BIOSIGNALS AS AN ADVANCED MAN-MACHINE INTERFACE

Egon L. van den Broek
Center for Telematics and Information Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
vandenbroek@acm.org

Viliam Lisý
Agent Technology Center, Dept. of Cybernetics, FEE, Czech Technical University
Technická 2, 16627 Praha 6, Czech Republic
viliam.lisy@agents.felk.cvut.cz

Joyce H. D. M. Westerink
User Experience Group, Philips Research Europe, High Tech Campus 34, 5656 AE Eindhoven, The Netherlands
joyce.westerink@philips.com

Marleen H. Schut, Kees Tuinenbreijer
Philips Consumer Lifestyle Advanced Technology, High Tech Campus 37, 5656 AE Eindhoven, The Netherlands
{*marleen.schut,kees.tuinenbreijer*}*@philips.com*

Keywords: Emotion, BioSignals, Man-Machine Interface, Automatic classification.

Abstract: As is known for centuries, humans exhibit an electrical profile. This profile is altered through various physiological processes, which can be measured through biosignals; e.g., electromyography (EMG) and electrodermal activity (EDA). These biosignals can reveal our emotions and, as such, can serve as an advanced man-machine interface (MMI) for empathic consumer products. However, such an MMI requires the correct classification of biosignals to emotion classes. This paper explores the use of EDA and three facial EMG signals to determine neutral, positive, negative, and mixed emotions, using recordings of 24 people. A range of techniques is tested, which resulted in a generic framework for automated emotion classification with up to 61.31% correct classification of the four emotion classes, without the need of personal profiles. Among various other directives for future research, the results emphasize the need for both personalized biosignal-profiles and the recording of multiple biosignals in parallel.

That men are machines (whatever else they may be) has long been suspected; but not till our generation have men fairly felt in concrete just what wonderful psycho-neuro-physical mechanisms they are.

William James (1842 – 1910)

1 INTRODUCTION

Despite the early work of William James and others, it still took until the last two decades before emotions were widely acknowledged and embraced by engineering. But, now it is generally accepted that emotions cannot be ignored; they influence us, with or without being aware, in all possible ways (Picard, 1997). Let us briefly denote three issues on how emotions influence our lives: 1) our long term physical well-being; e.g., Repetitive Strain Injury (RSI) (van Tulder et al., 2007), cardiovascular issues (Schuler and O'Brien, 1997; Frederickson et al., 2000), and our immune system (Ader et al., 1995; Solomon et al., 1974); 2) our physiological reactions/signals (Fairclough, 2009; Picard et al., 2001; van den Broek et al., 2009); and 3) our cognitive processes; e.g., perceiving, memory, reasoning (Critchley et al., 2000).

As is illustrated by the previous three issues, we are (indeed) "psycho-neuro-physical mechanisms" (James, 1893; Marwitz and Stemmler, 1998), who both send and perceive biosignals; e.g., electromyography (EMG), electrocardiography (ECG), and electrodermal activity (EDA). These biosignals can reveal a broad plethora of people's characteristics; e.g., workload, attention, and emotions. In this paper,

we will focus on biosignals that reveal people's emotional state. Such biosignals can act as a very useful interface between man and machine; e.g., computers or consumer products such as an MP3-player. Such an advanced Man-Machine Interface (MMI) would provide machines with empathic characteristics, capable of coping with the denoted issues.

For research on biosignals as an advanced MMI, traditional emotion research using interviews, questionnaires, and expert opinions are not sufficient (Fairclough, 2009; van den Broek et al., 2009). The recent progress in brain imaging techniques enables the inspection of brain activity while experiencing emotions; e.g., EEG and fMRI (Critchley et al., 2000; Grandjean and Scherer, 2008). The former research methods have as disadvantages that the measurements tend to be subjective, are very limited in explaining, and do not allow real time measurements: they can only be used before or after emotions are experienced. Although EEG techniques are slowly brought to practice; e.g., Brain Computer Interfacing (Bimber, 2008), these techniques are still very obtrusive. Hence, they are not usable in real world situations; e.g., for the integration in consumer products. As sort of a way between these two research methods, psychophysiological (or bio)signals can be used (Fairclough, 2009; Marwitz and Stemmler, 1998; van den Broek et al., 2009). These are not, or at least less, obtrusive, can be recorded and processed real time, are rich sources of information, and are relatively cheap to apply.

The traditional methods (e.g., questionnaires), brain imaging techniques, and biosignal measures used to infer people's emotional state, all share one thing: the problem of a lack of ground truth; i.e., a theoretically grounded, observable, operational definition of the construct(s) of interest (Fairclough, 2009; van den Broek et al., 2009). In addition, a range of other prerequisites should be taken into account when using such methods. In van den Broek et al. (2009), these are denoted for affective signal processing; however, most of them also hold for brain imaging techniques and traditional methods. The prerequisites include: 1) the validity of the research employed, 2) triangulation, 3) omitting the inference of emotion from the signals, if possible, and 4) inclusion and exploitation of signal processing knowledge. For a discussion on these topics, we refer to van den Broek et al. (2009). Let us now assume that all prerequisites are satisfied. Then, it is feasible to classify the biosignals in terms of emotions. In bringing biosignals-based emotion recognition to products, self-calibrating, automatic classification is essential to make it useful for Artificial Intelligence (AI) (Picard, 1997; Minsky, 2006), Ambient Intelligence (AmI) (Aarts, 2004), and MMI (Fairclough, 2009; Kim and André, 2008).

In the pursuit toward empathic technology for AI, AmI, and MMI purposes, we will discuss the work on classifying four biosignals signals: three facial EMGs and EDA. The research in which the data was gathered is discussed in both (van den Broek et al., 2006) and (Westerink et al., 2008). Therefore, we will now only provide a brief summary of it in the next section. After that, in Sections 3 and 4, we will briefly introduce the classification and preprocessing techniques employed. This is followed by Section 5 in which the classification results are presented. In Section 6, we reflect on our work, critically review it, and draw some final conclusions.

2 RECORDING EMOTIONS

An experiment was conducted in which the subjects' emotions were elicited using film fragments that are known to be powerful in eliciting emotions in laboratory settings; see also (Rottenberg et al., 2007). The physiological signals used, facial EMG and EDA, are commonly known to reflect emotions (Kreibig et al., 2007).

2.1 Participants

In the experiment, 24 subjects (20 females) participated (average age 43 years). The relative small number of males is due to an expected better facial emotion expression of females (Kring and Gordon, 1998).

2.2 Equipment and Materials

We selected 8 film fragments (duration: 120 sec. each) for their emotional content. For specifications of these film fragments, see (van den Broek et al., 2006; Westerink et al., 2008). The film fragments were categorized as being neutral or triggering positive, negative, or mixed (i.e., simultaneous negative and positive; (Carrera and Oceja, 2007)) emotions. This categorization was founded on Russell's valence-arousal model (Russell, 1980).

A TMS International Porti5-16/ASD system was used for the biosignals recordings, which was connected to a PC with TMS Portilab software (`http://www.tmsi.com/`). Three facial EMGs were recorded: the right corrugator supercilii, the left zygomaticus major, and the left frontalis muscle. The EMG signals were high-pass filtered at 20 Hz, rectified by taking the absolute difference of the two electrodes, and

average filtered with a time constant of 0.2 sec. The EDA was recorded using two active skin conductivity electrodes and average filtering with a time constant of about 2 sec.

2.3 Procedure

After the subject was seated, the electrodes were attached and the recording equipment was checked. The 8 film fragments were presented to the subject in pseudo-random order. A plain blue screen was shown between the fragments for 120 seconds; so, the biosignals returned to their baseline level for the next stimulus.

After the viewing session, the electrodes were removed. Next, the subjects answered a few questions regarding the film fragments viewed. To jog their memory, representative print-outs of each fragment were provided.

3 CLASSIFICATION TECHNIQUES

In this section, we briefly introduce the techniques used for those readers who are not familiar with (all of) them. First, ANalysis Of VAriance (ANOVA) and Principal Component Analysis (PCA) are briefly introduced, which will be both applied for preprocessing purposes. Second, the three classification techniques k-Nearest Neighbors (k-NN), Support Vector Machines (SVM), and Neural Networks (NN) are introduced. Third and last, the Leave-one-out cross validation (LOOCV) technique is introduced, which is used for the evaluation of the classifiers.

3.1 ANalysis Of VAriance (ANOVA)

ANalysis Of VAriance (ANOVA) is a statistical test to determine whether or not there is a significant difference between means of several populations. We will sketch the main idea here. For a more detailed explanation, see for example (King and Minium, 2007).

ANOVA assumes that the data of each population is independent and randomly chosen from a normal distribution. Moreover, it assumes that all the populations have the same variance. These assumptions usually hold with empirical data and the test is fairly robust against limited violations.

ANOVA examines the variance of population means compared to within class variance of the populations themselves. The result of the test is the probability p that all the populations were chosen from distributions with the same mean and variance. Hence, the smaller p, the higher the chance that there is a real difference between the populations.

3.2 Principal Component Analysis (PCA)

This linear transformation derives from an input data space a first base vector in the direction of the biggest variance in the data. Every next base vector is independent from the previous ones and represents the highest possible variance of the data with the independence constraint; see also (Everitt and Dunn, 2001).

Formally, if we have a data set $\left\{\vec{x^s}\right\}_{s\in\mathbf{Subj}}$ of n-dimensional vectors then the principal components of vector $\vec{x^s}$ from the data set are a sequence of components of vector $\vec{y^s}$ that are linear combinations of the components of vectors $\vec{x^s}$,

$$y_i^s = a_{i1}x_1^s + a_{i2}x_2^s + \cdots + a_{in}x_n^s = \vec{a_i}'\vec{x^s}$$

such that

$$\begin{array}{lll} \forall i \in \mathbb{N} \quad 1 \leq i \leq n & : & \vec{a_i}'\vec{a_i} = 1 \\ \forall i,j \in \mathbb{N} \quad 1 \leq i < j \leq n & : & \vec{a_j}'\vec{a_i} = 0 \end{array}$$

and subsequently, each $y_i = \left\{\vec{y_i^s}\right\}_{s\in\mathbf{Subj}}$ has the maximal possible variance with respect to the constraints. Variance covered by y_i is defined as

$$Var(y_i) = \vec{a_1}'\mathbf{S}\vec{a_1}$$

where $\mathbf{S}$ is the covariance matrix of the original data set.

At this point, we have to find vectors $\vec{a_i}$ that maximize the variance with respect to the constraints. This kind of optimization problems can be solved using the method of Lagrange multipliers.

In this case, the result is that $\vec{a_i}$ is the eigenvector of $\mathbf{S}$ corresponding to the i-th largest eigenvalue. Once we have the vectors $\vec{a_i}$, we can perform the transformation by mapping all the data vectors to its principal components.

$$\vec{y^s} = \begin{pmatrix} \vec{a_1}' \\ \vdots \\ \vec{a_n}' \end{pmatrix} \vec{x^s}$$

The principal components computed this way are very sensitive to scaling. In order to deal with the different scaling and capture the underlying structure of the data set, the components can be derived from the correlation matrix instead of the covariance matrix. It is equivalent to extracting the principal components in the described way after normalizing all the components of the original data set to have unit variance.

PCA is also a powerful tool for data inspection through visualization. For this purpose, often plots

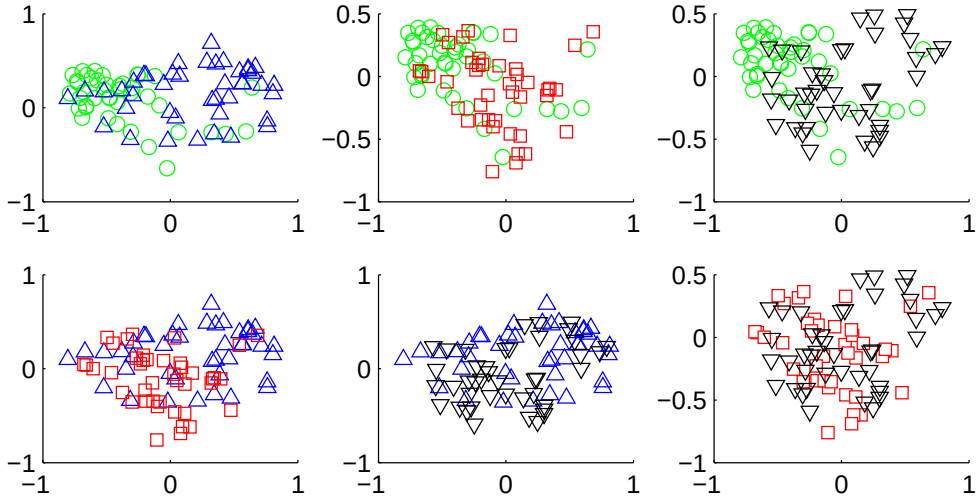

Figure 1: Visualization of the first two principle components of all six possible combinations of two emotion classes. The emotion classes are plotted per two to facilitate the visual inspection. The plots illustrate how difficult it is to separate even two emotion classes, where separating four emotion classes is the aim.

are made with the principle components on the axis. Figure 1 presents such a visualization. It presents for each set of two emotion classes, of the total of four, a plot denoting the first two principle components. The six resulting plots illustrate the complexity of separating the emotion classes perfectly.

3.3 k-Nearest Neighbors (k-NN)

We have decided to use this technique because it is a very intuitive and simple machine learning algorithm. The main idea is that each new feature vector, which is "close" to some of the vectors from the training set, will probably belong to the same class as most of these vectors. The training phase of the classifier is simply storing all (or a suitable subset) of the training samples with the correct classification category in a database.

A metric (e.g., Euclidean distance) is selected that assigns a non-negative real number to each pair of input vectors. The number represents how close the input vectors are to each other. When a new vector has to be classified, the metric is used to count the distance of the new sample from all the samples in the database. After this, the number of representatives of each class among the closest k samples are considered. If there is a class with a higher number of representatives than all the other classes; then, the new sample is classified to this class. If there is a tie of two or more classes; then, the sample is classified randomly to one of these classes.

k-NN is often applied. Consequently, various tutorials and introductions have been written. We refer to (Bishop, 2006), who provides an excellent introduction.

3.4 Support Vector Machine (SVM)

A Support Vector Machine (SVM) ensures the optimal division of a set of data to two classes with respect to the shape of the classifier and misclassification of the training samples. Using a suitable kernel function, it can create an optimally shaped classifier.

The main ideas of this classifier can be best explained through the example of a binary linear classifier; i.e., a separating hyperplane $\vec{w}.\vec{x}+b$, formally:

$$y_i(\vec{w}.\vec{x}_i+b) \geq 1-\xi_i, \text{ for } i=1,2,\ldots,N$$

where $\vec{x}_i$ are the data samples, $y_i \in \{-1,+1\}$ is the corresponding class of the i-th data sample, $\vec{w}$ is the normal vector of the hyperplane and b is the shift of the hyperplane. To make the plane optimal, the size of $\vec{w}$ and the sum of ξ_i must be minimized. It can be proved that minimization of these parameters can be solved by maximization of:

$$W(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j (\vec{x}_i.\vec{x}_j)$$

with constraints

$$0 \leq \alpha_i \leq C, \text{ for } i=1,\ldots,N$$

and

$$\sum_{i=1}^{N} \alpha_i y_i = 0.$$

where C is a constant determining the trade-off between minimizing of the size of $\vec{w}$ and the sum of ξ_i. This is a problem that can be solved using methods of quadratic programming.

After we have the Lagrange multipliers α_i, the classification is already easy:

$$f(\vec{x}) = \text{sgn}\left(\sum_{i=1}^{N} y_i \alpha_i . (\vec{x}.\vec{x}_i) + b\right)$$

It is possible to see from the derivation of this method that most of the α_is are usually equal to 0. The remaining relevant subset of the training data ($\vec{x}_i$) is called support vectors.

For a non-linear classification problem, we can transform the input space with an appropriate non-linear function Φ into a higher-dimensional feature space. For this, we only need the dot product of the transformed vectors. It is the only part of the computation where we have to work with the higher-dimensional space.

$$k(\vec{x},\vec{y}) = \left(\Phi(\vec{x}).\Phi(\vec{y})\right),$$

which results in a scalar. The function k is called the kernel function. Please note that the SVMs introduced above classify samples into two classes. However, usually we want to distinguish between multiple classes. This can be done using a separate binary classifier for each target class.

For more information on SVM, we refer to (Burges, 1998). This paper provides a gentle introduction on SVM for pattern recognition. Alternatively, (Bishop, 2006) and (Vapnik, 1999) can be consulted.

3.5 Neural Network (NN)

Neural Networks (NN) are the least intuitive group of approaches used. It has a solid theoretical basis; e.g., (Bishop, 2006). Nevertheless, their performance is not always satisfying in practice. Each neuron in a NN can perform only a trivial task, but after connecting more of them to a network, they can approximate any function (Leshno et al., 1993).

For our task, we will use the multilayer perceptron. Its perceptrons count weighed w_i sum of values on its inputs (x_i), subtracts a bias (b) and apply a sigmoid-shaped function (σ) to the result, producing single number as the output of the neuron.

$$y = \sigma\left(\sum_{i=1}^{n} w_i x_i - b\right)$$

Its neurons are divided into several layers. Inputs of one layer are all the outputs of the neurons in the previous layer. Its transfer function is crucial for the NN's functioning. For this, often gradient descent is used, which tries to minimize an overall error of the network expressed by the error function:

$$E(\vec{w}) = \sum_{k\in X}\sum_{j\in Y}\left(y_j(\vec{w},\vec{x}_k) - d_{kj}\right)^2$$

where X is the set of indexes of training samples, Y the set of output neurons, $y_j(\vec{w},\vec{x}_k)$ is the output of the j-th output neuron for input $\vec{x}_k$ and weight vector $\vec{w}$, and d_{kj} is the desired output on the j-th neuron for the k-th training sample.

Using theorems about derivation of a composite function, it is possible to derive the gradient of the error function. Following this vector with subsequent small adaptations of the weights, the algorithm will find a minimum of the error function. However, the risk remains that the error function reaches a local minimum and the learning process stops even thou the error is far from the global optimum. Although various methods exist that improve the gradient descend, none of them guarantees an optimal solution.

The strongest point of using neural network for our problem is its natural capability of incremental learning. Most of the algorithms used for neural network learning are intrinsically incremental.

For more information on NN, we refer to (Bishop, 2006). However, various other introductions have been published that differ with respect to both the provided details and their length.

3.6 Leave-one-out Cross Validation (LOOCV)

Leave-one-out Cross Validation (LOOCV) is a method to determine how good a classifier is. If we have some inputs with known correct classifications, for each of the data samples we:

- Make the classifier learn from the data without the selected sample.
- Classify the omitted sample.
- Compute the ratio of wrong classifications.

A little modification of this method in our case is that we do not leave out only one data sample, but we do not consider all data from one subject in the learning process. It is more accurate estimation of the classification error on an unknown subject.

The results reported in this paper are determined by this method if it is not specified another way. For more information on LOOCV, we refer to (Bishop, 2006).

Table 1: The best feature subsets for k-Nearest Neighbor (k-NN) classifier determined by ANalysis Of VAriance (ANOVA), using normalization per signal per participant.

		electrodermal activity (EDA)	facial electromyography (EMG)		
			Frontalis	Corrugator	Zygomaticus
Feature	Mean				o
	Absolute Deviation				o
	Standard Deviation		o		o
	Variance		o		o
	Skewness	o		o	o
	Kurtosis			o	

4 PREPROCESSING

The quest towards self-calibrating algorithms for consumer products and for AmI and AI purposes gave some constraints to processing the signals. For example, no advanced filters should be needed and the algorithms should be able to handle noisy and preferably also corrupt data. Therefore, we chose to refrain from advanced preprocessing schemes and only apply some basic preprocessing.

4.1 Normalization

Humans are known for their rich variety in all aspects, this is no different for their emotional reactions and their physiological derivatives. In developing generic classifiers, this required the normalization of the signals. This could boost its performance significantly (Rani et al., 2006).

For each person, for all his signals, and for all their features separately, two linear normalizations were applied:

$$x_n = \frac{x - min}{max - min},$$

where x_n is the normalized value, x the recorded value, and *max* and *min* the global maximum and minimum and

$$x_n^* = \frac{x - \bar{x}}{\sigma},$$

where x_n^* is the normalized value, x the recorded value, and $\bar{x}$ and σ the global mean and standard deviation.

The linear normalization of datasets (e.g., signals) has been broadly discussed, this resulted in a variety of normalization functions; e.g., see (Boucsein, 1992; Iglewicz, 1983).

4.2 Baseline Matrix

In their standard work, Picard et al. (2001) introduce a baseline matrix for processing biosignals for emotion recognition. This could tackle problems due to variation both within (e.g., inter day differences) and between participants. Regrettably, Picard et al. (2001) do not provide evidence for its working. However, their idea is appealing and, hence, was judged as worth trying.

The baseline matrix requires biosignals recordings while people are in a neutral state. Such recordings were, however, not available. Alternatively, one of the two neutral film fragments was chosen (van den Broek et al., 2006; Westerink et al., 2008).

In line with Picard et al. (2001), the input data was augmented with the baseline values of the same dataset. The results of some initial tests, using various weights, were far from convincing. A maximum performance improvement was achieved of 1.5%, using a kNN classifier. Therefore, the baseline matrix was excluded in the final processing pipeline.

4.3 Feature Selection

To achieve good classification results, the set of input features is crucial. This is no different with classifying emotions (Fairclough, 2009; van den Broek et al., 2009). To define an optimal set, a criterion function should be defined. However, no such criterion function is available in our case. Then, an exhaustive search in all possible subsets of input features is required to guarantee an optimal set (Cover and Campenhout, 1977). To limit this enormous search space, an ANOVA-based heuristic search was applied.

For both the normalizations, we performed feature-selection based on ANOVAs. We selected the features with ANOVA p values below 0.0013, as this led to the best precision. The features selected are in Table 1.

The last step of preprocessing is PCA. The improvement of the PCA is not that big compared to feature selection solely; but, it is positive for both normalizations; see also Table 2. Figure 1 presents for each set of two emotion classes, of the total of four, a plot denoting the first two principle components. As

such, the six resulting plots illustrate the complexity of separating the emotion classes.

5 CLASSIFICATION RESULTS

This section denotes the results achieved with the three classification techniques applied: k-Nearest Neighbors (k-NN), Support Vector Machines (SVM), and Neural Networks (NN). In all cases, the features extracted from the biosignals were used to classify participants' neutral, positive, negative, or mixed state of emotion.

5.1 k-Nearest Neighbors (k-NN)

For our experiments, we have used MATLAB[1] and k-NN implementation based on SOM Toolbox 2.0 (Vesanto et al., 2000). Besides the classification algorithm described in Section 3.3, we have used a modified version, more suitable for calculating the recognition rates. The output of the modified version is not the resulting class, but a probability of classification to each of the classes. This means that if there is a single winning class; then, the output is 100% for the winning class and 0% for all the other classes. If there is a tie of two classes then the result is 50% for each of them and 0% for the rest and so forth. All the recognition rates of the k-NN classifier in the current study are obtained by this modified algorithm.

A correct metric is a crucial part of a k-NN classifier. A variety of metrics provided by the pdist function in MATLAB was applied. Different feature subsets appeared to be optimal for different classes. Rani et al. (2006) denoted the same issue in their empirical review. If we use the feature subset optimized on the objective classes; then, the recognition precisions in other divisions lowers or improves only a little compared to the improvement in the optimized class division. The results of the best preprocessed input with respect to the four emotion classes (i.e., neutral, positive, negative, and mixed) is 61.31%, with a cityblock metric and $k = 8$.

Probability tables for the different classifications given a known emotion category are quite easy to obtain. They can be estimated from *confusion matrices* of the classifiers by transforming the frequencies to probabilities. Table 3 presents the confusion matrix of the classifiers used in this research.

[1] http://www.mathworks.com/products/matlab/

5.2 Support Vector Machines (SVM)

We have used MATLAB environment and SVM-KM Toolbox (Canu et al., 2005) for experimenting with SVMs. We use input enhanced with the best preprocessing described in the previous section. It was optimized for the k-NN classifier; however, we expect it to be a good input also for more complex classifiers, including SVM. This assumption was supported by several tests with other normalizations. The inputs in this section are normalized per signal per person. After feature selection, the first 5 principal components from the PCA transformation were used.

The kernel function of SVM characterizes the shapes of possible subsets of inputs classified into one category. We applied both polynomial kernels, defined as:

$$K_{Poly}(\vec{x},\vec{y}) = (\vec{x}.\vec{y}+1)^d$$

and Gaussian kernels, defined as:

$$K_{Gaus}(\vec{x},\vec{y}) = \exp\left(-\frac{|\vec{x}-\vec{y}|^2}{2\sigma^2}\right)$$

The correct kernel function is the most important part of SVM.

A Gaussian kernel ($\sigma = 0.7$) performed best with 60.71% correct classification. However, a polynomial kernel with $d = 1$ has a similar classification performance (58.93%). All the results are slightly worse than with the k-NN classifier.

5.3 Neural Networks (NN)

We have used a modified multi-layer perceptron trained by back-propagation algorithm that is implemented in the Neural Network Toolbox of MATLAB. It uses gradient descent with moment and adaptive training parameter. We have tried to recognize only the inputs that performed best with the k-NN classifier.

In order to assess what topology of NN is most suitable for the task, we ran a small test. In the experiments with one hidden layer, we have tried 2 to 16 neurons and we run LOOVC for each network 100 times. The networks were trained with the fixed number of 150 cycles and subsequently tested on the left-out subject. The experiments with two hidden layers were much slower; so, we made only 10 trials for each combination of sizes of the layers. We have 12×12 different topologies, 10 trials for each of them and one trial of LOOCV means to train and test 21 networks (total: 30240 networks). Each network was trained with 150 cycles.

Our experiments with different network topologies supported the claim from (Lawrence et al., 1996)

Table 2: The recognition precision of the k-Nearest Neighbor (k-NN) classifier, with and without ANalysis Of VAriance (ANOVA) feature selection (FS) and with and without Principle Component Analysis (PCA) transform. *# comp.* denotes the number of principal components used to reach the precision with FS.

Normalization	no FS	ANOVA FS	# comp.	ANOVA FS & PCA
no	45.54%			
yes	54.07%	60.71%	5	60.80%

Table 3: *Confusion matrix* of the k-NN based classifier of EDA and EMG signals for the best reported input preprocessing.

		Real			
		Neutral	Positive	Mixed	Negative
Classified	Neutral	71.43%	19.05%	9.52%	14.29%
	Positive	9.52%	57.14%	9.52%	21.43%
	Mixed	4.76%	4.76%	64.29%	11.90%
	Negative	14.29%	19.05%	16.67%	52.38%

that bigger NN does not always tend to over fit the data and the extra neurons are not used in the training process. Bigger networks showed good generalization capabilities. However, further enlargement of the network did not lead to better results. For this reason, we choose the topology with one hidden layer of 12 neurons.

An alternative method for stopping the adaptation of the NN is using validation data. The data set is split into three parts. In our case, we have used one subject for testing, three subjects for validation and seventeen subjects for training. The testing subject is completely removed from the training process at the beginning. After that, the network is trained using seventeen randomly chosen training subjects. At the end of each training iteration, the current network is tested on the three validation subjects.

In order to evaluate the NN on different desired outputs, we have performed the above described algorithm for each subject as the testing subject and for 15 random triples of the remaining 23 subjects as the validation data. The 15 networks trained on different data create an ensemble and the final result of the classifier is the most frequent output class. This way, we can profit from the early stopping of the back-propagation algorithm and still use all the training samples for training of the whole classifier. This procedure led to a 56.19% correct classification of the four emotion classes.

6 DISCUSSION & CONCLUSIONS

Successful automatic classification of biosignals could serve various purposes (Fairclough, 2009; van den Broek et al., 2009). One of them is to extend consumer products, AI, and AmI with empathic capabilities.

Throughout the last decade, various studies have been presented with similar aims, reporting good results on the automatic classification of biosignals. For example, Picard et al. (2001) reports 81% correct classification on the emotions of one subject. More recently, Kim and André (2008) reported a recognition accuracy of 95% and 70% for subject-dependent and subject-independent classification. Their study included three subjects.

In comparison with Picard et al. (2001) and Kim and André (2008), this research incorporated data of a large number (i.e., 24) of people, with the aim to develop a generic processing framework. At first glance, with a recognition accuracy of 61.31%, its success is questionable. However, taking in consideration the generic processing pipeline, it should be judged as (at least) reasonably good. Moreover, a broad range of improvements are possible. One of them would be to incorporate more biosignals in the processing framework. Another directive could be to question the need of identifying specific emotions, using biosignals for MMI. Hence, the use of alternative, rather rough categorizations, as used in the current research, should be further explored.

Also in this research, the differences among participants became apparent. They can be denoted on two levels: physiological and psychological. With this we mean that people have different physiological reactions on the same emotions and that people experience different emotions with the same stimuli (e.g., music or films). Moreover, these two levels interact (Fairclough, 2009; Marwitz and Stemmler, 1998). Although our aim was to develop a generic model, it seems to be questionable whether or not this can be

realized. Various attempts have been made to determine people's personal biosignals-profile; e.g., (Kim and André, 2008; Marwitz and Stemmler, 1998; Picard et al., 2001; Rani et al., 2006). However, no generally accepted standard has been developed so far.

With respect to processing the biosignals, the current research can be extended by a more detailed exploration of the time windows; e.g., with a span of 10 seconds (Fairclough, 2009; van den Broek et al., 2009). Then, data from different time frames can be combined and different normalizations can be better applied to create some new features that could easier reveal emotions. For example, the information concerning the behavior of the physiological signals could be more informative than only the integral features from the larger time window. Studying the short time frames also provides a better understanding on the relation between emotions and their physiological correlates.

Preprocessing of the biosignals could also be improved. First of all, we think that the feature selection based on an ANOVA is not sufficient for more complex classifiers such as Neural Networks. The ANOVA tests gather the centers of random distributions that would generate the data of different categories; hereby assuming that their variances are the same. However, a negative result of this test is not enough to decide that the feature does not contain any information. As an alternative for feature selection, the k-NN classifier can be extended by a metric that would weigh the features, instead of omitting the confusing or less informative features.

As can be derived from the discussion, various hurdles have to be taken in the development of a generic, self-calibrating, biosignal-driven classification framework for MMI purposes. The research and the directives denoted in this article could help in taking the first hurdles. When the remaining ones will also be taken; then, in time, the common denominators of people's biosignals can be determined and their relation with experienced emotions can be further specified. This would mark a new, biosignal-driven, era of advanced MMI.

ACKNOWLEDGEMENTS

The authors thank Frans van der Sluis (University of Twente, NL) and Joris H. Janssen (Eindhoven University of Technology, NL / Philips Research, NL) for reviewing earlier drafts of this article.

REFERENCES

Aarts, E. (2004). Ambient intelligence: Vision of our future. *IEEE Multimedia*, 11(1):12–19.

Ader, R., Cohen, N., and Felten, D. (1995). Psychoneuroimmunology: interactions between the nervous system and the immune system. *The Lancet*, 345(8942):99–103.

Bimber, O. (2008). Brain-Computer Interfaces. *IEEE Computer*, 41(10):[special issue].

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer.

Boucsein, W. (1992). *Electrodermal activity*. New York, NY, USA: Plenum Press.

Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167.

Canu, S., Grandvalet, Y., Guigue, V., and Rakotomamonjy, A. (2005). SVM and kernel methods Matlab Toolbox. Perception Systèmes et Information, INSA de Rouen, Rouen, France. URL: `http://asi.insa-rouen.fr/enseignants/~arakotom/toolbox/`.

Carrera, P. and Oceja, L. (2007). Drawing mixed emotions: Sequential or simultaneous experiences? *Cognition & Emotion*, 21(2):422–441.

Cover, T. M. and Campenhout, J. M. V. (1977). On the possible orderings in the measurement selection problem. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-7(9):657–661.

Critchley, H. D., Elliott, R., Mathias, C. J., and Dolan, R. J. (2000). Neural activity relating to generation and representation of galvanic skin conductance responses: A functional magnetic resonance imaging study. *The Journal of Neuroscience*, 20(8):3033–3040.

Everitt, B. S. and Dunn, G. (2001). *Applied Multivariate Data Analysis, Second Edition*. Arnold, London.

Fairclough, S. (2009). Fundamentals of physiological computing. *Interacting with Computers*, [in press].

Frederickson, B. L., Manusco, R. A., Branigan, C., and Tugade, M. M. (2000). The undoing effect of positive emotions. *Motivation and Emotion*, 24(4):237–257.

Grandjean, D. and Scherer, K. R. (2008). Unpacking the cognitive architecture of emotion processes. *Emotion*, 8(3):341–351.

Iglewicz, B. (1983). *Robust scale estimators and confidence intervals for location*, chapter 12, pages 404–432. New York, NY, USA: John Wiley & Sons, Inc.

James, W. (1893). Review: La pathologie des emotions by Ch. Féré. *The Philosophical Review*, 2(3):333–336.

Kim, J. and André, E. (2008). Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 30(12):2067–2083.

King, B. M. and Minium, E. W. (2007). *Statistical reasoning in psychology and education*. New York, NY, USA: John Wiley & Sons, Inc., 5th edition.

Kreibig, S. D., Wilhelm, F. H., Roth, W. T., and Gross, J. J. (2007). Cardiovascular, electrodermal, and respiratory response patterns to fear- and sadness-inducing films. *Psychophysiology*, 44(5):787–806.

Kring, A. M. and Gordon, A. H. (1998). Sex differences in emotion: Expression, experience, and physiology. *Journal of Personality and Social Psychology*, 74(3):686–703.

Lawrence, S., Giles, C. L., and Tsoi, A. (1996). What size neural network gives optimal generalization? Convergence properties of backpropagation. Technical Report UMIACS-TR-96-22 and CS-TR-3617.

Leshno, M., Lin, V. Y., Pinkus, A., and Schocken, S. (1993). Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, 6(6):861–867.

Marwitz, M. and Stemmler, G. (1998). On the status of individual response specificity. *Psychophysiology*, 35(1):1–15.

Minsky, M. (2006). *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. New York, NY, USA: Simon & Schuster.

Picard, R. W. (1997). *Affective Computing*. Boston MA, USA: MIT Press.

Picard, R. W., Vyzas, E., and Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10):1175–1191.

Rani, P., Liu, C., Sarkar, N., and Vanman, E. (2006). An empirical study of machine learning techniques for affect recognition in human-robot interaction. *Pattern Analysis & Applications*, 9(1):58–69.

Rottenberg, J., Ray, R. R., and Gross, J. J. (2007). *Emotion elicitation using films*, chapter 1, pages 9–28. New York, NY, USA: Oxford University Press.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178.

Schuler, J. L. H. and O'Brien, W. H. (1997). Cardiovascular recovery from stress and hypertension factors: A meta-analytic view. *Psychophysiology*, 34:649–659.

Solomon, G. F., Amkraut, A. A., and Kasper, P. (1974). Immunity, emotions and stress with special reference to the mechanisms of stress effects on the immune system. *Psychotherapy and Psychosomatics*, 23(1–6):209–217.

van den Broek, E. L., Janssen, J. H., Westerink, J. H. D. M., and Healey, J. A. (2009). Prerequisits for Affective Signal Processing (ASP). In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing*, page [in press], Porto – Portugal.

van den Broek, E. L., Schut, M. H., Westerink, J. H. D. M., van Herk, J., and Tuinenbreijer, K. (2006). Computing emotion awareness through facial electromyography. *Lecture Notes in Computer Science (Human-Computer Interaction)*, 3979:51–62.

van Tulder, M., Malmivaara, A., and Koes, B. (2007). Repetitive strain injury. *The Lancet*, 369(9575):1815–1822.

Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10(5):988–999.

Vesanto, J., Himberg, J., Alhoniemi, E., and Parhankangas, J. (2000). SOM toolbox for matlab. Technical Report A57, Helsinki University of Technology. URL: `http://www.cis.hut.fi/projects/somtoolbox/`.

Westerink, J. H. D. M., van den Broek, E. L., Schut, M. H., van Herk, J., and Tuinenbreijer, K. (2008). *Computing emotion awareness through galvanic skin response and facial electromyography*, volume 8 of *Philips Research Book Series*, chapter 14 (Part II: Probing in order to Feed Back), pages 137–150. Springer: Dordrecht, The Netherlands.

BRIEF BIOGRAPHY

Egon L. van den Broek obtained his MSc (2001) in Artificial Intelligence and his PhD (2005) in Content-Based Image Retrieval (CBIR), both from the Radboud University (RU), Nijmegen, The Netherlands (NL). Previously, he has been junior lecturer (RU), consultant, and assistant professor in Artificial Intelligence (AI) at the Vrije Universiteit (VU), Amsterdam, NL. Currently, he is head of a group on Advanced Interface Design (Center for Telematics and Information Technology, University of Twente, Enschede, NL), coordinates a MSc track, is member of the board of the post-doctoral professional study of ergonomics (VU), is consultant for Philips Research, and is visiting assistant professor in Artificial Intelligence (RU). He is involved in various national and EU projects and is specialized in engineering cognition, affective signal processing, cognitive computer vision, and perception. He has supervised 40+ BSc, MSc, and PhD students and published 100+ articles and book chapters, holds a patent, and developed the online image retrieval system http://www.m4art.org.