Univerzita Jana Evangelisty Purkyně, Ústí nad Labem

Počítačové zpracování signálu KI/PZS

Seminární práce

**Výpočet tepové frekvence z EKG signálu,**

**Výpočet korelace EKG signálů,**

**Klasifikace EMG signálů**

|  |  |
| --- | --- |
| ZS 2023/24 | Tomáš Ulrich  Osobní číslo: F21281 |

# Výpočet tepové frekvence z EKG signálů

## Zadání

Ve zdrojové databázi najdete celkem 17 měření EKG signálu. Signál je již filtrován a centralizován kolem podélné osy. EKG signál obsahuje dominantní píky, které se nazývají R vrcholy. Vzdálenost těchto vrcholů určuje dobu mezi jednotlivými tepy. Počet tepů za minutu je tedy počet R vrcholů v signálu o délce jedné minuty. Navrhněte algoritmus, který bude automaticky detekovat počet R vrcholů v EKG signálech a prezentujte tepovou frekvenci při jednotlivých jízdách/měřeních. Vás algoritmus následně otestujte na databázi MIT-BIH https://physionet.org/content/nsrdb/1.0.0/ a prezentujte jeho úspěšnost vzhledem k anotovaným datům z databáze.

## Postup řešení

### Načítání dat

Nejdříve je potřeba načíst data tak, abychom s nimi mohli pracovat. K tomu nám slouží funkce **„load\_data“,** která načításoubory databáze a vrací vzorky signálu EKG a seznam polí.

### Práh signálu

Pro odhad prahu signálu používáme funkci „estimate\_threshold“, která na základě zadaného percentilu (v našem případě 95%) odhaduje práh signálu.

### Filtr vysokých frekvencí

Pro filtraci signálu využíváme funkci „high\_pass\_filter“, která provede vysokofrekvenční filtraci signálu a vrátí nám vyfiltrovaný signál EKG.

### Odstranění baseline wander

Baseline wander způsobuje nežádoucí vertikální posunutí v signálu, což může komplikovat identifikaci prvků, jako jsou R píky. Pro odstranění tohoto nežádoucího prvku využijeme metodu „remove\_baseline\_wander\_fft“, která využívá Fourierovy transformace.

### Nalezení R píků

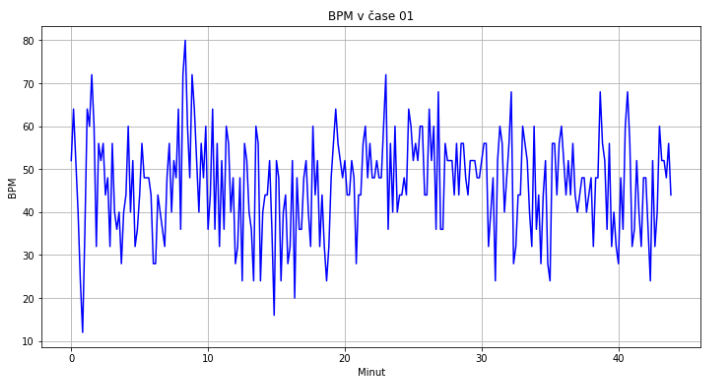
Pro nalezení R píků v EKG signálu využijeme funkce „find\_Rpeaks“, která hledá R píky v EKG signálu s dynamicky se měnící velikostí okna, což je klíčový prvek pří hledání R píků. Výstupem je seznam indexů R píků.

### Počet R píků v oknech

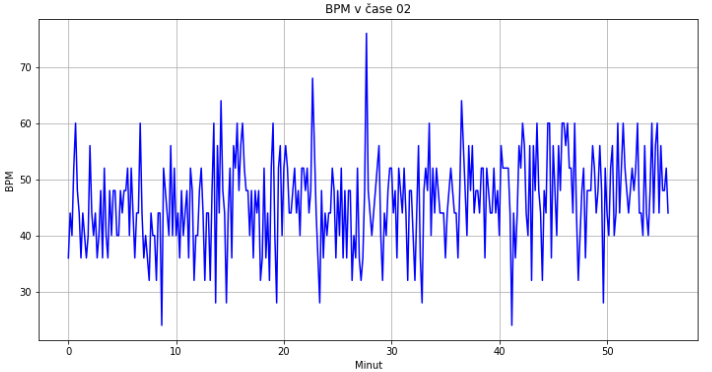
Pro nápočet R píků využijeme funkci „count\_rpeaks\_in\_windows“ počítá počet R píků v každých 15 sekundách. Výstupem je seznam počtu R píků v jednotlivých oknech.

## Výsledky

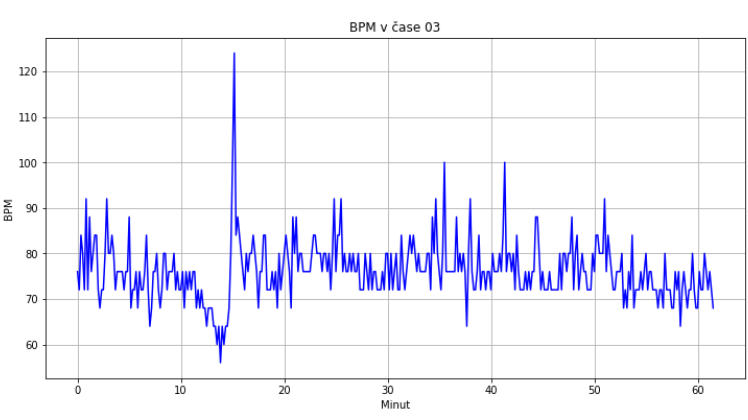
Grafy 1-3 zobrazují křivku průměrného srdečního tepu v čase.



**Graf 1**: BPM v průběhu času pro měření drive01

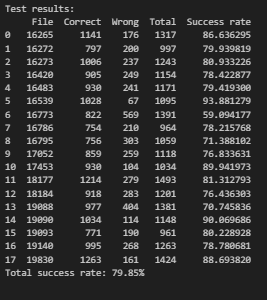


**Graf 2**: BPM v průběhu času pro měření drive02



**Graf 3**: BPM v průběhu času pro měření drive03

Algoritmus se otestoval na datech MIT-BIH <https://physionet.org/content/nsrdb/1.0.0/>. V tabulce níže jsou výsledky:



## Závěr

V grafech 1-3 vyčteme, že interval je reálný a odpovídá skutečnosti. Ovšem graf obsahuje nepřesnosti, kde hodnoty hodně oscilují za velmi krátký okamžik (např. 3. Graf).

Testování algoritmu na databázi MIT-BIH odhalilo celkovou úspěšnost 79,85%. Nejpřesnější měření bylo v souboru 16539 s přesností 93,89% a nejhorší přesnost měření se vyskytlo v souboru 16773 s přesností 59,09%.

# Výpočet korelace EKG signálů

## Zadání

Ve zdrojové databázi najdete celkem 17 měření obsahující EKG signál. Signály jsou již filtrované a centralizované kolem podélné osy. Různá měření jsou získána s různou vzorkovací frekvencí. U všech signálů analyzujte vzorkovací frekvenci a proveďte sjednocení na tu dominantní z nich. Pro tyto převzorkované signály proveďte korelační analýzu a prezentujte, jak jsou si signály napříč měřeními podobné. Pro smysluplné provedení této analýzy je potřeba nejprve srovnat signály na stejný počátek, např. dle pozice prvního dominantního R píku. Protože se délky jednotlivých signálů neshodují, je také nutné zvolit vhodnou délku korelační funkce, a to například analýzou autokorelačních funkcí samostatných signálů.

**Doplněk:**

V 1. seminární práci vidím dva signály 13 a 14, že jsou dle korelačního koeficientu stejné. Vezměte tyto dva signály a ukažte mi, že se korelační koeficient nemýlí. Využijte metody a přístupy, které uznáte za vhodné.

## Postup řešení

### Detekce R píků

Tato část algoritmu je stejná z předchozího řešení. Využíváme filtrování vysokých frekvencí a Fourierovu transformaci pro úpravu dat. Detekujeme R píky pro každý signál.

### Synchronizace

#### Převzorkování

V tomto kroku měníme vzorkovací frekvence signálu na jinou hodnotu (převzorkování). Využíváme převzorkování k přizpůsobení všech EKG signálů dominantní vzorkovací frekvenci, což umožňuje jejich vzájemné porovnání a analýzu na stejném časovém základě.

#### Synchronizace od prvního R píku

V tomto kroku nastavujeme časový počátek všech EKG signálů na okamžik detekce prvního R píku. Tato synchronizace zajistí, že všechny signály začínají ve stejný okamžik. To je dobré pro porovnávání a analýzu jejich charakteristik a vlastností.

#### Zkrácení na stejnou délku

Pro účely konzistentní analýzy se všechny synchronizované EKG signály zkracují na stejnou délku. Tímto způsobem se zajistí, že všechny signály mají stejný časový rozsah, což usnadňuje jejich porovnávání.

### Vytvoření korelační matice

Po provedení všech předchozích kroků, vytvoříme korelační matici pro všechny páry signálů. Využijeme funkci „numpy.corrcoef“, která takovou matici vytváří. Prvky na diagonále jsou nastaveny na „np.nan“, což zabraňuje samo-korelaci.

### Srovnání párů signálů s největší a nejmenší korelací

Po vytvoření korelační matice jsme srovnali absolutní hodnoty signálů a našli jsme dva páry signálů, kde jeden má maximální absolutní hodnotu korelačního koeficientu a ten druhý naopak nejmenší absolutní hodnotu korelačního koeficientu. Pro tyto dva páry jsme sestrojili korelační funkci pro vizualizaci jejich korelace.

### Ověření korelačního koeficientu

Signály 13 a 14 vyšli dle korelační matice s korelačním koeficientem 1. Pro ověření tohoto tvrzení jsem zvolil několik metod, které by měli toto tvrzení buď potvrdit nebo vyvrátit. Seznam zvolených metod: Určení směrodatných odchylek a průměrů obou signálů, posouzení odezvy nejvyšší cross-korelace, grafické znázornění obou signálů, knihovna scipy.stats (Pearsonův test, Spearmanův test, Kendall-Tau test), Scatter plot, Srovnání distribuce signálů

## Výsledky

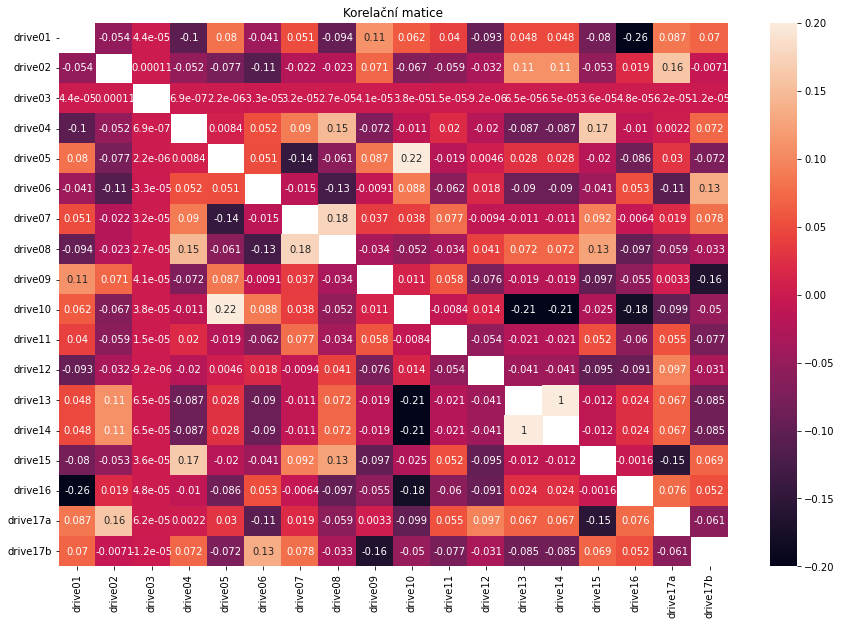
Níže jsou 3 korelační matice v zobrazení teplotní mapy:

* 1. matice je vytvořena jedním aplikováním funkce pro synchronizaci signálu
* 2. matice je vytvořena po třech opakováních funkce pro synchronizaci
* 3. matice po sedmi opakováních.



**Korelační matice 1**: aplikace funkce pro synchronizaci signálu jedenkrát



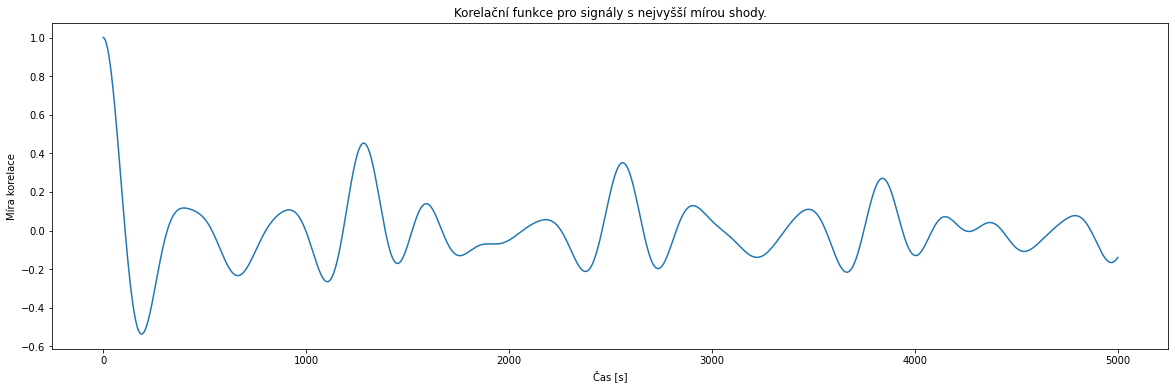


**Korelační matice 2**: aplikace funkce pro synchronizaci signálu třikrát

**Korelační matice 3**: aplikace funkce pro synchronizaci signálu sedmkrát

Z korelačních matic po sedmi opakováních funkce pro úpravu dat jsme usoudili dvě měření a ty znázornili v grafech:

1. S největší shodou to je měření drive13 a drive14 s mírou korelace 1 po sedmi opakováních funkce pro úpravu dat (znázorněno na grafu 1)



**Graf 1:** korelační funkce pro signály s nejvyšší mírou shody

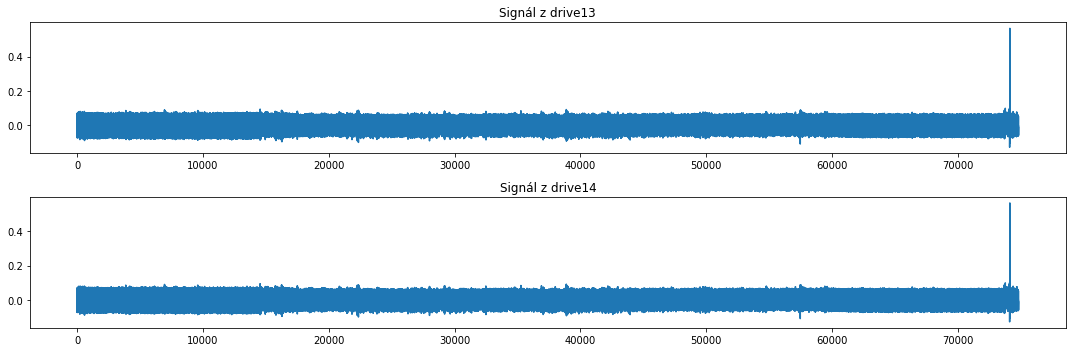
Obsah obrázku text, řada/pruh, Písmo, snímek obrazovky

Popis byl vytvořen automaticky2. S nejmenší shodou, a to jsou měření drive1 a drive16 s mírou korelace 6.85 ∙ 107(zobrazeno na grafu 2)

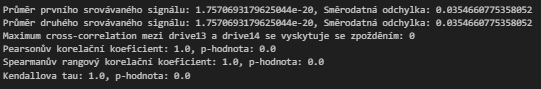
**Graf 2:** korelační funkce pro signály s nejnižší mírou shody

Zde jsou výsledky srovnání signálu 13 a 14:

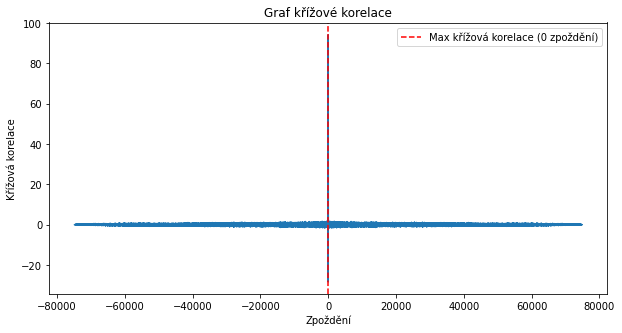
* Grafické zobrazení obou signálů

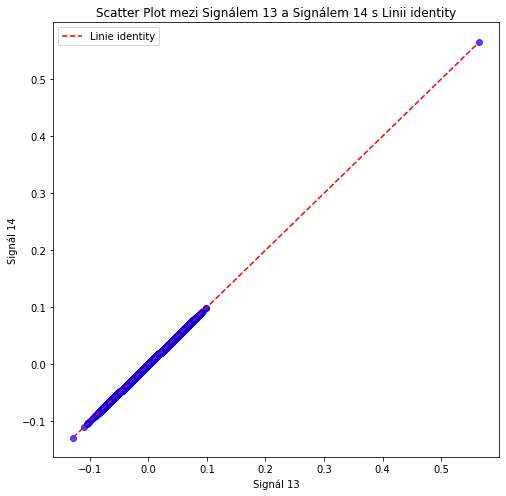


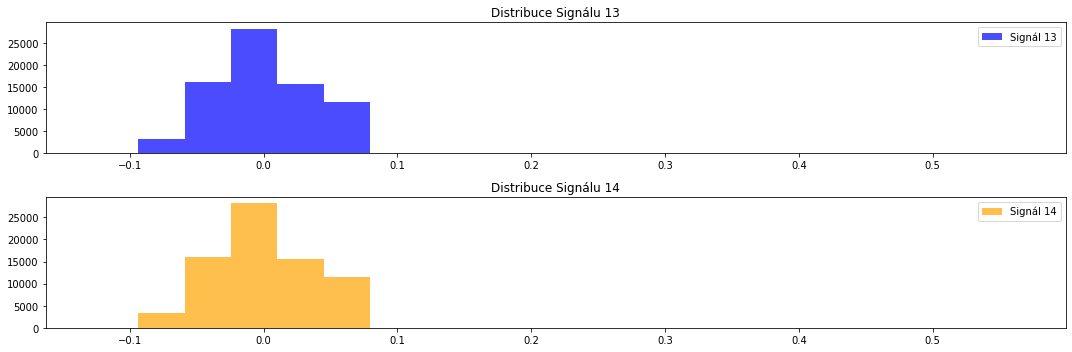
* Směrodatné odchylky a průměry, Odezva nejvyšší Cross-korelace, statistické testy



* Grafické znázornění:







## Závěr

Z korelační matice lze usoudit, že vzájemná podobnost mezi všemi signály je velmi nízká, což naznačuje, že neexistují výrazné strukturální podobnosti mezi těmito signály. Výjimkou jsou pouze signály 13 a 14, které vykazují výraznou shodu. Po opakované synchronizaci a úpravě signálů dochází k nárůstu korelace mezi signály. Musíme však vzít v úvahu, že úpravou dochází k částečné ztrátě jeho původní podoby, což může způsobit vyšší míru korelace při opakovaném zpracování.

Zajímavým pozorováním je, že měření 02 vykazuje téměř nulovou korelaci s ostatními měřeními i při opakované úpravě signálů.

Analyzujeme-li korelační funkce pro signály s nejvyšší a nejnižší mírou shody, lze pozorovat zřetelný rozdíl mezi nimi. Signály s nejvyšší mírou shody mají korelační funkce pomalé oscilace s výraznými amplitudami, naopak signály s nejnižší shodou mají oscilace frekvenčně vysoké a amplitudy jsou měnší.

Při zkoumání korelace mezi signálem 13 a 14 jsem dospěl k závěru, že korelační koeficient se nemýlí, protože všechny testy, které jsem provedl vedli ke stejnému závěru.

# Klasifikace EMG signálů

## Zadání

Ve zdrojové databázi najdete celkem 17 měření EMG signálu. Signál je již filtrován a centralizován kolem podélné osy. EMG signály zachycují aktivitu svalů během jízdy. Pro všech 17 měření spočítejte integrované EMG (iEMG) pomocí vzorce

kde, t je doba záznamu, je EMG signál a symbolizuje absolutní hodnotu. Dále detekujte oblasti, kde u jednotlivých signálů dochází k nárůstu a poklesu aktivity, a to pomocí okénkové varianty iEMG a derivace funkce. Velikost okénka zvolte tak, aby byly výsledky statisticky spolehlivé.

## Postup řešení

### Načítání dat

Načítání dat probíhá za pomocí funkce „**load\_data“**, která načte soubor databáze EKG signálu pomocí knihovny „**wfdb**“ a vrátí vzorky signálu EKG a seznam polí. Při načítání se také provede vysokofrekvenční filtrace signálu.

### Vysokofrekvenční filtr

Vysokofrekvenční filtr je v našem případě realizován funkcí „**high\_pass\_filter**“. Filtr pomáhá eliminovat nežádoucí nízkofrekvenční složky v EKG signálu, což zvyšuje kvalitu signálu a usnadňuje následnou analýzu.

### Odstínění signálu

Odstínění signálu je provedeno funkcí „**rectify\_signal**“. Tato funkce převádí všechny hodnoty signálu na jejich absolutní hodnoty. Tento převod umožňuje zachytit amplitudové změny bez ohledu na jejich směr.

### Integrace signálu

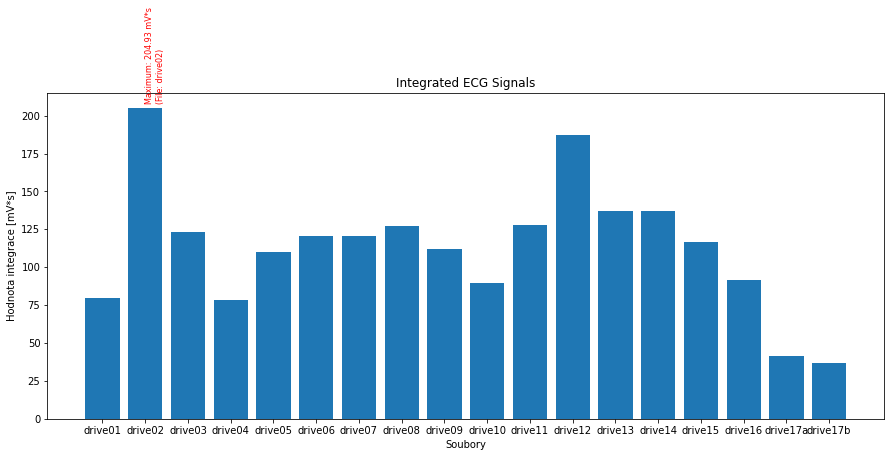
Integrace signálu je provedena funkcí „**integrate\_signal**“ tak, že použije konvoluce a okenního průměru. Integrace může sloužit k získání informací o trendu nebo celkové akumulaci signálu v daném časovém okně.

### Detekce změn v signálu

Posledním krokem je funkce „**detect\_changes\_in\_signal**“. Tato funkce detekuje změny v signálu na základě jeho gradientu. Pro každé okno signálu se vypočte gradient a rozhodne se, zda gradient přesahuje zadaný prah. Výsledek je seznam obsahující informace o změnách v podobě trojic (začátek okna, konec okna, jestli signál stoupá).

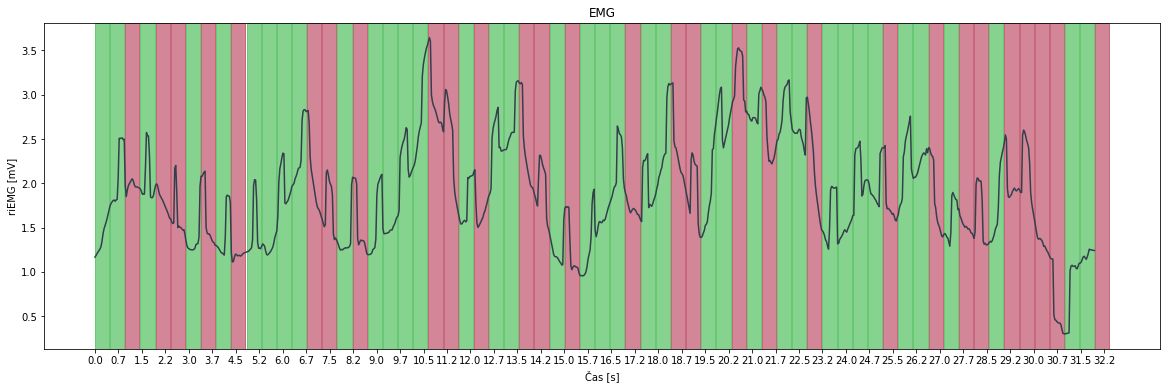
## Výsledky

Prvním krokem je tvorba histogramu. Z histogramu můžeme vidět, že soubor drive02 má nejvyšší hodnotu 204.93 mV ∙ s.



**Graf 1:** integrace EMG signálů

Dalším krokem je sestrojení grafu s vyznačenými oblastmi, ve kterých docházelo k nárůstu. Tento graf jsme vytvořili pro soubor drive03.



**Graf 2:** Oblastmi nárůstu a poklesu EMG aktivity

## Závěr

Práce s EKG signály byla náročná i přesto, že signály již byly dle zadání filtrovány a centralizovány dle podélné osy. Bylo potřeba udělat několik pokusů a dojít ke správně optimalizaci.

Z histogramu sice vyplívá, že soubor drive02 dosahuje nejvyšší hodnoty integrace, ale dle poznatků od kolegy, který mi při této seminární práci výrazně vypomáhal (Alex Schönfelder) vím, že tento soubor může být pro náš algoritmus obtížně zpracovatelný. Proto může docházet k zobrazení zkreslených hodnot integrovaného signálu pro tento konkrétní soubor.