Unsupervised Identification of Event-Related Brain Potentials via Competitive Learning

Daniel H. Lange[1], Gideon F. Inbar[2], Hillel Pratt[3] and Hava T. Siegelmann[4]

[1,2] Dept. Elec. Eng., Technion - I.I.T., Haifa 32000, Israel

[3] ERP Lab, Gutwirth Bldg, Technion - I.I.T., Haifa 32000, Israel

[4] Dept. Ind. Eng., Technion - I.I.T., Haifa 32000, Israel

[1] E-mail: lange@ee.technion.ac.il

Abstract

We present a novel approach to the problem of Event-Related Potential (ERP) identification, based on a competitive Artificial Neural Net (ANN). Our approach dismisses the need for stimulus- or event-related selective averaging, thus avoiding conventional assumptions on response invariability. The identifier is applied to real event-related potential data recorded during a common odd-ball type paradigm. For the first time, within-session variable signal patterns are automatically identified, dismissing the strong and limiting requirement of a-priori stimulus-related selective grouping of the recorded data. The results present new possibilities in ERP research.

1 Introduction

1.1 Event-Related Potentials

The ongoing electrical activity of the brain, the EEG, is comprised of relatively slow fluctuations, in the range of 0.1 - 100 Hz, with magnitudes of 10 - 100 uV. Event-Related Potentials are characterized by overlapping spectra with the EEG, but with significantly lower magnitudes of 0.1 - 10 uV. The unfavorable Signal to Noise Ratio (SNR) requires filtering of the raw signals to enable analysis of the time-locked evoked brain responses. The most common method used for this purpose is signal averaging, synchronized to repeated occurrences of a specific event [1]. Averaging-based techniques assume a deterministic signal within the averaged session, and thus signal variability can not be modeled unless a-priori stimulus- or response-based categorization is available. It is the purpose of this paper to provide an alternative working method to enhance conventional averaging techniques by automatic identification of the variable signal patterns, thus facilitating the analysis of variable brain responses.

1.2 Competitive Learning

Competitive learning is an established branch of the general theme of unsupervised learning [2]. The elementary principles of competitive learning are [3]:

- Start with a set of units that are all the same except for some randomly distributed parameter which makes each of them respond slightly differently to a set of input patterns.

- Limit the "strength" of each unit.

- Allow the units to compete in some way for the right to respond to a given subset of inputs.

Applying these three principles yields a learning paradigm where individual units learn to specialize on sets of similar patterns and thus become "feature detectors".

2 The Competitive ANN

2.1 Theory

A typical architecture of a competitive learning network appears in Fig. 1. The network consists of a set of hierarchically layered neurons in which each layer is connected via excitatory connections to the following layer. Within a layer, the neurons are divided into sets of inhibitory clusters in which all neurons within a cluster inhibit all other neurons in the cluster, which results in a competition among the neurons to respond to the pattern appearing on the previous layer; the stronger a neuron responds to an input pattern, the more it inhibits the other neurons of its cluster.

There are many variations of the competitive learning scheme. We have selected a single layer structure, where the output neurons are fully connected to the input nodes and the non-linearity is implemented in the learning-phase only. The advantage of using this simple structure lies in enhanced analysis capabilities of the converged network, as the weights actually
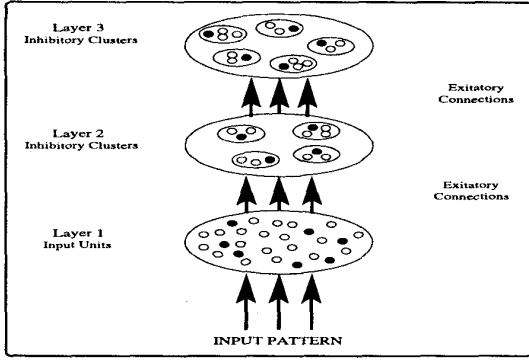
Figure 1: *Competitive learning takes place in a context of hierarchically layered units, which are presented as filled (active) and empty (inactive) dots. The winning neurons suppress the activity of neighboring neurons while exciting following layers.*

converge to the embedded signal patterns and thus form a Pattern Identification network. The general network structure is depicted in Fig. 2.
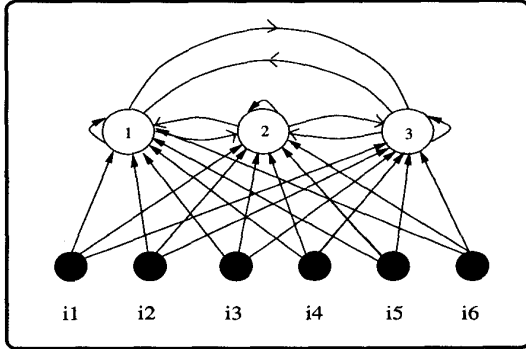

Figure 2: *Open arrows represent inhibitory connections and dark arrows represent excitatory connections. Lateral inhibition is used to suppress the activity of neighboring neurons.*

Let w_{ji} denote the synaptic weight connecting input node i to neuron j. Each neuron is given a fixed positive synaptic weight, which is distributed among its input nodes. A neuron learns by shifting synaptic weights from its inactive to active input nodes. If a neuron does not respond to some input pattern, no learning occurs in that neuron. According to the standard competitive learning rule, for a winning neuron, the change Δw_{ji} applied to synaptic weight w_{ji} is defined by:

$$\Delta w_{ji} = \eta(x_i - w_{ji}) \qquad (1)$$

where η is the learning rate coefficient. The effect of this rule is that the synaptic weight of a winning neuron is shifted towards the input pattern; thus once

converged, the network operates as a *matched filter bank* classifier [4].

2.2 Statistical Evaluation

2.2.1 Identification Property

The essential identification feature of the proposed network is, ideally, an inherent convergence of the network weights to the embedded ERP waveforms, thus operating as a Pattern Identification network. In an optimal scenario, in which the embedded ERP patterns and background EEG are uncorrelated, each of the competing neurons tends to *fixate* on a different signal type by mapping itself to a specific signal waveform [4]. Each single measurement can be represented as follows:

$$x_i(t) = \sqrt{E_i}s_i(t) + e_i(t) , \quad i = 1, 2, ..., N \qquad (2)$$

where x_i, E_i, s_i and e_i represent the recorded i-th single-sweep, the energy of the i-th EP, the normalized EP waveshape and the embedding background EEG, respectively. Assuming $P < N$ correctly identified signal categories, where $s_i \in \{S_j\}_{j=1}^P$, using normalized inputs and weights and a Gaussian model for the background EEG, it can be shown that in each iteration the winning neuron shifts its weights towards the respective EP pattern. First we calculate the neural outputs:

$$o^k = < x_i, w^k > , \quad k = 1..P$$

Then, selecting the winning neuron $l = argmax\{o^k\}$, we update the weights of the winning neuron only:

$$w_n^l = w_{n-1}^l + \eta \cdot (x_i - w_{n-1}^l) \qquad (3)$$

The winning neuron's output to a matching single-trial measurement is increasing monotonically (note: $|o^l(\cdot)| < 1$):

$$
\begin{aligned}
o_n^l &= < x_i, w_{n-1}^l + \eta \cdot (x_i - w_{n-1}^l) > &(4)\\
&= o_{n-1}^l + < x_i, \eta \cdot (x_i - w_{n-1}^l) >\\
&= o_{n-1}^l + \eta \cdot < x_i, x_i > -\eta \cdot < x_i, w_{n-1}^l >\\
&= o_{n-1}^l + \eta \cdot (1 - o_{n-1}^l)
\end{aligned}
$$

and thus:

$$o_n^l \geq o_{n-1}^l \qquad (5)$$

The pattern identification procedure is unbiased and the variance could be made as small as desired by decreasing the learning-rate coefficient, as presented in the following sections.

2.2.2 Identification Bias

The competing neurons are mapped to the input space, and the correlation of the neurons' weights

with their matching input patterns is ever increasing, i.e. the learning processes of the competing neurons are assumed to be independent. It is thus sufficient to evaluate a single neuron system detecting a constant signal pattern embedded within random noise realizations. Recalling the competitive learning rule, applied to the winning neuron, we have:

$$w_n = w_{n-1} + \eta \cdot (x_i - w_{n-1}) \quad ; \quad 0 < \eta << 1 \quad (6)$$

where x_i is an arbitrary input vector. Rearranging and using the additive signal and noise model yields:

$$w_n = w_{n-1} \cdot (1 - \eta) + \eta \cdot (s + e_i), \quad (7)$$

where s and e_i represent the embedded signal pattern and the embedding noise realizations respectively. Taking the limit as n approaches infinity, where $0 \leq \eta \leq 1$, $\tilde{e}_i = e_{n-i}$, yields the following solution [4]:

$$w_\infty = s + \eta \cdot \sum_{i=0}^{\infty} (1 - \eta)^i \cdot \tilde{e}_i \quad (8)$$

and calculating the expectancy provides the unbiased result:

$$E[w_\infty] = s + \eta \cdot \sum_{i=0}^{\infty} (1 - \eta)^i \cdot E[\tilde{e}_i] \quad (9)$$

$$E[w_\infty] = s \quad (10)$$

2.2.3 Identification Variance

Assuming zero-mean Gaussian EEG with σ^2 variance, and rewriting the solution of the learning rule equation (n indicates the iteration index):

$$w_n = s + \eta \cdot \sum_{i=0}^{n-1} (1 - \eta)^i \cdot \tilde{e}_i, \quad (11)$$

we can calculate the identification variance. Taking the limit as n approaches infinity yields the asymptotic identification variance (I denotes the identity matrix) [4]:

$$E(w_\infty - s)(w_\infty - s)^T = \frac{\eta}{2 - \eta} \cdot \sigma^2 I \quad (12)$$

2.2.4 Bounding the Learning Rate

The weight fluctuations due to the variance of estimation can be bounded as follows. Requiring $C_{ww} \leq \alpha \cdot E \cdot I$ where C_{ww}, α and E are the estimation covariance matrix, the distortion coefficient, and the energy of the signal, yields the bound:

$$\eta \leq \frac{2\alpha \frac{E}{\sigma^2}}{1 + \alpha \frac{E}{\sigma^2}} \quad (13)$$

e.g. for SNR's of 0dB,-20dB, and -40dB, and for a distortion coefficient of $\alpha = 0.1$, the learning rate coefficient should not exceed 0.18, 0.0198 and 0.002 respectively (see Fig. 3).
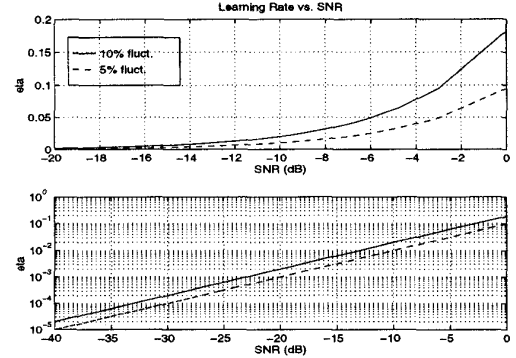

Figure 3: *Maximum Learning Rate as a function of the SNR, at noise fluctuation levels of 10% (solid) and 5% (dashed) of the signal energy (top: linear scale, bottom: logarithmic scale).*

3 Experimental Results

3.1 Experimental Paradigm

Cognitive event-related potential data were acquired during an odd-ball type paradigm from electrode Pz referenced to the mid-lower jaw, with a sample frequency of 250 Hz. The subject was exposed to repeated visual stimuli, consisting of the digits '3' and '5', appearing on a PC screen. The subject was instructed to press a push-button upon the appearance of '5' – the *Target* stimulus, and ignore the appearances of the digit '3' [5].

With odd-ball type paradigms, the Target stimulus is known to elicit a prominent positive component in the ongoing brain activity, related to the identification of a meaningful stimulus [6]. This component has been labeled P_{300}, indicating its polarity (positive) and timing of appearance (300 ms after stimulus presentation). The parameters of the P_{300} component (latency and amplitude) are used by neurophysiologists to assess, among other, effects related to the relevance of stimulus and level of attention (e.g. [7]).

3.2 Identification Results

The competitive network was trained with 80 input vectors, half of which were Target ERP's and the other half were Non Target. The network converged after approximately 300 iterations (per neuron). A sample of two single-trial post-stimulus sweeps, of the Target and Non-Target averaged ERP templates

and of the NN identified signal categories, are presented in Fig. 4. The automatic identification procedure has provided two signal categories, resembling the stimulus-related selective averaged signals, but requiring further examination as to the source of the slight differences between the selectively averaged waveforms and the categorization obtained by the ANN. The categorization process was consequently repeated, this time using Target and Non-Target data separately; the results are presented in Fig. 5. The categorization of Target data yielded 3 ERP patterns, increasing in latency and corresponding to our previous findings of increased latency with prolonged reaction times [5]. Non-Target ERP analysis yielded Target-like P_{300} waveform meaning that, at least occasionally, Target-like P_{300} appears even with Non-Target stimuli. This accounts for the above differences and obviously requires further investigation as to the reliability of selective event related data averaging when applied to cognitive brain function analysis.

Figure 5: *Top: Target and Non-Target ERP's. Bottom: The ANN categorizations. The categorized Non-Target patterns include a P_{300}-like waveform (dashed) indicating that some of the Non-Target trials may include a target-like P_{300} contribution.*
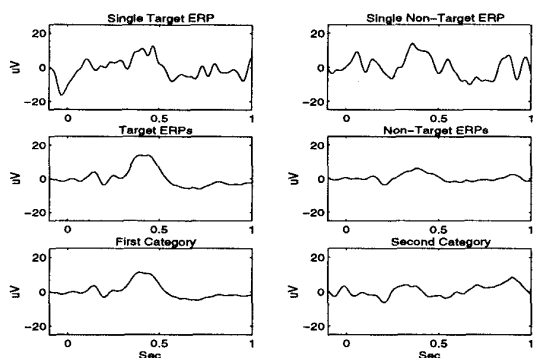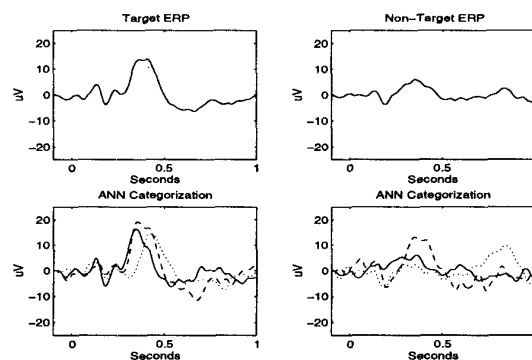

Figure 4: *Top: sample raw Target and Non-Target sweeps. Middle: Target and Non-Target ERP templates. Bottom: the NN categorized patterns.*

4   Conclusion

It has been shown that variable ERP waveforms can be identified and extracted from noisy realizations, overcoming the common assumption of response invariability which is essential for stimulus-related selective averaging. The identification process was evaluated statistically substantiating its credibility.

The experimental study presented an unsupervised identification of the raw Target and Non-Target responses, dismissing the requirement of stimulus- or event-related selective data grouping. The result is two-fold: (a) the identified patterns generally resemble conventional selective-average analysis, however (b) the obtained differences have been identified to be the result of unexpected appearance of P_{300}-like responses in the Non-Target data, further validating

the method and presenting its added value compared to conventional average-based analysis.

References

[1] J.I. Aunon, C.D. McGillem, and D.G. Childers, "Signal Processing in Evoked Potential Research: Averaging and Modeling," *CRC Crit. Rev. Bioeng.*, vol. 5, pp. 323-367, 1981.

[2] R.O. Duda and P.E. Hart. *Pattern Classification and Scene Analysis.* Wiley: New-York, 1976.

[3] D.E. Rumelhart and D. Zipser, "Feature Discovery by Competitive Learning," *Cognitive Science*, vol. 9, pp. 75-112, 1985.

[4] D.H. Lange, H.T. Siegelmann, H. Pratt, and G.F. Inbar, "Overcoming Selective Ensemble Averaging: Unsupervised Identification of Event Related Brain Potentials," submitted for publication to *IEEE Trans. Biomed. Eng.*

[5] D.H. Lange, H. Pratt, and G.F. Inbar, "Modeling and Estimation of Single Evoked Brain Potential Components," *IEEE Trans. Biomed. Eng.*, vol. BME-44, pp. 791-799, 1997.

[6] T.W. Picton and S.A. Hillyard, "Endogenous Event Related Potentials," in *Handbook of Electroencephalographic Clinical Neurophysiology,"* vol. 3, T. W. Picton, Ed. Amsterdam: Elsevier, 1988, pp. 361-426.

[7] D.H. Lange, H. Pratt, and G.F. Inbar, "Segmented Matched Filtering of Single Event Related Evoked Potentials," *IEEE Trans. Biomed. Eng.*, vol. BME-42, pp. 317-321, 1995.