NON-LINEAR SIGNAL PROCESSING TECHNIQUES AND PREDICTING PRE-TERM DELIVERY

- Selected non-linear signal processing techniques
- Peak frequency of the signal power spectrum
- Peak amplitude of the normalized power spectrum
- Median frequency of the signal power spectrum
- Evaluation of signal processing techniques
- Evaluation of median frequency of power spectrum
- Sample entropy
- Evaluation of sample entropy
- Evaluation of signal processing techniques
- Current performances
- Discussion
- (Autocorrelation zero-crossing)
- (Maximal Lyapunov exponent and correlation dimension)

Selected non-linear signal processing techniques

- Peak frequency of the signal power spectrum
- Peak amplitude of the normalized power spectrum
- Median frequency of the signal power spectrum
- Sample entropy (is a measure of regularity of finite length time series and estimates the extent to which the data did not arise from a random process)
- (Autocorrelation zero-crossing (estimates periodicity of time series))
- (Maximal Lyapunov exponent (estimates the amount of chaos in a system))
- (Correlation dimension (estimates the complexity of time series))

Peak frequency of the signal power spectrum

- The power spectrum reveals periodic components of a signal and it should always be employed in time series analysis whether the primary analysis is statistical or dynamical
- Peak frequency is a suitable estimate of the signal power spectrum
- The power spectrum, $P[i]$, is calculated using the fast discrete Fourier transform, then the peak frequency, f_{max}, of the power spectrum, $P[i]$, is calculated as follows:

$$f_{\max} = \frac{F_s}{N} \arg\left(\max_{i=i_{\text{low}}}^{i=i_{\text{high}}} P[i]\right)$$

where Fs and N denote the sampling frequency and the number of samples

Peak amplitude of the normalized power spectrum

- Peak amplitude of the normalized power spectrum is a suitable estimate of the signal power spectrum
- The power spectrum is calculated using the fast discrete Fourier transform, then the peak amplitude, p_{max}, of the *normalized* power spectrum, $P[i]$, is calculated as follows:

$$p_{\max} = \max_{i=i_{\text{low}}}^{i=i_{\text{high}}} P[i]$$

Median frequency of the signal power spectrum

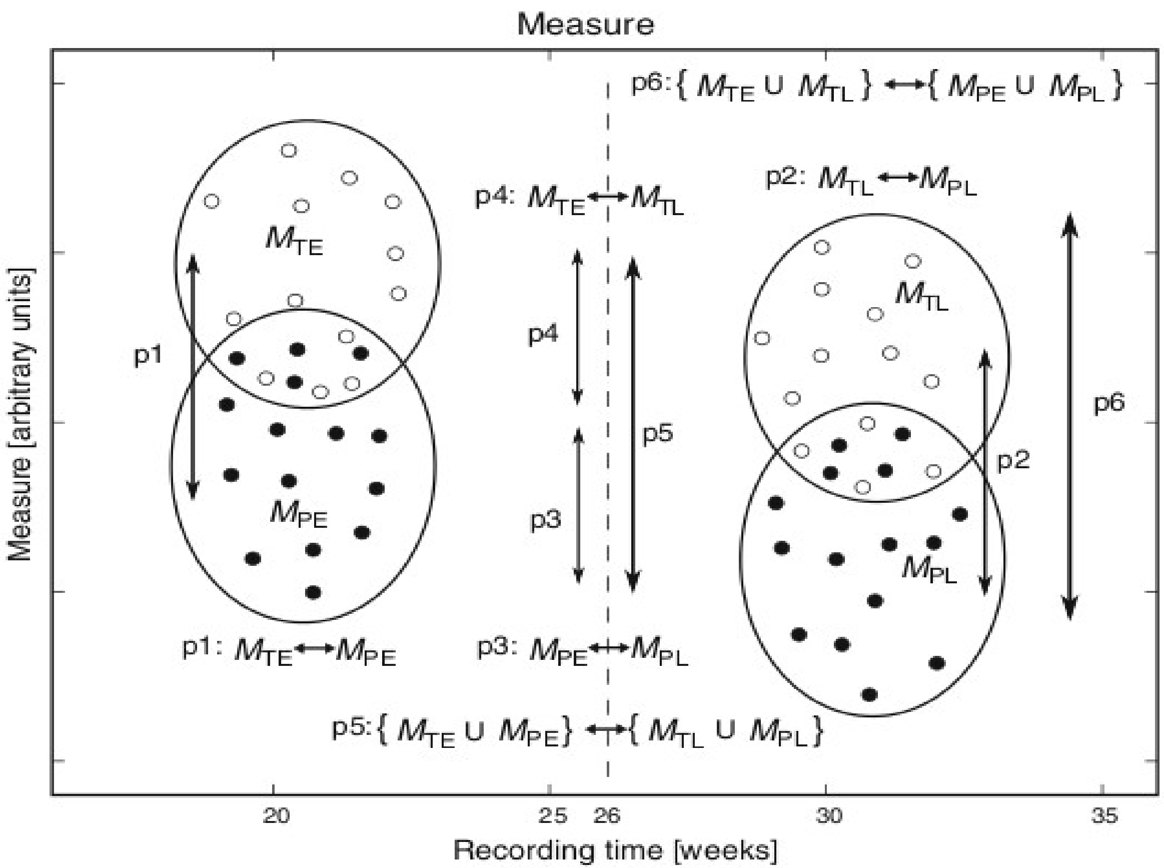
- Median frequency is a suitable estimate of the characteristic of the signal power spectrum
- The power spectrum, $P[i]$, is calculated using the fast discrete Fourier transform
- The median frequency, $f$med, is defined as the frequency where the sums of the parts above and below in the frequency power spectrum, $P[i]$, are approximately the same:

$$f_{\text{med}} = i_m \frac{F_S}{N}, \quad \sum_{i=i_{\text{low}}}^{i=i_m} P[i] \approx \sum_{i=i_{m+1}}^{i=i_{\text{high}}} P[i]$$

where Fs and N denote the sampling frequency and the number of samples

Evaluation of signal processing techniques

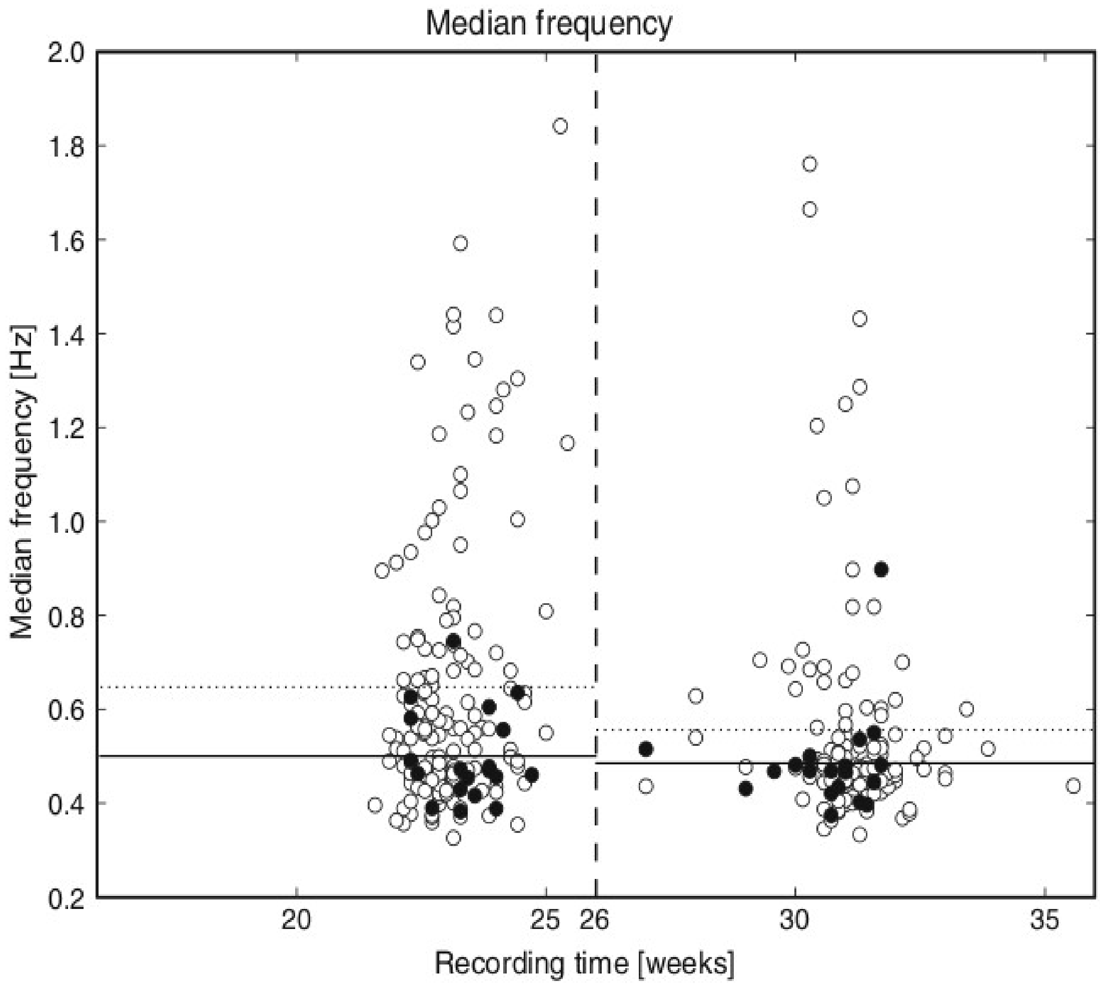
- *M* - measurements
  T - *term*
  P - *pre-term*
  E - measured *early*
  L - measured *late*
- p_1, ..., p_6 probabilities according to the Student's *t*-test when applied between the sets of measurements
- The Student's *t*-test produces the significance (probability), *p*, that two normally distributed sets belong to *the same population*



University of Ljubljana
Faculty of Computer and Information Science

Evaluation of median frequency of power spectrum

- Circles - measures obtained for *term* delivery records
- Filled circles - measures obtained for *pre-term* delivery records
- The dotted horizontal lines are the average median values for *term* delivery records (0.64 and 0.56 Hz)
- The full horizontal lines are the average median values for *pre-term* delivery records (0.5 and 0.49 Hz)



Evaluation of median frequency of power spectrum

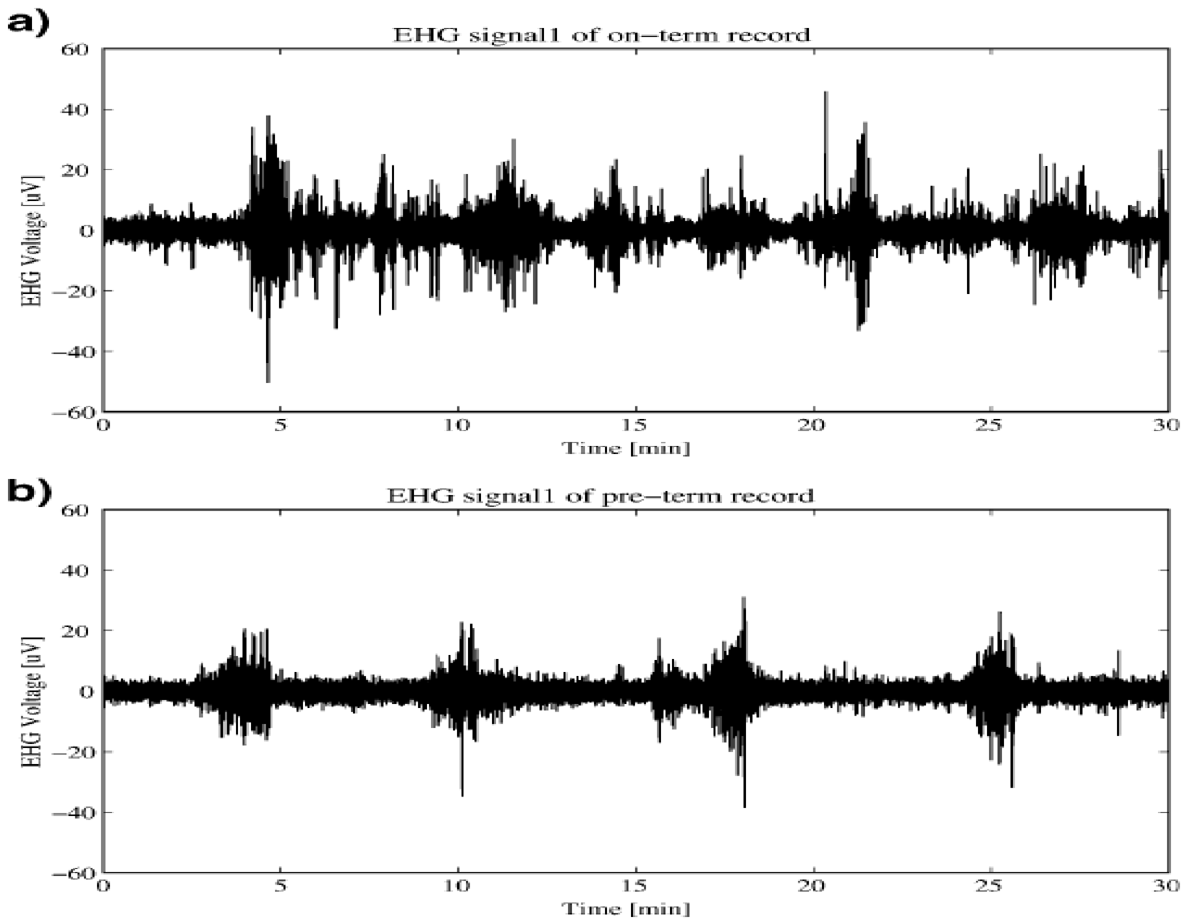
- Evaluation of median frequency of power spectrum to separate groups of records according to time of delivery (*term*, *pre-term*) and time of recording when the 0.3–3 Hz band-pass preprocessing filter was used

- *Sig:* Signal number
- p_1, ..., p_6: probabilities according to Student's *t*-tests
- Those probabilities <= 0.05 are bold
- The most important are p_1 and p_6

Technique	Preprocessing filter 0.3–3 Hz						
	Sig	p_1	p_2	p_3	p_4	p_5	p_6
Median	1	0.371	0.059	**0.012**	**0.002**	$\leq$**0.001**	0.055
Frequency	2	0.696	0.568	0.480	0.217	0.163	0.496
f_{med}	3	**0.030**	0.212	0.661	**0.007**	**0.005**	**0.012**

Evaluation of median frequency of power spectrum

- Exercises 2: Estimating time course of peak frequency and median frequency in the selected frequency bands along the spectrograms of uterine EMG records

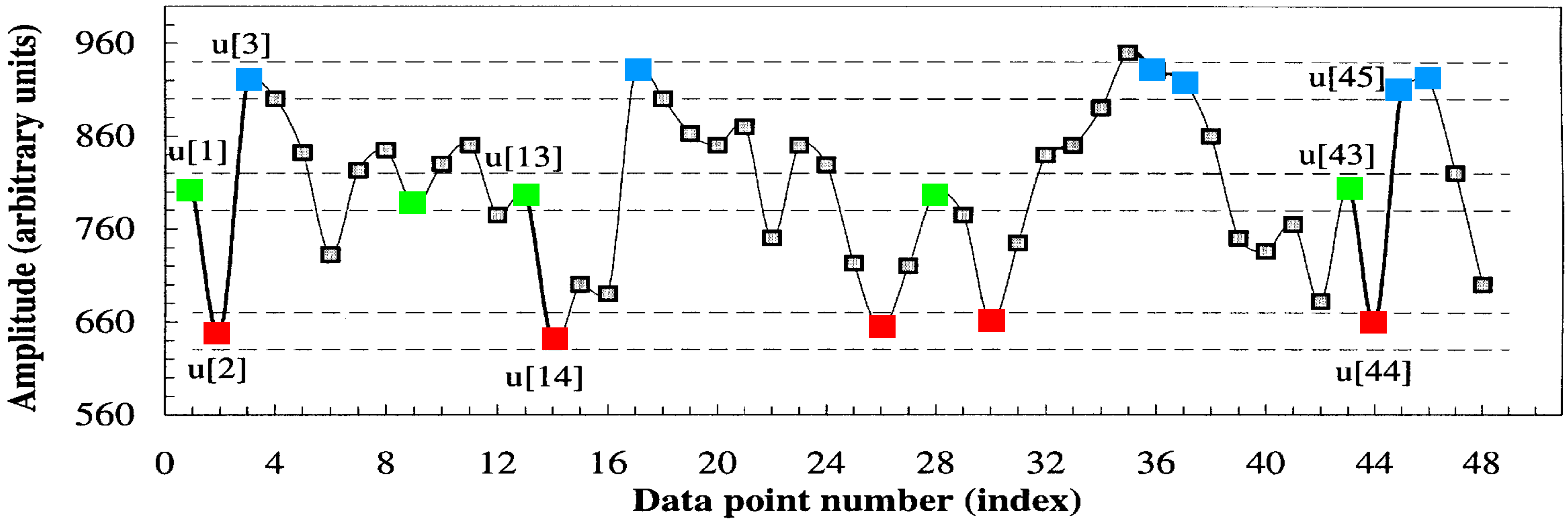


Sample entropy

- The sample entropy is a measure of regularity of finite length time series and estimates the extent to which the data did not arise from a random process
- Less predictable time series exhibit a higher sample entropy!
- Given a time series $u[n]$ of length N, and patterns $aj[0, \ldots, m - 1]$ of length m, $m < N$, where the patterns aj are taken from the time series $u[n]$, $aj[i] = u[i + j]$, $i = 0, \ldots, m-1$, $j = 0, \ldots, N - m$; the part of the time series $u[n]$ at time $n = ns$, $u[ns, \ldots, ns + m-1]$ is considered as a match for a given pattern aj if $abs\,(u[ns + i] - aj[i]) <= r$ for each $0 <= i < m$. The number of pattern matches (within a margin of r), cm, is constructed for each m.
- The sample entropy, *sampEn*, is then defined as:

$$SampEn_{m,r}(u) = \begin{cases} -\log\left(c_m / c_{(m-1)}\right), & c_m \neq 0 \wedge c_{m-1} \neq 0 \\ -\log\left((N-m)/(N-m-1)\right), & c_m = 0 \vee c_{m-1} = 0 \end{cases}$$

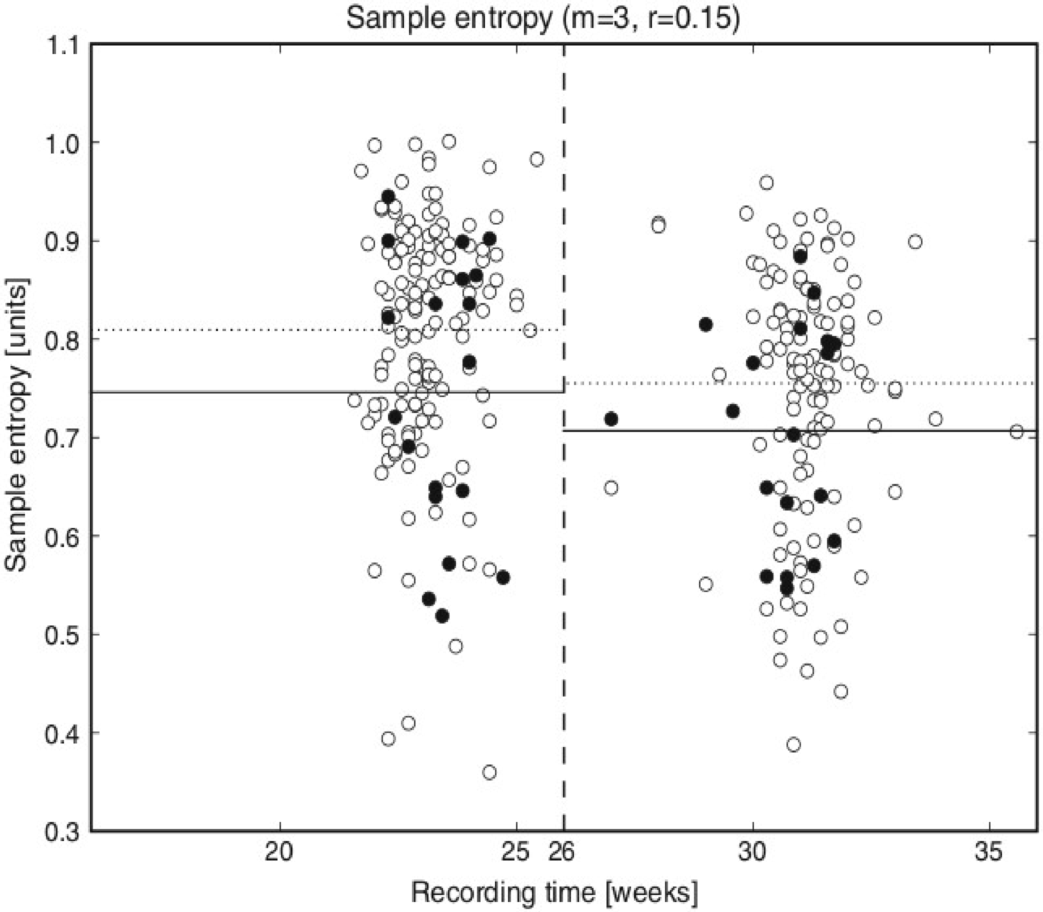
- Suitable parameters: $m = 2, 3, 4$; $2\,r = 10 - 20$ % of sample deviation, i.e., from 0.1 to 0.2.



- A simulated time series *u*[1], ..., *u*[*48*]; *m* = 3; 2 *r* = 40 (10 – 20% of sample deviation)
- Compose the first 2- and 3-component template matching sequences (*u*[1], *u*[2]) and (*u*[1], *u*[2], *u*[3])
- The number of sequences matching the 2-component template = 2
- The number of sequences matching the 3-component template = 1
- Repeat for the next 2- and 3-component template sequences (*u*[2], *u*[3]) and (*u*[2], *u*[3], *u*[4]) and add the number of matches to the previous values
- Repeat for all other possible template sequences (*u*[3], *u*[4], *u*[5]), ..., (*u*[*48*-2], *u*[*48*-1], *u*[*48*])
- The *SampEn* is the natural logarithm of the ratio between the total number of 3- and 2-component template matches

Evaluation of sample entropy

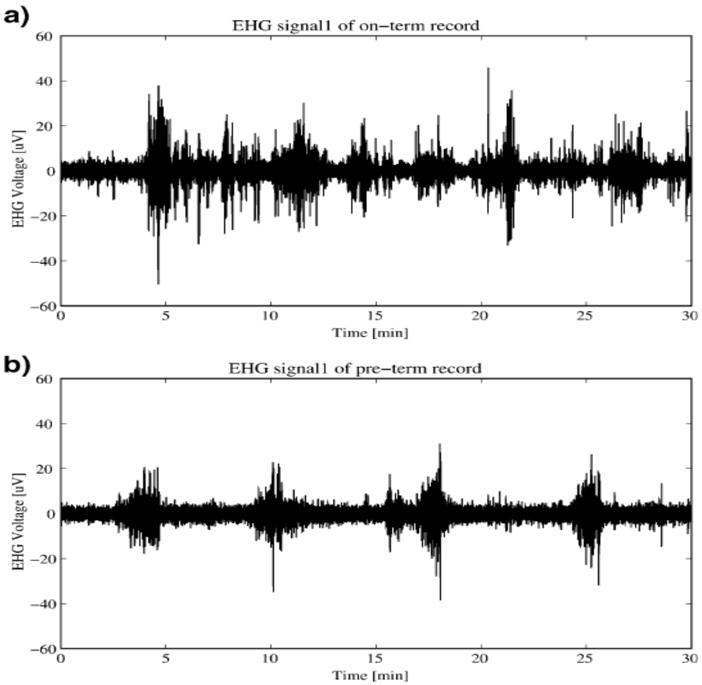
- Circles - measures obtained for *term* delivery records
- Filled circles - measures obtained for *pre-term* delivery records
- The dotted horizontal lines are the average sample entropy values for *term* records (0.81 and 0.76)
- The full horizontal lines are the average sample entropy values for *pre-term* records (0.75 and 0.71)



Evaluation of sample entropy

- Evaluation of sample entropy to separate groups of records according to time of delivery (*term*, *pre-term*) and time of recording when the 0.3–3 Hz band-pass preprocessing filter was used

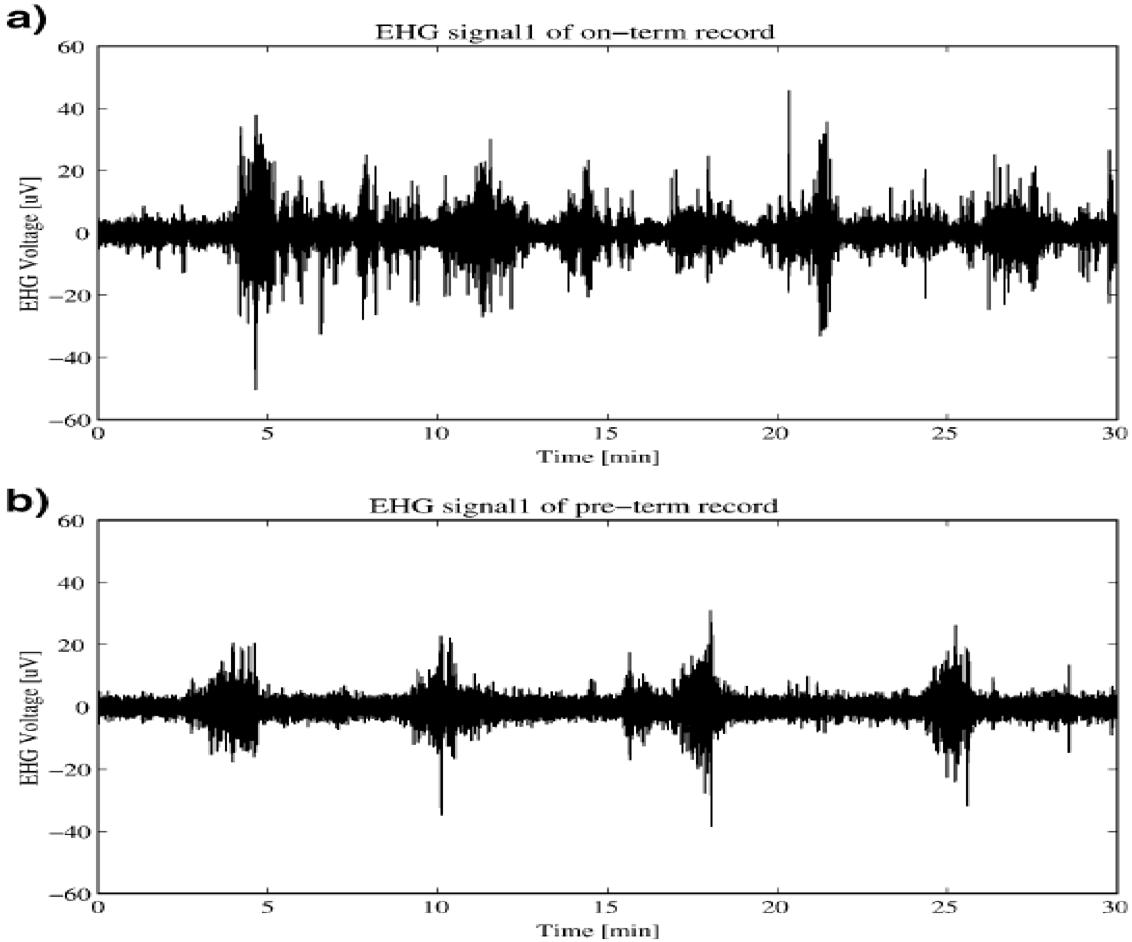
- *Sig:* Signal number
- $p_1, \dots, p_6$: probabilities according to Student's *t*-tests
- Those probabilities <= 0.05 are bold
- The most important are p_1 and p_6



Technique	Preprocessing filter 0.3–3 Hz						
	Sig	p_1	p_2	p_3	p_4	p_5	p_6
Sample entropy	1	0.326	0.172	0.272	**0.001**	**0.001**	0.084
sampEn	2	0.882	0.184	**0.017**	**≤0.001**	**≤0.001**	0.323
$m = 3, r = 1.5$	3	**0.035**	0.165	0.334	**≤0.001**	**≤0.001**	**0.011**

Evaluation of sample entropy

- Exercises 2: Separating uterine EMG records using sample entropy



Evaluation of signal processing techniques

- Evaluation of the techniques to separate groups of records according to time of delivery (*term*, *pre-term*) and time of recording when the 0.3–3 Hz band-pass preprocessing filter was used

- *Sig:* Signal number
- *p*1, ..., *p*6: probabilities according to Student's *t*-tests
- Those probabilities <= 0.05 are bold

Technique	Preprocessing filter 0.3–3 Hz						
	Sig	p_1	p_2	p_3	p_4	p_5	p_6
Root mean	1	0.586	0.349	0.247	0.838	0.529	0.769
Square	2	0.361	0.141	**0.016**	0.210	**0.044**	0.615
RMS	3	0.636	0.612	0.445	0.069	**0.045**	0.450
Peak	1	0.630	0.100	0.051	**0.020**	**0.005**	0.146
Frequency	2	0.252	0.201	0.371	0.093	0.256	0.705
f_{max}	3	0.138	0.176	0.416	**0.012**	**0.007**	**0.044**
Median	1	0.371	0.059	**0.012**	**0.002**	$\leq$**0.001**	0.055
Frequency	2	0.696	0.568	0.480	0.217	0.163	0.496
f_{med}	3	**0.030**	0.212	0.661	**0.007**	**0.005**	**0.012**
Autocorrelation	1	0.085	0.897	0.526	**0.033**	0.053	0.146
Zero crossing	2	0.089	0.340	0.223	0.658	0.499	0.059
$\tau_{R_{xx}}$	3	0.327	0.614	0.650	**0.045**	0.069	0.624
Maximal	1	0.543	0.518	0.339	0.991	0.726	1.000
Lyapunov exponent	2	0.533	0.175	0.056	0.421	0.156	0.591
λ_{max}	3	0.670	0.743	0.540	0.068	0.051	0.554
Correlation	1	0.150	0.961	0.131	0.413	0.209	0.334
Dimension	2	0.676	0.377	0.069	$\leq$**0.001**	$\leq$**0.001**	0.568
D_{corr}	3	0.790	0.976	0.446	0.113	0.079	0.882
Sample entropy	1	0.326	0.172	0.272	**0.001**	**0.001**	0.084
sampEn	2	0.882	0.184	**0.017**	$\leq$**0.001**	$\leq$**0.001**	0.323
$m = 3,\ r = 1.5$	3	**0.035**	0.165	0.334	$\leq$**0.001**	$\leq$**0.001**	**0.011**

Current performances

- Current performances in separating term and pre-term delivery EHG records of the TPEHG DB (using no additional clinical information)

- RMS, peak frequency, median frequency, sample entropy

  CA = 89% (CT) (Fergus at al, 2013)

- RMS, peak frequency, median frequency, sample entropy

  CA = 90% (NN) (Hussain et al, 2015)

- Sample entropy

  CA = 94.9% (SVM) (Ahmed et al, 2017)

- Wavelets

  CA = 96.25% (SVM) (Acharya et al. 2017)

- Normalized peak amplitude, median frequency, sample entropy

  CA = 96.33% (QDA)

  CA = 100% for early records (23rd week of pregnancy) (LBCSI, 2018)

  CA = (TP + TN) / (TP + FN +TN + FP)

- Realistic performances, CA about 65%

Discussion

- The median frequency shows a slight drop as the time of gestation progresses for *term* records, i.e., a slight decrease of the power spectra distribution (i.e., shift to lower frequencies)
- The sample entropy values are lower for both *early* and *later pre-term* delivery records and indicate that the signals of *pre-term* delivery records exhibit higher predictability than the signals of *term* delivery records (Less predictable time series exhibit a higher sample entropy)
- Peak frequency, peak amplitude, and median frequency of power spectrum, and sample entropy are promising techniques

(Evaluation of signal processing techniques)

- Student's *t*-test (the conventional statistic for measuring the significance (probability), *p*, of a difference of means):

  1) Estimate the standard deviation of the difference of the means:

$$s_D = \sqrt{\frac{\Sigma_{one}(x_i - \overline{x_{one}})^2 + \Sigma_{two}(x_i - \overline{x_{two}})^2}{N_1 + N_2 - 2}\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}$$

  2) Compute *t* by

$$t = \frac{\overline{x_{one}} - \overline{x_{two}}}{s_D}$$

  3) Evaluate the significance *p* of this value of *t* for Student's distribution $A(t \mid v)$ with $v = N_1 + N_2 - 2$ degrees of freedom, by: $p = 1 - A(t \mid v)$

- (Student's distribution estimates the probability that two normally distributed sets belong *to the same population*)
- A small numerical value of the significance ($p = 0.05$ or 0.01) means that the observed difference is "very significant"

18

(Autocorrelation zero-crossing)

- The autocorrelation provides a tool for discriminating between periodic and stochastic behavior of time series
- The autocorrelation zero-crossing is defined as the first zero-crossing starting at the peak in the autocorrelation, $R_{xx}(\tau)$, of the signal *x[i]*:

$$R_{xx}(\tau_{R_{xx}}) = 0; \quad R_{xx}(\tau) = \sum_{i=0}^{N-1} x[i]\, x[\tau + i]$$

(*For further reading see: Akay, 2000, Vol. I and II*)

(Maximal Lyapunov exponent and correlation dimension)

- The maximal Lyapunov exponent and the correlation dimension are both properties of non-linear systems
- Their calculation is based on a phase space, a construct which demonstrates the changes of the dynamical variables of the system
- The maximal Lyapunov exponent has ability to estimate the amount of chaos in the system
- The correlation dimension has ability to estimate the complexity of time series
- The phase space is a construct which demonstrates or visualizes the changes of the dynamical variables of the system
- Given a time series *x(t)* of length *N*, a Q-dimensional phase space is constructed from vectors ***y***(*t*):

$$\mathbf{y}(t) = \{y_d; d = 0, 1, \ldots, Q-1\},$$
$$y_d = (x(t+d), x(t+d+D_{\text{smp}}), \ldots,$$
$$x(t+d+(N/Q)D_{\text{smp}}))$$

where *Dsmp* is the sample delay and *Q* is the embedding dimension

(*For further reading see: Akay, 2000, Vol. I and II*)

(Maximal Lyapunov exponent and correlation dimension)

- The maximal Lyapunov exponent estimates the amount of chaos in a system and represents the maximal ‘‘velocity‘‘ with which different, almost identical states of the system, diverge
- The (maximum) Lyapunov exponent, λ, is a measure of how fast a trajectory converges from a given point into some other trajectory

$$\lambda = \lim_{t \to \infty} \lim_{\|\Delta y_0\| \to 0} \frac{1}{t} \log \frac{\|\Delta y_t\|}{\|\Delta y_0\|}$$

where $\|\Delta y_0\|$ represents the Euclidean distance between two states of the system at some arbitrary time t_0 and $\|\Delta y_t\|$ represents the Euclidean distance between the two states of the system at some later time t

- (*For further reading see: Akay, 2000, Vol. I and II*)

(Maximal Lyapunov exponent and correlation dimension)

- It is proportional to the probability of the distance between two points on a trajectory being less than some *r*:

$$D_{\mathrm{corr}} = \lim_{r \to 0} \frac{\log(C(r))}{\log(r)},$$

where

$$C(r) = \lim_{M \to \infty} \frac{1}{M^2} \sum_{i=1}^{M} \sum_{j=i+1}^{M} \Theta(r - |\mathbf{y}(i) - \mathbf{y}(j)|),$$

and

$$\Theta(r - |\mathbf{y}(i) - \mathbf{y}(j)|) = \begin{cases} 1 & : \ (r - |\mathbf{y}(i) - \mathbf{y}(j)|) \geq 0 \\ 0 & : \ (r - |\mathbf{y}(i) - \mathbf{y}(j)|) < 0 \end{cases}$$

(*For further reading see: Akay, 2000, Vol. I and II*)