



Sudan University of Science And  
Technology

**MSc. Program in Biomedical  
Engineering**

# **Discrimination of Four Classes In Brain Computer Interface Based on Motor Imagery**

Tasneem Mahmoud Salih

Supervisor:  
Dr.Omer E.H.Hamid



MSc. Thesis Defense  
Feb. 15, 2011



## Outline

### ❖ Introduction

- general introduction.
- Motivation.
- Thesis objectives.
- What is a BCI.
- GRAZ BCI system
- BCI Components.

### ❖ Signal processing.

- Dimension reduction.
- Feature Extraction.

### ❖ Classification and Pattern Recognition.

### ❖ Results & Discussion.

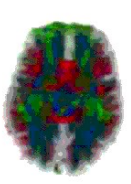
### ❖ Conclusion.

### ❖ Recommendation.

# Introduction



- ❑ Disabled people with a high spinal cord injury (SCI) not only the lower limbs, but also the upper extremities are paralyzed.
- ❑ A neuroprosthesis can be used to restore the loss of movement ability in those individuals. Brain-computer interface provides a voluntarily, non-manual control for artificial limb or device by translating brain activity patterns into control commands.



# Motivation

- Developing technologies for people with disabilities.
- Assist paralyzed people to operate external devices without physical movement.



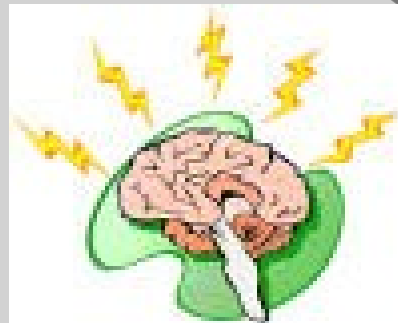
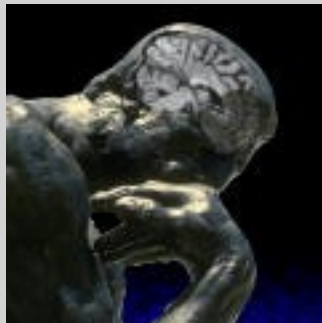
## Thesis Objectives

- In this research four classes of different motor imagery tasks are predicted and discriminated, namely the imagination of movement of the left hand (class 1), right hand (class 2), both feet (class 3), and tongue (class 4) by using data set 2a of BCI competition IV, that includes EEG data from 9 subjects.
- Principle Component Analysis (PCA) and Independent Component Analysis (ICA) are used in feature extraction. Support Vector Machine (SVM) is used in classification process.



# What is a BCI?

- Brain-computer interface (BCI) can be defined as a direct communication pathway between a brain and an external device.
- People can send information simply by thinking, without actual movement.
- BCIs translate brain's electrical activity into messages or commands.



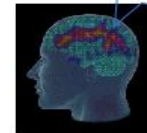
performing mental tasks produces electrical activity detectable with electrode caps

- ❑ Motor imagery can be used instead of actual movement due to similarity between action and imagination goes beyond motor movements.

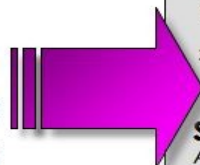


- ❑ “Motor imagery can be defined as a dynamic state during which an individual mentally simulates a given action.” This type of phenomenal experience implies that the subject feels himself performing the action.

## Functional equivalence between action performance and motor imagery



- Planning
- Preparing
- Simulating
- Observing



### Motor representation

As demonstrated by chronometric experiments involves two main components:

1. a representation of the body in action as a generator of forces.
2. a representation of the goal of action encoded in a pragmatic mode.

### Similar neurophysiological substrate

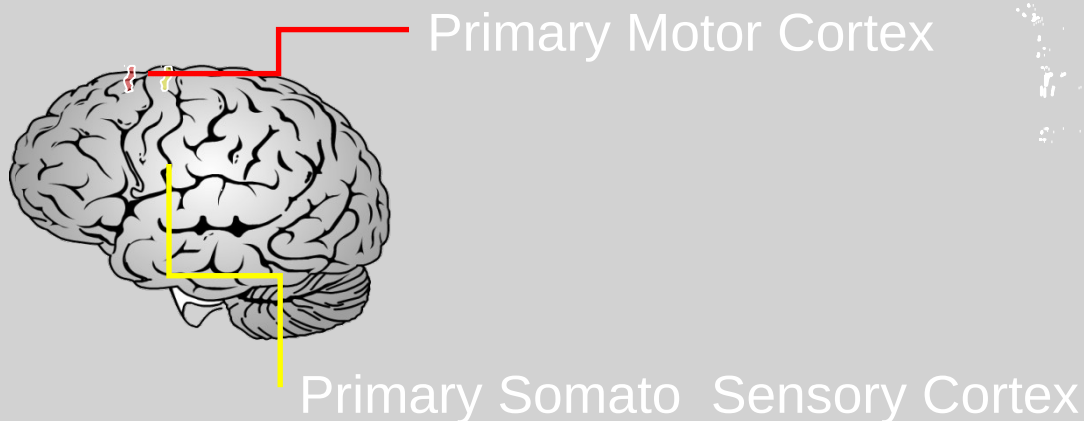
As demonstrated by:

- Measurements of the autonomic nervous system.
- Brain mapping techniques (PET and fMRI).
- Neuropsychological observations in patients.



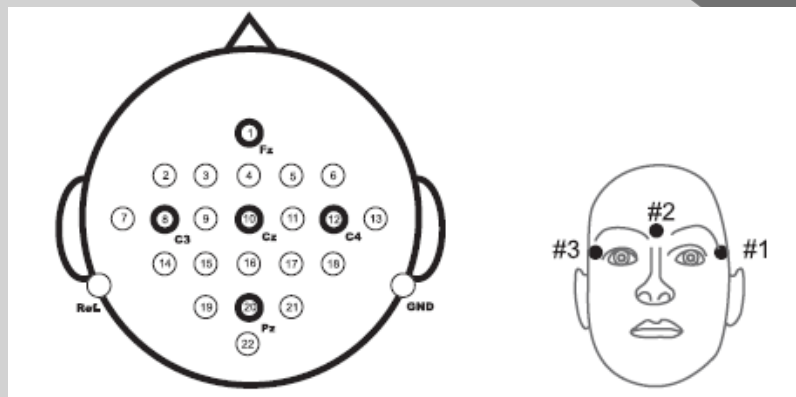
# Graz BCI system

- Graz-BCI system based on classification of motor imagery related brain activity changes in ongoing EEG
  - › Electrodes placed on the scalp
  - › Movement imagination (left hand, right hand, foot or tongue)

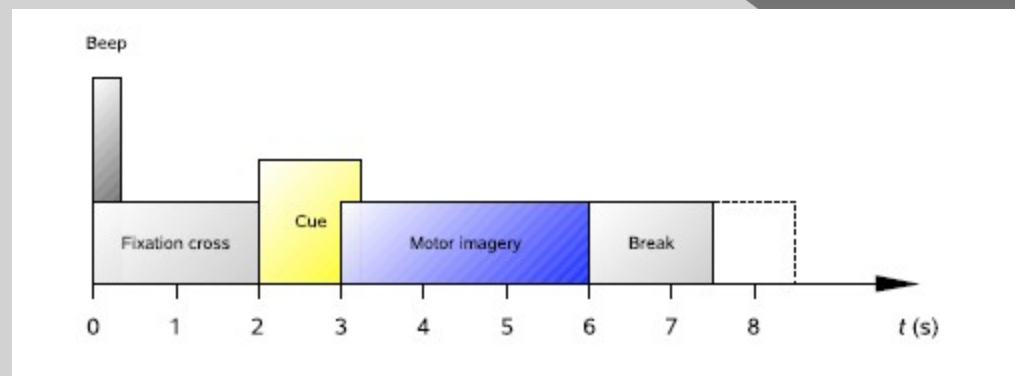




- Data set 2a from BCI competition IV (provided by the Graz group) comprises EEG signals from 9 subjects who performed left hand, right hand, foot and tongue MI by using twenty-two Ag/AgCl electrodes (with inter-electrode distances of 3.5 cm) were used to record the EEG.

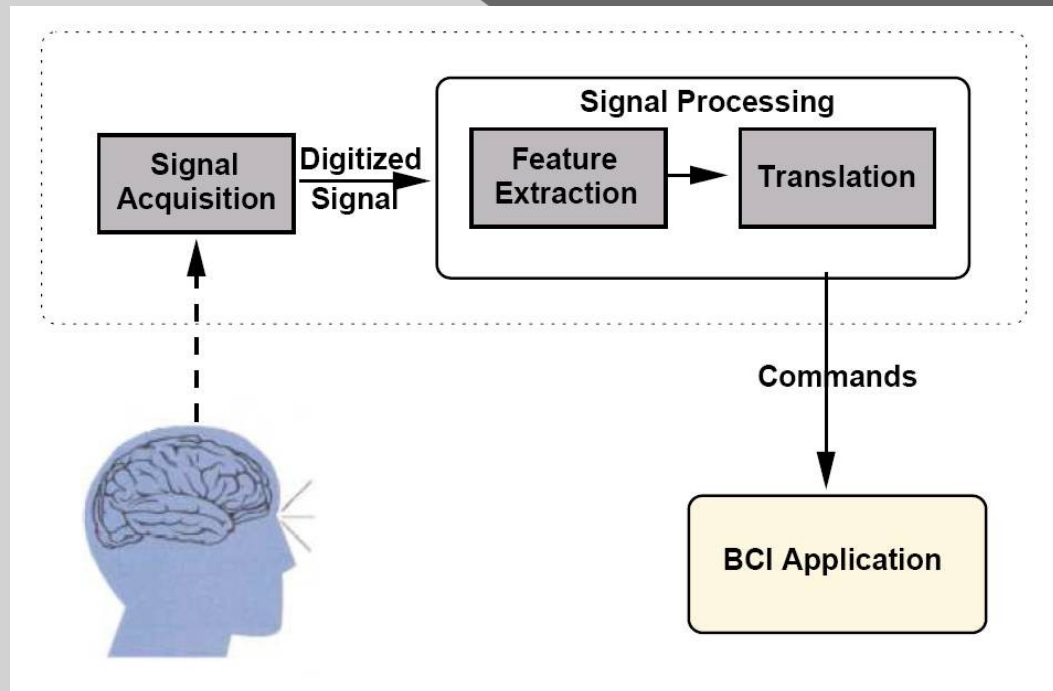


- The subjects were sitting in armchair in front of screen. At ( $t = 0$  s), the beginning of a trial, a fixation cross appeared on the black screen. In addition, a warning tone was presented.
- After two seconds a cue in the form of an arrow pointing either to the left, right, down or up (corresponding to one of the four classes: left hand, right hand, foot or tongue) appeared and stayed on the screen for 1.25 s



# Components of BCI System

- BCI is a control system

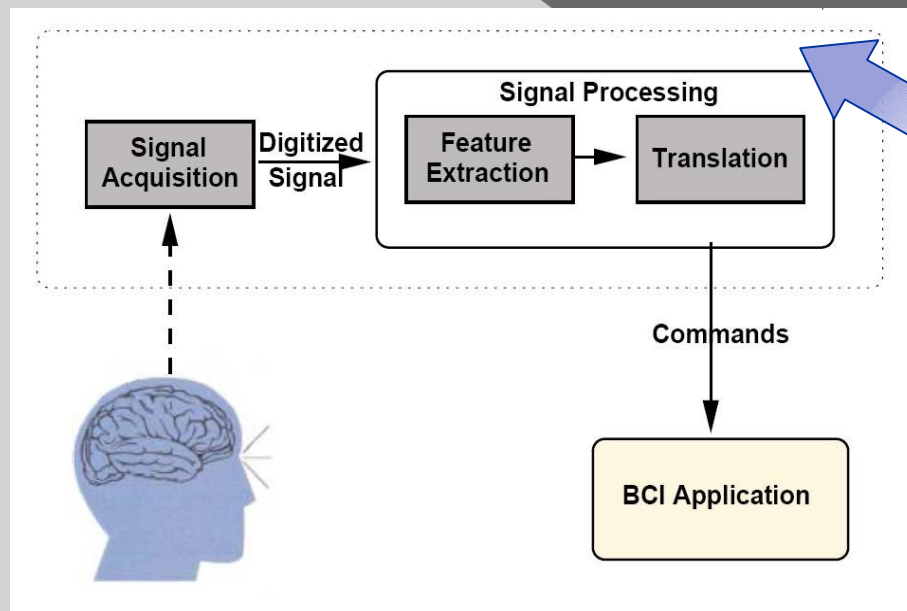


# Signal Acquisition

- Brain signals can be collected in different ways, one of these methods is EEG (Electroencephalography)
  - › Non-invasive
  - › Mu Rhythms
- In awake people, even when they are not producing motor output, motor cortical areas often display 8–12 Hz EEG activity (Mu Rhythm)
- Movement or preparation for movement typically causes a decrease in mu rhythms (motor imagery)

# Signal Processing

- Feature Extraction
- The Translation Algorithm



# Output Device

- ▣ Any controllable machines
  - › For answering yes/no questions
  - › For word processing at slow
  - › Wheelchair
  - › Virtual Reality
- ▣ Usually, Computer screen and the output is the selection of targets or cursor movement

## Problem definition



- EEG signals have been applied in augmentative communication and control technology for those with severe neuromuscular disorders such as amyotrophic lateral sclerosis and stroke. In a (BCI) system, these brain signals are directly translated into commands for controlling an external device.
- There are several techniques to achieve BCI depend on measurement techniques (invasive or non-invasive), Feature selection and Extraction (dimensionality reduction) and finally Classification.
- The Brain Computer Interface in this project is an EEG- based system.





- There are some difficulties of processing classifying acquired EEG signals related to responses to some visual stimuli and those are:
- The low signal- to- noise ratio of the signal.
- High subject variance in performance.
- The variability between different subjects.
- Low detection rates in mental tasks.
- Slow command speed.
- Low number of possible decision per command.
- Slow response times.
- Cumbersome preparation.
- Then, to achieve accepted results, it is so important to address these points when choosing a suitable classification technique.
- This project proposed (ICA,PCA and SVM) approaches to instruct the subject to perform tasks that generate multiple discriminative brain states rather than just two brain states.

# Dimension reduction

- When the dimensionality of data compared to the number of existed examples and information is very high, it will be hard of classification algorithms to extract the relevant information this is called the Curse of Dimensionality.



# Principal Component Analysis



- Principal Component Analysis (PCA) is often mentioned as a technique for reducing the number of variables in a data set without loss of essential information



- PCA works by transforming a set of correlated variables into new set of uncorrelated variables that are called principal components .
- In PCA eigenvalues are ordered in terms of their absolute values to find the dominant eigenvalue of the matrix in which is called the first principle component. Specifically, the first principle component represents the greatest amount of variability in the original data set.



□ Mathematically, PCA resulting from applied linear algebra allows to reduce the dimension of a matrix  $X_{N,M}$  containing multi-channel data. In the case of EEG records each of its column ( $j = 1, 2, \dots, M$ ) represent channel index and its row ( $i = 1, 2, \dots, N$ ) represent the observations. It is possible to convert this matrix into a new one using matrix  $P_{M,M}$  to find values:

$$Y_{N,M} = X_{N,M} P_{M,M}$$



- PCA solves the eigen problem:

$$COV_{X-\bar{X}} v = \lambda v$$

- Where:  $COV_{X-\bar{X}}$  is the covariance matrix of the zero mean data X.

$\lambda$  Is the eigenvalues of the matrix X.

$v$  Is a linear transformation matrix, and the columns of  $v$  is a characteristic vectors, or eigenvectors of the matrix.



## Independent component analysis (ICA)

- PCA has been most successful in applications where data reduction is more important than interpretation of this data. It is commonly used to provide information on the true dimensionality of the data set. And that is because most signals do not have a Gaussian distribution and so are not likely to be independent after they have been decorrelated using PCA. This inability to make signals independent through decorrelation affords the motivation for the independent component analysis methodology.





- ❑ ICA assumes that EEG signals from scalp electrodes represent combinations of underlying independent neural sources.
- ❑ Unfortunately, the EEG recording not only contains the desired signals but also ongoing brain activities beside muscular and ocular artifacts.
- ❑ Therefore, the signal to noise ratio become very low and classification task is difficult.

# Why?

These are neurons. Our brains has hundreds of billions of them.

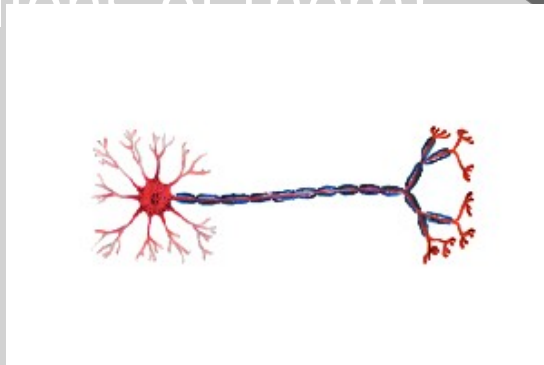
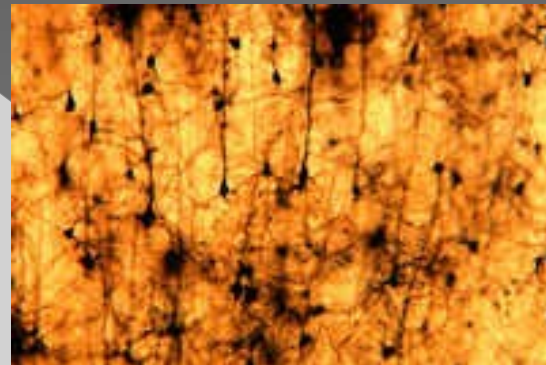
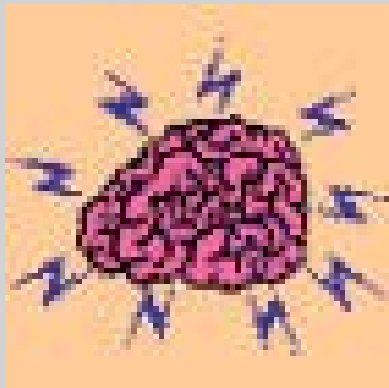


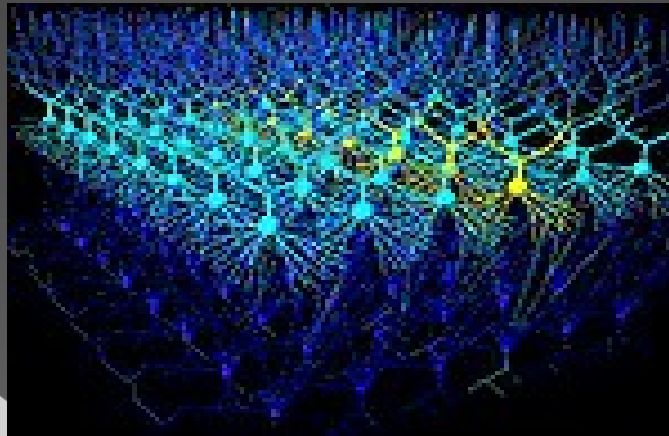
Diagram of a neuron.



A group of real neurons.

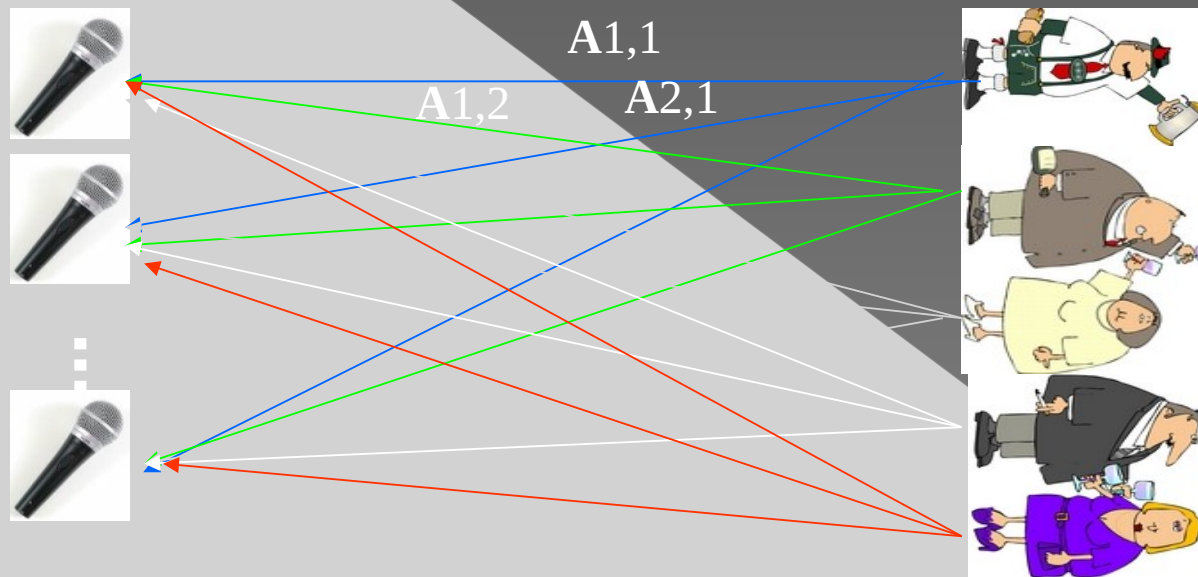
If millions of neurons all fire together, they produce enough electrical activity to be detectable to electrodes placed on the scalp.





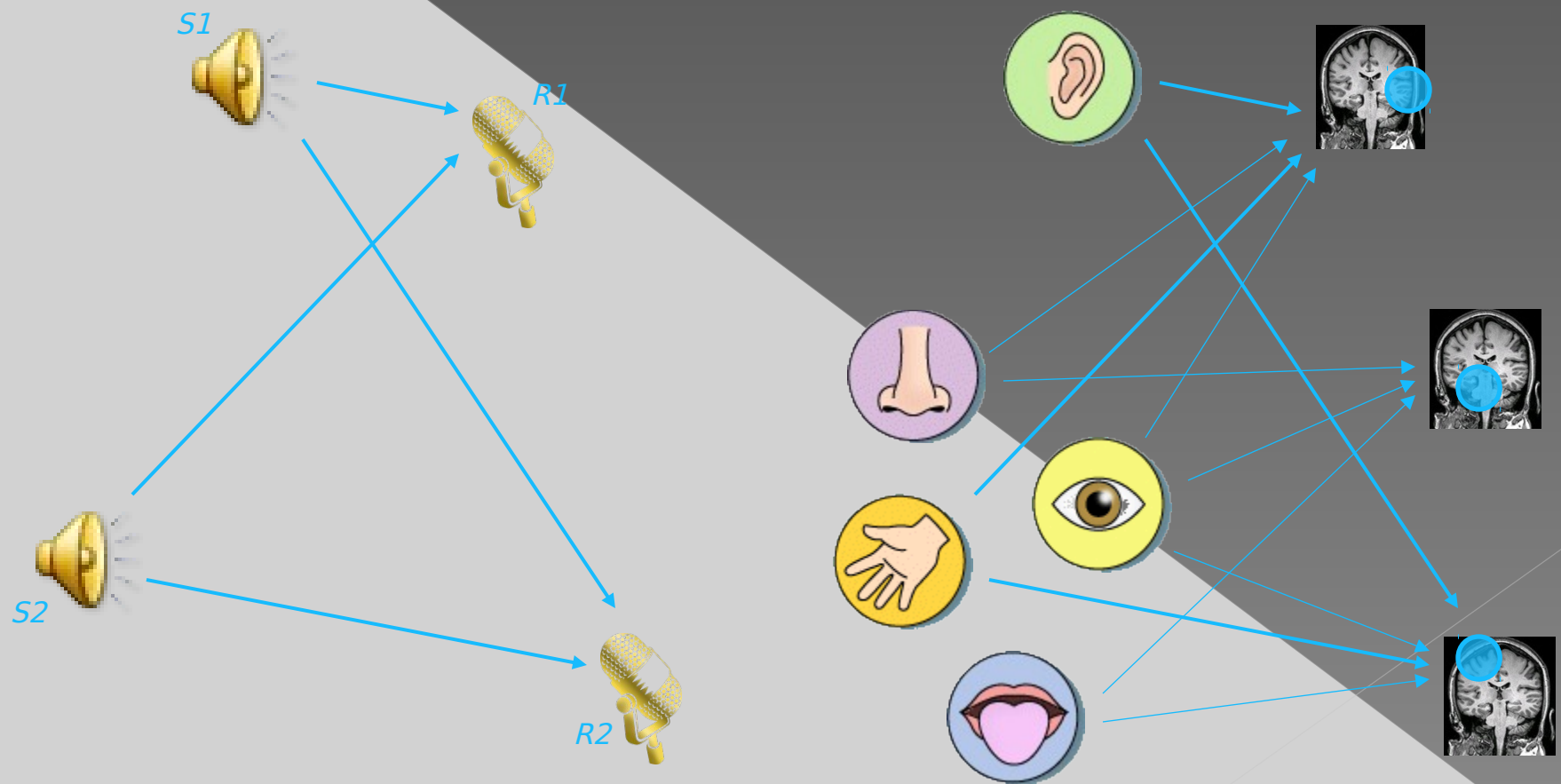
- we have no idea about the effective number of statistically independent brain signals contributing to the EEG recorded from the scalp.
- Thus the main problem in interpreting the output of ICA is the determining of proper dimension of input channels .

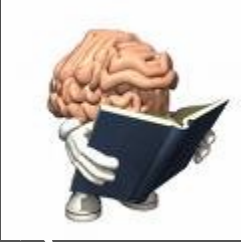
- What distinguishes ICA from other methods is that it looks for components that are both statistically independent, and nonGaussian



The Cocktail Party Problem

# What the ICA do?





- Independent component analysis is a method for solving the blind source separation problem.
- In BCI systems ICA is used to separate multichannel EEG into several components corresponding to neural sources inside the brain or noise.





□ If we consider a random source vector  $S(n)$  is defined by:

□ 
$$S(n) = [S_1(n), S_2(n), \dots, S_m(n)]^T$$

□ Where:  $s$  is a set of independent sources from 1 to  $m$ .

$n$  denotes discrete time.

□ And a nonsingular  $m$ -by- $m$  matrix  $A$ , is called mixing matrix, then the relation between  $X(n)$  which represents a vector of the measured signals and  $S(n)$  which represents a vector composed of all source signals, will be as follows:

□ 
$$X(n) = AS(n)$$



□ The source vector  $S(n)$  and the mixing matrix  $A$  are both unknown. The task of blind source separation is to solve a mixing matrix,  $A$ , from which independent components,  $s$ , can be recovered through simple matrix inversion:

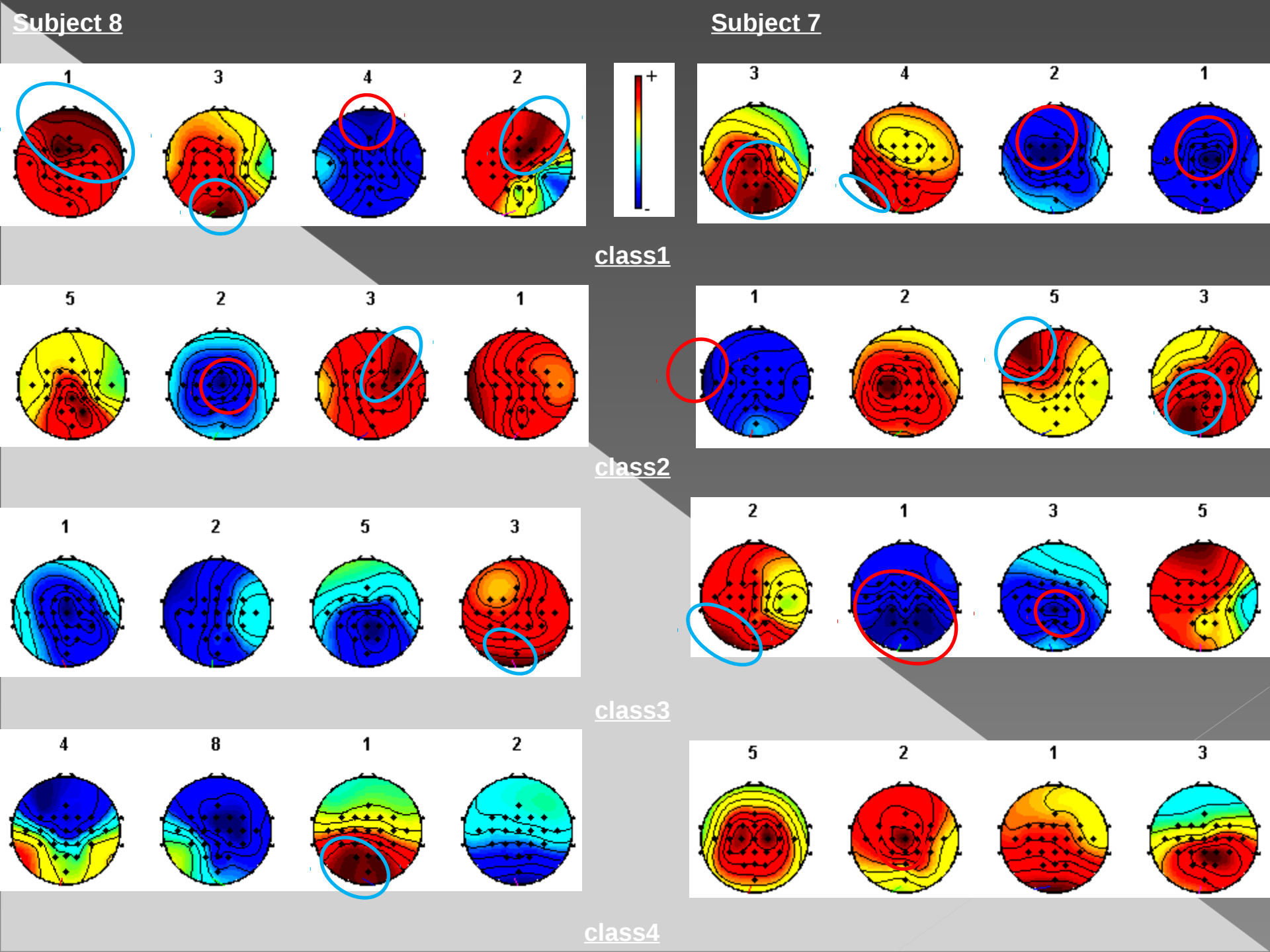
$$S(n) = A^{-1} X(n)$$



- The ICA method is based on the two assumptions: the source variables,  $s$ , are truly independent, and that they are non-Gaussian. The objective of an ICA algorithm is to find a demixing matrix,  $A^{-1}$ , such that components of  $S$  are statistically independent. We assume that the multichannel EEG can be modeled by  $X(n) = AS(n)$ , where  $X(n)$  is the recorded multichannel EEG at time  $n$ ,  $A$  is the mixing matrix, and  $S(n)$  is the source vector at time  $n$ .



- There are many algorithms to implement ICA such as FastICA, PearsonICA , Jade, MLICA...etc.
- In this work Jade MATLAB algorithm is used to calculate ICA components of the EEG signal.



# Feature Classification and Pattern Recognition



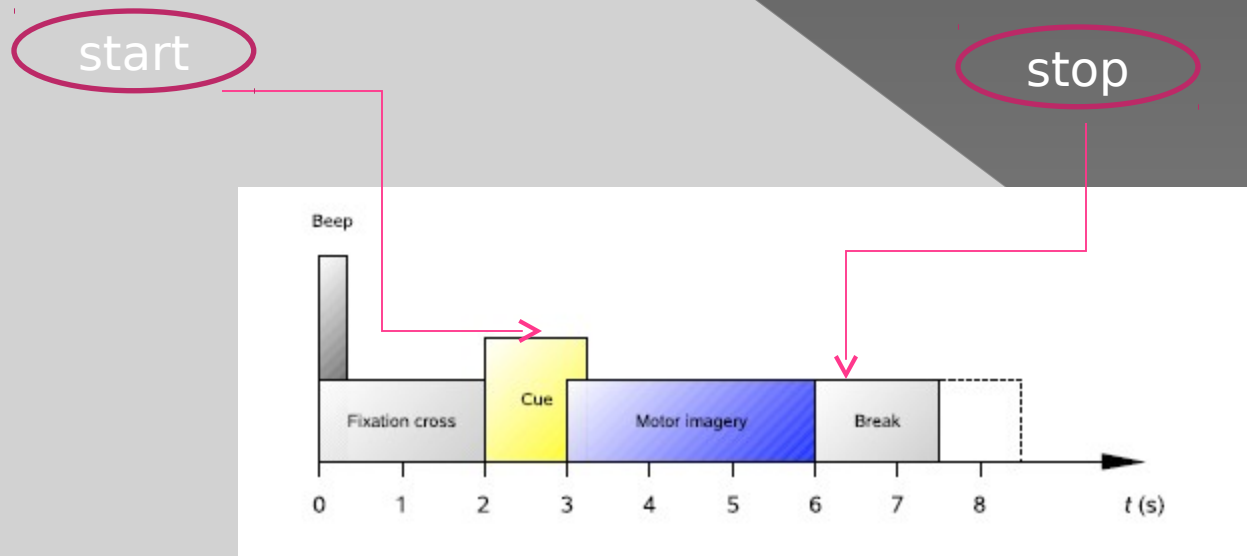
- classification is a supervised or unsupervised machine learning procedure in which individual items are placed into groups based on quantitative information on one or more characteristics inherent in the items (referred to as traits, variables, characters, etc.) and in case of supervised machine learning algorithm they are based on a training set of previously labeled items .
- Specifically, classification is a process of describing observations, relying on the extracted features.



- The classification is usually based on the availability of a set of patterns that have already been classified or described.” This set of patterns is termed the training set, and the resulting learning strategy is characterized as supervised learning. Learning can also be unsupervised, in the sense that the system is not given an a priori labeling of patterns, instead it itself establishes the classes based on the intrinsic regularities of the patterns.

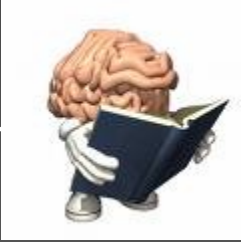


- the chosen interval for each trial started from 0.5 sec. after the onset of the visual cue and continued for 1.25 sec. Algorithm was trained on train\_time\_segment and evaluated on test\_time\_segment.





- Support Vector Machine is widely used in BCI field as a result of it is powerful algorithm for pattern recognition specifically for high-dimensional systems. It is early developed by Vapnik (1996) since then it has strongly performed in a large number of real world problems, including BCI.
- SVM model is a representation of the samples as points in space, mapped so that the samples of the separate categories are divided by a clear gap that is as wide as possible. New samples are then mapped into the same space to predict on which side of the gap they fall on

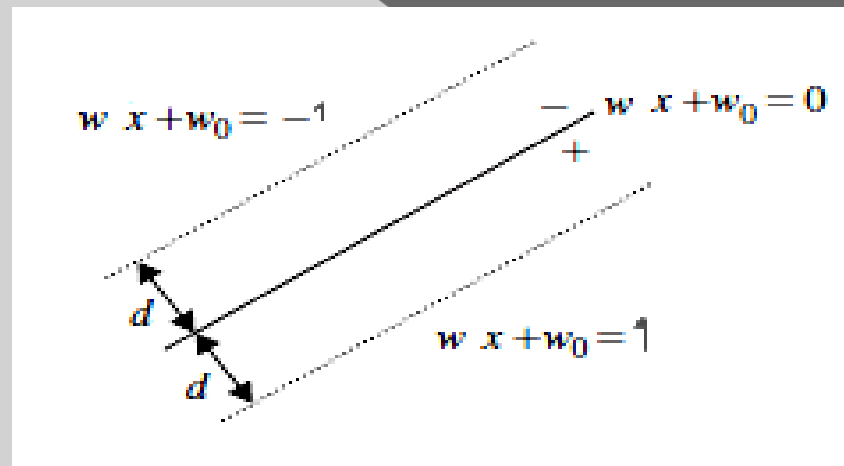


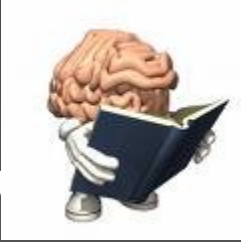
- In more specific meaning, support vector machine for a hyper plane or set of hyper-planes in a high dimensional space, which can be used for classification, regression or other tasks.” Intuitively, a good separation is achieved by the hyper-plane that has the largest distance to the nearest training data-points of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.”



# Mathematical estimation

- To represent the linear classifier mathematically; line and margin below must be specified:





- Consider the discriminant function  $f(x)$  of a linear class which can be defined by:  $f(x, w, w_0) = (w \cdot x) + w_0$ , the input vector  $x$  is assigned to class  $y \in \{1, -1\}$ , where  $y$  is a desired outputs, as follows:

$$f(x) = \begin{cases} 1 & \text{if } f(x) = (\omega \cdot x + \omega_0) + b \geq 0 \\ -1 & \text{if } f(x) = (\omega \cdot x + \omega_0) + b < 0 \end{cases}$$

- Then the plus plane in figure3.12 can be represented by equation:

$$\square \quad (w \cdot x) + w_0 = +1$$



- And the minus plane can be represented by equation:

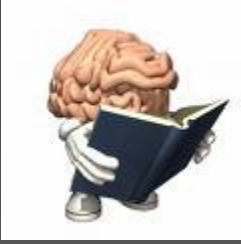
$$(\mathbf{w} \cdot \mathbf{x}) + w_0 = -1$$

- Let  $\mathbf{x}_-$  be any point on the minus plane, and  $\mathbf{x}_+$  be the closest plus plane point to  $\mathbf{x}_-$ . then the relation between them can be defined by the following equation:

$$\mathbf{x}_+ = \mathbf{x}_- + \lambda \mathbf{w}$$

- Where  $\lambda$  is a constant value.
- The margin  $M$ ,  $M=2d$  in the previous figure, which represents the distance between the minus and plus planes can be defined by the equation:

$$|\mathbf{x}_+ - \mathbf{x}_-| = M$$



Then, by substituting  $x_+$  in equation :

$$w \cdot x_+ + w_0 = +1$$

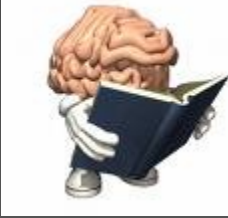
And by substituting  $x_-$  in equation :

$$w \cdot x_- + w_0 = -1,$$

By rearranging equations and we can prove the following:

$$\lambda w = M$$

□ then:



□  $\lambda = 2/w \cdot w$

□ Thus, we conclude:

□  $M = 2/\sqrt{w \cdot w}$

□ So by estimating  $w$  and  $w_0$  all points belongs to which plane can be computed and also the margin width  $M$ .

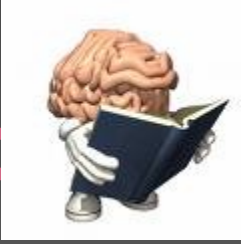
Therefore, the criteria is to minimize:

$$(w^*, w_0^*) = \arg_{w, w_0} \frac{1}{2} \min w \cdot w$$

□ Then, estimate the values of  $w$  and  $w_0$  which lead to the widest margin that matches all data points.



# Extending binary classifiers to solve multi class problem



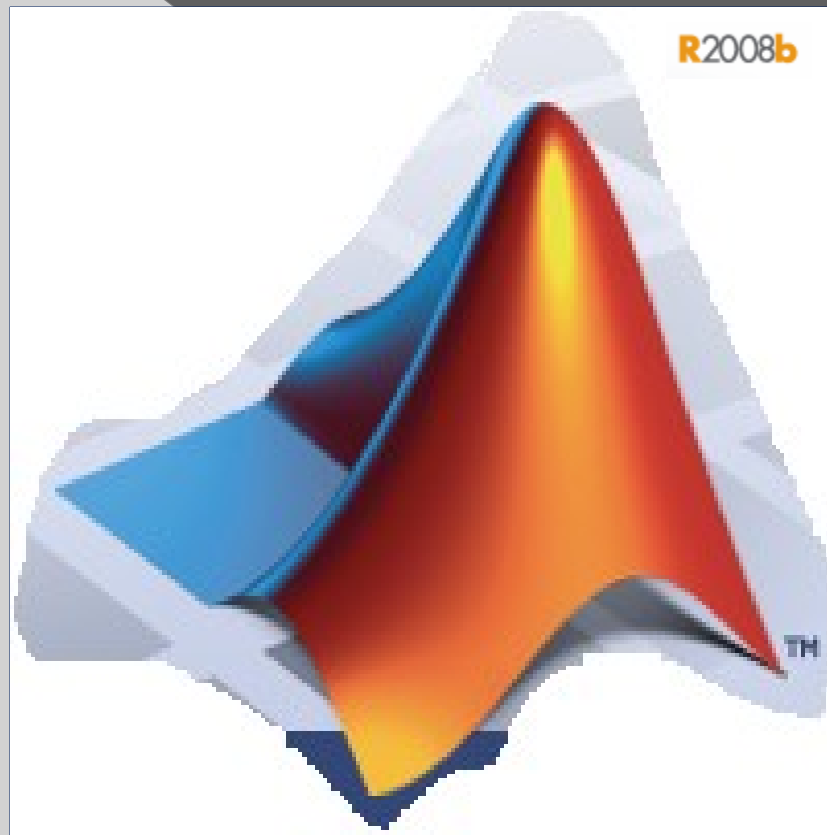
- There are two approaches to extend binary classifiers to solve a M-class problem:
- The one-versus-rest and the one-versus-one schemes . In the one-versus-rest scheme M binary classifiers are designed by training the  $i$ th classifier through labeling the samples of the  $i$ th class as positive and the remaining samples as negative, while in one-versus-one approach one classifier is designed for each pair of classes.
- In total:



$$\binom{k}{2} = \frac{k!}{(k-2)! \cdot 2!} = \frac{k \cdot (k-1)}{2}$$

- For both approaches, the class of a test sample is labeled by the class with the most numbers of votes. If contradictions arise in the voting process, the test sample can either be rejected or assigned to the class with largest prior probability .
- In this project the one-versus-one approach was used, following the assumption that it may be more suitable for practical use.

# MATLAB implementation



# Results & Discussion

- Results of classification phase:
- Session One: Table 1

	Right				
	hand	Left hand	Foot	Tongue	OA
Subject1	0.96	1	0.97	0.92	0.96
Subject2	0.98	0.97	0.97	0.91	0.96
Subject3	0.97	0.99	1	0.96	0.98
Subject4	0.96	0.98	0.97	0.90	0.95
Subject5	0.97	0.98	0.99	0.93	0.97
Subject6	0.96	0.98	1	0.93	0.97
Subject7	0.99	0.99	0.99	0.98	0.99
Subject8	0.97	0.98	0.98	0.93	0.97
Subject9	0.98	0.99	0.99	0.95	0.98
Class					
Acc.	0.97	0.98	0.98	0.93	0.97

## Session two: Tabel 2

	Right				
	hand	Left hand	Foot	Tongue	OA
Subject1	0.26	0.37	0.25	0.34	0.31
Subject2	0.26	0.26	0.30	0.31	0.28
Subject3	0.33	0.30	0.32	0.51	0.37
Subject4	0.38	0.27	0.35	0.35	0.34
Subject5	0.21	0.25	0.44	0.31	0.30
Subject6	0.26	0.30	0.39	0.26	0.30
Subject7	0.27	0.27	0.29	0.32	0.29
Subject8	0.33	0.26	0.39	0.30	0.32
Subject9	0.36	0.27	0.30	0.35	0.32
Class					
Acc.	0.30	0.28	0.34	0.34	0.31



- In our study we applied PCA in order to select just four channels before performing ICA. Furthermore, the great advantage of PCA is that it reduces the size of the data which makes learning of classifiers computationally more efficient. JADE (ICA) algorithm was used to separate two components from each trial, meaning it has calculated 288 (number of trials) for each data set.
- right hand can activate the same areas of left hand in the brain, while foot can activate the same areas of hands and tongue . Thus we did not have high expectations for separating the same signal sources in both sessions. And this also explains why we did not get the same accuracy in both sessions.
- Table 1 shows the results of discrimination of four classes when evaluated the classification algorithm by the training data (session one) instead of testing data (session two). However we didn't reject the trials marked as artifact, and obtained excellent accuracy of  $97 \pm 2\%$ .



- Table 2 shows the results of discriminating the four classes in session two which gives average accuracy around  $(31 \pm 4\%)$ .
- The best prediction accuracy (51%) happened when subject 3 performed tongue MI, and the least accuracy (21%) when subject 5 performed right hand MI.
- We evaluated all results by using Kappa function which estimates Cohen's kappa coefficient and related statistics. The specific accuracy (sACC) and the overall agreement (accuracy) were considered in our evaluation.
- Observed variability in the result may happen due to differences in the shape and position of MI with respect to skull, changing the association between channel location and MI, which making channels more or less effective at detecting MI activity, or it may have due to differences in the amount and location of cortical activity experienced when preparing for imagination of movement in such class.

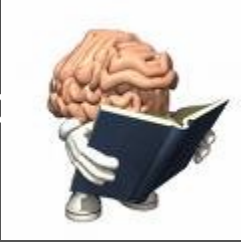
# conclusion



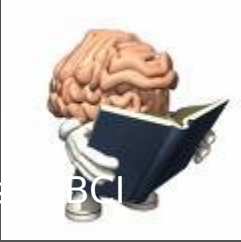
- This project investigated the classification of 4 different classes from 9 subjects who performed left hand, right hand, foot and tongue motor imageries (MI) using EEG measurements via a proposed PCA and ICA for pre-processing and feature extraction phase and SVM in classification phase. The study is performed on data set 2a from BCI competition IV (provided by the Graz group). The results exposed that the proposed techniques were successfully classified 4 different MI from EEG measurements. The results also demonstrated that using PCA could help in



reducing the size of the data and in reducing the time of classification which leads to help in using BCI system online. From the concept of the representation of Motor Signals in the human Brain, there is a motor representation of hands, tongue and foot in the motor cortex. Evidence in this project using EEG-LAB toolbox also illustrated that the motor cortex is activated during voluntary imagination of movements. The decrease in session two accuracy results suggested that there are relevant sources and signals related to the imagination of the four classes.

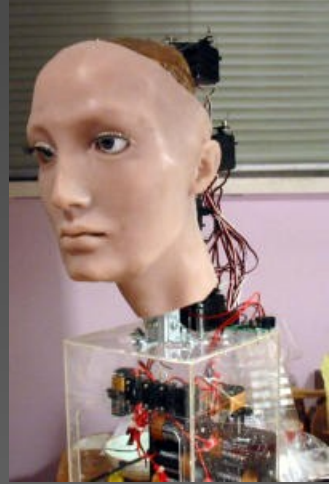


# Recommendations



- The limitation of applied technique in this project to classify the four classes in EEG-based BCI system could be improved in future works by the following recommendations:
- To readjust the selected segment from EEG data and chose it to start from 0.5 to 2.5 second after the onset of the visual cue.
- Try to Design two linear SVM as one-against-one classifiers and combine the classification outputs together, then, observe the change in accuracy.
- Try to use a compound technique: PCA in pre-processing phase and CSP (common spatial pattern) in feature extraction, or use Multi-class FBCSP (Filter Bank Common Spatial Pattern).
- The limitation of this thesis and such BCI study is that it provides only offline analysis and classification of EEG data, and online capacity still has to be achieved.

- Assuming the result of classification the four classes could be enhanced and improved, this BCI system could greatly impact lives of patients suffering from amyotrophic lateral sclerosis or spinal cord injuries and help them to communicate with outer world, to control mechanical and even entertainment devices.





**Nobody can look  
into the future; but  
we can invent it!!**

QUESTION?

