Biomedical Signal Processing and Computation

Zoran Nenadic, DSc

Department of Biomedical Engineering
Department of Electrical Engineering and Computer Science—by courtesy

EECS Colloquium
October 08 2008
Biomedical signals are generated by biomedical systems. These systems measure biomedical variables, which may be influenced by environmental/external variable(s). The signals recorded by sensors (measuring device) represent actions such as action potentials.
Properties of Biomedical Systems and Signals

Biomedical systems are:
- dynamic
- nonlinear
- stochastic
- non-stationary (time varying)
- spatially distributed
- most have multiple signaling pathways
- some of them are oscillatory

Consequently, biomedical signals are:
- of limited bandwidth with noise that is colored
- complex (attractors, chaos, fractals)
- noisy
- non-stationary
- high dimensional
- redundant (spatial correlations)
- some of them are periodic
Example 1  Neuronal action potentials

(data from monkey brain)

Temporal resolution <1 ms
Spatial resolution ~20 µm
Example 2  
Electroencephalogram (EEG)

(data from human brain)

Temporal resolution ~10 ms  
Spatial resolution ~2 cm
Example 3  Aortic blood velocity

(data from pig aorta)

Temporal resolution ~50 ms
Spacial resolution ~10 cm
Example 4  Intra-cranial pressure (ICP)

(human data--TBI patient)

Temporal resolution \sim 100 \text{ ms}
Spatial resolution ?
Research in this area is highly cross-disciplinary!
Action Potentials (APs)

Layer 5 adult cat pyramidal neuron


Extracellular recording

extracellular AP
Extracellular recordings can be pretty difficult!

Challenges:

- time consuming
- depends on the experience of the operator
- unmanageable for multiple electrodes
- requires constant human supervision and intervention

What can engineers do to help?

- automate the process of recording
- develop robust, adaptive signal processing tools
- uncover novel information from the data
- based on this give recommendations for electrode design and sensor layout
What problems do we need to solve for this to work?

To facilitate autonomous operation of the algorithm, all signal processing steps must be unsupervised.
Unsupervised AP detection and classification
Robust Unsupervised Detection of APs Using the Continuous Wavelet Transform


The presence of APs (spikes) must be detected in this sampled data.

The spike arrival times need to be estimated.

This is achieved through spike detection.
Challenges:

- Spike shapes will vary significantly over electrode’s movement range → unsupervised detection

- Spike amplitude (and therefore SNR) will vary significantly over electrode’s movement range → robust detection
Signal Detection Theory

Statistical detection theory is based on the hypothesis testing:

\[ \mathcal{H}_0 : x[n] = w[n] \quad \text{no signal present} \]
\[ \mathcal{H}_1 : x[n] = s[n] + w[n] \quad \text{signal + noise present} \]

The rejection/acceptance of \( H_0 \) based on the sufficient statistic \( T(x) \)

\[ T(x) = x[n] \quad \text{amplitude detection} \]
\[ T(x) = \sum_{n \in W} x^2[n] \quad \text{power detection} \]
\[ T(x) = \sum_{n \in W} x[n] s[n] \quad \text{matched filter} \]
\[ T(x) = \sum_{n \in W} x[n] \psi[n] \quad \text{wavelet detection} \]
Why wavelets?

There exist wavelet basis functions that provide a sparse representation of APs.

Wavelet functions—parameterized by scales and translations.

Vast majority of action potentials primate cortex are highly localized in time (0.5 – 1.0 ms)

This determines relevant scales.

Time–frequency analogy.
Five steps of wavelet detection:

1. **Step 1**: Perform multi-scale decomposition of the signal over the set of relevant scales.

2. **Step 2**: Separate the signal and noise at each scale using non-linear filtering techniques (Donoho & Johnstone, Biometrika, 1994).

3. **Step 3**: Perform Bayesian hypothesis testing.

4. **Step 4**: Combine decisions across scales.

5. **Step 5**: Estimate spike arrival times.
Spike Classification

Purpose: identify the sources of individual spikes in data containing multi-neuron activity.

Three steps of AP classification:
1. Spike Alignment
2. Feature Extraction
3. Model-based Clustering
Model–based Clustering with Gaussian Mixtures

Traditional clustering based on heuristic criteria, e.g.

- hierarchical clustering  
- k–means clustering  
  (Hartigan, *Clustering Algorithms*, 1975)

Deficiencies: can’t determine # of classes in the data, can’t handle outliers

Probabilistic framework → features are sampled from unknown distribution.

The corresponding density is modeled as a linear combination of an unknown number $G + 1$ of component densities $p_j$.

\[ \pi_j \triangleq P\{f_i \text{ drawn from } p_j\} \]

the density at $f_i$:

\[ p(f_i|\Pi, \Theta) = \sum_{j=0}^{G} \pi_j p_j(f_i|\theta_j) \]

mixture likelihood:

\[ \mathcal{L}_{\text{MIX}}(\Pi, \Theta|F) = \prod_i \sum_{j=0}^{G} \pi_j p_j(f_i|\theta_j) \]
Estimate the optimal parameters $\Pi^*$ and $\Theta^*$ of the mixture $\mathcal{L}_{\text{MIX}}$ through the Expectation–Maximization (EM) algorithm.

The class membership is decided via (hard clustering rule):

$$ f_i \in \text{Class } j^* \iff j^* = \arg \max_{0 \leq j \leq G} p_j(f_i | \theta_j^*) $$

Still, the number of clusters $G$ is unknown, and has to be found.

$$ \text{BIC}_G \triangleq 2 \log p(F | M_G, I) \approx 2 \log \mathcal{L}_{\text{MIX}}(\Pi^*, \Theta^* | F) - \nu_G \log s \quad \forall G $$

Optimal model is the one with the largest value of BIC.

In statistics this is called model selection problem.

MATLAB examples
Example of model selection

Data

BIC =

-1358.2801 (winner)  
-1371.5877  
-1403.0384  
-1424.8303
BIC =

-1518.6513
-1432.1608 (winner)
-1464.9499
-1490.1445
Signal Quality Metric

- choose signal quality metric, e.g. peak-to-peak amplitude or SNR.

- if SNR chosen $\rightarrow \text{SNR}_i \propto d_i$.

- evaluate signal quality over clusters $S_1, S_2, \ldots$, and select the dominant cluster, i.e. the cluster that provides the maximal average signal quality.

- spikes within the dominant cluster provide multiple observations of the objective function.
Q: How to find the maximum of the regression function from noisy observations?

A: Stochastic optimization.
Stochastic Optimization
(Kiefer & Wolfowitz, Annals of Math. Stat., 1952)

Find: \[ x^* = \arg \max \{ E_\omega[f(x, \omega)] \} \]

optimal electrode position

\[ x_{k+1} = x_k + \rho_k \xi_k \quad k = 0, 1, \ldots \]

(stochastic gradient ascent)

noisy observations of objective

The sequence \( \{\rho_k\} \) can be found so that \( x_k \to x^* \) with probability 1.

Problem: unbounded variance \( \text{Var}\{\xi_k\} \to \infty \) near peak implies excessive dithering–like movements.
Stochastic Optimization (basis function approach)

Key idea: estimate objective function \textit{adaptively}:

\[ E_\omega[f(x, \omega)] \approx \sum_{j=1}^{n} a_j \Psi_j(x) \]

Key challenge: choose \( n \) to avoid over-fitting. Bayesian probability theory used.

Two steps:

1) \textbf{Model Selection} (choose the order \( n \))
   - given a family of models \( \{M_1, M_2, \ldots, M_N\} \), find the optimal model order.

   \[ P(M_n | Y, I) = \frac{p(Y | M_n, I) P(M_n | I)}{p(Y | I)} \]

   For polynomial \( \Psi_j \) the posterior can be found \textit{analytically} (Nenadic & Burdick, \textit{IEEE Trans. Biomed. Eng.}, 2006).

   Optimal model:

   \[ n^* = \arg \max_{1 \leq n \leq N} P(M_n | Y, I) \]
2) **Parameter Estimation**: least squares estimate of the model $M_{n^*}$.

$$\{a_1^*, ..., a_n^*\} = \arg \max \sum_{i=1}^{k} \|M_{n^*}(x_i, a) - y_i\|^2$$

Newton’s method:

$$x_{k+1} = x_k + |H_k|^{-1} \xi_k \quad k = 0, 1, \ldots$$

$$\sum_j a_j^* \Psi_j''(x_k) \quad \sum_j a_j^* \Psi_j'(x_k)$$

Applications beyond movable electrode (2-D, 3-D)?
Experimental Results

Custom-made motorized microdrive:

(Cham, Branchaud, Nenadic, Greger, Andersen & Burdick, *J. Neurophysiol.*, 2005)

Data from monkey cortex
(Wolf et al., IEEE Int. Conf. Rob. Auto., 2008)

data from monkey cortex