# Identifying User State Using Electroencephalographic Data


Matthias Honal
University of Karlsruhe
76128 Karlsruhe, Germany
honal@ira.uka.de

Tanja Schultz
Carnegie Mellon University
407 S. Craig Street
15213 Pittsburgh, USA
tanja@cs.cmu.edu



## ABSTRACT

In modern meeting environments people interact a lot with electronic communication devices such as computers, cell-phones, PDAs etc. which notify their users of events like incoming phone calls, text messages or e-mails. Depending on the current user state (e.g. listening, talking, resting, reading e-mails etc.) a different behavior of such devices might be desirable.

In this paper we investigate the possibility to detect the user's state from his brain activity by measuring his electroencephalogram (EEG). Using neural networks and support vector machines for classification an average accuracy for the discrimination of six user states of 94.1% in subject and session dependent and 58.9% in subject independent experiments is obtained. To make the application of EEG more realistic in meeting environments, we developed a comfortable headband with four build-in electrodes with which an average accuracy of 83.0% for discrimination of three user states in subject dependent experiments could be achieved.



## Keywords

User state identification, EEG, Meetings


## 1. INTRODUCTION

Electronic devices like laptops, cell-phones or PDAs are ubiquitous in modern meeting environments. During meetings users receive different kinds of notifications from these devices such as incoming phone calls, text messages, e-mails etc. Depending on the current user state, a disturbance by electronic devices may be either welcome or unwanted: A user who is listening attentively to a talk might feel disturbed even by a vibrating alert of his cell phone, while such an alert would perhaps be welcome when he is resting since he is not interested in the topic currently discussed. Thus it would be desirable to detect changes in the user's state automatically so that the communication devices can configure themselves appropriately. Furthermore information about the user's state could improve communication between meeting participants: A speaker who could tell that his audience is little interested in his talk could immediately try to change his style of presentation to regain the audience's attention.

In this paper we investigate the possibility to detect the user's state from his brain activity, i.e. by measuring his electroencephalogram (EEG) using scalp electrodes. To make EEG recording applicable in a meeting scenario, the following issues must be addressed:

- Robustness: In contrast to clinical EEG recordings, people must be able to talk and to move in a real meeting environment which introduces artifacts in the data.
- Speed: Signal processing and classification must be fast enough, since user state identification should be possible in real time.
- Realistic scenarios: The user states considered here shall be likely to occur really in meeting scenarios.
- User comfort: A user state detection device must be comfortable to wear and the user wearing it must find his appearance still acceptable. This means that little electrodes at suitable positions should be used and conductive gel should not get in contact with the user's hair.

This work should be seen as a first step in the research on how these goals could be reached by relaxing the inconveniences of clinical EEG recording to make it practical for user state detection.

### 1.1 Motivation for user state identification

To motivate the benefit of an EEG-based user state identification system, some scenarios shall be illustrated here, where user state identification may be helpful.

Typical user states during a meeting are talking, listening or watching slides, reading or resting. A user might want to have different configurations of his mobile devices during all of these states without having to change them manually whenever his state changes. While during talking all kinds of alerts might be disturbing, during *attentive* listening only those alerts could be delayed which the user classified to be of minor importance. A user who is reading e-mails may be open for more kinds of alerts, but he may not want to see pop up chat requests from his friends on his computer screen in contrast to a user who is resting. To know whether a user is reading the slides which are just presented or for example his e-mail, the distinction of the states audio-visual perception and reading might help.

For a speaker the information about the *average* state of his audience could be of interest. The states resting or reading might indicate that the talk is too easy, too difficult or not of general interest. When the average user state changes from audio-visual perception to listening or resting although the speaker is actively explaining his slides, the audience might have already understood the current slide or requires more visual information. It is important that in this scenario information about the user state is never used to *control* particular users, since this would violate their privacy. For this reason the speaker should only see the *average* state of his audience. Additionally it is important that users share the information about their user state voluntarily (using their own user state identification device) because they want to have a better and more efficient presentation.

The combination of EEG-based user state detection with other information could improve the interaction with electronic devices or between meeting participants even more. Currently we are working on EEG-based workload measurement, but also other techniques like speech and image recognition and natural language understanding could be combined with the user state information.

## 1.2 Bio-Medical Background

Using metal electrodes on the scalp and highly sensitive amplifiers oscillations of electrical potentials with amplitudes between $1\mu V$ and $100\mu V$ and frequencies between 0Hz and 80Hz can be registered. These oscillations which are commonly known as electroencephalogram (EEG) show specific characteristics at different scalp positions, depending on the current mental state [20].

The sources of the EEG are the potential differences emerging from the neuronal activity in the brain's cortex. Information between neurons is transfered via the synapses where chemical reactions take place which cause ion movements resulting in excitatory or inhibitory electrical potentials in the post-synaptic neurons. The electrical fields emerging from the ion movements are called cortical field potentials and have a dipole structure. If the electrical activity of a huge number of neurons is synchronized, the corresponding dipoles point all in the same direction and their sum is large enough, potential differences between particular scalp positions and a constant reference point can be measured. EEG characteristics like frequency, amplitude, temporal and topographic relations of certain patterns can then be used to make inferences about underlying neural activities [23].

A higher mental task is characterized by activity at a particular position of the cortex (figure 1) [20]. The user states which are considered in this work however are characterized by activity patterns at *several* positions of the cortex, which suggests that the activity of the whole cortex must be considered to achieve best discrimination results. For a task like reading for example, the visual cortex, the area for understanding symbols (letters in this case) and the area for understanding the semantic meaning of the words will have a characteristic activity [20].

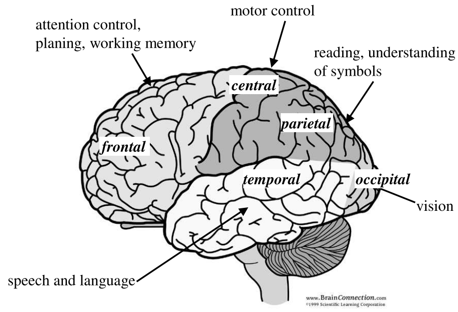


**Figure 1: Functional cortex divisions of the human brain according to [20]. Names of anatomical cortex areas are printed in italics.**

## 1.3 Related Work

A lot of work about detection of mental states using EEG data has been done in the past decade (see for instance [1], [2], [11], [8]) using the dataset recorded by Keirn and Aunon [17]. This dataset comprises recordings of five tasks which include resting with closed eyes, mental letter composing, imagination of counting numbers written on a board, rotation of an imaginary geometric figure and mental arithmetic, where each task had to be performed without vocalizing or moving.

Anderson et. al [1], [2] used data from six electrode channels from the central, parietal and occipital cortex to discriminate the tasks described above. In [1] neural networks were applied to classify a frequency-based representation of the signals resulting in an accuracy of up to 74% for discrimination of the baseline task from the arithmetic task. In [2] all five mental tasks were discriminated using a neural network with the coefficients of an $6^{th}$ order auto-regressive (AR) model for each of the six electrodes as input features. In experiments with four subjects an average accuracy of 54% was obtained when averaging over the network outputs of 20 consecutive half-second time windows. Ford [11] obtained an average accuracy of 88.5% for the discrimination of the baseline task from the arithmetic task using Learning Vector Quantization and the coefficients of the AR Model as described above as input features. Finally Culpepper [8] reports an accuracy of up to 94% for the discrimination of different pairs and and up to 86% for the discrimination of triplets out of all five mental tasks using independent component analysis for artifact removal and a frequency based representation of the EEG signals as input for neural networks. Note that all results were obtained in subject dependent experiments. Results in [1] and [2] were obtained across sessions, however they do not differ significantly from results reported for session dependent experiments.

More recently Chen et al. [6] introduced a Physiologically Attentive User Interface (PAUI) which uses heart rate variability to asses mental load and the $\mu$ power range (8Hz -

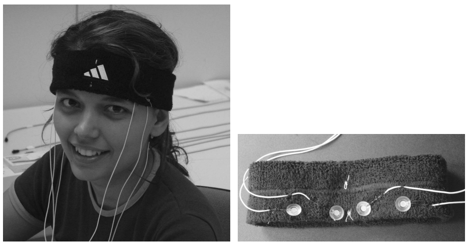

**Figure 2: Headband with four build-in electrodes.**

| Task | Avg. length | Range |
|---|---|---|
| (R) | 145 | $[143, 149]$ |
| (L) | 264 | $[169, 388]$ |
| (AV) | 360 | $[238, 495]$ |
| (RE) | 408 | $[223, 646]$ |
| (RS) | 206 | $[124, 345]$ |
| (C) | 252 | $[139, 496]$ |
| Total | 1636 | $[1267, 2132]$ |

**Table 1: Average length and range of task durations in seconds for all tasks over all subjects.**

30Hz) of a one-electrode EEG to measure motor activity. Using this information four attentional states of the user are distinguished: resting, moving, thinking and busy. As a sample application the automated regulation of notifications of a cellphone is proposed. In a six-person trial the correct state could be identified in 83% of all cases.

No research work could be found which attempts to identify user states beeing relevant in meeting scenarios which is the goal in this work. Therefore no direct comparison to the results obtained here can be given.

Other goals in computational processing of EEG data are monitoring of alertness (e.g. [16]) or task demand (e.g. [19], [5]). Furthermore EEG data is used in brain-computer interfaces where computers and other electronic devices should be controlled using only brain-activity (See for instance [22] for an overview).

## 2. DATA

For data acquisition two different devices were used: The major portion of the recordings was done with an ElectroCap$^{\text{TM}}$ EEG-cap [10] with electrode placements according to the international 10-20 system. 16 electrodes at the positions fp1, fp2, f3, f4, f7, f8, fz, t3, t4, t5, t6, p3, p4, pz o1 and o2 were recorded using the ElectroCap$^{\text{TM}}$ system. In order to show the feasibility of more comfortable EEG recording devices, we developed a headband with four built-in electrodes at the positions fp1, fp2, f7 and f8 (figure 2). The use of only these four electrodes has the advantage that no electrode paste gets into contact with the subjects hair and that only very little paste, which can be easily wiped off, remains after the recording on the forehead. Up to now only two recordings have been done using the headband.

In both cases reference electrodes were placed at or below the ear lobes and linked together, which results in a virtual reference point in the middle of the head. Amplification and A/D-conversion was done with a 16 channel VarioPort$^{\text{TM}}$ physiological data recorder [3]. Each channel had an amplification factor of 2775 and a frequency range from 0.9Hz to 60Hz. After amplification A/D conversion was performed using 4096 A/D-steps and a sampling rate of 256 Hz.

In contrast to recordings for clinical purposes, subjects were allowed to move freely during the recordings and the subject's head was *not* fixated to prevent muscular activity which usually introduces large artifacts in the data. During some of the recorded user states talking was even required.

User states were simulated by giving the subjects the following tasks to do: (R) resting, (L) listening to a talk, (AV) perceiving an audio-visual presentation, (RE) reading an article in a magazine, (RS) summarizing this article, (C) performing non-trivial calculations on a sheet of paper.

All tasks were recorded in a contiguous session lasting about 45 minutes. A resting period (task (R)) of about 70 seconds was inserted between each task or block of tasks (tasks (RE) and (RS) were performed as one block without resting in between). The order of the tasks or blocks of tasks between the resting periods was chosen randomly. Each task was adapted to the academical background of the subject to be recorded, so that mental demands were neither too high or too low. The talks (task (L)) contained a lot of facts, technical terms and complicated processes about selected topics of human physiology. This guaranteed a high mental load and a high memory demand which is typical for talks given without visual aid. For the audio-visual presentations topics from biology, medical imaging and computer science were chosen. Subjects were required to summarize the information obtained during tasks (L), (AV) and (A) which guaranteed their attention.

To give an idea about the amount of available data, table 1 shows the average length and the range of task durations (minimum and maximum) in seconds for all tasks over all subjects.

In order to avoid a bias of the classifiers towards one class (i.e. one task), for each class the same amount of training data had to be used. Thus the data collected during a resting period of only 70 seconds would limit the exploitation of the training data a lot. Therefore the data from two resting periods for each subject was concatenated, so that for each task at least 124 seconds of data (minimum for task (RS)) were available.

For the experiments reported in this work, the following data collected from volunteers was used:

**CMUSubjects** Six computer science students aged between 23 and 33 years (four males, two females) were recorded at Carnegie Mellon University in Pittsburgh (USA). None of them was a native English speaker, but they all had very good knowledge of the English language. 146 minutes of data in total were collected.

**UKASubjects:** Three university students with varying academic backgrounds aged between 23 and 26 years (two

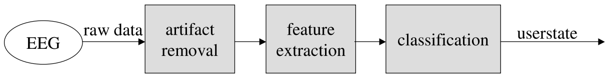


**Figure 3: Overview over the user state detection system**

males, one female) were recorded at University of Karlsruhe (Germany). One of them was a none-native German speaker. Two subjects were recorded twice. 153 minutes of data were available from this data collection.

**HeadBandSubjects:** 54 minutes of EEG data were collected at the University of Karlsruhe from two female university students aged between 21 and 23 years using the headband.

# 3. METHODS

Figure 3 shows the main components of our user state detection system. Independent component analysis (ICA) is applied for artifact removal. Then features are extracted representing the frequency content of the signal, simple feature normalization is performed and optionally linear discriminant analysis (LDA) is applied for feature reduction. The resulting features are then used by a multi-class support vector machine or a multilayer neural network for classification.

## 3.1 Artifact Removal

Due to eye movements or muscular activity, there is a number of artifacts which contaminate the EEG signals. Especially eye blinks cause large artifacts in the signal, since the corresponding muscles are very close to the (frontal) EEG electrodes and the electrical potentials caused by muscular activity are an order of magnitude larger than the sources of the EEG. ICA has been shown to be very efficient for the purpose of artifact removal (see, for instance [15]).

The signal measured at one EEG electrode can be seen as a linear combination of signals emerging from independent processes (i.e. cortical field potentials, muscular artifacts, 60Hz AC noise from electrical power lines etc.) [12]: Let $\vec{x}(t)$ be the vector of signals we are measuring at time $t$ at all electrodes and $\vec{s}(t)$ the real sources of the signals, i.e. the independent components. Then we can write

$$\vec{x}(t) = A \cdot \vec{s}(t)$$

where $A$ is called mixing matrix. The goal of ICA is now to determine the matrix $A$, or its inverse the unmixing matrix $W$, so that estimates for independent components can be obtained given the measured signals. Then components containing artifacts can then be rejected by visual inspection and the data can be projected back in the original coordinate system (figure 4).

The open source Matlab toolbox EEGLAB [9], was used for ICA computation is this work, which applies the Informax algorithm [4] for estimation of the ICA matrices. ICA estimation and component rejection was performed on the training data and the results were then applied to the test data.

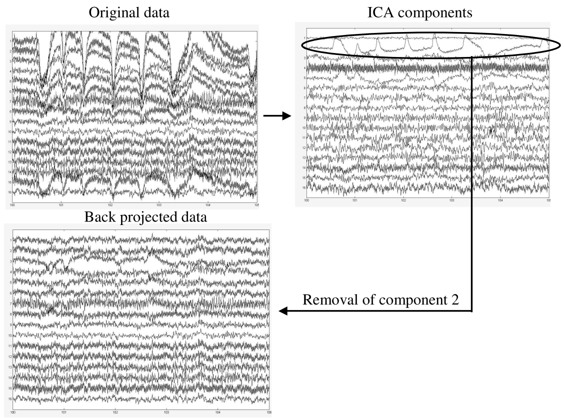


**Figure 4: Using ICA for artifact removal: independent components are computed from the original data, then the second component which contains eye blinking artifacts is rejected and the data is projected back to the original coordinate system.**

## 3.2 Feature Extraction and Normalization

After artifact removal features were extracted which represent the frequency content of the data. Two seconds long segments of the time signal which overlap one second were used to compute the power spectrum with a short time fourier transformation. This results in one feature per second and frequency band of 0.5Hz band width. In order to reduce the influence of single outliers among the features, averaging over the previous $k$ features was applied. The averaged feature for frequency band $f$ at time $t_0$ $x_f(t_0)$ is obtained from the features $x_f(t)$ as follows

$$x_f(t_0) = \sum_{i=0}^{k} \gamma(i) \cdot x_f(t_0 - i) \qquad (1)$$

In this work the value of $\gamma(i)$ was set constantly to 1, however it could also be used to decrease the influence of a features with increasing $i$.

Feature values for different frequency bands might have different ranges and they might fluctuate differently. However large fluctuations do not necessarily mean a large importance for classification. Therefore three simple normalization techniques were applied:

**GlobalNorm:** On the training data mean and variance is calculated for each electrode and each frequency band. The obtained values are then used for mean subtraction and variance normalization on the training, validation and test data.

**UserNorm:** Mean and variance normalization for each frequency band is performed separately on training, validation and test data. If more than one subject is in one of the datasets, normalization is performed separately for each subject.

**RelPower:** To preserve relations of feature values between frequency bands, the value for each frequency band is divided by the sum of values over all frequency bands.

Note that a possible drawback of this method is, that information about the relation of the power content between data points gets lost.

Although frequencies between 0Hz and 80Hz can be observed in normal EEG, according to [20] it is sufficient to consider the frequency content only up to 40Hz to monitor the mental processes we are interested in. Therefore we extracted features for the frequency range 0Hz to 45Hz, thus obtaining 90 features per electrode. Having 16 electrodes, this results in 1440 features total per data point which is huge compared to the relatively small amount of data (see table 1). Thus methods for feature reduction might help to obtain more reliable estimates for the models learned during classification. One such method is linear discriminant analysis (LDA) which can be used to find those features which discriminate best between given classes. The LDA algorithm ranks features according to their discriminative power. Thus feature reduction can be performed by taking the $n$ best ranked features.

## 3.3 Classification

Two classifiers are compared in this work: multilayer neural networks (NNs) and multi-class support vector machines (SVMs) which is a generalization of the standard SVM formulation to multiple classes proposed in [7].

We used NNs with one hidden layer and tanh activation functions. Training was performed with standard backpropagation with adaptive learning rate, early stopping using the validation set was applied to encounter the problem of overfitting. Since results obtained with neural networks fluctuate in a certain range due to random weight initializations, we trained several networks with the same data and used majority decisions over all networks to make predictions.

SVMs do not suffer from the problem of instable results due to different initializations. Furthermore "Xi-Alpha estimates" which represent and upper bound for the leaving-one-out error can be computed quickly from the SVM parameters which avoids time-consuming cross-validation[14].

Conventional SVMs are binary classifiers and multiple conventional SVMs must be trained to construct a classifier for more than two classes. In this work an SVM-like method which addresses explicitly the multi-class problem proposed in [7] is used. The classification problem is formulated here as follows: A data point $\vec{x}$ is assigned that class $k$ for which the similarity between its prototype $\vec{M}_k$ and $\vec{x}$ (expressed as $\vec{M}_k \cdot \vec{x}$) is maximal. The prototypes $M_k$ can be seen as the rows of a matrix $M$ whose norm must be minimized given some constraints. This problem can be solved using typical SVM optimization techniques.

In this work we used the multi-class version of the $SVM^{light}$ software [13] [21], which embeds the above formulation in a more general framework. Optimization algorithms of this software are designed to make SVM learning feasible for large scale problems which is essential for our data.

# 4. EXPERIMENTS AND RESULTS

Several types of experiments were conducted for this work:

**SD** User and session dependent experiments: Different data portions of the same recording were used for training (80% of the whole session, average length 19.5 minutes), testing and validation (each 10% of the whole data, average length 2.4 minutes). These types of experiments were conducted on the **CMUSubjects** data and on the **HeadBand** data.

**UI** User independent experiments: For the **CMUSubjects** data we trained our system on the data from five subjects (in average 97.4 minutes training data) and tested it on the data from the remaining subject in a round-robin manner. For comparability of the results the same test sets as in the **SD** setup were used. For validation 9.0 minutes of data from the **UKASubjects** data collection were available.

**SI** User dependent but session independent experiments: For the subjects from the **UKASubjects** data set which were recorded twice the system was trained using 80% of the the data of one session (in average 26.4 minutes) and tested on the test set from the other session (10% of the session data, in average 3.3 minutes) recorded from the same subject.

## 4.1 Baseline

### 4.1.1 Classifier Selection

Table 2 compares the results of a neural network (NN) with 20 neurons in the hidden layer and a linear SVM for experiment types **SD**, **UI** and **SI**. No averaging, artifact removal or feature reduction was done here. The normalization method **GlobalNorm** was applied to make input features suitable for the classifiers. Note that the neural network performance seems to be comparatively invariant against the number of neurons in the hidden layer. This parameter was varied between 15 and 25 without significant changes in results.

In case of experiment setup **UI**training can take up to one hour on a 3.0GHz Pentium 4 in the worst case for both classifiers. For the other setups training takes only a few minutes. For all three setups and both classifiers classification time for one sample is more than an order of magnitude below the sample length of one second, even on a 800MHz Pentium 3 laptop.

None of both classification methods performed significantly better than the other one. However neural networks have the problem, that results fluctuate depending on the weight initialization as mentioned above. Even when using seven networks to make a majority decision, there are standard deviations of about 6% for the average accuracies over all classes when repeating the same experiment five times for the same test set, particularly for experiment setups **SI** and **UI** where accuracies are comparatively low. For this reason we decided to use SVMs in the following experiments. Another finding is that, as one would expect, results decrease tremendously in session or even user independent experiments. Although the amout of training data is much larger for setup **UI** results are worse compared to setup **SI**, where the training set is about four times smaller. This suggests that without appropriate normalization or adaptation the generalization ability between sessions is much better than the generalization ability between subjects.

|  | NN | SVM |
|---|---|---|
| **SD** | 92.3% | 89.7% |
| **UI** | 37.55% | 38.2% |
| **SI** | 58.6% | 56.8% |

**Table 2: Average accuracies over all test sets and classes for different classifiers and different experiment setups.**

### 4.1.2 Performance for Individual User States

Figure 5 shows the accuracies for the single user states. Although accuracies for setup **UI** (black bars) are much below accuracies for setup **SD** (gray bars), the same tendencies concerning the variations between user states can be observed. For the setup **SI** (white bars), performance for the user states resting (R), reading (RE) and summarization of the read article (RS) is in the range of the session dependent experiments, while accuracies for the other user states are in the range for user independent experiments. This might indicate that user states like reading, talking and resting are more robust towards the variability of electrode placements or of the subject's mental state which is an inevitable problem for session independent experiments. Physiological reasons for this speculation remain to be investigated.

The user states listening to a talk (L) and perceiving an audiovisual presentation (AV) are mostly confused with each other and with resting (R). This is not surprising, since both user states involve the auditory system, and short resting-like periods might really occur during these states, since it is difficult to be constantly alert for a longer time. For the user independent experiments where the performance for user state (R) is bad as well, this state is mostly confused with user states (L) and (AV). The arithmetic task, i.e. user state (C) is mostly confused with reading (RE) and conversely user state (RE) is confused with user state (C) in user independent experiments. An explanation here might be that both states involve the understanding of symbols which is performed in the parietal cortex.

Note the standard deviation between subjects, depicted by the whiskers in figure 5, which is extremely high particularly for lower accuracies in setups **UI** and **SI**. Closer inspection of the results for particular subjects shows, that for each subject there are different user states which can be predicted extremely good or extremely bad using the training data from other subjects or sessions.

## 4.2 Normalization and Averaging

Figure 6 shows the impact of averaging over the previous $k$ samples according to equation 1 on the validation sets. For setup **SD** (solid line) gains for $k > 2$ are relatively small, while for setup **SI** (dotted line) considerable gains up to $k = 4$ can be observed. Therefore we decided to continue experiments with a value of $k = 2$ for setup **SD** which results in an accuracy of 93.4% on the test set and a value of $k = 4$ for setup **UI** resulting in an accuracy of 46.7% on the test set. For setup **SI** no appropriate validation set was available, since validation data should neither be taken from the test nor from the training session. Therefore we decided to use a value of $k = 3$ here, following the intuition that the properties for setup **SI** are somewhere between those of

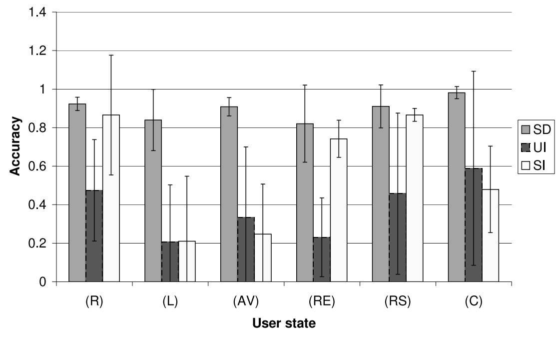


**Figure 5: Results for different user states. The standard deviation between subjects is shown by the whiskers.**

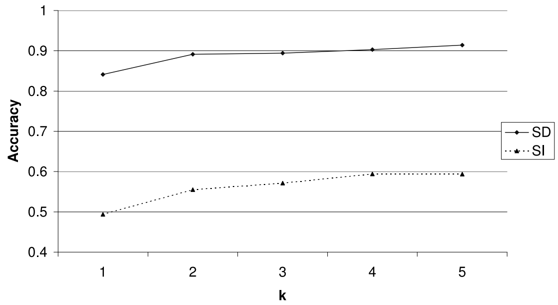


**Figure 6: Impact of averaging over previous samples for setup SD (solid line) and setup UI (dotted line).**

setups **SD** and **UI**. Thus an accuracy of 61.8% on the test set was obtained.

The results for different normalization methods for the different experiment setups are displayed in table 3. While there are no significant changes between method **GlobalNorm** and method **UserNorm** for setups **SD** and **SI** which contain only one subject in the training set, performance for setup **UI** improves significantly (relative gain 22.8%) when using method **UserNorm**. We conclude, that mean subtraction and variance normalization for each subject separately seems to reduce variabilities between subjects which are detrimental for classification performance. For normalization method **RelPower** results decrease tremendously for each setup. An explanation might be that the information about the relations between the power contents of the different samples gets lost. This information seems to be important however, since for most subjects clear differences in total power contents for the datapoints of different classes can be observed.

The difference between the best results for setup **UI** and **SI** is now very small compared to the results obtained for the baseline in table 2. However one can see that despite normalization efforts slightly better results are obtained when training on one session from the same subject, compared to five sessions from different subjects. Note furthermore that for no user state accuracies for setup **UI** and **SD** are in the same range, while this is still the case for the user states (R), (RE) and (RS) when comparing setups **SI** and **SD**.

| | GlobalNorm | UserNorm | RelPower |
|---|---|---|---|
| **SD** ($k = 2$) | 93.4% | 93.1% | 42.2% |
| **UI** ($k = 4$) | 46.7% | 58.9% | 34.1% |
| **SI** ($k = 3$) | 61.8% | 62.7% | 25.9% |

**Table 3: Results for different normalization methods for the different experiment setups. Results in the first column correspond to the baseline after finding the optimal value of $k$**

| | SD | SI |
|---|---|---|
| no ICA | 93.4% | 62.7% |
| ICA | 94.1% | 56.8% |

**Table 4: Best results without ICA application (see table 3) compared with results obtained with ICA-based artifact removal for setup SD and SI**

## 4.3 Artifact Removal

The use of ICA for artifact removal is straight forward as described in section 3.1 only for setup **SingleSub** where artifact contaminated components can be identified easily by visual inspection of the training data. (We concentrate here on eye-blinking artifacts, since their obvious impact on the EEG signal seems to be largest.) Furthermore only for this setup the ICA weights learned on the training data seem to be applicable to the test data in the sense that artifact components on the test data and the training data are the same. For setup **SI** this appears to be the case at first sight too, however the artifact components on the test data contain apparently to much of the actual EEG information, so that rejection of such components is detrimental to the results. Table 4 shows that for setup **SingleSub** small but significant improvements can be achieved (relative gain 10.6%), while for setup **SI** there is a significant performance loss. ICA-based artifact removal by visual inspection of the ICA components seems not to be possible for setup **UI**, since artifacts are spread over multiple channels. Removal of one or more such components would remove too much useful EEG information.

## 4.4 Feature Reduction

The computation of LDA coefficients involves the solution of an eigenvalue problem which does not have a stable solution in case of the user dependent data since not enough data points are available. Therefore LDA was only applied in the user independent experiments where the training sets were about four times larger. We tried to reduce the original data points with 1440 features down to a dimensionality of 25 features and compared the results to the case where no feature reduction was performed. Interestingly no significant differences in results on the validation data could be found for a reduction down to 75 features (figure 7). For testing we decided to reduce the dimensionality of the data only to 300 features. This "conservative" choice of the dimensionality was made to make sure that results are still stable, given the findings on the validation data. With 300 dimensional feature vectors an accuracy of 57.5% on the test set could be obtained, compared to 58.9% without feature reduction. Training time and memory consumption is reduced largely when 300 features instead of 1440 are used. We conclude that comparatively few features seem to be important for

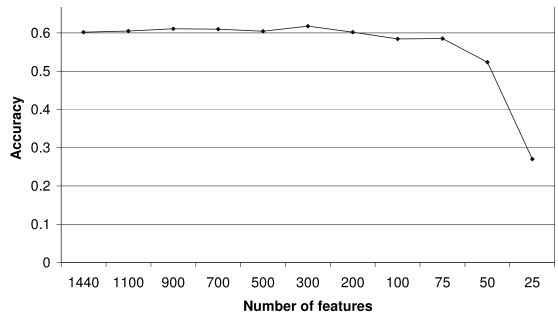


**Figure 7: Average performance on the validation set for a feature reduction from 1440 down to 25 features on setup UI. Note that non-equidistant scale on the x-axis for better visualization.**

the classification task. For practical applications this means particularly a gain in training and classification time and less memory consumption of the learned models. Furtermore models can be estimated more reliable with a smaller number of features.

Electrodes preferred by the LDA, i.e. electrodes from which most features are selected, are located over the frontal and the parietal cortex. This suggests that the EEG from these regions is particularly important for the user states we are considering in this work.

In future work other feature selection algorithms have to be applied, which can cope better with data sparseness in case of high dimensional data. Interesting results for a similar problem have been recently reported in [18]

## 4.5 Electrode Reduction

Findings from the previous section suggest, that the electrodes from the frontal *and* from the parietal cortex should be included in a set of reduced electrodes in any case. From a practical point of view the use of only the frontal electrodes fp1, fp2, f7 and f8 is more suitable however as argued in section 2.

Using only these electrodes from the EEG-cap recordings an average accuracy on the **CMUSubjects** data of 70.8% is obtained for setup **SD**. This compares to an accuracy of 65.3% obtained for the **HeadBand** data. When only the three user states resting (R), perceiving a presentation (AV) and reading (RE) are taken into account accuracies raise to 79.5% for the **CMUSubjects** data and to 83.0% for the **HeadBand** data for setup **SD**. The identification of these three user states is particularly important for applications in scenarios as described in section 1.1.

We conclude that user state detection with only four frontal electrodes is possible, however results drop significantly when no data from electrodes at other positions is available.

## 4.6 Adaptation to a Specific Subject

Since training data from only one session was available for session independent experiments, we tried to add the whole **CMUSubjects** training data to the original training data and tested in a round-robin manner on all four sessions for which we conducted session independent experiments. This resulted in an average accuracy of 61.1% which does not differ much from the 62.7% obtained when using only the training sets for setup **SI**This suggests that only additional noise is introduced when data from different subjects is added to enhance the training data for one particular subject.

## 5. CONCLUSIONS

In this paper we presented a system for user state identification using EEG data. Information about the current user state could be used to allow for more intelligent interaction between users and electronic communication devices or for more efficient interaction between users during meetings. Data for six user states which are typical for a meeting scenario was collected and several experiments were conducted.

Using support vector machines for classification and independent component analysis for artifact removal an average accuracy of 94.1% could be obtained for the discrimination of the six user states in user and session dependent experiments with a conventional EEG-cap. Average accuracies for session indepenent experiments (62.7%) and user indepent experiments (58.9%) are much below. In an experiment with a headband with four build-in electrodes an accuracy of 90.2% for the discrimination of three user states in a user and session dependent experiment was obtained. Improvements of electrodes and techniques for electrode attachment (e.g. on a glasses frame) are still needed however to make the use of wearable EEG devices even more acceptable and comfortable for users.

## 6. REFERENCES


[1] C. Anderson, S. Devulapalli, and E. Stolz. Determining Mental State from EEG Signals Using Neural Networks. *Scientific Programming, Special Issue on Applications Analysis*, 4(3):171–183, Fall 1995.

[2] C. Anderson and Z. Sijercic. Classification of EEG Signals from Four Subjects During Five Mental Tasks. In *Solving Engineering Problems with Neural Networks: Proceedings of the Conference on Engineering Applications in Neural Networks*, pages 407–414, 1996.

[3] K. Becker. VarioPort$^{TM}$. http://www.becker-meditec.com/.

[4] A. J. Bell and T. J. Sejnowski. An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation*, 7(6):1129–1159, 1995.

[5] C. Berka et al. Real-Time Analysis of EEG Indexes of Alertness, Cognition, and Memory Acquired With a Wireless EEG Headset. *International Journal of Human Computer Interaction*, 17(2):151–170, 2004.

[6] D. Chen and R. Vertegaal. Using Mental Load for Managing Interruptions in a Physiologically Attentive User Interface. In *Proceedings of the International Conference on Human Computer Interaction*, 2004.

[7] K. Crammer and Y. Singer. On the Algorithmic Implementation of Multiclass Kernel-Based Vector Machines. *Journal of Machine Learning Research*, 2(6):265–292, 2001.

[8] J. Culpepper. Discriminating Mental States Using EEG Represented By Spectral Power Density. Technical report, Harvey Mudd College, 1999.

[9] A. Delorme and S. Makeig. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods*, 134:9–21, 2004.

[10] Electro-Cap International, Inc. Electro-Cap$^{TM}$. http://www.electro-cap.com/.

[11] D. K. Ford. Analysis of LVQ in the Context of Spontaneous EEG Signal Classification. Master's thesis, Colorado State University, 1996.

[12] A. Hyvärinen and E. Oja. Independent Component Analysis: Algorithms and Applications. *Neural Networks*, 13(4-5):411–430, 2000.

[13] T. Joachims. *Making Large-Scale SVM Learning Practical*, chapter 11. MIT-Press, 1999.

[14] T. Joachims. Estimating the Generalization Performance of a SVM Efficiently. In *Proceedings of the ICML*, 2000.

[15] T. Jung et al. Removing Electroencephalographic Artifacts by Blind Source Separation. *Psychophysiology*, 37(2):163–178, 2000.

[16] T. Jung, S. Makeig, M. Stensmo, and T. Sejnowski. Estimating Alertness from the EEG Power Spectrum. *IEEE Transactions on Biomedical Engineering*, 4(1):60–69, January 1997.

[17] Z. Keirn and J. Aunon. A New Mode of Communication Between Man and his Surroundings. *IEEE Transactions on Biomedical Engineering*, 27:1209–1240, 1990.

[18] T. Mitchell et al. Learning to Decode Cognitive States from Brain Images. *Machine Learning*, 57(1–2):145–175, 2004.

[19] C. W. Pleydell-Pearce, S. E. Whitecross, and B. T. Dickson. Multivariate Analysis of EEG: Predicting Cognition on the basis of Frequency Decomposition, Inter-electrode Correlation, Coherence, Cross Phase and Cross Power. In *Proceedings of 38$^{th}$ HICCS*, 2003.

[20] R. F. Schmidt and G. Thews, editors. *Physiologie des Menschen*. Springer, 1997.

[21] I. Tsochantaridis, T. Hofmann, T. Joachims, and Y. Altun. Support Vector Machine Learning for Interdependent and Structured Output Spaces. In *Proceedings of the ICML*, 2004.

[22] T. Vaughan et al. Brain-computer interface technology: A review of the Second International Meeting. *IEEE Transactions on Neural Systems & Rehabilitation Engineering*, 11(2):94–109, June 2003.

[23] S. Zschocke. *Klinische Elektroenzephalographie*. Springer, 1995.