PARAMETERISATION COMPARISON FOR THE DETECTION OF PANIC DISORDER USING TIME-FREQUENCY TRANSFORMS AND SUPPORT VECTOR MACHINES

H. Dietl[1] and S. Weiss[1]

[1]Communications Research Group, ECS, University of Southampton, Southampton, United Kingdom
hwd01r@ecs.soton.ac.uk, s.weiss@ecs.soton.ac.uk

Abstract: In this paper we compare the effect of the parameterisation on the automatic detection of diseases based on biomedical data. Exemplarily, we study the analysis of event related brain potentials in patients suffering from panic disorder, whereby the data comprises responses to neutral and panic causing stimuli. This data is parameterised by time-frequency (TF) transforms, from which features are selected by a statistical test. The selected features represent the input to a support vector machine classifier yielding a detection rate for the TF parametrised data. This is compared with detection rates obtained for unparameterised time domain data.

Keywords: Transient, evoked, otoacoustic, emissions, wavelet, support, vector, machines.

INTRODUCTION

In medical facilities it is a common issue to judge the responses of patients to stimuli in order to determine a potential physiological or psychological illness [1, 2]. In cases where the response can be measured as an electrical signal, the signal evaluation used to be primarily based on an expert's decision regarding the waveforms of averaged signals. Such waveforms and often parameters derived from these as presented by standard clinical measurement devices were treated as additional information only. Recently, however, automated evaluation methods based on signal processing approaches have more and more frequently enhanced or even replaced the expert's judgement [1, 3]. As a result, many different propositions were made concerning statistical signal evaluation in an effort to enhance or perhaps even replace the decision of a human expert.

Here, we contribute to evaluate the separation of biomedical data to detect potential illnesses by firstly showing the application of TF methods and statistical tests to select features as introduced in [3]. Then, the selected features are used as an input to a support vector machine classifier which returns a trained support vector classification network. Using this network, a detection rate for a test data group is received. This method, as outlined in Fig. 1 is applied to panic disorder data collected from one patient. We are especially interested in investigating the influence of the parameterisation and therefore, we compare the detection results received for the selected features based on the TF parameterisations and unparameterised time domain data. As the amount of panic disorder data is limited, the TF transforms as well as the statistical tests are applied to all available data, before it is split into training and test data sets, as illustrated in Fig. 1 to ensure a robust parameterisation. However, as the main purpose of this contribution is to evaluate the parameterisation, the drawback compared to a study where the parameterisation is based on a training data set only can be accepted.

The paper is organised as follows. Firstly, we will introduce the TF transforms for the parameterisation of the data. Then, a method to isolate indicative parameters, which can be used for distinguishing is discussed. Next, we describe the support vector machine classifier which gives a detection rate for the test data based on a network resulting from the training data. The test and training data are received by splitting the data used for identifying distinctive coefficients. The paper closes with test results and conclusions for the application to panic disorder and acknowledgements.

TRANSFORMATION METHODS

In the following, we discuss transform methods to parametrise biomedical data with the aim of expressing its features by as few coefficients as possible.

To take the transient nature of biomedical waveforms into account, TF transforms are used for parameterisation of the data. For our application to panic disorder, TF transforms with a good time resolution are required [3]. The discrete wavelet transform (DWT) however generally yields a good frequency resolution and poor time resolution at low frequencies, resulting in a too coarse time segmentation in the frequency range of interest. Therefore, we concentrate on wavelet packet (WP) transform, whose level of decomposition can be adapted to fit the nature of the data, as well as the Gabor Frame (GF) decomposition, which yields a uniform tiling of the TF plane and hence can provide a desired resolution in a specific TF segment.

For the transforms considered, we choose a matrix notation

$$\mathbf{y} = \mathbf{T}_{\mathrm{j}} \cdot \mathbf{x} \quad , \tag{1}$$

where $\mathbf{x}$ represents one discrete and finite measurement in the time domain with N elements, $\mathbf{y}$ is vector holding the transformation coefficients and $j = \{\mathrm{WP}, \mathrm{GF}\}$ the potential transform method. While fast implementations of WP [4] and GF [5] avoid matrix implementations, the calculation of a limited number of significant elements in $\mathbf{y}$ can be performed faster by extracting the according rows from $\mathbf{T}_{\mathrm{j}}$. As the data in $\mathbf{x}$ is finite, a symmetric extension (extension by reflection) of the data is incorporated into $\mathbf{T}_{\mathrm{j}}$ according to [6].

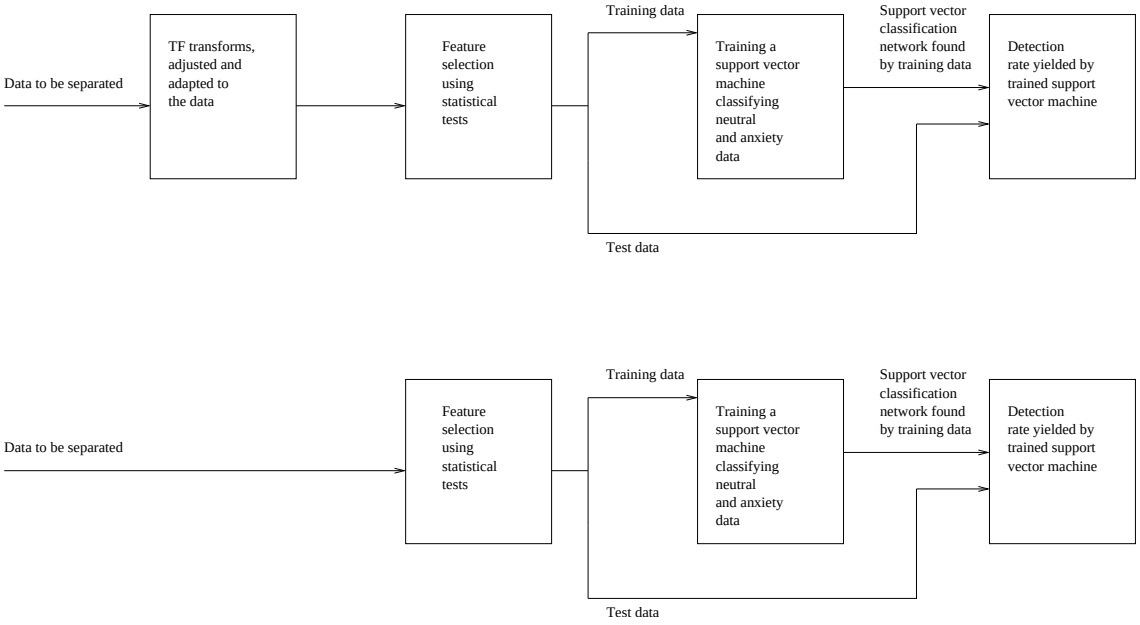

Fig. 1: Overview: Detection comparison study for (top) parameterised data and (bottom) time domain data.

Discrete Wavelet and Wavelet Packet Transformation

The WP is based on a discrete wavelet transformation (DWT) which is a fixed transform based on a "mother wavelet" from which the transformation coefficients are derived by scaling, translation and sampling. To comply with the symmetric extension in $\mathbf{T}_{\mathrm{j}}$, the mother wavelet must be symmetric. Here, we have chosen the Mallat wavelet [4] for which good results have been reported in similar studies [1]. We firstly review the DWT very briefly to lay the foundation for the description of the WP.

The DWT transform with a matrix $\mathbf{T}_{\mathrm{DWT}} \in \mathbb{R}^{K\times N}$ results in the vector $\mathbf{y}[k] \in \mathbb{R}^K$

$$\mathbf{y} = \begin{bmatrix} y[0] & y[1] & \cdots & y[K-1] \end{bmatrix}^{\mathrm{T}} \qquad (2)$$

containing the DWT coefficients $y[k]$, $k = 0(1)K-1$ of the vector $\mathbf{x}$ with $K = N$. Each coefficient approximately covers a TF tile in Fig. 2a).

The WP transform is an adaptive transformation similar to the DWT but with a different partitioning of the TF plane. The advantage of this approach compared to the DWT is that the entropy of $\mathbf{y}$ shall be minimised through variable levels of decomposition such that the energy is concentrated in as few coefficients as possible. One suitable measure for such a concentration is given by the Shannon entropy

$$\epsilon(\mathbf{y}) = -\sum_{k=0}^{K-1} \ln(\parallel \mathbf{y} \parallel^2) \cdot \parallel \mathbf{y} \parallel^2 , \qquad (3)$$

where ln is the natural logarithm and $\parallel \mathbf{y} \parallel$ the Euclidean vector norm of $\mathbf{y}$. The adaptive transformation approach is implemented in the following way: The measure in (3) is calculated for every wavelet decomposition level and if the entropy is decreased from one decomposition to the next, the decomposition will be continued. If the entropy from one decomposition level to the next does not decrease, the decomposition will be stopped.

Fig. 2 shows a sample DWT and WP decomposition, whereby the DWT can be considered as a special case

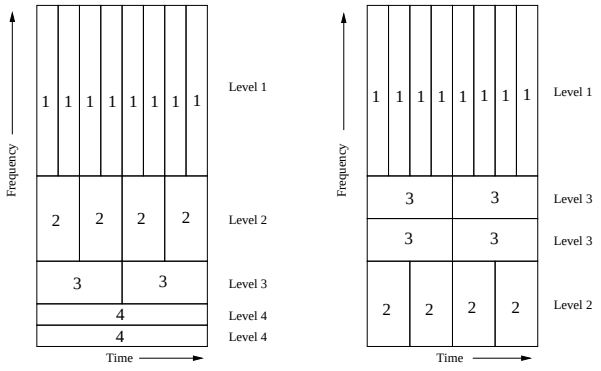

Fig. 2: Time-frequency tiling comparison between a) a DWT and b) a sample WP decomposition

of the WP transformation where the TF plane is segmented dyadically form one level to the next. The matrix $\mathbf{T}_{\mathrm{WP}} \in \mathbb{R}^{K\times N}$ is found as follows: For each measurement vector $\mathbf{x}$ to be transformed, the WP decomposition is calculated. Then, the decomposition that shows the minimum entropy averaged over all measurements is selected among the decompositions for each measurement as the optimal WP decomposition. The difference between the wavelet matrix $\mathbf{T}_{\mathrm{DWT}}$ and the wavelet packet matrix $\mathbf{T}_{\mathrm{WP}}$ lies in the change of some rows. For the example in Fig. 2, the rows that contain the level 2 coefficients in $\mathbf{T}_{\mathrm{DWT}}$ are replaced by level 3 coefficients in $\mathbf{T}_{\mathrm{WP}}$.

Gabor Frames

The GF are the second transform method applied to the data. This transform is fixed and its main difference compared to the WP is that the GF parametrisation is overcomplete and yields complex valued transform coefficients. We can also choose from various prototype functions to find the best one matching our analysis. The transform implemented here is an oversampled generalised DFT filter bank according to [5]. The GF transform can be regarded as a short time Fourier transform with more restrictions for the translated window, which is the prototype filter for the GF. Fig. 3 shows a sample time frequency tiling, which is characterised by the uniform resolution of the GF. Gabor functions are derived

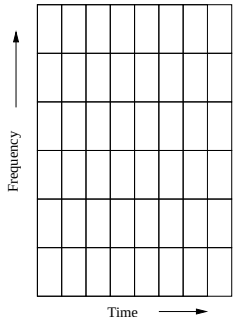

Fig. 3: Time frequency tiling for GF

from a prototype function $h[n]$ by modulation, usually

$$h_m[n] = h[n] \cdot e^{j\Omega mn}, \tag{4}$$

and give a decomposition according to

$$y_m[k] = \sum_n x[k \cdot D - n] \cdot h_m[n], \tag{5}$$

where D is the decimation factor.

As an example for a GF via an oversampled generalised DFT filter bank [5] let us assume a filter length of $W = 448$ for the prototype function, a frequency segmentation of $S = 32$ uniform scales, a decimation $D = 28$ and length of the vector $\mathbf{x}$ of $N = 224$. This setting yields a transform matrix $\mathbf{T}_{\mathrm{GF}} \in \mathbb{C}^{K \times N}$ with $K = (N/D + 1) \cdot S/2$, $K = (\frac{224}{28} + 1) \cdot \frac{32}{2} = 144$. Similar to a DFT, this type of transform yields a symmetry in the transform parameters for real valued input data. Hence, in general only half of the coefficients need to be retained, as the remainder is complex conjugated only and therefore redundant which is represented by the term $S/2$ in the formula above. Filters with varying length, frequency and time segmentations are used to determine a matrix $\mathbf{T}_{\mathrm{GF}}$ that optimises the parametrisation of the data. Apart from the restrictions that N needs to be an integer multiple of D and W/K needs to be an even integer, there are no other constraints in terms of the length of the signal to be analysed and the length of the analysing filter. This can be regarded as advantageous.

DIFFERENCE EVALUATION

Based on the various parameterisations derived in the previous section, we will identify a set of coefficients that allows us to differentiate between the presented neutral and anxiety words.

F-Test

Prior to the selection of significant coefficients that represent the main characteristics of the data, an F-test [7] is conducted to determine which method is used to identify them. The aim of this test is to determine whether two data sets are sampled from normal distributions with the same variances. If a value for the significance level P of lower than 0.05 is obtained by the F-test, we conclude that the hypothesis is rejected and the two data sets are sampled from normal distributions having different variances. The value of $P = 0.05$ is a limit commonly used in medical research [7]. When the sets $\mathbf{x}_1$ and $\mathbf{x}_2$ contain the series for one transformed coefficient k for all measurements taken for data set 1 representing the number of presented neutral words and data set 2 representing the number of presented anxiety words, they can be compared by the F-value, which is given by [7]

$$F = \frac{\sigma_1^2}{\sigma_2^2}, \tag{6}$$

with σ_1^2 and σ_2^2 being the variances of the two data sets. To receive the significance level P for the F-test, we need to define the degrees of freedom for the two data sets according to

$$\begin{aligned} \nu_1 &= L_1 - 1 \text{ and} \\ \nu_2 &= L_2 - 1, \end{aligned} \tag{7}$$

with L_1 and L_2 being the number of samples, ν_1 the degrees of freedom for the data set 1 and ν_2 the degrees of freedom for the data set 2. With the F value defined by (6) and the degrees of freedom ν_1 and ν_2, the significance level P for the F-test can be determined from lookup tables in literature, e.g. [7]. The tabulated values of F are all greater than 1, the two data sets in (6) need to be labelled such that $\sigma_1^2 \geq \sigma_2^2$. If the outcome of the F-test confirms that the two data sets are sampled from distributions with equal variances, we can subsequently conduct a t-test to determine distinctive coefficients. If the result of the F-test is that the underlying distributions from which the two data groups are sampled possess different variances we conduct a ut-test. The t-test and the ut-test are defined in the next subsection.

t- and ut-Tests

The t-test gives the probability that two data sets sampled from potentially two different distributions with identical variance possess different mean values, for which a sig-

nificance is returned. The t-value is defined as [7]

$$t = \frac{\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2}{\sqrt{\frac{\sigma_1^2}{L_1} + \frac{\sigma_2^2}{L_2}}} = \frac{\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2}{\sigma\sqrt{\frac{1}{L_1} + \frac{1}{L_2}}} \quad , \tag{8}$$

with $\sigma^2 = \sigma_1^2 = \sigma_2^2$. The values $\overline{\mathbf{x}}_1$ and $\overline{\mathbf{x}}_2$ represent the means for the two data sets, according to

$$\overline{\mathbf{x}}_i = \frac{1}{L_i} \cdot \sum_{l=0}^{L_i-1} x_i[n], \qquad i \in \{1, 2\}, \tag{9}$$

with $\mathbf{x}_i^T = [x_i[0]\ x_i[1]\ ...\ x_i[L_i - 1]]$.

The t-value also corresponds to a certain significance level P, which can be looked up from tables [7], with the degrees of freedom defined by $\nu_t = \nu_1 + \nu_2 = L_1 + L_2 - 2$. A smaller value for P indicates that the data sets have a significantly different mean. For example, for $P = 0.01$ the probability that the differences in the means are due to a sampling error is 1%. To identify distinctive coefficients for our study, the determination of the applied significance level will be discussed in the next subsection. The two tested distributions for our study were the distributions for a specific transform parameter over the two data sets.

For the case that the F-test yields a difference in variances such that the t-test cannot be used, we apply a ut-test for unequal variances defined as

$$ut = \frac{\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2}{\sqrt{\frac{\sigma_1^2}{L_1} + \frac{\sigma_2^2}{L_2}}} \quad . \tag{10}$$

According to [7], for data sets sampled from distributions with unequal variances, the t distribution can be approximated by the ut value if the t table is entered at the following defined degree of freedom:

$$\nu_{ut} = \frac{(\sigma_1^2/L_1 + \sigma_2^2/L_2)^2}{\frac{(\sigma_1^2/L_1)^2}{L_1 - 1} + \frac{(\sigma_2^2/L_2)^2}{L_2 - 1}}. \tag{11}$$

This test tends to be less powerful than the usual t-test, since it uses fewer assumptions [7]. However, for our application to panic disorder data, all identified distinctive coefficients have been isolated by the t-test. The main purpose of the ut-test is to have an analysis tool for all coefficients at hand whether they show equal variances or not.

To determine a significance level P, the relation of the t-test to the receiver operating characteristic (ROC) analysis is shown in the next subsection.

Relation Between ROC Analysis and t-Test

We firstly describe the receiver operating characteristic (ROC) analysis and then introduce the connection between ROC curves and the t-test and how we used the ROC analysis for our system.

A good measure for differentiation between two distributions are ROC curves [8], since the area under the ROC curve measures the separability independent of the selection of any threshold. Therefore, they have become remarkably useful in medical decision-making [8].

Firstly, we start by introducing the terms sensitivity and specificity [8] as we refer to these terms later when showing the results of our study. We assume we a have population consisting of healthy controls and patients that suffer from a certain disease but do not know or cannot express their suffering (e.g. hearing loss in infants). Our goal is to determine the patient group out of the population. For this, we run an imaginary test on the population. The outcome of the test consists of one test parameter which is either positive indicating the tested person is sick or negative meaning the tested person is healthy. In order to evaluate the performance of that test, the following values can be used, as illustrated in Table 1.

The interrelationship equations in the table result from the fact that each person is classified as healthy or sick by the test. In the following the terms represented in the table will be used equivalently, meaning that we always refer to the true positive rate when speaking of sensitivity or hit rate.

Secondly, we continue by introducing the ROC curves. An ROC curve is a graphical representation of the trade off between sensitivity and specificity for every possible cut off. By tradition, the plot of the ROC curve shows the false positive rate on the x axis and the hit rate on the y axis. However, based on the interrelationships shown in Table 1, the axis of the ROC curve can be modified. Suppose the above mentioned test parameter yields distributions for the sick and healthy groups as illustrated in Fig. 4 on the left.

The solid line represents the distribution of the test parameter for the patient group, the dashed line for healthy

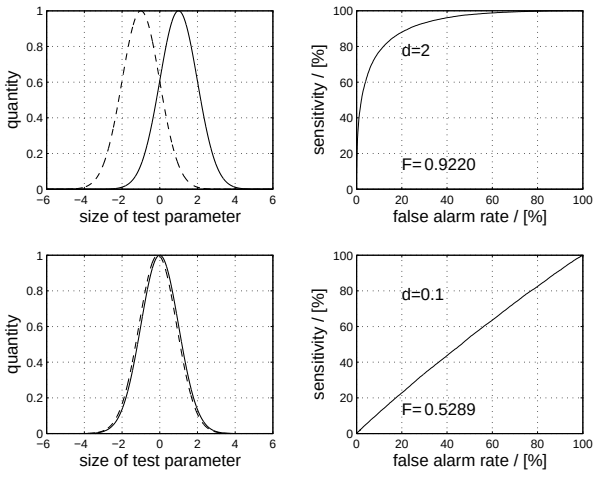

Fig. 4: ROC explanation: sample distributions (left) for sick (solid) and healthy (dashed) groups assuming an imaginary test parameter yielding ROC curves (right).

controls on the left in the figure. In the upper case the distributions have a distance of their means equalling two standard deviations, $d = 2$ whereby the distributions possess the same variances. The upper right of the figure shows the resulting ROC curve with the sensitivity and

	Test for disease		interrelationship
	sick group	healthy group	
Test result: positive	true positive (TP) rate in %, sensitivity, hit rate	false positive (FP) rate in %, false alarm rate	TP + FN = 100%
Test result: negative	false negative (FN) rate in %	true negative (TN) rate in %, specificity	TN + FP = 100%

Table 1: Definition of sensitivity and specificity.

the false alarm rate as axis labels. The false alarm rate describes the specificity as shown in Table 1. The lower case shows the same but for $d = 0.1$.

Ideally, for a good separation, the sensitivity and the specificity should be very high. As Fig. 4 illustrates, a value for the area under the ROC curve close to 1 yields a relatively good separation, whereas a value close to 0.5 yields a very poor performance when taking both the sensitivity and separability into account.

Having introduced the ROC analysis, we continue by showing a connection between the area under the ROC curve and the significance level received by the t-test.

Here, we make use of the ROC analysis to evaluate and obtained a significance level for the t-tests or ut-tests. Moreover, for finding a prototype filter for the GF transform, the ROC analysis is applied. The relation is investigated as follows. Different values for the area under the ROC curve are determined. For these values, two Gaussian distributions are generated. From these distributions, a certain number of random samples are taken out and based on a t-test or ut-test, the significance level is calculated for these samples originating from the Gaussian distributions. This calculation is repeated with random samples from the distributions and the significance level is averaged until it converges. The ROC analysis is independent of the sample size whereas the t-test and ut-test depend on it. Therefore, different sample sizes yield different relations, which is illustrated in Fig. 5 for a significance level P received by the t-test. When using the ut-test to investigate this relation, the results are very similar, e.g. no differences can be observed when adding the resulting curves to the ones illustrated in Fig. 5.

For our analysis of the panic disorder, we deal with a sample size of 24. Table 2 shows the areas under the ROC curve for the most commonly used significance levels P [7] in more detail.

Area under the ROC curve	Significance level P
0.717	0.05
0.778	0.01

Table 2: Area under ROC curve and significance levels P for a sample size of 24.

In most social research significance levels of $P = 0.05$ or $P = 0.01$ are used to determine difference between two sets of data [7]. In other studies such as [1], ROC values of ≈ 0.77 are found and stated to yield acceptable separation performance. Therefore, we choose

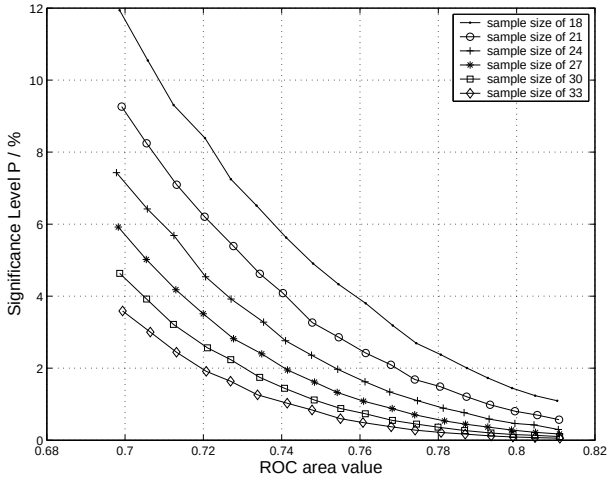

Fig. 5: Significance Level P for t-test over area under the ROC curve value for different sample sizes.

a significance level of $P = 0.01$ for our study to obtain distinctive transform coefficients. Also, when testing different prototype filters for the GF transform, we choose the one having the largest ROC value for the isolated transform coefficients. Next, we discuss the support vector machine classifier we applied.

SUPPORT VECTOR MACHINES

In the following, we give a brief introduction to support vector machines. For a detailed description, we refer to [9, 10]. We consider a two class classification problem, namely one class defined by anxiety causing data, and one class describing neutral data, which is split into two data groups, namely training data and test data, respectively. The training data is described as a set of training vectors $\{\mathbf{p}_i\}_{i=1\ \ldots\ M}$ with corresponding binary labels $S_i = 1$ for the one class, e.g. neutral data, and $S_i = -1$ for the second class, e.g. anxiety causing data. The SVM conducts a classification of a test vector $\mathbf{t}$ by assigning a label $\hat{S}$ by calculating

$$\hat{S} = \text{sign}(f(\mathbf{t})) \quad \text{with} \quad f(\mathbf{t}) = \sum_i \alpha_i S_i K(\mathbf{t}, \mathbf{p}_i) + b. \tag{12}$$

The α_i are called weights and b is the bias, which are SVM parameters and adopted during training by maximising

$$L_D = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j S_i S_j K(\mathbf{p}_i, \mathbf{p}_j) \tag{13}$$

under the constraints

$$0 \leq \alpha_i \leq C \quad \text{and} \quad \sum_i \alpha_i S_i = 0 \qquad (14)$$

with C being a positive constant which weighs the influence of training errors. $K(\cdot,\cdot)$ is called kernel of the SVM. If there is a solution for α_i, a value for b is determined. There are several commonly used kernels for SVM, which give some flexibility for the underlying application. Many implementations of kernels can be found in literature, whereby two popular ones are:

- Gaussian kernel:
  $K(\mathbf{p}_i, \mathbf{p}_j) = \exp(-\gamma \left\|\mathbf{p}_i - \mathbf{p}_j\right\|^2)$,
- polynomial kernel: $K(\mathbf{p}_i, \mathbf{p}_j) = (\mathbf{p}_i^T \cdot \mathbf{p}_j)^d$,

where γ is a kernel parameter for the Gaussian kernel and d the order of the polynomial kernel.

If $K(\cdot,\cdot)$ is positive definite, (13) and (14) is a convex quadratic optimisation problem, which converges towards the global optimum assuringly. This optimisation can be quite demanding in terms of computation time for real-world problems, and therefore, sophisticated algorithms like the sequential minimal optimisation (SMO) [9] are used for the solution.

Usually $\alpha_i = 0$ for the majority of i and thus the summation in (12) is limited to a subnet of the $\mathbf{p}_i$, which therefore is called the set of support vectors. For Gaussian kernels, when using stretched out values for the limitation of training errors defined by C and the kernel parameter γ, so called overfitting can occur meaning that all M training vectors are identified as support vectors. To avoid this for our application, we have chosen to use the polynomial kernel of order $d = 3$ as this is assumed to be the best comprise between computational time, avoiding overfitting and yielding a good detection rate for the test data.

RESULTS AND DISCUSSION

As discussed previously, we have different transform methods and a procedure to identify significant coefficients to being able to separate biomedical data. When splitting the data arbitrary in training and test data and applying a support vector machine as described in the previous section, we yield specific detection rates for training data as well as for the test data where the results for the test data describe the generalisation of the support vector classification network. We can also apply the tests described to unparameterised time domain data, deploy the support vector machine classification method and yield detection results for unparameterised data. By doing so, we arrive at an evaluation of the parameterisation methods. In the following, we will show this procedure applied to panic disorder data.

Description of the Data

Individuals with panic disorder are characterised by an abnormal fear of certain anxiety connected sensations such as palpitation, breathlessness, or dizziness [2]. The research into this disorder has led to studies investigating its symptoms by means of appropriate stimulation and measurement of the subsequent event related brain potentials [3]. In this context, visual stimulation has been performed with words causing panic disorder, whereby the EEG can be recorded showing event related potentials (ERP). Previous studies have resulted in revealing a low frequent transient waveform appearing approximately 300 ms after stimulus onset as a distinctive characteristic which is referred to as P300.

For our study, panic disorder ERP were measured for an anxiety patient who was presented with fear-inducing or neutral words tachistoscopically at the perception threshold of panic disorder. The patient's perception threshold for correctly identifying 50% of the words was determined with neutral words not used in the experiment. It can be assumed that the patient will recognise a greater number of anxiety words given at his perception threshold than neutral words [2]. Thus, it can be expected that the EEG exhibit an difference when neutral and anxiety words are presented.

The EEG was measured at the vertex electrode (Cz) synchronously to the stimuli, whereby the recordings were started 100 ms before the onset of the visual word stimulus. The data exemplary analysed in this study contains 24 neutral word presentations and 24 anxiety word presentations to one panic patient. Fig. 6 shows the average over the stimulus-synchronous EEG in reaction to the 24 words presented for each word category. The figure reveals a difference in the two averages with a stronger P300 and more positive EEG until approximately $t = 700$ ms in the panic disorder related data.

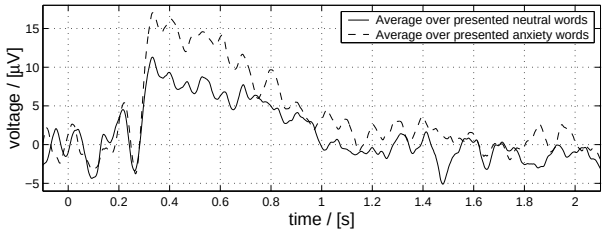

Fig. 6: Average over 24 EEG segments showing responses to anxiety related and neutral stimuli at the perception threshold.

Transform Adjustment

The optimal decomposition structure for the WP is found over minimising the entropy as described in the transformation section. The decomposition depth was limited to have at least 16 coefficients in one decomposition level as further decomposition would lead to a too coarse time segmentation. In terms of the Gabor transform, various filters were tested and it was found that using a prototype filter with length of 224, a frequency segmentation of 32 uniform scales and a time segmentation of 7 for the oversampling shows the best results for a ROC analysis.

Identified Coefficients and Difference Comparison

The coefficients to which the difference evaluation is applied were preselected whereby only coefficients are considered which contain 85% of the total energy. This is reasonable, as it reduces the probability to identify coefficients that contain noise only.

Fig. 7 shows the resulting coefficients when performing the difference evaluation on the parameterised data. The coefficients cover approximately the area of the P300 slow wave as it can be expected. They are all identified via a t-test according to a prior F-test whereby the threshold for the significance level for the F-test was $P = 0.05$, and for the t-test, it was set to $P = 0.01$.

For the unparameterised time domain data, 18 coefficients were identified, where the same statements as for the parameterised data apply.

Fig. 8 shows the difference of the averages of the neutral and anxiety EEG compared with its parameterisation by the identified coefficients for the two investigated transforms and for unparameterised time domain data.

It can be observed that the identified coefficients parameterise the P300 better for the WP than for the GF. However, the GF seems to show a better parameterisation than time domain data. In order to evaluate these results, a SVM analysis is applied next.

Detection Rates Yielded by SVM

As explained in the SVM section, a polynomial kernel of order 3 was used for the SVM. To find a significant value for the training error C, a leave-one-out (l-o-o) estimation of the error rate is applied as follows: From the training samples, remove the first example. Train the SVM on the remaining samples. Then test the removed example. If the example is classified incorrectly, it is said to produce a leave-one-out error.

In [9], an approach to estimate the maximum l-o-o error is shown avoiding training the SVM more than once, which is also used for our study. By changing the value for C stepwise, the minimum for the l-o-o error is found determining the SVM classification network.

Fig. 9 shows a SVM classification with a minimised l-o-o error for a WP parameterisation. This can be shown in a two dimensional plane as the for the WP, two coefficients were identified as significant. The two coefficients are illustrated in Fig. 7 (left).

In Fig 9, one class is defined by 12 points originating from arbitrary chosen neutral words and the second class

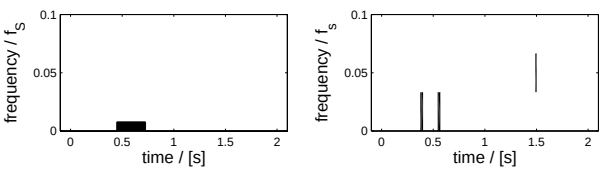

Fig. 7: Resulting coefficients for (left) WP and (right) Gabor transforms.

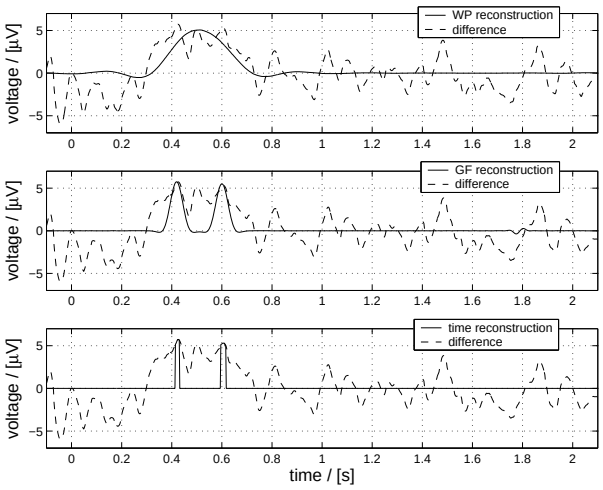

Fig. 8: Difference of neutral and anxiety EEG data compared with its parameterisation identified by the t-test for (top) WP, (middle) Gabor transforms and (bottom) unparameterised time domain data.

represents 12 panic causing words, also chosen arbitrary from the whole 24 defining one training data set. For this example, class one is completely assigned correctly; class two gets assigned incorrectly in $\frac{1}{3}$ of all cases. This rather asymmetric decision is due to the comparably small data size. Therefore, the SVM classification was conducted 100 times and averaged at the end meaning in loop, 12 arbitrary measurements for the two data groups were chosen, the SVM trained and the test group consisted of the remaining 12 measurements for each word category. Table 3 shows the results for this procedure for the training data.

It can be seen that the time domain data and the WP parameterisation yield comparably high values, whereas the GF identifies around two out of three words correctly for the training data. The number of support vectors is similar for each case.

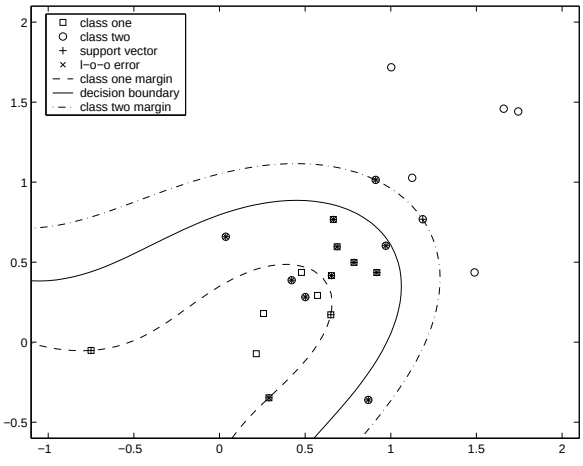

Fig. 9: SVM classification with two coefficients for a WP parameterisation on the axis.

	time domain	GF transform	WP transform
sensitivity	99.55 %	66.78 %	92.00%
specificity	93.13 %	65.88 %	96.28%
number of support vectors	15.68	16.65	17.39

Table 3: Results for training data.

Next, the more interesting results for the test groups are presented in Table 4. It can be seen that the WP

	time domain	GF transform	WP transform
sensitivity	67.34 %	78.49 %	82.32%
specificity	55.63%	54.39 %	52.82%

Table 4: Results for test data.

shows the overall best detection results for the test data, followed by the GF and the time domain data performing worst. However, for the specificity all three cases are similar around 50%.

What can be expected from Fig. 8 is confirmed: The WP parameterisation yields the best detection rates, followed by the GF transform and the simple time domain data performing worst. However, the specificity for the TF-transforms is not significant although the test data is used for the adjustment of the parameterisation methods. This can be due to the relatively small amount of data available. Moreover, the t-test for receiving distinctive coefficients may not be powerful. Therefore, the described system might show more encouraging results for the analysis of biomedical data which comprises more measurements than here and uses a different method for the extraction of the features from the parameterised data. However, recapitulating, it can be said that with both transforms an adequate overall detection of data of both categories, namely presented neutral and anxiety words, can be achieved better than without a parameterisation of the data.

CONCLUSIONS

We have presented an analysis comparing parameterised data by TF transforms with unparameterised data with the aim of differentiating between presented anxiety causing words and neutral words to a patient suffering from panic disorder.

The performance of the parameterisation methods were evaluated by identifying a distinctive coefficient set followed by a SVM classification. The obtained results where compared with the same analysis of time domain data without a parameterisation.

The results show that a parameterisation of biomedical data by TF transforms yield better detection rates than a mere time domain description. However, when using SVM for the classification and detection of diseases or abnormalities, a relatively high number of measurements is required.

ACKNOWLEDGEMENTS

The authors would like to thank Prof. Paul Pauli of the Dept. of Biological & Clinical Psychology, University of Würzburg, Würzburg, Germany, who kindly provided valuable expertise and the data.

REFERENCES

[1] S. Weiss, U. Hoppe, M. Schabert, and U. Eysholdt, "Wavelet Analysis of TEOAE for Differential Diagnosis of Cochlear Hearing Loss," in Asilomar Conf. on Signals, Systems, and Computers, Monterey, CA, November 2001.

[2] P. Pauli, G. Dengler, G. Wiedemann, P. Montoya, H. Flor, N. Birbaumer, and G. Buchkremer, "Behavioral and Neurophysiological Evidence for Altered Processing of Anxiety-Related Words in Panic Disorder," *Abnormal Psychology*, vol. 106, no. 2, pp. 213–220, 1997.

[3] H. Dietl and S. Weiss, "Categorisation of Panic Disorder by Time-Frequency Methods," in Asilomar Conf. on Signals, Systems, and Computers, Monterey, CA, November 2003.

[4] Stephane G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Tran. on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–692, July 1989.

[5] M. Harteneck, S. Weiss, and R. W. Stewart, "Design of Near Perfect Reconstruction Oversampled Filter Banks for Subband Adaptive Filters," *IEEE Tran. on Circuits & Systems II*, vol. 46, no. 8, pp. 1081–1086, August 1999.

[6] G. Strang and T. Nguyen, "Wavelets and Filter Banks," Wellesley–Cambridge Press, Wellesley, MA, 1996.

[7] P. Armitage, G. Berry, and J.N.S. Matthews, "Statistical Methods in Medical Research," Blackwell Science, Oxford, fourth edition, 2002.

[8] K. Chu, "An introduction to sensitivity, specificity, predictive values and likelihood ratios," *Emergency Medicine*, vol. 11, pp. 175–181, 1999.

[9] V.N. Vapnik, "Statistical Learning Theory," Cambridge University Press, New York: Wiley, 1998.

[10] C. Bahlmann, B. Haasdonk, and H. Burkhardt, "Online Handwriting Recognition with SVM – A Kernel Approach," in Proc. of the 8th Int. Workshop on Frontiers in Handwriting Recognition (IWFHR), pp. 49–54, 2002.