

# Principal component analysis of galvanic skin responses

M. P. Tarvainen<sup>1</sup>, P. A. Karjalainen<sup>1</sup>, A.S. Koistinen<sup>2</sup> and M. Valkonen-Korhonen<sup>3</sup>

<sup>1</sup>University of Kuopio, Department of Applied Physics, Kuopio, Finland

<sup>2</sup>Finnish Institute of Occupational Health, Brain Work Laboratory, Helsinki, Finland

<sup>3</sup>Kuopio University Hospital, Department of Psychiatry, Kuopio, Finland

*Abstract*— The galvanic skin response (GSR) is a simple method of capturing the autonomic nerve response as a parameter of the sweat gland function. Any stimulus capable of an arousal effect can evoke the response and the amplitude of the response is more dependent on the surprise effect of the stimulus than on the physical stimulus strength. In this paper principal component analysis (PCA) is used for the analysis of the evoked GSRs. Basis functions are obtained from the eigendecomposition of the data correlation matrix. Because PCA is the best mean square fit of a set of orthogonal functions to the set of measurements, the solution will depend upon the nature of measurements. The dimensionality of measurements can be estimated by the number of basis functions needed to estimate measurements in a certain accuracy. Hence the eigenvalues, corresponding to used basis functions, are a measure of similarity. The method was tested using 20 healthy subjects and 13 psychotic patients. 11 surprising auditory stimuli were delivered at irregular intervals and evoked GSRs were recorded from the hand. Observed similarities between adjacent waveforms were more remarkable within healthy subjects.

*Keywords*— Galvanic skin response, principal component analysis.

## I. INTRODUCTION

The galvanic skin response (GSR) is a simple, useful, and reproducible electrophysiological technique to investigate sympathetic nervous system function [1]. In history GSR is also known as, or closely related to, the psychogalvanic reflex (PGR), electrodermal response (EDR), skin conductance response (SCR) and sympathetic skin response. Physically GSR is a change in the electrical properties of the skin in response to different kinds of stimuli. In measurements changes in the voltage measured from the surface of the skin are recorded.

A well known phenomenon of GSR is habituation, where the amplitudes of GSRs decrease during repeated stimulations [2], [3], [4]. In recent years there has been many studies concerning the normal values of GSR [1], [5], [6], [2], [7], [3], [8], [4], [9]. The typical response waveforms have also been studied. GSR response is a long lasting (several seconds) waveform of simple shape, normally biphasic or triphasic.

The aim of this paper is to study the capabilities of PCA to discriminate the GSR signals of healthy and psychotic subjects. We measured GSRs from 20 healthy subjects and 13 psychotic patients. Observed degree of similarity was clearly higher for healthy subjects. We will show how PCA can be used to evaluate waveshape similarity. It turns out that eigenvalues of data correlation matrix reveal much about the nature of measured GSRs. A significant clustering of healthy and psychotic subject groups with respect

to eigenvalues is shown.

## II. PRINCIPAL COMPONENT ANALYSIS

Principal component analysis (PCA) is a multivariate statistical procedure. The central idea in PCA is to reduce the dimensionality of the data set, while retaining as much as possible of the information in the original data. The measured data is presented as a weighted sum of orthogonal basis vectors. Usually PCA is performed on the covariance, cross-product, or correlation matrix of the original data. Here we obtain the basis vectors from eigendecomposition of data correlation matrix.

In GSR measurements we use an observation model

$$z_j = s_j + v_j, \quad (1)$$

where  $z_j$  is a column vector  $z_j = (z_j(1), \dots, z_j(T))^T$  of sampled measured potential after  $j$ th stimulus,  $s_j$  is the corresponding response signal and  $v_j$  is measurement noise. The measurement noise is assumed to be a stationary zero mean process. If we make  $N$  measurements, the response signals  $s_j$  will span a vector space  $\mathcal{S}$ , which will be at most of  $N$  dimensions. In case there is similarities in measured waveshapes, the dimension of the vector space  $\mathcal{S}$  will be some  $K \leq N$  and measurements can be approximated well with some low dimensional subspace of  $\mathcal{S}$ . We can thus express each measurement as linear combinations

$$z_j = H_S \theta_j + v_j. \quad (2)$$

In PCA the basis vectors are obtained from eigendecomposition of data correlation matrix [10]. Eigenvectors  $(\psi_1, \dots, \psi_K)$  corresponding to largest eigenvalues  $(\lambda_1, \dots, \lambda_K)$  are used as basis. Approximation of correlation matrix is

$$R_z \approx \frac{1}{N} \sum_{j=1}^N z_j z_j^T. \quad (3)$$

Quantitatively the first basis vector is the best mean square fit of a single waveform to the entire set of measurements. The second basis vector is the best mean square fit to the residual from the fit of the first factor, with a constraint that it is orthogonal to the first basis vector etc. Hence by using eigenvectors  $(\psi_1, \dots, \psi_K)$  corresponding to largest eigenvalues  $(\lambda_1, \dots, \lambda_K)$  as basis, the best  $K$  dimensional approximation of measurements in the least squares sense is obtained. Because principal component

solution is a best fit of a set of orthogonal functions to the set of signals, the solution will depend upon the nature of signal set.

In this work our main interests are the similarities of waveshapes. By normalizing the data PCA becomes sensitive to waveshape but not the signal amplitude. Information about the nature of measurements can be obtained from coefficients  $\theta_{ij}$  and from eigenvalues  $\lambda_i$ . Whenever the observation matrix  $H_S$  is orthonormal, the coefficient  $\theta_{ij}^2$  has the property of being the mean square contribution of  $i$ th basis vector to  $j$ th measurement  $z_j$  [11], [12]. When measurements are normalized  $\sum_{i=1}^N \theta_{ij}^2 = 1$  and  $\theta_{ij}^2 \times 100\%$  can be directly interpreted as the percentage contribution. The expected value of coefficients  $(\theta_{i1}^2, \dots, \theta_{iN}^2)$  associated with  $i$ th basis vector is [13]

$$E \{(\theta_{i1}^2, \dots, \theta_{iN}^2)\} = \lambda_i. \quad (4)$$

So each eigenvalue  $\lambda_i$  represents the mean square contribution of the corresponding basis function  $\psi_i$  to the measurements. If the first  $K$  eigenvectors are used in the observation model (2) the mean square reconstruction error, averaged over all waveforms, will be  $\sum_{i=K+1}^N \lambda_i$ . This is also the smallest conceivable mean square error. If measurements are normalized  $\sum_{i=1}^N \lambda_i = 1$ . An ideal example of similarity is if all the measured waveforms are identical. Then there would be only one nonzero eigenvalue and the eigenvector corresponding to it would have the same shape as the measurements. So it is obvious that the magnitudes of largest eigenvalues describe the amount of similarity in measured waveforms.

One possible visual way to estimate the similarity of waveshapes is to plot the cumulative sum of the eigenvalues [11]. The shape of such a curve describes the degree of coupling between various waveforms. A highly coupled signal set will have a sharply rising curve, rapidly approaching to its maximum  $\sum_{i=1}^N \lambda_i$ . If there is only few similarities between various signals the curve will approach the maximum very slowly.

#### A. Procedure

A procedure for analysing similarities in GSR measurements with PCA goes as follows. Measure a set of GSR responses  $z = (z_1, \dots, z_N)$ . If waveshapes are of interest, it is recommendable to normalize the measured data. This can be done by setting the norm of each measurement to unity. Because of habituation you should ignore responses of smallest amplitudes. In order to reconstruct measurements you need to form an observation matrix  $H_S$ . In PCA the basis vectors are obtained from eigendecomposition of data correlation matrix. Eigenvectors corresponding to largest eigenvalues are selected. In discrete case the correlation matrix can be approximated by (3).

Information about the nature of individual responses is obtained from the eigenvalues. If similarity between measurements is high, the first eigenvalue will be relatively big. When similarities exists, but there is variation in response latencies, the second eigenvalue, which describes the derivative of the mean, will also be significant. The rest of the eigenvalues should be insignificant in case of similarities.

Thus the dominance of the first and second eigenvalues and the insignificance of the rest of the eigenvalues is a measure of similarity. The nature of eigenvalues can be presented visually by plotting the cumulative sum of the eigenvalues.

### III. MATERIALS AND METHODS

The galvanic skin response was recorded from 20 healthy subjects and 13 psychotic patients. Responses were recorded with metal electrodes placed in the palm of the hand. The analysis time for each measurement started from the stimulus onset and lasted 8 seconds.

The experimental procedure for all subjects was as follows. Three kinds of stimuli, standard, target and novelty, were used in the stimulation. These auditory beeps of different frequency were delivered to both ears of the subject. The subject was advised not to pay attention to standard stimulus and to push a button when hearing a target stimulus. Subject was not informed about the novelty sounds. The novelty sounds differed from both the standard and target sounds and the GSRs were measured after these deviant stimuli. The experiment procedure included 11 novelty sounds and lasted about 10 minutes. The time between consecutive novelties were random, but at least 30 seconds. It is well known that any stimulus capable of an arousal effect can evoke a galvanic skin response and the amplitude of the response is more dependent on the surprise effect of the stimulus than on the physical stimulus strength. Because of this fact it is preferable to investigate the responses of the novelty stimuli instead of target stimuli.

### IV. RESULTS

Differences in waveshape similarity was remarkable between the two subject groups. Similarity in adjacent responses was higher for healthy subjects. Response waveforms were usually unaltered for healthy subjects, but there was a tendency of habituation. Observed decrease in amplitudes was 67–99% within healthy subjects. For psychotic patients waveshapes were more random and amplitudes usually smaller. At first we will perform PCA to a typical pair of subjects (healthy and psychotic). Later on in cluster analysis we shall present some similarity parameters in group wise manner. Fig. 1 shows the measured responses

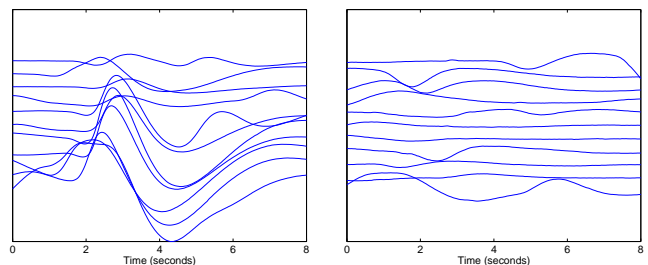


Fig. 1. Measured responses for healthy (left) and psychotic (right) subject as a waterfall plot.

for typical healthy and psychotic subject. Repeated measurements are plotted from bottom to top, with reversed vertical axis. Habituation phenomenon is seen in the responses of healthy subject.

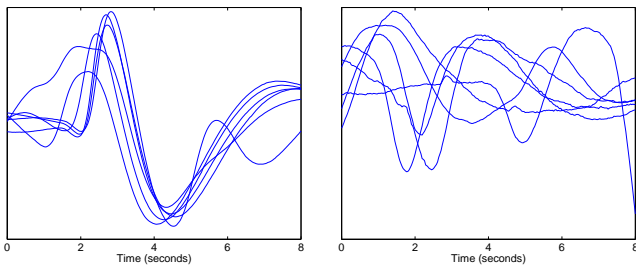


Fig. 2. Selected normalized responses for healthy (left) and psychotic (right) subject.

Our intention is to investigate these waveshape similarities with PCA and to obtain some criteria to separate the two subject groups from each others. Because of the habituation, we decided to analyse only six most substantial waveforms. The rest of the waveforms with small amplitudes are not significant when investigating waveshapes. The selected responses, normalized to unit norm, are presented in Fig. 2.

The mean values of the selected six waveforms are presented in Fig. 3. The mean value for healthy subject represents a very common response waveshape starting with a negative phase. For the psychotic subject the mean value is not very realistic and it seems to start immediately after the stimulus. Conclusions about the waveform coupling and normality of the responses could be drawn by simply looking at the mean of the measurements. Instead of stopping here we are aiming to a more sophisticated way to analyse GSR responses. We start by applying PCA into responses presented in Fig. 2.

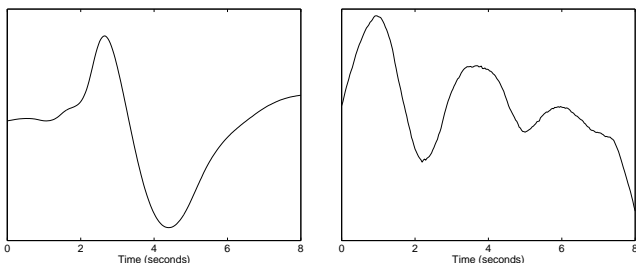


Fig. 3. The mean values of selected measurements for healthy (left) and psychotic (right) subject.

### A. Eigenvalues and eigenvectors

Three eigenvectors, obtained from eigendecomposition of data correlation matrix, corresponding to largest eigenvalues are presented in Fig. 4. For healthy subject the first eigenvector (upper left corner) is quite similar with the mean of the measurements in Fig. 3. The mean square contribution of the first eigenvector to the measurements is approximately 80% and the contribution of the first two eigenvectors is over 92%. This shows how effective the two most dominant eigenvectors can be when similarities exist. For the psychotic subject the total contribution of the first and second eigenvectors is only about 68%. So the first two eigenvalues seems to be a good measure of waveshape similarity.

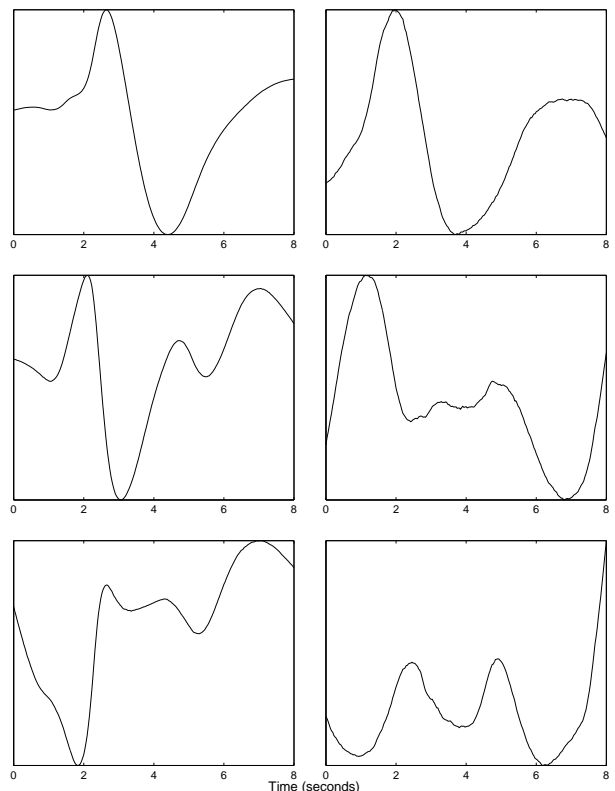


Fig. 4. Eigenvectors corresponding to the three largest eigenvalues of the correlation matrix of the measured GSRs for healthy (left) and psychotic (right) subject. The corresponding eigenvalues are  $\lambda_1 = 0.8090$ ,  $\lambda_2 = 0.1158$  and  $\lambda_3 = 0.0443$  for healthy subject and  $\lambda_1 = 0.4261$ ,  $\lambda_2 = 0.2495$  and  $\lambda_3 = 0.1366$  for psychotic subject.

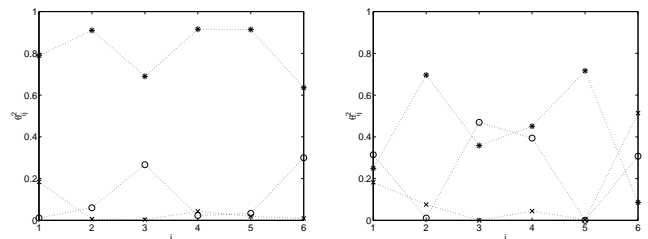


Fig. 5. Coefficients  $\theta_{1j}^2$  (\*),  $\theta_{2j}^2$  (o) and  $\theta_{3j}^2$  (x) for selected measurements ( $j = 1, \dots, 6$ ) for healthy (left) and psychotic (right) subject.

The mean square contributions of the three largest eigenvectors  $\theta_{1j}^2$ ,  $\theta_{2j}^2$  and  $\theta_{3j}^2$  to selected six measurements ( $j = 1, \dots, 6$ ) are presented in Fig. 5. For healthy subject the contribution of the first eigenvector is clearly the most substantial to all waveforms and the contribution of the third factor is effective only to the first measurement. For psychotic subject the difference between contributions of the first and second factor is not so evident and also the third factor is effective in few cases.

Next we will illustrate the similarity differences with the plot of the cumulative sum of the eigenvalues. The cumulative sum of six largest eigenvalues is presented in Fig. 6. The curve for healthy subject is more rapidly rising and achieves the maximum earlier. This is exactly what should happen when waveshape similarity is higher.

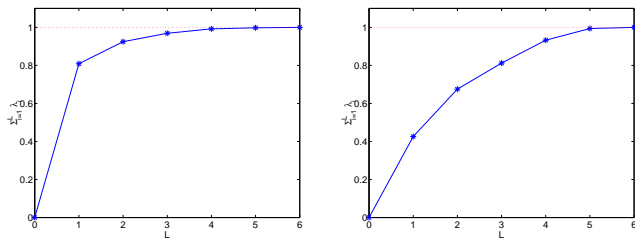


Fig. 6. The cumulative sum of largest eigenvalues for healthy (left) and psychotic (right) subject as a function of the number of eigenvalues summed (\*). The sum of all eigenvalues is 1 (- -).

### B. Response amplitudes

The most important clinical parameter of GSR is the amplitude. Amplitude was measured from peak to peak. For most subjects there was a habituation in amplitudes, even though we used novelties for recordings. We also observed a comprehensive difference in amplitudes between healthy and psychotic subject. The maximum amplitude was 7.75 mV for healthy subject and 2.23 mV for psychotic subject. We also calculated the mean value and standard deviation of five largest amplitudes. The results were  $7.00 \pm 0.90$  mV for healthy subject and  $1.59 \pm 0.50$  mV for the psychotic subject. In this case the difference in amplitudes is large, but the variation in amplitudes between healthy subjects is also rather substantial.

### C. Cluster analysis

In this section we will compare the parameters presented earlier between the two subject groups. Our main interests are the eigenvalues. The mean values and standard deviations of some calculated parameters for healthy and psychotic subject groups are presented in Table I. Usually

TABLE I

MEAN VALUES AND STANDARD DEVIATIONS (STD) OF SOME SPECIFIC PARAMETERS FOR HEALTHY AND PSYCHOTIC SUBJECTS.  $\lambda_i$  ( $i = 1, \dots, 6$ ) ARE THE SIX LARGEST EIGENVALUES.  $V_{\max}$  IS THE MAXIMUM AMPLITUDE OF MEASURED RESPONSES AND  $V_{\max}^{(5)}$  IS THE MEAN VALUE OF FIVE LARGEST AMPLITUDES.

	healthy subjects	psychotic subjects
$\lambda_1 \pm \text{std}$	$0.6673 \pm 0.1546$	$0.5133 \pm 0.1053$
$\lambda_2 \pm \text{std}$	$0.2076 \pm 0.0857$	$0.2154 \pm 0.0294$
$\lambda_3 \pm \text{std}$	$0.0742 \pm 0.0456$	$0.1428 \pm 0.0405$
$\lambda_4 \pm \text{std}$	$0.0327 \pm 0.0305$	$0.0816 \pm 0.0274$
$\lambda_5 \pm \text{std}$	$0.0138 \pm 0.0154$	$0.0352 \pm 0.0258$
$\lambda_6 \pm \text{std}$	$0.0044 \pm 0.0060$	$0.0116 \pm 0.0143$
$(\lambda_1 + \lambda_2) \pm \text{std}$	$0.8749 \pm 0.0902$	$0.7288 \pm 0.0878$
$V_{\max} \pm \text{std}$	$4.6205 \pm 3.5565$	$3.1794 \pm 2.9999$
$V_{\max}^{(5)} \pm \text{std}$	$3.1711 \pm 2.7830$	$1.6365 \pm 1.3157$

in PCA related cluster analysis a classification of an observation is made by plotting some principal component (coefficient  $\theta_{ij}$ ) as a function of another principal component. In order to visualize the clustering of healthy and psychotic subjects we will present plots of different eigenvalues (mean values presented in Table I). Fig. 7 shows a plot of the sum of two largest eigenvalues  $\lambda_1 + \lambda_2$  with

respect to sum  $\sum_{i=4}^N \lambda_i$  (reconstruction error when measurements are estimated with three most dominant eigenvectors). The clustering of groups is clearly seen, but is not complete. Most of the healthy subjects (14/20) are clearly clustered in the bottom right corner. But there is six healthy subjects which lie more or less inside psychotic group.

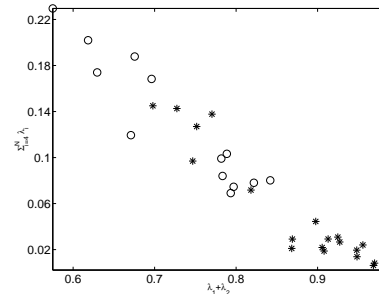


Fig. 7. Individual GSR measurements: plot of healthy (\*) and psychotic (o) subjects with respect to the sum of largest eigenvalues  $\lambda_1 + \lambda_2$  and sum  $\sum_{i=4}^N \lambda_i$ , where  $N = 6$  is the number of waveforms analysed.

The same kind of clustering is seen in a plot of third eigenvalue  $\lambda_3$  with respect to fourth eigenvalue  $\lambda_4$  in Fig. 8. Both  $\lambda_3$  and  $\lambda_4$  are small for a majority of healthy subjects. Again six healthy subjects are among psychotic subjects. These six are in fact the same ones which deviate from other healthy subjects in Fig. 7.

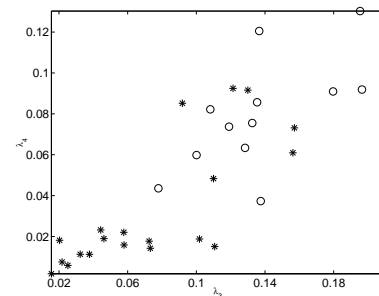


Fig. 8. Individual GSR measurements: plot of healthy (\*) and psychotic (o) subjects with respect to their third and fourth largest eigenvalues  $\lambda_3$  and  $\lambda_4$ .

## V. DISCUSSION

We have presented a principal component based analysis of GSR measurements. The nature of measurement set is revealed from eigenvalues of correlation matrix. If the degree of similarity is high, the first few eigenvalues are clearly dominative and rest of the eigenvalues are insignificantly small. By using eigenvalues as a measure of similarity a significant clustering of healthy and psychotic subject groups is obtained (Figs. 7 and 8). 70% of healthy subjects are clearly separated from all psychotic patients based on eigenvalues.

## REFERENCES

- [1] B. Shahani, J. Halperin, P. Boulu, and J. Cohen, "Sympathetic skin response—a method of assessing unmyelinated axon dysfunction in peripheral neuropathies," *J Neurol Neurosurg Psychiatr*, vol. 47, pp. 536–542, 1984.

- [2] B. Elie and P. Guiheneuc, "Sympathetic skin response: normal results in different experimental conditions," *Electroencephalogr Clin Neurophysiol*, vol. 76, pp. 258–267, 1990.
- [3] S. Aramaki, Y. Kira, and Y. Hirasawa, "A study of the normal values and habituation phenomenon of sympathetic skin response," *Am J Phys Med Reh*, vol. 76, pp. 2–7, January 1997.
- [4] M. Toyokura, "Waveform and habituation of sympathetic skin response," *Electroencephalogr Clin Neurophysiol*, vol. 109, pp. 178–183, 1998.
- [5] M. Baba, Y. Watahiki, M. Matsunaga, and K. Takebe, "Sympathetic skin response in healthy man," *Electromyogr Clin Neurophysiol*, vol. 28, pp. 277–283, 1988.
- [6] A. Uncini, S. Pullman, R. Lovelace, and D. Gambi, "The sympathetic skin response: normal values, elucidation of afferent components and application limits," *J Neur Sci*, vol. 87, pp. 299–306, 1988.
- [7] V. Drory and A. Korczyn, "Sympathetic skin response: age effect," *Neurology*, vol. 43, pp. 1818–1820, September 1993.
- [8] N. Kanzato, Y. Komine, and K. Fukiyama, "Sympathetic skin responses of the hand in normal subjects: shorter latency at distal phalanx," *Electroencephalogr Clin Neurophysiol*, vol. 105, pp. 165–170, 1997.
- [9] T. Deltombe, P. Hanson, J. Jamart, and M. Cl  rin, "The influence of skin temperature on latency and amplitude of the sympathetic skin response in normal subjects," *Muscle & Nerve*, vol. 21, pp. 34–39, January 1998.
- [10] I. Jolliffe, *Principal Component Analysis*. Springer-Verlag, 1986.
- [11] E. John, D. Ruchkin, and J. Villegas, "Experimental background: signal analysis and behavioral correlates of evoked potential configurations in cats," *Ann N Y Acad Sci*, vol. 112, pp. 362–420, 1964.
- [12] D. Ruchkin, J. Villegas, and E. John, "An analysis of average evoked potentials making use of least mean square techniques," *Ann N Y Acad Sci*, vol. 115, pp. 799–821, 1964.
- [13] E. Oja, *Subspace Methods in Pattern Recognition*. Wiley, 1983.