Optimization algorithms for brain computer interfaces

Roberto Santana

Cajal Blue Brain Project
Computational Intelligence Group
Technical University of Madrid

February 24, 2011

Index

1. **Brain Computer Interfaces**
   - Review
   - Interfaces
2. **Machine learning in BCIs**
   - Problems
   - Classification methods
   - Advanced machine learning methods
3. **Optimization applied to BCI**
   - Ensemble approach to P300 speller
   - Multi-objective MEG channel selection

Outline

1. **Brain Computer Interfaces**
   - Review
   - Interfaces
2. **Machine learning in BCIs**
3. **Optimization applied to BCI**

BCI review

What a brain computer interface is?

- Communication system that translates brain activity into commands for a computer or other devices.
- Allows users interaction with their environment by using only brain activity, without using peripheral nerves and muscles.

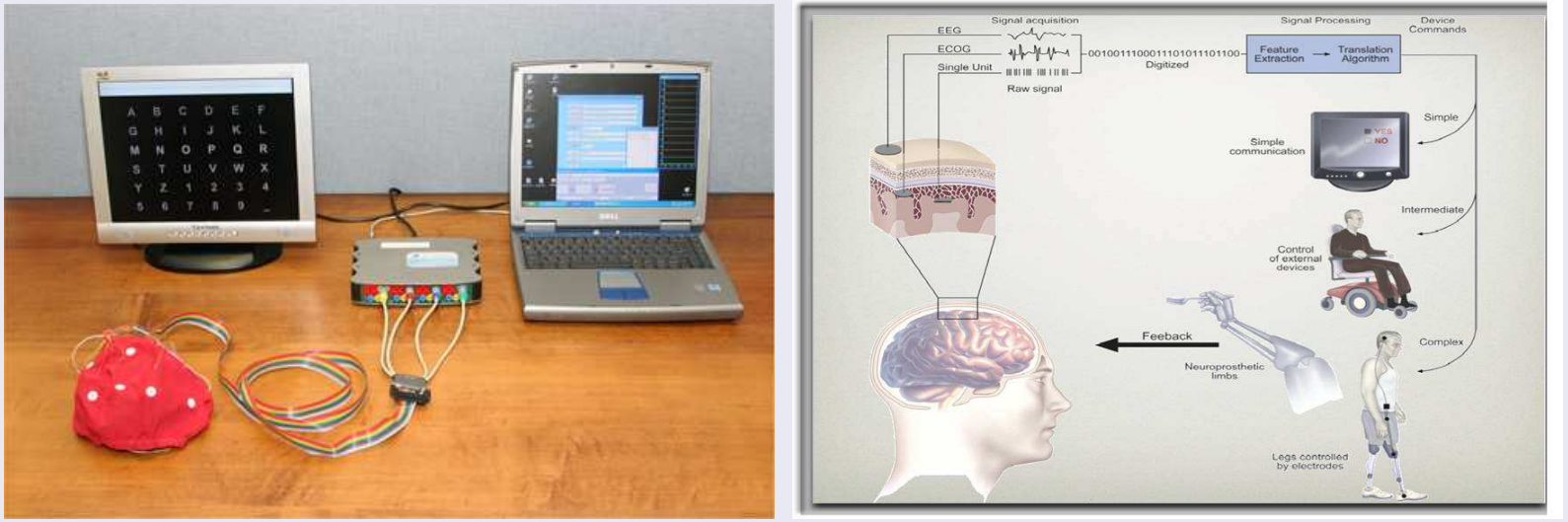
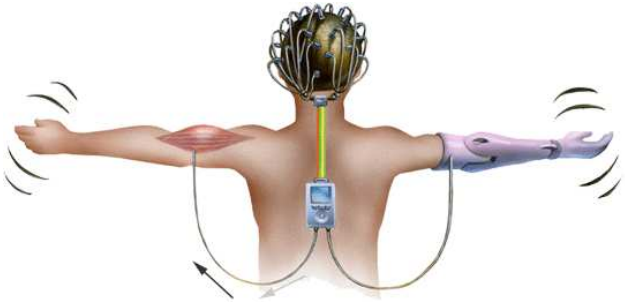


Figure: Left) Components of a brain computer interface. Right) BCI diagram



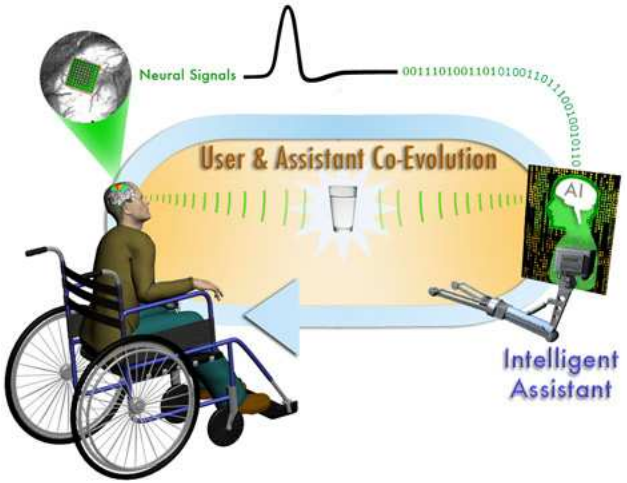


Figure: Different brain computer interfaces.

BCI applications

Applications

- Neuroprosthetic development.
- Assistive tools for disabled users.
- General man-machine communication.
- Control games or virtual environments (Lalor et al.:2005).
- Control driving and aircraft simulators (Bayliss and Ballard:2000).
- Moving maps in two dimensions (Trajo et al.: 2006).
- Evolving images in real real time (Bigdely and Makeig:2009).
- Space applications (Rossini et al: 2009)

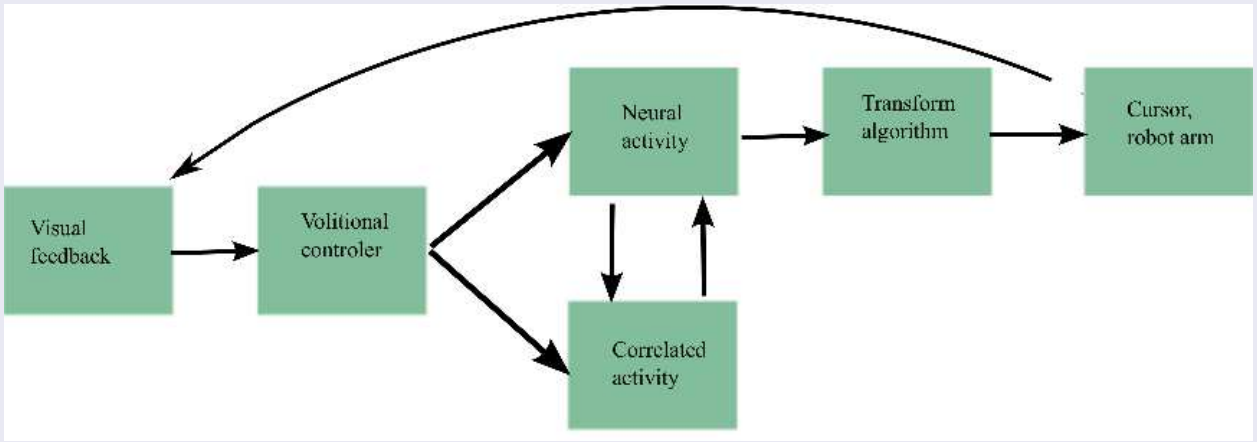


Figure: Basic BCI/BMI paradigm.

Outline

1 **Brain Computer Interfaces**
- Review
- Interfaces

2 **Machine learning in BCIs**
- Problems
- Classification methods
- Advanced machine learning methods

3 **Optimization applied to BCI**
- Ensemble approach to P300 speller
- Multi-objective MEG channel selection

BCI review

BCI classifications

- According to BCI paradigm.
- Offline vs Online analysis.
- Invasive vs Non-invasive.
- Evoked vs Spontaneous input.
- Type of signal: Time vs Frequency domain.
- Asynchronous vs Synchronous communication.
- According to device type.

BCI review

BCI classifications

- According to BCI paradigm.
- Offline vs Online analysis.
- Invasive vs Non-invasive.
- Evoked vs Spontaneous input.
- Type of signal: Time vs Frequency domain.
- Asynchronous vs Synchronous communication.
- According to device type.

BCI review

BCI paradigms

- P300 evoked potentials.
- Mu and beta rhythms and other activity from sensorimotor cortex.
- Visual evoked potentials.
- Slow cortical potentials.
- Neuronal ensemble activity.

BCI review

BCI paradigms

- P300 evoked potentials.
- Mu and beta rhythms and other activity from sensorimotor cortex.
- Visual evoked potentials.
- Slow cortical potentials.
- Neuronal ensemble activity.

BCI review

P300

- P300 is a positive deflection of EEG appearing approximately 300 ms after the presentation of rare or surprising, task-relevant stimuli.

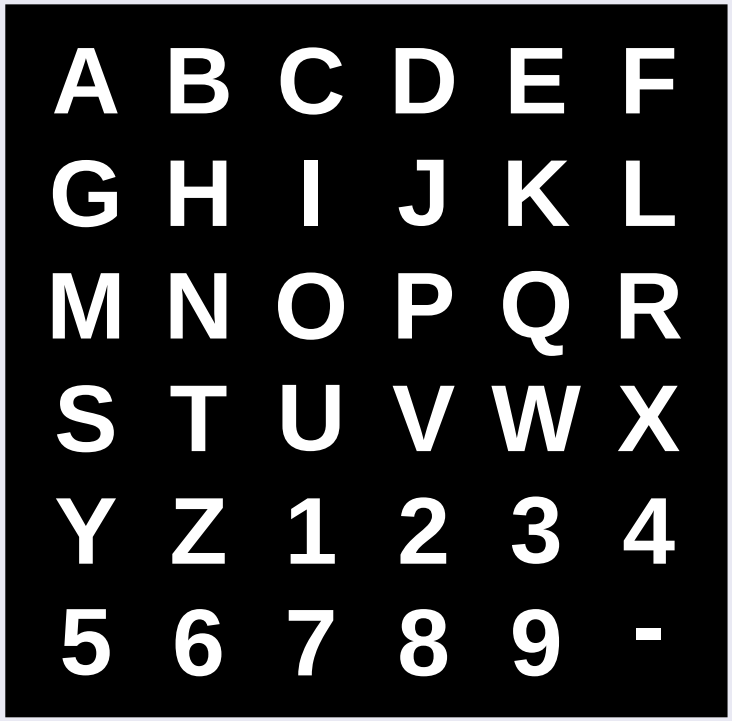


Figure: Schematic representation of the P300-speller as seen by the user on the display.

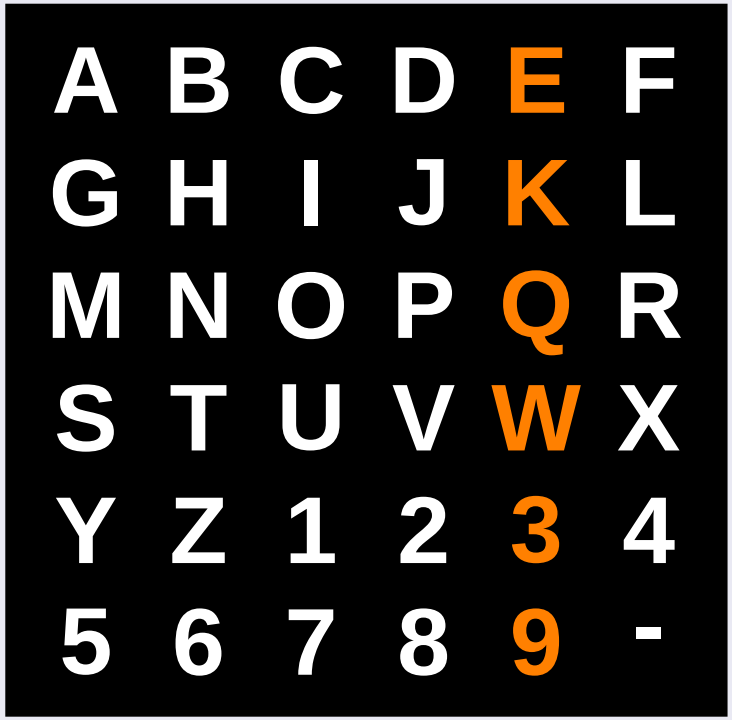


Figure: Target letter: N.

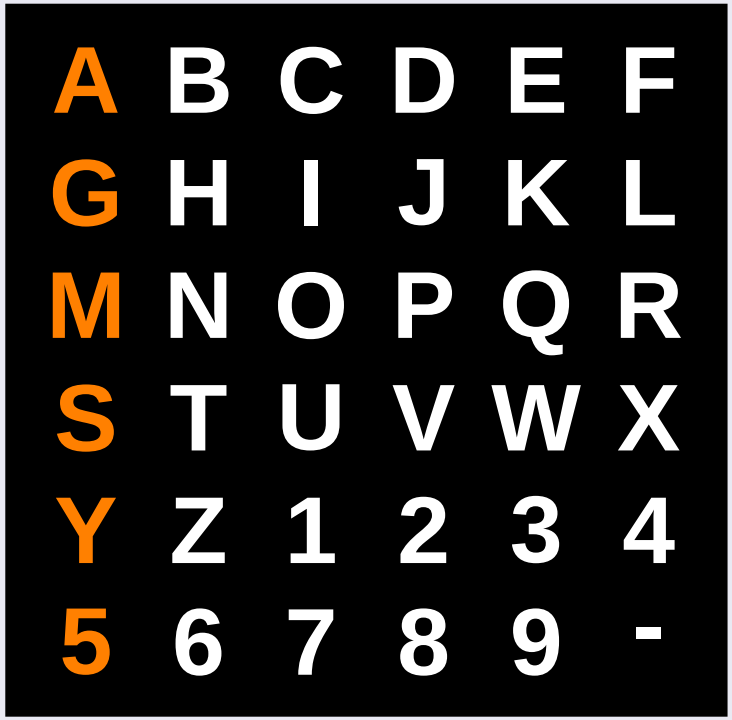


Figure: Target letter: N.

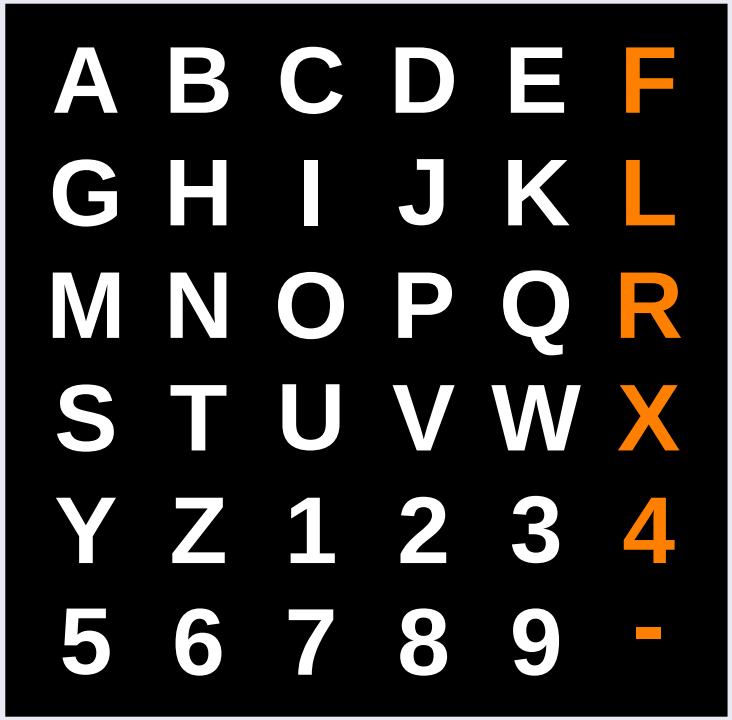


Figure: Target letter: N.

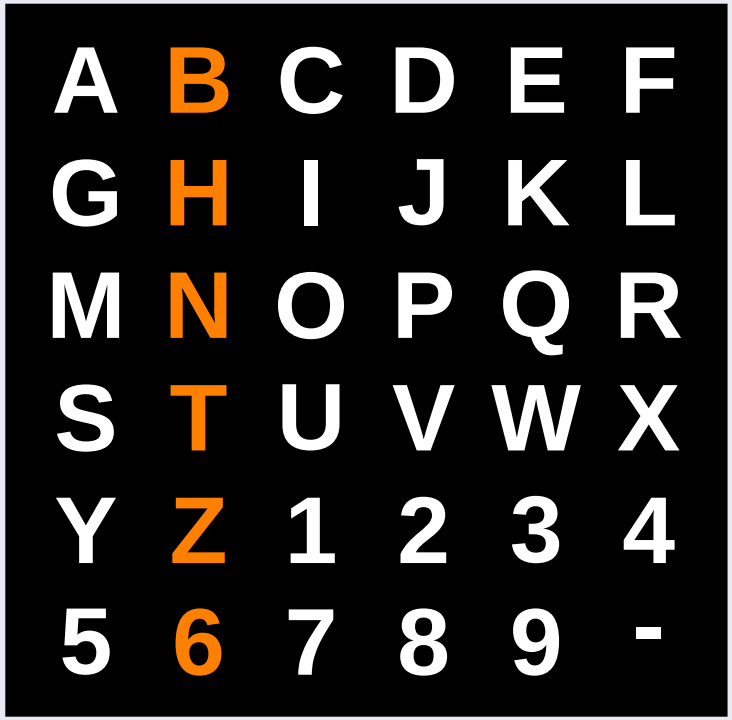


Figure: Target letter: N.

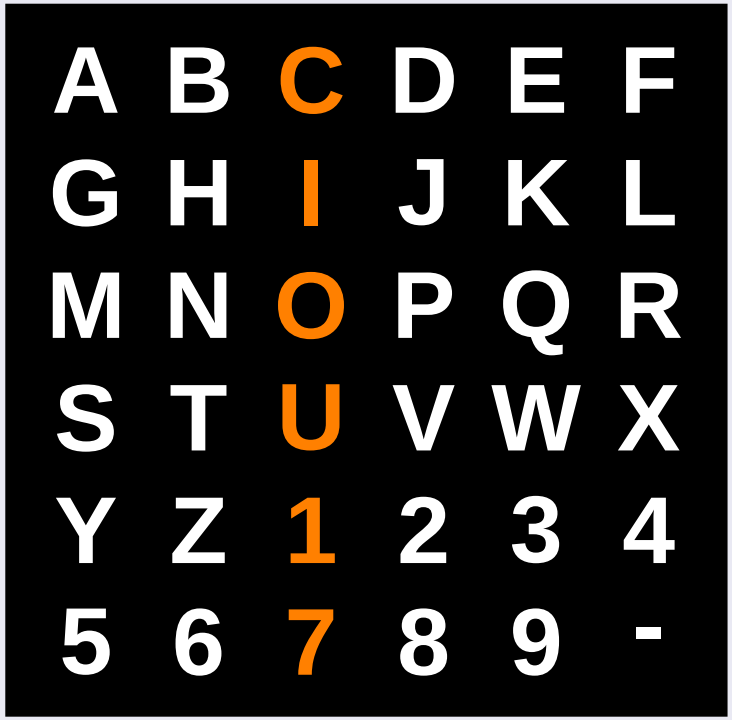


Figure: Target letter: N.

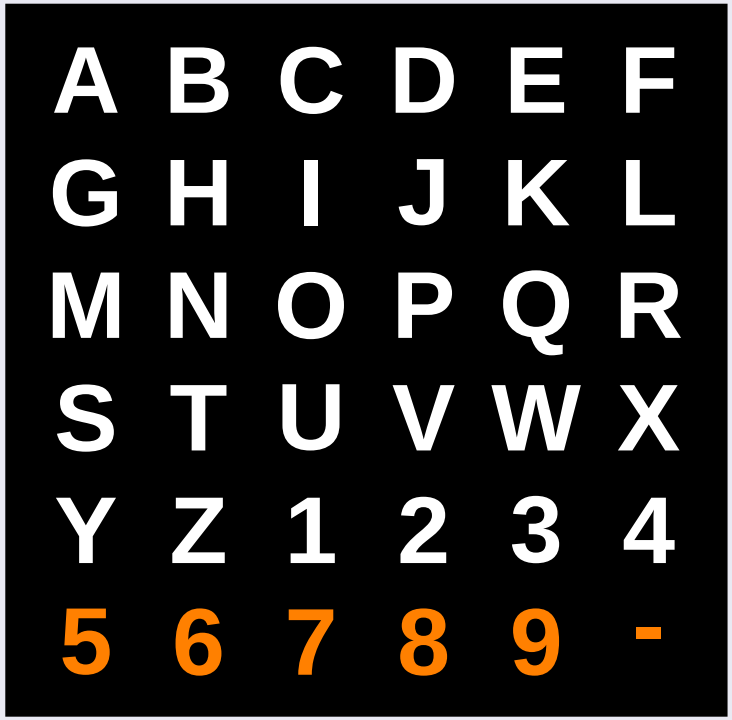


Figure: Target letter: N.

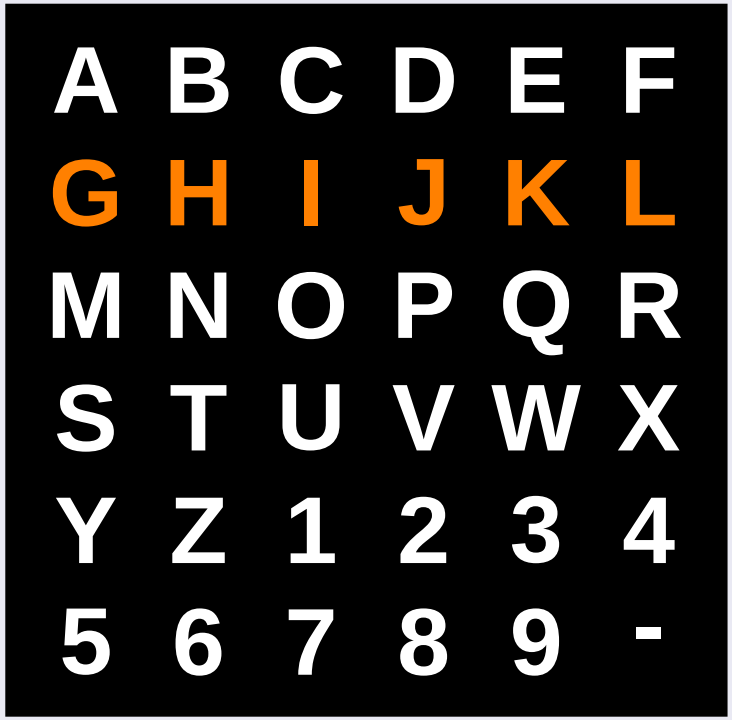


Figure: Target letter: N.

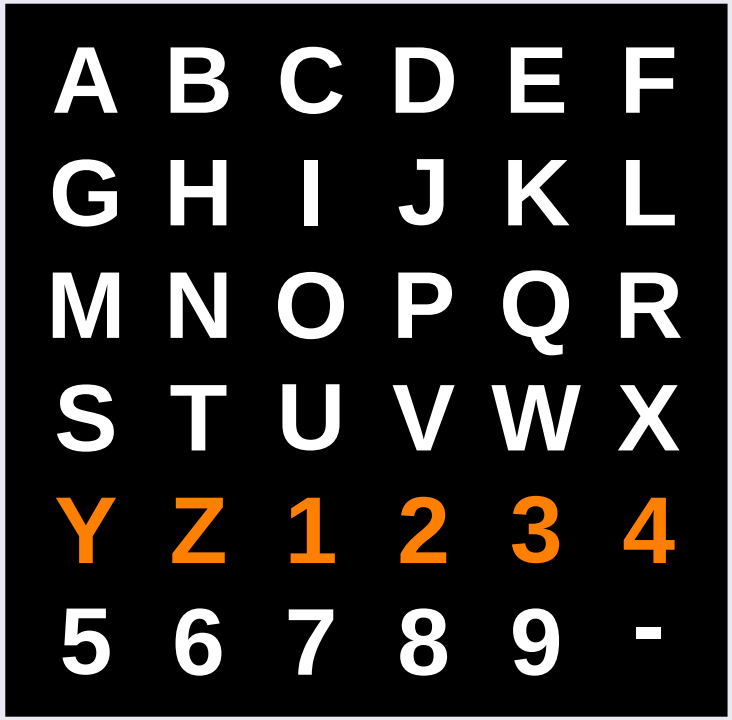
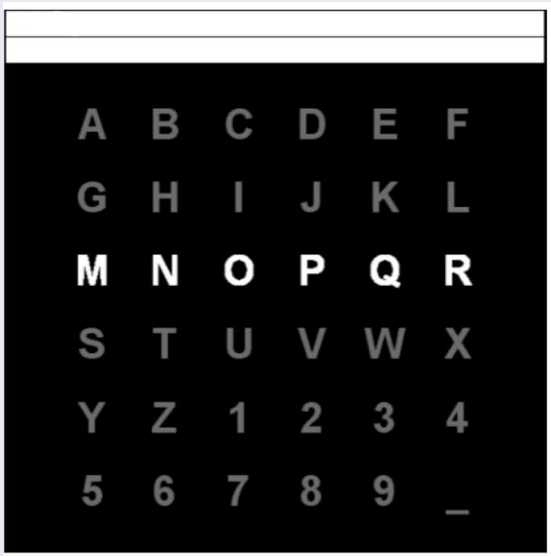


Figure: Target letter: N.

Figure: Target letter: N.



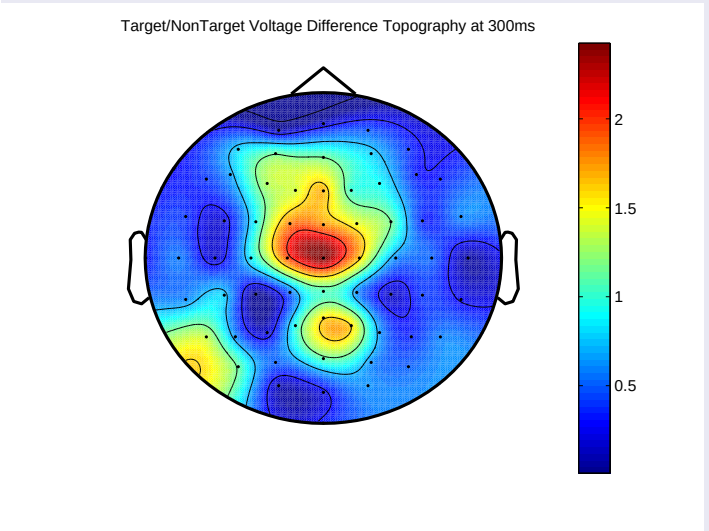


Figure: P300 speller and P300 response.

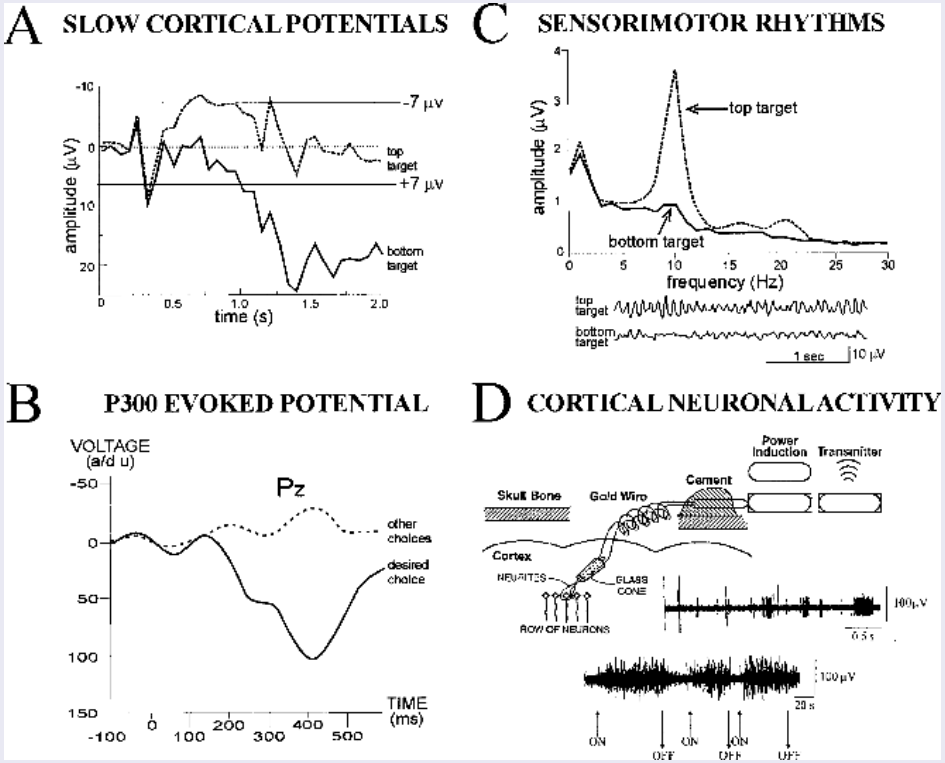


Figure: Different BCI paradigms (Wolpaw et al. 2002).

P300 speller

Common experimental framework

- Training phase: A classifier receives a training set of brain data recorded from a set of EEG channels during the identification of a set target characters which are also passed to the classifier.
- Test phase: New brain recordings are passed to the classifier. Based on previous training, the task is the prediction of the target characters.

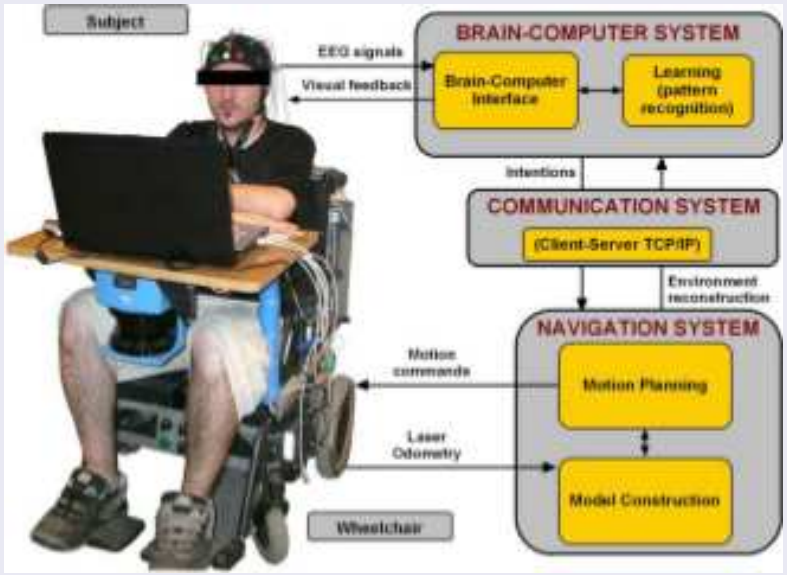


Figure: Brain-actuated wheelchair using P300 (Iturrate et al: 2008)

BCI review

BCI paradigms

- P300 evoked potentials.
- Mu and beta rhythms and other activity from sensorimotor cortex.
- Visual evoked potentials.
- Slow cortical potentials.
- Neuronal ensemble activity.

BCI review

Rhythms

- delta rhythm (1-4 Hz)
- theta rhythm (4-8 Hz)
- alpha and mu rhythms (8-13 Hz)
- beta rhythm (13-25 Hz)
- gamma rhythm (25-40 Hz)

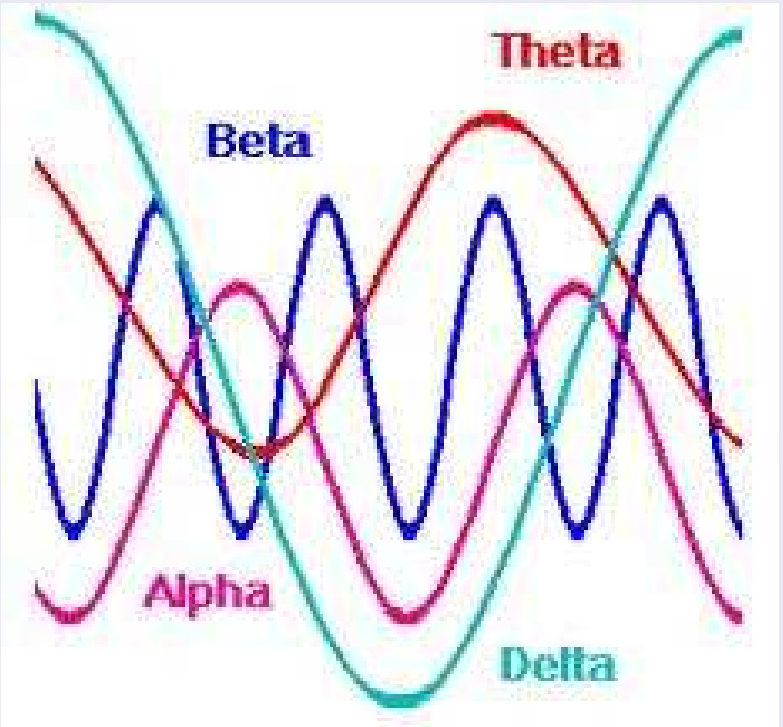


Figure: Brain rhythms.

BCI review

Oscillatory brain activity and BCI: Examples

- Mu-oscilations are decreased in amplitude when movements of body parts are imagined or performed.
- Changes in mu-rhythm are localized over the part of the sensorimotor cortex corresponding to the body part.
- Alpha rythms show high power over parietal and occipital areas in the absence of visual stimulation, which is then reduced in anticipation of visual targets.
- The topography of alpha also rhythms becomes spatially selective.

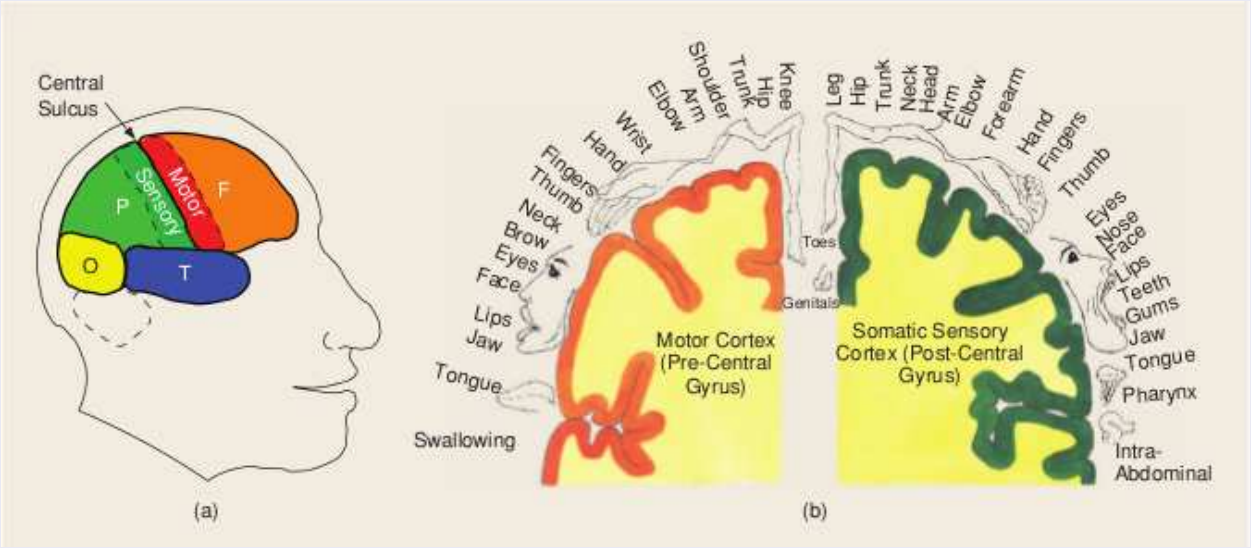


Figure: Geometric mapping between body-parts and motor/somatosensory cortex (Blankertz et al. 2008).

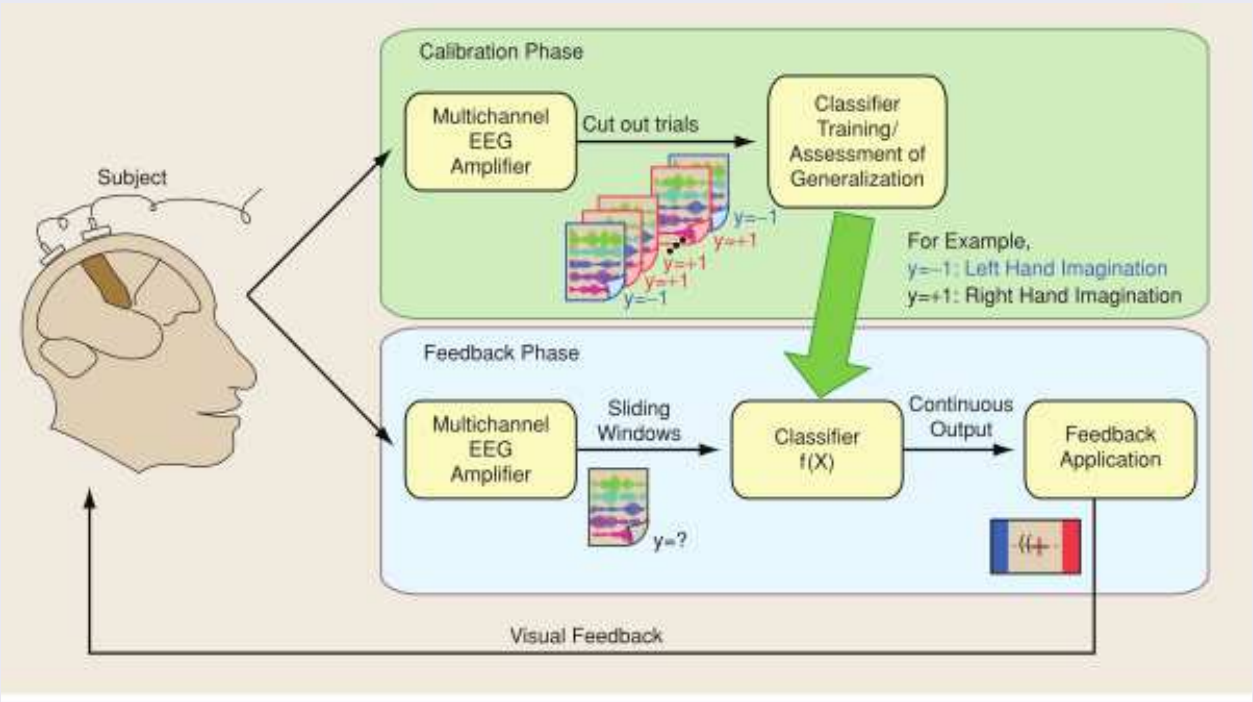


Figure: Machine learning based BCI system (Blankertz et al. 2008).

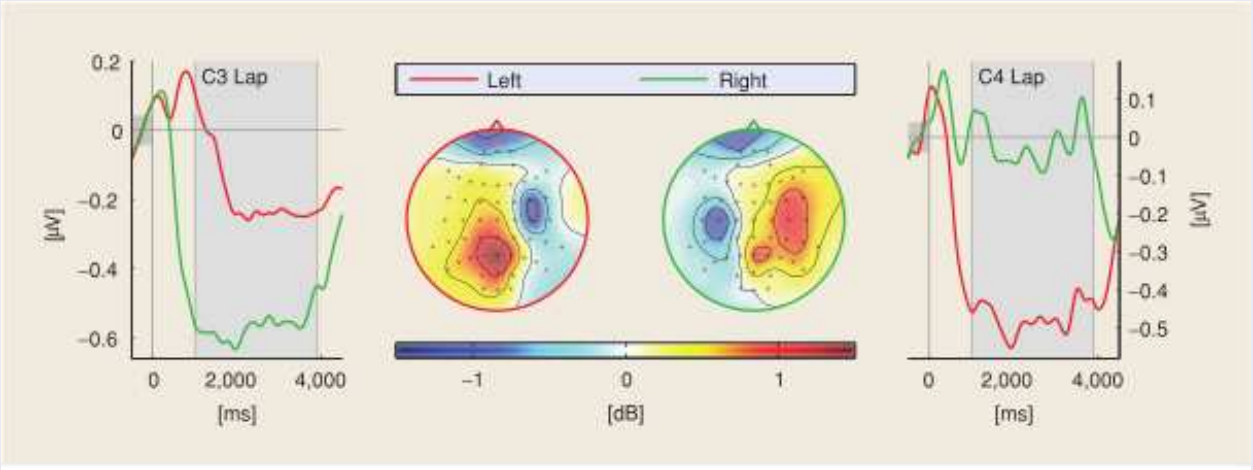


Figure: Motor imagery. Event related desynchronization (Blankertz et al. 2008).

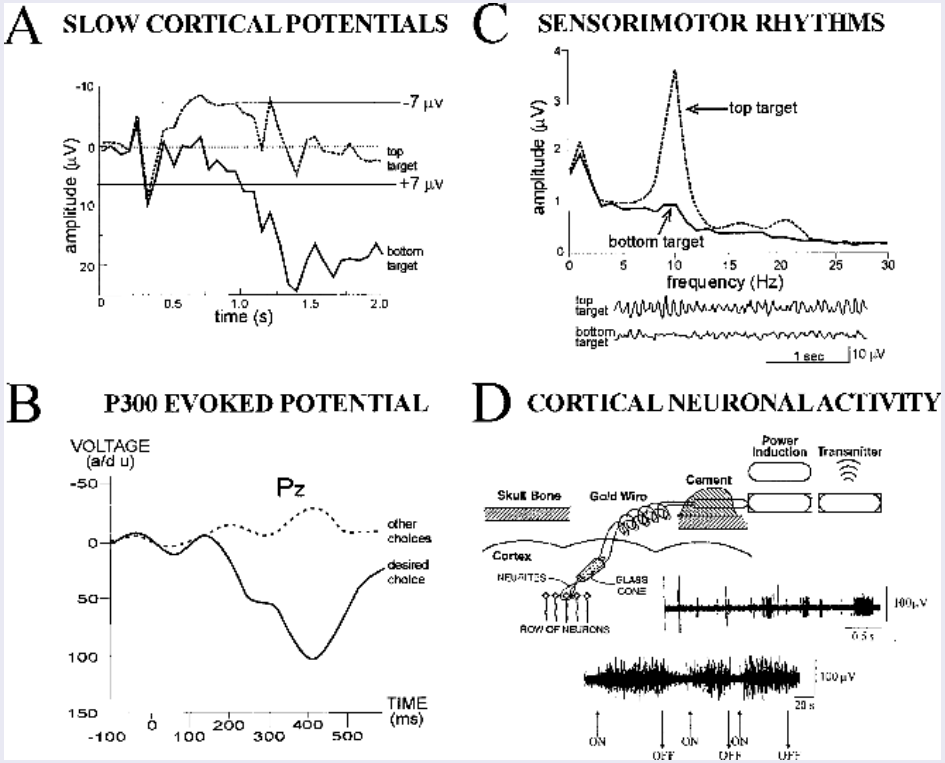


Figure: Different BCI paradigms (Wolpaw et al. 2002).

BCI review

BCI classifications

- According to BCI paradigm.
- Offline vs Online analysis.
- Invasive vs Non-invasive.
- Evoked vs Spontaneous input.
- Type of signal: Time vs Frequency domain.
- Asynchronous vs Synchronous communication.
- According to application.
- According to device type.

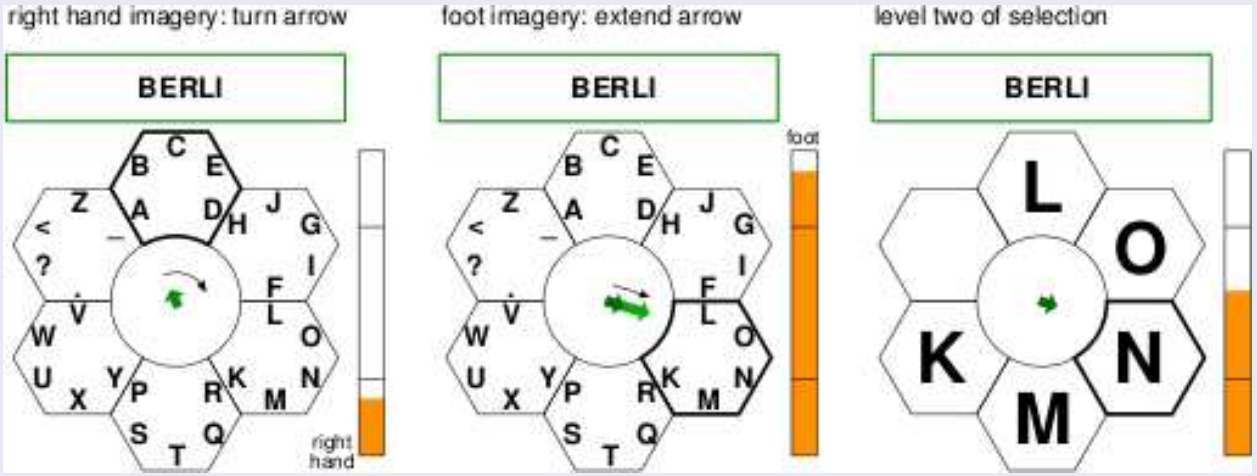


Figure: Motor-imagery-based speller (Muehler et al. 2008).

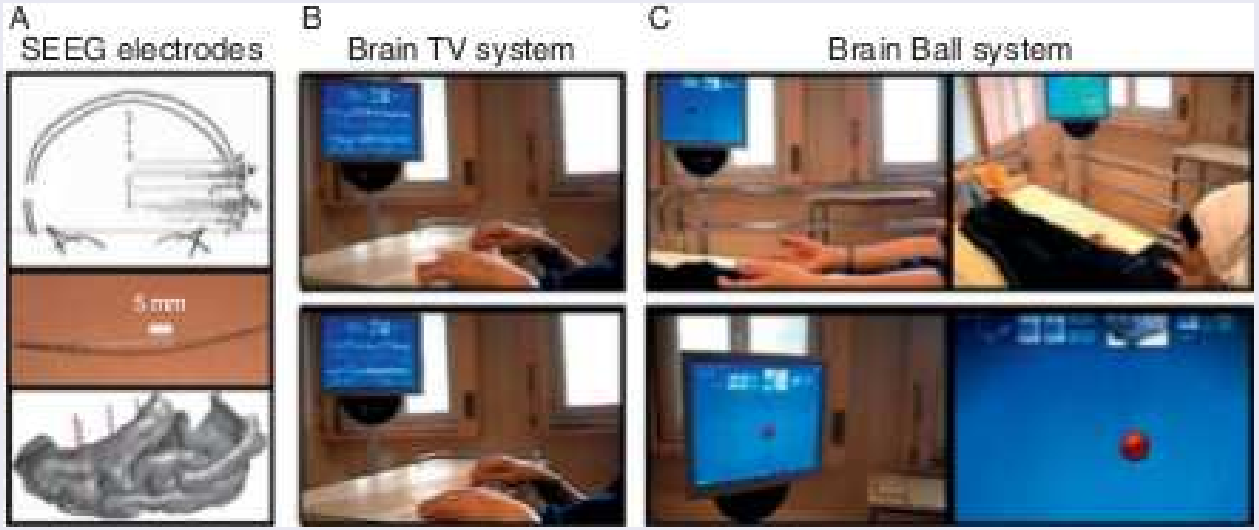


Figure: Brain TV setup (Jerbi: 2007).

BCI review

BCI classifications

- According to BCI paradigm.
- Offline vs Online analysis.
- Invasive vs Non-invasive.
- Evoked vs Spontaneous input.
- Type of signal: Time vs Frequency domain.
- Asynchronous vs Synchronous communication.
- According to application.
- According to device type.

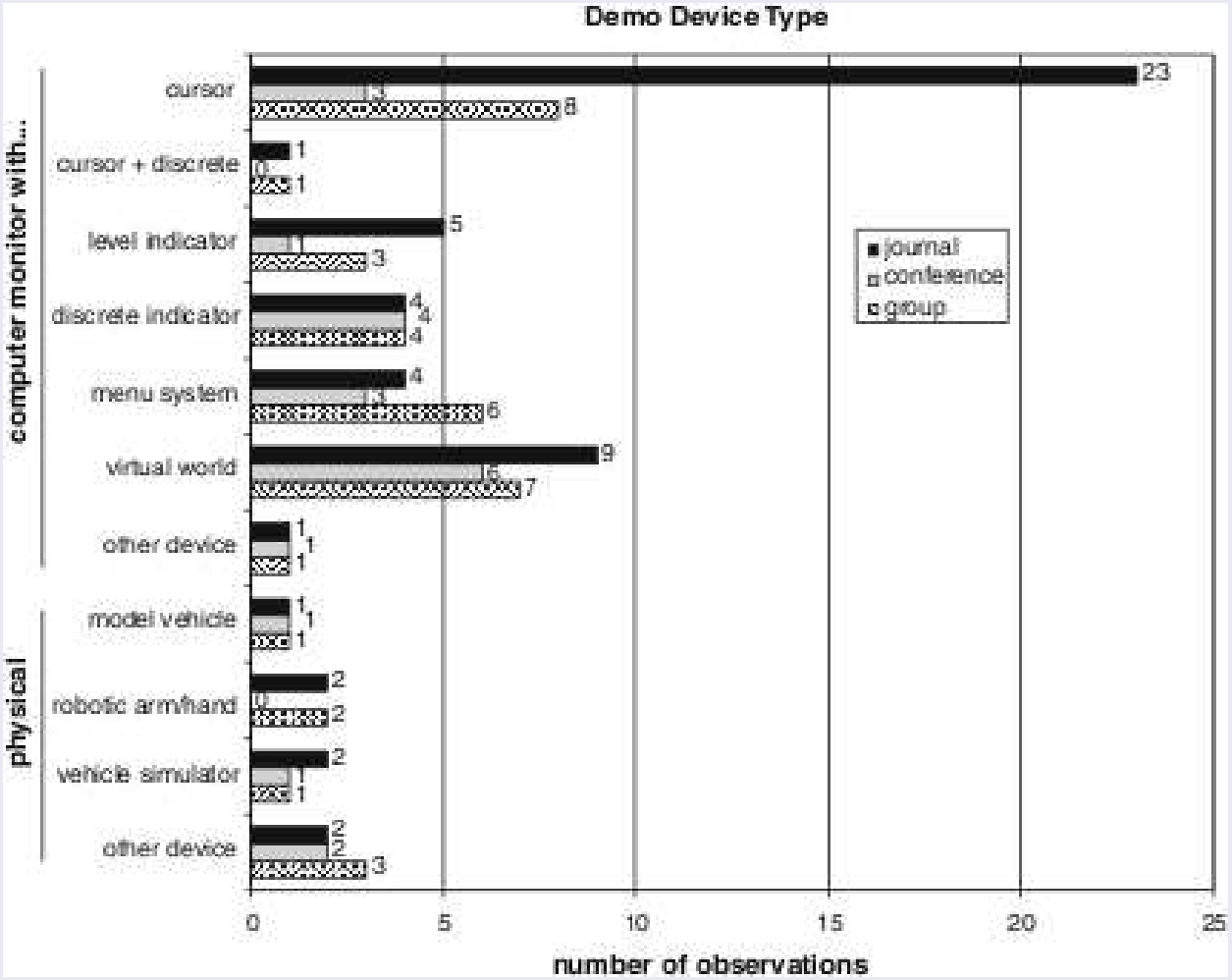


Figure: Distribution of demo device type (Mason et al: 2007).

Outline

1 **Brain Computer Interfaces**
- Review
- Interfaces

2 **Machine learning in BCIs**
- Problems
- Classification methods
- Advanced machine learning methods

3 **Optimization applied to BCI**
- Ensemble approach to P300 speller
- Multi-objective MEG channel selection

BCI data acquisition

BCI recording

- Electroencelographic data (EEG).
- Electrocorticography (ECog)
- Neuronal recording from implanted microelectrodes.
- Near-infrared functional brain imaging (NIR)
- Magnetoencelographic data (MEG).

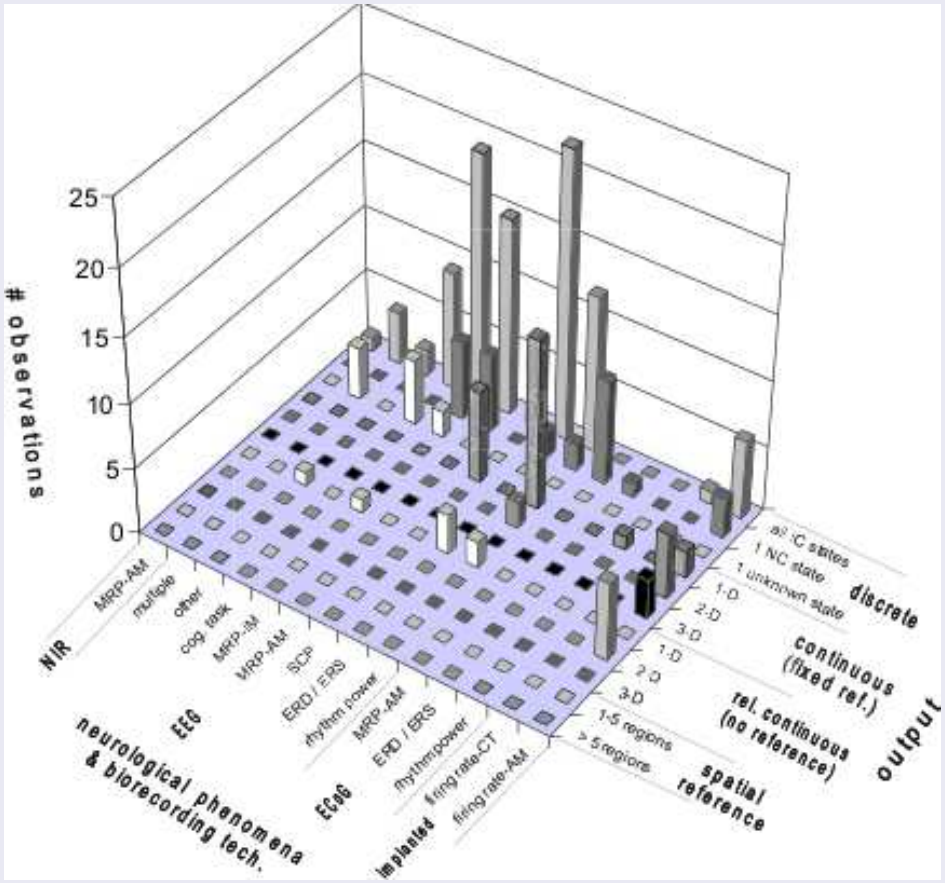


Figure: Bio-recording technology and neurological phenomena (Mason et al: 2007).

Outline

1. Brain Computer Interfaces
2. Machine learning in BCIs
   - Problems
   - Classification methods
   - Advanced machine learning methods
3. Optimization applied to BCI

Outline

1 Brain Computer Interfaces
- Review
- Interfaces

2 Machine learning in BCIs
- Problems
- Classification methods
- Advanced machine learning methods

3 Optimization applied to BCI
- Ensemble approach to P300 speller
- Multi-objective MEG channel selection

Machine learning for BCIs

Application areas

- Classification: Decoding information from brain signals.
- Human adaptation: Implementing adaptive feedback mechanisms to enhance proper control of electrophysiological signal.
- BCI adaptation: Adapting BCI to subject characteristics.

Machine learning for BCIs

Application areas

- Classification: Decoding information from brain signals.
- Human adaptation: Implementing adaptive feedback mechanisms to enhance proper control of electrophysiological signal.
- BCI adaptation: Adapting BCI to subject characteristics.

Machine learning for BCIs

Application areas

- Classification: Decoding information from brain signals.
- Human adaptation: Implementing adaptive feedback mechanisms to enhance proper control of electrophysiological signal.
- BCI adaptation: Adapting BCI to subject characteristics.

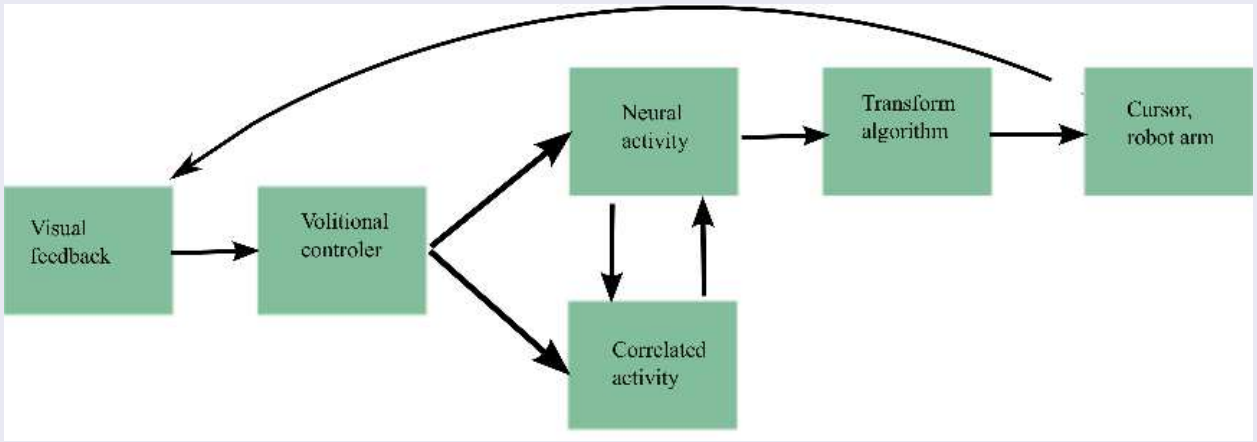


Figure: Basic BCI/BMI paradigm.

Outline

1 Brain Computer Interfaces
- Review
- Interfaces

2 Machine learning in BCIs
- Problems
- Classification methods
- Advanced machine learning methods

3 Optimization applied to BCI
- Ensemble approach to P300 speller
- Multi-objective MEG channel selection

Classification methods in BCIs

Classification

- BCI can be seen as a pattern recognition system (Lotte et al:2007).
- Classification is used to identify patterns of brain activity.
- Performance depend on the features and the classification algorithms.

Classification methods in BCIs

Types of features used for BCI

- Amplitude values of EEG signals (Kaper et al: 2003).
- Band powers (Pfurtscheller et al: 1997).
- Power spectral density values (Chiappa and Bengio: 2004).
- Autoregresive and adaptive autoregresive parameters (Penny et al: 2000).
- Time-frequency features (Wang et al: 2004).
- Inverse model-based features (Qin et al: 2004).

Classification methods in BCIs

Feature properties

- Noise and outliers.
- High dimensionality.
- Time information.
- Non-stationarity.
- Small training sets.

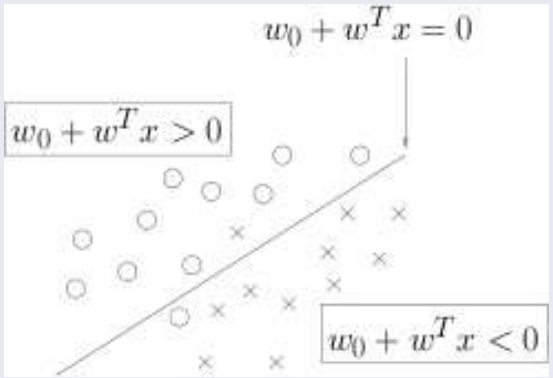
Current machine learning algorithms used for BCIs

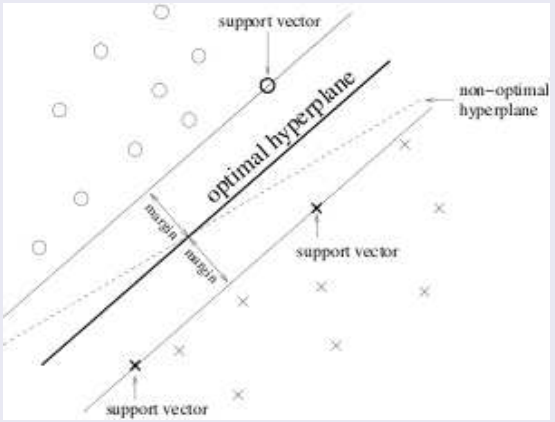
Linear classification methods

- Fisher linear discriminant analysis.
- Support Vector Machines.
- Common Spatial Pattern (CSP) algorithm.

Non-linear classification methods

- Neural networks.
- Fuzzy inference systems.





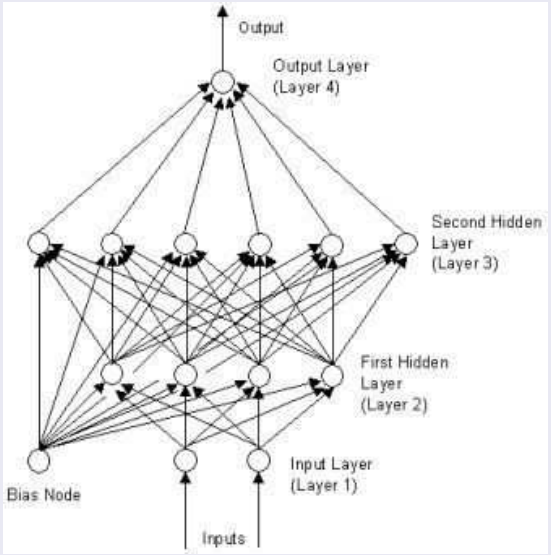


Figure: Classification approaches to BCI.

BCI

General BCI systems

- BCI2000 (Schalk et al: 2004)
- BioSig (Schloegl et al: 2007)
- BCI++ (Maggi et al: 2008)
- Openvibe (Renard et al: 2010)

Outline

1 Brain Computer Interfaces
- Review
- Interfaces

2 Machine learning in BCIs
- Problems
- Classification methods
- Advanced machine learning methods

3 Optimization applied to BCI
- Ensemble approach to P300 speller
- Multi-objective MEG channel selection

Advanced machine learning algorithms for BCIs

Characteristics

- Efficient.
- Scalable.
- Highly parallelized.
- Information-fusion-centered approaches.
- Model generators.
- Simultaneously-solving approaches.

Advanced machine learning algorithms for BCIs

Examples

- Probabilistic graphical models.
- Multi-objective optimization methods.
- Network-based inference methods.
- Regularization.
- Information-fusion strategies.
- Multi-dimensional classifiers.

Probabilistic graphical models

Potential PGM applications in BCI

- Modeling relationships between mental states and brain recordings.
- Representing probabilistic dependencies between brain regions learned from recorded signals.
- Unveil hidden relationships between channels signal output which could eventually serve as brain signatures.
- Modeling BCI demographics.

Regularization methods in BCI

Regularization

- Regularized common spatial patterns (Lotte and Guan: 2010).
- Regularized linear discriminant analysis (Lei et al: 2009).
- Regularized Fisher discriminant analysis (Hoffmann et al: 2007, Zhdanov et al: 2007).
- Sparse logistic regression (van Gerben et al: 2009).

Open questions in BCI research

Open problems

- Inter-subject and inter-session variability.
- Asynchronous BCI.
- Using BCI without training.
- Subject variability.
- BCI interpretability.
- Intuitive universal interface for BCI.
- BCI demographics.

Open questions in BCI research

Inter-subject and inter-session variability

- For any specific experimental paradigm, the underlying neural rhythmic activity associated with a task of interest remains invariant.
- Spatio-temporal characteristics of EEG signals vary greatly among subjects or repetitions of EEG experiments.

Outline

1. Brain Computer Interfaces
2. Machine learning in BCIs
3. Optimization applied to BCI
   - Ensemble approach to P300 speller
   - Multi-objective MEG channel selection

BCI optimization problems

Optimization approaches

- Many problems in BCI can be posed as optimization problems.
- Some classification problems can be addressed from the optimization point of view.
- Can we simultaneously optimize a problem and model the relationships between its components?

Outline

1 Brain Computer Interfaces
- Review
- Interfaces

2 Machine learning in BCIs
- Problems
- Classification methods
- Advanced machine learning methods

3 Optimization applied to BCI
- Ensemble approach to P300 speller
- Multi-objective MEG channel selection

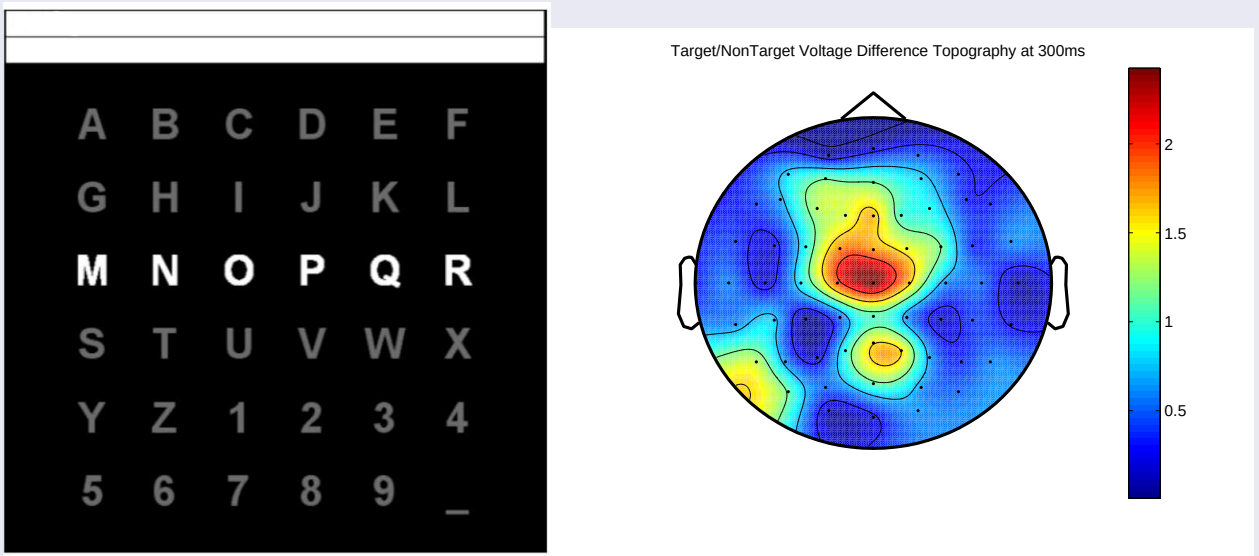


Figure: P300 speller and P300 response.

P300 speller

Experimental framework

- Training and test data from two different subjects.
- Training set: Data recorded from 64 EEG channels during the identification of 85 *known* characters.
- Test set: Similar data for 100 *unknown* characters.
- Problem: Using the training data, predict the characters in the test set.

P300 Speller

Decision hyper-plane

The presence or absence of a P300 evoked potential from EEG features can be considered a binary classification problem with a discriminant function having a decision hyper-plane defined by:

$$w \cdot f(\mathbf{x}) + b = 0 \tag{1}$$

where $\mathbf{x}$ is a feature vector, $f(.)$ is a transformation function, w is a vector of classification weights and b is the bias term.

P300 speller

Identification of target row and column

The resultant classification is taken as the maximum of the sum of scored feature vectors for the respective rows, as well as for the columns:

$$r^* = argmax_{rows} \left[\sum_{i_{row}} w \cdot f(x_{i_{row}}) \right] \tag{2}$$

$$c^* = argmax_{columns} \left[\sum_{i_{column}} w \cdot f(x_{i_{column}}) \right] \tag{3}$$

The predicted character is located at the intersection of the predicted row and column in the matrix.

P300 speller

Pure optimization approach

- Find a suitable transformation of the brain signal to a feature vector **x**.
- Using an optimization algorithm, find an optimal weight vector (maximal predicted accuracy) using the training data **w**.
- Evaluate the accuracy of the optimal solution using the test data.

P300 speller

Extended optimization approach: Ensembles

- Ensemble method: Weight vectors may specialize in different types of brain signatures.
- Ensembles may be more sensitive to the detection of the P300 evoked potential.
- A simple voting model is used: The character that receives the highest number of votes by the weight vectors is selected.

P300 speller

Extended optimization approach: Ensembles

- How to decide which weight vectors are part of the ensemble?
- k-best ensemble: The k *optimal* solutions with the highest accuracy (in the training set) are selected.
- k-worst ensemble: The k *optimal* solutions with the lowest accuracy (in the training set) are selected.
- For $k = 1$ we have the usual optimization approach.

P300 speller

Issues to be considered

- Set of channels ($\{C_1, \ldots, C_{nchannels}\}$): Select all or only a subset from the complete set of channels?
- Time window (*TW*): Determining the optimal time window in terms of the classification accuracy is also an optimization problem.
- Number of features per time window (*nFeat*): Deciding which is the optimal number of features to take is part of the optimization problem.
- The vector of weights has ($nchannels \times nFeat$) variables.

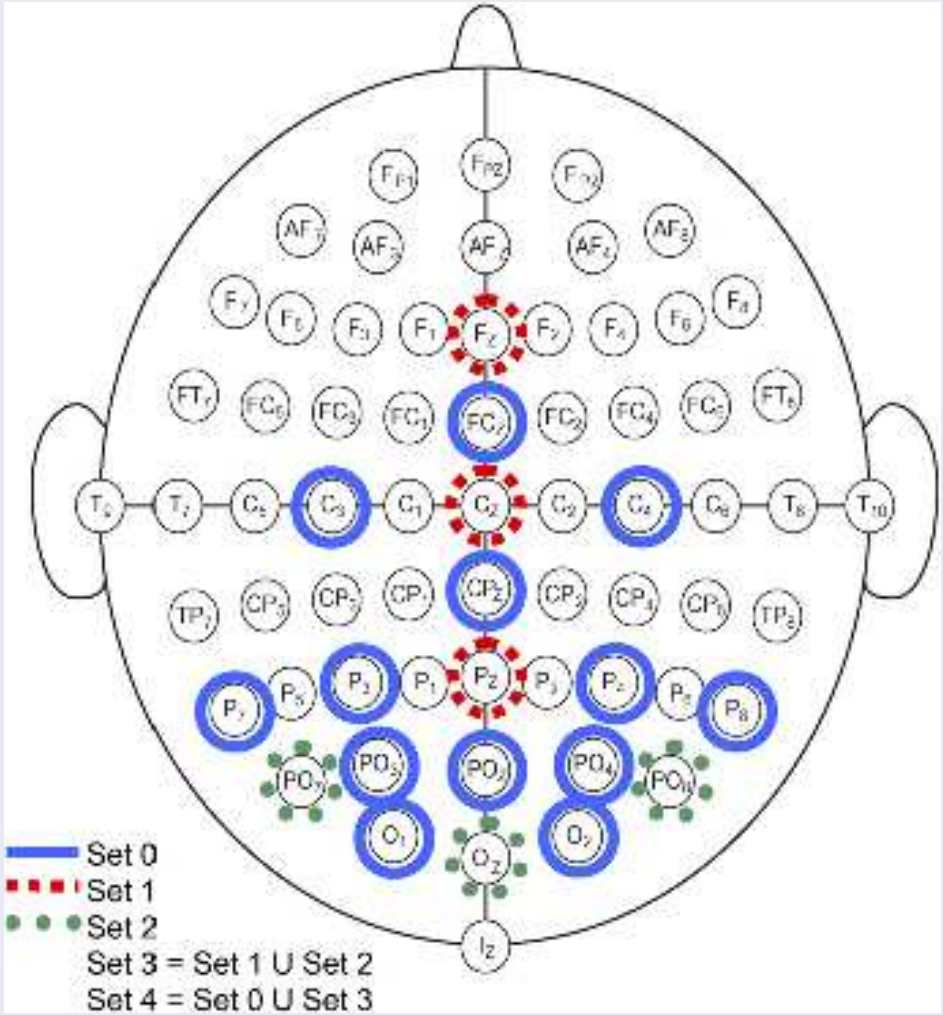


Figure: Position of the original 64 electrodes. The 23 selected channels were: P1, Pz, P2, PO7, PO3, POz, PO4, PO8, O1, Oz, O2, Iz, P3, P4, CP3, CP1, CPz, CP2, CP4, C1, Cz, C2 and FCz.

P300 speller

Parameters of the optimization problem

- Time window size: $755ms$.
- Number of feature per window: 5.
- Weight ranges $w_i \in [-10, 10]$.
- Number of variables $n = 23 \times 5 = 115$.

P300 speller

Optimization algorithms

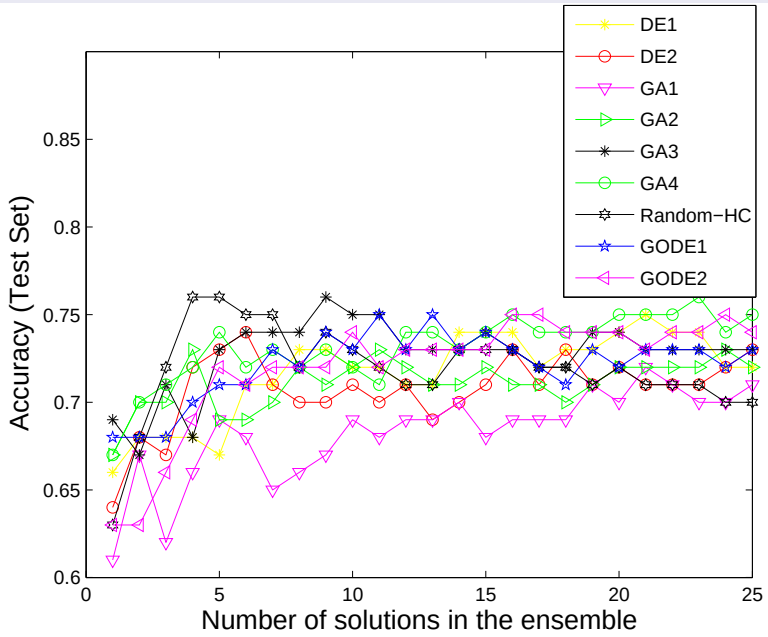
- DE-1 : Psize: 20, CR: 0.5, F: 0.5, Sel.: random
- DE-2 : Psize: 50, CR: 0.5, F: 0.5, Sel.: random
- GA-1 : Psize: 50, CX : BLX-α, Mutation: Gaussian, CX Prob.: 0.9, Mut. Prob.: 0.01, Sel.: tournament 2, elitism: 100%
- GA-2 : Psize: 50, CX : BLX-α, Mutation: Gaussian, CX Prob.: 0.9, Mut. Prob.: 0.01, Sel.: tournament 2, elitism: 50%
- GA-3 : Psize: 50, CX : BLX-α, Mutation operator: Gaussian, CX Prob.: 0.9, Mut. Prob.: 0.1, Sel.: tournament 2, elitism: 100%
- GA-4 : Psize: 50, CX : BLX-α, Mutation: Gaussian, CX Prob.: 0.9, Mut. Prob.: 0.1, Sel.: tournament 2, elitism: 50%
- RHC : The maximum number of random changes is equal to the maximum number of allowed evaluations.
- GODE-1 : Psize: 20, CR: 0.5, F: 0.5, Probability of OBL $p_o = 0.4$,Sel.: random
- GODE-2 : Psize: 50, CR: 0.5, F: 0.5, Probability of OBL $p_o = 0.4$,Sel.: random

P300 speller

Results

Alg/*k*	1	3	5	10	15	20	25
DE − 1	0.69	0.73	0.74	0.77	0.79	**0.80**	0.79
DE − 2	0.73	0.74	0.78	0.76	0.77	0.77	0.78
GA − 1	0.64	0.69	0.75	0.75	0.74	0.76	0.77
GA − 2	**0.75**	**0.78**	0.78	**0.80**	**0.80**	**0.80**	0.79
GA − 3	0.69	0.76	0.78	0.79	0.78	0.79	0.79
GA − 4	0.71	0.74	0.76	0.77	0.78	**0.80**	**0.80**
RHC	0.65	0.75	**0.81**	**0.80**	0.79	0.79	0.79
GODE − 1	0.73	0.76	0.77	0.78	0.79	0.79	**0.80**
GODE − 2	0.72	0.74	0.78	0.79	0.77	0.78	0.78

Table: Accuracy results computed from the test-data and averaged between the two subjects. Results have been achieved with different k-best ensembles.



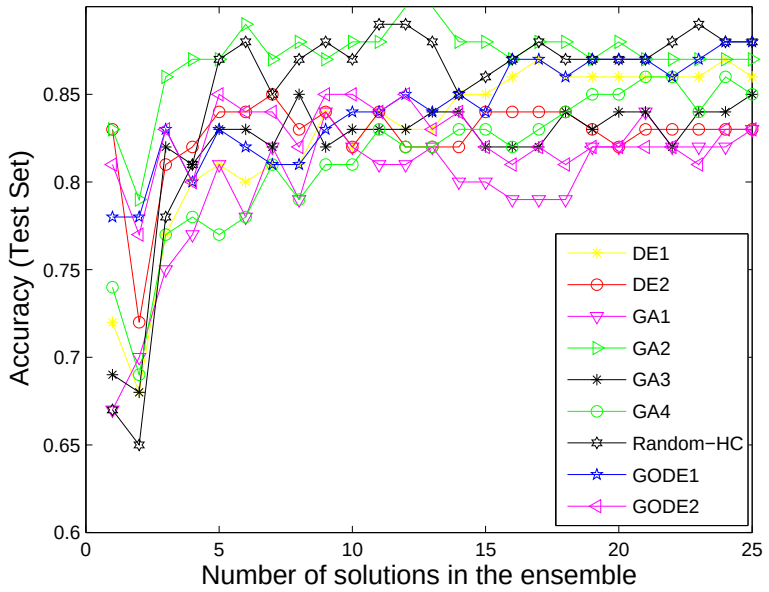


Figure: Results of the k-worst ensembles.

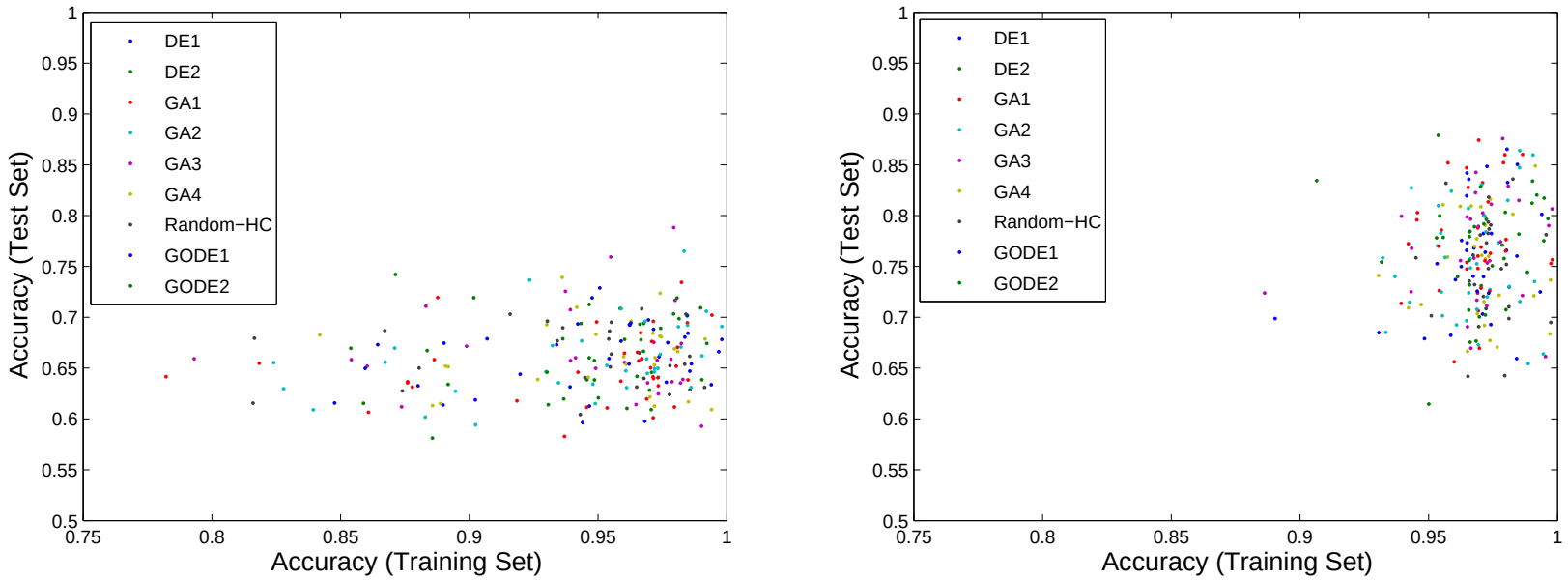


Figure: Mapping between training and test predicted accuracies.

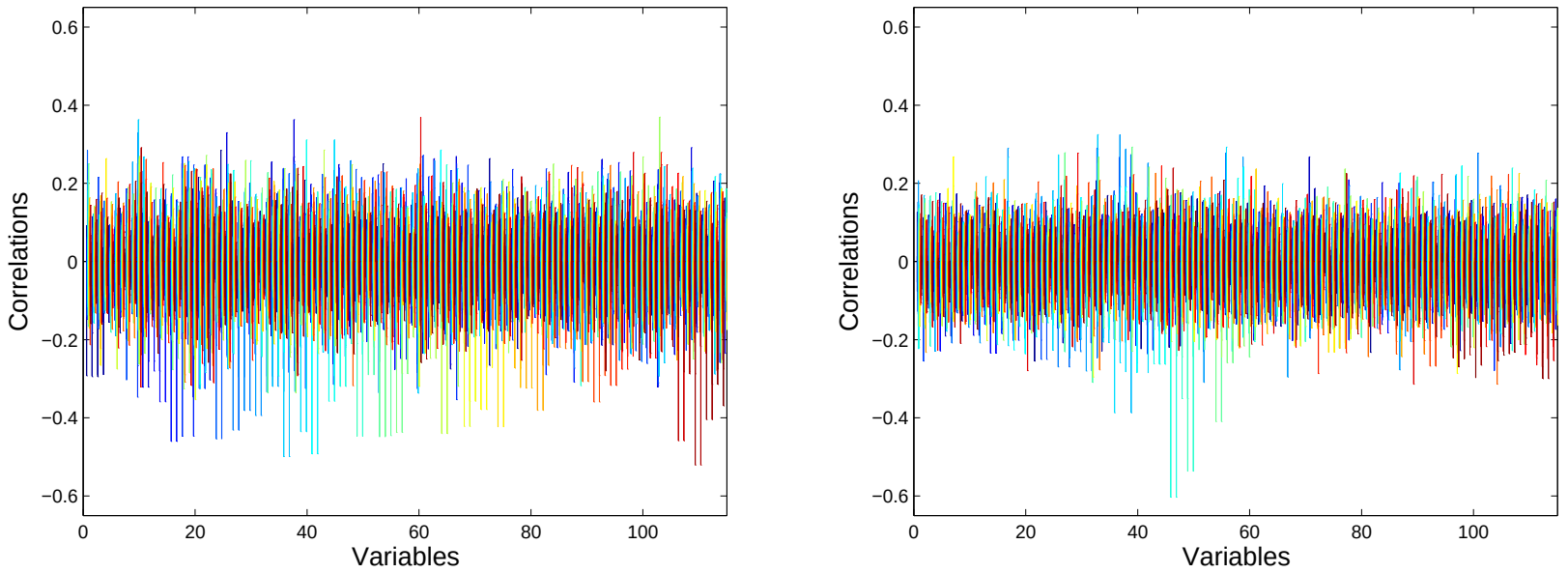
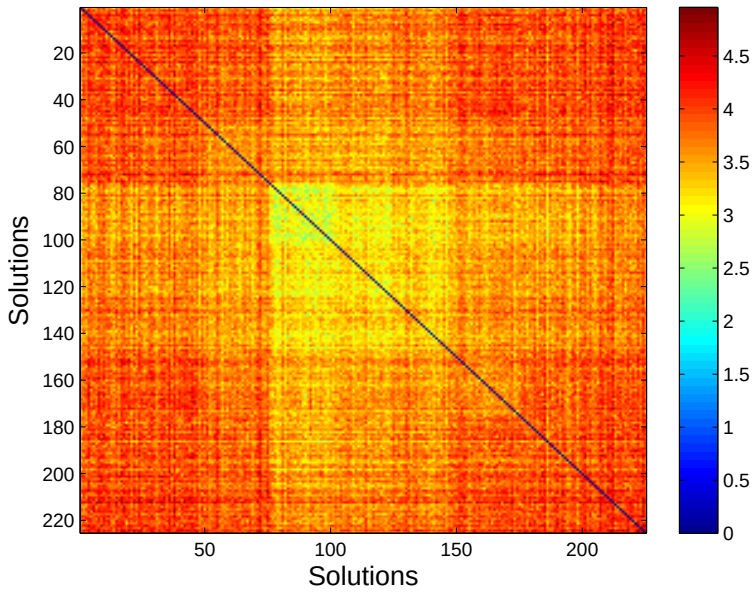


Figure: Correlation between the problem variables.



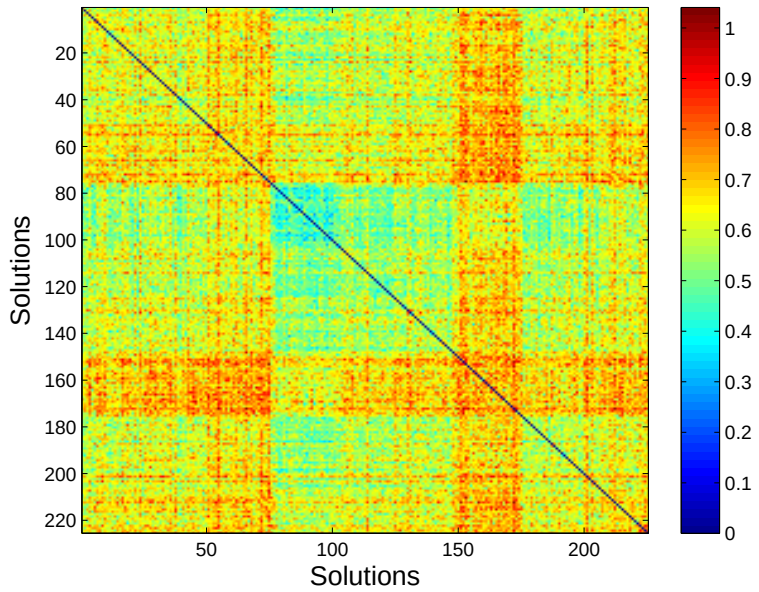


Figure: Similarities between the optimal solutions using the Euclidean and correlation distances.

Outline

1 Brain Computer Interfaces
- Review
- Interfaces

2 Machine learning in BCIs
- Problems
- Classification methods
- Advanced machine learning methods

3 Optimization applied to BCI
- Ensemble approach to P300 speller
- Multi-objective MEG channel selection

MEG channel selection problem

Covert attention paradigm

- Attention may be differentiated according to its status as *overt* versus *covert*.
- Overt attention is the act of directing sense organs towards a stimulus source.
- Covert attention is the act of mentally focusing on one of several possible sensory stimuli.
- Covert attention to spatial locations in the visual field alone may be sufficient to drive a BCI (van Gerven et al:2009).
- Covert shifts in visual attention are paired by alpha-desynchronisation contralateral and alpha-synchronisation ipsilateral in posterior sites.

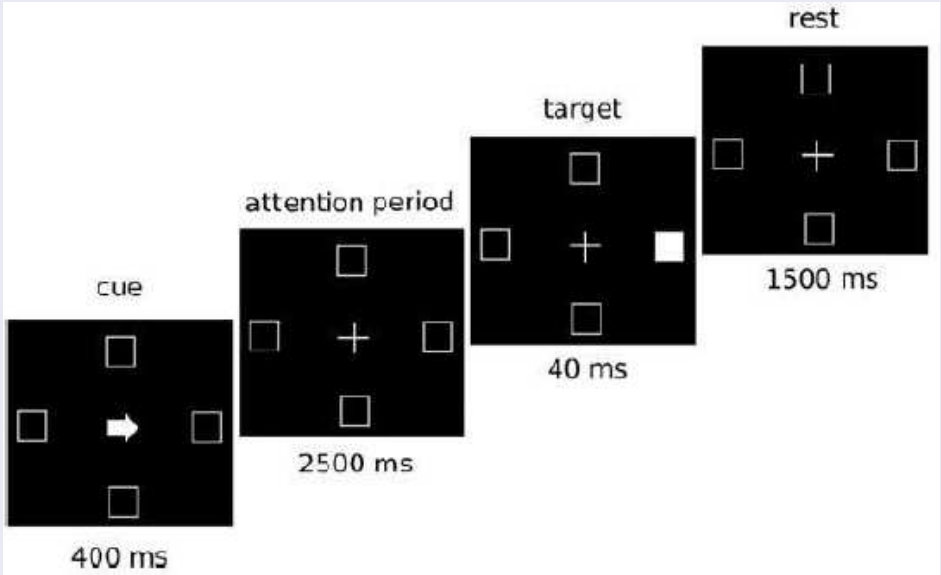


Figure: Timeline of a trial in the covert attention experiment.

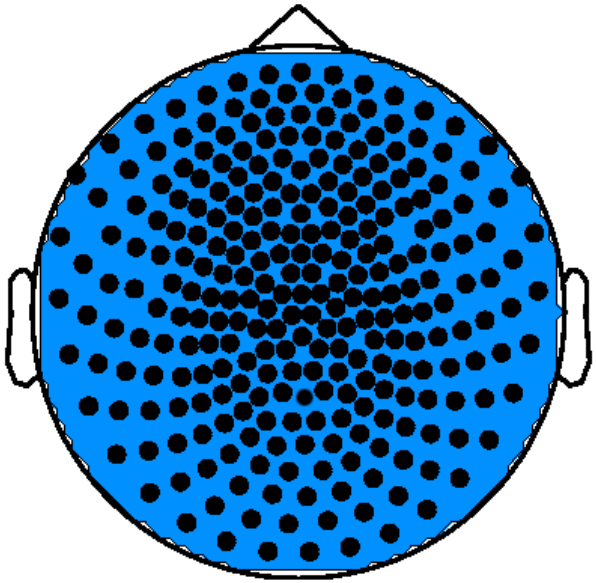
P300 speller

Experimental framework

- Training and test data from four different subjects.
- Data sets: Data recorded from 274 MEG channels during the presentation of 254 visual stimuli (128 left, 128 right).
- Problem: Based on the recorded data, predict the direction (left or right) the subject was covertly attending.

BCI optimization problems

Channel selection



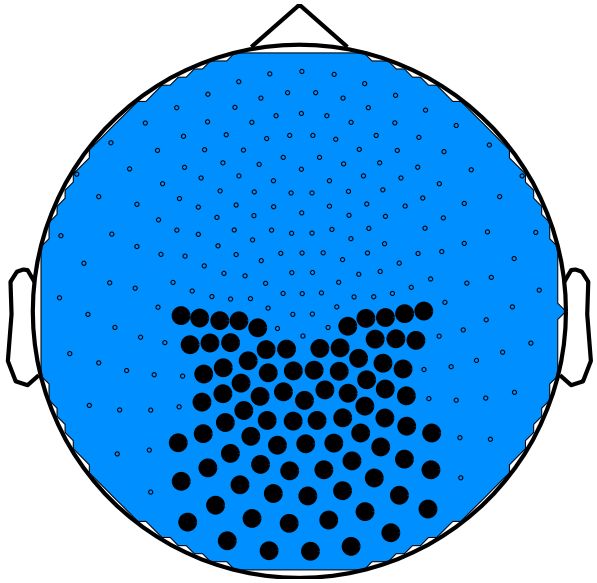


Figure: Complete set of channels (left) and occitoparietal channels(right).

Selection of optimal channels

Feature subset selection approach using optimization

- Find the set of channels (features) whose corresponding variables, passed to the classifier, give the best accuracy.
- A binary vector $\mathbf{x} = (x_1, \ldots, x_{274})$ represents a possible subset of channels.
- Estimate accuracy with (less costly) 2-fold cross-validation accuracy.
- Simultaneously optimize the accuracy for all the subjects with multi-objective optimization, (likely) sacrificing accuracy but increasing robustness.
- Reevaluate with leave-one-out cross-validation accuracy only solutions in the Pareto set.

Problem approach

Factors in MEG data analysis

- Type of information.
- Type of classifier.
- Channels from which information is extracted.

Solution aproach

Combined machine learning approach

- Apply a multi-variable channel information representation.
- Use and combine different type of information (raw data, channels correlation, interaction graphs).
- Apply an efficient classification method (deals with up to 53841 variables and 254 cases).
- Search for a subset of informative channels (out of 274) using a multi-objective evolutionary-based optimization algorithm over the four subjects.
- Use classification accuracy estimation to reduce the cost of the accuracy evaluation step.
- Apply leave-one-out cross-validation to the Pareto set of optimal solutions.

Representation and type of information

Representation

- Each channel is represented by a vector of m variables.

Type of information

- Raw data: Average data from raw time series ($m = 60$).
- Channel correlation: Correlation matrix from raw time series ($m \leq 273$).
- Interaction graph topological measures: ($m = 13$ plus 7 global variables).
- Combined representation: Raw data and channel correlation ($m \leq 333$).

Representation and type of information

Raw data

- The original set of 600 time points is reduced to a set of 60 components by averaging in each component a time window comprising 10 points.
- The classifier will receive a vector of $n = 60 \times 274 = 16440$ features.

Representation and type of information

Correlation matrix

- The correlation matrix between the time series corresponding to all the channels for a given trial is computed.
- The classifier will receive a vector of $n = \frac{274 \cdot 273}{2} = 37401$ features corresponding to the upper triangular part of the correlation matrix.

Representation and type of information

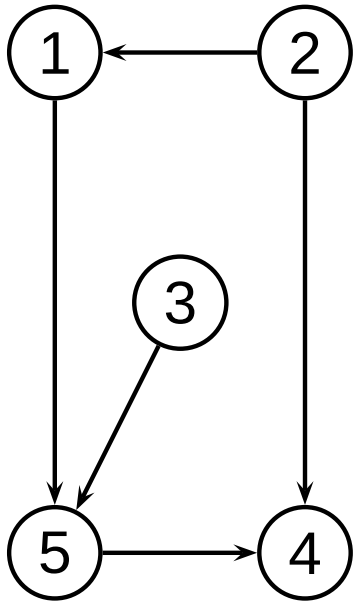
Correlation graph

The interaction graph $G = (V, A)$ is such that $V = \{v_1, \ldots, v_{274}\}$ is the set of vertices and arc $a_{i,j}$ between vertices v_i and v_j is defined in the following way:

$$a_{i,j} = \begin{cases} 1 & \text{if} \quad i < j \text{ and } cr_{i,j} > 0.5 \\ -1 & \text{if} \quad i < j \text{ and } cr_{i,j} < -0.5 \\ 0 & \text{if} \quad \text{otherwise} \end{cases}$$

- Network measures are extracted from the correlation graph.

$$\begin{pmatrix} 1 & -0.82 & 0.43 & 0.02 & 0.94 \\ -0.82 & 1 & 0.13 & 0.89 & -0.04 \\ 0.43 & 0.13 & 1 & 0.15 & 0.88 \\ 0.02 & 0.89 & 0.15 & 1 & -0.73 \\ 0.94 & -0.04 & 0.88 & -0.73 & 1 \end{pmatrix}$$



a) b)

Figure: a) Initial correlation matrix. b) Interaction graph.

Networks

How to construct networks from EEG and MEG data?

- Channel correlation.
- Mutual information.

Networks

Local network measures

1. *Betwenness centrality*
2. *Pair distance*
3. *Node eccentricity*
4. *Clustering coefficient*
5. *Indegree*
6. *Outdegree*
7. *Motif frequency,* $M = 3$
8. *Maximum modularity*
9. *Vertex participation coefficient*

Networks

Global network measures

1. *Assortativity coefficient*
2. *Characteristic path length*
3. *Network radius*
4. *Network diameter*
5. *Ndensity*
6. *Number of vertices*
7. *Number of edges*

Network measures

Motifs

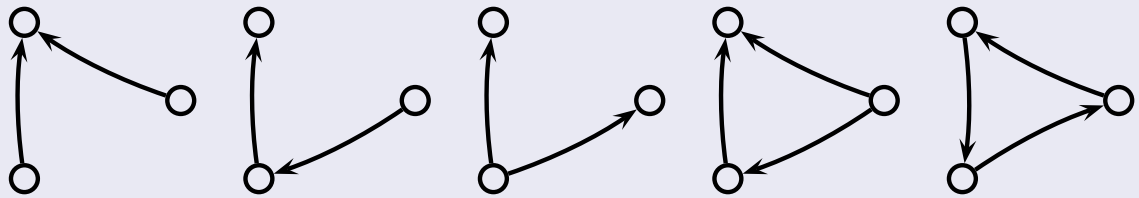


Figure: All motifs ($M = 3$) which appear in the correlation graphs learned from MEG data.

Classification

Classification

- Regularized logistic regression classifier using elastic net (Friedman et al. 2009).
- *lambda* selection: *lambda* that maximizes accuracy of the training set.

Multi-objective optimization

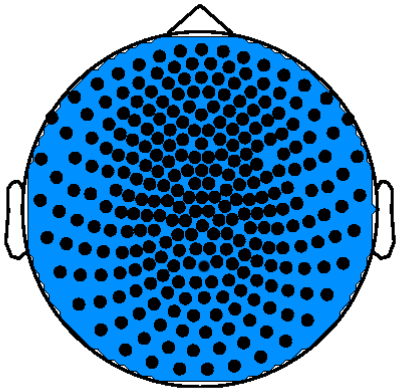
More than one objetive to optimize

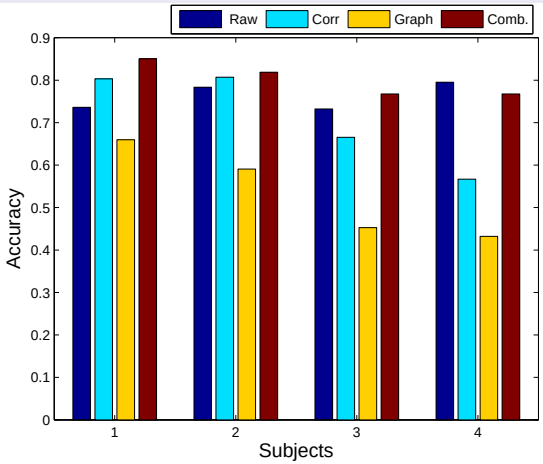
- What if more that one objective has to be optimized?
- E.g.: Minimize number of channels and maximize accuracy.
- How to find a compromise between different objectives?
- Find Pareto set approximations.

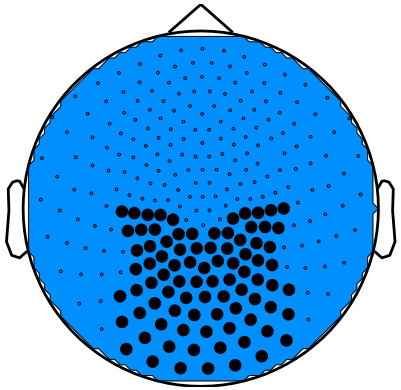
Multi-objective EBNA

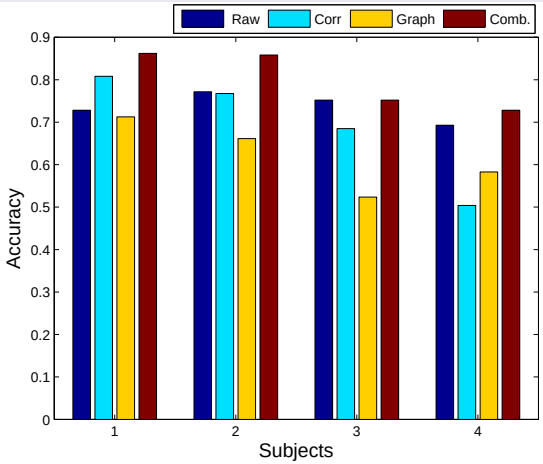
1. $BN_0 \leftarrow (S_0, \boldsymbol{\theta}^0)$ where $S_0$is an arcless dag and $\boldsymbol{\theta}^0$ is uniform
2. $p_0(\mathbf{x}) = \prod_{i=1}^{n} p(x_i) = \prod_{i=1}^{n} \frac{1}{r_i}$
3. $D_0 \leftarrow$ Sample M individuals from $p_0(\mathbf{x})$
4. t $\leftarrow 1$
5. **do** $\{$
6. $D_{t-1}^{Se} \leftarrow$ Select N individuals from D_{t-1} using Pareto-ranking selection.
7. $S_t^* \leftarrow$ Use local search to find one network structure that optimizes scoring metric
8. $\boldsymbol{\theta}^t \leftarrow$ Calculate θ_{ijk}^t using D_{t-1}^{Se} as the data set
9. $BN_t \leftarrow (S_t^*, \boldsymbol{\theta}^t)$
10. $D_t \leftarrow$ Sample M individuals from BN_t
11. $\}$ **until** Stop criterion is met

Accuracy results for the four types of information and all 274 channels (top) and occipitoparietal channels (bottom)









Results

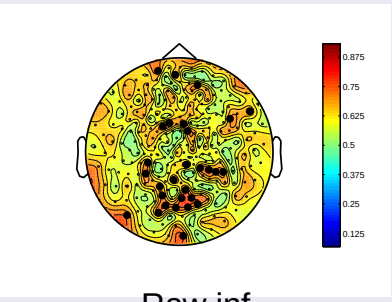
Answer to the challenge: Best achieved results using the combined type of information

Subjects	1	2	3	4
Accuracy	0.8701	0.8505	0.8031	0.8031

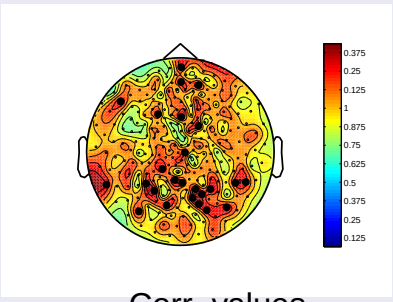
Analysis of the Pareto set solutions

- Most informative channels can be automatically extracted from Pareto set of solutions and compared with a priori known sets of involved brain areas.
- The accuracy provided by each channel can be estimated by averaging accuracies of solutions where each channel is involved.

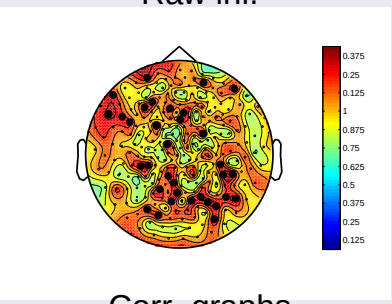
Channels that were in at least 80% of Pareto set solutions (black) and average channel classification accuracy (color)



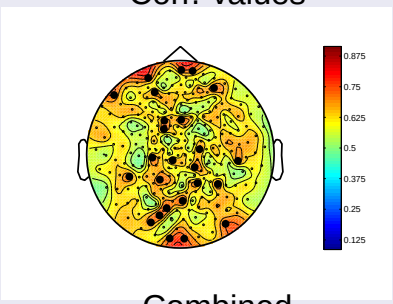
Raw inf.



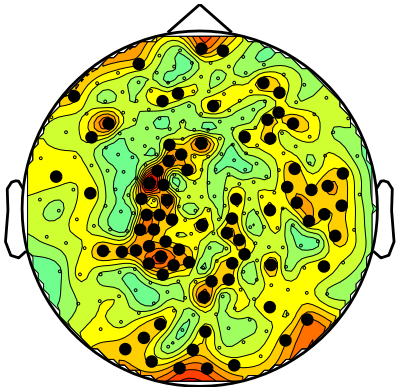
Corr. values



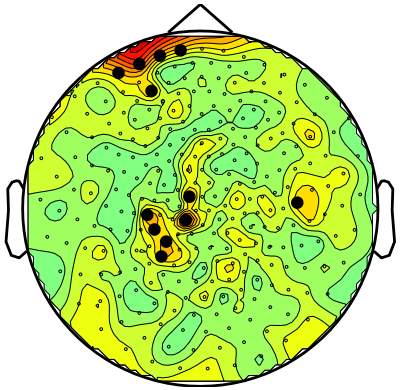
Corr. graphs



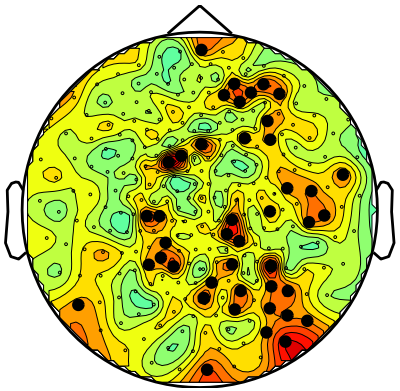
Combined



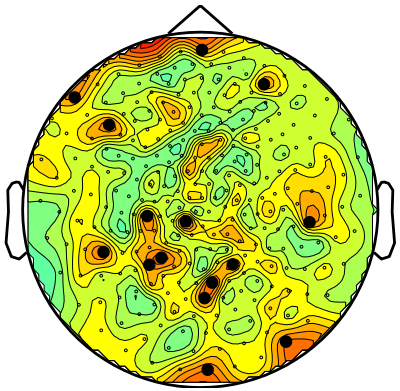
Subject 1



Subject 2



Subject 3



Subject 4

Figure: Channels that were identified as relevant.

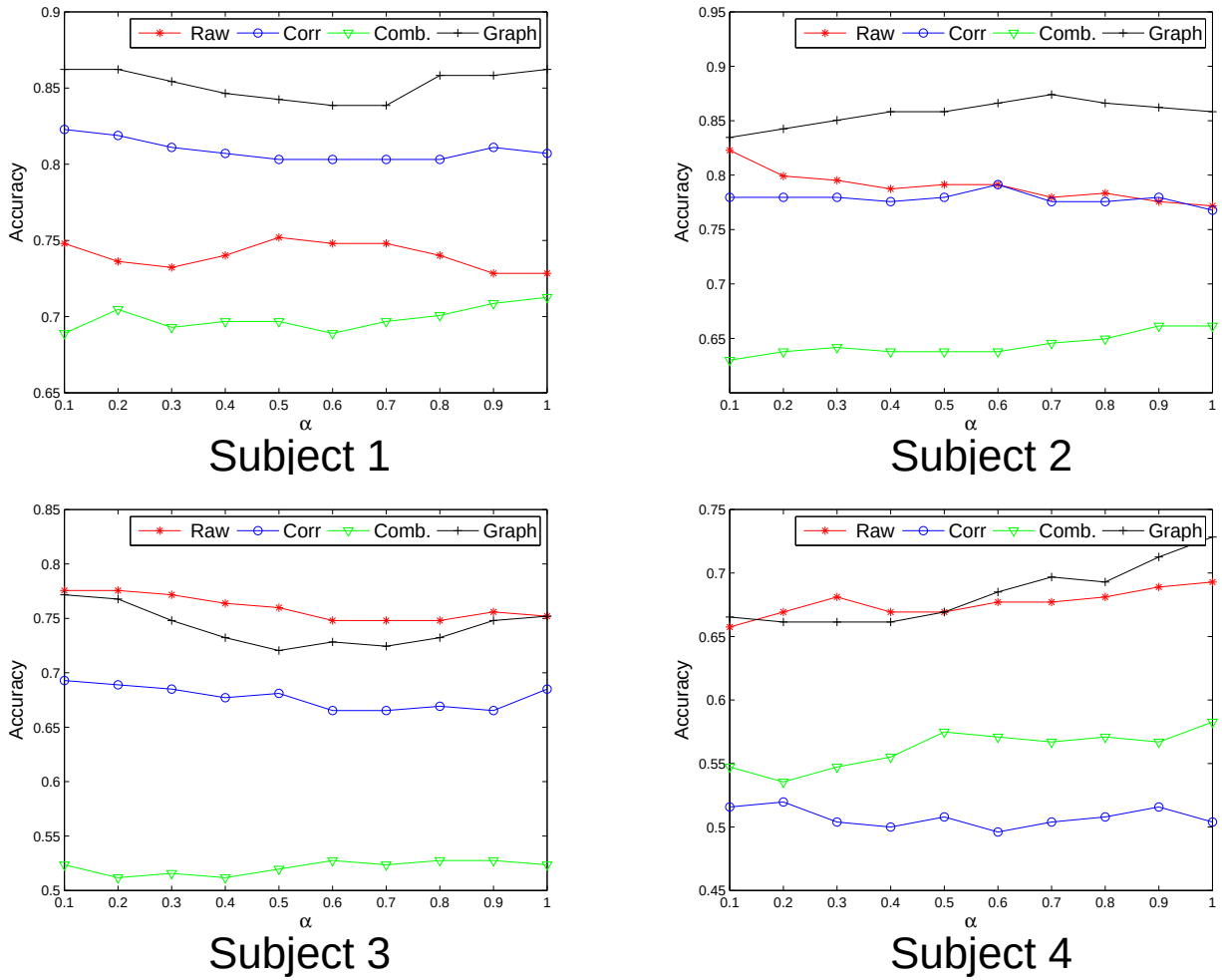
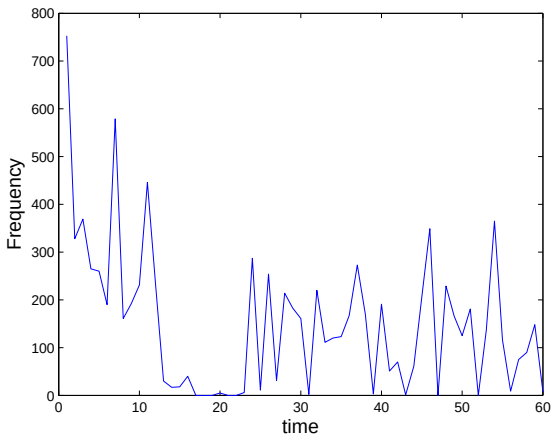
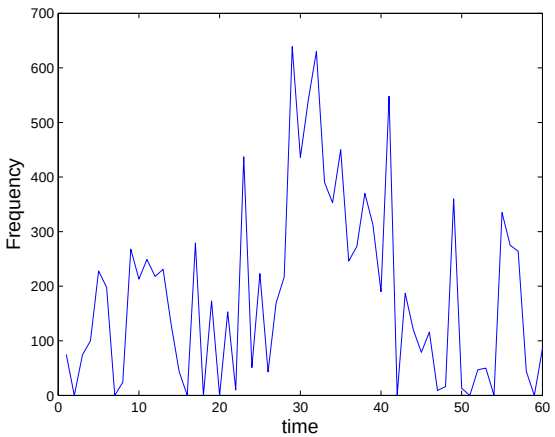


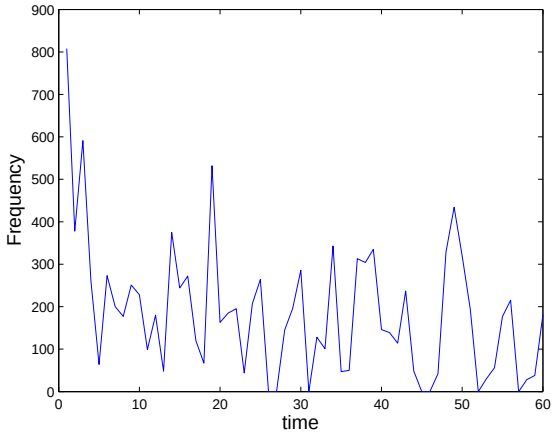
Figure: Accuracy of the classification results as a function of the parameter α.



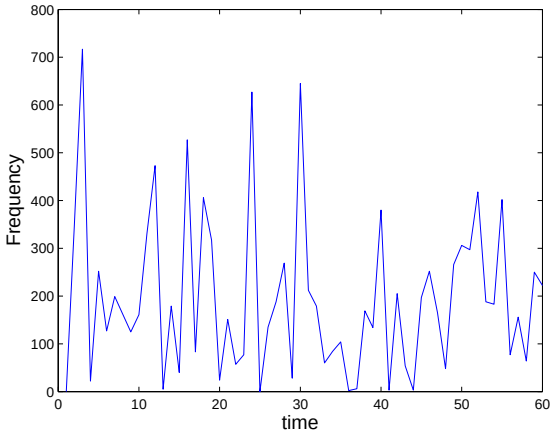
Subject 1



Subject 2



Subject 3



Subject 4

Figure: Number of non-zero coefficients for each of the 60 variables.

Optimization algorithms for brain computer interfaces

Roberto Santana

Cajal Blue Brain Project
Computational Intelligence Group
Technical University of Madrid

February 24, 2011