VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

ROZPOZNÁVÁNÍ A KLASIFIKACE EMOCÍ NA ZÁKLADĚ ANALÝZY ŘEČI

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE Bc. LUKÁŠ ČERNÝ
AUTHOR

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY
A KOMUNIKAČNÍCH TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND
COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

ROZPOZNÁVÁNÍ A KLASIFIKACE EMOCÍ NA ZÁKLADĚ ANALÝZY ŘEČI

EMOTIONAL STATE RECOGNITION AND CLASSIFICATION BASED ON SPEECH SIGNAL ANALYSIS

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE — Bc. LUKÁŠ ČERNÝ
AUTHOR

VEDOUCÍ PRÁCE — prof. Ing. ZDENEK SMÉKAL, CSc.
SUPERVISOR

BRNO 2010

oficialni zadání

ABSTRAKT

Diplomová práce se soustředí na klasifikaci emocí. Práce pojednává o parametrizaci zvukových souborů pomoci segmentálních a suprasegmentálních metod s ohledem na jejich další použití. Tato databáze obsahuje mnoho zvukových nahrávek s emocemi. Z těchto zvukových nahrávek jsou vytvořeny data, které jsou rozděleny do dvou částí. První část je použita pro trénik a druhá pro klasifikaci. Práce je soustředěna hlavně na samoorganizující sítě. Tato práce obsahuje programy v Matlabu, které mohou být použity pro parametrizaci jakékoliv databáze. Parametrizovaná data jsou předložena samoorganizující síťi ke klasifikaci. Dosažené výsledky jsou prezentovány na konci diplomové práce.

KLÍČOVÁ SLOVA

emoce, rozpoznávání, příznaky, neuronová síť, samorganizující neuronová síť, Kohenenova neuronová síť, prosodie, MFCC, základní tón řeči, energie, ZCR, výběr příznaků

ABSTRACT

The diploma thesis focuses on classification of emotions. Thesis deals about parameterization of sounds files by suprasegment and segment methods with regard for next used of these methods. Berlin database is used. This database includes many of sounds records with emotions. Parameterization creates files, which are divided to two parts. First part is used for training and second part is used for testing. Point of interest is self-organization network. Thesis includes Matlab´s program which can be used for parameterization of any database. Data are classified by self-organization network after parameterization. Results of hits rates are presented at the end of this diploma thesis.

KEYWORDS

emotion, clasification, recognation, artificial neural network, self-organizing network, Kohehen neural network, prosody, MFCC, fundamental frequence, energy, ZCR, extraction of characteristic

ČERNÝ, Lukáš *ROZPOZNÁVÁNÍ A KLASIFIKACE EMOCÍ NA ZÁKLADĚ ANALÝZY ŘEČI*: diplomová práce. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2010. 62 s. Vedoucí práce byl prof. Ing. ZDENEK SMÉKAL, CSc.

PROHLÁŠENÍ

Prohlašuji, že svou diplomovou práci na téma „ROZPOZNÁVÁNÍ A KLASIFIKACE EMOCÍ NA ZÁKLADĚ ANALÝZY ŘEČI“ jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené diplomové práce dále prohlašuji, že v souvislosti s vytvořením této diplomové práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

Brno ....

(podpis autora)

Děkuji vedoucímu práce prof. Ing. Zdeňku Smékalovi, CSc. ze velmi užitečné připomínky při zpracovávání diplomové práce. Také bych rád poděkoval Ing. Hichamu Atassimu za konzultace, během vypracovávání mé práce. Prof. Ing. Janě Tučkové,CSc. za konzultace a postřehy a také Ing. Petru Zetochovi za zpřístupnění programu SOM Laboratory, který sloužil jako předloha.

OBSAH

1 Úvod **11**

2 Tvorba řeči **12**
2.1 Dechové ústrojí 12
2.2 Hlasové ústrojí 12
2.3 Artikulační ústrojí 13
2.4 Modelování řečového signálu 14
2.5 Znázornění řečových signálů 15

3 Emoce **16**
3.1 Význam emocí 16
3.2 Vyjadřovaní emocí 18
3.3 Vlastnosti základních emocí se zaměřením na analýzu řeči 18
3.3.1 Neutrální projev (Neutral) 19
3.3.2 Vztek (Anger) 19
3.3.3 Smutek (Sadness) 19
3.3.4 Radost (Happieness) 19
3.3.5 Strach (Afraid, Fear) 19
3.3.6 Nuda (Bored) 19

4 Systém pro klasifikaci emocí **20**
4.1 Výpočet segmentálních a suprasegmentálních příznaků 20
4.2 Výběr vhodných příznaků 21
4.3 Výběr podle rozptylu 21
4.3.1 Redukce množiny příznaků dle grafické prezentace 22
4.3.2 Databáze emočních stavů 22
4.3.3 Klasifikace emocí 23

5 Prozodické informace obsažené v řečovém signálu **25**
5.1 Kmitočet základního tónu řeči 25
5.2 Energie 28
5.3 ZCR 29
5.4 MFCC 29

6 Neuronové sítě **32**
6.1 Principy učení umělých neuronových sítí a jejich rozdělení 33
6.2 Samoorganizující se neuronové sítě 35
6.2.1 Funkce souseda 37

6.2.2 Algoritmus SOM 37
6.2.3 Inicializace 38
6.2.4 Proces trénování 38
6.2.5 Trénovací parametr 40
6.2.6 Trénovací algoritmus 41

7 Praktická část práce **42**
7.1 Parametrizace databáze 42
7.2 Redukce množiny příznaků 43
7.3 Naměřené hodnoty 44

8 Závěr **51**

Literatura **53**

9 Zkratky **55**

A Příloha A – Obsluha programu **56**

B Příloha B – Jmený seznam pozic v proměné data **59**

C Příloha C – Vzorce statistických parametru **60**

D Příloha D – Obsah přiloženého CD **62**

SEZNAM OBRÁZKŮ

2.1 Hlasový trakt v lidském těle převzato z [12] . . . 13
2.2 Model hlasového traktu převzato z [9] . . . 14
2.3 Znázornění řečového signálů . . . 15
3.1 Emoční kružnice . . . 17
3.2 Základní parametry řečového signálu převzato z [8] . . . 18
4.1 Systém pro klasifikaci emocí převzato z [15] . . . 20
4.2 Prezentování vších parametru v barevném podání . . . 22
5.1 Reprezentace základního tónu z časové oblasti převzato z [8] . . . 26
5.2 Postup výpočtu metodou centralního klipování převzato z [1] . . . 28
5.3 Znázornění výpočtu průchodu nulou rovinou . . . 29
5.4 Prezentace MFCC příznaku . . . 31
6.1 Biologický vzor mozku převzato z [19] . . . 33
6.2 Dělení způsobu učení UNS převzato z [19] . . . 34
6.3 Topologie mřížky převzato z [16] . . . 36
6.4 Tvar mřížky zleva: sheet, cylinder, toroid převzato z [16] . . . 36
6.5 Tvar mřížky zleva: sheet, cylinder, toroid převzato z [16] . . . 37
6.6 Ovlivněné sousední neurony převzato z Matlab help . . . 37
6.7 Změna mřížky při nalezení vítězného neuronu převzato z [16] . . . 39
6.8 Trénovací funkce Lineární (červená), Inverzní v čase (modrá) a Power serie (černá) převzato z [16] . . . 40
7.1 Přehled emočních stavů : anglicky, německy, česky . . . 42
7.2 Kvalita příznaku . . . 43
7.3 Přehled nejlepších nalezených hodnot . . . 45
7.4 Zobrazení rozložení neuronu v mřížce pro I skupinu emocí . . . 46
7.5 Tabulka úspěšnosti pro I skupinu emocí (procentuální) . . . 46
7.6 Zobrazení rozložení neuronu v mřížce pro II skupinu emocí . . . 47
7.7 Tabulka úspěšnosti pro II skupinu emocí (procentuální) . . . 47
7.8 Zobrazení rozložení neuronu v mřížce pro všechny emoce . . . 48
7.9 Tabulka úspěšnosti pro II skupinu emocí (procentuální) . . . 48
7.10 Závislost nastavení funkce Neighbourhood . . . 49
7.11 Závislost nastavení délky trénování . . . 49
A.1 Okno programu parametrizace . . . 56
A.2 Přehled názvů emočních stavů . . . 57
A.3 Uživatelské prostředí funkce Emoce . . . 58

1 ÚVOD

Základním a jedním z nejstarších komunikačních prostředků pro komunikaci mezi lidmi je mluvená řeč. V řeči jsou obsaženy informace jako nápady a pocity, které chceme posluchači sdělit a další na první pohled skryté informace, které pro posluchače na první pohled nehrají klíčovou roli. Jedná se o informace jako jsou např. intenzita, intonace a zabarvení hlasu. Jednotně lze těmto informacím říkat emoční stav. Dříve než ke sdělení informace dojde, proběhne myšlenka mozkem a následně je vyprodukována hlasovým ústrojím. Tento proces je velice složitý a je předmětem zkoumání mnoha vědeckých pracovníků. Celé vědecké týmy se zabývají problematikou studování řečových signálů. S využitím výkonné počítačové techniky jsou dnes schopni s různou věrohodností např. vytvořit řeč z psaného textu STT. Většina takto vytvořených zvukových souborů má ovšem jeden problém a to, že pro posluchače nepůsobí věrohodně. Tento problém je především v tom, že se nedaří do této promluvy kvalitně implementovat emoční stavy. Lze nalézt mnoho oborů lidské činnosti, ve kterých se uplatní poznatky získané z tohoto velmi rozsáhlého vědního oboru. Jako příklad lze uvést zdravotnictví - rozpoznávání poruch řeči, bezpečnost - identifikace mluvčího a již zmíněný převod textu na řeč. Při zkoumání lidské řeči bylo popsáno mnoho technik, které lze použít ke zpracovávání a vyhodnocování různých problémů. Pokud se jedná o jednoduché úkoly, lze vystačit s omezeným, předem dobře zvoleným množstvím vstupních informací a dle sledování určitého množství výsledků se přiklonit k určitému závěru. Při řešení komplexnějšího problému, kde je zapotřebí zkoumat velké množství vstupních parametrů, je potřeba pro klasifikaci výsledků zapojit moderní metody. Jako moderní klasifikátory lze bezesporu považovat umělé neuronové sítě. Umělé neuronové sítě jsou typickým příkladem aplikace, kde je velké množství vstupních informací a pomocí hledaných podobností ve vstupních datech je lze klasifikovat do určité podmnožiny s podobnými výsledky. Existuje velké množství implementací umělých neuronových sítí. Jednou z nich je např. samoorganizující neuronová síť, která byla použita jako klasifikátor v této diplomové práci.

2 TVORBA ŘEČI

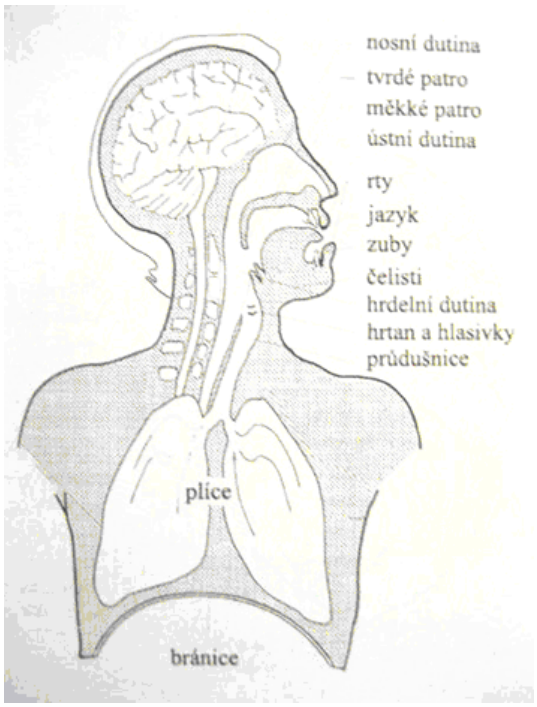
V této kapitole bude nastíněn proces tvorby řeči v hlasovém ústrojí. Jedná se o velmi komplikovaný systém, který je tvořen několika orgány, jejichž primární funkcí není tvorba řeči, ale každý z těchto orgánů má svou vlastní funkci (dutina nosní dýchání a čich; dutina ústní příjem potravy; hrtan dýchání a přijímání potravy). Když se tyto orgány spojí, vzniká hlasové ústrojí.

2.1 Dechové ústrojí

Dechové ústrojí je komplexní systém, který je tvořen především plícemi, průdušnici, hlasivkami, hrdelní dutinou, ústní dutinou a nosní dutinou. Na obrázku 2.1 lze vidět kompletní hlasový trakt. Při nádechu dochází k zaplnění plic vzduchem. Následně se při výdechu plíce stávají hlavním zásobníkem energie pro tvorbu řeči. Při vydechování prochází proud vzduchu průdušnicí a následně hrtanem a nadhrtanovými dutinami, kde se z něj stává řečový signál a je vyzařován rty do okolního prostoru. Síla s jakou je vzduch vydechován z plic má vliv na sílu hlasu a částečně i na jeho výšku. Pro vytvoření slyšitelné řeči je potřeba během několika sekund vytlačit více než 0,5l vzduchu [12]. Proces tvorby řeči je podobný, jako proces vytváření tónu u dechových hudebních nástrojů. [12]

2.2 Hlasové ústrojí

Hlasové ústrojí je uloženo v hrtanu, které je spojeno s plícemi pomocí průdušnice. Nejdůležitější části pro tvorbu hlasu jsou hlasivky, které se nachází v hrtanové dutině hned za ohryzkem“. Samotné hlasivky pak tvoří párový hlasivkový sval, hlasivkový vaz a slizniční hlasivková řasa. Typická délka hlasivek je pro muže 15mm a pro ženy 13mm. Hlasivky jsou pokryty sliznicí a jejich základ tvoří hlasivkový vaz a hlasivkový sval. Při vytváření hlasu se hlasivky nacházejí v hlasovém postavení, kdy jsou napnuty. Vydechovaný proud vzduchu prochází z plic až k hrtanu. V hrtanu se do cesty postaví hlasivky, které cestu vzduchu úplně uzavřou. Hlasivky se pod tlakem vzduchu stávají pružnými,začínají kmitat a střídavě se otvírají a uzavírají. V důsledku kmitání hlasivek se ze vzduchového proudu stává vzduchová vlna, kterou vnímáme jako zvuk. Tento periodický proud vzduchových pulsů tvoří základ lidského hlasu. Nazýváme ho základní (hlasivkový) tón. Frekvenci, jakou kmitají hlasivky označujeme $F0$ a nazývá se frekvence základního hlasivkového tónu. Tuto hodnotu vnímáme jako výšku hlasu. Převracenou hodnotou je pak $T0 = 1/F0$ a nazýváme ji periodou základního hlasivkového tónu.

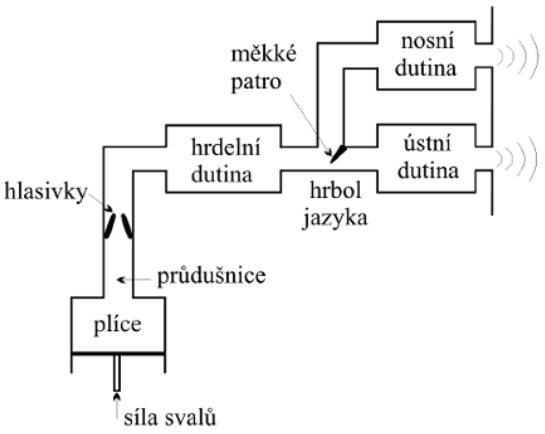


Obr. 2.1: Hlasový trakt v lidském těle převzato z [12]

Frekvence základního tónu řeči je u mužů, žen i dětí různá. Hodnota frekvence základního tónu se uvádí pro muže v rozmezí 80 - 160 Hz. Pro ženy v rozmezí 150 – 300Hz. Děti mívají tuto hodnotu mezi 200 – 600 Hz [12]. Základní tón se v průběhu života mění. Existují i jedinci, kteří mají tuto hodnotu pod (velmi hluboký mužský hlas) nebo nad touto hodnotou (cvičené operní pěvkyně okolo 1000 Hz). Frekvence základního hlasivkového tónu není konstantní, ale velice rychle se mění. Tuto změnu popisuje funkce jitter která je závislá na emočním stavu mluvčího (udává se v [%]). Dalším pozorovaným parametrem může být shimmer, který je závislý na kolísaní amplitudy (udává se v [dB]). Při normální promluvě se hodnota změny periody (jitter) pohybuje okolo 1%. Hodnoty, které jsou posluchačem postřehnutelné jsou nad 2% a hodnoty změny amplitudy (shimmer) 1dB. [12]

2.3 Artikulační ústrojí

Díky artikulačnímu ústrojí člověk může vytvářet velké množství různých zvuků. Artikulační ústrojí je tvořeno několika orgány. Jsou to různé dutiny jako hrdelní, ústní, nosní a nadhrtanové. Dutiny jsou vzájemně oddělené tzv. čípkem, který buďto umožňuje, nebo zamezuje přístupu vzduchu z dutiny hrdelní do dutiny nosní. Artikulační ústrojí lze rozdělit na aktivní a pasivní orgány podle toho jestli tvoří nebo



Obr. 2.2: Model hlasového traktu převzato z [9]

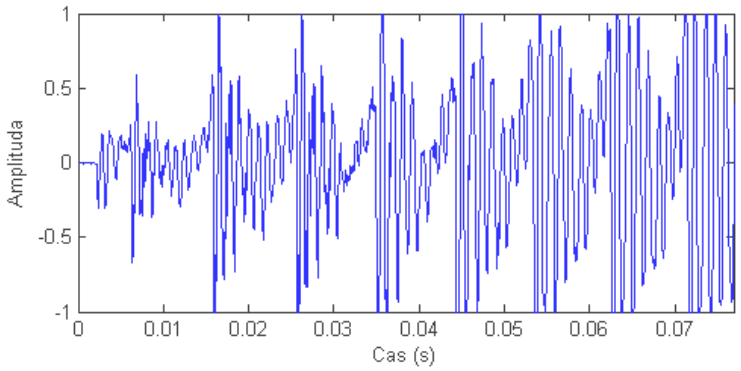
netvoří pohyblivou součást ústrojí. Jazyk, rty, a měkké patro se podílí na vytvoření největšího počtu různých zvuků. Podle umístění jazyka se mění tvary dutin a tím se vytváří různé zvuky. Menší měrou na tvorbě se na tvorbě zvuku podílí zuby, tvrdé patro a čelisti. Velice zajímavým ústrojím je hrtan, který kromě toho, že se zapojuje do tvory znělosti, se také pohybuje nahoru a dolů a mění délku proslovu [12].

2.4 Modelování řečového signálu

Při procesu zpracování řeči je užitečné tento velmi komplikovaný komplexní systém převést na matematický model, s kterým se pak lépe experimentuje. Cílem těchto experimentů je nalézt matematický model, který by co nejlépe popisoval celý proces a zároveň by byl co nejjednodušší. Pro navrhovaný systém by bylo velmi přínosné, kdyby se jednalo o systém lineární a časově invariantní. Lidská řeč ale těmto předpokladům neodpovídá, protože jde o proces souvislý a časově proměnný. Z tohoto důvodu doposud nebyl předložen univerzální model. Existuje ovšem model, který se užívá velice často a se kterým lze dosáhnout dobrých výsledků. Musíme si ovšem uvědomit že tento model funguje pouze na velice krátkých úsecích. Tento model funguje pro úseky dlouhé 10 - 30 ms. Jedná se o model lineárně časově invariální. Skládá se z lineárního modelu hlasového traktu s pomalu se měnícími parametry. Model je buzen dvěmi druhy signálu. Může se jednat o periodický sled pulzů, který nám vytvoří znělou řeč, nebo o šumový signál, který tvoří neznělou řeč. Model hlasového traktu reprezentuje především vlastnosti dutiny ústní, nosní a hrdelní. Ovšem jen pro velmi krátké úseky řeči.

2.5 Znázornění řečových signálů

Při určování řeči je vhodné řečový signál prezentovat graficky - vizualizovat. Časový průběh je velmi komplexní a nelze jej na první pohled jednoduchým způsobem popsat. Protože člověk mnohém lépe zpracovává obrazové informace, než informace numerické, je řečový signál nejčastěji prezentován jako závislost času na amplitudě (obrázek 2.3).V mnoha oblastech řečových signálů je účelné popsat řečový signál současně několika jednoduchými parametry. K takovému popisu se nabízejí především základní veličiny čas, energie a kmitočet, které lze znázornit v prostorové grafice přiřazením os podle obrázku 1.3[5].



Obr. 2.3: Znázornění řečového signálů

3 EMOCE

V této kapitole bude stručně pojednáno o významech emocí, jejich vyjadřovaní a základních vlastnostech řečového signálu v závislosti na emočním stavu. Dále bude každý emoční stav popsán zvlášť se zaměřením na jeho dominantní parametry, které se využívají pro klasifikaci emocí z řečového signálu.

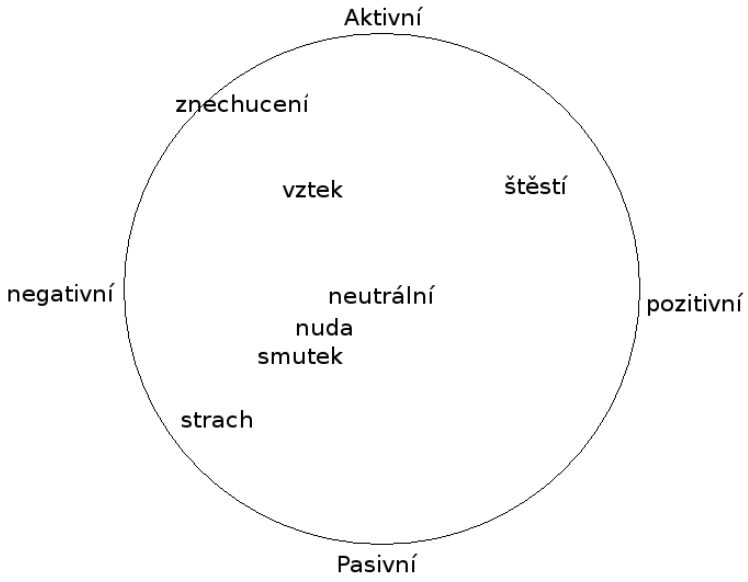
Emoce jsou psychologické procesy, které jsou výsledkem právě probíhajícího zážitku nebo zážitku, který se sice odehrál před delší dobou, ale při vzpomínce na něj, v nás vyvolávají silné vzpomínky. Emoce může být také způsobena dosažením nějakého cíle. Je-li cíle dosaženo, hovoříme o pozitivních emocích, pokud ne, jedná se o emoce negativní. Pro emoce platí, že každá emoce má svůj protiklad a může byt pociťovaná v různé intenzitě. Emoce se velice těžko dají ovlivňovat. Je to z toho důvodu, že jsou evolučně starší než rozumové chápání. Při různých výzkumech bylo shodně prokázáno, že určité vlastnosti zkoumaných signálů jsou si podobné a odpovídají základnímu rozdělení emocí na aktivní, pasivní a na příjemné a nepříjemné. Mezi základní kritéria při rozpoznání emocí patří fyziologické změny, např. změna srdečního tepu, rychlost dýchání, mimika obličeje, gestikulace (pohyb těla), a také mluvený projev. Právě z mluvené promluvy pochází asi jen 10% informací, které obsahují informaci o emočním stavu řečníka. Zkoumaných emočních stavů může být celá řada. Pokud by bylo potřebné rozpoznat příliš mnoho stavů, byla by pravděpodobně úspěšnost degradována na neuspokojivé hodnoty. Pro představu lze uvést některé výsledky, kterých bylo již dříve dosaženo, abychom měli představu jaké výsledky lze předpokládat. Například úspěšnost posluchačů při identifikaci emocí ve studiu Liebermana a Michaelse dosahovala 85% , při klasifikaci 4 emočních stavů

[20], dále při identifikaci 10 emočních stavů klesla úspěšnost už k 60% (studio Schererova 1981) [20]. V této práci budeme identifikovat 7 emočních stavů a předpokládáná úspěšnost by se mohla pohybovat v rozmezí 60-80%.

Některé studie se snaží jednotlivé emoční stavy zařadit do kružnice, která nese na své y-lonové ose informaci o aktivním, či pasivním zařazením emocí a na x-ové o zařazení do pozitivních, či negativních stavů. Ukázka této kružnice je vidět na obrázku 3.1.

3.1 Význam emocí

Jak již bylo zmíněno, emoce jsou považovány za starší než mluvený projev. Jako příklad lze uvést, že jako malé děti jsme pláčem uměli vyjádřit, že nám něco schází a až později, když jsme se naučili mluvit, jsme naše potřeby vyjadřovali mluvenou řečí. Je také důležité si uvědomit, že emoce se u každého jedince v průběhu celého života mění. Největší změny probíhají v dětství a v pubertě, kdy se setkáváme s



Obr. 3.1: Emoční kružnice

novými situacemi a postupně se na ně učíme reagovat. Existuje mnoho teorií, jak emoce vznikly popř. k čemu je potřebujeme.

Jako jednu z teorií lze uvést psycho-evoluční teorii Roberta Plutchika, která předpokládá, že se jedná o vrozené funkce, které nám pomáhají přežít. Jeho teorie je založena na emocionálním kole, kde jsou emoce rozděleny na oblasti pozitivní a negativní a dále dle intenzity, kde intenzivní emoce jsou rozpoznávány snaze, než emoce s nízkou intenzitou. [8] Další teorií je Cannon – Bardova teorie. Tato teorie předpokládá, že pokud člověk čelí události, která ovlivní jeho emoční stav, je zpráva z této události přenesena do mozku. Zde je následně posílána do kůry mozkové, která je spojena s emocemi. Další část putuje do hypothalamu, který má za následek fyziologické změny např. změna tepové frekvence, stahy svalů apod. [8] V přehledu lze uvést základní roli, kterou emoce mají:

- Vykonávají reakci na podněty, které potřebují okamžitou reakci
- Může se jednat o ukazatele toho, co máme v nejbližším okamžiku v plánu udělat
- Často mají komunikační funkci, kdy při rozhovoru poznáme náladu podle toho, jak mluvčí prezentoval svou myšlenku, ne přímo jak ji přesně formuloval
- Slouží k vyrovnání se se situacím která nastala

	neutralita	vztek	smutek	radost	strach	překvapení	obdiv	ironie
průměr F0	120Hz	180 Hz	110 Hz	200 Hz	120 Hz	120 Hz	150 Hz	190 Hz
min. a max. hodnoty F0 a jejich rozdíl	100 ; 170 E = 70 Hz	100 ; 130 E = 30 Hz	100 ; 130 E = 30 Hz	100 ; 250 E = 150 Hz	100 ; 130 E = 30 Hz	100 ; 340 E = 240 Hz	120 ; 200 E = 80 Hz	100 ; 400 E = 300 Hz
tvar intonační křivky								
průměrný rozdíl intenzity od neutr promluvy	0 dB	11 dB	0 dB	16 dB	0 dB	-1 dB	3 dB	3 dB
min., max. intenzita a rozdíl	+4 ; -2 E = 6 dB	+16 ; +3 E = 13 dB	+2 ; -3 E = 5 dB	+18 ; +8 E = 10 dB	+6 ; -3 E = 9 dB	0 ; -3 E = 3 dB	+10 ; +2 E = 8 dB	+5 ; -3 E = 8 dB
pauza	0 ms	0 ms	0 ms	750 ms	0 ms	250 ms	0 ms	0 ms
trvání věty	1700 ms	1650 ms	1950 ms	2700 ms	1850 ms	2300 ms	1920 ms	1800 ms

Obr. 3.2: Základní parametry řečového signálu převzato z [8]

3.2 Vyjadřovaní emocí

Hlavními prostředky pro vyjádření emočního stavu je především mimika obličeje. Jako příklad lze uvést úsměv, který signalizuje pozitivní náladu a je tak vnímán na celém světě, bez rozdílu etnika. Tímto druhem klasifikace emocí se zabývá mnoho akademických pracovníků po celém světě. Jako českého zástupce bychom mohli uvést Západočeskou univerzitu v Plzni a jejich řešení projektu: Rozpoznávání emocí z výrazu obličeje. Dalším druhem vyjádření emocí jsou gesta. Jedná se především o pohyb končetin, držení těla a oční kontakt. Při posuzování tohoto druhu emocí se většina odborníků v této oblasti neshoduje. Jejich názor není jednotný v tom, jak významnou roli mají tyto ukazatele při projevování emocí.

3.3 Vlastnosti základních emocí se zaměřením na analýzu řeči

Bylo již učiněno několik studií, které se zabývají klasifikací emocí. Vždy byly použity různé parametrické metody (bude podrobněji popsáno v kap. 5), které jsou pak přehledně prezentovány v tabulkách. Každá studie může mít v této tabulce jiné hodnoty které porovnává. Jako příklad je zde uvedena studie Léona [8]. Obecně se vždy popisují emoční stavy vzhledem k neutrální promluvě, která se bere jako referenční. Zde zobrazené výsledky jsou brány pouze jako informativní. Je důležité mít přibližnou představu, jakých parametru by jednotlivé emoční stavy měly nabývat.

3.3.1 Neutrální projev (Neutral)

Jak již bylo zmíněno, tento emoční stav se bere jako referenční a další emoční stavy jsou s tímto stavem srovnávány. Pro tento proslov se předpokládá, že neobsahuje žádné informace o emočním stavu a na emoční kružnici leží přesně v jejím středu. Průměrná hodnoty základního tónu řeči f dosahuje hodnoty 120 Hz.

3.3.2 Vztek (Anger)

Tento emoční stav je charakteristický svou vyšší průměrnou hodnotou základního tónu. Rozsah hodnot f je jeden z nejnižších ze všech pozorovaných emočních stavů. Jedná se o emoci aktivní a negativní. Značný je také rozdíl intenzity oproti neutrální promluvě, který dosahuje hodnoty 11dB.

3.3.3 Smutek (Sadness)

Smutek lze zařadit do emocí neutrálních, v rámci jeho aktivity a do velmi negativních. Hodnota f bývá podobná hodnotě neutrálního projevu, rozsah hodnot bývá velmi malý 30Hz. Nejvýraznější rozdíl oproti neutrálnímu projevu je v hodnotě intenzity, která bývá o 16dB vyšší. Tento emoční stav bývá nejčastěji klasifikován z největší procentuální pravděpodobností [15].

3.3.4 Radost (Happieness)

Tento stav je charakteristický vysokou hodnotou f , která nabývá průměrné hodnoty 200Hz. Její rozptyl je také jeden z nejvyšších. Dále je proslov charakteristický tempem promluvy. Tato emoce se na kružnici nachází v oblasti aktivních a velmi pozitivních emocí.

3.3.5 Strach (Afraid, Fear)

Jedná se o emoci aktivní a negativní. U tohoto stavu lze předpokládat f okolo hodnoty podobné neutrálnímu projevu. Rozsah hodnot je jeden z nejnižších ze všech emočních stavu.

3.3.6 Nuda (Bored)

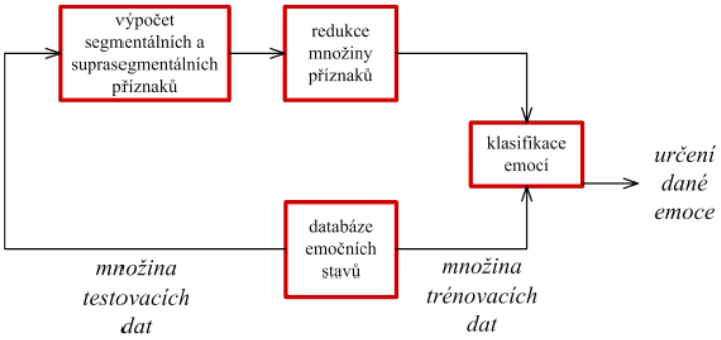
Tento emoční stav nabývá podobných hodnot f, jako referenční neutrální stav. Jeho rozptyl se také od neutrálního projevu moc nemění. Tempo tohoto emočního stavu bývá klesavé. Jedná se o pasivní, negativní stav.

4 SYSTÉM PRO KLASIFIKACI EMOCÍ

V této kapitole bude uveden postup, který se používá při klasifikaci emocí. Tento postup bývá podobný i pro jiné aplikace při práci s řečovým signálem.
Jako příklad lze uvést:

- Rozpoznávání slov
- Rozpoznávání řečníka
- Rozpoznání plynulé řeči
- Rozpoznávání pohlaví a věku řečníka
- Rozpoznávání logopedických vad

Při procesu klasifikace emocí je kladen důraz na zvýraznění emocí v řečovém signálu a na potlačení jiných nepotřebných parametrů. Při zpracování emočních stavů je proces klasifikace prováděn na dlouhém úseku, protože emoce jsou patrné až z dlouhé promluvy a ne z krátkého úseku jako jsou např. hlásky. Jednotlivé metody, které jsou v této práci použité, jsou popsány v kapitole 5. Při klasifikaci emocí se pro každé dva emoční stavy hodí jiné parametry, podle toho v jakých oblastech vykazují nejvyšší odchylky.



Obr. 4.1: Systém pro klasifikaci emocí převzato z [15]

4.1 Výpočet segmentálních a suprasegmentálních příznaků

Tomuto bloku je věnovaná celá kapitola 5. Celkově v této práci bylo vypočteno 41 příznaků (není započten první parametr, který odpovídá zařazení do emočního stavu).

4.2 Výběr vhodných příznaků

Přesný počet příznaků, které jsou potřeba, se v různých literaturách mění. Ovšem nejčastěji se uvádí počet příznaků 10-20 [13, 15].

4.3 Výběr podle rozptylu

Po parametrizaci nahrávek je potřeba snížit počet parametrů, protože práce s takto velkým počtem příznaků by byla zdlouhavá a zbytečná. Je potřeba vybrat pouze takové příznaky, které nám co nejvíce pomohly s klasifikací emočních stavů. Příznaky by měly splňovat určité požadavky. Parametry v rámci své třídy by měly mít co nejnižší rozptyl hodnot (soustředily by se co nejvíce kolem střední hodnoty) a zároveň střední hodnoty by měly dosahovaly co nejvyšších rozdílů. **Postup výpočtu:**

Kvadrát rozptylu třídy kolem střední hodnoty je definován: (funkce var **Matlabu)**

$$S_v^2 = E(x - \mu)^2 \tag{4.1}$$

Aritmetická střední hodnota všech počítaných tříd je:

$$S^2 = \frac{1}{V} \sum_{v=1}^{V} S_v^2 \tag{4.2}$$

Parametr V představuje počet klasifikovaných tříd. Aritmetická střední hodnota vzdálenosti je definovaná vztahem:

$$D^2 = \frac{1}{V(V-1)} \sum_{v=1}^{V} \sum_{u=1}^{V} D_{v,u}^2 \tag{4.3}$$

kde $D_v u$ **je kvadrát vzdálenost mezi třídami** v **a** u**.**

$$D_v^2, u = (\mu_v - \mu_u)^2, \tag{4.4}$$

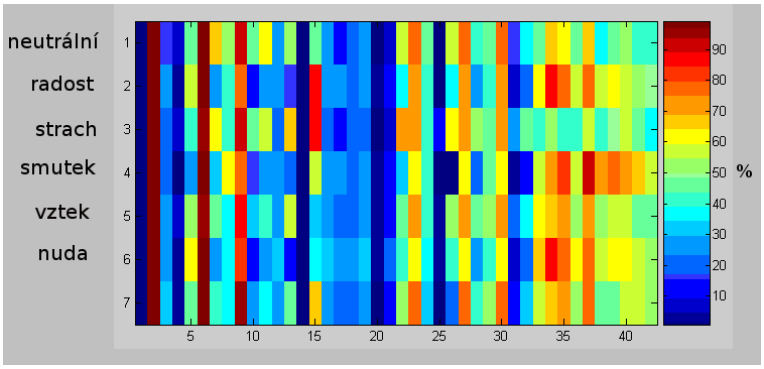
Informace o oddělitelnosti třídy je definována poměrem:

$$Q(.) = \frac{S^2}{S^2 + D^2} \tag{4.5}$$

Parametr Q **dosahuje hodnot** $0 - 1$**, kde hodnota** $Q = 0$ **představuje hodnotu, které by se mělo ideálně dosáhnout. Hodnota** $Q = 1$ **představuje hodnotu, která vypovídá o nevhodnosti použití tohoto příznaku**

4.3.1 Redukce množiny příznaků dle grafické prezentace

Pro lepší představu a následující výběr správných příznaků byla vytvořena funkce, která všechny příznaky normalizuje do stejného rozptylu hodnot (vždy v rámci jednoho příznaku je nalezena minimální a maximální hodnota a ty jsou převedeny na nové hodnoty, kde minimální hodnota se převede na 0 **a maximální hodnota např. na hodnotu** 100**). Dále jsou jednotlivé prvky emočních stavů v rámci své třídy zprůměrňovány a vyneseny na osu** y**. Pro jednodušší orientaci ve výstupních datech jsou data prezentována barevně, jak lze vidět na 4.2.**



Obr. 4.2: Prezentování vších parametru v barevném podání

Při takto prezentovaných datech už postupně můžeme odstranit ty oblasti, které nevykazují značné rozdíly z množiny příznaků. Zde se například může jednat o příznaky, které se nacházejí na pozicích $16 - 21$**. Příznaky vykazují shodné průměrné hodnoty a jsou zobrazeny modře.**

4.3.2 Databáze emočních stavů

Databáze emočních stavů hraje důležitou roli pro celý systém. Databáze je použita jak pro trénovaní sítě, tak pro data, která síť budou testovat. Je důležité, aby testovací data nebyly stejné jako data trénovaní. Pokud by tomu tak bylo, pak bychom hledali v natrénované množině dat identický prvek a věrohodnost těchto výsledků by degradovala. Doporučený poměr pro prvky pro trénovaní a klasifikaci by měl být $5:3$**,** $5:2$ **[15].**

Existují v podstatě jen 3 databáze:

- **Databáze emocí hraných**
- **Databáze emocí vyvolaných externím podmětem**
- **Databáze emocí spontálních**

Databáze emocí hraných : Převážně se jedná se o databázi vytvořenou profesionálními herci, která z hlediska věrohodnosti má nejmenší váhu. Nahrávky bývají nahrávány ve zvukových laboratořích, a proto bývají bez okolního šumu. Jako příklad lze uvést Berlínskou databázi [3], která bude detailněji popsána v kap. 7. Tato databáze je použita také v této práci. Databáze emocí vyvolaných externím podmětem : Pro tvorbu této databáze se většinou používá cizí podmět pro vyvolání emoce. Po tomto podmětu je osoba tázána na konkrétní otázky, které se týkají vyvolané emoce. Většinou se k vyvolání emocí používají krátké videozáznamy. Takto pořízené nahrávky už mají vyšší věrohodnost než databáze emocí hraných.
Databáze emocí spontánních : Tato databáze má nejvyšší věrohodnost, ovšem problém je tuto databázi vytvořit. Především by se člověk neměl dozvědět, že je nahráván. Toto by byl ovšem problém především z etického a právního hlediska. Také je možné použit nahrávky z televizního vysílání, nebo rozhovor posádky při problémech, které se objevily během letu letadla apod. [15].

4.3.3 Klasifikace emocí

Klasifikace emocí je posledním a jedním ze stěžejních bloků celého systému. Pokud máme celý systém hotový, pak je důležité vhodně zvolit klasifikátor o kterém si myslíme, že by mohl dosahovat nejlepší úspěšnosti.
Máme na výběr několik algoritmů:

- **Umělé neuronové sítě**
- **Metoda nejbližších sousedů**
- **Gaussovy smíšené modely**
- **Skryté Markovy modely**

Každý z těchto algoritmů lze pro klasifikaci emoce použít, ovšem s každým lze dosáhnout jiných hodnot, a proto je nutné každému z nich věnovat pozornost.

Umělým neuronovým sítím a především Samoorganizujícím sítím s Kohenenovým učením bude věnovaná celá kapitola 6.

Metoda nejbližších sousedů tato metoda nevyužívá trénování, ale pouze klasifikaci. Vstupní data jsou rozdělena do předem daného počtu oblastí, do kterých se při klasifikaci mají přiřadit. Jedním ze vstupních parametrů je vzdálenost okolí. Je to hodnota, ohraničující oblast, která je braná v úvahu. Tato oblast je převážně brána v eukleidovské metrice. Každý testovaný vzorek je umístěn do prostoru, kde jsou pak vypočteny jeho vzdálenosti k nejbližším hodnotám a dle maximálního počtu zařazení v dané třídě je tento prvek takto klasifikován.

Gaussovy smíšené modely tento algoritmus vychází z myšlenky modelovat trénovací příznaky jednou nebo více Gaussovými funkcemi rozložení pravděpodobnosti. Více se lze dočíst o této metodě v [15, 2].

Skryté Markovy modely jsou popsány např. v [4].

5 PROZODICKÉ INFORMACE OBSAŽENÉ V ŘEČOVÉM SIGNÁLU

Obsahem řeči může být segmentální popis, kde nás zajímá, co mluvčí říká (o čem mluví), ale také obsahuje prozodickou informaci, kterou lze v řeči poznat jak to mluvčí říká. Termín prozodie obsahuje několik vlastností řečového signálů. Jedná se především o frekvenci základního tónu řeči (výška hlasu, melodie), intenzitou (hlasitostí) a časováním. Informace spojené s časováním jsou rytmus (rozvržení přízvuku) a rychlost řeči (trvání slabik a hlásek). Změny, které se týkají základního tónu řeči tvoří melodii (intonaci). Souhrnně se těmto informacím také může říkat suprasegmentální jevy. Je to z toho důvodu, že zde popsané informace se vážou k delším časovým oblastem, jako jsou slabiky, slova, celé věty. Prozodie tvoří přirozenou součást komunikace, a proto je důležité se touto oblastí zabývat při automatickém zpracování řeči. Suprasegmentální vlastnosti hrají důležitou roli v syntéze řeči, kde významně ovlivňují především přirozenost, ale i srozumitelnost řeči. Bez modelování prozodických jevů by syntetická řeč působila uměle a ochuzeně. Je proto důležité tuto oblast nenechávát v ústranní zkoumání.

5.1 Kmitočet základního tónu řeči

Základní tón řeči je základním parametrem řečového signálu a projevuje se jako melodie řeči. Tento parametr nám určuje základní kmitočet na kterém nám kmitají hlasivky. Hodnota základního tónu řeči se vypočítá jako:

$$F0 = \frac{1}{T0}[Hz] \tag{5.1}$$

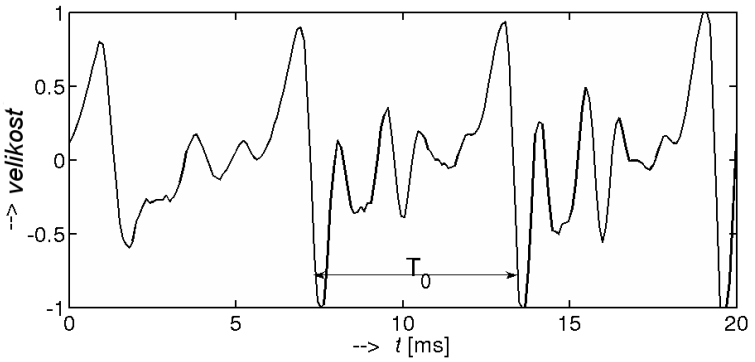
Obrácenou hodnotou je perioda základního tónu, který je vypočítán jako převrácená hodnota frekvence:

$$T0 = \frac{1}{F0}[s] \tag{5.2}$$

Lag označuje periodu vyjádřenou ve vzorcích,

$$L = T0.Fs[-], \tag{5.3}$$

kde Fs je vzorkovací perioda.



Obr. 5.1: Reprezentace základního tónu z časové oblasti převzato z [8]

Průměrný rozsah základního tónu je v rozmezí $60 - 400Hz$. Nejnižší kmitočty odpovídají velmi hrubým mužským hlasům. Kmitočty v rozmezí $150 - 300Hz$ odpovídají ženskému hlasu a nejvyšší kmitočty hlasu dětskému. Můžeme se setkat i se základní frekvencí vyšší, než je zde uvedeno. Ta pak odpovídá operním pěvcům, kteří mohou mít základní tón řeči až nad $700Hz$. Takový projev se ale může jevit jako nesrozumitelný.

Průměrná hodnota základního tónu řeči má i jiné uplatnění než je rozpoznání emočního stavu řeči, a to např. :

- **Detekce hlasové aktivity**
- **Identifikace mluvčího**
- **Syntéza řeči**
- **Kódování řeči pomocí LPC koeficientů**
- **Modelování prozodie**

Základní tón řeči lze vypočítat v různých oblastech signálu (časová, frekvenční, kepstrální) a lze použít různých výpočtů. Bude zde uveden pouze přehled metod a bude zde popsána pouze metoda použitá v této práci. Tato metoda se nazývá metoda centrálního klipování. Všechny metody jsou popsány v publikaci [1]. Metody výpočtu:

- **Metoda centrálního klipování**
- **Metoda založená na inverzní filtraci**
- **Syntéza řeči**
- **Spektrální metoda**
- **Kepstrální metoda**

Metoda centrálního klipování Tato metoda vypočítá hodnotu $F0$ v časové oblasti. Jedná se o jednu z nejrychlejších i nejpřesnějších metod. Právě proto byla tato metoda vybrána. Základním principem je, prozkoumání

navzorkovaného signálu. Hodnoty nad zvoleným prahem jsou zachovány a pod prahem jsou nulovány. Totéž platí pro záporné hodnoty signálu. Proto je možné použít vztah :

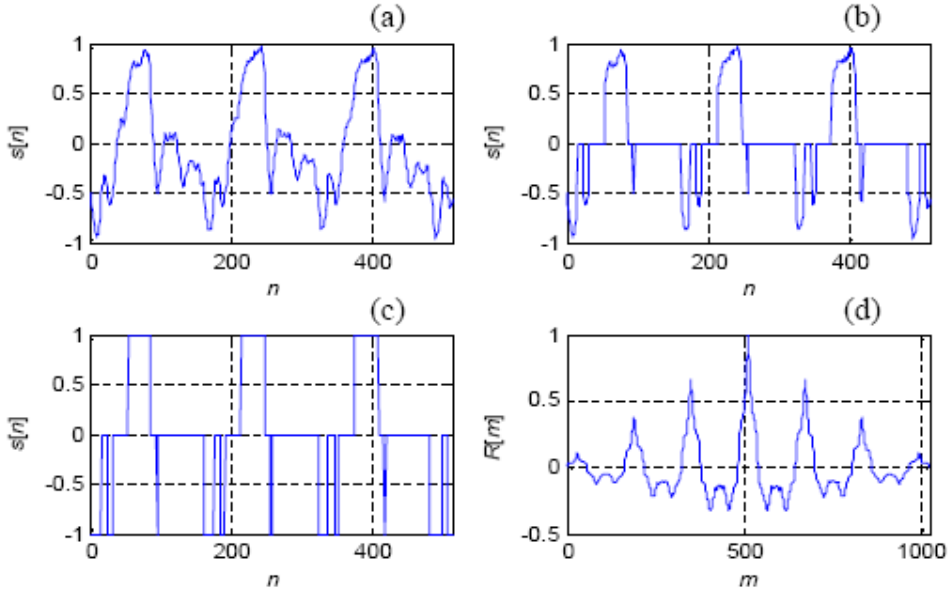
$$Pi = k.min(Max(i-1), Max(i+1), \tag{5.4}$$

Segmentace Před zpracováním se signál musí segmentovat. Je to z důvodu různé úrovně signálu v průběhu dlouhého úseku. Segmentace se provádí pomocí váhovacích oken a určitého překryvu. Základní okna pro segmentaci jsou:

- **Hammingovo**
- **Hannovo**
- **Pravoúhlé**

Postup výpočtu metodou centrálního klimpování [1]

1. **Segmentace signálu na délky** $32ms$ **s překryvem** $16ms$
2. **Vypočteme prahové hodnoty signálu pro každý segment**
3. **Signál je po prahovaní normalizován na hodnoty, které nabývají** $[-1;0;1]$
4. **Kmitočet** $F0$ **je získán z autokorelace signálu, který byl získán v kroku 3**
5. **U znělého signálu je** $F0$ **vypočteno ze vztahu**



Obr. 5.2: Postup výpočtu metodou centralního klipování převzato z [1]

a Vstupní segment
b Vstupní segment po prahování
c Vstupní segment po klipování
d Oboustranná autokorelační funkce klipovaného signálu

5.2 Energie

Energie, neboli intenzita je vnímána jako síla hlasu, neboli hlasitost. Před výpočtem krátkodobé energie je vhodné provést operaci, která nám zajistí, že hodnoty signálu budou v rozmezí $1-(-1)$**. Dále bude dlouhý signál segmentován na úseky o délce v rozmezí** $16-32ms$**. Jak je ze vzorce patrné, hodnoty které jsou v oblasti nulové úrovně přispívají malým dílem**

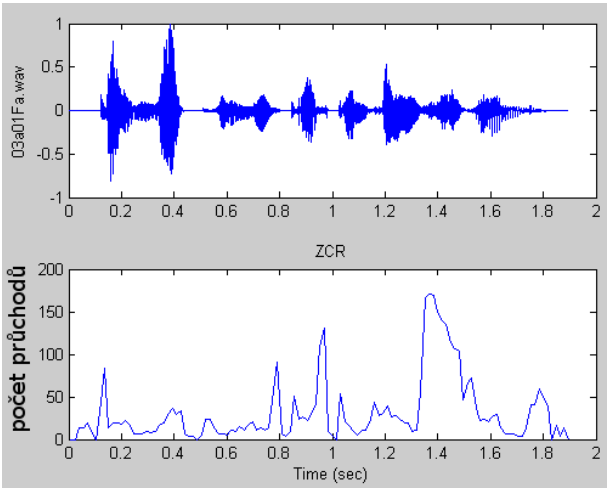
k navýšení celkové energie.

$$E = \frac{1}{N} \sum (x[n])^2 \tag{5.5}$$

5.3 ZCR

ZCR je zkratka z anglického Zero Crossing Ratio, což v překladu znamená počet průchodů nulovou hodnotou. Tato funkce se používá jako jednoduchý ukazatel změny základního tónu řeči. Nárůst průchodů nulou znamená, že se frekvence zvýšila a naopak, když počet průchodů klesá, pak se frekvence snížila. Dále se této funkce využívá pro rozdělení části na úseky s vysokou energii tzn. znělé úseky řeči (řeč, samohlásky) a úseky s nízkou energií tzn. neznělé úseky (ticho, souhlásky).

$$ZCR(m) = \sum |sgn(s(n)) - sgn(n-1)|\, w\,(m-n) \tag{5.6}$$



Obr. 5.3: Znázornění výpočtu průchodu nulou rovinou

5.4 MFCC

MFCC je zkratkou Mel-frequency cepstrum. Jak již z názvu vyplývá, tento výpočet je proveden v kepstrální oblasti. Tyto příznaky jsou používány jako jedny z nejčastějších při zpracování řeči. Kromě použití při rozpoznávání emocí lze uplatnit ve velkém množství aplikací.

Jako příklady lze uvést:

- **Rozpoznávání řečí**
- **Rozpoznání řečníka**
- **Rozpoznávání pohlaví**

Tento výpočetní proces obsahuje několik kroků:

- **Segmentace oken**

Segmentace oknem rozdělí dlouhý signál na několik částí o kratší délce

- **Váhování oken (volitelná)**

Váhování oknem je proces úpravy vzorku signálu. Nejčastěji se používá obdélníkové, trojúhelníkové, hammingovo, hannovo okno.

- **Banka mel filtrů**

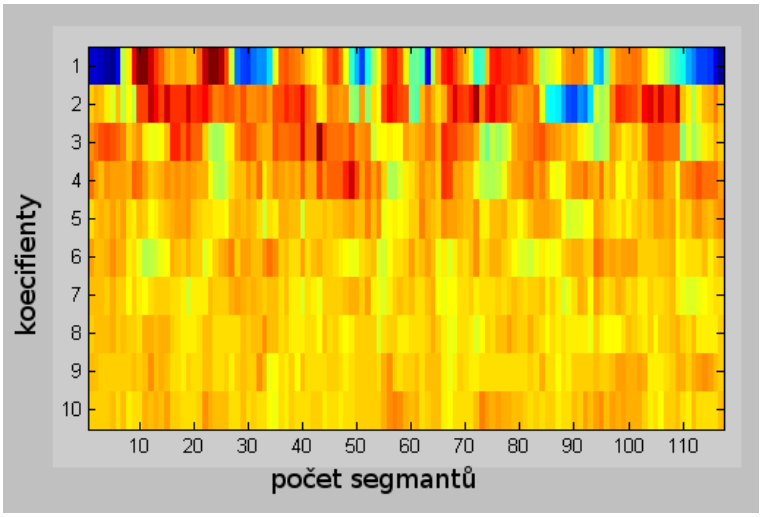
Banka melovských filtrů jako první brala v úvahu nelinearitu lidského sluchu. Například signál, který by měl 50Hz a jeho změna by byla o 10Hz, byla by změna posluchačem rozeznána snadněji, než změna z 1000Hz na 1010Hz. Proto byla vytvořena melovská stupnice, která tuto nelinearitu kmitočtů upraví a převede jí na lineární závislost. Pro převod z Hz na Mel se postupuje dle vztahu:

$$fm = 2595.log(\frac{f}{700}), \tag{5.7}$$

Po výpočtu výkonového spektra se následně provede vynásobení s bankou mel filtrů

- **Logaritmus**
- **Diskrétní kosínova transformace**

Diskrétní kosinovou transformací lze nahradit IFFT (Inverzní rychlou kosinovou transformací). Při tomto postupu je důležité dodržet délku okna, které nabývá hodnot $2^n (n \in R)$ a musí být počítáno s oboustranným spektrem. Po celém výpočtu se první koeficient smaže a je nahrazen logaritmem energie z časového průběhu signálu. Výsledná matice má velikosti $x.y$, kde x náleží počtu použitých mel filtrů a y náleží počtu segmentů. Výsledek se pak nejčastěji prezentuje barevně, jak je vidět na Obr. 5.4. Příloha C obsahuje vzorce, podle kterých byly vypočteny všechny příznaky.



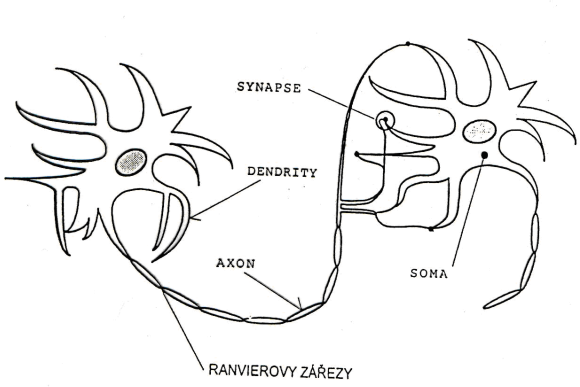
Obr. 5.4: Prezentace MFCC příznaku

6 NEURONOVÉ SÍTĚ

Vznik neuronových sítí je inspirován architekturou lidského mozku. Projevuje se u nich snaha napodobit chování lidského mozku při vyhodnocování určitých komplexně složitých problému, které by klasickými metodami nebyly řešitelné, nebo řešitelné jen velmi obtížně. Počátek vzniku UNS se datuje rokem 1943, kdy byl popsán matematický model neuronu. V roce 1958 byla vytvořena první neuronová síť, která byla nazvána Perceptron. Vymyslel ji psycholog Frank Rosenblatt a byla vytvořena jako model pro postup, při kterém lidský mozek zpracovává obraz a snaží se identifikovat předmět. Posléze se začalo využívat toho, že jsou UNS schopny samy vytvářet vzorce a samy se učit. Tohoto je využito pro mnoho různých odvětví, jako např. : energetika, medicína, telekomunikace, průmysl, finančnictví a bankovnictví, meteorologie, doprava, geologie, astronautika, obchod,...atd., kde se UNS dosud využívají. Pokud bychom chtěli najít jednotnou definici termínu UNS, tak bychom hledali marně. Jednou z nejpravděpodobnějších definicí by byla vydaná americkou společnosti DARPA. Tato definice která říká, že: "Neuronová síť je systém sestávající z mnoha jednoduchých procesorů, pracujících paralelně, jejichž funkce je definována strukturou sítě, intenzitou propojení a zpracováním ve výpočetních elementech , nebo uzlech." [19]

Základní pojmy Umělých neuronových sítí

Model biologické neuronové sítě Hlavním stavebním prvkem mozku je nervová buňka čili neuron. Mozek obsahuje okolo 100 miliard neuronů, jejichž velikost je $0,001$–$0,005$ milimetru v průměru a z nichž vystupuje několik dendritů a jeden axon. Každý z dendritu se větví podobně jako strom. Axon zůstává stejný, až ve své poslední části se dělí a jeho konce se přichycují k buněčnému tělu či na dendrity jiného axonu. Synapse je místo, kde je neuron v kontaktu s větvením axonu jiné buňky. Každý neuron vysílá svým axonem informace jiným neuronům. Pro lepší pochopení je na Obr. 6.1. zobrazen biologický vzor neuronu s nejdůležitějším popisem.



Obr. 6.1: Biologický vzor mozku převzato z [19]

Soma – **tělo neuronu s buněčným jádrem**
Dendrity **- výběžky vedoucí vzruchy směrem k buňce (délka** $1–3mm$**,** 10000 **vstupu)**
Axon – **výstup z neuronu (několik mikrometrů – délka až** $100cm$**)**
Ranvierovyzezy **- neboli opakovač na vedeních**
Synapse – **zprostředkují informační styk mezi navzájem spolupracujícími neurony = informační rozhraní (elektrochemické vazby)**

Matematický model neuronu si lze představit jako matematickou funkci do které vstupuje n**-rozměrný vektor parametrů (vstupních signálů). Výstupem je pak m-rozměrný skalár (výstupního signálu). Každý model neuronu obsahuje dvě funkce. Obvodovou (net function) a aktivační funkci (activation function). Obvodová funkce poskytuje informace o tom, v jaké závislosti budou uvnitř elektronu kombinovány vstupní parametry. Aktivační funkce jsou často připodobňovány k přenosovým funkcím, které známe z teorie obvodů [19].**

6.1 Principy učení umělých neuronových sítí a jejich rozdělení

Abychom byli schopni neuronových sítí využít, je také důležité vědět, co se děje při procesu učení. Jak poznat, jestli se ubíráme správným směrem či nikoliv. V této podkapitole bude popsán proces učení. Pojmy učení a paměť jsou spjaty s možností přizpůsobovat se vstupním podnětům. Tato vlastnost se nazývá plasticita synapsí. Je způsobena eliminací, nebo posilováním synapsí. U UNS se procesu učení říká trénink a lze

I	II	III
neasociativní	s učitelem	jednorázové
asociativní	bez učitele	opakované

Obr. 6.2: Dělení způsobu učení UNS převzato z [19]

jej definovat jako modifikaci synaptických vah a prahu podle zvoleného algoritmu učení. Základními charaktery učení jsou:

- **výběr charakteristických rysů a zkušeností ze vstupních signálů**
- **nastavení parametrů UNS tak, aby odchylka mezi požadovaným a skutečným výstupem při odezvě na soubor byla minimální**

6.2 Samoorganizující se neuronové sítě

Základním impulzem pro vytvoření tohoto druhu neuronové sítě, bylo dosáhnout podobnosti s fungováním lidského mozku. Lidský mozek přijímá ze svého okolí informace, které lze chápat jako vektory o různě velkých dimenzích. Tyto vektory jsou pak zpracovávány a postupně redukovány (odstraněny) nepotřebné informace. SOM pracují na totožném principu. Z vícedimenzionálních vstupních dat jsou vybrána ta data, která se jeví jako nejdůležitější přenositel informace. Tento proces je základním způsobem zpracování informací v přírodě a lze tedy konstatovat, že SOM zpracovává data totožně jako většina organizmů v přírodě. Funkce SOM je založena na soutěžním učení, kdy všechny neurony přijímají stejná data a pouze jeden nejpodobnější neuron se stává vítězem. Hodnoty kolem vítězného neuronu mívají podobné hodnoty a tvoří clustery (shluky). Procesu, kdy se neuronové sítí předkládají vstupní data, se nazývá trénování. Během trénování se mění struktura NS podle předem definovaných pravidel, které jsou zadávány na počátku trénovaní. Během procesu trénování jsou uchovávány nejdůležitější vazby a méně důležité vazby jsou odstraňovány. Jedná se tedy o kompresi informací. Proces trénování musí být zastaven ve správný okamžik, aby nedošlo k přetrénování. Přetrénování se rozumí takový okamžik, při kterém by NS důležitým informacím přikládala už menší váhu a soustředila by se na nepodstatné informace. Pro testování NS se užívají datá, které nebyla součástí trénovacích dat.

Využití SOM lze nalézt v mnoha oblastech lidské činnosti. Jako příklad lze uvést klasifikace dat, jejich kompresi a také možnost vizualizace. Nejznámějším typem SOM jsou Kohonenovy SOM, které jsou nazvány podle prof. Teuvo Kohonena z Helsinské technické univerzity. Na rozdíl od většiny ostatních neuronových sítí se SOM řadí do skupiny algoritmů s učením bez učitele (síti nejsou v průběhu učení předkládána data s požadovanými hodnotami).

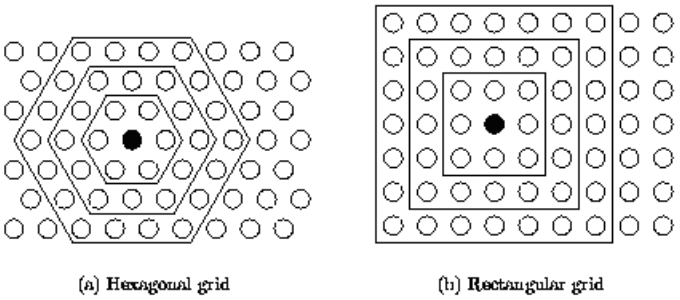
Existuje i metoda, která má informaci o přiřazení do určité skupiny dat a dle tohoto zařazení se v ní hledají podobnosti. Jedná se o metodu SSOM, která má oproti klasické SOM pravidelnou mřížku. Tento druh algoritmu má při grafické prezentaci zřetelnější hranice. Výsledek natrénované sítě bývá prezentován pomocí U-matice. U-matice bývá většinou prezentována ve škále šedi, kde tmavé oblasti mezi neurony reprezentují velké vzdálenosti tj. mezery a tím pádem hranice mezi oblastmi. Světlé oblasti pak oblasti blízké tzn. shluky (clustery).

Pro správné pochopení jak SOM fungují je potřeba si každý parametr popsat, abychom měli představu, jakým způsobem ovlivňují výsledek. Mřížka mapy Každý neuron je propojen se svými sousedy pomocí mřížky (lattice). Podle topologického způsobu uspořádání sousedních neuronů se dělí na :

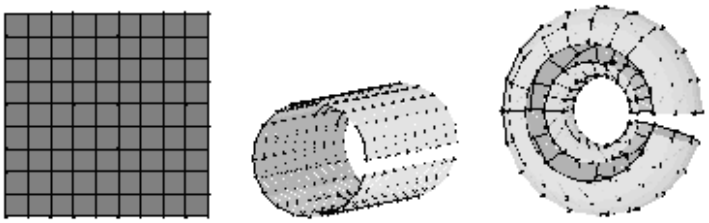
- **hexagonální**
- **pravoúhlé**

a následně podle mřížky rozprostžení mřížky v prostoru na:

- **dvourozměrnou**
- **válcovou**
- **toroidní**

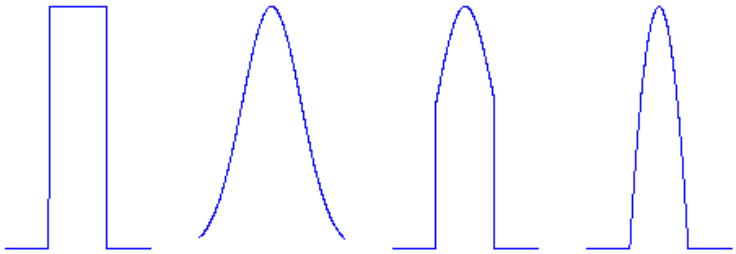


Obr. 6.3: Topologie mřížky převzato z [16]

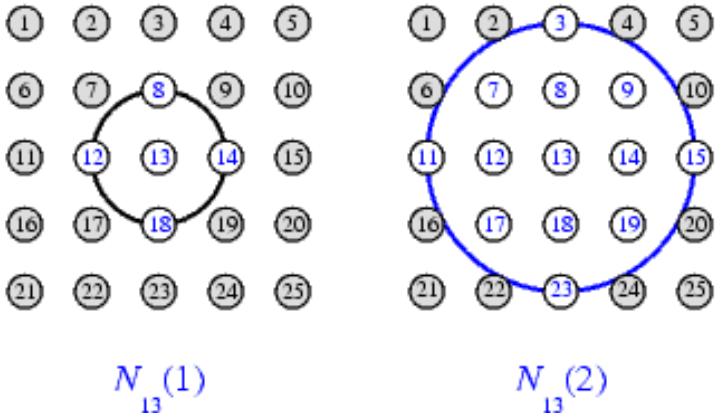


Obr. 6.4: Tvar mřížky zleva: sheet, cylinder, toroid převzato z [16]

Doporučením je volit hexagonální mřížku, která nepreferuje ani horizontální ani vertikální směry. Tvar mřížky se nejčastěji volí jako dvourozměrná, ostatní dva typy mřížky bývají voleny tam, kde předpokládáme že se budou vyskytovat data kruhovitého tvaru. Doporučením je, aby jedna hrana byla delší než druhá. Jako příklad lze uvést velikost mřížky [15 10]. Někde se doporučuje tvar mřížky volit čtvercově, zejména kvůli vizualizaci.



Obr. 6.5: Tvar mřížky zleva: sheet, cylinder, toroid převzato z [16]



Obr. 6.6: Ovlivněné sousední neurony převzato z Matlab help

6.2.1 Funkce souseda

Tato funkce představuje situaci, kdy předdefinované okolí je změněno v plném rozsahu hodot. Okolí ležící těsně za touto hranici není změnou poznamenáno vůbec. Dalšími funkcemi jsou:

- **Gausovská**
- **Gausovská ořízntutá**
- **Epanechicov**

V této funkci je také nastavováná hodnota vzdálenosti sousedů, kteří budou ovlivněni při nalezení a změně vítězného neuronu. Tato hodnota se mění v průběhu trénování z vyšších hodnot k hodnotám nižším. Je tím docílená prvotní plasticita tvaru neuronové sítě a po částečném ustálení dostatečná tuhost.

6.2.2 Algoritmus SOM

Ještě než je započato trénování NS, je potřeba nastavit tyto parametry:

- **Počet neuronů**
- **Topologie mřížky**
- **Tvar mřížky**
- **Dimenzi sítě**

Počet neuronů by měl byt volen tak velký, aby se co nejvíce blížil počtu trénovacích vektorů vstupujících do neuronové sítě. Nemělo by ovšem dojít k překročení počtu neuronů nad počet vstupních vektorů. Při počtu neuronů překračujících hodnotu 1000 se stává výpočetní proces velice výpočetně náročným a stává se nevhodným pro většinu aplikací. Doporučovaná hodnota počtu neuronů je daná dle [som], vztahem : $5 * sqrt(n)$**, kde** n **představuje počet vzorů vstupujicích do sítě.**

6.2.3 Inicializace

Inicializace, neboli nastavení počátečních hodnot lze u SOM provést třemi způsoby.

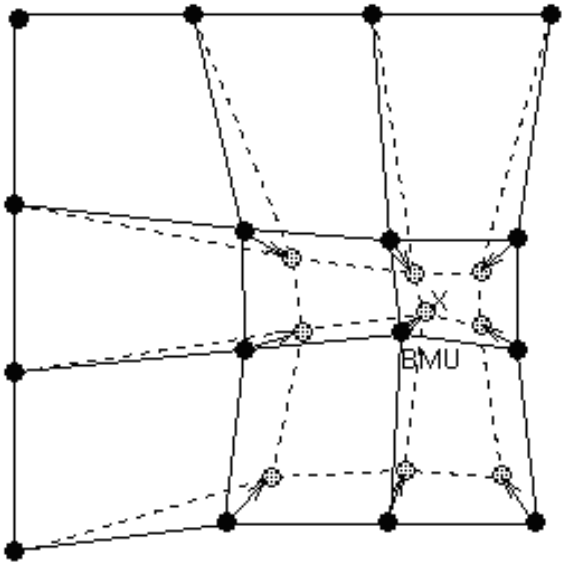
- **inicializace náhodnými čísly**
- **inicializace, která používá inicializační vzorky**
- **lineární inicializace.**

Náhodná inicializace se používá tam, kde o vstupním signálu víme příliš málo, nebo vůbec nic. Inicializace s inicializačními vzorky je výhodná v tom, že vzorky leží ve stejném vstupním prostoru s daty. Díky poslednímu druhu inicializace (lineární) můžeme dosáhnout rozpínání SOM do prostoru s nejvýznamnějším podílem energie. Je to dáno tím, že vstupní data korespondují nejpřesněji s vlastními daty. To je výhodné jen tehdy, když o zpracovávaném tématu hodně víme. Pokud SOM používáme hlavně proto, že neznáme zařazení jednotlivých trénovacích hodnot, pak je výhodnější použít inicializaci náhodnými čísly.

6.2.4 Proces trénování

Asi nejdůležitějším procesem je proces trénování. Tento proces je časově i výpočetně velice náročný. Při procesu trénování jde o to, že jsou NS postupně předkládány trénovací data, které se zobrazují v mapě. Při každém tomto zobrazení se vypočítává nejbližší neuron, který nejvíce odpovídá vstupnímu vzoru. Tento neuron je označen jako vítěz a jeho hodnoty spolu s hodnotami sousedních neuronů jsou přenastaveny tak, aby se co nejvíce blížily hodnotám vstupního vzoru. Nejbližší neuron se nejčastěji

počítá podle Euklidovy vzdálenosti. Vítězný neuron je obvykle označován jako BMU. Pro správný výpočet BMU je důležité, aby všechny vektory byly normovány tzn. vzdálenost mezi hodnotami v extrémech funkce by měly být stejné. Pokud by toto nebylo provedeno, pak by se nejvyšší důležitost přikládala dimenzi s nejvyšším rozdílem mezi maximální a minimální hodnotou funkce. Na obr. 6.7. lze vidět změnu (update), která se provádí při nalezení vítězného neuronu (BMU). V obrázku x značí místo, kde se zobrazí (promítnou) vstupní data. Plnou čarou jsou značeny topologické vazby mezi neurony před úpravami. Přerušovanou čarou jsou zobrazeny nově upravené vazby mezi neurony. Vítězným neuronem, jak je patrné z obrázku, se stane neuron pod vstupními daty vlevo od symbolu x", protože je nejblíže. Z tohoto pohledu je patrné, že nastavením vstupních parametrů, jako jsou počet ovlivňovaných sousedních neuronů, by při trénování vedl k odlišným výsledkům. Pokud by bylo okolí vítěze voleno příliš velké, pak by se v každém kroku měnilo velké množství hodnot a proces by byl velmi zdlouhavý. Pokud by ovšem bylo okolí volené příliš malé, pak by se mohlo stát, že budeme mít mnoho malých separovaných oblastí (clusteru). Tyto oblasti se vyznačují vyšší hustotou neuronů v tomto místě, oproti jiným místům v mapě.

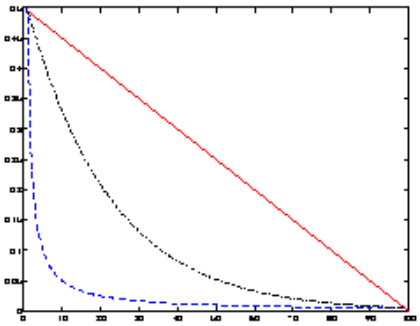


Obr. 6.7: Změna mřížky při nalezení vítězného neuronu převzato z [16]

6.2.5 Trénovací parametr

Jedná se o funkci, která nám vypovídá o tom, v jaké částí trénování se proces nachází. Pokud je trénování na počátku, pak je hodnota tohoto parametru blízká 1, na konci trénování pak 0. Tento parametr tedy nabývá hodnot $0-1$. Tato funkce může mít různý průběh podle toho, jestli jsou předkládaná data závislá na čase či nikoliv. V somtoolboxu, který je použit v této práci, jsou předdefinovány tři průběhy trénovacího parametru, jejich průběhy jsou vidět na obr. 6.6.

- **lineární**
- **inverzní v čase**
- **power series**



Obr. 6.8: Trénovací funkce Lineární (červená), Inverzní v čase (modrá) a Power serie (černá) převzato z [16]

Délka trénování by měla být přinejmenším 10 kráte delší, než je počet vzorků použitých pro trénování [toolbox]. Na počátku trénování by měla být mřížka co nejvíce přizpůsobivá (elastická) a na konci trénovaní by změny měly být přizpůsobeny co nejméně (mřížka tuhá). Většinou se proces trénování rozděluje na dvě fáze. V první fázi by měla být mřížka nastavena na co nejvíce elastickou (s velkým počtem ovlivnitelných sousedních neuronů). V druhé fázi by pak měla být mřížka méně elastická (ovlivnění pouze vítězného neuronu), hodnota klesá k 0. V [16] je doporučeno, aby byla druhá fáze trénování čtyřikrát delší než fáze první. Pokud je vybrán Lineární průběh trénování, pak je první fáze přeskočena a začíná se až s fází druhou. Doporučená hodnota při počátku trénování je definovaná vztahem (Neighborhood radius): $max(msize)/4$

Doporučená hodnota trénovacího parametru (Training rate) by měla začínat u první fáze na hodnotě 0.5 **a u druhé fáze na hodnotě** 0.05 **[16].**

6.2.6 Trénovací algoritmus

Tento algoritmus nám určuje, zda je topologie upravená vždy po předložení jednoho prvku (on-line), nebo až po předložení všech prvků během jednoho cyklu (off-line). Existují čtyři trénovací algoritmy a to:

- **lininit**
- **randinit**
- **batch**
- **seq**

Nejčastěji se využívá batch algoritmus z důvodu jeho rychlejšího výpočtu v prostředí Matlab. Jeho výsledky jsou většinou lepší než výsledky jiných algoritmů. Při trénování pomocí batch algoritmu jsou nejprve spočítány všechny vzdálenosti a pak se teprve hledá minimální hodnota, která se následně stává BMU. V sekvenčním algoritmu se naopak spočítá vzdálenost pro jeden vstupní vektor a s ním se porovnává následující hodnota. Rychleji konverguje sekvenční alg.

Po všech doporučeních je potřeba připomenout, že všechny přednastavené a doporučené hodnoty jsou pouze obecné a záleží na konkrétní aplikaci, pro kterou bude tento algoritmus použit. Je proto potřeba experimentovat.

7 PRAKTICKÁ ČÁST PRÁCE

Praktická část diplomové práce byla vytvořena v programu Matlab. Jako základ pro práci byla zvolena berlínská databáze, která obsahuje kvalitní nahrávky. Databáze obsahuje celkem 7 emočních stavů. V databázi je obsaženo více než pětsettřicet promluv, které jsou namluveny 5 muži a 5 ženami v 10 různých vět. Každá nahrávka byla předložena 20 posluchačům, kteří měli po poslechnutí klasifikovat nahrávku do jednoho ze sedmi emočních stavů. Pokud nahrávka byla klasifikována více než čtyřikrát špatně, byla odstraněna. Další informace o databázi lze získat z internetové adresy [3] , odkud ji lze i stáhnout. Prostřední sloupec obsahuje První písmeno Původního německého stavu, které je vždy obsaženo v názvu souboru Jednotlivé nahrávky jsou označeny dvěma čísly a písmenem. První číslo odpovídá řečníkovi, druhé označuje promluvu a písmeno je zkratka pro daný emoční stav. Přehled zkratek emočních stavů s původním německým, dále anglickým a českým překladem je vidět v tabulce na obrázku 7.1.

A	Anger	W	Arger (Wut)	Vztek
B	Bored	L	Langeweile	Nuda
D	Disgust	E	Ekel	Znechucení
F	Fear	A	Angst	Strach
H	Happiness	F	Freude	Štěstí
S	Sadness	T	Treuer	Smutek
N	Neutral	N	Neutral	Neutralní

Obr. 7.1: Přehled emočních stavů : anglicky, německy, česky

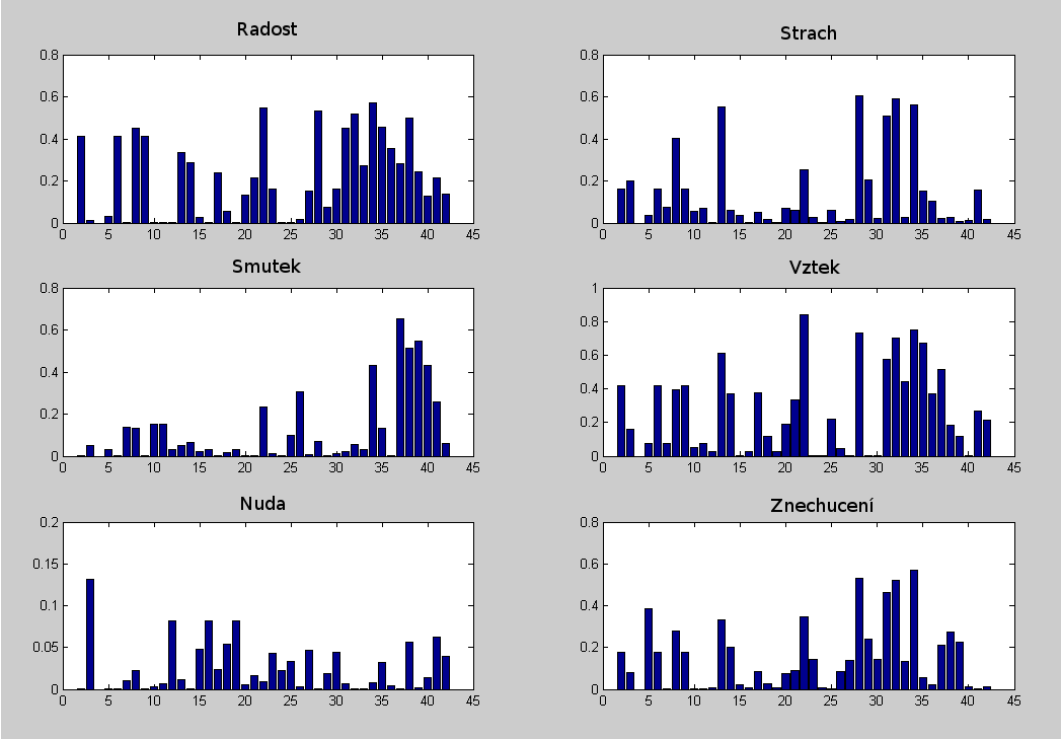
7.1 Parametrizace databáze

Prvním souborem, který je potřeba spustit, je soubor parametrizace.m, který si v první části načte obsah aktuálního adresáře obsahujícího nahrávky z Berlínské databáze (celkem 541 souboru). Postupně jsou v cyklu prováděny jednotlivé matematické operace. Pro každý soubor jsou vypočteny hodnoty popis, $F0$, En, ZCR, $MFCC$, které jsou uloženy v proměnné Data. Stručný popis ovládání lze nalézt v příloze A. Část popis obsahuje numerické hodnoty "1-7" získané z názvu souboru. Dále je vypočteno

vždy 10 statistických hodnot, které jsou uvedené v kapitole 5.5. kromě energie, kde je ještě přidán parametr rozdílu maximální a minimální hodnoty v decibelech. Hodnoty MFCC jsou vypočteny jako střední hodnota pro všechny rámce. Celkem tedy proměná Data, která je výstupem, obsahují matici o velikosti A x 42, kde A je počet klasifikovaných souborů a 42 je počet příznaků. Kompletní textový přehled příznaků lze nalézt v příloze B.

7.2 Redukce množiny příznaků

Pro redukci množiny příznaků bylo využito skriptu L2H redukce. Tento skript otestuje metodu hledání podle rozptylu, která je popsána v kapitole 4.2.1. jen s tím rozdílem, že nejlepší příznaky nabývají nejvyšších hodnot. Uvedené výsledky platí pro databázi (POUZIVANA DATABAZE.mat). Na obrázku 7.2 je vidět vzájemné porovnání nejlepších příznaků. X-ová osa reprezentuje pozici příznaků a y-ová reprezentuje kvalitu. Každá emoce je porovnávána s neutrální promluvou. Z obrázku



Obr. 7.2: Kvalita příznaku

je zřejmé, že nejhorší příznaky pro výběr vykazuje projev nudy. Nejvíce

kvalitních příznaků (nad hodnotu 0.5) vykazují emoční stavy : vztek, radost, strach. Jako nejvhodnější byly vybrány tyto pozice příznaků: Jejich slovní popis najdete v příloze B. Zde jsou uvedeny pouze číslelné pozice dle jejich nejčastějšího zastoupení. Jsou jimi příznaky z pozice 28, 32, 38, 34, 22, 41, 3, 16 . Po několika testech bylo zjištěno, že při takto vybraných příznacích SOM vykazuje horší výsledky, než při trénování se všemi 41 příznaky. Proto tato redukovaná množina nebyla zahrnuta do prezentovaných výsledků.

7.3 Naměřené hodnoty

Všechna měření se prováděly v programu Emoce, který byl vytvořen jako součást DP a je umístěn na přiloženém CD. Jeho návod lze nalést v příloze A. Pro závěrečné zhodnocení byla celá databáze rozdělena na dvě testovací množiny. Bylo to učiněno proto, aby bylo dobře prezentovatelné, že výsledky, kterých lze dosáhnout, jsou velice ovlivněny tím, s jakými konkrétními daty se pracuje. Zde prezentované výsledky byly vytvořeny z přiloženého souboru s názvem: POUZIVANA DATABAZE.mat, která je přiložena na CD. První test nalezení nejlepšího nastavení pro první, druhou množinu i celou databázi (celkem tři testy). První množina tvořila tyto emoční stavy:

- **Neutrální**
- **Radost**
- **Vztek**

Druhá množina tvořila tyto emoční stavy:

- **Strach**
- **Smutek**
- **Nuda**
- **Znechucení**

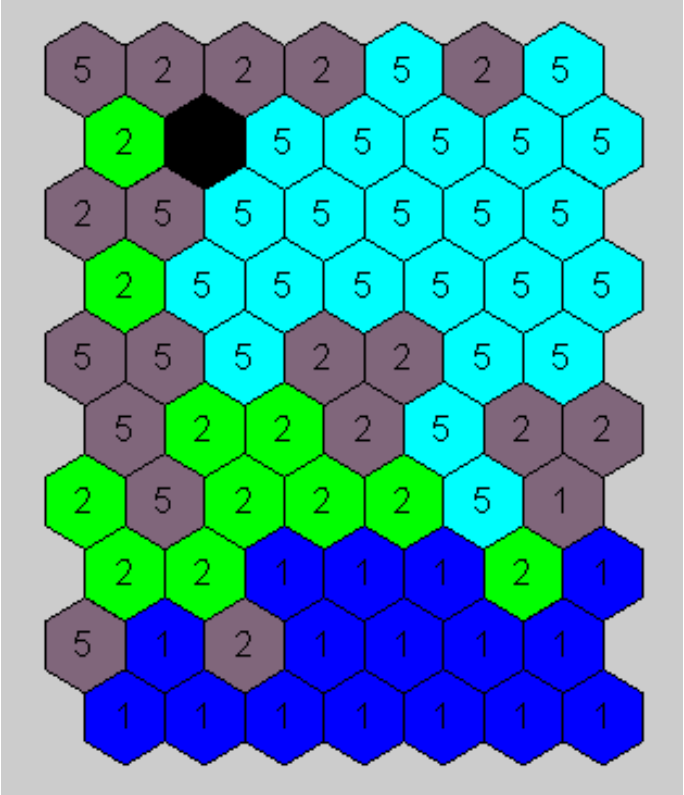
Výsledky prvního testu

V tomto testu jsou prezentovány nejlepší nalezené výsledky, které byly vytvořeny s daty obsaženými v souboru POUZIVANA DATABAZE.mat s poměrem trénovací a testovací matice 5:2. Z toho bylo použito 373 vektoru pro trénování a 153 pro testování. Počet příznaků obsažených v databázi je 41. Pozn. X označuje, že trénování Rough nebylo použito

První test	I	II	vše
Algorithm	SSOM	KSOM	KSOM
Type	linear	linear	linear
Map size	10 x 7	10 x 7	12 x 8
Training type	batch	batch	batch
Neighbourhood func	gaussian	ep	ep
Rough	X	X	X
Finetune	2;1;150	2.3;0.7;120	3.2;1.2;200
procentuální úspěšnost	84%	75%	52%

Obr. 7.3: Přehled nejlepších nalezených hodnot

Vždy bude pod sebou zobrazení rozložení neuronu v mřížce a následně tabulkdy obsahující procentuální hodnoty Neurony, které jsou zobrazeny černou barvou neobsahují popisek, protože nebyly vybrány jako nejlepší pro žádný z trénovacích vektorů. Na obr. 7.4 jsou šedou barvou zobrazeny neurony, které byly při testování špatně klasifikovány.

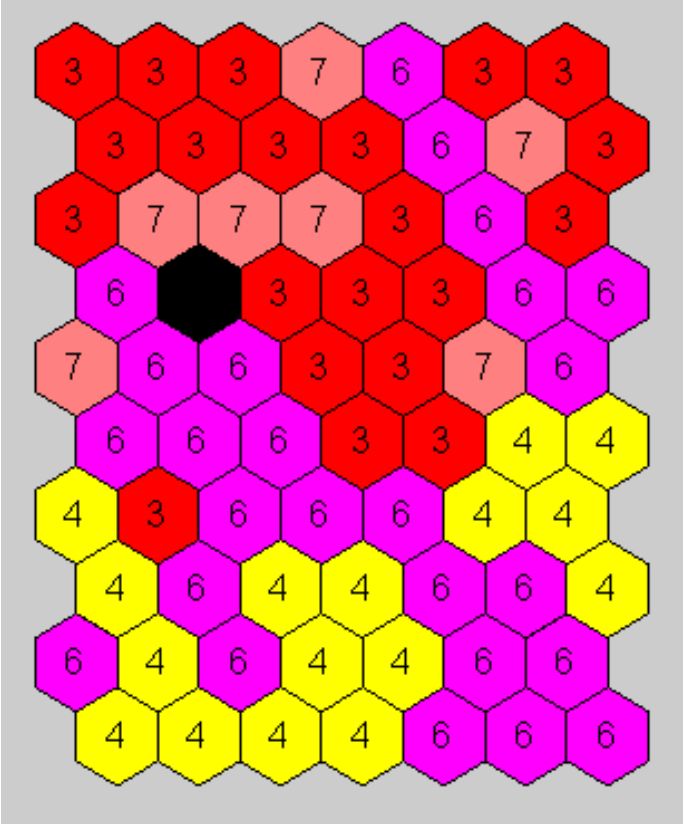


Obr. 7.4: Zobrazení rozložení neuronu v mřížce pro I skupinu emocí

I Test	'1' Neutrální	'2' Radost	'5' Vztek	Neoznačený neuron
Neutrální	91	4	0	4
Radost	5	65	30	0
Vztek	0	5	91	2

Obr. 7.5: Tabulka úspěšnosti pro I skupinu emocí (procentuální)

Tato skupina vykazovala hodnoty úspěšné klasifikace 84%

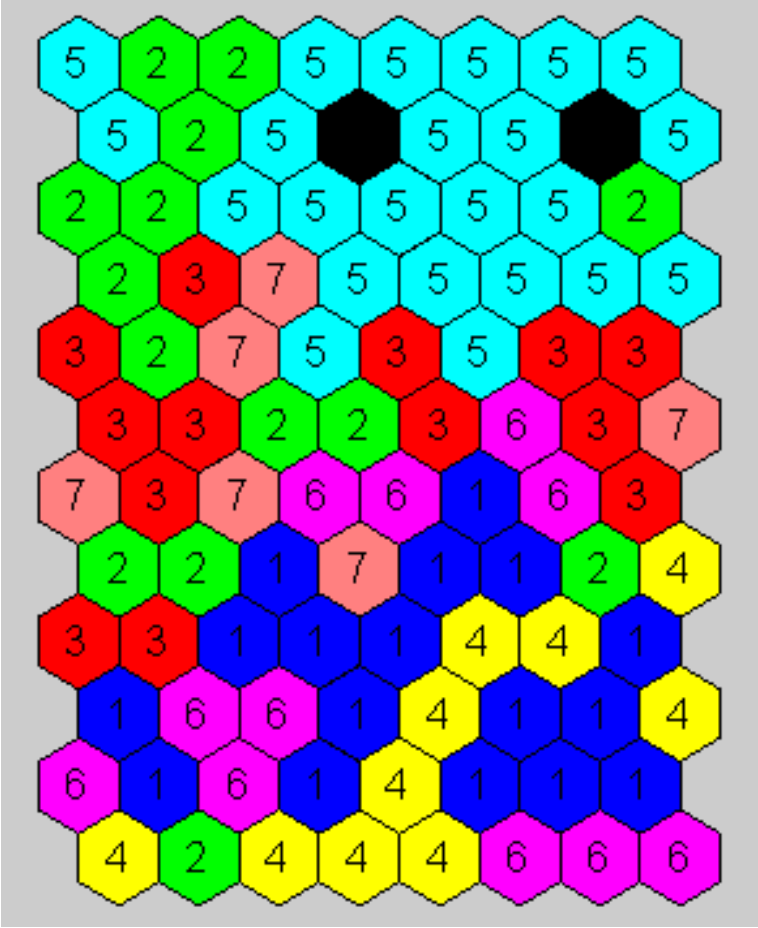


Obr. 7.6: Zobrazení rozložení neuronu v mřížce pro II skupinu emocí

	'3' Strach	'4' Smutek	'6' Nuda	'7' Znechucení	Neoznačený neuron
Strach	89	0	5	0	5
Smutek	0	88	11	0	0
Nuda	0	21	73	0	4
Znechucení	21	7	28	42	0

Obr. 7.7: Tabulka úspěšnosti pro II skupinu emocí (procentuální)

Tato skupina vykazovala hodnoty úspěšné klasifikace 75%



Obr. 7.8: Zobrazení rozložení neuronu v mřížce pro všechny emoce

	'1' Neutrální	'2' Radost	'3' Strach	'4' Smutek	'5' Vztek	'6' Nuda	'7' Znechucení	Neoznačený neuron
Neutrální	65	0	8	13	0	13	0	0
Radost	5	30	15	0	35	0	15	0
Strach	21	21	21	5	15	0	10	5
Smutek	11	0	0	83	0	5	0	0
Vztek	0	11	5	0	80	0	2	0
Nuda	39	4	8	21	0	26	0	0
Znechucení	14	0	42	7	0	0	35	0

Obr. 7.9: Tabulka úspěšnosti pro II skupinu emocí (procentuální)

Pro všechny emoční stavy úspěšnost klasifikace klesla k hodnotě 52%

Výsledky druhého testu

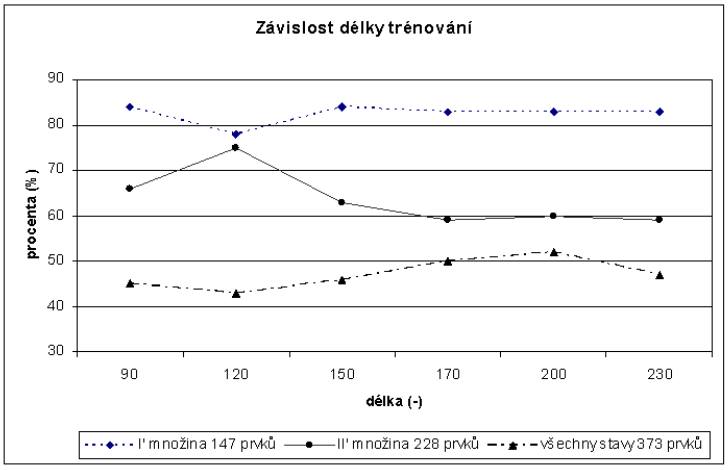
V tomto testu jsou prezentovány výsledky při změně volby Neighbourhood func (funkce souseda). Všechny ostatní hodnoty byly ponechány. Byl vždy změněn jen tento jediný parametr.

	I	II	vše
gaussian	84	60	36
cutgauss	72	47	36
ep	75	75	52
bubble	73	44	37

Obr. 7.10: Závislost nastavení funkce Neighbourhood

Výsledky třetího testu

V testu jsou prezentovány výsledky při změně parametru Length func (délka trénování). Ostatní parametry byly opět ponechány jako u předchozího testu. V popisech je napsan počet prvků, který množina obsahovala.



Obr. 7.11: Závislost nastavení délky trénování

Ze závislosti je patrné, že nejlepších výsledků dosahuje hodnota, která je blízká jedné polovině celkového počtu prvků v databázi. Všechny uvedené výsledky se vztahují k lineárnímu typu trénování. Náhodný typ zde

nebyl uveden, protože výsledky tohoto typu byly vždy pro stejné nastavení různé (celá mapa byla jinak nainicializována jinými hodnotami). Celkově lze tvrdit, že s tímto typem inicializace lze někdy možna dosáhnout lepších výsledku (jednotky procent), ale převážně tento typ trénovaní vykazoval horší výsledky (v řádu až desítek procent), ale při opakovaném tréninku nebude nikdy dosaženo stejných výsledků. Sekvenční typ trénování vykazuje delší trénovací časy. Obsahuje také vyšší počet nastavovacích parametrů, které se na výsledku podílí. Tento trénovací typ nebyl příliš důkladně prozkoumán. Jak je z tabulky 1 zřejmé, Rough (hrubé) trénování nebylo nikdy použito. Je to pravděpodobně proto, že mapy, na kterých bylo prováděno testování, měly malé rozměry a nastavování plastické“ mřížky vedlo k přesunu neuronu do míst, ze kterých už se v průběhu trénování Finetune (závěrečné) nepřesunuly na vhodnější místo v mapě. Funkce radius initial ovlivňuje počáteční vzdálenost okolí sousedních neuronů od BMU, na kterém se změny mají provádět. Funkce radius final ovlivňuje konečnou vzdálenost okolí sousedních neuronů od BMU, na kterém se změny mají provádět. Funkce training length určuje kolikrát budou předloženy trénovací data NS. V závěru lze dodat, že jako nejlépe klasifikované emoční stavy lze považovat smutek (83%), vztek (80%) a neutrální proslov (65%). Je zajimavé, že tyto stavy se neshodují se stavy, které vykazovaly nejlepší použitelné příznaky uvedené v podkapitole 7.2. Naproti tomu projev nudy, byl opět klasifikován jako nejhorší na rozpoznání.

8 ZÁVĚR

Cílem diplomové práce bylo otestovat možnost použití samoorganizujících se neuronových síťí ke klasifikaci emocí z řečového signálu. K tomuto byla využita berlínská databáze, která obsahuje přes pětsettřicet promluv. Počty jednotlivých promluv se u každého stavu liší z důvodu špatného rozpoznání posluchači.

V teoretickém rozboru je stručně popsán proces tvorby řečového signálu. Následuje kapitola, která se zabývá popsaním emočních stavů člověka, jejich vznikem a stručným popisem parametrů, podle kterých by mohlo být prováděno rozpoznávání. V další části je popsán teoretický model procesu klasifikáce emočních stavů se všemi jeho bloky. Jako základní příznaky byly vybrány frekvence základního tónu řeči, energie, počet průchodů nulovou hodnotou a melovské kepstrální koeficienty. Jedná se o příznaky, které jsou nejčastěji zmíňovány v literatuře. V celém procesu rozpoznávání emocí je jedním z klíčových prvků klasifikátor. V tomto případě SOM, který je popsán teoreticky v další kapitole spolu s různými doporučeními.

V kapitole Praktická část jsou prezentovány dosažené výsledky pomocí programu, které byly vytvořeny v této práci. Jsou jimi program Parametrizace, který umožňuje z databáze vytvořit příznaky. Jako základní předpřipravené možnosti příznaků jsou frekvence základního tónu řeči, energie, počet průchodů nulovou rovinou a melovské kepstrální koeficienty. Další možnosti tohoto programu dovolují doprogramovat si vlastní sadu příznaků.

Dalším programem je pak program Emoce. Tento program z parametrizovaného souboru, vytvořeného v první části dovoluje vybrat pouze určité emoční stavy a ty nechat klasifikovat pomocí SOM. Výsledky jsou prezentovány ve formě zobrazení neuronové sítě a tabulky zobrazující procentuální úspěšnost klasifikace jednotlivých stavů.

Jako zhodnocení lze uvést, že výsledky jsou velice závislé na konkrétních datech a způsobu nastavení SOM. Pro možnost prezentace výsledků byly vytvořeny tři různé kombinace porovnávaných emočních stavu Pro první množinu stavů se jednalo o kalsifikaci stavu neutrálního, radosti a vzteku. Druhá množina obsahovala strach, smutek, nudu a znechucení. Do poslední množiny byly zařazeny všechny emoční stavy. Nejlepší dosažený výsledek pro první kombinaci byl 84 % , pro druhou kombinaci 75 % a pro všechny emoční stavy pak úspěšnost klesla k hodnotě 52 % . Všech těchto výsledků bylo dosaženo při různém nastavení. Nelze proto

zhodnotit, jaké nastavení vede k nejlepším výsledkům. Z celkových výsledků lze jen tvrdit, že projev smutek, vztek a neutrální proslov byly klasifikovány vždy s nejvyšší přesností a projev nudy vykazoval nejhorší procentuální úspěšnost.

Závěrem lze dodat, že tyto programy lze použít i k další práci. Například mírnými úpravami lze tyto programy využít jako klasifikátory pro rozpoznání pohlaví, kde by bylo jako klasifikátoru opět využito samoorganizující neuronové síťě.

LITERATURA

[1] **ATASSI, H.** ***Metody detekce základního tónu reci. Elektrorevue - internetový časopis*** **[online]. 2008, č.4, s 1-17 [cit. 2010-05-24] ISSN 1213-1539 Dostupné z WWW: http://www.elektrorevue.cz.**

[2] **ATASSI, H.** ***Zavedení problematiky rozpoznání vzoru do výuky předmětu zpracování řeči. 1. 2009. s. 1-27.***

[3] **Berlin Database of Emotional Speech [online]. 2005, [cit. 2010-05-24] Dostupné z WWW: http://pascal.kgw.tu-berlin.de/emodb/index-1280.html.**

[4] **ČERMAK, J.** ***Rozpoznávání emočních stavů na základě analýzy řečového signálu. Brno:Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2009. 66 s.***

[5] **ČERNÝ, L.** ***Spektrální rozpoznávání vybraných úseků řečového signálu. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2007. 51 s.***

[6] **SOM Toolbox.** ***2005 [cit. 2010-05-24]. SOM Toolbox. Dostupné z WWW: http://www.cis.hut.fi/projects/somtoolbox/.***

[7] **HLAVICA, M.** ***Databáze emoční řeči. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2009. 52 s.***

[8] **KRČMOVÁ, M.** ***Fonetika. Dostupné z WWW: http://is.muni.cz/do/1499/el/estud/ff/js07/fonetika/materialy/index.html/.***

[9] **HOUDEK, M.** ***Rozpoznání emočního stavu člověka z řeči. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2009. 62 s. Vedoucí diplomové práce Ing. Hicham Atassi.***

[10] **MARTINOVIČ, J.** ***Information Retrieval a shlukování metodouWEBSOM Dostupný z WWW: http://mrkni.cz/martinovic/publications/doc/WEBSOM.pdf***

[11] **NAVRÁTIL,M.** ***Rozpoznávání emočních stavů pomocí analýzy řečového signálu. Diplomová práce, 2008. 66s.***

[12] **PSUTKA, J.** ***Mluvíme s počítačem česky. 1. vyd. Praha : ACADEMIA, 2006. 752 s. ISBN 80-200-1309-1***

[13] **SIGMUND, M.** ***Analýza řečových signálů : přednášky. 1. vyd. Brno : MJ servis, s.r.o., 2000. 86 s. ISBN 80-214-1783-8.***

[14] **SMÉKAL, Z.** ***Číslicové zpracování signálů : skripta k předmětu MCSI. Skripta [online]. 2007,s. 1-149. Dostupný z WWW: www.feec.vutbr.cz/et.***

[15] **SMÉKAL, Z.** ***Číslicové zpracování řeči : skripta k předmětu MZPR. Skripta [online]. 2007,s. 1-134. Dostupný z WWW: www.feec.vutbr.cz/et.***

[16] ***SOM Toolbox [online]. 2005. 2005 [cit. 2010-05-24]. SOM Toolbox. Dostupné z WWW: http://www.cis.hut.fi/projects/somtoolbox/.***

[17] **SYROVÝ, V.** ***Hudební akustika. Akademie múzických umení, Praha 2003. ISBN 80-7331-901-2***

[18] **TUČKOVÁ, J.** ***Aplikace umělých neuronových sítí při zpracování signálů vyd. Nakladatelství ČVUT, 2009. 135s. ISBN 978-80-01-04400-1***

[19] **TUČKOVÁ, J.** ***Úvod do teorie a aplikací umělých neuronových sítí. 2003. Praha : Vydavatelství ČVUT, 2003. 103 s.***

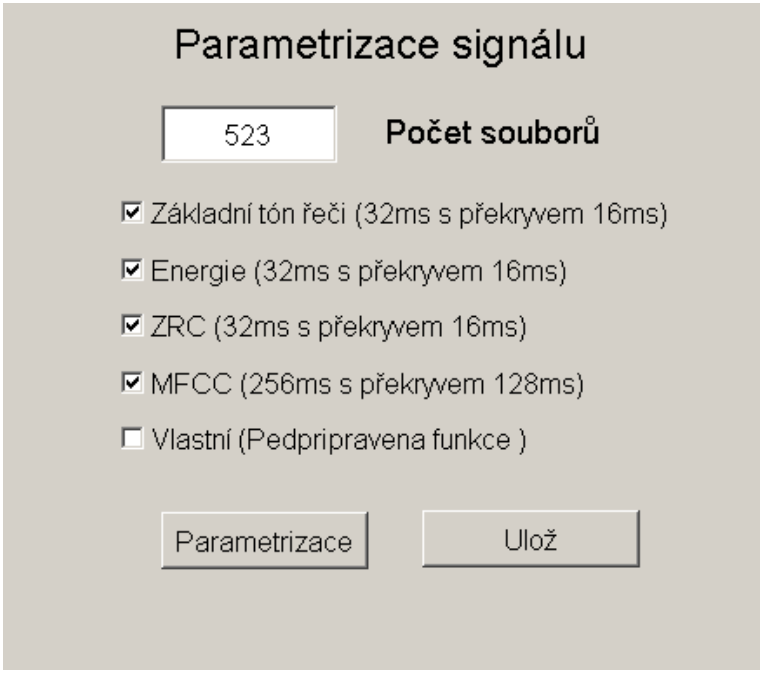
[20] **VLČKOVÁ, J.** ***Prozodie, cesta i mříž porozumění. Vyd. 1. Praha : Karolinum, 2006. 204 s. ISBN 80-246-1266-6.***

9 ZKRATKY

BMU Best match unit vítězný neuron
BNS Biologická neuronová síť
GSS Gausovy smíšené modely
NS Neuronová síť
SOM Samoorganizující se neuronové sítě
SMM Skryté Markovy modely
STT Speach to text předod mluvené řeči do textové podoby
UNS Umělé neuronové sítě
Rectangular pravouhlé
Lattice mžížka
Sheet plochá
Cylinder válcová
Updata změna
Cluster malá oblast s velkým výskytem podobných hodnot

A PŘÍLOHA A – OBSLUHA PROGRAMU

Program byl vytvořen v Matlabu verze 7.9.0 (R2009b), který pravděpodobně nebude kompatibilní se staršími verzemi, protože ty neobsahují některé funkce, které byly potřeba při tvorbě Grafického rozhraní, např. zobrazování tabulky. Program je rozdělen na dvě části. Jedná se o část pro parametrizaci hudebních (wav) souborů a o část sloužící ke klasifikaci pomocí SOM NS a zobrazování jejich výsledků. Nejdříve je nutné nastavit Matlab na aktuální adresář, ve kterém se programy nacházejí $(PC : SOMLabtvorbadatabaze)$**. První část se provádí spuštěním souboru parametrizace.m , kdy se zobrazí okno, které je vidět na Obr A.1. Nejprve**



Obr. A.1: Okno programu parametrizace

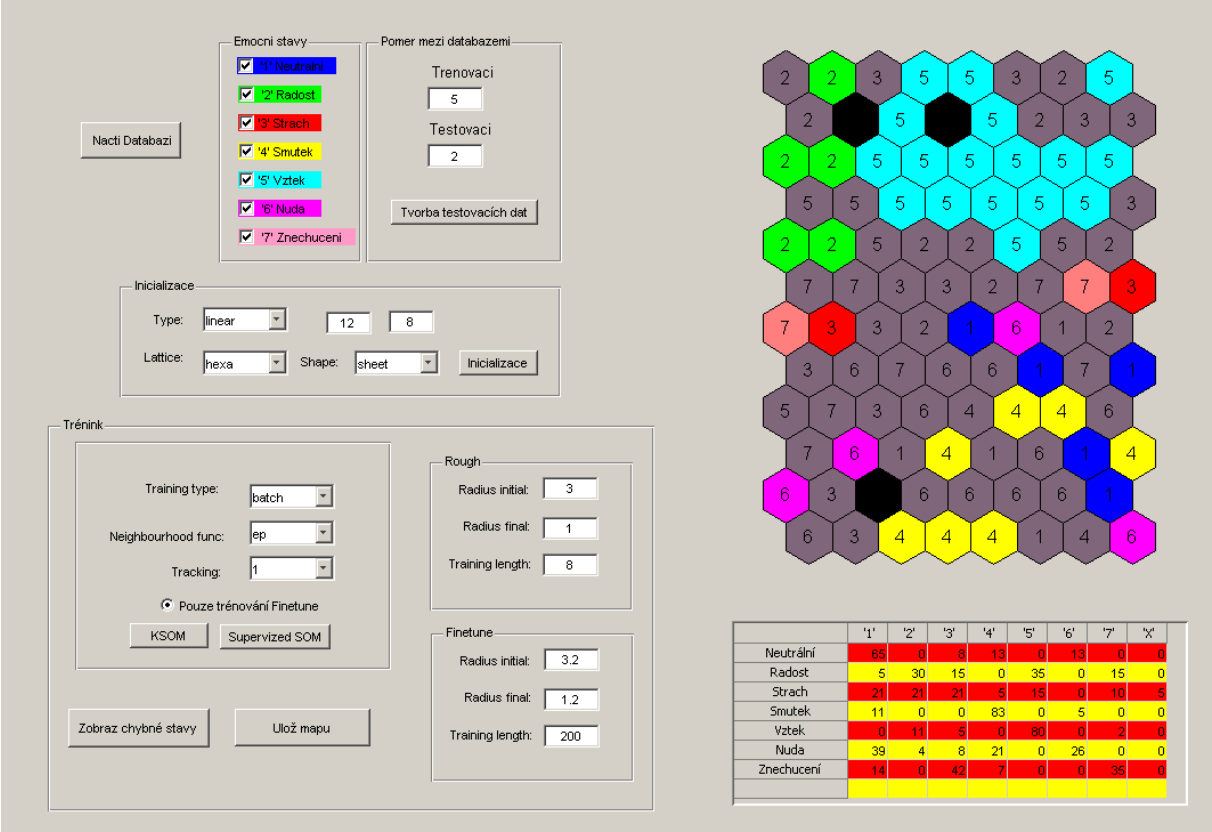
je nutné zvolit, kolik souborů máme připraveno k parametrizaci. Z Berlinské databáze bylo použito celkem 523 souborů. Při parametrizaci některých souborů docházelo k chybám a proto byly odstraněny. Protože program byl tvořen za účelem parametrizace této databáze, musí být pro parametrizaci jiné databáze dodrženo několik pravidel, aby vše proběhlo úspěšně. Soubory se musí vyskytovat na začátku adresáře. Toho docílíme např. přidáním číslice před původní název souboru.

Pro rozlišení o jakou emoci se jedná, musí být ve jménu souboru obsaženo písmeno dle konkrétního zařazení z této tabulky zobrezené na obr. A.2.: Například jedná-li se o soubor obsahující emoční stav strach, pak je

A	Anger	W	Arger (Wut)	Vztek
B	Bored	L	Langeweile	Nuda
D	Disgust	E	Ekel	Znechucení
F	Fear	A	Angst	Strach
H	Happiness	F	Freude	Štěstí
S	Sadness	T	Treuer	Smutek
N	Neutral	N	Neutral	Neutralní

Obr. A.2: Přehled názvů emočních stavů

nutné do jména souboru přidat velké písmeno A“ Při zaškrtnutí jednotlivých tlačítek budou do parametrizovaných dat přidány informace, které jsou zvoleny. Přehled jednotlivých parametrů je uveden v příloze B. Poslední zatrhávací tlačítko Vlastní“ obsahuje volbu vlastního parametru, který lze vložit do souboru L1Vlastni.m. Výstupem souboru L1Vlastni.m musí být řádkový vektor, který musí mít pro každý soubor stejnou délku. Po stisknutí tlačítka Parametrizuj se databáze parametrizuje. Čas, za jaký je parametrizace provedena, je závislý na počtu souborů a složitosti výpočtů. (523 souborů Berlinské databáze se všemi 4 předpřipravenými hodnotami trval přibližně minutu). Po parametrizaci a stisknutí tlačítka Ulož se výsledek uloží do mat souboru, jehož název bude zadán. Druhá částprogramu se nachází v jiném adresáři, proto je potřeba změnit cestu, například na $(PC : SOMLabsom)$ **a zadáním příkazu Emoce.m. Nejprve je nutné načíst parametrizovanou databázi tlačítkem Načti Databázi (převážně se jedná o soubor, který byl vytvořen v první části) a následně vybrat alespoň 2 ze 7 stavů, které chceme porovnávat. Dalším parametrem, který je potřeba zadat, je poměr mezi trenovací a testovací množinou. Tato hodnota je přednastavena na poměr 5:2. Pokračujeme stisknutím tlačítka Tvorba testovacích dat. Následuje krok inicializace, který přednastaví optimální velikost mapy vzhledem k počtu prvků použitých k trénování. Následuje volba parametrů, které jsou potřebné k trénování (jsou umísťěny v oblasti Trénink). Mezi hodnoty, které se mohou nastavovat patří: Nighbourhood func (funkce souseda), hodnoty Rough**



Obr. A.3: Uživatelské prostředí funkce Emoce

(hrubého tréninku) a hodnoty Finetude (závěrečného tréninku). Radius initial představuje hodnotu vzdálenosti od BMU, která má být upravována na počátku tréninku. Hodnota Radius final pak hodnota na konci tréninku. Training length pak počet opakování předkládaných dat. Po všech nastaveních je potřeba zvolit, jestli se bude trénovat pouze s hodnotami Finetude, nebo dvoufázově tzn. i s hodnotami Rought tréninku. Po všech těchto nastaveních je třeba zvolit možnost testování buď KSOM, nebo Supervised SOM. Po stisknutí jednoho z těchto tlačítek je síť natrénovaná a výsledky jsou zobrazeny v pravé části programu. V horní části je vidět natrénovaná mapa a v dolní části tabulka procentuální úspěšnosti natrénovaných hodnot. V tabulce jsou na řádcích uvedené emoční stavy, které jsou mapě předkládány a ve sloupcích je uvedené zařazení do jednotlivých emočních stavů, kde například '1' znamená neutrální stav. Postupuje se postupně, tak jak jsou emoční stavy uvedeny v oblasti Emoční stavy v levé části programu. Po stisknutí tlačítka Zobraz chybné stavy se v oblasti mapy zobrazí šedou barvou neurony, které byly špatně klasifikované jako správné.

B PŘÍLOHA B – JMENÝ SEZNAM POZIC V PROMĚNÉ DATA

1 Popis emočního stavu (hodnoty 1-7)
2 Hodnota maximálního parametru F0
3 Pozice hodnoty maximálního parametru F0
4 Hodnota minimálního parametru F0
5 Pozice hodnoty minimálního parametru F0
6 Rozdíl mezi maximální a minimální hodnotou parametru F0
7 Střední hodnota ze všech parametrů ve všech rámcích F0
8 Směrodatná odchylka ze všech parametrů ve všech rámcích F0
9 Koeficient variability vstupní hodnoty F0
10 Rozptyl vstupních hodnot F0
11 Směrodatná odchylka vstupních hodnot F0
12 Hodnota maximálního parametru Energie
13 Pozice hodnoty maximálního parametru Energie
14 Hodnota minimálního parametru Energie
15 Pozice hodnoty minimálního parametru Energie
16 Rozdíl mezi maximální a minimální hodnotou parametru Energie
17 Střední hodnota ze všech parametrů ve všech rámcích Energie
18 Směrodatná odchylka ze všech parametrů ve všech rámcích Energie
19 Koeficient variability vstupní hodnoty Energie
20 Rozptyl vstupních hodnot Energie
21 Směrodatná odchylka vstupních hodnot Energie
22 Rozdílu maximální a minimální hodnoty Energie v decibelech
23 Hodnota maximálního parametru ZCR
24 Pozice hodnoty maximálního parametru ZCR
25 Hodnota minimálního parametru ZCR
26 Pozice hodnoty minimálního parametru ZCR
27 Rozdíl mezi maximální a minimální hodnotou parametru ZCR
28 Střední hodnota ze všech parametrů ve všech rámcích ZCR
29 Směrodatná odchylka ze všech parametrů ve všech rámcích ZCR
30 Koeficient variability vstupní hodnoty ZCR
31 Rozptyl vstupních hodnot ZCR
32 Směrodatná odchylka vstupních hodnot ZCR
33-42 Střední hodnoty pro všech 10 rámců MFCC

Číselná hodnota udává pozici parametru v databázi.

Příloha C

1. Hodnota maximáního parametru

$$F_{0\max} = \max(\overline{F_0})[Hz]$$

2. Pozice hodnoty maximálního parametru

$$F_{0\max\ pos} = 100\ \frac{find\ (F_{0\max})}{N}[-]$$

3. Hodnota minimalního parametru

$$F_{0\min} = \min(\overline{F_0})[Hz]$$

4. Pozice hodnoty minimalního parametru

$$F_{0\min\ pos} = 100\ \frac{find\ (F_{0\min})}{N}[-]$$

5. Rozdíl mezi maximální a minimální hodnotou parametru

$$F_{0\max-\min} = F_{0\max} - F_{0\min}[Hz]$$

6. Střední hodnota ze všech parametrů ve všech rámcích

$$F_0 mean = \frac{1}{N}\sum_{i=0}^{N-1}(F_0[i])[-]$$

7. Směrodatná odchylka ze všech parametrů ve všech rámcích

$$F_0 std = \sqrt{(\frac{1}{N}\sum_{i=0}^{N-1}F_0[x] - F_0\bar{x})^2}$$

$$\text{kde } \bar{x} = \frac{1}{N}\sum_{i=0}^{N-1}F_0[x]$$

8. Koeficient variability vstupní hodnoty

$$V_{F0} = (F_{0\max} - F_{0\min})\frac{|F_{0mean} - F_{0median}|}{F_{0mean}}[Hz]$$

9. Rozptyl vstupních hodnot

$$D_{F0} = \frac{\sum_{i=0}^{N-1} (F_0[i])^2}{N} - (F_{0\,mean})^2 [Hz]$$

10. Směrodatná odchylka vstupních hodnot

$$\sigma_{F0} = \sqrt{D_{F0}}[-]$$

11. Rozdíl maximální a minimální hodnoty v dB (Pouze u výpočtu energie)

$$EndB = \log_{10}\left(\frac{E_n \max}{E_n \min}\right)[dB]$$

Příloha D

Součástí diplomové práce je program který je uložena na přiloženém CD. Na tomto CD jsou tři adresáře:

Parametrizace

> Tento adresář obsahuje stejnojmený soubor, který složí pro parametrizaci databáze

Som

> V adresáři se naléza soubor Emoce, který provádí klasifikaci z parametrizované databáze.

Ostatní

> V tomto adresáři jsou obsaženy balíky, které byly používány během tvorby programu. Jedná se o balík SOM TOOLBOX a Berlínskou databázi atd.

V kořenovém adresáři je uložen PDF soubor, který obsahuje DP.