Feature extraction for on-line EEG classification using principal components and linear discriminants

K. Lugger[1] D. Flotzinger[1] A. Schlögl[1] M. Pregenzer[2]
G Pfurtscheller[1,2]

[1]Ludwig Boltzmann-Institute for Medical Informatics & Neuroinformatics, Graz, Austria
[2]Institute for Biomedical Engineering, Department of Medical Informatics, University of Technology, Brockmanngasse 41, A-8010 Graz, Austria

Abstract—*The study focuses on the problems of dimensionality reduction by means of principal component analysis (PCA) in the context of single-trial EEG data classification (i.e. discriminating between imagined left- and right-hand movement). The principal components with the highest variance, however, do not necessarily carry the greatest information to enable a discrimination between classes. An EEG data set is presented where principal components with high variance cannot be used for discrimination. In addition, a method based on linear discriminant analysis (LDA), is introduced that detects principal components which can be used for discrimination, leading to data sets of reduced dimensionality but similar classification accuracy.*

Keywords—*Principal component analysis, Feature extraction, Linear discriminant, EEG classification, Brain–computer interface, Learning vector quantisation, Adaptive autoregressive model*

Med. Biol. Eng. Comput., 1998, **36**, 309–314

1 Introduction

ANALYSING AND classifying data is very often linked to the problem of high dimensionality. Questions arise such as: are all attributes useful for classification or can some be omitted? Is it possible to merge the information from all the attributes into only a few? There are two ways to handle this problem: either we can try to omit those attributes that do not contain considerable information (feature selection) or we can try to combine attributes to form new ones (feature extraction). One method for feature extraction is principal component analysis (PCA).

Online analysis and classification of EEG data is the main task in an EEG-based communication system (brain–computer interface, BCI). The BCI transforms specific mental activity (thoughts) into signals which can be used by subjects with severe motor disabilities to improve their communication with the environment (PFURTSCHELLER *et al.*, 1993, 1996; WOLPAW *et al.*, 1991, WOLPAW and MCFARLAND 1994). An EEG-based BCI can use a linear threshold (MCFARLAND *et al.*, 1993) or a neural-network classifier (FLOTZINGER *et al.*, 1994; KALCHER *et al.*, 1996). The latter method was implemented in the Graz BCI that tries to separate between imagination of left- and right-hand movement (PFURTSCHELLER *et al.* 1997). In both cases there are difficulties. First, the analysis of EEG data from two or more channels results in high-dimensional data vectors including all spatiotemporal information. Secondly, the number of examples available for training is relatively small compared to their dimensionality (e.g. about 100 examples).

Correspondence should be addressed to Dr. Pfurtscheller; e-mail: pfu@dpmi.tu-graz.ac.at

First received 2 April 1997 and in final form 10 December 1997

For the Graz BCI two different data preprocessing methods are used:

(i) Calculation of band power values in pre-selected frequency bands and time windows (KALCHER *et al.*, 1996; PFURTSCHELLER *et al.*, 1996; PREGENZER *et al.*, 1996)
(ii) Estimation of adaptive autoregressive (AAR) parameters using the least-mean-square (LMS) algorithm (SCHLÖGL *et al.*, 1997)

Also two different classification methods are used:

(i) Classification of the data according to different classes (e.g. right and left motor imagery) using linear discriminant analysis (LDA)
(ii) Classification with the nonlinear learning vector quantisation (LVQ) algorithm (FLOTZINGER *et al.*, 1992; PELTORANTA and PFURTSCHELLER 1994)

The series of BCI studies is needed to shed light on whether it is possible to discriminate between different types of motor imagery EEG-patterns in real-time by classifying features derived from EEG signals recorded from the intact scalp.

The dimensionality of the data vectors extracted from the EEG data needs to be reduced because for most classification algorithms it is very difficult to reliably estimate the parameters of a classifier in high dimensions when only few training examples are available. In the literature this problem is often referred to as the curse of dimensionality' (BISHOP, 1995). There is hope that by means of dimensionality reduction using linear methods, the error rate of non-linear classifiers can be improved, or at least be kept at the same level as without reduction, using fewer attributes.

The aim of this paper is to apply PCA to two different parameter sets (AAR parameters and band power values) extracted from the same raw EEG data and to investigate the impact of the parameter reduction on the results of

classification with LDA and LVQ. Data sets from three subjects are used. The projection of data vectors onto single principal components (PCs) will be compared to LDA and similarities will be pointed out. This leads to a new method for detecting class-information in PCs.

2 Methods

2.1 *Data preprocessing*

The source for the data reported here are raw EEG data taken from the latest Graz BCI experiments where subjects had to imagine either left- or right-hand movement in response to an arrow pointing to the according direction (PFURTSCHELLER *et al.*, 1997). The sampling frequency was 128 Hz. Data from two bipolar recordings close to electrode positions C3 and C4 (international 10–20 system) were used.

In a first set of features (BAND), EEG data from two EEG channels recorded during motor imagery were bandpass filtered using two subject-specific filter bands (subject S1: 9–13 Hz and 21–26 Hz; subject S2: 10–12 Hz and 20–23 Hz; subject S3: 10–14 Hz and 18–26 Hz) and a one-second time window (for details see PFURTSCHELLER *et al.*, 1997). The bandpass filtered values for one second were then squared and averaged over 32 consecutive power samples to obtain four values per second. This resulted in a 16-dimensional vector for the BAND data (four time points × two EEG channels × two frequency bands).

In the second set of features (AAR), and EEG data from each of the two channels were approximated using an AAR model of order six. Such an AAR model tries to describe a signal Y_t in the following form: $Y_t = a_{1,t}Y_{t-1}a_{2,t}Y_{t-2} + \cdots + a_{p,t}Y_{t-p} + E_t$. Where $a_{i,t}$ denotes the ith AAR parameter at timepoint t. E_t is considered as white noise with zero mean and finite variance. The difference from an AR model is that the AAR parameters can change as time changes.

The AAR parameters were estimated using the least mean square (LMS) method and an update coefficient of 0.007 (for details see SCHLÖGL *et al.*, 1997). The (time) average of the AAR parameters over one second per trial and channel, results in a 12-dimensional vector (model order of six × two EEG channels). Details of each data set can be found in Table 1.

The linear method for combining different features used in this paper is PCA. PCA derives the direction of a set of orthogonal vectors that successively point into the direction of the highest variance of the data set. The PCs of a data set are easily calculated as the eigenvectors of the covariance matrix of the data set (BISHOP, 1995). They can also be calculated by means of the singular value decomposition of the original data (RIPLEY, 1996). The eigenvalues denote the variance that the corresponding PCs (i.e. eigenvectors) account for.

For dimensionality reduction the m ($m <$ dimension of the data) PCs with highest variance of the data (above some threshold) are often selected. This strategy might neglect PCs that contribute to a better discrimination: as shown in the example depicted in Fig. 1, the first principal component points into the direction of the highest variance (along the x-axis), but the best classification would be along the y-axis. It will be shown here on two data sets that this selection strategy is, therefore, not suitable for classification problems.

2.2 *Classification*

Two classifiers were used, namely linear discriminant analysis (LDA) and learning vector quantisation (LVQ). A linear discriminant tries to find an optimal hyperplane to separate two classes (here, imaginary left- and right-hand movement). For a detailed explanation of LDA, see elsewhere (DUDA and HART, 1973; BISHOP, 1995).

Apart from LDA, the non-linear classification algorithm LVQ is used to see whether the proposed linear dimensionality reduction would also yield comparable or even better results with a non-linear classification. LVQ tries to represent classes with a set of labelled vectors, so-called codebook vectors. Those codebook vectors are fitted to the data in an iterative traning algorithm in a training phase. This can be carried out by means of several LVQ training algorithms that differ only slightly. In our case LVQ3 (KOHONEN, 1995) was used. When the training phase is completed, for each new example the Euclidean distance to each codebook vector is calculated. The resulting class of the new example is the same as the class of the closest codebook vector.

In this paper three codebook vectors (optimised for BAND data) were used for each class. Details on LVQ can be found elsewhere (KOHONEN, 1995; FLOTZINGER *et al.*, 1992).

It should be pointed out that LDA and LVQ were used as binary classifiers (i.e. examples were classified as either left- or right-hand motor imagery). By introducing a threshold, a third 'no decision' class can easily be obtained (KALCHER *et al.*, 1996).

The discrimination information of a particular feature set can be measured indirectly through the error rate calculated by averaging three times over a three-fold cross-validation. Three-fold cross-validation divides the data set into three equally sized, disjunct partitions. First, the two partitions are used to train the classifier (LDA or LVQ) and the third one is used for testing. Thereafter, the first and the last partition are used for training and the second is used for testing. Finally, the first partition is used for testing. The resulting cross-validation error rate is the average error rate of the three runs. This is done three times and each time the data set is permutated differently, resulting in nine runs. For further details on cross-validation see elsewhere (MICHIE *et al.*, 1994).

Calculating the angle a between LDA (v_1) and PC (v_2) was done using the following formula.

$$\cos(a) = \left| \frac{v_1' \cdot v_2}{|v_1| \cdot |v_2|} \right|$$

This results in angles between 0° and 90°. Only angles in this range are interesting, as angles close to 90° mean that the LDA and PC are orthogonal to each other and thus the PC does not contribute to discrimination. Conversely, angles close to 0° imply that PCs can be used for discrimination as they define

Table 1 Size of original data sets and their error rate using LDA and LVQ

				AAR			BAND		
subject	examples left	examples right	sum of examples	dim.	error rate LDA	error rate LVQ	dim.	error rate LDA	error rate LVQ
S1	298	291	589	12	16.92%	22.69%	16	14.15%	18.00%
S2	316	317	633	12	24.96%	29.02%	16	22.06%	21.22%
S3	318	315	633	12	28.01%	29.01%	16	22.01%	24.17%

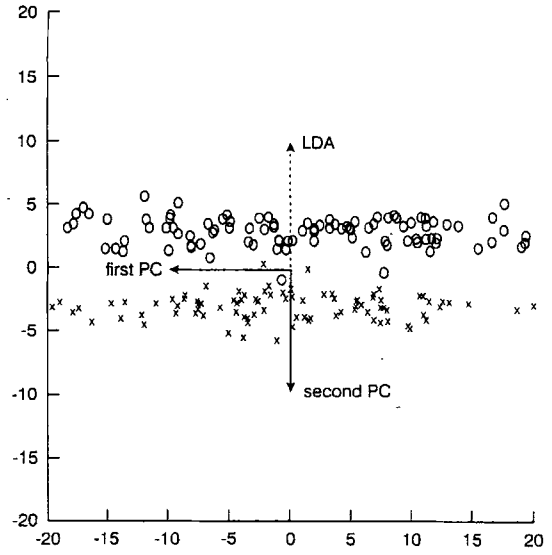

Fig. 1 *Example of two classes with corresponding LD and PCs. The angle between the LD and the first PC is about 88°, between the LD and the second PC about 2°*

essentially the same hyperplane. For illustration, an example with stimulated data is given in Fig. 1.

3 Experiments

3.1 *General*

In this paper, we will introduce a new method for feature extraction. Fig. 2 shows feature extraction as a two-stage process: after transformation of the original n attributes into n new ones (for example using PCA), feature selection is performed on these new attributes to reduce dimensionality from n to m (BISHOP, 1995; RIPLEY, 1996).

3.2 *Classification of original data*

First, the whole original data sets were classified using LDA and LVQ to see what error rate could be achieved on the originals.

3.3 *Classification of transformed data*

For the remaining experiments, the data sets were transformed into the direction of their PCs. Therefore, the first new attribute represented the highest amount of variation within the original data set, the second new attribute the second most variation etc. (see above).

Experiment 1: For each single new attribute, the error rate was calculated as described above. Additionally, the average angle between each PC and the LDs that resulted from the cross-validation of the original data set were calculated. This was done to see whether the PCA found on equally good

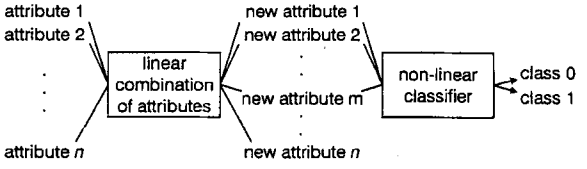

Fig. 2 *Feature extraction as a reduction from n old attributes to m new attributes by means of a linear combination of the old attributes (e.g. PCA); this allows a more reliable estimation of non-linear classifiers for the same number of training examples*

vector to discriminate between two classes: the lower the angle, the more similar the directions of two vectors. Therefore, the variance of the respective PC represents more class information.

Experiment 2: Next the data sets were classified, incrementing the number of PCs from 1 to 12 or 16, respectively. As in traditional PCA, where one selects a number of PCs according to the amount of variance in the data one intends to represent, at first only the first new attribute was used for classification, then the first two new attributes were used and so on.

Experiment 3: The last experiment was again made with the transformed (new) attributes. As before they were sorted, this time not according to decreasing explained variance but according to the increasing mean angle towards the LD.

4 Results

4.1 *Classification of original data*

The classification results of the original data sets can be found in Table 1.

4.2 *Classification of transformed data*

Experiment 1: The results on data sets from three subjects can be found in Fig. 3. Note that the y-axis denotes the error rate (percentage) as well as the angle (always below 90°) and the explained variance (percentage). It can be seen that most of the variance (nearly 100%) is explained within the first attributes, especially for the AAR parameters, while the error rate is only about 50%.

Experiment 2: The results of experiment 2 can be seen in Fig. 4. The x-axis denotes the dimensionality of the used data set. As in Fig. 3, the y-axis contains two different scales, one for the error rate and one for the accumulated variance. Note that in the Figures, a first point (zero) was inserted for all curves indicating that with zero dimensions, 0% of the variance of the data set is explained and an error rate of 50% (baseline error rate for a two-class problem where 50% of the examples belong to each class) can be achieved.

For the results in Table 2 (upper part) all the attributes were used that in sum explained at least 95% of the variance.

Experiment 3: As before, new attributes were added one by one, thus the dimensionality of the classified data set increased from zero to 12 or 16, respectively. The resulting graphs can

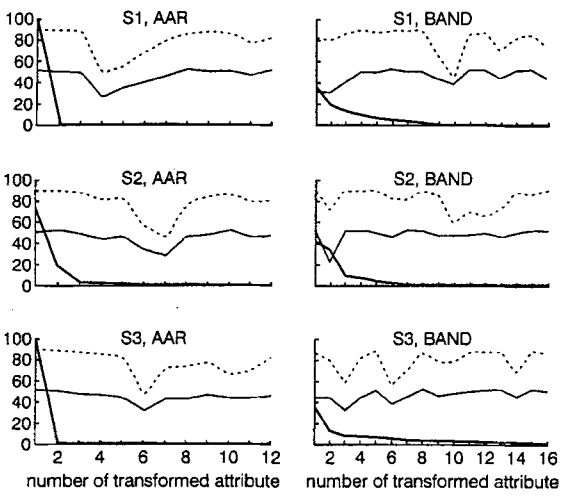

Fig. 3 *Error rate of LD (solid line, in percent), mean angle between the PCs and the LD (dashed line, in degrees) and explained variance (bold line, in percent) for each single new attribute of both data sets of subjects S1, S2 and S3*

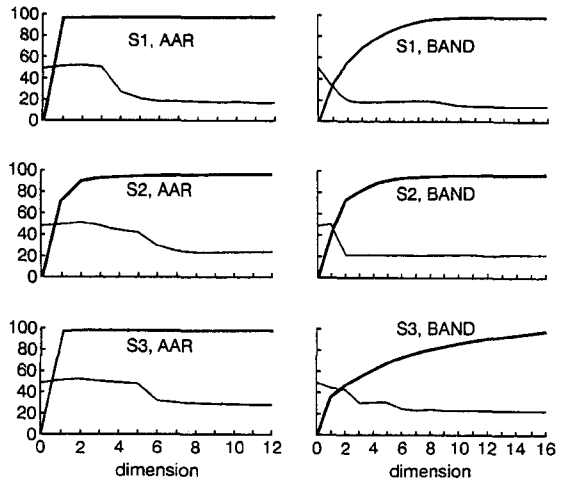

Fig. 4 *Error rate of LD (solid line, in percent) and explained variance (bold line, in percent) for the accumulated attributes (right figure) for subjects S1, S2 and S3, ordered by size of explained variance*

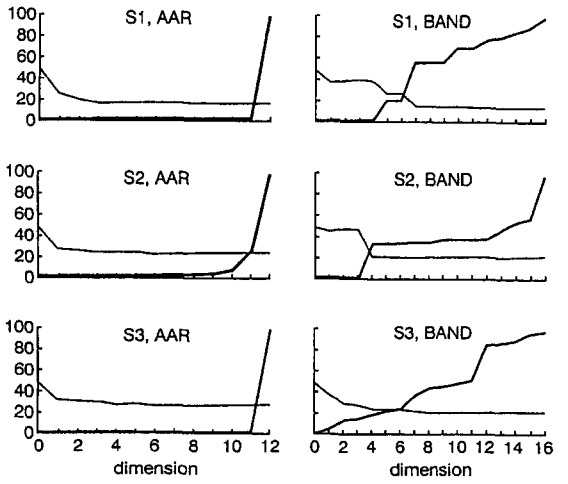

Fig. 5 *Error rate of LD (solid line, in percent) and explained variance (bold line, in percent) for the accumulated attributes (right figure) for subjects S1, S2 and S3, ordered by size of angle*

be seen in Fig. 5. As before, the y-axis contains the error rate (of LD) and the amount of accumulated variance. The x-axis depicts the dimensionality of the classified data set.

For the results in Table 2 (lower part), all those attributes were used where the average angle to the LD was smaller than 85°. Additionally, in this Table it can be seen how much of the variance is used for this discrimination accuracy and which of the new attributes were used for classification.

5 Discussion

5.1 *Experiment 1*

When feature sets are compared it can be seen that the first PC, in general, covers between 35 and 40% of the variance of the bandpass data, whereas for the autoregressive data it covers at least 74%. The high variance of the first PCs might stem from attributes which were not properly scaled, which is not the case in our data. All attributes for both data sets have about the same minimum, maximum and average value.

For the AAR data, most information is contained in components which represent only a very small part of the variation. Component 4 of subject S1, for instance, has an error rate of 27%, but is responsible only for 0.12% of the total variation. Conversely, for the BAND data, those new attributes that contain much information are responsible for a much higher part of the variation. For example, for subject S1, the new attribute 2 has an error rate of about 30% but is responsible for about 20% of the variation. Note that one cannot find a general PC index (e.g. the third or fourth PC) that is responsible for good classification results: since the last ten PCs of the AAR data are responsible for about the same amount of variation (which is also very small), their order is close to random.

It should be mentioned that for the AAR data in the angle between LD and the PCs is highly correlated with the respective error rate of each PC (correlation coefficients 0.97, 0.98 and 0.93 for subjects S1, S2 and S3, respectively), but not so strongly correlated for the BAND data, except for subject S3 (correlation coefficients 0.49, 0.41 and 0.86 for subjects S1, S2 and S3, respectively). This fact could stem from the different ways in which both data sets were preprocessed. For the BAND data, optimal time and frequency ranges were already known and the data were proprocessed accordingly, i.e. each attribute of the BAND data contains information that can be used for discrimination. The AAR data, on the other hand, were not preprocessed at all: the autoregressive parameters try to approximate the original signal as well as possible and, therefore, also model those parts of the signal that contain all discrimination information and 'noise' that distorts the data. This 'noise' might stem from the background EEG activity, artefacts (e.g. from muscles), variance of the stochastic parameter estimation method, intertrial variability, calculation inaccuracies etc.

5.2 *Experiment 2*

Traditional PCA builds on the (sometimes wrong) assumption that the first few PCs that represent, for example, 95% of the variance of the original data set already represent the data sufficiently. As can be seen in Table 2 the BAND data are classified much better than the AAR data, although about the same amount of variation is explained. In the AAR data 95%

Table 2 Classification results of the subsets, attributes chosen according to their variance (above) and according to their average angle (below)

	AAR					BAND				
Subject	attributes	dim.	variance explained	error rate LDA	error rate LVQ	attributes	dim.	variance explained	error rate LDA	error rate LVQ
S1	1	1	98.78%	52.12%	50.09%	1–7	7	95.46%	19.01%	18.68%
S2	1–3	3	95.80%	49.87%	47.76%	1–6	6	96.68%	22.12%	20.69%
S3	1	1	98.63%	52.08%	49.08%	1–13	13	95.08%	22.01%	25.17%
S3	4–7, 11–12	6	0.31%	17.65%	17.83%	1–2, 9–11, 13–16	9	57.72%	15.22%	18.90%
S2	4–9, 11–12	8	3.99%	24.49%	32.44%	2, 6–7, 9–16	11	38.60	21.91%	21.59%
S3	–12	8	0.28%	27.07%	29.75%	2–4, 6–7, 9–10, 14	8	45.13%	21.33%	25.07%

of the variation of the data yields error rates which are comparable to the baseline error rate on the whole data set. For subjects S1 and S3 even more than 98% of the variance is not able to allow discrimination between the two classes.

When the diagrams in Fig. 4 are compared for AAR and BAND data, it can be seen that the error rate for the BAND data increases abruptly while adding the first few new attributes; after the first or second PC, the error rate is always smaller than 50%. This is not the case for the AAR parameters: for the first few components the error rates stay at about 50% and then decrease sharply (e.g. while adding component 4 for subject S1, Fig. 4). This goes along with a very good discrimination result for the single transformed attribute 4 for subject S1 (Fig. 3). The same holds for subject S2 with attributes 6 and 7 (Fig. 4) and subject S3 with attribute 6 (Fig. 4). As can be seen, the method of choosing those new attributes that represent highest variance does not necessarily return those attributes that can best be used for discrimination.

5.3 *Experiment 3*

In Fig. 5, it is very interesting to see that for the AAR data, the new attribute that accounts for most of the variance is always rated last, with respect to the angle. Also, an error rate that is close to the optimal error rate (when using all attributes) is reached after only a few attributes. This indicates that for the AAR data, nearly all of the classification information is contained in only a few percent of the variance, whereas for the BAND data one needs about half of the variance.

Note that LDA results are about equal ($\pm$ 1%) to the results that were derived from the original data. This can be attributed to the linear transformation performed by PCA, followed by another linear combination of attributes by LDA. These two linear transformations can easily be merged into one without loss of performance, as long as those PCs that are chosen as the basis for classification contain most discrimination information. Note also that the non-linear LVQ scored nearly 5% better on the transformed AAR data for subject S1 (17.8% vs. 22.7%) than on the original data set. This substantiates the idea that dimensionality reduction can improve performance. It should also be pointed out that for subject S2, the AAR data classification was about 3% worse when based on the reduced data set than when based on the original data set.

The reason why LVQ seems to behave a little worse than LDA can stem from different reasons. First, LVQ parameters were optimised on 16-dimensional BAND data. Secondly, features were extracted using a linear method (PCA) but LVQ is a non-linear classification algorithm. Thus PCA might favour LDA. The last point to note is that only three subjects were used for these experiments, therefore no statistically valid comparison between the two classifiers can be given.

However, our experiments have generally shown that LVQ scores better than on BAND data, and LDA yields better results with AAR data.

It could be argued that the linear transformation stage of PCA is redundant because it merely redistributes the information content of the original attributes (i.e. it is a bijective mapping). Looking at experiments 2 and 3 on the original data it can be seen that the curve for the error rate is decreasing, but more slowly for the AAR data than for the transformed data. So for the same error rate, more attributes are needed.

6 Conclusion

The results presented in this paper indicate that the unsupervised method of extracting 'important' features (e.g. through PCA) only works with some data sets in the context of classificastion. In our case, the BAND data were reduced from 16 to 6, 7 and 13 dimensions (respectively, for subjects S1, S2 and S3) and still provided the same classification accuracy for LD and LVQ. For the AAR data, however, this method did not work at all: the dimensionality was heavily reduced (from 12 to 1 and 3 dimensions, respectively) but the classification accuracy was close to random.

By sorting the PCs according to their angle to the original LD, class information was added to each component, thus making the feature extraction procedure at least partly supervised. By using components with a smaller angle than 85° to the LD, satisfactory results were derived for both data sets. For the BAND data, the reduction went from 16 down to 8, 9 and 11, respectively. About the same classification accuracy was achieved as on the original data sets. For the AAR data, the dimensionality decreased from 12 to 6 and 8, also keeping the good classification accuracy from the original data. On one occasion LVQ scored even better (5% for subject S1) and once worse (3% for subject S2). In general, however, this new feature extraction procedure seems to perform more reliably than unsupervised PCA.

Instead of the angle to the LDA, the error rate of each PC could also be used as a sorting criterion. Calculating the error rate for each of the components, however, can be very time consuming, especially when there are many dimensions and examples. For the AAR data this method gives about the same performance due to the high correlation between the error rate and the angle, as given in Section 4. For the BAND data it yields slightly better results (i.e. fewer components to reach similar classification results as with the whole data set).

Deriving the LDA and consequently calculating angles towards it is not very difficult and can be performed quickly. Thus, the proposed method describes a reasonably quick and partly supervised feature extraction method. For the Graz BCI this method could significantly reduce the amount of data used for classification. By examining the weights of the PCs, one can also discover which attributes contribute more to a good classification. The most important aspect for the Graz BCI is that PCA has to be used in combination with another method (e.g. the method presented in this paper) to find those PCs that can best be used for classification.

Acknowledgments—This work was supported by the Fonds zur Förderung der wissenschaftlichen Forshung,' project P11208, and the 'Allgemeine Unfallversicherungsanstalt (AUVA),' Austria. Thanks are due to the psychological staff at the Department of Medical Informatics' for recording and preprocessing of the data.

References

BISHOP, C. M. (1995): 'Neural networks for pattern recognition; (Clarendon Press, Oxford)

DUDA, R. O. and HART, P. E. (1973): 'Pattern classification and scene analysis' (John Wiley & Sons)

FLOTZINGER, D., KALCHER, J. and PFURTSCHELLER, G. (1992): 'EEG classification by learning vector quantization,' *Biomedizinische Technik*, **37**, pp. 303–309

FLOTZINGER, D., PFURTSCHELLER, G., NEUPER, Ch., BERGER, J. and MOHL, W. 1994): 'Classification of non-averaged EEG data by learning vector quanTization and the influence of signal preprocessing,' *Med. Biol. Eng. Comput.*, **32**, (5), pp. 571–576

KOHONEN, T. (1995): 'Self-organizing maps' (Springer, Berlin)

KALCHER, J., FLOTZINGER, D., NEUPER, Ch., GÖLLY, S. and PFURTSCHELLER, G. (1996): 'Graz Brain-Computer Interface II—towards communication between humans and computers based on online classification of three different EEG patterns,' *Med. Biol. Eng. Comput.*, **34**, pp. 382–388

MCFARLAND, D. J., NEAT, G. W., READ, R. R. and WOLPAW, J. R. (1993): 'An EEG-based method for graded cursor control,' *Psychobiol.*, **21**, pp. 77–81

MICHIE, D., SPIEGELHALTER, D. J. and TAYLOR, C. C. (1994):

'Machine Learning, Neural and Statistical Classification' (Ellis Horwood Limited, Englewood Cliffs, N.J.)
PELTORANTA, M. and PFURTSCHELLER, G. (1994): 'Neural network based classification of non-averaged event-related EEG responses,' *Med. Biol. Eng. Comput.*, **32**, pp. 189–196
PFURTSCHELLER, G., FLOTZINGER, D. and KALCHER, J. (1993): 'Brain-computer interface—a new communication device for handicapped persons,' *J. Microcomp. Appl.*, **16**, p. 293–299
PFURTSCHELLER, G., KALCHER, J., NEUPER, Ch., FLOTZINGER, D. and PREGENZER, M. (1996): 'On-line EEG classification during externally-paced hand movements using a neural net-work-based classifier,' *Electroenceph. Clin. Neurophy.*, **99**, pp. 416–425
PFURTSCHELLER, G., NEUPER, Ch., FLOTZINGER, D. and PREGENZER, M. (1997): 'EEG-based discrimination between imagination of right and left hand movement,' *Electroenceph. Clin. Neurophysiol.* (in press)
PREGENZER, M., PFURTSCHELLER, G. and FLOTZINGER, D. (1996): 'Automated feature selection with a distinction sensitive learning vector quantizer,' *Neurocomput.*, **11**, pp. 19–29
RIPLEY, B. D. (1996): 'Pattern recognition and neural networks' (Cambridge University Press)
SCHLÖGL, A. FLOTZINGER, D. and PFURTSCHELLER, G. (1997): 'Using adaptive autoregressive parameters for a brain-computer-interface experiment', Proc. 19th Int Conf. IEEE/EMBS, Chicago, pp. 1533–1535
WOLPAW, J. R., MCFARLAND, D. J., NEAT, G. W. and FORNERIS, C. (1991): 'An EEG-based brain-computer interface for cursor control,' *Electroenceph. Clin. Neurophysiol.*, **78**, pp. 252–259
WOLPAW, J. R. and MCFARLAND, D. J. (1994): 'Multichannel EEG-based brain-computer communication,' *Electroenceph. Clin. Neurophysiol.*, **90**, pp. 444–449

Author's biography

Klaus Lugger received his MSc in Technical Mathematics from the University of Technology in Graz, Austria in 1995. After his thesis he joined the Biomedical Engineering Department at the University of Technology, Graz, Austria. He is currently working on the Graz-BCI.