Continuous EEG Classification During Motor Imagery—Simulation of an Asynchronous BCI

George Townsend, Bernhard Graimann, and Gert Pfurtscheller, *Member, IEEE*

***Abstract*—Nearly all electroencephalogram (EEG)-based brain–computer interface (BCI) systems operate in a cue-paced or synchronous mode. This means that the onset of mental activity (thought) is externally-paced and the EEG has to be analyzed in predefined time windows. In the near future, BCI systems that allow the user to intend a specific mental pattern whenever she/he wishes to produce such patterns will also become important. An asynchronous BCI is characterized by continuous analyzing and classification of EEG data. Therefore, it is important to maximize the hits (true positive rate) during an intended mental task and to minimize the false positive detections in the resting or idling state. EEG data recorded during right/left motor imagery is used to simulate an asynchronous BCI. To optimize the classification results, a refractory period and a dwell time are introduced.**

***Index Terms*—Asynchronous brain–computer interface (BCI), BCI, receiver operating characteristic (ROC), synchronous BCI, true-false (TF) differences.**

I. Introduction

BRAIN–COMPUTER interfaces (BCIs) can be divided into systems working in synchronous and asynchronous modes. In a synchronous BCI, the analysis and classification of brain potentials is limited to predefined fixed or variable time windows. BCI systems based on the analyses of evoked potentials [1] or slow cortical potentials [2] belong, in general, to this class. The BCI systems developed in Albany [3] and Graz [4] that analyze spontaneous EEG are also of the synchronous type. A characteristic feature of all of these systems is that the onset of mental activity is known in advance and associated with a specific cue or trigger stimulus. In the case of an ideal asynchronous BCI system, no cue stimulus is used, and the subject can intend whenever she/he wishes, a specific mental activity. The ongoing brain signals have to be analyzed and classified continuously. Mental events have to be detected and discriminated from noise and nonevents and transformed into a control signal as quickly and accurately as possible.

In training and feedback sessions with the synchronous Graz BCI, continuous electroencephalogram (EEG) data is stored, but only time periods of 4 s starting with the cue stimulus are processed [4]. The EEG outside the 4-s periods is characterized by unknown mental activity occurring during a resting or idling state. The subjects also perform blinks and other eye movements during these periods. Such artifacts are unavoidable, however, trained subjects are able to postpone such actions to the resting period. Therein lies one possible advantage of a synchronous approach and conversely a disadvantage or at least a difficulty in an asynchronous approach.

One goal of this paper is to accumulate knowledge for designing an asynchronous BCI by analyzing continuous EEG data consisting of defined cue-triggered mental states and rest or idling states with eye movement artifacts in between. Another goal is to use receiver operating characteristics (ROC) curves to define upper and lower thresholds and thereby to introduce a "zero class," that is a class which characterizes samples or periods not belonging to an intended activity, which are referred to as "resting periods."

Although there is existing literature in which asynchronous BCI performance is evaluated using ROC curves [5], only real movement execution was considered. This paper looks for the first time at the application of ROC curves for evaluating the performance of an asynchronous BCI during imagined movements. Special processing of the BCI output incorporating a dwell and refractory period is also introduced. This processing defines a minimum time that the detector must be above its threshold before declaring a positive output, and a refractory period to suppress the detector for a particular time once a positive is declared. This technique improves the performance of the detector by reducing false positives (FPs).

An important issue discussed here is that great care must be taken in the interpretation of results in an asynchronous environment. In such an environment, we need to evaluate the performance of the detector during both events and nonevents. An excellent detection method may appear to perform dismally when subjected to certain analyses. For example, a detector that generates a correct output 99% of the time based on a continuous evaluation against the correct behavior could potentially, in theory, fail miserably when subjected to an event-by-event analysis. This can happen when the detector produces a small proportion of FPs, but consistently produces at least one during a nonevent. This causes certain more stringent criteria to dismiss the detector as being incorrect for the entire nonevent despite its correctness during the greater part of such periods. Some researchers circumnavigate this problem by allowing the detector to dictate the number of events [6], but it was felt that the definition of an event and of a nonevent should be independent of the behavior of the detector being evaluated. (Despite this criticism, an analysis using a similar approach is also included here. A cautious explanation, however, is included in Section II-E to justify under what circumstances it might be appropriate to resort to such an approach.) It is important for the reader to recognize this when evaluating the results, and, in particular, when

Manuscript received January 17, 2003; revised July 1, 2003 and January 23, 2004.

G. Townsend is with the Department of Computer Science, Algoma University, Sault Ste. Marie, ON P6A 2G4, Canada (e-mail: townsend@auc.ca).

B. Graimann and G. Pfurtscheller are with the Department of Medical Informatics, Institute of Biomedical Engineering, University of Technology Graz, 8012 Graz, Austria.

Digital Object Identifier 10.1109/TNSRE.2004.827220

attempting to understand the significant differences in the performances of the same detector under the different evaluation criteria presented here.

A careful discussion is also given to explain the circumstances under which ROC curves may depart from the ideal smooth and continuous curves extending from corner to corner of the graph. Such ideal behavior requires that particular relationships exist between the rates of changes in true and false positives and negatives with respect to changes in the tested thresholds. As will be explained, not all detection methods will lead to such well-behaved curves.

The EEG data available for analysis consists of 4-s periods with right- or left-hand motor imagery (events) and resting periods of variable length. For the preprocessing of 27 EEG-channels, the method of common spatial patterns (CSP) is used [7]. With the ROC curves two thresholds are determined and based on this, three classes are defined: periods of right imagery, periods of left imagery, and periods of rest. Data from three able-bodied subjects is available for this study.

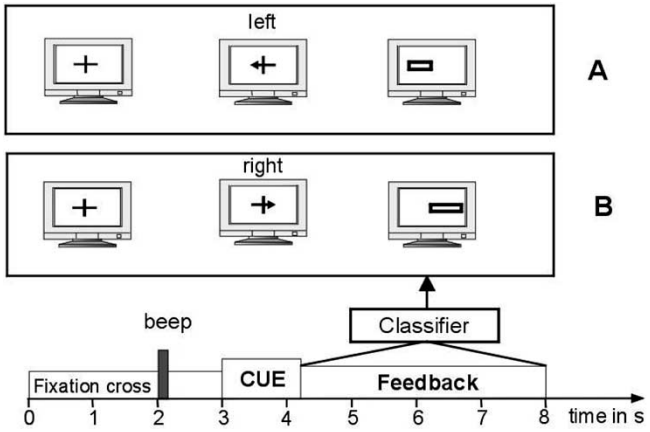

Fig. 1. Timing of one trial of the experiment with feedback. From second 3 until 4.25, a cue arrow pointing left or right appeared. The subject imagined a left- or right-hand movement, as per the arrow direction. Between second 4.25 and 8, the EEG was classified online and the result produced a horizontal bar appearing in center screen as feedback. The bar, varying in length, extends in the appropriate direction of the imagined movement as shown in A and B. The subject's task was to extend the bar toward the appropriate boundary of the screen.

II. Procedure

A. EEG Recording and Experimental Paradigm

Three male students (g3, g7, i2), all familiar with the BCI, participated in the experiments. Twenty-seven EEG electrodes (used to overlay the whole primary sensorimotor cortex), equally spaced with approximately 2.5-cm distance, were placed in three rows (nine columns) and referenced to the right ear. A referential recording was selected because a classification accuracy similar to other referencing methods (common average reference, bipolar, large Laplacian, and small Laplacian) can be achieved [7] and does not require additional processing time for rereferencing. The ground electrode was located on the forehead.

The amplified EEG was bandpass filtered between 0.5 and 50 Hz and sampled at 128 Hz. The resolution was 12 b. A notch filter was used to suppress the 50-Hz power line interference. Each experimental procedure consisted either of sessions without feedback or with feedback. For the subjects and sessions chosen for this study, all were feedback sessions except for g7. A fixation cross is shown on screen up to second 2, when a warning tone sounds. Next, a cue appears from second 3 to 4.25 indicating either a left- or right-hand movement imagination. From second 4.25 to 8 the subject has to imagine the respective hand movement. Each trial for the synchronous processing lasted 8 s. The time between trials was randomized in a range from 0.5 to 2.5 s to avoid adaptation. The timing of the nonfeedback sessions was the same as the feedback sessions except that the feedback bar was not shown on the screen. Fig. 1 depicts this paradigm.

Each session was divided into four experimental runs of 40 trials each, with randomized directions of the cues (20 left and 20 right).

Two subjects participated in six and one subject (i2) in seven sessions. Two sessions were without feedback and four with feedback. In these sessions, four common spatial filters (CSP) and five weight vectors (WV) were set up. Sessions were performed within three days (for more details, see [8]). For this study, two representative sessions from previous work showing a range of performances made by each of subjects g3, g7, and i2 were selected for analysis. In each case, the earlier of the two sessions was used to calculate the common spatial pattern filters and the weight vector. These parameters were applied to the later session to provide the simulated asynchronous analysis of the later of the two sessions.

B. Data Preprocessing

Prior to the calculations of spatial filters, all EEG channels are filtered (FIR filter) between 8–30 Hz, because this broad frequency range contains all mu and beta frequency components of the EEG which are important for the discrimination task [7]. It was shown that an increase of the classification accuracy can be achieved by using this broad range in comparison to narrow bands [9].

The method presented here uses the covariance between channels to design common spatial patterns and is based on the simultaneous diagonalization of two covariance matrices [10]. For the calculation, artifact-free 4-s epochs recorded during motor imagery (second 4–8) are used. The recordings were visually inspected to remove any trials containing artifacts. The decomposition (or filtering) of the EEG leads to new time series, which are optimal for the discrimination of two populations. The patterns are designed such that the spatially filtered EEG has maximum variance for left trials and minimum variance for right trials. In this way, the difference between left and right populations is maximized and the only information contained in these patterns is where the variance of the EEG varies most when comparing two conditions. More details in applying common spatial patterns to BCI are described in [7] and [8].

In the Graz-BCI, a SIMULINK model (The Mathworks Inc.) is used to simulate the preprocessing and classification during feedback sessions to provide real-time visual feedback to the subjects online. In this paper, the behavior of this SIMULINK model was emulated in MATLAB to provide the simulated asynchronous analysis discussed here. The function of the model is as follows: 27 EEG channels are bandpass filtered

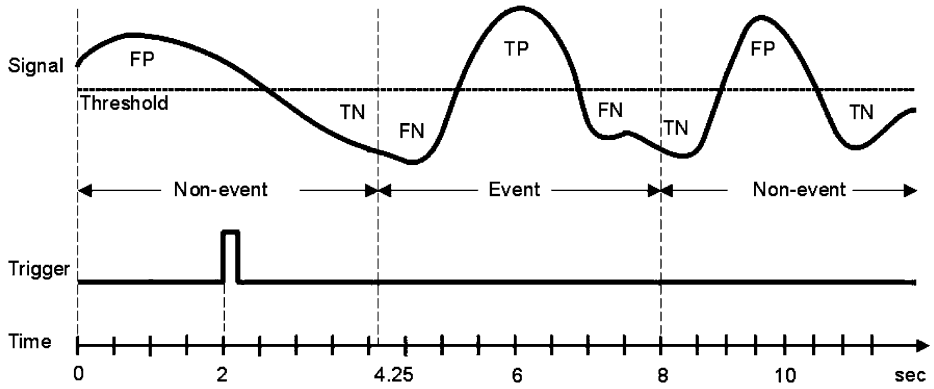

Fig. 2. Sample-by-sample evaluation of true and false positives (TPs and FPs, respectively) and true and false negatives (TNs and FNs, respectively). This describes left imagery only. Right imagery involves a lower threshold that the signal must drop below to indicate a detection. Events are defined as the period of time from 2.25 to 6.0 s after the rising edge of the trigger, which occurs 2.0 s after the start of each trial.

between 8 and 30 Hz. Output signals are then passed to the two most discriminating (1 and 27) and two second most discriminating (2 and 26) common spatial filters [7], [8]. After temporal and spatial filtering, the variances of the resulting four time series (VARp) were calculated sample-by-sample using a 1-s window, then normalized and also log-transformed: [logn(VARp/sum(VARp), $p = 1, \ldots, 4$)]. The four-dimensional feature vector is linearly combined by the weights of a former linear discriminant analysis (calculated from a former session) to form a one-dimensional feature vector that is then used for ROC analysis. A detailed description of hardware and software components is given elsewhere [8], [11].

C. Sample-by-Sample Analysis

In the continuous output signal of the SIMULINK model used in the Graz-BCI with a duration of about 1380 s simulating the preprocessing by spatial filtering with CSPs and linearly combining the feature vectors, left motor imagery is captured as a positive signal and right motor imagery as a negative signal. In this paper, this same signal is emulated off-line using MATLAB to provide the detector output used in ROC analysis. In this analysis, right and left motor imagery is tested separately. In one analysis right motor imagery (events) is compared with left imagery and resting periods (nonevents). In the other analyses, left motor imagery (events) is compared with right imagery and resting periods (nonevents). In the sample-by-sample analysis, the two axes of the ROC curves are true positive rate (TPR) and false positive rate (FPR). The former is a measure of sensitivity, while the later is a measure of selectivity. These quantities are captured by the following:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{1}$$

$$\text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}} \tag{2}$$

where TP, FN, TN, and FP are the number of true positives, false negatives, true negatives, and false positives, respectively. Note that all these values are counted in samples. Fig. 2 illustrates these parameters.

The area below an ROC curve varies between 0 and 1 and gives a measure about the separability of two classes with an area of 1 for complete separation [12].

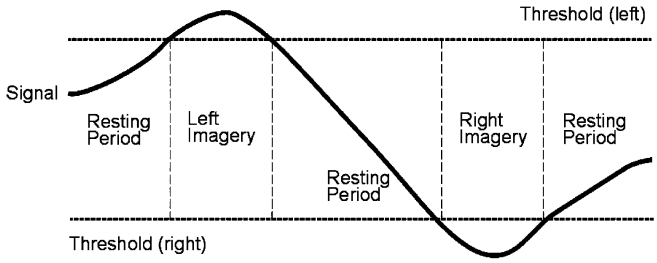

Fig. 3. Interpretation of the Graz-BCI signal. During left imagery, the signal goes more positive, while right imagery will drive the signal more negative. Upper and lower thresholds define the sense of the imagery (right or left) as well as resting periods (neither left nor right imagery).

In the ROC curves appearing in this paper, the left and right imagery analyses are combined with the labeled abscissa and ordinate values applying only to the right analysis. The left analysis is understood to use labels of 1-x and 1-y and, therefore, appears in the opposite corner of the graph for convenience. It should be noted that these curves are very symmetrical suggesting agreement between left and right results and giving evidence that the approach of combining resting periods with the opposite imagination to that being analyzed is a valid approach.

In the sample-by-sample basis, the relative values of the detector output and the actual status of the event are compared and give rise to all four of the possible measures mentioned earlier. From each ROC curve, a threshold corresponding to the point of the ROC curve closest to the line $y = 1 - x$ is selected as an indication of equal balance between true and FPs. As a result of these two thresholds, the three classes "right imagery," "left imagery," and "resting periods" are defined (Fig. 3).

D. Event-by-Event Analysis

An event-by-event analysis is undertaken with the goal of defining the number of true-positive classified events (TPE) and false-positive classified events (FPE) and also the true–false (TF)-difference defined as TPE minus FPE in percent for each simulated asynchronous BCI session. This event-by-event analysis is necessary to evaluate the performance of the BCI system. A valid ROC analysis requires that the number of events and nonevents are fixed rather than being determined by the number of detections. This is because if the detector is permitted to determine the number of events, there is no way of producing an

FN, since an FN is an event that went undetected by the detector. Since an ROC curve provides a measure of sensitivity versus selectivity, the inability to measure false negatives is clearly problematic, since it prevents a proper measure of selectivity from being evaluated.

It also becomes necessary to count multiple detections during an event as a single TP and, similarly, it becomes necessary to count multiple detections during a nonevent as a single FP for a valid ROC analysis. The reason for this is that if instead each detection was counted as a separate event, the illusion of many successful detections where in fact only one event occurred would skew the results. Similarly a single detection during a nonevent means that the entire nonevent was incorrectly recognized as a nonevent, and additional detections during such a period do not mean that additional nonevents occurred. Therefore, only one incorrect detection should be recorded. This results in a pass/fail approach, since even a single detection at any point in time during a nonevent results in the detector being declared incorrect for the entire time period of the nonevent.

When using a ROC curve as an evaluation tool, the area under the curve determines the performance of the detector. The criteria described earlier is far more stringent under this approach, than in the sample-by-sample approach and, therefore, the results may appear correspondingly poorer. Bear in mind that the same detector is evaluated in both cases and it is only the test criteria that has changed.

A modification of the event-by-event analysis is to introduce a "dwell time" and "refractory period." The first of these is the amount of time that the signal must cross the threshold to be considered a valid detection. Various dwell times were experimented with and from these results, 0.25 s was chosen for further analysis. The second parameter, the refractory period, is a period where the signal (once the "dwell time" has been met) will be ignored, whether above or below the threshold. Various refractory periods were tested, and from these values of 1.0, 1.25, and 1.35 s were analyzed further here. Fig. 4 should help to explain the use of dwell time and refractory periods. This method of processing the original signal will convert an output signal that is sustained above the threshold for some time into a train of spikes with a period equal to the sum of the dwell and refractory periods. If the original signal behaves ideally, however, the result of processing should produce a single spike for each event and no spikes for nonevents.

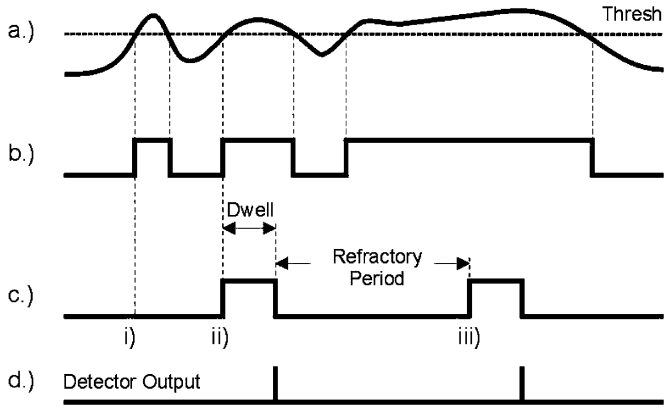

Fig. 4. Processing steps used to incorporate the dwell and refractory periods. (a) The original Graz-BCI output is compared to the threshold for the imagery in question. (b) The initial processing step consists of simple threshold detection (i). (c) The output is suppressed if the dwell time is not met or for the refractory period (iii) once dwell time is met (ii). (d) The output of step (c) is then converted into single spikes located where the dwell was met. Note: After the refractory period, the cycle may be repeated. This could potentially result in trains of spikes if the original signal remains above the threshold for some time.

E. TF Differences

The previous method of analysis was based on ROC curves and, therefore, required that multiple detections during a nonevent be counted as a single FP. It may not always be appropriate to do this because the period of a nonevent, or idle period, may be substantially longer than periods of imagery. It therefore becomes very significant how many false detections (i.e., FPs) occur during these time periods. To treat these as separate events prohibits a valid ROC analysis from being preformed. Therefore, a different method of analysis must be resorted to if this issue is significant.

To deal with such situations, we may use a system of true and FPs. Here, all detections during a nonevent are counted as FPs. A TP, however, is defined to be one or any number of detections during a particular event. Therefore, we will only count at most one TP per event. The total number of incorrect detections or FPs plus the total number of events then gives us a total number of "tests." That is, the total number of tests is then the sum of the number of events and FPs. Note that we do not count the number of nonevents in this system. As a result, the number of true and false positives is independent of the number of events and nonevents. This is in contrast to the methods discussed earlier where there was a relationship between all of these parameters that also involved true and false negatives. Under the scheme described here, care must be taken to recognize this subtle distinction. Much less subtle is the fact that there is no way to record or define a TN using the measure described here. For this reason, some caution should be exercised when resorting to this method, however when long resting periods make it unreasonable to declare the output of the detector invalid for the entire period due to a single false alarm, it would seem justifiable to adopt this approach as an alternative. Note that in this method of evaluation, the true rate corresponds precisely to the sensitivity measure as calculated in the event-by-event method, however the other parameter (false rate) differs from the selectivity measure. These distinctions prevent the parameters discussed here from being used to generate ROC curves. However, we may use these results to produce a TF difference, which is defined to be the difference in the ratio of TPs to events and FPs to tests. Ideally, the TF difference will be 100% as a result of one or more detections for each event and no detections during a nonevent. For a detection system with random behavior, these two parameters will both be 50% giving a TF difference of zero. The TF difference (TF%) is defined by the following:

$$\mathrm{TF}\% = \left(\frac{\mathrm{T}}{\mathrm{E}} - \frac{\mathrm{F}}{\mathrm{E}+\mathrm{F}}\right) \cdot 100 \tag{3}$$

where T, F, E are the number of TPs, FPs, and events, respectively. The minuend describes the TP ratio, the subtrahend the FP ratio.

The same range of thresholds were tested as before using a dwell time of 0.25 s and refractory times of 1.0, 1.25, and 1.35 s.

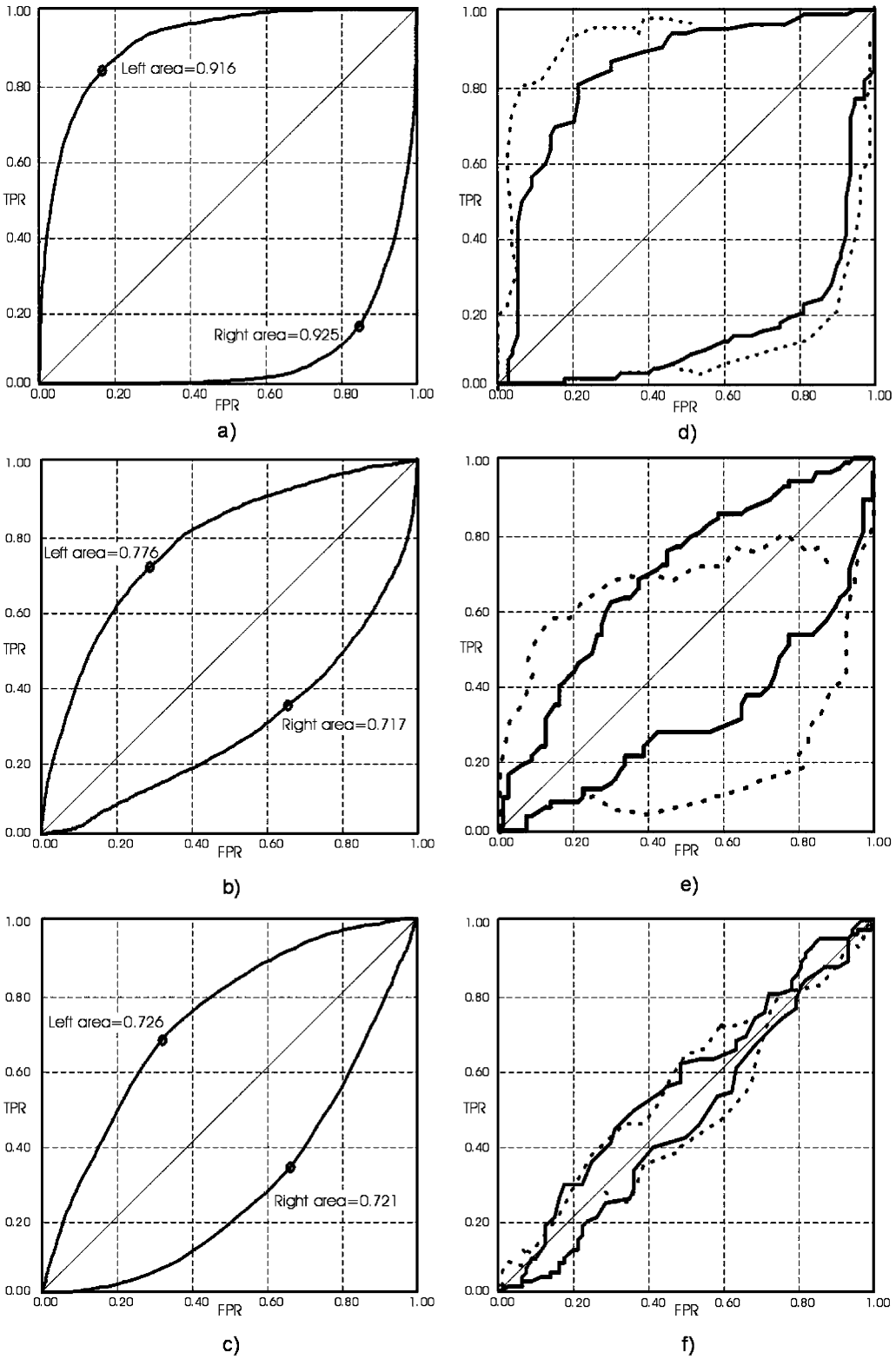

Fig. 5. (a)–(c) ROC curves generated on a sample-by-sample comparison of the classifier against the actual imagery. (d)–(f) ROC curves generated on an event-by-event basis. In the case of the later, the solid line was generated without the signal processing introduced for the dwell and refractory analysis. The dotted line is the result of the refractory period giving the best results. (a) Subject g3: sample by sample. (b) Subject g7: sample by sample. (c) Subject i1: sample by sample. (d) Subject g3: event by event. (e) Subject g7: event by event. (d) Subject i2: event by event.

III. Results

A. Sample-by-Sample Analysis

A few examples of ROC curves generated on the sample-by-sample basis using the simulated asynchronous output of the signal from the Graz-BCI are displayed in Fig. 5 (left column) as solid lines. In every case, the right imagery detection is shown in the customary orientation (above the main diagonal), while the left is shown in the opposite corner. The circled points on the curves mark the point of equal balance of sensitivity and

TABLE I
TP AND FP RATES (TPR% AND FPR%, RESPECTIVELY) CORRESPONDING TO THE BEST THRESHOLD FOR THE SOLID ROC CURVES SHOWN IN FIG. 5 AND THE AREA UNDER THE ROC CURVE (AUC) ARE GIVEN

Subject	Right Imagery			Left Imagery		
	TPR%	FPR%	AUC	TPR%	FPR%	AUC
g3	84	17	0.92	86	16	0.93
g7	72	29	0.78	66	35	0.72
i2	69	34	0.73	67	34	0.72

selectivity. For each of the three subjects, Table I captures the TP rate, FP rate, and the area under the ROC curve for the left and right imagery corresponding to the thresholds associated with those points. For ideal detection, the TP rate should be 1.0, the FP rate should be zero, and the area under the curve should be unity. Note the symmetry present in the table between left and right imagery for each of the three subjects. Subject g3 showed the highest rate of TPs and the lowest rate of FPs among all the subjects. Although the other two subjects appear to be quite similar in this analysis, they differ significantly when the data is subjected to different analysis methods.

The determination of the x and y coordinates plotted to form a ROC curve are based on a parameter, the test threshold. As the test threshold travels through the signal, both the false and true positive rates will increase as the threshold increases giving rise to a curve in which the x and y coordinates are strictly rising from 0 to 1. The ROC curves in the left column of Fig. 5, have these properties and therefore appear smooth and extend from the origin to the opposite corner of the graph.

B. Event-by-Event Analysis

There are only 80 events and 80 nonevents that were tested with 200 thresholds in the processing. Unlike the sample-by-sample comparison that effectively treats each of the thousands of samples separately, there are a relatively small number of events in the event-by-event analysis. This leads to curves in which often one of the coordinates being plotted remains fixed while the other changes. As a result of these two factors, the curves not only appear coarse, but often to "step" horizontally or vertically from point to point (Fig. 5 right column, solid line curve). There is no longer a guarantee that an increase in the test threshold will capture more FPs and more TPs. As a result, the ROC curves generated will not be the traditional well-behaved curves such as those shown in the left column of Fig. 5, but will exhibit such behavior as folding back in under themselves. Clearly, portions of the curve where such behavior takes place are not useful, and the range of the test thresholds was, therefore, limited to values that produced the portion of the curve which most closely approached the appropriate corner of the graph. This causes the curves to be truncated rather than stretching from corner to corner of the graph. Note that even though the curves were limited in this way, they still show sections of unevenness. This is also explained as a result of the true and FP rates not being both strictly increasing. The dotted lines in Fig. 5 (right column) represent the ROC curves corresponding to dwell and refractory times tested which produced the most accurate results.

TABLE II
EVENT-BY-EVENT COMPARISON OF DETECTOR OUTPUT WITH THE ACTUAL IMAGERY FOR DIFFERENT REFRACTORY PERIODS (RP). EIGHTY EVENTS WERE TESTED. THE DWELL TIME WAS 0.25 s IN ALL CASES EXCEPT FOR THOSE WITH NO REFRACTORY PERIOD IN WHICH CASE THE DWELL TIME WAS 0 s. T% REPRESENTS "TRUE-POSITIVE RATE," F% REPRESENTS "FALSE-POSITIVE RATE," AND TF% IS THE PERCENTAGE DIFFERENCE BETWEEN THESE PARAMETERS

Subject	RP	Right Imagery			Left Imagery		
		T%	F%	TF%	T%	F%	TF%
g3	none	81	10	71	76	16	60
	1.00	86	11	75	86	16	70
	1.25	91	16	75	86	15	71
	1.35	80	14	66	89	28	61
g7	none	58	16	42	79	37	42
	1.00	58	14	44	84	19	63
	1.25	65	27	38	86	24	62
	1.35	45	11	34	75	21	54
i2	none	93	77	16	94	75	19
	1.00	87	59	28	91	59	32
	1.25	81	52	29	88	59	29
	1.35	73	50	23	77	49	28

Refractory times of 1.0, 1.25, and 1.35 s were tested for all subjects and are summarized in Table II. In all cases where a dwell period was used, it was 0.25 s. For each of the subject/sessions, the maximum TF-difference for the tested thresholds was calculated. The table reports the true and FPs percentage rates and TF-differences for the three subjects. Various refractory periods were tested with and without dwell as described in the table. The best results generated were for the right imagery of subject g3 with a refractory period of 1.00 to 1.25 s, however, this imagery was generally good for all tested refractory periods for this subject. In the case of the left imagery of subject g7, however, the introduction of an appropriate refractory period made a more substantial improvement in the performance of this subject. Different refractory periods were best for each of the subjects.

IV. DISCUSSION AND CONCLUSION

Different issues arise when dealing with a simulated asynchronous BCI as compared with a synchronous one. It was shown that the evaluation criteria may be different and the requirements for good performance become more difficult to satisfy. The use of dwell and refractory to assist in determining the occurrence of a valid detection is an important contribution in achieving a more robust interface.

The use of ROC curves is a useful method to apply to a BCI providing a continuous evaluation of the output of a detector. We have also, however, demonstrated the limitations of this evaluation tool. A sample-by-sample analysis produces smooth curves that are easy to appraise, but may not be as useful for the evaluation of a practical application. The sample-by-sample approach effectively breaks each event or nonevent up into a large number of pieces. This allows, in a sense, the success or failure of the detector during that event or nonevent to be measured as a proportion of correctness rather than limiting the result to an absolute

failure or a perfect detection. Unfortunately, for a practical application, the later more stringent criteria is the one that must be imposed. In other words, we may be more concerned with the ability of the system to correctly determine the occurrence of events and nonevents rather than individual samples. For such situations, an event-based system may produce more practical results.

The TF-difference method of evaluation can not be used to produce ROC curves, however, it provides an alternative method of evaluation that does give a measure of the performance of a detection system. As we have seen, we may rely on this method for situations that are incompatible with the other approaches discussed here.

The introduction of a dwell and refractory period produced varying results in the performance of the Graz-BCI in the context of a simulated asynchronous application. Most notably, the left detection of subject g7 was dramatically improved [Fig. 5(e)], while it produced very poor results for subject i2 [Fig. 5(f)]. This could be related to the fact that subject i2 showed poor performance generally. The dwell time and the refractory times tested were selected on the basis of a manual examination of the original results. The results in the table suggest that the optimum refractory period varies according to the subject, but that a time between 1.0 and 1.25 s would give the best average performance over the subjects evaluated here. Alternate dwell times were not tested in this study. The time of 0.25 s was based on the observation that the method described with the initial event based evaluation failed when the detector output crossed test thresholds for very brief intervals at inappropriate times. Based on the manual evaluation that led to this conclusion, it is probably unlikely that altering the dwell time will have a very significant effect on the results; however, it does merit a more formal study.

The three subjects displayed different results. While subject g3 showed satisfactory results with TP around 90% and FP around 15% subject i2 showed an inferior behavior with FP around 60%. Inspection of the EEG data revealed many artifacts, partially originating from eye movements during the "resting" period. These artifacts are not unexpected because the subjects were only instructed to avoid eye movement during the imagery period after the cue presentation. In the case of an asynchronous BCI in real-world applications, the user has to be trained on the one hand to perform, for example, precise motor imagery tasks and on the other hand, to avoid such mental activity in between.

It can be stated that the introduction of a dwell time and a refraction period can enhance the overall classification performance of an asynchronous BCI system when simulated data taken from synchronous experiments is analyzed.

One difference between a synchronous and asynchronous BCI is that in the former case, a cue stimulus directs the subject to perform an imagery task. In this case, the time of the imagery task is known but this can be a disadvantage because the cue stimulus has to be processed. The latter means that the stimulus processing can induce changes in the ongoing EEG, superimposed to changes directly related to the imagery process. In this respect, it is of interest to note that a visual cue stimulus in synchronous BCI paradigm can modify sensorimotor rhythms as early as 250 ms after cue onset [13]. This can be interpreted as a result of a visually cue-triggered process in the primary sensorimotor area. In an asynchronous BCI, no external trigger is presented and, therefore, the mentioned interaction between trigger, induced, and mental task-induced processes does not exist. The Graz-BCI system tested was not sensitive to problems resulting from VEP, however, in the case of a general BCI, whatever effect visual evoked potentials might have is annulled by their absence in the asynchronous paradigm.

ACKNOWLEDGMENT

G. Townsend would like to thank his coauthors for their guidence and advice during the preparation of this paper. Prof. Pfurtscheller's invitation to join the Graz-BCI research group in the summer of 2000 eventually led to accpetance into the Ph.D. program at Graz Technical University. Prof. Pfurtscheller has proven to be an experienced and knowledgeable source of information and an understanding and patient supervisor. Dr. Graimann has been a kind and generous friend whose firm but gentle criticism during the preparation of this paper helped to improve it greatly.

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, 1998.

[2] N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kübler, J. Perelmouter, E. Taub, and H. Flor, "A spelling device for the paralyzed," *Nature*, vol. 398, pp. 297–298, 1999.

[3] J. R. Wolpaw and D. J. McFarland, "Multichannel EEG-based brain–computer communication," *Electroencephalogr. Clin. Neurophysiol.*, vol. 90, pp. 444–449, 1994.

[4] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain–computer communication," *Proc. IEEE*, vol. 89, pp. 1123–1134, July 2001.

[5] S. G. Mason and G. E. Birch, "A brain-controlled switch for asynchronous control applications," *IEEE Trans. Biomed. Eng.*, vol. 47, pp. 1297–1307, Oct. 2000.

[6] S. P. Levine, J. E. Huggins, S. L. BeMent, R. K. Kushwaha, L. A. Schuh, M. M. Rohde, E. A. Passaro, D. A. Ross, K. V. Elisevich, and B. J. Smith, "A direct brain interface based on event-related potentials," *IEEE Trans. Rehab. Eng.*, vol. 8, pp. 180–185, June 2000.

[7] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehab. Eng.*, vol. 8, pp. 441–446, Dec. 2000.

[8] C. Guger, H. Ramoser, and G. Pfurtscheller, "Real-time EEG analysis with subject-specific spatial patterns for a brain–computer interface (BCI)," *IEEE Trans. Rehab. Eng.*, vol. 8, pp. 447–456, Dec. 2000.

[9] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clin. Neurophysiol.*, vol. 110, pp. 787–798, 1999.

[10] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York: Academic, 1972.

[11] C. Guger, A. Schlögl, C. Neuper, D. Walterspacher, T. Strein, and G. Pfurtscheller, "Rapid prototyping of an EEG-based brain–computer interface (BCI)," *IEEE Trans. Rehab. Eng.*, vol. 9, pp. 49–58, Mar. 2001.

[12] R. O. Duda and P. E. Hard, *Pattern Classification and Scene Analysis*. New York: Wiley, 2000.

[13] G. Pfurtscheller, C. Neuper, H. Ramoser, and J. Müller-Gerking, "Visually guided motor imagery activates sensorimotor areas in humans," *Neurosci. Lett.*, vol. 269, pp. 153–156, 1999.

George Townsend received the M.Sc. degree in computer science from the University of Waterloo, Waterloo, ON, Canada, in 1989. He is currently on sabbatical working toward the Ph.D. degree at the University of Technology, Graz, Austria.

He is currently an Assistant Professor at Algoma University, Sault Ste. Marie, ON. His research interests include brain–computer interfaces and hardware design.

Bernhard Graimann received the M.S. degree in computer science and the Ph.D. degree in biomedical engineering from the University of Technology, Graz, Austria, in 1999 and 2002, respectively.

He is currently an Assistant Research Scientist with the Institute of Human Computer Interfaces, University of Technology, Graz. His research interests include biosignal processing, heuristic optimization, and, in particular, brain–computer communication systems.

Gert Pfurtscheller (M'00) received the M.S. and the Ph.D. degrees in electrical engineering from the Graz University of Technology, Graz, Austria.

He is currently a Professor of medical informatics and the Director of the Institute of Biomedical Engineering at the University of Technology, Graz. He is involved in a number of research programs including the design of a "brain–computer interface" and studies in functional brain topography using event-related desynchronization.