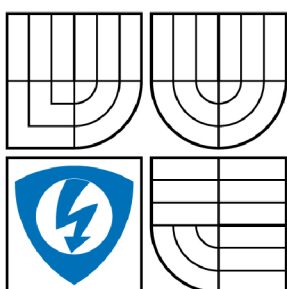


VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH
TECHNOLOGIÍ

ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

POROVNÁNÍ ANALÝZY ŘEČOVÉHO SIGNÁLU V ZÁVISLOSTI NA VĚKU A POHLAVÍ MLUVČÍHO

COMPARISON OF ANALYSIS OF SPEECH IN DEPENDENCE ON AGE AND GENDER

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

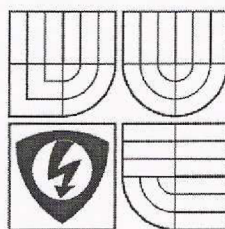
JOSEF BÁŇA

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. HICHAM ATASSI

BRNO 2008



VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

Ústav telekomunikací

Bakalářská práce

bakalářský studijní obor
Teleinformatika

Student: Báňa Josef
Ročník: 3

ID: 83877
Akademický rok: 2007/08

NÁZEV TÉMATU:

Porovnání analýzy řečového signálu v závislosti na věku a pohlaví mluvčího

POKYNY PRO VYPRACOVÁNÍ:

Prostudujte základní vlastnosti řečového signálu. Zaměřte se na vybrané příznaky, například základní tón, formanty, tempo řeči atd. Vytvořte vhodnou databázi řečových nahrávek od mluvčích různého pohlaví a různého věku. Analýzu řečových nahrávek je možné provést pomocí volně dostupných řečových analyzátorů. Dále navrhnete algoritmus, který je schopen automaticky nebo poloautomaticky odhadnout věk a pohlaví mluvčího.

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.

Termín zadání: 11.2.2008

Termín odevzdání: 4.6.2008

Vedoucí projektu: Ing. Hicham Atassi


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



UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

LICENČNÍ SMLOUVA

POSKYTOVANÁ K VÝKONU PRÁVA UŽÍT ŠKOLNÍ DÍLO

uzavřená mezi smluvními stranami:

1. Pan/paní

Jméno a příjmení: Josef Báňa
Bytem: Masarykova 21, 69172, Klobouky u Brna
Narozen/a (datum a místo): 25.10.1984, Hustopeče

(dále jen "autor")

a

2. Vysoké učení technické v Brně

Fakulta elektrotechniky a komunikačních technologií
se sídlem Údolní 244/53, 60200 Brno 2
jejímž jménem jedná na základě písemného pověření děkanem fakulty:
prof. Ing. Kamil Vrba, CSc.

(dále jen "nabyvatel")

Článek 1

Specifikace školního díla

1. Předmětem této smlouvy je vysokoškolská kvalifikační práce (VŠKP):

- disertační práce
- diplomová práce
- bakalářská práce

jiná práce, jejíž druh je specifikován jako

(dále jen VŠKP nebo dílo)

Název VŠKP: Porovnání analýzy řečového signálu v závislosti na věku a pohlaví mluvčího

Vedoucí/školitel VŠKP: Ing. Hicham Atassi

Ústav: Ústav telekomunikací

Datum obhajoby VŠKP:

VŠKP odevzdal autor nabyvateli v:

- tištěné formě - počet exemplářů 1
- elektronické formě - počet exemplářů 1

2. Autor prohlašuje, že vytvořil samostatnou vlastní tvůrčí činností dílo shora popsané a specifikované. Autor dále prohlašuje, že při zpracovávání díla se sám nedostal do rozporu s autorským zákonem a předpisy souvisejícími a že je dílo dílem původním.

3. Dílo je chráněno jako dílo dle autorského zákona v platném znění.

4. Autor potvrzuje, že listinná a elektronická verze díla je identická.

Článek 2

Udělení licenčního oprávnění

1. Autor touto smlouvou poskytuje nabyvateli oprávnění (licenci) k výkonu práva uvedené dílo nevýdělečně užit, archivovat a zpřístupnit ke studijním, výukovým a výzkumným účelům včetně pořizování výpisů, opisů a rozmnoženin.
2. Licence je poskytována celosvětově, pro celou dobu trvání autorských a majetkových práv k dílu.
3. Autor souhlasí se zveřejněním díla v databázi přístupné v mezinárodní síti
 - ihned po uzavření této smlouvy
 - 1 rok po uzavření této smlouvy
 - 3 roky po uzavření této smlouvy
 - 5 let po uzavření této smlouvy
 - 10 let po uzavření této smlouvy(z důvodu utajení v něm obsažených informací)
4. Nevýdělečné zveřejňování díla nabyvatelem v souladu s ustanovením § 47b zákona č. 111/1998 Sb., v platném znění, nevyžaduje licenci a nabyvatel je k němu povinen a oprávněn ze zákona.

Článek 3

Závěrečná ustanovení

1. Smlouva je sepsána ve třech vyhotoveních s platností originálu, přičemž po jednom vyhotovení obdrží autor a nabyvatel, další vyhotovení je vloženo do VŠKP.
2. Vztahy mezi smluvními stranami vzniklé a neupravené touto smlouvou se řídí autorským zákonem, občanským zákoníkem, vysokoškolským zákonem, zákonem o archivnictví, v platném znění a popř. dalšími právními předpisy.
3. Licenční smlouva byla uzavřena na základě svobodné a pravé vůle smluvních stran, s plným porozuměním jejímu textu i důsledkům, nikoliv v tísní a za nápadně nevýhodných podmínek.
4. Licenční smlouva nabývá platnosti a účinnosti dnem jejího podpisu oběma smluvními stranami.

V Brně dne:

.....

Nabyvatel

.....

Autor

ANOTACE

Obsah této práce je zaměřen na analýzu věku a pohlaví. Dále pak byli zkoumány vhodnosti příznaků a vhodnost jejich použití při klasifikaci mluvčích. Je v ní pojednáno o základní teorii řečového signálu a problematice zobrazení a zpracování řeči.

Je zde popsán volně dostupný program Praat, kterým se používá pro řečovou analýzu. Zaměřili jsme se na suprasegmentální příznaky řeči. Dále je zde popsána teorie příznaků které byly pro práci zvoleny.

Prvním z kroků v této práci bylo získání dostatečného počtu promluv (nahráváním) od mluvčích různého věku a pohlaví. Řečový korpus je poměrně obsáhlý. Mluvčí byli rozděleni do šesti věkových skupin. Dále byly nahrávky zpracovány v programu Praat. Průměrné hodnoty byly vyneseny do tabulek a sloupcových grafů pro větší přehlednost.

Dále bylo vybráno dvanáct nejvhodnějších příznaků podle kritéria kvality. S nimi poté byla provedena analýza zvolených příznaků pro automatické rozpoznávání pohlaví a věku. Příznaky byly při testování po jednom ubírány.

Jako klasifikátor pro rozpoznávání byla zvolena neuronová síť. Pro práci se sítí byl zvolen Neural Network Toolbox v programu Matlab. Vytvořené sítě byly použity pro klasifikaci mluvčích na základě věku a pohlaví. Výsledky byly rozebrány v závěru práce.

Klíčová slova: Řeč, korpus, analýza, příznaky, klasifikace, mluvčí, věk, pohlaví, Praat, neuronová síť.

ABSTRACT

This thesis deals with analysis of speech signal in dependence on the gender and the age of the speaker. We tried to investigate through the features to find the best set for the automatic classification of speakers. It also contains a brief discussion about the speech signal and its characteristics.

We used a program called Praat for the speech analysis purpose. This program is also described in this work. We mainly focused on the suprasegmental features of speech.

Our first step was to make our own speech corpus which should contain speech records from speakers with various age and gender. We made the analysis using Praat and reported it within this thesis.

For the automatic classification purpose, twelve features were selected basing on there quality criteria and used with a neural network to classify the speakers to classes with different age and gender.

As it was mentioned, a neural network was used as a classifier. We used "Neural Network Toolbox" in the Matlab program to create and train our network.

Keywords: Speech, corpus, analysis, features, classification, speaker, age, gender, Praat, neural network.

BÁŇA, J. Porovnání analýzy řečového signálu v závislosti na věku a pohlaví mluvčího. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2008. 67 s. Vedoucí bakalářské práce Ing. Hicham Atassi.

PROHLÁŠENÍ

Prohlašuji, že svou bakalářskou práci na téma „Porovnání analýzy řečového signálu v závislosti na věku a pohlaví mluvčího“ 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 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

Hluboce děkuji vedoucímu bakalářské práce Ing. Hichamu Atassimu, za důkladné a věcné konzultace, rady při zpracování bakalářské práce a trpělivý přístup. Děkuji svým rodičům, že mi umožnili studium a neustále mě při něm podporovali. V neposlední řadě děkuji všem mluvčím s kterými jsem nahrával za ochotu a pomoc získávání nahrávek pro moji práci. Také chci poděkovat vedení škol na kterých mi bylo nahrávání zprostředkováno. Děkuji všem lidem za rady a opravy při zpracování této práce.

V Brně dne

.....

(podpis autora)

Význam použité symboliky

T	[s]	celkový čas nahrávky
R	[slabik/s]	tempo řeči (n = počet)
Z_n	[-]	Počet znělých úseků
F_{0min}	[Hz]	minimální hodnota základního tónu
F_{0max}	[Hz]	maximální hodnota základního tónu
F_{0str}	[Hz]	střední hodnota základního tónu
F_{0std}	[Hz]	standardní výchylka základního tónu
F_{0sl}	[Půltóny/s]	strmost (slope without octave jumps)
H_{min}	[dB]	minimální hodnota harmonicity
H_{max}	[dB]	maximální hodnota harmonicity
H_{str}	[dB]	střední hodnota harmonicity
H_{std}	[dB]	standardní výchylka harmonicity
I_{min}	[dB]	minimální hodnota intenzity
I_{max}	[dB]	maximální hodnota intenzity
I_{str}	[dB]	střední hodnota intenzity
I_{std}	[dB]	standardní výchylka intenzity
F_1	[Hz]	střední frekvence 1. formantu
F_{1sp}	[Hz]	šířka pásma 1. formantu
F_2	[Hz]	střední frekvence 2. formantu
F_{2sp}	[Hz]	šířka pásma 2. formantu
F_3	[Hz]	střední frekvence 3. formantu
F_{3sp}	[Hz]	šířka pásma 3. formantu
F_4	[Hz]	střední frekvence 4. formantu
F_{4sp}	[Hz]	šířka pásma 4. formantu
x_i		i -tý příznak
N		počet
N_T		počet tříd
S		směrodatná odchylka
X_{STR}		střední hodnota veličiny X
VS1		první věková skupina, věk mluvčích 7-12 let
VS2		druhá věková skupina, věk mluvčích 13-15 let
VS3		třetí věková skupina, věk mluvčích 16-20 let
VS4		čtvrtá věková skupina, věk mluvčích 21-30 let
VS5		pátá věková skupina, věk mluvčích 31-50 let
VS6		šestá věková skupina, věk mluvčích nad 50 let

Obsah

1	Úvod.....	13
2	Řečový signál.....	14
2.1	Řeč a její vývoj.....	14
2.1.1	Vývoj řeči u dítěte	14
2.1.2	Informační obsah sdělení	15
2.2	Vznik řeči v lidském těle.....	15
2.3	Historie zkoumání a napodobování řeči.....	16
2.4	Zpracování řeči	17
2.4.1	Oblasti zpracování řečových signálů	18
2.5	Znázornění řečových signálů	19
2.5.1	Zobrazení v časové oblasti	19
2.5.2	: Zobrazení ve frekvenční oblasti	19
2.5.2.1	Spektrogram.....	20
3	Praat	21
3.1	O programu.....	21
3.2	Vzhled programu	21
3.3	Autoři	23
4	Řečový korpus	24
4.1	Technické parametry nahrávek.....	24
4.2	Rozčlenění na věkové skupiny	24
4.3	Získávání vzorků.....	25
4.4	Použité věty	25
4.5	Počet vzorků	26
4.6	Dotazník.....	27
5	Analyzované parametry	28
5.1	Základní tón	28
5.2	Harmonicita.....	29
5.3	Akustická intenzita	30
5.4	Formanty.....	31
6	Analýza	32
6.1	Tabulky hodnot	32
6.1.1	Výsledky analýzy pro ženy	32
6.1.2	Výsledky analýzy pro muže	34
6.2	Grafy vypočtených hodnot	36
7	Analýza použitých příznaků pro automatické rozpoznání věku a pohlaví ..	39
8	Výběr příznaků pro automatickou klasifikaci mluvčích	43
8.1	Výběr příznaků podle rozptylů	44
8.1.1	Příklady výpočtu	45
8.2	Korelace příznaků	50
9	Neuronové sítě.....	53
9.1	Výhody a nevýhody neuronových sítí	53
9.2	Formální neuron	54
9.3	Základní modely neuronových sítí	55
9.4	Parametry navržené neuronové sítě.....	55
10	Návrh neuronových sítí v Neural Network Toolbox	57
11	Postup rozdělení vzorků pro trénování a testování	58
12	Proces klasifikace.....	59
13	Závěr	65
14	Použitá literatura	67

Seznam obrázků

Obr. 2.1: Model lidského hlasového ústrojí.....	16
Obr. 2.2: Kratzensteinův systém rezonátorů	16
Obr. 2.3: Hlavní oblasti zpracování řeči	18
Obr. 2.4: Časový průběh	19
Obr. 2.5: Spektrogram věty „Hana tam nebude“	20
Obr. 3.1: Základní okno programu	21
Obr. 3.2: Editační okno programu	22
Obr. 3.3: Okno pro vykreslování a export obrázků	22
Obr. 5.1: Zobrazení základního tónu.....	28
Obr. 5.2: Průběh Harmonicity.....	29
Obr. 5.3: Průběh Intenzity	30
Obr. 5.4: Formanty	31
Obr. 6.1: Závislost hodnot různých příznaků na věkové skupině:	38
Obr. 6.2: Závislost hodnot různých příznaků na věkové skupině:	37
Obr. 6.3: Závislost hodnot různých příznaků na věkové skupině:	37
Obr. 6.4: Závislost střední hodnoty harmonicity na věkové skupině.....	38
Obr. 8.1: Princip selekce příznaků.....	43
Obr. 8.2: Kritérium kvality pro rozlišení pohlaví.....	48
Obr. 8.3: Kritérium kvality pro rozlišení věku.....	48
Obr. 8.4: Kritérium kvality pro rozlišení věku jen pro ženy.....	49
Obr. 8.5: Kritérium kvality pro rozlišení věku jen pro muže.....	49
Obr. 8.6: Míra korelace příznaku k_i s ostatními příznaky.....	52
Obr. 9.1: Základní struktura modelu umělého neuronu.....	54
Obr. 9.2: Blokové schéma jedné z vytvořených sítí s deseti příznaky.....	56
Obr. 10.1: Základní okno Neural Network Toolboxu.....	57
Obr. 12.1: Výsledky správného zařazení do tříd pro všechny věkové skupiny.....	60
Obr. 12.2: Výsledky správného zařazení do tříd s vynecháním prvních dvou věkových skupin.....	61

Seznam tabulek

Tabulka 4.1: Počet vytvořených nahrávek v jednotlivých věkových skupinách	26
Tabulka 4.2: Udávající počet mluvčích a zároveň počet analyzovaných vzorků v jednotlivých věkových skupinách.....	26
Tabulka 4.3: Použitý dotazník pro mluvčí.....	27
Tabulka 6.1: Průměry analyzovaných parametrů v dílčích věkových skupinách pro ženy	32
Tabulka 6.2: Směrodatné odchytky analyzovaných parametrů v dílčích věkových skupinách pro ženy	33
Tabulka 6.3: Průměry analyzovaných parametrů v dílčích věkových skupinách pro muže	34
Tabulka 6.4: Směrodatné odchytky analyzovaných parametrů v dílčích věkových skupinách pro ženy	35
Tabulka 7.1: Aritmetické průměry výsledků příznaků pro VS1 a VS2.	40
Tabulka 7.2: Aritmetické průměry výsledků příznaků pro VS3 a VS4.	41
Tabulka 7.3: Aritmetické průměry výsledků příznaků pro VS5 a VS6.	42
Tabulka 8.1: Hodnoty kritéria kvality příznaků.....	47
Tabulka 8.2: Částečné zobrazení matice korelačních koeficientů.....	51
Tabulka 8.3: Deset nejméně (vlevo) a deset nejvíce (vpravo) korelovaných příznaků.	51
Tabulka 8.4: Míra korelace příznaku k_i s ostatními příznaky.	52
Tabulka 11.1: Použité množství promluv pro trénování a testování NS.	58
Tabulka 12.1: Dvanáct nejvhodnějších příznaků vybraných podle Q	59
Tabulka 12.2: Výsledky správného zařazení do tříd pro všechny věkové skupiny (vlevo) a s vynecháním prvních dvou věkových skupin (vpravo).	60
Tabulka 12.3: Výsledky pro rozeznávání věku u mužů.	62
Tabulka 12.4: Výsledky pro rozeznávání věku u žen.	63
Tabulka 12.5: Střední hodnoty věkových skupin.....	64
Tabulka 12.6: Střední hodnoty a směrodatné odchytky vzdálenosti správné klasifikace.....	64

1 Úvod

Snahou této práce je zjištění některých vlivů (jako je například rýma, nachlazení, ochraptělost, frekventovanost konzumace alkoholu nebo zda-li je mluvčí kuřák) na řečové ústrojí, řečové parametry a obecně řeč mluvčího. Dále je snahou projektu sledování vlivu vzdělání, vlivu oboru vzdělání a zaměstnání či prostředí ve kterém se mluvčí převážně pohybuje. Tudiž i slovní vlivu zásoby na způsob vyjadřování a řeč dílčích jedinců.

Stěžejním cílem této práce je porovnání nahrávek stejných vět od různých mluvčích a snaha o zjištění směrodatného příznaku kterým by šla co nejjednoznačněji určit věková skupina a pohlaví mluvčího.

Poznání příznaků a směrodatných parametrů ovlivňujících řeč mluvčích je velmi důležité pro samotné využití v konkrétních aplikacích, jako jsou například detekce požití alkoholu nebo psychický stav jedince pomocí řeči.

Součástí práce je databáze nahrávek pro různé věkové skupiny a pohlaví.

Pro analýzu nahrávek byl použit program Praat. O tomto programu je níže samostatná kapitola.

Dále je využito zkoumaných příznaků pro klasifikaci mluvčích podle pohlaví a poté podle věku mluvčího. Bylo vybráno dvanáct nejvhodnějších příznaků podle kritéria kvality. S nimi poté byla provedena analýza zvolených příznaků pro automatické rozpoznávání pohlaví a věku. Příznaky byly při testování po jednom ubírány.

Jako klasifikátor pro rozpoznávání byla zvolena neuronová síť. Pro práci se sítí byl zvolen Neural Network Toolbox v programu Matlab. Vytvořené sítě byly použity pro klasifikaci mluvčích na základě věku a pohlaví.

2 Řečový signál

2.1 Řeč a její vývoj

Řečový signál vytváří každý z nás a setkáváme se s ním mezi lidmi denně. Řeč je totiž to co nás lidi odlišuje od zvířat. Je to zatím nejvyspělejší komunikační prostředek. Mnoho věcí si ohledně řeči neuvědomujeme, protože ji přirozeným způsobem vnímáme jako naši nedílnou součást. Řeč patří mezi jednu z lidských schopností, ale musí se náležitě rozvinout a „vycvičit“.

2.1.1 Vývoj řeči u dítěte

Rodiče na dítě promlouvají od narození a přibližně mezi rokem a půl a druhým rokem života se jeho biologické a psychické vlastnosti vyvinou natolik, že začne částečně mluvit. Podle vývojové psychologie se nejvíce člověk vyvíjí do 6. roku života. Mezi vlastnosti v tomto vývoji patří i řeč. Pokud se člověk nenaučí aspoň částečně mluvit a nebude kolem sebe vnímat řečové projevy do 7. roku života, již se nemůže plnohodnotně naučit mluvit.

Vývoj řeči začíná od našeho počátku. Od 6 týdne po narození si dítě začne broukat. [3] Poté začne produkovat některé samohlásky a po přidání některých zvuků se začnou objevovat i některé souhlásky. Kolem 2. měsíce začne napodobovat zvuky ve svém okolí – toto se nazývá stádiem žvatlání.

Jedním z prvních prvků, které dítě napodobí podle řeči svého okolí a který si rychle osvojí je melodie řeči. První slova bývají jednoslabičná nebo dvouslabičná a zpočátku jsou tvořena opakováním stejných slabik (mama, baba, pipi, ham apod.). Tedy, dítě nejdříve opakuje jednoduché zvuky, později se učí opakovat jednotlivé slabiky a slova a konečně celé věty a složitější větné výrazy.

Mezi prvním a druhým rokem života se slovní zásoba názvů osob, zvířat a věcí i činností výrazně zvětší. Mezi druhým a třetím rokem života má dítě nejzákladnější slovník výrazů vytvořen, je schopno se dorozumět a začíná své řeči používat podle vžitých gramatických principů. Toto období je ve vývoji dětské řeči nejzajímavější a nejdůležitější. Mezi třetím a čtvrtým rokem má již

dítě snahu o samostatné vyjadřování. V této době také prochází řeč dítěte kritickým údobím. Dítě zná již velké množství slov, ale celou řadu ještě nedovede vyslovit tak, jak by si přálo. Jeho rychlost vyjadřování se tak opožďuje za rychlostí přílivu jeho myšlenek. Dítě pak někdy začíná zadržávat. Tento jev ale ještě neznamena, že jde o koktavost - poruchu řeči, kterou je třeba léčit. Kolem pátého roku by měl být zpravidla vývoj řeči ukončen. Dítě je již schopno užívat hovorovou řeč, vyprávět jednoduchý děj. [3]

2.1.2 Informační obsah sdělení

Řečový signál nese jen vyslovenou zprávu (co nám chce někdo sdělit), ale i identitu mluvčího (kdo mluví) a stav mluvčího (jak mu je).

Obsah sdělení tvoří přibližně ze 75% vyslovená zpráva, z 15% identita mluvčího a z 10% stav mluvčího z celkové informace v řečovém signálu. [1]

Každý tvor vnímá zvuky. Ale řeč vnímá mnoho zvířat jinak než člověk. Například psi podle výzkumů vnímají pouze první jednu až dvě slabiky.

Řečový signál nese tyto informace:

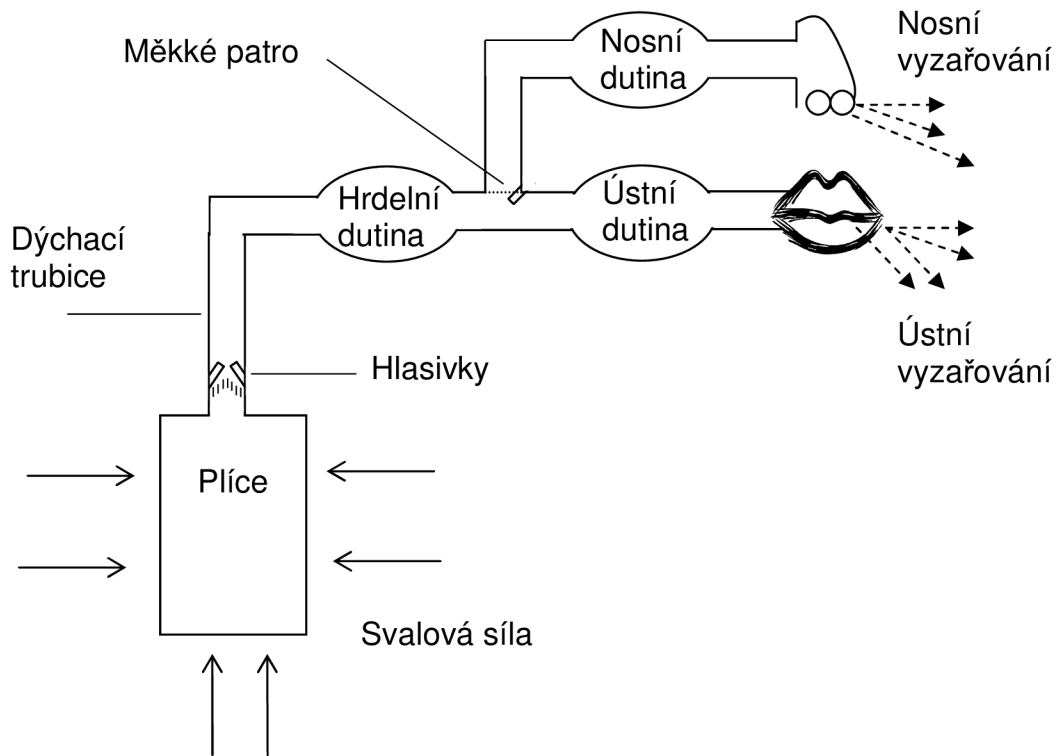
- vlastní vyslovená zpráva (skládá se ze slov)
- identitu mluvčího (mluvčí je rozpoznáván podle individuálního stylu promluvy)
- druh sdělení (sdělení je podle toho zesíleno nebo zeslabeno)
- nálada mluvčího (odráží se ve stylu promluvy)
- původ mluvčího (jazyk, dialekt, přízvuk)
- charakteristiky pozadí a prostředí, ve kterém je mluveno

2.2 Vznik řeči v lidském těle

V lidském těle frekvenci kmitání hlasivek určuje **základní hlasivkový tón**, který se pohybuje zhruba v rozmezí 100 - 400 Hz. [5]

Vytvářený zvuk se šíří dutinami hlasivkového traktu, kde dochází v závislosti od nastavení těchto dutin k rezonancím zesilujícím určité části generovaného zvukového spektra. Tyto části se ve fonetice nazývají **formanty**.

Zvuk, který (z výjimkou některých fonémů, např. sykavky s, š) vzniká v hrtanu, je modifikován v *rezonančních dutinách* - hrtanové, ústní a nosohltanové. [1]



Obr. 2.1: Model lidského hlasového ústrojí.

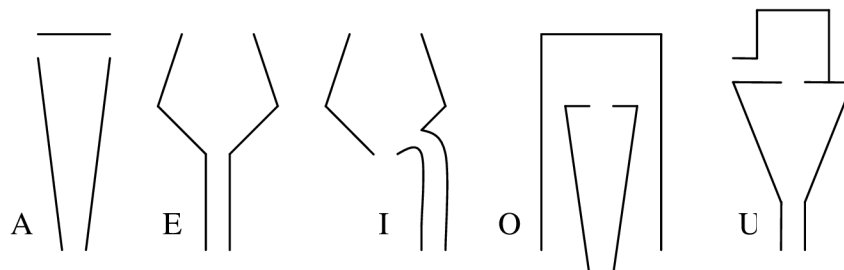
2.3 Historie zkoumání a napodobování řeči

V průběhu dějin se lidé snažili uměle vytvářet řečová ústrojí pomocí mechanických konstrukcí a poté i elektronických zařízení. [4]

-Řeč - australopithecus zhruba před třemi miliony let

-Ve starověku - budování mluvících soch a bůžků

-Kratzenstein - systému rezonátorů, které napodobovaly samohlásky a, e, i, o, u (1779).



Obr. 2.2: Kratzensteinův systém rezonátorů.

- von Kempelen – mluvicí stroj, který se dá označit za první řečový (mechanický) syntetizér (1791).
- devatenácté století – J. B. Fourier a H. Helmholtz, J. R. Ewald.
- 1924 –spektrální analýza řeči na bázi formantové analýzy samohlásek.
- Wheatstone – Pomocí rezonátoru a měchu (1922)
- Stewart
- Riesz (1937) – vytvoření tvarované cisterny která měla nahrazovat ústa a modulace přiváděného vzduchu se prováděla pomocí záklopek a škrťících ventilů.
- Voder (1939) – elektronický model vytvářející umělou řeč, který byl ovládaný “Vonderovou klávesnicí” a pedálem.

2.4 Zpracování řeči

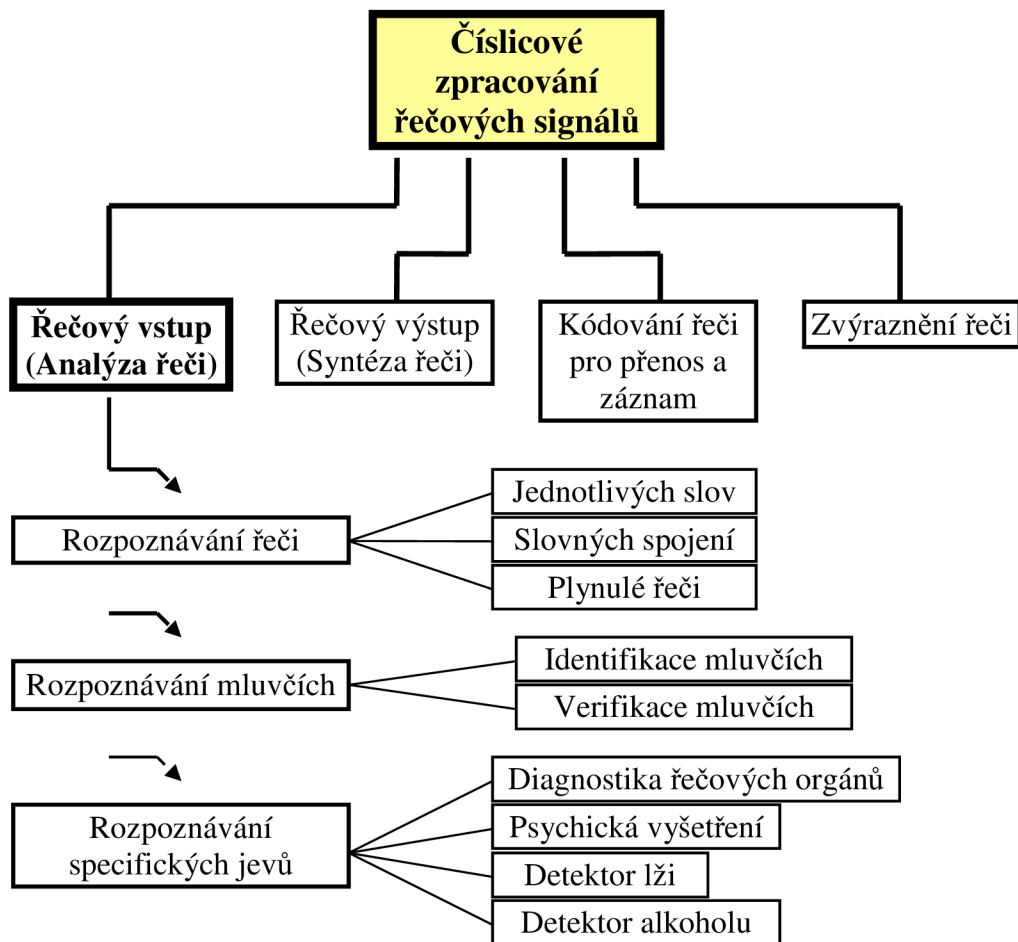
Zvukovou rovinou jazyka se zabývají dvě jazykovědné disciplíny. **Fonetika** zkoumající materiální stránku jazyka na nezobecněné úrovni i na prvním stupni zobecnění a **fonologie**, nauka o způsobech využívání zvukového materiálu v jazyce. Zatímco fonetika popisuje všechny zvuky, které se v řeči užívají, fonologie se zabývá jen těmi, které slouží rozlišování významu, tj. těmi, které mají distinktivní funkci.

Od konce 20. let 20. století fonologie bývá chápána jako samostatná vědní disciplína. Její základní jednotkou je foném, nejmenší zvukový prostředek schopný rozlišit minimální významovou jednotku jazyka, morfém.

Fonetika je jazykovědná disciplína zabývající se materiální stránkou zvukových výrazových prostředků jazyka, a to jak na nezobecněné úrovni poznání a popisu mechanismu vzniku řeči, její akustické stavby a percepce zvukového signálu sluchem, tak na zobecněné úrovni hlásek a prozodických prostředků, které jsou užívány v daném komunikačním společenství. [5]

2.4.1 Oblasti zpracování řečových signálů

Číslicové zpracování řečových signálů se dělí na mnoho odvětví. Podle historického vývoje bychom mohli vidět čtyři hlavní směry. Dnes nejvíce rozvinutou a zároveň historicky nejstarší oblastí je syntéza řeči (umělé vytváření řeči). [1]



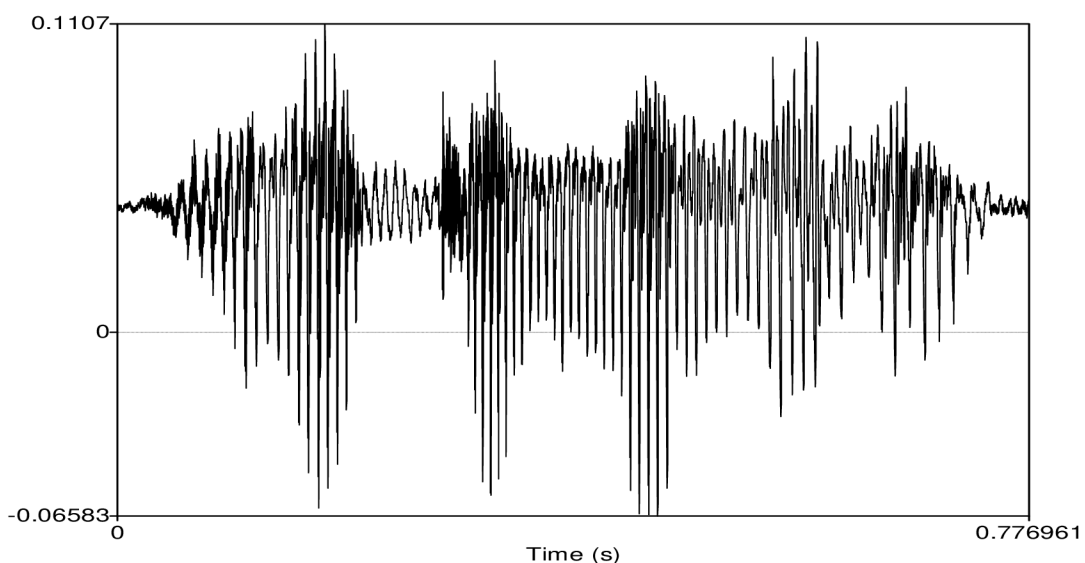
Obr. 2.3: Hlavní oblasti zpracování řeči.

2.5 Znázornění řečových signálů

Číslicový signál je pro nás někdy výhodnější v grafické podobě než jen jako vyjádření v číslech. Grafické zpracování signálů nám pomáhá při subjektivním posouzení. [1]

2.5.1 Zobrazení v časové oblasti

Mezi základní typy zobrazení řadíme zobrazení v časové oblasti. U jednoduchých průběhů se dá z tohoto zobrazení určit například délka periody, velikost úrovně signálu nebo zda-li se jedná o periodickou část některého úseku průběhu. Samotný časový průběh řeči však není vhodný pro zobrazení různých vlastností mluvené řeči, které jsou důležité pro kvalitu řeči a vnímání fonetických detailů [1]. Proto častěji používáme jiná zobrazení.



Obr. 2.4 Časový průběh.

2.5.2 : Zobrazení ve frekvenční oblasti

V technické praxi je zobrazení ve frekvenční oblasti mnohem používanější. Lze z průběhů více rozpoznat. Kmitočtové spektrum nám vznikne Furierovou transformací signálu. Podstatně jednodušší pro zkoumání vlastností akustických signálů je volit jako časovou - osu horizontální a jako kmitočtovou - osu vertikální.

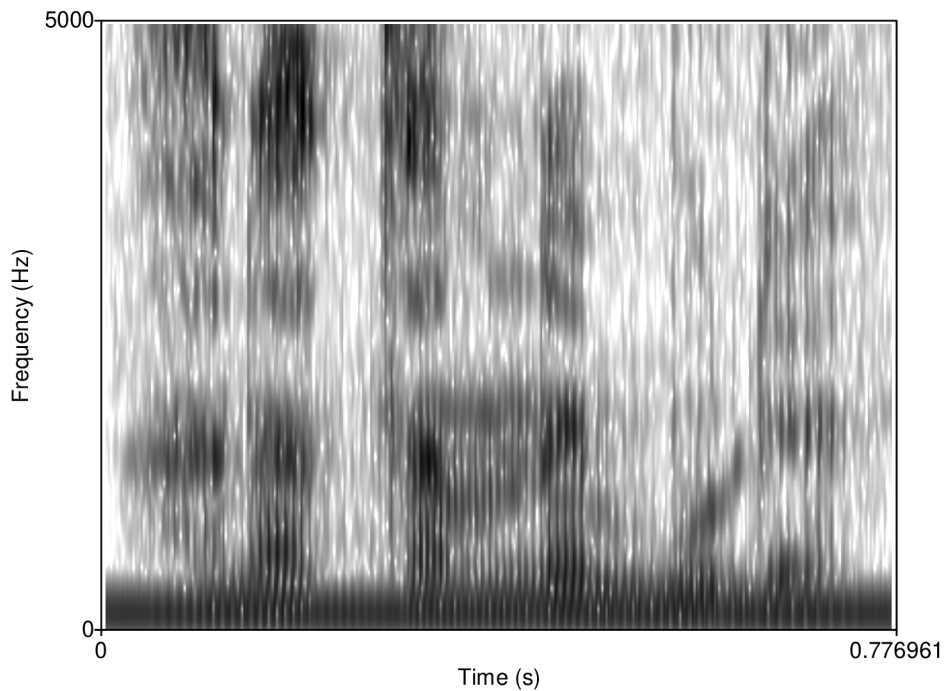
2.5.2.1 Spektrogram

Krátké spektrální intenzity kmitočtu zobrazujeme různou optickou intenzitou jedné barvy (stupně šedi). [1] Takovýto průběh nazýváme spektrogram.

Nejužívanější jsou 2 typy spektrogramů:

- úzkopásmový (šířka pásma 50Hz)
- širokopásmový (šířka pásma 300Hz)

V době, kdy byla používána výhradně analogová technika se u signálů porovnáváním spektrogramů provádělo ověřování mluvčích (často pro policejní účely speciálně vyškolenými odborníky). Níže je názorná ukázka spektrogramu.



Obr. 2.5: Spektrogram věty „Hana tam nebude“.

3 Praat

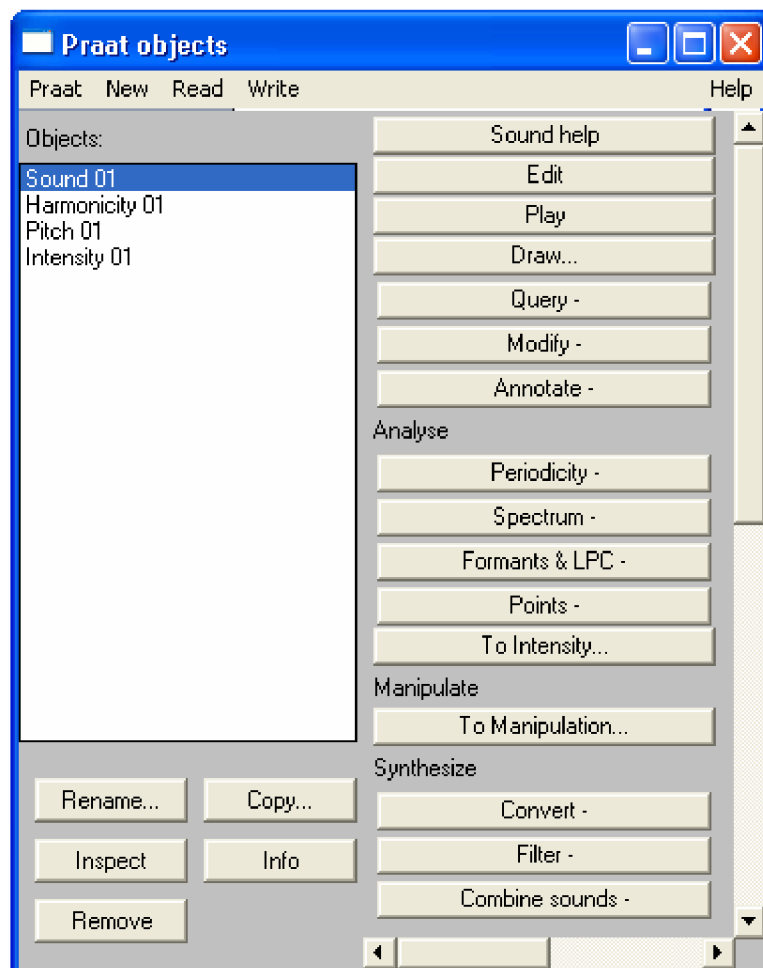
3.1 O programu

Program Praat dělá fonetickou transkripci na počítači. Je multiplatformní program pro fonetickou transkripci na počítači napsaný v jazyce C. [2]

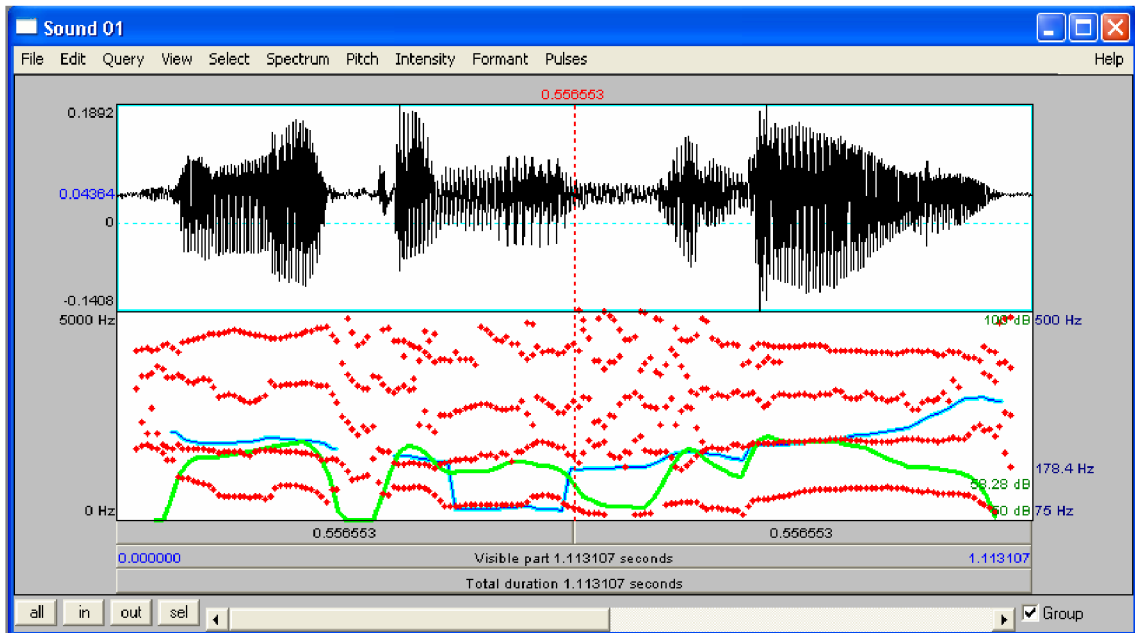
Existuje v různých verzích pro tyto operační systémy:

- Macintosh.
- Windows
- Linux, FreeBSD
- SGI, Solaris, HPUX (pro tyto platformy již nejsou vytvářeny nové verze)

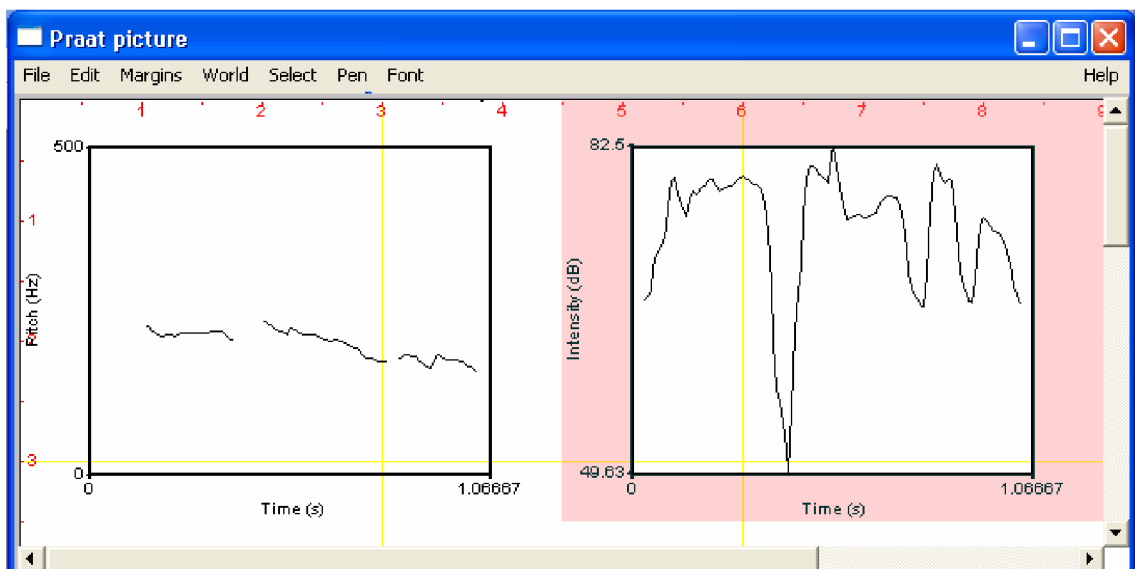
3.2 Vzhled programu



Obr. 3.1: Základní okno programu.



Obr. 3.2: Editační okno programu.



Obr. 3.3: Okno pro vykreslování a export obrázků.

Program byl použit pro výpočty střední hodnoty, minima, maxima a směrodatné odchylky pro tyto příznaky:

- základní tón
- intenzitu
- harmonicitu.

Dále pro výpočet počtu znělých úseků, celkového trvání řečového signálu a střední hodnoty a šířku prvních 4 formantů.

Má mnoho dalších funkcí, například syntéza intenzit a LPC.

3.3 Autoři

Oba autoři Paul Boersma a David Weenink pracují v Institutu fonetické syntézy, který se nachází na Amsterdamské Univerzitě v Holandsku. [2]

David Weenink

Vyučuje:

- zpracování řeči na počítači – pro studenty bakalářského programu v humanitních oborech.
- zpracování řeči a syntéza – pro studenty bakalářského programu v oboru umělé inteligence.
- zpracování řečových signálů - pro studenty magisterského programu v humanitních oborech.
- adaptivní identifikace samohlásek a skriptování – pro studenty doktorského studia

Paul Boersma

Zabývá se problematikou vztahů mezi fonologií a fonetikou.

A také vyučuje tyto oblasti:

- Fonetická transkripce
- Řeč
- Řečové vjemy
- Experimentální fonetika
- Obousměrná fonologie a fonetika

4 Řečový korpus

4.1 Technické parametry nahrávek

Nahrávky jsou všechny monofoní.

Vzorkovací kmitočet: 44100 Hz.

Počet bitů na jeden vzorek:16.

Zvukový formát nahrávek je PCM wave file (*.wav).

Pro záznam nahrávek byl použitý software GoldWave.

Použité záznamové zařízení: Počítačový mikrofon s elektretovou vložkou.

4.2 Rozčlenění na věkové skupiny

Rozdělení do věkových skupin jsem formuloval na základě konzultace s psychologem, který mi uvedl rozdělení věkových skupin podle celkového vývoje člověka. Jedinci v každé věkové skupině jsou v jiném stádiu biologického i psychického vývoje.

Členění skupin podle vývoje vypadá takto:

Předškolní věk 0-7 let.

Školní věk: 7-12 let.

Pubescence: 12-14 let.

Adolescence: 15-20 let.

Mladá dospělost: 20-30 let.

Střední dospělost: 30-45 (50) let.

Pozdní dospělost: 50-60 let.

Stáří: nad 60 let.

Poté jsem výběr věkových skupin uzpůsobil technické realizaci. Například kojenci ještě nemluví a dítě se do 7. roku života ještě vyvíjí, takže můžeme věkovou skupinu 0-7 let vypustit, protože pro nás statisticky nebude směrodatná. S velmi starými lidmi by se komplikovaně získávali nahrávky, takže jsem ji také vypustil.

Aby se daly vzorky mezi sebou porovnávat, je zapotřebí mít v každé věkové skupině dost mluvčích, minimálně deset mluvčích v každé věkové skupině. To s sebou nese velkou časovou náročnost a vzhledem k tomu, že jsem neměl k dispozici mobilní zařízení pro nahrávání, musel jsem použít stolních PC, které mi byli zapůjčeny v dané lokalitě. Po získání dostatečného počtu nahrávek jsem poupravil hranice skupin s přihlédnutím na změny řeči vlivem věku.

Výsledné rozdělení věkových skupin vypadá takto:

1. 7 – 12 let.
2. 13 – 15 let.
3. 16 – 20 let.
4. 21 – 30 let.
5. 30 – 50 let.
6. nad 50 let.

4.3 Získávání vzorků

Vzhledem k tomu že lidé ve věku 7-30 let jsou nejčastěji žáci a studenti, rozhodl jsem se oslovit školy. Tímto způsobem jsem si nepatrně usnadnil práci, protože jsem mohl dotyčné přesvědčit ke spolupráci hromadně. Bylo však zapotřebí mnoho věcí vysvětlovat individuálně a každého podrobně instruovat zvlášť. Nahrávání byla provedena na dvou základních školách, jedné středně škole, na vysokoškolských kolejích a také doma v rodině.

4.4 Použité věty

1. Hana tam nebude.
2. Vlák už pojede.
3. Včera jsem přišel pozdě.
4. Operace se jim nepovedla.
5. Tak dlouho se chodí se džbánem pro vodu, až se ucho utrhne.
6. Kdo chce, hledá způsoby, kdo nechce, hledá důvody.

7. Pohovor (nezávislá nahrávka do délky 1 minuty záznamu).

4.5 Počet vzorků

Celkový počet vytvořených nahrávek je **504**. Z toho 72 Pohovorů, které nejsou do tabulek zahrnuty.

Tabulka 4.1: Počet vytvořených nahrávek v jednotlivých věkových skupinách

Číslo skupiny	Věk	Ženy	Muži	Celkem
1.	7 – 12	60	18	78
2.	13 – 15	18	42	60
3.	16 – 20	18	96	114
4.	21 – 30	42	48	90
5.	30 – 50	36	30	66
6.	nad 50	18	6	24
Σ		192	240	432

Zpracovány byli vzorky pro 1. větu „Hana tam nebude.“ Celkový počet analyzovaných vzorků které jsou použity v semestrálním projektu je 72.

Tabulka 4.2: Udávající počet mluvčích a zároveň počet analyzovaných vzorků v jednotlivých věkových skupinách

Číslo skupiny	Věk	Ženy	Muži	Celkem
1.	7 – 12	10	3	13
2.	13 – 15	3	7	10
3.	16 – 20	3	16	19
4.	21 – 30	7	8	15
5.	30 – 50	6	5	11
6.	nad 50	3	1	4
Σ		32	40	72

4.6 Dotazník

Při analýze řeči existují parametry, které mohou ovlivnit hlasové ústrojí a řečové projevy. Na základě zformovaného seznamu potenciálních parametrů jsem vytvořil pracovní dotazník, do kterého jsem si zaznamenával informace o jednotlivých mluvčích. Vzhled dotazníku je uveden v následující tabulce.

Tabulka 4.3: Použitý dotazník pro mluvčí

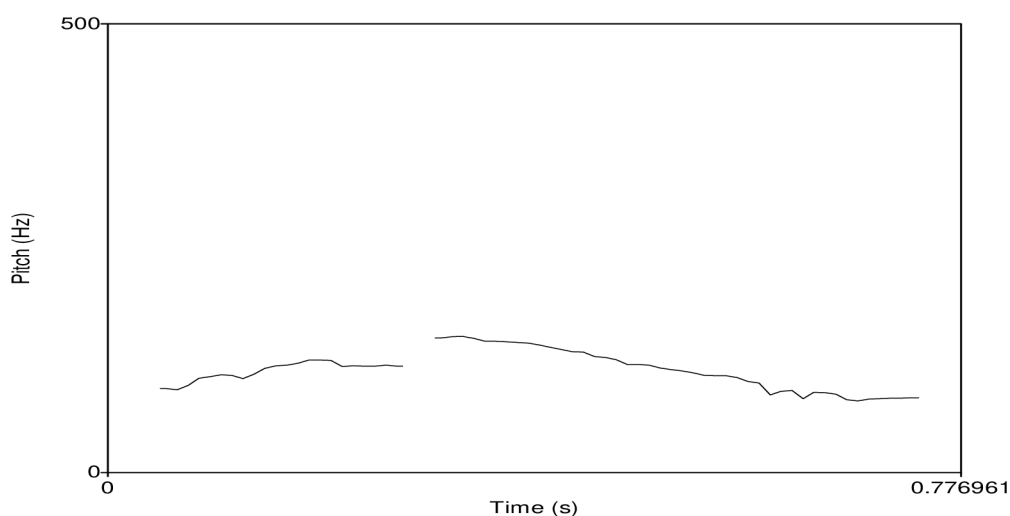
Dotazník číslo:	Pohlaví M / Ž	Jméno:	Datum:	Čas:	Národnost	
Věk:	7 -12	13 - 15	16 - 20	21 - 30	30-50	nad 50
Váha:			Výška:			
Odkud pocházíte						
Kde žijete (min 1 rok)						
Nejvyšší dosažené vzdělání	Základní __	SOU __	SŠ s mat. __	VOŠ __	VŠ __	Jiné __
Obor vzdělání:	Dosaženého:			Současného:		
Zaměstnání						
Kuřák	ANO __	NE __	Příležitost ně: __			
Jak často pijete alkohol	Abstinent __	1 za měsíc __	1 týdně __	2-3 x týdně __	Ob den __	Denně __
Cítíte se ospale	Ne __	Trochu __	Ano __			
Jak se cítíte? Je mi:	Skvěle __	Čile __	Mírně unaveně__	Unaveně __	Vyčerpaně __	
Máte nyní rýmu, nachlazení, rovnátka, piercing...						
Prostředí nahrávky						
Hlučnost prostředí						
Vizuální dojem pozorovatele	Únava:			Ospalost:		
Použité záznamové zařízení						

5 Analyzované parametry

5.1 Základní tón

Základní tón je zvuk, jehož zvukovou intenzitu lze v závislosti na čase popsat sinusoidou. Složený tón (nebo jen tón) je lineární kombinací základních tónů. Základní hlasivkový tón v lidském těle určuje frekvence kmitání hlasivek, která se pohybuje zhruba v rozmezí 100 - 400 Hz. [5]

Každý periodický signál je složen z několika harmonických složek, které lze vypočítat rozvojem do Fourierovy řady. Toto složení nazýváme spektrum. Spektrum u akustického signálu lze nazvat složený tón. Základní tón reprezentuje pravidelnost vzorku jako funkci v závislosti na čase. Tato pravidelnost nesouvisí s akustickým dojmem nebo chvěním hlasivek.



Obr. 5.1: Zobrazení základního tónu.

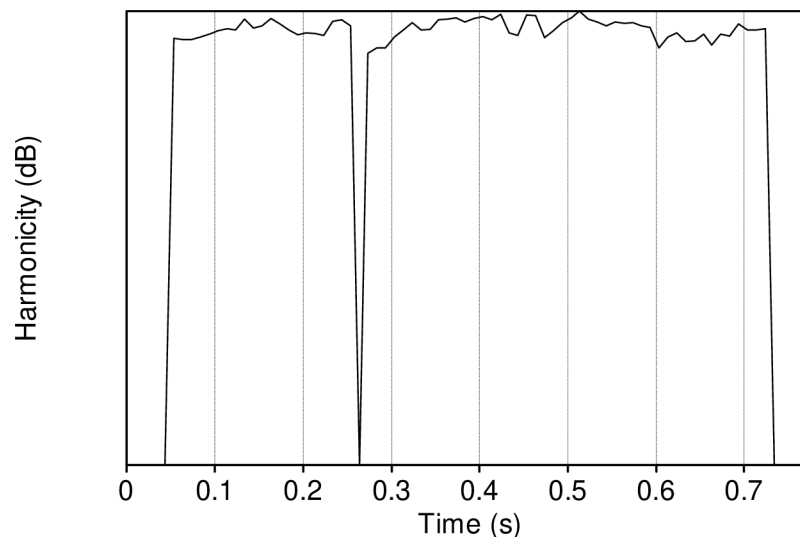
5.2 Harmonicita

Harmonicita reprezentuje stupeň akustické pravidelnosti, také pro ni bývá používána zkratka HRN (Harmonics - to - Noise Ratio).

Harmonicita je vyjádřena v dB: jestli 99% z energie signálu je v periodické části, a 1% je šum, HNR je $10 \cdot \log_{10}(99/1) = 20$ dB. HNR 0 dB znamená že se energie v harmonické části a v šumu rovnají.

Harmonicita může být užívána jako míra pro:

- Šumové procento z jakékoliv části periodického signálu.
- Hlasovou kvalitu. Například zdravý mluvčí může produkovat znělé [a] nebo [i] s harmoničností kolem 20 dB a [u] asi okolo 40 dB. Tento rozdíl pocházející z vysokých kmitočtů u [a] a [i], oproti nízkým kmitočtům u [u], má za následek velmi vyšší citlivost HNR pro vibrace v [a] a [i], než v [u]. Ochraptlí mluvčí bude mít u [a] harmoničnost mnohem nižší než 20 dB. Ví se o patologickém případě kde mluvčí měl HNR 40 dB pro [i], protože klesla frekvence jeho hlasu pod 2000 Hz. [2]



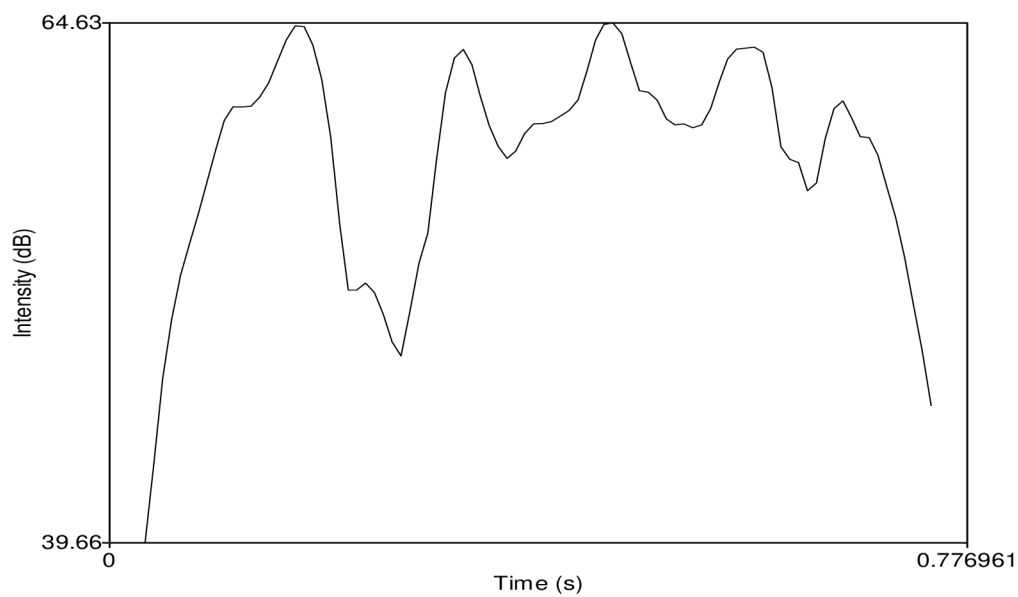
Obr. 5.2: Průběh Harmonicity

5.3 Akustická intenzita

Intenzita je fyzikální veličina vyjadřující množství akustické energie (energie kmitajících částic prostředí) které projde jednotkovou plochou za jednotku času.

Intenzita reprezentuje průběh síly v časovém rozložení $t_i = t_1 + (i - 1) dt$.

S hodnotami v dB SPL, t.j. dB vztažný pro $2 \cdot 10^{-5}$ Pa, což je normovaný sluchový práh pro sinusové vlny o $f = 1000\text{Hz}$. [2]



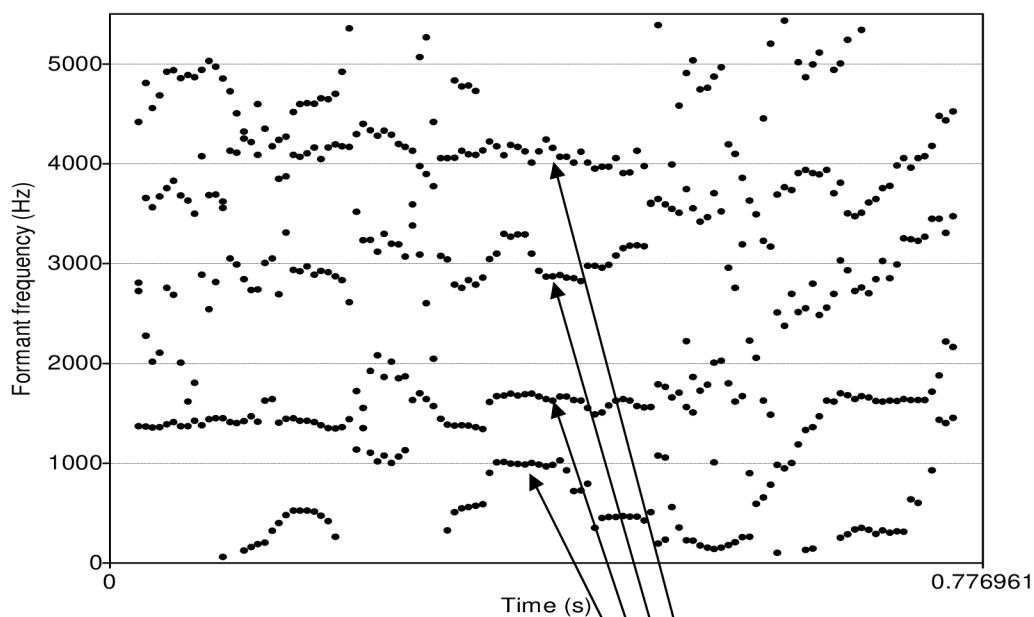
Obr. 5.3: Průběh Intenzity

5.4 Formanty

První tři formantové kmitočty obsahují důležité informace o charakteru samohlásek (vokálů) a znělých souhlásek. Jako samostatné příznaky při krátkodobém spektru lze použít kmitočty prvního formantu (tímto parametrem lze od sebe odlišit například fonémy „i“ a „a“), střední kmitočty prvních tří formantů (tímto parametrem lze rozlišit fonémy „i“ a „e“), velikost prvního formantu – vztaženo k 0 dB (u fonému „a“ nabývá vysokých hodnot), velikost prvního formantu – vztaženo k minimu (u fonému „a“ nabývá velmi malých hodnot, u nosovek naopak velmi velkých) [1]. První formanty F_1 - vznikající rezonancí v dutině ústní. Druhé formanty F_2 - v dutině hrdelní.

Níže jsou uvedeny typické frekvenční rozsahy pro první dva formanty F_1 a F_2 u českých samohlásek: [4]

	Formant F_1	Formant F_2
a	750 - 1100 Hz	1100 - 1500 Hz
e	500 - 700 Hz	1500 - 2000 Hz
i	300 - 500 Hz	2000 - 3000 Hz
o	500 - 700 Hz	900 - 1200 Hz
u	300 - 500 Hz	600 - 1000 Hz



Obr. 5.4: Průběhy formantů.

6 Analýza

6.1 Tabulky hodnot

Všechny tabulky s parametry jsou zpracováním dat od různých mluvčích v šesti věkových skupinách pro větu „Hana tam nebude“. V tabulkách jsou uvedeny střední hodnoty a rozptyly jednotlivých parametrů v dané věkové skupině. Nejprve pro ženy a poté pro muže.

6.1.1 Výsledky analýzy pro ženy

Tabulka 6.1: Průměry analyzovaných parametrů v dílčích věkových skupinách pro ženy

Věkové skupiny	7-12	13-15	16-20	21-30	30-50	nad 50
T [s]	1,2047	1,0487	0,9851	0,9751	1,1665	1,1878
R [slabik/s]	5,0916	5,7576	6,2168	6,2224	5,5462	5,0659
Z_n [-]	97	85	83	83	90	96
F_{0min} [Hz]	203	190	167	144	161	124
F_{0max} [Hz]	329	338	426	310	362	418
F_{0str} [Hz]	253	247	223	204	219	205
F_{0std} [Hz]	28	27	50	30	42	52
F_{0sl} [Půltóny/s]	20	22	26	27	30	36
H_{min} [dB]	-1	-6	-2	-2	-2	-6
H_{max} [dB]	25	26	27	28	22	25
H_{str} [dB]	14	15	14	13	12	11
H_{std} [dB]	6	6	7	6	6	7
I_{min} [dB]	44	53	52	39	43	42
I_{max} [dB]	74	87	77	67	69	74
I_{str} [dB]	66	78	72	61	62	66
I_{std} [dB]	13	7	6	6	7	7
F_1 [Hz]	678	601	560	663	759	719
F_{1sp} [Hz]	205	134	310	257	566	401
F_2 [Hz]	1767	1616	1617	1749	1707	1669
F_{2sp} [Hz]	321	280	405	287	411	226
F_3 [Hz]	3016	2844	2463	3031	3025	3084
F_{3sp} [Hz]	610	766	1133	552	1196	1093
F_4 [Hz]	4175	3972	4168	4176	4133	4260
F_{4sp} [Hz]	578	1380	320	724	481	737

Tabulka 6.2: Směrodatné odchytky analyzovaných parametrů v dílčích věkových skupinách pro ženy

Věkové skupiny	7-12	13-15	16-20	21-30	30-50	nad 50
T [s]	0,1878	0,0830	0,1365	0,1007	0,4334	0,2881
R [slabik/s]	0,7185	0,4585	0,9116	0,6768	0,8476	0,6771
Z_n [-]	17	17	9	11	4	4
F_{0min} [Hz]	20	3	28	43	4	2
F_{0max} [Hz]	91	0	119	129	10	0
F_{0str} [Hz]	23	31	20	35	5	6
F_{0std} [Hz]	12	9	13	19	3	3
F_{0sl} [Půltóny/s]	4	1	5	5	2	1
H_{min} [dB]	3	3	2	2	2	2
H_{max} [dB]	2	0	2	4	2	0
H_{str} [dB]	2	1	1	2	1	1
H_{std} [dB]	0	1	1	1	1	1
I_{min} [dB]	6	2	3	3	2	1
I_{max} [dB]	10	2	4	4	3	2
I_{str} [dB]	8	0	3	3	3	1
I_{std} [dB]	17	1	1	1	4	1
F_1 [Hz]	77	40	51	83	9	6
F_{1sp} [Hz]	74	53	151	163	9	7
F_2 [Hz]	93	46	89	58	10	7
F_{2sp} [Hz]	130	127	328	187	11	11
F_3 [Hz]	131	174	574	100	11	13
F_{3sp} [Hz]	207	263	1388	531	14	16
F_4 [Hz]	119	121	32	114	11	11
F_{4sp} [Hz]	391	975	164	819	20	31

6.1.2 Výsledky analýzy pro muže

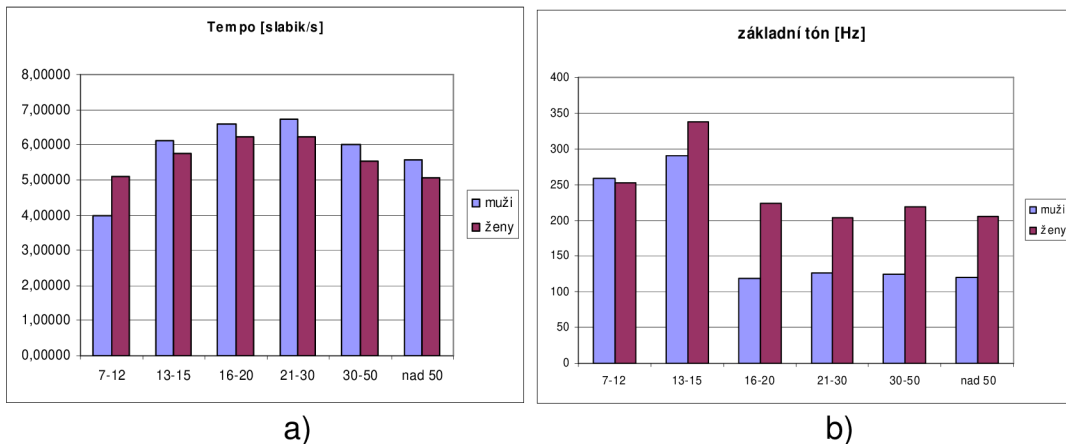
Tabulka 6.3: Průměry analyzovaných parametrů v dílčích věkových skupinách pro muže

Věkové skupiny	7-12	13-15	16-20	21-30	30-50	nad 50
T [s]	1,5563	0,9839	0,9200	0,9016	1,0129	1,0797
R [slabik/s]	3,9992	6,1227	6,5957	6,7262	6,0028	5,5571
Z_n [-]	110	79	75	73	83	100
F_{0min} [Hz]	168	134	95	99	89	86
F_{0max} [Hz]	400	291	166	190	155	144
F_{0str} [Hz]	259	166	118	126	125	121
F_{0std} [Hz]	37	23	16	23	18	18
F_{0sl} [Půltóny/s]	30	24	29	27	27	38
H_{min} [dB]	-1	-35	-2	-6	-2	-2
H_{max} [dB]	25	28	23	23	22	22
H_{str} [dB]	13	12	11	11	11	13
H_{std} [dB]	6	6	6	6	6	6
I_{min} [dB]	39	50	45	42	47	51
I_{max} [dB]	68	82	76	70	74	78
I_{str} [dB]	61	75	70	64	68	72
I_{std} [dB]	8	7	7	7	6	7
F_1 [Hz]	819	552	546	677	664	613
F_{1sp} [Hz]	185	421	324	392	388	256
F_2 [Hz]	1836	1470	1421	1672	1599	1654
F_{2sp} [Hz]	502	332	385	583	630	526
F_3 [Hz]	3189	2824	2693	2934	2963	3192
F_{3sp} [Hz]	537	783	462	524	562	768
F_4 [Hz]	4258	3857	3784	3963	3842	4161
F_{4sp} [Hz]	1496	684	758	541	284	692

Tabulka 6.4: Směrodatné odchytky analyzovaných parametrů v dílčích věkových skupinách pro ženy

Věkové skupiny	7-12	13-15	16-20	21-30	30-50	nad 50
T [s]	0,3149	0,0624	0,0939	0,0938	0,1189	0,0000
R [slabik/s]	0,7123	0,3878	0,7326	0,6853	0,6806	0,0000
Z_n [-]	8	7	11	8	8	0
F_{0min} [Hz]	50	32	8	13	6	0
F_{0max} [Hz]	108	187	99	97	22	0
F_{0str} [Hz]	5	38	10	10	10	0
F_{0std} [Hz]	14	15	15	22	4	0
F_{0sl} [Semitóny/s]	5	3	6	8	5	0
H_{min} [dB]	1	79	3	6	1	0
H_{max} [dB]	1	6	2	4	2	0
H_{str} [dB]	2	1	1	2	1	0
H_{std} [dB]	1	1	1	1	1	0
I_{min} [dB]	3	4	4	11	11	0
I_{max} [dB]	11	5	6	8	10	0
I_{str} [dB]	9	3	5	8	10	0
I_{std} [dB]	3	1	1	1	1	0
F_1 [Hz]	46	25	59	120	109	0
F_{1sp} [Hz]	48	364	139	164	165	0
F_2 [Hz]	24	109	96	112	208	0
F_{2sp} [Hz]	430	294	239	799	700	0
F_3 [Hz]	129	164	140	117	162	0
F_{3sp} [Hz]	145	651	338	266	271	0
F_4 [Hz]	127	228	156	130	167	0
F_{4sp} [Hz]	391	371	660	335	130	0

6.2 Grafy vypočtených hodnot

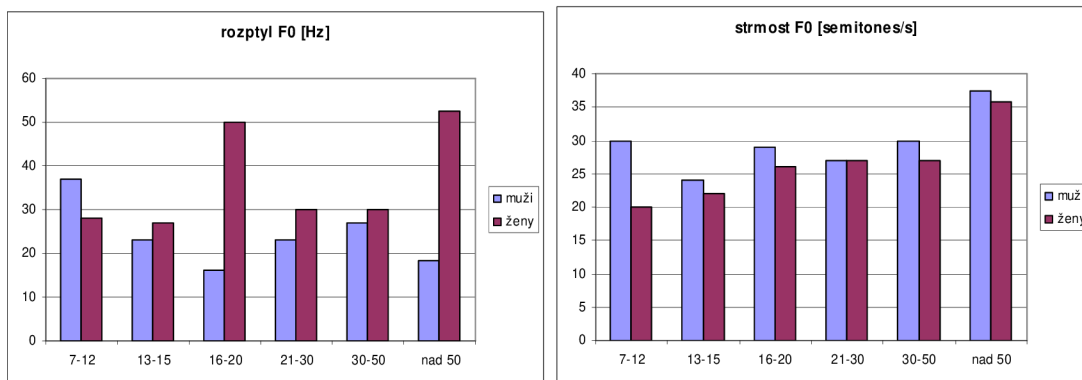


a) b)
**Obr. 6.1: Závislost hodnot různých příznaků na věkové skupině:
a) závislost střední hodnoty tempa řeči na věkové skupině
b) závislost střední hodnoty základního tónu na věkové skupině**

Z obrázku a) je patrné že tempo řeči je u žen ve většině věkových skupin nižší než u mužů. Což je překvapivý výsledek, protože předpoklad byl opačný. Jen ve skupině 7-12 let je u žen tempo vyšší. V této skupině jsou jen 3 chlapci. Při pořizování nahrávek měli problém s četbou vět na kterou se soustředili více než na samotnou reprodukci. Také byl jejich věkový průměr výrazně nižší než u dívek. Domnívám se, že při větším počtu mluvčích v mužské části by nám vyšel směrodatnější a vyšší výsledek tempa řeči.

V grafu je vidět, že tempo řeči (jak pro muže, tak pro ženy) s věkem roste, nabývá nejvyšších hodnot ve věku od 20 do 30 let a poté s věkem zase klesá. Nejvyšší tempo pro mluvčí ve věku 20-30 let je z psychologického hlediska logický. V tomto věku lze dosáhnout nejvyšších výkonů (fyzických i psychických) a lidé také žijí ve větším vytížení.

Z obrázku b) je vidět že ženy mají výrazně vyšší frekvenci základního tónu. Děti dosahují větších hodnot F_0 než dospělí. Nejvyšší frekvenci mají mluvčí v pubertálním věku (12-15let). U mužů v postpubertálním věku (16-20 let) po tom co přestanou mutovat a jejich biologické vlastnosti se měnit, hodnota frekvence F_0 výrazně klesne a ustálí se. U žen se frekvence základního tónu v průběhu stárnutí mírně mění.



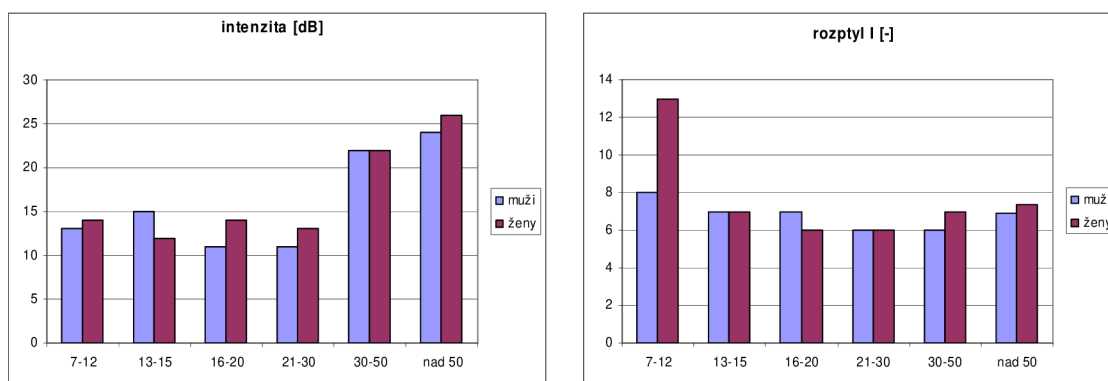
a)

b)

**Obr. 6.2: Závislost hodnot různých příznaků na věkové skupině:
a) závislost střední hodnoty rozptylu základního tónu na věkové skupině
b) závislost střední hodnoty strmosti základního tónu na věkové skupině**

Na obrázku a) je vidět rozptyl hodnot frekvence základního tónu. Nejvyšší hodnoty rozptylu jsou u žen ve věkových skupinách „16-20“ a nad „50 let“. V těchto věkových skupinách mají ženy individuálně dost odlišné velikosti hodnoty frekvence základního tónu.

Z obrázku b) je vidět, že u žen strmost základního tónu F_0 s přibývajícím věkem roste. U mužů do dvaceti let mění (v pubertálním věku je pokles) a potom též s věkem roste.



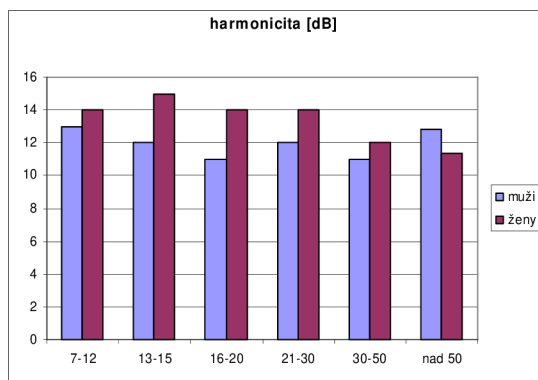
a)

b)

**Obr. 6.3: Závislost hodnot různých příznaků na věkové skupině:
a) závislost střední hodnoty intenzity na věkové skupině
b) závislost rozptylu intenzity na věkové skupině**

Z obrázku a) vidíme že hodnoty intenzity střídavě rostou a klesají do 30. roku života a poté s narůstajícím věkem roste, jak pro muže, tak pro ženy. Nejvyšších hodnot intenzity dosahuje skupina nad „50 let“.

Na obrázku b) je rozptyl hodnot intenzity. Je poměrně konstantní a dosahuje nízkých hodnot. Z toho lze usoudit že by intenzita mohla být směrodatným příznakem pro určení, jestli je mluvčí starší či mladší než 35 let. Větší výkyv hodnot je pouze pro ženy ve věku 7-12 let, což by mohlo činit problémy. Vzhledem k tomu že je rozptyl výrazně vyšší jen v této skupině, je možné že do výsledku byla zanesena chyba (například chybným zanesením hodnot do tabulky nebo při výpočtu).



Obr. 6.4: Závislost střední hodnoty harmonicity na věkové skupině

Z obrázku vidíme, že hodnoty harmonicity se s věkem mění. Ženy mají všeobecně vyšší harmonicitu než muži. Tudíž by šlo říct, že i větší znělost hlasu a tím pádem menší šumové procento. U mužů nad 50 let zřejmě mohlo dojít k chybě výsledku způsobenou nedostatečným počtem nahrávek.

7 Analýza použitých příznaků pro automatické rozpoznání věku a pohlaví mluvčího

V první části, která byla obhájena v semestrálním projektu byla provedena analýza příznaků pro jednu větu.

V druhé části (bakalářské práce) byla analýza dokončena pro zbývajících pět vět. Aritmetické průměry výsledků příznaků jsou uvedeny v tabulkách Tab. 7.1, 7.2 a 7.3. Těchto příznaků je využito pro automatické rozpoznání pohlaví a věku mluvčího pomocí vhodného klasifikátoru. V našem případě byla jako klasifikátor zvolena neuronová síť.

Následující kapitoly popisují výběr příznaků a rozdělení promluv pro klasifikaci. Dále budou popsány základní typy neuronových sítí, použité nástroje při práci a samotný proces klasifikace do tříd.

Tabulka 7.1: Aritmetické průměry výsledků příznaků pro VS1 a VS2.

VS1 Z	Věta:	1.	2.	3.	4.	5.	6.
	T	1,20	1,18	1,83	1,87	4,39	4,22
	R	5,09	4,37	3,97	5,51	4,64	3,78
	Z_n	97	81	111	127	294	301
	F_{0min}	203	193	193	196	161	175
	F_{0max}	329	503	392	426	588	532
	F_{0str}	253	259	258	256	264	252
	F_{0std}	28	62	36	45	70	54
	F_{0sl}	20	26	21	18	26	24
	H_{min}	-227	-227	-227	-227	-227	-227
	H_{max}	37	43	40	43	44	43
	H_{str}	14	12	12	13	12	12
	H_{std}	6	7	8	7	7	6
	I_{min}	44	41	40	39	38	38
	I_{max}	74	74	75	74	74	73
	I_{str}	65	65	65	64	63	63
	I_{std}	7	8	7	8	7	7
	F_1	693	799	832	740	768	796
	F_{1sp}	276	255	318	289	299	290
	F_2	1767	1870	2111	1871	1875	1939
	F_{2sp}	326	362	485	464	444	457
	F_3	3052	3202	3319	3213	3223	3245
	F_{3sp}	567	529	496	513	586	553
	F_4	4183	4304	4379	4341	4360	4361
	F_{4sp}	564	561	528	534	544	550
VS2 Z	Věta:	1.	2.	3.	4.	5.	6.
	T	1,05	1,01	1,42	1,46	3,11	2,92
	R	5,76	5,00	5,01	6,95	6,10	5,15
	Z_n	85	71	87	107	215	226
	F_{0min}	190	108	186	178	154	179
	F_{0max}	338	474	441	280	587	456
	F_{0str}	232	233	241	225	246	233
	F_{0std}	27	67	47	25	73	41
	F_{0sl}	22	27	23	21	35	28
	H_{min}	-227	-228	-227	-227	-227	-227
	H_{max}	40	46	46	43	45	40
	H_{str}	15	14	11	13	12	13
	H_{std}	6	6	8	8	8	7
	I_{min}	53	50	51	50	48	48
	I_{max}	87	86	82	84	83	81
	I_{str}	78	76	73	74	74	73
	I_{std}	7	8	7	7	8	7
	F_1	601	732	700	620	625	613
	F_{1sp}	199	157	240	166	188	165
	F_2	1616	1751	1946	1672	1682	1776
	F_{2sp}	201	307	455	285	275	309
	F_3	2844	2978	3214	2968	2971	2984
	F_{3sp}	623	593	519	540	608	557
	F_4	3972	4083	4212	4124	4081	4088
	F_{4sp}	555	536	487	535	512	535

VS1 M	Věta:	1.	2.	3.	4.	5.	6.
	T	1,56	1,21	2,04	2,18	6,27	6,64
	R	4,00	4,15	3,57	4,72	3,42	2,52
	Z_n	110	73	134	143	406	386
	F_{0min}	168	204	201	205	198	162
	F_{0max}	400	489	561	326	587	506
	F_{0str}	259	257	272	262	263	258
	F_{0std}	37	52	71	32	64	55
	F_{0sl}	30	24	25	22	24	22
	H_{min}	-226	-227	-228	-227	-228	-228
	H_{max}	34	44	43	44	43	48
	H_{str}	13	13	12	12	13	13
	H_{std}	6	7	7	7	6	6
	I_{min}	39	40	40	39	37	-91
	I_{max}	68	78	72	74	75	76
	I_{str}	61	67	61	64	63	62
	I_{std}	8	9	7	8	8	51
	F_1	819	834	940	757	789	864
	F_{1sp}	394	246	352	293	302	322
	F_2	1836	1818	2168	1939	1867	1996
	F_{2sp}	389	353	428	411	446	490
	F_3	3189	3188	3294	3290	3161	3319
	F_{3sp}	558	592	537	499	576	597
	F_4	4258	4289	4336	4316	4296	4306
	F_{4sp}	597	616	581	550	563	604
VS2 M	Věta:	1.	2.	3.	4.	5.	6.
	T	0,98	0,91	1,40	1,46	2,96	2,83
	R	6,12	5,55	5,00	6,89	6,44	5,34
	Z_n	79	61	86	94	199	208
	F_{0min}	134	123	135	132	100	116
	F_{0max}	291	320	460	286	521	521
	F_{0str}	166	162	181	165	181	172
	F_{0std}	23	32	64	28	71	53
	F_{0sl}	24	24	28	20	31	29
	H_{min}	-227	-227	-227	-227	-227	-227
	H_{max}	38	41	41	41	40	44
	H_{str}	12	11	9	11	10	11
	H_{std}	6	6	7	8	7	7
	I_{min}	50	50	51	50	49	46
	I_{max}	82	82	81	85	83	82
	I_{str}	75	74	72	74	73	73
	I_{std}	7	8	6	7	6	7
	F_1	552	663	769	617	610	621
	F_{1sp}	228	198	252	185	202	168
	F_2	1470	1648	1960	1669	1635	1730
	F_{2sp}	211	233	337	283	262	317
	F_3	2824	2914	3128	2957	2947	3005
	F_{3sp}	567	521	453	496	531	511
	F_4	3857	4001	4132	4030	4028	4028
	F_{4sp}	502	490	479	441	427	461

Tabulka 7.2: Aritmetické průměry výsledků příznaků pro VS3 a VS4.

VS3 Z	Věta:	1.	2.	3.	4.	5.	6.	VS3 M	Věta:	1.	2.	3.	4.	5.	6.
	T	0,99	1,31	1,84	1,48	3,24	3,19		T	0,92	0,93	1,34	1,38	2,92	2,90
	R	6,22	4,44	3,93	6,86	5,90	4,73		R	6,60	5,73	5,29	7,34	6,59	5,24
	Z_n	83	64	98	104	220	223		Z_n	75	54	73	81	196	203
	F_{0min}	167	152	166	150	105	123		F_{0min}	95	91	93	89	85	86
	F_{0max}	426	442	397	374	488	477		F_{0max}	166	367	384	194	527	480
	F_{0str}	223	227	228	221	218	210		F_{0str}	118	128	139	121	131	123
	F_{0std}	50	62	53	41	65	49		F_{0std}	16	58	69	24	65	50
	F_{0sl}	26	30	28	25	34	26		F_{0sl}	29	32	28	27	32	29
	H_{min}	-226	-226	-227	-227	-227	-228		H_{min}	-226	-227	-226	-226	-227	-227
	H_{max}	36	34	41	37	45	46		H_{max}	36	37	38	38	40	43
	H_{str}	14	10	9	12	11	12		H_{str}	11	9	8	10	10	11
	H_{std}	7	7	7	8	7	6		H_{std}	6	5	6	6	6	6
	I_{min}	52	46	46	46	45	44		I_{min}	46	18	41	42	39	39
	I_{max}	77	83	86	83	85	86		I_{max}	76	79	77	79	78	78
	I_{str}	72	73	74	73	75	73		I_{str}	70	70	68	69	68	69
	I_{std}	6	10	9	8	8	8		I_{std}	7	18	8	8	7	8
	F_1	560	778	760	578	635	683		F_1	546	675	809	640	668	630
	F_{1sp}	259	307	319	253	237	237		F_{1sp}	276	221	324	252	269	225
	F_2	1617	1695	1950	1591	1676	1773		F_2	1421	1584	1938	1649	1637	1678
	F_{2sp}	170	353	359	277	224	247		F_{2sp}	240	235	344	306	307	241
	F_3	2848	2943	3194	2951	3022	3098		F_3	2693	2844	3063	2872	2932	2887
	F_{3sp}	622	648	499	625	546	579		F_{3sp}	491	465	456	433	507	444
	F_4	4168	4112	4229	4097	4169	4195		F_4	3784	3871	4089	3945	3935	3901
	F_{4sp}	500	572	510	490	430	501		F_{4sp}	453	419	469	405	390	389
VS4 Z	Věta:	1.	2.	3.	4.	5.	6.	VS4 M	Věta:	1.	2.	3.	4.	5.	6.
	T	0,98	0,99	1,47	1,52	3,41	3,09		T	0,90	0,91	1,49	1,47	3,42	3,29
	R	6,22	5,11	4,79	6,63	5,65	4,94		R	6,73	5,53	4,82	6,91	5,66	4,66
	Z_n	83	58	79	89	225	216		Z_n	73	57	75	92	202	212
	F_{0min}	141	146	161	145	105	138		F_{0min}	98	100	97	96	91	95
	F_{0max}	311	433	442	302	450	379		F_{0max}	190	372	426	307	532	437
	F_{0str}	204	217	221	203	209	198		F_{0str}	126	150	142	131	141	126
	F_{0std}	30	60	50	35	46	36		F_{0std}	23	76	67	44	71	42
	F_{0sl}	27	26	26	21	28	25		F_{0sl}	27	32	33	25	29	28
	H_{min}	-227	-227	-227	-227	-227	-227		H_{min}	-226	-227	-227	-226	-227	-227
	H_{max}	35	40	35	40	42	43		H_{max}	34	41	44	39	41	45
	H_{str}	13	11	9	11	11	11		H_{str}	11	9	8	9	9	9
	H_{std}	6	7	7	9	7	6		H_{std}	6	6	6	7	6	6
	I_{min}	39	37	36	37	35	34		I_{min}	42	40	39	39	39	40
	I_{max}	67	70	68	74	68	70		I_{max}	70	73	73	72	73	75
	I_{str}	61	61	59	61	59	59		I_{str}	64	65	63	63	63	65
	I_{std}	6	8	7	7	7	7		I_{std}	7	8	7	7	7	7
	F_1	663	819	987	817	800	849		F_1	677	835	1020	818	887	860
	F_{1sp}	260	270	360	335	304	346		F_{1sp}	336	340	442	343	394	377
	F_2	1749	1879	2187	1922	1896	1991		F_2	1672	1829	2127	1925	1957	1965
	F_{2sp}	344	419	515	503	470	488		F_{2sp}	365	463	439	438	529	493
	F_3	3031	3067	3324	3160	3135	3190		F_3	2934	3081	3279	3143	3215	3192
	F_{3sp}	633	684	564	632	675	640		F_{3sp}	578	632	549	581	618	623
	F_4	4176	4207	4398	4308	4253	4307		F_4	3963	4082	4227	4130	4196	4185
	F_{4sp}	469	600	568	546	559	578		F_{4sp}	417	521	525	478	493	507

Tabulka 7.3: Aritmetické průměry výsledků příznaků pro VS5 a VS6.

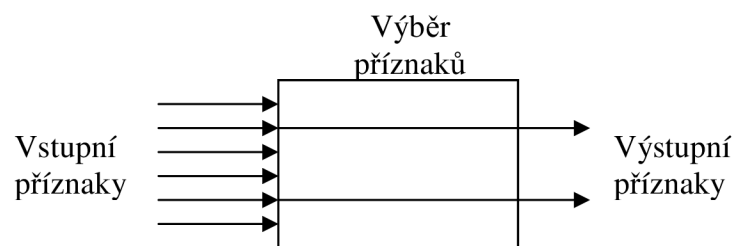
VS5 Z	Věta:	1.	2.	3.	4.	5.	6.	VS5 M	Věta:	1.	2.	3.	4.	5.	6.
	T	1,01	1,03	1,73	1,57	3,33	3,42		T	1,01	0,97	1,55	1,51	3,29	3,38
	R	5,98	4,90	4,18	6,58	5,76	4,49		R	6,00	5,19	4,59	6,67	5,80	4,46
	Z_n	89	75	106	121	247	264		Z_n	83	63	100	101	234	244
	F_{0min}	160	161	161	153	122	143		F_{0min}	88	88	85	88	90	86
	F_{0max}	318	371	393	367	506	422		F_{0max}	156	328	328	310	530	415
	F_{0str}	217	226	224	219	225	212		F_{0str}	125	135	129	128	137	129
	F_{0std}	36	56	47	44	65	45		F_{0std}	18	59	49	39	62	44
	F_{0sl}	30	31	25	26	34	28		F_{0sl}	27	36	34	28	35	32
	H_{min}	-226	-226	-227	-226	-227	-227		H_{min}	-226	-226	-227	-227	-227	-227
	H_{max}	36	36	39	34	40	38		H_{max}	36	38	41	38	42	42
	H_{str}	12	9	9	11	9	11		H_{str}	11	9	8	9	10	10
	H_{std}	6	5	6	6	6	6		H_{std}	6	5	6	6	6	6
	I_{min}	43	41	40	40	38	38		I_{min}	47	44	42	42	42	41
	I_{max}	69	66	67	69	68	69		I_{max}	74	76	74	76	79	77
	I_{str}	62	58	59	60	59	59		I_{str}	68	67	66	67	68	68
	I_{std}	6	6	7	7	6	7		I_{std}	6	8	7	7	7	8
	F_1	759	800	914	768	806	790		F_1	664	815	937	696	731	734
	F_{1sp}	392	393	444	369	406	403		F_{1sp}	366	329	370	365	320	311
	F_2	1707	1798	2143	1853	1867	1891		F_2	1599	1856	2077	1770	1804	1828
	F_{2sp}	255	396	392	368	410	379		F_{2sp}	270	369	346	311	335	300
	F_3	3025	3097	3343	3168	3220	3229		F_3	2963	3082	3225	3042	3126	3077
	F_{3sp}	609	669	513	686	664	637		F_{3sp}	447	671	536	518	492	530
	F_4	4133	4173	4304	4200	4237	4255		F_4	3842	3971	4098	3947	3972	3945
	F_{4sp}	521	597	573	603	547	542		F_{4sp}	448	489	477	454	438	421
VS6 Z	Věta:	1.	2.	3.	4.	5.	6.	VS6 M	Věta:	1.	2.	3.	4.	5.	6.
	T	1,19	1,28	1,98	1,87	4,10	4,54		T	1,08	1,22	2,16	1,77	3,53	3,55
	R	5,07	3,95	3,59	5,39	4,65	3,34		R	5,56	4,09	3,24	5,65	5,39	4,23
	Z_n	96	84	89	126	223	331		Z_n	100	75	132	120	237	232
	F_{0min}	124	135	134	123	106	123		F_{0min}	86	106	89	93	92	87
	F_{0max}	418	504	414	343	463	392		F_{0max}	144	568	198	168	213	165
	F_{0str}	205	198	221	183	197	182		F_{0str}	121	180	136	135	126	122
	F_{0std}	52	78	64	47	60	40		F_{0std}	18	129	20	19	15	19
	F_{0sl}	36	35	22	23	25	28		F_{0sl}	38	27	39	25	37	39
	H_{min}	-226	-227	-227	-227	-227	-227		H_{min}	-227	-226	-226	-226	-227	-227
	H_{max}	36	36	33	37	44	40		H_{max}	43	39	36	41	38	40
	H_{str}	11	9	8	11	9	10		H_{str}	13	12	6	10	8	8
	H_{std}	7	5	7	7	7	6		H_{std}	6	5	6	6	6	6
	I_{min}	42	44	41	43	40	39		I_{min}	51	48	50	48	49	46
	I_{max}	74	75	78	79	83	75		I_{max}	78	85	81	83	81	85
	I_{str}	66	67	66	68	70	63		I_{str}	72	75	72	74	73	73
	I_{std}	7	8	8	8	8	7		I_{std}	7	9	6	7	7	8
	F_1	719	788	920	695	736	795		F_1	613	746	856	637	630	670
	F_{1sp}	342	281	382	318	314	329		F_{1sp}	331	259	255	242	268	279
	F_2	1669	1763	2142	1826	1796	1936		F_2	1654	1616	1911	1787	1613	1738
	F_{2sp}	243	291	384	274	296	300		F_{2sp}	292	247	263	274	320	241
	F_3	3084	3113	3407	3155	3196	3282		F_3	3192	3113	3115	2971	3144	3018
	F_{3sp}	560	501	501	511	585	534		F_{3sp}	681	722	655	591	689	730
	F_4	4260	4292	4416	4280	4232	4362		F_4	4161	4177	4114	4055	4066	4048
	F_{4sp}	546	531	579	503	564	505		F_{4sp}	629	544	546	529	515	495

8 Výběr příznaků pro automatickou klasifikaci mluvních

Pro výběr příznaků byly použity především postupy popsané v [1].

Pro numerické zpracování je vhodné upravit vhodnými konstantami velikosti některých příznaků tak, aby hodnoty všech příznaků byly řádové stejné. Zabrání se tím některým numerickým potížím. Proto byla před učením neuronové sítě nejprve převedena matice vybraných příznaků dělením vhodným číslem na normovanou matici příznaků s hodnotami od jedné do sta.

Pro optimální výběr příznaků jsou stanovena kritéria kvality příznaků, která nám umožňují posoudit jak jsou pro daný účel vhodná pro rozpoznávání. V ideálním případě má kritérium kvality umožňovat již přímo odhad očekávané chybovosti celého systému. Použité metody pro výběr cílové skupiny příznaků vycházejí z principů selekce, kdy vybíráme ze stávajících příznaků jen některé příznaky a vybrané příznaky přitom zůstávají v původní podobě.



Obr. 8.1: Princip selekce příznaků.

Na obrázku 8.1 je znázorněn princip selekce, kde čáry vlevo představují vstupní soubor příznaků, blok představuje výběr vhodných příznaků podle určitých kritérií a šipky vpravo představují vybrané příznaky vhodné pro použití pro konkrétní práci.

8.1 Výběr příznaků podle rozptylů

Důležitým předpokladem výběru hodnotných příznaků je stanovení vhodného kritéria k vyjádření oddělitelnosti tříd. Třídou rozumíme podle konkrétního systému, který má rozlišovat například skupiny příbuzných hlásek.

V našem případě rozlišujeme dva typy tříd: 1. pohlaví, 2. věkové skupiny. Pozorováním hodnot příznaku x_i v příznakovém prostoru vznikne jednoduché kritérium $Q(x_i)$ pro vyjádření kvality příznaku. Přičemž x_i vyjadřuje i -tý příznak. Příznak je považován za zvláště kvalitní, pokud se na jednak prvky sejné třídy vyskytují blízko kolem střední hodnoty a zároveň se střední hodnoty jednotlivých tříd mezi sebou liší, pro různé třídy pokud možno co nejvíce.

Oba požadavky jsou vyjádřeny v následujících definicích. Kvadrát vzdáleností mezi středními hodnotami třídy v a třídy u je

$$D_{v,u}^2 = (\mu_v - \mu_u)^2 \quad \text{(8-1)}$$

Aritmetickou střední hodnotu vzdáleností mezi všemi třídami určíme podle

$$D^2 = \frac{1}{V(V-1)} \sum_{v=1}^V \sum_{u=1}^V D_{v,u}^2 \quad \text{(8-2)}$$

Kde V značí celkový počet tříd. Kvadrát rozptylu třídy v kolem střední hodnoty je

$$S_v^2 = E\{(x - \mu_v)^2\} \quad \text{(8-3)}$$

a jeho aritmetická střední hodnota

$$S^2 = \frac{1}{V} \sum_{v=1}^V S_v^2 \quad \text{(8-4)}$$

Informaci o geometrické oddělitelnosti (separabilitě) tříd v příznakovém prostoru dává míra definovaná poměrem

$$Q(\cdot) = \frac{S^2}{S^2 + D^2} \quad 0 \leq Q(\cdot) \leq 1 \quad \text{(8-5)}$$

pro jeden příznak $Q(x_1, x_2, \dots)$, která je omezena na intervalu od 0 do 1. Pokud vykazuje příznak x_i velké rozdíly pro různé třídy a malé rozdíly pro různé prvky v jedné třídě, má míra $Q(x_i)$ malou hodnotu, v ideálním případě je $Q(x_i) = 0$. Velké hodnoty naopak ukazují že příznaky jsou velmi rozptýleny a jsou tedy nevhodné pro rozpoznávání. Krajní případy vyjadřují:

$Q(\cdot) = 0$... optimální plná odlučitelnost

$Q(\cdot) = 1$... žádná oddělitelnost (totální chaos v příznacích)

Nevýhodou míry Q je, že nebere do úvahy korelaci mezi příznaky. Například mohou být vybrány jako nejlepší příznaky téměř dva identické příznaky s příslušnou vysokou hodnotou $Q(x_i)$. Blíže bude vysvětleno později. Další nevýhodou je klesající účinnost pro větší počet tříd.

8.1.1 Příklady výpočtu

Pro názornost bude následovně uveden příklad výpočtu kvality $Q(x_i)$ na střední hodnotě základního tónu $F_{0\text{str}}$ pro rozlišení pohlaví, kde rozlišujeme dvě třídy :

1. muži,
2. ženy.

Uvedené výpočty byli provedeny shodně s předešlými vzorci pomocí funkcí MS Excel a poté ověřeny v programu Matlab.

Nejprve byli rozděleny všechny vzorky do dvou tříd (množin hodnot) podle pohlaví na M – muži a Z – ženy.

Poté byly vypočteny střední hodnoty (aritmetický průměr) jednotlivých tříd

$$\begin{aligned} Z_{STR} &= 228,5943 \\ M_{STR} &= 147,1097 \end{aligned} \quad \text{(8-6)}$$

Kvadrát vzdáleností mezi středními hodnotami třídy M a třídy Z se vypočítá

$$D_{Z,M}^2 = (\mu_Z - \mu_M)^2 = (228,5943 - 147,1097)^2 = 6639,7412 \quad \text{(8-7)}$$

Aritmetickou střední hodnotu určíme jako aritmetický průměr, kdy při dvou třídách nám vychází jen jedna vzdálenost, tedy

$$D^2 = 6639,7412 \quad \text{(8-8)}$$

Dále byl vypočítán rozptyl pro jednotlivé třídy

$$\begin{aligned} S_Z^2 &= E\{(x - \underline{\mu}_Z)^2\} = 1192,0289 \\ S_M^2 &= E\{(x - \underline{\mu}_M)^2\} = 1845,7149 \end{aligned} \quad \text{(8-9)}$$

a aritmetická střední hodnota rozptylů

$$S^2 = \frac{1}{V} \sum_{v=1}^V S_v^2 = \frac{(S_Z^2 + S_M^2)}{2} = \frac{1192,0289 + 1845,7149}{2} = 1518,8719 \quad \text{(8-10)}$$

Výsledná míra kvality se vypočítá

$$Q(\cdot) = \frac{S^2}{S^2 + D^2} = \frac{1518,8719}{1518,8719 + 6639,7412} = 0,1862 \quad \text{(8-11)}$$

Stejně bylo postupováno při výpočtu kritéria kvality Q při rozdělení na šest tříd podle věkové skupiny.

Pro názornost bude následovně uveden příklad výpočtu kvality $Q(x_i)$ na střední hodnotě základního tónu F_{0str} pro rozlišení věkové skupiny, kde rozlišujeme 6 tříd pro něž střední hodnoty jsou:

$$\begin{aligned}
 VS1_{STR} &= 258,1784 \\
 VS2_{STR} &= 190,2006 \\
 VS3_{STR} &= 141,5767 \\
 VS4_{STR} &= 169,7555 \\
 VS5_{STR} &= 179,5484 \\
 VS6_{STR} &= 182,4389
 \end{aligned}
 \tag{8-12}$$

Kvadráty vzdáleností mezi středními hodnotami tříd jsou

$$\begin{aligned}
 D_{VS1,VS2}^2 &= (\mu_{VS1} - \mu_{VS2})^2 = (258,1784 - 190,2006)^2 = 4620,9822 \\
 D_{VS1,VS3}^2 &= (\mu_{VS1} - \mu_{VS3})^2 = 13595,9507 \\
 D_{VS1,VS4}^2 &= (\mu_{VS1} - \mu_{VS4})^2 = 7818,6046 \\
 D_{VS1,VS5}^2 &= (\mu_{VS1} - \mu_{VS5})^2 = 6182,6679 \\
 D_{VS1,VS6}^2 &= (\mu_{VS1} - \mu_{VS6})^2 = 5736,4636 \\
 D_{VS2,VS3}^2 &= (\mu_{VS2} - \mu_{VS3})^2 = 2364,2805 \\
 D_{VS2,VS4}^2 &= (\mu_{VS2} - \mu_{VS4})^2 = 418,0007 \\
 D_{VS2,VS5}^2 &= (\mu_{VS2} - \mu_{VS5})^2 = 113,4680 \\
 D_{VS2,VS6}^2 &= (\mu_{VS2} - \mu_{VS6})^2 = 60,2430 \\
 D_{VS3,VS4}^2 &= (\mu_{VS3} - \mu_{VS4})^2 = 794,0448 \\
 D_{VS3,VS5}^2 &= (\mu_{VS3} - \mu_{VS5})^2 = 1441,8524 \\
 D_{VS3,VS6}^2 &= (\mu_{VS3} - \mu_{VS6})^2 = 1669,7218 \\
 D_{VS4,VS5}^2 &= (\mu_{VS4} - \mu_{VS5})^2 = 95,9014 \\
 D_{VS4,VS6}^2 &= (\mu_{VS4} - \mu_{VS6})^2 = 160,8693 \\
 D_{VS5,VS6}^2 &= (\mu_{VS5} - \mu_{VS6})^2 = 8,3549
 \end{aligned}
 \tag{8-13}$$

Jejich počet musí vycházet $n_{D^2} = \binom{6}{2}$, aby byly vyčerpány všechny kombinace.

Aritmetická střední hodnota patnácti hodnot kvadrátů vzdáleností je

$$D^2 = 3005,4271 \tag{8-14}$$

Dále byly vypočítány rozptyly daných tříd

$$\begin{aligned}
S_{VS1}^2 &= E\{(x - \mu_Z)^2\} = 412,9778 \\
S_{VS2}^2 &= E\{(x - \mu_M)^2\} = 2036,7216 \\
S_{VS3}^2 &= E\{(x - \mu_M)^2\} = 1545,5867 \\
S_{VS4}^2 &= E\{(x - \mu_M)^2\} = 2116,7581 \\
S_{VS5}^2 &= E\{(x - \mu_M)^2\} = 2396,3180 \\
S_{VS6}^2 &= E\{(x - \mu_M)^2\} = 1862,7093
\end{aligned}
\tag{8-15}$$

a jejich aritmetická střední hodnota

$$S^2 = \frac{1}{V} \sum_{v=1}^V S_v^2 = \frac{(S_{VS1}^2 + S_{VS2}^2 + S_{VS3}^2 + S_{VS4}^2 + S_{VS5}^2 + S_{VS6}^2)}{6} = 1728,5119. \tag{8-16}$$

Výsledná míra kvality se vypočítá

$$Q = \frac{S^2}{S^2 + D^2} = \frac{1728,5119}{1728,5119 + 3005,4271} = 0,3651. \tag{8-17}$$

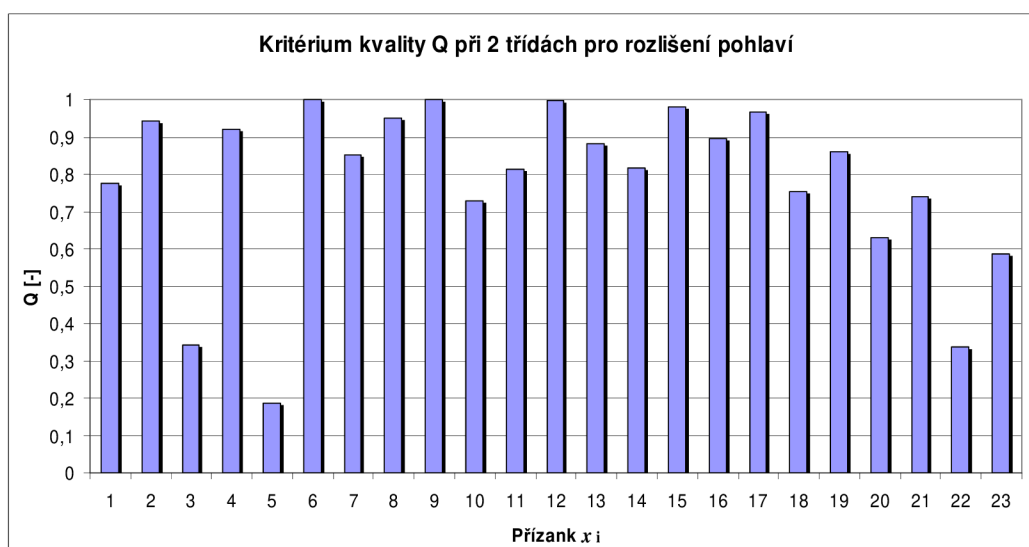
Pro kritérium kvality Q byla vytvořena tabulka Tab. 8.1 kde jsou zahrnuty výsledky pro rozlišení na třídy podle pohlaví a podle věku. Dále byli do tabulky vyneseny hodnoty, pro kritérium kvality pro rozpoznávání věku zvlášť pro muže a zvlášť pro ženy.

Tabulka 8.1: Hodnoty kritéria kvality příznaků.

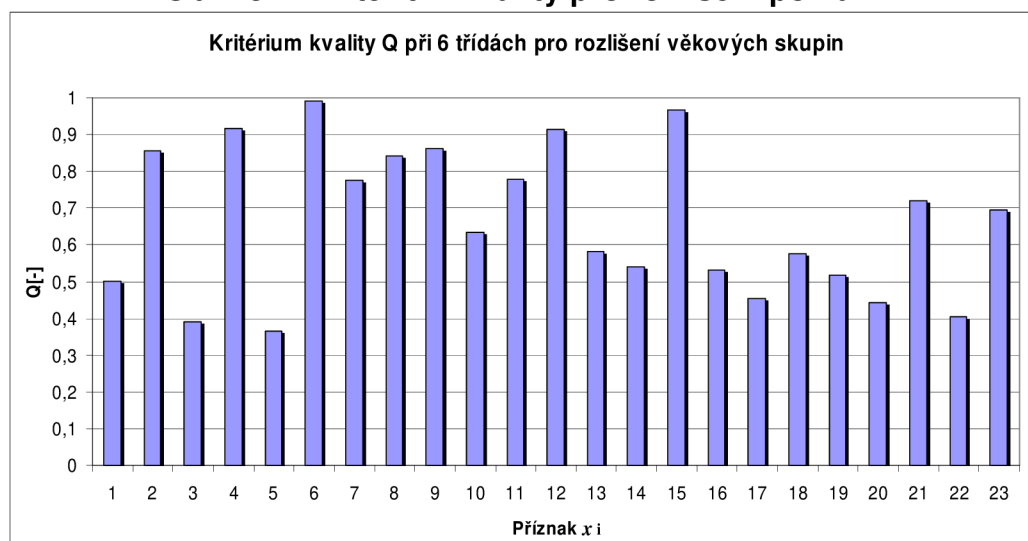
Q	Pro věk			Pro pohlaví
	Všichni	Ženy	Muži	
R	0,5015	0,6452	0,3682	0,7755
Z_n	0,8560	0,9185	0,7415	0,9416
F_{0min}	0,3909	0,6018	0,1108	0,3427
F_{0max}	0,9175	0,9351	0,7387	0,9216
F_{0str}	0,3651	0,4649	0,0823	0,1862
F_{0std}	0,9918	0,9380	0,9492	0,9995
F_{0sl}	0,7749	0,8087	0,6380	0,8530
H_{min}	0,8413	0,8327	0,7572	0,9512
H_{max}	0,8615	0,7563	0,8955	0,9999
H_{str}	0,6332	0,5814	0,4915	0,7289
H_{std}	0,7774	0,7237	0,5908	0,8133
I_{min}	0,9154	0,3722	0,8477	0,9970
I_{max}	0,5807	0,3453	0,6299	0,8813
I_{str}	0,5407	0,2992	0,5423	0,8153
I_{std}	0,9661	0,6499	0,8953	0,9813
F_1	0,5311	0,4906	0,4868	0,8948
F_{1sp}	0,4555	0,3687	0,4542	0,9676
F_2	0,5759	0,6044	0,5391	0,7543
F_{2sp}	0,5182	0,5173	0,4946	0,8602
F_3	0,4431	0,5717	0,4023	0,6297
F_{3sp}	0,7207	0,7339	0,4382	0,7387
F_4	0,4033	0,4664	0,3685	0,3372
F_{4sp}	0,6966	0,8745	0,5001	0,5866

Zvýrazněné hodnoty v tabulce jsou soubory dvanácti příznaků s nejnižšími hodnotami Q pro daný účel rozeznávání, které byli dále použity pro trénování neuronové sítě. Maximální počet příznaků bylo zvoleno 12, na základě doporučení vhodného počtu deseti až dvaceti příznaků z [1]. Natrénování neuronové sítě s dvanácti příznaky bude trvat delší dobu než u jiných klasifikátorů. Například u HMM (Skryté Markovovy modely), je trénování rychlejší. Samotné trénování i popis neuronové sítě je popsáno v kapitolách níže.

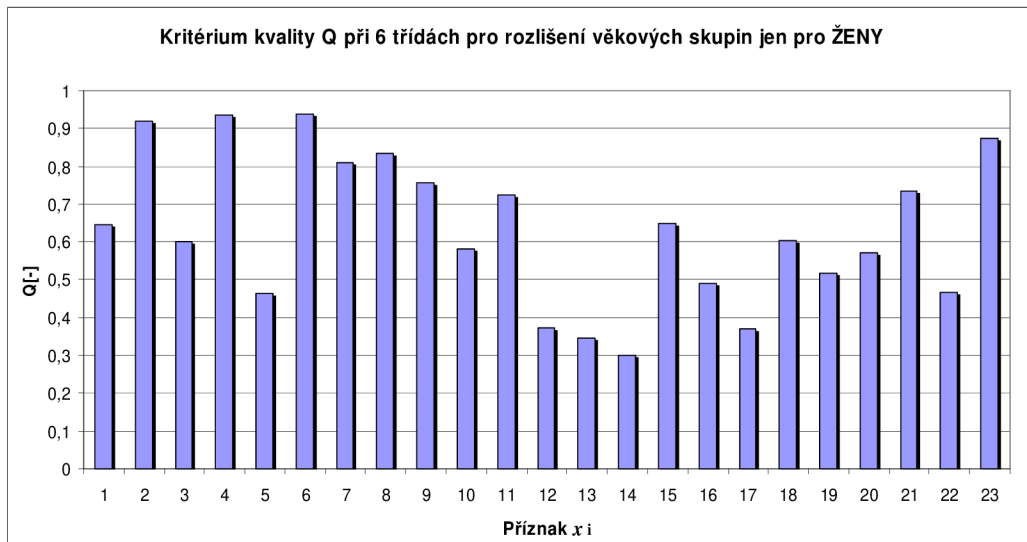
Výsledky hodnot kritéria kvality byly pro větší názornost zaneseny do čtyř sloupcových grafů.



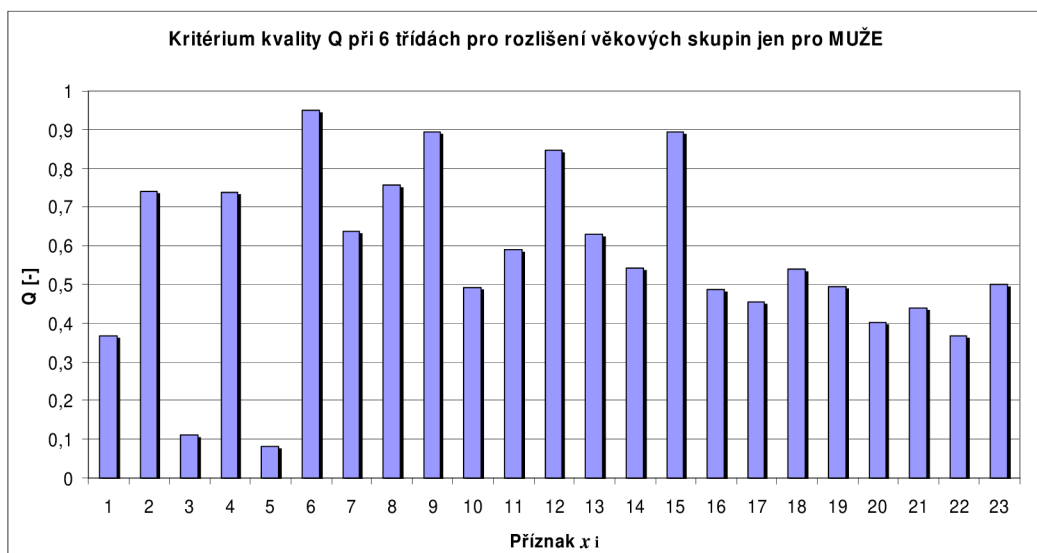
Obr. 8.2: Kritérium kvality pro rozlišení pohlaví.



Obr. 8.3: Kritérium kvality pro rozlišení věku.



Obr. 8.4: Kritérium kvality pro rozlišení věku jen pro ženy.



Obr. 8.5: Kritérium kvality pro rozlišení věku jen pro muže.

Kde i -tým členem příznaku x_i rozumíme:

1. R , 2. Z_n , 3. F_{0min} , 4. F_{0max} , 5. F_{0str} , 6. F_{0std} , 7. F_{0sl} , 8. H_{min} , 9. H_{max} , 10. H_{str} , 11. H_{std} ,
12. I_{min} , 13. I_{max} , 14. I_{str} , 15. I_{std} , 16. F_1 , 17. F_{1sp} , 18. F_2 , 19. F_{2sp} , 20. F_3 , 21. F_{3sp} , 22. F_4 ,
23. F_{4sp} .

V prvním grafu Obr. 8.2 jsou vyneseny hodnoty kritéria kvality, které popisují vhodnost jednotlivých příznaků pro rozpoznávání pohlaví. Graf Obr. 8.3 popisuje totéž pro rozpoznávání věku a v grafech Obr. 8.4, 8.5 bylo vyneseno kritérium kvality pro rozpoznávání věku pro muže a ženy zvlášť. Z uvedených výsledků plyne že je pro rozpoznání věkové skupiny vhodnější nejprve rozpoznat pohlaví mluvčích, protože hodnoty kvality příznaků pro muže a ženy při rozpoznávání věku se mohou zásadně lišit.

8.2 Korelace příznaků

Příznaky by neměly být vzájemně korelovány. Pokud dva příznaky silně korelují, je hodnota jednoho příznaku značně určena hodnotou jiného příznaku a jeden ze dvou korelovaných příznaků může být vyloučen z množiny příznaků, aniž by došlo k velké změně výsledků rozpoznávání [1], [9].

Kovariance $c(i,j)$ dvou příznaků x_i a x_j se dá vypočítat obdobně jako rozptyl v jednorozměrném rozložení. Z nadměrných hodnot používaných příznaků musí být určeny rozdělení příznaků v každé třídě střední hodnoty μ a kovarianční matice C . Podrobnější výpočty viz [1]

Kovariance se nehodí jako objektivní měřítko závislosti příznaků, protože sama závisí na velikosti středních hodnot, jak plyne z jejího vzorce pro výpočet kovarianční matice C_v

$$c_v(i, j) = \frac{1}{K_v} \sum_{k=1}^{K_v} x_c^i(k) x_c^j - \mu_c^i \mu_c^j, \quad (8-18)$$

kde i, j - jsou indexy příznaků (kde platí $1 \leq i, j \leq N$), N - je počet příznaků, K_v - je počet příznakových vektorů zpracovaných ve třídě v , v - značí měrnou třídu a μ_v je střední hodnota příznakového vektoru, která se pro každou rozpoznávanou třídu v určí takto

$$\underline{\mu}_v = \frac{1}{K_v} \sum_{k=1}^{K_v} \underline{x}_v(k) \quad (8-19)$$

Plnou kovarianční matici bychom získali, pokud bychom vypočítali kovariance pro všechny kombinace $i, j = 1, \dots, N$.

Normováním kovarianční matice v rozsahu od -1 od 1 vytvoříme vhodnější kritérium závislosti příznaků, kde na hlavní diagonále budou prvky rovny jedné. Prvky nad diagonálou a pod diagonálou jsou si rovny, zrcadlově podle diagonály. Prvky korelační matice jsou z matice kovarianční získány takto

$$\rho_v(i, j) = \frac{c_v(i, j)}{\sqrt{c_v(i, i) c_v(j, j)}} \quad (8-20)$$

Program Matlab přímo vypočítal matici korelačních koeficientů normovanou od -1 do 1. Kde příznaky blízkí -1 jsou nejméně korelovány a příznaky blízkí se 1 jsou nejvíce korelovány. Matici korelačních koeficientů

nebyla uvedena celá kvůli jejím velkým rozměrům při tak vysokém počtu příznaků. Ale je uvedena její část v Tab. 8.2, aby bylo zřejmé jak má její tvar vypadat.

Tabulka 8.2: Částečné zobrazení matice korelačních koeficientů.

	R	Z_n	F_{0min}	F_{0max}	F_{0str}	F_{0std}	.	.	.
R	1,00	-0,39	-0,31	-0,26	-0,40	-0,16			
Z_n	-0,39	1,00	0,05	0,34	0,20	0,10			
F_{0min}	-0,31	0,05	1,00	0,15	0,81	-0,05	.	.	.
F_{0max}	-0,26	0,34	0,15	1,00	0,40	0,83			
F_{0str}	-0,40	0,20	0,81	0,40	1,00	0,25			
F_{0std}	-0,16	0,10	-0,05	0,83	0,25	1,00	.	.	.
.		
.		
.		

V následujících tabulkách Tab. 8.3 vlevo je uvedeno deset nejvíce a v Tab. 8.3 vpravo deset nejméně korelovaných příznaků.

Tabulka 8.3: Deset nejméně (vlevo) a deset nejvíce (vpravo) korelovaných příznaků.

x_i	x_i	k_i
H_{max}	H_{min}	-0.6107
F_1	I_{str}	-0.5966
F_2	I_{str}	-0.5697
F_{2sp}	I_{str}	-0.5623
F_{1sp}	I_{str}	-0.5441
F_3	R	-0.5023
F_2	R	-0.4859
F_{1sp}	I_{max}	-0.4849
F_1	I_{max}	-0.4782
F_4	R	-0.4674

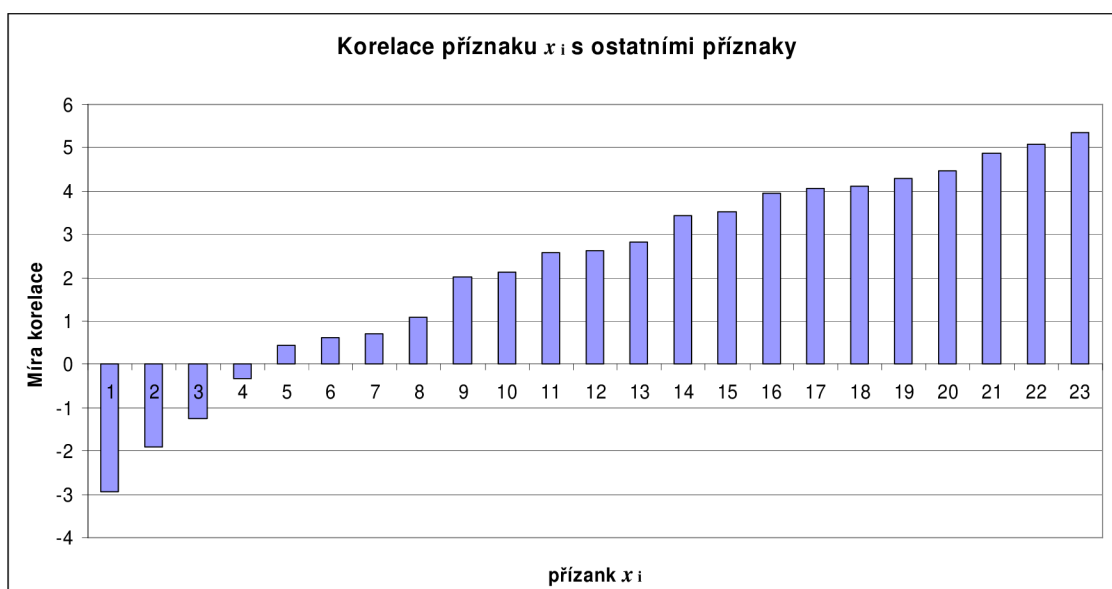
x_i	x_i	k_i
I_{str}	I_{max}	0.9448
F_2	F_1	0.8817
F_3	F_2	0.8401
F_{0std}	F_{0max}	0.8326
F_4	F_3	0.8101
F_{0str}	F_{0min}	0.8062
F_3	F_1	0.7771
F_4	F_2	0.7332
F_{2sp}	F_2	0.7211
F_{2sp}	F_1	0.6753

Dále byla pro bližší představu z matice korelačních koeficientů vytvořena míra korelace k_i příznaku x_i se všemi ostatními příznaky a to tak, že byla provedena suma korelačních koeficientů pro každý sloupec korelační matice, která obsahuje informaci o korelaci daného příznaku x_i se všemi ostatními příznaky. Prvky byly vzestupně seřazeny do Tab. 8.4 podle míry korelace příznaku k_i .

Tabulka 8.4: Míra korelace příznaku k_i s ostatními příznaky.

i	x_i	k_i	i	x_i	k_i	i	x_i	k_i
1	R	-2,9374	9	Z_n	2,0230	17	F_{4sp}	4,0609
2	H_{min}	-1,9089	10	F_{1sp}	2,1200	18	F_{0max}	4,0986
3	I_{str}	-1,2464	11	H_{max}	2,5676	19	F_{2sp}	4,2786
4	I_{max}	-0,3337	12	H_{std}	2,6305	20	F_2	4,4604
5	F_{0sl}	0,4385	13	F_{3sp}	2,8173	21	F_{0str}	4,8845
6	I_{min}	0,6216	14	F_{0std}	3,4243	22	F_3	5,0831
7	H_{str}	0,6967	15	F_{0min}	3,5185	23	F_4	5,3399
8	I_{std}	1,1004	16	F_1	3,9601			

Dále byly hodnoty z tabulky vyneseny do sloupcového grafu Obr. 8.6, kde význam příznaku x_i je uveden v tabulce Tab. 8.4.



Obr. 8.6: Míra korelace příznaku k_i s ostatními příznaky.

Z výsledků je patrné že nejméně korelující příznaky jsou tempo řeči – R a střední hodnota intenzity – I_{str} .

Naopak nejvíce korelujícími příznaky s ostatními příznaky jsou střední hodnota čtvrtého formantu F_4 , střední hodnota třetího formantu F_3 a střední hodnota základního tonu F_{0str} . Je vidět že střední hodnoty i šířky formantových příznaků mezi sebou vzájemně značně korelují, avšak na druhou stranu velmi nízce korelují s ostatními příznaky.

9 Neuronové sítě

Neuronová síť je jedním z výpočetních modelů používaných v umělé inteligenci. První matematický model neuronu vytvořili McCulloch a Pitts v roce 1943 a tento model se dodnes používá pro běžné aplikace. Vzorem pro umělý model neuronové sítě je biologická neuronová síť a chování jejích struktur. Umělá neuronová síť se skládá z umělých (formálních) neuronů. Má široké použití, především tam kde nelze použít známých funkčních algoritmů.

Neurony jsou vzájemně propojeny a navzájem si předávají signály a transformují je pomocí určitých přenosových funkcí. Neuron má libovolný počet vstupů, ale pouze jeden výstup.

Nejčastější použití neuronových sítí je pro rozpoznávání a kompresi obrazů nebo zvuků, předvídání vývoje časových řad (např. burzovních indexů), někdy dokonce k filtrování spamu (nežádoucí pošty). Také umělé neuronové sítě slouží v lékařství k prohlubování znalostí o fungování nervových soustav živých organismů. Například perceptronová síť vznikla původně jako simulace fyziologického modelu rozpoznávání vzorů na sítnici lidského oka [5].

9.1 Výhody a nevýhody neuronových sítí

Výhody:

- Schopnost adaptovat se - není nutné znát algoritmus řešení. Získávání znalostí probíhá učením pomocí množiny vzorů.
- Schopnost generalizace – jsou správně klasifikovány do tříd i vzory, které nejsou součástí tréninkového souboru hodnot.

Nevýhody:

- Značná doba potřebná k natrénování
- Je obtížné zjistit, zda neuronová síť správně generalizuje
- Nutnost zkušeností pro správnou volbu optimální topologie sítě
- Složitost a rozsah sítí

9.2 Formální neuron

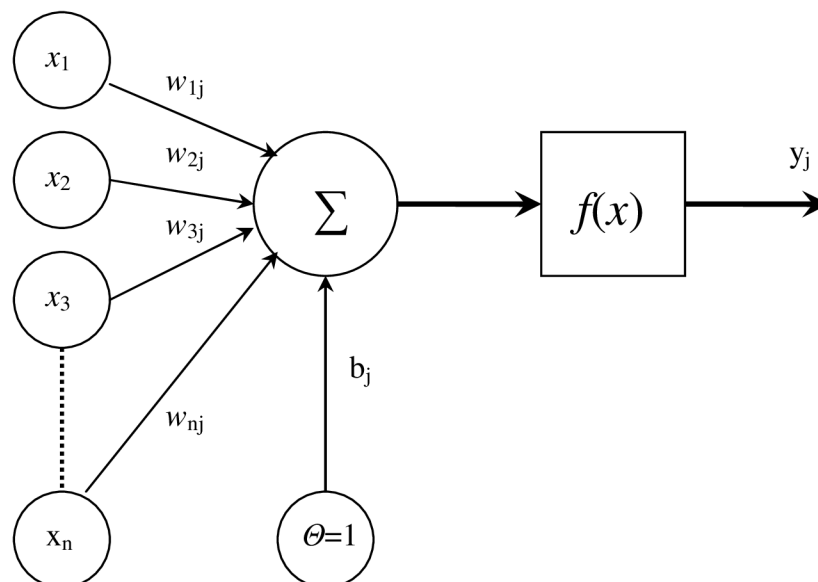
Je základem matematického modelu neuronové sítě. Strukturu popisuje obrázek Obr. 9.1, kde x_1, \dots, x_n jsou obecně reálné vstupy, modelující dendrity, kterých může být jedna až n . Vstupy jsou ohodnoceny reálnými synaptickými váhami w_{1j}, \dots, w_{nj} , které určují propustnost těchto vstupů. θ označuje práh (bias), který představuje zvláštní vstup, který není připojený k výstupu žádného neuronu, ale přivádí konstantní veličinu do neuronu. Tato veličina funguje jako prahová hodnota při aktivování výstupu. Když suma váženého součtu vstupů nepřesahuje prahovou hodnotu, tak se neuron neaktivuje a jeho výstup zůstane nezměněný. $f(x)$ je použitá přenosová funkce neuronu a y_j je výstup neuronu [5], [6], [7].

Potenciál j neuronu představuje vážená suma vstupních hodnot:

$$y_vstup_j = \sum_{i=1}^n w_{ij} x_i. \quad (9-1)$$

Práh může být do vztahu začleněn předáním komponent $\theta = 1$ do vektoru x . Práh je potom zpracováván jako jakákoliv jiná váha, tzn. $w_{0j} = b_j$. Vstup do neuronu Y_j je pak dán vztahem

$$y_vstup_j = w_{0j} + \sum_{i=1}^n x_i w_{ij} = b_j + \sum_{i=1}^n x_i w_{ij}. \quad (9-2)$$



Obr. 9.1: Základní struktura modelu umělého neuronu.

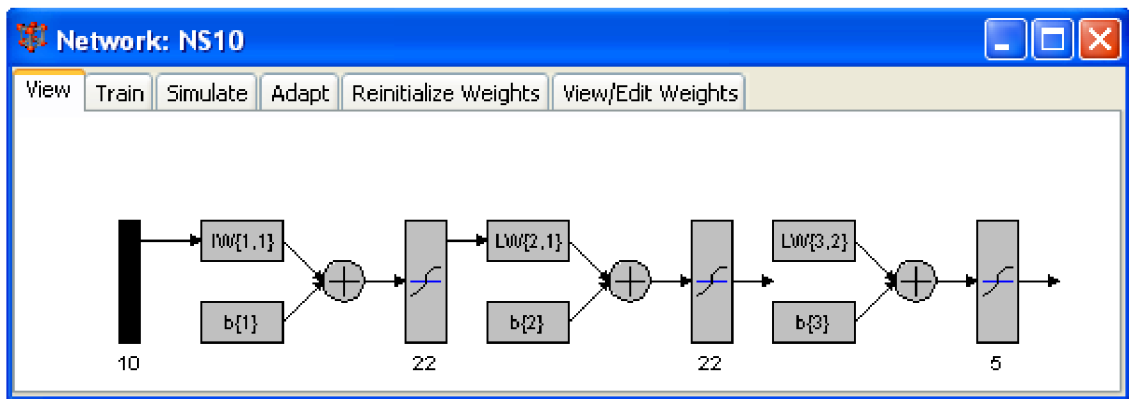
9.3 Základní modely neuronových sítí

Zde jsou uvedeny základní modely neuronových sítí, jež jsou popsány podle [6] a [8].

- Perceptron – je tvořena jednoduchým neuronem nebo jednovrstvou sítí s dopředným šířením. Nepoužívá učitele. Je jednou z nejznámějších a nejpoužívanějších sítí. Aktivační funkcí je jednotkový skok nebo znaménková funkce. Používá se jako klasifikátor pro lineárně separovatelné obrazy.
- Backpropagation – vícevrstvá síť vycházející z perceptronu. Obsahuje algoritmus se zpětným šířením chyby. Aktivační funkce mohou být spojitá, diferencovatelná nebo monotónně klesající.
- Hopfieldova síť – jednovrstvá rekurentní síť s pevnými vahami. Vstupní vektory musí být binární (0/1), popřípadě bipolární (-1/1). Aktivačními funkcemi jsou buď funkce jednotného skoku, nebo znaménkovou funkci. Používá se nejčastěji pro autoasociativní paměti, rekonstrukce neúplných a šumem poškozených obrazů, optimalizační problémy.
- Neuronová síť ART – dvouvrstvá rekurentní síť bez učitele. Vstupní vektor je buď binární nebo reálný. Používá se pro rozpoznávání znaků, řečových segmentů apod. a také pro shlukování.
- Kohonenova neuronová síť – jednovrstvová síť s dopředným šířením. Nepoužívá učitele. Výstupní hodnota neuronů je definována jako vzdálenost mezi vstupním a váhovým vektorem. Používá se pro analýzu dat, shlukování, vytváření sémantických map.

9.4 Parametry navržené neuronové sítě

Blokové schéma jedné z námi použitých neuronových sítí zachycuje Obr. 9.2, kde černý blok vlevo představuje soubor deseti vstupních příznaků převedených do vektoru v v bloku $LW\{1,1\}$. Blok $b\{1\}$ představuje práh a kulatý blok se symbolem plus představuje neurony který mají výstup přes přenosovou funkci. V našem případě je to přenosová funkce typu Hyperbolická tangenta. Dále cesta pokračuje na vstup dalších dvou vrstev.



Obr. 9.2: Blokové schéma jedné z vytvořených sítí s deseti příznaky.

Parametry a nastavení neuronových sítí, které byly použity v naší práci:

Typ modelu sítě: Dopředný perceptron s algoritmem backpropagation

Trénovací funkce: Levenberg-Marquardt backpropagation (Inovované vylepšení algoritmu backpropagation)

Úprava učicí se funkce: Moment vah se sestupným sklonem a s učicí funkcí prahu.

Vykonávající funkce: Minimalizace střední kvadratické chyby.

Počet vrstev: 3,

kde počet neuronů pro jednotlivé vrstvy byl zvolen podle následujícího pravidla.

Počet neuronů N_N se volí zpravidla pro vrstvu:

1. $N_N = N_T \cdot 3 + (2 \div 4)$
2. $N_N = N_T \cdot 3 + (2 \div 4)$ **(9-3)**
3. $N_N = N_T$

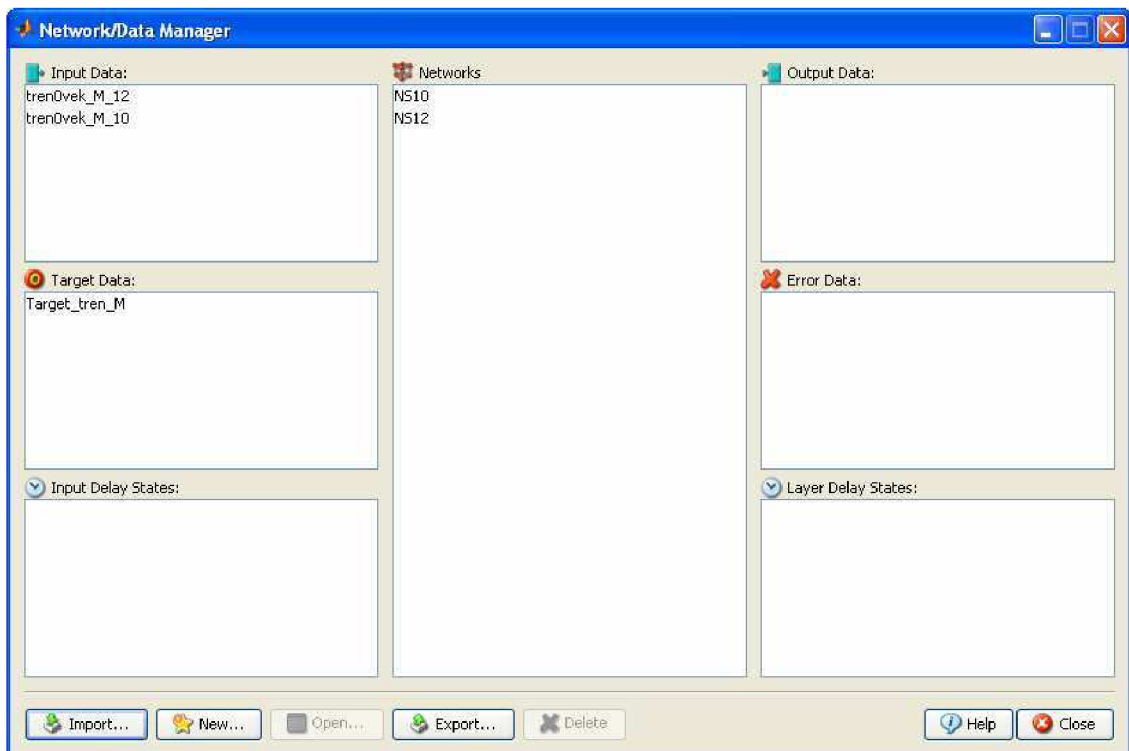
Kde N_T představuje počet tříd.

Pro rozeznávání pohlaví bylo použito 8 neuronů v prvních dvou vrstvách a 2 neurony ve třetí vrstvě. Pro rozeznávání věku bylo použito 22 neuronů v prvních dvou vrstvách, v třetí vrstvě 6 neuronů pro ženy a pouze 5 neuronů pro muže. Poslední třída byla pro rozeznání věku u mužů vynechána, protože v ní byl pouze jeden mluvčí.

U všech tří vrstev byla použita přenosová funkce neuronů typu Hyperbolická tangenta.

10 Návrh neuronových sítí v Neural Network Toolbox

NNToolbox je nástroj programu Matlab pro práci s neuronovými sítěmi. Je celý tvořen grafickým rozhraním, takže práci značně usnadňuje. V Matlabu se spouští příkazem `nntool`. Vstupní data lze do něj importovat jako matici nebo jako pole hodnot, jak jsme to činili my a program si sám data převede na matici. Import je možný z pracovního prostoru Matlabu (Workspace) nebo přímo ze souboru.



Obr. 10.1: Základní okno NN Toolboxu.

Na obrázku Obr. 10.1 je zachycen vzhled NNToolboxu, kde vlevo jsou vidět vstupní data s různým množstvím příznaků a matice s názvem `targetdata`, která určuje do které třídy patří vstupní trénovací vzorek. Obsahuje stejný počet řádků jako je počet tříd a stejný počet sloupců jako je počet trénovacích vzorků ve vstupních datech. V prostřed obrázku je seznam vytvořených neuronových sítí. Nová síť se vytvoří klepnutím na záložku `New...`, a poté můžeme zadat vstupní parametry. Parametry a nastavení které bylo použito v naší práci, jsou blíže specifikovány v kapitole o neuronových sítích.

Trénování sítě vybranými příznaky bylo také provedeno pomocí NNToolboxu. Množství epoch bylo zvoleno tak, aby byla pokud možno dosažena statistická chybovost při trénování menší než 10^{-5} .

11 Postup rozdělení vzorků pro trénování a testování

Promluvy byly rozděleny na dvě skupiny a to na **trénovací** S_{tren} a **testovací** S_{test} , tak aby byl větší počet promluv pro trénování sítě a menší pro testování sítě. Pro rozpoznávání pohlaví bylo rozdělení přibližně v poměru $S_{\text{tren}} - 5:3 - S_{\text{test}}$. Pro rozeznávání věku bylo v rámci možností v každé věkové skupině také dodrženo poměru 5:3. Pro rozeznávání pohlaví byli vynechány věkové skupiny VS1 – VS2, protože především chlapci než se dostanou do období adolescence jsou velmi obtížně rozeznatelní. Mají totiž neustálený hlas a hodnoty příznaků se neodlišují výrazně od dívek. K rozeznávání věku u mužů nebyla použita nejvyšší věková skupina VS6, jelikož v ní byl jen 1 mluvčí. U žen jsou v této věkové skupině pro trénování a testování pouze po jedné mluvčí. Je pravděpodobné že rozeznání této věkové skupiny nebude správné, protože počet mluvčích pro trénování je nízký.

V jednotlivých skupinách byl použit počet promluv pro trénování a testování pro rozpoznávání věku i pohlaví vyjádřený v tabulce Tab 11.1.

Tabulka 11.1: Použité množství promluv pro trénování a testování NS.

Pohlaví		S_{tren}	186
		S_{test}	108
Věk	Ženy	S_{tren}	120
		S_{test}	72
	Muži	S_{tren}	144
		S_{test}	90

V tabulce S_{tren} znamená trénovací skupinu a S_{test} testovací skupinu.

12 Proces klasifikace

Pro trénování jsme použili, jak bylo popsáno již dříve Neural Network Toolbox v programu Matlab. Pro trénování a testování neuronové sítě bylo vybráno podle kritéria kvality Q dvanáct nejvhodnějších příznaků které jsou zaneseny sestupně do Tab.12.1 na základě kritéria kvality pro daný účel rozpoznávání.

Tabulka 12.1: Dvanáct nejvhodnějších příznaků vybraných podle Q

Q	1	2	3	4	5	6	7	8	9	10	11	12
Pohlaví	F_{0str}	F_{4str}	F_{0min}	F_{4sp}	F_{3str}	H_{str}	F_{3sp}	F_{2str}	R	H_{std}	I_{str}	F_{2sp}
Věk	F_{0str}	F_{0min}	F_{4str}	F_{3str}	F_{1sp}	R	F_{2sp}	F_{1str}	I_{str}	F_{2str}	I_{max}	H_{str}

Trénování a následné testování s těmito příznaky bylo prováděno tak, že nejprve byly uvažovány všechny příznaky a postupně bylo ubíráno po jednom příznaku s nejhorším kritériem kvality Q , kdy s tímto novým souborem příznaků byla natrénována a otestována nová síť. Příznaky byli ubírány až do počtu jednoho příznaku, kdy jím byl F_{0str} . Při učení sítě s menším počtem příznaků než šest výrazně stoupá chybovost učení neuronové sítě u které je dobré aby byla v hodnotách nižších než 10^{-4} .

Testování toho jak dobře má neuronová síť naučeny vzory a jak je zařazuje do správných tříd jsme prováděli pomocí vhodného zobrazení, které nám výsledek zapsalo do tzv. matice zmatku, která se v mnoha literaturách uvádí jako „Confusion matrix“. Pro rozpoznávání pohlaví měla matice rozměr 2x2 a mohla vypadat například takto:

$$\begin{pmatrix} 83 & 17 \\ 27 & 73 \end{pmatrix},$$

kde prvky na hlavní diagonále znamenají procentuálně vyjádřený výsledek správného rozpoznání neuronové sítě. Na prvním řádku matice jsou výsledky rozpoznání žen a na druhém řádku pro rozpoznání mužů. Prvky mimo hlavní diagonály jsou procentuelním vyjádřením špatně rozeznávaných mluvčích ve své kategorii.

Procentuálně vyjádřené výsledky správného zařazení do daných tříd byly zaneseny do Tab. 12.2 (vlevo) a poté do grafu Obr. 12.1. Jsou zde zahrnuty všechny věkové skupiny. V tabulce Tab. 12.2 (vpravo) a grafu Obr. 12.2 jsou výsledky, kde pro trénování a testování byli vynecháni mluvčí z věkových skupiny VS1 a VS2 z důvodů popsaných v předchozí kapitole.

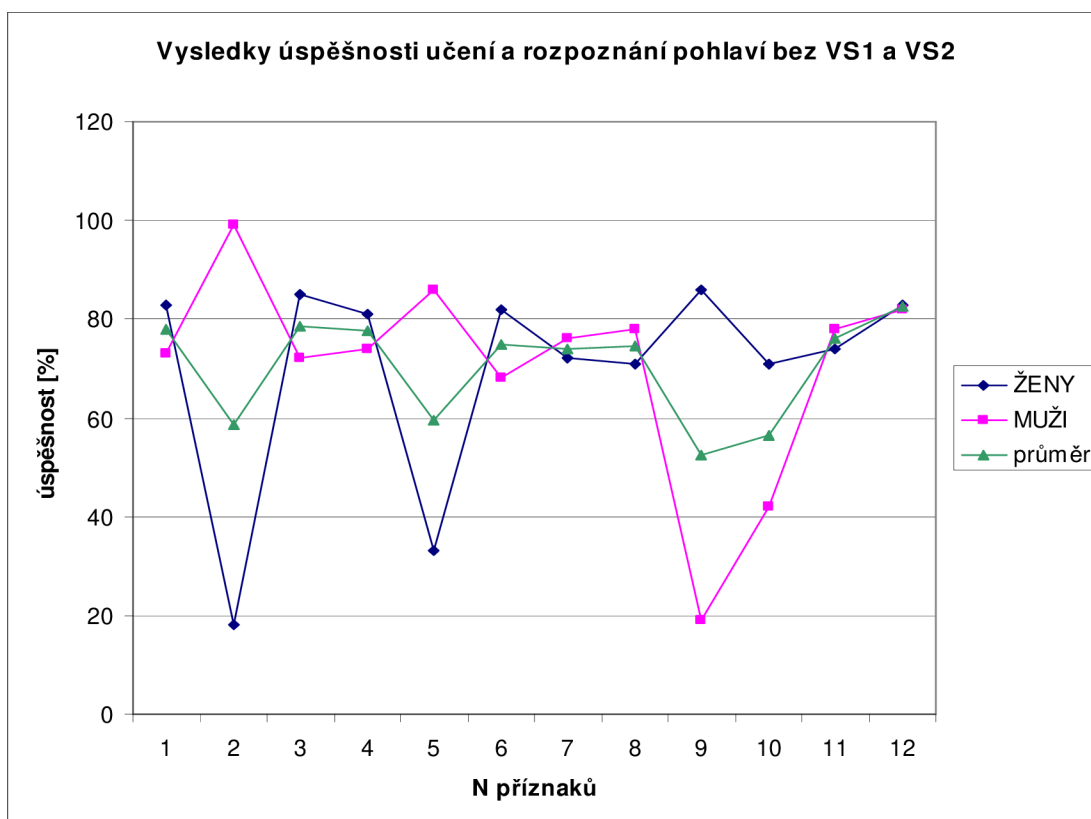
Výsledky pro rozpoznávání pohlaví:

Tabulka 12.2: Výsledky správného zařazení do tříd pro všechny věkové skupiny (vlevo) a s vynecháním prvních dvou věkových skupin (vpravo).

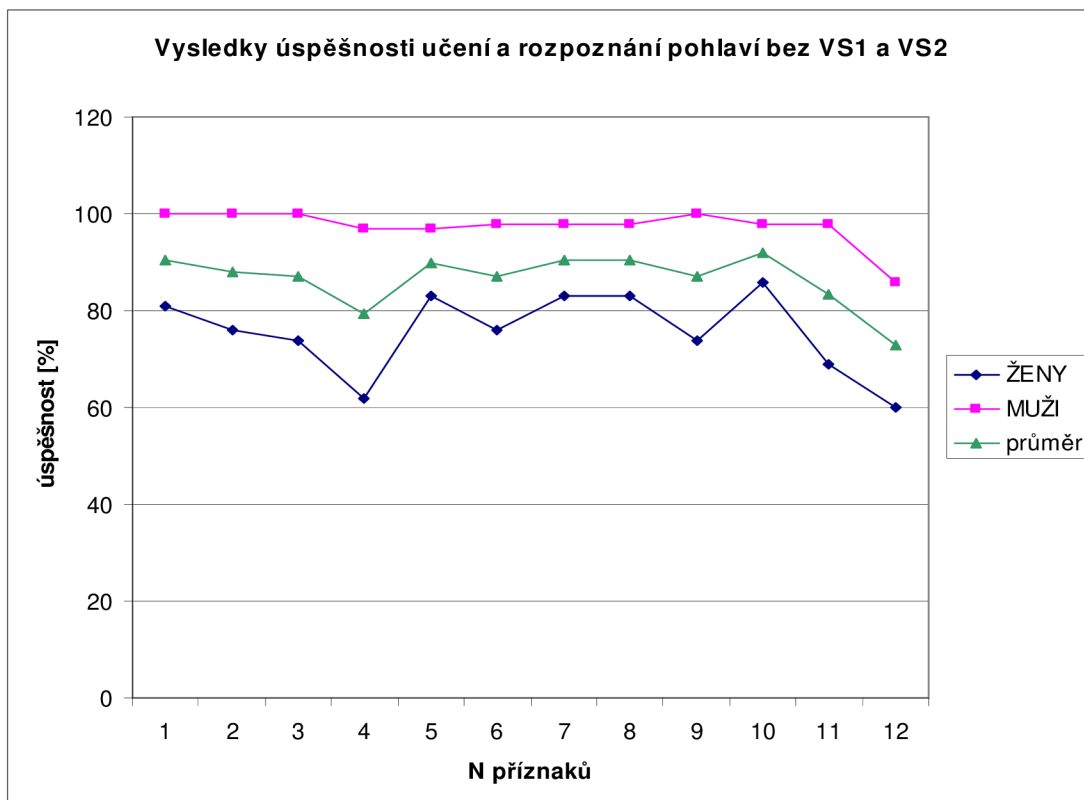
Počet použitých příznaků	Ženy	Muži	Průměr
1	83	73	78
2	18	99	58,5
3	85	72	78,5
4	81	74	77,5
5	33	86	59,5
6	82	68	75
7	72	76	74
8	71	78	74,5
9	86	19	52,5
10	71	42	56,5
11	74	78	76
12	83	82	82,5

Počet použitých příznaků	Ženy	Muži	Průměr
1	81	100	90,5
2	76	100	88
3	74	100	87
4	62	97	79,5
5	83	97	90
6	76	98	87
7	83	98	90,5
8	83	98	90,5
9	74	100	87
10	86	98	92
11	69	98	83,5
12	60	86	73

Podle střední hodnoty správného zařazení do třídy pro muže a ženy byly v tabulce Tab. 12.2 označeny tři soubory příznaků, s nimiž bylo dosaženo nejlepších výsledků.



Obr. 12.1: Výsledky správného zařazení do tříd pro všechny věkové skupiny.



Obr. 12.2: Výsledky správného zařazení do tříd s vynecháním prvních dvou věkových skupin.

Z grafů Obr. 12.1 a Obr. 12.2 je patrné, že při vynechání prvních dvou věkových skupin dochází mnohem více ke správné klasifikaci při různých počtech příznaků jak pro ženy, tak pro muže.

Pro rozpoznávání věku měla matice rozměr 5x5 pro muže a 6x6 pro ženy. Samotné výsledné matice jsou uvedeny v Tab. 12.3 pro muže a v Tab.12.4 pro ženy. Popis bude proveden pro větší názornost jen na matici typu 5x5 pro muže. Pro větší matici platí stejný princip. Každý řádek zahrnuje pět tříd, kde uvedená čísla jsou procentuelní vyjádření, do které třídy neuronová síť testované promluvy zařadila. Ke správnému zařazení do dané třídy došlo pokud je nejvyšší hodnota na hlavní diagonále. Ideální stav pro všechny třídy je hodnota 100 po celé hlavní diagonále.

Výsledky pro rozpoznávání věku

Tabulka 12.3: Výsledky pro rozeznávání věku u mužů.

M1	100	0	0	0	0	M7	67	33	0	0	0
	0	44	50	6	0		17	22	28	33	0
	0	14	69	17	0		0	22	69	3	6
	0	0	78	22	0		0	6	44	50	0
	0	17	75	8	0		0	8	33	50	8
M2	67	0	33	0	0	M8	50	33	0	17	0
	44	39	17	0	0		0	33	39	28	0
	0	19	72	3	6		0	25	61	8	6
	0	17	61	17	6		0	0	78	22	0
	0	8	67	0	25		0	0	50	42	8
M3	67	17	17	0	0	M9	67	17	17	0	0
	28	44	11	17	0		33	28	28	6	6
	0	25	64	3	8		0	28	50	11	11
	0	11	22	67	0		0	0	28	67	6
	0	42	33	25	0		0	0	92	0	8
M4	67	17	17	0	0	M10	67	17	17	0	0
	0	94	0	6	0		44	11	44	0	0
	0	28	58	6	8		0	25	61	14	0
	0	6	28	56	11		0	11	11	78	0
	0	0	33	33	33		0	0	33	58	8
M5	100	0	0	0	0	M11	17	83	0	0	0
	100	0	0	0	0		6	50	22	22	0
	86	0	8	3	3		0	28	64	6	3
	56	22	11	11	0		0	0	11	83	6
	42	17	0	42	0		0	0	8	83	8
M6	50	17	17	17	0	M12	50	17	17	0	17
	6	50	17	22	6		17	33	28	22	0
	0	17	72	11	0		0	14	67	19	0
	0	22	33	17	28		0	0	50	50	0
	0	17	8	25	50		0	0	50	50	0

Tabulka 12.4: Výsledky pro rozeznávání věku u žen.

Z1	83	4	4	4	4	0	Z7	63	8	4	17	8	0
	17	0	0	67	0	17		17	0	17	50	17	0
	17	0	33	50	0	0		33	50	0	0	17	0
	0	0	8	25	8	58		17	0	0	67	17	0
	17	0	17	50	17	0		8	25	0	8	58	0
	17	0	17	8	8	50		17	17	17	25	25	0
Z2	42	33	0	8	17	0	Z8	54	0	17	25	4	0
	33	0	0	0	67	0		67	0	0	33	0	0
	50	0	0	17	33	0		33	0	17	50	0	0
	8	0	8	0	33	50		8	0	0	42	0	50
	17	0	0	33	42	8		42	0	33	25	0	0
	25	17	25	0	17	17		33	0	17	33	8	8
Z3	42	8	0	25	25	0	Z9	71	8	0	17	4	0
	17	0	0	0	83	0		50	17	33	0	0	0
	17	0	0	33	50	0		67	17	17	0	0	0
	50	0	8	8	33	0		0	0	0	33	8	58
	17	0	0	42	33	8		17	17	25	25	17	0
	67	0	8	25	0	0		8	0	42	42	0	8
Z4	42	50	0	8	0	0	Z10	75	0	0	17	8	0
	0	50	0	17	33	0		67	17	17	0	0	0
	0	83	0	17	0	0		50	0	50	0	0	0
	0	25	0	17	17	42		0	0	0	42	0	58
	0	75	0	17	8	0		17	0	33	17	33	0
	0	67	8	8	17	0		17	0	17	25	0	42
Z5	42	50	0	8	0	0	Z11	63	13	0	21	4	0
	0	50	0	17	33	0		50	33	17	0	0	0
	0	83	0	17	0	0		83	0	17	0	0	0
	0	25	0	17	17	42		0	0	0	67	25	8
	0	75	0	17	8	0		33	0	33	17	8	8
	0	67	8	8	17	0		0	8	50	25	0	17
Z6	42	0	8	33	17	0	Z12	63	8	0	13	17	0
	17	17	17	50	0	0		50	50	0	0	0	0
	0	17	0	83	0	0		33	0	67	0	0	0
	0	0	0	58	0	42		0	0	0	50	0	50
	0	0	0	67	33	0		17	8	58	0	0	17
	17	0	33	42	0	8		50	0	33	17	0	0

Při rozpoznávání věkové skupiny byly dále výsledky zpracovány tak, že jsme určili z krajních hodnot jednotlivých skupin střední hodnoty. Vznikala nám tak pomyslná střední hodnota skutečného věku. Například VS1 má věkový rozsah 7-12 let. Takže z hodnot 7 a 12 jsme udělali aritmetický průměr a vyšla hodnota 9,5. Stejně tak bylo postupováno v ostatních věkových skupinách, kde výsledky jsou v Tab. 12.5.

Tabulka 12.5: Střední hodnoty věkových skupin.

Věková skupina	Střední hodnota
VS1	9.5
VS2	14
VS3	18
VS4	25.5
VS5	40
VS6	60

Dále byl vypočten aritmetický průměr a směrodatná odchylka hodnot jež vyjadřují věkovou vzdálenost klasifikované třídy od skutečného věku popsaného v Tab. 12.6 pro různé počty příznaků jak pro muže, tak pro ženy.

Tabulka 12.6: Střední hodnoty a směrodatné odchyly vzdálenosti správné klasifikace.

N příznaků	X_{STR}	S		N příznaků	X_{STR}	S
M1	5,38	6,55		Z1	13,57	13,60
M2	5,56	6,71		Z2	18,13	15,08
M3	6,11	7,59		Z3	19,80	15,33
M4	5,15	6,40		Z4	17,30	15,42
M5	8,69	7,36		Z5	19,64	19,10
M6	6,28	6,78		Z6	16,38	14,53
M7	5,47	6,40		Z7	13,60	13,78
M8	5,84	6,35		Z8	18,13	15,07
M9	6,61	7,48		Z9	16,42	15,77
M10	4,69	5,21		Z10	13,98	14,87
M11	4,84	5,11		Z11	14,81	14,92
M12	5,78	6,15		Z12	18,69	17,44

V Tab 12.6 písmeno M označuje muže a písmeno Z ženy. Číslo za těmito písmeny označuje počet použitých příznaků pro klasifikaci, X_{STR} je střední hodnota vzdáleností klasifikované věkové skupiny od skutečného věku.

s vyjadřuje směrodatnou odchylku hodnot vyjadřujících vzdálenost klasifikovaného věku od skutečného.

13 Závěr

V části která byla obhájena jako semestrální projekt jsou rozebrány a vysvětleny některé základní vlastnosti řečových signálů a jejich souvislosti s věkem, biologickým i psychickým vývojem člověka. Je zde popsán vznik řečových signálů v lidském těle. V projektu jsou stručně rozepsány dějiny zkoumání řeči a její umělé vytváření (syntéza). Pak je zde popsáno rozdělení do oblastí číslicového zpracování řečových signálů, způsoby zobrazování řečových signálů a práce s nimi. Je zde popsán program Praat, který byl stěžejním nástrojem pro analýzu nahrávek. Práce popisuje a odůvodňuje způsob rozdělení databáze na věkové skupiny, způsob získávání vzorků a práci s mluvčími. Jsou zde popsány možné souvislosti řečového signálu s biologií a návyky člověka, které by mohly řečový signál ovlivnit, jako například, zda-li je mluvčí kuřák. Hlavní zaměření projektu je zkoumání parametrů a příznaků řečových signálů v souvislosti s věkem a pohlavím mluvčích. V práci jsou popsány základní zkoumané příznaky řečových signálů, jako je například základní tón a intenzita signálu.

Semestrální projekt obsahuje výsledky analýzy rozdělené podle pohlaví mluvčích a popisuje rozdíly hodnot ve věkových skupinách. Grafické závislosti zobrazují výsledky analýzy zkoumané věty „Hana tam nebude“. Jedním z nejzajímavějších zkoumaných příznaků je závislost tempa řeči na věku a pohlaví mluvčích. A to především proto, že byla hodnota tempa očekávána vyšší u žen, byla vyšší u mužů. Hodnoty frekvence základního tónu vyšly podle očekávání vyšší u žen.

V bakalářské práci byla dokončena analýza příznaků ze semestrálního projektu pro všechny věty a bylo využito poznatků z obhájeného semestrálního projektu. Byly vybrány vhodné příznaky jak pro rozpoznávání pohlaví, tak i věku podle kritéria kvality. Vybraných příznaků bylo využito pro klasifikaci s využitím klasifikátoru. Jako klasifikátor byla zvolena neuronová síť. Byly popsány základní typy neuronových sítí a parametry námi zvolené neuronové sítě. Jako nástroj pro návrh neuronové sítě byl zvolen Neural Network Toolbox z programu Matlab. Vzorky byly rozděleny na trénovací, pomocí nichž byla síť trénována a na testovací, pomocí nichž bylo testováno, jak úspěšně síť

klasifikovala mluví do tříd podle pohlaví a zvláště pak do tříd podle věku. Pro testování byly použity promluvy od jiných mluvčích než pro trénování.

Výsledky byly vyneseny do tabulek, kde tři nejlépe klasifikované soubory příznaků byly zvýrazněny a následně zaneseny do grafů. Nejlépe síť klasifikovala mluví do tříd podle pohlaví s 83% úspěšností pro ženy a s 82% úspěšností pro ženy při použití souboru dvanácti příznaků. Mnohem lepších výsledků klasifikace bylo dosaženo po odstranění prvních dvou věkových skupin, které obsahovaly mluví ve věkovém rozsahu sedm až patnáct let. Úspěšnost klasifikace pro ženy se takřka nezměnila, pro muže se však téměř u všech sítí s různým počtem příznaků blížila stu procent. Takto patrný rozdíl je způsoben tím, že se u chlapců hlas vyvíjí a mění do šestnácti let velmi výrazně. Hodnoty příznaků mužů jsou pak velmi obtížně rozeznatelné od hodnot příznaků pro ženy.

Při klasifikaci mluví do věkových skupin byla pro muže největší úspěšnost při použití souboru jedenácti příznaků. Bylo dosaženo průměrné odchylky klasifikovaných věkových skupin od skutečných věků mluví 4,83 let. Směrodatná odchylka při použití sítě s těmito příznaky byla 5.11. Při klasifikaci žen do věkových skupin byla úspěšnost správného zařazování výrazně nižší. Nejmenší průměrná věková vzdálenost klasifikovaného věku od skutečného byla 13,57 let, se směrodatnou odchylkou 13,6 při použití jednoho příznaku, a to střední hodnoty základního tónu.

Jako celkové zhodnocení analýzy lze říct, že bylo dosaženo uspokojivých výsledků při zkoumání jednotlivých příznaků. Při samotném rozpoznávání pohlaví bylo dosaženo také více než uspokojivých výsledků. Při rozpoznávání věku byly výsledky lepší, než bylo předpokládáno vzhledem k tomu, že úloha správné klasifikace věku není v současnosti v oblasti zpracování řečových signálů považována za snadnou. Používá se proto při ní druhotných příznaků. My jsme se ale chtěli v této práci zaměřit na suprasegmentální příznaky.

Lze říci, že výrazně lepších výsledků při klasifikaci jak pohlaví, tak i věku, bylo dosaženo pro muže. Horší výsledky při klasifikaci žen mohou být způsobeny nedostatečně vysokým počtem ženských mluví v některých věkových skupinách.

14 Použitá literatura

- [1] SIGMUND, Milan. Rozpoznávání řečových signálů : přednášky. 1. vyd. Brno : MJ servis s.r.o., 2007. 86 s. ISBN 978-80-214-3526-1.
- [2] BOERSMA, Paul, WEENINK, David. Praat: doing phonetics by computer [online]. 2003- , December 10, 2007 [cit. 2007-12-15]. Dostupný z WWW: < <http://www.praat.org>>. <Http://www.fon.hum.uva.nl/praat/>.
- [3] NOVÁKOVÁ , Iva. Moje rodina a já : Vývoj řeči [online]. 2007. c2007- , Sunday, 26 August 2007 [cit. 2007-12-12]. Dostupný z WWW: <http://www.mojerodinaaja.cz/index.php?option=com_content&task=view&id=118&Itemid=143>.
- [4] KOPEČEK, Ivan. Pro mé studenty [online]. 2006 [cit. 2007-12-15]. Dostupný z WWW: <<http://www.fi.muni.cz/~kopecek/stud.htm>>.
- [5] Fonetika [online]. 2007 [cit. 2007-12-12]. Dostupný z WWW: <<http://is.muni.cz/elportal/estud/ff/js07/fonetika>>.
- [6] Neuronová síť [online]. c2000-2002 , 23. 4. 2008 [cit. 2008-05-29]. Dostupný z WWW: <http://cs.wikipedia.org/wiki/Neuronová_síť>.
- [7] MOLNÁR, Karol. Úvod do problematiky umělých neuronových sítí [online]. 2000 , 22.2.2000 [cit. 2008-05-30]. Dostupný z WWW: <http://www.elektrorevue.cz/clanky/00013/index.html>.
- [8] VESELOVSKÝ, Michal. Neuronové sítě [online]. [1999] [cit. 2008-05-30]. Dostupný z WWW: <<http://www.avari.cz/uir/index.php>>.
- [9] SMÉKAL, Zdeněk. Číslicové zpracování signálů. FEKT VUT Brno, 2005, elektronické texty.

POPISNÝ SOUBOR ZÁVĚREČNÉ PRÁCE

Autor: Josef Báňa

Název závěrečné práce: Porovnání analýzy řečového signálu v závislosti na věku a pohlaví mluvčího

Název závěrečné práce ENG: Comparison of analysis of speech in dependence on age and gender

Anotace závěrečné práce: Obsah této práce je zaměřen na analýzu věku a pohlaví. Dále pak byli zkoumány vhodnosti příznaků a vhodnost jejich použití při klasifikaci mluvčích. Je v ní pojednáno o základní teorii řečového signálu a problematice zobrazení a zpracování řeči.

Je zde popsán volně dostupný program Praat, kterým se používá pro řečovou analýzu. Zaměřili jsme se na suprasegmentální příznaky řeči.

Dále je zde popsána teorie příznaků které byly pro práci zvoleny.

Prvním z kroků v této práci bylo získání dostatečného počtu promluv (nahrávání) od mluvčích různého věku a pohlaví. Řečový korpus je poměrně obsáhlý. Mluvčí byli rozděleni do šesti věkových skupin.

Dále byly nahrávky zpracovány v programu Praat. Průměrné hodnoty byly vyneseny do tabulek a sloupcových grafů pro větší přehlednost.

Dále bylo vybráno dvanáct nejvhodnějších příznaků podle kritéria kvality. S nimi poté byla provedena analýza zvolených příznaků pro automatické rozpoznávání pohlaví a věku. Příznaky byly při testování po jednom ubírány.

Jako klasifikátor pro rozpoznávání byla zvolena neuronová síť. Pro práci se sítí byl zvolen Neural Network Toolbox v programu Matlab.

Vytvořené sítě byly použity pro klasifikaci mluvčích na základě věku a pohlaví. Výsledky byly rozebrány v závěru práce.

Anotace závěrečné práce ENG: This thesis deals with analysis of speech signal in dependence on the gender and the age of the speaker. We tried to investigate through the features to find the best set for the automatic classification of speakers. It also contains a brief discussion about the speech signal and its

POPISNÝ SOUBOR ZÁVĚREČNÉ PRÁCE

characteristics.

We used a program called Praat for the speech analysis purpose. This program is also described in this work. We mainly focused on the suprasegmental features of speech.

Our first step was to make our own speech corpus which should contain speech records from speakers with various age and gender. We made the analysis using Praat and reported it within this thesis.

For the automatic classification purpose, twelve features were selected basing on there quality criteria and used with a neural network to classify the speakers to classes with different age and gender.

As it was mentioned, a neural network was used as a classifier. We used “Neural Network Toolbox” in the Matlab program to create and train our networks.

Klíčová slova: Řeč, korpus, analýza, příznaky, klasifikace, mluvčí, věk, pohlaví, Praat, neuronová síť.

Klíčová slova ENG: Speech, corpus, analysis, features, classification, speaker, age, gender, Praat, neural network.

Typ závěrečné práce: bakalářská práce

Datový formát elektronické verze: pdf

Jazyk závěrečné práce: čeština

Přidělovaný titul: Bc.

Vedoucí závěrečné práce: Ing. Hicham Atassi

Škola: Vysoké učení technické v Brně

Fakulta: Fakulta elektrotechniky a komunikačních technologií

Ústav / ateliér: Ústav telekomunikací

Studijní program: Elektrotechnika, elektronika, komunikační a řídicí technika

Studijní obor: Teleinformatika