

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

Fakulta elektrotechnická

Katedra kybernetiky



Analýza EEG signálu

Diplomová práce

Bc. Robin Horniak

2010

Vedoucí diplomové práce: Ing. Václav Gerla

ZADÁNÍ DIPLOMOVÉ PRÁCE

Student: Bc. Robin Horniak
Studijní program: Elektrotechnika a informatika (magisterský), strukturovaný
Obor: Biomedicínské inženýrství
Název tématu: Analýza EEG signálu

Pokyny pro vypracování:

Cílem diplomové práce je realizovat framework umožňující klasifikaci EEG záznamů. Její hlavní účel je usnadnit zavádění již navržených a ověřených metod na spolupracujících lékařských pracovištích (FN Bulovka, spánková laboratoř v Bohnicích, ÚPMD).

Navržený a implementovaný nástroj by měl obsahovat následující funkce:

1. Datová analýza (uložení EEG dat a výsledků segmentace / extrakce příznaků / klasifikace).
2. Přehledné zobrazení dat (EEG signály, vybrané příznaky, spektrogramy a koherence, výsledky shlukové analýzy, klasifikace do tříd s možností uživatelské editace).
3. Výběr a realizace vybraných klasifikačních metod.
4. Volitelné: reimplementace vybraných metod (filtrace signálu, segmentace, extrakce příznaků, shluková analýza).
5. Testování a dokumentace.

Poznámka k implementaci:

- Java / C++ / .NET,
- modularita (jednotlivé výpočty budou probíhat v oddělených modulech a na ty se bude hlavní aplikace odkazovat),
- provázanost s již realizovanou EEG databází grafoelementů.

Velký důraz bude kladen na použitelnost navrženého řešení v odlišných úlohách měření EEG (spánek dospělých, epileptické/komatické/novorozenecké záznamy apod.).

Seznam odborné literatury:

- [1] Malmuvio J., Plonsey R.: Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields. New York: Oxford University Press.1995.
- [2] Liu, H., Motoda, H.: Feature Extraction, Construction and Selection: a Data Mining Perspective. National University of Singapore, Singapore, 1998, ISBN 0-7923-8196-3.
- [3] Schlesinger, M. I., Hlaváč, V.: Ten Lectures on Statistical and Structural Pattern Recognition", CTU, Prague, 1999.

Vedoucí diplomové práce: Ing. Václav Gerla

Platnost zadání: do konce letního semestru 2009/2010

prof. Ing. Vladimír Mařík, DrSc.
vedoucí katedry



doc. Ing. Boris Šimák, CSc.
děkan

Prohlášení

Prohlašuji, že jsem svou diplomovou práci vypracoval samostatně a použil k tomu úplný výčet citací použitých pramenů, které uvádím v seznamu přiloženém k práci.

Nemám závažný důvod proti užití tohoto školního díla ve smyslu §60 Zákona č.121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon).

V Praze dne 04.01.2010


.....
podpis

Poděkování

Na tomto místě bych chtěl poděkovat především vedoucímu mé diplomové práce Ing. Václavu Gerlovi za vstřícnost a cenné připomínky k práci. Mé díky patří také rodině a blízkým za podporu po dobu studia.

Abstrakt:

Cílem této práce je usnadnění vyhodnocování EEG záznamu. Největší pozornost je věnována přehledné vizualizaci EEG signálu a zobrazení výsledků některých analýz provedených nad EEG daty (adaptivní segmentace signálu, spektrální analýza, hierarchické shlukování). Druhou částí je klasifikace jednotlivých segmentů EEG záznamu s využitím databáze artefaktů. Důležitou součástí je možnost následného rozšiřování programu o zásuvné moduly, které mohou sloužit k dalšímu zpracování, vyhodnocování a vizualizaci EEG dat. Celá práce je navržena a implementována v jazyce C++ ve vývojovém prostředí Borland C++ Builder 6.

Abstract:

The main objective of this Diploma Thesis is to analyze EEG signals. EEG signal visualizing receives the most attention together with results of some analyzed EEG data (adaptive signal segmentation, spectrum analysis and hierarchical clustering). The second parts deals with the classification of individual EEG segments and using a database of artefacts. An important component is the possibility to further develop the program using additional modules which can be used for further EEG data processing, analyze and visualization. The whole thesis is designed and implemented in C++ using the Borland C++ Builder environment.

Obsah

1) Úvod	1
2) Současný stav problematiky	2
3) Elektroencefalografie	3
3.1 EEG signál	3
3.2 EEG signály - měření	4
3.3 EEG signály - artefakty	5
3.4 EEG signály - vzorce	5
4) Předzpracování EEG signálu	7
4.1 Odstranění rušení ze sítě 50Hz	7
4.2 Odstranění kolísání izolinie	7
4.3 Segmentace signálu	7
4.3.1 Konstantní segmentace	8
4.3.2 Adaptivní segmentace	8
5) Zpracování signálu	9
5.3 Výpočet příznaků	9
5.4 Selektce příznaků	12
5.5 Klasifikace	12
5.5.1 Shluková analýza	12
5.5.2 Klasifikace podle nejbližších sousedů [k-NN]	14
5.6 Vizualizace	14
6) PROGRAMOVÁ ČÁST	15
6.3 Úvod	15
6.4 Formáty vstupních dat	16
6.4.1 XML soubory	16
6.4.2 BIN - Binární soubory	17
6.4.3 TXT soubory	17
6.4.4 ARFF soubory	18
6.5 Výstupní data	18
6.6 Popis aplikace	18
6.6.1 Hlavní část programu	19
6.6.2 Zobrazení signálů	22

6.6.3	Zobrazení segmentace	27
6.6.4	Práce s klasifikovanými soubory	34
6.6.5	Výsledky shlukové analýzy	39
6.6.6	Zobrazení příznaků.....	41
6.6.7	Zobrazení spektrogramů	43
6.7	Klasifikace podle nejbližších sousedů [k-NN]	45
6.8	Nastavení programu	46
7)	Závěr.....	47
8)	Reference	48

Seznam obrázků

<i>obrázek 3-1 – rozložení elektrod</i>	3
<i>obrázek 3-2 – snímací zařízení</i>	3
<i>obrázek 3-3 - ukázka naměřeného EEG</i>	4
<i>obrázek 5-1 – hierarchické shlukování - dendrogram</i>	13
<i>obrázek 6-1 - struktura XML souboru</i>	16
<i>obrázek 6-2 – názvy příznaků</i>	18
<i>obrázek 6-3 – začátek bloku s daty</i>	18
<i>obrázek 6-4 - hlavní okno</i>	19
<i>obrázek 6-5 - stavový panel</i>	20
<i>obrázek 6-6 - načtené kanály</i>	21
<i>obrázek 6-7 - základní rozložení, zobrazení signálu</i>	22
<i>obrázek 6-8 - zobrazení vertikálního posuvníku</i>	23
<i>obrázek 6-9 - změna amplitudy</i>	24
<i>obrázek 6-10 - příklad změny intervalu zobrazení</i>	25
<i>obrázek 6-11 - změna barev signálů</i>	26
<i>obrázek 6-12 – zobrazení menu „Segments“</i>	27
<i>obrázek 6-13 - Adaptivní segmentace, bez popisku</i>	28
<i>obrázek 6-14 - zobrazení adaptivní segmentace - přehledně</i>	29
<i>obrázek 6-15 - zobrazení adaptivní segmentace - nepřehledně</i>	29
<i>obrázek 6-16 - výběr segmentace menu</i>	30
<i>obrázek 6-17 - výběr segmentace</i>	30
<i>obrázek 6-18 - odkryté objekty</i>	31
<i>obrázek 6-19 - výběr segmentů</i>	31
<i>obrázek 6-20 - výběr bloku vzorků</i>	32
<i>obrázek 6-21 - výběr zobrazení segmentace s popisem</i>	33
<i>obrázek 6-22 - PopupMenu</i>	33
<i>obrázek 6-23 - chybí automatická klasifikace</i>	34
<i>obrázek 6-24 – zobrazení klasifikace</i>	34
<i>obrázek 6-25 - zobrazená data</i>	35
<i>obrázek 6-26 – nepřehledné zobrazení</i>	36
<i>obrázek 6-27 – použitý ZOOM – přehledné zobrazení</i>	36
<i>obrázek 6-28 - aktuální pozice signál</i>	37
<i>obrázek 6-29 – editace signálu</i>	38
<i>obrázek 6-30 – editace signálu změna</i>	39
<i>obrázek 6-32 – Clusters – použití ZOOM X</i>	40
<i>obrázek 6-31 – Clusters – bez použití ZOOM X</i>	40
<i>obrázek 6-33 - načtení příznaků</i>	41
<i>obrázek 6-34- zobrazení příznaků</i>	42
<i>obrázek 6-35 – přehledné zobrazení příznaky</i>	43

<i>obrázek 6-36 – barevná maska pro spectrogram</i>	43
<i>obrázek 6-38 – Normovaný spektrogram</i>	44
<i>obrázek 6-37 – Welchův spektrogram</i>	44
<i>obrázek 6-39 – klasifikace k-NN</i>	45
<i>obrázek 6-40 – Nastavení souborů</i>	46

1) Úvod

Program VisualEEG (Visualization of Electroencephalographic Data) je součástí většího projektu. Celý projekt je rozdělen do tří částí: *EEG databáze* [1], *PSGLab Matlab Toolbox* [2] a zmiňovaný program VisualEEG. Program VisualEEG je zaměřen na práci s předzpracovanými signály v PSGLabu a je určen zejména pro přehlednou vizualizaci vypočtených výsledků.

Hlavní funkce programu

- zobrazení signálů (EEG i další druhy biologických signálů)
- zobrazení výsledků konstantní a adaptivní segmentace
- vizualizace vypočtených příznaků
- zobrazení klasifikace do tříd s možností editace
- zobrazení výsledky shlukové analýzy

V diplomové práci je nejprve stručně popsána problematika EEG, dále následuje současný stav v oblasti EEG a důvody vzniku tohoto projektu. V dalším bloku jsou uvedeny druhy segmentací a měřených příznaků. Samotný program je popsán podrobně v programové části. V závěru je uveden přínos a konečné zhodnocení.

Cílem této práce bylo navrhnout co nejpřehlednější a pro běžného uživatele co nejpohodlnější způsob zobrazení výsledků různých analýz nad EEG daty. Program je z důvodů kompatibility a návaznosti na současný projekt vytvořen v jazyce C++ (algoritmy z PSGLabu jsou průběžně převáděny do C++).

Veškerou práci je zdokumentována ve verzi HTML, která slouží taktéž jako nápověda pro práci s programem.

2) Současný stav problematiky

V současnosti existuje nemalý počet aplikací umožňující zobrazení nebo práci s EEG signálem (např. aplikace WaveFinder, dodávána společně s komerčními EEG systémy). Prozatím ale nebyla navržena volně šiřitelná aplikace, která by zahrnovala sběr, zpracování a zobrazení výsledných dat v použitelné podobě a současně by umožňovala další rozšiřitelnost jednotlivých částí aplikace. Celkový projekt by tedy měl pomáhat při celkové lékařské diagnóze ke snadnějšímu určení choroby, úrazu nebo jiné mozkové poruše.

Fakulta elektrotechnická ČVUT spolupracuje s pražskou nemocnicí Na Bulovce. Díky této spolupráci má k dispozici reálné EEG záznamy. První částí projektu, se v rámci své diplomové práce zabýval Ing. Martin Kytka. Jeho práce spočívala ve sběru a ukládání grafoelementů do databáze. Tato databáze nyní slouží pro automatickou detekci a případné odstranění nežádoucích grafoelementů ze EEG záznamů.

Druhou částí, se zabývá Ing. Václav Gerla. Jedná se o zpracování polysomnografických signálů a jejich následnou klasifikaci. Touto cestou vzniká velké množství výstupních dat, která jsou složitě zpracovávána pro jejich opětovné zobrazení.

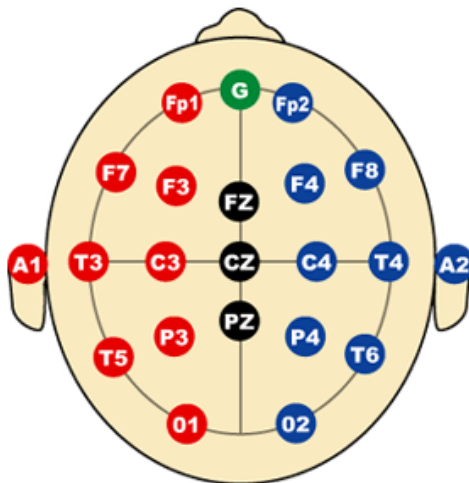
Tato diplomová práce je tedy zaměřena na zobrazení výstupních dat z PSGLab toolboxu. Při návrhu byl kladen velký důraz na použitelnost v medicínském prostředí. Tak bude možné doposud dosažené výsledky konzultovat i s lékaři. Nebylo opomenuto ani další rozšíření aplikace o další moduly zobrazení nebo předzpracování dat.

3) Elektroencefalografie

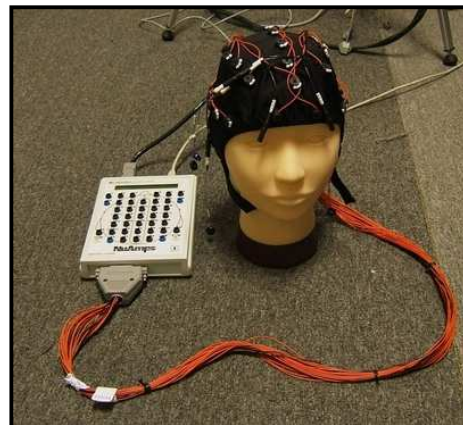
3.1 EEG signál

Elektroencefalogram (EEG) je záznam časové změny elektrického potenciálu způsobeného mozkovou aktivitou. Záznamy jsou pořizovány elektroencefalografem a to z povrchu skalpu lebky. Nejčastější zapojení elektrod pro měření je 10/20. Rozložení elektrod je uvedeno na *obr. 3-2*. Před měřením je nutno nanést na každou elektrodu gel, kterým se zvýší vodivost a tak dosáhneme přesnějších výsledků při měření.

Elektroencefalograf – přístroj zaznamenávající mozkovou aktivitu. Snímaná mozková aktivita má napětí řádově desítek mikrovoltů. Pro přesnější zpracování je třeba daný signál zesílit. Mezi největší nedostatky patří nežádoucí rušení signálu (špatné odstínění, špatný kontakt elektrod s pokožkou, elektromagnetické záření apod.). Přístroj se skládá ze dvou hlavních částí. První část tvoří „EEG čepice“ s rozložením elektrod a druhou část hlavní výkonová jednotka (viz. *obr. 3-1*).



obr. 3-2 - rozložení elektrod



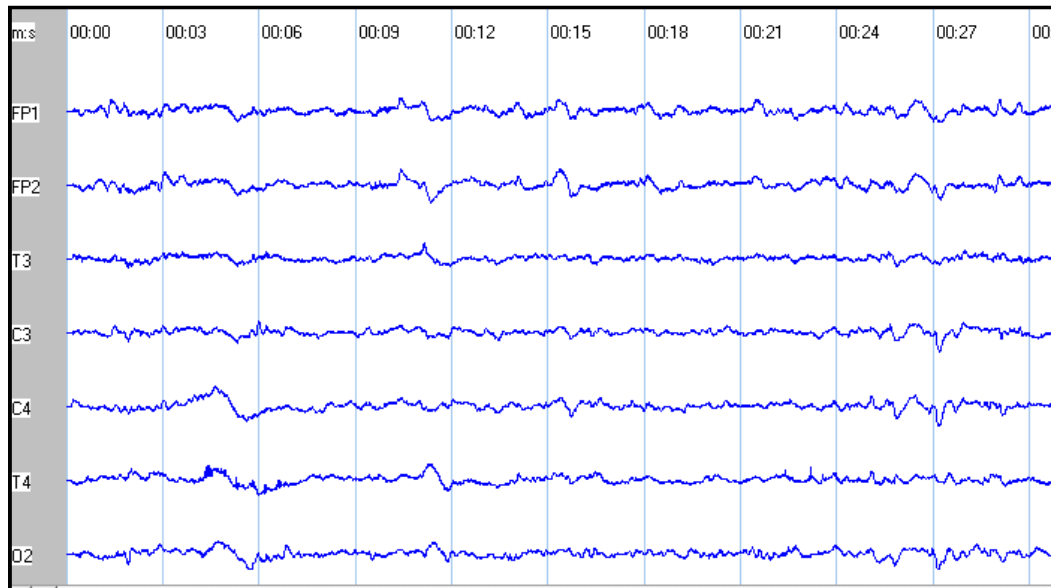
obr. 3-1 - snímací zařízení

3.2 EEG signály - měření

Měření EEG signálu je prováděno např. za účelem diagnózy chorob, epilepsií nebo nádorů. Jsou zaznamenávány několikaminutové až hodinové záznamy, které jsou posléze zpracovávány. Příklad naměřeného EEG signálu¹ je zobrazen na obr. 3-3.

V EEG signálu vznikají vlny o různých vlnových délkách, nejčastěji v rozpětí 0,1 – 40 Hz. U zdravého člověka jde velmi dobře zaznamenat usínání a posléze přechod mezi spánkem *REM (Rapid Eye Movement)* – pohyb očí během spánku a *NREM (Non Rapid Eye Movement)* – hluboký spánek.

Nízké frekvence odpovídají spánku (hluboký spánek → delta vlny: 0,1 – 3 Hz), vyšší frekvence pak bdělosti (Beta vlny: > 20Hz). Pro více informací lze čerpat např. z předešlé práce [1].



obr. 3-3 - ukázka naměřeného EEG

¹ Měřený signál se skládá z tzv. vzorků. Celkový počet vzorků odpovídá jednak délce záznamu a dále vzorkovací frekvenci. Máme-li tedy záznam dlouhý 30min, počet vzorků při vzorkovací frekvenci 128Hz bude $128 \times (30 \times 60) = 230400$ vzorků.

3.3 EEG signály - artefakty

Při měření signálu se vyskytují tzv. artefakty. Artefakty lze rozdělit do dvou skupin:

- a) biologické
- b) technické

Do kategorie biologické řadíme veškeré artefakty způsobené lidskou aktivitou (mrknutí, svalová aktivita, pocení při EEG vyšetření aj.).

Do kategorie technické pak řadíme artefakty způsobené ostatními vlivy (špatná elektroda, špatné odstínění aj.).

Artefakty komplikují hodnocení naměřených záznamů. Proto je vhodné tyto oblasti detekovat a usnadnit tak následnou klasifikaci dat.

3.4 EEG signály - vzorce

Normální EEG aktivita

U zdravého jedince se může vyskytovat několik základních EEG vzorců. Tyto základní vzorce jsou dále používány pro popis pozadí EEG signálu. Veškeré tyto vzorce jsou hodnoceny jako normální nález a svým původem přikládány bdělému relaxovanému stavu zdravého mozku.

Normální reaktivita

Reaktivita normálního grafu vytržená z kontextu základní aktivity může připomínat abnormitu epileptiformní, epileptogenní nebo pomalou, častěji generalizovanou. Například při fotostimulaci nacházíme ostré vlny nad zadními regiony.

Normální změny objevující se při kolísání vigility a ve spánku

Měření EEG během usínání a spánku patří mezi nejčastější EEG vyšetření. Výsledky jsou cíleně získávána ze spánkového grafu. Změny při usínání zahrnují zejména zpomalení základní aktivity s občasným výskytem vln tzv. ostrého charakteru. Pro rozlišení od epileptogenních změn je důležitá jejich vazba na určitá spánková stádia².

Novorozenecká aktivita

Novorozenecká aktivita je velmi odlišná od aktivity dospělého jedince. Liší se i samotné měření, protože se používá pouze 8 měřících elektrod (z důvodů malé hlavy novorozence). EEG záznamy je možno pořizovat od 20. týdne gestačního věku. V 22. týdnu lze pozorovat reakce na světlo, zvuk a jiné stimuly vnějšího prostředí. Zhruba v 27. týdnu můžeme pořídit záznam se stranovou synchronizací a od 30. týdne lze rozlišit bdění a spánek. U novorozenců se zejména zabýváme záznamem spánkového EEG. Novorozenec potřebuje pro zotavení organismu 20-22 hodin z celého dne oproti dospělého jedince, kterému stačí pro zotavení přibližně 8 hodin spánku.

Spánek novorozence má obecně cyklický charakter, kdy se střídají pomalý spánek (*IV stádia spánku, kde IV stádium spánku tvoří více než 50% aktivity delta vlnami*) s paradoxním spánkem (*II fáze, I. Fyzická – svalové a oční záškuby, II. Atonická – bez záškubů*). Rozdíl mezi spánkem novorozence a dospělým jedincem je v délce trvání jednotlivých cyklů. Po krátké fázi usínání nastává fáze pomalého spánku – stav, kdy mozková kůra odpočívá, činnost podkorových ústředí je zachována. U dospělého jedince tato fáze trvá kolem 90 minut, kdežto u novorozence 10 – 25 minut. Fáze paradoxního spánku u dospělého jedince tvoří přibližně 20% spánku s trváním asi 10minut. U novorozenců tato fáze pokrývá až 80% celkové doby spánku.

² některé záchvaty se vyskytují více ve spánku nebo při probouzení a usínání. Je tedy třeba je odlišit a neklasifikovat je jako součást spánku

4) Předzpracování EEG signálu

Metody předzpracování signálu mají za hlavní úkol eliminaci všech nežádoucích složek, které se mohou v EEG signálu objevit.

Během měření mozkové aktivity dochází jevům, která je potřeba eliminovat buď hned při měření (například pro docílení lepšího kontaktu s pokožkou je používán gel, který je nanášen na elektrody), nebo po naměření dat. Dochází například k rušení z elektrické rozvodné sítě frekvencí 50Hz a může docházet ke kolísání izolinie. Mezi další nežádoucí jevy můžeme zařadit také artefakty biologické (mrkání, svalová aktivita, dýchání).

4.1 Odstranění rušení ze sítě 50Hz

Protože v EEG signálu se frekvence vyšší než 50Hz nevyskytují, používá se nejčastěji pro odstranění tohoto typu rušení filtr typu dolní propust. Tedy filtr, který propustí frekvence menší než 50 Hz. Další možností je použití notch filtru.

Rušení lze také odstranit již během měření hardwarově, zapojením filtru do obvodu anebo softwarově odečtením 50 Hz frekvence ze signálu.

4.2 Odstranění kolísání izolinie

Během měření dochází ke kolísání signálu, pro odstranění tohoto jevu, je nejvhodnější použít průměrování signálu. Výsledek odečteme od původního signálu, a dostaneme námi požadovaný ustálený průběh.

4.3 Segmentace signálu

Segmentací signálu rozumíme rozdělení signálu na menší části. Provádí se za účelem dalšího zpracování signálu a případného odstranění nežádoucích artefaktů. Segmentaci rozdělujeme na:

- konstantní segmentace – konstantní délka segmentu
- adaptivní segmentace – proměnná délka segmentu

4.3.1 Konstantní segmentace

Při konstantní segmentaci se používá fixní délka segmentů – veškeré segmenty obsahují stejný počet vzorků a stejnou délku. Délka segmentů pro segmentaci EEG se volí v řádu jednotek sekund až minut. Nevýhodou tohoto rozdělení je nepřímý vztah k rozdělovanému signálu. Mezi výhodou jednoduchost implementace algoritmu na rozdělení.

4.3.2 Adaptivní segmentace

Zde je délka segmentu proměnná. Délka segmentu záleží na aktuální pozici v signálu a jeho aktuální změně. Hranice segmentů jsou voleny tak, že výsledné segmenty lépe korespondují se změnami v signálu, než při konstantní segmentaci. Hranice jsou hledány pomocí tzv. klouzavých oken. Jedna z možných realizací těchto oken je posouvání dvou oken vedle sebe o délce Nw v signálu a na základě vyhodnocování lokálních maxim z námi definovaných charakteristik pro segmenty. Adaptivní segmentace může být implementována pomocí různých algoritmů. V této práci byly použity segmentace založené na *FFT* a *Fuzzy c-means*.

Metoda adaptivní segmentace založená na FFT

Pomocí hlavních rytmů EEG (alfa, beta, gama, delta) nadefinujeme 4 oblasti ve frekvenčních oblastech. Hranice segmentu jsou tedy definovány změnami v dílčích frekvenčních oblastech.

$$X[k] = \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{-jkn(\frac{2\pi}{N})} \quad (4.1)$$

Pomocí diskrétní Fourierovi transformace (4.1) vypočteme amplitudy $X[k] = 0 \leq k \leq N-1$. Protože známe také vzorkovací frekvenci f_{vz} , můžeme spočítat jednotlivé frekvence (0 až $f_{vz}/2$) pro obdržené amplitudy. Délka FT N se volí shodně s délkou okna Nw .

Nakonec spočteme výkon (součet druhých mocnin amplitud všech frekvencí příslušejících do tohoto pásma.) Celkový průměr je dán jako vázaný průměr dílčích výkonů $p_\alpha \dots p_\theta$ koeficienty $k_\alpha \dots k_\theta$ a je funkcí n tedy proměnný s pozicí oken v rámci EEG signálu.

$$p_c[n] = k_\alpha p_\alpha[n] + k_\beta p_\beta[n] + k_\gamma p_\gamma[n] + k_\theta p_\theta[n] \quad (4.2)$$

Hranice odpovídají lokálním maximům posloupnosti. Více o segmentacích lze najít v [3].

Metoda adaptivní založená na Fuzzy c-means

Další možností je adaptivní segmentace Fuzzy c-means. Algoritmus přiřazuje jednotlivé vzorky do tříd. Více o této metodě a přesném popisu algoritmu *Fuzzy c-means* lze najít v publikaci docenta Vladimíra Krajčí [4].

5) Zpracování signálu

Po předzpracování signálu následují další metody zpracování signálu. Jelikož jsou data rozdělena do segmentů, můžeme s jednotlivými segmenty pracovat a provádět na nich výpočty, přiřadit je do tříd a podobně.

Nabízí se také celé řada vizualizačních metod (zobrazení spektrogramu, EEG koherencí a jiné), nebo použití shlukové analýzy.

5.3 Výpočet příznaků

Každý segment je popsán sadou příznaků. Výpis některých příznaků, které byly použity při testování v této práci:

- Statické parametry
 - Minimum (*Minimum*)
 - Maximum (*Maximum*)
 - Střední hodnota (*Mean*)
 - Koeficient špičatosti (*Kurtosis*) - charakteristika rozdělení náhodné veličiny, která porovnává náhodné rozdělení s normálním rozdělením pravděpodobnosti.
 - Koeficient šikmosti (*Skewness*) - je charakteristikou rozdělení pravděpodobnosti náhodné veličiny.
- Průměr a maximum hodnot první a druhé derivace

- RMS amplituda – efektivní hodnota, kdy se měří střední hodnota signálu za určitou dobu T.
- Shannonova entropie – entropie popisuje nejistotu výskytu objektu. Pokud diskretní náhodná veličina X nabývá konečného počtu x_i s pravděpodobnostmi

$$P(X = x_i) = p(x_i) \geq 0 \text{ pro } i = 1, 2, \dots, r, \sum p(x_i) = 1. \quad (5.1)$$

Chceme zde dosáhnout, aby nejistota o existenci hodnoty x_i byla větší než její pravděpodobnost $p_i(x_i)$ výskytu. Shannon $H(x)$ je popsán v diskretní náhodné veličině X jako entropie střední hodnoty elementární informace. Její vzorec:

$$H(X) = \sum_{i=1}^r p(x_i) I(x_i) \quad (5.2)$$

kde $I(x_i)$ je tzv. elementární informace:

$$I(x_i) = -\ln p(x_i) = \ln \frac{1}{p(x_i)} \quad (5.3)$$

Pokud klesá entropie, pak roste celková informace a naopak. Shannonova entropie tedy popisuje to, co by se mohlo v signálu vyskytovat.

- Absolutní a relativní výkonová spektrální hustota – Spektrální hustota je matematicky popsána jako Fourierova transformace autokorelačních časových úseků. Při vizualizaci spektrální hustoty lze vysledovat v EEG signálu aktivitu v těchto frekvenčních pásmech:

- *delta* 0,1 - 4 Hz
- *theta* 4 - 8 Hz
- *alfa* 8 - 13 Hz
- *beta* 13 - 30 Hz
- *gama* nad 30 Hz

- Statické hodnoty z vlnkových (wavelet) koeficientů – Waveletová transformace (WT) poskytuje oproti Fourierové transformaci informaci o časové lokalizaci spektrálních složek. Každá waveletová funkce osciluje pouze v okolí bodu, což poskytuje dobrou prostorovou lokalizaci. Cílem WT je rozložit vstupní signál do řady waveletových koeficientů. Toto je dosaženo filtrováním signálu párem

ortogonálních filtrů, které jsou označeny jako otcovský wavelet a mateřský wavelet. Otcovský wavelet určuje celkový trend signálu – rozklad na škálové koeficienty, zatímco mateřský wavelet zachycuje doplňkovou informaci o „jemnostech“ na jednotlivých úrovních – waveletové koeficienty. Základní rozdíl mezi Fourierovou a WT je v tom, že wavelety nejsou periodické funkce. Toto dělá WT velmi výhodnou pro analýzu nestacionárních nebo aperiodických signálů [5].

Wavelets musí splňovat základní podmínku – musí mít konečnou energii:

$$E = \int_{-\infty}^{\infty} |\Psi(t)|^2 dt < \infty \quad (5.4)$$

Dále musí platit, že její střední hodnota je rovna 0. $\int_{-\infty}^{\infty} \Psi(t) dt = 0$. Wavelets je detailně popsán v [6].

- Statické koeficienty korelace – Korelace vystihuje podobnost dvou signálů pro různé vzájemné posuvy. Pokud vypočteme celý vektor korelačních koeficientů $R(\tau)$, tzn. Pro všechny možné posuvy τ , bude jednomu určitému τ odpovídat největší korelační koeficient $R(\tau)_{max}$, kde τ je hledaná hodnota zpoždění v signálu [7].

$$R_{12}(\tau) = \sum_{k=-\infty}^{\infty} s_1(k + \tau) s_2^*(k), \tau \in Z \quad (5.5)$$

- Statické koeficienty koherence - koherence, je míra přesnosti vzájemné spektrální hustoty a je určena jako kvadrát vzájemné spektrální hustoty dělený součinem jednotlivých spektrálních hustot.

$$\gamma_{xy}(f) = \frac{|S_{xy}(f)|^2}{S_x(f)S_y(f)}, 0 \leq \gamma_{xy}^2 \leq 1. \quad (5.6)$$

kde 0 odpovídá nejmenší míře přesnosti a 1 maximální míře přesnosti vzájemné spektrální hustoty [7].

Všechny příznaky jsou podrobně popsány v práci [8].

5.4 Selekcce příznaků

Během zpracovávání příznaků dochází k nadměrnému množství dat. Je tedy potřeba vybrat pouze ty nejkvalitnější z nich. Jednou možností jak vybrat nejkvalitnější příznaky je selekcce příznaků. Existují různé metody selekcce příznaků. Vhodné příznaky je možné např. vybrat za pomoci lékaře, zobrazením klasifikace a vypočtených příznaků. Tím zaručíme, že vybereme opravdu ty nejlepší příznaky pro následnou klasifikaci.

5.5 Klasifikace

Klasifikace signálu je jednou z nejdůležitějších součástí procesu zpracování EEG záznamu. Klasifikací rozumíme rozdělení segmentů do příslušných tříd. Jedna konkrétní třída může odpovídat některé spánkové fázi, přítomnosti artefaktu, epileptické aktivitě apod.

Klasifikační metody

- S učitelem – klasifikátor je naučen za pomoci trénovací množiny. Pro správnou klasifikaci je potřeba velká trénovací množina. Typ výstupu je také předem známý.
Mezi tyto metody patří např. neuronové sítě [9].
- Bez učitele – klasifikátor se učí bez trénovací množiny. Je založen pouze na vnitřní struktuře dat a jeho výstupem jsou *shluky dat* s podobnými vlastnostmi. Příkladem takové klasifikace je *K-means* a *hierarchické shlukování*.

5.5.1 Shluková analýza

Pomocí vypočítaných příznaků a adaptivní segmentace lze signál rozdělit do několika skupin – tzv. clusterů. Na rozdělení do těchto skupin lze použít více typů algoritmů. Nejčastěji používané jsou K-means a hierarchické shlukování.

K-means

Princip K-means: Je potřeba rozdělit x segmentů do k -shluků (konstanta k je předem zvolená). Výstupem je pak množina K shluků. Postup rozdělení: Jako první se

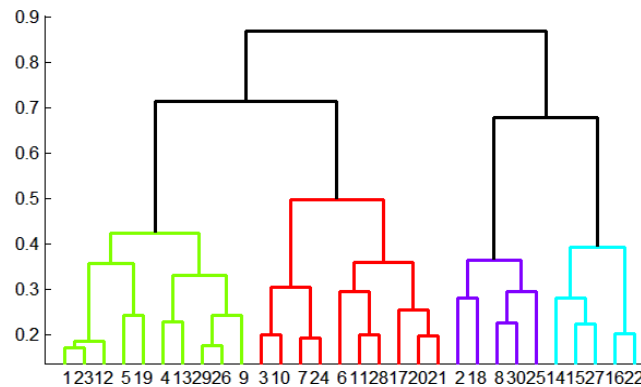
zvolí k náhodných segmentů a jejich hodnoty v příznacích. Pomocí euklidovské vzdálenost se pak

jednotlivé segmenty přesouvají mezi shluky. Konec algoritmu končí, pokud začne konvergovat nebo nastane jiná podmínka pro konec algoritmu (např. počet iterací).

Nevýhodou K-means je, že při každém startu může přiřadit segmenty do jiných shluků, musí se také zadat počet tříd na začátku algoritmu. Mezi klady patří zejména rychlost a malé nároky na operační paměť (jedná se o sekvenční algoritmus).

Hierarchické shlukování

Princip hierarchického shlukování: vytváří se zde stromová struktura známa jako dendrogram³. Na začátku algoritmu je každý segment jedním shlukem a je také potřeba definovat způsob hodnocení podobnostních vztahů mezi shluky, podle tohoto způsobu pak v každém kroku vybereme shluky, které si jsou nejpodobnější. Jeden krok algoritmu sníží počet shluků o 1. Algoritmus končí po $n-1$ krocích, kde n je počet segmentů, dostaneme tedy jeden výsledný shluk o velikosti n . Princip hierarchického shlukování je zobrazen na obr. 5-1. Detailnější popis K-means i hierarchického shlukování je v [10].



obr. 5-1 – hierarchické shlukování - dendrogram

³ Dendrogram = zobrazení hodnot ve dvourozměrném prostoru, jedná se o binární strom, který znázorňuje hierarchické shlukování, kde každý shluk je právě jeden uzel stromu. Osy zde tvoří zadané proměnné, horizontální řez představuje rozklad ze shlukovací sekvence a vertikální směr pak vzdálenost mezi shluky.

5.5.2 Klasifikace podle nejbližších sousedů [k-NN]

Klasifikace nejbližších sousedů spadá mezi neparametrické metody klasifikace. Tyto metody jsou založeny na podstatně slabších předpokladech než metody parametrické. Nepředpokládáme znalost tvaru pravděpodobnostních charakteristik tříd.

Metoda nejbližších sousedů je založena na hledání přímo aposteriozní pravděpodobnosti. Je myšlenkovým rozšířením metody klasifikace podle nejbližší vzdálenosti od etalonu.

Princip – známe trénovací množinu $\{(x_1, \omega_1)\}_{t=1, \dots, K}$, kde x_1 je vzorek, kterému je přiřazena třída ω_1 a K je velikost trénovací množiny. Pro neznámý prvek x hledáme x'_k takové, že $\|x'_k - x\| = \min_{t=1, \dots, K} \|x'_t - x\|$. Podle navoleného počtu sousedů „ k “ hledáme „ k “ vzorků s nejkratší vzdáleností. Třída ω_1 , která dominuje ve vybraných vzorcích je pak zvolena jako výsledná. [11].

5.6 Vizualizace

Veškerá data, která máme naměřena, či výsledky analýz nad těmito daty lze zobrazit. Zobrazovat lze například jednotlivé signály, vypočtené příznaky, spektrogramy a koherence. Nespornou výhodou při zobrazování EEG signálů je snadná přehlednost. V programové části je popsáno několik druhů vizualizace.

- ***Zobrazení signálů, příznaků, rozdělení do tříd a clusterů***

Jedná se o dvourozměrné zobrazení, výstupem je průběh signálu, nebo průběhy příznaků, clusterů a tříd.

- ***Zobrazení spektrogramu***

Spektrogram je zobrazení intenzity (spektrální hustoty) v časové a frekvenční ose. Horizontální osa určuje čas a vertikální osa určuje frekvenci. Třetím rozměrem je intenzita (používá se například barevné kódování). Jedná se o přehlednější způsob zobrazení než je tomu u zobrazení signálu v časové oblasti.

6) PROGRAMOVÁ ČÁST

6.3 Úvod

Pro tvorbu programu bylo zvoleno vývojové prostředí C++ BUILDER 6. Z důvodu, že program musí zvládat více zobrazení signálů, bylo třeba zvolit vhodné rozvržení oken. Program obsahuje pevné nastavení vstupních názvů souborů, díky pevné struktuře dat. S touto strukturou se počítá i do budoucna.

Program se skládá z těchto hlavních částí⁴:

- a) hlavní okno
- b) okno se signály
- c) okno s příznaky
- d) okno s třídami
- e) okno s výsledky shlukování
- f) okno se spektrogramy
- g) klasifikace podle nejbližších sousedů [k-NN]
- h) nastavení souborů pro aplikaci

Na tvorbu programu navazuji s předchozí bakalářské práce, kde jsem měl za úkol vytvořit databázi grafoelementů, které mají sloužit pro automatickou detekci jednotlivých artefaktů. Extrahovaná data z této databáze jsou tedy zobrazitelná a jsou kompatibilní s touto diplomovou prací.

Program slouží pouze jako vizuální nástroj s možností editací klasifikace do tříd u konstantní segmentace. Jeho možná rozšíření jsou programování dalších modulů, jako je například filtrace signálu, automatický výpočet adaptivní segmentace a příznaků, výpočet spektrogramu apod.

⁴ Všechny části jsou detailně popsány v následujících podkapitolách.

6.4 Formáty vstupních dat

V 2) jsem uvedl, že pracuji s výstupy programu Ing. Václava Gerly. Tento program (*toolbox*) je realizován ve vývojovém prostředí MATLAB a je třeba některé výstupy převést do tzv. binární podoby (6.4.2), protože samotné výstupní soubory „.mat“ jsou velmi obtížně zpracovatelné (jedná se o vyexportované kanály z jednotlivých záznamů).

Všechna data mají pevnou strukturu, která musí být pro další zpracování dodržena.

6.4.1 XML soubory

Hlavním přístupovým souborem, se kterým se pracuje je tzv. XML soubor (*eXtensible Markup Language*). Jedná se o soubor, který nese informaci o počtu segmentů, vzorkovací frekvenci, počtu elektrod i názvy všech obsažených polysomnografických kanálů. Struktura XML souboru je zobrazena na *obr. 6-1*.

```
<?xml version="1.0"?>
<!-- Written on 18-Nov-2008 17:24:11 using the XML Toolbox for Matlab -->
<root>
  <filename>09201319.asc</filename>
  <number_of_channels>25</number_of_channels>
  <number_of_samples>221272</number_of_samples>
  <sample_frequency>128</sample_frequency>
  <channel_001>MKR1+</channel_001>
  <channel_002>FP1</channel_002>
  <channel_003>FP2</channel_003>
  <channel_004>F7</channel_004>
  <channel_005>F3</channel_005>
  <channel_006>F2</channel_006>
  <channel_007>F4</channel_007>
  <channel_008>F8</channel_008>
  <channel_009>T3</channel_009>
  <channel_010>C3</channel_010>
  <channel_011>C2</channel_011>
  <channel_012>C4</channel_012>
  <channel_013>T4</channel_013>
  <channel_014>T5</channel_014>
  <channel_015>P3</channel_015>
  <channel_016>P2</channel_016>
  <channel_017>P4</channel_017>
  <channel_018>T6</channel_018>
  <channel_019>O1</channel_019>
  <channel_020>O2</channel_020>
  <channel_021>EOG1+</channel_021>
  <channel_022>EMG1+</channel_022>
  <channel_023>ECG1+</channel_023>
  <channel_024>PNG1+</channel_024>
  <channel_025>Mean</channel_025>
</root>
```

obr. 6-1 - struktura XML souboru

6.4.2 BIN - Binární soubory

Jak již bylo zmíněno výše, program pracuje s binárními soubory, nikoliv se soubory přímo z Matlabu, které mají odlišnou strukturu zápisu a také složitý přístup k datům. Binární soubor (binary file) může obsahovat jakákoliv data. V našem případě obsahuje řetězec znaků, tvořící signál k zobrazení. Na rozdíl od textového souboru, nelze soubor editovat ručně.

Pro převod souboru do binární podoby je použita funkce v MATLABU „*fwrite(zapis,soubor.data,'real*4')*“. Kde *zapis* je jméno souboru pro zápis, *soubor.data* soubor s uloženými daty a „*real*4*“ nám značí způsob zápisu do binárního souboru, konkrétně se jedná o reálná 32bitová čísla typu *float* [12].

6.4.3 TXT soubory

Soubory TXT (*text file*) jejich kódování je ASCII (*American Standard Code for Information Interchange*), tedy jeden byte odpovídá jednomu znaku v souboru. Tyto soubory nesou veškeré ostatní informace:

- a) *segments-XXX.txt* – nese informaci o výsledcích konstantní, nebo adaptivní segmentace pro daný kanál XXX
- b) *data-expert-XXX.txt* – pro každý segment je v tomto souboru uvedena třída, do které daný segment patří (ohodnoceno lékařem)
- c) *features-XXX.txt* – výsledky extrakce příznaků
- d) *data-classifier-XXX.txt* – automaticky oklasifikovaná data, možnost editace.
- e) *hierarchical-clusters-XXX.txt* – výsledky shlukové analýzy
- f) *spectrogram-welch-chXXX.txt* – data potřebná pro zpracování spektrogramu metodou „Welch“.
- g) *spectrogram-normalized-chXXX.txt* – data potřebná pro zobrazení normalizovaného spektrogramu.

6.4.4 ARFF soubory

Soubory ARFF obsahují informace o jednotlivých kanálech (vyextrahované příznaky pro jednotlivé segmenty a klasifikace lékařem). Soubor je typu ASCII (6.4.3) a je rozdělen do dvou částí. První část nese názvy jednotlivých kanálů s názvem měřeného příznaku. Druhá část pak nese samotná data. Důležitou informací, která je použita pro rozlišení dat od názvů jsou první atributy každého řádku. Pro popisky “@ATTRIBUTE” pro data “@DATA”. Na obr. 6-2 je zobrazen začátek souboru s atributy a na obr. 6-3 začátek bloku s daty.

```
@ATTRIBUTE ch_FP1---mean REAL
@ATTRIBUTE ch_FP1---std REAL
@ATTRIBUTE ch_FP1---skewness REAL
@ATTRIBUTE ch_FP1---kurtosis REAL
@ATTRIBUTE ch_FP1---1st_diff_mean REAL
@ATTRIBUTE ch_FP1---1st_diff_max REAL
@ATTRIBUTE ch_FP1---2st_diff_mean REAL
@ATTRIBUTE ch_FP1---2st_diff_max REAL
```

obr. 6-2 – názvy příznaků

```
@DATA
0.692 0.246 0.530 0.005 0.30
0.495 0.746 0.077 0.122 0.07
0.088 0.097 0.944 0.688 0.24
0.521 0.598 0.569 0.514 0.74
0.052 0.050 0.114 0.094 0.10
0.513 0.717 0.507 0.556 0.56
0.402 0.077 0.020 0.052 0.06
```

obr. 6-3 – začátek bloku s daty

6.5 Výstupní data

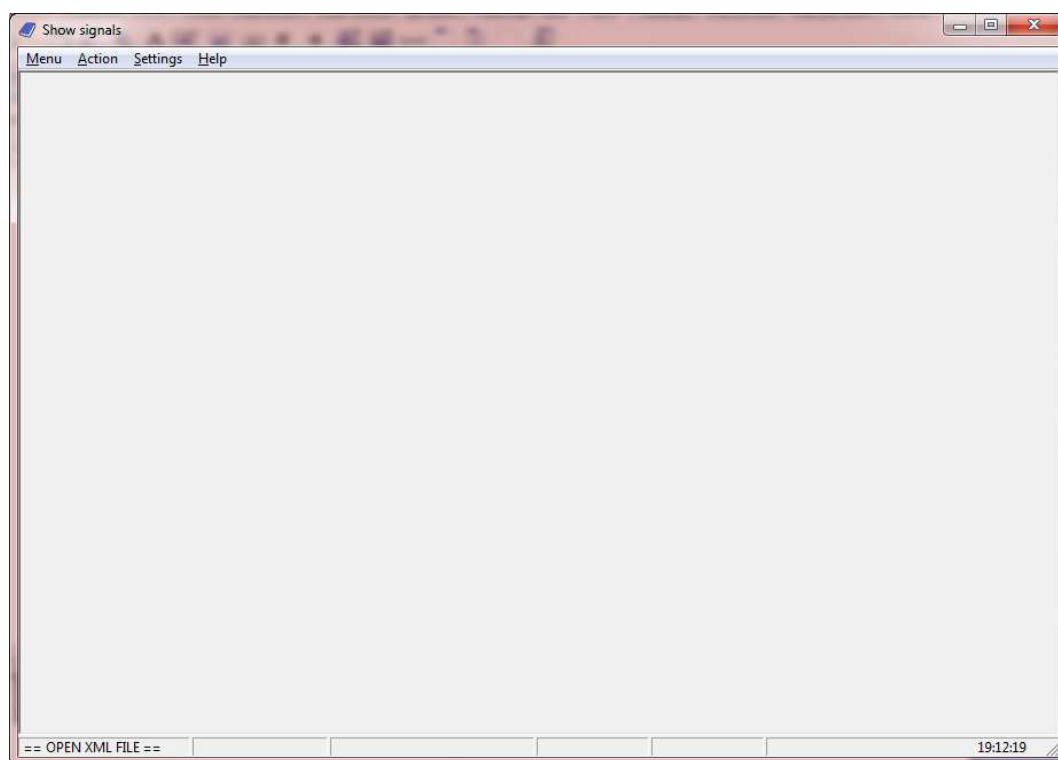
Mezi výstupní data programu patří zejména INI soubor, který nese veškeré informace o posledních změnách aplikace (změna pozice okna, velikost okna a další nastavení). Soubor se standardně ukládá do registrů operačního systému. Dalším výstupním souborem je případný TXT (6.4.3) soubor, vytvořený pro konstantní segmentaci a její třídy. Posledním výstupním souborem je obrázkový typ BMP (*Microsoft Windows Bitmap*).

6.6 Popis aplikace

Jak bylo zmíněno výše, kladl jsem důraz zejména na uživatelskou přehlednost. Po prvním spuštění a opuštění aplikace je zaznamenána aktuální pozice okna a při dalším spuštění bude hlavní okno zobrazeno na stejné pozici. Protože dat ke zpracování je spousta, zvolil jsem volbu postupného výběru funkcí, aby nedocházelo k nepřehlednosti aplikace. V programu nechybí podrobná nápověda pro první kroky a několik užitečných nastavení programu.

6.6.1 Hlavní část programu

Po spuštění aplikace se zobrazí prázdné hlavní okno, které obsahuje pouze roletové menu a stavový řádek. Viz *obr. 6-4*.



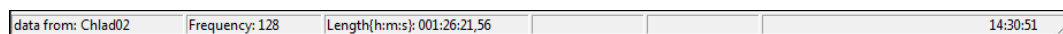
obr. 6-4 - hlavní okno

Načítání XML

Pro načítání jednotlivých kanálů (elektrod) je potřeba vybrat z roletového menu „Open XML...“, po té vybereme soubor. Po načtení XML dochází k načtení základních informací o měřených datech, jedná se o:

- délku měřeného signálu (počet segmentů)
- vzorkovací frekvence
- názvy elektrod
- název původního souboru
- počet kanálů

Ve stavovém řádku se pak tyto načtené informace zobrazí a všechny realizované metody pracují právě s těmito daty. Stavový řádek je uveden na *obr. 6-5*.



obr. 6-5 - stavový panel

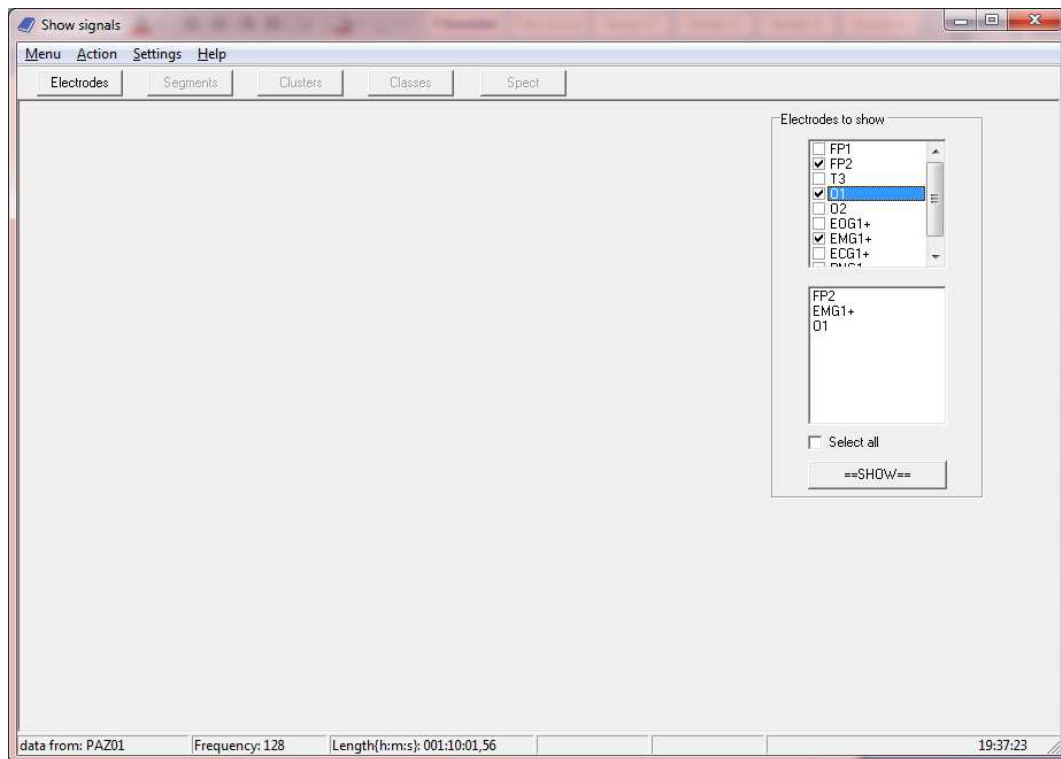
Načítání XML

XML Soubor obsahuje názvy všech elektrod, včetně těch, které nebyly dosud zpracovány. Proto bylo nutno pro další práci tyto kanály odfiltrovat. To jsem provedl pomocí funkce na ověřování existence souborů „channel-xxx.bin“. Pokud takový soubor existuje, načte se elektroda pro další výběr do seznamu. Současně s těmito daty se načítají výsledky adaptivní segmentace pro daný signál. Opět se testuje, jestli soubor se segmentací existuje.

Při výběru XML Souboru se odkryje podmenu (tvořeno tlačítky) pro práci s daty a menu pro výběr elektrod. V podmenu jsou tlačítka aktivní, jen pokud jsou načtena potřebná data pro zobrazení.

Vybraná elektroda k zobrazení se přidá do seznamu pro zobrazení. Data jsou dále zpracovávána ve stejném pořadí, jako jsou zobrazena. Tlačítko pro zobrazení „==SHOW==“ je zobrazeno za podmínky, že je vybrán minimálně jeden kanál pro zobrazení.

Pokud si elektrodu nepřejeme zobrazit stačí ji pouze odškrtnout z výběru a elektroda zmizí také ze seznamu. Seznam je tvořen pevnou strukturou dat – nelze s vybranými kanály dále pohybovat a měnit pořadí. Pořadí určuje postupné zaškrtování políček. Tento krok byl zvolen kvůli jednoznačnému pořadí a zobrazení zvolených kanálů a nedocházelo tak ke zbytečnému prohazování už jednou vytvořenému seznamu dat. Na obr. 6-6 je zobrazen příklad výběru tří elektrod.



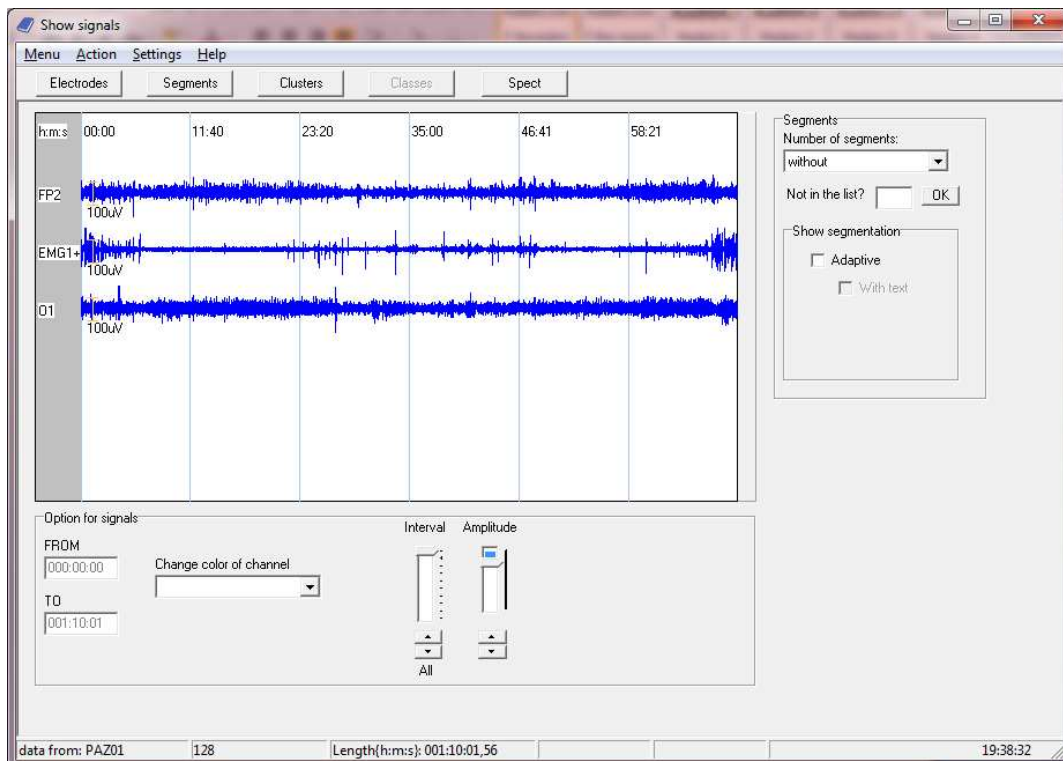
obr. 6-6 - načtené kanály

6.6.2 Zobrazení signálů

Po zvolení signálů k zobrazení a potvrzením tlačítkem „**==SHOW==**“ probíhá načítání jednotlivých kanálů, segmentů. Toto načítání je indikováno zobrazením tzv. loading baru. Dále je volána funkce pro vykreslení vybraných elektrod, které se předají veškeré potřebné údaje pro zobrazení, jako jsou naměřené hodnoty na jednotlivých elektrodách, počet vzorků, vzorkovací frekvence, načtené segmenty. Při zavolání funkce „vykresli“, je zobrazen loading bar ve stavovém řádku. Nutno podotknout, že čas pro vykreslení signálu je závislý na délce vykreslovaného signálu.

Jako první je vykreslen celý signál a aktivují se tlačítka v podmenu a odkrývá se další skryté menu „Option for signals“, které doposud nebylo potřeba zobrazovat. Viz *obr. 6-7*.

Skryté menu slouží pro práci s vykresleným signálem. Editační políčka jsou zde pouze informativní, tlačítka „*start*“ a „*end*“ slouží k nastavení na začátek a konec signálu, ostatní komponenty budou popsány v příslušných podkapitolách.



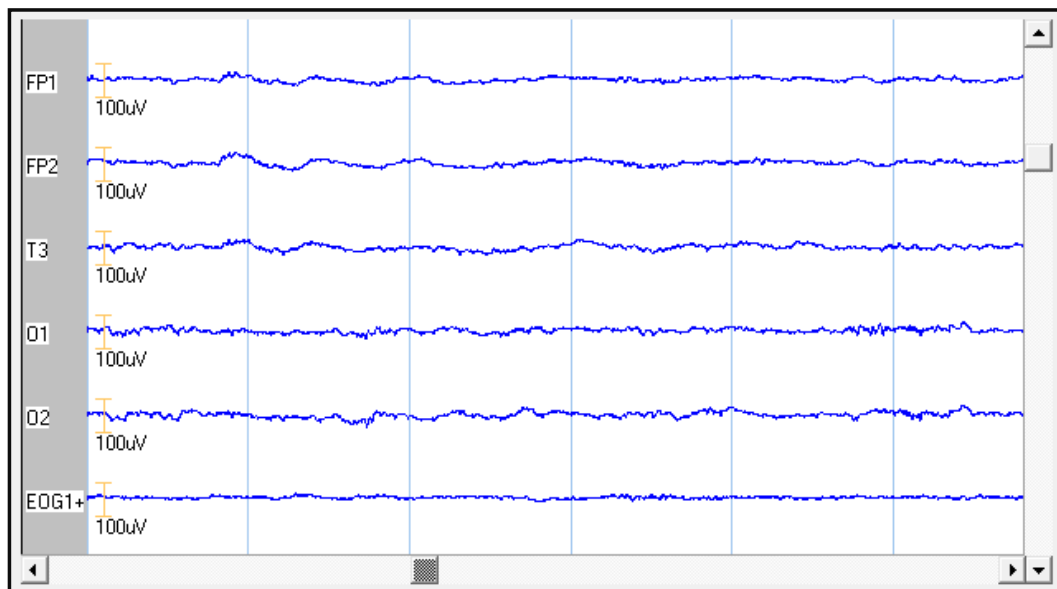
obr. 6-7 - základní rozložení, zobrazení signálu

Panel se signálem

Hlavní částí tohoto okna je prostor pro vykreslení naměřených signálů. Je vložen do samostatného panelu s vykreslovacím plátnem. Plátno je potřeba překreslovat. Aby bylo zabráněno zdlouhavému načítání dat a pak samotnému vykreslování po elektrodách, je pomocí funkce `TMainForm::vykresli(TCanvas Canvas)`, uloženo plátno do pomocné bitmapové proměnné. Tímto způsobem je mnohem snazší a rychlejší vykreslit plátno při jakémkoliv překreslování okna. Do pomocného bitmapového souboru ukládám po jakémkoliv změně ve vykreslování (změny měřítka, barvy signálů). Nejprve jsou vypsány názvy elektrod, poté plátno a nakonec veškeré signály. Časová osa na plátně je rozdělena konstantně, a záleží na délce signálu při vykreslení. Časová osa je rozdělena na hodiny, minuty a vteřiny, kdy hodiny se vypíší jenom v případě, že záznam je na pozici jedné hodiny a více. Formát času tedy bude buď `hh:mm:ss` nebo `mm:ss`. Kde `h,m,s` jsou celá čísla od 0 do 9.

Posuvníky

Protože signál může být dlouhý až několik hodin, je třeba ho rozdělit a zobrazovat po částech, proto je vložen posuvník horizontální, kterým se posunuje v signálu o určité úseky. Taktéž je vložen posuvník vertikální, ale ten je zobrazen pouze pokud je zobrazeno více signálů. Viz *obr. 6-8*.



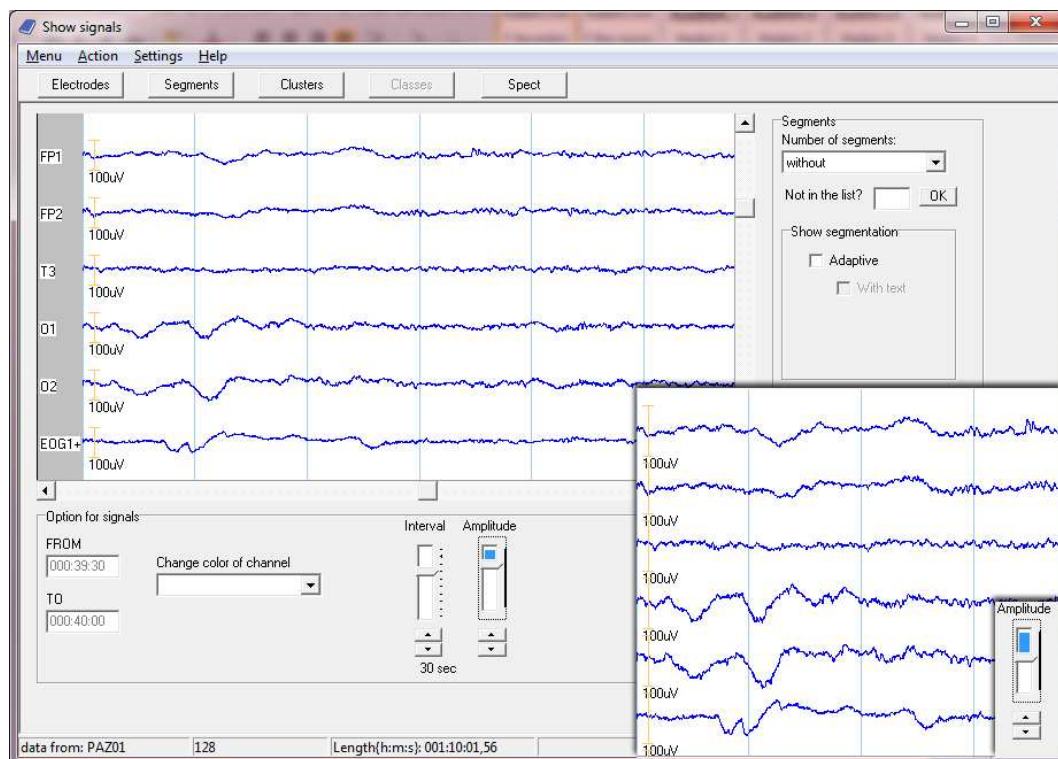
obr. 6-8 - zobrazení vertikálního posuvníku

Editační políčka „FROM“ a „TO“

Slouží pouze jako informační údaje pro zobrazení časové pozice odkud a kam je vykreslený signál zobrazen.

Změna amplitudy

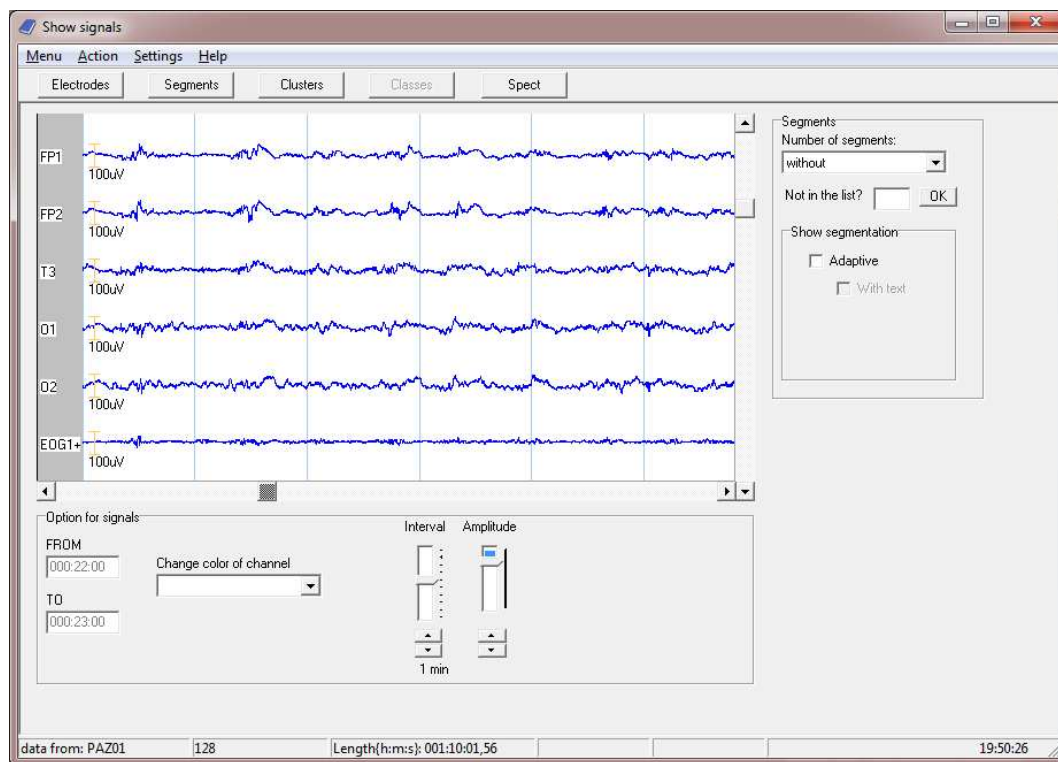
Změna amplitudy signálu je tvořena posuvníkem „*amplitude*“, Maximální hodnota je nastavena fixně na 100 a minimum na 0. Změnou pozice se předávají informace pro vykreslení. Příklad změny amplitudy je uvedena na *obr. 6-9*.



obr. 6-9 - změna amplitudy

Změna velikosti „Interval“

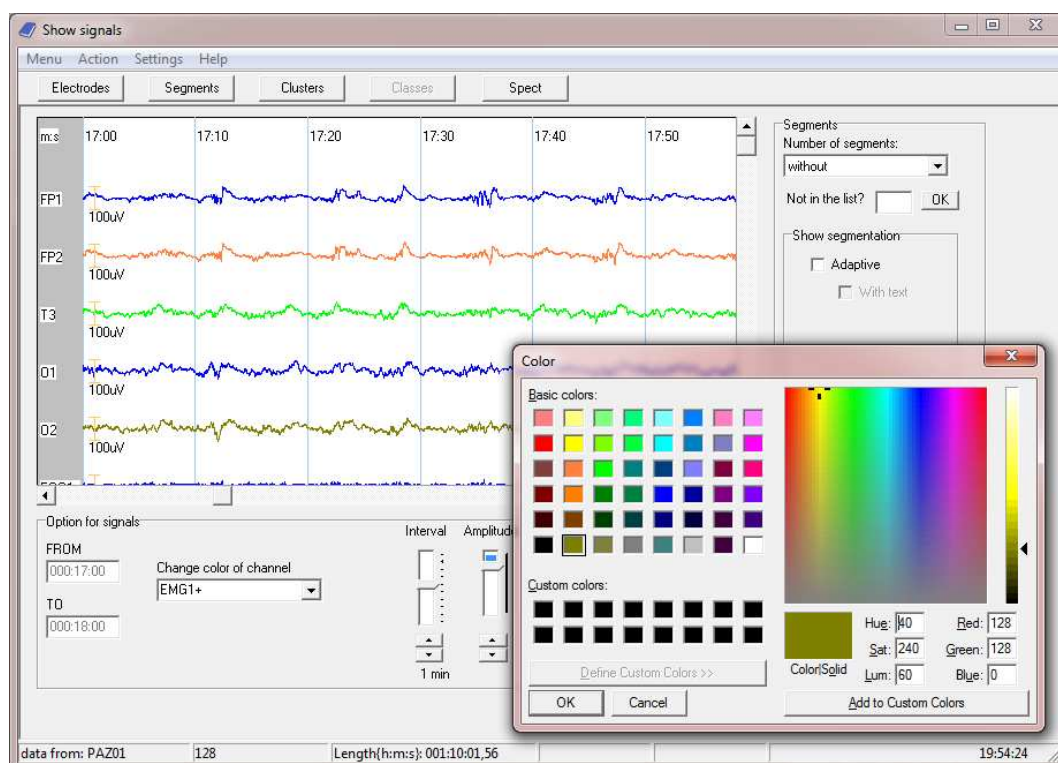
Změna velikosti signálu jde změnit více způsoby. Jedním z nich je tzv. „Interval“. Posuvníkem „Interval“ lze měnit časovou osu a délku vykreslovaného signálu. Interval je rozdělen do 10. stupňů: 5 sec, 10 sec, 30 sec, 1 min, 5 min, 10 min, 20 min, 30 min, 1 hod a celý signál. Podle zvoleného intervalu se také mění vykreslovaná časová osa, která je rozdělena konstantně na 5 až 6 dílků (vykreslujeme-li signál například po 1 minutě, je časová osa rozdělena po 10 vteřinách). Příklad zobrazení po intervalech je zobrazen na obr. 6-10.



obr. 6-10 - příklad změny intervalu zobrazení

Změna barvy signálu

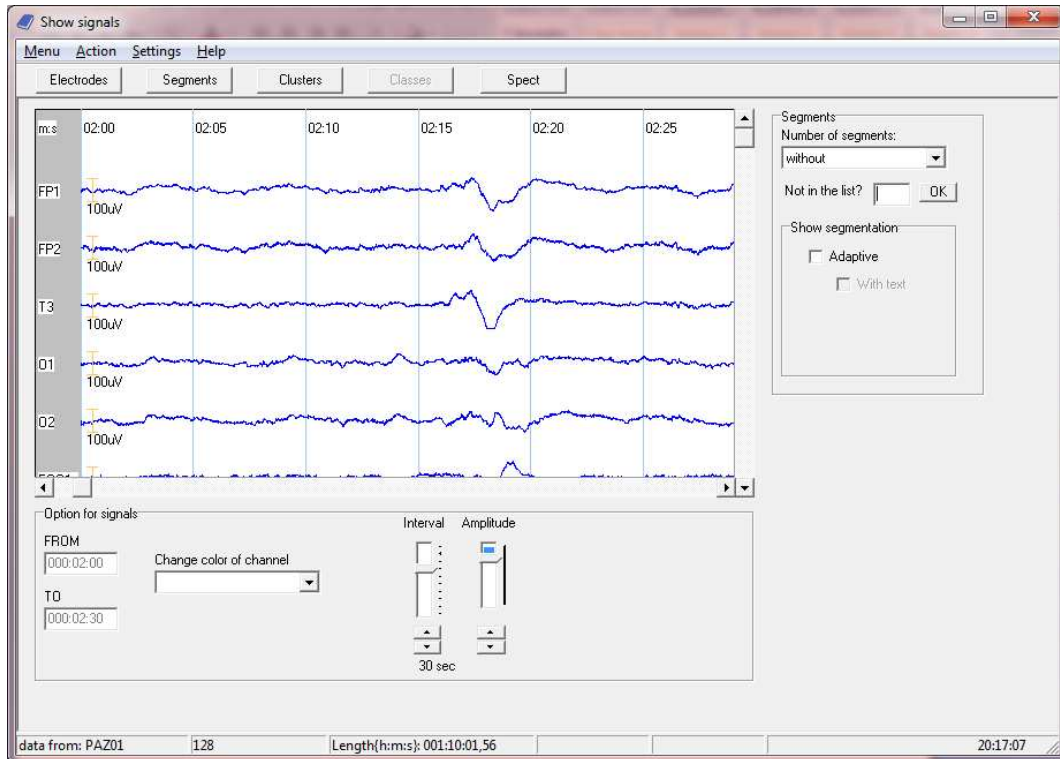
Pokud je potřeba zvýraznit nebo odlišit signál od ostatních, je zde přidána možnost volby barvy pro jednotlivé signály. Slouží k tomu výběrové menu s popisem „Change color of channel“. Z roletového menu se po výběru kanálu (elektrody) volá funkce pro změnu barvy. Tato funkce zavolá dialogové okno *Color*, po výběru barvy je barva pera uložena do proměnné *barvyPera* a následně se volá funkce pro vykreslení signálu, kde se použije právě vybraná barva. Příklad změny barvy je uvedena na *obr. 6-11*.



obr. 6-11 - změna barev signálů

6.6.3 Zobrazení segmentace

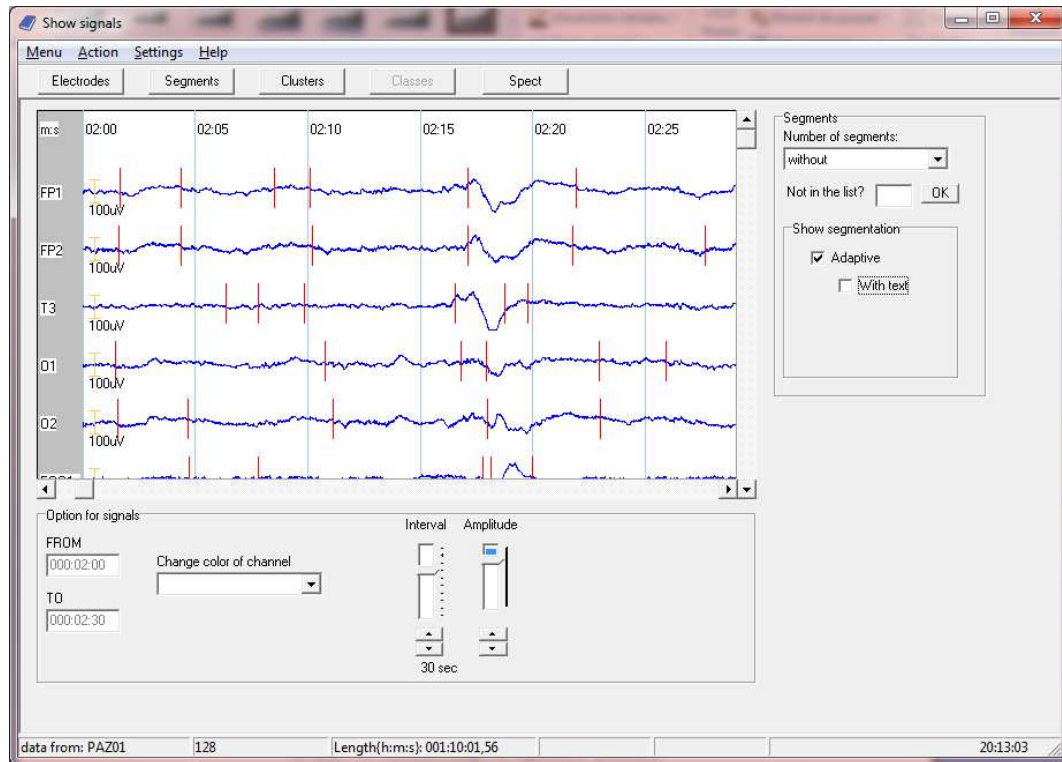
Pro další práci se signálem je nutno načítat výsledky segmentace. Jak již bylo zmíněno v sekci „Předzpracování EEG signálu“, načítáme segmentace adaptivní a konstantní. Při prvním načítání se menu se segmentacemi zobrazí automaticky, jinak je potřeba přepnout v podmenu pomocí tlačítka „Segments“. Po přepnutí se skryje nabídka, která je právě zobrazena. Viz. obr. 6-12.



obr. 6-12 – zobrazení menu „Segments“

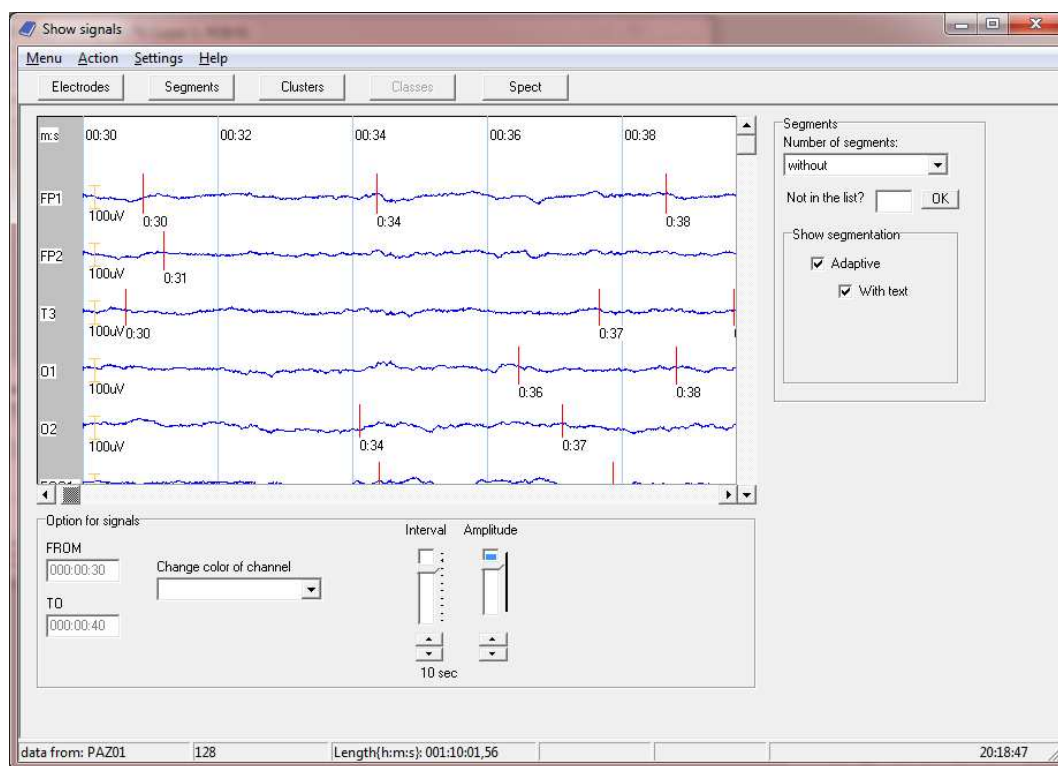
Adaptivní segmentace

Adaptivní segmentace je načítána zároveň s daty jednotlivých kanálů. Pokud však nebyla pro daný signál segmentace vypočtena je automaticky tato segmentace vynechána. Zobrazení segmentace je v nabídce „Show segmentation“. Na výběr jsou prvně dvě zaškrťovací políčka, jedno pro samotné zobrazení (obr. 6-13)

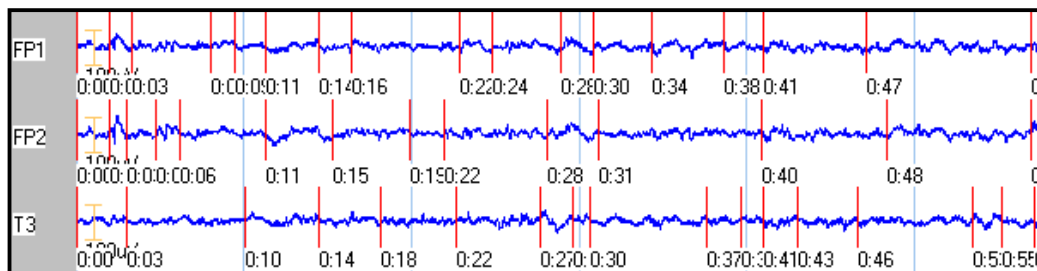


obr. 6-13 - Adaptivní segmentace, bez popisku

a druhé pro přesné určení pozice vzorku, kde končí a začíná nový segment pomocí popisku v signálu (*obr. 6-14*). Tato možnost zobrazení s/bez textu, je zde přidána díky větší přehlednosti v samotném signálu. Pokud je zobrazen větší úsek signálů dochází k nečitelnosti a nepřehlednosti popisků a je tedy výhodnější toto zobrazení vypnout viz *obr. 6-15*.



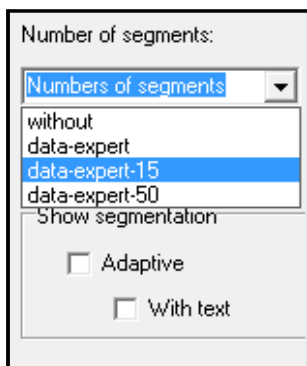
obr. 6-14 - zobrazení adaptivní segmentace - přehledně



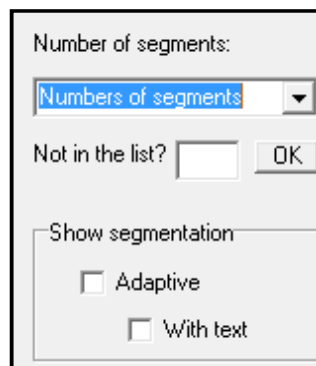
obr. 6-15 - zobrazení adaptivní segmentace - nepřehledně

Konstantní segmentace

Na rozdíl od segmentace adaptivní se výsledky segmentace konstantní načítají postupně. Nejdříve jsou v adresáři nalezeny soubory „*data-expert.txt*“, „*data-classifier-XXX.txt*“ a „*features-final-XXX.arff*“, kde XXX tvoří číslo počtu segmentů. Pro načítání konstantní segmentace je zde výběrové menu „*number of segments*“. Je zde navíc přidáno editační políčko pro upřesnění čísla XXX. (*výběr segmentace - obr. 6-16, roletové menu s editačním políčkem - obr. 6-17*). Tato možnost zde byla zařazena z důvodů přesnější volby souboru. Vložená hodnota se opět kontroluje, zda je číselná.

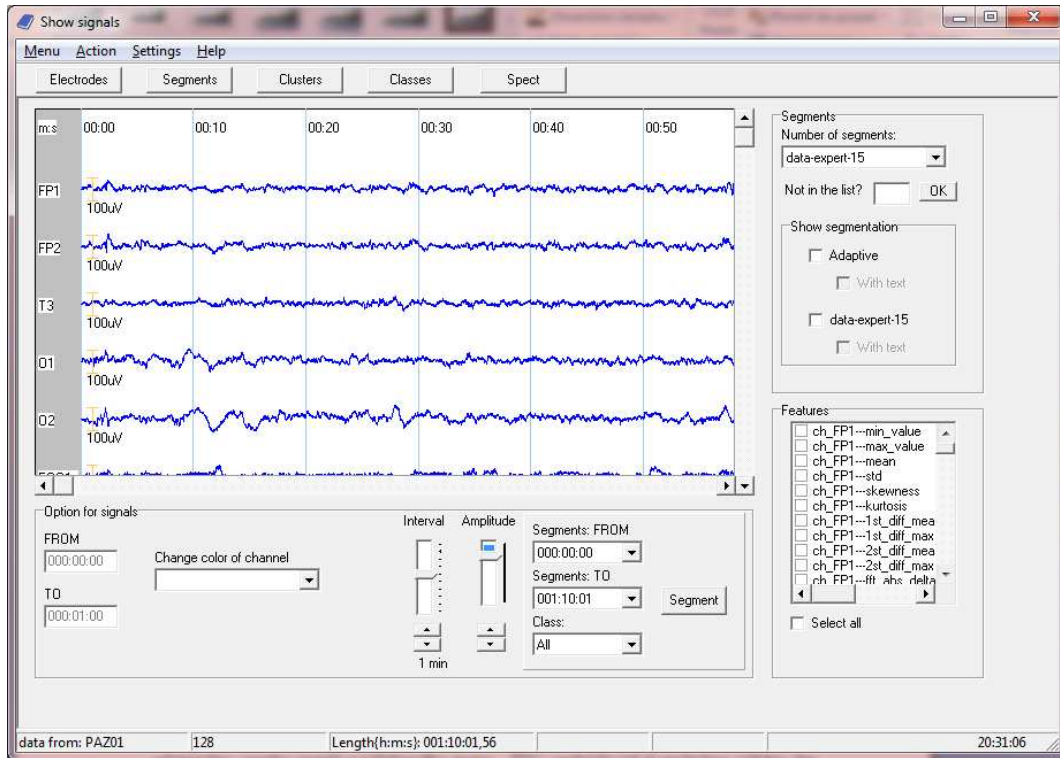


obr. 6-16 - výběr segmentace 1



obr. 6-17 - výběr segmentace 2

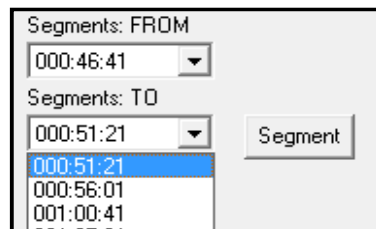
Po výběru typu segmentace jsou odhaleny skryté objekty, které jsou bez konstantní segmentace nepotřebné. (obr. 6-18).



obr. 6-18 - odkryté objekty

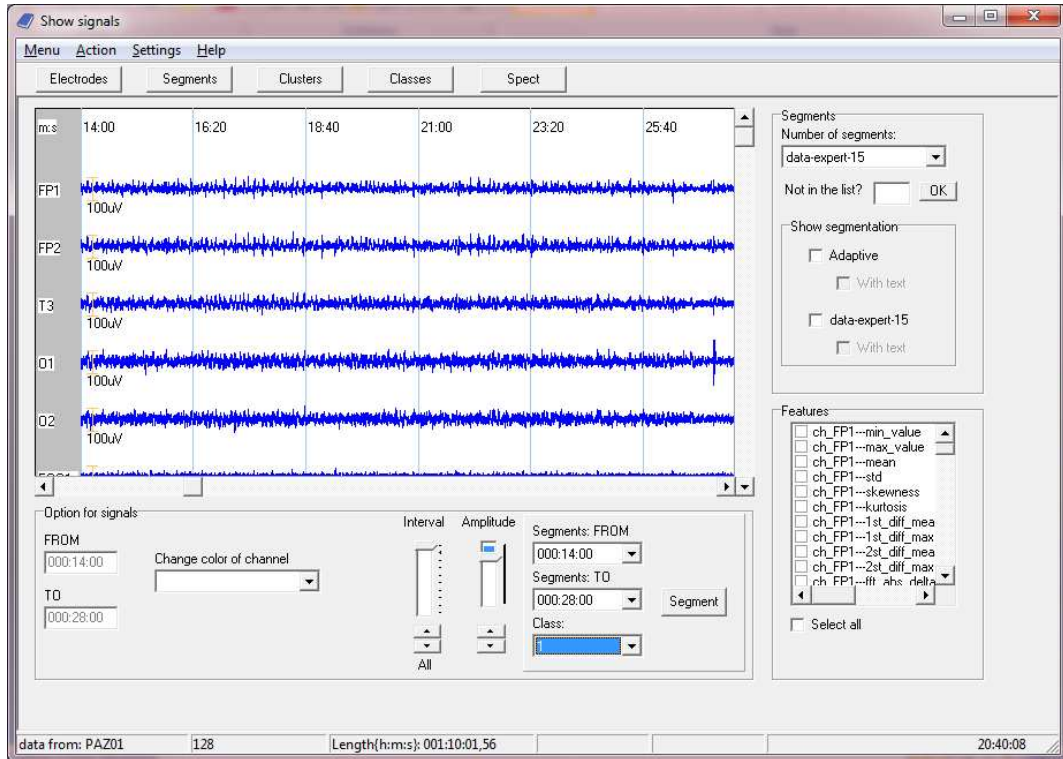
Classified segments FROM, TO, Class

Načtená data jsou pro snazší práci načtena do výběrového menu, odkud je možnost zobrazovat jednotlivé segmenty. Při výběru vzorku z „FROM“, jsou automaticky odstraněny vzorky menší z výběrového menu „TO“, nedojde tak k mylnému výběru dat. Také rozdělení do tříd filtruje příslušná data do výběrových polí „FROM, TO“ (obr. 6-19).



obr. 6-19 - výběr segmentů

Příklad výběru třídy 1 a zobrazení bloku dat, je znázorněno na *obr. 6-20*.

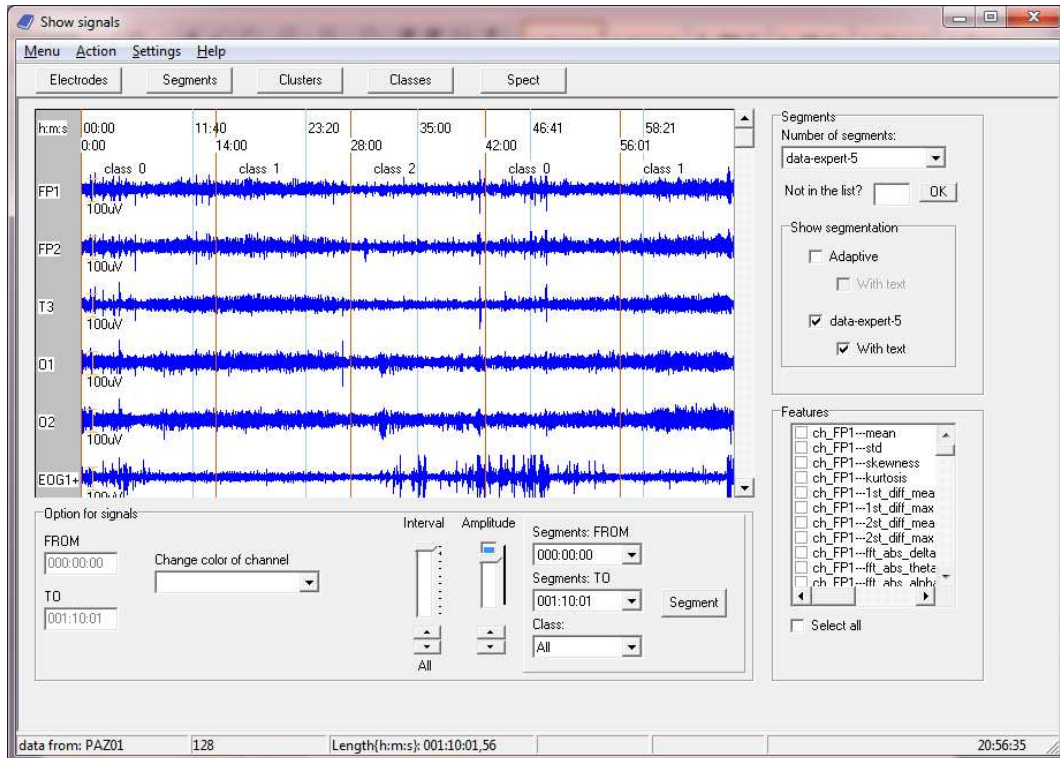


obr. 6-20 - výběr bloku vzorků

Zobrazení segmentace

Při načtení konstantní segmentace se zobrazí dodatečná zaškrťovací políčka pro segmentaci s nebo bez popisů. Také se mění jméno popisku, podle vybraného typu segmentace, aby bylo zřejmé, který soubor je používán. Protože je segmentace konstantní, platí tak pro veškeré kanály stejné segmenty.

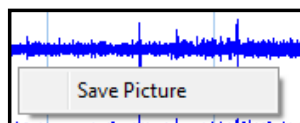
Pro vykreslení bloků byla použita stejná forma vykreslování, jako u časové osy. Pokud je vybrána možnost vykreslování s popisem, jsou také zobrazeny jednotlivé třídy segmentů. Viz. *obr. 6-21*.



obr. 6-21 - výběr zobrazení segmentace s popisem

Uložení vykresleného signálu

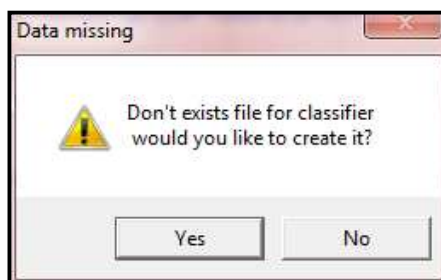
Při přejezdu myší přes vykreslený signál a vyvoláním „popup menu“ pravým stiskem tlačítka myši se vyvolá nabídka menu. Pro uložení zobrazených informací slouží „Save picture“. Obrázek je uložen jako BMP (6.5). Pro uložení se zobrazí dialogové okno pro výběr místa k uložení, obr. 6-22.



obr. 6-22 - PopupMenu

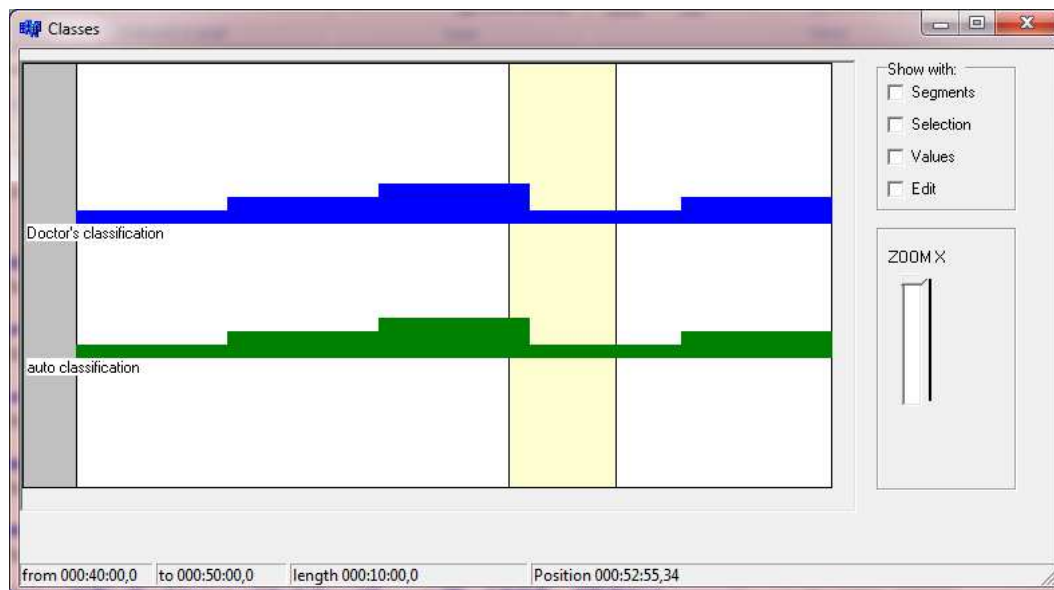
6.6.4 Práce s klasifikovanými soubory

Při výběru konstantní segmentace se automaticky načítají také klasifikační metody, pokud existují. A to jak lékařská diagnóza, tak automatická diagnóza. Pro zobrazení klasifikace slouží tlačítko z podmenu „Classes“. Tlačítko se aktivuje, pokud existuje alespoň lékařská diagnóza, při nepřítomnosti automatické klasifikace vyskočí okno (obr. 6-23) s dotazem, jestli si přejeme tuto klasifikaci vytvořit. Vytvořena klasifikace je kopie klasifikace lékařské.



obr. 6-23 - chybí automatická klasifikace

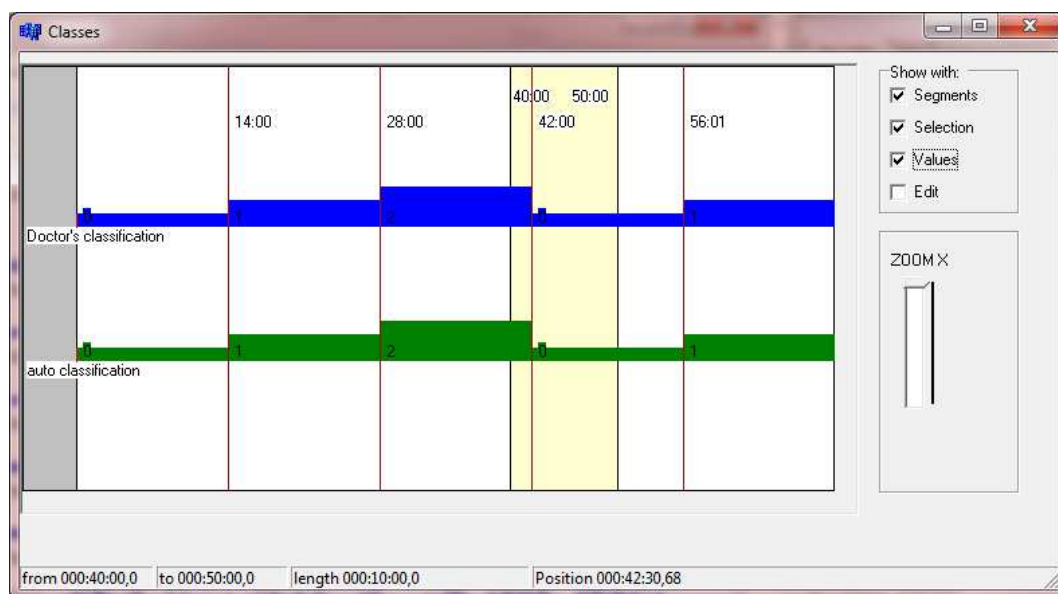
Jsou-li data v pořádku načtena, volá se funkce pro vytvoření okna s oklasifikovanými segmenty. Vygenerované okno je uvedeno na obr. 6-24. Okno obsahuje Statusbar s informacemi o signálu (pozice v signálu, začátek a konec vykresleného signálu, celková délka zobrazeného signálu v hlavní okně), menu se skrytými položkami, ZOOM X a samotnou klasifikaci.



obr. 6-24 – zobrazení klasifikace

Zobrazená data

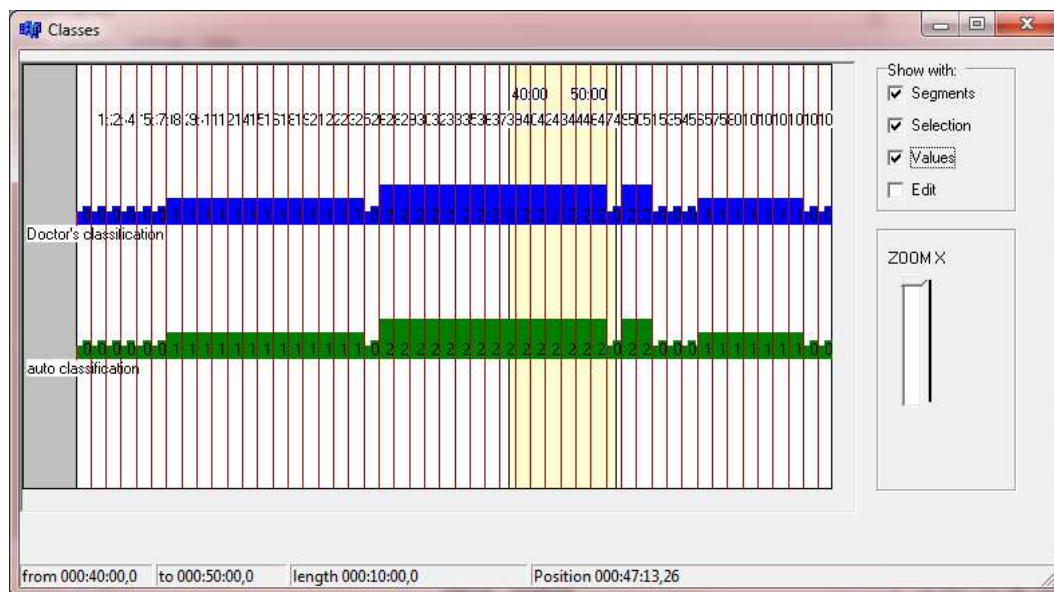
Nejprve je vykreslena lékařská klasifikace (modrý signál) a poté klasifikace automatická (zelený signál). Při výběru z postranního menu je možnost zobrazit data včetně popisků, rozdělení segmentů, výpis pozice v hlavním okně signálů a pak samotná editace. Na *obr. 6-25* jsou zobrazeny veškeré informace o vykresleném signálu.



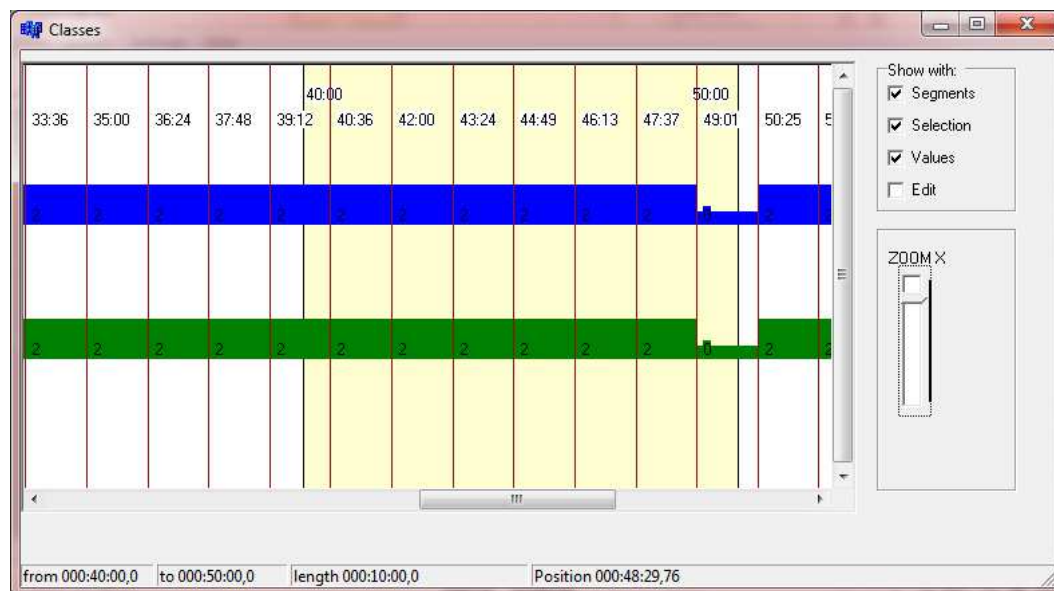
obr. 6-25 - zobrazená data

ZOOM X

Při větším počtu segmentů a výpisem informací v signálu (pozice, časová osa...) se ztrácí přehlednost. viz. obr. 6-26. Z toho důvodu je zde možnost zvětšení měřítka na vykreslovaném plátně (ZOOM X). Při inicializaci okna je vždy vykreslen celý signál. Posuvníkem se pak velikost vykreslovacího plátna mění obr. 6-27.



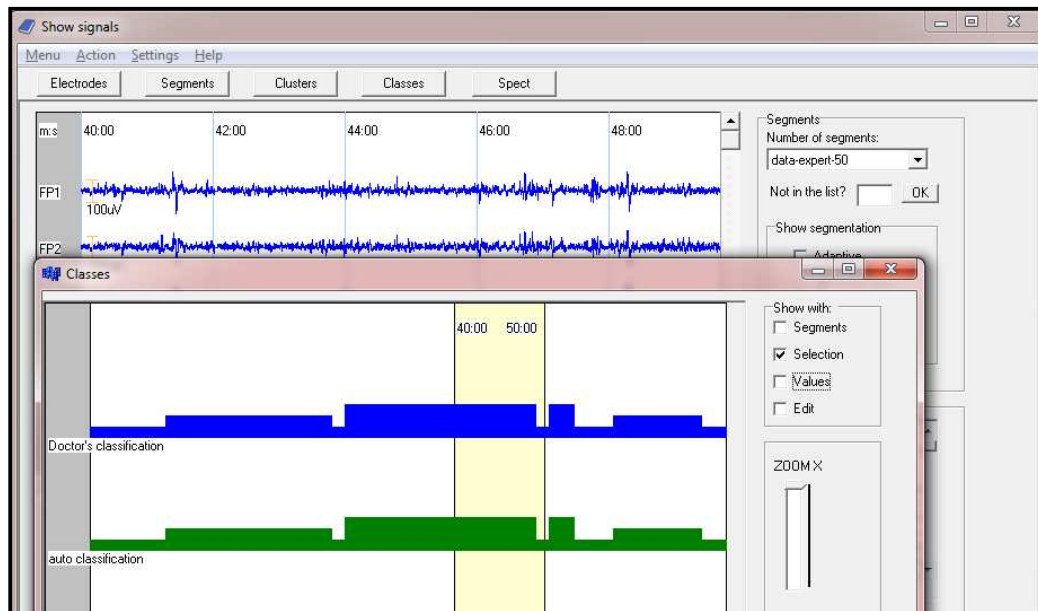
obr. 6-26 – nepřehledné zobrazení



obr. 6-27 – použitý ZOOM – přehledné zobrazení

Pozice v signálu

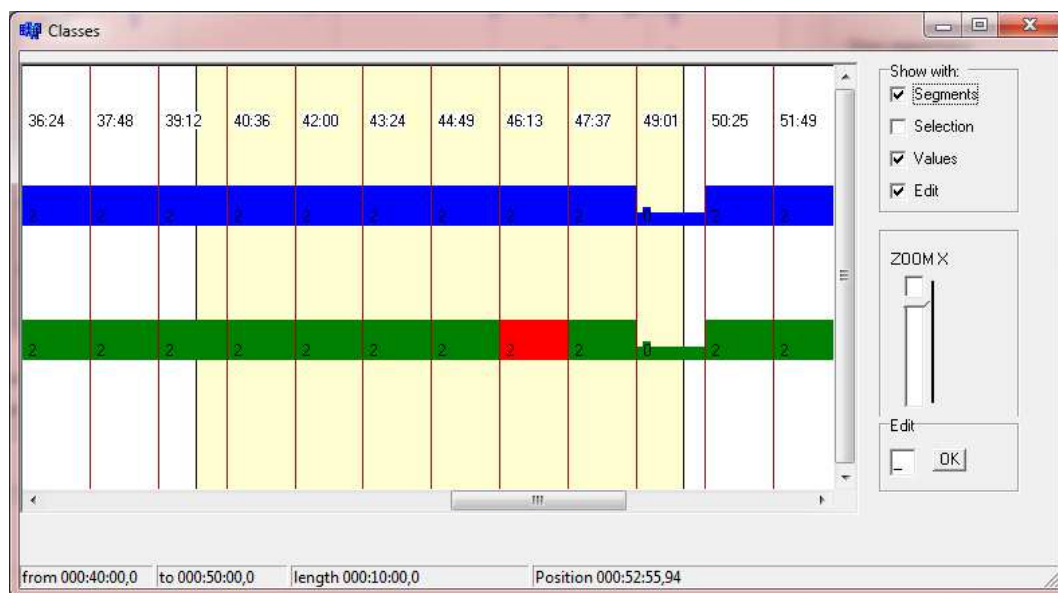
Pro orientaci v signálu a aktuální pozici, je vykreslováno obdélníkové označení pozice signálu. Na *obr. 6-28* je uvedena pozice v signálu a jeho vykreslení v klasifikaci. Označení se mění jakoukoliv změnou pozice v hlavním okně.



obr. 6-28 - aktuální pozice signál

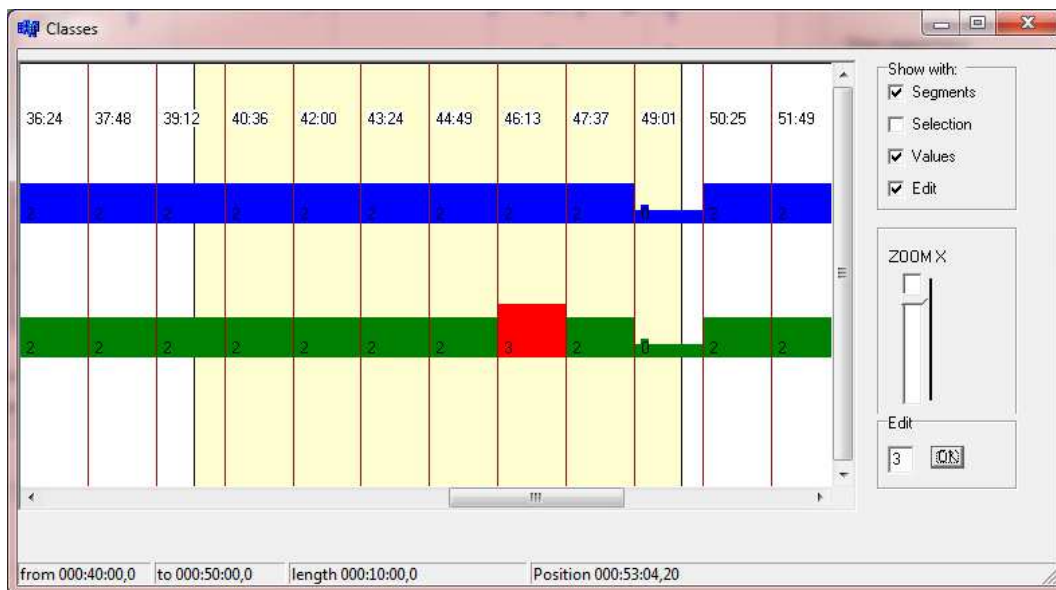
Editace klasifikace

Při výběru z menu „*Edit*“ se zobrazí doposud nepoužívané skryté menu pro editaci signálu. Editovat nelze lékařskou klasifikaci, ale pouze automatickou klasifikaci. Klasifikace, která má být změněna vybereme pomocí pozice v signálu myší, označená část je zvýrazněna červeně viz. *obr. 6-29*.



obr. 6-29 – editace signálu

Při editaci signálu pak už stačí zadat pouze hodnota do editačního políčka a potvrdit tlačítkem „*OK*“. Provedená změna je ihned zobrazena v signálu. Zadávaná hodnota je kontrolována, musí být zadáno číslo, a to v rozmezí 0-9, jakákoliv jiná hodnota je zakázaná. Viz. *obr. 6-30*.

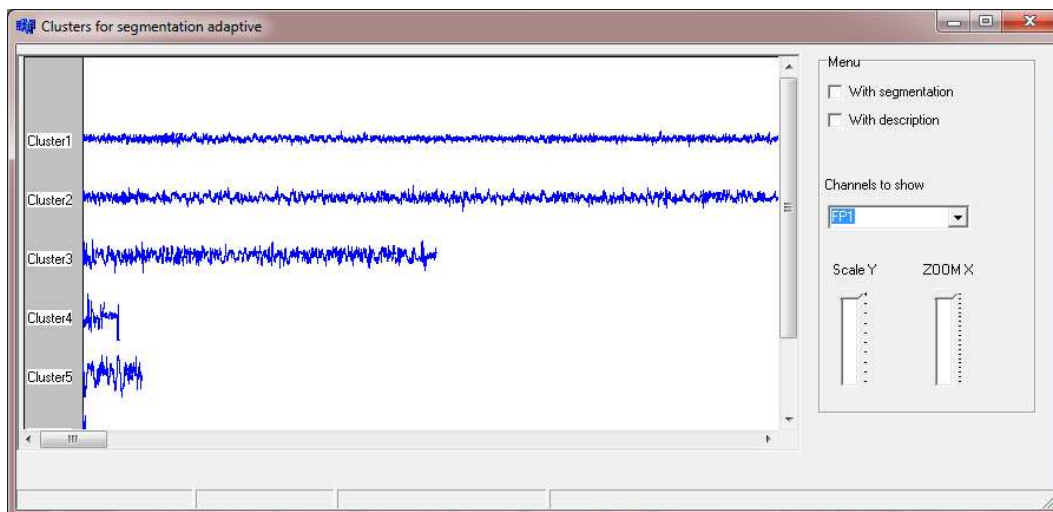


obr. 6-30 – editace signálu změna

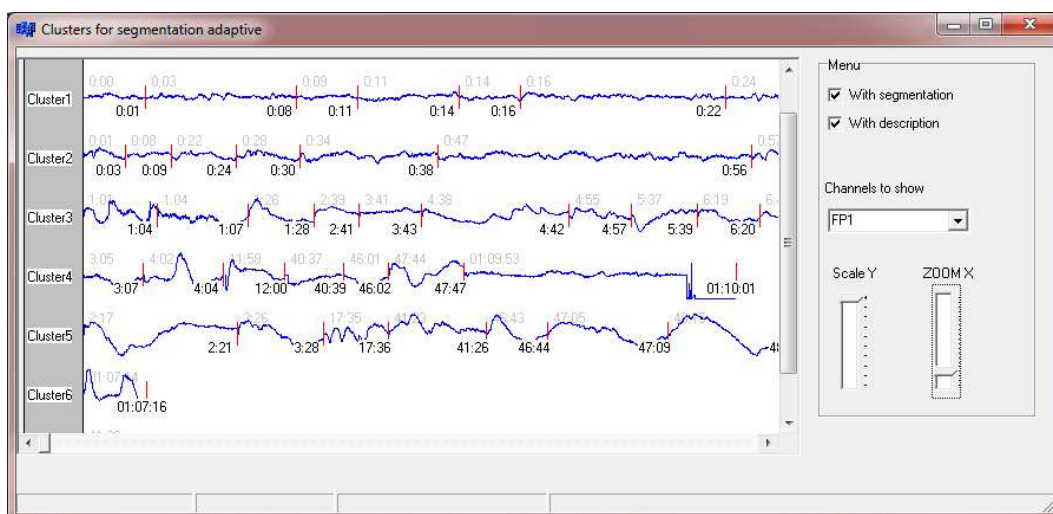
6.6.5 Výsledky shlukové analýzy

Výsledky shlukové analýzy jsou zobrazeny stejně jako samotná klasifikace dat do tříd. Shluková analýza pro signál se nachází v souboru „*hierarchical-clusters-XXX.txt*“. Soubory, pokud existují, jsou načteny už při samotném načítání kanálů. Pro zobrazení výsledků shlukové analýzy slouží tlačítko z podmenu „*Clusters*“. Stav vykreslení je opět indikován loading barem ve stavovém řádku.

Pokud shluková analýza pro nějaký signál neexistuje, signál se nevykreslí. Princip ovládání je stejný jako u zobrazování oklasifikovaných dat. přibylo pouze výběrové menu kanálů pro zobrazení z nabídky „*Channels to show*“. Posuvníky „*scale Y*“ a „*ZOOM X*“ slouží pro lepší orientaci ve shlucích. Zaškrťovací políčka „*with segmentation*“ zobrazuje začátek a konec segmentu a jeho shluk a „*with description*“ pak vypíše na jaké pozici se shluk nachází. Výsledek jde opět pomocí „*popUp*“ menu uložit do bitmapového souboru. Příklad zobrazeného výsledku shlukové analýzy je na obr. 6-32 a přehledné zobrazení na obr. 6-31.



obr. 6-32 – Clusters – bez použití ZOOM X

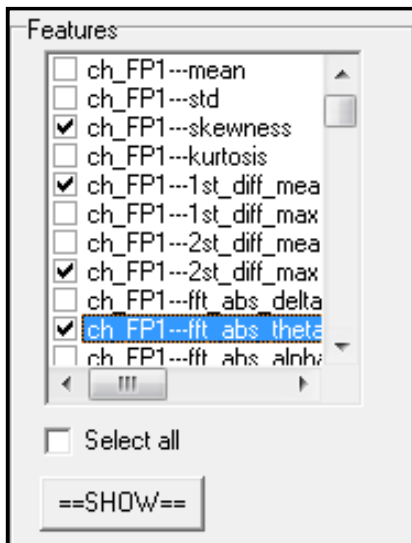


obr. 6-31 – Clusters – použití ZOOM X

6.6.6 Zobrazení příznaků

Při zobrazení bloků se segmentací je zároveň načítáno pole s dostupnými příznaky k zobrazení. Příznaky jsou načítány ze souboru „*features-final.arff*“ a „*features-final-XXX.arff*“, kde XXX je počet segmentů. Jak bylo zmíněno v „6.4.4. ARFF soubory“ soubor obsahuje jak popisky jednotlivých příznaků, tak samotná data. Veškerá data jsou načítána ve funkci *TStringList**⁵ *nactiPriznaky(AnsiString priznakyOdkud){...}*. Funkce otevře vybraný soubor, a po řádcích načítá data. Pokud řádek obsahuje *@ATTRIBUTE* načte data do seznamu příznaků. Pokud narazí na řádek s *@DATA*, přestane načítat příznaky a načtou se data. Také jsou načteny atributy třídy *@ATTRIBUTE Class*.

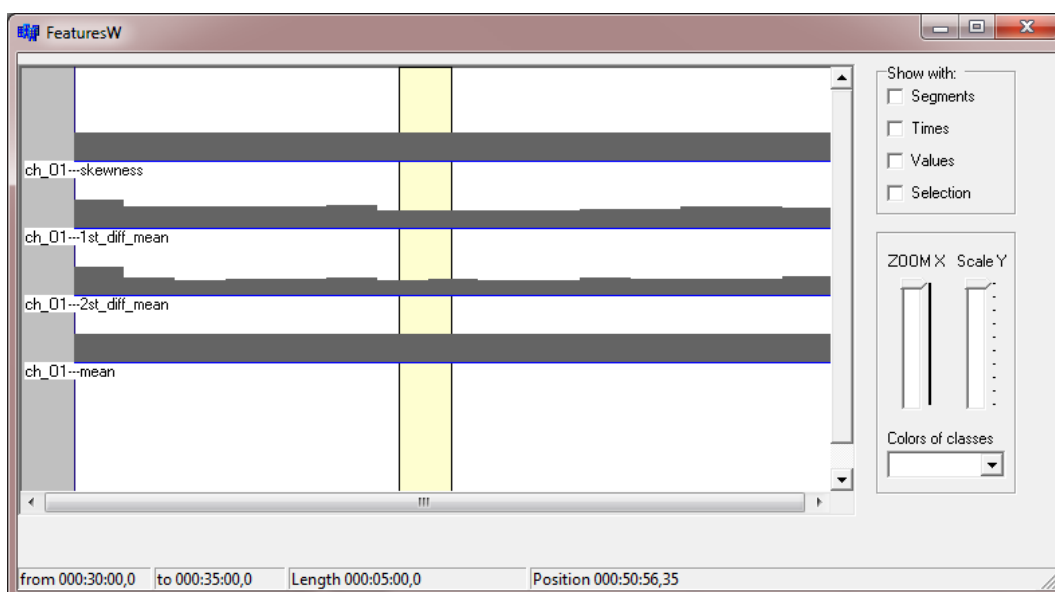
Další funkce zkontroluje přítomnost jednotlivých kanálů a vrátí aktuální seznam příznaků pro výpis do výběrového pole. Ve výběrovém poli po výběru minimálně jednoho příznaku, se zobrazí tlačítko pro zobrazení vybraných příznaků „*==SHOW==*“. Pro výběr všech příznaků slouží zaškrťovací políčko „*Select all*“, tuto funkci však nedoporučuji používat, jelikož při velkém množství dat dochází ke zdlouhavému načítání dat. Na obr. 6-33 je zobrazen seznam s načtenými daty.



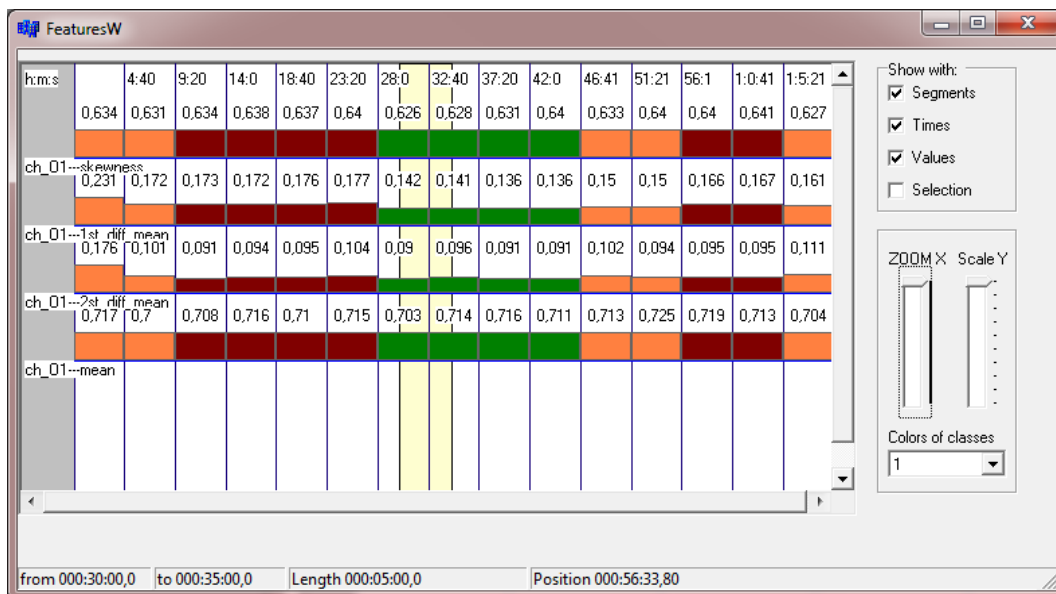
obr. 6-33 - načtení příznaků

⁵ *TStringList** vrací seznam řetězců. *StringList* má proměnou velikost, není třeba definovat velikost pole na začátku jako je tomu u obyčejných polí.

Po zvolení příznaků k zobrazení se opět předají veškeré informace pro vykreslení příznaků. Z vybraného seznamu, se najde index, na které pozici v souboru se nachází daný příznak pro segmenty. Data jsou načtena v řetězci. Jeden řetězec dat odpovídá jednomu segmentu. Protože data jsou v souboru oddělena prázdným znakem (mezerou), tak pro nalezení pozice vybraného příznaku, se počítají prázdné znaky. Po otevření okna z příznaky se zobrazí opět okno výše podobným. Mini menu „*Show with:*“ bylo již popsáno dříve, také „*ZOOM X*“ a „*ZOOM Y*“. Data po načtení a prvním spuštění je uvedeno na *obr. 6-34.* Roletové menu „*Colors of classes*“ slouží k barevnému rozlišení tříd. Vykreslená data jdou opět uložit do bitmapového souboru pomocí popUp menu na vykresleném signálu. Příklad zobrazeného signálu je také uveden na *obr. 6-35.*



obr. 6-34- zobrazení příznaků



6.6.7 Zobrazení spektrogramů⁶

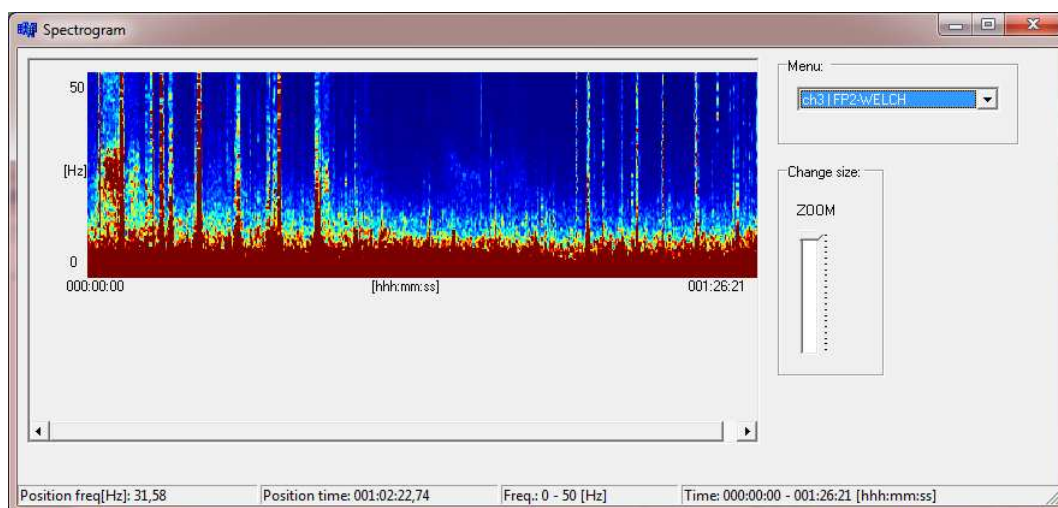
Při spuštění programu, se mimo jiné načítá také barevná maska pro vykreslení spektrogramu. Masku můžeme nazývat teplotní („JET“), protože nejslabší aktivitu (intenzitu) reprezentuje modrá, nejsilnější pak červená barva. Barevná maska je zobrazena na obr. 6-36.



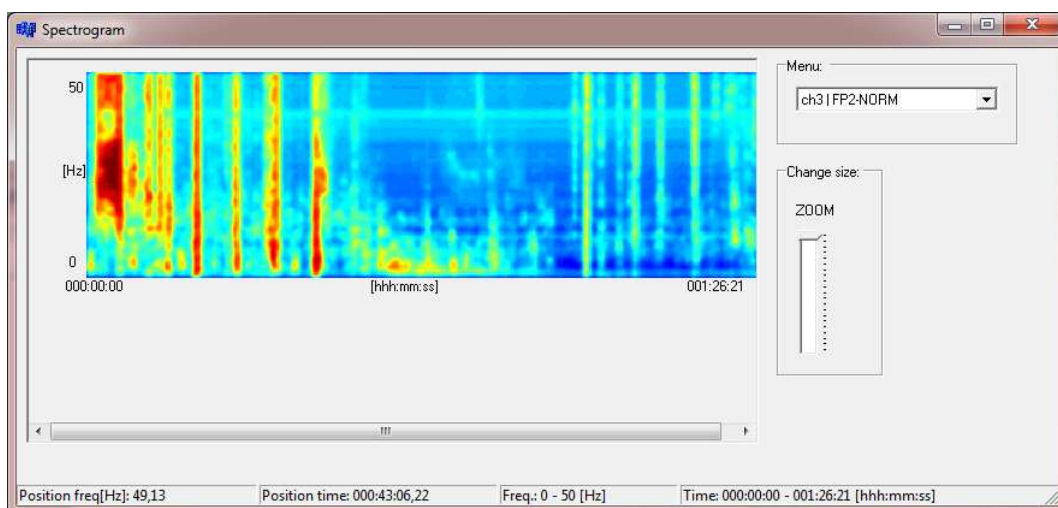
Pokud existuje soubor s uloženými daty pro zobrazení spektrogramu se aktivuje tlačítko v podmenu „Spect“. Při výběru tohoto tlačítka se zobrazí okno se spektrogramem.

⁶ Spektrogram obdržíme například výpočtem v programovém prostředí Matlab. Existuje více druhů spektrogramů, v této práci však použijeme Welchův výpočet spektrogramu, který byl obdržěn pomocí programového prostředí Matlab a to funkcí „pwelch(x)“. Funkce používá pro výpočet FFT (Fast Fourier transform, popsáno v [5]).

Výběrem s ComboBoxu můžeme vybírat veškeré dostupné spektrogramy. „*ZOOM*“ opět slouží ke zvětšení vykreslovaného spektrogramu. „*Statusbar*“ nese informace o spektrogramu a zároveň zobrazuje aktuální pozici. Při přejezdu přes vykreslený spektrogram se automaticky změní kurzor myši na křížek, pro přesnější určení polohy. Při vyvolání „*popUp*“ menu lze vykreslený spektrogram uložit do obrázku (formát *.bmp). Spektrogram vypočtený Welchovou metodou je na *obr. 6-38*. Normovaný spektrogram pak na *obr. 6-37*.



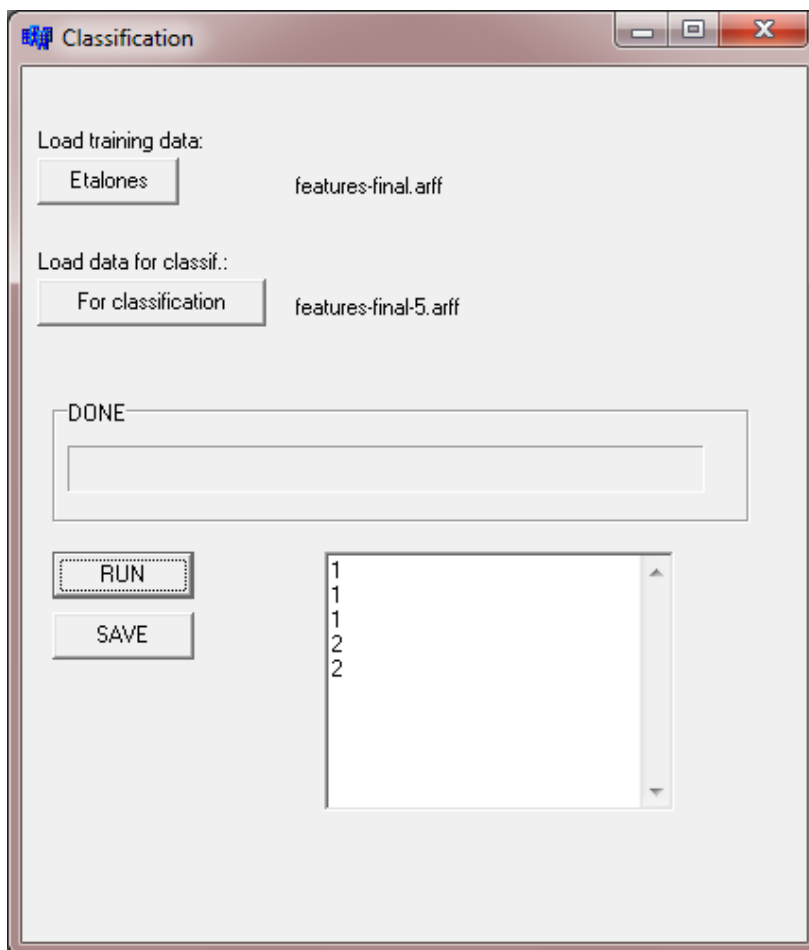
obr. 6-38 – Welchův spektrogram



obr. 6-37 – Normovaný spektrogram

6.7 Klasifikace podle nejbližších sousedů [k-NN]

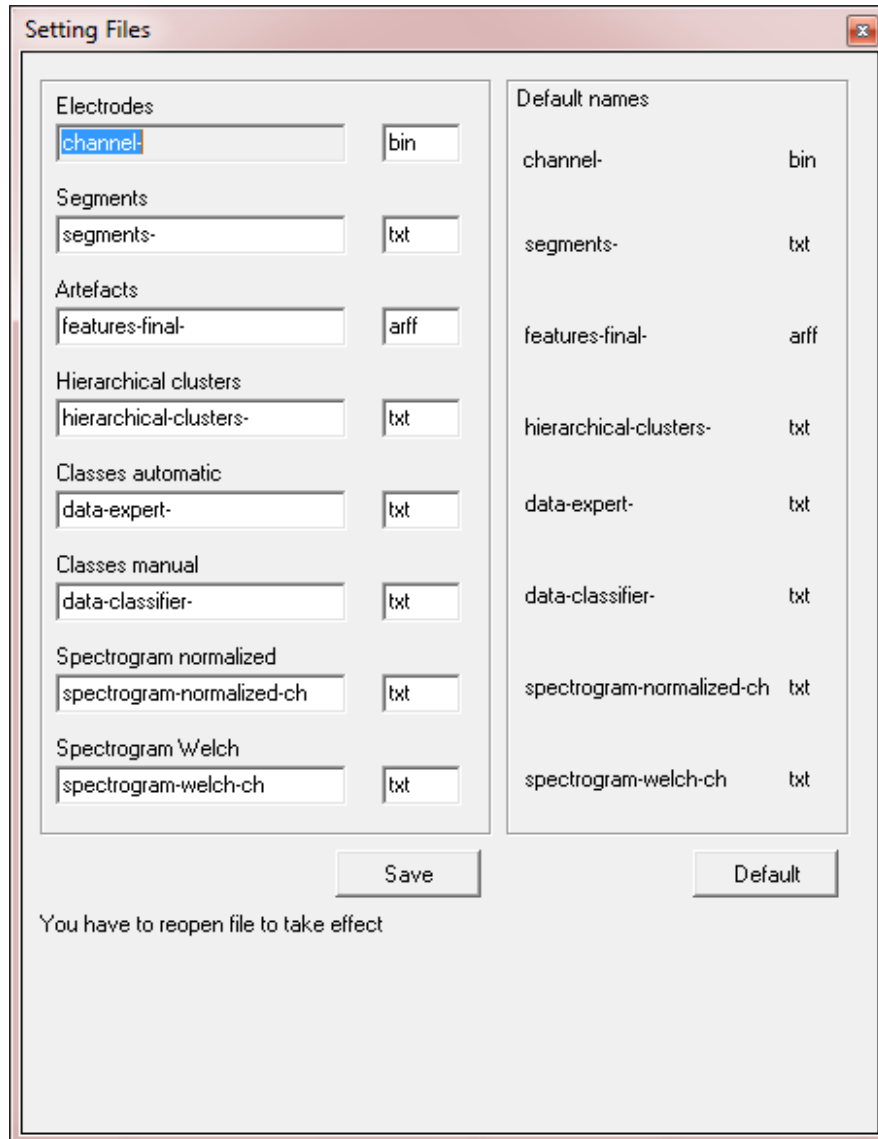
Klasifikaci podle nejbližších sousedů, lze vybrat z roletového menu „*Action* → *Classification*“ anebo ji lze spustit samostatně, protože tento modul byl programován jako samostatná aplikace. Pro klasifikaci je potřeba otevřít soubor s trénovací množinou a soubor pro klasifikaci (arff soubory, data s příznaky). Pokud jsou oba soubory úspěšně načteny je aktivováno tlačítko „*RUN*“ pro spuštění algoritmu pro klasifikaci. Počet nejbližších sousedů „*k*“ je volen automaticky a to podle počtu tříd ve trénovací množině (dvounásobný počet + 1). Pro přehled je veškeré načítání a zpracování indikováno pomocí „*Progress*“ baru. Pokud proběhne klasifikace úspěšně, lze výsledek klasifikace uložit do textového souboru pomocí tlačítka „*SAVE*“. Výsledek je také průběžně zobrazován přímo v aplikaci. Modul je zobrazen na obr. 6-39.



obr. 6-39 – klasifikace k-NN

6.8 Nastavení programu

Během testování jsem narazil na různé názvy jmen souborů. Proto byl doprogramován také modul pro nastavování vstupních souborů. Modul se nachází v roletovém menu „*Settings* → *Files*“. Tímto jsem odstranil nutnost přejmenovávání souborů. Názvy souborů jsou uloženy v textové podobě v souboru „*setfiles.set*“. Modul je zobrazen na *obr. 6-40*.



obr. 6-40 – Nastavení souborů

7) Závěr

Cílem práce bylo navrhnout program, který má sloužit jako vizuální část pro další dva programy, které jsou již realizovány (databáze grafoelementů a výpočetní operace nad EEG signály). Dále pak vybrat vhodnou klasifikační metodu a implementovat ji.

Výsledkem je zobrazovací nástroj EEG záznamů a výpočtů nad EEG záznamy. Důraz je kladen zejména na přehlednost všech obsažených prvků a možnost rozšíření o nové moduly. Aplikace byla navržena tak, aby bylo možné používat obecný polysomnografický záznam. Byly otestovány například záznamy spánkové, novorozenecké a epileptické. Byl naprogramován modul pro klasifikaci segmentů podle k-nejbližších sousedů, který lze spustit samostatně. Program je navíc doplněn o modul s nastavením názvů používaných souborů, protože ne vždy je možnost zachovat přesný název souboru v podobě, která byla předpokládána při zadávání této diplomové práce.

Struktura XML souboru, ze kterého jsou načítány všechny důležité informace o datech, nemusí být striktně dodržována, protože se počítá s budoucím rozšířením programu o nové moduly a funkce. Neměli by však být měněny názvy jednotlivých položek (názvy elektrod, vzorkovací frekvence apod.). Použitelnost programu není nijak omezena a může sloužit pro přehledné zobrazení dat (není potřeba načítat každý soubor zvlášť, jako tomu bylo dopsud).

Program je rozšířením projektu pro automatickou klasifikaci signálu EEG. Data použitelná a plně kompatibilní jsou jak ze samostatné databáze Bc. Martina Kytky, tak i výstupy z programu ing. Václava Gerly.

Program by měl také v budoucnu automaticky provádět adaptivní segmentaci, shlukování a klasifikaci dat za pomoci grafoelementů uložených v databázi. Mezi další přídatné moduly by mohly patřit například: automatické výpočty na signálech, filtrace signálů, segmentace a další typy vizualizací a jiné neméně důležité výpočetní operace.

8) Reference

- [1] **Horniak, R.** *Bakalářská práce: Aplikace databáze v EEG.* Praha : ČVUT FEL, 2006.
- [2] **Gerla, V.** PSGLab. *Polysomnographic Data Processing Matlab Toolbox.* [Online] [Citace: 2009-12-28] <http://bio.felk.cvut.cz/psglab>.
- [3] **Kantůrek, M.** *Adaptivní segmentace EEG signálu.* Praha : ČVUT, 2009.
- [4] **Krajča V., Petránek S., Patáková I., Väri A.** *Automatic Identification of Significant Graphoelements in Multichannel EEG Recordings by Adaptive Segmentation and Fuzzy Clustering.* místo neznámé : Biomed Comput. 28, 1991. stránky 71-89.
- [5] **Bílovský, P.** *Waveletová transformace a její použití při zpracování signálů.* [pdf dokument] Ostrava - Poruba : VŠB - TU Ostrava.
- [6] **Paul, S., A.** *The Illustrated Wavelet Transform Handbook.* London : Institute of Physics, 2002. 0 7503 0692 0.
- [7] **Drbal, O.** *Vybrané metody pro měření zpoždění mezi signály EEG.* [pdf dokument] Praha : České vysoké učení technické v Praze.
- [8] **Gerla V., Lhotská L., Krajča V., Paul K.** *Multichannel Analysis of the Newborn EEG Data.* [CD-ROM] Praha : autor neznámý, 2006. In International Special Topics Conference on Information Technology in Biomedicine.
- [9] **Christos, S. a Dimitrios, S.** NEURAL NETWORKS. *NEURAL NETWORKS.* [Online] [Citace: 2009-12-24] http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html.
- [10] **Moore, A., W.** *K-means and Hierarchical Clustering.* [PDF-slides] Carnegie Mellon University : School of Computer Science, 2001. 412-268-7599.
- [11] **Houdek, Michal, Svoboda, Tomáš a Procházka, Tomáš.** *Klasifikace podle nejbližších sousedů, Nearest Neighbours Classification [k-NN].* [Pdf dokument] Praha : ČVUT, 2001.
- [12] **Muller, J., M., Brisebarre, N. a Dinechin, F.** *Handbook of Floating-Point Arithmetic.* Sydney : University of Sydney, 2006. ISBN: 978-0-8176-4704-9.
- [13] **Malmivuo J., Plonsey R.** *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields.* New Yourk : Oxford University Press. 512 p., 1995. ISBN 0-19-505823-2.
- [14] **Stern J. M., Engel J.** *Atlas of EEG Patterns.* Philadelphia : 307p, 2005. ISBN 0-7817-4124-6.
- [15] **Hovorka J., Nežádal T., Herman E., Praško J.,.** *Klinická Elektroencefalografie.* Praha : Maxdorf, 2003.

- [16] **Golz M., Sommer D., Lembcke T., Kurella B., Schmalkalden F.** *Classification of Prestimulus EEG of K-complexes using Competitive Learning Networks*. Berlin : autor neznámý, 1995.
- [17] **Lewicke A. T., Sazonov E. S., Schuckers A. C.** *Sleep-Wake Identification in Infants: Heart Rate Variability Compared to Actigraphy*. San Francisco : IEEE Engineering Medicine and Biology Society Annual Conference Proceedings, 2004.
- [18] **Turnbull J. P., Loparo K. A., Johnson M. W., Scher M. S.** *Automated detection of trace alternant during sleep in healthy full-term neonates using discrete wavelet transform*. místo neznámé : *Clinical Neurophysiology* 112, 2001.
- [19] **Barlow J. S., Creutzfeldt O. D., Michael D., Houchin J., Epelbaum H.** *Automatic adaptive segmentation of clinical EEGs*. místo neznámé : *Electroencephalography and Clinical Neurophysiology*, 1981.
- [20] **Mařík V., Štěpánková O., Lažanský J.** *Umělá inteligence I*. Prague : Academia, 1993.
- [21] **Yamada Y., Suzuki E.** *Decision-tree Induction from Time-series Data Based on a Standard example Split Test*. Yokohama : autor neznámý, 2003. stránky 840-847. ICML 2003.
- [22] **Largo R., Rosa A.** *Sleep EEG Processing with Wavelets*. místo neznámé : Institute for System and Robotics.
- [23] **Liu H., Motoda H.,.** *Feature extraction, construction and selection: a data mining perspective*. Singapore : National University of Singapore, 1998. ISBN: 0-12-493290-8.
- [24] **Schlesinger M. I., Hlaváč V.** *Ten lectures on statistical and structural pattern recognition*. Prague : CTU, 1999.
- [25] **Windeatt T., Roli F.** *Multiple Classifier Systems*. Guildford : autor neznámý, 2003. ISBN: 0302-9743.
- [26] **Torralba A., Murphy K. P., William T.F.,.** *Sharing features: efficient boosting procedures for multiclass object detection*. MIT Cambridge : autor neznámý, 2004. IEEE Conf. Computer Vision and Pattern Recognition.
- [27] **Kamel M. S., Wanas N. M.** *Data Dependence in Combining Classifiers*. Ontario, Canada : University of Waterloo. N2L-3G1.
- [28] **Ting K. M., Witten I. H.,.** *Issues in Stacked Generalization*. místo neznámé : School of Computing, *Journal of Artificial Intelligence Research* 10, 1999.
- [29] **Gerla V., Lhotská L., Krajča V.** *Utilization of Time-Dependence in EEG Signal Classification*. Praha : Společnost biomedicínského inženýrství a lékařské informatiky ČLS JEP, 2005. ISSN: 1727-1983.
- [30] **Durka, J. P.** *Time-frequency analyses of EEG*. [editor] Warsaw University. Warszawa : Institute of Experimental Physics, 1996.

- [31] **Kytka, M.** *Bakalářská práce: Grafoelementy v EEG*. Praha : ČVUT FEL, 2006.
- [32] **Himananen, L., S.** *A Sleep Spindle Detection Algorithm*. Te Whare Wanaga o Waikato : The University of Waikato, 2002.
- [33] **Lederman, D.** *Automatic classification of Infants cry, diploma thesis*. Negev : Ben-Gurion University of the Negev, Faculty of Engineering Science, 2002.
- [34] **Matoušek, D.** *C++ Builder - vývojové prostředí*. Praha : BEN - technická literatura, 2000. ISBN: 80-86056-70-8.
- [35] **Mallat, S.** *Wavelet Tour of Signal Processing 2nd Edition*. San Diego, California : autor neznámý, 1999. ISBN: 978-0-12-466606-1.
- [36] **Mlčoch, Z.** Průběh EEG vyšetření, EEG epileptika, zdravého a usínajícího, spícího pacienta. [Online] [Citace: 14. 01 2009.]
http://www.zbynekmlcoch.cz/info/neurologie/prubeh_eeg_vysetreni_eeg_epileptika_z_draveho_a_usinajiciho_spiciho_pacienta.html.
- [37] DSP research group. *Fast Fourier Transform (FFT)*. [Online] [Citace: 2009-12-22]
<http://www.cmlab.csie.ntu.edu.tw/cml/dsp/training/coding/transform/fft.html>.
- [38] **Chen, P., Y. a Popovich, P., M.** *Parametric and Nonparametric Measures*. California : Thousand Oaks, 2002. ISBN: 0-7619-2228-8.