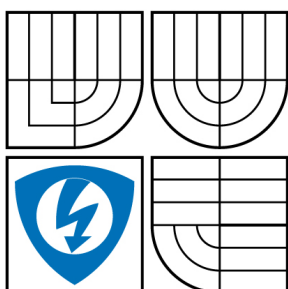




VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH
TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

ROZPOZNÁNÍ EMOČNÍHO STAVU ČLOVĚKA Z ŘEČI

AUTOMATIC VOCAL-ORIENTED RECOGNITION OF HUMAN EMOTIONS

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

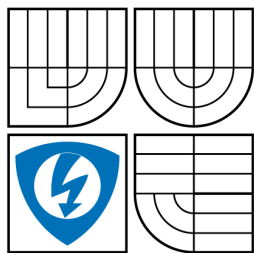
ALEŠ ADAMSKÝ

VEDOUcí PRÁCE

SUPERVISOR

Ing. HICHAM ATASSI

BRNO 2009



VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

Ústav telekomunikací

Bakalářská práce

bakalářský studijní obor
Teleinformatika

Student: Aleš Adamský

ID: 73101

Ročník: 3

Akademický rok: 2008/2009

NÁZEV TÉMATU:

Rozpoznání emočního stavu člověka z řeči

POKYNY PRO VYPRACOVÁNÍ:

Navrhněte a realizujte v libovolném prostředí algoritmus pro automatické rozpoznání emočního stavu člověka z řeči. Porovnejte různé metody předzpracování pro výpočet příznaků a metody klasifikace vstupního vzoru (neuronové sítě, skryté Markovovy modely atd.). Získané výsledky vhodným grafickým a numerickým způsobem reprezentujte.

DOPORUČENÁ LITERATURA:

- [1] Atassi H., Porovnání analýzy emočních stavů v závislosti na typu jazyka. Diplomová práce, Brno, VUT 2007.
- [2] Psutka J., Komunikace s počítačem mluvenou řeči. Academia, Praha 1995.
- [3] R. Duda, P. Hart, D. Stork, Pattern Classification, druhé vydání. Wiley, 2003.
- [4] Psutka J., Muller L., Matoušek J., Radová V., Mluvíme s počítačem česky. Academia Praha 2006.
- [4] Vlčková-Mejvaldová J., Prozodie, cesta i mříž porozumění. Karolinum, Praha 2006.

Termín zadání: 9.2.2009

Termín odevzdání: 2.6.2009

Vedoucí práce: Ing. Hicham Atassi

prof. Ing. Kamil Vrba, CSc.

Předseda oborové rady

UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být 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.

ANOTÁCIA

Práca sa zaoberá vlastnosťami, tvorením, reprezentáciou a analýzou reči a rečového signálu. Do detailov vysvetľuje pojmy ako emócia a prozódia. Rozoberá jednotlivé emočné stavy človeka, ako i prozodické parametre: intenzitu, harmonicitu a formanty. Jej súčasťou je i databáza emočných stavov človeka. Táto databáza bola analyzovaná pomocou programu Praat a boli vybrané vhodné príznaky pre rozpoznávanie emočných stavov človeka z reči. Ďalej sa zaoberá problematikou neurónových sietí. Obsahuje popis Java Neural Network Simulátoru, ktorý je využitý na rozpoznávanie emočných stavov človeka z reči. Výsledky rozpoznávania sú spracované v tabuľkách a grafoch pre jednoduchšiu orientáciu.

KLÚČOVÉ SLOVÁ

Príznak, emócia, rozpoznávanie, Java Neural Network Simulator, neurónová sieť, prozódia.

ABSTRACT

This work deals with the characteristics, formation, representation and analysis of speech and the speech signal. It explains the details of concepts such as emotions and prosody. It analyzes different emotion state of human, as well as prosodic parameters: intensity, harmonicity and formants. It includes a database emotional states of human. This database was analyzed using the Praat and suitable features were selected for detecting of emotional states of human speech. It also deals with neural networks. It contains a description of Java Neural Network Simulator, which is used to detect emotional states of human speech. Results of recognition are processed in tables and graphs for easier navigation.

KEY WORDS

Feature, emotion, recognition, Java Neural Network Simulator, neural network, prosody.

ADAMSKÝ, A. *Rozpoznání emočního stavu člověka z řeči*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2009. 45 s. Vedoucí bakalářské práce Ing. Hicham Atassi.

PREHLÁSENIE

Prehlasujem, že svoju bakalársku prácu na tému "Rozpoznání emočního stavu člověka z řeči" som vypracoval samostatne pod vedením vedúceho bakalárskej práce a s použitím odbornej literatúry a ďalších informačných zdrojov, ktoré sú citované v práci a uvedené v zozname literatúry na konci práce.

Ako autor uvedenej bakalárskej práce ďalej prehlasujem, že v súvislosti s vytvorením tejto bakalárskej práce som neporušil autorské práva tretích osôb, hlavne som nezasiahol nedovoleným spôsobom do cudzích autorských práv osobnostných a som si plne vedomý následkov porušení ustanovení § 11 a nasledujúcich autorského zákona č. 121/2000 Sb., vrátane možných trestnoprávných dôsledkov vyplývajúcich z ustanovení § 152 trestného zákona č. 140/1961 Sb.

V Brne dňa 25.5.2009

.....
(podpis autora)

POĎAKOVANIE

Moje poďakovanie patrí Ing. Hichamovi Atassimu za jeho odborné konzultácie, rady, zapožičané materiály a najmä námet tejto bakalárskej práce.

V Brne 25.5.2009

Ppodpis

OBSAH

Úvod.....	1
1 Emócie zo psychologického hľadiska	2
1.1 Definícia pojmov	2
1.1.1 Emócia.....	2
1.1.1.1 Kategórie emócií	3
1.1.1.2 Fyziologické zmeny pri emočnom podráždení	3
1.1.1.3 Analýzy emočného chovania	3
1.1.2 Postoj.....	4
2 Reč.....	5
2.1 Vlastnosti reči.....	5
2.2 Tvorenie reči človekom.....	5
2.2.1 Dych	6
2.2.2 Hlas.....	6
2.2.3 Artikulácia	7
2.3 Reprezentácia reči	8
2.4 Varianty reči	8
2.4.1 Šepot.....	9
2.4.2 Krik.....	9
2.4.3 Spev.....	9
3 Prozódia.....	10
3.1 Prozodické parametre	10
3.1.1 Intonácia	10
3.1.2 Intenzita	11
3.1.3 Trvanie	11
3.2 Funkcie prozódie	11
3.2.1 Jazykové funkcie prozódie	11
3.2.2 Fonoštylistické funkcie prozódie	12
3.3 Prozódia češtiny	12
3.3.1 Prízvuk češtiny	12
3.3.2 Tempo češtiny	12
4 Vytvorenie databázy emočných stavov	13
4.1 Zdroj databázy.....	13
4.2 Spôsob tvorby.....	13
4.3 Konečná forma	14
5 Analýza.....	15
5.1 Použitý software	15
5.2 Výsledky analýzy	15
5.2.1 Tabuľka výsledkov	15
5.2.2 Grafické znázornenie výsledkov	16
5.2.3 Zhrnutie výsledkov	25
6 Výber vhodných príznakov	26
6.1.1 Matlab.....	26
6.1.1.1 Rank features (hodnotenie príznakov)	26
7 Neurónové siete.....	27
7.1 História	27
7.2 Využitie	27
7.2.1 Rozdelenie oblastí využitia	28

7.2.1.1	Klasifikácia.....	28
7.2.1.2	Filtrácia.....	28
7.2.1.3	Predikcia.....	28
7.2.1.4	Kompresia	28
7.3	Neurón.....	28
7.4	Učenie.....	29
7.4.1	Typy učenia	29
7.4.1.1	Učenie s učiteľom.....	29
7.4.1.2	Učenie bez učiteľa.....	30
7.5	Používaný software	30
7.5.1	SNNS (Stuttgart Neural Network Simulator).....	30
8	Rozpoznávanie (klasifikácia)	31
8.1	JavaNNS.....	31
8.1.1	Inštalácia a prvé spustenie.....	31
8.1.2	Načítanie súborov.....	31
8.1.3	Práca so sieťou	31
8.1.4	Tvorenie novej siete	31
8.1.5	Inicializácia a tréning.....	32
8.1.6	Formát tréningových a vstupných súborov.....	32
8.1.7	Analýza.....	33
8.1.8	Výstupné dáta	33
8.2	Proces rozpoznávania emócií	33
8.2.1	Architektúra neurónových sietí	33
8.2.2	Tréningové a testovacie dáta	34
8.2.3	Tréning a testovanie.....	34
	Záver.....	41
	Použitá literatúra	42
	Zoznam skratiek	43
	Zoznam tabuliek.....	44
	Zoznam obrázkov.....	45

ÚVOD

Reč je jednou z najrozvinutejších metód prenosu informácie medzi ľuďmi. Preto sa v súčasnosti, v dobe zvyšujúcej sa úrovne vedy a techniky, snažia vedci a technici o to, aby sa počítač stal plnohodnotným, hovoriacim partnerom človeka. Ak chceme dosiahnuť tohoto výsledku, musíme tento systém najskôr algoritmicky a technicky vyriešiť. I keď dokonalý systém ešte neexistuje, čiastočné produkty vývinu sa už dnes úspešne využívajú v priemysle. Sú to napríklad dialógové systémy automatov vyhľadávajúcich v databázach a podobne. Veľké využitie majú aj hlasové systémy na ovládanie strojov, či už pri činnosti kedy máme zamestnané ruky aj nohy, alebo pre telesne hendikepovaných. Pracuje sa aj na automatickom prekladaní hlasu z a do cudzích jazykov.

Do tohoto mora sme sa rozhodli pridať svoju kvapku aj mojou bakalárskou prácou o rozpoznávaní emočného stavu človeka z reči. Táto bude pozostávať z databázy nahrávok a vzorkov hlasu ľudí v určitých hlavných druhoch emócií a z jej analýzy vhodným softwarom pre zistenie príznakov na spoľahlivé rozpoznanie emočného stavu. V ďalšej časti sa chceme zamerať na metódy rozpoznania emočných stavov pomocou neurónových sietí k čomu budeme využívať program Stutgard Neural Network Simulator (SNNS) a najmä jeho grafickú nadstavbu Java Neural Network Simulator. So systémom tohto nástroja pre tvorbu a používanie neurónových sietí sa aj bližšie zoznámime. Pokúsime sa vytvoriť, natrénovať a otestovať niekoľko neurónových sietí.

V závere zhodnotíme našu úspešnosť v rozpoznaní emočného stavu, kde verím že dosiahneme dobrých výsledkov, použiteľných pre ďalšie bádanie a výskum.

1 EMÓCIE ZO PSYCHOLOGICKÉHO HĽADISKA

Psychológovia majú vplyv na to, ako sú emócie vnímané v spoločnosti. Do konca 19. storočia boli emócie považované za prejavy psychických, alebo fyzických anomálií. Až Darwin vo svojej knihe *The Expression of Emotions in Man and Animals* v roku 1872 pripomenul emotívne chovanie primitívneho človeka a vrátil emócie do prirodzeného prejavu. Podľa neho je emócia prirodzenou reakciou na aktuálnu situáciu. Darwinov druhý zákon o vyjadrovaní emócií hovorí, že prejav, gestá a pohyby vyjadrovania kladnej emócie, sú presne opačné ako gestá a pohyby vyjadrovania zápornej emócie. Teda ak pri nehe je charakteristický pokles napätia dychového a hrtanového svalstva, tak pri hneve je charakteristické zvýšené napätie dychového a hrtanového svalstva. Podľa Wundta(1923) sa emócie od seba líšia podľa toho akým zážitkom bol sprevádzaný ich vznik. Prijemným, nepríjemným, stresujúcim, alebo uvoľňujúcim. Krech, Crutchfield & Balachey(1962) klasifikujú emóciu ako fyziologický prejav, zatiaľ čo cit sa orientuje iba na zážitkovú formu. Emócie za vzrušenie organizmu a city za vzrušenie mysli považuje aj Nakonečný (1997). Podľa Lazarusa(1966) je emócia psychologickým prejavom na dobré a zlé správy. Všetky tieto definície emócií nás uvádzajú do terminologickej neistoty, preto si ešte raz definujeme pojmy [1].

1.1 DEFINÍCIA POJMOV

Nasleduje definícia pojmov emócia a postoj, ktoré sa významným spôsobom podieľajú na finálnej podobe prozódie reči.

1.1.1 Emócia

V emóciách sú integrované psychické komponenty. Je to vlastne komplexný jav, ktorý zahŕňa zážitok, ale aj prejavy v správaní.

Samotný výraz emócia, môže byť v psychológii inak klasifikovaný ak sa jedná o skúmanie doprovodných telesných prejavov či správania a inak pri zameraní na aktivitu nervového systému.

Všeobecne definícia emócie hovorí, že je to zmena psychologického stavu, ktorý sa spúšťa pri dosahovaní nejakého cieľa. Ak cieľ úspešne dosiahneme, hovoríme o kladnej emócií. Ak nie, hovoríme o zápornej emócií.

Termín emočný býva veľmi často spätý s výrazom citový. Inak sa emócia prejavuje veľmi krátko v mimike, hlase, alebo gestikulácii. Vedomá emócia sa môže prejavovať aj niekoľko minút.

Emócia sa prejavuje najmä tam, kde nie je zabehnutý vzorec chovania, teda tam, kde sa objavujú nečakané a neznáme situácie. Preto deti majú viac emočných stavov ako dospelí. V situáciách, ktoré sú pre ne nové, nemajú zabehnutý vzorec chovania. Naopak zvieratá majú zabehnuté inštinktívne vzorce chovania a preto používajú emócie minimálne.

Prijateľnosť emočného prejavu je spätá so štatútom, aký emócia má v danom prostredí. Podľa Tousignanta a Habimana (1993) je prístup k emóciám do značnej miery etnologickou záležitosťou. Odtiaľto pramenia aj problémy s ich definíciou a charakteristikou. V rôznych kultúrach majú emócie rôzny zmysel a obsah.

Emócia sa prejavuje v chovaní, ale aj reči. Súčasťou rečového prejavu sú aj prozodické zmeny. K nim dochádza pod vplyvom emócie ako súčasť prežívanej emócie [1].

1.1.1.1 Kategórie emócií

Základné rozdelenie emócií, je do dvoch kategórií:

- fyziologické narušenie, šok, záchvat;
- chovanie, stála forma nejakého stavu.

Toto rozdelenie klasifikuje emócie na primárne a sekundárne. Potiaľto sa psychológovia zhodujú. Problémy nastávajú pri kategorizácii konkrétnych emócií. Spomeniem iba niektoré názory na rozdelenie emócií z histórie a súčasnosti. Watson (1919) rozlišuje strach, hnev a lásku. Dumas (1923) rozdeľuje emócie na štyri skupiny: radosť, smútok, strach, hnev a ďalej ich delí na pasívne a aktívne. Woodworth (1939) rozlišuje šesť kategórií: láska, prekvapenie, strach, hnev, zhnusenie, opovrhovanie. U niektorých pripúšťa i varianty.

Plutchik v roku 1980 na základe svojej teórie definoval osem základných z ktorých môžeme ostatné odvodiť: radosť, súhlas, strach, podriadenosť, smútok, hnus, hnev a očakávanie. Historický prehľad by nám ponúkal až 21 rôznych definícií emócie a ich rozdelenia. Vo všetkých sa v základnom delení nachádza strach a hnev. Avšak do základných ktoré sa zhodujú môžeme započítať aj radosť a smútok, keďže ich v anglickej literatúre môže vyjadrovať viac synonym, napr.: joy, pleasure, enjoyment, happiness. Počet kategórií emócií, ale i samotné definície sa menia od jednej štúdie ku druhej, ba priam od jedného výskumu k druhému. Nakonečný (1997) stavia emócie do dvojrozmerného pohľadu a to príjemný - nepríjemný a poddajnosť či povýšenosť. Neberie ovšem vôbec do úvahy aktivitu fyziologických prejavov pociťovanej emócie.

Všetci autori sa však zhodujú v tom, že žiadna emócia sa nevyskytuje úplne izolovaná, ale vždy ju tvorí viacero emócií [1].

1.1.1.2 Fyziologické zmeny pri emočnom podráždení

Skoro každú emóciu sprevádzajú globálne fyziologické zmeny ku ktorým patrí napríklad zmena tepu srdca či krvného tlaku, žalúdočné sťahy, bledosť, alebo červenanie. Rôznorodosť emócií a fyziologických prejavov ľudského tela vylučuje rozpoznávanie emócií na základe fyziologických reakcií.

Radosť je reakcia na úspech, potešenie z dosiahnutého dobra. Pri radosti sa u človeka prejavuje túžba podeliť sa s ďalšou osobou. Tu prichádza na rad reakcia okolia, ktorá môže byť kladná, ale i záporná. Najmä preto má radosť veľmi silný sociálny význam. Radosť sa vo fyziologických prejavoch vyznačuje zrýchlením tepu, zvýšením krvného tlaku a teploty.

Hnev je podobne ako zlosť či rozčúlenie reakcia na nejakú prekážku. Spoločne so strachom sa u oboch prejavuje zvýšenie fyzickej sily. Fyziologicky sa prejavuje zvýšeným svalovým napätím, zvyšuje sa frekvencia dychu, tlkot srdca a krvný tlak.

Smútok je reakciou na stratu. Intenzita smútku záleží najmä na význame strateného. Je to emócia pri ktorej subjekt stráca záujem. Fyziologické zmeny sú úzko späté so zmenami pri bolesti a to je spomalenie srdčej činnosti, výkyvy v tempe dychu a prejavovanie potreby sústrasti.

Strach je poslednou emóciou ktorá patrí do emócií primárnych. Je reakciou na hrozbu. Od strachu je však potrebné odlíšiť úzkosť, ktorá nemá príčinu. Fyziologicky je strach spojený s rastom frekvencie nervových impulzov. Ďalej podľa zmeny aktivácie s rastom, alebo poklesom tepu a dychu. Ich nepravidosťou či zmenou svalového napätia [1].

1.1.1.3 Analýzy emočného chovania

Do tejto kapitoly spadá najmä analýza rečového prejavu, ktorá spadá do niekoľkých modelov. Prvý z nich Trägerov hovorí o troch spôsoboch ako sa emócie v reči prejavujú:

- v emotívnych prejavoch sú viditeľné vlastnosti normálneho, neutrálneho hlasového prejavu hovoriaceho, ktorý je charakteristický pre jeho vek, pohlavie, zdravie či psychické rozpoloženie;
- emócie sa premietajú do kvality hlasu prostredníctvom prozodických parametrov (výška, tempo, intenzita) a farbou hlasu;
- konečné emócie sa prejavujú rôznymi zvukmi nejazykového charakteru.

Tento model bol však neskôr doplnený Davitzovým, ktorý hodnotil akustické prejavy emócií. Veľkou nevýhodou tohoto modelu je nepresnosť kategórií. Mnohé rozdielne emócie ako napríklad smútok a nuda sú v jednej kategórii a niektoré variácie sú zase v rozdielnych kategóriách [1].

Podľa Scherera môžeme emócie roztriediť do štyroch skupín.

- Aktívne príjemné emócie sú charakterizované vyššou globálnou úrovňou hlasu, väčšou intenzitou a vyšším tempom.
- Pasívne príjemné emócie sa prejavujú relatívne nízkym hlasovým registrom, pomalším tempom, pravidelným rytmom a prevažujúcou tendenciou k melodickému stúpaniu. Artikulácia je neistá.
- Aktívne nepríjemné emócie sú v hlasovom prejave rozpoznateľné vďaka vyššej priemernej hodnote základného tónu a vysokej intenzite. Rytmus a intonačný priebeh sú nepravidelné.
- Pasívne nepríjemné emócie sa vyznačujú prozodickými charakteristikami blízkymi prejavom pasívnych príjemných emócií v oblasti úrovne hlasového registra a tempa. Ich intonačné priebehy sú ovšem nepravidelné [1].

1.1.2 Postoj

Postoj je všeobecne charakterizovaný ako schopnosť reagovať kladne, alebo záporne voči spoločenským hodnotám. Je vždy na niečo zameraný. Subjekt svojim postojom vyjadruje svoj názor na niečo. Vždy má nejaký cieľ, alebo účel. Postoj môže byť pozitívny, alebo negatívny a to s rôznymi stupňami intenzity. Stretneme sa aj s definíciou postoja, ktorá o ňom hovorí, ako o označení stavu psychologickéj použiteľnosti, na základe skúseností s dynamickým vplyvom na správanie.

V inom poňatí postoj vyjadruje hodnotiace kritérium získané výchovou, alebo vlastnými skúsenosťami. V tomto prípade umožňuje vyhodnocovať udalosti a zvoliť vhodné chovanie [1].

2 REČ

Už v polovici 18. storočia sa prejavovali pokusy uplatniť rečový signál na komunikáciu so strojmi. Pre človeka je ovládanie stroja hlasom veľmi zaujímavým podnetom pre bádanie a výskum. Od tej doby prešlo veľa času a teda človek výskúmal mnoho spôsobov a metód používaných na analýzu ľudského hlasu. Avšak asi najväčší pokrok prišiel s nástupom číslicových počítačov, ktoré umožňujú digitalizáciu akustického rečového signálu. Príkladom metódy spracovania signálov je Furierova analýza. Táto bola známa už v prvej polovici 18. storočia, avšak používať sa začala až s nástupom prvých počítačov v druhej polovici päťdesiatych rokov. Jej využitie v reálnom čase prišlo s nástupom signálových procesorov v prvej polovici deväťdesiatych rokov.

V súčasnosti sa už pracuje na dialógových systémoch, ktoré dokážu rozumieť reálnej prirodzenej reči v určitom tematickom rozsahu a tak isto vďaka generátoru odpovedí viesť plnohodnotný dialóg [2].

2.1 VLASTNOSTI REČI

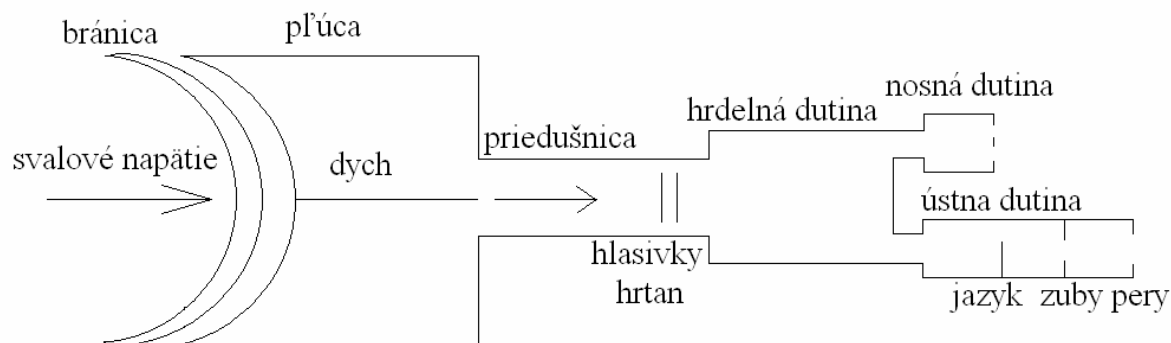
Jedným z dôležitých pojmov pri reči je jazyk. Tento pojem vyjadruje buď rečovú schopnosť dorozumieť sa, alebo schopnosť písomnú pomocou abecedy, alfabety a tak ďalej. Teda jazyk má dve formy: rečovú a písomnú. Obe z týchto foriem môžeme pokladať za rovnocenné a navzájom sa dopĺňujú. Je vlastne veľmi zložitým spôsobom komunikácie, ktorým disponujú iba ľudia.

Prvá forma jazyku a komunikácie medzi ľuďmi bola hovorená reč. Tá sa prenáša po komunikačnom kanále vo forme akustického signálu. Ten smeruje od úst hovoriaceho až k ušiam poslucháča. V tomto akustickom signále reči sa prenáša viac druhov informácie. Okrem samotného lingvistického významu (významu myšlienky) aj niektoré informácie o hovoriacom, ako charakter hlasového traktu, spôsob artikulácie, rečové vady a taktiež informácie o emočnom stave (stres, radosť, hnev...) hovoriaceho.

Písaná forma jazyku už nie je tak samozrejmá. Vyžaduje najmä zvládnutie procesu čítania a písania, ktorý sa zdá byť samozrejmý, avšak mnohým ľuďom robí ešte aj dnes problémy. Písanie vyžaduje sústredenosť a preto sa pri ňom nedá vykonávať aj iná činnosť ako pri hovorenej reči. Táto forma prenáša najmä informáciu lingvistickú. Ostatné informácie len veľmi obmedzene [2].

2.2 TVORENIE REČI ČLOVEKOM

Pre tvorenie reči má človek v tele skupinu orgánov ktoré svojou dokonalou spoluprácou reč vytvárajú. Tvorba reči však je iba ich vedľajšia činnosť a v tele zaujímajú úplne inú funkciu. Ich primárnou funkciou môže byť napríklad čuch, dýchanie, prijímanie potravy. Spája ich teda iba spoločný názov hlasový trakt. Hlasový trakt (obr.2.1), ktorý u dospelého jedinca môže dosiahnuť dĺžku 17 cm, môžeme rozdeliť do troch základných skupín: dychovej, hlasovej a artikuláčnej [2].



Obr. 2.1: Hlasový trakt.

2.2.1 Dych

Dych v ľudskom tele zabezpečuje dýchacie ústrojenstvo. To tvoria pľúca s okolitými svalmi (bránica) a vývod pľúc, ktorý je realizovaný priedušnicou. Objem ľudských pľúc je asi 4-5 litrov, avšak pri dýchaní musí v pľúcach zostať časť vzduchu čo tvoria asi 1-2 litre. Jedná sa o takzvanú zbytkovú kapacitu pľúc. Pri normálnom dýchaní človek obmieňa iba asi 0,5 litra vzduchu každých 3-5 sekúnd. Každým nádychom vytvárame energiu pre vznik hlasu (reči). Ten vzniká pri výdychu. Rýchlosť výdychu je asi 0,2 litra za sekundu a priamo ovplyvňuje silu hlasu, čiastočne i výšku hlasu. Dĺžka výdychu udáva akú dlhú nepretržitú reč sme schopní povedať. Opätovný nádych predstavuje v reči krátku pauzu. Tým doplníme do pľúc vzduch a teda materiál pre tvorbu ďalšej reči [2].

2.2.2 Hlas

Samotný hlas, respektíve základný tón hlasu vzniká vďaka hlasivkám v hrtane človeka. Hrtan je s pľúcami spojený priedušnicou. Hlasivky, dve ostré slizničné riasy, vedú naprieč hrtanom v jeho najužšom mieste. Zvyčajne majú dĺžku 15 mm u mužov a 13 mm u žien. Tvorí ich hlasový väz a hlasivkový sval. Sú na jednej strane upevnené na hlasivkové chrupavky a na strane druhej na štítnu chrupavku. V kľudovej polohe (pri nevydávaní hlasu) tvoria štrbinu v tvare trojuholníka, ktorá je asi 8 mm široká. (Obr. 2.2a) Pri vydávaní základného tónu pre reč sa napnú a štrbina sa úplne zatvorí. (Obr. 2.2b) Pri vydávaní hlasu, hlasivky vlastne zablokujú cestu vydychovanému vzduchu, čo ich začne rozkmitávať a čiastočne prudko otvárať a zatvárať. Pritom sa však najviac otvoria na plochu 20mm^2 u muža a 14mm^2 u ženy. Týmto otváraním hlasiviek vznikajú po sebe idúce časti redšieho a hustejšieho vzduchu, ktorý tvorí základný tón. Frekvencia otvárania hlasiviek sa nazýva základná frekvencia hlasu F_0 a jej prevrátená hodnota je perióda základného hlasového tónu T_0 .

Pri zmene emočného stavu človeka sa dĺžka periódy a amplitúda mierne mení i pri veľmi krátkych impulzoch. Tento jav zmeny periódy sa nazýva jitter a býva udávaný v percentách. Jav zmeny amplitudy sa nazýva shimmer a býva udávaný v dB. Aby tieto výkyvy spôsobené emóciami boli v reči vôbec počuť, musí byť jitter minimálne 2% a shimmer 1dB.

Dve základné polohy hlasivek, ktoré som spomínal rozdeľujú tvorbu reči na znelú a neznelú. Pri tvorbe znelých reči, čo sú samohlásky a niektoré znelé spoluhlásky, sú hlasivky v napnutej polohe a tvoria reč svojim kmitaním. Avšak pri tvorbe neznelých reči sú hlasivky v klude a reč je tvorená iba z vydychovaného vzduchu, artikuláciou v dutine ústnej. Neznelá reč nemá základný tón a teda ani základnú frekvenciu a periódu [2].



Obr 2.2: a) Hlasivky v kľude, b) napäté hlasivky.

2.2.3 Artikulácia

Artikulácia je tvorená takzvaným artikulačným ústrojenstvom. Toto ústrojenstvo umožňuje človeku vytváranie takmer všetkých zvukov, ktoré v reči používa. Je tvorené z nadhrtanových dutín a to sú nosná a ústna dutina, v ktorých sa nachádzajú artikulačné orgány. Medzi týmito dutinami sa nachádza lalok, ktorý je časťou mäkkého podnebia a zabraňuje či povoľuje prísun vzduchu do nosnej dutiny.

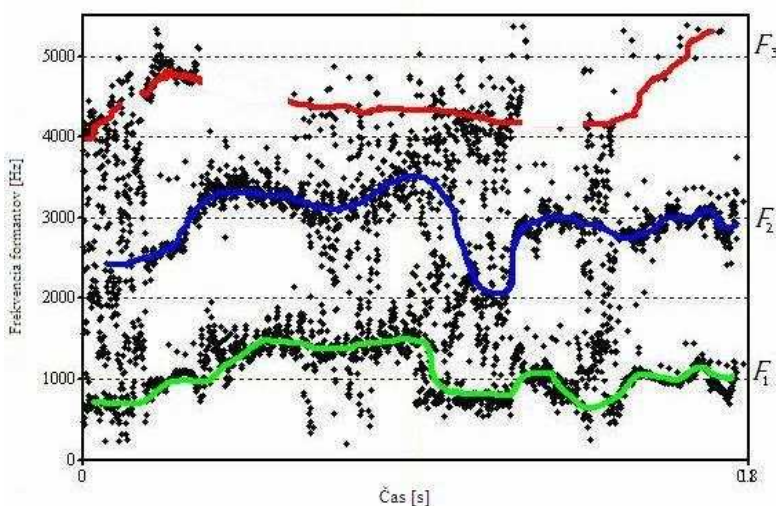
Dutiny sa na tvorbe zvuku podieľajú pasívne. To preto, že nie sú pohyblivé. Medzi aktívne artikulačné orgány patrí najmä jazyk, pery, mäkké a tvrdé podnebie, zuby, ktoré svojim pohybom menia veľkosť dutín. Posledným tvorcom zvuku je aj hrtan ktorý je pohyblivý a svojim pohybom môže meniť celkovú dĺžku hlasového traktu.

Unikanie vydychovaného vzduchu nadhrtanovými dutinami výrazne prispieva k zmenám zvuku. A to dvoma spôsobmi:

1. Vytvorenie tónovej štruktúry. Prechodom vzduchu dutinami vplyvom rezonancie sa mení rozloženie akustickej energie v signále. Energia sa sústreďuje okolo niektorých frekvencií, ktorým hovoríme formantové. Oblasť v ktorej sa táto akustická energia nahromadila, nazývame formantom. (Obr. 2.3) Ten označujeme postupne od najnižšej frekvencie F_1 , F_2 , $F_3, \dots, F_n$. Pri zapojení nosnej dutiny do procesu tvorenia reči vplývajú na hlas aj jej antirezonančné vlastnosti, ktoré spôsobujú potlačenie niektorých frekvencií a teda vznik anti-formantov, ktoré označujeme $A_1, A_2, \dots, A_m$.

2. Vytvorenie šumovej zložky. Všetky artikulačné orgány umiestnené v nadhrtanových dutinách svojimi pohybmi ovplyvňujú tvar hlasového traktu. Môžu ho až úplne zatarasiť, alebo úplne otvoriť. Pri prechode vzduchu touto časťou traktu vznikajú takzvané šumové zložky, ktorým pri opise jazyka hovoríme súhlásky. Rôzne druhy šumu vznikajú rôzne nastavenými prekážkami.

Treba ešte podotknúť, že hlasový trakt nieje diskretný systém a teda zmeny v ňom sú plynulé. To ovplyvňuje výsledný zvuk najmä prechody medzi vyslovovanými hláskami [2].



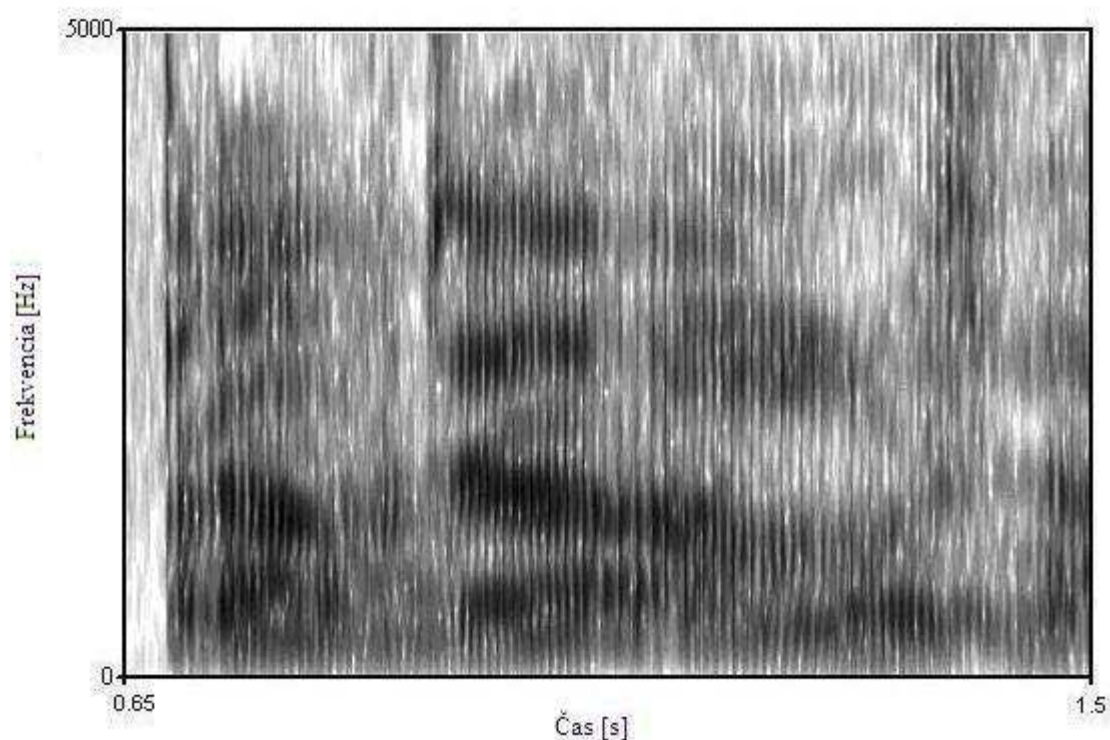
Obr. 2.3: Zobrazenie prvých troch formantov v slove "pravda" programom Praat.

2.3 REPREZENTÁCIA REČI

Reč je hovorenou časťou jazyka. Všeobecnou štúdiou jazyka sa zaoberá lingvistika. Tá reč rozdeľuje do lingvistických hľadísk: akustického, artikulačného, fonetického, fonologického, morfológického, syntaktického, sémantického a pragmatického. V tejto časti sa budeme venovať najmä akustickému hľadisku. Pre bližšie naštudovanie ostatných hľadísk môžeme doporučiť literatúru [2] str. 36.

Z akustického hľadiska je reč vlnenie elastického prostredia v šírke počuteľných frekvencií. Pre snímanie reči sa používa mikrofón, ktorý je vlastne menič a teda prevádza akustický signál na signál elektrický. Ten sa ďalej spracováva rôznymi spôsobmi, napríklad digitalizáciou. Hlavnými akustickými črtami reči sú: nosná frekvencia F_0 , intenzita a rozdelenie spektrálnej energie [2].

Na zobrazenie časovej a spektrálnej vlastnosti reči sa používa spektrogram. (Obr. 2.4) Zachytáva časovo-frekvenčnú reprezentáciu reči pomocou rôznej úrovne šedej v dvojrozmernom grafe. Spektrum teda obsahuje kombináciu frekvencií, z ktorých je daný signál vytvorený. Za základnú frekvenciu sa považuje najnižšia frekvencia [3].



Obr. 2.4 Spektrogram slova "pravda" vytvorený programom Praat.

2.4 VARIANTY REČI

Okrem klasickej reči, sa veľmi často v živote stretávame aj s jej rozdielnymi variantami. Sú to najmä šepot, krik a spev.

2.4.1 Šepot

Je to veľmi zvláštna varianta reči. Pri šepkaní je hlasivkový zdroj kmitov nahradený trením vzduchu v hlasivkách. Tým úplne zanikajú všetky znelé tóny a hlásky. Sú nahradené neznelými. Klesá aj hladina intenzity. Z bežných 40-60dB klesá o 20-40dB na 20-30dB. Šepot teda používame najmä ak nechceme aby našu reč bolo počuť intenzívne (nechceme rušiť alebo osloviť širšie okolie) [2].

2.4.2 Krik

Je ďalšou extrémnou variantou reči. Intenzita u neho stúpa z bežných 40-60dB až na 80dB. V kričanej reči sa zvýrazňujú samohlásky a to nie len intenzitou ale aj dĺžkou trvania. Tým že sa pri kričanej reči viac otvára hlasový trakt dochádza k zmene prvého formantu a tak isto k zmenám v spektre. Pri kričaní dochádza k zmene napätia hlasiviek a k zvýšeniu tlaku vydychovaného vzduchu. Tým sa skracuje doba otvárania hlasiviek. Tvar hlasivkového signálu je potom skôr štvorcový ako trojuholníkový u normálnej reči. Na vyšších frekvenciách tak prichádza k nárastu energie a teda sa zvyšuje aj nosná frekvencia F_0 .

Pri hlasových extrémoch sa vždy znižuje zrozumiteľnosť reči. U šepotu je to odstránením znej zložky. Pri kriku je to zasa zvyšovanie intenzity, na úkor zlej artikulácie [2].

2.4.3 Spev

Pri speve sa mení najmä trvanie rečových jednotiek a to tak aby sa dosiahlo určitého rytmu. Pri tom sa väčšinou mení dĺžka samohlások. Je to podobné ako u kriku, avšak nemení sa veľkosť nosnej frekvencie. Pridáva sa množstvo energie do 3. a 4. formantu, kde sa jej pri bežnej reči toľko nedostáva. Pri speve prichádza aj k zmene artikulácie samohlások, na základe čoho vzniká v okolí frekvencie 2,8 kHz tzv. spevný formant. Rozsah ľudského hlasu prechádza približne do dvoch oktáv. V bežnej reči je rozsah hlasu iba približne v jednej oktáve [2].

3 PROZÓDIA

Skôr ako začnem opisovať prozódriu ako takú, alebo prozodický systém je treba vysvetliť niektoré pojmy, ktoré s prozódriou súvisia.

Pojmy prozodický a suprasegmentálny sú synonymá a vyjadrujú zmeny základného tónu, intenzity a trvania v organizácii reči.

Pojem intonácia v užšom význame môže vyjadrovať melódiu (zmenu výšky hlasu). Vo význame širšom zahrňuje okrem výšky aj dynamický priebeh v časovom členení. V tomto širšom význame sa dá povedať, že sa jedná o synonymum slova prozódia.

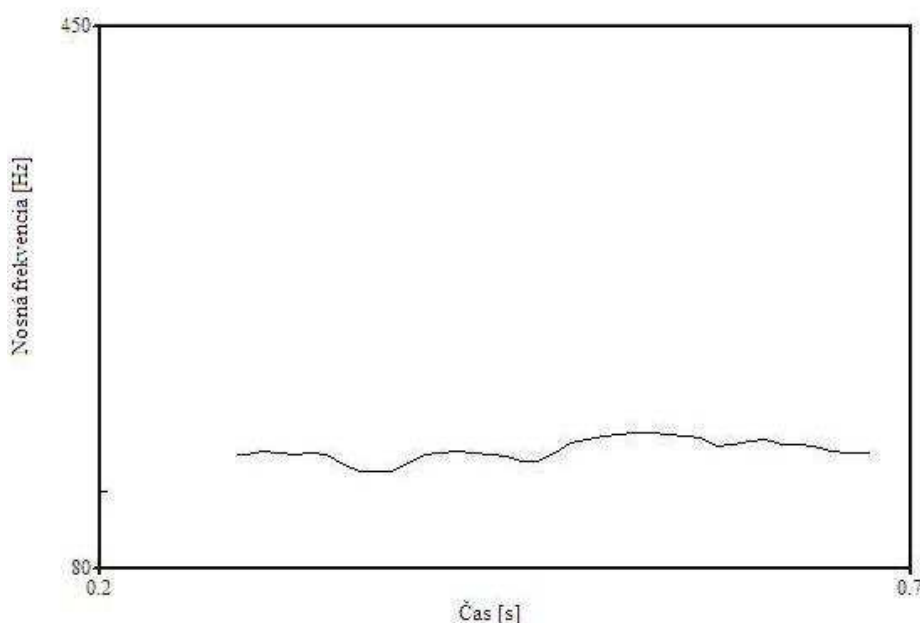
Do prozódie patrí aj to, akým spôsobom sa jednotlivé orgány konkrétne podieľajú na tvorení reči a akým spôsobom ovplyvňujú priebeh jednotlivých parametrov [1].

3.1 PROZODICKÉ PARAMETRE

Medzi základné parametre prozódie reči patria základná frekvencia, intenzita a trvanie. Základná frekvencia, meraná v Hertzoch, vyjadruje počet kmitov hlasiviek za sekundu. Jej zmeny vplývajú na zmeny výšky hlasu. Intenzita je sila hlasu vyjadrená v decibeloch. Trvanie, merané v sekundách, sa spolu s tempom podieľajú na rytme reči [1].

3.1.1 Intonácia

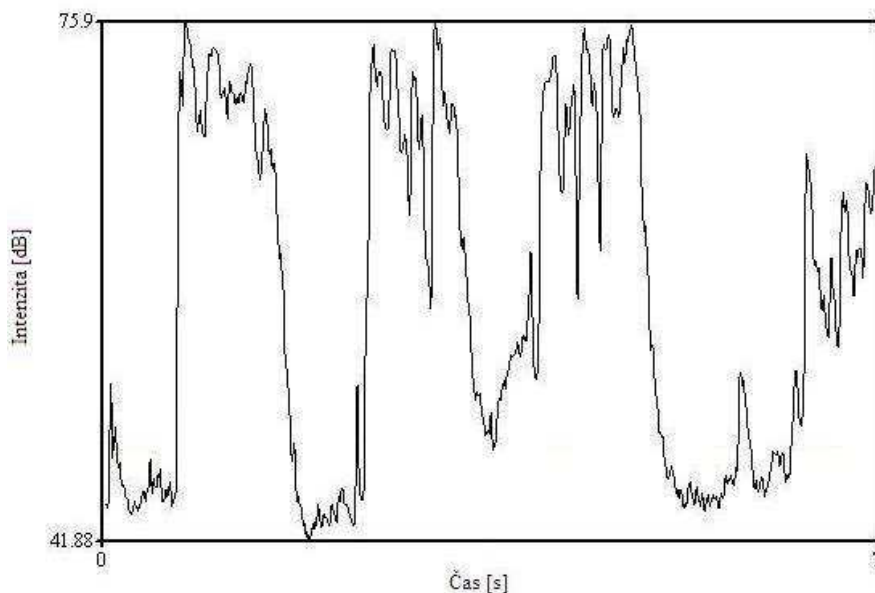
Tvorenie znelych tónov, ktoré vznikajú rezonanciou hlasiviek a tvorením základného tónu sa v reči nazýva melódiou. Priebeh a zmeny frekvencie základného tónu sa vo väčšine programov na analýzu reči dajú zistiť vďaka pitch detektoru. Základná frekvencia je pre človeka vnímaná ako výška hlasu. Základný tón je u človeka tvorený počtom hlasivkových kmitov za sekundu. Pohybuje sa priemerne v rozmedzí od 80 do 450 Hz. Jeho hodnoty sú závislé od veku a citového stavu hovoriaceho [1].



Obr. 3.1: Nosná frekvencia v bežnej reči, zobrazená programom Praat.

3.1.2 Intenzita

Je vnímaná ako sila hlasu, spojená s hladinou subglobálneho tlaku, ktorý je závislý od citového a emocionálneho stavu človeka. Rovnako ako zmeny v dýchaní, srdečnom tepe a krvnom tlaku sú fyziologické zmeny s tým súvisiace. Intenzita však počas nahrávania môže byť ovplyvnená zmenami nastavenia mikrofónov či zmenou vzdialenosti hovoriaceho. Preto analýzy pracujú iba s relatívnymi hodnotami v krátkych časových reláciách [1].



Obr. 3.2: Intenzita bežnej reči, zobrazená programom Praat.

3.1.3 Trvanie

V češtine treba pri tomto parametri spomenúť najmä fonologický rozdiel medzi krátkymi a dlhými vokálmi. Celok časového trvania slabík a páуз sa nazýva tempo reči.

3.2 FUNKCIE PROZÓDIE

Základné rozdelenie funkcií prozódie je na skupinu jazykovú a nejazykovú. Mimojazykové funkcie zabezpečujú informácie o skutočnostiach, ktoré hovoriaci poskytuje vedome, napríklad názory, postoje. Avšak aj informácie, ktoré sú v reči obsiahnuté automaticky. Sú to napríklad informácie o pohlaví, veku, či psychickom a fyzickom stave hovoriaceho [1].

3.2.1 Jazykové funkcie prozódie

Rozlíšenie deklaratívnej a interogatívnej vety zaisťuje sémantická funkcia. Podľa Fónagyho (1979) intonácia je mnohokrát jediným spôsobom odlišovania modalít, preto sa aj nazýva modálnou funkciou. Intonácia sa delí do štyroch hlavných intonačných schém:

1. Oznamovaciu vetu sprevádza klesavá melódia a stredný hlasový register.
2. Pre rozkazovaciu vetu je charakteristický priamy pokles finálnej melódie.
3. Zisťovacia otázka sa vyznačuje stúpaním melódie až do vyšších polôh rečového rozsahu.
4. Pri doplňovacej otázke klesavá melódia spája vyššiu polohu hlasového rozsahu s nižšou.

Mnoho zmien prozodických zložiek sprevádzajú modifikácie intelektuálneho významu. Zvýšenie nosnej frekvencie F_0 spojené s poklesom intenzity má podobnú funkciu ako ukončenie stúpavej melódie pri zisťovacích otázkach a teda vedie k zmene hovoriaceho. Súčasný pokles nosnej F_0 a intenzity na konci oznamovacej vety, ponúka prevzatie slova,

alebo ukončenie témy, monológu. Pokles nosnej F_0 a súčasný nárast intenzity naznačuje nedobrovoľné prerušenie, alebo váhanie. Samotné zvýšenie F_0 a pretrvanie v tomto stave sa zobrazuje na konci zvolacích viet.

Na úrovni skladby vety je úloha prozódie skôr organizačná. Prerozdeľuje reč bez vnútornej štruktúry na drobnejšie celky a prispieva k organizácii hovorenej reči [1].

3.2.2 Fonoštylistické funkcie prozódie

Fonoštylistické funkcie sú funkcie nejazykové. León do tejto kategórie pridáva ďalšie dve funkcie. 1. Identifikačnú funkciu ktorá charakterizuje hovoriaceho i bez jeho vedomia. Odvodzuje vlastnosti hovoriaceho ako napríklad jeho vek, pohlavie, alebo sociálny a regionálny pôvod. 2. Impresívnu funkciu ktorá vypovedá o zámere dať reči určitý štýl. Zaisťuje vyjadrenie postoja hovoriaceho.

Emotívne a expresné funkcie prozódie informujú adresáta hovorenej reči o postoji a citovom stave hovoriaceho.

Citový stav hovoriaceho je jedna z informácií, ktorá sa dostáva do každej hovorenej reči bez ohľadu na to či si toto hovoriaci uvedomuje.

Zámer, latinsky intencia je pre vyjadrovanie postojov relatívny. Avšak pri vyjadrovaní emócií je spontánny [1].

3.3 PROZÓDIA ČEŠTINY

Čeština je jazyk, kde veta predstavuje jasnú melodickú a ucelenú jednotku. Podľa Palkovej (1994) je hlavným prvkom štrukturujúcim vetu melódia. Avšak rovnako aj intenzita a časové členenie prispievajú k organizácii zvukovej jednotky. Štúdiá Romportlovej (1951) a Danešovej (1957) sa sústreďujú na rozoznanie všetkých druhov organizácií a sú založené na sluchovej analýze [1].

3.3.1 Prízvuk češtiny

Čeština má prízvuk viazaný na prvú slabiku rytmického celku. V niektorých nárečiach sa môže prízvuk objaviť na slabike predposlednej. To majú však väčšinou za následok iné jazyky v pohraničných oblastiach. V češtine prevažuje dynamický charakter prízvuku pretože na jeho tvorení sa podieľa najmä intenzita. Prízvuk tu neodlišuje význam slov, ako napríklad v ruštine. Intuitívny prízvuk je vytváraný nie len intenzitou, ale aj zmenou výšky a trvania.

V priaznivých podmienkach výslovnosti je vyššia intenzita a základná frekvencia. Tempo je však spomalené. I v nepriaznivých podmienkach kde nie je možné zvýšiť intenzitu a základnú frekvenciu sa ukázalo ako dostatočne významné spomalenie tempa a vyvolanie dojmu prízvučnej slabiky.

Jadro hovorenej reči býva zvyčajne umiestnené až v závere našej výpovede. S týmto jadrom je spojené aj umiestnenie prízvuku vo vete. Vetný prízvuk sa realizuje nízkym tónom ľahko rastúcim ku koncu prízvučnej slabiky [1].

3.3.2 Tempo češtiny

Tempo sa veľmi podieľa na vnímaní rytmu hovorenej reči. Zmeny tempa predstavujú zvukovú zmenu ktorá v češtine vyjadruje citové zaujatie hovoriaceho. Každý hovoriaci človek má svoju základnú rýchlosť reči, teda tempo. V súdobej češtine sa veľmi rozšírilo finálne predlžovanie. Napríklad ak sa v poslednej slabike nachádza dlhá samohláska, môže táto slabika dosiahnuť až 370% trvania priemerného krátkeho vokálu. Trvanie krátkej samohlásky môže dosiahnuť až 145% priemeru [1].

4 VYTVORENIE DATABÁZY EMOČNÝCH STAVOV

V mojej práci, asi najväčší problém tvorilo vytvorenie vhodnej prirodzenej databázy. Keďže zohnať väčšie množstvo nahrávok prirodzených emočných stavov je pravdepodobne nemožné, pri najmenšom veľmi časovo náročné rozhodol som sa pre alternatívny zdroj.

4.1 ZDROJ DATABÁZY

Zdrojom databázy nahrávok sa stali televízne seriály českej produkcie v ktorých sa nachádza veľmi veľa zápletiiek, problémov a intríg, vďaka ktorým sa v nich emočné stavy nachádzajú veľmi často. Za tieto seriály považujem napríklad seriály televízie Nova a Prima.

4.2 SPÔSOB TVORBY

Pri tvorbe sme sa v prvom rade potrebovali dostať k nahrávkam hore menovaných seriálov vo wav formáte. Tak sme začali postupne nahrávať Ordináciu v ružovej záhrade, Comeback a Ulicu z televízie Nova a Veľmi krehké vzťahy z televízie Prima. Získali sme už spomínané nahrávky po komprimácii s koncovkou .wav. Ďalším krokom tvorenia databázy bola úloha počúvať seriály a zapisovať trvania emočných prejavov do tabuľky. Ukážkou tabuľky je Tab. 4.1. Po vypočutí všetkých seriálov sme pomocou tabuľky a funkcie napísanej v programe Matlab nastihali všetky zapísané emočné prejavy.

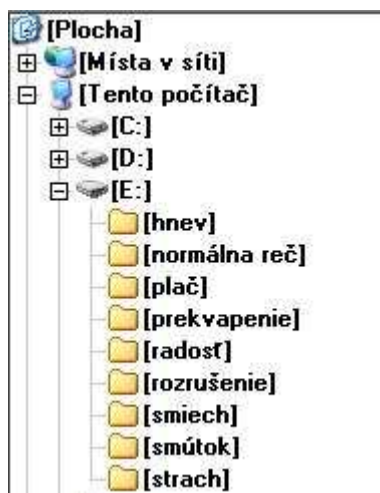
		od			do		
Seriál	skratka	hod.	min.	sek.	hod.	min.	sek.
ordinacia_comeback_2_10	s	0	1	54	0	1	57
ordinacia_comeback_2_10	s	0	3	0	0	3	20
ordinacia_comeback_2_10	p	0	3	25	0	3	32
ordinacia_comeback_2_10	r	0	7	13	0	7	19
ordinacia_comeback_2_10	r	0	11	17	0	11	25
ordinacia_comeback_2_10	v	0	9	52	0	9	56
ordinacia_comeback_2_10	v	0	47	35	0	47	37
ordinacia_comeback_2_10	v	1	3	40	1	3	44
ordinacia_comeback_2_10	v	0	15	46	0	15	52
ordinacia_comeback_2_10	k	0	14	58	0	15	1
ordinacia_comeback_2_10	k	1	21	27	1	21	35
ordinacia_comeback_2_10	k	0	38	31	0	38	35
ordinacia_comeback_2_10	b	0	37	50	0	37	55
ulice_2_10	m	0	2	36	0	2	41
ulice_2_10	v	0	0	5	0	0	16
ulice_2_10	s	0	0	20	0	0	30
ulice_2_10	v	0	9	37	0	9	47

emócie:	skratka
smútok	s
plač	p
smiech	m
hnev	v
radosť	r
prekvapenie	k
bežná reč	b
rozrušenie	rz
radosť	d

Tab. 4.1: Ukážka časového delenia emočných stavov v seriáli.

4.3 KONEČNÁ FORMA

Po nastrihaní zostávalo len uložiť nahrávky do istého sledu aby vytvárali tzv. databázu emočných stavov. Nahrávky som rozdelil do priečinkov podľa názvu emócie a jednotlivé vzorky pomenoval podľa poradí od 1 po n. (Obr. 4.1.)



Obr. 4.1: Ukážka databázy uloženej na disku.

5 ANALÝZA

Ďalej sme analyzovali vytvorenú databázu, za účelom získania príznakov pre neskoršie rozpoznávanie emočného stavu hovoriaceho. Za týmto účelom som pripravil tabuľku sledovaných príznakov. (Tab. 5.1)

Veličina	Sledovaná hodnota	Skratka	Jednotka
Základná frekvencia	minimálna hodnota	$F_{\min}$	Hz
	maximálna hodnota	$F_{\max}$	Hz
	stredná hodnota	F_{str}	Hz
	smerodatná odchýlka	F_{std}	Hz
1. formant	stredná frekvencia	F_1	Hz
	šírka pásma	F_{b1}	Hz
2. formant	stredná frekvencia	F_2	Hz
	šírka pásma	F_{b2}	Hz
3. formant	stredná frekvencia	F_3	Hz
	šírka pásma	F_{b3}	Hz
Intenzita	minimálna hodnota	$I_{\min}$	dB
	maximálna hodnota	$I_{\max}$	dB
	stredná hodnota	I_{str}	dB
	smerodatná odchýlka	I_{std}	dB
Harmonicita	minimálna hodnota	$H_{\min}$	dB
	maximálna hodnota	$H_{\max}$	dB
	stredná hodnota	H_{str}	dB
	smerodatná odchýlka	H_{std}	dB

Tab. 5.1: Tabuľka získavaných príznakov a ich jednotiek.

5.1 POUŽITÝ SOFTWARE

Na analyzovanie databázy som použil holandský freeware nazvaný Praat. Tento program bol vyvinutý na Amsterdamskej univerzite humanitných vied. Program je voľne prístupný na webovej stránke www.praat.org. Na tejto stránke je dostupný pre viacero platfôrm ako napríklad Windows, Unix, Macintosh a pravidelne aktualizovaný. Pracuje prevažne s mono signálom. Nie je užívateľsky náročný. Umožňuje aj písanie skriptov, čo som využil pri analyzovaní mojej databázy. Veľmi podrobný návod na používanie programu, tak ako aj písanie skriptov je v literatúre [3] od strany 24.

5.2 VÝSLEDKY ANALÝZY

5.2.1 Tabuľka výsledkov

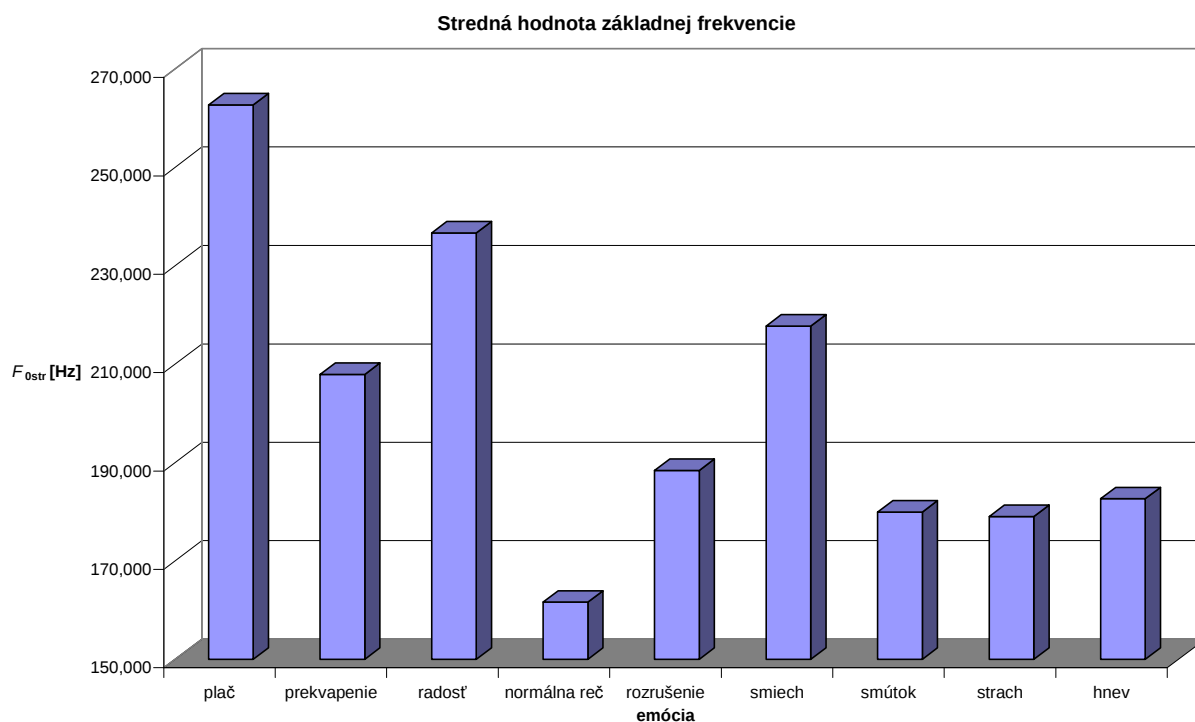
Výsledky analýzy z programu Praat som importoval do tabuľky . Dopočítal som priemerné hodnoty príznakov pre jednotlivé veličiny a vyniesol ich do tabuľky (Tab. 5.2).

príznak	jednotka	plač	prekvapenie	radosť	normálna reč	rozrušenie	smiech	smútok	strach	hnev
F_{0min}	[Hz]	81,546	89,809	97,753	85,846	79,867	88,772	78,069	81,398	84,970
F_{0max}	[Hz]	538,077	456,733	509,365	384,941	512,510	495,514	441,389	586,399	434,739
F_{0str}	[Hz]	262,644	207,908	236,638	161,590	188,353	217,688	179,897	178,978	182,610
F_{0std}	[Hz]	109,746	88,346	81,965	59,942	74,739	97,000	75,862	77,817	67,502
F_1	[Hz]	750,277	770,262	749,785	746,038	761,042	766,059	783,288	707,203	770,799
F_{b1}	[Hz]	340,919	364,102	316,612	409,136	373,013	334,596	430,111	318,900	392,551
F_2	[Hz]	1799,206	1827,626	1829,525	1805,606	1818,819	1816,781	1863,189	1789,137	1850,797
F_{b2}	[Hz]	437,989	476,768	389,296	507,348	460,314	413,064	494,523	486,331	469,756
F_3	[Hz]	2866,816	2904,316	2894,129	2907,249	2904,352	2904,224	2938,370	2895,701	2946,076
F_{b3}	[Hz]	491,576	556,261	479,016	548,558	515,851	509,892	552,056	528,664	517,525
I_{min}	[dB]	48,581	44,739	47,860	42,468	46,151	46,420	44,560	41,935	47,010
I_{max}	[dB]	81,718	78,776	80,090	77,191	80,450	80,204	79,024	80,275	79,997
I_{str}	[dB]	73,364	68,367	71,364	67,314	70,270	71,134	68,843	70,874	71,522
I_{std}	[dB]	7,867	8,570	7,855	9,183	8,304	8,695	9,143	10,737	8,450
H_{min}	[dB]	-226,378	-226,304	-226,518	-226,173	-226,683	-224,686	-226,291	-226,608	-220,704
H_{max}	[dB]	37,195	35,400	39,059	36,373	40,483	35,253	36,347	38,435	35,675
H_{str}	[dB]	6,248	5,964	7,424	5,928	7,408	5,670	5,646	6,736	5,482
H_{std}	[dB]	6,544	5,703	6,367	5,263	5,759	5,806	5,641	6,401	5,627

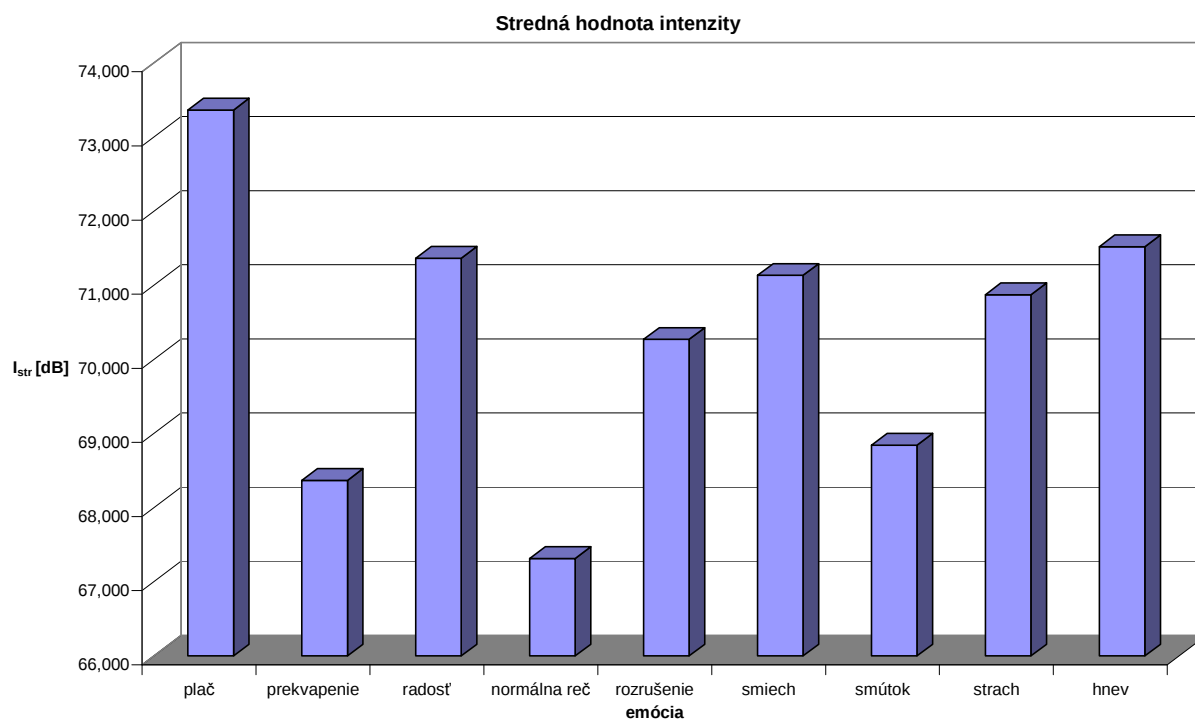
Tab. 5.2: Priemerné hodnoty príznakov pre jednotlivé emócie.

5.2.2 Grafické znázornenie výsledkov

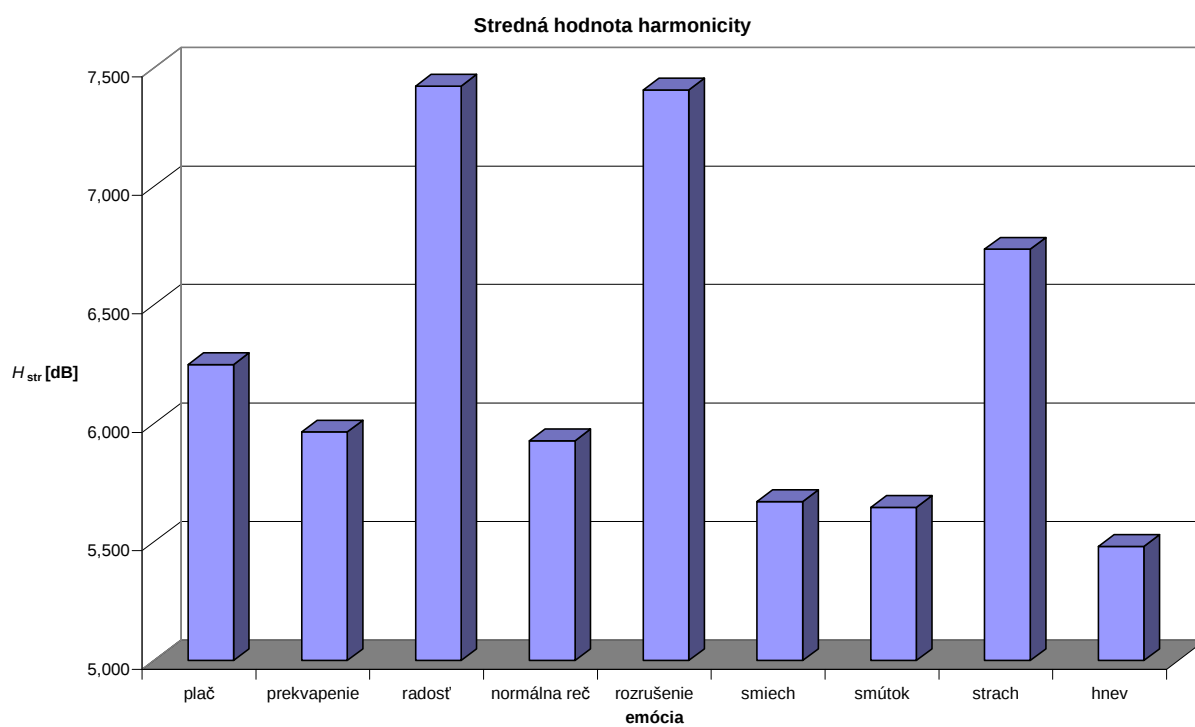
Výsledné hodnoty z tabuľky som pre lepšie orientovanie vyniesol do grafov.(Obr. 5.1)



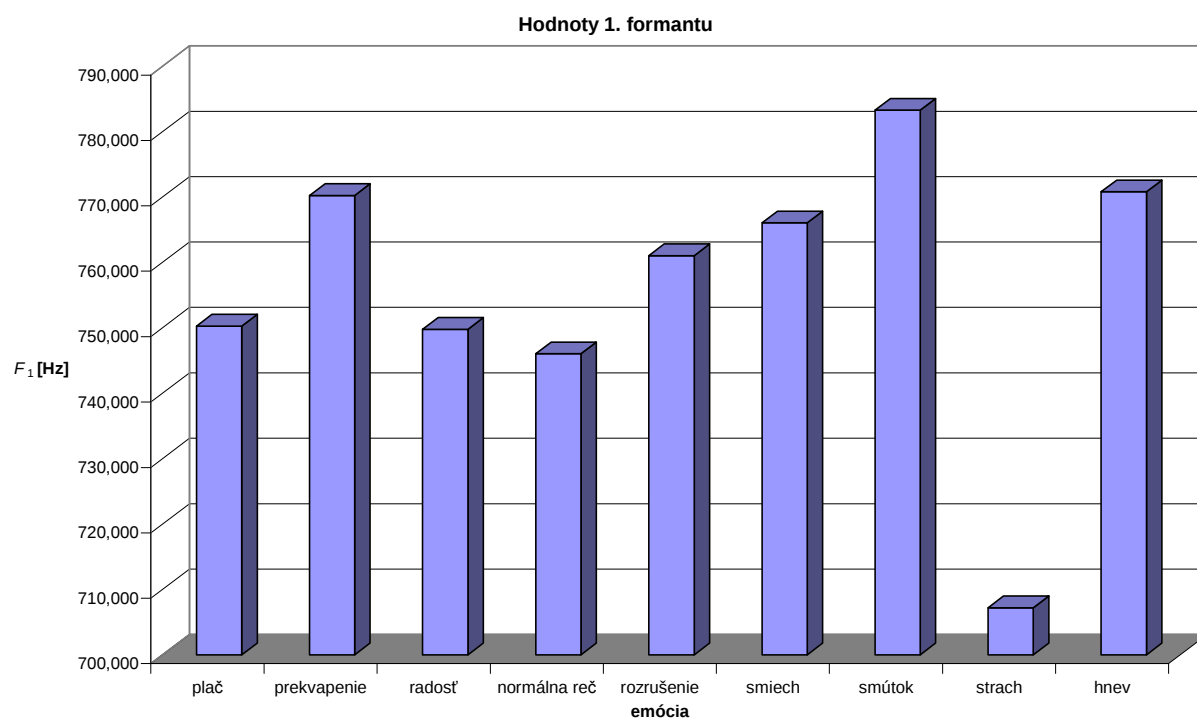
a)



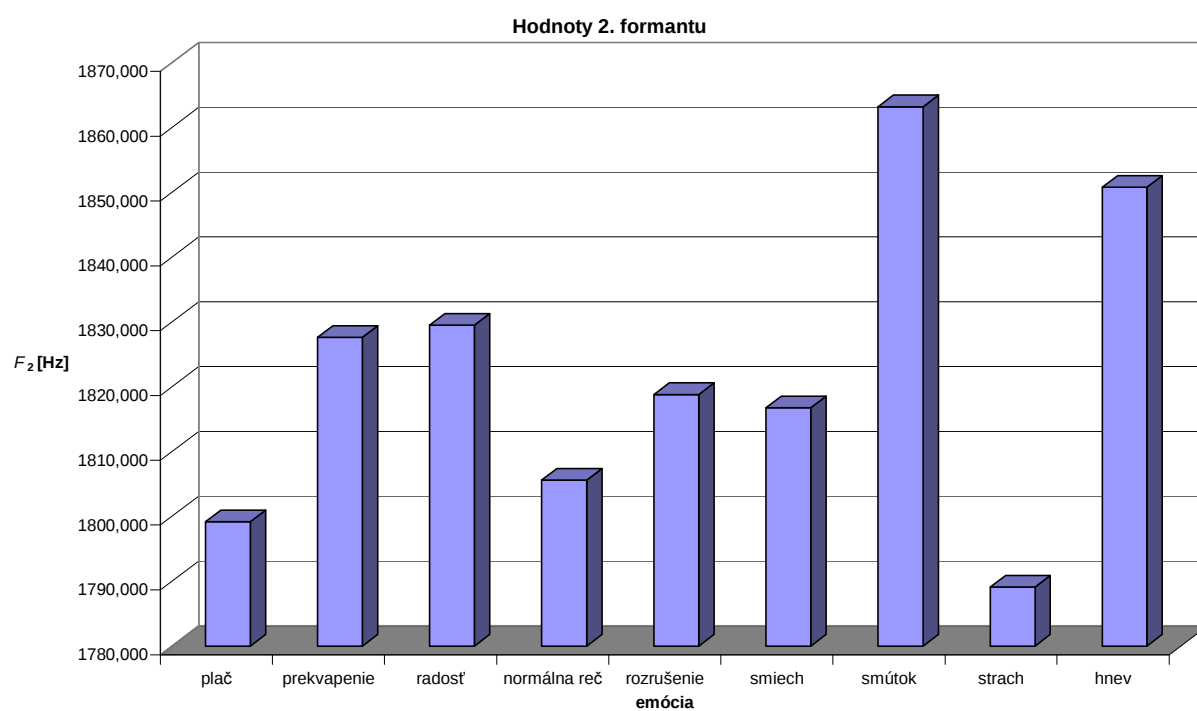
b)



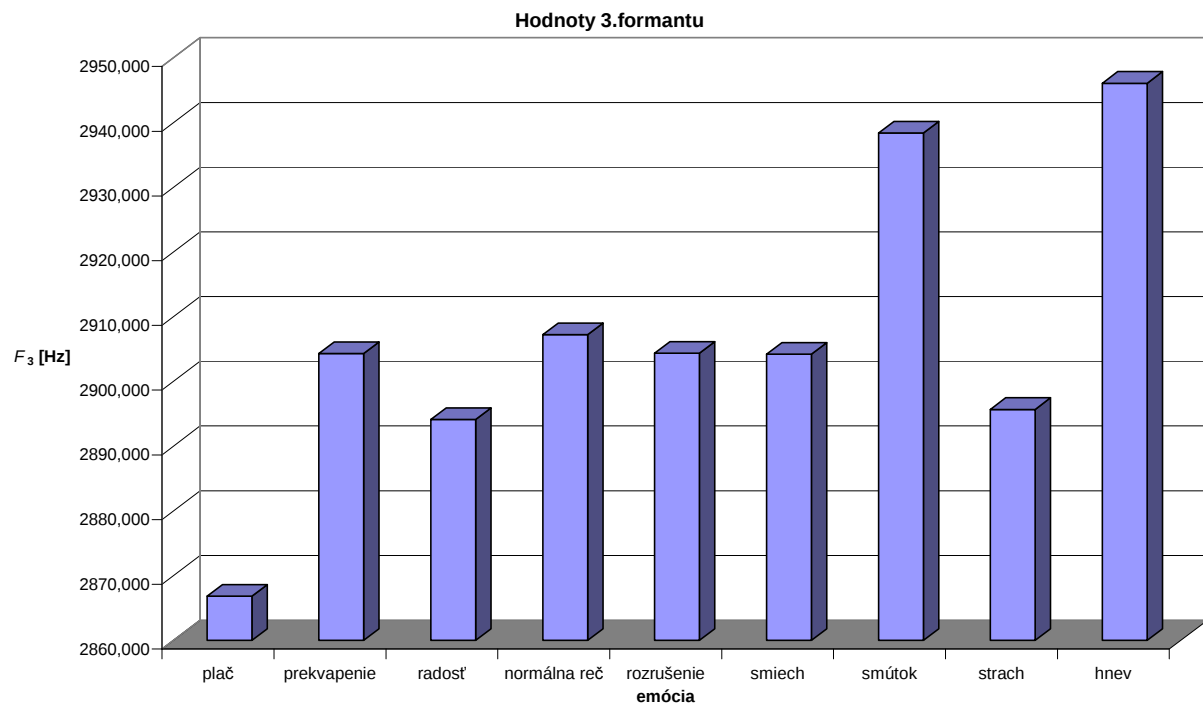
c)



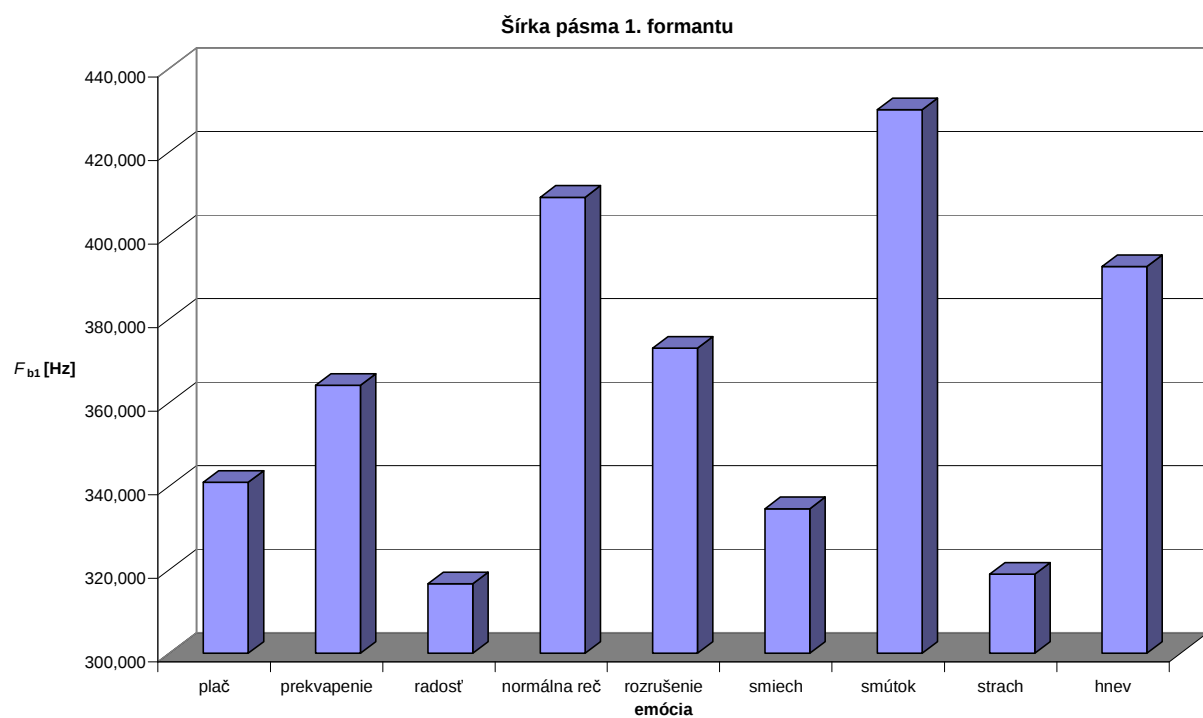
d)



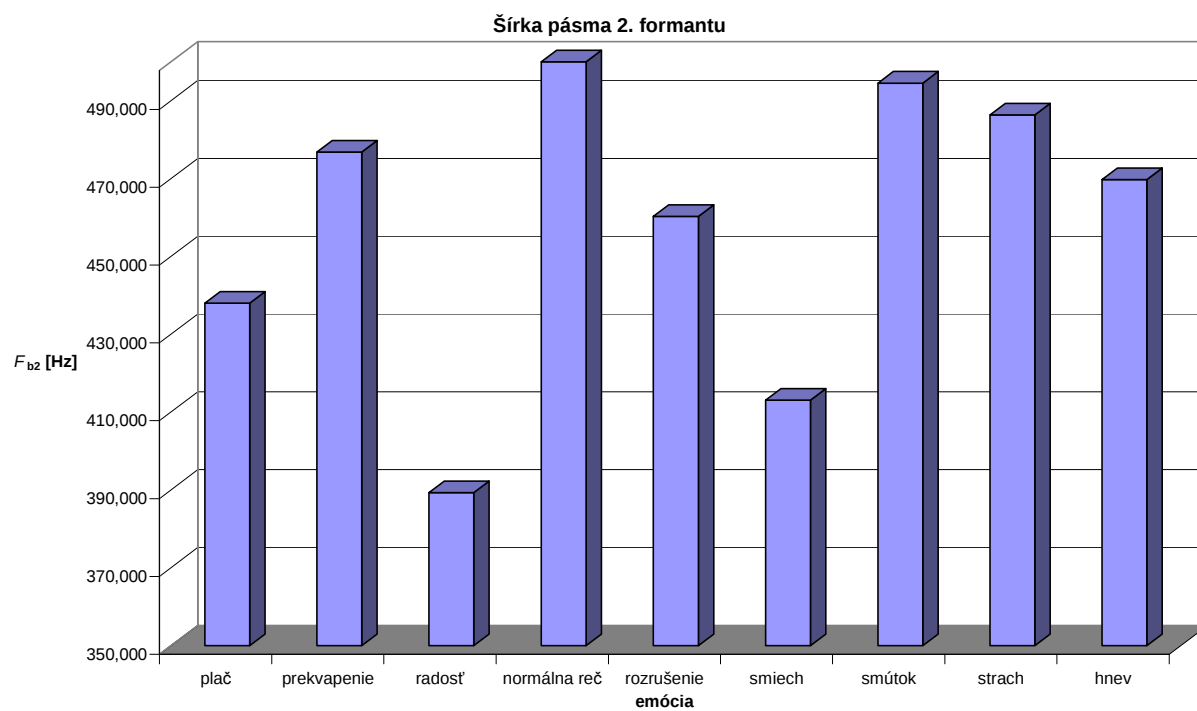
e)



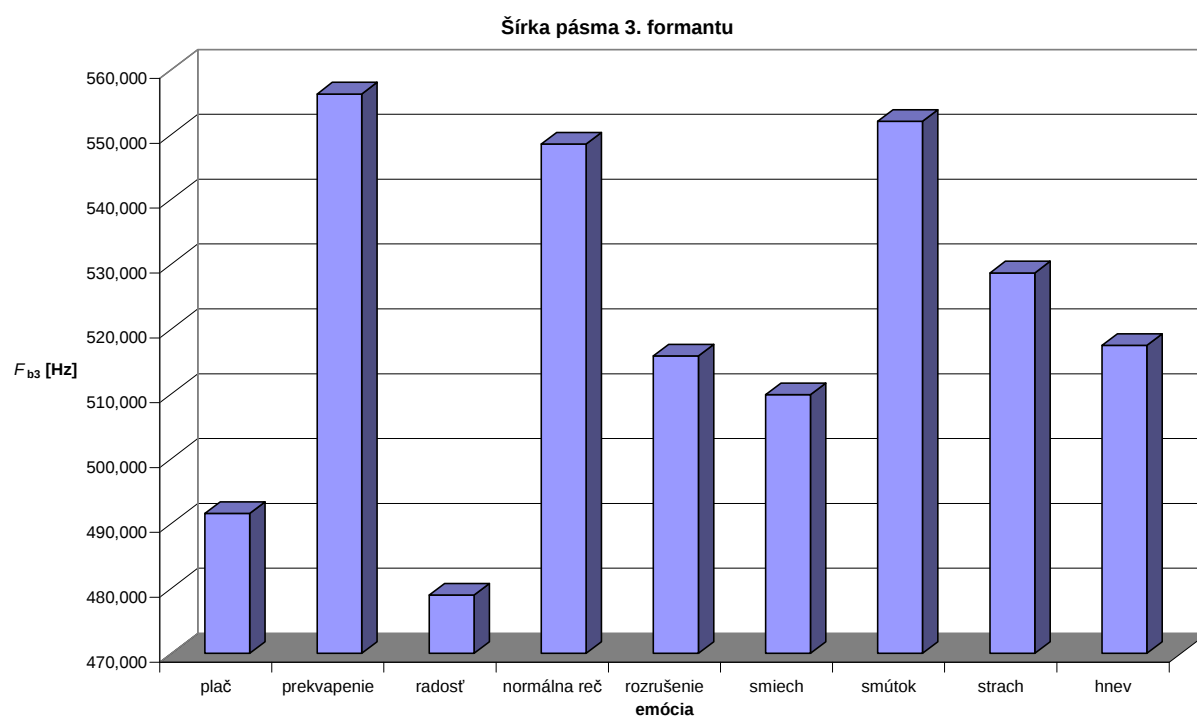
f)



g)

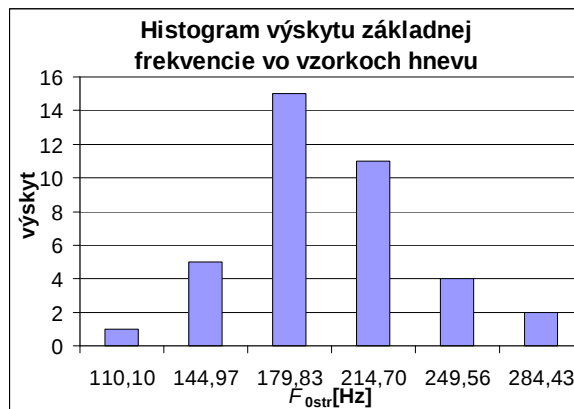
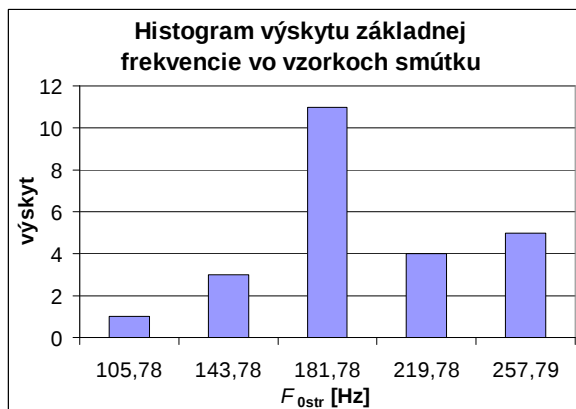
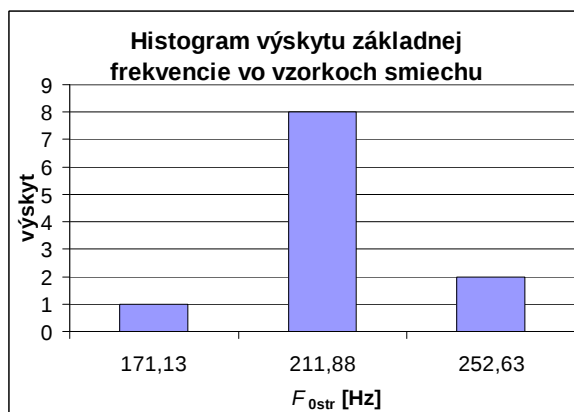
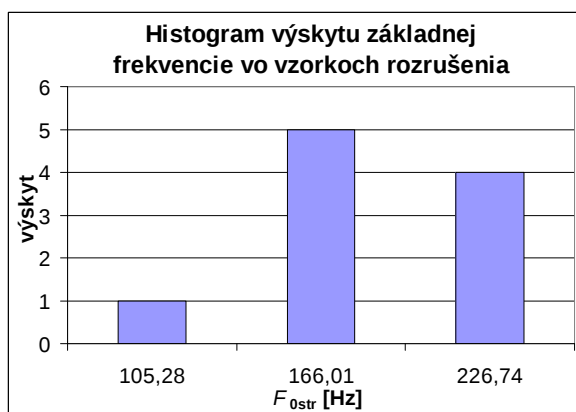
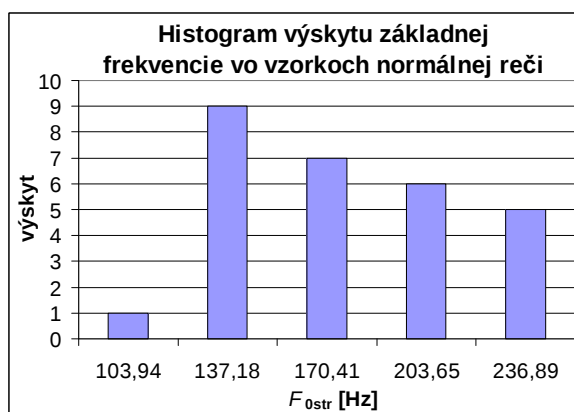
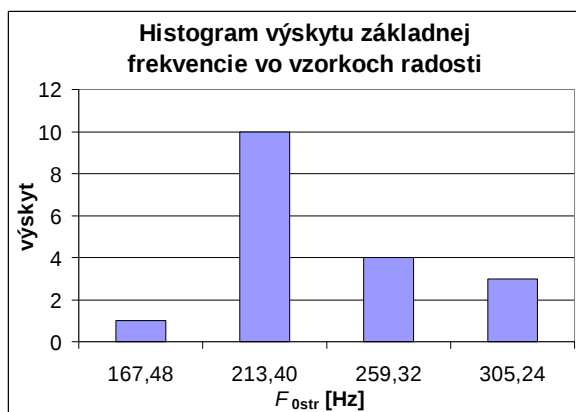
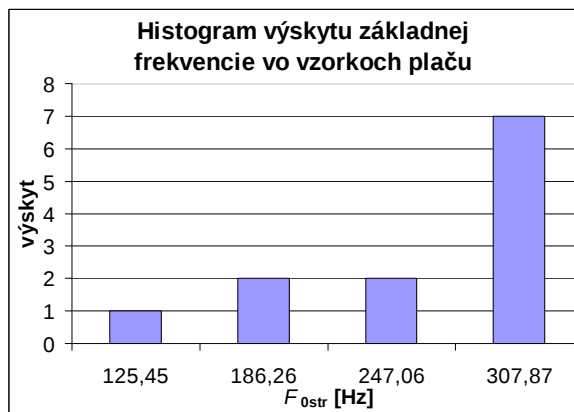
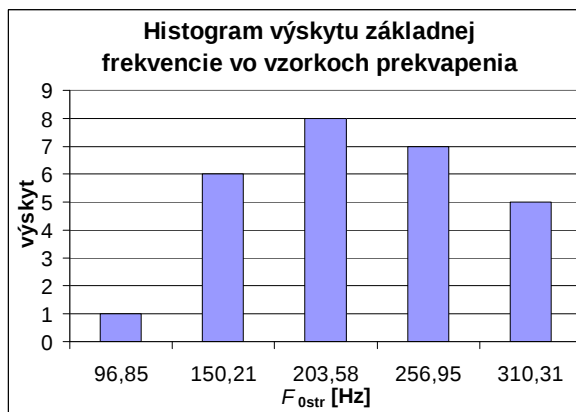


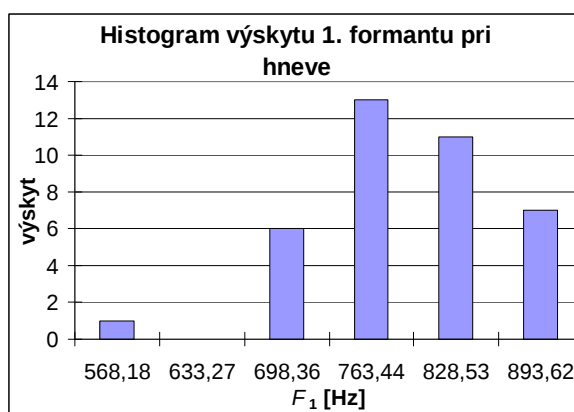
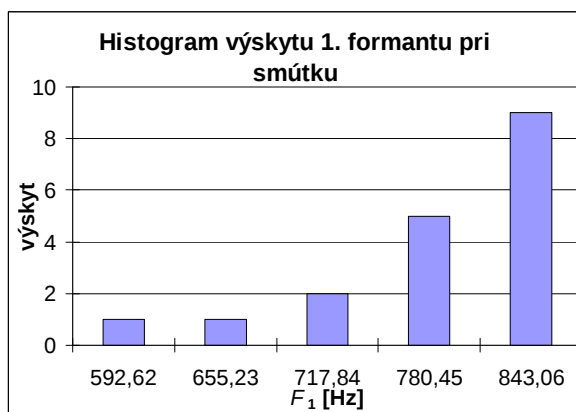
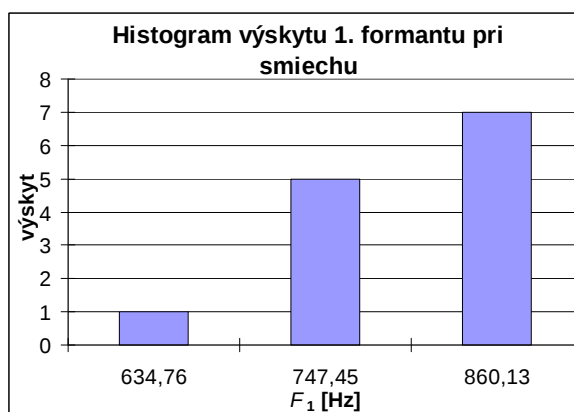
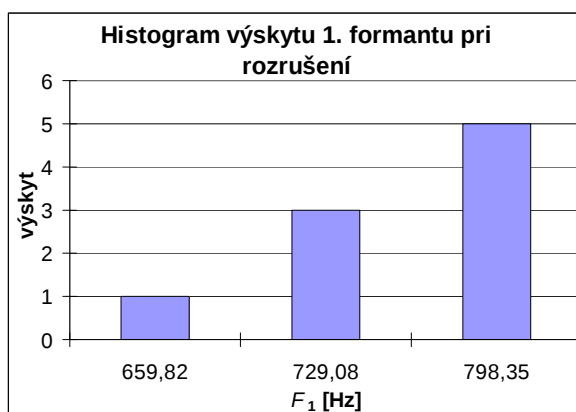
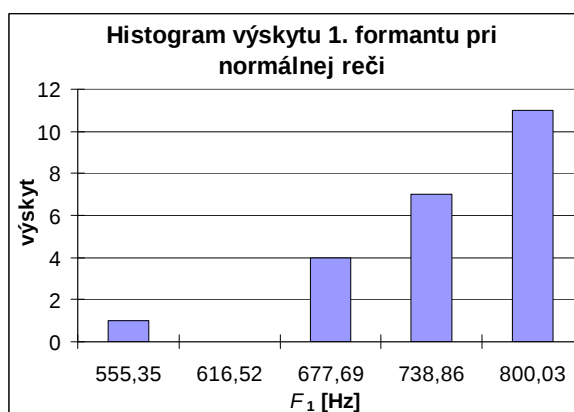
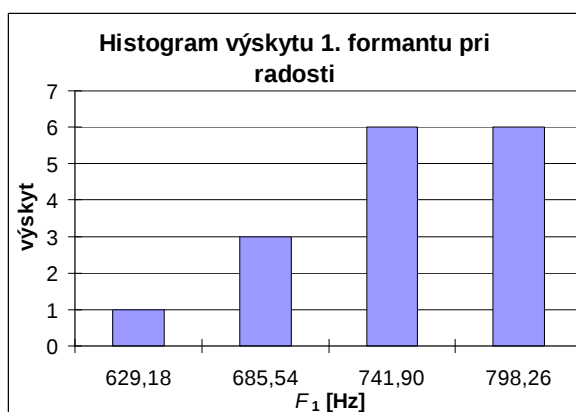
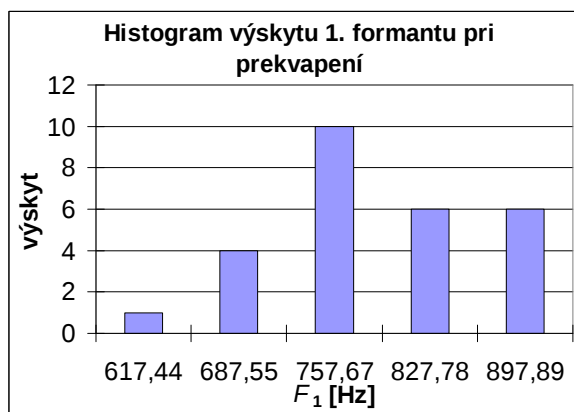
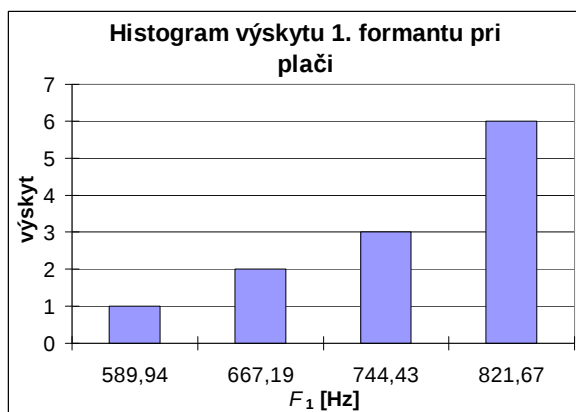
h)

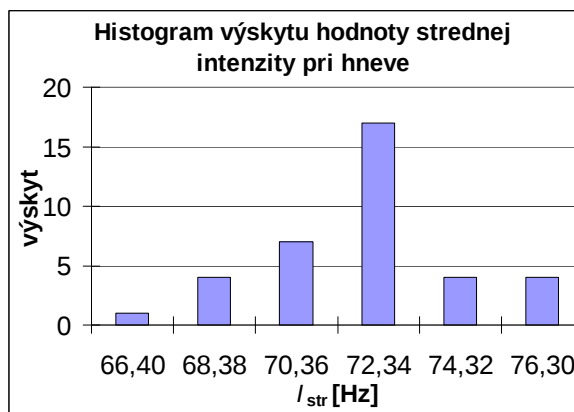
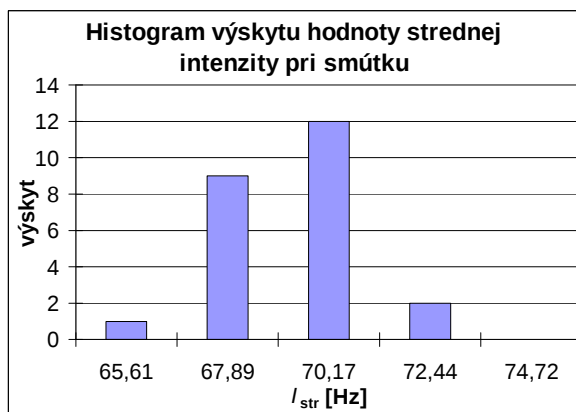
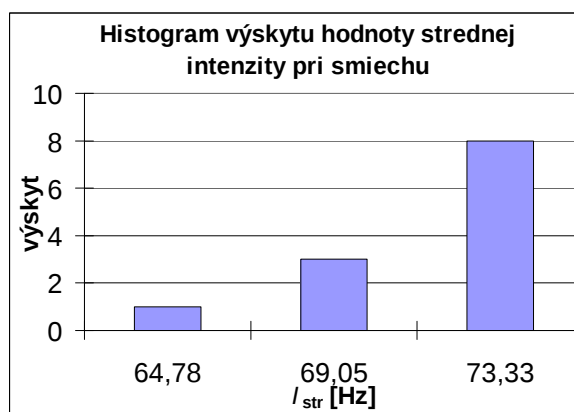
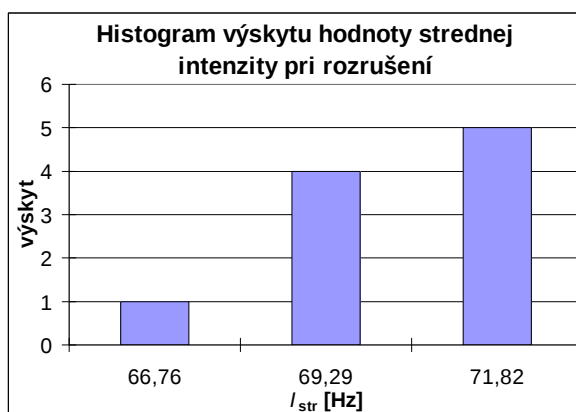
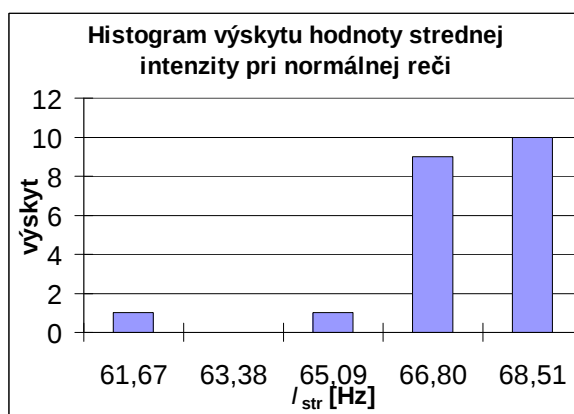
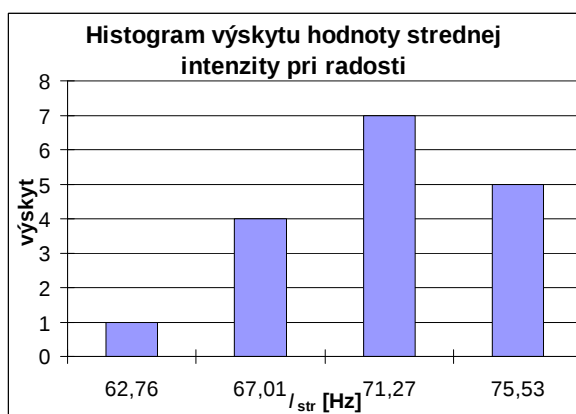
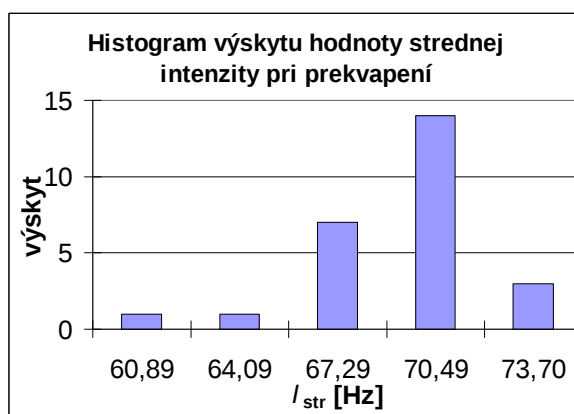
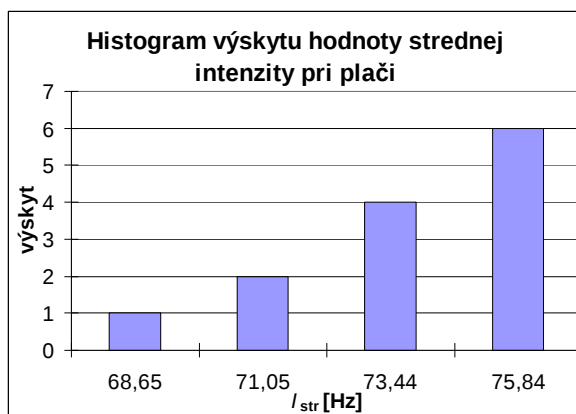


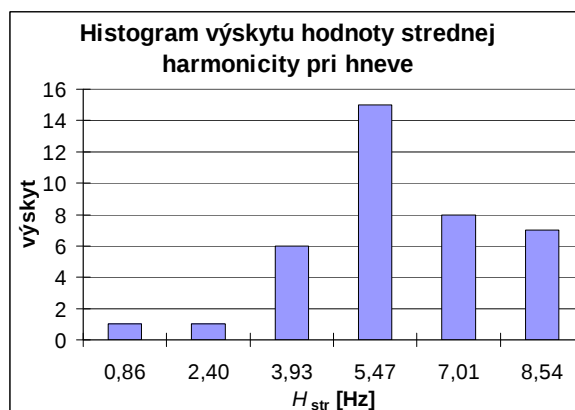
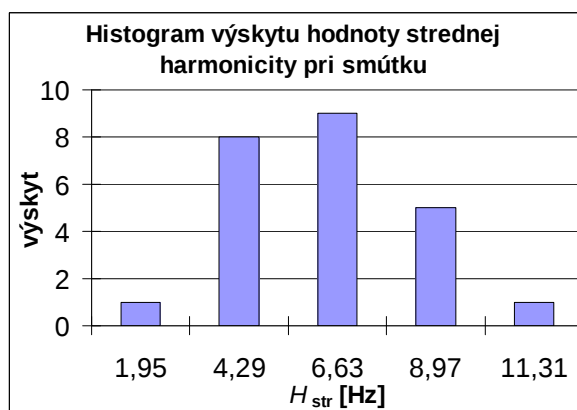
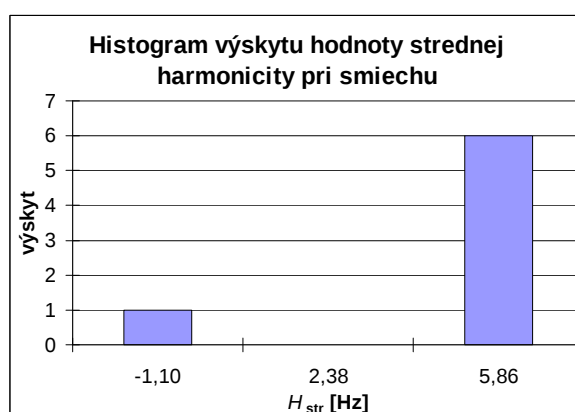
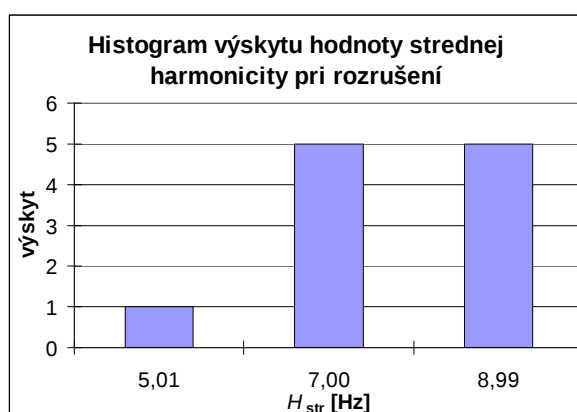
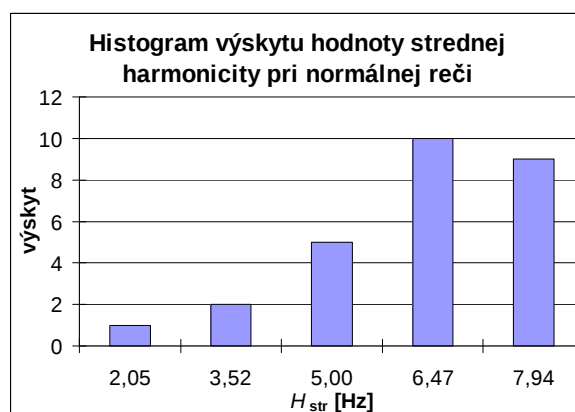
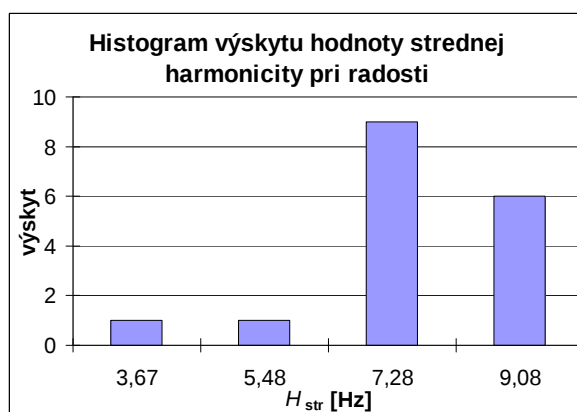
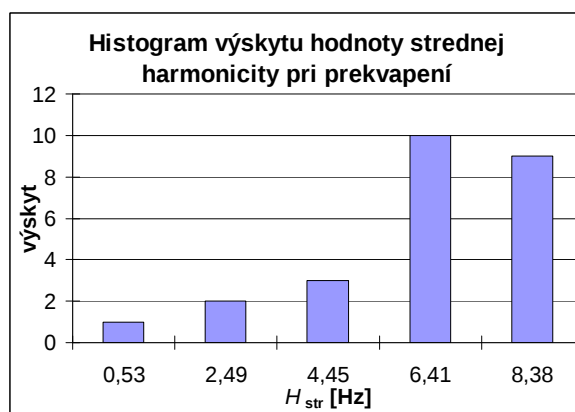
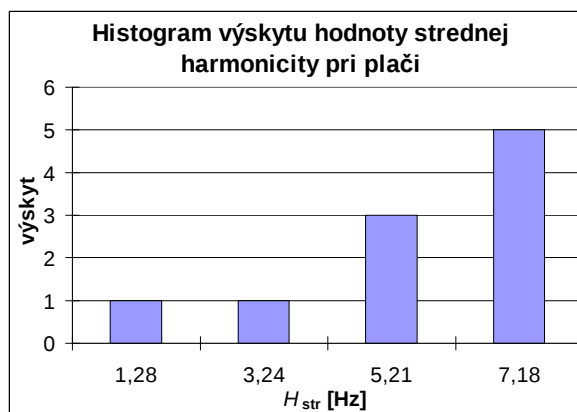
i)

Obr. 5.1: Grafy: a) stredná hodnota základnej frekvencie, b) stredná hodnota intenzity, c) stredná hodnota harmonicity, d) hodnoty 1. formantu, e) hodnoty 2. formantu, f) hodnoty 3. formantu, g) šírka pásma 1. formantu, h) šírka pásma 2. formantu, i) šírka pásma 3. formantu.









Obr. 5.2: Zobrazenie počtu výskytov jednotlivých príznakov v histogramoch.

5.2.3 Zhrnutie výsledkov

Z grafu strednej hodnoty základnej frekvencie je vidieť, že bežná reč má v porovnaní s ostatnými emóciami veľmi nízku základnú frekvenciu. Pri smiechu je naopak základná frekvencia podstatne vyššia, čo vyplýva z väčšieho hlasivkového napätia pri smiatí. Radosť má tak isto veľmi vysokú základnú frekvenciu, keďže pocit radosti spájame so ziskom, chceme týmto emočným stavom oznámiť tento zisk aj svojmu okoliu.

Stredná hodnota intenzity je nízka pri bežnej reči, smútku a prekvapení, kedy človek prejavuje emócie tichším hlasom. Naopak hnev, radosť či smiech sú emócie ktoré chceme aby dobre vnímalo okolie, a preto sa prejavujú intenzitou vysokou.

Stredná hodnota harmonicity predstavuje mieru akustickej periodicity, pomer signálu ku šumu, alebo kvalitu hlasu. Z grafu vyplýva že pomer šumu v normálnej reči tvorí asi 20%. Pokles miery šumu pod 20% sa prejavilo najmä u dobre zreteľných emócií, pri prejavovaní ktorých správne artikulujeme. Sú to strach, radosť, rozrušenie. Naopak väčšia miera šumu je pri hneve, pretože zvyšujeme intenzitu na úkor artikulácie. Pri smútku kde hovoríme potichu na úkor vynechania znelých tónov tvorených hlasivkami a tvorbou reči ústnou dutinou.

Z grafov hodnôt formantov môžeme pozorovať že hodnoty formantov sú poklesnuté u strachu. Toto môže byť spôsobené stresom a rozrušením pri ktorom sa tak isto hodnoty formantov držia v spodnej polovici poľa grafu. Naopak vyššie frekvencie nahromadenia hlasovej energie sa preukazujú pri hneve smútku a prekvapení.

Pri radosti je nahromadenie energie veľmi konkrétne, šírky pásiem u všetkých troch formantov sú veľmi úzke. Rovnako pomerne úzke sú šírky pásiem formantov aj u smiechu a plaču. Strach má veľmi zreteľný iba prvý formant kde je šírka pásma veľmi malá. Ďalšie dva formanty sa pri strachu strácajú, majú veľmi veľkú šírku. Rovnako je šírka pásiem všetkých formantov veľmi veľká u bežnej reči, smútku a hneve kde sa energia neviaže na konkrétnu frekvenciu.

Z histogramov môžeme vidieť najčastejší výskyt jednotlivých parametrov pri konkrétnych emóciách. Hodnoty ktoré mali najväčší výskyt môžeme považovať za smerodajné. Pri vzorkoch plaču sa najčastejšie hodnota pohybuje v okolí hodnoty 307 Hz. Tieto najčastejšie výskyty sa dajú zhrnúť v zvláštnej tabuľke. Rovnako sa dá zosumarizovať najčastejší výskyt formantov, strednej hodnoty intenzity a harmonicity.(Tab. 5.3)

príznak	Jednotka	plač	prekvapenie	radosť	normálna reč	rozrušenie	smiech	smútok	hnev
F_{0min}	[Hz]	308	203	213	138	166	212	180	195
F_1	[Hz]	820	758	775	800	790	860	843	764
I_{str}	[dB]	75	70	72	67	70	74	79	73
H_{str}	[dB]	6,5	7	7,5	6,8	8	6	5	5,5

Tab. 5.3: Tabuľka hodnôt v okolí ktorých je najčastejší výskyt jednotlivých veličín vo vzorkoch.

6 VÝBER VHODNÝCH PRÍZNAKOV

V ďalšej časti sme sa rozhodli pokračovať v rozpoznávaní štyroch vybraných emočných stavov, ktorými sú prekvapenie, normálna reč, smútok a hnev. Pri vyberaní vhodných príznakov, použiteľných na rozpoznávanie, sme použili software Matlab, konkrétne funkciu Rank features. Pri samotnom rozpoznávaní využijeme metódu neurónových sietí.

6.1.1 Matlab

Názov pochádza z MATrix LABoratory. Tento software je predovšetkým určený na technické výpočty, vizualizáciu, analýzu dát, ale aj programovanie algoritmov. Pracuje pomerne jednoduchým spôsobom. Príkazy priamo korešpondujú s názvami matematických operácií. K Matlabu existujú aj mnohé nadstavby a toolboxy. Sú to zbierky funkcií, ktorými sa rozširuje schopnosť Matlabu riešiť niektoré ďalšie problémy, ako napríklad štatistické spracovanie dát, optimalizácia, symbolické výpočty a podobne.

6.1.1.1 Rank features (hodnotenie príznakov)

Je to konkrétna funkcia Matlabu. Vyberá kľúčové prvky, podľa ktorých sa dajú určiť kritériá oddeliteľnosti.

Pri používaní Rank features funkcie v Matlabe, sme využili načítanie dát z programu Microsoft Excel. Zápis použitého skriptu vyzeral nasledovne:

```
prekv_rec_ne = xlsread('D:\skuska.xls', 'C4:T61'); .....načítanie dát z Exclu
prekv_rec = prekv_rec_ne'; .....transponovanie matice
classes = xlsread('D:\class.xls', 'A1:AC2'); .....načítanie delenia tried z Exclu
I = rankfeatures(prekv_rec,classes,'NumberOfIndices',18) .....telo funkcie
```

Pri zisťovaní vhodných príznakov sme určovali oddeliteľnosť každej emócie s každou, a teda vzniklo 6 poradí príznakov zoradených od najvhodnejšieho po najnevhodnejší použitý príznak.(Tab. 6.1)

	Príznak	Skratka	č.p.	prekv_rec	prekv_smut	prekv_hnev	rec_smut	rec_hnev	smut_hnev
Základná frekvencia	minimálna hodnota	F_{min}	1	9	4	9	3	18	11
	maximálna hodnota	F_{max}	2	7	9	12	11	3	15
	stredná hodnota	F_{str}	3	15	7	4	18	7	6
	smerodatná odchýlka	F_{std}	4	4	15	18	17	17	14
1. formant	stredná frekvencia	F_1	5	12	18	15	13	15	10
	šírka pásma	F_{b1}	6	1	17	10	14	1	7
2. formant	stredná frekvencia	F_2	7	3	12	6	1	4	13
	šírka pásma	F_{b2}	8	14	16	14	6	10	5
3. formant	stredná frekvencia	F_3	9	16	5	2	12	6	4
	šírka pásma	F_{b3}	10	5	1	7	7	5	17
Intenzita	minimálna hodnota	I_{min}	11	11	11	17	15	9	16
	maximálna hodnota	I_{max}	12	13	2	1	9	16	12
	stredná hodnota	I_{str}	13	10	10	11	4	13	2
	smerodatná odchýlka	I_{std}	14	6	13	16	10	8	8
Harmonicita	minimálna hodnota	H_{min}	15	2	8	5	2	12	1
	maximálna hodnota	H_{max}	16	18	14	13	5	14	3
	stredná hodnota	H_{str}	17	17	6	3	16	2	9
	smerodatná odchýlka	H_{std}	18	8	3	8	8	11	18

Tab. 6.1: Tabuľka zoradenia vhodnosti príznakov funkciou Rank features.

7 NEURÓNOVÉ SIETE

Umelé neurónové siete pracujú na princípe svojej biologickej predlohy, ktorou je nervová sústava človeka. Základným prvkom nervovej sústavy je neurón. Neurón je bunka určená na prenos, uchovávanie a spracovanie informácií. Takýchto buniek má ľudský mozog miliardy a každý z nich je prepojený so stovkami až tisíckami ďalších. Základom každého neurónu je bunečné jadro, z ktorého vychádzajú nervové výbežky (dendrity a axon). Dendrity tvoria vstup do neurónu. Každý neurón ich má veľké množstvo, zatiaľ čo axon je len jeden, zakončený terminálmi, ktoré tvoria výstup vzruchov. K prenosu vzruchov dochádza pomocou synapsií. Tie tvoria chemické reakcie založené na vylučovaní, alebo pohlcovaní neurochemickej látky z jedného neurónu na druhý. Synapsie majú rozdielnu priechodnosť. Tá sa s každým prenosom mení, čo je predpokladom učenia a pamätania informácie [4].

7.1 HISTÓRIA

História neurónových sietí siaha až do roku 1943, kedy Warren McCulloch a Walter Pitts vytvorili základný matematický model neurónu, ktorý tvorí základný stavebný prvok neurónových sietí. Donald Hebb v roku 1949 publikoval zákon, pomocou ktorého bolo možné vysvetliť učenie neurónových sietí. V roku 1958 Frank Rosenblatt objavil algoritmus učenia pre viac vrstvové siete s dopredným šírením signálu. Aj napriek rozvoju a veľkým nádejám, ktoré sa neurónovým sieťam pripisovali, dochádzalo k útlmu. Dôvod boli práve veľké nádeje, ale aj teória Marvinu Minského a Seymoura Paperta. Dokázali, že vtedy jediný známy model neurónu (perceptron) nedokáže riešiť funkciu XOR. Preto považovali ďalší výskum umelých neurónových sietí za neperspektívny.

Ďalšie obdobie až do začiatku osemdesiatych rokov sa v histórii neurónových sietí nazýva temným obdobím. V tomto čase sa výskumu a rozvoju neurónových sietí venovala iba veľmi malá časť vedcov. Neskôr americká organizácia DARPA (Defence Advance Research Project Association) začala vynakladať financie na výskum neurónových sietí. Vďaka tomu a vďaka objaveniu algoritmu spätného šírenia chyby, sa odborná verejnosť opäť začala zaujímať o tento obor. To samozrejme prinieslo ďalšie pokroky, ako napríklad objav Teuva Kohonena. Sieť, ktorá nepotrebuje učiteľa a sama mení svoje vnútorné chovanie. Dnes sa vývoj neurónových sietí zaoberá najmä vylepšením algoritmov a zmenšením výpočtovej náročnosti [4].

7.2 VYUŽITIE

Neurónové siete sa dnes využívajú v mnohých oblastiach. Ich najväčšou výhodou je uplatnenie v prípadoch, kde nie je známy konkrétny algoritmus na riešenie problému, ale aj tam kde nie sú známe presné, alebo sú neúplné informácie. Veľkú pozornosť vedci venovali najmä analýze a spracovaniu signálov. Napríklad vyhľadávanie určitých zložiek v EKG. Ide o klasifikačnú úlohu, ktorá dokáže lekárom zjednodušiť hľadanie konkrétnych vzorkov v zázname dlhom aj niekoľko desiatok hodín. Medzi najobľúbenejšie príklady využitia neurónových sietí patrí rozpoznávanie obrazov. Môžu sa použiť pri rozpoznávaní písma, znakov, ale aj rôznych objektov a osôb. Ďalším využitím je predikcia časových radov, kde sa na základe výberu vhodných vstupných dát zostavuje vhodný model pre predikciu. Tento model sa využíva pri predpovediach počasia, kurzov mien, spotreby elektrickej energie, hustoty premávky či vývoja zdravotného stavu pacienta. Uplatnenie neurónových sietí v riadení je najmä v automatickom riadení vozidiel, alebo riadení križovatiek. Významná je oblasť analýzy dát, ktorá sa využíva v lekárstve, ale aj v strojárstve. Vo svete majú neurónové siete aj množstvo ďalších využití ako napríklad pre vojenské účely, v robotike, bankovníctve

a v zábavnom priemysle. Za posledných 20 rokov bolo dosiahnutého významného pokroku vo využívaní neurónových sietí a ich ďalšie nasadzovanie je určite veľmi perspektívne [4].

7.2.1 Rozdelenie oblastí využitia

7.2.1.1 Klasifikácia

Úloha označovaná aj ako rozpoznávanie. Ide hlavne o rozhodovanie do ktorej triedy sa majú zaradiť vstupné dáta. Ak máme k dispozícii dostatočné množstvo učebných vzoriek, býva spoľahlivosť rozpoznania veľmi vysoká. Pre klasifikáciu sa používa množstvo rôznych typov sietí a učiacich metód [4].

7.2.1.2 Filtrácia

Je špeciálny prípad klasifikácie, jej cieľom je detekcia a izolácia známeho signálu od šumu. Vyskytujú sa u nej neuróny s nelineárnou prenosovou funkciou, organizované do viac vrstvovej siete. Miera úspešnosti filtrácie je závislá na podobnosti dát použitých na tréning a na samotnú filtráciu [4].

7.2.1.3 Predikcia

V mnohých praktických aplikáciách, je potrebné predpovedanie budúceho chovania veličín, na základe ich chovania v minulosti. Za týmto účelom bola už dlhšiu dobu rozvíjaná rada rôznych metodík. Niektoré z nich sú založené na algoritmických prístupoch a využívajú schopnosť neurónových sietí učiť sa. I keď v posledných rokoch sa vyvinulo a prakticky využilo veľa rôznych výkonných algoritmických predikčných metód a postupov, hlavnou výhodou neurónových sietí je ich schopnosť adaptovať sa na základe známych vzorov a možnosť využitia nelinearity [4].

7.2.1.4 Kompresia

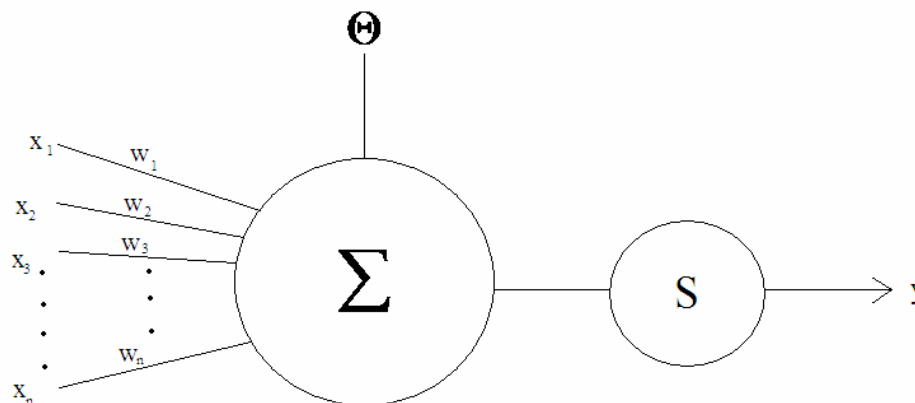
Za príklad môžeme použiť prenos televízneho signálu. V tomto prípade sa využíva viac vrstvová sieť s rovnakým počtom vstupných a výstupných neurónov s výrazne nižším počtom neurónov v skrytej vrstve. Po natrénovaní siete na určitú množinu dát je sieť rozdelená na vysielaciu a prijímaciu časť. Kompresnú (vysielaciu) časť tvorí vstupná vrstva a jedna skrytá vrstva, dekompresnú (prijímaciu) tvorí ďalšia skrytá vrstva a výstupná vrstva. Na ceste komprimovaného signálu je potrebné zaistiť čo najnižšie skreslenie. Úroveň kompresie a presnosť zrekonštruovaného signálu je závislá na rozložení neurónov v sieti [4].

7.3 NEURÓN

Najpoužívanejším neurónom v praxi je perceptrón vynájdený McCullochom a Pittsom, teda McCulloch-Pittsov perceptron. Tento je vybavený konečným počtom vstupov x_i . Každý vstup má svoju váhu w_i , prenosovú funkciu S , aktivačný prah Θ a výstup y . Každý neurón je popísaný rovnicou 7.1 [4].

$$y = S\left(\sum_{i=1}^n w_i x_i + \Theta\right) \quad (7.1)$$

Do vstupov neurónu sú privádzané dáta diskretných číselných množín, alebo dáta spojité. Všetky vstupy sú modifikované váhou. Váhy bývajú na začiatku učenia väčšinou generované náhodne a počas učenia sa menia. Úprava váh je vlastne podstatou učenia a uloženia skúseností neurónu [4].



Obr. 8.1: Model neurónu.

Pri prekročení prahu sa stáva neurón aktívny. Zisťuje sa ako vážený súčet všetkých vstupov. Prah býva často realizovaný ako ďalší vstup neurónu s hodnotou 1.

Aktivačná funkcia, tiež nazývaná ako prenosová, prevádza vnútorný potenciál neurónu do definovaného oboru výstupných hodnôt. Aktivačná funkcia tiež určuje či je neurón binárny, alebo spojitý. Toto rozlišovanie by sme ale nemali zamieňať s typmi neurónov [4].

7.4 UČENIE

Učenie, v mnohých prípadoch nazývané aj tréning, môžeme definovať ako nastavovanie synaptických váh a prahov podľa algoritmu učenia. Jeho podstatou je výber charakteristických znakov a skúseností zo vstupných signálov a parametrov neurónovej siete tak, aby odchýlka medzi požadovaným a skutočným výstupom bola minimálna.

Tak ako u živých organizmov prebieha učenie pomocou opakovania, tak i u umelej neurónovej siete prebieha opakovaný učiaci proces, až kým nie je dosiahnutá minimálna odchýlka výstupných dát od požadovaného výstupu. Existuje však aj učenie jednorazové, pri ktorom si sieť vzor zapamätá a už nedochádza k ďalším úpravám [4].

7.4.1 Typy učenia

Algoritmov zaoberajúcich sa učením neurónových sietí je veľké množstvo. Spoločným základom väčšiny z nich je Hebbov zákon učenia. Jeho základom je proces učenia v biologických sieťach a teda hovoríme, že pri učení dochádza k modifikácii synapsií. Predpokladá sa, že váha synapsie dvoch aktívnych, prepojených neurónov sa posilňuje a naopak [4].

7.4.1.1 Učenie s učiteľom

Je založené na porovnaní výstupu so správnym výsledkom, ktorý bol zadaný učiteľom. Aby sa výstup čo najviac podobal správne výsledku sú pomocou spätnej väzby nastavované váhy a prahy neurónov. Sieť sa učí na tréningovej skupine, ktorá obsahuje vstupné

dáta aj správne výsledky. Ak sieť dáva na dané vstupy stabilné výstupy s nízkou chybovosťou, môžeme sieť považovať za naučenú. Príkladom takéhoto učenia môže byť spätné šírenie chyby [4].

7.4.1.2 Učenie bez učiteľa

Môžeme ho tiež nazývať samo organizácia. Nepotrebuje žiadnu tréningovú skupinu dát. Princípom algoritmu učenia je, že vo vstupných dátach rozoznáva vzorky s podobnými vlastnosťami a tie potom zaraďuje do skupín. Príkladom môže byť Habbov zákon učenia [4].

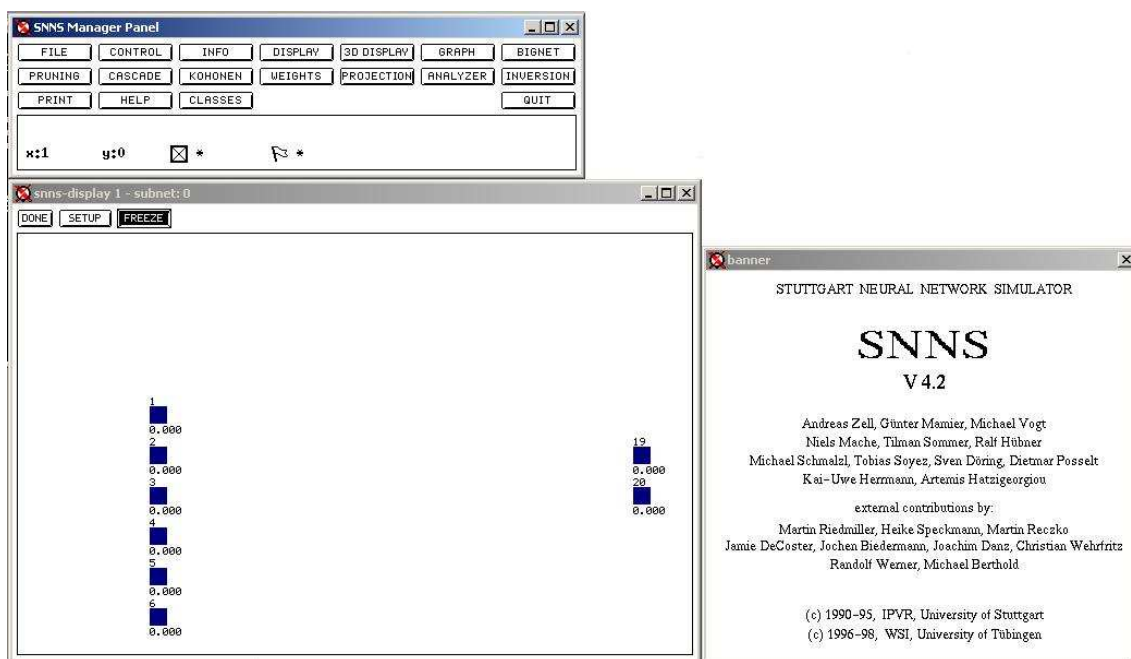
7.5 POUŽÍVANÝ SOFTWARE

7.5.1 SNNS (Stuttgart Neural Network Simulator)

Ide o simulátor pre neurónové siete vyvíjaný na Štutgartskej univerzite od roku 1989. Cieľom projektu je vytvoriť efektívne a flexibilné simulačné prostredie pre výskum a používanie neurónových sietí.

Simulátor má oddelené samotné jadro (kernel) od grafického rozhrania (GUI).

Dnes už existuje nová grafická nadstavba na jadre SNNS písaná v jazyku Java, JavaNNS s ktorou budeme pracovať aj v tejto práci [5].



Obr.8.2: Ukážka načítanej siete v SNNS.

8 ROZPOZNÁVANIE (KLASIFIKÁCIA)

8.1 JAVANNS

Java Neural Network Simulator je vytvorený na Wilhelm – Schickardovom inštitúte informatiky v Tübingene, v Nemecku. Je novým grafickým rozhraním, písaným v jazyku Java, pre Stuttgart Neural Network Simulator (SNNS) s jadrom 4.2. Má väčšinu schopností ako SNNS, pričom užívateľské rozhranie bolo navrhnuté jednoduchšie a intuitívnejšie na používanie. Niektoré zložité a nie často využívané funkcie SNNS boli vynechané a odložené do ďalších verzií. Vzhľadom k tomu že SNNS bol určený výhradne pre OS UNIX, JavaNNS beží na väčšine OS, kde je nainštalovaný Java Runtime Environment. Bol testovaný na: Windows NT, Windows 2000, RedHat Linux 6.1, Solaris 7, Mac OS X [6].

8.1.1 Inštalácia a prvé spustenie

V OS Windows XP software JavaNNS nie je nutné inštalovať. Spúšťa sa spustiteľným súborom JavaNNS.jar z hlavného adresára javanns. Po spustení programu, je úvodné okno prázdne, nie sú nahrané žiadne sieťové súbory, viditeľný je iba klasický panel hlavného menu.

8.1.2 Načítanie súborov

Prebieha rovnako ako u väčšiny programov cez menu *File->Open*. Otvárať môžeme súbory s rôznou koncovkou. Pre nás sú zaujímavé najmä súbory s koncovkou .net, ktoré predstavujú vytvorenú sieť a súbory s koncovkou .pat, ktoré môžu obsahovať sadu tréningových pravidiel, alebo rovnako vstupné dáta siete použiteľné na testovanie.

8.1.3 Práca so sieťou

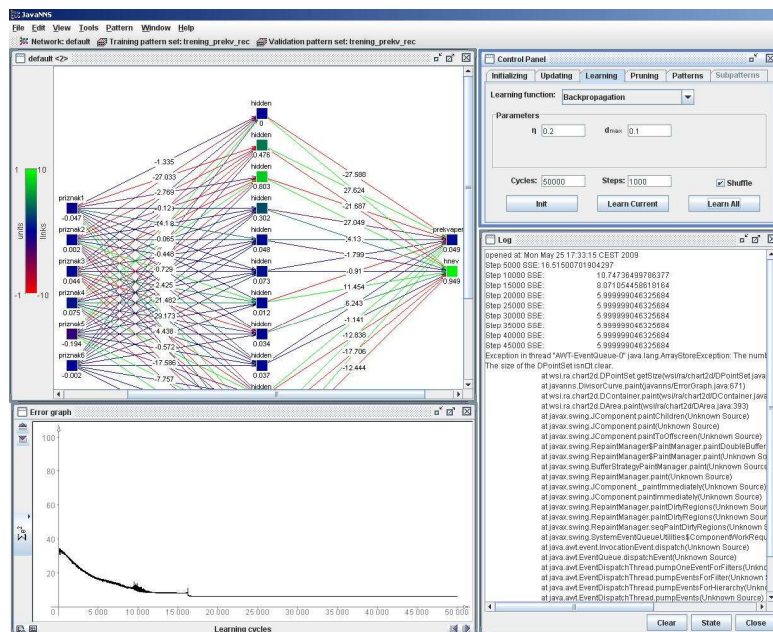
Pre zobrazenie načítanej siete zvolíme v menu *View->Network*. Otvorí sa pod okno v ktorom je schématicky zobrazená sieť. Neuróny a spoje majú rôzne farby, zastupujúce rôzne aktivačné hodnoty a váhy. Každá jednotka je koncipovaná ako štvorec s hranou 16 pixelov. Farebný pruh vľavo zobrazuje konkrétne hodnoty farieb. Tieto farby a hodnoty možno upravovať v menu *View->Display Settings*. Zobrazené pod okno obsahuje dve záložky: *General* v ktorej môžeme nastaviť farby jednotiek a spojov podľa aktivačnej hodnoty a váhy. *Units & Links* v ktorej užívateľ môže nastaviť vlastnosti jednotiek ako je meno, aktivačnú hodnotu, či popis jednotky. Niektoré vlastnosti sa dajú upravovať aj jednotlivo, mená cez menu *Edit->Names* a aktivačné hodnoty cez menu *Edit->Values*.

8.1.4 Tvorenie novej siete

Pri tvorení novej siete používame menu *File->New*. Táto voľba odstráni aktuálnu sieť. Hneď potom môžeme začať tvoriť novú sieť. Pri tvorení používame menu *Tools->Create->Layers* pre tvorbu vrstiev s jednotkami. V okne tvorby vrstvy nastavujeme hlavne počet riadkov a stĺpcov, ktoré má vrstva obsahovať a pozíciu vrstvy. Nastavujeme tiež typ vrstvy, aktivačnú a výstupnú funkciu. V menu *Tools->Create->Connections* tvoríme spoje. Pri tvorbe spojení môžeme vybrať jednotky, ktoré chceme spojiť, alebo zvoliť pospájanie feed-forward.

8.1.5 Inicializácia a tréovanie

Pri použití menu *Tools->Control Panel*, sa zobrazí Ovládací panel, ktorý obsahuje niekoľko záložiek. Prvou je *Initialzing*. Tu môžeme inicializovať novo vytvorenú sieť. To znamená priradiť náhodne generované váhy všetkým spojeniam. Druhou záložkou je *Learning*. Pomocou tejto záložky spúšťame učenie a najmä nastavujeme počet cyklov a krokov v kolíkych má byť sieť tréovaná. Ak stlačíme *Learn Current* učíme sieť z aktuálne nastavených hodnôt. *Learn All* učí so všetkých hodnôt z tréovacieho súboru. Poslednou dôležitou záložkou je *Patterns*, kde môžeme pracovať priamo so súborom tréovacích a vstupných hodnôt. Priebeh inicializácie, tréovania a ďalších operácií so sieťou môžeme sledovať pomocou *Log* okna, alebo *Error Grafu*, ktoré zobrazíme z menu *View*. (Obr.8.1)



Obr.8.1: JavaNN pri tréovaní neurónovej siete, so zobrazením error grafu a log listu.

8.1.6 Formát tréovacích a vstupných súborov

Formát súborov je daný podobne ako v SNNS. Jedná sa o textový súbor, ktorý obsahuje hlavičku a tréovacie, prípadne vstupné dáta. Súbor vyzerá nasledovne:

SNNS pattern definition file V3.2
generated at Mon Apr 25 15:58:23 1994

No. of patterns : 2
No. of input units : 2
No. of output units : 1

Input pattern 1:
0 0
Output pattern 1:
0
Input pattern 2:
0 1
Output pattern 2:
1

8.1.7 Analýza

Pri samotnom používaní siete počas tzv. života siete môžeme použiť analyzátor, ktorý zobrazíme z menu *Tools->Analyzer*. V prvej záložke *Setup* musíme nastaviť počet jednotiek v osiach x a y. V záložke *Test* nastavíme počet krokov a zmenu vstupu. Môžeme prípadne zaškrtnúť tréňovanie počas testu, a teda aj počet tréňovacích krokov.

8.1.8 Výstupné dáta

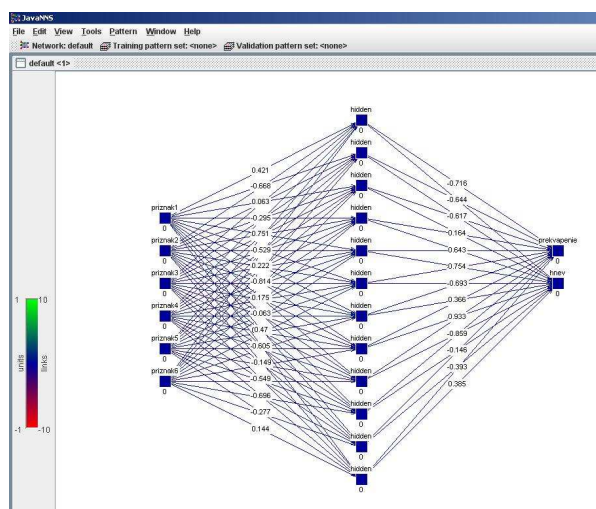
Súbor s výstupnými dátami uložíme cez menu *File->Save data* zvolením typu súboru .res.

8.2 PROCES ROZPOZNÁVANIA EMÓCIÍ

8.2.1 Architektúra neurónových sietí

Pri používaní programu Praat, sme za jeho pomoci získali so všetkých nahrávok 18 hodnôt príznakov. Za pomoci funkcie Rankfeatures sme zistili príznaky, ktoré sú pri rozpoznávaní jednotlivých emócií najvhodnejšie. Z predchádzajúcich bádani vieme, že nie je vhodné použiť všetkých 18 príznakov. Testy na neurónových sieťach sme teda robili s postupným pridávaním príznakov, a teda so 6, 8, 10 a 12 príznakmi na vstupe. Z toho vyplýva, použitie štyroch architektúr sietí:

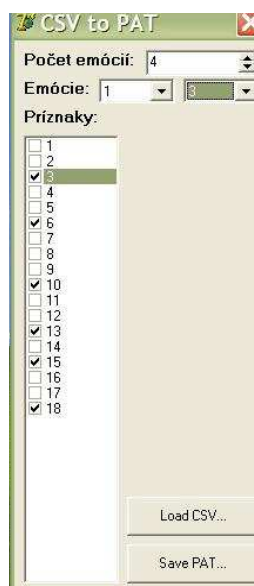
1. má vo vstupnej vrstve 6 neurónov podľa šiestich použitých príznakov. Do skrytej vrstvy sme sa na základe poznatkov rozhodli umiestniť 12 neurónov a do výstupnej vrstvy dva neuróny, ktoré predstavovali rozpoznávané emócie. (Obr. 8.2)
2. má vo vstupnej vrstve 8 neurónov podľa ôsmich použitých príznakov. Do skrytej vrstvy sme sa na základe poznatkov rozhodli umiestniť 16 neurónov a do výstupnej vrstvy dva neuróny, ktoré predstavovali rozpoznávané emócie.
3. má vo vstupnej vrstve 10 neurónov podľa desiatich použitých príznakov. Do skrytej vrstvy sme sa na základe poznatkov rozhodli umiestniť 18 neurónov a do výstupnej vrstvy dva neuróny, ktoré predstavovali rozpoznávané emócie.
4. má vo vstupnej vrstve 12 neurónov podľa dvanástich použitých príznakov. Do skrytej vrstvy sme sa na základe poznatkov rozhodli umiestniť 20 neurónov a do výstupnej vrstvy dva neuróny, ktoré predstavovali rozpoznávané emócie.



Obr.8.2: Ukážka architektúry neurónovej siete so šiestimi vstupnými príznakmi.

8.2.2 Trénovacie a testovacie dáta

Dáta do siete sa vkladajú pomocou textového súboru s koncovkou .pat. Ukážka tohto súboru je v kapitole 7.2.6. Vkladanie dát je asi najväčšou nevýhodou JavaNNS. Dáta do textového súboru treba ručne prepisovať čo je pri množstve trénovacích a testovacích dát veľmi zdĺhavé. My sme sa tento problém snažili aspoň čiastočne vyriešiť malým programom vyhotoveným v programovacom jazyku Delphi, ktorý dokáže dáta Microsoft Excelu uložené s koncovkou .csv, automaticky previesť do vstupného textového súboru JavaNNS s koncovkou .pat.



Obr. 8.3: Ukážka vytvoreného programu pre zjednodušenie prevodu .csv do .pat.

Tabuľku dát, v ktorej stĺpce predstavujú príznaky a riadky jednotlivé vzorky nahrávok, je nutné skopírovať do nového zošita programu Microsoft Excel, uložiť s koncovkou .csv a teda ako dáta oddelené bodkočiarkou. Dôležité je, aby počet vzoriek jednotlivých emócií bol rovnaký. Po otvorení programu je nutné nahráť súbor s koncovkou .csv, zadať koľko emočných stavov tabuľka obsahuje, ktoré dve emócie budeme od seba rozpoznávať a vybrať čísla príznakov ktoré chceme použiť pri rozpoznávaní. Na konci zostáva iba uložiť súbor s koncovkou .pat. Výstupom je pripravený trénovací súbor, vhodný pre JavaNNS ako aj pre SNNS.

8.2.3 Trénovanie a testovanie

Pri tréningu je možné nastavovať počet trénovacích cyklov a krokov. Zároveň aj to či majú dáta pri tréningu vstupovať postupne, alebo náhodným výberom. Pri testovaní našich neurónových sietí sme trénovali cez 50 000 cyklov a 1000 krokov. Toto je počet pri ktorom už zvyšovanie cyklov, alebo krokov zaručene neovplyvní náš výsledok testu. Každý tréning a test sme prevádzkali dva krát, pričom sme jedenkrát vkladali trénovacie dáta postupne a jeden krát náhodne. Výsledky testovania jednotlivých sietí sú zobrazené v tabuľkách a grafoch.

Počet použitých príznakov	Náhodný vstup tréningových dát	prekvapenie vs. normálna reč
		[%]
6	Nie	70
	Ano	75
8	Nie	60
	Ano	65
10	Nie	60
	Ano	80
12	Nie	65
	Ano	60
Priemer	Nie	63.75
Priemer	Ano	70

Počet použitých príznakov	Náhodný vstup tréningových dát	prekvapenie vs. smútok
		[%]
6	Nie	60
	Ano	60
8	Nie	45
	Ano	30
10	Nie	80
	Ano	70
12	Nie	60
	Ano	60
Priemer	Nie	61.25
Priemer	Ano	55

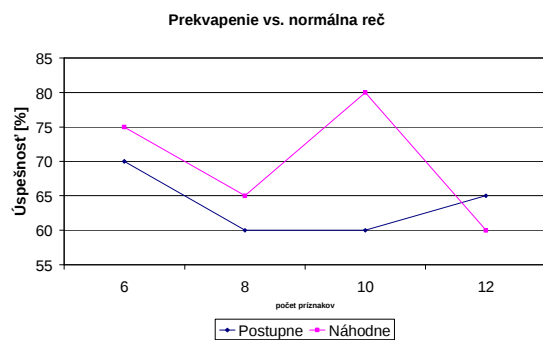
Počet použitých príznakov	Náhodný vstup tréningových dát	prekvapenie vs. hnev
		[%]
6	Nie	55
	Ano	65
8	Nie	80
	Ano	65
10	Nie	85
	Ano	80
12	Nie	75
	Ano	65
Priemer	Nie	73.75
Priemer	Ano	68.75

Počet použitých príznakov	Náhodný vstup tréningových dát	normálna reč vs. smútok
		[%]
6	Nie	55
	Ano	60
8	Nie	65
	Ano	50
10	Nie	50
	Ano	50
12	Nie	45
	Ano	55
Priemer	Nie	53.75
Priemer	Ano	53.75

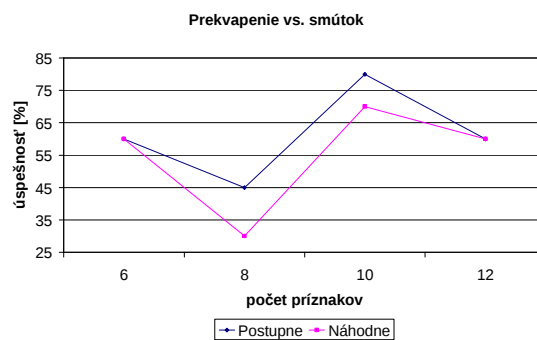
Počet použitých príznakov	Náhodný vstup tréningových dát	normálna reč vs. hnev
		[%]
6	Nie	55
	Ano	65
8	Nie	75
	Ano	75
10	Nie	65
	Ano	60
12	Nie	70
	Ano	80
Priemer	Nie	66.25
Priemer	Ano	70

Počet použitých príznakov	Náhodný vstup tréningových dát	smútok vs. hnev
		[%]
6	Nie	50
	Ano	60
8	Nie	60
	Ano	55
10	Nie	75
	Ano	65
12	Nie	80
	Ano	65
Priemer	Nie	66.25
Priemer	Ano	61.25

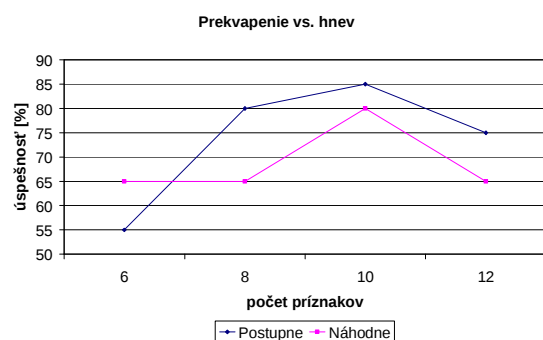
Tab.8.1: Prehľad úspešností rozpoznania emočných stavov v jednotlivých kombináciách.



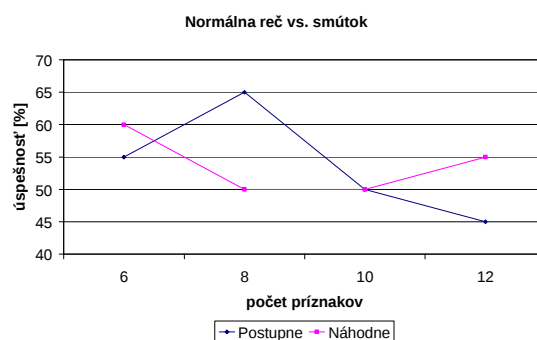
a)



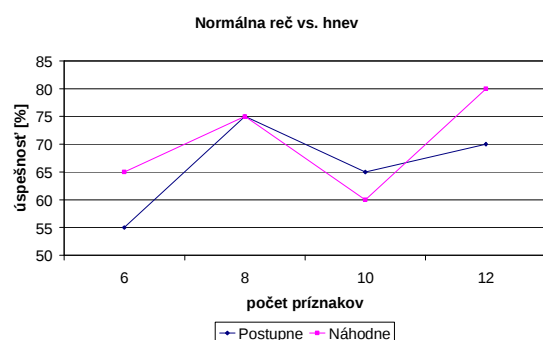
b)



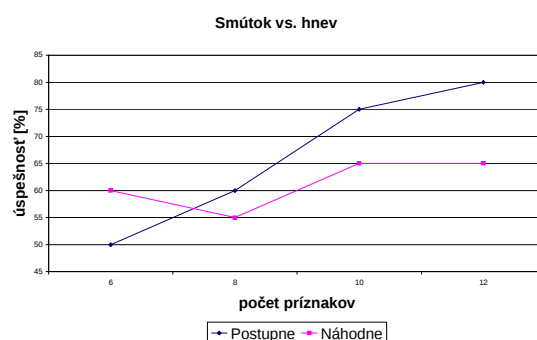
c)



d)

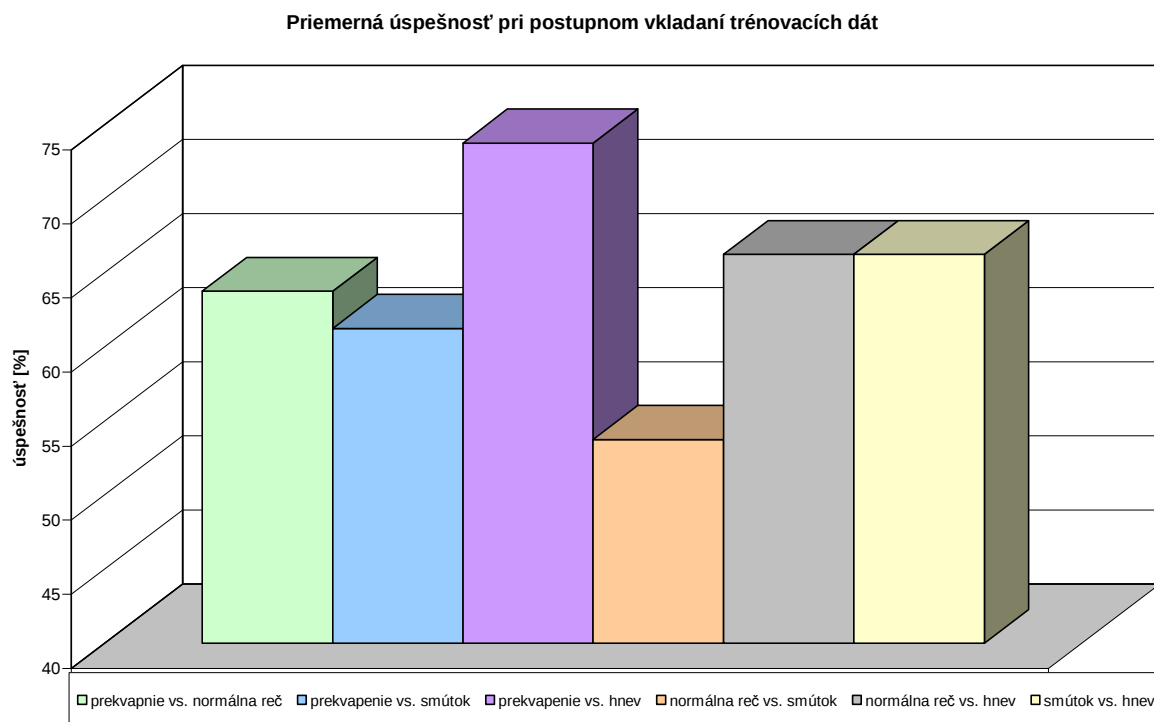


e)

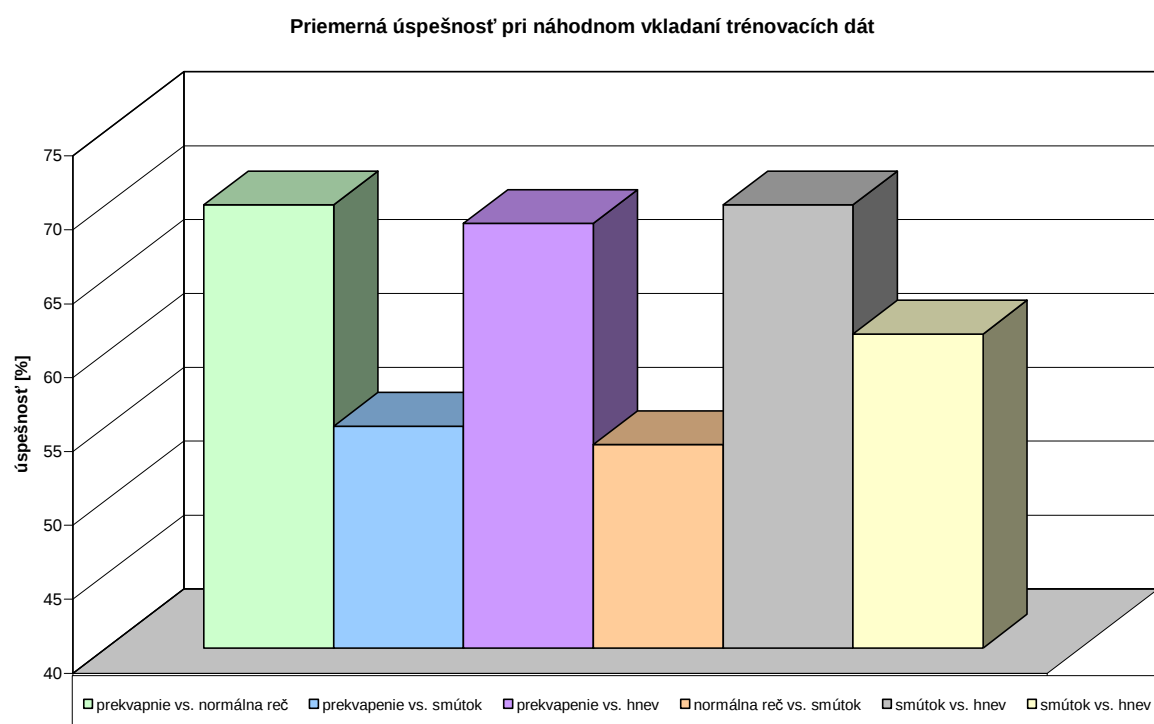


f)

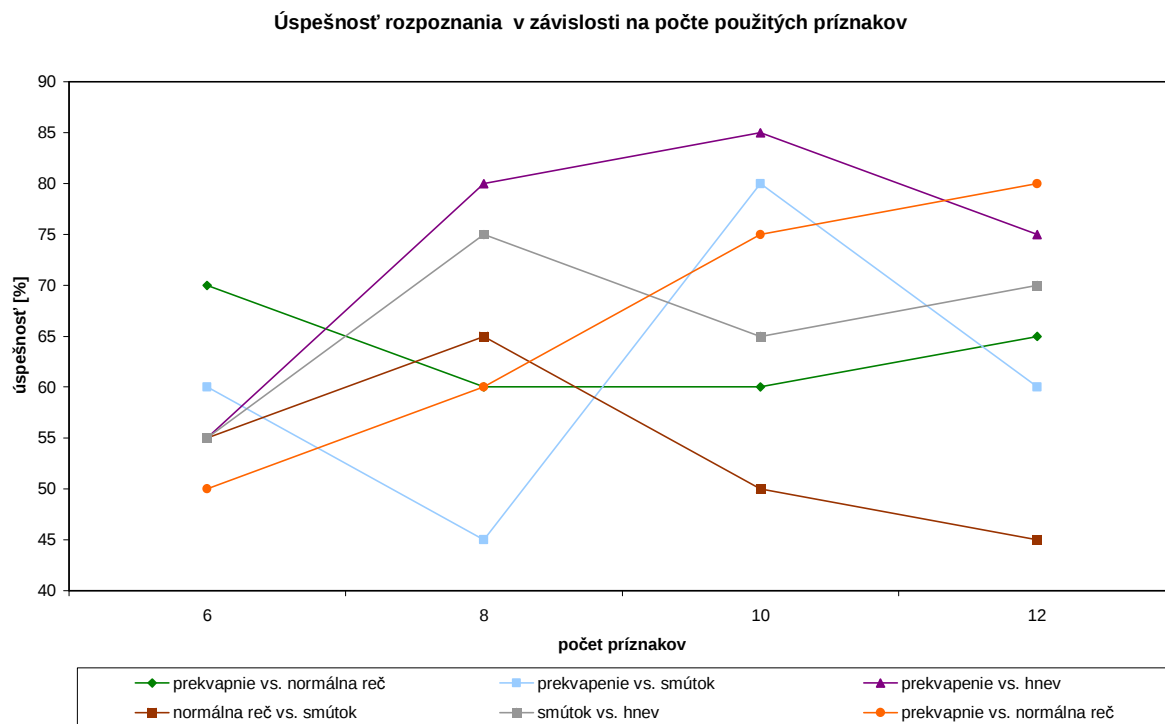
Obr.8.4: Grafy: a) prekvapenie vs. normálna reč, b) prekvapenie vs. smútok, c) prekvapenie vs. hnev, d) normálna reč vs. smútok, e) normálna reč vs. hnev, f) smútok vs. hnev.



Obr.8.5: Graf: Priemerná úspešnosť rozpoznania emócie, pri postupnom vkladaní tréningových dát.



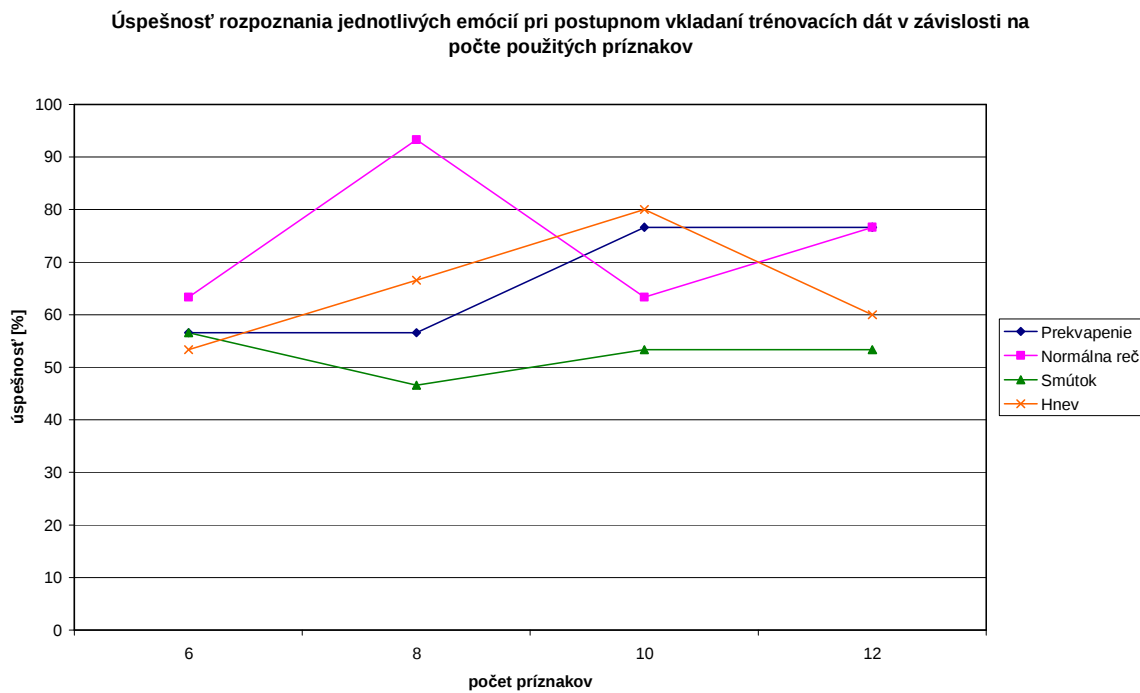
Obr.8.6: Graf: Priemerná úspešnosť rozpoznania emócie, pri náhodnom vkladaní tréningových dát.



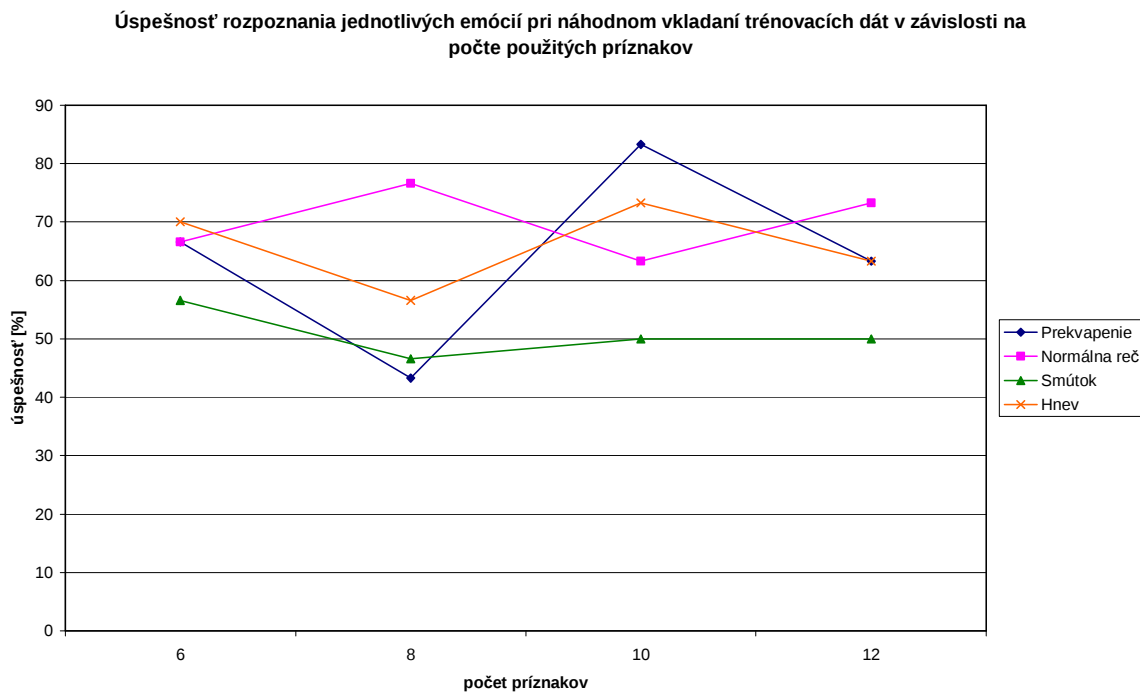
Obr.8.7: Graf: Úspešnosť rozpoznania v závislosti na počte použitých príznakov.

Počet použitých príznakov	Náhodný vstup tréningových dát	Prekvapenie	Normálna reč	Smútok	Hnev	Priemer
		[%]	[%]	[%]	[%]	[%]
6	Nie	56.6	63.3	56.6	53.3	57.450
	Ano	66.6	66.6	56.6	70	64.950
8	Nie	56.6	93.3	46.6	66.6	65.775
	Ano	43.3	76.6	46.6	56.6	55.775
10	Nie	76.6	63.3	53.3	80	68.300
	Ano	83.3	63.3	50	73.3	67.475
12	Nie	76.6	76.6	53.3	60	66.625
	Ano	63.3	73.3	50	63.3	62.475
Priemer	Nie	66.6	74.125	52.45	64.975	64.538
Priemer	Ano	64.125	69.95	50.8	65.8	62.669

Tab.8.2: Úspešnosť rozpoznania jednotlivých emočných stavov.

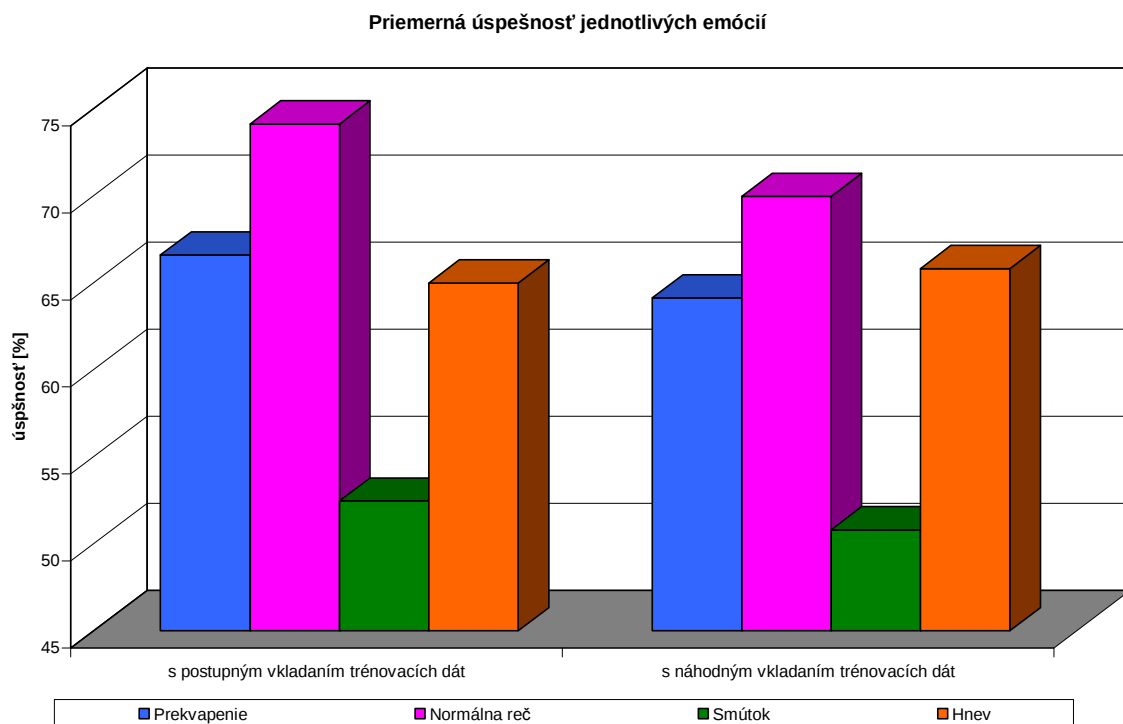


a)



b)

Obr.8.8: Grafy: a) Úspešnosť rozpoznania jednotlivých emócií pri postupnom vkladaní trénovacích dát v závislosti na počte použitých príznakov, b) Úspešnosť rozpoznania jednotlivých emócií pri náhodnom vkladaní trénovacích dát v závislosti na počte použitých príznakov.



Obr.8.9: Priemerná úspešnosť rozpoznania jednotlivých emočných stavov.

ZÁVER

Bakalárska práca sa zaoberala rozpoznávaním emočných stavov človeka od úplného počiatku, až po samotné rozpoznanie.

Proces rozpoznania emočných stavov pozostáva z vytvorenia databázy, jej analýzy programom Praat. Za pomoci tohoto programu sme z jednotlivých vzorkov databázy boli schopný získať hodnoty zvolených príznakov. Nasledovalo vybratie vhodných príznakov, použiteľných pri rozpoznávaní. V tejto práci sme pre každú dvojicu emócií vyberali zvlášť vhodné príznaky za pomoci funkcie Rank features, ktorá je súčasťou programu Matlab. Pri rozpoznávaní sme použili iba časť z pôvodných 18 príznakov a to konkrétne 6, 8, 10, a 12 najvhodnejších.

Samotné rozpoznávanie prebiehalo pomocou simulátoru JavaNNS (Java Neural Network Simulator). V tomto simulátore sme vytvorili 4 architektúry neurónových sietí, ktoré sme potom postupne trénovali metódou s učiteľom, na rozpoznávanie dvojíc emócií. Metóda trénovania siete s učiteľom pozostáva z trénovacích dát, ku ktorým sú priradené správne výstupné hodnoty. Po natrénovaní jednotlivých sietí sme testovali sieť za pomoci 20% množstva trénovacích vzorkov.

Výsledky rozpoznávania emočných stavov sú zaznamenané v tabuľkách a grafoch v 8. lekcii práce. Z testov jednotlivých neurónových sietí vyplýva, že najlepšie rozpoznateľné emócie navzájom od seba sú pri postupnom vkladaní trénovacích dát prekvapenie vs. hnev so 73,75% úspešnosťou a pri náhodnom vkladaní trénovacích dát prekvapenie vs. normálna reč a normálna reč vs. hnev so 70% úspešnosťou. Najhoršie rozpoznateľná dvojica bola v oboch prípadoch normálna reč vs. smútok 53,75% úspešnosťou.

Pri sledovaní úspešnosti rozpoznania jednotlivých emócií sa ako najlepšie rozpoznateľná ukázala v oboch prípadoch normálna reč so 74,125%. A najhoršie rozpoznateľný v oboch prípadoch smútok s úspešnosťou iba 50,8%.

Pri použití rôzneho počtu príznakov sa ako najvhodnejší počet príznakov ukázalo 10 príznakov kedy bola priemerná úspešnosť rozpoznania jednotlivých emócií 68,3% pri postupnom vkladaní trénovacích dát a 67,475% pri náhodnom vkladaní trénovacích dát.

V našom prípade sa ako úspešnejšie ukázalo postupné vkladanie dát s priemernou úspešnosťou 64,538%.

Kompletné tabuľky a grafy so všetkými hodnotami príznakov, vytvorené neurónové siete, trénovacie a testovacie súbory ako aj pomocný program vytvorený na tvorenie vstupných súborov s koncovkou .pat sú priložené na CD.

POUŽITÁ LITERATÚRA

- [1] VLČKOVÁ-MEJVALDOVÁ, Jana. *Prozodie, cesta i mříž porozumění*. Praha : Karolinum, 2006. 203 s. ISBN 80-246-1266-6.
- [2] PSUTKA, Josef , et al. *Mluvíme s počítačem česky*. Praha : Academia, 2006. 746 s. ISBN 8020013091 .
- [3] RENDEK, Tomáš. *Multijazyková analýza emočního stavu mluvčího*. VUT Brno, 2008. 59 s. Bakalářská práce.
- [4] CHLUMSKÝ, Petr. *Demonstrační aplikace pro podporu kurzu neuronových sítí*. ČVUT Praha, 2008. 41 s. Bakalářská práce.
- [5] ZELL, Andreas, et al. *SNNS -- Stuttgart Neural Network Simulator (User Manual, Version 4.2)* [online]. University of Stuttgart : \$2001-2005 , 22. 12. 2004 [cit. 2009-05-25]. Dostupný z WWW: <<http://www-ra.informatik.uni-tuebingen.de/downloads/SNNS/SNNSv4.2.Manual.pdf>>.
- [6] ZELL, Andreas, et al. *JavaNNS -- Java Neural Network Simulator (User Manual, Version 1.1)* [online]. University of Stuttgart, [cit. 2009-05-25]. Dostupný z WWW: <<http://www.ra.cs.uni-tuebingen.de/software/JavaNNS/manual/JavaNNS-manual.pdf>>.

ZOZNAM SKRATIEK

b	bežná reč
d	radosť
F	frekvencia
F_1	1. formant
F_2	2. formant
F_3	3. formant
F_{b1}	šírka 1. formantu
F_{b2}	šírka 2. formantu
F_{b3}	šírka 3. formantu
F_{0max}	maximálna hodnota základnej frekvencie
F_{0min}	minimálna hodnota základnej frekvencie
F_{0std}	smerodatná odchýlka základnej frekvencie
F_{0str}	stredná hodnota základnej frekvencie
H_{max}	maximálna hodnota harmonicity
H_{min}	minimálna hodnota harmonicity
H_{std}	smerodatná odchýlka harmonicity
H_{str}	stredná hodnota harmonicity
I_{max}	maximálna hodnota intenzity
I_{min}	minimálna hodnota intenzity
I_{std}	smerodatná odchýlka intenzity
I_{str}	stredná hodnota hodnota intenzity
k	prekvpenie
m	smiech
p	plač
r	radosť
rz	rozrušenie
s	smútok
v	hnev

ZOZNAM TABULIEK

Tab. 4.1 Ukážka časového delenia emočných stavov v seriáli.	16
Tab. 5.1 Tabuľka získavaných príznakov a ich jednotiek.	18
Tab. 5.2 Priemerné hodnoty príznakov pre jednotlivé emócie.	19
Tab. 5.3 Tabuľka hodnôt v okolí ktorých je najčastejší výskyt jednotlivých veličín vo vzorkoch.....	28
Tab. 6.1 Tabuľka zoradenia vhodnosti príznakov funkciou Rank features.....	31
Tab. 8.1 Prehľad úspešností rozpoznania emočných stavov v jednotlivých kombináciách.....	35
Tab. 8.2 Úspešnosť rozpoznania jednotlivých emočných stavov.....	38

ZOZNAM OBRÁZKOV

Obr. 2.1	Hlasový trakt.....	8
Obr. 2.2	a) Hlasivky v kľúde b) Napäté hlasivky.....	9
Obr. 2.3	Zobrazenie prvých troch formantov v slove "pravda" programom Praat.....	9
Obr. 2.4	Spektrogram slova "pravda" vytvorený programom Praat.....	10
Obr. 3.1	Nosná frekvencia v bežnej reči, zobrazená programom Praat.....	13
Obr. 3.2	Intenzita bežnej reči, zobrazená programom Praat.....	14
Obr. 4.1	Ukážka databázy uloženej na disku.....	17
Obr. 5.1	Grafy: a) stredná hodnota základnej frekvencie, b) stredná hodnota intenzity, c) stredná hodnota harmonicity, d) hodnoty 1. formantu, e) hodnoty 2. formantu, f) hodnoty 3. formantu.....	19
Obr. 5.2	Zobrazenie počtu výskytov jednotlivých príznakov v histogramoch.....	24
Obr. 7.1	Model neurónu.....	28
Obr. 7.2	Ukážka načítanej siete v SNNS.....	29
Obr. 8.1	JavaNNS pri trénovaní neurónovej siete, so zobrazením Error grafu a Log listu.....	32
Obr. 8.2	Ukážka architektúry neurónovej siete so šiestimi vstupnými príznakmi.....	33
Obr. 8.3:	Ukážka vytvoreného programu pre zjednodušenie prevodu .csv do .pat.....	34
Obr. 8.4	Grafy: a) prekvapenie vs. normálna reč, b) prekvapenie vs. smútok, c) prekvapenie vs. hnev, d) normálna reč vs. smútok, e) normálna reč vs. hnev, f) smútok vs. hnev.....	36
Obr. 8.5	Graf: Priemerná úspešnosť rozpoznania emócie, pri postupnom vkladaní trénovacích dát.....	36
Obr. 8.6	Graf: Priemerná úspešnosť rozpoznania emócie, pri náhodnom vkladaní trénovacích dát.....	37
Obr. 8.7	Graf: Úspešnosť rozpoznania v závislosti na počte použitých príznakov.....	37
Obr. 8.8	Grafy: a) Úspešnosť rozpoznania jednotlivých emócií pri postupnom vkladaní trénovacích dát v závislosti na počte použitých príznakov, b) Úspešnosť rozpoznania jednotlivých emócií pri náhodnom vkladaní trénovacích dát v závislosti na počte použitých príznakov.....	39
Obr. 8.9	Priemerná úspešnosť rozpoznania jednotlivých emočných stavov.....	39