EEG Analysis for Brain-Computer Interfacing and Other Applications

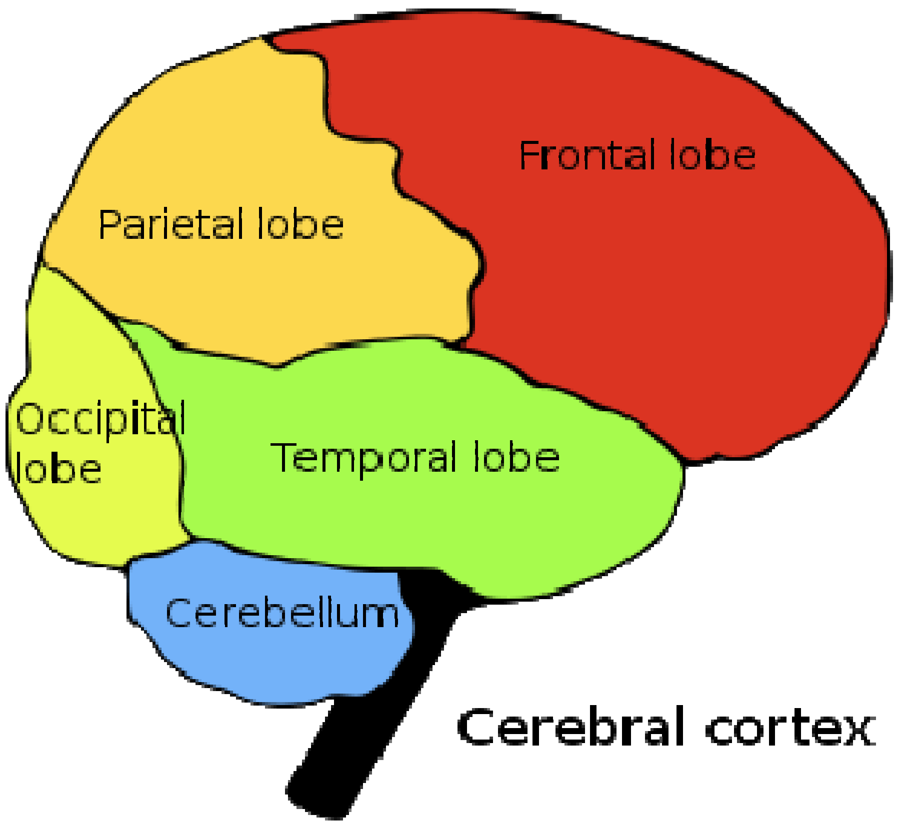
John Q. Gan
School of Computer Science & Electronic Engineering
Essex University
Colchester, UK
http://dces.essex.ac.uk/staff/jqgan

Outline

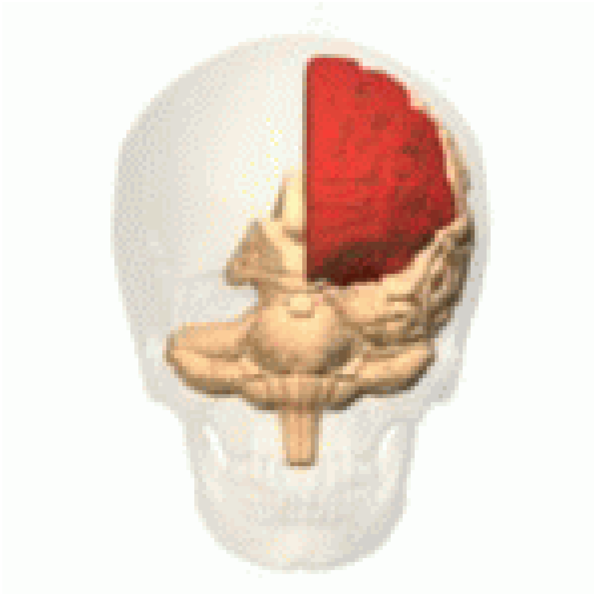
I. Neurophysiologic Background
II. Pattern Recognition Approach to EEG Data Analysis
III. EEG-based BCI Design
IV. On-line BCI Systems at Essex
V. Other Applications

I. Neurophysiological Background

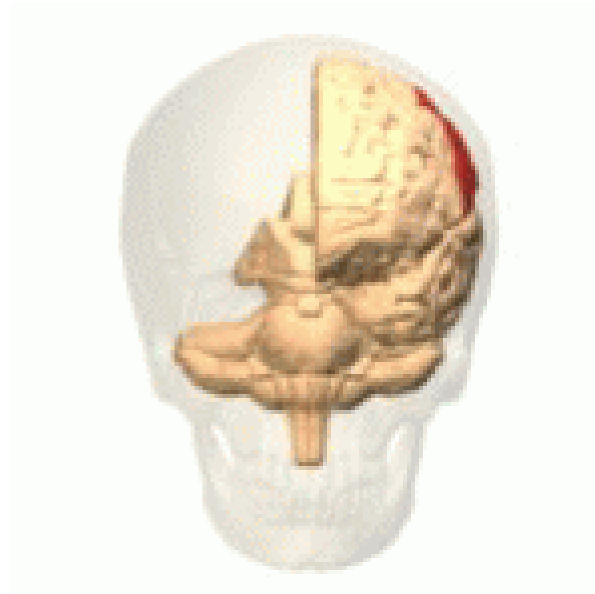
- Brain Regions



Frontal_lobe

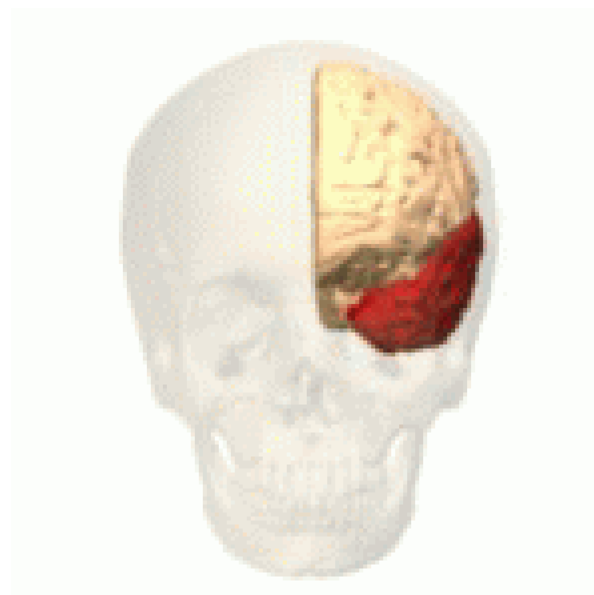


Parietal_lobe

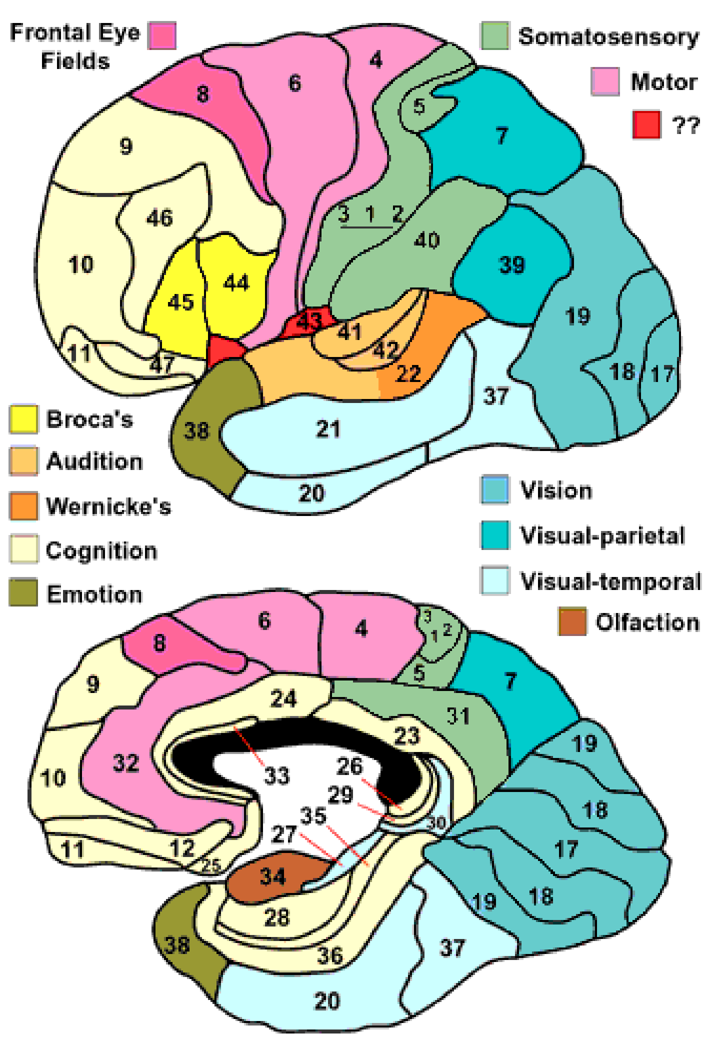


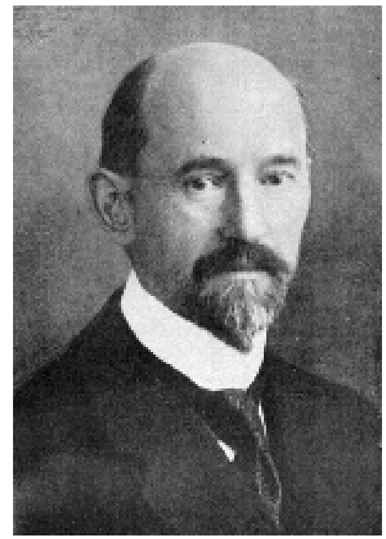
Occipital_lobe

Temporal_lobe



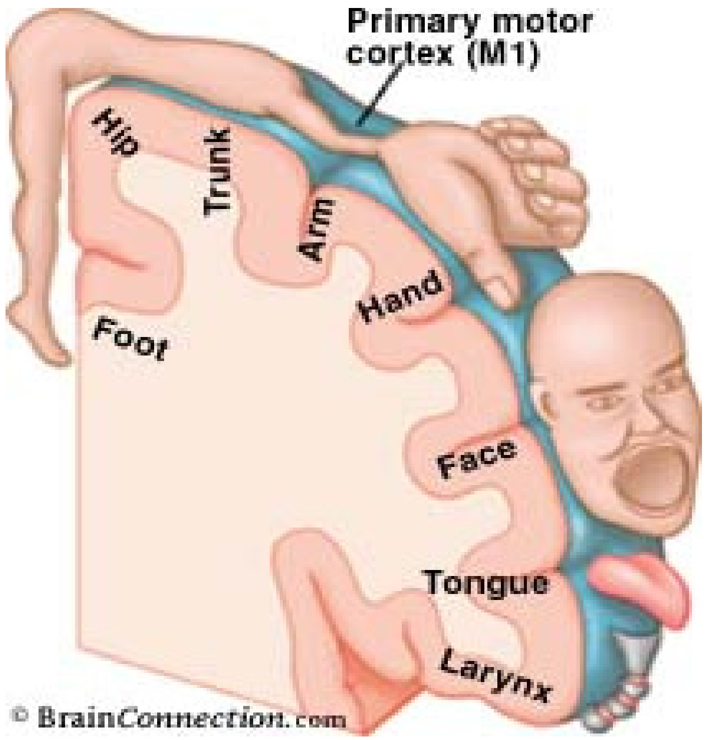
➢ Brodmann Areas



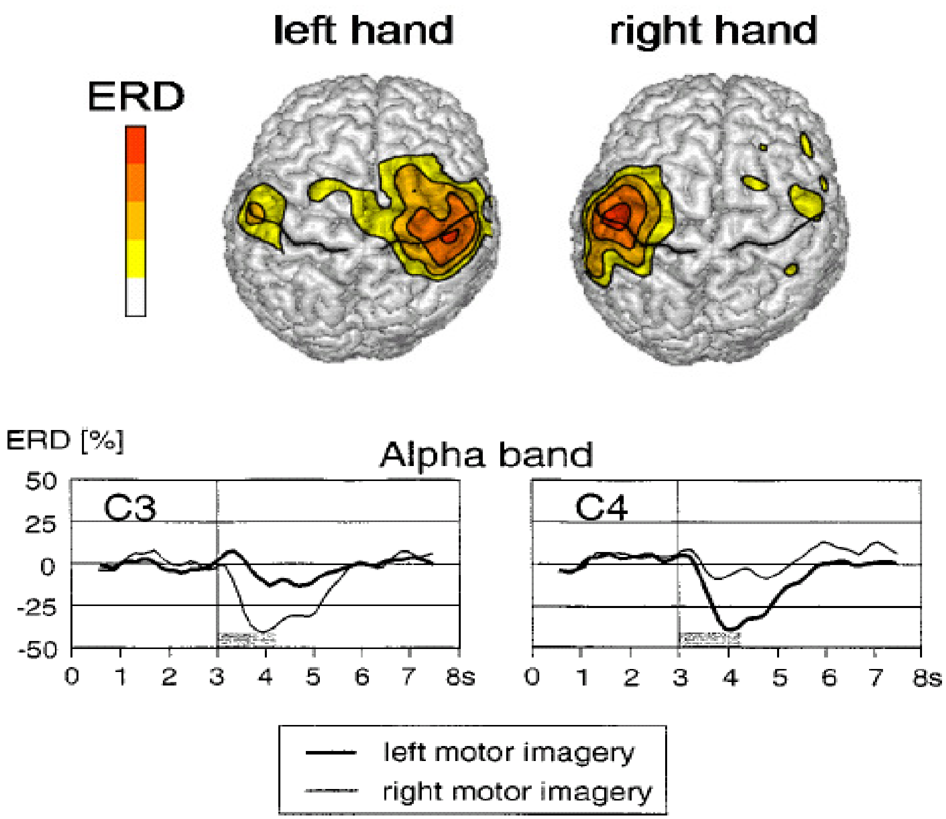


Korbinian Brodmann
(1968-1918)

➢ Motor Homunculus



➢ ERD/ERS



[Pfurtscheller et al. 2003]

➢ Measuring Brain Activity

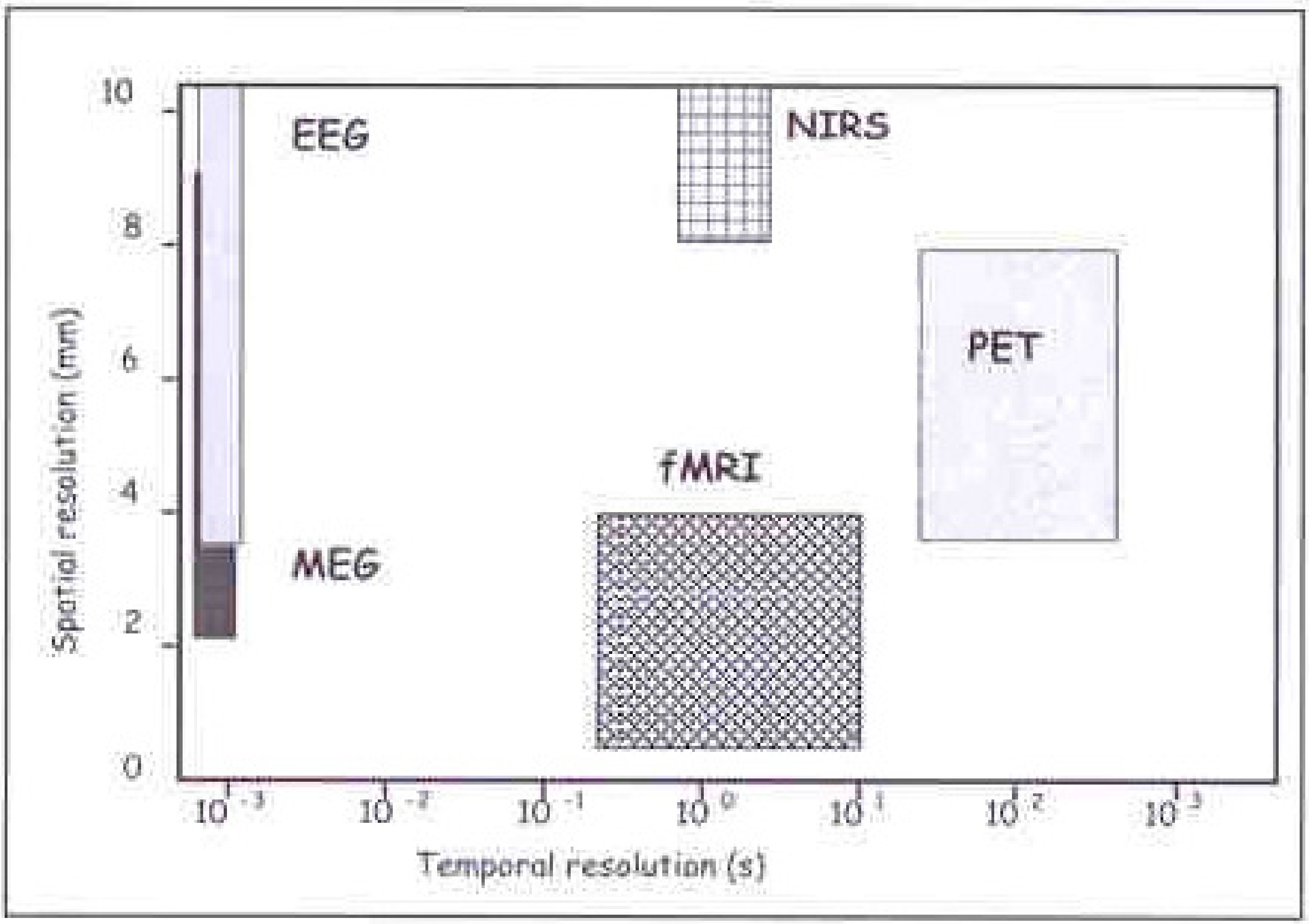
- Invasive:
  - Implanted systems ➔ *risk, cost, durability problems*
  - Positron emission tomography, PET ➔ *radiation, cost, slow response*
- Non-invasive:
  - Functional MRI ➔ *large equipment, cost, slow response*
  - Near-Infrared Spectroscopy ➔ *slow response, long term effects unknown*
  - Magneto-encephalogram, MEG ➔ *large equipment, cost*
  - Electroencephalogram, EEG ➔ *limited resolution, but*
    - *low cost*
    - *fast response (i.e., short latency events can be seen)*
    - *portable*

Neural activity →	Changes in … →	Changes in …	
	Changes in blood hemoglobin concentrations	Changes in reflected NIR light intensity	NIRS
	Changes in oxygen levels in blood	Changes in blood magnetic properties	fMRI
	Changes in neuronal action potentials	Changes in magnetic fields produced by …	MEG
	Changes in neuronal action potentials	Changes in scalp electric potentials	EEG

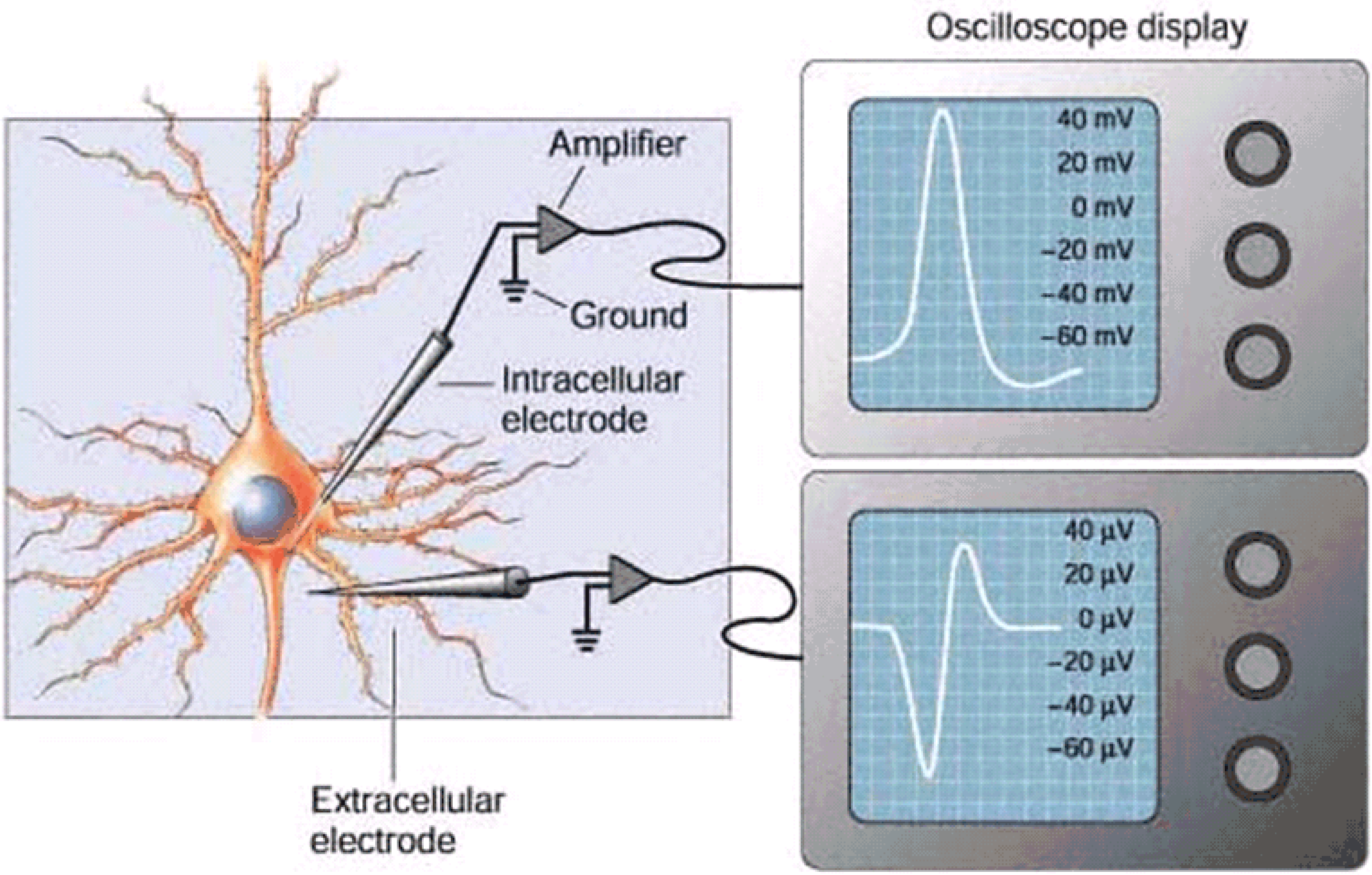
Neuro-imaging: An inverse problem

Which areas in the brain are activated by a stimulus or a mental task?

Comparison of Resolutions of EEG, MEG, fMRI, NIRS, PET



➢ Neurons and Action Potentials



➢ Neurons and Action Potentials

Stimuli (e.g. Action potentials from other neurons

↓

Neuro-transmitter release

↓

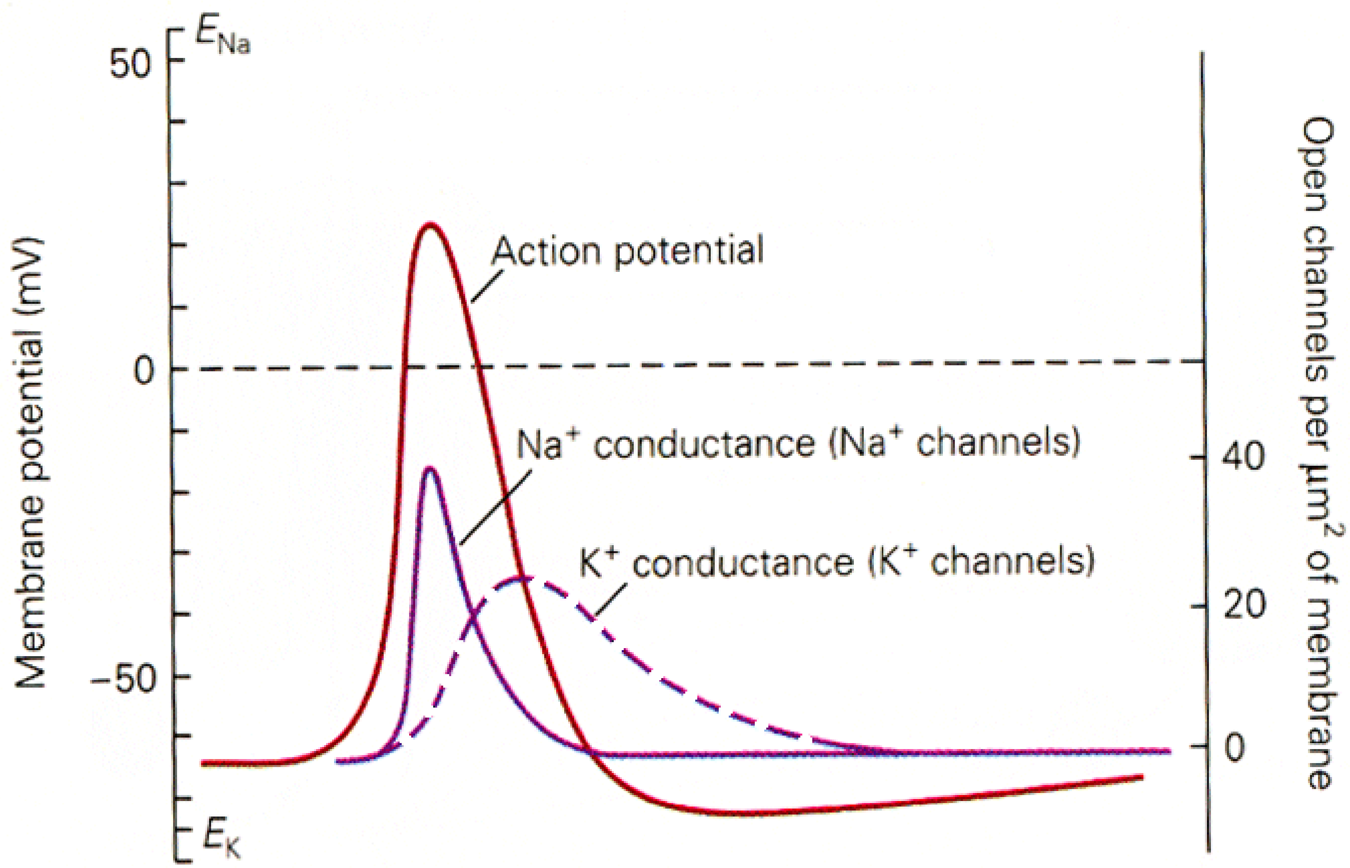
Flow of ions

↓

Membrane potential change

↓ If …

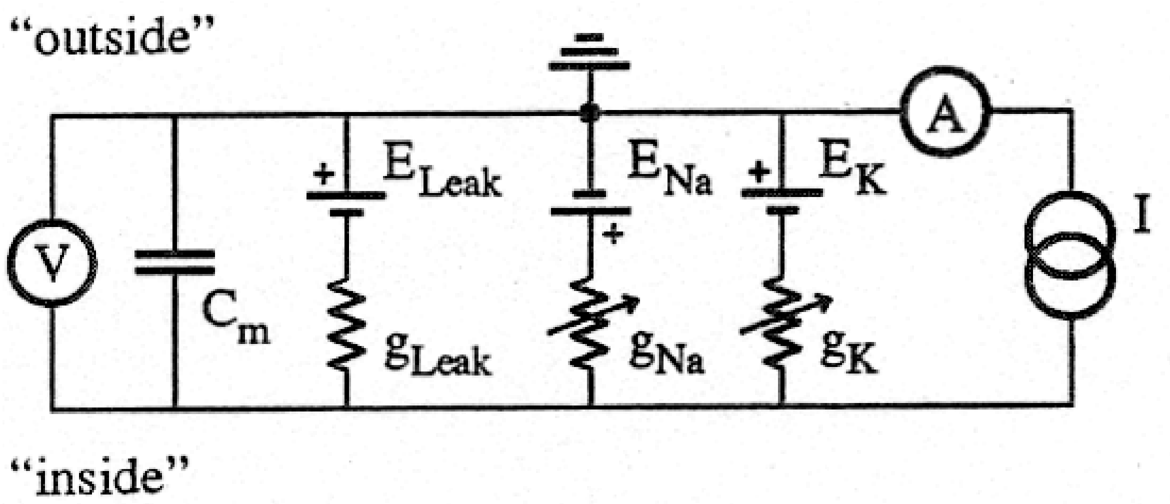
Action potential



(from Kandel, et al., 2000)

➢ Models for Membrane Potentials

Hodgkin-Huxley membrane model



The Nernst equilibrium potential (for single ions only)

$$E_X = (RT/ZF)\ \ln\ \{[X]_1/[X]_2\}$$

$[X]$ = ion X concentration in moles/cubic meter;

R (8.31 joules/Kelvin/mole) is the ideal gas constant;

T (293^{o}K at 20^{o}C) is the temperature in Kelvin;

F (96400 coulombs/mole) is the Faraday's constant;

Z is the valence of the permeant ions.

e.g., K^{+} $(Z=+1)$

$$E_K = 58 \log \{[K^+]_{out}/[K^+]_{in}\} \text{ (mV)}$$

For multiple ions, resting membrane potential:

$$E_m = \frac{RT}{F} ln \frac{P_K\left[K^+\right]_{out} + P_{Na}\left[Na^+\right]_{out} + P_{Cl}\left[Cl^-\right]_{in}}{P_K\left[K^+\right]_{in} + P_{Na}\left[Na^+\right]_{in} + P_{Cl}\left[Cl^-\right]_{out}}$$

Hodgkin-Huxley action potential model

$$C_m \frac{dV}{dt} = I_{ext} + I_{Na} + I_K + I_L$$

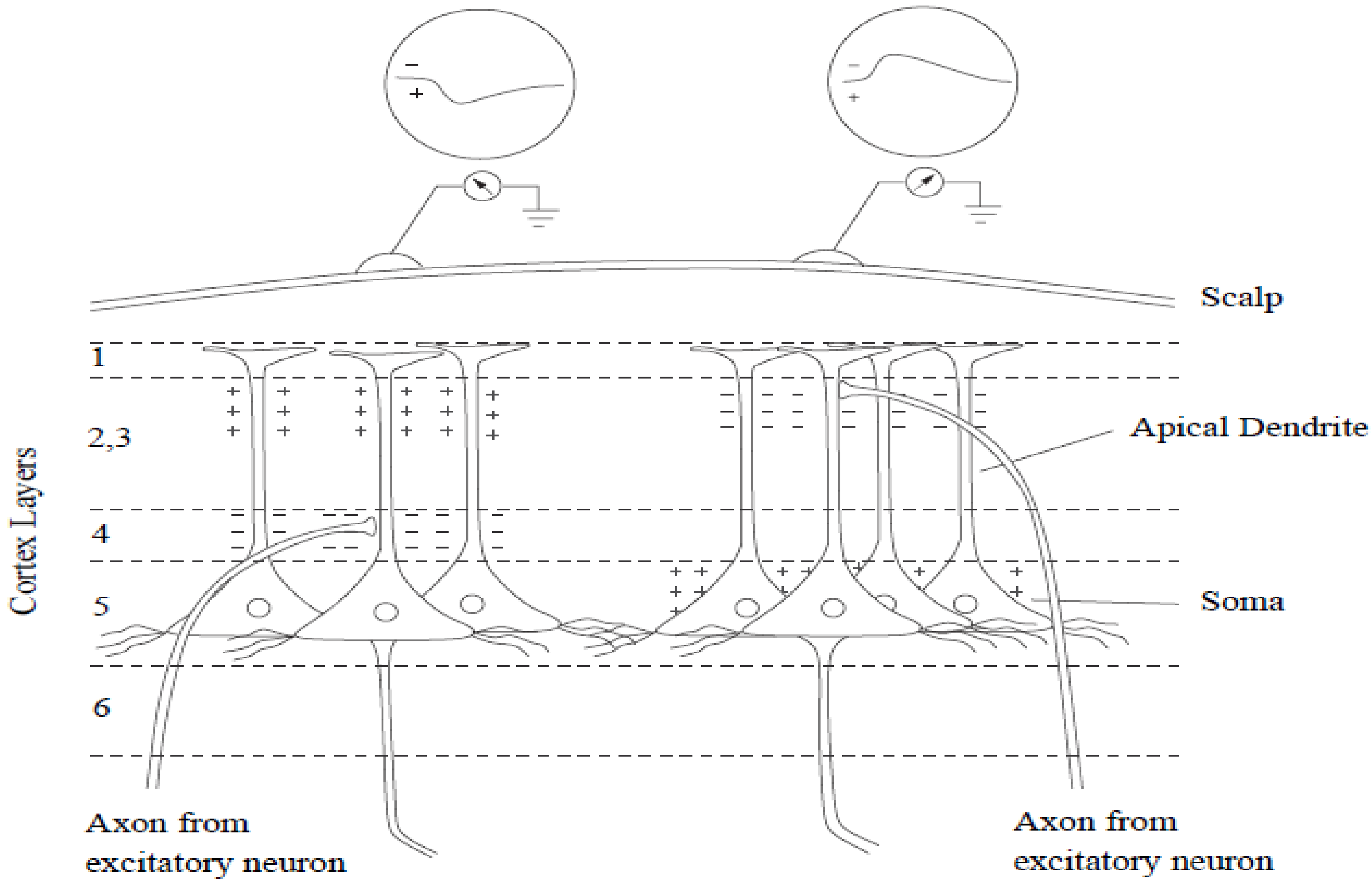
$$I_{Na} = g_{Na}(V - E_{Na})\, p_{open,Na}$$

$$I_K = g_K(V - E_K)\, p_{open,K}$$

$$I_L = g_{Leak}(V - E_{Leak})$$

$p_{open,Na}$ and $p_{open,K}$ are probabilities of ion channels being open, which are assumed to obey first - order kinetics.

➢ EEG



Generation of extra-cellular voltage fields from graded synaptic activity (J.H.Martin 1991)

➢ EEG

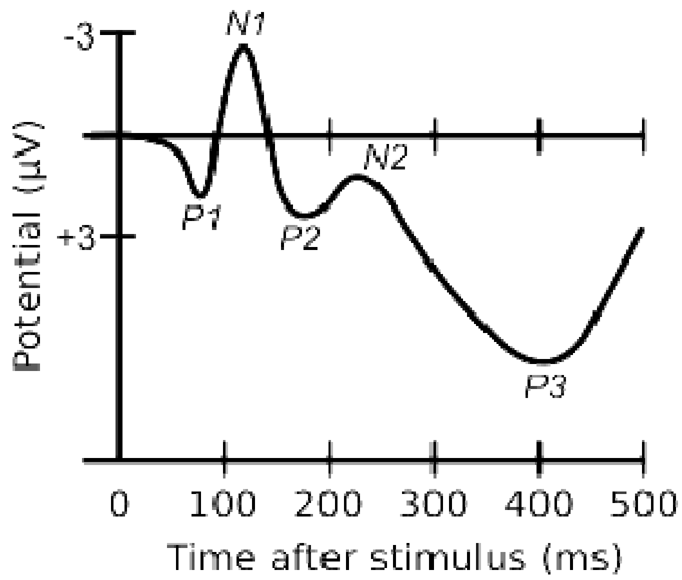
- EEG measures the current flow during synaptic excitation of the dendrites of pyramidal neurons in the cerebral cortex.
- EEG is a result of joint activity of millions of underlying neurons activated together.
- The amplitude of the EEG signal is proportional to the number of *synchronously* activated neurons.
- The EEG signal is "blurred" version of a real activity, as signal passes through several layers of non-neural tissue (meninges, fluid, skull, skin)

➢ Rhythms of Spontaneous EEG

- delta (<4 Hz): associated with deep sleep, brain disorders.
- theta (4~7 Hz): associated with drowsiness and sleep, stress.
- alpha (8~13 Hz): associated with visual relaxation while awake.
- mu (8~13 Hz): associated with motor relaxation while awake.
- beta (14~30 Hz): normally in sleep, especially in infants and young children. In BCI, it is usually associated with mu rhythms.
- gamma (30~80 Hz): associated with perception and consciousness.

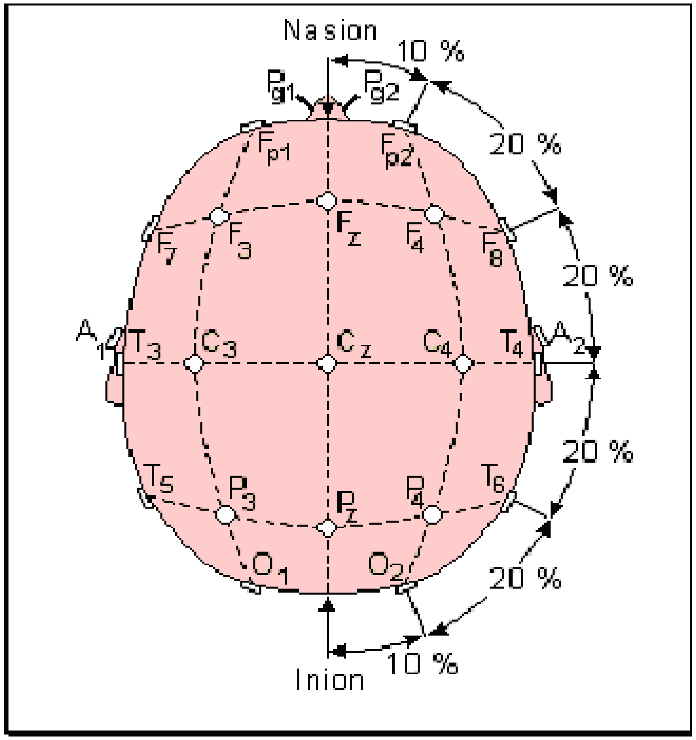
➢ ERP/EP

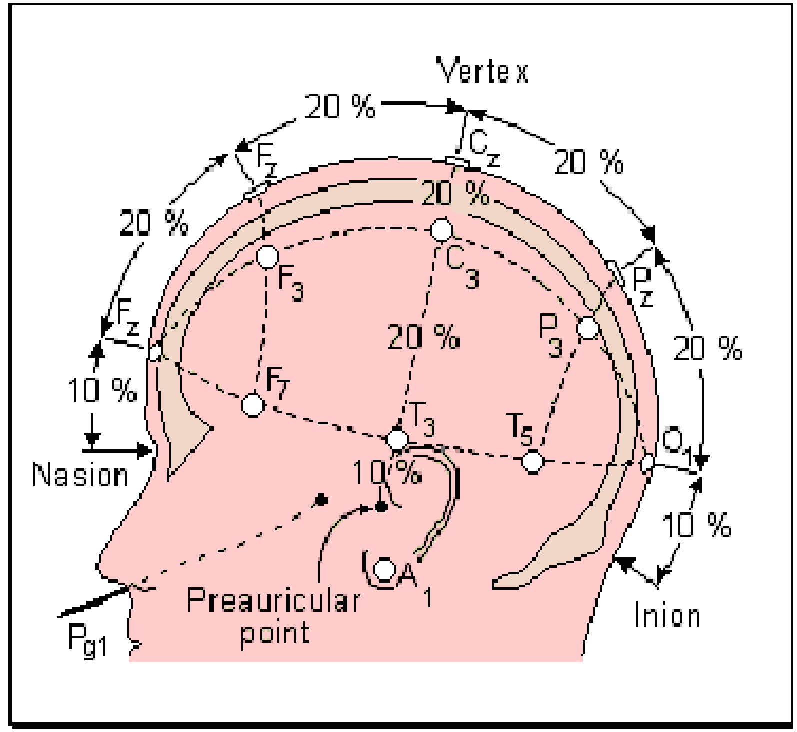
An event-related potential (ERP) is any measured brain response that is directly the result of a thought or perception.



Steady State Visually Evoked Potentials (SSVEP) are signals that are natural responses to visual stimulation at *specific* frequencies. When the retina is excited by a visual stimulus ranging from 3.5 Hz to 75 Hz, the brain generates electrical activity at the same (or multiples of) frequency of the visual stimulus.   [From Wikipedia]

➢ International 10-20 system of EEG electrode placement

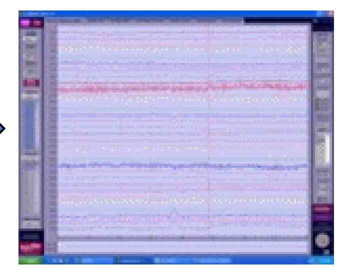




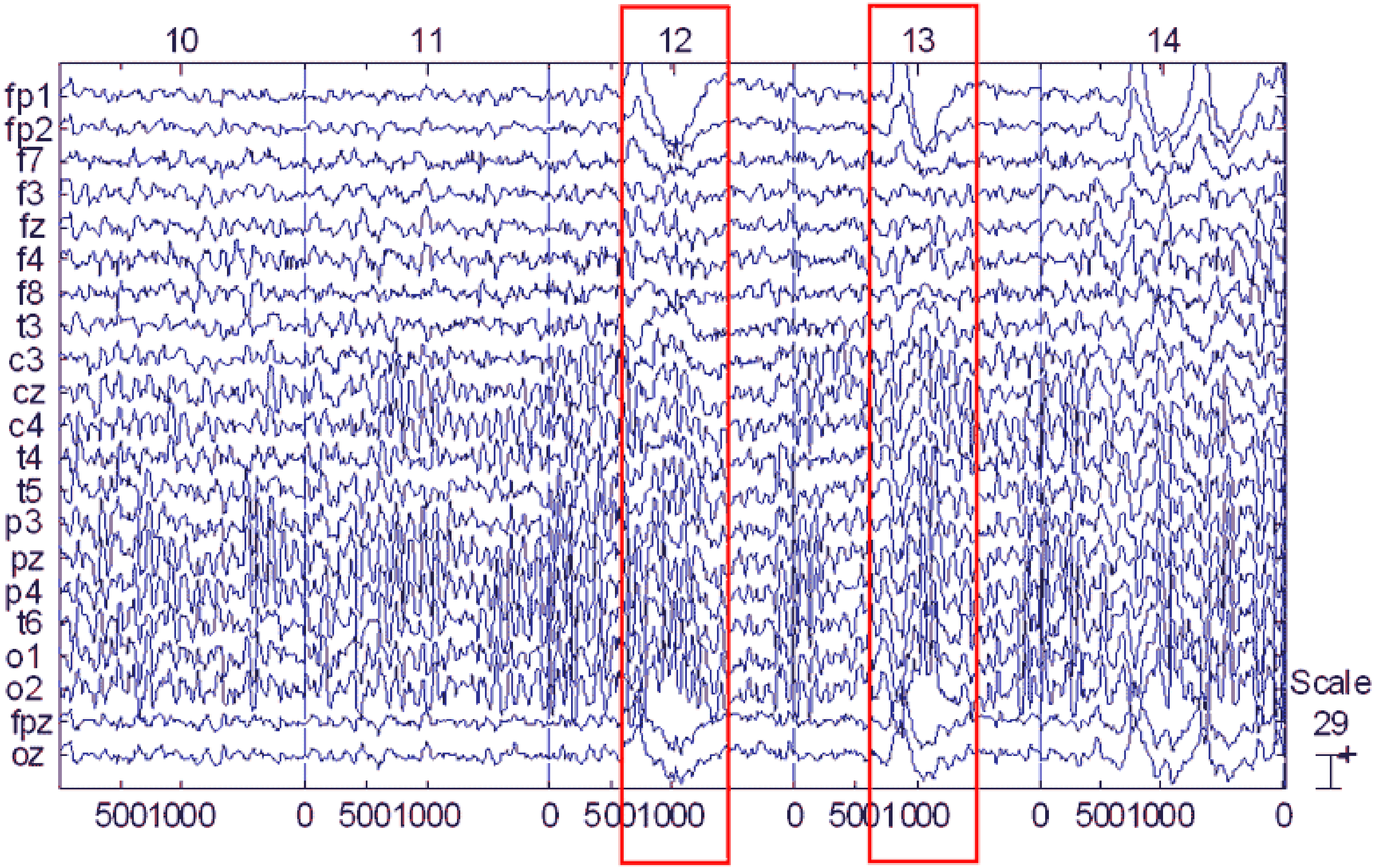
The "10" and "20" refer to the 10% or 20% inter-electrode distance. A - Ear lobe, C - central, Pg - nasopharyngeal, P - parietal, F - frontal, Fp - frontal polar, O – occipital, T - temporal.
[Malmivuo and Plonsey, 1995]

➢ Biosemi System for EEG Recording

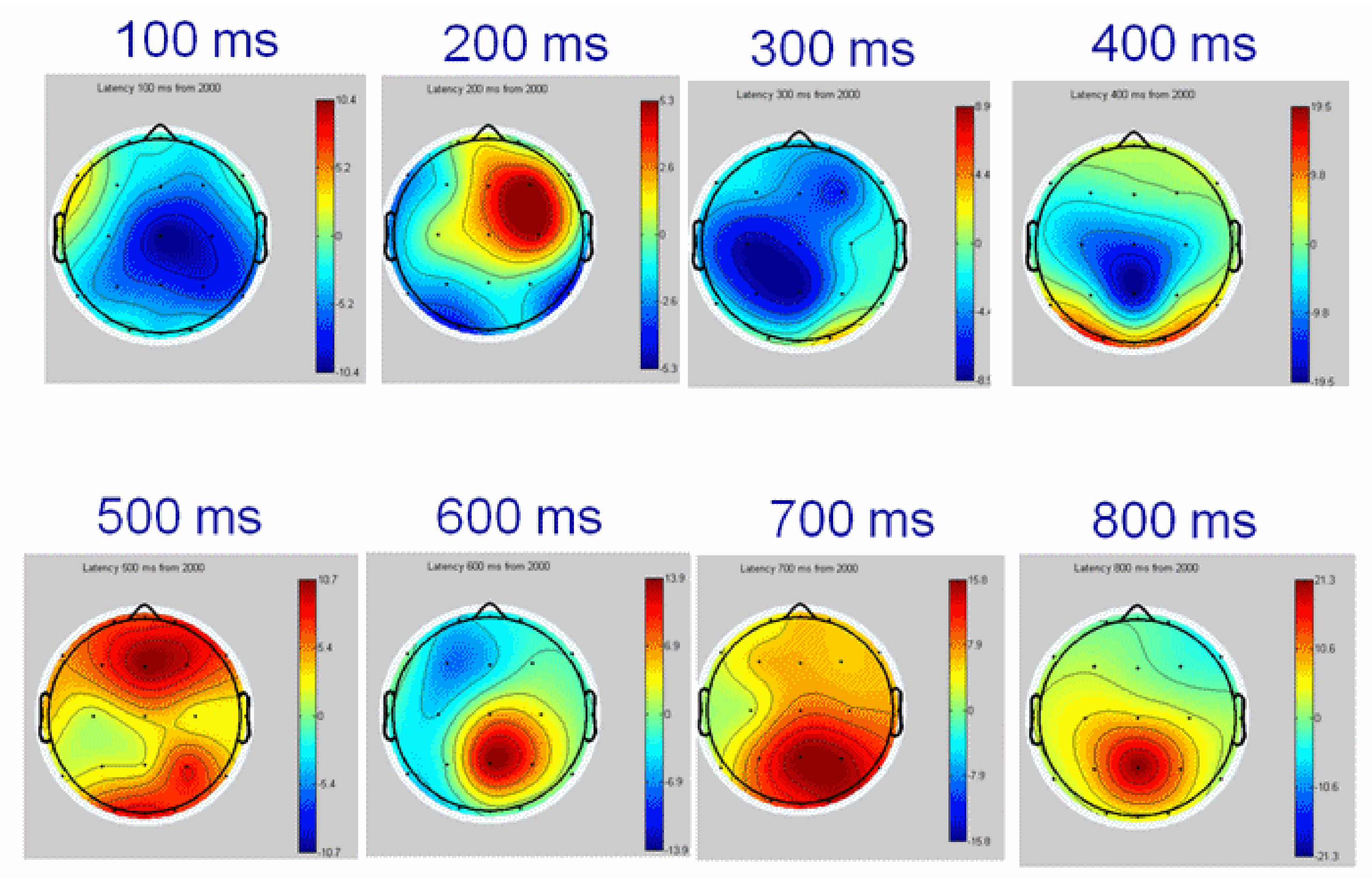
- Amplifier-A/D converter-PC



➢ EEG – Motor Imagery Example



➢ Time-Space Visualization of EEG



EEG of imaginary right hand movement

➢ EEG-based Neuroimaging (EEG inverse problem)

The forward equation:

$$\mathbf{\Phi} = \mathbf{K}\mathbf{J} + c\mathbf{1}$$

$\mathbf{\Phi} \in R^{N_E \times 1}$: scalp electric potentials

$\mathbf{J} = [\mathbf{J}_1^T \ \mathbf{J}_2^T \cdots \mathbf{J}_{N_v}^T]^T \in R^{(3N_v)\times 1}$: primary current density

$\mathbf{J}_l^T = (j_l^x \ \ j_l^y \ \ j_l^z) \in R^{1\times 3}$: three dipole moments at the l^{th} voxel

$$\mathbf{K} = \begin{bmatrix} \mathbf{k}_{1,1} & \cdots & \mathbf{k}_{1,N_V} \\ \vdots & \ddots & \vdots \\ \mathbf{k}_{N_E,1} & \cdots & \mathbf{k}_{N_E,N_V} \end{bmatrix} \in R^{N_E \times (3N_V)}$$: lead field (depending on head model)

$\mathbf{k}_{i,l} = (k_{i,l}^x \ k_{i,l}^y \ k_{i,l}^z) \in R^{1\times 3}$: scalp electric potential at the i^{th} electrode,

due to a unit strength $\{x,y,z\}$ - oriented dipole at the l^{th} voxel.

$\mathbf{1} \in R^{N_E \times 1}$: a vector of ones. c : arbitray constant.

Minimum norm inverse solution:

Find $\mathbf{J}$ by minimizing the following functional w.r.t. $\mathbf{J}$ and c, for given $\mathbf{K}$, $\boldsymbol{\Phi}$, and α.

$$F = \| \boldsymbol{\Phi} - \mathbf{KJ} - c\mathbf{1} \|^2 + \alpha \| \mathbf{J} \|^2$$

Using average reference transforms of Φ, *i.e.,* c**=0:**

$$\hat{\mathbf{J}} = \mathbf{T}\boldsymbol{\Phi}$$

$$\mathbf{T} = \mathbf{K}^T [\mathbf{KK}^T + \alpha \mathbf{H}]^+$$

$$\mathbf{H} = \mathbf{I} - \mathbf{11}^T / \mathbf{1}^T \mathbf{1}$$

$\mathbf{M}^+$ is Moore – Penrose pseudoinverse of $\mathbf{M}$.

$\mathbf{I}$: identity matrix.

Standardization of the estimate $\hat{\mathbf{J}}$ (sLORETA):

Estimation of the variance of $\hat{\mathbf{J}}$:

$$\mathbf{S}_J = \mathbf{K}^T[\mathbf{K}\mathbf{K}^T + \alpha\mathbf{H}]\mathbf{K}$$

sLORETA: Standardized low resolution brain electromagnetic tomography

Standardized current density power:

$$\hat{\mathbf{J}}_l^T\{[\mathbf{S}_J]_{ll}\}^{-1}\hat{\mathbf{J}}_l$$

$\hat{\mathbf{J}}_l \in R^{3\times1}$ is the current density estimate at the l^{th} voxel.

$[\mathbf{S}_J]_{ll} \in R^{3\times3}$ is the diagonal block of matrix $\mathbf{S}_\mathrm{J}$

Some issues: (http://www.uzh.ch/keyinst/loreta.htm)

How to choose head model and $\mathbf{K}$?
How to represent/visualize Φ and $\mathbf{J}$? ……

About Head Model:

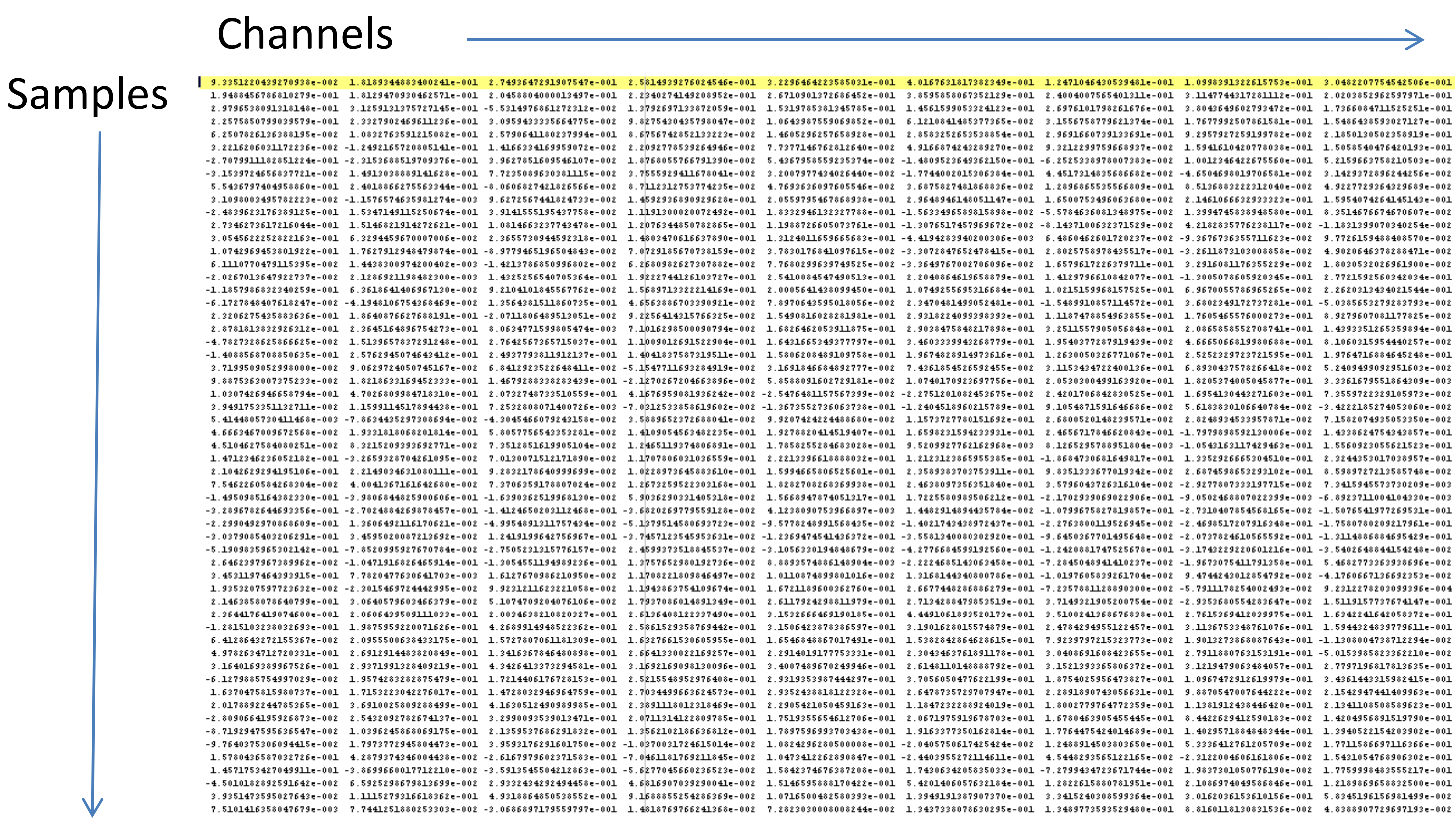
Not much choice in sLORETA.

The intracerebral volume is partitioned in 6239 voxels at 5 mm spatial resolution. Thus, sLORETA images represent the standardized electric activity at each voxel in neuroanatomic Montreal Neurological Institute (MNI) space as the exact magnitude of the estimated current density.

[Fuchs M, Kastner J, Wagner M, Hawes S, Ebersole JS. A standardized boundary element method volume conductor model. Clin Neurophysiol. 2002, 113:702-12.]

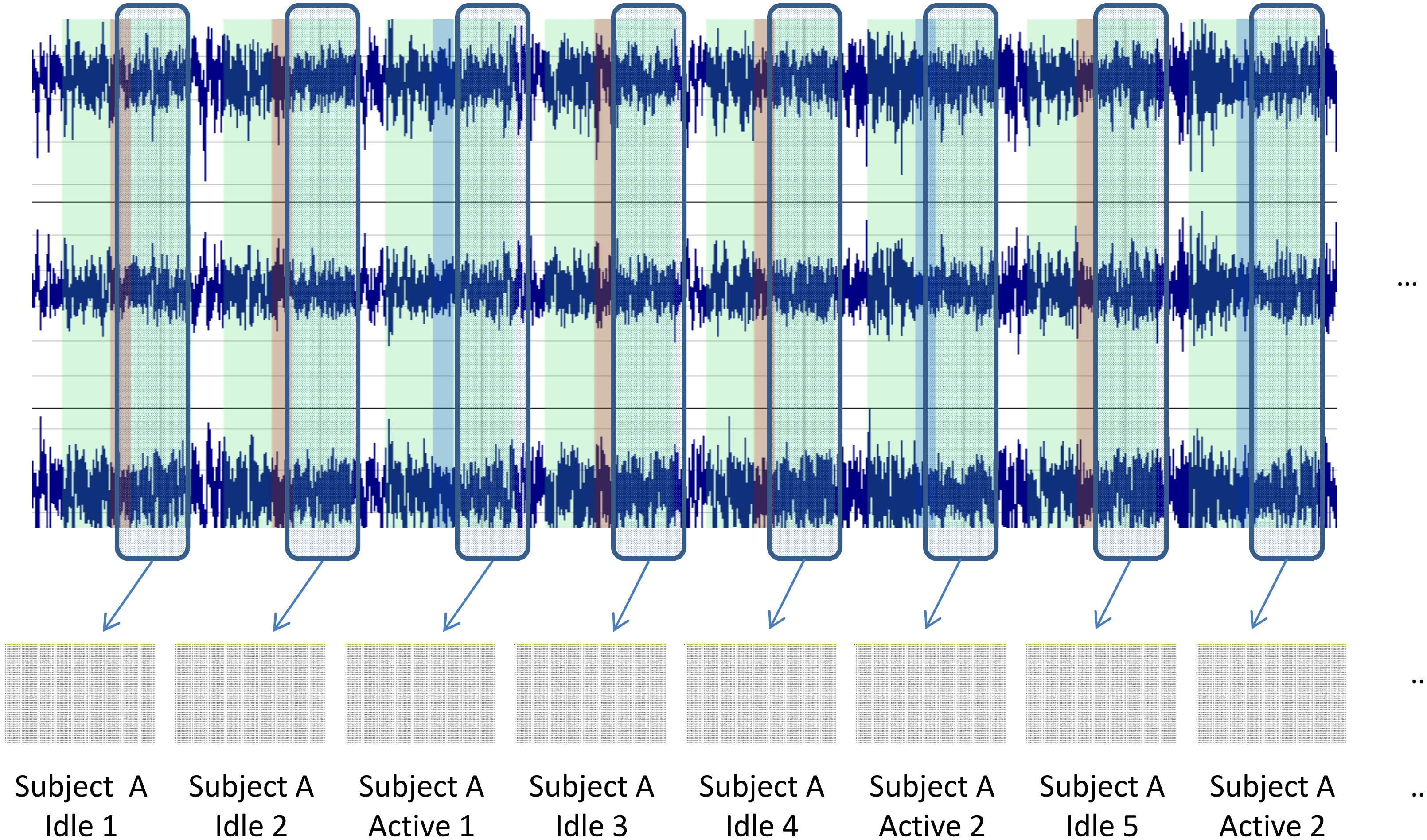
Formatting EEG Data for Input:

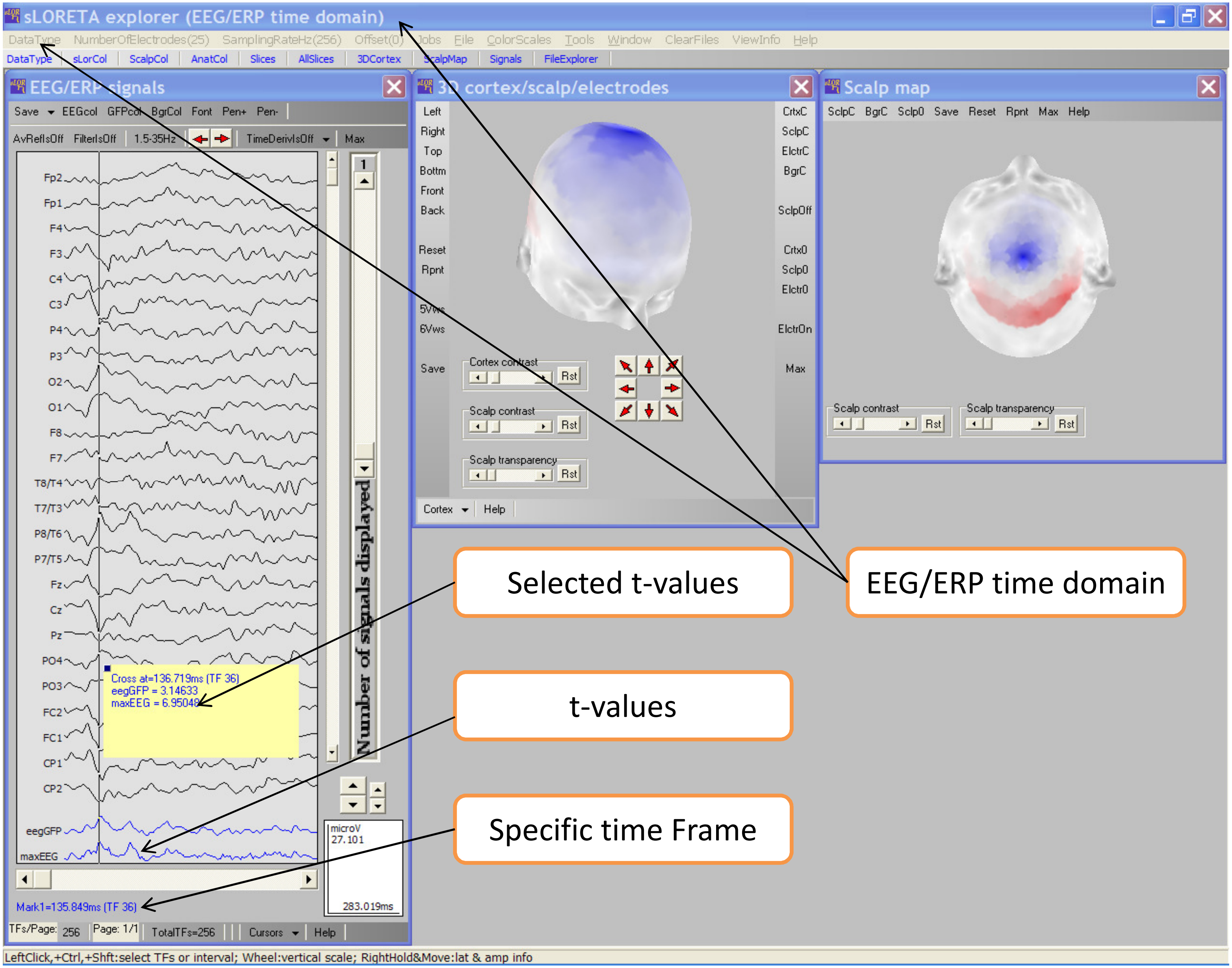
- EEG data can be input as ascii files.
  - In samples * channels format.

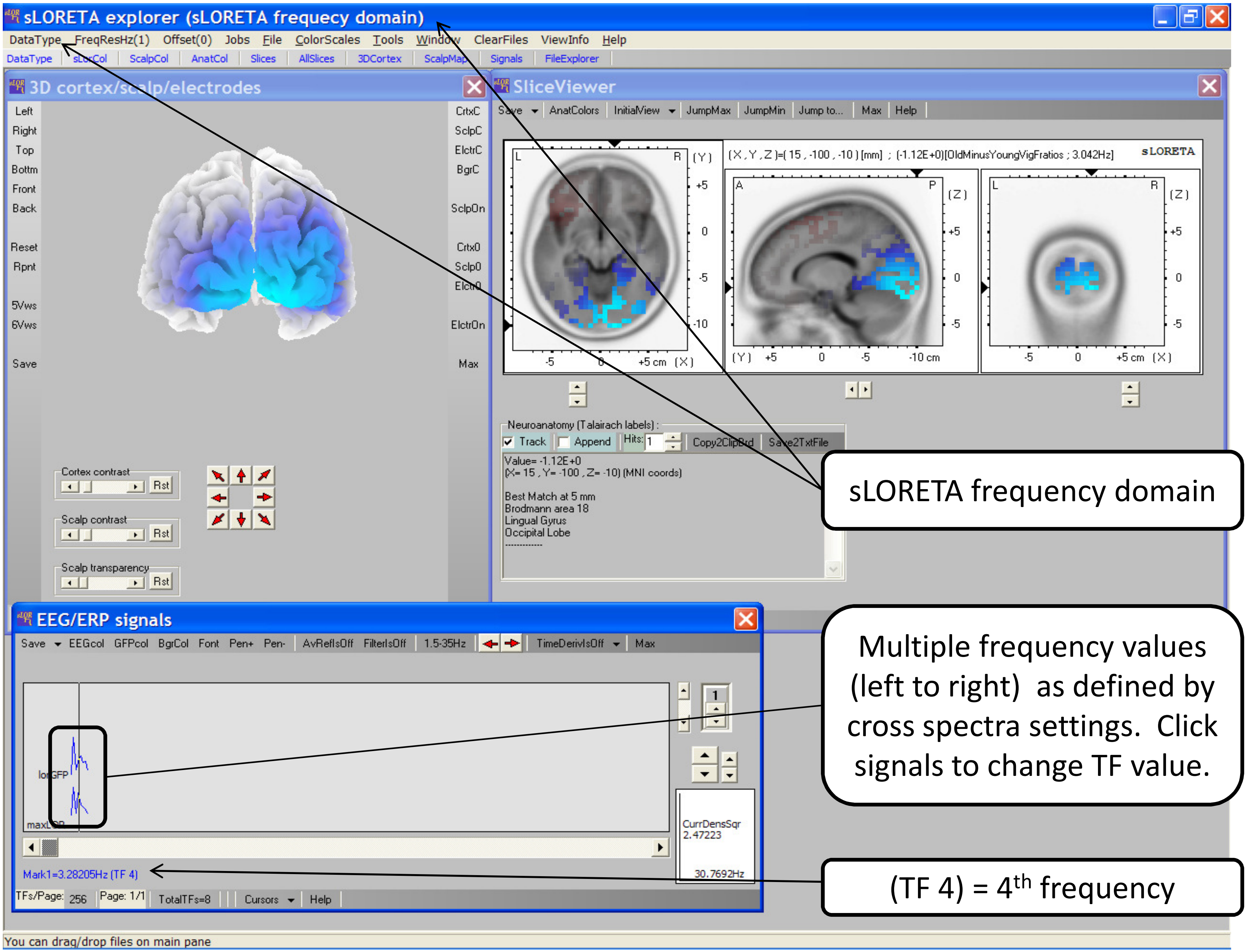


Formatting EEG Data for Input:

- One file is generated for each trial:




sLORETA explorer (EEG/ERP time domain)
EEG/ERP signals
3D cortex/scalp/electrodes
Scalp map
Cross at=136.719ms (TF 36)
eegGFP = 3.14633
maxEEG = 6.95048
Mark1=135.849ms (TF 36)
Selected t-values
EEG/ERP time domain
t-values
Specific time Frame
LeftClick,+Ctrl,+Shft:select TFs or interval; Wheel:vertical scale; RightHold&Move:lat & amp info


sLORETA explorer (sLORETA frequecy domain)
DataType FreqResHz(1) Offset(0) Jobs File ColorScales Tools Window ClearFiles ViewInfo Help
DataType sLorCol ScalpCol AnatCol Slices AllSlices 3DCortex ScalpMap Signals FileExplorer
3D cortex/scalp/electrodes
SliceViewer
(X, Y, Z)=(15, -100, -10) [mm] ; (-1.12E+0)[OldMinusYoungVigFratios ; 3.042Hz]
Neuroanatomy (Talairach labels) :
Value= -1.12E+0
(X= 15 , Y= -100 , Z= -10) (MNI coords)
Best Match at 5 mm
Brodmann area 18
Lingual Gyrus
Occipital Lobe
sLORETA frequency domain
EEG/ERP signals
Multiple frequency values (left to right) as defined by cross spectra settings. Click signals to change TF value.
Mark1=3.28205Hz (TF 4)
CurrDensSqr
2.47223
30.7692Hz
TFs/Page: 256 Page: 1/1 TotalTFs=8
(TF 4) = 4th frequency
You can drag/drop files on main pane

II. Pattern Recognition Approach to EEG Data Analysis

- Basic Steps

  Preprocessing/filtering/artefact removal

  Feature extraction (spatio-temporal-frequency, higher-order)

  Feature selection and dimensionality reduction

  Classification (offline training and online adaptation)

- Software Tools

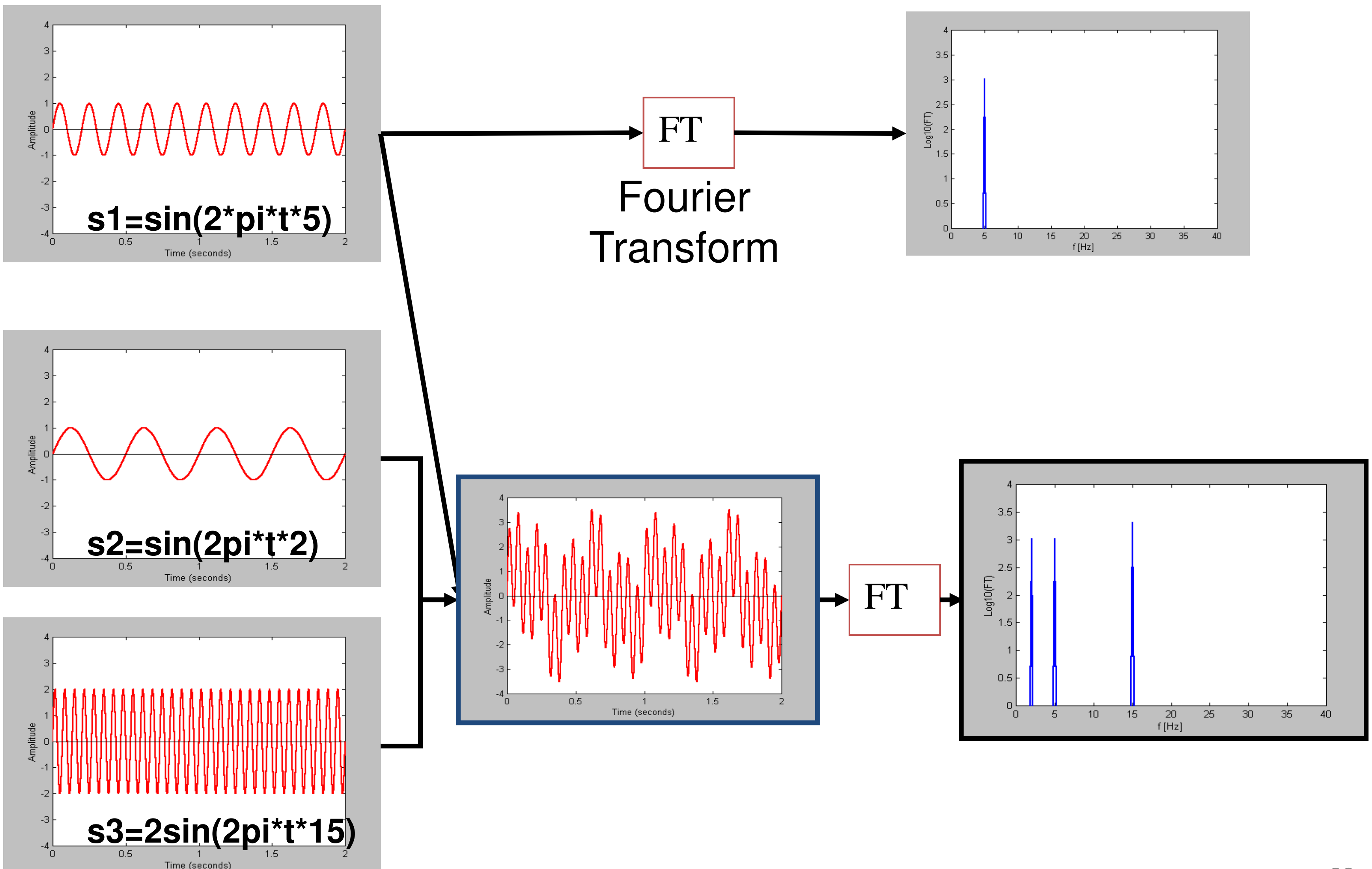
  BioSig

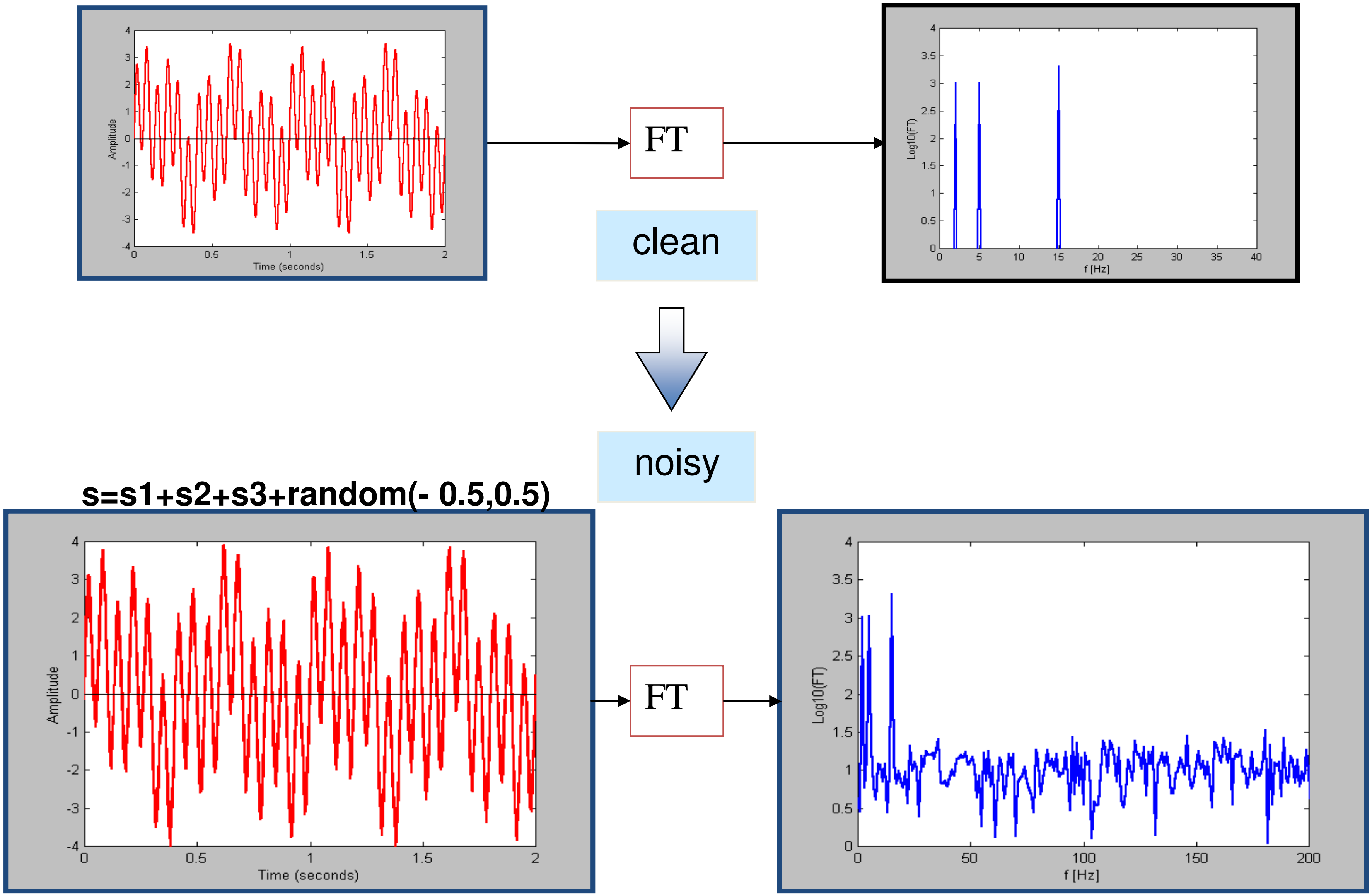
  EEGLab

  sLORETA

  Matlab Simulink (online analysis), Neuroscan software.

➢ Signal Processing - A Brief Primer





Amplitude
Time (seconds)
FT
clean
Log10(FT)
f [Hz]
noisy
s=s1+s2+s3+random(- 0.5,0.5)
Amplitude
Time (seconds)
FT
Log10(FT)
f [Hz]

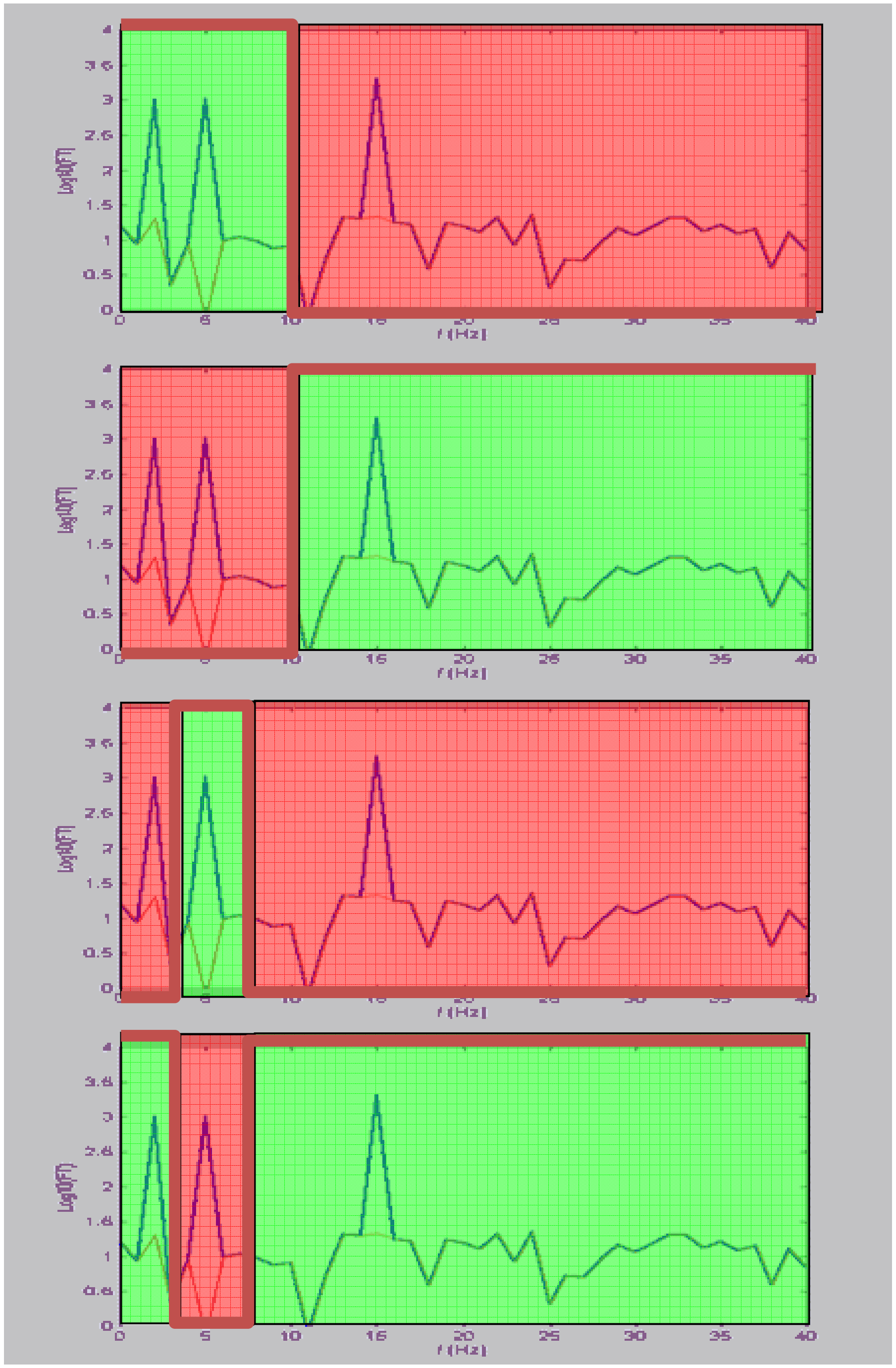
Ideal Filters

Low-Pass

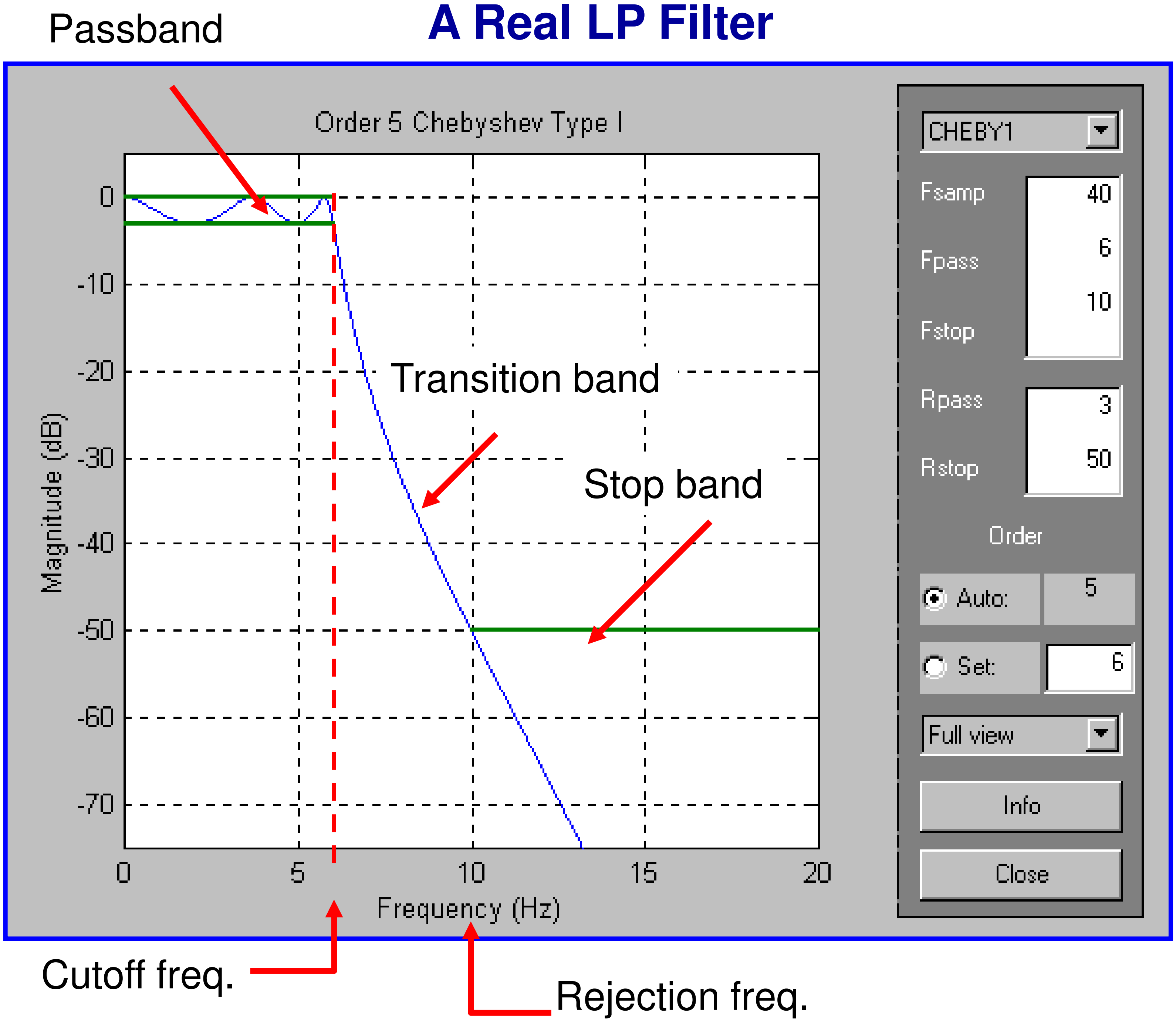
High-Pass

Band-Pass

Band-Stop



A Real LP Filter



Filter Design Using Transfer Function

$$H(s) = \frac{b(s)}{a(s)} = \frac{b(1)s^{nb} + b(2)s^{nb-1} + \cdots + b(nb+1)}{a(1)s^{na} + a(2)s^{na-1} + \cdots + a(na+1)}$$

or

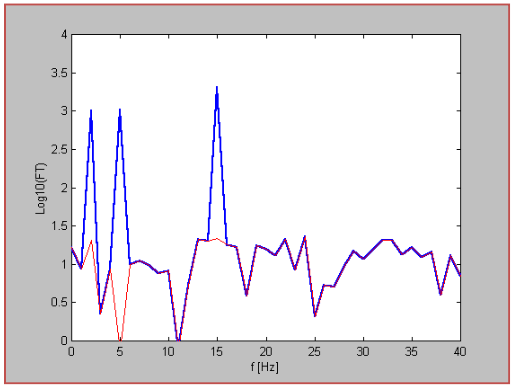
$$H(s) = \frac{z(s)}{p(s)} = k\frac{(s-z(1))(s-z(2))\cdots(s-z(n))}{(s-p(1))(s-p(2))\cdots(s-p(n))}$$

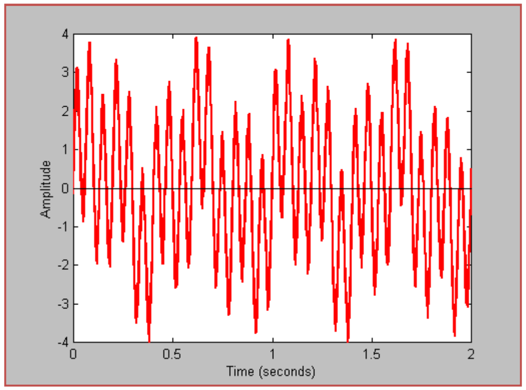
Cutoff frequency, (rejection frequency,) order, filter type

➔ values of transfer function parameters.

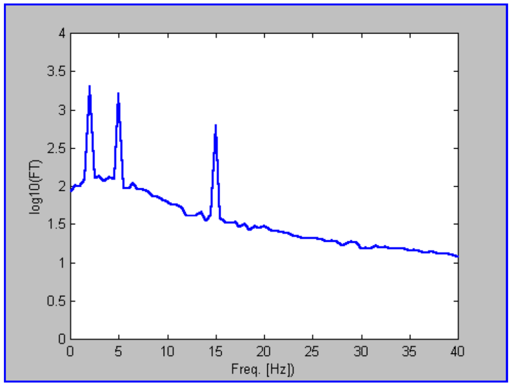
IIR (non-zero a(k) and b(k)) or FIR (only non-zero b(k)):

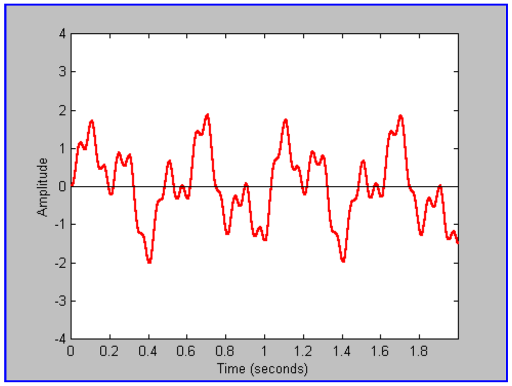
More accurate amplitude frequency response or linear phase unfiltered



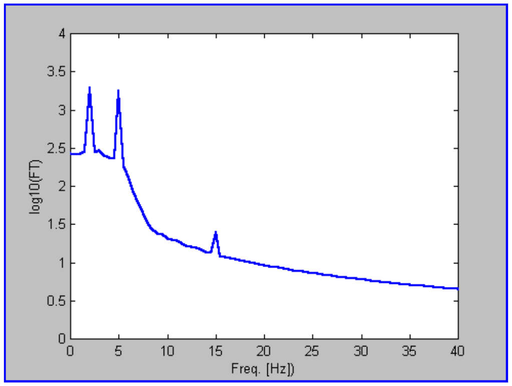


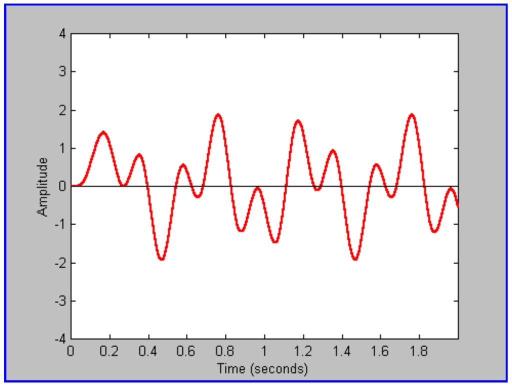
2nd order



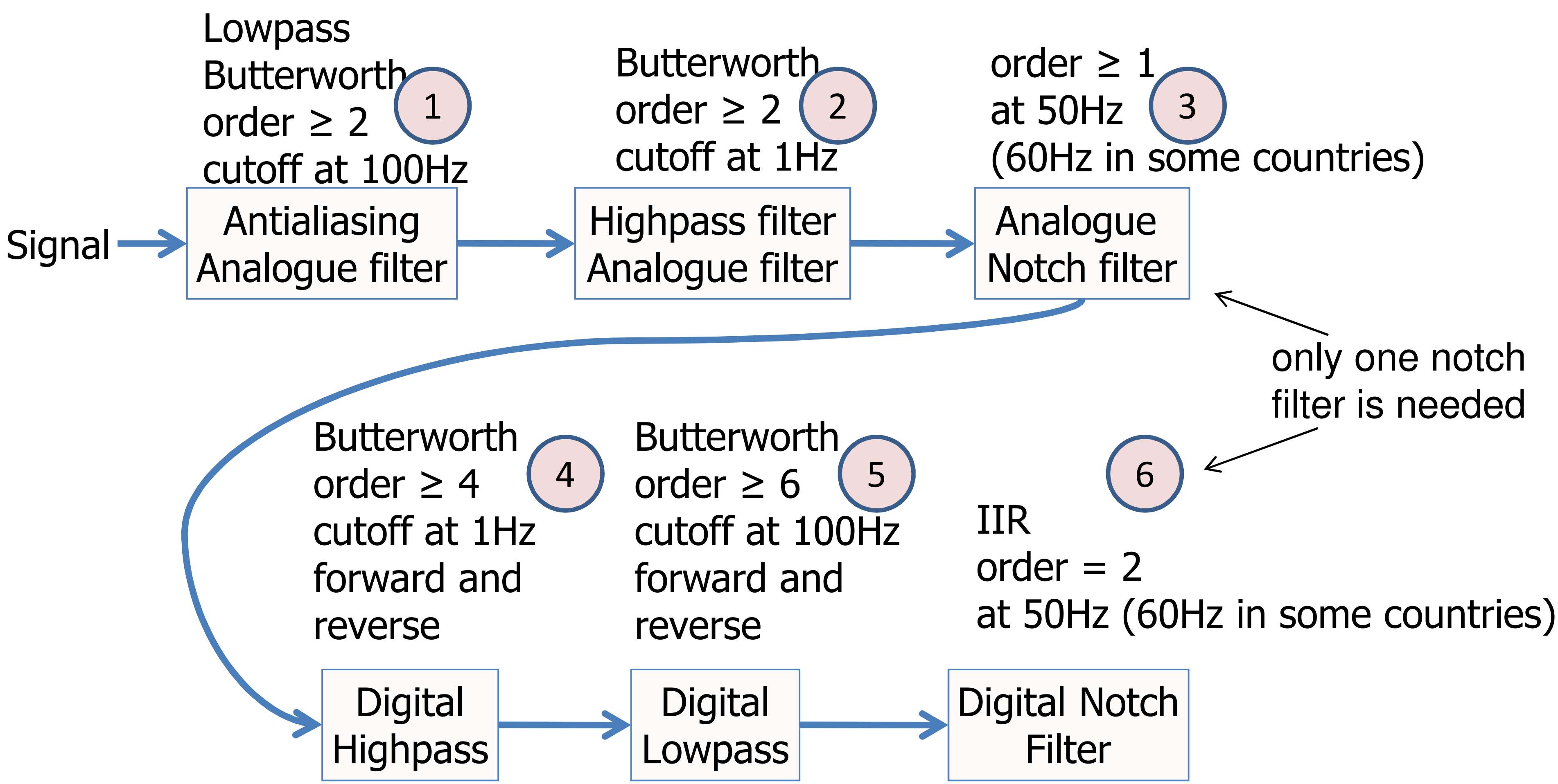


6th order





➢ EEG Pre-processing (Bandpass filtering, Artefacts removal: ICA, …)



Note: the ‘signal’ above is often subtracted from another common reference location

➢ Why Feature Extraction?

- Features are some values computed from the signals.
- Features should be
  - Representative of the signal
  - Reproducible
- Other criteria of the features will depend on the application
  - Smaller dimension than the signal
  - Inter-class variance high/intra-class variance low
  - Robust/enhanced representation of the signal (invariant to changes in noise, scale factors, etc.)
- It is usually much easier to classify features than raw signals.

➢ Commonly used ERP/EEG features

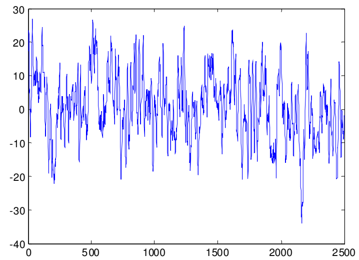
- Negativity/positivity amplitude, latency

  e.g., Socially withdrawn children have smaller mismatch negativity (MMN) amplitude and longer MMN latencies in their auditory ERP.

- Power over frequency bands
  (using bandpass filters or Fourier transform)

  e.g. Trait shyness is related to greater relative resting right frontal EEG alpha activity, whereas trait sociability is related to greater relative resting left frontal EEG alpha activity.

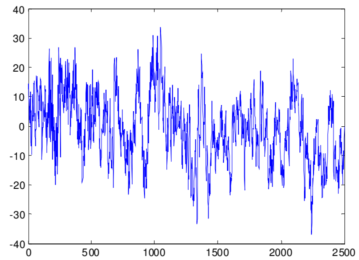
AR model coefficients – another example



EEG during math computation

→

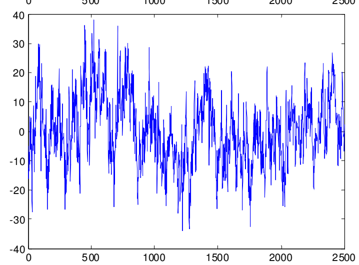
ARc=[-1.5661 0.7114 -0.1843 -0.0583 0.2179 -0.0769



EEG during math computation

→

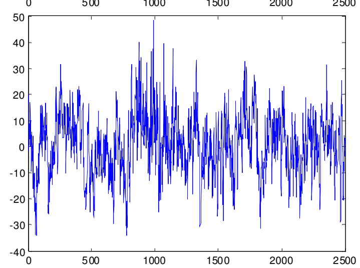
ARc=[-1.6091 0.603 -0.1931 -0.0432 0.2112 -0.0553]



EEG during object rotation

→

ARc=[-0.6128 -0.1677 -0.1159 -0.0733 0.0179 0.0299]



EEG during object rotation

→

ARc=[-0.5647 -0.2189 -0.0826 -0.0756 0.0083 0.0215]

- In the example, 4 EEG plots for one subject are shown from two mental activities (math's activity and imagining an object being rotated)
- Can you say, which is the maths and object rotation activity EEG from the plots?
- Now, use the AR coefficients (order 6); can you see which is which?

AR model:

$$x_t = \sum_{i=1}^{p} a_i \cdot x_{t-i} + \varepsilon_t$$

Yule-Walker equations:

$$\gamma_m = \sum_{i=1}^{p} a_i \cdot \gamma_{m-i} + \sigma_\varepsilon^2 \cdot \delta_{m,0}$$

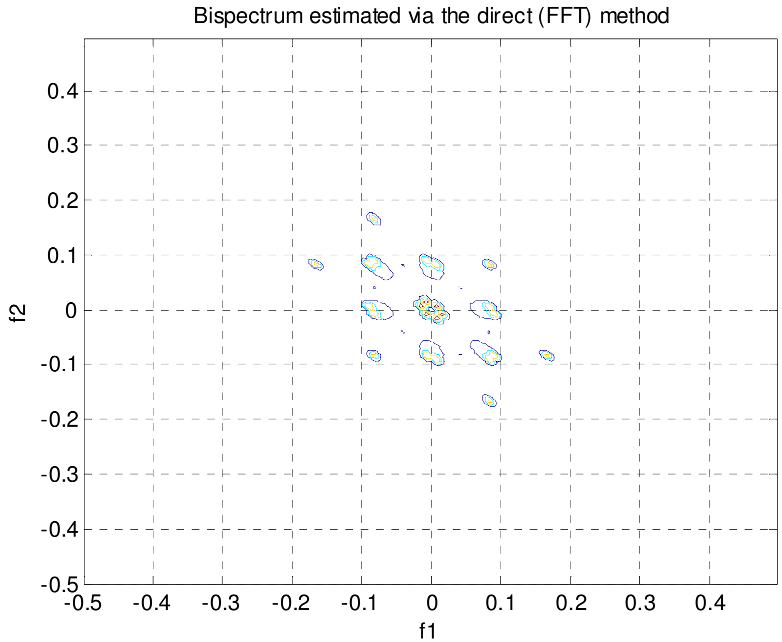
$$m = 0,1,\ldots, p, \ \sigma_\varepsilon^2 = E[\varepsilon_t^2],$$

$$\gamma_m = E[x_t \cdot x_{t-m}]$$

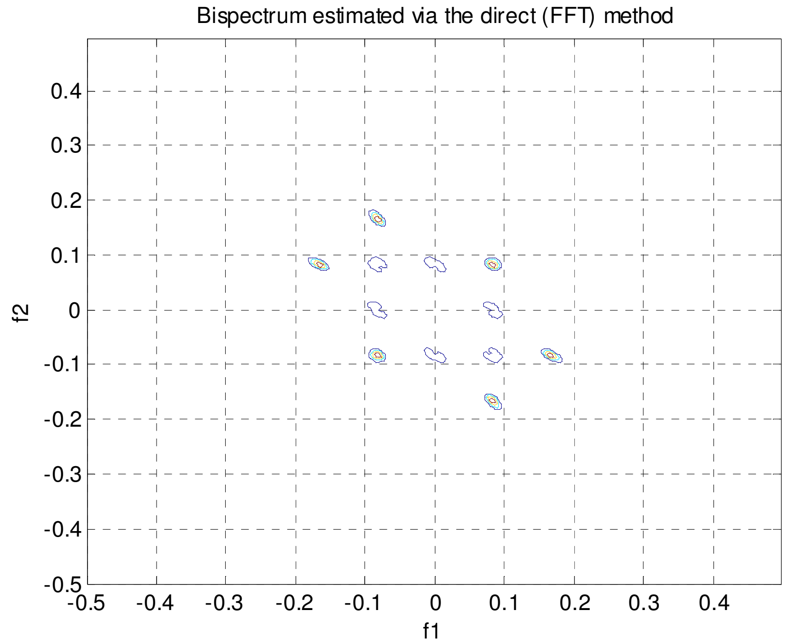
$$\begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_p \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_{-1} & \cdots & \gamma_{1-p} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{2-p} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}$$

AR model coefficients

➢ Higher-order Statistics as BCI Features



(a)



(b)

(a) Bispectrum of an EEG signal corresponding to a left-hand motor imagery;

(b) Bispectrum of an EEG signal from the same channel, corresponding to a right-hand motor imagery.

Bispectrum

The third-order cumulant:

$$C_{3x}(m,n) = E\left[x(k)x(k+m)x(k+n)\right]$$

Bispectrum is defined as the 2-D Fourier transform of the third-order cumulant:

$$B_x(\omega_1,\omega_2) = \sum_{m=-\infty}^{+\infty}\sum_{n=-\infty}^{+\infty} C_{3x}(m,n)\cdot\exp\left[-j2\pi(m\omega_1 + n\omega_2)\right]$$

$\xrightarrow{\text{Feature definition}}$ Higher-order statistics features

➢ Approximate Entropy as BCI Features

An EEG signal segment: $\mathbf{x}=[x(1), x(2), \ldots, x(N)]$

A sequence of vectors: $\mathbf{y}=[y_1, y_2, \ldots, y_M]$

$y_i=[x(i), x(i+\tau), x(i+2\tau), \ldots, x(i+(m-1)\tau)]$

$$C_i^m(r) = \sum_{j=1}^{N-(m-1)} \frac{\Theta(r - \|y_i - y_j\|)}{N-(m-1)}$$

$$\Phi^m(r) = \frac{1}{N-(m-1)} \sum_{i=1}^{N-(m-1)} \ln[C_i^m(r)]$$

$$ApEn(m, r) = \Phi^m(r) - \Phi^{m+1}(r)$$

m : embedding dimension, r : tolerance of comparison,

$\Theta(v)$: Heaviside function.

➢ Combination of Features

Band power or power spectrum density

AR coefficients / reflection coefficients

Wavelets

Entropy, approximate entropy, complexity

Higher-order statistics, e.g., bispectrum-based

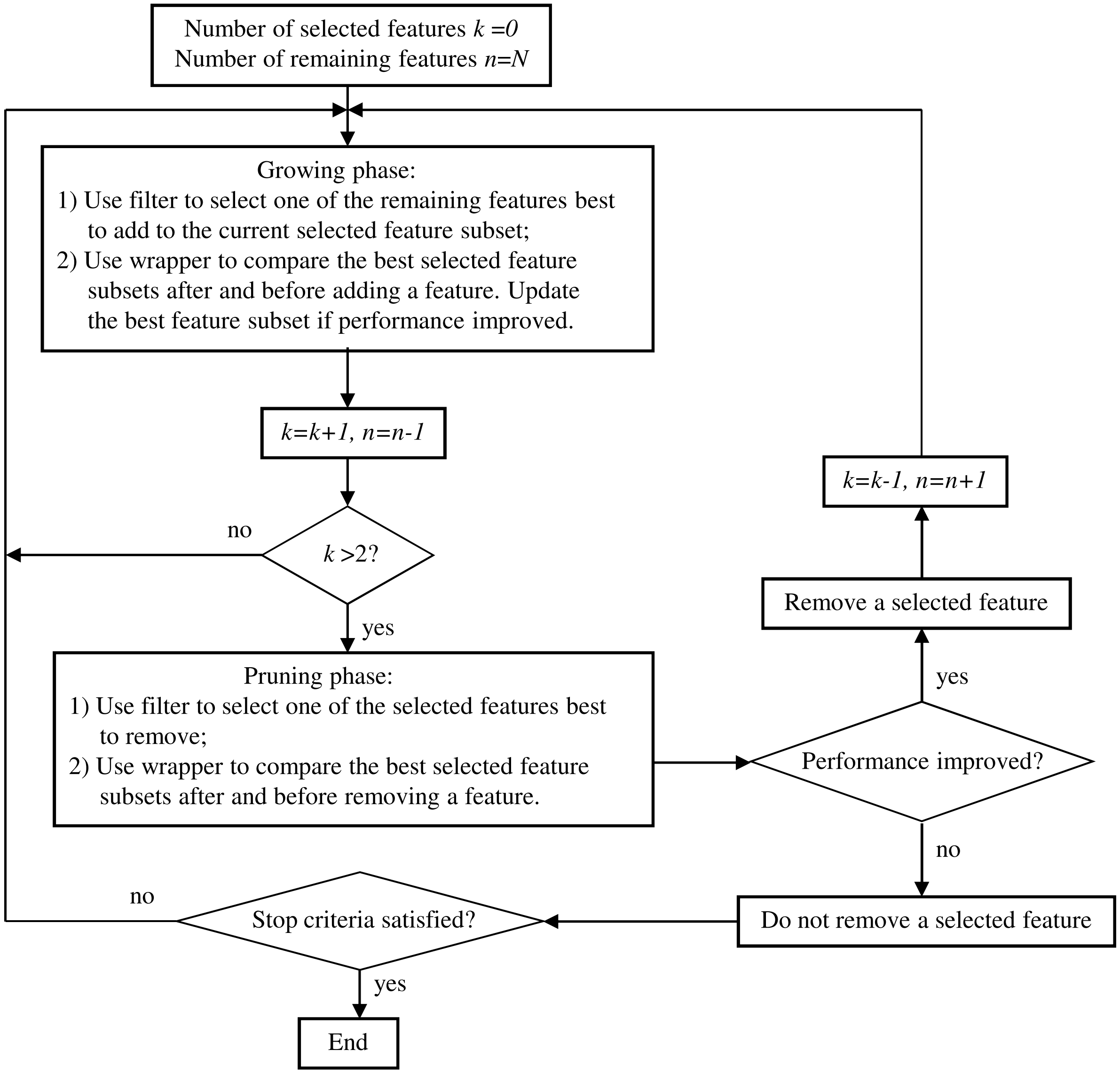
Other linear/nonlinear transformations, e.g., PCA, CSP

Spatio-temporal-frequency integration.

Other feature fusion methods.

Number of possible subsets: $\sum_{m=1}^{N} N!/m!/(N-m)!$

➢ FDHSFFS for Feature Subset Selection



Filters used:

DBI, MRMR

Wrappers used:

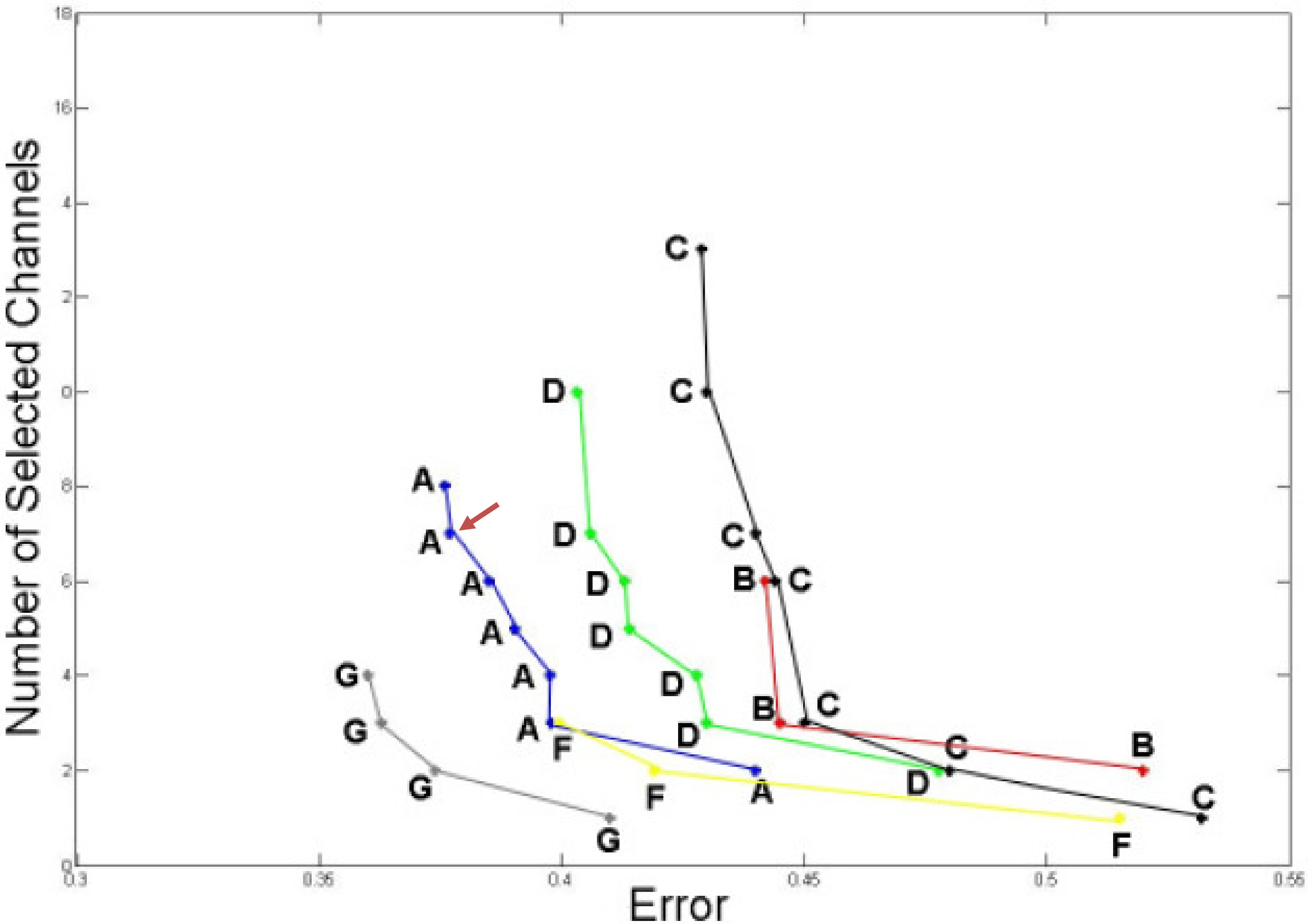
LDA, SVM, KNN

Two basic issues:
Search & evaluation

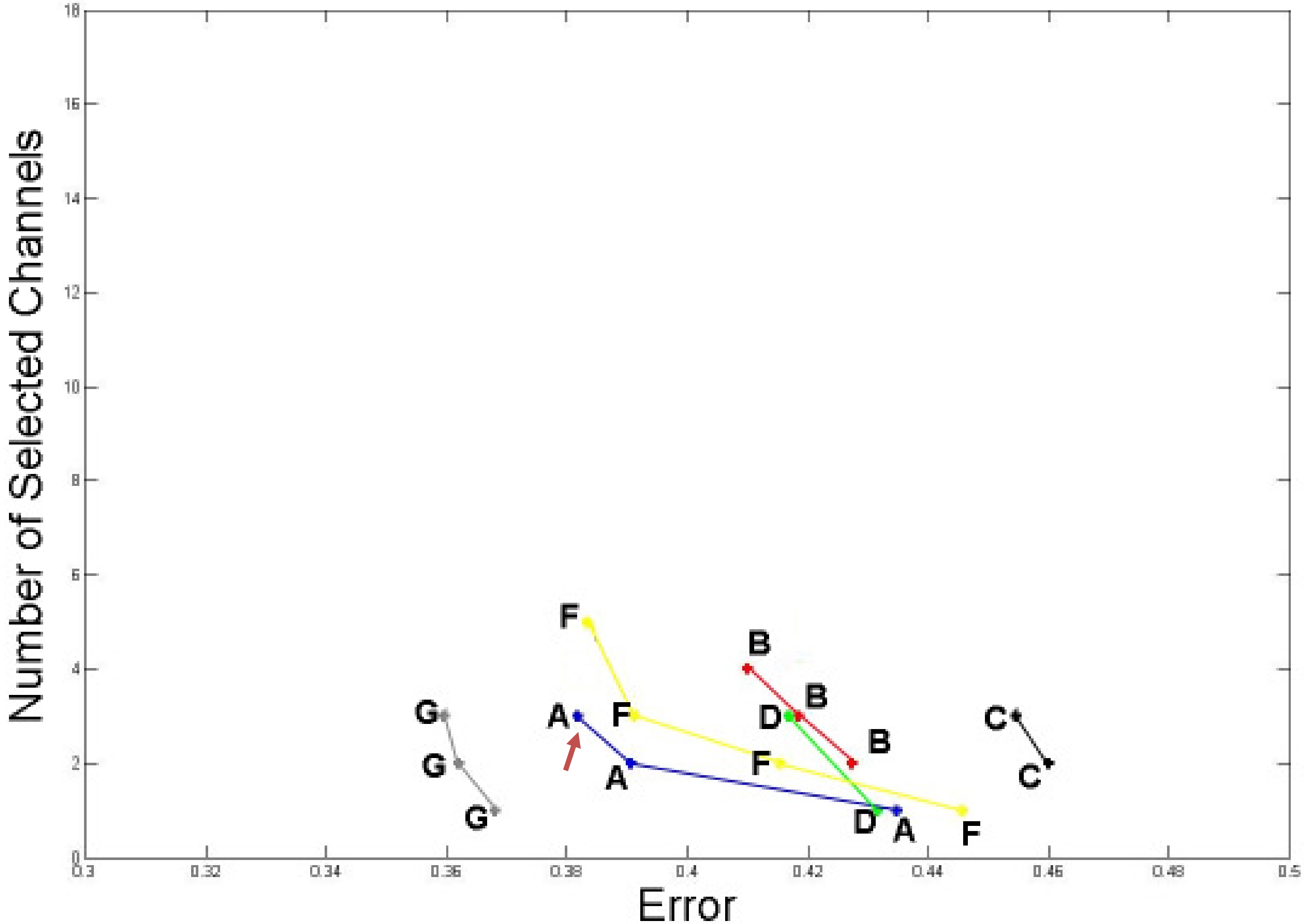
➢ Multi-Objective Evolutionary Methods for Channel Selection

- Every channel is modelled as a binary variable, with 0 for channel not selected and 1 for a selected channel.
- Each individual, a string of 0's and 1's, represents a possible solution.
- First objective is the error rate defined as:
  E = 1- CV, CV is the N-fold cross validation accuracy.
- Second objective is the number of selected channels.
- The goal is to find a set of solutions that minimize both objectives.
- Algorithms: Multi-Objective PSO (Reyes and Coello 2005),
  MOEA/D (Zhang and Li 2007)

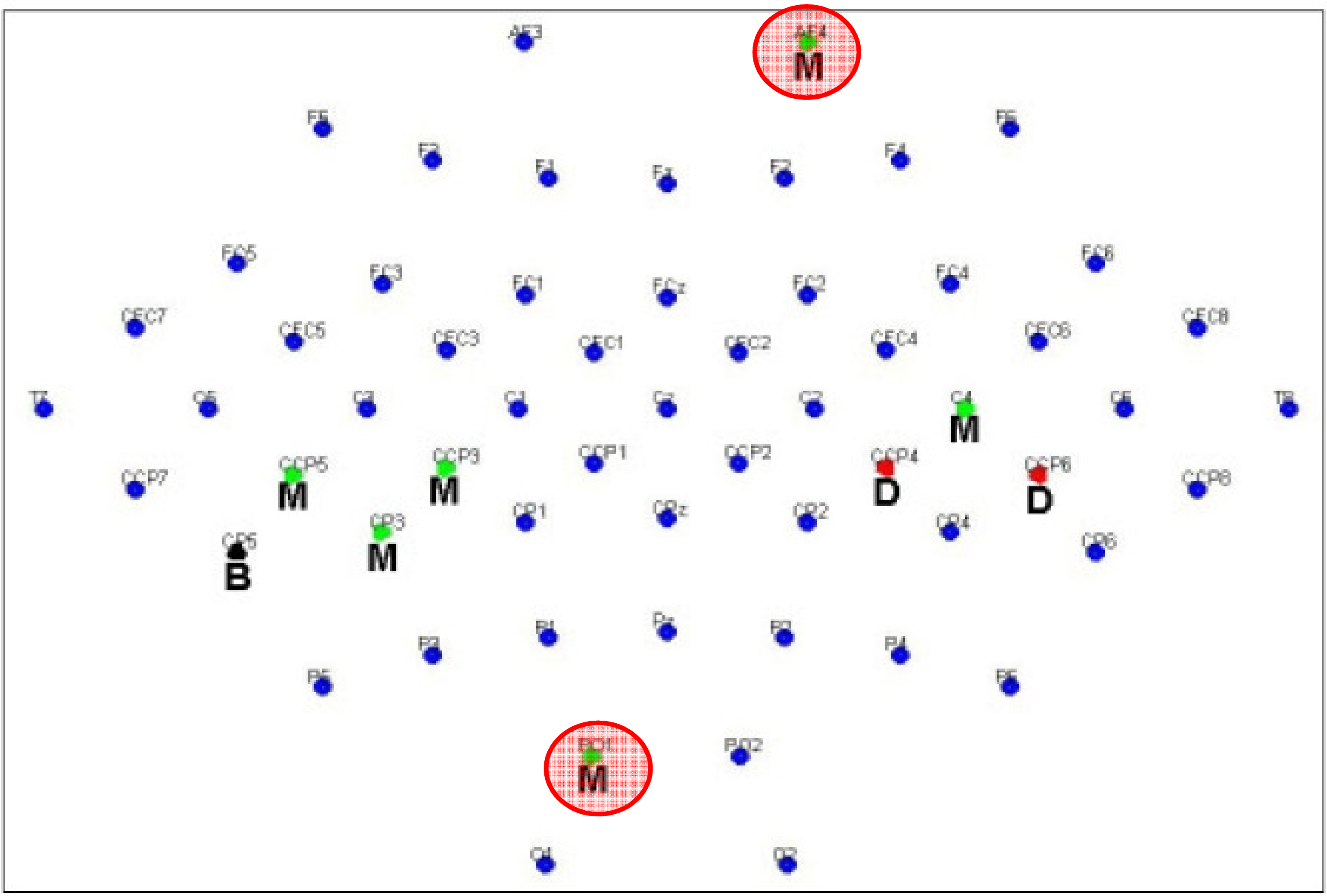
MOPSO Pareto Front



MOEA/D Pareto Front



Selected Channels for Subject A – An Example



➢ Classification (Pattern Recognition)

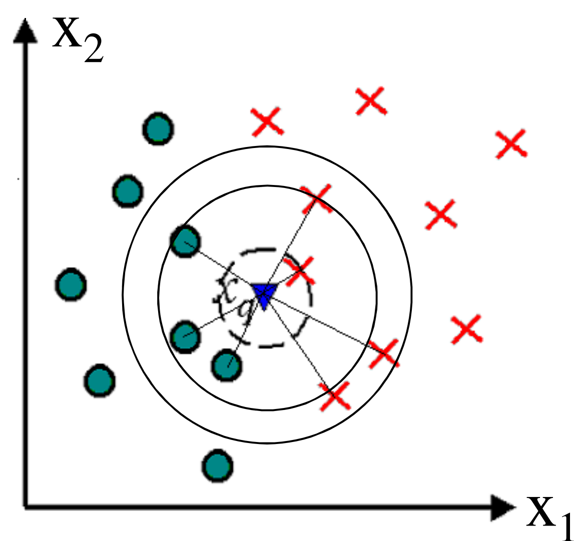
- ***Decision tree***

  Key issue: a good rule base (A hierarchical set of "If *conditions* Then *decision*"). It is usually difficult to design a complete rule base, e.g., thresholds selection.

- ***Similarity matching***

  Popular method: k-nearest neighbours (k-NN).

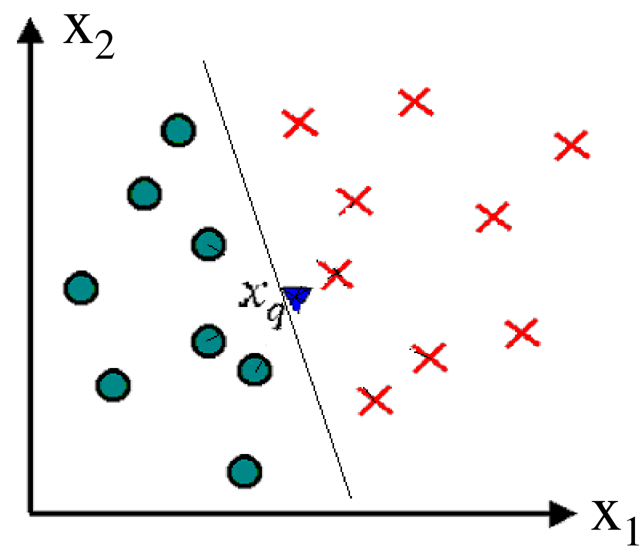
  The class of a data point or feature vector (x_q) is determined by the class of the majority in its k nearest neighbours in the sample data set. (*e.g.*, k=1, 5, or 7)



Key issues?

- ***Linear discriminant analysis (LDA)***

The class of a data point or feature vector (x_q) is determined by a decision line which is designed by the Fisher's discriminant criterion: to maximize between-class distance and minimize within-class covariance.



$$y = \mathrm{sgn}(\mathbf{w}^T \mathbf{x} + w_0), \quad \mathrm{sgn}(v) = \begin{cases} 1 & if \;\; v > 0 \\ 0 & if \;\; v \le 0 \end{cases}$$

Mean vectors of class 1 and class 2:

$$\mathbf{m}_1 = \frac{1}{N_1}\sum_{n\in C_1}\mathbf{x}_n \qquad \mathbf{m}_2 = \frac{1}{N_2}\sum_{n\in C_2}\mathbf{x}_n$$

N_1+N_2=N, C_k represents data from class k (k=*1,2*).

Projections of the mean vectors via $\mathbf{w}$:

$$V_1 = \mathbf{w}^T\mathbf{m}_1 \qquad V_2 = \mathbf{w}^T\mathbf{m}_2$$

Between-class distance on projected space: $|V_1 - V_2|$

Within-class covariance on projected space:

$$S_1^2 = \sum_{n\in C_1}(v_n - V_1)^2 \qquad S_2^2 = \sum_{n\in C_2}(v_n - V_2)^2$$

$$v_n = \mathbf{w}^T\mathbf{x}_n$$

Fisher's criterion:

$$J(\mathbf{w}) = \frac{(V_1 - V_2)^2}{S_1^2 + S_2^2} = \frac{\mathbf{w}^T S_B \mathbf{w}}{\mathbf{w}^T S_W \mathbf{w}}$$

$$S_B = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T$$

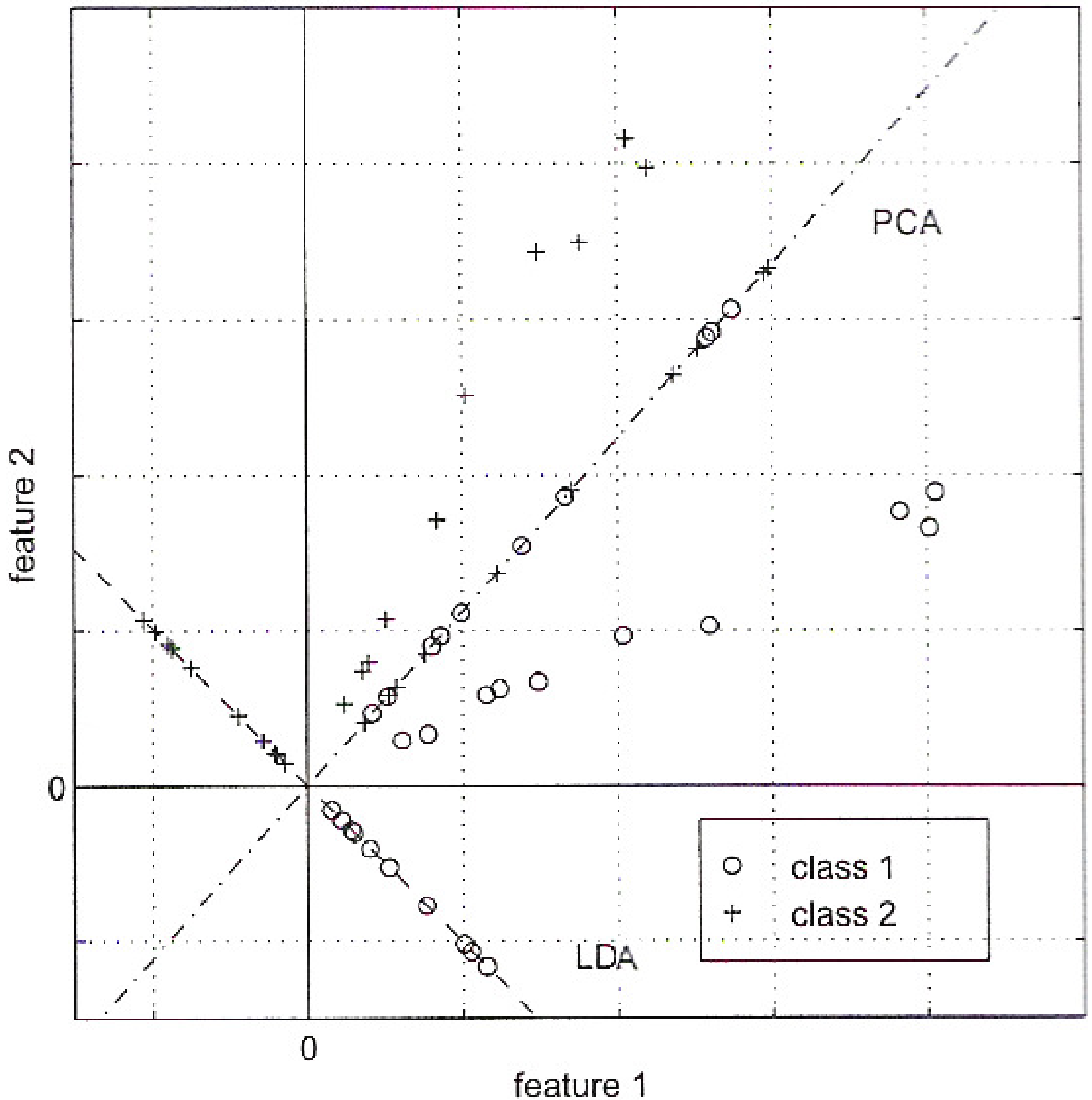
$$S_W = \sum_{n \in C_1} (\mathbf{x}_n - \mathbf{m}_1)(\mathbf{x}_n - \mathbf{m}_1)^T + \sum_{n \in C_2} (\mathbf{x}_n - \mathbf{m}_2)(\mathbf{x}_n - \mathbf{m}_2)^T$$

LDA solution:

$$\mathbf{w} = S_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2)$$

$$w_0 = \mathbf{w}^T \mathbf{m}, \quad \mathbf{m} = \frac{N_1 \mathbf{m}_1 + N_2 \mathbf{m}_2}{N_1 + N_2}$$

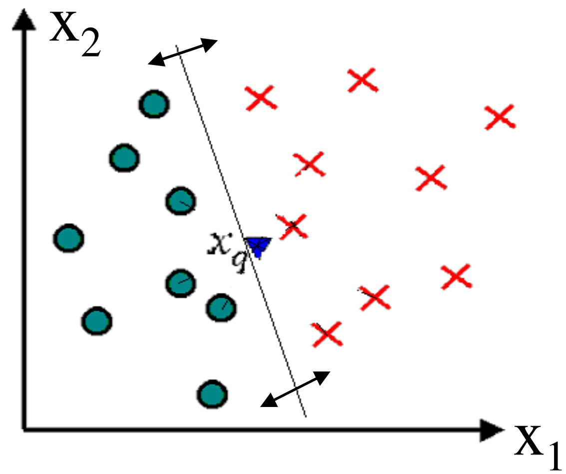
Means and covariance of the data → Weights of the classifier



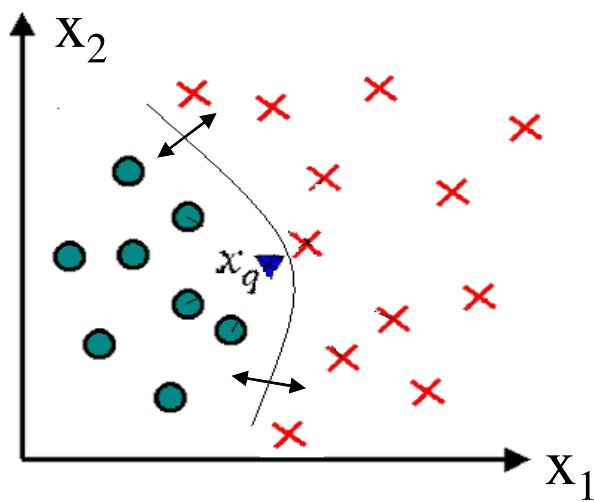
PCA
LDA
feature 2
feature 1
0
0
class 1
class 2

- ***Artificial neural networks***

The class of a data point or feature vector (x_q) is determined by a decision line (linear or nonlinear) which is designed by a learning process.



Key issues: collection of representative training data, appropriate learning process, e.g., back-propagation, Elman network.



$$y_k = \varphi[\sum_{i=0}^{n} w_{ki} h_i] = \varphi[\sum_{i=0}^{n} w_{ki} \varphi^h (\sum_{j=0}^{m} w_{ij}^h x_j)]$$

- ***Data distribution based classifiers***

Naïve Bayesian classifier:

"Naïve" assumption: Each feature is conditionally independent of every other feature.

$$p(c = class_i \mid x_1, \ldots, x_n) = p(c = class_i) \prod_{j=1}^{n} p(x_j \mid c = class_i)$$

$$class(x_1, \ldots, x_n) = \arg\max_{class_i} p(c = class_i \mid x_1, \ldots, x_n)$$

Gaussian Mixture Model (GMM).

Hidden Markov Model (HMM).

……

➢ Online Adaptation of Self-paced BCI Systems

1) Incremental updating of means and covariances for LDA, Naïve Bayesian, Gaussian Mixture Model (GMM), Hidden Markov Model (HMM), and Conditional Random Fields (CRF) adaptation.

- A standard approach: If a new input $\mathbf{x}_t$ is from the j^{th} class (by label or clustering):

$$\mu_j^t = \mu_j^{t-1} + \frac{\mathbf{x}_t - \mu_j^{t-1}}{N_j + 1}$$

$$\Sigma_j^t = \frac{(N_j - 1)\Sigma_j^{t-1} + (\mathbf{x}_t - \mu_j^t)(\mathbf{x}_t - \mu_j^t)^T}{N_j}$$

$$(N_j : number\ of\ samples\ from\ the\ j^{th}\ class)$$

- Incremental EM (Expectation-Maximization) -unsupervised approach

E-step:

$$\gamma(z_j) = p(z_j = 1 \mid \mathbf{x}_t) = \frac{\pi_j^{t-1} \aleph(\mathbf{x}_t \mid \mu_j^{t-1}, \Sigma_j^{t-1})}{\sum_{k=1}^{K} \pi_k^{t-1} \aleph(\mathbf{x}_t \mid \mu_k^{t-1}, \Sigma_k^{t-1})}$$

M-step:

$$\mu_j^t = \frac{1}{N_j^t}(N_j^{t-1} \cdot \mu_j^{t-1} + \gamma(z_j)\mathbf{x}_t)$$

$$\Sigma_j^t = \frac{1}{N_j^t}(N_j^{t-1} \cdot \Sigma_j^{t-1} + \gamma(z_j)(\mathbf{x}_t - \mu_j^t)(\mathbf{x}_t - \mu_j^t)^T)$$

$$N_j^t = N_j^{t-1} + \gamma(z_j), \quad \pi_j^t = \frac{N_j^t}{t}$$

Classification using GMM:

$$p(c = class_i \mid \mathbf{x}_t) = \sum_{k=1}^{K} p(c = class_i \mid z_k)\, p(z_k \mid \mathbf{x}_t)$$

$$p(z_k \mid \mathbf{x}_t) = \frac{\pi_k^t \aleph(\mathbf{x}_t \mid \mu_k^t, \Sigma_k^t)}{\sum_{j=1}^{K} \pi_j^t \aleph(\mathbf{x}_t \mid \mu_j^t, \Sigma_j^t)}$$

$p(c = class_i \mid z_k)$ - Probability of being class i if data from component k

$$class(\mathbf{x}_t) = \arg\max_{class_i} p(c = class_i \mid \mathbf{x}_t)$$

2) Extended Kalman filter based adaptation of LDA and dynamic logistic regression.

e.g.,

$$y = \frac{1}{1 + e^{\mathbf{w}^T \varphi}}$$

$$u = y(1 - y)$$

$$\mathbf{w}_{t|t-1} = \mathbf{w}_{t-1|t-1}$$

$$\boldsymbol{P}_{t|t-1} = \boldsymbol{P}_{t-1|t-1} + \boldsymbol{Q}_{t-1}$$

$$\boldsymbol{Q}_{t-1} = q_{t-1}\mathbf{I}$$

$$q_{t-1} = \max\{I_{t-1}, 0\}$$

$$I_{t-1} = u_{t-1|t} - u_{t-1|t-1}$$

$$\mathbf{w}_{t|t} = \mathbf{w}_{t|t-1} + \mathbf{K}_t (z_t - y_{t|t-1})$$

$$\boldsymbol{P}_{t|t} = \boldsymbol{P}_{t|t-1} - \mathbf{K}_t u_{t|t-1} (\boldsymbol{P}_{t|t-1} \varphi_t)^{\mathsf{T}}$$

$$\mathbf{K}_t = \frac{\boldsymbol{P}_{t|t-1}}{\mathrm{C} + u_{t|t-1} s^2_{t|t-1}} \varphi_t$$

$$s^2_{t|t-1} = \varphi_t^{\mathsf{T}} \boldsymbol{P}_{t|t-1} \varphi_t$$

3) Other methods:

Adaptive classification using sequential Monte Carlo sampling (idea similar to particle filtering)

[J.W. Yoon, S.J. Roberts, M. Dyson, and J.Q. Gan, “Adaptive classification for brain computer interface systems using sequential Monte Carlo sampling,” *Neural Networks*, vol. 22, no. 9, pp. 1286-1294, 2009.]

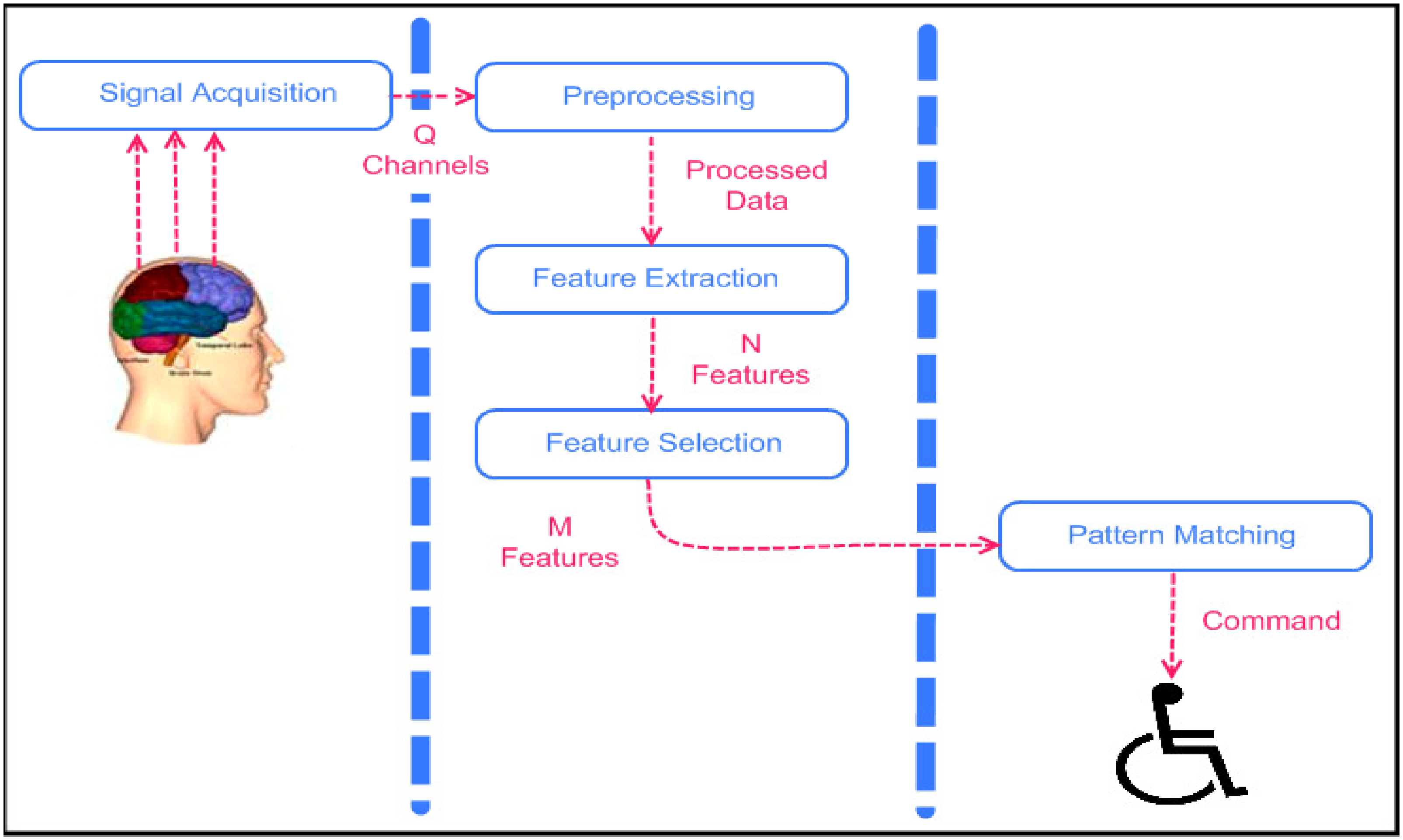
……

4) Some issues:

Overfitting/underfitting to new data, forgetting, …

III. EEG-based BCI Experiment Design

➢ Thought-Driven Control of Mobility Devices via BCI



➢ BCI Protocols

Cognitive/mental tasks

Electrode placement

Synchronous or asynchronous (self-paced)

Spontaneous EEG or evoked potentials (P300, SSVEP)

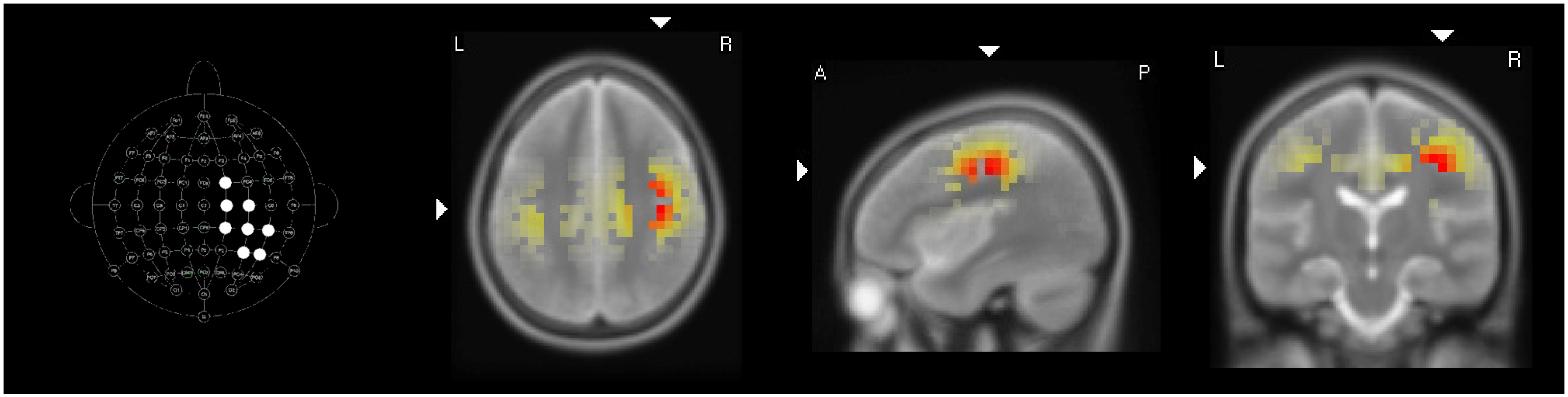
Data recording and labelling for offline analysis

Subjects: healthy or disabled, male or female, ethics, …

Online subject training with biofeedback

Online adaptation

Motor Imagery – Left Hand

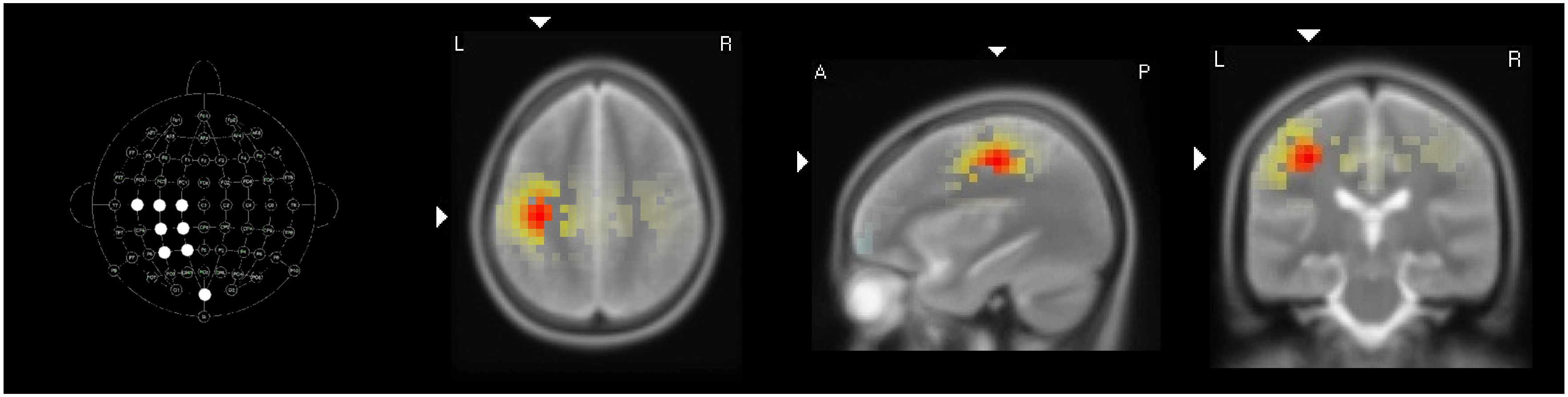


By SFFS

By sLORETA

- BA4: Primary Motor Cortex.
- Precentral Gyrus.
  - Upper Alpha / Mu (10.5 – 12 Hz).

Motor Imagery – Right Hand

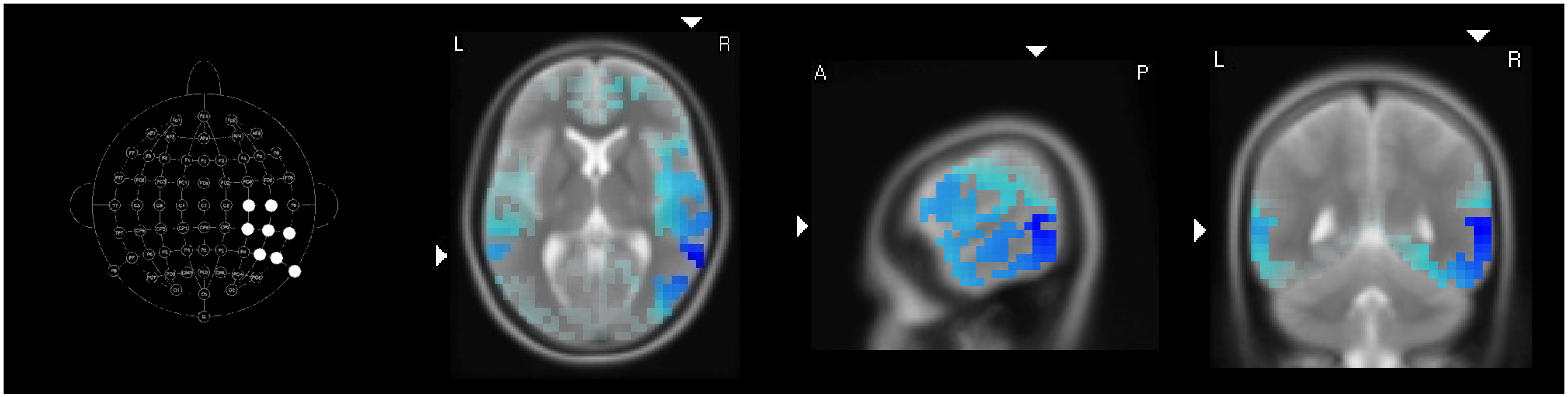


By SFFS

By sLORETA

- BA3: Primary Somatosensory.
- Postcentral Gyrus.
  - Upper Alpha / Mu (10.5 – 12 Hz).

Auditory Imagery

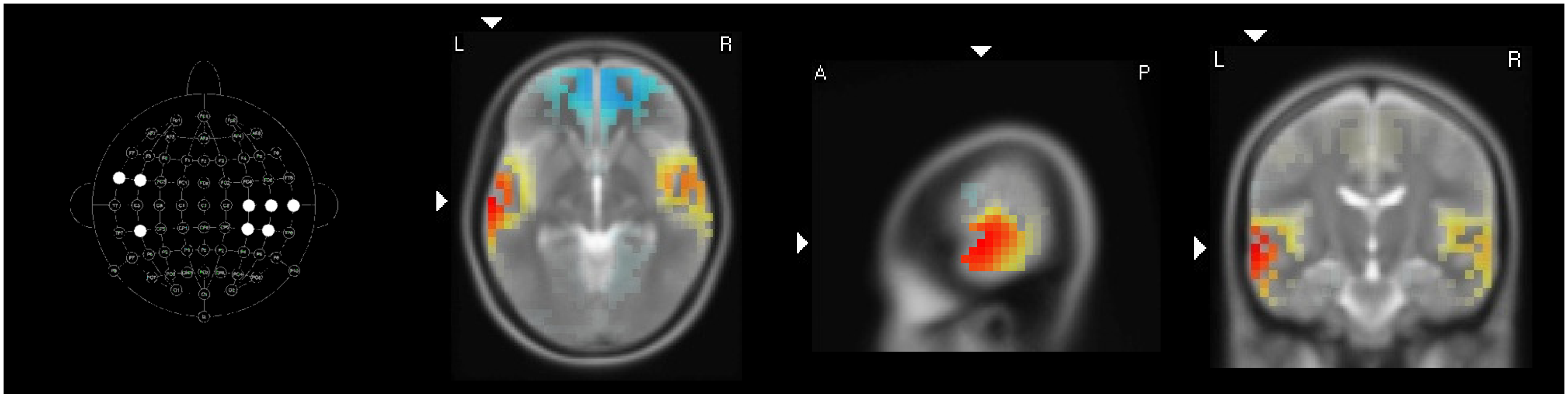


By SFFS

By sLORETA

- BA21: Auditory Association Area.
- Middle Temporal Gyrus.
  - High Beta (21 – 30 Hz).

Phone Imagery

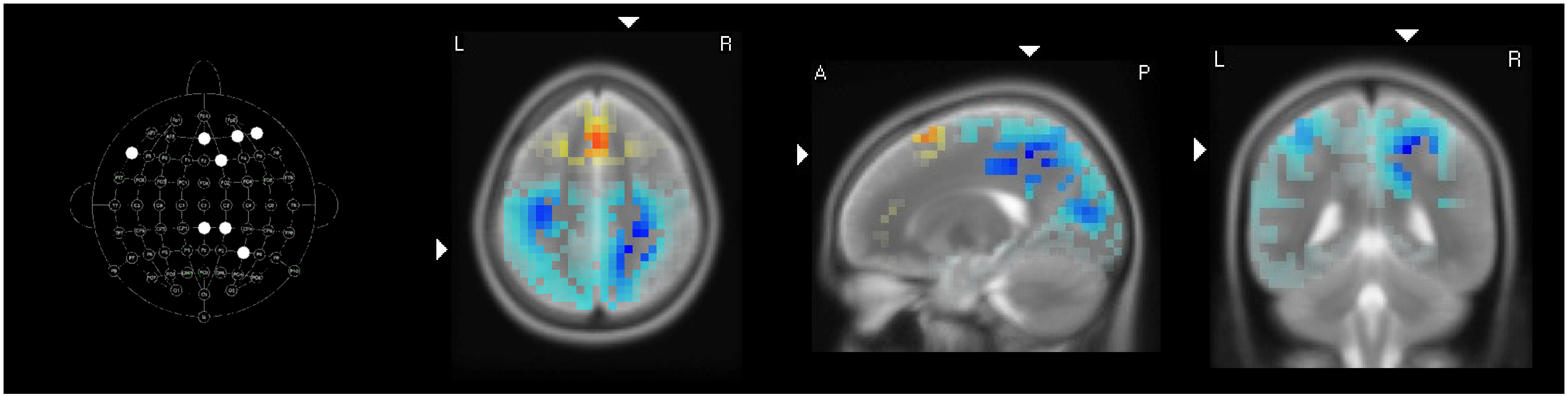


By SFFS

By sLORETA

- BA21: Auditory Association Area.
- Middle Temporal Gyrus.
  - High Beta (21 – 30 Hz).

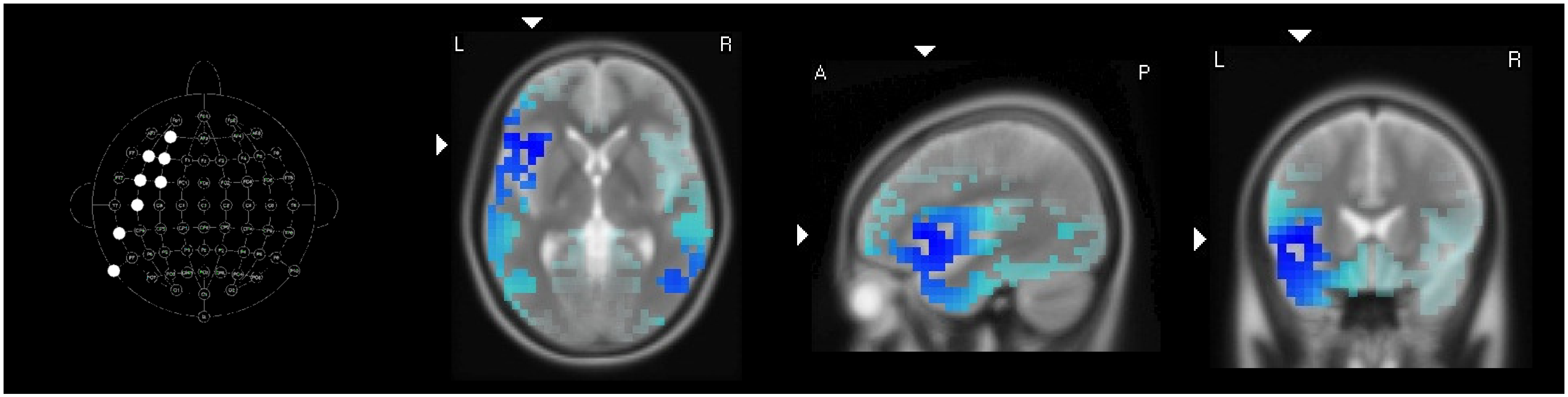
Navigation Imagery



By SFFS By sLORETA

- BA5: Somatosensory Association.
- BA32: Spiers et al: Activity correlated with proximity to the goal during navigation.
- Paracentral Lobule & Cingulate Gyrus.
  - Upper Alpha (10.5 – 12 Hz) & Beta (18.5 – 21 Hz).

Mental Arithmetic

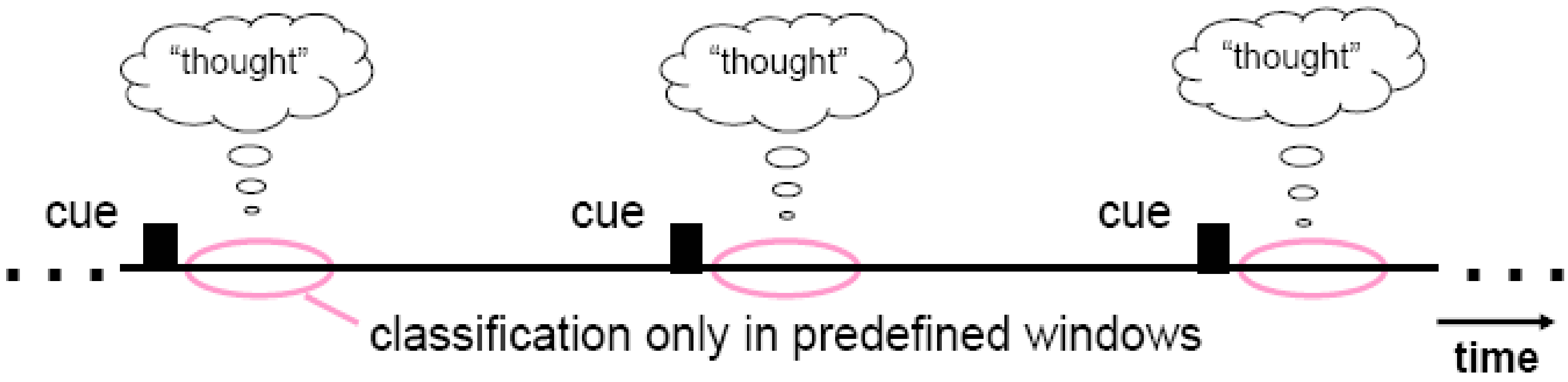


By SFFS

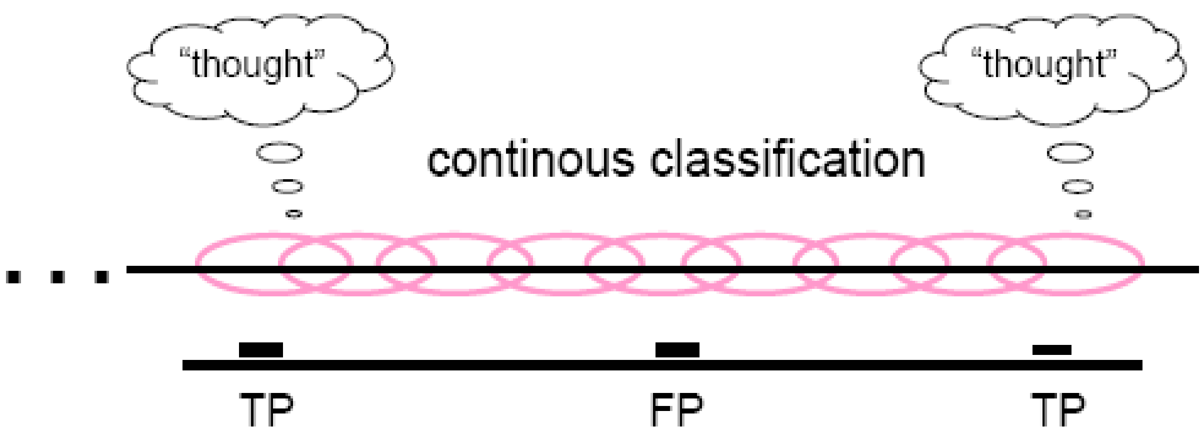
By sLORETA

- BA47: Semantics & Syntax.
- Inferior Frontal Gyrus.
  - Low Beta (12.5 – 18 Hz)

Synchronous BCI (cue-based, COMPUTER-driven)



Asynchronous BCI (uncued, USER-driven)

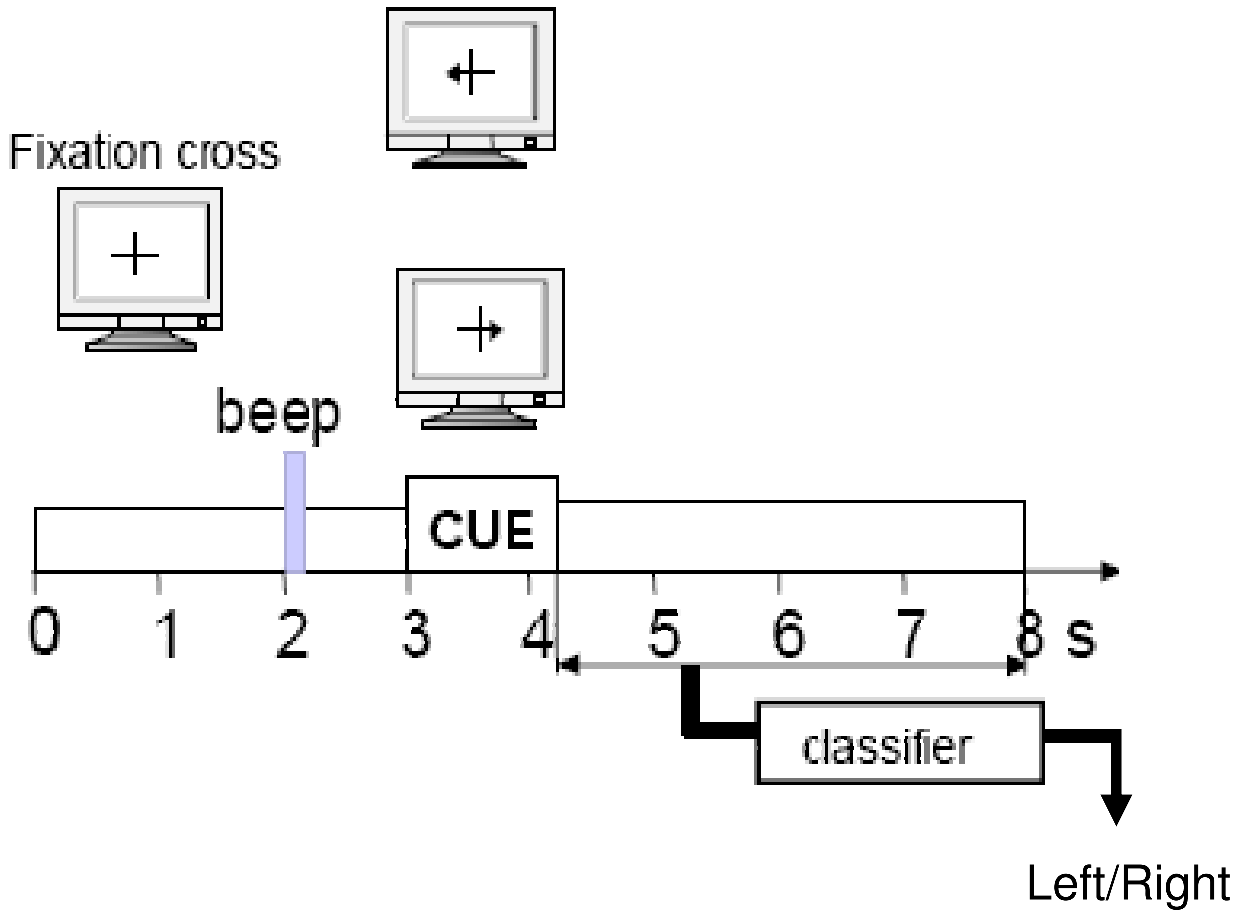


Problems:
minimization of FP
Differentiation between mental activity and rest

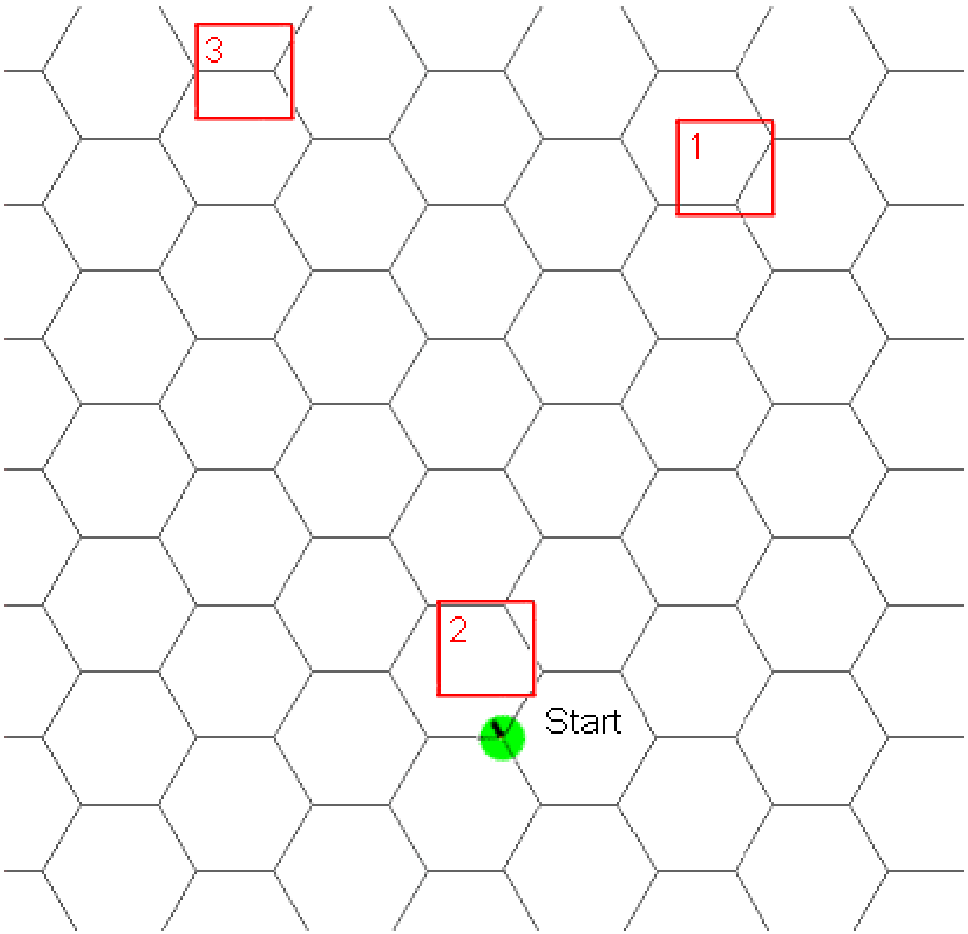
TP – True Positive
FP – False Positive

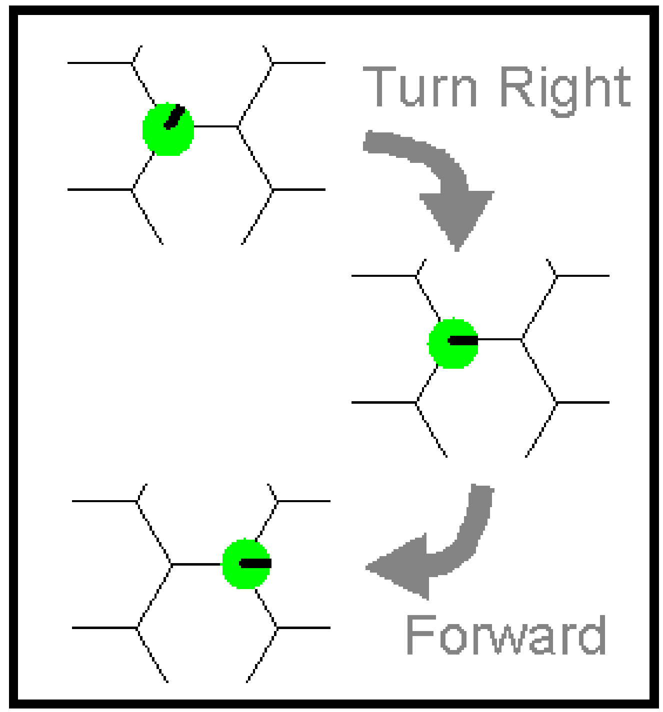
[Courtesy of BCI Lab, Graz.]

Graz Synchronous BCI Timing and Labelling

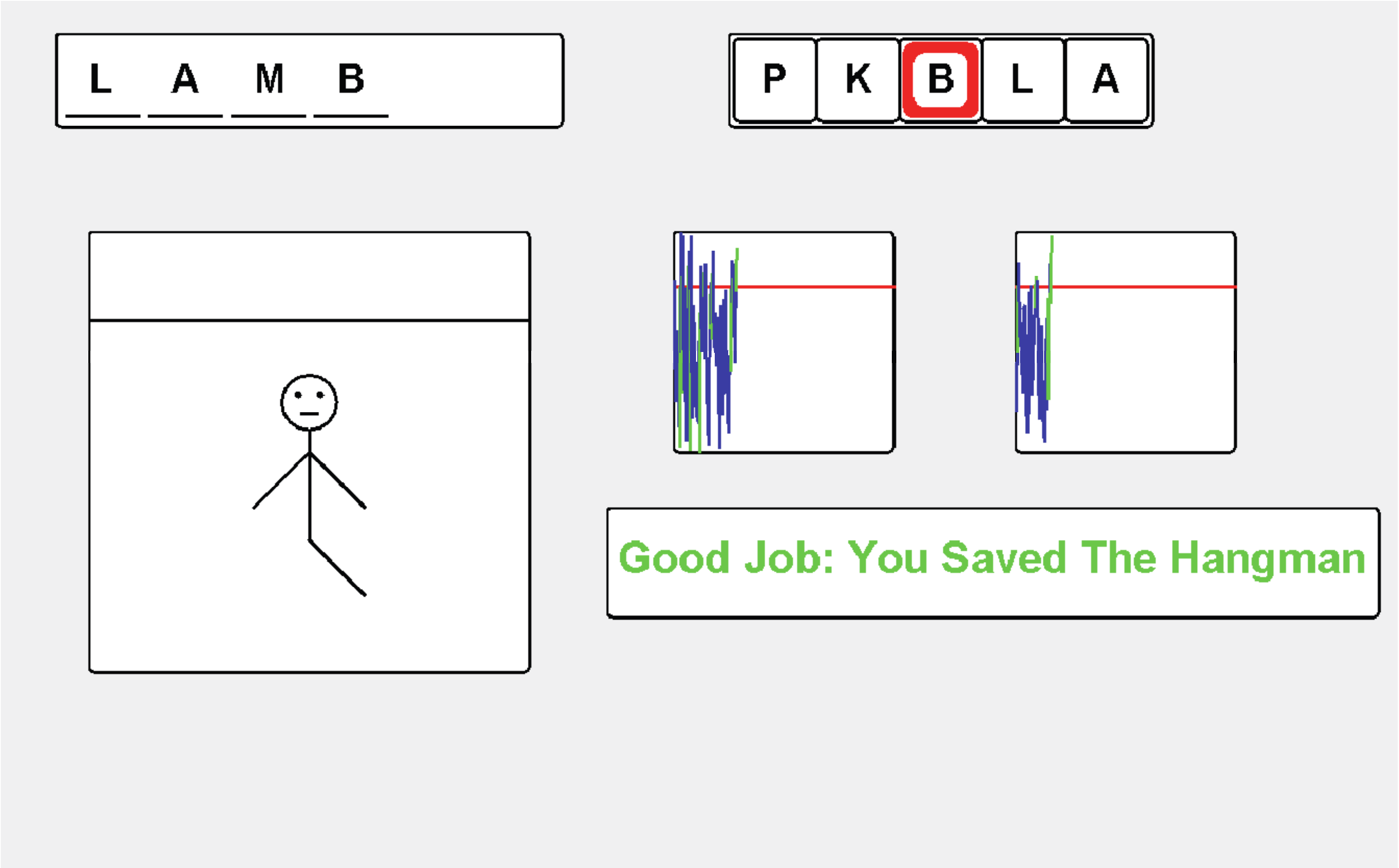


Specially designed scenario (hexagon grid) for online labelling and thus online training of self-paced BCI



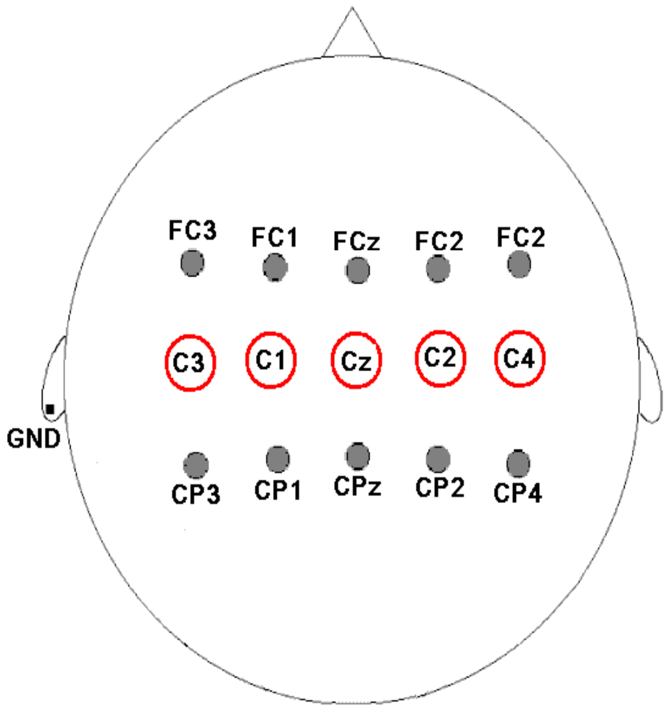


Hangman game for online labelling and thus online training of self-paced BCI



➢ Basic Setup of the Essex ***Self-paced*** Motor Imagery Based Online BCI

EEG data acquisition



5-channel bipolar electrodes; 250Hz

3 motor imagery tasks:

- ➢ left hand → turning left
- ➢ right hand → turning right
- ➢ feet → moving forward

Features:

Selected band power and...

Classification:

LDA classifiers and others.

Key to success:

Online training/adaptation.

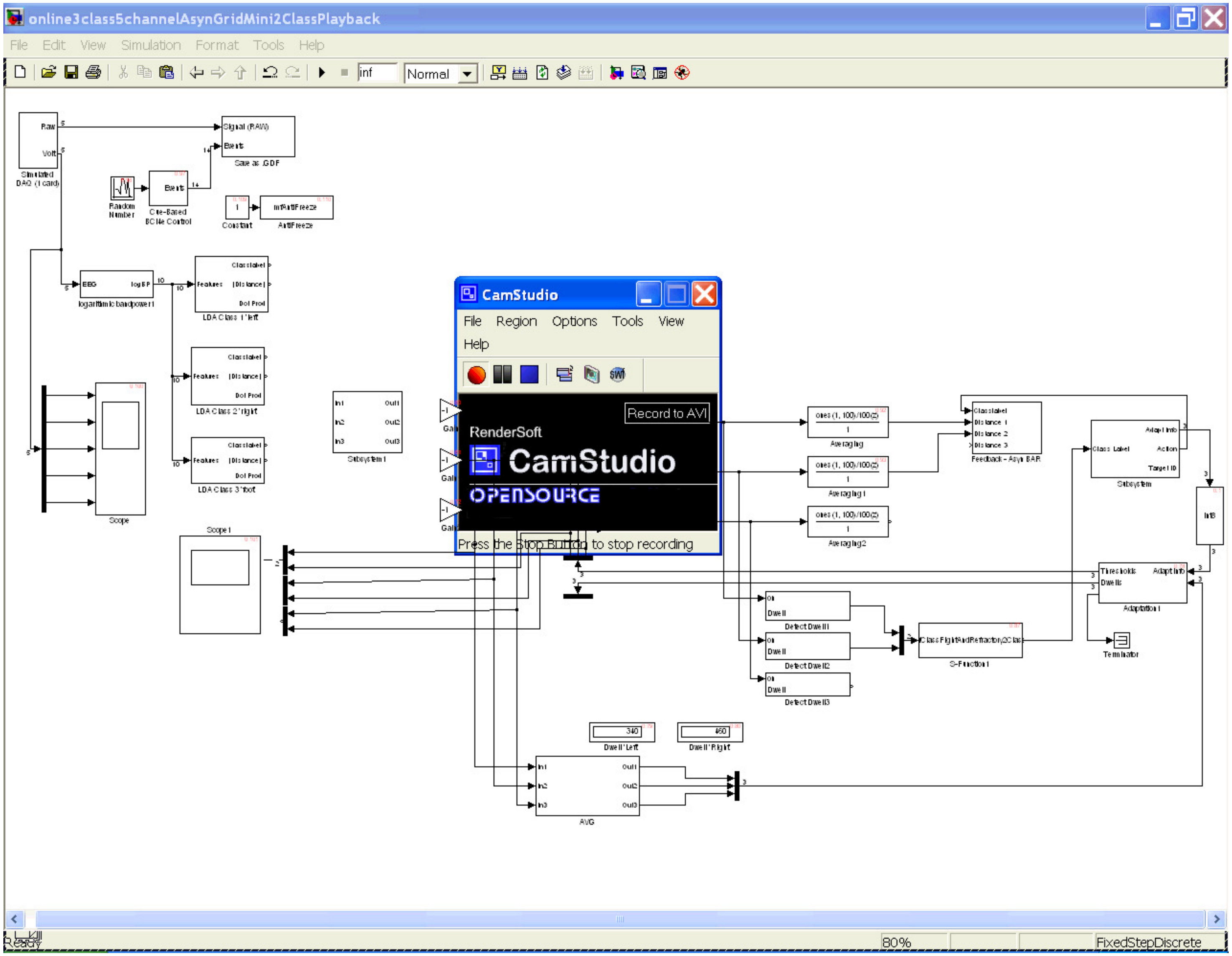
IV. On-line BCI Systems at Essex

Motor Imagery Based
Essex Online BCI for
Simulated Robot Control

By Essex AABAC Team

20th February 2007

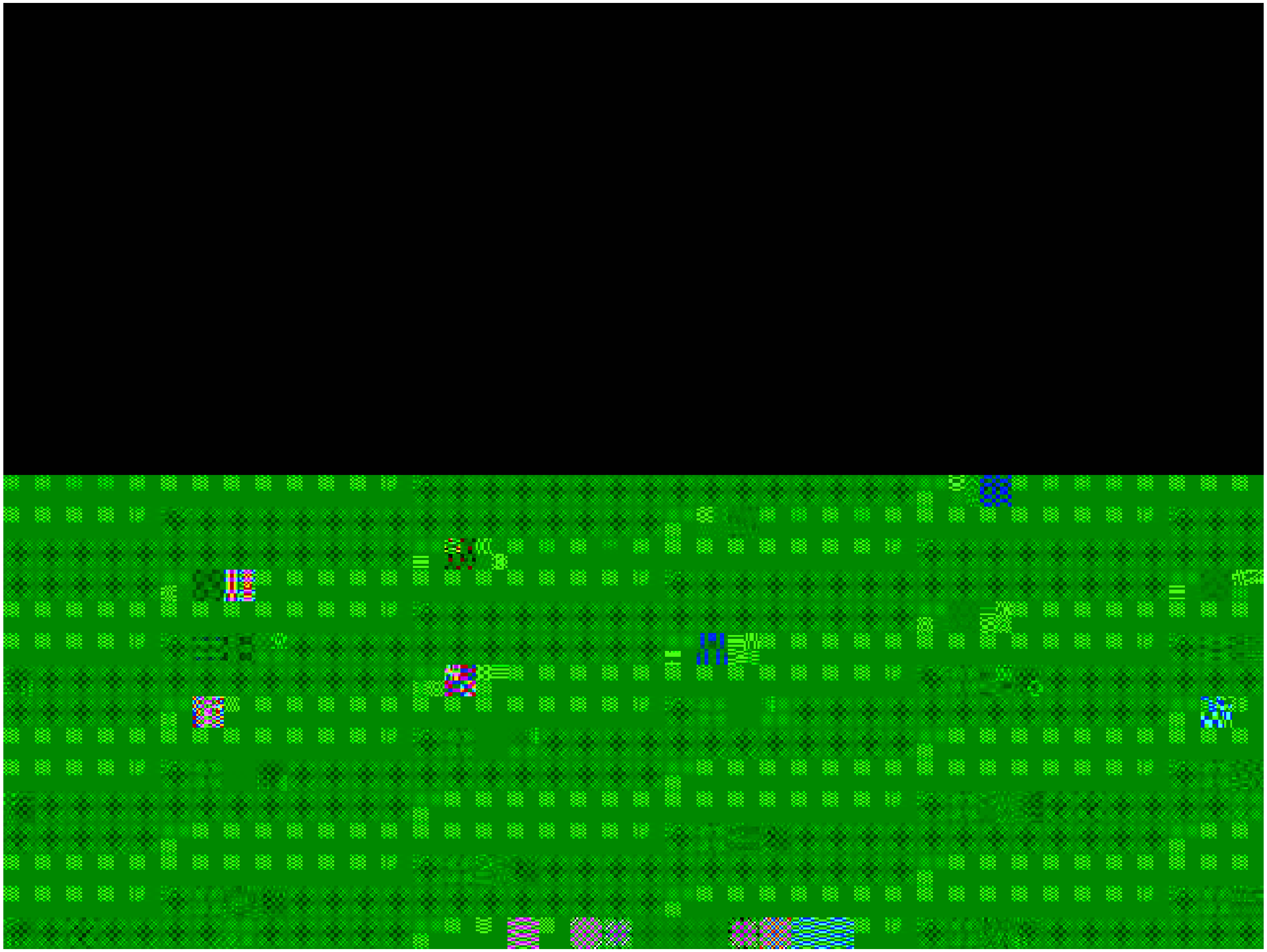
Essex Online Adaptive Self-paced BCI



Essex Online Self-paced BCI for Mobile Robot Control

Essex Self-paced BCI for Wheelchair Control

Essex Self-paced BCI for Playing Hangman Game



V. Other Applications

Early detection of social withdrawal in children

Early detection of learning difficulties in children

Early intervention via biofeedback and its evaluation

Medical diagnosis, Rehabilitation

…… (Any other suggestions?)

➔Salient features at specific locations/frequency bands/time, corresponding to well-designed cognitive tasks

➔ Effective feedback and evaluation

References

J.Q. Gan, B. Awwad Shiekh Hasan, and C.S.L. Tsui, “A filter-dominated hybrid sequential forward floating search for feature subset selection in high-dimensional space,” *Pattern Recognition* (to appear)

B. Awwad Shiekh Hasan and J.Q. Gan, “Conditional random fields as classifiers for 3-class motor-imagery brain-computer interfaces,” *Journal of Neural Engineering* (to appear)

M. Dyson, F. Sepulveda, and J.Q. Gan, “Localisation of cognitive tasks used in EEG-based BCIs,” *Clinical Neurophysiology*, vol. 120, no. 9, pp. 1481-1493, 2010.

B. Awwad Shiekh Hasan and J.Q. Gan, “Unsupervised movement onset detection from EEG recorded during self-paced hand movement,” *Medical & Biological Engineering & Computing*, vol. 48, no. 3, pp. 245-253, 2010.

T. Geng, J.Q. Gan, and H. Hu, “A self-paced online BCI for mobile robot control,” *International Journal of Advanced Mechatronic Systems,* vol. 2, no. 1-2, pp. 28-35, 2010.

J.W. Yoon, S.J. Roberts, M. Dyson, and J.Q. Gan, “Adaptive classification for brain computer interface systems using sequential Monte Carlo sampling,” *Neural Networks*, vol. 22, no. 9, pp. 1286-1294, 2009.

C.S.L. Tsui, J.Q. Gan, and S.J. Roberts, “A self-paced brain-computer interface for controlling a robot simulator: An online event labelling paradigm and an extended Kalman filter based algorithm for online training,” *Medical & Biological Engineering & Computing,* vol. 47, no. 3, pp. 257-265, 2009.

T. Geng, J.Q. Gan, M. Dyson, C.S.L. Tui, and F. Sepulveda, “A novel design of 4-class BCI using two binary classifiers and parallel mental tasks,” *Journal of Computational Intelligence and Neuroscience,* vol. 2008, doi: 10.1155/2008/437306.

S.-M. Zhou, J.Q. Gan, and F. Sepulveda, “Classifying mental tasks based on features of higher-order statistics from EEG signals in brain-computer interface,” *Information Sciences,* vol. 178, no. 6, pp.1629-1640, 2008.

More from http://dces.essex.ac.uk/staff/jqgan/Select_publications_byYear.html