Overcoming Selective Ensemble Averaging: Unsupervised Identification of Event-Related Brain Potentials

Daniel H. Lange, Hava T. Siegelmann, Hillel Pratt, and Gideon F. Inbar

***Abstract*—We present a novel approach to the problem of event-related potential (ERP) identification, based on a competitive artificial neural network (ANN) structure. Our method uses ensembled electroencephalogram (EEG) data just as used in conventional averaging, however without the need for *a priori* data subgrouping into distinct categories (e.g., stimulus- or event-related), and thus avoids conventional assumptions on response invariability. The competitive ANN, often described as a *winner takes all* neural structure, is based on dynamic competition among the net neurons where learning takes place only with the winning neuron. Using a simple single-layered structure, the proposed scheme results in convergence of the actual neural weights to the embedded ERP patterns.**

The method 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.

***Index Terms*—Artificial neural network, evoked electrical comprehensive learning.**

I. Introduction

A. Evoked Potentials

Ever since H. Berger's discovery, that the electrical activity of the brain can be measured and recorded via surface electrodes mounted on the scalp [1], there has been major interest in the relationship between such recordings and brain function. The first recordings conducted by Berger and his followers were concerned with the spontaneous electrical activity of the brain, appearing in the form of rhythmic voltage oscillations, which later received the term "electroencephalogram" or "EEG." More recent research has concentrated on time-locked brain activity, related to specific events, external or internal to the subject. Such evoked signals, also referred to as event-related potentials (ERP's), are regarded as manifestations of brain processes related to preparation for or in response to discrete events that are meaningful to the subject (e.g., [2]–[4].

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 μV. ERP's are characterized by overlapping spectra with the EEG, but with significantly lower magnitudes of 1–10 μV. The processing method described herein is applicable to a range of variable magnitude brain responses, such as visual-, cognitive-, and movement-related ERP's, whose signal-to-noise ratio (SNR) ranges from 0 dB downto -15 dB.

The unfavorable 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 [5]. 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. However, brain research has evolved to a point where interest is arising to analyze nonstationary brain processes, where deterministic repeating responses are not likely (e.g., [6]). In addition, cognitive neuroscience introduces experiments involving learning, habituation, and memory handling, where again the evoked brain responses need to be classified and analyzed on a trial-to-trial basis. 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.

B. Competitive Learning

Competitive learning is an established branch of the general theme of unsupervised learning [7]. The elementary principles of competitive learning are [8] as follows:

- 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." Competitive learning is a mechanism well-suited for regularity detection [9], where there is a population of stimulus patterns each of which is presented with some probability. The detector is supposed to discover statistically salient features of the input population, without requiring an *a priori* set of categories into which the patterns should be classified. Thus the detector needs to develop its own featural representation of the population of input patterns capturing its most salient features.

Finally, it is worth noting that competitive representations have some generic disadvantages over distributed representations [10]: they need one output neuron for each category, thus, N neurons can model only N categories, compared to 2^N for a binary code; they are not robust to neuron failure, which would cause loss of the whole respective category; and they cannot represent hierarchical knowledge—there is no way to have categories within categories (unless the *winner takes all* principle is relaxed).

C. Problem Statement

The major problem lies in the extremely unfavorable SNR of the evoked responses embedded within the ongoing background brain activity. Classification and estimation of the single evoked responses are, thus, difficult tasks, further complicated due to nonstationarities of the signal and noise.

A common assumption among most researchers is that the measured waveform is the sum of a signal component (ERP) and a statistically independent noise component (EEG). This is more of a definition than an assumption, since it is only natural to define the signal as the component which is correlated with the applied stimulus [11]. It should be noted that a different hypothesis was also proposed, referring to the phase spectrum of the post-stimulus EEG; while such phase values are random at the absence of stimulus, aggregated phase values appear with repeating stimulus presentation [12]. In practice, however, identical stimuli do not necessarily evoke identical responses [13]; trial-to-trial

Manuscript received February 9, 1998; revised February 15, 2000. *Asterisk indicates corresponding author*.

*D. H. Lange was with the Department of Electrical Engineering, Technion University—IIT Haifa 32002 Israel. He is now with R&D Medical Ltd., 2 HaMada Street, Synel Building, Yokneam, P.O. Box 338, Yokneam 20692, Israel (e-mail: lange@rdmed.com).

H. T. Siegelmann is with the Department of Industrial Engineering, Technion University—IIT Haifa 32002 Israel.

H. Pratt is with the Medical School, Technion University—IIT Haifa 32002 Israel.

G. F. Inbar is with the Department of Electrical Engineering, Technion University—IIT Haifa 32002 Israel.

Publisher Item Identifier S 0018-9294(00)04419-0.

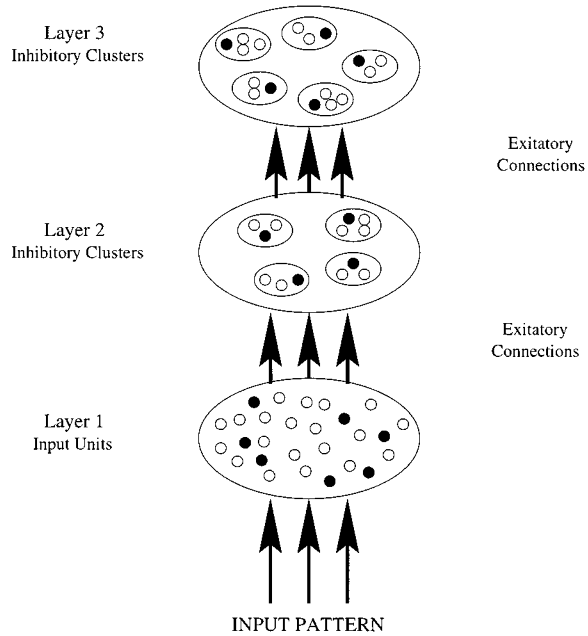

Fig. 1. The architecture of a competitive learning structure. Competitive learning takes place in a context of hierarchically layered units, which are presented as filled (active) and empty (inactive) dots. The winning neurons suppress the activity of neighboring neurons while exciting following layers.

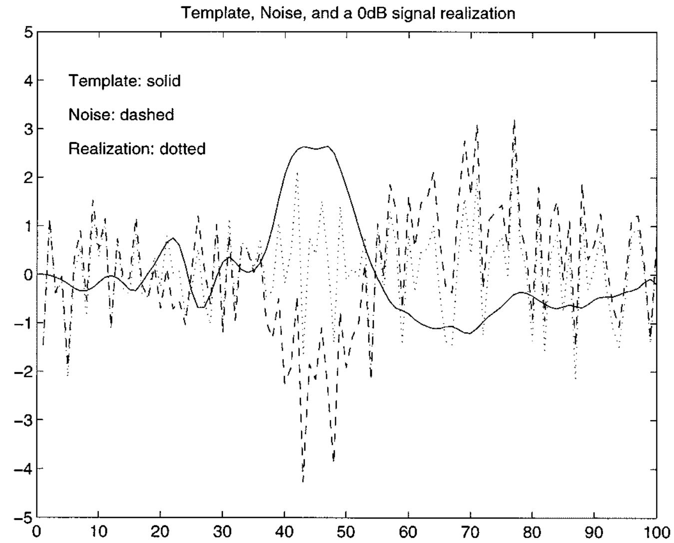

Fig. 2. A sample single realization (dotted) and its constituents (signal-solid, noise-dashed). SNR = 0 dB. Such realizations, at variable SNR levels, are used in the simulation study.

variability can be substantial, and ERP waveform, amplitude, and latency can change abruptly or progressively in time. Thus, the basic assumption underlying signal averaging and spectral analysis is generally violated. Several ERP classification procedures have been recently proposed (e.g., [14]–[16]), however none of them can actually identify embedded variable ERP waveforms without prior classification or clustering of the ensembled data and are, thus, unsuitable for on-line implementation.

The complicated, generally unknown relationships between the stimulus and its associated brain response, and the extremely low SNR of the brain responses which are practically *masked* by the background brain activity, make the choice of a self organizing structure for post-stimulus epoch analysis most appropriate. The competitive network, implemented so that its weights directly converge to the actual embedded signal patterns while inherently averaging out the additive background EEG is, thus, a powerful development [17], lending itself to real time on-line implementations.

II. The Competitive Neural Network Structure

A. Theory

A competitive neural 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 nonlinearity 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 converge to the embedded signal patterns and, thus, form a pattern identification network. The general network structure is depicted in Fig. 1.

For neuron j to be the winning neuron, its net internal activity level v_j for a specified input pattern x_i must be the largest among all neurons in the network. The output signal y_j of a winning neuron j is set equal to one, and all other neural outputs that lose the competition are set equal to zero.

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

$$\sum_i w_{ji}^2 = 1, \quad \text{for all } j. \tag{1}$$

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. When a single neuron wins the competition, each of its input nodes give up some proportion of its synaptic weight, which is distributed equally among the active input nodes. According to the standard competitive learning rule, for a winning neuron, the change Δw_{ji} applied to synaptic weight w_{ji} is defined by

$$\Delta w_{ji} = \eta(x_i - w_{ji}) \tag{2}$$

where η is the learning rate coefficient. The effect of this rule is that the synaptic weight of a winning neuron is shifted toward the input pattern; thus, in each learning cycle, the weights of the single winning neuron actually move toward the respective input pattern. Assuming zero-mean additive background EEG, once converged, the network operates as a *matched filter* bank classifier.

B. Simulaton Study

A simulation study was carried out to assess the performance of the competitive network classification system. A moving average (MA) process of order eight (selected according to the AIC condition applied to ongoing EEG [18]), driven by a deterministic realization of a Gaussian white noise series, simulated the ongoing background activity $x(n)$. An average of 40 single-trials from a cognitive odd-ball type experiment (to be explained in the Experimental Study), was used as the signal $s(n)$. Then, five 100-trial ensembles were synthesized,

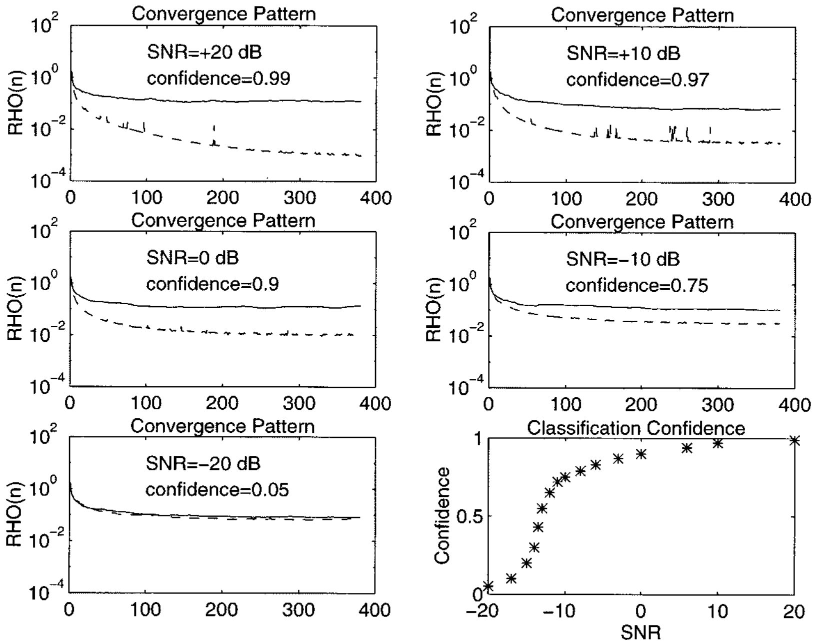

Fig. 3. Dynamic representation of the learning process. The solid line refers to the noise-only trials and the dashed line refers to the signal plus noise realizations. The convergence patterns and classification confidence values are shown for five SNR levels. The classification confidence breaks down with the fall of SNR, exhibiting a sharp fall in the range of -10 to -20-dB. RHO refers to (22), and the convergence pattern horizontal axes correspond to the number of iterations.

TABLE I
CLASSIFICATION RESULTS

	Pos	Neg	FP	FN
snr=+20dB	100%	100%	0%	0%
snr=+10dB	100%	100%	0%	0%
snr= 0dB	100%	100%	0%	0%
snr=-10dB	88%	92%	8%	12%
snr=-20dB	58%	54%	46%	42%

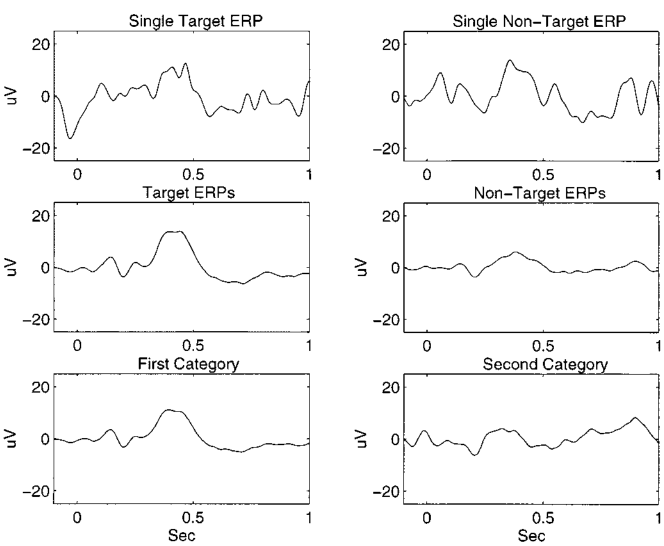

Fig. 4. Stimulus-related selective averaging versus spontaneous categorization. Top row: sample raw target and nontarget sweeps. Middle row: target and nontarget ERP templates. Bottom row: the neural-network categorized patterns. The spontaneously categorized ERP's appear similar to the stimulus-related averages.

to study the network performance under variable SNR conditions. A sample realization and its constituents, at an SNR of 0 dB, is shown in Fig. 2. The simulation included embedding the signal $s(n)$ in the synthesized background activity $x(n)$ at five SNR levels (-20, -10, 0, $+10$, and $+20$ dB), and training the network with 750 sweeps (per SNR level). Fig. 3 shows the convergence patterns and classification confidences of the two neurons, where it can be seen that for SNR's lower than -10 dB the classification confidence declines sharply.

The classification results, tested on 100 input vectors, 50 of each category, for each SNR, are presented in Table I; due to the competitive scheme, Positives and False Negatives as well as Negatives and False Positives are complementary. These empirical results are in agreement with corresponding analytical results presented in [19], where extensive statistical analysis of the proposed method is included.

III. Experimental Study

A. Motivation

An important task in ERP research is to identify effects related to cognitive processes triggered by meaningful versus nonrelevant stimuli (e.g., [4]). A common procedure to study these effects is the classic *odd-ball* paradigm, where the subject is exposed to a random sequence of stimuli and is instructed to respond only to the task-relevant (target) stimuli. Typically, the brain responses are extracted via selective averaging of the recorded data, ensembled selectively according to stimulus context. This method of analysis assumes that the brain responds equally to the members of each type of stimulus; however the validity of this assumption is unknown in the above case where cognition itself is being studied. Using our proposed approach, *a priori* grouping of the recorded data is not required, thus, overcoming the above severe assumption on cognitive brain function. The results of applying our method are described below.

B. Experimental Paradigm

Cognitive event-related potential data were acquired during an odd-ball type paradigm from electrode Pz referenced to the mid-lower jaw [20], 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 (20% of total stimuli), and ignore the appearances of the digit "3" (80% of stimuli) [21].

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. 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., [21]).

C. Identification Results

The competitive network was trained with 80 input vectors, half of which were target ERP's and the other half were nontarget. The network converged after approximately 300 iterations (per neuron), yielding a reasonable confidence coefficient of 0.7 [19]. A sample of two single-trial post-stimulus sweeps, of the target and nontarget averaged ERP templates and of the ANN 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 nontarget 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 [6]. Nontarget ERP analysis yielded **target-like** P_{300} waveform meaning that, at least occasionally, target-like P_{300} appears even with nontarget 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.

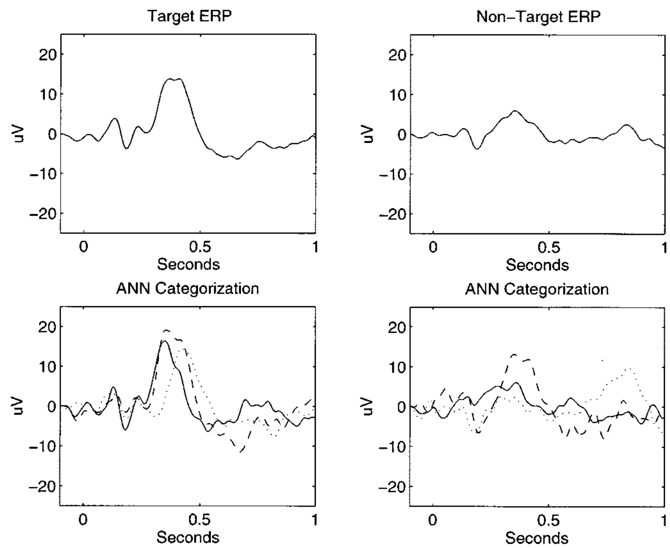

Fig. 5. Spontaneous categorization of separated target and nontarget ERP. (top) target and nontarget ERP (bottom). The neural network categorizations into three categories marked with solid, dashed, and dotted lines. The categorized nontarget patterns include a P_{300}-like waveform (dashed) indicating that some of the nontarget trials may include a target-like P_{300} contribution.

IV. Discussion

We have shown via simulation as well as with real ERP data 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 simulation study demonstrated the powerful capabilities of the proposed network in identifying and classifying the low amplitude signals embedded within the large background noise. The detection performance declined rapidly for SNR's lower than -10 dB. Empirically, high identification performance was maintained with SNR's of down to -10 dB, yielding confidences in the order of 0.7 or higher; thus, the method is applicable to a range of variable magnitude ERP's, such as visual- (0 dB), cognitive- (-5 dB), and movement-related (-10 dB) ERP's.

The experimental study presented an unsupervised identification and classification of the raw data into target and nontarget responses, dismissing the requirement of stimulus- or event-related selective data grouping. The result is twofold: 1) the identified patterns generally resemble conventional selective-average analysis, however 2) the obtained differences have been identified to be the result of unexpected appearance of P_{300}-like responses in the nontarget data, further validating the method and presenting its added value compared to conventional average-based analysis. The presented results indicate that the noisy single-trial brain responses may be identified and classified objectively in cases where relevance of the stimuli is unknown or needs to be determined (e.g., in lie-detection scenarios [22], and in man-machine communication [23]).

V. Conclusion

Common ensemble averaging suffers from two related main drawbacks; first, it requires *a priori* data categorization (e.g., according to type of stimulus or response), to increase coherence within each category. However, such categorization is not unbiased, as it assumes that a single experimental parameter controls all the experimental variance. This is obviously an oversimplification of brain function modeling, which might introduce erronous results, as demonstrated in this paper. Second, the analysis of variable responses, that may well be due to a gradual change in some known experimental parameter, is greatly limited with conventional averaging. Both of these drawbacks are treated in this work, presenting a powerful tool for automatic processing of the noisy single-trial data utilizing the actual embedded signals for spontaneous categorization and identification of the hidden ERP patterns. Overcoming stimulus-related selective averaging with a self-learning structure provides objective insight into brain function and, thus, opens new possibilities in ERP research.

References

[1] H. Berger, "On the electroencephalogram of man," in *Psychophysiology*, S. W. Porges and M. G. H. Coles, Eds, PA: Dawson, Hutchinson and Ross, Inc., 1967, pp. 9–14.

[2] G. McCarthy and E. Donchin, "A metric for thought: A comparison of R_{300}] latency and reaction time," *Science*, vol. 211, pp. 77–80, 1981.

[3] T. W. Picton and S. A. Hillyard, "Endogenous event related potentials," in *Handbook of Electroencephalographic Clinical Neurophysiology*, T. W. Picton, Ed. Amsterdam, The Netherlands: Elsevier, 1988, vol. 3, pp. 361–426.

[4] H. Pratt, H. J. Michalewski, G. Barrett, and A. Starr, "Brain potentials in memory-scanning task: Modality and task effects on potentials to the probes," *Electroencephalogr. Clin. Neurophysiol.*, vol. 72, pp. 407–421, 1989.

[5] 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.

[6] 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.

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

[8] D. E. Rumelhart and D. Zipser, "Feature discovery by competitive learning," *Cognitive Science*, vol. 9, pp. 75–112, 1985.

[9] S. Haykin, *Neural Networks: A Comprehensive Foundation*. New York: Macmillan College, 1994.

[10] J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation*. Reading, MA: Addison-Wesley, 1991.

[11] A. S. Gevins, "Analysis of the electromagnetic signals of the human brain: Milestones, obstacles, and goals," *IEEE Trans. Biomed. Eng.*, vol. BME-31, pp. 833–850, 1984.

[12] H. A. Beagly, B. M. Sayers, and A. J. Ross, "Fully objective ERA by phase spectral analysis," *Acta Otolaryngol.*, vol. 87, pp. 270–278, 1979.

[13] D. Popivanov and I. Krekule, "Estimation of homogeneity of a set of evoked potentials with respect to its dispersion," *Electroencephalogr. Clin. Neurophysiol.*, vol. 55, pp. 606–608, 1983.

[14] J. M. Moser and J. I. Aunon, "Classification and detection of single evoked brain potentials using time-frequency amplitude features," *IEEE Trans. Biomed. Eng.*, vol. BME-33, pp. 118–127, 1986.

[15] A. S. Gevins, N. H. Morgan, S. L. Bressler, J. C. Doyle, and B. A. Cutillo, "Improved event related potential estimation using statistical pattern classification," *Electroencephalogr. Clin. Neurophysiol.*, vol. 64, pp. 177–186, 1986.

[16] G. Zouridakis, B. H. Jansen, and N. N. Boutros, "A fuzzy clustering approach to EP estimation," *IEEE Trans. Biomed. Eng.*, vol. 44, pp. 673–680, 1997.

[17] D. H. Lange, H. T. Siegelmann, H. Pratt, and G. F. Inbar, "A generic approach for identification of event related brain potentials via a competitive neural network structure," presented at the NIPS '97—Neural Information Processing Systems, Denver, CO, 1998.

[18] H. Akaike, "Statistical predictor identification," *Ann. Inst. Stat. Math.*, vol. 22, pp. 203–217, 1970.

[19] D. H. Lange and G. F. Inbar, "Modern techniques in ERP research," in *Modern Techniques in Neuroscience Research*, U. Windhorst and H. Johansson, Eds. Berlin, Germany: Springer-Verlag, 1999, ch. 36, pp. 997–1023.

[20] H. H. Jasper, "The ten-twenty electrode system of the international federation," *Electroenc. Clin. Neurophysiol.*, vol. 10, pp. 371–375, 1958.

[21] D. H. Lange, H. Pratt, and G. F. Inbar, "Segmented matched filtering of single event related evoked potentials," *IEEE Trans. Biomed. Eng.*, vol. 42, pp. 317–321, 1995.

[22] D. H. Lange and G. F. Inbar, "Brain wave-based polygraphy," in *Proc. IEEE EMBS'96—The 18th Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society*, Amsterdam, The Netherlands, Oct. 1996, pp. 560–561.

[23] L. A. Farwell and E. Donchin, "Taking off the top of your head: Toward a mental prosthesis utilizing event-related potentials," *Electroencephalogr. Clin. Neurophysiol.*, vol. 70, pp. 510–523, 1988.