

BIOSIGNAL CHALLENGE 2017

detekce epileptických **mikrozáchvatů** v programovém prostředí MATLAB

Ing. Radek Janča, Ph.D.

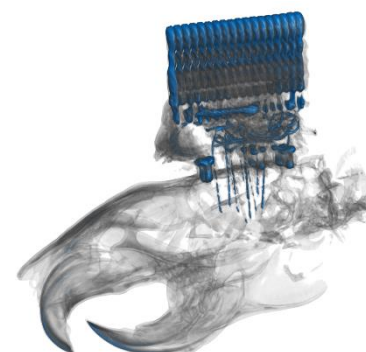
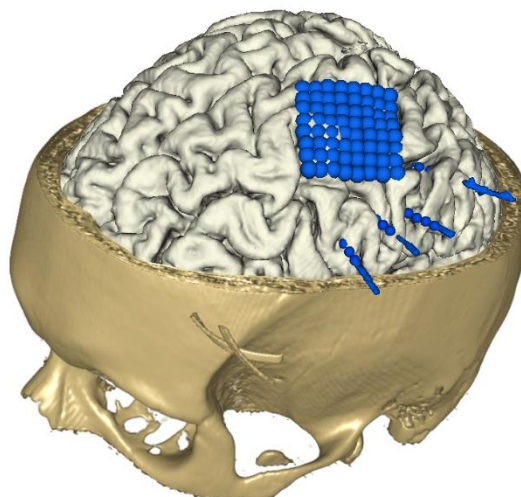
Ing. Jan Kudláček

1. Úvod

Epilepsie je závažné neurologické onemocnění, které ve vyspělých zemích postihuje okolo 0,5-1 % populace. Příčiny, průběh a závažnost nemoci jsou různorodé, avšak společným rysem je vznik spontánních epileptických záchvatů [1]. Výzkum v oblasti příčin vzniku epilepsie, sledování změn ve funkci neurální sítě vedoucích k náchylnosti k záchvatům a samotná iniciace záchvatů je limitován variabilitou mezi pacienty a ve většině případů neinvazivními diagnostickými metodami v humánní medicíně. Z těchto důvodů je nezbytné využívání animálních modelů (potkanů), u kterých lze navodit takové stavy, které umožňují sledovat změny vedoucí ke vzniku epilepsie [2]. Odhalení podstaty epileptogeneze je klíčové pro vývoj nových léků a terapeutických postupů, ze kterých budou profitovat budoucí pacienti.

Předmětem zkoumání jsou signály EEG měřené invazivně z oblasti hipokampu potkanů, které byly zasaženy tetanus-toxinem. Podobu několika týdnů byly sledovány změny v EEG, postupný nárůst patologické mozkové aktivity, a to do doby vzniku spontánních epileptických záchvatů [3]. V tomto meziobdobí se v EEG začínají objevovat patologické grafoelementy (mikrozáchvaty [4], interiktální výboje [5], vysokofrekvenční oscilace [3] aj.), které mění svoji četnost a parametry ve vztahu k fázi epileptogeneze.

Cílem soutěže je navrhnout autonomní algoritmus v prostředí MATLAB, který dokáže v dlouhodobých EEG záznamech **spolehlivě identifikovat jeden typ aktivity (mikrozáchvaty) a současně ignorovat další abnormní aktivitu (interiktální výboje, záchvaty)**. Návrh a vývoj algoritmu vyžaduje znalosti z oblasti digitálního zpracování signálů [6, 7] a umělé inteligence [8].



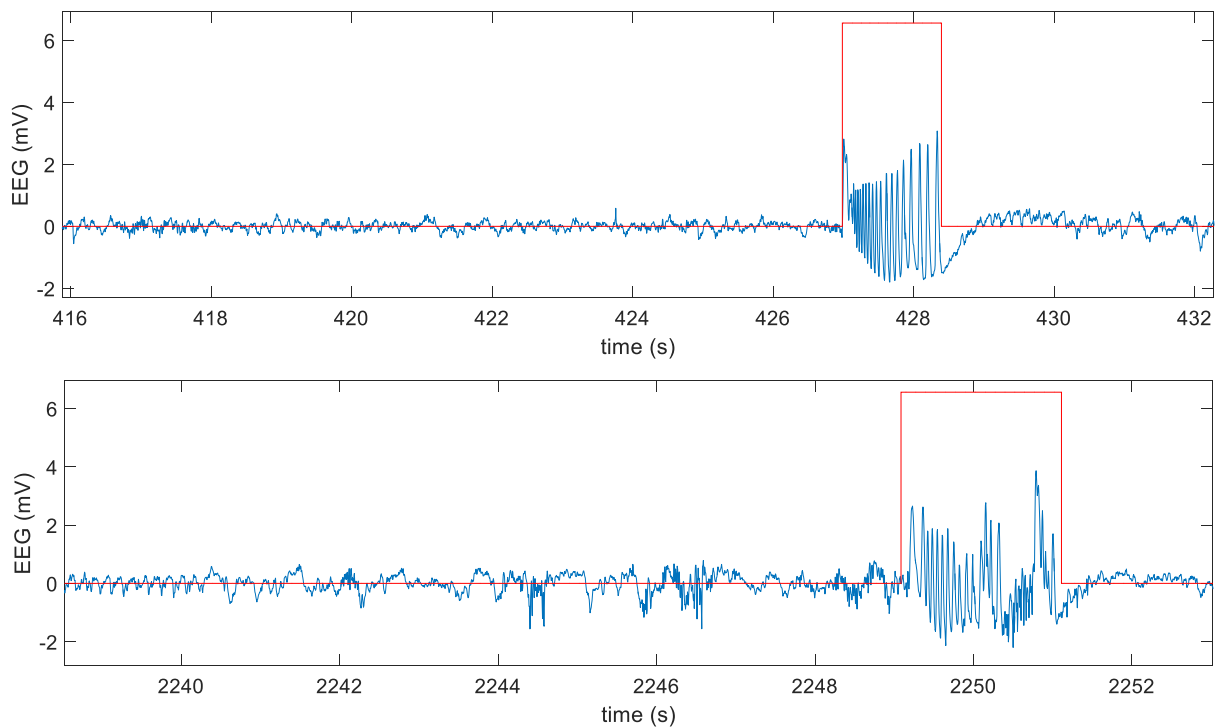
2. Data a metody

Dlouhodobé EEG záznamy čtyř zkoumaných subjektů byly vizuálně prohlédnuty a ručně označeny úseky mikrozáchvatů a záchvatů jako *zlatý standard* (Gold-Standard). Vhodné úseky signálů byly asi po hodině segmentovány. Vybrané segmenty byly náhodně rozděleny do *trénovací* (25 %) a *testovací* (75 %) skupiny. Každá skupina dat obsahuje záznamy pokrývající různé fáze epileptogeneze:

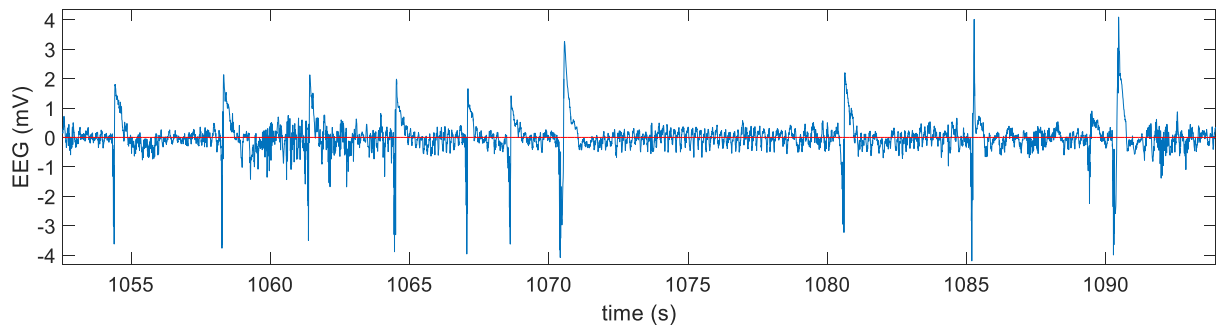
- 1) bez mikrozáchvatů a záchvatů
- 2) mikrozáchvaty společně se záchvaty
- 3) nízký počet mikrozáchvatů ($<q_{0.25}$)
- 4) střední počet mikrozáchvatů ($q_{0.25}-q_{0.75}$)
- 5) vysoký počet mikrozáchvatů ($>q_{0.75}$)

2.1. Mikrozáchvat

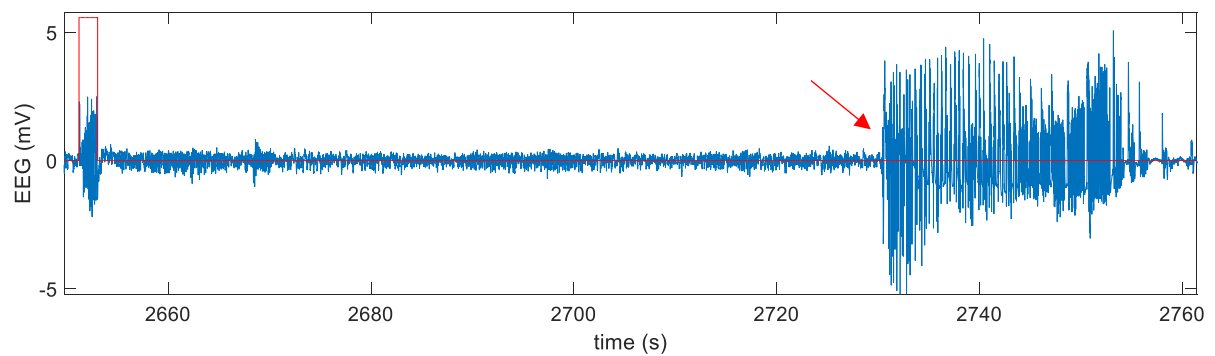
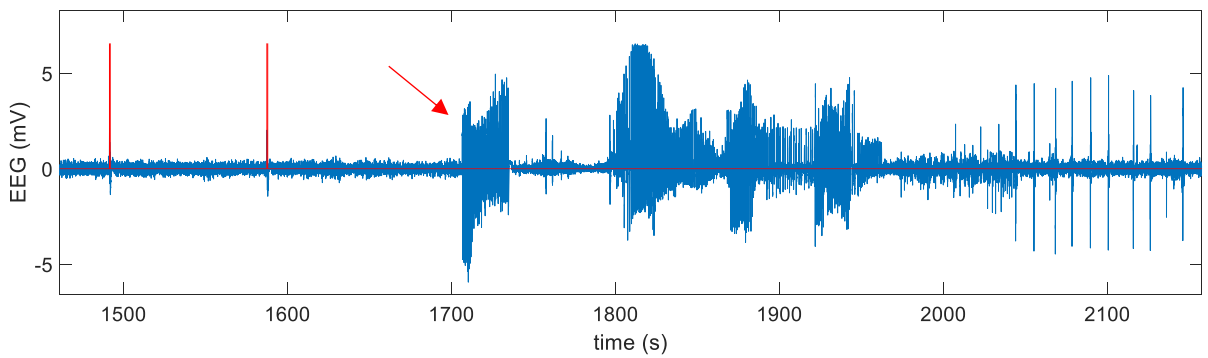
Jedná se o specifický grafoelement, který lze popsat jako ostrou vlnu, na kterou nasedá různý počet výbojů, vyznačeno červeným obdélníkem. Perioda mezi výboji s časem zpravidla roste, naopak amplituda výbojů roste. Délka mikrozáchvatů se pohybuje v průměru $1,3\pm 1,6$ sekundy, ale mohou dosahovat až několika sekund. Odlišitelnost mikrozáchvatu od aktivity pozadí kolísá v závislosti na bdění a spánku zvířete, fázi epileptogeneze a přítomnosti rušení.



Tvar a charakter interiktálních epileptiformních výbojů (IED) a jejich nakupení (až 2x za sekundu) může vézt k záměně s mikrozáchvaty. Výboje jsou nicméně ostře ohraničené s následnou pomalou vlnou, někdy s nasedající vysokofrekvenční aktivitou (HFO). Na rozdíl od mikrozáchvatu, po IED se neobjevuje rytmická zpomalující se aktivita.














Nejmarkantnější rozdíl mezi mikrozáchvaty a záchvaty je zejména v délce. Záchvaty trvají desítky sekund až minut, obsahují velké množství širokospektrální vysoce energetické aktivity, rytmických vzorců a interiktálních výbojů. Začátek záchvatu je označen šipkou. Mikrozáchvaty lze považovat za krátké, jasně ohraničené grafoelementy.



2.2. Struktura dat

Bylo připraveno 192 trénovacích souborů v MATLAB formátu. Název jednotlivých souborů zahrnuje kód subjektu, datum a čas monitorace a další systémové informace.

 jk20151109_1-151113_181603-A33c_A-d-006.mat	1. 2. 2017 16:22	Soubor MAT	9,864 kB
 jk20151109_1-151116_191406-A33c_A-d-027.mat	1. 2. 2017 16:22	Soubor MAT	9,864 kB
 jk20151109_1-151118_180451-A33c_A-d-055.mat	1. 2. 2017 16:22	Soubor MAT	5,923 kB
 jk20151109_1-151122_071635-A33c_A-d-133.mat	1. 2. 2017 16:23	Soubor MAT	5,928 kB
 jk20151109_1-151122_113844-A33c_A-d-137.mat	1. 2. 2017 16:23	Soubor MAT	5,926 kB
 jk20151109_1-151122_191728-A33c_A-d-144.mat	1. 2. 2017 16:23	Soubor MAT	5,930 kB
 jk20151109_1-151122_202300-A33c_A-d-145.mat	1. 2. 2017 16:23	Soubor MAT	5,932 kB
 jk20151109_1-151123_004508-A33c_A-d-149.mat	1. 2. 2017 16:23	Soubor MAT	5,924 kB
 jk20151109_1-151123_092925-A33c_A-d-157.mat	1. 2. 2017 16:23	Soubor MAT	5,924 kB
 jk20151109_1-151123_145045-A33c_A-d-160.mat	1. 2. 2017 16:23	Soubor MAT	9,864 kB
 jk20151109_1-151123_210035-A33c_A-d-164.mat	1. 2. 2017 16:23	Soubor MAT	9,864 kB

...

Každý soubor obsahuje následující proměnné:

d ... jednonábový segment EEG signál

fs ... vzorkovací kmitočet

tabs ... značka data+času v *datenum* formátu začátku segmentu signálu





sub_label ... struktura obsahující značky mikro-záchvatu a záchvatu v daném segmentu

sub_label.uSEIZ.posN ... začátky mikro-záchvatu v *datenum* formátu

sub_label.uSEIZ.durN ... trvání mikro-záchvatu v *datenum* formátu

sub_label.SEIZ.posN ... začátky záchvatu v *datenum* formátu

sub_label.SEIZ.durN ... trvání záchvatu v *datenum* formátu

 d	1310722x1 double	10485776
 fs	200	8
 sub_label	1x1 struct	1580
 tabs	7.3628e+05	8

Př. 1.: Převod časové osy relativního času na absolutní (*datenum* formát).

```
t_rel=linspace(0,length(d)-1)/fs,length(d);  
t_abs=datenum(0,0,0,0,0,t_rel)+tabs;
```

```
% index prvního mikro-záchvatu (ve vzorcích)
```

```
t_start=sub_label.uSEIZ.posN(1);
```

```
t_stop =sub_label.uSEIZ.posN(1) + sub_label.uSEIZ.durN(1);
```

```
idx=find(t_abs>=t_start & t_abs<=t_stop);
```

```
useiz_signal=d(idx);
```

2.3. Hodnotící kritéria

Výkon jednotlivých detektorů je stanoven skupinou parametrů, které vycházejí z porovnání expertního hodnocení a výsledků detekce. Ačkoliv je expertní hodnocení považováno za referenční tzv. zlatý standard (Gold Standard), nelze očekávat neomylnost v označování mikrozáchvatů a záchvatů. Vzhledem k celkovému popisu mnohatýdenních záznamů, obtížnému odlišení základní a patologické mozkové aktivity, subjektivních kritériích hodnotitele apod. lze očekávat např. přehlédnutí jasných grafoelementů expertem, nebo naopak označení aktivity, která kritéria mikrozáchvatu zcela nesplňuje. Od detektorů tedy nelze očekávat totožné výsledky, ale svými výkony by se měly referenci co nejvíce blížit [3, 9].

TP – (true positive) počet správných detekcí mikrozáchvatu

FP – (false positive) počet chybných detekcí

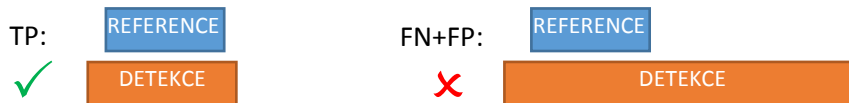
FN – (false negative) počet přehlédnutých mikrozáchvatů detektorem

Definice *TP*:

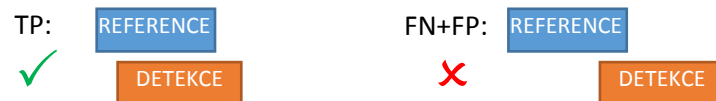
- 1) Detekce uvnitř reference. Délka detekce >50 % délky reference, jinak *FN+FP*.



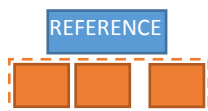
- 2) Detekce překrývá referenci. Celkový přesah detekce <50 % délky reference, jinak *FN+FP*.



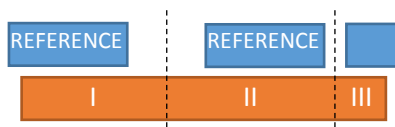
- 3) Detekce je posunutá od reference. Společný překryv s referencí >50 % a přesah <50 % délky reference.



- 4) Vícenásobná detekce uvnitř reference je sloučena a řídí se dle pravidel 1-3.



- 5) Detekce překrývající více referencí je rozdělena. Každá část je hodnocena pravidly 1-3.



Úspěšnost a výkonost detektoru je stanovena následujícími parametry:

- Senzitivita (*SEN*) definuje, kolik procent mikrozáchvatů bylo detekováno. Ideální hodnota je 100 %.

$$SEN = 100 \cdot \frac{TP}{TP + FN}$$

- Prediktivní hodnota pozitivního testu (Predictive Positive Value - *PPV*) definuje, kolik procent z detekcí bylo úspěšných. Ideální hodnota je 100 %.

$$PPV = 100 \cdot \frac{TP}{TP + FP}$$

Celkové hodnocení jednotlivých účastníků soutěže bude sestavena dle následujícího klíče:

- *SEN* na trénovacích datech (20 %)
- *PPV* na trénovacích datech (20 %)
- *SEN* na neveřejných testovacích datech (20 %)
- *PPV* na neveřejných testovacích datech (20 %)
- Zpracování průvodní dokumentace (teoretický rozbor, metodika, výsledky, diskuze), originalita řešení, výpočetní náročnost (20 %).

2.4. Detektor – vstupy a výstupy

Volací funkce detektoru musí být pojmenována `uSeizure_2017` a musí obsahovat předepsané vstupy a výstupy. Externí funkce mohou nést libovolný název, vyvarujte se ovšem přejmenování standardních MATLAB funkcí.

```
label=uSeizure_2017(d, fs, tabs, fileName);
```

Vstupy:

```
d ...analyzovaný signál  
fs ...vzorkovací kmitočet  
tabs ...absolutní čas začátku signálu  
fileName ...název analyzovaného souboru
```

Výstupy:

```
label ...struktura, obsahující detekce  
label.posN(1, :) ...řádkový vektor detekcí v datetime formátu (tabs+relativní časy detekcí)  
label.durN(1, :) ...řádkový vektor trvání detekce v datetime formátu  
label.fileName ...stringový název analyzovaného souboru
```

2.5. Validace

Pro snadné testování úspěšnosti detektoru je připravena validační funkce BSGvalidation2017_rev0.m, která výstup detektoru porovná s referenčními Gold-Standard daty uloženými v GS_train.mat. Validační funkce umí porovnat jednotlivé soubory i celou skupinu souborů.

```
label=uSeizure_2017(d,fs,tabs,fileName);
DET(1).label=label;
GS=load('GS_train.mat');

STAT=BSGvalidation2017_rev0(DET,GS)

% INPUTS:-----
% DET... struktura obsahující detekce a název souboru
%   DET.label.posN ...řádkový vektor detekcí v datenum formátu
%   DET.label.durN ...řádkový vektor trvání detekcí v datenum formátu
%   DET.fileName='jk20151109_1-151113_181603-A33c_A-d-006.mat'
%   ...stringový název souboru, který byl analyzován
% Pozn.: Statistiku více souborů lze provést vytvořením struktury DET(:,1)
%
%
% GS... struktura "gold standard" obsahuje referenční labaly k testovacím
% datům obsažených v souboru GS_train.mat
%   GS.fileName ... cell - stringový seznam souborů
%   GS.subject ... cell - stringový seznam subjektů
%   GS.label ... struct - obsahuje referenční labely
%   GS.label.posN ...řádkový vektor detekcí v datenum formátu
%   GS.label.durN ...řádkový vektor trvání detekcí v datenum formátu
%   GS.label.subjectID ...index v seznamu GS.subject{}
%   GS.label.fileNameID ...index v seznamu GS.fileName{}
% Pozn.: GS=load('GS_train.mat');
%
% OUTPUTS:-----
% STAT... struktura výsledků
%   STAT.TP(:,1) ...počet správných (true positive) detekcí.(*)
%   STAT.FP(:,1) ...počet chybných (false positive) detekcí.(*)
%   STAT.FN(:,1) ...počet přehlédnutí (false negative). (*)
%   STAT.total_TP ...celkový počet správných (true positive) detekcí.
%   STAT.total_FP ...celkový počet chybných (false positive) detekcí.
%   STAT.total_FN ...celkový počet přehlédnutí (false negative).
%   STAT.SEN ...sensitivita detektoru SEN=100*TP/(TP+FN).(*)
%   STAT.PPV ...přesnost detektoru PPV=100*TP/(TP+FP).(*)
%   STAT.total_SEN ...celková sensitivita detektoru
%   STAT.total_PPV ...celková přesnost detektoru (positive predictive value)
%   STAT.UP_err ...průměrná absolutní chyba začátku detekce (sekund).(*)
%   STAT.DOWN_err ...průměrná absolutní chyba konce detekce (sekund).(*)
%   STAT.avrg_UP_err ...celková průměrná abs. chyba začátku detekcí (sekund)
%   STAT.avrg_DOWN_err ...celková průměrná abs. chyba konce detekcí (sekund)
% (*)Hodnoty vektoru korespondují se vstupem DET(:,1)
```

Př. 2.: Hodnocení trénovací skupiny souborů

```
GS=load('GS_train.mat');

train_list=dir('*.mat'); % všechny trénovací soubory
for i=1:size(train_list,1)
    D=load(train_list(i).name);
    DET(i).label=uSeizure_2017(D.d,D.fs,D.tabs,train_list(i).name);
end
STAT=BSGvalidation2017_rev0(DET,GS);
```

3. Podmínky soutěže

- Soutěž je určena studentům všech vysokých škol České a Slovenské republiky
- Všechny algoritmy detekce musí být naprogramovány v prostředí MATLAB s uvedením verze produktu.
- Detektory navrhnete a optimalizujete na trénovacích datech <http://sami.fel.cvut.cz/biosignalchallenge2017/data.zip>
- Na adresu biosignalchallenge@fel.cvut.cz zašlete zdrojové kódy v podobě funkcí ve formátu (*.m) společně s protokolem ve formátu (*.doc, *.docx, *.pdf). Struktura protokolu se řídí dle standardních zvyklostí vědecké práce (Úvod, Metodika, Výsledky, Diskuze) a musí obsahovat popis všech použitých technik s odkazy na zdroje
- Své práce zasílejte do **19. května 2017, 24:00**
- Výsledky budou zveřejněny na webovských stránkách a v prostorách ČVUT FEL v týdnu od **29. května do 2. června 2017**. (Změna termínu vyhlášení je vyhrazena, všichni účastníci budou informováni)
- Na základě celkových výsledků budou oceněny tři nejlepší práce formou finančního daru: **1. místo 400,- USD, 2. místo 350,- USD, 3. místo 250,- USD**

Reference

- [1] Fisher, R.S., *et al.* ILAE official report: a practical clinical definition of epilepsy. *Epilepsia* **55**, 475-482 (2014).
- [2] Jiruska, P., *et al.* Synchronization and desynchronization in epilepsy: controversies and hypotheses. *The Journal of physiology* **591**, 787-797 (2013).
- [3] Jiruska, P., *et al.* Epileptic high-frequency network activity in a model of non-lesional temporal lobe epilepsy. *Brain: a journal of neurology* **133**, 1380-1390 (2010).
- [4] Stead, M., *et al.* Microseizures and the spatiotemporal scales of human partial epilepsy. *Brain*, awq190, (2010).
- [5] Janca, R., *et al.* Detection of interictal epileptiform discharges using signal envelope distribution modelling: application to epileptic and non-epileptic intracranial recordings. *Brain topography*, **28.1**: 172-183 (2015).
- [6] Alan, V. Oppenheim, W. Schafer Ronald, and R. B. John. "Discrete-time signal processing." New Jersey, Printice Hall Inc (1989).
- [7] Jan Uhlíř, Pavel Sovka, Roman čmejla. „Úvod do číslicového zpracování signálů“, Praha, Vydavatelství ČVUT (2003).
- [8] Duda, Richard O., Peter E. Hart, and David G. Stork. *Pattern classification*. John Wiley & Sons, (2012).
- [9] Casson, Alexander J., Elena Luna, and Esther Rodriguez-Villegas. "Performance metrics for the accurate characterisation of interictal spike detection algorithms." *Journal of neuroscience methods* **177.2** (2009): 479-487.