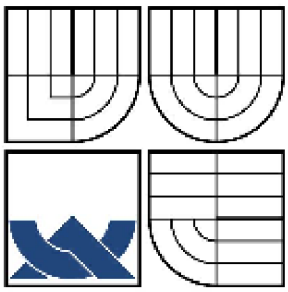
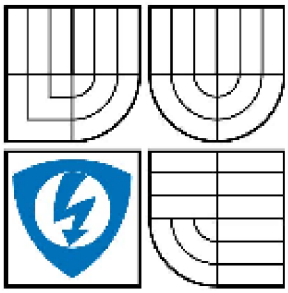


VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A KOMUNIKACNÍCH TECHNOLOGIÍ

ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

DEPARTMENT OF TELECOMMUNICATIONS

ROZPOZNÁVÁNÍ EMOČNÍCH STAVŮ NA ZÁKLADĚ ANALÝZY ŘEČOVÉHO SIGNÁLU

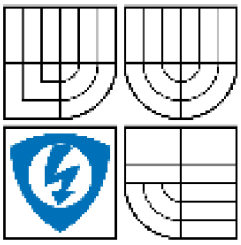
RECOGNITION OF EMOTION STATES BASED ON UTTEREANCE ANALYSIS

DIPLOMOVÁ PRÁCE
MASTER´S THESIS

AUTOR PRÁCE — Bc. JAN ČERMÁK
AUTHOR

VEDOUCÍ PRÁCE — Prof. Ing. ZDENĚK SMÉKAL, CSc.
SUPERVISOR

BRNO 2009



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky a komunikačních technologií

Ústav telekomunikací

Diplomová práce

magisterský navazující studijní obor
Telekomunikační a informační technika

Student:	Bc. Jan Čermák	***ID:***	89162
Ročník:	2	***Akademický rok:***	2008/2009

NÁZEV TÉMATU:

Rozpoznávání emočních stavu na základě analýzy řečového signálu

POKYNY PRO VYPRACOVÁNÍ:

Na základě vyhledání vhodných parametrů, které efektivně a komplexně popisují vlastnosti řečového signálu je možné rozpoznávat emoční stavy člověka, které se projeví ve změně základního tónu řeči, v prozodii, v mikrointonaci, přítomnosti chvění hlasu apod. Cílem projektu je nalézt vhodnou metodu, která bude nejprve poloautomaticky a nakonec automaticky rozpoznávat emoční stavy člověka s vysokou spolehlivostí.

DOPORUČENÁ LITERATURA:

[1] PSUTKA, J., MULLER, L., MATOUŠEK, J., RADOVÁ, V.: Mluvíme s počítačem česky. ACADEMIA, Praha 2006. ISBN 80-2100-1309-1
[2] SYROVÝ, V.: Hudební akustika. Akademie múzických umění, Praha 2003. ISBN 80-7331-901-2
[3] KRCMOVÁ, M.: Fonetika. Elektronické texty. Masarykova Univerzita, Brno 2003. http://is.muni.cz/do/1499/el/estud/ff/js07/fonetika/materialy/index.html

Termín zadání:	9.2.2009	***Termín odevzdání:***	26.5.2009

Vedoucí práce: prof. Ing. Zdeněk Smékal, CSc.

prof. Ing. Kamil Vrba, CSc.
předseda oborové rady

UPOZORNĚNÍ:

ANOTACE

Obsah této práce je zaměřen na klasifikaci emočních stavů s použitím neuronových sítí a klasifikátoru na bázi směsi Gaussových hustotních funkcí s využitím programu Matlab. Pojednává o problematice zpracování řečového signálu, z něhož byly extrahovány prozodické, spektrální příznaky a MFCC koeficienty. Práce se zabývá určením kvality jednotlivých příznaků a výběrem nejvhodnějších pro správnou klasifikaci emočních stavů. Pro určení emočních stavů byly použity dvě rozdílné metody. První metodou jsou neuronové sítě s různě zvolenými parametry. Druhou metodou klasifikace je použití smíšených Gaussových modelů tzv. GMM. U obou metod byla pro klasifikaci rozdělena databáze emočních promluv na trénovací a testovací skupinu. Při testování byla využita metoda nezávislá na mluvčím. Součástí práce je porovnání jednotlivých analyzovaných postupů, zobrazení a porovnání výsledků. Závěrem práce je návrh nejvhodnějších parametrů a klasifikátoru pro rozpoznání emočního stavu mluvčího.

Klíčová slova: klasifikace emočních stavů, neuronová síť, GMM, prozodie, příznaky, MFCC koeficienty, trénovací skupina, testovací skupina, Matlab.

ABSTRACT

The thesis is focused on the emotional states classification in the Matlab program, using neural networks and the classifier which is based on a combination of Gaussian density functions. It deals with the speech signal processing; the prosodic and spectral signs and the MFCC coefficients were extracted from the signal. The work also deals with the quality evaluation of individual signs of which the most suitable were chosen in order to provide the correct classification of emotional states. In order to identify the emotional states, two different methods were used. The first method of classification was the use of neural networks with differently selected parameters, and the second method was the use of the Gaussian mixture model (GMM). In both methods, a database of emotional utterances was divided into the training group and the test group. The testing was based on a method independent of the speaker. The work also includes the comparison of individual analyzed methods as well as the representation and comparison of the results. The conclusion comprises a proposition for the best parameters and the best classifier for the recognition of the speaker's emotional state.

Keywords: classification of emotional states, neural network, GMM, prosody, prosodic signs, spectral signs, MFCC coefficients, training group, test group, Matlab.

ČERMÁK, J. *Rozpoznávání emočních stavů na základě analýzy řečového signálu.* Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2009. 66 s. Vedoucí diplomové práce prof. Ing. Zdeněk Smékal, CSc.

PROHLÁŠENÍ

Prohlašuji, že svou diplomovou práci na téma „Rozpoznávání emočních stavu na nákladě analýzy řečového signálu“ 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.

V Brně dne…….. ……..

podpis autora

PODĚKOVÁNÍ

Děkuji konzultantovi diplomové práce Ing. Hichamovi Atassimu, za velmi užitečnou metodickou pomoc a cenné rady při zpracování diplomové práce.

V Brně dne…….. ……..

podpis autora

OBSAH

1 ÚVOD

Komunikace prostřednictvím mluvené řeči je základní a nejdůležitější prostředek přenosu informace mezi lidmi. Díky řeči je člověk schopen vyjádřit různé myšlenky, nápady, pocity a emoce. Umění mluvit a rozumět se učíme už jako děti a od té doby tuto schopnost považujeme za samozřejmou činnost. Jedná se však o velice složitou posloupnost akcí, jejichž úspěšné zvládnutí má vliv na průběh celé komunikace. Přenos informace obvykle začíná přípravou zprávy v mozku řečníka, pokračuje přenosem zprávy (tj. vlastní realizací promluvy akustickým řečovým signálem) a končí rozpoznáním akustického signálu posluchačem, včetně porozumění významu přenášené zprávy.

V dnešní době se klade stále větší důraz na výzkum v oblasti řečových signálů. Nové poznatky a samotný rozvoj techniky tak v různých oblastech otevřely cestu využití počítačových analýz řečového signálu (např. ve zdravotnictví, rozpoznávání řečových vad, rozpoznávání identity mluvčího, různé bezpečnostní prvky, aplikace potřebné k úpravě hlasu do přirozenější podoby).

Komunikace na dálku je již neodmyslitelnou součástí našeho života. Při dnešním rozvoji digitální techniky je řečový signál pomocí A/D převodníků převáděn do číslicové formy a nese s sebou mnoho druhů informací. Faktická informace je nositelkou myšlenky a podstatou komunikace samotné. Další, na první pohled podružnou, je informace emoční. Vyjadřuje buď okamžitý emoční stav mluvčího, nebo jeho emoční postoj. Emoční postoj je emoce, kterou v tomto případě mluvčí vědomě předává dále. Emoce jsou vnímány z prozodie řeči.

Rozpoznávání řeči stále více proniká do zařízení běžných potřeb ve všech možných odvětvích. Stále více uplatnění nalézají také poznatky o vlastnostech řeči v nenormálním emočním stavu (např. vlivem stresu, únavy, apod.). V oblasti rozpoznávání emocí se střetáváme s problémem nekvalitního materiálu pro samotný výzkum. Ten vychází z těžkostí získávání spontánního materiálu, který je však základem úspěchu výzkumu. Samotný výzkum se zaměřuje na získání příznaků, které by co nejlépe dokázaly charakterizovat emoční stav člověka. V práci jsme se zaměřili na analýzu řečových signálů v oblasti emocí. Samotná práce rozebírá vlastnosti řečového signálu a na řeč se dívá z pohledu technického i pohledu prozodického.

2 PROZODIE

2.1 Zvuková stránka souvislé řeči

Souvislá řeč není monotónní. Je modulována pomocí síly a výšky hlasu a získává určitý

rytmus díky proměnlivému časovému průběhu jednotlivých segmentů a jejich kombinací. Dalšími prostředky modulace je řečové tempo a existence různých typů pauz. Tyto zvukové prostředky se uplatňují na promluvě jako celku. Prostředky modulace jsou jednak přirozenou složkou zvukového signálu řeči (síla hlasu nejen přirozeně existuje, ale mění se i fyziologicky v průběhu řeči), jednak mají i komunikační funkce - zejména při vyjadřování pragmatických složek komunikace. Jsou velmi důležité, protože pouhá nápodoba hlásek či slabik bez náležité modulace je velmi těžko srozumitelná [KRČ-07].

2.2 Ústrojí hlasové (fonační)

Ústrojí hlasové (fonační) je uloženo v hrtanu. Je typicky lidským orgánem a jeho funkcí je vytvářet základní hlas, jehož dalšími úpravami vzniká hlasitá řeč. Základem hlasového ústrojí jsou dva hlasové valy pokryté sliznicí, tzv. hlasivky. Lidský hlas je (stejně jako jiné přírodní zvuky tónového charakteru) periodickým zhušťováním a zřeďováním vzduchu.

Základní vlastnosti lidského hlasu jsou dány fyziologií hlasového ústrojí. Význam má především délka hlasivek (u žen se udává pro soprán 14-19 mm, u mužů pro bas 24-25 mm); čím jsou hlasivky kratší, tím rychleji kmitají a běžný mluvní hlas je vyšší. Obvyklá výška hlasu jedince vychází z těchto anatomických předpokladů. Intenzita práce hlasivek je mimořádně velká, u velmi hlubokého hlasu jde sice jen o 50 Hz (kmitů za sekundu), ženské vysoké hlasy však dosahují až 480 Hz a při zpěvu může být počet kmitů daleko větší. Pro kojenecký věk se uvádí 400 Hz, což je poměrně vysoké.

Hlas vycházející z hlasivek nemá barvu lidského hlasu**.** Charakteristické znění, individuální pro jednotlivce, získává až průchodem nadhrtanovými prostorami - rezonátory, v nichž se mění některé ze svrchních tónů základního hlasu. Na konečném efektu se podílí i rezonance celé lebeční dutiny a lícních kostí.

Funkce hlasu při řeči spočívá především v tom, že vytváří základní tón řeči. Mluvní projev bez účasti hlasu je možný (např. šeptání), je však srozumitelný jen na malou vzdálenost. Činností hlasivek vzniká hlasnost: hlásky se pak odlišují podle toho, zda u nich dochází k chvění hlasivek, nebo ne. Hlasnost se označuje termínem znělost (sonorita), který odráží výsledný auditivní dojem hlásky. S tímto termínem se pracuje jak na fonetické úrovni (rozlišují se hlásky znělé a neznělé), tak na úrovni fonologické (znělostní rozdíl má v mnoha jazycích rozlišovací platnost) [KRČ-07].

2.3 Ústrojí modifikující

Ústrojí modifikující, jinak artikulační v užším smyslu slova, je uloženo nad hrtanem. Skládá se ze tří dutin : dutiny hrdelní, dutiny nosní a dutiny ústní.

Dutina hrdelní (laryngální) se rozkládá bezprostředně nad hlasivkami a končí (z fonetického hlediska) v místech, kde je jazyk při artikulaci nejblíže patru. Změny objemu hrdelní dutiny se uplatňují zvláště při tvoření vokálů. Vedle toho má při komunikaci funkci změna znění řeči závislá na sevření krčních a hrdelních svalů. Posluchač vnímá psychický

stav mluvčího, např. jeho nervozitu, která k napětí vedla, a to i v jazyce, kterému jinak nerozumí.

Dutina nosní (nazální) je také rezonančním prostorem, ten se však využívá jen u části hlásek.

Dutina ústní (orální) je vpředu ohraničena rty, vzadu přechází do dutiny hrdelní. Objem tohoto prostoru je proměnlivý - pozměňuje se jak pohybem jazyka, tak pohybem rtů a čelistí; spolu s proměnami hrdelní dutiny je tato změna využívána pro vytváření vokálů, spolupůsobí však i při tvoření konsonantů: přehrada v ústní dutině, jíž se konsonant tvoří, člení prostor úst a důsledkem je pak charakteristická výška šumu.

Pohybem jazyka se mění vzájemný poměr rezonátorů; ve výsledném znění se odrážejí i proměny výstupního otvoru rezonátorů dané postavením rtů. V této složité rezonanční soustavě dochází ke vzniku formantů, typických pásem zesílení zvukové energie. Pro vytvoření a identifikaci vokálů jsou nutné nejméně dva formanty, samohlásky lidské řeči jich však mají více (uvádí se až 6 formantů, k nim přibývají i další -svrchní tóny). Zvukové vlny vzniklé činností hlasivek podle ní rozechvívají sloupec vzduchu v hrdelní a ústní dutině. Tím vznikají oba základní formanty.

Základní tón řeči se naproti tomu může obměňovat. Proto lze i ve zpívaném textu rozeznávat jednotlivé samohlásky. Při extrémně vysokých polohách však může dojít u vokálů ke zkreslení. Při šeptaných samohláskách laryngální hlas chybí. Přesto mají i tyto samohlásky formantovou strukturu a jsou dobře poznatelné. Jako "budič charakteristiky" se u nich uplatňuje prostý výdechový proud doplněný šumem vzniklým při průchodu vzduchu mezi částečně sblíženými hlasivkami.

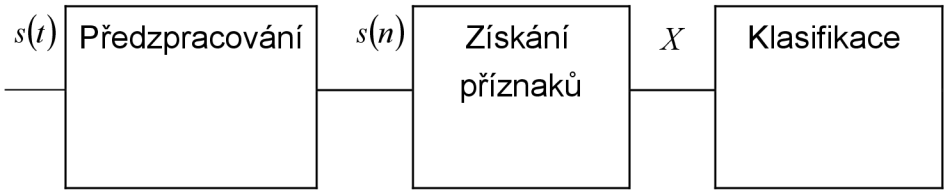
Výška formantů je pro jednotlivé vokály v daném jazyce omezena na jistá frekvenční pásma, nejde tedy o jediný tón. Dnes se dává přednost označení číslicemi vycházejícími z výšky formantů, jak ji zachycují objektivní metody analýzy zvuku řeči. Formant základního hlasu má označení F_0, nejbližší vyšší F_1, další F_2 atd. Např. v češtině se F_1 pohybuje v rozmezí 300 Hz [i:] - 800 Hz [a:], F_2 700 Hz [u:] - 2000 Hz [i:]; čísla jsou průměrná, skutečné realizace se pohybují v širokém rozmezí kolem těchto výšek. Formanty jednotlivých vokálů jsou navzájem odlišeny. Zřetelně se liší rezonance vznikající v ústní dutině, rezonance vznikající v dutině hrdelní jsou odlišeny méně [KRČ-07].

3 ŘEČOVÝ SIGNÁL

Řečový signál se ve všech oblastech zpracování řeči zpracovává výhradně v číslicové podobě. Využívají se k tomu s výhodou výkonné algoritmy a toolboxy, které jsou zejména od osmdesátých let k dispozici pro číslicové zpracování signálů.

Celý proces automatického rozpoznávání signálů můžeme rozdělit na tři základní

stupně podle obr. 3.1 a dále se samostatně zabývat jednotlivými stupni [SIG-00].



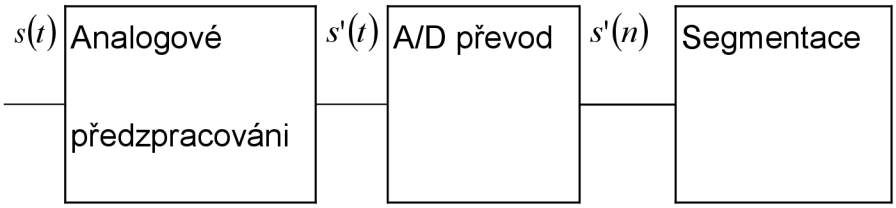
Obr. 3.1 Automatické rozpoznávání řečových signálů

Zpracování řeči začíná blokem předzpracování (pre-processing), který obsahuje digitalizaci a několik standardních operací na úpravu signálu. Přímá klasifikace podle časového průběhu signálu $s(t)$ popř. $s(n)$ není možná vzhledem ke značné variabilitě a velkému počtu vzorků. Proto je nutné z řečového signálu získat jen několik málo důležitých příznaků x, aniž by se přitom ztratily důležité části informace obsažené v signálu. Důležitost informace je dána konečným cílem zpracování, např. při kódování řeči je důležité zachování tvaru signálu, při rozpoznávání mluvčích jsou to naopak příznaky charakteristické pro jednotlivé mluvčí. Pod pojmem příznaky (features) rozumíme vlastnosti obrazce vyjádřené kvantitativně. Signál v této formě převeden ze „signálového prostoru" do „příznakového prostoru" je již připraven k vlastnímu zpracování informace a může být provedena klasifikace podle vektoru příznaků X [SIG-00].

Předzpracování

Lidská řeč a tím i řečový signál jsou značně variabilní. Nikdo není schopen vyslovit jedno slovo dvakrát naprosto stejně, tzn. dodržet stejný přízvuk, výšku tónu, hlasitost a rychlost promluvy. Ještě větší rozdíly jsou mezi různými mluvčími. Podstatný vliv na charakter řečového signálu mají také rušení, okolní zvuky a rovněž zkreslení při přenosu signálu, tzn. kmitočtové charakteristiky mikrofonů, filtrů a zesilovačů a vlastnosti přenosových cest při dálkovém přenosu řeči. Jmenované vlivy snižují celkovou úspěšnost rozpoznávacího procesu. Proto je účelné některé z uvedených vlivů potlačit hned na počátku procesu vhodným předzpracováním.

Číslicové předzpracování



Obr. 3.2 Blokový diagram operací předzpracování řečových signálů

3.1 Analogové předzpracování

Analogové předzpracování, tj. manipulace s řečovým signálem do té doby, než bude prezentován sledem vzorků, začíná převodem změn akustického tlaku na elektrický signál. Při „živém" snímání řeči může vhodný mikrofon zaručit velmi dobrý poměr signál/šum (řeč/zvuky pozadí). Pokud je odstup mikrofonu od zdroje řeči konstantní a velmi malý, pak lze zanedbat vliv akustiky místnosti (okolí).

Takto získaný signál je v rozsahu několika milivoltů a musí být zesílen pokud možno bez šumu a s lineární kmitočtovou závislostí v pracovním pásmu následujících stupňů zpracování. Přitom je nutné dbát na krátká spojovací vedení (použití koaxiálních kabelů). Je rozumné realizovat zesilovací jednotku dvěma stupni: předzesilovačem v přímé blízkosti mikrofonu a dalším stupněm před A/D převodníkem.

Analogový řečový signál $s(t)$ je nutné omezit dolní propustí s ohledem na následné vzorkování. Podle známého vzorkovacího teorému musí být vzorkovací kmitočet f_{vz} minimálně dvakrát vyšší než je kmitočet dolní propusti $f_{dp}\left(2 \cdot f_{vz} \succ f_{dp}\right)$.

Střední úroveň řečového signálu se při normální řeči mění obvykle o několik decibelů v časovém rozmezí několika sekund. Změnou polohy mikrofonu a úst mluvčího lze způsobit opět rozdíl několika decibelů. Mnoho parametrů signálu je závislých na kolísání hlasitosti. Protože tyto efekty nemají fonetický význam, je žádoucí vyrovnávat celkovou intenzitu řečového signálu hned na počátku zpracování ještě v analogové podobě [SIG-00].

3.2 Analogově číslicový převod

Analogově předzpracovaný řečový signál je nyní digitalizován obvykle vzorkováním na kmitočtu 8-22 kHz a kvantováním s rozlišením 8-16 bitů. Parametry digitalizace jsou buď dány zdrojem resp. přenosovou cestou signálu nebo si je při snímání řeči volíme sami s ohledem na účel zpracování řečového signálu. Obecně lze říci, že pro rozpoznávání obsahu řeči nám postačí nižší kvalita digitalizace (8-12 kHz, 8-10 bitů), zatímco při rozpoznávání mluvčích požadujeme kvalitnější číslicový signál. Nejvyšší kmitočtové a modulové rozlišení je vhodné pro rozpoznávání emočních stavů mluvčích a diagnostická vyšetření hlasu.

Při volbě kvantování jsme charakterem vzorkovaného signálu omezeni při stanovení maximálního smysluplného počtu kvantovacích úrovní. Maximální počet kvantizačních úrovní m je podle Shannona dán vztahem:

$$m = k\sqrt{1+\frac{P_s}{P_p}} \tag{3.1}$$

kde P_s je maximální výkon signálu, P_p je střední výkon poruch, k je konstanta typu šumu (pro bílý šum $k=1$).

Protože dynamický rozsah řečového signálu je asi 60 dB, je pro jeho kvalitní převod

(např. pro účely záznamu) zapotřebí B = 11-12 bitů. Jsou-li hodnoty signálu $s(t)$ rozloženy přibližně rovnoměrně v celém intervalu

$$|s(t)| \leq S_{\max} \tag{3.2}$$

nabízí se provést rovnoměrné (uniformní) kvantování celého rozsahu signálu do počtu pásem 2^B, při šířce pásma (kvantizační krok)

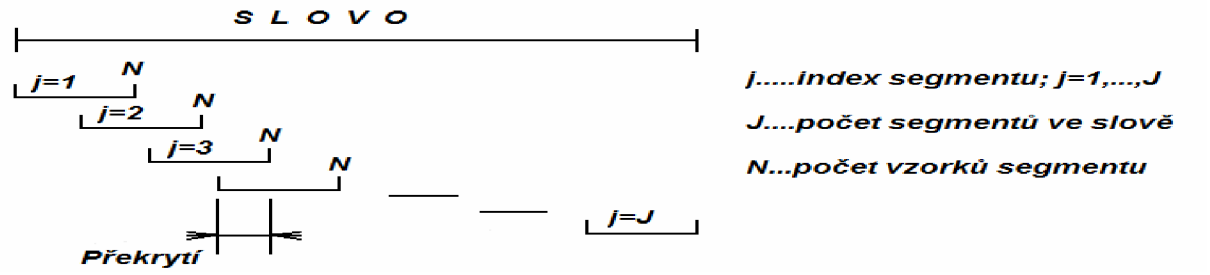
$$\Delta = \frac{2S_{\max}}{2^B} \tag{3.3}$$

Rozložení okamžitých hodnot řečových signálů však spíše připomíná exponenciální průběh kolem střední hodnoty. Informativní účinnost kódování takového signálu může být zvýšena logaritmickým kodérem [SIG-00].

3.3 Segmentace pomocí oken

Vzhledem ke své biologické povaze je řečový signál téměř výhradně zpracováván metodami tzv. krátkodobé analýzy. Tyto metody vycházejí z kvazistacionární podstaty řečového signálu, tj. možnosti přijetí předpokladu, že vlastnosti signálu se v čase mění „pomalu". Signál je za tím účelem rozdělen na ekvidistantní časové úseky - segmenty (frame) o délce N vzorků a každý segment je potom popsán vektorem příznaků.

Délka segmentu musí být na jedné straně dostatečně malá, aby bylo možné naměřené parametry uvnitř segmentu aproximovat konstantními hodnotami a na druhé straně dostatečně velká, aby bylo zaručeno, že požadované parametry budou bezchybně změřeny.



Obr. 3.3 Princip segmentace

Oba protichůdné požadavky jsou vcelku splněny pro úseky řeči dlouhé 10 až 35 ms, což souvisí se změnami nastavení lidského hlasového ústrojí, které probíhají v nejkratším intervalu 10 až 35 ms. U takových segmentů platí přibližně Gaussovo rozložení hustoty pravděpodobnosti okamžité velikosti řečového signálu. Princip segmentace je zobrazen na obr. 3.3. Celé slovo je rozděleno celkem na J segmentů, přičemž všechny segmenty mají stejnou délku odpovídající N vzorkům. Přitom se dva sousední segmenty mohou překrývat. Částečným překrýváním segmentů se dosáhne většího vyhlazení časových průběhů parametrů signálu, ale zpomalí se časový posun a částečně se zvýší výpočetní nároky.

Řečový segment $s(n)$ o N vzorcích může být vytvořen z řečového signálu pomocí váhové posloupnosti tzv. okna $w(n)$, kterým se vybírají resp. váží vzorky $s'(n)$. Matematicky provedeme tento úkon násobením:

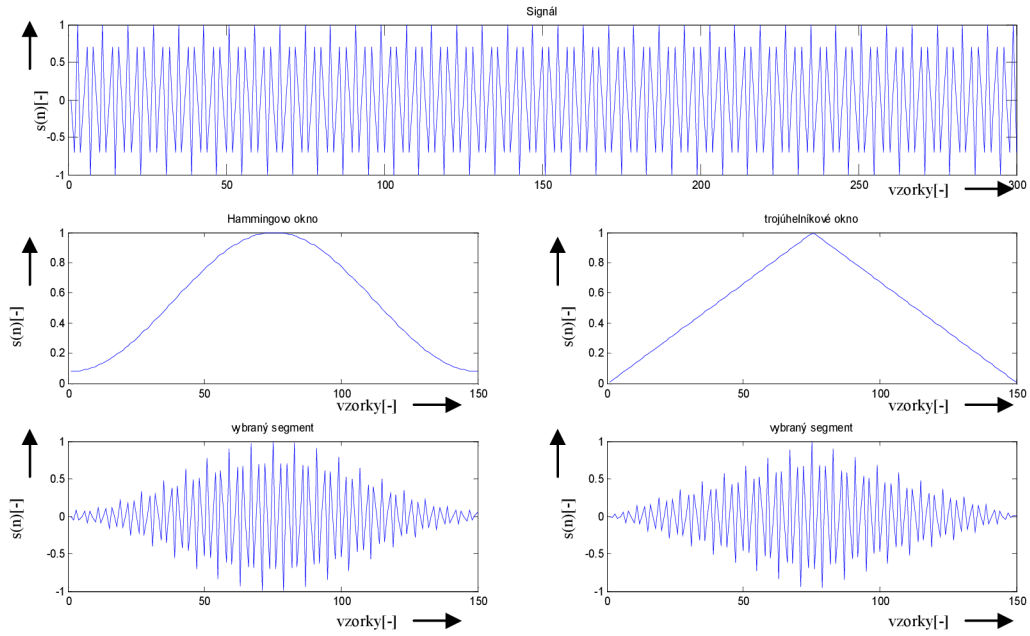
$$s(n) = s'(n)\, w(n) \tag{3.4}$$

Úkolem okna je vybrat po úsecích příslušné vzorky signálu a přidělit jim určitou váhu. Váhová funkce $w(n)$ určuje typ okna. Nejčastěji používanými typy oken při zpracování řečového signálu jsou:

pravoúhlé	$w(n) = 1$	pro $n = 1,2,...,N$
	$w(n) = 0$	pro ostatní n
Hammingovo	$w(n) = 0{,}54 - 0{,}46\cos(2\pi n / N)$	pro $n = 1,2,...,N$
	$w(n) = 0$	pro ostatní n

Tab. 3.1 Váhovací okna

V obou případech je N délka okna a tím současně také délka vybraného segmentu řeči vyjádřena v počtu vzorků. Časový průběh oken a jejich aplikace na řečový signál je znázorněn na obr. 3.4. Přestože pravoúhlé okno je jednodušší, často se upřednostňuje použití Hammingova okna vzhledem k tomu, že potlačuje vzorky na okrajích segmentů, čímž se zvyšuje stabilita některých výpočtů. Zvolené okno se pohybuje po časové ose s krokem N vzorků v případě, že segmenty na sebe navazují, nebo s krokem menším než N vzorků, pokud se segmenty překrývají.



Obr. 3.4 Segment řeči pomocí trojúhelníkového a Hammingova okna

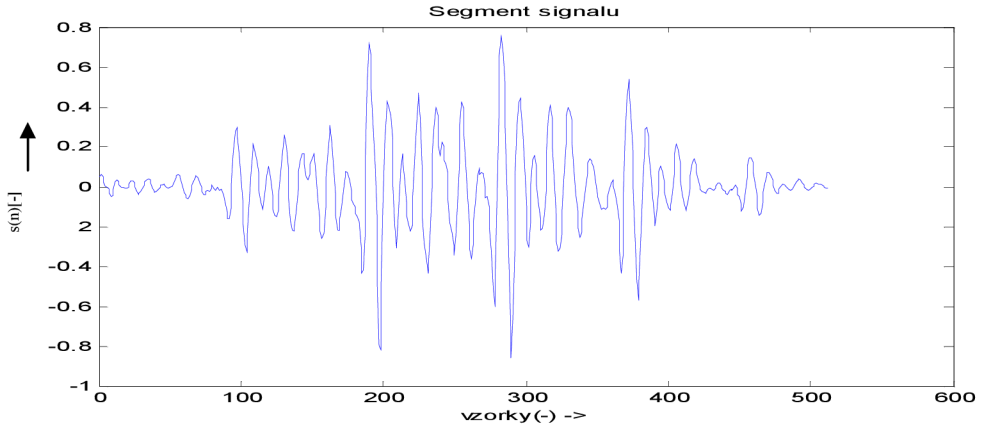
Obě okna v podstatě představují filtr typu dolní propust. Spektrum vybraného segmentu získané Fourierovou transformací reprezentuje výsledek konvoluce skutečného spektra daného úseku řečového signálu se spektrem použité okénkové funkce. V takovém případě je důležité znát, jak vypadá spektrum okénkové funkce a jaké závěry o skutečném

spektru řeči můžeme vyvodit z výsledné konvoluce.

Násobení řečového signálu pravoúhlým oknem vede ke dvěma nežádoucím efektům - rozmazání a rozptylu spektra. Oba efekty souvisejí s tím, že spektrum pravoúhlého okna je tvořeno jedním hlavním lalokem a větším množstvím vedlejších laloků. Konvolucí spektra okna se spektrem signálu se jediná spektrální čára ve spektru signálu rozšíří (rozmaže) na tvar hlavního laloku. Šířka hlavního laloku tak určuje kmitočtové rozlišení DFT a pro délku okna NT_{vz} je dána vztahem $2/NT_{vz}$, kde T_{vz} je vzorkovací perioda. Znamená to, že chceme-li dosáhnout velkého spektrálního rozlišení (při stejném vzorkování), musíme volit N co největší. Avšak při dlouhém analyzovaném úseku budou rychlé spektrální změny průměrovány a nemohou být detekovány. Druhý nežádoucí efekt (rozptyl spektra) je způsoben vedlejšími laloky ve spektru okna a projevuje se tím, že ve spektru navzorkovaného řečového signálu se objeví nové spektrální čáry vně hlavního laloku. Tento efekt nelze potlačit změnou délky okna, můžeme ho ovlivnit pouze tvarem okna. U pravoúhlého okna je výška prvního vedlejšího laloku 13 dB pod maximem hlavního laloku.

U řečového signálu (zejména v jeho znělých úsecích) se vyskytují rozdíly mezi nejsilnějšími a nejslabšími kmitočtovými komponenty více než 40 dB. Použitím pravoúhlého okna nemohou být slabé komponenty ve spektru signálu vůbec postihnuty. Řešením tohoto problému je použití jiného vhodnějšího typu okna, obvykle Hammingova. Toto okno má sice ve spektru zhruba dvojnásobně široký hlavní lalok, ovšem útlum vedlejších laloků 43 dB je podstatně lepší [SIG-00].

V našem případě byla každá jednotlivá nahrávka váhována Hammingovým oknem o délce 512 vzorků, což při vzorkovacím kmitočtu nahrávek f_{vz} = 16000Hz odpovídá délce 32ms. Tato krátká délka segmentů je vhodná pro všechny počítané příznaky. Přesah segmentů byl zvolen 50%.



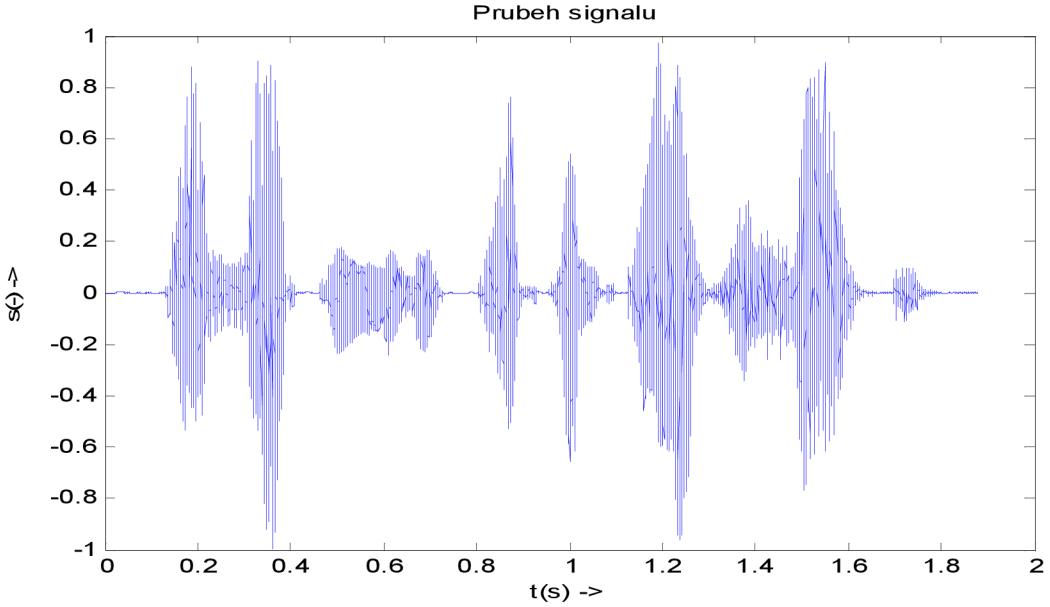
Obr. 3.5 Jeden segment signálu váhovaný Hammingovým oknem

4 PARAMETRY ŘEČOVÉHO SIGNÁLU

4.1 Databáze

Vzhledem k časové náročnosti získat použitelnou databázi emočních nahrávek, tzn. databázi, která by obsahovala alespoň desítky nahrávek pro každou emoci a to od obou pohlaví, byla pro tuto práci použita německá databáze emocí: Berlin Database of Emotional Speech <http://pascal.kgw.tu-berlin.de/emodb/index-1280.html>. Obsahuje 7 emočních stavů od 5ti mluvčích mužů a od 5ti mluvčích žen. Tvořena je z 10ti odlišných obsahů.

Vezměme například řečový signál, který není zabarven žádnou emocí a je tedy neutrálního charakteru. Tato příkladová promluva bude rozdělena na 117 segmentů o délce 32ms a každý segment váhován Hammingovým oknem. Následně jsou pro každý segment počítány všechny vypsané parametry, též nazývané příznaky.



Obr. 4.1 Průběh signálu neutrální promluvy

4.2 Střední počet průchodů signálu nulovou rovinou (ZCR)

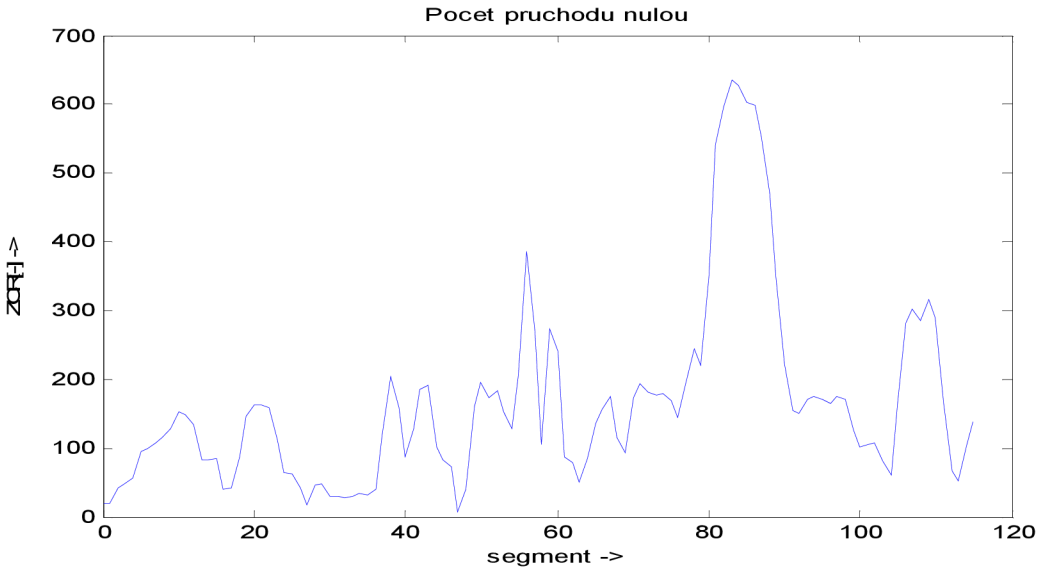
Díky tomuto parametru můžeme rozlišit zda je úsek promluvy řečí nebo šumem. Tento parametr je počítán pro každý rámec řeči. Předpokládá se, že před rozdělením do rámců je signál předzpracován metodami uvedenými v kapitole 3. Rámec řeči je označen $x(n)$, kde $0 \le n \le l_{seg} - 1$. Pro programování v Matlabu je vhodnější indexovat od 1 do l_{seg}.

$$ZCR = \sum_{n=0}^{l_{seg}-1} \left| sign\ x(n) - sign\ x(n-1) \right|, \tag{4.1}$$

kde znaménková funkce "*sign*" je definována:

$$sign\ x(n) = \begin{cases} +1\ pro\ x(n) \geq 0 \\ -1\ pro\ x(n) < 0 \end{cases} \tag{4.2}$$

ZCR je počítán jako součet jednotlivých průchodů nulovou rovinou(osa x) při prohledávání segmentu přes všechny vzorky.



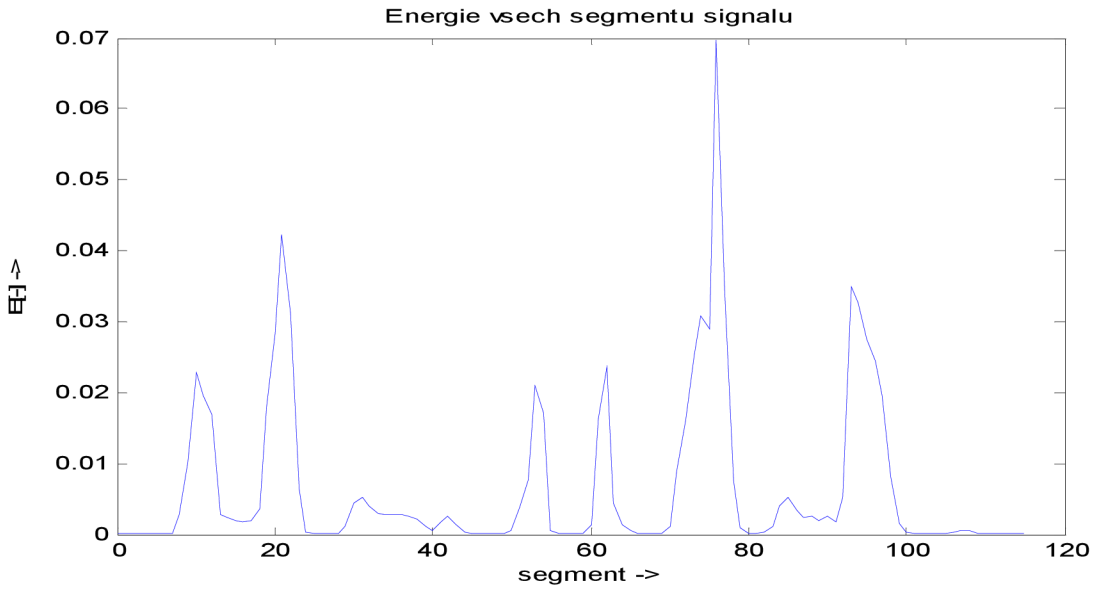
Obr. 4.2 Počet průchodů nulovou rovinou jednotlivých segmentů

4.3 Krátkodobá energie

Pomocí tohoto parametru může rozlišit úseky ticha (nízká energie) od úseků řeči (vysoká energie). Při měření krátkodobé energie je vhodnější volit kratší délku mikrosegmentů 20-40ms. Hodnoty funkce poskytují pro každý mikrosegment informaci o průměrné hodnotě energie v mikrosegmentu. Díky velkému dynamickému rozsahu energie (několik řádů) používáme často její logaritmus.

$$E = \frac{1}{l_{seg}} \sum_{n=0}^{l_{seg}-1} |x(n)|^2 \tag{4.3}$$

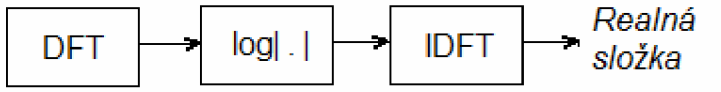
Energie byla získána jako součet kvadrátů jednotlivých vzorků segmentu.



Obr. 4.3 Energie jednotlivých segmentů signálu

4.4 Kepstrum

Model řeči je dán konvolucí buzení a impulsní charakteristiky filtru. Pokud jsou dva signály konvoluovány, je obtížné je získat zpět (provést dekonvoluci). Můžeme se o to pokusit tak, že na určitém místě zavedeme do transformací nelinearitu, která dokáže převést součin na součet. Jednotlivé komponenty součtu pak již lze oddělit.



Obr. 4.4 Postup výpočtu kepstrální analýzy

Komplexní kepstrum buzení sestává z pulsů, které se objevují v intervalech odpovídajících periodě základního hlasivkového tónu. Protože komplexní kepstrum impulsní odezvy hlasového traktu je soustředěno kolem $n = 0$ a komplexním kepstrem buzení jsou pulsy v intervalech úměrných periodě základního hlasivkového tónu, lze kepstrální hodnoty reprezentující hlasivkový trakt extrahovat z úplného kepstra pomocí lineárního systému, ve kterém jsou složky kepstra pro malé hodnoty $|n|$ násobeny hodnotou jedna a ostatní nulou. Postup kepstrální analýzy je na Obr. 4.4. Předzpracovaný signál je přiveden na vstup bloku DFT, jeho výstup přichází na blok $\log|\cdot|$, tento výsledek podrobíme IDFT. Výsledné kepstrum je reálná složka tohoto bloku.

První koeficient kepstra představuje energii signálu. Koeficienty s nízkým pořadím (dolní kvefrence) popisují pomalé změny ve spektru signálu, tzn. formantovou strukturu a tím i charakteristiku hlasového traktu. Koeficienty s vyšším pořadím (horní kvefrence)

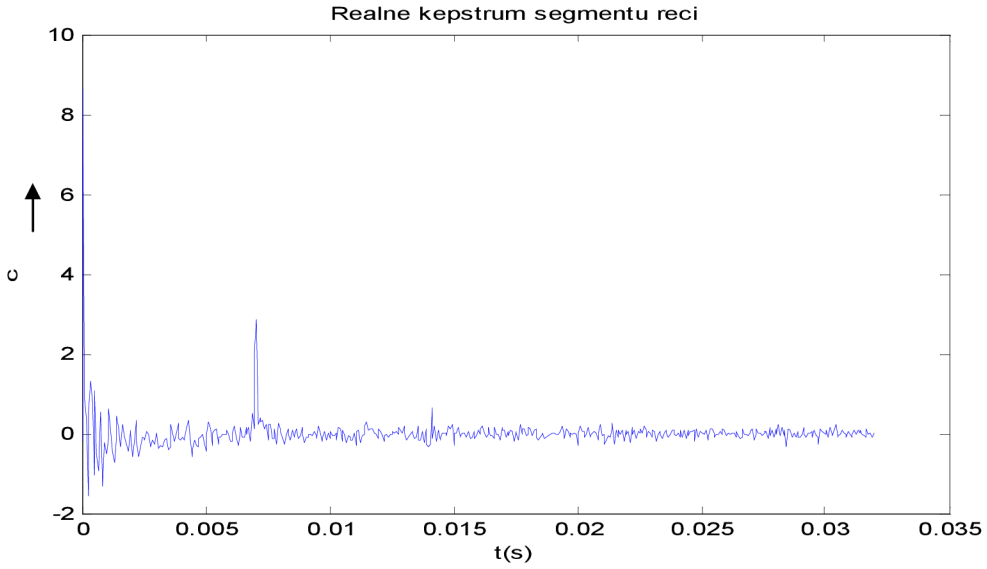
reprezentují rychlé změny ve spektru signálu, čímž specifikují změny buzení hlasového traktu. U znělých úseků řeči se v kepstru vyskytuje výrazná špička, která svou polohou určuje základní tón řeči. Odhad spektrální hustoty výkonu pomocí přímé DFT:

$$c(n) = DFT^{-1}\left\{\ln\left|DFT\left[s(n)\right]\right|^2\right\} \tag{4.4}$$

Diskrétní Fourierova transformace bývá samozřejmě implementována pomocí rychlého algoritmu FFT (Fast Fourier Transform). Jelikož nezávislá proměnná n v $c(n)$ má rozměr diskrétního času (vzorky), říká se tomuto rozměru slovní hříčkou kvefrence. Kepstrum vzniklo přesmyčkou ze slova spektrum.

Pro získání základního tónu řeči (ZTŘ) z kepstra použijeme znělé segmenty signálu. Na kepstrum je aplikován filtr, který odstraní z kepstra charakteristiku hlasového traktu, tzn. prvních 40 vzorků je nulováno. Základní tón je odpovídající kvefrence maximální hodnoty kepstra tohoto segmentu.

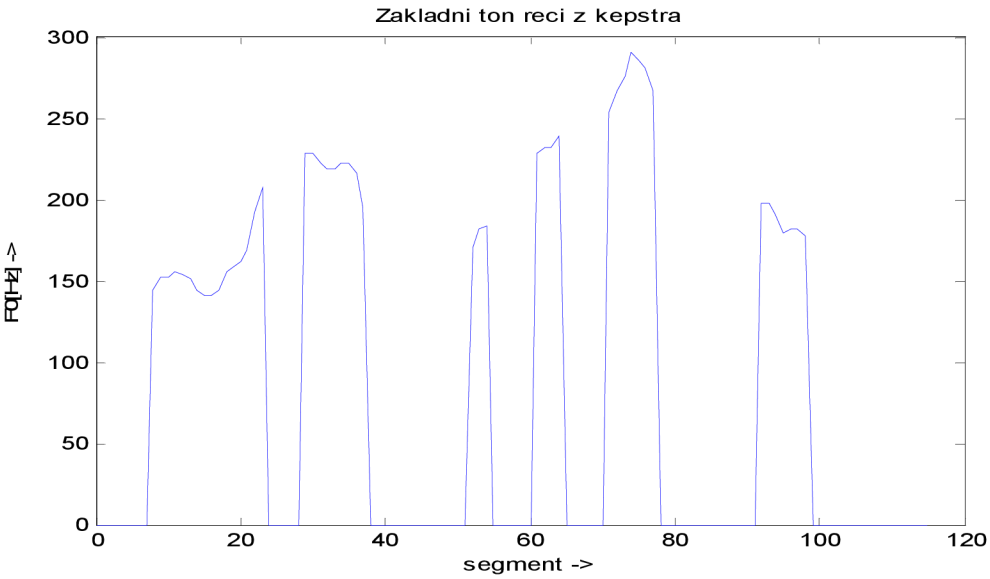
Na segment byla aplikována FFT(Fast Fourier Transformation) o délce 1024 vzorků. Reálná složka této FFT byla logaritmována a tento výsledek byl podroben IFFT(Inverse Fast Fourier Tranformation).



Obr. 4.5 Reálné kepstrum jednoho segmentu řeči

4.4.1 Základní tón řeči počítaný z kepstra (ZTŘ)

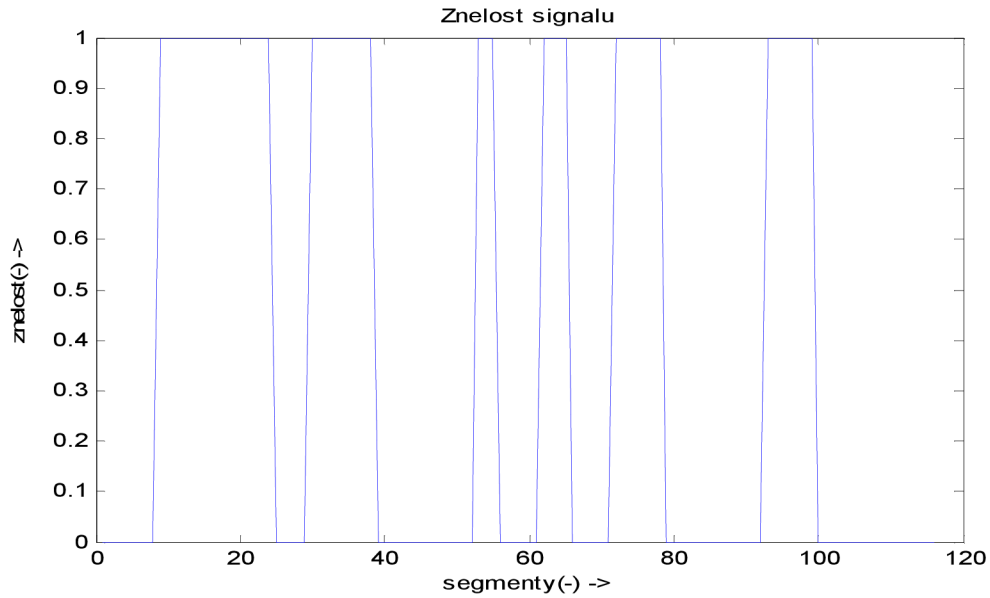
Na vypočtené znělé kepstrum byl použit filtr, který odstraní nízké složky (vysoké kvefrence). Ve zbylém signálu je nalezeno maximum a jeho index je námi hledaný ZTŘ.



Obr. 4.6 Výpočet základního tónu řeči z kepstra pro jednotlivé segmenty

4.5 Znělost

Pomocí poměru energie a počtu průchodů nulovou rovinou je možné hrubě určit, který segment obsahuje jen šum a který je pro nás důležitý řečový signál. Znělý segment má hodnotu 1, neznělý hodnotu 0.



Obr. 4.7 Znělé a neznělé segmenty

4.6 Základní tón řeči pomocí autokorelační funkce

Autokorelační funkce je definována pro rámec $x(n)$ jako:

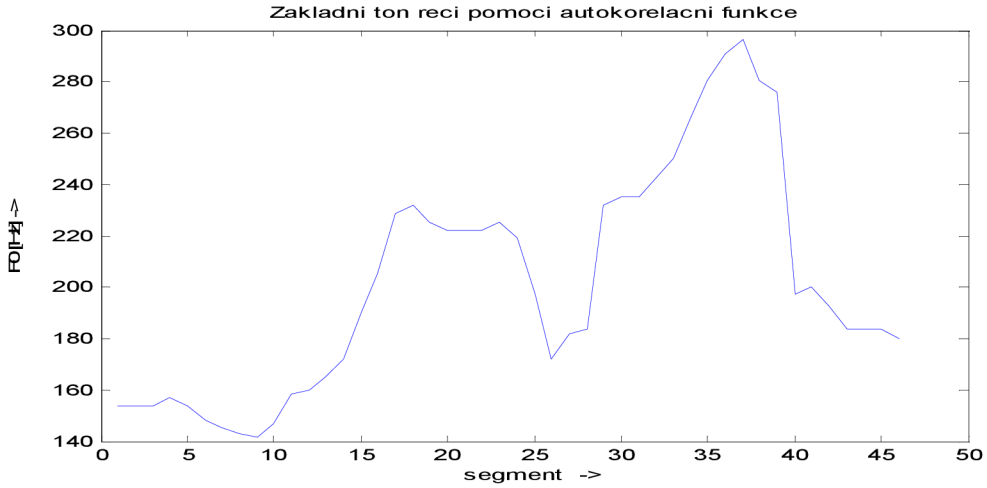
$$R(m) = \sum_{n=0}^{N-1-m} x(n)\, x(n+m) \tag{4.5}$$

Maximum této funkce hledáme pro $i \in [L_{\min}, L_{\max}]$, kde $L_{\min}$ je minimální povolená hodnota periody základního tónu ve vzorcích a $L_{\max}$ je maximum. Index maximální hodnoty označíme $i_{\max}$ a hodnotu $R_{\max}$. Pokud je :

$$\frac{R_{\max}}{R(0)} > \alpha\,, \tag{4.6}$$

kde α se volí asi 0,3, prohlásíme rámec za znělý a $i_{\max}$ udává periodu základního tónu řeči (ZTŘ). V opačném případě prohlásíme rámec za neznělý.

Nejprve je provedena autokorelace segmentu, vezme se její polovina a najde se maximum této korelace a jí odpovídající index. Dále se porovná toto maximum s prahovou hodnotou, aby byly získány pouze ZTŘ ze znělých úseků.



Obr. 4.8 Základní tón řeči pomocí autkorelační funkce – pouze znělé segmenty

4.7 Formanty

K výpočtu formantů použijeme metodu založenou na výpočtu pólů inverzního filtru. Vypočteme lineární predikční koeficienty a_i pomocí LPC analýzy (v Matlabu funkce lpc). Řád predikce nastavíme podle vztahu

$$p = \left\lfloor \frac{f_{vz}}{1000} \right\rfloor + 2\,, \tag{4.7}$$

kde f_{vz} je vzorkovací kmitočet. Vypočítají se póly přenosové funkce inversního filtru $H(z)$, přesněji se tedy vypočítají komplexní kořeny polynomu ve jmenovateli.

$$H(z) = \frac{G}{1 - \sum_{i=1}^{p} a_i z^{-i}}\,, \tag{4.8}$$

G je zesílení, pro náš výpočet nepodstatné. Získáme tedy komplexní tvar formantu (pólu)

$$z = |z| e^{j\varphi}, \tag{4.9}$$

kde $|z|$ je modul, φ je argument. Vybereme póly s komplexní složkou větší než 0. Následně se vypočte kmitočet formantu podle vztahu

$$F_x = \frac{\varphi}{2\pi} f_{vz} \tag{4.10}$$

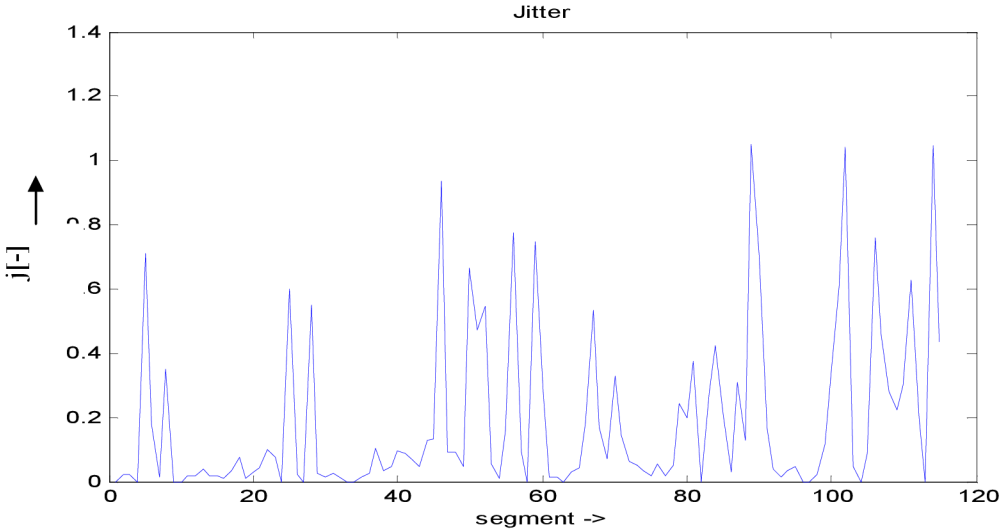
4.8 Kolísání základní periody (jitter)

Frekvence základního tónu ve skutečnosti není konstantní (hlasivkové pulsy nejsou ryze periodické). V delších řečových úsecích se totiž projeví vliv intonace promluvy. Délka periody i amplituda jednotlivých pulsů základního hlasivkového tónu se mírně liší i v rámci velmi krátkého signálu (obvykle již z periody na periodu). Takové mírné kolísání délky základní periody se nazývá jitter a je závislé na duševním (emocionálním) stavu mluvčího. Tento jitter je definován jako střední rozdíl délek sousedních period, dělený střední délkou periody. V tomto případě je ZTŘ počítán z kepstra.

$$j = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|T_0^{(i)} - T_0^{(i+1)}\right|}{\frac{1}{N}\sum_{i=1}^{N} T_0^{(i)}}, \tag{4.11}$$

$T_0^{(i)}, i=1,2,\ldots N$ je základní perioda a N = počet period

Jitter je počítaný podle vzorce (4.11) z periody základního tónu získaného pomocí kepstra.



Obr. 4.9 Kolísání základní periody počítané pro každé dvě sousední periody

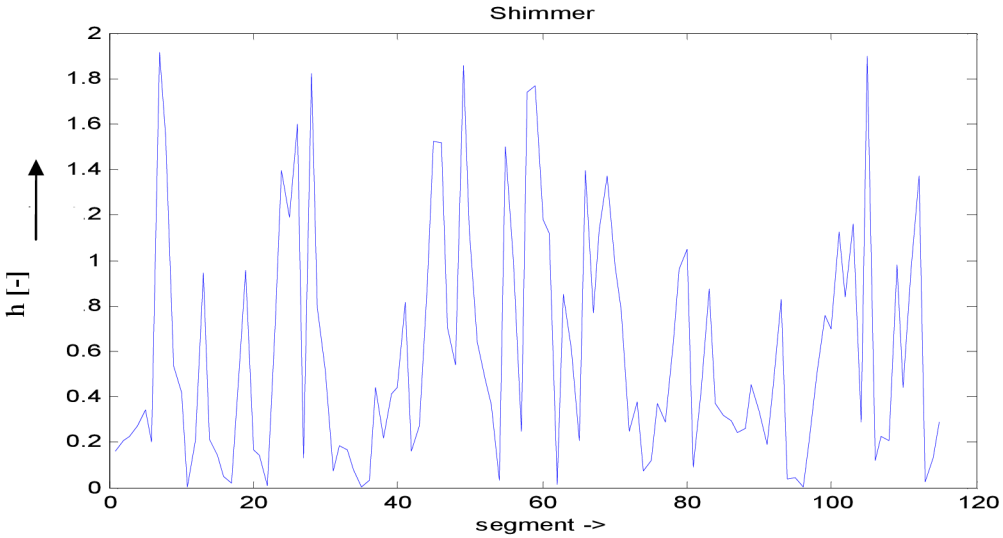
4.9 Kolísání amplitudy (shimmer)

Kolísání amplitudy hlasivkových pulsů se označuje jako shimmer. Je definován:

$$h = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|A^{(i)} - A^{(i+1)}\right|}{\frac{1}{N}\sum_{i=1}^{N} A^{(i)}} \tag{4.12}$$

$A^{(i)}, i=1,2,\ldots N$ je rozkmit modulu a N = počet impulsů

Tento parametr je počítán jako rozdíl maxim amplitud sousedních znělých segmentů.



Obr. 4.10 Kolísání amplitudy počítaný pro každé dva sousední segmenty

4.10 Spektrum

Periodické signály, jejichž analýza a syntéza je běžná (akustické aplikace), lze analyzovat pomocí Fourierových řad.

$$s(t) = \sum_{k=-\infty}^{\infty} c_k e^{jk\Omega t} \quad , \; \Omega = \frac{2\pi}{T_s} \tag{4.13}$$

$$c_k = \frac{1}{T_s}\int_0^{Ts} s(t) e^{-jk\Omega t} dt \tag{4.14}$$

Fourierova transformace má za určitých předpokladů k Fourierově řadě bezprostřední vztah. Mějme vzorkování zvoleno tak, aby perioda signálu byla celistvým násobkem vzorkovací periody T, tedy $T_S = NT$. (fázovým závěsem, dodatečným vzorkováním). Pokud nebude podmínka vzorkování splněna, nebudou koeficienty odpovídat skutečnosti a výsledné poskládání nebude tvořit přesnou funkci. Funkce $s(t)$ je frekvenčně omezená s horní mezní frekvencí $\omega_{\max}$ a vzorkování splňuje vzorkovací teorém $\omega_{\max} < \frac{\pi}{T} = \frac{N\Omega}{2}$, pak $c_k = 0$

pro $|k| > \frac{N}{2}$,

Spektrum počítáme z konečného počtu N vzorků. Původně nekonečný signál je vynásoben oknem o N vzorcích.

$$S(\omega) = DFT\{S(nT)\} = \sum_{n=0}^{N-1} S(nT) e^{-jk\Omega nT} \tag{4.15}$$

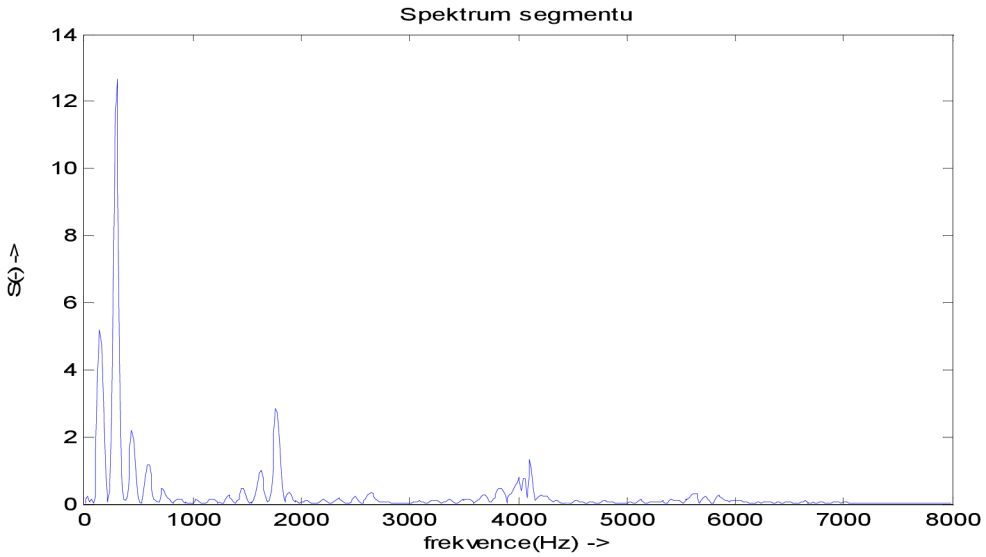
Diskrétní verze signálu pomocí zpětné DFT

$$s(nT) = \frac{1}{N} \sum_{k=0}^{N-1} S(k\Omega) e^{jk\Omega nT} = \frac{1}{N} \sum_{k=0}^{N-1} S(k\Omega) e^{jkn\frac{2\pi}{N}} \tag{4.16}$$

Pokud zaměníme spojitý čas t za diskrétní nT, dostaneme koeficienty Fourierovy transformace (jednotlivé frekvence)

$$c_k = \frac{1}{N} \sum_{N=0}^{N-1} s(nT) e^{-jk\Omega nt} \tag{4.17}$$

Spektrum je počítáno jako reálná složka pomocí funkce freqz v Matlabu.

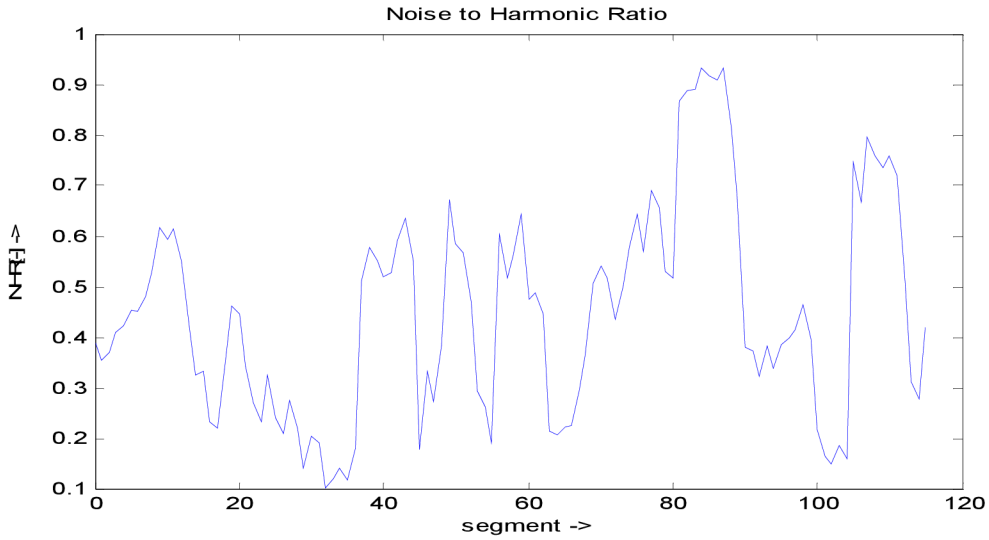


Obr. 4.11 Spektrum jednoho segmentu signálu

4.11 Poměr šumu a harmonické složky (NHR)

Noise to Harmonic Ratio – NHR je poměr hodnot neharmonické spektrální energie ve frekvenčním pásmu 1500-4500Hz, vzhledem k harmonické spektrální energii ve frekvenčním pásmu 70-4500Hz.

Tento příznak je počítán jako poměr energie neharmonické části spektra ku energii harmonické části spektra.



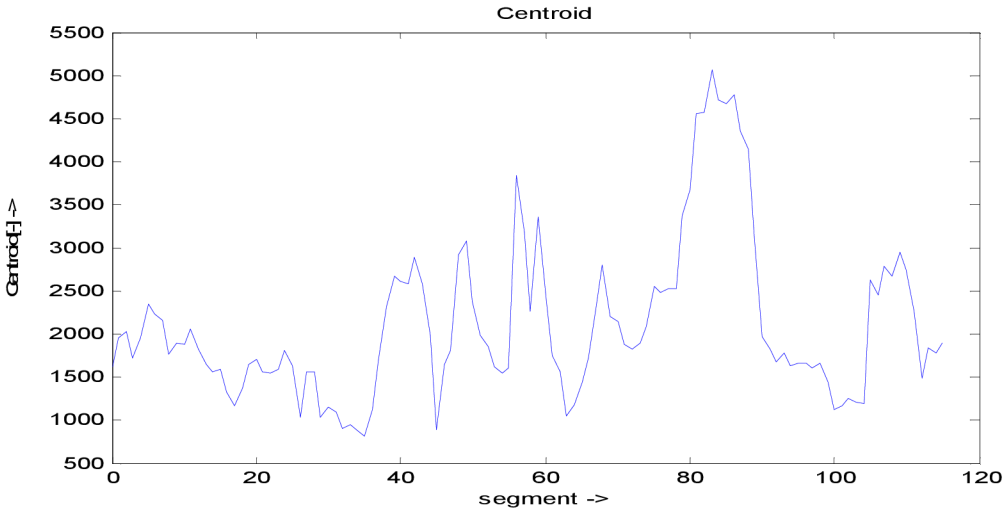
Obr. 4.12 Poměr šumu a harmonické složky počítaný pro všechny segmenty

4.12 Popis tvaru spektra

4.12.1 Spektrální centroid

Spektrální centroid je těžiště nebo též centrum spektra. Používá se v oblasti digitálního zpracování signálu, aby charakterizoval audio spektrum. Naznačuje, kde se nachází největší část spektra. Spektrální centroid je počítán jako vážený průměr z frekvencí, přítomných v signálu, které jsou vážené odpovídající amplitudou.

$$\mu = \frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{N=0}^{N-1} x(n)} \tag{4.18}$$



Obr. 4.13 Těžiště jednotlivých segmentů signálu

4.12.2 Spektrální rozptyl

Spektrální rozptyl je rozptyl spektra okolo jeho střední hodnoty.

$$\sigma^2 = \frac{\sum_{n=0}^{N-1} \left(f(n) - \mu\right)^2 \cdot x(n)}{\sum_{n=0}^{N-1} x(n)} \tag{4.19}$$

4.12.3 Spektrální šikmost

Šikmost udává hodnotu nesymetrie rozdělení okolo její střední hodnoty. Je počítána z momentu třetího řádu.

$$m_3 = \frac{\sum_{n=0}^{N-1} \left(f(n) - \mu\right)^3 \cdot x(n)}{\sum_{n=0}^{N-1} x(n)} \tag{4.20}$$

Šikmost je tedy:

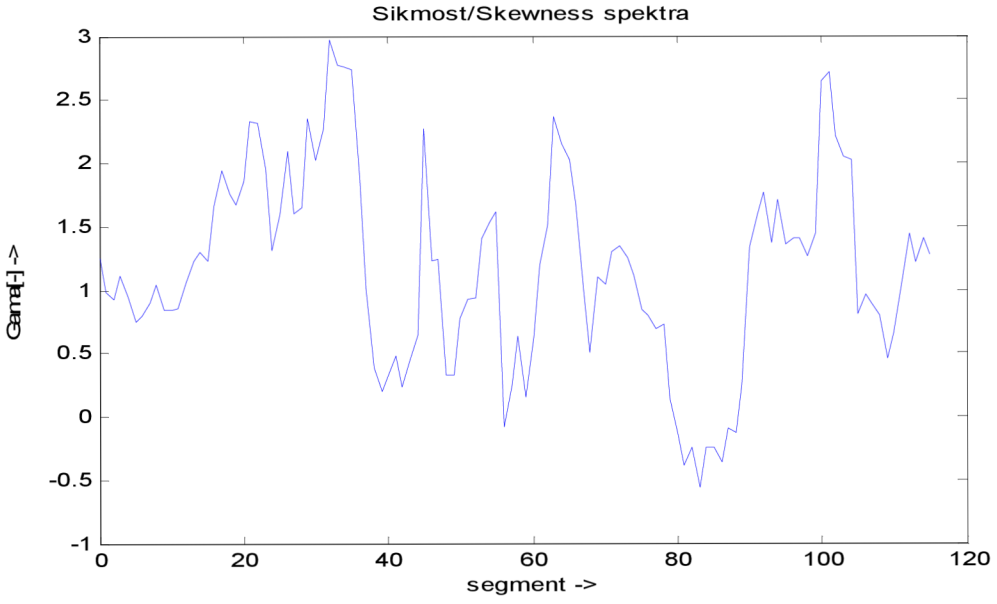
$$\gamma_1 = \frac{m_3}{\sigma^3} \tag{4.21}$$

Míra asymetrie rozložení tedy je:

$\gamma_1 = 0$, odpovídá symetrickému rozložení

$\gamma_1 < 0$, odpovídá rozložení více energie na pravé straně

$\gamma_1 > 0$, odpovídá rozložení více energie na levé straně



Obr. 4.14 Míra šikmosti spektra jednotlivých segmentů

4.12.4 Spektrální špičatost

Udává hodnotu špičatosti rozložení okolo jeho střední hodnoty. Je počítána z momentu 4.řádu.

$$m_4 = \frac{\sum_{n=0}^{N-1} (f(n) - \mu)^4 \cdot x(n)}{\sum_{n=0}^{N-1} x(n)} \tag{4.22}$$

Špičatost tedy je:

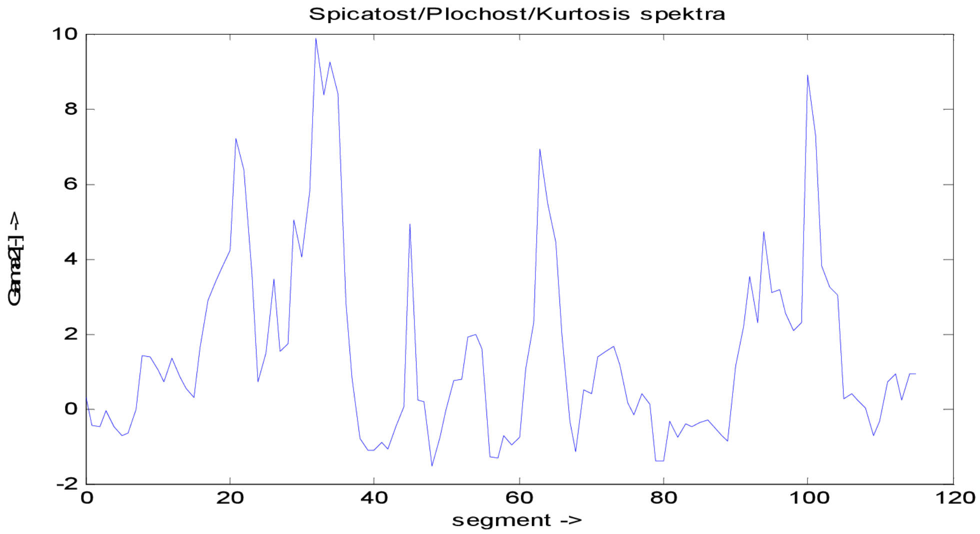
$$\gamma_2 = \frac{m_4}{\sigma^4} - 3 \tag{4.23}$$

Míra špičatosti rozložení tedy je:

$\gamma_2 = 0$, odpovídá normálnímu rozložení

$\gamma_2 < 0$, odpovídá ploššímu rozložení

$\gamma_2 > 0$, odpovídá špičatějšímu rozložení

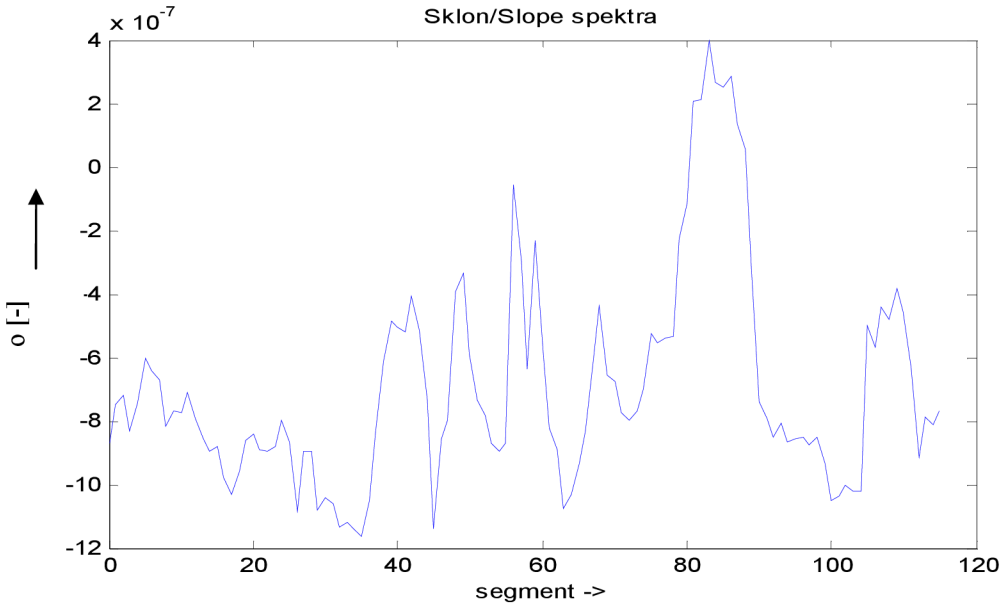


Obr. 4.15 Špičatost spektra počítaná pro jednotlivé segmenty

4.12.5 Spektrální sklon

Udává hodnotu snižující se spektrální amplitudy. Je počítán jako lineární regrese spektrální amplitudy, což představuje aproximaci daných hodnot polynomem prvního řádu (přímkou) metodou nejmenších čtverců.

$$o = \frac{1}{\sum_{n=0}^{N-1} x(n)} \frac{N \sum_{n=0}^{N-1} f(k) x(k) - \sum_{n=0}^{N-1} f(k) \sum_{n=0}^{N-1} x(k)}{N \sum_{n=0}^{N-1} f^2(k) - \left(\sum_{n=0}^{N-1} f(k) \right)^2} \tag{4.24}$$

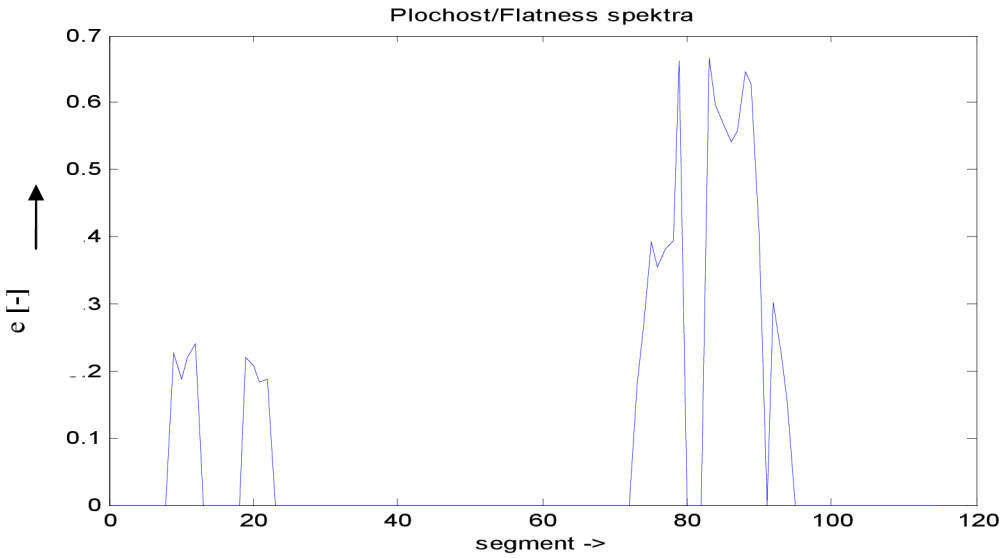


Obr. 4.16 Sklon spektra jednotlivých segmentů

4.12.6 Spektrální plochost

Spektrální plochost se používá v oblasti digitálního zpracování signálu za účelem charakterizace audio spektra. Vysoká hodnota spektrální plochosti naznačuje, že spektrum má podobné množství energie ve všech spektrálních pásmech – to znamená bílý šum, a graf spektra, by vypadal relativně ploše a hladce. Nízká hodnota spektrální plochosti naznačuje, že spektrum je soustředěno v relativně malém počtu pásem – to znamená směs sinusových vln a že spektrum vypadá "špičatě". Spektrální plochost se vypočítá vydělením geometrického průměru výkonového spektra střední hodnotou výkonového spektra:

$$e = \frac{\sqrt[N]{\prod_{n=0}^{N-1} x(n)}}{\frac{1}{N}\sum_{n=0}^{N-1} x(n)} \tag{4.25}$$



Obr. 4.17 Plochost spektra jednotlivých segmentů

4.13 Melovské kepstrální koeficienty (MFCC)

Mel Frequency Cepstral Coefficients (MFCC). Jedná se o nejpoužívanější příznaky používané v oblasti rozpoznávání řeči. Lidské ucho má na nízkých frekvencích větší rozlišovací schopnost než na frekvencích vysokých. Pokud se chceme co nejvíce přiblížit lidskému uchu, rozmístíme frekvenční charakteristiky na kmitočtové ose nelineárně. Frekvenční osu můžeme nelineárně upravit a na upravené ose pak filtry rozmístit rovnoměrně. Používaná nelineární úprava využívá převodu Hertzů na mely

$$f_m = 2565\log(1+\frac{f}{700}) \tag{4.26}$$

Převod $mel \rightarrow Hz$

$$f = 700(e^{\frac{f_m}{1127.01048}} - 1) \tag{4.27}$$

Obr. 4.18 Postup výpočtu MFCC koeficientů

Na vstup je systému je přiveden signál, který je váhován oknem o délce 512 vzorků, tato hodnota je volena vzhledem k následujícímu bloku výpočtu výkonového spektra pomocí FFT, tzn. mocnině 2. Klíčová část je melovská filtrace. Výpočetní algoritmus je realizován bankou trojúhelníkových filtrů. Lineárním rozmístěním filtrů na melovské ose má za následek nelineární rozmístění na standardní kmitočtové ose v Hz. Průchod signálu filtrem znamená, že každý koeficient FFT je násoben odpovídajícím ziskem filtru a výsledky jsou pro příslušné

filtry akumulovány. Další krok spočívá ve výpočtu logaritmu výstupů jednotlivých filtrů. Posledním krokem je provedení diskrétní kosinové transformace

$$c_m(j) = \sum_{i=1}^{M^*} \log y_m(i) \cos\left(\frac{\pi j}{M^*}(i-0{,}5)\right), \quad pro\ j = 0,1,...,M\,, \tag{4.28}$$

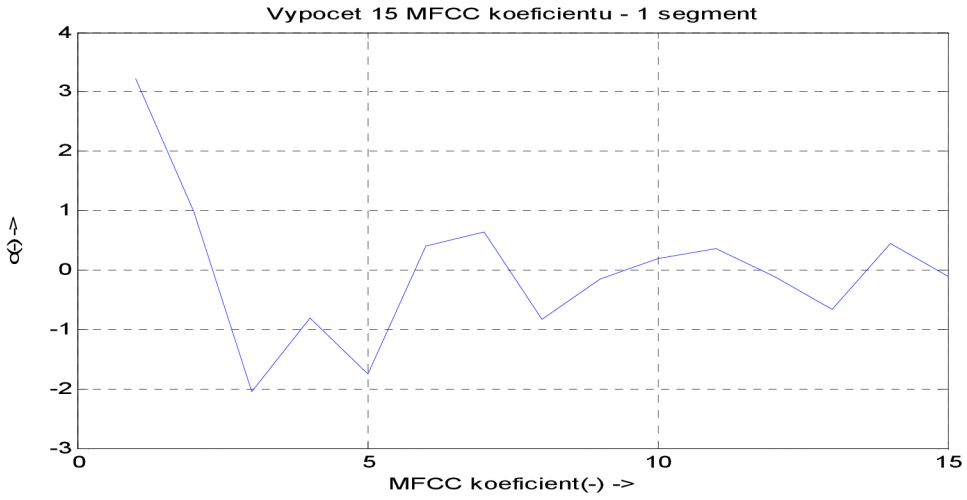
kde M^* je počet pásem melovského filtru, M je počet melovských kepstrálních koeficientů. Nulový koeficient $c_m(0)$ je úměrný logaritmu energie signálu, a proto je nahrazován výpočtem logaritmu krátkodobé energie přímo ze vzorků signálu.

Výpočet odezvy jednoho filtru

$$M(f,i) = \begin{cases} \dfrac{f - b_{i-1}}{b_i - b_{i-1}} & pro\ b_{i-1} \le f \le b_i \\ \dfrac{f - b_{i+1}}{b_i - b_{i+1}} & pro\ b_i \le f \le b_{i+1} \\ 0 \quad pro \quad ostatní \quad f \end{cases}, \tag{4.29}$$

kde M je funkce jednoho filtru, i je index filtru v bance, b_{i-1} je střední kmitočet předchozího filtru, b_{i+1} je střední kmitočet následujícího filtru.

Těmto příznakům bylo věnováno více testovacího prostoru, protože jsou nejpoužívanějšími pro rozpoznání řeči.



Obr. 4.19 MFCC koeficienty jednoho segmentu promluvy

4.14 Výběr suprasegmentálních příznaků

Pro správnou klasifikaci je potřeba velké množství suprasegmentálních příznaků, proto byly z výše uvedených příznaků vypočteny následující suprasegmentální příznaky:

- maximální hodnota- $\max(x)$ (4.30)
- minimální hodnota- $\min(x)$ (4.31)
- střední hodnota – $mean(x)$ (4.32)
- medián – $median(x)$ (4.33)
- směrodatná odchylka - $\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$ (4.34)
- relativní maximum - $\frac{\max(x)}{mean(x)}$ (4.35)
- relativní minimum - $\frac{\min(x)}{mean(x)}$ (4.36)

K dalšímu použití byly vybrány příznaky : ZTŘ, NHR, ZCR, energie, jitter, shimmer, centroid, spektrální nesouměrnost, spektrální špičatost, spektrální plochost, spektrální pokles.

5 EMOCE

Jen asi 10% informace z řečového signálu nám dává informaci o stavu mluvčího, hlavně emočním.

Při rozpoznávání řeči se lze soustředit na různé aspekty. Jedním z cílů bylo vytvořit syntetizovanou řeč co nejpřirozenější, dalším cílem bylo rozpoznat citový obsah z řeči.

Řečový signál zabarvený emocemi nám dává komplexnější pohled na řečníka. Na emotivní stav mluvčích reagují posluchači a přizpůsobují své chování podle druhu emoce, kterou mluvčí vyjadřuje. Například smutným lidem ukazujeme empatii, rozhněvaných se bojíme. Pro určení emočního stavu mluvčího na základě prozodických vlastností a kvality hlasu musíme roztřídit zvukové rysy v řeči a přiřadit je k náležejícím emocím.

Nalezení vhodných akustických vlastností k jednotlivým emocím není příliš jednoduché, také proto si výsledky občas odporují. Je těžké definovat, které příznaky se vztahují k emotivní řeči.

Nejjednodušší přístup k popisu emocí je použít kategorie používané v běžném hovorovém jazyce. Toto rozdělení umožňuje různé způsoby dělení kategorií, které mohou být použity pro popis emočních stavů a emocí.

Podle emočního výzkumu se emoce rozdělují do dvou kategorií: primární a sekundární.

Primární emoce

Kategorie obsahuje takové emoce, které jsou „čisté“ a „jednoduché“. Tyto emoce

mají jen několik forem, které jsou od sebe kvalitativně odlišné. Každá forma má příznaky, kterými se od ostatních odlišuje. Seznam základních emocí je vlastně jen dohoda – strach, vztek, štěstí / radost, smutek, nuda. Občas sem lze zařadit překvapení, hněv, pohrdání.

Sekundární emoce

Tato kategorie obsahuje emoční stavy, které jsou odvozeny z emocí primárních jejich smícháním. Tyto odvozeniny pokrývají velký rozsah emočních stavů, avšak málo z nich by mohlo být považováno za emoce základní. Mluvíme například o emocích: žal, zalíbení / něžnost, sarkasmus / ironie, překvapení / údiv, nenávist / odpor.

Pro náš výzkum bylo vybráno 7 následujících emocí:

- Neutralita - neutral
- Smutek - sadness
- Radost - happiness
- Vztek - anger
- Nuda - boredom
- Strach - fear
- Znechucení – disgust

6 KVALITA PŘÍZNAKŮ

V předchozí části jsme se zabývali výpočtem jednotlivých příznaků. Pro další postup bude potřeba zjistit, které z těchto příznaků jsou pro rozpoznání emocí nejvhodnější.

Podle [SIG-00] je příznak považován za kvalitní, jestliže se prvky jedné třídy vyskytují v okolí střední hodnoty a současně se střední hodnoty jednotlivých tříd co nejvíce liší. Proto bude nutné pro každý příznak vypočítat míru geometrické oddělitelnosti $Q(x_i)$

$$Q(x_i) = \frac{S^2}{S^2 + D^2}, 0 \le Q(.) \le 1, \tag{6.1}$$

kde S je aritmetická střední hodnota, D je aritmetická střední hodnota vzdálenosti. Jestliže příznak x_i vykazuje malé rozdíly v rámci své třídy a velké rozdíly v rámci jiných tříd, tak míra oddělitelnosti $Q(x_i)$ dosahuje malých hodnot. V tom případě je příznak dobře použitelný. V opačném případě, nabývá-li $Q(x_i)$ velkých hodnot (maximálně 1), značí to velký rozptyl hodnot a nevhodnosti příznaku pro rozpoznání jednotlivých tříd. Aritmetická střední hodnota S se určí ze vztahu

$$S = \frac{1}{V}\sum_{v=1}^{V} S_v^2 , \tag{6.2}$$

kde V označuje celkový počet tříd. Kvadrát rozptylu jedné třídy v okolo střední hodnoty je

určen

$$S_v^2 = (\underline{x} - \underline{\mu_v})^2, \tag{6.3}$$

kde $\underline{x}$ představuje vektor příznaků a $\underline{\mu}$ střední hodnotu. Aritmetickou střední hodnotu vzdálenosti vypočítáme podle

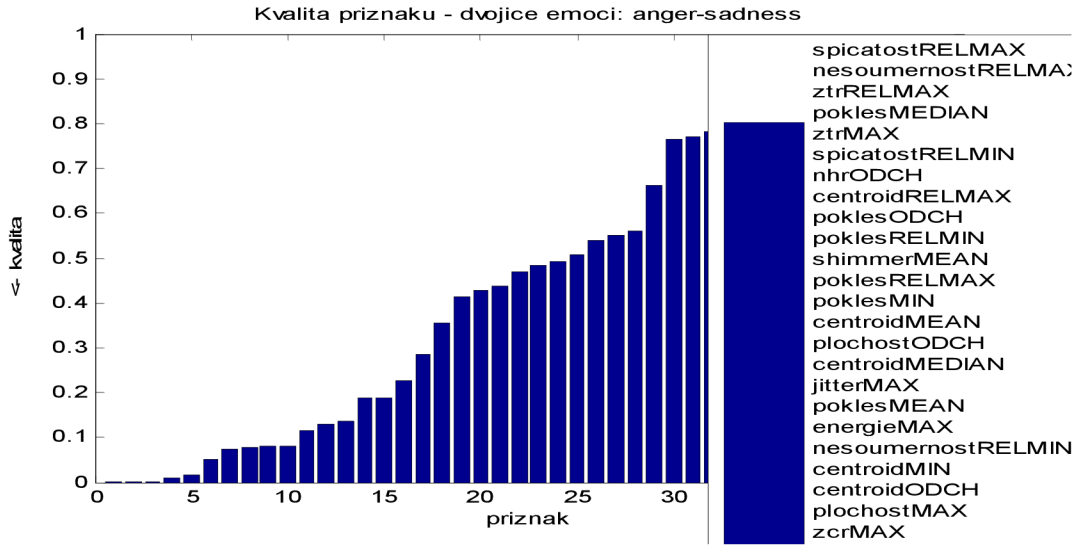
$$D = \frac{1}{V(V-1)} \sum_{v=1}^{V} \sum_{u=1}^{V} D_{v,u}^2 . \tag{6.4}$$

Kvadrát vzdálenosti mezi středními hodnotami dvou tříd se vypočte z

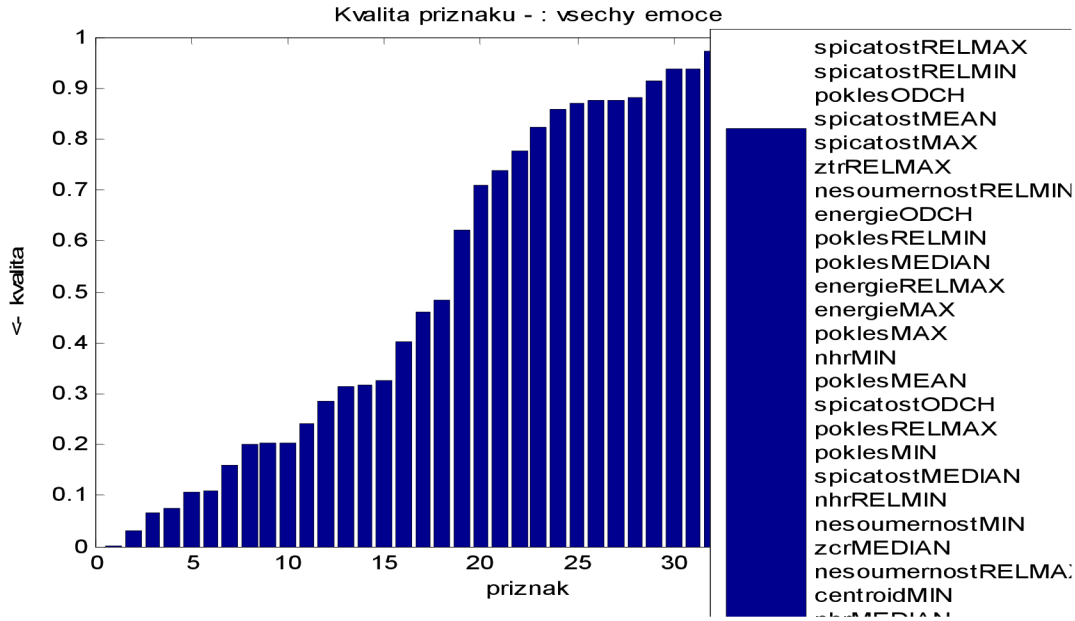
$$D_{v,u}^2 = (\underline{\mu_v} - \underline{\mu_u})^2 . \tag{6.5}$$

6.1 Kvalita suprasegmentálních příznaků

Pro každou promluvu z uvedené databáze bylo vypočítáno 77 výše zmíněných suprasegmentálních příznaků. Pro každý jednotlivý příznak byla spočtena jeho kvalita. Ta byla jednak počítána pro všechny možné dvojice emocí a také jako kvalita příznaku přes všechny uvažované emoce. Výsledek nám ukázal, že pro klasifikaci různých dvojic emocí by bylo vhodné použít různý specifický sled příznaků. Příklad dvojice emocí anger – sadness, happiness – boredom. Legenda popisuje posloupnost jednotlivých příznaků od nejlepšího.



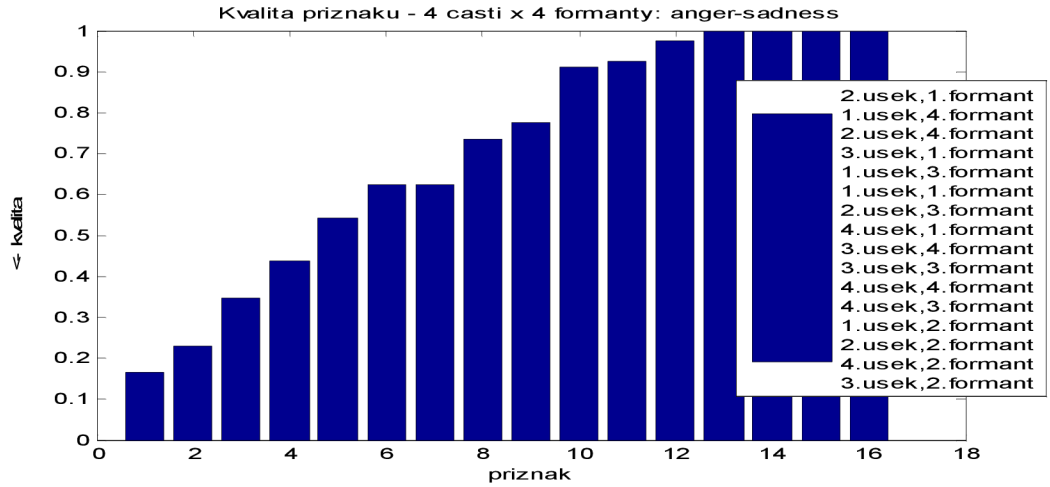
Obr. 6.1 Kvalita suprasegmentálních příznaků pro dvojici emocí anger – sadness



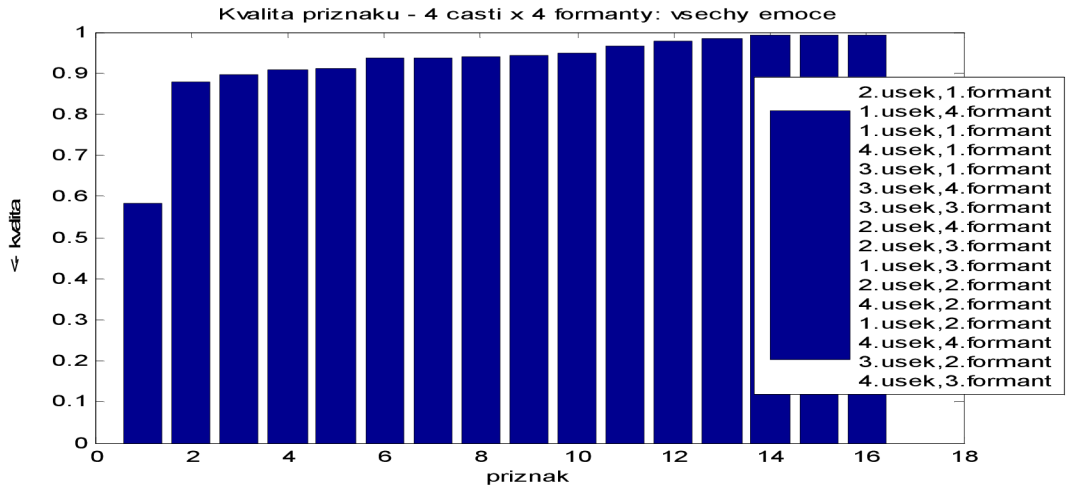
Obr. 6.2 Kvalita suprasegmentálních příznaků pro všechny emoce

6.2 Kvalita formantů

Z každého znělého segmentu signálu byly vypočítány 4 formantové frekvence. Promluva byla rozdělena na 4 části. V každé části byla vypočtena střední hodnota všech čtyřech formantů a to z odpovídajících segmentů. Tzn. 4 části po 4 formantech. Kvalita byla počítána opět pro dvojice emocí i všechny emoce. Z výsledných závislostí vyplývá, že by se tyto příznaky hodily k rozpoznání určitých dvojic emocí, avšak pro komplexní rozpoznávání se svojí kvalitou příliš moc nehodí. Opět legenda popisuje posloupnost příznaků od nejlepšího k nejhoršímu.



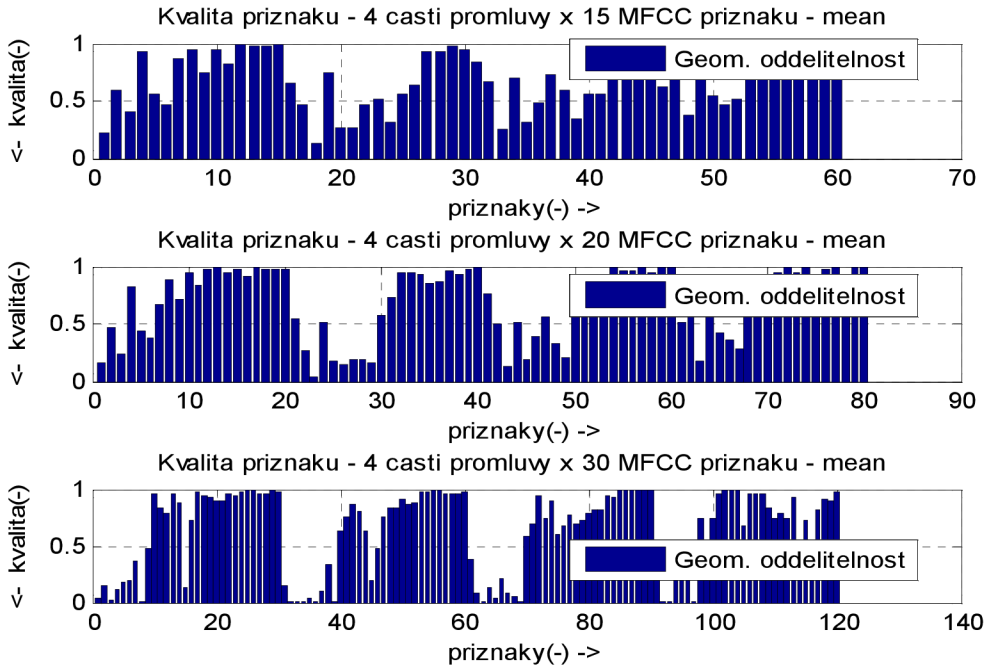
Obr. 6.3 Kvalita formantů pro dvojici emocí anger – sadness



Obr. 6.4 Kvalita formantů pro všechny emoce

6.3 Kvalita MFCC koeficientů

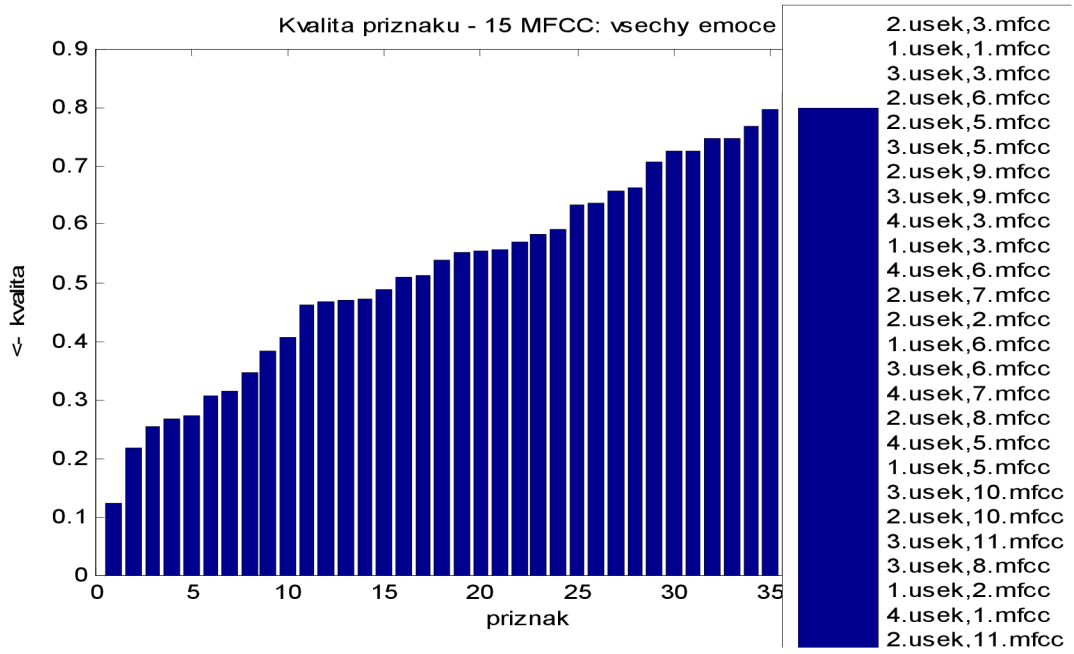
Kvalitu MFCC koeficientů jsme rozdělili do několika částí. V první části se zabýváme nejvhodnějším počtem MFCC koeficientů. Výpočet je prováděn pro počty 5, 10, 15, 20, 30, 40, 50 koeficientů. Počet koeficientů 5 a 10 je příliš nízký pro naše potřeby rozpoznávání, proto uvažujeme 15 a více koeficientů. V případech 30 a víc koeficientů je jich použitelných zhruba jen 10, proto bude vhodné počítat s 15 koeficienty, protože v tomto případě nebude výpočet zbytečný a zavádějící.



Obr. 6.5 Kvalita MFCC příznaků – nejvhodnější počet

6.3.1 Kvalita 15 MFCC příznaků

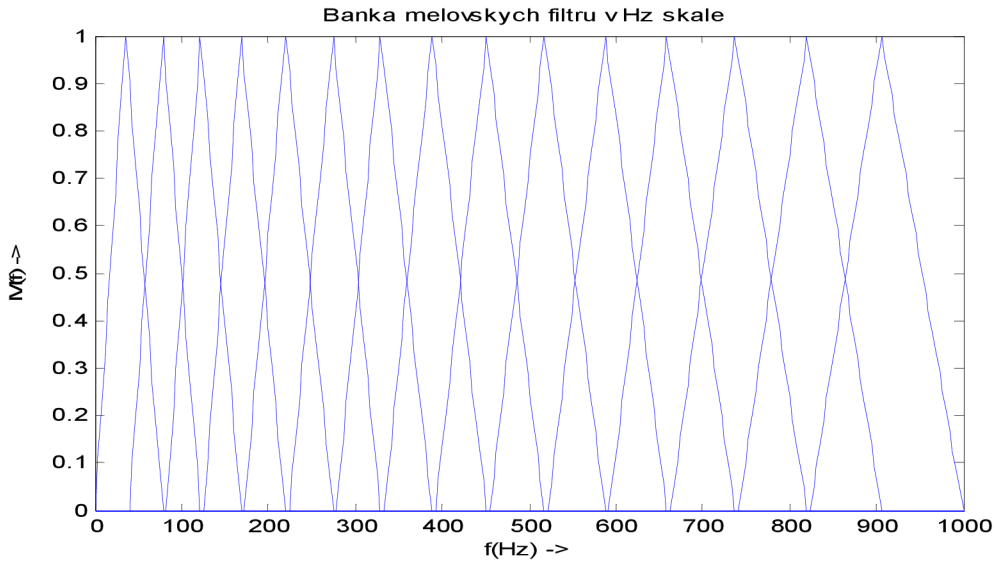
V této části jsme vypočetli 15 MFCC koeficientů pro každý segment a celou promluvu opět rozdělili na 4 úseky. V každém úseku byly z jednotlivých segmentů MFCC koeficienty průměrovány. Takto jsme získali 60 MFCC příznaků a vypočetli jejich kvalitu přes všechny emoce i pro jednotlivé dvojice emocí. Osa x udává jednotlivé příznaky, ty jsou od nejlepšího seřazeny v legendě.



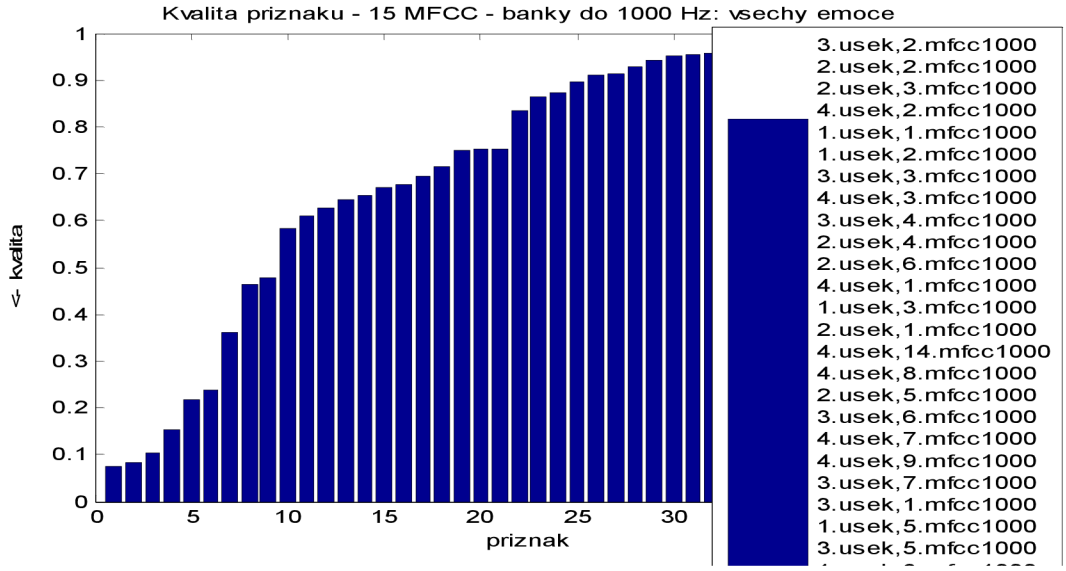
Obr. 6.6 Kvalita 15 MFCC koeficientů vypočtená pro všechny emoce

6.3.2 Kvalita 15 MFCC s omezením pásma do 500, 800, 1000 Hz

Z výše uvedených poznatků víme, že se k výpočtu MFCC koeficientů používá banka melovských filtrů. Pro rozpoznávání emocí by mohlo být vhodné získat lepší namodelování řeči na nízkých kmitočtech. To lze udělat tak, že melovské banky navrhneme pouze do určitého kmitočtu. V našem případě byly banky navrhnuty do kmitočtů 500, 800, 1000 Hz a kvalita příznaku vypočtena obdobně jako u předchozího bodu. Kvalita několika prvních koeficientů se oproti předchozímu případu zvýšila, avšak při výběru více než 10 příznaků jsou již výsledky neúměrně horší.



Obr. 6.7 Banka melovských filtrů v Hz škále do 1000Hz



Obr. 6.8 Kvalita 15 MFCC koeficientů –banky do 1000 Hz pro všechny emoce

6.4 Hodnocení příznaků

Další možností výpočtu kvality jednotlivých příznaků umožňuje Matlabovská funkce „rankfeatures“. Pro vybranou třídu (emoční stav) a všechny výše zmíněné příznaky funkce vypočítá kvalitu příznaku vzhledem k ostatním třídám (emočním stavům). Z takto získaných kvalit příznaků se již jednoduchým součtem jednotlivých pořadí příznaků u všech tříd (emocí) získá výsledné hodnocení jednotlivých příznaků. Příznaky seřazené funkcí rankfeatures jsou k dispozici v přílohách 3,4,5,6.

7 KLASIFIKÁTOR

7.1 Neuronová síť – se zpětným šířením

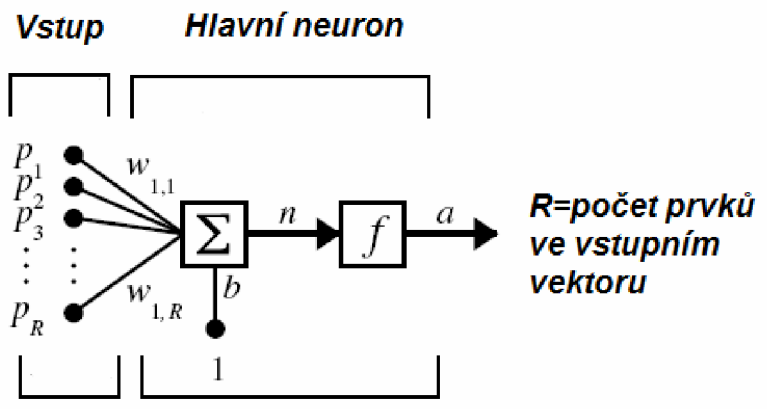
Umělé neuronové sítě jsou inspirovány biologickými neuronovými sítěmi, využívají distribuované, paralelní zpracování informace. Znalosti jsou ukládány především prostřednictvím síly vazeb. Učení je základní a podstatná jejich vlastnost.

Zpětné šíření je zobecněné pravidlo učení Widrow-Hoff pro vícevrstvé sítě a nelineární diferenciální aktivační funkce. Vstupní a odpovídající výstupní vektory jsou použity pro trénink sítě, dokud síť dokáže aproximovat funkci, přiřadit výstupní vektor k odpovídajícímu vstupnímu vektoru, nebo klasifikovat vstupní vektory.

Trénink může vést k lokálnímu chybovému minimu spíše než globálnímu. Potom může lépe pracovat síť s větším počtem neuronů, nebo síť s jinými počátečními podmínkami.

Umělou neuronovou síť charakterizuje: model neuronů, architektura sítě, způsob učení, způsob vybavování.

7.1.1 Model neuronu



Obr. 7.1Model neuronu

Neuron s R vstupy, váhami w, prahem b, vstupním vektorem p a aktivační funkcí f má výstup určen:

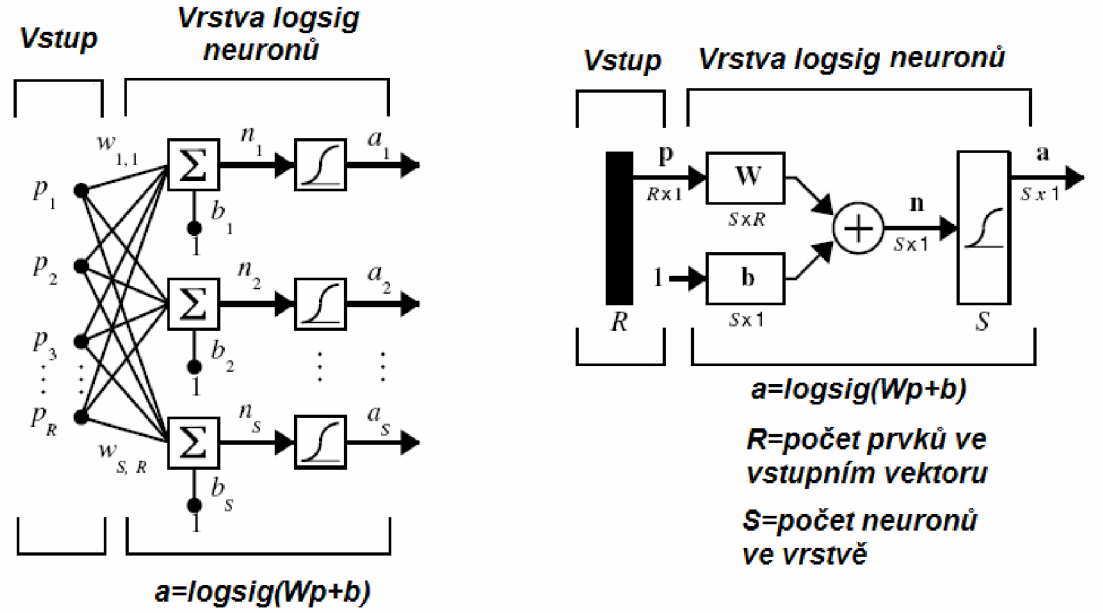
$$a = f(wp + b) \tag{7.1}$$

Aktivační funkce f mohou být: logsig (výstup 0 .. 1), tansig (výstup -1 .. 1), purelin (výstup libovolný)

Uvedené aktivační funkce jsou monotónně rostoucí a jsou diferencovatelné. V algoritmu backpropagation jsou důležité derivace aktivačních funkcí. Pro logsig, tansig a purelin existují v Matlabu jejich derivace: dtansig, dlogsig a dpurelin. Mimo uvedených funkcí je možné vytvořit vlastní aktivační funkce a jejich derivace [NEU-08].

7.1.2 Architektura sítě – dopředná síť

Způsob uspořádání neuronů v síti a jejich vzájemné propojení.

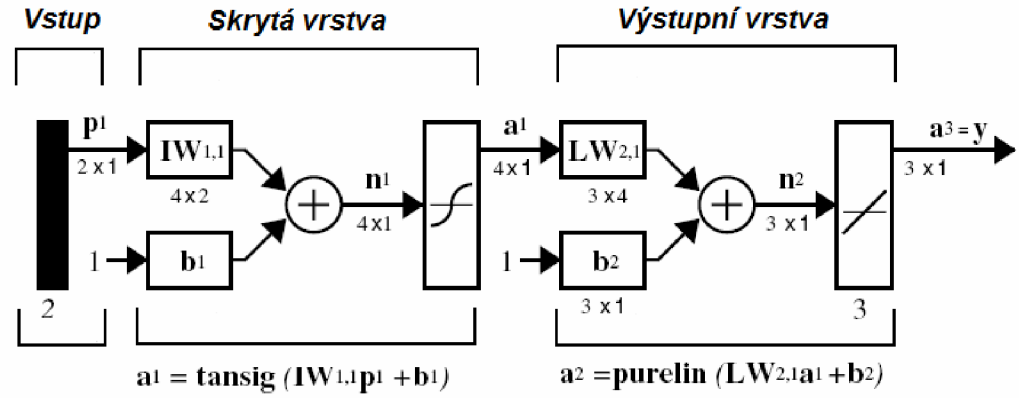


Obr. 7.2 Jednovrstvá dopředná síť:úplné zobrazení(vlevo), schématické zobrazení vrstvy(vpravo)

Jednovrstvá síť s S neurony s *logsig* aktivační funkcí, R vstupy, p vstupním vektorem, váhami W a prahy b má výstup:

$$a = \log sig(Wp + b) \tag{7.2}$$

Dopředné sítě obvykle obsahují jednu nebo více skrytých vrstev sigmoidních neuronů následovaných výstupní vrstvou lineárních neuronů. Další vrstvy neuronů s nelineární přenosovou funkcí umožňují síti, aby se učila nelineární a lineární vztahy mezi vstupními a výstupními vektory. Lineární výstupní vrstva umožňuje síti produkovat výstupní hodnoty v rozsahu <-1;1>.



Obr. 7.3 Vícevrstvá dopředná síť: schématické zobrazení vrstev

Takováto síť může být použita jako univerzální aproximátor funkce. Umožňuje aproximovat jakoukoliv funkci s konečným počtem nespojitostí, pokud má dostatečný počet neuronů ve skryté vrstvě [NEU-08].

7.1.3 Inicializace vah

Počáteční váhy a prahy jsou vytvořeny pomocí funkce *newff*, která funguje podobně, jako *newp*. Např.: *net=newff([-1 2; 0 5],[3,1],{'tansig','purelin'},'traingd');* Prvním argumentem určujeme meze vstupů a zároveň určujeme jejich počet. Vektor [3 1] stanovuje počty neuronů v jednotlivých vrstvách. Následují aktivační funkce jednotlivých vrstev ve složených závorkách [NEU-08].

7.1.4 Algoritmus učení vícevrstvé neuronové sítě

Krok 1. Inicializace vah a prahů

Nastavení vah a prahů všech neuronů sítě na malé náhodné hodnoty.

Krok 2. Předložení vzoru a jemu odpovídajícího výstupu

Na vstupy sítě předložíme nový trénovací vzor. Každý vzor je vektor vstupních elementů $X = \{x_1, x_2, \ldots, x_N\}$. Hodnoty těchto prvků jsou libovolná reálná čísla. Dále předložíme i vektor odpovídajících správných výstupů $D = \{d_1, d_2, \ldots, d_N\}$. Trénovací vzory vybíráme náhodně z uvažované globální množiny všech vzorů. Pokud použijeme tuto síť jako prostý klasifikátor, budou všechny požadované výstupní elementy nulové vyjma toho, který odpovídá správné odpovědi. Ten naopak bude jednotkový.

Krok 3. Výpočet aktuálních výstupů sítě

Dříve než budeme moci opravit váhy, musíme nejprve zjistit hodnotu chyby. Tu získáme z rozdílu požadované a skutečné hodnoty výstupu. Protože požadovaný výstup známe, zbývá nám určit skutečnou hodnotu výstupu. V tomto kroku budeme postupně procházet sítí od vstupů k výstupům a počítat výstupy z jednotlivých vrstev. Tyto hodnoty opět použijeme jako vstupy do následující vrstvy a pokračujeme tak až k poslední vrstvě. Uvědomíme si, že pro vstupy platí $x_j = \mu_j^1,\ 1 \le j \le P^1 = N$. Výstupy z poslední vrstvy jsou přímo naše očekávané hodnoty, tj. $y_j = \mu_j^{M+1}$.

$$\mu_j^{s+1} = f_s\left[\sum_{i=1}^{P^s} w_{ij}^s \mu_i^s - \Theta_j^s\right] \quad 1 \le j \le P^s, 1 \le s \le M \tag{7.3}$$

Proměnná s udává uvažovanou vrstvu ze které, resp. pro kterou počítáme výstupy. Její hodnota se pohybuje v intervalu $1 \le s \le M$, kde M je počet vrstev sítě. Parametr P, který udává počet neuronů ve vrstvě, není tentokrát konstanta, ale proměnná. V každé vrstvě může mít síť obecně různý počet neuronů. Proto také zápis P^s představuje právě aktuální počet neuronů v s-té vrstvě. Hodnota $P^{s=1}$ je rovna přímo počtu elementů N vstupního

vzoru a hodnota $P^{s=M+1}$ je rovna počtu výstupů. Symbolem w_{ij}^s označujeme váhy z i-tého do j-tého neuronu. Jako prahovou funkci f_s jsme použili sigmoidu.

Krok 4. Adaptace vah

Nyní, když známe hodnotu výstupu, známe i hodnotu chyby a můžeme přikročit k hlavní fázi učení. Tato fáze se nazývá back-propagation, protože budeme protlačovat získané odchylky zpět od výstupu ke vstupu a současně adaptovat váhy mezi jednotlivými vrstvami, aby lépe odpovídaly předloženému vzoru. Pro nové hodnoty vah platí:

$$w_{ij}^s(t+1) = w_{ij}^s(t) + \eta \delta_j^{s+1} \mu_i^s \tag{7.4}$$

V této rovnici jsou w_{ij}^s váhy z i-tého do j-tého neuronu, η je parametr učení, δ_j^s jsou chyby jednotlivých neuronů v uvažované vrstvě a μ_i^s jsou buď výstupy neuronů uvažované vrstvy, nebo hodnoty vstupů.

Jestliže se jedná o poslední vrstvu, tj. $s = M$, potom y_j je výstup z j-tého neuronu a pro zjištění δ_j^s použijeme tento vztah:

$$\delta_j^M = y_j(1-y_j)(d_j - y_j) \tag{7.5}$$

Je zřejmé, že tento vztah použijeme pouze jedenkrát a to hned na začátku učící fáze. Další hodnoty δ_j^s pro neurony v předcházejících vrstvách získáme ze vztahu:

$$\delta_j^s = \mu_j^{s+1}(1-\mu_j^{s+1}) \sum_{k=1}^{P^{s+1}} \delta_k^{s+1} w_{jk}^{s+1}, s = M-1, \ldots, 1, \tag{7.6}$$

kde k prochází všechny neurony $s+1$ vrstvy, tj. ty, které jsou pod j-tou vrstvou (tedy blíže k výstupu).

Poznámka: Pro urychlení a zlepšení konvergence se používají různé metody adaptace vah. O některých z nich se zmíníme dále. Na tomto místě si uvedeme pouze vztah pro adaptaci vah s využitím momentu:

$$w_{ij}^s(t+1) = w_{ij}^s(t) + \eta \delta_j^{s+1} \mu_i^s + \alpha \left[w_{ij}^s(t) - w_{ij}^s(t-1) \right], \tag{7.7}$$

kde $0 \le \alpha \le 1$ je momentový parametr.

Krok 5. Opakování procesu učení

Pokud se váhy ještě nestabilizovaly, tj. během jednoho cyklu došlo ke změně alespoň jedné váhy, nebo jsme nedosáhli požadované přesnosti $E < \varepsilon$, přejdeme opět ke Kroku 2. V opačném případě skončíme. Pro trénování algoritmem se zpětným šířením se užívá funkce *train* při nastavení *Net.trainFcn = 'traingd'* [JIR-02].

7.1.5 Vylepšené algoritmy se zpětným šířením

Učení s momentem, učení s adaptivním krokem učení, Levenberg-Marquardtovo pravidlo učení a další naleznete mezi demonstračními soubory Matlabu. V porovnání se základním algoritmem jsou jejich výsledky obvykle výrazně lepší.

Učení s momentem

Moment umožňuje síti reagovat nejen na lokální gradient, ale také na aktuální trendy v chybovém povrchu - umožňuje síti ignorovat jeho malé změny. Moment je do pravidla učení přidán následujícím způsobem: 1) změny vah jsou rovny součtu části posledních změn a nových změn vypočítaných pomocí zpětného šíření, 2) velikost efektu uvažování poslední změny vah je řízena konstantou momentu - mc (momentum constant), 3) pokud mc = 0, je moment ignorován, pokud mc = 1, jsou ignorovány nové změny a váhy jsou upraveny stejně, jako v předchozím kroku. Obvyklá hodnota je 0.95.

Nové váhy a prahy budou zamítnuty, pokud bude přírůstek chyby příliš velký - síť se pohybuje po chybovém povrchu do prudkého kopce. Učení je implementováno funkcí *traingdm,* která odpovídá funkci *traingd* s přidáním parametrů: konstanta momentu a maximální rozsah chyby. Při trénování jednoho neuronu může křivka chyby obsahovat lokální minima a díky zavedení momentu se lze přes tato minima dostat.

Moment a adaptivní krok učení

Pro zrychlení tréninku je použita funkce *traingda*, která obsahuje moment a adaptivní krok učení. Adaptivní krok učení je další metoda pro zrychlení učení. Síť se učí tím rychleji, čím je krok učení větší, nesmí však přesáhnout jistou hranici. Adaptivní krok učení proto zvětšuje krok učení, dokud není změna chyby příliš velká, potom je opět zmenšován, dokud není učení opět stabilní [JIR-02].

7.1.6 Trénovací data

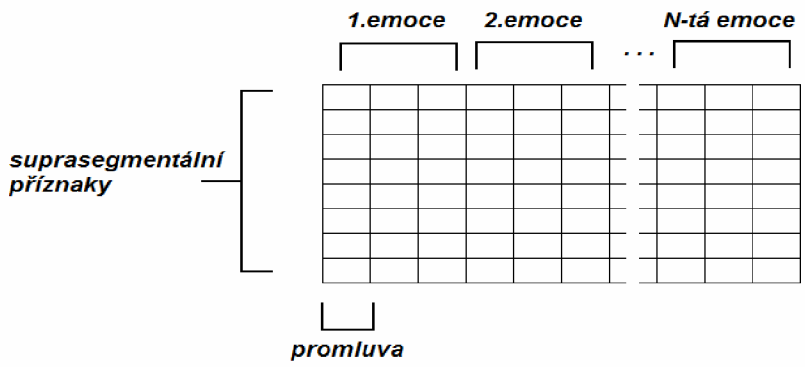
Z vybraných příznaků byla pro každou emoci a promluvu vytvořena matice suprasegmentálních příznaků, která obsahovala 28 suprasegmentálních příznaků. Počet sloupců koresponduje s počtem nahrávek. Např. matice suprasegmentálních příznaků pro neutralitu má tvar 28 řádků x 77 sloupců, což odpovídá 77 nahrávkám neutrální promluvy.

Z této matice vybereme 40 sloupců a použijeme je jako trénovací data a jiných 20 sloupců jako testovací data.

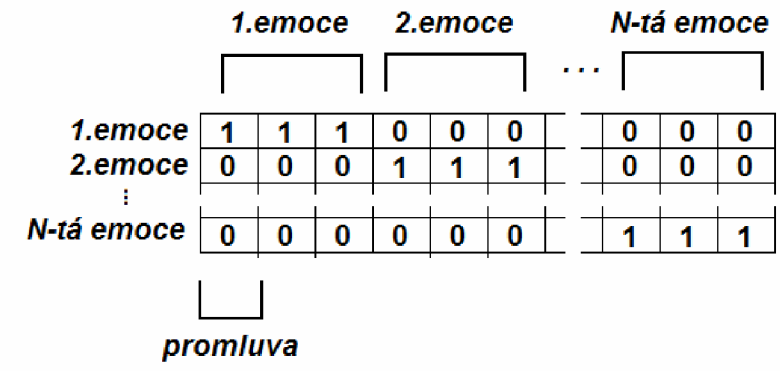
Trénovací matici vytvoříme tak, že za sebe seřadíme každých 40 sloupců suprasegmentálních příznaků jednotlivých sedmi emocí a získáme tak matici o rozměrech 28 řádků x 280 slupců. K této matici budeme ještě potřebovat matici target data. Tu vytvoříme

tak, že k jednotlivým sloupcům přiřadíme do odpovídajícího řádku správnou emoci, tzn. pro prvních 40 sloupců zadáme v prvním řádku samé jedničky, zbytek nuly, pro dalších 40 sloupců zadáme v druhém řádku samé jedničky a zbytek nuly atd. Až se naplní celá matice jedničkami a nulami. V našem případě vznikne matice 7 řádků x 280 sloupců.

Pro porovnání byla také vytvořena trénovací i testovací matice pro 4 emoce.



Obr. 7.4 Trénovací matice pro *N* emocí



Obr. 7.5 Target data matice pro *N* emocí

7.1.7 Testovací data

Testovací matici vytvoříme podobně jako trénovací, jen počet sloupců bude nižší. V našem případě 140 sloupců. Pokud budeme chtít testovat jen jednu promluvu bude mít logicky matice 28 řádků x 1 sloupec. Pro kontrolu si můžeme vytvořit i matici target data. Obdobným způsobem jako u trénovacích dat.

7.1.8 Klasifikace - NS

Klasifikovat budeme pomocí matlabovského toolboxu: nntool .

- **Vytvoření neuronové sítě**

Nejprve musíme vytvořit síť. Pro jednotlivé testování bude nutné mít sítě rozdílné. Pro rozpoznání čtyř emocí volíme typ dopředné sítě se zpětným šířením, trénovací algoritmus

trainrp, adaptivní funkce učení learngdm, počet vrstev 3, první vrstva 30 neuronů a přenosová fce tansig, druhá vrstva 30 neuronů a přenosová fce tansig, třetí vrstva 4 neurony a přenosová fce logsig.

Pro rozpoznávání sedmi emocí volíme typ sítě, trénovací algoritmus, adaptivní funkci učení, stejně jako v předchozím případě. Počet vrstev a neuronů bude měněn a zobrazeny výsledky testování viz níže.

- **Trénování neuronové sítě**

Pro čtyři emoce jsou jako vstupní data načtena trénovací 28x160 a target data 4x160 matice. Tato data zvolíme jako vstupní parametry vytvořené sítě a necháme síť trénovat.

Pro sedm emocí postupujeme obdobně, jen jako vstupní data načteme trénovací 28x280 matici a target data 7x280 matici.

- **Testování dat**

Na vstup sítě vložíme testovací matici. Výsledkem je matice čtyř respektive sedmi řádků a promluvám odpovídajících sloupců. Jednotlivá pole odpovídají pravděpodobnosti testovaných promluv.

7.1.9 Klasifikační matice – NS

Tato matice udává výsledná procenta úspěšného i neúspěšného rozpoznání jednotlivých emocí s ohledem na emoce ostatní. V horním rohu tabulky je rozložení neuronů. Na diagonále je procentuální úspěšnost správně určených emocí, v řádcích pak jejich chybné určení za emoci jinou.

30-30-4	Zlost	Smutek	Radost	Neutralita
Zlost	**59**	4	31	6
Smutek	3	**63**	10	24
Radost	60	10	**29**	1
Neutralita	2	21	7	**70**

Tab. 7.1 Klasifikační matice pro čtyři emoce, síť 30-30-4 neurony

50-50-4	Zlost	Smutek	Radost	Neutralita
Zlost	**60**	1	36	3
Smutek	5	**51**	19	25
Radost	63	0	**36**	1
Neutralita	4	6	10	**80**

Tab. 7.2 Klasifikační matice pro čtyři emoce, síť 50-50-4 neurony

100-100-4	Zlost	Smutek	Radost	Neutralita
Zlost	**58**	8	73	4
Smutek	3	**67**	0	13
Radost	38	2	**26**	1
Neutralita	1	23	1	**82**

Tab. 7.3 Klasifikační matice pro čtyři emoce, síť 100-100-4 neurony

150x7	Zlost	Smutek	Nuda	Strach	Radost	Neutralita	Znechucení
Zlost	**58**	0	0	22	8	1	11
Smutek	11	**44**	29	6	0	6	4
Nuda	0	1	**59**	4	5	25	6
Strach	18	0	14	**36**	16	4	12
Radost	46	0	0	18	**20**	0	16
Neutralita	1	4	50	6	4	**29**	6
Znechucení	7	0	1	30	19	8	**35**

Tab. 7.4 Klasifikační matice pro sedm emocí, síť 150-7 neuronů

30-50-50-7	Zlost	Smutek	Nuda	Strach	Radost	Neutralita	Znechucení
Zlost	**48**	2	0	9	33	1	7
Smutek	10	**28**	14	21	12	10	5
Nuda	1	15	**29**	7	10	33	5
Strach	14	6	3	**31**	16	4	26
Radost	61	4	0	7	**17**	0	11
Neutralita	6	12	31	7	4	**36**	4
Znechucení	14	5	1	39	17	1	**23**

Tab. 7.5 Klasifikační matice pro sedm emocí, síť 30-50-50-7 neuronů

100x150x150x1	Zlost	Smutek	Nuda	Strach	Radost	Neutralita	Znechucení
Zlost	**40**	0	0	35	20	0	5
Smutek	0	**34**	23	26	0	5	12
Nuda	1	17	**41**	0	4	26	11
Strach	12	0	1	**31**	21	10	25
Radost	40	5	0	16	**28**	0	11
Neutralita	0	5	43	4	0	**39**	9
Znechucení	7	0	2	21	14	14	**42**

Tab. 7.6 Klasifikační matice pro sedm emocí, síť 100-150-150-100-7 neuronů

7.2 GMM - Gaussian Mixture Models

GMM: Gaussian Mixture Models, Smíšené Gaussovy modely. Výkonný nástroj pro statistické modelování příznaků v příznakovém prostoru. Základní myšlenka GMM algoritmu vychází z modelování trénovacích příznaků jednou nebo více Gaussovými funkcemi rozložení pravděpodobnosti.

7.2.1 Gaussova funkce rozložení pravděpodobnosti

Jednorozměrná funkce rozložení pravděpodobnosti má rovnici:

$$f(x) = \frac{1}{\sigma\sqrt{2\cdot\pi}} \exp\left(-\frac{(x-\mu)^2}{2\cdot\sigma^2}\right), \tag{7.8}$$

kde μ je střední hodnota, σ^2 je rozptyl.

Obr. 7.6 Jednorozměrná Gaussova funkce rozložení hustoty pravděpodobnosti

Vícerozměrná Gaussova funkce rozložení pravděpodobnosti:

$$f(x) = \frac{1}{\sqrt{(2\cdot\pi)^d \det(\Sigma)}} \exp\left(-\frac{(x-\mu)^T \Sigma^{-1}(x-\mu)}{2}\right), \tag{7.9}$$

kde d je rozměr Gaussovy funkce rozložení pravděpodobnosti, Σ je kovarianční matice, μ je vektor středních hodnot. Kovarianční matice je následující:

$$\Sigma = \begin{bmatrix} E[(X_1-\mu_1)(X_1-\mu_1)] & E[(X_1-\mu_1)(X_2-\mu_2)] & \cdots & E[(X_1-\mu_1)(X_d-\mu_d)] \\ E[(X_2-\mu_2)(X_1-\mu_1)] & E[(X_2-\mu_2)(X_2-\mu_2)] & \cdots & E[(X_2-\mu_2)(X_d-\mu_d)] \\ \vdots & \vdots & \ddots & \vdots \\ E[(X_d-\mu_d)(X_1-\mu_1)] & E[(X_d-\mu_d)(X_2-\mu_2)] & \cdots & E[(X_d-\mu_d)(X_d-\mu_d)] \end{bmatrix}$$

Vektor středních hodnot :

$$\mu = [\mu_1 \quad \mu_2 \quad \cdots \quad \mu_d]$$

7.2.2 Gaussův smíšený model

Gaussův smíšený model (GMM) vzniká lineární kombinací více Gaussových funkcí rozložení pravděpodobnosti:

$$f(x) = \sum_{i=1}^{M} \alpha_i N_i(x), \tag{7.10}$$

M je počet Gaussových funkcí rozložení pravděpodobnosti, α_i váhovací parametr, $N_i(x)$ funkce rozložení pravděpodobnosti. Klasifikace GMM se vypočte podle následujícího vzorce:

$$f(x) = \sum_{i=1}^{M} \alpha_i \frac{1}{\sqrt{(2\cdot\pi)^d \det(\Sigma_i)}} \exp\left(-\frac{(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)}{2}\right) \tag{7.11}$$

Pro definici GMM bude třeba určit: kovarianční matici Σ, vektor středních hodnot μ, váhovací parametr α.

7.2.3 EM algoritmus

Parametry pro definici GMM určíme pomocí EM algoritmu. EM (Expectation-Maximization) je iterační algoritmus, který pracuje ve dvou krocích:

- Expectation (odhad): v tomto kroku se provede odhad parametrů GMM modelu a výpočet tzv. „likelihood“ funkce.
- Maximization (maximalizace): v tomto kroku jsou aktualizovány parametry GMM modelu za účelem maximalizace likelihood funkce.

Cílem EM algoritmu je najít takové parametry Θ^*, pro které je logaritmus likelihood funkce maximální:

$$\Theta^* = \arg\max_{\Theta}\left(\log L\left(\Theta|x\right)\right) \tag{7.12}$$

Logaritmus likelihood funkce je pro GMM definován jako:

$$\log L\left(\Theta|x\right) = \log\prod_{i=1}^{K} f\left(x_i|\Theta\right) = \log\prod_{i=1}^{K}\sum_{j=1}^{M}\alpha_j N_j\left(x_i|\theta_j\right), \tag{7.13}$$

kde K je počet trénovacích vektorů x, M je počet Gaussových fukcí ve směsi, $N(\ )$ je Gaussova funkce rozložení pravděpodobnosti, α je váhovací parametr, θ představuje parametry Gaussovy funkce rozložení pravděpodobnosti, tedy $\theta = (\mu, \Sigma)$, Θ představuje parametry GMM modelu, $(a_1, \ldots, \alpha_M, \theta_1, \ldots, \theta_M)$.

7.2.4 Inicializace parametrů

K inicializaci parametrů GMM modelu se využívá algoritmus K-means tzn. K-nejbližších sousedů. Snaha zařadit vstupní data do k klustrů $S = \{S_1, S_2, S_3, \ldots, S_k\}$

$\arg\min_{s}\sum_{i=1}^{k}\sum_{x_j \in S_i}\left\|x_j - \mu_i\right\|^2$, zde μ_i je střed klustru S_i

- Inicializace probíhá následovně: počet lustrů k - náhodně se vybere k vzorků, jako počáteční centroidy
- Přiřadí se každý vzorek nejbližšímu centroidu
- Obnoví se centroid
- Zopakuje se přiřazeni a obnovení centroidů

Proces trénování a testování je třeba několikrát zopakovat. Nahrávky použité pro trénování nesmí být použity pro testování. V této práci byla použita metoda nezávislá na mluvčím. To znamená, že pro trénovací a testovací skupinu byli vybráni rozdílní mluvčí. Část pro trénování a část pro testování byla rozdělena v poměru 8:2. Což odpovídá 10-ti rozdílným mluvčím. Pro následující testování byla použita metoda Cross validation. Což znamená, že z trénovací skupiny je odebrán jeden mluvčí, který je vyměněn s jedním mluvčím ze skupiny testovací. Tento postup je několikrát opakován a pokaždé je přepočítána klasifikační matice. Tímto způsobem se zajistí nezávislost na mluvčím.

7.2.5 Trénovací data

Pro všechny promluvy v databázi, rozdělené dle mluvčích, byly vypočteny všechny

sady příznaků (suprasegmentální, MFCC, formanty, atd). Jednotlivé úkoly jsou děleny dle těchto sad příznaků. Tzn. z jedné promluvy byly získány následující sady příznaků: 77 x suprasegmentálních příznaků, 16 x formanty, 60 x MFCC, 60 x MFCC do 1000Hz, 60 x MFCC do 800Hz, 60 x MFCC do 500Hz a to způsobem popsaným v kapitolách 4 a 6. Trénovací data pro jeden emoční stav mohou vypadat následovně. Př. Vezměme sadu suprasegmentálních příznaků, pak jedné promluvě odpovídá vektor (77x1). Pro 10 promluv stejné emoce (třídy) bude trénovací matice [77x10] atd. Databáze je roztříděna podle mluvčích, proto je možné do trénování i testování implementovat metodu Cross validation, což pro každé další trénování znamená nová trénovací data, získaná na základě výměny mluvčího.

7.2.6 Testovací data

Jsou získány z vypočtených sad příznaků metodou Cross validation. Z databáze mluvčích je 8 mluvčích vybráno pro trénování a zbylí 2 mluvčí pro testování. Při dalším trénování je vyměněn jeden z mluvčích z trénovací skupiny za jednoho mluvčího ze skupiny testovací. Tímto způsobem je možno pro každé nové trénování získávat jiné trénovací i testovací soubory.

7.2.7 Klasifikace – GMM

Klasifikovat budeme pomocí GMM toolboxu ”Copyright (c) Ian T Nabney (1996-2001)“

- **Vytvoření modelu**

Pomocí funkce „model = creategmmmodel(Data', ntr, ngauss)“ vytvoříme model jedné třídy (emočního stavu). Vstupními argumenty jsou: výše popsaná trénovací data jedné třídy, počet trénovacích cyklů, počet mixů Gaussových funkcí. Takovéto modely musí být vytvořeny (natrénovány) pro všech 7 emočních stavů.

- **Testování dat**

Pro testování dat byla použita funkce „`vysledek = gmmclassifier(model,Data')`“ Vstupem této funkce je natrénovaný model jedné třídy(emoce) a testovací data popsaná výše. Testovaná data necháme klasifikovat všemi 7-mi natrénovanými modely. Nejvyšší dosažený výsledek koresponduje s nejpravděpodobnějším emočním stavem.

Celá klasifikace je prováděna v několika cyklech, s ohledem na metodu nezávislou na mluvčím tzv. cross validation.

7.2.8 Klasifikační matice - GMM

Tato matice udává výsledná procenta úspěšného i neúspěšného rozpoznání jednotlivých emocí s ohledem na emoce ostatní. Na diagonále je procentuální úspěšnost správně klasifikovaných emocí. V řádcích je pak také v procentech udáno případné nesprávné určení jednotlivých emocí. V horním rohu tabulky je celková procentuální úspěšnost správného rozpoznávání emocí.

- **Suprasegmentální příznaky**

53,50%	zlost	nuda	znechucen	strach	radost	neutralita	smutek
zlost	**75**	1	1	5	15	0	0
nuda	2	**45**	1	17	3	26	3
znechucen	8	3	**25**	32	19	8	2
strach	18	1	3	**55**	10	10	0
radost	38	2	2	7	**38**	9	1
neutralita	0	23	3	10	3	**48**	10
smutek	0	8	4	12	0	11	**62**

Tab. 7.7 Klasifikační matice 77x suprasegmentální příznaky, směs 5x Gauss fcí

61,20%	zlost	nuda	znechucen	strach	radost	neutralita	smutek
zlost	**77**	0	1	5	15	0	0
nuda	1	**51**	3	12	0	19	10
znechucen	6	3	**43**	26	8	9	2
strach	13	0	7	**59**	10	7	1
radost	35	1	4	9	**45**	2	0
neutralita	0	12	1	12	0	**60**	13
smutek	0	1	2	12	0	8	**75**

Tab. 7.8 Klasifikační matice 28x suprasegmentální příznaky, směs 3x Gauss fcí

- **MFCC**

59,70%	zlost	nuda	znechucen	strach	radost	neutralita	smutek
zlost	**79**	0	1	5	13	0	0
nuda	0	**44**	3	6	2	33	9
znechucen	3	5	**46**	21	11	6	4
strach	15	2	8	**47**	10	8	6
radost	27	0	12	11	**43**	4	0
neutralita	0	27	4	5	0	**61**	0
smutek	0	8	4	4	0	5	**76**

Tab. 7.9 Klasifikační matice 60x MFCC , směs 3x Gauss fcí

56,20%	zlost	nuda	znechucen	strach	radost	neutralita	smutek
zlost	**70**	0	5	10	12	1	0
nuda	0	**63**	4	3	2	23	3
znechucen	15	3	**22**	25	19	10	3
strach	15	3	14	**45**	9	6	4
radost	28	0	16	13	**42**	0	0
neutralita	0	38	6	5	0	**47**	1
smutek	0	8	3	1	0	5	**81**

Tab. 7.10 Klasifikační matice 60x MFCC pásmo do 500Hz, směs 3x Gauss fcí

58,20%	zlost	nuda	znechucen	strach	radost	neutralita	smutek
zlost	**72**	0	3	11	11	0	0
nuda	0	**66**	3	4	0	22	2
znechucen	10	2	**36**	20	15	9	4
strach	15	5	13	**48**	9	4	3
radost	29	1	14	14	**38**	0	0
neutralita	0	30	9	6	0	**50**	2
smutek	0	10	5	6	0	6	**70**

Tab. 7.11 Klasifikační matice 60x MFCC pásmo do 800Hz, směs 5x Gauss fcí

56,10%	zlost	nuda	znechucen	strach	radost	neutralita	smutek
zlost	**70**	0	1	12	13	1	0
nuda	0	**66**	4	1	1	23	1
znechucen	8	6	**29**	18	19	13	4
strach	14	6	15	**44**	9	5	4
radost	27	2	10	20	**39**	0	0
neutralita	0	31	5	5	0	**55**	1
smutek	0	11	6	4	0	9	**67**

Tab. 7.12 Klasifikační matice 60x MFCC pásmo do 1000Hz, směs 3x Gauss fcí

- **Formanty**

39%	zlost	nuda	znechucen	strach	radost	neutralita	smutek
zlost	**59**	1	7	11	20	0	0
nuda	3	**26**	3	11	6	29	18
znechucen	5	8	**17**	11	29	16	10
strach	9	18	5	**32**	16	10	7
radost	30	3	17	15	**26**	4	2
neutralita	2	22	5	11	5	**42**	10
smutek	1	17	5	9	2	15	**47**

Tab. 7.13 Klasifikační matice 16x formanty, směs 7x Gauss fcí

- **Výběr příznaků**

66,00%	zlost	nuda	znechucen	strach	radost	neutralita	smutek
zlost	**85**	0	0	7	7	0	0
nuda	0	**50**	2	8	0	32	5
znechucen	0	0	**9**	59	9	9	13
strach	11	0	0	**68**	5	11	4
radost	33	0	0	14	**51**	0	0
neutralita	0	15	0	6	0	**76**	1
smutek	0	16	0	0	0	20	**64**

Tab. 7.14 Klasifikační matice 162x výběr příznaků, 1x Gauss fce, mluvčí muži

66,90%	zlost	nuda	znechucen	strach	radost	neutralita	smutek
zlost	**85**	0	2	7	4	0	0
nuda	0	**64**	3	4	1	23	3
znechucen	7	1	**51**	14	4	20	1
strach	22	1	13	**46**	4	3	7
radost	36	0	9	7	**46**	0	0
neutralita	0	21	1	0	0	**77**	0
smutek	1	4	6	4	0	1	**82**

Tab. 7.15 Klasifikační matice 162x výběr příznaků, 1x Gauss fce, mluvčí ženy

Nejlepší příznaky
'ztrRELMAX'
'ztrMEDIAN'
'zcrMEDIAN'
'ztrMEAN'
'shimmerRELMAX'
'jitterMEDIAN'
'ztrODCH'
'shimmerMEDIAN'
'plochostODCH'
'centroidMEAN'
'nhrRELMAX'
'spicatostMEAN'
'nhrMIN'
'shimmerMAX'
'centroidODCH'
'poklesMIN'

Tab. 7.16 Výběr nejlepších 16ti příznaků pro výslednou klasifikaci

Výběr příznaků pro jednotlivé klasifikační matice lze najít v přílohách. Detailní přehled všech počítaných klasifikačních matic, vybraných příznaků, všech pravděpodobností i porovnání mluvčích je k dispozici jako příloha v souboru klas_matice.xls na dodaném DVD.

Zde zobrazené výsledky jsou specificky vybrané s ohledem na obecnost výsledného řešení. V příloze je sestaveno pořadí nejvhodnějších příznaků pro klasifikaci všech sedmi emočních stavů.

8 ZÁVĚR

Cílem diplomové práce bylo navrhnout systém, který na základě analýzy řečové promluvy bude rozpoznávat emoční stavy mluvčího.

Základem pro tuto práci byla německá databáze emočních promluv o sedmi emočních stavech. Databáze obsahuje sedm emočních stavů: zlost, nuda, znechucení, strach, radost, neutralita, smutek. Proto bylo všech těchto sedm emočních stavů klasifikováno. Vzhledem k nepříliš vysokému počtu promluv pro jednotlivé emoce nebylo trénování a následné

testování statisticky dokonalé. Tento aspekt byl tedy z hlediska celé práce nejméně ovlivnitelný.

Po důkladném rozboru různých parametrů řečových signálů byly vypočteny důležité příznaky pro rozpoznávání, a to ze signálu, spektra signálu i jeho kepstra. Tyto příznaky byly rozděleny do několika skupin, které byly analyzovány každá zvlášť i všechny skupiny najednou. Jedná se o skupiny příznaků suprasegmentálních, MFCC, MFCC v určených kmitočtových pásmech, formanty.

V další části bylo potřeba zjistit kvalitu jednotlivých příznaků. Tento úkol byl opět rozdělen do několika částí. Kvalita příznaků byla počítána pomocí geometrické oddělitelnosti a pomocí Matlabovské funkce rankfeatures. Metoda geometrické oddělitelnosti umožnila zjistit kvalitu příznaku přes všechny uvažované emoce i všech možných dvojic emocí. Druhá metoda umožnila sestavit pro zvolenou emoci sestupné hodnocení příznaků pro oddělení. Zde se jako lepší zdála metoda rankfeatures, pomocí níž byly vybrány konečné nejvhodnější příznaky pro klasifikaci. Nejkvalitnějšími byla sada suprasegmentálních příznaků, z nichž bylo vybráno 16 pro výsledný program. Jsou to: ZTŘ, ZCR, jitter, shimmer, plochost, centroid, NHR, špičatost. Jako dostačující sada příznaků vyšly MFCC příznaky. Naopak pro klasifikaci nekvalitními příznaky jsou MFCC v určených kmitočtových pásmech i formanty.

K procesu samotné klasifikace byly využity neuronové sítě a GMM klasifikátor. V případě neuronových sítí byly využity pouze suprasegmentální příznaky a jejich výběr. Návrhy neuronové sítě se lišily v počtu vrstev i počtu neuronů. Jednotlivé třídy databáze byly jednoduše rozděleny na část trénovací a část testovací. Vytvořené sítě byly jednorázově natrénovány a při testování byly vytvořeny klasifikační matice. Protože průměrné procento rozpoznání emocí bylo 40%, proto není další testování neuronových sítí tak rozsáhlé.

Rozpoznávání pomocí GMM funguje na principu natrénování jednotlivých klasifikačních modelů a následné vybrání nejpodobnějšího testované promluvě. U této klasifikace byla při testování použita metoda nezávislá na mluvčím tzv. cross validation. Její smysl je takový, že pro trénovací a testovací data jsou voleni jiní mluvčí a při každém dalším testování je jiné složení mluvčích. U tohoto klasifikátoru byly testovány všechny sady příznaků zvlášť i dohromady, byly vybírány nejlepší příznaky a pro každé testování byla vypočítána klasifikační matice a její průměrná pravděpodobnost. Pro skupinu MFCC příznaků se pravděpodobnost správně rozpoznané emoce pohybovala kolem 57%, pro formanty jen 39%. Nejlepšího výsledku bylo dosaženo výběrem 162 příznaků ze všech 333 zkoumaných a to 66%. Detailní vyhodnocení jednotlivých testování je v tabulkách na přiloženém DVD.

Nejvhodnějším kompromisním řešením se jeví výběr 16ti suprasegmentálních příznaků s průměrným rozpoznáním 61%, a to s ohledem na výpočetní složitost a časovou náročnost.

Výsledkem této diplomové práce je návrh systému v prostředí Matlab, který s úspěšností 61% rozpoznává 7 emočních stavů.

Seznam použitých zdrojů:

[ČER-08] ČERNOCKÝ, J.: Zpracování řečových signálů. Studijní opora FIT VUT Brno. 2008. http://www.fit.vutbr.cz/~cernocky/speech/.cs

[HER-94] HERINGOVÁ, B., HORA, P.: Matlab 4.0 - popis grafického systému, grafická nadstavba a práce se soubory: Díl I. uživatelský manuál, díl II. referenční manuál. Technical Report 129VP, Institut technologie a spolehlivosti Západočeská univerzita v Plzni, 1994.

[JIR-02] JIRSÍK, V., HRÁČEK, P.: Neuronové sítě, expertní systémy a rozpoznávání řeči : Skripta VUT Brno. 2002. 106 s.

[KRČ-07] KRČMOVÁ, M.: Fonetika. Skripta Filosofická fakulta MU Brno. 2007. http://is.muni.cz/elportal/estud/ff/js07/fonetika/materialy/index.html

[NEU-08] Neural network toolbox[online].© 1984-2008. http://www.mathworks.com/access/helpdesk/help/toolbox/nnet/.

[PSU-06] PSUTKA, J., MÜLLER, L., MATOUŠEK, J., RADOVÁ, V.: Mluvíme s počítačem česky. Praha: Academia, 2006. ISBN 80-200-1309-1

[SIG-00] SIGMUND, M.: Analýza řečových signálů : přednášky. 1. vyd. Brno : MJ servis, s.r.o., 2000. 86 s. ISBN 80-214-1783-8.

Seznam použitých zkratek, veličin, symbolů:

F_0 ..... frekvence základního tónu řeči[Hz]
F_x ..... formant[Hz]
$s(t)$ ..... řečový signál se spojitým časem
$s(n)$ ..... řečový signál s diskrétním časem
X ..... vektor příznaků
f_{vz} ..... vzorkovací kmitočet
f_{dp} ..... kmitočet dolní propusti
m ..... počet kvantizačních úrovní
P_S ..... maximální výkon signálu
P_P ..... střední výkon poruch
B ..... počet bitů
Δ ..... kvantizační krok
N ..... počet vzorků v segmentu
$w(n)$ ..... váhovací okno
DFT ..... diskrétní Fourierova transformace
T_{vz} ..... perioda vzorkovacího kmitočtu
ZCR ..... počet průchodů nulovou rovinou
$x(n)$ ..... segment řeči
l_{seg} ..... délka segmentu
$sign\ x(n)$ ..... znaménková funkce
E ..... krátkodobá energie
IDFT ..... inverzní diskrétní Fourierova transformace
$c(n)$ ..... kepstrum
$R(m)$ ..... autokorelační funkce
ZTŘ ..... základní tón řeči
p ..... řád predikce LPC analýzy
$H(z)$ ..... frekvenční charakteristika filtru
LPC ..... lineární predikční analýza
z ..... komplexní číslo
φ ..... argument komplexního čísla
j ..... kolísání základní periody
T_0 ..... perioda základního tónu
h ..... kolísání amplitudy
A ..... rozkmit amplitudy
$\omega_{\max}$ ..... maximální kmitočet
c_k ..... koeficient Fourierovy transformace
$S(\omega)$ ..... spektrum signálu

NHR ..... poměr šumu a harmonické složky
μ ..... spektrální centroid
σ^2 ..... spektrální rozptyl
m_x ..... moment x-tého řádu
γ_1 ..... šikmost
γ_2 ..... špičatost
o ..... spektrální sklon
e ..... spektrální plochost
MFCC ..... melovské kepstrální koeficienty
FFT ..... rychlá Fourierova transformace
$c_m(j)$ ..... MFCC koeficienty
$M(f,i)$ ..... funkce MFCC filtru
$Q(x_i)$ ..... geometrická oddělitelnost
S ..... aritmetická střední hodnota
S_v ..... kvadrát rozptylu jedné třídy
D ..... aritmetická střední hodnota vzdálenosti
D_v ..... kvadrát vzdálenosti
p_i ..... vstupní vektor neuronu
$w_{1,i}$ ..... váhy neuronu
b ..... práh neuronu
a ..... výstup neuronu
δ_j^s ..... chyby jednotlivých neuronů
$f(x)$ ..... Gaussova funkce rozložení pravděpodobnosti
d ..... rozměr Gaussovy funkce
Σ ..... kovarianční matice
M ..... počet Gaussových funkcí
α_i ..... váhovací parametr
$N_i(x)$ ..... funkce rozdělení pravděpodobnosti
Θ^* ..... parametr EM algoritmu
$\log L(\Theta|x)$ ..... logaritmus likelihood
K ..... počet trénovacích vektorů
θ ..... parametry Gaussovy funkce rozložení pravděpodobnosti
k ..... počet klustrů
S_i ..... klustr

Seznam obrázků:

Seznam tabulek:

Seznam příloh:

Příloha 1: Výběr suprasegmentálních příznaků

Příloha 2: Výběr 162 nejlepších příznaků podle nejlepší pravděpodobnosti

Příloha 3: Příznaky seřazené podle funkce rankfeatures, 77 suprasegmentálních příznaků

Příloha 4: Příznaky seřazené podle funkce rankfeatures, všech 333 příznaků, 1.část

Příloha 5: Příznaky seřazené podle funkce rankfeatures, všech 333 příznaků, 2.část

Příloha 6: Příznaky seřazené podle funkce rankfeatures, všech 333 příznaků, 3.část

Příloha 1:

Výběr suprasegmentálních příznaků			
28 suprasegmentálních příznaků	77 suprasegmentálních příznaků		
'nesoumernostMAXXXX'	'nesoumernostMAXXXX'	'zcrMAX'	'jitterMAX'
'nesoumernostMIN'	'nesoumernostMIN'	'zcrMIN'	'jitterMIN'
'nesoumernostMEAN'	'nesoumernostMEAN'	'zcrMEAN'	'jitterMEAN'
'nesoumernostMEDIAN'	'nesoumernostMEDIAN'	'zcrMEDIAN'	'jitterMEDIAN'
'nesoumernostODCH'	'nesoumernostODCH'	'zcrODCH'	'jitterODCH'
'spicatostMIN'	'nesoumernostRELMAX'	'zcrRELMAX'	'jitterRELMAX'
'ztrMAX'	'nesoumernostRELMIN'	'zcrRELMIN'	'jitterRELMIN'
'ztrMIN'	'spicatostMAX'	'centroidMAX'	'shimmerMAX'
'ztrMEAN'	'spicatostMIN'	'centroidMIN'	'shimmerMIN'
'ztrMEDIAN'	'spicatostMEAN'	'centroidMEAN'	'shimmerMEAN'
'ztrODCH'	'spicatostMEDIAN'	'centroidMEDIAN'	'shimmerMEDIAN'
'ztrRELMIN'	'spicatostODCH'	'centroidODCH'	'shimmerODCH'
'nhrMEAN'	'spicatostRELMAX'	'centroidRELMAX'	'shimmerRELMAX'
'zcrMIN'	'spicatostRELMIN'	'centroidRELMIN'	'shimmerRELMIN'
'zcrMEDIAN'	'ztrMAX'	'energieMAX'	'poklesMAX'
'zcrODCH'	'ztrMIN'	'energieMIN'	'poklesMIN'
'energieMIN'	'ztrMEAN'	'energieMEAN'	'poklesMEAN'
'energieMEAN'	'ztrMEDIAN'	'energieMEDIAN'	'poklesMEDIAN'
'plochostMAX'	'ztrODCH'	'energieODCH'	'poklesODCH'
'plochostODCH'	'ztrRELMAX'	'energieRELMAX'	'poklesRELMAX'
'jitterMEDIAN'	'ztrRELMIN'	'energieRELMIN'	'poklesRELMIN'
'jitterRELMAX'	'nhrMAX'	'plochostMAX'	
'shimmerMAX'	'nhrMIN'	'plochostMIN'	
'shimmerMEAN'	'nhrMEAN'	'plochostMEAN'	
'shimmerMEDIAN'	'nhrMEDIAN'	'plochostMEDIAN'	
'shimmerRELMAX'	'nhrODCH'	'plochostODCH'	
'poklesMEAN'	'nhrRELMAX'	'plochostRELMAX'	
'poklesMEDIAN'	'nhrRELMIN'	'plochostRELMIN'	

Příloha 2:

Výběr 162 nejlepších příznaků podle nejlepší pradvděpodobnosti			
162 nejlepších příznaků			
'nesoumernostMAXXXX'	'4.usek,1.mfcc'	'3.usek,2.mfcc500'	'3.usek,8.mfcc800'
'nesoumernostMIN'	'4.usek,3.mfcc'	'3.usek,3.mfcc500'	'3.usek,11.mfcc800'
'nesoumernostMEAN'	'4.usek,5.mfcc'	'3.usek,4.mfcc500'	'3.usek,12.mfcc800'
'nesoumernostMEDIAN'	'4.usek,7.mfcc'	'3.usek,5.mfcc500'	'3.usek,13.mfcc800'
'nesoumernostODCH'	'4.usek,10.mfcc'	'3.usek,6.mfcc500'	'4.usek,2.mfcc800'
'spicatostMIN'	'4.usek,14.mfcc'	'3.usek,7.mfcc500'	'4.usek,3.mfcc800'
'ztrMAX'	'1.usek,1.formant'	'3.usek,8.mfcc500'	'4.usek,6.mfcc800'
'ztrMIN'	'1.usek,2.formant'	'3.usek,9.mfcc500'	'4.usek,7.mfcc800'
'ztrMEAN'	'1.usek,3.formant'	'3.usek,10.mfcc500'	'4.usek,8.mfcc800'
'ztrMEDIAN'	'2.usek,1.formant'	'3.usek,11.mfcc500'	'4.usek,10.mfcc800'
'ztrODCH'	'3.usek,1.formant'	'3.usek,12.mfcc500'	'4.usek,11.mfcc800'
'ztrRELMIN'	'3.usek,3.formant'	'3.usek,13.mfcc500'	'1.usek,1.mfcc1000'
'nhrMEAN'	'3.usek,4.formant'	'3.usek,14.mfcc500'	'1.usek,2.mfcc1000'
'zcrMIN'	'4.usek,1.formant'	'3.usek,15.mfcc500'	'2.usek,1.mfcc1000'
'zcrMEDIAN'	'4.usek,3.formant'	'4.usek,1.mfcc500'	'2.usek,2.mfcc1000'
'zcrODCH'	'1.usek,1.mfcc500'	'4.usek,2.mfcc500'	'2.usek,6.mfcc1000'
'energieMIN'	'1.usek,2.mfcc500'	'4.usek,3.mfcc500'	'2.usek,7.mfcc1000'
'energieMEAN'	'1.usek,3.mfcc500'	'4.usek,4.mfcc500'	'2.usek,12.mfcc1000'
'plochostMAX'	'1.usek,4.mfcc500'	'4.usek,5.mfcc500'	'3.usek,1.mfcc1000'
'plochostODCH'	'1.usek,5.mfcc500'	'4.usek,6.mfcc500'	'3.usek,2.mfcc1000'
'jitterMEDIAN'	'1.usek,6.mfcc500'	'4.usek,7.mfcc500'	'3.usek,3.mfcc1000'
'jitterRELMAX'	'1.usek,7.mfcc500'	'4.usek,8.mfcc500'	'4.usek,12.mfcc1000'
'shimmerMAX'	'1.usek,8.mfcc500'	'4.usek,9.mfcc500'	'4.usek,14.mfcc1000'
'shimmerMEAN'	'1.usek,9.mfcc500'	'4.usek,10.mfcc500'	'4.usek,15.mfcc1000'
'shimmerMEDIAN'	'1.usek,10.mfcc500'	'4.usek,11.mfcc500'	
'shimmerRELMAX'	'1.usek,11.mfcc500'	'4.usek,12.mfcc500'	
'poklesMEAN'	'1.usek,12.mfcc500'	'4.usek,13.mfcc500'	
'poklesMEDIAN'	'1.usek,13.mfcc500'	'4.usek,14.mfcc500'	
'1.usek,1.mfcc'	'1.usek,14.mfcc500'	'4.usek,15.mfcc500'	
'1.usek,2.mfcc'	'1.usek,15.mfcc500'	'1.usek,1.mfcc800'	
'1.usek,5.mfcc'	'2.usek,1.mfcc500'	'1.usek,2.mfcc800'	
'1.usek,6.mfcc'	'2.usek,2.mfcc500'	'1.usek,6.mfcc800'	
'1.usek,7.mfcc'	'2.usek,3.mfcc500'	'1.usek,7.mfcc800'	
'2.usek,1.mfcc'	'2.usek,4.mfcc500'	'2.usek,1.mfcc800'	
'2.usek,2.mfcc'	'2.usek,5.mfcc500'	'2.usek,4.mfcc800'	
'2.usek,3.mfcc'	'2.usek,6.mfcc500'	'2.usek,5.mfcc800'	
'2.usek,4.mfcc'	'2.usek,7.mfcc500'	'2.usek,6.mfcc800'	
'2.usek,9.mfcc'	'2.usek,8.mfcc500'	'2.usek,7.mfcc800'	
'3.usek,1.mfcc'	'2.usek,9.mfcc500'	'2.usek,8.mfcc800'	
'3.usek,3.mfcc'	'2.usek,10.mfcc500'	'3.usek,1.mfcc800'	
'3.usek,4.mfcc'	'2.usek,11.mfcc500'	'3.usek,2.mfcc800'	
'3.usek,5.mfcc'	'2.usek,12.mfcc500'	'3.usek,3.mfcc800'	
'3.usek,6.mfcc'	'2.usek,13.mfcc500'	'3.usek,4.mfcc800'	
'3.usek,7.mfcc'	'2.usek,14.mfcc500'	'3.usek,5.mfcc800'	
'3.usek,8.mfcc'	'2.usek,15.mfcc500'	'3.usek,6.mfcc800'	
'3.usek,10.mfcc'	'3.usek,1.mfcc500'	'3.usek,7.mfcc800'	

Příloha 3:

Příznaky seřazené podle rank features					
77 suprasegmentálních příznaků					
pořadí	součet umístění	příznak	pořadí	součet umístění	příznak
1	108	'centroidMEAN'	46	329	'nhrRELMIN'
2	108	'nhrRELMAX'	47	330	'zcrMAX'
3	110	'centroidMEDIAN'	48	335	'zcrODCH'
4	144	'nhrMEDIAN'	49	338	'zcrRELMIN'
5	145	'spicatostMEAN'	50	341	'spicatostMIN'
6	153	'centroidRELMAX'	51	348	'plochostRELMAX'
7	154	'nesoumernostMEAN'	52	352	'shimmerMAX'
8	155	'nesoumernostMEDIAN'	53	354	'shimmerMIN'
9	155	'zcrMEAN'	54	356	'poklesMAX'
10	157	'spicatostMEDIAN'	55	357	'shimmerODCH'
11	163	'ztrRELMAX'	56	358	'centroidMAX'
12	170	'nhrMEAN'	57	358	'centroidODCH'
13	176	'ztrMEDIAN'	58	363	'nesoumernostMIN'
14	178	'zcrMEDIAN'	59	371	'centroidRELMIN'
15	178	'ztrMEAN'	60	378	'energieODCH'
16	183	'zcrRELMAX'	61	378	'jitterMAX'
17	195	'shimmerRELMAX'	62	384	'jitterRELMAX'
18	210	'jitterMEDIAN'	63	401	'poklesMIN'
19	218	'ztrODCH'	64	409	'jitterMEAN'
20	225	'spicatostODCH'	65	412	'shimmerRELMIN'
21	232	'shimmerMEDIAN'	66	412	'spicatostRELMAX'
22	234	'nesoumernostODCH'	67	420	'energieMAX'
23	237	'plochostODCH'	68	430	'nhrODCH'
24	238	'nhrMIN'	69	433	'plochostMEDIAN'
25	251	'energieRELMAX'	70	437	'spicatostRELMIN'
26	251	'shimmerMEAN'	71	465	'poklesRELMIN'
27	254	'plochostMAX'	72	469	'nhrMAX'
28	254	'ztrMIN'	73	473	'poklesRELMAX'
29	259	'centroidMIN'			
30	260	'nesoumernostMAXXXX'			
31	271	'plochostMEAN'			
32	275	'spicatostMAX'			
33	279	'energieMEDIAN'			
34	281	'poklesODCH'			
35	285	'poklesMEAN'			
36	287	'poklesMEDIAN'			
37	288	'nesoumernostRELMAX'			
38	295	'ztrRELMIN'			
39	297	'ztrMAX'			
40	300	'jitterODCH'			
41	301	'energieRELMIN'			
42	302	'nesoumernostRELMIN'			
43	308	'energieMEAN'			
44	312	'energieMIN'			
45	324	'zcrMIN'			

Příloha 4:

Příznaky seřazené podle rank features všech 333 příznaků								
poř.	souč.	příznak	poř.	souč	příznak	poř.	souč.	příznak
1	251	'nhrRELMAX'	46	714	'2.usek,2.mfcc'	91	886	'4.usek,3.mfcc'
2	257	'centroidMEAN'	47	715	'1.usek,5.mfcc1000'	92	887	'1.usek,2.mfcc500'
3	308	'centroidMEDIAN'	48	722	'1.usek,5.mfcc800'	93	889	'3.usek,1.mfcc'
4	364	'4.usek,2.mfcc500'	49	723	'2.usek,6.mfcc'	94	894	'plochostMEAN'
5	388	'2.usek,3.mfcc'	50	727	'3.usek,6.mfcc'	95	896	'2.usek,6.mfcc1000'
6	390	'spicatostMEAN'	51	732	'1.usek,2.mfcc'	96	896	'3.usek,1.mfcc1000'
7	394	'zcrMEAN'	52	732	'3.usek,11.mfcc'	97	896	'3.usek,4.mfcc'
8	406	'spicatostMEDIAN'	53	735	'3.usek,6.mfcc1000'	98	898	'3.usek,7.mfcc'
9	408	'centroidRELMAX'	54	738	'4.usek,12.mfcc1000'	99	902	'2.usek,4.mfcc500'
10	415	'nhrMEDIAN'	55	742	'3.usek,6.mfcc800'	100	903	'2.usek,6.mfcc800'
11	424	'nesoumernostMEAN'	56	744	'1.usek,6.mfcc'	101	903	'3.usek,1.mfcc500'
12	425	'nesoumernostMEDIAN'	57	745	'2.usek,2.mfcc1000'	102	904	'2.usek,4.mfcc'
13	434	'ztrRELMAX'	58	745	'4.usek,12.mfcc800'	103	904	'3.usek,13.mfcc'
14	441	'4.usek,2.mfcc1000'	59	747	'3.usek,5.mfcc'	104	910	'3.usek,1.mfcc800'
15	448	'4.usek,2.mfcc800'	60	749	'plochostODCH'	105	919	'3.usek,12.mfcc'
16	452	'nhrMEAN'	61	752	'2.usek,2.mfcc800'	106	927	'3.usek,8.mfcc'
17	495	'3.usek,9.mfcc'	62	757	'4.usek,6.mfcc1000'	107	930	'1.usek,9.mfcc'
18	512	'zcrRELMAX'	63	757	'ztrMIN'	108	935	'2.usek,1.mfcc'
19	519	'2.usek,9.mfcc'	64	764	'4.usek,6.mfcc800'	109	941	'ztrRELMIN'
20	535	'4.usek,7.mfcc'	65	766	'energieRELMAX'	110	942	'2.usek,1.mfcc1000'
21	549	'ztrMEDIAN'	66	766	'nesoumernostMAXXXX'	111	947	'2.usek,3.mfcc500'
22	558	'ztrMEAN'	67	771	'4.usek,10.mfcc'	112	949	'2.usek,1.mfcc500'
23	563	'shimmerRELMAX'	68	772	'4.usek,7.mfcc1000'	113	953	'1.usek,4.mfcc500'
24	587	'3.usek,3.mfcc'	69	779	'4.usek,7.mfcc800'	114	953	'nesoumernostRELMAX'
25	596	'zcrMEDIAN'	70	780	'2.usek,2.mfcc500'	115	953	'poklesMEDIAN'
26	602	'jitterMEDIAN'	71	784	'shimmerMEDIAN'	116	956	'2.usek,1.mfcc800'
27	603	'2.usek,11.mfcc'	72	789	'2.usek,10.mfcc'	117	960	'energieMEDIAN'
28	616	'3.usek,10.mfcc'	73	793	'3.usek,4.mfcc500'	118	966	'2.usek,1.formant'
29	635	'4.usek,6.mfcc'	74	795	'4.usek,3.mfcc1000'	119	978	'4.usek,8.mfcc1000'
30	645	'4.usek,5.mfcc500'	75	799	'centroidMIN'	120	985	'4.usek,8.mfcc800'
31	647	'1.usek,1.mfcc'	76	802	'4.usek,3.mfcc800'	121	989	'poklesMEAN'
32	649	'1.usek,3.mfcc'	77	804	'4.usek,9.mfcc500'	122	996	'nesoumernostRELMIN'
33	649	'3.usek,2.mfcc500'	78	815	'2.usek,7.mfcc'	123	1005	'poklesODCH'
34	654	'1.usek,1.mfcc1000'	79	822	'1.usek,5.mfcc'	124	1009	'energieRELMIN'
35	658	'spicatostODCH'	80	827	'4.usek,9.mfcc'	125	1024	'jitterODCH'
36	661	'1.usek,1.mfcc500'	81	832	'3.usek,5.mfcc1000'	126	1025	'ztrMAX'
37	668	'1.usek,1.mfcc800'	82	839	'3.usek,5.mfcc800'	127	1043	'2.usek,3.mfcc1000'
38	668	'2.usek,5.mfcc'	83	841	'4.usek,5.mfcc'	128	1046	'energieMIN'
39	674	'nesoumernostODCH'	84	842	'spicatostMAX'	129	1049	'3.usek,4.formant'
40	684	'ztrODCH'	85	843	'2.usek,5.mfcc1000'	130	1050	'2.usek,3.mfcc800'
41	689	'nhrMIN'	86	850	'2.usek,5.mfcc800'	131	1061	'1.usek,11.mfcc'
42	691	'3.usek,2.mfcc1000'	87	865	'shimmerMEAN'	132	1068	'2.usek,12.mfcc'
43	696	'1.usek,2.mfcc1000'	88	869	'3.usek,2.mfcc'	133	1072	'3.usek,3.mfcc1000'
44	698	'3.usek,2.mfcc800'	89	879	'4.usek,2.mfcc'	134	1072	'4.usek,4.mfcc500'
45	703	'1.usek,2.mfcc800'	90	884	'plochostMAX'	135	1077	'3.usek,3.formant'

Příloha 5:

Příznaky seřazené podle rank features všech 333 příznaků								
poř.	souč.	příznak	poř.	souč	příznak	poř.	souč.	příznak
136	1079	'3.usek,3.mfcc800'	181	1265	'1.usek,4.mfcc800'	226	1433	'3.usek,14.mfcc500'
137	1096	'energieMEAN'	182	1271	'1.usek,2.formant'	227	1436	'4.usek,14.mfcc1000'
138	1108	'2.usek,13.mfcc'	183	1279	'nesoumernostMIN'	228	1441	'4.usek,2.formant'
139	1111	'3.usek,5.mfcc500'	184	1281	'poklesMAX'	229	1443	'4.usek,14.mfcc800'
140	1117	'2.usek,4.formant'	185	1282	'1.usek,6.mfcc1000'	230	1449	'1.usek,10.mfcc500'
141	1121	'4.usek,1.mfcc'	186	1288	'2.usek,14.mfcc1000'	231	1473	'4.usek,5.mfcc1000'
142	1122	'zcrODCH'	187	1288	'4.usek,4.mfcc'	232	1476	'4.usek,12.mfcc'
143	1126	'4.usek,7.mfcc500'	188	1289	'1.usek,6.mfcc800'	233	1478	'4.usek,15.mfcc1000'
144	1128	'4.usek,1.mfcc1000'	189	1290	'3.usek,7.mfcc1000'	234	1480	'4.usek,5.mfcc800'
145	1133	'nhrRELMIN'	190	1292	'plochostRELMAX'	235	1485	'4.usek,15.mfcc800'
146	1135	'4.usek,1.mfcc500'	191	1294	'3.usek,11.mfcc1000'	236	1489	'4.usek,15.mfcc500'
147	1137	'1.usek,4.formant'	192	1295	'2.usek,14.mfcc800'	237	1491	'1.usek,12.mfcc'
148	1139	'4.usek,13.mfcc'	193	1297	'3.usek,7.mfcc800'	238	1497	'2.usek,8.mfcc500'
149	1139	'zcrMIN'	194	1301	'3.usek,1.formant'	239	1498	'2.usek,9.mfcc1000'
150	1142	'4.usek,1.mfcc800'	195	1301	'3.usek,11.mfcc800'	240	1501	'1.usek,15.mfcc'
151	1159	'1.usek,1.formant'	196	1303	'4.usek,11.mfcc500'	241	1505	'2.usek,9.mfcc800'
152	1163	'shimmerODCH'	197	1306	'4.usek,11.mfcc'	242	1508	'1.usek,14.mfcc'
153	1165	'4.usek,6.mfcc500'	198	1308	'2.usek,15.mfcc'	243	1510	'3.usek,2.formant'
154	1166	'3.usek,15.mfcc'	199	1325	'energieODCH'	244	1514	'shimmerRELMIN'
155	1172	'2.usek,4.mfcc1000'	200	1328	'1.usek,9.mfcc500'	245	1522	'poklesMIN'
156	1179	'2.usek,4.mfcc800'	201	1332	'centroidMAX'	246	1530	'2.usek,12.mfcc1000'
157	1185	'spicatostMIN'	202	1337	'centroidRELMIN'	247	1533	'2.usek,9.mfcc500'
158	1187	'4.usek,13.mfcc500'	203	1338	'1.usek,3.mfcc500'	248	1533	'3.usek,7.mfcc500'
159	1192	'4.usek,10.mfcc1000'	204	1340	'4.usek,12.mfcc500'	249	1537	'2.usek,12.mfcc800'
160	1199	'4.usek,10.mfcc800'	205	1346	'2.usek,3.formant'	250	1538	'jitterMEAN'
161	1200	'2.usek,8.mfcc'	206	1353	'3.usek,4.mfcc1000'	251	1540	'2.usek,6.mfcc500'
162	1212	'zcrRELMIN'	207	1357	'1.usek,13.mfcc'	252	1544	'1.usek,7.mfcc500'
163	1214	'2.usek,10.mfcc500'	208	1357	'3.usek,13.mfcc1000'	253	1549	'2.usek,5.mfcc500'
164	1225	'4.usek,15.mfcc'	209	1359	'3.usek,10.mfcc500'	254	1554	'1.usek,14.mfcc1000'
165	1231	'4.usek,13.mfcc1000'	210	1360	'3.usek,4.mfcc800'	255	1561	'1.usek,14.mfcc800'
166	1234	'4.usek,8.mfcc'	211	1362	'4.usek,1.formant'	256	1568	'1.usek,10.mfcc1000'
167	1234	'zcrMAX'	212	1364	'3.usek,13.mfcc800'	257	1568	'2.usek,7.mfcc500'
168	1235	'1.usek,10.mfcc'	213	1365	'jitterMAX'	258	1572	'4.usek,8.mfcc500'
169	1235	'1.usek,4.mfcc'	214	1367	'1.usek,3.formant'	259	1574	'1.usek,7.mfcc1000'
170	1238	'4.usek,13.mfcc800'	215	1377	'2.usek,14.mfcc'	260	1575	'1.usek,10.mfcc800'
171	1244	'centroidODCH'	216	1379	'2.usek,2.formant'	261	1579	'1.usek,8.mfcc'
172	1245	'3.usek,3.mfcc500'	217	1381	'2.usek,13.mfcc1000'	262	1581	'1.usek,7.mfcc800'
173	1248	'shimmerMAX'	218	1388	'2.usek,13.mfcc800'	263	1589	'1.usek,12.mfcc1000'
174	1255	'1.usek,13.mfcc1000'	219	1393	'jitterRELMAX'	264	1594	'spicatostRELMAX'
175	1255	'1.usek,3.mfcc1000'	220	1395	'1.usek,7.mfcc'	265	1596	'1.usek,12.mfcc800'
176	1256	'shimmerMIN'	221	1398	'3.usek,14.mfcc'	266	1615	'4.usek,3.mfcc500'
177	1258	'1.usek,4.mfcc1000'	222	1401	'4.usek,4.mfcc1000'	267	1623	'energieMAX'
178	1262	'1.usek,13.mfcc800'	223	1408	'4.usek,4.mfcc800'	268	1626	'1.usek,5.mfcc500'
179	1262	'1.usek,3.mfcc800'	224	1420	'3.usek,9.mfcc500'	269	1628	'2.usek,7.mfcc1000'
180	1262	'4.usek,10.mfcc500'	225	1425	'3.usek,8.mfcc500'	270	1635	'2.usek,7.mfcc800'

Příloha 6:

Příznaky seřazené podle rank features všech 333 příznaků								
poř.	souč.	příznak	poř.	souč	příznak	poř.	souč.	příznak
271	1636	'1.usek,9.mfcc1000'	316	1823	'4.usek,14.mfcc500'			
272	1636	'nhrODCH'	317	1848	'3.usek,13.mfcc500'			
273	1636	'plochostMEDIAN'	318	1855	'nhrMAX'			
274	1643	'1.usek,9.mfcc800'	319	1858	'3.usek,6.mfcc500'			
275	1643	'3.usek,10.mfcc1000'	320	1868	'poklesRELMIN'			
276	1650	'3.usek,10.mfcc800'	321	1890	'poklesRELMAX'			
277	1654	'3.usek,12.mfcc500'	322	1892	'4.usek,14.mfcc'			
278	1654	'4.usek,9.mfcc1000'	323	1907	'2.usek,15.mfcc1000'			
279	1661	'4.usek,9.mfcc800'	324	1914	'2.usek,15.mfcc800'			
280	1662	'3.usek,9.mfcc1000'	325	1949	'3.usek,15.mfcc1000'			
281	1669	'2.usek,10.mfcc1000'	326	1956	'3.usek,15.mfcc800'			
282	1669	'3.usek,9.mfcc800'	327	1968	'1.usek,14.mfcc500'			
283	1676	'2.usek,10.mfcc800'	328	2089	'1.usek,15.mfcc1000'			
284	1679	'1.usek,8.mfcc500'	329	2096	'1.usek,15.mfcc800'			
285	1681	'3.usek,12.mfcc1000'						
286	1685	'3.usek,14.mfcc1000'						
287	1687	'2.usek,11.mfcc500'						
288	1688	'3.usek,12.mfcc800'						
289	1689	'2.usek,12.mfcc500'						
290	1689	'2.usek,8.mfcc1000'						
291	1692	'2.usek,14.mfcc500'						
292	1692	'3.usek,14.mfcc800'						
293	1696	'2.usek,8.mfcc800'						
294	1704	'1.usek,13.mfcc500'						
295	1707	'4.usek,3.formant'						
296	1713	'2.usek,13.mfcc500'						
297	1716	'spicatostRELMIN'						
298	1717	'1.usek,6.mfcc500'						
299	1723	'2.usek,15.mfcc500'						
300	1725	'3.usek,11.mfcc500'						
301	1727	'1.usek,8.mfcc1000'						
302	1734	'1.usek,8.mfcc800'						
303	1745	'3.usek,15.mfcc500'						
304	1748	'1.usek,11.mfcc500'						
305	1750	'4.usek,4.formant'						
306	1751	'2.usek,11.mfcc1000'						
307	1758	'2.usek,11.mfcc800'						
308	1765	'4.usek,11.mfcc1000'						
309	1771	'3.usek,8.mfcc1000'						
310	1772	'4.usek,11.mfcc800'						
311	1773	'1.usek,11.mfcc1000'						
312	1778	'3.usek,8.mfcc800'						
313	1780	'1.usek,11.mfcc800'						
314	1793	'1.usek,12.mfcc500'						
315	1811	'1.usek,15.mfcc500'						