PARTITION-BASED ADAPTIVE ESTIMATION OF SINGLE RESPONSE EVOKED POTENTIALS

Ahmad M. Sarhan†, *Russell C. Hardie*†, *and Kenneth E. Barner*‡

Department of Electrical Engineering†
University of Dayton
Dayton, OH 45469-0226

Department of Electrical Engineering‡
University of Delaware
Newark, Delaware 19716

ABSTRACT

In this paper we have introduced and analyzed a new class of adaptive nonlinear filters referred to as partition-based linear (Pl) filters. The operation of those filters depends on partitioning the observation space in some fashion. Specifically, we have used here scaler quantization as an example to illustrate the concept of partitioning the observation space. Each partition is then assigned an output based on a linear combinations of observed samples in a moving window of finite length N. The filters are shown to exhibit appealing robustness. Simulations include a novel approach to estimating response-to-response variations in evoked potentials (EP), buried in the on-going electroencephalogram (EEG). Unlike the multi-channel filters currently used in EP estimation, the Pl filters do not require a separate electrode to provide a reference signal. In addition, no repetition of the stimulus is needed and the time of the stimulus need not be known.

1. INTRODUCTION

The EP is an electrical signal produced by the brain in response to some effective sensory stimulus. They are generally buried in the ongoing EEG noise at signal-to-noise ratios (SNR) of less than -6 dB. This low SNR makes it very difficult to extract the EP response from the observed signal. Ensemble averaging has been the most widely used technique in evoked potential extraction and monitoring. Ensemble averaging estimates the underlying response by averaging thousands of ensembles. However, the estimated signal will loose the information about individual response amplitude and latencies. Aunon *et al.* [1] have proposed latency-corrected methods based on averaging. These methods still suffer from the same setback, namely they require repeating the stimulus and taking the average.

An adaptive filter approach for processing EP signals uses multi-channel adaptive signal enhancers [3, 4]. A separate electrode is usually required to provide a reference EEG noise signal. These adaptive filters operate under the assumption that the noise in the reference channel and the signal are stationary and uncorrelated. Unfortunately, the EEG data tend to be highly correlated among the scalp electrodes.

Wiener filtering of evoked potentials using *a posteriori* time-varying filter was proposed by deWeerd *et al* [2]. Wiener filtering was not able to provide better improvements in the SNR than that obtained by the simple averaging technique. Westerkamp *et al* [3] have recognized that the EP is a non-stationary signal. Since Wiener filtering assumes that the process is stationary, it does not produce satisfactory improvements in the SNR. The approach used in [5] is based on a temporal partitioning of the observed signal. A different linear filter is applied to the signal at each sample time after the stimulus. This technique can be effective, but the exact time of the stimulus must be known *apriori*. Furthermore, it assumes that the data are cyclostationary.

In this paper, we introduce and analyze the Pl filters, a new class of adaptive filters whose operation is based on partitioning the R^N observation space defined by a size N moving observation window. Here, we focus on quantization based partitioning. Each partition is then assigned an index and a corresponding set of filter weights. Given that an observation vector lies in partition i, the filter uses the corresponding set of weights $\mathbf{w}_i$ and forms an estimate by taking a linear combinations of the samples in the observation vector. The filter weights are derived to minimize the overall MSE between the filter output and the desired response.

Estimating a single response EP buried in EEG noise is considered as an application for the proposed filters. We show in the simulations that the proposed filters give significantly lower mean-squared error than the simple LMS filter in estimating single response evoked potentials.

1.1. Filter Definition

The filtering problem consists of a given input sequence $\{x(n)\}$ comprised of a desired sequence $\{d(n)\}$ representing the EP response and an undesired component $\{n(n)\}$, representing the EEG noise. The task is to find a system that will suppress the undesired noise component while preserving the characteristics of $\{d(n)\}$. The difference between $\{d(n)\}$ and the filter estimate $\{\hat{d}(n)\}$ constitutes the error sequence $\{e(n)\}$, and is given by, $\{e(n)\} = \{\hat{d}(n)\} - \{d(n)\}$.

Without loss of generality, an estimate is formed at time n by sliding a window that spans N signal sample

0-7803-2431-5/95 $4.00 © 1995 IEEE

points. At each time n, the moving window produces the observation vector,

$$\begin{aligned}\mathbf{x}(n) &= (x(n), x(n-1), \ldots, x(n-N+1)) \\ &= (x_1(n), x_2(n), \ldots, x_N(n)).\end{aligned} \quad (1)$$

For notational simplicity, the time index n will be dropped.

Definition 1 *The output of the Pl filter is defined as,*

$$F_{Pl}(\mathbf{x}) = \mathbf{w}_i^T \mathbf{x} \ , \quad (2)$$

where $\mathbf{x}$ *lies in partition* i, *and* $\mathbf{w}_i$ *is the weight vector assigned to partition* i.

1.2. Partitioning Schemes

Depending on the application, the observation space can be partitioned using several methods. A desirable property of the partitioning scheme is to track signal nonstationarities. Examples of the many possible partitioning schemes include vector quantization (VQ), scaler quantization, and rank permutations. Partitioning the observation space using sample rank permutations has been shown in [6, 7, 8] to be effective in impulse rejection applications. Here the mapping of the time-ordered observation vector to the rank-ordered observation vector is shown to partition the observation space with a set of $N!$ distinct partitions. For example, the 6 possible permutations (partitions) of R^3 ($N = 3$) are $x_1 \le x_2 \le x_3$, $x_1 \le x_3 \le x_2$, $x_2 \le x_1 \le x_3$, $x_2 \le x_3 \le x_1$, $x_3 \le x_1 \le x_2$, and $x_3 \le x_2 \le x_1$. Clearly, any observation vector $\mathbf{x} \in R^3$ must fall in one of these partitions.

Perhaps one of the simplest methods to partition the observation space is based on scaler quantization of R^N. Unlike the permutation partitioning of R^N, the scaler-based quantization process takes into account both the amplitudes and the time correlations among the data. The scaler-based quantization process is illustrated as following. Let t_k, for $k = 0, 1, 2, \ldots, L$ be a set of increasing quantization levels. These values are obtained according to some partitioning scheme such as simple uniform quantization or using the LBG algorithm. The levels t_0 and t_L represent the minimum and maximum values of the input sequence $\{x\}$, respectively. Define a vector of quantization level indices for the observation samples as $\mathbf{q} = [q_1, q_2, \ldots, q_N]$. The scalar quantization operation is given by

$$Q(\mathbf{x}) = \mathbf{q}, \quad (3)$$

where $q_i = k : t_k < x_i < t_{k+1}$, for $i = 1, 2, \ldots, N$. Each unique vector $\mathbf{q}$ defines a distinct partition. In this way, the observation space is divided into $z = L^N$ disjoint partitions whose union is the whole observation space R^N. These partitions can be sequentially indexed from 1 to z.

Consider also the case where the sample mean from the observation vector is subtracted from each observation sample prior to quantization. In this case, the number of partitions in which an observation vector may lie in is $L^N - 1$. This mean subtracted quantization can track signal variation with fewer levels than the straight quantization method.

Once the partition index for an observation vector $\mathbf{x}$ is determined, the Pl filter forms an estimate according to (2).

2. OPTIMIZATION

To derive the optimal set of weights for each partition, the Pl filter output in (2) is rewritten as

$$F_{Pl}(\mathbf{x}) = \sum_{i=1}^{z} \mathbf{w}_i^T \mathbf{x} I(p(\mathbf{x}) = i), \quad (4)$$

where $I(\cdot)$ is the indicator function defined by $I(true) = 1$ and $I(false) = 0$, and $p(\mathbf{x})$ is the partition index in which $\mathbf{x}$ lies. Using this representation of the filter output, the estimate MSE can be written as

$$J = E\left(d - \sum_{i=1}^{z} \mathbf{w}_i^T \mathbf{x} I(p(\mathbf{x}) = i)\right)^2. \quad (5)$$

Conditioning the expectation on the partition, and summing over all partitions, reduces the MSE J to

$$\sigma_d^2 - 2\sum_{i=1}^{z} \mathbf{w}_i^T \mathbf{P}_i Pr(p(\mathbf{x}) = i) + \sum_{i=1}^{z} \mathbf{w}_i^T \mathbf{R}_i \mathbf{w}_i Pr(p(\mathbf{x}) = i), \quad (6)$$

where $\sigma_d^2 = E(d)^2$, $\mathbf{R}_i$ and $\mathbf{P}_i$ are the conditional correlation and cross correlation matrices,

$$\mathbf{R}_i = E(\mathbf{x}\mathbf{x}^T | p(\mathbf{x}) = i) \ \text{ and } \ \mathbf{P}_i = E(d\mathbf{x} | p(\mathbf{x}) = i). \quad (7)$$

This follows by assuming that the desired and observed signals are jointly stationary. Differentiating J with respect to each weight vector $\mathbf{w}_i$,

$$\nabla_{\mathbf{w}_i} J = 0 \ \Rightarrow \ \mathbf{R}_i \mathbf{w}_i = \mathbf{P}_i. \quad (8)$$

Assuming each of the conditional correlation matrices are invertible, the optimal weight vectors are given by $\mathbf{w}_i^* = \mathbf{R}_i^{-1} \mathbf{P}_i$, for $i = 1, 2, \ldots, z$. Thus, the overall optimum weights are given by the Wiener solution for each partition.

The Wiener solution, however, requires *apriori* knowledge of the underlying random process which is not available in many situations. Alternatively, an iterative solution which uses no *apriori* information is obtained using the LMS algorithm described in [9]. Thus, the weight update at the beginning of l'th iteration is defined by,

$$\mathbf{w}_i^{l+1} = \mathbf{w}_i^j + 2\mu e^l \mathbf{x}^l, \quad (9)$$

where j is the largest iteration number less than $(l+1)$ such that both x^l and x^j occupy the same partition, and μ is the step size that regulates the rate of convergence. In forming the estimate, a bias weight w_b has been used such that $\mathbf{x} = [1, x_1, \ldots, x_N]^T$, and $\mathbf{w} = [w_b, w_1, \ldots, w_N]^T$ (Fig. 1). Provided that μ is properly chosen, the filter weights converge to the optimum weights $\mathbf{w}_i^*$ and minimum J. Note that for each new

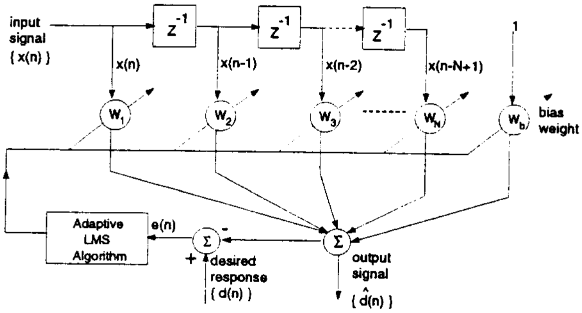

Figure 1: Schematic of the adaptive LMS filter. Filter comprises an N stage tapped delay line with N weights and a bias weight.

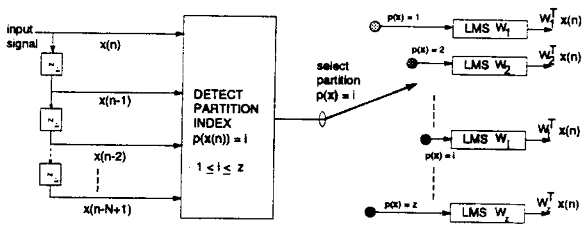

Figure 2: The Pl filtering algorithm. The system first determines the partition index i in which the observation vector is located and then uses the LMS algorithm to update the corresponding weight vector $\mathbf{w}_i$.

observation, only one set of filter weights are adapted (Fig. 2).

With partitions obtained using scaler quantization, the Pl filter can accurately track various signal nonstationarities. Consider as an example, the restoration of a chirp signal (Fig. 3) that linearly sweeps the frequency spectrum with time. The chirp signal, corrupted by zero-mean additive Gaussian white noise at a SNR of 0 dB, was used as the input to the Pl filter. After training the filter using the noisy and clean chirp signals as the input and desired response, respectively, the resulting frequency response of the filter at each time index n is a bandpass with a linearly increasing center frequency (Fig. 4). Clearly, the center frequency at time n corresponds to the local desired signal frequency. In fact, at each time n, $\mathbf{x}(n)$ is mapped to a different partition which shows that the scaler-based partitioning process preserves temporal correlations among the data. Furthermore, this example also illustrates that the Pl filter can be effectively trained to track a varying tone. The number of different tones that can be estimated using the Pl filter is, however, limited by $L^N - 1$, the number of distinct partitions.

3. SIMULATION RESULTS

The Pl filter was tested on simulated human EP data. The results have been compared with the standard LMS linear estimates. Mean square error results have been used as the criteria for comparison. The simulated EP

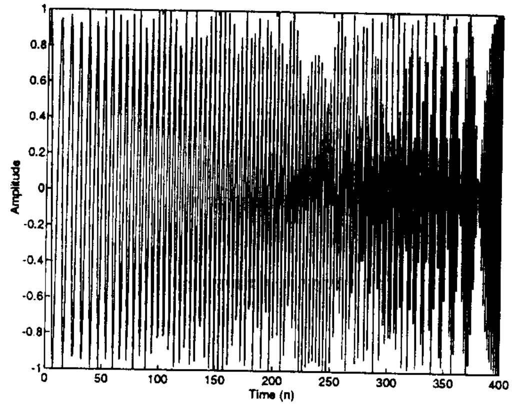

Figure 3: A chirp signal linearly sweeping the frequency spectrum with time.

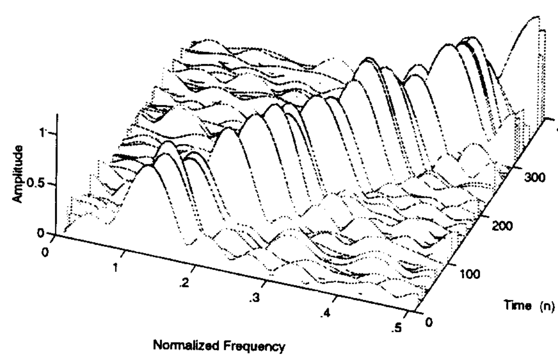

Figure 4: Frequency Response of the Pl filter at each time n, ($N = 17$, $t_0 = -\infty, t_1 = 0$, and $t_2 = \infty$).

data have been generated using raised cosine pulses to represent the individual peaks of the EP. Human EEG data have been added to the simulated scalp recorded responses. A total of 117 responses have been generated with 64 samples per response at a SNR of -6.12 dB. Fig. 5 shows three different simulated evoked responses along with the background EEG noise.

Figs. 6 and 7 show the standard LMS and Pl filter estimate, respectively, for one EP response that was not used in the training set. Both the LMS and the Pl filter have used a window size $N = 21$ to form the estimate. Two quantization levels ($L = 2$) have been used for the Pl filter after subtracting the local mean from the samples in each window. The quantization levels are given by $t_0 = -\infty$, $t_1 = 0$, and $t_2 = \infty$. With these parameters, the partition index will be invariant to scale and bias changes in $\mathbf{x}$. Note that the Pl filter preserves the peak and latency locations in the EP better than the straight LMS. Fig. 8 shows the MSE as a function of weights. Notice that the Pl filter produces a significantly lower mean-squared error than the standard linear filter.

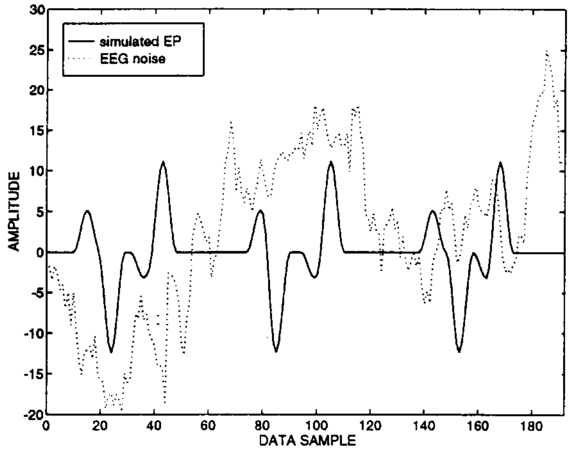

Figure 5: Examples of simulated EP showing three of the 117 evoked responses used in the simulations.

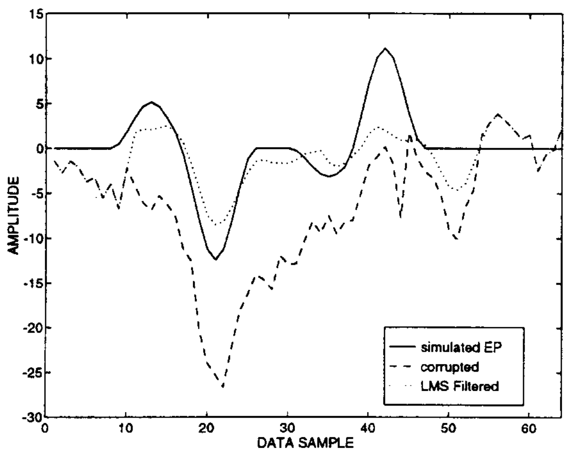

Figure 6: LMS filter estimate of a single EP response that was not used in the training data ($N = 21$).

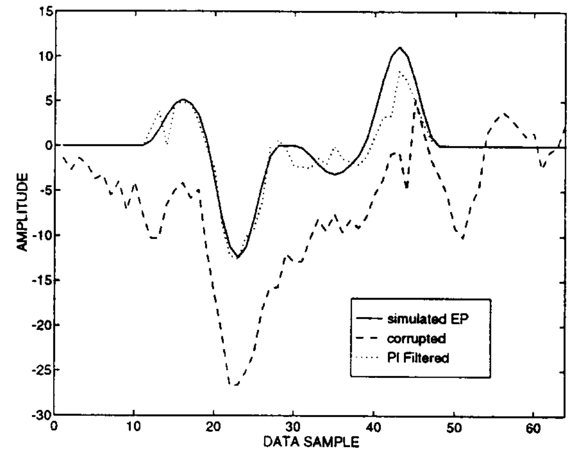

Figure 7: *Pl* filter estimate of an EP response that was not used in the training data, ($N = 21$, $t_0 = -\infty, t_1 = 0$, and $t_2 = \infty$).

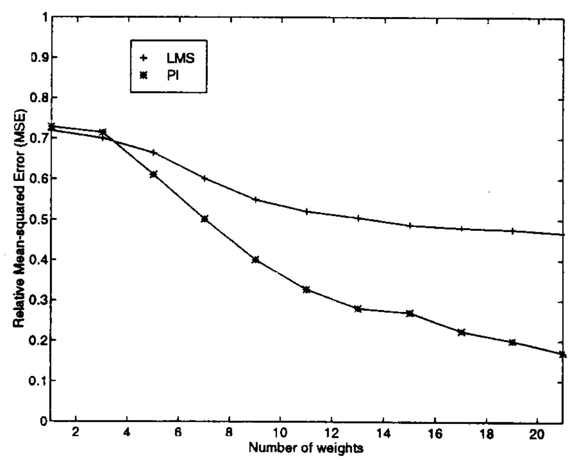

Figure 8: Relative MSE vs number of weights used.

4. REFERENCES

[1] J. I. Aunon, C. D. McGillem, and D. G. Childers, "Signal processing in evoked potential research: Averaging, principal components, modeling," in *Critical Reviews in Biomedical Engineering*. Cleveland, OH: CRC, 1981, pp.5:323-5:367.

[2] J. P. C. deWeerd and W. J. L. Martens, "Theorey and practice of a posteriory Wiener filtering of averaged evoked potentials," *Biol. Cybern.*, Vol. 30, p. 81, 1978.

[3] J. J. Westerkamp and J. I. Aunon, "Optimum multielectrode a posteriori estimates of a single-response evoked potentials," *IEEE Transactions on Biomedical Engineering* , 34(1), pp. 13-22, 1987.

[4] N. V. Thakor, "Adaptive filtering of evoked potentials," *IEEE Transactions on Biomedical Engineering,* 34(1), pp. 6-12, 1987.

[5] R. Williams, "Adaptive filtering of non-stationary signals using the modified P-vector algorithm," Dissertation, Univ. of Dayton, 1989.

[6] R. C. Hardie and K. E. Barner, "Rank conditioned rank selection filters for signal restoration," *IEEE Transactions on Image Processing,* 3(2), March, 1994.

[7] K. E. Barner and G. R. Arce, "Permutation fitlers: A class of non-linear filters based on set permutations," *IEEE Transactions on Signal Processing,* 42(4), April, 1994.

[8] K. E. Barner, R. C. Hardie, and G. R. Arce, "On the permutation and quantization partitioning of R^N and the filtering problem," *Proceedings of the 1994 CISS,* Princeton, New Jersey, March, 1994.

[9] B. Widrow and S. D. Stearns, *Adaptive Signal Processing.* Prentice-Hall, Englewood Cliffs, New Jersey, 1985.