Tfy-99.275 lecture 4

Modelling,
event- & trend detection

System investigation

- To obtain more knowledge about the behaviour of physiological systems that generate the signals we record, we can make (parametric) models and examine how their parameters change for different states of the physiological system.
- We can then use the parameters as input to classification/diagnosis methods.

general:

signal → rules or equations → derived information (e.g. model parameters)

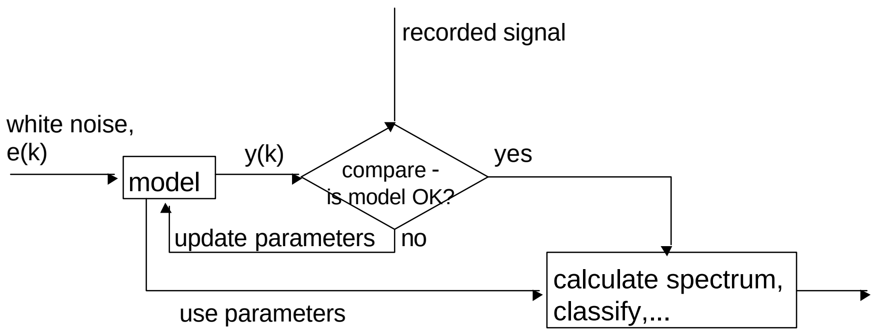
example, Fourier model:

signal → Fourier series analysis eqs. → $a_0, b_0, a_1, b_1, ..$ $\ldots, a_n, b_n$

Parametric methods

- assume the recorded time series to be the output of a given mathematical model (no assumptions about data outside the window are needed)
- Usage
  - choose a model and then estimate the model’s parameters
  - use parameters to calculate spectrum, classify the signal, perform compression, ..

- critical point: choice of an adequate model (adequate for the given signal)
  - not necessarily physiologic, anatomic, or physical characteristics, but simply input-output relations: black-box approach
  - **check validity of assumptions *a posteriori***



ARMA model

- Autoregressive moving average (ARMA) system is driven by white noise, e(k), with zero mean and constant power σ^2

$$y(k) = \underbrace{-\sum_{i=1}^{p} a_i \cdot y(k-i)}_{\text{AR}} + \underbrace{\sum_{j=1}^{q} b_j \cdot e(k-j)}_{\text{MA}} + e(k)$$

- output, y(k) is the output of the model
- AR (auto regressive) model $\rightarrow$ b are set to zero
  MA (moving average) model $\rightarrow$ a are set to zero

different models

- most prominent are AR (autoregressive) models, and ARMA (autoregressive moving average) models
- AR models show outstanding performance when the signal's frequency response has sharp peaks
- ARMA models are suitable for modelling signals with sharp peaks and valleys in their frequency contents
- AR models are popular because with them an accurate estimation of the PSD can be obtained by solving linear equations

  (Note: Matlab's Signal Processing Toolbox implements only AR models, for ARMA models you'll need the System Identification Toolbox)

AR process

- The transfer function from output to input can be written as

$$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_p z^{-p}$$

- The transfer function A(z) is stable and causal and known as the whitening inverse filter.
- The AR process can be written in the z-domain as E(z) = A(z) • Y(z) and in time domain as

$$e(k) = \sum_{i=0}^{\infty} a_i \cdot y(k-i)$$

$$\Leftrightarrow \; y(k) = -\sum_{i=1}^{\infty} a_i \cdot y(k-i) + e(k)$$

$$e(k) \rightarrow \boxed{\frac{1}{A(z)}} \rightarrow y(k)$$

- The **AR model** is also referred to as the ‘all-pole method’. Each sample of the signal can be expressed as a linear combination of previous samples and an error signal, e(k). The error signal is presumed to be independent of the previous samples.
- An estimation $\hat{y}(k)$ of the signal to be modelled can be obtained from:

$$\hat{y}(k) = -\sum_{i=1}^{P} a_i \cdot y(k-i)$$

- P is the model order, and the a_i are the AR coefficients.
- For an optimal AR model, the prediction error $e(k)$ must be white and independent from the previous outputs $y(k\text{-}i)$.

- When the noise power is σ^2 and the sampling interval Δt the PSD can be calculated as

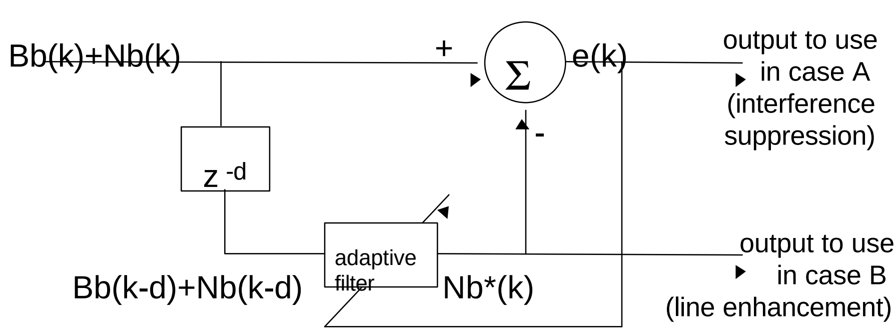
$$S_{AR}(w) = \frac{s_e^2}{\left|1+\sum_{i=1}^{P} a_i \cdot e^{-jwi\Delta t}\right|^2} = \frac{s_e^2}{\left|1+\sum_{i=1}^{P} a_i \cdot z^{-i}\right|^2_{z=e^{jw\Delta t}}} = \frac{s_e^2}{\left|\prod_{i=1}^{P} (z-z_l)\right|^2_{z=e^{jw\Delta t}}}$$

- This is called maximum entropy (ME) spectral estimate as it maximises the entropy of the spectrum.
- Since σ^2 is constant, the only values that are needed for calculating the shape of the PSD are the coefficients a_i.
- There are different methods to find those coefficients, e.g.,
  - The autocorrelation (Yule - Walker) method (can be implemented efficiently)
  - The Burg method (more appropriate for short data sequences)
  - Adaptive AR method that is more suitable if we have nonstationary signals (often the case in biomedical signals) – this can be done using the principle of the ALE filter

adaptive line enhancer (ALE)

- In practice it is not always possible to obtain a separate channel that provides reference noise.
- In such a situation we can create a reference signal ourselves from the main signal and then use the earlier described (lecture 3) adaptive noise canceller method
- ALE uses the assumption that the recorded signal contains a broadband part, Bb(k) (which can be signal or noise) and a narrowband part Nb(k) (which can be noise or signal)
  - Situation A (interference suppression): e.g., we have a complex biomedical signal corrupted by 50 Hz mains interference; the signal is Bb(k) and noise is Nb(k)
  - Situation B (line enhancement): e.g., we have a sinusoid-like signal corrupted by white noise; we have as signal Nb(k) and as noise Bb(k)

adaptive line enhancer



e(k)=Bb(k)+Nb(k)-Nb*(k)

ALE principle

- The idea is that while neighbouring samples in the Nb(k) part of the signal are correlated, those in the Bb(k) part are not.
- If we delay the combined signal by d samples Bb(k) will not be correlated with Bb(k-d) (whereas Nb(k) and Nb(k-d) most likely are correlated, if we choose d appropriately).
- If we subtract the filter's output from the combined input signal the result is only influenced by Nb(k). Minimising the error of the filter means making Nb* (the filter output) as close as possible to Nb.
- In interference removal we use the error as output (combined input minus estimation of Nb), in line enhancement we use the estimation of Nb as output.

input signal:

$$x(k) = Nb(k) + Bb(k)$$

filter input is a delayed replica of the input signal

$$x_d(k) = Nb(k-d) + Bb(k-d)$$

estimated narrowband part of the signal:

$$Nb^*(k) = \sum_{m=0}^{M} w_m(k) \cdot x_d(k-m)$$

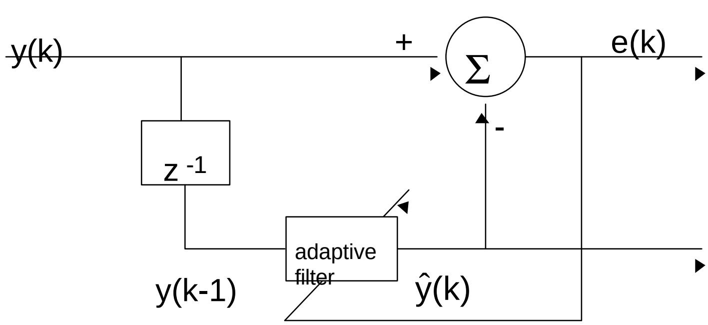
the error: $e(k) = x(k) - Nb^*(k) = Nb(k) + Bb(k) - Nb^*(k)$

- calculate Nb*(k)
- calculate e(k)
- update filter coefficients $w_m(k+1) = w_m(k) + a \cdot e(k) \cdot x(k-m-d)$
- repeat

- For example, if we have a signal that is a sinusoid buried in white noise.
- By choosing an appropriate delay, d, the ALE decorrelates the noise in the reference input from the noise in the main input.
- However, the signal (the sinusoid) in both inputs is still correlated. Eventually ALE will tune itself so that it produces produce a sharp peak at the frequency of the sinusoid.

Adaptive AR method

- develop a linear predictor that uses a reference channel derived from the main channel with a delay of one sample (this is the same principle as the ALE filter)



In the AR model estimation case, the part of the signal covered by the AR model is the ‘Nb’ part and using the AR coefficients instead of filter coefficients we get as update rule for the AR coefficients

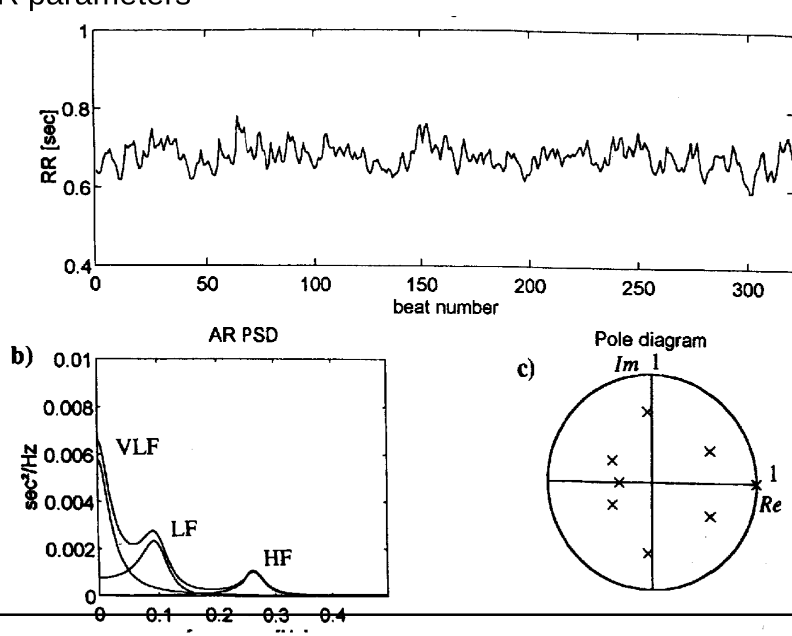
$$a_i(k+1) = a_i(k) + a \cdot e(k) \cdot y(k-i)$$

with α the learning rate.

After the adaptive filter/AR coefficients has converged, the PSD can be calculated according to:

$$S_{\mathrm{AR}}(\omega) = \frac{\sigma_e^2}{\left|1 + \sum_{i=1}^{P} a_i \cdot e^{-j \cdot \omega \cdot i \cdot \Delta t}\right|^2}$$

Example of an HRV spectrum calculated from a tachogram using AR parameters



AR model validation

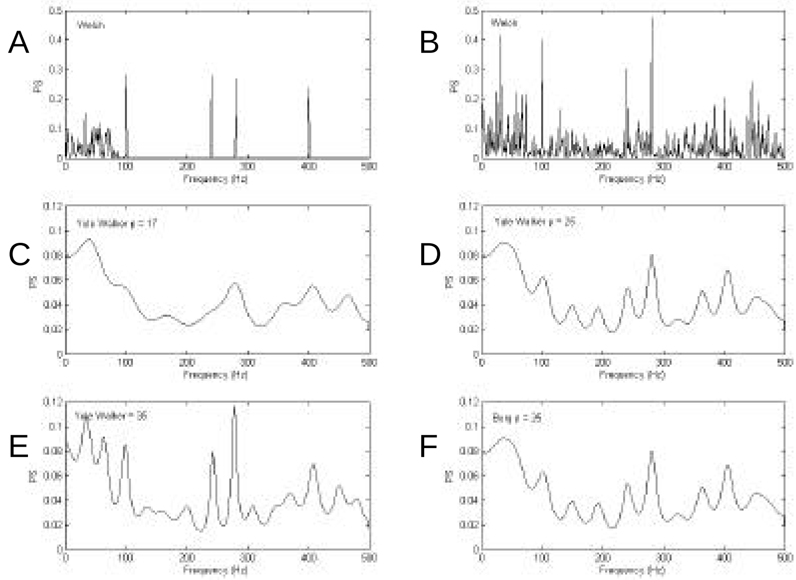
- AR modelling provides a good spectral resolution if the signal is well presented as an AR process. This may be validated *a posteriori* by checking the validity of assumptions:
  - whiteness of the residual $e(k)$
  - residual $e(k)$ and past outputs $y(k\text{-}i)$ do not correlate
  - (computing the prediction ratio: $R_{yi}^2 = 1 - \frac{\sum_{t=1}^{N} e_i^2(t)}{\sum_{t=1}^{N} y_i^2(t)}$ this is not very useful for validation but for comparison of different models)
- If the assumptions are not met there are two possibilities
  1. the model order is too low $\rightarrow$ try increasing it
  2. the signal may not be suited to be presented as an AR process

AR model order selection

- Model order selection for a real signal is a trade-off:
  - too low model order $\rightarrow$ the model cannot describe the signal
  - too high model order $\rightarrow$ the variance of the parameter estimates leads to increased variance e.g. in spectrum (spurious peaks)
- The objective criteria, such as Akaike's information criterion

$$AIC(P) = N\ln(E_P) + 2P$$

  may be applied. However, these criteria tend to underestimate the model order with real signals, which are not pure AR processes but may well be modelled as such with a sufficient model order.
- In practice:
  - get guidance from literature and objective criteria
  - try out many model orders
  - compare to non-parametric methods
  - use common sense!



comparison of PSD calculated with different AR models and the Welch method (DFT). A) spectrum using Welch method on a signal with 4 sinusoids (100, 240, 280, and 400Hz) with some low frequency noise. B) same method after white noise has been added (SNR=-8dB)
C, D, E) PSD using Yule-Walker method: p=17 cannot distinguish the two close peaks (240 and 280Hz), p=35 has sharper peaks, but also spurious peaks. F) Burg method shows results almost equal to Yule-Walker

Parametric vs. non-parametric methods in spectral estimation

- advantages over non-parametric methods:
  - more statistical consistency, even on short segments
  - spectrum is more easily interpretable
  - reliable calculation of (physiologically interpretable) spectral parameters
  - no need for windowing, no spectral leakage
  - frequency resolution "independent" of number of data
- disadvantages
  - more complex from a methodological and computational point of view
  - requires *a priori* definition of the kind of model and its order

AR modelling of seizure EEG

Analysis of ‘epileptic EEG’. This is often done by averaging segments processed with FFT. However this method does obscure dynamic changes. Left picture shows spectra calculated from fixed-length segments of rat-EEGs

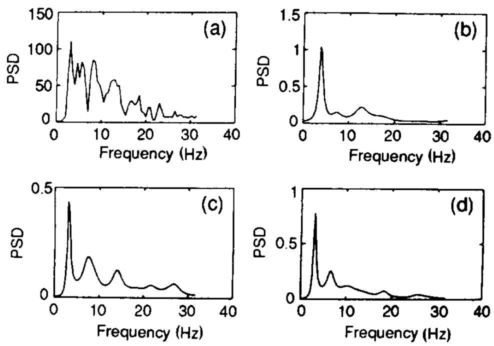


Fig. 10.4. Comparison between power spectra computed via the FFT and via parametric modeling of the EEG. Seizure EEG (left frontal lead) from a rat treated acutely with kainic acid to the right amygdala. (a) Power spectrum density of an 8-sec-long signal stretch computed via FFT. Averaging of the eight periodograms calculated on signal segments 1 sec long each. (b–d) Power spectrum calculated on three different 1-sec-long segments (included in the 8-sec-long signal stretch) through AR modeling. Model order 8, sampling rate 128 Hz. The *y*-axis is in arbitrary units. [From Gath *et al.* [24]].

AR modelling of seizure EEG (2)

For short data segments AR modelling provides better statistical consistency than FFT based methods

(provided that signal is well-presented by AR model and AR modelling is carried out adequately).

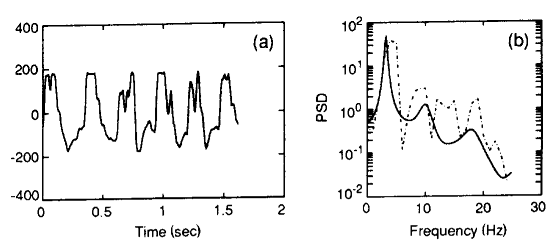


Fig. 10.5. Comparison between power spectra computed via FFT and via parametric modeling of the EEG. (a) EEG stretch 1.6 sec long derived by adaptive segmentation, recording conditions as in Fig. 1 in reference [24]. (b) Broken line, raw periodogram (FFT) of the EEG signal displayed above. No averaging of the periodogram was possible due to the short length of the signal at hand. Thus, the FFT estimate is inconsistent, and its standard deviation is as great as the quantity being measured. Solid line, power spectrum calculated on the EEG signal through AR modeling as in Fig. 1 in reference [24]. The estimate is consistent, and its standard error is approximately equivalent to that of averaging over 13 periodograms derived from 1.6 signal segments. [From Gath *et al.* [24]].

AR modeling of EEG during neurosurgery

Idea: to track changes in EEG during neurosurgery by tracking AR model poles.

Aim: to study central nervous system reaction on hypotensive drug (sodium nitro-prusside, SNP).

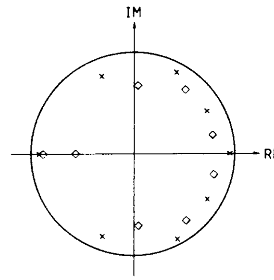


Fig. 10.30. Pole diagram of the identified model during the epoch of peak induction of SNP (aiming at evaluating the dynamics of poles position). ◇, actual epoch (No. 8); ×, preceding one. [From Cerutti *et al.* [54]].

AR modelling of EEG during neurosurgery

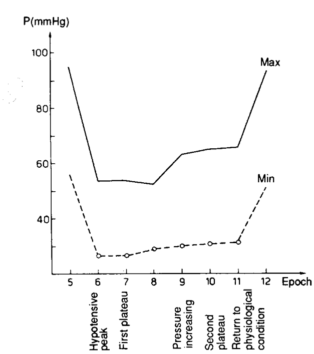


Fig. 10.31. Plotting of maximum and minimum pressure values for one patient during epochs from 5 to 12 (epoch 6 corresponds to the minimum value of pressures induced by SNP). [From Cerutti *et al.* [54]].

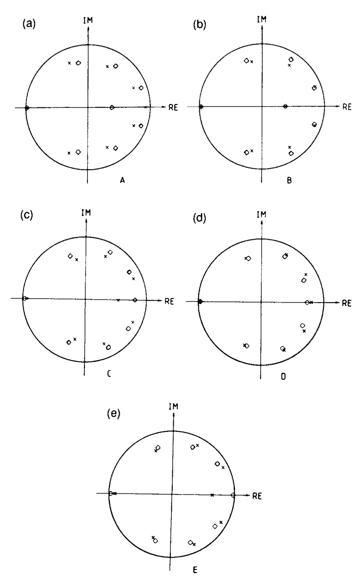


Fig. 10.32. Pole diagrams relative to the patient shown in Fig. 10.31 for epochs 5, 6, 8, 9, 11 indicated respectively by A, B, C, D, E. [From Cerutti *et al.* [54]].

ARMA modelling

- Sometimes, biomedical signals cannot be appropriately modelled using AR models. For example, for signals that are very heavily disturbed by background noise ARMA modelling may be better suited. The same holds for signals that have deep valleys in their frequency response.
- In this case we'll model

$$\hat{y}(k) = -\sum_{l=1}^{P} a_l y(k-l) + \sum_{m=0}^{Q} b_m e(k-m)$$

- solving the equations to find the coefficients requires solving non-linear equations, and the MLE approach requires minimising non-linear eqs. as well.

Event detection

- Rationale: biomedical signals carry signatures of physiological *events*. Analysis of physiological system behaviour often requires identification and investigation of specific events.
- Events are typically:
  - specific (almost deterministic or known) waveforms in the signal, often buried in noise or other activity
  - specific changes in the statistical properties of the signal, such as steps or trends in signal mean or power (variance)

EEG events

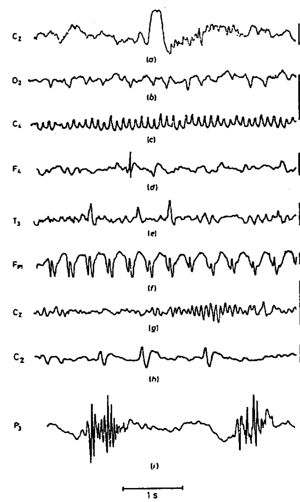


Figure 4.1: From top to bottom: (a) the K-complex; (b) the Lambda wave; (c) the mu rhythm; (d) a spike; (e) sharp waves; (f) spike-and-wave complexes; (g) a sleep spindle; (h) vertex sharp waves; and (i) polyspike discharges. The horizontal bar at the bottom indicates a duration of 1 s; the vertical bars at the right indicate $100\mu V$. Reproduced with permission from R. Cooper, J.W. Osselton, and J.C. Shaw, "EEG Technology", 3rd Edition, 1980, ©Butterworth Heinemann Publishers, a division of Reed Educational & Professional Publishing Ltd., Oxford, UK.

Detection based on descriptive properties

- Often events are characterised by specific descriptive properties that differ from other parts of the signal → use this in event detection
- Case: QRS-detection in ECG:
  - QRS complex has usually the largest slope in the signal
  - QRS complex has usually the largest amplitude in the signal
  - QRS complex has a certain frequency range different from other parts of the signal

QRS detection by slope detection

$y_2 = 1.3y_0 + 1.1y_1$

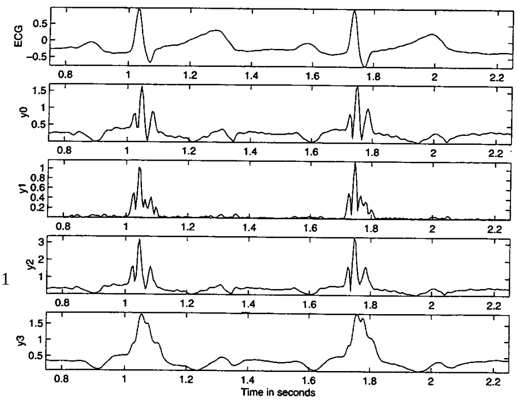


Figure 4.2: From top to bottom: two cycles of a filtered version of the ECG signal shown in Figure 3.4; output $y_0(n)$ of the first-derivative-based operator in Equation 4.1; output $y_1(n)$ of the second-derivative-based operator in Equation 4.2; the combined result $y_2(n)$ from Equation 4.3; and the result $y_3(n)$ of passing $y_2(n)$ through the 8-point MA filter in Equation 3.25.

Comment: sensitive to noise

Pan-Tompkins algorithm

QRS band

Derivative

Magnify & unidirect

Make unimodal

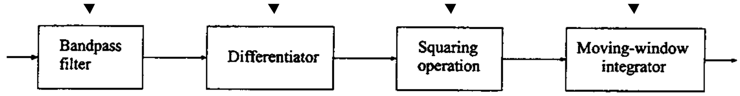


Figure 4.4: Block diagram of the Pan – Tompkins algorithm for QRS detection.

Pan-Tompkins algorithm

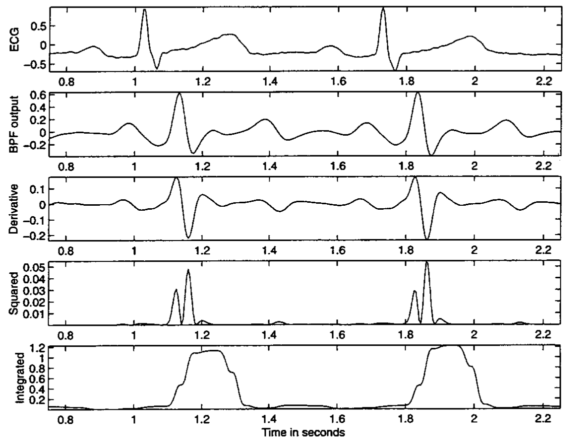


Figure 4.6: Results of the Pan – Tompkins algorithm. From top to bottom: two cycles of a filtered version of the ECG signal shown in Figure 3.4 (the same as that in Figure 4.2); output of the bandpass (lowpass and highpass) filter; output of the derivative-based operator; the result of squaring; and 100× the result of the final integrator.

Maximisation of the SNR (matched filtering)

- **rationale**:
  the goal is not to generate an output as close as possible to the true signal, but only to detect whether some signal occurs in some noisy 'environment' or not.

- For example, if we try to measure heart rate under noisy conditions we only have a need for detecting the occurrence of the R wave in ECG. What the exact shape of the ECG signal is, is not important in this particular case.

Matched filtering

Assume the sought signal $s(t)$ is some deterministic function, then the response of a filter with impulse response $h(t)$ is also deterministic.

$$\hat{s}(t) = h(t) \otimes s(t)$$

Define as the optimality criterion the maximization of the signal-to-noise ratio:

$$SNR_0(t) = \frac{\hat{s}(t)}{E\left\{n_0^2(t)\right\}}$$

Optimal filter maximises the output SNR at a given time t_0. This maximisation yields

$$\int_0^T h(x) \cdot r_{nn}(t-x)dx = a \cdot s(t_0 - t) \qquad 0 \le t \le T$$

with T the observation time and α any constant. This equation has to be solved for any given signal and noise.

Special case: white noise (r_{nn} is the delta function). The solution of the filter impulse response becomes:

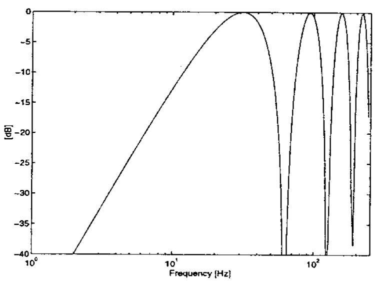
$$h(\tau) = \frac{1}{N} s(t_0 - \tau)$$

with N the noise power. The impulse response has the form of the sought signal run backwards, and shifted to time t_0: this is called a *matched filter*.

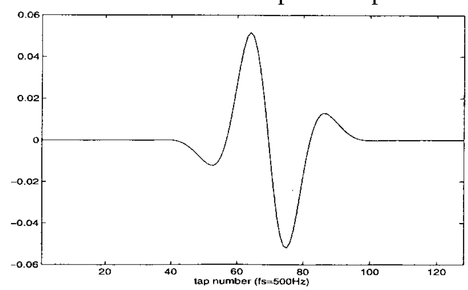
QRS detector using Matched Filter

Idea: matched filter impulse response matches with QRS waveform (flipped in time).

Bandpass filter (QRS band)



Matched filter impulse response



Ruha et al, IEEE T-BME, 1997

QRS detector by Matched filter

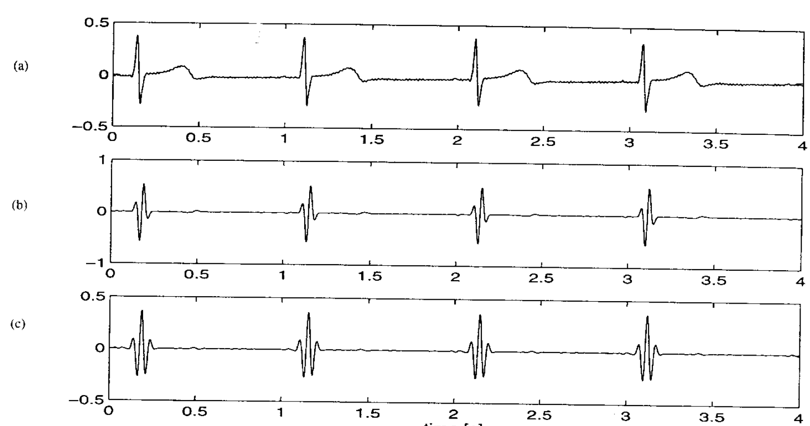


Fig. 5. (a) Almost noiseless ECG signal, which is (b) bandpass filtered, and (c) matched filtered.

Ruha et al, IEEE T-BME, 1997

QRS detector by Matched filter

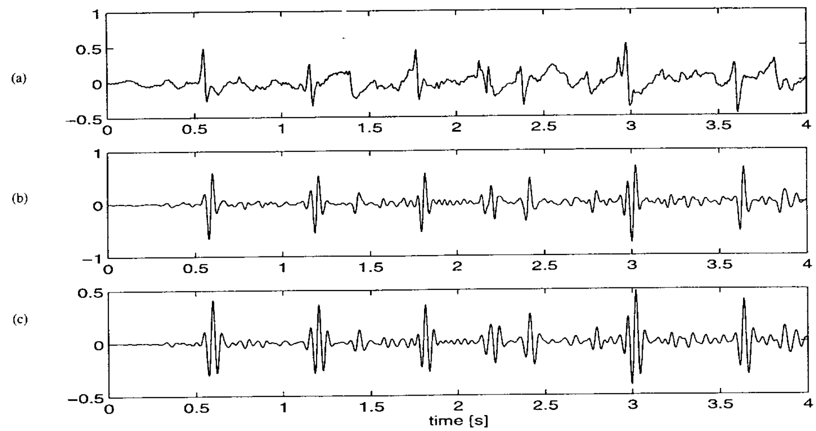


Fig. 6. (a) Noisy ECG signal, which is (b) bandpass filtered, and (c) matched filtered.

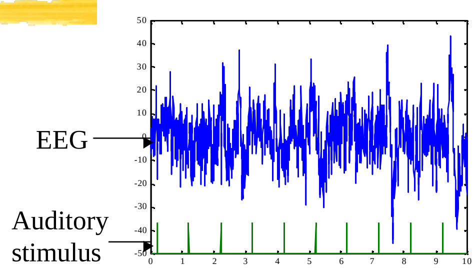
Ruha et al, IEEE T-BME, 1997

Signal averaging

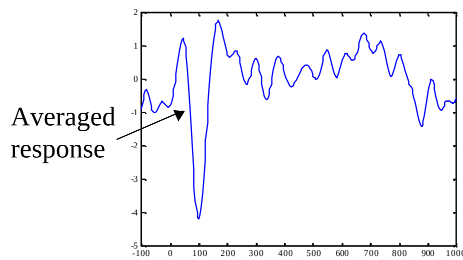
- Problem: what to do when frequency contents of signal and noise overlap?
  - No separate passbands and stopbands can be defined
  - A filter that cancels the noise will also cancel or at least distort the actual signal we want to study
- Typical problem fields: detection of brain potentials evoked by sensory stimulation (visual, acoustic, or somatosensory): evoked potentials (EPs), or detection of certain small potentials in ECG signals

Event-related and evoked potentials

Deterministic response to given stimulus buried in background activity and noise.



Usual case: brain electrical responses to sensory stimuli: very low SNR



- example: an auditory evoked potential (AEP)
- *signal*: evoked potential (amplitude 1-5 µV)
- *noise*: spontaneous EEG (amplitude 10-100 µV)
- SNR ≈ 0.1 and frequency contents of signal and noise coincide
- solution: if the evoked response stays the same with every measurement and the noise does not, one can extract the evoked response by *signal averaging*

- present stimuli at a fixed frequency to the subject
- record signal (including AEP, EEG and other noise) in a time interval locked to the stimulus: a *sweep*
- average the sweeps and the AEP can be extracted, but **only if:**
  1. the noise is a random, zero-mean signal whose stochastic properties do not change over time
  2. the response to stimuli, s(*t*), does not change over time (i.e., phase, form, latency and amplitude stay the same)
  3. signal, s(*t*), and noise, $n_i(t)$, are not correlated
- the measured signal in sweep *i*, $x_i(t)$, can be written as a sum of the AEP, $s(t)$, and the noise, $n_i(t)$:
  $x_i(t) = s(t) + n_i(t)$

If we present *N* stimuli, record after each stimulus *i* the signal x_i (consisting of the constant evoked response *s*, and a randomly varying n_i) and average the results we get as average of *x*

$$x_{average} = \frac{1}{N}(x_1 + x_2 + ... + x_N) = \frac{1}{N}((s+n_1)+(s+n_2)+...+(s+n_N)) = \frac{1}{N}(N \cdot s + n_1 + n_2 + ... + n_N),$$

or, rewrite it more compactly as :

$$x_{average} = s + \frac{1}{N}\sum_{i=1}^{N} n_i.$$

The expected value of $x_{average}$ is

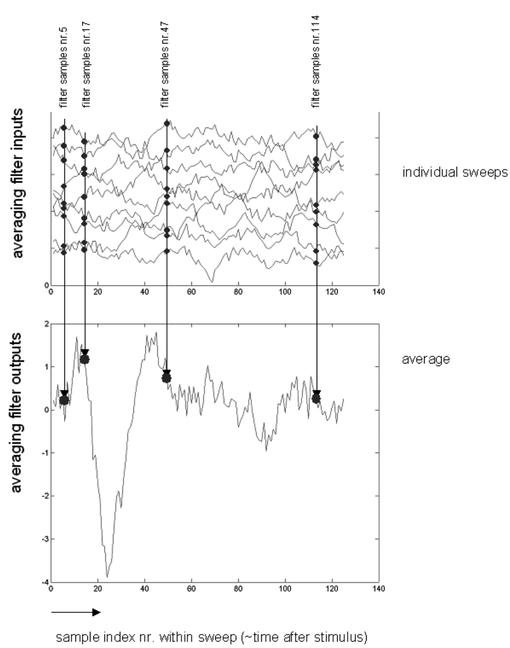
$$E[x_{average}] = E\left[s + \frac{1}{N}\sum_{i=1}^{N} n_i\right] = E[s] + E\left[\frac{1}{N}\sum_{i=1}^{N} n_i\right] = s + 0 = s\ .$$

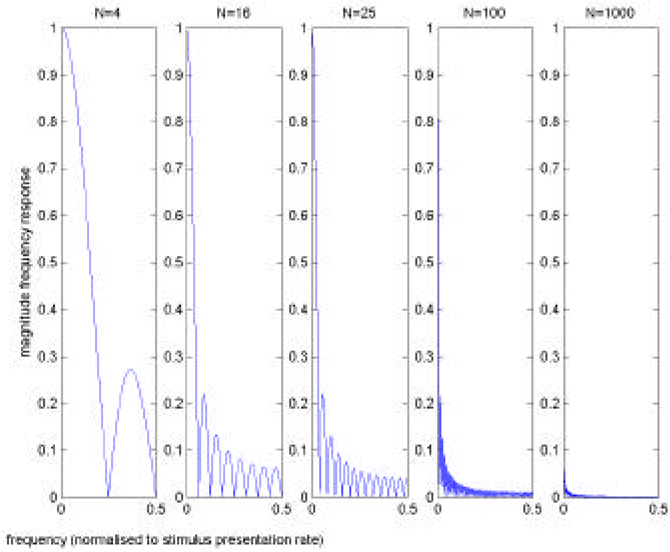
The variance of $x_{average}$ is (assuming *s* is constant, and n_i is uncorrelated with n_j so that $E[n_i \cdot n_j] = 0$ for $i \neq j$):

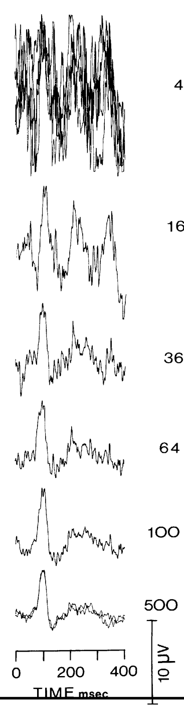
$$\mathrm{var}[x_{average}] = E\left[\left(\frac{1}{N}\sum_{i=1}^{N} n_i\right)^2\right] = \frac{1}{N}\mathrm{var}(n)\ .$$

- Eventually the x_average will converge to its expected value of s, and the variance of the x_average will decrease with increasing N.
- To put it in another way, and in amplitude terms (not in squared amplitudes that are used for the variance), this means that the signal-to-noise ratio (SNR) improves with √N.
- Thus, increasing the number of stimuli by a factor of 4 leads to an increase in SNR of 2 etc. The choice of the number of N in practice depends on the application. For some event related potential studies the use of several tens of stimuli may suffice, whereas for example for brainstem auditory evoked potentials several thousands of stimuli may be needed to 'lift' the small evoked potential (in the order of less than a few microvolts) from the spontaneous EEG background noise (with an amplitude that is in the order of ten times bigger).

observe: use of this technique is equal to use of a moving average filter over corresponding samples in the sweeps (rather than along the time axis).
The more sweeps the higher the filter order, and the narrower its low-pass band gets.







AEP average with
increasing number
of sweeps, getting more
and more low-pass filtered.
Eventually, only the components
of the signal that are constant
in all the sweeps will
be passed through.

Notes about EP averaging

- note: noise leakage through side-lobes (and also via dc-component, however this should not play a role since noise is supposed to have a zero mean)

- **recall: this works only if assumptions 1-3 are valid!**

- does an evoked potential (and noise) really stay the same during the whole measurement time?
- are EEG and EP really uncorrelated?
- is the noise really zero mean?

Step and trend detection

- Aim: to detect statistically significant trends or steps in stochastic signals (very common problem e.g. in patient monitoring). Problem is accurate and fast detection in the presence of other variability.

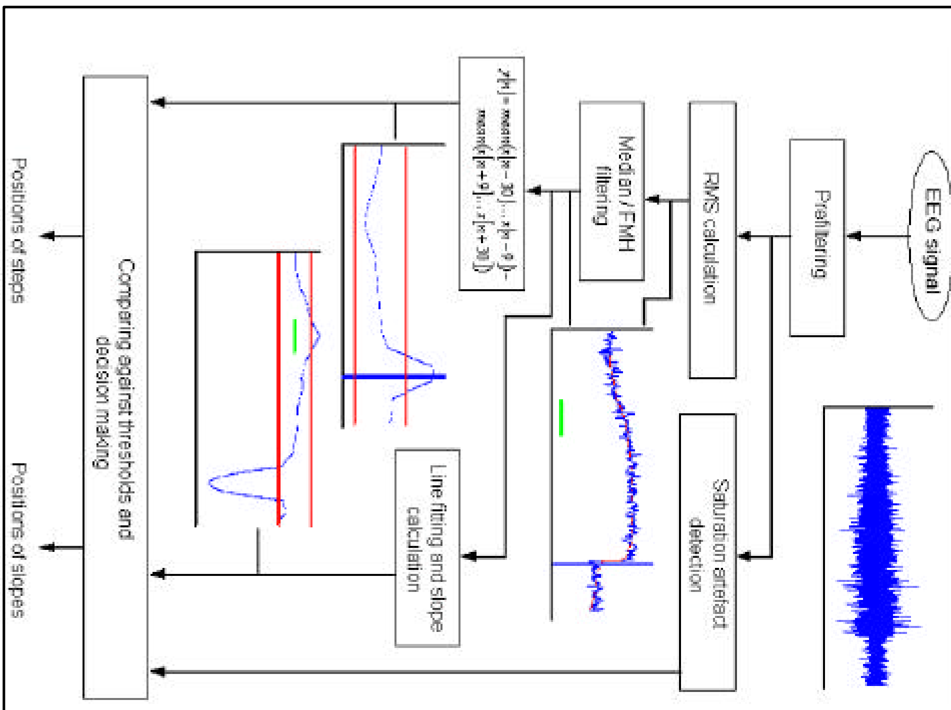
- Typical solution:
  1. track a statistical feature in the signal
  2. define a detection signal from a statistics between two successive windows
  3. derive thresholds experimentally or adaptively from the theoretical properties of the statistics

Step and trend detection

Example: simple step and trend detection in EEG RMS power during critical care:

- treats data in windows (L=2 sec); RMS (root mean squared) power computed for each window
- saturation artefacts are detected on the basis of high amplitudes in the analysis windows
- median filter smoothes the signal without affecting the steps
- Step detection: successive windows (each 30 samples) are compared; if their mean differs too much a step is detected
- Slope detection: a line is fitted (least squares fitting) to median filtered data; if the slope exceeds a threshold a trend is detected.

Motivation: (sudden) loss of EEG amplitude during surgery is often related to brain hypoxia



Statistical Process Control

- trends and sudden (significant) changes in mean are examples of a process going ’out of control’ when we use the context of for example a manufacturing process.
- a process being ’in control’ means that we can expect future values to be within certain limits; they are predictable (thanks to a stable mean and standard deviation)
- the theory behind ’detecting an out of control’ situation has been developed for monitoring production quality (weights, sizes etc. of products made in a factory) but can just as well be applied to patient monitoring

’out-of-control’ detection

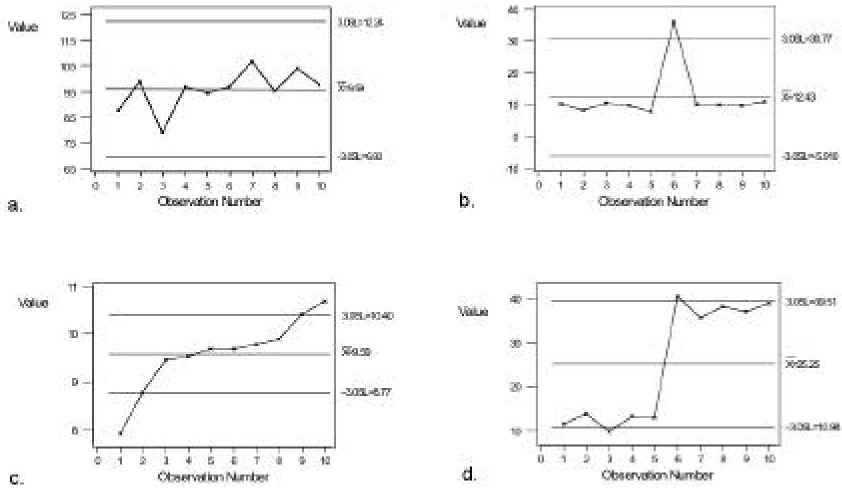
- A process can go out of control in many different ways, for example there may be a single exceptional value, a sudden shift in the mean, a gradual drift, or a sudden increase in variation.
- Different charts can be made to quickly detect these situations, these are called control charts. Examples are Shewhart and Cusum charts.
- 2 phases: in phase 1 we calculate a reliable mean and standard deviation and we derive control limits from them. In phase 2 we monitor the performance of the process.

Average Run Length (ARL)

- The performance of a control chart is assessed by calculating the probability of detecting an out-of-control point, but the result is usually expressed as the reciprocal of this probability, known as the average run length, or ARL.
- ARL is the average number of points plotted for each out-of-control point. If the process is in control a large ARL is desirable to minimize false alarms, but when the process goes out of control a short ARL is required for rapid detection.

Shewhart charts

- Estimate the standard deviation, σ, and set limits at 3σ on either side of the mean (empirically found to be useful).
- estimate mean and σ from a moving range
- for normally distributed data the chance that a value will lie outside the 3 σ limits is 0.0027 (probability of a false alarm if the process is in control)
- The average run length (ARL) for an in-control process is therefore 1/0.0027, or 370. This is the average number of points plotted for each out-of control point.

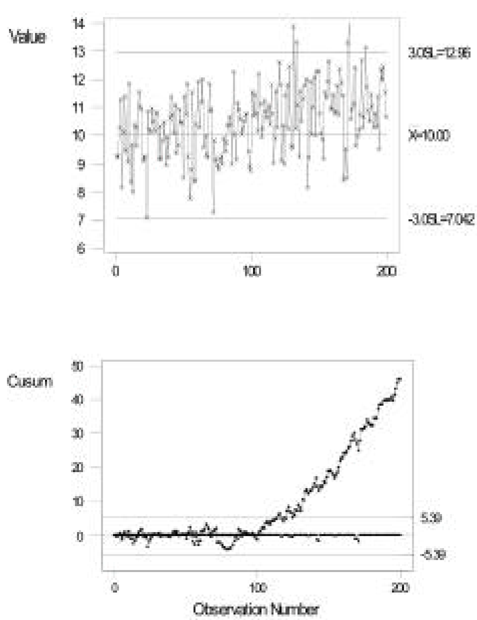


10 values plotted on Shewhart charts. In each case the mean and upper and lower control limits were calculated from all ten values. All points in (a) are within the control limits, although many more values would be required to be certain that the process was in control. A single exceptional value in (b) indicates that the process is out of control. In (c) there is a clear trend, while in (d) there is a sudden change in the mean; in each case some points are outside the control limits.

- The Shewhart chart is good for detecting a single outlier more than 3 σ from the mean (ARL=1), but it is not good for detecting small shifts in the mean (for a deviation of 2 σ, ARL=6 and for 1 σ ARL=44)
- Detection of shifts and trends can be improved by a number of additional criteria, called runs rules, such as ‘six points in a row all increasing or all decreasing’ or ‘two out of three points more than 2 σ from the mean’.

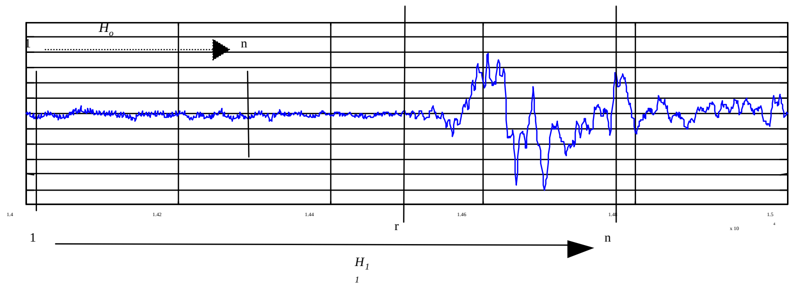
Cusum charts

- The Cusum chart is the appropriate Cuscore chart to detect a step change in the mean, even if it small.
- Here the plotted points are the cumulative sums of the deviations from the mean of the process established in Phase 1.
- Provided the process mean remains constant, the chart will be a line of constant slope, and any change in the mean produces a change in the slope.
- Tuning parameters can be set to define detection limits for ’out of control’. E.g., with appropriate settings to detect a 1σ change in mean the ARL=10 (ARL=465 for an in-control process).



Shewhart and Cusum charts consisting of 200 values. The first 100 values constitute phase 1, and are used to estimate the mean and standard deviation, and to calculate the control limits, which are applied to the next 100 values (phase 2). The mean of the second 100 values is 1σ greater that the mean of the first 100. In the Shewhart chart only a few points are above the upper control limit because of the relative lack of sensitivity of the Shewhart chart to small changes in the mean. The Cusum chart places nearly all points in Phase 2 above the upper limit, illustrating that the Cusum is the optimal chart to detect small changes in the mean.

Burst-suppression detection



Burst-suppression in EEG:
- high-amplitude burst in EEGfollowed by suppressed EEG
- occurs in pathological brain damage and deep anesthesia
- detection of major interest in monitoring of (depth of) anesthesia

Burst-suppression detection

EEG signal : s_n

L_n : log likelihood ratio

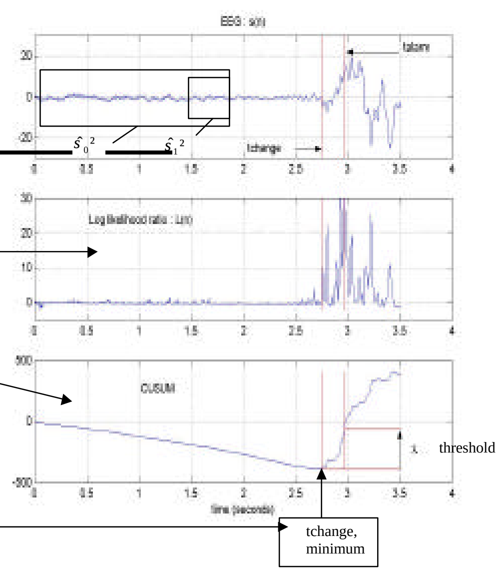
$$L_n = \ln \frac{p_1(s_n)}{p_0(s_n)}$$

$$= \ln \frac{\hat{s}_0}{\hat{s}_1} + \left(\frac{1}{\hat{s}_0^{\,2}} - \frac{1}{\hat{s}_1^{\,2}} \right) \frac{(s_n - m_0)^2}{2}$$

CUSUM

$$S_n = \sum_{k=1}^{n} L_k$$

Decision of change at talarm :
CUSUM - minimum > λ
Change time at tchange *i.e.* the minimum.



Burst-suppression detection