line). We observe the increase in performance as a function of training days, which indicates the ESOM's potential to represent and distinguish different types of stimuli based on single trial recordings.

### Hierarchy of Feature Subsets

One can question whether the projection onto a 2 dimensional ESOM is valid. To this end, we calculated the inner or effective dimension of the manifold consisting

of the data point in feature space $F$. We opted for the Grassberger-Procaccia correlation dimension [12]. For points $\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_N$ in our preselected feature space with the Euclidian distance between the points $\|\mathbf{f}_i - \mathbf{f}_j\|$, and for a positive number $r$ (radius), we estimated the correlation sum $C(r)$ as the fraction of pairs whose distance is smaller than $r$ according to:

$$C(r) = \frac{2}{N(N-1)} \sum_{i<j} \Theta(r - \|\mathbf{f}_i - \mathbf{f}_j\|), \tag{3}$$

where $\Theta(x)$ is the Heaviside step function. From the known relation $C(r) \sim r^D$, one can estimate the manifold dimension $D$ as the slope of the straight part of the *log-log* plot. After determining the dependency of $C(r)$ on $r$, and calculating the manifold dimension for all training days, we found that the dimensions of the rewarded and unrewarded stimuli are always similar and decrease over time (from 19.078 in the first day to 12.202 in the last training day).

Since the effective dimension is always above 2, we can question the faithfulness of the projection onto a two-dimensional ESOM. One way to address this problem, without loosing the advantage of a two-dimensional visualization, is to develop several ESOMs, trained on particular subsets of features, and to combine their classification outcomes. Hence, we used a hierarchical feature grouping strategy, and trained ESOMs on wavelet-, phase-, and wave features separately (primary case, thus, 3 ESOMs), but also ESOMs trained on pairs of features (wavelets+phases, wavelets+waves, waves+phases, thus, 3 ESOMs). We consider the following possible combinations of classifiers: 1) the primary classifiers (combination of 3 primary classifiers), 2) the classifiers for the pairs of features (combination of 3 paired classifiers), 3) the classifiers for the primary and the paired features (combination of 6 classifiers), and 4) the classifiers for the primary-, paired-, and all features (combination of 7 classifiers).

For every trained map, for a particular day, the optimal value of $k$ in the $k$NN classifier was estimated by means of a leave-one-out cross-validation. The optimal value of $k$ for the $j$th ESOM, is defined as $k_o^j$, and the corresponding maximum leave-one-out cross-validation performance is defined as $M_j$. The outcomes of the classifiers of the different ESOMs were combined as follows. For any new sample to be classified, we estimate $a_j(y) = \sum_{i=1}^{k_o^j} [y_{i,u} = y]\omega_i$ (notation: see above) for each ESOM considered in the combination. Based on these results, we assign the outcome of $V_j(y) = \dfrac{|a_j(1) - a_j(-1)|}{a_j(1) + a_j(-1)}$ to the class label $y$ for which $a_j(y)$ is maximal, and $V_j(y) = 0$ to the second class, for all ESOMs that are combined. The resulting class label for the new sample, based on the combined classifiers, is then defined as follows: $arg\max_{y \in Y} \sum_{j=1}^{n} M_j V_j(y)$. The prediction performance for the current day samples, based on the combined classifiers and ESOMs constructed for the previous day, for different combinations of feature sets, are shown in Fig. 3. It can be observed that the performance for some combinations of feature sets is significantly (based on ANOVA) better than that for the whole set of features (from day 18 onwards): the combination of 7 classifiers is significantly different from the classifier based on the full set of

features ($p = 7.6 \cdot 10^{-5}$); the combination of 6 classifiers also ($p = 1.2 \cdot 10^{-3}$), and combination of the paired classifiers also ($p = 2 \cdot 10^{-3}$).

## 6 Conclusion

We have used the ESOM for visualizing and classifying event-related responses recorded with a multi-electrode array implanted in area V4 of a rhesus monkey. Since the effective dimensionality of the feature space is still quite high, one can question the relevance of using a two-dimensional ESOM. The solution we propose is to combine the classification outcomes of ESOMs trained on various feature subsets. We have shown that, in this way, a significantly better decoding performance is obtained.

**Acknowledgments.** NVM is supported by IST-2004-027017. MMVH is supported by research grants EF 2005, CREA/07/027, G.0588.09, IAP P6/29, GOA 2000/11, and IST-2007-217077.

## References

1. Buzsáki, G.: Large-scale recording of neuronal ensembles. Nature Neuroscience 5, 446–451 (2004)
2. Frankó, E., Seitz, A.R., Vogels, R.: Dissociable Neural Effects of Long-term Stimulus-Reward Pairing in Macaque Visual Cortex. Journal of Cognitive Neuroscience (to appear)
3. Kohonen, T.: Self-organizing maps. Springer, Heidelberg (1995)
4. Ultsch, A., Hermann, L.: Architecture of emergent self-organizing maps to reduce projection errors. In: Proc. ESANN 2005, pp. 1–6 (2005)
5. Ultsch, A.: Data Mining and Knowledge Discovery with Emergent Self-Organizing Feature Maps for Multivariate Time Series. In: Kohonen Maps, pp. 33–46 (1999)
6. Ultsch, A.: Density Estimation and Visualization for Data containing Clusters of unknown Structure. In: Proc. GfKl 2004 Dortmund, pp. 232–239 (2004)
7. Databionic ESOM Tools, `http://databionic-esom.sourceforge.net/`
8. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. Journal of Machine Learning Research 3, 1157–1182 (2003)
9. Manyakov, N.V., Van Hulle, M.M.: Synchronization in monkey visual cortex analysed with information-theoretic measure. Chaos 18, 037130 (2008)
10. Rubino, D., Robbins, K.A., Hatsopoulos, N.G.: Propagating waves mediate information transfer in the motor cortex. Nature Neuroscience 9(12), 1549–1557 (2006)
11. Fleet, D.J., Jepson, A.D.: Computation of component image velocity from local phase information. Int. J. Comput. Vis. 7, 77–104 (1990)
12. Grassberger, P., Procaccia, I.: Measuring the strangeness of strange attractors. Physica D 9(1-2), 189–208 (1983)