Sensorimotor Rhythm-Based Brain–Computer Interface (BCI): Feature Selection by Regression Improves Performance

Dennis J. McFarland and Jonathan R. Wolpaw

***Abstract*—People can learn to control electroencephalogram (EEG) features consisting of sensorimotor rhythm amplitudes and can use this control to move a cursor in one or two dimensions to a target on a screen. In the standard one-dimensional application, the cursor moves horizontally from left to right at a fixed rate while vertical cursor movement is continuously controlled by sensorimotor rhythm amplitude. The right edge of the screen is divided among 2–6 targets, and the user's goal is to control vertical cursor movement so that the cursor hits the correct target when it reaches the right edge. Up to the present, vertical cursor movement has been a linear function of amplitude in a specific frequency band [i.e., 8–12 Hz (mu) or 18–26 Hz (beta)] over left and/or right sensorimotor cortex. The present study evaluated the effect of controlling cursor movement with a weighted combination of these amplitudes in which the weights were determined by an regression algorithm on the basis of the user's past performance. Analyses of data obtained from a representative set of trained users indicated that weighted combinations of sensorimotor rhythm amplitudes could support cursor control significantly superior to that provided by a single feature. Inclusion of an interaction term further improved performance. Subsequent online testing of the regression algorithm confirmed the improved performance predicted by the offline analyses. The results demonstrate the substantial value for brain–computer interface applications of simple multivariate linear algorithms. In contrast to many classification algorithms, such linear algorithms can easily incorporate multiple signal features, can readily adapt to changes in the user's control of these features, and can accommodate additional targets without major modifications.**

***Index Terms*—Brain–computer interface (BCI), electroencephalography, learning, mu rhythm, rehabilitation, sensorimotor cortex.**

I. Introduction

Many people with severe motor disabilities require alternative methods for communication and control. Numerous studies over the past two decades indicate that scalp-recorded electroencephalogram (EEG) activity can be the basis for nonmuscular communication and control systems, commonly called brain–computer interfaces (BCIs) [1]–[7]. EEG-based communication systems measure specific features of EEG activity and use the results as control signals. In some systems, these features are potentials evoked by stereotyped stimuli [1], [2]. Other systems, such as our own, use EEG components in the frequency or time domain that are spontaneous in the sense that they are not dependent on specific sensory events [3]–[8].

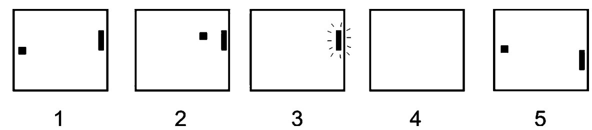

Fig. 1. Standard 1-D cursor control protocol with four possible targets. (1) Each trial begins when the cursor appears in the middle of the left edge of the screen and a target appears occupying one of the four quarters of the right edge. (2) 1 s later, the cursor begins to move steadily across the screen with its vertical movement controlled by the user's EEG. (3) In 2 s, the cursor reaches the right edge. If it hits the target, the target flashes for one s as a reward. If it misses the target, the screen simply goes blank. (4) Then, in either case, the screen is blank for 1 s. (5) Next trial begins.

With the Wadsworth BCI system, users learn over a series of training sessions to use sensorimotor rhythm amplitude in a mu (8–12 Hz) or beta (18–26 Hz) frequency band over left and/or right sensorimotor cortex to move a cursor on a video screen in one or two dimensions (see [9] for a full system description). Fig. 1 illustrates the standard one-dimensional (1-D) protocol. During each trial, the user is presented with a target somewhere along the right edge of the screen and a cursor on the left edge. The cursor moves across the screen at a steady rate, with its vertical movement controlled by sensorimotor rhythm amplitude. The user's task is to move the cursor to the height of the target so that it hits the target when it reaches the right edge.

Up to the present, vertical cursor movement (which occurs every 50 μs) has been a linear function of mu- or beta-rhythm amplitude over left or right sensorimotor cortex or of the sum of left and right amplitudes. That is, if V is the cursor movement, S is the control signal (i.e., one or the sum of two mu- or beta-rhythm amplitudes), b is the gain, and a is the mean control signal for the user's previous performance, then

$$V = b(S - a) \quad (1)$$

is the function that determines each cursor movement. This form of the linear equation is used so that the parameters a and b can be defined independently of each other. These two parameters are continually adjusted automatically so as to make all the targets equally accessible to the user [10], [11].

Manuscript received September 17, 2004; revised February 24, 2005; accepted March 17, 2005. This work was supported in part by the National Institutes of Health under Grant NICHD HD30146 and Grant NIBIB/NINDS EB00856 and in part by the James S. McDonnell Foundation.

The authors are with the Laboratory of Nervous System Disorders, Wadsworth Center, New York State Department of Health and State University of New York, Albany, NY 12201 USA (e-mail: mcfarlan@wadsworth.org).

Digital Object Identifier 10.1109/TNSRE.2005.848627

TABLE I
USERS AND THE FEATURES THEY USED FOR CURSOR CONTROL ONLINE

User	Sex	Age	Disability?	Features Used Online
A	F	32	None	Left and Right Mu
B	M	39	SCI	Left Mu and Beta
C	M	26	None	Left Mu
D	M	37	User	Left and Right Mu
E	M	30	None	Left and Right Mu
F	F	39	None	Left and Right Mu
G	M	20	SCI	Left and Right Mu

The control signal S in (1) can be expressed in terms of its constituent features (i.e, mu- or beta-rhythm amplitudes) as

$$S = \Sigma w_i x_i \tag{2}$$

where x_i is the ith feature and w_i is the weight given to that feature. Thus, S is the linear weighted sum of the features. Assuming that Σw_i is kept constant, the specific weights assigned to the different features can be controlled independently of the overall gain [i.e., b in (1)]. Our previous studies have used only one or two features, and the weights have always been 1.0.

The present study explored the impact on BCI performance of using four features (i.e., mu and beta rhythm amplitudes from right and left sides), including feature interaction terms as additional features, and adaptively adjusting the weights assigned to each feature. The addition of interaction terms projects the data into a higher dimensional space so as to take into account nonlinear relationships among the EEG features. Offline analyses and subsequent online evaluation indicate that these procedures can substantially improve BCI performance.

II. METHODS

A. Users

As shown in Table I, the BCI users were seven adults (five men and two woman, ages 20–39). Five had no disabilities, while two had spinal cord injuries (at C6 and T7) and were confined to wheelchairs. All gave informed consent for the study, which had been reviewed and approved by the New York State Department of Health Institutional Review Board. After an initial evaluation defined the frequencies and scalp locations of each person's spontaneous mu and beta rhythm activity, he or she learned EEG-based cursor control over several months (2–3 sessions/week). Table I shows the locations and frequencies of the features the users employed for cursor control at the end of training. The standard online protocol, which has been described in previous publications [9], [12], [13], is summarized here.

B. Standard Online Protocol

The user sat in a reclining chair facing a 51-cm video screen three meters away, and was asked to remain motionless during performance. Scalp electrodes recorded 64 channels of EEG [14], each referenced to an electrode on the right ear (amplification 20 000; bandpass 0.1–60 Hz; sampling rate 160 Hz).

A daily session had eight 3-min runs separated by 1-min breaks, and each run had 20–30 trials. As illustrated in Fig. 1, each trial consisted of a 1-s period from target appearance to the beginning of cursor movement, a 2-s period of cursor movement, a 1.5-s post-movement reward period, and a 1-s inter-trial interval. Users participated in 2–3 sessions per week each on a different day.

To control vertical cursor movement, one EEG channel over left sensorimotor cortex (i.e., electrode locations C3 or CP3 [14]) and/or one channel over right sensorimotor cortex (i.e., C4 or CP4) were derived from the digitized data according to a Laplacian transform [15]. Every 50 μs, the most recent 400-μs segment from each channel was analyzed by a 16th-order autoregressive model using the Berg algorithm [16] to determine the amplitude (i.e., square root of power) in a 3-Hz-wide mu- or beta-frequency band, and the amplitudes of the one or two channels were used in a linear equation that specified a cursor movement as described above. Thus, cursor movement occurred 20 times/s. Table I shows for each user the location(s) and frequency(ies) that controlled cursor movement. Complete EEG and cursor movement data were stored for later offline analysis.

C. Offline Analysis

In the standard online protocol, each user employed one or two mu or beta rhythm amplitudes over right or left sensorimotor cortex (i.e., Table I) to control cursor movement. To assess the potential value of controlling cursor movement with weighted combinations of mu and beta rhythm amplitudes from right and left sides, we calculated, in offline analyses of the data from each of the user's, the correlations with target location of each amplitude (i.e., left mu, left beta, right mu, right beta) singly and in weighted combinations using the multiple regression procedure from SAS [17]. Parameter estimates were determined using least-squares criteria and the normal equations

$$(X'X)b = X'Y \tag{3}$$

where X is a m by n matrix formed from the n observations of m predictor variables (i.e., EEG amplitudes at specific frequencies and locations) and Y is the vector of n values (i.e., target positions) to be predicted. Solving for b yields

$$b = (\mathrm{X'X})^{-1}\mathrm{X'Y}. \tag{4}$$

Correlation was expressed as r^2, the proportion of the total variance in target location that was accounted for by the model for the 2-s cursor movement period.

D. Generalization and Online Performance of Adaptive Feature Weights

Two users (B and E) were available for additional evaluation of adaptive feature weights. This evaluation had two goals. One goal was to determine whether the improvements achieved by weighted combinations of features persisted when the same weights were generalized to new data. For this purpose, we obtained a second data set (i.e., five more sessions) from Users B and E, analyzed these new data using either new weights calculated from the new data or the old weights obtained from the previous data set, and compared the resulting r^2 values. In these additional five sessions, online control of cursor movements was exactly the same as in the previous five sessions (i.e., it used the features shown in Table I without weights).

The other goal was to determine whether these r^2 effects were maintained when the weights calculated from previous data were actually used online, and whether they were accompanied by higher accuracy (i.e., better target-hit percentages) as expected from previous studies [18]. For this purpose, we obtained a third data set (i.e., eight additional sessions) from Users B and E. In these sessions, cursor movement in half of the runs was actually controlled online by the weighted combination of the best pair of features and their interaction. For comparison, cursor movement in the other half of the runs was controlled by the best single feature. The weights used online were those calculated from the second data set.

III. Results

Standard full topographical and spectral analyses of the data from each user showed that EEG control was sharply focused over sensorimotor cortex and in mu and beta frequency bands. Fig. 2 illustrates this control with data from Users B and E. Such sharp topographical and spectral localization is typical of sensorimotor rhythm control [19], and distinguishes it from non-EEG artifacts, such as EMG (which has a much broader frequency distribution and is prominent over peripheral scalp areas) or EOG (which is prominent at lower frequencies and near the forehead) [10], [20].

A. Offline Analysis

We analyzed the data of the final five consecutive sessions from each of the seven R^2 was based on the linear regression between vertical target position and the features in question. Table II shows the r^2 values for the best single feature, the best two features, the best two features and their interaction, and all four features and four interactions (i.e., the two-way interactions between right and left mu, right and left beta, right mu and beta, and left mu and beta). For all users, r^2 values increase from left to right: r^2 is lowest for the single best feature, larger for the best two features, still larger for the two best features and their interaction, and largest for all four features and four interactions. For the overall average, each of these increases in r^2 was significant (at least $p < 0.05$ for all comparisons by t-test). An analysis of variance with the individual r^2 values as the dependent variable and complexity of the statistical model as the independent variable was significant ($\mathrm{df} = 3/27$, $F = 17.16$, $p < 0.0001$). Subsequent post-hoc tests (Tukey's) indicated that the single predictor model was significantly below all other models ($p < 0.05$), and the model with all eight variables was significantly better than the best pair ($p < 0.05$). At the same time, the users differed substantially in regard to which change improved r^2 most. For User B, addition of a second feature increased r^2 slightly (0.64–0.67), while inclusion of the interaction increased it considerably more (0.67–0.77). In contrast, for User F, addition of a second feature greatly increased r^2 (0.19–0.49), while inclusion of the interaction produced only a tiny increase (0.489–0.491).

It is of interest to note that while User E's left beta rhythm has a much higher r^2 value than his right mu rhythm (Fig. 2), the best feature pair (Table II) was left mu and right mu, rather than left mu and left beta. The best pair is not necessarily the combination of the two best single features.

We have previously shown that r^2 is strongly correlated with actual performance, i.e., the percent of targets hit [18]. (With four targets as used in this study, accuracy in the absence of any user control (i.e., chance accuracy) would be 25%.) From the relationship observed in the present data between the r^2 values of the features used online and the online accuracies, we calculated that the best single feature would produce an average accuracy of 60% (range 49%–73% across the seven users) and the best pair and their interaction would produce an average accuracy of 68% (range 58%–82%). The corresponding bit rates would be 4.5 bits/min (range 2.0–8.0) and 6.7 bits/min (range 3.9–11.2) [7], [21]. Thus, the analysis suggests that using the weighted combination of the best pair of features and their interaction, instead of the best single feature, would improve BCI information transfer rate by about 50%.

B. Generalization of Adaptive Feature Weights to New Data

The impressive r^2 improvements seen in Table II could conceivably reflect overfitting of the data (i.e., the weights might reflect unique aspects of the data set that would not be present in subsequent data sets). If this were true, the weights in Table II would not provide similarly impressive results when applied to a new data set, and adaptive feature weights would probably not be valuable in actual online operation of the BCI system. To assess this question, we collected a second data set (i.e., five more sessions) from Users B and E, applied to this second data set the User-B and User-E weights in Table II (i.e., the weights calculated from the first data set), and compared the resulting r^2 values to those provided by using new weights calculated from the second data set.

Table III compares the r^2 values for the second data set obtained by calculating new feature weights from the second data set to the r^2 values for the second data set obtained by using the old feature weights calculated from the first data set. For both users, the r^2 values given by the old feature weights are equal or nearly equal to those given by the new feature weights. The actual weights are also nearly equal. For example, with the best pair and interaction for user B, weights for the second and third set were -0.420 and -0.399 for mu, -0.570 and -0.599 for beta, and 0.097 and 0.095 for their interaction. These results suggest that adaptive feature weights generalize well to new data, and thus should be useful for actual online BCI operation.

C. Online Assessment of Adaptive Feature Weights

Finally, we assessed how well the weights worked in actual online application. We collected eight further data sessions from Users B and E. In the sessions of this third data set, every other

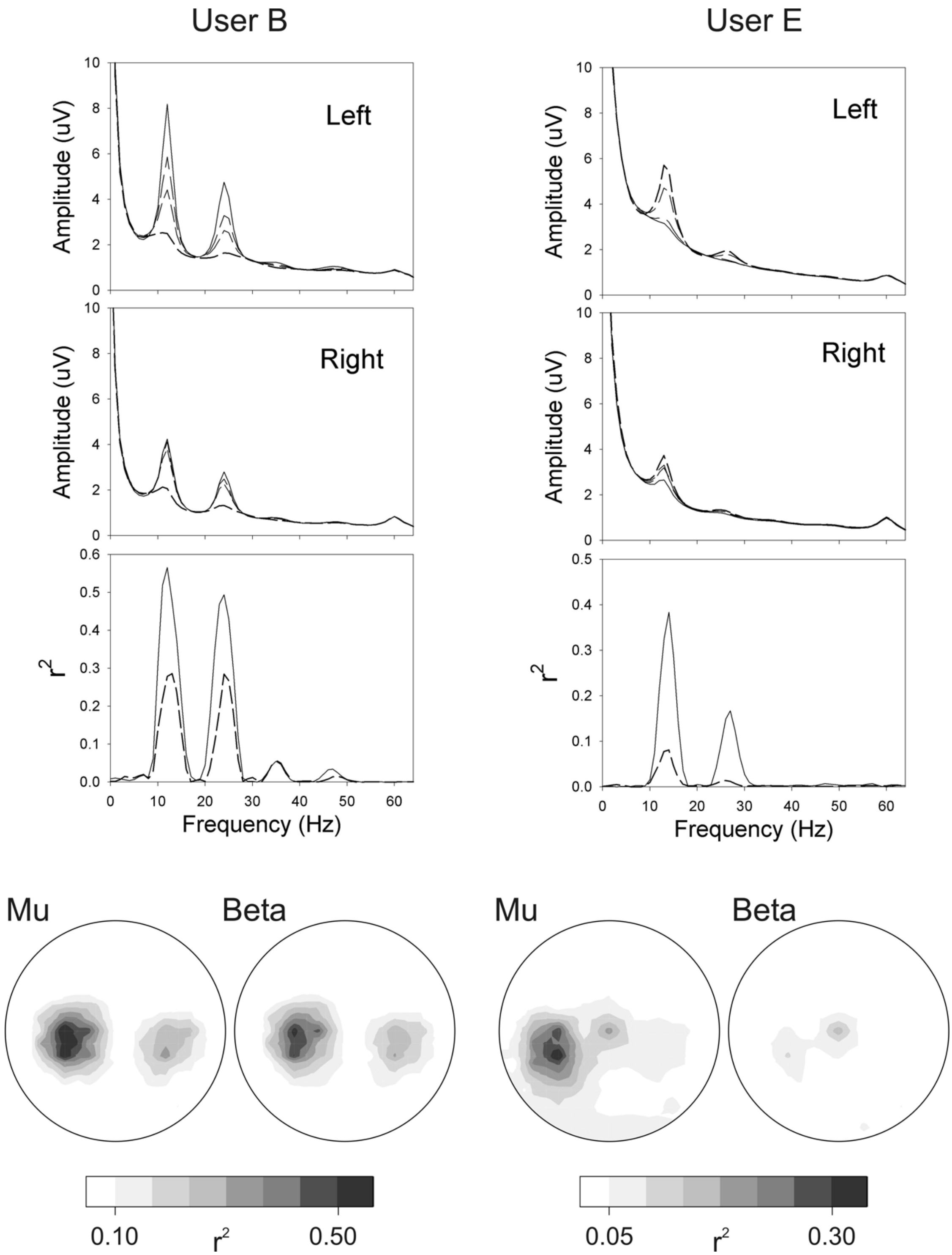

Fig. 2. Voltage and r^2 spectra, and r^2 topographies (nose at top) in the frequency band used for cursor control, for two users (B and E). In the amplitude graphs, the spectra for top, top-middle, bottom-middle, and bottom targets are solid, large dashed, medium-dashed, and small dashed, respectively. In the r^2 graphs, the spectra from left and right sides are solid and dashed, respectively. Polarity of the control signal was inverted for user E. Both users show control that is sharply focused spectrally in mu and beta frequency bands and sharply focused topographically over sensorimotor cortex.

3-min run used the best single feature to control cursor movement, while the intervening 3-min runs used the best pair and their interaction to control cursor movement. The best single feature and the best pair of features were those identified by analysis of the second data set, and the best pair and their interaction were weighted based on the second data set.

Table IV compares r^2 values for the third data set obtained using feature weights calculated from the second data set to the r^2 values for the second data set obtained using these same weights and to the r^2 values for the third data set obtained using r^2 values calculated offline from the third data set. As noted above, the weights calculated from the second

TABLE II

EACH USER'S VALUES OF r^2 FOR THE BEST SINGLE FEATURE, THE BEST TWO FEATURES, THE BEST TWO FEATURES AND THEIR INTERACTION, AND ALL FOUR FEATURES AND FOUR TWO-WAY INTERACTIONS. FOUR FEATURES WERE mu AND beta RHYTHM AMPLITUDES OVER LEFT (L) AND RIGHT (R) SENSORIMOTOR CORTICES. TWO-WAY INTERACTIONS BETWEEN L AND R FOR THE SAME RHYTHM AND BETWEEN mu AND beta RHYTHMS ON THE SAME SIDE WERE INCLUDED. IN THIS ANALYSIS, THE WEIGHTS OF THE FEATURES AND THEIR INTERACTIONS WERE CALCULATED BY A LINEAR MULTIPLE REGRESSION ANALYSIS (EXCEPT FOR THE SINGLE BEST FEATURE, WHICH SIMPLY HAD A WEIGHT OF 1.0)

User	Online Accuracy (%)	Best Feature	Best Feature r^2	Best Pair	Best Pair r^2	Best Pair & Interaction r^2	All 8 Features r^2
A	81	L mu	0.61	L&R mu	0.77	0.80	0.82
B	75	L mu	0.64	L mu&beta	0.67	0.77	0.78
C	66	L mu	0.40	L mu&beta	0.41	0.47	0.49
D	70	L mu	0.35	L&R mu	0.48	0.53	0.55
E	57	L mu	0.31	L&R mu	0.32	0.36	0.39
F	54	R mu	0.19	L&R mu	0.49	0.49	0.59
G	58	L mu	0.17	L&R mu	0.38	0.39	0.52
Mean (ÅSE)	66 4		0.38 0.07		0.50 0.06	0.54 0.07	0.59 0.06

TABLE III

VALUES OF r^2 FOR THE SECOND DATA SETS OF USERS B AND E FROM MULTIPLE REGRESSION ANALYSIS USING FEATURE WEIGHTS CALCULATED FROM THE SECOND DATA SET OR FROM THE USER'S FIRST DATA SET. FEATURES AND INTERACTIONS ARE AS DESCRIBED FOR TABLE II. THE OLD FEATURE WEIGHTS GENERALIZE TO THE NEW DATA: THEY GIVE r^2 VALUES EQUAL TO OR NEARLY EQUAL TO THOSE GIVEN BY NEW FEATURE WEIGHTS BASED ON THE NEW DATA

User	Feature Weights Calculated From	Best Feature (r^2)	Best Pair (r^2)	Best Pair & Interaction (r^2)	All 8 Features (r^2)
B	Second data set	0.65	0.67	0.77	0.79
	First data set	0.65	0.66	0.77	0.78
E	Second data set	0.23	0.26	0.27	0.32
	First data set	0.23	0.25	0.27	0.29

TABLE IV

VALUES OF r^2 FOR THE SECOND DATA SETS OF USERS B AND E USING FEATURE WEIGHTS CALCULATED FROM THE SECOND DATA SET, AND FOR THE THIRD DATA SETS USING FEATURE WEIGHTS CALCULATED FROM THE SECOND OR THIRD DATA SETS. WEIGHTS OF THE SECOND DATA SET WERE USED ONLINE TO CONTROL CURSOR MOVEMENT DURING COLLECTION OF THE THIRD DATA SET. FEATURES AND INTERACTIONS ARE AS DESCRIBED FOR TABLE II. WHEN ACTUALLY USED ONLINE, THE OLD FEATURE WEIGHTS GIVE r^2 VALUES EQUAL TO OR NEARLY EQUAL TO THOSE THEY GAVE OFFLINE FOR THE SECOND DATA SET OR TO THOSE CALCULATED OFFLINE FOR THE THIRD SET

User	Data Set	Feature Weights Calculated From	Best Feature (r^2)	Best Pair (r^2)	Best Pair & Interaction (r^2)	All 8 Features (r^2)
B	Second	Second data set	0.65	0.67	0.77	0.79
	Third	Second data set	0.60	0.64	0.72	0.73
	Third	Third data set	0.60	0.64	0.72	0.73
E	Second	Second data set	0.23	0.26	0.27	0.32
	Third	Second data set	0.27	0.28	0.29	0.33
	Third	Third data set	0.27	0.29	0.31	0.35

data set were used online to control cursor movement during collection of the third data set. When actually used online for collection of the third data set, the feature weights calculated from the second data set gave r^2 values equal to or nearly equal to those they gave offline for the second data set (i.e., the data set from which they were derived) and to those given by feature weights calculated offline from the third data set. These results further support the conclusion that adaptive feature weights generalize well to new data, and can improve actual online BCI operation.

As expected from previous studies [18], the higher r^2 values provided by the best pair and their interaction was accompanied by better accuracy (i.e., higher percentage of targets hit). User B had accuracies of 75% and 79% with the best single feature and with the best pair with interaction, respectively, and User E had accuracies of 51% and 56%, respectively.

IV. Discussion

A. Adaptive Multiple Regression for BCI Operation

In the present study, both offline analyses and subsequent online testing indicated that BCI performance can be improved by using adaptive multiple linear regression to combine and weight multiple features. Furthermore, the inclusion of feature interactions, which reflect nonlinear effects, further improves performance.

First, offline analyses of the first data set showed that adaptive multiple regression substantially improved r^2 values derived from the same data. These increases predicted correspondingly impressive improvements in accuracy. Second, offline analyses of a second data set showed that the feature weights calculated from the first data set generalized very well; they provided r^2 values equal to or nearly equal to those provided by feature weights derived from the second data set itself. This indicated that their impressive performance on the first data set was not merely a result of overfitting chance variation in the data. Third, actual online application of the feature weights calculated from the second data set showed that adaptive multiple regression improved online performance as expected from offline analysis. This result indicated that the features and feature weights calculated by offline analysis corresponded to what the user could actually control.

Prior to this study, our standard online protocol incorporated adaptive control of the intercept [a in (1)] and the slope (or gain) [b in (1)] of the linear equation that determines 1-D cursor movement. The intercept controls the overall vertical bias of cursor movement. Our translation algorithm continually calculates the intercept online using a weighted estimate of the mean of the features, based on past performance, so as to provide cursor movement that is not biased in one direction or another [11]. The slope controls the overall rate of cursor movement. In the standard four-alternative 1-D protocol, it controls the relative accessibility of the end (top and bottom) targets and the in-between targets. Low slopes favor the in-between targets while high slopes favor the end targets. The translation algorithm continually adjusts the slope online so as to make each target equally accessible [11]. The present study used adaptive feature weights to focus on those EEG features that the user is best able to control. The linear regression algorithm analyses each user's past data to select those weights that best predict target position, and then uses these weights online in subsequent BCI operation.

In theory, automatic online control of intercept, slope, and feature weights could be combined into a single process. However, it may be best to keep the control of intercept and slope separate from the control of feature weights. The physiological factors that influence intercept and slope (e.g., fluctuations in attention, fatigue, etc.) may differ from those that influence feature weights, and they may vary over different time scales. For example, fluctuations in overall feature amplitude (which determines the intercept) may be related to momentary differences in arousal and alertness, while changes in the optimal feature weights may be related to slower processes associated with gradual acquisition of the skill of EEG feature control. Consequently, intercept adaptation and feature weight adaptation probably need different time courses. Furthermore, they depend on different outcome measures (i.e., intercept adaption minimizes bias in predicting targets while feature weight adaptation minimizes overall target prediction error).

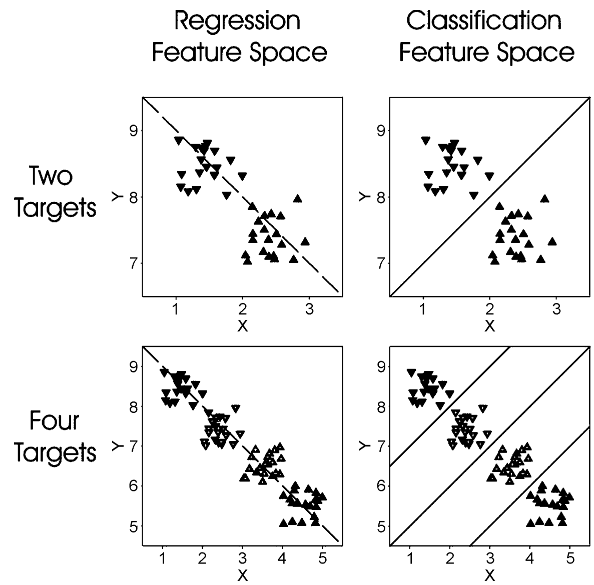

Fig. 3. Regression and classification approaches to BCI control of two-target and four-target applications. For the two-target application (targets are up and down triangles), both approaches must determine the parameters of a single function. In contrast, for the four-target application (targets are up and down closed and open triangles), the regression approach still needs only a single function while the classification approach needs three functions, one for each inter-target boundary.

It is important to note that the users who provided the data of Tables III and IV were well-trained. Data from less experienced users who are in the process of learning cursor control might produce different results. Finally, it should also be noted that for two-dimensional cursor control, adaptation of feature weights can serve an additional purpose: it can be used to decorrelate the signals controlling horizontal and vertical movements, respectively, so as to provide two independent control channels [8], [22].

Several groups have discussed the interaction between user and system. One approach relies on the capacity of machines to learn the subject's natural cerebral motor commands [23]. We have proposed that the mutual adaptation of system and user is a necessary feature of successful BCI operation [7]. The Graz group [24] found that fast adaptation of parameters during training was not necessary. They suggest that a classifier could be updated at the beginning of each session. This issue of how to adapt and at what rate is complex and will require more investigation.

B. Regression Versus Classification for BCI Translation Algorithms

The BCI translation algorithm used in our previous studies [3], [8], [13], [25], and further developed in this new study, employs a regression (or prediction) approach to determine which target the user wants to select. Features extracted from the EEG are used as independent variables to predict the location of the target. Other research groups [26], [27] (and ourselves as well [28]) have explored a classification (or discriminant) approach. Features extracted from the EEG are used as independent variables to define boundaries between the different targets in feature space.

Fig. 3 illustrates these two alternative approaches for two-target and four-target applications. For the two-target case, both the regression approach and the classification approach require that the parameters of a single function be determined. In the four-target case, the regression approach still requires only a single function (assuming that the targets are distributed along a single dimension (e.g., vertical position on the screen). In contrast, for the four-target case the classification approach requires that three functions be determined, one for each of the three boundaries between the four targets.

Thus, the classification approach might be most useful for two-target applications, such as a P300-based BCI [29]. The regression approach may be preferable for greater numbers of targets when these targets can be ordered along one or more dimensions. For example, the icons on a computer screen can be described in terms of two (horizontal and vertical) dimensions. In addition, the regression approach generalizes more readily to different numbers of targets. For example, the same regression equation derived for a four-target application could be applied to a six-target application. In contrast, with the classification approach, the six-target application would require different and additional boundaries. Furthermore, the regression approach provides a convenient source of continuous feedback in applications such as ours which involves continuous control of cursor movement.

Linear regression or linear classification approaches may not work well in cases in which linear prediction or separation is not sufficient. In response to this problem, some BCI researchers have explored nonlinear approaches such as neural networks [30], [31]. Although these approaches are more complex, such nonlinear methods may produce better results in some cases [30]. Explicit comparisons of these methods [32] is particularly useful. The present study addressed the problem of nonlinear effects by including feature interaction terms in the adaptive multiple regression. This strategy proved to be effective (e.g., Table II).

V. Conclusion

This study shows that adaptive multiple regression can markedly improve the performance of a sensorimotor rhythm-based BCI. Both offline analyses and subsequent online evaluation yielded results substantially superior to those obtained without adaptation. Feature weights obtained from previous data generalized well to new data sets and to online control. The adaptive multivariate regression approach appears to be a highly flexible and efficient method for solving the problem of translating a user's control of EEG features into accurate control of an output such as cursor movement.

Acknowledgment

The authors would like to thank G. Schalk and T. M. Vaughan for valuable comments on the manuscript.

References

[1] L. A. 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.
[2] E. E. Sutter, "The brain response interface: Communication through visually guided electrical brain responses," *J. Microcomput. Appl.*, vol. 15, pp. 31–45, 1992.
[3] J. R. Wolpaw, D. J. McFarland, G. W. Neat, and C. A. Forneris, "An EEG-based brain-computer interface for cursor control," *Electroencephalogr. Clin. Neurophysiol.*, vol. 78, pp. 252–259, 1991.
[4] N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kubler, J. Perlmouter, E. Taub, and H. Flor, "A spelling device for the paralyzed," *Nature*, vol. 398, pp. 297–298, 1999.
[5] G. Pfurtscheller, D. Flotzinger, and J. Kalcher, "Brain-computer interface—a new communication device for handicapped persons," *J. Microcomput. Appl.*, vol. 16, pp. 293–299, 1993.
[6] A. Kostov and M. Polak, "Parallel man–machine training in development of EEG-based cursor control," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 203–205, Jun. 2000.
[7] 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.
[8] J. R. Wolpaw and D. J. McFarland, "Multichannel EEG-based brain-computer communication," *Electroencephalogr. Clin. Neurophysiol.*, vol. 90, pp. 444–449, 1994.
[9] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: A general-purpose brain-computer interface (BCI) system," *IEEE Trans. Biomed. Eng.*, vol. 51, no. , pp. 1034–1043, Jun. 2004.
[10] D. J. McFarland, T. Lefkowicz, and J. R. Wolpaw, "Design and operation of an EEG-based brain-computer interface (BCI) with digital signal processing technology," *Behav. Res. Meth. Instrum. Comput.*, vol. 29, pp. 337–345, 1997.
[11] D. J. McFarland and J. R. Wolpaw, "EEG-based communication and control: Speed-accuracy relationships," *Appl. Psychophysiol. Biofeedback*, vol. 28, pp. 217–231, 2003.
[12] D. J. McFarland, W. A. Sarnacki, and J. R. Wolpaw, "Brain-computer interface (BCI) operation: optimizing information transfer rates," *Biol. Psychol.*, vol. 63, pp. 237–251, 2003.
[13] J. R. Wolpaw, D. J. McFarland, T. M. Vaughan, and G. Schalk, "The Wadsworth Center brain-computer interface (BCI) research and development program," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 204–207, Jun. 2003.
[14] F. Sharbrough, C. E. Chatrian, R. P. Lesser, H. Luders, 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] S. L. Marple, *Digital Spectral Analysis With Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
[17] SAS Institute Inc., , Cary, NC.
[18] H. Sheikh, D. J. McFarland, W. A. Sarnacki, and J. R. Wolpaw, "EEG-based communication: Characterizing EEG control and performance relationship," *Neurosci. Lett.*, vol. 345, pp. 89–92, 2003.
[19] J. R. Wolpaw, D. J. McFarland, and T. M. Vaughan, "Brain-computer interface research at the Wadsworth center," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 222–226, Jun. 2000.
[20] I. I. Goncharova, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "EMG contamination of EEG: spectral and topographical characteristics," *Clin. Neurophysiol.*, vol. 114, pp. 1580–1593, 2003.
[21] J. R. Pierce, *An Introduction to Information Theory*. New York: Dover, 1980.
[22] J. R. Wolpaw and D. J. McFarland, "Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans," *Proc. Nat. Acad. Sci.*, vol. 101, pp. 17 849–17 854, 2004.

[23] B. Blankertz, G. Dornhege, C. Schafer, R. Krepki, J. Kohlmorgen, K. R. Muller, 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 Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 100–104, Jun. 2005.
[24] G. Krausz, R. Scherer, G. Korisek, and G. Pfurtscheller, “Critical decision-speed and information transfer in the Graz brain-computer interface,” *Appl. Psychophysiol. Biofeedback*, vol. 28, pp. 233–240.
[25] D. J. McFarland, G. W. Neat, R. F. Read, and J. R. Wolpaw, “An EEG-based method for graded cursor control,” *Psychobiol.*, vol. 21, pp. 77–81, 1993.
[26] B. Blankertz, G. Dornhege, C. Schafer, R. Krepki, J. Kohlmorgen, K. R. Muller, 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 Syst. Rehabil. Eng.*, vol. 11, no. 6, pp. 127–131, Jun. 2003.
[27] F. Cincotti, D. Mattia, C. Babiloni, F. Carducci, S. Salinari, L. Bianchi, M. G. Marciani, and F. Babiloni, “The use of EEG modifications due to motor imagery for brain-computer interfaces,” *IEEE Trans. Neural Systems Rehabil. Eng.*, vol. 11, no. 2, pp. 131–133, Jun. 2003.
[28] G. E. Fabiani, D. J. McFarland, J. R. Wolpaw, and G. Pfurtscheller, “Conversion of EEG activity into cursor movement by a Brain-Computer Interface (BCI),” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 12, no. 3, pp. 331–338, Sep. 2004.
[29] E. Donchin, K. M. Spencer, and R. Wijesinghe, “The mental prostheses: Assessing the speed of a P300-based brain-computer interface,” *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 164–179, Jun. 2000.
[30] K. R. Muller, C. W. Anderson, and G. E. Birch, “Linear and nonlinear methods for brain-computer interfaces,” *IEEE Trans. Neural Syst. Rehabil.*, vol. 11, no. 2, pp. 165–169, Jun. 2003.
[31] B. O. Peters, G. Pfurtscheller, and H. Flyvbjerg, “Mining multi-channel EEG for its information content: An ANN-based method for a brain-computer interface,” *Neural Netw.*, vol. 11, pp. 1429–1433, 1998.
[32] D. Garrett, D. A. Peterson, C. W. Anderson, and M. H. Thaut, “Comparison of linear, nonlinear, and feature selection methods for EEG signal classification,” *IEEE Trans. Neural Syst. Rehabil.*, vol. 11, no. 2, pp. 141–144, Jun. 2003.

Dennis J. McFarland received the B.S. and Ph.D. degrees from the University of Kentucky, Lexington, in 1971 and 1978, respectively.

He is currently a Research Scientist with the Wadsworth Center for Laboratories and Research, New York State Department of Health, Albany. His research interests include the development of EEG-based communication and auditory perception.

Jonathan R. Wolpaw received the A.B. degree from Amherst College, Amherst, MA, in 1966 and the M.D. degree from Case Western Reserve University, Cleveland, OH, in 1970. He completed a residency in neurology at the University of Vermont, Burlington, and fellowship training in neurophysiology research at the National Institute of Health (NIH), Washington, DC.

He is currently the Chief of the Laboratory of Nervous System Disorders and a Professor at the Wadsworth Center, New York State Department of Health and the State University of New York, Albany. His research interests include operant conditioning of spinal reflexes as a new model for defining the plasticity underlying a simple form of learning in vertebrates, and development of EEG-based communication technology for those with severe motor disabilities. His research has been funded for many years by the NIH and many private foundations, and has received extensive recognition including a number of important awards.