

Západočeská univerzita v Plzni
Fakulta aplikovaných věd
Katedra informatiky a výpočetní techniky

Bakalářská práce

Biometrická autentizace osob na základě EEG signálu

Místo této strany bude
zadání práce.

Prohlášení

Prohlašuji, že jsem bakalářskou práci vypracoval samostatně a výhradně s použitím citovaných pramenů.

V Plzni dne 27. června 2016

Martin Šíp

Abstract

The main goal of this work is to study the possibilities of user authentication using biological signals such as ECG and EEG. It covers basic terms from machine learning, signal processing and theoretical principles of feature extraction from the above mentioned biological signals. The gained theoretical knowledge is afterwards used for measuring suitable EEG signals and implementation of feature extraction algorithms from measured data. Extracted features are thoroughly tested for living subjects authentication. The last part is the design and implementation of application helping in the practical part of the research.

Abstrakt

Hlavním cílem je studie možností využití biologických signálů jako jsou EEG a EKG potenciály pro autentikaci osob. Práce uvádí a vysvětluje základní termíny z oblasti strojového učení, zpracování signálů a teoretické principy extrakce parametrů z výše zmíněných biologických signálů. Získané znalosti jsou následně využity při návrhu scénářů pro získání vhodných EEG potenciálů a implementaci algoritmů pro extrakci parametrů z naměřených dat. Extrahované parametry jsou následně důkladně otestovány při autentikaci živých subjektů. Poslední část je věnována návrhu a vývoji aplikace nápomocné při praktické části výzkumu.

Obsah

1	Úvod	7
2	Biometrická autentizace	8
2.1	Co je autentizace	8
2.2	Biometrie	8
2.2.1	Úloha biometrie	9
2.2.2	Metriky biometrického systému	10
3	EKG biometrie	12
3.1	Úvod	12
3.2	Elektrické děje v srdci	12
3.3	Průběh EKG signálu	12
3.4	EKG v biometrii	14
3.5	Parametry EKG signálu	14
4	EEG biometrie	15
4.1	Úvod	15
4.2	EEG v biometrii	17
5	Základní pojmy	18
5.1	EEG dataset	18
5.2	Extrakce parametrů	18
5.2.1	Analýza hlavních komponent	18
5.2.2	Fourierova transformace	19
5.3	Vektor příznaků	19
5.4	Klasifikace	20
6	Realizace	22
6.1	Scénáře měření	23
6.1.1	Počítání barev	24
6.1.2	Představa pohybu rukou	25
6.1.3	Zavřená oční víčka v bdělém stavu	25
6.2	Parametry naměřených dat	26
6.2.1	Zavřená oční víčka	26
6.2.2	Počítání barev a představa pohybu	28
6.3	Volba klasifikátoru	28

7	Biometrický systém	30
7.1	O aplikaci	30
7.2	Specifikace požadavků	30
7.3	Návrh	31
7.3.1	Příjem dat	31
7.3.2	Databáze aplikace	31
7.3.3	Extrakce parametrů a autentikace	31
7.3.4	Analýza přijatých dat	32
7.3.5	Testování autentikačních algoritmů	32
7.4	Realizace	32
7.5	Testování	33
7.6	Použité nástroje	33
7.7	Uživatelská dokumentace	34
8	Autentizační testy	35
8.1	Formát výsledků	35
8.2	Zavřená oční víčka	36
8.2.1	Výsledky	37
8.2.2	Diskuze	37
8.3	Představa pohybu a počítání barev	38
8.3.1	Výsledky	38
8.3.2	Diskuze	38
9	Závěr	40
	Literatura	41
A	Uživatelská dokumentace	43
A.1	O aplikaci	43
A.2	Překlad	44
A.3	Požadavky aplikace	44
A.3.1	Softwarové požadavky	44
A.3.2	Hardwarové požadavky	44
A.4	BrainVision server	44
A.5	Ovládání	45
A.5.1	Administrátor	45
A.5.2	Autentikace datasetu	48
A.5.3	Běžný uživatel	52
B	Spektrogramy EEG signálu	53
B.0.1	Zavřená oční víčka	53

1 Úvod

Pod pojmem biometrie případně biometrické autentizace rozumíme metodu prokazování proklamované identity živého subjektu na základě jeho jedinečných a pokud možno neměnných biologických charakteristik. Tyto charakteristiky mohou být obsaženy například v oční duhovce, sítnici, otisku prstu, DNA nebo charakteru písma jedince.

Metody biometrické autentizace založené na skenu duhovky či rozpoznávání otisků prstů jsou již hojně nasazeny v bezpečnostních systémech a dosahují výborných výsledků. Na druhou stranu se také množí případy úspěšného obcházení těchto systémů za použití falešných otisků prstů, či vytisknutého snímku duhovky.

Proto se začaly hledat způsoby autentizace, jejichž případné padělání je nemožné nebo velmi obtížné. Možnými kandidáty splňujícími tuto požadovanou vlastnost jsou EEG a EKG potenciály, které je potřeba měřit přímo na těle. Vyžadují tedy bezprostřední přítomnost autentizovaného subjektu a jejich případné falšování je tak potenciálním podvodníkům výrazně znesnadněno.

Cílem této práce je tedy prostudování literatury týkající se biometrické autentizace osob na základě EEG a EKG potenciálů, následný návrh algoritmu pro získávání parametrů z EEG signálu, jeho implementace a praktické ověření. Zadavatel klade důraz na zhodnocení výsledků při zvoleném postupu a jejich zdokumentování. Výstupem budou navržené scénáře měření EEG aktivity, analýza dat získaných měřeními podle těchto scénářů, návrh a realizace algoritmů pro získání parametrů z analyzovaných dat, testování implementace a diskuze výsledků v závislosti na scénáři měření a algoritmu extrakce parametrů.

2 Biometrická autentizace

2.1 Co je autentizace

Autentizace (jinak také *autentifikace* či *autentikace*) je proces ověření proklamované identity subjektu. Jedná se o bezpečnostní opatření, kterým zajišťujeme ochranu před falšováním identity. Vlastnost, kterou autentizací dokazujeme se nazývá **autenticita**. Po autentizaci zpravidla následuje **autorizace**, pod tímto pojmem rozumíme souhlas s provedením nějaké akce, povolení k přístupu do fyzických prostor nebo informačního systému. Způsoby autentizace lze rozdělit do základních kategorií podle zkoumaných faktorů [8].

- **znalostní faktory** (angl. *the knowledge factors*) - něco, co uživatel **zná** (např. bezpečnostní **PIN**, kombinace uživatelského **jména** a **hesla** případně heslové fráze, bezpečnostní otázka a podobně)
- **vlastnické faktory** (angl. *the ownership factors*) - něco, co uživatel **vlastní** (např. identifikační karta, bezpečnostní token, hardwarový nebo privátní klíč etc.)
- **faktory inherence** (angl. *inherence factors*) - něco, co přímo pochází z osoby uživatele nebo jeho chování (např. řetězec **DNA**, hlas, otisky prstů, snímek duhovky, podpis a pod.)

Z průzkumů bylo zjištěno to - aby byl autentizační systém považován za spolehlivý, musí být založen **alespoň na dvou** faktorech, v takovém případě hovoříme o tzv. **dvoufázovém ověření** [1].

2.2 Biometrie

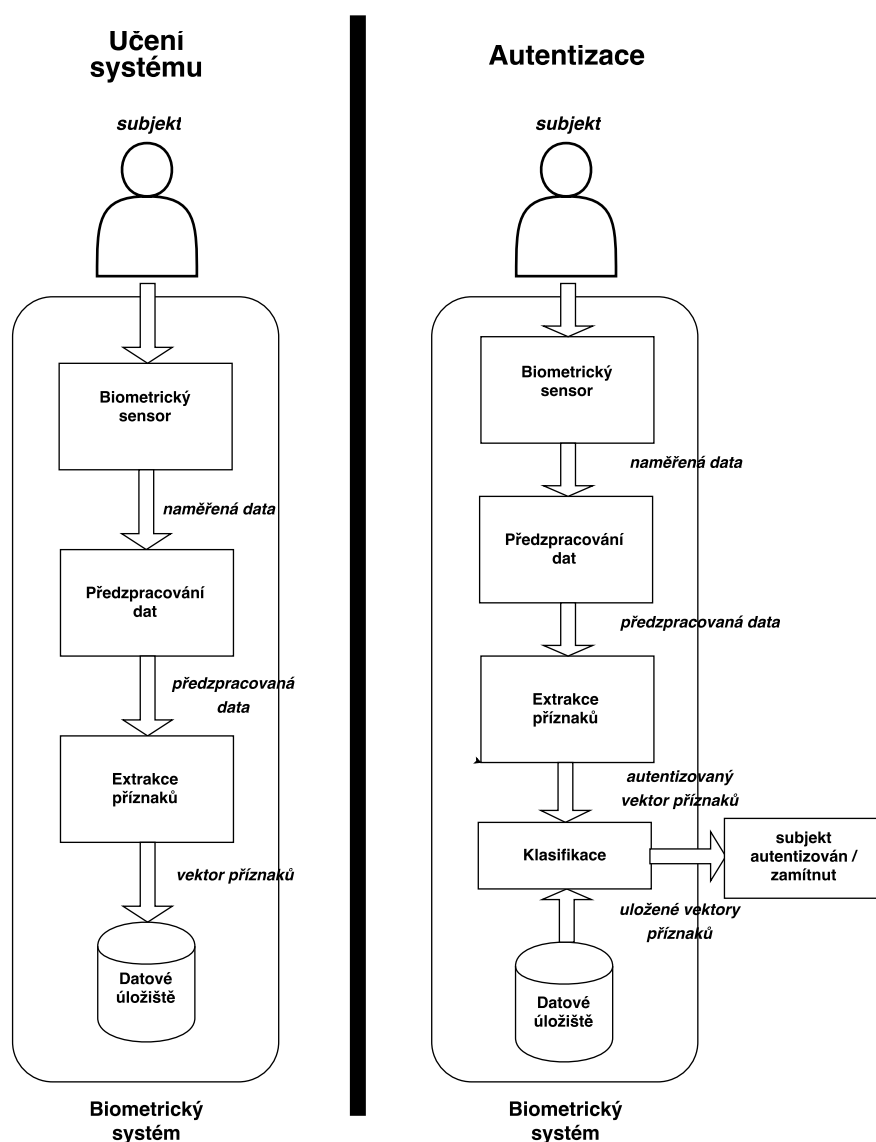
Pod pojmem biometrie rozumíme metodu výše zmíněné autentizace, která je založená na biologických charakteristikách subjektu. Abychom metodu založenou na takovéto charakteristice mohli považovat za perfektní, musela by splňovat tři následující **podmínky** [1]:

- **univerzálnost** - na každém subjektu lze tuto charakteristiku změřit, či jinak získat
- **unikátnost** - žádní dva jedinci nesdílejí stejnou charakteristiku

- **permanentnost** - charakteristika je neměnná v závislosti na čase

Existují metody biometrie založené na **fyziologických charakteristikách**, jako jsou ku příkladu výše zmíněné otisky prstů, DNA nebo sken duhovky. Další metody jsou založené na takzvaných **behaviorálních charakteristikách**. Mezi ně patří například Gaitova analýza studující tělesnou mechaniku pohybu, studie lidského hlasu a v neposlední řadě metody zkoumající dynamiku úderů do klávesnice nebo pohybu ruky při podpisu.

2.2.1 Úloha biometrie



Obrázek 2.1: Obecná úloha biometrie

Obrázek 2.1 zachycuje obecnou úlohu biometrie případně biometrického systému. Úloha biometrického systému se skládá ze dvou hlavních částí a to sice **učení systému** a **autentizace** samotné [1]. Použité pojmy jsou zde pro zachování plynulosti textu pouze povrchně vysvětleny. Jejich bližší vysvětlení bude následovat v 5. kapitole.

- **Učení systému** - Při učení systému jsou biometrické charakteristiky naměřeny pomocí senzoru. Naměřená data jsou následně předzpracována. Z předzpracovaných dat vyextrahujeme parametry ve formě **příznakového vektoru**, které uložíme do systému. Pod pojmem příznakového vektoru si můžeme představit množinu čísel, která určitým způsobem reprezentuje naměřená data. Proces extrakce parametrů opakujeme, dokud není databáze biometrického systému naplněna požadovaným množstvím vzorových vektorů reprezentujících charakteristiky měřeného subjektu.
- **Autentizace** - Během autentikace jsou první tři fáze stejné jako v případě učení. Rozdíl spočívá v tom, že extrahovaný příznak, který v tomto případě reprezentuje charakteristiky subjektu v roli například uživatele přihlašujícího se do autentikačního systému, je zaveden spolu se vzorovými příznaky získanými ve fázi učení do **klasifikačního bloku**. Na základě těchto vzorových příznaků patřících konkrétnímu uživateli systému pak klasifikační blok rozhodne, zda nově příchozí vektor patří do stejné **třídy** jako vzorové příznaky. V kladném případě je subjekt autentizován pod identitou vlastníka vzorových příznaků, v opačném je autentizace zamítnuta.

2.2.2 Metriky biometrického systému

Pro hodnocení výkonnosti a kvality biometrického systému jsou použity následující metriky spolu s dalšími [1]:

- **False match rate (FMR)** - Pravděpodobnost, že systém přiřadí vstupní vektor parametrů do neodpovídající třídy vzorových vektorů. To prakticky znamená pravděpodobnost, s jakou se je cizí subjekt např. potenciální podvodník schopen autentizovat jako oprávněný uživatel systému.
- **False non-match rate (FNMR)** - Pravděpodobnost s jakou systém nepřihradí vstupní vektor do třídy jemu odpovídajících vzorových vektorů. Jinými slovy řečeno se jedná o pravděpodobnost, s jakou je zamítnuta autentizace oprávněnému uživateli biometrického systému.

- **Failure to enroll rate (FER)** - Pravděpodobnost s jakou selže zařazení vstupního vektoru ve fázi učení systému mezi vzorové vektory v databázi. To může být způsobeno chybou při zpracování naměřených dat, jejich nízkou kvalitou a podobně.
- **Equal error rate (EER)** - Představuje hodnotu, při níž se **FMR** a **FNMR** přibližně rovnají. Prakticky se EER vypočte jako aritmetický průměr FMR a FNMR. Obecně řečeno, zařízení s nejnižší **EER** se dá považovat za nejpřesnější.

3 EKG biometrie

3.1 Úvod

EKG nebo **elektrokardiogram** je záznam časové změny elektrického potenciálu, způsobené srdeční aktivitou. Který zavedl začátkem 20. století holandský fyziolog a držitel Nobelovy ceny **W. Weinthovenem**. Jedná se většinou o neinvazivní vyšetření, kdy jsou elektrody připevněny na povrchu kůže měřeného subjektu, můžou být však umístěny přímo na stěně jícnu nebo v srdci [2].

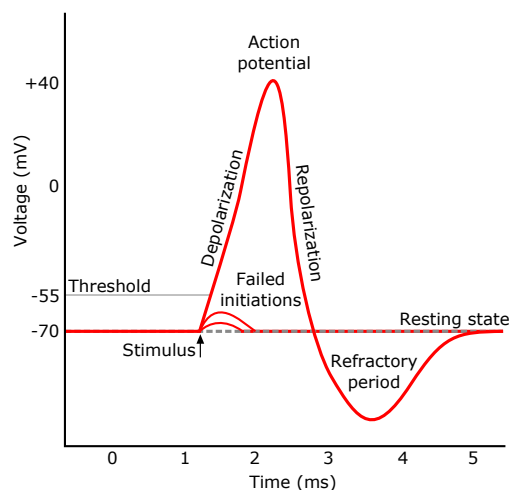
3.2 Elektrické děje v srdci

V buňkách srdečního svalu neboli **myokardu** dochází k transportu iontů nesoucích elektrický náboj mezi vnitřní a vnější stranou buněčné membrány. Tyto ionty jsou nerovnoměrně rozloženy. Nerovnoměrné rozložení je způsobeno převahou záporných aniontů uvnitř buňky a činností pumpy buněčné membrány, která přesouvá sodíkové kationty (Na^+) z buňky a draslíkové ionty (K^+) do buňky, čímž dochází ke změně elektrického potenciálu membrány.

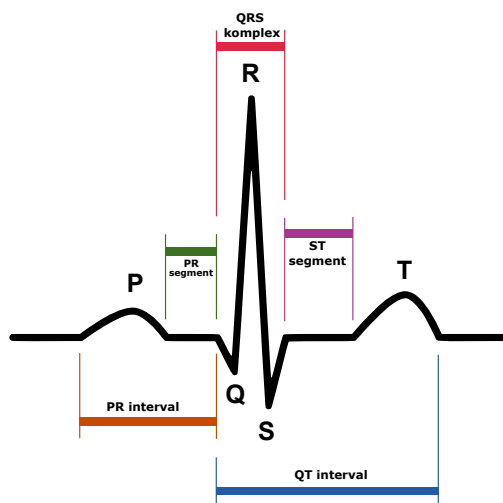
V klidové fázi je membrána vodivá pro draslík. Ten vstupuje do buňky a povrch membrány se stává pozitivním. Nastává tzv. fáze **klidové polarizace** membrány. Při dosažení nadprahové hodnoty začnou dovnitř buňky proudit sodíkovým kanálem (Na^+) kationty. Na vnější straně membrány tak dochází k úbytku kladných kationtů a stává se tak negativní. Nastává tedy tzv. fáze **depolarizace**. Při otevření vápníkového kanálu (Ca^{2+}) dochází rovněž k depolarizaci buněčné membrány a hovoříme o tzv. **plató** fázi. Tyto tři zmíněné fáze tvoří dohromady tzv. **akční potenciál** viditelný na obrázku 3.1. Díky akčnímu potenciálu dochází k přenosu vzruchu mezi jednotlivými buňkami a nastává svalový stah [2].

3.3 Průběh EKG signálu

Výše zmíněné elektrické děje v buňkách srdečního svalu se akumulují a šíří se jako signál myokardu do celého těla bez výraznějšího zeslabení. Pohybuje se také proto v poměrně velkých amplitudách (řádově jednotky až desítky mV) a je měřitelný téměř všude na povrchu kůže.



Obrázek 3.1: Akční potenciál



Obrázek 3.2: Průběh cyklu EKG

První vlna signálu viditelná elektrokardiogramem je **vlna P**. Ta se objevuje při depolarizaci srdečních předsíní, svědčí o jejich začínajícím stahu neboli **kontrakci**. Následuje tzv. **QRS komplex** způsobený depolarizací komor spolu s **T vlnou** zapříčiněnou jejich následnou repolarizací. Tento průběh se cyklicky opakuje, hovoříme tedy o tzv. **srdečním cyklu** viz obr. 3.2.

3.4 EKG v biometrii

Prvotní příčinou, proč se začalo uvažovat o EKG ve spojitosti s biometrií, byl fakt, že měřené potenciály neprocházejí pouze srdcem měřeného člověka, nýbrž i celým jeho tělem. Tvar výsledné křivky odráží tedy nejen fyziologické vlastnosti srdce měřeného člověka, ale i stavbu jeho těla. Další velkou výhodou je relativně snadné změření signálu elektrokardiogramu a technická nenáročnost měřících senzorů [11].

V případě EKG biometrie je bod **univerzálnosti** splněn, každému živému subjektu je možné změřit srdeční aktivitu. Podmínka **permanentnosti** splněna není, fyziologická stavba člověka je proměnná v závislosti na čase, dalším faktorem může být aktuální psychický stav subjektu ovlivňující jeho srdeční tep a podobně. Problém proměnlivé fyziologie se dá částečně opravit biometrickým systémem, který se bude pomalu dynamicky vyvíjet spolu s nově přichozími daty od autentizovaného subjektu. Míra splnění bodu **unikátnosti** je velmi závislá na metodě extrakce parametrů z naměřených potenciálů.

3.5 Parametry EKG signálu

Při návrhu parametrů se ve většině případů vychází z již zmíněného **QRS komplexu** vznikajícího při depolarizaci komor. Jako jednotlivé parametry EKG se pak často používají časové intervaly mezi výskytem po sobě jdoucích vln P, Q, R, S a jejich kombinace [11]. Některé metody využívají pro extrakci parametrů Fourierovu transformaci (např. [2, 12]), která bude blíže vysvětlena v kapitole 6.

4 EEG biometrie

4.1 Úvod

EEG neboli **elektroencefalogram** je neinvazivní diagnostická metoda sloužící k záznamu elektrické aktivity mozku. Změny potenciálů velkého množství synapsí (neuronů) jsou akumulovány a následně detekovány jako mozkové vlny za pomoci elektrod na pokožce hlavy. Rozlišujeme několik základních druhů **mozkových vln** [2] lišících se svým frekvenčním rozsahem a okolnostmi výskytu viz tabulka 4.1.

Název	Frekvenční rozsah [Hz]	Výskyt
Alfa	8-13	zavřené oči v bdělém stavu, klid, relaxace, pocit únavy
Beta	13-30	vnímaní, koncentrace, mentální aktivita, stres
Gama	30-42	strach, agrese, zloba, analytické myšlení
Theta	4-8	sny, hluboký spánek, relaxace
Delta	0.5-4	koma, regenerace mozku, hluboké uvolnění

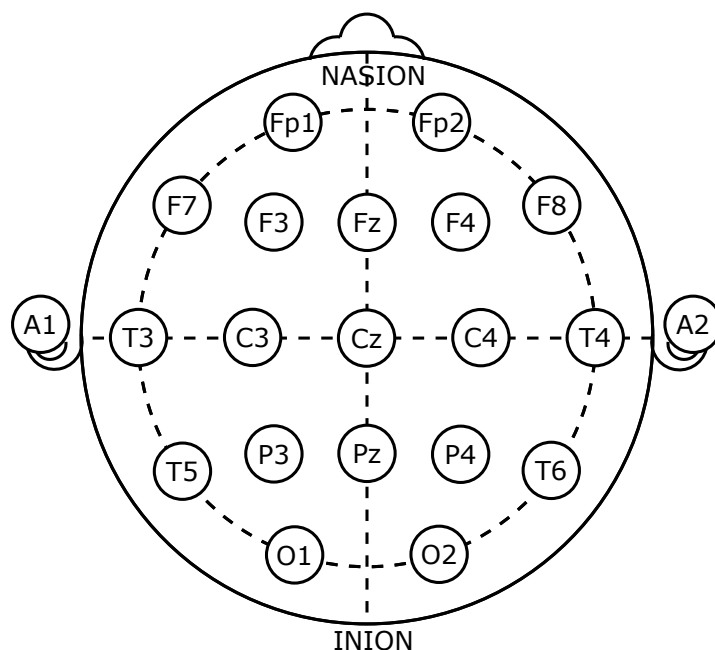
Tabulka 4.1: Druhy mozkových vln

Přístroj provádějící vyšetření se nazývá **encefalograf**. Ten se typicky skládá ze **čtyř částí** [3]:

- **Diferenční zesilovač** - *Vstupní zesilovač* zesiluje potenciály získané na elektrodách připevněných k lidské kůži z velmi nízkého napětí v řádu desítek mikrovoltů na napětí, se kterým je možno dále pracovat. *Diferenční zesilovač* neboli vstupní zesilovač v diferenčním zapojení obsahuje dva vstupy, jeden přímý a druhý invertovaný sloužící jako **referenční**. Pracuje tak, že zesiluje rozdíl napětí na obou vstupech. Smysl tohoto zapojení je v potlačení elektromagnetického rušení z okolí. Naindukuje-li se na obou vstupech stejně velké rušivé napětí, pak dojde k jeho odečtení a na výstupu se neprojeví.
- **Kanály** - Kanálem nazýváme výstup zesilovače. Platí jednoduché pravidlo, že kolika kanály je aparatura vybavena, tolik může zpracovat

vstupních signálů. V kanálu zesilovače existuje i možnost úpravy signálu pomocí **filtrů**.

- **Zápisové zařízení** - V dnešní době dochází k zápisu naměřených potenciálů nejčastěji v digitální podobě. Méně častěji se vyskytuje mechanický zápis na otočný válec s navinutým rastrovým papírem pomocí pisátka, v takovém to případě musí být aparatura vybavena výkonným zesilovačem schopným rozpohybovat zápisové pisátko.
- **Elektrody** - Nejčastěji se používají povrchové plovoucí elektrody. To znamená, že se připevňují na povrch kůže pomocí vodivých past. Elektrody se rovněž mohou vyskytovat v různém rozmístění. Mezi nejznámější patří tzv. mezinárodní **10-20 systém**. Čísla 10 a 20 v názvu odpovídají faktu, že sousední elektrody jsou od sebe vzdáleny 10 procent šířky a 20 procent délky lebky viz obrázek 4.1. Písmenné pojmenování jednotlivých elektrod **F**, **T**, **C**, **P** a **O** označuje jednotlivé mozkové laloky, nad kterými jsou elektrody umístěny, tedy *frontal* (čelní), *central*, *temporal* (spánkový), *parietal* (temenní) a *occipital* (týlní). Sudé číselné označení pak vyjadřuje umístění na pravé hemisféře, liché odpovídá levé hemisféře.



Obrázek 4.1: Systém 10-20

Při měření mozkové aktivity se dále setkáváme rovněž s nežádoucími jevy. V případě, že je rušivý nežádoucí signál navázán na EEG záznam, hovoříme

o takzvaném **artefaktu** [3]. Artefakty mohou být:

- **Technické** - Například rušení z okolních elektrospotřebičů zapojených do stejné elektrické sítě, způsobené nedostatečným uzemněním nebo elektromagnetický šum generovaný elektrickými zařízeními v okolí.
- **Biologické** - Nejčastěji vyvolané pohybem očí, mimických svalů, či přímo srdeční aktivitou.

4.2 EEG v biometrii

Měření mozkových vln je poměrně komplikovaná a časově náročná záležitost, existuje proto velice málo studií zabývajících se mozkovými vlnami v souvislosti s biometrií. Za zmínku stojí například výsledky R.B. Paranjapeho a kolektivu [4], kteří ve své studii kombinovali autoregresní modelování (AR) EEG signálu v kombinaci s klasifikací lineárně diskriminantní analýzou. Dosáhli úspěšnosti pohybující se mezi 49 a 85%. Rovněž je zajímavá studie Tottoriovy univerzity [5], kde se pokoušeli o autentizaci řidičů na základě mozkových vln spojených s řízením automobilu za využití zpracování signálu diskrétní Fourierovou transformací. Získali velmi uspokojivé **EER** pohybující se v rozmezí 27 až 37%.

Mozkové vlny může samozřejmě generovat pouze mozek živé bytosti, podmínka **univerzálnosti** je tedy automaticky splněna. Další velkou výhodou je, že mozkové vlny se nacházejí uvnitř těla, což velice znesnadňuje jakýkoliv pokus o jejich podvržení. Navíc mozek vydává tyto signály zcela automaticky a nevědomě, tento fakt umožňuje tzv. *on-demand autentizaci*, neboli teoreticky můžeme autentikaci provést okamžitě v případě potřeby. Tyto vlastnosti dělají z EEG signálů dobrého kandidáta na biometrickou charakteristiku použitelnou v biometrickém systému. Cílem této práce je proto návrh algoritmu pro získávání biometrických parametrů z EEG signálu, jeho implementace a následné otestování provedením **sady měření**.

Z dosud provedených výzkumů je patrné, že metoda algoritmu získávání biometrických potenciálů z EEG signálu je silně závislá na scénáři měření mozkové aktivity a způsobu **stimulace** měřeného subjektu. Setkáváme se tak s **dvěma pomyslnými přístupy** získání vhodných dat EEG pro biometrické účely, od nichž se odvíjí návrh scénářů měření mozkové aktivity:

- **Resting brainwaves biometrics** - Ve volném překlade biometrie na základě klidových mozkových vln. (např. [6])
- **Tasked condition biometrics** - Biometrie založená na základě mozkových vln vznikajících při mentální a/nebo fyzické činnosti.

5 Základní pojmy

V následujících kapitolách budou uvedeny základní pojmy z oblasti **strojového učení**, **rozpoznávání** a **teorie zpracování signálu** nutné pro plné porozumění následujícího textu a k osvětlení textu předchozího.

5.1 EEG dataset

Jedná se většinou o binární soubor nejčastěji ukončený koncovkou **.eeg**. Obsahuje naměřené potenciály na jednotlivých elektrodách v závislosti na čase. Pro naše potřeby takto pojmenujeme výstup jednoho měření subjektu elektroencefalografem z digitálního záznamového zařízení.

5.2 Extrakce parametrů

Termín **extrakce parametrů** někdy též **extrakce příznaků** (angl. *feature extraction*) pochází z oblasti strojového učení a rozpoznávání. Jedná se o proces, kdy z původních dat, které jsme získali například měřením, odvodíme novou množinu dat, takzvaných **parametrů**. Tyto parametry nesou pokud možno co největší informační hodnotu původní množiny dat a mají velmi nízkou redundanci neboli informační nadbytečnost [8]. V následujících podkapitolách bude detailněji uvedeno několik základních metod užívaných při extrakci parametrů z diskrétního **biologického signálu** například z **elektroencefalografu** nebo **elektrokardiogramu**.

5.2.1 Analýza hlavních komponent

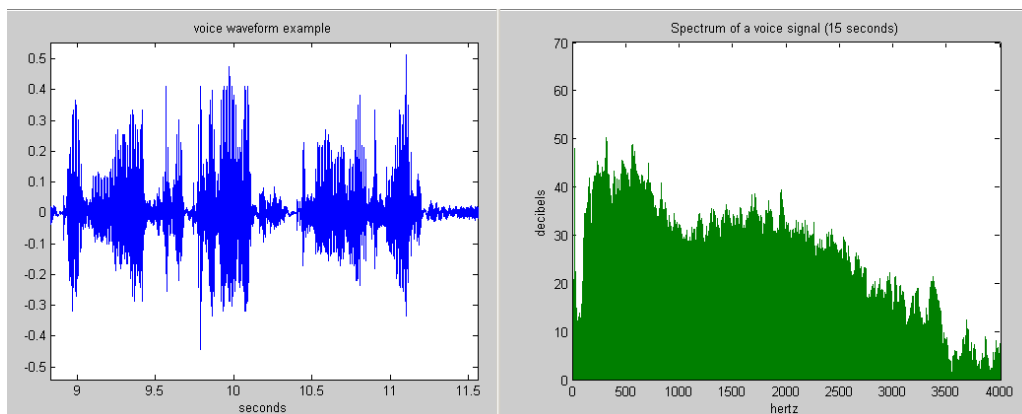
Analýza hlavních komponent (*Principal Component Analysis, PCA*) je pojmem z teorie signálu. Jedná se o ortogonální transformaci, která převádí množinu pravděpodobně lineárně závislých číselných pozorování na množinu lineárně nezávislých proměnných neboli množinu **hlavních komponent**.

Analýza hlavních komponent se používá v případech, kdy počet proměnných u jednotlivých pozorování je velmi vysoký, nepřehledný a znemožňuje tak jejich jednoduchou interpretaci. Cílem PCA je tak nahrazení původních proměnných pozorování menším počtem jiných proměnných s co nejmenší ztrátou informace. Jednoduše řečeno analýza hlavních komponent snižuje dimenzi vstupních dat za co nejmenší ztráty informace [10].

5.2.2 Fourierova transformace

Fourierova transformace je jeden z hlavních nástrojů teorie zpracování signálů a teorie řízení. Dovoluje vzájemně jednoznačný převod signálu z časové reprezentace $\mathbf{f}(t)$ do reprezentace frekvenční $\mathbf{F}(\zeta)$. Převod mezi časově na frekvenčně závislé vyjádření je prováděn pomocí harmonických signálů **sin** a **cos**. Klasická Fourierova transformace vyžaduje znalost matematického vyjádření zpracovávaného signálu. Pokud však známe pouze vzorky signálu z konečného intervalu, neboli hovoříme o **diskrétním signálu**, pro který chceme získat vzorky frekvenčního spektra, musíme použít modifikované numerické metody známé pod jménem **diskrétní Fourierova transformace (DFT)**. Výpočet DFT je poměrně náročná operace vyžadující N^2 komplexní součtů a N^2 komplexních součinů. V roce 1965 proto došlo k optimalizaci algoritmu vyžadujícího pouze $N \log_2 N$ součtů a $N \log_2 N$ součinů. Tento optimalizovaný algoritmus byl nazván **rychlou Fourierovou transformací**.

Zobrazení signálu ve frekvenční oblasti se nazývá **spektogramem** někdy též **periodogramem**. Na obrázku 5.1 můžeme vidět signál lidského hlasu pohybující se v časové oblasti. Napravo od něj je zobrazen jeho periodogram s frekvencí na ose x a amplitudou v decibelech na ose y.



Obrázek 5.1: Periodogram signálu

Z periodogramu se často vychází při analýze signálu a následného **návrhu jeho parametrů** [2].

5.3 Vektor příznaků

Vektor příznaků (angl. *Feature Vector*) je opět pojem ze strojového učení a rozpoznávání. Jedná se o n-rozměrný vektor nejčastěji numerických hodnot (výše zmíněných parametrů), které reprezentují určitý objekt [8].

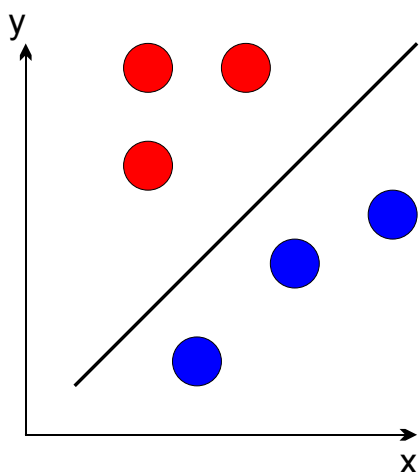
5.4 Klasifikace

Klasifikace v oblasti strojového učení znamená druh problému, kdy máme určit do které z kategorií neboli **tříd** spadá určité pozorování. Pro každou třídu je však nutné nejprve určit trénovací množinu pozorování, která do ní opravdu patří. Protože předkládáme trénovací množinu správně určených příkladů, hovoříme o tzv. **učení s učitelem**. Jednotlivá pozorování je nutné však nejprve kvantifikovat, tzn. číselně vyjádřit jejich vlastnosti. Algoritmus provádějící klasifikaci se nazývá **klasifikátor**. Z matematického hlediska se jedná o zobrazení vstupních dat na třídy. Existuje několik druhů klasifikací [8].

- Binární klasifikace - Klasifikuje pozorování do dvou tříd. Jedná se o nejjednodušší druh klasifikace.
- Diskrétní klasifikace - Třídí pozorování obecně do několik tříd.
- Vícetřídová klasifikace - Přiřazuje každému z pozorování obecně jednu a více tříd.
- **Jednotřídí klasifikace** - Na základě trénovacích pozorování má za úkol určit odlehle hodnoty jinak také **anomálie**. Z tohoto důvodu je obzvláště zajímavá pro účely **autentikace**. Založíme-li autentikační metodu například na otisku prstu, pak může odhalená anomálie znamenat otisk prstu neoprávněné osoby, ba dokonce podvodníka.

Surová číselná data pozorování se většinou nepoužívají na vstupu klasifikačního algoritmu. Vstupní data je vhodné nejprve **předzpracovat**, vyloučit anomálie a snížit jejich redundanci např. výše zmíněnou **extrakcí vektoru parametrů**. Snížíme tím i výpočetní dobu klasifikačního algoritmu.

Nejnámějším druhem klasifikačních algoritmů jsou takzvané **lineární klasifikátory** [8]. Tento druh klasifikátorů se dá zjednodušeně označit za lineární funkce. Tyto funkce vstupnímu klasifikovanému vektoru \vec{x} přiřazují skóre příslušnosti k jednotlivým třídám na základě skalárního součinu tohoto vektoru a vektoru vah klasifikátoru \vec{w} , který získáme naučením algoritmu pomocí množiny příznaků extrahovaných z trénovacích pozorování. Klasifikovaný vektor pak patří do třídy, pro níž dosáhl nejvyššího skóre. Při problému binární klasifikaci si můžeme představit operaci tohoto klasifikátoru jako rozdělování lineárního prostoru vstupních příznaků pomocí nadroviny viz obr. 5.2. Všechny body (vektory klasifikovaných pozorování) na jedné straně nadroviny jsou klasifikovány do první třídy, ostatní do třídy druhé. Mezi lineární klasifikátory patří například perceptron, LDA a v této práci použité **support vector machines** (SVM).



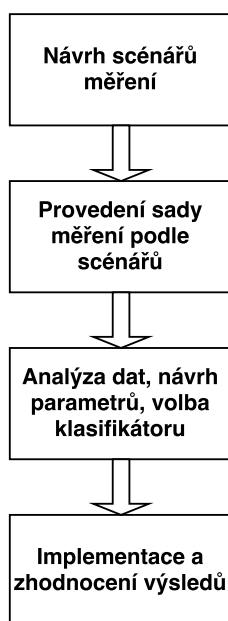
Obrázek 5.2: Binární lineární klasifikátor

Pokud máme však například třídy, které nejsou lineárně separovatelné nadrovinou, musíme využít klasifikátor nelineární. Příkladem takových algoritmů jsou SVM využívající kernelové metody [10].

6 Realizace

Cílem praktické části bakalářské práce je, jak už bylo zmíněno, provedení řady měření, návrh algoritmu pro získávání biometrických potenciálů z EEG signálu vhodných pro autentikaci, jeho implementace a následné otestování.

Vlastní práce se tedy skládá z několika částí zobrazených na obrázku 6.1. První částí je naměření vhodných EEG potenciálů použitelných pro biometrické účely. Pro potřeby akvizice těchto dat je nutno navrhnout scénáře měření, při nichž bude měřený subjekt stimulován určitým podmětem. Druhá část je věnována analýze naměřených dat a následnému návrhu parametrů na ní založeném. Třetí část je volba vhodného klasifikačního bloku, na němž bude postavena autentikační logika. Čtvrtou a poslední částí je implementace předešlých bodů, testování a zhodnocení získaných výsledků.



Obrázek 6.1: Schéma postupu při realizaci

Reálný postup samozřejmě **nebyl lineární**, vyžadoval řadu návratů od implementace zpět k novému návrhu. Schéma je zidealizované a slouží k vytvoření základní představy o dané úloze. Jsou uváděny hlavně návrhy, jejichž implementace dosahovala adekvátních výsledků. Proto v dalším textu navzdory tomu, že hovořím o návrzích, používám již minulý čas.

6.1 Scénáře měření

Při návrhu scénářů měření jsem se rozhodl zaměřit na oba dříve zmíněné přístupy při měření EEG aktivity. Tedy na snímání mozkových vln, jež vznikají, pokud je měřený subjekt zatížen určitou **fyzickou či mentální aktivitou** a na měření **klidových mozkových vln**. Ve všech případech byl měřený subjekt posazen do odhlučněné místnosti osvětlené pouze tlumeným denním světlem. Pokojová teplota se pohybovala **kolem 23 °C**. Subjekt dostal rovněž instrukce omezit co nejvíce fyzickou aktivitu v průběhu měření tzn. pohybu končetin, hlavy a především obličejových svalů (mrkání, pohyby čelistí etc.) pro snížení výskytu biologických artefaktů v nasnímaných datech.

Všechna měření probíhala v dopoledních hodinách. Pro každý měřený subjekt byly provedeny **2-3 sady měření s daným scénářem**. Sada se skládá ze **4-8 po sobě jdoucích měření** stejného scénáře. Jejich počet se odvíjel od aktuálního fyzického stavu měřeného jedince. Na svůj stav byl dotazován v pravidelných intervalech. Mezi jednotlivými měřeními byla ponechána doba pro odpočinek měřeného subjektu. Celková doba měření sady nikdy nepřesahovala 20 minut, aby nedošlo k únavě měřeného jedince a tím zkreslení výsledků. Mezi měřeními jednotlivých sad byl ponechán minimálně týdenní odstup. Cílem takto velké časové prodlevy bylo ukázat, že měřené charakteristiky nejsou krátkodobé. Měření se účastnilo celkem 8 subjektů. S každým scénářem bylo provedeno měření za účasti alespoň 5 subjektů.

U všech měřeních scénářů byl použit **elektroencefalograf** spolu s čepicí systémem 10-20 a nastavením uvedeným v tabulce 6.1.

Vzorkovací frekvence [Hz]	1000
Referenční elektroda	Fz
Uzemnění	levé ucho

Tabulka 6.1: Obecné nastavení encefalografu

Namapování elektrod čepice systému 10-20 na výstupní kanály uvádí tabulka 6.2. Při měření scénářů bylo použito **následující příslušenství**.

- **elektroencefalograf**

- EEG čepice odpovídající systému 10-20
- diferenční zesilovač
- abrazivní pasta
- elektrovodivý gel
- zemnicí elektroda

- ostatní

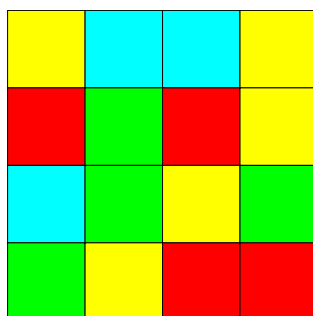
- osobní počítač
- Full HD monitor
- BrainVision Recorder 1.20

Číslo kanálu	Elektroda	Rozlišení [mV]
1	Fp1	100
2	Fp2	100
3	F3	100
4	F4	100
5	C3	100
6	C4	100
7	P3	100
8	P4	100
9	O1	100
10	O2	100
11	F7	100
12	F8	100
13	T3	100
14	T4	100
15	T5	100
16	T6	100
17	Cz	100
18	Fz	100
19	Pz	100

Tabulka 6.2: Namapování elektrod na kanály

6.1.1 Počítání barev

Měřenému subjektu byla na před ním stojícím monitoru promítána matice barevných čtverců velikosti 4x4. Rozložení barev v matici bylo generováno zcela náhodně. Cílem subjektu připojeného k elektroencefalografu bylo spočítat počet čtverců určité předem dané barvy. Po 4 sekundách se barevné rozložení matice přegenerovalo a úkolem měřeného subjektu bylo opět zjistit zobrazený počet dané barvy. Délka jednoho měření byla 50 sekund. Referenční barva, kterou měl subjekt spočítat, byla pouze jedna a byla dána na počátku měření. Tímto scénářem bylo naměřeno 5 subjektů.



Obrázek 6.2: Barevná matice

6.1.2 Představa pohybu rukou

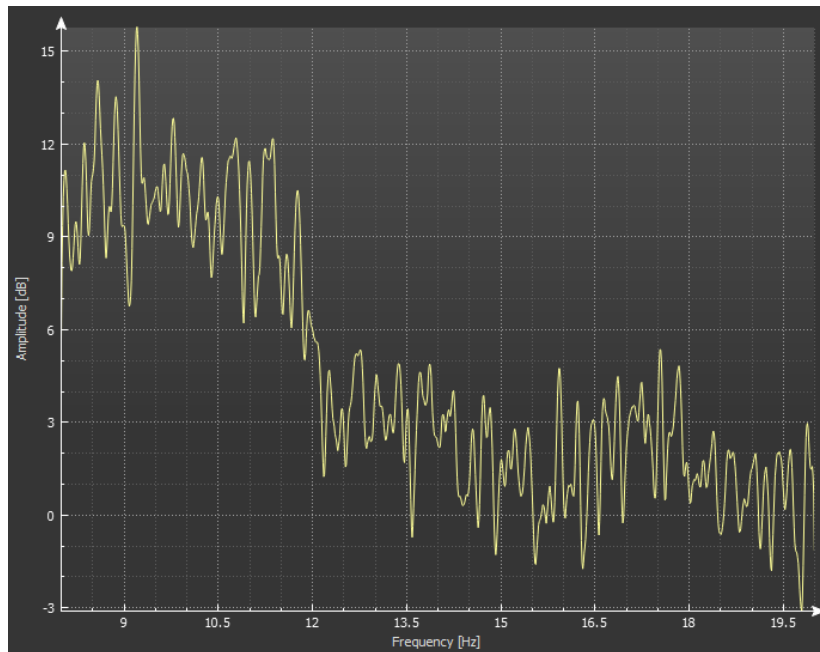
Měřenému subjektu byly na předním stojícím displeji zobrazovány na bílém pozadí střídavě siluety levé a pravé ruky. V jeden okamžik byla zobrazena pouze jedna ruka. Cílem měřeného subjektu bylo představovat si, jak se dotýká ruky zobrazované na obrazovce. Pokud je zobrazena pravá ruka, představoval si dotyk pravou rukou, pokud byla zobrazená levá ruka, představoval si dotyk rukou levou. Způsob dotyku byl ponechán úplně na měřeném subjektu. Jedinou podmínkou, bylo že zvolený postup by měl zůstat během všech měření přibližně stejný. Zobrazované ruce se střídaly po 4 sekundách. Celková doba jednoho měření byla 50 sekund. Při návrhu tohoto scénáře jsem vycházel z poznatku, že představa pohybu a pohyb samotný vyvolávají v mozku velmi podobnou reakci [9]. Na měření scénáře se podílelo 5 subjektů.

6.1.3 Zavřená oční víčka v bdělém stavu

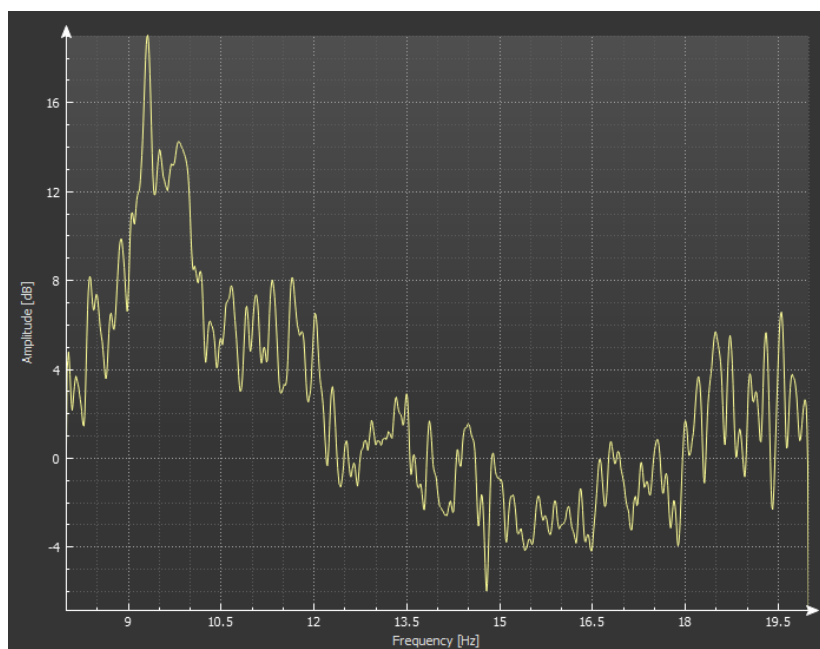
Měřený subjekt byl posazen do odhlučněné tmavé místnosti. Po připojení k encefalografu byl instruován zůstat nehybně sedět a zavřít oční víčka, po zavření očních víček započalo snímání mozkové aktivity subjektu. Po 50 sekundách bylo měření ukončeno a subjekt mohl oční víčka opět otevřít. Tímto způsobem bylo naměřeno 8 subjektů.

6.2 Parametry naměřených dat

6.2.1 Zavřená oční víčka



Obrázek 6.3: Spektrogram subjektu č.1

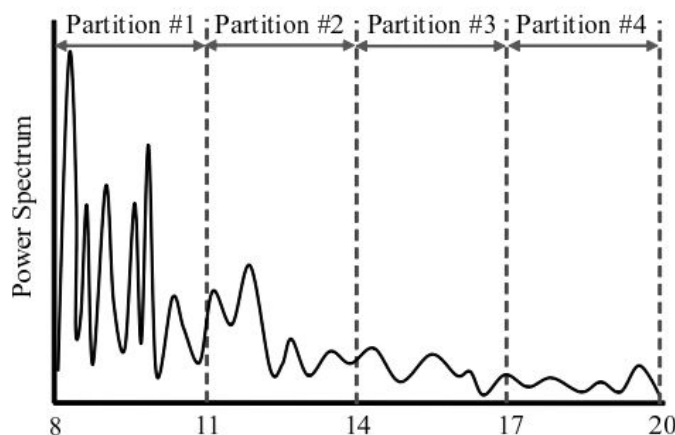


Obrázek 6.4: Spektrogram subjektu č.2

Analýza

Při analýze naměřených dat byl použit spektrogram postavený na rychlé Fourierově transformaci. Hodnoty spektrogramu byly navíc filtrovány **metodou klouzavého průměru**. Analýzou spektrogramů v závislosti na kanálu a frekvenčním rozsahu jsem došel z k zajímavému pozorování. U každého měřeného subjektu se ve frekvenčním rozsahu **8 až 20Hz** objevoval jistý vzor. Tento vzor byl nejvýraznější na kanálech odpovídajícím elektrodám **P3, P4, O1, O2**. Pozorovaný vzor se v rámci různých měření stejného subjektu téměř nelišil. Naopak se projevovaly odlišnosti v rámci vzorů různých subjektů. Na snímcích 6.3 a 6.4 jsou zobrazeny spektrogramy dvou rozdílných subjektů (více v příloze B).

Volba parametrů



Obrázek 6.5: Frekvenční pásy

Při návrhu parametrů získávaných z měření mozkové aktivity ze zavřených očních víčekách jsem vycházel ze studie Tottoriovy univerzity [5]. Hlavní myšlenkou jejich výzkumu je, že různý tvar křivky spektrogramu u navzájem různých jedinců zcela jistě znamená i různé rozložení energie na frekvenční ose těchto jedinců. Parametry jsem proto navrhl jako poměrné rozložení energie na frekvenčním rozsahu 8 až 20Hz, kde se vyskytuje inkriminovaný vzor. Toto frekvenční pásmo bylo rozděleno na několik intervalů (viz obr. 6.5) o různých velikostech. Na jednotlivých úsecích byla vypočtená energie definovaná jako součet kvadrátů amplitud odpovídajících frekvencím na daném intervalu. Energie v jednotlivých frekvenčních pásmech pak byly použity

jako složky příznakového vektoru. Pro potlačení rozdílností v rámci různých měření ovlivněných například různým odporem na elektrodách jsem zavedl do návrhu **normalizaci**, jednotlivé složky příznakového vektoru tak byly vyděleny celkovou energií na intervalu **8-20Hz**.

6.2.2 Počítání barev a představa pohybu

Analýza

Vlastní analýza naměřených dat ve frekvenčním spektru neukázala žádný pravidelně se vyskytující vzor spektrogramu či jiný viditelný úkaz, na němž by se dala postavit metoda extrakce parametrů.

Volba parametrů

Protože jsem se při návrhu nemohl opřít o žádné z mých vizuálních pozorování, musel jsem vycházet z již ozkoušených postupů. Uvádí se (např. [7]), že dříve popsané vlnové pásy pojmenované **delta**, **theta**, **alfa**, **beta** a **gama** mají vysoký biometrický potenciál ve frekvenčním spektru, jak je uvedeno i v následující citaci.

A spectral analysis approach examines the power spectrum of these various frequency bands - and can be used for person authentication. [7]

Návrh parametrů počítání barev a představy pohybu byl proto zvolen obdobně jako v případě návrhu parametrů signálu zavřených očních víček. Složky příznakového vektoru jsou opět tvořeny energiemi v pásích alfa, beta, gama, theta a delta. Do návrhu byla znovu zapojena normalizace příznakového vektoru.

6.3 Volba klasifikátoru

Při volbě klasifikátoru byl vybrán algoritmus umožňující **jednotřídní klasifikaci**. Jinými slovy klasifikační algoritmus, který dokáže na základě trénovacích pozorování, v našem případě na základě vektorů příznaků extrahovaných z jednotlivých měření určitého subjektu, rozhodnout, zda nově příchozí příznak patří do stejné třídy jako vzorové příznaky (tzn. patří tomuto subjektu), nebo do stejné třídy nepatří (tzn. nepřísluší tomuto subjektu nebo se jedná o anomálii). V kladném případě bude osoba, jíž patří nově příchozí příznak, autentizována pod identitou subjektu, jemuž patří vzorová data. Rovněž je pravděpodobné, že vstupní vektor nemusí být ve všech případech

lineárně oddělitelný od trénovacích vektorů. Hledáme tedy **nelineární klasifikátor**.

7 Biometrický systém

7.1 O aplikaci

Vlastní program je vzhledově a základní funkcionalitou koncipován jako **dvoufázový biometrický systém** (autentikace na základě jména a hesla následovaná biometrickou autentikací) s možností přihlášení administrátora. Cílem aplikace je převedením poznatků nabytých v teoretické části bakalářské práce do praktické roviny. Tedy implementace algoritmů pro extrakci parametrů, jejich následné využití pro autentikaci měřených subjektů a umožnění podrobného testování implementovaných metod.

7.2 Specifikace požadavků

Pro aplikaci pojmenovanou **Recognitor** byl vytvořen seznam požadavků, které by aplikace měla splňovat.

1. Přehrávání grafických stimulací a zvukových signálů potřebných při měření scénářů. V našem případě se jedná o zobrazování měnící se barevné matice, střídání siluet levé a pravé ruky. V neposlední řadě také přehrávání zvuků oznamující uživateli pokyny pro zavření a otevření očních víček.
2. Příjem dat z elektroencefalografu, jejich následná správa a archivace.
3. Vytvoření databáze přijatých dat, ta by měla rovněž obsahovat, kterému uživateli nasnímaná data patří a jaké jsou jejich parametry.
4. Pomocné nástroje pro analýzu naměřených dat.
5. Extrakce parametrů ze získaných dat.
6. Využití extrahovaných příznaků při autentikaci jednotlivých měřených subjektů.
7. Běh autentikačních testů a export výsledků.
8. Přívětivé GUI umožňující jednoduchou a intuitivní manipulaci s aplikací členům odborných skupin provádějících rozsáhlejší testy.

Jednotlivé požadavky byly vybrány tak, aby sloužily jako podpora výzkumné části bakalářské práce a ke zjednodušení manipulace s velkým množstvím získaných dat.

7.3 Návrh

Architektura programu bude rozdělena do tří vrstev podle klasického vzoru MVC (*Model-View-Controller*). Pro implementaci bude použit kompilovaný objektový programovací jazyk umožňující zapojení matematických knihoven a knihoven pro zpracování signálu.

7.3.1 Příjem dat

Podle požadavků bude Recognitor schopný příjmu dat přímo z elektroencefalografu. K tomuto účelu bude sloužit třída `RDARceiver` (*Remote data access receiver*) připojující se na pevně daném TCP/IP portu 51244 k serveru aplikace **BrainVision Recorder 1.20**, které dostává data přímo z encefalografu. Po připojení k tomuto serveru budou data přeposílána následně do Recognitoru, kde budou po dokončení příjmu uložena v binárním formátu `.eeg` patentovaném firmou BrainVision.

Přijatá data budou uvnitř programu reprezentována třídou `Dataset` obsahující jméno přijatého souboru, jeho umístění na disku, jméno subjektu, jehož měřením dataset vznikl a informaci o obsažených datech, např. vzorkovací frekvence, počet kanálů a formát `short` nebo `float` binární dat, které soubor obsahuje

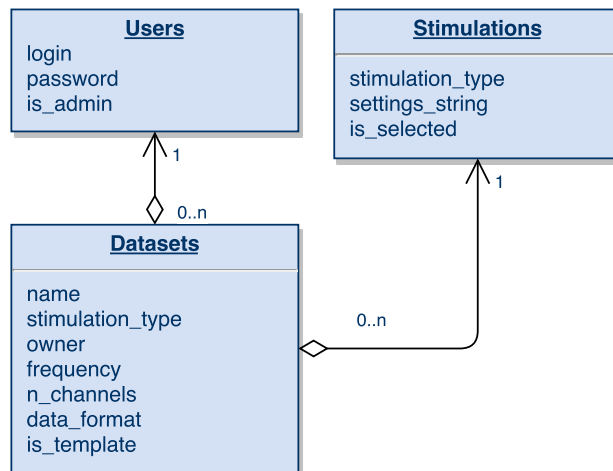
7.3.2 Databáze aplikace

V aplikaci bude vytvořena databáze podle dále uvedeného schématu na obrázku 7.1. Ta bude sloužit k ukládání přijatých dat, informací o měřených subjektech a nastavení stimulací. Přijatá data (`.eeg` soubory) jsou reprezentovány entitou `Datasets`, měřené subjekty, respektive uživatele, reprezentuje entita `Users`. Informace o stimulacích je uchována v entitě `Stimulation`.

Pro práci s databází bude v programu implementována třída `DbManager` zjednodušující vkládání a výběr dat.

7.3.3 Extrakce parametrů a autentikace

Nejprve bude navržena třída `AuthenticationEngine`, ta bude obsahovat metodu, jíž bude prvním parametrem předán seznam vzorových datasetů



Obrázek 7.1: Schéma databáze

a druhým argumentem bude dataset, o kterém chceme zjistit, zda patří do stejné třídy jako vzorové datasety. Toho docílíme pomocí třídy implementující rozhraní `IAuthenticator`, ve které se bude nacházet logika pro extrakci konkrétních příznaků a autentikaci. Pro urychlení autentikačního mechanismu bude rozhraní `IAuthenticator` obsahovat i metody pro cachování extrahovaných příznaků.

7.3.4 Analýza přijatých dat

Pro analýzu přijatých dat bude v programu implementována třída `SpectralDensityFFT` využívající **spektogram** postavený na rychlé Fourierově analýze. Bude umožňovat zobrazení spektogramu pro konkrétní kanál a frekvenční rozsah spolu s dalšími nástroji pro filtrování pohledu.

7.3.5 Testování autentikačních algoritmů

Pro potřeby testů bude v aplikaci vytvořena třída `TestThread`, jíž poskytnutá funkcionálníta poběží v samostatném vlákne. Pro zobrazení výsledků testů bude využita třída `HTMLTestResultExporter` zapisující výsledky testů do přehledných tabulek ve formátu `html`.

7.4 Realizace

Program byl implementován v programovacím jazyku Microsoft Visual C++. Tento jazyk byl vybrán pro efektivní překlad kódu a s ním spojenou výkonnost, jež byla při zpracování velkého množství binárních dat potřeba. Další

výhodou je jeho objektový přístup a možnost využití velkého množství externích knihoven, například použité knihovny Dlib obsahující klasifikační algoritmy. Spolu s Visual C++ jsem užil knihovnu Qt 5.6.0 pro její pokročilé možnosti tvorby GUI, práce s multimédií, práce s vlákny a databázemi.

Při realizaci prezenční vrstvy jsem aplikoval výše zmíněnou knihovnu Qt 5.6.0. GUI umožňuje zobrazení seznamu měřených subjektů a jejich datasetů, dále editaci nastavení stimulací, zobrazování spektrogramů, zjednodušuje správu velkého množství dat a mnohé další. Celé GUI je intuitivní a uživatelsky přívětivé. Veškeré jeho možnosti jsou popsány v uživatelském manuálu.

O realizaci datové vrstvy se stará třída `DbManager` manipulující s databází. Pro svou jednoduchost a souborový přístup byla použita databáze SQLite 3. Celou databázi aplikace reprezentuje tedy pouze jeden soubor opatřený koncovkou `.db`.

Původní návrh aplikace byl plně dodržen a plnil všechny předem stanovené požadavky.

7.5 Testování

Aplikaci jsem inkrementálně testoval během jejího vývoje pomocí debuggeru ve vývojovém prostředí Qt Creator. Program rovněž obsahuje i trasovací výpisy zpřehledňující tok programu. Na jejich implementaci byla použita knihovna `QDebug`, prokázala se jako obzvláště užitečná. Finální verze byla testována několik studenty Fakulty aplikovaných věd ZČU. Nebyly objeveny žádné chyby.

7.6 Použité nástroje

Program byl původně vyvíjen pod systémem Ubuntu 14.04LTS. Postupem času jsem přešel pod vývoj ve Windows 7 Professional SP1. Program navzdory použitému frameworku není multiplatformní. Jeho překlad a spuštění je možné pouze na platformě Windows a s nástroji Microsoft Visual C++.

Recognizer byl vyvíjen ve vývojovém prostředí Qt Creator 3.6.1. Pro překlad finální verze programu byl použit překladač Microsoft Visual Studio C++ 2015 x64. Pro spuštění programu je tedy potřeba nainstalovat distribuované balíčky jazyka Visual C++ obsahující runtime moduly potřebné ke spouštění aplikací napsaných v tomto jazyce.

7.7 Uživatelská dokumentace

Uživatelská dokumentace obsahující ovládání programu a pokyny pro překlad a spuštění je součástí příloh.

8 Autentizační testy

Logika testů je následující. V aplikaci existuje pro každého uživatele seznam jeho naměřených datasetů v závislosti na použitém scénáři měření. Konkrétně má každý uživatel 3 skupiny datasetů podle použité stimulace, tzn. datasety zavřených očních víček, počítání barev a datasety představy pohybu. V každé skupině existuje několik datasetů (alespoň 3) označených jako **vzorové** (template). Vektory příznaků extrahovaných z těchto datasetů jsou použity pro naučení klasifikačního algoritmu, který na tomto základě je schopen autentizovat či zamítnout nově příchozí dataset určitému subjektu. Aby testy prokázaly, že získávané charakteristiky skutečně náležejí měřenému subjektu a nejsou pouze charakteristikou daného měření, byl výběr vzorových datasetů omezen pouze na první sadu měření subjektu pro každou stimulaci. Celkový počet vzorových datasetů byl limitován na **3 až 5**.

Uváděné data jsou výsledkem autentizačních testů prováděných a vyexportovaných vlastní aplikací Recognitor. Konkrétní implementace zvoleného nelineárního jednotřídního klasifikátoru je tzv. **radial basis kernel svm** z knihovny **Dlib**. Volba konkrétní implementace klasifikátoru a jeho nastavení na rozlišení **kernel 4.0** bylo zvoleno empiricky na základě průběžných výsledků exportovaných z aplikace.

Po naučení trénovací množinou **příznakových vektorů** určitého subjektu vrací použitý klasifikátor matematickou funkci, kde na jejím vstupu je nově příchozí vektor a výsledkem funkce je skalární hodnota vyjadřující jeho příslušnost danému subjektu. Pro účely testů byla tato hodnota nazvána jako **score**. Dále definujeme hranici, kterou musí **score** dosáhnout, aby byl dataset klasifikačním algoritmem autentikován tomuto subjektu, nazvanou jako **threshold**. Pokud hovoříme o autentikaci určitého datasetu uživateli, myslíme tím rozhodnutí klasifikačního algoritmu, že dataset skutečně obsahuje mozkovou aktivitu tohoto subjektu.

8.1 Formát výsledků

Vzhledem k relativně malému množství měřených subjektů (5 až 8) jsou výsledky testů uváděny jednotlivě pro všechny měřené subjekty, proto byl zvolen tabulkový zápis výsledků viz tabulka 8.1.

Jednotlivé procentuální údaje úspěšnosti v tabulce mají dvojitou interpretaci v závislosti na jejich poloze.

	Subjekt 1	Subjekt 2	Subjekt 3
Subjekt 1	90%	94%	82%
Subjekt 2	82%	95%	91%
Subjekt 3	71%	96%	72%

Tabulka 8.1: Způsob zápisu výsledků

- **diagonální hodnoty** - Vyjadřují úspěšnost autentikace uživatelových datasetů na jeho identitu. Tato hodnota je dána matematickým vztahem viz vzorec 8.1.

$$\text{úspěšnost} = 100 * \frac{\text{počet_autentikovaných_datasetů}}{\text{počet_všech_datasetů}} \quad (8.1)$$

Kde ve jmenovateli figuruje počet všech datasetů tohoto uživatele a čísel vyjadřuje, kolik z těchto měření bylo tomuto uživateli úspěšně autentikováno.

- **nediagonální hodnoty** - Vyjadřují odolnost subjektu, uvedeném ve sloupci dané hodnoty, proti falešné autentikaci subjektu v odpovídající řádce. Výpočet probíhá pomocí vzorce 8.2.

$$\text{odolnost} = 100 * \frac{\text{počet_zamítnutých_datasetů}}{\text{počet_všech_datasetů}} \quad (8.2)$$

V čitateli se nachází počet měření subjektu uvedeného v odpovídající řádce zamítnutých při pokusu o autentikaci na subjekt v odpovídajícím sloupci. Jmenovatel vyjadřuje celkový počet datasetů uživatele v řádce.

Například hodnota **71%** v tabulce 8.1 označuje, že 71 procent datasetů subjektu č. 3 se nepodařilo autentikovat na uživatele č. 1.

8.2 Zavřená oční víčka

Implementovaná extrakce parametrů z datasetů naměřených pomocí stimulace zavřených očních víček vychází z návrhu zmíněného v pododdíle 6.2.1. Signál na dříve zmíněných elektrodách **P3**, **P4**, **O1** a **O2** byl zprůměrován a pro výsledný průměr byl proveden spektrogram zachycující interval 8-20 Hz. Tento interval byl rozdělen na šest nepřesahujících pásů o velikosti 2Hz. Tyto frekvenční pásy pak byly použity ke zformování vektoru příznaků podle předešlého návrhu. Výsledný vektor měl 6 složek.

8.2.1 Výsledky

Uváděné výsledky se drží schématu uvedeného v oddílu 5.2. Měření zavřených očních víček se zúčastnilo celkem 8 subjektů. Každý jedinec měl v době provádění testů 8-15 datasetů pocházejících z 2-3 sad měření (viz oddíl 5.1). **Threshold** byla stanovena na **-0.001**.

	Martin	Matěj	Petr	Livien	Anastasia	Kristýna	Štěpán	Tony
Martin	100%	100%	100%	100%	0%	100%	100%	100%
Matěj	100%	100%	43%	100%	79%	100%	100%	100%
Petr	91%	64%	91%	82%	100%	73%	100%	100%
Livien	100%	100%	86%	100%	100%	0%	100%	44%
Anastasia	17%	83%	100%	100%	100%	100%	100%	100%
Kristýna	100%	100%	83%	75%	100%	83%	100%	92%
Štěpán	100%	100%	100%	100%	100%	100%	83%	92%
Tony	100%	100%	100%	100%	100%	100%	100%	100%

Tabulka 8.2: Výsledky zavřených očních víček

8.2.2 Diskuze

Z dosažených výsledků uvedených v tabulce 8.2 je patrné, že navržená a následně implementovaná metoda jeví určitý biometrický potenciál. Úspěšná autentikace subjektu na svou skutečnou identitu se pohybuje v rozmezí 83-100%. **FNMR** v rámci různých subjektů leží tedy mezi 0-17%.

Problém této metody se ukazuje v případě autentizace subjektů na identitu jiného subjektu v biometrickém systému. Ačkoliv se v drtivé většině případů nepodaří jednomu subjektu autentikovat pod identitou subjektu druhého, nastávají případy označené v tabulce 8.2 červenou barvou, kdy docházelo ve většině případů k úspěšné autentizaci jednoho subjektu na identitu jiného subjektu. Jako příklad můžeme uvést dvojici uživatelů Martin a Anastasia. Pouze 17% případů autentikace uživatele Anastasia na identitu Martina bylo zamítnuto, v opačném případě byl uživatel Martin vždy autentikován pod identitou Anastasie. Tento jev pramení z faktu, že pozorovaný vzor ve frekvenčním pásu 8-20Hz není unikátní, a proto rozlišení některých subjektů pomocí této metody je velmi obtížné. V těchto případech metoda selhává. Za výhodu tohoto přístupu naopak považují použití nízkého počtu elektrod, jenž by ulehčily zapojení v reálném provozu.

8.3 Představa pohybu a počítání barev

Pro měření získané pomocí stimulace představy pohybu a počítání barev byla navržena a implementována stejná metoda extrakce parametrů. Implementace vychází z návrhu v pododdílu 6.2.2. Pro každou z 19 elektrod čepice systému 10-20 bylo zjištěno poměrné rozložení energie v pásech alfa, beta, gama, delta a theta. Získal jsem tak pro každou elektrodu vektor o pěti složkách. Postup byl opakován pro všech 19 elektrod a složky získaných vektorů byly spojeny do výsledného vektoru příznaků celkově o 95 složkách (5x19).

8.3.1 Výsledky

Uváděné výsledky se opět drží schématu uvedeného v oddílu 5.2. Měření se zúčastnilo celkem 5 subjektů. Každý jedinec měl v době provádění testů 7-15 datasetů pocházejících z 2-3 sad měření (viz oddíl 5.1). **Threshold** byla stanovena na **-0.006**.

	Anastasia	Kristýna	Livien	Petr	Jitka
Anastasia	90%	100%	100%	100%	100%
Kristýna	100%	80%	100%	100%	100%
Livien	100%	84%	91%	100%	100%
Petr	100%	100%	86%	100%	100%
Jitka	100%	100%	100%	100%	90%

Tabulka 8.3: Výsledky počítání barev

	Anastasia	Kristýna	Livien	Petr	Jitka
Anastasia	72%	100%	100%	100%	100%
Kristýna	100%	80%	54%	100%	93%
Livien	100%	84%	91%	100%	100%
Petr	100%	100%	100%	100%	100%
Jitka	100%	100%	89%	100%	78%

Tabulka 8.4: Výsledky představování pohybu ruky

8.3.2 Diskuze

Z výsledků uvedených v tabulkách 8.3 a 8.4 je opět patrné, že tyto metody vykazují biometrický potenciál. Úspěšná autentikace jedinců na svou skuteč-

nou identitu se pohybuje pro obě metody v rozmezí 72-100%. V závislosti na jedinci se tedy **FNMR** pohybuje mezi 0-28%.

Obě metody vykazují i odolnost vůči pokusům o autentikaci jednoho subjektu, například potenciálního podvodníka, na identitu druhého, jehož data jsou zanesena v biometrickém systému. **FMR** se tedy v závislosti na dvojici subjektů pohybuje v rozmezí 0-46%. Nejhorších výsledků dosahovala dvojice Kristýna a Livien v tabulce 8.4, kdy bylo pouze 54% pokusu o autentikaci Kristýny na identitu Livien zamítnuto. Na druhou stranu nebyl objeven případ, kdy by navržené metody nebyly schopny rozlišit dva různé subjekty a docházelo by tak k jejich naprostému selhání. Pro podrobné zhodnocení těchto metod je potřeba dalších měření. Na základě zatím dosažených výsledků se jeví obě metody perspektivně z hlediska biometrické autentizace. Doporučuji jejich další průzkum případným zájemcům o danou tematiku.

9 Závěr

Tato bakalářská práce si vytyčila několik důležitých cílů. Mezi ně patřilo prostudování literatury týkající se biometrické autentizace osob na základě EEG a EKG potenciálů, následný návrh a implementace algoritmů pro extrakci parametrů z EEG signálu, důkladné otestování implementace provedením sady měření a dokumentace dosažených výsledků při zvoleném postupu. Hlavním smyslem bakalářské práce bylo zjištění, zda lze počítat s těmito implementacemi algoritmů extrakce parametrů pro následné rozsáhlé testy prováděné odbornou skupinou.

Vlastní práce je tedy rozdělena do několika částí. První část je věnována obecným pojmům z biometrie a autentizace. Následuje část zabývající se výše zmíněnými biologickými potenciály EEG a EKG. Získané teoretické znalosti jsou následně použity při návrhu algoritmů pro extrakci parametrů z EEG záznamu a návrhu scénářů měření sloužících pro stimulaci mozkové aktivity, které jsou vhodné pro získání biologických parametrů ze signálu. Navržené metody jsou plně implementovány ve vlastní aplikaci a výsledná implementace otestována.

Testy implementovaných metod stimulace a extrakce ukázaly, že navržené metody skutečně mají biometrický potenciál a stojí za zevrubnější otestování provedené odbornou skupinou. K tomuto účelu byla i přizpůsobena vlastní aplikace nazvaná Recognitor. Slouží k uchování naměřených potenciálů, promítání stimulací a k automatizaci autentikačních testů a exportu jejich výsledků usnadňujících práci odborné skupiny.

Práci považuji za úspěšnou. Splňuje jak cíle stanovené v zadání bakalářské práce, tak i mnou předem stanovené požadavky a návrh.

Literatura

- [1] DRAHANSKÝ, Martin; ORSÁG, Filip. *Biometrie*. 1. vyd. Brno: Computer Press, 2011. 294 s. ISBN 978-80-254-8979-6.
- [2] SÖRNMO, Leif; LAGUNA, Pablo. *Bioelectrical Signal Processing in Cardiac and Neurological Applications*, 1. vyd. New York: Academic Press, 2005. 688 s. ISBN 978-0124375529.
- [3] TEPLAN, Michal. Fundamentals of EEG measurement. In *Measurement Science Review*, 2002, volume 2, section 2, s. 90-102.
- [4] PARANJAPE, B. Raman; BENEDICENTY, L.; MAHOVSKY, Jan. The electroencephalogram as a biometric. In *Canadian Conference on Electrical and Computer Engineering*, 2001, volume 2, s. 1363-1366.
- [5] NAKANISHI, Isao; OZAKI, Koutaro; SHIGANG, Li. Evaluation of the Brain Wave as Biometrics in a Simulated Driving Environment. In *International Conference of the Biometrics Special Interest Group*, 2012, volume 1, s. 350-357. ISBN 978-1-4673-1010-9.
- [6] CAMPISI, P.; SCARANO, G.; BABILONI, F.; COLONNESE, S.; MAIORANA, Enrico. Brain waves based user recognition using the Eyes Closed Resting Conditions protocol. In *IEEE International Workshop on Information Forensics and Security*, 2011, volume 1, pp. 1-6.
DOI:10.1109/WIFS.2011.6123138
- [7] TRAORE, Issa; AHMED, Ahmed E. Awad. *Continuous Authentication Using Biometrics: Data, Models, and Metrics*. 1. vyd. Pennsylvania: IGI Publishing Hershey, 2011, 294 s. ISBN-13 978-1613501290
- [8] XINDONG, Wu; VIPIN, Kumar. *The Top Ten Algorithms In Data Mining*. 1. vyd. Abingdon: Chapman and Hall, 2009, 232 s. ISBN 9781420089646
- [9] PFURTSCHELLER, C.; NEUPER, C.; Flotzinger, D. EEG-based discrimination between imagination of right and left hand movement. In *Electroenceph. Clin. Neurophys.*, Vol. 103, 1997, pp. 642-651.

- [10] ELDÉN, Lars. *Matrix Methods in Data Mining and Pattern Recognition*. 1. vyd. Philadelphia: SIAM, 2007. ISBN-13 978-0898716269.
- [11] DEY, Monalisa; DEY, Nilanjay; ACHARJEE, Suvojit. Electrocardiogram Feature Based Inter-human Biometric Authentication System. In *International Conference on Electronic Systems, Signal Processing and Computing Technologies*, 2004, volume 1, s. 300-304. DOI:10.1109/ICESC.2014.57
- [12] FOURNIER, R.; NAIT-ALI, A.. A novel biometric authentication approach using ECG and EMG signals. In *Journal of Medical Engineering*, 2005, volume 4, s. 226-238. DOI:10.3109/03091902.2015.1021429

A Uživatelská dokumentace

A.1 O aplikaci

Aplikace Recognitor 1.0 byla vyvinuta v rámci bakalářské práce *Biometrická autentizace osob na základě EEG signálu* na Fakultě aplikovaných věd Západočeské univerzity. Cílem bakalářské práce bylo:

- Návrh vhodného algoritmu pro získávání biometrických parametrů z EEG signálu.
- Implementace navrženého algoritmu a důsledné otestování výsledné implementace
- Provedení sady měření, pomocí nichž zjistíte, zda je možné počítat s implementací této metody pro rozsáhlé testy odbornou skupinou.
- Usnadnění práce odborné skupiny.

Aplikace Recognitor má za úkol: promítání grafických stimulací měřenému subjektu, přímý příjem dat z elektroencefalografu v podobě datasetů obsahujících měřenou mozkovou aktivitu, správu naměřených datasetů, extrakci parametrů z naměřených dat, autentikaci extrahovaných parametrů, běh autentikačních testů a export jejich výsledků. **Pro plné porozumění funkčnosti a kontextu aplikace je silně doporučeno přečtení textu bakalářské práce.** Svou formou je aplikace pojata jako dvoufázový biometrický systém. **Měřené subjekty** zde vystupují v roli jejich **uživatelů** a musejí být instruováni o scénářích navržených v bakalářské práci. Subjekt/uživatel připojený k elektroencefalografu se zadáním kombinace jména a hesla pokouší o přihlášení do aplikace. Aplikace vyzve uživatele k přípravě na scénář nastavený v aplikaci. Může se jednat o představu pohybu rukou, počítání barev nebo zavřená oční víčka (viz. text bakalářské práce), poté započne samotné měření. Pokud již byl uživatel v minulosti naměřen touto aplikací pro zvolený scénář, tak dojde k pokusu o jeho autentikaci na identitu zadanou kombinací přihlašovacího jména a hesla. Dále zde vystupuje osoba **administrátora**. Administrátor vybírá stimulaci použitou pro autentikaci subjektů, nastavuje parametry stimulací, spravuje naměřená data a má právo spustit autentikační testy.

A.2 Překlad

Překlad aplikace je možný pouze na platformě **Windows**. Pro úspěšný překlad zdrojových kódů potřebujeme mít nainstalován **Qt Creator 3.6.1** postavený na knihovně Qt 5.6.1 spolu s překladačem **Microsoft Visual C++ Compiler 2015**.

- Nejdříve se v příkazovém řádku přepneme do adresáře obsahující zdrojové kódy aplikace. Poté vygenerujeme **makefile** aplikace následujícím příkazem.

```
/> qmake -o Makefile recognitor.pro
```

- Finální překlad provedeme příkazem `/> nmake` ve stejné adresáři.

A.3 Požadavky aplikace

A.3.1 Softwarové požadavky

Operační systém **Microsoft Windows 7** (a vyšší) s nainstalovaným balíčkem runtime modulů **Microsoft Visual 2015 C++ Redistributable**.

A.3.2 Hardwarové požadavky

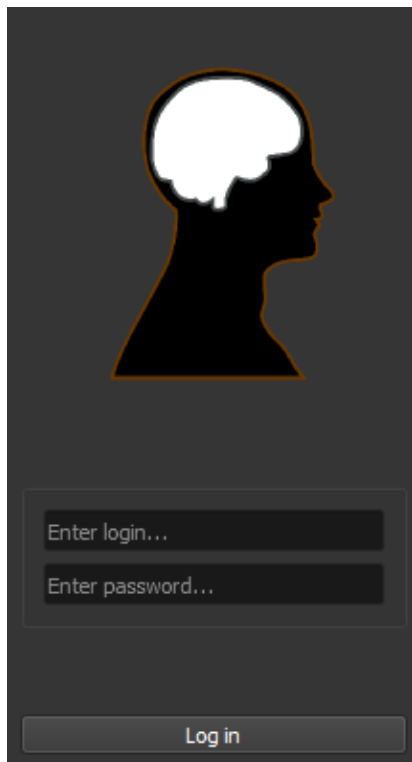
- **Processor:** Intel nebo AMD x86 procesor podporující instrukční sadu SSE2
- **RAM:** 1024MB (doporučená pamět 2048MB)
- **HDD:** 1GB + místo pro ukládání datasetů
- **GPU:** doporučená grafická karta podporující OpenGL 3.3

A.4 BrainVision server

Aby aplikace mohla přijímat data měřeného subjektu z encefalografu, musí mít možnost připojení k BrainVision serveru. V případě, že máme nainstalovaný software **BrainVision 1.20** na lokálním počítači nebo počítači přístupném po síti, musíme provést jeho konfiguraci. Nejprve vybereme workspace aplikace (**File > Open Workspace**) obsahující nastavení měření. Workspace najdeme jako součást instalace. Poté zapneme BrainVision server kliknutím na ikonu oka v levém horním rohu. Recognitor je nyní po zadání správné IP adresy počítače s BrainVision softwarem schopný příjmu dat z encefalografu.

A.5 Ovládání

Spuštění provedeme dvojklikem na soubor **Recognitor.exe** nesoucí ikonu s logem programu. Dojde k zobrazení přihlašovací obrazovky programu viz obr. A.1. Ovládání programu se dále liší, podle toho zda se přihlašujeme na účet uživatele nebo administrátora. Pro prvotní přihlášení administrátora použijte login **admin** a heslo **admin**.



Obrázek A.1: Přihlašovací okno

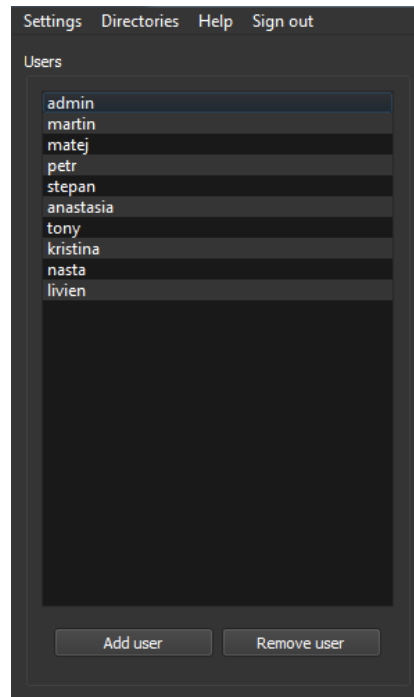
A.5.1 Administrátor

Administrátorské okno

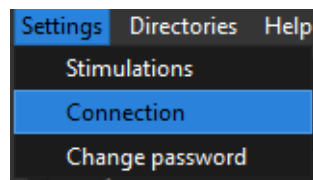
Při přihlášení na administrátorský účet dojde k zobrazení administrátorského okna viz obrázek A.2.

V centrální části je **seznam registrovaných uživatelů**, v dolní části jsou **tlačítka pro přidání a mazání** uživatelů. V horní liště se nachází menu *Settings* (viz obr. A.3) umožňující editovat nastavení stimulací, připojení k BrainVision serveru a podobně. Vedle se nachází menu *Directories* (viz obr. A.4). To umožňuje zobrazení **nejdůležitějších složek programu**:

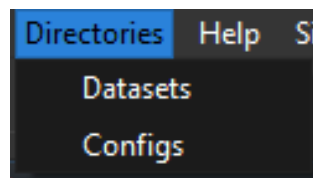
- **Datasets** - Složka kam se umisťují veškeré datasey, se kterými aplikace pracuje.
- **Configs** - Složka obsahující konfigurační *xml* soubor a databázi aplikace.



Obrázek A.2: Administrátorské okno



Obrázek A.3: Settings menu



Obrázek A.4: Directories menu

Poslední menu *Sign out* (viz obr. A.5) umožňuje odhlášení z administrátorského okna zpět do přihlašovací obrazovky.



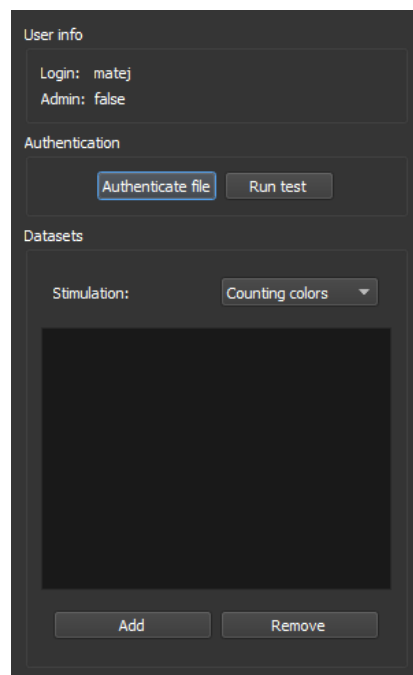
Obrázek A.5: Sign out menu

Přidání datasetu

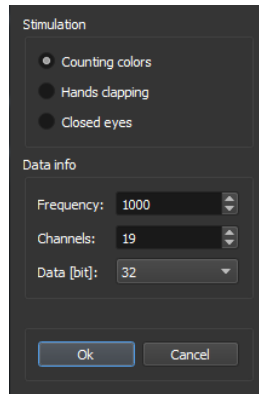
V **administrátorském okně** vybereme uživatele, jemuž chceme dataset přidat, a dvojklikem otevřeme detail uživatele (viz obr. A.6). Dole stiskneme tlačítko *Add*. Dojde k zobrazení dialogu pro výběr souboru ve **složce s datasety**. Po výběru datasetu dojde k zobrazení dialogu pro výběr jeho parametrů (viz A.7) pro potřeby aplikace.

V dialogu vybereme **stimulaci**, za jejíž působení byl dataset změřen, a informace o datasetu, tzn. **vzorkovací frekvenci** a **počet kanálů** použitých při měření datasetu encefalografem. Následuje výběr **binárního formátu** datasetu - 16b nebo 32b slova.

Pokud chceme přidat externí dataset do databáze, musíme ho nejprve vložit do složky s datasety aplikace.



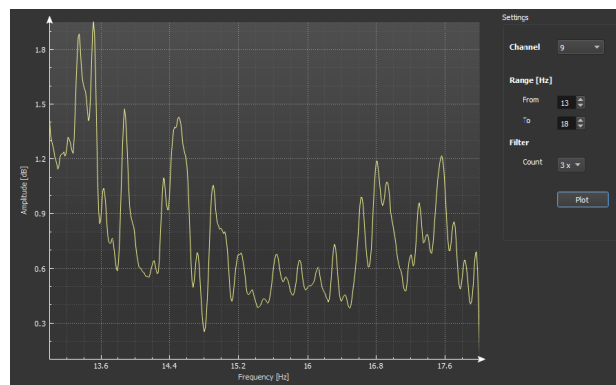
Obrázek A.6: Detail uživatele



Obrázek A.7: Dialog parametrů datasetu

Spektrogram

Recognitor umožňuje zobrazení **spektrogramu** (viz obr. A.8) pro jednotlivé datasety. Spektrogram je vizuální reprezentace spektra frekvencí v signálu. K zobrazení spektrogramu se dostaneme dvojklikem na jméno datasetu v detailu uživatele, dojde k zobrazení informačního dialogu. V dolní části dialogu je tlačítko *View spektrogram*, které provede zobrazení spektrogramu pro vybraný dataset. V pravém horním rohu spektrogramu můžeme vidět, výběr **kanálů**, zobrazovaný **frekvenční rozsah** a **násobnost filtrace**, kterou chceme nad spektrogramem provést.



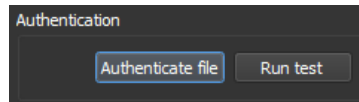
Obrázek A.8: Spektrogram

A.5.2 Autentikace datasetu

Aby bylo možné provedení autentikace datasetu uživateli, musíme nejdříve naměřit alespoň 3 vzorové datasety sloužící k natrénování klasifikátoru a označit je jako **template** v detailu datasetu. Template označuje, že vek-

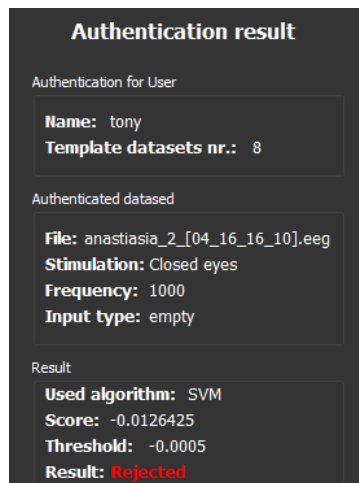
tor parametrů extrahovaný z daného datasetu bude použit pro natrénování klasifikátoru.

Pokus o autentikaci datasetu probíhá obdobně jako přidání datasetu do aplikace. Nejříve si v *administračním okně* vybere uživatele, kterému chceme dataset autentikovat. V detailu uživatele poté klikneme na tlačítko *Authenticate file* viz obr. A.9.



Obrázek A.9: Autentikace datasetu

Po krátké chvíli dojde k zobrazení okna s výsledky (viz obr. A.10). V okně s výsledkem autentikace jsou nejprve zobrazeny informace o uživateli, na něhož jsme se pokoušeli daný soubor autentikovat. Následují informace o autentikovaném datasetu a úplně na konci jsou uvedeny hodnoty výsledku, tedy **algoritmus** použitý při autentikaci, **dosažené skóre** a **hranice** nutná k dosažení, aby byl dataset autentikován.

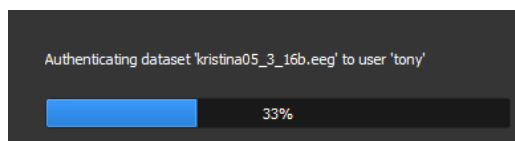


Obrázek A.10: Výsledek autentikace

Pokud chceme autentikovat externí dataset, musíme ho nejprve vložit do složky s datasy aplikace.

Autentikační testy

V detailu uživatele je vedle dříve zmíněného tlačítka *Authenticate file* tlačítko pro spuštění testů *Run test*. Při kliknutí nám dá program rozhodnout pro jakou stimulaci chceme provést autentikační test. Po spuštění testů se zobrazí



Obrázek A.11: Postup testů

okno s jejich postupem (viz obr. A.11). Po dokončení testů dojde k zobrazení dialogu pro uložení výsledků. Výsledky jsou ukládány ve formátu *html*.

Export výsledků

Export výsledků autentikačních testů pro určitého uživatele se skládá ze tří částí. První částí je tabulka výsledků, kdy se uživatel, pro něhož provádíme testy, pokoušel autentikovat na ostatní uživatele programu (viz obrázek A.12). Přesnost obsažená v tabulce vyjadřuje, kolik datasetů z celkového počtu měření tohoto uživatele se nepodařilo autentikovat na jiného uživatele uvedeného v řádce tabulky.

Authentication on other users

User	Accuracy [%]
admin	NETD
anastasia	100 (12/12)
livien	83.3333 (10/12)
martin	100 (12/12)
matej	100 (12/12)
nasta	NETD
petr	100 (12/12)
stepan	100 (12/12)
tony	100 (12/12)

Obrázek A.12: Autentikace na ostatní uživatele

Druhou část tvoří tabulka výsledků (viz obr. A.13), kdy se ostatní uživatelé systému pokoušeli autentikovat na uživatele, pro kterého je prováděn export. Přesnost je vyjadřována stejnou mírou jako v předcházejícím případě. Poslední tabulka (viz. A.14) vyjadřuje, kolik datasetů uživatele, pro něhož jsme spustili testy, se podařilo autentikovat na jeho osobu. Přesnost je zde vyjádřena jako podíl počtu jeho datasetů, které se povedlo autentikovat na jeho identitu, a celkového počtu jeho datasetů.

Pokud místo procent dojde k zobrazení **NETD** (*Not enough template datasets*) nebo **NEUD** (*Not enough users datasets*) oznamuje tím aplikace, že pro dané uživatele není zanesen dostatečný počet datasetů pro provedení testů.

Other users on subject

User	Accuracy [%]
admin	NEUD
anastasia	100 (6/6)
livien	42.8571 (3/7)
martin	100 (6/6)
matej	100 (14/14)
nasta	NEUD
petr	100 (11/11)
stepan	100 (12/12)
tony	100 (12/12)

Obrázek A.13: Autentikace na ostatní uživatele

Self authentication

Accuracy [%]
100 (12/12)

Obrázek A.14: Autentikace sama na sebe

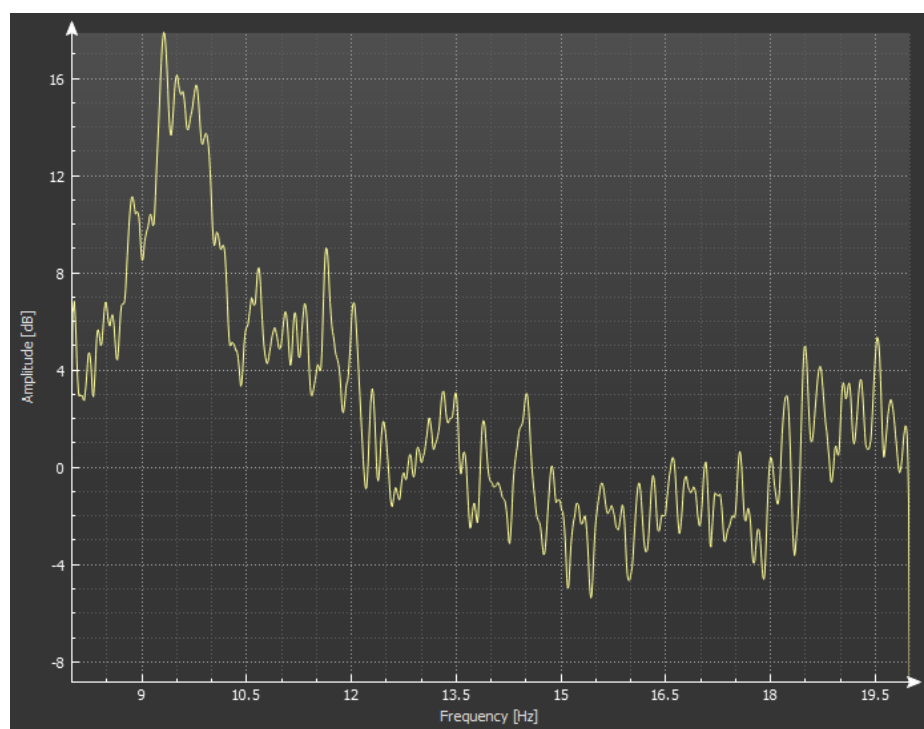
A.5.3 Běžný uživatel

V roli běžného uživatele vystupuje v aplikaci Recognitor měřený subjekt. Jediným jeho úkolem je pokus o přihlášení do biometrického systému, znalost scénářů a jejich dodržování.

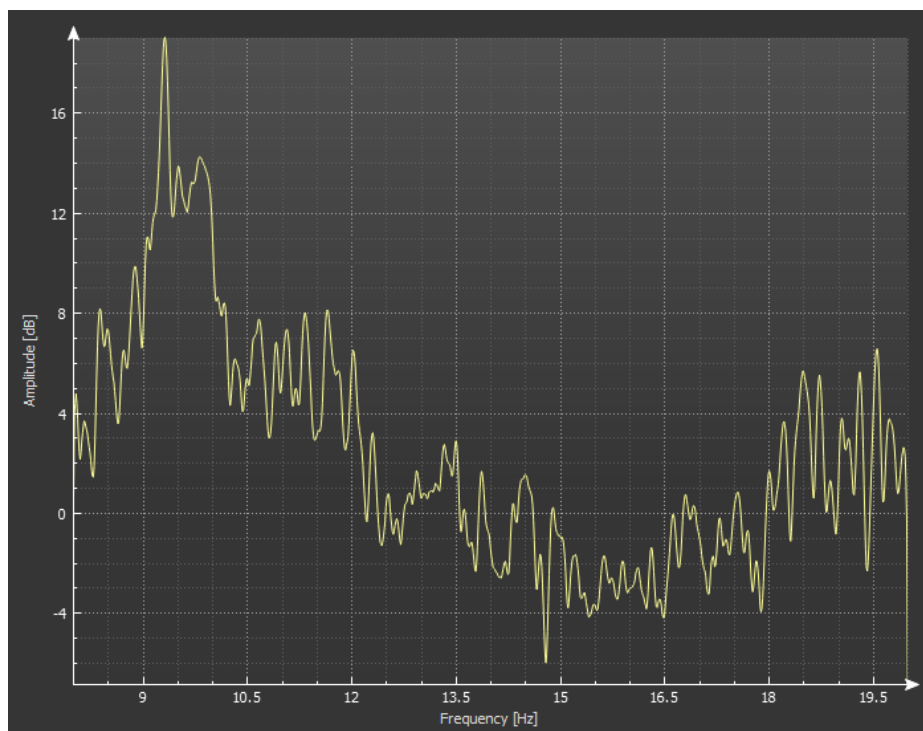
B Spektrogramy EEG signálu

B.0.1 Zavřená oční víčka

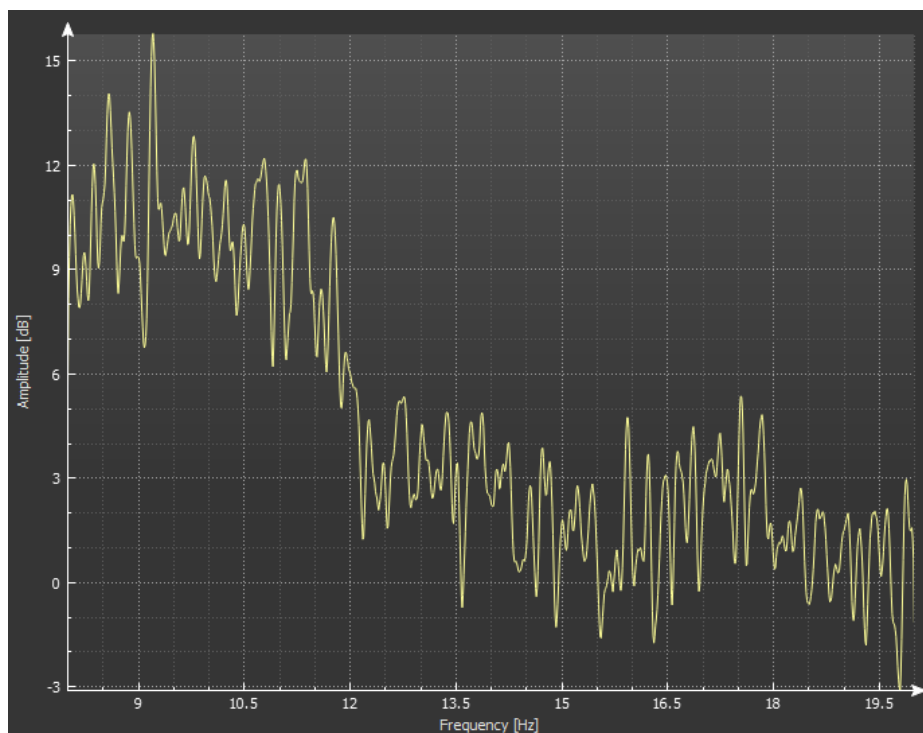
Na následujících snímcích jsou zobrazeny spektrogramy pro zavřená oční víčka ve frekvenčním rozmezí 8 až 20Hz. Signál pro nějž je spektrogram vytvořen je dán aritmetickým průměrem elektrod O1, O2, P3 a P4.



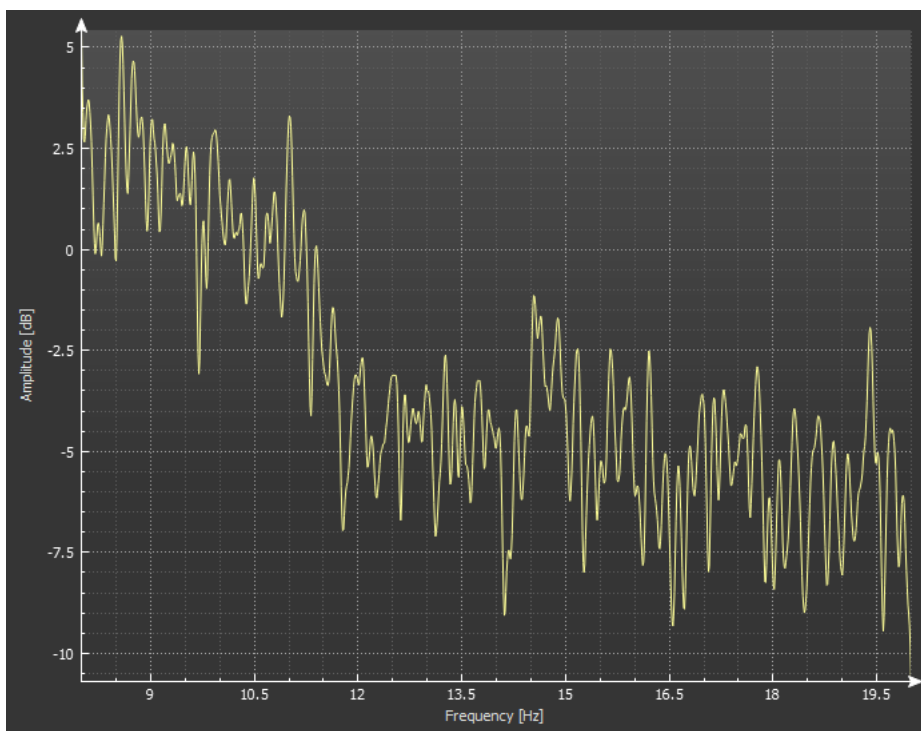
Obrázek B.1: Subjekt č.1 měření 1



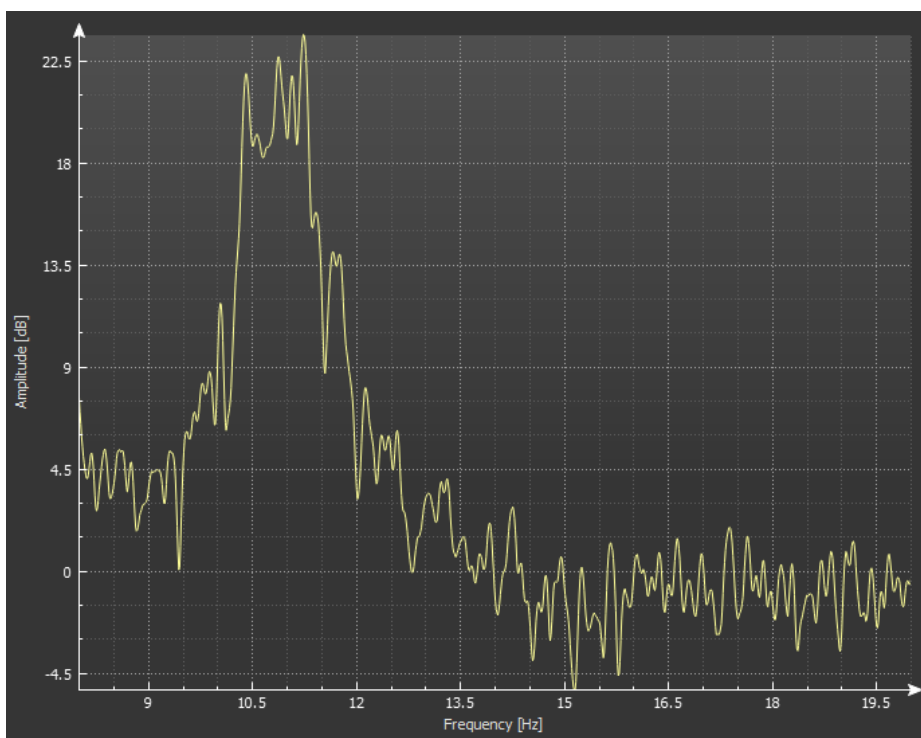
Obrázek B.2: Subjekt č.1 měření 2



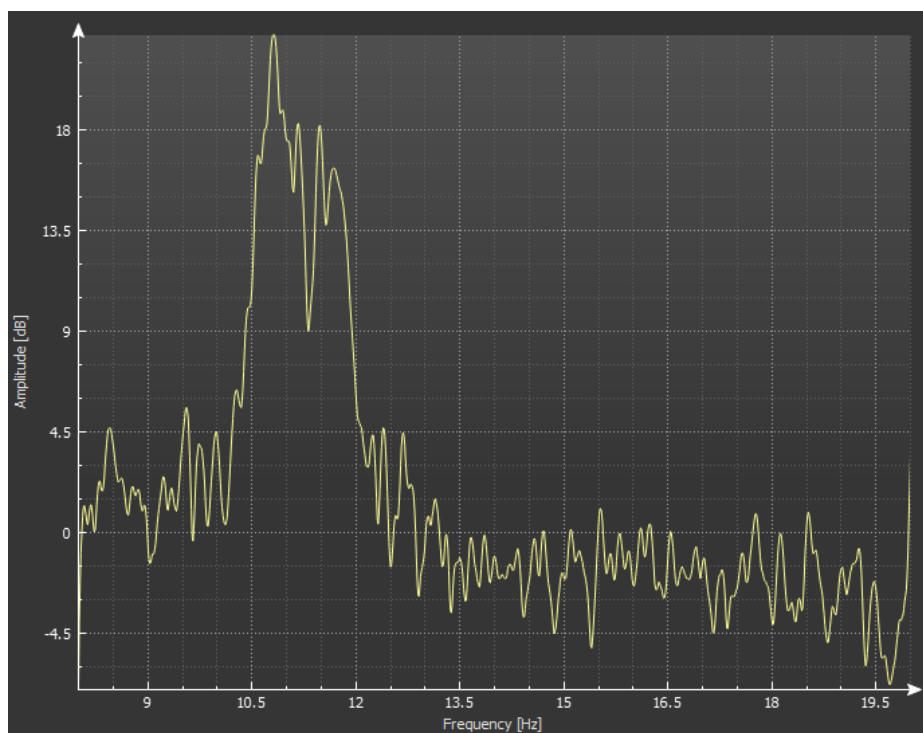
Obrázek B.3: Subjekt č.2 měření 1



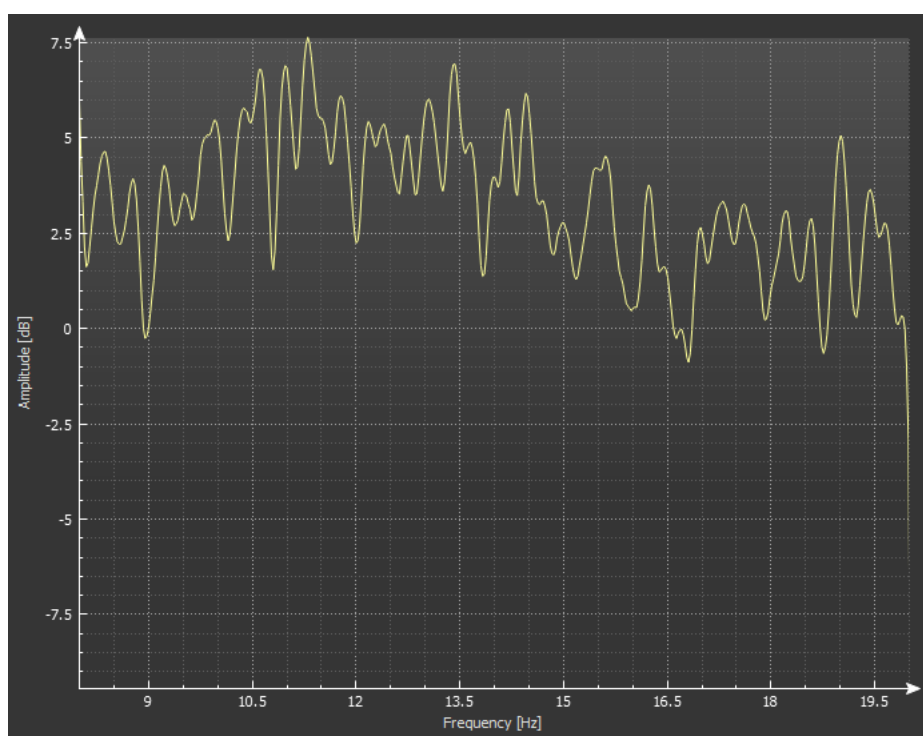
Obrázek B.4: Subjekt č.2 měření 2



Obrázek B.5: Subjekt č.3 měření 1



Obrázek B.6: Subjekt č.3 měření 2



Obrázek B.7: Subjekt č.4 měření 1