

**DESIGN OF A SELF-PACED BRAIN COMPUTER  
INTERFACE SYSTEM USING FEATURES EXTRACTED  
FROM THREE NEUROLOGICAL PHENOMENA**

by

Mehrdad Fatourech

B.Sc., University of Tehran, 1998

M.Sc., University of Tehran, 2001

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

**The Faculty of Graduate Studies**

(Electrical and Computer Engineering)

THE UNIVERSITY of BRITISH COLUMBIA

January 2008

© Mehrdad Fatourech, 2008

## ABSTRACT

Self-paced Brain computer interface (SBCI) systems allow individuals with motor disabilities to use their brain signals to control devices, whenever they wish. These systems are required to identify the user's "intentional control (IC)" commands and they must remain inactive during all periods in which users do not intend control (called "no control (NC)" periods).

This dissertation addresses three issues related to the design of SBCI systems: 1) their presently high false positive (FP) rates, 2) the presence of artifacts and 3) the identification of a suitable evaluation metric.

To improve the performance of SBCI systems, the following are proposed: 1) a method for the automatic user-customization of a 2-state SBCI system, 2) a two-stage feature reduction method for selecting wavelet coefficients extracted from movement-related potentials (MRP), 3) an SBCI system that classifies features extracted from three neurological phenomena: MRPs, changes in the power of the Mu and Beta rhythms; 4) a novel method that effectively combines methods developed in 2) and 3) and 5) generalizing the system developed in 3) for detecting a right index finger flexion to detecting the right hand extension. Results of these studies using actual movements show an average true positive (TP) rate of 56.2% at the FP rate of 0.14% for the finger flexion study and an average TP rate of 33.4% at the FP rate of 0.12% for the hand extension study. These FP results are significantly lower than those achieved in other SBCI systems, where FP rates vary between 1-10%.

We also conduct a comprehensive survey of the BCI literature. We demonstrate that many BCI papers do not properly deal with artifacts. We show that the proposed BCI achieves a good performance of TP=51.8% and FP=0.4% in the presence of eye movement artifacts. Further tests of the performance of the proposed system in a pseudo-online environment, shows an average TP rate =48.8% at the FP rate of 0.8%.

Finally, we propose a framework for choosing a suitable evaluation metric for SBCI systems. This framework shows that Kappa coefficient is more suitable than other metrics in evaluating the performance during the model selection procedure.

# TABLE OF CONTENTS

Abstract.....	ii
Table of Contents .....	iii
List of Tables.....	viii
List of Figures .....	xi
List of Abbreviations .....	xv
Acknowledgements.....	xvii
Dedication.....	xviii
Co-authorship Statement .....	xix
Chapter 1    Introduction and background.....	1
1.1       Introduction and motivation .....	1
1.1.1    High false positive rates (FPR) .....	3
1.1.2    Presence of artifacts.....	4
1.1.3    Evaluation metrics.....	5
1.2       Functional model of a brain computer interface system.....	5
1.3       Background.....	7
1.3.1    Signal recording .....	7
1.3.2    Choice of neurological phenomenon .....	8
1.3.3    Timing of BCI control .....	13
1.4       Design of self-paced BCI systems .....	14
1.5       Use of multiple neurological phenomena in BCI systems.....	18
1.5.1    Simultaneous application of MRPs and changes in the power of Mu/Beta rhythms.....	18
1.5.2    Using multiple neurological phenomena in BCI systems .....	19
1.6       Artifacts in BCI systems.....	21
1.6.1    Artifact avoidance .....	23
1.6.2    Artifact rejection.....	23
1.6.3    Artifact removal .....	25
1.7       Evaluating the performance of SBCI systems.....	27
1.8       Thesis contributions.....	31
1.8.1    Reducing high false positive rates .....	32
1.8.2    Addressing artifacts in SBCI systems.....	33
1.8.3    Finding a suitable evaluation metric for SBCI systems.....	34
1.9       Organization of the thesis.....	34
1.10      References.....	40

Chapter 2	Automatic user customization for improving the performance of a self-paced brain computer interface system .....	50
2.1	Introduction .....	50
2.2	Background.....	53
2.3	Problem statement.....	56
2.4	Methods.....	58
2.5	Experimental results.....	62
2.6	Discussion and conclusions .....	69
2.7	Acknowledgements .....	71
2.8	References.....	72
Chapter 3	Application of a hybrid wavelet feature selection method in the design of a self-paced brain computer interface system.....	75
3.1	Background.....	75
3.2	Data collection.....	79
3.3	Method.....	81
3.4	Results.....	86
3.5	Discussion and conclusions .....	89
3.6	Acknowledgements .....	95
3.7	References.....	96
Chapter 4	A self-paced brain computer interface system that uses movement related potentials in changes in the power of brain rhythms .....	99
4.1	Introduction .....	99
4.2	Background.....	102
4.2.1	Neurological phenomenon background.....	102
4.2.2	Multiple neurological phenomena in BCI systems.....	104
4.3	Data collection.....	106
4.4	Methods.....	108
4.4.1	Feature extraction .....	110
4.4.2	Feature classifier .....	114
4.4.3	Feature selection.....	117
4.4.4	Performance evaluation .....	119
4.5	Results.....	120
4.6	Discussion.....	127
4.6.1	Observations on the BCI designs based on a single neurological phenomenon .....	127
4.6.2	Observations on Study 1 .....	128
4.6.3	Observations on Study 2 .....	128
4.6.4	Statistical analysis.....	128
4.7	Acknowledgements .....	131
4.8	References.....	133
Chapter 5	A self-paced brain computer interface system with a low false positive rate .....	138
5.1	Introduction .....	138
5.2	Methods.....	141

5.2.1	Feature extraction .....	141
5.2.2	Feature classification.....	147
5.2.3	Hybrid genetic algorithm (HGA).....	149
5.3	Experimental results.....	152
5.3.1	Data collection and evaluation.....	152
5.3.2	Results.....	155
5.4	Discussion and conclusions.....	157
5.5	Acknowledgements .....	162
5.6	References.....	163
Chapter 6	EMG and EOG artifacts in brain computer interface systems: a survey .....	167
6.1	Introduction .....	167
6.2	Current neurological phenomena and associated artifacts.....	168
6.2.1	Current neurological phenomena .....	168
6.2.2	Artifacts in BCI systems .....	171
6.3	Methods of handling artifacts .....	172
6.3.1	Artifact avoidance .....	172
6.3.2	Artifact rejection.....	173
6.3.3	Artifact removal .....	175
6.4	Literature survey .....	178
6.4.1	EOG artifacts .....	185
6.4.2	EMG artifacts .....	185
6.5	Discussion and conclusions.....	186
6.6	References.....	189
Chapter 7	Performance of a self-paced Brain computer Interface on data contaminated with eye blinks and on data recorded in subsequent sessions .....	209
7.1	Introduction .....	209
7.2	Methods.....	212
7.2.1	Self-paced brain computer interface design .....	212
7.2.2	Data collection .....	213
7.2.3	Evaluation.....	215
7.3	Results.....	217
7.3.1	Analysis of SBCI performance on artifact-contaminated data .....	217
7.3.2	Test on data recorded in subsequent sessions .....	218
7.3.3	The effect of adding a debounce component .....	222
7.4	Discussion.....	227
7.5	Acknowledgements .....	230
7.6	References.....	231
Chapter 8	Selection of a suitable evaluation metric for a self-paced brain computer interface system.....	234
8.1	Introduction .....	234
8.2	Problem statement.....	239
8.3	A framework for comparing evaluation metrics.....	244

8.3.1	Suitability of an evaluation metric.....	245
8.3.2	Guidelines for comparing two evaluation metrics .....	247
8.3.3	Degree of consistency (DoC) .....	248
8.3.4	The Degree of discriminancy (DoD).....	251
8.3.5	Comparison of two evaluation metrics.....	252
8.3.6	Using sub-sampling grids for calculating the comparison measures.....	252
8.4	Selected evaluation metrics in SBCIs.....	255
8.4.1	Overall accuracy (OA) .....	255
8.4.2	Information transfer rate (mutual information) .....	256
8.4.3	Kappa.....	256
8.4.4	HF-difference .....	257
8.4.5	$\frac{TPR}{FPR}$ ratio .....	257
8.4.6	ROC curve and related metrics .....	258
8.5	Simulations .....	259
8.5.1	Application.....	259
8.5.2	Results.....	259
8.6	Discussion and conclusions .....	271
8.7	References.....	274
Chapter 9	New studies on the design of a 2-state self-paced brain computer interface system with a low false activation rate .....	276
9.1	Introduction .....	276
9.2	Experimental paradigm .....	280
9.2.1	Data recording .....	280
9.2.2	Artifact monitoring .....	282
9.3	System design methods .....	283
9.3.1	Generating the IC and NC data .....	283
9.3.2	Feature extraction .....	284
9.3.3	Feature classification.....	287
9.3.4	Multiple classifier system .....	288
9.3.5	Calculating the TPs and FPs.....	288
9.3.6	Metric selection for model evaluation .....	289
9.3.7	Model selection .....	292
9.3.8	Evaluation.....	293
9.3.9	Using ROC curves for summarizing the performance on test sets.....	293
9.4	Results.....	294
9.4.1	Choosing the evaluation metric for model selection.....	295
9.4.2	Performance of the system .....	295
9.5	Discussion and future work .....	299
9.5.1	Discussion.....	299
9.5.2	Future works .....	305
9.6	Acknowledgements .....	308
9.7	References.....	309

Chapter 10	Summary and conclusions .....	312
10.1	Summary .....	312
10.1.1	Chapter 2: Improving the performance of LF-ASD by automatic user-customization .....	313
10.1.2	Chapter 3: Using DWT to extract features .....	314
10.1.3	Chapter 4: Using three neurological phenomena as the source of control.....	315
10.1.4	Chapter 5: Design of an automated SBCI system with low FP rates .....	317
10.1.5	Chapter 6: Analysis of the effect of artifacts in BCI systems .....	318
10.1.6	Chapter 7: Analysis of the performance of the proposed SBCI on artifact-contaminated data .....	319
10.1.7	Chapter 8: A framework for evaluating the performance of SBCI systems .....	320
10.1.8	Chapter 9: Applying the proposed SBCI with hand extension data .....	320
10.2	Summary of contributions.....	321
10.2.1	Reducing high false positive rates .....	321
10.2.2	Addressing artifacts in SBCI systems.....	323
10.2.3	Finding a suitable evaluation metric for SBCI systems.....	323
10.3	Future research directions.....	324
10.4	References.....	326
Appendix A-	UBC Research Ethics Board Certificate .....	327
Appendix b-	Theoretical analysis of the proposed SBCI .....	328
B.1.	Formulating the problem .....	328
B.2.	Constraints .....	329
B.3.	Objective functions.....	330
B.4.	Results .....	332
B.5.	References .....	336

## LIST OF TABLES

Table 1-1. Comparison of the TPR and FPR rates achieved in different SBCI studies. ....	17
Table 2-1. The confusion matrix for a 2-state self-paced BCI system.....	60
Table 2-2. Comparison of the fitness value of the initial and final populations (tested on the validation sets).....	66
Table 2-3. TP rates of the LF-ASD and the ALF-ASD (FP=2%). ....	68
Table 2-4. Delay parameter values used in the design of the LF-ASD based on the ensemble averages of the MRP patterns in the training data set. Note that $\beta_i$ and $\beta_j$ are set to zero and that the same delay parameter values are used for the rest of the bipolar channels. The table is reproduced from [15]. ....	69
Table 3-1. The confusion matrix for a 2-state self-paced BCI system.....	84
Table 3-2. Comparison of the average TP, average FP rates, average $\frac{TPR}{FPR}$ and the average number of features.....	87
Table 3-3. The average number of selected features per channel after applying the hybrid feature selection algorithm. ....	88
Table 4-1. The average TP and FP rates (%) for Study 1 (the numbers in parenthesis show the standard deviation).....	121
Table 4-2. The average TP and FP rates (%) for each neurological phenomenon in Study 2 (the numbers in parenthesis show the standard deviation).....	121
Table 4-3. The average TP and FP rates (%) for User AB1 in Study 2 (the numbers in parenthesis show the standard deviation).....	125
Table 4-4. The average TP and FP rates (%) for User AB2 in Study 2 (the numbers in parenthesis show the standard deviation).....	126
Table 4-5. The average TP and FP rates (%) for User AB3 in Study 2 (the numbers in parenthesis show the standard deviation).....	126
Table 4-6. The average TP and FP rates (%) for User AB4 in Study 2 (the numbers in parenthesis show the standard deviation).....	127
Table 5-1. The time schedule of recording the data.....	153



Table 5-2. The performance results for the proposed SBCI system. ....	156
Table 5-3. Comparison of the performance results. ....	158
Table 6-1. Methods of handling artifacts in BCI literature.....	181
Table 6-2. Methods of automatic EOG rejection in BCI studies. ....	183
Table 6-3. Methods of automatic EMG rejection in BCI studies.....	184
Table 6-4. Methods of automatic EOG removal in BCI studies. ....	184
Table 6-5. Methods of automatic EMG removal in BCI studies.....	185
Table 7-1. The time schedule of recording the data. For each participant, Day 1 is the first day that a participant attended the experiments. The rest of days are numbered with respect to Day 1 of that particular participant.....	215
Table 7-2. Comparison of the average test results on artifact-contaminated and non-contaminated data. The averages are calculated over 5 outer validation sets. The numbers in the parentheses indicate standard deviations. ....	217
Table 7-3. Comparison of the average results using data recorded in the first five sessions with those using data recorded in subsequent sessions. The averages are calculated over 5 outer validation sets. The numbers in the parentheses indicate standard deviations.....	219
Table 8-1. DoS results for the evaluation metrics studied in this paper. <i>Res1</i> stands for the finest resolution, <i>Mean</i> stands for the average of 10 resolution values and <i>Res10</i> stands for the coarsest resolution. ....	261
Table 8-2. The DoC results for the evaluation metrics studied in this paper (the first three comparisons). <i>Res1</i> , <i>Mean</i> and <i>Res10</i> stand for the finest resolution, the average of 10 resolution values and the coarsest resolution, respectively.....	263
Table 8-3. The DoC results for the evaluation metrics studied in this paper (the last three comparisons). <i>Res1</i> , <i>Mean</i> and <i>Res10</i> stand for the finest resolution, the average of 10 resolution values and the coarsest resolution, respectively.....	264
Table 8-4. The DoD results for the evaluation metrics studied in this paper (the first three comparisons). <i>Res1</i> , <i>Mean</i> and <i>Res10</i> stand for the finest resolution, the average of 10 resolution values and the coarsest resolution, respectively.....	265
Table 8-5. The DoD results for the evaluation metrics studied in this paper (the last three comparisons). <i>Res1</i> , <i>Mean</i> and <i>Res10</i> stand for the finest resolution, the average of 10 resolution values and the coarsest resolution, respectively.....	266

Table 8-6 . The effect of weights on average values of DoC.....	267
Table 8-7. The effect of weights on average values of DoD. ....	268
Table 9-1. Comparison of the TP rates of monopolar and bipolar montages for different false activation rates. The numbers in parentheses show the standard deviations.....	299
Table 9-2. Comparison of the TPR and FAR rates achieved in different SBCI studies. ....	302

## LIST OF FIGURES

Figure 1-1. A BCI system allows users to control a device using their brain signals only.....	2
Figure 1-2. A typical SBCI system that identifies an IC command related to the execution of right finger flexion.....	3
Figure 1-3. High false positive rates can significantly impact the performance of an SBCI system, even if the TP rates are high. (a) Brain states of a user; (b) The output of the SBCI system. ....	4
Figure 1-4. Functional model of a BCI system. Note the control display is optional.....	6
Figure 1-5. Two examples of neurological phenomena. (a) Changes in the power of Beta rhythms over time; (b) A movement-related potential. Vertical line shows the time of activation of the movement. Note that these shapes are generated by averaging over many epochs. ....	10
Figure 1-6. Synchronized vs. self-paced control. (a) In a synchronized BCI system, control can be done only in certain intervals specified by the system; (b) In a self-paced BCI system, the control is done at the user's own pace. ....	14
Figure 1-7. An example of how artifacts can affect the performance of an SBCI system. (a) The brain state of the user; (b) The periods when artifacts have occurred; (c) The output of the SBCI system (note: FP: false positive, TN: true negative, FN: false negative and TP: true positive).....	21
Figure 1-8. Types of evaluation metrics used in synchronized and self-paced BCI systems. ....	29
Figure 1-9. The overall schematic of the SBCI system developed and studied in Chapters 4, 5, 7, and 9.....	36
Figure 1-10. Outline of the thesis. ....	39
Figure 2-1. Components of the LF-ASD system (from [32]). ....	54
Figure 2-2. Points selected by the feature generator when applied to a sample bipolar EEG signal. ....	55

Figure 2-3. The fitness of the best chromosomes as a function of the generation number for two representative individuals. a) AB2; b) SCI4.....	65
Figure 3-1. The overall structure of the proposed hybrid method for extracting MRP features.....	81
Figure 3-2. Spatial distribution of the average number of selected features for AB1. ....	90
Figure 3-3. Spatial distribution of the average number of selected features for AB2. ....	90
Figure 3-4. Spatial distribution of the average number of selected features for AB3. ....	91
Figure 3-5. Spatial distribution of the average number of selected features for AB4. ....	91
Figure 3-6. Comparison of the fitness of the best chromosome vs. other subset of features.....	95
Figure 4-1. Functional model of a BCI system (adapted from [1]). ....	99
Figure 4-2. Synchronized vs. SBCI systems. (a) In a synchronized BCI system control is only possible during <i>System Ready</i> periods; (b) In an SBCI system, the system continuously accepts the input signals. ....	100
Figure 4-3. NC periods are generated by shifting a window over NC datasets. ....	108
Figure 4-4. The overall structure of the SBCI system implemented in Study 1. ....	109
Figure 4-5. The overall structure of the two-stage MCS implemented in Study 2. ....	110
Figure 4-6. The process of generating templates. ....	112
Figure 4-7. The spatial distribution of the selected features for individuals in Study 2. (a) User AB1; (b) User AB2; (c) User AB3; and (d) User AB4. ....	125
Figure 5-1. The overall structure of the improved SBCI incorporating three neurological phenomena.....	142
Figure 5-2. An example of how features are extracted using the proposed cross-covariance method. ....	145
Figure 5-3. (a) The structure of a chromosome; (b) Representation of the parameter values for each SVM in a chromosome.....	150
Figure 5-4. Method of calculating the TP rate; (a) EEG Signal; (b) Output of the finger switch; (c) Output of the SBCI.....	155

Figure 6-1. The functional model of a BCI system depicting its principle functional components.....	167
Figure 6-2. The number of papers published on different methods of handling EOG artifacts in BCI studies.....	180
Figure 6-3. The number of papers published on different methods of handling EMG artifacts in BCI studies. ....	180
Figure 7-1. The overall structure of the improved SBCI .....	213
Figure 7-2. An example of extracting the maximum of the cross-correlogram using the proposed cross-covariance method. ....	214
Figure 7-3. Method of calculating the TP rate; (a) EEG Signal; (b) Output of the hand switch; (c) Output of the SBCI.....	216
Figure 7-4. The SBCI output during periods when finger movements were executed for a) Participant AB1; b) Participant AB2; and c) The output of the SBCI during NC sessions when movements did not occur for Participant AB1. ....	222
Figure 7-5. The operation of a debounce component.....	224
Figure 7-6. The TP rate, FP rate and the $\frac{TPR}{FPR}$ ratio as a function of the length of the debounce window for (a) Participant AB1; (b) Participant (AB2); (c ) Participant AB3; (d) Participant (AB4); (e) Averages of all four participants.....	227
Figure 7-7. The output of the SBCI during periods when finger movements were executed for participant AB4; (b) the output of the SBCI during NC sessions when movements did not occur for participant AB4. ....	229
Figure 8-1. a) An example of a confusion matrix for a balanced dataset; b) An example of a confusion matrix for an imbalanced dataset; c) A second example of a confusion matrix for an imbalanced dataset. ....	236
Figure 8-2. A sample fitness landscape for a classification problem with two classes. ....	242
Figure 8-3. An example of dividing the (TPR,FPR) domain into regions. Different movements on the (TPR, FPR) space may be associated with different weights. Note that the numbers on each axis denote (%). ....	244
Figure 8-4. Two examples of more complex break-down of the (TPR, FPR) domain with more complex partition and weighting schemes.....	245
Figure 8-5. An example of using grids on the (TPR, FPR) domain. ....	253

Figure 8-6. Consistency between different resolutions in reaching the same conclusion for the cases studied here. The chart attributes to DoD results. ....	271
Figure 9-1. The overall structure of the SBCI (from [17]). The dashed lines show the parts of the system whose values are determined by the hybrid genetic algorithm (HGA). ....	284
Figure 9-2. An example of how features are extracted using the proposed cross-covariance method. ....	286
Figure 9-3. Method of calculating the TP rate; (a) EEG Signal; (b) Output of the finger switch; (c) Output of the SBCI. ....	289
Figure 9-4. An example of dividing the (TPR,FPR) domain into regions. Different movements on the (TPR, FPR) space may be associated with different weights. Note that the numbers on each axis denote (%). ....	291
Figure 9-5. ROC plots for (a) AB1; (b)AB2;(c)AB3;(d) AB4;(e)AB5. ....	298
Figure 9-6. Average MRPs for Channel C1 over two sessions. (a) participant AB1; (b)participant AB2 ; (c) participant AB3; (d) participant AB4;. ....	305

## LIST OF ABBREVIATIONS

<b>1-NN</b>	One nearest neighbor
<b>AB</b>	Able-bodied
<b>AEP</b>	Auditory evoked potentials
<b>ANC</b>	Activity of neural cells
<b>AR</b>	Auto-regressive
<b>AUC</b>	Area under the receiver operating characteristic curve
<b>BCI</b>	Brain computer interface
<b>BI</b>	Brain Interface
<b>BSS</b>	Blind source separation
<b>CAR</b>	Common average reference
<b>CBR</b>	Changes in the brain rhythms
<b>CPBR</b>	Changes in the power of Beta rhythms
<b>CPMR</b>	Changes in the power of Mu rhythms
<b>CT</b>	Cognitive task
<b>DoC</b>	Degree of consistency
<b>DoD</b>	Degree of discriminancy
<b>DoS</b>	Degree of suitability
<b>DROP</b>	Desired region of operation
<b>DWT</b>	Discrete wavelet transform
<b>ECG</b>	Electrocardiography
<b>ECoG</b>	Electro-corticogram
<b>EEG</b>	Electroencephalogram
<b>EMG</b>	Electromyogram
<b>ENT</b>	Energy normalization transform
<b>EOG</b>	Electrooculgram
<b>ERD</b>	Event-related desynchronization
<b>ERP</b>	Event-related potential
<b>ERS</b>	Event-related synchronization
<b>FAR</b>	False activation rate
<b>FDR</b>	False discovery rate
<b>FIR</b>	Finite impulse response
<b>FP</b>	False positive
<b>FPR</b>	False positive rate
<b>GA</b>	Genetic algorithm
<b>HGA</b>	Hybrid genetic algorithm
<b>IC</b>	Intentional control
<b>ICA</b>	Independent component analysis
<b>ITR</b>	Information transfer rate
<b>k-NN</b>	k-nearest neighbor
<b>LF-ASD</b>	Low frequency- asynchronous switch design
<b>MCS</b>	Multiple classifier system
<b>MEG</b>	Magnetoencephalography
<b>MI</b>	Mutual information

<b>MN</b>	Multiple neurological phenomena
<b>MRA</b>	Movement-related activity
<b>MRP</b>	Movement-related potential
<b>NC</b>	No control
<b>NN</b>	Neural networks
<b>OA</b>	Overall accuracy
<b>OPM</b>	Outlier processing method
<b>PCA</b>	Principal component analysis
<b>RMS</b>	Root mean square
<b>ROC</b>	Receiver operating characteristic
<b>SBCI</b>	Self-paced brain computer interface
<b>SCI</b>	Spinal cord injury
<b>SNR</b>	Signal-to-noise ratio
<b>SSEP</b>	Somatosensory evoked potential
<b>SSVEP</b>	Steady stated visual evoked potential
<b>STD</b>	Standard deviation
<b>SVM</b>	Support vector machine
<b>SWT</b>	Stationary wavelet transform
<b>TEM</b>	Time of expected attempted movement
<b>TP</b>	True positive
<b>TPR</b>	True positive rate
<b>VEP</b>	Visual evoked potential



## ACKNOWLEDGEMENTS

This thesis is the result of nearly five years of research. I would like to express my sincere gratitude to all those who have supported me for completing my thesis. First, I would like to thank my research supervisors, Prof. Rabab K. Ward and Dr. Gary E. Birch for giving me the opportunity to work in their research group. I am greatly indebted to their guidance, support and encouragement throughout the course of my studies. I would also like to give thanks to my committee members: Dr. Dave Michelson, Dr. Jane Wang, and Dr. Tim Salcudean for investing their time to read and give me valuable feedback on this thesis.

Next, I extend my thanks to the researchers in the brain interface laboratory of the Neil Squire Society and the Image and Signal Processing lab of UBC for their support and help. Especially, I would like to thank Dr. Lino Coria, Dr. Steven G. Mason, Dr. Jaimie Borisoff, Gordon Handford, Borna Nouredin, Xin Yi Yong, Angela Chuang, Qiang Tang, and Zicong Mai for their valuable technical comments and support on my work.

Most importantly, I would like to express my deepest gratitude to my wife, Nona, my parents, my sister, my family and my friends for their love, help and endless support. If it were not for their sincere encouragement, I would have not made it as far as I did.

This work was supported in part by NSERC under Grant 90278-06 and CIHR under Grant MOP-72711.

## **DEDICATION**

**This thesis is dedicated to my lovely wife, Nona,**

Who has offered me unconditional love and support and has stood by me all along. She believed in me, when I did not and for that, I shall always remain grateful.

**And it is also dedicated to my parents, Nasrin and Hassan,**

Who have raised me to be the person I am today. They have been with me in every step. I hope I have made them proud.

**And it is also dedicated to Shahnaz and Hossein,**

Who have always supported me and encouraged me.

## CO-AUTHORSHIP STATEMENT

1- Fatourech, M., Bashashati, A., Birch, G.E. and Ward, R.K. “Automatic User Customization for Improving the Performance of an Asynchronous Brain Interface System”, *Journal of Medical & Biological Engineering and Computing*, Vol.44, No.12, Dec 2006, pp.1093-1104.

MF developed the method, analyzed the data, interpreted the results, wrote the manuscript and acted as the corresponding author.

AB contributed to the study design, helped in shaping the manuscript , interpreting the results and evaluating the manuscript.

RKW and GEB supervised the development of the work, helped in writing and evaluating the manuscript.

2- Fatourech, M., Birch, G. E., and Ward, R. K., "Application of a Hybrid Wavelet Feature Selection Method in the Design of a Self-paced Brain Interface System", *Journal of NeuroEngineering and Rehabilitation*, Vol.4, No.1, Apr 2007.

MF developed the method, analyzed the data, interpreted the results, wrote the manuscript and acted as the corresponding author.

RKW and GEB supervised the development of the work, helped in writing and evaluating the manuscript.

3- Fatourech, M., Birch, G. E., and Ward, R. K., “A Self-paced Brain Interface System that Uses Movement Related Potentials and Changes in the Power of Brain Rhythms”, *Journal of Computational Neuroscience*, Vol.23, No.1, Aug 2007,pp.21-37.

MF developed the method, analyzed the data, interpreted the results, wrote the manuscript and acted as the corresponding author.

RKW and GEB supervised the development of the work, helped in writing and evaluating the manuscript.

4- Fatourehchi, M., Birch, G. E., Ward, R. K., “A Self-paced Brain Interface System with a Low False Positive Rate”, *Journal of Neural Engineering*, vol.5, 2008, pp.9-23.

MF developed the method, analyzed the data, interpreted the results, wrote the manuscript and acted as the corresponding author.

RKW and GEB supervised the development of the work, helped in writing and evaluating the manuscript.

5- Fatourehchi, M., Bashashati, A., Ward, R. K., and Birch, G. E., "EOG and EMG Artifacts in Brain Interface Systems: a Survey", *Clinical Neurophysiology*, Vol.118, No.3, Mar 2007, pp.480-494 (*Invited Paper*).

MF reviewed the papers, created the tables, interpreted the results, wrote the manuscript and acted as the corresponding author.

AB helped in writing the manuscript, interpreting the results and evaluating the manuscript.

RKW and GEB supervised the development of the work, helped in writing and evaluating the manuscript.

6- Fatourehchi, M., Birch, G.E., Ward, R.K., “Performance of a Self-paced Brain Computer Interface on Data Contaminated with Eye Blinks and on Data Recorded in Subsequent Sessions”, submitted.

MF developed the method, analyzed the data, interpreted the results, wrote the manuscript and acted as the corresponding author.

RKW and GEB supervised the development of the work, helped in writing and evaluating the manuscript.

7- Fatourehchi, M., Mason, S. G., Ward, R.K., and Birch, G. E., “A New Framework for Comparing Metrics Used in Pattern Classification Problems with Large Test Samples”, submitted.

MF developed the method, analyzed the data, interpreted the results, wrote the manuscript and acted as the corresponding author.

SGM contributed to the development of the initial concept, interpreting the results and evaluating the manuscript.

RKW and GEB supervised the development of the work, helped in writing and evaluating the manuscript.

8- Fatourech, M., Ward, R.K., and Birch, G. E., “New studies on the design of a 2-state self-paced brain computer interface system with a low false positive rate”, submitted.

MF developed the method, analyzed the data, interpreted the results, wrote the manuscript and acted as the corresponding author.

RKW and GEB supervised the development of the work, helped in writing and evaluating the manuscript.

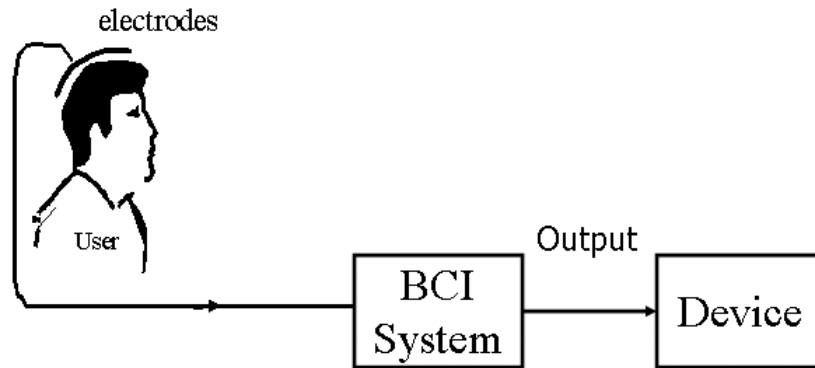
# **CHAPTER 1 INTRODUCTION AND BACKGROUND**

## **1.1 Introduction and motivation**

Many physiological disorders such as Amyotrophic Lateral Sclerosis (ALS) or injuries such as high-level spinal cord injury can disrupt the communication path between the brain and the body. People with severe motor disabilities may lose all voluntary muscle control, including eye movements. These people are forced to accept a reduced quality of life, resulting in dependence on caretakers and escalating social costs [1]. Most of the existing assistive technology devices for these patients are not possible because these devices are dependant on motor activities from specific parts of the body. Alternative control paradigms for these individuals are thus desirable.

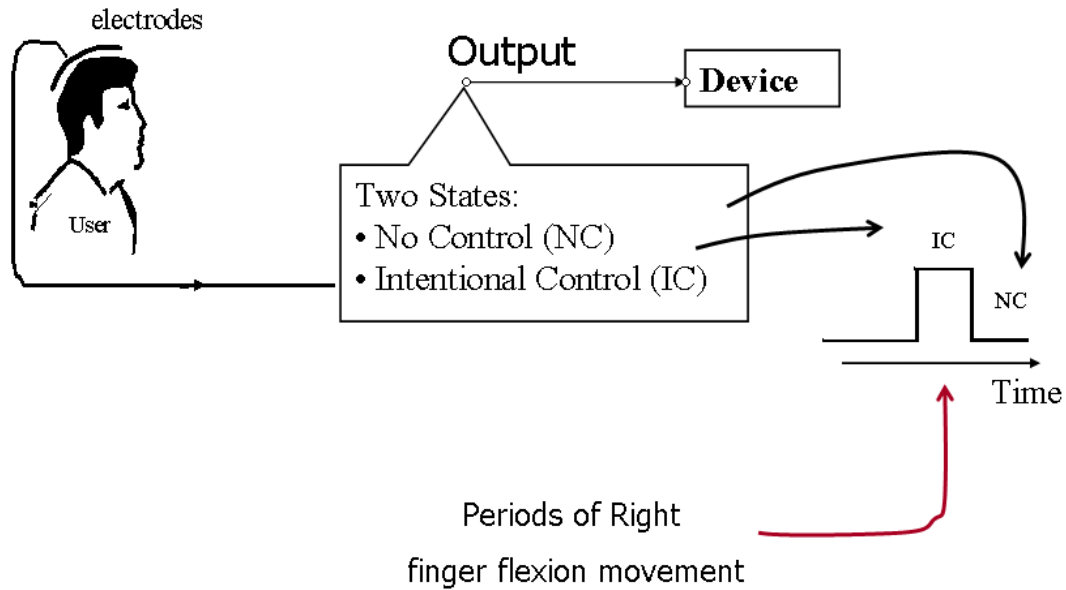
Over the last two decades, brain-computer interface (BCI) has emerged as a new frontier in assistive technology since it could provide an alternative communication channel between a user's brain and the outside world [2](see Figure 1-1 for a high-level block diagram of a BCI system). Other terms that are also used in the literature for referring to a BCI system include: brain interface (BI), direct brain interface (DBI), and brain machine interface (BMI). A successful BCI design would enable people to control objects in their environment (such as a light switch in their room or television, wheelchairs, neural prosthesis and computers) by thought only. This could be accomplished by measuring specific features of the user's brain activity that relate to his/her intent to perform the control. This specific type of brain activity is termed a "neurological phenomenon". As an example, when a particular movement such as right index finger flexion is performed, specific neurological phenomena that correspond to

that movement are generated. The corresponding neurological phenomena are then translated into signals that are eventually used to control devices [3].



**Figure 1-1. A BCI system allows users to control a device using their brain signals only.**

Currently, two different approaches are pursued in the design of BCI systems: *synchronized* and *self-paced* [4]. In the *synchronized* approach, which forms the traditional approach to the design of BCI systems, the user can only perform the control in certain time intervals that are specified by the system. While synchronized BCI systems can achieve high classification accuracy (>90%), their application is limited. This is because the user cannot perform the control at all times. Moreover, many of these systems assume that the user will exert an intentional control (IC) command during specified control periods. In other words they do not consider periods for which the user does not wish to exert control (called *no control*, NC, periods). As a result, they may become unstable during NC periods [3]. To address these shortcomings of synchronized BCI systems, the concept of self-paced BCI (SBCI) has been proposed. An SBCI system is constantly available for a user to use, as it should be able to identify IC patterns from the NC periods. Figure 1-2 shows a typical example of a 2-state SBCI system that should recognize IC patterns generated as the result of right finger flexion from the NC states. The output of this SBCI system is 'IC' when the system detects an IC command and is "NC" at all other times.



**Figure 1-2. A typical SBCI system that identifies an IC command related to the execution of right finger flexion**

The main aim of this thesis is to study the following three issues pertaining to SBCI systems:

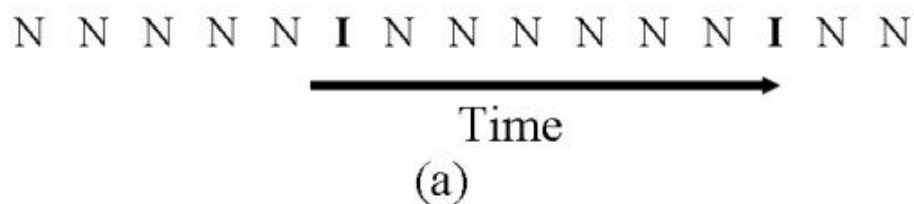
### **1.1.1 High false positive rates (FPR)**

The performance of SBCI systems is usually summarized by two measures: 1) the correct detection rate of IC commands (denoted as the true positive or TP rate), 2) the amount of false activations during NC periods (false positive or FP rate). At present, the performance of the SBCI technology is not high enough so that it can be used in a practical setting. While these systems can achieve an arguably good detection rate of  $TP > 50\%$ , their FP rates remain too high for practical applications (e.g., a false positive occurs every few seconds [5, 6]). For example, it has been argued that for an SBCI system that makes a decision every  $\frac{1}{16}^{th}$  of a second, FP rates higher than 2% can cause excessive user frustration, since the SBCI system generates a false positive every 6.25 seconds on average [5]. As another example, consider the self-paced control of lighting in a room using an SBCI system. The system has two states: I (*turn on/turn off the light*)

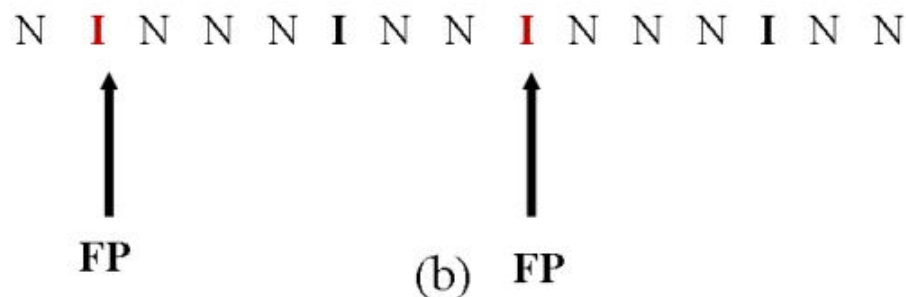


and N (*no control*). Figure 1-3 (a) and Figure 1-3 (b) show the brain states of a user and the output of the SBCI system, respectively. As seen, the system generated an FP at the beginning of monitoring the brain signals. The user then attempted to compensate for this error by issuing an intentional control command. After a short period, the system generated a second false positive, and the user had to compensate for it again. Clearly, during this period, the user only managed to compensate for the errors generated by the system. This process becomes frustrating when errors happen frequently and especially if the TP rate is not very high. Based on these arguments, it is clear that the ultimate value of this new technology will largely depend on the degree to which its performance can be improved, e.g., false positives shall occur no more than once per minute.

### • Brain States



### • Output of the SBCI System



**Figure 1-3. High false positive rates can significantly impact the performance of an SBCI system, even if the TP rates are high. (a) Brain states of a user; (b) The output of the SBCI system.**

#### 1.1.2 Presence of artifacts

A second factor that limits the application of SBCI systems is the presence of artifacts. Artifacts are unwanted signals that can degrade the performance of the system. If artifacts occur at the same time of the initiation of an IC command by the user, they may change the shape of a neurological phenomenon, and decrease the TP rate. If they

occur during the NC periods, they can generate false positives and increase the FP rate. Since some artifacts such as physiological artifacts (e.g., eye blinks) frequently occur, methods should be developed to effectively handle them. Unfortunately, most BCI systems do not handle artifacts at all (or at least efficiently). This is a serious drawback in online applications of BCI systems in general and SBCI systems in particular.

### **1.1.3 Evaluation metrics**

Yet another important factor in the design of SBCI systems is the availability of a suitable “evaluation metric”. In synchronized BCI systems, the overall classification accuracy (OA) and the information transfer rate (ITR) are metrics that are widely used. They are also accepted by the research community as reliable measures for comparing the performances of different synchronized BCI systems. This is not the case for SBCIs. It is very difficult to compare the performances of different SBCI systems. A wide variety of metrics such as OA[7], HF-difference[8], the mutual information (and ITR) [9], Kappa [10], the area under the receiver operating characteristic (ROC) curve [10], the TP rate at a fixed FP rate [5] and others have been proposed in the literature. However, no consensus yet exists amongst self-paced BCI researchers regarding which metric is more suitable for summarizing the performance and how a suitable evaluation metric should be chosen for a particular self-paced BCI systems [4].

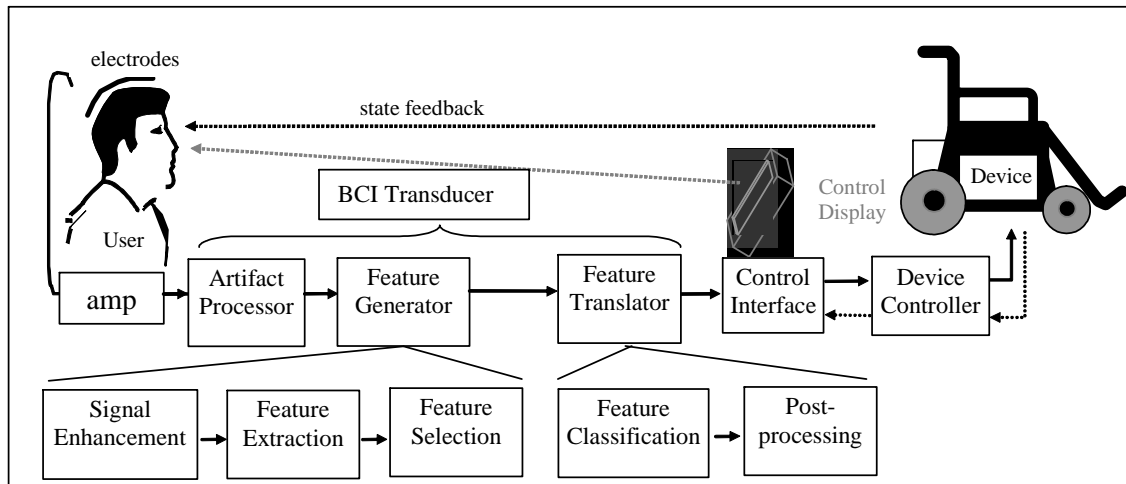
Please note that the neurological phenomena generated as the results of attempted movements by able-bodied individuals are similar to those generated by individuals with motor disabilities, as discussed later in this section. For this reason, in this thesis, data collected from able-bodied individuals are used for the analysis.

Before we address the existing work in the literature related to the above topics, we first provide some background information about the operation of BCI systems. This is done in the next two sections.

## **1.2 Functional model of a brain computer interface system**

Figure 1-4 shows a traditional BCI system in which a person controls a device in an operating environment (e.g., a powered wheelchair in a house) through a series of

functional components [11]. In this context, the user's brain activity is used to generate IC commands that operate the BCI system. The user monitors the state of the device to determine the result of his/her control efforts.



**Figure 1-4. Functional model of a BCI system. Note the control display is optional.**

The building components of a BCI system (shown in Figure 1-4 ) have the following tasks: the electrodes placed on the head of the user record the brain signal (e.g., electroencephalography (EEG) signals from the scalp, electrocorticography (ECoG) signals from the brain or neuronal activity recorded using microelectrodes implanted in the brain). The ‘artifact processor’ block deals with artifacts in the EEG signals after the signals have been amplified. This block can either remove artifacts from the EEG signals or can simply mark some EEG epochs as artifact-contaminated. The ‘feature generator’ block transforms the resultant signals into feature values that relate to the underlying neurological phenomena employed by the user for control. For example, if the user is using the power of his/her Mu (8-12Hz) rhythm for the purpose of control, the feature generator could continually generate features relating to the power-spectral estimates of the user’s Mu rhythms. The feature generator generally consists of three components: the ‘signal enhancement’, the ‘feature extraction’, and the ‘feature selection’ components, as shown in Figure 1-4.

In some BCI designs, ‘signal enhancement’ or some of form of ‘pre-processing’ is performed to increase the signal-to-noise ratio of the brain signal(s) prior to extracting the

features. To reduce the dimensionality of the problem, it is desired to reduce the number of features and/or the number of EEG channels. ‘Feature selection’ could be performed after or at the feature extraction stage to reduce the number of features and/or EEG channels used. Ideally, the features that are meaningful or useful in the classification stage are identified and chosen, while others are omitted.

The ‘feature translator’ block translates the features into logical control signals, e.g., 0 and 1 where 0 denotes NC and 1 denotes IC. The translation algorithm uses linear classification methods (e.g., linear discriminant analysis) or nonlinear ones (e.g., neural networks). As shown in Figure 1-4, a feature translator may consist of two components: ‘feature classification’ and ‘post-processing’. The main aim of the feature classification component is to classify the features into logical control signals. Post-processing methods such as a moving average may be used after feature classification to reduce the number of activations of the system.

The control interface translates the logical control signals from the feature translator into semantic control signals that are appropriate for the particular type of device used. Finally, the device controller translates the semantic control signals into physical control signals that are used by the device. For more detail refer to [3].

In the next section, we provide a brief review of some of the work done in the literature.

## **1.3 Background**

Since the introduction of the concept of BCI control in early 70’s (e.g., [12]), many BCI systems have been developed. Despite these efforts, many design issues remain under debate. In this section, we briefly review these design issues.

### **1.3.1 Signal recording**

An activity in a normal human brain can generate various responses including electrical, magnetic, and metabolic responses. These signals can be detected by appropriate sensors and they can be used for controlling a BCI system. For example, brain activity can produce magnetic fields that can be recorded using

magnetoencephalographic (MEG) activity. Brain activity can also result in some metabolic consequences in terms of changes in the blood flow and metabolism. Imaging methods such as functional magnetic resonance imaging (fMRI) can image these activities. At present, because of the cost and physical dimensions, methods that measure the electrical activities of the brain are more favored [1].

There are various ways to record the electrical activities of the brain. Non-invasive BCI approaches mostly use the EEG signals as the source of information. EEG signals are recorded by means of electrodes placed on the scalp. Invasive approaches, on the other hand, use electrocorticography (ECoG) signals recorded from the surface of the brain or action potentials of single neurons in the cerebral cortices, using implanted microelectrodes.

EEG signals have good temporal resolution, but their spatial resolution is not good compared to other recording technology methods [1]. A recent study showed that only 12% of published BCI studies use implanted electrodes, 5% use microelectrode arrays, and more than 80% use EEG signals [3]. The main reason is that the EEG recording equipment is commercially produced and their cost is lower than other brain signal recording technologies. Also, since no surgery is necessary for placing electrodes, more individuals are willing to participate in such BCI experiments.

### **1.3.2 Choice of neurological phenomenon**

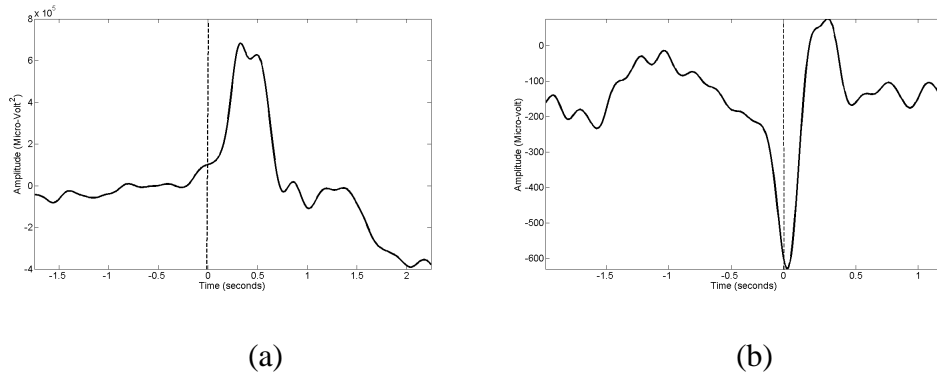
Neurological phenomena are specific features of the brain activity that appear in the brain signals and can be used to control a BCI system. They are time-locked to a physical stimulus or to the cognitive responses of the brain. Neurological phenomena are characterized by their voltage amplitude, their latency which is related to the internal or external stimuli and their spatiotemporal distribution. Their amplitude is usually much smaller than the background EEG signal.

The more common neurological phenomena in BCI systems are:

*Changes in the brain rhythms (CBR) such as the Mu, and Beta rhythms related to a movement:* Mu ([8-12] Hz) and Beta ([18-30] Hz) rhythms are frequency bands in the brain signals that are known to be suitable neurological phenomena for controlling BCI

systems. The reason is that they are closely associated with those cortical areas directly connected to normal motor channels of the brain. Voluntary movement results in a circumscribed desynchronization in the Mu and Beta bands that are localized close to the sensorimotor areas ([13, 14]). This desynchronization, termed “event-related desynchronization (ERD)”, starts about two seconds prior to the onset of movement [15]. The enhanced rhythmic activity following the movement is called “event-related synchronization (ERS)”. The post-movement Beta ERS is found in the first second after the termination of a voluntary movement, when the Mu rhythm might still display a desynchronized pattern [15]. The Beta ERS is a relatively robust phenomenon and is found in nearly all users after a finger, hand, arm or foot movement (see Figure 1-5 (a) for an example of the Beta ERS) [16].

Many research groups have developed BCI systems using the features extracted from the Mu and Beta rhythms. However, the works of two research groups are more prominent. Wolpaw and McFarland and their associates in Wadsworth Center have focused on developing such a CBR-based synchronized BCI system. Their proposed BCI system allows users to control the amplitude of Mu and Beta rhythms. This amplitude is then used to move a cursor on the computer screen [17-20]. Users of this system usually need training that may take up to a few weeks, but eventually they can achieve high accuracies (e.g., above 90%) [21]. The other research group, the Graz BCI, uses the ERD and the ERS of the Mu/Beta rhythms in the design of synchronized BCI systems[22-26]. Similar to the first group, after a few sessions of training, the users of the Graz BCI can also achieve high accuracies.



**Figure 1-5. Two examples of neurological phenomena. (a) Changes in the power of Beta rhythms over time; (b) A movement-related potential. Vertical line shows the time of activation of the movement. Note that these shapes are generated by averaging over many epochs.**

*Movement related potentials (MRPs):* Averaging the EEG data with respect to movement onset results in the generation of slow potentials called “movement-related potentials” (MRPs) [27]. MRPs start about 1.5–1 seconds before the onset of a particular movement and have bilateral distribution (see Figure 1-5 (b) for an example of an MRP) [27-31]. High-resolution EEG studies have modeled the main sources of MRPs arising in the supplementary motor area and the primary sensorimotor cortex [32, 33]. MRPs have been used for the neurological phenomenon in several BCI studies. These studies include the work that has been carried out by Mason and Birch’s research group [5, 7, 34-36], Muller and Blankertz et. Al [37, 38] as well as Yom-tov and Inbar [6, 39, 40] .

*Other movement related activities (OMRAs):* We categorized the movement-related activities that do not belong to any of the preceding categories as OMRA. OMRAs are usually not restricted to a particular frequency band or scalp location and usually cover different frequency ranges. They may be a combination of specific and non-specific neurological phenomena. Levine and Huggins’ research group are amongst the prominent research groups that have used OMRAs related to different movements to design their ECoG-based BCI system. They recorded ECoG activity from patients with 16-126 subdural electrodes prior to an epilepsy surgery. They have used topographically focused potentials associated with different movements to develop various 2-state self-paced BCI designs [8, 41, 42].

*Slow cortical potentials (SCPs):* SCPs are slow usually non-movement potential changes generated by the user. They reflect changes in the cortical polarization of the EEG, lasting from 300 ms up to several seconds [2, 43]. Birbaumer and his colleagues have developed a BCI system called “Thought Translation Device (TTD)” that uses an SCP as the source of control [44-47]. They have shown that patients with severe motor disabilities such as late-stage ALS can learn to control their SCPs and thus use TTD to communicate with the outside world.

*Cognitive tasks (CTs):* Changes in the brain signals as a result of non-movement mental tasks (e.g., mental counting, solving a multiplication problem) are usually categorized as CTs [48]. The works of Milan et.al [49] and Anderson et. al [50] are amongst the prominent BCI research carried out using cognitive tasks. Millan et.al’s work involves using the mental tasks to control a mobile robot, while Anderson et.al have focused on the design of a multi-class BCI system that detects cognitive tasks associated with different tasks such as 3D object recognition, mental counting, etc [51, 52].

*P300:* When infrequent or particularly significant auditory, visual or somatosensory stimuli are interspersed with frequent or routine stimuli, they evoke a positive peak at about 300 ms after the stimulus is received. This peak is called P300 [48, 53]. Using this so-called “oddball” response, Donchin and his colleagues have used P300 to build a successful BCI system [54, 55]. More recently, a number of studies have shown that P300-based control can be used as an alternative communication channel for people with spinal cord injury and ALS [56, 57]. Also, for individuals with visual impairments, solutions based on auditory or tactile stimuli have been proposed [58, 59].

*Visual evoked potentials (VEP):* VEPs are small changes in the brain signal, generated in response to a visual stimulus such as flashing lights. They display properties whose characteristics depend on the type of the visual stimulus [48]. Many BCI systems use VEPs to control the BCI system including the works of Vidal [60], Sutter [61] and Middendorf [62].

*Steady-State visual evoked potentials (SSVEP):* If a visual stimulus is presented repetitively at a rate of 5-6 Hz or greater, a continuous oscillatory electrical response is elicited in the visual pathways. Such a response is termed SSVEP. The distinction



between VEP and SSVEP depends on the repetition rate of the stimulation [63]. The works carried out by Gao's research group are noticeable in this area [63-66].

*Multiple neurological phenomena (MNs):* BCI systems based on multiple neurological phenomena use a combination of two or more of the above neurological phenomena for the purpose of control. We will review this category of BCI systems in more details later in this chapter.

*Activity of neural cells (ANC):* Some BCI research groups have used microelectrode arrays to record the activity of single neurons in the motor cortex for the purpose of BCI control [67-73]. These BCI systems are usually based on reconstructing a movement from recorded spike trains. Experiments with monkeys have shown a relatively good ability of control in multiple directions in these systems [74]. Recently, there have been reports of a patient learning to use his neuronal activity to move a computer cursor to several directions using the ANCs [73]. These encouraging results provide hope for BCI control with multiple options and high accuracy. The downside is the invasive nature of the microelectrode implants, which may result in infection and side effects in the brain.

The above neurological phenomena can be categorized in two groups based on the origin of the phenomenon in the brain. Those neurological phenomena generated as the result of cognitive responses of the brain are called endogenous. The ones evoked by an external stimulus are called exogenous.

BCI systems that use exogenous neurological phenomena, usually do not need any user training [54]. The downside of using these systems is that they require a constant commitment of one of sensory pathways to an external stimulus [75]. Furthermore, not all users may tolerate repetitive sensory stimulation. On the other hand, endogenous-based BCI systems rely on the generation of a phenomenon that is more natural and is thus expected to cause the users less fatigue. This may be the reason why more than 80% of BCI studies use endogenous neurological phenomena to control BCI systems [3].

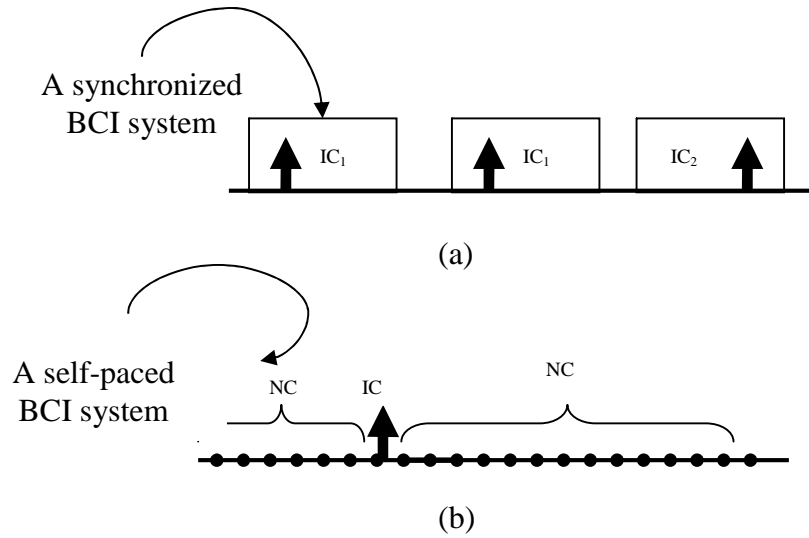
To generate a suitable neurological phenomenon, endogenous-based BCI systems usually need user training. This training may take a long time, sometimes even up to few months. The use of complex signal processing schemes for detecting weak neurological

phenomena can greatly reduce or even eliminate the training process [76]. Another advantage of employing endogenous neurological phenomena is that it is possible to select and use a combination of some of them to improve the performance of the system.

### **1.3.3 Timing of BCI control**

So far most BCI researchers have focused their attention on “synchronized” control applications. In synchronized applications, a user can initiate a command only during specific times specified by the system (see Figure 1-6(a)). In these systems, the users are required to generate an intentional control (IC) command during the periods allowed by the BCI system. In the example shown in Figure 1-6 (a), the user should generate one of  $IC_1$  or  $IC_2$  commands during the control period (the control period is shown as a ‘box’).

In contrast, in self-paced BCI system, a user does not need to be constantly engaged in initiating the control command. In these systems, the users only consciously control their state when they desire to control the device (see Figure 1-6(b)) [35]. In the example shown in Figure 1-6 (b), the user is in the no control (NC) state at all times, except for those periods when he/she initiates an IC command. In the latter case, the system will be in an IC state. During NC periods, the user can be idle, thinking about a problem or performing any action other than attempting to control the device. This property of SBCI systems that allows them to support the presence of NC periods is called “NC support”. Whenever a BCI system involves control actions with periods of inaction, it needs to have NC support.



**Figure 1-6. Synchronized vs. self-paced control. (a) In a synchronized BCI system, control can be done only in certain intervals specified by the system; (b) In a self-paced BCI system, the control is done at the user's own pace.**

Synchronized BCI systems usually require the user to initiate an IC command during the control periods. In other words, during the control periods, the users are expected to be engaged with controlling the device. For this reason, they usually do not support the “NC” periods. In some cases, the output of the system might even become unstable if an IC command is not issued.

In the next three sections, we briefly describe the literature related to this thesis. First, in Section 1.4, the previous work pertaining to the design of SBCI systems is discussed. In Section 1.5, we address the use of more than one neurological phenomenon as the control source. Then in Section 1.6, we address the previous work on handling artifacts in the design of SBCI systems. Finally in Section 1.7, we discuss the previous research related to evaluating the performance of SBCI systems.

## 1.4 Design of self-paced BCI systems

Self-paced BCI systems provide the user more freedom and more control flexibility. From the signal processing point of view, they are much more challenging to design compared to synchronized BCI systems. The main reason is that there are many

types of NC states (e.g., idle, different mental tasks, etc.). As a result, an SBCI system should be able to handle various types of different NC signals, once the SBCI system is turned on. For this reason, only a few BCI groups have pursued the design of self-paced BCI systems [7, 36, 40, 77, 78].

The concept of self-paced control started in early 90's with the development of the outlier processing method (OPM), which aimed at detecting movement-related potentials (MRPs) in the EEG signals [36]. The results from this work were promising as true positive (TP) rates greater than 90% were achieved on a thumb movement task. However, its poor performance over NC epochs (FP rates ranging from 10% to 30%) restricted its use as a BCI system.

To overcome the vulnerabilities of OPM, another SBCI system called the low frequency- asynchronous switch design (LF-ASD) was later proposed in 2000 by Mason and Birch [35]. Similar to OPM, LF-ASD is also designed to detect MRPs in the EEG signals. It uses features extracted from the 0.1- 4Hz band in six bipolar EEG channels recorded from  $F_1-FC_1$ ,  $F_z-FC_z$ ,  $F_2-FC_2$ ,  $FC_1-C_1$ ,  $FC_z-C_z$  and  $FC_2-C_2$  on the scalp, sampled at 128 Hz. A detector that was a simplified version of the discrete wavelet transform was applied as the feature extractor and a 1-nearest neighbor (1-NN) classifier was used as the feature classifier. By analyzing the EEG signals of five individuals, the features related to MRP (or IC) periods showed a definite difference from those in NC periods [35]. During the past few years, several changes have been applied to the structure of LF-ASD to improve its performance. These changes include the addition of an energy normalization transform [79], the addition of a debounce window as a post-processing component to decrease the FP rate [5], user-customization of the feature extractor's parameter values[80], and adding the knowledge of the past paths of features [81].

Despite these improvements, the performance of the LF-ASD is still not suitable for many practical applications. The most recent design of the LF-ASD achieves an average TP rate of 54.0 % at the false positive rate of 1%[82]. Since LF-ASD generates an output every  $1/16^{\text{th}}$  of a second, this is translated into, on average, one false positive every 6.25 seconds, while the detection rate of IC commands is less than 50%. For most

practical applications, generating such a high FP rate, may result in excessive user frustration.

Another SBCI design, which improves upon the feature extractor of LF-ASD, is proposed by Yom-tov et.al[6]. The proposed method combines the LF-ASD feature extractor with a matched filter, resulting in a hybrid detector. This method also results in a high FP rate. For FP rates <2%, the TP rates are lower than 30%. This system generates an output every  $1/25^{\text{th}}$  of a second. An FP rate of 2% is translated into one false positive, every two seconds. As a result, the high amount of FPs limits the application of the proposed design.

While the above studies are based on features extracted from EEG signals, researchers from the University of Michigan have focused on extracting features from ECoG signals [8, 41, 77, 83-85]. To detect IC commands, their designs either use the cross-correlation with a template [83] or the energy of wavelet packet transform [8]. In these studies, a threshold-based classifier is used for classifying the features. While these systems usually achieve  $TP > 50\%$ , their performance on NC epochs is not very clear. First, none of these studies has determined the number of NC epochs. Moreover, to quantify the false positives, a new metric called the *false discovery rate* (FDR, i.e., the percentage of total activations of the switch that were false) was used [86]. Since the number and the length of NC epochs is not determined in these studies, it is impossible to calculate the FP rate for these systems. In a recent study by this group, the reported FDR were in the range of 0% to 82% with 24 out of the 31 reported FDRs being higher than 10% [8]. However, since the numbers of IC and NC epochs were not specifically determined, no comment can be made on the performance of these systems over NC data.

Table 1-1 compares the TPR and FPRs achieved in selected SBCI studies. Please keep in mind that although a direct comparison is not possible, this table roughly shows the performances of some of the existing SBCI systems. The rows of this table show the different SBCI studies. The columns show the rate at which the system generates an output, TPR, and FPR, respectively. As shown, with the exception of the first study that uses ECoG signals, the rest of these studies, have low TPR for  $FPR < 2\%$ . Please note that,

given the rates at which that these SBCI systems generate outputs, these FP rates are translated into one false positive every few seconds on average.

**Table 1-1. Comparison of the TPR and FPR rates achieved in different SBCI studies.**

Paper\Study		Frequency	TPR(%)	FPR(%)
Graimann, et.al [8]		100	Up to 100%	?
Mason and Birch[35]	LF-ASD	16	<20%	2
	OPM		<10%	
	Mu-ASD <sup>*</sup>		<10%	
Yom-tov and Inbar[6]		25	30%	2
Townsend et.al [87]		?	<20%	2
Bashashati, et.al [82]		16	54.0	1

\* Mu-ASD is a self-paced BCI system that uses Mu rhythms as the neurological phenomenon.

A review of self-paced endogenous BCI studies shows that with the exception of one paper [8] (which will be discussed in more detail later in this section), all the proposed designs have relied on a single neurological phenomena. In the next Section, we bring evidence from the literature that supports the advantage of using the following three neurological phenomena (instead of only one) in a self-paced BCI system: MRPs, changes in the power of the Mu and Beta rhythms.

## **1.5 Use of multiple neurological phenomena in BCI systems**

### **1.5.1 Simultaneous application of MRPs and changes in the power of Mu/Beta rhythms**

A number of papers provide some evidence that MRPs and changes in the power of brain rhythms [usually characterized as the event-related desynchronization (ERD) and event-related synchronization (ERS)] provide complementary information for exploring the cognitive functions of the brain. In [88], the analysis of subdural EEG recordings from primary sensorimotor in epileptic patients showed that the amplitude of the ERD of the Alpha rhythm recorded from subdural areas was not always correlated with the corresponding MRPs. It is suggested in the same paper that these neurological phenomena represent different aspects of cortical motor processes. In [89], the ERD of the Alpha rhythm is not always detected in cortical sites generating MRPs. In [31], through a high-resolution EEG study, it is shown that MRPs and the ERD of the Mu rhythm provide complementary information on human brain responses accompanying the preparation and execution of a finger movement. Further evidence from the analysis of EEG signals [90, 91] and magnetoencephalography (MEG)[92-94] strengthens these findings.

There is also some evidence regarding the differences between the Mu and the Beta rhythms. Several papers show that the reactivities of the Mu and Beta rhythms related to the movement onset are different [95, 96]. Both the Mu and Beta rhythms desynchronize before the occurrence of a voluntary self-paced movement. However, after the movement, the ERD of the Mu rhythm is usually followed by a slow return to baseline (and sometimes by a slight synchronization), while the Beta rhythms synchronize rapidly after the movement onset [96].

This evidence from the literature shows that MRPs, Mu and Beta rhythms provide complementary information that can be used for improving the performance of BCI systems. In the next sub-section, we review the simultaneous use of these phenomena in the BCI literature.

### 1.5.2 Using multiple neurological phenomena in BCI systems

The main advantage of using more than one neurological phenomenon at the same time is that more information is available for the BCI system to detect an IC command related to a particular movement. The downside is that as the size of the input data increases, the complexity of the pattern recognition algorithm increases as well.

Although most BCI researchers use a single neurological phenomenon as the source of control, there have been reports of using multiple neurological phenomena in BCI systems [8, 76, 92, 97-99]. In [92], the authors analyzed different combinations of 1) features extracted from an early component of the MRP called Bereitschaftspotential (BP), 2) features extracted from the ERD of neurological phenomena above 4Hz (through autoregressive modeling) and 3) features extracted from the common spatial patterns (CSP) features related to the ERD of Mu rhythms. The BCI system had to discriminate between left and right index finger movements. A linear discriminant analysis (LDA) classifier was used for classification. Different combination schemes were explored. The study showed that a certain combination of classifiers could result in a lower error rate than the case where a single classifier is used. The results of combining the ERD of the Mu rhythm and the BP were not reported, although the authors mention that those results were slightly worse than the results obtained when all three neurological phenomena were used in the design of the BCI system. In [97], the authors applied a combination of microstate analysis and common spatial subspace decomposition to extract features belonging to three different frequency bands: Theta + Delta, Mu and Beta. MRPs were not treated as a separate neurological phenomenon. Instead, features were extracted from the frequency band covering both the Delta and Theta rhythms. These features were then used to discriminate between left and right hand movements. Using data of three participants, the proposed method achieved an average accuracy higher than 80%. In [100], the authors used the BP and the ERD of the brain rhythms in the 10 to 33 Hz frequency band (including both the Mu and Beta rhythms) to classify left vs. right finger movement. The features extracted from all neurological phenomena and from all EEG channels were then combined, the dimension of the feature vector was reduced and the final vector was classified using a perceptron neural network. The results showed classification accuracy of 84% on the test set. In [101], the authors used features



extracted from the BP and the ERD of the Mu rhythms for classifying the left and right index finger movements. It was shown that combining features results in decreasing the classification error for four out of five subjects whose data were studied.

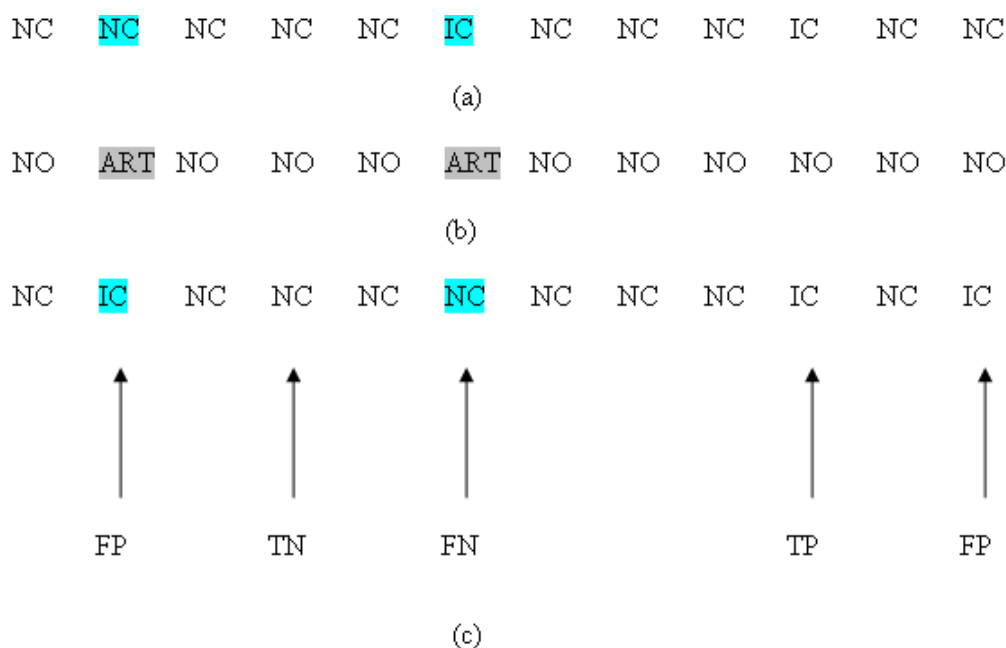
The above studies all pertain to synchronized BCI systems. Only one SBCI system that uses multiple neurological phenomena has been reported so far [8]. In this study, the authors combined a number of neurological phenomena in order to design an ECoG-based SBCI system. Using a wavelet packet transform, ECoG signals were divided into 18 different frequency bands covering the range from 0 - 100 Hz. This range covered a wide range of neurological phenomena including Mu, Beta and Gamma rhythms, as well as other movement-related activities (OMRAs). Then for each band, wavelet-filtered signals were reconstructed. The wavelet filtered signals were then squared to achieve power values, and a genetic algorithm was applied to reduce the dimension of the feature space to one. Using a thresholding classifier, the test samples were classified as movement or no movement. As mentioned earlier in Table 1-1, the reported false discovery rates of this study were in the range of 0% to 82% with 24 out of the 31 reported FDRs being higher than 10%. This study, however, did not consider MRPs as one of the neurological phenomena. Instead it solely focused on detecting the power of different frequency bands. Furthermore, it extracted features from ECoG signals instead of EEG signals. As noted earlier in Section 1.3.1, recording ECoG signals needs surgery and it has an invasive nature. For this reason, this method of recording brain signals may not be fully accepted by the research community until the health-related issues are fully investigated.

As we will discuss in Section 1.9, in this thesis we will design a new SBCI system that simultaneously uses information extracted from MRPs as well as changes in the power of the Mu and the Beta rhythms. To the best of our knowledge, this is the first time in the BCI literature that such a study is carried out in the context of self-paced BCI systems.

## 1.6 Artifacts in BCI systems

Artifacts are undesirable potentials that contaminate brain signals and are mostly of non-cerebral origin. Unfortunately, they can modify the shape of a neurological phenomenon that drives a BCI system. They can also mistakenly result in an unintentional control of the device [102]. Therefore, there is a need to avoid, reject or remove artifacts from the recordings of brain signals.

In an SBCI system, artifacts can impact the performance of the system in two ways: 1) by changing the shape of the neurological phenomenon during an IC period, they cause a decrease in the TP rate. 2) By mimicking the shape/properties of the neurological phenomenon during the NC periods, artifacts results in an increase in the FP rates.



**Figure 1-7. An example of how artifacts can affect the performance of an SBCI system. (a) The brain state of the user; (b) The periods when artifacts have occurred; (c) The output of the SBCI system (note: FP: false positive, TN: true negative, FN: false negative and TP: true positive).**

Figure 1-7 shows how this can happen. Figure 1-7 (a) shows the brain states of a user during a specific time frame. As seen, the user is in an NC state, however, at two time instants the user initiates an IC command. Figure 1-7 (b) shows the periods of EEG signals that are contaminated with artifacts. The term “ART” denotes “artifact-

contaminated periods” and “NO” refers to the periods not contaminated with artifacts. The second period coincides with the first IC command. Figure 1-7 (c) shows the output of the SBCI system. The occurrence of the first artifact results in a false positive. The second artifact, results in masking the first IC command (a false negative or FN).

Artifacts originate from non-physiological as well as physiological sources. Non-physiological artifacts originate from outside the human body (such as 50/60 Hz power-line noise or changes in electrode impedances), and are usually avoided by proper filtering, shielding, etc.

Physiological artifacts arise from a variety of bodily activities. Electrocardiography (ECG) artifacts are caused by heart beats and may introduce a rhythmic activity into the EEG signal. Respiration can also cause artifacts by introducing a rhythmic activity that is synchronized with the body’s respiratory movements. Skin responses such as sweating may alter the impedance of electrodes and cause artifacts in the EEG signals [103]. The two physiological artifacts that have been most examined in BCI studies, however, are ocular (Electrooculography or EOG) and muscle (Electromyography or EMG) artifacts.

EOG artifacts are generally high-amplitude patterns in the brain signal caused by blinking of the eyes, or low-frequency patterns caused by movements (such as rolling) of the eyes [104]. EOG activity has a wide frequency range, being maximal at frequencies below 4Hz, and is most prominent over the anterior head regions [105].

EMG activity (movement of the head, body, jaw or tongue) can cause large disturbances in the brain signal. EMG activity has a wide frequency range, being maximal at frequencies higher than 30 Hz [104, 105]. Difficult tasks may cause an increase in EMG activity related to the movement of facial muscles [106, 107].

Some studies have shown that EOG and EMG activities may generate artifacts that affect the neurological phenomena used in a BCI system [108, 109]. For example, [109] demonstrated that brain rhythms are contaminated with EMG artifacts during the early training sessions of their proposed BCI system that used Mu and Beta rhythms as sources of control.

Physiological artifacts such as EOG and EMG artifacts are much more challenging to handle than non-physiological ones. Moreover, controlling them during the signal acquisition stage is not easy. There are different ways of handling artifacts in BCI systems. In this Section, we briefly examine the reported methods for handling EOG and EMG artifacts, as these are among the most important sources of contamination in BCI systems.

### **1.6.1 Artifact avoidance**

The first step in handling artifacts is to avoid their occurrence by issuing proper instructions to users. For example, users are instructed to avoid blinking or moving their bodies during the experiments.

Instructing individuals to avoid generating artifacts during data collection has the advantage of being the least computationally demanding among the artifact handling methods, since it is assumed that no artifact is present in the signal (or that the presence of artifacts is minimal). However, it has several drawbacks. First, since many physiological signals, such as the heart beats, are involuntary, artifacts will always be present in brain signals. Even in the case of EOG and EMG activities, it is not easy to control eye and other movement activities during the process of data recording. Second, the occurrence of ocular and muscle activity during an online operation of any BCI system is not avoidable. Third, collecting sufficient amount of data without artifacts may be difficult, especially in cases where a user has a neurological disability [110]. Finally, avoiding artifacts may introduce an additional cognitive task for the individual. For example, it has been shown that refraining from eye blinking results in changes in the amplitude of some evoked potentials [111, 112].

### **1.6.2 Artifact rejection**

Artifact rejection refers to the process of rejecting the trials affected by artifacts. It is perhaps the simplest way of dealing with brain signals contaminated with artifacts. It has some important advantages over the “artifact avoidance” approach. For example, it would be easier for individuals to participate in the experiments and perform the required tasks, especially those individuals with motor disabilities. Also, the “secondary”

cognitive task, resulting from an individual trying to avoid generating a particular artifact, will not be present in the EEG signal.

”Artifact rejection” is usually done by visually inspecting the EEG or the artifact signals, or by using an automatic detection method [113].

### **Manual rejection**

Manual rejection of epochs contaminated with artifacts is a common practice in the BCI field. Trials are visually checked by an expert, and those that are contaminated with artifacts are removed from the analysis.

Similar to “artifact avoidance”, manual rejection also has the advantage of not being computationally demanding, as it is assumed that a human expert has identified all the artifact-contaminated epochs and removed them from the analysis. On the other hand, there are many disadvantages in using “manual rejection”. First, “manual rejection” comes at the cost of intensive human labor, especially if the study involves a large number of individuals or a large amount of recorded data. Second, the process of selecting the artifact-free trials may become subjective. It has been argued that because of the selection bias, the sample trials that are artifact-free may not be representative of the entire population of the trials [113]. Third, in the case of offline analysis, the rejection of artifact-contaminated trials, may lead to a substantial loss of data. This may become a huge drawback, especially in the case of individuals with motor disabilities, where offline data recording is not as convenient as it is for able-bodied individuals.

### **Automatic rejection**

In the “automatic rejection”, the BCI system automatically discards the epochs of brain signals that are contaminated with particular artifacts. This procedure is commonly carried out in offline investigations.

Automatic rejection of epochs can be done in the following two ways:

***Rejection using the EOG (EMG) signal:*** When one of the characteristics of the EOG (EMG) signal in an epoch exceeds a pre-determined threshold, the epoch is considered as artifact-contaminated and is automatically rejected.

***Rejection using the EEG signal:*** This rejection methodology is similar to the above; only the EEG signal is used instead of the EOG (EMG) signal. This approach has the advantage of being independent of the EOG (EMG) signal, and is useful if the EOG (EMG) signal is not recorded during data collection.

An advantage of the “automatic rejection” approach over that of “manual rejection” is that it is less labor intensive. However, automatic rejection still suffers from loss of valuable data [114, 115]. In the case of EOG artifacts, the automatic rejection approach also does not allow the rejection of contaminated trials when the EOG amplitude is small [116, 117].

Two issues need to be addressed for the BCI systems which reject artifacts:

Because of the vast number of artifacts that exist in BCI systems (eye blinking, eye movements, movements of different parts of the body, breathing, etc.), not all the artifact-contaminated trials can be rejected. Usually only the epochs with a strong presence of artifacts are excluded from the analysis. Therefore, the so-called “clean” data are unfortunately not completely free of artifacts.

The second issue is that the rejection of artifact-contaminated data during an offline analysis may generate “cleaner” data. However, for online real-time applications of a BCI system, this may pose a huge drawback. In online applications, artifacts are unavoidable. If artifacts are rejected during the offline analysis, the same rejection mechanism can be used to reject them during the online analysis. The only problem is that during the specific time periods when artifact-contaminated signals are rejected, the system is unreachable and cannot be used for controlling the device.

### **1.6.3 Artifact removal**

Artifact removal is the process of identifying and removing artifacts from brain signals. An artifact-removal method should be able to remove the artifacts as well as keep the related neurological phenomenon intact. Common methods for removing the artifacts in EEG signals are linear filtering [118, 119], linear combination and regression [116], blind source separation [120], principal component analysis [121], wavelet transform [122], nonlinear adaptive filtering [123] and source dipole analysis (SDA) [124].

A survey of all BCI studies published before January 2006 shows that most BCI papers do not report whether or not they have considered EMG and/or EOG artifacts in their analysis. This is an important issue, since offline analysis methods that do not account for physiological artifacts may probably face some problems when tested during an online study. As a result, it is important that BCI researchers pay more attention to this important issue and address the method that they have employed for handling artifacts.

A number of BCI studies state that EMG activity will not be present in the EEG signal when the EEG signal is analyzed before a movement has occurred [125]. This argument may not be valid for BCI systems. This is because peripheral changes such as EMG tension can affect the EEG signal, even though the amount by which the EEG signal is affected remains unclear [126]. It is pointed out in [126] that even when the individuals are very restricted, they still preserve motor control over some muscle groups. Although the activities of several muscle groups are monitored in BCI studies, there remain some muscles whose activities are not recorded.

The BCI systems that employ “manual rejection” of EOG and EMG artifacts should also consider the fact that “manual rejection” is only a preliminary step in the design of a BCI system. “Manual rejection” can only be used for offline analysis. In order for a particular BCI system to work in an online fashion, a scheme for handling artifacts should be incorporated. Requesting the individuals to avoid artifacts should be only considered as a temporary solution. In a practical application, EMG and EOG artifacts do happen, so methods of handling these artifacts during an online experiment should be investigated.

One solution for handling artifacts, which is not explored well in the BCI studies, is to design a BCI that is robust in the presence of artifacts. If such a BCI is designed, then the need for having a method of handling artifacts will be minimized. Another solution that has not been explored well in the BCI literature, is that of using more than one neurological phenomenon may lead to increasing the robustness to the occurrence of artifacts[76]. Since EOG artifacts mostly affect the low-frequency components of the EEG signals, BCI systems that use low-frequency ERPs, such as MRP and SCP are mostly affected by EOG artifacts. EMG artifacts on the other hand, mainly affect the

high-frequency components of the EEG signals, hence BCI systems that use high-frequency ERPs, such as Mu and Beta rhythms are mostly affected by EMG artifacts. Thus, it can be concluded that a BCI system that uses multiple neurological phenomena from whose frequency span both the low as well as the high frequency bands, may become more robust to the presence of artifacts.

## 1.7 Evaluating the performance of SBCI systems

Model selection is the process of finding or adjusting the model parameters for any classification problem. For BCI systems, model selection is a crucial part of the design. This process may include selecting the features, the type of the feature extractor, the classifier, the EEG channels, the neurological phenomenon, the frequency band of interest, the values of the classifier's parameters and the preprocessing and post-processing components. As an example, to find the optimal set of features for a certain BCI, different sets of features are considered. For every set, the performance of the system is calculated and different performances are compared. The set of features that yields the best performance is then selected. The performance of this best model can then be compared with those achieved by similar BCI systems (i.e., systems with the same experimental as well as evaluation protocols). Therefore, the performance of an SBCI must be evaluated in the following two cases, 1) during the model selection procedure and 2) when comparing the performance with other systems.

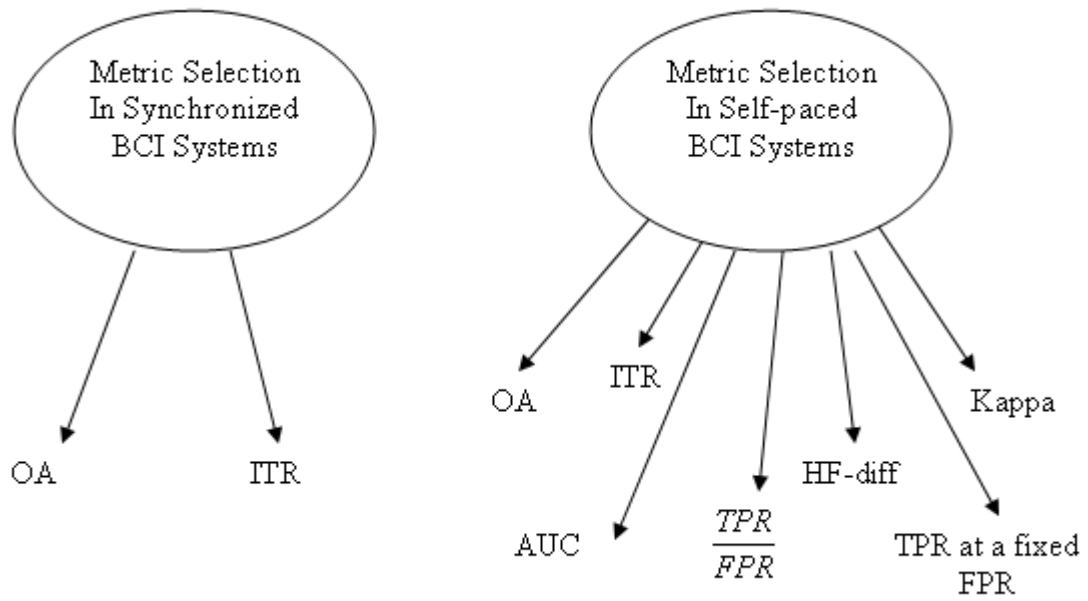
The performance of a BCI with discrete states is usually summarized by a confusion matrix. The  $(i,j)$  entry of this matrix represents the number of samples from class  $i$  that are classified as belonging to class  $j$ . A confusion matrix provides valuable information regarding how well each class is classified by the BCI system. It is, however, not usually straightforward to compare different confusion matrices. Evaluation metrics are thus needed to summarize a confusion matrix into a single value. For classification problems with balanced datasets such as synchronized BCI systems (where  $prob(Class_1) \approx prob(Class_2) \approx \dots \approx prob(Class_N)$  for an  $N$ -class problem), the overall classification accuracy (OA) is the most common evaluation metric presently used to summarize the performance [10]. The use of OA for problems with highly



imbalanced classes (e.g.,  $prob(Class_1) \gg prob(Class_2)$  for a two-class problem) is not satisfactory [127].

The choice of the evaluation metric is of great importance and is application-dependent. A poorly defined evaluation metric may guide the model selection procedure to a far-from-optimal model or it can lead to erroneous conclusions when comparing the performances of two SBCI systems. As a result, all the effort spent in the design of a sophisticated SBCI may be lost, simply because of the poor choice of the evaluation metric. Recently, the choice of OA as the default evaluation metric has been questioned, even in classification applications with balanced datasets. Specifically, it was shown that for many applications, the area under the receiver operating characteristic (AUC) can summarize the performance better than OA [128].

Although OA is not suitable for classification problems with imbalanced classes, the choice of an alternative evaluation metric is not obvious. Several attempts have been made to define more suitable evaluation metrics for these problems. Examples of such evaluation metrics include weighted overall accuracy (WOA) [129], the use of receiver operating characteristic (ROC) curves and related measures such as area under the ROC (AUC) [130] and the Kappa coefficient [131]. In the SBCI literature, some of the evaluation metrics used include overall accuracy [7], HF-difference[8], mutual information (information transfer rate) [9], Kappa [10], AUC [10], the true positive rate (TPR) at a fixed false positive rate (FPR) [5] and  $\frac{TPR}{FPR}$  [132]. Figure 1-8 shows the proposed evaluation metrics for synchronized and self-paced BCI systems. As seen, the number of proposed evaluation metrics is significantly higher for self-paced BCI systems than synchronized BCI systems.



**Figure 1-8. Types of evaluation metrics used in synchronized and self-paced BCI systems.**

OA shows the total number of test samples correctly classified by an SBCI system. It has been frequently used in evaluating many synchronized BCI systems [133-135]. Its use in SBCIs, however, has so far been limited [7]. This is because, for an SBCI system, OA assigns a huge weight on the more frequent class (NC) and a very small weight on the less frequent classes (IC). This may lead to misleading conclusions about the performance of the system.

The information transfer rate (ITR) has been specifically proposed for evaluating the performance of synchronized BCI systems [136]. This metric is proposed based on the similarities between an SBCI and a communication channel, and using Shannon's communication theory. The rationale is that ITR measures the amount of information transferred between two reference points. The output  $\mathbf{Y}$  of an SBCI is the interpretation (information) of the current state of the brain, and  $\mathbf{Y}$  conveys this information to the downstream components. It is thus argued in [136] that the amount of information in  $\mathbf{Y}$  is a useful tool for comparing the results obtained from different synchronized BCI designs. It is also argued that ITR by itself is "not" a suitable single evaluation metric for an SBCI system. This is because of the unique nature of this metric having more than one maximum (see [137] for a detailed discussion).

Cohen's Kappa coefficient is a measure of agreement between two estimators [138]. Since it considers chance agreements, it is regarded as a more robust measure than OA [10].

The HF-difference is a newly proposed metric that summarizes the confusion matrix [85]. It is defined as the difference between the TP rate and the percentage of total activations that are incorrect (the false discovery rate (FDR)[86]). The advantage of using HF-difference is that it is sensitive to the ratio of FPs to the total number of detections. The downside of using the HF-difference is that it does not consider the length of NC periods.

The  $\frac{TPR}{FPR}$  is another evaluation metric that was recently proposed for 2-class SBCI systems [132, 139]. This metric gives more weight to cases with low FPRs. As a result, during the model tuning process, any model with a high FPR is assigned a low fitness, even though TPR might have a high value. The downside is that for  $FPR=0$ , the system cannot differentiate between confusion matrices with different TPRs.

The receiver operating characteristics (ROC) curve is a popular metric for evaluating systems with imbalanced classes. The ROC curve depicts the relationship between TPR and FPR. Popular methods that use the ROC curve for measuring the performance employ one of the following two criteria 1) The area under the ROC curve (AUC) which is used as the fitness of the system [10]; 2) Defining a critical FPR value ( $FPR_{Critical}$ ), and then using the value of the TP rate at  $FPR_{Critical}$  as the fitness [5]. The advantage of using the ROC curve over previous metrics is that a whole range of solutions (in terms of a tradeoff between TPR and FPR) is provided.

One problem with using the ROC curve is that when it is plotted over the whole range of TPR and FPR, most SBCI systems produce a curve that is similar to a perfect ROC curve [4]. The other problem with using the ROC curve (and perhaps more important) is that it is computationally more demanding than other evaluation metrics. Several points need to be evaluated until a partial ROC curve that is accurate enough for estimating the AUC is drawn. Similarly, several points need to be calculated in order to obtain the value of TPR at  $FPR_{Critical}$ . Even if the ROC curve is estimated using the more

computationally efficient algorithm as described in [128], it remains much more time consuming than the metrics described above as these only need the value of a single point to assess the performance. When these metrics are used to evaluate the performance and select a model from thousands of confusion matrices during a model selection procedure, the computational burden becomes problematic. For these reasons, evaluation metrics that summarize the performance based on a single evaluation of a confusion matrix are more desirable during the model selection procedure.

Each of these metrics has strengths and weaknesses [10], however, the published SBCI studies do not usually discuss why a particular evaluation metric is chosen for evaluating the performance. This leads to the obvious conclusion that finding suitable evaluation metrics forms an important and a needed study for SBCI systems. This need has been emphasized in a recently published technical report on evaluating SBCI systems [4].

## **1.8 Thesis contributions**

As discussed above, this thesis addresses three issues of importance to the designs of SBCI systems:

- 1) Decreasing the false positive rates in SBCI systems.
- 2) Handling artifacts in SBCI systems, and
- 3) Evaluating the performance of SBCI systems.

To address the high FP rates and the presence of artifacts, in Chapters 4, 5, 7, and Chapter 9, we propose and evaluate a new 2-state SBCI system that can distinguish an IC command related to a specific movement pattern from the NC state in EEG signals. In the design of this system, the main focus is to improve the performance over those of previous EEG-based SBCIs. To achieve this goal, we propose and investigate the simultaneous detection of three neurological phenomena to recognize IC commands. These three neurological phenomena are movement-related potentials (MRPs), changes in the power of Mu rhythms (CPMR) and changes in the power of Beta rhythms (CPBR). These neurological phenomena are known to be time-locked to the onset of movement, so

we postulated that detecting all of them at the same time improves the system's performance. This is the first time in the BCI literature that the analysis of these neurological phenomena at the same time is proposed for detecting the IC commands. A systematic approach for feature extraction, selection and classification for each neurological phenomenon is presented. We also propose a 2 –stage multiple classifier system (MCS) to efficiently combine the information extracted from these neurological phenomena.

The performance of the proposed system is compared with those of other state-of-the-art EEG-based SBCI systems. It is shown that the proposed method results in a superior performance. A theoretical analysis of the performance of the proposed SBCI is presented and it is shown that under certain conditions the proposed methodology can theoretically approach perfect classification accuracy.

Since the proposed SBCI relies on detecting more than one neurological phenomenon at the same time, it is expected that its performance is robust in the presence of most artifacts. This is because artifacts are usually more prominent over a certain frequency band and do not affect other frequency bands as much. In Chapter 7, we show that the proposed SBCI has a good performance over periods contaminated with artifacts.

Finally, in Chapter 8 a framework for comparing and selecting evaluation metrics for SBCI systems is also proposed. It is shown that this framework can be successfully applied to select from a number of available metrics, the evaluation metric that is most suitable for evaluating SBCI systems. The findings of this chapter are applied in Chapter 9 to select the most suitable evaluation metric for evaluating the performance.

The thesis provides a detailed description of our methods and results. The main contributions of this thesis fall into three categories as follows:

### **1.8.1 Reducing high false positive rates**

- 1) Introducing the idea of using features from MRPs, CPMR and CPBR at the same time to detect the possible presence of IC commands.
- 2) Developing a new SBCI system that extracts and classifies features extracted from the above neurological phenomena efficiently. A new two-stage multiple classifier

system (MCS) is proposed. At the first stage, an MCS is separately designed for each neurological phenomenon. At the second stage, another MCS combines the outputs of MCSs in the first stage.

- 3) Studying the performance of the 2-stage MCS using the Linear Programming theory.
- 4) Investigating of the performance of the proposed SBCI system on two datasets: one dataset related to the right finger flexion and the other dataset related to the right hand extension. It will be shown that the proposed SBCI system achieves error rates that are significantly lower than those of other state-of-the-art EEG-based SBCI systems.
- 5) Comparing the use of monopolar and bipolar EEG electrodes for detecting right hand extension movements and demonstrating that bipolar electrodes provide superior results.
- 6) Studying the effect of automatic user-customization in the performance of a state-of-the-art self-paced BCI system previously developed in the brain interface lab of the Neil Squire society. It will be demonstrated that automatic user customization significantly improves the performance compared to manual customization by an expert.

### **1.8.2 Addressing artifacts in SBCI systems**

- 1) Presenting a detailed review of the methods that handle artifacts in BCI systems. Surprisingly, this review shows that most BCI systems do not address the presence of artifacts properly.
- 2) Investigating the performance of the proposed SBCI system over periods contaminated with eye-blink artifacts. It will be shown that the system has a reasonably good performance over periods contaminated with large eye movement artifacts.
- 3) Investigating the performance of the proposed SBCI system using the data from the session recorded few days after the data used for training the SBCI system. Again, it will be demonstrated that the proposed SBCI system achieves a good performance for three out of four participants whose data are studied.

- 4) Proposing an artifact monitoring system that detects large eye movement artifacts as well as EMG artifacts related to frontalis muscles at the same time.

### **1.8.3 Finding a suitable evaluation metric for SBCI systems**

- 1) Proposing a framework for comparing the evaluation metrics during the model selection process in SBCI systems.
- 2) Applying the proposed framework to a particular SBCI system and finding the most suitable evaluation metric.
- 3) Demonstrating that the Kappa coefficient is the most suitable evaluation metric for the proposed SBCI system.

## **1.9 Organization of the thesis**

The organization of this thesis is as follows:

We first study the Low Frequency - Asynchronous Switch Design (the LF-ASD) in Chapter 2. The LF-ASD is a state-of-the-art EEG-based SBCI system that is used as the basis for performance comparison in some of the following chapters. It is thus reasonable that at the first stage of the research, the structure of LF-ASD as well as its performance are examined in detail.

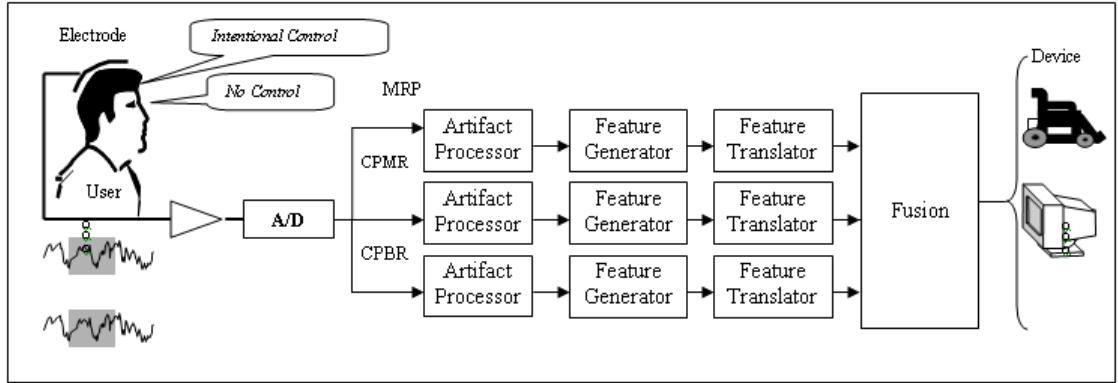
The parameter values of the feature generator of the LF-ASD have been usually determined by the designer based on trial and error. This process is suboptimal, subjective and time-consuming for the researchers. In Chapter 2, we propose the use of a genetic algorithm (GA) to automatically tune the parameter values of the feature generator of one of the designs of the LF-ASD. The purpose of this study is 2-fold: 1) to automate the tuning process of the feature generator and 2) to improve the performance of the LF-ASD. Specifically, we are interested in finding an upper limit for the performance of this design of the LF-ASD. This could be a starting point for the next stage of the research as it would decide whether the current feature generator should be kept or if it should be replaced by a more powerful one. In Chapter 2, we show that only moderate improvements in the performance of the LF-ASD occur, after automatically

tuning the parameters of the feature generator. For this reason, the subsequent chapters, the use of the wavelet transform for extracting the features is explored.

In Chapter 3, we study the use of the discrete wavelet transform (DWT) for extracting MRP features. The reason behind choosing DWT is 2-fold: 1) LF-ASD uses a transform for feature extraction that is a simplified version of the wavelet transform. We thus postulate that a more sophisticated version of this detector (i.e., the DWT) should be able to better extract features related to MRPs and 2) DWT provides both time as well as frequency information, so it can provide more information than the traditional frequency-based approaches and can improve the performance [11]. The evidence from the BCI literature also supports this hypothesis [140, 141]. We compare two different variations of this design. The first is based on MRP features extracted from monopolar EEG channels and the other is based on MRP features extracted from bipolar EEG channels. We argue that the system based on the bipolar MRP features yields a superior performance.

Parallel to the research carried out in Chapter 3, we carried out a second study which focused on a simple design of an SBCI system that is based on features extracted from three neurological phenomena: MRP, CPMR and CPBR. This study is carried out as a proof of concept to show that the combination of the three neurological phenomena discussed earlier in this chapter would improve the performance of the system. For this reason, we only applied simple feature extraction and classification methods (matched filtering and K-nearest neighbor classifier). A 2-stage multiple classifier system (MCS) is proposed to “fuse” the classification results attributed to each neurological phenomenon. Figure 1-9 shows the overall structure of the system studied in Chapter 4.





**Figure 1-9.** The overall schematic of the SBCI system developed and studied in Chapters 4, 5, 7, and 9.

In Chapter 5, we use the findings from Chapters 3 and 4 and propose an improved design. This design uses a new feature extraction method (a combination of stationary wavelet transform (SWT) followed by matched filtering) and proposes the use of a hybrid genetic algorithm (HGA) to simultaneously select the features, the parameter values of the classifiers and the combination method for the 2-stage MCS. We demonstrate that the new design achieves much lower FP rates than previous EEG-based SBCI systems, while it maintains a modest TP rate. These promising results facilitate the practical applications of the proposed SBCI system.

In Chapters 6 and 7 we focus on artifacts in BCI systems in general and in the proposed SBCI system in particular. Artifacts are unwanted potentials that can change the shape of the neurological phenomena and thus decrease the system's performance. As a result, handling artifacts is an important part in the design of BCI systems. In Chapter 6 artifacts in BCI literature are addressed. The results of this review study show that the BCI literature does not properly report artifacts handling. In other words, BCI researchers do not report whether or not they have considered the presence of artifacts. A large number of studies reject the artifact-contaminated periods either manually or automatically. We argue that a proper solution is to either efficiently remove artifacts or to design a BCI system whose performance is robust to the presence of artifacts.

In Chapter 7 we further analyze the performance of the proposed SBCI. We first consider the performance over periods contaminated with eye-blink artifacts. Next, we

test the performance of the system on data collected in a session recorded after the sessions used for training and testing the performance of the system. In both cases, we demonstrate that our proposed system maintains its good performance in the presence of artifacts. These results also demonstrate that during online testing, the system does not need to reject periods marked with artifacts. This, in turn, greatly increases the periods during which the system is available for control.

In Chapter 8, we address the critical issue of selecting a suitable evaluation metric in the design of SBCI systems. We revise and improve a framework that was proposed earlier for comparing the classification accuracy and the AUC metrics [128]. Our revised model can be used to compare various metrics as well as studying new metrics. It can also be used for selecting the metric(s) that is (are) most suitable for evaluating a certain classification system. We also analyze the application of the proposed framework to the field of SBCI systems. In particular, we consider four evaluation metrics: overall classification accuracy (OA),  $\frac{TPR}{FPR}$ , Kappa's coefficient, and HF-difference and compare their performances during the model selection procedure for a particular SBCI system. We demonstrate that some evaluation metrics such as Kappa and HF-difference are more suitable and some such as OA and  $\frac{TPR}{FPR}$  are less suitable evaluation metrics for SBCI systems.

In Chapters 2 to 8, the type of movement that was considered for the generation of IC commands was the index finger flexion. In order for the system to be generalized to more control options, its performance on new mental tasks (related to other types of movements) should also be investigated. It is also desired that the same system also performs well on other types of movements. In Chapter 9, we examine the performance of the SBCI system proposed in Chapter 5 on a new dataset. In this dataset, IC commands are generated by hand extension movements. NC data are also recorded in a more engaging environment than those used in previous studies for training the system. It is demonstrated that our proposed design maintains a superior performance compared to other EEG-based SBCI designs in the literature. Secondly, electromyography (EMG) signals from frontalis muscles are recorded to rule out the activation of such muscles

during the movement executions. Furthermore, we use the framework developed in Chapter 8 to select the most suitable evaluation metric for the system. We conclude that Cohen's Kappa coefficient is the most suitable evaluation metric for the model selection procedure of the proposed SBCI system. Finally, we compare the performance of monopolar and bipolar EEG electrode montages. This study shows that the bipolar montage generates more suitable features and thus a superior performance than the monopolar montage.

In Chapter 10, we summarize the contributions of this thesis to the field of SBCI systems. We also present some of the potential research subjects that can immediately follow this research in this chapter. Figure 1-10 shows a summary of the organization of this thesis.

In Appendix. A, we provide a copy of the approval from the Behavioral Research Ethics Board (BREB) of the University of British Columbia to conduct this study. In Appendix B, we use the linear programming theory to show how the proposed multiple classifier system could achieve perfect classification accuracy under certain conditions.

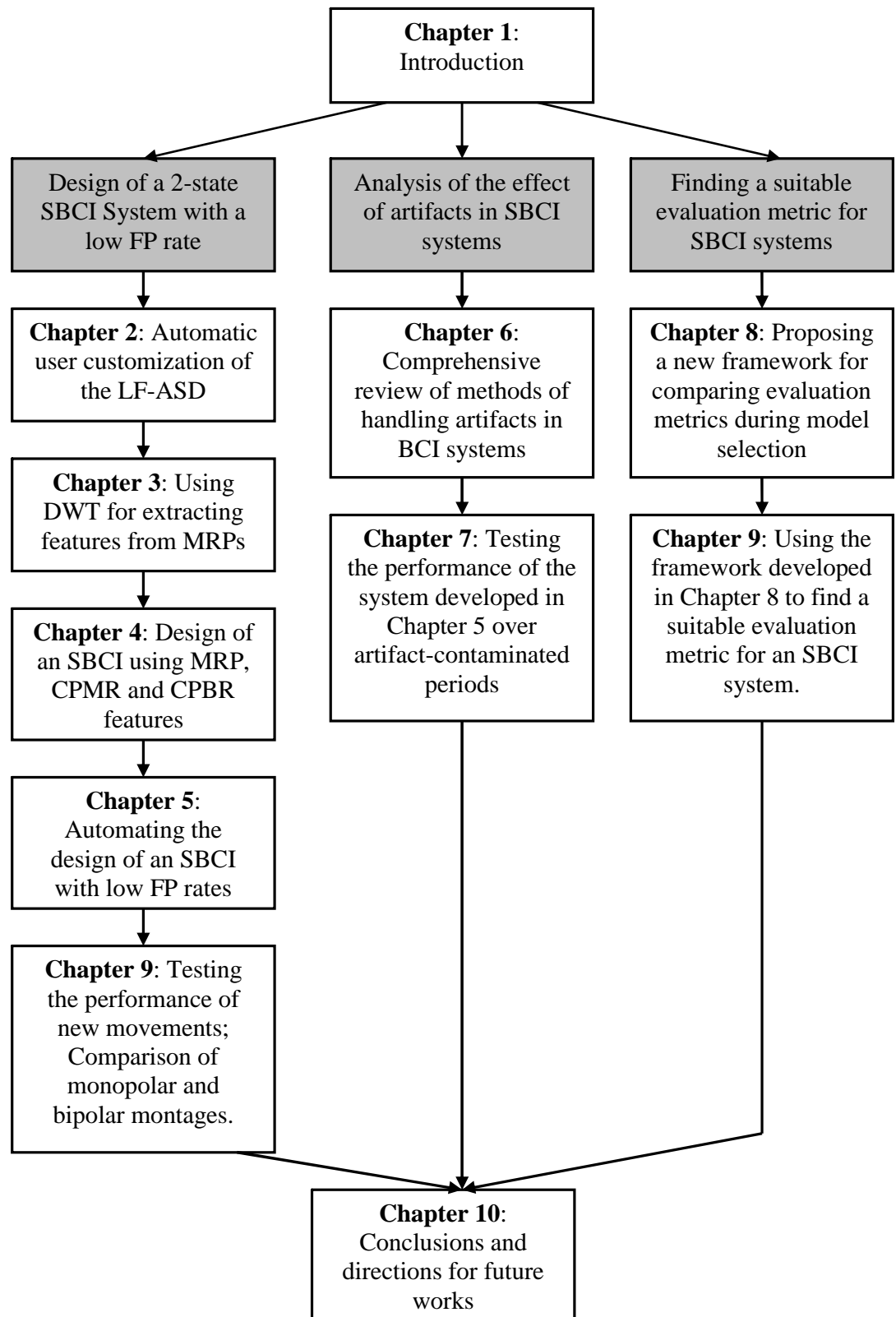


Figure 1-10. Outline of the thesis.

## 1.10 References

- [1] T. Vaughan, W. J. Heetderks, L. J. Trejo, W. Z. Rymer, M. Wienrich, M. M. Moore, A. Kubler, B. H. Dobkin, N. Birbaumer, E. Donchin, E. W. Wolpaw and J. W. R. "Brain-computer interface technology: a review of the second international meeting", *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 11, no.2, pp. 94-109, 2003.
- [2] 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, no.6, pp. 767-791, Jun.2002.
- [3] S. G. Mason, A. Bashashati, M. Fatourehchi, K. F. Navarro and G. E. Birch, "A comprehensive survey of brain interface technology designs", *Ann. Biomed. Eng.*, vol. 35, no.2, pp. 137-169, Feb.2007.
- [4] S. G. Mason, J. Kronegg, J. Huggins, M. Fatourehchi and A. and Schloegl, "Evaluating the performance of self-paced BCI technology", Technical Report, available online: [http://www.bci-info.tugraz.at/Research\\_Info/documents/articles/self\\_paced\\_tech\\_report-2006-05-19.pdf](http://www.bci-info.tugraz.at/Research_Info/documents/articles/self_paced_tech_report-2006-05-19.pdf), 2006.
- [5] J. F. Borisoff, S. G. Mason, A. Bashashati and G. E. Birch, "Brain-computer interface design for asynchronous control applications: improvements to the LF-ASD asynchronous brain switch", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 985-992, Jun.2004.
- [6] E. Yom-Tov and G. F. Inbar, "Detection of Movement-Related Potentials from the Electro-Encephalogram for possible use in a Brain-Computer Interface", *Medical and Biological Engineering and Computing*, vol. 41, no.1, pp. 85-93, Jan.2003.
- [7] G. E. Birch, Z. Bozorgzadeh and S. G. Mason, "Initial on-line evaluations of the LF-ASD brain-computer interface with able-bodied and spinal-cord subjects using imagined voluntary motor potentials", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 10, no.4, pp. 219-224, Dec.2002.
- [8] B. Graimann, J. E. Huggins, S. P. Levine and G. Pfurtscheller, "Toward a direct brain interface based on human subdural recordings and wavelet-packet analysis", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 954-962, Jun.2004.
- [9] J. Kronegg, s. Voloshynovskiy and P. Pun, "Analysis of bit rate definitions for brain-computer interfaces," in the Proc. *Int. Conf. on Human-Computer Interaction (HCI'05)*, Las Vegas, Nevada, 2005.
- [10] A. Schlögl, J. Kronegg, J. Huggins and S. G. Mason, "Evaluation criteria in BCI research," in *Towards Brain-Computer Interfacing* (G. Dornhege, J. R. Millan, T. Hinterberger, D. McFarland and K. R. Muller, Eds.), MIT Press, 2007.
- [11] A. Bashashati, M. Fatourehchi, R. K. Ward and G. E. Birch, "A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals", *J. Neural Eng.*, vol. 4, no.2, pp. R32-57, Jun.2007.
- [12] J. J. Vidal, "Toward direct brain-computer communication", *Annu. Rev. Biophys. Bioeng.*, vol. 2, pp. 157-180, 1973.
- [13] G. Pfurtscheller and A. Aranibar, "Event-related cortical desynchronization detected by power measurements of scalp EEG", *Electroencephalogr. Clin. Neurophysiol.*, vol. 42, no.6, pp. 817-826, Jun.1977.

- [14] L. Leocani, C. Toro, P. Manganotti, P. Zhuang and M. Hallett, "Event-related coherence and event-related desynchronization/synchronization in the 10 Hz and 20 Hz EEG during self-paced movements", *Electroencephalogr. Clin. Neurophysiol.*, vol. 104, no.3, pp. 199-206, May.1997.
- [15] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles", *Clin. Neurophysiol.*, vol. 110, no.11, pp. 1842-1857, Nov.1999.
- [16] G. Pfurtscheller, K. Pichler-Zalaudek, B. Ortmayr, J. Diez and F. Reisecker, "Postmovement beta synchronization in patients with Parkinson's disease", *J. Clin. Neurophysiol.*, vol. 15, no.3, pp. 243-250, May.1998.
- [17] D. J. McFarland, W. A. Sarnacki, T. M. Vaughan and J. R. Wolpaw, "Brain-computer interface (BCI) operation: signal and noise during early training sessions", *Clin. Neurophysiol.*, vol. 116, no.1, pp. 56-62, Jan.2005.
- [18] J. R. Wolpaw and D. J. McFarland, "Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans", in *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no.51, pp. 17849-17854, Dec 21.2004.
- [19] J. R. Wolpaw and D. J. McFarland, "Multichannel EEG-based brain-computer communication", *Electroencephalogr. Clin. Neurophysiol.*, vol. 90, no.6, pp. 444-449, Jun.1994.
- [20] D. J. McFarland, L. M. McCane and J. R. Wolpaw, "EEG-based communication and control: short-term role of feedback", *IEEE Trans. Rehabil. Eng.*, vol. 6, no.1, pp. 7-11, Mar.1998.
- [21] L. A. Miner, D. McFarland and J. R. Wolpaw, "Answering Questions with an Electroencephalogram-Based Brain Computer Interface", *Arch. Phys. Med. Rehabil.*, vol. 79, pp. 1029-1033, 1998.
- [22] G. R. Müller-Putz, R. Scherer, G. Pfurtscheller and R. Rupp, "EEG-based neuroprosthesis control: a step towards clinical practice", *Neurosci. Lett.*, vol. 382, no.1-2, pp. 169-174, Jul 2005.
- [23] G. Pfurtscheller, G. R. Müller-Putz, J. Pfurtscheller and R. Rupp, "EEG-Based Asynchronous BCI Controls Functional Electrical Stimulation in a Tetraplegic Patient", *EURASIP Journal on Applied Signal Processing*, vol. 2005, no.19, pp. 3152-3155, 2005.
- [24] G. Pfurtscheller, B. Graimann, J. E. Huggins and S. P. Levine, "Brain-computer communication based on the dynamics of brain oscillations", *Suppl. Clin. Neurophysiol.*, vol. 57, pp. 583-591, 2004.
- [25] G. Pfurtscheller, C. Neuper, C. Guger, W. Harkam, H. Ramoser, A. Schlogl, B. Obermaier and M. Pregenzer, "Current trends in Graz Brain-Computer Interface (BCI) Research", *IEEE Trans. Rehab. Eng.*, vol. 8, no.2, pp. 216-219, Jun. 2000.
- [26] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication", *Proc. IEEE*, vol. 89, no.7, pp. 1123-1134, 2001.
- [27] L. Deecke, B. Grozinger and H. H. Kornhuber, "Voluntary finger movement in man: cerebral potentials and theory", *Biol. Cybern.*, vol. 23, no.2, pp. 99-119, Jul 14.1976.
- [28] H. Shibasaki, G. Barrett, E. Halliday and A. M. Halliday, "Components of the movement-related cortical potential and their scalp topography", *Electroencephalogr. Clin. Neurophysiol.*, vol. 49, no.3-4, pp. 213-226, Aug.1980.

- [29] I. M. Tarkka and M. Hallett, "Cortical topography of premotor and motor potentials preceding self-paced, voluntary movement of dominant and non-dominant hands", *Electroencephalogr. Clin. Neurophysiol.*, vol. 75, no.2, pp. 36-43, Feb.1990.
- [30] M. Hallett, "Movement-related cortical potentials", *Electromyogr. Clin. Neurophysiol.*, vol. 34, no.1, pp. 5-13, Jan-Feb.1994.
- [31] C. Babiloni, F. Carducci, F. Cincotti, P. M. Rossini, C. Neuper, G. Pfurtscheller and F. Babiloni, "Human movement-related potentials vs desynchronization of EEG alpha rhythm: a high-resolution EEG study", *Neuroimage*, vol. 10, no.6, pp. 658-665, Dec.1999.
- [32] A. Urbano, C. Babiloni, P. Onorati and F. Babiloni, "Human cortical activity related to unilateral movements. A high resolution EEG study", *Neuroreport*, vol. 8, no.1, pp. 203-206, Dec 20.1996.
- [33] A. Urbano, C. Babiloni, P. Onorati, F. Carducci, A. Ambrosini, L. Fattorini and F. Babiloni, "Responses of human primary sensorimotor and supplementary motor areas to internally triggered unilateral and simultaneous bilateral one-digit movements. A high-resolution EEG study", *Eur. J. Neurosci.*, vol. 10, no.2, pp. 765-770, Feb.1998.
- [34] Z. Bozorgzadeh, G. E. Birch and S. G. Mason, "The LF-ASD brain computer interface: On-line identification of imagined finger flexions in the spontaneous EEG of able-bodied subjects," in *Proc. IEEE ICASSP'00*, vol.6, pp. 2385-2388, 2000.,
- [35] S. G. Mason and G. E. Birch, "A brain-controlled switch for asynchronous control applications", *IEEE Trans. Biomed. Eng.*, vol. 47, no.10, pp. 1297-1307, Oct.2000.
- [36] G. E. Birch, P. D. Lawrence and R. D. Hare, "Single Trial Processing of Event Related Potentials Using Outlier Information", *IEEE Trans. Biomed. Eng.*, vol. 40, no.1, pp. 59-73, 1993.
- [37] B. Blankertz, G. Dornhege, C. SchÅœfer, R. Krepki, J. Kolmorgen, K. R. MÃ¼ller, 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," in *IEEE Trans. Neural Sys. Rehab. Eng.*, vol.11, no.2, 2003,
- [38] G. Dornhege, B. Blankertz and G. Curio, "Speeding up classification of multi-channel brain-computer interfaces: Common spatial patterns for slow cortical potentials," in *Proc. 1<sup>st</sup> IEEE EMBS Int. Conf. on Neural Engineering*, pp. 595-598. 2003,
- [39] E. Yom-Tov and G. F. Inbar, "Feature Selection for the Classification of Movements From Single Movement-Related Potentials", *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 10, no.3, pp. 170-177, Sep.2002.
- [40] E. Yom-Tov and G. F. Inbar, "Selection of relevant features for classification of movements from single movement-related potentials using a genetic algorithm," in *the Proc. 23<sup>rd</sup> IEEE/EMBS Int. Conf.*, vol.2, pp. 1364-1366, 2001.
- [41] S. P. Levine, J. E. Huggins, S. L. Bement, R. K. Kushwaha, L. A. Schuh, E. A. Passaro, M. M. Rohde and D. A. Ross, "Identification of Electroencephalogram Patterns as the Basis for a Direct Brain Interface", *J Clinical Neurophysiol*, vol. 16, no.5, pp. 439-447, Sep.1999.
- [42] J. E. Huggins, S. P. Levine, R. Kushwaha, S. L. Bement, L. A. Schuh and D. A. Ross, "Identification of cortical signal patterns related to human tongue protrusion," in pp. 670-672. 1995.

- [43] N. Neumann, A. Kubler, J. Kaiser, T. Hinterberger and N. Birbaumer, "Conscious perception of brain states: mental strategies for brain-computer communication", *Neuropsychologia*, vol. 41, no.8, pp. 1028-1036, 2003.
- [44] T. Hinterberger, B. Wilhelm, J. Mellinger, B. Kotchoubey and N. Birbaumer, "A device for the detection of cognitive brain functions in completely paralyzed or unresponsive patients", *IEEE Trans. Biomed. Eng.*, vol. 52, no.2, pp. 211-220, Feb.2005.
- [45] N. Birbaumer, "The thought-translation-device (TTD): Taming cognition for action", *Brain Cogn.*, vol. 54, no.2, pp. 130-130, Mar.2004.
- [46] T. Hinterberger, S. Schmidt, N. Neumann, J. Mellinger, B. Blankertz, G. Curio and N. Birbaumer, "Brain-computer communication and slow cortical potentials", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 1011-1018, Jun.2004.
- [47] N. Birbaumer, A. Kubler, N. Ghanayim, T. Hinterberger, J. Perelmouter, J. Kaiser, I. Iversen, B. Kotchoubey, N. Neumann and H. Flor, "The thought translation device (TTD) for completely paralyzed patients", *IEEE Trans. Rehabil. Eng.*, vol. 8, no.2, pp. 190-193, Jun.2000.
- [48] A. Kubler, B. Kotchoubey, J. Kaiser, J. R. Wolpaw and N. Birbaumer, "Brain-Computer Communication: Unlocking the Locked In", *Psych Bulletin*, vol. 127, no.3, pp. 358-375, May.2001.
- [49] J. d. R. Millan, J. Mourino, M. G. Marciani, F. Babiloni, F. Topani, I. Canale, J. Heikkinen and K. Kaski, "Adaptive brain interfaces for physically-disabled people," in Proc. IEEE EMBS Conf, vol.4., pp. 2008-2011, 1998.
- [50] C. W. Anderson, S. V. Devulapalli and E. A. Stolz, "Signal Classification with Different Signal Representations", *Neural Networks for Signal Processing*, pp. 475-483, 1995.
- [51] 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. Rehab. Eng.*, vol. 11, no.2, pp. 141-144, Jun. 2003.
- [52] C. W. Anderson, E. A. Stolz and S. Shamsunder, "Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks", *IEEE Trans. Biomed. Eng.*, vol. 45, no.3, pp. 277-286, Mar.1998.
- [53] B. Z. Allison and J. A. Pineda, "ERPs evoked by different matrix sizes: Implications for a brain computer interface (BCI) system", *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 11, no.2, pp. 110-113, Jun.2003.
- [54] E. Donchin, K. M. Spencer and R. Wijesinghe, "The mental prosthesis: assessing the speed of a P300-based brain-computer interface", *IEEE Trans. Rehabil. Eng.*, vol. 8, no.2, pp. 174-179, Jun.2000.
- [55] 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, no.6, pp. 510-523, Dec.1988.
- [56] E. W. Sellers, A. Kubler and E. Donchin, "Brain-computer interface research at the University of South Florida Cognitive Psychophysiology Laboratory: the P300 Speller", *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 14, no.2, pp. 221-224, Jun.2006.
- [57] F. Piccione, F. Giorgi, P. Tonin, K. Priftis, S. Giove, S. Silvoni, G. Palmas and F. Beverina, "P300-based brain computer interface: reliability and performance in healthy and paralysed participants", *Clin. Neurophysiol.*, vol. 117, no.3, pp. 531-537, Mar.2006.



- [58] A. A. Glover, M. C. Onofrj, M. F. Ghilardi and I. Bodis-Wollner, "P300-like potentials in the normal monkey using classical conditioning and an auditory 'oddball' paradigm", *Electroencephalogr. Clin. Neurophysiol.*, vol. 65, no.3, pp. 231-235, May.1986.
- [59] B. Roder, F. Rosler, E. Hennighausen and F. Nacker, "Event-related potentials during auditory and somatosensory discrimination in sighted and blind human subjects", *Brain Res. Cogn. Brain Res.*, vol. 4, no.2, pp. 77-93, Sep.1996.
- [60] J. J. Vidal, "Real-Time Detection of Brain Events in EEG," *Proc IEEE*, vol. 65, pp. 633-641, 1977.
- [61] E. E. Sutter, "The brain response interface: communication through visually-induced electrical brain responses", *J Micro Comp App*, vol. 15, pp. 31-45, 1992.
- [62] M. Middendorf, G. McMillan, G. Calhoun and K. S. Jones, "Brain-Computer Interfaces Based on the Steady-State Visual-Evoked Response", *IEEE Trans. Rehab. Eng.*, vol. 8, no.2, pp. 211-214, Jun. 2000.
- [63] X. Gao, D. Xu, M. Cheng and S. Gao, "A BCI-based environmental controller for the motion-disabled", *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 11, no.2, pp. 137-140, Jun. 2003.
- [64] Y. Wang, R. Wang, X. Gao and S. Gao, "Brain-computer interface based on the high-frequency steady-state visual evoked potential," in the *Proc. 1st Int. Conf. in Neural Interface and Control*, pp. 37-39, 2005.
- [65] Cheng Ming, Gao Xiaorong, Gao Shangkai and Wang Boliang, "Stimulation frequency extraction in SSVEP-based brain-computer interface," in in the *Proc. 1st Int. Conf. Neural Interface and Control*, ,pp. 64-67. 2005,
- [66] Yijun Wang, Zhiguang Zhang, Xiaorong Gao and Shangkai Gao, "Lead selection for SSVEP-based brain-computer interface," in the *Proc. 26<sup>th</sup> IEEE/EMBS Int. Conf.*,vol.2,pp. 4507-4510 , 2004.
- [67] J. K. Chapin, K. A. Moxon, R. S. Markowitz and M. A. Nicolelis, "Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex", *Nat. Neurosci.*, vol. 2, no.7, pp. 664-670, Jul.1999.
- [68] M. A. L. Nicolelis and J. K. Chapin, "Controlling robots with the mind", *Sci. Am.*, vol. 287, no.4, pp. 46-53, Oct.2002.
- [69] J. T. Francis and J. K. Chapin, "Force field apparatus for investigating movement control in small animals", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 963-965, Jun.2004.
- [70] S. Darmanjian, Sung Phil Kim, M. C. Nechyba, S. Morrison, J. Principe, J. Wessberg and M. A. L. Nicolelis, "Bimodal brain-machine interface for motor control of robotic prosthetic," in the *Proc. IEEE Int. Conf. Intelligent Robots and Systems*, vol.4,pp. 3612-3617 , 2003.
- [71] J. P. Donoghue, A. Nurmikko, G. Friehs and M. Black, "Development of neuromotor prostheses for humans", *Suppl. Clin. Neurophysiol.*, vol. 57, pp. 592-606, 2004.
- [72] F. Wood, M. J. Black, C. Vargas-Irwin, M. Fellows and J. P. Donoghue, "On the variability of manual spike sorting", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 912-918, Jun.2004.
- [73] L. R. Hochberg, M. D. Serruya, G. M. Friehs, J. A. Mukand, M. Saleh, A. H. Caplan, A. Branner, D. Chen, R. D. Penn and J. P. Donoghue, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia", *Nature*, vol. 442, no.7099, pp. 164-171, Jul 13.2006.

- [74] M. D. Serruya, N. G. Hatsopoulos, L. Paninski, M. R. Fellows and J. P. Donoghue, "Instant neural control of a movement signal", *Nature*, vol. 416, no.6877, pp. 141-142, Mar.2002.
- [75] T. M. Vaughan, J. R. Wolpaw and E. Donchin, "EEG-Based Communication: Prospects and Problems", *IEEE Trans. Rehab. Eng.*, vol. 4, no.4, pp. 425-430, 1996.
- [76] G. Dornhege, B. Blankertz, G. Curio and K. R. Muller, "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 993-1002, Jun.2004.
- [77] B. Graimann, J. E. Huggins, A. Schlogl, S. P. Levine and G. Pfurtscheller, "Detection of movement-related desynchronization patterns in ongoing single-channel electrocorticogram", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no.3, pp. 276-281, Sep.2003.
- [78] R. Scherer, G. R. Muller, C. Neuper, B. Graimann and G. Pfurtscheller, "An asynchronously controlled EEG-based virtual keyboard: improvement of the spelling rate", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 979-984, Jun.2004.
- [79] Z. Yu, S. G. Mason and G. E. Birch, "Enhancing the performance of the LF-ASD brain-computer interface," in *Proc. of the 2<sup>nd</sup> Joint EMBS/BMES Conference*, vol.3, pp. 2443-2444, 2002.
- [80] A. Bashashati, M. Fatourehchi, R. K. Ward and G. E. Birch, "User customization of the feature generator of an asynchronous brain interface", *Ann. Biomed. Eng.*, vol. 34, no.6, pp. 1051-1060, Jun.2006.
- [81] A. Bashashati, R. K. Ward and G. E. Birch, "A new design of the asynchronous brain computer interface using the knowledge of the path of features," in *Proc. 2<sup>nd</sup> IEEE EMBS Int. Conf. on Neural Engineering*, pp. 101-104. 2005.
- [82] A. Bashashati, R. K. Ward and G. E. Birch, "Towards development of a 3-state self-paced brain computer interface", *Computational Intelligence and Neuroscience*, Vol.2007, pp.1-8, Oct. 2007.
- [83] J. E. Huggins, S. P. Levine, J. A. Fessler, W. M. Sowers, G. Pfurtscheller, B. Graimann, A. Schloegl, D. N. Minecan, R. K. Kushwaha, S. L. BeMent, O. Sagher and L. A. Schuh, "Electrocorticogram as the basis for a direct brain interface: Opportunities for improved detection accuracy," in *Proc. 1<sup>st</sup> IEEE EMBS Int. Conf. on Neural Engineering*, pp. 587-590. 2003,
- [84] 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, no.2, pp. 180-185, Jun.2000.
- [85] J. E. Huggins, S. P. Levine, S. L. Bement, R. K. Kushwaha, L. A. Schuh, E. A. Passaro, M. M. Rohde, D. A. Ross, K. V. Elisevich and B. J. Smith, "Detection of Event-Related Potentials for Development of a Direct Brain Interface", *J Clinical Neurophysiol*, vol. 16, no.5, pp. 448-455, Sep.1999.
- [86] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing", *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 57, no.1, pp. 289-300, 1995.
- [87] G. Townsend, B. Graimann and G. Pfurtscheller, "Continuous EEG classification during motor imagery--simulation of an asynchronous BCI", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 12, no.2, pp. 258-265, Jun.2004.

- [88] C. Toro, G. Deuschl, R. Thatcher, S. Sato, C. Kufta and M. Hallett, "Event-related desynchronization and movement-related cortical potentials on the ECoG and EEG", *Electroencephalogr. Clin. Neurophysiol.*, vol. 93, no.5, pp. 380-389, Oct.1994.
- [89] S. Arroyo, R. P. Lesser, B. Gordon, S. Uematsu, D. Jackson and R. Webber, "Functional significance of the mu rhythm of human cortex: an electrophysiologic study with subdural electrodes", *Electroencephalogr. Clin. Neurophysiol.*, vol. 87, no.3, pp. 76-87, Sep.1993.
- [90] G. Pfurtscheller and A. Aranibar, "Evaluation of event-related desynchronization (ERD) preceding and following voluntary self-paced movement", *Electroencephalogr. Clin. Neurophysiol.*, vol. 46, no.2, pp. 138-146, Feb.1979.
- [91] L. Defebvre, J. L. Bourriez, K. Dujardin, P. Derambure, A. Destee and J. D. Guieu, "Spatiotemporal study of Bereitschaftspotential and event-related desynchronization during voluntary movement in Parkinson's disease", *Brain Topogr.*, vol. 6, no.3, pp. 237-244, Spring.1994.
- [92] K. R. Muller, G. Curio, B. Blankertz and G. Dornhege, "Combining features for BCI," in the *Proc. Advances in Neural Inf. Proc. Systems (NIPS 02)*, vol.15,2003.
- [93] L. Narici, V. Pizzella, G. L. Romani, G. Torrioli, R. Traversa and P. M. Rossini, "Evoked alpha- and mu-rhythm in humans: a neuromagnetic study", *Brain Res.*, vol. 520, no.1-2, pp. 222-231, Jun 18.1990.
- [94] B. Feige, R. Kristeva-Feige, S. Rossi, V. Pizzella and P. M. Rossini, "Neuromagnetic study of movement-related changes in rhythmic brain activity", *Brain Res.*, vol. 734, no.1-2, pp. 252-260, Sep 23.1996.
- [95] G. Pfurtscheller, "Central beta rhythm during sensorimotor activities in man", *Electroencephalogr. Clin. Neurophysiol.*, vol. 51, no.3, pp. 253-264, Mar.1981.
- [96] W. Szurhaj, P. Derambure, E. Labyt, F. Cassim, J. L. Bourriez, J. Isnard, J. D. Guieu and F. Mauguire, "Basic mechanisms of central rhythms reactivity to preparation and execution of a voluntary movement: a stereoelectroencephalographic study", *Clin. Neurophysiol.*, vol. 114, no.1, pp. 107-119, Jan.2003.
- [97] H. S. Liu, X. Gao, F. Yang and S. Gao, "Imagined hand movement identification based on spatio-temporal pattern recognition of EEG," in *Proc. of the 1<sup>st</sup> Joint EMBS/BMES Conference*, pp. 599-602. 2003,
- [98] B. D. Mensh, J. Werfel and H. S. Seung, "BCI Competition 2003--Data set Ia: combining gamma-band power with slow cortical potentials to improve single-trial classification of electroencephalographic signals", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 1052-1056, Jun. 2004.
- [99] T. Hinterberger and G. Baier, "Parametric orchestral sonification of EEG in real time", *Multimedia, IEEE*, vol. 12, no.2, pp. 70-79, 2005.
- [100] Y. Wang, Z. Zhang, Y. Li, X. Gao, S. Gao and F. Yang, "BCI Competition 2003--Data set IV: an algorithm based on CSSD and FDA for classifying single-trial EEG", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 1081-1086, Jun. 2004.
- [101] M. Krauledat, G. Dornhege, B. Blankertz, F. Losch, G. Curio and K. -. Muller, "Improving speed and accuracy of brain-computer interfaces using readiness potential features," in the *Proc. 26<sup>th</sup> IEEE/EMBS Int. Conf.*, vol.2, pp. 4511-4515 , 2004.
- [102] T. M. Vaughan, W. J. Heetderks, L. J. Trejo, W. Z. Rymer, M. Weinrich, M. M. Moore, A. Kubler, B. H. Dobkin, N. Birbaumer, E. Donchin, E. W. Wolpaw and J. R. Wolpaw, "Brain-

- computer interface technology: a review of the Second International Meeting", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no.2, pp. 94-109, Jun.2003.
- [103] J. S. Barlow, "Artifact processing (rejection and minimization) in EEG data processing", *Handbook of Electroencephalography and Clinical Neurophysiology (Revised Series Ed.)*, Amsterdam: Elsevier, vol.2., pp.15-62, 1986.
- [104] P. Anderer, S. Roberts, A. Schlogl, G. Gruber, G. Klosch, W. Herrmann, P. Rappelsberger, O. Filz, M. J. Barbanoj, G. Dorffner and B. Saletu, "Artifact processing in computerized analysis of sleep EEG - a review", *Neuropsychobiology*, vol. 40, no.3, pp. 150-157, Sep.1999.
- [105] 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, no.3, pp. 386-394, Sep.1997.
- [106] W. Waterink and A. van Boxtel, "Facial and jaw-elevator EMG activity in relation to changes in performance level during a sustained information processing task", *Biol. Psychol.*, vol. 37, no.3, pp. 183-198, Jul.1994.
- [107] B. H. Cohen, R. J. Davidson, J. A. Senulis, C. D. Saron and D. R. Weisman, "Muscle tension patterns during auditory attention", *Biol. Psychol.*, vol. 33, no.2-3, pp. 133-156, Jul.1992.
- [108] 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, no.9, pp. 1580-1593, Sep.2003.
- [109] D. J. McFarland, W. A. Sarnacki, T. M. Vaughan and J. R. Wolpaw, "Brain-computer interface (BCI) operation: signal and noise during early training sessions", *Clin. Neurophysiol.*, vol. 116, no.1, pp. 56-62, Jan.2005.
- [110] R. N. Vigario, "Extraction of ocular artefacts from EEG using independent component analysis", *Electroencephalography and Clinical Neurophysiology*, vol. 103, no.3, pp. 395-404, Sep. 1997.
- [111] R. Verleger, "The instruction to refrain from blinking affects auditory P3 and N1 amplitudes", *Electroencephalogr. Clin. Neurophysiol.*, vol. 78, no.3, pp. 240-251, Mar.1991.
- [112] C. J. Ochoa and J. Polich, "P300 and blink instructions", *Clin. Neurophysiol.*, vol. 111, no.1, pp. 93-98, Jan.2000.
- [113] G. Gratton, "Dealing with artifacts: The EOG contamination of the event-related brain potential", *Behavior Research Methods, Instruments, & Computers*, vol. 30, no.1, pp. 44-53, 1998.
- [114] H. Ramoser, J. Muller-Gerking and G. Pfurtscheller, "Optimal Spatial Filtering of Single Trial EEG During Imagined Hand Movement", *IEEE Trans. Rehab. Eng.*, vol. 8, no.4, pp. 441-446, Dec.2000.
- [115] J. Millan, M. Franze, J. Mourino, F. Cincotti and F. Babiloni, "Relevant EEG features for the classification of spontaneous motor-related tasks", *Biol. Cybern.*, vol. 86, no.2, pp. 89-95, Feb.2002.
- [116] R. J. Croft and R. J. Barry, "Removal of ocular artifact from the EEG: a review", *Neurophysiol. Clin.*, vol. 30, no.1, pp. 5-19, Feb.2000.
- [117] V. Rowland, "Cortical steady potential (direct current potential) in reinforcement and learning", *Progress in Physiological Psychology*, vol. 2, pp. 1-77, 1968.

- [118] J. S. Barlow, "EMG artifact minimization during clinical EEG recordings by special analog filtering", *Electroencephalogr. Clin. Neurophysiol.*, vol. 58, no.2, pp. 161-174, Aug.1984.
- [119] J. R. Ives and D. L. Schomer, "A 6-pole filter for improving the readability of muscle contaminated EEGs", *Electroencephalogr. Clin. Neurophysiol.*, vol. 69, no.5, pp. 486-490, May.1988.
- [120] S. Choi, A. Cichocki, H. M. Park and S. Y. Lee, "Blind Source Separation and Independent Component Analysis: A Review", *Neural Information Processing-Letters and Review*, vol. 6, no.1, pp. 1-57, 2005.
- [121] T. D. Lagerlund, F. W. Sharbrough and N. E. Busacker, "Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition", *J. Clin. Neurophysiol.*, vol. 14, no.1, pp. 73-82, Jan.1997.
- [122] M. Browne and T. R. Cutmore, "Low-probability event-detection and separation via statistical wavelet thresholding: an application to psychophysiological denoising", *Clin. Neurophysiol.*, vol. 113, no.9, pp. 1403-1411, Sep.2002.
- [123] P. He, G. Wilson and C. Russell, "Removal of ocular artifacts from electro-encephalogram by adaptive filtering", *Med. Biol. Eng. Comput.*, vol. 42, no.3, pp. 407-412, May.2004.
- [124] P. Berg and M. Scherg, "A multiple source approach to the correction of eye artifacts", *Electroencephalogr. Clin. Neurophysiol.*, vol. 90, no.3, pp. 229-241, Mar.1994.
- [125] D. Burke, S. Kelly, P. de Chazal and R. Reilly, "A simultaneous filtering and feature extraction strategy for direct brain interfacing," in *Proc. of the 2<sup>nd</sup> Joint EMBS/BMES Conference*, vol.1, pp. 279-280, 2002.
- [126] A. Kuebler, B. Kotchoubey, H. P. Salzmann, N. Ghanayim, J. Perelmouter, V. Homberg and N. Birbaumer, "Self-regulation of slow cortical potentials in completely paralyzed human patients", *Neurosci. Lett.*, vol. 252, no.3, pp. 171-174, Aug.1998.
- [127] F. Provost and T. Fawcett, "Robust Classification for Imprecise Environments", *Mach. Learning*, vol. 42, no.3, pp. 203-231, 2001.
- [128] J. Huang and C. X. Ling, "Using AUC and accuracy in evaluating learning algorithms", *IEEE Trans. Knowled. Data Eng.*, vol. 17, no.3, pp. 299-310, 2005.
- [129] J. Zhu and T. Yao, "An evaluation of statistical spam filtering techniques", *ACM Transactions on Asian Language Information Processing (TALIP)*, vol. 3, no.4, pp. 243-269, 2004.
- [130] A. P. Bradley, "Use of the area under the ROC curve in the evaluation of machine learning algorithms", *Pattern Recognit*, vol. 30, no.7, pp. 1145-1159, 1997.
- [131] N. T. Choplin and D. C. Lundy, "The sensitivity and specificity of scanning laser polarimetry in the detection of glaucoma in a clinical setting", *Ophthalmology*, vol. 108, no.5, pp. 899-904, May.2001.
- [132] M. Fatourechi, G. E. Birch and R. K. Ward, "A self-paced brain interface system that uses movement related potentials and changes in the power of brain rhythms", *J. Comput. Neurosci.*, vol.23, no.1, pp.21-37, Aug. 2007.
- [133] N. Yamawaki, C. Wilke, Z. Liu and B. He, "An enhanced time-frequency-spatial approach for motor imagery classification", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no.2, pp. 250-254, Jun.2006.

- [134] A. Buttfeld, P. W. Ferrez and R. Millan Jdel, "Towards a robust BCI: error potentials and online learning", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no.2, pp. 164-168, Jun.2006.
- [135] G. R. Muller-Putz, R. Scherer, C. Neuper and G. Pfurtscheller, "Steady-state somatosensory evoked potentials: suitable brain signals for brain-computer interfaces?", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no.1, pp. 30-37, Mar.2006.
- [136] J. R. Wolpaw, D. McFarland and G. Pfurtscheller, "EEG-based Communication: Improved Accuracy by Reponse Verification", *IEEE Trans. Rehab. Eng.*, vol. 6, no.3, pp. 326-333, 1998.
- [137] M. Fatourehchi, S. G. Mason, G. E. Birch and R. K. Ward, "Is information transfer rate a suitable performance measure for self-paced brain interface systems?" in Proc. IEEE Int. Symp. Signal Processing and Information Technology, pp. 212-216, 2006.
- [138] J. Cohen, "A coefficient of agreement for nominal scales", *Educational and Psychological Measurement*, vol. 20, no.1, pp. 37-46, 1960.
- [139] M. Fatourehchi, G. E. Birch and R. K. Ward, "Applying a hybrid genetic algorithm in the design of a self-paced brain interface with a low false positive rate," in Proc. IEEE ICASSP'07, vol.4, pp. IV-1157; IV-1160, Apr. 2007.
- [140] V. Bostanov, "BCI Competition 2003--Data sets Ib and I Ib: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 1057-1061, Jun.2004.
- [141] L. Qin and B. He, "A wavelet-based time-frequency analysis approach for classification of motor imagery for brain-computer interface applications", *J. Neural Eng.*, vol. 2, no.4, pp. 65-72, Dec.2005.

## **CHAPTER 2    AUTOMATIC USER CUSTOMIZATION FOR IMPROVING THE PERFORMANCE OF A SELF- PACED BRAIN COMPUTER INTERFACE SYSTEM<sup>1</sup>**

### **2.1 Introduction**

A self-paced brain computer interface (BCI) system allows individuals with severe motor disabilities to control objects in their environment using their brain signals only and at any time, i.e., at their own pace [1-11]. The output of a self-paced BCI system should only be activated when the user intends to control, and should remain inactive at all other times. Implementing such a BCI system is much more difficult than implementing a traditional synchronized BCI system, in which the user can only control a device at certain periods of time specified by the system [12].

BCI systems use specific features of a neurological phenomenon in the brain activity for the purpose of control. Various neurological phenomena can be used, including neural firing rates, changes in the Mu and Beta rhythms, movement-related potentials (MRPs), slow cortical potentials (SCPs) and P300. For a complete list of neurological phenomena used in BCI systems and pertinent references, please see [13]. In designing a feature extractor for a BCI system, an important factor that needs to be addressed is the variability in the chosen neurological phenomenon; i.e., the specifications of the neurological phenomenon may change from one user to another. For example, it has been shown that the Mu and Beta frequency bands [14] and the shape of an MRP [15] may vary from one user to another. As a result, if the features extractor does not extract user-specific features, the performance of the BCI system may degrade [16],

---

<sup>1</sup> A version of this chapter has been published. Fatourehchi, M., Bashashati, A., Birch, G.E. and Ward, R.K. "Automatic User Customization for Improving the Performance of an Asynchronous Brain Interface System", Journal of Medical & Biological Engineering and Computing, Vol.44, No.12, Dec 2006, pp.1093-1104.

or even detect an incorrect pattern [15]. A successful BCI system must therefore select features that correctly characterize the underlying neurological phenomenon of the specific user. We call this process user *customization of the feature extractor*.

Traditionally, BCI systems have not employed user customization for extracting features. Recent studies, however, showed that user customization of the feature extractor leads to improved performance for most users [5, 15-19]. User customization can be achieved either “manually” or “automatically”. In the manual user customization, the neurological phenomenon of interest is visually inspected by a human expert (usually through inspecting the ensemble average of many single trials); this is then followed by the expert determining the parameter values of the feature extractor [15]. This customization process has two main advantages: it is relatively fast and it is not computationally demanding. Thus, when the total number of users and EEG channels is small and the signal-to-noise ratio (SNR) of the neurological phenomenon is sufficiently high, the manual approach can be used for customizing the parameter values.

When the number of users grows, however, this process becomes increasingly time-consuming and exhausting. The problem becomes more challenging when features are extracted from a large number of EEG channels, since many EEG channels (and not only one or two) need to be visually observed. If the SNR of the neurological phenomenon is low, visual estimation becomes subjective and inaccuracies are introduced in the estimates of the parameter values. Furthermore, if some kind of preprocessing that changes the shape of the neurological phenomenon of interest is employed, then this change should be considered in the design of the feature extractor. For these reasons, an automatic user customization algorithm is desired.

In this Chapter, we employ automatic user customization of the feature extractor of a self-paced BCI system called the Low Frequency-Asynchronous Switch Design (the LF-ASD) [1]. The LF-ASD detects an Intentional Control (IC) command in the EEG signal. The IC command corresponds to an MRP pattern generated by the flexion of the right index finger. When the users are not in an IC state, they are said to be in a no-control (NC) state. In an NC state, a user may be idle or perform some action other than trying to control the BCI system. We chose the LF-ASD for this study because of our



intimate knowledge of this BCI system. Also, the LF-ASD has been used as the basis for potential design improvements by other researchers [6], and it is one of the few BCI systems that has been successfully tested online [20].

Because the shape of MRP patterns differs from one user to another, determining the specific design parameter values for each individual is expected to improve the performance of a BCI system. In [15] the parameter values of the feature extractor of the LF-ASD are estimated by a human expert (see Section 2.3 for details). It is shown that such user customization results in improved performance of the LF-ASD. There are some limitations, however, in the application of the proposed method. First, as mentioned above, the process can become very time-consuming, especially for a relatively large number of EEG channels, which is the case here (the LF-ASD uses 6 bipolar EEG channels). Second, the LF-ASD incorporates a pre-processing component that changes the shapes of MRPs. Third, the SNR of the MRPs is usually very low; this makes estimating the parameter values from the ensemble averages unreliable for some individuals [15].

In this study, we propose the use of a genetic algorithm (GA) to automatically estimate the shape of MRPs for each user and thus user customize the parameter values of the LF-ASD. A GA is a heuristic search method that provides a framework for effectively sampling large search spaces [21]. GAs are designed based upon the genetic processes of biological organisms, which evolve over many generations according to the principles of *natural selection* and *survival of the fittest*. By mimicking this process, they are able to *evolve* solutions to real-world problems. They have been shown to be effective in optimization problems where a large-dimensional feature space is involved, especially when the optimization problem cannot be solved by analytical tools [21, 22]. Since in this study we plan to automatically estimate the shape of the MRP pattern for each EEG channel, and thus we are dealing with a high-dimensional parameter space, we employ GAs. The use of a GA for automatic user customization of the LF-ASD was also motivated by the results of our earlier work in [23]. There, we used a GA to automatically customize the parameter values of the post-processing component in the LF-ASD for two individuals. The improvements in performance of the two individuals studied, demonstrated the effectiveness of employing a GA.

This study demonstrates that automatic user customization of the LF-ASD results in statistically significant improvements in the performance over the BCI system whose design parameter values are user customized by a human expert [15]. This finding further supports existing evidence that automatic user customization leads to performance improvement in BCI systems.

## **2.2 Background**

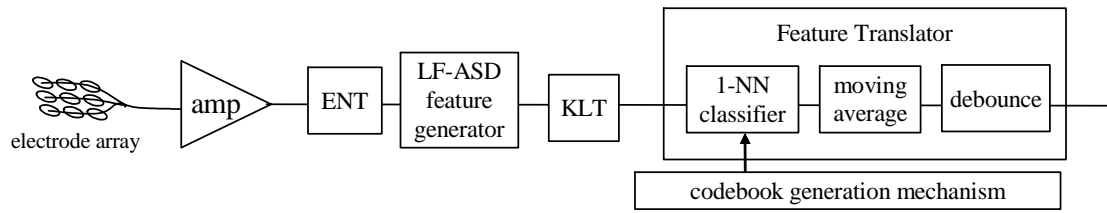
This section briefly reviews MRPs and the overall structure of the LF-ASD. MRPs are low-frequency potentials that start about 1-1.5 seconds before a movement. They have bilateral distribution and present maximum amplitude at the vertex [24-26]. An MRP is a robust phenomenon observed in the brain signal. It has been shown that there are similarities between the shapes of MRPs resulting from a real execution of a movement and those resulting from an attempt to perform a movement [1]. In some BCI systems, MRPs have thus been chosen as the neurological phenomenon, from which the presence of an IC command is extracted [1, 27-29]. An MRP consists of different components, such as Bereitschaftspotential, a motor potential (MP), post-movement positive potential (PMPP), etc. [30]. Different BCI systems focus on the detection of different components. For example, in [31], Bereitschaftspotential are detected, whereas in [1], the whole MRP is detected. There are various methods for detecting the components of MRPs, such as using autoregressive parameters [27], wavelet transform [28] and Fourier transform [29]. One solution is to use a simple feature extractor that detects the peaks of MRPs, since the peaks are found to be robust over different individuals [1]. The LF-ASD system and its variations have used this idea for detecting MRP patterns [1, 15, 32].

The block diagram of the LF-ASD [15, 32] is shown in Figure 2-1. This design uses features extracted from six bipolar EEG channels located on the sensorimotor cortex. After amplification and low-pass filtering using a low-pass, linear phase FIR filter with a 4 Hz cut-off frequency, all six EEG channels are normalized with an energy normalization transform (ENT) [33].

The ENT, (see Figure 2-1), normalizes the input energy and has been shown to result in a better class separation by increasing the difference between the means of the IC and NC features [32-34]. The output of the ENT is calculated using

$$e(n) = \frac{x(n)}{\sqrt{\sum_{s=-(W_N-1)/2}^{s=(W_N+1)/2} x(n-s)^2}} \quad (2-1)$$

where  $x(n)$  is the input EEG channel,  $W_N$  is the width of the "sliding" window used to normalize  $x(n)$ , and  $e(n)$  is the normalized EEG channel (the output of the ENT). The only design parameter of ENT is  $W_N$ , i.e., the "normalization parameter". Its value was originally determined through using an exhaustive search on the data collected from one individual. This value was then used for all other individuals [33].

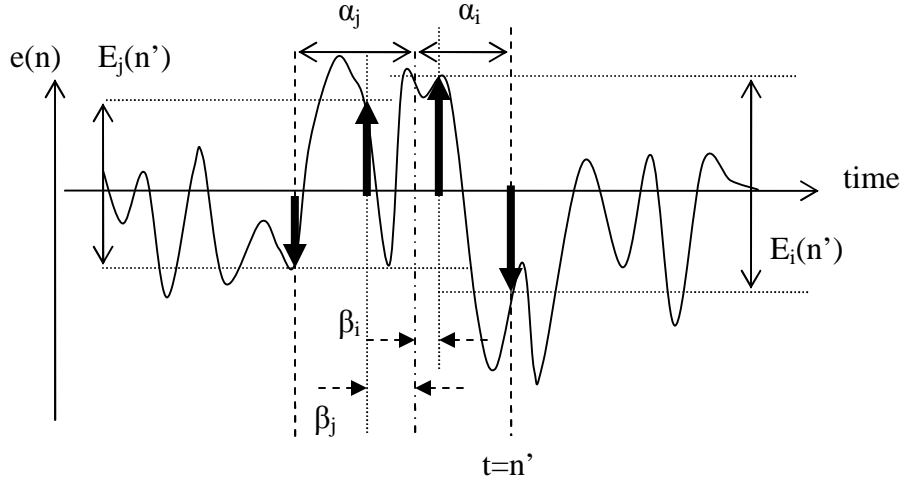


**Figure 2-1. Components of the LF-ASD system (from [32]).**

A specific feature generator is then applied to detect the presence of an MRP pattern in the single trial bipolar EEG signals [1]. Figure 2-2 shows the points used to calculate the features from a sample EEG signal at a particular point in time ( $t=n'$ ). As shown in Figure 2-2, each of the elemental features  $E_i(n')$  and  $E_j(n')$  is defined as the difference in  $e(n)$  at two points in time, described by (2-2) and (2-3) below:

$$E_i(n') = e(n' - \alpha_i + \beta_i) - e(n') \quad (2-2)$$

$$E_j(n') = e(n' - \alpha_i - \beta_j) - e(n' - \alpha_i - \alpha_j) \quad (2-3)$$



**Figure 2-2. Points selected by the feature generator when applied to a sample bipolar EEG signal.**

where  $e(n)$  is the ENT-normalized EEG signal. Throughout this Chapter, the above parameters  $(\alpha_i, \alpha_j, \beta_i, \beta_j)$  are referred to as the “delay parameters” and are used to estimate the shape of a bipolar MRP.

To emphasize the samples for which two large elemental features appear concurrently, compound features are defined by pairing the elemental features  $(E_i, E_j)$ , as shown below:

$$\begin{aligned} g_{ij}(n') &= E_i(n') \times E_j(n') & \text{if } E_i(n') \times E_j(n') \geq 0 \\ &= 0 & \text{otherwise} \end{aligned} \quad (2-4)$$

For robustness, the compound features are maximized over a window as follows:

$$G_{ij}(n') = \max \{g_{ij}(n'-8), g_{ij}(n'-7), \dots, g_{ij}(n'-1), g_{ij}(n')\} \quad (2-5)$$

Since there are six pairs of bipolar EEG channels, this procedure is repeated for each of these channels. Compound features of each of the six EEG signals then form a 6-dimension feature vector. The Karhunen-Loève Transform (KLT) is used to reduce the 6-dimensional feature space produced by the feature generator to a 2-dimensional space[35]. A 1-nearest neighbor (1-NN) classifier is used as the feature classifier. The

codebook generation mechanism for the classifier is explained elsewhere [1]. Finally, a moving average and a debounce algorithm are employed to improve the classification accuracy of the system by reducing the number of false activations (for details, see [1, 32]). After training, the LF-ASD classifies the input patterns as one of two classes: NC or IC.

## 2.3 Problem statement

In designing the LF-ASD, the parameter values of the ENT and the feature generator must be estimated. The aim is to determine these estimates so that it is possible to detect MRP patterns in a single trial. The ENT has one parameter to be determined. This is the window size,  $W_N$ . Its value should be estimated for each of the six EEG channels. The feature generator has four delay parameters ( $\alpha_i, \alpha_j, \beta_i, \beta_j$ ) for each of the six EEG channels, resulting in a total of 24 delay parameters whose values should be estimated. This means that, to detect the presence of an MRP pattern, the values of 30 parameters should be determined. For the rest of this Chapter, we refer to these 30 parameters as the “design parameters”. These parameters were originally estimated by a human expert from the ensemble average of MRP patterns for one individual and then were used for all subsequent individuals [1, 32]. As the MRP pattern related to a specific movement may differ from one individual to another, using the same design parameter values for all individuals may lead to erroneous results. Therefore, the design parameter values should be estimated for each individual. The same argument applies to any BCI system that uses a user-dependant pattern for its IC state.

When determining the design parameter values, two points should be considered. First, these values could not be determined using an exhaustive search approach. Without having an efficient automatic method, it is prohibitively time consuming to determine all parameter values simultaneously by using an exhaustive search method. Second, an improper choice of design parameter values may lead the BCI system to detect an incorrect pattern in the EEG signal. This, in turn, degrades the performance of the system, since the detected pattern would not correspond to an MRP pattern as it may have resulted from a particular artifact.

To improve the performance of the LF-ASD, the  $(\alpha_i, \alpha_j)$  delay parameter values were user customized by a human expert in [15]. The  $\beta_i$ 's and  $\beta_j$ 's were set to 0, equal  $\alpha_i$  and  $\alpha_j$  values were used for each of the six EEG channels and the size of the normalization window was fixed for all individuals and for all EEG channels. For each individual, the MRP pattern associated with the flexion of the right index finger was determined using the ensemble average of the MRP pattern. The delay parameter values were then estimated by visually inspecting the user's ensemble average of the MRP patterns. The rationale behind using the ensemble average was that it enhanced the SNR, and that the resulting waveform better showed the desired pattern that the LF-ASD aimed at detecting. As for the normalization parameters, since no analytical method for estimating these values existed, the values found earlier in [33] by trial and error were used. The data of eight individuals were analyzed [15]. Improvements from 2.0% to 6.8% were reported for four individuals, but the results for the rest of the individuals did not improve [15].

Although implementing the above customization approach seems straightforward, there were some problems associated with it. First, estimating the delay parameter values from the ensemble averages was not trivial. The number of available trials had a significant effect on the quality of the generated ensemble averages and ultimately on the estimated values of the delay parameters. For some individuals, there were a number of closely located peaks in the ensemble averages that made the estimation of the delay parameter values very difficult. As a result, several points had to be tested before the desired delay parameter values were estimated [15]. Also, the values of the normalization parameters and the delay parameters were not estimated simultaneously. Since the ENT was applied first, the delay parameter values were estimated subsequently. *Thus, for each value of the normalization parameter of the ENT, the delay parameter values had to be estimated.* Since no analytical method currently exists for estimating the normalization parameter values, the resulting estimates of the delay parameter values may not be reliable. Finally, the amount of improvement in the performance of the system found in [15] over that of the non-user customized system [32], was not as high as expected. This

is probably due to the fact that *estimating the delay parameter values based on the ensemble averages does not guarantee optimal performance in single-trial analysis*.

To address these limitations, in the next section we propose the use of a GA to automatically user customize the design parameter values.

## 2.4 Methods

In applying GAs to select the parameter values, each parameter of interest is first coded in the form of a randomly generated binary string. Each bit in this binary string is called a gene. The concatenation of all the binary strings forms a “chromosome”, and the set of “chromosomes” forms a “population”. Each chromosome is then evaluated and a fitness value assigned. For example, the fitness value can be the classification accuracy of the BCI system for a particular set of parameter values. The chromosomes are then combined using operators such as “selection”, “crossover” and “mutation” in order to generate new chromosomes. The “selection” operator selects a proportion of the existing population to breed a new generation. The selected chromosomes are usually the ones with higher fitness compared to other chromosomes in the population. After selection of the “fitter” chromosomes, a pair of "parent" chromosomes is selected for generating the “child” chromosomes. A child chromosome is a new solution that typically shares many of the characteristics of its "parents". The “crossover” operator ensures that this is the case by copying some of the genes of each parent to the child. The “mutation” operator is used to maintain genetic diversity from one generation of a population to the next. This process is repeated until a new population of chromosomes is generated. It is expected that the population evolves gradually and that fitness improves over generations. This process is continued until some criteria for stopping the GA is met [21].

The GA we apply for user customization has the following characteristics. Each chromosome consists of a concatenated binary version of 31 parameter values. These parameters comprise the 30 design parameters previously stated and the “scale factor” parameter, which determines the operating point of the BCI system on the receiver operating characteristic (the ROC) curve. The ROC curve shows the relationship between

the true positive (TP) and the false positive (FP) results for each parameter configuration (for more details on scale factor and plotting ROC curve for the LF-ASD, see [32]). The function of the scale factor is explained below.

The width of the normalization window was chosen to be from 0 to 1.5 seconds. The initial values of the delay parameters were visually estimated from the ensemble averages of the MRP patterns in the training data with the ENT removed from the system. For simplicity, the same initial delay values were chosen for all channels. The ranges for the delay parameter values were then chosen as follows (all numbers refer to sample numbers):

$$\text{Range of } \alpha_i: [\alpha_{i-est} - 32 \text{ to } \alpha_{i-est} + 96]$$

$$\text{Range of } \alpha_j: [\alpha_{j-est} - 96 \text{ to } \alpha_{j-est} + 32]$$

$$\text{Range of } \beta_i: [-32 \text{ to } +32]$$

$$\text{Range of } \beta_j: [-32 \text{ to } +32] \tag{2-6}$$

where,  $\alpha_{i-est}$  and  $\alpha_{j-est}$  are the approximate values of the delay parameters estimated from the ensemble averages. These parameter ranges were chosen to cover the range over which the peaks, associated with the pattern shown in Figure 2-2, are expected to occur. Their values, thus, give an estimation of the shape of the MRP pattern. The range of the scale factor (which determines the operating point on the ROC curve) was chosen as from 0.1 to 4. Our experience has shown that this selection covers the range of the operating points on the ROC curve of the LF-ASD that should be at low FP rates [32].

Following an initial estimate of the delay parameter values, a suitable fitness function for the GA was chosen as follows. A confusion matrix, shown in Table 2-1, was used to summarize the classification performance of a 2-state self-paced BCI system. In Table 2-1, the FP rate is the percentage of misclassifying a NC state as an IC state, the true negative (TN) rate is the percentage of correctly classifying an NC, the TP rate is the percentage of correctly classifying an IC and the false negative (FN) rate is the percentage of misclassifying an IC state as an NC state. A suitable fitness function for a self-paced BCI should be able to effectively summarize the confusion matrix. For a two-state self-paced BCI system such as the LF-ASD, we have



$$FN(\%) = 100 (\%) - TP (\%)$$

and

$$TN(\%) = 100 (\%) - FP (\%) \quad (2-7)$$

Based on (2-7) , the fitness function needs to contain only TP and FP rates. One choice of a good fitness function can be one that maximizes the TP rate for a reasonably low fixed FP rate. This choice is based on our previous results, where it was found that an FP rate above 2% caused excessive frustration and distraction in users using a self-paced BCI system [32]. Thus, it is important to keep the FP rates below 2%.

**Table 2-1. The confusion matrix for a 2-state self-paced BCI system.**

<div style="text-align: center;"> <div style="display: inline-block; transform: rotate(-45deg);"> Actual Class Predicted Class </div> </div>	IC	NC
	TP	FN
NC	FP	TN

Our earlier attempts at calculating a suitable performance measure based on the confusion matrix were based on reporting the TP rate at a fixed FP rate (which was set at 2%; see [32] for details). In order to achieve this, various points on the ROC curve were analyzed by varying the scale factor until a desired point, with an FP rate of 2%, was found. Such an approach is undesired for calculating the fitness function because of the huge computational load involved. Currently, each evaluation of the fitness function, including training the classifier and evaluating the system on the validation set, takes about two minutes on a PC with a Pentium IV 2.8 GHz CPU and 512 MB of RAM. Since finding a specific point on the ROC curve requires several such evaluations, and this process should to be repeated for all chromosomes in the population, the computational load increases dramatically. To be more specific, if the time needed for each evaluation of the fitness function is denoted by  $T_{Evaluations}$  , and  $N_{Chromosome}$  evaluations are needed to find a specific point on the ROC curve, and the GA needs to evaluate

$N_{Evaluations}$  chromosomes during its operation, the running time of the GA can be calculated as follows:

$$T_{GA} = N_{Evaluations} \times N_{Chromosome} \times T_{Evaluations} \quad (2-8)$$

Since  $T_{Evaluations} \approx 2$ , and  $N_{Evaluations}$  is in the order of thousands (e.g., 5000), it is evident that even for a small  $N_{Chromosome}$ ,  $T_{GA}$  will become very large. For the same reason, using the area under the ROC curve is not practical at this stage, since several points on the ROC curve should be estimated for a single evaluation of the fitness function. Our final configuration incorporated the FP rate as a constraint in the fitness function. We defined the fitness function as follows:

$$fitness(Chromosome) = \begin{cases} TP, & \text{if } FP \leq 2\% \\ 0.1TP, & \text{if } FP > 2\% \end{cases} \quad (2-9)$$

where the TP and the FP rates are expressed in %. In (2-9), the TP rates remain intact only for FP values less than 2%. For  $FP > 2\%$ , we attenuated the fitness of these chromosomes dramatically in order to prevent the less fit chromosomes from becoming active members of the population. Although such chromosomes had high TP rates, they also had high FP rates, and were considered “unfit” from a practical point of view.

The scale factor was added to the structure of the chromosome because of the expectation that the algorithm is able to find the value of the scale factor that yields the highest TP rate when  $FP \leq 2\%$ . In [23], we showed that this was indeed the case. The GA was able to find the scale factor value yielding the highest TP rate for  $FP \approx 2\%$ .

The remaining operators of the GA were chosen as follows. Tournament-based selection (tournament size =3) was used as the selection operator. Uniform crossover ( $p=0.9$ ) and uniform mutation ( $p=0.01$ ) operators were used. The sizes of the initial population and of the population in the next generations were chosen as 200 and 100, respectively. We used random initialization for initializing the GA. The number of evaluations was set to 5000 and this criterion was used for the termination of the algorithm. If the improvement in the best solution was found to be less than 1% for more than 10 consecutive generations, before reaching the total number of evaluations the

algorithm was terminated. Because of the computational load involved, we did not tune the GA parameter values such as the mutation and crossover rates.

## 2.5 Experimental results

In this section, the performance of the proposed algorithm is evaluated using the data collected from eight individuals. Off-line data were collected from users positioned 150 cm in front of a computer monitor. The EEG signals were recorded from six bipolar electrode pairs positioned over the users' supplementary motor area and primary motor cortex at  $F_1$ - $FC_1$ ,  $F_z$ - $FC_z$ ,  $F_2$ - $FC_2$ ,  $FC_1$ - $C_1$ ,  $FC_z$ - $C_z$ , and  $FC_2$ - $C_2$  in accordance with the International 10-20 System. Features extracted from these channels had been shown to provide more discriminant information for the separation of IC and NC features [1]. Electrooculography (EOG) activity was measured as the potential difference between two electrodes, placed at the corner of and below the right eye. The ocular artifacts were automatically rejected when the difference between the EOG electrodes exceeded  $\pm 25$   $\mu$ V. All signals were sampled at 128 Hz and referenced to the ear electrodes (see [1, 36] for details).

Data from four individuals with a high-level spinal cord injury (location of injury between C4-5 and C6-7 on the spinal cord) and four able-bodied individuals were used in this study. The individuals with spinal cord injury were coded as SCI (spinal cord injury) individuals and the able-bodied individuals were coded as AB individuals. None of the individuals with spinal cord injury had residual sensation or motor function in their hands. The users' descriptions are shown in Table 2-3.

The data were collected from the users as they performed a guided task. At each interval, a white circle of 2 cm diameter was displayed on the user's monitor for  $\frac{1}{4}$  second, prompting the user to attempt a movement. In response to this cue, the user had to attempt to flex his right index finger one second after the cue appeared. The 1-second delay was used to avoid visual evoked potential (VEP) effects from the cue, and the users were trained to estimate it. The 1-second time after the cue is denoted by the "time of the expected attempted movement (TEM)". Note that this is the time when the user is expected to attempt to perform the movement, and that this time may vary from one user

to another and from trial to trial. This task resulted in an attempted movement in individuals with spinal cord injury i.e., no physical finger movement, and an actual finger flexion in able-bodied individuals (see [36] for more details). For each user, an average of 80 trials was collected every day over a period of 6 days.

The data in the EEG signals were divided into segments, each of length equal to seven seconds. A 7-second window was wide enough to contain an MRP pattern as well as NC periods. A training set, a validation set and a test set were then randomly generated for each user from these 7-second windows. The training set was used to train the classifier. The validation set was used to select the optimal values of the design parameters using the proposed GA. The parameter values yielding the least error on the validation set were then selected. The performance of the system was evaluated using the test set. For each user, the epochs were randomly divided into five non-overlapping sets of equal size. The data in the first set were used for training, the data in sets two and three were used for estimating the parameters and the data in sets four and five were used for testing the performance of the selected model. The number of epochs in the training, validation and test sets for each user is reported in the fifth column of Table 2-3.

The features in (2-4) were generated by moving the feature generator over epochs, each of a 7-second length. Since the EEG signal is filtered to frequencies below 4Hz, the feature generator was shifted by 0.0625 seconds (8 samples), resulting in a total of 112 features in a 7-second epoch.

To determine whether or not an IC command was detected by the system, we defined a sliding window around the TEM. The length of this window was 1.5 seconds (from 0.5 seconds before the TEM to 1 second after the TEM). If an MRP pattern was detected at any time within any such window, the output of the BCI system was activated. This method is similar to those used by other researchers [3, 6, 7]. False positives were assessed in the periods before the system cue appeared and after the user was expected to perform the movement.

In [15], a 5-fold stratified cross-validation process was used to assess the performance of the LF-ASD. The trials in the training sets, validation sets and test sets were chosen randomly. The performance over different validation sets varied very little.

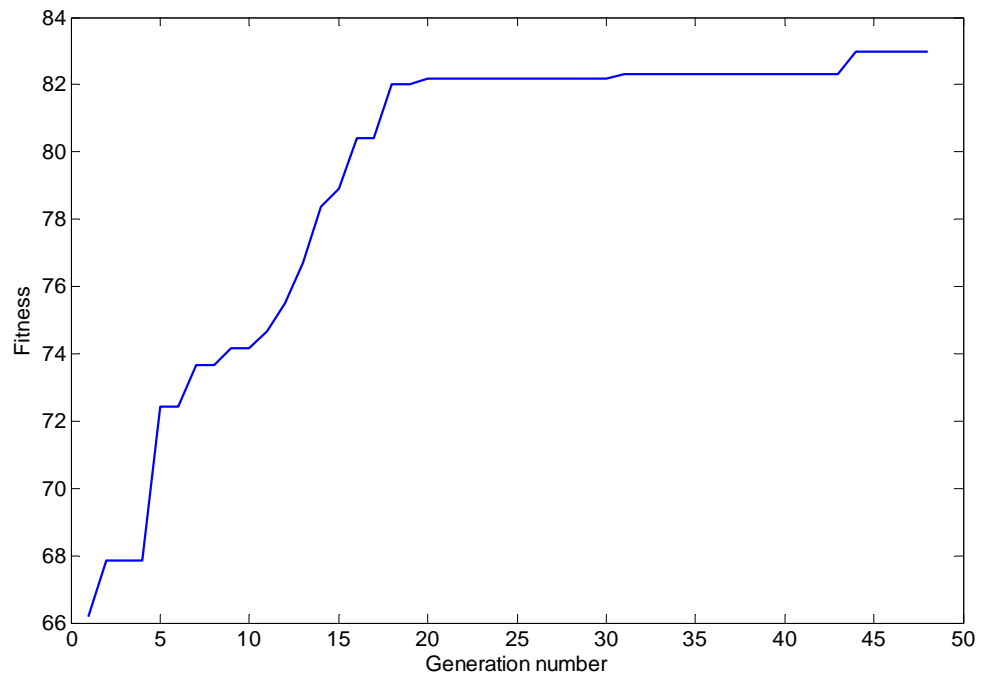
Thus, to save on computational time, we did not perform cross-validation over the different validation sets in this study, saving about 20% of the time needed for a 5-fold stratified cross-validation.

Figure 2-3 shows the fitness of the best chromosome in each generation as a function of the generation number for two representative individuals (AB2 and SCI4). Figure 2-3 clearly shows the evolution of the fitness of the best chromosome as the GA explores the search space. Please note that in the early stages of the GA, the improvement rate of the fitness of the best chromosome is fast. As the population evolves, the rate of improvement drops. This is because in the early stages of the GA, the value of the scale factor is not properly chosen. The design parameter values are also far from optimal. As the population evolves, the GA is able to find the scale factor value that yields the highest TP rate for FP=2%. This in turn results in a significant improvement in the fitness. As the generation number increases, the scale factor value is more properly set. The rate of improvement thus drops.

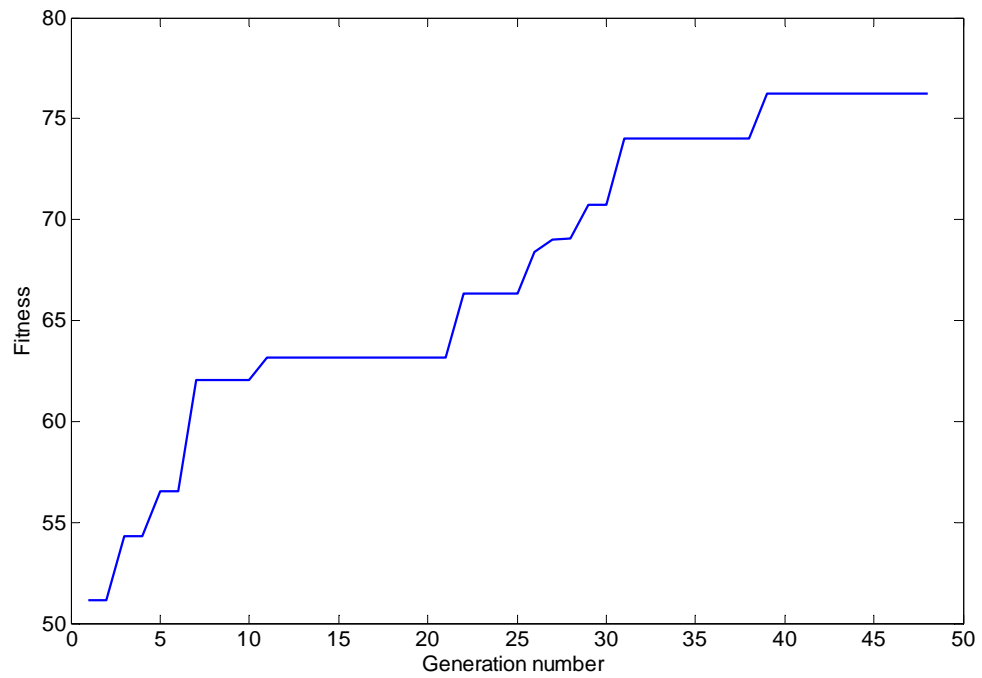
Table 2-2 summarizes the performance of the GA. In this table, the average fitness of the population, the fitness of the best chromosome and the fitness of the worst chromosome are reported for both initial and the final populations. As Table 2-2 shows, the average fitness of the initial population is very low. This result may be due to the following reasons:

- (1) The parameter values are randomly selected and are far from optimal.

- (2) The scale factor value is not properly set. Many chromosomes in the population are thus assigned a fitness value equal to zero since their FP rates are above the threshold of FP=2% (see (2-9)).



(a)



(b)

**Figure 2-3. The fitness of the best chromosomes as a function of the generation number for two representative individuals. a) AB2; b) SCI4.**

As the population evolves through generations, the GA is able to find the optimal value of the scale factor that yields the highest TP rates for FP=2%. Moreover, the choice of optimal parameter values leads to the generation of chromosomes with high fitness, resulting in an increased average fitness of the population. Since the GA found the suitable scale factor values for the chromosomes, the fitness of the weakest chromosome in the population is also dramatically increased.

**Table 2-2. Comparison of the fitness value of the initial and final populations (tested on the validation sets).**

<b>User</b>	<b>Initial population</b>			<b>Final population</b>		
	<b>Worst Fitness</b>	<b>Mean Fitness</b>	<b>Best Fitness</b>	<b>Worst Fitness</b>	<b>Mean Fitness</b>	<b>Best Fitness</b>
<b>AB1</b>	0	13.45	63.75	76.79	76.90	77.65
<b>AB2</b>	0	13.74	66.19	81.31	81.50	82.99
<b>AB3</b>	0	6.33	54.55	78.03	79.22	80.65
<b>AB4</b>	0	13.66	65.42	82.93	83.82	86.51
<b>SCI1</b>	0	16.00	64.68	75.88	77.46	78.34
<b>SCI2</b>	0	8.93	63.69	79.21	81.74	83.19
<b>SCI3</b>	0	15.77	64.46	70.76	72.42	73.33
<b>SCI4</b>	0	11.94	51.14	73.12	75.81	76.24
<b>Average</b>	0	12.48	61.73	77.25	78.61	79.86

The performance of the proposed “Automatically User Customized LF-ASD” system or ALF-ASD on the test sets is shown in Table 2-3 . We compared the performance of the ALF-ASD with that of the latest design of the LF-ASD whose parameter values tuned by a human expert [15]. The estimates of the delay parameter values in [15] are shown in Table 2-4. We tested both designs on 10 different randomly

chosen datasets. The TP results were then averaged over 10 sets for a fixed FP rate of 2%. Table 2-3 shows the results of running both algorithms on the data of all individual. The numbers in parentheses show the standard deviations. The last column shows the difference in TP rate for each user as well as the significance levels of the results, found by applying a two-sample t-test. Before carrying out the t-test, the Levene's test for equality of variances was used to determine whether the estimates of means in the t-test should be equal or unequal [37]. The results of Levene's test showed the homogeneity of the variances.

As Table 2-3 shows, the average TP rate was increased to 67.78% from 61.13% achieved using the method described in [15]. Such an improvement was statistically significant for 5 users ( $p < 0.01$ ) and non-significant for the remaining three ( $p > 0.05$ ). The average improvement in the TP rate for individuals with spinal cord injury was more than that of able-bodied individuals. To be more specific, the average TP rate for individuals with spinal cord injury was increased to 64.90% in the current study from 55.08% achieved using the customization by a human expert (an increase of 9.82%). As for able-bodied users, the average TP rate was increased to 70.76% in the current study from 67.17% achieved using the customization by a human expert (an increase of 3.58%).

Interestingly, the standard deviations of the TP rate also dropped from those achieved using the customization by a human expert. For individuals with spinal cord injury, the standard deviation of the TP rate decreased to 4.58% from 12.32% achieved using the customization by a human expert; while for able-bodied users, the standard deviation fell to 2.76% from 3.39% achieved using the customization by a human expert. Overall the standard deviation of the TP rate was reduced to 4.65% compared to 10.57% achieved using the customization by a human expert. These findings indicate that as we remove the inaccuracies introduced as the result of estimating the design parameter values by a human expert, the performance of individuals gets closer to each other. In other words, these results indicate that if the parameter values of the feature generator are correctly determined, the inter-subject variability in terms of performance will decrease.



**Table 2-3. TP rates of the LF-ASD and the ALF-ASD (FP=2%).**

User	Disability Description	Age	Gender	Number of epochs			LF-ASD (%)	ALF-ASD(%)	Difference in the TP(%)
				Train	Validation	Test			
<b>AB1</b>	N/A	56	M	128	256	256	65.5 (3.6)	70.0 (1.9)	4.5 ( $p<0.01$ )
<b>AB2</b>	N/A	43	M	103	206	206	72.2 (2.3)	72.6 (3.2)	0.5 ( $p>0.05$ )
<b>AB3</b>	N/A	31	F	133	266	266	66.2 (1.4)	67.5 (3.2)	1.2 ( $p>0.05$ )
<b>AB4</b>	N/A	45	M	97	194	194	64.7 (3.4)	72.9 (3.8)	8.2 ( $p<0.005$ )
<b>Average- (AB users)</b>	N/A	-	-	115.2 (17.9)	230.5 (35.8)	230.5 (35.8)	67.2 (3.4)	70.8 (2.8)	3.6 ( $p=0.07$ )
<b>SCI1</b>	C4/5 (17 y <sup>2</sup> )	53	M	128	256	256	63.5 (2.0)	64.6 (3.5)	1.1 ( $p>0.05$ )
<b>SCI2</b>	C4/5 (23 y)	56	M	103	206	206	66.0 (4.4)	70.6 (2.6)	4.5 ( $p<0.005$ )
<b>SCI3</b>	C5/6 (4 y)	33	M	91	182	182	39.1 (5.1)	59.3 (4.1)	20.2 ( $p<0.0001$ )
<b>SCI4</b>	C4/5 (5 y)	35	M	85	170	170	51.7 (5.7)	65.0 (5.3)	13.3 ( $p<0.0001$ )
<b>Average (SCI)</b>	-	-	-	101.7 (19.0)	203.5 (38.1)	203.5 (38.1)	55.1 (12.3)	64.9 (4.6)	9.8 ( $p=0.09$ )
<b>Overall Average</b>	-	-	-	108.5	217	217	61.1 (10.6)	67.8(4.6)	6.7 ( $p=0.06$ )

<sup>2</sup> Indicates number of years since injury.

**Table 2-4. Delay parameter values used in the design of the LF-ASD based on the ensemble averages of the MRP patterns in the training data set. Note that  $\beta_i$  and  $\beta_j$  are set to zero and that the same delay parameter values are used for the rest of the bipolar channels. The table is reproduced from [15].**

User	$\alpha_i$	$\alpha_j$
AB1	95	87
AB2	83	114
AB3	37	21
AB4	128	43
SCI1	112	99
SCI2	95	53
SCI3	39	64
SCI4	89	69

## 2.6 Discussion and conclusions

An important issue in the design of many BCI systems is the correct detection of the IC pattern (if present) for each user. Since the shape of a neurological phenomenon varies to some extent from one individual to another, it is necessary to consider this variation in the design of BCI systems. As a result, adjusting the parameter values of the feature extractor (user customization of the feature generator of the BCI system) is necessary for each user. If such user customization is done visually by a human expert, the results may have a subjective bias and unreliable; the customization process also becomes time consuming and exhausting. An automatic method therefore needs to be developed to perform user customization without the interference of a human expert.

In this Chapter, the effect of automatic user customization of the design parameter values of a self-paced BCI system was analyzed. More specifically, we proposed an automatic method for estimating the shape of an MRP used to drive the output of a self-paced BCI. Since MRPs have been used as the neurological phenomenon in a number of BCI systems, an automatic algorithm to estimate their shape can be used as an effective feature extraction method in those systems.

A GA was implemented to user customize a self-paced BCI called the LF-ASD. The LF-ASD is one of the few self-paced BCI systems that have been successfully tested online [20] and has been used by other researchers as well [6]. In design of the LF-ASD,

estimates of the delay parameter values obtained from the ensemble averages may be far from optimal because of the noisy nature of the EEG signals, the presence of artifacts and the psychological factors of each user. In addition, no analytical method currently exists for estimating the normalization parameter values. Until recently, these have been estimated in an ad-hoc manner through an exhaustive search of possible values. Automatic customization resolves this problem, since it estimates the parameter values depending on their associated cost functions.

We showed that by using a GA, the performance of the LF-ASD is improved to a great extent over the case where the design parameter values were estimated by a human expert [15]. This finding provides additional evidence that automatic user customization boosts the performance of a BCI system. Moreover, the designer is relieved from the cumbersome task of choosing the values of the feature extractor for each user.

One of the interesting findings of this Chapter is that the highest improvements were achieved in the performance of individuals with spinal cord injury when the delay parameter values were automatically customized. When the customization is done by a human expert, the highest improvements were achieved for able-bodied users [15]. However, the performance of individuals with spinal cord injury did not improve much. On the other hand, the results presented in Table 2-3 show that when the automatic user customization is used, the highest improvements were achieved for individuals with spinal cord injury. The average improvement in the TP rate was 3.58% for able-bodied users and 9.82% for individuals with spinal cord injury (resulting in the overall improvement of 6.68% ( $p=0.06$ )). This is probably due to the fact that individuals with spinal cord injury did not perform an actual movement, thus their MRP patterns were not as strong as those of able-bodied users. This resulted in noisier ensemble average MRP templates for the latter users, where visual estimation of the delay parameter values was not straightforward. The proposed automatic user customization method, however, deals with the optimization of the performance over single epochs and thus was able to find more suitable delay parameter values. Because of the low number of users, these findings cannot be generalized. They do, however provide some preliminary evidence that automatic user customization is necessary for achieving acceptable BCI performance for individuals with spinal cord injury.

We also found that for every individual, the values of the delay parameters found by the GA differed from one channel to another. There are two reasons for this result. First, the spatial distribution of the measured EEG signals was taken into consideration. Since the spatial distribution differs from one channel to another, it is expected that the delay parameter values should also differ. The other reason is the presence of the ENT. The value chosen for each normalization parameter changes the shape of the resultant EEG signals to some extent. Thus, for every value of the normalization window, a new set of delay parameter values should be estimated to correctly detect the presence of a bipolar MRP pattern in the EEG signal. The design parameter values found by the GA also differed from one user to another, providing further evidence that user customization is necessary to achieve acceptable performance values.

Comparison of the average results on test sets in Table 2-2 and Table 2-3 shows a drop of 12.05% in the performance. This drop in the performance indicates that the use of more sophisticated classifiers may be beneficial. For example, a support vector machines (SVM) can be used as a classifier, since not only it minimizes the empirical risk (the training error), it minimizes the confidence error as well (the test error) [38].

Future work includes finding better cost functions. Such a study has not been well explored in self-paced BCI systems. Finding better cost functions that can summarize the confusion matrix more effectively, is especially desired in optimization problems. Future work should also include online testing of the ALF-ASD. Specifically we shall investigate the performance of the ALF-ASD over time. Since the literature indicates that the shapes of MRPs may change from one day to another, a method that locally tunes the parameter values of the feature generator ahead of each session should be developed.

## **2.7 Acknowledgements**

This work was supported in part by the NSERC under Grant 90278-06 and the CIHR under Grant MOP-72711. The authors also would like to thank Mr. Craig Wilson for his valuable comments on this Chapter.

## 2.8 References

- [1] S. G. Mason and G. E. Birch, "A brain-controlled switch for asynchronous control applications", *IEEE Trans. Biomed. Eng.*, vol. 47, no.10, pp. 1297-1307, Oct. 2000.
- [2] G. E. Birch, P. D. Lawrence and R. D. Hare, "Single-trial processing of event-related potentials using outlier information", *IEEE Trans. Biomed. Eng.*, vol. 40, no.1, pp. 59-73, Jan. 1993.
- [3] 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. Rehabil. Eng.*, vol. 8, no.2, pp. 180-185, Jun. 2000.
- [4] R. Millan Jdel and J. Mourino, "Asynchronous BCI and local neural classifiers: an overview of the Adaptive Brain Interface project", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no.2, pp. 159-161, Jun. 2003.
- [5] R. Scherer, G. R. Muller, C. Neuper, B. Graimann and G. Pfurtscheller, "An asynchronously controlled EEG-based virtual keyboard: improvement of the spelling rate", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 979-984, Jun. 2004.
- [6] E. Yom-Tov and G. F. Inbar, "Detection of movement-related potentials from the electroencephalogram for possible use in a brain-computer interface", *Med. Biol. Eng. Comput.*, vol. 41, no.1, pp. 85-93, Jan. 2003.
- [7] G. Townsend, B. Graimann and G. Pfurtscheller, "Continuous EEG classification during motor imagery--simulation of an asynchronous BCI", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 12, no.2, pp. 258-265, Jun. 2004.
- [8] L. R. Hochberg, M. D. Serruya, G. M. Friehe, J. A. Mukand, M. Saleh, A. H. Caplan, A. Branner, D. Chen, R. D. Penn and J. P. Donoghue, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia", *Nature*, vol. 442, pp. 164-171, Jul 13. 2006.
- [9] J. F. Borisoff, S. G. Mason and G. E. Birch, "Brain interface research for asynchronous control applications", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no.2, pp. 160-164, Jun. 2006.
- [10] J. T. Francis and J. K. Chapin, "Neural ensemble activity from multiple brain regions predicts kinematic and dynamic variables in a multiple force field reaching task", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no.2, pp. 172-174, Jun. 2006.
- [11] G. Pfurtscheller, G. R. Muller-Putz, A. Schlogl, B. Graimann, R. Scherer, R. Leeb, C. Brunner, C. Keinrath, F. Lee, G. Townsend, C. Vidaurre and C. Neuper, "15 years of BCI research at Graz University of Technology: current projects", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no.2, pp. 205-210, Jun. 2006.
- [12] S. G. Mason and G. E. Birch, "Temporal control paradigms for direct brain interfaces - rethinking the definition of asynchronous and synchronous", in Proc. HCI International Conference, Las Vegas, USA, 2005.
- [13] S. G. Mason, A. Bashashati, M. Fatourehchi, K. F. Navarro and G. E. Birch, "A Comprehensive Survey of Brain Interface Technology Designs", *Annals of Biomedical Engineering*, vol. 35, no. 2, pp. 137-69, Feb 2007.
- [14] M. Pgegenzer and G. Pfurtscheller, "Frequency component selection for an EEG-based brain to computer interface", *IEEE Trans. Rehabil. Eng.*, vol. 7, no.4, pp. 413-419, Dec. 1999.

- [15] A. Bashashati, M. Fatourehchi, R. K. Ward and G. E. Birch, "User customization of the feature generator of an asynchronous brain interface", *Ann. Biomed. Eng.*, vol. 34, no.6, pp. 1051-1060, Jun. 2006.
- [16] M. Pregenzer and G. Pfurtscheller, "Frequency component selection for an EEG-based brain to computer interface", *IEEE Trans. Rehabil. Eng.*, vol. 7, no.4, pp. 413-419, Dec. 1999.
- [17] G. Blanchard and B. Blankertz, "BCI Competition 2003--Data set IIa: spatial patterns of self-controlled brain rhythm modulations", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 1062-1066, Jun. 2004.
- [18] Wenjie Xu, Cuntain Guan, Chng Eng Siong, S. Ranganatha, M. Thulasidas and Jiankand Wu, "High accuracy classification of EEG signal", in Proc. 17<sup>th</sup> Int. Conf. Pattern Recognition (ICPR 2004), vol.2, pp. 391-394, 2004.
- [19] T. N. Lal, M. Schroder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer and B. Scholkopf, "Support vector channel selection in BCI", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 1003-1010, Jun. 2004.
- [20] S. G. Mason, R. Bohringer, J. F. Borisoff and G. E. Birch, "Real-time control of a video game with a direct brain--computer interface", *J. Clin. Neurophysiol.*, vol. 21, no.6, pp. 404-408, Nov-Dec. 2004.
- [21] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, MA: Addison-Wesley Publishing Company, 1989.
- [22] T. Back, D. B. Fogel and T. Michalewicz, *Evolutionary Computation*. Bristol and Philadelphia: Institute of Physics Publishing, 2000.
- [23] M. Fatourehchi, A. Bashashati, R. K. Ward and G. E. Birch, "A hybrid genetic algorithm approach for improving the performance of the LF-ASD brain computer interface", in the Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, (ICASSP '05), vol. 5, pp. 345-348, 2005.
- [24] C. Babiloni, F. Carducci, F. Cincotti, P. M. Rossini, C. Neuper, G. Pfurtscheller and F. Babiloni, "Human movement-related potentials vs desynchronization of EEG alpha rhythm: a high-resolution EEG study", *Neuroimage*, vol. 10, no.6, pp. 658-665, Dec. 1999.
- [25] L. Deecke, B. Grozinger and H. H. Kornhuber, "Voluntary finger movement in man: cerebral potentials and theory", *Biol. Cybern.*, vol. 23, no.2, pp. 99-119, Jul 14. 1976.
- [26] M. Hallett, "Movement-related cortical potentials", *Electromyogr. Clin. Neurophysiol.*, vol. 34, no.1, pp. 5-13, Jan-Feb. 1994.
- [27] D. P. Burke, S. P. Kelly, P. de Chazal, R. B. Reilly and C. Finucane, "A parametric feature extraction and classification strategy for brain-computer interfacing", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 13, no.1, pp. 12-17, Mar. 2005.
- [28] E. L. Glassman, "A wavelet-like filter based on neuron action potentials for analysis of human scalp electroencephalographs", *IEEE Trans. Biomed. Eng.*, vol. 52, no.11, pp. 1851-1862, Nov. 2005.
- [29] M. Krauledat, G. Dornhege, B. Blankertz, F. Losch, G. Curio and K. -. Muller, "Improving speed and accuracy of brain-computer interfaces using readiness potential features", in Proc. EMBC Int. Conf., vol.6, pp.4511-4515, 2005.
- [30] R. Q. Cui and L. Deecke, "High resolution DC-EEG analysis of the Bereitschaftspotential and post movement onset potentials accompanying uni- or bilateral voluntary finger movements", *Brain Topogr.*, vol. 11, no.3, pp. 233-249, Spring. 1999.

- [31] B. Blankertz, C. Schürer, G. Dornhege and G. Curio, "Single trial detection of EEG error potentials: A tool for increasing BCI transmission rates", in Proc. Int. Conf. Artificial Neural Networks (ICANN'02), pp. 1137-1143, 2002.
- [32] J. F. Borisoff, S. G. Mason, A. Bashashati and G. E. Birch, "Brain-computer interface design for asynchronous control applications: improvements to the LF-ASD asynchronous brain switch", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 985-992, Jun. 2004.
- [33] Z. Yu, S. G. Mason and G. E. Birch, "Enhancing the performance of the LF-ASD brain-computer interface", in Proc. 2nd Joint IEEE-EMBS/BMES Conference, Houston, TX, USA, vol. 3, pp.2443-2444, Oct. 2002.
- [34] Z. Yu, S. G. Mason and G. E. Birch, "Impact of an energy normalization transform on the performance of the LF-ASD brain computer interface", in the Proc. Advances in Neural Information Processing Systems (NIPS'03), 16, pp. 725-732, 2003.
- [35] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Prentice Hall, 1984,
- [36] G. E. Birch, Z. Bozorgzadeh and S. G. Mason, "Initial on-line evaluations of the LF-ASD brain-computer interface with able-bodied and spinal-cord subjects using imagined voluntary motor potentials", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 10, no.4, pp. 219-224, Dec. 2002.
- [37] K. A. Brownlee, "Statistical theory and methodology in science and engineering", *A Wiley Publication in Applied Statistics*, New York: Wiley, 1965, 2nd Ed., 1965.
- [38] H. Yoon, K. Yang and C. Shahabi, "Feature subset selection and feature ranking for multivariate time series", *IEEE Trans. Knowledge and Data Eng.*, vol. 17, no.9, pp. 1186-1198, 2005.

## CHAPTER 3 APPLICATION OF A HYBRID WAVELET FEATURE SELECTION METHOD IN THE DESIGN OF A SELF-PACED BRAIN COMPUTER INTERFACE SYSTEM<sup>3</sup>

### 3.1 Background

A successful brain computer interface (BCI) system enables individuals with severe motor disabilities to control object in their environment (such as a light switch, a neural prosthesis or a computer) by using only their brain signals. Such a system measures specific features of a person's brain signal that relate to his or her intent to affect control, and then translates them into control signals that are used to control a device [1, 2].

Brain computer interface systems are implemented in two ways: system-paced (synchronized) or self-paced (asynchronous). In system-paced BCI systems, a user can initiate a command only during certain periods specified by the system. In a self-paced BCI system, users can affect the output of the BCI system whenever they want, by intentionally changing their brain state. The state in which a user is intentionally attempting to control a device is called an *intentional control* (IC) state. At other times, users are said to be in a *no-control* (NC) state, where they may be idle, thinking about a problem, or performing some action other than trying to control the device[3, 4]. To operate in this paradigm, BCI systems should be designed to respond only when the user is in an IC state and to remain inactive when the user is in an NC state. So far, only a few BCI systems (e.g. [3, 5-10]) have been specifically designed and tested for self-paced

---

<sup>3</sup> A version of this chapter has been published. Fatourehchi, M., Birch, G. E., and Ward, R. K., "Application of a Hybrid Wavelet Feature Selection Method in the Design of a Self-paced Brain Interface System", *Journal of NeuroEngineering and Rehabilitation*, Vol.4, No.1, Apr 2007.



control applications. But as recognized in [2], self-paced BCI systems deserve more attention.

The discrete wavelet transform (DWT) can be used as a powerful feature extraction tool to extract time-frequency features similar in shape to that of a particular wavelet function. It therefore has an advantage over other feature extraction methods that operate in only one domain, such as the Fourier transform, or autoregressive modeling.

The DWT has been extensively applied in the analysis of event-related potential (ERP) because of its ability to effectively explore both the time and frequency information of these signals [11, 12]. It has also been successfully used to generate wavelet features in BCI systems. In [13], DWT was employed in the design of a synchronized BCI system that used wavelet coefficients extracted from slow cortical potentials (SCPs) as well as other ERPs. This system performed better than other designs that used EEG time series and a mixed filtering method. In [14], the energies of various frequency bands decomposed by a wavelet packet transform (18 frequency bands in total) were used as features in detecting different movement patterns in a self-paced BCI system. These features were linearly combined to generate a single feature, with coefficients of the linear mapping determined by a genetic algorithm (GA). In [15], a custom-made wavelet function was employed in two different studies: the detection of P300 in a single EEG channel, and the detection of the Bereitschaftspotential from two EEG channels. In [16], a weighted linear combination of all available wavelet coefficients (15 in total) extracted from a single EEG channel was used to detect P300 patterns. To estimate weights for each feature in the linear combination, a neural network was employed. Finally, in [17], investigators applied DWT to extract the 0-4Hz component of the EEG signal in a P300-based BCI system. Based on the above encouraging results, in this study we explore applying DWT to extract movement-related potential (MRP) features for driving a self-paced BCI system.

Although the above BCI studies provide promising evidence that DWT can be employed to extract features in BCI systems, two main issues still need to be addressed. First, studies that used discrete wavelet coefficients as features (rather than wavelet-filtered EEG signals), used only one or two EEG channels. In these cases, the resulting

dimensionality of the space does not pose a serious problem, since it is not very large. Having a BCI system that uses data recorded from only one or two electrodes seems very appealing, since the setup is fast and uses less hardware/software infrastructure. Most of the above-mentioned papers, however, achieved a relatively high degree of classification error when only one or two EEG channels were used. For example, in [16], the reported error rates were relatively high (nearly 40% error). In [17], where wavelet-filtered EEG signals were used, the system did not perform well (30% misclassification). For the only self-paced BCI system that has applied wavelet coefficients so far [14], false discovery rates (the percentage of hits that were not true positives) varied up to 67%, however, the authors did not indicate the number of NC epochs used in their study, so critical commentary on the performance of their BCI system cannot be made. The invasiveness of the recording technology of the BCI system in [14] is also an important issue that needs to be considered.

The above observations strongly motivate the use of additional EEG electrodes in BCI systems. With signals recorded from multiple channels, we can explore spatial information, which is expected to yield improvements in classification performance.

Another issue that must be addressed when using DWT to extract features in BCI systems is the feature selection procedure. That is, how many features should be selected and how should they be selected? In [13], all of the 64 wavelet features used for classification were extracted from only one EEG channel. In [15], because of the computational limitations affecting the classifier, only a number of top wavelet features (ranked by the amount of discriminability) were selected. None of the above-mentioned approaches yielded best results (since the feature selection process used was necessarily not optimal). Using all features does not necessarily provide the best results, because some of the less discriminant features may degrade the classifier's performance [18]. On the other hand, using only few features that have the highest rank (and filtering out the rest of features) does not necessarily lead to the optimal classification performance, since there is no guarantee that using only top-ranked features leads to the best classifier performance [19].

Based on the related literature review, we postulate that the information extracted from *multiple- electrode signals* is necessary for achieving acceptable performance. This in turn leads us to the high dimensionality problem of the feature space; since the feature space dimension is directly affected by the number of electrodes used as well as by the number of features per EEG signal. Since not all the wavelet coefficients provide discriminatory information between the output classes, we postulate that features that better discriminate between the output classes need to be selected to obtain better classification performance. A mechanism for selecting the most discriminating features is thus needed.

Wrapper methods, such as GAs, use the classifier's performance to evaluate a particular feature vector. They provide a good solution for finding the features that work well together by choosing the ones that lead to better classifier performance [20]. The downside of using wrapper methods is time inefficiency. As the dimension of the search space increases, it becomes harder for a wrapper method to find a suitable subset of features that lead to a high performance.

In order to benefit from the advantages of both filter and wrapper methods, we decided to employ a hybrid approach. Features carrying the least discriminative information about the output classes were filtered out first. Then a wrapper method was applied to the reduced feature space to find the features that work well together, i.e., the combination that leads to the best classification performance. We used mutual information (MI) in the filtering stage. Mutual information is a powerful tool for ranking features based on the amount of discriminative information each carries [21]. We then applied a GA in a wrapper approach to select the features that lead to the best classification performance. Genetic algorithms are heuristic methods that can effectively sample large search spaces [22]. They are implemented based on the principles of evolutionary biology, and evolve over many generations. By mimicking this process, GAs are able to evolve solutions to real-world problems. They have been shown to be useful tools in automatically customizing many practical systems [22, 23].

We used a support vector machine (SVM) to classify the selected features into one of two classes: no control (NC) or intentional control (IC). The results of this study

show that applying the proposed approach to the offline data collected from four able-bodied individuals yields low false positive (FP) rates at a reasonably high true positive (TP) rate. We also examine the spatial distribution of the selected features. We show that this distribution varies considerably from one individual to another. This finding shows the importance of user customization of BCI systems.

### 3.2 Data collection

People with severe motor disabilities cannot physically execute certain movements such as a finger flexion, but they are usually able to *attempt* it. Several studies have shown that recordings of brain signals obtained from attempted and real movements for able-bodied individuals bear many similarities [14, 24-29]. Based on these studies, both attempted and executed movements have been shown to activate similar cortical areas and to generate similar movement patterns. This evidence enables us to base our analysis on the data of able-bodied individuals, who actually execute a particular movement. It is then possible to detect the occurrence of the control command by analyzing signals such as electromyography (EMG) signal or the output of an actual switch. Such signals can be used to label the brain signals and to evaluate the performance of a BCI. The data analysis of individuals with motor disabilities was thus left to future studies.

The data of four (three male and one female) able-bodied individuals were used in this study. All individuals were right-handed and between 31 and 56 years old. They had all signed consent forms prior to participation in the experiment.

Individuals were positioned 150 cm in front of a computer monitor. The EEG signals were recorded from 13 monopolar electrodes positioned over the individuals' supplementary motor area and primary motor cortex (according to the International 10-20 System at F<sub>1</sub>, F<sub>z</sub>, F<sub>2</sub>, FC<sub>3</sub>, FC<sub>1</sub>, FC<sub>z</sub>, FC<sub>2</sub>, FC<sub>4</sub>, C<sub>3</sub>, C<sub>1</sub>, C<sub>z</sub>, C<sub>2</sub> and C<sub>4</sub> locations). Electrooculography (EOG) activity was measured as the potential difference between two electrodes, placed at the corner of and below the right eye. An ocular artifact was considered present when the difference between the EOG electrodes exceeded  $\pm 25 \mu\text{V}$ . All signals were sampled at 128 Hz and referenced to ear electrodes (see [30] for details

of the data recording). The recorded signals were then saved on the computer and converted to bipolar EEG signals by calculating the difference between the adjacent EEG channels. This procedure was used since it has been shown that bipolar electrodes generate more discriminating MRP features than monopolar electrodes do [3]. This conversion generated the following 18 bipolar EEG channels:  $F_1-FC_1$ ,  $F_1-F_z$ ,  $F_2-F_z$ ,  $F_2-FC_2$ ,  $FC_3-FC_1$ ,  $FC_3-C_3$ ,  $FC_1-FC_z$ ,  $FC_1-C_1$ ,  $FC_z-FC_2$ ,  $C_1-C_z$ ,  $C_2-C_4$ ,  $FC_2-FC_4$ ,  $FC_4-C_4$ ,  $FC_2-C_2$ ,  $FC_z-C_z$ ,  $C_3-C_1$ ,  $C_z-C_2$  and  $F_z-FC_z$ .

Data were collected from individuals as they performed the following guided task. At each interval, a white, 2cm diameter circle was displayed on the individual's monitor for  $\frac{1}{4}$  second, prompting the individual to attempt a movement. In response to this cue, the user had to perform a right index finger flexion one second after the cue appeared. The 1-second delay was used to avoid visual evoked potential (VEP) effects caused by the cue (see [31] for more details). For each individual, an average of 80 IC epochs were collected every day over a period of 5 days.

An IC epoch consisted of data collected over an interval containing the movement onset (measured as the finger switch activation) if no artifact was detected in that particular interval. The interval starts at  $t_{start}$  seconds before movement onset and ends at  $t_{finish}$  seconds after it. There were limitations in choosing the total length of  $(t_{start} + t_{finish})$ . If the length of  $(t_{start} + t_{finish})$  increases, more artifacts may be present in an IC epoch. As a result, the number of training epochs that are artifact-free based on the criterion used to reject ocular artifacts will be reduced. If the length of  $(t_{start} + t_{finish})$  is too short, a poor exploration of potential features results. Since a simple finger flexion MRP usually starts about 1.5 seconds before the movement and returns back to the normal baseline around 1 second after the movement [32], data obtained from 1.5 seconds before to 1.0 second after the movement onset were analyzed (i.e.,  $t_{start}=1.5$  seconds and  $t_{finish}=1.0$  second).

NC epochs were selected as follows. A window of width  $(t_{start} + t_{finish})$  seconds was considered ( $t_{start}=1.5$  seconds and  $t_{finish}=1.0$  second). To extract NC epochs, the window was shifted over each EEG signal recorded during NC sessions by a step of 16 samples (0.1250 sec). Wavelet coefficients were extracted for each epoch that did not contain artifacts.

### 3.3 Method

The overall structure of the proposed scheme is shown in Figure 3-1. EEG signals were checked for the presence of EOG artifacts. The contaminated epochs were rejected, as explained in Section 3.2.

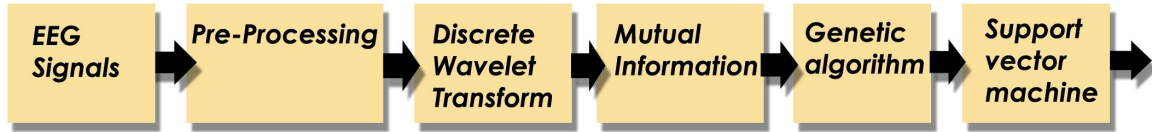


Figure 3-1. The overall structure of the proposed hybrid method for extracting MRP features.

The continuous wavelet transform (CWT) is defined as the convolution of the signal  $x(t)$  with the wavelet functions  $\psi_{a,b}(t)$ , where  $\psi_{a,b}(t)$  is the dilated and shifted version of the *wavelet function*  $\psi(t)$  and is defined as follows:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (3-1)$$

where  $a$  and  $b$  are the scale and translation parameters, respectively. The CWT maps a signal of one independent variable  $t$  into a function of two independent variables  $a, b$ . This procedure is redundant and not efficient for algorithmic implementations. Therefore, it is more practical to define the wavelet transform at a discrete scale  $a$  and a discrete time  $b$  by choosing the set of parameters (such a transform is called a discrete wavelet transform, or DWT), such that

$$a_j = 2^{-j}, b_{j,k} = 2^{-j}.k \quad (j, k \text{ are integers}) \quad (3-2)$$

The contracted versions of the wavelet function will match the high-frequency components of the original signal and the dilated versions will match the low-frequency oscillations. Then by correlating the original signal with the wavelet functions of different sizes, the details of the signal at different scales are obtained. The resulting correlation features can be arranged in a hierarchical scheme called multi-resolution decomposition [33] which separates the signal into “details” at different frequency bands and a coarser representation of the signal called an “approximation”.

In this study, the *rbio3.3* wavelet from the B-spline family was chosen as the wavelet function because it has some similarities with the shape of the classic bipolar MRP pattern. Using a 5-level decomposition method resulted in wavelet coefficients corresponding to the following frequency bands (the sampling frequency was 128 Hz): [32-64], [16-32], [8-16], [4-8], [2-4], and [0-2] Hz.

Based on the previous findings in [3], which showed that MRP features are mostly located in the frequency range below 4Hz, only the lowest frequency bands (i.e., 0-2Hz and 2-4Hz) were considered for further analysis of MRPs. Even with this reduced feature space, the resulting feature space dimension ( $N_{features}$ ), which is the product of the number of electrodes ( $N_{electrodes}$ ) and the number of wavelet features per EEG signal ( $N_{wavelet}$ ). That is,  $N_{features} = N_{electrodes} \times N_{wavelet}$  remained very high. Thus, a feature selection procedure had to be used that could select the features that lead to optimal classification performance. This procedure should specify the selected EEG channels as well as the features selected per channel.

We devised a hybrid feature selection algorithm to meet these requirements. Mutual information (MI) was employed in the filtering stage and a GA was then used to select the optimal set of features.

Although MI has been used elsewhere to filter out the less informative features [21, 34], it is *not* usually successful at finding features that lead to optimal classification performance. This is because when there are more than three feature dimensions, the calculation of MI is computationally demanding, and impossible for large feature spaces (since the calculation of MI requires the joint probability of features in a high dimension) [21, 34]. Thus, MI was only used in our algorithm to discard the least informative features based on the amount of information that each feature carries regarding the output classes.

The MI between the input feature vector  $\mathbf{X}$  and the output classes  $\mathbf{Y}$  was calculated as follows:

$$I(\mathbf{X}, \mathbf{Y}) = H(\mathbf{Y}) - H(\mathbf{Y}|\mathbf{X}) \quad (3-3)$$

where

$$H(\mathbf{Y}) = -\sum_{j=1}^M P(y_j) \cdot \log_2 P(y_j) \quad (3-4)$$

$$H(\mathbf{Y}|\mathbf{X}) = -\sum_{i=1}^N \sum_{j=1}^M P(x_i) \cdot P(y_j|x_i) \cdot \log_2 P(y_j|x_i) \quad (3-5)$$

$$P(y_j) = \sum_{i=1}^N P(x_i) \cdot P(y_j|x_i) \quad (3-6)$$

In these formulae,  $I$  represents the mutual information between  $\mathbf{X}$  and  $\mathbf{Y}$ , where  $\mathbf{X} = \{x_i\}$ , ( $i = 1, 2, 3, \dots, N$ ) and  $\mathbf{Y} = \{y_j\}$ , ( $j = 1, 2, 3, \dots, M$ ),  $N$  is the number of input states and  $M$  is the number of outputs states ( $M=N=2$ , since the input and output can only take two values: IC and NC),  $P(x_i)$  is the probability of occurrence of an input state  $x_i$ ,  $P(y_j)$  is the probability of the output class  $y_j$  when the input is unknown, and  $P(y_j|x_i)$  is the probability of the output class  $y_j$  when the input state  $x_i$  is known.

For each individual, the wavelet coefficient (feature) values corresponding to all the training set data were calculated. Then, using histograms with 10 bins each, the probability function of each feature was estimated and its mutual information with each of the output classes was calculated. The values of MI were calculated for all  $N_{features}$  features and then ranked in descending order. The top  $L$  features were then selected. In this study, we arbitrarily chose  $L=50$  to avoid having a feature space with a very high dimension.

After reducing the dimension of the feature space, a GA was used to select a subset of  $m$  features from the top  $L$  features. To represent each possible combination of features, a binary chromosome of length  $L$  was defined. The bit  $i$  of the binary chromosome specified whether or not the feature  $i$  was selected by the GA. A value of “1” indicated the presence of feature  $i$  and a value of “0” indicated its absence in a chromosome.

An important decision in the design of a GA is the definition of a proper fitness function. In the proposed design, a suitable fitness function should consider at least three



objectives: maximizing the TP rate, minimizing the FP rate and minimizing the number of features selected by the hybrid feature selection procedure.

The classification performance of a 2-state, self-paced BCI system is usually determined by a confusion matrix, as shown in Table 3-1. In Table 3-1, the FP rate is the percentage of instances for which an NC epoch is misclassified as an IC epoch, the true negative (TN) rate is the percentage of NC epochs being correctly classified, the true positive (TP) rate is the percentage of IC epochs being correctly classified and the false negative (FN) rate is the percentage of misclassifying an IC epoch as an NC epoch. The fitness function should summarize this confusion matrix. For a 2-state self-paced BCI system, we have

$$FN(\%) = 100(\%) - TP(\%) \quad (3-7)$$

and

$$TN(\%) = 100(\%) - FP(\%) \quad (3-8)$$

**Table 3-1. The confusion matrix for a 2-state self-paced BCI system.**

Actual Class \ Predicted Class	IC	NC
IC	TP	FN
NC	FP	TN

Based on 3-7) and 3-8), only TP rates (TPR) and FP rates (FPR) need to be included in the fitness function. One example of a fitness function is a function that maximizes the  $\frac{TPR}{FPR}$  ratio. In this paper, the following objective function was used:

$$f(Z) = \begin{cases} 0, & TPR < 20\% \\ \frac{TPR(Z)}{FPR(Z)}, & TPR \geq 20\% \end{cases} \quad (3-9)$$

where  $Z$  is a chromosome and  $f$  is the fitness function. This fitness function gives a higher fitness level to chromosomes that generate a higher  $\frac{TPR}{FPR}$  ratio. We also postulated that TP rates below 20% were too low for the successful operation of a self-paced BCI system (since they correspond to detection of less than one IC out of every five IC states, which may lead to user frustration, even though the FP rates might be very low). Such chromosomes were considered “unfit” and were assigned a “0” fitness value.

Next, a lexicographic approach was applied for multi-objective optimization of the GA population [23]. Very briefly, in this approach, the objectives were ranked according to the priorities assigned to them prior to optimization. The objective with the highest priority was used first for comparing the members of the population. In our case, the average of  $\frac{TPR}{FPR}$  over the validation sets was first selected as the objective function with the highest priority. The chromosomes were then ranked in a single-objective fashion. Any ties were resolved by comparing the relevant chromosomes again with respect to objectives that were assigned lower priority. The other three objectives were chosen as (1) the average of FP rate over the validation sets, (2) the average of TP rate over the validation set, and (3) the number of features, resulting in four objectives per chromosome in the GA population. The 2<sup>nd</sup> and 3<sup>rd</sup> objectives were ordered such that for two chromosomes with the same  $\frac{TPR}{FPR}$  ratio, the one with the lower FP rate was considered to be the fit chromosome.

The remaining operators of the GA were tournament-based selection (tournament size =3), uniform crossover and uniform mutation. The sizes of the initial population and the population in the next generations were chosen as 100 and 50, respectively. We used random initialization to initialize the GA. Elitism was used to keep the best performing chromosome of each population in the subsequent populations.

The number of evaluations was set to 2000. If the improvement in the  $\frac{TPR}{FPR}$  ratio of the best solution was found to be less than 1% for more than 10 consecutive

generations, the algorithm was terminated. Because of the computational load, tuning the GA parameter values (such as the mutation and crossover rates) was not performed.

A support vector machine (SVM) that uses kernel-based learning was chosen to classify each chromosome in the GA population. In kernel-based learning, all of the beneficial properties of linear classification methods, such as simplicity, are maintained, but the overall classification is nonlinear in the input space, since the feature and input spaces are nonlinearly related [35]. Another reason for selecting a SVM as a classifier is that SVMs not only minimize the empirical risk (training error), they also minimize the confidence error (test error) [36]. We used the *LIBSVM* software [37], which has also been used in other BCI papers [38, 39].

The evaluation process was as follows. For each individual, IC and NC epochs were randomized and divided into training, validation and test sets. The training set was used to train the classifier, and the validation set was used to select the best set of features. The configuration yielding the best results on the validation set in the multi-objective sense mentioned above was selected, and the performance of the system calculated on the test set was reported. We used a five-fold nested cross-validation for evaluating the performance of the system. For each outer cross-validation set, 20% of the data were used for testing and the rest were used for training and model selection (selection of optimal subset of features). In order to select the models, the datasets were further divided into five folds. For each fold, 80% of the data were used for training the classifier and 20% were used for model selection.

To deal with the problem of unbalanced training sets (there were at least 20 times more NC epochs than IC epochs), the size of the NC training feature set was reduced to be the same as the size of the training IC feature sets. This was done by randomly selecting epochs from the NC training set.

### 3.4 Results

In this section, we present the offline analysis of the data of the four individuals described in Section 3.2. We performed a search on the classifier's parameters during the model selection. Our findings showed that a 5<sup>th</sup> degree polynomial kernel function

performed better than other kernel functions studied (linear, polynomial with a degree other than 5 (3, 4, 6 and 7) and RBF kernel).

Since a five-fold nested cross-validation was used for the performance evaluation, the results were averaged over five runs of the outer validation sets. The columns 1 to 5 of Table 3-2 show the individual identification number, the average TP rate on the test sets, the average FP rate on the test sets, the average  $\frac{TPR}{FPR}$  ratio and the average number of features selected by the hybrid feature selection process. The latest performance results of another state-of-the art self-paced BCI system (the LF-ASD) [40], applied to the data of individuals AB1 to AB4 are presented in columns 6 to 9 of Table 3-2. The numbers in parentheses are the standard deviations. As Table 3-2 shows, our proposed design achieved low FP rates for three of the four individuals (individuals AB1, AB2 and AB4) for a relatively high TP rate. For individual AB3, the TPR results on the test sets were low (although the FP rates remained less than 4%).

**Table 3-2. Comparison of the average TP, average FP rates, average  $\frac{TPR}{FPR}$  and the average number of features.**

Individual ID	Test Set (Current Study)			Number of features (Current Study)	Test Set ([40])			Number of Features ([40])
	TPR	FPR	$\frac{TPR}{FPR}$		TPR	FPR	$\frac{TPR}{FPR}$	
<b>AB1</b>	68.0 (4.8)	1.0 (0.3)	<b>68.0</b>	30.6 (1.1)	67.8 (1.4)	2.0	33.9	6
<b>AB2</b>	73.3 (2.6)	1.4 (0.4)	<b>52.4</b>	29.2 (3.3)	74.0 (1.7)	2.0	37.0	6
<b>AB3</b>	33.1 (14.0)	3.9 (1.0)	8.5	23.4 (2.4)	64.0 (1.3)	2.0	<b>32.0</b>	6
<b>AB4</b>	56.1 (4.9)	1.4 (0.7)	<b>40.0</b>	27.0 (2.8)	73.1 (1.8)	2.0	36.6	6
<b>Average</b>	57.4	1.9	30.2	27.5	69.7	2.0	34.9	6

Next, the spatial distributions of the selected features were examined. The average number of selected features per channel is shown in Table 3-3. The numbers in parentheses show the standard deviation over five runs of outer cross-validation. Figure 3-2 to Figure 3-5 show the number of selected features per channel for all individuals after applying the hybrid selection method (averaged over the number of cross-validation

**Table 3-3. The average number of selected features per channel after applying the hybrid feature selection algorithm.**

Individual ID Channel	AB1	AB2	AB3	AB4
F <sub>1</sub> -FC <sub>1</sub>	3.6 (1.1)	3 (1.2)	1.8 (0.8)	3 (0.7)
F <sub>1</sub> -F <sub>z</sub>	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	3.4 (0.5)
F <sub>2</sub> -F <sub>z</sub>	0.0 (0.0)	1.6 (0.9)	0.4 (0.5)	0.0 (0.0)
F <sub>2</sub> -FC <sub>2</sub>	0.2 (0.4)	2 (0.7)	0.8 (0.8)	0.4 (0.5)
FC <sub>3</sub> -FC <sub>1</sub>	1.0 (0.0)	1.0 (0.0)	1.6 (0.9)	0.0 (0.0)
FC <sub>3</sub> -C <sub>3</sub>	1 (0.71)	3.0 (0.0)	2.4 (1.14)	1.6 (0.5)
FC <sub>1</sub> -FC <sub>z</sub>	0.0 (0.0)	1.0 (0.0)	0.6 (0.5)	1.2 (0.84)
FC <sub>1</sub> -C <sub>1</sub>	4.6 (0.5)	2.8 (0.4)	0.0 (0.0)	1.2 (0.4)
FC <sub>z</sub> -FC <sub>2</sub>	0.0 (0.0)	2.2 (0.4)	0.6 (0.5)	0.0 (0.0)
C <sub>1</sub> -C <sub>z</sub>	1.6 (0.5)	0.4 (0.5)	3.6 (1.1)	1.2 (0.4)
C <sub>2</sub> -C <sub>4</sub>	0.6 (0.5)	2.2 (0.4)	4.4 (0.9)	2.6 (0.9)
FC <sub>2</sub> -FC <sub>4</sub>	4.2 (0.4)	1.6 (0.9)	2.2 (1.1)	3.4 (1.1)
FC <sub>4</sub> -C <sub>4</sub>	3.2 (0.45)	2 (1.0)	1.8 (0.8)	4.4 (0.5)
FC <sub>2</sub> -C <sub>2</sub>	2.0(0.0)	2.2 (0.4)	0.6 (0.5)	2.2 (0.4)
FC <sub>z</sub> -C <sub>z</sub>	1.6 (0.9)	0.6 (0.5)	0.2 (0.4)	0.8 (0.4)
C <sub>3</sub> -C <sub>1</sub>	1 (0.7)	2.0 (0.0)	2.0 (0.0)	0.0 (0.0)
C <sub>z</sub> -C <sub>2</sub>	3.8 (0.4)	0.0 (0.0)	0.0 (0.0)	0.6 (0.5)
F <sub>z</sub> -FC <sub>z</sub>	2.2 (1.3)	1.6 (0.5)	0.4 (0.5)	1.0 (0.7)

sets). The low standard deviation obtained for all cases shows the robustness of the proposed method over different runs of the algorithm.

### 3.5 Discussion and conclusions

Discrete wavelet transform (DWT) is a useful feature extraction tool since it explores the time as well as the frequency information of the signal. Although DWT has been employed to some degree of success in a number of synchronized BCI systems, there remain some limitations in its application to self-paced BCI systems (in terms of the large size of the feature space).

Brain computer interface systems that use DWT features have mostly employed only one or two channels (perhaps due to the large dimensionality of the feature space or to limitations imposed by the experimental protocol). To simultaneously explore the wavelet coefficients (features) of BCIs with more channels (so as to explore the spatial information) and to avoid the problems associated with the resultant large feature space, a two-stage (hybrid) feature selection algorithm is proposed. The first stage uses mutual information (MI) to discard the least informative features. In the second stage, a genetic algorithm (GA) selects those remaining features that lead to better system performance in the sense of meeting multiple objectives.

In our study, the features selected per channel varied considerably from one individual to another, as shown in Figure 3-2 to Figure 3-5. For example, for individual AB1, more features were selected from channels  $FC_1-C_1$ ,  $F_1-FC_1$ ,  $F_z-FC_z$ ,  $FC_4-C_4$ ,  $FC_2-FC_4$  and  $C_z-C_2$ , while for individual AB4, more features were selected from channels  $FC_4-C_4$ ,  $FC_2-FC_4$ ,  $F_1-F_z$ ,  $C_2-C_4$ ,  $F_1-FC_1$ , and  $FC_2-C_2$ . These results support the hypothesis that proper channel selection for every individual is necessary to obtain superior performance.

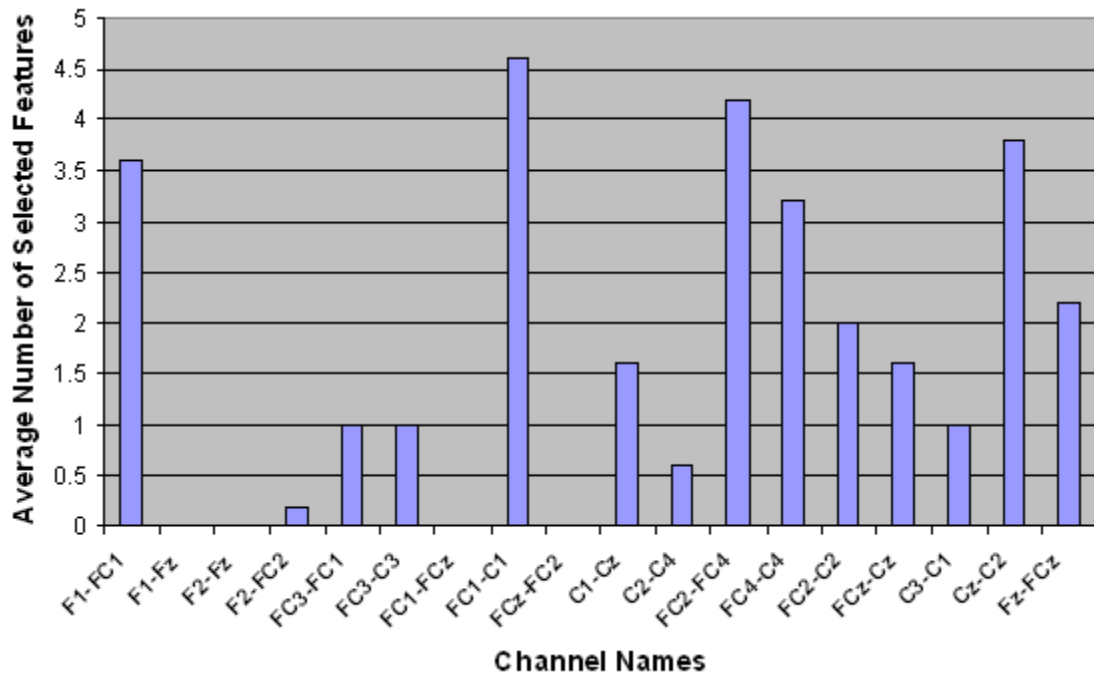


Figure 3-2. Spatial distribution of the average number of selected features for AB1.

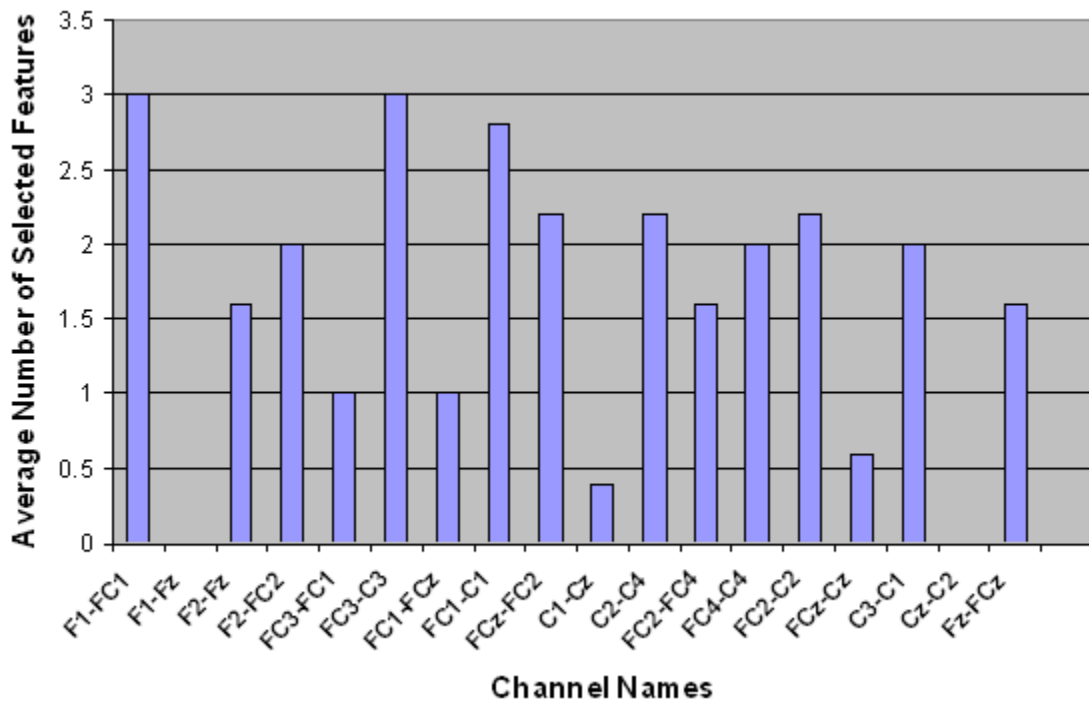


Figure 3-3. Spatial distribution of the average number of selected features for AB2.

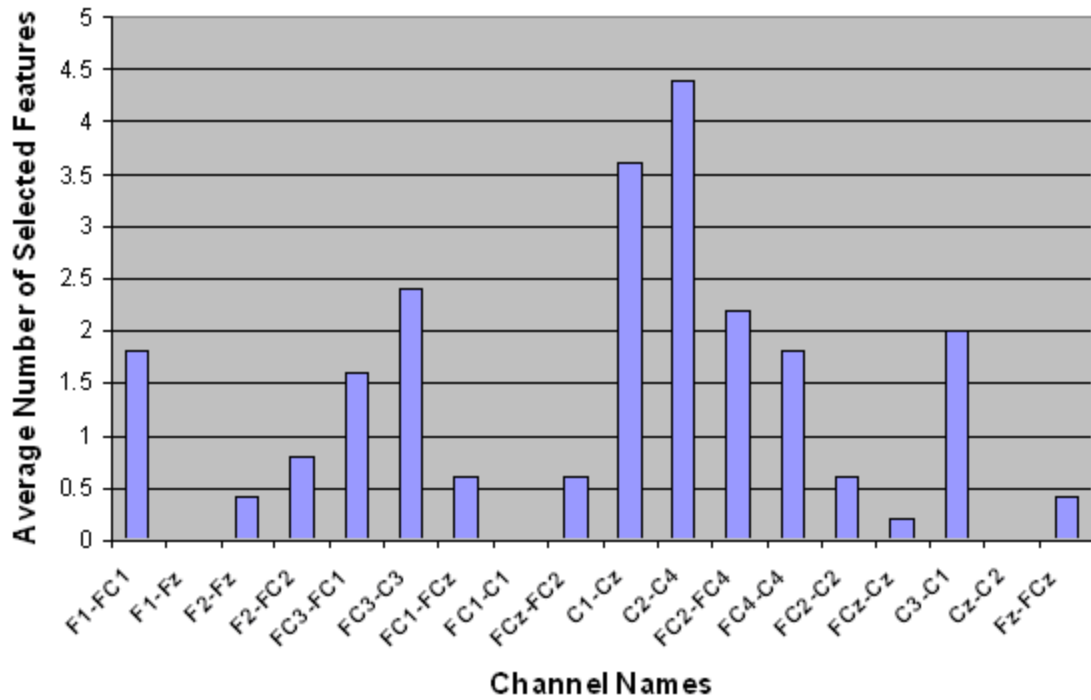


Figure 3-4. Spatial distribution of the average number of selected features for AB3.

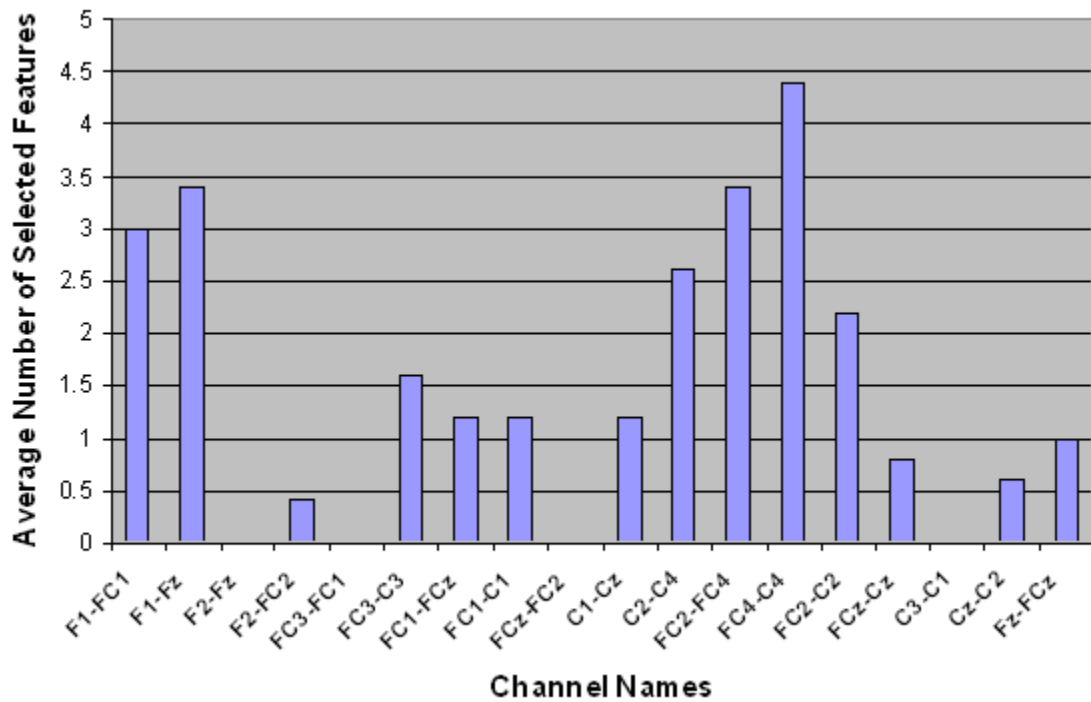


Figure 3-5. Spatial distribution of the average number of selected features for AB4.



Another finding from Figure 3-2 to Figure 3-5 is that the relevant features for each individual were unique. These findings are in contrast to an earlier study done by our group that empirically determined six pairs of electrodes for all individuals (channels  $F_1$ - $FC_1$ ,  $F_2$ - $FC_2$ ,  $FC_1$ - $C_1$ ,  $FC_2$ - $C_2$ ,  $FC_z$ - $C_z$ , and  $F_z$ - $FC_z$ ) [3]. Our findings in this regard are not surprising. The evidence from the literature supports the hypothesis that there is a significant amount of inter-subject variability in terms of generating MRP patterns [41]. The literature also shows that the selected features are not necessarily located in the standard frequency bands or on specific scalp locations, and that the set of selected features differs from individual to individual [42]. These studies support the notion that a customized BCI system should be designed for each individual.

Table 3-3 shows that for each individual, a number of bipolar channels were not selected by the feature selection process (such as channel  $F_1$ - $F_z$  for individuals AB1, AB2 and AB3, and channel  $FC_3$ - $FC_1$  for individual AB4). These results indicate that these channels can be eliminated from the analysis in future studies. Moreover, Table 3-3 and Figure 3-2 to Figure 3-5 show that the degree of contribution to the classification performance varies from one channel to another. These results indicate that a channel elimination methodology could be incorporated into the proposed method to further decrease the number of channels used for the operation of the system. Such an approach would rank the channels according to the number of selected features. It would then repeatedly eliminate the channel with the lowest contribution to fitness until the performance drops below a certain threshold (recursive elimination of channels). Systematic elimination of channels can lead to a faster setup of the system as well as decreased computational time. This could be part of future research works aimed at moving towards a more practical system.

It should be mentioned that it is difficult to directly compare the results of our study with other BCI studies. This is because the user population (whether or not individuals are able-bodied), the experimental protocols, the evaluation protocol and the neurological phenomenon differ from one study to another. In addition, the degree of training individuals receive before participating in a BCI experiment, vary among studies.

We can, however, compare our current results with the latest design of a state-of-the-art self-paced BCI system called the low frequency–asynchronous switch design (the LF-ASD) [40]. Both studies use the same individuals, the same experimental protocol, the same EEG data and similar evaluation protocol.

The LF-ASD (originally reported in [3] and later modified as reported in [40]) uses a feature extractor with a shape similar to a wavelet function, and extracts features from six bipolar EEG channels. The Karhunen-Loève Transform (KLT) is used to reduce the 6-dimensional feature space produced by the feature generator to a 2-dimensional space. A 1-NN classifier is used as the feature classifier. A moving average and a debounce algorithm are employed to improve the performance of the system by reducing the number of false activations. The parameter values of the system were estimated by an expert (for details, see [3, 30, 40]). The latest performance results of the LF-ASD [40], applied to the data of individuals AB1 to AB4 are presented in columns 6 to 9 of Table 3-2. As can be seen from the table, our proposed system has resulted in an increased  $\frac{TPR}{FPR}$  ratio for all individuals (with the exception of individual AB3). Specifically, the  $\frac{TPR}{FPR}$  ratio increased from 33.9 to 67.7 for individual AB1 (relative improvement of 99.5%), from 37.0 to 52.4 for individual AB2 (relative improvement of 41.6%), and from 36.5 to 39.8 for individual AB4 (relative improvement of 8.9%). These results show that our proposed approach improved the performance of most individuals compared with the latest design of the LF-ASD. The degree of improvements in the  $\frac{TPR}{FPR}$  ratio, however, is not statistically significant ( $p > 0.05$ ), so tests on the data of more individuals are needed to further substantiate this improvement. Note that the improved performance was achieved at the expense of using more features (please see columns 6 and 9 in Table 3-2).

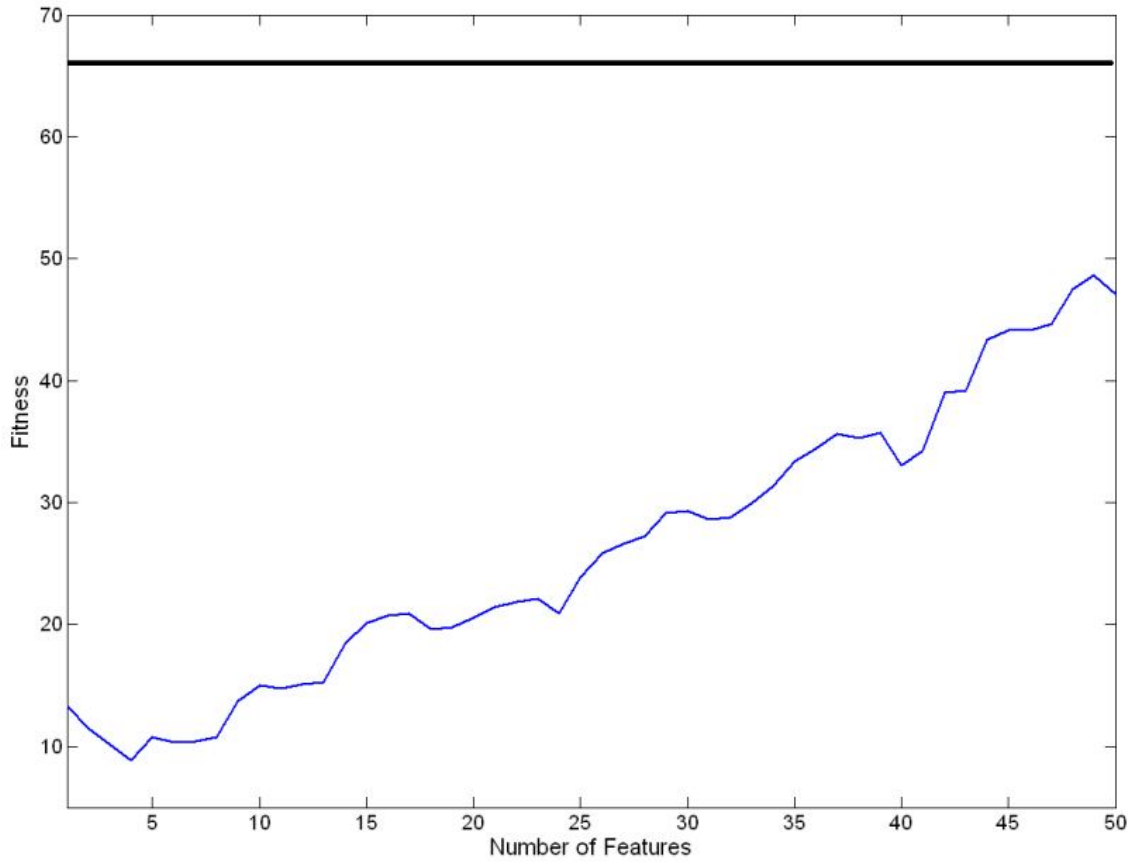
The relatively poor results obtained for individual AB3 may be partly related to our choice of wavelet function. Note that the wavelet function chosen for this study was based on the similarities between the chosen wavelet function and a typical bipolar MRP ensemble average pattern. However, there is substantial inter-subject variability in the shape of MRPs, especially in single trials [40]. It is expected that by analyzing a more

diverse family of wavelet functions, a different wavelet function might be chosen for each individual that would produce superior results.

As mentioned in Section 3.3, we designated the number of features chosen by the MI to be  $L=50$ . Fewer features would have sped up the process of feature selection at the second stage, but might have resulted in a lower fitness value. To test this possibility, we compared the fitness of the best subset of features (see Table 3-2) with that of all features for individual AB1 (see Figure 3-6). In this figure, the **thick** line shows the fitness of the best configuration (calculated from Table 3-2). The thin line shows the fitness of the classifier as a function of the number of top features. We began by training and testing the classifier using only the feature with the highest MI score, and then calculated the fitness. Then we added features one at a time (according to their MI scores) and trained and tested the classifier using the new set of features. This process was repeated until we reached  $L=50$ . Although the fitness of the classifier increased as more features were added, it stayed well below the optimal value achieved by the GA. These results indicate that a lower  $L$  (especially when only limited top features are used for training the classifier) does not necessarily lead to better performance.

A useful area to explore is the automation of the classifier. Currently, the feature selection procedure is automated but the selection of other parameters, such as those of the classifier, is carried out through cross-validation. Incorporating these parameters into the automation process would relieve the designer from the tiresome process of selecting the classifier's parameter values, while potentially yielding better classification results. Expanding the current results to continuous signals and ultimately online testing are also worthwhile topics for future work.

These results should be considered as preliminary results in the development of a self-paced brain computer interface system with a low FP rate. Our future work will also include testing of the proposed system on a larger pool of individuals to further investigate its usability.



**Figure 3-6. Comparison of the fitness of the best chromosome vs. other subset of features.**

### **3.6 Acknowledgements**

This work was supported in part by NSERC under Grant 90278-06 and CIHR under Grant MOP-72711. This research has been enabled by the use of WestGrid computing resources, which are funded in part by the Canada Foundation for Innovation, Alberta Innovation and Science, BC Advanced Education, and the participating research institutions. The authors would like to thank Mr. Craig Wilson for his valuable comments on this paper.

### 3.7 References

- [1] T. M. Vaughan, W. J. Heetderks, L. J. Trejo, W. Z. Rymer, M. Weinrich, M. M. Moore, A. Kubler, B. H. Dobkin, N. Birbaumer, E. Donchin, E. W. Wolpaw and J. R. Wolpaw, "Brain-computer interface technology: a review of the Second International Meeting," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no.2, pp. 94-109, Jun. 2003.
- [2] 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, no.6, pp. 767-791, Jun. 2002.
- [3] S. G. Mason and G. E. Birch, "A brain-controlled switch for asynchronous control applications," *IEEE Trans. Biomed. Eng.*, vol. 47, no.10, pp. 1297-1307, Oct. 2000.
- [4] S. G. Mason and G. E. Birch, "Temporal control paradigms for direct brain interfaces - rethinking the definition of asynchronous and synchronous," in *Proc. HCI Int. Conf.*, 2005, Las Vegas, USA.
- [5] G. E. Birch, P. D. Lawrence and R. D. Hare, "Single-trial processing of event-related potentials using outlier information," *IEEE Trans. Biomed. Eng.*, vol. 40, no.1, pp. 59-73, Jan. 1993.
- [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. Rehabil. Eng.*, vol. 8, no.2, pp. 180-185, Jun. 2000.
- [7] R. Millan Jdel and J. Mourino, "Asynchronous BCI and local neural classifiers: an overview of the Adaptive Brain Interface project," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no.2, pp. 159-161, Jun. 2003.
- [8] R. Scherer, G. R. Muller, C. Neuper, B. Graimann and G. Pfurtscheller, "An asynchronously controlled EEG-based virtual keyboard: improvement of the spelling rate," *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 979-984, Jun. 2004.
- [9] E. Yom-Tov and G. F. Inbar, "Detection of movement-related potentials from the electroencephalogram for possible use in a brain-computer interface," *Med. Biol. Eng. Comput.*, vol. 41, no.1, pp. 85-93, Jan. 2003.
- [10] G. Townsend, B. Graimann and G. Pfurtscheller, "Continuous EEG classification during motor imagery--simulation of an asynchronous BCI," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 12, no.2, pp. 258-265, Jun. 2004.
- [11] T. Demiralp, J. Yordanova, V. Kolev, A. Ademoglu, M. Devrim and V. J. Samar, "Time-frequency analysis of single-sweep event-related potentials by means of fast wavelet transform," *Brain Lang.*, vol. 66, no.1, pp. 129-145, Jan. 1999.
- [12] V. J. Samar, A. Bopardikar, R. Rao and K. Swartz, "Wavelet analysis of neuroelectric waveforms: a conceptual tutorial," *Brain Lang.*, vol. 66, no.1, pp. 7-60, Jan. 1999.
- [13] T. Hinterberger, A. Kubler, J. Kaiser, N. Neumann and N. Birbaumer, "A brain-computer interface (BCI) for the locked-in: comparison of different EEG classifications for the thought translation device," *Electroencephalogr. Clin. Neurophysiol.*, vol. 114, no.3, pp. 416-425, Mar. 2003.
- [14] B. Graimann, J. E. Huggins, S. P. Levine and G. Pfurtscheller, "Toward a direct brain interface based on human subdural recordings and wavelet-packet analysis," *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 954-962, Jun. 2004.

- [15] E. L. Glassman, "A wavelet-like filter based on neuron action potentials for analysis of human scalp electroencephalographs," *IEEE Trans. Biomed. Eng.*, vol. 52, no.11, pp. 1851-1862, Nov. 2005.
- [16] S. Fukuda, D. Tatsumi, H. Tsujimoto and S. Inokuchi, "Studies of input speed of word inputting system using event-related potential," in the Proc. 20<sup>th</sup> Annual Int. Conf. IEEE Engineering in Medicine and Biology Society, vol.3, pp. 1458-1460, 1998.
- [17] B. H. Jansen, A. Allam, P. Kota, K. Lachance, A. Osho and K. Sundaresan, "An exploratory study of factors affecting single trial P300 detection," *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 975-978, Jun. 2004.
- [18] C. H. Ding, "Unsupervised feature selection via two-way ordering in gene expression analysis," *Bioinformatics*, vol. 19, no.10, pp. 1259-1266, Jul 1. 2003.
- [19] L. E. -. Talavera, *An Evaluation of Filter and Wrapper Methods for Feature Selection in Categorical Clustering.*, vol. 3646, 2005, pp. 440-451.
- [20] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, pp. 273-324, 1997.
- [21] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Trans. Neural Networks*, vol. 5, no.4, pp. 537-550, 1994.
- [22] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison-Wesley Publishing Company, Reading, MA, 1989.
- [23] T. Back, D. B. Fogel and T. Michalewicz, *Evolutionary Computation.* Institute of Physics Publishing, Bristol and Philadelphia, 2000.
- [24] R. Beisteiner, P. Hollinger, G. Lindinger, W. Lang and A. Berthoz, "Mental representations of movements. Brain potentials associated with imagination of hand movements," *Electroencephalogr. Clin. Neurophysiol.*, vol. 96, no.2, pp. 183-193, Mar. 1995.
- [25] G. E. Chatrian, M. C. Petersen and J. A. Lazarte, "The blocking of the rolandic wicket rhythm and some central changes related to movement," *Electroencephalogr. Clin. Neurophysiol. Suppl.*, vol. 11, no.3, pp. 497-510, Aug. 1959.
- [26] G. Pfurtscheller and C. Neuper, "Motor imagery activates primary sensorimotor area in humans," *Neurosci. Lett.*, vol. 239, pp. 65-68, Dec 19. 1997.
- [27] G. Pfurtscheller, C. Neuper, D. Flotzinger and M. Pregenzer, "EEG-based discrimination between imagination of right and left hand movement," *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no.6, pp. 642-651, Dec. 1997.
- [28] C. A. Porro, M. P. Francescato, V. Cettolo, M. E. Diamond, P. Baraldi, C. Zuiani, M. Bazzocchi and P. E. di Prampero, "Primary motor and sensory cortex activation during motor performance and motor imagery: a functional magnetic resonance imaging study," *J. Neurosci.*, vol. 16, no.23, pp. 7688-7698, Dec. 1996.
- [29] R. Cunnington, R. Iansel, J. L. Bradshaw and J. G. Phillips, "Movement-related potentials associated with movement preparation and motor imagery," *Exp. Brain Res.*, vol. 111, no.3, pp. 429-436, Oct. 1996.
- [30] J. F. Borisoff, S. G. Mason, A. Bashashati and G. E. Birch, "Brain-computer interface design for asynchronous control applications: improvements to the LF-ASD asynchronous brain switch," *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 985-992, Jun. 2004.

- [31] G. E. Birch, Z. Bozorgzadeh and S. G. Mason, "Initial on-line evaluations of the LF-ASD brain-computer interface with able-bodied and spinal-cord subjects using imagined voluntary motor potentials," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 10, no.4, pp. 219-224, Dec. 2002.
- [32] C. Babiloni, F. Carducci, F. Cincotti, P. M. Rossini, C. Neuper, G. Pfurtscheller and F. Babiloni, "Human movement-related potentials vs desynchronization of EEG alpha rhythm: a high-resolution EEG study," *Neuroimage*, vol. 10, no.6, pp. 658-665, Dec. 1999.
- [33] S. G. Mallat, "Multifrequency channel decompositions of images and wavelet models," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 37, no.12, pp. 2091-2106, 1989.
- [34] N. Kwak and Chong-Ho Choi, "Input feature selection for classification problems," *IEEE Trans. Neural Networks*, vol. 13, no.1, pp. 143-159, 2002.
- [35] K. R. Muller, C. W. Anderson and G. E. Birch, "Linear and Nonlinear Methods for Brain-Computer Interfaces," *IEEE Trans. Neural Syst. and Rehab. Eng.*, vol. 11, no.2, pp. 165-169, Jun. 2003.
- [36] H. Yoon, K. Yang and C. Shahabi, "Feature subset selection and feature ranking for multivariate time series," *IEEE Trans Knowledge Data Eng.*, vol. 17, no.9, pp. 1186-1198, 2005.
- [37] C. Chang and C. Lin, *LIBSVM: A Library for Support Vector Machines*. 2001, s Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [38] M. Kaper, P. Meinicke, U. Grossekhoefer, T. Lingner and H. Ritter, "BCI Competition 2003--Data set IIB: support vector machines for the P300 speller paradigm," *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 1073-1076, Jun. 2004.
- [39] M. Kaper and H. Ritter, "Generalizing to new subjects in brain-computer interfacing," in Proc. 26<sup>th</sup> Annual Int. Conf. of Engineering in Medicine and Biology Society (EMBC'04), vo.6, , pp. 4363-4366, 2004.
- [40] A. Bashashati, M. Fatourechi, R. K. Ward and G. E. Birch, "User customization of the feature generator of an asynchronous brain interface," *Ann. Biomed. Eng.*, vol. 34, no.6, pp. 1051-1060, Jun. 2006.
- [41] V. G. Evidente, J. N. Caviness, B. Jamieson, A. Weaver and N. Joshi, "Intersubject variability and intrasubject reproducibility of the Bereitschaftspotential," *Mov. Disord.*, vol. 14, no.2, pp. 313-319, Mar. 1999.
- [42] J. Millan, M. Franze, J. Mourino, F. Cincotti and F. Babiloni, "Relevant EEG features for the classification of spontaneous motor-related tasks," *Biol. Cybern.*, vol. 86, no.2, pp. 89-95, Feb. 2002.

## CHAPTER 4 A SELF-PACED BRAIN COMPUTER INTERFACE SYSTEM THAT USES MOVEMENT RELATED POTENTIALS IN CHANGES IN THE POWER OF BRAIN RHYTHMS<sup>4</sup>

### 4.1 Introduction

In a brain computer interface (BCI) system, specific features of a person's brain signal relating to his/her intent are used to generate a control command that controls/actuates a device (see Figure 4-1 for a functional model of a BCI system).

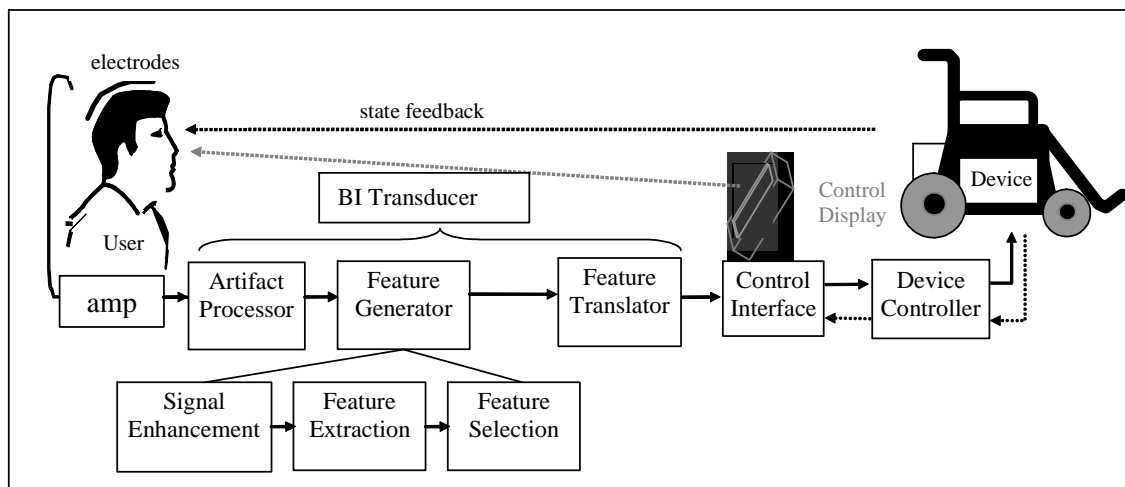


Figure 4-1. Functional model of a BCI system (adapted from [1]).

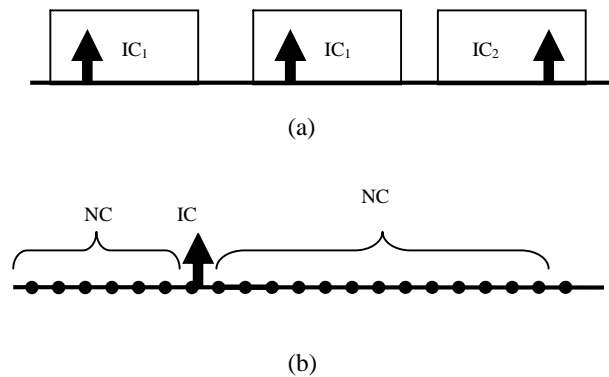
BCI designs are implemented in two ways: synchronized (system-paced) or asynchronous (self-paced). In synchronized BCI systems, a user can initiate a command

<sup>4</sup> A version of this chapter has been published. Fatourehchi, M., Birch, G. E., and Ward, R. K., "A Self-paced Brain Interface System that Uses Movement Related Potentials and Changes in the Power of Brain Rhythms", *Journal of Computational Neuroscience*, Vol.23, No.1, pp.21-37, Aug 2007.



only during certain periods specified by the system. It is assumed that a user only intends some control action during these specified times. Figure 4-2(a) shows a synchronized BCI system that can detect two intentional control (IC) commands ( $IC_1$  and  $IC_2$ ).

In a self-paced BCI (SBCI) system, users can affect the BCI transducer's output whenever they want, by intentionally changing their brain state. The state in which a user is intentionally attempting to control a BCI transducer is called an IC state. At other times, users are said to be in a no-control (NC) state, where they may be idle or performing some action other than trying to control the BCI transducer [2, 3]. To operate in this paradigm, BCI transducers are designed to respond only when a user is in an IC state and to remain inactive when a user is in an NC state. Figure 4-2(b) shows the output of an SBCI system.



**Figure 4-2. Synchronized vs. SBCI systems. (a) In a synchronized BCI system control is only possible during *System Ready* periods; (b) In an SBCI system, the system continuously accepts the input signals.**

So far, only a few BCI transducers (e.g., [2, 4-9]) have been specifically designed and tested for self-paced control applications. But as recognized in [10], the SBCI systems deserve more attention.

In this chapter, we focus on the issue of improving the performance of an SBCI system. A two-state SBCI system should be able to discriminate an IC command from an NC state. The performance of this system is usually evaluated through two metrics: a true positive (TP) rate and a false positive (FP) rate. An FP rate is the percentage of incorrectly classifying NC periods as IC periods, and a TP rate is the percentage of

correctly classifying IC periods. The FP rates of current SBCI systems are still very high for practical applications. The main reason is the very noisy nature of the brain's electrical signals, which makes correct detection of patterns associated with control commands very difficult. Nevertheless, it is crucial to keep the FP rate as low as possible in order to prevent user frustration.

We propose the use of multiple neurological phenomena as sources of control to improve the performance of an SBCI system (in terms of increasing the  $\frac{TPR}{FPR}$  ratio). The proposed SBCI system uses features extracted from three neurological phenomena: movement-related potentials (MRP), changes in the power of Mu rhythms (CPMR) and changes in the power of Beta rhythms (CPBR). The main rationale behind using these three neurological phenomena is that they are time-locked to movement onset [11, 12]. Thus, when a movement happens, it is expected that all three will be present. As a result, we postulate that a detector that considers the simultaneous occurrence of these phenomena should be more robust to the presence of transient non-control related changes in the brain signals (which may affect the performance of an SBCI system) than those that just look at one of the above-mentioned neurological phenomena.

Increasing the number of neurological phenomena considered from one to three has the disadvantage of increasing the dimension of the feature space. When there are restrictions on the number of sample sizes, the training data are likely to fall in a very small fraction of the sample space. This will limit the generalization property of the classifier [13]. To address this issue, we developed a new algorithm to reduce the dimensionality of the feature space. The proposed algorithm uses a two-stage multiple-classifier system (MCS) to classify the brain signals. We use the spatial information to develop the first stage of the MCS and the information from the neurological phenomena to develop the second stage of the MCS. The use of an MCS allows us to design a strong classifier by using an ensemble of “weak” classifiers. The evidence from the literature shows that in many cases this approach yields superior results over those of the best individual classifier [14, 15] and in some cases over those of a single powerful classifier [16-18]. For a BCI system, where the number of training IC patterns is usually limited, using an MCS provides us with a great opportunity to explore more features.

The proposed system showed superior performance to that of a simpler approach whereby all the features were combined into a single feature vector and then classified into IC or NC classes. Its performance was also found to be superior to that of an MCS that was based on the spatial information of a single neurological phenomenon.

To maximize performance of the proposed design, we employ a user-customized approach. It is well known that the spatiotemporal characteristics of a particular neurological phenomenon change from one individual to another [19-21]. As a result, customizing the BCI system for each user is very important in achieving consistently good results in all users. The improvements in the performance for users after employing user-customization (e.g., [7, 22-25]) further emphasize the importance of such customization.

In order to reduce the inter-subject variability of neurological phenomena in the proposed BCI design, a genetic algorithm was implemented to select the set of EEG channels that resulted in the best classification accuracy for each MCS. The results showed that the configuration of EEG channels leading to optimal performance varies from one user to another.

In the next section, we discuss the neurological phenomena under consideration and provide evidence from the literature that their combination would be useful in the design of an SBCI system.

## **4.2 Background**

### **4.2.1 Neurological phenomenon background**

It is known that an internally paced movement generates the following responses in the EEG signal: a movement-related potential (MRP), an event-related desynchronization (ERD) and an event-related synchronization (ERS). The MRP on the one hand and the ERD and ERS on the other are different responses of neural structures in the brain [26].

Averaging EEG data with respect to movement onset results in the generation of typical slow potentials, “movement-related potentials” (MRPs), from background

oscillatory electrical activity [27]. MRPs start about 1.5–1 seconds before the onset of a particular movement and have bilateral distribution [12, 27-30]. High-resolution EEG studies have modeled the main sources of MRPs arising in the supplementary motor area and the primary sensorimotor cortex [31, 32].

Voluntary movement results in a circumscribed desynchronization in the Mu and Beta bands, localized close to the sensorimotor areas ([33, 34]). This desynchronization, termed as “event-related desynchronization” (ERD), starts about 2 seconds prior to the onset of movement [26].

The enhanced rhythmic activity following the movement is called “event-related synchronization” (ERS). The post-movement Beta ERS is found in the first second after the termination of a voluntary movement, when the Mu rhythm still displays a desynchronized pattern [26]. The Beta ERS is a relatively robust phenomenon and is found in nearly all individuals after a finger, hand, arm or foot movement [35].

A number of papers provide some evidence that MRPs and changes in the power of brain rhythms (usually characterized as ERD and ERS) provide complementary information for exploration of the cognitive functions of the brain. It is suggested that MRPs can be considered as a series of transient post-synaptic responses of main pyramidal neurons triggered as a result of a specific event [26]. The same paper also states that the ERD and ERS phenomena can be viewed as being generated by changes in one or more parameters that control oscillations in neural networks. In [36], analysis of subdural EEG recordings from primary sensorimotor in epileptic patients showed that the amplitude of the ERD of the Alpha rhythm recorded from subdural areas was not always correlated with corresponding MRPs. It is suggested in the same paper that these neurological phenomena represent different aspects of cortical motor processes. In [37], the ERD of the Alpha rhythm is not always detected in cortical sites generating MRPs. In [12], through a high-resolution EEG study, it is shown that MRPs and the ERD of the Alpha rhythm provide complementary information on human brain responses accompanying the preparation and execution of a finger movement. Further evidence from the analysis of EEG signals [38, 39] and magnetoencephalography (MEG) [18, 40, 41] strengthens these findings.

There is also some evidence regarding the differences between Mu and Beta rhythms. Several papers show that the reactivities of the Mu and Beta rhythms related to the movement onset are different [20, 42]. Both the Mu and Beta rhythms desynchronize before a voluntary self-paced movement. However, after the movement, the ERD of the Mu rhythm is followed by a slow return to baseline (and sometimes by a slight synchronization), while the Beta rhythms synchronize rapidly after the movement onset [20].

#### **4.2.2 Multiple neurological phenomena in BCI systems**

Although most BCI researchers use a single neurological phenomenon as the source of control, there have been reports of using multiple neurological phenomena in BCI systems [11, 18, 43-46].

In [18], the authors analyzed combinations of features extracted from an early component of the MRP called Bereitschaftspotential (BP), features extracted from the ERD of neurological phenomena above 4Hz (through AR modeling) and features extracted from the common spatial patterns (CSP) features related to the ERD of the Mu rhythms. The BCI system had to discriminate between left and right index finger movements. A linear discriminant analysis (LDA) classifier was used for classification. Different combination schemes were explored. The study showed that a certain combination of classifiers could result in a lower error rate than the case where a single classifier is used. The results of combining the ERD of the Mu rhythm and the BP were not reported, although the authors mention that those results were slightly worse than the results obtained when all three neurological phenomena were used in the design of the BCI system. In [43], the authors applied a combination of microstate analysis and common spatial subspace decomposition to extract features belonging to three different frequency bands: Theta + Delta, Mu and Beta. The MRPs were not regarded as a separate neurological phenomenon. Instead, the features were extracted from the frequency band covering both the Delta and Theta rhythms. These features were then used to discriminate between left and right hand movements. In [47], the authors used the BP and the ERD of the brain rhythms from 10 to 33 Hz (including both the Mu and Beta rhythms) to classify left vs. right finger movement. The features extracted from all

neurological phenomena and all channels were then combined, the dimension of the feature vector was reduced and the final vector was classified using a perceptron neural network. The results showed classification accuracy of 84% on the test set, but the amount of contribution of each neurological phenomenon is not exactly known. In [48], the authors used features extracted from the BP and the ERD of the Mu rhythms for classifying the left and right index finger movements.

The above studies all pertain to synchronized BCI systems. To the best of our knowledge, only one SBCI system that uses multiple neurological phenomena has been reported so far [44]. In [44], the authors studied combining a number of neurological phenomena in order to design an ECoG-based SBCI system. Using a wavelet packet, the ECoG signal was divided into 18 different frequency bands covering a range from 0 - 100 Hz. This range covered a wide range of neurological phenomena including the Mu, Beta and Gamma rhythms, as well as other movement-related activities. Then for each band, wavelet-filtered signals were reconstructed. The wavelet filtered signals were then squared to achieve power values, and a genetic algorithm applied to reduce the dimension of the feature space to one. Using a thresholding classifier, the test samples were classified as movement or no movement.

Aside from the different signal processing approach, there are the following neurological phenomena-related differences between our proposed approach and that proposed in [44]:

- 1) While [44] focuses on the power of signals in a wide range of frequency bands, including the Theta, Beta, Mu and Gamma rhythms among others, our approach focuses on three specific neurological phenomena: MRPs, and changes in the power of the Mu and Beta rhythms.
- 2) In [44], the power in the Delta rhythms was used as one of the features. In this chapter, we intend to detect the shape of the MRP pattern in the ongoing EEG signal.
- 3) In [44], the contribution of each neurological phenomenon is not evident. Only the most significant features (those with the largest weight) were highlighted. In

our study, we specifically show which neurological phenomena are present in the design of the proposed user-customized SBCI system.

4) In [44], only the powers of signals at different frequency bands were used as features, while in this chapter, we are interested in detecting the time course of three distinct neurological phenomena.

### 4.3 Data collection

People with severe motor disabilities cannot physically execute a movement such as a finger flexion, but they are usually able to attempt a movement execution (by thinking that they are executing it). Several studies have shown that the EEG recordings obtained from attempted and real movements for able-bodied individuals bear many similarities [49-52]. These studies demonstrated that attempted and executed movements both result in the activation of similar cortical areas and generation of similar patterns. This evidence enables us to take the initial steps towards the development of an SBCI system, using the data recorded from able-bodied individuals. A similar rationale can be found in the design of other SBCI systems when the data of able-bodied individuals were employed [53, 54]. By using the data of able-bodied individuals, it is then possible to detect the occurrence (if any) of a control command by analyzing signals such as the electromyography (EMG) or the output of an actual switch. The signals can be used for labeling the brain signals and to evaluate the system's performance. The data analysis of individuals with motor disabilities was left to future studies.

The data of four able-bodied individuals (three males and one female) were used in this study. All individuals were right-handed and between 31 and 56 years old. They had all signed consent forms prior to participation in the experiment.

The individuals were positioned 150 cm in front of a computer monitor. The EEG signals were recorded from 13 monopolar electrodes positioned according to the International 10-20 System at  $F_1, F_z, F_2, FC_3, FC_1, FC_z, FC_2, FC_4, C_3, C_1, C_z, C_2$  and  $C_4$  locations. The cutoff frequency of the amplifier was set at 30Hz. Electrooculography (EOG) activity was measured as the potential difference between two electrodes, placed at the corner of and below the right eye. The ocular artifact was detected when the

difference between the EOG electrodes exceeded  $\pm 25 \mu\text{V}$ . All signals were sampled at 128 Hz and referenced to the ear electrodes (see [55] for details of the data recording). The recorded signals were then saved on a computer for further analysis.

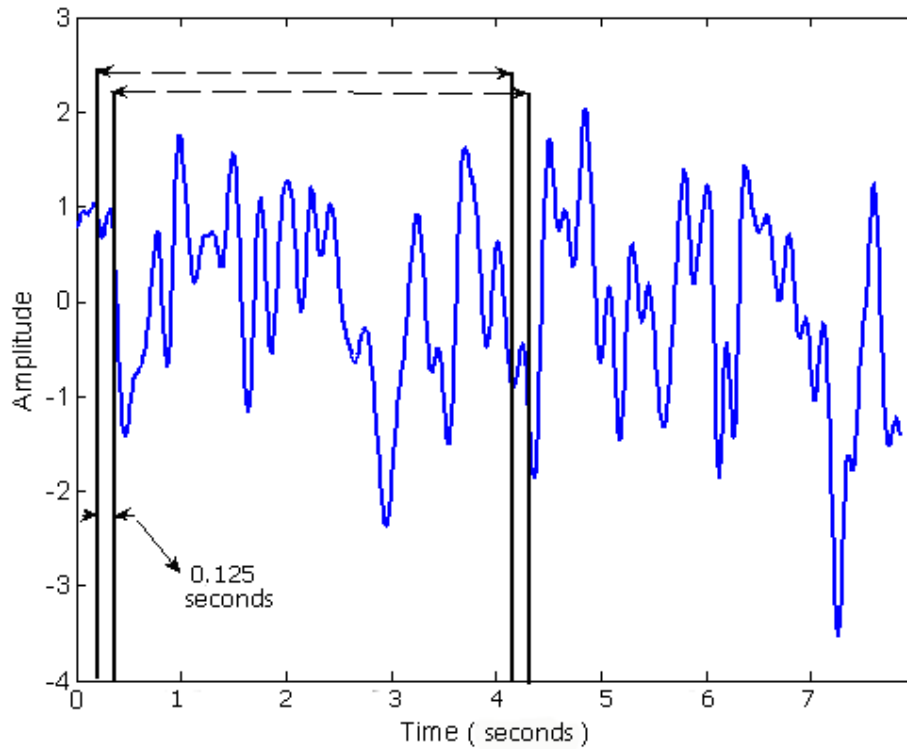
The individuals performed a guided task. At each interval, a white circle of 2cm diameter was displayed on the user's monitor for  $\frac{1}{4}$  second, prompting the users to attempt a movement. In response to this cue, the user had to perform a right index finger flexion one second after the cue appeared. The 1-second delay was used to avoid visual evoked potential (VEP) effects caused by the cue. This is the time that the user is expected to attempt the movement, but this time may vary from one user to another and from one movement attempt to another (see [56] for more details). For each individual, we used an average of 80 trials per day, over a period of 5 days.

An IC trial consisted of data collected from  $-t_{start}$  seconds before to  $t_{finish}$  seconds after the movement onset (measured as the finger switch activation), if no EOG artifact was detected in that particular interval. There were limitations in choosing the total length of  $(t_{start}+t_{finish})$ . As the length of  $(t_{start}+t_{finish})$  increases, IC periods have a higher probability of getting contaminated by the artifacts. As a result, the number of training trials that are not affected by artifacts will be reduced. On the other hand, if the length of  $(t_{start}+t_{finish})$  is too short, it results in a poor exploration of potential features. Since MRP and the changes in the power of Mu/Beta rhythms usually occur from approximately  $t_{start}=1$  before and last up to  $t_{finish}=2$  seconds after the movement, the data from 1 second before to 2 seconds after the activation of the finger switch was collected and marked as an IC trial.

NC sessions were recorded as follows: During NC sessions, the user was asked to count the number of times that a white ball bounced off the screen. As a result, the NC sessions contained attentive as well as non-attentive NC data. Each NC session lasted approximately three minutes and during each recording day, up to 2 of such NC sessions were recorded. The NC periods were then selected as follows: a window of width  $(t_{start}+t_{finish})$  seconds was considered. This window was slid over EEG signals collected during NC sessions by a step of 16 samples (0.1250 sec). For each window where artifacts were not detected, features were extracted (see Figure 4-3). Please note that



although the evidence from the literature suggests the Beta ERS occurs in the first second after the movement (which was confirmed in this chapter that by analyzing the averages of the of Beta rhythms over the trials in the training set),  $t_{finish}=2$  seconds was chosen because the return of the power of the Mu rhythm to the baseline activity usually takes a much longer time. In this chapter, we didn't study the optimal (or at least a sub-optimal) choice for  $t_{start}$  and  $t_{finish}$ . This research is left to future work.



**Figure 4-3.** NC periods are generated by shifting a window over NC datasets.

## 4.4 Methods

To analyze the information extracted from the different neurological phenomena, we carried out two separate studies. In the first study (“Study 1”), the features extracted from all the neurological phenomena were combined into a single feature vector and then the best subset of features (that led to optimal performance) was selected (see Figure 4-4). In the second study (“Study 2”), the classification was performed using a two-stage multiple classifier system (MCS).

In the first stage of the proposed system, each neurological phenomenon is classified separately. The dimension of each feature space for each classifier is then one-third of the original feature space. For each neurological phenomenon, an MCS that uses the spatial information obtained from the multi-channel EEG signals is designed. This approach further reduces the dimension of the feature space for each classifier in the MCS by a factor proportional to the number of EEG channels (as explained later in this section). The output of each MCS in the first stage, is a classification label for the input pattern that can be either an “IC” (coded as logical “1”) or an “NC” (coded as logical “0”). In the second stage, another MCS combines the outputs of the three MCSs designed separately for each neurological phenomenon and generates the final classification label for the input signal (see Figure 4-5). The details of the signal processing blocks will be explained in the rest of this section.

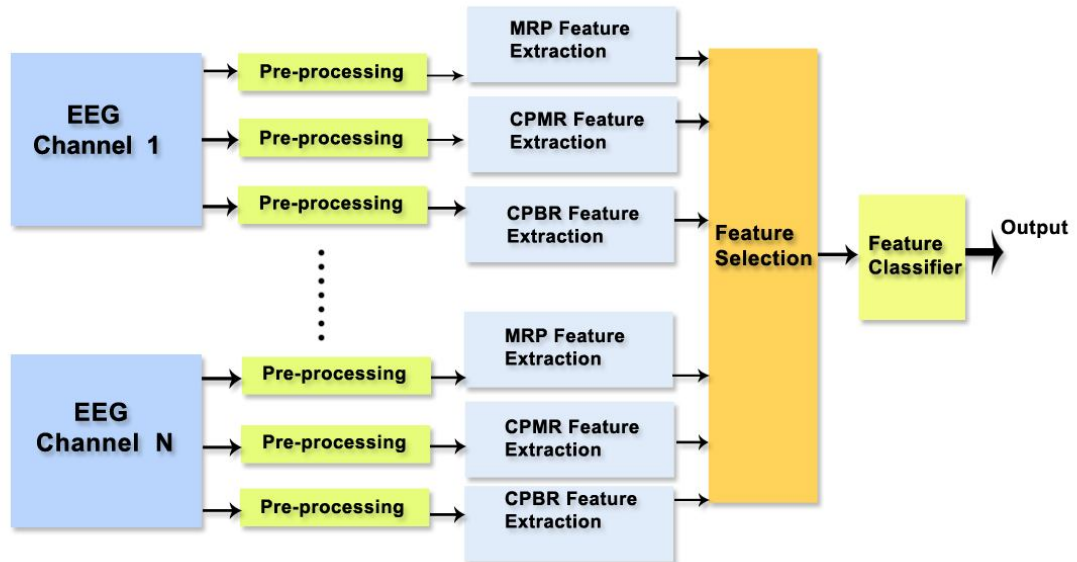


Figure 4-4. The overall structure of the SBCI system implemented in Study 1.

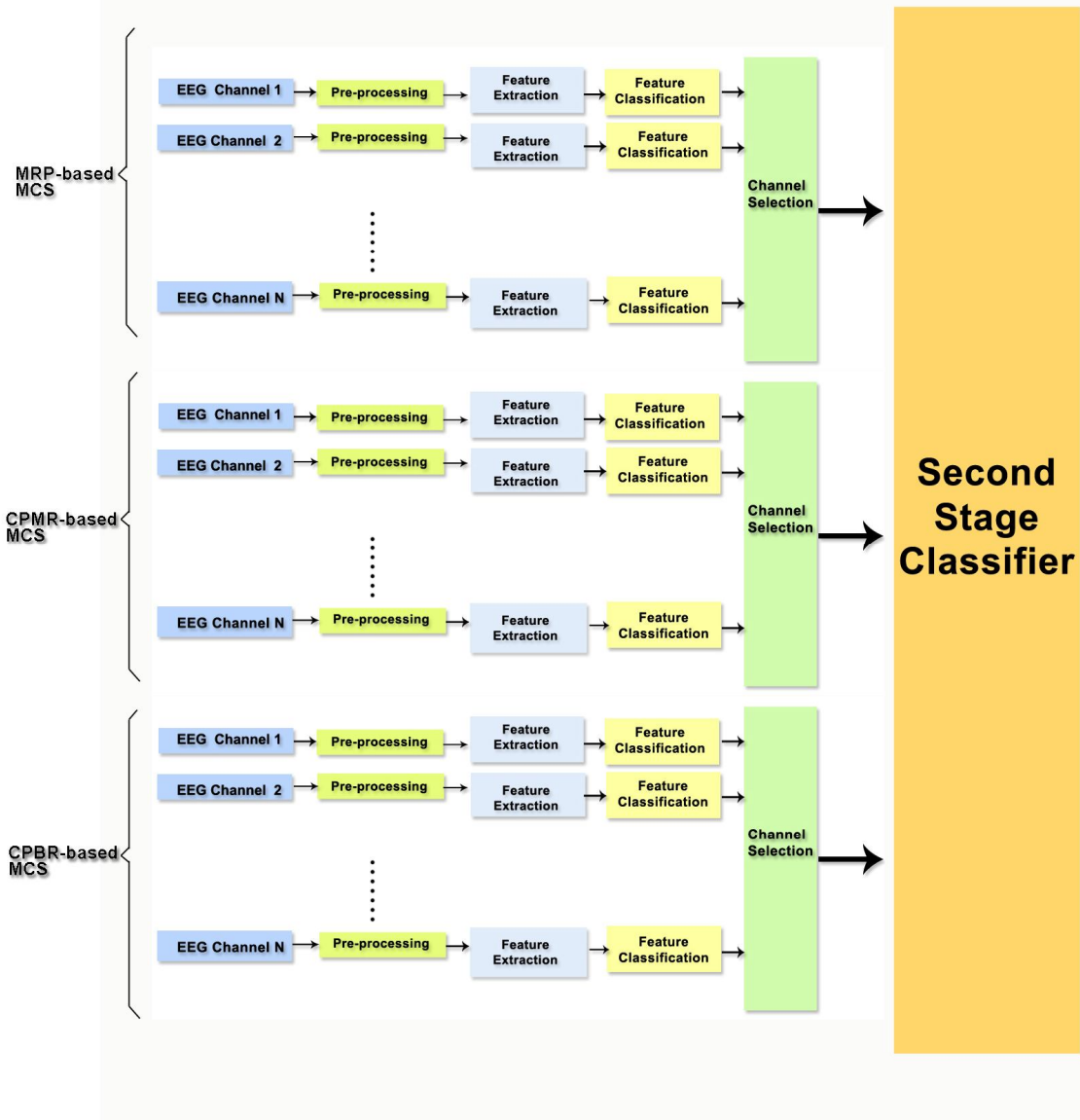


Figure 4-5. The overall structure of the two-stage MCS implemented in Study 2.

#### 4.4.1 Feature extraction

We have employed matched filtering to extract the features of the three neurological phenomena. It has been stated elsewhere that the best analyzing function for an event-related potential (ERP) is the one that matches that event as closely as possible in the temporal shape [57]. Matched filtering is known to be a simple yet useful tool for measuring the similarities between two sequences. A linear matched filter provides the maximum signal-to-noise power ratio at its output for a given template. In this chapter,

matched filtering is performed through cross-covariance. One of the sequences was the template signal (obtained by ensemble averaging of the trials in the training set, as will be explained later). The other sequence was a single-trial sequence.

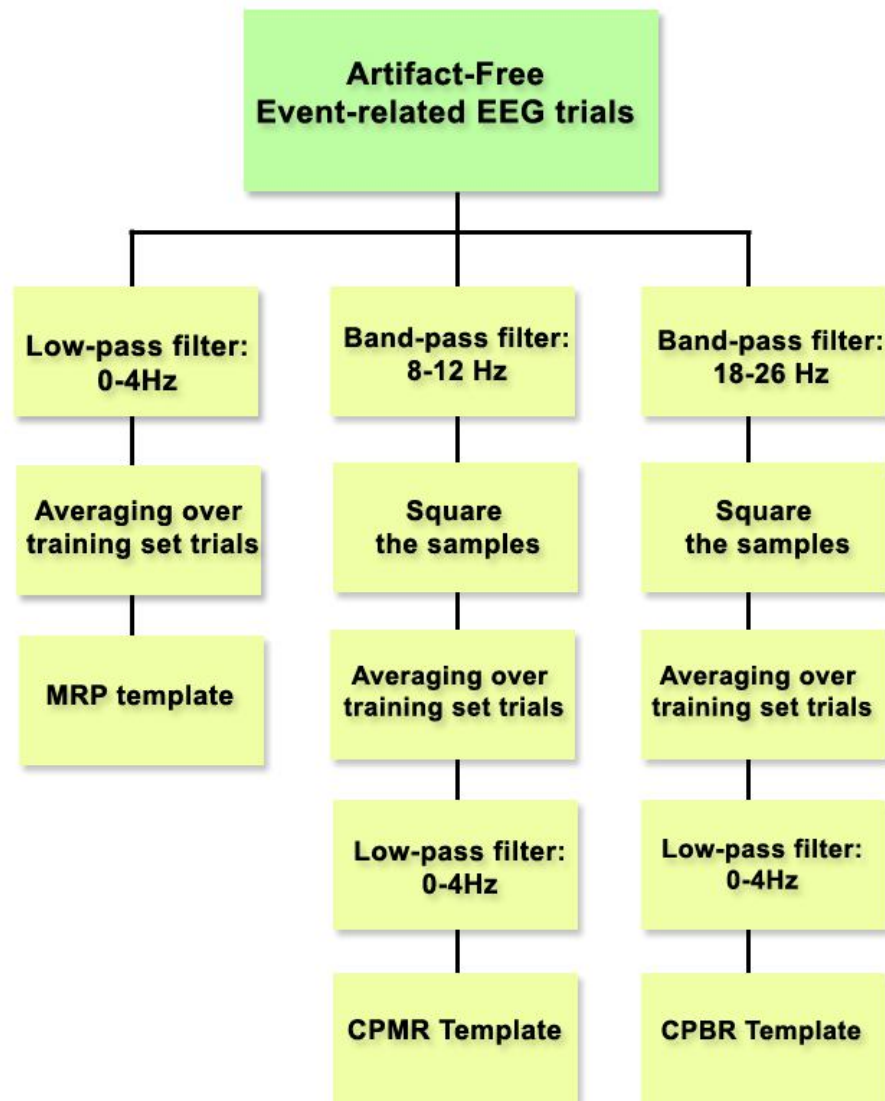
Our rationale for creating a template of the neurological phenomena based on the averages of the neurological phenomena of interest is similar to that in clinical studies. In clinical studies, averaging methods are widely used in the study of ERPs. The general assumption is that ERPs are time-locked to movement onset, while the ongoing EEG activity acts as additive noise. The averaging procedure thus enhances the signal-to-noise ratio (SNR) of the template. Below, we describe the process of creating the (see Figure 4-6).

#### **Creating MRP templates**

First, the EEG signals were filtered using a low-pass, linear-phase 32-point FIR filter with a 4 Hz cut-off frequency. The selection of 4Hz as the cutoff frequency was based on previous studies, which showed that MRPs have features mostly located in the frequency range below 4Hz [2]. We refer to this filter as *filterMRP*. Next, the intentional control (IC) periods were extracted (as explained in Section 4.3), and the IC periods in the training session were averaged. This process was repeated for all EEG channels.

#### **Creating CPMR and CPBR templates**

In the literature, a reference interval is defined some seconds before the occurrence of the movement in order to describe the ERD and the ERS. The ERD is then defined as the percentage of power decrease and the ERS is defined as the percentage of power increase with respect to the power of the signal in the reference interval [58]. However, it has been shown that such an approach may yield misleading results depending on the value of the reference signal's power [59]. In order to consider this referencing issue, we did not calculate the ERD or ERS templates as suggested in [26]



**Figure 4-6.** The process of generating templates.

(which are dependent on the value of the power signal in a reference interval). Instead we focused on the time courses of the power of the Mu and Beta rhythms themselves, as recommended and successfully applied in [59]. For the rest of this chapter, we use the terminology “changes in the power of the Mu rhythms” (CPMR) and “changes in the power of the Beta rhythms” (CPBR) to refer to these neurological phenomena.

The CPMR and CPBR templates were calculated as follows. All signals were band-pass filtered. For the Mu rhythms, the band-pass was chosen from 8 to 12Hz. For the Beta rhythms, the band-pass was chosen from 18 to 26Hz . Both filters were linear-phase 32-point FIR filters. The amplitudes of the samples were then squared to obtain the power samples. The resulting power signals were then low-pass filtered with a linear-phase 32-point FIR filter called *filterPower*. The cutoff frequency of the filter was selected as 4 Hz. This filtering stage ensured that a smooth shape was obtained for the power templates. The power templates in the training sets were then averaged over the available trials. The power at time instant  $j$  was calculated as follows:

$$P_j = \frac{1}{L} \sum_{i=1}^L (x_{ij})^2 \quad (4-1)$$

where  $P_j$  is the  $j$ -th power sample of the template,  $L$  is the total number of IC periods in the training set and  $x_{ij}$  is the  $j$ -th sample of the  $i$ -th trial of the band-pass filtered data. This process was repeated for all EEG channels.

### Cross-covariance

The cross-covariance between two stationary sequences is calculated as follows:

$$C_{xy}(n) = E((X_m - \mu_x)(Y_{m+n} - \mu_y)^*) \quad (4-2)$$

where  $X_n$  and  $Y_n$  are stationary random processes and  $E$  is the expected value operator. In our analysis, one of the sequences is the IC template for a particular channel and the other sequence is a single-trial sequence. After calculating  $C_{xy}(n)$  for each trial, the following was extracted:

$$F_i = \max(C_{xy}(n)), \quad n \in [t_{start} + t_{finish} - 0.0625, t_{start} + t_{finish} + 0.0625] \quad (4-3)$$

When applying “matched filtering” between a template and a test sequence, the two sequences must be aligned. If the two sequences are not aligned, there is a possibility that the particular movement pattern is not detected. This issue was resolved in (4-3) by calculating the maximum of the cross-correlogram over a window of length 16 samples ( $\frac{1}{8}$ <sup>th</sup> of a second) instead of calculating it using a single epoch. Please note that in order

to generate the NC periods, a window was slid over the NC datasets by the amount of shift, which was  $\frac{1}{8}$ <sup>th</sup> of a second .

#### 4.4.2 Feature classifier

##### K-nearest neighbor classifier

We used a K-nearest neighbor (KNN) classifier for classification. The LVQ3 learning algorithm [60] was used to create the codebook vectors for each class. Prior to applying the LVQ3 learning algorithm, K-means was used for forming the initial clusters.

Since the number of generated NC features was much larger than the number of IC features, not all NC features were used. The number of NC features was reduced to the same number as that of IC features by randomly choosing from the NC periods available in each training set. However, the number of NC features in the validation or test sets was not reduced.

Prior to classification, each feature  $f_i$  was normalized according to the following equations:

$$\begin{aligned}
 f_{iAVERAGE} &= \frac{1}{P} \cdot \sum_{j=1}^P f_{ij} \\
 f_{iSTD} &= \sqrt{\frac{1}{(P-1)} \cdot \sum_{i=1}^P (f_{ij} - f_{iAVERAGE})^2} \\
 f_{ijNormalized} &= \frac{f_{ij} - f_{iAVERAGE}}{f_{iSTD}}
 \end{aligned} \tag{4-4}$$

where  $f_{ij}$  is the  $j$ -th value of feature  $i$ ,  $f_{iAVERAGE}$  is the average of feature  $i$ ,  $f_{iSTD}$  is the standard deviation of feature  $i$ ,  $f_{ijNormalized}$  is the normalized value of  $f_{ij}$  and  $P$  is the total number of NC and IC values available in the training set for feature  $f_i$ .

### **Multiple classifier systems**

As mentioned in the beginning of this section, in the second study a two-stage MCS that combines the information extracted from different neurological phenomena was used. Although an MCS seems to be an efficient tool to classify the spatial information in BCI systems, its application so far has been limited. In [61], an MCS is used to decrease the bias of the neural network classifiers with respect to the initial conditions. In [62], the individual EEG channels were ranked according to their classification accuracy. Starting with the channel with the best accuracy, channels were added to the MCS until no improvement in accuracy was achieved. In [63], the authors used different combination strategies to combine the information in different channels in an ECoG-based SBCI system. No channel selection was performed.

As will be explained later, in our study, we propose a different approach that not only chooses classifiers with higher accuracies, but also the ones that are diverse. We use both the spatial information of channels and the presence of different neurological phenomena as the sources of control to create a diverse set of classifiers.

In Study 2, each neurological phenomenon is first classified separately. For each neurological phenomenon, a K-NN classifier (see Section 4.4.2) is designed for each EEG channel. Then the outputs of those classifiers are combined using an MCS to generate the output label corresponding to that neurological phenomenon. A genetic algorithm (GA) is used to select the best combination of channels for each neurological phenomenon. The selected channels then participate in a majority voting scheme to generate the final classification label for that particular neurological phenomenon. Once an individual MCS is designed for each neurological phenomenon, the outputs of MCSs (the classification labels) are combined using a second MCS, as shown in Figure 4-5.

There are various methods to combine the outputs of these expert systems. To determine the proper output combination scheme, one needs to make a compromise between the number of true positives and the number of false positives. Since having a low FP rate is very important in the design of SBCI systems, we considered having a very low probability of misclassifying NC periods at the expense of having a higher



probability of missing IC commands. This rationale has formed the overall structure of the second-stage MCS shown in Figure 4-5.

We explored the following configurations for this classifier:

1. Configuration 1- Combined MRP and CPBR -based classifiers: In this configuration, only the outputs of MCSs using MRP and CPBR are combined. The output of the second stage MCS is then calculated as the product of the outputs of each of these classifiers:

$$O_{Final}[n] = O_{MCS1}[n] * O_{MCS2}[n] \quad (4-5)$$

where  $O_{Final}[n]$  is the output of the second-stage MCS at the time instant n,  $O_{MCS1}[n]$  is the output of the first MCS at the time instant n and  $O_{MCS2}[n]$  is the output of the second MCS at the time instant n. Using the product rule, we can ensure that the output of the second-stage MCS is activated only when both the MRP-based and CPBR-based MCSs detect the presence of movement in the pattern.

2. Configuration 2- Combined MRP and CPMR -based classifiers: In this configuration, only the outputs of MCSs based on MRP and CPMR are combined, using (4-5).

3. Configuration 3- Combined CPMR and CPBR -based classifiers: In this configuration, only the outputs of MCSs based on CPMR and CPBR are combined, using (4-5).

4. Configuration 4- Combined outputs of all classifiers according to the majority voting scheme: In this scheme, the outputs of all three classifiers are combined according to the following equation:

$$O_{Final}[n] = \begin{cases} 1, & \text{if } \sum_{i=1}^3 MCS_i[n] > 1 \\ 0, & \text{if } \sum_{i=1}^3 MCS_i[n] \leq 1 \end{cases} \quad (4-6)$$

where  $O_{Final}[n]$  is the output of the second-stage MCS at time instant n and  $O_{MCSi}[n]$  is the output of the ith MCS at that time instant.

5. Configuration 5- Combined outputs of all classifiers according to the product scheme: In this scheme, the outputs of all three classifiers are combined according to the following equation.

$$O_{Meta}[n] = \prod_{i=1}^3 (O_{MCSi}[n]) \quad (4-7)$$

where  $O_{Final}[n]$  is the output of the second-stage MCS at time instant  $n$ ,  $O_{MCSi}[n]$  is the output of the  $i$ th MCS at the same time instant and  $\prod$  is the product operator.

#### 4.4.3 Feature selection

In both studies, a genetic algorithm (GA) is used for feature selection. GAs are heuristic methods that provide a framework for effectively sampling large search spaces [64]. They are based upon the genetic processes of biological organisms, which evolve over many generations according to the principles of natural selection and survival of the fittest. By mimicking this process, GAs are able to evolve solutions to real-world problems and have been shown to be useful tools in automatically customizing many practical systems [64, 65].

In applying GA's to select features, the features of interest are first coded in the form of a set of randomly generated binary numbers (that can be either "0" or "1", with a "1" indicating the presence of a feature and a "0" indicating the absence of a feature in the binary string. Each set of the binary strings is called a "chromosome" and the set of "chromosomes" forms a "population". The classifier then classifies the features specified by the chromosome. The classification accuracy of the classifier is then used to provide the fitness value of that particular chromosome. The chromosomes are then combined using operators such as selection, crossover and mutation in order to generate new chromosomes. This recombination process results in the production of a new generation. It is expected that the population evolves gradually and the fitness improves over generations. This process is continued until some criteria for stopping the GA is met [64].

In Study 1, a GA was used to select a subset of  $q$  features from the total available  $3N$  features (where  $N$  is the number of EEG channels). Then a K-NN classifier, as

described in sub-section 4.4.2, was trained using the selected feature vector. In Study 2, for each neurological phenomenon, a K-NN classifier was designed for each EEG channel. For each neurological phenomenon, a GA was then used to select amongst all classifiers (each classifier corresponds to one EEG channel) those that when combined gave the best performance using a majority voting scheme. The outputs of the MCSs developed for the three neurological phenomena were then combined through the method described in sub-section 4.4.2.

To develop a suitable fitness function for the GA, the criteria considered were maximizing the TP rate, minimizing the FP rate and minimizing the complexity of the system (measured here by the number of features). Four objective functions were defined so as to achieve these goals. The first objective function was defined to combine the TP and FP rates:

$$f(x) = \begin{cases} 0, & TPR < 20\% \\ \frac{TPR(x)}{FPR(x)}, & otherwise \end{cases} \quad (4-8)$$

In (4-8),  $x$  is a chromosome and  $f$  is the objective function. This objective function assigns a higher rank to chromosomes that yield higher TP rates and lower FP rates. We also postulated that TP rates below 20% (which can be translated to the correct detection less than one out of five control attempts) would be too low for the successful operation of an SBCI system (even though their FP rates might also be very low), so they were considered “unfit”. This objective function was selected as the highest-priority objective function in the optimization of the system.

The second objective was the average of the FP rate over the validation sets. The third objective was the average of the TP rate over the validation sets and the fourth objective was the number of features, respectively, resulting in four objectives per chromosome.

A lexicographic approach was then used for optimizing the multi-objective fitness function [65]. The objectives were ranked according to their priorities. The average of  $\frac{TPR}{FPR}$  over the validation sets was selected as the objective function with the highest

priority. The chromosomes were then ranked in a single-objective fashion. Any ties were resolved by comparing the relevant chromosomes again with respect to the second, third and fourth objectives. The orders of the second and 3<sup>rd</sup> objective were chosen such that for two chromosomes with the same  $\frac{TPR}{FPR}$  ratio (the same value for the first objective function), the chromosome with the lower FP rate is considered more fit than the other chromosome.

The operators of the GA were tournament-based selection (tournament size =3) , uniform crossover and uniform mutation. The size of the initial population and of the population in the next generations was chosen as 100 and 50, respectively. We used random initialization for initializing the GA. The best chromosome of each population was retained in the subsequent population.

For Study 1, the number of evaluations was set to 5000. For Study 2 the number of evaluations was set to 1500, since for each individual neurological phenomenon we were dealing with a smaller feature space compared with that in “Study 1”. If the improvement in the best solution was found to be less than 1% for more than 10 consecutive generations, the algorithm was terminated.

#### **4.4.4 Performance evaluation**

The IC and NC datasets were each randomized and divided into training, validation and test sets. The training set was used for training the classifiers and the validation set was used to select those channels whose combination led to the best performance of the system. The configuration that yielded the best average performance on the validation set was selected and using the test sets, the results of classification were reported. We used a 5-fold nested cross-validation for evaluating the performance of the system, since in self-paced BI designs the number of labeled IC epochs are limited compared to the complexity of the data. It has been argued that using a fixed split of data into training, validation and test sets is not robust. It is thus recommended that when sufficient computing resources are available, a nested cross-validation should be performed [66]. The inner cross-validation is used for model selection (feature selection, in our case) and the outer cross-validation is used to estimate the generalization error.

For each outer cross-validation set (test set), 20% of the data were used as the test set and the rest were used for training and model validation. In order to select the models, the datasets were further divided into five folds themselves. For each fold, 80% of the data were used for generating a template to be used for cross-correlation, and 20% were used for model validation.

The value of  $K$  in the  $K$ -NN classifiers, the number of codebooks and the parameters' values for the LVQ learning algorithm were all determined through parallel runs of the algorithm on a cluster of computers for different parameters' values.

## 4.5 Results

The results of Study 1 (averaged over 5 outer cross-validation sets) are presented in Table 4-1. The numbers in parenthesis show the standard deviation. The method employed in Study 1 resulted in an average TP rate of 60.91% and an average FP rate of 4.63% for the four individuals. Although the average TP rate can be considered as acceptable (corresponding to the average detection of 6 out of every 10 IC commands), the average FP rate is too high for most practical applications since it results in a false activation in almost every three seconds. This is because as the system is designed so that it generates an output every  $\frac{1}{8}$  of a second (see sub-section 4.4.1).

The findings of Study 2 are presented in Table 4-2 to Table 4-6. In Table 4-2, the system's performance on the validation sets for each neurological phenomenon is shown. For all individuals, the average FP rates were simply too high for any practical application. Although the MRP-based classifier was able to detect more than half of the IC commands, the CPMR-based and CPBR-based classifiers were successful in detecting one out of four IC commands.

**Table 4-1. The average TP and FP rates (%) for Study 1 (the numbers in parenthesis show the standard deviation).**

Individual	Test		
	<i>TPR</i>	<i>FPR</i>	$\frac{TPR}{FPR}$
<b>AB1</b>	47.0 (12.0)	1.6 (1.3)	28.5
<b>AB2</b>	60.9 (10.9)	2.8 (0.4)	21.8
<b>AB3</b>	80.9 (7.0)	11.5 (4.5)	7.0
<b>AB4</b>	54.9 (11.7)	2.6 (0.5)	21.0
<b>Average</b>	60.9	4.6	13.2

**Table 4-2. The average TP and FP rates (%) for each neurological phenomenon in Study 2 (the numbers in parenthesis show the standard deviation).**

Individual	Validation - MRP			Validation- CPMR			Validation- CPBR		
	<i>TPR</i>	<i>FPR</i>	$\frac{TPR}{FPR}$	<i>TPR</i>	<i>FPR</i>	$\frac{TPR}{FPR}$	<i>TPR</i>	<i>FPR</i>	$\frac{TPR}{FPR}$
<b>AB1</b>	54.9 (1.8)	3.0 (0.1)	18.3	21.2 (1.1)	2.6 (0.5)	8.11	24.0 (4.3)	1.1 (0.5)	22.3
<b>AB2</b>	49.8 (1.1)	4.6 (0.7)	10.8	21.6 (1.6)	4.3 (0.9)	5.1	31.2 (6.3)	2.6 (0.9)	11.9
<b>AB3</b>	46.2 (6.6)	5.9 (1.3)	7.8	22.4 (3.1)	8.0 (1.5)	2.8	23.4 (3.9)	2.4 (0.6)	9.6
<b>AB4</b>	59.9 (5.6)	6.9 (1.1)	8.7	25.1 (6.6)	2.7 (0.9)	9.2	27.1 (2.4)	5.2 (5.6)	5.2
<b>Average</b>	52.7	5.1	10.3	22.6	4.4	5.1	26.4	2.8	5.5

Table 4-3 to Table 4-6 show the average of the TP and FP rates and their ratios over the inner validation sets (called Validation in the tables) and the outer validation sets (called Test in the tables) for all five configurations described in Section 4.4.2. The numbers in parenthesis are the standard deviations calculated over 5 outer validation sets.

The classification accuracy obtained using the validation sets are also reported. This is because the validation set determines the choice of the best configuration. For each individual, the configuration highlighted in ***BOLD&ITALICS*** was the one that yielded the highest average of the  $\frac{TPR}{FPR}$  rates and also, it had an average of  $TPR > 20\%$  over the “inner validation sets”.

As seen from Table 4-3 to Table 4-6, the configuration that resulted in a superior performance for each individual was subject-dependent, i.e., not always the same as those of the other individuals. Table 4-3 shows that for user AB1, the combination of all three neurological phenomena through majority voting (Configuration 4) resulted in the highest  $\frac{TPR}{FPR}$  rate (with average  $TPR > 20\%$ ). Although the TP rate showed an average detection of one out of four IC commands, the average FP rate was 0.13%, corresponding to a true negative rate of 99.9%. This FP value is 15.4 times smaller than the FP value (for the same user) achieved in a previous study by our group [55]. However, a direct comparison is not possible because of the difference in the experimental designs.

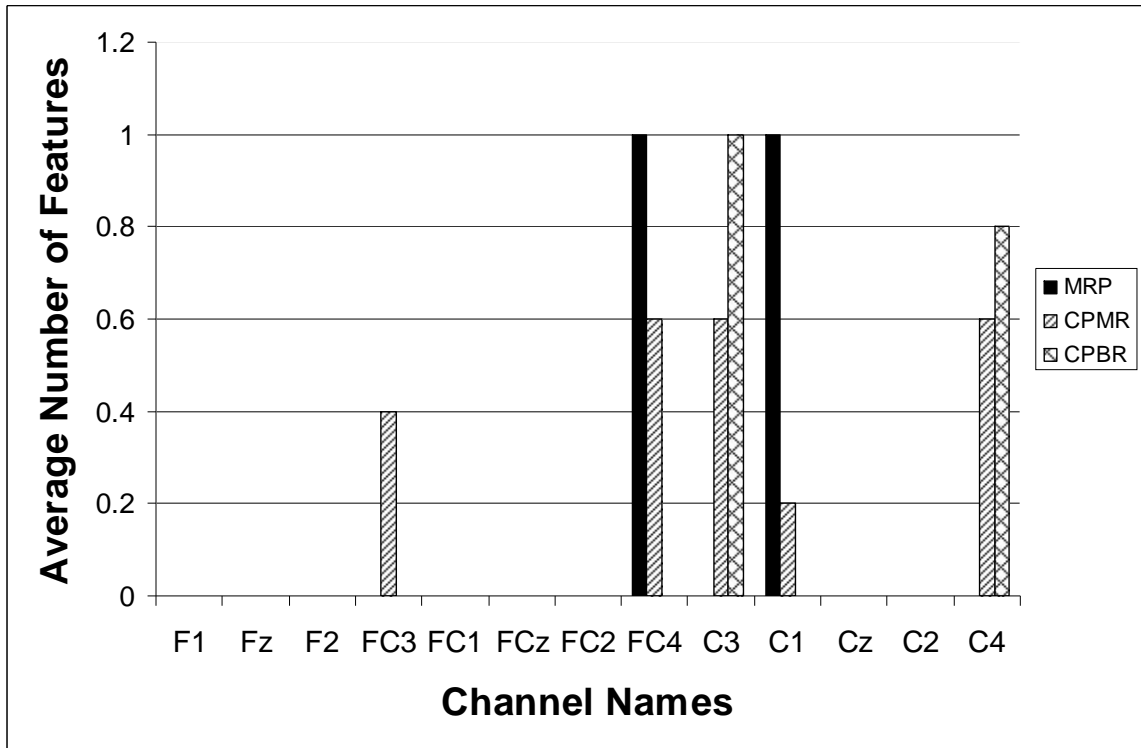
For users AB2 and AB3, Configuration 1 (the combination of MRP-based and CPBR-based classifiers) resulted in superior performance. More specifically, for AB2, an FP rate of 0.3% (corresponding to  $TN = 99.7\%$ ) was achieved for a TP rate of 25.5% (see Table 4-4). As seen in Table 4-5, for user AB3, an FP rate of 1.1% (corresponding to  $TN = 98.9\%$ ) was achieved at  $TP = 36.7\%$  (corresponding to the detection of one out of three IC commands).

Table 4-6 shows that for user AB4, the combination of the MRP-based and CPMR-based classifiers (Configuration 2) resulted in superior performance ( $TP = 21.4\%$  and  $TN = 99.6\%$ ). However, the achieved performance was very close to that of the combination of MRP-based and CPBR-based classifiers, with  $TP = 24.7\%$  and  $TN = 99.5\%$ .

For all individuals, the combination of CPMR-based and CPBR-based classifiers (Configuration 3) resulted in poor  $\frac{TPR}{FPR}$  ratios. This finding suggests that in this study,

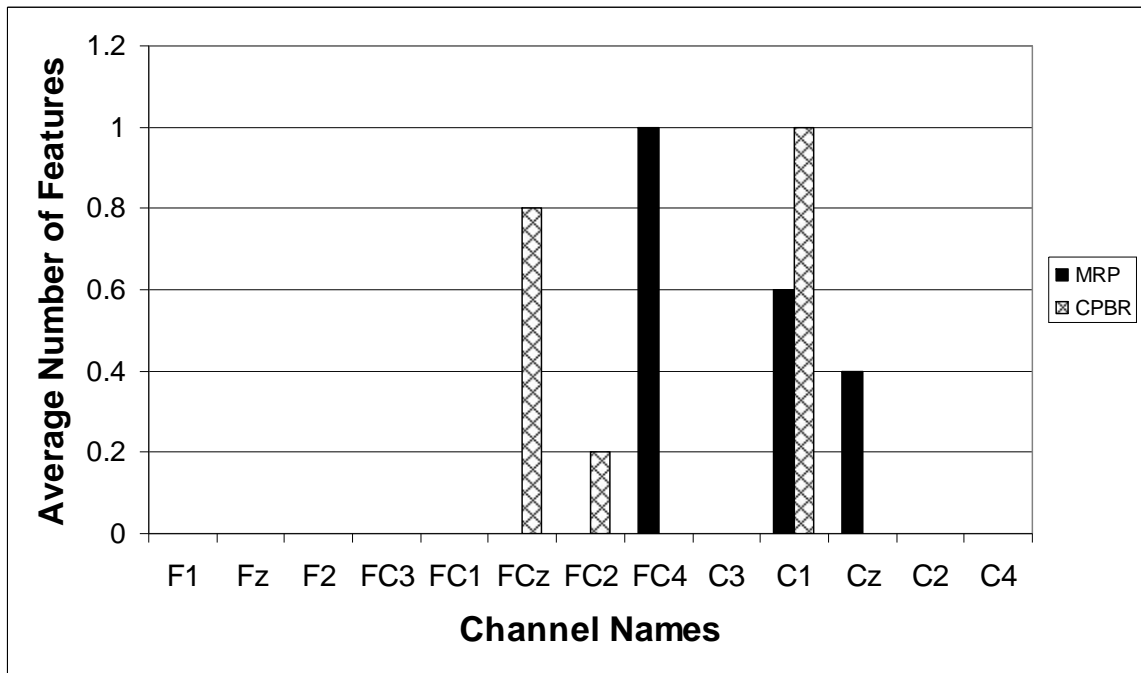
the presence of an MRP was necessary for the relatively successful operation of the BCI system.

Figure 4-7 shows the average number of selected features per channel for each individual. As can be seen, there is a great amount of inter-subject variability in terms of EEG channels selected in the best configuration.

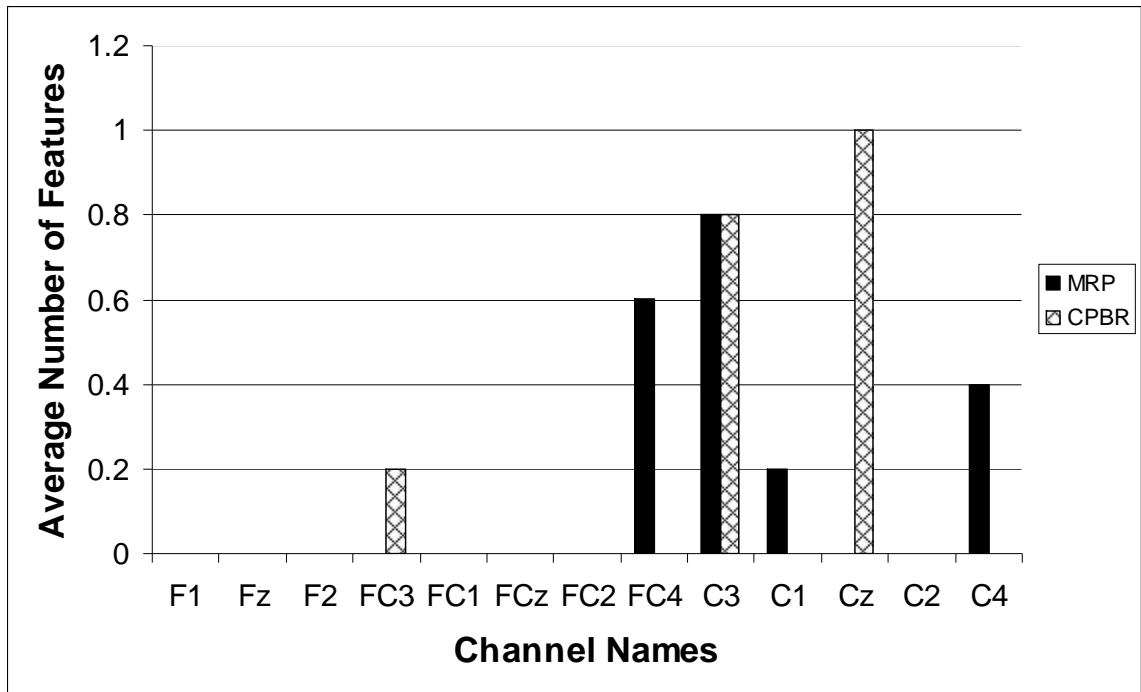


(a)

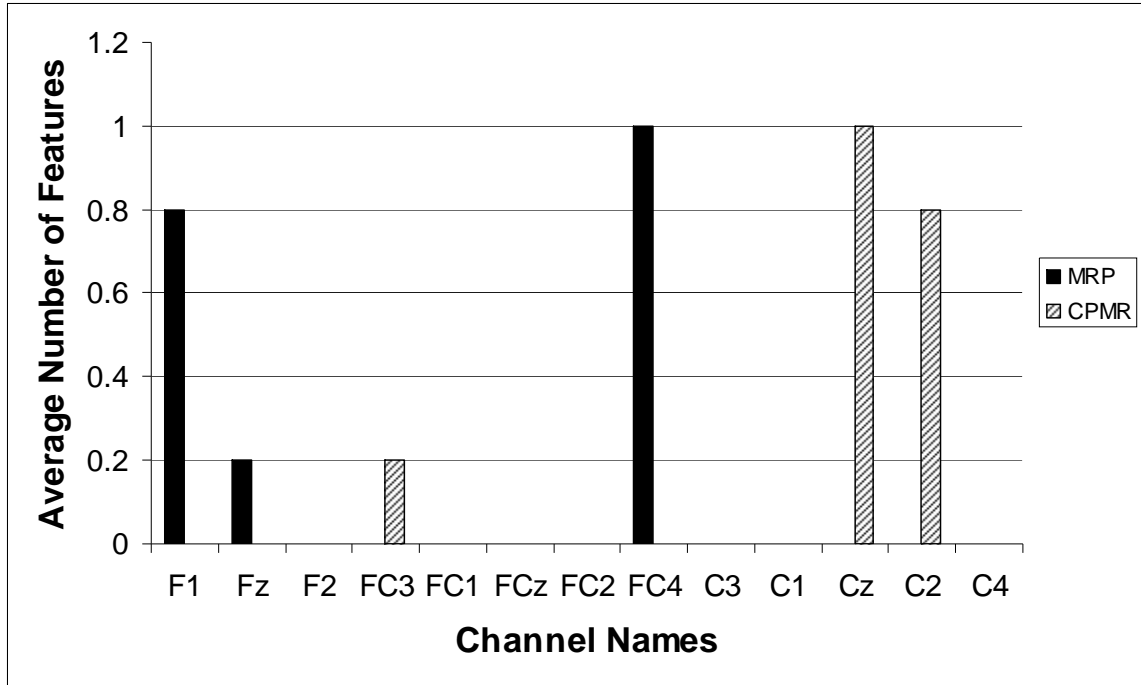




(b)



(c)



(d)

Figure 4-7. The spatial distribution of the selected features for individuals in Study 2. (a) User AB1; (b) User AB2; (c) User AB3; and (d) User AB4.

Table 4-3. The average TP and FP rates (%) for User AB1 in Study 2 (the numbers in parenthesis show the standard deviation).

Configuration	Validation			Test		
	<i>TPR</i>	<i>FPR</i>	$\frac{TPR}{FPR}$	<i>TPR</i>	<i>FPR</i>	$\frac{TPR}{FPR}$
<b>Configuration1</b>	24.3 (2.00)	0.4 (0.1)	59.1	26.4 (6.6)	0.4 (0.1)	68.3
<b>Configuration2</b>	25.2 (3.0)	0.3 (0.1)	72.8	25.1 (9.3)	0.4 (0.2)	57.4
<b>Configuration3</b>	16.1 (0.8)	0.7 (0.1)	23.3	16.1 (5.1)	0.6 (0.3)	26.0
<b>Configuration4</b>	<b>27.9</b> (2.8)	<b>0.2</b> (0.1)	<b>131.9</b>	<b>26.0</b> (9.5)	<b>0.1</b> (0.1)	<b>202.0</b>
<b>Configuration5</b>	11.4 (1.6)	0.0 (0.0)	297.3	10.3 (5.0)	0.0 (0.1)	200.7

**Table 4-4.** The average TP and FP rates (%) for User AB2 in Study 2 (the numbers in parenthesis show the standard deviation).

Configuration	Validation			Test		
	<i>TPR</i>	<i>FPR</i>	$\frac{TPR}{FPR}$	<i>TPR</i>	<i>FPR</i>	$\frac{TPR}{FPR}$
<b>Configuration1</b>	25.2 (2.2)	0.3 (0.1)	83.9	25.2 (5.7)	0.3 (0.1)	85.3
<b>Configuration2</b>	27.4 (2.2)	1.2 (0.2)	22.7	27.0 (7.9)	1.3 (0.5)	21.0
<b>Configuration3</b>	22.3 (1.9)	1.1 (0.2)	19.9	23.5 (6.8)	1.5 (0.2)	16.1
<b>Configuration4</b>	32.5 (2.6)	0.7 (0.2)	45.2	27.0 (3.9)	0.3 (0.1)	97.7
<b>Configuration5</b>	17.1 (3.0)	0.1 (0.0)	192.8	16.1 (6.8)	0.2 (0.1)	101.9

**Table 4-5.** The average TP and FP rates (%) for User AB3 in Study 2 (the numbers in parenthesis show the standard deviation).

Configuration	Validation			Test		
	<i>TPR</i>	<i>FPR</i>	$\frac{TPR}{FPR}$	<i>TPR</i>	<i>FPR</i>	$\frac{TPR}{FPR}$
<b>Configuration1</b>	34.9 (5.5)	1.0 (0.1)	33.5	36.7 (3.7)	1.1 (0.4)	33.4
<b>Configuration2</b>	20.3 (4.5)	1.4 (0.3)	14.7	16.0 (4.1)	1.2 (0.7)	13.2
<b>Configuration3</b>	21.5 (3.1)	1.9 (0.3)	11.3	18.3 (8.3)	2.0 (0.6)	9.3
<b>Configuration4</b>	27.9 (1.3)	1.3 (0.1)	21.7	31.3 (13.5)	3.4 (4.0)	9.3
<b>Configuration5</b>	12.7 (3.8)	0.2 (0.1)	64.8	16.8 (5.9)	0.2 (0.1)	85.5

**Table 4-6. The average TP and FP rates (%) for User AB4 in Study 2 (the numbers in parenthesis show the standard deviation).**

Configuration	Validation			Test		
	<i>TPR</i>	<i>FPR</i>	$\frac{TPR}{FPR}$	<i>TPR</i>	<i>FPR</i>	$\frac{TPR}{FPR}$
<b>Configuration1</b>	23.2 (2.4)	0.6 (0.1)	41.7	24.7 (8.7)	0.5 (0.5)	46.9
<b>Configuration2</b>	22.2 (3.3)	0.4 (0.1)	54.6	21.4 (6.2)	0.4 (0.3)	50.2
<b>Configuration3</b>	18.4 (1.8)	1.3 (0.1)	14.0	16.5 (6.9)	1.4 (0.6)	11.9
<b>Configuration4</b>	32.5 (4.8)	0.9 (0.2)	35.0	31.9 (9.0)	1.1 (0.3)	27.6
<b>Configuration5</b>	13.9 (2.6)	0.1 (0.1)	101.1	13.0 (0.5)	0.1 (0.1)	103.5

## 4.6 Discussion

In this chapter, we used the spatiotemporal information extracted from different neurological phenomena to design a new SBCI system. Our main rationale was to achieve low FP rates at an acceptable TP rate. The results presented in Table 4-1 to Table 4-6 and in Figure 4-7 led us to the following interesting observations.

### 4.6.1 Observations on the BCI designs based on a single neurological phenomenon

When only one neurological phenomenon was used in the design of the SBCI system, the FP rates (even when calculated on the validation sets) were simply too high, especially for the case of CPMR and CPBR where the TP rates were also relatively low (see Table 4-2). The finding that each of the CPBR and CPMR by itself is not suitable for a successful operation of an SBCI is not surprising. It was reported in [26] that the ERD of the Mu rhythms is not equally strong in all individuals. Control that is solely based on the Mu rhythms might even fail if the intervals between consecutive periods are too short to allow for proper synchronization. For example, in [25, 67], the results of three individuals had to be excluded from further analysis, because the pre-analysis showed very poor error rates. Moreover, an SBCI system based on the application of Mu rhythms had shown poor performance [2].

#### **4.6.2 Observations on Study 1**

In Study 1, all the features were combined into a single feature vector. As seen in Table 4-1 and Table 4-2, the results obtained in Study 1 show improvements over systems based on a single neurological phenomenon. However, perhaps with the exception of User AB1 (who had an FP rate of 1.65%) the FP rates for all other users still remained high (2.6, 2.8 and 11.5%). FP rates higher than 2% are simply too high, as they cause excessive user frustration [55].

#### **4.6.3 Observations on Study 2**

- Table 4-3 to Table 4-6 show that when only MRPs were included in the structure of the two-stage MCS, an acceptable performance was achieved.
- Table 4-3 to Table 4-6 show that the selection of the best configuration is subject-dependent in the sense that the configuration leading to superior performance is different for different users. Further evidence from the literature strengthens this observation. For example, in [50], it was shown that for better classification between left and right movements, the Mu rhythms were the best to use for one user, the Beta rhythms were more suitable for another user, while for the third user, both rhythms were suitable. This strongly suggests that the amount of discriminatory information is subject-dependent.
- Figure 4-7 clearly shows that not all EEG channels are necessarily needed for the successful operation of the BCI system. The use of fewer channels will certainly lead to faster setup of the system.

Moreover, as shown in Figure 4-7, the spatial distribution of the selected features is different for different users. These results imply that it is necessary to develop a user-customized BCI system for each user.

#### **4.6.4 Statistical analysis**

A two-way analysis of variance (ANOVA) was carried out to compare the results of Study 1 and Study 2. The results achieved in Study 1 were compared with the results of the best Configuration for each user in Study 2. The independent variables were

“Individual “ and “Study” and  $\frac{TPR}{FPR}$  was the dependant variable. The results showed a significant main effect of Individual ( $p < 0.001$ ), a significant main effect of Study ( $p < 0.0005$ ) and a marginally significant effect of the interaction of both ( $p = 0.062$ ). ANOVA confirms that the proposed method in Study 2 results in a superior performance compared to the more conventional scheme implemented in Study 1.

To further check the configuration effect, we conducted a two-way ANOVA on the results. “Individual” and “Configuration” were used as independent variables and the performance (expressed as  $\frac{TPR}{FPR}$ ) was used as the dependent variable. “Configuration 5” was not included in this analysis, because it resulted in low TP rates, thus making its application very limited. The ANOVA results revealed a significant main effect of “Individual” ( $p < 10^{-4}$ ), a significant main effect of “Configuration” ( $p < 10^{-4}$ ) and a significant interaction of both ( $p < 10^{-4}$ ). Further analysis showed that Configuration 1 (the combination of MRP and CPBR) resulted in a superior performance than Configuration 3 (the combination of CPMR and CPBR) with  $p < 0.05$ . Similarly, we compared the performance of Configuration 2 (the combination of MRP and CPMR) vs. Configuration 3. Although the average  $\frac{TPR}{FPR}$  for Configuration 2 was higher with that of Configuration 3, the amount of improvement was not statistically significant ( $p > 0.05$ ).

Furthermore, Configuration 1 resulted in a higher  $\frac{TPR}{FPR}$  ratio compared to Configuration 2. However, these results were not statistically significant ( $p > 0.05$ ), so future work should further explore this issue in order to determine whether or not the combination of MRP and CPBR is indeed superior to the combination of MRP and CPMR.

The combination of all three neurological phenomena through majority voting (Configuration 4) resulted in a statistically significant superior performance in terms of  $\frac{TPR}{FPR}$  compared to Configurations 2 and 3 ( $p < 0.05$ ). However, compared to that of

Configuration 1 the improvements in the performance were not statistically significant ( $p > 0.05$ ).

We also carried out a more detailed analysis on Configuration 5 . A two-way ANOVA with “FPR” as the dependent variable and “Configuration” and “Individual” as independent variables, showed that this Configuration achieved the lowest false positive rates compared to other Configurations ( $p < 0.05$ ) with average true negative rate of 99.9%. However, the true positive rates were also the lowest amongst all the configurations ( $p < 0.05$ ) with an average TP rate of 14.1%. This low FP rate was achieved because this “conservative” configuration generated an IC control command only when MRP, CPBR and CPMR were all detected simultaneously. It is expected, however, if more powerful feature extraction schemes are used, the TP performance of this configuration would improve and still benefit from the very low FP performance.

It is difficult to directly compare the results of our study with those of other SBCI studies, as the type of participants in the experiments (whether of not they are able-bodied), the recording equipment, recording protocols, classification protocols, the neurological phenomena considered, and even the evaluation methods differ from study to study. One crucial difference between this study and many other SBCI studies is that here the TP rate is only considered at the time of movement. However, in other studies, any detection, from  $x$  seconds before to  $y$  seconds after the movement trigger is considered as a TP [55, 68], since the exact time of movement was not known. On the other hand, since in our studies, the time of movement was known with a good approximation, we only considered the IC periods collected at the time of movement.

Since this study is based on the use of executed movements, the results should be considered as preliminary steps toward the development of an SBCI system. This is because although the evidence from the literature supports the similarities between executed and attempted movements (see Section 4.3 ), there are also some important differences between them. For example the negative potential shift before the movement onset is of smaller amplitude in the imagined movements compared to executed movements [69]. Also the reafferent feedback after the movement offset is missing in the

imagined movements [53]. Future studies should then extend the obtained results to individuals with motor disabilities.

Future work should also study extension of the proposed method to continuous classification of the EEG signals. The definition of a debounce window [55], for example, may lead to a lower FP rate when the proposed algorithm is tested on the continuous classification of EEG signals.

Although the GA was used to automatically select the channels, the values of the classifiers' parameters were found using cross-validation. This procedure is obviously suboptimal, as it is computationally demanding. Currently none of the SBCI designs have studied simultaneous optimization of both feature selection and parameter selection procedures. Another part of our future work will thus examine development of a fully automated user-customized self-paced BCI system.

Future work would also involve studying more complex feature extraction and feature classification strategies. For example, the wavelet transform could form a suitable candidate for exploring different neurological phenomena because it allows for simultaneous analysis of both time and frequency features. On the other hand, more complex classifiers, such as support vector machines (SVMs), have recently been successfully applied to a number of BCI systems. The application of these tools is certainly worthwhile exploring. In this chapter, we only considered binary outputs for the classifiers. To increase the performance, future work could also explore continuous outputs. To minimize the presence of other cognitive tasks, future work will also include testing the proposed BCI system in a complete self-paced environment (without the presence of any preparation cue).

## **4.7 Acknowledgements**

This work was supported in part by the NSERC under Grant 90278-06 and the CIHR under Grant MOP-72711. This research was enabled by the use of WestGrid computing resources, which are funded in part by the Canada Foundation for Innovation, Alberta Innovation and Science, BC Advanced Education, and the participating research institutions. WestGrid equipment is provided by IBM, Hewlett Packard and SGI. The



authors also would like to thank Mr. Craig Wilson for his valuable comments on this chapter.

## 4.8 References

- [1] S. G. Mason and G. E. Birch, "A general framework for brain-computer interface design," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no.1, pp. 70-85, Mar. 2003.
- [2] S. G. Mason and G. E. Birch, "A brain-controlled switch for asynchronous control applications," *IEEE Trans. Biomed. Eng.*, vol. 47, no.10, pp. 1297-1307, Oct. 2000.
- [3] S. G. Mason and G. E. Birch, "Temporal control paradigms for direct brain interfaces - rethinking the definition of asynchronous and synchronous," in *Proc. HCI Int. Conf.*, Las Vegas, USA, 2005.
- [4] G. E. Birch, P. D. Lawrence and R. D. Hare, "Single-trial processing of event-related potentials using outlier information," *IEEE Trans. Biomed. Eng.*, vol. 40, no.1, pp. 59-73, Jan. 1993.
- [5] 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. Rehabil. Eng.*, vol. 8, no.2, pp. 180-185, Jun. 2000.
- [6] R. Millan Jdel and J. Mourino, "Asynchronous BCI and local neural classifiers: an overview of the Adaptive Brain Interface project," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no.2, pp. 159-161, Jun. 2003.
- [7] R. Scherer, G. R. Muller, C. Neuper, B. Graimann and G. Pfurtscheller, "An asynchronously controlled EEG-based virtual keyboard: improvement of the spelling rate," *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 979-984, Jun. 2004.
- [8] E. Yom-Tov and G. F. Inbar, "Detection of movement-related potentials from the electroencephalogram for possible use in a brain-computer interface," *Med. Biol. Eng. Comput.*, vol. 41, no.1, pp. 85-93, Jan. 2003.
- [9] G. Townsend, B. Graimann and G. Pfurtscheller, "Continuous EEG classification during motor imagery--simulation of an asynchronous BCI," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 12, no.2, pp. 258-265, Jun. 2004.
- [10] 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, no.6, pp. 767-791, Jun. 2002.
- [11] G. Dornhege, B. Blankertz, G. Curio and K. R. Muller, "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms," *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 993-1002, Jun. 2004.
- [12] C. Babiloni, F. Carducci, F. Cincotti, P. M. Rossini, C. Neuper, G. Pfurtscheller and F. Babiloni, "Human movement-related potentials vs desynchronization of EEG alpha rhythm: a high-resolution EEG study," *Neuroimage*, vol. 10, no.6, pp. 658-665, Dec. 1999.
- [13] W. J. Krzanowski, P. Jonathan, W. V. McCarthy and M. R. Thomas, "Discriminant Analysis with Singular Covariance Matrices: Methods and Applications to Spectroscopic Data," *Applied Statistics*, vol. 44, no.1, pp. 101-115, 1995.
- [14] J. Kittler, M. Hatef, R. P. W. Duin and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no.3, pp. 226-239, 1998.

- [15] A. Verikas, A. Lipnickas, K. Malmqvist, M. Bacauskiene and A. Gelzinis, "Soft combination of neural classifiers: A comparative study," *Pattern Recog. Lett.*, vol. 20, no.4, pp. 429-444, 1999.
- [16] D. M. J. Tax, M. V. Breukelen, Duin, R. P. W. and J. Kittler, "Combining multiple classifiers by averaging or multiplying?" *Pattern Recognition*, vol. 33, pp. 1475-1485, 2000.
- [17] A. Tsymbal, S. Puuronen and D. W. Patterson, "Ensemble feature selection with the simple Bayesian classification," *Information Fusion*, vol. 4, no.2, pp. 87-100, 2003.
- [18] K. R. Muller, G. Curio, B. Blankertz and G. Dornhege, "Combining features for BCI," in the *Proc. Advances in Neural Inf. Proc. Systems Conf. (NIPS 02)*, 2002.
- [19] S. Lemm, B. Blankertz, G. Curio and K. R. Muller, "Spatio-spectral filters for improving the classification of single trial EEG," *IEEE Trans. Biomed. Eng.*, vol. 52, no.9, pp. 1541-1548, Sep. 2005.
- [20] W. Szurhaj, P. Derambure, E. Labyt, F. Cassim, J. L. Bourriez, J. Isnard, J. D. Guieu and F. Mauguiere, "Basic mechanisms of central rhythms reactivity to preparation and execution of a voluntary movement: a stereoelectroencephalographic study," *Clin. Neurophysiol.*, vol. 114, no.1, pp. 107-119, Jan. 2003.
- [21] A. Bashashati, M. Fatourehchi, R. K. Ward and G. E. Birch, "User Customization of the Feature Generator of an Asynchronous Brain Interface," *Annals of Biomedical Engineering*, vol. 34, no.6, pp. 1051-1060, Jun. 2006.
- [22] G. Blanchard and B. Blankertz, "BCI Competition 2003--Data set IIa: spatial patterns of self-controlled brain rhythm modulations," *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 1062-1066, Jun. 2004.
- [23] W. Xu, C. Guan, C. E. Siong, S. Ranganatha, M. Thulasidas and M. Wu, "High accuracy classification of EEG signal," in *Proc. IEEE Int. Conf. Pattern Recognition (ICPRA'04)*, pp. 391-394, 2004.
- [24] M. Pregenzer and G. Pfurtscheller, "Frequency component selection for an EEG-based brain to computer interface," *IEEE Trans. Rehabil. Eng.*, vol. 7, no.4, pp. 413-419, Dec. 1999.
- [25] T. N. Lal, M. Schroder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer and B. Scholkopf, "Support vector channel selection in BCI," *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 1003-1010, Jun. 2004.
- [26] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clin. Neurophysiol.*, vol. 110, no.11, pp. 1842-1857, Nov. 1999.
- [27] L. Deecke, B. Grozinger and H. H. Kornhuber, "Voluntary finger movement in man: cerebral potentials and theory," *Biol. Cybern.*, vol. 23, no.2, pp. 99-119, Jul 14. 1976.
- [28] H. Shibasaki, G. Barrett, E. Halliday and A. M. Halliday, "Components of the movement-related cortical potential and their scalp topography," *Electroencephalogr. Clin. Neurophysiol.*, vol. 49, pp. 213-226, Aug. 1980.
- [29] I. M. Tarkka and M. Hallett, "Cortical topography of premotor and motor potentials preceding self-paced, voluntary movement of dominant and non-dominant hands," *Electroencephalogr. Clin. Neurophysiol.*, vol. 75, no.2, pp. 36-43, Feb. 1990.
- [30] M. Hallett, "Movement-related cortical potentials," *Electromyogr. Clin. Neurophysiol.*, vol. 34, no.1, pp. 5-13, Jan-Feb. 1994.

- [31] A. Urbano, C. Babiloni, P. Onorati and F. Babiloni, "Human cortical activity related to unilateral movements. A high resolution EEG study," *Neuroreport*, vol. 8, no.1, pp. 203-206, Dec. 1996.
- [32] A. Urbano, C. Babiloni, P. Onorati, F. Carducci, A. Ambrosini, L. Fattorini and F. Babiloni, "Responses of human primary sensorimotor and supplementary motor areas to internally triggered unilateral and simultaneous bilateral one-digit movements. A high-resolution EEG study," *Eur. J. Neurosci.*, vol. 10, pp. 765-770, no.2, Feb. 1998.
- [33] G. Pfurtscheller and A. Aranibar, "Event-related cortical desynchronization detected by power measurements of scalp EEG," *Electroencephalogr. Clin. Neurophysiol.*, vol. 42, no.6, pp. 817-826, Jun. 1977.
- [34] L. Leocani, C. Toro, P. Manganotti, P. Zhuang and M. Hallett, "Event-related coherence and event-related desynchronization/synchronization in the 10 Hz and 20 Hz EEG during self-paced movements," *Electroencephalogr. Clin. Neurophysiol.*, vol. 104, no.3, pp. 199-206, May. 1997.
- [35] G. Pfurtscheller, K. Pichler-Zalaudek, B. Ortmayr, J. Diez and F. Reisecker, "Postmovement beta synchronization in patients with Parkinson's disease," *J. Clin. Neurophysiol.*, vol. 15, no.3, pp. 243-250, May. 1998.
- [36] C. Toro, G. Deuschl, R. Thatcher, S. Sato, C. Kufta and M. Hallett, "Event-related desynchronization and movement-related cortical potentials on the ECoG and EEG," *Electroencephalogr. Clin. Neurophysiol.*, vol. 93, no.5, pp. 380-389, Oct. 1994.
- [37] S. Arroyo, R. P. Lesser, B. Gordon, S. Uematsu, D. Jackson and R. Webber, "Functional significance of the mu rhythm of human cortex: an electrophysiologic study with subdural electrodes," *Electroencephalogr. Clin. Neurophysiol.*, vol. 87, no.3, pp. 76-87, Sep. 1993.
- [38] G. Pfurtscheller and A. Aranibar, "Evaluation of event-related desynchronization (ERD) preceding and following voluntary self-paced movement," *Electroencephalogr. Clin. Neurophysiol.*, vol. 46, no.2, pp. 138-146, Feb. 1979.
- [39] L. Defebvre, J. L. Bourriez, K. Dujardin, P. Derambure, A. Destee and J. D. Guieu, "Spatiotemporal study of Bereitschaftspotential and event-related desynchronization during voluntary movement in Parkinson's disease," *Brain Topogr.*, vol. 6, no.3, pp. 237-244, Spring. 1994.
- [40] L. Narici, V. Pizzella, G. L. Romani, G. Torrioli, R. Traversa and P. M. Rossini, "Evoked alpha- and mu-rhythm in humans: a neuromagnetic study," *Brain Res.*, vol. 520, pp. 222-231, Jun 18. 1990.
- [41] B. Feige, R. Kristeva-Feige, S. Rossi, V. Pizzella and P. M. Rossini, "Neuromagnetic study of movement-related changes in rhythmic brain activity," *Brain Res.*, vol. 734, pp. 252-260, Sep 23. 1996.
- [42] G. Pfurtscheller, "Central beta rhythm during sensorimotor activities in man," *Electroencephalogr. Clin. Neurophysiol.*, vol. 51, no.3, pp. 253-264, Mar. 1981.
- [43] H. S. Liu, X. Gao, F. Yang and S. Gao, "Imagined hand movement identification based on spatio-temporal pattern recognition of EEG," in *Proc. 1st Int. EMBS/BMES Conf. on Neural Eng.*, pp. 599-602, 2003.
- [44] B. Graimann, J. E. Huggins, S. P. Levine and G. Pfurtscheller, "Toward a direct brain interface based on human subdural recordings and wavelet-packet analysis," *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 954-962, Jun. 2004.

- [45] B. D. Mensh, J. Werfel and H. S. Seung, "BCI Competition 2003--Data set Ia: combining gamma-band power with slow cortical potentials to improve single-trial classification of electroencephalographic signals," *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 1052-1056, Jun. 2004.
- [46] T. Hinterberger and G. Baier, "Parametric orchestral sonification of EEG in real time," *Multimedia, IEEE*, vol. 12, no.2, pp. 70-79, 2005.
- [47] Y. Wang, Z. Zhang, Y. Li, X. Gao, S. Gao and F. Yang, "BCI Competition 2003--Data set IV: an algorithm based on CSSD and FDA for classifying single-trial EEG," *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 1081-1086, Jun. 2004.
- [48] M. Krauledat, G. Dornhege, B. Blankertz, F. Losch, G. Curio and K. - . Muller, "Improving speed and accuracy of brain-computer interfaces using readiness potential features," in Engineering in Medicine and Biology Society, 2004.EMBC 2004.Conference Proceedings.26th Annual International Conference of the2004, pp. 4511-4515 Vol.6.
- [49] R. Beisteiner, P. Hollinger, G. Lindinger, W. Lang and A. Berthoz, "Mental representations of movements. Brain potentials associated with imagination of hand movements," *Electroencephalogr. Clin. Neurophysiol.*, vol. 96, no.2, pp. 183-193, Mar. 1995.
- [50] G. Pfurtscheller, C. Neuper, D. Flotzinger and M. Pregenzer, "EEG-based discrimination between imagination of right and left hand movement," *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no.6, pp. 642-651, Dec. 1997.
- [51] C. A. Porro, M. P. Francescato, V. Cettolo, M. E. Diamond, P. Baraldi, C. Zuiani, M. Bazzocchi and P. E. di Prampero, "Primary motor and sensory cortex activation during motor performance and motor imagery: a functional magnetic resonance imaging study," *J. Neurosci.*, vol. 16, no.23, pp. 7688-7698, Dec 1. 1996.
- [52] R. Cunnington, R. Iansel, J. L. Bradshaw and J. G. Phillips, "Movement-related potentials associated with movement preparation and motor imagery," *Exp. Brain Res.*, vol. 111, no.3, pp. 429-436, Oct. 1996.
- [53] B. Graimann, J. E. Huggins, S. P. Levine and G. Pfurtscheller, "Toward a direct brain interface based on human subdural recordings and wavelet-packet analysis," *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 954-962, Jun. 2004.
- [54] S. G. Mason and G. E. Birch, "A brain-controlled switch for asynchronous control applications," *IEEE Trans. Biomed. Eng.*, vol. 47, no.10, pp. 1297-1307, Oct. 2000.
- [55] J. F. Borisoff, S. G. Mason, A. Bashashati and G. E. Birch, "Brain-computer interface design for asynchronous control applications: improvements to the LF-ASD asynchronous brain switch," *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 985-992, Jun. 2004.
- [56] G. E. Birch, Z. Bozorgzadeh and S. G. Mason, "Initial on-line evaluations of the LF-ASD brain-computer interface with able-bodied and spinal-cord subjects using imagined voluntary motor potentials," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 10, no.4, pp. 219-224, Dec. 2002.
- [57] V. J. Samar , A. Bopardikar, R. Rao, and K. Swartz, "Wavelet analysis of neuroelectric waveforms: a conceptual tutorial," Technical report, 1996.
- [58] G. Pfurtscheller, "Graphical display and statistical evaluation of event-related desynchronization (ERD)," *Electroencephalogr. Clin. Neurophysiol.*, vol. 43, no.5, pp. 757-760, Nov. 1977.

- [59] S. Lemm, C. Schafer and G. Curio, "BCI Competition 2003--Data set III: probabilistic modeling of sensorimotor mu rhythms for classification of imaginary hand movements," *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 1077-1080, Jun. 2004.
- [60] T. Kohonen, "The self-organizing map," *Proceedings of IEEE*, vol. 78, pp. 1464-1480, 1990.
- [61] E. Haselsteiner and G. Pfurtscheller, "Using Time-Dependent Neural Networks for EEG Classification," *IEEE Trans. Rehab. Eng.*, vol. 8, no.4, pp. 457-463, Dec.. 2000.
- [62] B. O. Peters, G. Pfurtscheller and H. Flyvbjerg, "Automatic Differentiation of Multichannel EEG Signals," *IEEE Trans. Biomed. Eng.*, vol. 48, no.1, pp. 111-116, Jan. 2001.
- [63] U. H. Balbale, J. E. Huggins, S. L. Bement and S. P. Levine, "Multi-channel analysis of human event-related cortical potentials for the development of a direct brain interface," in Proc. 21<sup>st</sup> IEEE EMBS Conf. Atlanta, GA, USA, pp. 447, 1999.
- [64] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Publishing Company, Reading, MA ,1989.
- [65] T. Back, D. B. Fogel and T. Michalewicz, *Evolutionary Computation*. Institute of Physics Publishing, Bristol and Philadelphia, 2000.
- [66] K. R. Muller, M. Krauledat, G. Dornhege, G. Curio and B. Blankertz, "Machine learning techniques for brain-computer interfaces," *Biomed. Tech.*, vol. 49, no.1, pp. 11-22, 2004.
- [67] H. Yoon, K. Yang and C. Shahabi, "Feature subset selection and feature ranking for multivariate time series," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 17, no.9, pp. 1186-1198, 2005.
- [68] J. E. Huggins, S. P. Levine, J. A. Fessler, W. M. Sowers, G. Pfurtscheller, B. Graitmann, A. Schloegl, D. N. Minecan, R. K. Kushwaha, S. L. Bement, O. Sagher and L. A. Schuh, "Electrocorticogram as the basis for a direct brain interface: Opportunities for improved detection accuracy," in Proc. IEEE EMBS Conf. Neural Engineering, Capri Island, Italy, pp. 587-590 , 2003.
- [69] R. Cunningham, R. Ianssek and J. L. Bradshaw, "Movement-related potentials associated with movement preparation and motor imagery," *Experimental Brain Research*, vol. 111, pp. 429-436, 1996.

## **CHAPTER 5 A SELF-PACED BRAIN COMPUTER INTERFACE SYSTEM WITH A LOW FALSE POSITIVE RATE<sup>5</sup>**

### **5.1 Introduction**

Self-paced brain computer interface (SBCI) systems allow individuals to control a device or an object using their brain signals only, and at their own pace, i.e., whenever they wish. This is unlike the traditional synchronized approach, where the user is only able to control the device during periods specified by the system [1].

The performance of SBCIs is usually determined via two objective functions: (1) the true positive (TP) rate, i.e., the rate at which the SBCI detects intentional control (IC) commands sent by the user, and (2) the false positive (FP) rate, i.e., the rate of false detections by the system when the user does not intend control. The latter periods are called No Control (NC) periods.

Currently, the performance of EEG-based SBCIs is not suitable for most practical applications. For example, the latest variation of an SBCI design, called the low frequency – asynchronous switch design (the LF-ASD) generates a false positive every 6 seconds on average (with average TP rate= 41.1%) [2]. Such frequent false activations may cause user frustration and limit the application of the system. In this Chapter, we focus on improving the performance of SBCI systems in terms of decreasing the FP rate so that the system is more suitable for practical applications. There are several ways of improving the performance of EEG-based SBCIs. These include the use of sophisticated signal processing schemes; exploring the spatial, temporal and frequency-related

---

<sup>5</sup> A version of this chapter has been submitted for publication. Fatourehchi, M., Birch, G. E., Ward, R. K., “A Self-paced Brain Interface System with a Low False Positive Rate”, *Journal of Neural Engineering*, vol.5, 2008, pp.9-23.

information of the EEG signal; and taking advantage of the information provided by different control sources (neurological phenomena) of the brain (see [3] for a review of current neurological phenomenon used in BCI systems).

To improve the performance of SBCIs, the simultaneous use of three neurological phenomena as sources of control has been recently proposed [4]. These phenomena consisted of movement-related potentials (MRP) [5-7], changes in the power of Mu rhythms (CPMR) and changes in the power of Beta rhythms (CPBR) [8, 9]. The main rationale behind using these specific neurological phenomena is that they are time-locked to the onset of a movement. Thus, when a movement occurs, they are expected to be present in the EEG. A number of papers provide some evidence that these MRP and changes in the Mu rhythms provide complementary information to explore the cognitive functions of the brain [7, 10-14]. There is also some evidence regarding the differences between the Mu and Beta rhythms [15, 16]. See [4] for more details.

In [4], an EEG-based SBCI is proposed that uses information extracted from these three neurological phenomena and achieves low FP rates. One feature is extracted for each phenomenon in each EEG channel, resulting in the generation of three features per channel. Each feature is extracted by matched filtering (MF) the signal with a template of the corresponding neurological phenomenon (created through averaging the IC epochs). Each feature is classified using a K-nearest neighbor (K-NN) classifier. Increasing the number of neurological phenomena from one to three has the disadvantage of tripling the dimensionality of the feature space. To reduce the dimensionality of the feature space, therefore, a new algorithm is developed in [4] that uses a two-stage multiple classifier system (MCS) to classify the features. An MCS forms a strong classifier by using an ensemble of “weak” classifiers. For an SBCI, the number of training IC patterns is usually limited. Therefore, the proposed two-stage MCS allows the system to examine a large number of features, thus exploring as much information as possible. To reduce the dimensionality of the feature space, a genetic algorithm (GA) is used to select a subset of features that yield near optimal performance. For simplicity, the parameter values of all classifiers are assumed to be the same and are selected through an exhaustive search. The proposed system is shown to achieve low FP rates (an average FP rate, FPR, of 0.49% for four individuals). The TP rate, however, is also low (the average



TP rate, TPR, is 27.33%). To improve the performance of the system, we notice that it has a total of  $3N$  classifiers, where  $N$  is the number of EEG channels. In [4], it is assumed that the parameters of all classifiers have the same value. The parameter values are then found using an exhaustive search. This process is clearly suboptimal. Furthermore, because of the computational complexity involved, the corresponding MCS for each neurological phenomenon is designed separately. For each MCS, a separate GA is employed to select the features that produce the best performance. A better design would be to have the process of feature selection carried out simultaneously for all three MCSs.

In this Chapter, we introduce improvements to this system to boost its performance. A method that uses a stationary wavelet transform (SWT) and matched filtering is developed for feature extraction. Support vector machines (SVMs) are used for classification because they have the advantage of minimizing the empirical risk (the training error), as well as the confidence error (the test error) [17]. We also used bipolar EEG signals instead of monopolar EEG signals as in [4]. This is done by first recording the EEG signals in a monopolar fashion (e.g., electrodes F1 and FC1, referenced to ear electrodes). Then the bipolar EEG signals are generated by calculating the differences between each adjacent pairs of EEG electrodes (e.g., F1-FC1 in the above example). Bipolar signals were calculated because it has been shown that bipolar EEG signals may result in the generation of more discriminant wavelet features (extracted from MRPs) than when monopolar EEG signals are used [18]. Since using bipolar EEG signals leads to an increase in the number of EEG signals, the dimensionality of the feature space as well as the number of classifier parameters whose values need to be estimated increases. A hybrid genetic algorithm (HGA) is proposed to automate the design process of the improved SBCI. The HGA simultaneously selects the features, estimates the classifiers' parameters and chooses how the outputs of MCSs developed for each neurological phenomenon, should be combined together. Analysis of the data obtained from four able-bodied individuals (coded as AB1 to AB4) showed that the improved SBCI performs significantly better than previous SBCIs.

## 5.2 Methods

The structure of the improved SBCI is shown in Figure 5-1. For each neurological phenomenon in every EEG signal (there are  $N$  EEG signals in total), features are extracted using an SWT. To reduce the dimensionality of the wavelet feature space, we propose the use of a matched filter. For each neurological phenomenon in an EEG channel, an SVM is designed (resulting in a total of  $3N$  classifiers). The output of each SVM is a logical state '1', when an IC pattern is detected and is '0' in other cases. For each neurological phenomenon, an MCS classifies the outputs of  $N$  SVMs using the majority voting rule. A 2nd-stage MCS uses the outputs of the three MCSs to decide the outputs of which MCSs should be combined together and how this combination shall be done. An HGA is employed to simultaneously find (1) the subset of features, (2) the parameter values for each SVM and (3) the configuration of the three MCSs that leads to near optimal performance (defined as the  $\frac{TPR}{FPR}$  ratio). In the rest of this section, we describe the details of the components of this two-stage MCS.

### 5.2.1 Feature extraction

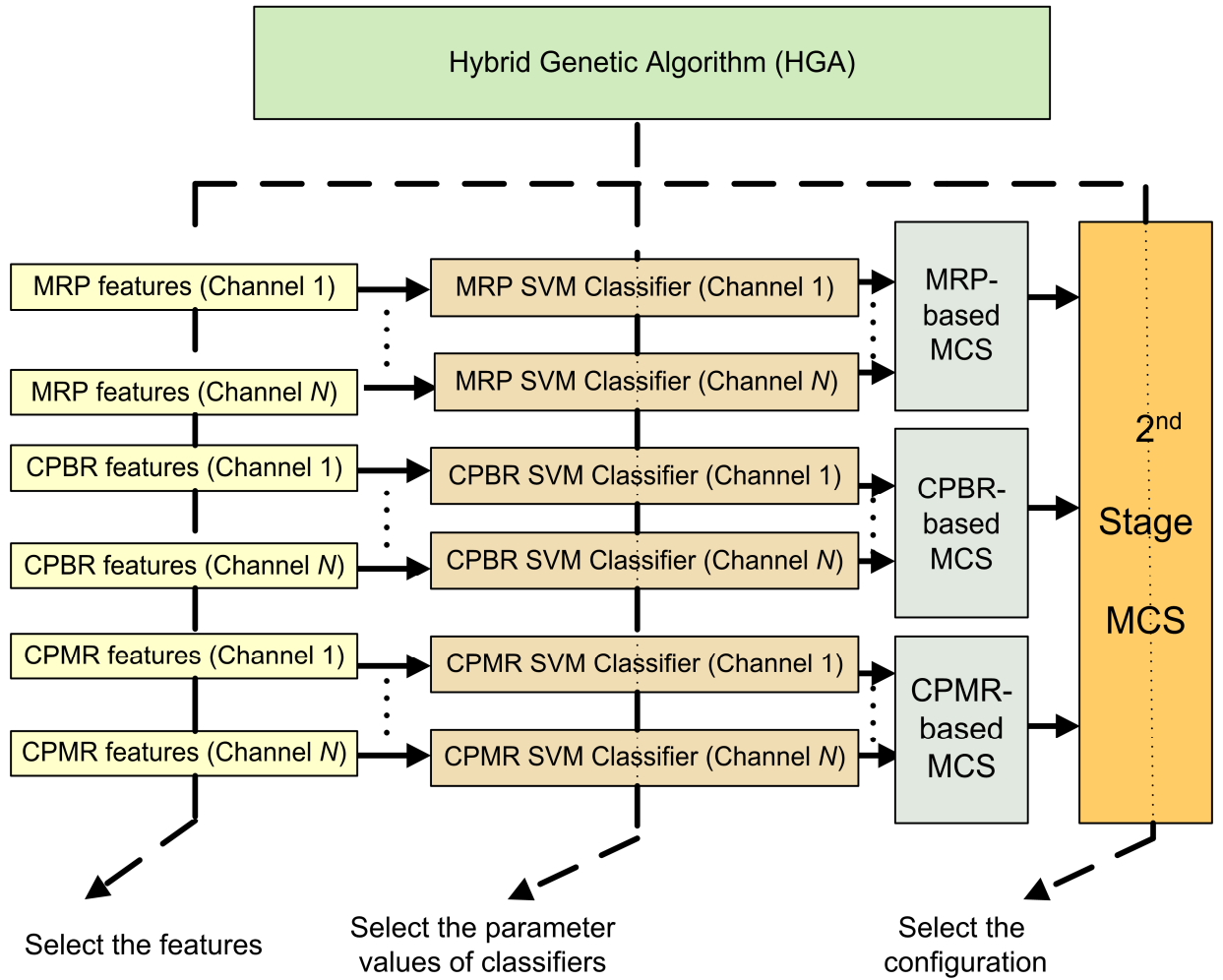
The discrete wavelet transform (DWT) is a powerful tool for extracting time-frequency features. It has been extensively applied in the analysis of event-related potentials (ERPs) [19, 20], as well as in the design of BCI systems [21-25].

The DWT is defined as the convolution of the signal  $x(t)$  with wavelet functions  $\psi_{a,b}(t)$ , where  $\psi_{a,b}(t)$  is the dilated and shifted version of the *wavelet function*  $\psi(t)$  and is defined as follows:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (5-1)$$

where  $a$  and  $b$  are the scale and translation parameters, respectively. The DWT thus maps a signal of one independent variable  $t$  into a function of two independent variables  $a, b$ , such that

$$a_j = 2^{-j}, b_{j,k} = 2^{-j} \cdot k \quad (j, k \text{ are integers}) \quad (5-2)$$



**Figure 5-1. The overall structure of the improved SBCI incorporating three neurological phenomena.**

The contracted versions of the wavelet function will match the high-frequency components of the original signal and the dilated versions will match the low-frequency oscillations. Then by correlating the original signal with the wavelet functions of different sizes, the details of the signal at different scales are obtained. The resulting correlation features can be arranged in a hierarchical scheme called multi-resolution decomposition [26] which separates the signal into “details” at different frequency bands and a coarser representation of the signal called an “approximation”. See [26] more details.

DWT, however, is shift-variant, and the values of wavelet coefficients may vary even with small shifts in time [27]. Therefore, we propose using the shift-invariant stationary wavelet transform (SWT) to detect the neurological phenomenon of interest.

An SWT resolves the shift-variancy problem associated with the DWT by eliminating the downsampling operator from the multi-resolution analysis [28]. We first describe application of the SWT to extract features from MRPs and then discuss feature extraction from CPMR and CPBR.

Consider the set of all training data consisting of  $N_{IC}$  IC commands. Suppose that each of the training epochs is decomposed using a wavelet function  $\psi(t)$ . If the wavelet coefficients are to be used as features, the number of features becomes:

$$N_{Features} = (N_{Level} + 1) \times N_{Samples} \quad (5-3)$$

where  $N_{Features}$ ,  $N_{Level}$  and  $N_{Samples}$  denote the total number of wavelet features per EEG signal, the number of decomposition levels and the number of samples per epoch, respectively. It is apparent that the size of the feature space becomes very large, even for a small number of EEG signals. To reduce the dimensionality of the feature space, we propose using a matched filter. A linear matched filter is known to be a simple yet useful tool for measuring the similarities between two sequences. Assuming that  $c_{j,k,p}$  and  $d_{j,k,p}$  are the approximation and detail coefficients at scale  $j$  and translation  $k$  of the  $p^{th}$  epoch in the training set of the IC commands, the averages of the approximation and detail coefficients at scale  $j$  and translation  $k$  ( $\bar{c}_{j,k}$ ,  $\bar{d}_{j,k}$ ) are:

$$\bar{c}_{j,k} = \frac{1}{N_{IC}} \cdot \sum_{p=1}^{N_{IC}} c_{j,k,p} \quad (5-4)$$

$$\bar{d}_{j,k} = \frac{1}{N_{IC}} \cdot \sum_{p=1}^{N_{IC}} d_{j,k,p} \quad (5-5)$$

The approximation template at scale  $j$  ( $TemplateC_j$ ) and the detail template at scale  $j$  ( $TemplateD_j$ ) are then obtained using the following formulae:

$$TemplateC_j = \{\bar{c}_{j,k}\} \quad (j=1,2,\dots,N_{Level}, k=1,2,\dots,N_{Samples}) \quad (5-6)$$

$$TemplateD_j = \{\bar{d}_{j,k}\} \quad (j=1,2,\dots,N_{Level}, k=1,2,\dots,N_{Samples}) \quad (5-7)$$

Let  $C_{j,p} = \{c_{j,k,p}\}$  ( $j = 1, 2, \dots, N_{Level}$ ,  $k = 1, 2, \dots, N_{Samples}$ , and  $p = 1, 2, \dots, N$ ) be the set of all approximation coefficients at scale  $j$  of the  $p^{th}$  epoch. The cross-covariance between  $TemplateC_j$  and  $C_{j,p}$  is then calculated as follows:

$$XCOR_{j,p}(n) = E((TemplateC_{j_m} - \mu_{TemplateC_j})(C_{j,p_{m+n}} - \mu_{C_j})^*) \quad (5-8)$$

where  $E$  is the expected value operator. After calculating  $XCOR_{j,p}(n)$  for each epoch, the following features, representing the maximum of the cross-correlogram over a period of 0.125 seconds, are extracted [4]:

$$F_{j,p} = \text{Max}_n[XCOR_{j,p}(n)], \quad \text{where } n \in [t_{start} + t_{finish} - 0.0625, t_{start} + t_{finish} + 0.0625] \quad (5-9)$$

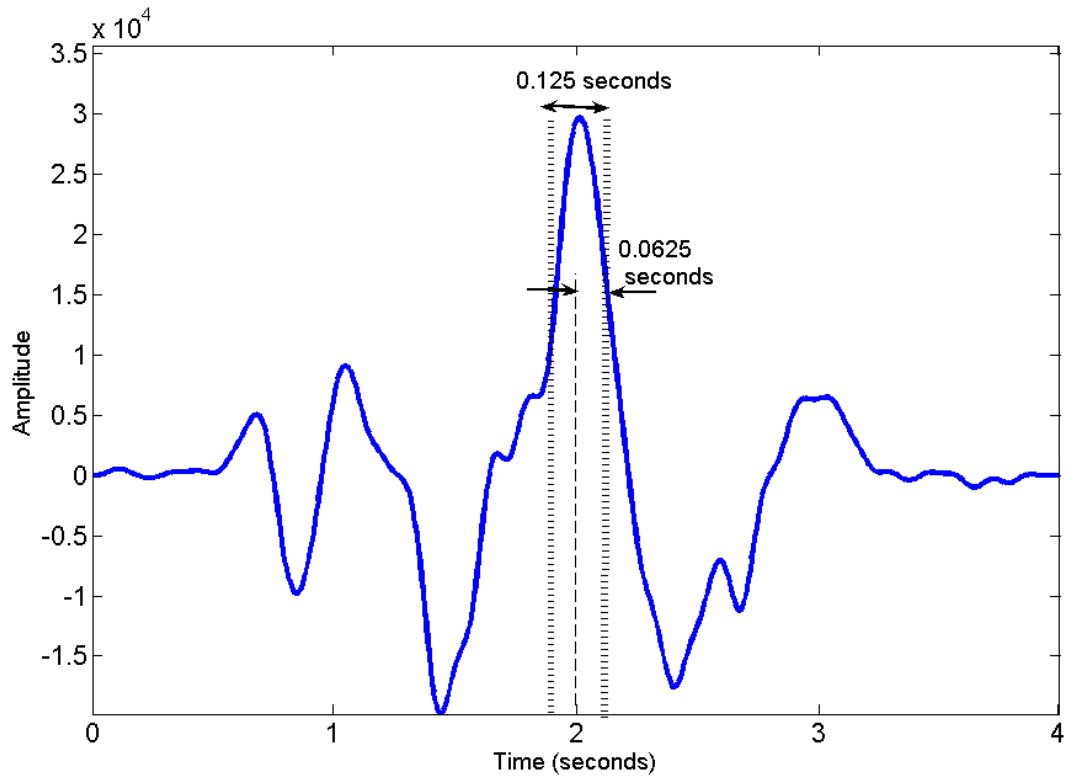
where  $(t_{finish} - t_{start})$  is the length of the epoch, and  $t_{finish}$  and  $t_{start}$  show the start and finish of an epoch as discussed in detail in Section 5.3. Figure 5-2 demonstrates an example of this feature extraction method, assuming that the length of the epoch and the template are both  $t_{finish} - t_{start} = 2$  seconds (the duration of the cross-covariance signal will thus be four seconds). The feature extractor considers a window of width 0.125 seconds around the middle point of the cross-correlogram (i.e., at time  $t = t_{finish} - t_{start} = 2$  seconds). This window covers from  $t_1 = \frac{0.125}{2} = 0.0625$  seconds before to  $t_2 = \frac{0.125}{2} = 0.0625$  after  $t = 2$  seconds. The maximum value in (5-9) is then calculated over this window with width of 0.125 seconds, because MRPs lie in frequencies below 4 Hz [18], and the sampling rate is 128Hz. Features are then generated by sliding a window over the EEG signal by shifts of 0.125 seconds.

Apart from the above features, the following features are also extracted:

$$T_{j,p} = t(XCOR_{j,p}(n) = F_{j,p}) \quad (5-10)$$

where  $t$  is the time operator. This feature provides information about the time instant when the maximum of the cross-correlogram occurs. Similar formulae can be obtained for the detail coefficients as well as for the features extracted from the NC epochs. This process is repeated for all EEG channels. We select the features belonging to the coarsest

approximation and detail levels. As a result, four MRP features are generated for each EEG channel.



**Figure 5-2.** An example of how features are extracted using the proposed cross-covariance method.

For the CPMR and CPBR phenomena, all epochs are band-pass filtered before feature extraction. For CPMR, the band pass is chosen from 8 to 12Hz. For CPBR, because of the relatively wide range of the Beta rhythms, a user-customized band pass is chosen for each individual, as explained below. Both filters are linear phase 32-point FIR filters. The amplitudes of the bandpass-filtered signals are squared to obtain the power values. The SWT is then applied and the wavelet coefficients of the power signals are calculated. The rest of the feature extraction process is similar to that used for MRPs and it yields four CPMR features and four CPBR features for each EEG channel.

***The choice of the proper wavelet function.*** In the analysis of ERPs, the wavelet function is usually chosen solely based on the similarity between the neurological phenomenon and the shape of the wavelet function [21, 29, 30]. The downside of this

approach is that the choice of wavelet function may become subjective. Moreover, it has been shown that the shape of the neurological phenomenon may vary from one individual to another[31]. As a result, to achieve better performance, this process needs to be carried out separately for each individual. Even if a separate wavelet function is chosen for each individual, the use of a single wavelet function for all channels may not be optimal because the amount of information varies from one EEG channel to another. For each individual, and for each neurological phenomenon in each EEG channel, a Fisher ratio is defined, as follows[32]:

$$C(p, q, r, s) = \frac{(\mu_{IC}(p, q, r, s) - \mu_{NC}(p, q, r, s))^2}{\sigma_{IC}^2(p, q, r, s) + \sigma_{NC}^2(p, q, r, s)} \quad (5-11)$$

( $p = 1, 2, 3$  ;  $q = 1, 2, \dots, N$  ;  $r = 1, 2, \dots, N_{Features}$  ;  $s = 1, 2, \dots, N_{WaveletFunctions}$ )

where  $\mu_{IC(p,q,r,s)}$  and  $\mu_{NC(p,q,r,s)}$  are the means and  $\sigma_{IC}^2(p, q, r, s)$  and  $\sigma_{NC}^2(p, q, r, s)$  are the variances of the IC and NC classes for feature  $r$  of neurological phenomenon  $p$  and for channel  $q$  extracted using wavelet function  $s$ . For each channel  $q$  and neurological phenomenon  $p$ , the wavelet function that achieves the following objective is chosen for that particular pair:

$$Max_{r,s} [(C(p, q, r, s))] \quad (p = 1, 2, 3, \quad q = 1, 2, \dots, N) \quad (5-12)$$

The wavelet functions are selected from a pool of Daubechies, Biorthogonal, Symlet and Coiflet wavelet functions (46 wavelet functions in total). Some of these wavelet functions were chosen because of their similarities with the shape of neurological phenomena. As an example, in [22] Symlet wavelets were found to be suitable for the analysis of the event-related desynchronization of the brain rhythms. Similarly, in [21], Daubechies wavelets were chosen for the analysis of event-related potentials. Biorthogonal wavelets also carry resemblance to the shape of bipolar MRPs. However, as stated earlier, we used an automatic method for selecting the type of wavelet functions to minimize the subjectivity in the choice of wavelet functions. Features are normalized prior to the calculation of the Fisher ratios.

***The choice of the proper CPBR frequency band.*** Because of the wide range of the Beta rhythms, and to select more discriminant features for CPBR, (5-11) is calculated

for seven frequency bands: [14-18], [18-22], [22-26], [26-30], [18-26], [22-30] and [14-30] Hz. Each frequency band is analyzed separately. The averages of the Fisher ratios are compared, and the frequency band that results in the highest average is selected. The reason different Beta frequency bands were considered for this study was to find user-specific frequency bands that resulted in more discriminant features (based on the Fisher's ratio). Please note that although some of the frequency bands described above are covered by other frequencies (e.g.,  $f_1 = [22 - 26]$  is covered by  $f_2 = [18 - 26]$ ), this does not mean that features extracted from  $f_2$  are necessarily more discriminant. This is because if the feature extracted from frequency band  $f_3 = [18 - 22]$  do not provide discriminant information, adding  $f_3$  to  $f_1$  may even result in decreasing the amount of discriminancy between the classes.

### 5.2.2 Feature classification

The features for each neurological phenomenon in an EEG channel are classified as an IC or NC state using an SVM classifier. For each neurological phenomenon, the classifiers' outputs are combined using an MCS. Prior to classification, outliers were removed as follows. Suppose the Mahalanobis distance for a feature vector with  $K$  variables,  $\mathbf{x} = [x_1, x_2, \dots, x_K]$  with an assumed central point  $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_K]$  is defined as

$$Mahal(\mathbf{x}, \boldsymbol{\mu}) = (\mathbf{x} - \boldsymbol{\mu}) \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})^T \quad (5-13)$$

where  $\boldsymbol{\Sigma}$  is the covariance matrix evaluated from the data. The outliers are then removed using the following algorithm [33, 34]:

1. Round  $p$ . If there exists  $\mathbf{x}$  such that  $Mahal(\mathbf{x}, \boldsymbol{\mu}) > \lambda$ , let  $FS = \{\mathbf{x} \mid Mahal(\mathbf{x}, \boldsymbol{\mu}) \leq \lambda\}$ . Retain only the points in  $FS$ . The value of  $\lambda$  was chosen such that the training samples that were further than 4 standard deviations from the mean, were considered as outliers [35].

2. Repeat until the above condition is not met.



After applying this algorithm, the maximum percentage of features recognized as outliers was 3% for NC features and 1% for IC features.

### **Support vector machines (SVMs)**

A total of  $3N$  SVM classifiers are used for each individual. Kernel-based learning combines the beneficial properties of the linear classification methods, such as simplicity, but since the feature and input spaces are nonlinearly related the overall classification is nonlinear in the input space [36]. We used the LIBSVM software for implementing the SVMs [37], and a Gaussian kernel as the kernel function. The classifier's performance depends on the regularization parameter  $C$  and the bandwidth  $\sigma$  of the kernel. Since there are  $3N$  classifiers,  $3N$  values had to be estimated for each parameter. The output of each SVM is a binary label that indicates if the input pattern belongs to an IC or an NC class.

### **Multiple classifier systems (MCSs)**

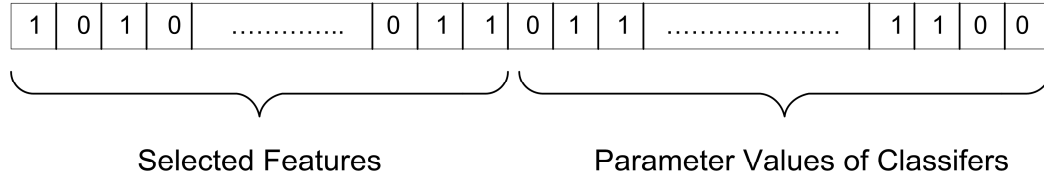
For each neurological phenomenon, an MCS with a majority voting rule classifies the binary outputs of the SVMs (there are  $N$  SVMs for each neurological phenomenon). In the case of “even” number of classifiers and if both classes have equal number of votes, the more-frequent class (NC) is chosen as the label for the input pattern. The outputs of the three MCSs are then combined using a 2nd-stage MCS as shown in Figure 5-1. This MCS can have five configurations for combining the outputs of the three MCSs as follows: (1) Configuration 1- uses the AND rule to combine the binary outputs of MCS1 and MCS2 related to MRP and CPBR, respectively. The default class is an NC (the logical state ‘0’), unless both MCS1 and MCS2 identify an IC command (the logical state ‘1’); (2) Configuration 2- uses the AND rule to combine the binary outputs of MCS1 and MCS3 that are related to MRP and CPMR, respectively; (3) Configuration 3- the AND rule is used to combine the binary outputs of MCS2 and MCS3 related to MRP and CPBR, respectively; (4) Configuration 4- the outputs of all three MCSs are combined according to the majority voting rule; (5) Configuration 5 - the AND rule is used to combine the outputs of all MCSs. The choice of the best configuration is done by an HGA as explained in the next section.

### 5.2.3 Hybrid genetic algorithm (HGA)

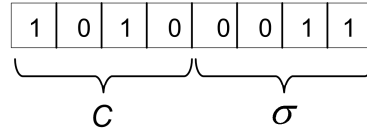
A hybrid genetic algorithm (HGA) is designed so it (1) selects the best features; (2) determines the values of the classifiers' parameters; and (3) selects the best of the five MCS configurations described in Section 5.2.2. In applying genetic algorithms to optimize the performance of the system, each parameter of interest is first coded in the form of a randomly generated binary string. Each bit in this binary string is called a gene. The concatenation of all the binary strings forms a "chromosome", and the set of "chromosomes" forms a "population". Each chromosome is then evaluated and a fitness value assigned. The chromosomes are then combined using operators such as "selection", "crossover" and "mutation" to generate new chromosomes. The "selection" operator selects a proportion of the existing population to breed a new generation. The selected chromosomes are usually the ones with higher fitness compared to other chromosomes in the population. After selection of the "fitter" chromosomes, a pair of "parent" chromosomes is selected for generating the "child" chromosomes. A child chromosome is a new solution that typically shares many of the characteristics of its "parents". The "crossover" operator ensures that this is the case by copying some of the genes of each parent to the child. The "mutation" operator is used to maintain genetic diversity from one generation of a population to the next. This process is repeated until a new population of chromosomes is generated. It is expected that the population evolves gradually and that fitness improves over generations. This process is continued until some criteria for stopping the GA is met [38].

To represent each possible combination of features, a binary chromosome of length  $L_{Chromosome}$  is defined (see Figure 5-3(a)). Bit  $i$  of the first  $N_{features}$  bits of the binary chromosome specifies whether or not feature  $i$  is selected by the HGA. A value of "1" indicates the presence of feature  $i$  and a value of "0" indicates its absence in the chromosome. The second part of the chromosome is used to select the parameter values of the classifiers. For each of the  $3N$  SVM classifiers, two parameter values need to be determined: the regularization parameter  $C$  and the bandwidth of the Gaussian kernel ( $\sigma$ ). A portion of the chromosome with length 8 bits is used for the two parameter values (see Figure 5-3(b)). The first four bits are used to represent the value of  $C$  and the second

half is used to represent the value of  $\sigma$ . Exponentially growing sequences are used for  $C$  and  $\sigma$ , i.e., their values vary from  $2^{-8}$  to  $2^7$ .



(a)



(b)

**Figure 5-3. (a) The structure of a chromosome; (b) Representation of the parameter values for each SVM in a chromosome.**

For each chromosome, a local exhaustive search is then carried out to find the best of the five configurations in the 2<sup>nd</sup>-stage MCS. Suppose  $x$  denotes a model in Figure 5-1.

In order to add a larger weight to solutions with lower false activation rates, the objective function for the HGA is defined as in (5-14).

$$f_1 : \text{Max}_x \left[ K \times \frac{\text{mean}(TPR(x))}{\text{mean}(FAR(x)) \times (\text{Var}(TPR(x)) + \text{Var}(FAR(x)))} \right]$$

$$K = \begin{cases} 1, & \text{if } \text{mean}(TPR(x)) \geq T_2 \\ \frac{\text{mean}(TPR(x)) - T_1}{T_2 - T_1}, & \text{if } T_1 \leq \text{mean}(TPR(x)) < T_2 \\ 0, & \text{if } \text{mean}(TPR(x)) < T_1 \end{cases} \quad (5-14)$$

where  $FA$  (false activations) is the percentage of NC epochs that are affected by one or more false detections. The main difference between the  $FA$  and  $FP$  rate is that multiple  $FP$ s in an epoch are counted as one  $FA$ . The values of  $T_1$  and  $T_2$  in (5-14) are selected as

50% and 80%, respectively, for all individuals except for user AB3. Please note that currently there is no consensus amongst BCI researchers, as what is the acceptable threshold for the performance of a BCI system (this is especially the case for SBCI systems). For this study, the value of  $T_1$  was chosen as 50%, for two reasons: first, we wanted to prevent the solutions with low TP rates that also had low FA rates to become dominant in the population. An example of such solution is the one with  $TPR=20\%$ ,  $FAR=1\%$ . In this example, although the FAR is very low, the TPR is also low (corresponding to the successful identification of only one out of 5 IC commands). Second, we postulated that the value of  $T_1=50\%$  will be a reasonable minimum requirement for the TP rate, as it corresponds to the identification of one out of every two IC commands on average (please note that IC commands should be separated from the periods of No Control). Any configuration that resulted in the average TP rate of less than  $T_1=50\%$  was then penalized with a zero fitness value. The value of  $T_2=80\%$  was chosen, as it has been stated that (at least for synchronized BCI systems) accuracies above 70% are considered to be acceptable [39, 40]. We thus chose the value of  $T_2=80\%$  and none of the solutions whose performances yield  $TPR \geq 80\%$  were penalized. The solutions whose fitness lies between these two extremes were penalized according to the formula described by (5-14). For user AB3, the value of  $T_1$  resulted in the generation of chromosomes with high FAR values. For this individual, the values of  $T_1$  and  $T_2$  were chosen as 33% and 50%, respectively. The “mean” operator was applied over the inner-validation sets (see Section 5.3).

We implemented a lexicographic approach for sorting the chromosomes in the HGA population [41]. In this approach, the chromosomes are compared and ranked according to the values of  $f_1(x)$  in (5-14). Any ties were resolved by comparing the relevant chromosomes again with respect to another objective. If there is also a tie, a 3<sup>rd</sup> objective function is used for comparison and so on. A total of six objective functions were used as follows (in the order of priorities):

$$f_2 : \text{Min}_x [\text{mean}(FAR(x)) \times \text{Var}(FAR(x))] \quad (5-15)$$

$$f_3 : \text{Max}_x \left[ \frac{\text{mean}(TPR(x))}{\text{Var}(TPR(x))} \right] \quad (5-16)$$

$$f_4 : \text{Min}_x[\text{mean}(FPR(x)) \times \text{Var}(FPR(x))] \quad (5-17)$$

$$f_5 : \text{Min}_x[(N(x))] \quad (5-18)$$

$$f_6 : \text{Min}_x[(N_{Features}(x))] \quad (5-19)$$

where  $N_{Features}$  is the number of features. The “mean” operator is applied over the results obtained from the inner-validation sets (see Section 5.3).

The remaining operators of the HGA are tournament-based selection (tournament size =3), uniform crossover ( $p=0.9$ ) and uniform mutation ( $p=0.01$ ). The sizes of the initial population and the rest of the populations are chosen as 200 and 100, respectively. The HGA is randomized initially. Elitism is used to keep the best performing chromosome of each population in the subsequent populations. The number of evaluations is set to 5000. If for more than 10 consecutive generations, the improvement in the first objective of the best solution was found to be less than 1%, the algorithm is terminated.

## 5.3 Experimental results

In this section, the results of the experimental analysis of the data of four able-bodied individuals are presented and the results are compared to those reported in previous BCI studies. A theoretical analysis of the performance of the proposed two-stage MCS is addressed in Appendix B.

### 5.3.1 Data collection and evaluation

The data of three male and one female able-bodied individual (denoted as users AB1 to AB4) were used in this study. The individuals were right-handed and between 31 and 56 years old. They had signed consent forms prior to participation in the experiment. The data were collected as users performed a guided right index finger flexion movement. The EEG signals were recorded from 13 monopolar electrodes positioned over the  $F_1, F_z, F_2, FC_3, FC_1, FC_z, FC_2, FC_4, C_3, C_1, C_z, C_2$  and  $C_4$  locations according to the International 10-20 System. The cutoff frequency of the amplifier was set at 30 Hz. An ocular artifact was detected when the difference between the electrooculogram (EOG)

electrodes (placed at the corner of and below the right eye) exceeded  $\pm 25 \mu\text{V}$ . This threshold was determined during data recording and by carefully monitoring the recorded EOG activity during the calibration stage. It was chosen such that most of the prominent eye movement activities were captured (see [42] for details). All signals were sampled at 128 Hz and referenced to linked ears.

The recorded signals were converted to bipolar EEG signals, since it has been shown that bipolar EEG signals may result in the generation of more discriminant wavelet features for MRPs compared to the case where monopolar EEG signals are used [18]. The conversion was carried out by calculating the difference between adjacent EEG channels, and resulted in the following 18 bipolar EEG signals:  $F_1\text{-}FC_1$ ,  $F_1\text{-}F_z$ ,  $F_2\text{-}F_z$ ,  $F_2\text{-}FC_2$ ,  $FC_3\text{-}FC_1$ ,  $FC_3\text{-}C_3$ ,  $FC_1\text{-}FC_z$ ,  $FC_1\text{-}C_1$ ,  $FC_z\text{-}FC_2$ ,  $C_1\text{-}C_z$ ,  $C_2\text{-}C_4$ ,  $FC_2\text{-}FC_4$ ,  $FC_4\text{-}C_4$ ,  $FC_2\text{-}C_2$ ,  $FC_z\text{-}C_z$ ,  $C_3\text{-}C_1$ ,  $C_z\text{-}C_2$  and  $F_z\text{-}FC_z$ . For each individual, an average of 80 IC epochs was collected every day over a period of 5 days. Table 5-1 shows the timetable of recording the data for all individuals. For each individual, “Day 1” was considered as the origin date, and the dates when the rest of the data were collected, were numbered relative to “Day1”.

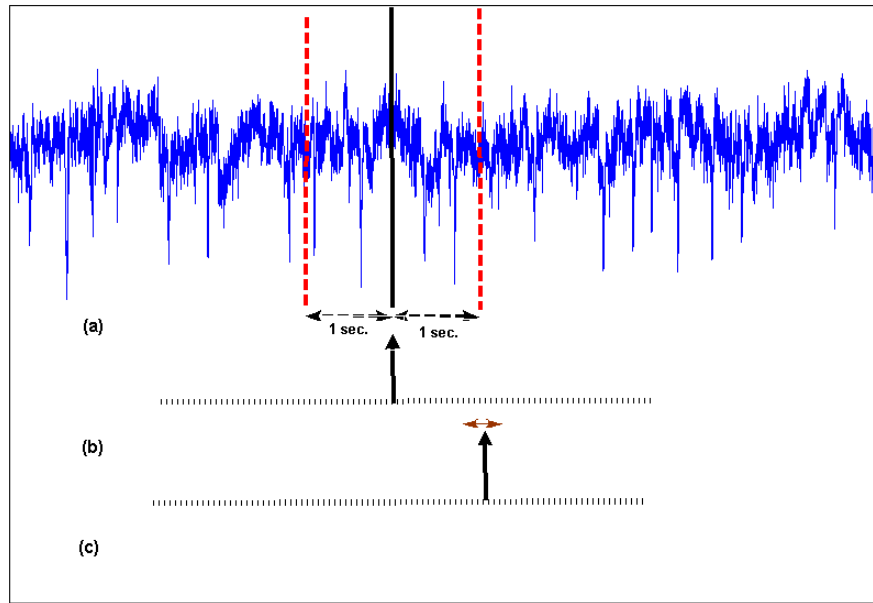
An IC epoch consisted of data collected over an interval containing the onset of movement (measured as the switch activation) as long as no artifact was detected in that particular interval. The interval started at  $t_{start} = -1$  second, i.e., 1 second before the onset of movement, and ended at  $t_{finish}$ , i.e., 1 second after the onset of movement.

**Table 5-1. The time schedule of recording the data**

User IDs	1 <sup>st</sup> session	2 <sup>nd</sup> session	3 <sup>rd</sup> session	4 <sup>th</sup> session	5 <sup>th</sup> session
AB1	1	3	5	8	10
AB2	1	3	4	8	9
AB3	1	2	4	8	9
AB4	1	3	5	8	10

As mentioned in the Introduction, an SBCI should differentiate between IC and NC epochs (in contrast to synchronized BCI systems that need to differentiate different IC commands from each other). For this reason, the data in NC sessions are also needed to represent the epochs that the user did NOT intend to control. During an NC session, users were asked to count the number of times that a white ball bounced off the monitor's screen. The NC sessions thus contained attentive as well as non-attentive NC data. Each NC session lasted approximately two minutes and during each recording day, up to two such NC sessions were recorded. The NC epochs were selected as follows: a window of width  $(t_{finish} - t_{start})$  seconds was slid over each EEG signal collected during an NC session by a step of 16 time samples (0.1250 sec), resulting in 8 classifier decisions per second. For each 1-second window that artifacts were not detected, features were extracted.

The method of calculating the TP rate is shown in Figure 5-4. In Figure 5-4(a), a sample EEG signal and in Figure 5-4(b) the output of the physical switch are shown. As stated earlier, data (from 1 second before to 1 second after a decision point) are used for classification. Assuming the system has no processing delay and the SBCI system has the ideal detection rate, the output of the SBCI system should be as demonstrated in Figure 5-4(c). In other words, the IC command should be detected one second after pressing the switch. Although, the exact timing of the switch activation is known, the neurological phenomena may not be completely time-locked to the switch activation. As a result, we have also considered any activation in the time range  $[-0.125, +0.125]$  seconds around the expected activation of the switch as a true positive (see Figure 5-4(c)). The rest of activations were treated as false activations.



**Figure 5-4. Method of calculating the TP rate; (a) EEG Signal; (b) Output of the finger switch; (c) Output of the SBCI.**

### 5.3.2 Results

A 5-level SWT decomposition resulted in the generation of wavelet coefficients in the following frequency bands: [32-64], [16-32], [8-16], [4-8], [2-4], and [0-2] Hz. For all neurological phenomena, the features were calculated for the lowest approximation and detail levels (which are attributed to the [0-2] and [2-4] Hz frequency bands, respectively). For users AB1 to AB4, the selected CPBR frequency bands were [22-30], [14-30], [22-30] and [14-18] Hz, respectively (see Section 5.2.1 for details). Although the selected frequency bands resulted in more discriminant features compared to features selected from other frequency bands, the results were not necessarily significant ( $p > 0.05$ ). This observation was consistent among all individuals.

We used a nested cross-validation to analyze the performance of the SBCI. The inner cross-validation set was used for selecting the best chromosome and the outer cross-validation set was used to test the performance. For each outer cross-validation set, 20% of the data were used for testing and the rest were used for training. The training



datasets were further divided into five folds. For each fold, 80% of the data were used for training the SVM and 20% were used for choosing the best chromosome.

The test results are shown in Table 5-2. The first row shows the selected configuration for the two-stage MCS. For users AB1 and AB4, the combination of MRP and CPBR led to superior results (Configuration 1), while for users AB2 and AB3, the combination of all three neurological phenomena using the AND rule was the best configuration (Configuration 5). The next three rows in Table 5-2 show the total number of selected bipolar signals, the number of selected channel-neurological phenomenon combinations (please note that there are three neurological phenomena and 18 bipolar EEG signals, resulting in a total of  $18 \times 3 = 54$  EEG channel-neurological phenomena combinations) and the total number of selected features (averaged over the five outer validation sets). The last three rows show the average of the TP, FA and FP rates on the

**Table 5-2. The performance results for the proposed SBCI system.**

	<b>AB1</b>	<b>AB2</b>	<b>AB3</b>	<b>AB4</b>
<b>Selected Configuration</b>	1	5	5	1
<b>Average Number of Selected Bipolar Channels</b>	18	18	18	18
<b>Average Number of Selected Channel- Neurological Phenomenon Combinations</b>	32.8 (1.8)	50.8 (2.2)	50.8 (1.5)	34.4 (1.5)
<b>Average Number of Selected Features</b>	71.0 (5.7)	112.0 (3.8)	119.0 (8.3)	81.4 (5.3)
<b>Average TP</b>	58.6 (8.6)	64.2 (7.5)	46.9 (10.4)	55.1 (5.3)
<b>Average FA</b>	1.2 (1.1)	0.2 (0.3)	2.3 (1.4)	0.8 (0.8)
<b>Average FP</b>	0.1 (0.1)	0.0 (0.0)	0.3 (0.2)	0.1 (0.1)

five outer cross-validation test sets. The numbers in parentheses show the standard deviations. These results indicate that the proposed SBCI achieves a very low FP rate at a modest TP rate. In the next section, we will show that these results are significantly better than those of previous related SBCI systems.

## 5.4 Discussion and conclusions

It is theoretically possible to design a multiple classifier system such that very good classification accuracy can be obtained (see Appendix .A). This can be achieved even if the performance of individual classifiers is only slightly better than chance. To achieve high performance, the classifiers need to be diverse. In this Chapter, we explored the information from three neurological phenomena (movement-related potentials and changes in the power of Mu and Beta rhythms) and EEG channel locations to create diverse classifiers. A hybrid genetic algorithm (HGA) was designed to maximize the performance under the computational constraints (i.e., time and computational resources). The proposed design is denoted as SBCI<sub>Fully-Automated</sub>, as in this design, features, parameter values of classifiers and the method of classifier combination have all been automatically determined.

We showed that the proposed SBCI achieves low FP rates at a modest TP rate. To our knowledge this is the first time that such low FP rate has been reported for a modest TP rate in an EEG-based SBCI. This brings the design of a practical EEG-based SBCI system with low false positive rate closer to the reality.

It is, however, difficult to directly compare the result of this study with those of other SBCI studies. This is because the recording protocol, the neurological phenomena used, the decision rate and the evaluation methodology vary amongst different studies. Furthermore, the method of labeling the output samples varies between different SBCI studies. This difficulty in comparing SBCI systems has been discussed in detail in a technical report recently published by researchers from leading research laboratories in the field of SBCI systems [43].

We compared the results of SBCI<sub>Fully-Automated</sub> with those of SBCI<sub>Semi-Automated</sub> as reported in [4]. Both studies use similar experimental paradigms (we denote the latter

design as  $SBCI_{Semi-Automated}$  as in this design, only feature extraction was automated for each neurological phenomenon.). The performance of both studies is summarized in Table 5-3. A two-way analysis of variance (ANOVA) using “User” and “Study” as independent variables shows that the TP rate increases from 27.1% to 56.2% ( $p < 10^{-5}$ ) in  $SBCI_{Fully-Automated}$  and the average FP rate decrease from 0.5% to 0.1% ( $p < 10^{-5}$ ). These results indicate that  $SBCI_{Fully-Automated}$  achieves a superior performance compared to  $SBCI_{Semi-Automated}$ .

**Table 5-3. Comparison of the performance results.**

User IDs	Results obtained in $SBCI_{Fully-Automated}$			Results obtained in $SBCI_{Semi-Automated}$			Results obtained in $LFASD_{User-Customized}$			Results obtained in $LFASD_{Path}$		
	TPR	FPR	$\frac{TPR}{FPR}$	TPR	FPR	$\frac{TPR}{FPR}$	TPR	FPR	$\frac{TPR}{FPR}$	TP R	FPR	$\frac{TPR}{FPR}$
<b>AB1</b>	58.6	0.1	390.9	26.0 (9.5)	0.1 (0.1)	200	<10.0	0.2	<50.0	42.7	1.0	42.7
<b>AB2</b>	64.2	0.0	3256.8	25.2 (5.7)	0.3 (0.1)	87.0	<10.0	0.2	<50.0	47.6	1.0	47.6
<b>AB3</b>	46.9	0.3	167.5	36.7 (3.7)	1.1 (0.4)	33.30	<8.0	0.2	<40.0	45.8	1.0	45.8
<b>AB4</b>	55.1	0.1	458.8	21.4 (6.2)	0.4 (0.4)	49.80	?	0.2	?	28.3	1.0	28.3
<b>Mean</b>	56.2	0.1	401.3	27.3	0.5	55.8	<10.0*	0.2	<50.0*	41.1	1.0	41.1

\*Average over the data of three individuals.

Compared to the SBCI<sub>Semi-Automated</sub>, the performance of the SBCI was improved because of:

1) ***Automation of the design***: In [4], the classifier parameter values and the structure of the 2nd-stage classifier were not automatically determined. The proposed method in this Chapter achieves full automation by reformulating the chromosomes and incorporating these parameters in the structure of each chromosome. We have proposed a hybrid GA for this automation process.

2) ***New feature extraction method***: We have proposed a new feature extraction method that applies a stationary wavelet transform (SWT) as a pre-processing stage and matched filtering (MF) for the final feature extraction stage. We also proposed a criterion for the automatic selection of the wavelet function (which is usually done subjectively by the designer). However, this improvement comes at the expense of increased system complexity.

The LF-ASD is another state-of-the-art SBCI, previously developed by the brain interface laboratory of the Neil Squire society [18]. During the past few years, different variations of the LF-ASD have been proposed by the members of our research group [2, 31, 44] as well as by others [45]. The LF-ASD uses features extracted from six bipolar EEG channels to distinguish an IC command (if present) from the background NC states. In Table 5-3, we also compare our results with those of two of the latest variations of the LF-ASD. In the first variation (denoted as LFASD<sub>User-Customized</sub>), the effects of user-customization of the system's parameter values by an expert are studied [31]. In the second (denoted as LFASD<sub>Path</sub>), the knowledge of the path of features is used to improve the performance[2]. Both papers focused on improving the TP rate at a fixed FP rate.

As shown in Table 5-3, the TP rates of LFASD<sub>User-Customized</sub> drop below 10% for FP rates equal to or below 0.2%. The values of TP rates at 0.2% were estimated from the receiver operating characteristic (ROC) curves plotted in[31]. Since the ROC curves for able-bodied individuals were not available in [2], we used the reported TPR results for

FPR=1%. The results in Table 5-3 show the improvement achieved in terms of  $\frac{TPR}{FPR}$  rates. As seen from this table, for low FP rates,  $SBCI_{Fully-Automated}$  achieves higher TP rates than both the  $LFASD_{User-Customized}$  and  $LFASD_{Path}$ . A t-test between the performance obtained by  $SBCI_{Fully-Automated}$  and those achieved by  $LFASD_{Path}$  shows that the TP rates in  $SBCI_{Fully-Automated}$  are higher ( $p < 0.02$ ), while the FP rates are lower ( $p = 0$ ). The same comparison with the results obtained for FPR=0.2% in  $LFASD_{User-Customized}$  shows a highly significant improvements in the TP rates ( $p < 10^{-5}$ ), while the decrease in FP rates is not statistically significant ( $p=0.16$ ).

The results in Table 5-3 also show that  $SBCI_{Fully-Automated}$  has an average of 1.11 FPs every 100 seconds. The original design of the LF-ASD had an average of one FP every six seconds [44] and the improved design had an average of one FP every 12 seconds[2]). Thus,  $SBCI_{Fully-Automated}$  is able to recognize a longer period of an NC state without having a false activation.

Although the results in Table 5-3 show that  $SBCI_{Fully-Automated}$  achieved a superior performance compared to the rest of SBCIs examined in this study, the results also indicate a great deal of inter-subject variability in terms of performance. As an example, the TP and FP rate for user AB2 were 64.2% and 0.0%, respectively, while the values obtained for user AB3 were TPR=46.9% and 0.3%, respectively. One reason that can be stated for this is the variability of the quality of the neurological phenomenon from one individual to another. For example, when the IC epochs of user AB2 were averaged, very distinct MRP patterns emerged, however, for user AB3, the MRP patterns were less pronounced (see the discussion in [31] for the variability of MRPs amongst different individuals). An interesting area that needs further exploration is to see how the qualities of the neurological phenomena will improve after individuals get more training. This is left to future studies.

One concern in BCI studies is the effect of artifacts on the performance of the system. Particularly, systems that use slow potentials such as MRP, may be vulnerable to the presence of eye movement artifacts. One advantage of our proposed system is that it uses three neurological phenomena each belonging to a different frequency band. While

eye movements are mostly low frequency components that may affect MRPs, their effect on the changes in the power of Mu and Beta rhythms is much less significant. Since our system depends on observing movement-related patterns in more than one neurological phenomenon when detecting an IC pattern, it is thus more robust to the presence of artifacts. Nevertheless, when detecting EOG artifacts using a thresholding scheme, smaller EOG artifacts may not be detected. Thus in our future studies, we plan to explore the use of more sophisticated artifact-removal methods such as Independent Component Analysis (ICA) to improve the artifact monitoring system.

Another research area that needs further attention is the choice of a suitable evaluation metric for SBCIs. The evaluation of the performance of any SBCI system greatly depends on the evaluation metric used. Currently there is no consensus amongst BCI researchers as to which performance metric summarizes the performance of a given SBCI more efficiently [43]. As an example, although in a number of SBCI papers, receiver operating characteristic (ROC) curve and the area under ROC (AUC) have been used for evaluating the performance, the suitability of this metric in the field of SBCI was recently questioned. This is because when the ROC curve is plotted over the whole range of the (TPR, FPR) domain, the solution looks like a perfect answer, which is usually not the case [43]. Future research in this area can result in the generation and selection of more suitable cost functions that guide the model search procedure more efficiently.

Table 5-2 also shows that in  $SBCI_{Fully-Automated}$ , almost all bipolar channels are selected. The use of fewer EEG channels is preferable, since it reduces the complicity of the feature space and may also hasten setup of the data recording. Future work explores decreasing the number of bipolar EEG channels used by the  $SBCI_{Fully-Automated}$ .

Scale-up is another important issue that is part of our future research. Scaling can be done in two stages. In stage one, the system detects IC commands (using the proposed method). In the second stage, a second detector differentiates different IC patterns (e.g., related to left/right movements) from each other. This approach has been successfully implemented by another research group [46].

This study is based on the use of executed movements. Future studies investigate the performance of the proposed SBCI system during attempted (imagined) movements

performed by able-bodied individuals. After that, the performance will be evaluated using the data of individuals with motor disabilities. Online testing of the performance of our system is also part of the future works.

## **5.5 Acknowledgements**

This work was supported in part by NSERC under Grant 90278-06 and CIHR under Grant MOP-72711. The authors also would like to thank Mr. Craig Wilson for his valuable comments on this paper. This research has been enabled by the use of WestGrid computing resources, which are funded in part by the Canada Foundation for Innovation, Alberta Innovation and Science, BC Advanced Education, and the participating research institutions.

## 5.6 References

- [1] S. G. Mason and G. E. Birch, "Temporal control paradigms for direct brain interfaces - rethinking the definition of asynchronous and synchronous," in *Proc. HCI Int. Conf.*, Las Vegas, USA, 2005.
- [2] A. Bashashati, S. Mason, R. K. Ward and G. E. Birch, "An improved asynchronous brain interface: making use of the temporal history of the LF-ASD feature vectors", *J. Neural Eng.*, vol. 3, no.2, pp. 87-94, Jun.2006.
- [3] S. G. Mason, A. Bashashati, M. Fatourech, K. F. Navarro and G. E. Birch, "A Comprehensive Survey of Brain Interface Technology Designs", *Annals of Biomedical Engineering*, vol. 35, no. 2, pp. 137-69, Feb 2007.
- [4] M. Fatourech, G. E. Birch and R. K. Ward, "A self-paced brain interface system that uses movement related potentials and changes in the power of brain rhythms", *J. Comput. Neurosci.*, vol.23, no.1, pp.21-37, Aug 2007.
- [5] L. Deecke, B. Grozinger and H. H. Kornhuber, "Voluntary finger movement in man: cerebral potentials and theory", *Biol. Cybern.*, vol. 23, no.2, pp. 99-119, Jul 14.1976.
- [6] M. Hallett, "Movement-related cortical potentials", *Electromyogr. Clin. Neurophysiol.*, vol. 34, no.1, pp. 5-13, Jan-Feb.1994.
- [7] C. Babiloni, F. Carducci, F. Cincotti, P. M. Rossini, C. Neuper, G. Pfurtscheller and F. Babiloni, "Human movement-related potentials vs desynchronization of EEG alpha rhythm: a high-resolution EEG study", *Neuroimage*, vol. 10, no.6, pp. 658-665, Dec.1999.
- [8] G. Pfurtscheller and A. Aranibar, "Event-related cortical desynchronization detected by power measurements of scalp EEG", *Electroencephalogr. Clin. Neurophysiol.*, vol. 42, no.6, pp. 817-826, Jun.1977.
- [9] L. Leocani, C. Toro, P. Manganotti, P. Zhuang and M. Hallett, "Event-related coherence and event-related desynchronization/synchronization in the 10 Hz and 20 Hz EEG during self-paced movements", *Electroencephalogr. Clin. Neurophysiol.*, vol. 104, no.3, pp. 199-206, May.1997.
- [10] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles", *Clin. Neurophysiol.*, vol. 110, no.11, pp. 1842-1857, Nov.1999.
- [11] C. Toro, G. Deuschl, R. Thatcher, S. Sato, C. Kufta and M. Hallett, "Event-related desynchronization and movement-related cortical potentials on the ECoG and EEG", *Electroencephalogr. Clin. Neurophysiol.*, vol. 93, no.5, pp. 380-389, Oct.1994.
- [12] S. Arroyo, R. P. Lesser, B. Gordon, S. Uematsu, D. Jackson and R. Webber, "Functional significance of the mu rhythm of human cortex: an electrophysiologic study with subdural electrodes", *Electroencephalogr. Clin. Neurophysiol.*, vol. 87, no.3, pp. 76-87, Sep.1993.
- [13] L. Narici, V. Pizzella, G. L. Romani, G. Torrioli, R. Traversa and P. M. Rossini, "Evoked alpha- and mu-rhythm in humans: a neuromagnetic study", *Brain Res.*, vol. 520, no.1-2, pp. 222-231, Jun 18.1990.



- [14] B. Feige, R. Kristeva-Feige, S. Rossi, V. Pizzella and P. M. Rossini, "Neuromagnetic study of movement-related changes in rhythmic brain activity", *Brain Res.*, vol. 734, no.1-2, pp. 252-260, Sep 23.1996.
- [15] G. Pfurtscheller, "Central beta rhythm during sensorimotor activities in man", *Electroencephalogr. Clin. Neurophysiol.*, vol. 51, no.3, pp. 253-264, Mar.1981.
- [16] W. Szurhaj, P. Derambure, E. Labyt, F. Cassim, J. L. Bourriez, J. Isnard, J. D. Guieu and F. Mauguiere, "Basic mechanisms of central rhythms reactivity to preparation and execution of a voluntary movement: a stereoelectroencephalographic study", *Clin. Neurophysiol.*, vol. 114, no.1, pp. 107-119, Jan.2003.
- [17] H. Yoon, K. Yang and C. Shahabi, "Feature subset selection and feature ranking for multivariate time series", *IEEE Trans. Knowledge and Data Eng.*, vol. 17, no.9, pp. 1186-1198, 2005.
- [18] S. G. Mason and G. E. Birch, "A brain-controlled switch for asynchronous control applications", *IEEE Trans. Biomed. Eng.*, vol. 47, no.10, pp. 1297-1307, Oct.2000.
- [19] T. Demiralp, J. Yordanova, V. Kolev, A. Ademoglu, M. Devrim and V. J. Samar, "Time-frequency analysis of single-sweep event-related potentials by means of fast wavelet transform", *Brain Lang.*, vol. 66, no.1, pp. 129-145, Jan.1999.
- [20] V. J. Samar, A. Bopardikar, R. Rao and K. Swartz, "Wavelet analysis of neuroelectric waveforms: a conceptual tutorial", *Brain Lang.*, vol. 66, no.1, pp. 7-60, Jan.1999.
- [21] T. Hinterberger, A. Kubler, J. Kaiser, N. Neumann and N. Birbaumer, "A brain-computer interface (BCI) for the locked-in: comparison of different EEG classifications for the thought translation device", *Clin. Neurophysiol.*, vol. 114, no.3, pp. 416-425, 2003.
- [22] B. Graimann, J. E. Huggins, S. P. Levine and G. Pfurtscheller, "Toward a direct brain interface based on human subdural recordings and wavelet-packet analysis", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 954-962, Jun.2004.
- [23] E. L. Glassman, "A wavelet-like filter based on neuron action potentials for analysis of human scalp electroencephalographs", *IEEE Trans. Biomed. Eng.*, vol. 52, no.11, pp. 1851-1862, Nov.2005.
- [24] S. Fukuda, D. Tatsumi, H. Tsujimoto and S. Inokuchi, "Studies of input speed of word inputting system using event-related potential," in the Proc. 20<sup>th</sup> Annual Int. Conf. IEEE Engineering in Medicine and Biology Society, vol.3, pp. 1458-1460, 1998.
- [25] B. H. Jansen, A. Allam, P. Kota, K. Lachance, A. Osho and K. Sundaresan, "An exploratory study of factors affecting single trial P300 detection", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 975-978, Jun. 2004.
- [26] S. G. Mallat, "Multifrequency channel decompositions of images and wavelet models," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 37, no.12, pp. 2091-2106, 1989.
- [27] A. P. Bradley and W. J. Wilson, "On wavelet analysis of auditory evoked potentials", *Clinical Neurophysiology*, vol. 115, no.5, pp. 1114-1128, 2004.
- [28] G. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical applications", *Wavelets and Statistics*, vol. 103, pp. 281-299, 1995.
- [29] L. Citi, R. Poli, C. Cinel and F. Sepulveda, "Feature selection and classification in brain computer interfaces by a genetic algorithm," in Proc. of Genetic and Evolutionary Computation Conference (GECCO'2004), 2004.

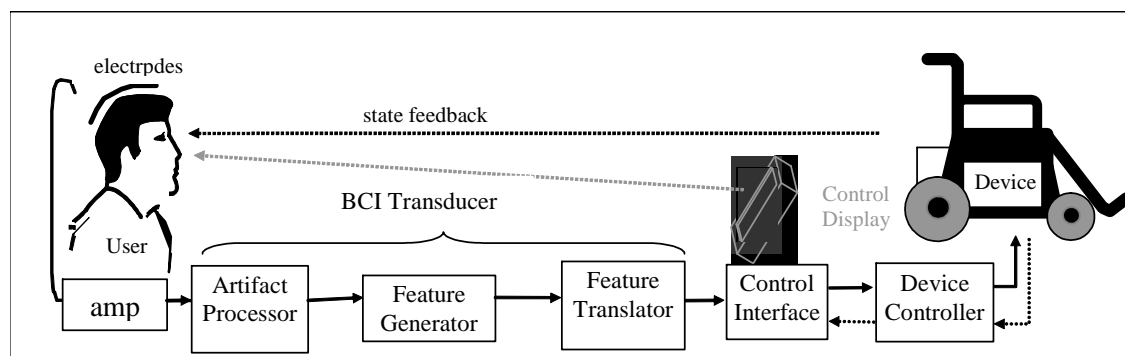
- [30] B. Graimann, J. E. Huggins, S. P. Levine and G. Pfurtscheller, "Detection of ERP and ERD/ERS patterns in single ECoG channels", In Proc.1<sup>st</sup> *IEEE/EMBS Int. Conf. Neural. Eng.*, pp. 614-617, 2003.
- [31] A. Bashashati, M. Fatourehchi, R. K. Ward and G. E. Birch, "User customization of the feature generator of an asynchronous brain interface", *Ann. Biomed. Eng.*, vol. 34, no.6, pp. 1051-1060, Jun.2006.
- [32] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. Wiley New York, 1973.
- [33] A. Blum, A. M. Frieze, R. Kannan and S. Vempala, "A Polynomial-Time Algorithm for Learning Noisy Linear Threshold Functions.", *Algorithmica*, vol. 22, no.1/2, pp. 35-52, 1998.
- [34] J. Dunagan and S. Vempala, "Optimal outlier removal in high-dimensional spaces", *J. Comput. Syst. Sci.*, vol. 68, no.2, pp. 335-373, 2004.
- [35] M. Prastawa, E. Bullitt, S. Ho and G. Gerig, "A brain tumor segmentation framework based on outlier detection", *Med. Image Anal.*, vol. 8, no.3, pp. 275-283, Sep.2004.
- [36] K. R. Muller, C. W. Anderson and G. E. Birch, "Linear and Nonlinear Methods for Brain-Computer Interfaces", *IEEE Trans. Neural. Syst. Rehab. Eng.*, vol. 11, no.2, pp. 165-169, Jun. 2003.
- [37] C. Chang and C. Lin, *LIBSVM: A Library for Support Vector Machines*. 2001,s Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [38] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Publishing Company, Reading, MA: 1989.
- [39] E. W. Sellers, A. Kubler and E. Donchin, "Brain-computer interface research at the University of South Florida Cognitive Psychophysiology Laboratory: the P300 Speller", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no.2, pp. 221-224, Jun.2006.
- [40] A. Kubler, V. K. Mushahwar, L. R. Hochberg and J. P. Donoghue, "BCI Meeting 2005--workshop on clinical issues and applications", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no.2, pp. 131-134, Jun.2006.
- [41] T. Back, D. B. Fogel and T. Michalewicz, *Evolutionary Computation*. Institute of Physics Publishing, Bristol and Philadelphia: 2000.
- [42] G. E. Birch, Z. Bozorgzadeh and S. G. Mason, "Initial on-line evaluations of the LF-ASD brain-computer interface with able-bodied and spinal-cord subjects using imagined voluntary motor potentials", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 10, no.4, pp. 219-224, Dec.2002.
- [43] S. G. Mason, J. Kronegg, J. Huggins, M. Fatourehchi and A. and Schloegl, "Evaluating the performance of self-paced BCI technology", Technical Report, available online at [http://www.bci-info.tugraz.at/Research\\_Info/documents/articles/self\\_paced\\_tech\\_report-2006-05-19.pdf](http://www.bci-info.tugraz.at/Research_Info/documents/articles/self_paced_tech_report-2006-05-19.pdf), 2006.
- [44] J. F. Borisoff, S. G. Mason, A. Bashashati and G. E. Birch, "Brain-computer interface design for asynchronous control applications: improvements to the LF-ASD asynchronous brain switch", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 985-992, Jun.2004.
- [45] E. Yom-Tov and G. F. Inbar, "Detection of Movement-Related Potentials from the Electro-Encephalogram for possible use in a Brain-Computer Interface", *Medical and Biological Engineering and Computing*, vol. 41, no.1, pp. 85-93, Jan. 2003.

- [46] R. Scherer, G. R. Muller, C. Neuper, B. Graimann and G. Pfurtscheller, "An asynchronously controlled EEG-based virtual keyboard: improvement of the spelling rate", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 979-984, Jun.2004.

## CHAPTER 6 EMG AND EOG ARTIFACTS IN BRAIN COMPUTER INTERFACE SYSTEMS: A SURVEY<sup>6</sup>

### 6.1 Introduction

A brain computer interface (BCI) system provides a communication channel between a user's brain and a device the user intends to control. A successful BCI system enables a person to control some aspects of his or her environment (such as lights in the room, a television, a neural prosthesis or a computer) by analyzing his or her brain signals (see Figure 6-1). Specific features of the user's brain activity (or "neurological phenomenon") that relate to their intent to control a device are measured. These features are then translated to control commands that are used to control the device.



**Figure 6-1. The functional model of a BCI system depicting its principle functional components.**

Artifacts are undesired signals that can introduce significant changes in brain signals and ultimately affect the neurological phenomenon. Artifacts are attributed either

<sup>6</sup> A version of this chapter has been published. Fatourechi, M., Bashashati, A., Ward, R. K., and Birch, G. E., "EOG and EMG Artifacts in Brain Interface Systems: a Survey", *Clinical Neurophysiology*, Vol.118, No.3, Mar 2007, pp.480-494 (Invited Paper).

to non-physiological sources (such as 50/60 Hz power-line noise, changes in electrode impedances, etc.) or physiological sources, such as potentials introduced by eye or body movements. Although BCI researchers usually take necessary precautions for handling non-physiological artifacts, physiological artifacts, especially those generated by eye or body movements, remain a significant problem in the design of BCI systems.

In this paper, artifacts caused by eye movements (electrooculography [EOG] artifacts) or muscle movements (electromyography [EMG] artifacts) are reviewed in the context of BCI systems. The aim of the current study is to find out how the BCI community has addressed EMG and EOG artifacts and what outstanding issues still remain. This review is part of a broader attempt to review the field of BCI systems using the framework proposed in [1, 2].

In Section 6.2, we briefly address the current neurological phenomena used in the BCI literature and their associated artifacts. In Section 6.3, we address the existing methods for handling these artifacts, with special focus on EOG and EMG artifacts. In Section 6.4, we present a review of artifact handling methods in the BCI literature. Discussion and conclusions are presented in Section 6.5.

## **6.2 Current neurological phenomena and associated artifacts**

In this section, we briefly review the current neurological phenomena in BCI systems and their associated artifacts.

### **6.2.1 Current neurological phenomena**

Although several strategies exist for sensing the brain signal used for direct communication between the brain and a computer, not all such strategies have been extensively explored. For example, applications of BCI systems based on functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG) are currently limited, as such systems are large, expensive, and require a magnetically shielded environment [3].

The electrical signals of the brain provide suitable representations of the sources of the control signals used in BCI systems. The technology needed for recording the brain's electrical signals can be relatively cheap, especially when these signals are recorded from the scalp. Brain signals typically have fast responses and co-vary with cognitive processes [4]. Hence, the focus of this paper is on the neurological phenomena embedded in the electrical signals of the brain. Our survey showed that artifacts related to direct cortical recording (DCR), which uses microelectrodes that penetrate the brain, have not yet been addressed in the BCI literature. Since this paper is a review study, no critical commentary will be given on artifacts in DCR-based systems. Rather, we focus on BCI systems that use recordings from the surface of the scalp (electroencephalography [EEG]) or from the surface of the brain (electrocorticography [ECoG]) for recording brain activity.

The current neurological phenomena in EEG/ECoG-based BCI systems are as follows:

- Changes in the Brain Rhythms such as Mu, Beta and Gamma rhythms related to a movement (CBR): A voluntary movement results in a circumscribed desynchronization in the Mu and lower Beta bands [4-8]. This desynchronization starts in the contralateral rolandic region about two seconds prior to the onset of a movement, and becomes bilaterally symmetrical immediately before execution of movement [9]. After a voluntary movement, the power in the brain rhythms as well as the amplitude of gamma rhythms increases.
- Movement related potentials (MRPs): MRPs are low-frequency potentials that start about 1- 1.5 seconds before a movement. They have bilateral distribution and present maximum amplitude at the vertex [10-12].
- Other movement related activities (OMRAs): The movement-related activities that do not belong to any of the preceding categories are categorized as OMRA. They are usually not restricted to a particular frequency band or scalp location and usually cover different frequency ranges. They may be a combination of specific and non-specific neurological phenomena.

- Slow cortical potentials (SCPs): SCPs are slow non-movement potential changes generated by the individual. They reflect changes in the cortical polarization of the EEG, lasting from 300 ms up to several seconds. Functionally, a SCP reflects a threshold regularization mechanism for local excitory mobilization [13, 14].
- Cognitive tasks (CTs): Changes in the brain signals as a result of non-movement mental tasks (e.g., mental counting, solving a multiplication problem) are categorized as CTs [4].
- P300: Infrequent or particularly significant auditory, visual or somatosensory stimuli, when interspersed with frequent or routine stimuli, typically evoke a positive peak at about 300 ms after the stimulus is received. This peak is called P300 [4, 15].
- Visual evoked potentials (VEP): VEPs are small changes in the brain signal, generated in response to a visual stimulus such as flashing lights. They display properties whose characteristic depends on the type of the visual stimulus [4].
- Steady-State visual evoked potentials (SSVEP): If a visual stimulus is presented repetitively at a rate of 5-6 Hz or greater, a continuous oscillatory electrical response is elicited in the visual pathways. Such a response is termed SSVEP. The distinction between VEP and SSVEP depends on the repetition rate of the stimulation [16].
- Auditory evoked potentials (AEPs): AEPs are small electrical activity changes that are generated in response to an auditory stimulus (such as clicks or tones).
- Somatosensory evoked potentials (SSEPs): SSEPs are potentials generated in response to the stimulation of somatic sensation.
- Multiple neurological phenomena (MNs): BCI systems based on multiple neurological phenomena use a combination of two or more of the above-mentioned neurological phenomena.

Each neurological phenomenon has unique spatiotemporal characteristics. This fact should be taken into consideration when addressing the presence of artifacts, as a particular neurological phenomenon may be more vulnerable to the presence of certain artifacts.

### 6.2.2 Artifacts in BCI systems

Artifacts are undesirable potentials that contaminate brain signals, and are mostly of non-cerebral origin. Unfortunately, they can modify the shape of a neurological phenomenon used to drive a BCI system. Thus, even cerebral potentials may sometimes be considered as artifacts. For example, in an MRP-based BCI system, a visual evoked potential (VEP) is considered as an artifact. Visual alpha rhythms can also appear as artifacts in a Mu-based BCI system [17]. One problem with such artifacts is that they could mistakenly result in controlling the device [3]. Therefore, there is a need to avoid, reject or remove artifacts from recordings of brain signals.

Artifacts originate from non-physiological as well as physiological sources. Non-physiological artifacts originate from outside the human body (such as 50/60 Hz power-line noise or changes in electrode impedances), and are usually avoided by proper filtering, shielding, etc. For reviews of non-physiological artifacts and the methods of avoiding, rejecting or removing them, the reader can refer to biomedical books [7, 18, 19].

Physiological artifacts arise from a variety of bodily activities. Electrocardiography (ECG) artifacts are caused by heart beats and may introduce a rhythmic activity into the EEG signal. Respiration can also cause artifacts by introducing a rhythmic activity that is synchronized with the body's respiratory movements. Skin responses such as sweating may alter the impedance of electrodes and cause artifacts in the EEG signals [20]. The two physiological artifacts that have been most examined in BCI studies, however, are ocular (EOG) and muscle (EMG) artifacts.

EOG artifacts are generally high-amplitude patterns in the brain signal caused by blinking of the eyes, or low-frequency patterns caused by movements (such as rolling) of the eyes [21]. EOG activity has a wide frequency range, being maximal at frequencies below 4Hz, and is most prominent over the anterior head regions [17].

EMG activity (movement of the head, body, jaw or tongue) can cause large disturbances in the brain signal. EMG activity has a wide frequency range, being maximal at frequencies higher than 30 Hz [17, 21]. Difficult tasks may cause an increase in EMG activity related to the movement of facial muscles [22, 23].



A number of studies have shown that EOG and EMG activities may generate artifacts that affect the neurological phenomena used in a BCI system [24, 25]. For example, [25] demonstrated that brain rhythms are contaminated with EMG artifacts during the early training sessions of a BCI system that used Mu and Beta rhythms as sources of control.

Physiological artifacts such as EOG and EMG artifacts are much more challenging to handle than non-physiological ones. Moreover, controlling them during signal acquisition is not easy. There are different ways of handling artifacts in BCI systems. In Section 3, we examine the reported methods for handling EOG and EMG artifacts, as they are among the most important sources of contamination in BCI systems.

### **6.3 Methods of handling artifacts**

In this section, we briefly address methods of handling artifacts. Our focus throughout this section will be on EOG and EMG artifacts.

#### **6.3.1 Artifact avoidance**

The first step in handling artifacts is to avoid their occurrence by issuing proper instructions to users. For example, users are instructed to avoid blinking or moving their body during the experiments.

Instructing individuals to avoid generating artifacts during data collection has the advantage of being the least computationally demanding among the artifact handling methods, since it is assumed that no artifact is present in the signal (or that the presence of artifacts is minimal). However, it has several drawbacks. First, since many physiological signals, such as the heart beats, are involuntary, artifacts will always be present in brain signals. Even in the case of EOG and EMG activities, it is not easy to control eye and other movement activities during the process of data recording. Second, the occurrence of ocular and muscle activity during an online operation of any BCI system is not avoidable. Third, the collection of a sufficient amount of data without artifacts may be difficult, especially in cases where a user has a neurological disability [26]. Finally, avoiding artifacts may introduce an additional cognitive task for the

individual. For example, it has been shown that refraining from eye blinking results in changes in the amplitude of some evoked potentials [27, 28].

### **6.3.2 Artifact rejection**

Artifact rejection refers to the process of rejecting the trials affected by artifacts. It is perhaps the simplest way of dealing with brain signals contaminated with artifacts. It has some important advantages over the “artifact avoidance” approach. For example, it would be easier for individuals to participate in the experiments and perform the required tasks, especially those individuals with motor disabilities. Also, the “secondary” cognitive task, resulting from an individual trying to avoid generating a particular artifact, will not be present in the EEG signal.

”Artifact rejection” is usually done by visually inspecting the EEG or the artifact signals, or by using an automatic detection method [29].

#### **Manual rejection**

Manual rejection of epochs contaminated with artifacts is a common practice in the BCI field. Trials are visually checked by an expert, and those that are contaminated with artifacts are removed from the analysis.

Similar to “artifact avoidance”, manual rejection also has the advantage of not being computationally demanding, as it is assumed that a human expert has identified all the artifact-contaminated epochs and removed them from the analysis. On the other hand, there are many disadvantages in using “manual rejection”. First, “manual rejection” comes at the cost of intensive human labor, especially if the study involves a large number of individuals or a large amount of recorded data. Second, the process of selecting the artifact-free trials may become subjective. It has been argued that because of the selection bias, the sample trials that are artifact-free may not be representative of the entire population of the trials [29]. Third, in the case of offline analysis, the rejection of artifact-contaminated trials, may lead to a substantial loss of data. This may become a huge drawback, especially in the case of subjects with motor disabilities, where offline data recording is not as convenient as it is for able-bodied individuals.

## Automatic rejection

In the “automatic rejection”, the BCI system automatically discards the epochs of brain signals that are contaminated with particular artifacts. This procedure is commonly carried out in offline investigations.

Automatic rejection of epochs can be done in the following two ways:

***Rejection using the EOG (EMG) signal:*** When one of the characteristics of the EOG (EMG) signal in an epoch exceeds a pre-determined threshold, the epoch is considered as artifact-contaminated and is automatically rejected.

***Rejection using the EEG signal:*** This rejection methodology is similar to the above; only the EEG signal is used instead of the EOG (EMG) signal. This approach has the advantage of being independent of the EOG (EMG) signal, and is useful if the EOG (EMG) signal is not recorded during data collection.

An advantage of the “automatic rejection” approach over that of “manual rejection” is that it is less labor intensive. However, automatic rejection still suffers from sampling bias and loss of valuable data [30, 31]. In the case of EOG artifacts, the automatic rejection approach also does not allow the rejection of contaminated trials when EOG amplitude is small [32, 33].

Two issues need to be addressed for the BCI systems that reject artifacts:

1) Because of the vast number of artifacts that exist in BCI systems (eye blinking, eye movements, movements of different parts of the body, breathing, etc.), not all the artifact-contaminated trials can be rejected. Usually only the epochs with a strong presence of artifacts are excluded from the analysis. Therefore, the so-called “clean” data are unfortunately not free of artifacts.

2) The rejection of artifact-contaminated data during an offline analysis may generate “cleaner” data, however for online real-time applications of a BCI system, this may pose a huge drawback. In online applications, artifacts are unavoidable. If artifacts are rejected during the offline analysis, the same rejection mechanism can be used to reject them during the online analysis. The only problem is that during the specific time periods when artifact-contaminated signals are rejected, the system is unreachable and

cannot be used for controlling the device. On the other hand, if artifacts are rejected during the offline analysis and the design of the BCI system is not robust to artifacts, false responses may occur in an online application due to the presence of artifacts.

### **6.3.3 Artifact removal**

Artifact removal is the process of identifying and removing artifacts from brain signals. An artifact-removal method should be able to remove the artifacts as well as keep the related neurological phenomenon intact. Common methods for removing the artifacts in EEG signals are as follows.

#### **Linear filtering**

Linear filtering is useful for removing artifacts located in certain frequency bands that do not overlap with those of the neurological phenomena of interest [34, 35]. For example, low-pass filtering can be used to remove EMG artifacts and high-pass filtering can be used to remove EOG artifacts. Linear filtering was commonly used in early clinical studies to remove artifacts in EEG signals [36, 37].

The advantage of using filtering is its simplicity. Also the information from the EOG signal is not needed to remove the artifacts. This method, however, fails when the neurological phenomenon of interest and the EMG or EOG artifacts overlap or lie in the same frequency band [38]. A look at the frequency range of neurological phenomena used in BCI systems unfortunately shows that this is usually the case. As a result, a simple filtering approach cannot remove EMG or EOG artifacts without removing a portion of the neurological phenomenon. More specifically, since EOG artifacts generally consist of low-frequency components, using a high-pass filter will remove most of the artifacts. Such methods are successful to some extent in BCI systems that use features extracted from high-frequency components of the EEG (e.g., Mu and Beta rhythm). However, for BCI systems that depend on low-frequency neurological phenomena (such as MRPs), these methods are not as desirable, since these neurological phenomena may lie in the same frequency range as that of the EOG artifacts.

In the case of removing EMG artifacts from EEG signals, filtering specific frequency bands of the EEG can be used to reduce the EMG activity. Since artifacts

generated by EMG activity generally consist of high-frequency components, using a low-pass filter may remove most of these artifacts. Again, such methods may be successful to some extent for BCI systems that rely on low-frequency components (e.g., MRPs), but they cannot be effective for BCI systems that use neurological phenomena with high-frequency content (such as Beta rhythms).

### **Linear combination and regression**

Using a linear combination of the EOG-contaminated EEG signal and the EOG signal is the most common technique for removing ocular artifacts from EEG signals [32]. The linear combination technique is based on the following model [29]:

$$EEG_{nc}^i(t) = EEG_{ac}^i(t) - K.EOG(t) \quad (6-1)$$

where  $EEG_{ac}^i(t)$  is the EOG-contaminated EEG signal of channel  $i$ ,  $EEG_{nc}^i(t)$  is the non-contaminated EEG signal of channel  $i$ ,  $EOG(t)$  is the EOG signal and  $K$  is an unknown constant.

Based on the model in (6-1), a fraction of the EOG signal(s) should be subtracted from  $EEG_{ac}^i(t)$  to generate  $EEG_{nc}^i(t)$ . The question then is how to estimate the value of  $K$ . A popular method that aims at minimizing the effect of noise on the estimates employs linear regression using least square criterion to estimate the value of  $K$  [39].

A question as arises as to whether the value of  $K$  should be calculated separately for each type of EOG artifact [29] and for the different frequencies of a particular EOG artifact [40]. Both cases have been discussed extensively in the literature, but some papers have shown that similar results are obtained by the standard linear regression method [32].

One problem with using the above linear combination and regression approach is that the EOG signal to be subtracted from the EEG signal is also contaminated with the EEG signal. However, subtracting the EOG signal may also remove part of the EEG signal. Nevertheless, [41] argues in favor of using correction methods over rejection methods.

This problem becomes more challenging for EMG artifacts, since they have no reference channels [20] and applying regression using signals from multiple muscle groups requires multiple reference channels [42]. The papers use regression techniques for the removal of head-movement artifacts (Bayliss and Ballard ,1999;Bayliss and Ballard ,2000a;Bayliss and Ballard ,2000b), but they do not explain how this was done. The validity of the results is also not verified.

### **Blind source separation (BSS)**

BSS techniques separate the EEG signals into components that “build” the EEG signals. They identify the components that are attributed to artifacts and reconstruct the EEG signal without these components (for a review, see [43]). Among the BSS methods, Independent Component Analysis (ICA) is more widely used. ICA is a method that blindly separates mixtures of independent source signals, forcing the components to be independent. It has been widely applied to remove ocular artifacts from EEG signals [44-46]. Preliminary studies have shown that ICA increases the strength of motor-related signal components in the Mu rhythms, and is thus useful for removing artifacts in BCI systems [47].

Although BSS methods have been used to remove EOG artifacts in EEG clinical studies, only a few studies have used BSS methods to remove EMG artifacts [42, 48, 49].

One advantage of using BSS methods such as ICA is that they do not rely on the availability of reference artifacts for separating the artifacts from the EOG signals[36]. One disadvantage of ICA, along with other BSS techniques, is that they usually need prior visual inspection to identify artifact components [45, 46]. However, some automatic methods have been proposed [50-52].

### **Principal component analysis (PCA)**

PCA uses the eigenvectors of the covariance matrix of the signal to transform the data to a new coordinate system and to find the projection of the input data with greater variances. The components of the signal are then extracted by projecting the signal onto

the eigenvectors. PCA has been shown to be an effective method for removing ocular artifacts from EEG signals [53-55].

One advantage of PCA is the requirement that artifacts are uncorrelated with the EEG signal. This is a weaker requirement than the independency requirement of ICA. On the other hand, it has been shown that PCA cannot completely separate eye-movement artifacts from the EEG signal, especially when they have comparable amplitudes [53]. PCA also does not necessarily decompose similar EEG features into the same components when applied to different epochs [53].

#### **Other methods**

Other methods have also been proposed for removing artifacts from EEG signals in clinical studies with varying degrees of success. Examples include the wavelet transform [56, 57], nonlinear adaptive filtering [58, 59] and source dipole analysis (SDA) [60]. However, their application in BCI systems has so far been limited.

### **6.4 Literature survey**

In this section, we review how artifacts are addressed in the BCI literature. Since it is expected that proper measures for avoiding non-physiological artifacts were taken during the BCI experiments, the focus in this section will be on EMG and EOG artifacts. These are the physiological artifacts that have been addressed in detail in the BCI literature.

The following criteria were used for selecting the papers reviewed in this study:

- Since the focus of this paper concerns the design and evaluation of BCI systems, a study that does not include a BCI transducer (i.e., a BCI transducer with the general structure depicted in Figure 6-1) is not considered. In searching the literature, the keywords “brain interface”, “brain-computer interface”, “brain-machine interface”, “direct brain interface”, “direct brain connection”, “direct neural control” and “brain-actuated control” were used.
- Only papers published in English in refereed international journals or conference proceedings prior to January 2006 were considered for the analysis.

- Only BCI systems that use neurological phenomena embedded in the EEG or ECoG signals were chosen for this study (approximately 250 papers).

For each neurological phenomenon, we grouped the artifacts into one of five categories: “Not Mentioned”, “No Rejection/Removal”, “Manual Rejection”, “Automatic Rejection” and “Automatic Removal”. The “*Not Mentioned*” category signifies cases where the authors did not explicitly mention whether or not they dealt with EMG or EOG artifacts in their BCI designs. “*No Rejection/Removal*” refers to cases where the authors acknowledged the presence of artifacts in their data, but did not apply any method to handle them.

Table 6-1 lists the methods of handling EOG and EMG artifacts. Each neurological phenomenon is highlighted by a gray color. The white rows below each neurological phenomenon show the artifacts, methods of handling them and the bibliography. For each of the artifact handling categories and for each of the neurological phenomena (or groups of them), Figure 6-2 and Figure 6-3 show the number of published papers on how EMG and EOG artifacts were handled, respectively. Table 6-2 and Table 6-3 identify the methods used for the automatic rejection of EOG and EMG artifacts (along with references). Table 6-4 and Table 6-5 list the methods used for the automatic removal of EOG and EMG artifacts (along with references).



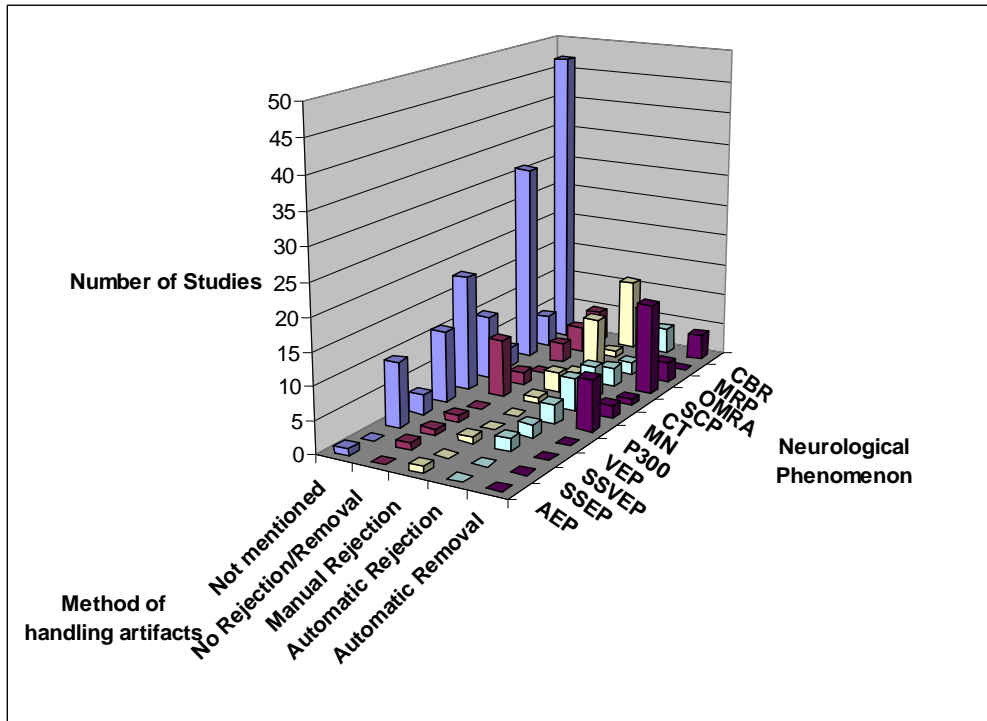


Figure 6-2. The number of papers published on different methods of handling EOG artifacts in BCI studies.

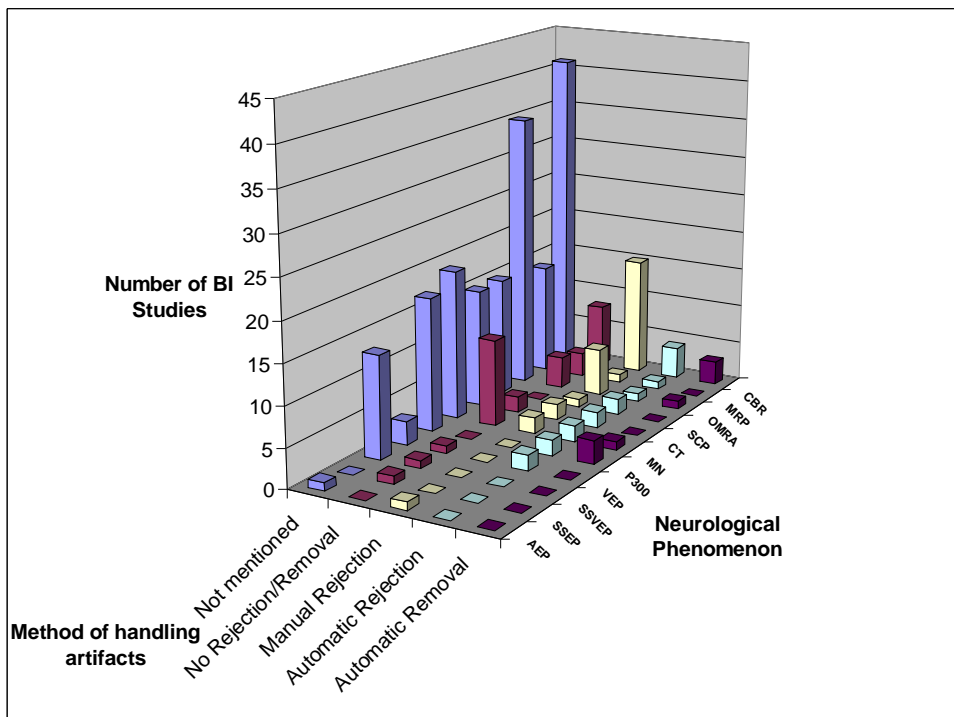


Figure 6-3. The number of papers published on different methods of handling EMG artifacts in BCI studies.

**Table 6-1. Methods of handling artifacts in BCI literature**

<b>Neurological phenomenon</b>		
<b>CBR</b>		
Artifact	Method	Bibliography
EOG	Not mentioned	[17, 25, 61-105]
	No Rejection/Removal	[106-110]
	Manual Rejection	[30, 31, 111-119]
	Automatic Rejection	[120-123]
	Automatic Removal	[124-127]
EMG	Not mentioned	[17, 61-72, 74-85, 88-90, 92, 94-100, 102, 103, 105, 116, 125]
	No Rejection/Removal	[25, 93, 101, 106-110]
	Manual Rejection	[30, 31, 73, 86, 87, 91, 104, 111-115, 117-119]
	Automatic Rejection	[120-123]
	Automatic Removal	[124, 126, 127]
<b>MRP</b>		
EOG	Not mentioned	[128-132]
	No Rejection/Removal	[108, 133-135]
	Manual Rejection	[136]
	Automatic Rejection	[123, 137-144]
EMG	Not mentioned	[128-132, 135, 137-144]
	No Rejection/Removal	[108, 133, 134]
	Manual Rejection	[136]
	Automatic Rejection	[123]
<b>OMRA</b>		
EOG	Not Mentioned	[69, 71, 92, 95, 96, 99, 145-169]
	No Rejection/Removal	[106, 170, 171]
	Manual Rejection	[111, 171-177]
	Automatic Rejection	[178, 179]
	Automatic Removal	[125, 126, 180]
EMG	Not Mentioned	[69, 71, 92, 95, 99, 125, 145-169, 176-178, 180]
	No Rejection/Removal	[96, 106, 170, 171]
	Manual Rejection	[111, 171-175]
	Auto rejection	[179]
	Automatic Removal	[126]
<b>SCP</b>		
EOG	Not Mentioned	[181-183]

Neurological phenomenon		
	Manual Rejection	[151]
	Automatic Rejection	[184-186]
	Automatic Removal	[13, 151, 184-195]
EMG	Not Mentioned	[13, 151, 181-183, 186-195]
	Manual Rejection	[151]
	Automatic Rejection	[184, 185]
CT		
EOG	Not mentioned	[196-205]
	No Rejection/Removal	[206, 207]
	Manual Rejection	[173, 208, 209]
	Automatic Rejection	[210-214]
	Automatic Removal	[215, 216]
EMG	Not mentioned	[196-205, 208, 211, 213-216]
	No Rejection/Removal	[206, 207]
	Manual Rejection	[173, 209]
	Automatic Rejection	[210, 212]
MN(CBR and MRP)		
EOG	Not mentioned	[217]
	No Rejection/Removal	[108, 218, 219]
EMG	Not mentioned	[217]
	No Rejection/Removal	[108, 218, 219]
MN(CBR and OMRA)		
EOG	Not mentioned	[153, 211, 220-222]
EMG	Not mentioned	[153, 211, 220-222]
MN(CBR,MRP, and OMRA)		
EOG	No Rejection/Removal	[223]
EMG	No Rejection/Removal	[223]
MN ( CBR and SCP)		
EOG	Not mentioned	[224]
	Automatic Removal	[225]
EMG	Not mentioned	[224, 225]
MN(CBR and CT)		
EOG	Not mentioned	[226-229]
	No Rejection/Removal	[230-233]
	Automatic Rejection	[234]
EMG	Not mentioned	[227-229]
	No Rejection/Removal	[230-234]
	Manual Rejection	[226]
MN(OMRA and CT)		
EOG	Not mentioned	[235-241]
	No Rejection/Removal	[233]
	Manual Rejection	[242]
	Automatic Rejection	[234, 243-245]
EMG	Not mentioned	[235-241, 245]
	No Rejection/Removal	[233, 234]
	Manual rejection	[242]
	Automatic Rejection	[243, 244]
MN(CBR, OMRA and CT)		
EOG	Automatic Removal	[246]
EMG	Not mentioned	[246]
P300		
EOG	Not Mentioned	[130, 181, 247-255]
	Automatic Rejection	[256, 257]
	Automatic Removal	[258-265]

Neurological phenomenon		
EMG	Not mentioned	[130, 181, 247-256, 258, 259, 262, 264, 265]
	Automatic Rejection	[257]
	Automatic Removal	[260, 261, 263]
VEP		
EOG	Not mentioned	[266-268]
	No Rejection/Removal	[269]
	Automatic Rejection	[270, 271]
EMG	Not mentioned	[267, 268, 270]
	No Rejection/Removal	[269]
	Automatic Rejection	[266, 271]
SSVEP		
EOG	Not mentioned	[16, 272-280]
	No Rejection/Removal	[281]
	Manual Rejection	[282]
	Automatic Rejection	[283, 284]
EMG	Not mentioned	[16, 272-275, 277-284]
	No Rejection/Removal	[276]
AEP		
EOG	Not mentioned	[267]
	Manual Rejection	[285]
EMG	Not mentioned	[267]
	Manual Rejection	[285]
SSEP		
EOG	No Rejection/Removal	[286]
EMG	No Rejection/Removal	[286]

**Table 6-2. Methods of automatic EOG rejection in BCI studies.**

Neurological phenomenon	Rejection method	Bibliography
CBR	Threshold on EEG Amplitude	[120, 121]
	Threshold on EEG Power Spectra	[122]
	Not mentioned	[123]
MRP	Threshold on EOG amplitude	[137-144]
	Not mentioned	[123]
OMRA	Threshold on EOG Amplitude	[178, 179]
SCP	Threshold on the Amplitude of(EOG-EEG)	[184-186]
CT	Threshold on EOG Amplitude	[211, 214]
	Not mentioned	[210, 212, 213]
CT and OMRA	Threshold on EEG power	[243, 244]
	Threshold on EEG Amplitude	[245]
	Not mentioned	[234]
CT and CBR	Not mentioned	[234]
P300	Threshold on EOG Amplitude	[256]
	Not mentioned	[257]
VEP	Threshold on EOG Amplitude	[270]
	Threshold on EEG amplitude	[271]
SSVEP	Threshold on EOG Amplitude	[283, 284]

**Table 6-3. Methods of automatic EMG rejection in BCI studies.**

Neurological phenomenon	Rejection method	Bibliography
CBR	Threshold on EEG Amplitude	[120, 121]
	Threshold on EEG Power	[122]
	Not mentioned	[123]
MRP	Not mentioned	[123]
OMRA	Threshold on EMG amplitude	[179]
SCP	Threshold on EEG Amplitude	[184, 185]
CT	Not mentioned	[210, 212]
CT and OMRA	Threshold on EEG power	[243, 244]
P300	Not mentioned	[257]
VEP	Threshold on EEG Amplitude	[266, 271]

**Table 6-4. Methods of automatic EOG removal in BCI studies.**

Neurological phenomenon	Removal method	Bibliography
CBR	Filter (no details)	[124]
	Neural adaptive filter	[125]
	PCA	[126]
	ICA	[127]
OMRA	Linear Combination and Regression	[180]
	Neural adaptive filter	[125]
	PCA	[126]
SCP	Linear Combination and Regression	[13, 184-195]
CT	Linear Combination and Regression	[215]
	ICA	[216]
CT, CBR and OMRA	Adaptive noise canceling using neural networks	[246]
SCP and CBR	Linear Combination and Regression	[225]
P300	Not mentioned	[265]
	Blind Source Separation	[259]
	Linear Combination and Regression	[258-264]

**Table 6-5. Methods of automatic EMG removal in BCI studies.**

Neurological phenomenon	Removal method	Bibliography
CBR	Filter (no details)	[124]
	PCA	[126]
	ICA	[127]
OMRA	PCA	[126]
P300	Linear Combination and Regression	[260, 261, 263]

Based on the results presented in Table 6-1 to Table 6-5 and Figure 6-2 and Figure 6-3, the following observations about the methods of handling artifacts in BCI systems were made:

#### **6.4.1 EOG artifacts**

1. More than half (53.7%) of the BCI studies considered, do not mention whether or not they handle EOG artifacts, 10.0% do not remove EOG artifacts, 10.4% manually reject them, 13.5% use an automatic rejection method and 12.4% use an automatic removal method to handle EOG artifacts.

2. Among the 13.5% of the studies that used automatic EOG artifact rejection methods, nearly half (45.7%) reject trials when the EOG amplitude reaches a certain pre-defined threshold, 14.3% employ this rejection strategy but use the EEG amplitude instead of the EOG amplitude for rejecting the contaminated trials, 22.9% do not mention the rejection method and 17.1% use other EOG rejection methods.

3. Among the 12.4% of the studies that use EOG artifact removal methods, around 69.7% use a linear method of combination of EEG and EOG signals, 9.1% use BSS techniques, 6.1% use PCA, 3.0% use linear filtering methods , 9.1% use other EOG removal methods and one paper does not mention the details of its automatic artifact removal method.

#### **6.4.2 EMG artifacts**

1. Approximately 67.6% of the BCI studies considered, do not mention whether or not they handle EMG artifacts, 12.1% do not remove EMG artifacts, 10.9% manually

remove them, 6.2% use automatic rejection methods and 3.2% use an automatic method for removal of EMG artifacts.

2. Among the 6.2% of the papers that use EMG artifact rejection, close to half (43.8%) reject trials when the EEG amplitude reaches a certain pre-defined threshold, 6.2% employ this rejection strategy but use the EMG amplitude instead of the EEG amplitude, 18.7% reject trials when the EEG power reaches a certain threshold, and the rest (31.2%) do not specify their rejection method.

3. Only seven of all studies reviewed, use an automatic method for removal of EMG artifacts. Two papers use PCA, one uses ICA, one uses linear filtering and three papers use regression.

## **6.5 Discussion and conclusions**

In this paper, we have addressed EOG and EMG artifacts associated with neurological phenomena in EEG/ECOG-based brain computer interface (BCI) systems. We have also discussed the common methods of handling them in BCI systems and we presented a detailed review as to how BCI studies have addressed this issue.

Our survey of the BCI studies (published until January 2006) shows that most BCI papers do not report whether or not they considered EMG and/or EOG artifacts in their analysis. The number of studies that do not report these artifacts in their systems is higher for EMG artifacts (65.9%) compared with EOG artifacts (55.6%). This is an important issue, since offline analysis methods that do not account for physiological artifacts may probably face some problems when tested during an online study. As a result, it is important that BCI researchers pay more attention to this very important issue and address the method that they have employed for handling artifacts.

A number of BCI studies state that EMG activity will not be present in the EEG signal when the EEG signal is analyzed before a movement has occurred [165]. This argument may not be valid for BCI systems. This is because peripheral changes such as EMG tension can affect the EEG signal, even though the amount by which the EEG signal is affected remains unclear [192]. It is pointed out in [192] that even when the individuals are very restricted, they still preserve motor control over some muscle groups.

Although the activities of several muscle groups are monitored in BCI studies, there remain some muscles whose activities are not recorded.

The BCI systems that employ “manual rejection” of EOG and EMG artifacts should also consider the fact that “manual rejection” is only a preliminary step in the design of a BCI system. “Manual rejection” can only be used for offline analysis. In order for a particular BCI system to be able to work in an online fashion, a scheme for handling artifacts should be incorporated. Requesting the individuals to avoid artifacts should be only considered as a temporary solution. In a practical application, EMG and EOG artifacts do happen, so methods of handling these artifacts during an online experiment should be investigated.

One solution for handling artifacts, which is not explored well in the BCI studies, is to design a BCI that is robust in the presence of artifacts. If such a BCI is designed, then the need for having a method of handling artifacts will be minimized. Our literature survey showed that 10.0% of BCI papers reported that they did not remove or reject EOG artifacts, and 12.1% did not reject EMG artifacts. Although one reason for not removing/rejecting these artifacts may be that these BCI designs are robust to them, the performance of such BCI systems when contaminated by artifacts is not well explored in the BCI literature. Future work for such BCI systems should also include the analysis of the robustness of the performance of the method in the presence of artifacts.

Another solution that has not been explored well in the BCI literature is that of using more than one neurological phenomenon may lead to increasing the robustness to the occurrence of artifacts [218]. Since EOG artifacts mostly affect the low-frequency components of the EEG signals, BCI systems that use low-frequency ERPs, such as MRP and SCP are mostly affected by EOG artifacts. EMG artifacts on the other hand, mainly affect the high-frequency components of the EEG signals, hence BCI systems that use high-frequency ERPs, such as Mu and Beta rhythms are mostly affected. Thus, it can be concluded that a BCI system that uses multiple neurological phenomena from both low and high frequency bands, may become more robust to the presence of artifacts. This promising research area has remained unexplored but it needs attention and further research.



Whatever solution is proposed for handling the artifacts, it is necessary to show that the proposed BCI system can work well in an online manner and it is not the artifacts that are controlling the BCI system. Recently, this issue concerning in whether or not a few particular BCI systems, an artifact (and not the neurological phenomenon) is the source of control in the BCI system has been discussed [25, 287] .

Finally, the comparison between different artifact removal methods is not straightforward because it is generally not clear what a correct EEG waveform should look like. It is therefore important to note that no analytical method for validating artifact removal algorithms is available at this time [32]. This means that to validate a certain artifact removal method, visual inspection of the cleaned signal is used. Development of criteria for validation of artifact removal methods is thus clearly necessary and need to be explored in the future studies.

## 6.6 References

- [1] S. G. Mason and G. E. Birch, "A general framework for brain-computer interface design", *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 11, no.1, pp. 70-85, Mar. 2003.
- [2] S. G. Mason, A. Bashashati, M. Fatourehchi, K. F. Navarro and G. E. Birch, "A Comprehensive Survey of Brain Interface Technology Designs", *Annals of Biomed. Eng.*, vol. 35, no. 2, pp. 137-69, Feb 2007.
- [3] T. M. Vaughan, W. J. Heetderks, L. J. Trejo, W. Z. Rymer, M. Weinrich, M. M. Moore, A. Kubler, B. H. Dobkin, N. Birbaumer, E. Donchin, E. W. Wolpaw and J. R. Wolpaw, "Brain-computer interface technology: a review of the Second International Meeting", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no.2, pp. 94-109, Jun.2003.
- [4] A. Kubler, B. Kotchoubey, J. Kaiser, J. R. Wolpaw and N. Birbaumer, "Brain-Computer Communication: Unlocking the Locked In", *Psych Bulletin*, vol. 127, no.3, pp. 358-375, May. 2001.
- [5] G. Pfurtscheller and A. Aranibar, "Event-related cortical desynchronization detected by power measurements of scalp EEG", *Electroencephalogr. Clin. Neurophysiol.*, vol. 42, no.6, pp. 817-826, Jun.1977.
- [6] H. Jasper and W. Penfield, "Electrocortigrams in man: Effect of voluntary movement upon the electrical activity of the precentral gyrus", *Arch Psychiat Nervenkr*, vol. 183, pp. 163-174, 1949.
- [7] E. Niedermeyer and F. L. Da Silva, *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins, 2004.
- [8] J. W. Kozelka and T. A. Pedley, "Beta and mu rhythms", *J. Clin. Neurophysiol.*, vol. 7, no.2, pp. 191-207, Apr.1990.
- [9] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles", *Clin. Neurophysiol.*, vol. 110, no.11, pp. 1842-1857, Nov.1999.
- [10] C. Babiloni, F. Carducci, F. Cincotti, P. M. Rossini, C. Neuper, G. Pfurtscheller and F. Babiloni, "Human movement-related potentials vs desynchronization of EEG alpha rhythm: a high-resolution EEG study", *Neuroimage*, vol. 10, no.6, pp. 658-665, Dec.1999.
- [11] L. Deecke, B. Grozinger and H. H. Kornhuber, "Voluntary finger movement in man: cerebral potentials and theory", *Biol. Cybern.*, vol. 23, no.2, pp. 99-119, Jul .1976.
- [12] M. Hallett, "Movement-related cortical potentials", *Electromyogr. Clin. Neurophysiol.*, vol. 34, no.1, pp. 5-13, Jan-Feb.1994.
- [13] N. Neumann, A. Kubler, J. Kaiser, T. Hinterberger and N. Birbaumer, "Conscious perception of brain states: mental strategies for brain-computer communication", *Neuropsychologia*, vol. 41, no.8, pp. 1028-1036, 2003.
- [14] 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, no.6, pp. 767-791, Jun.2002.

- [15] B. Z. Allison and J. A. Pineda, "ERPs evoked by different matrix sizes: Implications for a brain computer interface (BCI) system", *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 11, no.2, pp. 110-113, Jun.2003.
- [16] X. Gao, D. Xu, M. Cheng and S. Gao, "A BCI-based environmental controller for the motion-disabled", *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 11, no.2, pp. 137-140, Jun.2003.
- [17] 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, no.3, pp. 386-394, Sep. 1997.
- [18] J. Moore and G. Zouridakis, *Biomedical Technology and Devices Handbook*. CRC Press, 2004.
- [19] B. J. Fisch, *Fisch and Spehlmann's Eeg Primer: Basic Principles of Digital and Analog Eeg*. Elsevier Publishing Company, 2000.
- [20] J. S. Barlow, "Artifact processing (rejection and minimization) in EEG data processing", *Handbook of Electroencephalography and Clinical Neurophysiology (Revised Series Ed.)*, Amsterdam: Elsevier, vol.2., pp.15-62, 1986.
- [21] P. Anderer, S. Roberts, A. Schlogl, G. Gruber, G. Klosch, W. Herrmann, P. Rappelsberger, O. Filz, M. J. Barbanoj, G. Dorffner and B. Saletu, "Artifact processing in computerized analysis of sleep EEG - a review", *Neuropsychobiology*, vol. 40, no.3, pp. 150-157, Sep.1999.
- [22] W. Waterink and A. van Boxtel, "Facial and jaw-elevator EMG activity in relation to changes in performance level during a sustained information processing task", *Biol. Psychol.*, vol. 37, no.3, pp. 183-198, Jul.1994.
- [23] B. H. Cohen, R. J. Davidson, J. A. Senulis, C. D. Saron and D. R. Weisman, "Muscle tension patterns during auditory attention", *Biol. Psychol.*, vol. 33, no.2-3, pp. 133-156, Jul.1992.
- [24] 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, no.9, pp. 1580-1593, Sep.2003.
- [25] D. J. McFarland, W. A. Sarnacki, T. M. Vaughan and J. R. Wolpaw, "Brain-computer interface (BCI) operation: signal and noise during early training sessions", *Clin. Neurophysiol.*, vol. 116, no.1, pp. 56-62, Jan.2005.
- [26] R. N. Vigario, "Extraction of ocular artefacts from EEG using independent component analysis", *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no.3, pp. 395-404, Sep. 1997.
- [27] R. Verleger, "The instruction to refrain from blinking affects auditory P3 and N1 amplitudes", *Electroencephalogr. Clin. Neurophysiol.*, vol. 78, no.3, pp. 240-251, Mar.1991.
- [28] C. J. Ochoa and J. Polich, "P300 and blink instructions", *Clin. Neurophysiol.*, vol. 111, no.1, pp. 93-98, Jan.2000.
- [29] G. Gratton, "Dealing with artifacts: The EOG contamination of the event-related brain potential", *Behavior Research Methods, Instruments, & Computers*, vol. 30, no.1, pp. 44-53, 1998.
- [30] H. Ramoser, J. Muller-Gerking and G. Pfurtscheller, "Optimal Spatial Filtering of Single Trial EEG During Imagined Hand Movement", *IEEE Trans. Rehab. Eng.*, vol. 8, no.4, pp. 441-446, Dec.2000.

- [31] J. Millan, M. Franze, J. Mourino, F. Cincotti and F. Babiloni, "Relevant EEG features for the classification of spontaneous motor-related tasks", *Biol. Cybern.*, vol. 86, no.2, pp. 89-95, Feb.2002.
- [32] R. J. Croft and R. J. Barry, "Removal of ocular artifact from the EEG: a review", *Neurophysiol. Clin.*, vol. 30, no.1, pp. 5-19, Feb.2000.
- [33] V. Rowland, "Cortical steady potential (direct current potential) in reinforcement and learning", *Progress in Physiological Psychology*, vol. 2, pp. 1-77, 1968.
- [34] J. S. Barlow, "EMG artifact minimization during clinical EEG recordings by special analog filtering", *Electroencephalogr. Clin. Neurophysiol.*, vol. 58, no.2, pp. 161-174, Aug.1984.
- [35] J. R. Ives and D. L. Schomer, "A 6-pole filter for improving the readability of muscle contaminated EEGs", *Electroencephalogr. Clin. Neurophysiol.*, vol. 69, no.5, pp. 486-490, May.1988.
- [36] W. Zhou and J. Gotman, "Removing Eye-movement Artifacts from the EEG during the Intracarotid Amobarbital Procedure", *Epilepsia*, vol. 46, no.3, pp. 409-414, 2005.
- [37] J. Gotman, D. R. Skuce, C. J. Thompson, P. Gloor, J. R. Ives and W. F. Ray, "Clinical applications of spectral analysis and extraction of features from electroencephalograms with slow waves in adult patients", *Electroencephalogr. Clin. Neurophysiol.*, vol. 35, no.3, pp. 225-235, Sep.1973.
- [38] N. A. de Beer, M. van de Velde and P. J. Cluitmans, "Clinical evaluation of a method for automatic detection and removal of artifacts in auditory evoked potential monitoring", *J. Clin. Monit.*, vol. 11, no.6, pp. 381-391, Nov.1995.
- [39] R. J. Croft, J. S. Chandler, R. J. Barry, N. R. Cooper and A. R. Clarke, "EOG correction: a comparison of four methods", *Psychophysiology*, vol. 42, no.1, pp. 16-24, Jan.2005.
- [40] T. Gasser, L. Sroka and J. Mocks, "The transfer of EOG activity into the EEG for eyes open and closed", *Electroencephalogr. Clin. Neurophysiol.*, vol. 61, no.2, pp. 181-193, Aug.1985.
- [41] R. J. Croft and R. J. Barry, "Issues relating to the subtraction phase in EOG artefact correction of the EEG", *Int. J. Psychophysiol.*, vol. 44, no.3, pp. 187-195, Jun.2002.
- [42] T. P. Jung, C. Humphries, T. W. Lee, S. Makeig, M. J. McKeown, V. Iragui and T. J. Sejnowski, "Extended ICA removes artifacts from electroencephalographic recordings", in the Proc. *Advances in Neural Information Processing Systems (NIPS'98)*, vol. 10, pp. 894-900, 1998.
- [43] S. Choi, A. Cichocki, H. M. Park and S. Y. Lee, "Blind Source Separation and Independent Component Analysis: A Review", *Neural Information Processing-Letters and Review*, vol. 6, no.1, pp. 1-57, 2005.
- [44] R. Vigario, J. Sarela, V. Jousmiki, M. Hamalainen and E. Oja, "Independent component approach to the analysis of EEG and MEG recordings", *IEEE Trans. Biomed. Eng.*, vol. 47, no.5, pp. 589-593, 2000.
- [45] T. P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne and T. J. Sejnowski, "Analysis and visualization of single-trial event-related potentials", *Hum. Brain Mapp.*, vol. 14, no.3, pp. 166-185, Nov.2001.
- [46] T. P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne and T. J. Sejnowski, "Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects", *Clin. Neurophysiol.*, vol. 111, no.10, pp. 1745-1758, Oct.2000.

- [47] S. Makeig, S. Enghoff, T. P. Jung and T. J. Sejnowski, "A natural basis for efficient brain-actuated control", *IEEE Trans. Rehabil. Eng.*, vol. 8, no.2, pp. 208-211, Jun.2000.
- [48] W. De Clercq, A. Vergult, B. Vanrumste, J. Van Hees, A. Palmi, W. Van Paesschen and S. Van Huffel, "A new muscle artifact removal technique to improve the interpretation of the ictal scalp electroencephalogram", in the *Proc. 27<sup>th</sup> IEEE/EMBS Int. Conf.*, pp. 944-947, 2005.
- [49] J. Iriarte, E. Urrestarazu, M. Valencia, M. Alegre, A. Malanda, C. Viteri and J. Artieda, "Independent component analysis as a tool to eliminate artifacts in EEG: a quantitative study", *J. Clin. Neurophysiol.*, vol. 20, no.4, pp. 249-257, Jul-Aug.2003.
- [50] S. Makeig, S. Enghoff, T. P. Jung and T. J. Sejnowski, "Moving-window ICA decomposition of EEG data reveals event-related changes in oscillatory brain activity", In the *Proc. 2<sup>nd</sup> Int. Workshop on Independent Component Analysis and Signal Separation*, pp. 627-632, 2000.
- [51] C. J. James and O. J. Gibson, "Temporally constrained ICA: an application to artifact rejection in electromagnetic brain signal analysis", *IEEE Trans. Biomed. Eng.*, vol. 50, no.9, pp. 1108-1116, 2003.
- [52] C. A. Joyce, I. F. Gorodnitsky and M. Kutas, "Automatic removal of eye movement and blink artifacts from EEG data using blind component separation", *Psychophysiology*, vol. 41, no.2, pp. 313-325, 2004.
- [53] T. D. Lagerlund, F. W. Sharbrough and N. E. Busacker, "Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition", *J. Clin. Neurophysiol.*, vol. 14, no.1, pp. 73-82, Jan.1997.
- [54] O. G. Lins, T. W. Picton, P. Berg and M. Scherg, "Ocular artifacts in EEG and event-related potentials. I: Scalp topography", *Brain Topogr.*, vol. 6, no.1, pp. 51-63, Fall.1993.
- [55] O. G. Lins, T. W. Picton, P. Berg and M. Scherg, "Ocular artifacts in recording EEGs and event-related potentials. II: Source dipoles and source components", *Brain Topogr.*, vol. 6, no.1, pp. 65-78, Fall.1993.
- [56] M. Browne and T. R. Cutmore, "Low-probability event-detection and separation via statistical wavelet thresholding: an application to psychophysiological denoising", *Clin. Neurophysiol.*, vol. 113, no.9, pp. 1403-1411, Sep.2002.
- [57] T. Zikov, S. Bibian, G. A. Dumont, M. Huzmezan and C. R. Ries, "A wavelet based denoising technique for ocular artifact correction of the electroencephalogram," in *Proc. of the 2<sup>nd</sup> Joint EMBS/BMES Conference*, vol.1, pp. 98-105. 2002,
- [58] P. He, G. Wilson and C. Russell, "Removal of ocular artifacts from electro-encephalogram by adaptive filtering", *Med. Biol. Eng. Comput.*, vol. 42, no.3, pp. 407-412, May.2004.
- [59] S. Selvan and R. Srinivasan, "Removal of ocular artifacts from EEG using an efficient neuralnetwork based adaptive filtering technique", *IEEE Signal Processing Letters*, vol. 6, no.12, pp. 330-332, 1999.
- [60] P. Berg and M. Scherg, "A multiple source approach to the correction of eye artifacts", *Electroencephalogr. Clin. Neurophysiol.*, vol. 90, no.3, pp. 229-241, Mar.1994.
- [61] M. Cheng, W. Jia, X. Gao, S. Gao and F. Yang, "Mu rhythm-based cursor control: an offline analysis", *Electroencephalogr. Clin. Neurophysiol.*, vol. 115, no.4, pp. 745-751, Apr.2004.
- [62] L. Qin, J. Deng, L. Ding and B. He, "Motor imagery classification by means of source analysis methods," in the *Proc. 26<sup>th</sup> EMBC Int. Conf.*, vol.2, pp. 4356-4358 Vol.6. 2004.

- [63] Wenyan Jia, Xianghua Zhao, Hesheng Liu, Xiaorong Gao, Shangkai Gao and Fusheng Yang, "Classification of single trial EEG during motor imagery based on ERD," *in the Proc. 26<sup>th</sup> EMBC Int. Conf.*, vol.1, pp. 5-8 Vol.1. 2004.
- [64] Tao Wang, Jie Deng and Bin He, "Classification of motor imagery EEG patterns and their topographic representation," *in the Proc. 26<sup>th</sup> EMBC Int. Conf.*, vol.2, pp. 4359-4362 Vol.6. 2004.
- [65] S. Lemm, C. Schafer and G. Curio, "BCI Competition 2003--Data set III: probabilistic modeling of sensorimotor mu rhythms for classification of imaginary hand movements", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 1077-1080, Jun. 2004.
- [66] J. A. Pineda, D. S. Silverman, A. Vankov and J. Hestenes, "Learning to control brain rhythms: making a brain-computer interface possible", *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 11, no.2, pp. 181-184, Jun.2003.
- [67] C. Neuper, G. R. Muller, A. Kubler, N. Birbaumer and G. Pfurtscheller, "Clinical application of an EEG-based brain-computer interface: a case study in a patient with severe motor impairment", *Electroencephalogr. Clin. Neurophysiol.*, vol. 114, no.3, pp. 399-409, Mar.2003.
- [68] H. Sheikh, D. J. McFarland, W. A. Sarnacki and J. R. Wolpaw, "Electroencephalographic(EEG)-based communication: EEG control versus system performance in humans", *Neurosci. Lett.*, vol. 345, no.2, pp. 89-92, Jul.2003.
- [69] B. O. Peters, G. Pfurtscheller and H. Flyvbjerg, "Automatic Differentiation of Multichannel EEG Signals", *IEEE Trans. Biomed. Eng.*, vol. 48, no.1, pp. 111-116, Jan. 2001.
- [70] G. R. Muller, C. Neuper and G. Pfurtscheller, "Implementation of a Telemonitoring System for the Control of an EEG-Based Brain-Computer Interface", *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 11, no.1, pp. 54-59, Mar.2003.
- [71] G. Pfurtscheller, C. Neuper, C. Guger, W. Harkam, H. Ramoser, A. Schlogl, B. Obermaier and M. Pgegenzer, "Current trends in Graz Brain-Computer Interface (BCI) Research", *IEEE Trans. Rehab. Eng.*, vol. 8, no.2, pp. 216-219, Jun. 2000.
- [72] I. Ivanova, G. Pfurtscheller and C. Andrew, "AI-based classification of single-trial EEG data," *in the Proc. 26<sup>th</sup> EMBC Int. Conf.*, vol.1, pp. 703-704. 1995.
- [73] A. Kubler, F. Nijboer, J. Mellinger, T. M. Vaughan, H. Pawelzik, G. Schalk, D. J. McFarland, N. Birbaumer and J. R. Wolpaw, "Patients with ALS can use sensorimotor rhythms to operate a brain-computer interface", *Neurology*, vol. 64, no.10, pp. 1775-1777, May.2005.
- [74] G. Pfurtscheller, D. Flotzinger and J. Kalcher, "Brain-Computer Interface - a new communication device for handicapped persons", *J Microcomputer App.*, vol. 16, pp. 293-299, 1993.
- [75] S. Lemm, B. Blankertz, G. Curio and K. R. Muller, "Spatio-spectral filters for improving the classification of single trial EEG", *IEEE Trans. Biomed. Eng.*, vol. 52, no.9, pp. 1541-1548, Sep.2005.
- [76] L. Qin and B. He, "A wavelet-based time-frequency analysis approach for classification of motor imagery for brain-computer interface applications", *J. Neural Eng.*, vol. 2, no.4, pp. 65-72, Dec.2005.

- [77] G. R. Muller-Putz, R. Scherer, G. Pfurtscheller and R. Rupp, "EEG-based neuroprosthesis control: a step towards clinical practice", *Neurosci. Lett.*, vol. 382, no.1-2, pp. 169-174, Jul 1-8.2005.
- [78] Yuanqing Li, Andrzej Cichocki, Cuntai Guan and Jianzhao Qin, "Sparse factorization preprocessing-based offline analysis for a cursor control experiment," in *IEEE Int. Workshop Biomedical Circuits and Systems*, pp. S3/5/INV-S3/5/5-8. 2004,
- [79] J. R. Wolpaw and D. J. McFarland, "Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans", *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no.51, pp. 17849-17854, Dec 21.2004.
- [80] L. Qin, B. Kamousi, Z. M. Liu, L. Ding and B. He, "Classification of motor imagery tasks by means of time-frequency-spatial analysis for brain-computer interface applications," in *Proc. IEEE EMBS Int. Conf. on Neural Engineering*, pp. 374-376. 2005.
- [81] R. Leeb and G. Pfurtscheller, "Walking through a virtual city by thought in the *Proc. 26<sup>th</sup> EMBC Int. Conf.*, vol.2, pp. 4503-4506 , 2004.
- [82] G. Pfurtscheller, C. Neuper, D. Flotzinger and M. Pregenzer, "EEG-based discrimination between imagination of right and left hand movement", *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no.6, pp. 642-651, Dec.1997.
- [83] G. Pfurtscheller, C. Neuper, G. R. Muller, B. Obermaier, G. Krausz, A. Schlogl, R. Scherer, B. Graimann, C. Keinrath, D. Skliris, M. Wortz, G. Supp and C. Schrank, "Graz-BCI: state of the art and clinical applications", *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 11, no.2, pp. 177-180, Jun.2003.
- [84] F. Babiloni, F. Cincotti, L. Lazzarini, J. Millan, J. Mourino, M. Varsta, J. Heikkinen, L. Bianchi and M. G. Marciani, "Linear classification of low-resolution EEG patterns produced by imagined hand movements", *IEEE Trans. Rehab. Eng.*, vol. 8, no.2, pp. 186-188, Jun .2000.
- [85] T. Wang, J. Deng and B. He, "Classifying EEG-based motor imagery tasks by means of time-frequency synthesized spatial patterns", *Clin. Neurophysiol.*, vol. 115, no.12, pp. 2744-2753, Dec.2004.
- [86] F. Babiloni, L. Bianchi, F. Semeraro, J. d. R. Millan, J. Mourino, A. Cattini, S. Salinari, M. G. Marciani and F. Cincotti, "Mahalanobis distance-based classifiers are able to recognize EEG patterns by using few EEG electrodes," in pp. 651-654. 2001,
- [87] F. Cincotti, L. Bianchi, J. del R Millan, J. Mourino, S. Salinari, M. G. Marciani and F. Babiloni, "Brain computer interface: The use of low resolution surface laplacian and linear classifiers for the recognition of imagined hand movements," in *Engineering in Medicine and Biology Society, 2001.Proceedings of the 23rd Annual International Conference of the IEEE*, vol.1, pp. 655-658 vol.1. 2001,
- [88] F. Cincotti, A. Scipione, A. Timperi, D. Mattia, A. G. Marciani, J. Millan, S. Salinari, L. Bianchi and F. Babiloni, "Comparison of different feature classifiers for brain computer interfaces," in *Neural Engineering, 2003.Conference Proceedings.First International IEEE EMBS Conference on*, pp. 645-647. 2003,
- [89] 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)", *Neural Systems and Rehabilitation Engineering, IEEE Transactions on [See also IEEE Trans. on Rehabilitation Engineering]*, vol. 12, no.3, pp. 331-338, 2004.

- [90] L. Qin, L. Ding and B. He, "Motor imagery classification by means of source analysis for brain-computer interface applications", *J. Neural Eng.*, vol. 1, no.3, pp. 135-141, Sep.2004.
- [91] 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 Syst. Rehab. Eng.*, vol. 11, no.2, pp. 131-133, Jun.2003.
- [92] R. Boostani and M. H. Moradi, "A new approach in the BCI research based on fractal dimension as feature and Adaboost as classifier", *J. Neural Eng.*, vol. 1, no.4, pp. 212-217, Dec.2004.
- [93] G. Pfurtscheller, G. R. Muller, J. Pfurtscheller, H. J. Gerner and R. Rupp, "'Thought'--control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia", *Neurosci. Lett.*, vol. 351, no.1, pp. 33-36, 2003.
- [94] J. R. Wolpaw, D. J. McFarland and T. M. Vaughan, "Brain-Computer Interface Research at the Wadsworth Center", *IEEE Trans. Rehab. Eng.*, vol. 8, no.2, pp. 222-226, Jun.2000.
- [95] C. Guger, G. Edlinger, W. Harkam, I. Niedermayer and G. Pfurtscheller, "How many people are able to operate an EEG-based brain-computer interface (BCI)?", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no.2, pp. 145-147, Jun.2003.
- [96] J. Kalcher, D. Flotzinger and G. Pfurtscheller, "Graz brain-computer interface: An EEG-based cursor control system," in the *Proc. 15<sup>th</sup> IEEE EMBS Int. Conf.*, pp. 1264-1265. 1993.
- [97] J. Kalcher, D. Flotzinger and G. Pfurtscheller, "A new approach to a brain-computer-interface (BCI) based on learning vector quantization (LVQ3)," in the *Proc. 14<sup>th</sup> IEEE EMBS Int. Conf.*, vol.4, pp. 1658-1659. 1992.
- [98] 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. Rehab. Eng.*, vol. 11, no.2, pp. 204-207, Jun.2003.
- [99] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication", *Proce. IEEE*, vol. 89, no.7, pp. 1123-1134, 2001.
- [100] G. Pfurtscheller, D. Flotzinger and C. Neuper, "Differentiation between finger, toe and tongue movement in man based on 40 Hz EEG", *Electroencephalogr. Clin. Neurophysiol.*, vol. 90, no.6, pp. 456-460, Jun.1994.
- [101] G. Pfurtscheller, G. R. Müller-Putz, J. Pfurtscheller and R. Rupp, "EEG-Based Asynchronous BCI Controls Functional Electrical Stimulation in a Tetraplegic Patient", *EURASIP Journal on Applied Signal Processing*, vol. 2005, no.19, pp. 3152-3155, 2005.
- [102] D. Coyle, G. Prasad and T. M. McGinnity, "A Time-Frequency Approach to Feature Extraction for a Brain-Computer Interface with a Comparative Analysis of Performance Measures", *EURASIP Journal on Applied Signal Processing*, vol. 2005, no.19, pp. 3141-3151, 2005.
- [103] D. J. McFarland, L. M. McCane and J. R. Wolpaw, "EEG-based communication and control: short-term role of feedback", *IEEE Trans. Rehabil. Eng.*, vol. 6, no.1, pp. 7-11, Mar.1998.
- [104] F. Babiloni, F. Cincotti, L. Bianchi, G. Pirri, J. del R Millan, J. Mourino, S. Salinari and M. G. Marciani, "Recognition of imagined hand movements with low resolution surface Laplacian and linear classifiers", *Med. Eng. Phys.*, vol. 23, no.5, pp. 323-328, Jun.2001.



- [105] B. Graimann, J. E. Huggins, A. Schlogl, S. P. Levine and G. Pfurtscheller, "Detection of movement-related desynchronization patterns in ongoing single-channel electrocorticogram", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no.3, pp. 276-281, Sep.2003.
- [106] T. N. Lal, M. Schroder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer and B. Scholkopf, "Support vector channel selection in BCI", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 1003-1010, Jun.2004.
- [107] G. Blanchard and B. Blankertz, "BCI Competition 2003--Data set Ila: spatial patterns of self-controlled brain rhythm modulations", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 1062-1066, Jun.2004.
- [108] M. Krauledat, G. Dornhege, B. Blankertz, F. Losch, G. Curio and K. -. Muller, "Improving speed and accuracy of brain-computer interfaces using readiness potential features," in the *Proc. 26<sup>th</sup> EMBC Int. Conf.*, vol.2, pp. 4511-4515, 2004.
- [109] J. R. Wolpaw and D. J. McFarland, "Multichannel EEG-based brain-computer communication", *Electroencephalogr. Clin. Neurophysiol.*, vol. 90, no.6, pp. 444-449, Jun.1994.
- [110] 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, no.3, pp. 252-259, Mar.1991.
- [111] G. Pfurtscheller, C. Neuper, A. Schlogl 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, Sep.1998.
- [112] M. Pregenzer and G. Pfurtscheller, "Frequency component selection for an EEG-based brain to computer interface", *IEEE Trans. Rehabil. Eng.*, vol. 7, no.4, pp. 413-419, Dec.1999.
- [113] C. Neuper, R. Scherer, M. Reiner and G. Pfurtscheller, "Imagery of motor actions: differential effects of kinesthetic and visual-motor mode of imagery in single-trial EEG", *Brain Res. Cogn. Brain Res.*, vol. 25, no.3, pp. 668-677, Dec.2005.
- [114] 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, no.4, pp. 447-456, Dec.2000.
- [115] G. Townsend, B. Graimann and G. Pfurtscheller, "Continuous EEG classification during motor imagery--simulation of an asynchronous BCI", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 12, no.2, pp. 258-265, Jun.2004.
- [116] M. Pregenzer and G. Pfurtscheller, "Distinction sensitive learning vector quantization (DSLQ) application as a classifier based feature selection method for a brain computer interface," in the *Proc. 4<sup>th</sup> Int. Conf. Artificial Neural Networks*, pp. 433-436. 1995.
- [117] P. J. Durka, "Finding Significant Correlates of Conscious Activity in Rhythmic EEG", *EURASIP Journal on Applied Signal Processing*, vol. 2005, no.19, pp. 3122-3127, 2005.
- [118] 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, no.3, pp. 233-240, Sep.2003.
- [119] L. J. Trejo, K. R. Wheeler, C. C. Jorgensen, R. Rosipal, S. T. Clanton, B. Matthews, A. D. Hibbs, R. Matthews and M. Krupka, "Multimodal neuroelectric interface development", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no.2, pp. 199-204, Jun.2003.

- [120] J. R. Wolpaw, D. Flotzinger, G. Pfurtscheller and D. J. McFarland, "Timing of EEG-based cursor control", *J Clinical Neurophysiol*, vol. 14, no.6, pp. 529-538, Nov.1997.
- [121] G. Schalk, J. R. Wolpaw, D. J. McFarland and G. Pfurtscheller, "EEG-based communication: presence of an error potential", *Electroencephalogr. Clin. Neurophysiol.*, vol. 111, no.12, pp. 2138-2144, Dec.2000.
- [122] G. Pfurtscheller, J. Kalcher, C. Neuper, D. Flotzinger and M. Pregenzer, "On-line EEG classification during externally-paced hand movements using a neural network-based classifier", *Electroencephalogr. Clin. Neurophysiol.*, vol. 99, no.5, pp. 416-425, Nov.1996.
- [123] S. G. Mason and G. E. Birch, "A brain-controlled switch for asynchronous control applications", *IEEE Trans. Biomed. Eng.*, vol. 47, no.10, pp. 1297-1307, Oct.2000.
- [124] J. R. LaCourse and E. Wilson, "Brainiac: A brain computer link," in Proc. IEEE Instrumentation and Measurement Technology Conference, ,pp. 587. 1995.
- [125] B. Mahmoudi and A. Erfanian, "Single-channel EEG-based prosthetic hand grasp control for amputee subjects," in *Proc. of the 2<sup>nd</sup> Joint EMBS/BMES Conf.*,vol.3,pp. 2406-2407 2002.
- [126] S. Kelly, D. Burke, P. de Chazal and R. Reilly, "Parametric models and spectral analysis for classification in brain-computer interfaces," in Proc. 14<sup>th</sup> Int. Conf. DSP, ,vol.1,pp. 307-310,2002.
- [127] C. I. Hung, P. L. Lee, Y. T. Wu, L. F. Chen, T. C. Yeh and J. C. Hsieh, "Recognition of motor imagery electroencephalography using independent component analysis and machine classifiers", *Ann. Biomed. Eng.*, vol. 33, no.8, pp. 1053-1070, Aug.2005.
- [128] M. Fatourech, A. Bashashati, J. F. Borisoff, G. E. Birch and R. K. Ward, "Improving the performance of the LF-ASD brain computer interface by means of genetic algorithm," in Proc. IEEE ISSPIT'04, pp. 38-41. 2004.
- [129] M. Fatourech, A. Bashashati, R. K. Ward and G. E. Birch, "A hybrid genetic algorithm approach for improving the performance of the LF-ASD brain computer interface," in Proc.ICASSP'05, vol.5,pp. 345-348. 2005.
- [130] E. L. Glassman, "A wavelet-like filter based on neuron action potentials for analysis of human scalp electroencephalographs", *IEEE Trans. Biomed. Eng.*, vol. 52, no.11, pp. 1851-1862, Nov.2005.
- [131] A. B. Barreto, A. M. Taberner and L. M. Vicente, "Classification of spatio-temporal EEG readiness potentials towards the development of a brain-computer interface," in *Proc. IEEE Southeastern Biomed. Eng. Conf*, pp. 99-102. 1996.
- [132] A. B. Barreto, A. M. Taberner and L. M. Vicente, "Neural network classification of spatio-temporal EEG readiness potentials," in *Proc. 15<sup>th</sup> IEEE Southeastern Biomed. Eng. Conf.*, ,pp. 73-76. 1996,
- [133] B. Blankertz, G. Dornhege, C. SchÅfer, R. Krepi, J. Kolmorgen, K. R. MÅller, 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," in *IEEE Trans. Neural Sys. Rehab. Eng.*, vol.11(2),2003,
- [134] G. Dornhege, B. Blankertz and G. Curio, "Speeding up classification of multi-channel brain-computer interfaces: Common spatial patterns for slow cortical potentials," in *Proc. IEEE EMBS Int. Conf. on Neural Engineering*,pp. 595-598. 2003.

- [135] E. Yom-Tov and G. F. Inbar, "Detection of Movement-Related Potentials from the Electro-Encephalogram for possible use in a Brain-Computer Interface", *Medical and Biological Engineering and Computing*, vol. 41, no.1, pp. 85-93, Jan. 2003.
- [136] B. Blankertz, G. Curio and K. R. Müller, "Classifying single trial EEG: Towards brain computer interfacing," in *Proc. Advances in Neural Inf. Proc. Systems (NIPS 01)*, vol.14, pp. 157-164, 2002.
- [137] D. P. Burke, S. P. Kelly, P. de Chazal, R. B. Reilly and C. Finucane, "A parametric feature extraction and classification strategy for brain-computer interfacing", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 13, no.1, pp. 12-17, Mar.2005.
- [138] A. Bashashati, R. K. Ward and G. E. Birch, "A new design of the asynchronous brain computer interface using the knowledge of the path of features," in *Proc. IEEE EMBS Int. Conf. on Neural Engineering*, pp. 101-104. 2005.
- [139] G. E. Birch, S. G. Mason and J. F. Borisoff, "Current Trends in Brain-Computer Interface Research at the Neil Squire Foundation", *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 11, no.2, pp. 123-126, Jun. 2003.
- [140] D. Lisogurski and G. E. Birch, "Identification of finger flexions from continuous EEG as a brain computer interface," in *the Proc. 20<sup>th</sup> IEEE/EMBS Int. Conf.*, vol.4, pp. 2004-2007 vol.4. 1998.
- [141] Z. Bozorgzadeh, G. E. Birch and S. G. Mason, "The LF-ASD brain computer interface: On-line identification of imagined finger flexions in the spontaneous EEG of able-bodied subjects," in *Proc. IEEE ICASSP'05*, vol.6, pp. 2385-2388, 2000.
- [142] J. F. Borisoff, S. G. Mason, A. Bashashati and G. E. Birch, "Brain-computer interface design for asynchronous control applications: improvements to the LF-ASD asynchronous brain switch", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 985-992, Jun.2004.
- [143] S. G. Mason, R. Bohringer, J. F. Borisoff and G. E. Birch, "Real-time control of a video game with a direct brain-computer interface", *J. Clin. Neurophysiol.*, vol. 21, no.6, pp. 404-408, Nov-Dec.2004.
- [144] G. E. Birch, Z. Bozorgzadeh and S. G. Mason, "Initial on-line evaluations of the LF-ASD brain-computer interface with able-bodied and spinal-cord subjects using imagined voluntary motor potentials", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 10, no.4, pp. 219-224, Dec.2002.
- [145] W. Xu, C. Guan, C. E. Siong, S. Ranganatha, M. Thulasidas and M. Wu, "High accuracy classification of EEG signal," in *IEEE Int. Conf. Pattern Recognition (ICPRA'04)*, pp. 391-394. 2004.
- [146] D. Coyle, G. Prasad and T. M. McGinnity, "Extracting features for a brain-computer interface by self-organising fuzzy neural network-based time series prediction," in *the Proc. 26<sup>th</sup> IEEE/EMBS Int. Conf.*, vol.2, pp. 4371-4374, 2004.
- [147] R. Scherer, G. R. Muller, C. Neuper, B. Graimann and G. Pfurtscheller, "An asynchronously controlled EEG-based virtual keyboard: improvement of the spelling rate", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 979-984, Jun.2004.
- [148] C. Neuper, A. Schlögl and G. Pfurtscheller, "Enhancement of Left-Right Sensorimotor EEG Differences during Feedback-Regulated Motor Imagery", *Journal of Clinical Neurophysiology*, vol. 16, no.4, pp. 373-382, 1999.

- [149] G. Pfurtscheller and C. Guger, "Brain-Computer Communication System: EEG-based control of hand orthosis in a tetraplegic patient", *Acta Chir. Austriaca*, vol. 31, pp. 23-25, 1999.
- [150] S. J. Roberts, W. Penny and I. Rezek, "temporal and spatial complexity measures for EEG-based Brain Computer Interfacing", *Medical and Biological Engineering and Computing*, vol. 37, no.1, pp. 93-99, 2003.
- [151] M. Schroder, M. Bogdan, T. Hinterberger and N. Birbaumer, "Automated EEG feature selection for brain computer interfaces," in *Proc. 1<sup>st</sup> IEEE EMBS Int. Conf. on Neural Engineering*, pp. 626-629. 2003.
- [152] A. Schlogl, C. Keinrath, R. Scherer and P. Furtscheller, "Information transfer of an EEG-based brain computer interface," in *Proc. 1<sup>st</sup> IEEE EMBS Int. Conf. on Neural Engineering*, pp. 641-644. 2003.
- [153] J. E. Huggins, S. P. Levine, J. A. Fessler, W. M. Sowers, G. Pfurtscheller, B. Graimann, A. Schloegl, D. N. Minecan, R. K. Kushwaha, S. L. BeMent, O. Sagher and L. A. Schuh, "Electrocorticogram as the basis for a direct brain interface: Opportunities for improved detection accuracy," in *Proc. 1<sup>st</sup> IEEE EMBS Int. Conf. on Neural Engineering*, pp. 587-590. 2003.
- [154] J. E. Huggins, S. P. Levine, S. L. Bement, R. K. Kushwaha, L. A. Schuh, E. A. Passaro, M. M. Rohde, D. A. Ross, K. V. Elisevich and B. J. Smith, "Detection of Event-Related Potentials for Development of a Direct Brain Interface", *J Clinical Neurophysiol*, vol. 16, no.5, pp. 448-455, 09//.1999.
- [155] S. P. Levine, J. E. Huggins, S. L. Bement, R. K. Kushwaha, L. A. Schuh, E. A. Passaro, M. M. Rohde and D. A. Ross, "Identification of Electrocorticogram Patterns as the Basis for a Direct Brain Interface", *J Clinical Neurophysiol*, vol. 16, no.5, pp. 439-447, Sep.1999.
- [156] G. N. Garcia, T. Ebrahimi and J. -. Vesin, "Correlative exploration of EEG signals for direct brain-computer communication," in *Proc. IEEE ICASSP'03*, vol.5, pp. 816-19, 2003.
- [157] B. Obermaier, G. R. Muller and G. Pfurtscheller, ""Virtual keyboard" controlled by spontaneous EEG activity", *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 11, no.4, pp. 422-426, Dec. 2003.
- [158] E. J. Costa and E. F. C. Jr, "EEG-based discrimination between imagination of left and right hand movements using Adaptive Gaussian Representation", *Med. Eng. Phys.*, vol. 22, no.5, pp. 345-348, Jun. 2000.
- [159] U. H. Balbale, J. E. Huggins, S. L. BeMent and S. P. Levine, "Multi-channel analysis of human event-related cortical potentials for the development of a direct brain interface in *Proc. of the 1<sup>st</sup> Joint EMBS/BMES Conference*, vol.1, pp. 447, 1999.
- [160] 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. Rehabil. Eng.*, vol. 8, no.2, pp. 180-185, Jun.2000.
- [161] E. Yom-Tov and G. F. Inbar, "Selection of relevant features for classification of movements from single movement-related potentials using a genetic algorithm," in *Proc. 23<sup>rd</sup> IEEE EMBS Int. Conf. on Neural Engineering*, vol.2, pp. 1364-1366, 2001.
- [162] E. Yom-Tov and G. F. Inbar, "Feature Selection for the Classification of Movements From Single Movement-Related Potentials", *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 10, no.3, pp. 170-177, Sep.2002.

- [163] B. H. Cho, S. Kim, D. I. Shin, J. H. Lee, S. M. Lee, I. Y. Kim and S. I. Kim, "Neurofeedback training with virtual reality for inattention and impulsiveness", *Cyberpsychol Behav.*, vol. 7, no.5, pp. 519-526, Oct.2004.
- [164] H. Yoon, K. Yang and C. Shahabi, "Feature subset selection and feature ranking for multivariate time series", *IEEE Trans. Knowled. Data Eng.*, vol. 17, no.9, pp. 1186-1198, 2005.
- [165] D. Burke, S. Kelly, P. de Chazal and R. Reilly, "A simultaneous filtering and feature extraction strategy for direct brain interfacing in *Proc. of the 2<sup>nd</sup> Joint EMBS/BMES Conference*, vol.1, pp. 279-280 , 2002.
- [166] H. Lee and S. Choi, "PCA-based linear dynamical systems for multichannel EEG classification," in *Proc.ICONIP '02*, vol.2, pp. 745-749 , 2002.
- [167] Hyekyung Lee and Seungjin Choi, "PCA+HMM+SVM for EEG pattern classification," in *Proc. 7<sup>th</sup> Int. Symp. Signal Processing and Its Applications*, vol.1, pp. 541-544 , 2003.
- [168] M. Schröder, T. Navin Lal, T. Hinterberger, M. Bogdan, N. J. Hill, N. Birbaumer, W. Rosenstiel and B. Schölkopf, "Robust EEG Channel Selection across Subjects for Brain-Computer Interfaces", *EURASIP Journal on Applied Signal Processing*, vol. 2005, pp. 3103-3112, 2005.
- [169] M. M. Rohde, S. L. BeMent, J. E. Huggins, S. P. Levine, R. K. Kushwaha and L. A. Schuh, "Quality estimation of subdurally recorded, event-related potentials based on signal-to-noise ratio", *IEEE Trans. Biomed. Eng.*, vol. 49, no.1, pp. 31-40, Jan.2002.
- [170] A. Schloegl, K. Lugger and G. Pfurtscheller, "Using adaptive autoregressive parameters for a brain-computer-interface experiment," in *the Proc. 19<sup>th</sup> IEEE EMBS Int. Conf.*, vol.4, pp. 1533-1535 , 1997.
- [171] B. Obermaier, C. Guger, C. Neuper and G. Pfurtscheller, "Hidden Markov models for online classification of single trial EEG data", *Pattern Recognition Letters*, vol. 22, no.12, pp. 1299-1309, 2001.
- [172] E. Haselsteiner and G. Pfurtscheller, "Using Time-Dependent Neural Networks for EEG Classification", *IEEE Trans. Rehab. Eng.*, vol. 8, no.4, pp. 457-463, Dec.2000.
- [173] P. Sykacek, S. Roberts, M. Stokes, E. Curran, M. Gibbs and L. Pickup, "Probabilistic methods in BCI research", *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 11, no.2, pp. 192-195, Jun.2003.
- [174] B. Obermaier, C. Munteanu, A. Rosa and G. Pfurtscheller, "Asymmetric Hemisphere Modeling in an Offline Brain-Computer Interface", *IEEE Trans. Syst., Man and Cyber. - Part C*, vol. 31, no.4, pp. 536-540, Nov.2001.
- [175] A. Schloegl, C. Neuper and G. Pfurtscheller, "Subject specific EEG patterns during motor imaginary", *the Proc. 19<sup>th</sup> IEEE EMBS Int. Conf.*, vol.4, pp. 1530-1532 , 1997.
- [176] L. Parra, C. Alvino, A. Tang, B. Pearlmutter, N. Yeung, A. Osman and P. Sajda, "Single-trial detection in EEG and MEG: Keeping it linear", *Neurocomputing*, vol. 52-54, pp. 177-183, 2003.
- [177] L. Parra, C. Alvino, A. Tang, B. Pearlmutter, N. Yeung, A. Osman and P. Sajda, "Linear Spatial Integration for Single-Trial Detection in Encephalography", *Neuroimage*, vol. 17, pp. 223-230, 2002.

- [178] G. N. Garcia, T. Ebrahimi and J. -. Vesin, "Support vector EEG classification in the fourier and time-frequency correlation domains," in *Proc. IEEE EMBS Int. Conf. Neural Eng.*, pp. 591-594. 2003.
- [179] G. E. Birch, P. D. Lawrence and R. D. Hare, "Single-trial processing of event-related potentials using outlier information", *IEEE Trans. Biomed. Eng.*, vol. 40, no.1, pp. 59-73, Jan.1993.
- [180] L. C. Parra, C. D. Spence, A. D. Gerson and P. Sajda, "Response error correction--a demonstration of improved human-machine performance using real-time EEG monitoring", *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 11, no.2, pp. 173-177, Jun.2003.
- [181] V. Bostanov, "BCI Competition 2003--Data sets Ib and Iib: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 1057-1061, Jun.2004.
- [182] J. Kaiser, J. Perelmouter, I. H. Iversen, N. Neumann, N. Ghanayim, T. Hinterberger, A. Kubler, B. Kotchoubey and N. Birbaumer, "Self-initiation of EEG-based communication in paralyzed patients", *Electroencephalogr. Clin. Neurophysiol.*, vol. 112, no.3, pp. 551-554, Mar.2001.
- [183] T. Hinterberger, N. Weiskopf, R. Veit, B. Wilhelm, E. Betta and N. Birbaumer, "An EEG-driven brain-computer interface combined with functional magnetic resonance imaging (fMRI)", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 971-974, Jun.2004.
- [184] T. Hinterberger, S. Schmidt, N. Neumann, J. Mellinger, B. Blankertz, G. Curio and N. Birbaumer, "Brain-computer communication and slow cortical potentials", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 1011-1018, Jun.2004.
- [185] J. Kaiser, A. Kubler, T. Hinterberger, N. Neumann and N. Birbaumer, "A non-invasive communication device for the paralyzed", *Minimally Invasive Neurosurgery*, vol. 45, no.1, pp. 19-23, Mar.2002.
- [186] T. Hinterberger, N. Neumann, M. Pham, A. Kubler, A. Grether, N. Hofmayer, B. Wilhelm, H. Flor and N. Birbaumer, "A multimodal brain-based feedback and communication system", *Exp. Brain Res.*, vol. 154, no.4, pp. 521-526, Feb.2004.
- [187] T. Hinterberger, B. Wilhelm, J. Mellinger, B. Kotchoubey and N. Birbaumer, "A device for the detection of cognitive brain functions in completely paralyzed or unresponsive patients", *IEEE Trans. Biomed. Eng.*, vol. 52, no.2, pp. 211-220, Feb.2005.
- [188] N. Neumann, T. Hinterberger, J. Kaiser, U. Leins, N. Birbaumer and A. Kubler, "Automatic processing of self-regulation of slow cortical potentials: evidence from brain-computer communication in paralysed patients", *Electroencephalogr. Clin. Neurophysiol.*, vol. 115, no.3, pp. 628-635, Mar.2004.
- [189] A. Kubler, B. Kotchoubey, T. Hinterberger, N. Ghanayim, J. Perelmouter, M. Schauer, C. Fritsch, E. Taub and N. Birbaumer, "The thought translation device: a neurophysiological approach to communication in total motor paralysis", *Experimental Brain Research*, vol. 124, no.2, pp. 223-232, Jan.1999.
- [190] T. Hinterberger, A. Kubler, J. Kaiser, N. Neumann and N. Birbaumer, "A brain-computer interface (BCI) for the locked-in: comparison of different EEG classifications for the thought translation device", *Electroencephalogr. Clin. Neurophysiol.*, vol. 114, no.3, pp. 416-425, Mar.2003.

- [191] A. Kubler, N. Neumann, J. Kaiser, B. Kotchoubey, T. Hinterberger and N. P. Birbaumer, "Brain-computer communication: self-regulation of slow cortical potentials for verbal communication", *Arch. Phys. Med. Rehabil.*, vol. 82, no.11, pp. 1533-1539, Nov, 2001.
- [192] A. Kuebler, B. Kotchoubey, H. P. Salzmann, N. Ghanayim, J. Perelmouter, V. Homberg and N. Birbaumer, "Self-regulation of slow cortical potentials in completely paralyzed human patients", *Neurosci. Lett.*, vol. 252, no.3, pp. 171-174, Aug. 1998.
- [193] T. Hinterberger, R. Veit, B. Wilhelm, N. Weiskopf, J. J. Vatine and N. Birbaumer, "Neuronal mechanisms underlying control of a brain-computer interface", *Eur. J. Neurosci.*, vol. 21, no.11, pp. 3169-3181, Jun.2005.
- [194] N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kubler, J. Perelmouter, E. Taub and H. Flor, "A spelling device for the paralysed", *Nature*, vol. 398, no.6725, pp. 297-298, Mar.1999.
- [195] N. Birbaumer, A. Kubler, N. Ghanayim, T. Hinterberger, J. Perelmouter, J. Kaiser, I. Iversen, B. Kotchoubey, N. Neumann and H. Flor, "The thought translation device (TTD) for completely paralyzed patients", *IEEE Trans. Rehabil. Eng.*, vol. 8, no.2, pp. 190-193, Jun.2000.
- [196] R. Palaniappan, "Brain computer interface design using band powers extracted during mental tasks," in *Proc. of the 2<sup>nd</sup> Joint EMBS/BMES Conference*, pp. 321-324. 2005.
- [197] Nai-Jen Huan and R. Palaniappan, "Classification of mental tasks using fixed and adaptive autoregressive models of EEG signals," in *Proc. of the 2<sup>nd</sup> Joint EMBS/BMES Conference*, pp. 633-636. 2005.
- [198] A. Kostov and M. Polak, "Parallel Man-Machine Training in Development of EEG-Based Cursor Control", *IEEE Trans. Rehab. Eng.*, vol. 8, no.2, pp. 203-205, Jun.2000.
- [199] M. Polak and A. Kostov, "Development of brain-computer interface: Preliminary results," in *the Proc. 19<sup>th</sup> IEEE EMBS Int. Conf.*, vol.4, pp. 1543-1546 vol.4. 1997.
- [200] M. Polak and A. Kostov, "Feature extraction in development of brain-computer interface: A case study," in *the Proc. 20<sup>th</sup> IEEE EMBS Int. Conf.*, vol.4, pp. 2058-2061 vol.4. 1998.
- [201] M. Polak and A. Kostov, "Training setup for control of neural prosthesis using brain-computer interface," in *Proc. of the 1<sup>st</sup> Joint EMBS/BMES Conference*, vol.1, pp. 446, 1999.
- [202] N. J. Huan and R. Palaniappan, "Neural network classification of autoregressive features from electroencephalogram signals for brain-computer interface design", *J. Neural Eng.*, vol. 1, no.3, pp. 142-150, Sep.2004.
- [203] R. Palaniappan, R. Paramesran, S. Nishida and N. Saiwaki, "A New Brain-Computer Interface Design Using Fuzzy ARTMAP", *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 10, no.3, pp. 140-148, Sep.2002.
- [204] A. Kostov and M. Polak, "Prospects of computer access using voluntary modulated EEG signal," in *Proc. ECPD Symp. Brain & Consciousness*, pp. 233-236. 1997.
- [205] C. W. Anderson, E. A. Stolz and S. Shamsunder, "Discriminating mental tasks using EEG represented by AR models," in *the Proc. 17<sup>th</sup> IEEE EMBS Int. Conf.* vol.2, pp. 875-876 1995.
- [206] C. W. Anderson, S. V. Devulapalli and E. A. Stolz, "Signal Classification with Different Signal Representations", *Neural Networks for Signal Processing*, pp. 475-483, 1995.

- [207] D. A. Peterson, J. N. Knight, M. J. Kirby, C. W. Anderson and M. H. Thaut, "Feature Selection and Blind Source Separation in an EEG-Based Brain-Computer Interface", *EURASIP Journal on Applied Signal Processing*, vol. 2005, no.19, pp. 3128-3140, 2005.
- [208] K. Tavakolian, A. M. Nasrabadi and S. Rezaei, "Selecting better EEG channels for classification of mental tasks," in *Proc. ISCAS '04*, vol.3, pp. III-537-40, 2004.
- [209] A. Panuccio, M. Bicego and V. Murino, "A hidden markov model-based approach to sequential data clustering," in *Structural, Syntactic and Statistical Pattern Recognition* (A. A. T. Caelli R.P.W. Duin, Ed.), Springer, 2002, pp. 734-742.
- [210] D. Wang, T. Kochiyama, S. Lu and J. Wu, "Measurement and analysis of electroencephalogram (EEG) using directional visual stimuli for brain computer interface," in *Proc. AMT 2005*, pp. 34-39. 2005.
- [211] 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. Rehab. Eng.*, vol. 11, no.2, pp. 141-144, Jun.2003.
- [212] Z. A. Keirn and J. I. Aunon, "A new mode of communication between man and his surroundings", *IEEE Trans. Biomed. Eng.*, vol. 37, no.12, pp. 1209-1214, Dec.1990.
- [213] J. d. R. Millan, J. Mourino, M. G. Marciani, F. Babiloni, F. Topani, I. Canale, J. Heikkinen and K. Kaski, "Adaptive brain interfaces for physically-disabled people," in *the Proc. 20<sup>th</sup> IEEE EMBS Int. Conf.*, vol.4, pp. 2008-2011, 1998.
- [214] C. W. Anderson, E. A. Stolz and S. Shamsunder, "Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks", *IEEE Trans. Biomed. Eng.* vol. 45, no.3, pp. 277-286, Mar.1998.
- [215] Liu Hailong, Wang Jue and Zheng Chongxun, "Mental tasks classification and their EEG structures analysis by using the growing hierarchical self-organizing map," in *Proc. 1<sup>st</sup> Int. Conf. Neural Interface and Control*, pp. 115-118. 2005.
- [216] B. J. Culpepper and R. M. Keller, "Enabling computer decisions based on EEG input", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no.4, pp. 354-360, Dec.2003.
- [217] Y. Wang, Z. Zhang, Y. Li, X. Gao, S. Gao and F. Yang, "BCI Competition 2003--Data set IV: an algorithm based on CSSD and FDA for classifying single-trial EEG", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 1081-1086, Jun.2004.
- [218] G. Dornhege, B. Blankertz, G. Curio and K. R. Muller, "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 993-1002, Jun.2004.
- [219] Y. Li, X. Gao, H. Liu and S. Gao, "Classification of single-trial electroencephalogram during finger movement", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 1019-1025, Jun.2004.
- [220] H. S. Liu, X. Gao, F. Yang and S. Gao, "Imagined hand movement identification based on spatio-temporal pattern recognition of EEG," in *Proc. 1<sup>st</sup> Int. EMBS/BMES Conf. Neural Eng.*, pp. 599-602. 2003.
- [221] B. Graimann, J. E. Huggins, S. P. Levine and G. Pfurtscheller, "Detection of ERP and ERD/ERS patterns in single ECoG channels," in *1<sup>st</sup> Int. IEEE/EMBS Conf. . Neural Eng.*, pp. 614-617. 2003.
- [222] B. Graimann, J. E. Huggins, S. P. Levine and G. Pfurtscheller, "Toward a direct brain interface based on human subdural recordings and wavelet-packet analysis", *IEEE Trans. Biomed. Eng.* vol. 51, no.6, pp. 954-962, Jun. 2004.



- [223] K. R. Muller, G. Curio, B. Blankertz and G. Dornhege, "Combining features for BCI," in *Proc. Advances in Neural Inf. Proc. Systems (NIPS 02)*, vol.15,2003.
- [224] B. D. Mensh, J. Werfel and H. S. Seung, "BCI Competition 2003--Data set Ia: combining gamma-band power with slow cortical potentials to improve single-trial classification of electroencephalographic signals", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 1052-1056, Jun. 2004.
- [225] T. Hinterberger and G. Baier, "Parametric orchestral sonification of EEG in real time", *Multimedia, IEEE*, vol. 12, no.2, pp. 70-79, 2005.
- [226] J. Millan, J. Mourino, M. Franze, F. Cincotti, M. Varsta, J. Heikkonen and F. Babiloni, "A Local Neural Classifier for the Recognition of EEG Patterns Associated to Mental Tasks", *IEEE Trans. Neural Networks*, vol. 13, no.3, pp. 678-686, May. 2002.
- [227] J. d. R. Millan, F. Renkens, J. Mourino and W. Gerstner, "Non-invasive brain-actuated control of a mobile robot," in *Proc. 18<sup>th</sup> Joint Int. Conf. Artificial Intelligence*, vol.51,2004.
- [228] J. d. Millan, J. Mourino, F. Cincotti, M. Varsta, J. Heikkonen, F. Topani, M. G. Marciani, K. Kaski and F. Babiloni, "Neural networks for robust classification of mental tasks," in *the Proc. 22<sup>nd</sup> IEEE EMBS Int. Conf.*, vol.2, pp. 1380-1382 vol.2. 2000,
- [229] J. d. R. Millan, J. Mourino, F. Babiloni, F. Cincotti, M. Varsta and J. Heikkonen, "Local neural classifier for EEG-based recognition of mental tasks," in *Prpc. IEEE IJCNN 2000*, vol.3, pp. 632-636 , 2000.
- [230] R. Millan Jdel and J. Mourino, "Asynchronous BCI and local neural classifiers: an overview of the Adaptive Brain Interface project", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no.2, pp. 159-161, Jun.2003.
- [231] J. R. Millan, "On the need for on-line learning in brain-computer interfaces," in *Proc. IEEE Int. Conf. Neural Networks*, vol.4, pp. 2877-2882 , 2004.
- [232] R. Millan Jdel, F. Renkens, J. Mourino and W. Gerstner, "Noninvasive brain-actuated control of a mobile robot by human EEG", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 1026-1033, Jun.2004.
- [233] E. Gysels and P. Celka, "Phase synchronization for the recognition of mental tasks in a brain-computer interface", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 12, no.4, pp. 406-415, Dec.2004.
- [234] E. Gysels, P. Renevey and P. Celka, "SVM-based recursive feature elimination to compare phase synchronization computed from broadband and narrowband EEG signals in Brain-Computer Interfaces", *Signal Process*, vol. 85, no.11, pp. 2178-2189, 11.2005.
- [235] E. Curran, P. Sykacek, M. Stokes, S. J. Roberts, W. Penny, I. Johnsrude and A. M. Owen, "Cognitive Tasks for Driving a Brain-Computer Interface System: A Pilot Study", *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 12, no.1, pp. 48-54, 2004.
- [236] J. Mourino, S. Chiappa, R. Jane and J. d. R. Millan, "Evolution of the mental states operating A brain-computer interface," *International Federation for Medical and Biological Engineering*, pp. 400-401. 2002.
- [237] I. Rezek, S. Roberts and P. Sykacek, "Ensemble coupled hidden markov models for joint characterization of dynamic signals," in *Proc. 9<sup>th</sup> Int. Workshop . Artificial Intelligence and Statistics*, 2003.
- [238] G. Garcia and T. Ebrahimi, "Time-frequency-space kernel for single EEG-trial classification," in *the Proc. NORSIG Conf.*, 2002.

- [239] S. J. Roberts and W. Penny, "Real-time Brain-Computer Interfacing: a preliminary study using Bayesian learning", *Medical and Biological Engineering and Computing*, vol. 38, no.1, pp. 56-61, 2003.
- [240] M. Varsta, J. Heikkonen, J. d. Millan and J. Mourino, "Evaluating the performance of three feature sets for brain-computer interfaces with an early stopping MLP committee," in Proc. 15<sup>th</sup> Int. Conf. Pattern Recognition, vol.2, pp. 907-910 , 2000.
- [241] W. D. Penny, S. J. Roberts, E. A. Curran and M. J. Stokes, "EEG-Based Communication: A Pattern Recognition Approach", *IEEE Trans. Rehab. Eng.*, vol. 8, no.2, pp. 214-215 Jun. 2000.
- [242] P. Sykacek, S. J. Roberts and M. Stokes, "Adaptive BCI based on variational Bayesian Kalman filtering: an empirical evaluation", *IEEE Trans. Biomed. Eng.*, vol. 51, no.5, pp. 719-727, May.2004.
- [243] G. N. Garcia, T. Ebrahimi and J. M. Vesin, "Joint Time-Frequency-Space Classification of EEG in a Brain-Computer Interface Application", *EURASIP Journal on Applied Signal Processing*, vol. 2003, no.7, pp. 713-729, 2003.
- [244] G. N. Garcia, T. Ebrahimi, J. -. Vesin and A. Villca, "Direct brain-computer communication with user rewarding mechanism," in *Proc. IEEE Int. Symp. Information Theory*, ,pp. 221-221. 2003,
- [245] G. Garcia, T. Ebrahimi and J. -. Vesin, "Classification of EEG signals in the ambiguity domain for brain computer interface applications," in Proc. 14<sup>th</sup> Int. Conf. Digital Signal Processing, vol.1,pp. 301-305 ,2002.
- [246] A. Erfanian and A. Erfani, "ICA-based classification scheme for EEG-based brain-computer interface: The role of mental practice and concentration skills," in *the Proc. 26<sup>th</sup> EMBC Int. Conf.*, vol.1, pp. 235-238 , 2004.
- [247] M. Kaper, P. Meinicke, U. Grossekhoefer, T. Lingner and H. Ritter, "BCI Competition 2003--Data set IIB: support vector machines for the P300 speller paradigm", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 1073-1076, Jun..2004.
- [248] H. Serby, E. Yom-Tov and G. F. Inbar, "An improved P300-based brain-computer interface", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 13, no.1, pp. 89-98, Mar.2005.
- [249] Gao Xiaorong, Xu Neng, Hong Bo, Gao Shanghai and Yang Fusheng, "Optimal selection of independent components for event-related brain electrical potential enhancement," in Proc. IEEE Int. Workshop. Biomedical Circuits and Systems, pp. S3/5/INV-S3/5/1-4, 2004.
- [250] B. H. Jansen, A. Allam, P. Kota, K. Lachance, A. Osho and K. Sundaresan, "An exploratory study of factors affecting single trial P300 detection", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 975-978, Jun.2004.
- [251] M. Kaper and H. Ritter, "Generalizing to new subjects in brain-computer interfacing," in *the Proc. 26<sup>th</sup> EMBC Int. Conf.*, vol.2, pp. 4363-4366 , 2004.
- [252] M. Kaper and H. Ritter, "Progress in P300-based brain-computer interfacing," in Proc. IEEE Int. Workshop Biomedical Circuits and Systems, ,pp. S3/5/INV-S3/5/9-12. 2004.
- [253] T. Kawakami, M. Inoue, Y. Kobayashi and K. Nakashima, "Application of event related potentials to communication aids," in, pp. 2229-2231. 1996.
- [254] S. Fukuda, D. Tatsumi, H. Tsujimoto and S. Inokuchi, "Studies of input speed of word inputting system using event-related potential," in *the Proc. 20<sup>th</sup> IEEE EMBS Int. Conf.*, vol.3, pp. 1458-1460 ,1998.

- [255] N. Xu, X. Gao, B. Hong, X. Miao, S. Gao and F. Yang, "BCI Competition 2003--Data set IIb: enhancing P300 wave detection using ICA-based subspace projections for BCI applications", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 1067-1072, Jun.2004.
- [256] J. B. Polikoff, H. T. Bunnell and W. Borkowski, "Toward a P300-based computer interface," in Proc. *RESNA '95 Annual Conf.*, pp. 178-180. 1995.
- [257] 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, no.6, pp. 510-523, Dec.1988.
- [258] Cuntai Guan, M. Thulasidas and Jiankang Wu, "High performance P300 speller for brain-computer interface," in Proc. IEEE Int. Workshop Biomedical Circuits and Systems, pp. S3/5/INV-S3/13-16, 2004.
- [259] M. Thulasidas, C. Guan, S. Ranganatha, J. K. Wu, X. Zhu and W. Xu, "Effect of ocular artifact removal in brain computer interface accuracy," in the Proc. 26<sup>th</sup> IEEE EMBS Int. Conf., vol.2, pp. 4385-4388, 2004.
- [260] J. D. Bayliss and D. H. Ballard, "Recognizing evoked potentials in a virtual environment," in Proc. *Advances in Neural Information Processing Systems*, vol.12, 2000.
- [261] J. D. Bayliss and D. H. Ballard, "Single trial P300 recognition in a virtual environment," in Proc. *CIMA'99 (Soft Computing in Biomedicine)*, 1999.
- [262] J. D. Bayliss, "Use of the evoked potential P3 component for control in a virtual apartment", *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 11, no.2, pp. 113-116, Jun. 2003.
- [263] J. D. Bayliss and D. H. Ballard, "A Virtual Reality Testbed for Brain-Computer Interface Research", *IEEE Trans. Rehab. Eng.*, vol. 8, no.2, pp. 188-190, Jun.2000.
- [264] J. D. Bayliss, S. A. Inverso and A. Tentler, "Changing the P300 brain computer interface", *Cyberpsychol Behav.*, vol. 7, no.6, pp. 694-704, Dec.2004.
- [265] E. Donchin, K. M. Spencer and R. Wijesinghe, "The mental prosthesis: assessing the speed of a P300-based brain-computer interface", *IEEE Trans. Rehabil. Eng.*, vol. 8, no.2, pp. 174-179, Jun.2000.
- [266] Jinan Guan, Yaguang Chen, Jiarui Lin, YunYuan and Ming Huang, "N2 components as features for brain computer interface," in Proc. 1<sup>st</sup> Int. Conf. Neural Interface and Control, pp. 45-49. 2005.
- [267] Chang Su Ryu, Yoonseon Song, Done-Sik Yoo, Sangsup Choi, Sung Sill Moon and Jin-Hun Sohn, "EEG-based discrimination between yes and no," in Proc. of the 1<sup>st</sup> Joint EMBS/BMES Conference, vol.1, pp. 444, 1999.
- [268] E. E. Sutter, "The brain response interface: communication through visually-induced electrical brain responses", *J Micro Comp App*, vol. 15, pp. 31-45, 1992.
- [269] Yan Wang, M. T. Sutherland, L. L. Sanfratello and A. C. Tang, "Single-trial classification of ERPS using second-order blind identification (SOBI)," in Proc. Int. Conf. Machine Learning and Cybernetics, vol.7, pp. 4246-4251, 2004.
- [270] P. L. Lee, C. H. Wu, J. C. Hsieh and Y. T. Wu, "visual evoked potential actuated brain computer interface: a brain-actuated cursor system", *Electronic Letters*, vol. 41, no.15, pp. 832-834, Jul.2005.
- [271] J. J. Vidal, "Real-Time Detection of Brain Events in EEG", *Proc IEEE*, vol. 65, no.5, pp. 633-641, 1977.

- [272] Y. Wang, R. Wang, X. Gao and S. Gao, "Brain-computer interface based on the high-frequency steady-state visual evoked potential," in *Proc. 1<sup>st</sup> Int. Conf. Neural Interface and Control*, pp. 37-39. 2005.
- [273] Cheng Ming, Gao Xiaorong, Gao Shangkai and Wang Boliang, "Stimulation frequency extraction in SSVEP-based brain-computer interface," in *Proc. 1<sup>st</sup> Int. Conf. Neural Interface and Control*, pp. 64-67. 2005.
- [274] Yijun Wang, Zhiguang Zhang, Xiaorong Gao and Shangkai Gao, "Lead selection for SSVEP-based brain-computer interface," in *the Proc. 26<sup>th</sup> IEEE EMBS Int. Conf.* vol.2, pp. 4507-4510, 2004.
- [275] M. Cheng, D. Xu, X. Gao and S. Gao, "Brain-computer interface with high transfer rates," in *Proc. 8th Int. Conf. Neural Information Proc. (ICONIP 2001)*, vol.49, pp. 1181-1186, 2001.
- [276] G. L. Calhoun and G. R. McMillan, "EEG-based control for human-computer interaction," in *Proc. 3<sup>rd</sup> Int. Symp. Human Interaction with Complex Systems*, pp. 4-9. 1996.
- [277] M. Cheng, X. Gao, S. Gao and D. Xu, "Design and Implementation of a Brain-Computer Interface With High Transfer Rates", *IEEE Trans. Biomed. Eng.*, vol. 49, no.10, pp. 1181-1186, Oct. 2002.
- [278] M. Middendorff, G. McMillan, G. Calhoun and K. S. Jones, "Brain-Computer Interfaces Based on the Steady-State Visual-Evoked Response", *IEEE Trans. Rehab. Eng.*, vol. 8, no.2, pp. 211-214, Jun.2000.
- [279] M. Cheng and S. Gao, "An EEG-based cursor control system," in *in Proc. of the 1<sup>st</sup> Joint EMBS/BMES Conference*, vol.1, pp. 669-669, 1999.
- [280] S. P. Kelly, E. C. Lalor, R. B. Reilly and J. J. Foxe, "Visual spatial attention tracking using high-density SSVEP data for independent brain-computer communication", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 13, no.2, pp. 172-178, Jun.2005.
- [281] E. C. Lalor, S. P. Kelly, C. Finucane, R. Burke, R. Smith, R. B. Reilly and G. McDarby, "Steady-State VEP-Based Brain-Computer Interface Control in an Immersive 3D Gaming Environment", *EURASIP Journal on Applied Signal Processing*, vol. 2005, no.19, pp. 3156-3164, 2005.
- [282] S. P. Kelly, E. C. Lalor, C. Finucane, G. McDarby and R. B. Reilly, "Visual spatial attention control in an independent brain-computer interface", *IEEE Trans. Biomed. Eng.*, vol. 52, no.9, pp. 1588-1596, Sep.2005.
- [283] S. P. Kelly, E. Lalor, R. B. Reilly and J. J. Foxe, "Independent brain computer interface control using visual spatial attention-dependent modulations of parieto-occipital alpha," in *Proc. IEEE EMBS Int. Conf. Neural Eng.*, pp. 667-670, 2005.
- [284] S. P. Kelly, E. Lalor, C. Finucane and R. B. Reilly, "A comparison of covert and overt attention as a control option in a steady-state visual evoked potential-based brain computer interface," in *the Proc. 26<sup>th</sup> IEEE EMBS Int. Conf.*, vol.2, pp. 4725-4728, 2004.
- [285] N. J. Hill, T. N. Lal, K. Bierig, N. Birbaumer and B. Scholkopf, "Attention modulation of auditory event-related potentials in a brain-computer interface," in *Proc. IEEE Int. Workshop Biomedical Circuits and Systems*, pp. S3/5/INV-S3/17-20, 2004.
- [286] Y. Wang, M. T. Sutherland, L. L. Sanfratello and A. C. Tang, "Single-trial classification of ERPs using second-order blind identification (SOBI)," in *Proc. Third International Conference on Machine Learning and Cybernetics*, Shanghai, China, pp. 4246-4251, 2004.

- [287] A. Bashashati, M. Fatourehchi, R. K. Ward and G. E. Birch, "User customization of the feature generator of an asynchronous brain interface", *Ann. Biomed. Eng.*, vol. 34, no.6, pp. 1051-1060, Jun.2006.

## **CHAPTER 7 PERFORMANCE OF A SELF-PACED BRAIN COMPUTER INTERFACE ON DATA CONTAMINATED WITH EYE BLINKS AND ON DATA RECORDED IN SUBSEQUENT SESSIONS<sup>7</sup>**

### **7.1 Introduction**

A brain computer interface (BCI) system allows individuals to control devices using their brain signals only. A self-paced BCI (SBCI) allows such control to be done at the user's own pace, i.e., at any time he/she chooses. This is unlike traditional synchronized BCI systems where users are only allowed to control the system at certain time intervals. This makes the design of SBCI systems more challenging than synchronized BCI systems. Of particular interests in this study are periods for which the user's control signals are contaminated with physiological artifacts.

Physiological artifacts are undesired potentials that contaminate EEG signals. These artifacts can modify the shape of a neurological phenomenon that drives a BCI system. As a result, artifacts may prevent an SBCI system from correctly recognizing a control command or they may results in the system identifying an artifact-related pattern as an intentional control command (resulting in a false activation). The two physiological artifacts that have been examined most in BCI studies are electrooculogram (EOG) and electromyogram (EMG) artifacts. A number of studies have shown that EOG and EMG activities may generate artifacts that affect the neurological phenomena used in a BCI [1, 2]. These artifacts are often involuntary, and controlling them during signal acquisition is not easy. Therefore, there is a need to avoid, reject or remove them from the EEG signals.

---

<sup>7</sup> A version of this chapter has been submitted for publication. Fatourehchi, M., Birch, G.E., Ward, R.K., "Performance of a Self-paced Brain Computer Interface on Data Contaminated with Eye Blinks and on Data Recorded in Subsequent Sessions", submitted.

Many BCI studies do not report the performance of BCIs had data contaminated with artifacts are used [3-8]. In some studies, data contaminated with artifacts are rejected either manually by a human expert [9-12] or by using an automatic method [13-16]. One shortcoming of automatically rejecting artifacts is that the periods for which the system is not available for control greatly increases. This is because physiological artifacts such as eye blinks happen frequently, and consequently, many epochs are removed by the artifact detection method. The SBCI system is thus not available for control during those periods. Only a few studies have reported the use of an automatic method for removing artifacts [17-19]. See [20] for a detailed review.

An alternative approach to removing artifacts from EEG signals is to design a BCI system that is robust in the presence of artifacts. In that case, the presence of artifacts would not affect the output of the BCI system. The latter approach has not yet received much attention in the BCI literature. One solution to increasing the robustness of an SBCI in the presence of artifacts is for it to simultaneously detect more than one neurological phenomenon. Every neurological phenomenon has its own spatiotemporal characteristics that are more prominent in a particular frequency band. A neurological phenomenon is expected to be robust to artifacts whose frequency contents are concentrated in frequency bands other than that of the neurological phenomenon. As an example, consider movement-related potentials (MRPs). MRPs have low frequency contents  $< 4$  Hz [15], while muscle artifacts usually lie within frequency bands  $> 10$  Hz [1, 21]. It is thus possible to design an MRP-based SBCI system with robust performance in the presence of muscle artifacts. Similarly, Mu and Beta rhythms cover frequencies above 8 Hz, while eye movement artifacts such as eye blinks mostly affect lower frequency components of EEG signals. As a result, BCI systems based on Mu and Beta rhythms are expected to have a robust performance in the presence of EOG artifacts. Based on these arguments, we postulate that an SBCI system that simultaneously uses features extracted from MRPs and changes in the power of Mu/Beta rhythms should be more robust in the presence of muscle and EOG artifacts. To test this hypothesis, we examine the performance of the system proposed in [22]. This SBCI system uses features extracted from three neurological phenomena ( movement-related potentials (MRPs), changes in the power of

Mu rhythms (CPMR) and changes in the power of Beta rhythms(CPBR)) to distinguish between intentional control (IC) commands and No Control (NC) EEG segments.

In [22], a threshold-based detector was used to mark any epoch with EOG amplitude above 25  $\mu$ V as “an epoch contaminated with eye blinks”. The amplitude was measured as the difference between two electrodes, one placed at the eye level and the other below the right eye. The amplitude was conservatively determined after a careful review of the bipolar EOG signals. The performance of the system was tested on data of four able-bodied participants. Using these data, an average TP rate of 56.2% was achieved while the average FP rate was 0.1%. However, only epochs that were not contaminated with eye blink artifacts were used for evaluating the performance.

To test the performance of the system during periods contaminated with artifacts, two studies are carried out in this paper. In the first study, we evaluate the performance of the system on epochs that were originally marked as contaminated with eye-blink artifacts. It should also be noted that no artifact rejection method based on a thresholding scheme is perfect. This is because the artifact detector may fail to detect artifacts with amplitudes lower than the set threshold. The artifact detector we used in [22] (and which has also been used in our previous studies [15, 23, 24]) is no exception.

For this reason, we carry a second study where the performance of the system is tested on data recorded in “subsequent sessions”. In [22], we studied the performance of the SBCI on data collected in five sessions carried using a cross-validation method. For each participant, data were recorded in sessions held on separate days and were collected over a period of approximately 12 days. We reserved the data collected on the last day for the present study. These data, denoted the data recorded in “subsequent sessions”, were recorded one to six days after the data in the first 5 sessions were recorded. No part of the subsequent sessions data was used in the training of the system, so it can be regarded as a complete “blind” test set. To test the performance, all epochs, whether or not contaminated with artifacts, are considered in this study.

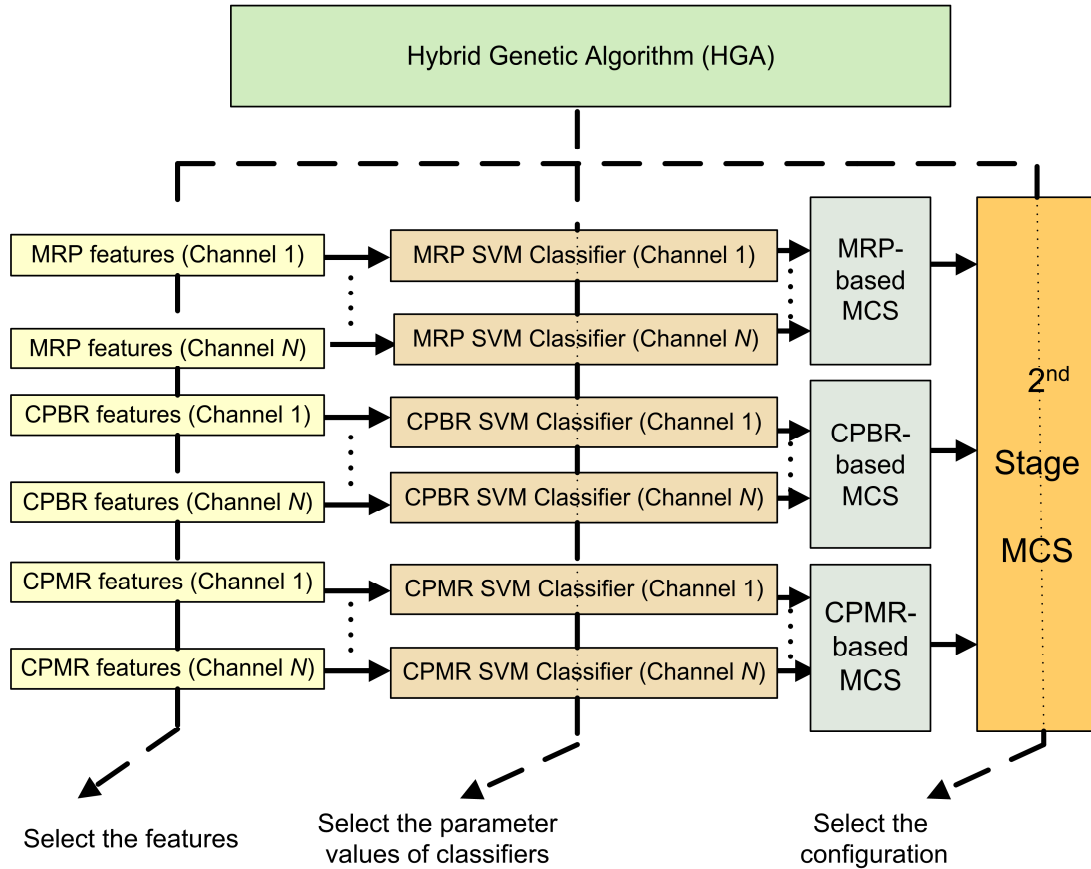
In the next section, the SBCI system, the data collection and the evaluation method are discussed briefly.



## 7.2 Methods

### 7.2.1 Self-paced brain computer interface design

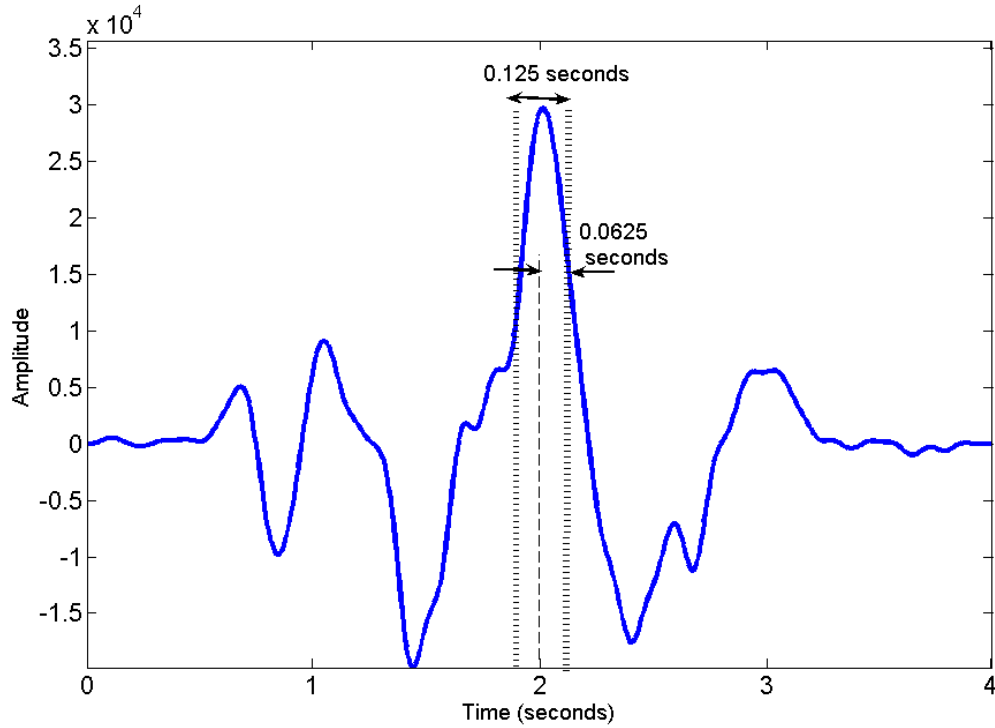
The structure of the SBCI is shown in Figure 7-1. The system uses features extracted from  $N=18$  bipolar EEG signals. For each neurological phenomenon and every EEG channel, features are extracted using a stationary wavelet transform (SWT) with a 5-level decomposition. Feature extraction from Mu and Beta bands involves bandpass filtering and squaring the sample values before applying SWT so that power values are calculated. To reduce the dimensionality of the wavelet feature space, matched filtering (using the cross-covariance function) with a template is performed. The template is created for each neurological phenomenon and for each EEG channel, by averaging the data in the training epochs. After calculating the cross-covariance for each epoch, the features representing the maximum of the cross-correlogram over a period of 0.125 seconds as well as the time this maximum occurred are extracted (see Figure 7-2). This process is only carried out for the lowest approximation and detail level of the SWT decomposition. For each neurological phenomenon and for each EEG channel, a support vector machine (SVM) is designed (resulting in a total of  $3N$  classifiers). The output of each SVM is a logical state '1' when an IC pattern is detected and is '0' in other cases. For each neurological phenomenon, a multiple classifier system (MCS) classifies the outputs of  $N$  SVMs using the majority voting rule. A 2nd-stage MCS combines the outputs of the three MCSs (each MCS is attributed to one neurological phenomenon) to generate the final classification label. A hybrid genetic algorithm (HGA) is then employed. This algorithm simultaneously finds (1) the subset of features, (2) the parameter values for each SVM and (3) the configuration of combining the three MCSs that leads to near optimal performance (measured as the  $\frac{TPR}{FPR}$  ratio). Please see [22], for more details.



**Figure 7-1. The overall structure of the improved SBCI**

### 7.2.2 Data collection

The data were collected from four right-handed (three males and one female) able-bodied participants between 31 and 56 years old. They had all signed consent forms prior to participation in the experiment. The EEG signals were recorded from 13 monopolar EEG channels (according to the International 10-20 system at  $F_1, F_z, F_2, FC_3, FC_1, FC_z, FC_2, FC_4, C_3, C_1, C_z, C_2$  and  $C_4$  locations) and were sampled at 128 Hz. The signals were then converted to bipolar EEG signals since such electrodes are more likely to generate more discriminant MRP features than monopolar electrodes [15]. The conversion was carried out by calculating the difference between adjacent EEG channels and resulted in the generation of the following 18 bipolar EEG:  $F_1-FC_1, F_1-F_z, F_2-F_z, F_2-FC_2, FC_3-FC_1, FC_3-C_3, FC_1-FC_z, FC_1-C_1, FC_z-FC_2, C_1-C_z, C_2-C_4, FC_2-FC_4, FC_4-C_4, FC_2-C_2, FC_z-C_z, C_3-C_1, C_z-C_2$  and  $F_z-FC_z$ .



**Figure 7-2.** An example of extracting the maximum of the cross-correlogram using the proposed cross-covariance method.

During the experiment, the participants performed a right index finger flexion. Epochs of the IC type consisted of data collected over an interval containing the onset of movement (measured as the finger switch activation). The interval started at  $t_{start} = -1$  second, i.e., 1 second before the onset of movement, and ended at  $t_{finish} = 1$ , i.e., 1 second after the onset of movement. For each participant, an average of 80 IC epochs was collected over a period of 6 different days. Table 7-1 shows the timetable of recording the data for all participants. For each participant, “Day 1” was considered as the origin date, and the dates when the rest of the data were collected, were numbered relative to “Day1”.

**Table 7-1. The time schedule of recording the data. For each participant, Day 1 is the first day that a participant attended the experiments. The rest of days are numbered with respect to Day 1 of that particular participant.**

<b>Participant ID</b>	<b>1<sup>st</sup> session</b>	<b>2<sup>nd</sup> session</b>	<b>3<sup>rd</sup> session</b>	<b>4<sup>th</sup> session</b>	<b>5<sup>th</sup> session</b>	<b>6<sup>th</sup> session</b>
<b>AB1</b>	Day 1	Day 3	Day 5	Day 8	Day 10	Day 12
<b>AB2</b>	Day 1	Day 3	Day 4	Day 8	Day 9	Day 10
<b>AB3</b>	Day 1	Day 2	Day 4	Day 8	Day 9	Day 15
<b>AB4</b>	Day 1	Day 3	Day 5	Day 8	Day 10	Day 12

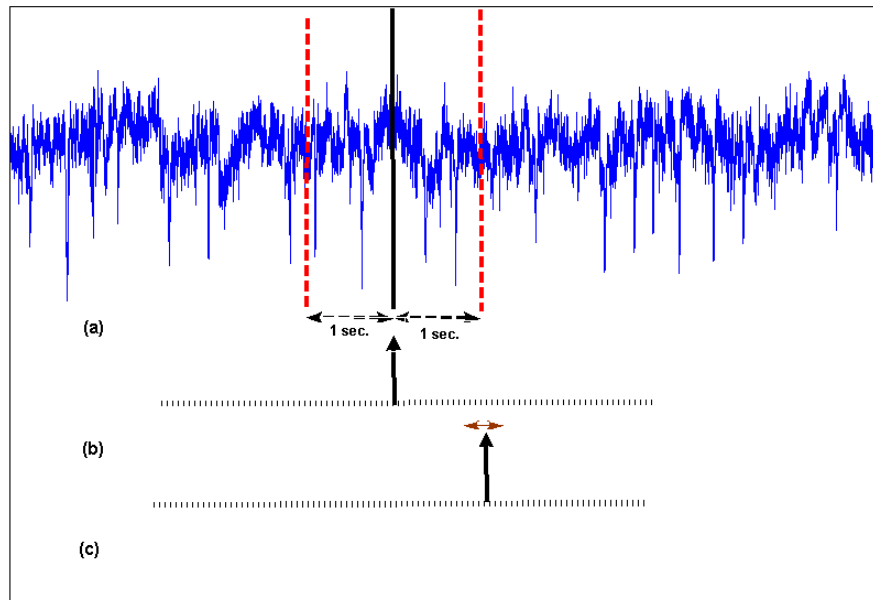
An SBCI should differentiate between IC and NC epochs. For this reason, data in NC sessions are also needed to represent the epochs during which the user did NOT intend to control. During an NC session, participants were asked to count the number of times that a white ball bounced off the monitor's screen. The NC sessions thus contained attentive as well as non-attentive NC data. Each NC session lasted approximately two minutes. During each recording day, up to two such NC sessions were recorded. The NC segments were selected as follows: a window of width ( $t_{\text{finish}} - t_{\text{start}}$ ) seconds was slid over each EEG signal collected during an NC session by a step of 16 time samples (0.1250 sec). For each NC epoch, features were extracted and classified by the SBCI. This resulted in 8 classification decisions per second by the system. For each 1-second window where artifacts were not detected, features were extracted for training the SBCI system. The IC and NC epochs for which the EOG activity exceeded a pre-defined threshold ( $\pm 25 \mu\text{V}$ ) were marked as contaminated with eye-blink artifacts and were not used in the training process.

### **7.2.3 Evaluation**

A five-fold nested cross-validation was used to evaluate the performance of the system. The inner cross-validation set was used for model selection and the outer cross-validation set was used to estimate the generalization error. For each outer cross-

validation set, 20% of the data were used for testing and the rest were used for training and model validation. In order to select the models, the datasets were further divided into five folds. For each fold, 80% of the data were used for training the classifier and 20% were used for model validation.

The method of calculating the TP rate is shown in Figure 7-3. In Figure 7-3(a), a sample EEG signal and in Figure 7-3(b), the output of the physical switch are shown. As stated earlier, data from 1 second before to 1 second after a decision point is used for classification. Assuming the system has no processing delay and the SBCI system has the ideal detection rate, the output of the SBCI system should be as demonstrated in Figure 7-3(c). In other words, the IC command is initiated by the system one second after pressing the switch. Although, the exact timing of the switch activation is known, the neurological phenomena may not be completely time-locked to the switch activation. As a result, we also considered any activation in the time range  $[-0.125, +0.125]$  seconds around the expected activation of the switch as a true positive (see Figure 7-3 (c)). The rest of activations were treated as false activations.



**Figure 7-3. Method of calculating the TP rate; (a) EEG Signal; (b) Output of the hand switch; (c) Output of the SBCI.**

## 7.3 Results

### 7.3.1 Analysis of SBCI performance on artifact-contaminated data

Table 7-2 shows the averages of the TP and FP rates for both non-contaminated and artifact-contaminated data. The averages are calculated over five outer validation sets. The numbers in parentheses show the standard deviations.

**Table 7-2. Comparison of the average test results on artifact-contaminated and non-contaminated data. The averages are calculated over 5 outer validation sets. The numbers in the parentheses indicate standard deviations.**

Participant ID	Test on non-contaminated data		Test on contaminated data	
	TPR (%)	FPR (%)	TPR (%)	FPR (%)
<b>AB1</b>	58.6(8.6)	0.1(0.1)	47.7(7.9)	0.5(0.3)
<b>AB2</b>	64.2 (7.5)	0.0 (0.0)	51.0 (4.0)	0.1 (0.0)
<b>AB3</b>	46.9 (10.4)	0.3 (0.2)	43.7 (4.8)	0.7 (0.2)
<b>AB4</b>	55.0 (5.3)	0.1 (0.1)	64.7 (5.7)	0.4 (0.1)
<b>Mean</b>	56.2 (7.2)	0.1 (0.1)	51.8 (9.1)	0.4 (0.1)

To examine the effect of artifact-contamination on the performance, we carried out a two-way analysis of variance (ANOVA). First, the “TP rate” was considered as the dependant variable and “Participant” and “Artifact Contamination” were considered as the independent variables. As for “Artifact Contamination”, there were two cases: “contaminated” and “non-contaminated”. ANOVA showed a highly significant main effect of “Participant” ( $p < 0.001$ ). The main effect of “Artifact Contamination” was not significant ( $p > 0.05$ ). The average TP rate over all participants was 56.2% for non-contaminated data and 51.8% for artifact-contaminated data. The average TP rate thus

decreased only by 4.4% when the system's performance was tested using artifact-contaminated data.

Next, the "FP rate" was considered as the dependant variable and "Participant" and "Artifact-Contamination" were considered as the independent variables. ANOVA showed a highly significant main effect of "Participant" ( $p < 10^{-4}$ ), and a highly significant main effect of "Artifact Contamination" ( $p < 10^{-4}$ ). The effect of the interaction of both was not significant ( $p > 0.1$ ). The average of FP rates for all four participants for non-contaminated test sets was 0.1% and 0.4% for artifact-contaminated data. As a result, the average FP rate of the system was increased only by 0.3% for contaminated data.

### **7.3.2 Test on data recorded in subsequent sessions**

In [22] and in Section 7.3.1 , we studied the performance of the system on data collected in the first five sessions using a nested cross-validation method. We reserved the data recorded on the last (sixth) session for this study and denoted them as data recorded in "subsequent sessions". For participants AB1, AB2, AB3, and AB4, the "subsequent sessions" data were respectively recorded on days 2, 1, 6 and 2 after recording the data used for designing the system (see Table 7-1).

The "subsequent session" data can be considered as a complete blind test set, as no part of these data was used for designing the system. Furthermore, all epochs of the "subsequent session" data (whether or not contaminated with artifacts) were used in the present study.

The performance of the SBCI when tested on the "subsequent session" data is reported in Table 7-3. The columns represent the TP and FP rates and the rows represent the participant IDs. Two sets of NC data were considered: the NC data collected during the NC sessions, i.e. during the two-minute sessions where movements were not performed, and the NC data collected during the sessions where intentional movements were performed. For the latter data, the NC data were collected for epochs not between movement attempts.

The TPR and FPR results of the system calculated using data recorded in the first five days are also reported in Table 7-3. Please note that these values are the combined results of both artifact-contaminated and non-contaminated data. In [22], the five-fold nested cross-validation analysis resulted in five different sets of features and classifier parameter values for each participant. The results in Table 7-3 are thus shown after averaging over the five outer cross-validation sets for each participant. The numbers in parenthesis are the standard deviations.

**Table 7-3. Comparison of the average results using data recorded in the first five sessions with those using data recorded in subsequent sessions. The averages are calculated over 5 outer validation sets. The numbers in the parentheses indicate standard deviations.**

Participant IDs	Combined test results on the first five session		Test results on the 6 <sup>th</sup> session	
	TPR(%)	FPR(%)	TPR(%)	FPR(%)
<b>AB1</b>	50.3(8.0)	0.5(0.2)	62.9(8.8)	0.3(0.2)
<b>AB2</b>	55.0(5.0)	0.1(0.0)	46.3(5.1)	0.1(0.1)
<b>AB3</b>	44.2(5.7)	0.7(0.2)	46.6(6.0)	0.8(0.2)
<b>AB4</b>	61.2(5.3)	0.4(0.1)	39.7(8.0)	1.8(1.0)
<b>Mean</b>	52.7(7.2)	0.4(0.2)	48.8(9.9)	0.8(0.7)

To further examine the performance of the SBCI, we carried out a two-way ANOVA study. First, the “TP rate” was considered as the dependant variable and “Participant” and “Session” were considered as the independent variables. As for “Session”, there were two values: “Current” and “Subsequent”. We compared the TP rates attributed to classifying epochs in the “Current” test set with those attributed to classifying the epochs in “Subsequent sessions”. ANOVA showed a significant main effect of “Participant” ( $p<0.01$ ), but it did not show a significant main effect of



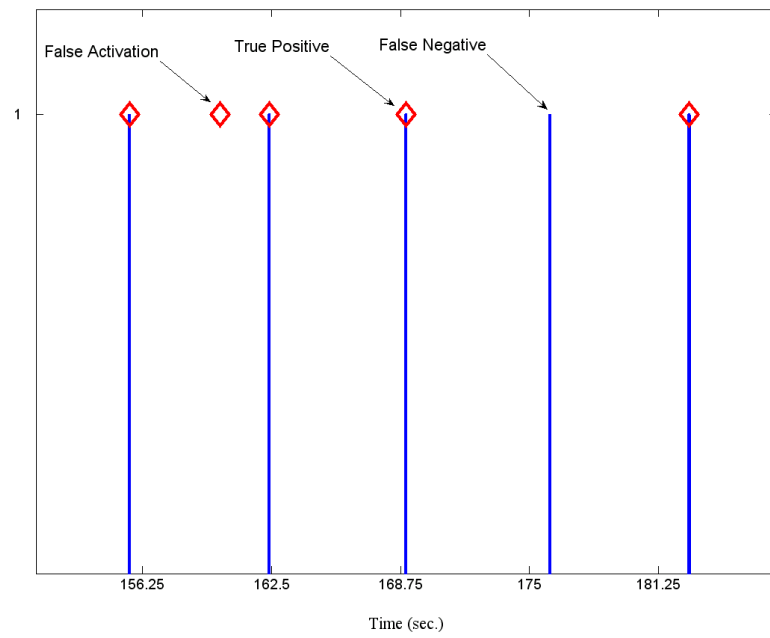
“Session” ( $p>0.05$ ). The average TP rate on the current test sets was 52.7%, and the average TP rate on the data in “Subsequent sessions” was 48.8%.

Next, we compared the FP rates on the “Current” test sets, with the FP rates of the data labeled “Subsequent sessions”. ANOVA showed a highly significant main effect of “Participant” ( $p<10^{-4}$ ), a significant main effect of “Session” ( $p<0.01$ ) and a highly significant effect of the interaction of both ( $p<10^{-4}$ ). The average FP rate on the “Current” test sets was 0.4% and 0.8% on the data in a “Subsequent session”.

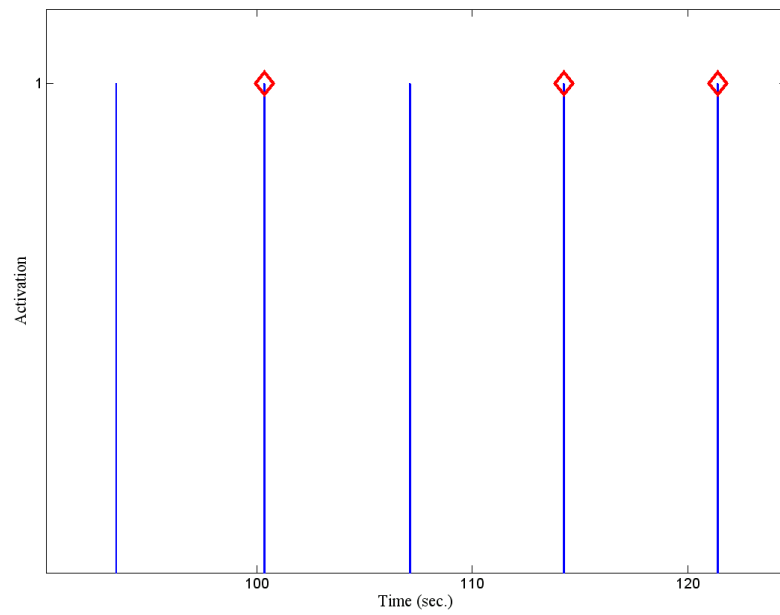
Please note that the average TP rate over participants AB1 to AB3 was 49.8% for the first 5 sessions vs., 51.9% for the 6<sup>th</sup> session, while the average FP rate was 0.4% for these 3 participants for both conditions. These results show that with the exception of AB4, the system did maintain its performance when tested with the new data sets.

We plot the output of the SBCI system for two participants to show that the detection of an IC command did coincide with the onset of the movement (see Figure 7-4). In order to have a clearer picture, the output is plotted for a small representative time duration (around 20-30 seconds). The onset of movement is plotted as a solid line and the output of the SBCI is plotted as a diamond. For one participant (AB1), TP and FN are also shown (see Figure 7-4 (a)). Please note that the  $x$ -axis represents seconds. These results indicate that the SBCI does indeed detect the IC command, since the SBCI detects the pattern around the time of activation of the switch.

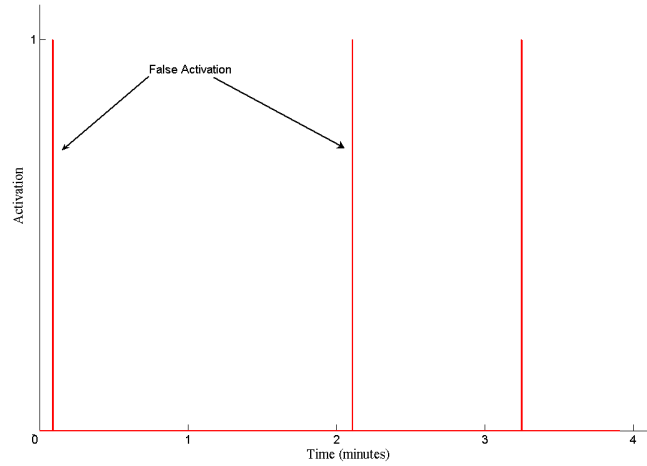
We have also shown the output of the SBCI during NC sessions for a representative user (AB1). For clarity, the output of the SBCI system is plotted as a solid line (see Figure 7-4 (c)). Please note that here the  $x$ -axis shows minutes and not seconds. These plots show that the system is able to maintain an NC state for a long period of NC data. This is a noteworthy advantage as relatively high FP rates are known to be frustrating to users [23].



(a)



(b)



(c)

**Figure 7-4.** The SBCI output during periods when finger movements were executed for a) Participant AB1; b) Participant AB2; and c) The output of the SBCI during NC sessions when movements did not occur for Participant AB1.

### 7.3.3 The effect of adding a debounce component

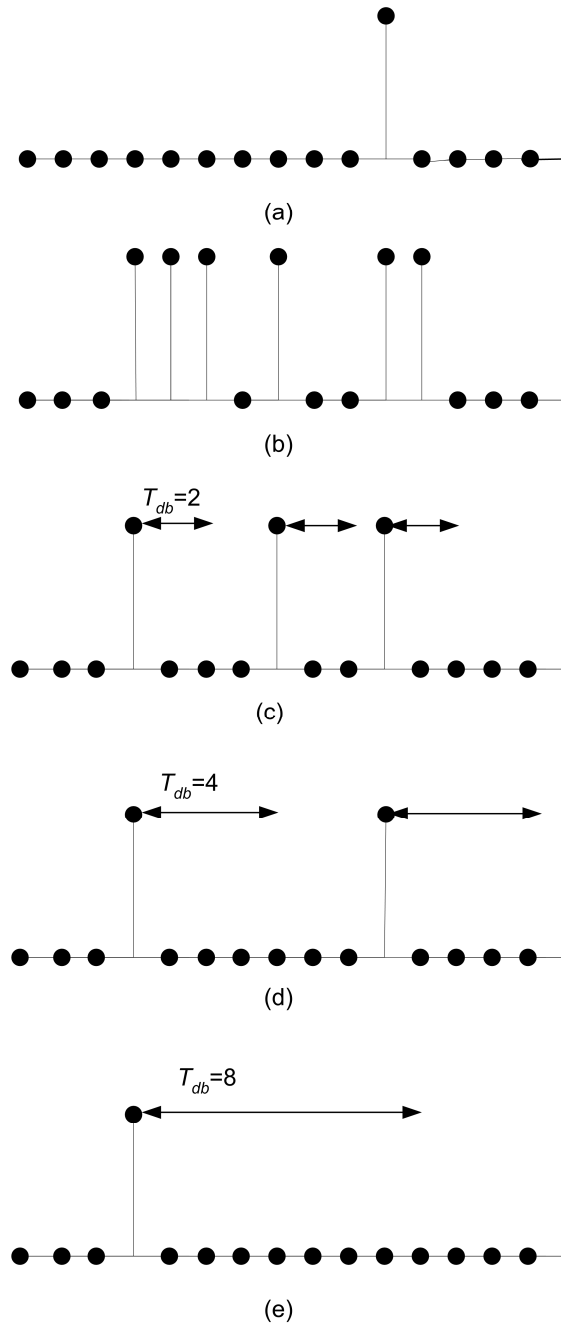
As discussed in Section 7.3.2, when the system was tested on subsequent sessions, its performance deteriorated for participant AB4. To decrease the FP rate, some studies have suggested the use of post-processing such as introducing the concept of refractory periods [25] or a debounce component [23]. In this section, we examine the effect of adding a debounce component to the output of the SBCI.

Debouncing the output of an BCI system in a manner similar to that of debouncing physical switches has been shown to improve the FP rate [23]. The debounce component continuously monitors the output. After an activation is detected (i.e., a change in state from 0 to 1), the output is automatically set to the logical state ‘1’ (IC) for one sample. However, if a debounce window is present, then the output is forced to stay at ‘0’ (NC) for a period of  $T_{db}$  samples, where  $T_{db}$  is the length of the debounce window in time samples. The function of a debounce component is demonstrated in Figure 7-5. Figure 7-5(a) shows the output of the switch. There is one activation towards the end of the window shown in this figure. The points marked by a black circle show the time

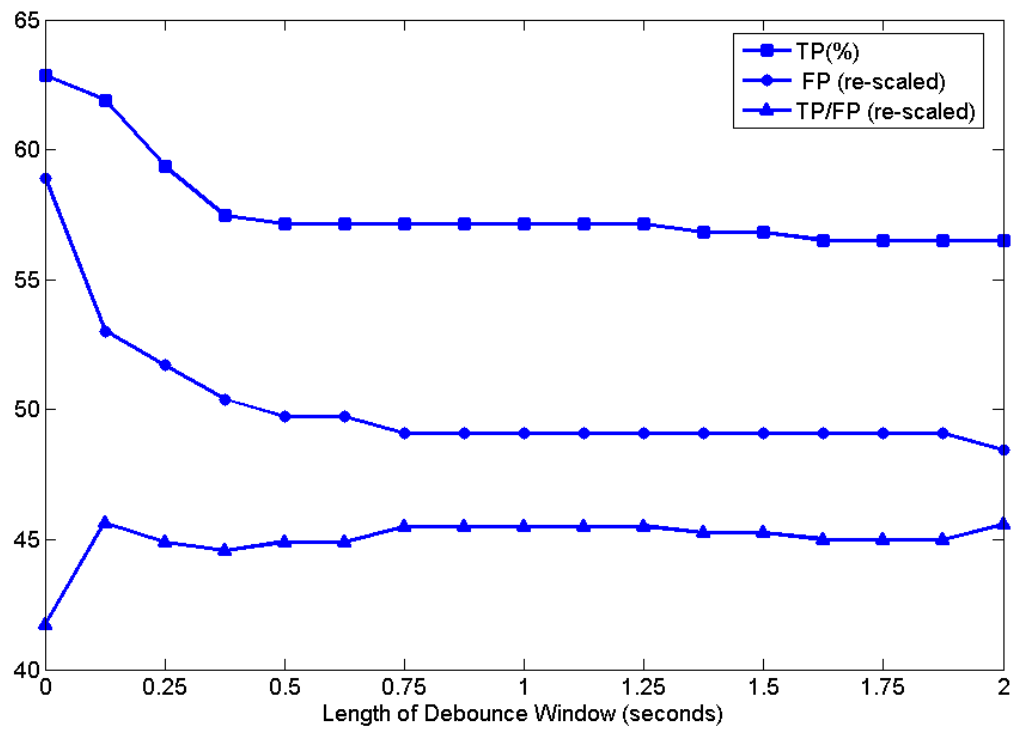
samples. Please note that the time between two consecutive decisions (black circles) by the SBCI system is 0.125 seconds, as the SBCI makes 8 decisions per second [22]. Now consider the output of the SBCI as shown in Figure 7-5 (b). There are many FPs, even though the system was able to correctly detect the IC command. Figure 7-5 (c) shows the output of the system when a debounce component with a length of  $T_{db} = 2$  time samples is added to the system. As can be seen, some FPs are blocked by the debounce component. As the length of the debounce window increases to  $T_{db} = 4$  time samples (Figure 7-5 (d)) the number of FPs sharply drops. However, as Figure 7-5 (e) shows, when the length of the debounce window is  $T_{db} = 8$  time samples, the presence of an earlier FP activation forces the output of the SBCI to have the logical value of '0' for 8 samples. Thus, the TP activation at the end of this epoch is unfortunately no longer detected. Figure 7-5 clearly shows that a tradeoff exists in choosing the length of the debounce window.

We have calculated the performance of the SBCI system when the debounce component is added. Figure 7-6 (a) to Figure 7-6 (d) show the TPR, FPR and  $\frac{TPR}{FPR}$  rates for all participants, respectively, as a function of the length of the debounce window (in seconds). Figure 7-6(e) shows the average plot for all participants. Please note that as the scales of TPR, FPR and  $\frac{TPR}{FPR}$  are different; we had to re-scale the FPR and  $\frac{TPR}{FPR}$  so that all plots can be shown in the same graph. As these figures show, as the length of the debounce window goes from 0 to 1 sample (0.125 seconds), a drop occurs in both TP and FP rates (especially a larger drop in the FP rates). As a result, the  $\frac{TPR}{FPR}$  ratio increases for all participants. These results indicate that a very small debounce window ( $T_{db} = 0.125$  seconds) had a positive effect on the performance of the SBCI system. However, a further increase in the length of the debounce component did not increase the  $\frac{TPR}{FPR}$  ratio by more than the value achieved by  $T_{db} = 1$  output time sample (or 0.125 seconds). This is with the exception of participant AB4, where a larger debounce window improved performance. Thus, we conclude that in this study a minimal debounce window

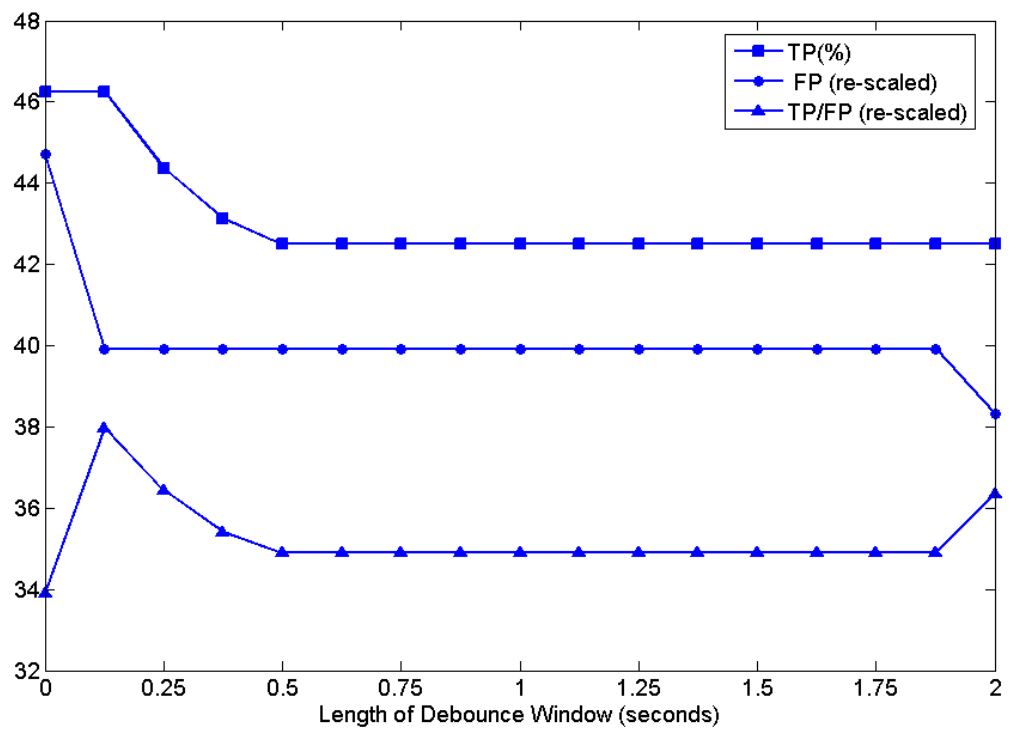
yielded superior results for most participants, but a customized debounce window would further improve the performance of some individuals.



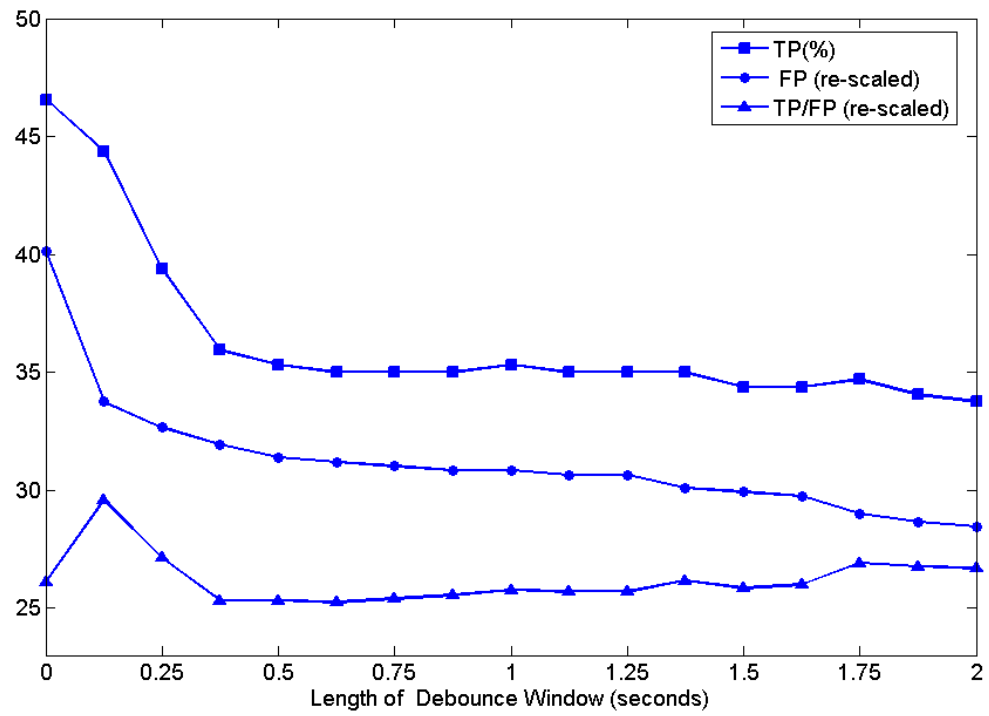
**Figure 7-5. The operation of a debounce component.**



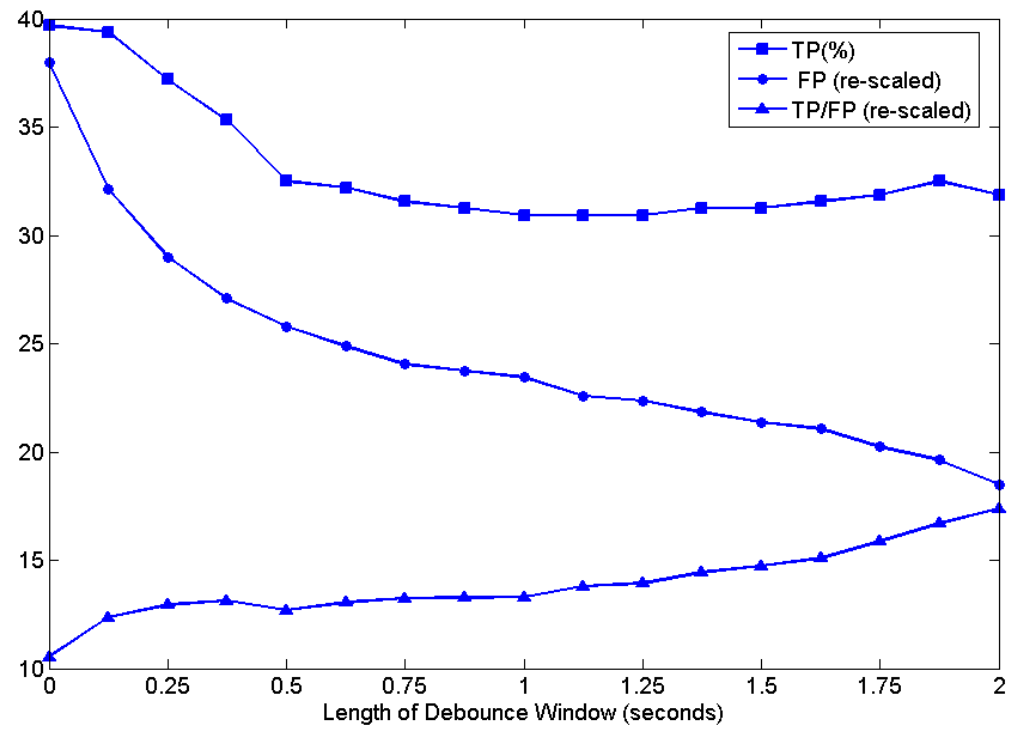
(a)



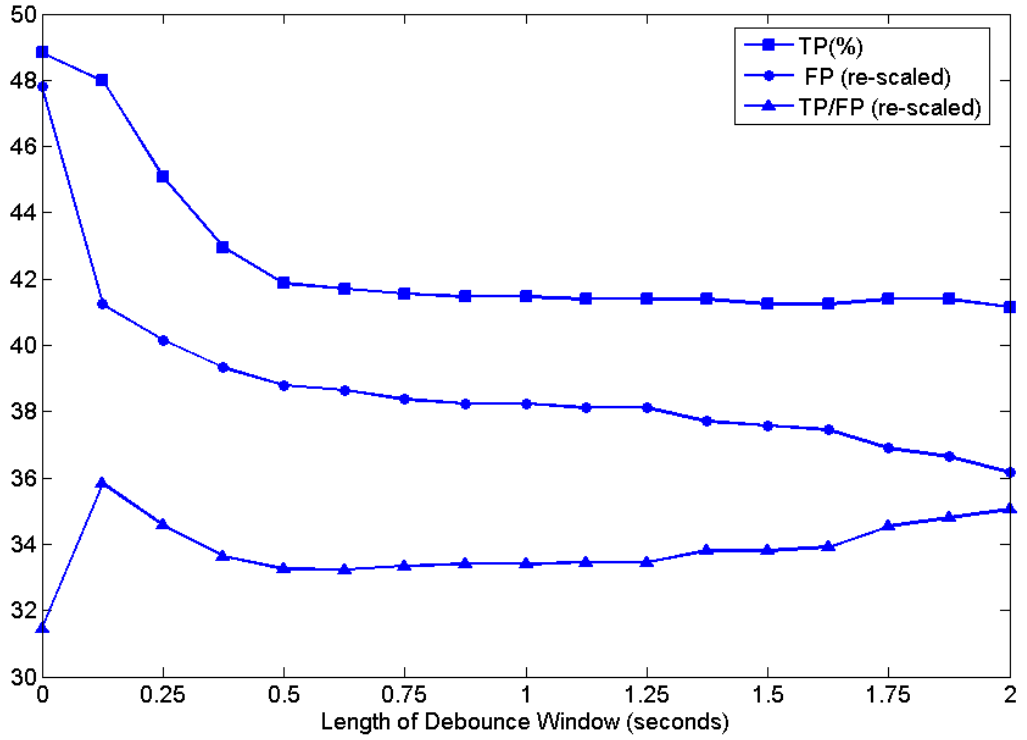
(b)



(c)



(d)



( e )

Figure 7-6. The TP rate, FP rate and the  $\frac{TPR}{FPR}$  ratio as a function of the length of the debounce window for (a) Participant AB1; (b) Participant (AB2); (c ) Participant AB3; (d) Participant (AB4); (e) Averages of all four participants.

## 7.4 Discussion

In this paper, we carry out two studies to further explore the performance of an SBCI system described in [22]. Specifically, we have analyzed its performance on data contaminated with eye-blink artifacts and on data recorded in subsequent sessions. Furthermore, we analyzed the effect of adding a debounce component to the output of the system.

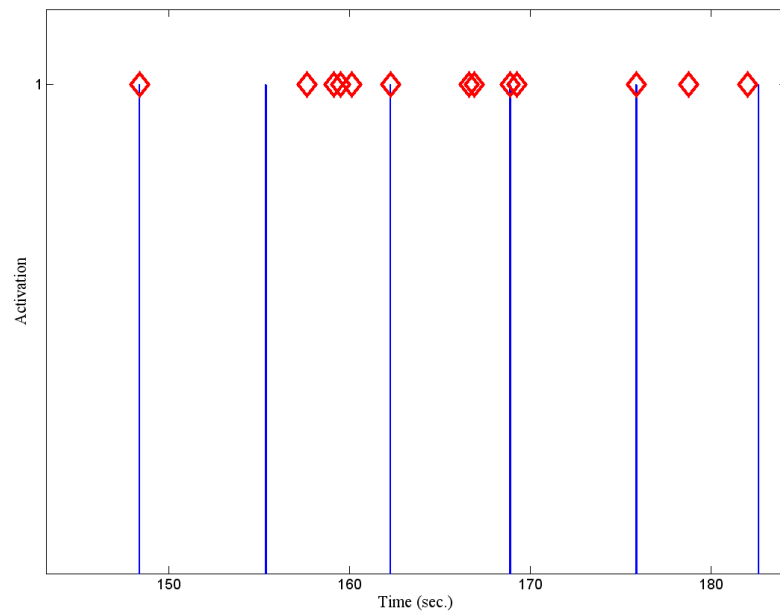
The results of our analysis show that the average TP rate did not show much change when eye blinks artifacts were removed or were present in the test set. The fact that the average TP rate dropped only by 4.4% when artifact-contaminated data were used (from 56.2% to 51.8%) shows further evidence that eye blink artifacts do not have a



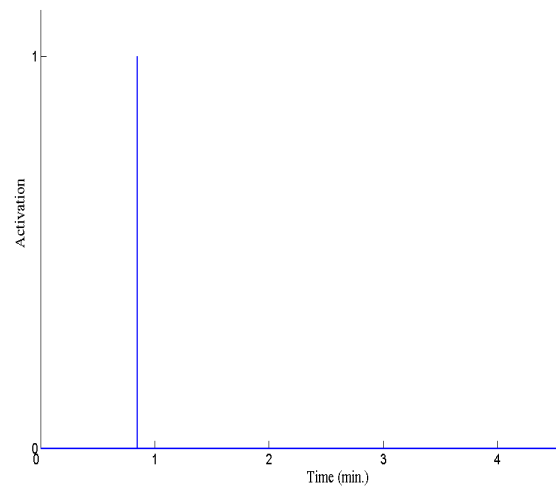
great impact on the detection capability of this system. This robustness in the performance can be attributed to the use of matched filtering with a template but mainly the use of three neurological phenomena for detecting the intentional control pattern.

Although the average FP rate was increased from 0.1% to 0.4%, the resultant values remained lower than the FP rates of some of the recently developed EEG-based SBCI systems with the same amount (or higher) output rates, and with relatively the same TP rate [23, 24, 26-28]. This means that on average this SBCI system generates lower error rates compared to these systems. This change in the FP rate is a trade-off as the system is now available for use at all times, i.e., even when artifacts such as eye blinks occur. While the SBCI system in [22] has a better performance, it is only available during certain time intervals when eye blinks do not occur. Thus, by accepting a moderate decrease in the performance, the system becomes available for control at all times.

It should also be noted that the performance of three participants did not change much when tested on the data of subsequent session (see Table 7-3). This forms preliminary evidence that the system's performance is robust given the fact that event-related potentials (ERPs) may change with time [29-31]. As for participant AB4, we have observed many false positives between successive movement attempts (see Fig.7 (a)), but during NC sessions the FP rate was very low (i.e.,  $FP < 0.1\%$ , see Fig. 7(b)). We suspect the reason why the system had so many FPs between different IC commands is due to the fact that NC data between movement attempts were not used for training the system.. These observations raise the issue of what kind of NC data should be used for optimally training an SBCI system. This investigation is left for future studies.



(a)



(b)

**Figure 7-7. The output of the SBCI during periods when finger movements were executed for participant AB4; (b) the output of the SBCI during NC sessions when movements did not occur for participant AB4.**

Adding a debounce component results in a decrease in the FP rates, since it masks multiple consecutive FPs. Our study showed that a small debounce window is needed for most participants for improving the performance. The optimal size of this debounce window (which was found to be only one output sample) indicates that there were occasions that the system mistakenly identified two consecutive NC epochs as IC. By adding a debounce window, multiple consecutive FPs are treated as one, thus the FP rate decreases. There are two problems with using a larger debounce window. First, if the width of the debounce window is relatively large (e.g., 2 seconds, as recommended in [23]), and if an FP occurs closely prior to an IC command, the IC command is blocked by the debounce component. The TP rate of the system thus drops (see Figure 7-5(e)). Second, as the width of the debounce window increases, periods for which the SBCI system becomes unavailable grow (see Figure 7-5 (d) and Figure 7-5 (e)). This limits the practicality of the SBCI, as there will be long periods in which the system is not available. Our analysis also shows that for the proposed SBCI, using a very short duration of debounce window improves its performance in general. Larger debounce windows, however, might be needed for particular individuals with higher FP rates that occur in a short time frame (e.g., participant AB4, see Figure 7-7(a)).

## **7.5 Acknowledgements**

This work was supported in part by the NSERC under Grant 90278-06 and the CIHR under Grant MOP-72711. The authors would like to thank Mr. Craig Wilson for his valuable comments on this paper.

## 7.6 References

- [1] 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, Sep. 2003.
- [2] D. J. McFarland, W. A. Sarnacki, T. M. Vaughan and J. R. Wolpaw, "Brain-computer interface (BCI) operation: signal and noise during early training sessions," *Clin. Neurophysiol.*, vol. 116, pp. 56-62, Jan. 2005.
- [3] D. J. McFarland, L. M. McCane and J. R. Wolpaw, "EEG-based communication and control: short-term role of feedback," *IEEE Trans. Rehabil. Eng.*, vol. 6, pp. 7-11, Mar. 1998.
- [4] F. Babiloni, F. Cincotti, L. Bianchi, G. Pirri, J. del R Millan, J. Mourino, S. Salinari and M. G. Marciani, "Recognition of imagined hand movements with low resolution surface Laplacian and linear classifiers," *Med. Eng. Phys.*, vol. 23, pp. 323-328, Jun. 2001.
- [5] B. Graimann, J. E. Huggins, A. Schlogl, S. P. Levine and G. Pfurtscheller, "Detection of movement-related desynchronization patterns in ongoing single-channel electrocorticogram," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, pp. 276-281, Sep. 2003.
- [6] A. B. Barreto, A. M. Taberner and L. M. Vicente, "Neural network classification of spatio-temporal EEG readiness potentials," in *Proc. 15<sup>th</sup> Southern Biomed. Eng. Conf.*, pp. 73-76, 1996.
- [7] M. M. Rohde, S. L. BeMent, J. E. Huggins, S. P. Levine, R. K. Kushwaha and L. A. Schuh, "Quality estimation of subdurally recorded, event-related potentials based on signal-to-noise ratio," *IEEE Trans. Biomed. Eng.*, vol. 49, pp. 31-40, Jan. 2002.
- [8] T. Hinterberger, N. Weiskopf, R. Veit, B. Wilhelm, E. Betta and N. Birbaumer, "An EEG-driven brain-computer interface combined with functional magnetic resonance imaging (fMRI)," *IEEE Trans. Biomed. Eng.*, vol. 51, pp. 971-974, Jun. 2004.
- [9] 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, Sep. 2003.
- [10] L. J. Trejo, K. R. Wheeler, C. C. Jorgensen, R. Rosipal, S. T. Clanton, B. Matthews, A. D. Hibbs, R. Matthews and M. Krupka, "Multimodal neuroelectric interface development," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, pp. 199-204, Jun. 2003.
- [11] L. Parra, C. Alvino, A. Tang, B. Pearlmutter, N. Yeung, A. Osman and P. Sajda, "Linear Spatial Integration for Single-Trial Detection in Encephalography," *Neuroimage*, vol. 17, pp. 223-230, 2002.
- [12] M. Schroder, M. Bogdan, T. Hinterberger and N. Birbaumer, "Automated EEG feature selection for brain computer interfaces," in *Proc. 1<sup>st</sup> IEEE EMBS Int. Conf. on Neural Engineering*, pp. 626-629, 2003.
- [13] T. Hinterberger, N. Neumann, M. Pham, A. Kubler, A. Grether, N. Hofmayer, B. Wilhelm, H. Flor and N. Birbaumer, "A multimodal brain-based feedback and communication system," *Exp. Brain Res.*, vol. 154, pp. 521-526, Feb. 2004.

- [14] G. Pfurtscheller, J. Kalcher, C. Neuper, D. Flotzinger and M. Pregenzer, "On-line EEG classification during externally-paced hand movements using a neural network-based classifier," *Electroencephalogr. Clin. Neurophysiol.*, vol. 99, pp. 416-425, Nov. 1996.
- [15] 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.
- [16] G. N. Garcia, T. Ebrahimi and J. -. Vesin, "Support vector EEG classification in the fourier and time-frequency correlation domains," in *Proc. 1<sup>st</sup> IEEE EMBS Int. Conf. on Neural Engineering*, pp. 591-594, 2003.
- [17] S. Kelly, D. Burke, P. de Chazal and R. Reilly, "Parametric models and spectral analysis for classification in brain-computer interfaces," in *Proc. 14<sup>th</sup> Int. Conf. Digital Signal Processing*, vol.1, pp. 307-310, 2002.
- [18] C. I. Hung, P. L. Lee, Y. T. Wu, L. F. Chen, T. C. Yeh and J. C. Hsieh, "Recognition of motor imagery electroencephalography using independent component analysis and machine classifiers," *Ann. Biomed. Eng.*, vol. 33, pp. 1053-1070, Aug. 2005.
- [19] B. J. Culpepper and R. M. Keller, "Enabling computer decisions based on EEG input," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, pp. 354-360, Dec. 2003.
- [20] M. Fatourechi, A. Bashashati, R. K. Ward and G. E. Birch, "EMG and EOG artifacts in brain computer interface systems: A survey," *Clin. Neurophysiol.*, vol. 118, pp. 480-494, Mar. 2007.
- [21] P. V. Komi and P. Tesch, "EMG frequency spectrum, muscle structure, and fatigue during dynamic contractions in man," *Eur. J. Appl. Physiol. Occup. Physiol.*, vol. 42, pp. 41-50, Sep. 1979.
- [22] M. Fatourechi, G. E. Birch and R. K. Ward, "Applying a hybrid genetic algorithm in the design of a self-paced brain interface with a low false positive rate," in *Proc. IEEE ICASSP'07*, vol.4,, pp. IV-1157; IV-1160, Apr. 2007.
- [23] J. F. Borisoff, S. G. Mason, A. Bashashati and G. E. Birch, "Brain-computer interface design for asynchronous control applications: improvements to the LF-ASD asynchronous brain switch," *IEEE Trans. Biomed. Eng.*, vol. 51, pp. 985-992, Jun. 2004.
- [24] M. Fatourechi, A. Bashashati, G. E. Birch and R. K. Ward, "Automatic user customization for improving the performance of a self-paced brain interface system," *Med. Biol. Eng. Comput.*, Vol.44, No.12, pp.1093-1104, Dec 2006.
- [25] G. Townsend, B. Graimann and G. Pfurtscheller, "Continuous EEG classification during motor imagery--simulation of an asynchronous BCI," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 12, pp. 258-265, Jun. 2004.
- [26] A. Bashashati, M. Fatourechi, R. K. Ward and G. E. Birch, "User customization of the feature generator of an asynchronous brain interface," *Ann. Biomed. Eng.*, vol. 34, pp. 1051-1060, Jun. 2006.
- [27] A. Bashashati, S. Mason, R. K. Ward and G. E. Birch, "An improved asynchronous brain interface: making use of the temporal history of the LF-ASD feature vectors," *J. Neural Eng.*, vol. 3, pp. 87-94, Jun. 2006.
- [28] E. Yom-Tov and G. F. Inbar, "Detection of Movement-Related Potentials from the Electro-Encephalogram for possible use in a Brain-Computer Interface," *Medical and Biological Engineering and Computing*, vol. 41, pp. 85-93, Jan. 2003.

- [29] L. T. Mainardi, J. Kupila, K. Nieminen, I. Korhonen, A. M. Bianchi, L. Pattini, J. Takala, J. Karhu and S. Cerutti, "Single sweep analysis of event related auditory potentials for the monitoring of sedation in cardiac surgery patients," *Comput. Methods Programs Biomed.*, vol. 63, pp. 219-227, 2000.
- [30] J. E. Herron and M. D. Rugg, "Retrieval orientation and the control of recollection," *J. Cogn. Neurosci.*, vol. 15, pp. 843-854, Aug 15. 2003.
- [31] E. A. Jones, V. I. Giger-Mateeva, D. Reits, F. C. Riemsdag, B. Liberov and H. Spekrijse, "Visual event-related potentials in cirrhotic patients without overt encephalopathy: the effects of flumazenil," *Metab. Brain Dis.*, vol. 16, pp. 43-53, Jun. 2001.

## **CHAPTER 8   SELECTION OF A SUITABLE EVALUATION METRIC FOR A SELF-PACED BRAIN COMPUTER INTERFACE SYSTEM<sup>8</sup>**

### **8.1 Introduction**

Brain computer interface (BCI) systems allow individuals with severe motor disabilities to communicate with the outside world using their brain signals only. A self-paced BCI (SBCI) is a BCI system that facilitates this communication at the users' own pace, i.e., at any time they wish to do so [1]. An SBCI classifies each epoch in the user's EEG signals as either containing a pattern associated with an intentional control (IC) command or belonging to a no control (NC) state.

Model selection (or model tuning/validation) is the process of finding or adjusting the model parameters for any classification problem. For BCI systems, model selection is a crucial part of the design. This process may include selecting the features, the type of the feature extractor, the classifier, the EEG channels, the neurological phenomenon, the frequency band of interest, the values of the classifier's parameters and the preprocessing and post-processing components. As an example, to find the optimal set of features for a certain BCI, different sets of features are considered. For every set, the performance of the system is calculated and different performances are compared. The set of features that yields the best performance is then selected. The performance of this best model can then be compared with those achieved by similar BCI systems (i.e., systems with the same experimental as well as evaluation protocols). Therefore, the performance of an SBCI must be evaluated in the following two cases, 1) during the model selection procedure and 2) when comparing the performance with other systems.

---

<sup>8</sup> A version of this chapter has been submitted for publication. Fatourehchi, M., Mason, S. G., Ward, R.K., and Birch, G. E., "A New Framework for Comparing Metrics Used in Pattern Classification Problems with Large Test Samples", submitted.

The performance of a BCI with discrete states is usually summarized by a confusion matrix. The (i,j) entry of this matrix represents the number of samples from class i that are classified as belonging to class j. Figure 8-1(a) shows an example of a confusion matrix for a BCI. This system is of the balanced datasets type, i.e., the probability of occurrence of the first class is comparable to that of the second class. A typical example of such system is a “synchronized” BCI system with two IC classes, e.g., classifying right and left index finger flexions. In Figure 8-1, the total number of test samples of  $Class_{IC1}$  and  $Class_{IC2}$  are 87 and 92, respectively. Out of 87 samples that belong to  $Class_{IC1}$ , 76 test samples are correctly classified as  $Class_{IC1}$  and 11 of them are misclassified as  $Class_{IC2}$ . Similarly, 84 out of 92 test samples that belong to  $Class_{IC2}$  are correctly classified as  $Class_{IC2}$ , however, eight samples are misclassified as  $Class_{IC1}$ .

A confusion matrix provides valuable information regarding how well each class is classified by the BCI system. It is, however, not usually straightforward to compare different confusion matrices. Evaluation metrics are thus needed to summarize a confusion matrix into a single value. As an example, in Figure 8-1.(a), the overall classification accuracy (OA) can be used to summarize the confusion matrix, as follows:

$$OA = \frac{\text{Total number of test samples correctly classified}}{\text{Total number of test samples}} = \frac{84 + 76}{179} = 89.39\%$$

For classification problems with balanced datasets such as synchronized BCI systems (where  $prob(Class_1) \approx prob(Class_2) \approx \dots \approx prob(Class_N)$  for an  $N$ -class problem), OA is the most common evaluation metric presently used to summarize the performance [2].



<i>Predicted</i> <i>Actual</i>	<i>Class<sub>IC1</sub></i>	<i>Class<sub>IC2</sub></i>	<i>Total</i>
<i>Class<sub>IC1</sub></i>	76	11	87
<i>Class<sub>IC2</sub></i>	8	84	92
<i>Total</i>	84	95	179

(a)

<i>Predicted</i> <i>Actual</i>	<i>Class<sub>NC</sub></i>	<i>Class<sub>IC</sub></i>	<i>Total</i>
<i>Class<sub>NC</sub></i>	999	0	999
<i>Class<sub>IC</sub></i>	1	0	1
<i>Total</i>	999	0	1000

(b)

<i>Predicted</i> <i>Actual</i>	<i>Class<sub>NC</sub></i>	<i>Class<sub>IC</sub></i>	<i>Total</i>
<i>Class<sub>NC</sub></i>	998	1	999
<i>Class<sub>IC</sub></i>	0	1	1
<i>Total</i>	998	2	1000

(c)

Figure 8-1. a) An example of a confusion matrix for a balanced dataset; b) An example of a confusion matrix for an imbalanced dataset; c) A second example of a confusion matrix for an imbalanced dataset.

The use of OA for problems with highly imbalanced classes (e.g.,  $prob(Class_1) \gg prob(Class_2)$  for a two-class problem) is not satisfactory [3]. An example of such systems is an SBCI system which must detect infrequent IC patterns amongst the more frequent NC epochs. To illustrate why OA is an inadequate metric for such applications, consider the two examples of an SBCI system shown in Figure 1 (b) and Figure 1 (c). Suppose in these examples,  $p(Class_{NC}) \approx 0.999$  and there are 999 test

samples from  $Class_{NC}$  and only one test sample from  $Class_{IC}$  available for testing the performance. In the case shown in Figure 8-1.(b), the SBCI system correctly classifies 998 out of 999 test samples of the NC class ( $Class_{NC}$ ). It also correctly classifies the only test sample in the IC class ( $Class_{IC}$ ), yielding  $OA = 99.9\%$ . Now consider the example shown in Figure 8-1(c). The SBCI correctly classifies all the 999 test samples that belong to  $Class_{NC}$ , but it misclassified the only  $Class_{IC}$  test sample. Again,  $OA = 99.9\%$ . The difference is that the high value of OA is not representative of the true performance of the classification problem, since its performance in detecting  $Class_{IC}$  is 0%.

The choice of the evaluation metric is thus of great importance and is application-dependent. A poorly defined evaluation metric (such as OA illustrated in the above example) may guide the model selection procedure to a far-from-optimal model or it can lead to erroneous conclusions when comparing the performances of two SBCI systems. As a result, all the effort spent in the design of a sophisticated SBCI may become in vain, simply because of the poor choice of the evaluation metric. Recently, the choice of OA as the default evaluation metric has been questioned, even in classification applications with balanced datasets. Specifically, it was shown that for many applications, the area under the receiver operating characteristic (AUC) can summarize the performance better than OA [4].

Although OA is not suitable for classification problems with imbalanced classes, the choice of an alternative evaluation metric is not obvious. Several attempts have been made to define more suitable evaluation metrics for these problems. Examples of such evaluation metrics include weighted overall accuracy (WOA) [5], the use of receiver operating characteristic (ROC) curves and related measures such as area under the ROC (AUC) [6], and Kappa coefficient [7]. In the SBCI literature, some of the evaluation metrics used include overall accuracy [8], HF-difference[9], mutual information (information transfer rate) [10], Kappa [2], AUC [2], the true positive rate (TPR) at a fixed false positive rate (FPR) [11] and  $\frac{TPR}{FPR}$  [12]. Each of these metrics have strengths and weaknesses [2], however, SBCI studies do not usually discuss why a particular evaluation metric is chosen for the evaluation of the performance. This leads to the

obvious conclusion that the study of finding suitable evaluation metrics is important for SBCI systems. This need has been emphasized in a recently published technical report on evaluating SBCI systems [13].

Comparing evaluation metrics is not easy. Consider the case where two evaluation metrics are proposed to summarize two confusion matrices for an SBCI system. As sometimes happens, one evaluation metric yields the conclusion that the confusion matrix  $A$  is better than the confusion matrix  $B$ , while according to the second metric, the opposite case is concluded. This uncertainty causes a problem, as it is unclear which metric should be used to summarize the performance. As a result, a test measure is needed for comparing evaluation metrics.

Comparing two metrics should be carried out directly, i.e., without the need of a 3<sup>rd</sup> metric. In other words, if two evaluation metrics  $f$  and  $g$  are to be compared, and if a third evaluation metric  $h$  is to be used for this comparison, then the validity of the conclusion depends on how valid the third evaluation metric is. Besides, if the evaluation metric  $h$  is a good evaluation metric, why do we need to introduce  $f$  and  $g$ ?

The aim of this study is to discover a general solution for comparing the evaluation metrics that are or could be used for SBCI systems. The proposed solution expands the method in [4] which directly compares evaluation metrics. In [4], two comparison measures [the degree of consistency (DoC) and the degree of discriminancy (DoD)] are defined. Briefly, DoC examines the degree of consistency between two metrics in evaluating the performance of a classification problem. If two metrics are found to be consistent with each other, then DoD is applied to find if one metric is better than the other in discriminating the performance. In [4], DoC and DoD were applied to compare OA and AUC. The conclusion was that AUC is superior to OA. See [4, 14] for more details. In this study, we expand the framework proposed in [4] so that

- 1) It is applicable to other evaluation metrics. We introduce a new measure, named as the “degree of suitability” (DoS) to calculate how suitable an evaluation metric is in summarizing the performance of a classification problem.

2) It can evaluate classification problems with large test samples (independent of whether the class distribution is balanced or imbalanced).

3) It can facilitate the model selection/tuning procedure.

To calculate the DoC and DoD values, we introduce the idea of dividing the [true positive rate (TPR), false positive rate (FPR)] domain into sub-regions and using a grid-based approach. This framework is used to compare selected evaluation metrics in the field of SBCI system (see Section 8.4 for more discussion on the performance metrics studied in this paper). We examine which of these evaluation metrics (if any) is better than others in summarizing the confusion matrix of a particular SBCI. The best evaluation metric can then be chosen as the most suitable one for evaluating the performance of that SBCI. We also examine the effect of class ratio (ratio of NC class to IC class) in comparing metrics. We show that the number of test samples as well as the ratio of test samples are important in reaching the proper conclusion.

## 8.2 Problem statement

Consider a classifier designed for classifying  $M$  classes. Suppose the number of test samples for  $Class_1$  is  $N_1$ , the number of test samples for  $Class_2$  is  $N_2$ , and so on. If  $N_{Total}$  is the total number of test samples, then,

$$N_{Total} = \sum_{i=1}^M N_i = \sum_{i=1}^M \sum_{j=1}^M N_{ij} \quad (8-1)$$

where  $N_{ij}$  is the number of test samples in  $Class_i$  classified as  $Class_j$  ( $i, j = 1, 2, \dots, M$ ).

The performance of the classification problem can be represented by the following  $M \times M$  confusion matrix:

$$\begin{bmatrix} N_{11} & N_{12} & \cdots & N_{1M} \\ N_{21} & & & N_{2M} \\ \vdots & & \ddots & \vdots \\ N_{M1} & \cdots & & N_{MM} \end{bmatrix} \rightarrow \begin{matrix} N_1 \\ N_2 \\ \vdots \\ N_M \end{matrix} \quad (8-2)$$

An evaluation metric  $f(N_{11}, \dots, N_{MM})$  summarizes the confusion matrix described by (8-2) by a single evaluation metric (i.e., by a single value). In the rest of this section, we describe the framework proposed by [4] for comparing evaluation metrics and we address how to modify and generalize the framework so that it could be applied to SBCI systems.

When comparing two evaluation metrics, these metrics should be as consistent with each other as possible. For example, let  $f(a)$  and  $f(b)$  be the values by which metric  $f$  summarizes confusion matrices  $a$  and  $b$ . If  $f(a) > f(b)$ , then for metric  $g$  to be consistent with metric  $f$ ,  $g$  should reach the same conclusion (i.e.,  $g(a) > g(b)$ ). The degree of consistency between  $f$  and  $g$ ,  $\text{DoC}(f, g)$ , counts the number of times when both metrics reach the same conclusion that  $f(a) > f(b)$  as well as  $g(a) > g(b)$ . If for a certain application,  $\text{DoC}$  yields a reasonable consistency between both metrics (i.e.,  $\text{DoC}(f, g) > 0.5$  as explained later in this paper), then  $\text{DoD}(f, g)$ , the degree of discriminancy of  $f$  over  $g$ , is applied to measure how much  $f$  is better than  $g$  in discriminating between the performances. If  $\text{DoC}(f, g) > 0.5$  and  $\text{DoD}(f, g) > 1$ , it can then be stated that metric  $f$  is a better measure for that application than metric  $g$  [4].

Before stating the problem of concern of this study, we need to define three terms that will be used throughout the paper: *(TPR, FPR) Domain*, *Fitness Landscape*, and *Moving on the (TPR, FPR) domain*. Please note that in this paper, for simplicity of presentation we only use the notation for the 2-class SBCI system, however, the definitions and notations can be easily generalized to an  $L$ -class SBCI system.

Assume the “NC” class is the class with the frequently occurring test samples and the “IC” class is the class with the seldom occurring test samples. We denote  $N_{NC}$  and  $N_{IC}$  as the number of test samples in the majority (negative) class and minority (positive) class, respectively. Thus,

$$N_{Total} = N_{IC} + N_{NC} \quad (8-3)$$

Let  $N_{TP}$ ,  $N_{FP}$ ,  $N_{TN}$  and  $N_{FN}$  respectively denote the number of test samples that the SBCI has decided as true positive (TP), false positive (FP), true negative (TN), and false negative (FN) cases. Then the  $2 \times 2$  confusion matrix is  $\begin{bmatrix} TPR & FNR \\ FPR & TNR \end{bmatrix}$  where:

$$TPR = \frac{N_{TP}}{N_{IC}} \quad (8-4)$$

$$FNR = \frac{N_{FN}}{N_{IC}} \quad (8-5)$$

$$FPR = \frac{N_{FP}}{N_{NC}} \quad (8-6)$$

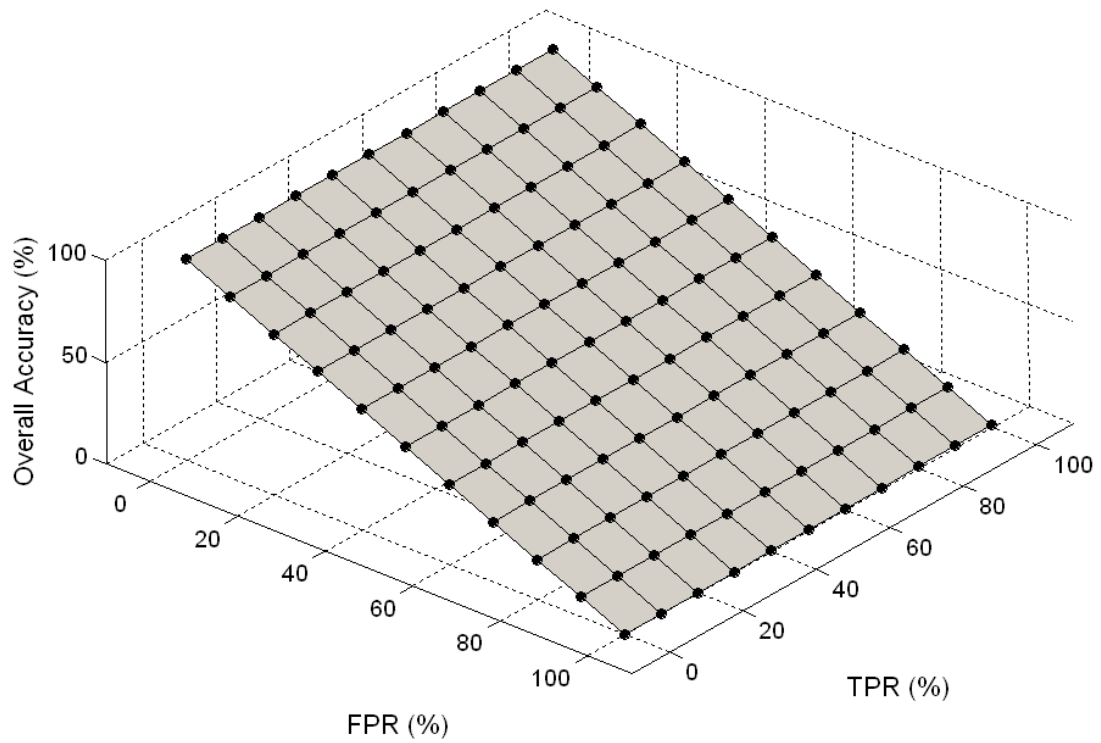
$$TNR = \frac{N_{TN}}{N_{NC}} \quad (8-7)$$

If  $N_{NC}$  and  $N_{IC}$  are known, FNR and TNR can then be directly computed from TPR and FPR. An evaluation metric can then summarize the confusion matrix using the TPR and FPR values only.

*Definition of (TPR, FPR) domain:* consider a 2D space, where one axis denotes TPR and the other axis denotes FPR. We call this domain the (TPR,FPR) domain (see Figure 8-2).

*Definition of fitness landscape:* If an evaluation metric  $f$  is defined over the (TPR,FPR) domain, then the plot of  $f(\text{TPR}, \text{FPR})$  is a 3D surface that we refer to it as the fitness landscape. For specific  $N_{NC}$  and  $N_{IC}$  values, Figure 8-2 shows the fitness landscape of  $OA = \frac{N_{TP} + N_{TN}}{N_{Total}}$  as a function of TPR and FPR. Thus for every (TPR,FPR) point, the fitness landscape of the evaluation metric denotes the value of the evaluation metric at that point.

*Moving on the (TPR, FPR) domain:* We denote a transition path from point  $A = (\text{TPR}_1, \text{FPR}_1)$  to point  $B = (\text{TPR}_2, \text{FPR}_2)$  as a move on the (TPR,FPR) domain.



**Figure 8-2.** A sample fitness landscape for a classification problem with two classes.

To expand the framework in [4] to the more general case, the following points should be considered:

1) ***Suitability of an evaluation metric***: It is necessary that we first investigate if an evaluation metric is indeed suitable for summarizing the performance. In [4], two well-known metrics (OA and AUC) are compared under the assumption that both metrics are suitable evaluation metrics for the applications examined. This assumption, however, may not be applicable to a different metric. Thus, as an initial step it is necessary to investigate whether or not the new evaluation metric is suitable for evaluating the problem at hand. We propose a new measure called the *degree of suitability* (DoS) for this purpose. Once a metric is determined as suitable, it can then be compared with other suitable evaluation metrics.

2) Expansion to classification problems with large test samples (including classification problems with highly imbalanced datasets): To calculate the values of DoC and DoD introduced in the framework proposed in [4] the classification label result of every test sample is needed. As shown in [4], if the number of test samples goes beyond 20, the number of calculations needed for measuring DoC and DoD becomes extremely large, this makes this approach infeasible. Even if the proposed framework uses all the possible configurations of the confusion matrix (instead of all the possible classification labels), it still becomes computationally demanding when the number of test samples is large. This is especially the case for classification problems with highly imbalanced classes such as in SBCI systems. A more general approach is therefore needed for testing the performance of such real-world applications. We propose applying a grid that subsamples the (TPR,FPR) domain so that only representative confusion matrices are used for comparing the evaluation metrics. From the computational point of view, the calculation of comparison metrics such as DoC and DoD using this approach becomes less demanding.

3) *Considering different regions on the (TPR, FPR) domain*: When evaluating two evaluation metrics for model tuning, it is important to consider how to select the region(s) of the (TPR,FPR) domain over which the comparison of evaluation metrics is made. For example, consider the case where we move from a point with a very low TPR and a very high FPR to another nearby point (i.e., a point with also a very low TPR and a very high FPR value). In this case, we may not care much if the metrics are consistent with each other or if one is more discriminant than the other. Now consider the case where we move from a point with a low TPR and a high FPR to a point with a high TPR and a low FPR. In this case we like both metrics to be as consistent with each other as much as possible. We also wish to know if one metric is more discriminant than the other, and hence is more suitable for guiding the model towards better solutions. In this study, we propose a new weighting scheme to emphasize that the comparisons carried over the different regions in the (TPR, FPR) domain can have different degrees of importance. We also generalize the definitions of DoC and DoD so that they are more suitable for comparing the evaluation metrics *during model tuning*.

In the next Section, the framework for comparing evaluation metrics is proposed.



### 8.3 A framework for comparing evaluation metrics

Before describing our proposed framework, we need to introduce another definition, *the Desired Region of Operation (DROP)*. For every SBCI system, the ideal goal is to achieve the perfect classification accuracy (in the 2-class example, this is  $TPR=100\%$ , and  $FPR=0\%$ ). Since this ideal performance is difficult to achieve, more relaxed goals in terms of classification performance are set. For example, these goals can be determined using the designer's expertise or the knowledge in the published material. For a 2-class SBCI, these goals correspond to regions on the (TPR, FPR) domain, where TPR is above a certain pre-defined value and FPR is below another pre-defined threshold. We refer to such a region as the *Desired Region of Operation (DROP)*. In the example shown in Figure 8-3, region  $\psi_1$  is defined by  $TPR \geq TPR_{Threshold}$  and  $FPR \leq FPR_{Threshold}$ .  $TPR_{Threshold}$  usually has a much higher value than the  $FPR_{Threshold}$ .

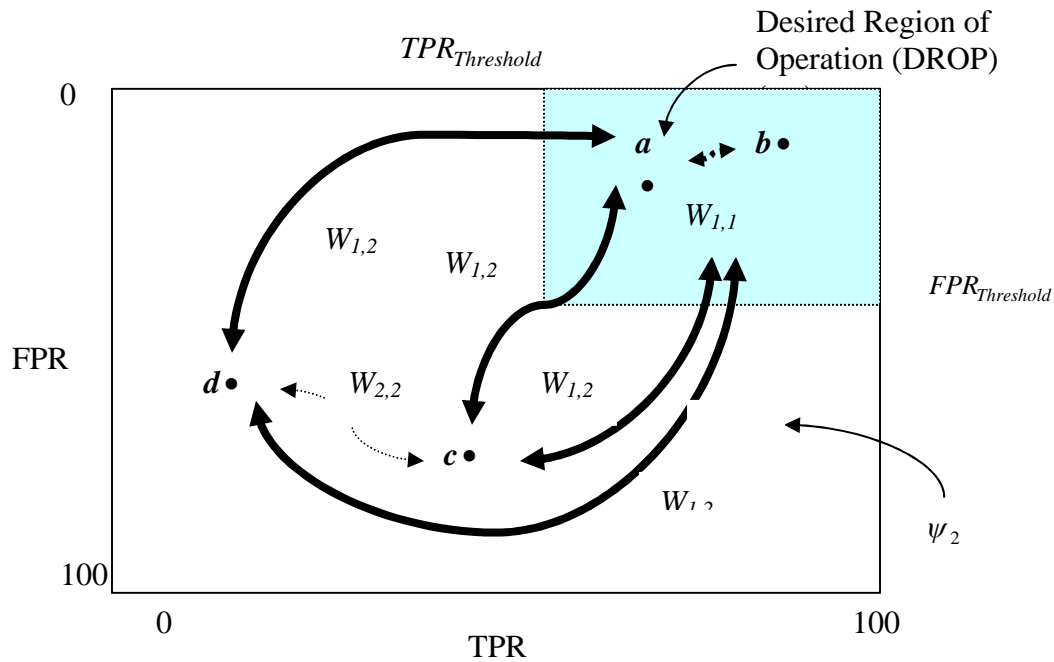
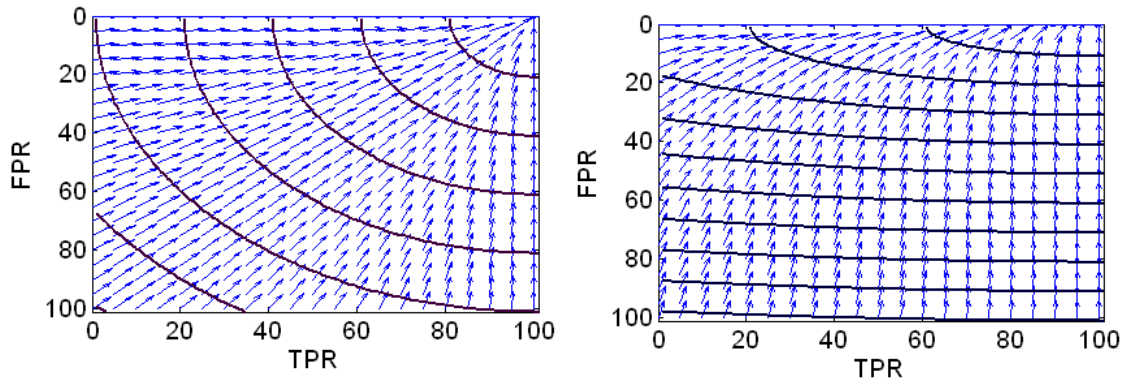


Figure 8-3. An example of dividing the (TPR,FPR) domain into regions. Different movements on the (TPR, FPR) space may be associated with different weights. Note that the numbers on each axis denote (%).

Please note that depending on the problem statement, the (TPR, FPR) domain outside the DROP,  $\psi_2$ , may also be sub-divided into smaller sub- regions (see Figure 8-4

for some examples). However, in this study we focus on the example shown in Figure 8-3.



**Figure 8-4.** Two examples of more complex break-down of the (TPR, FPR) domain with more complex partition and weighting schemes.

In the rest of this section, we present our proposed framework for comparing evaluation metrics. For simplicity, the formulae and discussion are written for a 2-class SBCI with two regions on the (TPR, FPR) domain. The first region is DROP and the other region contains the points outside DROP.

We first introduce “Degree of Suitability” (DoS) of an evaluation metric. Then we evolve DoC and DoD notions so that they can be applied for model tuning of classification problems with large test samples.

### 8.3.1 Suitability of an evaluation metric

If the number of variables contained in an evaluation metric  $f$  is low, the first step in investigating the suitability of  $f$  in evaluating an SBCI system is to visualize  $f$ . This is done by plotting the fitness landscape of  $f$ . This step visually demonstrates the characteristics of the evaluation metric. For example, it reveals if the values of the evaluation metric increase, as the performance of the model move closer to the more desirable points on the (TPR,FPR) domain and if its values decrease as the performance of the model gets closer to the less desirable points. In brief, visualization shows the distribution of the fitness landscape as a function of TPR and FPR.

This approach, while beneficial for the 2-dimensional (TPR, FPR) domain, cannot be used for the general case, as the  $L$ -dimension domain cannot be visualized. Accordingly, another approach which can be used in the general case should be defined. In the rest of this sub-section, we will focus on developing this approach.

We argue that if an evaluation metric attains its higher values inside the DROP region, it is logical to consider it as a suitable metric. In the general case, if the (TPR, FPR) domain is divided into  $K$  regions, then the evaluation metric should have higher values over the regions where it is more desirable to operate in.

To investigate if an evaluation metric indeed has higher values for points located inside the DROP region, we now define the degree of suitability (DoS) of an evaluation metric. In what follows, the definitions related to an SBCI system with two classes will be presented, however, they can be easily generalized to the  $L$ -class problem.

**Definition 1: Degree of suitability (DoS)**

Let  $\psi_1$  and  $\psi_2$  be two non-overlapping sets defined over the (TPR, FPR) domain  $\xi$  such that  $\xi = \psi_1 \cup \psi_2$  (see Figure 3). Let  $\psi_1$  be the DROP region. As a result,  $\psi_1$  is more desirable to operate in than  $\psi_2$ . Let  $f$  be an evaluation metric defined over  $\xi$ , and let  $x = (TPR_1, FPR_1)$  be an arbitrary point in  $\psi_1$  and  $y = (TPR_2, FPR_2)$  an arbitrary point in  $\psi_2$ . The fitnesses of points  $x$  and  $y$  are  $f(x)$  and  $f(y)$ , respectively. Let  $A$  and  $B$  represent all the  $(x, y)$  pairs such that  $A = \{(x, y) | x \in \psi_1, y \in \psi_2, f(x) > f(y)\}$  and  $B = \{(x, y) | x \in \psi_1, y \in \psi_2, f(x) \leq f(y)\}$ . Then DoS( $f$ ) is defined as follows:

$$\text{DoS}(f) = \frac{|A|}{|A| + |B|}, \text{ i.e., } (0 \leq \text{DoS} \leq 1) \quad (8-8)$$

where  $|A|$  measures the number of elements (the cardinality) of the set  $A$ .

DoS specifies the degree to which the evaluation metric yields higher values to points located inside the DROP region than to points outside this region. It is desirable that the resultant value of DoS be as close to unity as possible. Once an evaluation metric is determined as suitable, it can be compared with other suitable evaluation metrics.

During model tuning, it is desired that the evaluation metric guides the model tuning procedure towards the global optimum (or at least towards the DROP region). For example, in Figure 8-3, it is usually desired that the model tuning algorithm moves from points d or c towards a or b. Once the (TPR,FPR) point reaches DROP, it is also important that it keeps moving towards the global optimum (e.g., from point a to point b in Figure 8-3). It should be noted that moving from a to b may not be as desirable as moving from d to a. Likewise, outside the DROP region, it is important that the model tuning procedure does not stall and moves towards DROP (e.g., from point d towards point c). Unfortunately, it is not always straightforward to determine if a particular move on the fitness landscape is a good or a bad one. This is especially the case when there exist local minima/ maxima on the fitness landscape, since moving away from a local minimum or a local maximum might be considered as a good move, only if it guides the algorithm towards a better point in the (TPR,FPR) domain. We can, however, define some guidelines for comparing two metrics. These guidelines are addressed in the next sub-section.

### 8.3.2 Guidelines for comparing two evaluation metrics

It is important to consider the following guidelines when comparing different evaluation metrics:

1) A metric should generally attain a higher value for (TPR, FPR) points inside a region with higher importance (e.g., DROP) than for points located in a region with less importance. Referring to Figure 8-3, consider evaluation metrics  $f$  and  $g$ . Assuming that  $a$ , and  $b$  lie inside the DROP region, and  $c$ , and  $d$  lie outside it. Then it is desirable that  $f$  and  $g$  be such that

$$\begin{aligned} f(a) > f(c), f(a) > f(d), f(b) > f(c), f(b) > f(d), \\ g(a) > g(c), g(a) > g(d), g(b) > g(c), g(b) > g(d). \end{aligned}$$

2) To compare  $f$  and  $g$ , it is important that both metrics behave in a similar fashion, when moving from any arbitrary point to another arbitrary point. It is desirable that both metrics move in the same direction and, if metric  $f$  increases, metric  $g$  increases

as well, i.e.,  $f(a) > f(c) \Leftrightarrow g(a) > g(c)$ . If the metrics are not consistent, it would be very difficult to draw any conclusion regarding their relative performance.

3) It is preferable that the evaluation metric can differentiate between two points lying in the same region. During the model tuning procedure, this will guide the search to move from one point to another and prevents it from getting stuck in the same area. This also means that the evaluation metric can differentiate between the performances of different classification problems that are under examination. As a result, if the value of metric  $f$  changes when moving from point  $a$  to point  $b$  and the value of  $g$  does not, then from the point of view of comparing  $f$  and  $g$ , metric  $f$  is preferred to metric  $g$  [4], as  $f$  is more discriminant. This argument, however, may not be valid when comparing a model from one region to another model from a different region, as explained later in this section.

4) When moving from one region to another, the metric that yields a higher value for a point inside the region with a higher degree of importance is preferable to the metric that yields a lower value, even if the latter can better discriminate between the two points. As an example, in Figure 8-3, if  $f(a) = 10, f(c) = 80, g(a) = 96.0$ , and  $g(c) = 95.0$ , although metric  $f$  better separates the two points, metric  $g$  is more suitable since it assigns a higher fitness to the point located inside DROP.

5) As mentioned in Section 8.2, the comparison between two evaluation metrics depends on the location where the comparison is made. As an example, we may prefer that two metrics are *more* consistent over points belonging to a certain region than over points belonging to another region. This implies that a weighting scheme must be used. The weighting can be done in different ways. It can be discrete (as is shown in Figure 8-3) or there can be continuous weights as shown in Figure 8-4.

Based on these observations, we now modify the notions of DoC and DoD (defined in [4]) for our problem description.

### 8.3.3 Degree of consistency (DoC)

The consistency between two evaluation metrics  $f$  and  $g$  is defined in [4] as follows: If according to metric  $f$ , a point  $a$  on the (TPR, FPR) domain is better than a

point  $b$ , then metric  $g$  should not yield that point  $b$  is better than point  $a$ . If this observation is true for all points on the (TPR, FPR) domain, we can state that metrics  $f$  and  $g$  are *strictly consistent* with each other [4].

Unfortunately, this is not always true. There are many examples for which metric  $f$  increases as we move from point  $a$  to point  $b$  and at the same time, metric  $g$  decreases or remains unchanged. Accordingly, a stochastic version of this definition is proposed in [4].

The original definition of DoC [4] does not consider the cases for which  $f$  or  $g$  have equal values for both points  $a$  and  $b$  (i.e., when  $f(a) = f(b)$  and  $g(a) = g(b)$ ). Thus, using the guidelines from discussed earlier in this section, the definition of DoC can be improved for our problem description as follows:

**Definition 2: Degree of consistency (DoC)**

Let  $\psi_1$  and  $\psi_2$  be two non-overlapping sets defined over the (TPR, FPR) domain  $\xi$  such that  $\xi = \psi_1 \cup \psi_2$  (see Figure 8-3). Let  $f$  and  $g$  be two evaluation metrics and  $x$  and  $y$  be two arbitrary points in  $\xi$ . We define the following sets:

$$R_{(i,j)} = \{(x, y) | x \in \psi_i, y \in \psi_j, \text{sign}((f(x) - f(y)) \times \text{sign}(g(x) - g(y))) = +1\} \quad (i, j = 1, 2) \quad (8-9)$$

$$R'_{(i,j)} = \{(x, y) | x \in \psi_i, y \in \psi_j, f(x) = f(y), g(x) = g(y)\} \quad \text{for } (i, j = 1, 2) \quad (8-10)$$

$$S_{(i,j)} = \{(x, y) | x \in \psi_i, y \in \psi_j, \text{sign}((f(x) - f(y)) \times \text{sign}(g(x) - g(y))) = -1\} \quad (i, j = 1, 2) \quad (8-11)$$

$$S'_{(i,j)} = \{(x, y) | x \in \psi_i, y \in \psi_j, f(x) \neq f(y), g(x) = g(y) \text{ or } f(x) = f(y), g(x) \neq g(y)\} \quad \text{for } (i, j = 1, 2) \quad (8-12)$$

$$R''_{(i,j)} = R_{(i,j)} \cup R'_{(i,j)} \quad (8-13)$$

$$S''_{(i,j)} = S_{(i,j)} \cup S'_{(i,j)} \quad (8-14)$$

$$L_{(i,j)} = \{(x, y) | x \in \psi_i, y \in \psi_j, x \neq y\} \quad \text{for } (i, j = 1, 2) \quad (8-15)$$

$L_{(i,j)}$  is used for normalizing the sets of  $(x, y)$  pairs defined by (8-13) and (8-14). This normalization decreases the degree to which the degree of consistency is dependent on the number of points located in region  $\psi_i (i = 1, 2)$ .

Let  $WR''_{(i,j)}$  and  $WS''_{(i,j)}$  be non-negative weights associated with  $R''_{(i,j)}$  and  $S''_{(i,j)}$ . Using (8-9) to (8-15), DoC(f, g) is defined as follows:

$$\text{DoC}(f, g) = \frac{\sum_{i=1}^2 \sum_{j=1}^2 WR''_{(i,j)} \cdot |R''_{(i,j)}| \cdot |L_{(i,j)}|^{-1}}{\sum_{i=1}^2 \sum_{j=1}^2 WR''_{(i,j)} \cdot |R''_{(i,j)}| \cdot |L_{(i,j)}|^{-1} + \sum_{i=1}^2 \sum_{j=1}^2 WS''_{(i,j)} \cdot |S''_{(i,j)}| \cdot |L_{(i,j)}|^{-1}},$$

$$(WR''_{(i,j)} \geq 0, WS''_{(i,j)} \geq 0, \text{ for } (i, j = 1, 2)) \quad (8-16)$$

Please note that if the restriction  $i \geq j$  is added when calculating (8-16), the comparison between any two regions is carried out only once.

DoC forms an enhanced definition of that in [4]. It compares two metrics based on the number of times they consistently recognize an arbitrary point  $x$  on the (TPR, FPR) domain to be better than another point  $y$ . DoC also incorporates a weighting scheme which is useful when comparing two metrics. For example, during model selection, the metrics are *more* consistent over certain regions (e.g., when moving towards the DROP region from another region), but not as important when moving between or within other regions. The values of  $WR''_{(i,j)}$  and  $WS''_{(i,j)}$  are application-dependent and are chosen according to how the (TPR, FPR) domain is divided and the relative importance assigned to each region.

**Lemma 1-** If two metrics  $f$  and  $g$  defined over the domain  $\xi$  are consistent with each other, then  $\text{DoC}(f, g) > 0.5$ .

**Proof-** If two metrics  $f$  and  $g$  are consistent with each other, then we have:

$$\sum_{i=1}^2 \sum_{j=1}^2 WR''_{(i,j)} \cdot |R''_{(i,j)}| \cdot |L_{(i,j)}|^{-1} > \sum_{i=1}^2 \sum_{j=1}^2 WS''_{(i,j)} \cdot |S''_{(i,j)}| \cdot |L_{(i,j)}|^{-1} \quad (8-17)$$

In this case, it can easily be deducted from (8-17) that  $0.5 < \text{DoC}_{LS} \leq 1$ .

The closer DoC is to '1', the more consistent are the two metrics.

If  $\sum_{i=1}^2 \sum_{j=1}^2 WS''_{(i,j)} \cdot |S''_{(i,j)}| \cdot |L_{(i,j)}|^{-1} = 0$ , the two metrics are strictly consistent and DoC = 1.

If DoC > 0.5, it is possible to determine which metric is more suitable. If DoC ≤ 0.5, this framework cannot be used for comparing evaluation metrics.

### 8.3.4 The Degree of discriminancy (DoD)

Let  $\psi_1$  and  $\psi_2$  be two non-overlapping sets defined over the (TPR, FPR) domain  $\xi$  such that  $\xi = \psi_1 \cup \psi_2$ . Let  $\psi_1$  be the DROP region.  $\psi_1$  is thus more desirable to operate in than  $\psi_2$ . Let  $f$  and  $g$  be two evaluation metrics and  $x$  and  $y$  be two arbitrary points in  $\xi$ . We then define the following sets:

$$U_{(i,j)} = \left\{ (x, y) \mid \begin{array}{l} x \in \psi_i, y \in \psi_j, f(x) > f(y), \text{ and } g(x) \leq g(y) \\ (x, y) \mid x \in \psi_i, y \in \psi_j, f(x) = f(y), \text{ and } g(x) < g(y) \end{array} \right\}, \text{ for } (i, j = 1, 2), (i > j) \quad (8-18)$$

$$U'_{(i,j)} = \left\{ (x, y) \mid x \in \psi_i, y \in \psi_j, f(x) \neq f(y), \text{ and } g(x) = g(y) \right\}, \text{ for } (i, j = 1, 2), (i = j) \quad (8-19)$$

$$V_{(i,j)} = \left\{ \begin{array}{l} (x, y) \mid x \in \psi_i, y \in \psi_j, g(x) > g(y), \text{ and } f(x) \leq f(y) \\ (x, y) \mid x \in \psi_i, y \in \psi_j, g(x) = g(y), \text{ and } f(x) < f(y) \end{array} \right\} (i, j = 1, 2), (i > j) \quad (8-20)$$

$$V'_{(i,j)} = \left\{ (x, y) \mid x \in \psi_i, y \in \psi_j, g(x) \neq g(y), \text{ and } f(x) = f(y) \right\}, \text{ for } (i, j = 1, 2), (i = j) \quad (8-21)$$

$$U'' = U \cup U' \quad (8-22)$$

$$V'' = V \cup V' \quad (8-23)$$

Let  $WU''_{(i,j)}$  and  $WV''_{(i,j)}$  be non-negative weights associated with  $U''_{(i,j)}$  and  $V''_{(i,j)}$ . Using (8-18) till (8-23), DoD(f, g) is defined as follows:

$$\text{DoD}(f, g) = \begin{cases} 1, & \text{if } \sum_{i=1}^2 \sum_{j=1}^2 WU''_{(i,j)} \cdot |U''_{(i,j)}| = 0 \text{ and } \sum_{i=1}^2 \sum_{j=1}^2 WV''_{(i,j)} \cdot |V''_{(i,j)}| = 0 \\ \frac{\sum_{i=1}^2 \sum_{j=1}^2 WU''_{(i,j)} \cdot |U''_{(i,j)}| \cdot |L_{(i,j)}|^{-1}}{\sum_{i=1}^2 \sum_{j=1}^2 WV''_{(i,j)} \cdot |V''_{(i,j)}| \cdot |L_{(i,j)}|^{-1}} & (WR''_{(i,j)} \geq 0, WS''_{(i,j)} \geq 0), \text{ Otherwise} \end{cases} \quad (8-24)$$



Please note that the restriction  $i \geq j$  is added in calculating (8-24), so that the comparison between any two regions is carried out only once.

**Lemma 2-** *Consider two metrics  $f$  and  $g$  defined over domain  $\xi$ . If metric  $f$  is more discriminant than metric  $g$ , then  $\text{DoD}(f, g) > 1$ .*

**Proof-** If  $f$  is more discriminant than  $g$ , then:

$$\sum_{i=1}^2 \sum_{j=1}^2 W U''_{(i,j)} \cdot \left| U''_{(i,j)} \right| \left| L''_{(i,j)} \right|^{-1} > \sum_{i=1}^2 \sum_{j=1}^2 W V''_{(i,j)} \cdot \left| V''_{(i,j)} \right| \left| L''_{(i,j)} \right|^{-1} \quad (8-25)$$

From (8-24), it is easy to conclude that  $\text{DoD}(f, g) > 1$ .

$\text{DoD}(f, g)$  not only captures the cases where one metric fails to differentiate between two points but the other one does not, but also gives a higher score to a metric that yields higher values for the points located in a region with more importance.

### 8.3.5 Comparison of two evaluation metrics

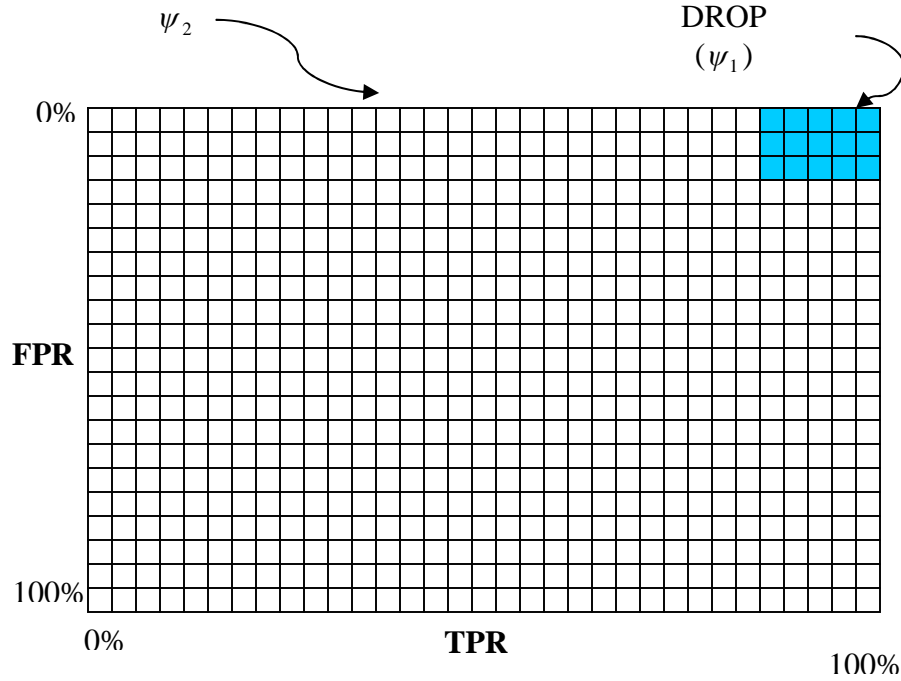
The following definition determines if metric  $f$  is superior to metric  $g$ :

**Definition 4:** - Let  $\psi_1$  and  $\psi_2$  be two non-overlapping sets defined over the (TPR, FPR) domain  $\xi$  such that  $\xi = \psi_1 \cup \psi_2$ . Let  $f$  and  $g$  be two evaluation metrics defined on the domain  $\xi$ . If  $\text{DoC}(f, g) > 0.5$  and  $\text{DoD}(f, g) > 1$ , then metric  $f$  is said to be superior to metric  $g$  on the domain  $\xi$ .

It is evident that the higher the values of  $\text{DoC}(f, g)$  and  $\text{DoD}(f, g)$ , the stronger our confidence will be regarding the superiority of metric  $f$  over metric  $g$ .

### 8.3.6 Using sub-sampling grids for calculating the comparison measures

The most reliable approach for comparing two evaluation metrics is to generate all the possible classification outcomes in a confusion matrix and then calculate the DoS, DoC and DoD measures using all the points on the (TPR, FPR) domain. As an example, consider a case where the number of IC test samples is 10 and the number of NC test samples is 100. The number of TPs can thus vary between 0 and 10 and the number of



**Figure 8-5. An example of using grids on the (TPR, FPR) domain.**

FPRs can vary between 0 and 100. As a result, there are  $11 \times 101 = 1111$  points on the (TPR, FPR) domain corresponding to  $11 \times 101 = 1111$  confusion matrices. For the calculation of DoC and DoD, all these points should be compared with each other, so roughly 617,000 comparisons should be made in order to calculate the DoC or DoD measures. For the case when the number of test samples is large, calculating these measures becomes computationally demanding. We thus propose the use of *sub-sampling grids* to decrease the number of computations. In this approach, the (TPR, FPR) domain is evenly divided into smaller grids (see Figure 8-5 for a typical example). The number of nodes can be determined based on the available computational resources. The DoC and DoD values (equations (8-16) and (8-24)) are then calculated for the points belonging to each point located on the grid, instead of calculating them for each point generated by the confusion matrix. For example in the above example, we can use every other sample for calculating DoC and DoD. As the result, the values of TP will be 0, 2, ..., 10 and the values of FP will be 0, 2, 4, ..., 100. Accordingly, (TPR, FPR) domain now has  $6 \times 51 = 306$  nodes (instead of 1111 points in the first example). The number of calculations thus becomes to be 46,600, which is more than 13 times smaller than the original case. If the results from

both approaches are consistent with each other, the second approach is more desirable as it is computationally more efficient.

To further demonstrate that the grid-based approach significantly reduces the number of comparisons needed (especially for calculating DoC and DoD), suppose we denote the number of points in the  $\xi$  by  $N_{points}$  and the number of points in the grid by  $N_{nodes}$ . Since calculating DoC and DoD involves comparing every two points on the grid, then the number of comparisons needed to obtain the DoC(f,g) or DoD(f,g) values is:

$$N_{Grid} = \frac{N_{nodes} \cdot (N_{nodes} - 1)}{2} \quad (8-26)$$

The number of comparisons needed to calculate DoC(f,g) or DoD(f,g) values for every point in  $\xi$  is:

$$N_{NonGrid} = \frac{N_{points} \cdot (N_{points} - 1)}{2} \quad (8-27)$$

Let the ratio of the number of confusion matrix configurations to the number of nodes be defined:

$$K = \frac{N_{points}}{N_{nodes}} \quad (8-28)$$

Using (8-26) to (8-28), the ratio of the number of calculations in the non-grid approach to the grid approach is calculated as follows:

$$\frac{N_{NonGrid}}{N_{Grid}} = \frac{\frac{N_{points} \cdot (N_{points} - 1)}{2}}{\frac{N_{nodes} \cdot (N_{nodes} - 1)}{2}} = \frac{KN_{nodes} \cdot (KN_{nodes} - 1)}{N_{nodes} \cdot (N_{nodes} - 1)} \cong K^2 \quad (N_{nodes} \gg 1) \quad (8-29)$$

If  $K > 1$ , then the non-grid approach needs to calculate  $K^2$  comparisons more than the grid approach. This suggests that if the resolution of the grid-based approach is coarser than the resolution of the (TPR, FPR) domain (which is usually the case), then much less computations is required. As a simple example, consider the case, where  $N_{NC} = 10^3, N_{IC} = 10^2, N_{nodes} = 10^2$ . The number of possible values for the confusion

matrix is then calculated as  $N_{points} = N_{NC} \times N_{IC} = 10^5$ . As a result,  $\frac{N_{NonGrid}}{N_{Grid}} \approx 10^6$ . The grid-based approach thus reduces the number of calculations significantly.

In the next section, we briefly review the metrics that will be compared in this study using the proposed framework.

## 8.4 Selected evaluation metrics in SBCIs

In this Section, we describe some of the current evaluation metrics in SBCI systems (for a more detailed discussion see[2]). For simplicity, in the rest of this Section, we focus on SBCI systems with only two classes. Let  $N_{TP}$ ,  $N_{FP}$ ,  $N_{TN}$  and  $N_{FN}$  respectively denote the number of epochs an SBCI system has determined as to belong to the true positive(TP), false positive (FP), true negative (TN), and false negative (FN) cases.

Then the confusion matrix is represented as  $\begin{bmatrix} N_{TP} & N_{FP} \\ N_{FN} & N_{TN} \end{bmatrix}$  and TPR, FPR, TNR and FNR are defined as in equations (8-4) to (8-7).

If  $N_{IC}$  and  $N_{NC}$  are known, FNR and TNR can be directly computed from TPR and FPR only. The evaluation metric can thus summarize the confusion matrix using TPR and FPR values only. We now briefly review selected evaluation metrics used in SBCI systems. It should be noted that evaluation metrics described in Sections 8.4.1 to 8.4.3 below, are used for synchronized as well as self-paced BCI systems. However, metrics described Sections 8.4.4 to 8.4.6 have been mostly used for SBCI systems.

### 8.4.1 Overall accuracy (OA)

OA shows the total number of test samples correctly classified by an SBCI system. For a 2-class SBCI, OA is calculated as follows:

$$OA = \frac{N_{TN} + N_{TP}}{N_{IC} + N_{NC}} \quad (8-30)$$

OA has been frequently used in evaluating many synchronized BCI systems [15-17]. Its use in SBCIs, however, has so far been limited [8]. This is because, for an SBCI system, OA assigns a huge weight on the more frequent class (NC) and a very small

weight on the less frequent classes (IC). This may lead to misleading conclusions about the performance of the system, as shown by the examples in the Introduction. Nevertheless, since in general OA is the most obvious evaluation metric for classification problems, it is thus investigated in this paper.

#### **8.4.2 Information transfer rate (mutual information)**

The information transfer rate (ITR) has been specifically proposed for evaluating the performance of synchronized BCI systems [18]. Based on the similarities between an SBCI and a communication channel, [18] used Shannon's communication theory and proposed this metric. The rationale is as follows: ITR measures the amount of information transferred between two reference points. The output  $\mathbf{Y}$  of an SBCI is the interpretation (information) of the current state of the brain, and  $\mathbf{Y}$  conveys this information to downstream components. It was thus argued in [18] that the amount of information in  $\mathbf{Y}$  is a useful tool for comparing the results obtained from different synchronized BCI designs.

The original ITR formula has some shortcomings that make it unsuitable for SBCI systems [19]. Based on Shannon's theorem, a more complex formula for the calculation of ITR is proposed in [20] and [10]. It is shown that the proposed measure yielded a better estimation of ITR for several synchronized BCI systems than the definition proposed in [18]. Later, the ITR of an SBCI with two states is calculated and analyzed in details [19]. It is argued that ITR by itself is "not" a suitable single evaluation metric for an SBCI system. This is because of the unique nature of this metric that has more than one maximum (see [19] for a detailed discussion). ITR (or mutual information) is thus not considered as an evaluation metric in this paper.

#### **8.4.3 Kappa**

Cohen's Kappa coefficient is a measure of agreement between two estimators [21]. Since it considers chance agreements, it is regarded as a more robust measure than OA [2]. Kappa's coefficient is calculated as [21]:

$$\kappa = \frac{OA - p_e}{1 - p_e} \quad (8-31)$$

where  $p_e$  is the chance agreement and is defined as follows:

$$p_e = \frac{\sum_{i=1}^M N_{:i} N_{i:}}{(N_{Total})^2} \quad (8-32)$$

Here,  $N_{:i}$  is the sum of the elements in the  $i$ -th column and  $N_{i:}$  is the sum of the elements in the  $i$ -th row in the confusion matrix.

#### 8.4.4 HF-difference

The HF-difference is a newly proposed metric that summarizes the confusion matrix [22]. It is defined as

$$HF = H\% - F\%; \quad (8-33)$$

where  $F\%$  is the percentage of total activations that are incorrect (false discovery rate [23]) and  $H\%$  is the percentage of the total number of IC commands that are correctly detected by the system ( $TPR = \frac{N_{TP}}{N_{IC}}$  or Hit rate).  $F$  is calculated as follows:

$$F = \frac{N_{FP}}{N_{TP} + N_{FP}} \quad (8-34)$$

The advantage of using HF-difference is that it is sensitive to the ratio of FPs to the total number of detections. The downside of using the HF-difference is that no information about the length of NC periods has been incorporated in calculating (8-33).

#### 8.4.5 $\frac{TPR}{FPR}$ ratio

The  $\frac{TPR}{FPR}$  is another evaluation metric that was recently proposed for 2-class SBCI systems [12, 24]. This metric gives more weight to cases with low FPRs. As a result, during the model tuning process, any model with a high FPR is assigned a low

fitness, even though TPR might have a high value. The downside is that for  $FPR=0$ , the system cannot differentiate between confusion matrices with different TPRs.

#### 8.4.6 ROC curve and related metrics

The receiver operating characteristics (ROC) curve is a popular metric for evaluating systems with imbalanced classes. The ROC curve depicts the relationship between TPR and FPR. Popular methods that use the ROC curve for measuring the performance employ one of the following two criteria 1) The area under the ROC curve (AUC) which is used as the fitness of the system [2]; 2) Defining a critical FPR value ( $FPR_{Critical}$ ), and then using the value of the TP rate at  $FPR_{Critical}$  as the fitness [11]. The advantage of using the ROC curve over previous metrics is that a whole range of solutions (in terms of a tradeoff between TPR and FPR) is provided.

One problem with using the ROC curve is that when it is plotted over the whole range of TPR and FPR, most SBCI systems produce a curve that is similar to a perfect ROC curve [13]. One solution is to restrict the ROC curve to a small region corresponding to low FPRs. The other problem (and perhaps more important) with using the ROC curve is that it is computationally more demanding than other evaluation metrics. Several points need to be evaluated until a partial ROC curve that is accurate enough for estimating the AUC is drawn. Similarly, several points need to be calculated in order to obtain the value of TPR at  $FPR_{Critical}$ . Even if ROC is estimated using the more computationally efficient algorithm as described in [4], it remains much more time consuming than the metrics described above as these only need the value of a single point to assess the performance. When these metrics are used to evaluate the performance and select a model from thousands of confusion matrices during a model selection procedure, the computational burden becomes problematic. For these reasons, evaluation metrics that summarize the performance based on a single evaluation of a confusion matrix are more desirable and are thus considered in this study.

## 8.5 Simulations

In this section, we use the framework described in Section 8.3 to compare selected evaluation metrics employed in evaluating SBCI systems. The aim is to analyze and find which (if any) of these metrics is more suitable for an SBCI system. The evaluation metrics studied here are: overall accuracy (OA) [8], Kappa [2], HF- difference [22] and  $\frac{TPR}{FPR}$  [12].

### 8.5.1 Application

Consider an SBCI system with two output states: IC and NC. We consider three values for the number of IC epochs ( $N_{IC} = 10, 50$ , or  $100$ ). As for the NC class, we assume that  $N_{NC}$ , the number of NC epochs, varies between  $100$  to  $500$  in a step of  $100$ . Thus,  $1 \leq \frac{N_{NC}}{N_{IC}} \leq 50$ .

The (TPR,FPR) domain is divided into two regions: a) the DROP region, formulated as  $\psi_1 = \{(TPR, FPR) | TPR \geq 50\%, \text{ and } FPR \leq 2\%\}$  and b) the rest of the (TPR,FPR) domain, formulated as  $\psi_2 = \{(TPR, FPR) | TPR < 50\%, \text{ or } FPR > 2\%\}$ . The reason the DROP region was chosen as above is because the solutions found in a previous SBCI study lied in this region [11].

### 8.5.2 Results

#### Applying a grid to the (TPR,FAR) domain

When using the finest grid resolution, the resolution is  $(\frac{1}{N_{IC}}, \frac{1}{N_{NC}})$ . This is the case when all the possible configurations of the confusion matrix are studied. We studied 10 different resolutions for comparing the evaluation metric. The resolutions varied between  $(\frac{1}{N_{IC}}, \frac{1}{N_{NC}})$  to  $(\frac{10}{N_{IC}}, \frac{10}{N_{NC}})$  with a step of 1. If the results obtained by using different resolutions are consistent, then coarser resolutions can be used to approximate



the values of DoS, DoC and DoD. The most suitable evaluation metric can thus be selected with less number of computations.

### Calculating DoS

DoS values are shown in Table 8-1. The columns denote the evaluation metric. For each evaluation metric, three sub-columns are presented: *Res1* stands for the finest resolution when all the possible combinations of the confusion matrix are calculated. *Mean* stands for the average resolution, i.e., over the 10 considered resolutions. Finally, *Res10* stands for the coarsest resolution, i.e., when one out of every 10 points is used for calculating DoS. The rows show different  $N_{IC}$  and  $N_{NC}$  configurations (15 configurations in total).

As seen from Table 8-1, all metrics have  $\text{DoS} > 0.9$ . Especially when the classes are imbalanced (except for the case where  $N_{IC}=100$  and  $N_{NC}=100$ ),  $\text{DoS} > 0.98$ . This shows that the points inside the defined DROP region have higher fitness values compared to the points located in the other region. Furthermore, comparing the results obtained from *Res1* with the average of other resolutions and even with the coarsest resolution shows a high degree of consistency. Since by decreasing the resolution, the amount of computational complexity significantly decreases, these findings are very important. To be more specific, by using the coarsest resolution, we could reach the same conclusion as with using the highest resolution. However, the amount of computation will be significantly less.

**Table 8-1. DoS results for the evaluation metrics studied in this paper. *Res1* stands for the finest resolution, *Mean* stands for the average of 10 resolution values and *Res10* stands for the coarsest resolution.**

		OA			HF			$\frac{TPR}{FPR}$			Kappa		
		Res 1	Mean	Res 10	Res 1	Mean	Res 10	Res 1	Mean	Res 10	Res 1	Mean	Res 10
$N_{IC} = 10$	$N_{NC}=100$	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00
	$N_{NC}=200$	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$N_{NC}=300$	1.00	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$N_{NC}=400$	1.00	1.00	1.00	0.98	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$N_{NC}=500$	1.00	1.00	1.00	0.98	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$N_{IC} = 50$	$N_{NC}=100$	0.98	0.98	0.98	0.97	0.98	0.97	0.99	0.98	0.97	0.98	0.98	0.98
	$N_{NC}=200$	0.99	0.99	0.99	0.99	0.99	0.99	1.00	0.99	0.98	0.99	0.99	0.99
	$N_{NC}=300$	0.99	0.99	0.99	0.99	0.99	0.99	1.00	0.99	0.99	0.99	0.99	0.99
	$N_{NC}=400$	1.00	1.00	0.99	0.99	0.99	0.99	1.00	0.99	0.99	0.99	0.99	0.99
	$N_{NC}=500$	1.00	1.00	1.00	0.99	0.99	0.99	1.00	1.00	0.99	0.99	1.00	1.00
$N_{IC} = 100$	$N_{NC}=100$	0.96	0.96	0.95	0.95	0.95	0.94	0.99	0.98	0.97	0.96	0.96	0.95
	$N_{NC}=200$	0.98	0.98	0.97	0.97	0.98	0.97	1.00	0.99	0.98	0.98	0.98	0.97
	$N_{NC}=300$	0.99	0.99	0.98	0.98	0.98	0.98	1.00	0.99	0.99	0.98	0.98	0.98
	$N_{NC}=400$	0.99	0.99	0.99	0.99	0.99	0.98	1.00	0.99	0.99	0.99	0.99	0.98
	$N_{NC}=500$	0.99	0.99	0.99	0.99	0.99	0.98	1.00	0.99	0.99	0.99	0.99	0.99
<b>Av e.</b>	-	<b>0.99</b>	<b>0.99</b>	<b>0.99</b>	<b>0.98</b>	<b>0.99</b>	<b>0.99</b>	<b>1.00</b>	<b>0.99</b>	<b>0.99</b>	<b>0.99</b>	<b>0.99</b>	<b>0.99</b>

### Calculating DoC and DoD

DoC and DoD are calculated next. As discussed in the previous discussion, the different movements on the (TPR,FPR) domain can be associated with different weights (see equations (8-16) and (8-24)). In this section, we also examine the relative importance of moving from the DROP region to the other region that contains the points outside the DROP region. Four weighting cases are considered for equations (8-16) to (8-24). In these cases, we have increased the importance of moving between the DROP and the region outside the DROP. The cases considered are as follows:

$$\begin{aligned} \text{Case 1 : } W_R'' = W_S'' = W_U'' = W_V'' &= \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \text{Case 2: } W_R'' = W_S'' = W_U'' = W_V'' = \begin{bmatrix} 1 & 10 \\ 10 & 1 \end{bmatrix} \\ \text{Case3: } W_R'' = W_S'' = W_U'' = W_V'' &= \begin{bmatrix} 1 & 50 \\ 50 & 1 \end{bmatrix}, \text{Case4: } W_R'' = W_S'' = W_U'' = W_V'' = \begin{bmatrix} 1 & 100 \\ 100 & 1 \end{bmatrix} \end{aligned} \quad (8-35)$$

The DoC results for Case 1 are shown in Table 8-2 and Table 8-3. The columns show the DoC values. For each column, three sub-columns are presented similar to Table 8-1 *Res1* stands for the finest resolution; *Mean* stands for the average resolution over the 10 resolutions that were studied and *Res10* stands for the coarsest resolution. Rows show different  $N_{IC}$  and  $N_{NC}$  configurations (15 configurations in total).

Based on these tables, the following observations can be made (please note that we have presented our observations based on the pair-wise comparison of metrics):

1 ) *OA* vs. *HF*: As seen from Table 8-2 , both metrics are consistent with each other and  $DoC(HF, OA) > 0.75$  for all cases and for all resolutions studied. Especially when the number of IC samples is large enough to allow lower resolution analysis,  $DoC(HF, OA) > 0.85$ .

The analysis of the DoD results is more complex. When the number of IC samples is low ( $N_{IC}=10$ ), the results from different resolutions are highly inconsistent (see Table 8-4). This may be due to the fact that a low number of IC epochs yields highly variable results, especially when the resolution is low. On the other hand, when the number of IC epochs grows, HF consistently outperforms OA, with  $DOD(OA, HF) < 0.5$  for most resolutions studied here (see Table 8-4). Based on these arguments, for the cases studied here, ***HF is preferable to OA.***

**Table 8-2. The DoC results for the evaluation metrics studied in this paper (the first three comparisons). *Res1*, *Mean* and *Res10* stand for the finest resolution, the average of 10 resolution values and the coarsest resolution, respectively.**

		DoC(OA,HF)			DoC(OA, $\frac{TPR}{FPR}$ )			DoC(OA,Kappa)		
		Res1	Mean	Res10	Res1	Mean	Res10	Res1	Mean	Res10
<b><math>N_{IC}=10</math></b>	$N_{NC}=100$	0.84	0.79	0.76	0.79	0.69	0.68	0.89	0.88	0.86
	$N_{NC}=200$	0.82	0.78	0.75	0.83	0.72	0.67	0.87	0.87	0.87
	$N_{NC}=300$	0.81	0.78	0.75	0.85	0.75	0.67	0.87	0.87	0.87
	$N_{NC}=400$	0.80	0.79	0.75	0.87	0.78	0.67	0.87	0.88	0.87
	$N_{NC}=500$	0.79	0.80	0.80	0.87	0.81	0.78	0.87	0.88	0.89
<b>Mean</b>		<b>0.81</b>	<b>0.79</b>	<b>0.76</b>	<b>0.84</b>	<b>0.75</b>	<b>0.69</b>	<b>0.87</b>	<b>0.88</b>	<b>0.87</b>
<b><math>N_{IC}=50</math></b>	$N_{NC}=100$	0.90	0.90	0.89	0.80	0.64	0.61	0.95	0.95	0.94
	$N_{NC}=200$	0.88	0.88	0.87	0.81	0.67	0.60	0.93	0.92	0.92
	$N_{NC}=300$	0.87	0.86	0.86	0.81	0.71	0.60	0.92	0.91	0.91
	$N_{NC}=400$	0.86	0.85	0.86	0.82	0.75	0.60	0.91	0.90	0.91
	$N_{NC}=500$	0.86	0.83	0.82	0.82	0.79	0.77	0.91	0.89	0.87
<b>Mean</b>		<b>0.87</b>	<b>0.86</b>	<b>0.86</b>	<b>0.81</b>	<b>0.71</b>	<b>0.63</b>	<b>0.92</b>	<b>0.91</b>	<b>0.91</b>
<b><math>N_{IC}=100</math></b>	$N_{NC}=100$	0.94	0.94	0.93	0.81	0.64	0.59	1.00	1.00	1.00
	$N_{NC}=200$	0.91	0.91	0.90	0.82	0.68	0.60	0.96	0.95	0.95
	$N_{NC}=300$	0.89	0.89	0.89	0.81	0.71	0.59	0.94	0.94	0.94
	$N_{NC}=400$	0.88	0.88	0.88	0.81	0.75	0.59	0.93	0.92	0.93
	$N_{NC}=500$	0.88	0.86	0.86	0.81	0.78	0.77	0.93	0.91	0.90
<b>Mean</b>		<b>0.90</b>	<b>0.89</b>	<b>0.89</b>	<b>0.81</b>	<b>0.71</b>	<b>0.63</b>	<b>0.95</b>	<b>0.94</b>	<b>0.94</b>
<b>Overall Mean</b>		<b>0.86</b>	<b>0.85</b>	<b>0.84</b>	<b>0.82</b>	<b>0.73</b>	<b>0.65</b>	<b>0.92</b>	<b>0.91</b>	<b>0.91</b>

**Table 8-3. The DoC results for the evaluation metrics studied in this paper (the last three comparisons). *Res1*, *Mean* and *Res10* stand for the finest resolution, the average of 10 resolution values and the coarsest resolution, respectively.**

		DoC(HF, $\frac{TPR}{FPR}$ )			DoC(HF,Kappa)			DoC( $\frac{TPR}{FPR}$ ,Kappa)		
		Res1	Mean	Res10	Res1	Mean	Res10	Res1	Mean	Res10
<b><math>N_{IC}=10</math></b>	$N_{NC}=100$	0.83	0.91	1.00	0.94	0.91	0.88	0.86	0.82	0.83
	$N_{NC}=200$	0.86	0.93	1.00	0.93	0.91	0.88	0.90	0.85	0.83
	$N_{NC}=300$	0.87	0.94	1.00	0.93	0.91	0.88	0.92	0.88	0.83
	$N_{NC}=400$	0.87	0.94	1.00	0.92	0.91	0.88	0.93	0.89	0.83
	$N_{NC}=500$	0.87	0.95	1.00	0.92	0.91	0.90	0.94	0.91	0.89
<b>Mean(std)</b>		<b>0.86</b>	<b>0.93</b>	<b>1.00</b>	<b>0.93</b>	<b>0.91</b>	<b>0.88</b>	<b>0.91</b>	<b>0.87</b>	<b>0.84</b>
<b><math>N_{IC}=50</math></b>	$N_{NC}=100$	0.81	0.66	0.63	0.94	0.95	0.95	0.83	0.67	0.64
	$N_{NC}=200$	0.83	0.71	0.65	0.94	0.94	0.95	0.85	0.73	0.66
	$N_{NC}=300$	0.84	0.76	0.66	0.94	0.94	0.95	0.87	0.78	0.67
	$N_{NC}=400$	0.84	0.80	0.66	0.94	0.93	0.95	0.88	0.82	0.68
	$N_{NC}=500$	0.84	0.84	0.82	0.93	0.93	0.95	0.88	0.86	0.85
<b>Mean (std)</b>		<b>0.83</b>	<b>0.75</b>	<b>0.68</b>	<b>0.94</b>	<b>0.94</b>	<b>0.95</b>	<b>0.86</b>	<b>0.77</b>	<b>0.70</b>
<b><math>N_{IC}=100</math></b>	$N_{NC}=100$	0.79	0.63	0.58	0.94	0.94	0.93	0.81	0.63	0.58
	$N_{NC}=200$	0.82	0.69	0.61	0.95	0.95	0.95	0.84	0.70	0.62
	$N_{NC}=300$	0.83	0.73	0.61	0.94	0.94	0.95	0.85	0.75	0.63
	$N_{NC}=400$	0.83	0.78	0.61	0.94	0.94	0.95	0.86	0.80	0.64
	$N_{NC}=500$	0.83	0.82	0.81	0.94	0.93	0.93	0.86	0.84	0.82
<b>Mean</b>		<b>0.82</b>	<b>0.73</b>	<b>0.64</b>	<b>0.94</b>	<b>0.94</b>	<b>0.94</b>	<b>0.84</b>	<b>0.75</b>	<b>0.66</b>
<b>Overall Mean</b>		<b>0.84</b>	<b>0.81</b>	<b>0.78</b>	<b>0.94</b>	<b>0.93</b>	<b>0.92</b>	<b>0.87</b>	<b>0.80</b>	<b>0.73</b>

**Table 8-4. The DoD results for the evaluation metrics studied in this paper (the first three comparisons). *Res1*, *Mean* and *Res10* stand for the finest resolution, the average of 10 resolution values and the coarsest resolution, respectively.**

		DoD(OA,HF)			DoD(OA, $\frac{TPR}{FPR}$ )			DoD(OA,Kappa)		
		Res1	Mean	Res10	Res1	Mean	Res10	Res1	Mean	Res10
<b><i>N<sub>IC</sub></i>=10</b>	<i>N<sub>NC</sub></i> =100	0.09	4.22	5.00	1.14	8.52	5.00	0.00	0.06	0.10
	<i>N<sub>NC</sub></i> =200	0.09	7.98	10.00	0.55	11.35	10.00	0.00	0.03	0.05
	<i>N<sub>NC</sub></i> =300	0.12	11.01	15.00	0.36	11.23	15.00	0.00	0.02	0.03
	<i>N<sub>NC</sub></i> =400	0.15	14.68	20.00	0.38	14.91	20.00	0.00	0.01	0.03
	<i>N<sub>NC</sub></i> =500	0.20	18.38	25.00	0.30	18.59	25.00	0.00	0.01	0.02
<b>Mean(std)</b>		<b>0.13</b>	<b>11.25</b>	<b>15.00</b>	<b>0.55</b>	<b>12.92</b>	<b>15.00</b>	<b>0.00</b>	<b>0.02</b>	<b>0.05</b>
<b><i>N<sub>IC</sub></i>=50</b>	<i>N<sub>NC</sub></i> =100	0.02	0.24	0.41	3.29	19.18	13.27	0.01	0.08	0.14
	<i>N<sub>NC</sub></i> =200	0.02	0.37	0.72	1.22	19.75	22.09	0.01	0.03	0.07
	<i>N<sub>NC</sub></i> =300	0.02	0.43	1.05	0.63	15.73	31.23	0.00	0.02	0.06
	<i>N<sub>NC</sub></i> =400	0.02	0.35	1.38	0.43	9.64	40.55	0.00	0.01	0.03
	<i>N<sub>NC</sub></i> =500	0.02	0.10	0.20	0.29	1.22	1.55	0.00	0.00	0.00
<b>Mean (std)</b>		<b>0.02</b>	<b>0.30</b>	<b>0.75</b>	<b>1.17</b>	<b>13.10</b>	<b>21.74</b>	<b>0.00</b>	<b>0.03</b>	<b>0.06</b>
<b><i>N<sub>IC</sub></i>=100</b>	<i>N<sub>NC</sub></i> =100	0.03	0.10	0.20	5.37	24.66	18.07	1.00	1.00	1.00
	<i>N<sub>NC</sub></i> =200	0.02	0.12	0.25	2.04	23.36	26.70	0.01	0.04	0.08
	<i>N<sub>NC</sub></i> =300	0.02	0.13	0.33	1.06	18.73	36.41	0.01	0.02	0.06
	<i>N<sub>NC</sub></i> =400	0.02	0.11	0.41	0.70	11.76	46.90	0.01	0.01	0.04
	<i>N<sub>NC</sub></i> =500	0.01	0.05	0.12	0.47	2.05	2.57	0.01	0.00	0.01
<b>Mean</b>		<b>0.02</b>	<b>0.10</b>	<b>0.26</b>	<b>1.93</b>	<b>16.11</b>	<b>26.13</b>	<b>0.21</b>	<b>0.22</b>	<b>0.24</b>
<b>Overall Mean</b>		<b>0.06</b>	<b>3.88</b>	<b>5.34</b>	<b>1.22</b>	<b>14.05</b>	<b>20.96</b>	<b>0.07</b>	<b>0.09</b>	<b>0.11</b>

**Table 8-5. The DoD results for the evaluation metrics studied in this paper (the last three comparisons). *Res1*, *Mean* and *Res10* stand for the finest resolution, the average of 10 resolution values and the coarsest resolution, respectively.**

		DoD( $\frac{TPR}{FPR}$ ,HF)			DoD(HF,Kappa)			DoD( $\frac{TPR}{FPR}$ ,Kappa)		
		Res1	Mean	Res10	Res1	Mean	Res10	Res1	Mean	Res10
$N_{IC}=10$	$N_{NC}=100$	0.07	0.00	1.00	0.01	0.00	0.00	0.00	0.00	0.00
	$N_{NC}=200$	0.06	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	$N_{NC}=300$	0.12	0.02	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	$N_{NC}=400$	0.15	0.01	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	$N_{NC}=500$	0.33	0.03	1.00	0.00	0.00	0.00	0.00	0.00	0.00
<b>Mean(std)</b>		<b>0.13</b>	<b>0.15</b>	<b>0.01</b>	<b>1.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
$N_{IC}=50$	$N_{NC}=100$	0.01	0.00	0.00	0.14	0.21	0.21	0.01	0.00	0.00
	$N_{NC}=200$	0.02	0.00	0.00	0.07	0.06	0.06	0.01	0.00	0.00
	$N_{NC}=300$	0.03	0.00	0.00	0.06	0.03	0.04	0.01	0.00	0.00
	$N_{NC}=400$	0.03	0.00	0.00	0.03	0.01	0.01	0.01	0.00	0.00
	$N_{NC}=500$	0.04	0.00	0.00	0.02	0.01	0.01	0.01	0.00	0.00
<b>Mean (std)</b>		<b>0.02</b>	<b>0.03</b>	<b>0.00</b>	<b>0.00</b>	<b>0.06</b>	<b>0.07</b>	<b>0.07</b>	<b>0.01</b>	<b>0.00</b>
$N_{IC}=100$	$N_{NC}=100$	0.02	0.00	0.01	36.23	14.10	4.96	0.19	0.06	0.06
	$N_{NC}=200$	0.02	0.00	0.00	0.19	0.22	0.24	0.01	0.00	0.00
	$N_{NC}=300$	0.03	0.00	0.00	0.26	0.16	0.14	0.02	0.00	0.00
	$N_{NC}=400$	0.03	0.00	0.00	0.16	0.08	0.07	0.02	0.00	0.00
	$N_{NC}=500$	0.04	0.00	0.07	0.20	0.06	0.02	0.03	0.01	0.00
<b>Mean</b>		<b>0.02</b>	<b>0.03</b>	<b>0.00</b>	<b>0.02</b>	<b>7.41</b>	<b>2.92</b>	<b>1.09</b>	<b>0.05</b>	<b>0.02</b>
<b>Overall Mean</b>		<b>0.06</b>	<b>0.07</b>	<b>0.01</b>	<b>0.34</b>	<b>2.49</b>	<b>1.00</b>	<b>0.38</b>	<b>0.02</b>	<b>0.01</b>

**Table 8-6 . The effect of weights on average values of DoC.**

		<b>Case 1</b>	<b>Case2</b>	<b>Case 3</b>	<b>Case 4</b>
<b>DoC(OA,HF)</b>	Res1	0.86	0.96	0.98	0.99
	Mean	0.85	0.96	0.98	0.99
	Res10	0.84	0.96	0.98	0.99
<b>DoC(OA, <math>\frac{TPR}{FPR}</math>)</b>	Res1	0.81	0.94	0.98	0.98
	Mean	0.71	0.91	0.97	0.97
	Res10	0.65	0.89	0.96	0.97
<b>DoC(OA,Kappa)</b>	Res1	0.91	0.97	0.99	0.99
	Mean	0.91	0.98	0.99	0.99
	Res10	0.91	0.98	0.99	0.99
<b>DoC(HF, <math>\frac{TPR}{FPR}</math>)</b>	Res1	0.83	0.94	0.97	0.98
	Mean	0.80	0.93	0.97	0.97
	Res10	0.78	0.92	0.96	0.97
<b>DoC(HF,Kappa)</b>	Res1	0.94	0.98	0.99	0.99
	Mean	0.93	0.98	0.99	0.99
	Res10	0.92	0.98	0.99	0.99
<b>DoC(<math>\frac{TPR}{FPR}</math>,Kappa)</b>	Res1	0.86	0.95	0.98	0.98
	Mean	0.79	0.93	0.97	0.97
	Res10	0.73	0.91	0.96	0.97



**Table 8-7. The effect of weights on average values of DoD.**

		Case 1	Case2	Case 3	Case 4
<b>DoD(OA, HF)</b>	Res1	0.10	0.14	0.33	0.56
	Mean	4.31	4.34	4.47	4.62
	Res10	5.34	5.43	5.78	6.15
<b>DoD(OA, <math>\frac{TPR}{FPR}</math>)</b>	Res1	1.83	1.61	1.23	1.05
	Mean	15.47	13.73	10.49	8.89
	Res10	20.96	18.54	13.89	11.52
<b>DoD(OA, Kappa)</b>	Res1	0.07	0.09	0.18	0.29
	Mean	0.09	0.12	0.27	0.45
	Res10	0.11	0.17	0.43	0.76
<b>DoD(<math>\frac{TPR}{FPR}</math>, HF)</b>	Res1	0.01	0.05	0.23	0.41
	Mean	0.00	0.01	0.04	0.05
	Res10	0.00	0.01	0.05	0.10
<b>DoD(HF, Kappa)</b>	Res1	2.03	0.43	0.25	0.24
	Mean	0.83	0.33	0.27	0.32
	Res10	0.38	0.19	0.16	0.19
<b>DoD(<math>\frac{TPR}{FPR}</math>, Kappa)</b>	Res1	0.01	0.05	0.18	0.33
	Mean	0.00	0.02	0.06	0.11
	Res10	0.00	0.01	0.05	0.08

2)  $OA$  vs.  $\frac{TPR}{FPR}$ : Based on the results in Table 8-2, both metrics are consistent with each other ( $DoC(OA, \frac{TPR}{FPR}) > 0.6$  for all the cases studied). However, the consistency between these metrics is lower than what was achieved between OA and HF.

In other words, OA is more consistent with HF than it is with  $\frac{TPR}{FPR}$ . However, note that for finer resolutions,  $\text{DoC}(\text{OA}, \frac{TPR}{FPR})$  increases above the 0.8 level. This important observation shows that, based on coarse resolutions, one cannot make a general statement as to “how much” two metrics are consistent with each other. Only a general statement about whether or not the two metrics are consistent with each other can be made. The  $\text{DoD}(\text{OA}, \frac{TPR}{FPR})$  from Table 8-4 shows that the amount DoD of one metric over another depends on the values of  $N_{IC}$  and  $N_{NC}$ . Although most of the results (especially from the coarser resolutions) indicate that OA is a better option, a general statement cannot be made. We thus conclude that depending on the values of  $N_{IC}$  and  $N_{NC}$ , the superior metric should be determined.

3) *OA vs. Kappa*: OA and Kappa have a high DoC for all the cases studied here (see Table 8-2). This is not surprising. According to (8-31), A, the definition of Kappa depends on the definition of OA. Table 8-4 shows that with the exception of the balanced case (where  $\frac{N_{NC}}{N_{IC}} = 1$ ), ***Kappa is more discriminant than OA and is thus more suitable.***

4) *HF vs.  $\frac{TPR}{FPR}$* :  $\text{DoC}(\text{HF}, \frac{TPR}{FPR})$  is above the 0.6 level for all the cases studied here, so both metrics are consistent with each other. As Table 8-5 shows, ***HF is more discriminant than  $\frac{TPR}{FPR}$  and is thus considered as more suitable.***

5) *HF vs. Kappa*: HF and Kappa are consistent with each other with  $\text{DoC}(\text{HF}, \text{Kappa}) > 0.85$  for most cases studied (see Table 8-3). As shown in Table 8-5, with the exception of the balanced case (where  $\frac{N_{NC}}{N_{IC}} = 1$ ), in all other cases ***Kappa is found to be more discriminant and thus more suitable than HF.***

6) *Kappa* vs.  $\frac{TPR}{FPR}$ : These metrics are consistent with each other with DoC>0.58 as shown in Table 8-5. For all cases studied here, Table 8-5 demonstrates that  $\text{DoD}(\frac{TPR}{FPR}, \text{Kappa}) \gg 1$ , and thus ***Kappa is found superior than  $\frac{TPR}{FPR}$*** .

Overall, the performances of Kappa and HF was superior and those of OA and  $\frac{TPR}{FPR}$  were inferior in directing the model selection algorithm on the (TPR,FPR) domain. This argument does not hold for all  $N_{NC}$  and  $N_{IC}$  values, although the results indicate that for a particular  $\frac{N_{NC}}{N_{IC}}$ , the most suitable evaluation metric can be selected from a pool of available evaluation metrics.

As seen from Table 8-6, when the weights of moving between the DROP region and the region outside the DROP increases (from Case 1 to Case 4), DoC significantly increases for all cases. This observation indicates that these metrics are more consistent with each other when moving from the DROP region to the other region and vice versa. Similarly the value of DoD changes as the weights of moving between the two regions increase (from Case 1 to Case 4 in Table 8-7). Again, this indicates that when it comes to comparisons between two regions, DoD can significantly vary. This is especially the case for  $\frac{TPR}{FPR}$ . While this metric did not perform very well in general, it performed much better when moving from one region to another. This is demonstrated by observing the increase in the DoD value of  $\frac{TPR}{FPR}$  compared to the other three metrics in Table 8-7.

In all the cases studied, the DoC values were consistently above the 0.5 threshold, no matter which resolution was used. However, the DoD results were less consistent. We have plotted the percentage of consistency between different resolutions in Figure 8-6. For ALL the cases studied here, this figure shows how many times *Resolution i* and *Resolution j* reached the same conclusion about DoD (only the results of Case 1 are presented here). As seen from this figure, there is a high correlation here. This indicates that instead of using high resolutions, lower resolutions can also be used. This

results in significant savings in simulation time. As an example, going from *Resolution 1* to *Resolution 2* results in approximately  $4 \times 3 = 12$  times less computation and going from *Resolution 1* to *Resolution 3* results in reducing the amount of computation by a factor of  $9 \times 8 = 72$ .

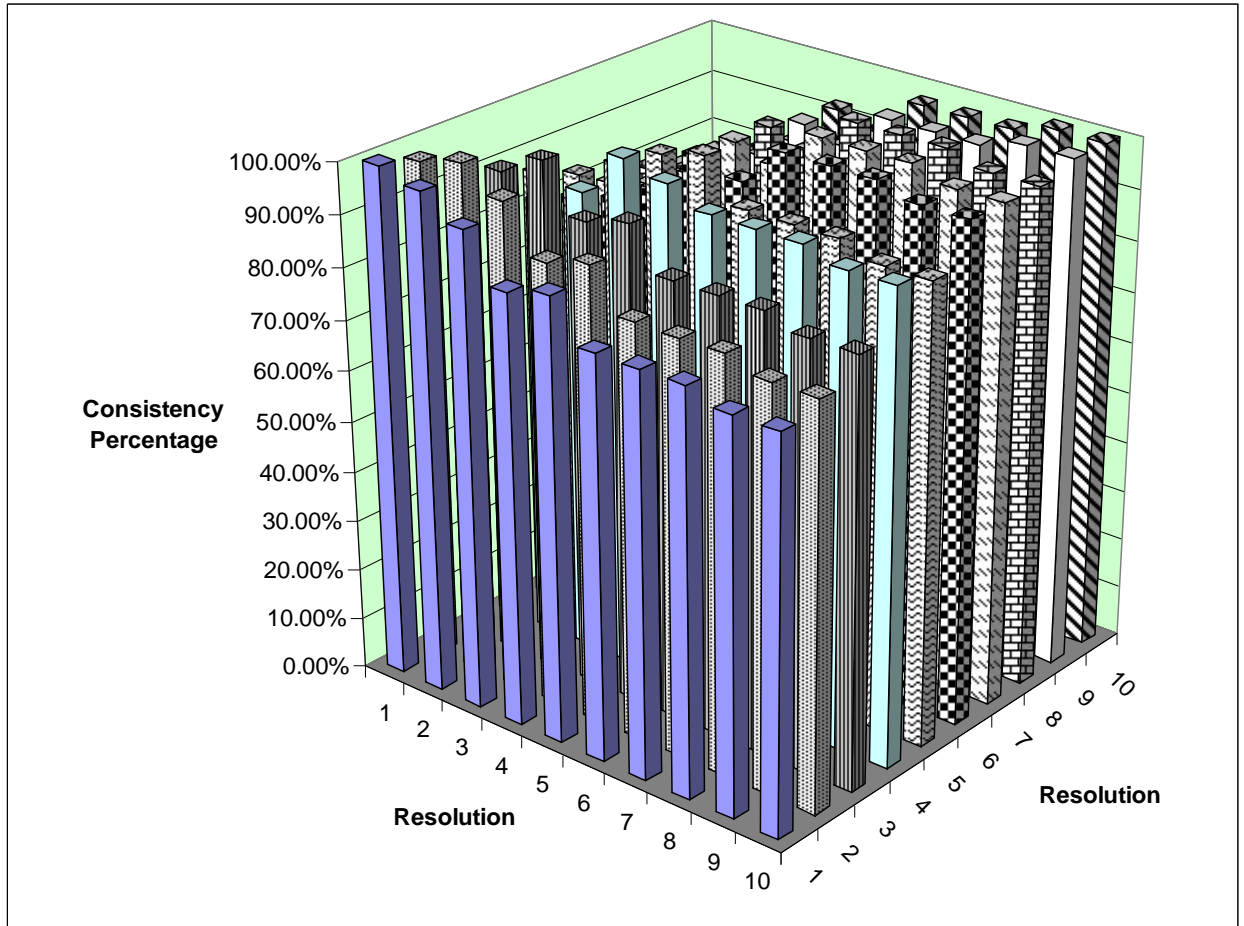


Figure 8-6. Consistency between different resolutions in reaching the same conclusion for the cases studied here. The chart attributes to DoD results.

## 8.6 Discussion and conclusions

Model tuning is a crucial step in the design of most classification algorithms. The results, however, depend on the evaluation metric used. As argued in the Introduction section, finding a suitable evaluation metric is not always easy, especially for classification algorithms with highly imbalanced classes such as in self-paced brain computer interface (SBCI) systems. A poorly chosen evaluation metric may mislead the model tuning procedure. Determining a suitable evaluation metric is thus of great

importance, but until recently it has been very difficult to compare the different performances of different evaluation metrics.

This paper addresses the problem of finding the most suitable metric from a given set of metrics. We first propose a framework for comparing evaluation metrics that are or could be used for classification problems with a large number of test samples. To decrease the computational complexity, a sub-sampling grid-based approach is then proposed. As a case study, we apply the proposed framework to compare four evaluation metrics defined for SBCI systems.

Based on the results presented in Section 8.5.2, it can be concluded that in finding a suitable evaluation metric, the number of available positive and negative samples (and their ratios) may impact the outcome of whether a particular evaluation metric is preferred over another. Depending on the requirement of the problem statement, the number of available test samples of each class and their ratios, must be taken into consideration when choosing a metric that can guide the solution to the DROP region. However, in most cases studied here, Kappa and HF outperformed OA and  $\frac{TPR}{FPR}$ .

Weights can be used to increase the relative importance of moving from the DROP to another region containing the points outside the DROP. They form a strong tool for analyzing how well an evaluation metric behaves when moving between different regions in the (TPR,FPR) domain. The results in Table 8-6 and Table 8-7 show that as the weights assigned to moves between the two regions increase, DoC, the degree of consistency, between the metrics significantly increases as well. This observation shows that when comparing evaluation metrics over different regions, the performance of all metrics was highly consistent with each other.

From the computational point of view, the proposed framework is more efficient than the original framework in [4], as explained in Section 8.3. This is an important advantage, especially when exploring classification problems with highly imbalanced classes, i.e., where the number of test samples in one class is significantly larger than those in the other class.

An important part of using the proposed framework is that of determining the DROP region. Very few SBCI studies have determined their classification goal, or the DROP region [24, 25]. We suggest the following approaches for defining DROP: 1) Using previous knowledge: if some knowledge exists that a particular TPR and FPR are suitable for the operation a particular SBCI system, DROP can be defined using this knowledge. 2) Simulated experiments can be carried out with test pilots to find the DROP region that the users find acceptable. 3) The results published in previous studies can be considered as a starting point for future research. Those results can be used to determine as the DROP region. Since the knowledge referred to in approach (1) is usually not readily available and approach (2) needs conducting an extensive research, we propose the use of approach (3) for determining the DROP region.

The main focus of this paper has been on choosing metrics for model tuning. The same principles can be applied when comparing the performance of different classification problems. Suppose algorithm A achieves a performance equal to  $x$  using evaluation metric  $f$  and algorithm B achieves a performance equal to  $y$  using evaluation metric  $g$ . By using DoC we can examine whether both metrics are consistent. If they are, DoD can be used to determine which evaluation metric is better than the other.

The proposed framework does not only apply to SBCI systems, it can also be applied to other classification problems with discrete states and a large number of test samples. Imbalanced dataset frequently arise in many other real-world situations including fraud detection [3], event classification in high energy physics [26], rare diagnosis such as thyroid diseases [27], and detection of oil spills in satellite radar images [28]. The proposed framework is expected to aid the researchers to choose a more suitable evaluation metric for the validation procedure. It should also be stressed that the proposed framework can be used for comparing evaluation metrics in any classification problem with a large number of test samples, i.e., independent of the class ratios.

## 8.7 References

- [1] S. G. Mason, A. Bashashati, M. Fatourehchi, K. F. Navarro and G. E. Birch, "A Comprehensive Survey of Brain Interface Technology Designs", *Annals of Biomedical Engineering*, vol. 35, no. 2, pp. 137-69, Feb 2007.
- [2] A. Schlögl, J. Kronegg, J. Huggins and S. G. Mason, "Evaluation criteria in BCI research," in *Towards Brain-Computer Interfacing* (G. Dornhege, J. R. Millan, T. Hinterberger, D. McFarland and K. R. Muller, Eds.), MIT Press, 2007.
- [3] F. Provost and T. Fawcett, "Robust Classification for Imprecise Environments", *Mach. Learning*, vol. 42, no.3, pp. 203-231, 2001.
- [4] J. Huang and C. X. Ling, "Using AUC and accuracy in evaluating learning algorithms", *IEEE Trans. Knowled. Data Eng.*, vol. 17, no.3, pp. 299-310, 2005.
- [5] J. Zhu and T. Yao, "An evaluation of statistical spam filtering techniques", *ACM Transactions on Asian Language Information Processing (TALIP)*, vol. 3, no.4, pp. 243-269, 2004.
- [6] A. P. Bradley, "Use of the area under the ROC curve in the evaluation of machine learning algorithms", *Pattern Recognit*, vol. 30, no.7, pp. 1145-1159, 1997.
- [7] N. T. Choplin and D. C. Lundy, "The sensitivity and specificity of scanning laser polarimetry in the detection of glaucoma in a clinical setting", *Ophthalmology*, vol. 108, no.5, pp. 899-904, May.2001.
- [8] G. E. Birch, Z. Bozorgzadeh and S. G. Mason, "Initial on-line evaluations of the LF-ASD brain-computer interface with able-bodied and spinal-cord subjects using imagined voluntary motor potentials", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 10, no.4, pp. 219-224, Dec.2002.
- [9] B. Graimann, J. E. Huggins, S. P. Levine and G. Pfurtscheller, "Toward a direct brain interface based on human subdural recordings and wavelet-packet analysis", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 954-962, Jun.2004.
- [10] J. Kronegg, s. Voloshynovskiy and P. Pun, "Analysis of bit rate definitions for brain-computer interfaces," in the *Proc. Int. Conf. on Human-Computer Interaction (HCI'05)*, Las Vegas, Nevada, 2005.
- [11] J. F. Borisoff, S. G. Mason, A. Bashashati and G. E. Birch, "Brain-computer interface design for asynchronous control applications: improvements to the LF-ASD asynchronous brain switch", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 985-992, Jun.2004.
- [12] M. Fatourehchi, G. E. Birch and R. K. Ward, "A self-paced brain interface system that uses movement related potentials and changes in the power of brain rhythms", *J. Comput. Neurosci.*, vol.23, no.1, pp.21-37, Aug 2007.
- [13] S. G. Mason, J. Kronegg, J. Huggins, M. Fatourehchi and A. and Schloegl, "Evaluating the performance of self-paced BCI technology", Technical Report, available online: [http://www.bci-info.tugraz.at/Research\\_Info/documents/articles/self\\_paced\\_tech\\_report-2006-05-19.pdf](http://www.bci-info.tugraz.at/Research_Info/documents/articles/self_paced_tech_report-2006-05-19.pdf), 2006.
- [14] C. X. Ling, J. Huang and H. Zhang, "AUC: a better measure than accuracy in comparing learning algorithms", in *Proc.of IJCAI*, vol. 3, 2003.

- [15] N. Yamawaki, C. Wilke, Z. Liu and B. He, "An enhanced time-frequency-spatial approach for motor imagery classification", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no.2, pp. 250-254, Jun.2006.
- [16] A. Buttfeld, P. W. Ferrez and R. Millan Jdel, "Towards a robust BCI: error potentials and online learning", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no.2, pp. 164-168, Jun.2006.
- [17] G. R. Muller-Putz, R. Scherer, C. Neuper and G. Pfurtscheller, "Steady-state somatosensory evoked potentials: suitable brain signals for brain-computer interfaces?", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no.1, pp. 30-37, Mar.2006.
- [18] J. R. Wolpaw, D. McFarland and G. Pfurtscheller, "EEG-based Communication: Improved Accuracy by Reponse Verification", *IEEE Trans. Rehab. Eng.*, vol. 6, no.3, pp. 326-333, 1998.
- [19] M. Fatourech, S. G. Mason, G. E. Birch and R. K. Ward, "Is information transfer rate a suitable performance measure for self-paced brain interface systems?" in the Proc. IEEE ISSPIT'04, pp. 212-216, Aug. 2006.
- [20] T. Nykopp, "Statistical modeling issues for the adaptive brain interface", M.Sc. Thesis, Helsinki University of Technology, 2001.
- [21] J. Cohen, "A coefficient of agreement for nominal scales", *Educational and Psychological Measurement*, vol. 20, no.1, pp. 37-46, 1960.
- [22] J. E. Huggins, S. P. Levine, S. L. Bement, R. K. Kushwaha, L. A. Schuh, E. A. Passaro, M. M. Rohde, D. A. Ross, K. V. Elisevich and B. J. Smith, "Detection of Event-Related Potentials for Development of a Direct Brain Interface", *J Clinical Neurophysiol*, vol. 16, no.5, pp. 448-455, Sep.1999.
- [23] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing", *Journal of the Royal Statistical Society.Series B (Methodological)*, vol. 57, no.1, pp. 289-300, 1995.
- [24] M. Fatourech, G. E. Birch and R. K. Ward, "Applying a hybrid genetic algorithm in the design of a self-paced brain interface with a low false positive rate," in the *Proc. IEEE ICASSP'07*, vol.4, pp. IV-1157; IV-1160, 2007.
- [25] E. Yom-Tov and G. F. Inbar, "Detection of Movement-Related Potentials from the Electro-Encephalogram for possible use in a Brain-Computer Interface", *Medical and Biological Engineering and Computing*, vol. 41, no.1, pp. 85-93, Jan. 2003.
- [26] S. Clearwater and E. Stern, "A rule-learning program in high energy physics event classification", *Comput. Phys. Commun.*, vol. 67, no.2, pp. 159-182, 1991.
- [27] M. Kubat, R. Holte and S. Matwin, "Learning when negative examples abound", *Machine Learning: ECML-97, Lecture Notes in Artificial Intelligence*, vol. 1224, 1997.
- [28] M. Kubat, R. C. Holte and S. Matwin, "Machine Learning for the Detection of Oil Spills in Satellite Radar Images", *Mach. Learning*, vol. 30, no.2, pp. 195-215, 1998.



## **CHAPTER 9 NEW STUDIES ON THE DESIGN OF A 2-STATE SELF-PACED BRAIN COMPUTER INTERFACE SYSTEM WITH A LOW FALSE ACTIVATION RATE<sup>9</sup>**

### **9.1 Introduction**

Brain-computer interface (BCI) systems provide an alternative communication channel for individuals with severe motor disabilities. A BCI system bypasses the traditional sensorimotor pathways and uses the brain signals to control objects (see [1, 2] for reviews of BCI systems).

Most of the current BCI systems are designed such that they can only be activated during certain periods specified by the system. These types of BCI systems are called synchronized BCI systems. Their main characteristic is that they require constant user attention during the control periods, i.e., the periods when they are activated. As a result, these systems can not usually handle the “no control (NC)” state of the user, i.e., when the user does not intend to control (i.e., activate the system). In contrast, self-paced BCI (SBCI) systems are BCI systems that allow the user to perform a control action whenever he or she desires. These systems should thus be able to distinguish an intentional control (IC) command from the NC states. Please see [1, 3-5] for more information on synchronized and self-paced BCI systems.

While SBCI systems allow the users to have more freedom as to when to control the system, from the pattern recognition point of view, they create a much more challenging design problem compared to synchronized BCI systems. On one hand, an SBCI system should be able to recognize IC commands and thus should have a

---

<sup>9</sup> A version of this chapter will be submitted for Publication. Fatourehchi, M., Ward, R.K., and Birch, G. E., “New studies on the design of a 2-state self-paced brain computer interface system with a low false positive rate”, submitted.

reasonably high true positive rate, TPR. On the other hand, the system should not be activated during NC periods and should generate very low false positive (FP) activations. A review of current SBCI systems demonstrates that most of these designs achieve modest TP rates, but their FP rates (FPR) are unfortunately too high for many practical applications [5-13]. The main reason is that SBCI systems generate an output at a relatively high rate, which is usually higher than 10Hz as in [6, 8, 14]. This translates into a false positive activation every few seconds, even when the FP rate is low (e.g.,  $FPR \approx 2\%$  as in [11, 15]). As an example, the most recent design of an SBCI system called the low frequency- asynchronous switch design (LF-ASD) achieves an average TPR of 41% at the false positive rate of 1% [16]. Since LF-ASD generates an output every  $1/16^{\text{th}}$  of a second, this FP rate translates into one false positive every 6.25 seconds. For many applications, such a high FPR can result in user frustration, especially if the TP rate is not very high.

In [17], we proposed a new SBCI system, which achieved a low FP rate (average FP rate was 0.1%), at a modest TP rate (average TP rate was 56.2%). Since the system generated an output every  $1/8^{\text{th}}$  of a second, this error rate results in a false activation roughly every “two minutes”. This is much lower than those reported in previous EEG-based SBCI systems. Because of the good performance of this system, in this Chapter, we examine its performance when used under more general situation. First, we investigate if the system can be generalized to detect patterns related to another IC command, besides that of a right index finger flexion movement. The ability of an SBCI system to detect more IC commands provides the users with more control options. For example, during a “menu selection” application, one IC command can be used to control the movement of the selection box, while the other can be used to actually make the selection. Second, the performance of the system is investigated during periods for which the user is engaged with other mental activities, i.e., the system is investigated with NC data that are more engaging than those of previous studies. NC data in SBCI studies have usually been recorded over periods during which users are in an “idle” state. Examples of these data are those recorded between two successive movement attempts by a user or during some periods over which the user is not performing any particular task [11, 15]. In real-life applications, however, the user of an SBCI system may be engaged with any mental NC

activity when not controlling a device. Many such activities are of an engaging nature, i.e., the activity may require attention and mental activities other than that attributed to controlling the SBCI system. These activities may unfortunately generate patterns that can generate false positives in the system. As an example, studies have shown that “attention” can block the Mu rhythm. This can result in the generation of an event-related desynchronization in the Mu rhythm, similar to the one generated as the result of a movement intended to control a device [5]. Based on these discussions, it is necessary that the performance of the system is also tested with “engaging NC data”, e.g., data collected when the user is engaged with mental activities that require a higher level of thinking. Because of the more engaging nature of such NC data, we expect that the performance of the system would deteriorate. However, in this Chapter, we show that the system’s good performance is maintained even when NC data of engaging nature are used.

In this Chapter, we study the performance of the system using data related to the right hand extension movement. The data were recorded in a self-paced fashion but under a more engaging environment than those used in previous studies. The data were recorded as participants were playing a Tetris-like game. The NC data were collected as the users were mentally engaged with the game, as explained in more details in the next section. To improve the performance of the system, we carried out the following:

**1) Implementing a better artifact monitoring system:** Artifacts are unwanted potentials that can degrade the system’s performance. For this study, a new artifact monitoring system that monitors the eye movement activities as well as the movements of the frontalis muscles was built. The aim is to ensure that the system is not activated by eye blinks or the movements of the frontalis muscles. In our previous studies, only eye movement activities were monitored while recording the data. Because MRPs and eye blink activities share the same frequency bands, this monitoring scheme may have been good enough for some of our earlier SBCI designs which only relied on the detection of the MRP neurological phenomenon [5, 6, 11, 15, 18]. A better artifact monitoring scheme is however needed for a system that uses neurological phenomena whose frequency bands are different from that of MRPs, as is the case with the system under investigation here. Such monitoring is necessary to ensure that the source of the control initiates from

the brain and not from the peripheral muscles such as the eye or facial muscles. Evidence from the literature shows that muscle activities from the arm and leg muscles are not concurrent with EEG-based communication [19], so we these types of artifacts were not studied here. However, the effects of facial muscles need to be investigated more thoroughly, as they are closer to the locations where EEG electrodes are placed.

**2) More suitable evaluation metric for model selection:** Selecting a suitable evaluation metric for SBCI systems during the model selection procedure is still an issue under debate. Various metrics such as the overall classification accuracy[6], HF-difference[20], the mutual information (information transfer rate) [21], Kappa [22], the area under the receiver operating characteristics (ROC) curve [22] , the TP rate at a fixed FP rate [11] have been proposed. SBCI researchers, however, have yet to agree as to which evaluation metric is more suitable for evaluating a particular SBCI application. In Chapter 8, we proposed a new framework for selecting an evaluation metric for an SBCI system during the model selection procedure. Given a number of evaluation metrics, this framework can help the designer choose the most suitable one for the specific application considered. In this paper, we apply this framework to choose a suitable metric for summarizing the performance during the model selection procedure.

**3) More suitable referencing method:** The choice of a suitable referencing method can affect the quality of EEG signals used for control. In the BCI literature, different montages have been proposed to effectively extract the “localized” EEG activity and thus provide suitable means of control. The proposed referencing methods include monopolar, bipolar, common average reference (CAR), small and large Laplacian [23-25]. Thus, a number of studies that compare different referencing methods have been discussed in the literature [23, 24].

In the previous studies of our group, two different referencing methods have been used: monopolar (referenced to common ear electrodes) and bipolar. Given the relatively low number of EEG channels used in these studies, these two montages seemed to be more suitable than others. Some evidence from our previous research suggests that MRP features extracted from bipolar EEG signals are more suitable than monopolar EEG signals for the purpose of SBCI control [5, 18]. However, a specific study has not been

conducted yet to compare these two referencing methods. In this paper, we compare the performance of the SBCI system based on monopolar and bipolar montages during the classification of IC and NC epochs. The aim is to determine if the bipolar montage is indeed superior to the monopolar montage as suggested by earlier indications.

In the next section, the experimental paradigm used for data recording is described in detail.

## **9.2 Experimental paradigm**

### **9.2.1 Data recording**

Five right-handed individuals (four males and one female) between 20 and 30 years old (mean 26 years) participated in this study. These individuals were denoted by AB1 to AB5 (AB stands for able-bodied). The study consisted of two sessions, each session was of two hours duration. All the participants signed consent forms prior to the beginning of the experiment. This study has been approved by the Behavioral Research Ethics Board (BREB) of the University of British Columbia.

EEG data were collected from the participants while they played a version of a game called PowerWord [26]. This game has a similar structure as Tetris, a falling puzzle video game. In Tetris, shapes composed of 4-square blocks fall down on the playing field. The object of the game is that users should move or rotate these blocks to make a horizontal line of blocks without gaps. If such line is created, it disappears and any blocks that rest above this line (if any) would fall down. PowerWord has the same structure as of Tetris, however, instead of cubes, letters fall down. The aim is to make meaningful words by moving the letters to the right position. At the start of each run, the first four rows of the playing field were randomly filled by letters. The playing field consists of six columns and 12 rows. This means that at the time of initialization, at most 24 letters would have been randomly placed on the playing field. A letter then starts falling down from the upper-left corner of the playing field. It takes 4.5 seconds as a letter moves from one row to another. The user's task is to move the letters to the right position such that after falling down, the letter constitutes a meaningful word with the letters that have already been in the playing field. In order for a combination of letters to be recognized as

a word, it should have 3, 4, 5 or 6 letters and should be meaningful. A local database is used for searching and identifying the meaningful words. After a word is recognized by the system, it appears on the right hand column and a score in the form of *(number of letters in the created word-2) × 10* is given to the user. The letters constituting that word would then disappear from the playing field. The user's task is to create as many words as possible in a 5-minute period. No timer was shown on the playing field so as to eliminate any possible source of anxiety. The participants were assured repeatedly that their performance is not evaluated against other users. Instead, they were asked to solely focus on the game and perform the required movements.

Once a letter appeared on the screen, the participants had to first think and decide the location the letter should be moved to. When the user decided on this location, then he/she was asked to solely focus on moving the letter to the desired position. This was done by activating a custom-made hand switch. The switch was made such that in the rest position, it was deactivated, but as soon as the user extended his/her right hand, the switch became activated. The participants were asked to avoid pressing the switch successively. Instead, after each movement, they were asked to wait until the letter goes down for at least one row before they made their subsequent movement attempt (if needed). The interval between two successive movement attempts was chosen to be 4.5 seconds, to ensure that the neurological phenomena related to any two successive movements do not overlap. The individuals were asked to avoid blinking or moving other parts of their body, when they activated the switch. After moving a specific letter to its desired location, the user was free to perform other mental tasks including thinking about where to place the next letter (the next letter is displayed in a column on the right hand side of the playing field).

We collected NC data in two ways. First, we analyzed the periods between two successive movement (IC) attempts and selected the NC data as the periods that were not contaminated with artifacts. These data mostly involved the periods when the user was thinking about creating a word with the next letter. We also recorded up to two separate NC recordings during each session. Each NC recording lasted about three minutes. During an NC recording, the user was asked to watch a movie of the game as played by

another user. The user was asked to follow the game and think about the locations he/she would have moved the letters to, had he/she was playing the game.

EEG data were collected from 13 monopolar electrodes positioned over the primary sensory motor cortex and supplementary motor area, references to linked ear electrodes. The electrodes were placed according to the International 10-20 system at  $F_1$ ,  $F_z$ ,  $F_2$ ,  $FC_3$ ,  $FC_1$ ,  $FC_z$ ,  $FC_2$ ,  $FC_4$ ,  $C_3$ ,  $C_1$ ,  $C_z$ ,  $C_2$  and  $C_4$  locations. EOG signals were recorded by measuring the activity of both eyes. Two electrodes were placed at the corner and below each eye. The EOG activity of each eye was measured as the difference between these two electrodes. EMG activity was measured as the difference between two electrodes placed on the forehead above the right eye. The first electrode was placed about 1 cm above the eyebrows and the second one was placed 3 cm above the first electrode (as recommended in [27] ). All signals were sampled at 256 Hz.

After the recorded signals were stored on the hard disk of a personal computer, they were converted to bipolar EEG signals. The conversion was carried out by calculating the difference between adjacent EEG channels and resulted in the following 18 bipolar EEG signals:  $F_1-FC_1$ ,  $F_1-F_z$ ,  $F_2-F_z$ ,  $F_2-FC_2$ ,  $FC_3-FC_1$ ,  $FC_3-C_3$ ,  $FC_1-FC_z$ ,  $FC_1-C_1$ ,  $FC_z-FC_2$ ,  $C_1-C_z$ ,  $C_2-C_4$ ,  $FC_2-FC_4$ ,  $FC_4-C_4$ ,  $FC_2-C_2$ ,  $FC_z-C_z$ ,  $C_3-C_1$ ,  $C_z-C_2$  and  $F_z-FC_z$ .

### 9.2.2 Artifact monitoring

EOG signals were first low-pass filtered to the [0, 5] frequency band using a 32-point linear-phase FIR filter. The eye blinks were detected when the difference between the electrodes around an eye exceeded a certain threshold. This threshold was determined separately for each participant and it was determined after careful examination of the data segments which contained eye blinks.

EMG artifacts from the frontalis muscles were detected using a procedure similar to the one suggested in [28]. The EMG signal was first filtered so it lies between [5,100] Hz. The root mean square (RMS) of the signal was calculated after applying an averaging sliding window of 25 ms width. In order to detect an EMG activation, the level of noise should first be estimated. This was done by treating the baseline EMG activity as

stochastic noise. The average RMS value of the EMG in the first 5 seconds of each recording was then calculated and the value of the noise level was estimated as three standard deviations from the mean value. If in an epoch, the amplitude of the EMG signal exceeded this value for a minimum of 20 ms, that epoch was marked as contaminated with EMG artifacts. We tested the performance of this detector as follows. The participants were asked to slightly raise their eyebrows at certain intervals specified by the monitoring system. The amount of raising the eyebrow was demonstrated to each participant by the experimenter before the experiment. Overall, raising eyebrows created distinct patterns in the EMG as well as EEG signals. The proposed detector was then applied to the data and it was able to successfully detect the periods when the users raised their frontalis muscles.

The outputs of EMG and EOG artifact detectors were then combined using an “AND” operator. Any artifact-contaminated period was removed from the analysis.

### **9.3 System design methods**

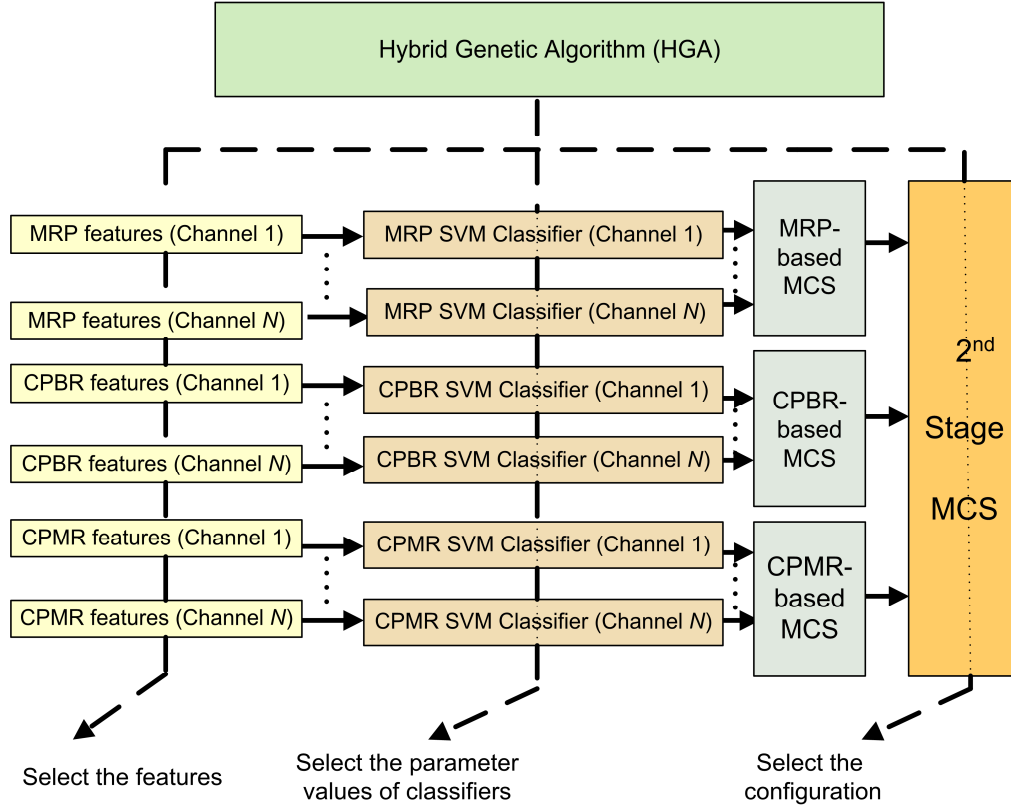
#### **9.3.1 Generating the IC and NC data**

An epoch was considered to be an intentional control (IC) if the data collected over the epoch’s time interval contained the onset of a movement, i.e., a hand switch was activated and no artifact was detected during that epoch. The interval started at  $t_{start} = -1$  second, i.e., 1 second before the onset of movement, and ended at  $t_{finish} = 1$ , i.e., 1 second after the onset of movement.

NC data were selected from the NC recordings as well as the EEG data between IC attempts. NC epochs were formed of 1-second segments as follows: a window of width  $(t_{finish} - t_{start})$  seconds was slid over each EEG signal by a step of 8 time samples (i.e., every 0.0625 sec), resulting in 16 classification decisions per second. If no artifacts were detected within that 1-second window, then that window was declared to be an NC epoch.

The structure of the SBCI is shown in Figure 9-1 (please refer to [17] for more information about this system).





**Figure 9-1.** The overall structure of the SBCI (from [17]). The dashed lines show the parts of the system whose values are determined by the hybrid genetic algorithm (HGA).

### 9.3.2 Feature extraction

We first describe the feature extraction method for MRPs, followed by the method for feature extraction for Changes in the Power of the Mu Rhythms (CPMR) and Changes in the Power of the Beta Rhythms (CPBR). For each neurological phenomenon in each of  $N$  EEG signal, features are extracted using the stationary wavelet transform (SWT). A 6-level SWT decomposition resulted in the generation of wavelet coefficients in the following frequency bands: [64-128], [32-64], [16-32], [8-16], [4-8], [2-4], and [0-2] Hz.

Suppose  $c_{j,k,p}$  and  $d_{j,k,p}$  are the approximation and detail coefficients at scale  $j$  and translation  $k$  of the  $p^{\text{th}}$  epoch in the training set of the IC commands. The averages of the approximation and detail coefficients at scale  $j$  and translation  $k$  ( $\bar{c}_{j,k}$ ,  $\bar{d}_{j,k}$ ) are:

$$\bar{c}_{j,k} = \frac{1}{N_{IC}} \cdot \sum_{p=1}^{N_{IC}} c_{j,k,p} \quad (9-1)$$

$$\bar{d}_{j,k} = \frac{1}{N_{IC}} \cdot \sum_{p=1}^{N_{IC}} d_{j,k,p} \quad (9-2)$$

The approximation template at scale  $j$  ( $TemplateC_j$ ) and the detail template at scale  $j$  ( $TemplateD_j$ ) are then obtained using the following formulae:

$$TemplateC_j = \{\bar{c}_{j,k}\} \quad (j=1,2,\dots,N_{Level}, k=1,2,\dots,N_{Samples}) \quad (9-3)$$

$$TemplateD_j = \{\bar{d}_{j,k}\} \quad (j=1,2,\dots,N_{Level}, k=1,2,\dots,N_{Samples}) \quad (9-4)$$

Let  $C_{j,p} = \{c_{j,k,p}\}$  ( $j = 1,2,\dots,N_{Level}, k = 1,2,\dots,N_{Samples}$ , and  $p = 1,2,\dots,N$ ) be the set of all approximation coefficients at scale  $j$  of the  $p^{\text{th}}$  epoch. The cross-covariance between  $TemplateC_j$  and  $C_{j,p}$  is then calculated as follows:

$$XCOR_{j,p}(n) = E((TemplateC_{j_m} - \mu_{TemplateC_j})(C_{j,p_{m+n}} - \mu_{C_j})^*) \quad (9-5)$$

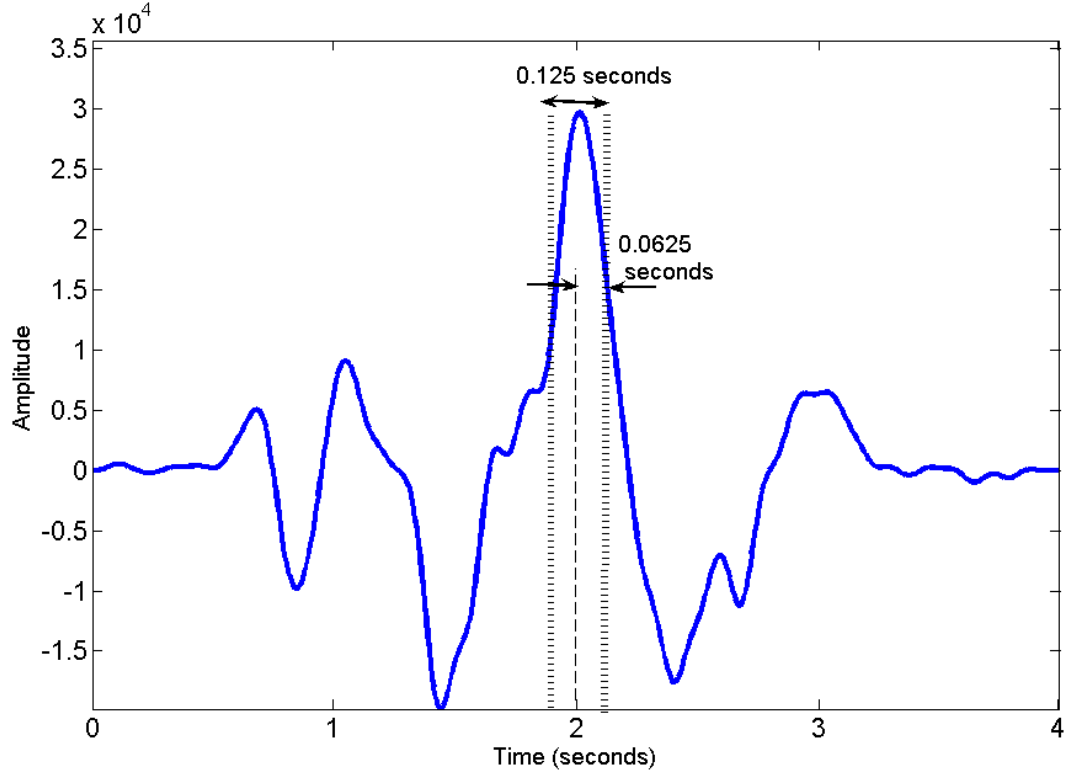
where  $E$  is the expected value operator. After calculating  $XCOR_{j,p}(n)$  for each epoch, the following features, representing the maximum of the cross-correlogram over a period of 0.125 seconds, are extracted:

$$F_{j,p} = \text{Max}_n[XCOR_{j,p}(n)], \quad n \in [t_{start} + t_{finish} - 0.0625, t_{start} + t_{finish} + 0.0625] \quad (9-6)$$

where  $(t_{finish} - t_{start})$  is the length of the epoch, and  $t_{finish}$  and  $t_{start}$  show the start and finish of an epoch. Figure 9-2 demonstrates an example of this feature extraction method, assuming that the length of the epoch and the template are both  $t_{finish} - t_{start} = 2$  seconds (the duration of the cross-covariance signal will thus be four seconds). The feature extractor considers a window of width 0.0625 seconds around the middle point of the cross-correlogram (i.e., at time  $t = t_{finish} - t_{start} = 2$  seconds). This window covers the time from  $t_1 = \frac{0.125}{2} = 0.0625$  seconds before to  $t_2 = \frac{0.125}{2} = 0.0625$  after  $t = 2$  seconds.

The maximum value in (9-6) is then calculated over this window with width of 0.125

seconds,. Features are then generated by sliding a window over the EEG signal in steps of 0.125 seconds (see Figure 9-2).



**Figure 9-2.** An example of how features are extracted using the proposed cross-covariance method.

Apart from the above features, the following features are also extracted:

$$T_{j,p} = t(XCOR_{j,p}(n) = F_{j,p}) \quad (9-7)$$

where  $t$  is the time operator. This feature provides information about the time instant when the maximum of the cross-correlogram occurs. Similar formulae can be obtained for the detail coefficients and for the features extracted from the NC epochs. This process is repeated for all EEG channels. Because MRPs are of low-frequency nature, we select the features belonging to the coarsest approximation and detail levels. As a result, four MRP features are generated for each EEG channel.

To extract features related to the CPMR and CPBR, EEG signals are band-pass filtered before feature extraction. For CPMR, the band pass is chosen from 8 to 12Hz and for CPBR, the band-pass is chosen as 22 Hz to 30 Hz. Both filters are linear phase 32-

point FIR filters. The amplitudes of the bandpass-filtered signals are squared to obtain their power values. SWT is then applied and the wavelet coefficients of the power signals are calculated. The rest of the feature extraction process is similar to that used for MRPs and it yields four CPMR features and four CPBR features for each EEG channel.

The wavelet function for each neurological phenomenon and for each EEG channel is selected from a pool of Daubechies, Biorthogonal, Symlet and Coiflet wavelet functions (46 wavelet functions in total). This is done by calculating the Fisher ratios for each channel. Please refer to [17] for more details.

### 9.3.3 Feature classification

The features for each neurological phenomenon in an EEG channel are classified as an IC or NC state using a support vector machine (SVM) classifier. Prior to classification, outliers were removed as follows. Suppose the Mahalanobis distance for a feature vector with  $K$  variables,  $\mathbf{x} = [x_1, x_2, \dots, x_K]$  with an assumed central point  $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_K]$  is defined as

$$Mahal(\mathbf{x}, \boldsymbol{\mu}) = (\mathbf{x} - \boldsymbol{\mu}) \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})^T \quad (9-8)$$

where  $\Sigma$  is the covariance matrix evaluated from the data. The outliers are then removed using the following algorithm [29, 30]:

1. Round  $p$ . If there exists  $\mathbf{x}$  such that  $Mahal(\mathbf{x}, \boldsymbol{\mu}) > \lambda$ , let  $FS = \{\mathbf{x} \mid Mahal(\mathbf{x}, \boldsymbol{\mu}) \leq \lambda\}$ . Retain only the points in  $FS$ . The value of  $\lambda$  was chosen such that all training samples further than 4 standard deviations from the mean were considered as outliers[31].

2. Repeat until the above condition is not met.

After applying this algorithm, the maximum percentage of features recognized as outliers was 1% for both IC and NC features.

The proposed configuration resulted in a total of  $3N$  classifiers. The output of each SVM is a logical state ‘1’, when an IC pattern is detected and is ‘0’ in other cases. We use the LIBSVM software for implementing the SVMs [32] and a Gaussian kernel

for the kernel function. The classifier's performance depends on the regularization parameter  $C$  and the bandwidth  $\sigma$  of the kernel. Since there are  $3N$  classifiers,  $3N$  values had to be estimated for each parameter. The method of determining these values is described later in this section.

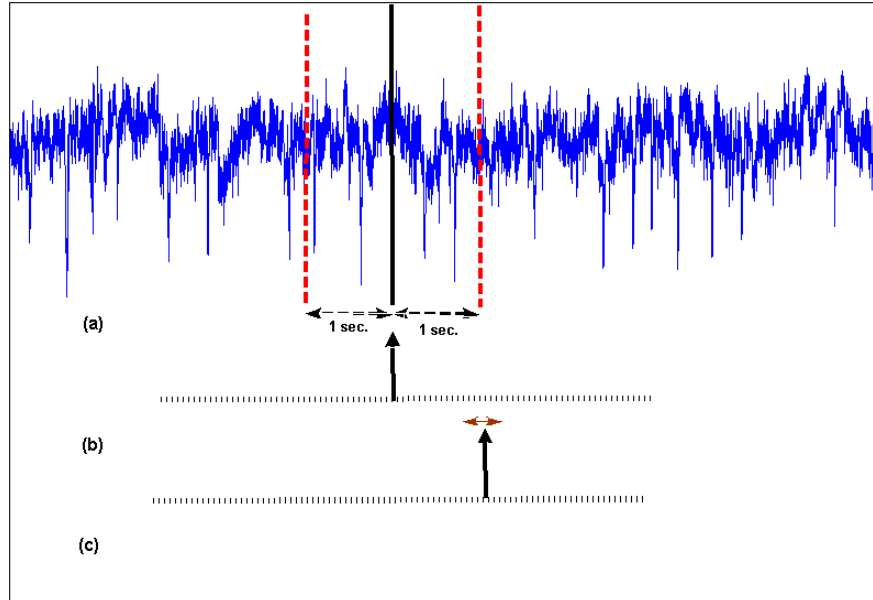
#### 9.3.4 Multiple classifier system

For each neurological phenomenon, a multiple classifier system (MCS) with a majority voting rule classifies the binary outputs of the SVMs (there are  $N$  SVMs for each neurological phenomenon). In the case the number of classifiers was even and if both classes have equal number of votes, the more-frequent class (NC) is chosen as the label for the input pattern. The outputs of the three MCSs are then combined using a 2nd-stage MCS as shown in Figure 9-1. This MCS can have the following five configurations for combining the outputs of the three MCSs: (1) Configuration 1: the AND rule is used to combine the binary outputs of MCS1 and MCS2 related to MRP and CPBR, respectively. The default class is an NC (the logical state '0'), unless both MCS1 and MCS2 identify an IC command (the logical state '1'); (2) Configuration 2: the AND rule is used to combine the binary outputs of MCS1 and MCS3 that are related to MRP and CPMR, respectively; (3) Configuration 3: the AND rule is used to combine the binary outputs of MCS2 and MCS3 related to MRP and CPBR, respectively; (4) Configuration 4: the outputs of all three MCSs are combined according to the majority voting rule; (5) Configuration 5 : the AND rule is used to combine the outputs of all MCSs. The choice of the best configuration is done during the model selection procedure as explained later in this section.

#### 9.3.5 Calculating the TPs and FPs

The method of calculating the TP rate is shown in Figure 9-3. In Figure 9-3(a), a sample EEG signal and in Figure 9-3(b) the output of the physical switch are shown. As stated earlier, data from one second before to one second after a decision point are used for classification. Assuming the system has no processing delay and the SBCI system has the ideal detection rate, the output of the SBCI system should be as demonstrated in Figure 9-3(c). In other words, the IC command should be detected one second after

pressing the switch. Although, the exact timing of the switch activation is known, the neurological phenomena may not be completely time-locked to the switch activation. As a result, we also consider any activation in the time range  $[-0.125, +0.125]$  seconds around the expected activation of the switch as a true positive activation (see Figure 9-3(c)). Any activation in the NC epochs is considered as a false positive one.



**Figure 9-3. Method of calculating the TP rate; (a) EEG Signal; (b) Output of the finger switch; (c) Output of the SBCI.**

### 9.3.6 Metric selection for model evaluation

The choice of the evaluation metric for summarizing the performance of SBCI systems has been under debate by BCI researchers. In synchronized BCI systems, the Overall classification Accuracy (OA) and the Information Transfer Rate (ITR) are widely used metrics for comparing the performance of synchronized BCI systems. This is not the case with self-paced BCI systems, though. It is very difficult to compare the performance of different self-paced BCI systems, as a wide variety of metrics such as OA[6], HF-difference[20], mutual information (information transfer rate) [21], Kappa [22], area under the Receiver Operating Characteristic (ROC) curve [22], the TP rate at a fixed FP

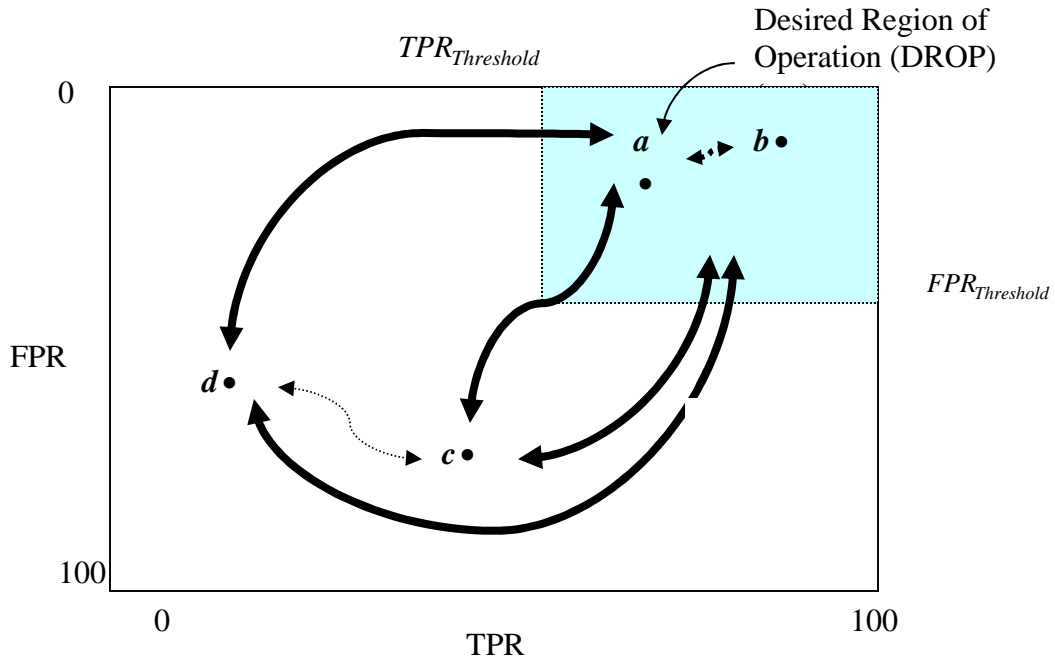
rate [11] among others have been proposed in the literature. Currently, no consensus exists amongst SBCI researchers as to which metric is more suitable for summarizing the performance and how a suitable evaluation metric should be chosen for a particular SBCI system [3].

In this chapter, we apply the framework in Chapter 8 to find the most suitable evaluation metric for the model selection procedure. Consider a 2D space, where one axis denotes TPR and the other axis denotes FPR. We call this domain the (TPR,FPR) domain. To apply the framework proposed in Chapter 8, first a desired region of operation (termed as DROP) is defined on the (TPR,FPR) domain. This region defines the area we desire the BCI system to operate in. In other words, the model selection procedure should guide the model of the system to this region (see Figure 9-4 for a demonstration). Next, we compare the performances of  $K$  evaluation metrics to determine which is the most suitable in summarizing the performance of the model. The following measures should then be calculated for this purpose (for the mathematical definitions and the formulae pertained to this sub-section, please refer to Chapter 8):

***Degree of Suitability (DoS)***: This metric determines if an evaluation metric yields higher values for the points inside the desired region of operation (DROP) region compared to the points outside the DROP region. The values of DoS vary between 0 and 1 where ‘0’ means that all points inside DROP have lower values compared to the points outside DROP. A value of “1” indicates the opposite case. It is thus desirable that evaluation metrics have DoS values close to unity.

After calculating DoS, the evaluation metrics that are found to be suitable, can be compared using the following two measures:

***Degree of Consistency (DoC)***: When moving on the (TPR,FPR) domain, this measure determines how consistent two metrics are with each other. In other words, when moving from an arbitrary point  $a$  to an arbitrary point  $b$  on the (TPR,FPR) domain,



**Figure 9-4.** An example of dividing the (TPR,FPR) domain into regions. Different movements on the (TPR, FPR) space may be associated with different weights. Note that the numbers on each axis denote (%).

DoC indicates whether the changes in the values of the evaluation metrics are in the same direction or not. The values of DoC vary between 0 to 1. The value of 1 means that both metrics vary in the same direction while a value of “0” means that the metrics vary in opposite directions. In order for two metrics to be considered as consistent with each other, it is desired that  $DoC > 0.5$  [33]. If  $DoC \leq 0.5$ , then this framework cannot be used for comparing the evaluation metrics. However, based on our previous simulations in Chapter 8 and the simulations in this paper, DoC values between of all metrics tried in SBCI systems were higher than 0.5.

***Degree of Discriminancy (DoD):*** Once two metrics are found to be consistent with each other ( $DoC > 0.5$ ), DoD is applied in order to measure how discriminant one metric is over the other. This is done by calculating if one metric is better than the other in guiding the model selection procedure towards DROP. In order for a metric  $f$  to be more discriminant than metric  $g$ , it is desired that  $DoD(f, g) > 1$  and vice versa [33].



Let  $f$  and  $g$  be two evaluation metrics defined on the (TPR,FPR) domain. If  $\text{DoC}(f,g) > 0.5$  and  $\text{DoD}(f,g) > 1$ , then metric  $f$  is said to be superior to metric  $g$  on the (TPR,FPR) domain. In this Chapter, we use this framework to select amongst many metrics the most suitable one.

### 9.3.7 Model selection

A hybrid genetic algorithm (HGA) is designed to guide the model selection procedure towards the DROP area. In other words, the HGA performs the following tasks: 1) it selects the features; 2) determines the values of the classifiers' parameters; and 3) selects the best of the five MCS configurations described in the previous subsections.

To represent each possible combination of features, a binary chromosome of length  $L_{\text{Chromosome}}$  is defined. Bit  $i$  of the first  $N_{\text{features}}$  bits of the binary chromosome specifies whether or not feature  $i$  is selected by HGA. A value of "1" indicates the presence of feature  $i$  and a value of "0" indicates its absence in the chromosome. The second part of the chromosome is used to select the parameter values of the classifiers. For each of the  $3N$  SVM classifiers, two parameter values need to be determined: the regularization parameter  $C$  and the bandwidth of the Gaussian kernel ( $\sigma$ ). A portion of the chromosome with length 8 bits is used for the two parameter values. The first four bits are used to represent the value of  $C$  and the second half is used to represent the value of  $\sigma$ . Exponentially growing sequences are used for  $C$  and  $\sigma$ , i.e., their values vary from  $2^{-8}$  to  $2^7$ . For each chromosome, a local search is then carried out to find the best of the five configurations in the 2<sup>nd</sup>-stage MCS.

The operators of HGA are selected as follows: tournament-based selection (tournament size =3), uniform crossover ( $p=0.9$ ) and uniform mutation ( $p=0.01$ ). The sizes of the initial population and the rest of the populations are chosen as 200 and 100, respectively. HGA is randomized initially. Elitism is used to keep the best performing chromosome of each population in the subsequent populations. The number of evaluations is set to 5000. If for more than 10 consecutive generations, the improvement

in the fitness of the best solution was found to be less than 1%, the algorithm is terminated. We discuss the choice of the fitness function in the next Section.

### **9.3.8 Evaluation**

A nested cross-validation was used to analyze the performance of the SBCI system. The inner cross-validation set was used for selecting the best chromosome and the outer cross-validation set was used to test the performance. For each outer cross-validation set, 20% of the data were used for testing and the rest were used for training. The training datasets were further divided into five folds. For each fold, 80% of the data were used for training the SVM and 20% were used for choosing the best chromosome.

### **9.3.9 Using ROC curves for summarizing the performance on test sets**

A receiver operating characteristic (ROC) curve depicts the relationship between TPR and FPR and thus provides a tradeoff between TPR and FPR. One problem with using the ROC curve is that when it is plotted over the whole range of TPR and FPR, most SBCI systems produce a curve that is similar to a perfect ROC curve [3]. In this study, we propose a new form of the ROC curve for SBCI systems. We define the false activation rate (FAR) as the ratio (percentage) of 1-second NC epochs that contain at least one false positive decision to the total of number of 1-second NC epochs [18]. For example, FAR of  $x\%$  shows that false positive decisions occurred in  $x$  different seconds in every 100 seconds of NC data on average. In the proposed ROC curve, the  $x$ -axis is the “partial FAR(%)”, i.e., only values of FAR% that are within the acceptable operating range of the system. The  $y$ -axis is to the same as the traditional ROC curve, i.e., TPR values are plotted. One advantage of this new ROC plot over the traditional ROC plot is that the interpretation of false positive values is more easily apparent. Please note that in the traditional ROC curves, TPR values are plotted vs. FPR. Since the rate at which the system generates an output is not considered in the ROC plot, it is not evident whether the reported values of TPR at a particular FPR are meaningful. For example, from the practical point of view, a TPR of 50% at an FPR of 1% can be meaningless, if the system generates 100 outputs per second, i.e., at the rate of 100Hz. This is because this system generates one false positive decision every second. So even though the FPR value is low,

this does not generate a useful SBCI system. On the other hand, the use of FAR as the  $x$ -axis is meaningful as it clearly shows the percentage of false positives as well as “how frequently” false positive decisions have occurred.

ROC plots were created as follows: as described earlier, each MCS consists of  $L$  classifiers, each representing the results of classifications from one classifier corresponding to a particular EEG channel. Thus, for the monopolar montage, each MCS consists of 13 classifiers and for the bipolar montage each MCS consists of 18 classifiers. Since one MCS is designed for each neurological phenomenon, there are three MCSs in total. The final classification label of each MCS, as discussed earlier, depends on the majority voting of the individual classifiers in that MCS. For example in an MCS with 13 classifiers, at least 7 votes are needed to label an input pattern as class “IC” or else it is labeled as class “NC”. The main rationale is that after training the classifier, IC epochs get the majority number of votes compared to NC epochs. The ROC curve can thus be plotted by changing the number of votes from the “majority voting” to “every number of votes”. In other words, the values of TPR and FAR are calculated when the number of votes that are necessary for labeling an epoch as an IC varies from “1” vote to  $L$  votes in step of 1. For each specific number of votes, TPR and FAR values are calculated and then the results are plotted. It is apparent that when the threshold is set to “1” vote, most outputs are labeled as “IC”, since only one vote out of all  $L$  votes is necessary to label an epoch as IC. In contrast, when the threshold is set at  $L$ , all classifiers should label the input pattern as “IC’s”. It is thus expected that only few patterns (if any) will be labeled as IC. The rest of the cases fall between these two extremes.

## 9.4 Results

In this section, we present the results of this study. First, the results of applying the framework to select a suitable evaluation metric is presented. Then the ROC curves for all participants are presented and the results from monopolar and bipolar montages are compared.

### 9.4.1 Choosing the evaluation metric for model selection

We considered four evaluation metrics which have been proposed in the literature for the evaluation of SBCI systems.: Overall Accuracy (OA) [6], Kappa [22], HF-difference [20] and  $\frac{TPR}{FPR}$  [17, 34]. We defined DROP as  $TPR \geq 50\%$  and  $FPR \leq 1\%$ .

These values were chosen based on the results obtained in the previous studies of our research group [35]. After applying the proposed framework, all four metrics were found to yield high DoS values ( $DoS \approx 1$ ). The value of DoC between all metrics was also found to be above 0.8, so all metrics were highly consistent with each other. After calculating DoD, the Kappa coefficient was found to be the most discriminant amongst the four metrics with  $DoD(Kappa, HF) > DoD(Kappa, \frac{TPR}{FPR}) > DoD(Kappa, OA) > 1$ .

Thus, the Kappa coefficient was chosen as the evaluation metric for this study.

The Kappa coefficient is a measure of agreement between two estimators [36] and is calculated as follows:

$$\kappa = \frac{OA - p_e}{1 - p_e} \quad (9-9)$$

where  $p_e$  is the chance agreement and is defined as follows:

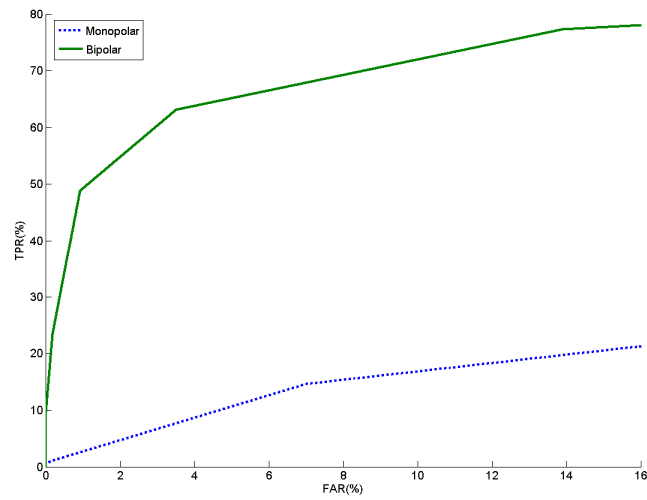
$$p_e = \frac{\sum_{i=1}^M N_{:i} N_{i:}}{(N_{Total})^2} \quad (9-10)$$

Here,  $N_{:i}$  is the sum of the elements in the  $i$ -th column and  $N_{i:}$  is the sum of the elements in the  $i$ -th row in the confusion matrix.

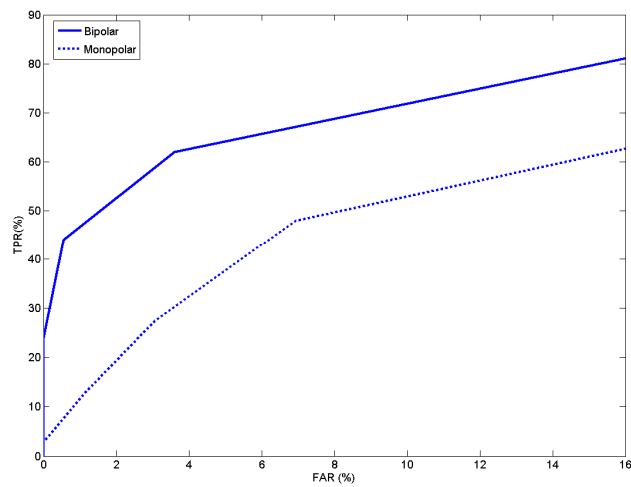
### 9.4.2 Performance of the system

The ROC curves of all participants are shown in Figure 9-5. The plots are averaged over the outer validation sets and thus each represents the average of 5 plots. As seen for AB1, AB2 and AB4, the bipolar montage resulted in a superior performance compared to the monopolar montage. In contrast, for AB3, the performances of both

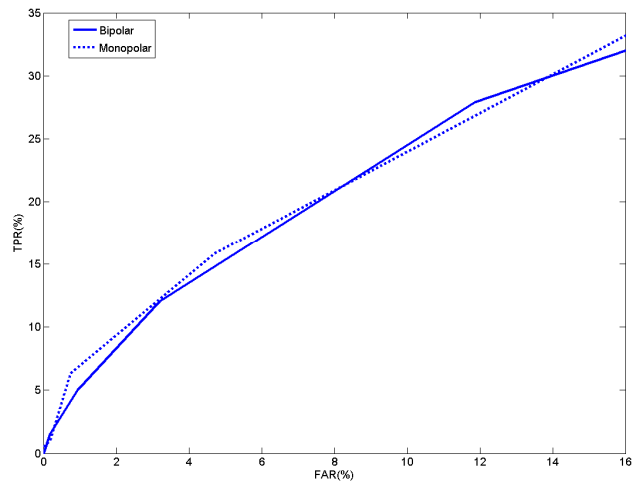
montages were the same and below the average. Participant AB5, was the only one where monopolar montage outperformed the bipolar montage for most FAR values.



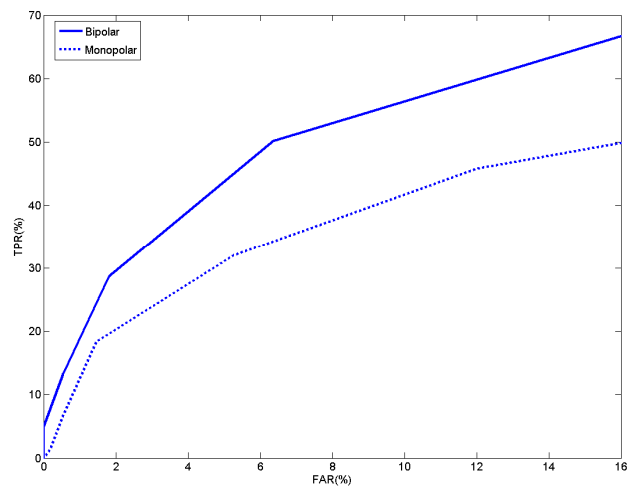
(a)



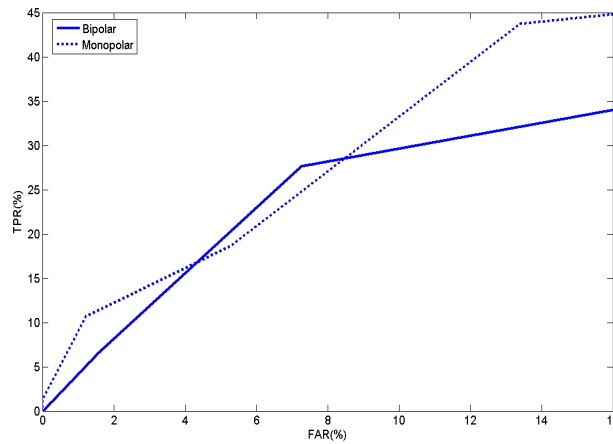
(b)



(c)



(d)



(e)

**Figure 9-5. ROC plots for (a) AB1; (b)AB2;(c)AB3;(d) AB4;(e)AB5.**

Table 9-1 compares TPR of monopolar and bipolar montages for different false activation rates. The 2<sup>nd</sup> column thus shows the average length of NC data between false positives (in seconds). The numbers in the table show the TPR values of each participant for a specific montage and a specific FAR. The numbers in parentheses are the standard deviations. The results are averaged over the 5 outer validation sets.

Figure 9-5 and Table 9-1 show that as we move from higher FAR values of 16% towards FAR=5%, the bipolar montage keeps modest TPR values. For the case of monopolar montage, the TP rate drops to 24.6%, corresponding to the detection of one out of every 4 IC commands. For FAR =2% (corresponding to an average of one false positive activation every 50 seconds), the bipolar montage can still detect 33.4% of the IC commands on average. However, the performance of the monopolar montage drops below 20%. When FAR =1 % (corresponding to an average of false positive activation every 100 seconds), the bipolar montage still recognizes one out of four IC commands on average, while for the case of monopolar montage, the TPR corresponds to the detection of one out of 10 IC commands, which is very low. Overall, the bipolar montage outperformed the monopolar montage for all FAR values studied.

**Table 9-1. Comparison of the TP rates of monopolar and bipolar montages for different false activation rates. The numbers in parentheses show the standard deviations.**

FAR (%)	Average length between false positives	Referencin g	TP Rate					
			AB1	AB2	AB3	AB4	AB5	Average
16	6.25	Monopolar	23.3 (5.2)	62.1 (7.6)	37.1 (7.7)	48.9 (4.1)	44.4 (12.4)	43.2 (14.4)
		Bipolar	77.7 (3.4)	80.5 (6.8)	34.1 (8.3)	68.6 (4.3)	35.1 (9.8)	<b>59.2</b> <b>(22.9)</b>
10	10	Monopolar	17.4 (5.2)	53.2 (7.2)	30.7 (8.0)	42.4 (5.0)	31.0 (15.3)	34.9 (13.5)
		Bipolar	74.3 (4.3)	73.9 (8.3)	25.8 (6.8)	59.0 (2.0)	30.2 (10.3)	<b>52.6</b> <b>(23.4)</b>
5	20	Monopolar	11.3 (5.7)	35.2 (6.8)	19.9 (5.9)	35.7 (6.0)	20.7 (14.3)	24.6 (10.6)
		Bipolar	70.9 (4.7)	65.5 (9.6)	15.5 (3.5)	44.1 (2.7)	19.2 (14.7)	<b>43.0</b> <b>(25.5)</b>
2	50	Monopolar	5.8 (3.5)	19.5 (8.6)	12.6 (4.9)	22.9 (10.2)	13.7 (10.5)	14.9 (6.6)
		Bipolar	62.0 (6.4)	55.1 (12.1)	8.3 (2.7)	27.4 (5.7)	14.2 (11.9)	<b>33.4</b> <b>(24.1)</b>
1	100	Monopolar	3.1 (2.5)	14.7 (7.9)	7.1 (4.4)	15.6 (9.6)	12.2 (9.0)	10.5 (5.3)
		Bipolar	56.4 (7.8)	47.1 (14.8)	5.1 (2.0)	19.7 (8.7)	9.7 (9.0)	<b>27.6</b> <b>(22.9)</b>

## 9.5 Discussion and future work

### 9.5.1 Discussion

In this paper, we carried out a new study using the SBCI system we had previously proposed in [17]. First, we trained the system using data recorded from a new type of movement (hand extension vs. the traditional finger flexion previously used in [15, 17, 18, 34]). Our aim was to study if the system would perform well on a new type of movement as it did on the right index finger flexion. This is a necessary step towards generalizing the system to more IC commands. Secondly, we recorded NC signals in a self-paced environment (compared to the cue-based recording that was done in our previous studies). For that, we used more engaging NC data (along with IC data) to train



and evaluate the performance of the system. Since online tests may involve periods that require the user's attention, but do not involve an initiation of control by him, this part of the study forms an important step towards running online tests on the system. We also applied a better artifact monitoring system that does not only monitor the activities of eyes, but also those of the frontalis muscles as well. To find a suitable evaluation metric for the model selection procedure, we also applied the framework proposed in Chapter 8. This is the first time in the SBCI literature, that the choice of the evaluation metric is quantitatively justified. Finally, we compared the performances of the monopolar and bipolar montages. Although using the bipolar montage in our proposed scheme requires more computational effort than the monopolar montage, however, from the performance point of view, this is justified. As shown in Figure 9-5 and Table 9-1, for different FAR values, the bipolar montage outperformed the monopolar montage by a wide margin.

It is difficult to compare the results obtained from this paper directly with those obtained from previous SBCI studies. Various reasons can be cited for this. The protocols for conducting the experiments, gathering IC and NC epochs, evaluating the performance, type of movement, the neurological phenomena used, and the number of participants vary significantly amongst the studies. For this reason, most SBCI papers simply compare the results with their previous work. Having said that, in Table 1-1, we report a summary of the results obtained in various EEG-based SBCI studies. Please keep in mind that although a direct comparison is not possible, this table roughly hints at the relative performance of different SBCI systems. The rows of this table show different SBCI studies. The columns show the rate at which the system generates an output, TPR, FPR and FAR rates, respectively. Please note that studies 1,2, and 3 reported high TPR values for high FPR values using the ROC curve. However, in order to make a fair comparison, the TPR values had to be estimated for low FPR values from the ROC plots in these papers.

Study 1 (the first three rows in Table 1-1) compares three 2-state SBCI designs[5], the low frequency –asynchronous switch design (LF-ASD), another SBCI system that uses Mu rhythms (“Mu-ASD”) and a third SBCI system termed “Outlier Processing Method” (OPM)[7]. All three systems generated outputs at the rate of 16Hz and TPR values were based on FPR of 2%. This is because in another study on an SBCI

with a similar decision rate, FPR higher than 2% were found to be irritating for the users [6].

Study 2 examines some modifications in the design of LF-ASD and proposes a new Hybrid LF-ASD [13]. The 3<sup>rd</sup> study proposes two separate 2-state SBCI designs [37]. One design is based on the detection of the right index finger movement and the other design is based on the detection of the left index finger movement. The following row refers to the results of the 4<sup>th</sup> study involving automation of the feature extractor of the LF-ASD [15]. The results were obtained from a table in [15] that reported TPR at a fixed FPR of 2%.

The 5<sup>th</sup> row presents the results of a 3-state SBCI system recently proposed by our research group [35]. The design identified IC commands related to right and left hand extension movements. This is done by implementing a 2-stage classification scheme. In the first stage, IC classes were separated from the NC class and in the second stage, the IC classes were classified into IC1 and IC2. The results pertaining to the classification of IC state from NC state were reported and the TPR values were reported at a fixed FPR of 1%.

The 6<sup>th</sup> study refers to the results of the SBCI system used in the current study, when originally applied to the data collected from the right index finger flexion movements [17]. The final study shows the results obtained in this paper.

Please note that in studies 1 to 5, when the FP rates are low (e.g., less than 0.5 %), the TP rates are very low as well. As a result, from a practical point of view, the application of these systems is limited. In contrast, the system we proposed in [17] and further investigated its performance here, has a good performance even for low FP rates. The results of Study 7 indicate that for three participants (AB1, AB2 and AB4), an average TP rate higher than 50% can be obtained even for FAR=5%. These results further demonstrate that our proposed design can achieve very low FP rates at moderate TP rates. They also mirror our previous findings in [17] that showed modest TP rates of

**Table 9-2. Comparison of the TPR and FAR rates achieved in different SBCI studies.**

<b>Study</b>	<b>Paper</b>		<b>Frequency (Number of classifier's decisions per second)</b>	<b>TPR(%)</b>	<b>FPR(%)</b>	<b>FAR(%)</b>
<b>Study 1</b>	[5]	LF-ASD	16	<20%	2	33
		OPM		<10%		
		Mu-ASD		<10%		
<b>Study 2</b>	[13]		25	30%	2	25
<b>Study 3</b>	[37]		?	<20%	2	?
<b>Study 4</b>	[15]		16	67.8	2	33
<b>Study 5</b>	[35]		16	54.0	1	16
<b>Study 6</b>	[17]		8	56.2	0.14	1.2
<b>Study 7</b>	Current study		16	33.4	0.12	2

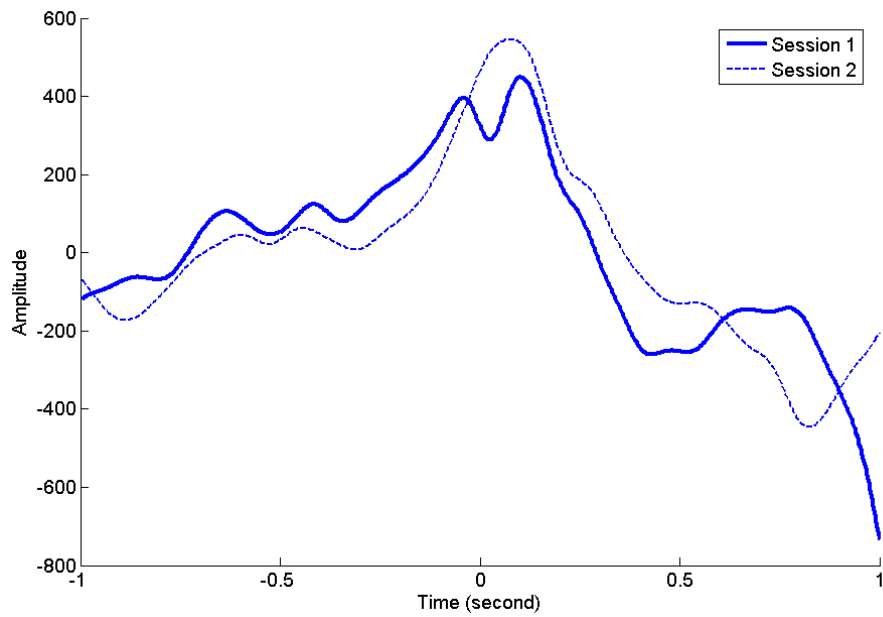
56.2% at FAR=1.2%. Although a direct comparison between our study and those reported in Table 9-1 is not totally valid, the results from this table show that our proposed SBCI system generates far less false positives compared to other EEG-based SBCI systems, at a modest TP rate. This is an important step in moving towards the online implementation of SBCI systems.

The only study, whose results outperformed the results obtained in this study, was the one we conducted using the same SBCI system earlier on finger flexion movements [17]. We believe the drop in the performance is due to the use of more engaging NC data in this study compared to [17]. As mentioned earlier, the “NC” term is applied to any

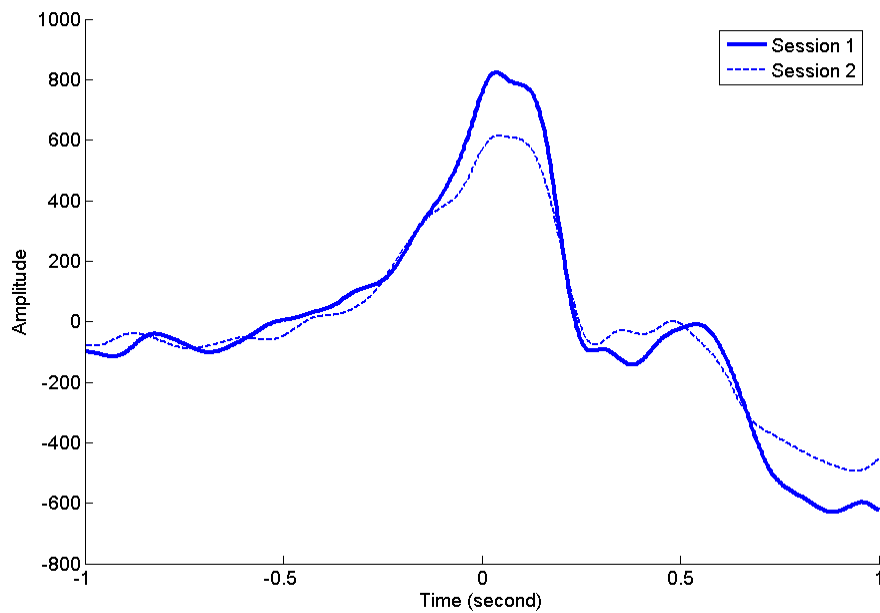
epoch that does not contain an IC attempt. Obviously, a long range of activities including being idle, thinking about a problem, doing mental calculations, watching an event, amongst others can all be categorized as NC. The question is that given the engaging nature of some of these tasks (such as thinking about a problem), what are their effects on EEG signals? In other words, would it be possible that they also generate patterns that are similar to those of neurological phenomena and thus affect the performance of the system? If this is the case, what is the extent of this effect? The evidence from our earlier research supports the hypothesis that more engaging NC data may result in desynchronization in the Mu and Beta rhythms, which would in turn affect the performance of the system [5]. In other words, we believe the decrease in the average performance of the system compared to [17] can be due to the more engaging nature of NC data used in this study compared to those used in our previous studies. Even though more engaging data were used, the performances of 3 participants (AB1, AB2 and AB4) are still comparable to the average performance achieved in [17]. More specifically, in the case of AB1, the proposed SBCI achieved a TPR=56.4% at FAR=1%, while for AB2, it achieved a TPR of 47.1% at FAR=1%. For AB4, the TPR results for FAR=1% are relatively low (the average TPR is 19.7%). However, when moving to FAR=5%, the average TPR rises to 44.1%.

As shown in the previous Section, the performances of AB3 and AB5 were worse than the other three participants. We believe the reasons for the below-the-average performance of these individuals are as follows. During the experiments, we observed that AB3 was less enthusiastic than the others and thus was less engaged with the experiment. Changes in the neurological phenomena of this individual from Session 1 to Session 2 may also have contributed to the poor performance. To further explore this hypothesis, we have plotted the averages of MRP patterns of a selected EEG channel (channel C1) for AB1 to AB4 in Figure 9-6. As seen, while for AB1, AB2, and AB4 the averages are fairly consistent between both sessions, this was not the case for AB3. These changes in the neurological phenomena could thus have contributed to the poor performance of the system on the data of this individual. AB5 was engaged in the experiment, however, this participant only attended in one session (the orientation session). The amount of data collected for AB5 was thus significantly less than the other

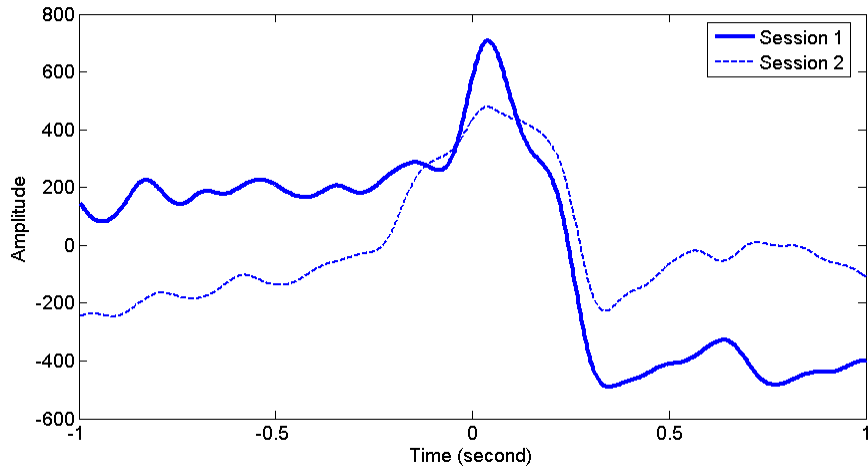
four participants. The lower amount of data might have resulted in the poor training of the SBCI system and thus the below average performance for this individual.



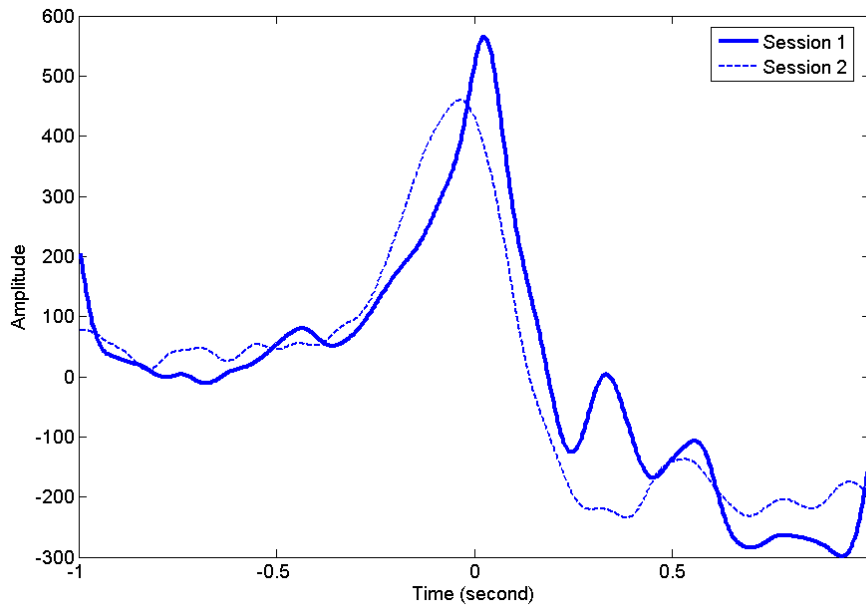
(a)



(b)



(c)



(d)

**Figure 9-6. Average MRPs for Channel C1 over two sessions. (a) participant AB1; (b) participant AB2 ; (c) participant AB3; (d) participant AB4;.**

### 9.5.2 Future works

A review of BCI literature shows that while many synchronized BCI systems have been tested online [38-41], only couple of SBCI studies have been conducted online[12, 42]. In other words, besides the above studies, almost all other SBCI designs have been tested in an offline fashion. The reason for conducting offline analysis in SBCI systems is that their FP rates are still very high. While from the pattern recognition point of view, offline SBCI designs can achieve very high classification accuracies, from the

practical point of view, they need further improvements until their FP rates become very low (especially when using the engaging NC data). Synchronized BCI systems, on the other hand, do not deal with NC data, as users can only control the device in certain periods determined by the system. Moreover, if the users of synchronized BCI systems are in NC state during the control periods, it is not evident how the output of the SBCI system can handle the NC state. Below, we will examine the performance of online SBCI studies in more details.

In [42], an SBCI system with three IC states is implemented. The system identifies patterns related to the right hand, left hand and foot movements and the resultant IC commands are used to control a virtual keyboard. While the system can be operated in a self-paced manner, i.e., the user does not need to wait for the system's cue to initiate the control, its performance on NC data was not evaluated. The users of this system were continuously engaged in controlling the virtual keyboard, so it is not known how well this system could perform on NC data (especially *engaging* NC data).

In [12], another online experiment was conducted using the LF-ASD SBCI system. Two able-bodied and two individuals with spinal-cord injury (SCI) participated in this experiment that consisted of two sessions. The experiment involved changing the direction of a center ball in a "pong" style display. The users were asked to use a sip and puff switch to report the classification errors. While this SBCI system considered NC states, the NC epochs were mostly collected between different IC attempts and in a relatively non-engaging environment. Even though less engaging NC data were used, the FP rates for the two sessions varied between 2-3% for one able-bodied (AB) participant (TP rate varied between 72.0% and 81.2% ) and varied between 1-1.5% for the other participant (TP rate for this participant was between 69.6% and 77.3%). Interestingly, when NC data related to reporting the classification errors by the users were added to the analysis, the FP rate was increased by 0.4% (Session 1) and 0.3% (Session 2) for the first AB participant and was increased by 0.9% (Session 1) and 0.8% (Session 2) for the second AB participant. These observations provide further evidence that more engaging environments may further decrease the performance of an SBCI system. For individuals with spinal cord injury (SCI), the average TPR was below 45%, since higher TP rates were associated with FP rates in the range of 2-3% and resulted in the fatigue and user

frustration. The FP rates of both SCI participants were, however, lower compared to able-bodied individuals. For the first SCI participant, the FP rate varied between 0.7 -1% and for the other participant, it varied between 0.1 and 0.5%. Similar to able-bodied subjects, FP rates were increased during the NC data associated with using the sip and puff switch. The amount of increase was 0.1-0.2% for the first SCI participant and 0.2-0.3% for the second participant.

As seen, online tests of SBCI systems have been conducted under very specific conditions so far. In other words, either NC data were eliminated from the analysis [42] or less engaging NC data that resulted in relatively high FP rates for high TP rates were used [12]. Based on these discussions, SBCI researchers still needs to focus on further improving the performance, before the system is ready for online experiments. This is especially the case when engaging NC data are used.

Since our proposed SBCI system has resulted in much lower false positives compared to other EEG-based SBCI systems (as demonstrated in the last two rows in Table 9-2), for future research we propose the following studies before the system is tested online:

- 1) As noted above, more analysis using NC data is necessary to ensure that the SBCI system performs well in different (and perhaps more engaging) NC environments. Our future work will thus explore NC data in more detail. Different NC data shall be recorded and the performance of the system once trained and tested using different NC data will be investigated. This would be a very important step towards developing practical SBCI systems.

- 2) Our studies so far were based on the actual execution of movements. This is because the evidence from the literature shows that motor imagery or attempted movements can modify the neuronal activity in the primary sensorimotor cortex areas very similarly to what is observed from the execution of real movements [43, 44]. Based on this view, the main difference between the real and attempted movements is that in the latter case, the execution of the movement is blocked at some cortico-spinal level [45]. Before the system can be tested online, another study is needed to demonstrate the performance for attempted movements.



3) Some research groups have shown the importance of user training in improving the performance of a BCI system [42]. These findings indicate that user training can further improve the performance of our proposed SBCI design. This training can be especially useful in the cases of AB3 and AB5, whose performances were lower compared to the other participants in the experiment.

4) Finding a proper application for testing the performance is another necessary step. As the results from Table 9-1 show, the system managed to maintain a good performance on the data of AB1, AB2 and AB4 even for  $FAR \approx 5\%$ . For AB3 and AB5, the error rates are higher, so a completely self-paced environment would result in multiple false activations and thus user frustration. As a result, the test environment implemented for these two individuals may need more constraints than the environment implemented for other participants. Future studies should thus involve finding “proper” test applications for different individuals.

## **9.6 Acknowledgements**

This work was supported in part by NSERC under Grant 90278-06 and CIHR under Grant MOP-72711. This research has been enabled by the use of WestGrid computing resources, which are funded in part by the Canada Foundation for Innovation, Alberta Innovation and Science, BC Advanced Education, and the participating research institutions.

## 9.7 References

- [1] S. G. Mason, A. Bashashati, M. Fatourehchi, K. F. Navarro and G. E. Birch, "A comprehensive survey of brain interface technology designs", *Ann. Biomed. Eng.*, vol. 35, no.2, pp. 137-169, Feb.2007.
- [2] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller and T. M. Vaughan, "Brain-computer interfaces for communication and control", *Electroencephalogr. Clin. Neurophysiol.*, vol. 113, no.6, pp. 767-791, Jun.2002.
- [3] S. G. Mason, J. Kronegg, J. Huggins, M. Fatourehchi and A. and Schloegl, "Evaluating the performance of self-paced BCI technology", Technical Report, available online: [http://www.bci-info.tugraz.at/Research\\_Info/documents/articles/self\\_paced\\_tech\\_report-2006-05-19.pdf](http://www.bci-info.tugraz.at/Research_Info/documents/articles/self_paced_tech_report-2006-05-19.pdf), 2006.
- [4] S. G. Mason and G. E. Birch, "Temporal control paradigms for direct brain interfaces - rethinking the definition of asynchronous and synchronous," in *Proc. HCI Int. Conf.*, Las Vegas, USA, 2005.
- [5] S. G. Mason and G. E. Birch, "A brain-controlled switch for asynchronous control applications", *IEEE Trans. Biomed. Eng.*, vol. 47, no.10, pp. 1297-1307, Oct.2000.
- [6] G. E. Birch, Z. Bozorgzadeh and S. G. Mason, "Initial on-line evaluations of the LF-ASD brain-computer interface with able-bodied and spinal-cord subjects using imagined voluntary motor potentials", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 10, no.4, pp. 219-224, Dec.2002.
- [7] G. E. Birch, P. D. Lawrence and R. D. Hare, "Single Trial Processing of Event Related Potentials Using Outlier Information", *IEEE Trans. Biomed. Eng.*, vol. 40, no.1, pp. 59-73, 1993.
- [8] E. Yom-Tov and G. F. Inbar, "Selection of relevant features for classification of movements from single movement-related potentials using a genetic algorithm," in *the Proc. 23 IEEE EMBS Int. Conf.*, vol.2, pp. 1364-1366, 2001.
- [9] B. Graimann, J. E. Huggins, A. Schloegl, S. P. Levine and G. Pfurtscheller, "Detection of movement-related desynchronization patterns in ongoing single-channel electrocorticogram", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no.3, pp. 276-281, Sep.2003.
- [10] A. Bashashati, M. Fatourehchi, R. K. Ward and G. E. Birch, "User customization of the feature generator of an asynchronous brain interface", *Ann. Biomed. Eng.*, vol. 34, no.6, pp. 1051-1060, Jun.2006.
- [11] J. F. Borisoff, S. G. Mason, A. Bashashati and G. E. Birch, "Brain-computer interface design for asynchronous control applications: improvements to the LF-ASD asynchronous brain switch", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 985-992, Jun.2004.
- [12] Z. Bozorgzadeh, G. E. Birch and S. G. Mason, "The LF-ASD brain computer interface: On-line identification of imagined finger flexions in the spontaneous EEG of able-bodied subjects," in *the Proc. IEEE ICASSP'00*, vol.6, pp. 2385-2388, 2000.
- [13] E. Yom-Tov and G. F. Inbar, "Detection of Movement-Related Potentials from the Electro-Encephalogram for possible use in a Brain-Computer Interface", *Medical and Biological Engineering and Computing*, vol. 41, no.1, pp. 85-93, Jan. 2003.

- [14] J. E. Huggins, S. P. Levine, S. L. Bement, R. K. Kushwaha, L. A. Schuh, E. A. Passaro, M. M. Rohde, D. A. Ross, K. V. Elisevich and B. J. Smith, "Detection of Event-Related Potentials for Development of a Direct Brain Interface", *J Clinical Neurophysiol*, vol. 16, no.5, pp. 448-455, Sep. 1999.
- [15] M. Fatourehchi, A. Bashashati, G. E. Birch and R. K. Ward, "Automatic user customization for improving the performance of a self-paced brain interface system", *Med. Biol. Eng. Comput.*, Vol.44, No.12, pp.1093-1104, Dec 2006.
- [16] A. Bashashati, S. Mason, R. K. Ward and G. E. Birch, "An improved asynchronous brain interface: making use of the temporal history of the LF-ASD feature vectors", *J. Neural Eng.*, vol. 3, no.2, pp. 87-94, Jun.2006.
- [17] M. Fatourehchi, G. E. Birch and R. K. Ward, "Applying a hybrid genetic algorithm in the design of a self-paced brain interface with a low false positive rate," in *Proc. IEEE ICASSP'07*, ,vol.4,pp. IV-1157; IV-1160, 2007.
- [18] M. Fatourehchi, G. E. Birch and R. K. Ward, "Application of a hybrid wavelet feature selection method in the design of a self-paced brain interface system", *J. Neuroengineering Rehabil.*, vol. 4, pp. 11, Apr 30.2007.
- [19] 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, no.6, pp. 428-433, Dec.1998.
- [20] B. Graimann, J. E. Huggins, S. P. Levine and G. Pfurtscheller, "Toward a direct brain interface based on human subdural recordings and wavelet-packet analysis", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 954-962, Jun.2004.
- [21] J. Kronegg, s. Voloshynovskiy and P. Pun, "Analysis of bit rate definitions for brain-computer interfaces," in the *Proc. Int. Conf. on Human-Computer Interaction (HCI'05)*, Las Vegas, Nevada, 2005.
- [22] A. Schlögl, J. Kronegg, J. Huggins and S. G. Mason, "Evaluation criteria in BCI research," in *Towards Brain-Computer Interfacing* (G. Dornhege, J. R. Millan, T. Hinterberger, D. McFarland and K. R. Muller, Eds.), MIT Press, 2007.
- [23] 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, no.3, pp. 386-394, Sep.1997.
- [24] H. Ramoser, J. Muller-Gerking and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement", *IEEE Trans. Rehabil. Eng.*, vol. 8, no.4, pp. 441-446, Dec.2000.
- [25] M. Essl and P. Rappelsberger, "EEG coherence and reference signals: experimental results and mathematical explanations", *Med. Biol. Eng. Comput.*, vol. 36, no.4, pp. 399-406, Jul.1998.
- [26] D. Cawley, "Power Word: Tetris-Like Game Using Words (v 2.0)", available online at <http://www.freevbcode.com/ShowCode.asp?ID=2427>.
- [27] 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, no.9, pp. 1580-1593, Sep.2003.
- [28] C. J. De Luca, "The use of surface electromyography in biomechanics", *J Appl Biomech*, vol. 13, no.2, pp. 135-163, 1997.

- [29] A. Blum, A. M. Frieze, R. Kannan and S. Vempala, "A Polynomial-Time Algorithm for Learning Noisy Linear Threshold Functions.", *Algorithmica*, vol. 22, no.1/2, pp. 35-52, 1998.
- [30] J. Dunagan and S. Vempala, "Optimal outlier removal in high-dimensional spaces", *J. Comput. Syst. Sci.*, vol. 68, no.2, pp. 335-373, 2004.
- [31] M. Prastawa, E. Bullitt, S. Ho and G. Gerig, "A brain tumor segmentation framework based on outlier detection", *Med. Image Anal.*, vol. 8, no.3, pp. 275-283, Sep.2004.
- [32] C. Chang and C. Lin, *LIBSVM: A Library for Support Vector Machines*. 2001.
- [33] J. Huang and C. X. Ling, "Using AUC and accuracy in evaluating learning algorithms", *IEEE Trans. Knowled. Data Eng.*, vol. 17, no.3, pp. 299-310, 2005.
- [34] M. Fatourechi, G. E. Birch and R. K. Ward, "A self-paced brain interface system that uses movement related potentials and changes in the power of brain rhythms", *J. Comput. Neurosci.*, vol.23, no.1, pp.21-37, Aug 2007.
- [35] A. Bashashati, R. K. Ward and G. E. Birch, "Towards development of a 3-state self-paced brain computer interface", *Computational Intelligence and Neuroscience*, In press.
- [36] J. Cohen, "A coefficient of agreement for nominal scales", *Educational and Psychological Measurement*, vol. 20, no.1, pp. 37-46, 1960.
- [37] G. Townsend, B. Graimann and G. Pfurtscheller, "Continuous EEG classification during motor imagery--simulation of an asynchronous BCI", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 12, no.2, pp. 258-265, Jun.2004.
- [38] A. Buttfeld, P. W. Ferrez and R. Millan Jdel, "Towards a robust BCI: error potentials and online learning", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no.2, pp. 164-168, Jun.2006.
- [39] E. C. Leuthardt, G. Schalk, J. R. Wolpaw, J. G. Ojemann and D. W. Moran, "A brain-computer interface using electrocorticographic signals in humans", *J. Neural Eng.*, vol. 1, no.2, pp. 63-71, Jun.2004.
- [40] 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.and Rehab. Eng.*, vol. 11, no.2, pp. 204-207, Jun.2003.
- [41] B. Obermaier, C. Guger, C. Neuper and G. Pfurtscheller, "Hidden Markov models for online classification of single trial EEG data", *Pattern Recognition Letters*, vol. 22, no.12, pp. 1299-1309, 2001.
- [42] R. Scherer, G. R. Muller, C. Neuper, B. Graimann and G. Pfurtscheller, "An asynchronously controlled EEG-based virtual keyboard: improvement of the spelling rate", *IEEE Trans. Biomed. Eng.*, vol. 51, no.6, pp. 979-984, Jun.2004.
- [43] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication", *Proceedings of the IEEE*, vol. 89, no.7, pp. 1123-1134, 2001.
- [44] J. A. Pineda, B. Z. Allison and A. Vankov, "The Effects of Self-Movement, Observation, and Imagination on mu Rhythms and Readiness Potentials (RP's): Toward a Brain-Computer Interface (BCI)", *IEEE Trans. Rehab. Eng.*, vol. 8, no.2, pp. 219-222, Jun. 2000.
- [45] M. Jeannerod and V. Frak, "Mental imaging of motor activity in humans", *Curr. Opin. Neurobiol.*, vol. 9, no.6, pp. 735-739, Dec.1999.

## CHAPTER 10 SUMMARY AND CONCLUSIONS

### 10.1 Summary

This research was motivated by key shortcomings in the design of self-paced brain computer interface systems (SBCI) systems. SBCI systems are assistive technology devices that allow users to control objects in their environment using their brain signals only and at their own pace. This is done by measuring specific features of the brain signal that pertain to intentional control (IC) commands issued by the user. As an example, when a user attempts to perform a specific movement such as flexion of the index finger, a desynchronization in the power of the Mu ([8-12Hz]) and the Beta ([18-30Hz]) rhythms of the EEG signals occur. Since the duration of these phenomena are time-locked to the time of an IC command, they can be used for detecting the initiation of control by the user.

In Chapter 1, we identified three research areas that need to be addressed in the design of SBCI systems: 1) Current SBCI systems have high false positive rates that make them unsuitable for most practical applications, even though they may obtain high true positive rates. As an example, the latest design of an SBCI system developed in our lab achieves average true positive rate of 73.4% at the false positive rate of 1% [1]. However, the performance of the system deteriorates for very low FP rates [2, 3]. , 2) Most BCI systems do not address the presence of physiological artifacts adequately. As a result, the performance of BCI systems might be vulnerable in the presence of artifacts; 3) Researchers in the field of SBCI have not decided on the choice of a suitable evaluation metric yet. Although several evaluation metrics have been proposed, SBCI

researchers have not yet agreed on how to choose a suitable evaluation metric for their specific application.

The goal of this thesis is to devise an SBCI system that achieves considerably lower false positives than other SBCI systems. This system should also achieve good performance in the presence of artifacts such as eye blinks that can change the shape of the neurological phenomena. Finally a suitable performance metric for evaluating the system's performance needs to be developed.

We summarize the findings of this thesis according to each chapter as follows:

### **10.1.1 Chapter 2: Improving the performance of LF-ASD by automatic user-customization**

In order to gain some insight into the operation of SBCI systems, we first analyzed and improved the performance of a state-of-the-art SBCI system. This system, called the low frequency – asynchronous switch design (LF-ASD), was previously developed in the brain interface laboratory of the Neil Squire society in 2001. It detects bipolar movement-related potential (MRP) patterns from six bipolar EEG electrodes. Since then, it has been subject to several improvements and different versions of it are implemented. In Chapter 2, we automatically user customized the feature generator of one of the most recent versions of the LF-ASD. For each user of this system, 31 parameter values had to be specified. These parameter values were originally determined by trial and error and we argued that this procedure is clearly sub-optimal. We proposed the use of a genetic algorithm for tuning the parameter values of the system. The system's performance was tested using the data attributed to the right index finger flexion (the same data are used as the basis of analysis in Chapter 3s, 4, 5 and 7). The contributions of this research can be summarized as follows:

- 1) Offline analysis of the data of eight individuals revealed that automatic user customization improves the true positive (TP) rate of the system by an average of 6.7% over that whose customization was carried out by a human expert (at a fixed FP rate of 2%). In other words, automatic user-customization leads to

improvements in the performance compared to the non user-customized LF-ASD and the LF-ASD version that was user customized by an expert.

2) It was shown that by using automatic user-customization, the performance of the individuals with spinal cord injury (SCI) improved more than that of able-bodied individuals. In the case of individuals with SCI, the average improvement in the TP rate was 9.8% compared to 3.6% that was achieved for able-bodied individuals. This is because the averages of MRPs for individuals with SCI were weaker than those of able-bodied people, so it was more difficult for a human expert to estimate the parameter values using the ensemble averages of MRPs.

3) The automation procedure relieves the designer of the cumbersome task of tuning the parameter values for each user. Especially since LF-ASD uses an energy normalization transform (ENT) that changes the shapes of MRPs, this automation method is necessary.

4) The results show great variations in the parameter values from one individual to another. These results emphasize the importance of user-customizing the feature generator in SBCI systems.

5) The proposed method also provides an upper threshold on the performance of that particular LF-ASD design. In other words, to achieve better performance for LF-ASD, some changes in its structure must be done. For this reason, subsequent studies of our research group on LF-ASD have mostly focused on changing its structure so as to obtain superior performance [3, 4].

### **10.1.2 Chapter 3: Using DWT to extract features**

In Chapter 3, we explored the use of discrete wavelet transform (DWT) as the feature extractor to replace the user-customized detector in LF-ASD. Since the wavelet transform explores both time and frequency information, is expected to be a more suitable feature extractor than those which work in the time or frequency domain only. Using DWT coefficients as features, however, has one disadvantage: the size of the feature space increases dramatically. As a result, given the low number of test samples from the IC class, a method for handling the large size feature space should be implemented. The contributions of this research are as follows:

1) We proposed the use of a 2-stage hybrid feature selection algorithm to solve the high-dimensionality problem. In the first stage, mutual information between the input features and the output classes is used to remove the less informative features, while in the second stage, a genetic algorithm is used to select the features that yield high performances. The proposed method was then tested on the data of four able-bodied individuals.

2) The performance of the system was tested using the features extracted from 13 monopolar EEG signals and from 18 bipolar EEG signals. The bipolar EEG signals were mathematically calculated as the voltage differences between two adjacent monopolar EEG electrodes. During our tests, the performance of the system based on monopolar features was weak and it resulted in high false positive rates. However, the system based on bipolar features achieved a performance that was superior to that of LF-ASD in three out of four individuals whose data were studied.

3) It was shown that the location of the selected features for each EEG channel varied considerably from one individual to another. These results support the hypothesis that obtaining superior performance depends on the proper channel selection for every individual.

4) The main deficiency in using DWT as a feature extractor is that it is shift variant. In other words, decisions should be made in very short time intervals if the IC pattern is to be correctly identified during online experiments. From the computational point of view, this is an important shortcoming of using DWT as feature extractor in an SBCI system. A solution to this problem is proposed in Chapter 5.

### **10.1.3 Chapter 4: Using three neurological phenomena as the source of control**

In Chapter 4, we discussed another research direction to improve the performance of SBCI systems. We proposed the idea of using three neurological phenomena (MRPs, CPMR and CPBR) instead of only one. Our rationale was as follows: extracting features from more than one neurological phenomenon increases the amount of information contained in the system. Thus, the performance of such an SBCI should be higher than



those that extract features from a single phenomenon. We also cited evidence from the literature as to why combining these neurological phenomena is suitable for improving the performance of an SBCI system. The summary of this research are as follows:

- 1) We proposed the idea of combining features extracted from three specific neurological phenomena (MRPs, CPMR and CPBR) for the first time in the BCI literature.
- 2) By using features from three neurological phenomena (instead of using features from one neurological phenomenon as done in many BCI studies), the size of the feature space greatly increases. We thus proposed a 2-stage multiple-classifier system (MCS) to effectively combine the features extracted from the above neurological phenomena. First, features extracted from each EEG channel and each neurological phenomenon were classified separately. Then, in the first stage of the proposed MCS, three MCSs combined the outputs of classifiers for each neurological phenomenon. In the second stage, another MCS combined the outputs of the MCSs developed in the first stage. By implementing this 2-stage MCS, we achieved a strong classifier, although each of the participant classifiers was weak. We showed that the proposed scheme achieved low FP rates ( $FPR < 1\%$ ) for three out of four individuals whose data were studied. However, the TP rates were relatively low, as they corresponded to the detection of one out of every four IC commands on average. The low FP rates were promising and they indicated that with better pattern recognition schemes, a superior performance can be achieved.
- 3) We demonstrated that the MCS that was designed in the first stage for classifying MRP features was always selected in the final configuration. In other words, the presence of MRP was necessary in achieving a good performance. However, the choice of the other MCS depended on the individual whose data were studied and it varied from one person to another.
- 4) Again, it was shown that the selected EEG channels varied from one individual to another. Also for each person, there were a number of EEG channels that were not selected during the analysis, so they could be omitted during the future

research. This will speed up the preparation process of these individuals in future studies.

#### **10.1.4 Chapter 5: Design of an automated SBCI system with low FP rates**

To achieve better TP rates at very low FP rates, in Chapter 5, we proposed an improved design for SBCI systems. This was done by combining the ideas presented in Chapters 3 and 4 to design an SBCI with low FP rates at modest TP rates. In this chapter, we accomplished the following:

- 1) We propose a new feature extraction method. This feature extractor uses a stationary wavelet transform (SWT) followed by matched filtering with a template. For each neurological phenomenon and for each EEG channel, we extract new features for each of the two lowest frequency bands: the maximum cross-covariance at the center of the correlogram and the time this maximum has occurred. As a result, a total of  $12N$  features are extracted where  $N$  is the number of EEG channels.
- 2) Similar to Chapter 4, for each EEG channel and for each neurological phenomenon, a classifier was designed. In total,  $3N$  classifiers were implemented. We chose support vector machines (SVMs), not only because they minimize the empirical risk (training error), but they can also minimize the confidence error (test error). Similar to the one proposed in Chapter 4, a 2-stage MCS that combines the outputs of the classifiers and generate the final classification label of the input pattern was implemented. A hybrid genetic algorithm (HGA) was proposed to automate the design process of the improved SBCI. The HGA simultaneously selects the features, estimates the classifiers' parameters and chooses how the outputs of MCSs developed for each neurological phenomenon, should be combined together.
- 3) The value of calculated FP rates depends on the rate by which the system generates an output. In order to remove this dependency, we proposed the notion of the false activation rate (FAR). FAR is the percentage of 1-second NC epochs that have at least one false positive. As seen, this definition does not depend on

the output rate of the system and is thus more meaningful than the traditional FP rate definition.

4) Analysis of the data obtained from four able-bodied individuals showed that the improved SBCI performs significantly better than previous ones. In other words, while the average TP rates are above 50%, the proposed SBCI system produced far less FAR than other EEG-based SBCI systems.

5) The results also showed that using either the combination of MRP and CPBR-based classifiers or the “AND” configuration of all three MCSs results in a performance that is better than those of other configurations studied (e.g., majority voting of all three MCSs). However, the choice of the configuration varied from one individual to another.

6) We also provide a theoretical analysis of the performance of the MCS in Appendix A. Using the linear programming theory, it was shown that even if the performance of each classifier is not high, the overall performance can still approach the perfect classification accuracy by using the proposed hybrid MCS method (assuming that the classifiers are sufficiently diverse).

#### **10.1.5 Chapter 6: Analysis of the effect of artifacts in BCI systems**

In this chapter, we provided a detailed review of artifacts in SBCI systems. The summary of this chapter is as follows:

1) A detailed survey of the methods of handling artifacts in BCI systems was carried out. We reviewed how different BCI research groups have handled EOG and EMG artifacts in their proposed designs. This survey showed that the majority of BCI systems did not properly report if they have handled artifacts. This would cause a serious drawback when these systems are applied in an online fashion, as it is not evident how these systems would perform when artifacts occur.

2) We suggested two solutions for the proper handling of artifacts: the SBCI researchers should either design efficient artifact-removal methods or SBCI systems that are robust to the presence of artifacts.

### 10.1.6 Chapter 7: Analysis of the performance of the proposed SBCI on artifact-contaminated data

The performance of BCI systems are usually reported on data not contaminated by artifacts. This is not a totally realistic approach, specifically if a BCI system is to be used in an online environment. In this chapter, we analyzed the performance of the SBCI system proposed in Chapter 5 when datasets contaminated with artifacts were used. Although EOG and EMG artifacts have a large frequency range, they are more prominent over certain frequency bands. We postulated that the proposed SBCI should have a good performance when artifact-contaminated data are used. This is because the proposed system uses features extracted from three neurological phenomena that belong to different frequency bands. The summary of this chapter is as follows:

- 1) We studied the performance of the proposed SBCI system over the data contaminated with EOG artifacts. It was shown that the performance of the proposed SBCI slightly deteriorates, but remains superior to those achieved by other SBCI systems. Furthermore, the advantage of including the artifact-contaminated epochs in the analysis is that the system is available for control at *all* times.
- 2) We studied the performance of the system over data from a session recorded a few days after recording the data used in training the system (called the data in “a subsequent session”). The importance of this study is to analyze how the system performs using pseudo-online data. No part of the data (including those with artifacts) was rejected in this analysis. The results show that the proposed SBCI system still achieves a good performance on these datasets, although no tuning in its parameter values was performed.
- 3) We also studied the effect of adding a debounce component as a post-processing component. A debounce component reduces consecutive false activations from activating the system repeatedly. It was shown that a very small debounce window was needed to increase the performance of the proposed SBCI system. In other words, the number of consecutive false activations in the

proposed SBCI was less than those of LF-ASD whose window size was much larger.

#### **10.1.7 Chapter 8: A framework for evaluating the performance of SBCI systems**

The issue of how to evaluate SBCI systems remains an unresolved one. Several evaluation metrics have been proposed, but no consensus yet exists amongst researchers as to how best select a suitable evaluation metric for a particular SBCI. In this chapter, we address this issue. In particular the contributions of this chapter are:

- 1) The definitions of the degree of consistency (DoC) and degree of discriminancy (DoD) measures proposed in [5] were modified so as to they are applicable to pattern recognition problems with a large number of test samples. These measures can be used to select the most suitable evaluation metric given a number of evaluation metrics for a particular classification problem.
- 2) We proposed a new measure, the degree of suitability (DoS), to calculate how suitable an evaluation metric is in summarizing the performance of a classification problem.
- 3) To decrease the number of calculations needed in estimating the above measures, we proposed the idea of grid-based computing.
- 4) We applied the proposed framework to a particular SBCI system and demonstrated that metrics such as Kappa and HF-difference outperform OA (overall accuracy) and  $\frac{TPR}{FPR}$  in guiding the model selection procedure to a desired region of operation.

#### **10.1.8 Chapter 9: Applying the proposed SBCI with hand extension data**

In order for an SBCI system to be generalized to detect different types of IC states IC states, it should be able to detect new types of mental tasks. In this chapter, we tested and analyzed the performance of the SBCI proposed in Chapter 5 using data collected from performing hand movement extensions, instead of finger flexion, the movement the system was originally designed to detect. The findings of this research are:

- 1) A new artifact monitoring system was proposed. This system monitors the activities of eye movements, and those of the frontalis muscle movements. This monitoring system was specially built so that eye blinks and activities of frontalis muscles do not control the SBCI system.
- 2) We studied the performance of the proposed SBCI system over data related to hand extension movements. These data also contained more periods of engaging NC data compared to previous studies. Despite the presence of the latter data, the SBCI system still achieved a superior performance compared to those achieved by the traditional SBCI systems. The performances of three out of the five individuals who participated in the experiments, were especially good and much higher than the other two individuals.
- 3) We used the framework developed in Chapter 8 to determine which performance measure is more suitable for the model selection procedure. The Kappa coefficient was found to be as a suitable evaluation metric for summarizing the performance of this system.
- 4) We compared the performance of monopolar and bipolar montages and showed that the bipolar montage results in superior performances to those of the monopolar montage.

## **10.2 Summary of contributions**

The main contributions of this thesis fall under three main topics and are summarized as follows:

### **10.2.1 Reducing high false positive rates**

- 1) Introducing the idea of using features from three specific neurological phenomena (MRPs, CPMR and CPBR) to detect the possible presence of IC commands in an SBCI system.
- 2) Developing a new SBCI system that extracts and classifies features extracted from three neurological phenomena (MRPs, CPMR and CPBR) efficiently. For this purpose, a new two-stage multiple classifier system (MCS) as well as a

hybrid genetic algorithm (HGA) are proposed to effectively deal with the large dimensionality of the input feature space. It was shown that the proposed method can yield very low FP rates compared to other state-of-the-art EEG-based SBCI systems.

3) Proposing a new feature extraction method that consists of a stationary wavelet transform followed by a matched filter. This method was applied to significantly reduce the dimension of the wavelet feature space. Two new types of features, the maximum of the cross correlogram over a certain period and the time that this maximum has occurred are extracted. The contributions discussed above can also be used in order to improve the classification accuracy of synchronized BCI systems.

4) Using the linear programming theory, it was shown that an MCS can theoretically achieve high  $\frac{TPR}{FPR}$  values, even though the performance of individual classifiers is low.

5) We showed that the proposed algorithm can detect different types of movements and thus it has the potential to generalize to more IC commands. This was done by analyzing the performance of the proposed SBCI system on two datasets: one dataset related to the right finger flexion and the other dataset related to the right hand extension.

6) We demonstrated the superiority of using bipolar EEG channels to monopolar EEG channels. To do this, we compared the performances of the systems developed using monopolar and bipolar EEG signals during the detection of right hand extension movements.

7) We showed that automatic user-customization results in improvements in the performance of this SBCI system, especially in the case of individuals with spinal-cord injury. This was carried out by analyzing the effect of automatic user-customization in the performance of the LF-ASD SBCI system. Similar automatic customization can be applied to the feature extractor of other BCI system in order to further improve their performances.

### 10.2.2 Addressing artifacts in SBCI systems

- 1) We presented a detailed review of the methods that handle artifacts in BCI systems. We showed that the majority of the BCI papers do not report if they have considered artifacts during their analysis. We argued that since physiological artifacts occur frequently, these systems may face problems during the online implementations.
- 2) We investigated the performance of our proposed SBCI system in Chapter 5 over periods contaminated with eye-blink artifacts. We showed that although the performance of the system deteriorates slightly, the performance is still much higher than other EEG-based SBCI systems. The advantage of considering artifact-contaminated periods during the analysis is that the system is available for control during all times. This study highlights the importance of designing BCI systems that are robust to the presence of artifacts.
- 3) We investigated the performance of our proposed SBCI system in Chapter 5 in a pseudo – online fashion, where no part of the data has been rejected. In other words, the SBCI system was available for control at *all times*. This was done using the data from sessions recorded in later dates than the data used for training the system were recorded. This analysis is very close to online testing of the performance of the system. The only difference is that in our analysis, users do not receive any feedback on their actions.
- 4) We also implemented an artifact monitoring system that not only considers eye blink artifacts (similar to the one built in the previous studies of our group), it also considers the EMG artifacts related to frontalis muscles as well. This artifact monitoring system can be implemented for other BCI systems in order to ensure that artifacts such as EOG and EMG are not used to control the BCI system.

### 10.2.3 Finding a suitable evaluation metric for SBCI systems

- 1) Proposing a framework for comparing the evaluation metrics during the model selection process in SBCI systems. This framework can successfully find a suitable evaluation metric given a number of evaluation metrics. Furthermore, the



proposed framework can be applied to *any* classification problem with imbalanced datasets.

2) Applying the proposed framework to our proposed SBCI system and determining the Kappa coefficient as the most suitable evaluation metric.

### 10.3 Future research directions

There are a number of studies that can be followed as an extension to this thesis. We have summarized them in this section:

1) With the exception of Chapter 2, all the research in this thesis has been conducted using executed movements. While strong evidence from the literature supports the similarity between attempted movements and real movements, a study is needed to verify the performance of the proposed SBCI on attempted movements.

2) Additional experiments should be carried out to determine the sensitivity to the design steps and the parameter values used. As an example, an interesting research area would be to study the effect of different penalty functions in the design of an SBCI system (e.g., Eq 2.9 in Chapter 2).

3) Currently, not much research has been concluded to find out what the acceptable performance of a particular SBCI application is. This is an important research issue, as determining the pre-defined goals of an SBCI system helps the researchers to decide if the results are suitable or not. A study that determines such performance thresholds for different applications is thus of great value for SBCI researchers.

4) The studies reported in this thesis have all been conducted offline. Such analysis is necessary to evaluate the performance of the system prior to online tests. A study is therefore needed to examine the performance of the proposed SBCI in online experiments. For this purpose, a suitable test environment should be developed so that false positives do not cause user frustration. The evidence that we have provided in Chapter 7 indicates that the proposed design can be used in the near future in an online testing environment.

- 5) Although the performance of the system was tested using different types of NC data, further tests are also needed to further investigate the performance of the system during other types of NC periods. These tests may include periods when users are engaged in various tasks such as talking, daydreaming, performing different types of mental tasks, etc.
- 6) The designs presented in this thesis identified one type of IC command from an NC state. For future studies, we would like to give more control power to the users, so that they can send different types of IC commands. Thus the proposed SBCI should be able to identify different IC states (related to different types of movements) from each other as well as from NC states. We have already shown in this thesis that the SBCI system proposed in Chapter 5 can detect right index finger flexion and right hand extension movements. Future work should focus on distinguishing these IC commands from each other. Also, it is of interest to study the performance of the system on new types of movements.
- 7) In these studies we did not evaluate the performance of the system over a long time period. A study that analyzed the performance of the system over long periods (examples months) is thus of great interest. Since neurological phenomena may change over time, an adaptation algorithm may be needed to tune the parameter values of the system. Also is of interest is to study the effect of training the participants in improving the performance of the system.
- 8) In this thesis, we studied artifacts related to eye blinks and frontalis muscle movements. Future work should also analyze the performance of the system on other types of artifacts including eye rolling, saccades, chewing and swallowing as they can also affect the performance of the system.
- 9) The application of artifact-removal methods such as independent component analysis (ICA) in removing artifacts and possibly improving the performance of the system should also be investigated as a part of future studies.
- 10) The framework proposed in Chapter 8 can also be extended to multi-objective evaluation metrics. It would be of interest to find out if a multi-objective evaluation metric can outperform single objective evaluation metrics.

## 10.4 References

- [1] A. Bashashati, "Towards development of a 3-state self-paced brain computer interface (BCI)", Ph.D. thesis, 2007.
- [2] A. Bashashati, M. Fatourechi, R. K. Ward and G. E. Birch, "User Customization of the Feature Generator of an Asynchronous Brain Interface", *Annals of Biomedical Engineering*, vol. 1, pp. 1-1, 2006.2006.
- [3] A. Bashashati, S. Mason, R. K. Ward and G. E. Birch, "An improved asynchronous brain interface: making use of the temporal history of the LF-ASD feature vectors", *J. Neural Eng.*, vol. 3, no.2, pp. 87-94, Jun.2006.
- [4] A. Bashashati, R. K. Ward and G. E. Birch, "Towards development of a 3-state self-paced brain computer interface", *Computational Intelligence and Neuroscience*, In press.
- [5] J. Huang and C. X. Ling, "Using AUC and accuracy in evaluating learning algorithms", *IEEE Trans. Knowled. Data Eng.*, vol. 17, no.3, pp. 299-310, 2005.

## APPENDIX A- UBC RESEARCH ETHICS BOARD CERTIFICATE



The University of British Columbia  
Office of Research Services  
**Behavioural Research Ethics Board**  
Suite 102, 6190 Agronomy Road, Vancouver, B.C. V6T 1Z3

### CERTIFICATE OF APPROVAL- MINIMAL RISK RENEWAL

<b>PRINCIPAL INVESTIGATOR:</b> Gary E. Birch	<b>DEPARTMENT:</b> UBC/Applied Science/Electrical and Computer Engineering	<b>UBC BREB NUMBER:</b> H06-80316
<b>INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT:</b>		
<b>Institution</b> Vancouver Coastal Health (VCHRI/VCHA)		<b>Site</b> GF Strong Rehabilitation Centre
Other locations where the research will be conducted: N/A		
<b>CO-INVESTIGATOR(S):</b> Rabab K. Ward Mehrdad Fatourehchi		
<b>SPONSORING AGENCIES:</b> Canadian Institutes of Health Research (CIHR) - "Single-Trial Detection of Motor-Related Potentials for Use in a Direct Brain-Computer Interface for Persons with Severe Disabilities" Natural Sciences and Engineering Research Council of Canada (NSERC) - "Single-Trial Detection of Motor-Related Potentials for Use in a Direct Brain-Computer Interface for Persons with Severe Disabilities"		
<b>PROJECT TITLE:</b> Single-Trial Detection of Motor-Related Potentials for Use in a Direct Brain-Computer Interface for Persons with Severe Disabilities		

**EXPIRY DATE OF THIS APPROVAL:** April 15, 2008

**APPROVAL DATE:** April 15, 2007

The Annual Renewal for Study have been reviewed and the procedures were found to be acceptable on ethical grounds for research involving human subjects.

Approval is issued on behalf of the Behavioural Research Ethics Board  
and signed electronically by one of the following:

Dr. Peter Suedfeld, Chair  
Dr. Jim Rupert, Associate Chair  
Dr. Arminee Kazanjian, Associate Chair  
Dr. M. Judith Lynam, Associate Chair  
Dr. Laurie Ford, Associate Chair

## APPENDIX B- THEORETICAL ANALYSIS OF THE PROPOSED SBCI

The majority of studies that theoretically analyzed the combinations of classifiers have made some assumption about the independency of classifiers (for more details, see [1-3]). In this Appendix, the performance of the proposed SBCI is analyzed theoretically using the framework developed in [3, 4]. This theoretical framework applies linear programming to determine the lower and higher bounds of performance of an MCS, but it does not make any assumption about the independence of classifiers. For simplicity, we focus on the upper and lower bounds of the fitness function formulated as a  $\frac{TPR}{FPR}$  ratio. To obtain these bounds, the maximum and minimum of the TP and FP rates of the two-stage MCS are determined by linear programming.

### B.1. Formulating the problem

Let *2ndStageMCS* denote the second-stage MCS that combines the outputs of the MCSs in the first stage. The first-stage MCSs are denoted by  $\{MCS_1, MCS_2, \dots, MCS_K\}$ , where  $K$  is the number of MCSs in the first stage (in the proposed SBCI,  $M$  is 2 or 3). To calculate the maximum and minimum of the  $\frac{TPR}{FPR}$ , we represent the classification labels generated by all  $K$  classifiers with a binary string. Let  $bit(j, K)$  be such a bit string that denotes the  $K$  bit binary expansion of  $j$ . Each classifier is represented by a bit in  $bit(j, K)$ . A value of “0” indicates that the classifier didn’t correctly classify an IC command (a FN) and a value of “1” indicates that the classifier correctly identified an IC command (a TP). We use the convention that if there are  $K$  classifiers  $MCS_1, MCS_2, \dots, MCS_K$ ,  $MCS_1$  is the least significant bit (LSB) and  $MCS_K$  is the most significant bit (MSB). Let  $x = [x_0, x_1, \dots, x_{(2^K-1)}]^T$  be the vector of joint probabilities of the correct detection of an IC

command. Since for  $K$  classifiers, there are  $2^K$  possible combination of correct/incorrect classifiers, vector  $x$  will be of length  $2^K$ . This combination of the classifier can be shown using a Venn diagram, as in Figure B.1. In this figure,  $x_0$  shows the percentage of IC commands that all MCSs failed to correctly identify, and  $x_i$  shows the percentage of IC commands that the first classifier ( $MCS_1$ ) correctly identified but that the rest of the classifiers ( $MCS_2$  and  $MCS_3$ ) could not correctly identify, and so on.

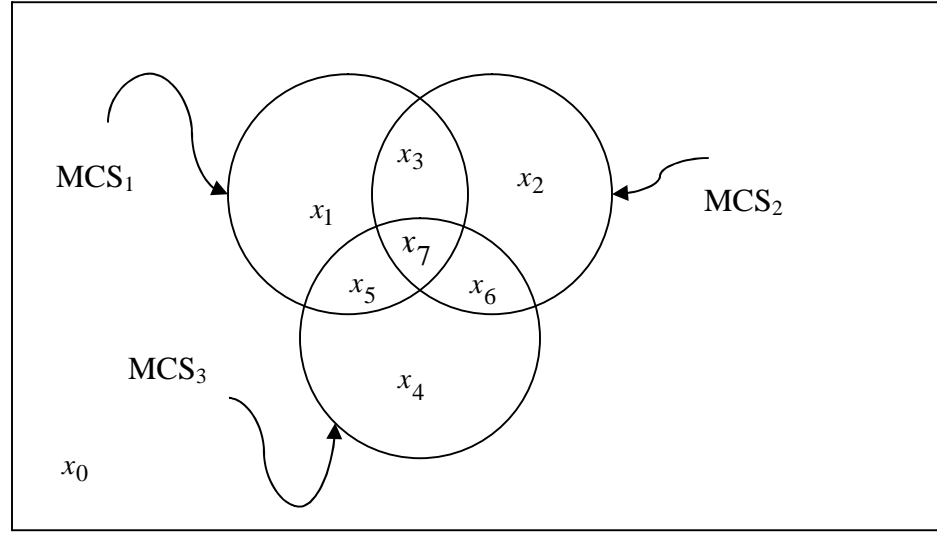


Figure B1. The Venn diagram for three MCSs.

## B.2. Constraints

Let  $TP_{SBCI}(\mathbf{x})$  represent the probability of the correct classification of the IC commands in the proposed SBCI. We wish to find the values of  $\mathbf{x}$  that yield the maximum and the minimum of  $TP_{SBCI}(\mathbf{x})$ . The constraints of this optimization problem are as follows:

1. The values  $x_i$  are non-negative and are smaller than 1:

$$0 \leq x_i \leq 1, \quad i = 0, 1, \dots, (2^K - 1) \quad (\text{B.1})$$

2. The sum of the joint probabilities is 1:

$$\sum_{i=0}^{2^K-1} x_i = 1 \quad (\text{B.2})$$

3. The sum of the joint probabilities for which classifier  $r$  could correctly identify an IC command must equal  $p_r$ , the normalized TP rate of the classifier  $r$ . Or, mathematically,

$$A_{eq} \mathbf{x} = \mathbf{d} \quad (\text{B.3})$$

where  $\mathbf{d}$  is the vector of the normalized TP rates of the classifier, as represented below:

$$\mathbf{d} = [p_1, p_2, \dots, p_K]^T \quad (\text{B.4})$$

$$p_j = \frac{TP_j}{N_{IC}} \quad j = 1, 2, \dots, K \quad (\text{B.5})$$

$N_{IC}$  is the number of IC commands and  $TP_j$  is the TP rate of the  $j^{\text{th}}$  classifier, and  $A_{eq}$  is a  $K \times 2^K$  matrix, whose  $r^{\text{th}}$  row corresponds to the  $r^{\text{th}}$  classifier.  $A_{eq}$  is defined as follows:

$$A_{eq} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_K]^T \quad (\text{B.6})$$

where  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_K$  are bit strings of length  $2^K$  and can be calculated as follows:

$$\begin{aligned} \mathbf{b}_1 &= [01, \dots, 01]^T \\ \mathbf{b}_2 &= [0011, \dots, 0011]^T \\ &\vdots \\ \mathbf{b}_K &= [\underbrace{00 \dots 00}_{2^{(K-1)}}, \underbrace{11 \dots 11}_{2^{(K-1)}}]^T \end{aligned} \quad (\text{B.7})$$

### B.3. Objective functions

Let  $f_{TP(SBI)}(i)$  denote the entry at  $i^{\text{th}}$  position in  $TP_{SBI}(\mathbf{x})$ . We can then define the following fitness function for the two-stage MCS (Configurations 1-3 and 5 in Chapter 5) as follows.

$$f_{TP(SBCI)}(i) = \begin{cases} 1, & \text{if } N_1 = K \\ 0, & \text{if } N_1 = 0 \\ P_{IC}^{N_0}, & \text{otherwise} \end{cases} \quad (\text{B.8})$$

where  $N_I$  is the number of ones in  $bit(i,K)$  and  $N_0$  is the number of zeros in  $bit(i,K)$  and  $p_{IC}$  is the probability of the IC state, calculated as follows:

$$P_{IC} = \frac{N_{IC}}{N_{Total}} \quad (B.9)$$

where  $N_{IC}$  is the number of IC epochs and  $N_{Total}$  is the total number of epochs.  $N_{Total}$  is calculated as follows:

$$N_{Total} = N_{IC} + N_{NC} \quad (B.10)$$

Eq. (B.8) implies that only when all the classifiers participating in the two-stage MCS correctly identify an IC command, the output of the two-stage MCS will be “1”. If all of them fail to recognize an IC command, the output is zero. In other cases, the decision is made based on the probability of the IC state. It can be seen that as  $N_0$  increases, the SBCI has a higher probability of generating an FN. When  $p_{IC}$  is sufficiently (e.g.,  $p_{IC} < .01$ ), and  $N_0$  is sufficiently large,  $p_{IC}^{N_0} \approx 0$ , and  $f_{TP(SBCI)}$  will take a form of an AND operator, as described in Chapter 5, For Configuration 4 in Chapter 5, (B.8) becomes

$$f_{TP(SBCI)}(i) = \begin{cases} 1, & \text{if no. of 1s in } bit(i,K) > \frac{K}{2} \\ 0, & \text{Otherwise} \end{cases} \quad (B.11)$$

Similarly, by applying the following replacements in Eqs. (B.1)-(B.11), the optimization problem for the NC state can be formulated:

$$\begin{cases} x \rightarrow y \\ P_{IC} \rightarrow P_{NC} \\ TP \rightarrow TN \\ N_{IC} \rightarrow N_{NC} \\ N_{NC} \rightarrow N_{IC} \end{cases} \quad (B.12)$$

As can be seen, the formulation is the same as that for the TP rate. The main difference is in the formulation of (B.8). Here, because of the high probability of  $p_{NC}$  (e.g.,  $p_{NC} > 0.99$ ), the number of elements closer to “1” grows compared to the optimization problem for the TP rate. Similar formulations for each of the MCSs in the



first stage can be developed. The only difference will be in the case when the number of classifiers is even. In this case, the function  $f$  for the case of TP is formulated as

$$f_{even}(i) = \begin{cases} 1, & \text{if } N_1 > \frac{K}{2} \\ P_{IC}^{N_0}, & \text{if } N_0 = \frac{K}{2} \\ 0, & \text{otherwise} \end{cases} \quad (\text{B.13})$$

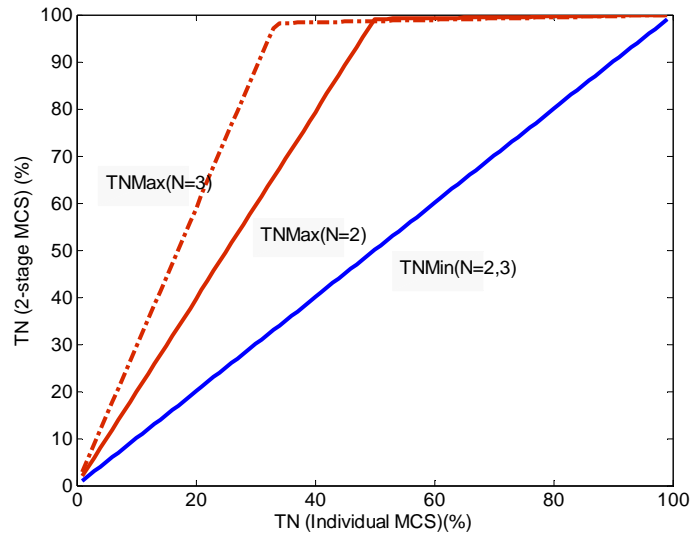
In the case of a tie and for a relatively large  $N_0$ , the probability of correct identification of an IC command will be close to zero. The opposite case is true for NC trials.

## B.4. Results

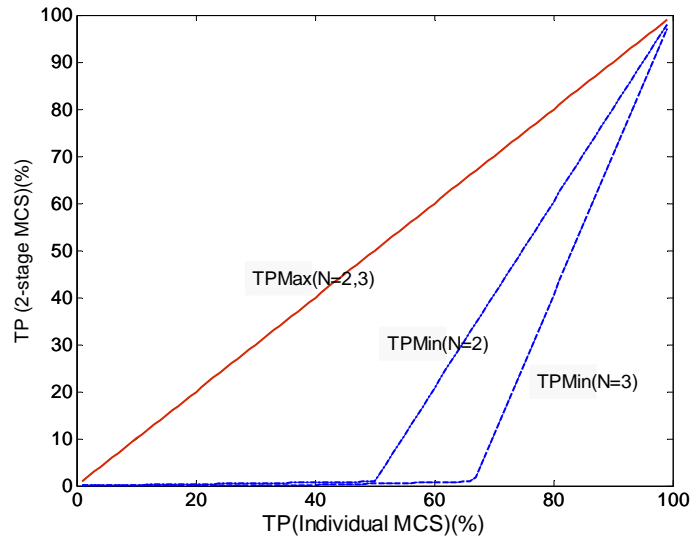
Figure B.2 depicts the  $TN_{Max}$  and  $TN_{Min}$  values of the SBCI as functions of the  $TN$  rates of the individual MCSs in the first stage (for simplicity, it is assumed that all individual MCSs have the same  $TN$  rates). The optimal values are found by maximizing and minimizing the TP and TN values using linear programming. The value of  $P_{NC}$  is estimated from the experimental protocol (described in Chapter 5) to be  $P_{NC} \approx 0.99$ . Figure B2 (a) shows that even for MCSs with high FP rates (e.g.,  $20\% < FP < 50\%$ ), it is theoretically possible that the proposed SBCI will achieve low FP rates. For  $FP < 10\%$ , the FP rate can theoretically approach zero.

Figure B2(b) shows the  $TP_{Max}$  and  $TP_{Min}$  values. It is assumed that all individual MCSs have the same TP rate. As Figure B3(a) shows, the theoretically low FP rate of the SBCI comes at the expense of a lower TP rate. Assuming all MCSs have the same TP and FP rates, the  $\frac{TPR}{FPR}$  of the SBCI can be demonstrated graphically. Figure B4 shows the  $\frac{TPR}{FPR}$  of an individual MCS. Figures B3(b) and B3(c) show the  $\min(\frac{TPR}{FPR})$  and  $\max(\frac{TPR}{FPR})$  of the SBCI, respectively for two MCSs in the first stage. Similar figures were obtained for three MCSs in the first stage. As Figure B3(c) shows the proposed

SBCI can theoretically have a much higher  $\frac{TPR}{FPR}$  than that of an individual MCS ( Figure B3(a)).

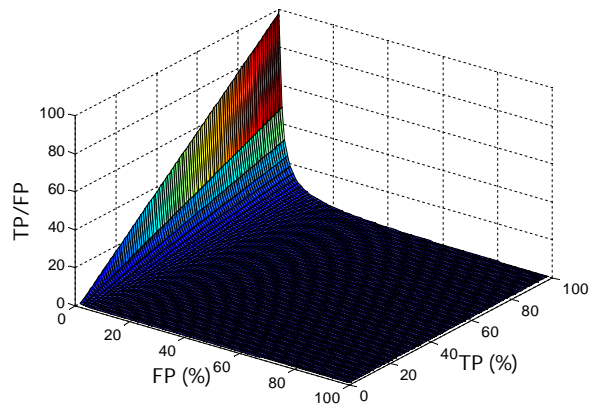


(a)

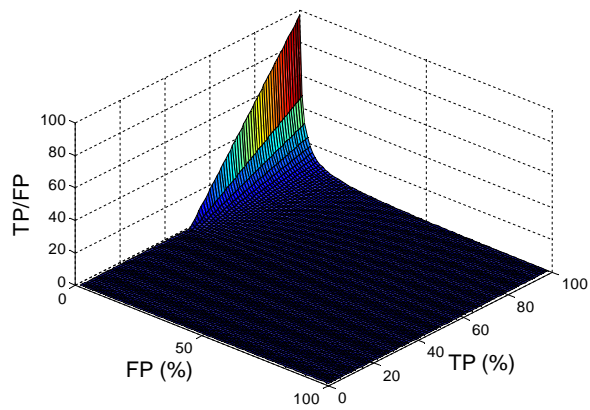


(b)

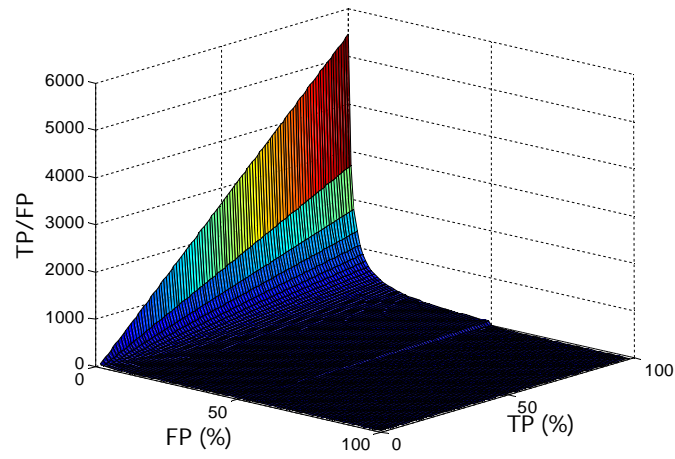
**Figure B2.** (a)  $TN_{Max}$  and  $TN_{Min}$  for two and three MCSs; (b)  $TP_{Max}$  and  $TP_{Min}$  for two and three MCSs.



(a)



(b)



(c )

Figure B3. (a)  $\frac{TPR}{FPR}$  for an individual MCS ; (b)  $\text{Min}(\frac{TPR}{FPR})$  of the two-stage MCS (for two MCSs in the first stage); (c)  $\text{Max}(\frac{TPR}{FPR})$  of the two-stage MCS (for two MCSs in the first stage).

## B.5. References

- [1] J. Kittler, M. Hatef, R. P. W. Duin and J. Matas, "On combining classifiers", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no.3, pp. 226-239, 1998.
- [2] L. Lam and S. Y. Suen, "Application of majority voting to pattern recognition: an analysis of its behavior and performance", *Systems, Man and Cybernetics, Part A, IEEE Transactions on*, vol. 27, no.5, pp. 553-568, 1997.
- [3] A. Narasimhamurthy, "Theoretical Bounds of Majority Voting Performance for a Binary Classification Problem", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no.12, pp. 1988-1995, 2005.
- [4] M. Demrekler and H. Altincay, "Plurality voting-based multiple classifier systems: statistically independent with respect to dependent classifier sets", *Pattern Recognit*, vol. 35, no.11, pp. 2365-2379, 2002.