



**ČESKÉ VYSOKÉ
UČENÍ TECHNICKÉ
V PRAZE**

F3

**Fakulta elektrotechnická
Katedra kybernetiky**

Bakalářská práce

Analýza spánkových EOG signálů

Svetlana Ulitina

leden 2019

Vedoucí práce: Ing. Václav Gerla, Ph.D.

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Ulitina** Jméno: **Svetlana** Osobní číslo: **453544**
Fakulta/ústav: **Fakulta elektrotechnická**
Zadávající katedra/ústav: **Katedra kybernetiky**
Studijní program: **Otevřená informatika**
Studijní obor: **Informatika a počítačové vědy**

II. ÚDAJE K BAKALÁŘSKÉ PRÁCI

Název bakalářské práce:

Analýza spánkových EOG signálů

Název bakalářské práce anglicky:

Analysis of Sleep EOG Signals

Pokyny pro vypracování:

Hlavním těžištěm práce je provést analýzu EOG signálů zaznamenaných během spánku a sestavit algoritmy, které umožní rozpoznat klinicky významné vzory v těchto signálech. Sestavené metody se stanou součástí širšího řešení pro rozpoznávání spánkových tříd (bdělost, REM, NREM) a umožní přesnější analýzu dlouhodobých PSG záznamů.

Cíle práce:

- implementace vybrané metody adaptivní segmentace EOG signálu,
- implementace sady vhodných popisných příznaků (např. statistické příznaky, frekvenční příznaky, nebo koeficienty waveletové transformace),
- implementace vybraného klasifikátoru,
- validace navržených metod nad reálnými EOG signály.

Studentka pro svou práci využije databázi EOG signálů, např. The CAP Sleep Database

(<https://physionet.org/pn6/capslpdb/>), případně použije EOG signály, které jí poskytne vedoucí práce.

Seznam doporučené literatury:

- [1] Iber C, Ancoli-Israel S, Chesson A, and Quan SF for the American Academy of Sleep Medicine. The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, 1st. Ed.: Westchester, Illinois: American Academy of Sleep Medicine, 2007.
- [2] Malaekah E, Shahrababaki SS, Cvetkovic D. Automatic Sleep Stage Detection and Classification: Distinguishing Between Patients with Periodic Limb Movements, Sleep Apnea Hypopnea Syndrome, and Healthy Controls Using Electrooculography (EOG) Signals. J Bioprocess Biotech 5:216, 2015.
- [3] Estrada , Nazeran H, Barragan J, Burk J R, Lucas E A, Behbehani K. EOG and EMG: Two Important Switches in Automatic Sleep Stage Classification. Engineering in Medicine and Biology Society, 2006. EMBS '06. IEEE, 2006.
- [4] Moser D; Anderer P; Gruber G; Parapatics S; Loretz E; Boeck M; Kloesch G; Heller E; Schmidt A; Danker-Hopfe H; Saletu B; Zeitlhofer J; Dorffner G. Sleep classification according to AASM and rechtschaffen & kales: effects on sleep scoring parameters. SLEEP;32(2):139-149, 2009.

Jméno a pracoviště vedoucí(ho) bakalářské práce:

Ing. Václav Gerla, Ph.D., kognitivní neurověda CIIRC

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) bakalářské práce:

Datum zadání bakalářské práce: **12.01.2018**

Termín odevzdání bakalářské práce: **08.01.2019**

Platnost zadání bakalářské práce: **30.09.2019**

Ing. Václav Gerla, Ph.D.
podpis vedoucí(ho) práce

doc. Ing. Tomáš Svoboda, Ph.D.
podpis vedoucí(ho) ústavu/katedry

prof. Ing. Pavel Ripka, CSc.
podpis děkana(ky)

III. PŘEVZETÍ ZADÁNÍ

Studentka bere na vědomí, že je povinna vypracovat bakalářskou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v bakalářské práci.

Datum převzetí zadání

Podpis studentky

Poděkování / Prohlášení

Chtěla bych poděkovat vedoucímu práce Ing. Václavu Gerlovi, Ph.D. za věcné připomínky a účinnou metodickou, pedagogickou a odbornou pomoc.

Děkuji také Mgr. Lucii Černé za pomoc při gramatické kontrole práce.

Mé díky patří i rodině, především mému manželovi, za podporu a pomoc během dosavadního studia.

Prohlašuji, že jsem předloženou práci vypracovala samostatně, a že jsem uvedla veškeré použité zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

V Praze dne 08.01.2019

.....

Abstrakt / Abstract

Závěrečná práce se zabývá analýzou celonočních spánkových EOG signálů a klasifikací segmentů signálů do dvou spánkových stavů: bdělost a REM.

Práce zahrnuje popis provedené segmentace signálu a extrakce sady vhodných popisných příznaků pomocí vlnkové transformace.

Dále je popsána selekce nejvýznamnějších příznaků pomocí algoritmu sekvenčního dopředného výběru příznaků.

Následuje učení klasifikátoru k-NN na základě množiny nejvýznamnějších příznaků a porovnávají se výsledky klasifikace segmentů o různých délkách.

Praktická část závěrečné práce je implementována v prostředí Matlab. Navržené metody jsou ověřeny na reálných EOG signálech.

Klíčová slova: spánek, EOG, spánkové stavy, segmentace, CWT, k-NN

This bachelor thesis deals with an analysis of the whole-night EOG recordings and also with a classification of signals' segments of two sleep stages: wake and REM.

The thesis includes a description of a performed signal segmentation and a feature set extraction using a wavelet transformation.

Further, it describes a sequential forward selection algorithm to reduce the feature set.

The previous step is followed by a k-NN classifier training and an evaluation of classification results of signals' segments with different length.

The practical part of the bachelor thesis was implemented using a Matlab environment. Proposed methods were validated on real EOG signals.

Keywords: sleep, EOG, sleep stages, segmentation, CWT, k-NN

Title translation: Analysis of Sleep EOG Signals

Obsah /

1 Úvodní rozbor zadané problematiky	1
1.1 Spánek	1
1.1.1 Spánkové stavy	2
1.1.2 Polysomnografie	3
1.1.3 Význam EOG v analýze spánku	3
1.2 Extrakce příznaků	4
1.2.1 Fourierova transformace	5
1.2.2 Vlnková transformace	5
1.3 Segmentace signálu	6
1.4 Selektce příznaků	6
1.4.1 Sekvenční dopředný výběr příznaků	6
1.4.2 PCA pro vizualizaci dat	7
1.5 Klasifikace	8
1.5.1 k-NN klasifikátor	8
1.5.2 Křížová validace	8
1.5.3 Hodnocení klasifikátorů	9
1.5.4 ROC prostor a ROC křivka	10
2 Popis řešení zadaného úkolu	11
2.1 Data set	11
2.2 Spektrální analýza	12
2.3 Transformace signálů	14
2.3.1 Segmentace signálů	14
2.3.2 První úroveň dvouúrovňové segmentace signálů	15
2.3.3 Druhá úroveň dvouúrovňové segmentace signálů	18
2.4 Selektce příznaků	18
2.4.1 Vizualizace příznakového prostoru	19
2.5 k-NN klasifikace	20
2.6 Popis programu	23
3 Závěr	25
Literatura	26
A Zkratky	29
B Schematické znázornění implementační části závěrečné práce	30
C Časově frekvenční analýza WAKE fáze	31
D Časově frekvenční analýza REM fáze	34
E SFS pro segmenty různé velikosti	37
E.1 Průběh SFS pro 5s segment ...	37
E.2 Průběh SFS pro 10s segment ..	38
E.3 Průběh SFS pro 15s segment ..	38
E.4 Průběh SFS pro 30s segment ..	39
F Matice záměn pro klasifikaci segmentů různé velikosti	40
G Adresářová struktura DVD disku ..	41

Tabulky / Obrázky

1.1. Obecný tvar matice záměn	10
2.1. Počet 30s oken pro jednotlivé subjekty	17
2.2. SFS výsledky pro segmenty různé délky	19
2.3. Obsah data setu pro segmenty různé délky	21
2.4. Optimální hodnoty parametru k pro k-NN klasifikátor	22
2.5. Hodnocení k-NN klasifikátoru pro testovací subjekty 19, 20	22
F.1. Matice záměn k-NN klasifikace 5s segmentů	40
F.2. Matice záměn k-NN klasifikace 10s segmentů	40
F.3. Matice záměn k-NN klasifikace 15s segmentů	40
F.4. Matice záměn k-NN klasifikace 30s segmentů	40
1.1. Hypnogram - celonoční profil spánku dle R&K	3
1.2. Lokalizace elektrod pro spánkovou polysomnografie	3
1.3. Časově frekvenční rozložení definované STFT	6
1.4. Časově frekvenční rozložení definované WT	6
1.5. Grafová reprezentace SFS algoritmu	7
1.6. Náhodný klasifikátor v ROC prostoru	10
2.1. Časově frekvenční analýza WAKE fáze	12
2.2. Časově frekvenční analýza REM fáze	13
2.3. Algoritmus provedení 1. úrovně dvouúrovňové segmentace	16
2.4. Provedení 2. úrovně dvouúrovňové segmentace	17
2.5. Výpočet hodnoty příznaku na příkladě 15s segmentu	18
2.6. Matice příznaků sestavená ze submatic	18
2.7. Vizualizace nejvýznamnějších příznaků	20
2.8. ROC křivka pro klasifikaci vzorků o délce 5 s	22
2.9. ROC křivka pro klasifikaci vzorků o délce 10 s	22
2.10. ROC křivka pro klasifikaci vzorků o délce 15 s	23
2.11. ROC křivka pro klasifikaci vzorků o délce 30 s	23

Kapitola 1

Úvodní rozbor zadané problematiky

Daná kapitola obsahuje přehled teoretických základů metod a algoritmů, které byly použité v závěrečné práci.

Kapitola 1.1 poskytuje základní informace o spánku. Zahrnuje přehled nejdůležitějších historických událostí v oblasti výzkumu spánku, včetně existujících metodologií - R&K a AASM. V závěrečné práci se klade důraz na metodologii, jež je používána v medicínské praxi i v současné době - AASM. Kapitola vysvětluje význam polysomnografie a rozebírá role EOG signálu v analýze spánku. Ukazuje, že analýza EOG signálu má velký význam zejména při odlišení REM fáze od fáze probouzení. Zdůrazňuje, že pro odlišení NREM stavů mezi sebou již nestačí mít pouze EOG signál a že je třeba zohlednit jiné signály polysomnografie.

Kapitola 1.2 obsahuje teoretické základy o selekci příznaků. Zahrnuje přehled příznaků, na základě nichž se provádí automatická klasifikace. Věnuje se popisu Fourierové transformace, vlnkové transformace a přehledu vybraných publikací, v nichž se píše o selekci příznaků na základě zmíněných transformací.

Kapitola 1.3 vysvětluje problém segmentace signálu.

Kapitola 1.4 se věnuje problému redukce množiny příznaků. Zahrnuje teoretický základ sekvenčního dopředného výběru příznaků, který je dále aplikován v praktické části závěrečné práce. Uvede teoretický základ algoritmu PCA, který je aplikován pro vizualizaci klasifikačního problému.

Kapitola 1.5 obsahuje základní informace o klasifikaci a klasifikátoru k-NN, který byl použit pro rozpoznání spánkových stavů v praktické části závěrečné práce. Zahrnuje popis křížové validace jakožto přístupu, který dovoluje zabránit vzniku problému přeučení. Poté se kapitola věnuje způsobu hodnocení klasifikátoru pomocí ROC prostoru. Obsahuje definice ROC metrik pro binární klasifikaci spánkových stavů, které byly použity v závěrečné práci.

1.1 Spánek

V poslední době analýza spánku a spánkových stavů nabývá velké popularity. Nauka o spánku je ale relativně mladý obor. Vědci v tomto oboru se shodují, že za posledních 50 let se o spánku získalo více poznatků než za staletí vědy předtím. Josef Faber nazývá rok 1953, kdy byl objeven a popsán paradoxní spánek autory Aserinskym a Kleitmanem, „začátkem nové éry rozkvětu cílevědomého výzkumu spánku“. Období předcházející tomuto významnému roku se považuje za „Starověk hypnologie“ [1].

■ 1.1.1 Spánkové stavy

Série experimentů, kterou provedli američtí vědci Eugene Aserinsky a Nathaniel Kleitman v roce 1953, ukázala, že se mozková aktivita mění v průběhu spánku. Jejich pozorování uvádí nový typ očního pohybu během spánku - [rychlé pohyby očí](#) (angl. Rapid Eye Movement - **REM**). Účastníci experimentů byli pozorováni během celonočního spánku bez přerušení. Ukázalo se, že během noci se střídají typy oční aktivity několikrát [2].

V publikaci Aserinského a Kleitmana se poprvé rozlišují dvě základní fáze spánku: **REM** - spánek s rychlým očním pohybem a **NREM** - spánek bez očního pohybu.

Kritéria pro rozlišení různých fází spánku byla standardizována v roce 1968, kdy byl publikován [manuál Rechtschaffen a Kales \(R&K\)](#). Metodologie R&K rozdělila spánek na 5 disjunktních množin:

- REM;
- NREM1;
- NREM2;
- NREM3;
- NREM4.

Spánek v REM fázi se nazývá [asynchronní spánek](#), v NREM fázi - [synchronní spánek](#).

V roce 2007 byl Americkou akademií spánkové medicíny publikován [manuál pro hodnocení spánku a souvislých událostí](#) (angl. The AASM Manual for the Scoring of Sleep and Associated Events - **AASM**), podle kterého se provádí klasifikace spánku dodnes [3]. AASM změnila některá z pravidel pro hodnocení fází spánku. Nově se nerozlišuje mezi stavy NREM3 a NREM4 - metodologie je spojuje do jednoho spánkového stavu. Dle AASM se spánek rozděluje na 4 disjunktní množiny [3]:

- REM;
- NREM1;
- NREM2;
- NREM3.

Aserinský a Kleitman objevili, že fáze spánku se navzájem střídají. Pozorování celonočního spánku přineslo inovační objev: během noci se objevuje několik fází spánku s rychlým pohybem očí, které se střídají se spánkem bez rychlého pohybu [2].

Následující výzkumy v dané oblasti definují [spánkový cyklus](#) jako přechod mezi fázemi NREM - REM a ukazují, že během jednoho cyklu se fáze navzájem střídají v určitém pořadí.

Během noci nastává 4 až 6 cyklů spánku. Každý cyklus začíná fází **NREM1** a postupně se spánek prohlubuje, do **NREM2** a poté **NREM3** synchronního spánku. Před a po asynchronním spánku dochází k probouzení - fáze bdělosti (angl. wake - **WAKE**). Probouzení u zdravých subjektů zpravidla trvá několik vteřin [1]. Pro znázornění průběhu spánku se používá [hypnogram](#). Schematický příklad hypnogramu je uveden na obrázku 1.1.

1.1.2 Polysomnografie

Polysomnografie (angl. Polysomnography **PSG**) je multiparametrický medicínský test, který se provádí za účelem analýzy spánku. PSG zahrnuje snímání hodnot z elektrod napojených na pokožku spícího subjektu.

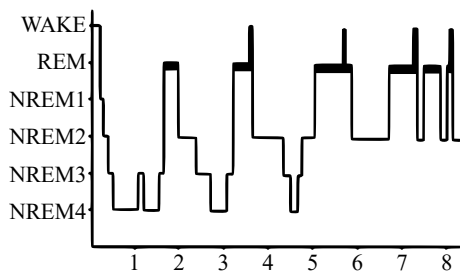
AASM poskytuje detailní instrukce jak pro napojení elektrod, tak i pro následnou analýzu záznamů [4].

Metodologie AASM definuje **minimální seznam signálů**, které mají být snímány pro analýzu spánkových stavů [5]:

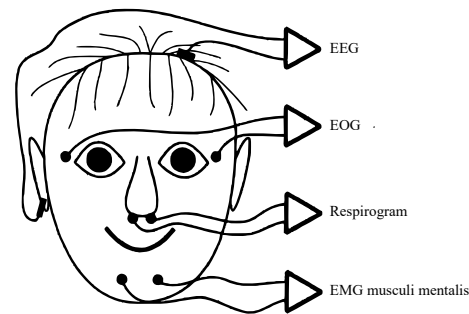
- elektroencefalografie **EEG** - snímání signálů generovaných z mozkové kůry;
- elektrookulografie **EOG** - registrace očních pohybů;
- elektromyografie **EMG** - snímání aktivity obličejových svalů.

Snímání hodnot dalších signálů (například respirogram pro registraci dechu) není povinné, ale dané záznamy mohou expertovi poskytnout doplňující informace.

Schematické rozmístění elektrod je na obrázku 1.2, kde je také zobrazeno rozmístění elektrod pro registraci dechu.



Obrázek 1.1. Hypnogram - celonoční profil spánku dle R&K, [1] Obr.100 s. 108



Obrázek 1.2. Lokalizace elektrod pro spánkovou polysomnografií, [1] Obr.111 s. 116

1.1.3 Význam EOG v analýze spánku

Zkoumání očního pohybu začíná na začátku 18. století. Thomas Eggert popisuje metody, které se používají pro nahrání záznamů pohybu oka [6]. První metody byly kontaktní, t.j. napojení elektrod bylo provedeno přímo na rohovku oka. Metoda EOG, která umožnila snímat elektropotenciál bez mechanického zásahu do oka, byla popsána Elwinem Margem v roce 1951 [7] a má řadu výhod:

- neohrožuje oko mechanickým poškozením, jelikož elektrody jsou napojené na pokožku ve vzdálenosti cca 1 cm od zevních očních koutků, viz obrázek 1.2;
- umožňuje nahrávat charakteristiky očního pohybu při zavřených očích;
- metoda je postavena na fyziologii lidského oka, které lze popsat jako elektrický dipól. EOG využívá skutečnost, že tento dipól se otáčí pohybem oka. To způsobuje malé rozdíly mezi elektrickým potenciálem na povrchu kůže v závislosti na poloze oka;
- umožňuje rozlišit horizontální a vertikální pohyby okem.

Hlavní výhoda EOG pro analýzu spánku spočívá v tom, že umožňuje snímat oční pohyb i při zavřených očích [6].

Zápis a klasifikaci spánkových stavů detailně popsal Alon Y. Avidan [5]. Zpravidla se prvních několik minut každého PSG záznamu plní **WAKE** fází. V této fázi EOG kanály zahrnují mrkání a rychlé pohyby očí. Signály se postupně zpomalují ve frekvenci a snižuje se jejich amplituda. Pokud jsou oči subjektu zavřené, EOG kanály mají podobné charakteristiky jako v REM fázi, což způsobuje potíže při manuální klasifikaci fáze spánku.

Z fáze **WAKE** je typický přechod do spánkové fáze **NREM1**, ale také je možné rychlé přepnutí z **WAKE** fáze do **NREM2** nebo do **REM**.

V průběhu upadání do spánku a fáze **NREM1** se oční pohyb začíná zpomalovat. Samotný pohyb očí ale nelze popsat univerzálně, jelikož se u různých subjektů projevuje jinak: škubavý a nepravidelný, nebo naopak mírný a jemný.

Ve fázích **NREM2**, **NREM3** a **NREM4** EOG zrcadlí chování EEG. Proto také byl spánek v těchto fázích pojmenován „synchronní“. Speciální kritéria, která by určovala a klasifikovala zmíněné fáze na základě EOG signálů, nejsou definovaná. Jelikož dané spánkové fáze jsou popsány EEG signály, není možné analyzovat spánkové stavy na základě samotných EOG, bez porovnání s ostatními PSG signály.

Během „asynchronního“ spánku přichází sny. Subjekt začíná rychle pohybovat očima pod zavřenými očními víčky.

Chování EOG signálů ve fázi **REM** je odlišné od EEG kvůli velké oční aktivitě a rychlému očnímu pohybu.

EOG signály mají velkou roli při analýze **REM** a **WAKE** fází. Je to způsobeno tím, že EEG v daných fázích mohou mít podobné charakteristiky a vizuálně si být podobné [5].

1.2 Extrakce příznaků

První pokusy o automatické zpracování spánkových záznamů vznikly před 30 lety [8].

V současné době se pro automatické zpracování a analýzu signálů používají různé metodologie. Jedna z nich je na základě množiny popisujících příznaků (angl. feature set).

Automatická klasifikace může být provedena na základě extrakce popisujících příznaků ze signálů a následující klasifikace.

Pro automatickou klasifikaci PSG dat je používána velká škála příznaků. Pro různé klinické studie v oblasti signálů se používají různé příznaky [9]:

- **příznaky ve frekvenční oblasti** (angl. frequency domain features), například frekvence vrcholů, koeficienty spektrální analýzy, energie po použití vlnkové transformace, hodnoty první a druhé derivace signálu v segmentech;
- **příznaky v časové oblasti** (angl. time domain features), například délka křivky, počet lokálních extrémů;

- **entropické příznaky** (angl. entropy based features), například spektrální entropie.

Reprezentace signálu v **časově frekvenční oblasti** (angl. Time-Frequency domain) může být zároveň nalezena pomocí integrální transformací (například **vlnkové**).

■ 1.2.1 Fourierova transformace

Fourierova transformace (angl. Fourier transform - **FT**) je jedním ze základních algoritmů číslicového zpracování signálu [9]. Je to metoda, která transformuje signál do frekvenční oblasti, čímž transformovaný signál ztrácí informaci o časové složce. Jeden ze způsobů výpočtu FT je algoritmus **rychlé Fourierovy transformace** (angl. Fast Fourier transform - **FFT**).

Fourierova transformace je významná v analýze signálů. V [10–13] můžeme nalézt použití FT k analýze EEG a v dalších publikacích slouží FT pro **výběr nejvýznamnějších příznaků** (angl. Feature Selection - **FS**) [11, 14–16].

Krátkodobá Fourierova transformace (angl. Short-time Fourier transform - **STFT**) je odvozena z FFT metody a je jednou z nejvíce používaných metod při analýze v časově frekvenční oblasti signálů, včetně EEG [10–11, 17]. Je to mechanismus, který umožňuje popsat spektra neboli frekvence nestacionárního signálu v libovolný okamžik času. Signál je segmentován na N částí (oken), na které se aplikuje FFT. Schematické znázornění segmentace je uvedeno na obrázku 1.3.

Frekvenční složka je obvykle reprezentována v podobě **spektrogramu** [18].

■ 1.2.2 Vlnková transformace

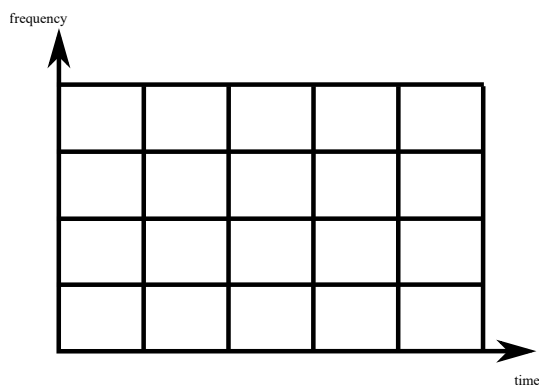
Vlnková transformace (angl. Wavelet transform - **WT**) je alternativou STFT metody. WT lze použít pro analýzu nestacionárního signálu v časově frekvenční oblasti. Na rozdíl od STFT, která aplikuje stejně velkou velikost okna pro transformaci celého úseku signálu, WT používá krátké okno pro vysokou frekvenci a dlouhé okno pro nízké frekvence [19].

Převod do časově frekvenční oblasti je realizován na základě vlnkové ortonormální báze pomocí škálování specializovaného vzoru vlny se změnou příslušných vah a postupnou aplikací na originální signál (obrázek 1.4). Tak vzniká matice koeficientů, jejichž hodnoty korespondují s hodnotou pro příslušné měřítko odpovídajícího vzoru [20].

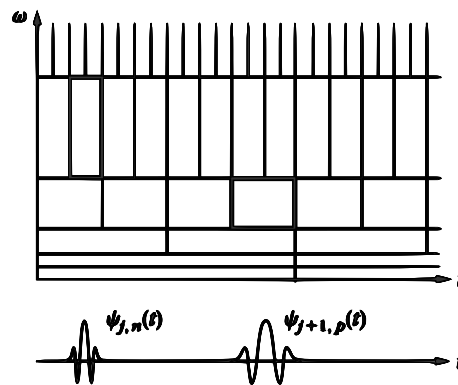
Existují dva hlavní typy WT, první je **spojitá vlnková transformace** (angl. Continuous Wavelet Transform - **CWT**), druhá je **diskrétní vlnková transformace** (angl. Discrete Wavelet Transform - **DWT**) [20].

Využití WT pro analýzu EEG je dnes velice populární. V článku Olivera Fausta [21] najdeme seznam publikací z oblasti analýzy EEG, které používají FS založený na CWT a DWT (více než 50 publikací mezi roky 2003 a 2014).

WT je lépe přizpůsobena pro analýzu rychlých změn signálu díky transformaci v různých frekvenčních pásmech. Publikace Amjeda S. Al-Fahouma porovnává různé metody FS pro EEG signály. Provádí klasifikaci na základě různých příznaků a porovnává výsledky. Ukazuje, že **klasifikace nestacionárního signálu** na základě příznaků vybraných



Obrázek 1.3. Časově frekvenční rozložení definované STFT [19] Obr. 2 a) s.16



Obrázek 1.4. Časově frekvenční rozložení definované WT [20] Obr. 1.6 s.20

z WT poskytuje lepší klasifikační výsledky než klasifikace na základě příznaků vybraných z FFT [16].

1.3 Segmentace signálu

Nestacionární biologické signály je nutné před zpracováním rozdělit do úseků kratší délky. Následně se provádí klasifikace daných úseků.

V literatuře se rozlišují dva druhy segmentace [9]:

- segmentace do úseků konstantní délky;
- adaptivní segmentace, v níž se délka segmentů mění na základě detekce změny stacionarity a odhadu přesného okamžiku změny.

Segmentace do úseku konstantní délky se nedoporučuje pro detekci grafoelementů (tvarových charakteristik signálu), jelikož hranice segmentů mohou rozdělit grafoelement do dvou segmentů.

Segmentace do úseku konstantní délky se používá při výběru příznaků z frekvenční oblasti (například STFT) nebo časově frekvenční oblasti (například CWT).

1.4 Selektce příznaků

Výběr nejvýznamnějších příznaků je prováděn za účelem snížení dimenzionality prostoru výběrem nejrelevantnější informace z originálních dat.

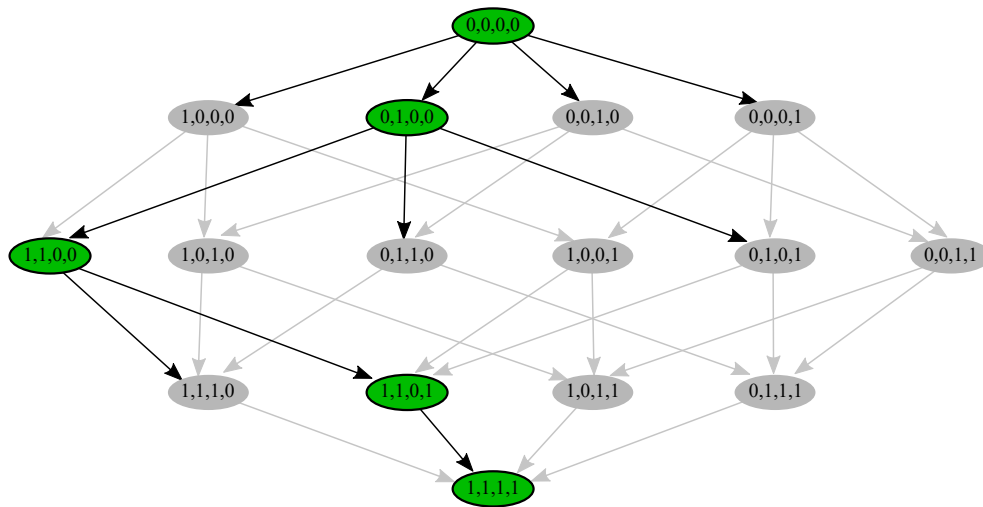
FS je úloha vyhledávání (angl. a search problem), v níž se hledá optimální nebo suboptimální podmnožina m příznaků z M původních pomocí ztrátové funkce.

1.4.1 Sekvenční dopředný výběr příznaků

Jedním z algoritmů na FS je **sekvenční dopředný výběr příznaků** (angl. Sequential Forward Selection - **SFS**).

Použití SFS pro výběr nejlepší sady příznaků ve výzkumu epileptických EEG signálů umožňuje dosáhnout vysoké přesnosti (angl. accuracy) klasifikace [22].

SFS je iterační algoritmus, který má grafovou reprezentaci, například schéma, které je znázorněno na obrázku 1.5.



Obrázek 1.5. Grafová reprezentace SFS algoritmu [23]. Hodnoty uzlů reprezentují vektor příznaků, kde 1 označuje příznak, který byl vybrán v jedné z předchozích iterací.

Na začátku první iterace se nastaví seznam významných příznaků na prázdný.

V prvním kroku se provede klasifikace tříd pro každý příznak zvlášť. Vypočte se chyba klasifikace. Příznak, pro který je chyba klasifikace nejmenší, je považován za nejlepší v této iteraci a je přidán na seznam vybraných příznaků.

V každé další iteraci se provádí klasifikace na základě sady již vybraných příznaků s přidáním jednoho nového. Příznak, jehož přidáním se chyba klasifikace zmenší nejvíce, bude přidán na seznam.

Algoritmus se zastaví, pokud nastane jeden ze tří případů:

- přidáním dalšího příznaku se nezmenší chyba;
- byl dosažen nastavený počet maximálních iterací;
- byly vybrány všechny příznaky.

1.4.2 PCA pro vizualizaci dat

Rozvoj podle hlavních komponent (angl. Principal Component Analysis, **PCA**) je jedna z metod, která se používá pro redukci dimenzionality.

Cílem PCA je odvození hlavních komponent, které nesou maximální informaci o původních proměnných. PCA je **optimalizační úloha hledání lineárního podprostoru dimenze q** k původnímu prostoru dimenze n , který minimalizuje součet čtverců vzdáleností k daným bodům v původním prostoru.

Jelikož ve vysokých dimenzích jsou data těžce reprezentovatelná, PCA se používá jako způsob vizualizace podobnosti a odlišnosti dat pomocí redukce dimenze.

Metoda je založena na reprezentaci bodů $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$ v prostoru dimenze n . Dále matice $\mathbf{A} = [\mathbf{a}_1 \mathbf{a}_2 \dots \mathbf{a}_m]^T \in R^{m \times n}$.

Spektrální rozklad matice A se vypočítá dle vzorce $\mathbf{A}^T \mathbf{A} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T$, kde $\mathbf{\Lambda}$ je diagonální matice vlastních čísel seřazených podle velikosti a \mathbf{V} odpovídá vlastním vektorům matice \mathbf{A} .

Dle věty o spektrálním rozkladu matice posledních q sloupců tvoří ortonormální bázi lineárního podprostoru dimenze q [24].

1.5 Klasifikace

Klasifikační úlohy patří k úlohám strojového učení. Algoritmy strojového učení se dělí podle způsobu učení do následujících kategorií [9]:

- **učení s učitelem** (angl. supervised learning) - učení, při kterém se systém trénuje z vnějšku. „Učitel“ zadává vstupní a výstupní vektory dat, vyhodnocuje výsledky a provádí změny;
- **učení bez učitele** (angl. unsupervised learning) - učení, při kterém systém nedostává informace z vnějšku. Učení je založeno pouze na informacích, které systém získal během samotného učení;
- **posílené učení** (angl. reinforcement learning) - způsob učení pomocí zpětné vazby.

1.5.1 k-NN klasifikátor

Algoritmus **k-nejbližších sousedů** (angl. k-Nearest Neighbours - **k-NN**) patří ke skupině klasifikátorů s učitelem. Nastavitelný parametr klasifikátoru k určuje požadovaný počet sousedů, který řídí přiřazení vzorku do jedné z možných tříd.

Definice klasifikační úlohy v obecném tvaru:

- množina tříd: $Y = \{y_j\}, j \in M$;
- množina vzorků: $X = \{x_i\}, i \in N$;
- trénovací množina: $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, kde každý prvek z X byl jednoznačně ohodnocen jednou z tříd Y ;
- metrika pro výpočet vzdálenosti mezi prvky: d .

Řada výzkumů používá k-NN klasifikátor pro práci s biologickým signálem. Ukazuje se, že je vhodný i pro klasifikaci spánkových stavů na základě EEG, jelikož poskytuje vysokou přesnost klasifikace [25].

1.5.2 Křížová validace

Křížová validace (angl. cross-validation) je jedna ze základních metod strojového učení.

Princip metody spočívá v rozdělení vstupní množiny dat na podmnožiny, které se považují za trénovací a testovací, kde testovací podmnožina je doplňkem trénovací podmnožiny v množině vstupních dat. Klasifikační úloha využije trénovací podmnožinu pro učení klasifikátoru a vyhodnotí přesnost klasifikátoru pomocí testovací podmnožiny. Rozdělení vstupní množiny a vyhodnocení klasifikátoru se opakuje několikrát.

Pro rozdělení vstupní množiny na podmnožiny se používá několik přístupu. Jeden z nich je princip založený na **vynechání jednoho vzorku z dat** (angl. Leave-One-Out - **LOO**), který představuje testovací množinu.

Křížová validace umožňuje **zabránit vzniku problému přeučení** (angl. overfitting) [26].

1.5.3 Hodnocení klasifikátorů

Hodnocení binárního klasifikačního systému lze provést pomocí ROC analýzy (angl. Receiver Operating Characteristic - **ROC**). ROC umožňuje vyhodnotit klasifikační metody.

Princip ROC analýzy detailně popsal Tom Fawcett [27].

Pro binární klasifikaci spánkových stavů WAKE a REM lze definovat ROC metriky:

- **Skutečné REM fáze** (angl. true REM - T_REM) - vzorek náleží fázi REM a je klasifikován jako REM;
- **Falešné REM fáze** (angl. false REM - F_REM) - vzorek náleží fázi WAKE, ale je klasifikován jako REM;
- **Skutečné WAKE fáze** (angl. true WAKE - T_WAKE) - vzorek náleží fázi WAKE a je klasifikován jako WAKE;
- **Falešné WAKE fáze** (angl. false WAKE - F_WAKE) - vzorek náleží třídě REM, ale je klasifikován jako třída WAKE;
- **Senzitivita** (angl. sensitivity) - je pravděpodobnost ohodnocení jako REM fázi pro skutečnou REM fázi

$$sensitivity = \frac{T_REM}{T_REM + F_WAKE};$$

- **Specifita** (angl. specificity) - je pravděpodobnost ohodnocení jako WAKE fázi pro skutečnou WAKE fázi

$$specificity = \frac{T_WAKE}{T_WAKE + F_REM};$$

- **Poměr správně ohodnocených REM fází** (angl. true REM rate - T_REM_RATE) - je poměr správně ohodnocených vzorků REM fází vůči celkovému počtu vzorku REM

$$T_REM_RATE = sensitivity;$$

- **Poměr špatně ohodnocených WAKE fází** (angl. false REM rate - F_REM_RATE) - je poměr špatně ohodnocených vzorků WAKE fází vůči celkovému počtu vzorků WAKE

$$F_REM_RATE = 1 - specificity;$$

- **Přesnost** (angl. accuracy) - je pravděpodobnost, s jakou klasifikátor poskytuje správné výsledky. Odhadujeme ji pro n vzorků jako

$$accuracy = \frac{T_REM + T_WAKE}{n};$$

- **Prediktivní hodnota WAKE fáze** (angl. predictive value WAKE) - je pravděpodobnost, že fáze je skutečně WAKE, když byla klasifikována jako WAKE

$$WAKE_PRED_VALUE = \frac{T_WAKE}{T_WAKE + F_WAKE};$$

- **Prediktivní hodnota REM fáze** (angl. predictive value REM) - je pravděpodobnost, že fáze je skutečně REM, když byla klasifikována jako REM

$$REM_PRED_VALUE = \frac{T_REM}{T_REM + F_REM}.$$

Uvedené případy se reprezentují v podobě 2x2 **maticí záměn** (angl. confusion matrix), obecný tvar je nastíněn v tabulce 1.1.

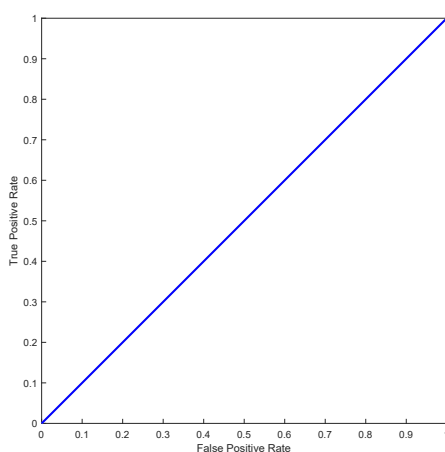
		Ohodnoceno expertem	
		REM	WAKE
Odhadnutá hodnota	REM	T_REM	F_REM
	WAKE	F_WAKE	T_WAKE

Tabulka 1.1. Obecný tvar matice záměn

■ 1.5.4 ROC prostor a ROC křivka

ROC prostor je dvoudimenzionální graf. ROC prostor je zobrazen na obrázku 1.6, kde

- T_REM_RATE se zobrazí na ose Y;
- F_REM_RATE na ose X;
- diagonála Y=X reprezentuje strategii náhodného klasifikátoru.



Obrázek 1.6. Náhodný klasifikátor v ROC prostoru

ROC křivka (angl. Receiver Operating Characteristic Curve) slouží ke grafickému znázornění chování klasifikátoru v závislosti na měnícím se parametru použitým pro binární klasifikaci. Klasifikátor má nejlepší klasifikační schopnost pro parametry, které mají nejmenší vzdálenost k bodu [0,1] ROC prostoru.

Kapitola 2

Popis řešení zadaného úkolu

Tato kapitola obsahuje popis praktické části závěrečné práce. Schematické znázornění kroků, které byly provedeny v rámci implementační části, je představeno v příloze B.

Závěrečná práce je zaměřena na rozlišování mezi dvěma fázemi - **REM** a **WAKE** na základě EOG signálů: EOG1 a EOG2. Úseky signálů, které nepatří k jedné z těchto fází byly ignorovány. Metody, které byly použity, se stanou součástí širšího řešení pro rozpoznání všech spánkových fází.

Kapitola 2.1 popisuje data set, který byl použit pro analýzu EOG signálů zaznamenaných během spánku.

Kapitola 2.2 zahrnuje popis provedené spektrální analýzy. Na základě výsledků spektrální analýzy bylo rozhodnuto použít pro extrakci příznaků metodu CWT.

Kapitola 2.3 popisuje dvouúrovňovou segmentaci signálů a zahrnuje popis algoritmu. Algoritmus v první úrovni najde v signálu 30s úseky spánkových stavů WAKE a REM, dočasně je doplní o další úseky signálu, aby vypočítal koeficienty CWT pro EOG1 a EOG2. V druhém kroku algoritmus rozdělí matice koeficientů na menší segmenty, které dále budou použity pro extrakci příznaků.

Kapitola 2.4 popisuje selekci příznaků pomocí algoritmu SFS. Kapitola obsahuje redukcí dimenze prostoru pomocí PCA provedenou s cílem vizualizace dat v třídímním prostoru.

Kapitola 2.5 poskytuje výsledky provedené k-NN klasifikací. Ukazuje, že algoritmus k=NN poskytuje velkou klasifikační přesnost při klasifikaci WAKE a REM fází na základě zvolených příznaků.

Kapitola 2.6 obsahuje popis programu závěrečné práci.

2.1 Data set

Pro analýzu a klasifikaci spánkových stavů byla použita online databáze celonočních spánkových PSG záznamů The DREAMS Subjects Database [28]. Vybraná databáze obsahuje PSG data 20 zdravých subjektů ve věku od 20 do 65 let. 16 z nich jsou ženy, 4 jsou muži.

Každý záznam obsahuje alespoň dva EOG kanály (P8-A1, P18-A1), alespoň tři EEG kanály (CZ-A1 nebo C3-A1, FP1-A1 a O1-A1) a alespoň jeden z musculus mentalis na bradě EMG. Pro ukládání dat je použit standard European Data Format (EDF).

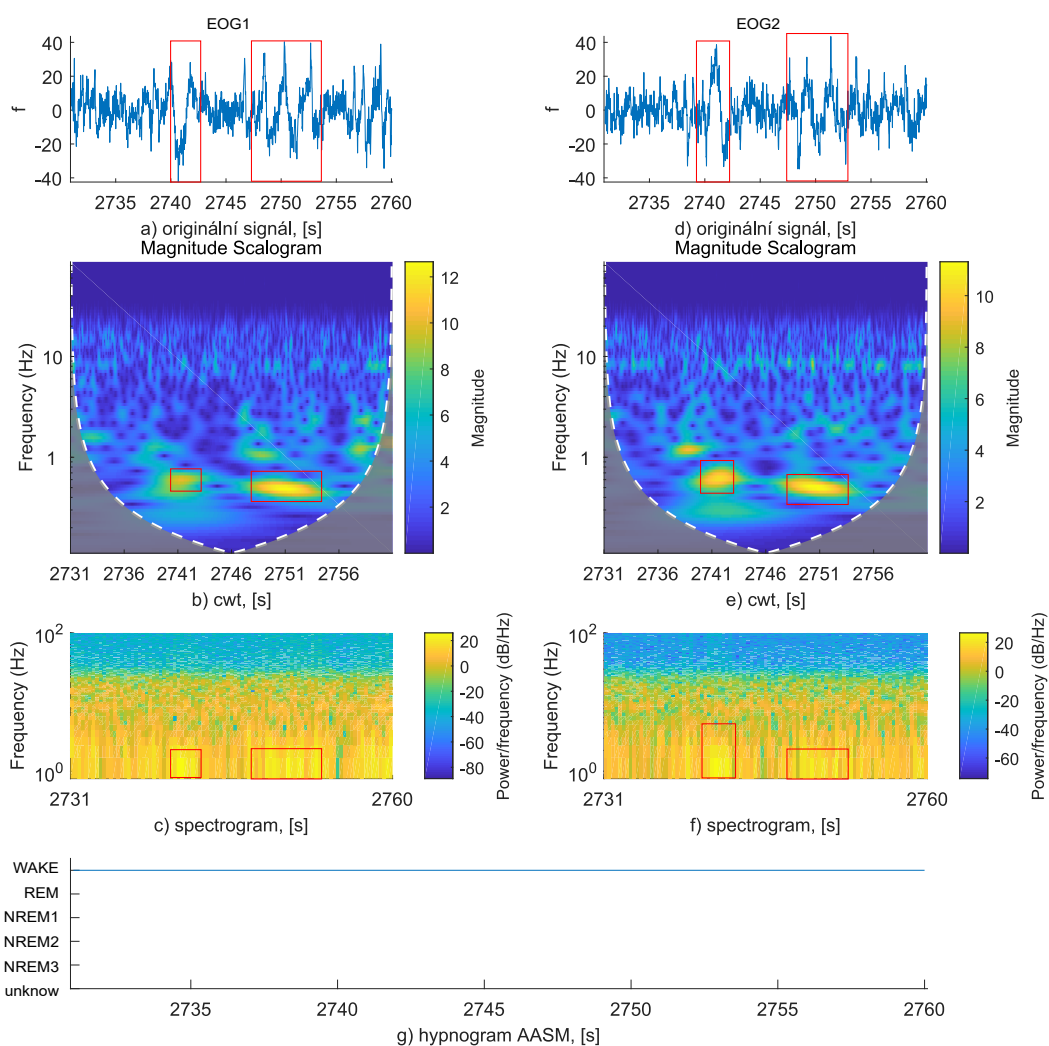
Pro signál EOG, snímány z kanálu P8-A1 dále se používá zkratka **EOG1**, snímány z kanálu P18-A1 - zkratka **EOG2**.

On-line databáze poskytuje také hodnocení spánkových stavů na základě vizuálního hodnocení expertem laboratoře podle obou R&K a AASM metod [28]. Dané ohodnocení experta na základě metody AASM bylo použito pro učení klasifikátoru a pro ověření kvality klasifikací.

2.2 Spektrální analýza

V rámci této práce byla provedena spektrální analýza EOG signálů za účelem výběru metody, která bude použita pro selekci příznaků.

Na obrázcích 2.1 a 2.2 jsou vizualizace, které znázorňují ukázky výstupu provedených analýz. Další příklady jsou uvedeny v příloze C a v příloze D.



Obrázek 2.1. Časově frekvenční analýza WAKE fáze

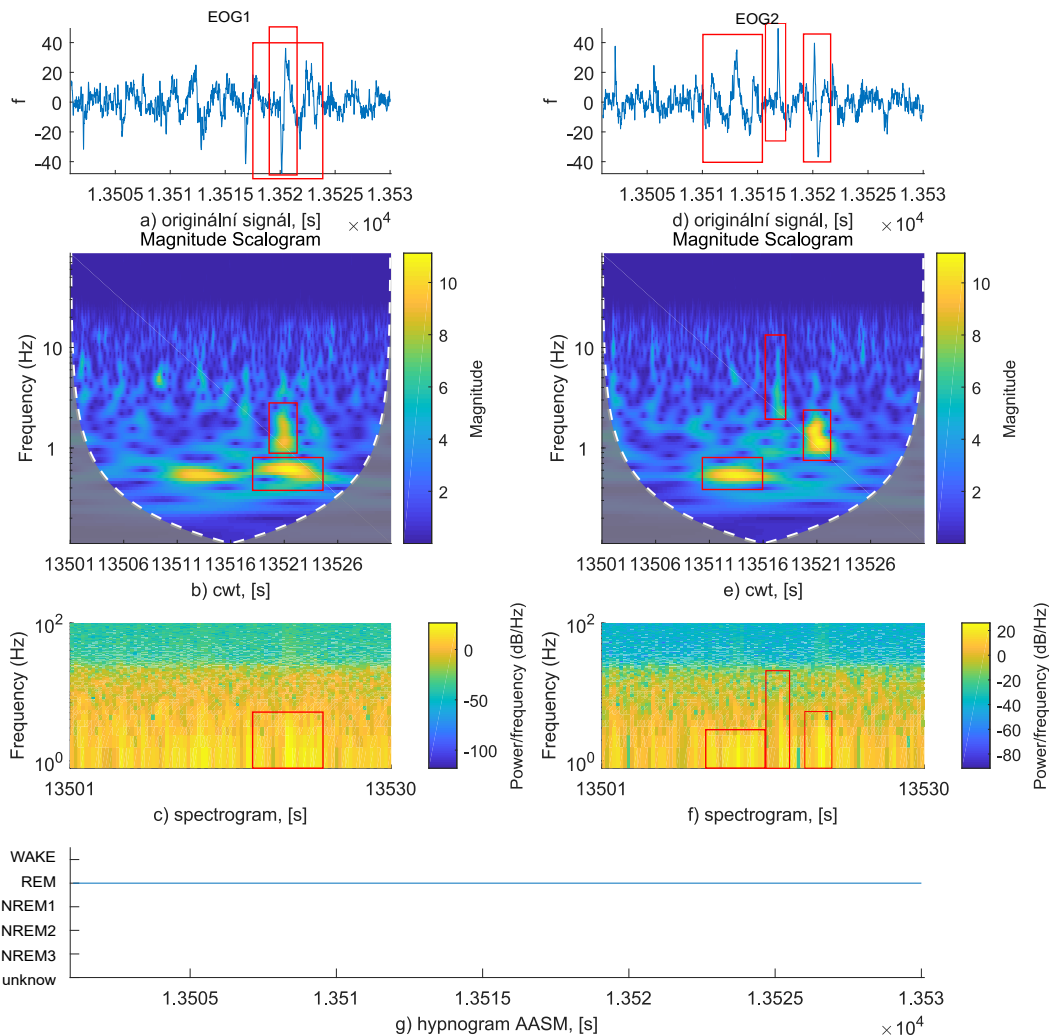
Vizualizace průběhu signálů v časovém okně 30 s je znázorněna na obrázcích 2.1.a), 2.2.a) pro signál EOG1 a 2.1.d), 2.2.d) pro signál EOG2 pro **WAKE** a **REM** zvlášť.

Obrázky 2.1.g) a 2.2.g) odpovídají hypnogramu příslušného časového okna ohodnoceného metodou AASM.

Transformace příslušných úseků signálů byla provedena pomocí Matlab funkcí ze Signal Processing Toolbox:

- `cwt`, která používá spojitou 1-D WT, je vizualizována na obrázcích 2.1.b), 2.2.b) pro EOG1, a na 2.1.e), 2.2.e) pro EOG2;
- `spectrogram`, která používá STFT, je znázorněna na obrázcích 2.1.c), 2.2.c) pro EOG1, a na 2.1.f), 2.2.f) pro EOG2.

Na základě vizuálního vjemu nelze snadno posoudit, které z úseků odpovídají REM fázi a které WAKE fázi.



Obrázek 2.2. Časově frekvenční analýza REM fáze

EOG signály v porovnávaných fázích mohou mít přibližně stejné rozmezí frekvenčních hodnot (například obrázky 2.2.a) a 2.1.a)). Avšak v REM fázi může mít signál větší amplitudu a menší frekvenční hodnoty než ve fázi WAKE. Například obrázky D.11.a) a D.11.d) znázorňují rychlou aktivitu v malém frekvenčním pásmu pro REM fázi. Pro WAKE fázi tak malé pásmo není charakteristické (například příloha D).

Veličina vlnkových koeficientů vrácených `cwt` funkcí je barevně kódována. Teplé barvy mají větší hodnoty magnitudy než studené barvy. Z uvedených příkladů 2.2.b) a 2.1.e)

je vidět, že rozmezí hodnot koeficientů magnitudy pro zkoumané spánkové fáze WAKE a REM mohou být ve stejném rozsahu.

Matlab metoda `spectrogram` rozdělí signál na krátké úseky, na kterých se provádí STFT pro nalezení frekvenčního spektra každého úseku. Nalezená spektra jsou rozložena vedle sebe a konvertována do podoby obrázku, na kterém jsou nízké amplitudy vizualizovány chladnými barvami a vysoké amplitudy teplými barvami.

Vizualizace úseku signálu EOG1 v časové oblasti ukazuje, že signál má dvě odlišné oblasti zvýrazněné na obrázku 2.1.a). Podobná aktivita je přítomna i na příslušném časovém úseku EOG2. Analýza dané oblasti metodou `spectrogram` ukazuje, že dané aktivity jsou téměř neodlišitelné. Ze spektrogramu lze určit, že změna proběhla v nízkých frekvenčních hodnotách, ale přesnou lokalitu je těžké určit. Analýza výstupu Matlab metody `cwt` ukazuje, že obě aktivity se nacházejí ve stejném frekvenčním pásmu a že mají podobné hodnoty magnitudy.

Na obrázku 2.2.b) jsou zvýrazněny dvě aktivity, které metoda `cwt` detekovala při aplikaci transformace v různých frekvenčních pásmech, viz obrázek 1.4. Hodnoty magnitud jsou pro dané oblasti blízké maximu. Na výstupu metody `spectrogram` ale není možné rozpoznat přítomnost těchto dvou událostí.

Na obrázku 2.2.d) jsou zvýrazněny úseky, které detekovala metoda `cwt` v různých frekvenčních pásmech. Na spektrogramu nejsou až tak zřejmé.

Na základě provedené spektrální analýzy je vidět, že matice koeficientů Matlab metody `cwt` má řadu výhod oproti matici koeficientů Matlab metody `spectrogram`:

- lepší lokalizace významných oblastí;
- dokáže rozpoznat změny pro různá frekvenční pásma;
- vhodnější pro nestacionární a aperiodický signál, jímž je EOG.

Po provedení spektrální analýzy bylo rozhodnuto extrahovat příznaky pouze za využití koeficientů CWT.

2.3 Transformace signálů

Pro účely klasifikace původní signál je rozdělen na segmenty konstantní délky. Dané segmenty jsou transformovány pomocí CWT. Z koeficientů transformace jsou extrahovány příznaky.

Tento proces je realizován pomocí dvouúrovňové segmentace s transformací.

2.3.1 Segmentace signálů

Po dohodě s vedoucím práce nebyla provedena adaptivní segmentace EOG signálu, která byla definována v zadání bakalářské práce, ale [dvouúrovňová segmentace do úseků konstantní délky](#):

- 1. úroveň - rozdělení signálu na segmenty konstantní délky 30 s. Výpočet matice CWT koeficientů pro vzniklé segmenty.

- 2. úroveň - rozdělení matice koeficientů z 1. úrovně segmentace na segmenty o délkách 5 s, 10 s, 15 s, 30 s. Výpočet matice příznaků pro každý uvedený segment.

Provedení dvouúrovňové segmentace připraví vzorky pro klasifikaci.

■ 2.3.2 První úroveň dvouúrovňové segmentace signálů

První úroveň dvouúrovňové segmentace je provedena za účelem extrahování ze signálu 30s segmentů, které odpovídají jednomu z hledaných spánkových stavů: WAKE nebo REM.

Pro dané segmenty jsou vypočítány koeficienty CWT pomocí Matlab funkce `cwt`.

Matlab funkce `cwt` pro doplnění matice koeficientů aproximuje signál v obou směrech pro nízká frekvenční pásma. Danou chybu je vidět na obrázku 2.2.b), na kterém je označena šedou barvou. V rámci oblasti „kuželu vlivu“ jsou odhady koeficientů spolehlivé. Mimo oblast „kuželu vlivu“ jsou koeficienty dopočítávány z aproximovaného signálu.

Daný signál nelze spolehlivě aproximovat kvůli náhlým frekvenčním změnám. Aproximace by mohla způsobit velkou chybu klasifikace.

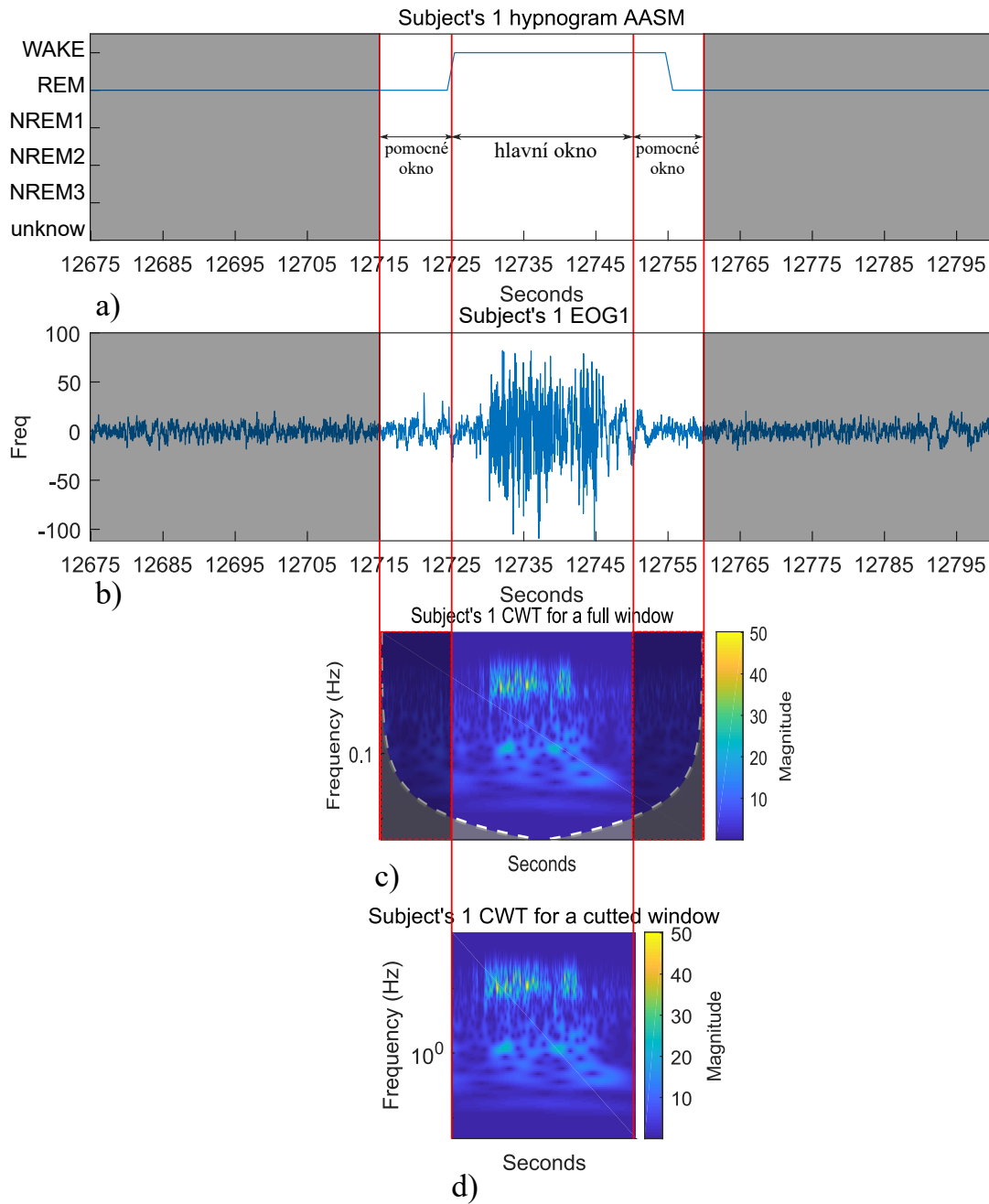
```

1. Nastavit spánkový stav:
   S = REM / WAKE
2. Nastavit délku hlavního okna:
   WM = 30s
3. Nastavit délku pomocného okna:
   WH = 10s
4. Nastavit index na délku pomocného okna:
   idx = WH

5. WHILE není konec signálu
5.1. V hypnogramu najít další úsek o délce WM, v němž
    se vyskytuje pouze stav S:
   h_I = [idx ; idx+WM-1]
5.2. Rozšířit interval o délku pomocného okna:
   h_I' = [idx-WH ; idx+WM-1+WH]
5.3. FOR EACH EOG signál e
5.3.1. Ze signálu e extrahovat segment odpovídající
    intervalu h_I'
   e_I' = e(h_I'(1) : h_I'(2) - h_I'(1) + 1)
5.3.2. Vypočítat CWT pro e_I'
5.3.3. Ze získané matice odstranit submatice koeficientů
    o délce pomocných oken z obou stran
5.3.4. Uložit segment pro následující zpracování
   END FOR
5.4. Posunout index o délku hlavního okna:
   idx += WM
   END WHILE

```

Výpis 2.1. Algoritmus výpočtu matice CWT koeficientů pro 1. úroveň segmentace



Obrázek 2.3. Algoritmus provedení 1. úrovně dvouúrovňové segmentace: a) označení hlavního a pomocných oken v hypnogramu, b) rozšířený segment EOG signálu, c) matice CWT koeficientů rozšířeného segmentu, d) výsledný segment

Za účelem zmenšení této chyby byla provedena CWT nad delšími segmenty, na které byl aplikován algoritmus představený ve výpisu 2.1

Koeficienty **hlavního okna** byly dále zpracovány v 2. úrovni dvouúrovňové segmentace signálů. Nepřesnosti okrajových koeficientů vzniklé po aplikování algoritmu jsou zanedbatelné.

Koeficienty **pomocného okna** byly dále ignorovány.

Počet nalezených hlavních oken pro délku hlavního okna **30 s** je uveden v tabulce 2.1.

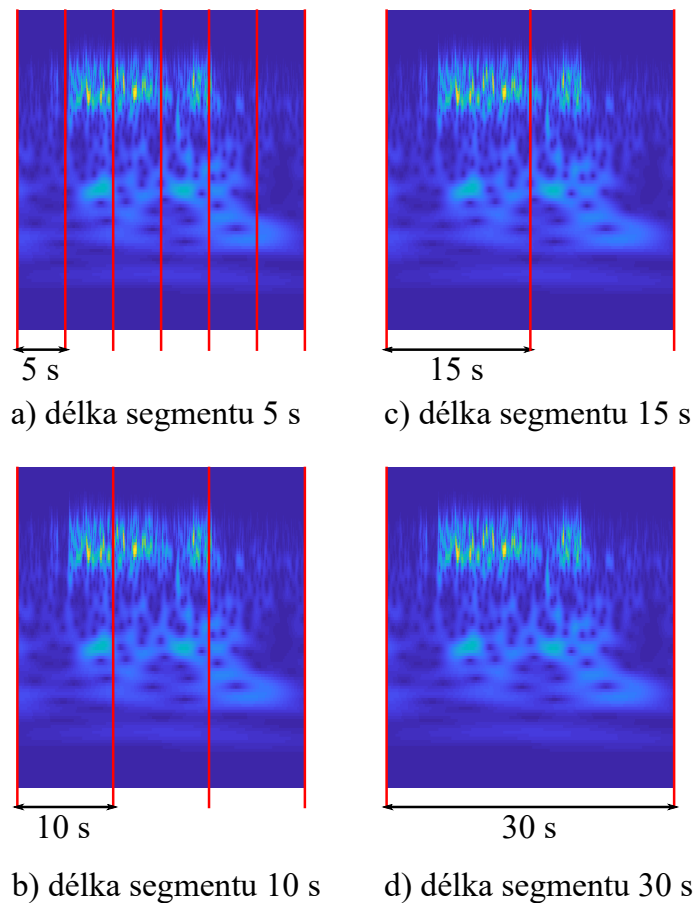
Subjekt	REM	WAKE	Subjekt	REM	WAKE
1	113	229	11	212	120
2	122	146	12	87	391
3	212	121	13	89	179
4	155	91	14	163	206
5	196	83	15	123	112
6	187	177	16	147	257
7	131	392	17	162	65
8	162	179	18	166	167
9	131	214	19	162	128
10	146	70	20	153	196
			Celkem:	3019	3523

Tabulka 2.1. Počet 30s oken pro jednotlivé subjekty

Každá matice CWT koeficientů v 1. úrovni má velikost 105x6000, kde

- 105 je počet frekvenčních pásem, ve kterých byla provedena CWT;
- 6000 je počet vzorků závislý na frekvenci, při které byl signál nahrán (200Hz), a délce hlavního okna (30 s).

Vizualizace první úrovně segmentace je znázorněna na obrázku 2.3.



Obrázek 2.4. Provedení 2. úrovně dvouúrovňové segmentace: a) segment 5 s, b) segment 10 s, c) segment 15 s, d) segment 30 s.

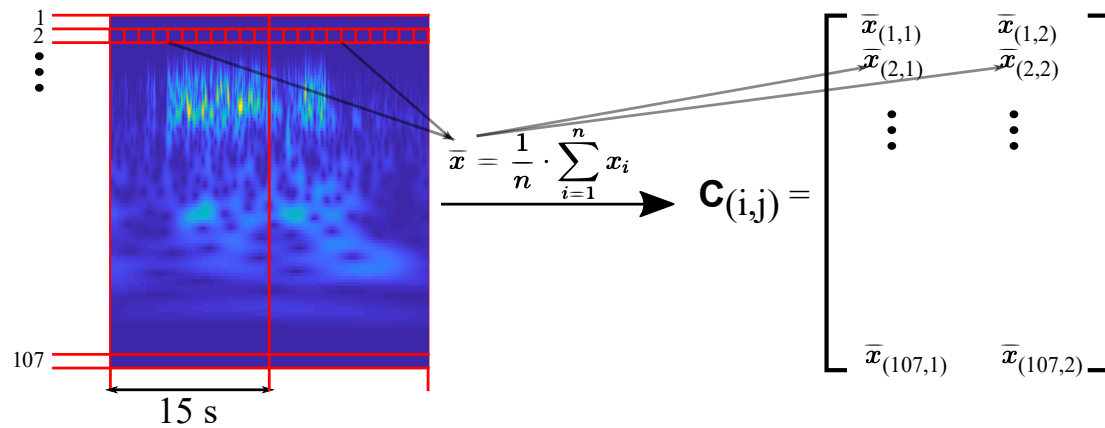
2.3.3 Druhá úroveň dvouúrovňové segmentace signálů

Druhá úroveň dvouúrovňové segmentace je provedena za účelem přípravy sady příznaků, na základě kterých se dále bude provádět klasifikace.

Segmenty z předchozího kroku byly postupně rozděleny do menších segmentů o délce 5 s, 10 s, 15 s a 30 s. Rozdělení na segmenty je znázorněno na obrázku 2.4 pro segmenty různé délky.

V každém frekvenčním pásmu byla vypočtena hodnota příznaku (angl. feature value) jako průměr koeficientů pro daný segment. Schematické znázornění provedeného výpočtu je zachyceno na obrázku 2.5, na kterém je naznačen výpočet pro 15s segment. Submatice $\mathbf{C}_{(i,j)}$ má velikost 105x2.

Pro segmenty jiné délky je výpočet analogický. Pro 5s segment má submatice $\mathbf{C}_{(i,j)}$ rozměry 105x6, pro 10s segment - 105x3 a pro 30s segment - 105x1.



Obrázek 2.5. Výpočet hodnoty příznaku na příkladě 15s segmentu

2.4 Selektce příznaků

Pro výběr nejvýznamnějších příznaků byl aplikován algoritmus **SFS** popsáný v kapitole 1.4. Matice \mathbf{M}_l , kde $l \in \{5, 10, 15, 30\}$ je délka segmentu. \mathbf{M}_l se skládá ze všech příznaků a byla sestavena ze submatic $\mathbf{C}_{(i,j)}$, kde $i \in \{1, \dots, P\}$ odpovídá indexu hlavního okna, $j \in \{1, 2\}$ odpovídá EOG1 nebo EOG2.

Schematické znázornění matice \mathbf{M}_l je uvedeno na obrázku 2.6.

$$\mathbf{M}_l = \begin{bmatrix} \mathbf{C}_{(1,1)} & \mathbf{C}_{(2,1)} & \dots & \mathbf{C}_{(P,1)} \\ \mathbf{C}_{(1,2)} & \mathbf{C}_{(2,2)} & \dots & \mathbf{C}_{(P,2)} \end{bmatrix}$$

Obrázek 2.6. Matice příznaků sestavená ze submatic

sfs funkce Matlab byla aplikována na matici \mathbf{M}_l pro každé l zvlášť. Metoda vždy vybírala ze seznamu 210 příznaků.

Porovnání výsledků SFS pro segmenty různých délek je znázorněno v tabulce 2.2. Detailní přehled SFS je uveden v příloze E.

Délka segmentu	Počet nejlepších příznaků	Dosažená nejmenší chyba	Doba trvání SFS [s]
5 s	29	0,05450	39364
10 s	22	0,04210	9656
15 s	12	0,03965	1586
30 s	15	0,03172	917

Tabulka 2.2. SFS výsledky pro segmenty různé délky

Nejlepší příznaky pro každou délku segmentu formulují množinu vybraných příznaků F . Kardinalita $|F| = w$, kde w se rovná počtu nalezených příznaků pro danou délku.

Jednotlivé segmenty při klasifikaci je možno chápat jako body v příslušném Z dimenzionálním prostoru. Dimenze prostoru také odpovídá kardinalitě množiny příznaků $Z = |F|$. Pro 5s segment dimenze prostoru $Z = |F| = 29$.

2.4.1 Vizualizace příznakového prostoru

Pro vizualizaci vypočtených příznaků byla použita transformace PCA. Algoritmus PCA byl aplikován na všechny příznaky pro všechny segmenty dat patřících jednomu subjektu. Teoretické základy metody jsou uvedeny v kapitole 1.4.2.

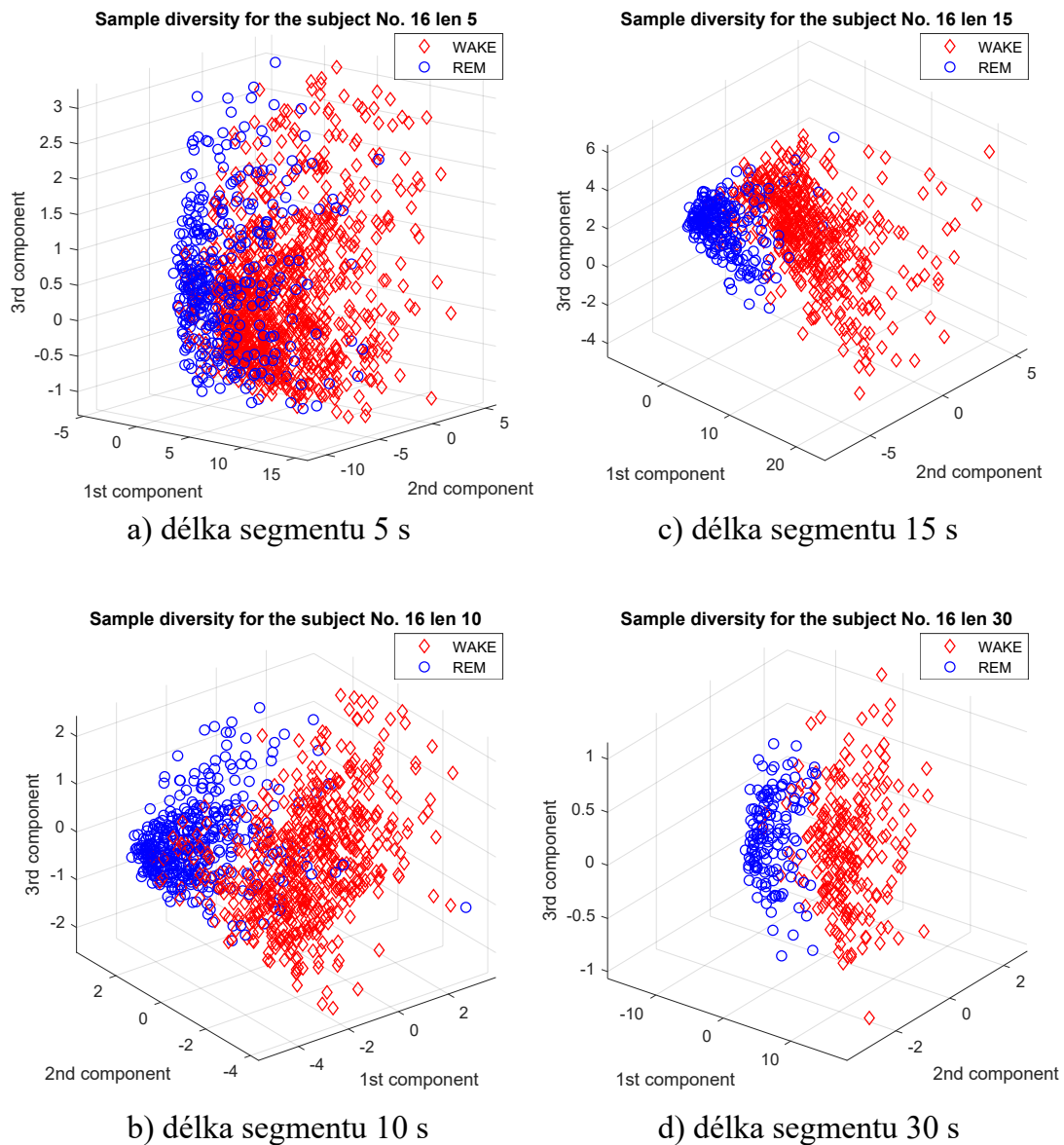
Množina 5s segmentů představuje body v prostoru dimenze $Z = 29$ viz tabulka 2.2. Daný prostor byl proložen podprostorem dimenze $q = 3$ tak, aby součet kvadrátů vzdáleností původních a promítnutých bodů byl minimální. Následně promítnuté body byly zobrazeny v souřadnicích báze nalezeného afinního podprostoru [24]. Body v novém prostoru jsou zobrazeny na obrázku 2.7.a).

Stejný postup byl proveden pro množinu:

- 10s segmentů, které představují body v prostoru dimenze $Z = 22$ viz tabulka 2.2. Body v novém prostoru jsou zobrazeny na obrázku 2.7.b);
- 15s segmentů, které představují body v prostoru dimenze $Z = 12$ viz tabulka 2.2. Body v novém prostoru jsou zobrazeny na obrázku 2.7.c);
- 30s segmentů, které představují body v prostoru dimenze $Z = 15$ viz tabulka 2.2. Body v novém prostoru jsou zobrazeny na obrázku 2.7.d).

Dle vizuálního hodnocení obrázků 2.7 lze předpokládat, že klasifikace metodou k-NN bude na základě delších segmentů poskytovat lepší výsledky. Příznaky pro segmenty délky 5 s se seskupí na vizualizaci podle typu spánkového stavu WAKE nebo REM. Existuje ale velká část prostoru, která sdílí vzorky jak WAKE fáze, tak i REM fáze. Bude tedy těžké klasifikovat vzorky, které se rozmístí do dané části prostoru.

Zvýšení délky segmentů zlepšuje rozdělení spánkových stavů. Obrázek 2.7.d) znázorňuje, že množiny vzorků jsou dobře rozlišitelné a že klasifikace podle nejbližších sousedů by měla poskytnout dobré výsledky.



Obrázek 2.7. Vizualizace nejvýznamnějších příznaků: a) segment 5 s, b) segment 10 s, c) segment 15 s, d) segment 30 s

2.5 k-NN klasifikace

Formulace klasifikační úlohy pro odlišení REM a WAKE spánkových stavů:

- množina tříd: $Y = \{WAKE, REM\}$;
- množina vzorků: $X = \{x_i\}, i \in N$;
- trénovací množina: $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, kde každý prvek z X byl jednoznačně ohodnocen jednou ze tříd Y ;
- metrika pro výpočet vzdálenosti mezi prvky: d - Euklidova vzdálenost.

Pro ohodnocení kvality klasifikátoru z trénovací množiny byla „schována“ data dvou subjektů. Množina vzorků patřící subjektům 19 a 20 byla definována jako **testovací množina**.

Délka segmentu	Trénovací subjekty	Počet REM	Počet WAKE	Testovací subjekty	Počet REM	Počet WAKE
5 s	1-18	16224	19194	19-20	1834	1652
10 s	1-18	8112	9597	19-20	917	826
15 s	1-18	5408	6398	19-20	611	551
30 s	1-18	2704	3199	19-20	306	275

Tabulka 2.3. Obsah data setu pro segmenty různé délky

Celkové počty vzorků pro učení (trénovací množina) a ověření kvality klasifikátoru (testovací množina) jsou uvedeny v tabulce 2.3.

Pro každou délku segmentu byla nalezena **optimální hodnota k** pomocí algoritmu, který používá **LOO křížovou validaci**. Daný algoritmus je představen ve výpisu 2.2. Indexy testovacích subjektů byly vygenerovány pomocí Matlab funkce `crossvalind`.

```

1. Nastavit iterační proměnnou k_set
   k_set = 1:20 // 1 je minimální počet sousedů a 20 je maximální
2. Nastavit velikost trénovací množiny, dle počtu trénovacích
   subjektů:
   M = 18
3. Vygenerovat indexy testovacích subjektů
   za použitím LOO křížové validace:
   tst_idx = crossvalind('Kfold', M, M)
4. Nastavit pole validačních chyb pro k_set:
   err = []
5. FOR EACH k_cur v k_set:
   5.1 Nastavit pole validačních chyb pro k_cur:
       err_k_cur = []
   5.1 FOR EACH tst v tst_idx:
       5.1.1. Nastavit indexy trénovacích subjektů:
           trn = [1:(tst_ind-1) (tst_ind+1):M]
       5.1.2. Naučit klasifikátor na trénovací množině trn
       5.1.3. Vyhodnotit klasifikátor na testovací množině tst
       5.1.4. Uložit chybu klasifikace do err_k_cur
           na příslušný index
       END FOR
   5.2 Uložit průměrnou chybu pole err_k_cur
       err(k_cur) = mean(err_k_cur)
   END FOR
6. Najít optimální k jako index minimálního prvku z pole err

```

Výpis 2.2. Algoritmus nalezení optimální hodnoty k

Nalezené optimální hodnoty k jsou uvedeny v tabulce 2.4.

Výsledky klasifikace pro různě dlouhé segmenty jsou uvedeny v tabulce 2.5.

Pro klasifikační úlohu byly vypočítány ROC metriky. ROC křivka byla proložena ROC prostorem pro každý klasifikátor zvlášť. ROC křivka pro 5s segmenty je znázorněna na obrázku 2.8, ROC křivka pro 10s segment je na obrázku 2.9, ROC křivka pro

Délka segmentu	Optimální počet sousedů k-best
5 s	15
10 s	15
15 s	7
30 s	13

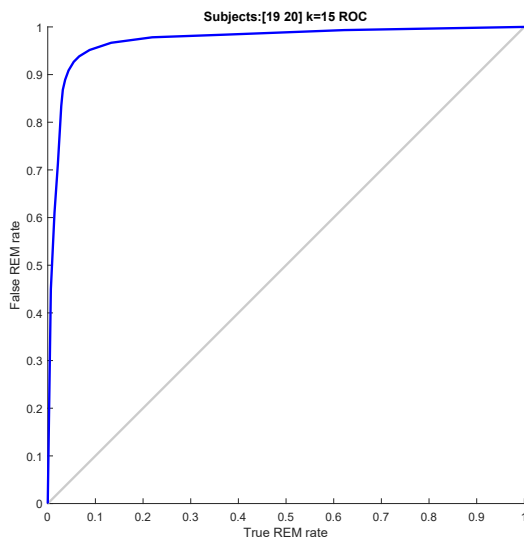
Tabulka 2.4. Optimální hodnoty parametru k pro k-NN klasifikátor

Hodnota parametru	Okno 5 s	Okno 10 s	Okno 15 s	Okno 30 s
sensitivity	0,95638	0,95965	0,91489	0,96078
specificity	0,90860	0,93099	0,91470	0,93091
accuracy	0,93373	0,94607	0,91480	0,94664
REM predictive value	0,92073	0,93917	0,92244	0,93930
WAKE predictive value	0,94940	0,95409	0,90647	0,95522

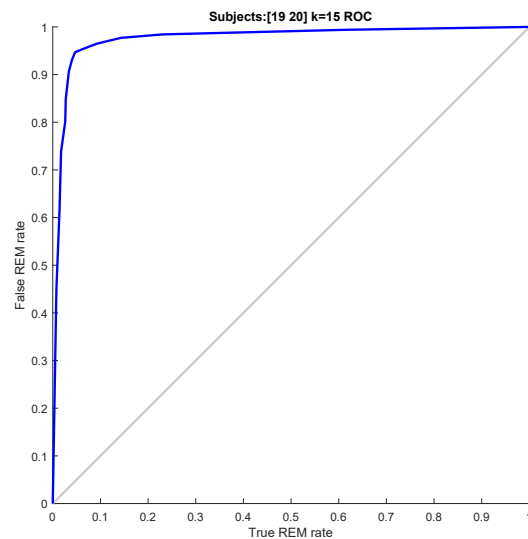
Tabulka 2.5. Hodnocení k-NN klasifikátoru pro testovací subjekty 19, 20

15s segment je na obrázku 2.10 a pro 30s segment je na obrázku 2.11. Matice záměn jsou uvedeny v příloze F:

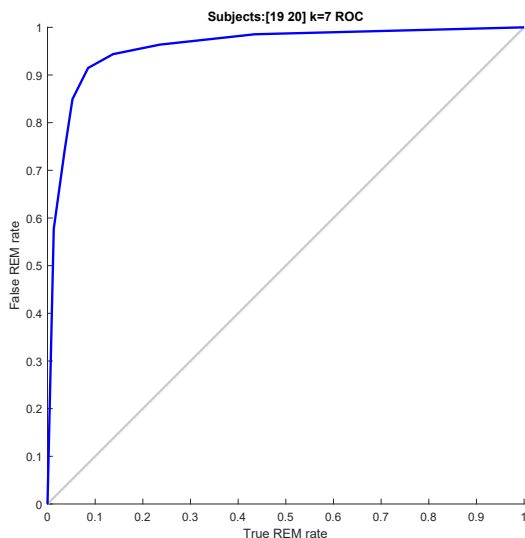
- pro 5s segmenty je uvedena v tabulce F.1;
- pro 10s segmenty je uvedena v tabulce F.2;
- pro 15s segmenty je uvedena v tabulce F.3;
- pro 30s segmenty je uvedena v tabulce F.4.



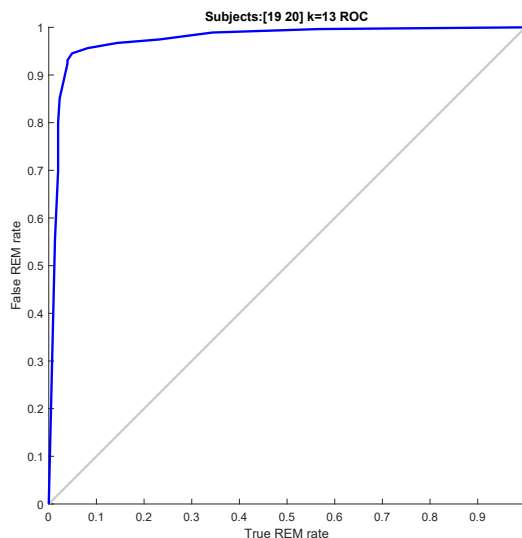
Obrázek 2.8. ROC křivka pro klasifikaci vzorků o délce 5 s



Obrázek 2.9. ROC křivka pro klasifikaci vzorků o délce 10 s



Obrázek 2.10. ROC křivka pro klasifikaci vzorků o délce 15 s



Obrázek 2.11. ROC křivka pro klasifikaci vzorků o délce 30 s

2.6 Popis programu

Program pro analýzu EOG a klasifikaci segmentů signálů do dvou spánkových stavů byl implementován v prostředí Matlab R2018a.

Kompletní program je k dispozici na přiloženém disku. Adresářová struktura disku je představena v příloze G.

Na disku je také k dispozici soubor `thesis.pdf`, který představuje kopii dané závěrečné práce.

Program je umístěn do složky `program` a obsahuje tři Matlab skripty:

- `program/main_time_frequency_analysis.m` zahrnuje spektrální analýzu, popsanou v kapitole 2.2. Skript ukládá výsledky analýzy na disk do složky `program/data/internal/time_frequency_analysis`.
- `program/main_train.m` připravuje data pro klasifikátor k-NN. Výsledky každého kroku jsou ukládány na disk do složky `program/data/internal/matlab`. Skript provádí transformaci signálů (viz kapitola 2.3), selekci příznaků (viz kapitola 2.4), přípravu trénovacích a testovacích podmnožin (viz kapitola 2.5) a nalezení optimální hodnoty k pro k-NN klasifikátor (viz kapitola 2.5).
- `program/main_test.m` klasifikuje segmenty 5 s, 10 s, 15 s nebo 30 s pomocí k-NN klasifikátoru. Skript používá data ze složky `program/data/internal/matlab`, která byla uložena pomocí skriptu `program/main_train.m`.

Všechny funkce, které jsou potřebné pro správný běh programu, jsou k nalezení ve složce `program/functions`. Daná složka obsahuje implementované metody pro závěrečnou práci a také složku `program/functions/3rd_party`, která obsahuje metody implementované třetími stranami.

Program je funkční bez dat subjektů (popsaných v kapitole 2.1), na kterých byl naučen klasifikátor. Složka `program/data/db` obsahuje připravenou adresářovou strukturu, se soubory `../README.txt`, které popisují případné přidání dat.

Kapitola 3

Závěr

V bakalářské práci byly zanalyzovány celonoční EOG záznamy, které byly měřeny během spánku. Nad EOG signály byla provedena spektrální analýza. Ze dvou analyzovaných metod pro extrakci příznaků (STFT a CWT) byla zvolena CWT jako vhodnější pro nestacionární signál.

Signál EOG byl rozdělen na segmenty konstantní délky pomocí dvouúrovňové segmentace. Tento přístup se ukázal jako vhodný pro účely této práce a umožnil provést sérii nezávislých klasifikací pro segmenty o různých délkách (5 s, 10 s, 15 s, 30 s).

Segmenty signálů získané z 1. úrovně segmentace byly transformovány pomocí CWT. Pro každý segment 2. úrovně segmentace (5 s, 10 s, 15 s, 30 s) byla vypočítána sada popisujících příznaků jako průměrná hodnota koeficientů CWT. Pomocí algoritmu SFS byly zvoleny pro každý segment zvláště nejvýznamnější příznaky pro následující klasifikaci.

Klasifikace segmentů do dvou spánkových stavů (WAKE a REM) byla provedena pomocí k-NN algoritmu. Pro každý z definovaných segmentů měl implementovaný klasifikátor na testovací sadě průměrnou přesnost klasifikace 93,53 %. Klasifikátor pro segment o délce 30 s poskytuje nejpřesnější výsledek klasifikace 94,66 %.

Přesnost klasifikace by mohla být zvýšena, pokud by bylo použito více přesnějších příznaků. Bylo by například možné zahrnout do klasifikace další příznaky, které jsou extrahovány z jiných PSG signálů. Přesnost klasifikace by dále mohla být zlepšena například pomocí odstranění nežádoucích artefaktů. To by mohlo být provedeno například pomocí sady vhodných filtrů.



Literatura

- [1] Josef Faber. *Elektroencefalografie a psychofyziologie*. Praha: ISV, 2001. ISBN 80-85615-91-6.
- [2] Eugene Aserinsky a Nathaniel Kleitman. Regularly Occurring Periods of Eye Motility, and Concomitant Phenomena, during Sleep. *Science, New Series*. 1953, 118 (3062), 273-274.
- [3] Roebuck A., Monasterio V., Geder E. a others. A review of signals used in sleep analysis. *Physiological Measurement*. 2014, 35 (1),
- [4] Berry RB, Brooks R, Gamaldo CE a et al. for the American Academy of Sleep Medicine. *The AASM Manual for the Scoring of Sleep and Associated Events RULES, TERMINOLOGY AND TECHNICAL SPECIFICATIONS. VERSION 2.4*. 2017.
- [5] Sudhansu Chokroverty, Meeta Bhatt a Robert J. Thomas. *Atlas of Sleep Medicine*. : Elsevier Inc., 2005. ISBN .
- [6] Thomas Eggert. Eye Movement Recordings: Methods. *Neuro-Ophthalmology*. 2007, 40 15–34.
- [7] Hema L. Ramkumar. *Electrooculogram*. 2017. EyeWiki, a service of the American Academy of Ophthalmology, http://eyewiki.aaopt.org/Electrooculogram#Principle_of_EOG.
- [8] Crawford C. Sleep recording in the home with automatic analysis of results. *Eur Neurol*. 1986, 25 30-35.
- [9] Mohzlova Jitka Krajča Vladimír. *Číslíkové zpracování neurofyziologických signálů*. : ČVUT v Praze, 2011. ISBN .
- [10] Nicolae Cristian PAMPU. *Study of effects of the Short time Fourier transform configuration on EEG spectral estimates*. 2011.
- [11] Ugur Halici Dilan Gorur a others. *Sleep spindles detection using Short time Fourier transform and neural networks*. 2002.
- [12] Henry C. E. Knott J. R., Gibbs F. A. Fourier transforms of the electroencephalogram during sleep. *Journal of Experimental Psychology*. 1942, 31 (6), 465-477.
- [13] Caffo B Zhang L, Samet J a others. Power spectral analysis of EEG activity during sleep in cigarette smokers. *Chest*. 2007, 133 (2), 27-32.
- [14] Yuerong Wang. *Feature Extracting of Weak Signal in Real-Time Sleeping EEG with Approximate Entropy and Bispectrum Analysis*. 2009.

- [15] Shah Ahsanul Haque Md. Riyasat Azim, Md. Shahedul Amin a others. *Feature extraction of human sleep EEG signals using wavelet transform and Fourier transform*. 2009.
- [16] Amjed S. Al-Fahoum a Ausilah A. Al-Fraihat. *Methods of EEG Signal Features Extraction Using Linear Analysis in Frequency and Time-Frequency Domains*. 2014. <https://www.hindawi.com/journals/isrn/2014/730218/>.
- [17] W. Mansor A. Zabidi a others. *Short-time Fourier Transform analysis of EEG signal generated during imagined writing*; . 2012.
- [18] S.S. Hussin a R. Sudirman. EEG Interpretation through Short Time Fourier Transform for Sensory Response Among Children. *Aust. J. Basic and Appl. Sci.* 2014, 8 (5), 417-422.
- [19] Olivier Rioul a Martin Vetterli. Wavelets and signal processing. *IEEE Signal Processing Magazine*. 1991,
- [20] Stéphane Mallat. *A wavelet tour of signal processing*. : Academic Press is an imprint of Elsevier, 2009. ISBN .
- [21] U. Rajendra Acharya Oliver Faust a others. Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis. *Seizure*. 2015, 26 56-64.
- [22] Abdulla S Ghayab HRA, Li Y a others. Classification of epileptic EEG signals based on simple random sampling and sequential feature selection. *Brain Informatics*, 3:1–7. doi: 10.1007/s40708-016-0039-1.. 2016,
- [23] Isabelle Guyon. *Lecture 4, Embedded methods* . <https://slideplayer.com/slide/4646748/>.
- [24] Tomáš Werner. *Elektronická skripta předmětu A4B33OPT*. https://cw.fel.cvut.cz/old/_media/courses/a4b33opt/opt.pdf.
- [25] P.A. Castillo A.M. Mora, L.J. Herrera a others. Sleeping with Ants, SVMs, ANNs and SOMs. *10th International Conference on Intelligent Systems Design and Applications, ISDA*. 2010,
- [26] Michael Love Rafael Irizarry. *Cross-validation*. Harvard School of Public Health PH525x series - Biomedical Data Science <http://genomicsclass.github.io/book/pages/crossvalidation.html>.
- [27] Tom Fawcett. Introduction to ROC analysis. *Pattern Recognition Letters*. 2006, 27 (8), 861-874.
- [28] Devuyt Stéphanie. *The DREAMS Subjects Database*. 2003. <http://www.tcts.fpms.ac.be/~devuyt/Databases/DatabaseSubjects/>.

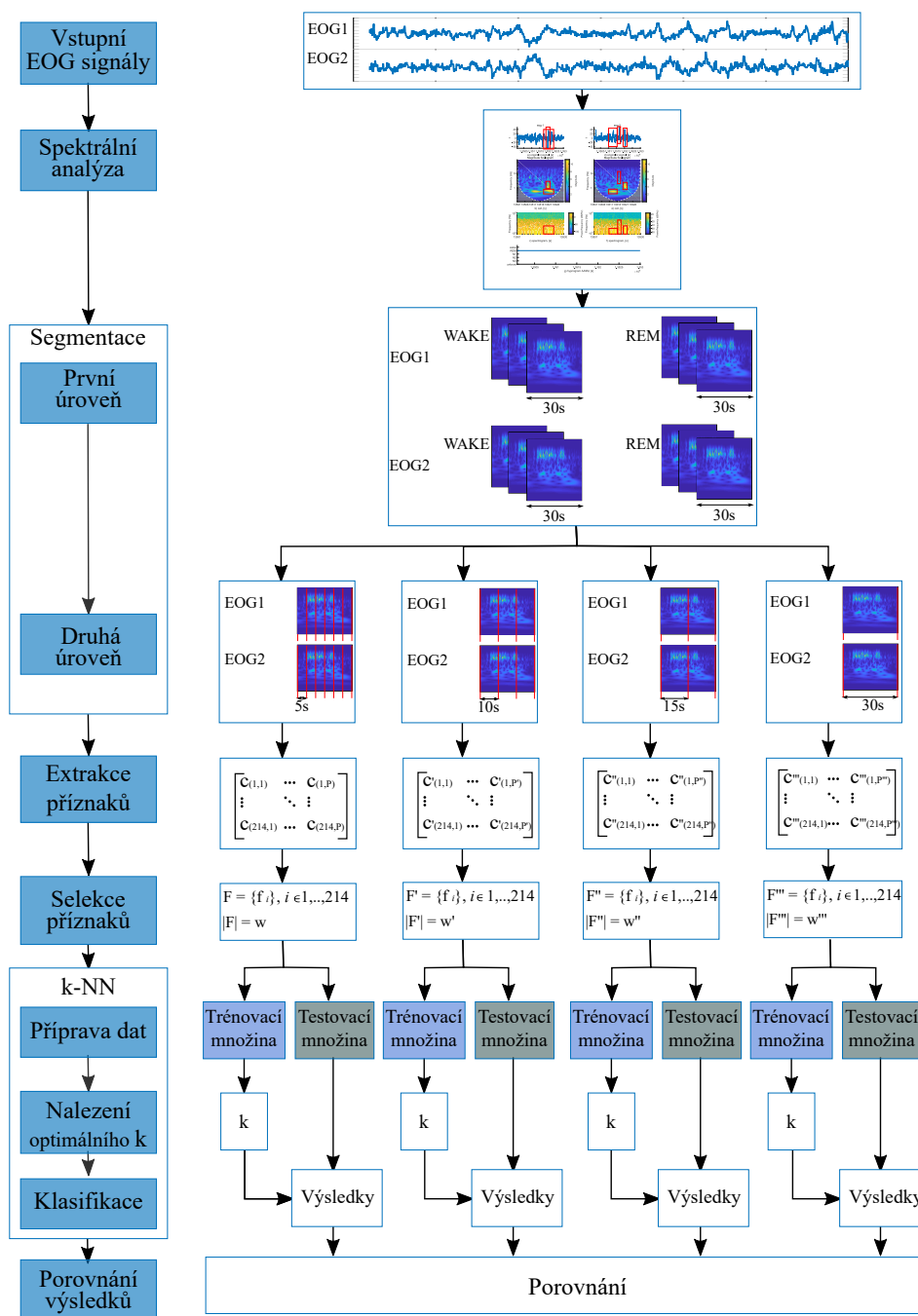
Příloha A

Zkratky

AASM	■ American Academy of Sleep Medicine, také metodologie hodnocení spánkových stavů
CWT	■ Continuous Wavelet Transform
DWT	■ Discrete Wavelet Transform
EEG	■ Elektroencefalografie
EMG	■ Electromyografie
EOG	■ Elektrookulografie
EOG1	■ EOG signál, snímány z kanálu P8-A1
EOG2	■ EOG signál, snímány z kanálu P18-A1
FFT	■ Fast Fourier Transform
FS	■ Feature Selection
FT	■ Fourier Transform
k-NN	■ k-Nearest Neighbours algoritmus
LOO	■ Leave-One-Out
NREM	■ Non Rapid Eye Movement
NREM1	■ 1. fáze NREM spánku
NREM2	■ 2. fáze NREM spánku
NREM3	■ 3. fáze NREM spánku
NREM4	■ 4. fáze NREM spánku
PCA	■ Principal Component Analysis
PSG	■ Polysomnografie
REM	■ Rapid Eye Movement fáze spánku
ROC	■ Receiver Operating Characteristic
R&K	■ Rechtschaffen a Kalese
SFS	■ Sequential Forward Selection
STFT	■ Short-time Fourier Transform
WAKE	■ Fáze bdělosti
WT	■ Wavelet Transform

Příloha B

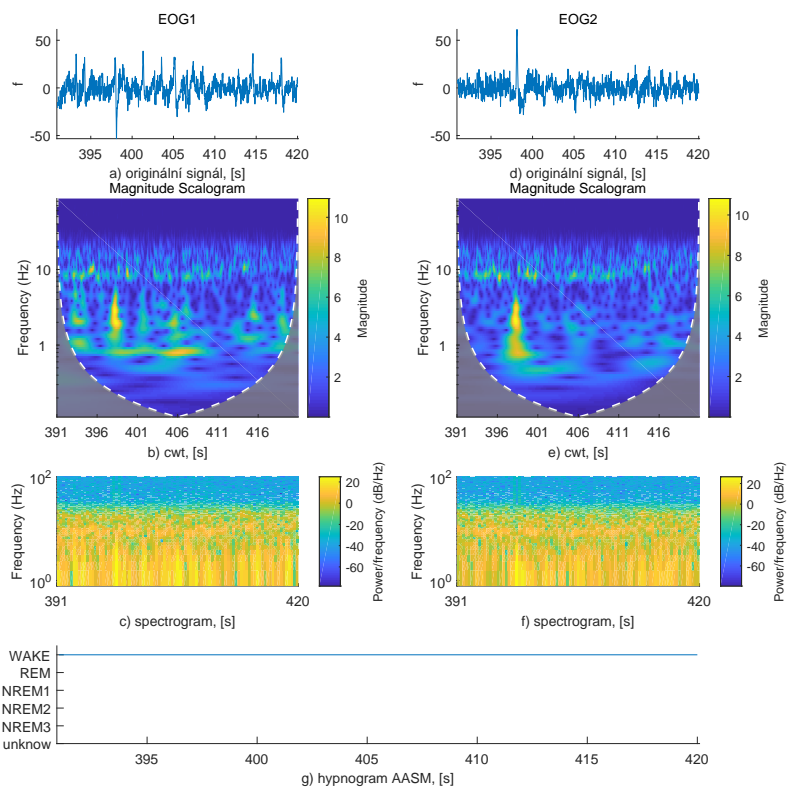
Schematické znázornění implementační části závěrečné práce



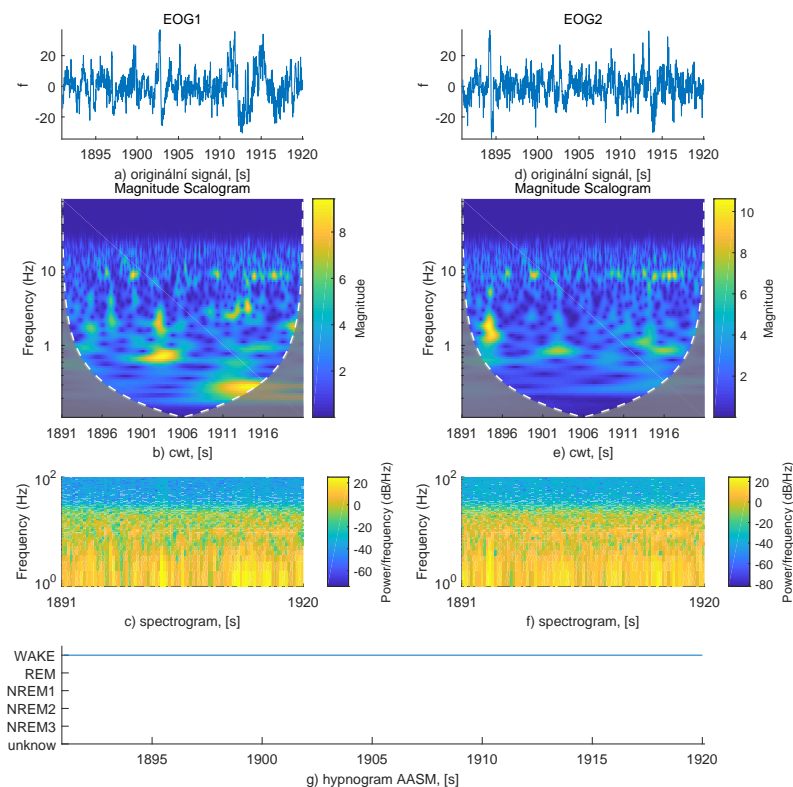
Obrázek B.1. Schematické znázornění implementační části závěrečné práce

Příloha C

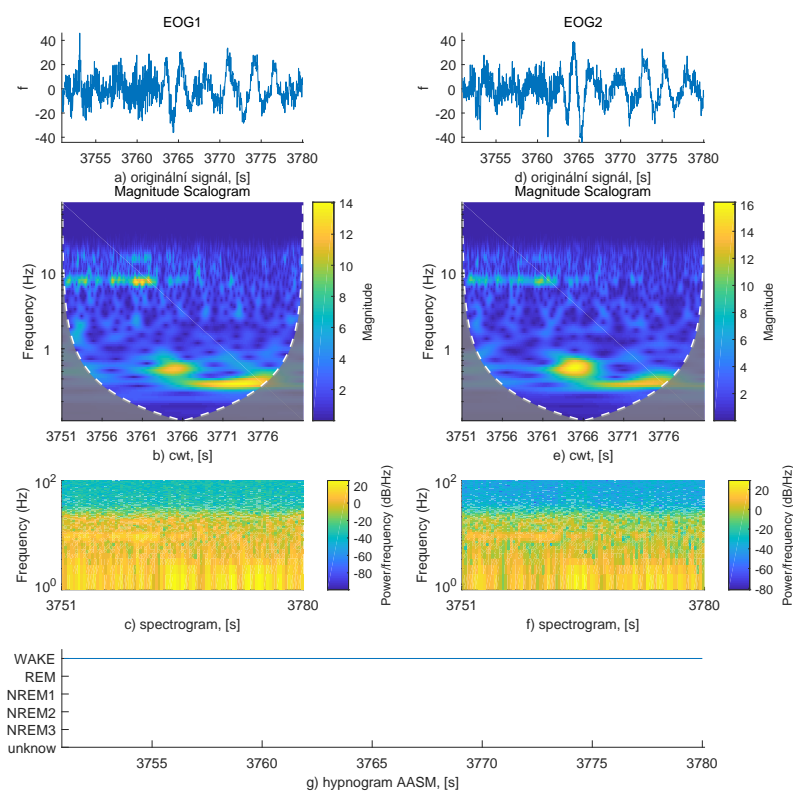
Časově frekvenční analýza WAKE fáze



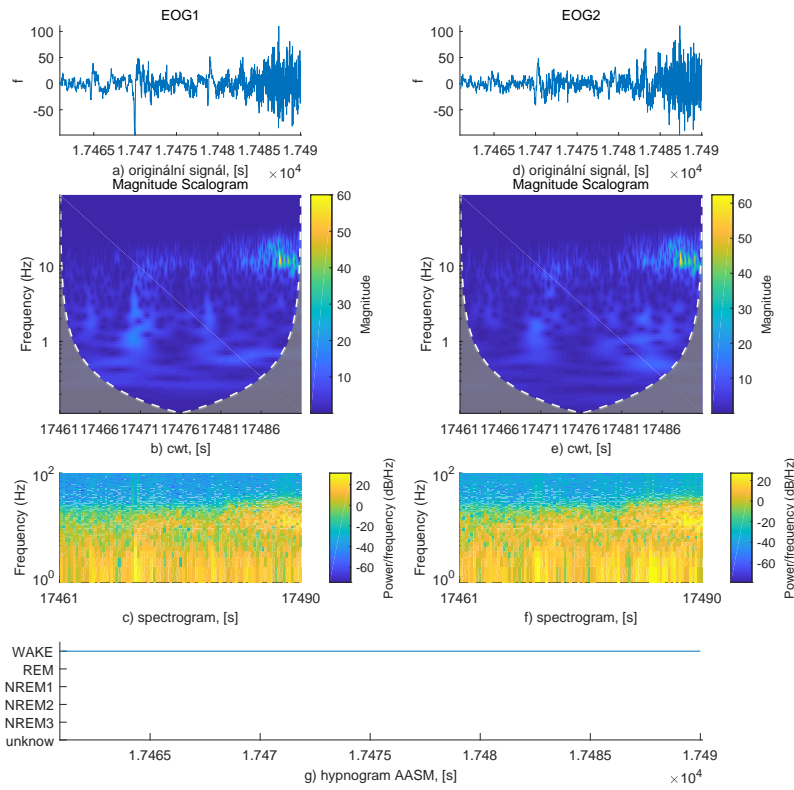
Obrázek C.2. Subjekt č.1, fáze spánku WAKE, délka segmentu 30 s



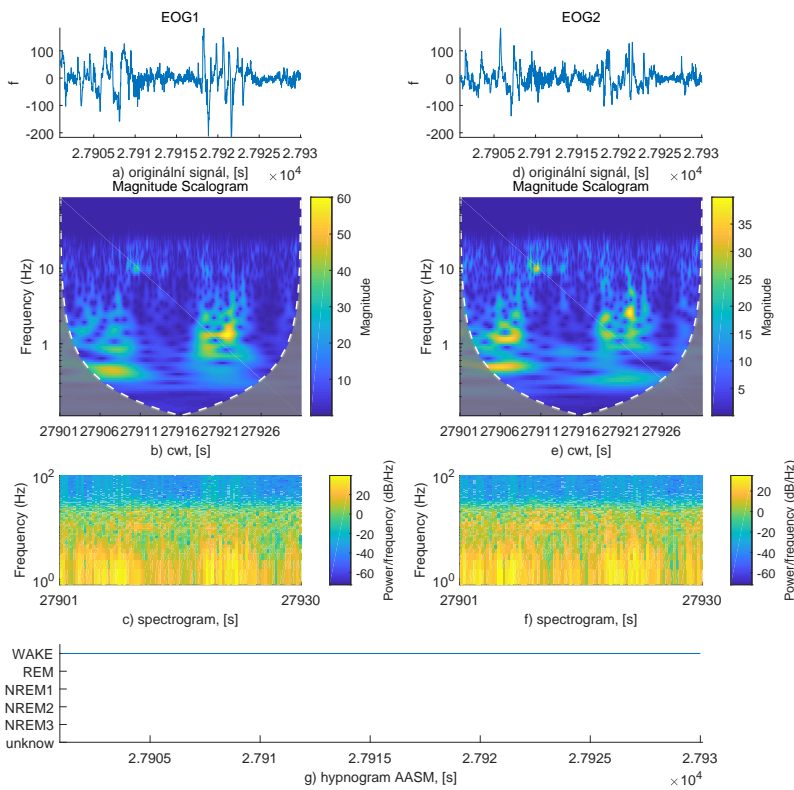
Obrázek C.3. Subjekt č.1, fáze spánku WAKE, délka segmentu 30 s



Obrázek C.4. Subjekt č.1, fáze spánku WAKE, délka segmentu 30 s



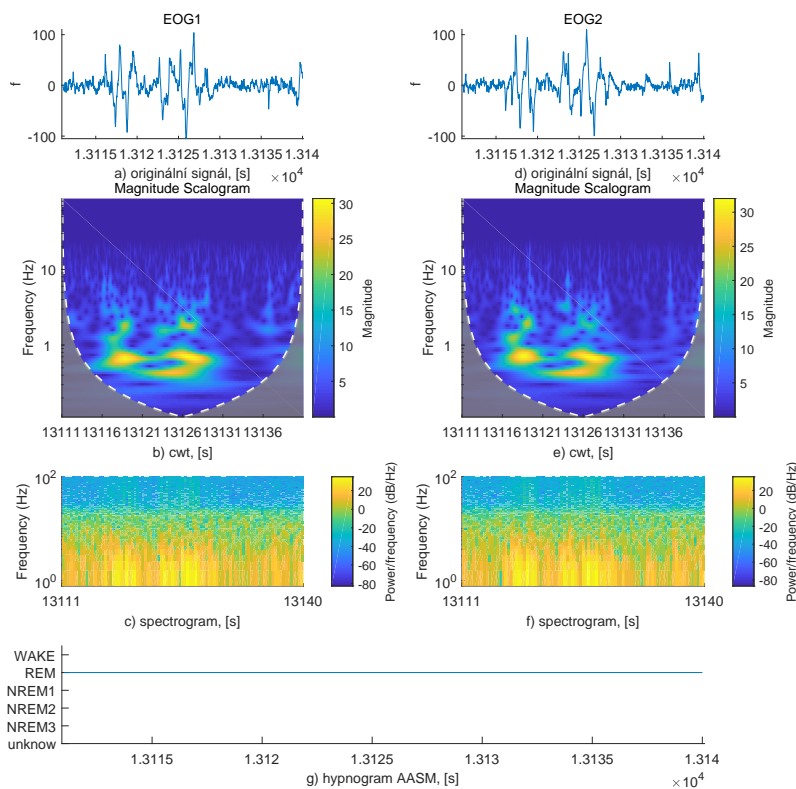
Obrázek C.5. Subjekt č.1, fáze spánku WAKE, délka segmentu 30 s



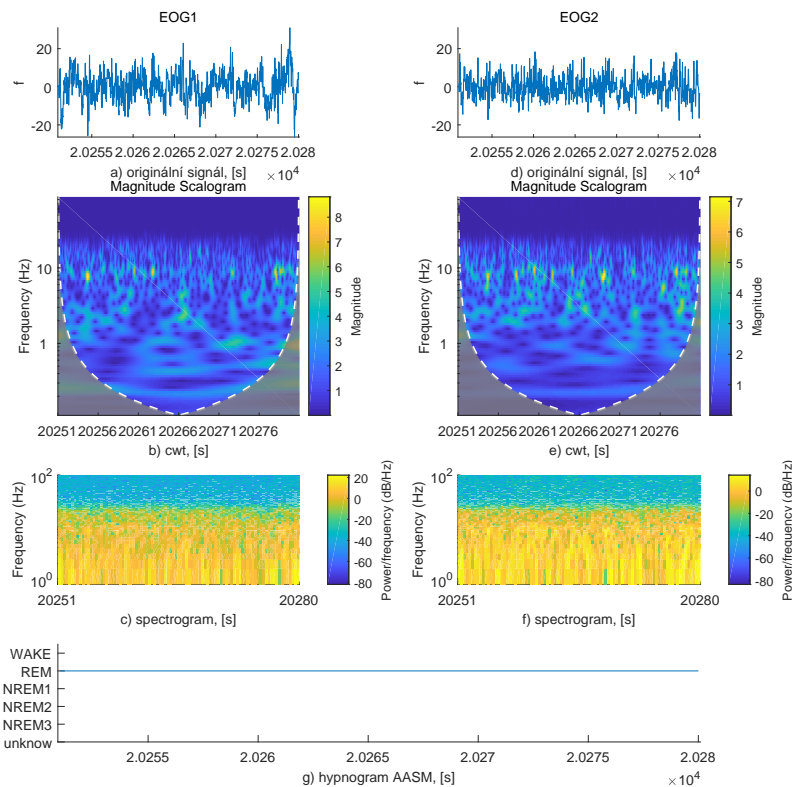
Obrázek C.6. Subjekt č.1, fáze spánku WAKE, délka segmentu 30 s

Příloha D

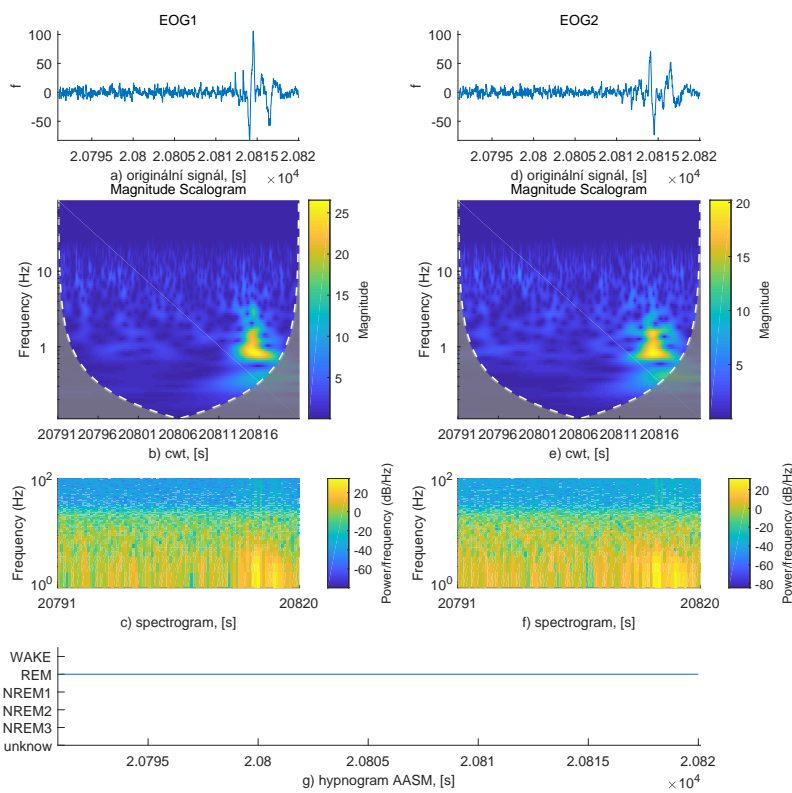
Časově frekvenční analýza REM fáze



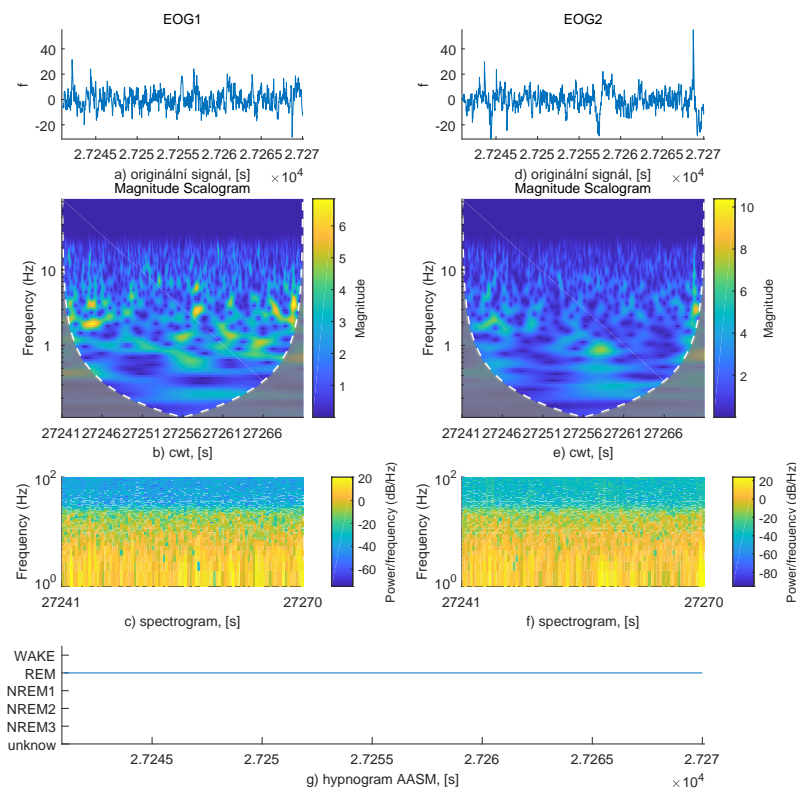
Obrázek D.7. Subjekt č.1, fáze spánku REM, délka segmentu 30 s



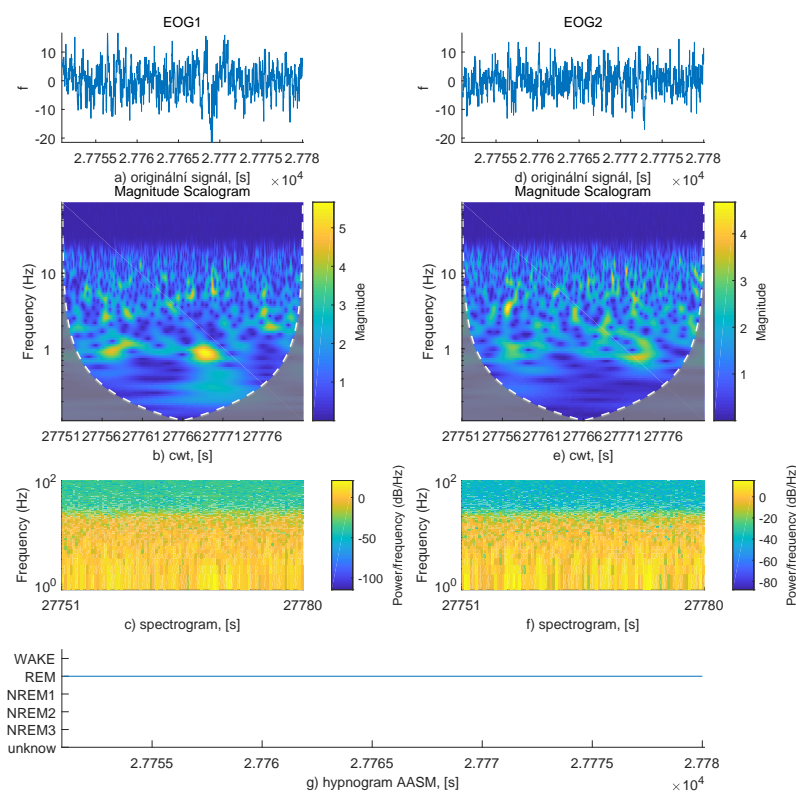
Obrázek D.8. Subjekt č.1, fáze spánku REM, délka segmentu 30 s



Obrázek D.9. Subjekt č.1, fáze spánku REM, délka segmentu 30 s



Obrázek D.10. Subjekt č.1, fáze spánku REM, délka segmentu 30 s



Obrázek D.11. Subjekt č.1, fáze spánku REM, délka segmentu 30 s

Příloha E

SFS pro segmenty různé velikosti

E.1 Průběh SFS pro 5s segment

```
Start forward sequential feature selection:
Initial columns included:  none
Columns that can not be included:  none
Step 1, added column 117, criterion value 0.142572
Step 2, added column 137, criterion value 0.120813
Step 3, added column 143, criterion value 0.100652
Step 4, added column 13, criterion value 0.0907554
Step 5, added column 40, criterion value 0.0813203
Step 6, added column 125, criterion value 0.0785236
Step 7, added column 136, criterion value 0.0750814
Step 8, added column 12, criterion value 0.073176
Step 9, added column 24, criterion value 0.0709939
Step 10, added column 14, criterion value 0.0691192
Step 11, added column 154, criterion value 0.0685967
Step 12, added column 116, criterion value 0.0672752
Step 13, added column 10, criterion value 0.0658
Step 14, added column 49, criterion value 0.0646629
Step 15, added column 11, criterion value 0.0638945
Step 16, added column 29, criterion value 0.0632799
Step 17, added column 8, criterion value 0.0618047
Step 18, added column 120, criterion value 0.0606368
Step 19, added column 187, criterion value 0.0599914
Step 20, added column 34, criterion value 0.0578708
Step 21, added column 130, criterion value 0.0567644
Step 22, added column 119, criterion value 0.0563341
Step 23, added column 15, criterion value 0.0558731
Step 24, added column 118, criterion value 0.0553507
Step 25, added column 23, criterion value 0.0546745
Step 26, added column 158, criterion value 0.0545731
Step 27, added column 17, criterion value 0.0545168
Step 28, added column 56, criterion value 0.0545052
Step 29, added column 121, criterion value 0.0544982
Final columns included:  8 10 11 12 13 14 15 17 23 24 29 34 40 49
 56 116 117 118 119 120 121 125 130 136 137 143 154 158 187
sfs completed, took 39364.32 seconds
```

E.2 Průběh SFS pro 10s segment

```
Start forward sequential feature selection:
Initial columns included:  none
Columns that can not be included:  none
Step 1, added column 117, criterion value 0.134673
Step 2, added column 152, criterion value 0.107444
Step 3, added column 137, criterion value 0.0908476
Step 4, added column 144, criterion value 0.0795992
Step 5, added column 12, criterion value 0.0703792
Step 6, added column 35, criterion value 0.0650317
Step 7, added column 22, criterion value 0.0615895
Step 8, added column 9, criterion value 0.0585162
Step 9, added column 119, criterion value 0.0560575
Step 10, added column 127, criterion value 0.0539062
Step 11, added column 115, criterion value 0.0519393
Step 12, added column 29, criterion value 0.0504641
Step 13, added column 13, criterion value 0.0486815
Step 14, added column 131, criterion value 0.0472063
Step 15, added column 49, criterion value 0.0461614
Step 16, added column 27, criterion value 0.0448706
Step 17, added column 126, criterion value 0.0445018
Step 18, added column 171, criterion value 0.0435798
Step 19, added column 118, criterion value 0.0430881
Step 20, added column 155, criterion value 0.0427193
Step 21, added column 116, criterion value 0.0425963
Step 22, added column 15, criterion value 0.0421046
Final columns included:  9 12 13 15 22 27 29 35 49 115 116
    117 118 119 126 127 131 137 144 152 155 171
sfs completed, took 9656.35 seconds
```

E.3 Průběh SFS pro 15s segment

```
Start forward sequential feature selection:
Initial columns included:  none
Columns that can not be included:  none
Step 1, added column 117, criterion value 0.13286
Step 2, added column 147, criterion value 0.100406
Step 3, added column 136, criterion value 0.0772635
Step 4, added column 139, criterion value 0.0668449
Step 5, added column 12, criterion value 0.0596533
Step 6, added column 24, criterion value 0.0530149
Step 7, added column 49, criterion value 0.0494191
Step 8, added column 157, criterion value 0.0454545
Step 9, added column 123, criterion value 0.0437027
Step 10, added column 10, criterion value 0.0420431
```

```
Step 11, added column 13, criterion value 0.0408446
Step 12, added column 115, criterion value 0.039646
Final columns included: 10 12 13 24 49 115 117 123 136 139 147 157
sfs completed, took 1585.65 seconds
```

E.4 Průběh SFS pro 30s segment

```
Start forward sequential feature selection:
Initial columns included: none
Columns that can not be included: none
Step 1, added column 117, criterion value 0.113037
Step 2, added column 150, criterion value 0.0859303
Step 3, added column 136, criterion value 0.0626959
Step 4, added column 138, criterion value 0.0547667
Step 5, added column 12, criterion value 0.0512631
Step 6, added column 39, criterion value 0.0438871
Step 7, added column 127, criterion value 0.039646
Step 8, added column 13, criterion value 0.0379864
Step 9, added column 10, criterion value 0.035958
Step 10, added column 115, criterion value 0.0346672
Step 11, added column 22, criterion value 0.034114
Step 12, added column 135, criterion value 0.0335608
Step 13, added column 15, criterion value 0.03227
Step 14, added column 11, criterion value 0.0320856
Step 15, added column 131, criterion value 0.0317168
Final columns included: 10 11 12 13 15 22 39 115 117 127 131 135
136 138 150
sfs completed, took 916.90 seconds
```

Příloha F

Matice záměn pro klasifikaci segmentů různé velikosti

		Ohodnoceno expertem	
		REM	WAKE
Odhadnutá hodnota	REM	1754	151
	WAKE	80	1501

Tabulka F.1. Matice záměn k-NN klasifikace 5s segmentů

		Ohodnoceno expertem	
		REM	WAKE
Odhadnutá hodnota	REM	880	57
	WAKE	37	769

Tabulka F.2. Matice záměn k-NN klasifikace 10s segmentů

		Ohodnoceno expertem	
		REM	WAKE
Odhadnutá hodnota	REM	559	47
	WAKE	52	504

Tabulka F.3. Matice záměn k-NN klasifikace 15s segmentů

		Ohodnoceno expertem	
		REM	WAKE
Odhadnutá hodnota	REM	294	19
	WAKE	12	256

Tabulka F.4. Matice záměn k-NN klasifikace 30s segmentů

Příloha G

Adresářová struktura DVD disku

```
| thesis.pdf
|
\---program
|   main_test.m
|   main_train.m
|   main_time_frequency_analysis.m
|
+---data
|   +---db
|   |   |   DREAMS_databases_License.txt
|   |   |
|   |   +---hypnogram_AASM
|   |   |   README.txt
|   |   |
|   |   \---subjects
|   |   |   README.txt
|   |   |
|   \---internal
|   |
|   |   +---matlab
|   |   |   k_best_5.mat
|   |   |   k_best_10.mat
|   |   |   k_best_15.mat
|   |   |   k_best_30.mat
|   |   |   main_features_ids_5.mat
|   |   |   main_features_ids_10.mat
|   |   |   main_features_ids_15.mat
|   |   |   main_features_ids_30.mat
|   |   |   test_data_set_5.mat
|   |   |   test_data_set_10.mat
|   |   |   test_data_set_15.mat
|   |   |   test_data_set_30.mat
|   |   |   train_data_set_5.mat
|   |   |   train_data_set_10.mat
|   |   |   train_data_set_15.mat
|   |   |   train_data_set_30.mat
|   |   |
|   \---time_frequency_analysis
|   |   +---R
|   |   |   012061_012090_R_1.pdf
```

```
|          |          ...
|          |          027751_027780_R_1.pdf
|          |
|          \---W
|          000001_000030_W_1.pdf
|          ...
|          028801_028830_W_1.pdf
|
|
\---functions
|  calculate_cwt.m
|  calc_k_best.m
|  classification_quality.m
|  classify_knn.m
|  collect_data.m
|  extract_eog_signals.m
|  extract_main_features.m
|  extract_signal_part.m
|  feature_extracting.m
|  get_seconds.m
|  get_start_indices_hypno.m
|  labels_compression.m
|  labels_extracting.m
|  plot_cwt.m
|  plot_main_components.m
|  plot_time_frequency_analysis.m
|  prepare_test_data.m
|  prepare_train_data.m
|  read_edf_hypnogram_data.m
|
|
\---3rd_party
|  edfread.m
|  fprintf_r.m
|  panel.m
```