



**ČESKÉ VYSOKÉ
UČENÍ TECHNICKÉ
V PRAZE**

F3

**Fakulta elektrotechnická
Katedra řídicí techniky**

Bakalářská práce

Akustické charakteristiky hlasu v závislosti na věku

Jana Zázvorková
Kybernetika a robotika

2020

Vedoucí práce: Ing. Tomáš Bořil Ph.D.

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Zázvorková** Jméno: **Jana** Osobní číslo: **474400**
Fakulta/ústav: **Fakulta elektrotechnická**
Zadávající katedra/ústav: **Katedra řídicí techniky**
Studijní program: **Kybernetika a robotika**

II. ÚDAJE K BAKALÁŘSKÉ PRÁCI

Název bakalářské práce:

Akustické charakteristiky hlasu v závislosti na věku

Název bakalářské práce anglicky:

Age-dependent acoustic features of speech

Pokyny pro vypracování:

- 1) Seznamte se s problematikou akustických charakteristik hlasu se zaměřením na ty, které by mohly záviset na věku, proveďte rešerši a shrňte důležité principy.
- 2) Zpracujte dříve pořízené nahrávky řeči zdravých osob v rozmezí 20 až 80 let věku - úloha "čtený text".
- 3) Vyberte a změňte vhodné akustické charakteristiky v závislosti na věku.
- 4) Vyhodnoťte, zda jsou jednotlivé parametry použitelné pro účely plně automatické analýzy nahrávek, nebo zda vyžadují kroky manuálního zpracování.

Seznam doporučené literatury:

- [1] Eichhorn JT, Kent RD, Austin D, Vorperian HK. Effects of Aging on Vocal Fundamental Frequency and Vowel Formants in Men and Women. *J Voice*. 2018;32(5).
- [2] Li M, Han KJ, Narayanan S. Automatic speaker age and gender recognition using acoustic and prosodic level information fusion. *Comput Speech Lang*. 2013;27(1):151-67.
- [3] Volín, J., Tykalová, T., Bořil, T. (2017). Stability of prosodic characteristics across age and gender groups. In *Proceedings of Interspeech 2017*, pp. 3902–3906.
- [4] Volín, J., Bořil, T. (2017). Acoustic correlates of prosodic dimensions in younger and older speakers of Czech. *Acta Universitatis Carolinae – Philologica 3/2017, Phonetica Pragensia XIV*, pp. 45–54.
- [5] Xue SA, Hao GJP. Changes in the human vocal tract due to aging and the acoustic correlates of speech production: A pilot study. *J Speech Lang Hear R*. 2003;46(3):689-701.
- [6] Cox VO, Selent M. Acoustic and Respiratory Measures as a Function of Age in the Male Voice. *J Phonet and Audiol*. 2015;1:105.
- [7] Fletcher AR, McAuliffe MJ, Lansford KL, Liss JM. The relationship between speech segment duration and vowel centralization in a group of older speakers. *J Acoust Soc Am*. 2015;138(4):2132-9.
- [8] Harnsberger JD, Shrivastav R, Brown WS, Rothman H, Hollien H. Speaking rate and fundamental frequency as speech cues to perceived age. *J Voice*. 2008;22(1):58-69.
- [9] Torre P, Barlow JA. Age-related changes in acoustic characteristics of adult speech. *J Commun Disord*. 2009;42(5):324-33.
- [10] 25. Harrington J, Palethorpe S, Watson CI. Age-related changes in fundamental frequency and formants: a longitudinal study of four speakers. *Interspeech 2007: 8th Annual Conference of the International Speech Communication Association, Vols 1-4*. 2007:1081.
- [11] Kent RD, Vorperian HK. Static measurements of vowel formant frequencies and bandwidths: A review. *J Commun Disord*. 2018;74:74-97.
- [12] Sapir, S., Ramig, L. O., Spielman, J. L., & Fox, C. (2010). Formant centralization ratio: a proposal for a new acoustic measure of dysarthric speech. *Journal of speech, language, and hearing research: JSLHR*, 53(1), 114–125.
- [13] Grabe, E., Kochanski, G., Coleman, J. (2007). Connecting intonation labels to mathematical descriptions of fundamental frequency. *Lang Speech*. 2007;50(Pt 3):281-310.
- [14] Linville SE, Fisher HB. Acoustic Characteristics of Womens Voices with Advancing Age. *J Gerontol*. 1985;40(3):324-30.
- [15] Liss JM, Weismer G, Rosenbek JC. Selected Acoustic Characteristics of Speech Production in Very Old Males. *J Gerontol*. 1990;45(2):P35-P45.
- [16] Rastatter MP, Jacques RD. Formant Frequency Structure of the Aging Male and Female Vocal-Tract. *Folia Phoniatr*. 1990;42(6):312-9.

- [17] Schötz S. Acoustic analysis of adult speaker age. In C. Müller (Ed.). Speaker classification I. Fundamentals, features, and methods (pp. 88-107). Berlin: Springer. 2007.
- [18] Jacob Benesty, M. Mohan Sondhi, Yiteng Huang. Handbook of Speech Processing. Springer, 2008. ISBN: 978-3-540-49125-5.
- [19] Ramig LA. Effects of physiological aging on speaking and reading rates. J Commun Disord. 1983;16(3):217-226.

Jméno a pracoviště vedoucí(ho) bakalářské práce:

Ing. Tomáš Bořil, Ph.D., Fonetický ústav Filozofické fakulty UK

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

prof. Ing. Pavel Sovka, CSc., katedra teorie obvodů FEL

Datum zadání bakalářské práce: **20.01.2020** Termín odevzdání bakalářské práce: _____

Platnost zadání bakalářské práce:
do konce letního semestru 2020/2021

Ing. Tomáš Bořil, Ph.D.
podpis vedoucí(ho) práce

prof. Ing. Michael Šebek, DrSc.
podpis vedoucí(ho) ústavu/katedry

prof. Mgr. Petr Páta, Ph.D.
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

Poděkování / Prohlášení

Chtěla bych především poděkovat svému vedoucímu bakalářské práce Ing. Tomáši Bořilovi Ph.D za vedení, rady, podporu a poskytnutí studijních materiálů při zpracování práce.

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 18.05.2020

.....

Abstrakt / Abstract

Tato bakalářská práce se zabývá analýzou akustických charakteristik hlasu v závislosti na věku u zdravých českých rodilých mluvčích. Vlivem stárnutí dochází k rozsáhlým fyziologickým změnám organismu a změnám v akustických charakteristikách hlasu. V práci jsou zkoumány závislosti jednotlivých parametrů na objektivním a subjektivním věku mluvčích. Součástí práce je zhodnocení jednotlivých analýz parametrů a porovnání, které charakteristiky jsou použitelné pro účely plně automatické analýzy nahrávek a které vyžadují kroky manuálního zpracování. Jednotlivé parametry vhodné pro zkoumání v závislosti na věku se liší u mužů a žen. V případě mužů se jedná např. o tempo řeči, CSI vypočítané z půltónů a spektrální sklon. U žen patří mezi významné parametry základní frekvence f_0 , tempo řeči a CSI vypočítané z intenzity.

Klíčová slova: hlas, akustické charakteristiky hlasu, základní frekvence, tempo řeči, formanty, CSI, subjektivní věk, objektivní věk

This bachelor thesis deals with the analysis of age-dependent acoustic features of the speech of Czech native speakers. Due to the process of ageing, there are extensive physiological changes in organisms as well as in acoustic features of speech. In this thesis, dependencies of individual parameters to both objective and perceived age were studied. Part of the thesis is the evaluation of individual analyses and comparison of suitable features for fully automatic analysis of recordings. Final suitable features are different for males and females. Some of the suitable features for analysing male recordings are for example speaking rate, CSI determined from semitones and spectral slope. Suitable features for analysing female recordings are for instance fundamental frequency f_0 , speaking rate, CSI determined from intensity.

Keywords: speech, acoustic features of speech, fundamental frequency, speaking rate, formants, CSI, perceived age, objective age

Title translation: Age-dependent acoustic features of speech

Obsah /

1 Úvod	1
1.1 Motivace.....	1
2 Teoretický úvod do problematiky ..	2
2.1 Tvorba řeči.....	2
2.1.1 Stavba fonačního sys- tému.....	2
2.2 Filtrová teorie produkce řeči	3
2.3 Akustické dimenze zvuku.....	3
2.4 Akustické charakteristiky hlasu.....	4
2.4.1 Fonační charakteristi- ky hlasu.....	4
2.4.2 Artikulační tempo řeči.....	5
2.4.3 Dynamika intonace a intenzity.....	5
2.4.4 Spektrální sklon.....	5
2.5 Hlasy v českém jazyce.....	5
3 Související práce a současná situace problematiky	7
3.1 Fyziologické projevy stárnutí.....	7
3.1.1 Rozdíly ve stárnutí mužů a žen.....	7
3.2 Pozorované jevy v závislosti na věku mluvčího.....	7
3.2.1 Změny f_0 v závislosti na věku.....	8
3.2.2 Tempo řeči.....	8
3.2.3 Počátek znělosti VOT.....	8
3.2.4 Formanty.....	8
3.2.5 Kumulativní index str- mosti CSI.....	9
3.2.6 Přehled výsledků vy- braných parametrů.....	9
4 Metodika	10
4.1 Data.....	10
4.2 Praat.....	11
4.3 Matlab a mPraat.....	11
4.4 Analýza jednotlivých para- metrů.....	11
4.4.1 Základní frekvence f_0	11
4.4.2 Tempo řeči.....	12
4.4.3 CSI z f_0 v Hz, v pultó- nech a z intenzity.....	12
4.4.4 Poměr trvání vokálů k trvání nahrávky.....	13
4.4.5 Průměrné trvání sa- mohlásky.....	13
4.4.6 Spektrální sklon.....	13
4.4.7 Dynamika intenzity a intonace.....	13
4.4.8 Formanty.....	14
4.5 Mnohonásobná lineární re- grese.....	14
5 Výsledky	15
5.1 Základní frekvence f_0	15
5.2 Tempo řeči.....	16
5.3 Kumulativní index strmosti ...	16
5.4 Poměr délky vokálů k trvání nahrávky.....	16
5.5 Průměrná doba trvání sa- mohlásky.....	18
5.6 Spektrální sklon.....	19
5.7 Dynamika intenzity a into- nace.....	19
5.8 Formanty.....	20
5.9 Mnohonásobná lineární re- grese.....	21
6 Diskuze	23
6.1 Zhodnocení akustických cha- rakteristik hlasu žen.....	23
6.2 Zhodnocení akustických cha- rakteristik hlasu mužů.....	24
6.3 Porovnání výsledků pro obě pohlaví.....	25
7 Závěr	28
Literatura	29
A Přílohy	31

Tabulky / Obrázky

3.2.1. Přehled výsledků změn vybraných parametrů vzhledem k objektivnímu a subjektivnímu věku9	2.1.1. Schéma jednotlivých systémů podílejících se na procesu tvorby řeči2
6.1.1. Přehled výsledných Pearsonových korelačních koeficientů vybraných parametrů pro ženy 24	2.1.2. Schéma činnosti hlasivek3
6.2.2. Přehled výsledných Pearsonových korelačních koeficientů vybraných parametrů pro muže 25	2.2.3. Schéma filtové teorie produkce řeči3
6.3.3. Přehled a porovnání výsledků modelování odhadovaného objektivního a subjektivního věku mužů 26	2.4.4. Vyznačené formanty ve spektrogramu4
6.3.4. Přehled a porovnání výsledků modelování odhadovaného objektivního a subjektivního věku žen 26	2.4.5. Měření spektrálního sklonu5
	3.2.1. Výsledky zkoumání CSI z intenzity9
	5.1.1. Výsledky analýzy závislosti průměrné základní frekvence f_0 na objektivním věku mluvčího 15
	5.1.2. Výsledky analýzy závislosti průměrné základní frekvence f_0 na subjektivním věku mluvčího 15
	5.2.3. Výsledky analýzy závislosti průměrného tempa řeči na objektivním věku mluvčího 16
	5.2.4. Výsledky analýzy závislosti průměrného tempa řeči na subjektivním věku mluvčího ... 16
	5.3.5. Výsledky analýzy závislosti CSI vypočítaného ze základní frekvence na objektivním věku mluvčího 17
	5.3.6. Výsledky analýzy závislosti CSI vypočítaného ze základní frekvence na subjektivním věku mluvčího 17
	5.3.7. Výsledky analýzy závislosti CSI vypočítaného ze základní frekvence převedené na půltóny na objektivním věku mluvčího 17
	5.3.8. Výsledky analýzy závislosti CSI vypočítaného ze základní frekvence převedené na půltóny na subjektivním věku mluvčího 17
	5.3.9. Výsledky analýzy závislosti CSI vypočítaného z intenzity na objektivním věku mluvčího . 18

5.3.10.	Výsledky analýzy závislosti CSI vypočítaného z intenzity na subjektivním věku mluvčího	18
5.4.11.	Výsledky analýzy závislosti poměru doby trvání samohlásek v segmentu C na objektivním věku mluvčího	18
5.4.12.	Výsledky analýzy závislosti poměru doby trvání samohlásek v segmentu C na subjektivním věku mluvčího ...	18
5.5.13.	Výsledky analýzy závislosti průměrné doby trvání samohlásky na objektivním věku mluvčího	19
5.5.14.	Výsledky analýzy závislosti průměrné doby trvání samohlásky na subjektivním věku mluvčího	19
5.6.15.	Výsledky analýzy závislosti průměrného spektrálního sklonu na objektivním věku mluvčího	20
5.6.16.	Výsledky analýzy závislosti průměrného spektrálního sklonu na subjektivním věku mluvčího	20
5.7.17.	Výsledky analýzy závislosti dynamiky intenzity na objektivním věku mluvčího	20
5.7.18.	Výsledky analýzy závislosti spektrálního sklonu na subjektivním věku mluvčího	20
5.7.19.	Výsledky analýzy závislosti dynamiky intonace v púltónech na objektivním věku mluvčího	21
5.7.20.	Výsledky analýzy závislosti dynamiky intonace v púltónech na subjektivním věku mluvčího	21
5.8.21.	Výsledky analýzy závislosti frekvence formantu F1 pro samohlásku [i] na objektivním věku mluvčího	22

- 5.8.22.** Výsledky analýzy závislosti frekvence formantu F1 pro samohlásku [i] na subjektivním věku mluvčího 22
- 5.8.23.** Výsledky analýzy závislosti frekvence formantu F2 pro samohlásku [o] na objektivním věku mluvčího 22
- 5.8.24.** Výsledky analýzy závislosti frekvence formantu F2 pro samohlásku [o] na subjektivním věku mluvčího 22

Kapitola 1

Úvod

Tato bakalářská práce se zabývá problematikou analýzy akustických charakteristik hlasu v závislosti na věku u zdravých mluvčích. První část práce se věnuje především teoretickému úvodu do problematiky (kapitola 2) a souhrnu poznatků ze souvisejících prací (kapitola 3). Kapitola 4 popisuje zvolenou metodiku, způsob zpracování dat a programy použité pro jejich analýzu. Následně jsou stručně popsány způsoby získávání a zkoumání jednotlivých parametrů a jejich implementace. Výsledky jednotlivých analýz jsou uvedeny v kapitole 5, kde je zkoumána především závislost vybraných parametrů na věku objektivním a subjektivním. Výsledky z kapitoly 5 jsou podrobněji rozebrány v kapitole 6.

1.1 Motivace

Řečová analýza v dnešní době slouží jako nástroj při diagnostice nemocí mezi které patří např. Alzheimer, Parkinson a další neurodegenerativní onemocnění. Diagnostika se nemusí zaměřovat pouze na neurodegenerativní onemocnění, ale i například na PTSD (*angl. Posttraumatic Stress Disorder*), či ischemickou chorobu srdeční. [17] [21] Pro různé diagnostiky je ovšem zapotřebí mít přístup i k referenčním hodnotám řečových analýz zdravých jedinců. Je důležité zjistit, co je již indikátorem nějaké anomálie, nemoci a co je pouhým projevem stáří. Vlivem stárnutí dochází k rozsáhlým fyziologickým změnám organismu a změnám v akustických charakteristikách hlasu. Těmito změnami se zabývalo již mnoho studií, které sloužily především jako inspirace ve výběru vhodných parametrů pro tuto práci.

Další motivací pro zkoumání vztahu akustických charakteristik hlasu vzhledem k věku mluvčího je i zlepšení automatizace klasifikace věku a tzv. speech emotion recognition, neboli rozpoznání emocí z hlasu. Hlavním cílem této práce je odhalit vhodné akustické charakteristiky pro zjednodušení klasifikace věku a zároveň pro určení referenčních hodnot zdravých Čechů.

Kapitola 2

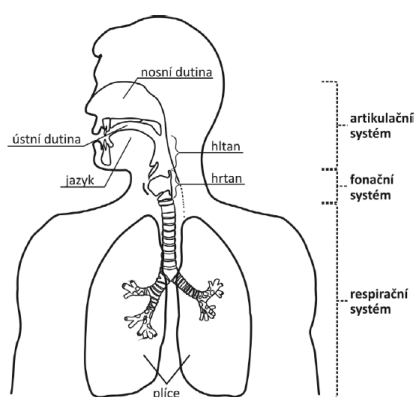
Teoretický úvod do problematiky

Pro lepší porozumění akustických charakteristik hlasu v závislosti na věku je nutné se seznámit s procesem tvorby řeči a jejími základními vlastnostmi. Kromě jednotlivých fyzikálních vlastností řečového zvuku je v této kapitole obsažen i krátký úvod do filtrové teorie produkce řeči a do systematiky hlásek v českém jazyce.

2.1 Tvorba řeči

Proces tvorby lidské řeči začíná v plicích (viz Obrázek 2.1.1), které jsou součástí respiračního systému. Respirační systém zajišťuje dýchání, přičemž vydechovaný vzduch z plic slouží jako energetický zdroj pro tvorbu zvuku v systému fonačním. Fonační systém, jehož součástí je hrtan s hlasivkami, zpracovává vydechovaný vzduch a tvoří zvuk rozkmitáváním hlasivek. Hrtanové svalstvo nastavuje tvar, délku, napětí, tuhost a polohu hlasivek, čímž jsme schopni produkovat různé zvuky.

Proces tvorby lidské řeči končí v artikulačním systému, kdy zvuk z fonačního systému prochází vokálním traktem. Vokální trakt je tvořen ze tří nadhrtanových dutin – hltanu, dutiny ústní a nosní. Artikulační orgány zajišťují polohu nastavení mluvidel, které takto mění tvar a objem rezonančních prostor. Různá nastavení vokálního traktu určují specifické akustické charakteristiky pro jednotlivé hlásky. Většina řečových zvuků, které je schopen člověk vytvořit, vzniká v dutině ústní. [25]

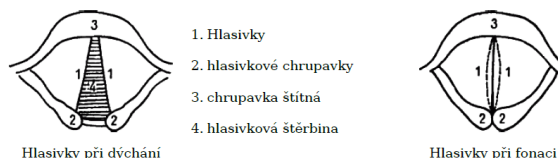


Obrázek 2.1.1. Schéma jednotlivých systému podílejících se na procesu tvorbě řeči. Převzato z [25].

2.1.1 Stavba fonačního systému

Nejdůležitější částí fonačního systému jsou hrtan a hlasivky. Hrtan (*larynx*) je součástí dolních cest dýchacích, tvoří jej krátká trubice složená z chrupavek, kterou proudí vzduch do průdušnice a následně do plic. Hlavními chrupavkami jsou chrupavka prstencová a štítná. Hrtan má v podstatě stejnou stavbu u žen i mužů, ovšem u mužů vystupuje štítná chrupavka tvořící podklad ohryzku, a proto je nápadnější.

Hlasivky (viz Obrázek 2.1.2), hlasové vazy, jsou uloženy symetricky uvnitř hrtanu. Vpředu jsou pevně uchycené k hrtanu, vzadu jsou pohyblivě upevněny na hlasivkové chrupavky hrtanu. Hlasivková štěrbina (*glottis*) se nachází v prostoru mezi hlasivkami. Při normálním dýchání má hlasivková štěrbina tvar protáhlého rovnoramenného trojúhelníku a nedochází k žádným slyšitelným efektům. Délka hlasivek se liší podle věku a pohlaví. [19]

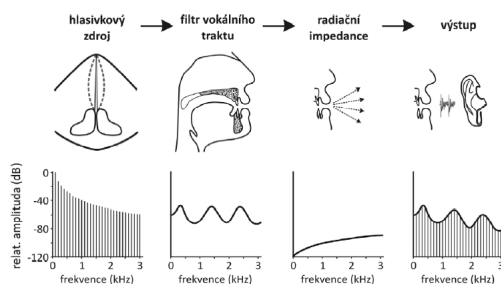


Obrázek 2.1.2. Schéma změny polohy hlasivek při dýchání a fonaci. Převzato z [12].

2.2 Filtrová teorie produkce řeči

Filtrová teorie produkce řeči (*angl. Source-filter Theory*, viz Obrázek 2.2.3) je teorie vynalezena švédským fonetikem Gunnarem Fantem v roce 1960. Jak již bylo řečeno v sekci 2.1, hlasivky přetváří vydechaný vzduch na kmitání – zvuk. Hlasivky jsou dostatečným zdrojem kmitání pro rozkmitávání vokálního traktu, jelikož vytváří složený zvuk s různorodým frekvenčním spektrem. V případě tvorby neznělých hlásek, kdy nedochází ke kmitání hlasivek, je zdrojovým signálem šum. [25]

Supraglotální vokální trakt v této teorii slouží jako filtr. Základní frekvence f_0 , také primární hlasivkový tón, je v nadhrtanových dutinách filtrována a rezonanční frekvence jednotlivých nastavení vokálního traktu, tzv. formanty, jsou chápány jako frekvenční pásma, která jsou nejméně utlumená. Formanty můžeme také popsat jako lokální maxima spektra složených tónů. Celý proces produkce řeči končí radiační impedancí, způsobem jak se šíří zvuková energie z vokálního traktu do okolí. [6] [25]



Obrázek 2.2.3. Schéma filtrové teorie produkce řeči s jednotlivými fázemi. Převzato z [25].

2.3 Akustické dimenze zvuku

Lidská řeč je jistou formou zvuku, jedná se tedy o mechanické vlnění, které se šíří vzduchem nebo jiným prostředím od svého zdroje. V následující sekci uvažujeme pouze tzv. čistý tón, který není složený z více frekvencí zároveň.

Mezi základní akustické dimenze zvuku patří frekvence f , která je definována jako počet kmitů za sekundu a udává se v hertzech (Hz). Další základní akustickou dimenzí zvuku je doba trvání zvuku, která se udává v sekundách (s). Jako poslední uvádíme amplitudu, jež je definována v akustice jako výchylka od atmosférického tlaku (rovnovážné polohy).

Jinými slovy, amplituda popisuje sílu zvuku, jeho hlasitost. Síla (intenzita) zvuku se vyjadřuje pomocí logaritmické stupnice v dB SPL (*Sound Pressure Level*). Popis v dB SPL vyjadřuje poměr měřené hodnoty k referenční hodnotě akustického tlaku, přičemž hodnota 0 dB SPL je obvykle považována za hranici sluchu. [25] [19]

2.4 Akustické charakteristiky hlasu

Nositelem akustického signálu je zvuková podélná vlna. Jednoduché tóny jsou pravidelné periodické zvukové vlny, nejběžnější řečové zvuky jsou ovšem zvuky složené, tzn. neobsahují pouze jediný sinusový tón. Složené zvuky mohou být buď šumového charakteru (neznělé frikativy), nebo kvaziperiodického charakteru (samohlásky), nebo kombinací obou (znělé frikativy). U kvaziperiodických průběhů dochází k opakování se základní periodou, která se v čase mění. Odpovídající frekvence k základní periodě se nazývá základní frekvence f_0 .

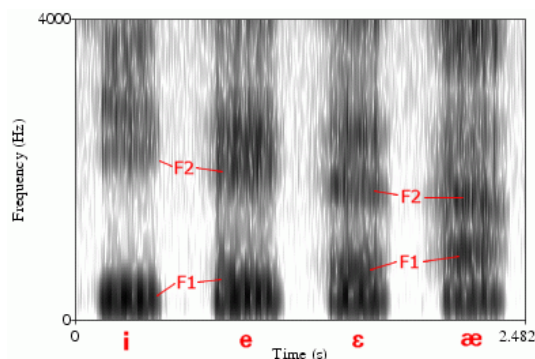
2.4.1 Fonační charakteristiky hlasu

Fonační charakteristiky hlasu silně souvisí s fyziologií hlasového ústrojí, kde má především význam délka hlasivek určující výšku hlasu – čím kratší, tím hlasivky rychleji kmitají a hlas je vyšší.

Základní frekvence f_0 patří mezi jednu z nejčastěji analyzovaných akustických charakteristik hlasu v souvislosti s věkem. Zobrazením zvukového spektra vykreslujícího intenzitu zvuku v závislosti na frekvenci, můžeme samohlásky, neboli tóny ze specifických konfigurací vokálního traktu, považovat za složené tóny charakterizované formanty, což jsou jejich rezonanční frekvence.

Formanty značíme F_i , kde $i \in \{1, 2, 3, 4\}$. Vyšší formanty jsou v podstatě stejné u všech vokálů, a proto je ani neuvádíme. Tyto rezonance dále tvarují složený tón a jeho spektrum. Opakem formantů jsou tzv. antiformanty, neboli lokální minima, propady signálového spektra. [12] [25] [30]

Za nejvýznamnější formanty se považují F_1 a F_2 (viz Obrázek 2.4.4), ve kterých dochází k největším změnám. Pomocí těchto formantů v akustických analýzách vyhodnocujeme vokální prostor.



Obrázek 2.4.4. Vyznačené jednotlivé formanty ve spektrogramu. Převzato¹.

¹ <https://home.cc.umanitoba.ca/~kruss11/phonetics/acoustic/formants.html>

2.4.2 Artikulační tempo řeči

Artikulační tempo řeči souvisí s tvorbou hlásek a funkcí artikulačních orgánů. Je dáno rychlostí artikulace a v případě v češtině jej udáváme počtem slabik za sekundu. Tempo řeči je ovlivněno několika faktory. Záleží, jestli se jedná o projev čtený, či spontánní, na délce pauz v promluvě apod.

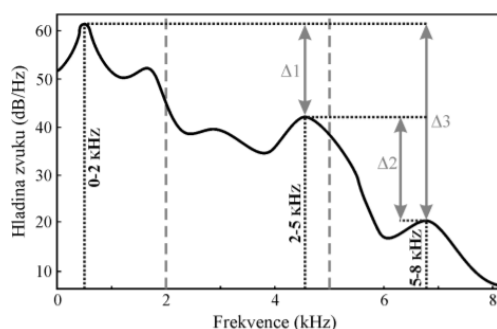
2.4.3 Dynamika intonace a intenzity

Dynamika intonace (*angl. pitch dynamics*) vyjadřuje průměrný absolutní intonační sklon v půltónech bez oktákových skoků. Každý lokální skok větší než 6 půltónů je snížen o jednu, či více oktáv. Oktáva má 12 půltónů, takže skok například o 10 půltónů nahoru se vyhodnocuje jako skok o 2 půltóny dolů.¹

Dynamika intenzity, neboli dynamika síly zvuku (*angl. intensity dynamics*) je definována jakožto průměrný absolutní sklon intenzity v dB/s. Měří se pouze přes znělé části, jelikož v neznělých může mluvíci mlčet. V těchto případech by se intenzita řeči rovnala intenzitě zvuků v pozadí, které jsou zároveň silně závislé na nahrávacím prostředí.

2.4.4 Spektrální sklon

Spektrální sklon (*angl. spectral slope*, viz Obrázek 2.4.5) je termín, kterým se označuje spád křivky interpolující závislost energie řeči [dB] na frekvenci [Hz] v řečovém spektru a je ovlivněn několika faktory, kde nejvýznamnějšími pro forenzní rozpoznávání jsou kvalita a barva hlasu. Spektrální sklon se ovšem obtížně měří, mezi nejčastější metodologické problémy patří kvantizace digitálních signálů při záznamu řeči, správná vzdálenost mluvíčího od mikrofonu, jenž musí mít dostatečný frekvenční a dynamický rozsah. V literatuře se setkáváme s hodnotou -6 dB/oktávu při normální fonaci. [30] Normální fonace (*angl. modal phonation*) je tzv. normální kmitání hlasivek a odpovídá nejefektivnějšímu způsobu rozkmitávání hlasivek. Když jsou uzavřené hlasivky a zároveň člověk vydechuje, tak se pod hlasivkami tvoří subglotální tlak, který po dosažení určité meze hlasivky rozrazí a skrze vzniklou hlasivkovou šterbinu začne proudit rychle vzduch. Hlasivky se následně přitáhnou zpět k sobě vlivem vzniklého podtlaku v hlasivkové šterbině. [8] [25]



Obrázek 2.4.5. Ukázka měření spektrálního sklonu jako rozdílu maximálních energií LTAS ve vyznačených frekvenčních pásmech. Převzato z [30].

2.5 Hlásky v českém jazyce

Hlásky obecně se v každém jazyce dělí na vokály (samohlásky) a konsonanty (souhlásky). Tyto dvě skupiny můžeme porovnávat různými fonetickými perspektivami.

¹ <https://developers.vokaturi.com/algorithms/acoustic-features>

Kapitola 3

Související práce a současná situace problematiky

Proces stárnutí je velice komplexní a ovlivňuje mimo jiné i způsob řečové komunikace. Způsoby lidské mluvy se mění po celou dobu našeho života – od brzkého dětství až po pokročilý věk. I když k většině změn dochází následkem významných hormonálních a fyziologických změn především v dětství a pubertě, pozorujeme i další věkově závislé proměny hlasových charakteristik v průběhu celého života.

Celé problematice se věnuje nespočet studií, některé se věnují spíše fyziologickým projevům stárnutí a jejich následkům, jiné se zabývají objevováním parametrů pro zjednodušení automatizace práce s hlasem, nebo zkoumají různé parametry hlasu, které fungují jako indikátory budoucích projevů různých onemocnění.

3.1 Fyziologické projevy stárnutí

Určitý vliv na akustické charakteristiky hlasu v závislosti na věku má i fyzický proces stárnutí. Bohužel je často neskutčné vymezit hranice mezi změnami v důsledku věku, různých nemocí, či změn v okolním prostředí.

Hlas nejvíce ovlivňují změny týkající se dýchací soustavy člověka. Respirační systém se dovyvine po pubertě, ale ovšem v průběhu let dochází k poklesu plicní kapacity nejčastěji ze ztráty elasticity plicní tkáně. Mezi další projevy stárnutí patří oslabení dýchacích svalů. Fyziologické změny hrtanu ovlivňují především základní frekvenci f_0 a kvalitu hlasu. [22]

Jevy degenerace a atrofie svalů jsou průvodnými jevy stárnutí, dochází i k jisté změně v motorické činnosti jak u periferní, tak centrální nervové soustavy. Tyto změny mohou ovlivnit tempo řeči a koordinaci artikulačních orgánů. [13]

3.1.1 Rozdíly ve stárnutí mužů a žen

Přestože muži a ženy jsou z biologického hlediska jedním druhem, rozdíly v procesu stárnutí u obou pohlaví jsou markantní. U mužů je nejvýraznější hormonální změnou puberta, ženy kromě puberty ještě absolvují menopauzu. Osifikace hrtanu se vyskytuje později u žen než u mužů. Nicméně věkově závislé změny u dospělých jsou většinou výraznější u mužů než u žen, např. v případě rozsahu změn v hrtanové struktuře, pohybu jazyka a tempa řeči. [22]

3.2 Pozorované jevy v závislosti na věku mluvčího

Pro zkoumání hlasových charakteristik jsou dva základní typy studií. Prvním typem jsou studie podélné (dlouhodobé, longitudiální), kde jsou účastníci studie pozorováni po dobu několika let. Druhým a častějším typem jsou studie příčné, ve kterých je vybraná určitá skupina lidí splňující požadovaná kritéria. Většina studií zkoumá různé

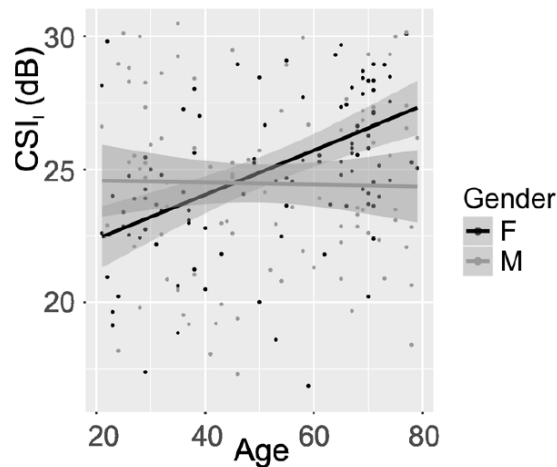
3.2.5 Kumulativní index strmosti CSI

Kumulativní index strmosti (*angl. Cumulative Slope Index (CSI)*) se používá pro zachycení variací kontur základní frekvence f_0 a intezity. CSI je definováno následujícím vzorcem:

$$CSI = \frac{1}{N_{syll}} \sum_{n=2}^N |x(n) - x(n-1)| \quad (3.1)$$

kde N_{syll} je počet slabik v daném segmentu (může reprezentovat i počet hlásek), N je počet diskretních bodů v analyzovaném segmentu, $x(n)$ je hodnota n -tého bodu (pro intenzitu dB, pro výpočet ze základní frekvence obvykle v pŕiltónech).[28]

Kumulativní index strmosti je poměrně nový a zatím ne příliš zkoumaný parametr. CSI počítané z intezity zvuku ukazuje nárůst v závislosti na věku pro ženy (viz obrázek 3.2.1), u mužů ovšem zůstává konstantní. U CSI vypočítaného z manuálně opravených kontur f_0 naopak byl nárůst u mužů větší než u žen.[28]



Obrázek 3.2.1. Výsledky zkoumání CSI vypočítaného z intezity. Převzato z [28].

3.2.6 Přehled výsledků vybraných parametrů

Změny ve vybraných akustických parametrech jak pro objektivní, tak subjektivní věk jsou přehledně uvedeny v Tabulce 3.2.1 převzaté z [22].

parametr	OV ženy	SV ženy	OV muži	SV muži
tempo řeči [syll/s]	↘ →	↘ →	↘	↘
průměrná f_0 [Hz]	(↘ →) poté ↘	↘	↘ poté ↗	↘ poté ↗
rozsah f_0 [Hz]	↗ poté ↘	↗ →	↗ poté → ↘	↗ poté (→ ↘)
frekvence F1 [Hz]	↘ →	↘	↘ →	↘
frekvence F2 [Hz]	↗ poté (→ ↘)	↘	↗ poté (→ ↘)	↘

Tabulka 3.2.1. Přehled výsledků zkoumání změn vybraných parametrů vzhledem k vzrůstajícímu objektivnímu (OV) a subjektivnímu věku (SV). → = konstantní průběh, ↘ = pokles, ↗ = nárůst, | = nebo. Slovo poté označuje změnu ve vyšším věku (u žen menopauzu). Přeloženo a upraveno z [22].

Kapitola 4

Metodika

V následujících podkapitolách je uveden způsob zpracování dat a následně je popsána analýza vybraných akustických parametrů.

4.1 Data

Jako dataset pro odhalení akustických charakteristik hlasu v závislosti na věku byly použity nahrávky pořízené v rámci grantu *Věkově závislé změny akustických charakteristik řeči dospělých mluvčích*, GAČR 16-19975S, hlavní řešitelé Roman Čmejla (ČVUT), za FF UK Jan Volín, 2016 – 2018.

Tyto nahrávky obsahují 240 mluvčích ve věku od 18 do 90 let, kteří namluvili úryvek z vybraného díla *Měl jsem psa a kočku* od Karla Čapka. Tento úryvek se skládá ze 137 slov a při jeho výběru byl kladen důraz na vynechání nízkofrekvenčních lexikálních jednotek a neobvyklých syntaktických skladeb. Úryvek byl následně rozdělen do 12 úseků označených A-L.

A: „*I na tom, že člověk si opatří psa, aby nebyl sám, je mnoho pravdy.*“

B: „*Pes opravdu nechce být sám.*“

C: „*Jen jednou jsem nechal Mindu o samotě v předsíni.*“

D: „*na znamení protestu sežrala všechno, co našla, a bylo jí pak poněkud nedobře.*“

E: „*Podruhé jsem ji zavřel do sklepa s tím výsledkem, že rozkousala dveře.*“

F: „*Od té doby nezůstala sama ani po jedinou minutu.*“

G: „*Když píše, chce, abych si s ní hrál.*“

H: „*Když si lehnu, považuje to za znamení, že si mně smí lehnout na prsa a kousat mě do nosu.*“

I: „*Přesně o půlnoci s ní musím provádět Velikou Hru, při níž se s velikým hlukem honíme, koušeme a kutálíme po zemi.*“

J: „*Když se uřítí, jde si lehnout.*“

K: „*pak si smím lehnout i já, ovšem s tou podmínkou,*“

L: „*že nechám dveře do ložnice otevřené, aby se Mindě nestýskalo.*“ [1]

Zastoupení jednotlivých pohlaví mluvčích bylo skoro vyvážené – 105 mužů a 135 žen. Všichni mluvčí byli zdraví a jejich mateřským jazykem byla čeština.

Před nahráváním měli mluvčí dostatečný čas se s textem seznámit. Nahrávání probíhalo v zařízené a tiché místnosti, kde mluvčí četli plynule celý úryvek. Při nahrávání byl použit kondenzátorový hlavový mikrofon Bayerdynamic Opus 55, zapojený přímo do kapesního záznamníku nastaveného do 16bitového režimu nekomprimovaných 48 kHz. Jednotlivé nahrávky byly posléze manuálně zpracovány, nejprve byly rozčleněny na jednotlivé segmenty A-L, následně k nim byly vytvořeny pomocí Praatu [2] jednotlivé Textgridy s několika ručně segmentovanými úrovněmi – hlásky, slova a celá věta, dle fonetických pravidel v souladu s [16].

Dále byly extrahovány průběhy základní frekvence f_0 a následně manuálně opraveny

s ohledem na chyby oktávových skoků a interpolovány v oblastech třepené fonace. Takto manuálně opravené kontury f_0 byly exportovány do souborů typu .PitchTier. Mezi vyexportované soubory patří i intenzita jednotlivých segmentů v dB, která byla měřená každých 10 ms pomocí kubické interpolace.

Pro zjištění subjektivního věku mluvčích bylo vybráno 37 respondentů, jejichž úkolem bylo určit věk mluvčích pouze na základě poslechu. Výsledný subjektivní věk je aritmetickým průměrem výsledných odhadů respondentů.

Pro práci s těmito daty v této bakalářské práci proběhla i poslechová kontrola všech nahrávek a z výběru pro další zkoumání byly vyřazeny nahrávky obsahující libovolné přerušky. Z důvodu časté chybovosti a tedy nedostatečného množství validních úseků byly nakonec vyřazeny všechny nahrávky segmentu I.

4.2 Praat

Praat [2] je volně dostupný program používán pro analýzu a rekonstrukci řeči ve fonetice. Autoři jsou prof. Paul Boersma a dr. David Weenink z University v Amsterdamu. Tento program umožňuje základní i pokročilé analýzy zvuku. Pracuje s formáty .wav, mezi nejvhodnější funkce pro analýzu řeči patří spektrální analýza, analýza průběhu f_0 , analýza formantů a další. Dále je vhodný pro tvorbu TextGridů, obsahuje i dopředné neuronové sítě, různé filtry, či multidimenzionální škálování zvuku.

4.3 Matlab a mPraat

Většina implementační části práce byla prováděna v Matlabu [18] za pomoci toolboxu mPraat [3]. Matlab je program vyvíjen společností MathWorks, jedná se o interaktivní programovací prostředí a skriptovací programovací jazyk. Díky rozsáhlé databázi toolboxů je Matlab univerzálním nástrojem v mnoha oborech – od strojového učení, přes optimalizaci po neuronové sítě a deep learning.

Toolbox mPraat umožňuje upravovat a zpracovávat soubory typu .TextGrid, .PitchTier a .IntensityTier v Matlabu, přičemž manipulace s nimi je snazší než v Praat skriptech. Více technicky zaměření odborníci používají spíše nástroje jako je Matlab a R, zatímco lingvisticky orientovaní odborníci používají Praat a tento toolbox slouží k jejich vzájemnému propojení. [3]

4.4 Analýza jednotlivých parametrů

Analýza jednotlivých parametrů probíhala především v Matlabu, až na výjimky, kdy bylo snazší použít již implementované metody v Praatu, které toolbox mPraat neobsahuje. Grafy byly generovány pomocí R a knihovny ggplot2 [31]. Jak již bylo zmíněno v 4.1, pro validaci dat byla vyhotovena validační matice $A_{240 \times 13}$, kde prvním sloupci odpovídají jednotlivá ID mluvčích a zbývající sloupce odpovídají jednotlivým segmentům A-L. Pro validitu dat byla zvolena binární stupnice – 1 pro validní nahrávky, 0 pro vyřazené.

4.4.1 Základní frekvence f_0

Výpočet průměrné základní frekvence byl implementován v Matlabu. Procházením validační maticí byl vždy určitý segment vyhodnocen jako validní, či chybný. V případě validního segmentu byl načten odpovídající soubor .PitchTier pomocí toolboxu mPraat

a byl vyhotoven průměr kontur základní frekvence za celý segment. Pokud došlo ke čtení chybné nahrávky, výsledný průměr základní frekvence daného segmentu byl nastaven na hodnotu *NaN*. Z dílčích validních výsledků jednotlivých mluvčích se na závěr vyhotovila i průměrná celková základní frekvence f_0 .

4.4.2 Tempo řeči

Pro výpočet průměrného tempa řeči ve slabikách za sekundu bylo nutné přidat úroveň slabik do již vyhotovených souborů typu `.TextGrid`. Nová úroveň s názvem *syll* byla vytvořena pouze u validních vzorků pomocí toolboxu `mPraat`. Po načtení validní nahrávky s odpovídajícím souborem `.TextGrid` bylo nutné spojit celou úroveň hlásek do jednoho dlouhého textového řetězce, který se následně musel ručně oslabikovat.

Jelikož zde nebyla jedna správná verze a sloučené řetězce hlásek měly více verzí, veškeré nahrávky, jejichž verze řetězce se vyskytovaly méně než třikrát, byly označeny za chybné ve validační matici.

Po manuálním oslabikování bylo nutné přiřadit správnou verzi nové úrovni slabik k původnímu souboru `.TextGrid`.

Implementace výpočtu průměrného tempa řeči byla implementována v `Matlabu`. Po načtení validní nahrávky byl načten odpovídající soubor `.TextGrid` s úrovní *syll*. Soubory `.TextGrid` obsahují prázdné intervaly odpovídajícím pauzám na začátku a na konci každého segmentu. Celková doba trvání segmentu použitá pro výpočet průměrného tempa řeči byla zkrácena o dobu trvání těchto pauz. Průměrné tempo řeči (*syll_s*) pro segment *X* dostaneme pomocí vzorce:

$$syll_s_X = \frac{(numInt - numEmpty)}{t} \quad (4.1)$$

kde *numInt* je celkový počet intervalů úrovně *syll*, *numEmpty* je celkový počet prázdných intervalů v této úrovni (pauz) a *t* je celková doba trvání po oříznutí o trvání první a poslední pauzy. Z jednotlivých výsledků pro validní segmenty bylo posléze vypočítáno průměrné tempo řeči (*syll_s*).

4.4.3 CSI z f_0 v Hz, v půltónech a z intenzity

Implementace výpočtu kumulativního indexu strmosti byla realizována v `Matlabu` dle rovnice (3.1). Celkem byl CSI vypočítáván z validních segmentů pro 3 veličiny, ze základní frekvence f_0 v Hz, v půltónech a z intenzity zvuku.

CSI ze základní frekvence bylo počítáno ze souborů typu `.PitchTier` obdobně jako v 4.4.1. Po načtení validního segmentu s příslušným souborem `.PitchTier` pro jednotlivé mluvčí byla aplikována již zmíněná definice kumulativního indexu strmosti (3.1). V případě nevalidních nahrávek byl výstup nastaven na hodnotu *NaN*. Z validních dílčích výsledků pro každého mluvčího byl posléze vyhotoven aritmetický průměr pro stanovení průměrného CSI ze základní frekvence f_0 .

Určení CSI z půltónů probíhal identicky jako pro CSI ze základní frekvence f_0 v Hz s tím rozdílem, že po načtení příslušného validního souboru typu `.PitchTier` byly frekvence převedeny na půltóny pomocí funkce z toolboxu `mPraat`.

Vstupními daty pro CSI z intenzity nebyly soubory typu `.PitchTier`, ale vyexportované textové soubory s intenzitou v dB uváděné po 10 ms z `Praatu`. Samotný výpočet pro validní segmenty probíhal opět dle (3.1), kde bylo jen nutné vyloučit časové úseky, které byly alespoň z jedné strany obklopeny hodnotou *NaN*, které se vyskytují před zahájením čtení mluvčího a po jeho skončení.

■ 4.4.4 Poměr trvání vokálů k trvání nahrávky

Určení poměru trvání vokálů k celkovému trvání čtení daného úseku úryvku bylo implementováno v Matlabu za pomoci toolboxu mPraat. Pro každou validní nahrávku dle validační matice byl načten příslušný soubor typu .TextGrid s úrovní phone obsahující jednotlivé hlásky. Stejně jako u výpočtu tempa řeči v 4.4.2 se délka trvání jednotlivého segmentu musela oříznout o trvání počáteční a koncové pauzy na délku t . Následně byly v jednotlivých intervalech hlásek vyhledány samohlásky a dvojhásky a určila se jejich doba trvání t_{vow} . Z celkové doby trvání t byly ještě odečteny veškeré délky intervalů t_{emp} odpovídající pauze. Výsledný poměr trvání vokálů k trvání nahrávky (vow2con) pro segment X dostáváme pomocí vzorce:

$$vow2con_X = \frac{t_{vow}}{t - t_{emp}} \quad (4.2)$$

Oproti ostatním zkoumaným parametrům zde nelze udělat celkovou průměrnou hodnotu, jelikož je každý dílčí výsledek pevně vázán na konkrétní formulaci věty X.

■ 4.4.5 Průměrné trvání samohlásky

Zkoumání průměrného trvání samohlásek ve vztahu k věku bylo přejato ze studie [7]. Řešení bylo implementováno v Matlabu. Pro validní segmenty dle validační matice byl načten příslušný soubor typu .TextGrid s úrovní phone. Následně byly vyhledány všechny intervaly daného segmentu I_{vow} obsahující samohlásky. Výsledné průměrné trvání samohlásky (vow_on) pro segment X se spočítalo dle vzorce:

$$vow_on_X = \frac{t_{vow}}{I_{vow}} \quad (4.3)$$

kde t_{vow} je doba trvání všech samohlásek za daný segment. Z dílčích validních segmentů pro každého mluvčího byla aritmetickým průměrem vypočítána průměrná délka samohlásky [ms] pro konkrétního mluvčího.

■ 4.4.6 Spektrální sklon

Výpočet spektrálního sklonu byl implementován v Praatu a výsledky dále zpracovány v Matlabu.

Spektrální sklon byl určen pomocí tzv. pitch-corrected LTAS (*Long-Term Average Spectrum*) již implementované v Praatu. Tato metoda bere v potaz pouze znělé části nahrávky a automaticky opravuje nehladkou distribuci intonace. Spektrální sklon byl vypočítáván jako rozdíl v dB mezi frekvenčním pásmem od 50 do 1000 Hz a frekvenčním pásmem od 1 do 4 kHz.

V Matlabu bylo nutné tabulku výsledků přeformátovat tak, aby se daly údaje přidat do hromadné tabulky výsledků všech ostatních parametrů. Hodnoty spektrálního sklonu pro nevalidní hodnoty byly nastaveny dle validační matice na hodnotu NaN a následně byl z validních segmentů vypočítán průměrný spektrální sklon jednotlivých mluvčích.

■ 4.4.7 Dynamika intenzity a intonace

Dynamika intenzity a intonace byla analyzována v Praatu a výsledky dále zpracovány v Matlabu. Výpočet dynamiky intenzity i intonace je již implementován v Praatu.

Zpracování výsledné tabulky v Praatu probíhalo stejně jako v 4.4.6. Byly vyřazeny výsledky chybných nahrávek a vypočítána průměrná hodnota dynamiky intenzity a intonace pro každého mluvčího.

4.4.8 Formanty

Základní kostra skriptu pro určení formantů byla převzata od Daniela Riggse¹. Po úpravě, která především spočívala ve změně formátu výsledků a přidání alternativního způsobu výpočtu formantů pomocí třetin intervalu dané hlásky, byly výsledky obou způsobů výpočtu pro všechny nahrávky vyexportovány do tabulky pro následné zpracování v Matlabu.

Nevýhodou převzatého skriptu bylo, že formanty určil u každé nahrávky pro každou hlásku, tudíž vygenerovaná tabulka výsledků měla přes 150 000 řádků. V Matlabu byl implementován skript, který vyřadil nejprve všechny řádky týkající se segmentu I a sloučil je vždy do jednoho s hodnotou NaN pro každého mluvčího. Následně byly z tabulky vyřazeny řádky obsahující libovolné souhlásky a hlásky v okolí nazálních hlásek viz sekce 3.2.4.

Dále bylo nutné vyřadit ostatní chybné nahrávky pomocí validační matice a zároveň u validních nahrávek pro každého mluvčího vyhotovit průměr frekvencí jednotlivých formantů F1 a F2 za každý segment pro každou samohlásku. Ze všech validních dílčích výsledků byl poté zhotoven průměr frekvencí formantů pro danou samohlásku x , kde $x \in \{ [a], [a:], [e], [e:], [i], [i:], [o], [o:], [u], [u:] \}$ pro každého mluvčího.

4.5 Mnohonásobná lineární regrese

Pro modelování objektivního a subjektivního věku byl použit lineární model s mnohonásobnou lineární regresí bez interakcí. Samotné modelování bylo provedeno pomocí skriptu v R naimplementovaného panem Ing. Tomášem Bořilem PhD. Pro validaci modelu a vyhodnocení odchylek byla použita křížová validace opakovaná celkově patnáctkrát v konfiguraci 10-*fold*. Data byla tedy během každého *fold* rozdělena náhodně v poměru 9:1 na trénovací a testovací skupinu a v rámci všech 10 *folds* jednoho opakování křížové validace bylo zajištěno, že pro testování byla každá položka použita právě jednou. Patnáct opakování celého procesu zajistí větší statistickou stabilitu, a tím realnost změřených odchylek. Výsledné intervaly odchylek znázorňují mezikvartilový rozsah (IQR), což odpovídá rozdílu mezi 75. a 25. percentilem, tedy 50 % nejčastějších hodnot. Modelování slouží pro zhodnocení a porovnání výsledků [28] v kapitole 6 s výsledky analýz popsaných v sekci 4.4.

¹ <http://praascriptingtutorial.com/filesExtendedExample>

Kapitola 5

Výsledky

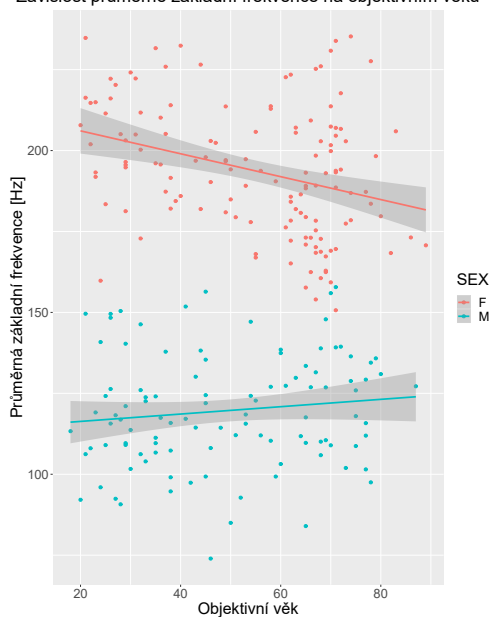
Tato kapitola se věnuje výsledkům jednotlivých analýz vybraných akustických parametrů dle uvedené metodiky v kapitole 4. Výsledky analýz jsou nejpřehlednější ve formě statistického grafu pomocí lineární regrese s vyznačeným intervalem spolehlivosti pro konfidenční hladinu 95 %, ovšem v této kapitole bude uvedena jen jejich část, zbývající grafy budou v kapitole Příloha A. Dále bude uveden u každého parametru Pearsonův korelační koeficient popisující lineární vztah mezi vybraným parametrem a věkem objektivním (příp. subjektivním) pro ženy a muže.

Pro dosažení co nejobecnějších výsledků bez ohledu na konkrétní věty jsou zde uvedeny grafy průměrovaných výsledků (pokud šly průměrovat) analýz veličin pro jednotlivé mluvčí, tzn. jeden bod v grafu odpovídá jednomu mluvčímu.

5.1 Základní frekvence f_0

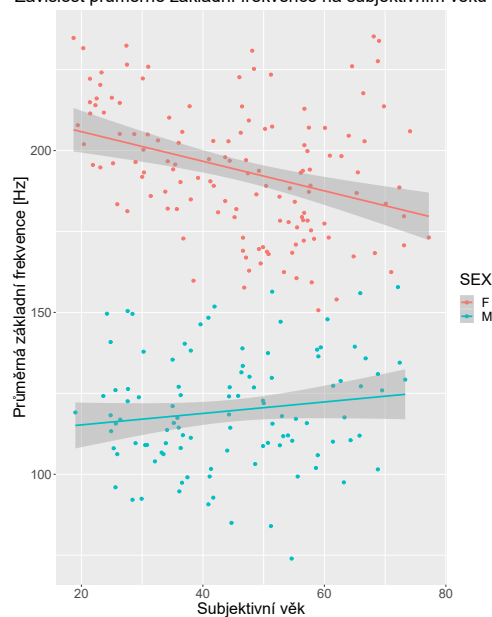
Na grafech 5.1.1, 5.1.2 vidíme průběhy průměrné základní frekvence f_0 v závislosti na objektivním a subjektivním věku mluvčích. Pearsonův korelační koeficient pro muže $r_M = 0.12$ pro objektivní věk a pro subjektivní věk $r_M = 0.14$. U žen vychází korelační koeficient $r_F = -0.32$ pro objektivní věk a pro subjektivní věk $r_F = -0.34$.

Závislost průměrné základní frekvence na objektivním věku



Obrázek 5.1.1. Závislost mezi průměrnou základní frekvencí f_0 [Hz] a objektivním věkem.

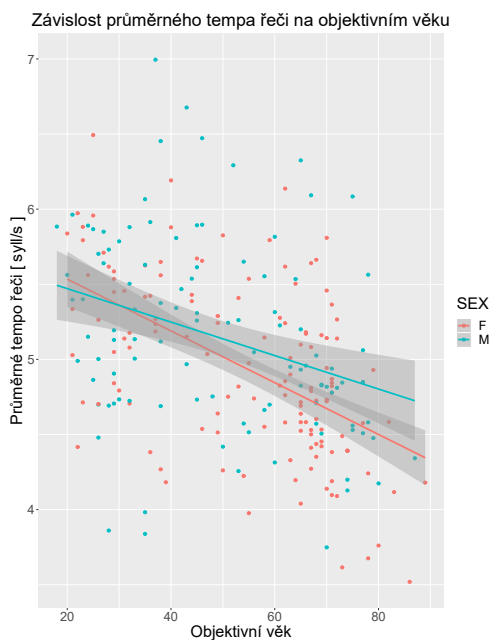
Závislost průměrné základní frekvence na subjektivním věku



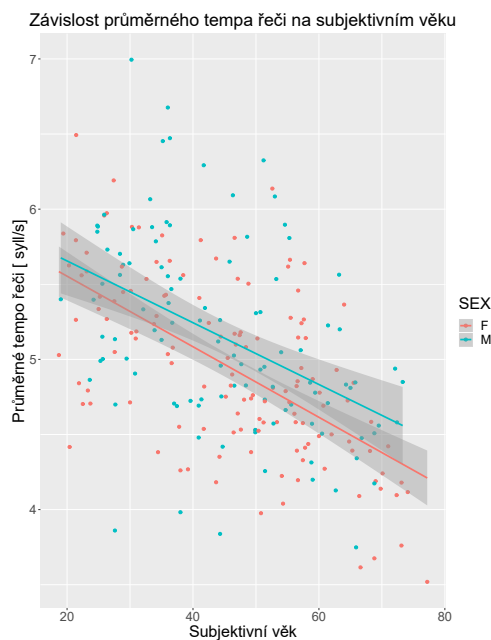
Obrázek 5.1.2. Závislost mezi průměrnou základní frekvencí f_0 [Hz] a subjektivním věkem.

5.2 Tempo řeči

Grafy 5.2.3, 5.2.4 zobrazují průběhy průměrného tempa řeči v závislosti na objektivním a subjektivním věku mluvčích. Pearsonův korelační koeficient pro muže $r_M = -0.33$ pro objektivní věk a pro subjektivní věk $r_M = -0.44$. U žen vychází korelační koeficient $r_F = -0.53$ pro objektivní věk a pro subjektivní věk $r_F = -0.6$.



Obrázek 5.2.3. Závislost průměrného tempa řeči [syll/s] na objektivním věku.



Obrázek 5.2.4. Závislost průměrného tempa řeči [syll/s] na subjektivním věku.

5.3 Kumulativní index strmosti

Kumulativní index strmosti CSI byl spočítán ze tří veličin – ze základní frekvence (grafy 5.3.5, 5.3.6), ze základní frekvence převedené na půltóny (grafy 5.3.7, 5.3.8) a z intenzity (grafy 5.3.9, 5.3.10).

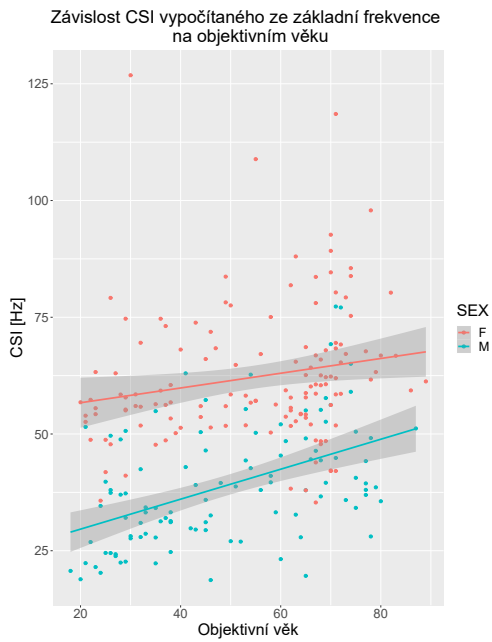
Korelační koeficient pro CSI ze základní frekvence f_0 vychází pro muže $r_M = 0.47$ pro objektivní věk a pro subjektivní věk $r_M = 0.52$. U žen vychází korelační koeficient $r_F = 0.2$ pro objektivní věk a pro subjektivní věk $r_F = 0.24$.

Korelační koeficient CSI z frekvence převedené na půltóny vychází pro muže $r_M = 0.6$ pro objektivní věk a pro subjektivní věk $r_M = 0.65$. U žen vychází korelační koeficient $r_F = 0.36$ pro objektivní věk a pro subjektivní věk $r_F = 0.41$.

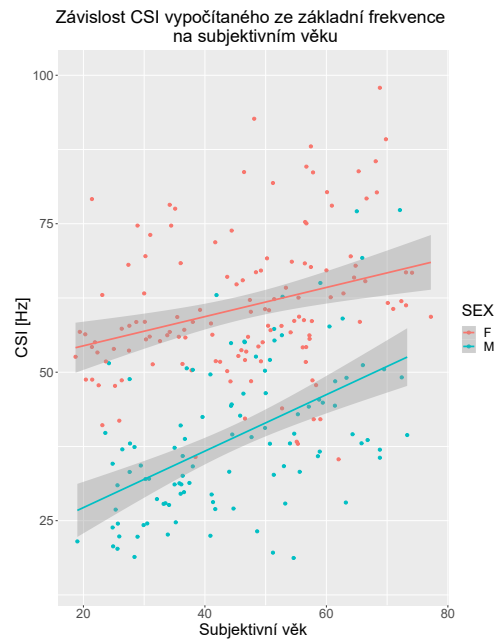
Korelační koeficient pro CSI z intenzity vychází pro muže $r_M = 0.09$ pro objektivní věk a pro subjektivní věk $r_M = 0.18$. U žen vychází korelační koeficient $r_F = 0.46$ pro objektivní věk a pro subjektivní věk $r_F = 0.52$.

5.4 Poměr délky vokálů k trvání nahrávky

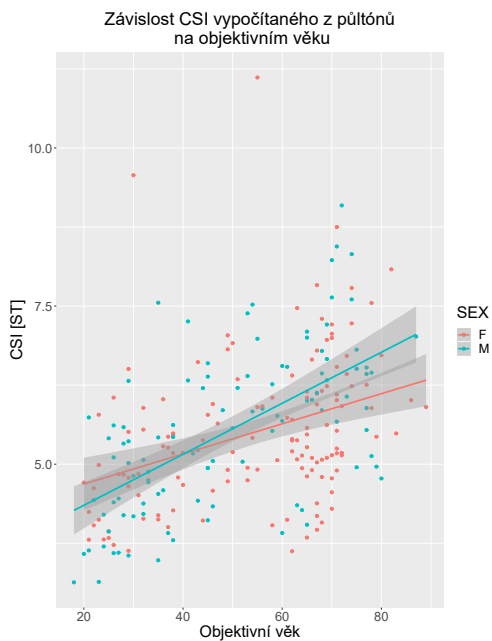
Tento parametr se nedá zobecnit průměrováním a je tedy závislý na konkrétní větě a její formulaci. Grafy 5.4.11, 5.4.12 zobrazují poměr doby trvání vokálů pro segment C, zbytek grafů bude přiložen v kapitole A. Korelační koeficient pro muže $r_M = 0.09$



Obrázek 5.3.5. Závislost průměrného kumulativního indexu strmosti CSI [Hz] vypočítaného ze základní frekvence f_0 na objektivním věku.



Obrázek 5.3.6. Závislost průměrného kumulativního indexu strmosti CSI [Hz] vypočítaného ze základní frekvence f_0 na subjektivním věku.

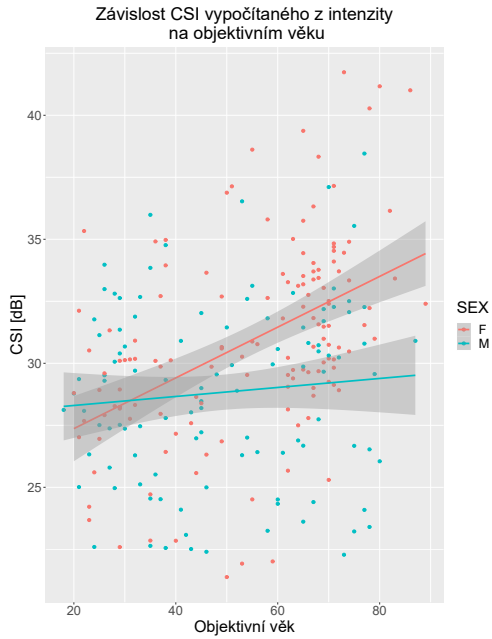


Obrázek 5.3.7. Závislost průměrného kumulativního indexu strmosti CSI [ST] vypočítaného z púltónů na objektivním věku.

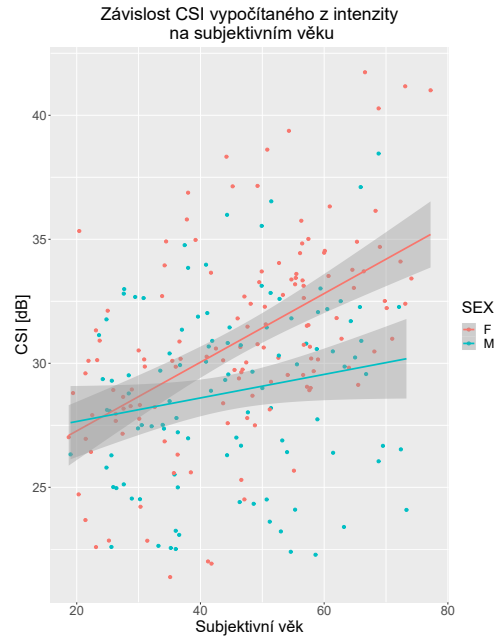


Obrázek 5.3.8. Závislost průměrného kumulativního indexu strmosti CSI [ST] vypočítaného z púltónů na subjektivním věku.

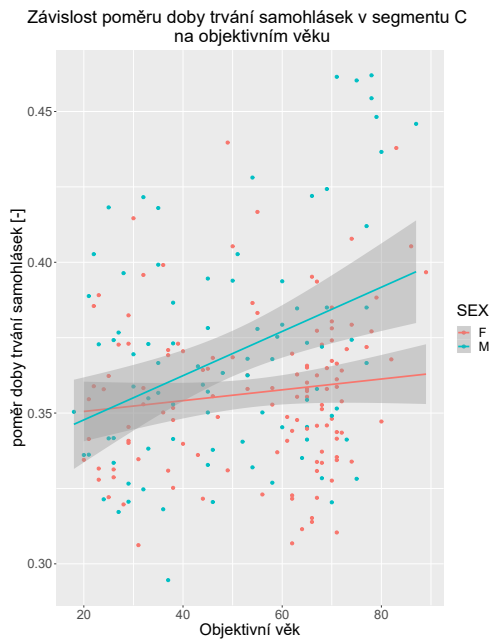
pro objektivní věk a pro subjektivní věk $r_M = 0.18$. U žen vychází korelační koeficient $r_F = 0.46$ pro objektivní věk a pro subjektivní věk $r_F = 0.52$.



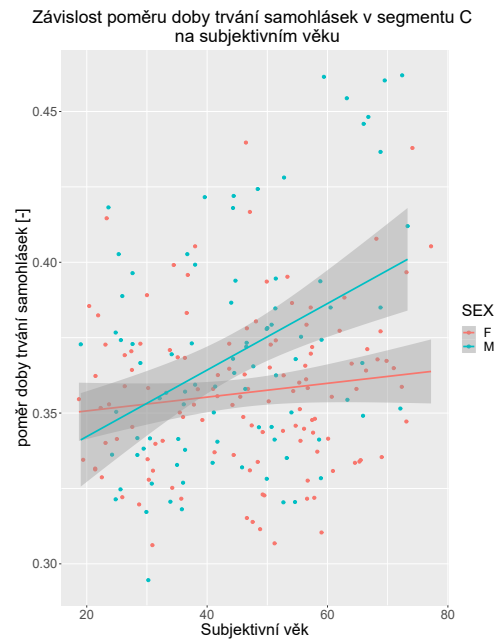
Obrázek 5.3.9. Závislost průměrného kumulativního indexu strmosti CSI [dB] vypočítaného z intenzity na objektivním věku.



Obrázek 5.3.10. Závislost průměrného kumulativního indexu strmosti CSI [dB] vypočítaného z intenzity na subjektivním věku.



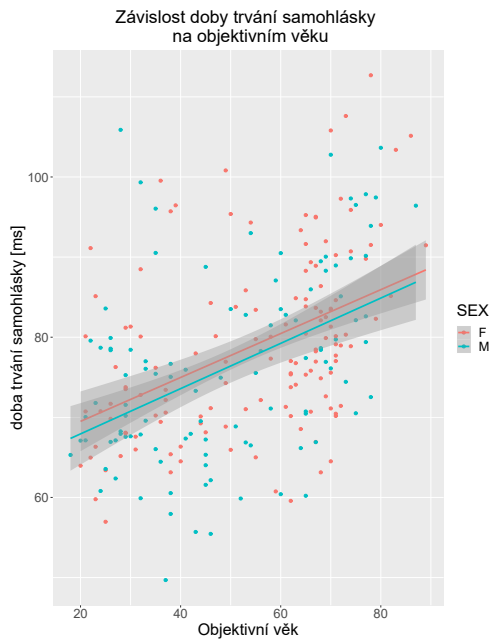
Obrázek 5.4.11. Závislost poměru doby trvání samohlásek v segmentu C [-] na objektivním věku.



Obrázek 5.4.12. Závislost poměru doby trvání samohlásek v segmentu C [-] na subjektivním věku.

5.5 Průměrná doba trvání samohlásky

Grafy 5.5.13, 5.5.14 zobrazují průběhy průměrné délky trvání samohlásky v závislosti na objektivním a subjektivním věku mluvčích. Korelační koeficient pro muže $r_M = 0.45$ pro objektivní věk a pro subjektivní věk $r_M = 0.54$. U žen vychází korelační koeficient



Obrázek 5.5.13. Závislost průměrné doby trvání samohlásky [ms] na objektivním věku.



Obrázek 5.5.14. Závislost průměrné doby trvání samohlásky [ms] na subjektivním věku.

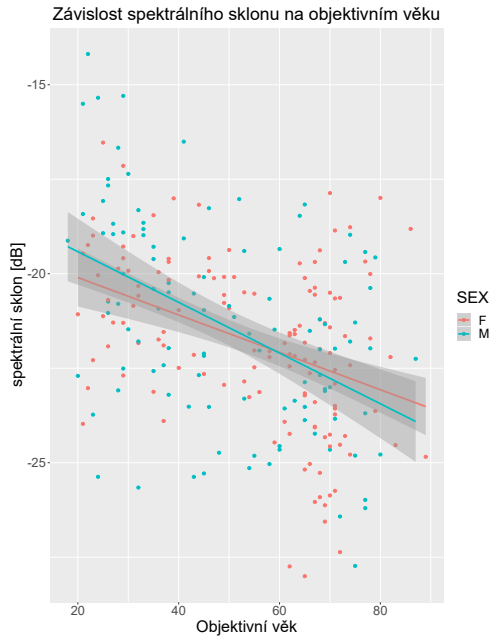
$r_F = 0.44$ pro objektivní věk a pro subjektivní věk $r_F = 0.54$.

5.6 Spektrální sklon

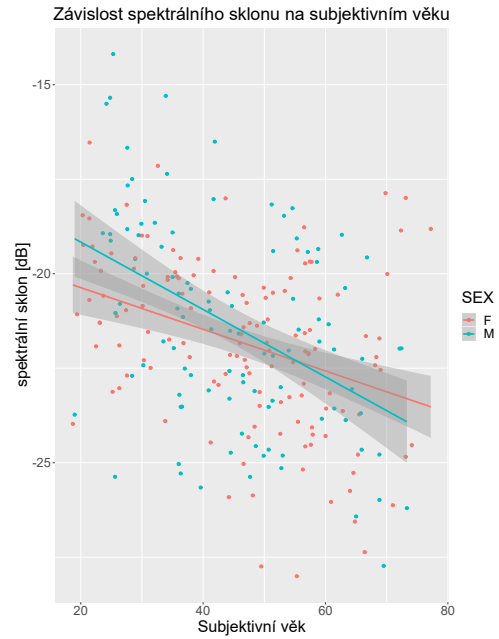
Grafy 5.6.15, 5.6.16 zobrazují průběhy průměrného spektrálního sklonu v závislosti na objektivním a subjektivním věku mluvčích. Korelační koeficient pro muže $r_M = -0.46$ pro objektivní věk a pro subjektivní věk $r_M = -0.45$. U žen vychází korelační koeficient $r_F = -0.4$ pro objektivní věk a pro subjektivní věk $r_F = -0.37$.

5.7 Dynamika intenzity a intonace

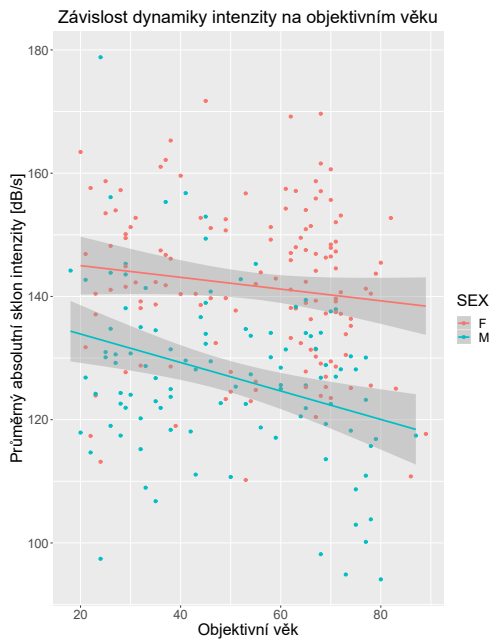
Grafy 5.7.17, 5.7.18 zobrazují průběhy průměrné dynamiky intenzity v závislosti na objektivním a subjektivním věku mluvčích. Korelační koeficient dynamiky intenzity pro muže $r_M = -0.32$ pro objektivní věk a pro subjektivní věk $r_M = -0.33$. U žen vychází korelační koeficient $r_F = -0.13$ pro objektivní věk a pro subjektivní věk $r_F = -0.19$. Grafy 5.7.19 5.7.20 zobrazují průběh průměrné dynamiky intonace v závislosti na objektivním a subjektivním věku mluvčích. Korelační koeficient dynamiky intonace pro muže $r_M = 0.45$ pro objektivní věk a pro subjektivní věk $r_M = 0.36$. U žen vychází korelační koeficient $r_F = 0.25$ pro objektivní věk a pro subjektivní věk $r_F = 0.28$.



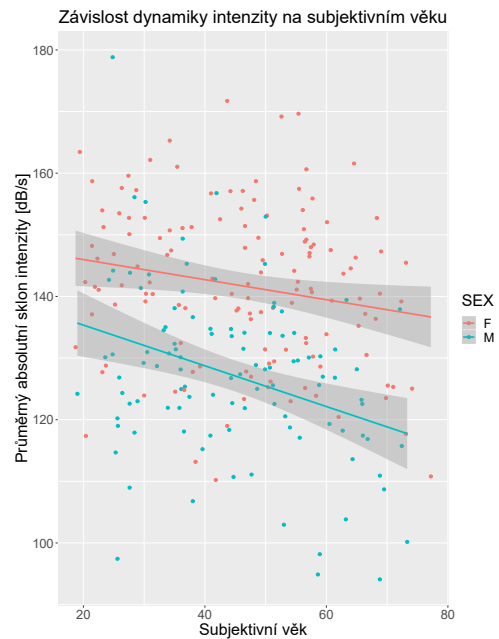
Obrázek 5.6.15. Závislost průměrného spektrálního sklonu [dB] na objektivním věku.



Obrázek 5.6.16. Závislost průměrného spektrálního sklonu [dB] na subjektivním věku.



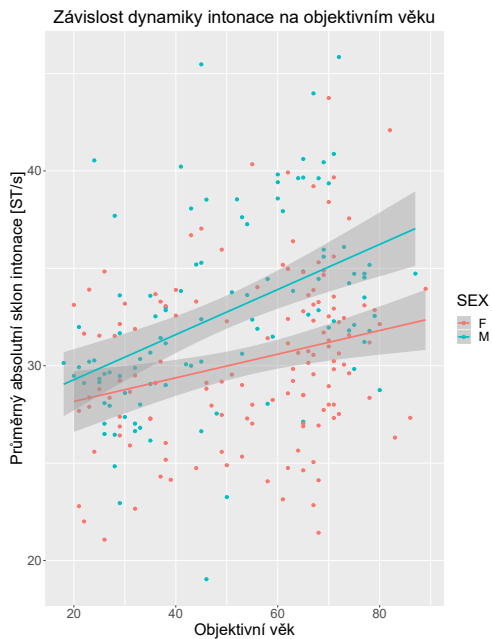
Obrázek 5.7.17. Závislost průměrné dynamiky intenzity [dB/s] na objektivním věku.



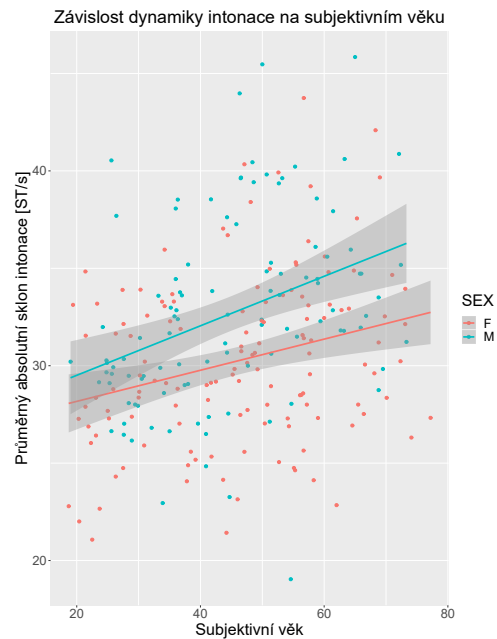
Obrázek 5.7.18. Závislost průměrné dynamiky intenzity [dB/s] na subjektivním věku.

5.8 Formanty

Z důvodu nedostatku výskytu samohlásek [a:], [e:], [i:], [o:], [u:] bylo grafické zpracování provedeno pouze pro [a], [e], [i], [o], [u]. Výsledné grafy zobrazují průběhy průměrných frekvencí F1 a F2, uvedeme pouze výsledky F1 pro [i] (5.8.21, 5.8.22) a F2 pro [o] (5.8.23, 5.8.24). Zbylé grafy budou uvedeny v kapitole Příloha A. Průměrné frekvence



Obrázek 5.7.19. Závislost průměrné dynamiky intonace v púltónech [ST/s] na objektivním věku.



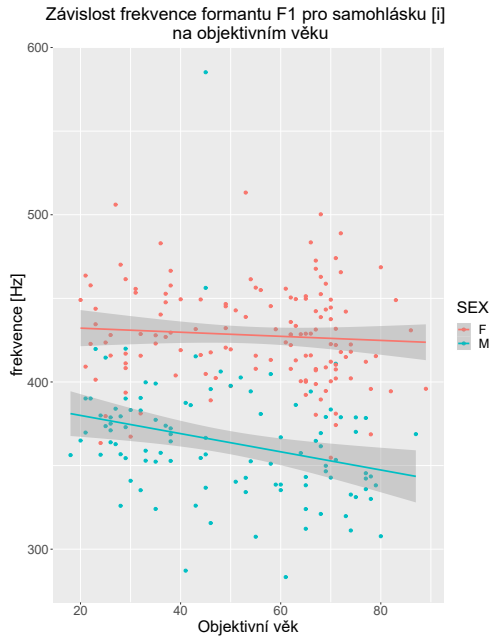
Obrázek 5.7.20. Závislost průměrné dynamiky intonace v púltónech [ST/s] na subjektivním věku.

formantů byly určeny dvěma způsoby, dále uvádíme pouze výsledky z výpočtu pomocí třetin. Korelační koeficienty počítané z výpočtu pomocí poloviny intervalu vycházejí téměř stejně.

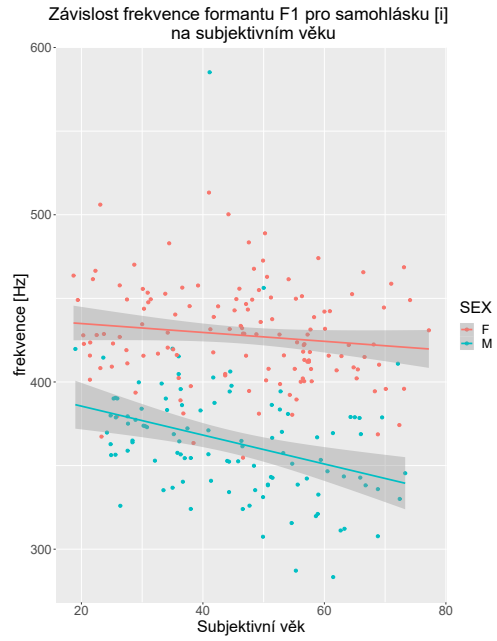
Korelační koeficient F1 pro [i] pro muže $r_M = -0.28$ pro objektivní věk a pro subjektivní věk $r_M = -0.33$. U žen vychází korelační koeficient $r_F = -0.08$ pro objektivní věk a pro subjektivní věk $r_F = -0.14$. Korelační koeficient F2 pro [o] pro muže $r_M = -0.11$ pro objektivní věk a pro subjektivní věk $r_M = -0.03$. U žen vychází korelační koeficient $r_F = -0.41$ pro objektivní věk a pro subjektivní věk $r_F = -0.41$.

5.9 Mnohonásobná lineární regrese

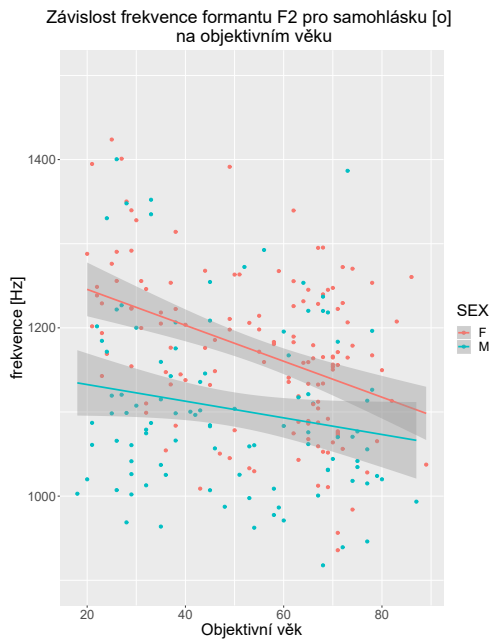
Interval odchylek výsledného modelu pro objektivní věk mužů je $\langle -7.971, 9.707 \rangle$, pro ženy $\langle -10.448, 10.267 \rangle$. Pro subjektivní věk vychází intervaly odchylek pro muže $\langle -6.130, 6.215 \rangle$ a pro ženy $\langle -6.025, 5.923 \rangle$.



Obrázek 5.8.21. Závislost průměrné frekvence F1 [Hz] pro samohlásku [i] na objektivním věku.



Obrázek 5.8.22. Závislost průměrné frekvence F1 [Hz] pro samohlásku [i] na subjektivním věku.



Obrázek 5.8.23. Závislost průměrné frekvence F2 [Hz] pro samohlásku [o] na objektivním věku.



Obrázek 5.8.24. Závislost průměrné frekvence F2 [Hz] pro samohlásku [o] na subjektivním věku.

Kapitola 6

Diskuze

Výsledky analýz vybraných parametrů uvedené v kapitole 5 ukazují, že ne všechny analyzované charakteristiky hlasu jsou závislé na věku mluvčího. Nevhodným parametrem se nakonec ukázal poměr délky vokálů k trvání samohlásky popsany v sekci 4.4.4, jelikož není možné tento parametr zautomatizovat – mluvčí musí říkat přesně stejné věty jako v datasetu, nelze ani tento parametr nějak zobecnit, zprůměrovat jako např. průměrnou základní frekvenci mluvčího.

Pro rozhodování, zda existuje nějaká závislost mezi daným parametrem a věkem, používáme Pearsonův korelační koeficient. Nevýhodou takto počítaného korelačního koeficientu je, že zkoumá pouze lineární závislost. Jak již vyplynulo z kapitoly 3, jednotlivé charakteristiky hlasu závislé na věku mohou být rozdílné pro muže a ženy. Je důležité zdůraznit, že věkové rozpětí se liší pro objektivní a subjektivní věk. U žen se mluvčí pohybovaly ve věku 20 – 89 let. U rozpětí průměrného subjektivního věku kvůli průměrování dochází i k výskytu necelých čísel, tzn. 18.7 – 77 let je rozpětí pro subjektivní věk žen. U mužů byli mluvčí ve věku 18 – 87 let a rozpětí pro subjektivní věk je 19 – 73.3 let. Pro snazší popis bude v následujících podkapitolách hovořeno o poklesu, či nárůstu určitého parametru během stárnutí mluvčích. Období stárnutí bude vymezeno výše uvedenými intervaly tzn. pro objektivní věk u mužů se bude jednat o rozpětí 18 – 87 let, pro subjektivní věk u žen naopak o rozpětí 18.7 – 77 let, pokud nebude řečeno jinak.

6.1 Zhodnocení akustických charakteristik hlasu žen

Z Pearsonových korelačních koeficientů a grafického znázornění jednotlivých parametrů vyplývá, že akustickými charakteristikami hlasu nejvíce lineárně závislými na věku, ať subjektivním, či objektivním, jsou průměrná základní frekvence, průměrné tempo řeči, kumulativní index strmosti vypočítán z intenzity a pultónů, spektrální sklon, průměrná doba trvání samohlásky a průměrná frekvence formantu F2 pro samohlásku [o] a [u]. Přehled těchto korelačních koeficientů je v tabulce 6.1.1.

Závěry zřejmé z grafu 5.1.1 se shodují s hypotézami ze studií [5], [22], kde u žen dochází ke kontinuálnímu poklesu v průměrné základní frekvenci f_0 , tento trend se potvrdil i u věku subjektivní (graf 5.1.2), což je ve shodě se studií [9]. Z výsledků analýzy vyplývá, že v průměru dochází u žen k poklesu průměrné základní frekvence o 10.43 – 38.4 Hz u objektivního věku. V případě subjektivního věku u žen je pokles výraznější, a to o 12.69 – 40.43 Hz.

Tempo řeči u žen vykazuje pokles pro objektivní a subjektivní věk (grafy 5.2.3, 5.2.4). Tyto výsledky se shodují s výsledky studií [22], [9] a [28]. Zpomalování průměrného tempa řeči je nejčastěji přisuzováno celkovému zpomalení motorické činnosti, které provází proces stárnutí.

Z grafů 5.3.9, 5.3.10 je zřejmé, že kumulativní index strmosti CSI vypočítaný z intenzity roste u žen výrazněji než u mužů a nárůst je větší pro subjektivní věk. Pro objektivní

věk se nárůst CSI pohybuje mezi 4.44 – 9.65 dB, u subjektivního věku 5.56 – 10.72 dB. Nárůst hodnoty tohoto parametru je v souladu s výsledky studie [28]. K nárůstu dochází také v případě CSI vypočítaného z pŕltónů pro objektivní i subjektivní věk (5.3.7, 5.3.8). Naopak z výsledků analýzy spektrálního sklonu je patrné, že hodnota spektrálního sklonu klesá s věkem jak subjektivním, tak objektivním (5.6.15, 5.6.16). Průměrná doba trvání samohlásky se prodlužuje v případě objektivního i subjektivního věku. Výraznější nárůst zaznamenáváme u subjektivního věku (5.5.14). Prodlužování doby trvání samohlásek je nejspíše do jisté míry zapřičiněno obdobně jako zpomalování tempa řeči kvůli zpomalování motorických činností.

Výsledky analýz frekvencí formantů F1 a F2 ukazují, že jedinými souhláskami, které vykazují závislost na objektivním, či subjektivním věku u žen, jsou souhlásky [o] (5.8.23, 5.8.24) a [u] (6.3.37, 6.3.38). Dle hodnot korelačních koeficientů byly vybrány pouze formanty F2, jelikož korelační koeficienty pro frekvenci formantu F1 se pohybují pouze okolo 0.1. Korelační koeficient určující lineární závislost mezi frekvencí F2 pro samohlásku a objektivním věkem u žen $r_F = -0.29$, tj. dochází k poklesu hodnot v průběhu stárnutí. Tento výsledek se shoduje s výsledky [22].

Ostatní vybrané parametry mají vyšší korelační koeficienty, a proto není frekvence F2 pro samohlásku [a] uvedena v Tabulce 6.1.1.

parametr	r_F OV ženy	r_F SV ženy
průměrná f_0 [Hz]	-0.32	-0.34
tempo řeči [syll/s]	-0.53	-0.6
CSI z intenzity [dB]	0.46	0.52
CSI z pŕltónů [ST]	0.36	0.41
spektrální sklon [dB]	-0.4	-0.37
průměrná doba trvání samohlásky [ms]	0.44	0.54
průměrná frekvence formantu F2 pro [o]	-0.41	-0.41
průměrná frekvence formantu F2 pro [u]	-0.38	-0.37

Tabulka 6.1.1. Přehled korelačních koeficientů vyjadřující nejsilnější lineární závislost vybraných akustických parametrů s objektivním a subjektivním věkem pro ženy. OV označuje objektivní věk, SV věk subjektivní. Souhrn výsledků z kapitoly 5.

6.2 Zhodnocení akustických charakteristik hlasu mužů

Akustické charakteristiky hlasu, které se mění s věkem a jejich korelační koeficienty naznačují lineární závislost jsou u mužů rozdílné než u žen. V případě mužů se jedná o průměrné tempo řeči, kumulativní index strmosti vypočítaný ze základní frekvence a pŕltónů, dynamika intonace a intenzity, spektrální sklon, průměrná doba trvání samohlásky a průměrná frekvence formantu F2 pro samohlásku [i]. Pro lepší přehlednost jsou korelační koeficienty uvedeny v Tabulce 6.2.2.

Výsledky analýzy průměrné základní frekvence f_0 u mužů se shodují s výsledky studií [5], [22], ovšem mezi vybrané vhodné parametry nepatří, jelikož je korelační koeficient $r_F = 0.12$ pro objektivní věk a $r_F = 0.14$ pro věk subjektivní. Z grafů 5.2.3, 5.2.4 je zřejmé, že u průměrného tempa řeči dochází i u mužů ke kontinuálnímu poklesu v průběhu procesu stárnutí. Tyto závěry se shodují s výsledky studií [22], [9] a [28], kde je kontinuální pokles přisuzován nejčastěji zpomalování motorických činností se zvyšujícím se věkem. CSI vypočítané ze základní frekvence a z pŕltónu

jsou na sobě závislé parametry, jelikož CSI vypočítané z pŕltónů získáme převodem základní frekvence a následným výpočtem samotného CSI. U žen měl vyšší korelační koeficient pouze CSI z pŕltónů, zde je určitá lineární závislost i mezi věkem mužů a CSI vypočítaného ze základní frekvence. Z grafů 5.3.7, 5.3.8 je patrný kontinuální nárůst hodnot pro CSI z pŕltónů, tento parametr má zároveň nejvyšší korelační koeficient pro muže pro objektivní i subjektivní věk, tzn. lineární závislost na věku je nejsilnější mezi nalezenými parametry.

Z výsledků analýzy dynamiky intonace (grafy 5.7.19, 5.7.20) je patrné, že v průběhu stárnutí dochází k nárůstu, v tomto případě zrychlení, a to v rozsahu 4.43 – 11.55 ST/s pro objektivní věk a v případě subjektivního věku nárůstu v rozsahu 3.01 – 10.81 ST/s, který je menší, což zároveň odpovídá i nižšímu korelačnímu koeficientu viz 6.2.2.

Hodnoty parametru dynamika intenzity naopak rostou s vyšším věkem, jak je zřejmé z grafů 5.7.17 a 5.7.18. Pro objektivní věk dochází k nárůstu v rozsahu 5.32 – 26.56 dB/s, pro subjektivní věk je rozsah větší (6.86 – 29 dB/s).

Výsledky analýzy spektrálního sklonu ukazují, že u mužů dochází také k poklesu se zvyšujícím se objektivním i subjektivním věkem. (5.2.3, 5.2.4) Ve shodě se ženami jsou také výsledky analýzy průměrné doby trvání samohlásky, kde s narůstajícím objektivním a subjektivním věkem dochází k prodlužování samohlásek; tento trend nejspíše také souvisí s postupným zpomalováním tempa řeči, jak je již uvedeno v sekci 6.1.

Jediná analýza jednotlivých frekvencí formantů F1 a F2 pro samohlásky, u které vyšel korelační koeficient srovnatelný s ostatními vybranými parametry, byla u mužů analýza frekvence formantu F2 pro samohlásku [i] (6.3.35, 6.3.36). Parametr frekvence formantu F1 samohlásky [i] by se také mohl zařadit, jelikož korelační koeficient vychází -0.28 pro objektivní věk, ale byly upřednostněny parametry udávající silnější lineární závislost na věku mužů. Bohužel žádná z uvedených studií v kapitole 3 se nevěnovala přímo analýze této samohlásky a není zde žádné porovnání s ostatními výsledky jednotlivých studií.

parametr	r_F OV muži	r_F SV muži
tempo řeči [syll/s]	-0.33	-0.44
CSI ze základní frekvence [Hz]	0.47	0.52
CSI z pŕltónů [ST]	0.6	0.64
dynamika intonace [ST/s]	0.45	0.36
dynamika intenzity [dB/s]	-0.32	-0.33
spektrální sklon [dB]	-0.46	-0.45
průměrná doba trvání samohlásky [ms]	0.45	0.54
průměrná frekvence formantu F2 pro [i]	0.33	0.4

Tabulka 6.2.2. Přehled korelačních koeficientů vyjadřující nejsilnější lineární závislost vybraných akustických parametrů s objektivním a subjektivním věkem pro muže. OV označuje objektivní věk, SV věk subjektivní. Souhrn výsledků z kapitoly 5.

6.3 Porovnání výsledků pro obě pohlaví

Z tabulek 6.1.1 a 6.2.2 je patrné, že pro zkoumání věku jak objektivního, tak subjektivního jsou pro obě pohlaví vhodné následující parametry: tempo řeči, CSI vypočítané z pŕltónů, spektrální sklon a průměrná doba trvání samohlásky. Jelikož

tempo řeči je nejčastěji přisuzované přirozenému zpomalování motorických činností spojených se stárnutím, je změna v tomto parametru zřejmá pro obě pohlaví. S tímto souvisí i průměrná doba trvání samohlásky. Naopak naprosto nevhodným parametrem je poměr délky trvání vokálů, kde mimo jiné se limitujeme na konkrétní formulace a nelze tedy tento parametr zobecnit. Dle výsledků analýz frekvence formantů F1 a až na výjimky F2 můžeme usoudit, že závislost vzhledem objektivnímu a subjektivnímu věku je v porovnání s ostatními parametry slabá.

Výsledné parametry byly vždy vhodné pro objektivní i subjektivní věk a až na výjimky (např. spektrální sklon u žen viz 6.1.1) byly nalezené závislosti silnější pro věk subjektivní. Problémem určování subjektivního věku je především tvorba trénovacích dat. Objektivní věk je jednoznačně dán, a proto i přes slabší závislosti na parametrech bych označila automatizaci analýzy nahrávek a určování objektivního věku snazší než určování subjektivního věku.

Studie [28] pracovala s nahrávkami použitými jako dataset v této bakalářské práci. V článku ovšem nebylo modelování prováděno s křížovou validací a nebyl vytvořen model pro odhadování subjektivního věku. Vedoucí mé práce, pan Ing. Tomáš Bořil Ph.D, ochotně přidal křížovou validaci a model pro odhadování subjektivního věku pro porovnání s výsledky mnou vybraných parametrů analyzovaných v této bakalářské práci. Výsledky z [28] a 5.9 jsou pro přehlednost uvedeny v Tabulkách 6.3.3 a 6.3.4, kde intervaly znázorňují mezikvartilové rozpětí (IQR).

	OV muži	SV muži
interval odchylek výsledků z [28]	$\langle -12.154, 10.543 \rangle$	$\langle -7.789, 8.712 \rangle$
interval odchylek výsledků z 5.9	$\langle -7.971, 9.707 \rangle$	$\langle -6.130, 6.215 \rangle$

Tabulka 6.3.3. Přehled výsledných intervalů odchylek modelu odhadovaného věku mužů ze studie [28] a modelu na základě výsledku analýzy akustických parametrů vybraných pro tuto bakalářskou práci. OV označuje objektivní věk, SV věk subjektivní.

	OV ženy	SV ženy
interval odchylek výsledků z [28]	$\langle -11.811, 10.681 \rangle$	$\langle -8.703, 8.780 \rangle$
interval odchylek výsledků z 5.9	$\langle -10.448, 10.267 \rangle$	$\langle -6.025, 5.923 \rangle$

Tabulka 6.3.4. Přehled výsledných intervalů odchylek modelu odhadovaného věku žen ze studie [28] a modelu na základě výsledku analýzy akustických parametrů vybraných pro tuto bakalářskou práci. OV označuje objektivní věk, SV věk subjektivní.

Je zřejmé, že v případě objektivního i subjektivního věku dosahuje lepších výsledků model rozšířený i o zde vybrané parametry než pouze s tempem řeči, kumulativního indexu strmosti vypočítaného z intenzity a púltónů, jak tomu je v případě modelu z [28]. Oba modely mají lepší výsledky pro subjektivní věk, což by se dalo částečně také přisoudit faktu, že nahrávky namlouvali mluvčí i z univerzity 3. věku a často se jednalo o lidi, kteří se aktivně celý život věnují sborovému zpěvu apod. Kondice jejich hlasu neodpovídá objektivnímu věku, což mohlo zkreslit a zhoršit výsledky odhadu objektivního věku.

Všecké parametry, které ve svých výpočtech používají soubory typu .TextGrid, vyžadují manuální předzpracování. Jedná se o tempo řeči, všechny varianty CSI, průměrná délka samohlásky a formanty. Parametry průměrná základní frekvence f_0 ,

průměrný spektrální sklon, průměrná dynamika intenzity a intonace vyžadují jako vstupní data pouze soubory typu .wav, kde je manuální předzpracování minimální. V případě průměrné základní frekvence ovšem je lepší pracovat s manuálně opravenými konturami, takže by bylo vhodné tento parametr zařadit spíše do kategorie parametrů, které potřebují manuální předzpracování.

Kapitola 7

Závěr

Cílem této bakalářské práce bylo seznámit se s problematikou a vybrat a změřit vhodné akustické charakteristiky v závislosti na věku. Dále na základě výsledků měření určit, zda jsou jednotlivé parametry vhodné pro účely automatické analýzy nahrávek.

Po nastudování problematiky a souvisejících studií byly vybrány parametry vhodné pro analýzu nahrávek. Dle výsledků studií (kapitola 3) je zřejmé, že vhodné parametry pro zkoumání mužských a ženských hlasů se liší. Nahrávky bylo nutné zvalidovat a analýzy provádět pouze na validních a bezchybných položkách. K nahrávkám byl znám objektivní věk mluvčích a dále byly poskytnuty i odhady věku subjektivního, které byly zprůměrovány. Následně byly implementovány výpočty parametrů pomocí Matlabu a Praatu. Na základě výsledků jednotlivých analýz byly vyřazeny některé parametry – poměr délky trvání samohlásek v segmentu, frekvence formantů F1 a až na výjimky frekvence formantů F2, protože nevykazovaly závislost vzhledem k věku. Za vhodné parametry byly označeny tempo řeči, základní frekvence, kumulativní index strmosti, spektrální sklon, dynamika intenzity, dynamika intonace a průměrná délka trvání samohlásky.

Pro určení použitelnosti parametrů k analýze nahrávek v závislosti na věku byl použit Pearsonův korelační koeficient, který vyjadřuje míru lineární závislosti mezi danými parametry. Obecně vycházely závislosti silnější pro subjektivní věk. Nejsilnější lineární závislost pro muže vykazuje kumulativní index strmosti (CSI) vypočítaný z pŕltónů s korelačním koeficientem $r_F = 0.64$ pro subjektivní věk.

Na závěr práce byly na základě hodnot korelačních koeficientů vybrány vhodné parametry pro zkoumání ženského a mužského hlasu. Analýza parametrů a jejich závislosti na subjektivním věku byly provedeny nad rámec zadání této práce, závislosti na subjektivním věku vyšly většinou silnější, ale pro automatizaci je nutné zvážít určení subjektivního věku, které je výrazně složitější než pro věk objektivní, jenž je jednoznačně daný.

Z vybraných parametrů by bylo možné vytvořit modely, které by mohly predikovat věk objektivní a subjektivní u zdravých jedinců. S parametry zkoumanými v této práci se interval odchylek odhadu modelu zúžil oproti modelu na základě výsledků práce [28], která pracovala se stejnými nahrávkami. Lepších výsledků dosahují modely v predikci subjektivního věku a v případě objektivního věku je přesnější predikce mužů. Zároveň dílčí analýzy lze využít jako referenční hodnoty v detekci a zkoumání vlivů různých onemocnění na hlase.

Literatura

- [1] ČAPEK, K. *Měl jsem psa a kočku*. Melantrich, Praha, 1996.
- [2] BOERSMA, P., AND WEENINK, D. Praat: doing phonetics by computer [Computer program]. Version 6.1.10, 2020.
<http://www.praat.org/>.
- [3] BOŘIL, T., AND SKARNITZL, R. Tools rPraat and mPraat. In *Text, Speech, and Dialogue: 19th International Conference, TSD 2016, Brno, Czech Republic, September 12-16, 2016, Proceedings* (Cham, 2016), P. Sojka, A. Horák, I. Kopeček, and K. Pala, Eds., Springer International Publishing, pp. 367–374.
- [4] COX, V. O., AND SELENT, M. Acoustic and respiratory measures as a function of age in the male voice. *Journal of Phonetics & Audiology* 1, 1 (2015).
- [5] EICHHORN, J. T., KENT, R. D., AUSTIN, D., AND VORPERIAN, H. K. Effects of aging on vocal fundamental frequency and vowel formants in men and women. *Journal of Voice* 32, 5 (Sept. 2018), 644.e1–644.e9.
- [6] FANT, G. *Acoustic theory of speech production with calculations based on X-ray studies of Russian articulations*. Mouton, s'Gravenhage, 1960.
- [7] FLETCHER, A. R., MCAULIFFE, M. J., LANSFORD, K. L., AND LISS, J. M. The relationship between speech segment duration and vowel centralization in a group of older speakers. *The Journal of the Acoustical Society of America* 138, 4 (Oct. 2015), 2132–2139.
- [8] GICK, B. *Articulatory phonetics*. John Wiley & Sons Inc, Malden, MA, 2013.
- [9] HARNSBERGER, J. D., SHRIVASTAV, R., BROWN, W., ROTHMAN, H., AND HOLLIEN, H. Speaking rate and fundamental frequency as speech cues to perceived age. *Journal of Voice* 22, 1 (Jan. 2008), 58–69.
- [10] HARRINGTON, J., PALETHORPE, S., AND WATSON, C. I. Age-related changes in fundamental frequency and formants: a longitudinal study of four speakers. In *in InterSpeech* (2007).
- [11] KENT, R. D., AND VORPERIAN, H. K. Static measurements of vowel formant frequencies and bandwidths: A review. *Journal of Communication Disorders* 74 (July 2018), 74–97.
- [12] KRČMOVÁ, M. *Fonetika a fonologie*. Masarykova Univ, 2008.
- [13] LINVILLE, S. *Vocal aging*. Singular Thomson Learning, Australia San Diego, 2001.
- [14] LINVILLE, S. E., AND RENS, J. Vocal tract resonance analysis of aging voice using long-term average spectra. *Journal of Voice* 15, 3 (Sept. 2001), 323–330.
- [15] LISS, J., WEISMER, G., AND ROSENBEK, J. Selected acoustic characteristics of speech production in very old males. *Journal of Gerontology: Social Sciences* 45, 2 (1990).
- [16] MACHAČ, P., AND SKARNITZL, R. *Fonetická segmentace hlásek*. Epocha, Praha, 2010.

- [17] MAOR, E., SARA, J. D., ORBELO, D. M., LERMAN, L. O., LEVANON, Y., AND LERMAN, A. Voice signal characteristics are independently associated with coronary artery disease. *Mayo Clinic Proceedings* 93, 7 (July 2018), 840–847.
- [18] MATLAB. *version 9.5.0 (R2018b)*. The MathWorks Inc., Natick, Massachusetts, 2018.
- [19] MICHAEL, A., AND JOHN, M. *Úvod do obecné fonetiky*. Charles University in Prague, Karolinum Press, 2016.
- [20] PALKOVÁ, Z. *Fonetika a fonologie češtiny: s obecným úvodem do problematiky oboru*. Univerzita Karlova, vydavatelství Karolinum, 1994.
- [21] SADEGHIAN, R., SCHAFFER, J. D., AND ZAHORIAN, S. A. Speech processing approach for diagnosing dementia in an early stage. In *Interspeech 2017* (Aug. 2017), ISCA.
- [22] SCHÖTZ, S. Acoustic analysis of adult speaker age. In *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2007, pp. 88–107.
- [23] SCHÖTZ, S., AND MÜLLER, C. A study of acoustic correlates of speaker age. In *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2007, pp. 1–9.
- [24] ŠIMÁČKOVÁ, Š., PODLIPSKÝ, V. J., AND CHLÁDKOVÁ, K. Czech spoken in Bohemia and Moravia. *Journal of the International Phonetic Association* 42, 2 (2012), 225–232. Cambridge University Press.
- [25] SKARNITZL, R., ŠTURM, P., AND VOLÍN, J. *Zvuková báze řečové komunikace*, první ed. nakladatelství Karolinum, Praha, 2016.
- [26] TORRE, P., AND BARLOW, J. A. Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders* 42, 5 (Sept. 2009), 324–333.
- [27] VOLÍN, J., AND BOŘIL, T. Acoustic correlates of prosodic dimensions in younger and older speakers of Czech. *Acta Universitatis Carolinae – Philologica* 3, *Phonetica Pragensia XIV* (2017), 45–54.
- [28] VOLÍN, J., TYKALOVÁ, T., AND BOŘIL, T. Stability of prosodic characteristics across age and gender groups. In *Proceedings of Interspeech 2017* (2017), pp. 3902–3906.
- [29] WATSON, P. J., AND MUNSON, B. A comparison of vowel acoustics between older and younger adults. In *International Congress on Phonetic Sciences 13* (2007), pp. 561–564.
- [30] WEINGARTOVÁ, L., BOŘIL, T., AND VAŇKOVÁ, J. Spektrální sklon. In *Fonetická identifikace mluvího*, R. Skarnitzl, Ed. Univerzita Karlova v Praze, Filozofická fakulta, Praha, 2014, pp. 77–94.
- [31] WICKHAM, H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.

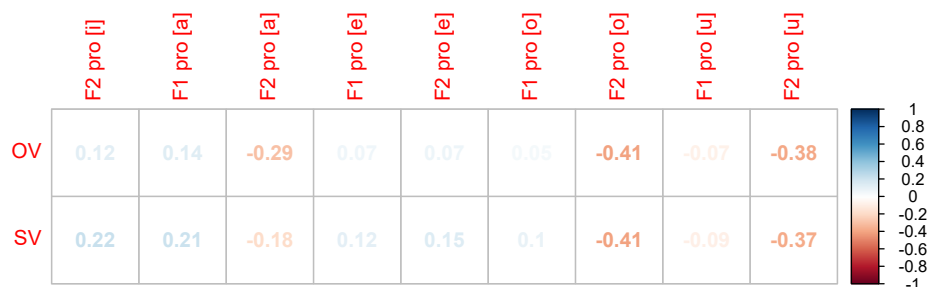
Příloha A

Přílohy

Korelační matice jednotlivých parametrů pro ženy

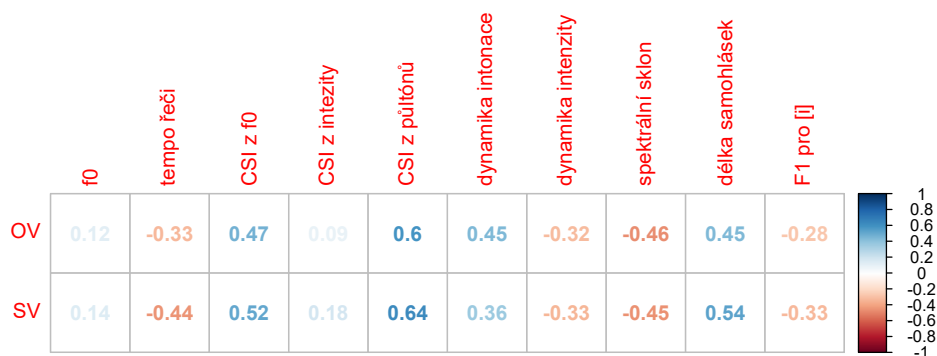


Korelační matice jednotlivých parametrů pro ženy

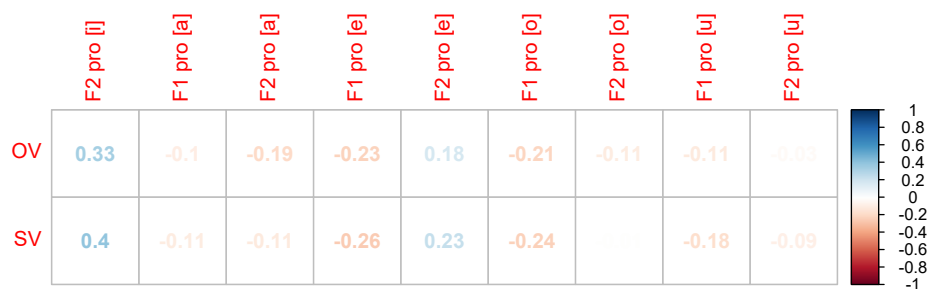


Obrázek 6.3.1. Výsledná korelační matice parametrů pro ženy. OV = objektivní věk, SV = subjektivní věk.

Korelační matice jednotlivých parametrů pro muže

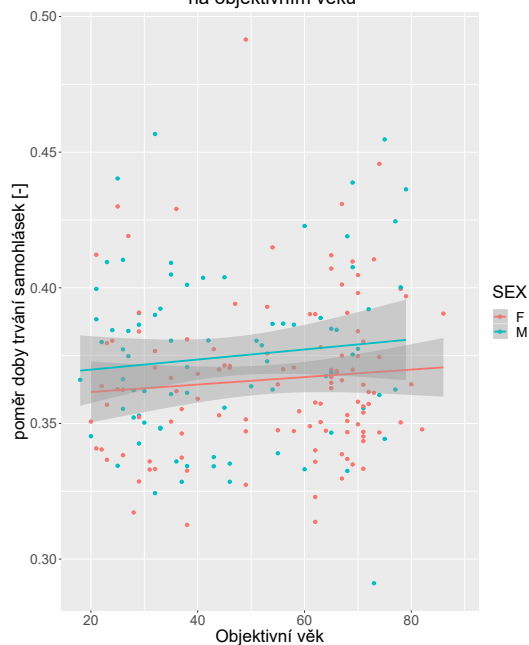


Korelační matice jednotlivých parametrů pro muže



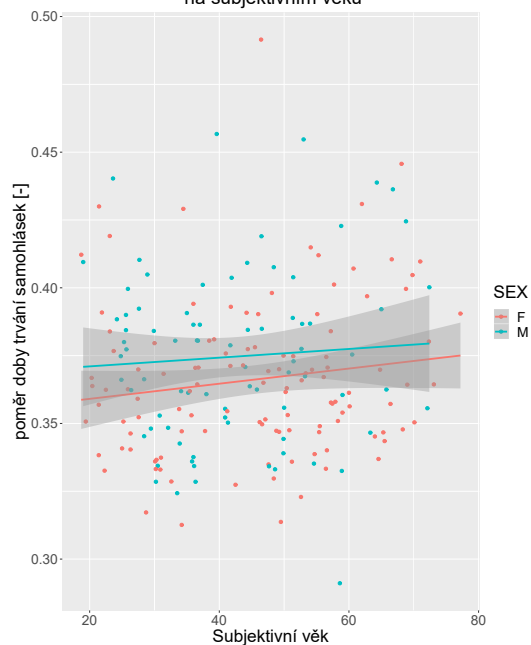
Obrázek 6.3.2. Výsledná korelační matice parametrů pro muže. OV = objektivní věk, SV = subjektivní věk.

Závislost poměru doby trvání samohlásek v segmentu A na objektivním věku



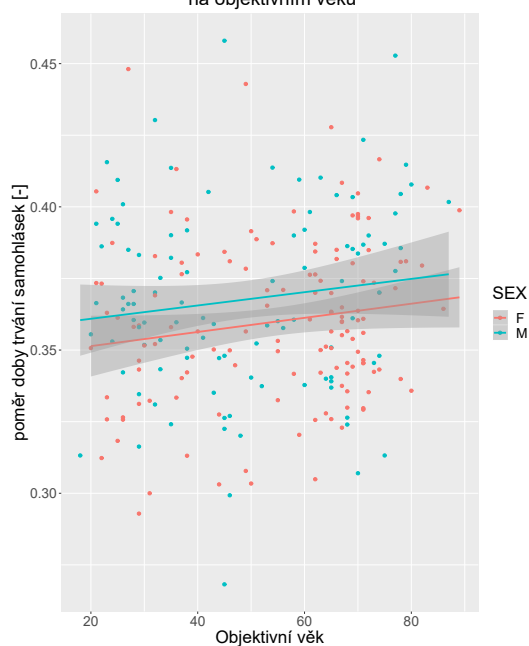
Obrázek 6.3.3. Závislost poměru délky trvání samohlásek v segmentu A [-] na objektivním věku.

Závislost poměru doby trvání samohlásek v segmentu A na subjektivním věku



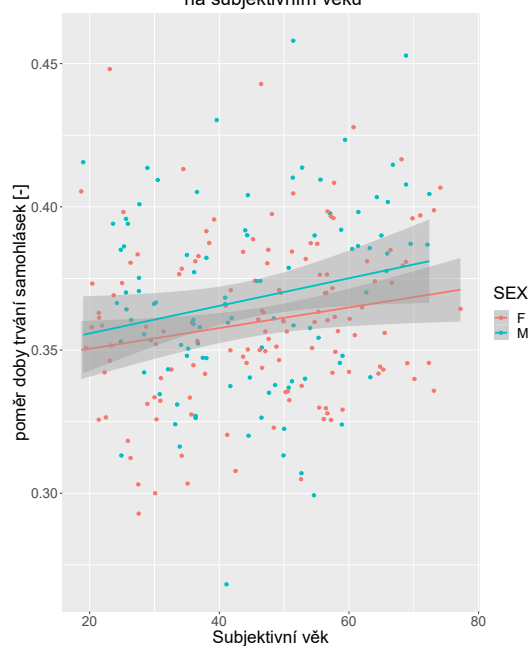
Obrázek 6.3.4. Závislost poměru délky trvání samohlásek v segmentu A [-] na subjektivním věku.

Závislost poměru doby trvání samohlásek v segmentu B na objektivním věku

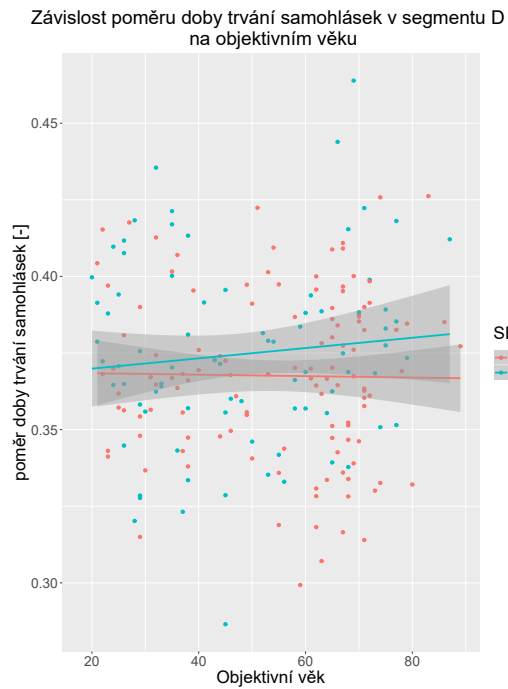


Obrázek 6.3.5. Závislost poměru délky trvání samohlásek v segmentu B [-] na objektivním věku.

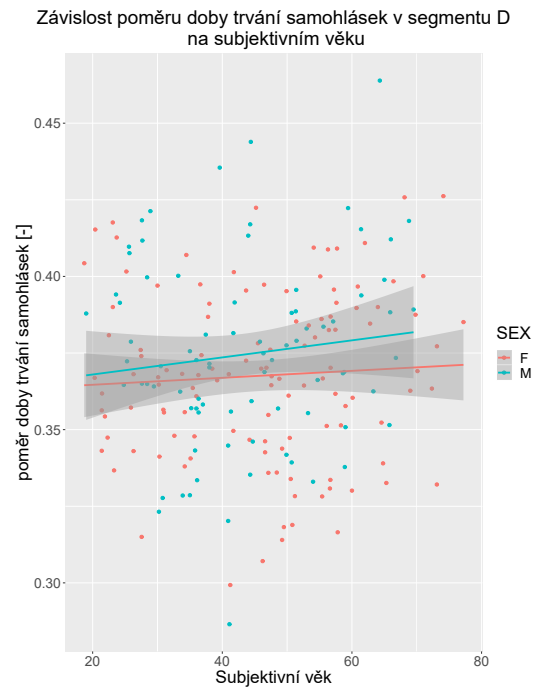
Závislost poměru doby trvání samohlásek v segmentu B na subjektivním věku



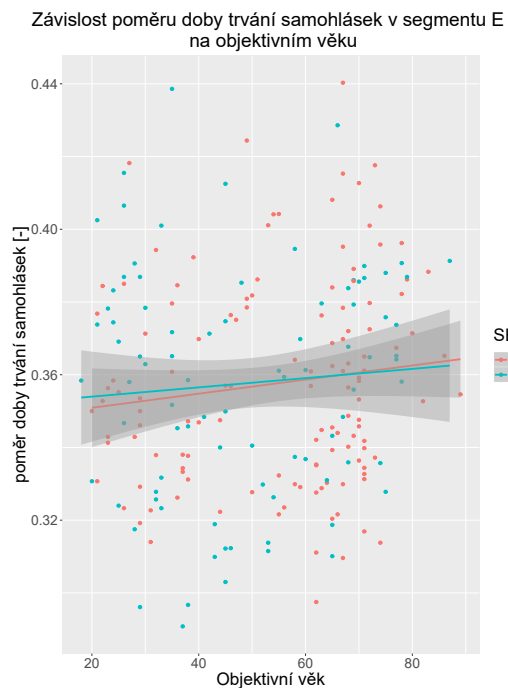
Obrázek 6.3.6. Závislost poměru délky trvání samohlásek v segmentu B [-] na subjektivním věku.



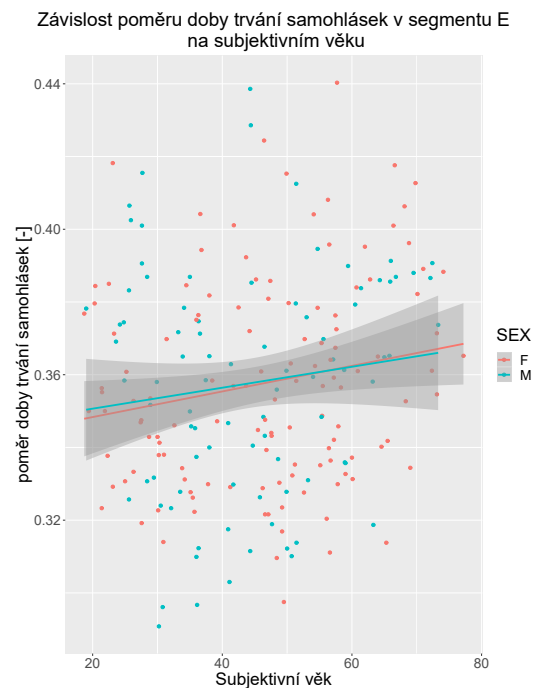
Obrázek 6.3.7. Závislost poměru délky trvání samohlásek v segmentu D [-] na objektivním věku.



Obrázek 6.3.8. Závislost poměru délky trvání samohlásek v segmentu D [-] na subjektivním věku.

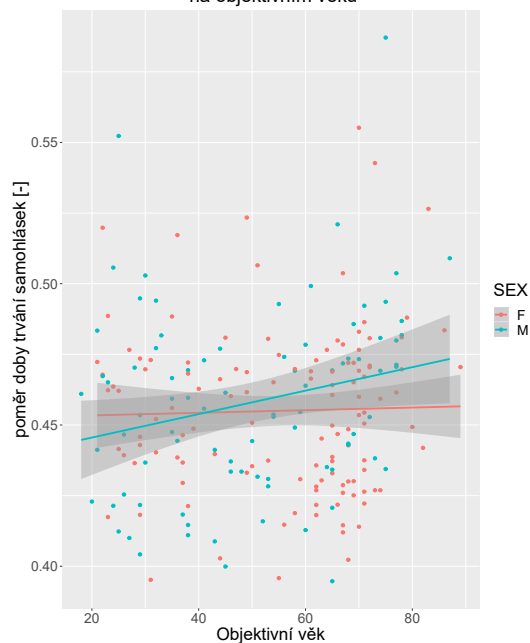


Obrázek 6.3.9. Závislost poměru délky trvání samohlásek v segmentu E [-] na objektivním věku.



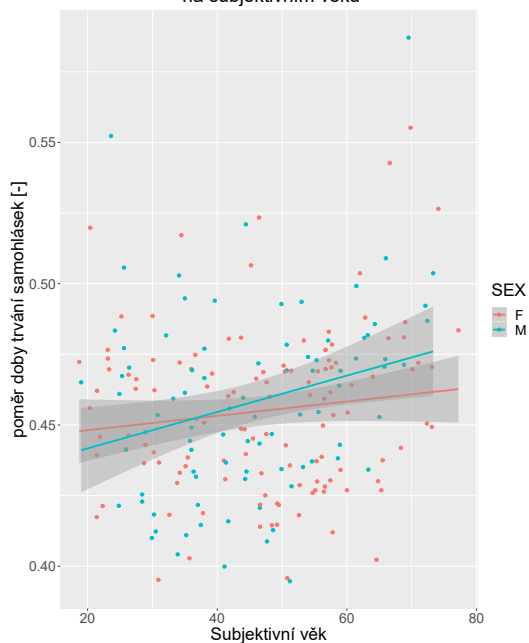
Obrázek 6.3.10. Závislost poměru délky trvání samohlásek v segmentu E [-] na subjektivním věku.

Závislost poměru doby trvání samohlásek v segmentu F na objektivním věku



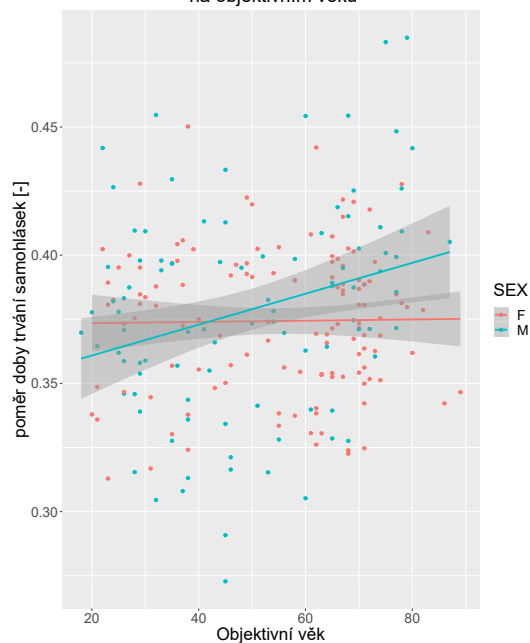
Obrázek 6.3.11. Závislost poměru délky trvání samohlásek v segmentu F [-] na objektivním věku.

Závislost poměru doby trvání samohlásek v segmentu F na subjektivním věku



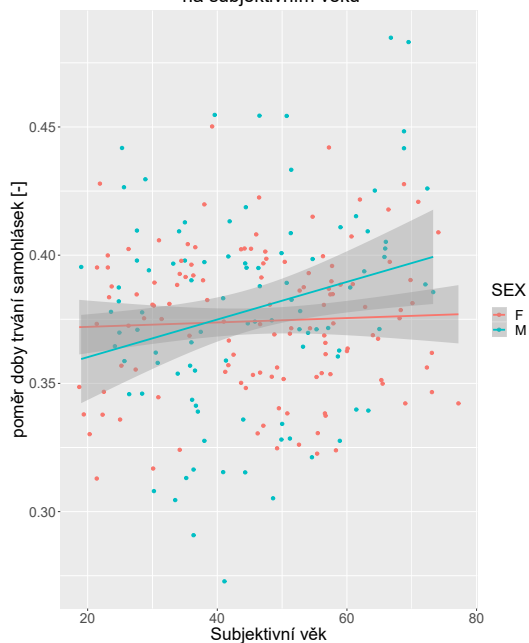
Obrázek 6.3.12. Závislost poměru délky trvání samohlásek v segmentu F [-] na subjektivním věku.

Závislost poměru doby trvání samohlásek v segmentu G na objektivním věku

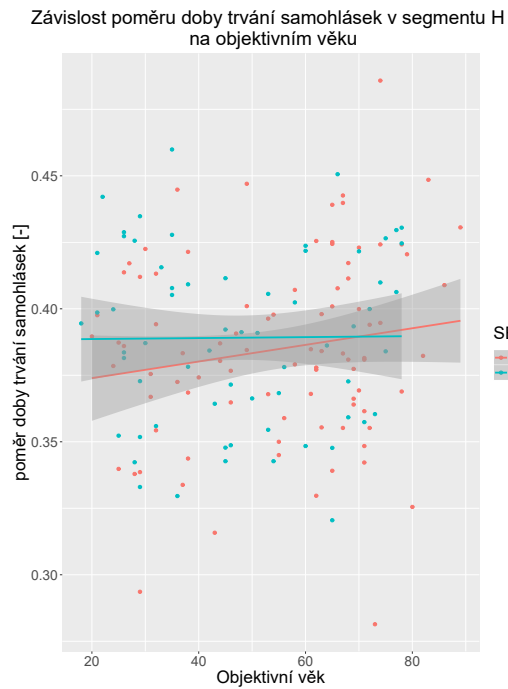


Obrázek 6.3.13. Závislost poměru délky trvání samohlásek v segmentu G [-] na objektivním věku.

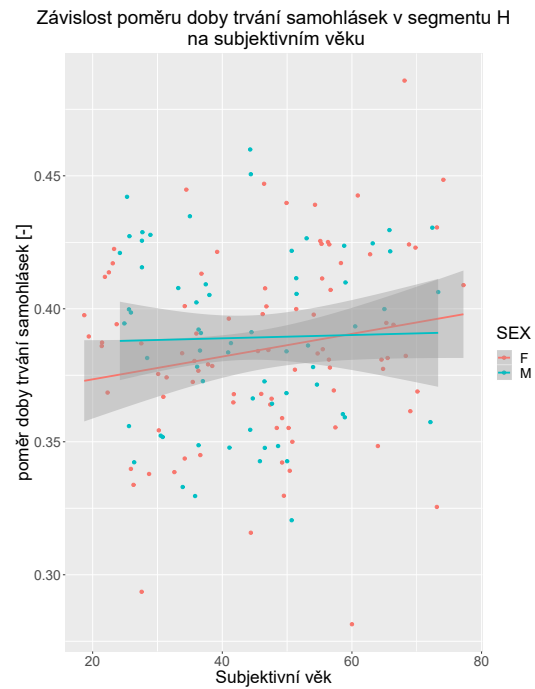
Závislost poměru doby trvání samohlásek v segmentu G na subjektivním věku



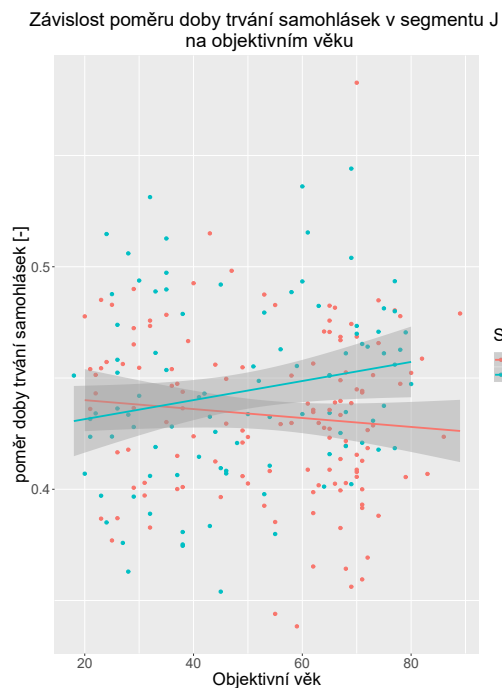
Obrázek 6.3.14. Závislost poměru délky trvání samohlásek v segmentu G [-] na subjektivním věku.



Obrázek 6.3.15. Závislost poměru délky trvání samohlásek v segmentu H [-] na objektivním věku.



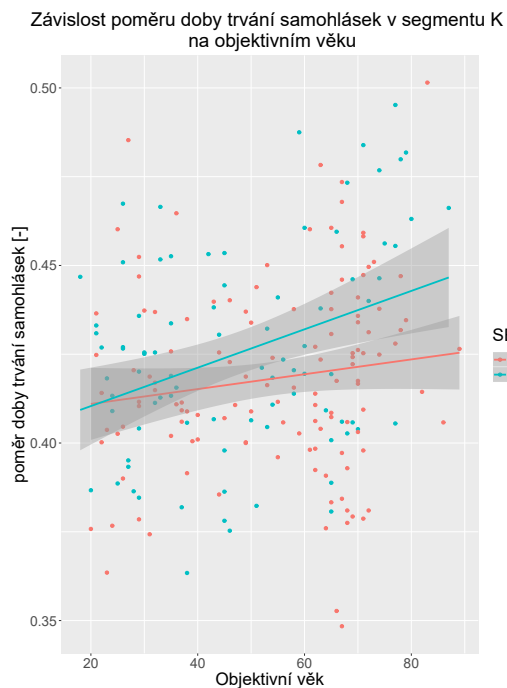
Obrázek 6.3.16. Závislost poměru délky trvání samohlásek v segmentu H [-] na subjektivním věku.



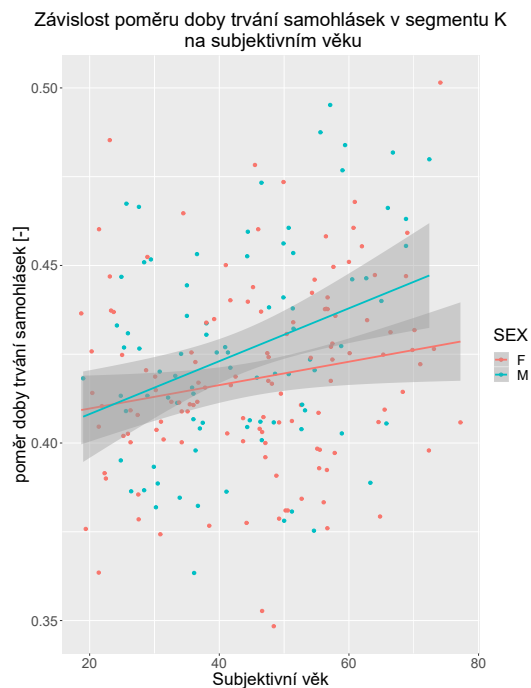
Obrázek 6.3.17. Závislost poměru délky trvání samohlásek v segmentu J [-] na objektivním věku.



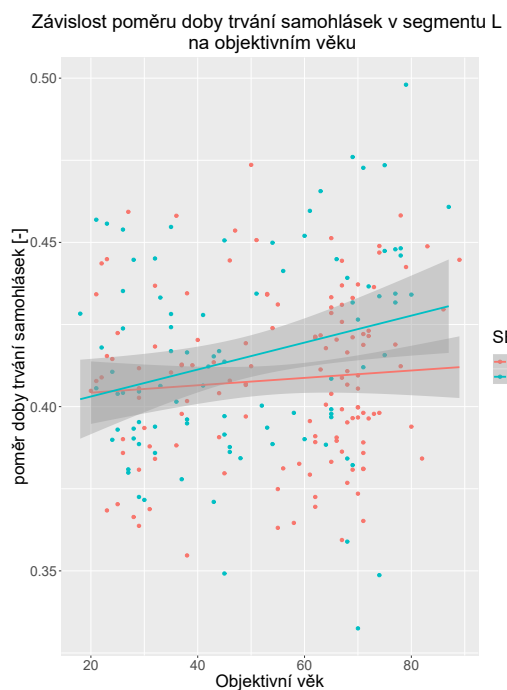
Obrázek 6.3.18. Závislost poměru délky trvání samohlásek v segmentu J [-] na subjektivním věku.



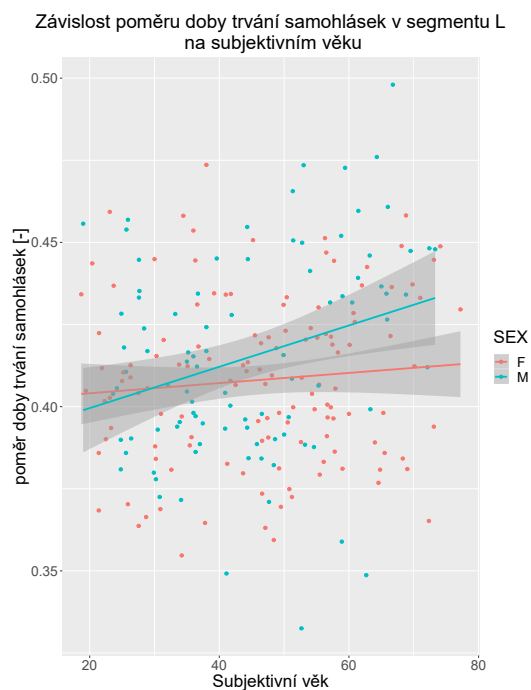
Obrázek 6.3.19. Závislost poměru délky trvání samohlásek v segmentu K [-] na objektivním věku.



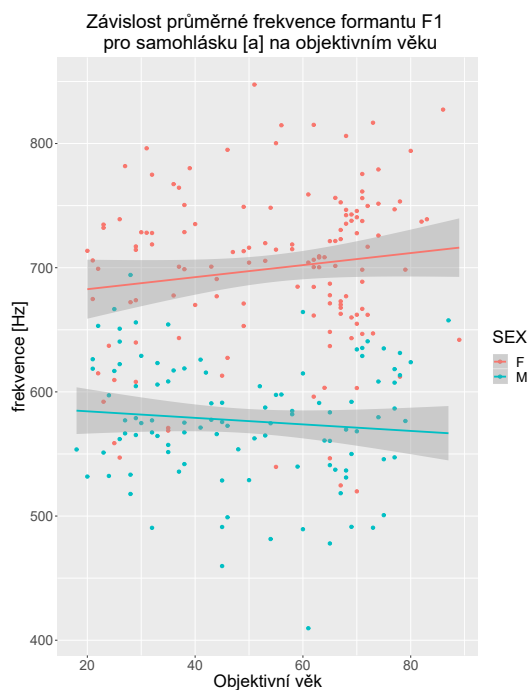
Obrázek 6.3.20. Závislost poměru délky trvání samohlásek v segmentu K [-] na subjektivním věku.



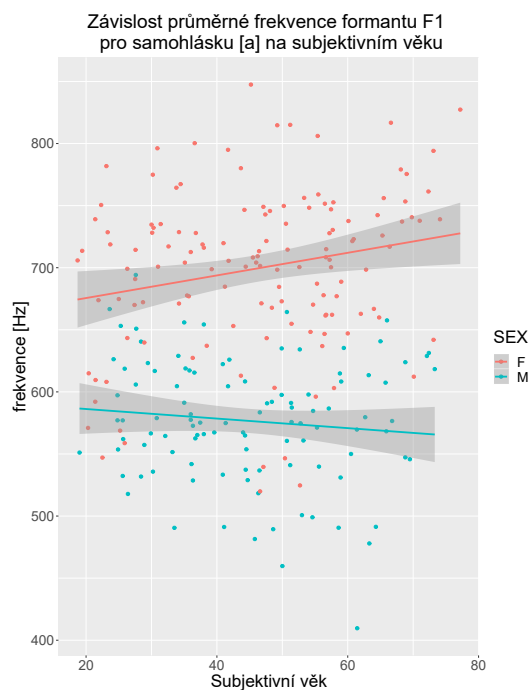
Obrázek 6.3.21. Závislost poměru délky trvání samohlásek v segmentu L [-] na objektivním věku.



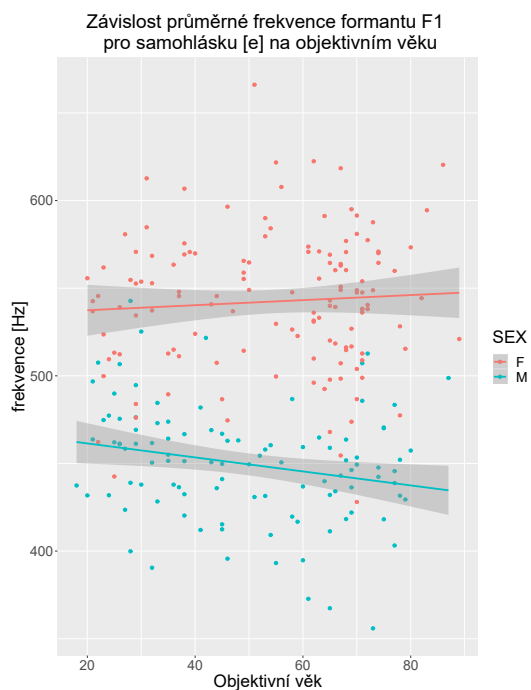
Obrázek 6.3.22. Závislost poměru délky trvání samohlásek v segmentu L [-] na subjektivním věku.



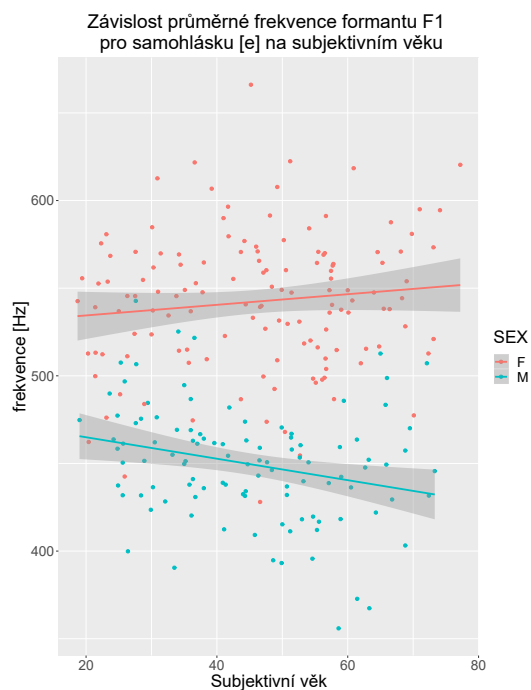
Obrázek 6.3.23. Závislost průměrné frekvence F1 [Hz] pro samohlásku [a] na objektivním věku.



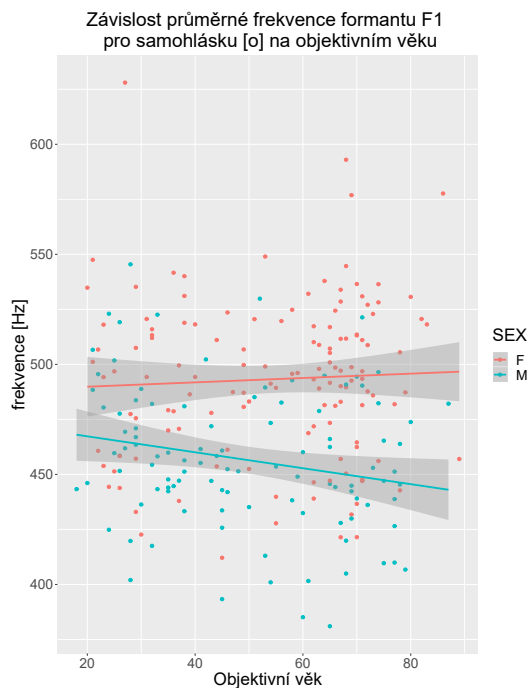
Obrázek 6.3.24. Závislost průměrné frekvence F1 [Hz] pro samohlásku [a] na subjektivním věku.



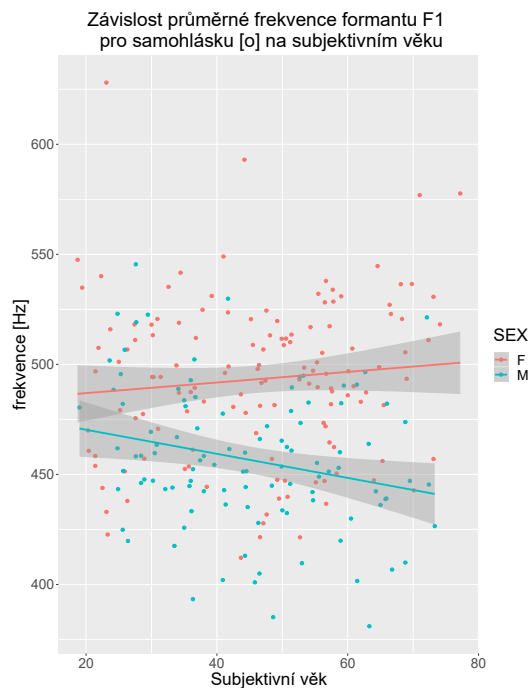
Obrázek 6.3.25. Závislost průměrné frekvence F1 [Hz] pro samohlásku [e] na objektivním věku.



Obrázek 6.3.26. Závislost průměrné frekvence F1 [Hz] pro samohlásku [e] na subjektivním věku.



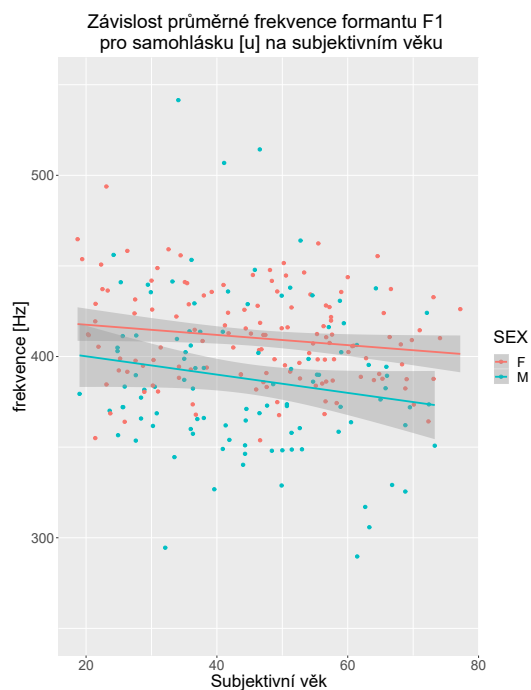
Obrázek 6.3.27. Závislost průměrné frekvence F1 [Hz] pro samohlásku [o] na objektivním věku.



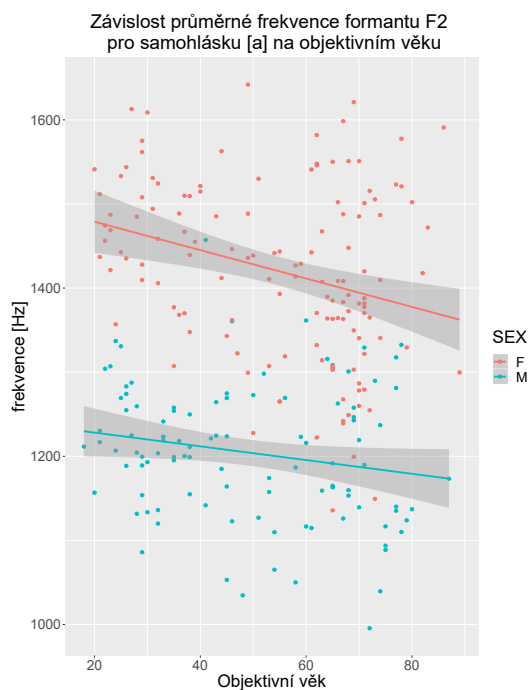
Obrázek 6.3.28. Závislost průměrné frekvence F1 [Hz] pro samohlásku [o] na subjektivním věku.



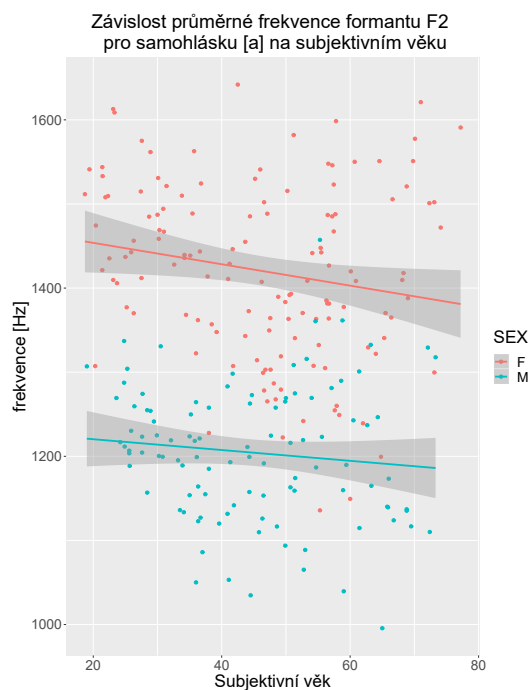
Obrázek 6.3.29. Závislost průměrné frekvence F1 [Hz] pro samohlásku [u] na objektivním věku.



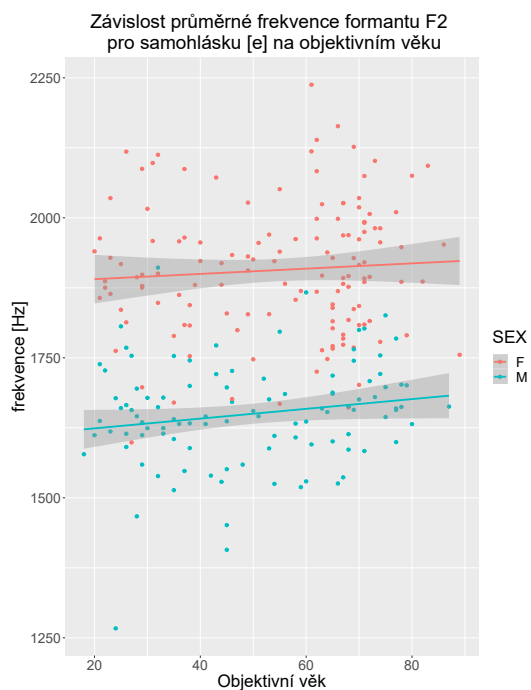
Obrázek 6.3.30. Závislost průměrné frekvence F1 [Hz] pro samohlásku [u] na subjektivním věku.



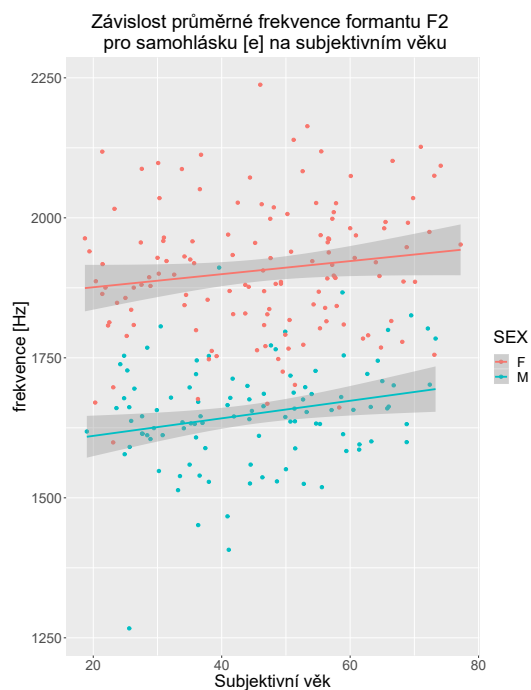
Obrázek 6.3.31. Závislost průměrné frekvence F2 [Hz] pro samohlásku [a] na objektivním věku.



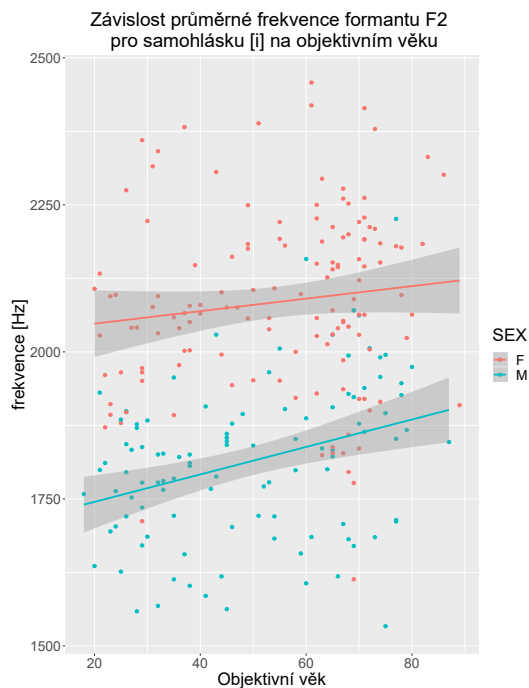
Obrázek 6.3.32. Závislost průměrné frekvence F2 [Hz] pro samohlásku [a] na subjektivním věku.



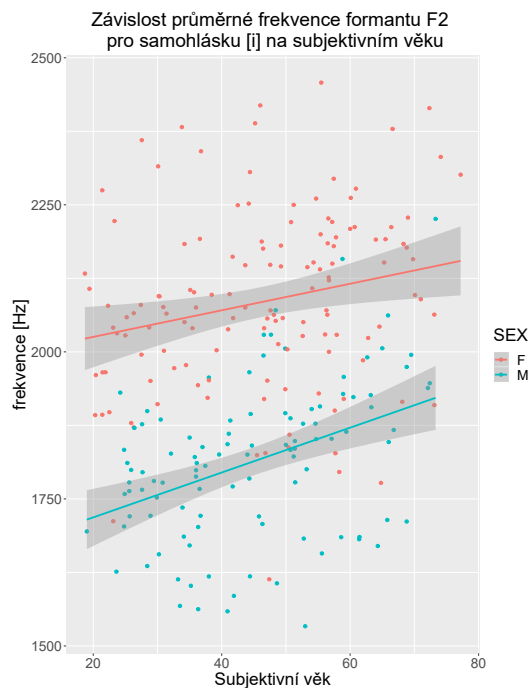
Obrázek 6.3.33. Závislost průměrné frekvence F2 [Hz] pro samohlásku [e] na objektivním věku.



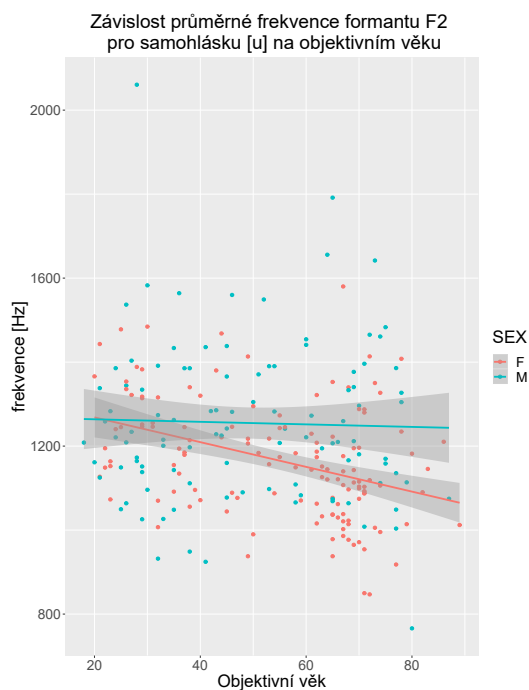
Obrázek 6.3.34. Závislost průměrné frekvence F2 [Hz] pro samohlásku [e] na subjektivním věku.



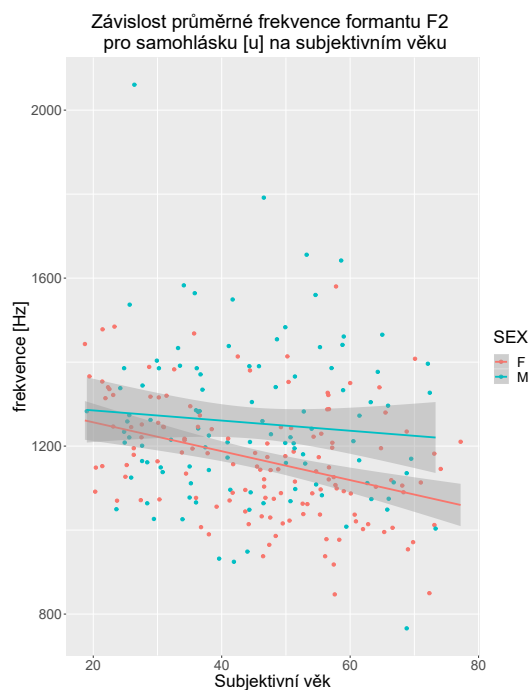
Obrázek 6.3.35. Závislost průměrné frekvence F2 [Hz] pro samohlásku [i] na objektivním věku.



Obrázek 6.3.36. Závislost průměrné frekvence F2 [Hz] pro samohlásku [i] na subjektivním věku.



Obrázek 6.3.37. Závislost průměrné frekvence F2 [Hz] pro samohlásku [u] na objektivním věku.



Obrázek 6.3.38. Závislost průměrné frekvence F2 [Hz] pro samohlásku [u] na subjektivním věku.