

Bakalářská práce



České
vysoké
učení technické
v Praze

FEL

Fakulta elektrotechnická
Katedra kybernetiky

Věková závislost akustických charakteristik řeči v opakovaných slabikách

Tereza Lukášová

Vedoucí práce: prof. Ing. Roman Čmejla, CSc.
Leden 2018

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Lukášová** Jméno: **Tereza** Osobní číslo: **420327**
Fakulta/ústav: **Fakulta elektrotechnická**
Zadávající katedra/ústav: **Katedra kybernetiky**
Studijní program: **Kybernetika a robotika**
Studijní obor: **Robotika**

II. ÚDAJE K BAKALÁŘSKÉ PRÁCI

Název bakalářské práce:

Věková závislost akustických charakteristik řeči v opakovaných slabikách

Název bakalářské práce anglicky:

Age Dependence of Acoustic Speech Characteristics in Repeated Syllables

Pokyny pro vypracování:

1. Proveďte rešerši literatury pojednávající o věkové závislosti exploziv a samohlásek v řeči dospělých a na základě rešerše literatury vyberte vhodné akustické parametry.
2. Na databázi slabik /pa/-/pa/-/pa/ proveďte akusticko-fonetické analýzy se zaměřením na věkovou závislost.
3. Posuďte možnosti klasifikace a realizujte experiment rozpoznávání věku ze slabik /pa/-/pa/-/pa/ na základě akustických charakteristik.
4. Analýzu proveďte v prostředích Praat a MATLAB a klasifikační experiment v MATLABu.

Seznam doporučené literatury:

- [1] Roman Čmejla, Jarmila Behenská, Tereza Tykalová a Jan Rusz: Dlouhodobá studie věkově závislých akustických charakteristik řeči. Akustické listy, ročník 22, číslo 1-4, 3-8.
[2] Roman Čmejla, Jan Rusz a Hana Růžičková: Oslabení okluzí v promluvách pacientů s Parkinsonovou chorobou. Akustické listy, ročník 15, číslo 3-4, str. 12-18.
[3] Jan Janda: Posuzování logopedického věku dítěte. PhD práce, ČVUT FEL 2012.

Jméno a pracoviště vedoucí(ho) bakalářské práce:

prof. Ing. Roman Čmejla, CSc., katedra teorie obvodů FEL

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) bakalářské práce:

Datum zadání bakalářské práce: **20.09.2017**

Termín odevzdání bakalářské práce: **09.01.2018**

Platnost zadání bakalářské práce: **17.02.2019**

prof. Ing. Roman Čmejla, CSc.
podpis vedoucí(ho) práce

doc. Ing. Tomáš Svoboda, Ph.D.
podpis vedoucí(ho) ústavu/katedry

prof. Ing. Pavel Ripka, CSc.
podpis děkana(ky)

III. PŘEVZETÍ ZADÁNÍ

Studentka bere na vědomí, že je povinna vypracovat bakalářskou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v bakalářské práci.

Datum převzetí zadání

Podpis studentky

/ Prohlášení

Prohlašuji, že jsem předloženou práci vypracovala samostatně a že jsem uvedla veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

.....
V Praze dne

.....
Podpis autora práce

Abstrakt / Abstract

Tato práce se zabývá analýzou možných závislostí mezi akustickými charakteristikami řeči a věkem mluvčího. Přítomnost či absence závislosti je zkoumána pro hodnoty charakteristik, které byly získány z nahrávek řečové databáze, které obsahují opakovanou slabiku /pa/.

V práci je nejprve shrnuto základní pojmosloví dané problematiky a očekávané akustické vlastnosti zkoumaných zvukových nahrávek. Dále je navržen detektor, umožňující automatické nalezení znělých úseků nahrávek použité databáze. Z takto nalezených úseků je možné získat hodnoty následujících charakteristik: základní hlasové frekvence, jitteru, shimmeru, HNR, formantů a spektrálních momentů. Po zjištění hodnot parametrů je diskutována jejich závislost na věku mluvčího. Nakonec je proveden pokus klasifikace věku mluvčího pomocí dříve získaných hodnot parametrů.

Detekce znělých úseků je implementována pomocí programu MATLAB. Hodnoty jednotlivých akustických parametrů jsou získávány z řečového signálu za pomoci programu MATLAB a programu Praat.

Klíčová slova: řečový signál, věková závislost, spektrum signálu, fundamentální hlasová frekvence, hlasivky, jitter, shimmer, spektrální momenty, formantové frekvence

This paper aims to analyse possible dependencies of acoustic characteristics of speech and the age of the speaker. The (non)existence of the dependencies will be analysed using characteristics gathered from a database containing records of repeatedly spoken syllable /pa/.

Firstly we will summarise the domain terminology and assumptions about acoustic properties of the records in the used database. Then we will propose a detector for automatic detection of sonant parts of the records. From these sonant parts we will be able to extract values of these characteristics: fundamental voice frequency, shimmer, jitter, HNR, formants and spectral moments. After we extract the values we will analyse and discuss their dependency on the speaker's age. Finally, we will perform conduct an experiment attempting to classify speakers by their age using previously gathered parameter values.

Detection of sonant parts will be implemented using MATLAB. Values of acoustic parameters will be extracted from the records using MATLAB and Praat.

Keywords: speech signal, age dependence, signal spectrum, fundamental voice frequency, vocal cords, jitter, shimmer, spectral moments, formant frequencies

Title translation: Age Dependence of Acoustic Speech Characteristics in Repeated Syllables

/ Obsah

1 Úvod	1
2 Řečová databáze	2
3 Vlastnosti mluvené řeči	4
3.1 Reprezentace řeči	4
3.2 Proces utváření řeči	5
3.3 Vlastnosti samohlásek	5
3.4 Vlastnosti souhlásek	6
3.4.1 Vlastnosti souhlásky /p/ ..	6
4 Detekce znělých úseků v nahrávkách	8
4.1 Detekce pomocí krátkodobé energie signálu	9
4.2 Detekce pomocí spektra	10
5 Akustické parametry z oblasti fonace	13
5.1 Frekvence základního hlasivkového tónu	13
5.1.1 F0 v programu MATLAB	14
5.1.2 F0 v programu Praat	17
5.1.3 F0 závěr	17
5.2 Shimmer	20
5.3 Jitter	22
5.4 HNR	23
6 Akustické parametry z oblasti artikulace: formanty	25
7 Spektrální momenty	30
7.1 Spektrální momenty neznělé explozivy /p/	31
7.2 Spektrální momenty znělé části signálu	34
8 Klasifikace věku mluvčího na základě jeho akustických charakteristik	37
9 Závěr	41
Literatura	42
A Zkratky a symboly	45
B Grafy	46
C Doplnující ilustrace	49

Tabulky / Obrázky

<p>3.1. Tabulka rozdělení souhlásek6</p> <p>5.1. Intervaly předpokládaných hodnot F0 získaných z kepstra pro muže a ženy 15</p> <p>5.2. Intervaly předpokládaných hodnot F0 získaných pomocí autokorelace pro muže a ženy . 16</p> <p>5.3. Průměrné výsledné hodnoty F0 18</p> <p>5.4. Korelace F0 a věku mluvčích .. 19</p> <p>5.5. Korelace F0 a věku žen 19</p> <p>5.6. Test významnosti závislosti F0 na věku žen 20</p> <p>5.7. Korelace Shimmeru vůči věku, děleno dle nahrávky <i>a</i> a <i>b</i> 21</p> <p>5.8. Korelace Shimmeru vůči věku . 21</p> <p>5.9. Korelace Shimmeru vůči věku ženy 21</p> <p>5.10. Korelace jitteru u mužů i žen.. 23</p> <p>5.11. Korelace jitteru žen, děleno dle věku..... 23</p> <p>5.12. Korelace hodnot HNR s věkem mužů a žen 23</p> <p>5.13. Míra korelace HNR s věkem ženy 24</p> <p>6.1. Korelace F1,F2,F3 a F1·F2 vůči věku mluvčích..... 26</p> <p>6.2. Korelace F1 a F2..... 27</p> <p>7.1. Korelace ST a ostatních momentů 32</p> <p>7.2. Korelace spektrálních momentů a věku mluvčího 32</p> <p>7.3. Korelace spektrálních momentů ženy menopauza 33</p> <p>7.4. Korelace spektrálních momentů a F0 34</p> <p>7.5. Korelace spektrálních momentů a formantů 35</p> <p>7.6. Korelace spektrálních momentů ze znělého signálu a věku mluvčích 35</p> <p>8.1. Úspěšnost pokusu klasifikace a 40</p> <p>8.2. Úspěšnost pokusu klasifikace b 40</p>	<p>2.1. Histogram databáze, 5 let2</p> <p>3.1. Plocha slyšení akustických signálů4</p> <p>3.2. Spektrogram českých samohlásek.....6</p> <p>3.3. Spektrogram slabiky /pa/7</p> <p>4.1. Detekce znělých úseků krátkodobou energií 10</p> <p>4.2. Neúspěšná detekce znělých úseků krátkodobou energií..... 10</p> <p>4.3. Hrubá segmentace pomocí spektra..... 11</p> <p>4.4. Hrubá segmentace pomocí spektra a podmínek 12</p> <p>4.5. Hrubá segmentace pomocí spektra na zašuměném signálu 12</p> <p>5.1. Střední hodnoty a směrodatné odchylky F0 získané z kepstra 16</p> <p>5.2. F0 získané z kepstra ve všech slabikách proluvy F46HC357a . 16</p> <p>5.3. F0 získané z kepstra ve všech slabikách proluvy M27HC504 . 16</p> <p>5.4. Střední hodnoty a směrodatné odchylky F0 získané pomocí autokorelace 17</p> <p>5.5. Hodnoty F0 mužů a žen získané programem Praat 17</p> <p>5.6. Střední hodnoty a směrodatné odchylky F0 z programu Praat..... 18</p> <p>5.7. Shimmer(apq3)..... 21</p> <p>5.8. Střední hodnoty a směrodatné odchylky jitteru(ppq5)..... 23</p> <p>5.9. Střední hodnoty HNR..... 24</p> <p>5.10. Graf všech získaných hodnot parametru HNR 24</p> <p>6.1. Formanty v MATLABu 25</p> <p>6.2. Formanty ve spektrogramu 26</p> <p>6.3. F1, F2 a F3 vůči věku mluvčích 27</p> <p>6.4. Střední hodnoty a směrodatné odchylky F1·F2 27</p> <p>6.5. Pohyb vrcholu vokálního trojúhelníku, ženy 28</p>
--	--

8.3. Úspěšnost pokusu klasifikace celé databáze	40
6.6. Pohyb vrcholu vokálního trojúhelníku, muži.....	28
6.7. Hodnoty F1 vůči F2	29
7.1. Ilustrace k hodnotám SE a SP.....	30
7.2. Závislost spektrálních momentů	32
7.3. Spektrální momenty /p/ muži .	33
7.4. Spektrální momenty /p/ ženy .	34
7.5. SE pro znělý signál.....	35
7.6. SP pro znělý signál.....	35
8.1. Klasifikační strom, ženy	38
8.2. Výsledek klasifikace žen ve věku do 43 let	38
8.3. Výsledek klasifikace žen ve věku mezi 43 a 57 lety	39
8.4. Výsledek klasifikace žen nad 57 let.....	39
B.1. Histogram databáze, 10 let	46
B.2. Amplitudové spektrum exploze /p/	46
B.3. Amplitudové spektrum /a/	47
B.4. Amplitudové spektrum pauzy .	47
B.5. Získání hodnoty F0 z kepstra..	48
B.6. Získání hodnoty F0 pomocí autokorelace	48
C.7. Schématické znázornění lidského hlasového ústrojí	49
C.9. Formanty ve spektru samohlásek.....	49
C.8. Distribuce věku menopauzy u žen	50
C.10. Grafické zobrazení vztahu F1 a F2.....	50

Kapitola 1

Úvod

Cílem práce je prozkoumat existenci možných závislostí akustických charakteristik mluvené řeči na věku řečníka. Hodnoty těchto charakteristik budou získány z řečové databáze obsahující promluvy mužů a žen, u kterých je známé jejich stáří v době pořízení nahrávek. Obsahem nahrávek z řečové databáze jsou opakované slabiky /pa/. S přihlédnutím k povaze analyzovaných dat budou zkoumány akustické parametry získané buď z okamžitých hodnot, nebo z velice krátkých úseků řečového signálu. Pro získané hodnoty parametrů zjistíme jejich střední hodnoty, směrodatné odchylky a míru korelace vzhledem k věku mluvčích. Následně bude provedena diskuze, zda lze odhalené trendy průběhů charakteristik využít ke zpětnému rozpoznání věku mluvčího, tedy jestli jde o statisticky významné trendy. Nakonec možnost rozpoznání věku otestujeme pokusem klasifikace.

Kromě možnosti případné klasifikace věku mluvčího je cílem analýzy akustických parametrů zmapování běžných rozsahů hodnot, kterých tyto parametry u zdravých dospělých jedinců nabývají. Mnoho prací se zabývá možností diagnózy rozličných nemocí, které se projevují výkyvy některých akustických parametrů hlasu. Pro tento druh výzkumu je důležité mít dobrou představu o hodnotách parametrů kontrolní skupiny. Navíc, pokud by parametr používaný k diagnostice dané choroby byl závislý na věku mluvčího, pak by přihlédnutí k tomuto trendu při diagnóze mohlo vést k přesnějším výsledkům. Výhoda metody odhalování patologií z akustických parametrů tkví v její rychlosti, neinvazivnosti a cenové efektivitě. Proto je přínosné se touto problematikou zabírat.

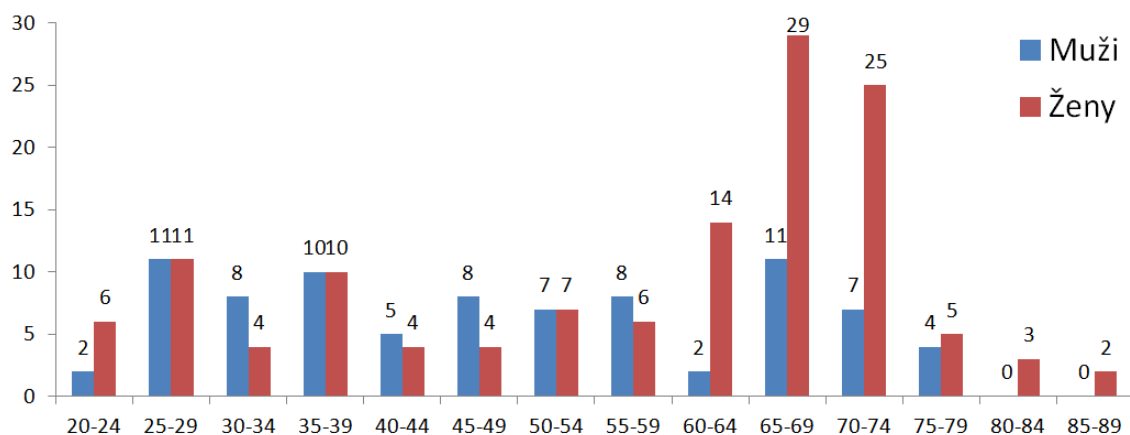
Kapitola 2

Řečová databáze

Obsahem použité řečové databáze jsou promluvy 83 mužů a 130 žen ve věku od dvaceti do devadesáti let, ve kterých je opakována slabika /pa/. Jednotlivé nahrávky nemají jednotné délky trvání, počet repetice slabiky /pa/, délky pauz mezi slabikami, ani frekvence opakování slabiky. Liší se i délka vyslovované samohlásky. U většiny záznamů je slabika /pa/ vyslovována krátce. V některých je však vyslovováno dlouhé /á/.

V databázi je vždy pro každého mluvčího k dispozici dvojice nahrávek. Nahrávky ve dvojici se rozlišují koncovkami *a* a *b* v názvu záznamu konkrétního mluvčího. Předmětem této práce je zkoumání závislosti parametrů promluv řečníka na věku, případně pohlaví, proto jsem názvy nahrávek v databázi pro snazší orientaci opatřila předponami. Názvy nahrávek mužů začínají písmenem *M* a názvy nahrávek žen písmenem *F*. Po tomto označení pohlaví následuje dvojčíslí věku mluvčího. Ostatní znaky názvů jsou ponechány z původních označení nahrávek v řečové databázi.

Většina nahrávek má vzorkovací frekvenci $F_s=48000\text{Hz}$, v databázi je ale přítomno i několik výjimek, kde má nahrávka vzorkovací frekvenci $F_s=44100\text{Hz}$. Proto je důležité, nestanovit vzorkovací frekvenci záznamu jako konstantu, ale pro každého mluvčího ji ověřit, než bude použita k dalším výpočtům.



Obrázek 2.1. Histogram počtů mužů a žen v řečové databázi v intervalech po pěti letech

Zkoumání akustických charakteristik řeči v závislosti k věku mluvčích bude ve většině následujících kapitol prováděno pro intervaly věku o velikosti pěti let. Vzhledem k počtu nahrávek v použité řečové databázi by totiž při uvažování konkrétního věku mluvčího, při určování středních hodnot a směrodatných odchylek z dat získaných akustických parametrů, mělo za následek mnoho skupin obsahujících jednoho nebo dva mluvčí. Pokud existují dlouhodobé trendy pro zkoumané akustické charakteristiky v závislosti na věku mluvčího, měly by být pozorovatelné i při agregaci věku mluvčích do pětiletých intervalů, neboť případné fyziologické změny hlasového traktu by měly mít graduální

nástup. Na obrázku 2.1 je histogram použité databáze, kde jsou mluvčí rozděleni do kategorií podle věku v intervalech po pěti letech. Databáze mužských nahrávek má relativně vyvážené zastoupení jednotlivých věkových kategorií, bohužel je ale jejich celkový počet velice nízký a nejsou k dispozici žádné nahrávky mužů starších osmdesáti let. U žen je k dispozici sice více nahrávek, ale je zde problémem nevyváženost v rámci věkových intervalů. Navíc ze statistického hlediska je bohužel i 130 jedinců relativně málo. Pokud však existují silnější závislosti akustických parametrů mluvčích na jejich věku, je možné, že budou pozorovatelné i na takto malém vzorku populace a přesnější povaha těchto závislostí může být v budoucnu lépe prozkoumána na větším počtu jedinců.

V příloze je pro ilustraci připojen obrázek histogramu řečové databáze B.1, kde je velikost intervalu věku stanovena na deset let.

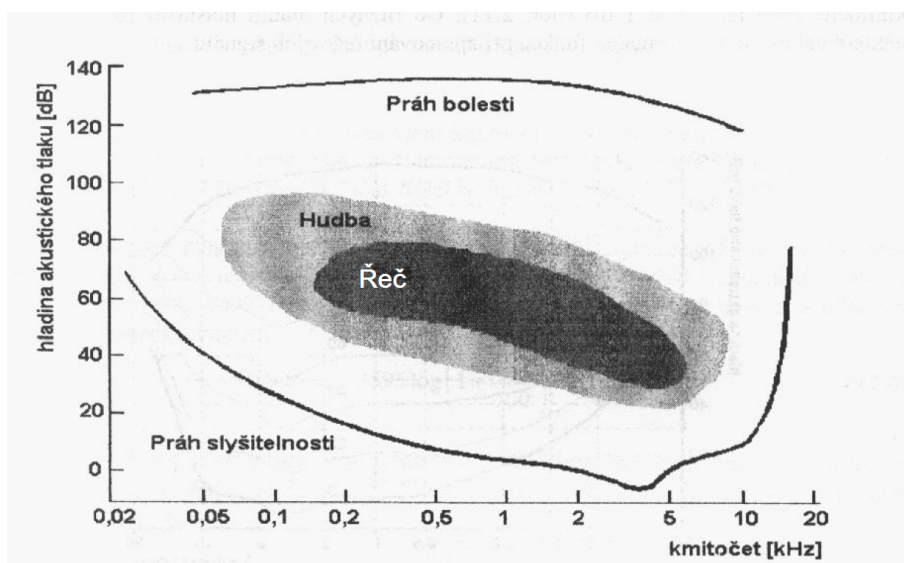
Řečová databáze byla před získáváním charakteristik akustického signálu zkontrolována z hlediska kvality jejích nahrávek. V několika případech byly vyřezány úseky nahrávek, které obsahovaly kromě mluvčího i cizí rušivé zvuky jako bouchání dveří, nebo hovor na pozadí nahrávky.

Kapitola 3

Vlastnosti mluvené řeči

Obecné informace o řeči prezentované v této kapitole jsou získány ze zdrojů [1], [2] a [3].

Mluvená řeč nám dává schopnost vyjádřit myšlenky pomocí množiny akustických symbolů. Přenáší se komunikačním kanálem ve formě akustického signálu, tedy díky vlnění elastického prostředí v oboru slyšitelných frekvencí. Slyšitelné vlnění, jehož frekvence leží v rozsahu, které je schopno vnímat sluchové ústrojí člověka, se liší u každého jedince a mění se s věkem. Obvykle se však uvádí, že zdravý člověk ve věku do 20-25 let je schopen vnímat zvuk o frekvenci přibližně 16Hz - 20kHz. Šířka pásma mluvené řeči je ale o mnoho menší, než šířka pásma slyšitelnosti. Udává se zhruba mezi 180Hz - 6kHz (pro trénované umělce to může být až od 80Hz do 16kHz). Pro věrnou reprodukci řeči se tedy počítá s frekvenčním pásmem do 12kHz, pro dobrou srozumitelnost s pásmem 300Hz až 5000Hz a pro nejběžnější konverzaci postačí i pásmo 300Hz až 2500Hz.



Obrázek 3.1. Plocha slyšení akustických signálů, převzato z [2]

Na rozdíl od psané řeči je v řeči mluvené obsažena i informační složka závislá na konkrétním mluvčím. Jedná se o charakteristiky jeho hlasového traktu, způsob artikulace, emocionální stav či řečové vady.

3.1 Reprezentace řeči

Lingvistika je věda zabývající se studiem lidského jazyka. Mluvenou řeč je možné popisovat z různých lingvistických hledisek. V této práci se budeme zabírat primárně hlediskem akustickým, fonetickým a fonologickým.

Fonetika zkoumá zvukový signál lidské řeči v celé jeho složitosti. Definuje možný

inventář hlásek a popisuje jejich artikulační a akustické vlastnosti.

Fonologie se zabývá hláskami z hlediska jejich výskytu a schopnosti vytvářet vzájemné kombinace. Fonologické jednotky mají hierarchické členění:

promluva → fráze → slova → slabiky → difony → fonémy → alofony

Foném je nejmenší jednotkou řeči, která je významotvorná. Čeština má 36 fonémů.

Fonémy se dělí do dvou skupin: Samohlásky (vokály) a souhlásky (konsonanty).

Slabika se skládá ze slabičného jádra (nukleus) a svahů slabiky. Slabičné jádro je obvykle tvořeno samohláskou nebo slabikotvornou souhláskou, tedy souhláskami s převládající tónovou složkou (/m/, /n/, /l/, /r/). Svahy slabiky jsou převážně tvořeny šumovými souhláskami a obklopují jádro slabiky. Počáteční svah slabiky nazýváme praetura. Koncový svah slabiky nazýváme koda. V českém jazyce může být slabika složena z jednoho až pěti fonémů.

Na řeč můžeme nahlížet dvěma způsoby. První je, že můžeme řeč popisovat jako sekvenci jednotlivých fonémů, tedy akustických segmentů. Tomuto popisu se říká segmentální popis. V druhém případě se zajímáme ne tolik co, ale jak něco mluvčí říká. Zkoumáme pak takzvanou prozodickou informaci. Termínem prozodie se označují vlastnosti řečového signálu související s výškou hlasu, hlasitostí a časováním. Prozodické rysy se většinou vztahují k větším řečovým jednotkám než k fonémům. Jsou proto označovány jako suprasegmentální jevy a mluvíme o suprasegmentálním popisu jazyka. Vzhledem k povaze použité řečové databáze v této práci se v následujících kapitolách budeme zabírat pouze segmentálním popisem.

3.2 Proces utváření řeči

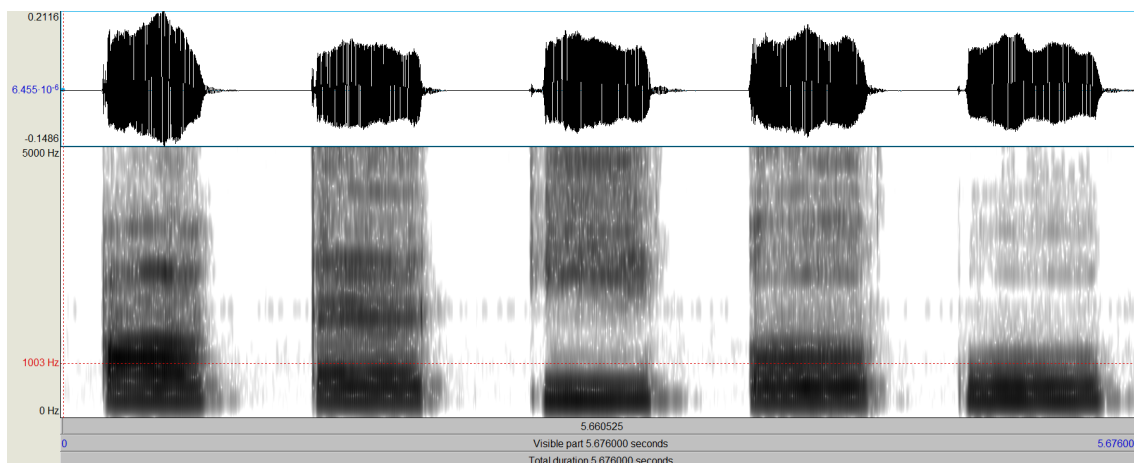
Skupina orgánů podílející se na tvorbě lidské mluvené řeči se nazývá řečové (artikulační) orgány. Na rozdíl od orgánů sluchových, které slouží v první řadě k příjmu akustického signálu, orgány řečové byly vyvinuty primárně pro jiný účel, než je tvorba akustického signálu.

Zdrojem energie řeči je dechové ústrojí. Při běžném dýchání má proud vzduchu z plic volnou cestu, neboť hlasivky jsou v klidovém postavení. Během řeči se však hlasivky stáhnou do fonačního postavení a cestu proudu vzduchu úplně uzavřou. Stažené hlasivky se pod tlakem vzduchu stanou pružnými a začnou kmitat. Tím se proud vzduchu z plic rozdrobí na kvanta hustšího a řidšího vzduchu a z dříve homogenního proudu vzduchu se stává pravidelný budící signál. Tento proces, při kterém je vytvářen zvuk, se nazývá fonace. Budící signál je v následujícím hlasovém traktu (dutina hrdelní, dutina nosní a dutina ústní) zpracován pomocí artikulace. Artikulací vznikne konečný řečový signál. Doplnující ilustrace k procesu utváření řeči je v příloze na obr. C.7.

3.3 Vlastnosti samohlásek

Všechny samohlásky jsou znělými hláskami, které vznikají v ustálené poloze hlasového traktu a při fonačním postavení hlasivek. Hlasivky kmitají a vydechovanému proudu vzduchu se v hlasovém traktu nad hlasivkami nestaví do cesty žádná překážka, která by proud přerušila. Artikulace jednotlivých samohlásek je docílena pomocí jazyka, rtů a měkkého patra.

Samohlásky se díky tomu oproti ostatním fonémům vyznačují vysokou kvaziperiodicitou, vysokou amplitudou, vysokou energií signálu, delším trváním a dobře



Obrázek 3.2. Spektrogram samohlásek /á/ /é/ /í/ /ó/ /ú/, vytvořeno v programu Praat.

pozorovatelnou budící frekvencí hlasu v jejich kmitočtovém spektru. Na spektrogramu na obr 3.2 je vidět, že jejich energie se soustřeďuje především pod 1kHz.

Čeština má 5 krátkých a 5 dlouhých samohláskových fonémů. Krátké a dlouhé samohlásky se v českém jazyce liší především dobou trvání. Jejich artikulace a spektrální vlastnosti jsou v zásadě stejné. Pro artikulaci českých samohlásek je nejvýznamější poloha a pohyb jazyka.

3.4 Vlastnosti souhlásek

Souhlásky vznikají v přechodových stavech hlasového traktu a obsahují ve svém spektru charakteristický šum. Ten je následkem zhoršení průchodnosti hlasového traktu během artikulace pro proud vzduchu, který vychází z plic. Podle svých vlastností se dělí do podskupin. Jedno z možných dělení je podle přítomnosti či absence základního hlasového tónu (budící frekvence) ve spektru, tedy dělení na souhlásky znělé a neznělé. Dále se dají dělit na souhlásky párové a nepárové. Párové jsou takové souhlásky, které mají protějšek jehož artikulace se odlišuje jen přidáním či odebráním základního hlasového tónu.

Párově neznělé	p	t	ť	k	f	s	š	ch	c	č	r
Párově znělé	b	d	ď	g	v	z	ž	h	dz	dž	ř
Nepárově znělé		m	n	ň	l	r	j				

Tabulka 3.1. Souhlásky v českém jazyce rozdělené do skupin podle znělosti a párovosti

Další možná dělení souhlásek jsou odvozena od konkrétní artikulace dané souhlásky nebo od jejích akustických vlastností.

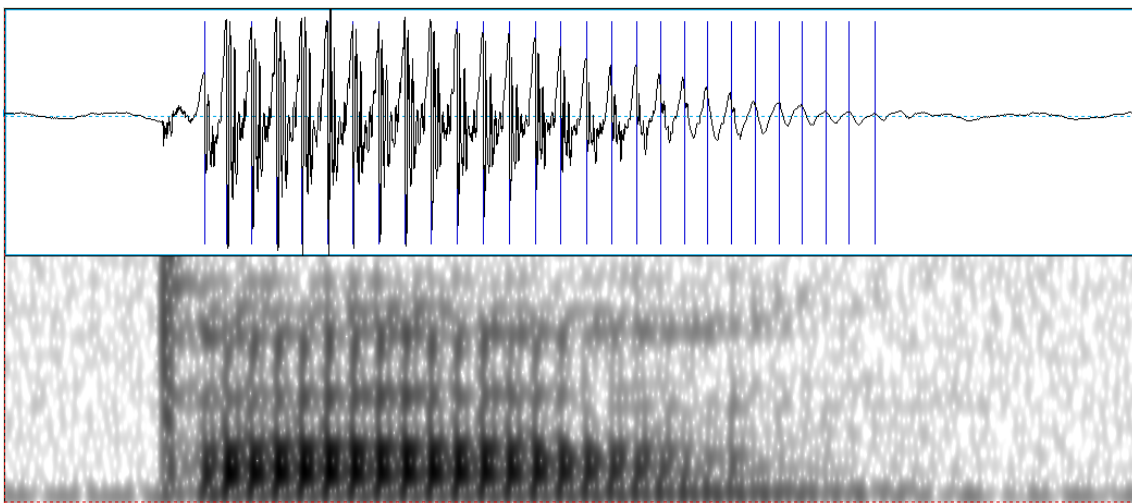
3.4.1 Vlastnosti souhlásky /p/

Protože foném /p/ je jediná souhláska vyskytující se v použité řečové databázi, rozebereme v této sekci jeho charakteristické vlastnosti.

Souhláska /p/ se řadí mezi souhlásky závěrové, neboli okluzivy. Jedná se o skupinu souhlásek, kde při artikulaci dojde k úplnému přerušení proudu vzduchu. Dále se řadí mezi skupinu souhlásek obouretných, takzvaných bilabiál, protože to je místo, kde při jeho artikulaci dojde k onomu přerušení.

Při artikulaci /p/ tedy nejprve nastává v řečovém signálu krátká pauza způsobená přerušením proudu vzduchu, po které následuje takzvaná exploze, tedy únik nahromaděného vzduchu. Exploze má velice krátkou dobu trvání (kolem 10ms). Exploze fonému /p/ se na spektrogramu projeví jako charakteristický a téměř celistvý vertikální pruh. Proto se také /p/ často označuje jako neznělá exploziva. Bohužel vzhledem k jeho krátkému trvání je obtížnější tento foném ve zvukovém signálu identifikovat a dále ho analyzovat.

Na obrázku 3.3 je spektrogram slabiky /pa/, na kterém je dobře viditelný vertikální pruh vybuzený explozí souhlásky /p/.



Obrázek 3.3. Spektrogram slabiky /pa/, vytvořeno v programu Praat.

Kapitola 4

Detekce znělých úseků v nahrávkách

Před samotnou analýzou charakteristik akustického signálu v programu matlab je nezbytné určit, kde se v nahrávkách z řečové databáze nacházejí znělé úseky a kde se nachází pauzy mezi slabikami /pa/. Tento proces budeme nazývat hrubou segmentací, a na jeho konci je výstupem počet slabik /pa/ na nahrávku a časy jejich počátků a konců, přičemž úsek řečového signálu obsahující jednu slabiku budeme nazývat segment. Hlavními zdroji této kapitoly byly [1] [2] [3] [4].

Protože artikulační orgány mluvího jsou fyzicky schopné pohybu jen do určitých rychlostí, můžeme považovat dostatečně krátký úsek řečového signálu za stacionární [3]. Takový úsek signálu se nazývá mikrosegment a jeho typická délka je 10ms, což je kompromis zajišťující zároveň dostatečnou délku pro bezchybné měření a přitom stále zachová stacionaritu úseku. Je však možné stanovit i jiné hodnoty délky mikrosegmentu (běžný rozsah je zhruba 10-25ms). Nahrávka se rozdělí na mikrosegmenty diskrétní konvolucí vzorků řečového signálu $s(n)$ a okénka $w(n)$. Okénka jsou impulsními funkcemi, které váží vzorky signálu. Mezi nejčastěji používané tvary okének patří pravoúhlé okénko, Hammingovo okénko a Hannovo okénko. Platí:

Pravoúhlé okénko:

$$w(n) = \begin{cases} 1 & \text{pro: } 1 \leq n \leq N, \\ 0 & \text{pro ostatní } n \end{cases} \quad (1)$$

Hammingovo okénko:

$$w(n) = \begin{cases} 0,54 - 0,46\cos(2\pi n/N) & \text{pro: } 1 \leq n \leq N, \\ 0 & \text{pro ostatní } n \end{cases} \quad (2)$$

Hannovo okénko:

$$w(n) = \begin{cases} \sin^2(\pi n/N) & \text{pro: } 1 \leq n \leq N, \\ 0 & \text{pro ostatní } n \end{cases} \quad (3)$$

Kde N je počet vzorků signálu na jeden mikrosegment.

Vybraný typ okna se během tvorby mikrosegmentů posunuje po časové ose. Velikost kroku tohoto posunu může být N vzorků na posun, pokud na sebe jednotlivé mikrosegmenty mají přímo navazovat. Lze však zvolit i krok menší než N vzorků aby se vytvořené mikrosegmenty překrývaly. Po vytvoření mikrosegmentů z celého řečového signálu můžeme signál považovat za posloupnost po sobě jdoucích stacionárních jevů,

které jsou svázány s odvozenými krátkodobými charakteristikami.

Pravoúhlé okénko je výhodné použít, má-li při analýze periodických vlastností znělých úseků řečového signálu perioda delší dobu trvání. To proto, že pro analýzu za pomoci tohoto okénka stačí, aby mikrosegment obsahoval jedinou periodu signálu. Oproti tomu pro analýzu za pomoci Hammingova okénka je potřeba mít periody obsaženy v mikrosegmentu alespoň dvě. Je tedy zapotřebí dvoujnásobně dlouhý mikrosegment. To může mít zejména pro mužské hlasy, které jsou hlubší a mohou tedy mít základní frekvenci až 50Hz (délka trvání periody 20ms) za následek ztrátu stacionarity signálu v rámci jednoho mikrosegmentu a získané hodnoty svázány s takovýmto mikrosegmentem jsou pouze průměry, které neodráží aktuální stav řečového signálu.

Nevýhodou pravoúhlého okénka je, že slabé komponenty řečového signálu nejsou postihnuty ve spektru. Hammingovo okénko poskytuje větší útlum pro vyšší frekvence, což tomuto jevu zamezí a proto je většinou v úlohách zpracovávání řeči Hammingovo okénko před pravoúhlým upřednostňováno. [1] [2]

Během hrubé segmentace se budou analyzovat různé vlastnosti mikrosegmentů a podle stanovených kritérií pak budou mikrosegmenty klasifikovány jako součást znělého nebo neznělého úseku signálu. Mikrosegmenty se používají zejména pro klasifikaci pomocí parametrů signálu, které nelze určit z jediného vzorku signálu $s(n)$. U parametrů, které takto určit lze, je použití mikrosegmentů možné, ale nikoliv nezbytné.

4.1 Detekce pomocí krátkodobé energie signálu

Jedná se o metodu krátkodobé analýzy v časové oblasti.

U periodických signálů je počítána průměrná energie periody signálu. Pro neperiodický signál, jakým je i řečový signál, je však energie signálu E definována jako hodnota okamžitého výkonu signálu $p(n)$, který je roven druhé mocnině hodnoty řečového signálu $s(n)$ ve vzorku n . Pro účely detekce je vhodné převést získané hodnoty na decibely.

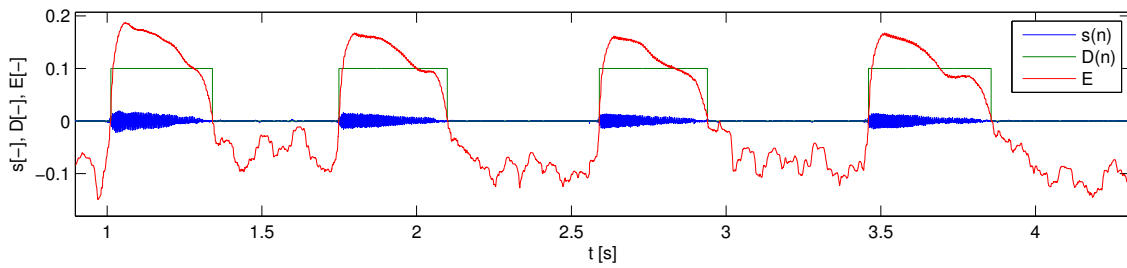
$$E(n) = 10 \log(p(n)) = 10 \log(s^2(n)) = 20 \log(|s(n)|) [dB] \quad (4)$$

Pro získání energetické obálky signálu je použit filtr počítající klouzavý průměr okamžitých hodnot energie signálu E . Tím se zahradí prudké výkyvy hodnot energie signálu na krátkých časových úsecích a výsledné hodnoty energie znělých úseků signálu jsou výrazně vyšší než hodnoty úseků neznělých. Díky tomu je možné energetickou obálku signálu použít k hrubé segmentaci signálu a určit hodnoty detektoru segmentů $D(n)$ dle následující logiky:

$$D(n) = \begin{cases} 1 & \text{pro: } E_{obal}(n) > E(E), \\ 0 & \text{pro: } E_{obal}(n) \leq E(E) \end{cases} \quad (5)$$

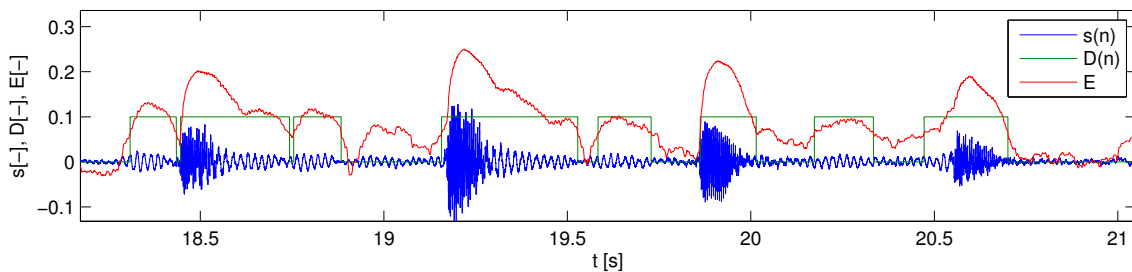
kde $E_{obal}(n)$ je hodnota energetické obálky n -tého vzorku signálu $s(n)$ a $E(E)$ značí střední hodnotu energie celého řečového signálu.

Poznámka k obrázkům 4.1 a 4.2: energetická obálka E signálu je posunuta po ose y a přenásobena konstantou, aby bylo možné průběhy zobrazit přes sebe v jednom grafu. Detektor $D(n)$ byl také přenásoben konstantou.



Obrázek 4.1. Detekce znělých úseků v nahrávce za použití krátkodobé energie. Vytvořeno v MATLABu

Na obrázcích 4.1 a 4.2 je vidět, že tato metoda detekce funguje dobře, je-li analyzovaný signál málo zašuměn. V opačném případě však není příliš spolehlivá, neboť hodnoty energetické obálky signálu se v neznělých úsecích dostatečně neliší od znělých úseků.



Obrázek 4.2. Příklad neúspěšné detekce znělých úseků v příliš zašuměné nahrávce za použití krátkodobé energie. Vytvořeno v MATLABu

4.2 Detekce pomocí spektra

Stejně jako libovolný jiný signál, lze i mikrosegment řečového signálu lze popsat jako součet sinusových funkcí. K jednoznačné definici průběhu funkce sinus stačí znát tři hodnoty: Amplitudu A , frekvenci f (kde platí, že úhlová rychlost ω je rovna $2\pi f$) a fázový posun φ . Z toho plyne, že mikrosegment lze popsat množinou trojic těchto parametrů. Zobrazení fázového posunu a amplitudy vůči frekvenci, tedy zobrazení ve frekvenční oblasti, se nazývá spektrem a vztah mezi časovou a frekvenční oblastí řečového signálu je popsán krátkodobou diskrétní Fourierovou transformací:

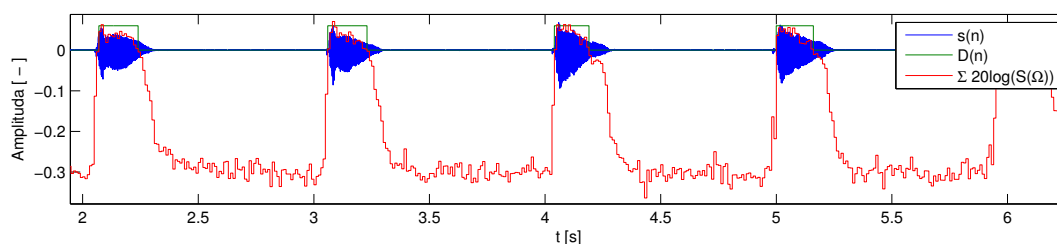
$$S(\Omega) = \sum_{k=-\infty}^{\infty} s(k)w(n-k)e^{-j\omega k} \quad (6)$$

kde funkce $w(n)$ je vybraná funkce okénka pro tvorbu mikrosegmentu signálu. K numerickému výpočtu spektrálních složek je použit algoritmus FFT v MATLABu. Existují způsoby klasifikace ve spektru, které odhadují konkrétní foném, ze kterého je mikrosegment vybrán. Pro účely hrubé segmentace na konkrétní použité řečové databázi však není nutné určení konkrétního fonému. Navíc je známo, že obsahem každé nahrávky použité řečové databáze jsou jen slabiky /pa/, tedy fonémy /p/ a /a/ a neznělé úseky mezi těmito slabikami. Je tedy možné se podívat na spektra vytvořená z mikrosegmentů těchto tří stavů, a stanovit vhodný způsob detekce znělých úseků pro

tuto specifickou databázi. Ukázka řečového signálu a jeho amplitudového spektra pro mikrosegment obsahující explozi písmene /p/ je v příloze na obr. B.2, mikrosegment obsahující část fonému /a/ je v příloze na obr. B.3 a mikrosegment obsahující neznělý úsek je rovněž v příloze na obr. B.4. Z těchto grafů je patrné, že největší rozdíly mezi znělým a neznělým úsekem řečového signálu se nacházejí ve frekvenčním pásmu od 0 do 1500Hz. V logaritmickém amplitudovém spektru se v tomto pásmu hodnoty znělého signálu na rozdíl od neznělého pohybují v převážně kladných hodnotách. Jako parametr mikrosegmentu, dle kterého budeme určovat, zda pochází z neznělého či znělého úseku řečového signálu, zvolíme sumu hodnot, kterých nabývá logaritmické amplitudové spektrum signálu ve frekvenčním pásmu od 0Hz do 1500Hz. Pokud hodnota tohoto součtu bude větší než nula, budeme považovat mikrosegment za znělý úsek signálu.

$$D(n) = \begin{cases} 1 & \text{pro: } \sum 20\log(|S(\Omega)|) > 0 \\ 0 & \text{pro: } \sum 20\log(|S(\Omega)|) \leq 0 \end{cases} \quad \text{pro hodnoty } S(\Omega) \text{ kde } 0 \leq f \leq 1500\text{Hz}, \quad (7)$$

Samozřejmě hodnota, podle které se klasifikuje znělost mikrosegmentu signálu, nemusí být konstanta. Lze jí stanovit například jako střední hodnotu všech parametrů mikrosegmentů v dané nahrávce, případně jako střední hodnotu přenásobenou konstantou.



Obrázek 4.3. Hrubá segmentace řečového signálu za použití frekvenčního spektra

Na obrázku 4.3 je detekce pomocí spektra otestována. Funkce jsou opět přenásobeny konstantami, aby bylo možné je zobrazit v jednom grafu.

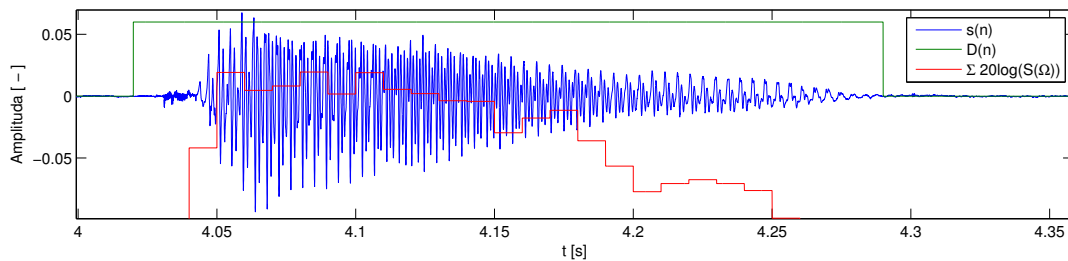
Nyní je vhodné přidat podmínky ke generování detektoru $D(n)$, které zajistí co největší pravděpodobnost, že jednotlivé výstupní segmenty signálu budou skutečně obsahovat celou slabiku /pa/, aniž by bylo nutné manuálně upravovat časy počátků a konců. Tyto podmínky mohou být s případnými menšími obměnami přidány k libovolnému způsobu detekce.

Většina metod detekce spolehlivě odhalí počátek vokálu ve slabice. Chceme-li zajistit, že součástí segmentu bude slabika včetně počáteční exploze fonému /p/, je vhodné od stanoveného času vzestupné hrany detektoru $D(n)$ odečíst konstantu. V této práci byla jako konstanta zvolena doba trvání tří mikrosegmentů, tedy 30ms.

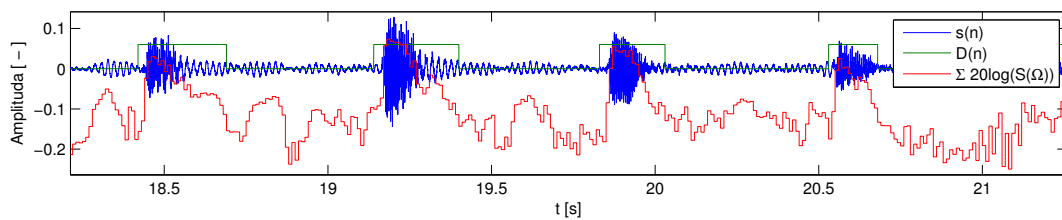
Také vzhledem k průběhu signálu slabik /pa/ je vhodné stanovit různá kritéria pro klasifikaci vzestupných a sestupných hran $D(n)$. Z grafu 4.3 je obzvláště dobře patrné, že počátek slabiky se projeví prudkou změnou ve spektru, zatímco v druhé polovině slabiky a po jejím skončení jsou změny pozvolnější. Pro klasifikaci vzestupné hrany je tedy možné ponechat přísnější kritérium, zatímco pro hranu sestupnou musí parametr klesnout pod hodnotu nižší. Díky tomu nebude docházet k useknutí části konce slabiky

při segmentaci a zároveň u zašuměných nahrávek nebude docházet k falešně pozitivním segmentům klasifikovaným z šumu mezi slabikami.

Dále lze přidat pravidlo, které nedovolí příliš časté střídání vzestupných a sestupných hran. Pro tuto práci byl zvolen minimální počet mikrosegmentů mezi hranami na pět, tedy 50ms, protože žádná slabika ani mezera mezi slabikami v databázi nebude mít kratší trvání. Toto pravidlo zamezí špatné segmentaci v případě kolísání hodnot, podle kterých se mikrosegmenty klasifikují, k čemuž obzvláště často dochází na koncích slabik nebo u zašuměných signálů. Příklad hrubé segmentace na zašuměném a nezašuměném řečovém signálu, za použití spektra signálu a dodatečných podmínek na obrázcích 4.4 a 4.5.



Obrázek 4.4. Hrubá segmentace řečového signálu za použití frekvenčního spektra a dodatečných podmínek



Obrázek 4.5. Hrubá segmentace zašuměného řečového signálu za použití frekvenčního spektra a dodatečných podmínek

Kapitola 5

Akustické parametry z oblasti fonace

Cílem této práce je prozkoumat akusticko-fonetické parametry se zaměřením na jejich věkovou závislost. U akustických parametrů z oblasti fonace, tedy procesu vytváření zvuku, mohou být příčiny závislosti parametrů na věku spojeny se fyziologickými změnami hrtanu a hlasivek. Tyto změny mohou být spojeny nejen se stářím, ale i s životospřávou mluvího (významným faktorem je například kouření) a mnohými dalšími podružnými faktory, nebo dokonce patologiemi. Mezi hlavními příčinami změn hlasu s věkem se udává zmenšování objemu hrtanových svalů, tuhnutí a změna tloušťky hlasivek, kalcifikace až případné kostnatění chrupavek hrtanu a celkově menší pružnost hlasového ústrojí [5] [6]. Mnohé z těchto změn hlasového aparátu mohou být zapříčiněny stárnutím organismu, ale také hormonálními změnami, přičemž oddělit, které změny hlasu jsou důsledkem čeho, je velice obtížné.

Hlasivky jsou velice citlivé na pohlavní hormony. Mezi ty nejdůležitější patří androgeny, progesteron a estrogen. U žen tato citlivost zapříčiňuje změny v poloze základní hlasové frekvence F_0 v průběhu menstruačního cyklu. Změny F_0 jsou způsobeny i pokud žena používá hormonální antikoncepci. Změna hormonů spojená s věkem ženy nastává kromě puberty ještě během menopauzy. Menopauza nastupuje obvykle mezi 49 až 52 lety ve vyspělých zemích a mezi 41 až 47 lety v zemích rozvojových[7](ilustrační obrázky v příloze C.8). Změny hlasu spojené s menopauzou jsou individuální pro každou ženu podle jejich fyzických dispozic. Také lze oddálit jejich nástup užitím hormonální antikoncepce[6]. I mužům se s věkem mění množství a poměr pohlavních hormonů. Po třicátém roku života postupně klesá množství testosteronu v organismu. Bohužel andropauza má na rozdíl od menopauzy graduální nástup a hůře se tak určuje konkrétní věková hranice[8]. Navíc je poměr jednotlivých pohlavních hormonů u mužů mnohem individuálnější a méně vázaný na konkrétní věk. U mužů je také v případě diagnostiky hormonální nerovnováhy možná léčba. Proto je u nich, na rozdíl od žen, mnohem obtížnější stanovit věkový interval očekávané změny a zkoumat tak možnou závislost pohlavních hormonů na jejich hlas.

5.1 Frekvence základního hlasivkového tónu

Při fonačním postavení hlasivek, dochází díky jejich periodickému kmitání k rozdrobení vzduchového proudu vycházejícího z plic na kvanta. Perioda tohoto jevu T_0 je nazývána periodou základního hlasivkového tónu a frekvence F_0 , pro kterou platí $T_0 = 1/F_0$, se nazývá frekvence základního hlasivkového tónu. Kmitání hlasivek můžeme aproximovat jako stojaté mechanické vlnění struny upevněné na obou koncích. Frekvence vlnění takové struny je definováno jako:

$$f = \frac{1}{2l} \sqrt{\frac{F}{\rho_L}} \quad (1)$$

kde l je délka napnuté struny, F je napínající síla působící na strunu a ρ_L lineární hustota struny, která je definována jako:

$$\rho_L = \frac{m}{L} \quad (2)$$

kde m je hmotnost struny a L je délka struny.

Ze vztahu (1) se můžeme pokusit vyvodit příčiny odlišných hodnot F_0 pro různé mluvčí.

Vidíme, že hodnota frekvence f bude nepřímo úměrná délce napnuté struny L , což znamená, že hodnota fundamentální frekvence F_0 by měla být nepřímo úměrná délce efektivní části hlasivek (tedy té části, která kmitá). Touto závislostí se můžeme pokusit vysvětlit rozdíl hodnot fundamentální frekvence F_0 mezi muži a ženami. V průměru mají muži délku efektivní části hlasivek o 60% delší než ženy [9]. Odtud vyplývá, že by měl platit vztah $F_{0F} = 1,6F_{0M}$ (uvažujeme-li, že parametry síly napínající hlasivky F a lineární hustota hlasivek ρ_L nejsou ovlivněny pohlavním dimorfismem člověka a tudíž jsou shodné pro průměrného muže i průměrnou ženu). A skutečně, ověříme-li tento údaj s průměrnými hodnotami F_0 mužů a žen z literatury [1], kde se průměr F_0 pro muže udává jako $F_{0M} = 132\text{Hz}$ a průměr F_0 pro ženy jako $F_{0F} = 223\text{Hz}$, vyjde, že $F_{0F} \approx 1,69F_{0M}$. Z toho je patrné, že délka efektivní části hlasivek má značný vliv na hodnoty fundamentální frekvence mluvčího. Vzhledem k tomu, že mluvčí v použité řečové databázi jsou již dospělí jedinci, nepředpokládáme, že změny v F_0 budou zapříčiněny růstem. Na druhou stranu nelze vyloučit změnu délky efektivní části hlasivek jedince s věkem, kvůli komplexním změnám, které v organismu s nástupem stáří probíhají.

Běžný rozsah hodnot fundamentální frekvence je 50Hz až 500Hz, kde nejnižší hodnoty mají hluboké mužské hlasy a nejvyšší hlásky malých dětí (v některých případech může dosahovat F_0 novorozenců až 1000Hz [10]). Hodnota F_0 většiny mužů spadá do intervalu 80 až 160Hz a hodnota F_0 většiny žen spadá do intervalu 150 až 300Hz[1].

Je také důležité podotknout, že hodnotu fundamentální frekvence promluvy může mluvčí vědomě ovlivňovat (například intonací ve větě, nebo při zpěvu podle melodie písně). Každý mluvčí však má nepříliš rozsáhlý interval hodnot hlasové frekvence, které jsou jeho hlasovému ústrojí nejpřirozenější, tedy tyto hlasové frekvence jsou pro něho nejméně namáhavé vyprodukovat (za použití uvolněného hlasu). Jako hodnotu fundamentální frekvence mluvčího pak chápeme spíše průměrnou hodnotu tohoto přirozeného rozsahu mluvčího během promluvy[10]. V této kapitole se tedy snažíme ověřit závislost průměru přirozeného rozsahu fundamentální hlasové frekvence mluvčích na věku.

■ 5.1.1 F_0 v programu MATLAB

Metody výpočtu F_0 v této kapitole byly čerpány z [11].

Obecná úloha při odhadování základní hlasové frekvence F_0 z řečového signálu je v úseku analyzovaného signálu identifikovat dominantní frekvenci, se kterou se signál opakuje. Při řešení této úlohy narážíme na ztěžující okolnosti. Nejprve je pro řešení úlohy podstatné analyzovat úsek signálu, který obsahuje periodický signál. Tento problém byl řešen v kapitole o detekci znělých úseků v náhrávkách a není třeba se jím

zde dále zaobírat. Další skupina překážek úspěšného odhadu je nekvalita signálu. Tou může být vysoká míra šumu. Navíc i tento šum může mít vlastní dominantní frekvenci, která správné určení F_0 dále ztěžuje. Navíc se F_0 mluvčího může v čase měnit. Čím rychleji mění svojí hodnotu, tím obtížnější je pak její správné určení. Nakonec je značný problém správně určit periodu T_0 a ne nějaký její násobek.

V této kapitole budou testovány odhady fundamentální hlasové frekvence dvěma různými přístupy v programu MATLAB.

První metodou je odhad ze stacionárních řečových segmentů signálu v kepstru. Kepstrum získáme fourierovou transformací logaritmického amplitudového spektra signálu. Z původního signálu je tedy kepstrum vytvořeno dvojnásobnou fourierovou transformací.

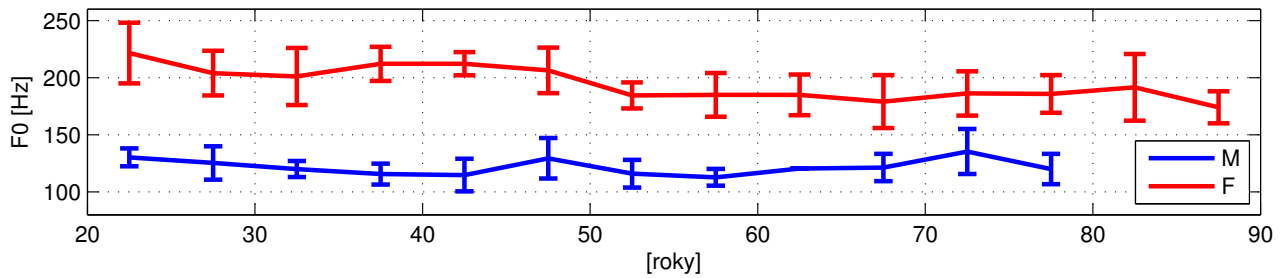
Metoda určení základní hlasové frekvence z kepstra vychází z následující úvahy. Je-li zkoumaný signál periodický, pak jeho amplitudové spektrum bude obsahovat mnoho harmonických frekvencí s konstantními rozestupy. V následující fourierově transformaci tak v zásadě analyzujeme další periodický signál, jehož velikost periody odpovídá fundamentální frekvenci původního řečového signálu. Ve výsledném kepstru by se tudíž mělo objevit výrazné lokální maximum v oblasti odpovídající fyzicky možným hodnotám F_0 lidské řeči viz tabulka 5.1. Veličina osy x v kepstru je Q efrency a udává se v sekundách. Pokud toto lokální maximum úspěšně identifikujeme, je možné dopočítat výslednou fundamentální frekvenci. Hodnota osy x kepstra v místě maxima odpovídá hodnotě periody základního hlasivkového tónu T_0 . F_0 je tedy $1/T_0$. Ilustrační graf k metodě je v příloze B.5.

Nevýhoda metody spočívá v její vysoké chybovosti, pokud analyzovaný segment signálu není téměř bez šumu. I z tohoto důvodu jsou předpokládané hodnoty F_0 v tabulce 5.1 v relativně úzkých intervalech. U zašuměného signálu má logaritmické amplitudové spektrum signálu nižší energii harmonické složky a zvýšenou energii šumové složky, což vede k horší harmonické struktuře a v důsledku toho se pak v kepstru objevovala další lokální maxima, která vedla ke špatnému určení F_0 . I v takto úzkých intervalech mají hodnoty F_0 u stejného mluvčího v různých slabikách značný rozptyl, jak můžeme pozorovat na obrázcích 5.3 a 5.2. V obou grafech jsou zjištěné hodnoty F_0 pro daného mluvčího z každé slabiky. U promluvy ženy, jejíž hodnoty F_0 jsou na obrázku 5.2, lze předpokládat a poslechem nahrávky potvrdit, že klesající trend F_0 během nahrávky je dán intonací. U muže, jehož hodnoty F_0 jsou na obrázku 5.3, se však některé hodnoty liší příliš mnoho od střední hodnoty, aby byly způsobeny intonací. V těchto bodech selhala metoda odhadu F_0 .

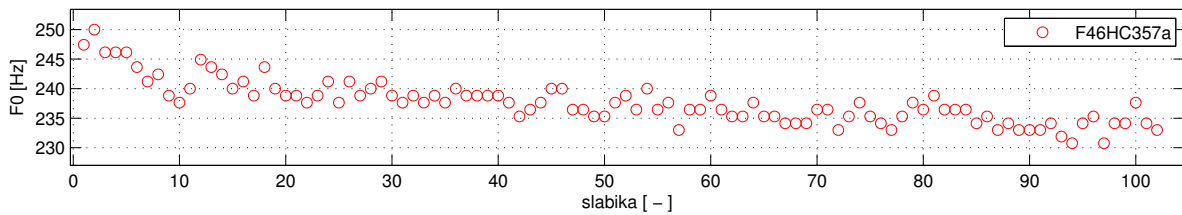
	Předpokládané hodnoty F_0	Oblast hledání lokálního maxima v kepstru
Muži	50Hz - 200Hz	0,005s - 0,02s
Ženy	100Hz - 300Hz	0,003s - 0,01s

Tabulka 5.1. Intervaly předpokládaných hodnot F_0 pro muže a ženy a korespondující oblasti hledání lokálního maxima kepstra.

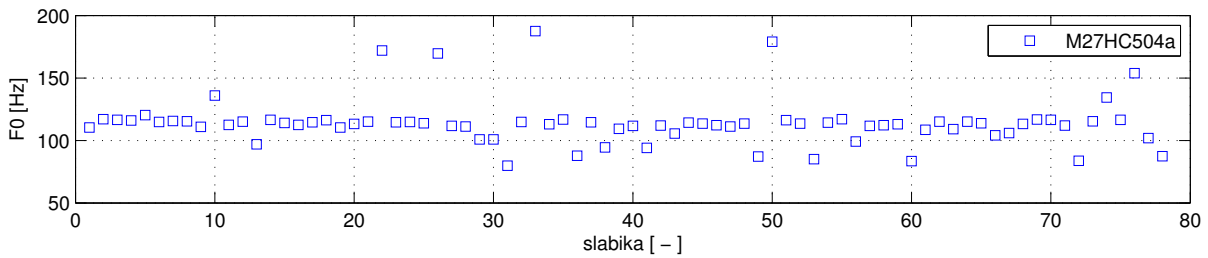
Druhou metodou získání F_0 je odhad v časové oblasti pomocí autokorelace. Princip této metody spočívá v hledání periodičnosti v samotném segmentu řečového signálu. Periodičnost lze odhalit použitím funkce autokorelace, která udává, jak dobře signál koreluje sám se sebou na množině různých hodnot zpoždění průběhu. U periodického



Obrázek 5.1. Střední hodnoty a směrodatné odchylky parametru F0 získaného pomocí kepstra, vztažené vůči věku mluvčích



Obrázek 5.2. Hodnoty F0 získané pomocí kepstra signálu ve všech slabikách proluvy F46HC357a



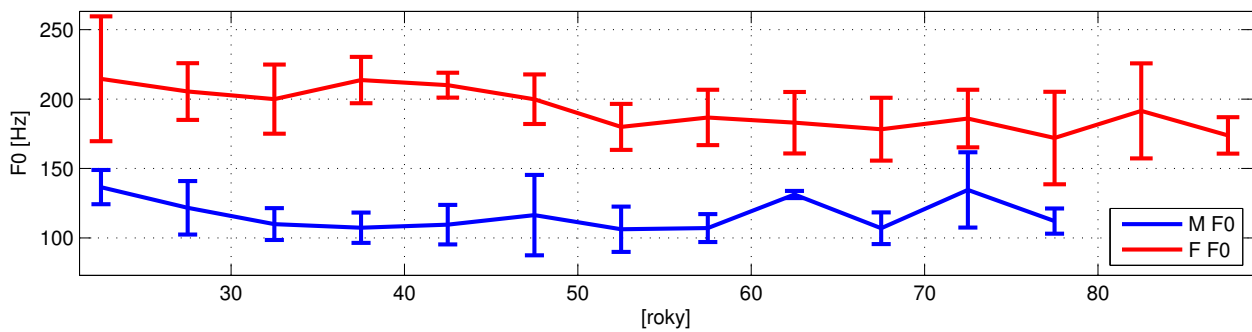
Obrázek 5.3. Hodnoty F0 získané pomocí kepstra signálu ve všech slabikách proluvy M27HC504

signálu lze očekávat, že signál bude dobře korelovat pro všechny hodnoty zpoždění průběhu, které jsou násobky základní hlasové periody T_0 . Z toho plyne hlavní nevýhoda metody, kterou je občasné půlení F_0 , kvůli detekci dvojnásobné hodnoty T_0 . Stejně jako u předchozí metody, základní hlasovou frekvenci odhadujeme v intervalu fyzicky dosažitelných hodnot pro lidskou řeč viz tabulka 5.2. Na rozdíl od metody odhadu za pomoci kepstra je zde možné povolit větší intervaly, neboť metoda je robustnější. Ilustrační graf k metodě je v příloze B.6.

Na obrázku 5.4 jsou střední hodnoty a směrodatné odchylky fundamentální frekvence F_0 mužů a žen získané pomocí funkce autokorelace.

	Předpokládané hodnoty F_0	Oblast hledání lokálního maxima autokorelace
Muži	60Hz - 250Hz	0,004s - 0,016s
Ženy	100Hz - 500Hz	0,002s - 0,01s

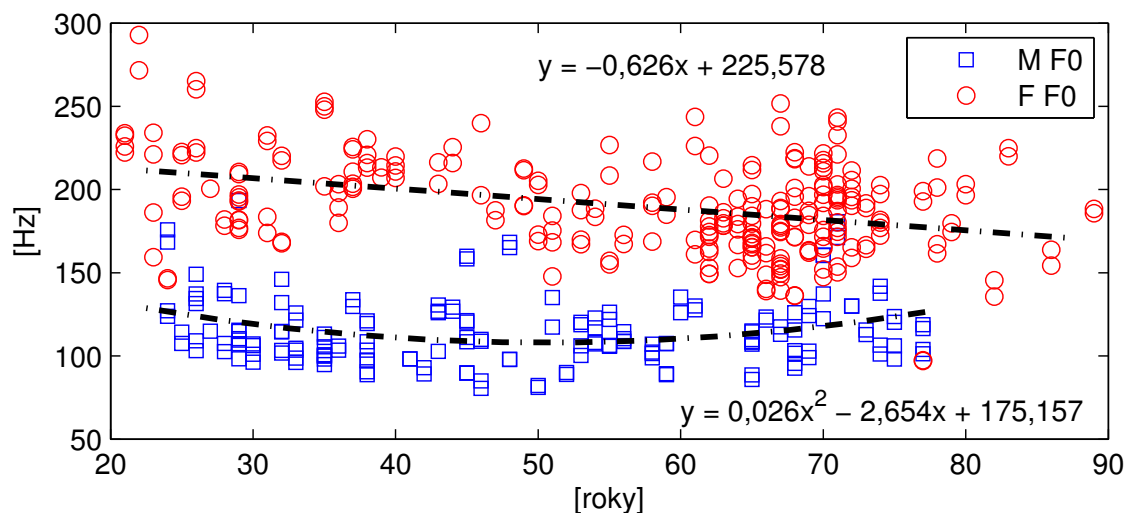
Tabulka 5.2. Intervaly předpokládaných hodnot F_0 pro muže a ženy a korespondující oblasti hledání lokálního maxima autokorelace.



Obrázek 5.4. Střední hodnoty a směrodatné odchylky parametru F_0 získaného pomocí autokorelační funkce, vztažené vůči věku mluvčích

5.1.2 F_0 v programu Praat

Třetí způsob získání hodnot fundamentální frekvence F_0 byl pomocí funkce v programu Praat. Na obrázku 5.5 jsou všechny získané hodnoty F_0 pro muže i ženy, kde hodnoty fundamentální frekvence žen byly aproximovány lineární funkcí a hodnoty F_0 mužů byly aproximovány pomocí polynomu druhého řádu.



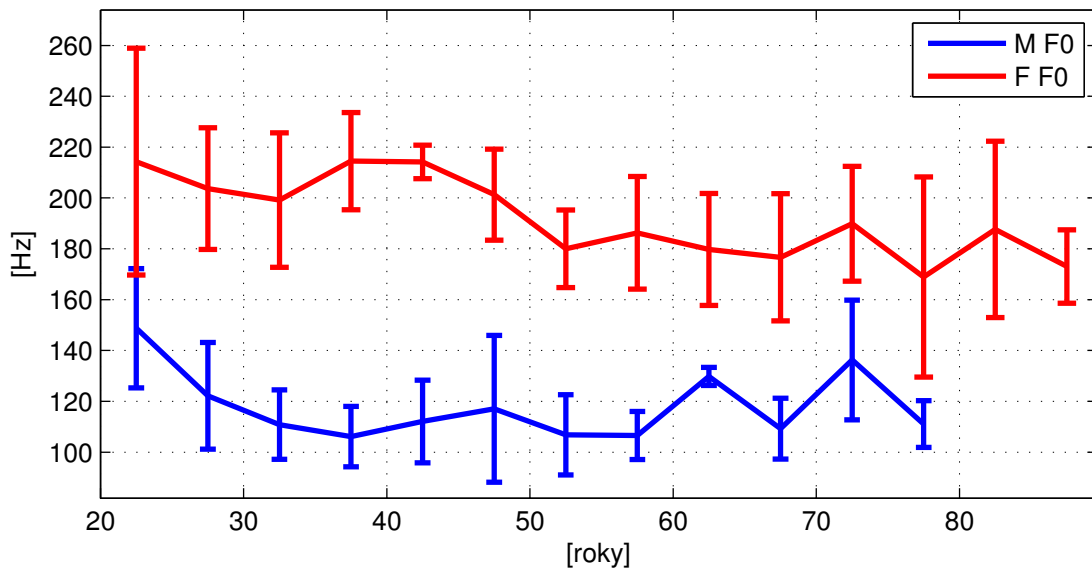
Obrázek 5.5. Všechny hodnoty F_0 získané programem Praat

Na obrázku 5.6 jsou střední hodnoty a směrodatné odchylky fundamentální frekvence získané programem Praat, které jsou vztažené vůči věku mluvčích.

5.1.3 F_0 závěr

Pro zajímavost se ještě vrátíme k tématu aproximace hlasivek jako strun upevněných na obou koncích. Tvzení, že délka efektivní části hlasivek je z velké části zodpovědná za výslednou hodnotu F_0 , a proto by měl platit vztah $F_{0F} = 1,6F_{0M}$, jsme otestovali na průměrných hodnotách F_0 mužů a žen ze zdroje [1]. Nyní můžeme porovnat poměry hodnot F_0 mužů a žen na průměrných frekvencích získaných z použité řečové databáze. V tabulce 5.3 jsou průměrné hodnoty získané jednotlivými metodami odhadu F_0 .

Budeme-li tedy uvažovat hodnotu $F_{0F} \cong 189$ a $F_{0M} \cong 114$, pak dojdeme k poměru $F_{0F} \cong 1,658F_{0M}$. Tento výsledek je zajímavý i proto, že průměry F_0 vyšly v použité řečové databázi pro obě pohlaví o něco nižší než v použité literatuře. Odlišnost průměrů



Obrázek 5.6. Střední hodnoty a jejich směrodatné odchylky parametru F0 získaného programem Praat

	F0 muži [Hz]	F0 ženy [Hz]
Kepstrum	114,4	187,8
Autokorelace	114,1	189,7
program Praat	115,0	189,6

Tabulka 5.3. průměrné hodnoty F0 získány jednotlivými metodami

hodnot fundamentální frekvence, ale zachování poměru se můžeme pokusit vysvětlit nevyvážeností řečové databáze z hlediska věku mluvčích. Pokud u žen s věkem hodnota F0 skutečně klesá, jak výsledky této práce naznačují, pak velký počet šedesátnic může způsobit nižší průměr F0, než jaký by nastal při rovnoměrném zastoupení napříč věkovými skupinami. Průměrná hodnota F0 žen ve věku do 43 let byla 209Hz, průměrná hodnota F0 žen ve věku nad 57 let byla 181Hz. Nejnižší průměrné hodnoty F0 u mužů se opět nacházejí ve věkových kategoriích s větším počtem jedinců, což vede k nižšímu průměru hodnot F0 mužů v této práci.

Nyní můžeme diskutovat výsledky hodnot F0 všech tří postupů jejich získání. Hodnoty získané pro jednotlivé mluvčí se pro tyto tři metody liší většinou jen velice málo, zvláště pokud uvažujeme, že hodnota fundamentální hlasové frekvence se přirozeně pohybuje v určitém rozsahu. Získaná trojice grafů středních hodnot F0 a jejich směrodatných odchylek si je pak také velice podobná.

Pro zhodnocení míry závislosti fundamentální frekvence řečového signálu na věku mluvčích použijeme nejprve výběrový Pearsonův korelační koeficient $r(X, Y)$, který je odhadem korelačního koeficientu $\rho(X, Y)$. Pro danou charakteristiku X a věk muže nebo ženy Y je míra korelace dána jako [12] [13]:

$$r(X, Y) = \frac{\text{cov}(X, Y)}{S_x S_Y} = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_i (X_i - \bar{X})^2} \sqrt{\sum_i (Y_i - \bar{Y})^2}} \quad (3)$$

V tabulce 5.4 jsou míry korelace hodnot F0 vůči věku mluvčích pro muže i ženy a pro všechny tři sady hodnot F0. Podle očekávání pro ženy vychází záporná hodnota míry korelace, ukazující na klesající trend hodnot dat, zatímco trend vývoje hodnot F0 mužů s věkem nelze uspokojivě aproximovat lineární přímkou.

Korelace věku mluvčích s	F0 muži	F0 ženy
Kepstrum	0,017	-0,443
Autokorelace	0,000	-0,423
program Praat	-0,006	-0,393

Tabulka 5.4. Hodnoty korelace F0 vůči věku mluvčích, dělené podle jednotlivých metod získání hodnot F0

Míru korelace F0 s věkem v tabulce 5.5 budeme určovat pro určité věkové intervaly. U žen vycházíme z předpokladu, že hlavní příčina poklesu hlasu ve stáří je daná pohlavními hormony. Proto je rozdělíme do tří skupin. První skupina obsahuje ženy ve věku od 20 do 43 let, pro které předpokládáme, že ještě menopauzou neprošly. Druhá skupina obsahuje ženy ve věku od 44 do 56 let, ve které jsou statisticky některé ženy před a některé již po přechodu. A nakonec třetí skupina obsahuje ženy od 57 let do 90 let, u kterých je velice pravděpodobné, že menopauzou již prošly. Pokud je naše hypotéza platná, potom by první a třetí skupina žen měla mít velice malou hodnotu korelace jejich věku a hodnoty fundamentální frekvence. K poklesu F0 by mělo docházet hlavně u druhé věkové skupiny. Výsledné hodnoty v tabulce tuto teorii nevyvrací.

Korelace věku mluvčích s	F0 ženy 20-43let	F0 ženy 44-56let	F0 ženy 57-90let
Kepstrum	-0,068	-0,451	-0,017
Autokorelace	-0,071	-0,453	-0,017
Program Praat	-0,031	-0,484	-0,008

Tabulka 5.5. Hodnoty korelace F0 vůči věku žen, které jsou rozděleny do tří skupin podle věku

Nakonec zhodnotíme statistickou významnost klesajícího trendu fundamentální frekvence žen s jejich rostoucím věkem. Stanovíme hypotézu H_0 , která závislost zkoumané charakteristiky F0 na věku mluvčího popírá: $H_0 : \rho(X, Y) = 0$. Hypotézu H_0 se pak můžeme pokusit zamítnout použitím Studentova t-testu. Alternativní hypotéza k H_0 je $H_A : \rho(X, Y) \neq 0$. Testovací statistika T bude mít tvar:

$$T = \frac{r(X, Y)}{\sqrt{1 - r(X, Y)^2}} \sqrt{n - 2} \quad (4)$$

	hodnota T	hladina p
Všechny ženy	-7,498	p < 0,001
Ženy 20-43 let	-0,578	p < 0,005
Ženy 44-56 let	-2,689	
Ženy 57-90 let	-0,215	

Tabulka 5.6. Výsledky studentova t-testu, vztah věku žen k hodnotě jejich F0

a bude porovnávána s kritickou hodnotou t-rozdělení $t_\alpha(n-2)$ na hladině významnosti α a $n-2$ stupni volnosti (převzato z [12]). Získané hodnoty jsou v tabulce 5.6.

Z grafů středních hodnot a směrodatných odchylek F0 mužů je patrné, že nejspíše také existuje nějaká závislost tohoto parametru na věku mluvčího. Bohužel však vzhledem k nedostatku informací o hormonálních změnách v populaci mužů vztažených k jejich věku nejsem schopna stanovit hypotézu, na kterých věkových úsecích by měly existovat jaké tendence hodnoty F0. Je také možné, že u mužů nemají hormony na frekvenci F0 vliv, a za změny hodnoty F0 může jiný fyziologický jev. Z grafů středních hodnot F0 můžeme usuzovat, že hodnoty fundamentální frekvence mužů mají tvar písmene V popisovaný ve zdroji [5], kde nejprve dochází k dlouhodobému pozvolnému poklesu hodnoty F0 a poté naopak k dlouhodobému pozvolnému růstu.

5.2 Shimmer

Čerpáno z [14] a [1].

Velikost amplitud zvukového signálu v po sobě jdoucích periodách základního hlasového tónu nebývá u žádného mluvčího zcela konstantní. Parametr, který toto kolísání popisuje se nazývá shimmer. Jedná se o střední rozdíl mezi amplitudami A sousedících hlasivkových pulsů a obvykle je vyjádřený v decibelech [dB], popřípadě relativní shimmer je vyjádřen v procentech [%].

Malé hodnoty shimmeru běžný posluchač nezaznamená, shimmer je dobře postřehnutelný od jednoho decibelu, přičemž při normální řeči se běžně pohybuje v rozsahu 0,04-0,21dB.[1] Shimmer je užitečný parametr při určování řečových patologií, neboť takové hlasy mají často vyšší hodnoty shimmeru než zdravé. [15]

Program praat počítá několik různých variant shimmeru.

$Shimmer(local, dB)$ je základní shimmer, který určuje průměrné kolísání amplitud v decibelech.

$$Shimmer(local, dB) = \frac{1}{N-1} \sum_{n=1}^{N-1} |A_n - A_{n+1}| \quad (5)$$

$Shimmer(local)$ je průměrný rozdíl v absolutní hodnotě mezi amplitudami po sobě jdoucích period, vydělený průměrnou amplitudou

$$Shimmer(local) = \frac{\frac{1}{N-1} \sum_{n=1}^{N-1} |A_n - A_{n+1}|}{\frac{1}{N} \sum_{n=1}^N A_n} \quad (6)$$

$$Shimmer(apq3) = \frac{\frac{1}{N-2} \sum_{n=1}^{N-2} \left| \frac{A_n + A_{n+1} + A_{n+2}}{3} - A_{n+1} \right|}{\frac{1}{N} \sum_{n=1}^N A_n} \quad (7)$$

Shimmer amplitude pertubation quotient(apq3) je shimmer určený ze trojic (případně pětic pro apq5, nebo z jedenácti pro apq11) po sobě jdoucích period.

U těchto pěti shimmerů byla v závislosti na věku mluvčích otestována míra korelace. Hodnoty, pro které následný t-test prokázal věkovou závislost alespoň na hladině $p < 0,01$ jsou v tabulkách tučně.

Shimmer	(local, dB)	(local)	(apq3)	(apq5)	(apq11)
Muži A	0,167	0,217	0,272	0,232	0,183
Muži B	0,204	0,203	0,259	0,241	0,192
Ženy A	0,150	0,174	0,210	0,167	0,140
Ženy B	0,126	0,144	0,188	0,144	0,146

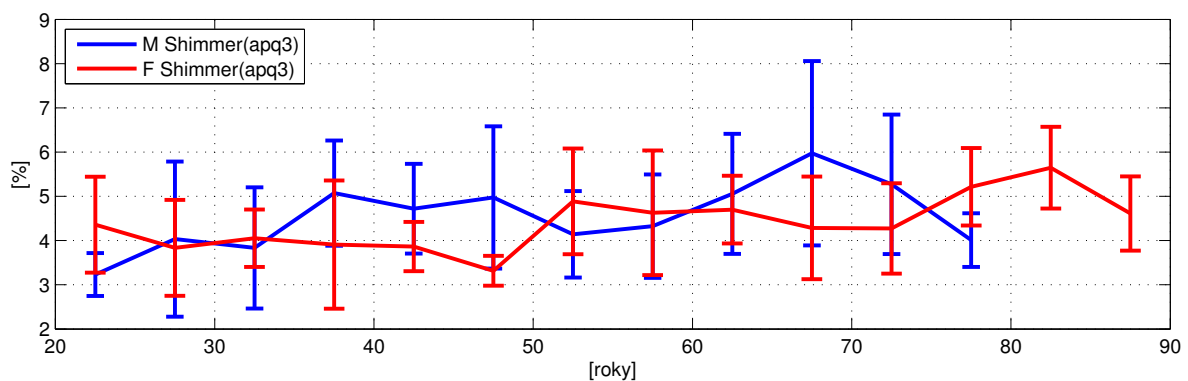
Tabulka 5.7. Korelace věku mluvčích a Shimmeru v nahrávkách s koncovkou *a* a *b*

Shimmer	(local, dB)	(local)	(apq3)	(apq5)	(apq11)
Muži A i B	0,186	0,209	0,265	0,237	0,187
Ženy A i B	0,138	0,159	0,198	0,156	0,143

Tabulka 5.8. Korelace věku mluvčích a Shimmeru z obou nahrávek pro každého mluvčího

Shimmer	(local, dB)	(local)	(apq3)	(apq5)	(apq11)
Ženy 20-43let	-0,037	-0,072	-0,073	0,027	0,079
Ženy 44-56let	0,183	0,285	0,416	0,408	0,237
Ženy 57-90let	0,038	0,068	0,080	0,036	-0,018

Tabulka 5.9. Korelace věku žen a hodnot shimmeru v intervalech podle věku



Obrázek 5.7. Střední hodnoty a směrodatné odchylky parametru shimmer(apq3) mužů(M) a žen(F) ve věkových skupinách po pěti letech

5.3 Jitter

Čerpáno z [14] a [1].

Délky trvání po sobě následujících period $T0$ základního hlasového tónu také nebývají u žádného mluvčího zcela konstantní. Parametr, který toto kolísání popisuje se nazývá jitter. Jitter je definován jako střední rozdíl délek sousedních period dělený střední délkou periody a je obvykle vyjadřován v procentech, případně lze vyjádřit absolutní jitter v sekundách. Hodnoty jitteru v běžné řeči jsou 0,5 až 1 procento a běžně postřehnutelný při poslechu je zhruba od dvou procent [1]. I jitter je užitečným parametrem při určování patologií řeči, protože patologické hlasy mívají vyšší hodnoty jitteru [15]. Program Praat počítá několik různých variant jitteru [14].

$Jitter(local)$ je průměrný absolutní rozdíl mezi po sobě jdoucími periodami podělený průměrnou délkou periody $T0$.

$$Jitter(local) = \frac{\frac{1}{N-1} \sum_{n=1}^{N-1} |T0_n - T0_{n+1}|}{\frac{1}{N} \sum_{n=1}^N T0_n} \quad (8)$$

$Jitter(local, abs)$ je průměrný absolutní rozdíl mezi po sobě jdoucími periodami udaný v sekundách.

$$Jitter(local, abs) = \frac{1}{N-1} \sum_{n=1}^{N-1} |T0_n - T0_{n+1}| \quad (9)$$

$Jitter\ relative\ average\ pertubation$ (rap), neboli relativní průměrná porucha je průměrný absolutní rozdíl mezi periodou a průměrem jí a jejích dvou sousedů. Je tedy určován ze trojic po sobě jdoucích period.

$$Jitter(rap) = \frac{\frac{1}{N-2} \sum_{n=1}^{N-2} \left| \frac{T0_n + T0_{n+1} + T0_{n+2}}{3} - T0_{n+1} \right|}{\frac{1}{N} \sum_{n=1}^N T0_n} \quad (10)$$

$Jitter\ five\ point\ Period\ Perturbation\ Quotient$ (ppq) je velmi podobný jitter(rap), rozdíl je jen, že průměr se počítá z periody a jejích čtyř nejbližších sousedů.

$$Jitter(ppq5) = \frac{\frac{1}{N-4} \sum_{n=1}^{N-4} \left| \frac{T0_n + T0_{n+1} + T0_{n+2} + T0_{n+3} + T0_{n+4}}{5} - T0_{n+2} \right|}{\frac{1}{N} \sum_{n=1}^N T0_n} \quad (11)$$

Výsledné korelace věku a hodnot variant jitteru jsu v tabulkách 5.10 a 5.11. Hodnoty, pro které následný t-test prokázal věkovou závislost alespoň na hladině $p < 0,01$ jsou opět v tabulkách tučně.

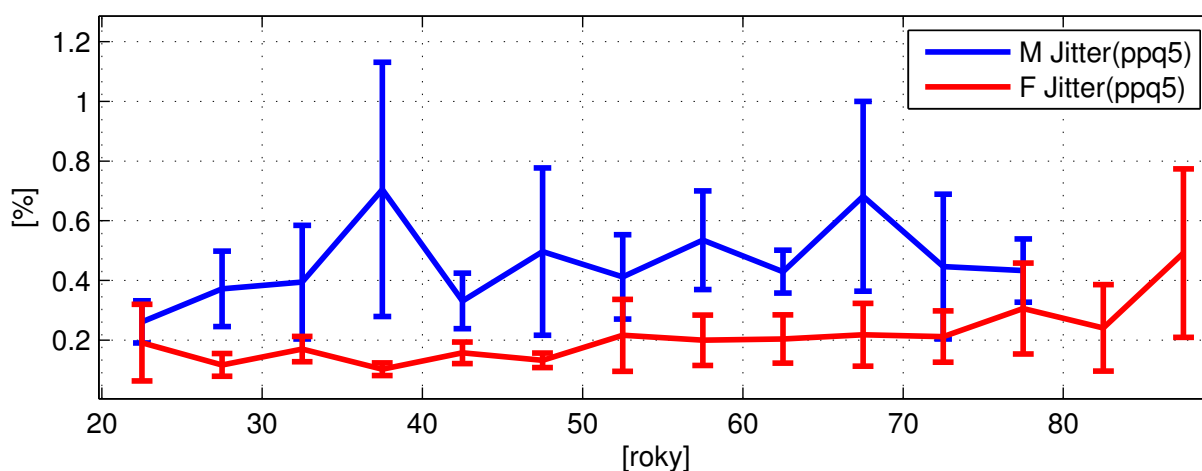
Na obrázku 5.8 je graf středních hodnot a směrodatných odchylek jitteru(ppq). Všechny ostatní varianty jitteru mají velice podobné průběhy.

Jitter	(local)	(local, abs)	(rap)	(ppq5)
Muži A i B	0,174	0,124	0,119	0,156
Ženy A i B	0,392	0,366	0,361	0,382

Tabulka 5.10. Korelace různých variant parametru jitter a věku mužů i žen, neděleno na dvě části databáze

Jitter	(local)	(local, abs)	(rap)	(ppq5)
Ženy 20-43 let	-0,141	-0,162	-0,184	-0,154
Ženy 44-56 let	0,368	0,385	0,335	0,350
Ženy 57-90 let	0,307	0,261	0,284	0,290

Tabulka 5.11. Korelace různých variant parametru jitter a věku žen



Obrázek 5.8. Střední hodnoty a směrodatné odchylky parametru jitter(ppq5) pro muže(M) a ženy (F), vztažené k věku mluvčích

5.4 HNR

Čerpáno z [14] a [16].

Parametr HNR je veličina popisující poměr harmonických kmitočtů k šumu. Hodnota HNR bude klesat pro chraptivější mluvčí a pro mluvčí s hlasitějším dýcháním. Dále bude klesat pro jedince s různými patologiemi hrtanu [15]. Hodnoty, pro které následný t-test prokázal věkovou závislost alespoň na hladině $p < 0,01$, jsou opět v tabulkách 5.12 a 5.13 tučně.

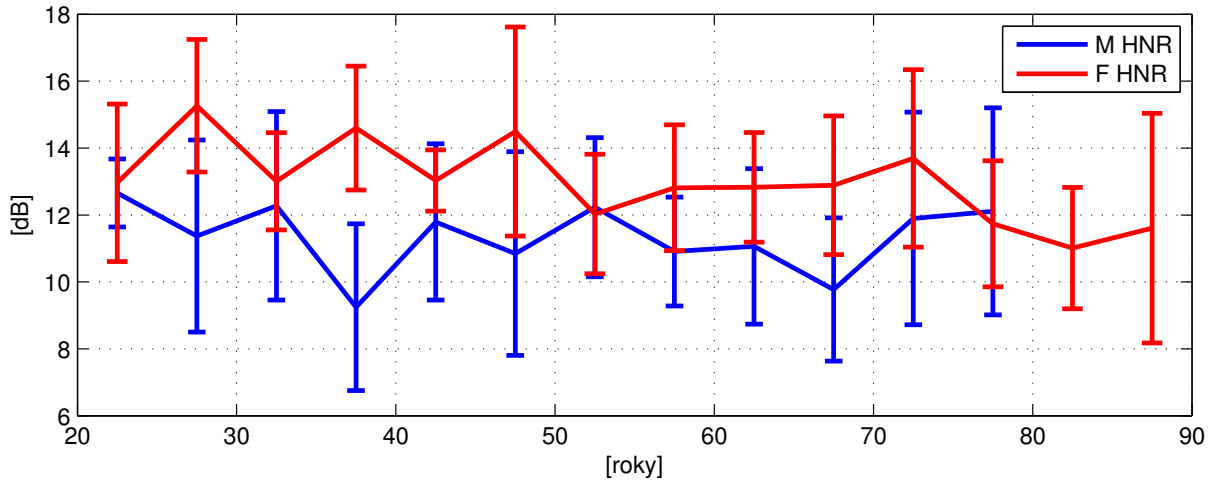
Korelace věku a	HNR
Muži	-0,039
Ženy	-0,234

Tabulka 5.12. Míra korelace hodnot HNR s věkem mužů a žen

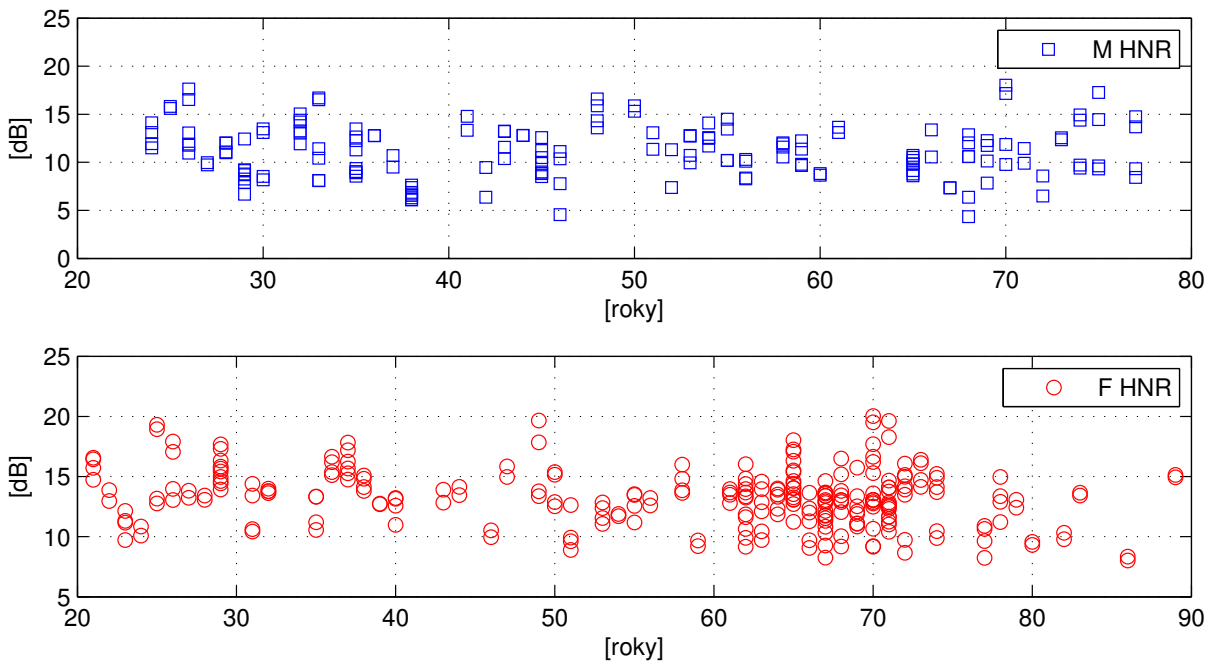
Na obrázku 5.9 je graf středních hodnot a směrodatných odchylek parametru HNR.

Korelace věku a	HNR
Ženy 20-43 let	-0,020
Ženy 44-56 let	-0,225
Ženy 57-90 let	-0,115

Tabulka 5.13. Korelace HNR věk ženy menopauza



Obrázek 5.9. Střední hodnoty a směrodatné odchylky parametru HNR mužů(M) a žen(F) ve věkových skupinách po pěti letech



Obrázek 5.10. Všechny hodnoty HNR získané programem Praat, pro muže(M) a ženy(F) vztahované k věku

Kapitola 6

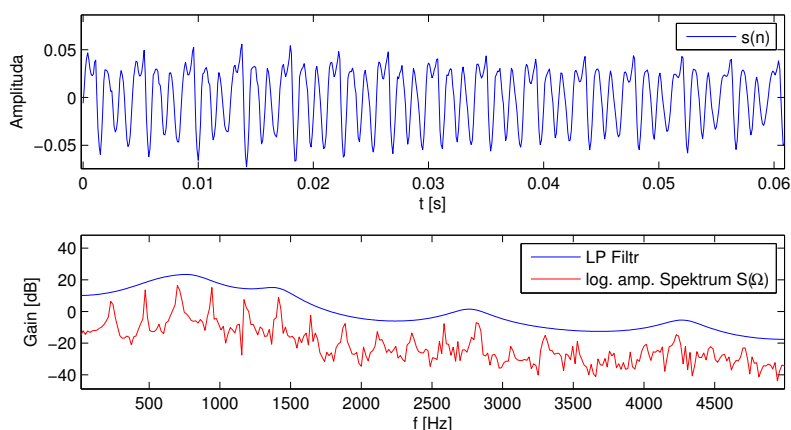
Akustické parametry z oblasti artikulace: formanty

Hlavní zdroje informací prezentovaných v této kapitole jsou [17], [11] a [14].

V této kapitole bude prozkoumávána závislost formantů samohlásky /a/ na věku mluvčích. Formantové frekvence vznikají při tvorbě řečového signálu změnou rozložení akustické energie základního hlasivkového tónu během artikulace. Akustická energie se soustředí právě v okolí formantových frekvencí, viz ilustrační obrázek v příloze C.9, kde je znázorněna spektrální obálka signálu, jejíž polohy lokálních maxim odpovídají formantovým frekvencím. První tři formanty jsou považovány za rezonanční kmitočty největších dutin hlasového traktu. Formant F1 je přiřazen k dutině hrdelní, formant F2 k dutině ústní a nakonec formant F3 k dutině nosní.

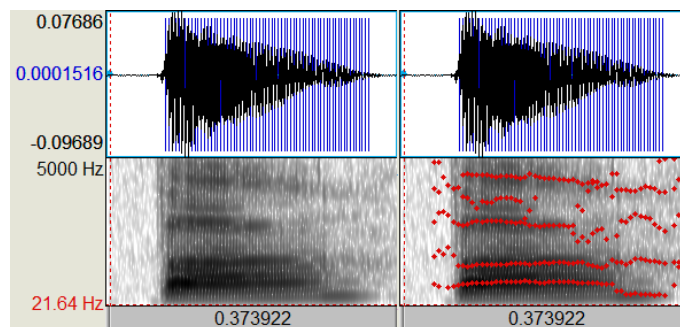
Odhad hodnoty jednotlivých formantů je obecně náročnější úloha než odhad fundamentální frekvence F0. To z důvodu, že formanty jsou vlastnostmi hlasového traktu mluvčího a ne samotného řečového signálu. Z toho musí být spíše odvozeny, nelze je přímo měřit. Často využívaná metoda odhadu formantů je založena na modelování segmentu řečového signálu, jako by byl generován konkrétním zdrojem signálu a filtrem. Formanty jsou poté odhadnuty z rezonančních frekvencí modelovaného systému. Metoda analýzy, kde se hledá nejlépe pasující systém, se nazývá lineární predikce.

Na obr. 6.1 je demonstrován postup hledání rezonančních frekvencí v MATLABu. Je na něm patrné, že lokální maxima frekvenční odezvy LP filtru korespondují s tvarem logaritmickeho amplitudového spektra, které bylo přímo vytvořené z daného úseku řečového signálu $s(n)$. Konkrétní hodnoty rezonančních frekvencí získáme nalezením kořenů polynomu LP filtru pro hodnoty koeficientů filtru, které byly získané z analyzovaného řečového úseku.



Obrázek 6.1. Segment řečového signálu $s(n)$, frekvenční odezva LP filtru a pro porovnání logaritmicke amplitudové spektrum signálu $S(\Omega)$ vytvořené z původního řečového signálu $s(n)$

Výsledné hodnoty formantů prezenované v této práci byly získány pomocí programu Praat. Ve spektrogramu vytvořeném ze znělých úseků řeči můžeme pozorovat formanty jako tmavé horizontální pásy. Na obr 6.2 je ukázka spektrogramu vytvořeném ze slabiky /pa/ v programu Praat. V levé části obrázku můžeme pozorovat horizontální pruhy frekvencí, ve kterých se soustředí energie signálu. Napravo jsou ve spektrogramu také červeně vyznačeny hodnoty formantových frekvencí, jejichž odhad provedl program Praat.



Obrázek 6.2. Slabika /pa/, její spektrogram a napravo spektrogram s vyhledanými formantovými frekvencemi programem Praat

Program Praat umožňuje získat tabulky hodnot formantů v čase pro celý analyzovaný záznam. Z těchto dat můžeme vybrat hodnoty formantů odpovídající časům fonému /a/ v jednotlivých slabikách /pa/ ve zkoumané nahrávce. Časy slabik jsme získali pomocí hrubé segmentace nahrávek a nyní je můžeme znovu využít.

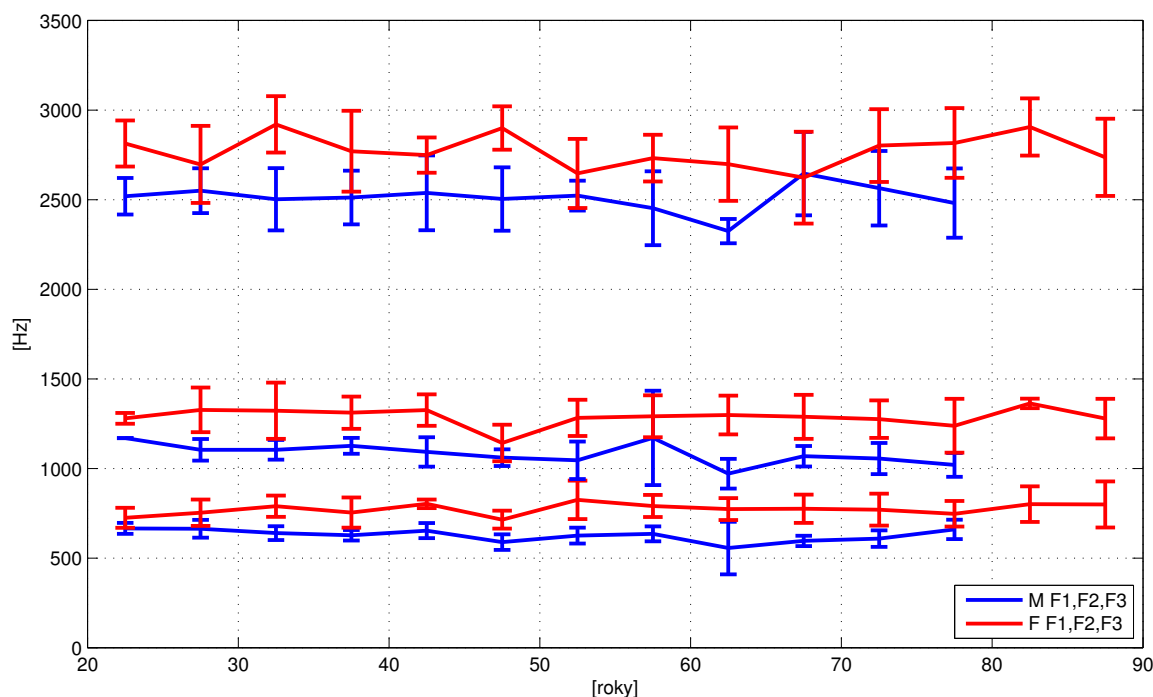
Na obrázku 6.3 jsou střední hodnoty a směrodatné odchylky formantů F1, F2 a F3 ve věkových skupinách po pěti letech.

Korelace věku a	F1	F2	F3	F1·F2
Muži	-0,249	-0,211	0,055	-0,278
Ženy	0,100	-0,076	-0,043	0,026

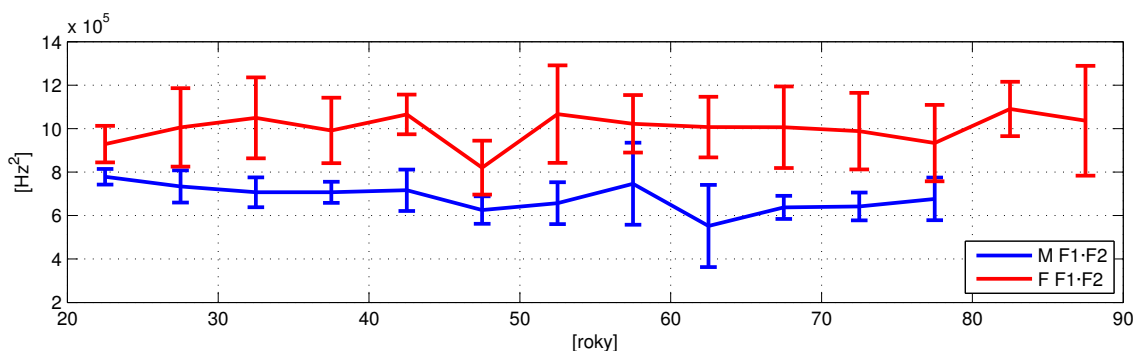
Tabulka 6.1. Korelace F1, F2, F3 a F1·F2 vůči věku mužů a žen

První dvě formantové frekvence se při analýze řeči považují za nejdůležitější. Často je zkoumán vztah mezi těmito dvěma parametry. Každá samohláska má charakteristická pásma hodnot, ve kterých se pohybují její hodnoty F1 a F2. Vyznačením těchto pásem v grafu získáme takzvaný vokální trojúhelník, viz obr. C.10. Hodnoty formantů fonému /a/ tedy tvoří jeden z vrcholů tohoto trojúhelníku. Vzhledem k nepřítomnosti ostatních samohlásek v použité řečové databázi se není možné věnovat změnám obsahu vokálního trojúhelníku v závislosti na věku mluvčích. Je ale možné prozkoumat změny obsahu obdélníku, který vznikne vynásobením hodnoty F1 s hodnotou F2 mluvčího viz obrázek 6.4.

Dále můžeme prozkoumat míru korelace mezi hodnotami F1 a F2 jednotlivých mluvčích. V tabulce 6.2 jsou hodnoty korelace F1 a F2 pro muže a ženy. Hodnoty, pro které následný t-test prokázal věkovou závislost alespoň na hladině $p < 0,01$ jsou opět v tabulce tučně. Na obrázku 6.7 jsou vyznačeny získané hodnoty formantů všech mluvčích v databázi, které jsou proloženy přímkami lineární regrese.



Obrázek 6.3. Střední hodnoty a směrodatné odchytky parametrů F1, F2 a F3 vůči věku mluvčích, ve věkových skupinách po pěti letech.

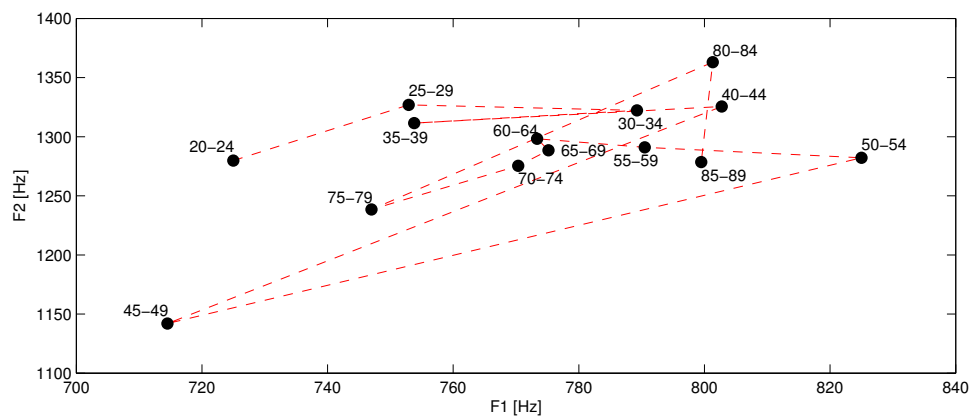


Obrázek 6.4. Střední hodnoty a směrodatné odchytky hodnot F1·F2 vztažené k věku mluvčích na intervalech po pěti letech.

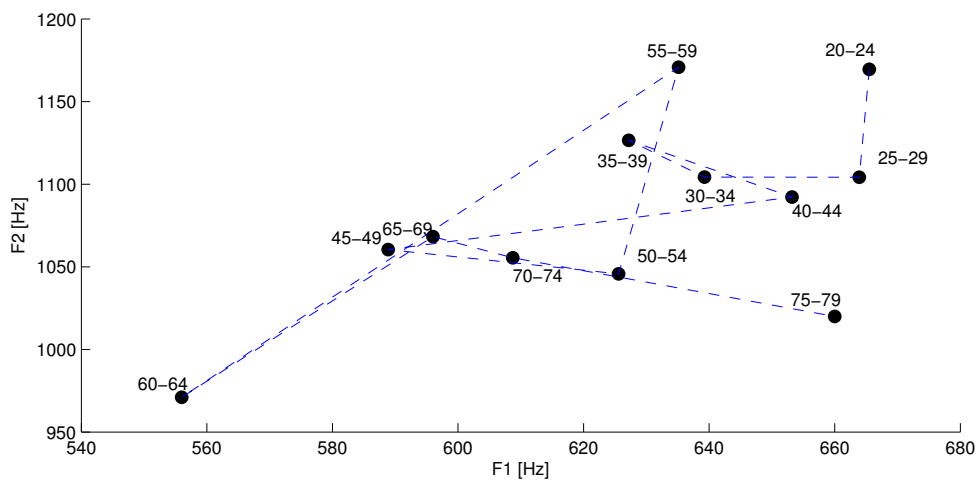
	Korelace F1 a F2
Muži	0,343
Ženy	0,597

Tabulka 6.2. Korelace F1 a F2 u mužů a žen

Nakonec můžeme prozkoumat posun těžiště vrcholu vokalického trojúhelníku fonému /a/ v závislosti na věku mluvčích. Na obrázku 6.5 jsou zobrazeny střední hodnoty F1 vůči středním hodnotám F2 pro jednotlivé věkové kategorie žen. Po sobě následující věkové kategorie jsou propojeny čarou. Stejným způsobem byl vytvořen obrázek 6.6 pro muže.

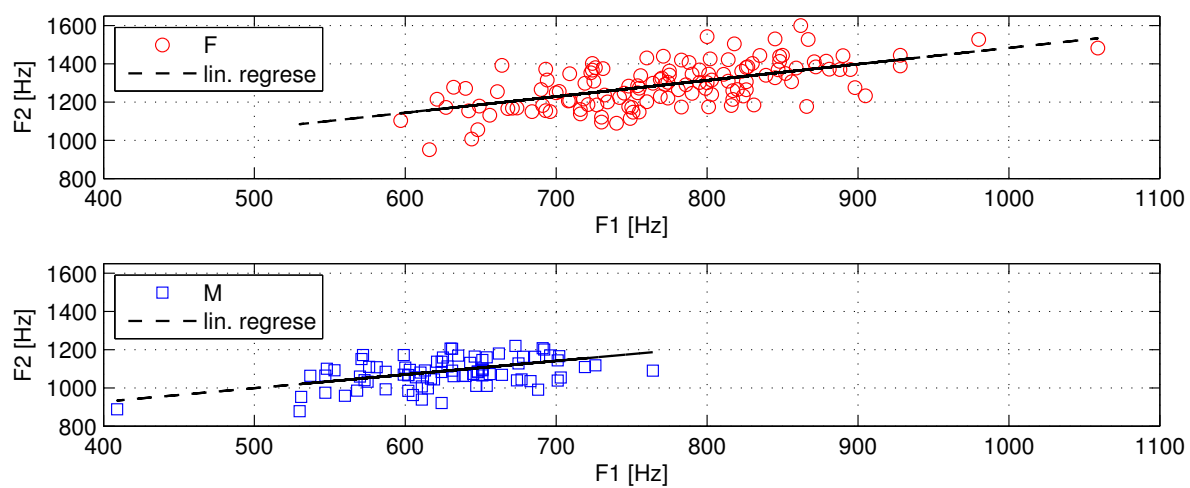


Obrázek 6.5. Pohyb vrcholu vokálního trojúhelníku pro foném /a/ podle věku žen



Obrázek 6.6. Pohyb vrcholu vokálního trojúhelníku pro foném /a/ podle věku mužů

Závěr: pro ženy jsme neodhalili žádnou statisticky významnou závislost mezi hodnotami formantových frekvencí a věkem mluvčí. U mužů existuje klesající trend hodnot F1 a F2 s věkem, díky čemuž dosáhneme největší míry korelace pro parametr F1·F2.



Obrázek 6.7. Hodnoty F1 vůči F2 pro jednotlivé muže a ženy

Kapitola 7

Spektrální momenty

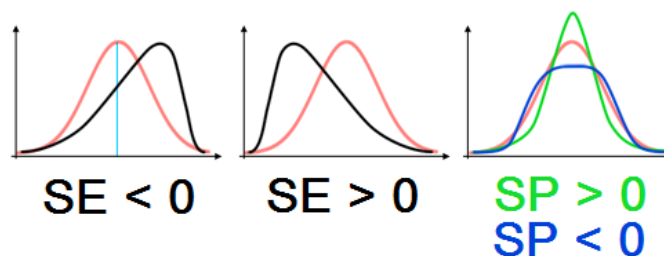
Informace o spektrálních momentech v této kapitole byly čerpány ze zdrojů [18] a [14].

Jedná se o čtveřici parametrů, která je určena z vlastností spektra úseku řečového signálu. Spektrální momenty jsou často používány při analýze charakteristického šumu souhlásek, zejména pak souhlásek neznělých, které kvůli nepřítomnosti základní hlasové frekvence nelze analyzovat pomocí formantů, ani žádného dalšího akustického parametru odvozeného od F_0 diskutovaného v páté kapitole. Pro neznělé souhlásky tak spektrální momenty poskytují možný statistický popis jejich akustických charakteristik.

Prvním spektrálním momentem je **spektrální těžiště** (v angličtině center of gravity), označováno zkratkou **ST**. Spektrální těžiště udává průměrnou frekvenci celé zkoumané frekvenční oblasti a uvádí se proto v hertzech. Ostatní spektrální momenty jsou odvozeny vzhledem ke spektrálnímu těžišti.

Směrodatná odchylka spektra (standard deviation), která je označována zkratkou **SO**, je hodnotou vyjadřující míru odchylky frekvencí spektra od spektrálního těžiště. Například pokud bychom analyzovali spektrum sinusového signálu, které by obsahovalo pouze frekvenci této funkce (která by tím pádem byla hodnotou spektrálního těžiště), pak by hodnota směrodatné odchylky byla rovna nule. U šumových signálů, jakým je i neznělá exploziva /p/ se dá předpokládat, že směrodatná odchylka bude nabývat vysoké hodnoty. I směrodatná odchylka je uváděna v hertzech.

Třetím spektrálním momentem je **sešikmenost hodnot spektra** (skewness), označováno zkratkou **SE**. Hodnota vyjadřuje stupeň asymetrie spektra zkoumané oblasti signálu vzhledem k hodnotě spektrálního těžiště (tento stupeň asymetrie se porovnává vůči Gaussově křivce normální distribuce). Porovnává tedy oblast frekvencí vyšších než je hodnota ST, které se říká pravé křídlo spektra, vůči oblasti frekvencí nižších než je ST, která se označuje jako levé křídlo spektra. SE může nabývat kladných i záporných hodnot. U neznělé explozivy /p/ se dá obecně očekávat hodnota SE kladná, ale výrazně nižší než u znělé části řečového signálu, jehož stupni asymetrie přispívá přítomnost základní hlasové frekvence a formantů ve spektru.



Obrázek 7.1. Ilustrace k hodnotám SE a SP

Posledním spektrálním momentem je **špičatost hodnot spektra** (kurtosis), označována zkratkou **SP**. Jak vyplývá z názvu momentu, hodnota SP vyjadřuje stupeň špičatosti případně plochosti spektra vzhledem ke spektrálnímu těžišti (tento stupeň

špičatosti se porovnává vůči Gaussově křivce normální distribuce). Podobně jako SE může nabývat kladných i záporných hodnot. I u tohoto momentu platí, že hodnota SP u neznělé exploziv /p/ nabývá kladných hodnot, které jsou nižší než hodnoty SP pro znělý úsek signálu. Pro spektrum bílého šumu je hodnota $SP \cong \frac{6}{5}$.

7.1 Spektrální momenty neznělé exploziv /p/

Hodnoty spektrálních momentů neznělé exploziv /p/ byly získány z řečové databáze manuálně za pomoci funkcí programu Praat. Neznělé explozivы obecně patří k jedněm z nejkratších přechodových akustických jevů, byla tedy dána přednost přesnosti zkoumaných segmentů signálu před automatizací sběru dat.

Program Praat definuje hodnotu ST jako

$$ST[\text{Hz}] = \frac{\int_0^{\infty} f |S(f)|^p df}{\int_0^{\infty} |S(f)|^p df} \quad (1)$$

kde $S(f)$ je komplexní spektrum signálu, f je frekvence a p je parametr, jehož běžné hodnoty jsou 2, 1 nebo 2/3. Pro $p = 1$ je vážení průměrné hodnoty frekvence spektra prováděno ve spektru amplitud, pro $p = 2$ pak ve výkonovém spektru. V této práci byla používána hodnota $p = 2$. [14]

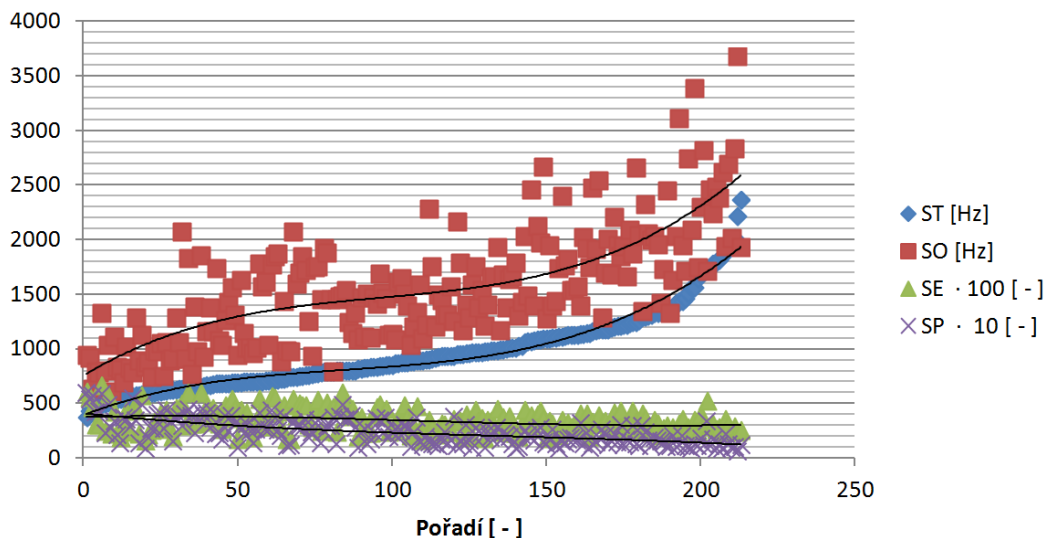
K určení hodnot SO, SE a SP využívá program Praat hodnotu **centrálního spektrálního momentu**. Ten je určen jako

$$CSM = \frac{\int_0^{\infty} (f - ST)^n |S(f)|^p df}{\int_0^{\infty} |S(f)|^p df} \quad (2)$$

kde n udává řád centrálního spektrálního momentu a ST udává hodnotu spektrálního těžiště.

Hodnota směrodatné odchylky spektra je pak stanovena jako druhá odmocnina centrálního spektrálního momentu. Hodnota SE je rovna podílu CSM řádu $n = 3$ s CSM řádu $n = 2$ umocněného na tři poloviny. A nakonec hodnota SP je dána podílem CSM řádu $n = 4$ s druhou mocninou CSM řádu $n = 2$, jehož hodnota je zmenšena o tři.

Hodnoty spektrálních momentů SO, SE a SP jsou skrze výpočet centrálního spektrálního momentu vázány na hodnotu spektrálního těžiště. Proto nejprve prozkoumáme vztahy hodnot ST s hodnotami ostatních spektrálních momentů, bez zohlednění věku jednotlivých mluvčích. Hodnoty spektrálního těžiště byly seřazeny vzestupně a zobrazeny v grafu spolu s odpovídajícími hodnotami ostatních momentů z nich odvozených, viz obr. 7.2. Hodnoty SE a SP byly přenásobeny konstantami, aby bylo možné zobrazit všechny čtyři momenty v jednom grafu. Hodnoty všech čtyř momentů jsou pro



Obrázek 7.2. Zobrazení vzestupně seřazených hodnot spektrálního momentu spektrálního těžiště mužů i žen vůči ostatním spektrálním momentům. Hodnoty jsou proloženy polynomy 3. řádu.

přehlednější zobrazení trendu jejich vývoje aproximovány polynomickou regresí pomocí polynomu 3. řádu.

Na obrázku jsou patrné následující trendy: pokud ST vzrůstá, pak SO vzrůstá, SE klesá a SP klesá. Tedy nabývá-li spektrální těžiště hodnot vyšších frekvencí, pak je sice směrodatná odchylka spektra vyšší, ale spektrum se co se týče míry špičatosti a asymetrie blíží více normální distribuci. V obrázku 7.2 jsou zobrazeny hodnoty získané z promluv žen i mužů, neboť při zobrazení zvlášť byly trendy mezi jednotlivými spektrálními momenty velice podobné. V tabulce 7.1 jsou přiloženy vypočítané hodnoty míry korelace mezi ST a ostatními spektrálními momenty pro muže, ženy i obě skupiny dohromady.

Korelace ST a	SO	SE	SP
Muži	0,718	-0,172	-0,559
Ženy	0,753	-0,359	-0,598
Muži a ženy	0,738	-0,286	-0,579

Tabulka 7.1. Korelace hodnot ST s hodnotami SO, SE a SP

Nyní budeme analyzovat hodnoty spektrálních momentů explozivní fonému /p/ s ohledem na věk a pohlaví mluvčích. Nejprve prozkoumáme míru korelace jednotlivých momentů vůči věku v tabulce 7.2

Korelace věku a	ST	SO	SE	SP
Muži	-0,175	-0,166	-0,222	-0,119
Ženy	0,051	0,043	-0,265	-0,246

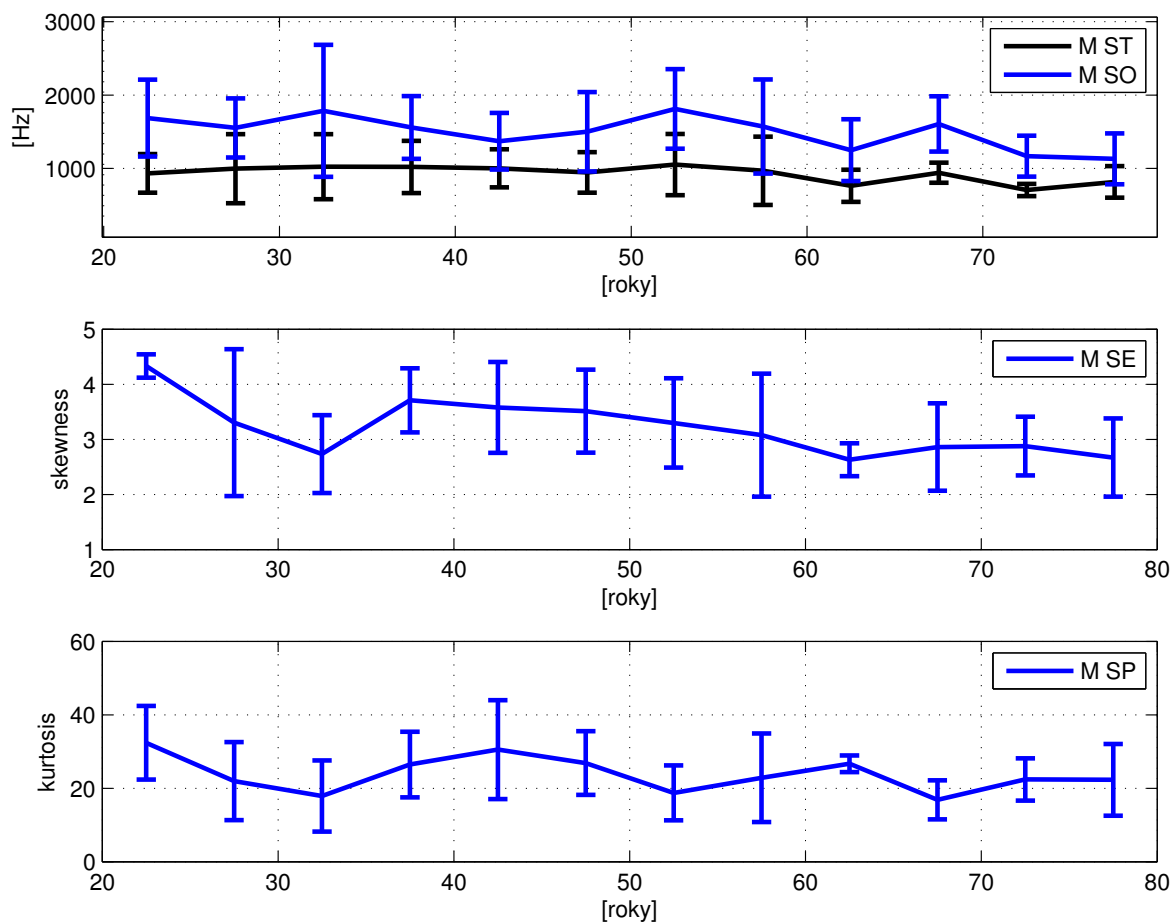
Tabulka 7.2. Korelace hodnot spektrálních momentů s věkem mluvčích

V grafech 7.3 a 7.4 jsou střední hodnoty a směrodatné odchylky jednotlivých spektrálních momentů neznělé explozivní fonémy /p/ vztaheny k věkovým intervalům mluvčích po

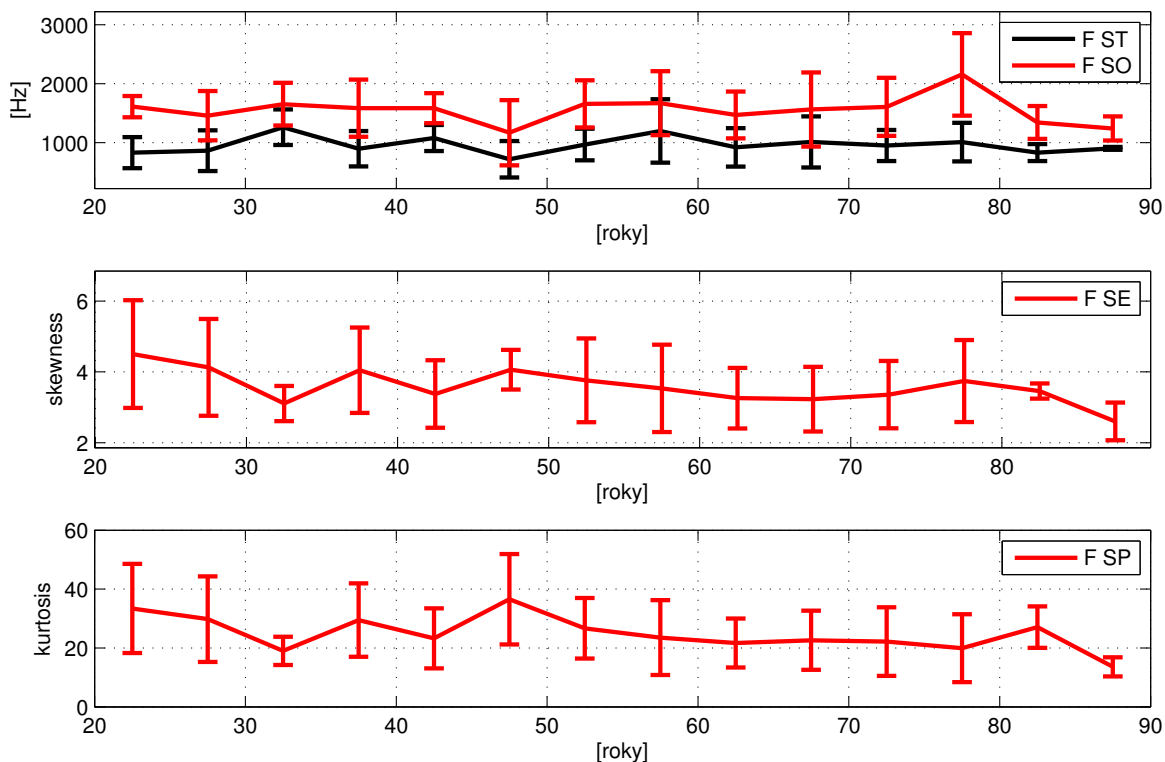
Korelace věku a	ST	SO	SE	SP
Ženy 20-43 let	0,146	-0,042	-0,087	-0,033
Ženy 44-56 let	0,601	0,533	0,020	-0,272
Ženy 57-90 let	-0,014	0,073	0,012	-0,058

Tabulka 7.3. Korelace věku žen v intervalech podle věku

pěti letech. V grafech je dobře patrná závislost momentů vůči spektrálnímu těžišti, která již byla v této kapitole diskutována. Spektrální moment SP má tendenci u obou pohlaví po 45 letech mluvčích klesat. Celkově závislosti spektrálních momentů neznělé explozivy /p/ nejsou silné. Pro budoucí zkoumání problematiky by bylo vhodné se mluvčích dotázat, zda a případně jak dlouho využívají zubní náhradu. U starších mluvčích je vyšší pravděpodobnost používání takové náhrady a zejména než si mluvčí na náhradu zvykne, je možné, že ovlivní jeho artikulaci, což se projeví na výsledcích měření hodnot spektrálních momentů souhlásek.



Obrázek 7.3. Střední hodnoty a směrodatné odchylky spektrálních momentů neznělé explozivy /p/ u mužů



Obrázek 7.4. Střední hodnoty a směrodatné odchytky spektrálních momentů neznělé explozivy /p/ u mužů

7.2 Spektrální momenty znělé části signálu

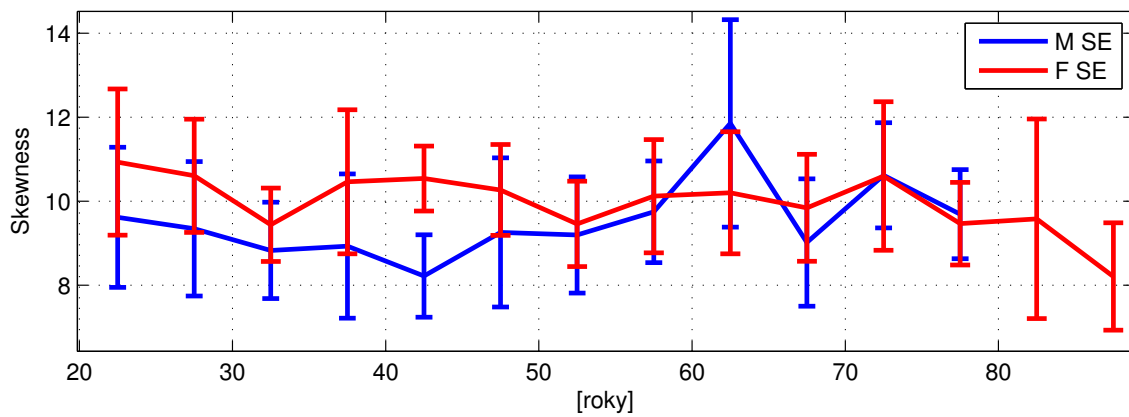
Byť jsou spektrální momenty běžně používány především pro analýzu charakteristického šumu souhlásek, dají se určit pro libovolný úsek signálu. Vzhledem k existujícím závislostem, mezi věkem mluvčího a hodnotami základní hlasové frekvence F_0 a formantů, prozkoumaných v předcházejících kapitolách, se dá předpokládat, že i spektrální momenty určené ze znělých úseků řečového signálu budou mít určitou závislost na věku. Hodnoty SE a SP byly určeny v programu MATLAB a jejich střední hodnoty a směrodatné odchytky na intervalech po pěti letech věku mluvčích jsou zobrazeny v grafech na obrázcích 7.5 a 7.6.

Stejně jako u spektrálních momentů neznělé explozivy /p/ můžeme i u znělého signálu pozorovat podobnost tendencí dvojice průběhů hodnot SE a SP. Dále můžeme prozkoumat korelaci hodnot SE a SP s hodnotami F_0 pro muže i ženy v tabulce 7.4.

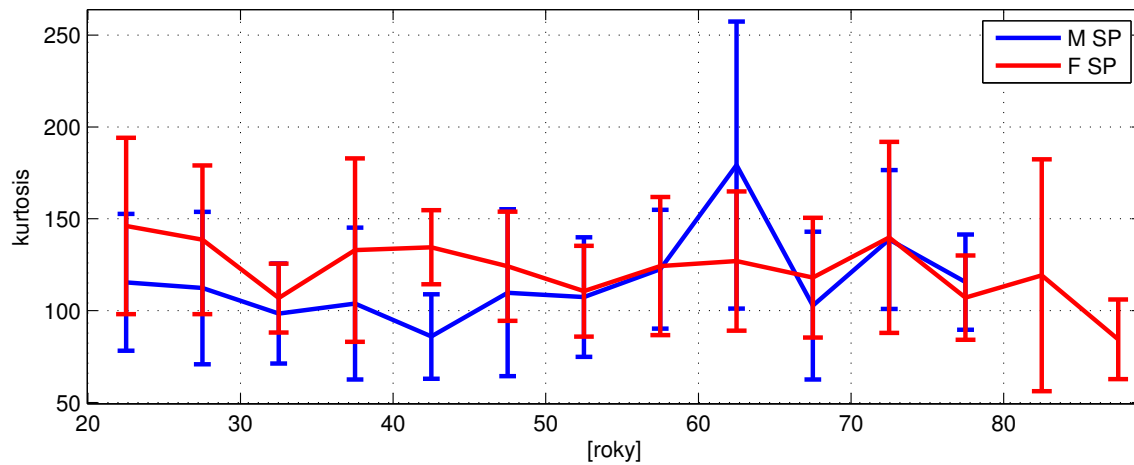
Korelace F_0 a	SE	SP
Muži	0,338	0,267
Ženy	0,248	0,201

Tabulka 7.4. korelace hodnot spektrálních momentů s F_0 mluvčích

A v tabulce 7.5 jsou hodnoty míry korelace mezi formanty mluvčích a spektrálními momenty SE a SP znělého úseku signálu.



Obrázek 7.5. Střední hodnoty a směrodatné odchytky SE ze znělého úseku signálu



Obrázek 7.6. Střední hodnoty a směrodatné odchytky SP ze znělého úseku signálu

Korelace F1 a	SE	SP
Muži	-0,225	-0,228
Ženy	-0,257	-0,225
Korelace F2 a	SE	SP
Muži	-0,158	-0,162
Ženy	-0,095	-0,053
Korelace F3 a	SE	SP
Muži	0,063	0,074
Ženy	0,063	0,075

Tabulka 7.5. Korelace hodnot spektrálních momentů s formanty mluvčích

Korelace věku a	SE	SP
Muži	0,219	0,187
Ženy	-0,146	-0,112

Tabulka 7.6. Korelace hodnot spektrálních momentů ze znělého signálu s věkem mluvčích

Nakonec v tabulce 7.6 zhodnotíme míry korelace hodnot SE a SP získaných ze znělých úseků signálu vůči věku mluvčích. Z hodnot je patrné, že určitá závislost mezi veličinami existuje, není však příliš silná.

Kapitola 8

Klasifikace věku mluvčího na základě jeho akustických charakteristik

Nyní máme k dispozici hodnoty akustických charakteristik mluvčích použité řečové databáze. Ze získaných dat je patrné, že kvůli nepříliš silným závislostem jednotlivých parametrů na věku mluvčího bude velice obtížné použít tyto parametry k přesné klasifikaci věku. Při pokusu o tvorbu modelového regresního stromu pro klasifikaci žen do věkových intervalů po pěti letech, byl výsledný strom pouze jediný list, který zařadil veškeré ženy do kategorie 65 až 69 let. Patrně vzhledem k slabým závislostem jednotlivých parametrů, bylo nejmenší klasifikační chyby docíleno automatickým klasifikováním žen do nejčastěji zastoupené věkové kategorie v databázi.

Z důvodů nedostatečně vyváženého zastoupení mluvčích v rámci jednotlivých věkových kategorií a protože většina akustických charakteristik se statisticky významnou závislostí na věku byla zjištěna u žen, a to často s prudkým poklesem případně nárůstem hodnot mezi čtyřicátým a padesátým rokem, byl provedený klasifikační pokus zaměřen na rozpoznání žen mladších 43 let a starších 57 let (tedy na skupinu žen ve věku před nástupem menopauzy a na skupinu žen ve věku která jí již prošla). Skupinu žen, která spadá do věkové kategorie mezi 43 a 57 let, nebudeme při tvorbě klasifikačního stromu využívat, protože nemůžeme s relativní jistotou označit, do které ze dvou skupin patří. Skupinu žen ve věku do 43 let označíme zkratkou PRE a skupinu žen ve věku nad 57 let označíme zkratkou POST.

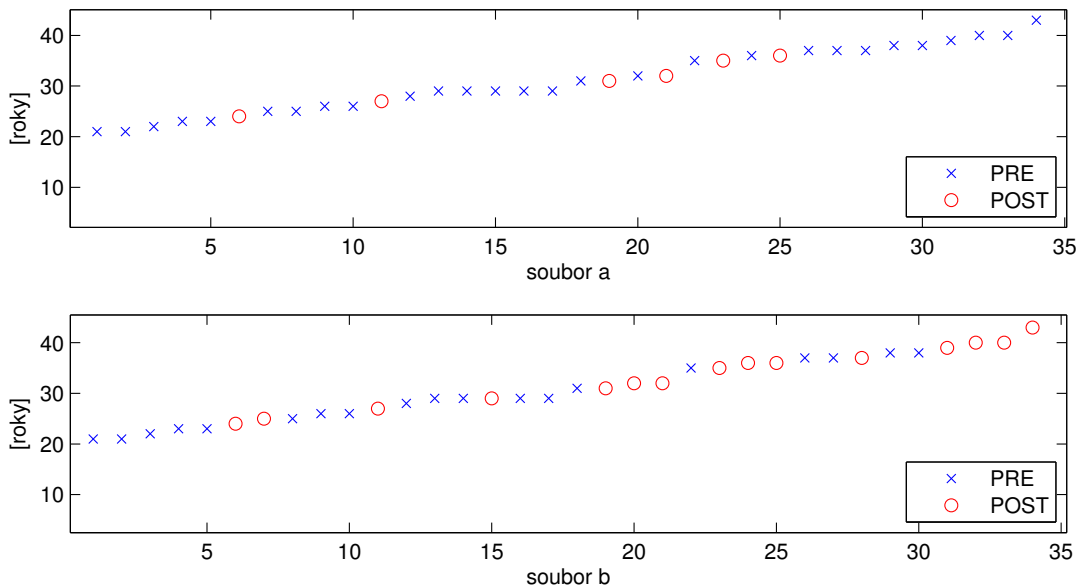
Rozhodovací strom prezentovaný v této práci byl vytvořen v programu Rapid Miner. Data použitá k tvorbě stromu obsahovala hodnoty parametrů získané ze všech nahrávek žen ve věku do 43 let s koncovkou *a* a části nahrávek žen nad 57 let s koncovkou *a*. Hodnoty získané z nahrávek s koncovkou *b* v názvu při tvorbě použity nebyly vůbec. Z kategorie žen nad 57 let byly vypuštěny některé záznamy, aby došlo k menší nevyváženosti dvou klasifikovaných skupin. Záznamy nepoužité ke klasifikaci byly vybrány náhodně z nejčastěji zastoupených věkových kategorií. Strojové učení bylo tedy provedeno na vzorku dat získaných od 35 žen kategorie PRE a 72 žen kategorie POST. Na obrázku 8.1 vidíme výsledný rozhodovací strom vytvořený programem Rapid Miner z poskytnutých dat. Strom využívá ke klasifikaci následujících akustických charakteristik: Fundamentální frekvence F_0 (získaná pomocí autokorelace), Jitter(local,abs), Jitter(rap), Shimmer(local) a hodnotu SE získanou ze znělé části signálu.

Nyní klasifikační strom implementujeme v MATLABu a použijeme ho ke klasifikaci dat ze všech dostupných nahrávek v databázi. Zajímavé bude pozorovat zejména výsledky klasifikace pro nahrávky z databáze označené koncovkou *b*, z nichž žádné nebyly použity k tvorbě stromu. Na obrázku 8.2 jsou dva grafy. Osy *x* grafů značí pouze pořadová čísla nahrávek v řečové databázi. Osa *y* zobrazuje věk mluvčího na klasifikované nahrávce. Pokud jsou akustické parametry nahrávky klasifikovány stromem do skupiny PRE, pak je v grafu zobrazen modrý křížek. Pokud parametry spadají do skupiny POST, potom je vyneseno červené kolečko. Nahoře je graf klasifikace nahrávek s kon-

```

Jitter Local Absolute E-6 seconds > 13.430
| jitter rap [%] > 0.032
| | F0 autokorelace [Hz] > 235.400: Pre {Pre=2, Post=0}
| | F0 autokorelace [Hz] ≤ 235.400
| | | jitter rap [%] > 0.093
| | | | shimmer apq11 [%] > 11.736: Pre {Pre=1, Post=1}
| | | | shimmer apq11 [%] ≤ 11.736: Post {Pre=1, Post=44}
| | | jitter rap [%] ≤ 0.093
| | | | skewness (znela) > 8.832
| | | | | skewness (znela) > 9.678
| | | | | shimmer local [%] > 11.094: Post {Pre=0, Post=5}
| | | | | shimmer local [%] ≤ 11.094
| | | | | F0 autokorelace [Hz] > 171.450
| | | | | | skewness (znela) > 11.959: Post {Pre=0, Post=3}
| | | | | | skewness (znela) ≤ 11.959: Pre {Pre=15, Post=6}
| | | | | F0 autokorelace [Hz] ≤ 171.450: Post {Pre=0, Post=3}
| | | | | skewness (znela) ≤ 9.678: Post {Pre=0, Post=10}
| | | | | skewness (znela) ≤ 8.832: Pre {Pre=5, Post=0}
| jitter rap [%] ≤ 0.032: Pre {Pre=2, Post=0}
Jitter Local Absolute E-6 seconds ≤ 13.430: Pre {Pre=9, Post=0}
    
```

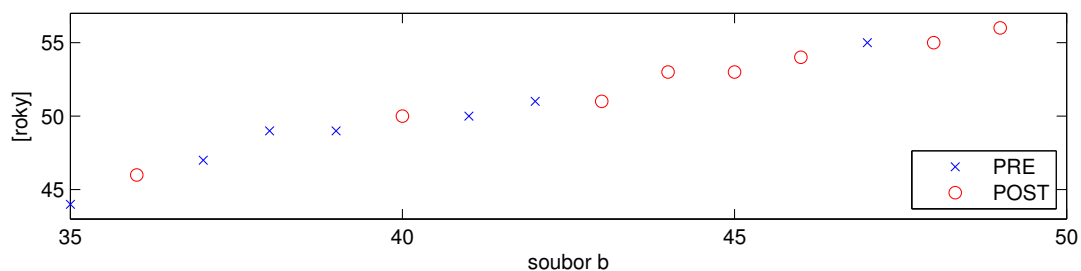
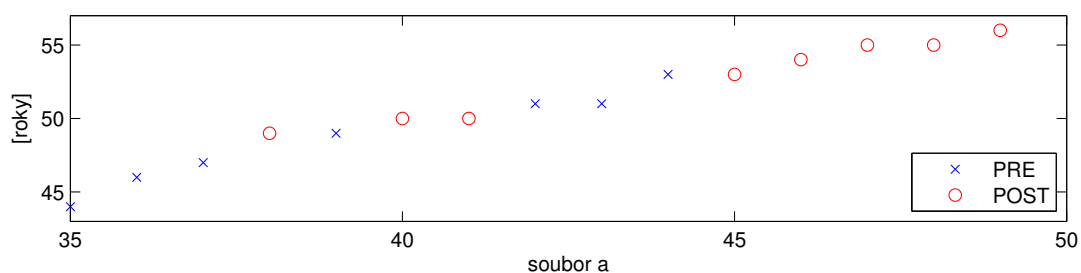
Obrázek 8.1. Klasifikační strom, rozdělující ženy do dvou věkových kategorií - PRE(před menopauzou) a POST(po menopauze)



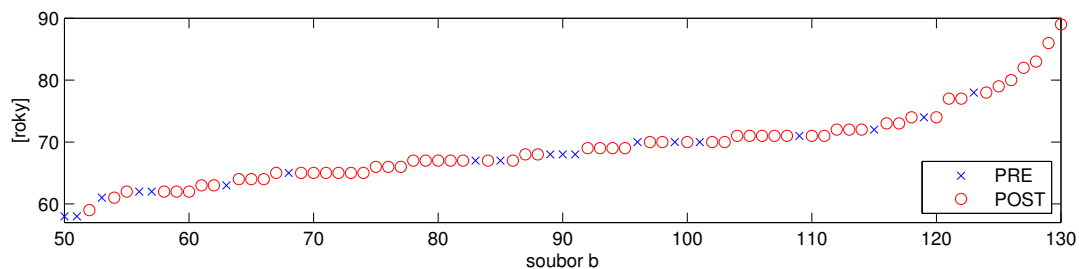
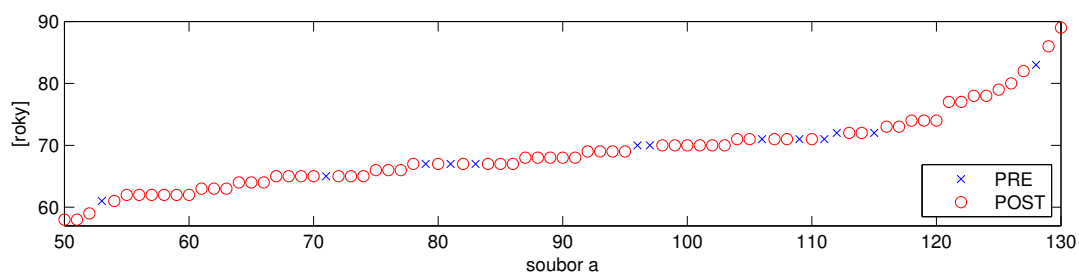
Obrázek 8.2. Výsledek klasifikace skupiny žen do 43 let, modrý křížek značí klasifikaci do věkové skupiny před menopauzou, červené kolečko do skupiny po menopauze. V prvním grafu jsou výsledky pro hodnoty nahrávek s koncovkou *a*, ve druhém pro koncovku *b*.

covkami *a*, pod ním je graf z nahrávek s koncovkami *b*. Pro ideální klasifikační strom by tento obrázek obsahoval pouze modré křížky.

Na obrázku 8.3 jsou zobrazeny výsledky klasifikačního experimentu pro ženy ve věku mezi 43 a 57 lety. A na obrázku 8.4 jsou výsledky klasifikačního experimentu pro ženy ve věku nad 57 let. Pro ideální klasifikační strom by tento graf obsahoval pouze červená kolečka.



Obrázek 8.3. Výsledek klasifikace skupiny žen ve věku od 43 do 57 let.



Obrázek 8.4. Výsledek klasifikace skupiny žen nad 50 let

V tabulkách 8.1, 8.2 a 8.3 pak zhodnotíme úspěšnost klasifikačního experimentu. Je patrné, že klasifikace dosahuje lepší přesnosti pro data, na kterých byl strom vytvořen. V ideálním případě, kdy by byl strom tvořen z dostatečně velkého množství dat, by k tomuto jevu nemělo docházet. Naznačuje to totiž, že strom tvoří větve jen aby klasifikoval několik specifických krajních případů. Při použití toho samého stromu na jiných datech pak dochází k chybám. S přihlédnutím na velikost vzorku trénovacích dat je však výsledný klasifikační strom uspokojivý.

Klasifikováno jako:	PRE	POST	Úspěšnost klasifikace kategorie
Kategorie PRE	28	6	82,35%
Kategorie POST	13	68	83,95%
Celková úspěšnost klasifikace	83,48%		

Tabulka 8.1. Úspěšnost pokusu klasifikace na nahrávkách s příponou *a*

Klasifikováno jako:	PRE	POST	Úspěšnost klasifikace kategorie
Kategorie PRE	19	15	55,88%
Kategorie POST	19	62	76,54%
Celková úspěšnost klasifikace	70,43%		

Tabulka 8.2. Úspěšnost pokusu klasifikace na nahrávkách s příponou *b*

Klasifikováno jako:	PRE	POST	Úspěšnost klasifikace kategorie
Kategorie PRE	47	21	69,11%
Kategorie POST	32	130	80,25%
Celková úspěšnost klasifikace	76,96%		

Tabulka 8.3. Úspěšnost pokusu klasifikace na nahrávkách s příponou *a* i *b*

Kapitola 9

Závěr

V práci byly prozkoumány hodnoty vybraných akustických parametrů a míra jejich závislosti na věku mluvčích. Dále byly diskutovány možné fyzikální a biologické příčiny získaných výsledků.

Nejdříve jsme v kapitole 4 prováděli hruhou segmentaci řečového signálu. Díky tomu jsme zjistili časy v nahrávkách databáze, ve kterých můžeme analyzovat akustické charakteristiky signálu.

Kapitola 5 byla zaměřena na získání a následnou analýzu akustických charakteristik, které jsou spjaty s procesem tvorby zvuku. Mezi veličiny s nejsilnější závislostí na věku mluvčích se řadila hodnota fundamentální hlasové frekvence F_0 žen a parametr kolísání amplitud zvukového signálu shimmer (nejsilnější závislost měla varianta parametru shimmer(apq3)) mužů i žen. Dále pak byla závislost na věku zjištěna u parametru kolísání délky period zvukového signálu jitter a to zejména u žen nad 40 let. A u žen byla také zjištěna klesající tendence parametru poměru harmonických kmitočtů k šumu HNR.

V kapitole 6 jsme se věnovali parametrům popisujícím rezonanční vlastnosti řečového ústrojí mluvčích, neboli formantovými frekvencemi. Významnou závislost parametru na věku jsme zjistili pouze u $F_1 \cdot F_2$ mužů, což je parametr získaný vynásobením hodnoty F_1 hodnotou F_2 . Nakonec jsme se v kapitole 7 zaobírali spektrálními momenty, tedy parametry, které se určují z vlastností spektra úseku řečového signálu. Ty umožňují statistický popis charakteristického šumu souhlásky /p/. U mužů i žen byla zjištěna závislost parametru spektrálního momentu SE na věku.

V kapitole 8 jsou získané hodnoty jednotlivých akustických charakteristik využity k pokusu zpětné klasifikace věku mluvčích. Protože většina akustických charakteristik se statisticky významnou závislostí na věku byla zjištěna u žen, a to často s prudkým poklesem případně nárůstem hodnot mezi čtyřicátým a padesátým rokem, byl klasifikační pokus zaměřen na rozpoznání žen mladších 43 let a starších 57 let.


Pro budoucí zkoumání problematiky by bylo dobré pracovat s databází, která odráží zastoupení mužů a žen různých stáří v populaci. Kvůli nevyváženosti databáze a nedostatečném zastoupení některých věkových kategorií je pak výsledné trendy nutné brát s jistou rezervou. Nevíme totiž, zda jsou přítomné i pro širší populaci, nebo jde o náhodný jev přítomný pouze u zkoumané skupiny mluvčích. Navíc při dostatečném počtu jedinců by bylo možné charakteristiky vázat ke konkrétnímu věku mluvčího, namísto agregace do pětiletých intervalů, které se používají v této práci.

Literatura

- [1] J. Psutka L. Müller J. Matoušek V. Radová. *Mluvíme s počítačem česky*. Academia, 2006. ISBN 80-200-1309-1.
- [2] CSc. Doc. Ind. Milan Sigmund. *Analýza Řečových Signálů*. FEI VUT v Brně, 2000.
- [3] Lenka Podloucká. *Identifikace pauz v rušeném řečovém signálu*. Diplomová práce, Fakulta elektrotechniky a komunikačních technologií, VUT Brno, 2008.
- [4] R. Čmejla M. Novotný, J. Ruzs. *Automatické hledání významných pozic v Parkinsonových promluvách založených na rychlém opakování slabik /pa/-/ta/-/ka/*. ČVUT, FEL Katedra teorie obvodů,
- [5] R. Čmejla J. Ruzs J. Behenská T. Tykalová. *Dlouhodobá studie věkově závislých akustických charakteristik řeči*. Akustické listy, ročník 22, číslo 1-4, 3-8,
- [6] Sameep Kadakia, Dave Carlson a Robert T. Sataloff. *The Effect of Hormones on the Voice*. Journal of Singing, May/June 2013 Volume 69, No. 5, pp. 571–574, National Association of Teachers of Singing,
- [7] Maria Kaczmarek. *Variation in Age at Natural Menopause among Polish Women in Relation to Biological and Social Factors*. Department of Human Biological Development, Adam Mickiewicz University, Poznań, Poland,
- [8] Andropause. In: Encyclopedia of Genetics, Genomics, Proteomics and Informatics. Springer, Dordrecht, 2008,
- [9] The National Center for Voice a Speech.
<http://www.ncvs.org/ncvs/tutorials/voiceprod/tutorial/influence.html>, čerpáno 27.11 2017,
- [10] John Eulenberg. *Fundamental Frequency and the Glottal Pulse*. Department of Communicative Sciences and Disorders, Michigan State University, https://msu.edu/course/asc/232/study_guides/F0_and_Glottal_Pulse_Period.html čerpáno 6.12 2017,
- [11] UCL Department of Phonetics a Linguistics. *Speech Signal Analysis*.
<http://www.phon.ucl.ac.uk/courses/spsci/matlab/lect10.html>, čerpáno 24.11 2017,
- [12] Jan Janda. *Posuzování logopedického věku dítěte*. Disertační práce, Fakulta elektrotechnická, ČVUT Praha, 2012.
- [13] Josef Štěpán Karel Zvára. *Pravděpodobnost a matematická statistika*. matfyzpress, 2012. ISBN 978-80-7378-218-4.
- [14] *Tutorial for voice analysis with Praat*.
<http://www.fon.hum.uva.nl/praat/manual> čerpáno 8.11 2017,
- [15] Dr. Will Styler. *Using Praat for Linguistic Research*.
<http://savethevowels.org/praat>, 2017.
- [16] Niranjan U. Cholayya Kumara Shama, Anantha krishna. *Study of Harmonics-to-Noise Ratio and Critical-Band Energy Spectrum of Speech as Acoustic Indicators*

of Laryngeal and Voice Pathology. Hindawi Publishing Corporation, EURASIP Journal on Advances in Signal Processing, Volume 2007, Article ID 85286, 9 pages, doi:10.1155/2007/85286,

- [17] Ian Vince McLoughlin. *Speech and audio processing, A MATLAB-based approach*. Cambridge University Press, 2016. ISBN 978-1-107-08546-6.
- [18] Pavel Machač. *Temporální a spektrální struktura českých explozí*. Disertační práce, Univerzita Karlova v Praze, Filozofická fakulta, Fonetický ústav, 2006.



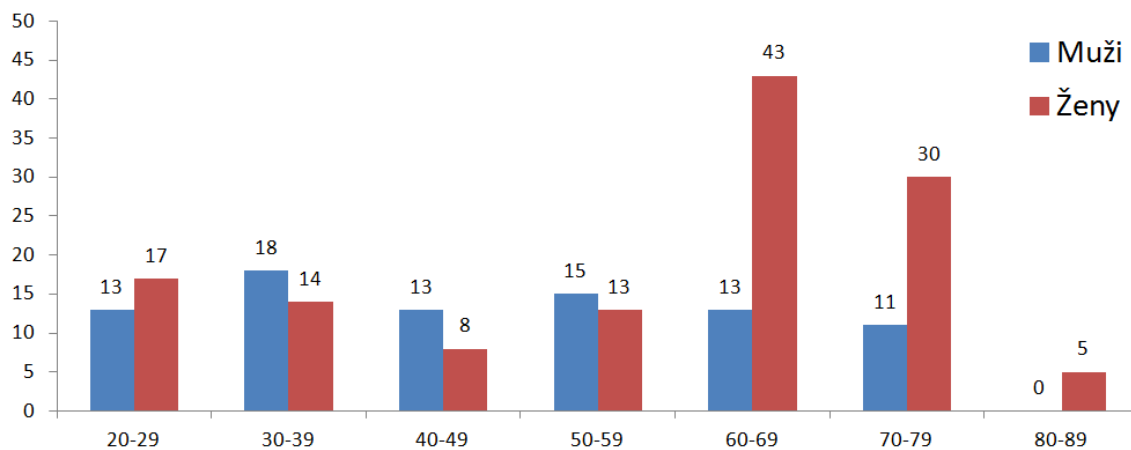
Příloha A

Zkratky a symboly

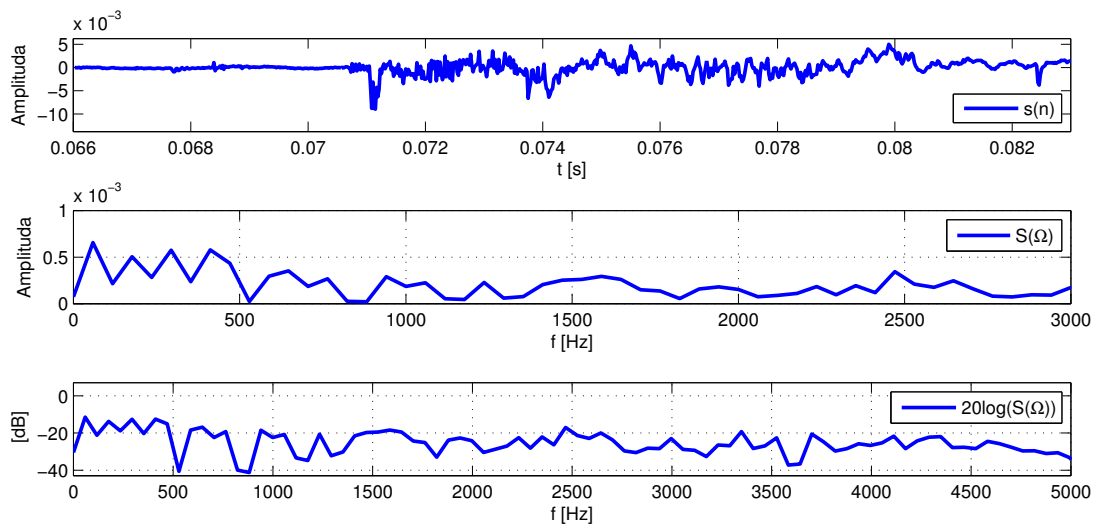
apq	Koeficient poruchy amplitudy
DFT	Diskrétní Fourierova transformace
F	Ženy (females)
FFT	Rychlá Fourierova transformace
HNR	Poměr harmonických kmitočtů k šumu (harmonics-to-noise ratio)
LP	Lineární predikce
M	Muži (males)
ppq5	Pětibodový koeficient poruchy periody
rap	Relativní průměrná porucha
SNR	Poměr signálu a šumu

Příloha B

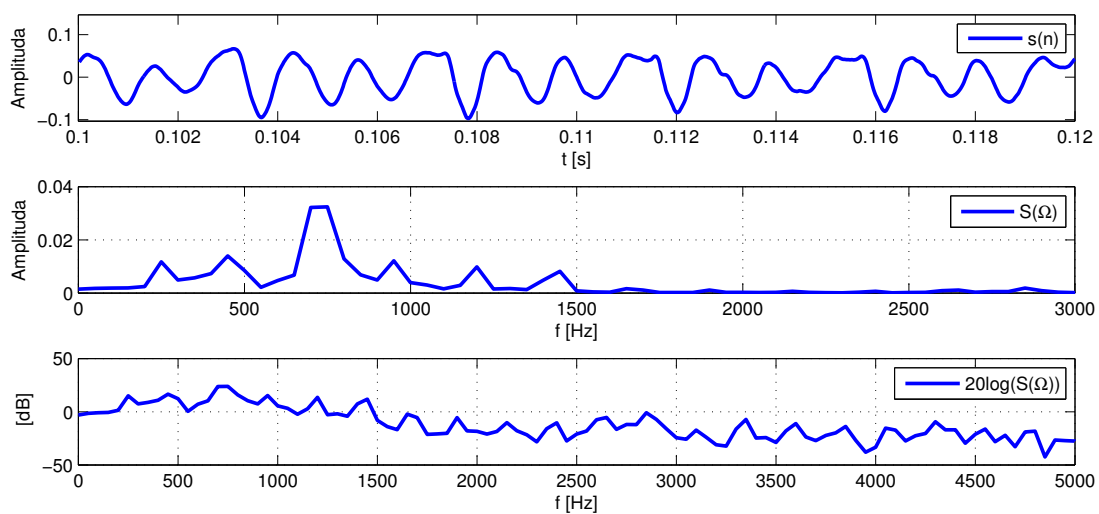
Grafy



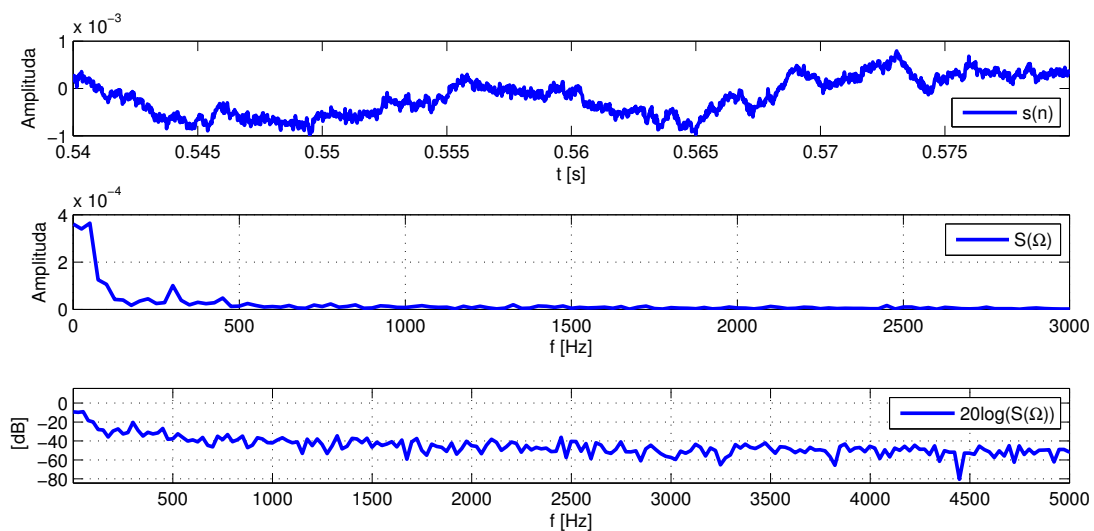
Obrázek B.1. Histogram počtů mužů a žen v řečové databázi v intervalech po deseti letech



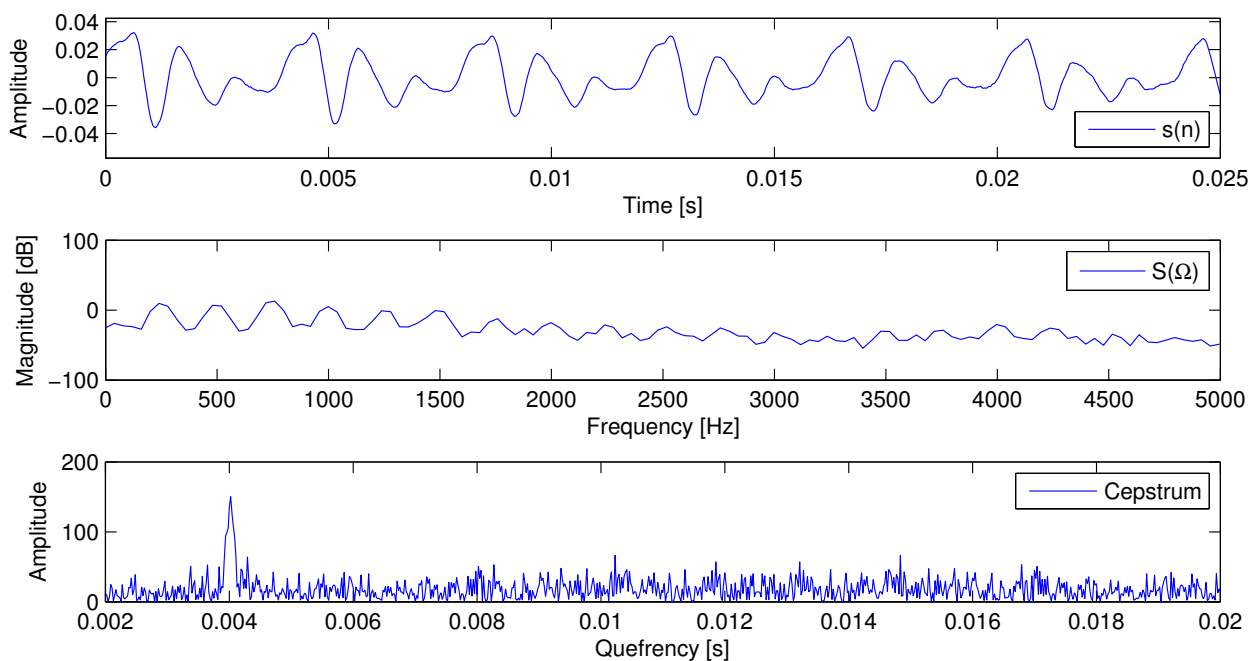
Obrázek B.2. Počáteční exploze fonému /p/ v časové oblasti a v amplitudovém spektru. Použito pravoúhlé okénko



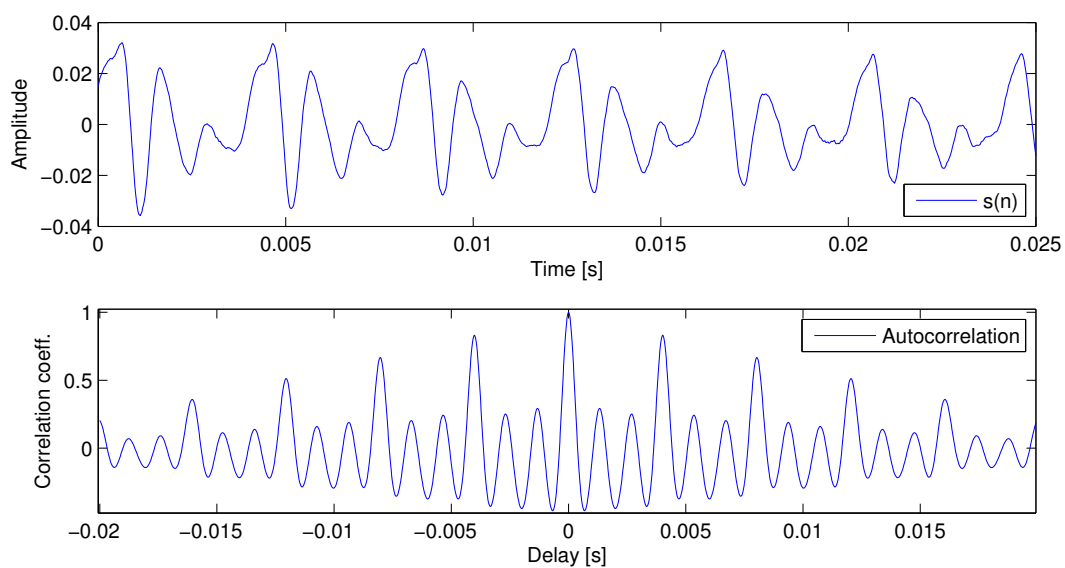
Obrázek B.3. Průběh řečového signálu fonému /a/ v časové oblasti a v amplitudovém spektru. Použito pravoúhlé okénko



Obrázek B.4. Průběh řečového signálu v neznělém úseku mezi slabikami v časové oblasti a amplitudovém spektru. Použito pravoúhlé okénko



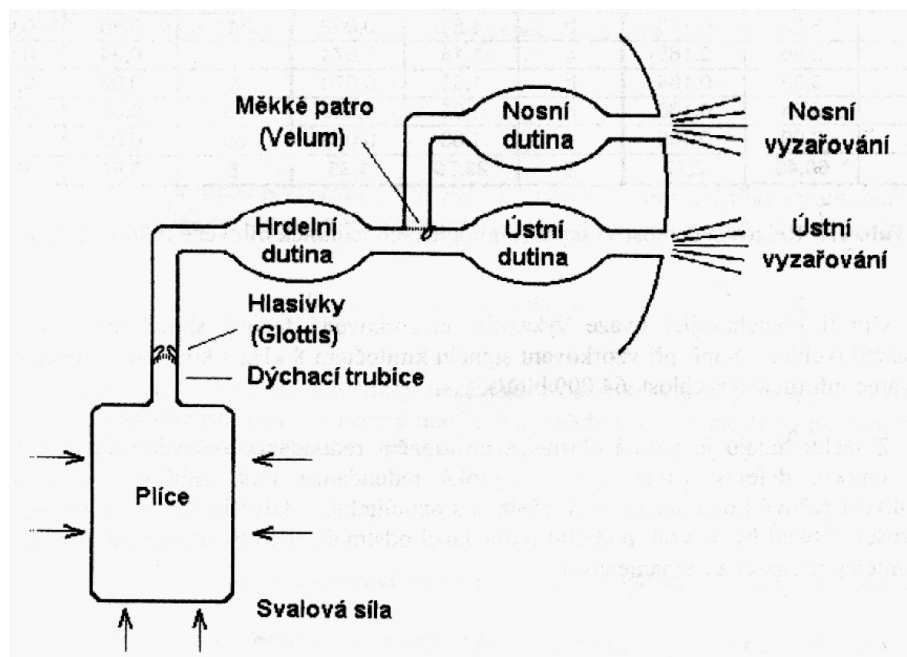
Obrázek B.5. Postup získání hodnoty F0 z kepra



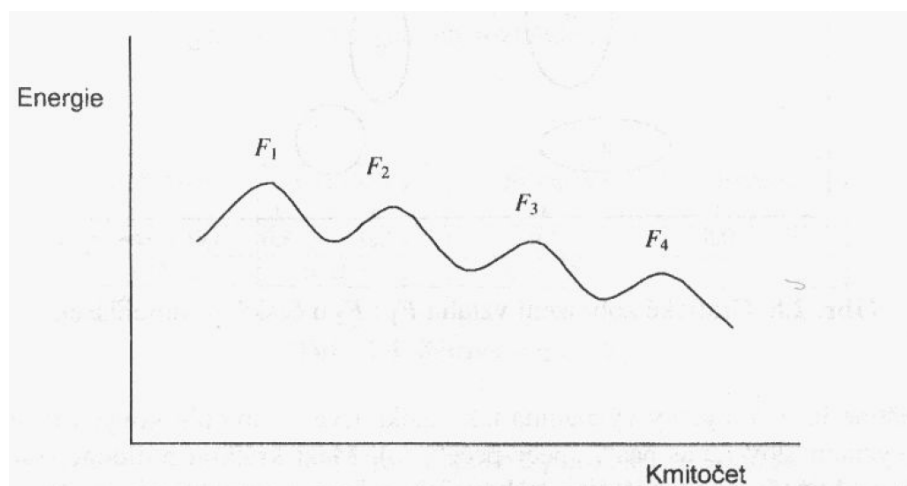
Obrázek B.6. Postup získání hodnoty F0 pomocí autokorelace

Příloha C

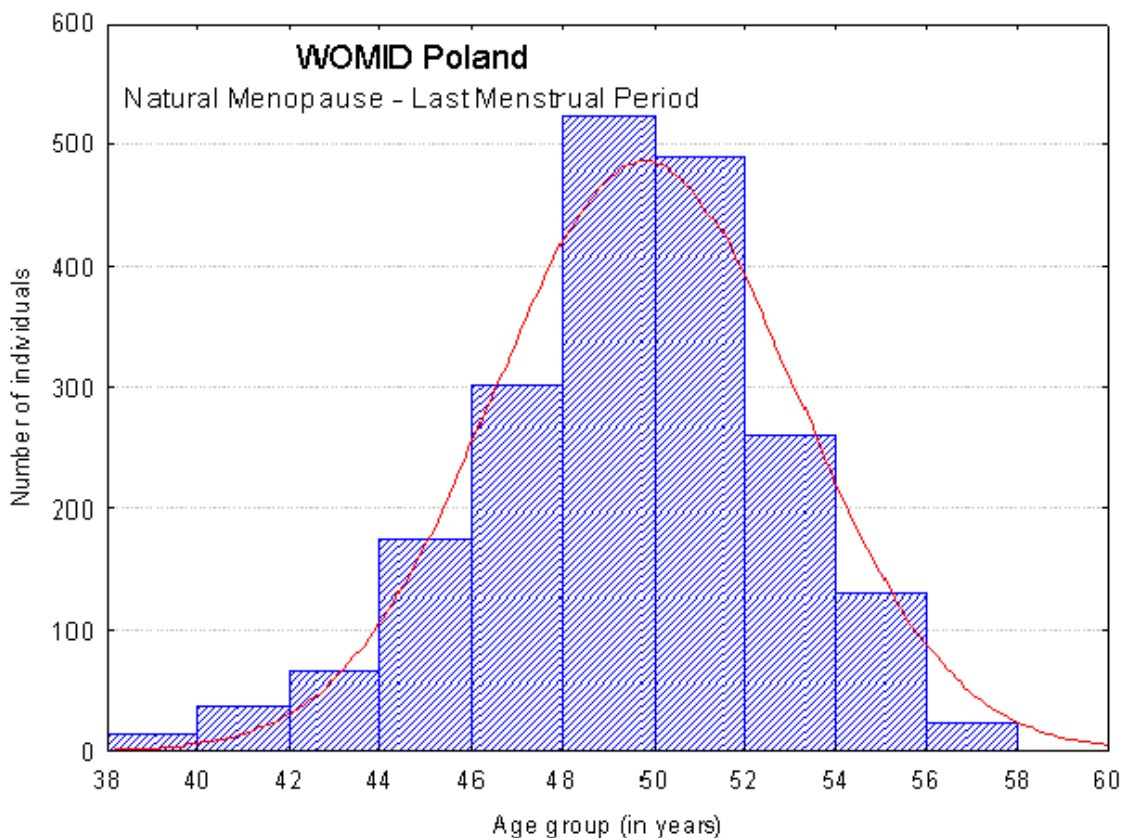
Doplňující ilustrace



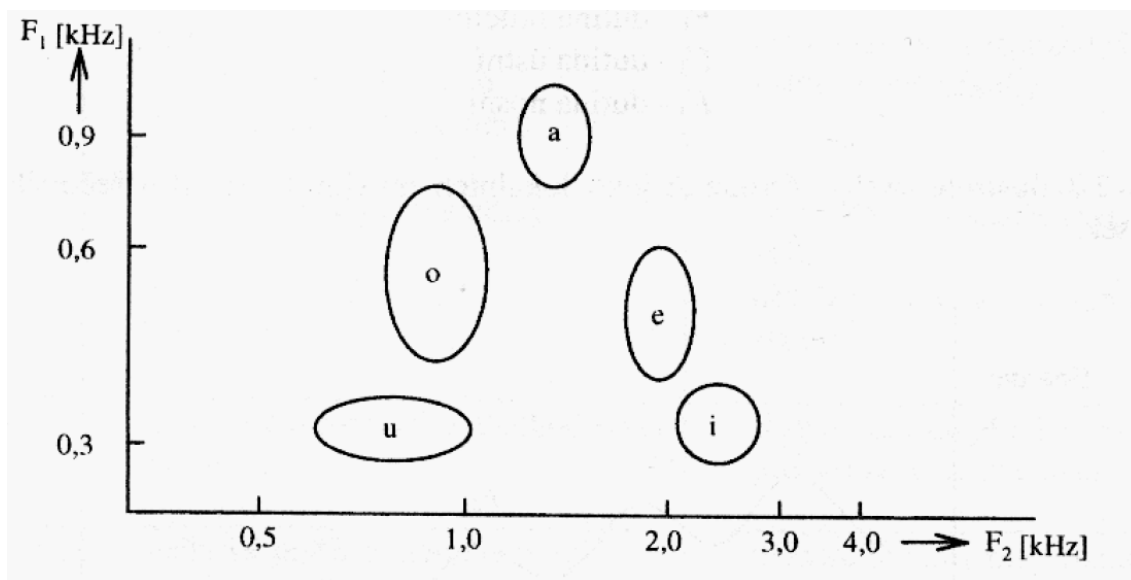
Obrázek C.7. Schématické znázornění lidského hlasového ústrojí, převzato z [2]



Obrázek C.9. Formanty ve spektru samohlásek. Převzato z [2]



Obrázek C.8. Distribuce věku poslední menstruace u polských žen po dvouletých intervalech. Převzato z [7]



Obrázek C.10. Grafické zobrazení vztahu F₁ : F₂ u českých samohlásek, převzato z [2]