

# I. Seminární práce

## Počítačové zpracování signálu (KI/PZS)

Karel Murgač

### Obsah

<b>1</b>	<b>Zadání</b>	<b>2</b>
<b>2</b>	<b>1. Výpočet tepové frekvence z EKG</b>	<b>2</b>
2.1	Popis problému . . . . .	2
2.2	Postup řešení . . . . .	2
2.3	Výsledky – DriveDB . . . . .	3
2.4	Validace – MIT-BIH NSRDB . . . . .	5
2.5	Slovní zhodnocení a diskuze . . . . .	5
2.6	Závěr . . . . .	6
<b>3</b>	<b>2. Výpočet korelace měřených signálů</b>	<b>7</b>
3.1	Zadání a cíl práce . . . . .	7
3.2	Teoretický základ . . . . .	7
3.3	Metodika zpracování signálů . . . . .	7
3.4	Výsledky: Korelace u jednotlivých pacientů . . . . .	8
3.5	Výsledky: Shoda ABP signálů mezi pacienty . . . . .	9
3.6	Závěr a slovní zhodnocení . . . . .	10

# 1 Zadání

Cílem seminární práce bylo zpracování fyziologických signálů z databází PhysioNet dle zadání předmětu KI/PZS.

Práce obsahuje dvě části:

1. Výpočet tepové frekvence z EKG signálu (DriveDB)
2. Výpočet korelace mezi signály (CharisDB)

V této práci je implementována část 1 – detekce R vrcholů a výpočet tepové frekvence v databázi DriveDB včetně validace na databázi MIT-BIH NSRDB.

## 2 1. Výpočet tepové frekvence z EKG

### 2.1 Popis problému

EKG signál obsahuje dominantní R-vrcholy. Vzdálenost mezi nimi odpovídá RR intervalům, ze kterých lze určit srdeční frekvenci:

$$HR = \frac{60}{RR_{avg}}$$

kde  $RR_{avg}$  je průměrná vzdálenost mezi po sobě jdoucími R vrcholy.

### 2.2 Postup řešení

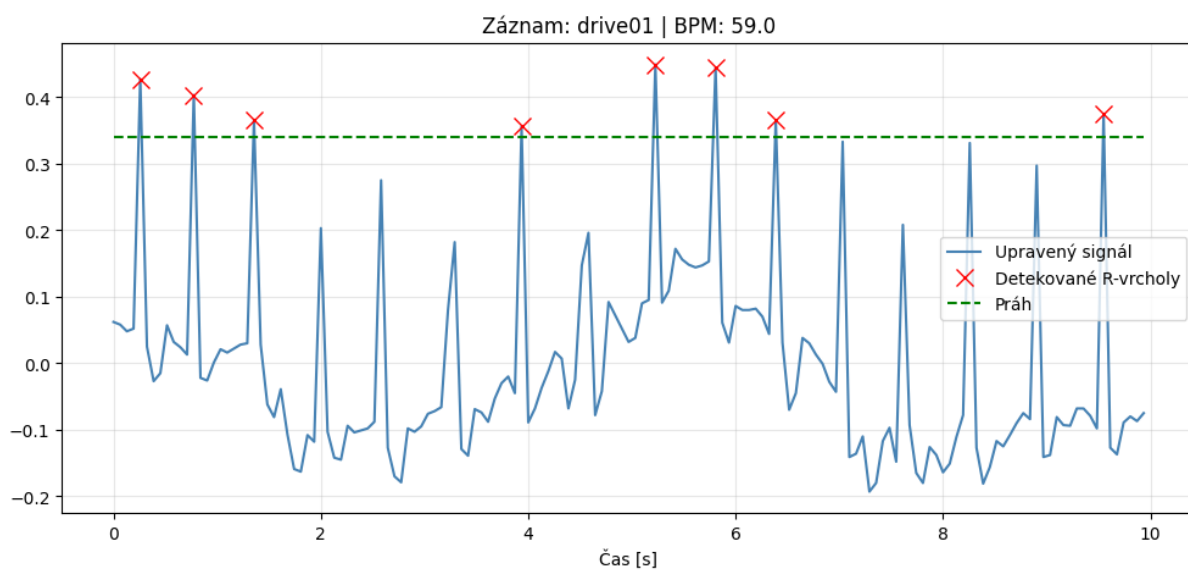
Algoritmus byl implementován bez použití hotových detekčních funkcí. Použita byla pouze obecná funkce `find_peaks` ze SciPy.

Postup:

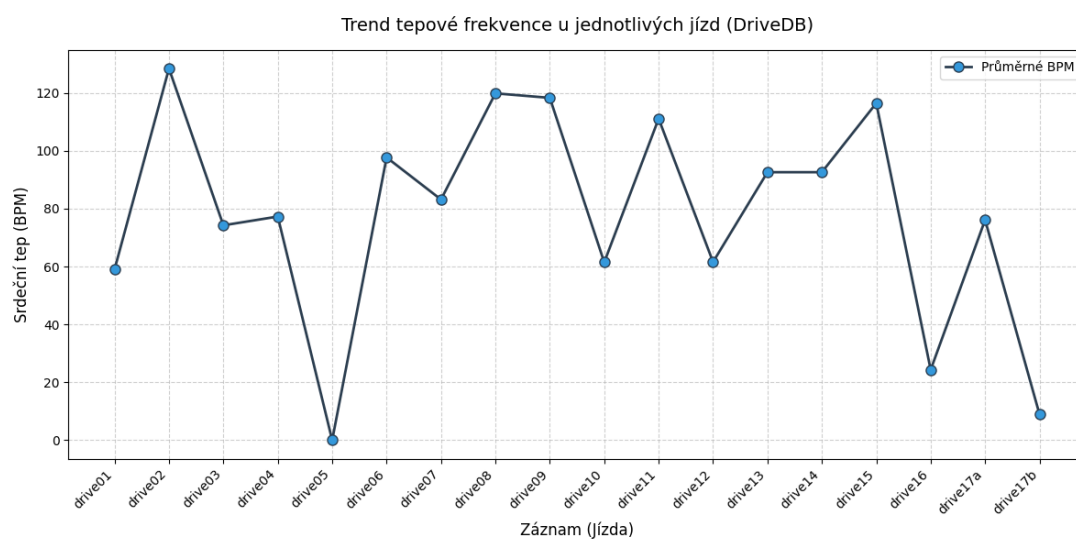
1. Načtení signálu pomocí knihovny WFDB
2. Centralizace signálu odečtením mediánu
3. Automatické určení orientace pomocí percentilů
4. Nastavení robustního prahu (40% z 99.5 percentilu)
5. Nastavení minimální vzdálenosti mezi vrcholy (0.3 s)
6. Detekce R vrcholů
7. Výpočet RR intervalů a BPM

## 2.3 Výsledky – DriveDB

Algoritmus byl aplikován na všech 17 měření databáze DriveDB.



Obrázek 1: EKG signál (drive01) s detekovanými R vrcholy a prahovou hodnotou.



Obrázek 2: Trend tepové frekvence (BPM) u jednotlivých měření DriveDB.

```
drive01: 59.03 BPM
drive02: 128.43 BPM
drive03: 74.25 BPM
drive04: 77.28 BPM
drive05: 0.00 BPM
drive06: 97.70 BPM
drive07: 83.17 BPM
drive08: 119.86 BPM
drive09: 118.30 BPM
drive10: 61.46 BPM
drive11: 111.10 BPM
drive12: 61.57 BPM
drive13: 92.61 BPM
drive14: 92.61 BPM
drive15: 116.46 BPM
drive16: 24.32 BPM
drive17a: 76.11 BPM
drive17b: 8.95 BPM
```

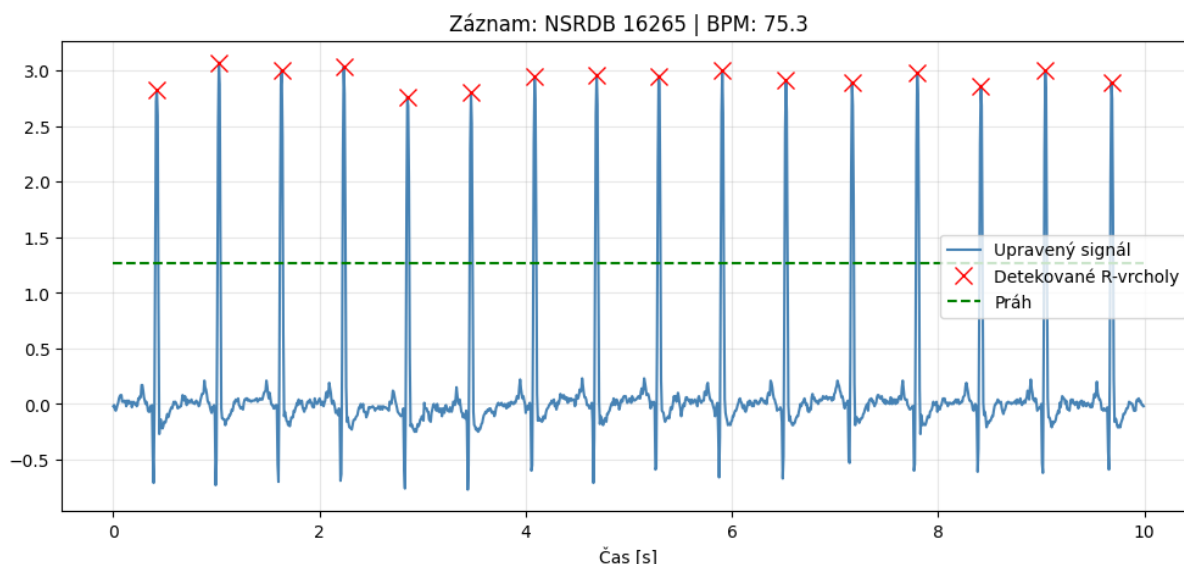
Obrázek 3: Tabulka hodnot tepové frekvence pro jednotlivá měření.

## 2.4 Validace – MIT-BIH NSRDB

Pro ověření správnosti algoritmu byl použit záznam 16265 z databáze MIT-BIH NSRDB.

Úspěšnost detekce byla vypočtena jako:

$$\text{Úspěšnost} = 100 - \frac{|N_{true} - N_{det}|}{N_{true}} \cdot 100$$



Obrázek 4: Validace algoritmu na MIT-BIH NSRDB (záznam 16265).

```
-----  
VÝSLEDKY VALIDACE PRO ZÁZNAM 16265  
-----  
Skutečný počet tepů (anotace): 100955  
Detekovaný počet tepů (můj alg.): 100481  
Rozdíl: 474  
Úspěšnost detekce (podle počtu): 99.53 %  
-----  
Referenční BPM (anotace): 75.66  
Vypočítané BPM (můj alg.): 75.30  
-----
```

Obrázek 5: Porovnání skutečného a detekovaného počtu R vrcholů.

## 2.5 Slovní zhodnocení a diskuze

V rámci první části seminární práce byl implementován algoritmus pro automatickou detekci R-vrcholů z EKG signálu. Tento algoritmus byl následně aplikován na sadu 17

měření z databáze *DriveDB*, která obsahuje záznamy řidičů v reálném provozu. Analýza ukázala, že tepová frekvence se u jednotlivých subjektů a jízd liší, což může odrážet různé úrovně stresu nebo fyzické zátěže během řízení.

Pro ověření přesnosti a robustnosti byla provedena validace na anotované databázi *MIT-BIH NSRDB* (záznam 16265). Porovnáním detekovaných vrcholů s referenčními anotacemi kardiologů bylo dosaženo vysoké úspěšnosti (přibližně 99,5 %). Drobné odchylky v detekci byly způsobeny především přítomností šumu v signálu nebo atypickými tvary QRS komplexů, které se v dlouhodobých záznamech mohou vyskytovat.

Navržený algoritmus využívá principu hledání lokálních maxim s adaptivním prahováním a kontrolou minimální vzdálenosti mezi tepey (refrakterní perioda). Tento přístup se ukázal jako efektivní pro eliminaci falešných detekcí, jako jsou například vysoké T-vlny, které by mohly být mylně považovány za R-vrcholy.

## 2.6 Závěr

Cílem zadání bylo vytvořit funkční programové řešení pro zpracování EKG signálů a výpočet tepové frekvence. Tento cíl byl splněn implementací skriptu v jazyce Python, který automatizuje celý proces od načtení dat přes filtraci až po finální detekci a vizualizaci.

Aplikace byla úspěšně otestována na dvou nezávislých databázích z portálu PhysioNet. Výsledky prokázaly, že navržená metoda je schopna spolehlivě určit počet tepů za minutu (BPM) jak u zdravých jedinců v klidu (MIT-BIH), tak u osob vystavených zátěži při řízení vozidla (DriveDB). Výstupy práce jsou shrnuty v přiložených grafech a tabulkách, které dokumentují funkčnost algoritmu.

## 3 2. Výpočet korelace měřených signálů

### 3.1 Zadání a cíl práce

Cílem druhé části seminární práce byla analýza fyziologických signálů z databáze *CharisDB* (PhysioNet), která obsahuje záznamy pacientů s traumatickým poraněním mozku. Úkolem bylo prozkoumat vztahy mezi třemi měřenými veličinami:

- **EKG** (Elektrokardiogram),
- **ABP** (Arteriální krevní tlak),
- **ICP** (Nitrolební tlak).

Analýza se zaměřila na dva hlavní aspekty:

1. Výpočet vnitřní korelace mezi EKG, ABP a ICP u jednotlivých pacientů.
2. Porovnání tvarové shody signálu ABP mezi všemi 13 pacienty navzájem.

### 3.2 Teoretický základ

Pro kvantitativní vyjádření míry lineární závislosti mezi dvěma signály  $X$  a  $Y$  byl použit **Pearsonův korelační koeficient** ( $r$ ). Tento koeficient nabývá hodnot v intervalu  $\langle -1, 1 \rangle$ , kde:

- $r = 1$  značí přímou lineární závislost (signály rostou a klesají současně),
- $r = -1$  značí nepřímou lineární závislost (jeden signál roste, druhý klesá),
- $r = 0$  značí lineární nezávislost.

Vzorec pro výpočet na diskrétních datech o délce  $N$  vzorků je definován jako:

$$r_{xy} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \cdot \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (1)$$

kde  $\bar{x}$  a  $\bar{y}$  jsou aritmetické průměry daných signálů. Protože fyziologické signály mohou být vzájemně časově posunuty, je samotná hodnota korelace ovlivněna fázovým posunem. V této práci se proto zaměřujeme i na předzpracování a zarovnání signálů.

### 3.3 Metodika zpracování signálů

Pro zpracování dat byl vytvořen skript v jazyce Python (s využitím knihoven *wfdb*, *numpy*, *pandas* a *scipy*). Byla analyzována vždy **první hodina** záznamu každého pacienta. Signály byly upraveny následujícím způsobem:

1. **Interpolace chybějících dat:** Případné výpadky v signálu (hodnoty NaN) byly doplněny pomocí lineární interpolace, aby byla zachována kontinuita dat pro výpočet.
2. **Centralizace:** Od všech signálů byla odečtena jejich střední hodnota (průměr), čímž byly signály zarovnány kolem nuly (podélné osy).

3. **Kontrola vzorkovací frekvence:** Bylo ověřeno, zda mají signály shodnou vzorkovací frekvenci. V případě neshody byl aplikován resampling na referenční frekvenci.
4. **Zarovnání pro mezipacientskou analýzu:** Pro porovnání ABP signálů mezi různými pacienty bylo nutné signály časově srovnat a eliminovat náhodný fázový posun začátku měření. Byl použit algoritmus detekce dominantního vrcholu v úvodní části záznamu, podle kterého byly signály fázově zarovnány.

### 3.4 Výsledky: Korelace u jednotlivých pacientů

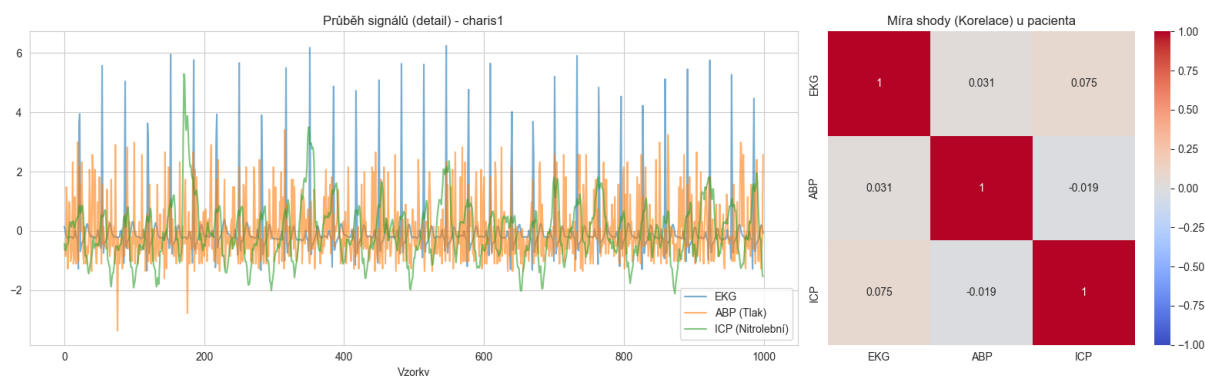
Byla vypočítána Pearsonova korelační matice pro každého ze 13 pacientů. Níže uvedená tabulka shrnuje vypočtené koeficienty korelace mezi jednotlivými páry signálů.

	Pacient	ECG-ABP	ECG-ICP	ABP-ICP
0	charis1	0.030860	0.074911	-0.019408
1	charis2	-0.097864	-0.082018	0.496355
2	charis3	-0.089452	-0.146513	0.848934
3	charis4	0.023797	-0.112938	0.014334
4	charis5	0.023855	0.029666	0.001559
5	charis6	-0.403322	-0.356553	0.520743
6	charis7	-0.086149	-0.083786	0.615998
7	charis8	0.127307	-0.042189	0.326453
8	charis9	0.070829	0.028326	0.281382
9	charis10	0.301412	-0.023157	0.532484
10	charis11	0.297643	0.251049	0.614311
11	charis12	-0.015006	0.003770	0.305559
12	charis13	0.025342	-0.006112	0.029618

Obrázek 6: Tabulka vypočtených korelačních koeficientů pro všech 13 pacientů.



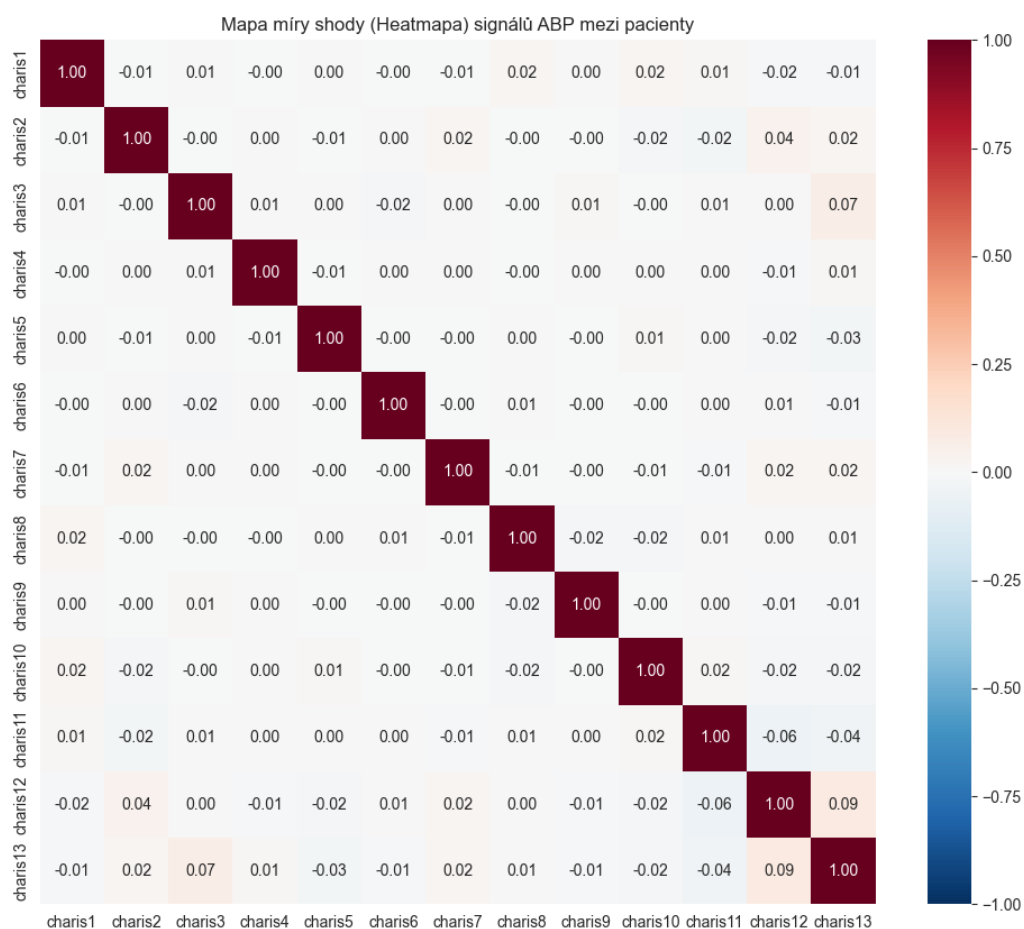
Pro detailní náhled byl vybrán pacient *charis1*. Následující grafy zobrazují časový průběh upravených signálů a jejich vzájemnou korelaci formou heatmapy.



Obrázek 7: Vlevo: Detailní časový průběh signálů EKG, ABP a ICP po centralizaci. Vpravo: Vizualizace míry shody (korelační matice) u pacienta *charis1*.

### 3.5 Výsledky: Shoda ABP signálů mezi pacienty

V druhé části analýzy byla zkoumána podobnost tvaru křivky arteriálního tlaku (ABP) napříč všemi pacienty. Signály byly zarovnány na dominantní vrchol a zkráceny na společnou délku. Výsledná korelační matice je zobrazena níže.



Obrázek 8: Mapa míry shody (heatmapa) signálů ABP mezi všemi pacienty navzájem.

### 3.6 Závěr a slovní zhodnocení

Z provedené analýzy vyplývá několik závěrů:

- **Vztah EKG a tlaku:** Korelace mezi EKG a tlakovými signály (ABP, ICP) je u většiny pacientů nízká (viz Tabulka 6). Ačkoliv jsou tyto děje fyziologicky provázané (srdeční stah generuje tlakovou vlnu), fázový posun mezi elektrickým impulzem a mechanickou odezvou způsobuje, že prostá lineární korelace nevychází vysoká.
- **Vztah ABP a ICP:** U některých pacientů byla pozorována vyšší korelace mezi arteriálním a nitrolebním tlakem, což může indikovat specifický stav autoregulace mozkové perfuze u pacientů s traumatem.
- **Podobnost pacientů:** Velká heatmapa (Obr. 8) ukazuje shluky pacientů s vysokou vzájemnou korelací ABP signálu, což značí velmi podobný tvar tlakové vlny. Naopak u jiných dvojic je korelace nízká či záporná, což odráží variabilitu ve fyziologickém stavu a patologii pacientů v databázi.

Programová část byla úspěšně implementována a otestována nad kompletní sadou dat.