The BCI Competition 2003:
Progress and Perspectives in Detection and Discrimination of EEG Single Trials

Benjamin Blankertz, Klaus-Robert Müller, Gabriel Curio, Theresa M. Vaughan, Gerwin Schalk, Jonathan R. Wolpaw, Alois Schlögl, Christa Neuper, Gert Pfurtscheller, Thilo Hinterberger, Michael Schröder, Niels Birbaumer

***Abstract*— Interest in developing a new method of man-to-machine communication—a brain-computer interface or BCI—has grown steadily over the past few decades. BCIs create a new communication channel between the brain and an output device by bypassing conventional motor output pathways of nerves and muscles. These systems use signals recorded from the scalp, the surface of the cortex, or from inside the brain to enable users to control a variety of applications including simple word-processing software and orthotics. BCI technology could therefore provide a new communication and control option for individuals who cannot otherwise express their wishes to the outside world. Signal processing and classification methods are essential tools in the development of improved BCI technology. We organized the BCI Competition 2003 to evaluate the current state of the art of these tools.**

Four laboratories well versed in EEG-based BCI research provided six data sets in a documented format. We made these data sets (i.e., labeled training sets and unlabeled test sets) and their descriptions available on the Internet. The goal in the competition was to maximize the performance measure for the test labels. Researchers worldwide tested their algorithms and competed for the best classification results. This article describes the six data sets and the results and function of the most successful algorithms.

***Index Terms*— brain-computer interface, BCI, single-trial classification, slow cortical potentials, mu-rhythm, beta-rhythm, P300, rehabilitation, augmentative communication, ERP, EEG, imagined hand movements, lateralized readiness potential.**

I. Introduction

THE aim of Brain-Computer Interface (BCI) research is to establish a new augmentative communication system that translates human intentions—reflected by suitable brain signals—into a control signal for an output device such as a computer application or a neuroprothesis [1]. According to the definition put forth at the first international meeting for BCI technology in 1999, a BCI "must not depend on the brain's normal output pathways of peripheral nerves and muscles" [2]. While BCI research is a relatively young field, interest is increasing; researchers from 38 groups attended the second International BCI Meeting held in 2002, as compared to only six groups in 1994.

A BCI data competition was initiated in 2001 in an attempt to present common, relevant, well-defined data sets in order to evaluate and compare algorithms [3]. The BCI Competition 2003 was prompted by the success of that first competition, the recent growth of interest in BCI research, and the desire to address several key issues.

Three key issues underlie much present-day EEG-based BCI research: 1) data quality (is all task-relevant performance independent of conventional motor output?); 2) generalization/overfitting (do off-line results generalize to online experiments?); and 3) feedback (will methods developed on data collected without feedback work when feedback is provided?).

The design of the BCI Competition 2003 encompasses the first two issues. The third issue, feedback, is largely an empirical matter that must be addressed at least in part in online experiments.

To ensure that the contributions of competitors would be based solely on outputs of the central nervous system and not on artifacts arising from motor actions, we used data sets from four groups experienced in EEG-based BCI research who had significantly addressed the issue of such artifacts and eliminated them from their data [4], [5], [6].

Although each test set was unlabeled, methods might be adapted to the test set itself, e.g., by evaluating event-related potentials (ERPs) with respect to the estimated labels. This was especially the case for data set IIb, in which correct classification resulted mostly in complete English words. Therefore, hints about the targets were implicit in the data itself.

To evaluate the submissions to the competition with regard to a feedback mode, it would be necessary to implement online versions of some successful algorithms and to perform further experiments in the hosting BCI laboratories with the same subject(s) from which the competition data have been recorded. Given the present positive experiences with a substantial number of innovative and successful submissions, it could be a new and ambitious objective of future BCI competitions to integrate this online feedback evaluation into the competition. In addition to technical difficulties, the variable performance of any single subject would necessitate performing many sessions with different subjects to measure system performance.

BB and KRM are with Fraunhofer FIRST (IDA), Berlin, Germany, E-mail: `benjamin.blankertz@first.fraunhofer.de`. KRM is also with University of Potsdam, Germany.

GC is with the Dept. of Neurology, Campus Benjamin Franklin, Charité University Medicine Berlin, Germany.

TMV, GS, and JRW are with the Labatory of Nervous System Disorders, Wadsworth Center, New York State Dept. of Health, Albany, NY, USA. JRW is also with the State University of New York, Albany, NY, USA.

AS, CN, and GP are with the Dept. of Medical Informatics, Institute for Biomedical Engineering, Graz University of Technology, Austria.

MS is with the Dept. of Technical Computer Science, Eberhard-Karls-Universität Tübingen, Germany.

TH and NB are with the Institute of Medical Psychology and Behavioral Neurobiology, University of Tübingen, Germany. NB is also with the University of Trento, Italy.

TABLE I
IN THIS TABLE THE WINNING TEAMS FOR ALL COMPETITION DATA SET ARE LISTED. REFER TO SEC. IV TO SEE WHY THERE ARE MULTIPLE WINNERS OF DATA SET IIB.

data set	research lab	contributor(s)
Ia	Massachussets Institute of Technology, Boston	Brett Mensh, Justin Werfel, Sebastian Seung
Ib	University of Tübingen	Vladimir Bostanov
IIa	Fraunhofer FIRST (IDA), Berlin	Gilles Blanchard, Benjamin Blankertz
IIb	University of Bielefeld	Matthias Kaper, Peter Meinicke, Ulf Großekathöfer, Thomas Lingner, Helge Ritter
IIb	Tsinghua University, Beijing	Xiaorong Gao, Neng Xu, Xiaobo Miao, Bo Hong, Shangkai Gao, Fusheng Yang
IIb	University of Tübingen	Vladimir Bostanov
IIb	Fraunhofer FIRST (IDA), Berlin	Benjamin Blankertz, Gabriel Curio (Charité Berlin, CBF)
IIb	Fraunhofer FIRST (IDA), Berlin	David Tax, Benjamin Blankertz
III	Fraunhofer FIRST (IDA), Berlin	Christin Schäfer, Steven Lemm (Charité Berlin, CBF)
IV	Tsinghua University, Beijing	Zhiguang Zhang, Yijun Wang, Yong Li, Xiaorong Gao

A. Ranking of competition results

The ranking of results from Internet competitions cannot be taken at face value since they may not provide a completely objective assessment of quality for several reasons:

(1) There is great variance in how much effort contributors put into preparing their submissions.

(2) When test sets (and the number of classes) are relatively small, luck may also play a big role. For example, if there are 15 methods in a binary problem that are able to classify correctly 60 % of the ideal set of all trials with random output on the remaining 40 %, the expected accuracy of all these methods is 80 %. However, on a fixed test set consisting of 100 trials, the expected difference between the best and the worst result is greater than 10 % (assuming independence between methods and test trials).

In Sec. II–VI of this paper, we will describe the six data sets comprising the competition and we will report on and comment on the submissions. The results of all submissions are more fully reported on the web (`http://ida.first.fhg.de/~blanker/competition/results`) where we also list short descriptions of the applied methods. A list of the winning teams for each data set is reported in table reftab:winners. The winning labs are publishing individual articles on their approaches, see [7], [8], [9], [10], [11], [12], [13]. Note that two of winning teams on data set IIb (Tax et al. and Blankertz et al.) agreed not to publish articles on their approaches in order to have not so many articles on that particular data set. Nevertheless you can find information on their methods on the results page of the BCI Competition web site, see above.

II. DATA SETS IA AND IB: SELF-REGULATION OF SCPS

These data sets were provided by the Institute of Medical Psychology and Behavioral Neurobiology, University of Tübingen (head: Niels Birbaumer).

A. Description of the data set

Data set Ia was taken from a healthy subject. Data set Ib was taken from an artificially respirated completely paralyzed (locked-in) patient with amyotrophic lateral sclerosis (ALS). The subjects were asked to move a cursor up or down on a computer screen, while their slow cortical potentials (SCPs) were recorded. The subjects received visual feedback of their SCPs (Cz-Mastoids) which were corrected for vertical eye movements. Cortical positivity (negativity) led to a downward (upward) movement of the cursor on the screen. Each trial lasted 6 s in data set Ia, and 8 s in set Ib.

During each trial, the task to produce cortical negativity or positivity was visually presented by a highlighted goal at either the top or bottom of the screen from 0.5 s on. In addition, for data set Ib, the task was vocalised (“up” or “down”) at 0.5 s. The visual feedback was presented from second 2 to second 5.5 for set Ia and from second 2 to second 6.5 for set Ib. For the competition, only this interval of every trial was provided for training and testing in order to avoid the classification of brain responses related to task presentation or reinforcement. Brain activity was recorded from the following scalp positions at a sampling rate of 256 Hz: A1, A2, C3f, C3p, C4f, C4p, all referenced to Cz and the vEOG (A1/A2 = left/right mastoid, C3f means 2 cm frontal of C3, C3p 2 cm parietal of C3). vEOG data were not published for data set Ia to avoid classification of artifact data. To help a completely paralyzed patient data set Ib was provided with vEOG as it could provide useful information.

For data set Ia, 268 trials were recorded on two different days and mixed randomly. Of the total 268 trials, 168 originated from day 1 and the remaining 100 trials from day 2. For data set Ib, the training and test set each contain 200 trials recorded on the same day and permuted randomly. Competition participants had to submit their estimated class ratings for every trial of the test set. The performance measure was the correct response rate defined by the number of correctly classified trials divided by the total number of trials.

B. Outcome of the competition

We received fifteen submissions for data set Ia and eight submissions for data set Ib. For data set Ia, the competition winner was Brett Mensh from the Massachussets Institute of Technology with his co-workers Justin Werfel and Sebastian Seung. They won with an error rate of 11.3 % by using a linear discriminant analysis on the DC potentials of the first two channels and high beta power as additional dimensions for classification. Error rates below 12 % were also achieved by Guido Dornhege and co-workers from the Fraunhofer FIRST (IDA), Berlin, using regularized linear discriminant classifiers and by Kai-Min Chung and his group from the National Taiwan University, Taipei, who applied a support vector machine (SVM) classification on the data after downsampling to 25 Hz.

For data set Ib, the best result was achieved by Vladimir Bostanow from the Institute of Medical Psychology, University of Tübingen, Germany. He applied a stepwise linear discriminant analysis method on wavelet transformed data and achieved an error rate of 45.6 %. This result is close to chance level (i.e., 50 %), which might indicate that these data may not contain task-related information. The results for these data demonstrate the difficulties entailed in training a locked-in patient who cannot provide information about his state of consciousness, abilities, and level of motivation necessary for successful brain computer communication.

III. Data set IIa: self-regulation of mu- and/or central beta-rhythm

This data set was provided by the Wadsworth Center, New York State Department of Health (head: Jonathan R. Wolpaw).

A. Description of the data set

This comprehensive data set represents a complete record of actual BCI performance from 3 trained subjects in 10 sessions each.

In each trial, the subject sat in a reclining chair facing a video screen and was asked to remain motionless during performance. Scalp electrodes recorded 64 channels of EEG [14], each referred to an electrode on the right ear (amplification 20,000; band-pass 0.1–60 Hz). All 64 channels were digitized at 160 Hz and stored. Only a small subset of channels was used to control cursor movement online as described below.

The subjects used mu or beta rhythm amplitude (i.e., frequencies between 8–12 Hz or 18–24 Hz, respectively) to control vertical cursor movement toward the vertical position of a target located at the right edge of the video screen. Data were collected from each subject for 10 sessions of 30 min each. Each session consisted of six runs, separated by one-minute breaks, and each run consisted of about 32 individual trials. Each trial began with a 1 s period during which the screen was blank. Then the target appeared at one of four possible positions on the right edge of the screen. One second later, a cursor appeared at the middle of the left edge of the screen and started traveling across the screen from left to right at a constant speed. Its vertical position was controlled by the subject's EEG as described below. The subject's goal was to move the cursor to the height of the correct target. When the cursor reached the right edge, the screen went blank. This event signaled the end of the trial.

Cursor movement was controlled as follows: Ten times/sec, the preceding 200 ms of digitized EEG from 1-3 channels over sensorimotor cortex was re-referenced to a common average reference or a Laplacian derivation [15] and then submitted to frequency analysis by an autoregressive algorithm [16] to determine amplitude (i.e., the square root of power) in a mu and/or beta rhythm frequency band. The amplitudes for the 1-3 channels were combined to give a control signal that was used as the independent variable in a linear equation that controlled vertical cursor movement. Electrode position and center frequency remained constant for a particular subject, but certain parameters were updated online after each trial (e.g., parameters that estimate the signal's dynamics (i.e., the slope and the intercept of the linear equation that translated rhythm amplitude into cursor movement [17], [18]).

The objective in this contest was to use the labeled sessions (i.e., session 1-6) to train a classifier and to test this classifier by predicting the correct class (i.e., the target position) for each trial in the unlabeled sessions (i.e., sessions 7-10 for each subject). Participants were required to submit results only from causal classifiers (i.e., algorithms that only use preceding data to make a prediction).

For each contest participant, the average classification accuracy was calculated over all three subjects and four test sessions (sessions 7-10) by comparing the predicted target position in each trial with the actual target position in the trial during online operation. The participant with the highest average accuracy won the contest.

B. Outcome of the competition

We received five submissions to this data set (2 of these 5 submissions submitted results for only 2 subjects and were discarded). The best submission was by Gilles Blanchard and colleagues from Fraunhofer FIRST (IDA), Berlin with 71.8 % correct target prediction (compared to the 25.0 % that represents chance accuracy without any control) by using bandpass filtering, common spatial patterns and regularized linear discriminant analysis. This winning result is close to the results achieved online (73.2 %) using the linear equation described above.

IV. Data set IIb: P300 speller paradigm

This data set was provided by the Wadsworth Center, New York State Department of Health (head: Jonathan R. Wolpaw).

A. Description of the data set

This data set represents a complete record of P300 evoked potentials (3 sessions from one subject) recorded with the Wadsworth BCI2000 software [19], [20], using a paradigm described in [21] and originally by Farwell and Donchin [22]. In these experiments, a user focused on one of 36 different characters. The objective in the contest was to use the data from two sessions (i.e., 42 characters) to train a classifier, and to then predict the 31 characters in the one remaining session.

The user was presented with a 6 by 6 matrix of characters. The user's task was to focus attention on characters in a word that was prescribed by the investigator (i.e., one character at a time). The 6 rows and 6 columns of this matrix were successively and randomly intensified at a rate of 5.7 Hz. Two out of 12 intensifications of rows or columns highlighted the desired character (i.e., one particular row and one particular column). The responses evoked by these infrequent stimuli (i.e., the 2 out of 12 stimuli that did contain the desired character) are different from those evoked by the stimuli that did not contain the desired character and they are similar to the P300 responses previously reported [21], [22].

Signals were collected from one subject in three sessions and digitized at 240 Hz. Each session consisted of a number of runs. In each run, the subject focused attention on a series of characters. For each character, the user saw a matrix displayed for a 2.5 s period, and during this time each character had the same intensity (i.e., the matrix was blank). Subsequently, each row and column in the matrix was randomly intensified for 100 ms (i.e., resulting in 12 different stimuli, 6 rows and 6 columns). After intensification of a row/column, the matrix was blank for 75 ms. Row/column intensifications were block randomized in blocks of 12. Sets of 12 intensifications were repeated 15 times (i.e., 15 sequences) for each character (i.e., any specific row/column was intensified 15 times and thus there were 180 total intensifications for each character). Each sequence of 15 sets of intensifications was followed by a 2.5 s period, and during this time the matrix was blank. This period informed the user that this character was completed and to focus on the next character in the word that was displayed on the top of the screen (the current character was shown in parentheses).

The objective in the contest with this data set was to use the two labeled sessions to train a classifier and to test this classifier by predicting the 31 target characters in the one unlabeled session. Participants were also encouraged to report the minimum number of sequences that produced the same result.

B. Outcome of the competition

We received 7 submissions to this data set. Five of them predicted all 31 characters correctly, i.e., 100 % accuracy. (By comparison, the accuracy expected by chance was 2.8 %.) In addition, submissions needed only as few as 5 sequences (out of the 15 in the data file) to produce the same result. The winners are listed in table reftab:winners.

V. DATA SET III: MOTOR IMAGERY

This data set is provided by the Department of Medical Informatics, Institute for Biomedical Engineering, University of Technology Graz (head: Gert Pfurtscheller).

A. Description of the data set

This data set was recorded from a healthy subject (female, 25 yrs) during a feedback session. The subject sat in a relaxing chair with armrests. The task was to control a feedback bar in one dimension by imagination of left- or right-hand movements. The order of left and right cues was random.

The experiment included 7 runs with 40 trials each. All runs were conducted on the same day with breaks of several minutes in between. The data set consists of 280 trials of 9 second length. The first 2 s were quiet. At t=2 s, an acoustic stimulus indicated the beginning of the trial, and a cross ('+') was displayed for 1 s. Then at t=3 s, an arrow (left or right) was displayed as a cue stimulus. The subject was asked to use imagination as described above to move the feedback bar into the direction of the cue. The feedback was based on AAR parameters calculated from channels C3 and C4. The AAR parameters were combined with a discriminant analysis into one output parameter (similar to [23], [24]). The recording was made using a g.tec amplifier and Ag/AgCl electrodes. Three bipolar EEG channels were measured over C3, Cz and C4. EEG was sampled with 128 Hz and was filtered between 0.5 and 30 Hz. Similar experiments are described in [23], [24], [25], [26], [27]. The trials for training and testing were randomly selected to prevent any systematic effect due to the feedback.

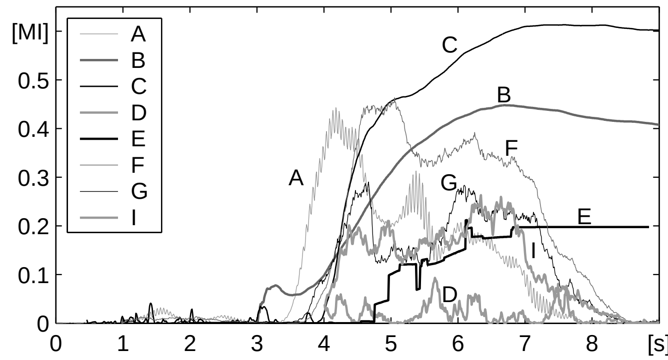

Fig. 1. Time course of the mutual information. At t=3 s the cue (left or right in random order) was presented. The increase of the mutual information indicates an increase in separability between left and right trials.

The task was to provide an analysis system to be used to control a continuous feedback, i.e., continuous values (<0 class '1', >0 class '2', 0 non-decisive) for each time point. The magnitude of the value should reflect the confidence of the classification with the sign indicating the class. A description of the analysis system was required.

Since there is a close relationship between the error rate and the mutual information (MI) [26], MI was used because it also takes into account the magnitude of the outputs. The criterion was the ratio between the maximum of the mutual information and the time delay since the cue (t=3 s).

B. Outcome of the competition

We received 9 submissions from 7 groups. One of these submissions contained only class labels for each trial and no continuous information on magnitude or time so that no time-variation could be obtained. Fig. 1 compares the time courses of the mutual information with the 9 submissions labeled Methods A through I. Because of the similarity in the MI time course, Methods A and F might use a very similar property of the EEG. (The delay might be explained when the different delay times are considered.) Methods G and I reach 0.26 and 0.21 bits at t=4.66 s and 6.34 s, respectively. Method H did not provide any time information and, moreover, the result did not correlate with the true class labels.

In evaluating these submissions, several issues were considered. First, although it is quite common to use the error rate for comparing different methods, the error rate takes only the sign of the classifier output but not the magnitude into account. For this reason, the mutual information was used to compare the different results [26], [27]. Moreover, it was important to decide whether or not to consider the time delay. Although the time-delay does not matter in offline analysis, it becomes important for fast and accurate online feedback. For offline analysis, one needs to compare just the maximum separability of the data; for online analysis, the steepness of the increase of MI is of interest. In this respect, Methods A, C and F have a similar steepness. Method A is 0.5 s earlier, but this might be due to a non-causal filter. (In real-time processing, one has to add this 0.5 s). Method I has a similar steepness, but does not reach a comparable maximum. Method G starts earlier, but the increase is not as steep. Method B starts at t=3 s, reaches its first peak at t=3.3 s, decreases and starts a slow increase. Physiological considerations suggest that the first peak does not represent deliberate activity, because conscious brain activity requires more time. Thus, the first peak might reflect a stimulus response. For this reason, only the second and larger peak was considered. Here, Methods A, C, F, and for a short period even Method G, are superior to Method B. In summary, the methods A, C and F provide the fastest increase in MI. Moreover, Method G deserves further investigation because of its early start.

Since this is an offline analysis and obviously not all results are based on causal algorithms, the time delay of the different methods cannot be compared. Hence, the final evaluation criterion is based on the maximum separability. According to this criterion, Method C submitted by Christin Schäfer and Steven Lemm, Fraunhofer FIRST (IDA), gave the best result, with an MI of 0.61 bits (error = 10.7 %).

VI. Data set IV: self-paced tapping

This data set was provided by Fraunhofer FIRST, Intelligent Data Analysis Group (head: Klaus-Robert Müller), and Charité University Medicine Berlin, Campus Benjamin Franklin, Department of Neurology, Neurophysics Group (head: Gabriel Curio).

A. Description of the data set

This data set was recorded from a healthy subject during a session with no feedback. The subject sat in a normal chair, relaxed arms resting on the table, fingers in the standard typing position at the computer keyboard. The task was to press with either the index or the little finger of either the left or the right hand one of four assigned keys in self-chosen order and timing (‘self-paced tapping’). The experiment consisted of 3 runs of 6 minutes each. All runs were conducted in one session with some minutes break in between. Typing was performed at an average speed of 1 key tap per second.

For the competition, 416 epochs of 500 ms EEG were provided, each ending 130 ms before an actual key press. (This choice of an early endpoint ensured that almost all trials were free of EMG activity.) The epochs were randomly shuffled and split into a training set (316 epochs) which is labeled ‘0’ for upcoming left hand and ‘1’ for upcoming right hand movements, and an unlabeled test set (100 epochs). EEG was recorded from 28 scalp positions, mainly covering the primary (sensori-)motor cortices bilaterally. Signals were provided with the original 1000 Hz sampling rate as well as in a version downsampled at 100 Hz. The goal for the competition was to submit estimated labels for the test set with minimum number of misclassifications.

B. Outcome of the competition

There were 15 submissions for this data set, 4 of which had a performance close to chance level (⩾43 % error). The best submission was given by Zhiguang Zhang and colleagues (see Sec. I-A) with an error of 16 %; the employed method is described in [13]. The second-best submission came from Radfort Neal (University of Toronto), who reached an error of 19 % by a Bayesian logistic regression classifier applied to a 188-dimensional feature vector (time- and frequency-domain and correlational features) that was compiled by hand. In addition, there were 6 contributions with an error between 23 and 29 %. In comparison, the Berlin BCI method described in [6], using standard parameters established for a larger group of subjects, achieved a classification with 17 % error. This error could be further reduced by specifically adapting the parameters to this set of training data.

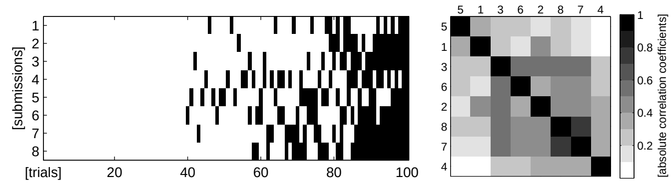

Fig. 2. The plot on the left shows for each trial and each of the 8 best submissions the classification success, i.e., whether the estimated label was correct (white) or not (black). Submissions are labeled by their rank (sorted by number of misclassifications on the test set). Trials are sorted along the x-axis according to the number of misclassifications in these 8 methods. The figure on the right illustrates the normalized covariance matrix (absolute values) between the classification success of the 8 best submissions. The sorting of submissions was done deliberately to reveal some block structure.

In Fig. 2 we show for each trial and for each of the 8 best submissions (error$<$30 %), whether or not the estimated label was correct. We sorted the trials according to the number of misclassifications for these 8 best methods. More than half of the trials were classified correctly by all or by all but one method. Remarkably, 10 % of the trials were misclassified by at least 6 of the 8 methods. This suggests that: (a) artifacts could render the EEG movement signals irretrievable, or (b) movements are not yet reflected adequately in the chosen set of EEG parameters.

VII. General Discussion

As described in the introduction a major objective of the BCI Competition 2003 was to learn about the state of the art how the common problem of overfitting can be handled. The common solutions (e.g., cross-validation or leave-one-out estimation) might fail to prevent overfitting in some important cases, as, for example, in the instance where parameters (or features) are chosen by selecting those with minimum cross-validation error. Nested cross-validations could be considered, but when parameters are selected from a huge search space it is often very difficult to rule out overfitting effects. Most helpful in such situations is verify whether the selected parameters match with some a-priori knowledge about the problem.

Looking at all the results of the competition, it is notable that for each data set there were submissions with an accuracy near chance level. It can be speculated that those contributors had a considerably better validation error on the training set, i.e., they expected their algorithms to perform well also on the test set (except for data set Ib, where in some contributions it was noted that only an accuracy at chance level was achieved and the contribution would be 'just for fun'.) Although some of the failures may also have been due to technical problems, overfitting problems are presumably the main cause. This indicates the need to be alert when reading or reviewing articles reporting results from offline analyses. On the other hand, for all data sets (except for set Ib, which might comprise no useful information, as discussed above), there have also been promising results using sophisticated approaches. This shows that it is in fact possible to adapt complex models to intricate data like EEG with good generalizability.

VIII. Outlook

We organized the BCI Competition 2003 to evaluate the current state of the art of signal processing and classification methods. We received many submissions to our posted data sets that illustrate the interest in BCI communication. The results demonstrate that complex models can capture essential features of intricate data such as EEG, but much care has to be taken in choosing and adapting the models. The accompanying papers of the winner teams provide a toolbox documenting a rich diversity of algorithmic approaches to the most important EEG-based BCI paradigms under study today.

The data sets and their descriptions will continue to be available on the competition web page [28]. Other researchers interested in EEG single-trial analysis are welcome to test their algorithms on these data sets and to report their results. To imitate competition conditions, all selections of method, features and model parameters must be confined to the training sets. However, due to the current availability of the labels of the test data and the publication of thorough analyses of these data, future classification results of the competition data cannot fairly be compared to the original submissions.

Since we received very positive responses to the BCI Competition 2003, it is highly probable that we will organize another competition in the future.

Acknowledgment

We thank all people who contributed to this competition, either by submitting classification results, or by giving feedback about the competition.

Authors BB, KRM and GC were partially supported by a grant of the BMBF, FKZ 01IBB02A and 01IBB02B.

Authors TMV, GS and JRW were partially supported by NIH grants HD30146 and EB00856.

Authors AS, CN and GP were partially supported by "Bundesministerium für Verkehr, Innovation und Technologie" (Austria) grant 140.587/2-V/B/9b/2000 and Amt der Steiermärkischen Landesregierung FA6A-10Vh00/1.

References

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, pp. 767–791, 2002.

[2] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, and T. M. Vaughan, "Brain-computer interface technology: A review of the first international meeting," *IEEE Trans. Rehab. Eng.*, vol. 8, no. 2, pp. 164–173, 2000.

[3] P. Sajda, A. Gerson, K.-R. Müller, B. Blankertz, and L. Parra, "A data analysis competition to evaluate machine learning algorithms for use in brain-computer interfaces," *IEEE Trans. Neural Sys. Rehab. Eng.*, vol. 11, no. 2, pp. 184–185, 2003.

[4] B. Kotchoubey, H. Schleichert, W. Lutzenberger, and N. Birbaumer, "A new method for self-regulation of slow cortical potentials in a timed paradigm," *Appl. Psychophysiol. Biofeedback*, vol. 22, no. 2, pp. 77–93, 1997.

[5] T. M. Vaughan, L. A. Miner, D. J. McFarland, and J. R. Wolpaw, "EEG-based communication: analysis of concurrent EMG activity," *Electroencephalogr. Clin. Neurophysiol.*, vol. 107, pp. 428–433, 1998.

[6] B. Blankertz, G. Dornhege, C. Schäfer, R. Krepki, J. Kohlmorgen, K.-R. Müller, V. Kunzmann, F. Losch, and G. Curio, "Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis," *IEEE Trans. Neural Sys. Rehab. Eng.*, vol. 11, no. 2, pp. 127–131, 2003.

[7] B. Mensh, J. Werfel, and S. Seung, "Bci competiton 2003 – data set Ia," *IEEE Trans. Biomed. Eng.*, 2004, submitted.

[8] V. Bostanov, "Bci competiton 2003 – data sets Ib and IIb," *IEEE Trans. Biomed. Eng.*, 2004, submitted.

[9] G. Blanchard and B. Blankertz, "BCI competition 2003 – data set IIa: Spatial patterns of self-controlled brain rhythm modulations," *IEEE Trans. Biomed. Eng.*, 2004, to appear.

[10] M. Kaper, P. Meinicke, U. Großekathöfer, T. Lingner, and H. Ritter, "Bci competiton 2003 – data set IIb," *IEEE Trans. Biomed. Eng.*, 2004, submitted.

[11] X. Gao, N. Xu, X. Miao, B. Hong, S. Gao, and F. Yang, "Bci competiton 2003 – data set IIb," *IEEE Trans. Biomed. Eng.*, 2004, submitted.

[12] S. Lemm, C. Schäfer, and G. Curio, "On-line classification of imagery hand movements. BCI competition 2003 - data set III," *IEEE Trans. Biomed. Eng.*, 2004, in revision.

[13] Z. Zhang, Y. Wang, Y. Li, and X. Gao, "Bci competiton 2003 – data set IV," *IEEE Trans. Biomed. Eng.*, 2004, submitted.

[14] F. Sharbrough, G.-E. Chatrian, R. Lesser, H. Lüders, M. Nuwer, and T. W. Picton, "American electroencephalographic society guidelines for standard electrode position nomenclature," *J. Clin. Neurophysiol.*, vol. 8, pp. 200–202, 1991.

[15] D. J. McFarland, L. M. McCane, S. V. David, and J. R. Wolpaw, "Spatial filter selection for EEG-based communication," *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, pp. 386–394, 1997.

[16] D. J. McFarland, A. T. Lefkowicz, and J. R. Wolpaw, "Design and operation of an EEG-based brain-computer interface with digital signal processing technology," *Behav. Res. Methods Instrum. Comput.*, vol. 29, no. 3, pp. 337–345, 1997.

[17] H. Ramoser, J. R. Wolpaw, and G. Pfurtscheller, "EEG-based communication: Evaluation of alternative signal prediction methods," *Biomed. Technik*, vol. 42, no. 9, pp. 226–233, 1997.

[18] D. J. McFarland and J. R. Wolpaw, "EEG-based communication and control: Speed-accuracy relationships," *Appl. Psychophysiol. Biofeedback*, vol. 28, no. 3, pp. 217–231, 2003.

[19] J. R. Wolpaw, G. Schalk, D. J. McFarland, T. Hinterberger, J. Perelmouter, B. Godde, N. Birbaumer, and G. Pfurtscheller, "BCI2000: A general purpose brain-computer interface," *Soc. Neurosci. Abstr.*, vol. 26, no. 1, p. 1229, 2000.

[20] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. Wolpaw, "BCI2000: A general-purpose brain-computer interface (BCI) system," *IEEE Trans. Biomed. Eng.*, 2004, submitted.

[21] E. Donchin, K. M. Spencer, and R. Wijesinghe, "The mental prosthesis: Assessing the speed of a P300-based brain-computer interface," *IEEE Trans. Rehab. Eng.*, vol. 8, no. 2, pp. 174–179, June 2000.

[22] L. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalogr. Clin. Neurophysiol.*, vol. 70, pp. 510–523, 1988.

[23] A. Schlögl, K. Lugger, and G. Pfurtscheller, "Using adaptive autoregressive parameters for a brain-computer-interface experiment," in *Proceedings 19th International Conference IEEE/EMBS*, 1997, pp. 1533–1535.

[24] C. Neuper, A. Schlögl, and G. Pfurtscheller, "Enhancement of left-right sensorimotor EEG differences during feedback-regulated motor imagery," *J. Clin. Neurophysiol.*, vol. 16, no. 4, pp. 373–382, 1999.

[25] G. Pfurtscheller, C. Neuper, A. Schlögl, and K. Lugger, "Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters," *IEEE Trans. Rehab. Eng.*, vol. 6, no. 3, pp. 316–325, 1998.

[26] A. Schlögl, C. Neuper, and G. Pfurtscheller, "Estimating the mutal information of an EEG-based brain-computer interface," *Biomed. Technik*, vol. 47, no. 1-2, pp. 3–8, 2002.

[27] A. Schlögl, C. Keinrath, R. Scherer, and G. Pfurtscheller, "Information transfer of an EEG-based brain-computer interface." in *Proceedings of the 1st International IEEE EMBS Conference on Neural Engineering*, 2003, pp. 641–644.

[28] B. Blankertz, "BCI competition 2003 (web page)." [Online]. Available: http://ida.first.fhg.de/projects/bci/competition/