Analysis of galvanic skin responses with principal components and clustering techniques

Mika P. Tarvainen, Anu S. Koistinen, Minna Valkonen-Korhonen, Juhani Partanen, Pasi A. Karjalainen

Abstract— **An advanced method for analyzing the patterning of successive galvanic skin responses (GSR) is presented. The proposed method is based on principal component analysis (PCA) in which the vector containing the measured signal is presented as a weighted sum of orthogonal basis vectors. The method is tested using measurements from 20 healthy controls and 13 psychotic patients. For each subject 11 surprising auditory stimuli were delivered to right ear at irregular intervals and evoked GSRs were recorded from the hand. For most of the healthy controls there was a clear pattern in successive GSRs, whereas within psychotic patients the lack of time-locking of GSRs seemed to be characteristical. These between group differences can be revealed by the proposed method. With application to clustering a significant discrimination, with overall correct ratings of 82%, of healthy controls and psychotic patients is achieved. A significant fact is that all patients were ranked correctly giving the proposed method a sensitivity of 100%.**

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

I. Introduction

The galvanic skin response (GSR) is a simple, useful and reproducible method of capturing the autonomic nerve response as a parameter of the sweat gland function [1]. Physically GSR is a change in the electrical properties of the skin in response to different kinds of stimuli. 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 measurements changes in the voltage measured from the surface of the skin are recorded.

In history GSR is also known as, or closely related to, the sympathetic skin response (SSR) and skin conductance response (SCR). Most of the GSR studies in last decades are concerned with the normal values of response amplitude and latency [1], [2], [3], [4], [5], [6], [7]. Also the habituation of response amplitudes during repeated stimulations has been studied [4], [5], [6]. Response amplitudes vary substantially, depending on the experimental conditions. In [4] an auditory stimulus was delivered to both ears and a mean amplitude of (2.8 ± 1.2) mV measured from the palm was observed. Observed latency measured from the palm to auditory stimulus was (1.50 ± 0.09) seconds in [4] and (1.49 ± 0.17) seconds in [3].

Typical GSR waveshapes have been studied e.g. in [5], [6]. The waveshape is usually biphasic or triphasic and lasts several seconds. Because GSRs are such long lasting waveforms interstimulus intervals (ISIs) should be long enough. When using short ISIs response overlapping should be considered by decomposing the overlapped responses. Such a decomposition for SCRs is presented in [8].

Normally reproduced within-subject GSRs have common features in waveshapes. Amplitudes tend to habituate, latencies might increase slightly and waveshapes remain fairly unaltered in repetitions. The decrease in amplitudes and increase in latencies is affected by the weakening of the surprise effect of stimulation and by the weakening of alertness of the subject during the experiment. The influence of alertness and emotional state of subject to evoked responses is discussed in [9], [6].

The aim of this paper is to present a new systematic method for analyzing the patterning of successive GSRs. The proposed method is based on principal component analysis (PCA) and on regression methods. As a specific application the method is applied to GSRs measured from 20 healthy controls and 13 psychotic patients. Observed degree of similarity in successive GSRs was clearly higher for healthy subjects. For psychotic patients no clear time-locking or pattern was observed in the measured responses. In this paper we will show how PCA can be used to evaluate the nature of measurements and thereby discriminate GSR signals of healthy controls and psychotic patients. With application to clustering a significant discrimination of the two subject groups is obtained.

II. Principal components and clustering

Principal component analysis (PCA) is a multivariate statistical procedure, where the vector containing the measured signal is presented as a weighted sum of orthogonal basis vectors. 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 dimensionality of measurements can be estimated by the number of basis vectors needed to estimate measurements in a certain accuracy. 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.

Mika P. Tarvainen and Pasi A. Karjalainen are with University of Kuopio, Department of Applied Physics, P.O.Box 1627, FIN-70211 Kuopio, Finland

Anu S. Koistinen is with Finnish Institute of Occupational Health, Brain@Work Laboratory, Helsinki, Finland

Minna Valkonen-Korhonen is with Kuopio University Hospital, Department of Psychiatry, Kuopio, Finland

Juhani Partanen is with Kuopio University Hospital, Department of Clinical Neurophysiology, Kuopio, Finland

A. Formulation of PCA

We use here the following notation for GSR measurements. The sampled potential after j'th stimulus is denoted with a length T column vector

$$z_j = \begin{pmatrix} z_j(1) \\ \vdots \\ z_j(T) \end{pmatrix} \tag{1}$$

As an observation model for the measurements we use the so-called additive noise model

$$z_j = s_j + v_j \tag{2}$$

where s_j is the response signal corresponding to j'th stimulus 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 $\min\{N, T\}$ dimensions. In the case that there is similarities in measured waveshapes, the dimension of the vector space $\mathcal{S}$ will be $K \leq N$ and measurements can be well approximated with some lower dimensional subspace of $\mathcal{S}$. We can thus express each measurement as linear combination

$$z_j = H_{\mathcal{S}}\theta_j + v_j \tag{3}$$

where $H_{\mathcal{S}} = (\psi_1, \ldots, \psi_K)$ is a $T \times K$ matrix of basis vectors which span the K dimensional subspace of $\mathcal{S}$ and $\theta_j \in \mathbb{R}^K$ is a column vector of weights to j'th measurement. By defining a measurement matrix $z = (z_1, \ldots, z_N)$ the observation model (3) can be written in the form

$$z = H_{\mathcal{S}}\theta + v \tag{4}$$

where $\theta = (\theta_1, \ldots, \theta_N)$ and $v = (v_1, \ldots, v_N)$.

The critical point in the use of model (4) is the selection of basis vectors ψ_i. A variety of ways to select these basis vectors exist. For example a selection of trigonometric basis is the Fourier series approach. Here we will concentrate on a special case when the basis vectors are orthonormal to each other. This means that

$$H_{\mathcal{S}}^T H_{\mathcal{S}} = I \tag{5}$$

and hence the ordinary least squares solution becomes

$$\hat{\theta}_{\mathrm{LS}} = (H_{\mathcal{S}}^T H_{\mathcal{S}})^{-1} H_{\mathcal{S}}^T z = H_{\mathcal{S}}^T z \tag{6}$$

and the measurements z can be reconstructed by

$$\hat{z}_{\mathrm{LS}} = H_{\mathcal{S}}\hat{\theta}_{\mathrm{LS}} = H_{\mathcal{S}} H_{\mathcal{S}}^T z = \sum_{i=1}^{K} \psi_i c_i \tag{7}$$

where $c_i = \psi_i^T z$. If the measurements z are random, the coefficients c_i are also random parameters. Hence we will require the coefficients c_i to be uncorrelated and we can write

$$\begin{aligned} E\left\{cc^T\right\} &= E\left\{H_{\mathcal{S}}^T z z^T H_{\mathcal{S}}\right\} \\ &= H_{\mathcal{S}}^T R_z H_{\mathcal{S}} = \mathrm{diag}\left(\lambda_1, \ldots, \lambda_K\right) \end{aligned} \tag{8}$$

This is an eigenproblem and the basis vectors are obtained as eigenvectors of data correlation matrix R_z. The sum (7) is called the discrete Karhunen-Loeve or principal component transform.

B. Characteristics of PCA

The fact that the basis vectors are orthogonal has lead to interpretations that the basis vectors could represent some independent physiological generators. This interpretation is however questionable because the fact that basis vectors are orthogonal is because they are eigenvectors of a symmetric matrix. Signals from independent physiological generators however do not necessarily have this property as shown in [10]. Even though the basis vectors have no clear meaning it is reasonable to argue, based on [11], that variations in GSR latencies is the main reason for the existence of additional components in PCA.

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 $(\psi_1, \ldots, \psi_K)$ corresponding to largest eigenvalues $(\lambda_1, \ldots, \lambda_K)$ 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 we are interested on the morphology of successive GSR responses. By normalizing the measured data PCA becomes sensitive to waveshape but not the signal amplitude. Information about the nature of measurements can be obtained from the principal components $\theta_j(i)$ and from eigenvalues λ_i. Whenever the observation matrix $H_{\mathcal{S}}$ is orthonormal, the coefficient $\theta_j^2(i)$ has the property of being the mean square contribution of i'th basis vector to j'th measurement [12], [13]. When measurements are normalized $\sum_{i=1}^{N} \theta_j^2(i) = 1$ and $\theta_j^2(i) \times 100\%$ can be directly interpreted as the percentage contribution. The expected value of coefficients $(\theta_1^2(i), \ldots, \theta_N^2(i))$ associated with i'th basis vector is [14], [15]

$$\begin{aligned} E\left\{\left(\theta_1^2(i), \ldots, \theta_N^2(i)\right)\right\} &= E\left\{(\psi_i^T z)^2\right\} \\ &= E\left\{\psi_i^T z z^T \psi_i\right\} \\ &= \psi_i^T E\left\{z z^T\right\} \psi_i \\ &= \psi_i^T R_z \psi_i = \lambda_i \end{aligned} \tag{9}$$

So that each eigenvalue λ_i represents the mean square contribution of the corresponding basis vector ψ_i to the measurements. If the first K eigenvectors are used in the observation model (4) 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 the case that 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 [12]. The shape of such a curve describes the degree of similarity 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 common features between various signals the curve will approach the maximum very slowly.

C. Graphical representation of data

The main idea of PCA is to reduce the dimensionality of the data set. This is useful in obtaining a straightforward graphical representation of multi-dimensional data in some lower dimensional subspace. If two or three first principal components account for most of the total variation in data, the data can be represented comprehensively with respect to these first two or three PCs. In this study we want to represent a set of measurements instead of a single measurement in e.g. two dimensional coordinates. Hence we need to use the eigenvalues of correlation matrix of data set instead of principal components of a single measurement in the graphical representation. A typical choice would be to represent data sets with respect to first few eigenvalues, but also some sensible combinations of the eigenvalues can be used. Several techniques, which can be used in conjunction with PCA, for representing high dimensional data are discussed in [15].

D. Clustering methods

In cluster analysis the objective is to divide the given observations into groups or clusters in some sensible way. Little or nothing about the groups is known *a priori*. Instead of clustering the original observations it could be more appropriate to perform PCA and clusterize the obtained PCs $\theta_j(i)$ or eigenvalues λ_i. Using this kind of approach the obtained clusters can be visually presented in two or three dimensions. This kind of approach has been used e.g. in [16].

In this study we use a simple hierarchical agglomerate clustering method. The initial partition is singleton partition, where all clusters contain only one component. Then two closest clusters are merged into one cluster repeatedly, until the desired number of clusters is achieved or some defined criterion is satisfied. The crucial point of the method is the definition of the distance $D(X, Y)$ between clusters X and Y. Different types of distance definitions are presented in [17]. The most natural cluster-wise distance function is the Euclidean distance of cluster centers $\overline{X}$ and $\overline{Y}$.

$$D(X, Y) = \left\| \overline{X} - \overline{Y} \right\|_2^2 \tag{10}$$

A Clustering method using this as a distance function is called the *centroid method*.

III. A systematic method for analyzing GSRs

In this section we describe a systematic method for analyzing the nature of GSR measurements. Procedure for discriminating response sets of different degrees of waveshape similarity is presented. The proposed method is described stepwise as follows.

1. Measure a set of GSR responses $z = (z_1, \ldots, z_N)$. Because GSR is a long lasting waveform, the interstimulus interval should be long enough to avoid response overlapping. Novelty stimuli are preferable in order to obtain decent responses.

2. If habituation of amplitudes is strong it is reasonable to reject the weakest GSRs from the analysis.

3. When waveshapes are of interest the measured data should be normalized in some manner. One way to do this is to set the norm

$$\|z_j\| = 1, \quad \forall j = 1, \ldots, N \tag{11}$$

This normalization makes PCA sensitive to waveshapes, but not to signal amplitudes.

4. Calculate the data correlation matrix R_z. For discrete case the correlation matrix can be approximated by

$$R_z \approx \frac{1}{N} \sum_{j=1}^{N} z_j z_j^T \tag{12}$$

Note that when correlation matrix is used here, an approximation for the mean of the GSRs is modeled automatically as the first eigenvector of the matrix. If covariance matrix is used instead, the mean has to be included in the equations explicitly.

5. Solve the ordinary eigendecomposition $R_z \Psi = \lambda \Psi$ of the correlation matrix. Form observation matrix $H_{\mathcal{S}} = (\psi_1, \ldots, \psi_K)$, where ψ_i is the i'th eigenvector of R_z corresponding to i'th largest eigenvalue λ_i. Principal components θ can be solved from (6) and estimates $\hat{z}_{\mathrm{LS}}$ for GSRs from (7).

6. Information about the nature of a set of GSRs is obtained from the eigenvalues. If similarity between consecutive measurements is high, the first eigenvalue will be relatively large. 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. Corresponding information about single responses is obtained from coefficients $\theta_j^2(i)$. If waveforms are not normalized these coefficients depend on amplitudes and can be used to describe habituation. We define a parameter C_h describing the relative dominance of the first three waveforms as

$$C_h = \frac{\Sigma_{j=1}^{3} |\theta_j(1)| + |\theta_j(2)|}{\Sigma_{j=1}^{N} |\theta_j(1)| + |\theta_j(2)|} \tag{13}$$

IV. Materials and experimental procedure

We examined 13 first-episode patients with acute psychosis (PANSS scores 99–120) admitted for hospital eval-

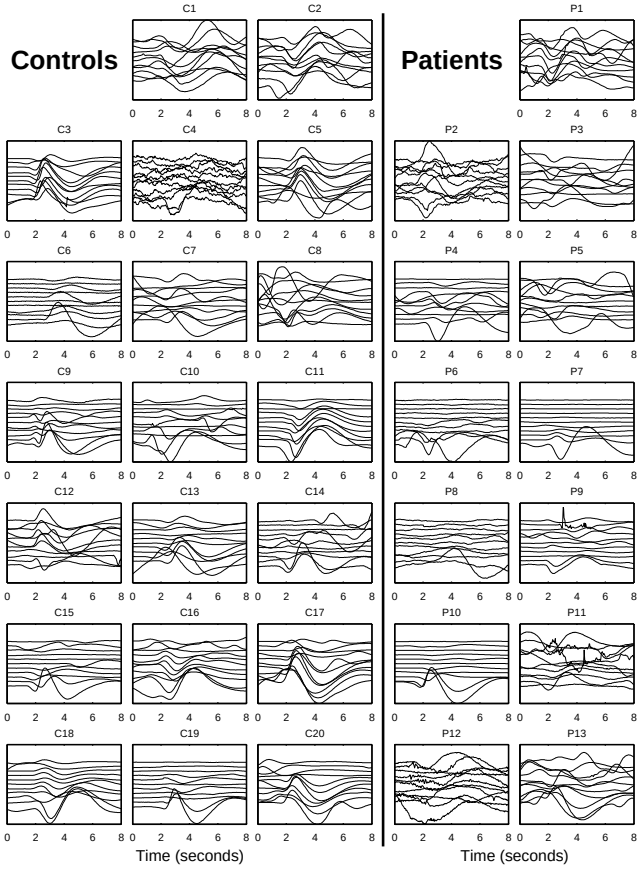

Fig. 1. GSR measurements for healthy controls C1, C2, . . . , C20 and for psychotic patients P1, P2, . . . , P13 as waterfall plots. The amplitude scale is different for each subject.

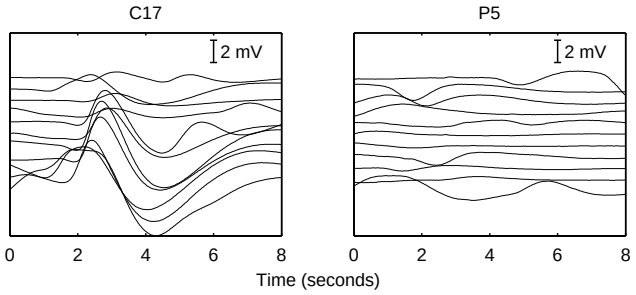

Fig. 2. Measured responses for healthy control (C17) and psychotic patient (P5) as a waterfall plot.

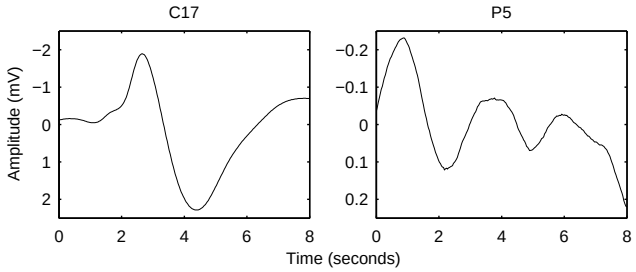

Fig. 3. The mean values of measurements for healthy control (C17) and psychotic patient (P5). Note that the amplitude scale for psychotic patient is 10 times smaller than for healthy control.

uation compared to 20 healthy controls. None of the patients had never received medication for psychiatric problems. Although main inclusion criterion was disturbance of reality-testing, the diagnostic confirmation was verified by SCID interview, carried out by a trained interviewer. All subjects provided written informed consent and the study was approved by the local ethical committee. The galvanic skin responses were recorded using Ag-AgCl electrodes affixed to the palm and dorsal side of the non-dominant hand. As a recording device a NeuroScan^{TM} system by NeuroSoft Inc. was used. The sampling frequency of the device is 500 Hz, but GSR signals were decimated to 31 Hz before the analysis.

The experimental procedure for all subjects was as follows. Three kinds of auditory stimuli, standard, target and novelty, were used in the stimulation according to classical oddball paradigm (85% standard stimuli of 800 Hz and 15% pitch deviants of 560 Hz with interstimulus intervals of 1 second). The 11 unique novelty sounds were human sounds randomly presented among standard and target stimuli. Tone intensity was individually adjusted 55 dB above the hearing level. Subjects were advised not to pay attention to standard stimulus and to push a button when hearing a target stimulus. They were not informed about the novelty sounds. GSR was recorded continuously during the session and 8 second epoch after each novelty stimulus was extracted for analysis. Time between consecutive novelties were random but at least 30 seconds. Because of the long interstimulus intervals of novelty stimuli no response overlapping appeared in the measurements.

V. Results

The proposed method is applied to GSRs measured from 20 healthy controls and 13 psychotic patients. Measurement sets for all subjects are presented in Fig. 1. At first we introduce the method by applying it to GSR sets of typical control and psychotic subject. Afterwards we show the discrimination of the two subject groups with respect to different parameters obtained from the method.

A. Measurements

Typical sets of measurements for control subject (C17) and psychotic patient (P5) are presented in Fig. 2. Repeated measurements are plotted from bottom to top, with reversed vertical axis. Amplitude scale is the same in both figures. Measurement time lasts 8 seconds after stimulus onset.

The degree of similarity between these 11 successive responses is clearly higher for the control subject, latency variation is small and the response waveshape does not change prominently between repetitions. The approximated onset latency is about 1.8 seconds. For psychotic patient the lack of time-locking in GSRs seems to be characteristic and there is not any observable pattern in the measurements. Mean values of the 11 responses are presented in Fig. 3. The mean value for healthy subject rep-

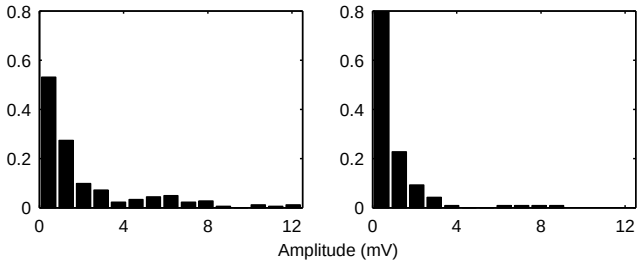

Fig. 4. Peak to peak amplitudes for healthy controls (left) and psychotic patients (right).

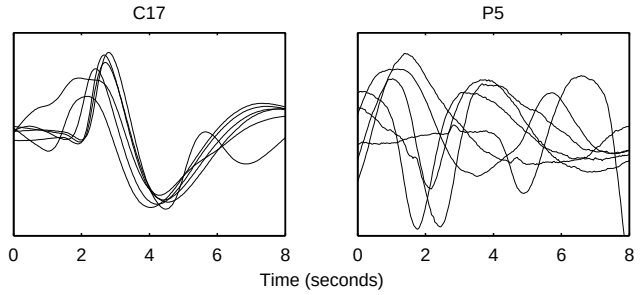

Fig. 5. Selected normalized responses for healthy control (C17) and psychotic patient (P5).

resents 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 arise 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 analyze GSR responses.

B. Amplitudes and habituation

GSR amplitudes were defined from peak to peak at time window from 1 to 6 seconds after stimulus onset. This is the time frame in which GSR peaks normally appear. Response amplitudes in Fig. 2 seem to be much larger for control subject. There is in fact a slight difference in amplitudes between the two subject groups as seen from Fig. 4, which shows the histograms of peak to peak amplitudes of all responses of all subjects in the two groups. However the amplitude variation within control subjects is rather substantial and therefore response amplitudes can not be used to discriminate the subject groups.

For most subjects a significant habituation of amplitudes was observed. Habituation is seen from responses of the control subject in Fig. 2. The largest response amplitude is 7.75 mV and the smallest only 0.85 mV. So the decrease in amplitudes is about 89%. Within all control subjects the habituation varies between 67–99%, even though we used novelties in stimulation.

Observed strong habituation must be taken into account in the analysis. It is possible that the weakest measurements in a set do not feature responses to specific stimuli at all. Instead a weak measurement can include a response

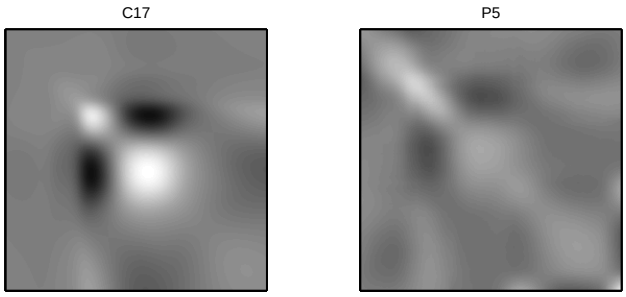

Fig. 6. Correlation matrix for healthy control (C17) and psychotic patient (P5) displayed as an gray-scale image. The values of positive correlation are printed in white and the values of negative correlation in black.

to some preceding target stimulus and the contribution of measurement noise is also notable. Based on these observations it is reasonable to reject the weakest responses from the analysis and focus on the similarities between the most substantial waveforms.

C. PCA analysis

Because of the strong habituation, we decided to analyze only 6 most substantial waveforms. The rejected 5 measurements of smallest norms are usually responses to last stimuli and do not necessarily feature responses to specific stimuli. Because waveshapes are of interest each selected response is normalized to unit norm. This normalization makes PCA sensitive to waveshapes only. The normalized selected responses are presented in Fig. 5.

First step in PCA is to calculate the correlation (or covariance) matrix of data. Correlation matrices of selected responses are presented in Fig. 6. For control subject one can distinguish regions of strong positive and negative correlations. The two regions of positive correlation (printed in white) in the diagonal are due to the time-locking of the two peaks in GSR signals. For the regions of negative correlation (printed in black) the peaks in signals are in opposite phase. For psychotic patient there is not as significant strongly correlated regions as for control subject.

Three eigenvectors, obtained from eigendecomposition of data correlation matrix, corresponding to largest eigenvalues are presented in Fig. 7. For the control subject the first eigenvector in the 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 81% 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 patient the total contribution of the first two eigenvectors is only about 67%. These results indicate how much the two largest eigenvalues can vary when the similarities of response sets are clearly different.

The nature of the eigenvalues can be presented visually by plotting the cumulative sum of the eigenvalues. The cumulative sum of six largest eigenvalues is presented in Fig. 8. The curve for control subject is more rapidly rising

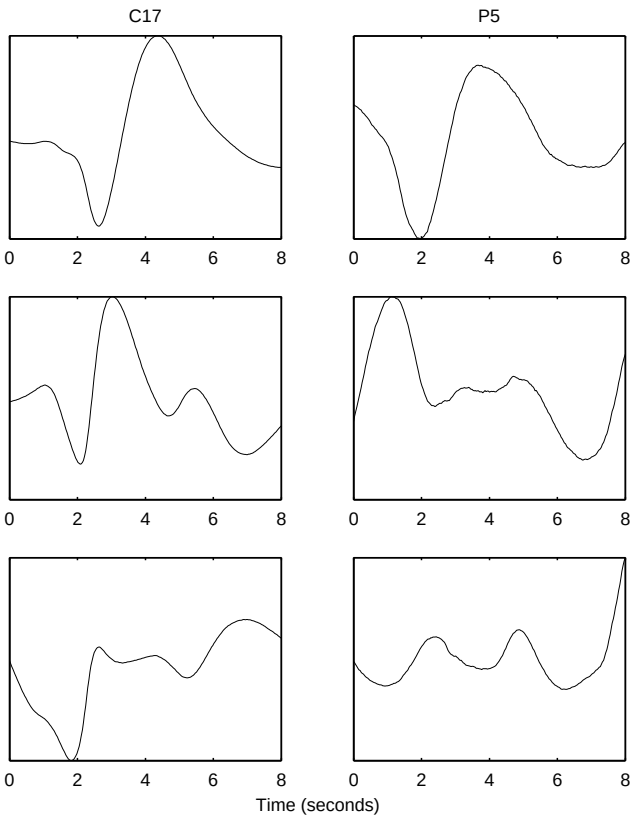

Fig. 7. Eigenvectors corresponding to the three largest eigenvalues of the correlation matrix of the measured GSRs for healthy control (C17) and psychotic patient (P5). The corresponding eigenvalues are $\lambda_1 = 0.8089$, $\lambda_2 = 0.1159$ and $\lambda_3 = 0.0443$ for control subject and $\lambda_1 = 0.4235$, $\lambda_2 = 0.2447$ and $\lambda_3 = 0.1442$ for psychotic patient.

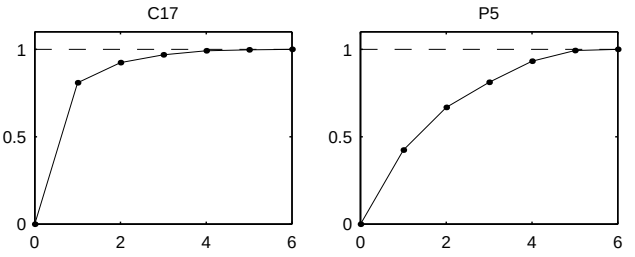

Fig. 8. The cumulative sum of largest eigenvalues for healthy control (C17) and psychotic patient (P5) as a function of the number of eigenvalues summed ($\bullet$). The sum of all eigenvalues is 1 ($--$).

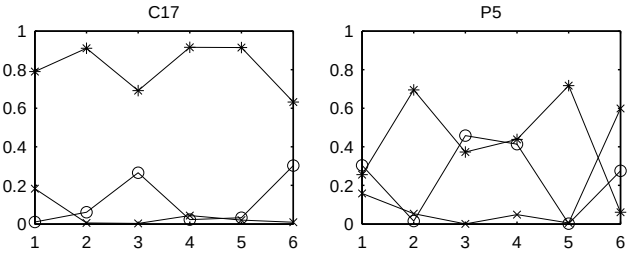

Fig. 9. Coefficients $\theta_j^2(1)$ ($*$), $\theta_j^2(2)$ ($\circ$) and $\theta_j^2(3)$ ($\times$) for selected measurements ($j = 1, \ldots, 6$) for healthy control (C17) and psychotic patient (P5).

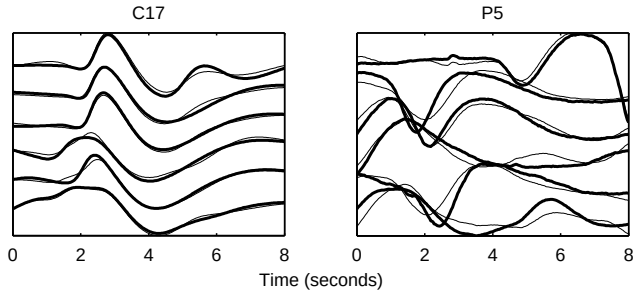

Fig. 10. Selected normalized measurements (bold line) and estimates (thin line) for healthy control (C17) and psychotic patient (P5).

and approaches the maximum earlier. This is characteristic for a highly coupled signal set according to [12].

When calculating parameters θ and estimates for measurements the number of eigenvectors to be used in the observation matrix H_S has to be decided. Some rules for this decision are discussed in [15]. Here we use three most dominant eigenvectors, which are presented in Fig. 7. Parameters θ are solved from (6) and the squares $\theta_j^2(1)$, $\theta_j^2(2)$ and $\theta_j^2(3)$ that describe the mean square contributions of the three largest eigenvectors to selected measurements ($j = 1, \ldots, 6$) are presented in Fig. 9. For control 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 patient the difference between contributions of the first and second factor is not so evident and also the third factor is effective in few cases.

Estimates for the six selected waveforms calculated from equation (7) are presented in Fig. 10.

D. Cluster analysis

In this section we examine the clustering of the two subject groups. Since our purpose is to discriminate GSR sets, as opposed to the single responses, the clusterization will mainly consist on the eigenvalues, that describe the mean square contributions of corresponding eigenvectors. However the regression parameters $\theta_j(i)$ are also considered, because they contain all the information about the nature of single GSRs. All clusters are formed with a hierarchical agglomerate clustering algorithm, with cluster-wise distance defined by equation (10).

First we show the clusterization of the groups when only six most substantial waveforms, normalized to unit norms, are analyzed. Fig. 11 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 control subjects (14/20) are clearly clustered in the bottom right corner. For these controls the contribution of two largest eigenvectors is over 85% and the reconstruction error when three largest eigenvectors are used in reconstruction is less than 5%. But there is also six controls among psychotic patients.

CLASSIFICATION

Group	Ω_1	Ω_2
Controls (+)	14	6
Patients (∘)	0	13

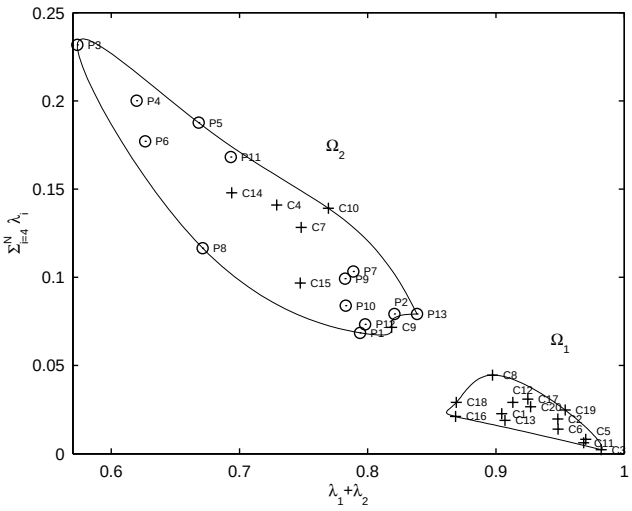

Fig. 11. Clusterization of healthy controls (+) and psychotic patients (∘) with respect to the sum of largest eigenvalues $\lambda_1+\lambda_2$ and sum $\sum_{i=4}^{N}\lambda_i$. Six most substantial waveforms ($N = 6$), normalized to unit norms, were analyzed.

CLASSIFICATION

Group	Ω_1	Ω_2
Controls (+)	12	8
Patients (∘)	0	13

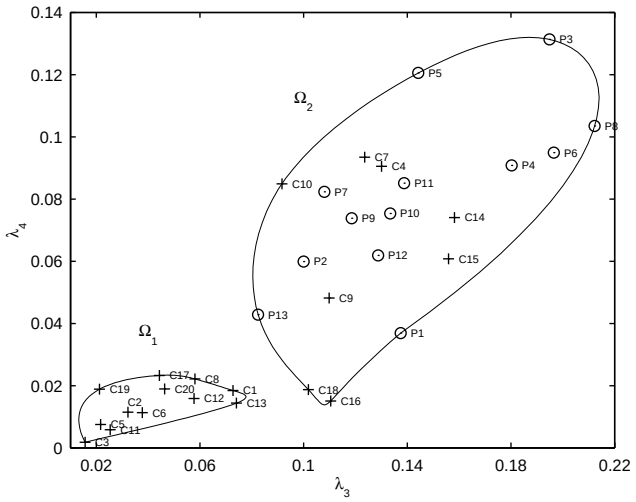

Fig. 12. Clusterization of healthy controls (+) and psychotic patients (∘) with respect to their third and fourth largest eigenvalues λ_3 and λ_4. Six most substantial waveforms ($N = 6$), normalized to unit norms, were analyzed.

The same kind of clustering is seen in a plot of third eigenvalue λ_3 with respect to fourth eigenvalue λ_4 in Fig. 12. Both λ_3 and λ_4 are small for a majority of control subjects. Now eight control subjects are among psychotic subjects and six of these are in fact the same ones which deviated from other controls in Fig. 11. Figs. 11 and 12 clarify the capabilities of PCA to discriminate GSR sets, even though the clusterizations are not complete.

For comparison we perform PCA to all 11 waveforms without normalization. Then PCA is sensitive to signal amplitudes and thus the strongest signals dominate the shapes of the eigenvectors. The clusterization with respect to the sum $\lambda_1+\lambda_2$ and $\sum_{i=4}^{N}\lambda_i$ is shown in Fig. 13. There is four psychotic patients among the controls group. This is due to a strong habituation in waveform amplitudes as can be seen from Fig. 14 where parameter C_h with respect to $\lambda_1+\lambda_2$ is presented. Figure shows a strong habituation for three of the four patients clustered into controls group.

VI. Discussion

We have proposed an analysis method for galvanic skin responses. GSR responses of different nature can be discriminated with principal component analysis based procedure. Regression methods are also needed when single responses are of interest. Previously PCA has been applied to SCRs in [18] without application to clustering, but the procedure presented in this paper is a totally new approach in GSR analysis. As a specific application the method was applied to GSR sets of healthy controls and psychotic patients and a good discrimination of the two groups was obtained.

The best obtained discrimination of healthy controls and psychotic patients is presented in Fig. 11. All patients were ranked correctly, but six controls were ranked as patients. This gives an overall correct ratings of 82%. A significant fact is that none of the psychotic patients were ranked into controls group. The same data sets were visually examined and rated into patient and normal categories by 20 evaluators (10 physicists and 10 physicians) and an overall average for correct ratings of 78% was attained [19]. All evaluators were shown a picture of the single trial trend of a typical control and patient beforehand. None of the evaluators had any previous experience in GSR analysis, but 10 of them were professionals in signal processing and EEG analysis.

There were altogether six false-positive classifications (subjects C4, C7, C9, C10, C14 and C15 were classified as patients) in Fig. 11. These misclassifications are partly due to the habituation of response amplitudes suggesting the extinction of GSR in repetitions. Responses of subject C4 seem to present abnormally large measurement noise, since the amplitudes of the responses are exceptionally small. Subject C4 was not however rejected because there were no problems in the recordings in our knowledge. Despite these misclassifications it should be noted that all patients were ranked correctly giving the proposed method a sensitivity of 100%. Thus, there were no false-negative rankings. This fact enables the method to be applied in clinical practice as a screening test for specific risk-groups and prodromal patients in future. The method is inexpensive and noninvasive and therefore appropriate for the screening of larger

CLASSIFICATION

Group	Ω_1	Ω_2
Controls (+)	17	3
Patients (∘)	4	9

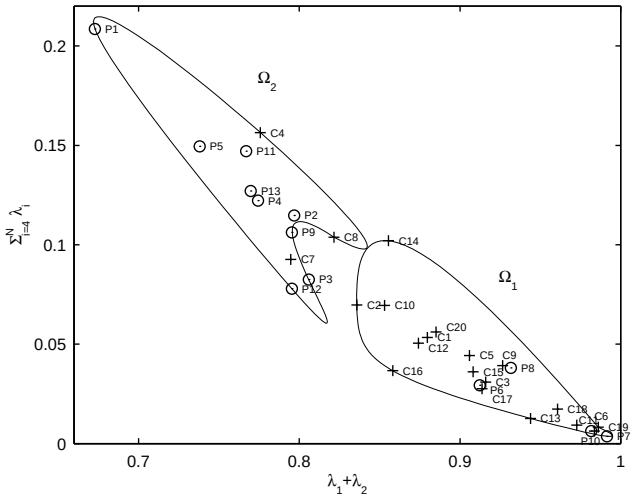

Fig. 13. Clusterization of healthy controls (+) and psychotic patients (∘) with respect to the sum of largest eigenvalues $\lambda_1 + \lambda_2$ and sum $\sum_{i=4}^{N} \lambda_i$. All 11 non-normalized waveforms ($N = 11$) were analyzed.

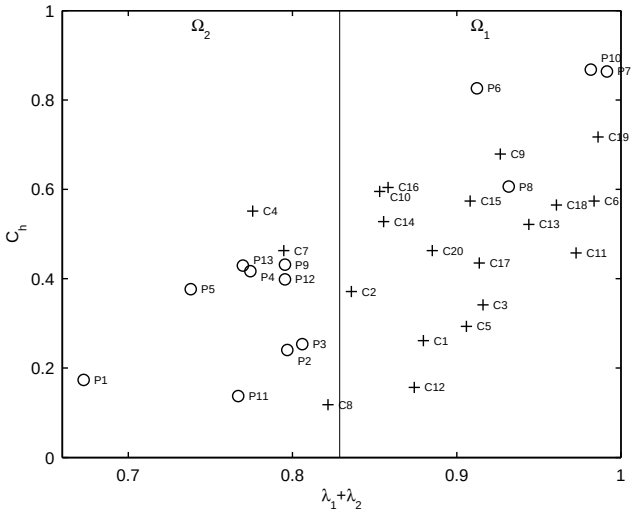

Fig. 14. Plot of healthy controls (+) and psychotic patients (∘) with respect to coefficient C_h and sum of largest eigenvalues $\lambda_1 + \lambda_2$.

populations as well. Subjects rated falsely as positive can be sorted out in more focused further follow-up studies (e.g. psychiatrist's interviews or MRI).

One significant factor effecting on the proposed method is the selection of the number of waveforms N per subject to be analyzed. It is practically impossible to define a generic criterion for the selection of N. Instead N should be selected based on the experimental procedure. For example the number of stimuli and the surprising effect of these are both decisive in selecting N. When the stimulus is repeated several times it naturally loses its surprising effect and is probably not able to evoke any GSR response. This habituation should be taken into account when selecting N, since totally habituated responses do not carry information about the true GSR and can therefore cause distortion to analysis. In this study the selection $N = 6$ seemed to be appropriate considering the number of stimuli and the habituation rate.

It should be emphasized that the basis for the proposed method is the experimental procedure. For the used test the 11 successive GSR responses were typically clearly patterned for healthy controls whereas for psychotic patients the lack of time-locking was characteristical. Because of such clear distinctions in morphologies of GSR sets for healthy controls and psychotic patients the second order statistics of the proposed method is sufficient for obtaining good results in practice. However for more complicated test procedures some nonlinear methods could be necessary. One advantage of the proposed method is that the dimensionality of data set is reduced while retaining as much as possible of the information in the original data. This reduction enables the graphical representation of the data and further analysis of clusterization. Another important property is the formability of the method according to the problem. Adjustments of the method can and should be done based on the signals under consideration. In this study we used the sum of two largest eigenvalues as a measure of similarity of a GSR set instead of just the first eigenvalue. The second eigenvalue was included because it allows some variation in response latencies, which is thought to be normal in GSRs. The capability of the normalized correlation coefficient, which is a measure of similarity between two waveforms, to discriminate GSR sets was also tested, but results were not as good as were results obtained from PCA.

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] M. Baba, Y. Watahiki, M. Matsunaga, and K. Takebe, "Sympathetic skin response in healthy man," *Electromyogr Clin Neurophysiol*, vol. 28, pp. 277–283, 1988.

[3] 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.

[4] B. Elie and P. Guiheneuc, "Sympathetic skin response: normal results in different experimental conditions," *Electroencephalogr Clin Neurophysiol*, vol. 76, pp. 258–267, 1990.

[5] 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.

[6] M. Toyokura, "Waveform and habituation of sympathetic skin response," *Electroencephalogr Clin Neurophysiol*, vol. 109, pp. 178–183, 1998.

[7] T. Deltombe, P. Hanson, J. Jamart, and M. Clérin, "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.

[8] C. Lim, C. Rennie, R. Barry, H. Bahramali, I. Lazarro, B. Manor, and E. Gordon, "Decomposing skin conductance into tonic and phasic components," *Int J Psychophysiol*, vol. 25, pp. 97–109, 1997.

[9] W. Knezevic and S. Bajada, "Peripheral autonomic surface po-

tential. a quantitative technique for recording sympathetic skin conduction in man," *J Neur Sci*, vol. 67, pp. 239–251, 1985.

[10] A. V. Rotterdam, "Limitation and difficulties in signal processing by means of the principal-components analysis," *IEEE Trans Biomed Eng*, vol. 17, pp. 268–269, 1970.

[11] J. Möcks, "The influence of latency jitter in principal component analysis of event-related potentials," *Psychophysiol*, vol. 23, pp. 480–484, 1986.

[12] 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.

[13] 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.

[14] E. Oja, *Subspace Methods in Pattern Recognition*. Wiley, 1983.

[15] I. Jolliffe, *Principal Component Analysis*. Springer-Verlag, 1986.

[16] A. B. Geva and H. Pratt, "Unsupervised clustering of evoked potentials by waveform," *Med Biol Eng Comput*, vol. 32, pp. 543–550, September 1994.

[17] J. Boberg, *Cluster Analysis, A Mathematical Approach with Applications to Protein Structures*. PhD thesis, University of Turku, Department of Computer Science, 1999.

[18] W. Roth, K. Dorato, and B. Kopell, "Intensity and task effects on evoked physiological responses to noise bursts," *Psychophysiol*, vol. 21, pp. 466–481, July 1984.

[19] M. Valkonen-Korhonen, A. Koistinen, P. Karjalainen, J. Lehtonen, and J. Karhu, "Loss of time-organized sympathetic skin responses in acute psychosis," *J Nerv Ment Disease*, 2001. In Press.