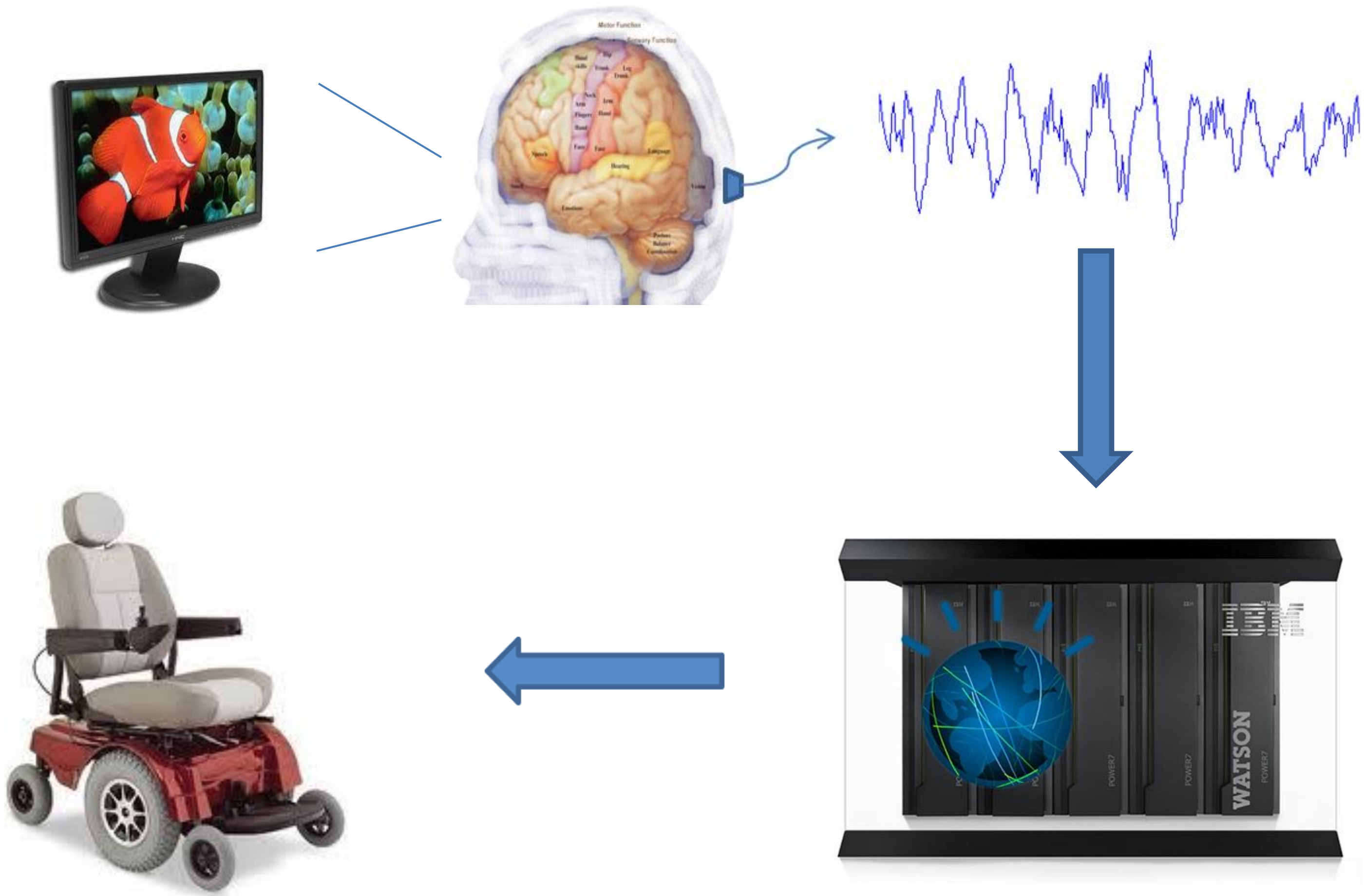
Designing Brain Computer Interfaces Using Visual Evoked Potentials

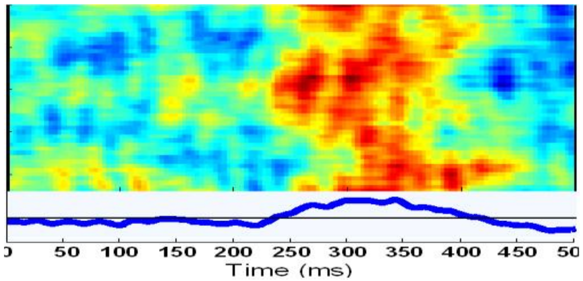
Deniz Erdogmus
Cognitive Systems Laboratory
Northeastern University

The EEG-BCI Concept

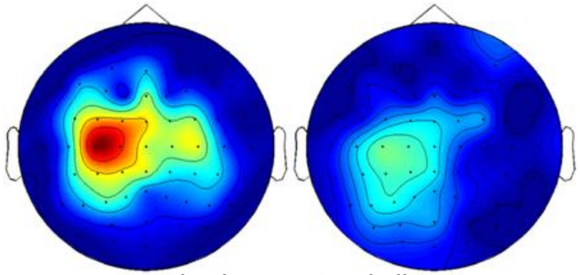


Typical Brain Signals Exploited in EEG-BCI Design

- Visual or auditory evoked P300 response (ERP)

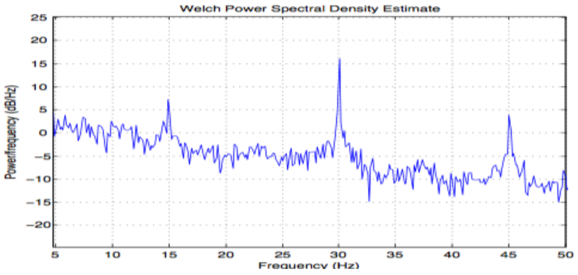


- Motor imagery (MI) or other synchronization / dysynchronization activity in the motor cortex



Thanks: Gerwin Schalk

- Steady state visual evoked potentials (SSVEP)



Qualitative Comparison: Training Time versus Bandwidth

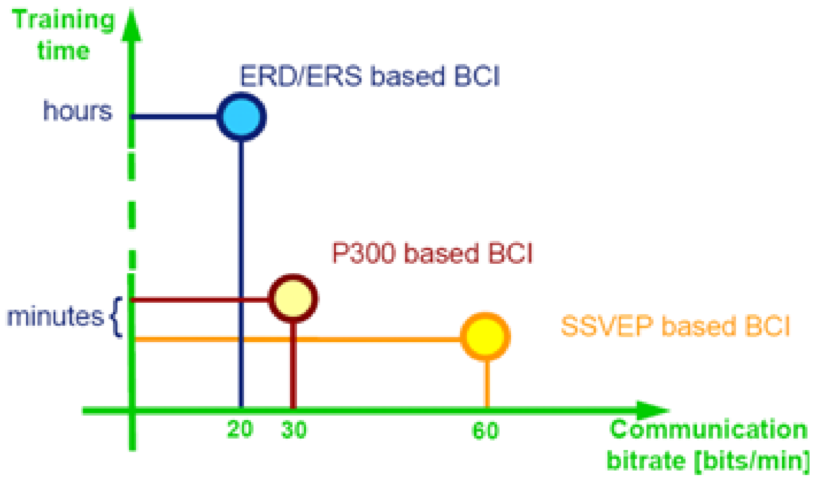


Table. Area under the ROC curve (AUC) for binary intent classification when subject is focused and distracted.

	P3a	MI	VEP
Focused	0.96	0.57	0.94
Distracted	0.91	0.49	0.94

- We performed a binary (left/right) intent communication experiment using these three signals with a **naïve** subject.
  - P3a: Left/right square flashes
  - MI: Left/right hand-tap imagined
  - SSVEP: Left/right square flickers
- Experiment conducted in two modes: focused & distracted
  - Focused: Sit still on chair and focus on task as prompted
  - Distraction: Tap feet on floor as if walking while simultaneously attending task as prompted

Experiment Details for Table on Previous Slide

- P300 task
  - Short white square pulses randomly every [500,600)ms. Left/right flashes desynchronized. Post-stimulus onset 500-ms EEG from [O1, O2, POZ, OZ, FC1, CZ, P1, P2, C1, C2, C3, C4, CP3, CP4] used in RDA.
- MI task
  - Subject visually instructed to imagine left/right hand motion. Bipolar C3-C4 & CP3-CP4 used in RDA with 4-sec windowed PSD features.
- VEP task
  - M-seq VEP obtained using two flickering checkerboards. M-seq 31-periodic, presented at 15bits/sec (decision time of 2.3 seconds) . Template matching classifier using signals from [O1, O2, POz, Oz].

Image Triage BCI using P3 and RSVP

- Present sequence of images using rapid serial visual presentation (RSVP).
- RSVP performed twice to prevent misses.
- Tag potential target images using single-trial P3 detection.
- Result: 6-fold speed-up compared to manual tagging.

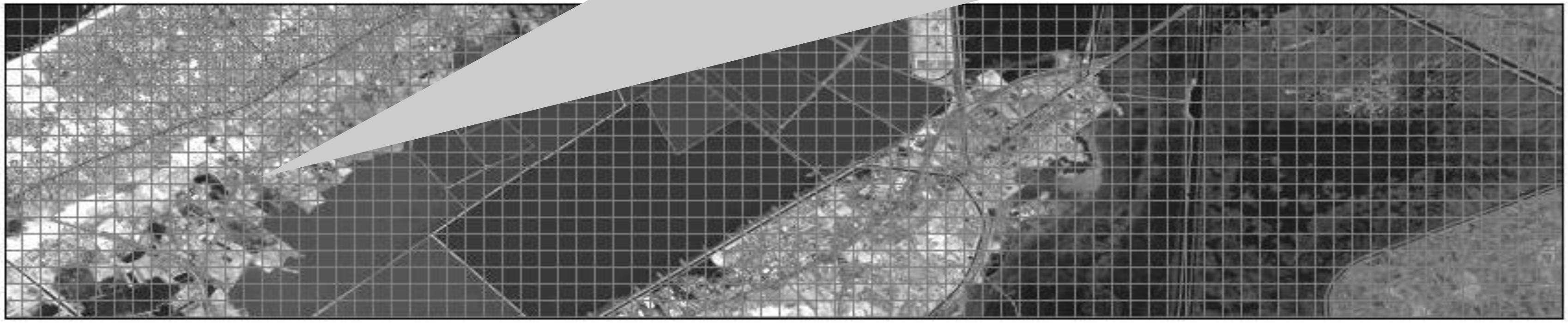
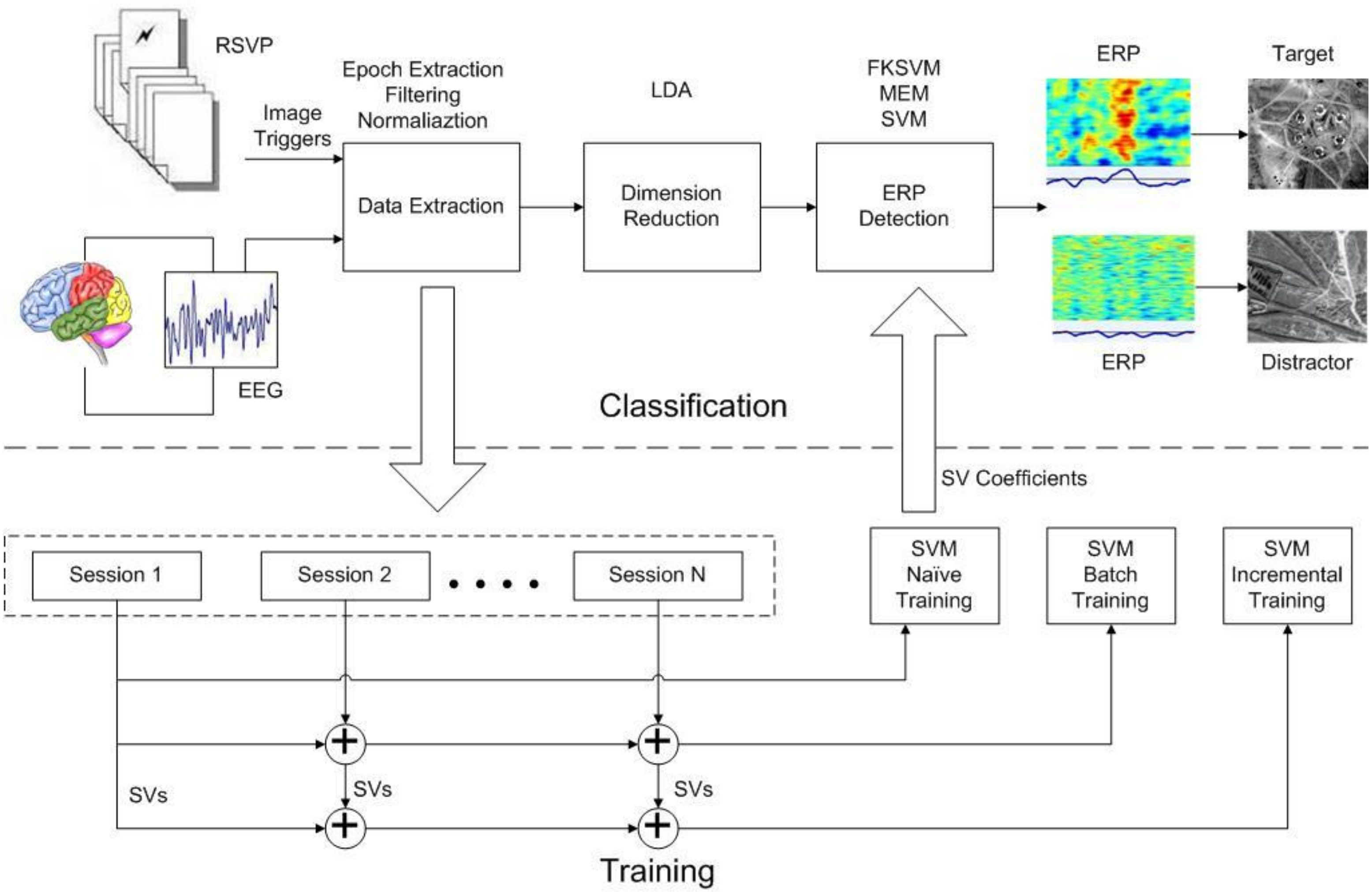
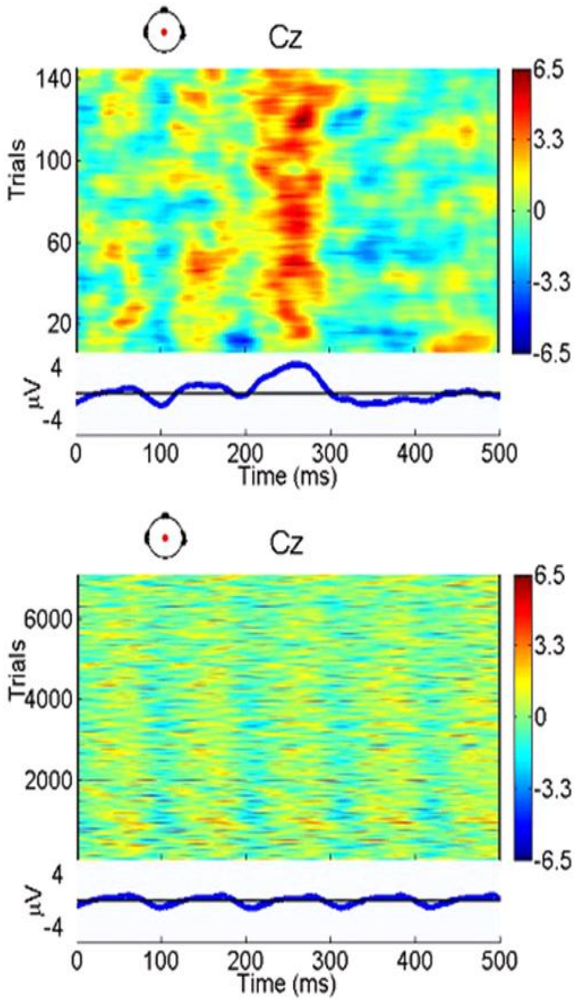
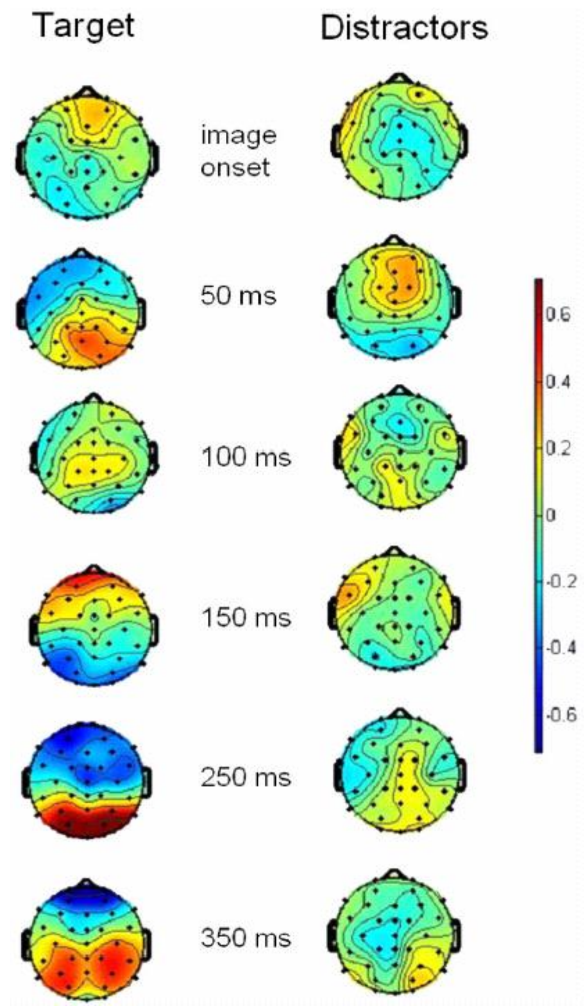


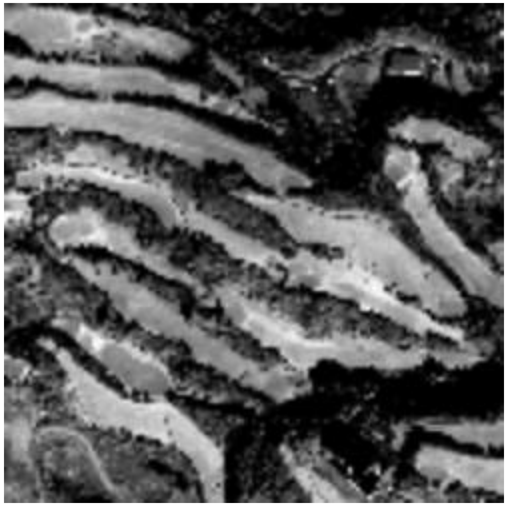
Image Triage BCI System Overview



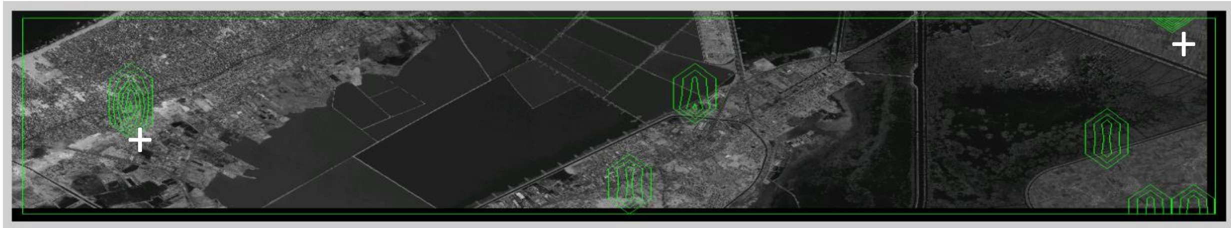
RSVP and P3


Cz
Trials
140
100
60
20
6.5
3.3
0
-3.3
-6.5
µV
4
-4
0
100
200
300
400
500
Time (ms)
Cz
Trials
6000
4000
2000
6.5
3.3
0
-3.3
-6.5
µV
4
-4
0
100
200
300
400
500
Time (ms)

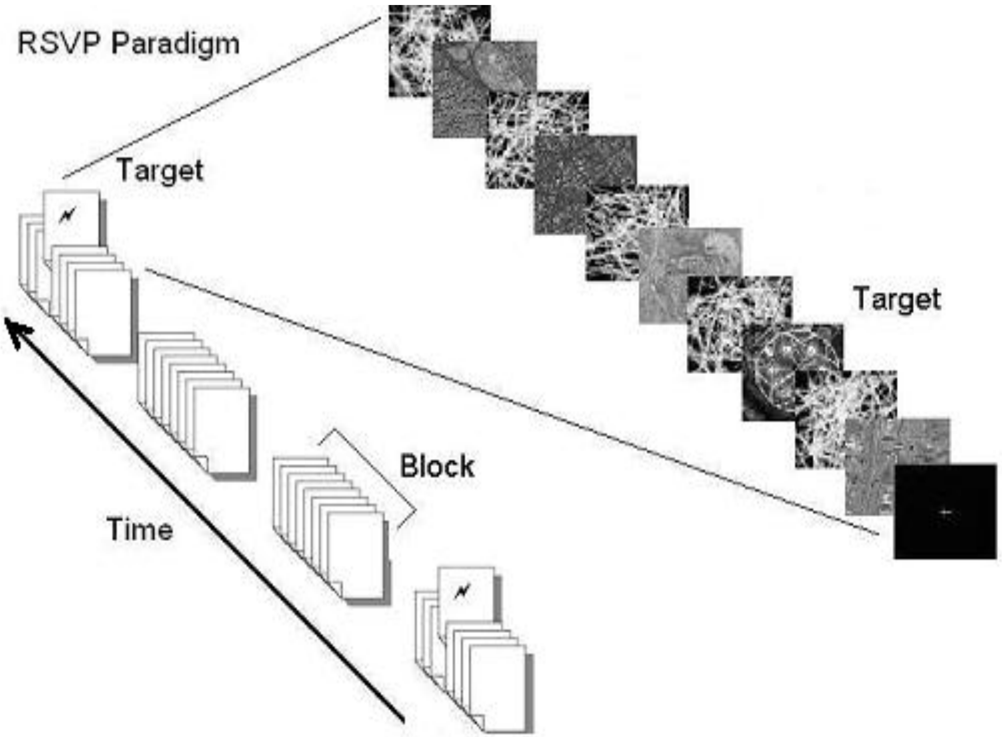

Target
Distractors
image onset
50 ms
100 ms
150 ms
250 ms
350 ms
0.6
0.4
0.2
0
-0.2
-0.4
-0.6

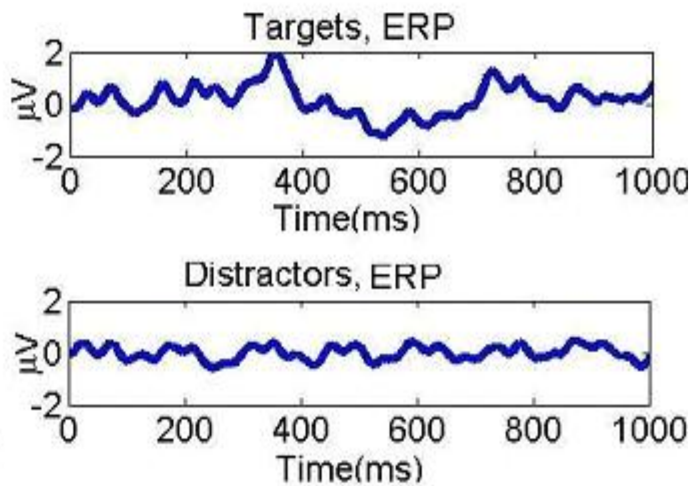


RSVP



- Targets are rare (<1%), non-target distractors are numerous
- Image presentation duration between 50-200 ms/image





Support Vector Machines

- Our baseline classifier is a Gaussian-SVM. SVMs
  - Map input EEG features to a high-dimensional space via kernel eigenfunctions
  - Identify optimal linear binary classification boundary that maximizes the margin
  - Find a small number of support vectors that are closest to the boundary, such that SV-to-boundary distances are equal to this maximal margin

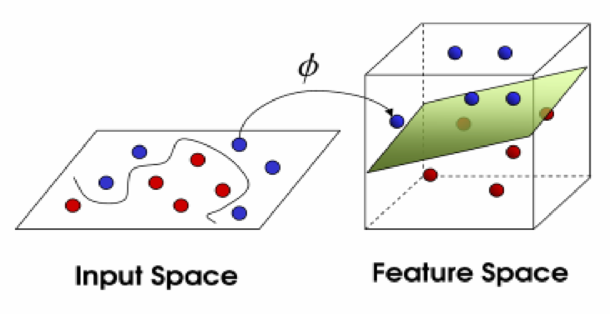
$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 - 2^{-1}\boldsymbol{\alpha}^T \mathbf{D_c K D_c} \boldsymbol{\alpha} - (2C)^{-1}\|\boldsymbol{\alpha}\|_2^2 \quad \text{s.t.} \ \ 0 \le \alpha_i \ \forall i \ \text{and} \ \boldsymbol{\alpha}^T \mathbf{c} \ge 0$$

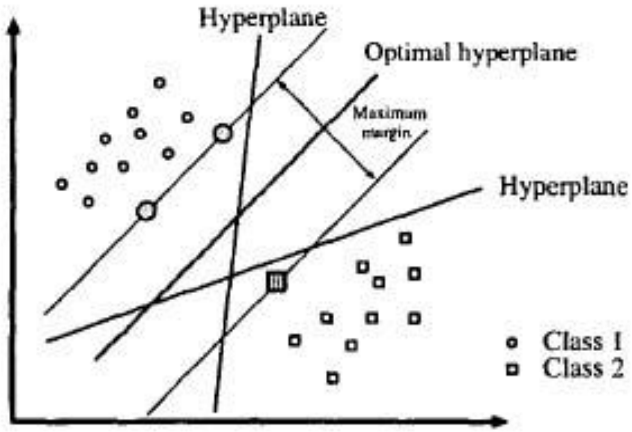
  - Evaluates the class label for new samples by comparing them to these SVs only

$$f(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i c_i < \boldsymbol{\varphi}(\mathbf{x}_i), \boldsymbol{\varphi}(\mathbf{x}) >$$

- Gaussian-SVM and Linear-SVM to detect the presence of P3:

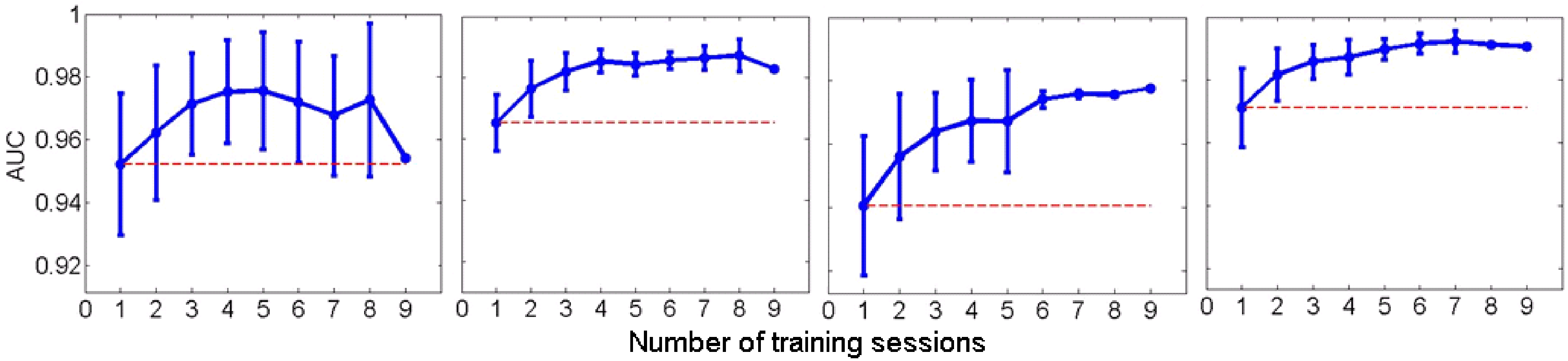
ϕ: **Gaussian kernel eigenfunctions**



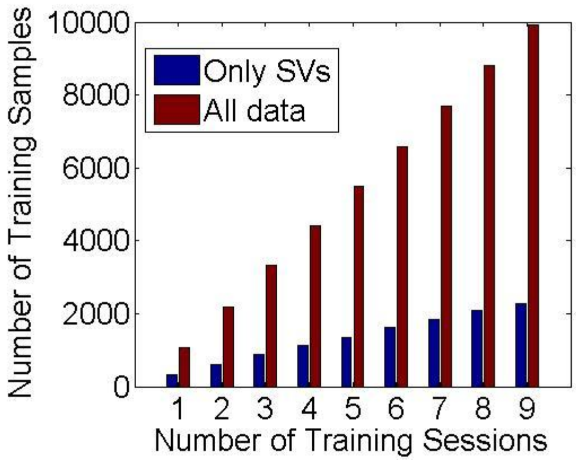


Multi-session Calibration of GSVM P3 Detector

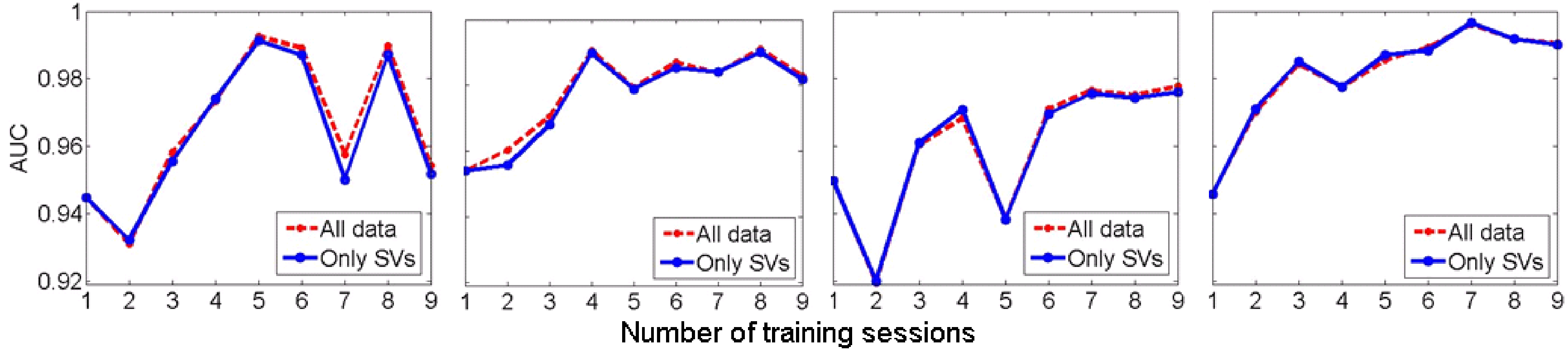
- We performed a BCI image triage experiment using 4 naïve subjects, each for 10 sessions (5 consecutive days, morning/afternoon).
- Trained 9 Gaussian-SVM P3 detectors as follows: $GSVM_i$ is trained on data from sessions $\{1,\ldots,i\}$ and tested on session $(i+1)$.
- The average AUC on the test data is reported here for each subject/session with estimated error bars.



Multi-session Incremental SVM Training



- SVM training complexity increases superlinearly with number of samples.
- Incremental SVM is based on the premise that only support vectors of a previously examined training set are necessary to remember.
- Incremental SVM (iSVM) training proceeds as follows: train iSVM$_i$ using support vectors of {iSVM$_1$,…, iSVM$_{(i\text{-}1)}$} and training set *i*.



Mixed Effects Modeling (MEM) of ERP

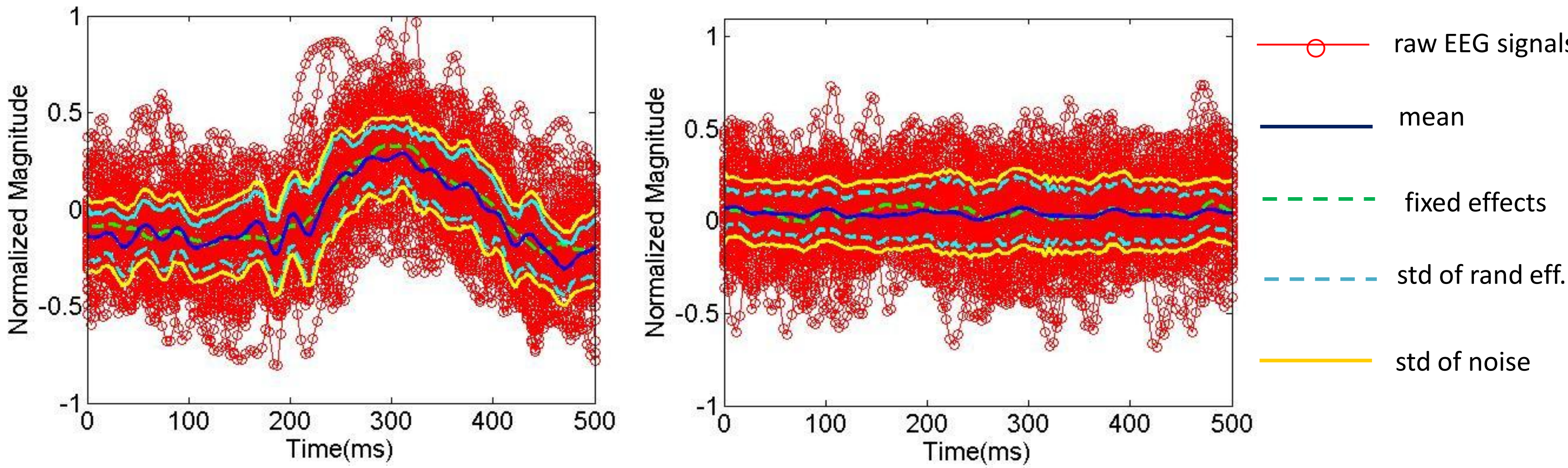
$$\mathbf{y}_i \sim N(\mathbf{X}_i\boldsymbol{\alpha}, \sigma^2\mathbf{I}_{n_i} + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T) \qquad \mathbf{b}_i \sim N(0, \mathbf{D})$$

$$\boldsymbol{\varepsilon}_i \sim N(0, \sigma^2\mathbf{I}_{n_i})$$

$$\underbrace{\mathbf{y}_i}_{\text{observations}} = \underbrace{\overbrace{\mathbf{X}_i}^{\text{design matrix}}\boldsymbol{\alpha}}_{\text{population effects}} + \underbrace{\mathbf{Z}_i\mathbf{b}_i}_{\text{random effects}} + \underbrace{\boldsymbol{\varepsilon}_i}_{\text{errors}}$$

$\mathbf{y}_i$, $\mathbf{X}_i$, and $\mathbf{Z}_i$ are known
$\boldsymbol{\alpha}$, $\mathbf{b}_i$, and $\boldsymbol{\varepsilon}_i$ are unknown

- Single trial ERP/non-ERP responses to each image are variable.
- MEM tries to capture this variability using a simple hierarchical Bayesian approach. In particular, we used a Gaussian graphical model.



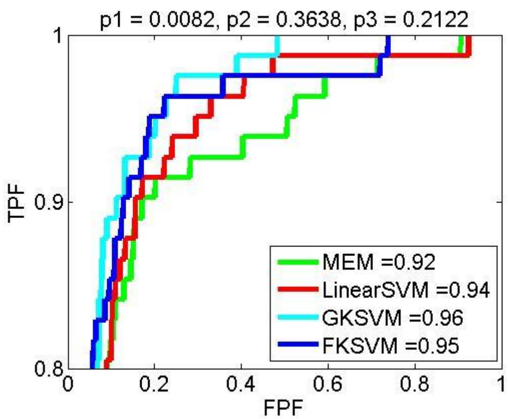
Fisher Kernel

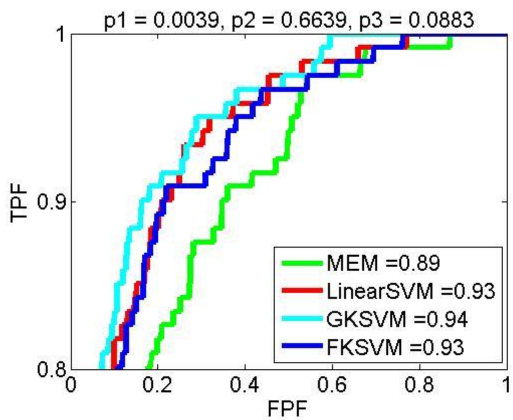
- Fisher score as feature transform: $\mathbf{U_X} = \nabla_{\boldsymbol{\theta}} \log p(\mathbf{X} \mid \boldsymbol{\theta})$
- Fisher information matrix inverse as Riemanian metric: $\mathbf{I}^{-1} = \mathbf{E_X}\left[\mathbf{U_X U_X}^T\right]^{-1}$
- Linear Fisher kernel: $\mathbf{K}\left(\mathbf{x}_i, \mathbf{x}_j\right) = \mathbf{U}_{\mathbf{x}_i} \mathbf{I}^{-1} \mathbf{U}_{\mathbf{x}_j}^T$
- Gaussian Fisher kernel: $\mathbf{K}\left(\mathbf{x}_i, \mathbf{x}_j\right) = \exp(-\mathbf{U}_{\mathbf{x}_i} \mathbf{I}^{-1} \mathbf{U}_{\mathbf{x}_j}^T)$

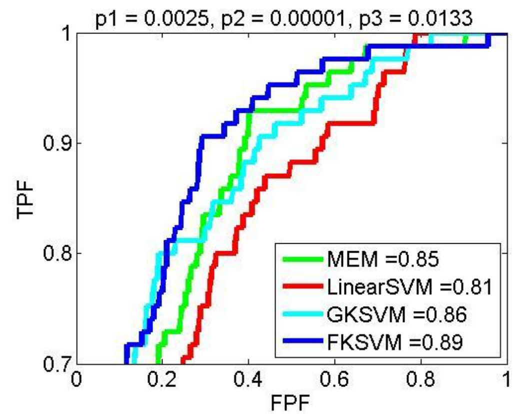
$$\boxed{\mathbf{X} \longrightarrow \mathbf{U_X} \longrightarrow \phi(\mathbf{U_X})}$$

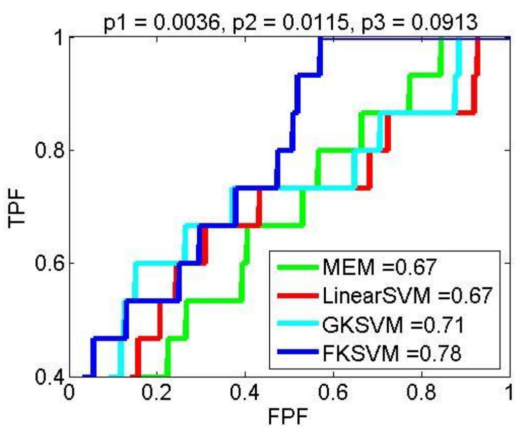
ROC Curves:
MEM Likelihood Ratio, LFK-SVM, L-SVM, G-SVM

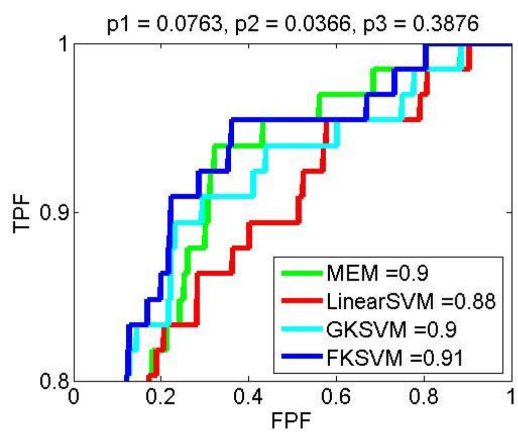
P1: MEM vs. FKSVM; P2: LinearSVM vs. FKSVM; P3: GKSVM vs. FKSVM





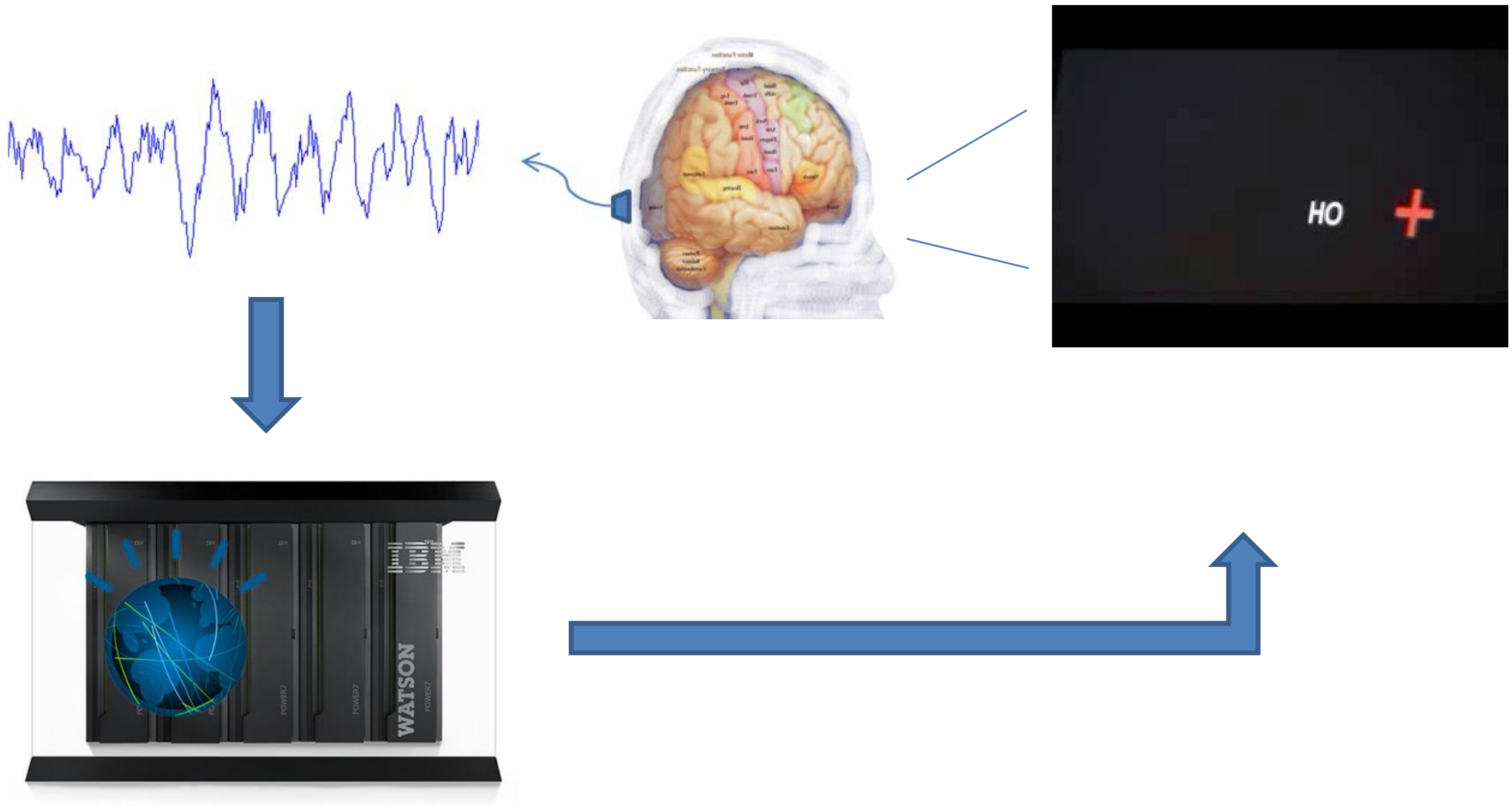






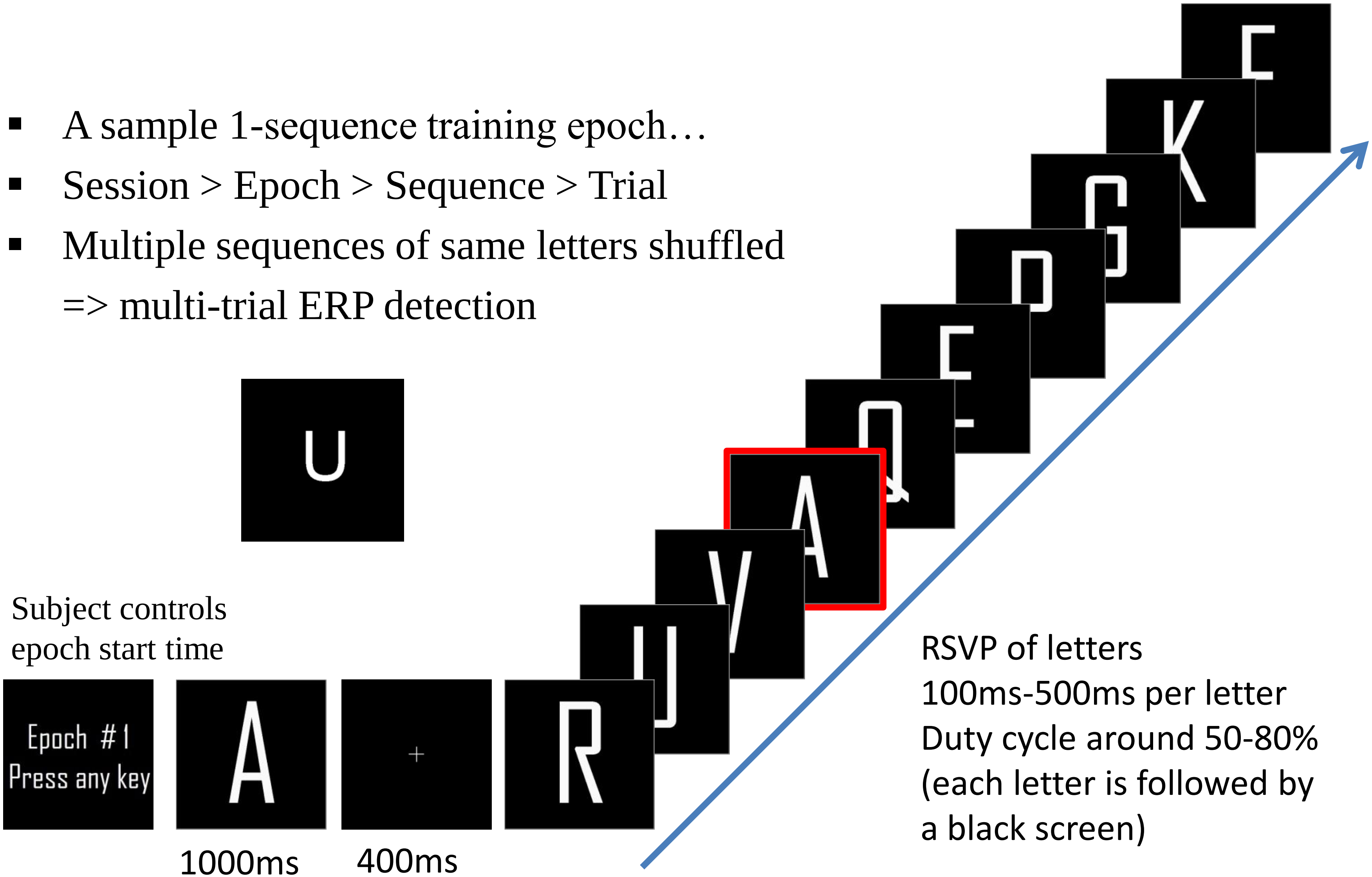
Mean_MEM=0.846
Mean_LinearSVM=0.846
Mean_GKSVM=0.874
Mean_FKSVM=0.892

RSVP Keyboard:
A Spelling Interface based on the P3 Signal



RSVP Keyboard Terminology

- A sample 1-sequence training epoch…
- Session > Epoch > Sequence > Trial
- Multiple sequences of same letters shuffled
  => multi-trial ERP detection

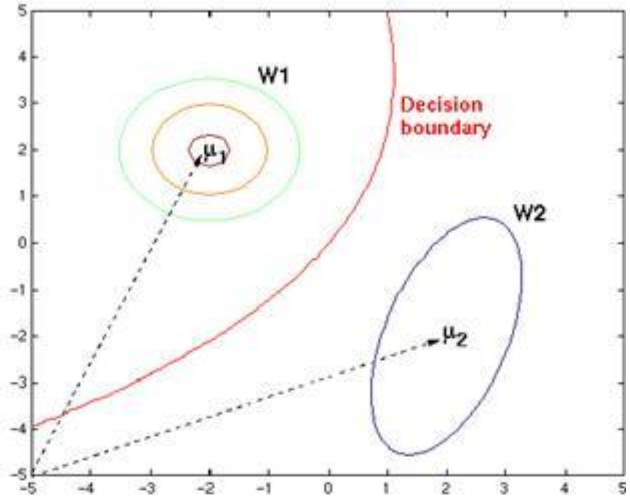


Subject controls
epoch start time

RSVP of letters
100ms-500ms per letter
Duty cycle around 50-80%
(each letter is followed by
a black screen)

Regularized Discriminant Analysis

- **Quadratic Discriminant Analysis:** Bayes classifier assuming feature vectors have multivariate normal distribution for each class.
- **Regularized Discriminant Analysis:** A modification of QDA to eliminate the singularities of covariance matrices.
  - Shrinkage is applied as



$$\hat{\mathbf{\Sigma}}_c(\lambda) = (1-\lambda)\hat{\mathbf{\Sigma}}_c + \lambda\hat{\mathbf{\Sigma}},$$

where λ is the shrinkage parameter; $\hat{\mathbf{\Sigma}}_c$ is the class covariance matrix estimated for class $c \in \{0,1\}$ with $c = 0$ for non-target class and $c = 1$ for target class; $\hat{\mathbf{\Sigma}}$ is the weighted average of class covariance matrices.

  - Regularization is administered as

$$\hat{\mathbf{\Sigma}}_c(\lambda,\gamma) = (1-\gamma)\hat{\mathbf{\Sigma}}_c(\lambda) + \frac{\gamma}{d}\mathrm{tr}[\hat{\mathbf{\Sigma}}_c(\lambda)]\mathbf{I},$$

where γ is the regularization parameter, $\mathrm{tr}[\cdot]$ is the trace function and d is the dimension of the data vector.

Classification Score from RDA

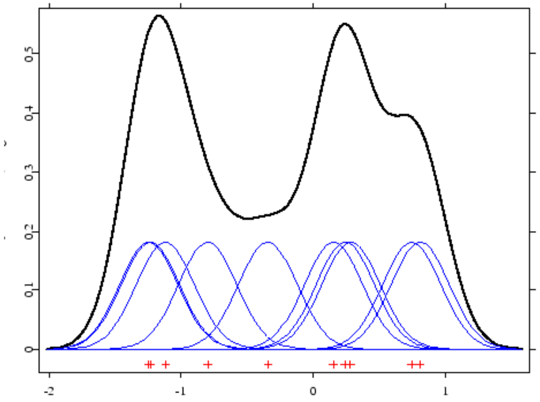
- Log-of-posterior-ratio is used as the ERP classification score to be used in the fusion of language model after estimating means and covariances from the training samples.

$$\delta_{\text{RDA}}(\mathbf{x}) = \log \frac{f_{\mathcal{N}}(\mathbf{x}; \hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\Sigma}}_1(\lambda, \gamma))\hat{\pi}_1}{f_{\mathcal{N}}(\mathbf{x}; \hat{\boldsymbol{\mu}}_0, \hat{\boldsymbol{\Sigma}}_0(\lambda, \gamma))\hat{\pi}_0},$$

where $\boldsymbol{\mu}_c$, $\hat{\pi}_c$ are estimates of class means and priors respectively; $\mathbf{x}$ is the data vector to be classified and $f_{\mathcal{N}}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the pdf of a multivariate normal distribution.

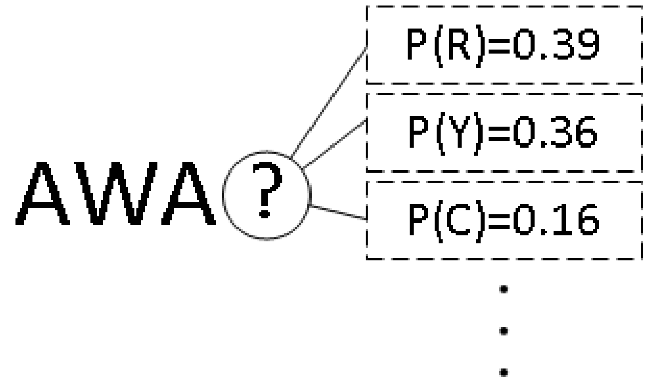
- The parameters of RDA are determined via Nelder-Mead simplex reflection using K-fold validation.

- $P(\delta_{\text{RDA}}(\mathbf{x}_{W_i,n_s}))|c_{W_i} = c)$ is estimated using kernel density estimation on the training data, using a Gaussian kernel whose bandwidth is selected using Silverman's rule of thumb.



Language Modeling

- Language modeling → improve typing accuracy and consequently speed.
- Symbol prediction: The current target symbol is predicted using some of the previously written symbols.
- n-gram sequence modeling:
  Estimates the conditional probability of any letter in a sequence given n-1 previously selected letters using a Markov model of order n-1.



- W is a sequence of letters where W_i is the ith letter in the sequence. For an n-gram model the estimate of the conditional probability of the letter W_i is obtained from

$$P(W_i|W_{i-1}, W_{i-2}, \cdots, W_{i-n+1}) = \frac{P(W_i, W_{i-1}, \cdots, W_{i-n+1})}{P(W_{i-1}, \cdots, W_{i-n+1})},$$

where the joint probabilities are estimated by regularized relative frequency estimation from a large text corpus.

Off-line Analysis by Simulated Typing: Fusing Language Model & EEG Evidence

- RSVP Keyboard makes letter selections based on joint evidence from an n-gram language model at the symbol level and EEG evidence from RSVP of letter sequences as illustrated before.

$$P(c_{W_i} = c | \delta_{\text{RDA}}(\mathbf{x}_{W_i,1}), \delta_{\text{RDA}}(\mathbf{x}_{W_i,2}), \cdots, \delta_{\text{RDA}}(\mathbf{x}_{W_i,N_S}), W'_{i-1}, W'_{i-2}, \cdots, W'_{i-n_{LM}+1})$$

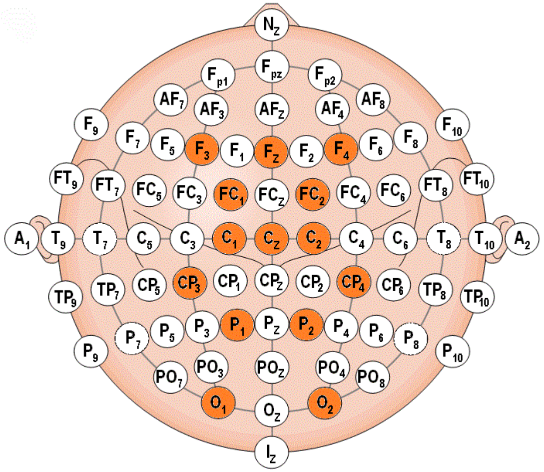
- Language model is trained using one or a combination of Wall Street Journal, Enron Emails, and self-provided previous conversation scripts from subject.

- For off-line analysis of simulated typing data, we use Bayes' theorem to obtain a likelihood ratio test based decision mechanism.

$$L = \frac{\left(\prod_{n_s=1}^{N_S} P(\delta_{\text{RDA}}(\mathbf{x}_{W_i,n_s}) | c_{W_i} = 1)\right) P(c_{W_i} = 1 | W'_{i-1}, W'_{i-2}, \cdots, W'_{i-n_{LM}+1})}{\left(\prod_{n_s=1}^{N_S} P(\delta_{\text{RDA}}(\mathbf{x}_{W_i,n_s}) | c_{W_i} = 0)\right) P(c_{W_i} = 0 | W'_{i-1}, W'_{i-2}, \cdots, W'_{i-n_{LM}+1})}$$

Off-line Analysis by Simulated Typing: Experiment Setup

- Experimental setup:
  - 14 EEG electrodes:



- Simulation study:
  - Subjects are presented multiple epochs consisting of a target symbol to look for, followed by a fixation sign and 3 sequences of randomly ordered 26 letters of the English alphabet with 150 ms inter-stimuli interval.
  - **Assumption:** Each target symbol is assumed to be a part of a typing process of a letter where all the previous n-1 letter are written correctly using BCI.
  - Using 10-fold validation ERP classifier is trained and language model fusion is applied on the test epochs. Area under ROC curves are calculated for different scenarios.
  - Each epoch gives 1 target class sample and 25 non-target class samples.

Off-line Analysis by Simulated Typing: Results

		1 sequence	2 sequences	3 sequences
0-gram		(0.812, 0.884)	(0.907, 0.956)	(0.957, 0.985)
1-gram		(0.892, 0.922)	(0.944, 0.973)	(0.972, 0.986)
4-gram	Word-initial	(0.892, 0.941)	(0.954, 0.983)	(0.977, 0.991)
	Word-internal	(0.975, 0.983)	(0.985, 0.992)	(0.991, 0.997)
8-gram	Word-initial	(0.905, 0.945)	(0.960, 0.984)	(0.979, 0.992)
	Word-internal	(0.991, 0.993)	(0.995, 0.997)	(0.995, 0.998)

The minimum and the maximum values of the area under the ROC curves.

		1 sequence	2 sequences	3 sequences
0-gram		(0.453, 0.548)	(0.700, 0.810)	(0.828, 0.889)
1-gram		(0.556, 0.660)	(0.767, 0.841)	(0.900, 0.953)
4-gram	Word-initial	(0.606, 0.688)	(0.740, 0.884)	(0.886, 0.971)
	Word-internal	(0.842, 0.899)	(0.912, 0.966)	(0.960, 0.989)
8-gram	Word-initial	(0.614, 0.716)	(0.766, 0.905)	(0.899, 0.971)
	Word-internal	(0.951, 0.971)	(0.972, 0.990)	(0.986, 0.996)

The minimum and the maximum values of the detection rates for 5% false alarm rate using fusion classifier under different scenarios.

Closed-loop Performance Analysis: Fusing Language Model & EEG Evidence

- **Main assumption:** Previously written symbols and ERP scores from RDA are conditionally independent given the class label of the candidate symbol.

- **Posterior probability of the class label of a symbol:**

$$P(c_s = c|\delta_{\text{RDA}}(\mathbf{x}_s), \mathbf{W}'_{i-1}) = \frac{\left(\prod_{n_s=1}^{N_S} P(\delta_{\text{RDA}}(\mathbf{x}_{s,n_s})|c_s = c)\right) P(c_s = c|\mathbf{W}'_{i-1}) P(\mathbf{W}'_{i-1})}{P(\delta_{\text{RDA}}(\mathbf{x}_{s,1}), \delta_{\text{RDA}}(\mathbf{x}_{s,2}), \cdots, \delta_{\text{RDA}}(\mathbf{x}_{s,N_S}), \mathbf{W}'_{i-1})}$$

- **Posterior of target symbol:** We assume there is one and only one target symbol for each epoch and labels for symbols are independent given all the evidence.

$$P(W_i = s|\delta_{\text{RDA}}, \mathbf{W}'_{i-1}) = \frac{P(c_s = 1|\delta_{\text{RDA}}(\mathbf{x}_s), \mathbf{W}'_{i-1}) / P(c_s = 0|\delta_{\text{RDA}}(\mathbf{x}_s), \mathbf{W}'_{i-1})}{\sum_{t \in \mathcal{S}} P(c_t = 1|\delta_{\text{RDA}}(\mathbf{x}_t), \mathbf{W}'_{i-1}) / P(c_t = 0|\delta_{\text{RDA}}(\mathbf{x}_t), \mathbf{W}'_{i-1})}$$

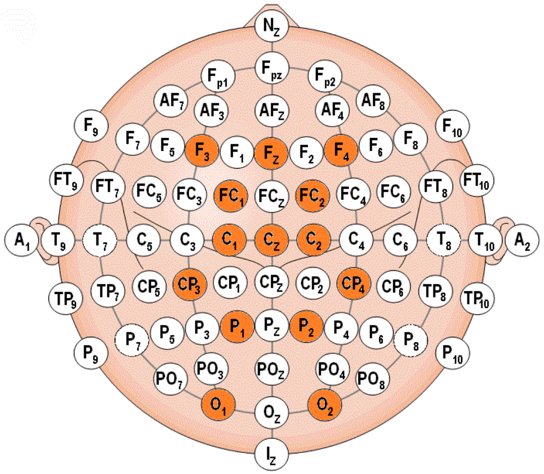
- **Final Decision Rule:**

$$\hat{W}_i = \arg\max_{s \in \mathcal{S}} P(W_i = s|\delta_{\text{RDA}}, \mathbf{W}'_{i-1})$$

- **Confidence Metric:** $P(W_i = \hat{W}_i|\delta_{\text{RDA}}, \mathbf{W}'_{i-1})$ is used as a stopping criterion for the epoch with a threshold of 0.9.

Closed-Loop Typing Experiments

- Experimental setup:
  - 14 EEG electrodes:



- Training session 50 epochs x 3 sequences.

- Backspace symbol probability for language model: 0.1

- Typing experiment: Free typing with correction.

- Number of sequences per epoch is upper bounded by 6.

- The epoch is stopped if posterior of the symbol ≥ 0.9.

Closed-Loop Typing Results

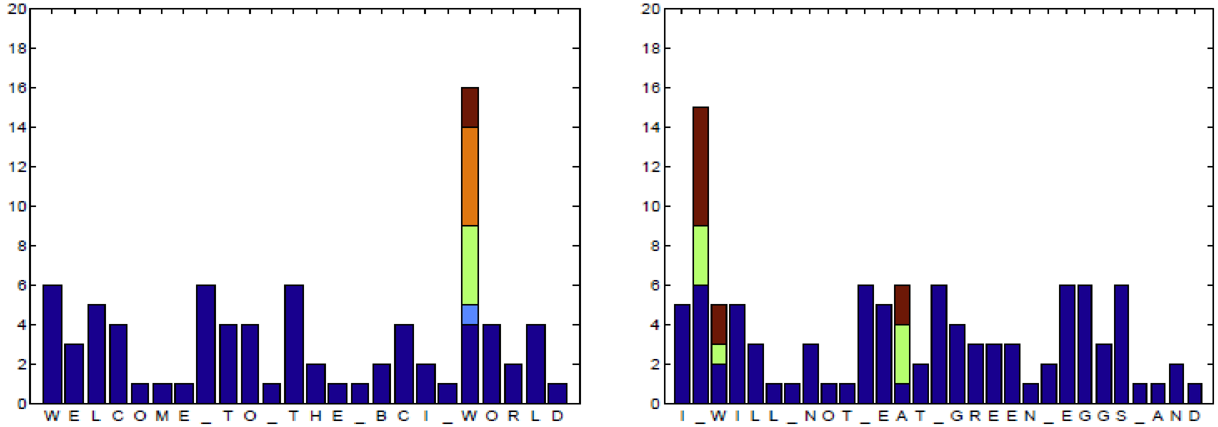


Fig. 2. HS1 number of sequences used to type each symbol.

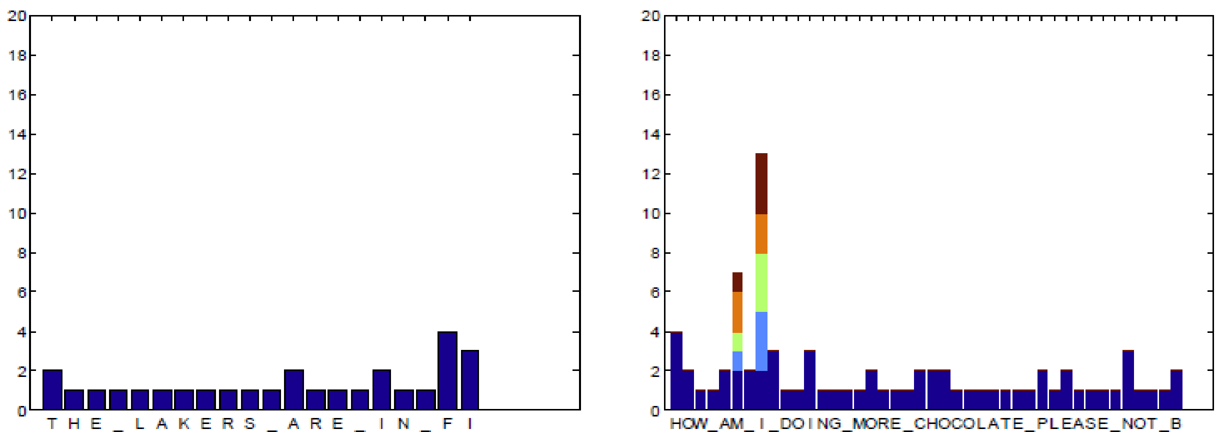


Fig. 3. HS2 number of sequences used to type each symbol.

HS1:

2.93 seqs/symbol

1.58 seqs/symbol

HS2:

1.40 seqs/symbol

3.06 seqs/symbol

LS1:

6.12 seqs/symbol

“HELLO_THERE_PEOP”

3.69 seqs/symbol

“THIS_IS_FAMILY”

RSVP Keyboard: Videos of Locked-in Consultant Typing

- 2011-02-02 Preparing for a session
- 2011-05-01 A typical calibration phase
- 2011-02-02 Locked-in subject G
  - Calibration; subject's face
  - Free typing; screen
- 2011-02-02 Healthy subject K
  - Free typing; subject's face
  - Free typing; screen
- 2011-06-21 Computational linguistics conference (Portland, OR)
  - OPB News
  - Fox12 News

SSVEP-BCI

This paradigm exploits visual cortex response patterns observed due to periodic flickering of visual stimulus (spatial pattern or light source).

Frequency and pseudorandom binary sequence (PRBS) variations are used.

Frequency

- Temporal stimulus pattern is:

  0101010101010…

- Different symbols have different bit presentation rates.
- Frequency resolution of PSD estimation imposes a limit.
- Artifacts and background brain activity overlap with stimulus response in Fourier domain.

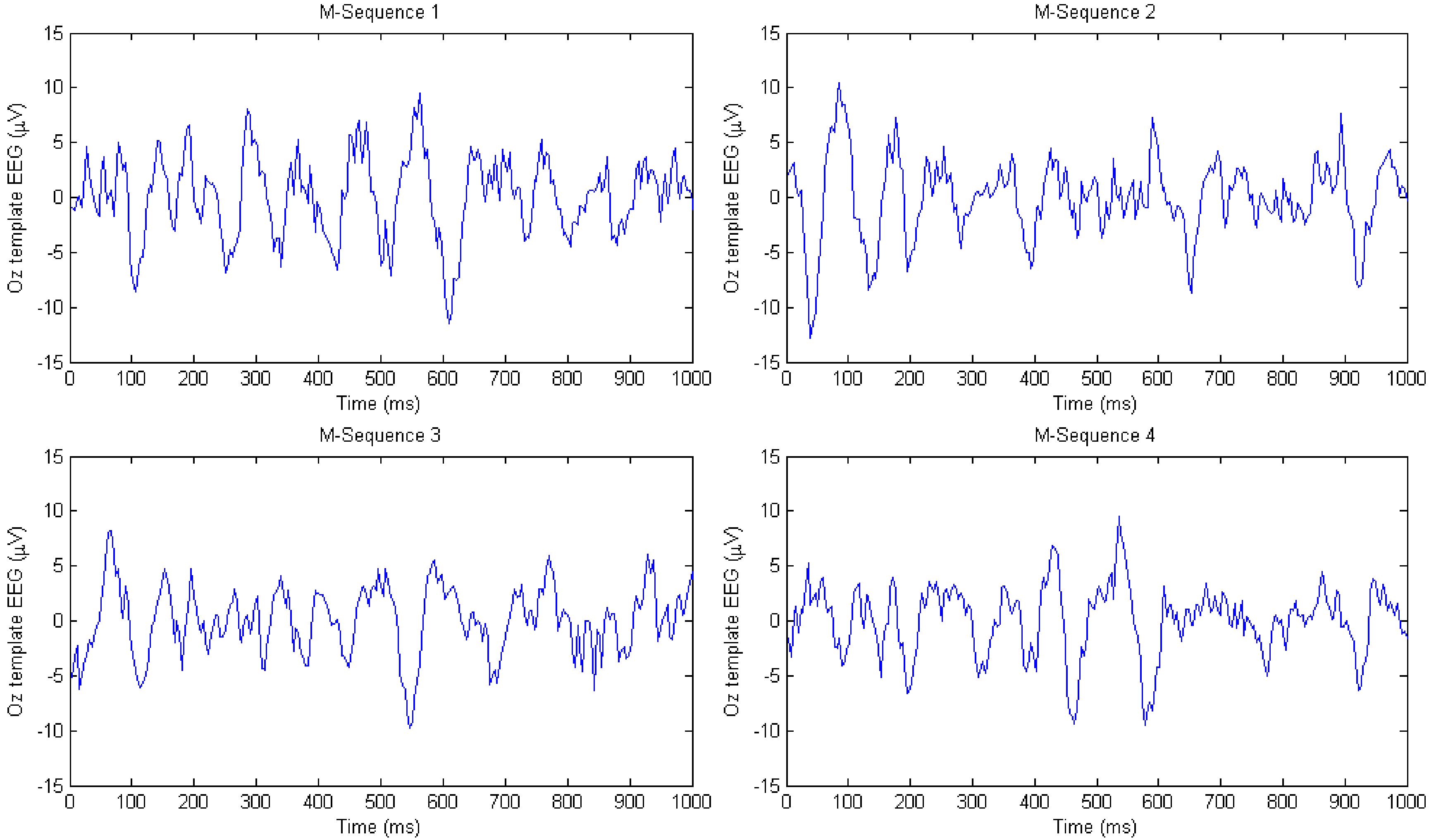
PRBS

- Temporal stimulus pattern is ~:

  1110101000010 0…

- Different symbols have different bit sequences.
- Number of distinct sequences imposes a limit.
- Codes are ultra-wideband. Narrow-band artifacts present a smaller problem.

Average Oz Response to M-sequence Stimuli



Single-channel Template Matching Classifier Accuracy as a Scalp Distribution

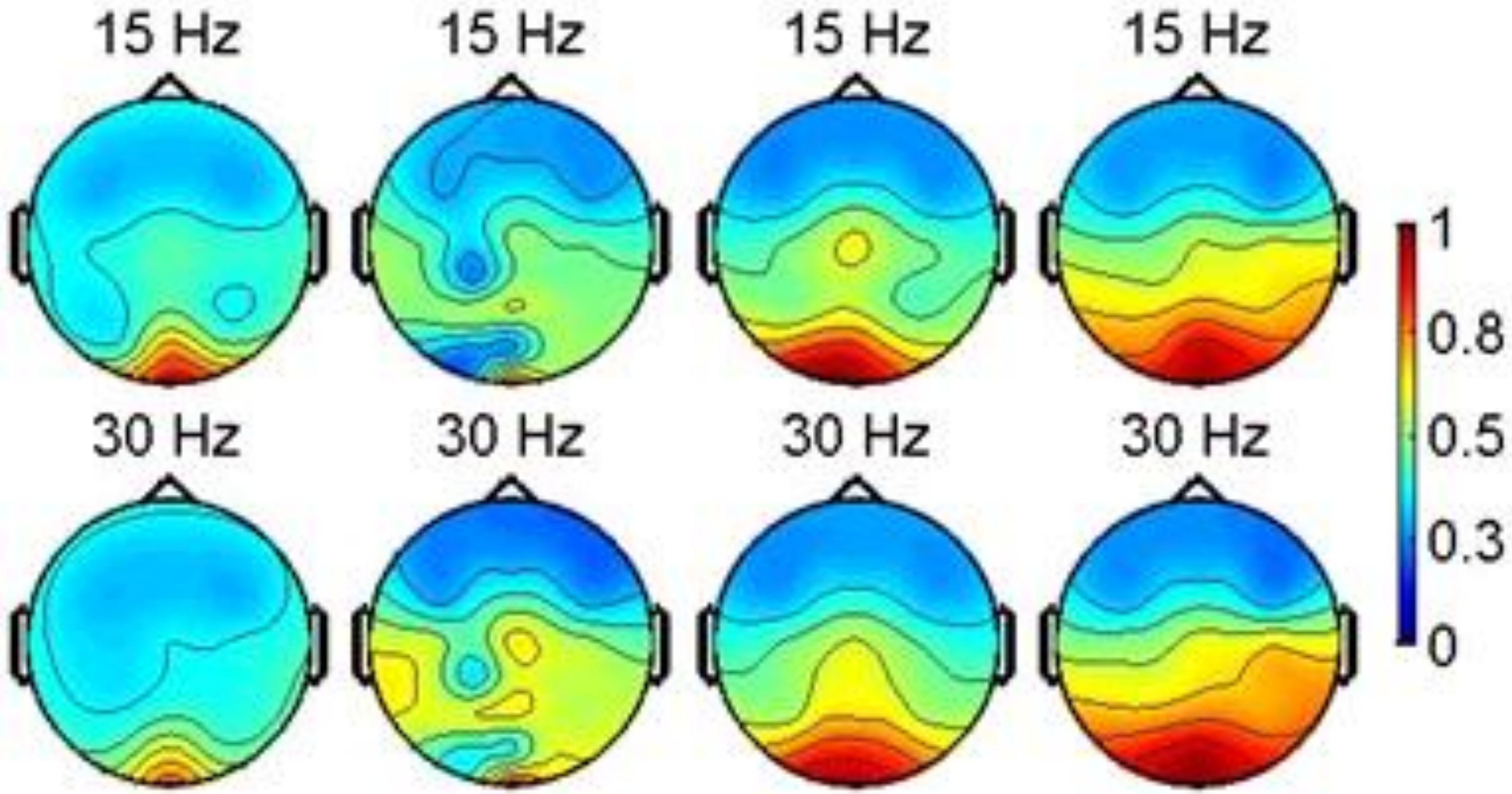


Figure. Spatial distribution of single-channel correct classification probability among 4 m-sequences for 4 subjects, 15 & 30 bits/sec.