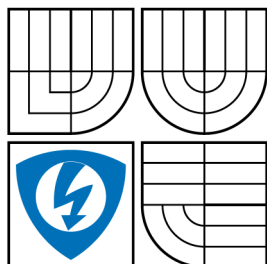


VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY
A KOMUNIKAČNÍCH TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ



FACULTY OF ELECTRICAL ENGINEERING AND
COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

ROZPOZNÁNÍ TYPU PŘENOSOVÉHO KANÁLU Z ŘEČOVÉHO SIGNÁLU

SPEECH-SIGNAL-BASED RECOGNITION OF TYPE OF TRANSMISSION CHANNEL

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

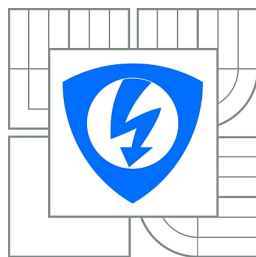
AUTOR PRÁCE
AUTHOR

TOMÁŠ KOPŘIVA

VEDOUcí PRÁCE
SUPERVISOR

Ing. HICHAM ATASSI

BRNO 2013



**VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ**

**Fakulta elektrotechniky
a komunikačních technologií**

Ústav telekomunikací

Bakalářská práce

bakalářský studijní obor
Teleinformatika

Student: Tomáš Kopřiva

ID: 136542

Ročník: 3

Akademický rok: 2012/2013

NÁZEV TÉMATU:

Rozpoznání typu přenosového kanálu z řečového signálu

POKYNY PRO VYPRACOVÁNÍ:

Prostudujte Základní vlastnosti řečového signálu. Zaměřte se na vybrané typy příznaků, například melovské keprstrální koeficienty, Lineární predikční koeficienty a spektrální charakteristiky. Vytvořte databázi řečových promluv pro různé typy přenosových kanálů, jako například PSTN a GSM. Navrhněte algoritmus v prostředí Matlab, který bude schopen automaticky rozpoznat typ přenosového kanálu z řečového signálu.

DOPORUČENÁ LITERATURA:

- [1] Psutka J.. Komunikace s počítačem mluvenou řečí. Academia, Praha 1995.
- [2] Psutka J., Müller L., Matoušek J., Radová V.. Mluvíme s počítačem česky. Academia, Praha 2006.
- [3] Sigmund M.. Analýza řečových signálů. Skripta, VUT, Brno 2000.
- [4] R. Duda, P. Hart, D. Stork, Pattern Classification, druhé vydání. Wiley, 2003.
- [5] Gao, Di, Xiong Xiao, Guangxi Zhu, Eng Siong Chng, and Haizhou Li. "Classification of speech transmission channels: Landline, GSM and VoIP networks." In Signal Processing, 2008. ICSP 2008. 9th International Conference on, pp. 671-675. IEEE, 2008.

Termín zadání: 11.2.2013

Termín odevzdání: 5.6.2013

Vedoucí práce: Ing. Hicham Atassi

Konzultanti bakalářské práce:

prof. Ing. Kamil Vrba, CSc.

Předseda oborové rady

UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

Výzkum popsáný v této bakalářské práci byl realizován v laboratořích podpořených z projektu SIX; registrační číslo CZ.1.05/2.1.00/03.0072, operační program Výzkum a vývoj pro inovace.

ABSTRAKT

Práce se zabývá rozpoznáním pěti odlišných řečových přenosových kanálů. Použité kanály jsou: GSM, dva PSTN a dva VoIP kanály. Pro účely testování a trénování je vytvořena pro přenosové kanály řečová databáze SPLAB_TranCh. Řečové nahrávky z tohoto korpusu pochází z databáze TIMIT a každá promluva prošla každým přenosovým kanálem.

Hlavním cílem práce je nalezení optimálních příznaků a klasifikátorů, které poskytují nejvyšší úspěšnost klasifikace. Je vyzkoušeno několik typů příznaků, včetně MFCC, LPCC a spektrálních charakteristik. Nejlepší suprasegmentální příznaky byly určeny algoritmem mRMR. Také bylo otestováno několik klasifikátorů. Výsledky ukazují, že rozpoznání přenosových kanálů může být realizováno s vysokou úspěšností (okolo 90 %). Práce také zkoumá vliv zkreslení, které může nastat během přenosu, na úspěšnost klasifikace. Jsou uvažovány následující typy zkreslení: saturace, prahování, ozvěna, praskání a vliv filtrů a různých barev šumů.

KLÍČOVÁ SLOVA

Rozpoznání kanálu, MFCC, LPCC, GMM, KNN, SVM, mRMR, suprasegmentální příznaky, zkreslení kanálu

ABSTRACT

This work deals with the classification of five different transmission channels by speech signal processing. The channels considered are: GSM, two PSTN channels and two VoIP channels. For the training and testing purposes, a speech database for the transmission channels called SPLAB_TranCh was constructed. The speech signals of this corpus originally come from well-known TIMIT database, where each utterance passed through each mentioned transmission channel.

The main objective of this work is to find optimal features and classification accuracy that yield best classification accuracy. Several types of features, including MFCC, LPCC and spectral characteristics were put under examination. The best suprasegmental features were identified by using mRMR algorithm. Several classifiers were tested as well. The results suggested that the classification of transmission channel can be performed with high accuracy (around 90 %). Influence of adverse effects, which can occur during transmission, is also examined. Considered types of distortions are: saturation, thresholding, echo, crackling noises and different colors of noises and filters.

KEYWORDS

Channel classification, MFCC, LPCC, GMM, KNN, SVM, mRMR, suprasegmental features, channel distortion

KOPŘIVA, Tomáš *Rozpoznání typu přenosového kanálu z řečového signálu*: bakalářská práce. BRNO: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2013. 60 s. Vedoucí práce byl Ing. Hicham Atassi

PROHLÁŠENÍ

Prohlašuji, že svou bakalářskou práci na téma „Rozpoznání typu přenosového kanálu z řečového signálu“ jsem vypracoval samostatně pod vedením vedoucího bakalářské 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 bakalářské práce dále prohlašuji, že v souvislosti s vytvořením této bakalářské práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

BRNO

.....

(podpis autora)

PODĚKOVÁNÍ

Děkuji vedoucímu bakalářské práce Ing. Hichamu Atassimu za konzultace, odborné vedení a připomínky k práci.

BRNO

.....

(podpis autora)

OBSAH

Úvod	9
1 Rozvoj problematiky	10
2 Příznaky	12
2.1 Mel-frekvenční keprální koeficienty	12
2.2 MELBS	13
2.3 Koeficienty lineární predikce	13
2.4 Keprální koeficienty LPC	15
2.5 Spektrální příznaky	16
2.6 Suprasegmentální příznaky	17
3 Redukce příznaků	20
4 Klasifikátory	21
4.1 Gaussovy smíšené modely	21
4.2 k -nejbližších sousedů	21
4.3 Algoritmus podpůrných vektorů	22
5 Tvorba databáze	23
5.1 Použitá databáze	23
5.2 Výběr přenosových kanálů	23
5.3 Nahrávání	23
5.4 Databáze SPLAB_TrnC	25
6 Poškození signálu	27
6.1 Vliv zkreslení na klasifikaci	27
6.2 PESQ	31
7 Výsledky	32
7.1 Popis metody	32
7.2 Výsledky pro jeden typ příznaku	33
7.3 Výsledky pro kombinace příznaků	35
7.4 Výsledky po zkreslení signálu	39
7.4.1 Saturace	39
7.4.2 Prahování	40
7.4.3 Ozvěna	41
7.4.4 Praskání	42

7.4.5	Šum	44
7.4.6	Vliv filtrů	44
8	Závěr	46
	Literatura	47
	Seznam symbolů, veličin a zkratk	49
	Seznam příloh	52
A	Modul Midprocessing	53
B	Obsah CD	54
C	Kompletní výsledky	55

ÚVOD

Tématem bakalářské práce je rozpoznání typu přenosového kanálu z řečového signálu. Problematika zpracování řečových signálů nabývá na významu. Díky rozvoji výpočetní techniky a výhradně digitálnímu zpracování signálů se neustále otevírají nové možnosti využití. S prudkým rozvojem telekomunikačních systémů se zvyšuje i jejich datová náročnost. Do popředí se tak dostávají otázky zpracování velkého množství dat, jejich analýza a archivace.

Otázce rozpoznání přenosových kanálů se v současnosti nevěnuje mnoho vědeckých prací. Přesto jsou k dispozici rozsáhlé možnosti využití této problematiky – analýza hovorů v call centrech, zdokonalení systémů pro rozpoznání řeči a zlepšení mnoha bezpečnostních aplikací např. pro ověřování osob.

Cílem práce je vytvoření databáze několika přenosových kanálů a nalezení optimálních příznaků a klasifikátorů pro jejich rozpoznání. Další část práce se zabývá poškozením signálu zkreslením a jeho vlivu na úspěšnost klasifikace.

V první kapitole je popsán současný stav problematiky a dosud použité přístupy k řešení. V dalších kapitolách jsou postupně popsány použité příznaky, jejich redukce a zvolené klasifikátory. Pátá kapitola se věnuje tvorbě databáze SPLAB_TrCh, na které je klasifikace realizována. Následující kapitoly popisují dosažené výsledky a vlivy zkreslení na úspěšnost klasifikace. V závěru jsou diskutovány dosažené výsledky a zvolený přístup k řešení dané problematiky.

1 ROZVOJ PROBLEMATIKY

Klasifikací přenosových kanálů se zabývá několik článků publikovaných v letech 2002–2008. Celkem 3 články o dané problematice pochází od autorů z Taiwanu, Číny a Singapuru. V následujících odstavcích jsou popsány jejich přístupy a úspěšnost řešení problému klasifikace, shrnutí je uvedeno v tab. 1.1.

Nejstarší práce z roku 2002 [5] se věnuje klasifikaci dvou přenosových kanálů: pevné veřejné telefonní sítě (PSTN) a sítě pro mobilní komunikaci GSM. Výsledky klasifikace zde slouží pro dosažení vyšší úspěšnosti při rozpoznání řeči. Použité databáze jsou v mandarínské čínštině. Pro trénovací databázi PSTN bylo použito 36 427 promluv od 1 969 mluvčích z databáze MAT (řečové nahrávky taiwanských telefonních sítí [5]). Pro GSM je použito 23 534 promluv od 492 mluvčích. Nahrávky byly pořízeny v taiwanských telefonních sítích ($f_{vz} = 8 \text{ kHz}$). Dále byla nahrána jedna testovací databáze pro pevnou telefonní síť a čtyři pro síť mobilní. Ty se liší prostředím, ve kterém byly pořízeny. Vektor příznaků se skládá z: 12 MFCC a 12 Δ MFCC koeficientů, Δe_{log} a $\Delta^2 e_{log}$. Rámce vstupní řeči jsou nejdříve zpracovány a rozděleny do následujících tříd: *počáteční*, *závěrečná* a *neřečová*. Toto rozdělení se provádí pomocí rekursivních neuronových sítí. Ty jsou dále využity pro klasifikaci kanálů, kde dosahují úspěšnosti 86 % (bez předchozího třídění rámců). Při testování jednotlivých kombinací tříd *počáteční*, *závěrečná* a *neřečová* dosahuje systém úspěšnosti v rozmezí 78,1–89,3 %. Pro použití tříd *počáteční* a *neřečová* je úspěšnost nejvyšší.

Tématem další práce [14] je klasifikaci čtyř telefonních kanálů pro zdokonalení procesu rozpoznání řeči. Byla použita databáze čínské mluvené řeči 863. Část databáze pro trénování obsahuje 50 promluv vyslovených 15 muži, včetně 201 „základních“ slabik (slouží pro rozpoznání řeči). Testovací část je stejných 50 promluv vyslovených jinými 10 muži. Původní vzorkovací kmitočet databáze 16 kHz byl nejprve převzorkován na 8 kHz. Následně byla databáze „přefiltrována“ pomocí telefonních simulátorů pro získání skupiny čtyř typů kanálů. Vektor příznaků obsahuje 12 MFCC koeficientů a logaritmus energie. Gaussovy smíšené modely slouží k modelování příznaků. Úspěšnost klasifikace lineárně filtrovaného řečového signálu je vyšší než 96,8 %, pro některé kanály i 100 %.

Nejpodrobněji se rozpoznáním kanálů zabývá práce [2]. Zkoumá problematiku klasifikace tří přenosových kanálů – pevné telefonní linky, mobilní sítě GSM a sítě typu VoIP. Mel-frekvenční kepstrální koeficienty jsou použity jako příznaky, které jsou modelovány Gaussovými smíšenými modely.

Práce využívá různé řečové databáze pro trénování a testování, celkem je použito 12 databází: 7 existujících standardních databází a 5 vytvořených. Existující databáze jsou rozděleny následovně: 5 z nich bylo vytvořeno v LDC (Linguistic Data Consortium) – 3 s přenosovým kanálem pevné telefonní linky a 2 s kanálem GSM.

Další 2 databáze jsou z I2R (Institute for Infocomm Research) – jedna s kanálem pevné linky a další je typu GSM. Z každé databáze byly náhodně vybrány 4 hodiny řečových nahrávek, z nichž celkově je 6 databází v anglickém jazyce a 1 v mandarínské čínštině.

Soubor vytvořených nahrávek obsahuje: 3 databáze od každého kanálu (tj. pevná linka, GSM a Skype), které byly nahrány zároveň různými mluvčími, v délce 4 hodin. Dále další 2 databáze VoIP kanálů – pro Skype a místního poskytovatele VoIP služeb, oba s jinými mluvčími a rozdílnými promluvami. Vzdálenost vysílání při nahrávání byla 10 km, jazyk všech nahrávek je anglický a nejsou nahrány rodilými mluvčími.

Pro trénování byly použity kombinace dvou typů databází (tj. standardní a vytvořená) pro pevnou linku, GSM a jedné pro VoIP. Ze zbylých 7 databází bylo z každé vybráno 500 promluv pro testování. Průměrná úspěšnost klasifikace kanálů je 96,63 %. Z výsledků je patrná vysoká úspěšnost při použití kombinace MFCC-GMM, dokonce i pro klasifikaci mandarínské čínštiny na anglicky trénovaném systému. Dále testování prokázalo, že úspěšnost rozpoznání přenosového kanálu roste s délkou nahrávek a složitostí modelu GMM.

Tab. 1.1: Shrnutí dosavadních prací

Reference	[5]	[14]	[2]
Rok publikace	2002	2004	2008
Kanály	PSTN, GSM	4 telefonní kanály	PSTN, GSM, VoIP
Příznaky	12 MFCC, 12 Δ MFCC, Δe_{log} , $\Delta^2 e_{log}$	12 MFCC, e_{log}	MFCC
Klasifikace	RNN	GMM	GMM
Použité databáze	MAT + nahrávky z taiwanských telefonních sítí	863	12 databází: LDC, I2R, vlastní nahrávky
Jazyk databází	mandarínská čínština	mandarínská čínština	anglický, 1 v mandarínské čínštině
Úspěšnost	78,1–89,3 %	96,8–100 %	průměrná 96,63 %

2 PŘÍZNAKY

Výběr vhodných příznaků je klíčový za účelem rozpoznání určitých parametrů z řeči – např. identita mluvčího, určení pohlaví, emocí nebo v našem případě při rozpoznání přenosových kanálů. Kapitola popisuje příznaky použité při klasifikaci. Celý postup jejich výpočtu včetně rozměrů matic v každém kroku je znázorněn na obr. 2.4. Signál je nejprve segmentován se zvolenou délkou okna a překrytím. Z matice signálu po segmentaci jsou po sloupcích vypočítány příznaky – získáme matici příznaků. Z té se po řádcích vypočítají suprasegmentální příznaky. V posledním kroku se řádky matice suprasegmentálních příznaků operací zřetězení umístí do výsledného vektoru příznaků.

2.1 Mel-frekvenční keprstrální koeficienty

Důležitou metodou krátkodobé keprstrální analýzy jsou mel-frekvenční keprstrální koeficienty (MFCC). Ty jsou navrženy tak, aby respektovaly nelineární vnímání zvuku lidským uchem. K tomu využívají banku trojúhelníkových pásmových filtrů (obr. 2.1) s lineárním rozložením frekvencí v tzv. melovské frekvenční škále

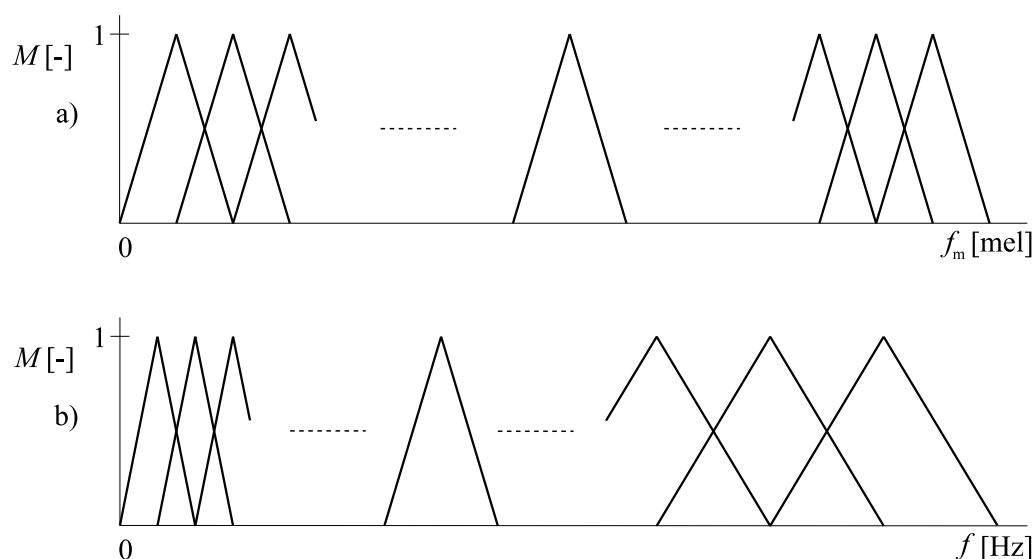
$$f_m = 2595 \log_{10} \left(1 + \frac{f}{700} \right), \quad (2.1)$$

kde $f[\text{Hz}]$ je frekvence v lineární škále a $f_m[\text{mel}]$ odpovídá frekvenci v melovské škále [7].

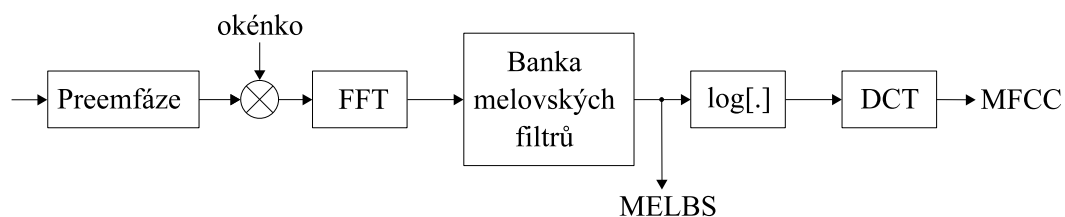
V první fázi výpočtu koeficientů MFCC je provedena preemfáze – kompenzace poklesu spektrálních složek řečového signálu na vyšších frekvencích. Dále je aplikováno Hammingovo okénko na segment o délce 10–30 ms. Počet vzorků signálu v segmentu se volí jako mocnina 2 z důvodu použití rychlé Fourierovy transformace (FFT). Algoritmem FFT je signál převeden do frekvenční oblasti a je vypočítáno jeho modulové spektrum. Hodnoty spektra jsou vynásobeny ziskem filtru a sečteny pro jednotlivé filtry. Sečtené hodnoty jsou následně logaritmovány, což vhodně omezí dynamiku signálu. Posledním krokem je výpočtu koeficientů je použití zpětné FFT. Ta se z důvodu reálného a symetrického spektra redukuje na diskrétní kosinovou transformaci [7]

$$c_m(j) = \sum_{i=1}^{M^*} \log y_m(i) \cos \left(\frac{\pi j}{M^*} (i - 0,5) \right), \quad \text{pro } j = 0, 1, \dots, M, \quad (2.2)$$

kde M^* je počet pásem melovského pásmového filtru a M je počet MFCC koeficientů. Celý proces výpočtu mel-frekvenčních keprstrálních koeficientů je zobrazen na obr. 2.2.



Obr. 2.1: Banka trojúhelníkových filtrů a) v melovské škále b) v lineární škále



Obr. 2.2: Schéma výpočtu MFCC a MELBS koeficientů

2.2 MELBS

Melovské spektrální příznaky banky (MELBS - mel frequency bank spectrum) mají stejný postup výpočtu jako MFCC. Jsou získány po vynásobení hodnot spektra ziskem filtru viz obr. 2.2.

2.3 Koeficienty lineární predikce

Metoda lineárního prediktivního kódování (LPC) odhaduje z řečového signálu parametry modelu vytváření hlasu. LPC umožňuje k -tý vzorek signálu $s(k)$ popsat lineární kombinací Q předchozích vzorků buzení $u(k)$ (následující vzorce jsou pře-

vzaty z [7]):

$$s(k) = - \sum_{i=1}^Q a_i s(k-i) + Gu(k), \quad (2.3)$$

kde G je koeficient zesílení a Q je řád modelu. Přenosová funkce hlasového traktu má tvar

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{A(z)} = \frac{G}{1 + \sum_{i=1}^Q a_i z^{-i}}. \quad (2.4)$$

Model vytváření řeči s lineárním číslicovým filterm je zobrazen na obr. 2.3. Pokud signál na daném časovém úseku bude přibližně stacionární, využijeme pro výpočet a_i a G metodu nejmenších čtverců. V rovnici 2.3 je neznámý člen $Gu(k)$, kvůli kterému vzniká chyba mezi hodnotou skutečnou $s(k)$ a předpovězenou $\hat{s}(k)$. Funkce

$$E = \sum_k e^2(k) = \sum_k [s(k) - \hat{s}(k)]^2 = \sum_k [s(k) + \sum_{i=1}^Q a_i s(k-i)]^2 \quad (2.5)$$

má minimum v bodě

$$\frac{\partial E}{\partial a_j} = 0, \quad 1 \leq j \leq Q. \quad (2.6)$$

Řešení rovnice 2.6 při použití autokorelačního přístupu lze po úpravách zapsat ve tvaru

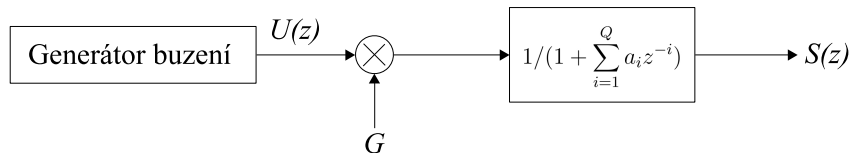
$$\sum_{i=1}^Q a_i R_n(|j-i|) = -R_n(j), \quad 1 \leq j \leq Q, \quad (2.7)$$

nebo v maticovém zápisu

$$\begin{bmatrix} R_n(0); & R_n(1); & R_n(2); & \dots; & R_n(Q-1); \\ R_n(1); & R_n(0); & R_n(1); & \dots; & R_n(Q-2); \\ \vdots & & & & \\ R_n(Q-1); & R_n(Q-2) & R_n(Q-3); & \dots; & R_n(0); \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_Q \end{bmatrix} = \begin{bmatrix} -R_n(1) \\ -R_n(2) \\ \vdots \\ -R_n(Q) \end{bmatrix}, \quad (2.8)$$

kde R_n je krátkodobá autokorelační funkce.

Matice 2.8 je symetrická v Töplitzově tvaru, lze použít Levinsonův algoritmus modifikovaný Durbinem. Řešení pro koeficienty soustavy 2.5 je vyjádřeno rekurzivně



Obr. 2.3: Model vytváření řeči

pro $i = 1, 2, \dots, Q$ jako

$$\begin{aligned}
E_n^{(0)} &= R_n(0), \\
k_i &= -[R_n(i) + \sum_{j=1}^{i-1} a_j^{(i-1)} R_n(i-j)] / E_n^{(i-1)}, \\
a_i^{(i)} &= k_i, \\
a_j^{(i)} &= a_j^{(i-1)} + k_i a_{i-j}^{(i-1)}, \quad 1 \leq j \leq i-1, \\
E_n^{(i)} &= (1 - k_i^2) E_n^{(i-1)},
\end{aligned} \tag{2.9}$$

kde $a_j^{(i)}$ je j -tý parametr prediktoru řádu i , k_i jsou koeficienty odrazu.

Tvar spektra signálu vypočítaného pomocí koeficientů a_i má charakter vyhlazené spektrální obálky skutečného spektra původního diskretizovaného signálu $s(k)$. Po substituci $z = e^{j\omega}$ pro frekvenční přenos modelu vytváření akustického signálu platí

$$H(j\omega) = \frac{G}{1 + \sum_{i=1}^Q a_i e^{-j\omega i}}. \tag{2.10}$$

2.4 Kepstrální koeficienty LPC

Kepstrální koeficienty popisují přenosovou funkci hlasového traktu. Pro jejich výpočet nejdříve podle pravidel homomorfní analýzy určíme logaritmus přenosové funkce 2.4. Pokud polynom $A(z)$ proměnné z^{-1} je Q -tého řádu a všechny kořeny tohoto polynomu leží uvnitř jednotkové kružnice a $A(\infty) = 1$, lze provést rozvoj Taylorovou řadou pro $\log(G/A(z))$

$$\log(G/A(z)) = c(0) + c(1)z^{-1} + \dots = \sum_{k=0}^{\infty} c(k)z^{-k}, \tag{2.11}$$

kde $c(k)$ představují kepstrální koeficienty LPC. Po derivaci rovnice 2.11 a následných úpravách, při $a_0 = 1$ a porovnání členů u stejných mocnin z , jsou kepstrální koeficienty LPC definované [7]

$$\begin{aligned}
c(1) &= -a_1, \\
c(k) &= -a_k - \sum_{i=1}^{k-1} \left(\frac{i}{k}\right) c(i) a_{k-i}, \quad \text{pro } 2 \leq k \leq Q, \\
c(k) &= -\sum_{i=1}^Q \left(\frac{k-i}{k}\right) c(k-i) a_i, \quad \text{pro } k = Q+1, Q+2, \dots
\end{aligned} \tag{2.12}$$

Důležitou vlastností kepstrálních koeficientů LPC je, že jsou obecně de Korelované. Z tohoto důvodu jsou vhodné jako příznaky v rozpoznávacích systémech.

2.5 Spektrální příznaky

Spektrální centroid

Je to hodnota, která se využívá ve zpracování signálů a popisuje spektrum. Charakterizuje místo, kde leží největší část spektra – centrum neboli těžiště spektra. Spektrální centroid je definovaný jako frekvence obsažené v signálu vážené amplitudou vydělené součtem amplitud

$$\mu_c = \frac{\sum_{n=0}^{N-1} f(n)M(n)}{\sum_{n=0}^{N-1} M(n)}, \quad (2.13)$$

kde $M(n)$ je modul a $f(n)$ frekvenční složka signálu.

Spektrální rozptyl

Rozptyl spektra kolem jeho střední hodnoty:

$$\sigma_r = \frac{\sum_{n=0}^{N-1} (f(n) - \mu_c)^2 M(n)}{\sum_{n=0}^{N-1} M(n)}. \quad (2.14)$$

Spektrální šikmost

Popisuje nesymetričnost rozložení kolem střední hodnoty. Jedná se o třetí normalizovaný moment:

$$\gamma_s = \frac{\sum_{n=0}^{N-1} (f(n) - \mu_c)^3 M(n)}{\sum_{n=0}^{N-1} M(n)}. \quad (2.15)$$

Pokud $\gamma_s = 0$, hodnoty spektra jsou přibližně rovnoměrně rozděleny. Při $\gamma_s < 0$ je více energie soustředěno na pravé straně, pro $\gamma_s > 0$ je energie koncentrována na levé straně.

Spektrální špičatost

Charakterizuje tvar rozložení kolem střední hodnoty:

$$\gamma_k = \frac{\sum_{n=0}^{N-1} (f(n) - \mu_c)^4 M(n)}{\sum_{n=0}^{N-1} M(n)}. \quad (2.16)$$

Tvar vypadá shodně s Gaussovým v případě $\gamma_k = 0$. Pro $\gamma_k < 0$ je rozložení plošší a pro $\gamma_k > 0$ je špičatější než Gaussovo rozdělení.

Spektrální sklon

Vyjadřuje, jak rychle spektrum signálu klesá směrem k vysokým frekvencím. Počítá se pomocí lineární regrese spektra signálu metodou nejmenších čtverců

$$s_s = \frac{1}{\sum_{n=0}^{N-1} M(n)} \frac{\sum_{n=0}^{N-1} f(n)M(n) - \sum_{n=0}^{N-1} f(n) \sum_{n=0}^{N-1} M(n)}{\sum_{n=0}^{N-1} f^2(n) - \left(\sum_{n=0}^{N-1} f(n)\right)^2}. \quad (2.17)$$

2.6 Suprasegmentální příznaky

Suprasegmentální příznaky jsou vypočítány po řádcích z matice příznaků (obr. 2.4). Z tohoto důvodu zachycují změny řečového signálu v čase a jejich výpočtem se odstraňuje redundantní informace obsažená v signálu.

Průměr

Průměr je definován

$$y_m = \frac{1}{N} \sum_{i=1}^N x_i, \quad (2.18)$$

kde N počet sloupců matice (neboli počet segmentů promluvy) a x_i je prvek řádku matice.

Medián, percentil

Medián rozděluje řadu seřazených výsledků na stejně početné poloviny. Pro neklesající posloupnost liché délky se jedná o hodnotu uprostřed posloupnosti

$$y_{med} = \text{Med}(x). \quad (2.19)$$

Obdobně je definován percentil, který se vztahuje k setinám pořadí. Např. u 10. percentilu (decilu) se hodnota nachází v jedné desetíně pořadí. Pod 10. percentilem se nachází 10 % všech hodnot a nad ním je 90 % všech ostatních hodnot.

Směrodatná odchylka

Je vypočítána dle vztahu

$$y_s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - y_m)^2}. \quad (2.20)$$

Relativní směrodatná odchylka

Je vážena průměrnou hodnotou

$$y_{sr} = \frac{y_s}{y_m}. \quad (2.21)$$

Maximum, relativní maximum

Maximální hodnota řady a maximální hodnota řady vážená průměrem:

$$y_{max} = \text{Max}(x_i), \quad y_{maxr} = \frac{\text{Max}(x_i)}{y_m}. \quad (2.22)$$

Minimum, relativní minimum

Minimální hodnota řady a minimální hodnota řady vážená průměrem:

$$y_{min} = \text{Min}(x_i), \quad y_{minr} = \frac{\text{Min}(x_i)}{y_m}. \quad (2.23)$$

Rozsah, relativní rozsah

Rozdíl maxima a minima a rozdíl maxima a minima vážený průměrem:

$$y_r = y_{\max} - y_{\min}, \quad y_{rr} = \frac{y_{\max} - y_{\min}}{y_m}. \quad (2.24)$$

Šikmost

Jedná se o třetí standardizovaný moment vydělený třetí mocninou směrodatné odchylky

$$y_{\gamma s} = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - y_m)^3}{\left(\frac{1}{N} \sum_{i=1}^N (x_i - y_m)^2\right)^{3/2}}. \quad (2.25)$$

Špičatost

Čtvrtý standardizovaný moment vydělený čtvrtou mocninou směrodatné odchylky

$$y_{\gamma k} = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - y_m)^4}{\left(\frac{1}{N} \sum_{i=1}^N (x_i - y_m)^2\right)^2} - 3. \quad (2.26)$$

Regresní koeficienty, regresní chyba

Metody, které slouží odhadu hodnoty náhodné veličiny. Regresní koeficienty jsou vypočítány ze vztahu

$$y_{rc} = \frac{\sum_{i=1}^N (x_i) i - \frac{1}{N} \sum_{i=1}^N i \sum_{i=1}^N x_i}{\sum_{i=1}^N i^2 - \frac{1}{N} \left(\sum_{i=1}^N i\right)^2} \quad (2.27)$$

a regresní chyba

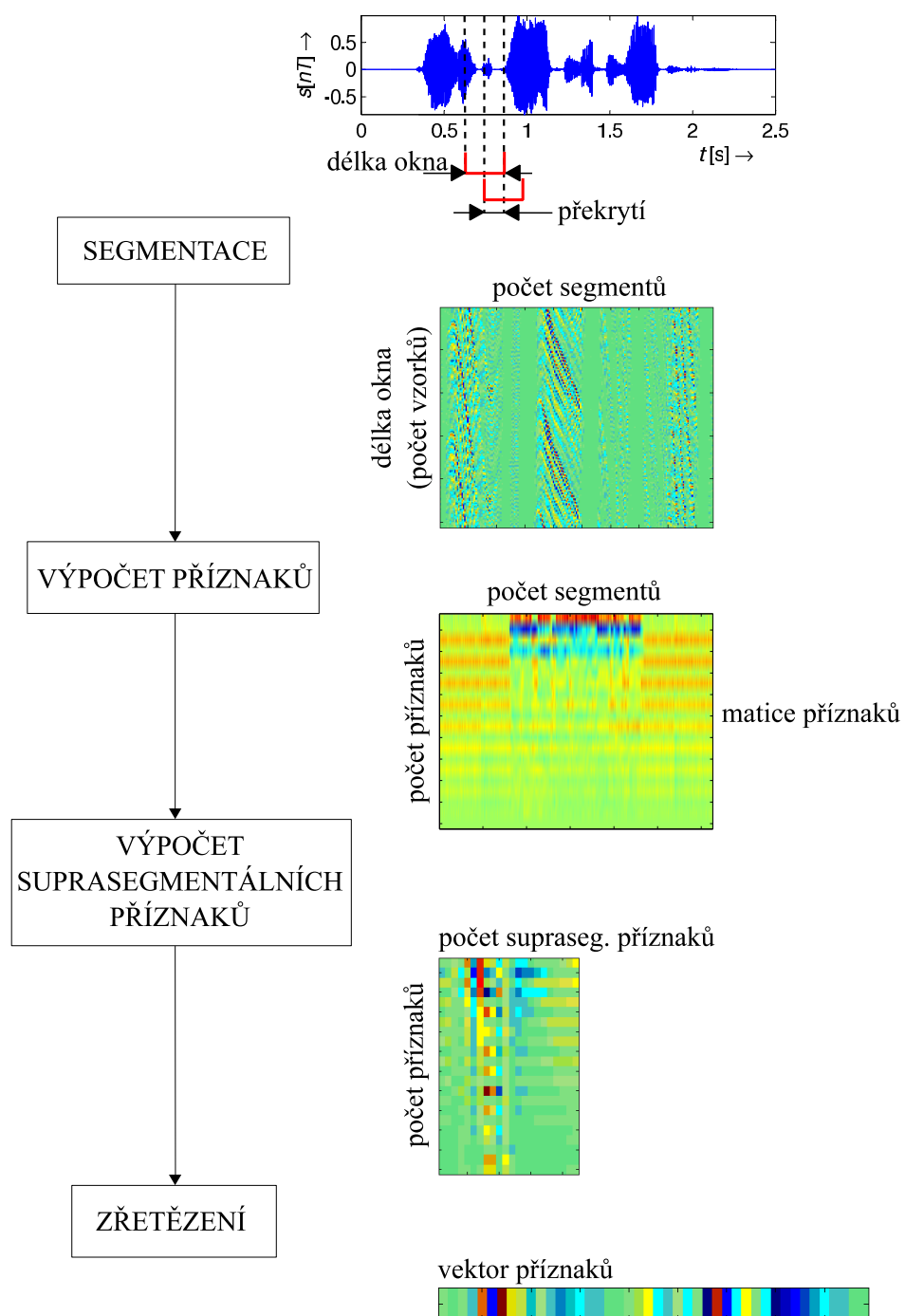
$$y_{re} = \frac{\sum_{i=1}^N (x_i - (y_m - y_{rc} \bar{i}) - y_{rc} i)^2}{N}. \quad (2.28)$$

5-tý moment, 6-tý moment

Momenty jsou definované obdobně jako 2.25, kde obecně k -tý moment lze zapsat ve tvaru

$$y_{\gamma} = \frac{\mu_k}{\sigma^k}, \quad (2.29)$$

kde μ_k je k -tý centrální moment a σ je směrodatná odchylka.



Obr. 2.4: Výpočet příznaků

3 REDUKCE PŘÍZNAKŮ

Z vypočítaných suprasegmentálních příznaků se vybere pouze jejich část. Algoritmus minimální redundance maximální relevance (mRMR - minimum Redundancy Maximum Relevance) je efektivním nástrojem pro redukci příznaků. Z názvu vyplývá, že funkcí algoritmu je zmenšit počet nadbytečných příznaků a vybrat maximální počet nejvhodnějších příznaků patřících k dané třídě.

Je využita informace o podobnosti příznaků. Podmnožina S označuje redukované příznaky a množina Ω všechny příznaky. Podmínka minimální redundance $\min W$ [4]:

$$W = \frac{1}{|S|^2} \sum_{i,j \in S} I(f_i, f_j), \quad (3.1)$$

kde $I(f_i, f_j)$ je vzájemná informace mezi příznaky f_i a f_j , $|S|$ je počet příznaků v množině S . Podmínka maximální relevance $\max V$ [4]:

$$V = \frac{1}{|S|} \sum_{i \in S} I(c, f_i), \quad (3.2)$$

kde c je cílová třída a $I(c, f_i)$ je vzájemná informace mezi třídou a příznakem. Sadu příznaků vybranou mRMR získáme kombinací podmínek 3.1 a 3.2.

Mezi výhody mRMR patří výpočetní nenáročnost, nezávislost na klasifikátoru a nekorelovanost příznaků.

4 KLASIFIKÁTORY

Klasifikátor je algoritmus, který na základě vstupních příznaků rozhodne o přiřazení do třídy – v našem případě se jedná o typ přenosového kanálu. V kapitole jsou popsány použité klasifikátory.

4.1 Gaussovy smíšené modely

Gaussovy smíšené modely (GMM) pracují na principu statistického modelování příznaků v příznakovém prostoru. Vícerozměrnou Gaussovu funkci rozdělení pravděpodobnosti můžeme zapsat

$$f(x) = \frac{1}{\sqrt{(2\pi)^d \det(\Sigma_i)}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right), \quad (4.1)$$

kde d je rozměr Gaussovy funkce, μ je vektor středních hodnot a Σ je kovariační matice [9]. Pokud je kovariační matice diagonální, osy Gaussových funkcí jsou rovnoběžné s osami příznakového prostoru. Prvky mimo hlavní diagonálu určují otočení funkce vůči osám prostoru.

GMM vznikají lineární kombinací Gaussových funkcí

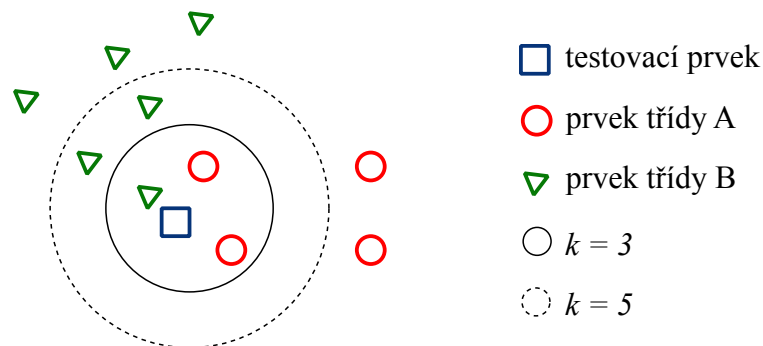
$$f(x) = \sum_{i=1}^{M_G} \alpha_i N_i(x), \quad (4.2)$$

kde M_G je počet Gaussových funkcí rozložení pravděpodobnosti, α_i váhovací parametr a $N_i(x)$ je funkce rozložení pravděpodobnosti [9].

K výpočtu GMM modelu je nutné určit kovariační matici Σ , váhovací parametr α_i a vektor středních hodnot μ . Parametry pro definici a trénování GMM modelů se určují pomocí iteračního algoritmu EM (Expectation-Maximization). Ten pro svou inicializaci vyžaduje k -means algoritmus pro vypočítání počátečních hustot rozložení pravděpodobnosti. Pokud je použita jedna Gaussova funkce a typ kovariační matice je diagonální, algoritmus EM se nemusí použít.

4.2 k -nejbližších sousedů

Metoda k -nejbližších sousedů (KNN z anglického k -Nearest Neighbor) patří mezi nejjednodušší způsoby klasifikace. Třída prvku je klasifikována na základě vzdálenosti k nejbližším natrénovaným vzorům. Testovací prvek je zařazen do třídy, ve které leží nejvíce nejbližších k prvků. Vliv má také velikost k . Tento proces je znázorněn na obr. 4.1. Pro $k = 3$ je prvek zařazen do třídy A . Za předpokladu $k = 5$ převažují prvky třídy B , kam by nyní náležel testovací prvek. Je vhodné volit lichý



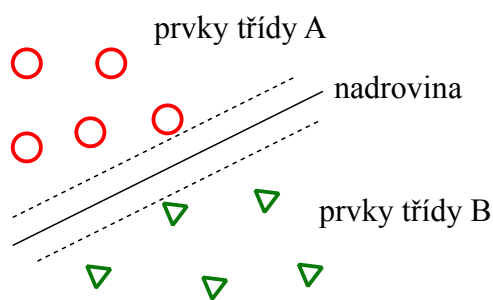
Obr. 4.1: Princip klasifikace KNN

počet k z důvodu možnosti výběru stejného počtu prvků pro různé třídy. Pro $k = 1$ se klasifikace provede přiřazením k nejbližšímu prvku.

4.3 Algoritmus podpůrných vektorů

Princip algoritmu podpůrných vektorů (SVM - Support Vector Machine) spočívá v rozdělení prostoru příznaků nadrovinou, která odděluje vstupní trénovací data – viz obr. 4.2. Nadrovina je zvolena tak, aby prostor mezi prvky tříd byl co největší. Rozdělující nadrovina rozděluje data do dvou tříd.

SVM patří do skupiny tzv. jádrových algoritmů. Způsobem na principu jádrové transformace lze převést příznaky, které nelze oddělit lineárně, do prostoru vyšší dimenze. V tomto prostoru jsou příznaky již lineárně separovatelné. Toho se využívá při klasifikaci více tříd.



Obr. 4.2: Algoritmus SVM

5 TVORBA DATABÁZE

5.1 Použitá databáze

Pro práci byl použit korpus čtené řeči TIMIT [3]. Ten byl vytvořen v roce 1993 pro studium a vývoj systémů umožňujících rozpoznání řeči. Databáze v našem případě obsahuje 526 mluvčích osmi hlavních dialektů americké angličtiny. Od každého mluvčího je k dispozici 10 promluv, z nichž 2 jsou vždy stejné a ostatní se opakují s velmi malou četností. Podle pohlaví je databáze roztržena na 170 žen a 356 mužů. Každá promluva je v jednokanálovém zvukovém formátu wav – kódování 16 bitů PCM se vzorkovacím kmitočtem 16 kHz. Průměrná délka promluvy je 3,1 s se směrodatnou odchylkou 0,9 s.

Z databáze byla náhodně vybrána jedna promluva od každého mluvčího v celkovém počtu 300 promluv. Ty jsou rozděleny na 150 promluv od žen a 150 promluv od mužů. Celková délka vybraných promluv je 18 minut a 9 sekund.

5.2 Výběr přenosových kanálů

Pro studium problematiky je vybráno 5 přenosových kanálů. Jedná se o kanál místní telefonní sítě uvnitř budovy (PSTN1), kanál pevné telefonní sítě meziměstského hovoru (PSTN2) a mobilní sítě (GSM). Další kanály zastupují VoIP služby. Skype představuje jeden z nejpoužívanějších bezplatných komunikátorů. Alternativu k tomuto programu zastupuje program Rynga.

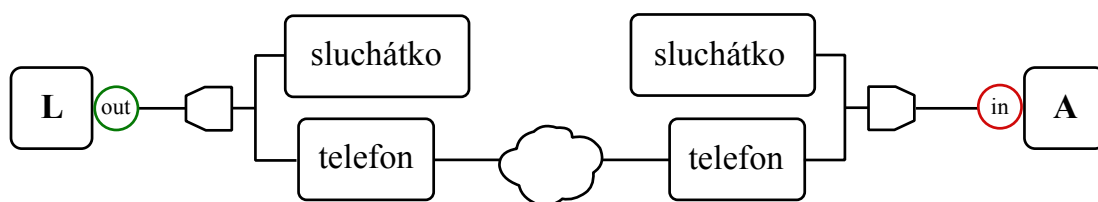
Ve všech případech sloužila pevná linka pro příjem hovorů. Vyjmenované kanály jsou tedy kombinací daného kanálu a kanálu pevné telefonní linky (tj. GSM-PSTN, PSTN1-PSTN, PSTN2-PSTN, Skype-PSTN, Rynga-PSTN). V textu jsou tyto kanály označeny kratším zápisem ve tvaru GSM, PSTN1, PSTN2, Skype nebo Rynga.

5.3 Nahrávání

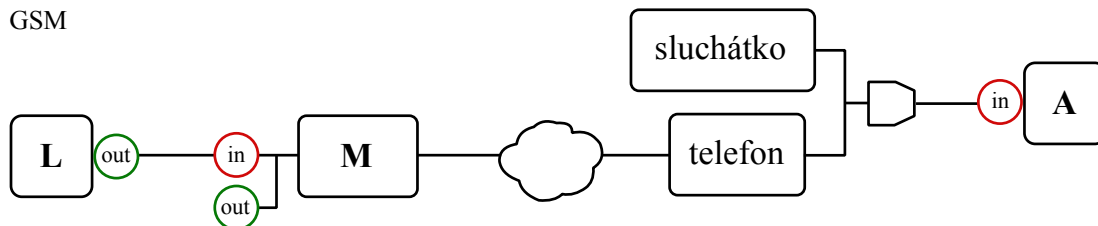
Prvotním úkolem bylo nahrát vybrané promluvy přes každý přenosový kanál. Z důvodu následného zpracování v Matlabu, byly promluvy odděleny signalizačním tónem. Ten se skládal ze součtu dvou harmonických signálů s kmitočtem $f_1 = 2$ kHz a $f_2 = 2,5$ kHz v délce trvání jedné sekundy. Tento signál byl ohraničen z obou stran sekundou ticha.

Schéma zapojení každého kanálu při nahrávání popisuje obr. 5.1. Písmeno L označuje zdrojové zařízení Lenovo ThinkPad SL510. To sloužilo pro přehrávání promluv z multimediálního přehrávače VLC. Pro záznam byl zvolen notebook Asus

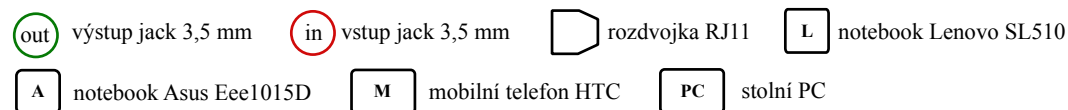
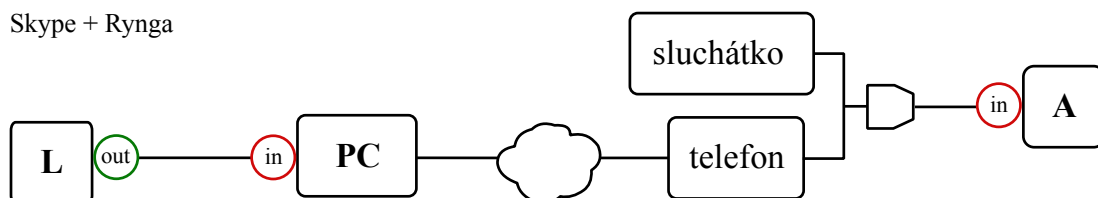
PSTN1 + PSTN2



GSM



Skype + Rynga



Obr. 5.1: Schéma zapojení při nahrávání

Eee 1015PD (označený jako *A*). Důvodem byla jeho nízká vstupní impedance zvukové karty. Pro nahrávání byl ve všech případech připojen ke klasickému telefonu přes rozdvojku RJ11. Ta sloužila k vyvedení signálu z reproduktoru sluchátka na vstup jack 3,5 mm. Všechny promluvy byly nahrány na stejném telefonním přístroji v budově Ústavu telekomunikací. Mezi telefonem (sluchátkem) a rozdvojkou je použitý telefonní kabel s konektory RJ10–RJ11, sluchátko při nahrávání mělo odpojený mikrofon. Jako nahrávací program byl použit bezplatný All2WAV Recorder. Ten umožňoval nastavení všech potřebných parametrů: formát wav, vzorkovací kmitočet 16 kHz, nahrávání v mono kanálu s rozlišovací schopností 16 bitů.

První část obr. 5.1 zobrazuje stejné schéma zapojení pro oba telefonní kanály. Kanál PSTN1 představuje telefonní hovor uvnitř budovy – v našem případě UTKO. Pro porovnání s tímto kanálem je vytvořen další telefonní kanál PSTN2. Jedná se o hovor Holešov-Brno. Vzdálenost měst je přibližně 85 km. Signál je do telefonu

přiveden přes telefonní rozdvojku RJ11 obdobným způsobem, jako při záznamu. Propojovací kabel lze vytvořit přivedením dvou vodičů (kanál + zem) z konektoru jack do telefonního konektoru RJ11. Jedná se o vnitřní nebo vnější páry konektoru (závisí na konstrukci telefonu). Pro kvalitu nahrávek bylo důležité odpojení vstupního zařízení *L* od elektrické sítě 230 V. Připojení k elektrické síti se projevovalo slyšitelným rušením řečového signálu. Zařízení pro záznam *A* také nebylo napájeno ze sítě. Tabulka 5.1 popisuje nastavené hlasitosti zvukových karet (všechna nastavení byla provedena v ovladači zvukové karty Realtek) a programu VLC pro všechny kanály a zařízení. Toto nastavení je důležité pro dosažení vhodné hlasitosti pro záznam a zabránění saturaci signálu.

Pro vytvoření kanálu mobilní sítě GSM byl hovor realizován z telefonu značky HTC. Signál z výstupu počítače byl do něho veden přes rozdvojku jack 3,5 mm.

Poslední část obr. 5.1 znázorňuje zapojení pro oba VoIP komunikátory – Skype a Rynga. Ty jsou nainstalovány na počítači *PC*, ze kterého je veden hovor na pevnou linku. Z důvodu stejných podmínek v rámci celého hovoru byla v programech vypnuta funkce automatického nastavení hlasitosti mikrofونů a reproduktorů.

Tab. 5.1: Nastavení hlasitostí při nahrávání

Zařízení	L		A	PC	
Kanál	VLC	zvuková karta	zesílení mikrofونu	hlasitost nahrávání	zesílení mikrofونu
PSTN1	77 %	54 %	+10 dB/100 %	–	–
PSTN2	79 %	80 %	+10 dB/100 %	–	–
GSM	59 %	54 %	+10 dB/100 %	–	–
Skype	79 %	80 %	+10 dB/100 %	55 %	0 dB
Rynga	79 %	80 %	+10 dB/100 %	45 %	0 dB

5.4 Databáze SPLAB_TrانCh

Z každého kanálu byla získána nahrávka o délce 33 minut a 12 sekund, která obsahovala všechny vybrané promluvy oddělené pípnutím. V Matlabu byla nahrávka rozdělena do samostatných promluv. Ty byly následně pojmenovány následujícím způsobem: *číslo nahrávky_identifíkátor mluvčího_označení promluvy_pohlaví mluvčího*. Např. 010_AJP0_SI2334_M.

Databáze *SPLAB_TrانCh* obsahuje 5 přenosových kanálů (GSM, PSTN1, PSTN2, Rynga, Skype) s celkovým počtem 1500 promluv (tj. 300 promluv pro 1 kanál) – viz tab. 5.2. Je rozdělena na část pro testování *SPLAB_TrانCh_testing*,

pro kterou bylo náhodně vybráno 50 promluv z každého kanálu (celkem 250). Zbylá část, pojmenovaná *SPLAB_Tranch_training*, slouží k trénování (250 promluv pro každý kanál, dohromady 1250).

Tab. 5.2: Použité databáze

Databáze	TIMIT	SPLAB_Tranch
Počet promluv	5260	1500 (tj. 300 pro 1 kanál)
Celkový počet mluvčích	526	300
Počet mluvčích žen	170	150
Počet mluvčích mužů	356	150

6 POŠKOZENÍ SIGNÁLU

6.1 Vliv zkreslení na klasifikaci

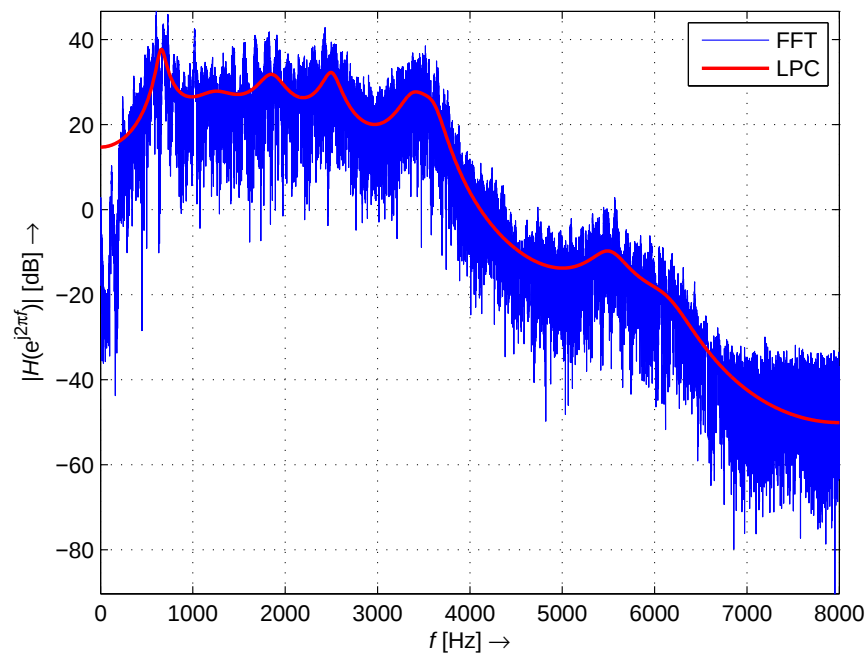
Důležitým parametrem systémů, které jsou určeny pro úlohy rozpoznání, je robustnost. Ta představuje schopnost systému spolehlivě pracovat i při změněných podmínkách. V práci jsou simulovány rozdílné podmínky pro trénování a testování tím, že testovací část databáze je úmyslně poškozena různými vlivy prostředí a je otestována na původním natrénovaném systému bez poškození. Hlavní zdroje zkreslení signálu lze rozdělit následovně: na vliv řečníka, vliv prostředí a vliv přenosového kanálu [7]. Při rozpoznání přenosových kanálů je systém na řečnickovi nezávislý. Vliv řečníka lze z tohoto důvodu zanedbat. Tento parametr je důležitý v systémech pro rozpoznání řeči.

Další dva zdroje zkreslení signálu budou mít vliv na charakteristiku daného kanálu. Do kategorie vliv prostředí patří šumy a hluky, které nejsou korelované a mají buď stacionární nebo nestacionární průběh. Stacionární průběhy mají v čase konstantní spektrální hustotu výkonu (např. zvuk větráku, deště) a lze je simulovat různými barvami šumů. Spektrální charakteristiky jsou u nestacionárního šumu proměnlivé v čase (např. výkřik, bouchnutí dveří). Dalším možným vlivem prostředí je ozvěna (echo).

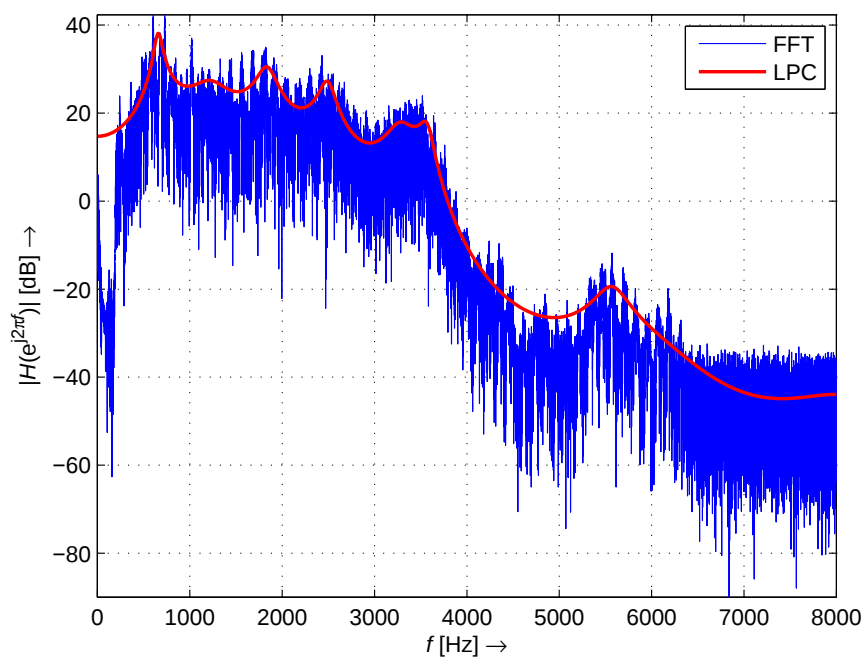
Zkreslení přenosového kanálu může mít mnoho podob. Je to poškození signálu, které je způsobeno mikrofonom a vlastnostmi kanálu – např. zkreslení a odlišnost mikrofونů, praskání, saturace, zkreslení a šum kanálu, šířka pásma nebo ozvěna v telefonních sítích. Práce se zabývá otázkou schopnosti klasifikace přenosových kanálů, pokud jsou promluvy cíleně poškozeny některými výše uvedenými vadami. Zkreslení ovlivní průběh signálu v čase, jeho modulové spektrum a tvar obálky. Nezkreslené průběhy jsou zobrazeny pro každý kanál na obr. 6.1 až 6.5. Srovnání všech kanálů je na obr. 6.6.

Testované vlivy zkreslení:

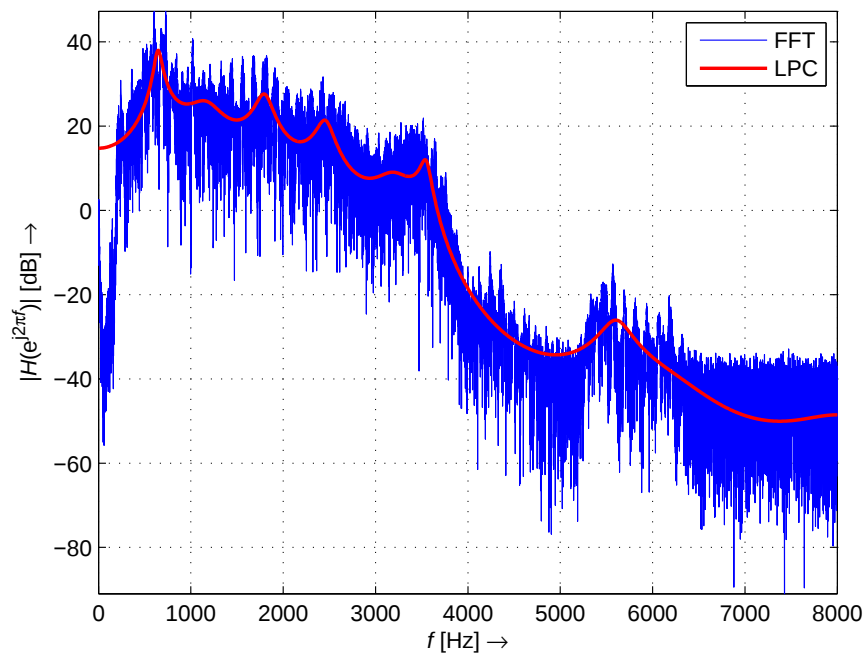
- bílý šum, růžový šum, Brownův šum
- ovlivnění signálu napětím sítě s frekvencí 50 Hz
- saturace, prahování (thresholding)
- ozvěna
- praskání
- filtry (horní a dolní propust)



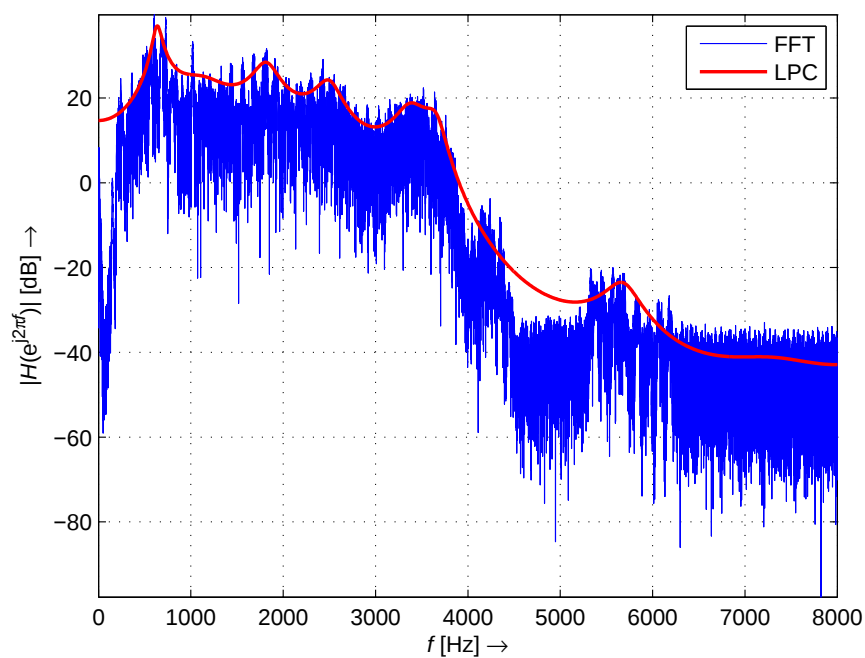
Obr. 6.1: Spektrum promluvy pro kanál GSM vypočítané pomocí LPC koeficientů a Fourierovy transformace



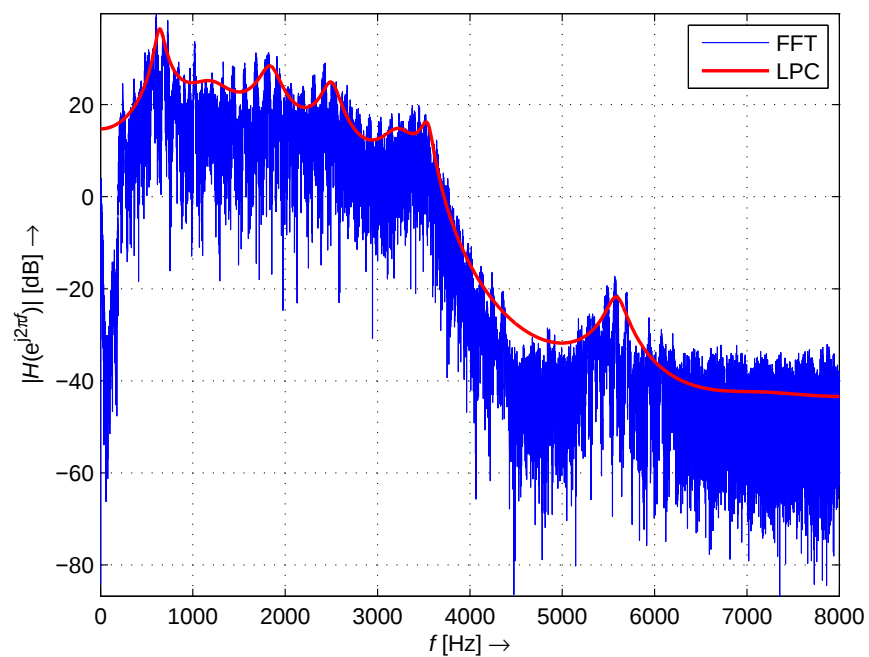
Obr. 6.2: Spektrum promluvy pro kanál PSTN1 vypočítané pomocí LPC koeficientů a Fourierovy transformace



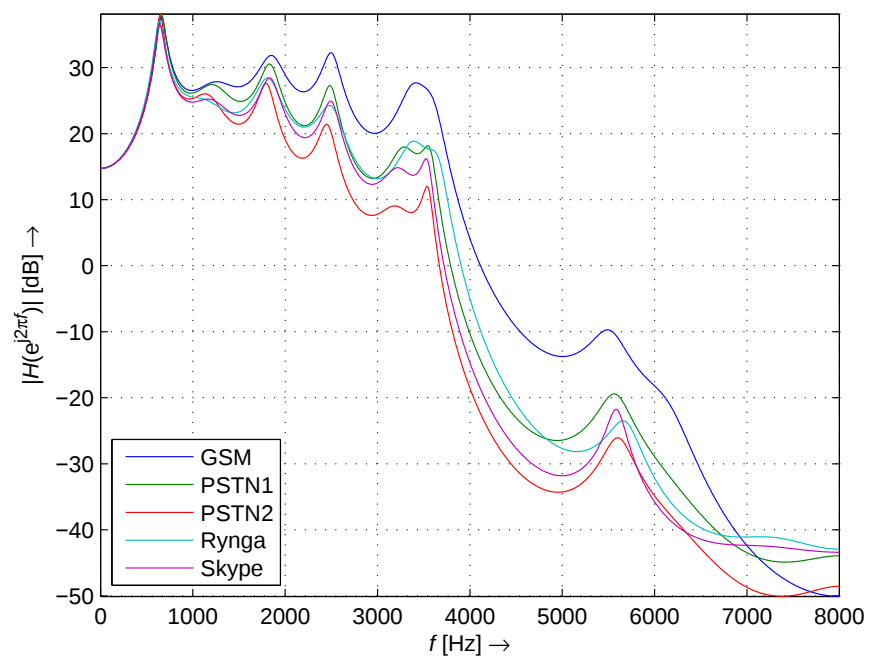
Obr. 6.3: Spektrum promluvy pro kanál PSTN2 vypočítané pomocí LPC koeficientů a Fourierovy transformace



Obr. 6.4: Spektrum promluvy pro kanál Rynka vypočítané pomocí LPC koeficientů a Fourierovy transformace



Obr. 6.5: Spektrum promluvy pro kanál Skype vypočítané pomocí LPC koeficientů a Fourierovy transformace



Obr. 6.6: Obálky spekter pro všechny kanály

6.2 PESQ

Pro vyhodnocení kvality signálu po degradaci je využita metoda PESQ (Perceptual Evaluation of Speech Quality). Patří do skupiny objektivních metod. Ty byly vyvinuty z důvodu zřejmých nevýhod subjektivních metod (velká časová i finanční náročnost, nutná skupina posluchačů, která hodnotí kvalitu). PESQ využívá dvoukanálové měření tj. kvalita je posouzena porovnáním referenčního (čistého) a testovaného (degradovaného) signálu. Metoda je velmi rozšířená (standard ITU-T P.862) a používá se při hodnocení kvality řečových kodeků a také v paketových sítích. Výsledky PESQ jsou korelované se subjektivními poslechovými testy na stupnici MOS (Mean Opinion Scores): 1 (nejhorší kvalita) až 4,5 (nejlepší kvalita) [8].

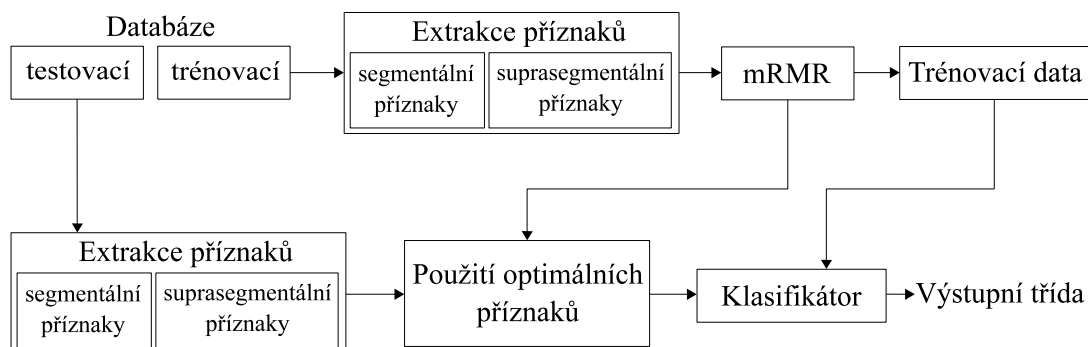
7 VÝSLEDKY

7.1 Popis metody

Pro veškerou práci s příznaky a následnou klasifikaci byl použit matlabovský program Hila. Jeho autorem je Ing. Hicham Atassi. Výsledky kapitoly 7.4 byly získány s využitím programového modulu Midprocessing, který byl vytvořen v rámci bakalářské práce. Jeho popis je uveden v příloze A.

Celý proces určení přenosových kanálů znázorňuje obr. 7.1. Z trénovací části databáze *SPLAB_TrnCh_training* jsou promluvy rozděleny na segmenty o délce 16 ms (256 vzorků), vážené Hammingovým okénkem a vzájemně se překrývají o 128 vzorků. Z těchto segmentů se vypočítá nastavený typ příznaků a umístí se do matice o rozměru *počet příznaků (koeficientů) × počet segmentů*. Pro příznaky MFCC a MELBS je nastaven výpočet 20 koeficientů, pro LPCC 10 koeficientů. Použité spektrální příznaky jsou popsány v kapitole 2.5. Dalším krokem je výpočet supra-segmentálních příznaků a jejich redukce algoritmem mRMR s výběrem 50 příznaků. Práce porovnává tři odlišné typy klasifikátorů – nejjednodušší je metoda KNN pro $k = 7$, následně je použito GMM s jednou Gaussovou funkcí a jako poslední algoritmus SVM. Pro trénovací část databáze je provedena validace typu vynechání jednoho vzorku (leave one-out). Při ní je použito jediné promluvy pro testování a ostatních pro trénování systému. Tento postup se následně opakuje – zvolí se další promluva a dochází k novému trénování dokud není každá promluva použita pro validaci. Tímto způsobem je vypočítána tzv. zmatková matice (confusion matrix), která zobrazuje úspěšnosti rozpoznání jednotlivých kanálů.

Obdobným způsobem se celý systém testuje. Z testovací databáze, která se nesmí shodovat s trénovací, klasifikátor z již natrénovaných výsledků určí výslednou třídu. Výsledkem je také zmatková matice, nyní již však pro testování.



Obr. 7.1: Proces klasifikace

7.2 Výsledky pro jeden typ příznaku

Při klasifikaci byly otestovány všechny kombinace zmíněných příznaků a klasifikátorů s výše uvedeným nastavením. V tabulce 7.1 jsou uvedeny průměrné úspěšnosti rozpoznání přenosových kanálů – použité příznaky, typ klasifikátoru a procentuální průměrná úspěšnost při validaci a testování.

Výsledky jasně ukazují, že nejvhodnějšími příznaky z námi použitých jsou MFCC. Úspěšnost jejich použití dokládají nejvyšší průměrné úspěšnosti jak pro validaci, tak i pro testování. Podobné úspěšnosti dosahuje pouze kombinace MELBS a SVM. Naopak použití příznaků spektrálních (označené jako SF – spectral features) a LPCC se nejvíce jeví jako vhodné – úspěšnost je v rozmezí 64,40 % – 77,20 %.

Další jev, který lze z výsledků vypožorovat, je závislost úspěšnosti klasifikace na použitém klasifikátoru. Ve všech případech pro každý typ příznaku dosahuje nejvyšší úspěšnosti klasifikátor SVM, následovaný KNN a GMM. Liší se ale jejich rozdíl úspěšností pro odlišné příznaky – např. u MFCC je úspěšnost všech klasifikátorů kolem 85 %, u příznaků MELBS je již rozdíl mezi nejlepší a nejhorší úspěšností testování 22,4 %.

Tab. 7.1: Výsledky klasifikace – průměrná úspěšnost

Příznaky	Klasifikátor	Validace (%)	Testování (%)	Číslo testu
MFCC	GMM	82,40	83,60	1
	KNN	83,52	85,20	2
	SVM	88,16	86,80	3
MELBS	GMM	62,48	61,20	4
	KNN	76,32	75,60	5
	SVM	83,52	83,60	6
LPCC	GMM	68,24	66,80	7
	KNN	71,44	67,60	8
	SVM	77,52	77,20	9
SF	GMM	64,80	64,40	10
	KNN	68,00	72,00	11
	SVM	73,20	77,20	12

Zmatkové matice poskytují podrobnější pohled na výsledky. Na hlavní diagonále se nachází úspěšnost rozpoznání každého přenosového kanálu. Prvky mimo hlavní diagonálu poskytují procentuální informace o promluvách, které byly klasifikovány nesprávně.

Příznaky MFCC

Ve výsledcích je zřejmý problém rozpoznání podobných kanálů PSTN1 a PSTN2. Při rozpoznání PSTN1 bylo velmi časté špatné zařazení promluv – u klasifikátorů GMM tab. 7.2 a SVM tab. 7.4 do tříd GSM a PSTN2. Na rozdíl od KNN (tab. 7.3), který promluvy PSTN1 chybně zařazoval navíc i do kanálu Rynga. Naopak předpokládaný problém klasifikace dvou VoIP kanálů se příliš nepotvrdil. Chybná klasifikace obou VoIP kanálů se pohybuje v řádu jednotek procent, nejlepších výsledků v tomto případě dosáhl klasifikátor KNN. Dalším pozorovatelným jevem je chybovost rozpoznání GSM a PSTN1 – přibližně kolem 10 %.

Ostatní příznaky

Ostatní příznaky dosahují horších průměrných výsledků. Přesto kombinace příznaků MELBS a klasifikátoru SVM poskytuje dobrou úspěšnost klasifikace a může sloužit v dalším výzkumu. Kompletní výsledky jsou uvedeny v příloze C.

Tab. 7.2: Zmatková matice pro kombinaci MFCC-GMM

Testování	GSM	PSTN1	PSTN2	Rynga	Skype
GSM	88	12	0	0	0
PSTN1	10	74	14	2	0
PSTN2	0	8	82	0	10
Rynga	0	6	4	84	6
Skype	0	2	8	0	90

Průměr: **83,60 %**

Tab. 7.3: Zmatková matice pro kombinaci MFCC-KNN

Testování	GSM	PSTN1	PSTN2	Rynga	Skype
GSM	86	12	0	0	2
PSTN1	8	66	14	12	0
PSTN2	0	8	88	0	4
Rynga	0	4	2	92	2
Skype	0	0	6	0	94

Průměr: **85,20 %**

Tab. 7.4: Zmatková matice pro kombinaci MFCC-SVM

Testování	GSM	PSTN1	PSTN2	Rynga	Skype
GSM	90	8	0	0	2
PSTN1	12	72	10	6	0
PSTN2	0	8	84	2	6
Rynga	0	2	0	94	4
Skype	0	2	4	0	94

Průměr: **86,80 %**

7.3 Výsledky pro kombinace příznaků

V předešlých testech SVM dosahoval ve všech případech nejlepších výsledků. Z tohoto důvodu je dále použitý pro klasifikaci vzájemných kombinací příznaků.

Výsledky pro kombinace různých typů příznaků jsou zobrazeny v tab. 7.5. Kromě jediného případu (kombinace MFCC-SF) je průměrná úspěšnost klasifikace vždy vyšší než při použití jediného typu příznaku. Pro kombinaci dvou typů příznaků mají nejvyšší úspěšnost 90,8 % příznaky MFCC-LPCC. LPCC, které dosahují samostatně úspěšnosti pouze 77,20 %, v kombinaci s jinými příznaky zvyšují úspěšnost klasifikace. Druhou nejvyšší úspěšnost má kombinace MFCC-MELBS. Použitím příznaků typu MFCC-MELBS-LPCC (mRMR = 50, stejně jako v předchozích případech), bylo dosaženo úspěšnosti testování 92 %.

Pro tuto nejlepší kombinaci příznaků byla také otestována závislost průměrné úspěšnosti klasifikace na počtu vybraných suprasegmentálních příznaků - viz obr. 7.2. Výběrem 80 suprasegmentálních příznaků algoritmem mRMR došlo k mírnému zlepšení o 0,8 % na 92,8 %, jak lze vidět v tab. 7.6. U kombinací příznaků je zřetelná také výrazně lepší schopnost rozpoznání kanálů PSTN1 a PSTN2 než při použití jediného typu příznaku. Na obr. 7.3 je znázorněno procentuální zastoupení suprasegmentálních příznaků. Největší měrou se na rozpoznání přenosových kanálů podílejí percentily (použity percentily 1 %, 5 %, 10 %, 20 %, 30 %, 40 %, 60 %, 70 %, 80 %, 90 %, 95 %, 99 % a rozsah percentilů).

Souhrnné výsledky pro všechny použité příznaky a každý kanál jsou zobrazeny v následujících grafech: GSM – 7.4, PSTN1 – 7.5, PSTN2 – 7.6, Rynga – 7.7, Skype – 7.8. Kompletní výsledky všech zmatkových matic jsou seřazeny dle čísla testu a uvedeny v příloze C.

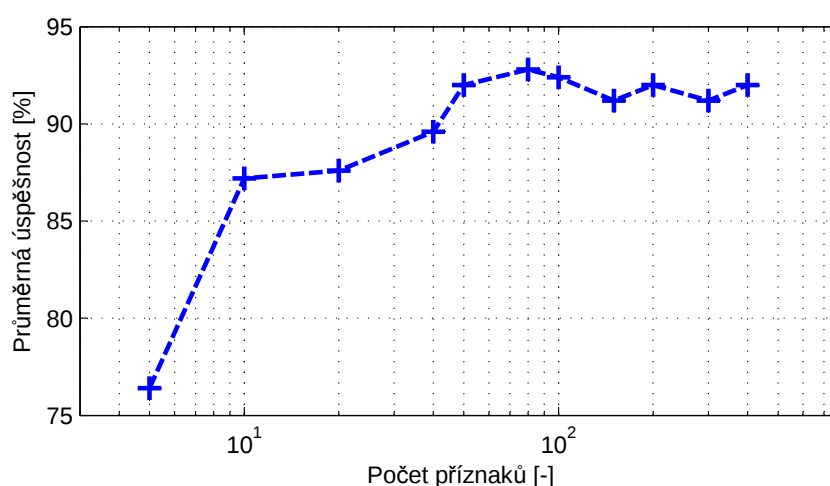
Tab. 7.5: Kombinace příznaků, klasifikátor SVM – průměrná úspěšnost

Příznaky	Validace (%)	Testování (%)	Číslo testu
MFCC-MELBS	90,56	89,60	13
MFCC-LPCC	89,36	90,80	14
MFCC-SF	87,60	86,00	15
MELBS-LPCC	86,24	85,20	16
MELBS-SF	85,84	86,80	17
LPCC-SF	81,04	82,80	18
MFCC-MELBS-LPCC (mRMR = 50)	89,68	92,00	19
MFCC-MELBS-LPCC (mRMR = 80)	90,96	92,80	20
MFCC-MELBS-SF	90,24	88,80	21
MELBS-LPCC-SF	87,20	87,60	22
LPCC-SF-MFCC	88,40	90,40	23

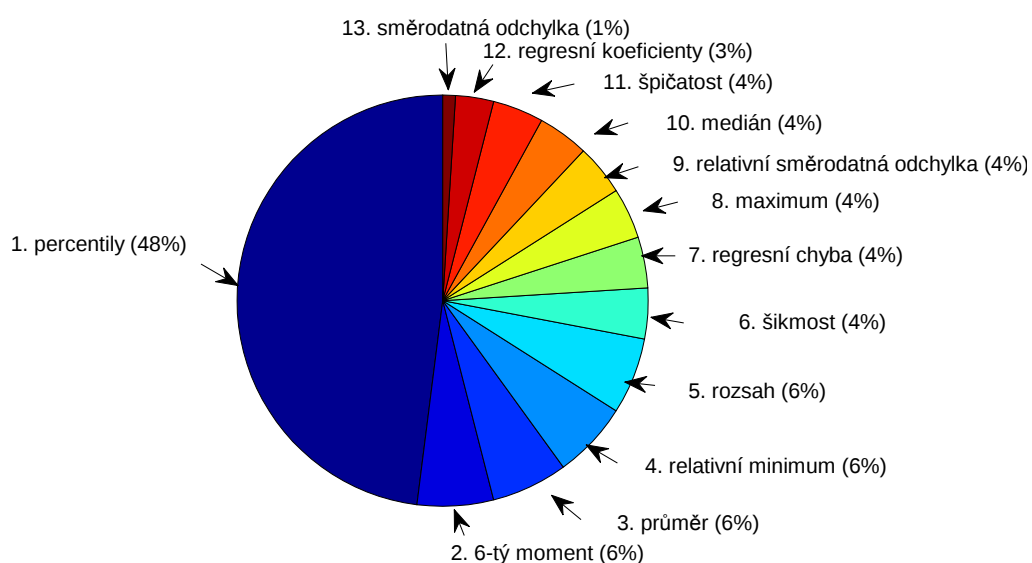
Tab. 7.6: Nejlepší dosažený výsledek – MFCC-MELBS-LPCC, mRMR = 80

Testování	GSM	PSTN1	PSTN2	Rynga	Skype
GSM	94	6	0	0	0
PSTN1	2	92	4	2	0
PSTN2	0	6	92	0	2
Rynga	0	6	0	90	4
Skype	0	0	4	0	96

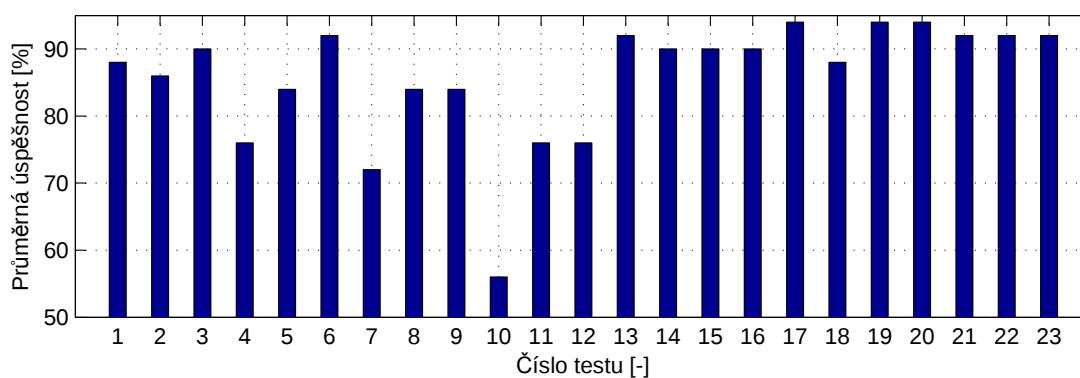
Průměr: **92,80 %**



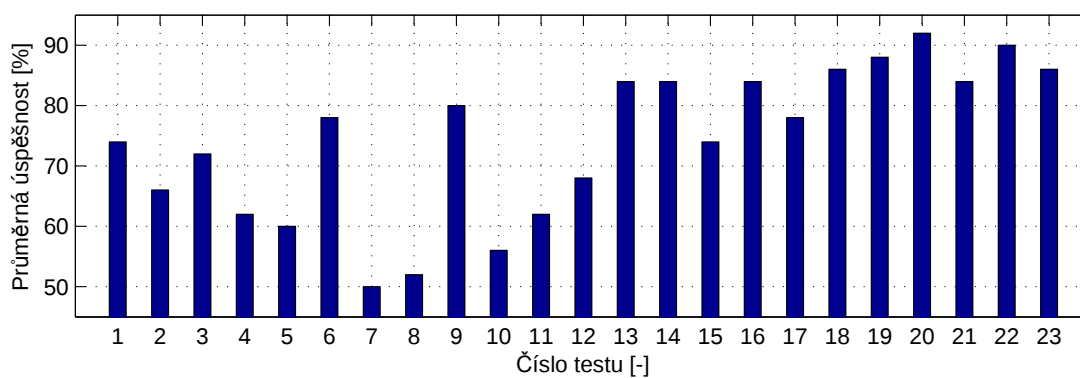
Obr. 7.2: Závislost úspěšnosti klasifikace na počtu vybraných suprasegmentálních příznaků (pro kombinaci MFCC-MELBS-LPCC)



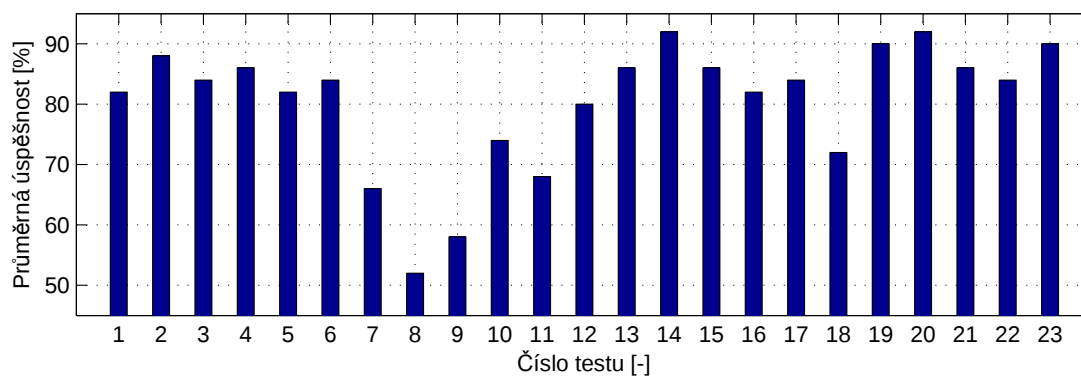
Obr. 7.3: Procentuální zastoupení suprasegmentálních příznaků



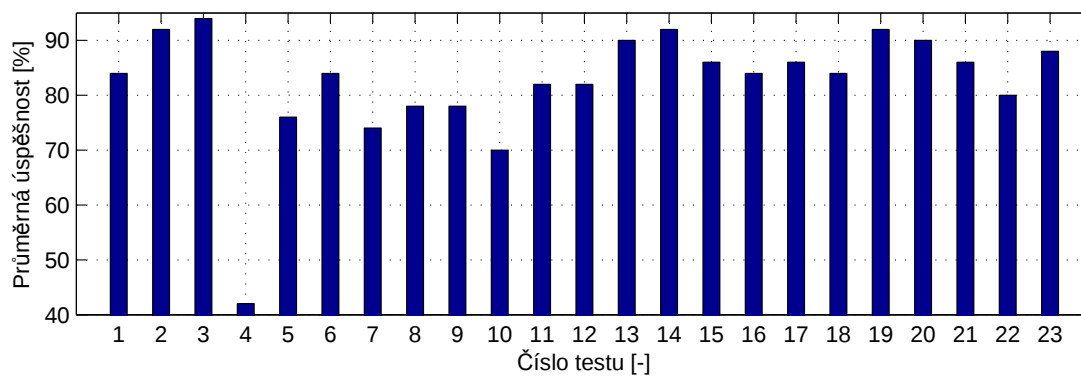
Obr. 7.4: Úspěšnost klasifikace pro kanál GSM



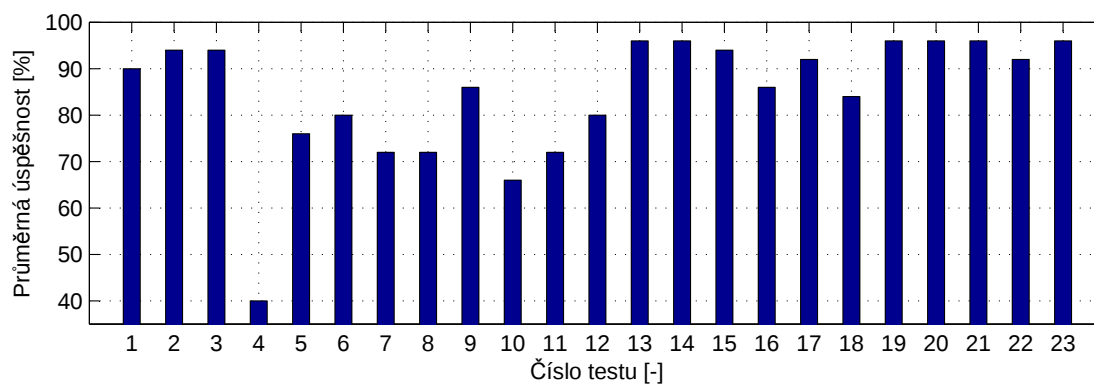
Obr. 7.5: Úspěšnost klasifikace pro kanál PSTN1



Obr. 7.6: Úspěšnost klasifikace pro kanál PSTN2



Obr. 7.7: Úspěšnost klasifikace pro kanál Rynga

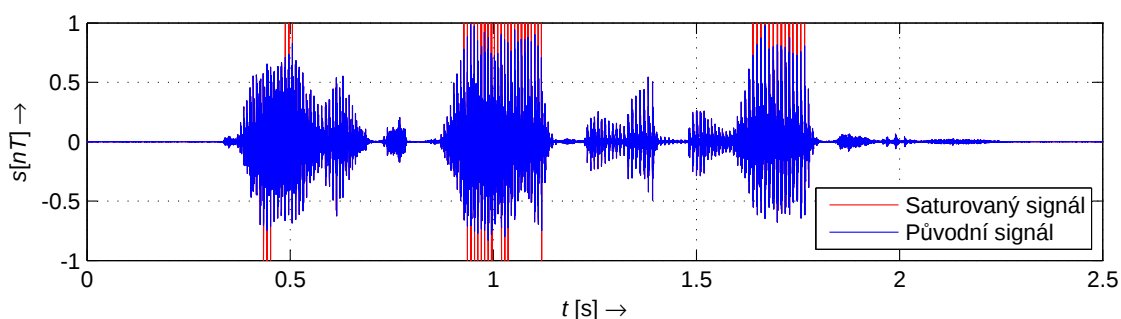


Obr. 7.8: Úspěšnost klasifikace pro kanál Skype

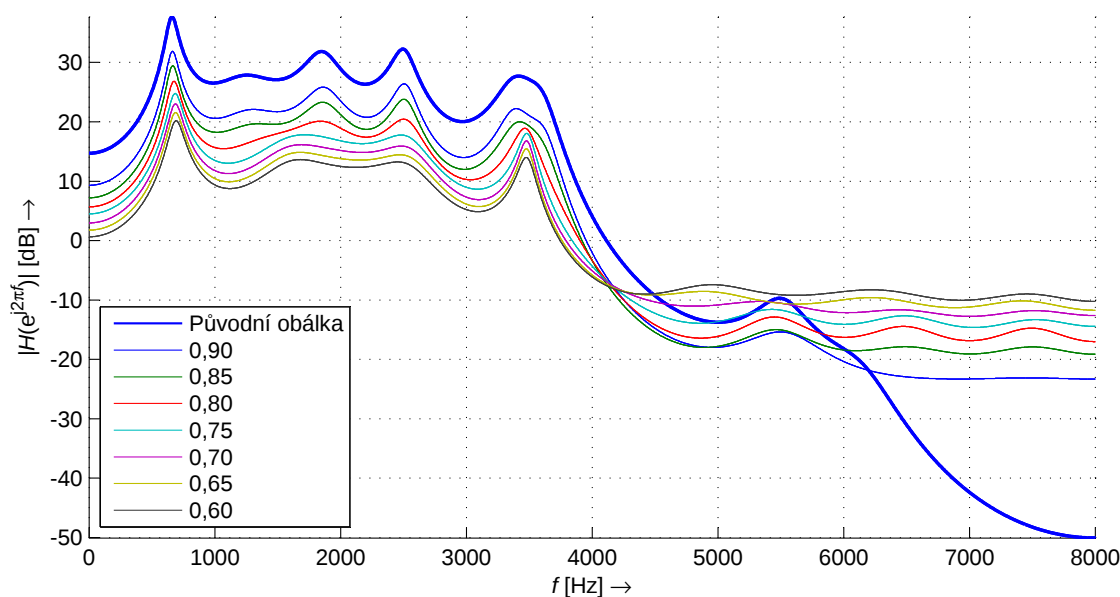
7.4 Výsledky po zkreslení signálu

7.4.1 Saturace

Saturace signálu je stav, při kterém je signál omezen po překročení určitého rozsahu (obr. 7.9). Při snižování prahu, při kterém saturace nastane, dochází k postupnému snižování průměrné úspěšnosti klasifikace a kvality nahrávky – viz tab. 7.7. Obr. 7.10 ilustruje podstatnou změnu obálky spektra při testovaných prazích. Ve spektrální oblasti vlivem saturace lze pozorovat „zploštění“ obálky spektra a výrazný nárůst v oblasti vyšších kmitočtů (přibližně od 6000 Hz). Při úrovních pod 0,7 (přibližně MOS = 4) dochází k výrazné deformaci spektra a systém nedokáže spolehlivě pracovat, což se projeví klasifikací promluv do jediné třídy.



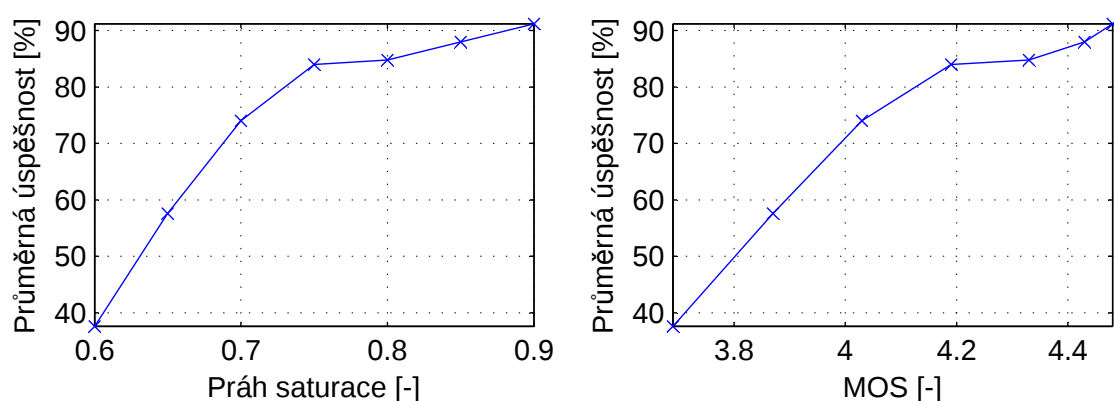
Obr. 7.9: Příklad saturace signálu pro práh 0,7



Obr. 7.10: Vliv saturace na obálku spektra

Tab. 7.7: Vliv saturace na klasifikaci

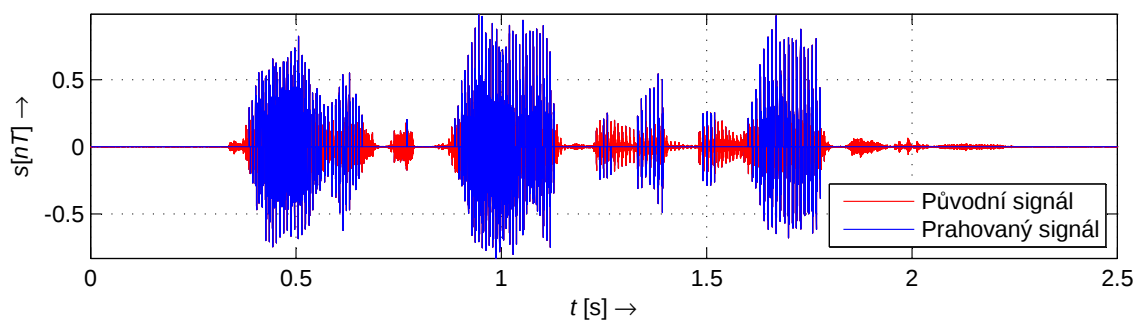
Práh	Prům. úspěšnost (%)	MOS	Číslo testu
0,90	91,2	4,48	24
0,85	88	4,43	25
0,80	84,8	4,33	26
0,75	84	4,19	27
0,70	74	4,03	28
0,65	57,6	3,87	29
0,60	37,6	3,69	30



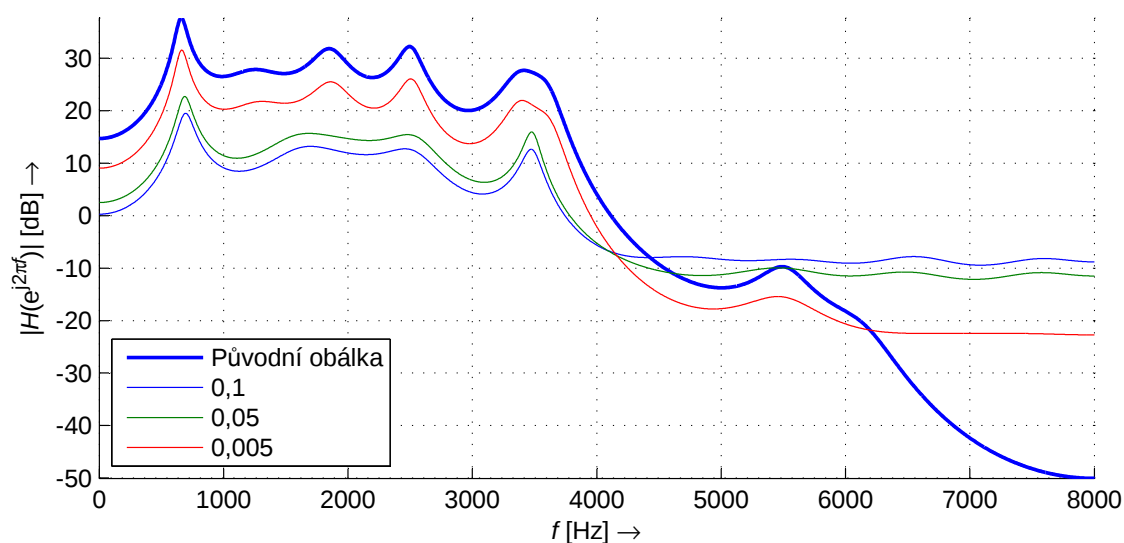
Obr. 7.11: Průměrná úspěšnost klasifikace v závislosti na prahu a MOS

7.4.2 Prahování

Prahování neboli thresholding se projeví vynulováním signálu, pokud nedosahuje určité úrovně (obr. 7.12). Vliv tohoto typu zkreslení je ve frekvenční oblasti více patrný než u saturace. Ani při nastavení nízkých hodnot systém nebyl schopen pracovat a všechny promluvy byly přiřazeny do třídy GSM.



Obr. 7.12: Vliv prahování signálu pro práh 0,2



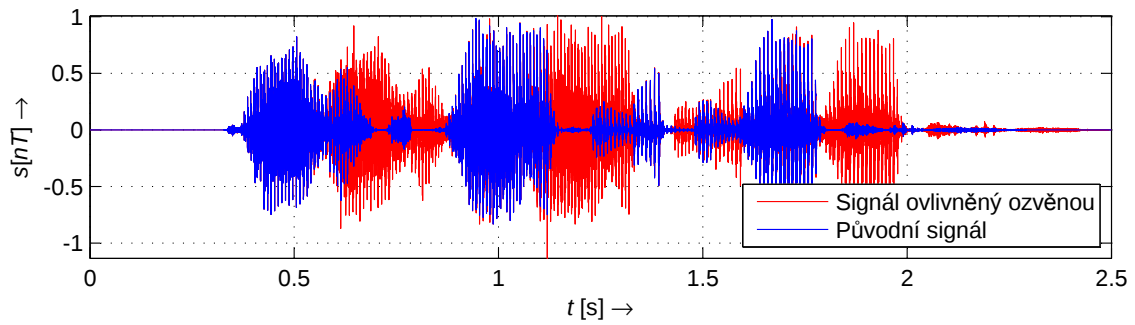
Obr. 7.13: Vliv prahování na obálku spektra

7.4.3 Ozvěna

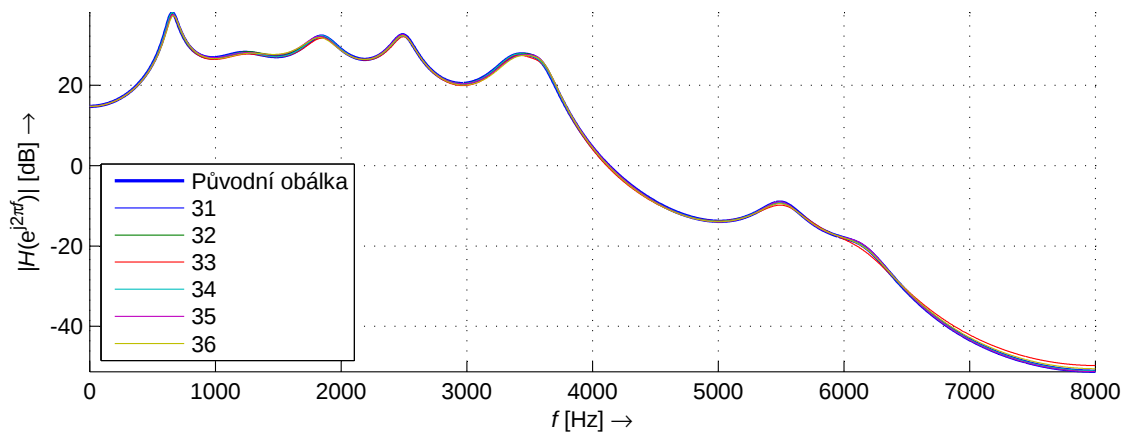
U tohoto typu zkreslení jsou testovány dva parametry – zpoždění a počet ozvěn. Jak ukazuje tab. 7.8, ozvěnou se sníží jak úspěšnost klasifikace (přibližně na 75 %), tak především kvalita (nízká hodnota MOS kolem 1,7). Nicméně délka zpoždění dále nemá velký vliv na klasifikaci ani na změnu hodnoty MOS. Naopak velké změny v časové oblasti způsobené větším počtem ozvěn nepříznivě ovlivňují výsledky. Na obr. 7.14 je příklad poškození signálu ozvěnou v čase, ve frekvenční oblasti se změny ve všech testovaných případech projeví pouze minimálně (obr. 7.15).

Tab. 7.8: Vliv ozvěny na klasifikaci

Počet	Zpoždění	Prům. úspěšnost (%)	MOS	Číslo testu
1	0,1	77,2	1,65	31
1	0,2	75,6	1,66	32
1	0,3	76,4	1,72	33
1	0,4	74,4	1,76	34
2	0,1	52,8	1,40	35
3	0,1	33,6	1,27	36



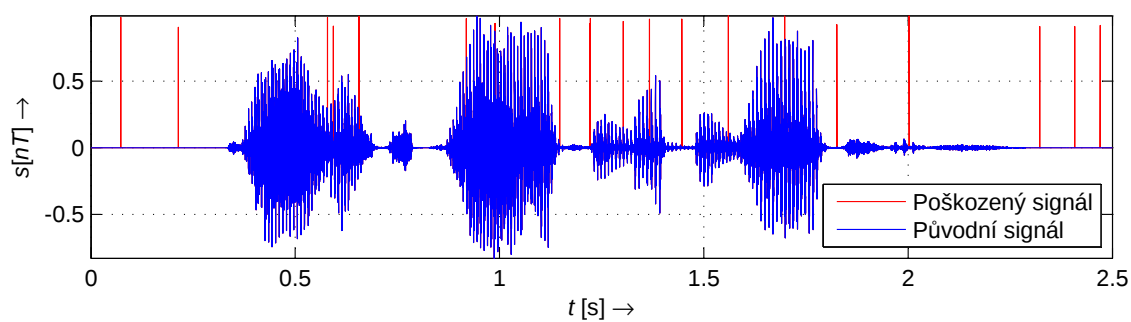
Obr. 7.14: Vliv ozvěny (zpoždění 0,2 s) na signál



Obr. 7.15: Vliv ozvěny na obálku spektra

7.4.4 Praskání

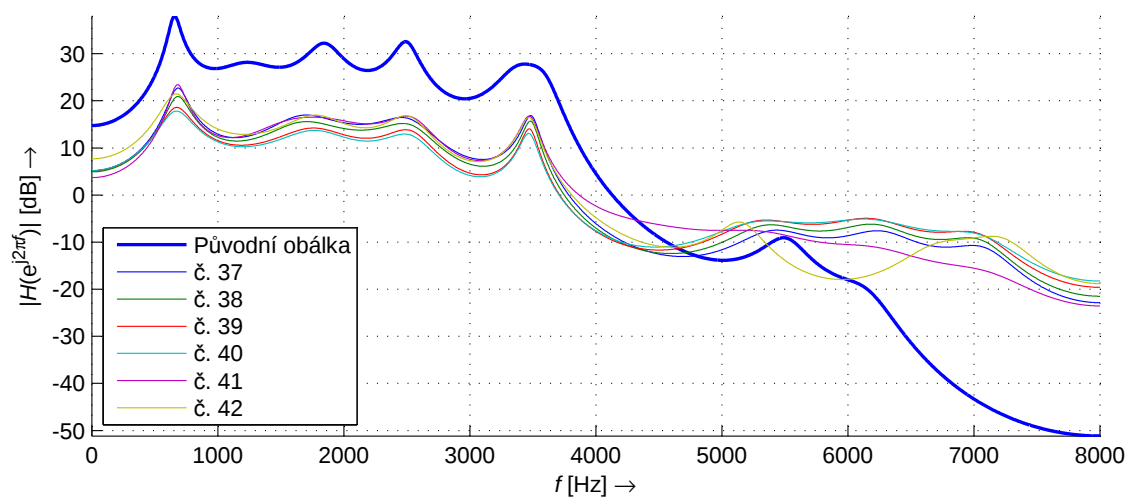
Dalším zkreslujícím jevem, který může být například způsoben vadou mikrofону, je praskání (obr. 7.16). Testování ukázalo, že při poškození signálu více než 10 prasknutími již dochází k podstatnému poškození signálu a následné neschopnosti klasifikace. Podstatný vliv má také „charakter“ praskání, který je realizován různou délkou prasknutí (ve vzorcích – viz sloupec *Délka* v tab. 7.9). Např. prasknutí o délce 2 vzorky není tak zvukově výrazné jako prasknutí délky 8 vzorků. I tento parametr má významný vliv na proces rozpoznání.



Obr. 7.16: Vliv praskání na signál

Tab. 7.9: Vliv praskání na klasifikaci

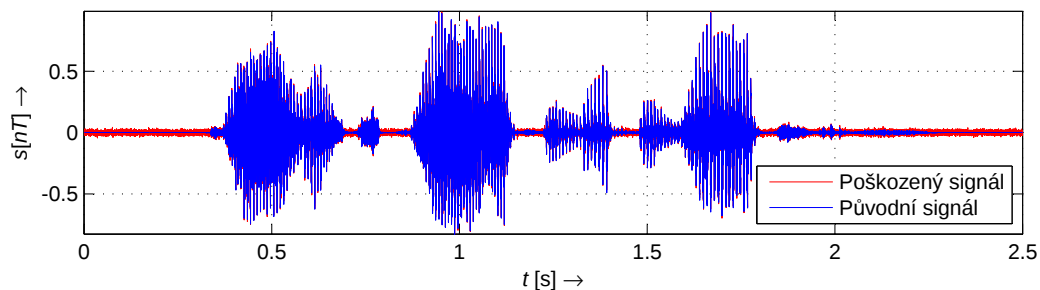
Počet	Délka	Prům. úspěšnost (%)	MOS	Číslo testu
5	4	87,2	3,07	37
10	4	78	2,55	38
20	4	64,8	1,96	39
25	4	53,2	1,73	40
5	2	24,8	3,24	41
5	8	84,4	3,05	42



Obr. 7.17: Vliv praskání na obálku spektra

7.4.5 Šum

Přidání šumu (příklad na obr. 7.18) způsobí podstatné změny ve spektru, které však nejsou tak patrné z průběhu spektrálních obálek – viz obr. 7.19. Zkreslení obálek je zřejmé především na vyšších kmitočtech, v závislosti na použité barvě šumu. Z výsledků uvedených v tab. 7.10 a na obr. 7.20 vyplývá velká citlivost použitých příznaků na přidání šumu.



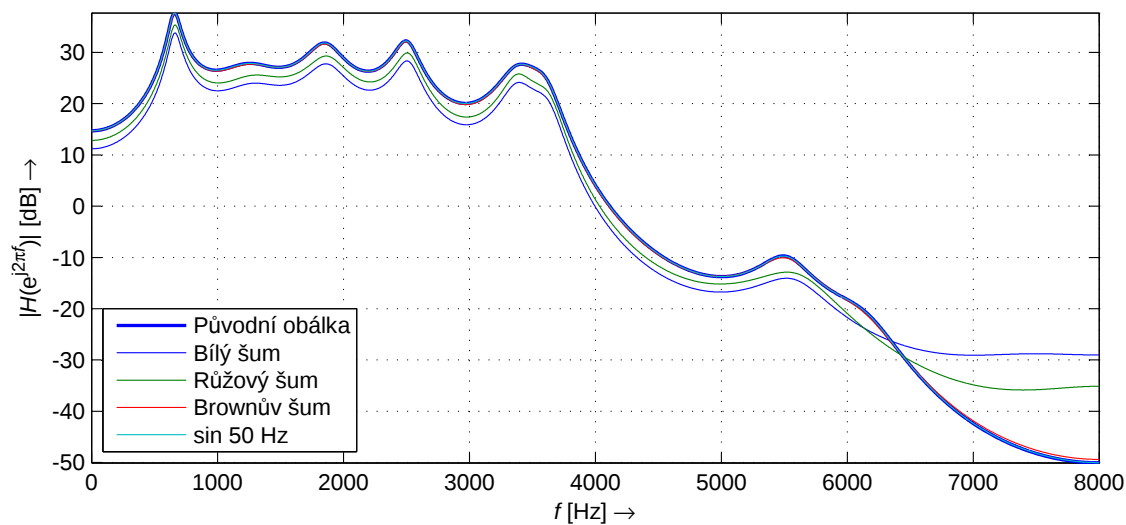
Obr. 7.18: Poškození signálu bílým šumem (SNR = 20)

Tab. 7.10: Vliv šumů na klasifikaci

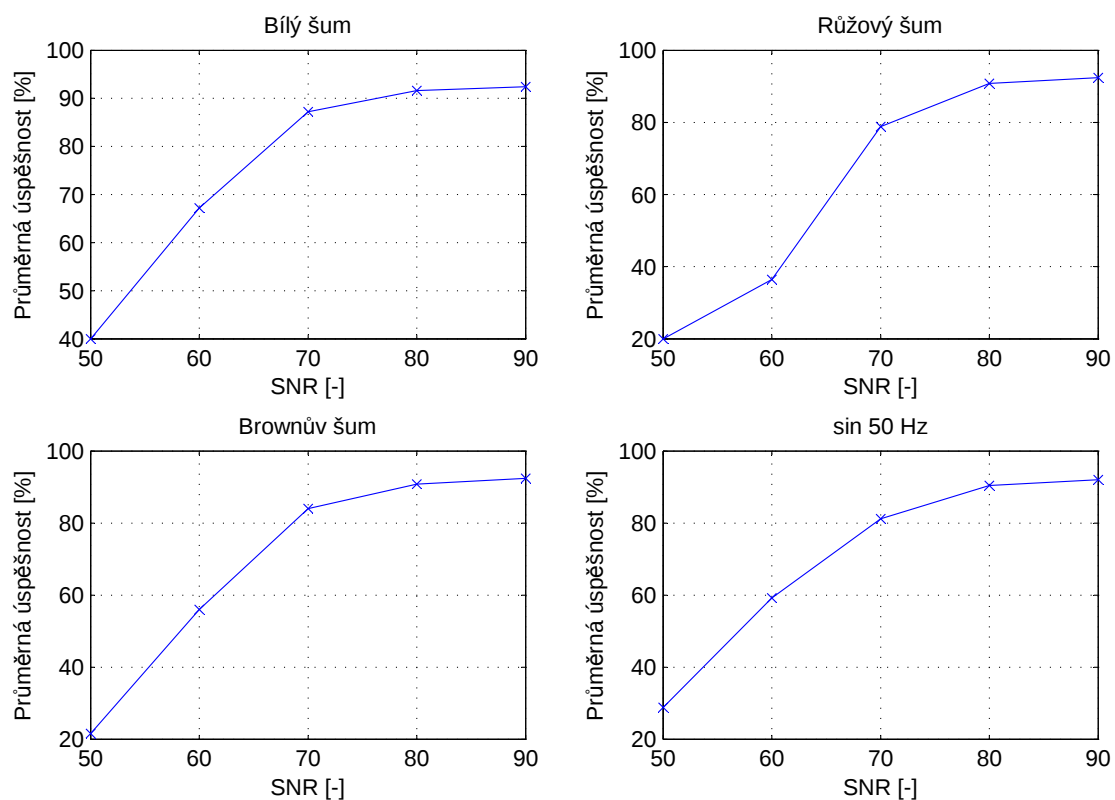
SNR	Průměrná úspěšnost			
	Bílý šum	Růžový šum	Hnědý šum	sin 50 Hz
90	92,4	92,4	92,4	92
80	91,6	90,8	90,8	90,4
70	87,2	78,8	84	81,2
60	67,2	36,4	56	59,2
50	40	20	21,6	28,8
Číslo testů:	43–47	48–52	53–57	58–62

7.4.6 Vliv filtrů

Použitím filtru (testována horní a dolní propust) se změnil celkový charakteristiky přenosových kanálů (porovnání na obr. 6.6). Po změně těchto charakteristik systém nedokáže správně pracovat a všechny promluvy jsou zařazeny do jediné třídy (GSM).



Obr. 7.19: Porovnání obálek signálu poškozeného šumem (SNR = 50)



Obr. 7.20: Vliv šumů na klasifikaci

8 ZÁVĚR

Práce se zabývala rozpoznáním přenosových kanálů z řečového signálu. V rámci řešení byla vytvořena databáze SPLAB_TranCh. Vznikla výběrem 300 promluv z řečového korpusu TIMIT. Tyto promluvy byla následně nahrány v pěti kanálech – GSM, dva různé kanály PSTN a dvě různé VoIP služby (Rynga a Skype). Dohromady SPLAB_TranCh tvoří 5 kanálů, 1500 promluv od 300 mluvčích se stejným zastoupením mužů a žen. Je rozdělena na část pro trénování a testování. Tato databáze byla použita pro klasifikaci.

Cílem práce bylo najít nejvhodnější kombinaci příznaků (MFCC, MELBS, LPCC, SF) a klasifikátorů (GMM, KNN, SVM) pro dosažení nejvyšší úspěšnosti klasifikace. V programu Hila byl realizován výpočet příznaků, suprasegmentálních příznaků, jejich redukce algoritmem mRMR a klasifikace. Po validaci typu vynechání jednoho vzorku následovalo testování s využitím testovací části databáze SPLAB_TranCh.

Při použití jediného typu příznaku dosahují nejlepších výsledků příznaky MFCC s nejvyšší průměrnou úspěšností testování 86,8 % a ve všech případech je nejvhodnější klasifikátor SVM. Ten je následně vybrán pro další testy s více typy příznaků. Jako nejvhodnější se ukázala kombinace MFCC-MELBS-LPCC s optimálním výběrem 80 příznaků a úspěšností 92,8 %. Vysoké procentuální úspěšnosti dokazují schopnost rozpoznání jednotlivých kanálů – dokonce i těch, které jsou stejného druhu např. PSTN1, PSTN2 a Rynga, Skype.

Následující část práce se zabývala jevy, které mohou poškodit signál na přenosové trase a tím ovlivnit výsledky klasifikace. Zkreslení signálu bylo realizováno modulem Midprocessing vytvořeným v programu Matlab. Některé jevy svou přítomností úspěšnost rozpoznání pouze úměrně snižují (saturace, ozvěna), jiné svou povahou klasifikaci úplně znemožňují (prahování, vliv filtrů).

Tato práce ukázala, jak zkreslení ovlivňuje frekvenční charakteristiky přenosových kanálů, především na vyšších kmitočtech. Kromě frekvenční oblasti hraje důležitou roli také časová oblast, ve které jsou změny sledovány suprasegmentálními příznaky. Ty poznatky je třeba brát v úvahu při dalším zdokonalování navrženého systému. Nalezení vhodnějších příznaků a nových klasifikačních postupů pro zvýšení robustnosti systému bude cílem dalšího výzkumu.

Výsledky této práce [6] byly publikovány v přijatém článku na *36th International Conference on Telecommunications and Signal Processing (TSP), 2013*.

LITERATURA

- [1] ČERNOCKÝ, J. *Zpracování řečových signálů – studijní opora*. VUT FIT v Brně, 2006. 129 s.
- [2] DI, G., XIONG, X., GUANGXI, Z., ENG, S. C., HAIZHOU L. Classification of speech transmission channels: Landline, GSM and VoIP networks. *Signal Processing, 2008. ICSP 2008. 9th International Conference on*, pp.671–675, 26–29 Oct. 2008.
- [3] GAROFOLO John S., et al. *TIMIT Acoustic-Phonetic Continuous Speech Corpus*. Linguistic Data Consortium, Philadelphia, 1993.
- [4] HANCHUAN, P., FUHUI, L., DING, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.
- [5] HONG, W. An RNN-based channel classification for Mandarin speech recognition over GSM/PSTN transmission environments. *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, vol. 1, pp. I-1033–I-1036, 13–17 May 2002.
- [6] KOPŘIVA, T., ATASSI, H., HUSSAIN, A. Classification of Transmission Channels by Speech Signal Processing. *Telecommunications and Signal Processing (TSP), 2013 36th International Conference on*, 2–4 July 2013.
- [7] PSUTKA, J., MÜLLER, L., MATOUŠEK, J., RADOVÁ, V. *Mluvíme s počítačem česky*. 1. vyd. Praha: Academia, 2006. ISBN 80-200-1309-1.
- [8] RIX, A. W., BEERENDS, J. G., HOLLIER, M. P., HEKSTRA, A. P. Perceptual evaluation of speech quality (PESQ) – a new method for speech quality assessment of telephone networks and codecs. *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on*, vol. 2, pp. 749–752, 2001.
- [9] REYNOLDS, D., QUATIERI, T., DUNN, R. *Speaker Verification Using Adapted Gaussian Mixture Models*. Digital Signal Processing 10, 19–41, 2000.
- [10] Rynga, VoIP aplikace. Dostupné z: <http://www.rynga.com>.
- [11] SMÉKAL, Z. *Číslicové zpracování řeči*. Skriptum VUT FEKT v Brně, 2010. 135 s.

- [12] Skype, VoIP aplikace. Dostupné z: <http://www.skype.com>.
- [13] SMITH, J. O. *Spectral Audio Signal Processing* [online]. [cit. 2013-05-17]. Dostupné z: <http://ccrma.stanford.edu/~jos/sasp/>.
- [14] XU, W., PENG, X., WANG, B. A GMM-based telephone channel classification for Mandarin speech recognition. *Signal Processing, 2004. Proceedings. ICSP '04. 2004 7th International Conference on*, vol. 1, pp. 642–645, 31 Aug.–4 Sept. 2004.
- [15] ZAPLATÍLEK, K., DOŇAR, B. *MATLAB – tvorba uživatelských aplikací*. 1. vydání. Praha: BEN – technická literatura, 2005. 216 s. ISBN 80-7300-133-0.
- [16] ZAPLATÍLEK, K., DOŇAR, B. *MATLAB – začínáme se signály*. 1. vydání. Praha: BEN – technická literatura, 2006. 272 s. ISBN 80-7300-200-0.

SEZNAM SYMBOLŮ, VELIČIN A ZKRATEK

EM Expectation-Maximization

FFT Fast Fourier Transform – rychlá Fourierova transformace

GMM Gaussian Mixture Models – Gaussovy smíšené modely

GSM Groupe Spécial Mobile – globální systém pro mobilní komunikaci

KNN k -Nearest Neighbor – k -nejbližší soused

LPC Linear Predictive Coding – lineární prediktivní kódování

MELBS Mel-frequency bank spectrum – melovské spektrální příznaky banky

MFCC Mel-frequency cepstral coefficients – mel-frekvenční cepstrální koeficienty

MOS Mean Opinion Scores

mRMR minimum Redundancy Maximum Relevance

PCM Pulse-code modulation – pulsně kódová modulace

PESQ Perceptual Evaluation of Speech Quality

PSTN Public Switched Telephone Network – veřejná telefonní síť

RNN Recurrent Neural Network – rekurzivní neuronová síť

SF Spectral Features – spektrální příznaky

SNR Signal-to-Noise Ratio - poměr signálu k šumu

SVM Support Vector Machine – algoritmus podpůrných vektorů

VoIP Voice over Internet Protocol – přenos digitalizovaného hlasu v IP paketech

Δe_{log} derivace logaritmu krátkodobé energie

$\Delta^2 e_{log}$ druhá derivace logaritmu krátkodobé energie

Δ MFCC derivace MFCC

$a_j^{(i)}$ j -tý parametr řádu

α_i váhovací parametr

c cílová třída

$c_m(j)$	j -tý koeficient MFCC
$c(k)$	kepstrální koeficienty LPC
d	rozměr Gaussovy funkce
σ	směrodatná odchylka
σ_r	spektrální rozptyl
f	frekvence
f_m	frekvence v melovské škále
f_{vz}	vzorkovací kmitočet
$f(n)$	n -tá složka signálu
G	koeficient zesílení
γ_k	spektrální špičatost
γ_s	spektrální šikmost
$H(z)$	přenosová funkce hlasového traktu
$I(c, f_i)$	vzájemná informace mezi třídou a příznakem
$I(f_i, f_j)$	vzájemná informace mezi příznaky
k_i	koeficienty odrazu
M	počet MFCC
$M(n)$	n -tý vzorek modulu
M_G	počet Gaussových funkcí rozložení pravděpodobnosti
M^*	počet pásem melovského filtru
μ	vektor středních hodnot
μ_c	spektrální centroid
N	počet sloupců matice
$N_i(x)$	funkce rozložení pravděpodobnosti
Q	řád modelu

R_n	krátkodobá autokorelační funkce
$s(k)$	k -tý vzorek signálu
s_s	spektrální sklon
S	množina redukováných příznaků
$ S $	počet příznaků v množině S
V	relevance
W	redundance
Ω	množina všech příznaků
x_i	prvek řádku matice
y_γ	k -tý centrální moment
$y_{\gamma k}$	špičatost
$y_{\gamma s}$	šikmost
y_m	průměrná hodnota
$y_{\max}$	maximum
$y_{\max r}$	relativní maximum
y_{med}	medián
$y_{\min}$	minimum
$y_{\min r}$	relativní minimum
y_r	rozsah
y_{rc}	regresní koeficienty
y_{re}	regresní chyba
y_{rr}	relativní rozsah
y_s	směrodatná odchylka
y_{sr}	relativní směrodatná odchylka
Σ	kovariační matice

SEZNAM PŘÍLOH

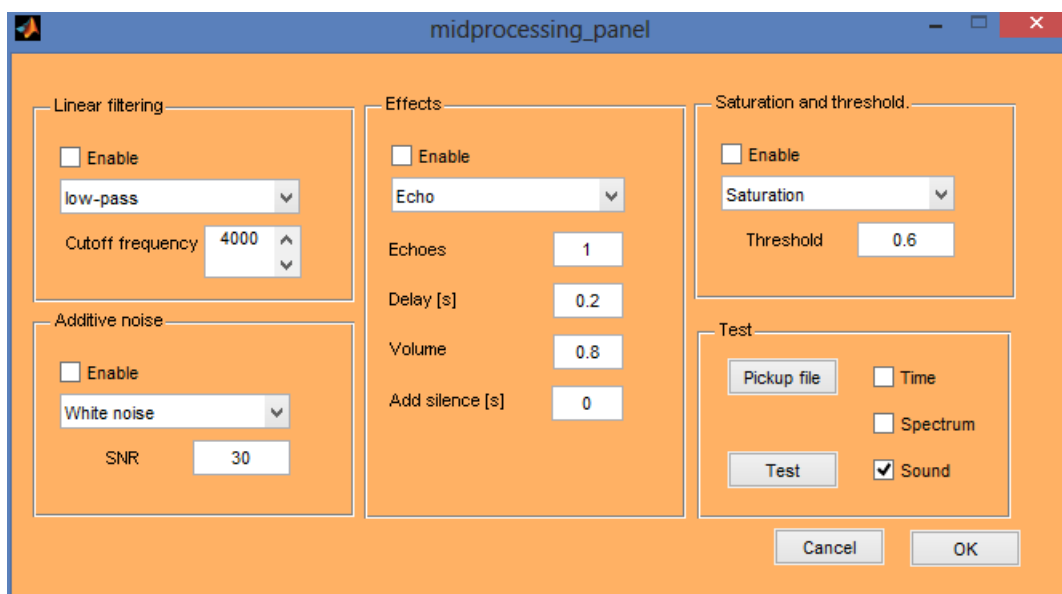
A Modul Midprocessing	53
B Obsah CD	54
C Kompletní výsledky	55

A MODUL MIDPROCESSING

V rámci řešení bakalářské práce byl vytvořen matlabovský program Midprocessing, který je součástí přiloženého CD. Tento program je součástí modulového nástroje Hila, jehož autorem je Ing. Hicham Atassi.

Program umožňuje poškodit testované promluvy zvoleným zkreslením. Okno modulu Midprocessing (obr. A.1) je rozděleno do několika částí:

- **Linear filtering** – výběr typu filtru (horní, dolní a pásmová propust, pásmová zádrž) a mezních kmitočtů.
- **Additive noise** – výběr šumu (bílý, růžový, hnědý, sin 50 Hz), a odstup signál-šum.
- **Effects** – výběr ozvěny (parametry: počet ozvěn, zpoždění, hlasitost ozvěny, přidání ticha) nebo praskání (parametry: počet prasknutí, délka prasknutí ve vzorcích).
- **Saturation and Thresholding** – výběr saturace nebo prahování, nastavení prahu.
- **Test** – otevření okna pro výběr promluvy (tlačítko Pickup file), kterou lze použít pro vyzkoušení nastavených parametrů. Volbou zaškrťovacích políček a tlačítkem Test lze zobrazit původní a poškozený signál v čase, modulové kmitočtové spektrum včetně obálek a přehrát poškozený signál.



Obr. A.1: Okno modulu Midprocessing

B OBSAH CD

- bakalářská práce ve formátu pdf
- složka Midprocessing se zdrojovými kódy modulu Midprocessing
- databáze SPLAB_TrnCh

C KOMPLETNÍ VÝSLEDKY

Tab. C.1: Test č. 1

85,2	12	0,4	0	2,4
7,2	71,2	13,2	6,4	2
0	8	83,6	1,6	6,8
1,2	5,6	4	85,6	3,6
0,4	2,8	9,2	1,2	86,4
Validace - průměr: 82,40 %				
88	12	0	0	0
10	74	14	2	0
0	8	82	0	10
0	6	4	84	6
0	2	8	0	90
Testování - průměr: 83,60 %				

Tab. C.2: Test č. 2

84,8	11,6	0,8	1,6	1,2
2,8	72,4	13,6	10	1,2
0	10,8	81,6	2,4	5,2
1,6	2,8	2,8	92,4	0,4
0,4	2	8	3,2	86,4
Validace - průměr: 83,52 %				
86	12	0	0	2
8	66	14	12	0
0	8	88	0	4
0	4	2	92	2
0	0	6	0	94
Testování - průměr: 85,20 %				

Tab. C.3: Test č. 3

92	5,6	0,8	0,4	1,2
4	79,6	10,4	3,6	2,4
0	7,2	84,4	2,4	6
1,6	3,2	2,4	92	0,8
0,4	1,6	4,8	0,4	92,8
Validace - průměr: 88,16 %				
90	8	0	0	2
12	72	10	6	0
0	8	84	2	6
0	2	0	94	4
0	2	4	0	94
Testování - průměr: 86,80 %				

Tab. C.4: Test č. 4

80	17,6	0,4	2	0
3,2	59,6	30	4,4	2,8
0,8	6	86,8	0,4	6
2,8	28,4	11,6	47,2	10
0	10,8	45,2	5,2	38,8
Validace - průměr: 62,48 %				
76	18	2	4	0
6	62	22	6	4
0	10	86	0	4
2	34	12	42	10
0	16	38	6	40
Testování - průměr: 61,20 %				

Tab. C.5: Test č. 5

88	10	0,4	1,2	0,4
3,2	68	15,2	8	5,6
1,2	6	82	3,2	7,6
2,4	7,2	6	71,6	12,8
0	5,6	15,6	6,8	72
Validace - průměr: 76,32 %				
84	16	0	0	0
6	60	22	10	2
0	6	82	0	12
2	8	6	76	8
0	8	8	8	76
Testování - průměr: 75,60 %				

Tab. C.6: Test č. 6

94	5,6	0	0,4	0
4,8	76	10	5,2	4
0,8	6,4	85,2	2	5,6
1,6	6	2	82	8,4
2	5,2	7,6	4,8	80,4
Validace - průměr: 83,52 %				
92	8	0	0	0
6	78	6	8	2
0	6	84	0	10
2	6	0	84	8
0	6	6	8	80
Testování - průměr: 83,60 %				

Tab. C.7: Test č. 7

74	18,4	1,2	3,6	2,8
8,8	39,6	34,4	11,2	6
1,6	9,2	76,8	2,4	10
2,4	5,6	3,6	78	10,4
1,6	3,6	14,4	7,6	72,8
Validace - průměr: 68,24 %				
72	16	4	6	2
6	50	24	10	10
0	22	66	2	10
8	8	6	74	4
2	4	20	2	72

Testování - průměr: **66,80** %

Tab. C.8: Test č. 8

84,4	9,6	1,2	2,8	2
12,4	52,8	22	10	2,8
1,6	21,2	67,2	2	8
1,6	9,2	2,8	80,4	6
3,2	8	10,4	6	72,4
Validace - průměr: 71,44 %				
84	8	2	6	0
10	52	30	6	2
6	34	52	0	8
2	12	0	78	8
4	6	10	8	72

Testování - průměr: **67,60** %

Tab. C.9: Test č. 9

87,6	8,8	0,4	1,6	1,6
11,2	56,8	20	9,6	2,4
1,6	15,2	76	0,8	6,4
2	6,4	0,8	85,2	5,6
2	5,6	7,2	3,2	82
Validace - průměr: 77,52 %				
textbf84	6	4	4	2
4	80	14	2	0
6	28	58	0	8
4	10	2	78	6
2	8	2	2	86

Validace - průměr: **77,20** %

Tab. C.10: Test č. 10

69,2	18	6	5,6	1,2
18,8	43,2	21,6	10,4	6
0,8	11,6	72	1,2	14,4
4,4	7,2	3,6	72	12,8
1,2	5,6	14	11,6	67,6
Validace - průměr: 64,80 %				
56	28	10	4	2
14	56	12	12	6
4	8	74	2	12
6	12	2	70	10
0	8	12	14	66

Validace - průměr: **64,40** %

Tab. C.11: Test č. 11

70,8	16,8	3,6	4,8	4
19,2	53,2	10,4	10,8	6,4
3,2	14,4	68	1,6	12,8
2,8	6,4	0,8	82,8	7,2
5,2	7,2	12,8	9,6	65,2

Validace - průměr: **68,00** %

Tab. C.12: Test č. 12

78,8	13,2	4,8	2	1,2
15,2	55,6	14,4	9,2	5,6
2	9,2	80,4	0	8,4
2,4	7,6	2	81,2	6,8
4	5,6	12	8,4	70
Validace - průměr: 73,20 %				
76	14	2	4	4
14	68	4	8	6
6	8	80	2	4
4	10	4	82	0
4	0	12	4	80

Validace - průměr: **77,20** %

Tab. C.13: Test č. 13

96	3,6	0,4	0	0
1,6	81,6	12,4	2,8	1,6
0,4	5,6	88,8	1,2	4
1,2	4	1,6	92,8	0,4
0,4	0,4	5,2	0,4	93,6
Validace - průměr: 90,56 %				
92	6	0	0	2
6	84	8	2	0
0	8	86	2	4
0	2	0	90	8
0	0	4	0	96

Validace - průměr: **89,60** %

Tab. C.14: Test č. 14

92,8	6,4	0	0,4	0,4
4	78,4	12	2,8	2,8
0,4	4,4	89,2	0,4	5,6
0	4,4	1,2	93,6	0,8
0,4	1,2	4,4	1,2	92,8
Validace - průměr: 89,36 %				
90	10	0	0	0
4	84	10	2	0
0	6	92	0	2
0	6	0	92	2
0	0	4	0	96

Validace - průměr: **90,80** %

Tab. C.15: Test č. 15

89,6	8,4	0,8	0	1,2
4,8	79,2	10,8	3,6	1,6
0	8	86	1,2	4,8
0,8	4,4	2,4	91,6	0,8
0,4	1,2	5,6	1,2	91,6
Validace - průměr: 87,60 %				
90	8	2	0	0
12	74	8	6	0
0	10	86	0	4
0	8	0	86	6
0	0	6	0	94
Validace - průměr: 86,00 %				

Tab. C.16: Test č. 16

94	5,2	0,4	0	0,4
4,8	77,2	11,2	4,4	2,4
0,8	7,2	83,6	0,8	7,6
1,2	4	1,6	90,8	2,4
0,8	3,2	6,8	3,6	85,6
Validace - průměr: 86,24 %				
90	8	0	0	2
4	84	6	2	4
0	10	82	0	8
0	10	0	84	6
0	6	4	4	86
Validace - průměr: 85,20 %				

Tab. C.17: Test č. 17

94	5,6	0,4	0	0
4,8	78	9,6	4	3,6
1,2	4,4	84,4	0,8	9,2
0,8	2,8	2,4	88,8	5,2
0,8	3,6	9,6	2	84
Validace - průměr: 85,84 %				
94	4	0	0	2
8	78	2	4	8
0	8	84	0	8
2	6	0	86	6
0	4	2	2	92
Validace - průměr: 86,80 %				

Tab. C.18: Test č. 18

88,40	9,2	0	1,6	0,8
11,2	63,6	13,6	8,8	2,8
1,2	11,2	79,6	0,4	7,6
0,8	6	1,6	90,4	1,2
2,4	5,2	7,6	1,6	83,2
Validace - průměr: 81,04 %				
88	6	2	2	2
4	86	4	2	4
0	16	72	2	10
8	6	0	84	2
4	6	6	0	84
Validace - průměr: 82,80 %				

Tab. C.19: Test č. 19

96,4	3,6	0	0	0
2,8	78,4	13,6	3,6	1,6
0	8,4	87,2	0,4	4
0	4	1,2	94	0,8
0,4	1,2	4,4	1,6	92,4
Validace - průměr: 89,68 %				
94	4	2	0	0
2	88	8	2	0
0	6	90	0	4
0	6	0	92	2
0	0	4	0	96
Testování - průměr: 92,00 %				

Tab. C.20: Test č. 20

96,4	3,2	0	0	0,4
1,2	81,6	13,6	2,4	1,2
0,4	6	90	0,8	2,8
0	3,6	1,6	94,4	0,4
0,8	1,6	4	1,2	92,4
Validace - průměr: 90,96 %				
94	6	0	0	0
2	92	4	2	0
0	6	92	0	2
0	6	0	90	4
0	0	4	0	96
Testování - průměr: 92,80 %				

Tab. C.21: Test č. 21

95,6	3,6	0,4	0,4	0
1,6	84	10	3,2	1,2
0,4	6	88	2	3,6
0,8	4,8	1,6	91,6	1,2
0,4	0,8	4,8	2	92
Validace - průměr: 90,24 %				
92	6	2	0	0
4	84	6	6	0
0	8	86	2	4
2	6	0	86	6
0	0	4	0	96
Validace - průměr: 88,80 %				

Tab. C.22: Test č. 22

95,2	4,4	0,4	0	0
4	78	11,2	4	2,8
0	6,4	85,2	0,4	8
0,8	4	1,6	91,2	2,4
0,8	4	7,2	1,6	86,4
Validace - průměr: 87,20 %				
92	6	0	0	2
4	90	4	0	2
0	8	84	0	8
4	10	0	80	6
0	4	2	2	92
Validace - průměr: 87,60 %				

Tab. C.23: Test č. 23

94	4,8	0	0,4	0,8
4,8	77,2	12	3,2	2,8
0,4	7,2	87,2	0,8	4,4
0,4	4,4	1,6	92,4	1,2
0,4	1,6	4,8	2	91,2

Validace - průměr: **88,40 %**

92	8	0	0	0
4	86	8	2	0
0	10	90	0	0
2	10	0	88	0
0	0	4	0	96

Validace - průměr: **90,40 %**

Tab. C.24: Test č. 24

94	6	0	0	0
4	90	4	2	0
0	6	90	0	4
0	6	0	86	8
0	0	4	0	96

Testování - průměr: **91,20 %**

Tab. C.25: Test č. 25

94	6	0	0	0
8	84	6	2	0
0	8	80	0	12
0	6	0	86	8
0	0	4	0	96

Testování - průměr: **88,00 %**

Tab. C.26: Test č. 26

96	2	0	0	2
16	68	10	4	2
2	4	78	0	16
0	6	0	86	8
0	0	4	0	96

Testování - průměr: **84,80 %**

Tab. C.27: Test č. 27

96	2	0	0	2
20	70	6	4	0
12	4	66	0	18
0	4	0	88	8
0	0	0	0	100

Testování - průměr: **84,00 %**

Tab. C.28: Test č. 28

98	0	0	0	2
46	48	4	2	0
26	0	52	0	22
6	4	0	82	8
10	0	0	0	90

Testování - průměr: **74,00 %**

Tab. C.29: Test č. 29

100	0	0	0	0
76	20	0	2	2
62	0	30	0	8
24	2	0	70	4
32	0	0	0	68

Testování - průměr: **57,60 %**

Tab. C.30: Test č. 30

100	0	0	0	0
92	6	0	2	0
94	0	4	0	2
56	0	0	44	0
66	0	0	0	34

Testování - průměr: **37,60 %**

Tab. C.31: Test č. 31

90	10	0	0	0
18	80	0	2	0
12	12	76	0	0
10	6	0	80	4
2	8	16	14	60

Testování - průměr: **77,20 %**

Tab. C.32: Test č. 32

94	6	0	0	0
18	78	4	0	0
18	10	72	0	0
14	10	2	72	2
8	4	20	6	62

Testování - průměr: **75,60 %**

Tab. C.33: Test č. 33

90	10	0	0	0
18	76	6	0	0
16	8	76	0	0
12	8	0	78	2
10	4	16	8	62

Testování - průměr: **76,40 %**

Tab. C.34: Test č. 34

92	8	0	0	0
22	70	8	0	0
22	8	70	0	0
22	4	0	74	0
12	6	20	6	56

Testování - průměr: **72,40 %**

Tab. C.35: Test č. 35

98	2	0	0	0
52	48	0	0	0
42	12	46	0	0
54	2	0	44	0
38	8	10	16	28

Testování - průměr: **52,80 %**

Tab. C.36: Test č. 36

98	2	0	0	0
76	24	0	0	0
80	4	16	0	0
78	4	0	18	0
70	6	6	6	12

Testování - průměr: **33,60** %

Tab. C.37: Test č. 37

96	0	2	0	2
8	70	20	2	0
0	2	88	0	10
0	8	0	84	8
0	0	2	0	98

Testování - průměr: **87,20** %

Tab. C.38: Test č. 38

96	0	2	0	2
20	42	24	2	12
2	2	74	0	22
2	6	2	78	12
0	0	0	0	100

Testování - průměr: **78,00** %

Tab. C.39: Test č. 39

96	0	0	0	4
34	20	12	2	32
8	0	36	0	56
10	0	2	72	16
0	0	0	0	100

Testování - průměr: **64,80** %

Tab. C.40: Test č. 40

96	0	0	0	4
50	2	4	6	38
14	0	14	0	72
22	0	2	60	16
6	0	0	0	94

Testování - průměr: **53,20** %

Tab. C.41: Test č. 41

100	0	0	0	0
100	0	0	0	0
100	0	0	0	0
92	0	0	8	0
84	0	0	0	16

Testování - průměr: **24,80** %

Tab. C.42: Test č. 42

96	0	2	0	2
14	62	22	2	0
0	2	84	0	14
4	4	0	82	10
0	0	2	0	98

Testování - průměr: **84,40** %

Tab. C.43: Test č. 43

94	6	0	0	0
2	90	6	2	0
0	6	92	0	2
0	6	0	90	4
0	0	4	0	96

Testování - průměr: **92,40** %

Tab. C.44: Test č. 44

94	6	0	0	0
2	90	6	2	0
0	10	88	0	2
0	6	0	90	4
0	0	4	0	96

Testování - průměr: **91,60** %

Tab. C.45: Test č. 45

88	8	0	4	0
2	96	0	2	0
0	18	68	4	10
0	6	0	88	6
0	0	4	0	96

Testování - průměr: **87,20** %

Tab. C.46: Test č. 46

52	14	0	32	2
0	78	0	22	0
0	30	20	42	8
0	2	0	96	2
0	4	2	4	90

Testování - průměr: **67,20** %

Tab. C.47: Test č. 47

58	0	0	42	0
30	2	0	68	0
28	0	0	72	0
12	0	0	88	0
12	0	0	36	52

Testování - průměr: **40,00** %

Tab. C.48: Test č. 48

94	6	0	0	0
2	90	6	2	0
0	6	92	0	2
0	6	0	90	4
0	0	4	0	96

Testování - průměr: **92,40** %

Tab. C.49: Test č. 49

94	6	0	0	0
2	84	12	2	0
0	6	92	0	2
0	6	0	90	4
0	2	4	0	94

Testování - průměr: **90,80** %

Tab. C.50: Test č. 50

98	2	0	0	0
8	78	14	0	0
40	4	56	0	0
2	8	0	84	6
2	2	18	0	78

Testování - průměr: **78,80 %**

Tab. C.51: Test č. 51

100	0	0	0	0
82	18	0	0	0
96	4	0	0	0
56	6	0	36	2
68	2	2	0	28

Testování - průměr: **36,40 %**

Tab. C.52: Test č. 52

100	0	0	0	0
100	0	0	0	0
100	0	0	0	0
100	0	0	0	0
100	0	0	0	0

Testování - průměr: **20,00 %**

Tab. C.53: Test č. 53

94	6	0	0	0
4	90	4	2	0
0	6	92	0	2
0	6	0	90	4
0	0	4	0	96

Testování - průměr: **92,40 %**

Tab. C.54: Test č. 54

94	6	0	0	0
2	86	10	2	0
0	6	92	0	2
0	6	0	88	6
0	2	4	0	94

Testování - průměr: **90,80 %**

Tab. C.55: Test č. 55

94	4	2	0	0
2	82	14	2	0
6	6	88	0	0
2	16	6	72	4
0	4	12	0	84

Testování - průměr: **84,00 %**

Tab. C.56: Test č. 56

98	0	2	0	0
40	50	8	2	0
64	4	32	0	0
30	26	2	38	4
32	0	6	0	62

Testování - průměr: **56,00 %**

Tab. C.57: Test č. 57

100	0	0	0	0
100	0	0	0	0
100	0	0	0	0
94	0	0	6	0
98	0	0	0	2

Testování - průměr: **21,60 %**

Tab. C.58: Test č. 58

94	6	0	0	0
4	88	6	2	0
0	6	92	0	2
0	6	0	90	4
0	0	4	0	96

Testování - průměr: **92,00 %**

Tab. C.59: Test č. 59

94	6	0	0	0
2	90	6	2	0
0	6	92	0	2
0	8	2	84	6
0	4	4	0	92

Testování - průměr: **90,40 %**

Tab. C.60: Test č. 60

94	4	2	0	0
2	84	14	0	0
0	6	94	0	0
4	30	10	50	6
2	4	10	0	84

Testování - průměr: **81,20 %**

Tab. C.61: Test č. 61

98	0	2	0	0
26	58	16	0	0
50	4	42	0	4
24	36	12	24	4
14	0	12	0	74

Testování - průměr: **59,20 %**

Tab. C.62: Test č. 62

100	0	0	0	0
90	8	2	0	0
96	2	2	0	0
74	12	6	6	2
72	0	0	0	28

Testování - průměr: **28,80 %**