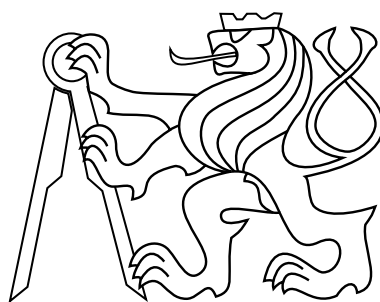


Diplomová práce

Automatické hodnocení poruch artikulace samohlásek u Parkinsonovy nemoci

Bc. Petr Bukovský



Květen, 2017

Vedoucí práce: Ing. Jan Ruzs, Ph.D.

České vysoké učení technické v Praze
Fakulta elektrotechnická, Katedra teorie obvodů

ZADÁNÍ DIPLOMOVÉ PRÁCE

Student: Bc. Petr B u k o v s k ý

Studijní program: Biomedicínské inženýrství a informatika

Obor: Biomedicínské inženýrství

Název tématu: Automatické hodnocení poruch artikulace samohlásek u Parkinsonovy nemoci

Pokyny pro vypracování:

1. Seznamte se s možnostmi potenciálního využití hodnocení poruch artikulace samohlásek u Parkinsonovy nemoci. Na základě literatury nastudujte vhodné parametry pro hodnocení poruch artikulace samohlásek a jejich možnosti automatické detekce. Dále se seznamte s algoritmy pro detekci formantových frekvencí a možnosti využití rozpoznávače řeči pro segmentaci fonémů v řeči.
2. Na základě dostupné literatury navrhnete vhodné metody pro automatickou detekci poruch artikulace samohlásek u Parkinsonovy nemoci. Pro řešení návrhu využijte dostupné detektory formantových frekvencí. Vypočítejte parametry artikulace samohlásek s využitím metody bez nutnosti segmentace i metody zahrnující segmentaci hlásek pomocí rozpoznávače řeči. Pro návrh metod využijte výpočetní prostředí MATLAB a softwarový nástroj Praat.
3. Navržené metody otestujte na vybraném vzorku řečových promluv pacientů s výskytem Parkinsonovy nemoci a zdravých jedinců kontrolní skupiny. Porovnejte výsledky jednotlivých parametrů pro hodnocení poruch artikulace samohlásek s využitím různých formantových detektorů, metod na bázi segmentace i bez jejího využití a zároveň výsledky porovnejte s ruční analýzou. Na dostupných datech proveďte klasifikační experiment a vyhodnoťte nejvhodnější parametry artikulace samohlásek pro odlišení zdravých mluvčích od pacientů s Parkinsonovou nemocí.

Seznam odborné literatury:

- [1] Sapir S, Ramig LO, Spielman JL, Fox C. Formant Centralization Ratio: A proposal for a new acoustic measure of dysarthric speech. J Speech Lang Hear Res 2010;53:114-125.
- [2] Ruzs J, Cmejla R, Tykalova T, Ruzickova H, Klempir J, Majerova V, Picmausova J, Roth J, Ruzicka E. Imprecise vowel articulation as potential early marker of Parkinson's disease: effect of speaking task. J Acoust Soc Am 2013;134:2171-2181.
- [3] Sandoval S, Berisha V, Utianski RL, Liss JM, Spanias A. Automatic assessment of vowel space area. J Acoust Soc Am 2013;134:EL477-EL483.
- [4] Mustafa K, Bruce IC. Robust formant tracking for continuous speech with speaker variability. IEEE T Audio Speech 2006;14:435-444.
- [5] Boersma P, Weenink D. PRAAT, a system for doing phonetics by computer. Glot International 2001;5:341-345.
- [6] Schwarz P. Phoneme Recognition based on Long Temporal Context. PhD Thesis, Brno University of Technology, 2009.

Vedoucí diplomové práce: Ing. Jan Ruzs, Ph.D.

Platnost zadání: do konce letního semestru 2017/2018

L.S.

prof. Ing. Pavel Sovka, CSc.
vedoucí katedry

prof. Ing. Pavel Ripka, CSc.
děkan

V Praze dne 1. 10. 2016

Poděkování

Mé poděkování si zaslouží především Ing. Jan Ruzs, Ph.D. za projevenou trpělivost a odborné vedení této práce. Dále bych chtěl poděkovat všem, kteří mi nějakým způsobem pomohli nejen v době sepisování této práce, ale i v rámci celého mého studia.

Prohlášení

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

V Praze dne 20. 5. 2017

.....
Petr Bukovský

Abstrakt

Parkinsonova nemoc (PN) je obvykle spojena s poruchou řeči tzv. hypokinetickou dysartrií, která se mimo jiné projevuje také nepřesnostmi v artikulaci samohlásek. Cílem práce bylo navrhnout automatickou metodu pro hodnocení poruch artikulace samohlásek u PN. Pro detekci poruch artikulace se vycházelo z hodnot první a druhé formantové frekvence a od nich odvozených parametrů VSA (obsah vokální plochy), VAI (artikulační index) a F_{2i}/F_{2u} (poměr druhých formantů samohlásek i a u). Pro návrh automatické metody bylo využito dvou přístupů: s rozpoznávačem řeči pro segmentaci fonémů a bez rozpoznávače řeči. Využití rozpoznávače řeči umožnilo zpracovávat signál pouze v úsecích řeči, v nichž byla vyslovena samohláska. Testování proběhlo na databázi čítající 72 subjektů, která obsahovala data od 36 zdravých lidí a 36 pacientů s PN. U všech lidí byly pořízeny řečové záznamy čteného textu a spontánního monologu, na kterých proběhla ruční analýza formantových frekvencí. Ze statistické analýzy ruční detekce byly vyhodnoceny jako signifikantní parametry pro odlišení zdravých a nemocných zejména F_{2u} , VAI a F_{2i}/F_{2u} . U výsledků automatické metody bez rozpoznávače řeči nebyly shledány žádné signifikantní parametry. Parametry, které vzešly z výsledků automatické metody využívajícího rozpoznávač řeči, korespondují s výsledky ruční analýzy co do výčtu signifikantních parametrů i jejich statistické významnosti. Na výsledcích automatické metody s rozpoznávačem řeči proběhla klasifikace, jejíž celková přesnost pro odlišení zdravých jedinců a pacientů s PN dosáhla přes 70 %. Na základě výsledků automatické metody je patrné, že celý proces hodnocení artikulace samohlásek může být zautomatizován. Automatická metoda by mohla najít své potencionální využití v objektivním hodnocení úspěšnosti léčby, progresu nemoci, pomoci při diagnóze PN či u jiných neurologických nemocí, které souvisí s poruchou artikulace samohlásek.

Klíčová slova

Parkinsonova nemoc, porucha řeči, artikulace samohlásek, formantová frekvence, rozpoznávač řeči, klasifikace

Abstract

The aim of this work was to propose an automatic method for assessment of imprecise vowel articulation in Parkinson's disease (PD). For the detection of articulation disorders the first and second formant frequencies and derived parameters VSA (vowel space area), VAI (vowel articulation index) and F_{2i}/F_{2u} (the ratio of the second formants of the vowels i and u) were used. The approach based on speech recognizer using phoneme segmentation as well as approach without recognizer were used to design automated method for vowel articulation quality extraction. Speech recognizer allowed the processing of signal in vowel speech parts only. Testing of two approaches was performed on a database of 72 subjects including data from 36 healthy people and 36 PD patients. Speech recordings of reading passage and spontaneous monologue were analyzed using manual labels with respect to formant frequencies. Statistical analysis of manual detection showed that parameters such as F_{2u} , VAI and F_{2i}/F_{2u} were able to distinguish between healthy people and PD patients. No statistically significant parameters were found in the results based on the automatic method without recognizer. Parameters resulting from the automatic method with recognizer correlated with the results of the manual analysis. The results of the automatic method with recognizer were subsequently used for a classification with overall reached accuracy of 70 % in separating between PD and healthy groups. Based on the results of the automatic method with recognizer we proved that the entire vowel articulation evaluation process can be automated. The features automatic method can provide biomarkers for early disease detection, monitoring of treatment efficacy or disease progression in PD as well as in other diseases manifesting imprecise vowel articulation.

Keywords

Parkinson's disease, speech disorder, vowel articulation, formant frequency, speech recognizer, classification

Seznam zkratek

ACC	Přesnost (z anglického accuracy)
DDK	Diadochokinetický úkol
DP	Filtr typu dolní propust
F_0	Základní formantová frekvence
F_1	První formantová frekvence
F_2	Druhá formantová frekvence
F_3	Třetí formantová frekvence
FN	Falešně negativní (z anglického false negative)
FP	Falešně pozitivní (z anglického false positive)
GMM	Směs Gausovských modelů (z anglického Gaussian mixture model)
HP	Filtr typu horní propust
HY	Hoehn-Yahr
KS	Kontrolní skupina
LPC	Predikční koeficienty (z anglického linear predictive control)
RACF	Opakovaná autokorelační funkce
SE	Senzitivita
SP	Specificita
TN	Skutečně negativní (z anglického true negative)
TP	Skutečně pozitivní (z anglického true positive)
VAI	Artikulační index (z anglického vowel articulation index)
VOT	Počátek znělosti (z anglického voice onset time)
VSA	Obsah vokální plochy (z anglického vowel space area)

Obsah

1. Úvod	1
1.1. Řeč	1
1.1.1. Proces tvorby řeči	1
1.1.2. Fonémy	2
Samohlásky	2
Souhlásky	4
1.2. Parkinsonova nemoc	5
1.2.1. Parametry pro popis PN	6
1.2.2. Poruchy řeči u PN	7
Respirace	7
Fonace	7
Artikulace	7
Prozódie	8
Vliv léčby na řeč	8
1.3. Porucha artikulace samohlásek u PN	8
1.4. Ruční analýza formantových frekvencí	10
1.5. Metody pro automatický výpočet formantových frekvencí	11
1.5.1. Praat	11
1.5.2. Ghosh	12
1.5.3. Mustafa & Bruce	12
1.6. Metoda pro automatický výpočet VSA	14
1.7. Rozpoznávač řeči pro segmentaci fonémů	15
1.8. Cíle práce	16
2. Metody	17
2.1. Databáze	17
2.2. Nahrávání řeči	18
2.3. Ruční analýza formantových frekvencí	18
2.4. Algoritmy pro výpočet formantových frekvencí	18
2.5. Algoritmy pro automatickou detekci artikulace samohlásek	19
2.5.1. Metoda bez rozpoznávače řeči	20
Výpočet první a druhé formantové frekvence	20
Odlehlé hodnoty	21
Shlukování	21
2.5.2. Metoda s rozpoznávačem řeči	23
Výpočet vektorů formantových frekvencí pro vybrané samohlásky	23
Úprava formantových vektorů vybraných samohlásek	23
Odlehlé hodnoty ve formantových párech stejných samohlásek	25
Shlukování	26
2.6. Metody pro vyhodnocení výsledků	27
2.6.1. Test spolehlivosti	27
2.6.2. Test rozpoznávače řeči	27
2.6.3. Test středních hodnot	28
2.6.4. Test minimální délky řečového signálu	29
2.6.5. Klasifikační metody	29

3. Výsledky	31
3.1. Ruční analýza	31
3.1.1. Intra-judge a inter-judge reliabilita	31
3.1.2. Porovnání parametrů mužů a žen	31
3.1.3. Porovnání parametrů mezi skupinami KS a PN	32
3.1.4. Porovnání formantů dlouhých a krátkých samohlásek	34
3.2. Automatické metody	35
3.2.1. Test spolehlivosti algoritmů pro výpočet formantů	35
3.2.2. Test rozpoznávače řeči	36
3.2.3. Automatická metoda bez rozpoznávače řeči	38
3.2.4. Automatická metoda s rozpoznávačem řeči	38
3.2.5. Porovnání ruční analýzy s automatickou metodou	41
3.2.6. Opakovatelnost automatické metody	43
3.2.7. Minimální délka řečového signálu	44
3.3. Klasifikace	44
4. Diskuze	47
5. Závěr	51
Literatura	53
Přílohy	
A. Text	57
B. Obsah přiloženého CD	59

1. Úvod

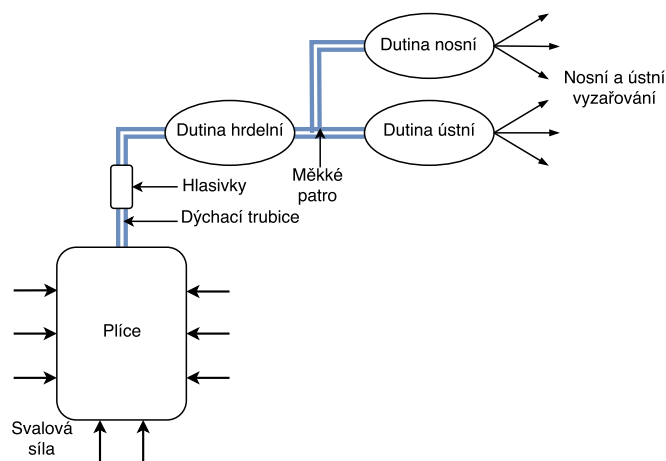
1.1. Řeč

Řeč je přirozenou formou komunikace mezi lidmi. Řeč však nemusí být brána jenom jako prostředek k dorozumívání, ale má své využití i ve vědě a technice. Analýza řečového signálu se mimo jiné používá k detekci poruch řeči, které mohou být spojeny s různými onemocněními. Na základě akustické analýzy lze poskytnout kvantitativní, objektivní a přesné prostředky pro popis přítomnosti, závažnosti a charakteristik řečových poruch a ke sledování zlepšování či zhoršování řeči. [1] Tato práce se zabývá poruchou řeči u lidí postižených Parkinsonovou chorobou.

1.1.1. Proces tvorby řeči

Schématické znázornění hlasového ústrojí člověka je uvedeno na obr. 1. Proud vzduchu vycházející z plic prochází tzv. hlasivkovou štěrbinou, která je obklopena hlasivkami. Hlasivky se díky procházejícímu vzduchu rozkmitají, čímž otevírají a zavírají hlasivkovou štěrbinu. Tím vzniká budící periodický signál, který je charakteristický pro znělé hlásky, o frekvenci F_0 odpovídající výšce hlasu. [2] Frekvence F_0 souvisí mimo jiné s délkou hlasivek. Délka hlasivek u mužů se pohybuje v rozmezí 17 – 25 mm a u žen v rozmezí 12,5 – 17,5 mm. Ženy mají zpravidla vyšší hodnotu F_0 než muži. [3] Při produkci neznělých hlásek nedochází k rozkmitání hlasivek a je generován pouze šumový signál.

Budící signál prochází do hlasového traktu, jenž je tvořen dutinou hrdelní, dutinou nosní a dutinou ústní. V hlasovém traktu dochází pomocí artikulace k modifikaci budícího signálu. Artikulací se rozumí pohyb řečových orgánů (jazyka, rtů a čelistí), díky čemuž je signál namodulován na správný tvar a amplitudu. Z ústní dutiny pak vychází srozumitelný řečový signál. [2]



Obrázek 1. Zjednodušené schématické znázornění hlasového ústrojí.

1. Úvod

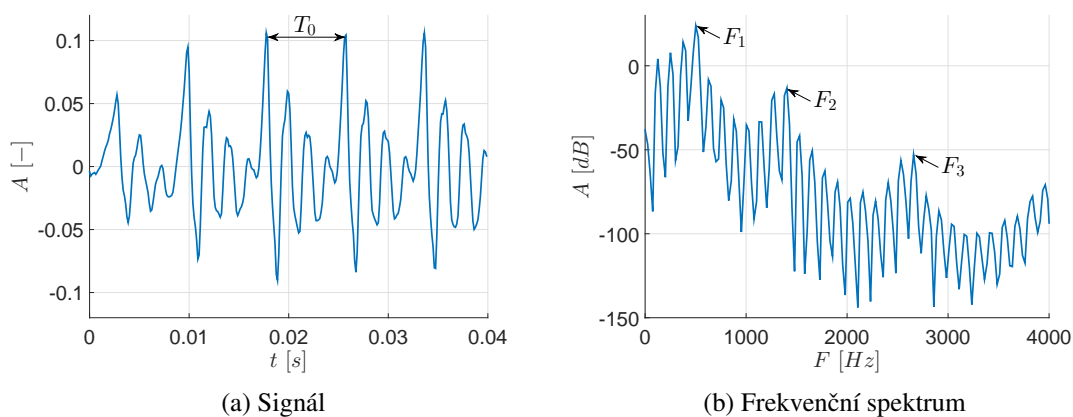
1.1.2. Fonémy

Foném je nejmenší významotvornou jednotkou řeči. V českém jazyce se nachází celkem 39 fonémů, které se dělí na samohláskové (vokalizké) a souhláskové (konsonantické).

Samohlásky

Čeština obsahuje celkem 13 samohláskových fonémů - 5 krátkých (a, e, i, o, u), 5 dlouhých ($a:, e:, i:, o:, u:$) a tři difony (au, eu, ou). Dlouhé a krátké samohlásky tvoří páry. Vyslovení dlouhé samohlásky trvá zhruba dvakrát déle než vyslovení krátké. [4]

Všechny samohlásky jsou znělé. Průběh signálu u samohlásek je periodický s periodou $T_0 = \frac{1}{F_0}$. Na obr. 2 (a) je znázorněn signál samohláskového fonému a , ve kterém je zřetelná perioda T_0 . Ve frekvenčním spektru samohlásek jsou patrné rezonanční frekvence, tzv. formanty (viz obr. 2 (b)). Formanty jsou tvořeny rezonancí v dutinách hlasového ústrojí. První formantová frekvence (F_1) má ve spektru nejvyšší energii. Vzniká rezonancí v dutině hrdelní. Druhá formantová frekvence (F_2) je tvořena dutinou ústní a třetí formantová frekvence (F_3) se vytváří v dutině nosní. [2]



Obrázek 2. Signál a frekvenční spektrum samohlásky a .

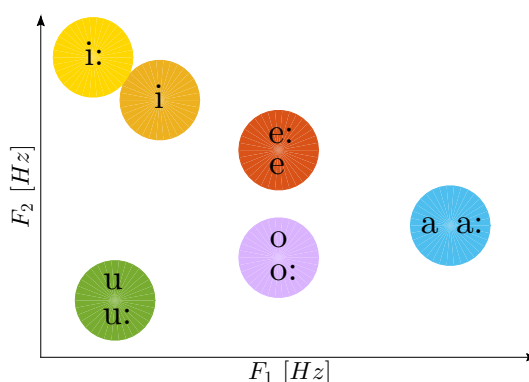
Formantové frekvence F_1 a F_2 souvisí s umístěním a pohybem jazyka. Obecným pravidlem akusticko-artikulačních vztahů je, že se frekvence F_1 mění nepřímo úměrně s výškou nadzvednutí jazyka a frekvence F_2 se zvyšuje s posunem jazyka směrem vpřed. Dále se obě formantové frekvence snižují se zaokrouhleností rtů. [5] Poloha jazyka a postavení rtů českých samohlásek je zahrnuto v tab. 1. Samohlásky jsou v závislosti na postavení jazyka popisovány umístěním ve vokalizkém prostoru, který se dá pro české samohlásky aproximovat trojúhelníkem. Obr. 3 vyjadřuje postavení samohlásek v prostoru první a druhé formantové frekvence. Samohlásky a , i a u jsou vzhledem ke svému postavení ve vokalizkém prostoru nazývány rohovými samohláskami. [4]

Při vyslovování samohlásek u a $u:$ je jazyk nadzvednutý a umístěn vzadu, podle čehož jsou tyto samohlásky nazývány zadní vysoké. Samohlásky u a $u:$ jsou artikulovány se zaokrouhlenými rty, což společně s polohou jazyka přispívá k nízkým formantovým frekvencím F_1 a F_2 . U středních nízkých samohlásek a a $a:$ jsou rty otevřené a nevykonávají jinak žádnou aktivní činnost. Díky nízkému nadzvednutí jazyka mají tyto samohlásky vysokou první formantovou frekvenci. U artikulace předních vysokých samohlásek i a $i:$ vykonávají rty aktivní činnost tak, že jsou koutky úst roztaženy od sebe. Tato poloha rtů je označována jako zaostřená a je charakteristická i pro samohlásky e a $e:$. Samohláskový pár i a $i:$ se kromě délkou trvání liší postavením artikulačních orgánů. U vyslovování dlouhé samohlásky $i:$ má jazyk tendenci se

více zvedat a být více vpředu. To má za následek odlišné postavení ve vokalickém prostoru. U ostatních samohláskových párů nastávají podobné jevy akorát v menší míře, proto je postavení těchto samohláskových párů v rámci vokalického prostoru podobné. [4]

Samohláska	Poloha jazyka		Postavení rtů
	Nadzvednutí	Posunutí	
a	Nízké	Střední	Neutrální
e	Středové	Přední	Zaostřené
i	Vysoké	Přední	Zaostřené
o	Středové	Zadní	Zaokrouhlené
u	Vysoké	Zadní	Zaokrouhlené

Tabulka 1. Postavení artikulačních orgánů při vyslovování samohlásek.

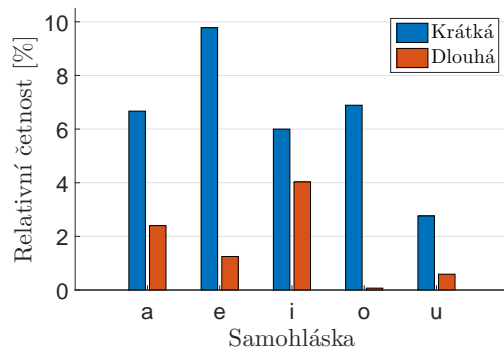


Obrázek 3. Pozice samohlásek ve vokalickém prostoru.

Hodnoty formantových frekvencí jsou ovlivněné pohlavím a věkem. První a druhá formantová frekvence dosahuje vyšších hodnot u žen než u mužů (viz např. [6]). Studie [7] zabývající se porovnáním prvních dvou formantů mladších a starších mluvčích ukázala, že hodnoty formantů všech samohlásek jsou u starších mužů nižší než u mladších. U starších žen byly pozorovány nižší formantové frekvence pouze u fonémů *a* a *u*. Naopak formanty fonémů *e*, *i* a *o* dosahovaly u starších žen vyšších hodnot oproti mladším ženám.

Obr. 4 znázorňuje procentuální zastoupení samohláskových fonémů vůči všem fonémům v českém jazyce. V češtině je mnohem větší zastoupení krátkých samohláskových fonémů než dlouhých. Nejčastější je foném *e*, jehož zastoupení činí téměř 10 % všech českých fonémů. Následují fonémy *a*, *i* a *o*, jejichž relativní četnost se pohybuje mezi 6 a 7 procenty. Poměrně velké zastoupení má dlouhý foném *i:*, kterého se v češtině nachází 4 % a navíc má větší relativní četnost než krátký foném *u*. Ostatní dlouhé fonémy *a:*, *e:*, *o:* a *u:* mají relativní zastoupení do 2,5 %. [8] Vůbec nejmenší zastoupení ze samohláskových fonémů má foném *o:*, který se vyskytuje převážně ve slovech převzatých z cizích jazyků. [4]

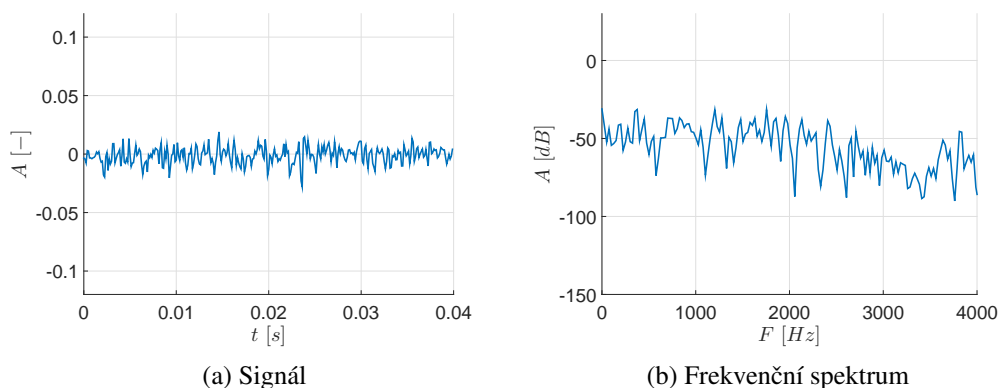
1. Úvod



Obrázek 4. Relativní četnost samohláskových fonémů vůči všem fonémům v českém jazyce.

Souhlásky

Souhlásky mohou být na rozdíl od samohlásek znělé i neznělé a jsou artikulovány pomocí vytvoření překážek proudu vzduchu jdoucímu z plic. Příklad signálu neznělé souhlásky znázorňuje obr. 5 (a). Obrázek zobrazuje průběh souhlásky *s*. Na rozdíl od samohláskových fonémů nejsou u neznělých souhlásek patrné rezonanční frekvence ve spektru signálu (viz obr. 5 (b)). Rozdělení souhlásek na znělé a neznělé uvádí tab. 2. V tabulce není zahrnuta souhláska *ř*, která se může chovat jako znělá i jako neznělá. Dále lze do skupiny znělých fonémů zařadit fonémy převzaté z cizích jazyků *dz* a *dž*. [4]



Obrázek 5. Signál a frekvenční spektrum neznělé souhlásky *s*.

Skupina	Souhlásky
Znělé	b, v, d, d', z, ž, g, h, m, n, ň, j, l, r
Neznělé	p, f, t, t', s, š, k, ch, c, č

Tabulka 2. Znělé a neznělé souhlásky.

Souhlásky se dají kromě kritéria, zda jsou znělé či neznělé, rozdělit také podle typu překážky a místa vokálního traktu, kde byla překážka vytvořena. Rozdělení souhlásek podle typu překážky a místa, kde překážka vzniká, udává tab. 3. *Explozivny* jsou artikulovány pomocí přetlaku vzduchu, který vzniká díky přepážce zabráňující proudu vzduchu vyjít z úst. Při uvolnění překážky vzniká charakteristický zvukový efekt tzv. exploze. U vyslovování *frikativ* dochází v místě vytvoření překážky ke zúžení vokálního traktu. Při procházení proudu vzduchu vzni-

kají v místech zúžení turbulence, které jsou v dalších částech vokálního traktu filtrovány. *Afrikáty* jsou tvorbou své výslovnosti mezi explozívy a frikativy. Při artikulaci afrikátů je v první fázi utvořen závěr (explozívy), který následně přechází ve zúžení (frikativy). *Vibranty* se při artikulaci vyznačují charakteristickými zákmity špičky jazyka. *Nazály* jsou podobně jako explozívy tvořeny závěrem. Na rozdíl od explozív však u nazálů nevzniká přetlak vzduchu před překážkou, protože vzduch uniká nosní dutinou. U *aproximantů* dochází stejně jako u frikativ k vytvoření úžiny. Tato úžina je však o něco širší, než je tomu u frikativ a nevzniká zde turbulentní proudění vzduchu. Překážky jsou tvořeny nejčastěji pomocí jazyka a rtů, které tvoří překážku společně se zuby, dásněmi či tvrdým nebo měkkým patrem. [4]

	B	Ld	Al	Po	Pa	V	La
E	p,b		t,d		t',d'	k,g	
F		f,v	s,z	š,ž		ch	h
Af			c,dz	č,dž			
V		r	ř				
N	m	m	n		ň	n	
Ap			l		j		

Tabulka 3. Rozdělení souhlásek podle způsobu a místa vytvoření překážky. *Místo vytvoření překážky:* E = explozívy (hlásky závěrové ústní), F = frikativy (hlásky třené), Af = afrikáty (hlásky polotřené), V = vibranty (hlásky kmitavé), N = nazály (nosovky), Ap = aproximanty. *Způsob vytvoření překážky:* B = bilabiální (obouretné), Ld = labiodentální (retozubé), Al = alveolární (dásňové), Po = postalveolární (dásňové zadní), Pa = palatální (tvrdopatrové), V = velární (měkkopatrové), La = laryngální (hrtanové).

1.2. Parkinsonova nemoc

Parkinsonova nemoc (PN) je po Alzheimerově chorobě druhým nejčastějším neurodegenerativním onemocněním. [9] PN vzniká v důsledku odumírání dopaminergních neuronů lokalizovaných zejména v buňkách černého jádra (substantia nigra), které se nachází ve středním mozku. [10, 11, 12] Nadměrným odumíráním těchto neuronů vzniká nedostatek dopaminu, což vede k nesprávné regulaci funkce bazálních ganglií a tím dochází k projevům onemocnění. [12] Prvotní projevy PN bývají obvykle necharakteristické. Dochází ke ztrátě výkonnosti, bolesti zad a ramen, poruchám spánku, stavům deprese atd. Jedním z projevů je i změna hlasu projevující se zejména tichostí a monotónností. [13] Teprve později se začínají objevovat typické příznaky nemoci: klidový třes (tremor), svalová ztuhlost (rigidita), snížený rozsah pohybů (hypokineze), pohybové zpomalení (bradykineze), porucha startu pohybu (akineze) a posturální nestabilita. [10, 14]

Třes je nejtypičtější příznakem PN. Objevuje se především na končetinách, když je postižený PN v klidu. Při cílevědomé činnosti třes ustupuje. Svalová ztuhlost omezuje jedince trpící PN v pohybu. Projevuje se vyšším svalovým napětím, které je potřeba k vykonání pohybu. Třes a svalová ztuhlost se na začátku PN projevují unilaterálně. S postupem nemoci se však oba příznaky stávají symetričtějšími a postihují tak levou i pravou stranu. Akineze je neschopnost zahájení pohybu (například zahájení chůze). Bradykineze se projevuje velmi pomalými a chudými pohyby. Hypokineze, bradykineze ani akineze není závislá na míře svalové ztuhlosti. Posturální nestabilita souvisí s poruchou stoje a chůze. Postižení PN obvykle mají sehnutý trup a pokrčená ramena. S postupem nemoci se pacienti pohybují drobnějšími kroky a obvyklé jsou i ztráty rovnováhy. Výskyt těchto příznaků a jejich vývoj během nemoci je u pacientů postižených PN individuální. [10, 13, 14]

1. Úvod

Nejběžnější a nejefektivnější léčba PN probíhá pomocí léku levodopa (L-DOPA). L-DOPA působí v mozku, kde z něho vzniká dopamin. I když nelze PN vyléčit, díky dopaminergní léčbě lze potlačovat typické projevy nemoci. Léčba dlouhodobě eliminuje zejména bradykinezi, hypokynezi a rigiditu. Klidový třes má u každého pacienta jinou odezvu na léčbu. Ostatní příznaky reagují na dopaminergní léčbu obvykle hůře. [13, 15]

1.2.1. Parametry pro popis PN

Míra postižení motorických funkcí se u PN obvykle popisuje pomocí dvou parametrů: *Hoehn-Yahr* (HY) a *UPDRS III*. HY je široce používaná stupnice klinického hodnocení, která definuje kategorie motorických funkcí u PN. Stupnice zachycuje typické vzorce pro popis poškození motoriky, které lze aplikovat bez ohledu na léčbu pacienta. Bylo zjištěno, že hodnota HY koreluje s poklesem motorických funkcí a zhoršením kvality života. Vzhledem k jednoduchosti nebere stupnice HY v potaz všechny příznaky, které jsou spojeny s PN. Často se tedy používá modifikovaná stupnice HY, která uvažuje i některé další mezistupně spojené s PN. Modifikovaná HY stupnice je uvedena v tab. 4. Nejnižší stupeň je stupeň 1, u kterého se příznaky onemocnění projevují pouze unilaterálně. U stupně 1,5 se objevují i axiální příznaky a začínají se projevovat i poruchy řeči a poruchy držení těla. U stupně 2 se začínají objevovat bilaterální příznaky. U stupňů 2,5 a 3 pacienti trpí pokročilejšími bilaterálními příznaky a ztrátou rovnováhy. Stupně 4 a 5 již představují závažná postižení. Stupeň 5 značí, že pacient není schopen samostatné chůze a je upoután na vozík. [14, 16]

Stupeň	Popis postižení
1	Pouze jednostranné příznaky
1,5	Jednostranné a axiální příznaky
2	Mírně oboustranné příznaky, bez narušení rovnováhy
2,5	Mírně pokročilé oboustranné příznaky, možná posturální nestabilita
3	Pokročilé oboustranné příznaky, posturální nestabilita
4	Těžké postižení, schopnost samostatného stoje a chůze
5	Upoutání na lůžko nebo vozík

Tabulka 4. Modifikovaná stupnice Hoehn-Yahr.

Hodnota *UPDRS III* (z anglického *unified Parkinson's disease rating scale*) je dána součtem dílčích hodnot charakterizující různá motorická omezení. V hodnocení se popisuje míra dysfunkce různých částí těla na škále 0 až 4, kde 0 znamená, že je pacient bez postižení a 4 označuje nejvyšší stupeň postižení. Hodnotí se rigidita, bradykineze a tremor vyskytující se na různých částech těla. Dále se hodnotí příznaky spojené s posturální nestabilitou a axiální příznaky. Do axiálních příznaků patří mimo jiné i porucha řeči. Hodnocení poruchy řeči je zahrnuto v tab. 5. Celkové rozmezí *UPDRS III* je 0 – 108. [17]

Stupeň	Popis postižení
0	Normální řeč
1	Mírná ztráta vyjadřovací schopnosti a/nebo hlasitosti
2	Monotónní, nezřetelná, ale srozumitelná řeč
3	Značně nezřetelná řeč, těžká k porozumění
4	Nesrozumitelná řeč

Tabulka 5. *UPDRS III* - hodnocení míry postižení řečové funkce.

1.2.2. Poruchy řeči u PN

Řečová dysfunkce způsobená PN je označována jako *hypokinetická dysartrie*. Hypokinetickou dysartrií trpí přibližně 70 až 90 % lidí s PN. [18] PN ovlivňuje různé aspekty řeči, které přispívají ke snížení srozumitelnosti. Jedná se o respiraci v průběhu mluvy, fonaci, artikulaci a prozódii. [19, 20, 21, 22]

Respirace

Svalová ztuhlost spojená s PN může mít za následek ovlivnění dýchacího procesu, který slouží k vytváření proudu vzduchu, jenž je pro řeč nezbytný (viz kap. 1.1.1). Deficity v dýchání mohou ovlivňovat schopnost normálního frázování a hlasitost řeči. Kromě toho může nižší respirační tlak způsobit nedostatky i ve fonaci a artikulaci. [23]

Ve studii [24] bylo zjištěno, že respirační změny spojené s PN mají vliv na řeč při čtení i při vedeném monologu. Pacienti s PN vyprodukují na jeden nádech méně slabik a délka promluvy je na jeden nádech kratší, než je tomu u zdravých mluvčích. Lidé trpící PN mají také tendenci zkracovat délku samohlásek. Průměrná délka samohlásek oproti zdravým jedincům může být až o 50 % kratší. Doba trvání samohlásek se snižuje s postupem PN. [21, 22, 25]

V závislosti na poruchách respirace u PN byla dále zkoumána hlasitost řeči a schopnost ji měnit. Pacienti s PN mají celkově nižší úroveň hlasitosti [26, 27], mají deficity v udržení hlasové úrovně a hůře upravují intenzitu řeči v reakci na vnější podněty. [28, 29] Dále byl zjištěn nižší rozsah hlasitosti řeči u pacientů s PN oproti zdravým jedincům, což poukazuje na nižší schopnost podpory dechu a celkového ovládní řeči u PN. [22]

Fonace

Řada průzkumů, která se zabývala poruchami fonace, zjistila vyšší rozsah a variabilitu základní frekvence F_0 u pacientů s PN oproti zdravým jedincům. [30, 31, 32] Vyšší rozsah a variace F_0 u PN je přisuzováno zhoršené schopnosti udržet hrtanové svaly v pevné konstantní poloze pro prodloužené samohlásky.

U další studie se zkoumal počátek znělosti (VOT z anglického *voice onset time*). VOT je definováno jako trvání doby od artikulační vydání souhlásky k nástupu vyjádření následující samohlásky. [33] U pacientů s PN byla zjištěna vyšší průměrná hodnota VOT oproti zdravým mluvčím. Tento deficit je přisuzován horší koordinaci svaloviny hrtanu. [34]

Pacienti s PN mají deficity v koordinaci fonace a artikulace. Jak ukazují studie [12, 35], u některých pacientů nastává problém při přecházení ze znělých na neznělé fonémy. Zjistilo se, že při přechodu ze znělého fonému je následující neznělý foném vyjadřován z části jako znělý, což má za následek zpoždění ve vyjádření neznělého fonému.

Artikulace

V této sekci jsou popsány pouze některé poruchy artikulace. Poruchy artikulace samohlásek jsou podrobněji popsány v další kapitole (kap. 1.3).

Studie [35, 36, 37], které se zabývaly artikulací souhlásek, prokázaly, že pacienti s PN mají zhoršenou artikulaci exploziv. Explozivy byly artikulovány nepřesně a vyslovovány jako frikativy. Pacienti s PN nedokázali vytvořit úplný závěr, který by způsobil přetlak nutný k artikulaci exploziv. Místo toho docházelo k pouhému zúžení vokálního traktu a tím i k nepřesné artikulaci. Tento deficit může být způsoben například nedostatečným nadzvedáním jazyka.

Jednou z metod pro hodnocení artikulační schopnosti je diadochokinetický (DDK) úkol sloužící ke zkoumání rychlosti střídání artikulačních pohybů. Typický DDK úkol zahrnuje vyslovení několika sekvencí po sobě jdoucích slabik. Slabiky, které jsou tvořeny souhláskou a sa-

1. Úvod

mohláskou, jsou vybírány tak, aby byly využity různé artikulační pohyby - bilabiální (obouretný, např. „papapa...“), alveolární (dásňový, např. „tatata...“) a velární (měkkopatrový, např. „kakaka...“). [38] Výsledky ukázaly, že pacienti s PN mají při stejné rychlosti vyslovování slabik menší amplitudu pohybu rtů než zdraví jedinci. Někteří pacienti v brzkém stádiu PN úspěšně potlačovali bradykinezi a svými výsledky se pak dostávali blízko výsledkům zdravých jedinců a to jak v rychlosti řeči, tak i v amplitudě pohybu rtů. Naopak pacienti v pozdějším stádiu PN při snaze dosáhnout stejné amplitudy pohybu rtů vyslovovali slabiky abnormálně pomalu. [39]

Prozódie

Prozódie je termín pro označení přirozených změn hlasitosti, výšky hlasu a rytmu během mluvy. Řada studií prokázala zhoršenou prozódii u pacientů trpících PN. Studie [40] se zabývala měřením základní frekvence F_0 na poslední slabice tázacích vět. Bylo zjištěno, že pacienti s PN mají na poslední slabice tázací věty nižší F_0 než zdraví jedinci. Frekvence F_0 byla zkoumána dále v průběhu celé věty. Zde byl zjištěn signifikantně nižší rozsah i variabilita F_0 v průběhu mluvy oproti zdravým mluvčím. [21, 22]

Další studie [41] byla zaměřena na změnu prozodické intenzity. Zkoumána byla změna intenzity v průběhu vět mající různý emocionální význam. V tomto případě byly porovnávány věty neutrální s větami vyjadřující vztek. U skupiny nemocných byla změřena menší změna intenzity mezi oběma druhy vět oproti skupině zdravých jedinců. Dále byla zjištěna signifikantně nižší změna intenzity hlasu u skupiny PN v průběhu čtení standardního textu. [42]

Vliv léčby na řeč

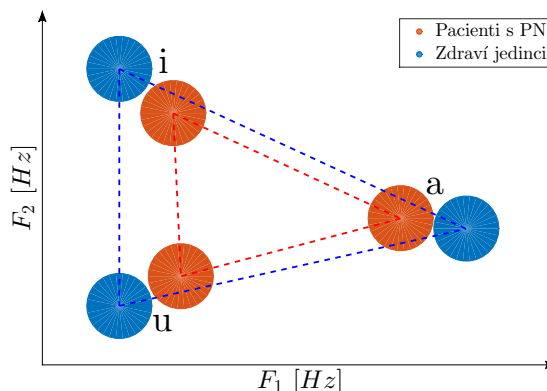
Několik studií ukázalo, že se různé aspekty řeči při léčbě lékem L-DOPA zlepšují. Studie [43] odhalila v úloze spontánního monologu významné zvýšení srozumitelnosti řeči po zahájení léčby. Lepší řečovou srozumitelnost po zahájení léčby odhalila také studie [44], ve které měli pacienti za úkol říkat slova či krátké fráze. Tyto studie však přinesly pouze subjektivní výsledky, kde byla srozumitelnost řeči hodnocena nezávislými posluchači. Studie [45], která zkoumala pohyblivost rtů, odhalila vyšší amplitudu pohybu rtů v průběhu mluvy po nasazení léku. Studie [46] ukázala, že po dvou letech od nasazení léku se u pacientů s PN zlepšila kvalita hlasu a artikulace a dále se zvýšila variabilita intenzity hlasu a variabilita základní frekvence F_0 . U pacientů bylo pozorováno zlepšení artikulace samohlásek a zvýšení variability F_0 společně se zlepšením bradykineze a rigidity.

1.3. Porucha artikulace samohlásek u PN

Hypokinetická dysartrie obvykle souvisí s nepřesnou artikulací samohlásek, což přispívá ke snížení srozumitelnosti řeči. [47] Nepřesná artikulace samohlásek souvisí zejména s nižším rozsahem a rychlostí pohybu artikulačních orgánů. Nízký rozsah pohybu jazyka a rtů ovlivňuje hodnoty prvních dvou formantových frekvencí (viz kap. 1.1.2). Omezený rozsah pohybu artikulátorů má za následek centralizaci samohláskových fonémů. To znamená, že formanty s přirozeně vyšší frekvencí mají tendenci se snižovat a naopak formanty s přirozeně nižší frekvencí nabývají vyšších hodnot než je obvyklé u zdravých lidí. [48, 49]

Celkové snížení pracovního prostoru samohlásek u PN se odráží v nižším obsahu trojúhelníka tvořeného první a druhou formantovou frekvencí rohových samohlásek (viz obr. 6). Tento obsah se označuje *VSA* (z anglického *vowel space area*) a je dán vzorcem (1). Některé studie (např. [49, 50]) však parametr *VSA* neshledávají statisticky významným pro rozlišení mezi

pacienty s PN a zdravými jedinci. Tento fakt může být pravděpodobně způsoben vysokou variabilitou mluvčích. Hodnota VSA je ovlivněna individuálními vlastnostmi mluvčích a souvisí zejména s pohlavím a věkem. [51] Tyto skutečnosti pak zastiňují skutečné rozdíly mezi VSA u zdravých lidí a pacientů s PN.



Obrázek 6. Centralizace rohových samohlásek u pacientů s PN ve srovnání se zdravými jedinci.

$$VSA = \frac{1}{2} |F_{1i}(F_{2a} - F_{2u}) + F_{1a}(F_{2u} - F_{2i}) + F_{1u}(F_{2i} - F_{2a})| \quad (1)$$

Z tohoto důvodu byly zavedeny další parametry, které by spolehlivě oddělily skupinu zdravých jedinců od nemocných. Jedním z takových parametrů je VAI (z anglického *vowel articulation index*). VAI vyjadřuje poměr prvních dvou formantů rohových samohlásek tak, aby co nejlépe popisoval centralizaci samohláskových fonémů. Ze vzorce (2) je patrné, že přirozeně vyšší formantové frekvence jsou umístěny v čitateli a přirozeně nižší formantové frekvence se nachází ve jmenovateli. Další z nejčastěji používaných parametrů pro popis hypokineticke dysartrie je poměr F_{2i}/F_{2u} , jenž představuje relativní vzdálenost mezi druhými formanty samohlásek i a u . Oba tyto parametry byly shledány více citlivými pro rozlišení mezi skupinami zdravých jedinců a skupinou PN. [1, 49, 52, 53]

$$VAI = \frac{F_{1a} + F_{2i}}{F_{1i} + F_{1u} + F_{2a} + F_{2u}} \quad (2)$$

Nepřesnosti artikulace samohlásek byly zkoumány během různých řečových úloh. Mezi běžné úlohy patří čtení zadaného textu, vedení spontánního monologu, opakování zadané věty a trvalá fonace samohlásek. Studie [53], která se zabývala mimo jiné porovnáním různých úloh z hlediska citlivosti na nepřesnost artikulace samohlásek pro rozlišení mezi zdravými jedinci a pacienty trpící PN, prokázala nejvyšší citlivost u spontánního monologu. Naopak nejméně citlivou úlohou byla shledána trvalá fonace samohlásek. Studie [54] ukázala, že u pacientů s PN je spontánní řeč méně srozumitelná než například čtení nebo opakování frází. Z toho lze také usoudit, že vyšší rozdíly mezi zdravými a nemocnými budou patrné u úlohy spontánního monologu.

Studie [53] ukázala, že poruchy artikulace samohlásek jsou mezi prvními příznaky spojenými s poruchou řeči u PN. Poruchy artikulace tak mohou být jedním z prvních měřitelných projevů PN a jejich vyhodnocení může pomoci v brzké diagnóze. Ze studií [55] a [56] vyplývá, že závažnost dysartrie u PN je ovlivněna závažností motorických symptomů a délkou onemocnění. Díky hodnocení poruch artikulace samohlásek tak lze objektivně určit míru svalového postižení. Dále bylo zjištěno, že zlepšení artikulace samohlásek koreluje se změnami rigidity

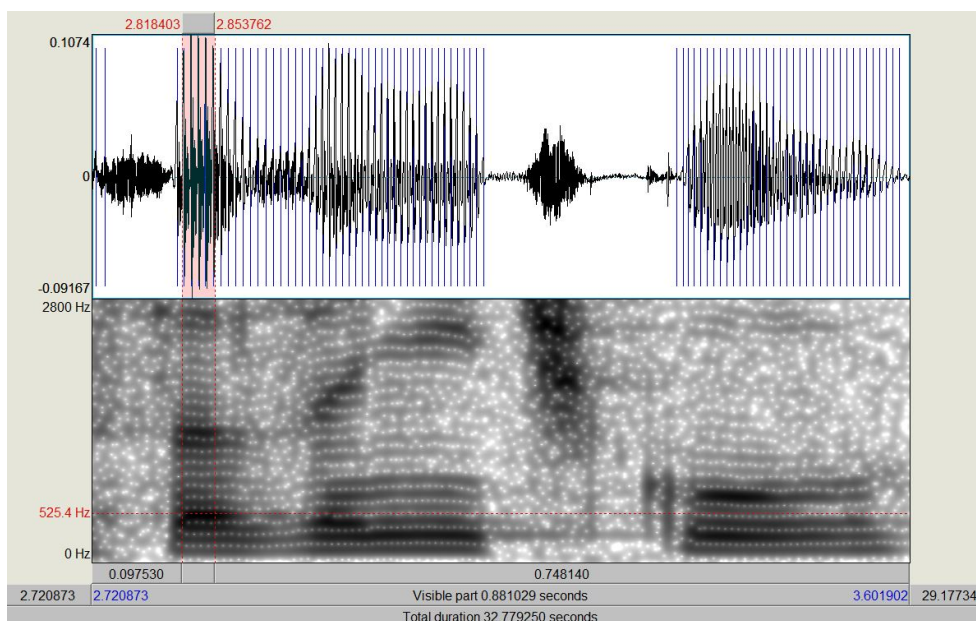
1. Úvod

a bradykineze v průběhu léčby. [46] Proto může hodnocení artikulace přinést objektivní informaci o účincích léčby.

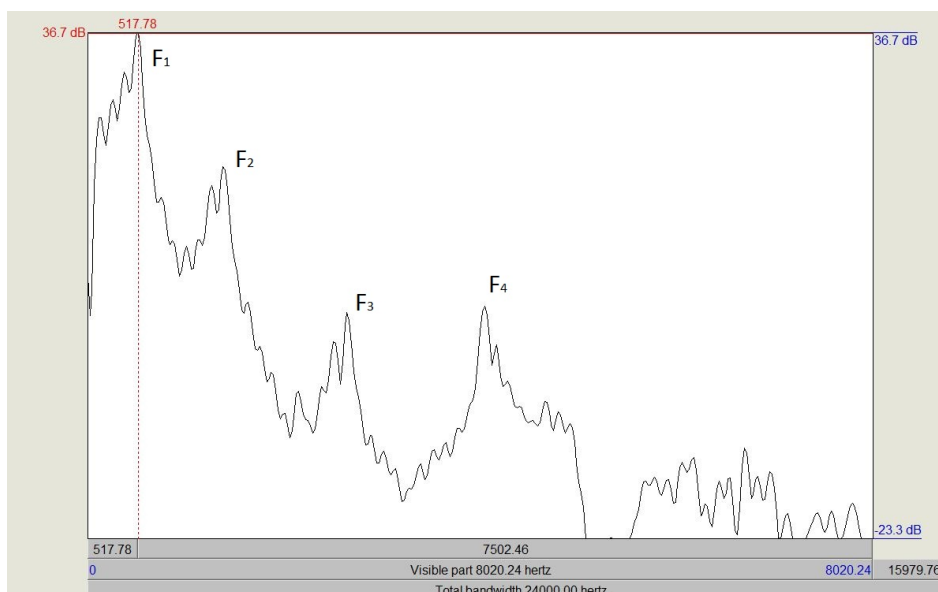
1.4. Ruční analýza formantových frekvencí

Hodnocení poruch artikulace samohlásek probíhá tradičně ruční analýzou formantových frekvencí (viz např. [1, 52, 53]). Ruční analýza se provádí většinou pomocí specializovaného softwarového nástroje, jenž umožňuje pro vybranou část signálu zobrazit spektrogram a spektrum. Příkladem takového nástroje je software *Praat*. [57] Z řečového záznamu je vybráno několik vhodných samohlásek *a*, *i* a *u*. V řečovém signálu jsou následně vyhledány části odpovídající vybraným samohláskám, ze kterých jsou na základě spektrogramu nebo spektra určeny první dvě formantové frekvence. Formanty jsou zvláště pro každou samohlásku zprůměrovány. Z průměrných hodnot jsou dopočítány další parametry (VSA , VAI a F_{2i}/F_{2u}), na základě kterých probíhá hodnocení artikulace.

Na obr. 7 nahoře je znázorněn úsek řečového signálu *i* s vybranou částí odpovídající samohlásce (červeně vyznačená oblast). Na stejném obrázku dole je zobrazen spektrogram. Ve spektrogramu jsou tmavě znázorněny energeticky vyšší frekvence pro daný čas, ze kterých lze odhadnout formantové frekvence. Někdy však nemusí být tyto frekvence zřetelné nebo nelze přesně určit jejich hodnotu. Z toho důvodu probíhá odhad formantů z frekvenčního spektra, kde jsou špičky odpovídající rezonančním frekvencím zřetelnější a lze tak přesněji určit jejich hodnotu (viz obr. 8).



Obrázek 7. Řečový signál a spektrogram slova „sazeničku“ s vybranou částí odpovídající fonému *a*.



Obrázek 8. Frekvenční spektrum vybrané části signálu s vyznačenými formantovými frekvencemi.

1.5. Metody pro automatický výpočet formantových frekvencí

Algoritmy pro automatický výpočet formantových frekvencí slouží k výpočtu časových průběhů formantů. V této práci jsou porovnány tři metody pro výpočet formantů, které jsou volně dostupné. Ostatní představené metody mají buď jistá omezení, nebo nejsou volně dostupné.

Nejběžnější algoritmy pro výpočet formantů jsou založeny na spektrální analýze a hledání špiček ve spektrální oblasti. Tyto algoritmy však nejsou dostatečně přesné a jsou ovlivněné šumem v pozadí. (viz např. [58])

Metoda [59] je založená na filtrování řeči časově proměnným filtrem pro každý formant. Následně je použit výpočet pro odhad špiček ve spektrální oblasti. Filtrování omezuje oblast výpočtu pro každou formantovou frekvenci. Tím nedochází k ovlivnění sousedních formantů a zároveň je tím i potlačen hluk v pozadí. Tento algoritmus pracuje spolehlivě pro zřetelně vyjádřené znělé úseky řeči. Pro neznělé úseky řeči nebo pauzy v průběhu mluvy může docházet k nenávratnému ovlivnění algoritmu a tím k nesprávnému výpočtu následujících formantů pro znělé úseky řeči.

Na předešlou metodu navazuje metoda [60], která přináší některá vylepšení. Algoritmus zahrnuje detektor energie formantů a detektor znělých úseků řeči. Algoritmus tedy nesleduje formanty během neznělých úseků řeči a během pauz a/nebo pokud formant nemá dostatečnou energii pro spolehlivý výpočet.

Nejnovější metoda [61] využívá toho, že se hlasový trakt chová jako autoregresivní systém. Pro každé frekvenční pásmo signálu, který je filtrován bankou filtrů, je počítána opakovaná autokorelační funkce (RACF). Díky RACF, která je použita na frekvenčně omezený signál, je pro každé pásmo výrazná jedna frekvence odpovídající právě hledané formantové frekvenci.

1.5.1. Praat

Jednou z metod pro automatický výpočet formantů použitý v práci je implementován v softwarovém nástroji Praat. [57] Algoritmus pracuje v několika krocích. Nejprve je signál kvůli kompresi převzorkován na frekvenci, která je dvakrát větší než je maximální formantová frekvence (F_{Max}). Převzorkování je bráno s ohledem na Shannonův vzorkovací teorém $F_s > 2F_{max}$, kde F_s je nová vzorkovací frekvence.

1. Úvod

Převzorkovaný signál je následně filtrován preemfázovým filtrem. Preemfázový filtr je filtr typu horní propust (HP). Filtr je aplikován z toho důvodu, že největší část energie řečového signálu se nachází v rozmezí 0 – 300 Hz, kdežto podstatná část signálu, která obsahuje formantové frekvence, má frekvenci nad 300 Hz. Díky preemfázovému filtru tak vyniknou energeticky nižší ale důležitější složky ve spektru signálu. [2]

Výpočet formantů probíhá v oknech, přičemž je každé okno váhováno Gaussovským oknem. Následně jsou vypočítány predikční koeficienty (LPC z anglického *linear predictive coding*). Metoda LPC odhaduje n -tý vzorek signálu $\hat{s}(n)$ pomocí lineární kombinace M předchozích vzorků signálu podle vzorce (3) znázorňující tzv. pólový model. Úloha výpočtu LPC koeficientů je označována jako predikční analýza. Při predikční analýze je snahou určit koeficienty a_m tak, aby bylo dosaženo co nejmenšího součtu kvadratických chyb ε mezi skutečným signálem $s(n)$ a signálem odhadnutým $\hat{s}(n)$. Součet kvadratických chyb je dán vzorcem (4). LPC koeficienty mají kromě predikčních schopností také vlastnosti spektrální. Normované LPC spektrum se získá pomocí z-transformace (5). Počet pólů, které jsou algoritmem počítány odpovídá dvojnásobku maximálního počtu formantů.

$$\hat{s}(n) = \sum_{m=1}^M a_m s(n-m) \quad (3)$$

$$\varepsilon = \sum_{n=1}^N [s(n) - \hat{s}(n)]^2 \quad (4)$$

$$S(f) = \left| \frac{1}{1 - \sum_m a_m z^{-m}} \right|_{z=e^{j\frac{2\pi f}{f_s}}} \quad (5)$$

Algoritmus z počátku nachází tolik formantů, kolik bylo zvoleno jako maximum v celém rozsahu od 0 Hz do F_{Max} . Takto nalezené formanty však mohou mít buď velmi nízkou frekvenci (blízko 0 Hz), nebo velmi vysokou frekvenci (blízko F_{Max}). Takto nízké, respektive vysoké frekvence pak odpovídají spíše artefaktům, které jsou způsobené výpočtem LPC koeficientů než-li skutečným formantovým frekvencím. Z tohoto důvodu jsou všechny frekvence, které jsou nižší než 50 Hz nebo vyšší než $F_{Max} - 50$ Hz, odstraněny.

1.5.2. Ghosh

Druhá metoda pro výpočet formantových frekvencí, která je zahrnutá v této práci, je implemtoována ve vypočetním prostředí *Matlab* a je volně dostupná. (viz [62]) Stejně jako metoda Praat, která byla popsána v kap. 1.5.1, i tato metoda počítá formantové frekvence z LPC koeficientů. Algoritmus pracuje na základě průměrování formantových frekvencí vypočítaných pro různá časově-frekvenční rozlišení spektrogramu.

Výpočet probíhá v iteracích, kdy je v každé iteraci změněno časově-frekvenční rozlišení. U vysokého časového rozlišení může vzniknout chyba odhadu formantů způsobena nízkým frekvenčním rozlišením a naopak u vysokého frekvenčního rozlišení se projevují nedostatky spojené s nízkým časovým rozlišením. Výsledné formanty jsou tedy vypočítány jako průměry formantů získaných ze všech iterací.

1.5.3. Mustafa & Bruce

Posledním z algoritmů pro výpočet formantových frekvencí, který je použit v této práci, má další vylepšení oproti metodě uvedené v [60]. Metoda [63] je navržena tak, aby lépe sledovala frekvenci v průběhu souvislé řeči. Dále si tato metoda poradí s variabilitou mluvčích a s různými zvuky v pozadí. Díky tomu je tento algoritmus použitelný i v běžném prostředí.

Blokové schéma algoritmu je znázorněno na obr. 9. Nejprve jsou pomocí preemfázového filtru odstraněny nízké frekvence signálu. Filtrovaný signál je převeden na analytickou reprezentaci signálu pomocí Hilbertovy transformace. Analytická reprezentace signálu umožňuje použití komplexních hodnot filtrů v bance filtrů. Konverze do analytické oblasti dále snižuje množství aliasingu v signálu, což vede k přesnějšímu odhadu formantových frekvencí.

Adaptivní banka filtrů vychází z metody [59], ale byla modifikována tak, aby byl potlačen účinek frekvence F_0 pro lepší odhad prvního formantu. Každé pásmo z banky filtrů je zodpovědné za pásmovou filtraci signálu pro odhad jednotlivých formantových frekvencí. Hodnoty nul a pólů každého filtru jsou aktualizovány v průběhu času na základě odhadů předchozích hodnot formantových frekvencí. To umožňuje dynamické potlačení rušení ze sousedních formantů a zdrojů hluku vyskytujících se v pozadí.

První čtyři formantové frekvence jsou odhadovány ze čtyř pásem adaptivní banky filtrů pomocí lineární predikce prvního řádu, která je použita pro každé pásmo. Analytický signál vycházející ze všech čtyř filtrů je váhován Hammingovým oknem. LPC koeficienty jsou vypočítány pouze pro znělé úseky řeči mající v daných pásmech dostatečnou energii.

Po filtrování signálu bankou filtrů je pro každé pásmo spočítána energie signálu. Aby bylo možné odhadnout konkrétní formantovou frekvenci, musí být energie vypočtena v dané formantovém pásmu vyšší než je prahová úroveň. Prahová úroveň energie se pro každé pásmo liší a je přizpůsobena pro dlouhodobé změny spektrálních energií v daném pásmu. Postupná úprava prahové úrovně zabraňuje dlouhodobé chybě detektoru energie a umožňuje rychlé zotavení algoritmu z krátkých hlasitých zvuků v pozadí.

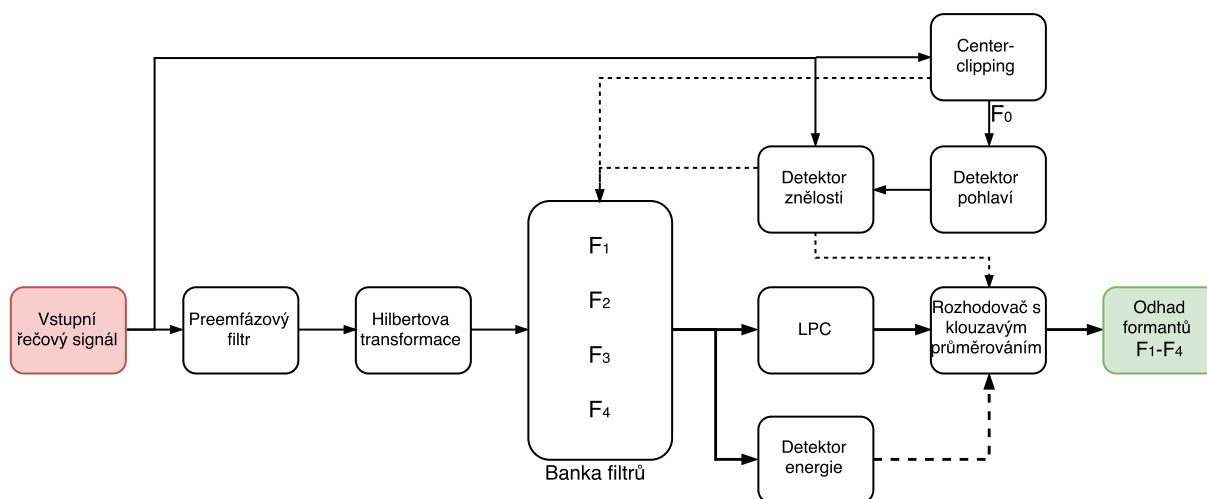
Detektor znělosti řeči slouží pro posouzení toho, zda je daný úsek řeči znělý či neznělý. Znělost řečového úseku se posuzuje na základě poměru energie nízkofrekvenčního a vysokofrekvenčního signálu, které se dostanou pomocí filtrování signálu dolní a horní propustí. Parametry detektoru musí být přizpůsobeny pro ženské i mužské mluvčí. Z tohoto důvodu je nutné detekovat i pohlaví mluvčího.

Detekce pohlaví vychází z odhadu frekvence F_0 . Algoritmus používá autokorelaci vycházející z metody *center-clipping*. [64] Metoda center-clipping je založena na hledání špiček v signálu, které udávají jeho periodu. Metoda potlačuje první formantovou frekvenci, čímž zamezuje záměně frekvence F_0 za F_1 a zvýší tak přesnost odhadu.

Posledním blokem ve schématu je rozhodovač s klouzavým průměrováním, jehož užití má dva účely. Prvním účelem je, že jsou za pomoci něho vypočítány a aktualizovány hodnoty každé z formantových frekvencí. Druhým účelem je rozhodnutí, zda aktuální formantovou frekvenci určit z odhadnutých LPC koeficientů, nebo zda formantovou frekvenci vypočítat jako průměrnou hodnotu z předchozích hodnot příslušného formantu. Pokud se v analyzovaném okně vyskytuje pouze znělá řeč a signál má dostatečnou energii, je aktuální hodnota formantu určena z LPC koeficientů. V opačném případě je aktuální formant vypočítán jako průměr z předchozích hodnot.

Tento algoritmus má omezení, které se týká hodnot jednotlivých formantových frekvencí. Pokud jsou nuly a póly filtrů příliš blízko u sebe dochází ke zkreslení frekvenční odezvy. Z toho důvodu jsou limitace pro to, jak blízko mohou být u sebe jednotlivé formanty, aby mohly být detekovány. Algoritmus neumožňuje detekovat první formantovou frekvenci, která je vzdálená méně než 150 Hz od frekvence F_0 . Podobně musí existovat rozestupy u ostatních frekvencí. Mezi F_1 a F_2 musí být minimálně 300 Hz, mezi F_2 a F_3 400 Hz a mezi F_3 a F_4 500 Hz. Pokud tato podmínka není splněna je hodnota x -té formantové frekvence vypočítána podle (6), kde ΔF je minimální rozestup mezi formantovými frekvencemi F_x a F_{x-1} .

$$F_x = F_{x-1} + \Delta F \quad (6)$$



Obrázek 9. Blokové schéma algoritmu Mustafa & Bruce.

1.6. Metoda pro automatický výpočet VSA

Algoritmus, který je podrobněji popsán v literatuře [65], slouží k automatickému výpočtu parametru VSA. Běžné metody počítají VSA jako obsah trojúhelníku nebo čtyřúhelníku (v závislosti na jazyku), který je dán rohovými samohláskami. Tento algoritmus bere v potaz skutečnost, že se nerohové samohlásky (v češtině *e* a *o*) mohou nalézat i vně trojúhelníku ohraničující vokální prostor. Tyto samohlásky jsou následně také zahrnuty do výpočtu VSA.

Algoritmus pracuje s řečovými signály, které obsahují dostatečný počet samohlásek. Signál je analyzován po oknech, kde v každém okně je vypočítána první a druhá formantová frekvence (formantový pár). Pro odhad prvních dvou formantových frekvencí je zde použita metoda dostupná v softwaru Praat (viz kap. 1.5.1).

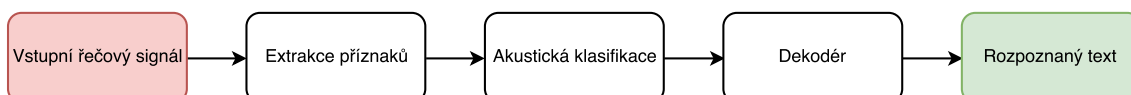
Z vypočtených hodnot formantových párů jsou vyloučeny odlehlé hodnoty. K identifikaci odlehlých hodnot je pomocí směsi Gaussovských modelů (GMM z anglického *Gaussian mixture model*) modelováno pravděpodobnostní rozdělení. Odlehlé hodnoty jsou určeny pomocí věrohodnostní funkce. Pro každý formantový pár je spočítána logaritmická věrohodnost a všechny formantové páry s nižší logaritmickou věrohodností než je $0,3L(F_p)$ jsou identifikovány jako odlehlé hodnoty a vyloučeny. Hodnota $L(F_p)$ odpovídá průměrné věrohodnosti všech formantových párů.

Zbývající hodnoty jsou shlukovány. Počet shluků odpovídá počtu samohlásek analyzovaného jazyka. Shlukování je prováděno metodou *k-means*. Střediska shluků jsou inicializovány podle očekávaných hodnot formantových frekvencí pro jednotlivé samohlásky. VSA je dáno obsahem konvexního obalu, který vznikne spojením příslušných středů shluků.

Tato metoda je zatím jedinou automatickou metodou počítající jeden z parametrů, který se dá využít pro hodnocení poruchy artikulace. Nicméně metoda byla testována pouze na zdravých anglicky mluvících lidech. Navíc parametr VSA není nejvhodnějším parametrem pro odlišení zdravých lidí od jedinců s PN (viz kap. 1.3). Tato metoda má i další omezení týkající se řečového signálu. Metoda vyžaduje v řečovém signálu velký počet samohlásek oproti souhláskám, což nemusí být v běžné mluvě splněno. Výsledky pak mohou být ovlivněny formantovými frekvencemi znělých souhlásek.

1.7. Rozpoznávač řeči pro segmentaci fonémů

Rozpoznávač řeči slouží k segmentaci fonémů vstupního řečového signálu. Schéma obecného rozpoznávače řeči je uvedeno na obr. 10. Rozpoznávač se skládá ze tří základních bloků: *extrakce příznaků*, *akustická klasifikace* a *dekodér*. Na vstupu rozpoznávače je řečový signál. *Extrakce příznaků* slouží k popisu řečového signálu. Tím, že jsou vynechány irelevantní informace o signálu, dochází k jeho kompresi. V bloku *akustická klasifikace* dochází ke klasifikaci části řečového signálu. Část měřeného signálu je porovnávána s uloženými ukázkami řeči. *Dekodér* následně přiřadí dané části signálu význam (například slovo nebo foném). Na výstupu rozpoznávače je rozpoznatý text.



Obrázek 10. Blokové schéma obecného rozpoznávače řeči.

V této práci je použit rozpoznávač fonémů, který pracuje na základě dlouhodobého časového kontextu. [66] Nejprve jsou v bloku *extrakce příznaků* vypočítány parametry pro popis signálu. Analýza signálu probíhá v okně o délce 25 ms, které se posouvá po signálu s krokem 10 ms. Na signál je aplikován preemfázový filtr. Následně je signál váhován Hammingovým oknem. V každém kroku je vypočítán vektor parametrů, který popisuje danou část signálu. Pro popis signálu jsou použity Melovské frekvenční keprální koeficienty.

V bloku *akustická klasifikace* se každé akustické jednotce (fonému) přiřazuje skóre na základě porovnání parametrů analyzovaného signálu s parametry pro jednotlivé fonémy uložené v databázi. Pro akustickou klasifikaci je využit hybridní systém, který kombinuje neuronové sítě a skryté Markovovské modely.

Blok *dekodér* slouží k transkripci řetězce vektorů do řetězce lexikálních jednotek (fonémů). Dekódování probíhá na základě znalostí akustických a lingvistických vlastností. Akustické vlastnosti se posuzují z vypočítaného vektoru parametrů popisující signál v bloku *akustická klasifikace*. Lingvistické vlastnosti souvisí s analyzovanou řečí a jsou posuzovány v návaznosti na předchozí hodnoty fonému. U lingvistických znalostí se například předpokládá malá pravděpodobnost dvou po sobě jdoucích samohláskových fonémů.

Vybraný rozpoznávač řeči je přizpůsoben pro 4 světové jazyky: češtinu, angličtinu, ruštinu a maďarštinu. Pro každý jazyk byla změřena jiná chyba rozpoznávače (viz tab. 6). Nejnížší chybu u rozpoznání fonémů vykazuje český a anglický jazyk. Naopak nejvyšší chyba byla zjištěna u ruského jazyka.

Jazyk	Chyba [%]
Čeština	24,24
Angličtina	24,24
Ruština	39,27
Maďarština	33,32

Tabulka 6. Relativní chyba použitého rozpoznávače řeči pro různé jazyky.

V práci je rozpoznávač řeči pro segmentaci fonémů použit v implementaci automatické metody pro hodnocení poruch artikulace samohlásek. Díky detekovaným fonémům, které vychází z rozpoznávače, mohou být analyzovány pouze části signálu odpovídající samohláskám.

1.8. Cíle práce

Hlavním cílem práce je navrhnout metodu pro automatické hodnocení poruch artikulace samohlásek u pacientů s PN. Zatím jediná dosud vytvořená metoda, která by mohla mít potenciál pro automatickou detekci poruch artikulace počítá pouze parametr *VSA* (viz kap. 1.6). Tato metoda má několik omezení, kvůli kterým nemusí být detekce poruch dostatečně přesná a tím pádem přínosná.

Pro návrh automatické metody budou použity algoritmy pro automatický výpočet formantových frekvencí. Jedním z cílů bude porovnat dostupné algoritmy (viz kap. 1.5) s výsledky ruční analýzy a vybrat z nich nejlepší, který bude sloužit k následné implementaci metody pro automatické hodnocení poruch artikulace.

U návrhu metody bude využito dvou přístupů - s využitím rozpoznávače řeči pro segmentaci fonémů a bez jeho využití. Metoda bez využití rozpoznávače bude částečně vycházet z popsané metody pro výpočet *VSA* (viz kap. 1.6). Cílem bude porovnat oba přístupy z hlediska citlivosti pro rozlišení mezi zdravými lidmi a jedinci s PN.

Dále bude automatická metoda analyzována z hlediska podobnosti výsledků s těmi, které byly pořízeny ruční analýzou a také z hlediska dalších omezení týkajících se algoritmu (z hlediska opakovatelnosti a délky promluv).

Vzhledem k různým formantovým frekvencím krátkých a dlouhých samohlásek (zejména samohlásek *i* a *i:*, viz kap. 1.1.2) bude zkoumán tento vliv na citlivost pro rozlišení mezi zdravými a nemocnými.

Nakonec bude testováno, jak dobře lze od sebe klasifikovat skupiny nemocných a zdravých lidí na základě výsledků automatické metody. Cílem je zjistit, jaké parametry jsou pro klasifikaci vhodné a jaké naopak nikoliv.

2. Metody

2.1. Databáze

Pro analýzu rozdílů mezi zdravými jedinci a pacienty trpících PN byla poskytnuta databáze čítající celkem 72 naměřených subjektů. Pro analýzu řečových signálů byly u subjektů pořízeny zvukové záznamy čteného textu a vedeného monologu. Čtený text byl pro všechny jedince stejný. Jednalo se o úryvek z díla Karla Čapka „O lidech“ (viz kap. A). U většiny jedinců byly pořízeny dva záznamy čteného textu.

U všech subjektů je uveden věk a pohlaví. U pacientů s PN jsou dále uvedeny informace o délce trvání nemoci, léčbě a míře postižení. Míra postižení je popsána parametry Hoehn-Yahr a UPDRS III (viz kap. 1.2.1).

Databáze je rozdělena na kontrolní skupinu zdravých jedinců (KS) a skupinu pacientů s PN. Databáze byla vytvořena tak, aby bylo v obou skupinách stejné zastoupení mužů a žen a zároveň tak, aby bylo v obou skupinách podobné věkové zastoupení. Kontrolní skupina obsahuje data od 36 zdravých jedinců a skupina PN 36 dat od jedinců postižených Parkinsonovou chorobou. V každé skupině je zastoupeno 15 žen a 21 mužů. Průměrný věk u jedinců z KS je $62,6 \pm 9,1$ a u jedinců z PN $63,0 \pm 9,6$ let. Rozmezí věku u KS činí 45 – 81 let a u PN 41 – 82 let.

Informace o skupině PN jsou uvedeny v tab. 7. Pacienti mají mírný až mírně pokročilý stupeň motorického postižení, což udává parametr Hoehn-Yahr, který se pohybuje v rozmezí 1 – 2,5 a parametr UPDRS III pohybující se v rozmezí 4 – 36. Důležitým parametrem, který popisuje míru postižení řeči je subskóre řeči u UPDRS III. Průměrné subskóre řeči nabývá hodnoty $0,8 \pm 0,6$, což znamená, že nejčastěji mají pacienti pouze mírnou ztrátu vyjadřovací schopnosti nebo hlasitosti řeči. Průměrná doba onemocnění je 8 let.

Parametr	\bar{x}	σ	Rozmezí
Doba onemocnění [roky]	8,0	4,7	1 – 24
L-DOPA ekvivalent [mg]	775	416	280 – 2080
Hoehn-Yahr	2,1	0,4	1 – 2,5
UPDRS III	17,4	8,6	4 – 36
UPDRS III subskóre řeči	0,8	0,6	0 – 2

Tabulka 7. Informace o nemoci u skupiny PN. \bar{x} je označení pro průměrnou hodnotu a σ pro směrodatnou odchylku. *Rozmezí* značí minimum a maximum daného parametru zjištěného u PN.

Všichni účastníci jsou čeští rodilí mluvčí. U zdravých mluvčích nebyly zjištěny vady řeči ani sluchu, které by mohly nepříznivě ovlivnit měření, ani jim v minulosti nebyla diagnostikována žádná neurologická nemoc. U skupiny PN byly pozorovány pouze vady řeči spojené s Parkinsonovou nemocí.

2.2. Nahrávání řeči

Pořizování řečových záznamů probíhalo v tiché místnosti s nízkou hladinou hluku z okolního prostředí. K nahrávání byl použit náhlavní kondenzátorový mikrofon (Beyerdynamic Opus 55) umístěn přibližně 5 cm od úst mluvčího. Pro zajištění srovnatelných podmínek pro všechny subjekty bylo nastaveno zesílení mikrofonu na optimální úroveň, které bylo pro všechny účastníky stejné. Zvuková data byla digitalizována z kazety do počítače se vzorkovací frekvencí 48 kHz a 16 bitovým kvantováním. Všichni účastníci byli zaznamenáváni ve stejné oblasti neurologické kliniky. Všichni účastníci byli obeznámeni s úkoly čteného textu a monologu i s postupem nahrávání. V případě chyby mohl účastník daný úkol opakovat. Během nahrávání nebyli účastníci nijak časově omezeni.

2.3. Ruční analýza formantových frekvencí

U pořízených řečových signálů byly pro vybrané samohlásky a , i a u určeny hodnoty první a druhé formantové frekvence. Vybrané samohlásky pro text jsou vyznačeny v kap. A. U textu bylo analyzováno celkem 30 samohlásek - 10 a , 10 i a 10 u .

U monologu měli mluvčí za úkol spontánně hovořit o tom, co dělají během aktuálního dne nebo týdne, o jejich rodinách, o zaměstnání nebo o zájmech. Pro každého účastníka bylo vybráno opět celkem 30 samohlásek - 10 a , 10 i a 10 u . Vzhledem k rozdílnému monologu každého mluvčího byly samohlásky vybrány podle těchto pravidel:

1. Slovo, ze kterého mohla být vybrána samohláska, muselo být jako celek srozumitelné a vjemově normální.
2. Z jednoho slova mohly být vybrány pouze stejné samohlásky.
3. Vybraná samohláska nesměla být ovlivněna vedlejšími vlivy jako je např. koartikulace okolních hlásek. Pro zajištění této podmínky mohly být vybrány pouze samohlásky, které byly vysloveny samostatně nebo následovaly po neznělé souhlásce.
4. Minimální délka samohlásky musela být alespoň 40 ms a obsahovat úsek o délce alespoň 30 ms, který by mohl být považován za stabilní část samohlásky. Za stabilní část samohlásky je považován segment bez výrazných změn první a druhé formantové frekvence.

Ruční analýza probíhalo v softwarovém nástroji Praat (viz kap. 1.4). Pomocí spektrogramu a výkonového spektra části signálu byly odhadovány první dvě formantové frekvence pro vybrané samohlásky. Formanty byly odhadovány ze středu stabilní části signálu o délce alespoň 30 ms tak, aby se zamezilo vlivu okolních fonémů. Naměřené formantové frekvence byly pro každého jedince a pro každý úkol odděleně zprůměrovány. Výsledkem pro každého účastníka pro text a pro monolog byly průměrné hodnoty prvních dvou formantů samohlásek a , i a u , z kterých byly vypočítány parametry VSA , VAI a F_{2i}/F_{2u} .

2.4. Algoritmy pro výpočet formantových frekvencí

V této práci jsou porovnány 3 metody pro automatický výpočet formantových frekvencí. Metoda uvedená v kap. 1.5.1 je implementována v softwarovém nástroji Praat. Metoda *Ghosh* (viz kap. 1.5.2) a metoda *Mustafa & Bruce* (viz kap. 1.5.3) jsou implementovány v Matlabu. V práci jsou tyto metody použity k automatickému výpočtu průběhů prvních dvou formantových frekvencí.

Pro výpočet formantů metodou Praat byly nastaveny parametry, které jsou uvedené v tab. 8. Časový krok 0,001 s udává velikost kroku, s jakým okno o délce 0,05 s klouže po signálu. Maximální velikost formantů je nastavena na hodnotu 5250 Hz. Preemfázový filtr, který je

nastaven na 50 Hz, značí mezní frekvenci filtru typu HP. Formantové frekvence jsou detekovány v rozmezí 50 až 5200 Hz, ve kterém je vypočítáno maximálně 5 formantů. Z nich jsou vybrány hodnoty dvou nejmenších formantů odpovídající první a druhé formantové frekvenci. Pokud se v některé části signálu nepodařilo vypočítat žádnou formantovou frekvenci nebo se podařilo určit pouze jednu, je daný úsek vyloučen.

Parametr	Hodnota
Krok [s]	0,001
Maximální počet formantů	5
Maximální formant [Hz]	5250
Délka okna [s]	0,05
Preemfázový filtr [Hz]	50

Tabulka 8. Nastavení algoritmu pro výpočet formantových frekvencí metodou *Praat*.

Metoda *Mustafa & Bruce* má nastavení stejné, jako je popsáno v literatuře [63]. Jediná hodnota, jež byla upravena, je délka okna, ze kterého jsou počítány formantové frekvence. Z důvodu rychlé reakce na změny byla délka okna snížena z původních 50 ms na 20 ms. Ostatní nastavení zůstalo nezměněno. Algoritmus počítá celkem 4 formantové frekvence, z nichž jsou použity dvě nejnižší. Výstupem algoritmu je také označení znělých a neznělých úseků řeči. Vzhledem k tomu, že se tato práce zabývá hodnocením samohlásek, které jsou všechny znělé, jsou části neznělých úseků odstraněny. Stejně tak jsou odstraněny úseky, ve kterých je vzdálenost první a druhé formantové frekvence menší než 315 Hz. Toto opatření je učiněno z důvodu, které plynou z rovnice (6) uvedené v kap. 1.5.3. Při poklesu rozdílů mezi frekvencemi prvních dvou formantů pod 300 Hz je hodnota druhého formantu dána vzorcem $F_2 = F_1 + 300$. Nepředpokládá se, že by hodnota druhého formantu byla skutečně vzdálena 300 Hz od prvního formantu, ale jedná se spíše o chybu algoritmu, kdy skutečná hodnota druhé formantové frekvence nebyla správně detekována. Z toho důvodu jsou odstraněny úseky, ve kterých se rozdíl obou frekvencí pohybuje pod 315 Hz.

U metody *Ghosh* bylo zachováno původní nastavení uvedené v [62]. Metoda počítá tři formantové frekvence, z nichž jsou použity první dvě. Výpočet probíhá v 8 iteracích, kde je v každé iteraci měněno časově-frekvenční rozlišení. Výsledné formantové frekvence jsou dány průměrem ze všech iterací.

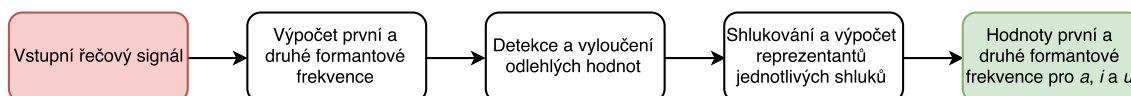
2.5. Algoritmy pro automatickou detekci artikulace samohlásek

V této práci byly využity dva přístupy pro automatickou detekci artikulace samohlásek - s využitím rozpoznávače řeči pro segmentaci fonémů řečového signálu a bez jeho využití. Vstupem do obou metod je řečový signál \vec{x} a výstupem jsou hodnoty první a druhé formantové frekvence pro samohlásky *a*, *i* a *u* (viz (7)). Z formantových frekvencí jsou následně vypočítány parametry VSA, VAI a F_{2i}/F_{2u} , které slouží k hodnocení poruch artikulace samohlásek.

$$\vec{x} \rightarrow \begin{pmatrix} F_{1a} & F_{2a} \\ F_{1i} & F_{2i} \\ F_{1u} & F_{2u} \end{pmatrix} \rightarrow \begin{pmatrix} VSA \\ VAI \\ F_{2i}/F_{2u} \end{pmatrix} \quad (7)$$

2.5.1. Metoda bez rozpoznávače řeči

Metoda bez využití rozpoznávače řeči částečně vychází z algoritmu pro automatický výpočet VSA, který je popsán v kap. 1.6. Blokové schéma upraveného algoritmu je znázorněno na obr. 11. Nejprve jsou z řečového signálu \vec{x} vypočítány vektorové páry prvních dvou formantových frekvencí \vec{F}'_1 a \vec{F}'_2 (viz (8)). Z vektorových párů jsou vyloučeny odlehlé hodnoty. Výsledkem tohoto kroku jsou formantové páry bez odlehlých hodnot \vec{F}_1 a \vec{F}_2 . Pomocí shlukování jsou formantové páry rozděleny do shluků odpovídající jednotlivým samohláskám. Pro formantový pár samohlásky x jsou výsledkem příslušné vektory \vec{F}_{1x} a \vec{F}_{2x} . Z těchto vektorů jsou následně vypočítány reprezentanti jednotlivých shluků (F_{1x} a F_{2x}) odpovídající hledaným formantovým frekvencím samohlásek.

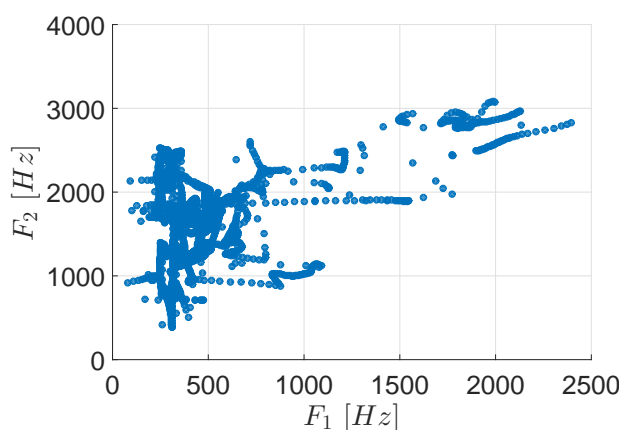


Obrázek 11. Blokové schéma automatické metody bez využití rozpoznávače řeči.

$$\vec{x} \rightarrow \begin{pmatrix} \vec{F}'_1 \\ \vec{F}'_2 \end{pmatrix} \rightarrow \begin{pmatrix} \vec{F}_1 \\ \vec{F}_2 \end{pmatrix} \rightarrow \begin{pmatrix} \vec{F}_{1a} & \vec{F}_{2a} \\ \vec{F}_{1e} & \vec{F}_{2e} \\ \vec{F}_{1i} & \vec{F}_{2i} \\ \vec{F}_{1o} & \vec{F}_{2o} \\ \vec{F}_{1u} & \vec{F}_{2u} \end{pmatrix} \rightarrow \begin{pmatrix} F_{1a} & F_{2a} \\ F_{1i} & F_{2i} \\ F_{1u} & F_{2u} \end{pmatrix} \quad (8)$$

Výpočet první a druhé formantové frekvence

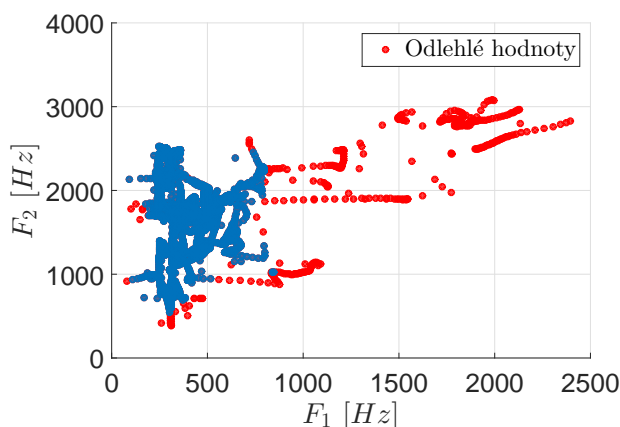
Pomocí zvolené metody pro výpočet formantů jsou pro vstupní signál vypočítány průběhy prvních dvou formantových frekvencí. Vypočtené páry formantových frekvencí jsou zachyceny na obr. 12. Formantové frekvence obsahují spoustu odlehlých hodnot, které jsou způsobeny zejména výskytem souhlásek v řečovém signálu a také chybami výpočetní metody.



Obrázek 12. Metoda bez rozpoznávače řeči: vypočtené formanty ze vstupního signálu.

Odlehlé hodnoty

Ve vypočtených vektorech jsou detekovány a následně vyloučeny odlehlé hodnoty. Nejprve jsou vyloučeny hodnoty, které neodpovídají fyziologickým předpokladům hlasového ústrojí. Takto jsou vyloučeny formantové páry mající hodnotu první formantové frekvence vyšší než 1000 Hz a/nebo formantové páry, které mají druhou formantovou frekvenci vyšší než 2800 Hz. Následně je vytvořen model Gaussovských směsí obsahující 5 shluků, z nichž každý odpovídá jedné české samohlásce. V tomto případě jsou dlouhé a krátké samohlásky považovány za jednu. Na základě logaritmické věrohodnosti jsou detekovány a poté vyloučeny odlehlé hodnoty. Kritérium pro vyloučení je takové, že jsou vyloučeny formantové páry s nižší logaritmickou věrohodností než $1, 2l(\bar{F}_p)$, kde $l(\bar{F}_p)$ je průměrná logaritmická věrohodnost všech formantových párů \vec{F}_1 a \vec{F}_2 . Vyloučené formantové páry znázorňuje obr. 13.



Obrázek 13. Metoda bez rozpoznávače řeči: detekce odlehlých hodnot.

Shlukování

Vektory bez odlehlých hodnot jsou shlukovány. Opět je shlukování prováděno do 5 shluků tentokrát metodou *k-means*. Středů shluků jsou inicializovány podle tab. 9. Počáteční středů shluků jsou zvoleny podle očekávané polohy formantů jednotlivých fonémů. To znamená, že například u samohlásky *u* se očekávají nejmenší hodnoty obou formantových frekvencí. U samohlásky *i* se očekává nejnižší hodnota prvního formantu a nejvyšší hodnota druhého formantu. Samohlásky *a* je inicializována na nejvyšší hodnotu prvního formantu a střední hodnotu druhého formantu.

	F_1	F_2
<i>a</i>	$\max(\vec{F}_1)$	$K_{50}(\vec{F}_2)$
<i>e</i>	$K_{50}(\vec{F}_1)$	$K_{75}(\vec{F}_2)$
<i>i</i>	$\min(\vec{F}_1)$	$\max(\vec{F}_2)$
<i>o</i>	$K_{50}(\vec{F}_1)$	$K_{25}(\vec{F}_2)$
<i>u</i>	$\min(\vec{F}_1)$	$\min(\vec{F}_2)$

Tabulka 9. Inicializace středů shluků u metody bez rozpoznávače řeči. *K* je označení pro hodnoty kvartilů. K_{25} - dolní kvartil, K_{50} - medián, K_{75} - horní kvartil.

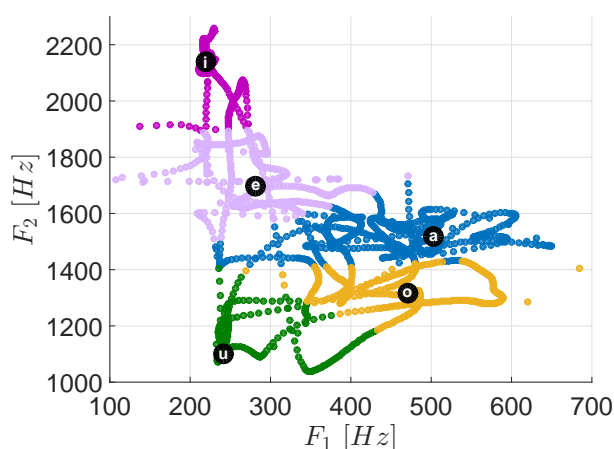
Každý z formantových párů je přiřazen jedné z 5 samohlásek. Samohlásce *x* odpovídají vektory formantových párů \vec{F}_{1x} a \vec{F}_{2x} , z kterých jsou vypočítány reprezentanti pro jednotlivé

2. Metody

samohlásky podle tab. 10. Výsledné shluky i s vypočtenými formantovými frekvencemi jednotlivých samohlásek uvádí obr. 14.

	F_1	F_2
a	$K_{75}(\vec{F}_{1a})$	$K_{50}(\vec{F}_{2a})$
e	$K_{50}(\vec{F}_{1e})$	$K_{50}(\vec{F}_{2e})$
i	$K_{25}(\vec{F}_{1i})$	$K_{75}(\vec{F}_{2i})$
o	$K_{50}(\vec{F}_{1o})$	$K_{50}(\vec{F}_{2o})$
u	$K_{25}(\vec{F}_{1u})$	$K_{25}(\vec{F}_{2u})$

Tabulka 10. Výpočet konečných formantových frekvencí jednotlivých shluků. K je označení pro hodnoty kvartilů. K_{25} - dolní kvartil, K_{50} - medián, K_{75} - horní kvartil.



Obrázek 14. Metoda bez rozpoznávací řeči: shluky s výslednými formantovými frekvencemi jednotlivých samohlásek.

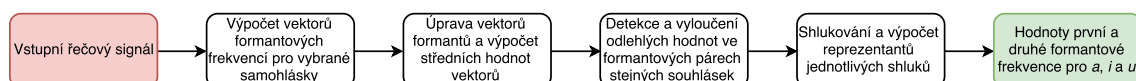
Formantové frekvence jednotlivých samohlásek musí splňovat podmínky uvedené v tab. 11. Pokud není splněna nějaká z podmínek 1 až 5 jsou obě hodnoty formantů nesplňující příslušnou podmínku prohozeny. Pokud by nebyla splněna jedna z podmínek 6 byla by hodnotě formantu F_{1a} přiřazena větší z hodnot F_{1o} a F_{1e} . Při nesplnění jedné z podmínek 7 bude hodnota formantu F_{2a} vypočítána jako průměrná hodnota z hodnot F_{2e} a F_{2o} . Tímto jsou určeny konečné hodnoty formantových frekvencí samohlásek a , i a u .

	Podmínka
1	$F_{1u} < F_{1o}$
2	$F_{2u} < F_{2o}$
3	$F_{1i} < F_{1e}$
4	$F_{2i} > F_{2e}$
5	$F_{2e} > F_{2o}$
6	$F_{1a} > F_{1o} \ \& \ F_{1a} > F_{1e}$
7	$F_{2a} > F_{2o} \ \& \ F_{2a} < F_{2e}$

Tabulka 11. Podmínky, které musí splňovat vypočtené formantové frekvence.

2.5.2. Metoda s rozpoznávačem řeči

Blokové schéma navrženého automatického detektoru využívajícího rozpoznávač řeči je uvedeno na obr. 15. Nejdůležitější částí algoritmu je blok *Výpočet vektorů formantových frekvencí pro vybrané samohlásky*, ve kterém jsou pro detekované samohlásky počítány vektorové páry prvních dvou formantů. Pro samohlásku x je vypočítáno N_x vektorů první a druhé formantové frekvence (viz (9)). Následuje úprava vektorů a vypočítání středních hodnot pro každý vektor. Pro každou samohlásku x jsou vypočítány dva vektory (\vec{F}'_{1x} a \vec{F}'_{2x}) o délce N_x odpovídající středním hodnotám vypočtených vektorů. Ve formantových párech stejných samohlásek jsou následně detekovány a vyloučeny odlehlé hodnoty (\vec{F}'_{1x} a \vec{F}'_{2x}). Posledním krokem je shlukování párů formantů a výpočet reprezentantů shluků, které odpovídají vypočteným formantům rohových samohlásek (F_{1x} a F_{2x}).



Obrázek 15. Blokové schéma automatické metody s využitím rozpoznávače řeči.

$$\vec{x} \rightarrow \begin{pmatrix} \vec{F}'_{1a} & \dots & \vec{F}'_{1a}^{N_a} & \vec{F}'_{2a} & \dots & \vec{F}'_{2a}^{N_a} \\ \vec{F}'_{1e} & \dots & \vec{F}'_{1e}^{N_e} & \vec{F}'_{2e} & \dots & \vec{F}'_{2e}^{N_e} \\ \vec{F}'_{1i} & \dots & \vec{F}'_{1i}^{N_i} & \vec{F}'_{2i} & \dots & \vec{F}'_{2i}^{N_i} \\ \vec{F}'_{1o} & \dots & \vec{F}'_{1o}^{N_o} & \vec{F}'_{2o} & \dots & \vec{F}'_{2o}^{N_o} \\ \vec{F}'_{1u} & \dots & \vec{F}'_{1u}^{N_u} & \vec{F}'_{2u} & \dots & \vec{F}'_{2u}^{N_u} \end{pmatrix} \rightarrow \begin{pmatrix} \vec{F}'_{1a} & \vec{F}'_{2a} \\ \vec{F}'_{1e} & \vec{F}'_{2e} \\ \vec{F}'_{1i} & \vec{F}'_{2i} \\ \vec{F}'_{1o} & \vec{F}'_{2o} \\ \vec{F}'_{1u} & \vec{F}'_{2u} \end{pmatrix} \rightarrow \begin{pmatrix} \vec{F}_{1a} & \vec{F}_{2a} \\ \vec{F}_{1e} & \vec{F}_{2e} \\ \vec{F}_{1i} & \vec{F}_{2i} \\ \vec{F}_{1o} & \vec{F}_{2o} \\ \vec{F}_{1u} & \vec{F}_{2u} \end{pmatrix} \rightarrow \begin{pmatrix} F_{1a} & F_{2a} \\ F_{1i} & F_{2i} \\ F_{1u} & F_{2u} \end{pmatrix} \quad (9)$$

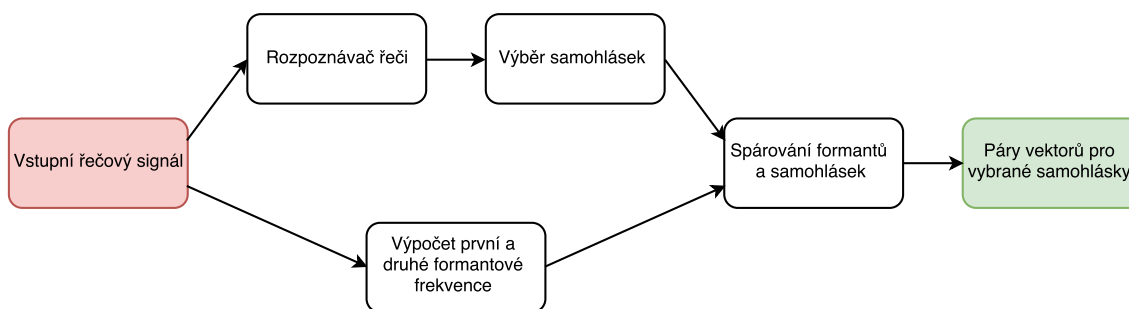
Výpočet vektorů formantových frekvencí pro vybrané samohlásky

Blokové schéma části algoritmu *Výpočet vektorů formantových frekvencí pro vybrané samohlásky* je zobrazeno na obr. 16. Řečový signál vstupuje jednak do rozpoznávače řeči a jednak do bloku pro výpočet formantových frekvencí. V rozpoznávači řeči je signál segmentován na soubor časově ohraničených fonémů. Z těchto fonémů jsou vybrány samohlásky. Výběr samohlásek může proběhnout podle různých pravidel (např. pouze po neznělých souhláskách, pouze krátké či dlouhé samohlásky apod.). V bloku pro výpočet první a druhé formantové frekvence jsou pomocí vybrané metody počítány průběhy prvních dvou formantů řečového signálu. Formantové frekvence a časové ohraničení odpovídající vybraným samohláskám jsou spárovány. Pro každou samohlásku x je detekováno N_x vektorových párů, kde každý pár odpovídá průběhu formantů jedné zachycené samohlásky. Minimální doba trvání samohlásky je 40 ms. Výsledek tohoto kroku představuje obr. 17, na kterém jsou znázorněny ohraničené formantové frekvence vybraných samohlásek.

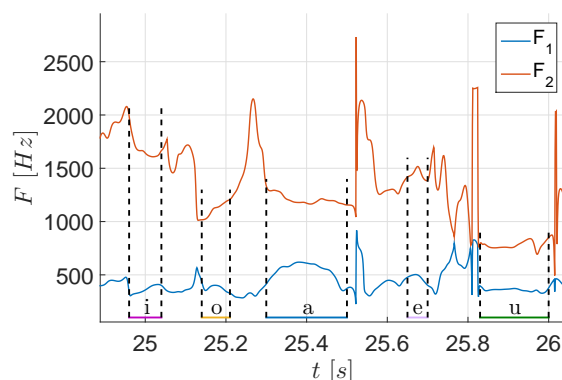
Úprava formantových vektorů vybraných samohlásek

Úprava vektorů probíhá v několika krocích (viz obr. 18). Nejprve je v každém vektoru oříznut začátek a konec. Důvodem toho je snaha eliminovat momenty, kdy dochází k vyslovování samohlásky z předešlého fonému, respektive následujícího fonému po samohlásce tak, aby vypočítané formantové frekvence byly těmito momenty co nejméně ovlivněny. Dalším důvodem je zpoždění výpočetní metody počítající formanty, se kterým je detekována správná formantová frekvence. Z každého vektoru je oříznuto prvních a posledních 10 ms.

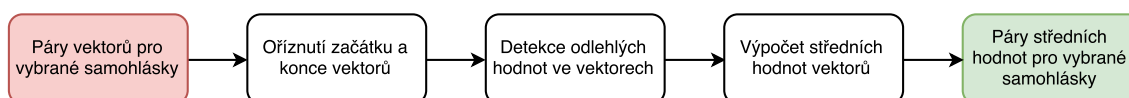
2. Metody



Obrázek 16. Blokové schéma výpočtu vektorů formantových frekvencí pro vybrané samohlásky.



Obrázek 17. Metoda s rozpoznávačem řeči: průběh formantových frekvencí s časovým ohraničením vybraných samohlásek.



Obrázek 18. Úprava formantových vektorů samohlásek.

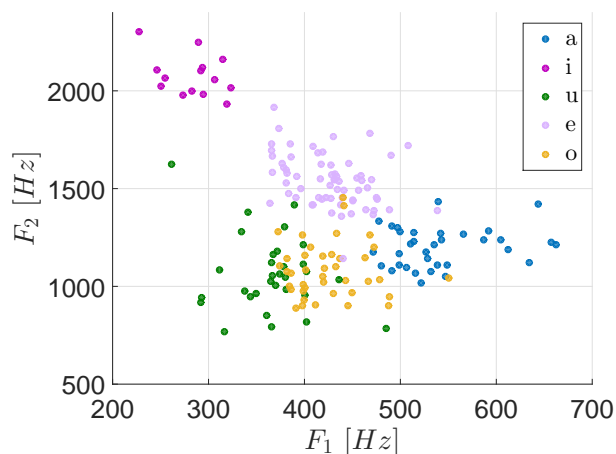
Dalším krokem je vyloučení odlehlých hodnot. Odlehlé hodnoty se ve vektorech mohou objevovat například chybou výpočetní metody počítající formanty, šumem v pozadí nebo zakolísáním hlasu mluvčího. Bude se předpokládat normální rozdělení dat ve vektorech. Odlehlé hodnoty jsou vyloučeny na základě *Mahalanobisovy vzdálenosti*. [67, 68] Mahalanobisova vzdálenost udává vzdálenost bodu \vec{x} od distribučního rozdělení X . V rovnici (10) odpovídá $\vec{\mu}$ středním hodnotám a S kovariační matici rozdělení X . Mahalanobisova vzdálenost pro normální rozdělení tvoří χ^2 distribuci s N stupni volnosti. Počet stupňů volnosti odpovídá dimenzi rozdělení X . V tomto případě jsou dimenze tvořeny první a druhou formantovou frekvencí, proto bude $N = 2$.

$$D_M(\vec{x}) = \sqrt{(\vec{x} - \vec{\mu})^T S^{-1} (\vec{x} - \vec{\mu})} \quad (10)$$

Detekce odlehlých hodnot probíhá ve dvou krocích. První krok je inicializační. V inicializačním kroku je pro každý bod z rozdělení X vypočítána Mahalanobisova vzdálenost. Jako odlehlé hodnoty jsou označeny body, jejichž Mahalanobisova vzdálenost je větší než $\chi_N^2(q)$, kde q je kvantil nabývající hodnoty 0,8. Body, které nebyly posouzeny jako odlehlé hodnoty tvoří nové rozdělení X' . V druhém kroku je počítána Mahalanobisova vzdálenost mezi novým rozdělením X' a mezi body, které byly v inicializačním kroku posouzeny jako odlehlé hodnoty. Body, které mají menší vzdálenost než $\chi_N^2(q)$, kde q je v tomto případě 0,9, jsou navraceny do rozdělení.

Ostatní hodnoty jsou označeny jako odlehlé a jsou z vektorů vyloučeny.

Z upravených vektorů bez odlehlých hodnot jsou vypočítány střední hodnoty. Pro každou samohlásku x je vypočítána dvojice vektorů o délce N_x odpovídající středním hodnotám vektorů prvních dvou formantů. Znázornění formantových frekvencí jednotlivých samohlásek je zobrazeno na obr. 19.



Obrázek 19. Metoda s rozpoznávacím řeči: vypočítané formantové frekvence z upravených vektorů.

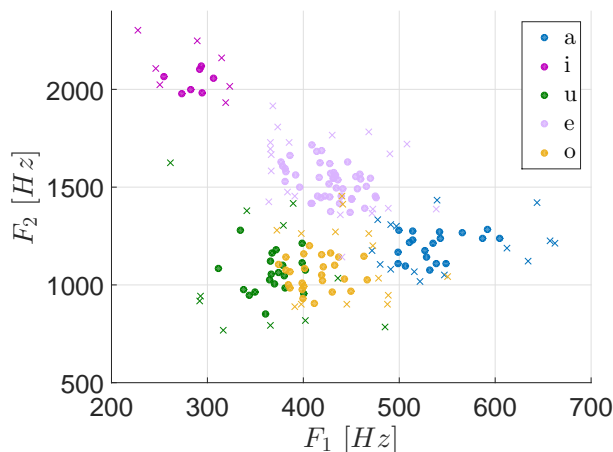
Odlehlé hodnoty ve formantových párech stejných samohlásek

U rozpoznávací řeči mohou nastat 3 druhy chyb. První chybou je, že rozpoznávací nezachytí samohlásku, kterou mluvčí v daném okamžiku řekl. V případě, že i přesto dokáže rozpoznávací v promluvě zachytit dostatek samohlásek, nenastává problém v tom, že by mohl být výsledek ovlivněn. Druhá chyba, která může být rozpoznávacím způsobena, je, že zachytí určitou samohlásku, i když mluvčí ve skutečnosti v daném čase vyslovil souhlásku. Třetí chybou je, že rozpoznávací zachytí jinou samohlásku než tu, která je ve skutečnosti vyslovena.

Další z chyb, která může v průběhu výpočtu vzniknout, je způsobena při výpočtu formantových frekvencí. Chyba se projevuje nejčastěji u samohlásek u a i . U samohlásky u může docházet k přiblížení první a druhé formantové frekvence. Z toho důvodu pak mohou být tyto frekvence považovány za jednu frekvenci a za hodnotu druhé formantové frekvence je považován třetí formant. U samohlásky i může docházet k přiblížení druhé a třetí formantové frekvence, která se při výpočtu považuje za jednu frekvenci. Tím dochází k určení vyšší hodnoty formantové frekvence, než je skutečná hodnota. Podobné chyby mohou vznikat i při výpočtu frekvencí u ostatních samohlásek.

Druhou a třetí chybou rozpoznávací a chybu výpočtu formantových frekvencí je nutné potlačit. U druhé chyby rozpoznávací většinou nastává případ, že jedna z formantových frekvencí (případně obě) je vzdálena od ostatních formantových frekvencí stejné samohlásky. Tuto chybu lze potlačit detekcí a následným vyloučením odlehlých hodnot. Tímto způsobem jsou eliminovány také chyby vzniklé při výpočtu formantových frekvencí. Detekce odlehlých hodnot probíhá stejným způsobem jako je popsáno v kap. 2.5.2 pomocí Mahalanobisovy vzdálenosti. Odlehlé hodnoty jsou detekovány pro každou samohlásku zvlášť. Mezní hodnota pro Mahalanobisovu vzdálenost u inicializačního kroku je nastavena na $\chi_N^2(0, 5)$ a u druhého kroku na $\chi_N^2(0, 6)$. Detekované odlehlé hodnoty znázorňuje obr. 20.

2. Metody



Obrázek 20. Metoda s rozpoznávačem řeči: detekce odlehlých hodnot ve formantových párech stejných samohlásek. Odlehlé hodnoty jsou znázorněny křížkem.

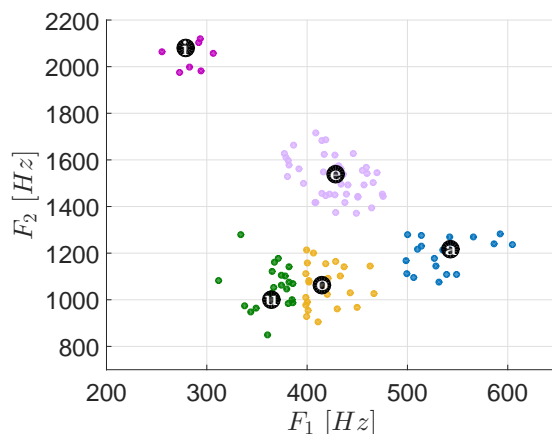
Shlukování

U třetí chyby rozpoznávače může dojít k tomu, že jsou hodnoty formantů chybně detekovaných samohlásek blíže správné hodnotě, tudíž nemusí dojít k vyloučení za pomoci detekce odlehlých hodnot. Tento případ nastává zejména pro dvojice samohlásek, které mají blízké formantové frekvence (*u-o*, *o-a*, *a-e*, *e-i*, *e-o*). Z toho důvodu je prováděno shlukování tak, aby souhlásky, které mohly být špatně rozpoznány, byly zařazeny do správné skupiny. Shlukování probíhá pomocí metody k-means. Inicializační hodnoty pro jednotlivé středy shluků jsou uvedeny v tab. 12.

	F_1	F_2
<i>a</i>	$\max(\vec{F}_{1a})$	$K_{50}(\vec{F}_{2a})$
<i>e</i>	$K_{50}(\vec{F}_{1e})$	$K_{75}(\vec{F}_{2e})$
<i>i</i>	$\min(\vec{F}_{1i})$	$\max(\vec{F}_{2i})$
<i>o</i>	$K_{50}(\vec{F}_{1o})$	$K_{25}(\vec{F}_{2o})$
<i>u</i>	$\min(\vec{F}_{1u})$	$\min(\vec{F}_{2u})$

Tabulka 12. Inicializace středů shluků u metody s rozpoznávačem. *K* je označení pro hodnoty kvartilů. K_{25} - dolní kvartil, K_{50} - medián, K_{75} - horní kvartil.

Reprezentanti pro daný shluk se vypočítají stejně jako u metody bez rozpoznávače (viz tab. 10). Stejně jako u metody bez rozpoznávače musí být splněny podmínky, které jsou uvedeny v tab. 11. Výsledek shlukování i s vypočtenými formanty jsou znázorněny na obr. 21.



Obrázek 21. Metoda s rozpoznávačem řeči: shluky s výslednými formantovými frekvencemi jednotlivých samohlásek.

2.6. Metody pro vyhodnocení výsledků

2.6.1. Test spolehlivosti

Test spolehlivosti (reliability) se využívá ke zjištění, zda jsou výsledky testu konzistentní. Spolehlivost se určuje pomocí *Pearsonova korelačního koeficientu* (viz rovnice (11)). Čím blíže jedné je koeficient korelace r , tím je test spolehlivější.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (11)$$

V práci bude měřena spolehlivost ruční analýzy i automatické metody. U ruční analýzy bude zjišťována tzv. *intra-judge reliability* a *inter-judge reliability*. Hodnota *intra-judge reliability* udává, s jakou spolehlivostí byly změřeny formantové frekvence jedním hodnotitelem. Hodnotitel provádí dvě měření pro stejná data po delším časovém intervalu, aby se co nejvíce zamezilo závislosti mezi měřeními. *Inter-judge reliability* udává spolehlivost výsledků, které byly vypočítány dvěma nezávislými hodnotiteli. Oba testy proběhnou na 25 % dat.

Obdobně budou testovány algoritmy pro automatický výpočet formantů. Na 25 % dat budou označeny časové údaje, v nichž se nacházejí samohlásky vyznačené v textu (viz kap. A). V těchto časových intervalech budou spočítány průměrné hodnoty formantů získaných pomocí různých algoritmů, které budou porovnávány s výsledky manuální metody. Z výsledků bude patrné, který z algoritmů počítá formantové frekvence nejspolehlivěji.

Pomocí testu spolehlivosti bude testována automatická metoda v porovnání s výsledky ruční analýzy. Bude vyhodnoceno, jaké formanty jsou pro jaké samohlásky pomocí algoritmů počítány spolehlivě a jaké nikoliv. Výsledky budou pouze orientační, protože u automatické metody nemohou být vybírány samohlásky podle stejných pravidel, jako tomu bylo v ruční analýze.

Dále bude proveden test spolehlivosti pro zjištění opakovatelnosti automatických metod. Test opakovatelnosti proběhne na úloze čtení, kde u většiny měřených subjektů byly pořízeny dva záznamy.

2.6.2. Test rozpoznávače řeči

Na 25 % dat bude proveden test rozpoznávače řeči. Test bude spočívat v posouzení, s jakou chybou dokáže rozpoznávač rozlišit samohlásky *a*, *i* a *u* vyznačené v textu (viz kap. A). Bude posuzována celková chyba rozpoznávače, tedy v kolika procentech případů nedokáže správně roz-

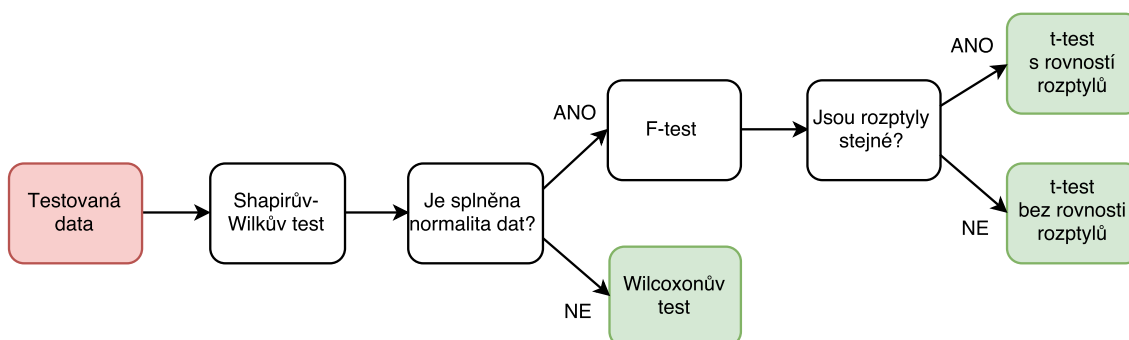
2. Metody

poznat hledanou samohlásku. Dále bude analyzováno, u jakých samohlásek je největší chyba a zda se liší chyby u KS a PN.

U testu rozpoznávače se dále bude testovat relativní četnost rozpoznávaných samohláskových fonémů vůči těm, které se vyskytují v textu. Tento test ukáže kolik procent samohlásek vyskytujících se v textu dokáže rozpoznávač identifikovat. Opět budou analyzovány rozdíly mezi KS a PN.

2.6.3. Test středních hodnot

Pro posouzení rozdílů mezi skupinami PN a KS budou provedeny testy středních hodnot pro vypočítané parametry (první dva formanty rohových fonémů, VSA , VAI a F_{2i}/F_{2u}). Pro test středních hodnot bude použit buď dvouvýběrový nepárový *Studentův t-test*, nebo dvouvýběrový nepárový *Wilcoxonův test*. Dvouvýběrový nepárový test předpokládá nezávislost dvou výběrů, což je v případě testování skupin KS a PN splněno. Co nemusí být splněno je normální rozdělení dat, které je předpokladem pro použití t-testu. Pro posouzení normality dat bude proveden Shapirův-Wilkův test. Při použití t-testu je dále potřeba zjistit, zda mají porovnávané výběry stejný nebo rozdílný rozptyl. Pro porovnání rozptylů bude použit *F-test*. Diagram výběru testu středních hodnot znázorňuje obr. 22.



Obrázek 22. Diagram výběru testu středních hodnot.

Jak u t-testu, tak Wilcoxonového testu je nulová hypotéza H_0 definována podle (12) a alternativní hypotéza H_A podle (13), kde μ_1 je střední hodnota testovaného parametru skupiny PN a μ_2 je střední hodnota parametru u KS. Výsledky testů budou hodnoceny na základě hladiny významnosti. Pro zamítnutí nulové hypotézy H_0 bude stanovena hladina významnosti 5 %. Za statisticky významné parametry pro odlišení skupin PN a KS tedy budou považovány takové, jejichž *p-hodnota* bude menší než 0,05.

$$H_0 : \mu_1 = \mu_2 \quad (12)$$

$$H_A : \mu_1 \neq \mu_2 \quad (13)$$

Pomocí testu rovnosti středních hodnot budou nejprve analyzovány rozdíly mezi muži a ženami u KS. Z výsledků bude patrné, jaké parametry jsou závislé na pohlaví. V návaznosti na výsledky budou analyzovány rozdíly mezi KS a PN jak pro celé skupiny, tak i pro rozdělené skupiny pouze na muže a pouze na ženy. Bude hodnoceno, jak se liší výsledky při rozdělení skupin na muže a ženy oproti výsledkům, které budou dosaženy u nerozdělených skupin.

V kap. 1.1.2 bylo zmíněno, že se mohou lišit formantové frekvence dlouhých a krátkých fonémů, což je patrné zejména u samohláskového páru $i, i:$. Na manuálních datech proto budou u úlohy čteného textu analyzovány rozdíly formantových frekvencí krátkých a dlouhých

samohlásek. Test rozdílů středních hodnot bude proveden pomocí *párového testu*, protože je předpoklad závislosti mezi hodnotami formantů dlouhých a krátkých fonémů. Budou porovnávány hodnoty formantových frekvencí mezi i a i : (v textu 4., 7. a 10. vyznačená samohláska i) a mezi u a u : (v textu 4., 6. a 10. vyznačená samohláska u). Výsledky prokáží, jestli je nutné brát v úvahu délku samohláskových fonémů, nebo jestli jsou formantové frekvence na délce fonémů nezávislé.

Nakonec budou analyzovány rozdíly mezi KS a PN dosažené automatickými metodami. U metody s rozpoznávačem řeči bude využito různých možností výběru samohlásek (viz tab. 13). Výběr samohlásek je přizpůsoben délce řečového signálu u obou úloh. Výběr po neznělých souhláskách zahrnuje také výběr po pauzách v mluvě. Tento způsob výběru zamezuje tomu, že formantové frekvence vybrané samohlásky nebudou ovlivněny formanty předchozí souhlásky (podobně jako v ruční analýze monologu, viz kap. 2.3). Dosažené výsledky budou porovnány s výsledky manuální metody.

Úloha	Výběr samohlásek
Text	Všechny samohlásky Všechny samohlásky kromě krátkého fonému i
Monolog	Všechny samohlásky Všechny samohlásky kromě krátkého i a dlouhého u : Samohlásky po neznělých souhláskách Samohlásky po neznělých souhláskách s vyloučením krátkého i

Tabulka 13. Způsoby výběru samohlásek pro úlohu textu a monologu.

2.6.4. Test minimální délky řečového signálu

Test minimální délky řečového signálu ukáže, jaký nejkratší signál má být zpracováván, aby bylo dosahováno spolehlivých výsledků u automatické metody. Test bude probíhat iterativně. V každém kroku se zvýší délka analyzovaného signálu. Pro každou délku řečového signálu t bude počítána chyba δ_t podle vzorce (14), kde F je vypočítaná hodnota formantu z celého řečového signálu a F_t je hodnota formantu pro řečový signál délky t . Jako minimální délka řečového signálu bude považována taková délka t , při kterém se ustálí hodnota chyby δ_t pod 5 %. Délka řečového signálu bude vyhodnocována pro oba formanty každé rohové samohlásky zvlášť. U metody s rozpoznávačem řeči bude analyzováno, kolik je potřeba detekovat fonémů, aby bylo dosaženo dostatečně přesných výsledků formantových frekvencí.

$$\delta_t = \frac{|F - F_t|}{F} \times 100 \quad (14)$$

2.6.5. Klasifikační metody

U nejlepších výsledků automatické metody, které vyplynou na základě statistické analýzy, bude proveden klasifikační experiment pro posouzení, s jakou přesností lze od sebe odlišit nemocné jedince od zdravých lidí. Bude vyhodnoceno, s jakou přesností lze klasifikovat skupiny KS a PN jednak na základě všech formantů dohromady a jednak na základě odvozených parametrů VSA , VAI a F_{2i}/F_{2u} . Z výsledků bude patrné, zda je pro odlišení obou skupin vhodnější použít formantové frekvence či odvozené parametry. Vzhledem k menší množině subjektů bude pro klasifikaci využito metody *leave-one-out*, což je speciální případ křížové validace, kdy je jeden subjekt použit jako testovací množina a na ostatních subjektech je klasifikátor natrénován.

2. Metody

Jako klasifikátor bude použit *Adaboost*. [69] *Adaboost* je založen na váženém hlasování několika slabších klasifikátorů, které se učí na iterativně se měnících vzorcích trénovacích dat. Zvolený počet iterací bude 5. Jako dílčí klasifikátor *Adaboostu* bude využit rozhodovací pařez, což je rozhodovací strom o hloubce 1.

Klasifikační experiment proběhne pro úlohy textu a monologu. Bude zhodnoceno, jaké parametry jsou vhodné pro oddělení KS a PN u obou úloh. Dále bude posouzeno, zda se bude lišit úspěšnost klasifikace v případě, že budou klasifikováni muži a ženy zvlášť. U klasifikace bude zohledněna jednak celková přesnost klasifikace a dále také senzitivita a specificita, jejichž hodnoty vycházejí z tzv. čtyřpolní tabulky (viz tab. 14). Čtyřpolní tabulka uvádí správně klasifikované subjekty, tedy kolik nemocných jedinců bylo klasifikováno jako nemocní (*TP* z anglického *true positive*) a kolik zdravých jako zdraví (*TN* z anglického *true negative*). Dále uvádí počet chybných klasifikací, tedy počet zdravých klasifikováno jako nemocní (*FP* z anglického *false positive*) a počet nemocných klasifikováno jako zdraví (*FN* z anglického *false negative*). Celková přesnost klasifikace (*ACC* z anglického *accuracy*) je dána vzorcem (15). Senzitivita (*SE*) udává úspěšnost, s jakou jsou rozpoznány skutečně nemocné osoby (viz vzorec (16)). Specificita (*SP*) na druhou stranu udává schopnost rozpoznat zdravé jedince (viz vzorec (17)).

		Skutečnost	
		Nemocní	Zdraví
Test	Nemocní	<i>TP</i>	<i>FP</i>
	Zdraví	<i>FN</i>	<i>TN</i>

Tabulka 14. Čtyřpolní tabulka.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

$$SE = \frac{TP}{TP + FN} \quad (16)$$

$$SP = \frac{TN}{TN + FP} \quad (17)$$

3. Výsledky

3.1. Ruční analýza

3.1.1. Intra-judge a inter-judge reliabilita

Výsledky testu spolehlivosti ruční analýzy jsou uvedeny v tab. 15. Hodnoty Pearsonových korelačních koeficientů u všech formantových frekvencí se statisticky významně liší od nuly ($p < 0,001$), což značí vysokou spolehlivost provedené ruční analýzy. Ze všech formantových frekvencí vykazuje nejnižší spolehlivost odhad prvního formantu fonému i . To může být způsobeno pravděpodobně nízkým rozsahem tohoto fonému, proto i menší odchylky mohou výrazněji ovlivnit hodnotu korelace. Hodnota F_{1i} nabývá společně s F_{1u} nejnižší průměrné hodnoty všech analyzovaných formantových frekvencí a případná menší nespolehlivost ve výpočtu této frekvence nebude mít téměř žádný vliv na spolehlivost výpočtu odvozených parametrů VSA a VAI.

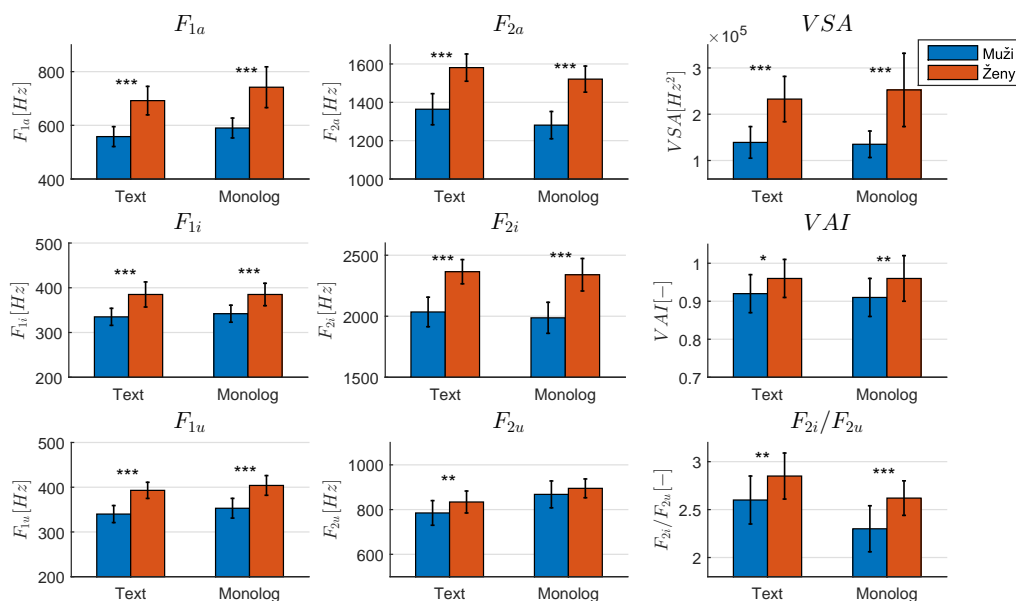
	<i>Intra-judge</i>		<i>Inter-judge</i>	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
F_{1a}	0,95	< 0,001	0,95	< 0,001
F_{2a}	0,95	< 0,001	0,96	< 0,001
F_{1i}	0,84	< 0,001	0,79	< 0,001
F_{2i}	0,97	< 0,001	0,96	< 0,001
F_{1u}	0,94	< 0,001	0,93	< 0,001
F_{2u}	0,93	< 0,001	0,95	< 0,001

Tabulka 15. Výsledky intra-judge a inter-judge reliability pro jednotlivé formantové frekvence. r - Pearsonův korelační koeficient; p - p-hodnota.

3.1.2. Porovnání parametrů mužů a žen

Výsledky testu středních hodnot mezi muži a ženami pro úlohu čteného textu a monologu jsou uvedeny na obr. 23. S výjimkou formantu F_{2u} u úlohy monologu byly rozdíly středních hodnot všech parametrů shledány statisticky významnými. Tento fakt může způsobit, že parametry pro posouzení rozdílů mezi KS a PN mohou být ovlivněny pohlavím a tím pádem budou zastíněny rozdíly mezi oběma skupinami. Z obrázku lze vyčíst, že ženy mají vyšší formantové frekvence u všech samohlásek. Nejmenší rozdíly mezi oběma skupinami byly pozorovány u parametrů F_{2u} u úlohy monologu a VAI a F_{2i}/F_{2u} u úlohy textu.

3. Výsledky

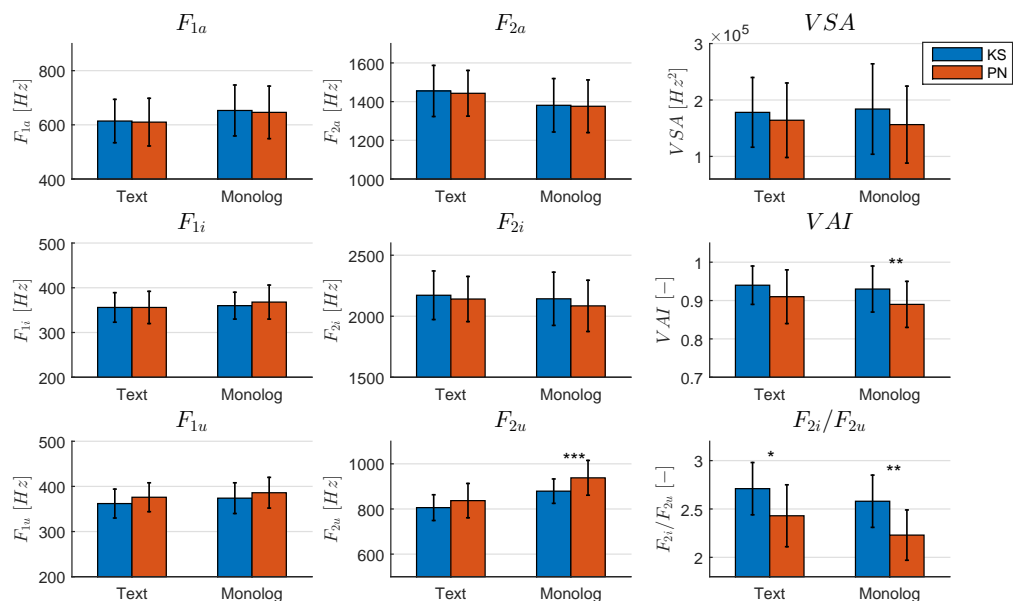


Obrázek 23. Porovnání středních hodnot parametrů mužů a žen u KS pro text a monolog. Hvězdičky indikují signifikantní rozdíly mezi ženami a muži: * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$.

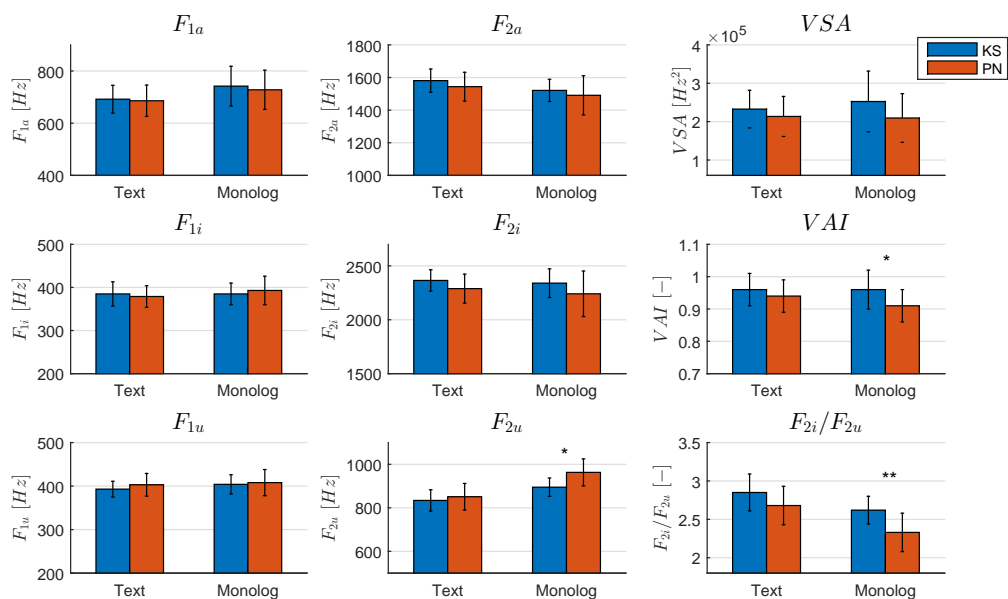
3.1.3. Porovnání parametrů mezi skupinami KS a PN

Obr. 24 zobrazuje rozdíly středních hodnot parametrů mezi skupinami KS a PN pro úlohu textu a monologu. U úlohy textu byly pozorovány největší rozdíly mezi oběma skupinami u parametrů F_{1u} , F_{2u} , VAI a F_{2i}/F_{2u} . Pouze parametr F_{2i}/F_{2u} ($p < 0,05$) však byl z těchto parametrů vyhodnocen jako statisticky významný pro odlišení obou skupin. U úlohy monologu byly zjištěny 3 statisticky významné parametry, které od sebe odlišují obě skupiny: F_{2u} ($p < 0,001$), VAI ($p < 0,01$) a F_{2i}/F_{2u} ($p < 0,01$). Z obrázku jsou dále patrné rozdíly středních hodnot u jednotlivých formantových frekvencí. Rozdíly jsou patrné především u formantových frekvencí F_{2i} , F_{1u} a F_{2u} . Podle očekávání jsou průměrné hodnoty frekvencí F_{1u} a F_{2u} u skupiny PN vyšší než u KS. Naopak průměrná frekvence F_{2i} nabývá vyšší hodnoty u KS. Tento fakt je způsoben centralizací formantových frekvencí, která se projevila i u parametrů VSA , VAI a F_{2i}/F_{2u} , které dosahují u skupiny PN nižších hodnot.

Jelikož jsou téměř všechny parametry závislé na pohlaví mluvčího, proběhla analýza rozdílů mezi skupinami KS a PN pro obě úlohy zvlášť u žen a zvlášť u mužů. Výsledky jsou shrnuty v obr. 25 u žen a v obr. 26 u mužů. Z výsledků lze vyčíst, že nebyly shledány žádné signifikantní parametry u žen u úlohy textu. Naopak u té samé úlohy byly u mužů shledány statisticky signifikantními parametry F_{1u} ($p < 0,01$) a F_{2i}/F_{2u} ($p < 0,05$). U úlohy monologu byly pozorovány statisticky významné parametry jak u mužů, tak i u žen. U žen byly nalezeny největší rozdíly u parametrů F_{2u} ($p < 0,05$), VAI ($p < 0,05$) a F_{2i}/F_{2u} ($p < 0,01$). U mužů byly kromě parametrů F_{2u} ($p < 0,05$), VAI ($p < 0,05$) a F_{2i}/F_{2u} ($p < 0,05$) shledány jako statisticky významné parametry také F_{1u} ($p < 0,05$) a VSA ($p < 0,05$). Vzhledem k těmto skutečnostem budou u automatických metod vyhodnocovány především parametry F_{1u} , F_{2u} , VSA , VAI a F_{2i}/F_{2u} i s ohledem na pohlaví mluvčích.

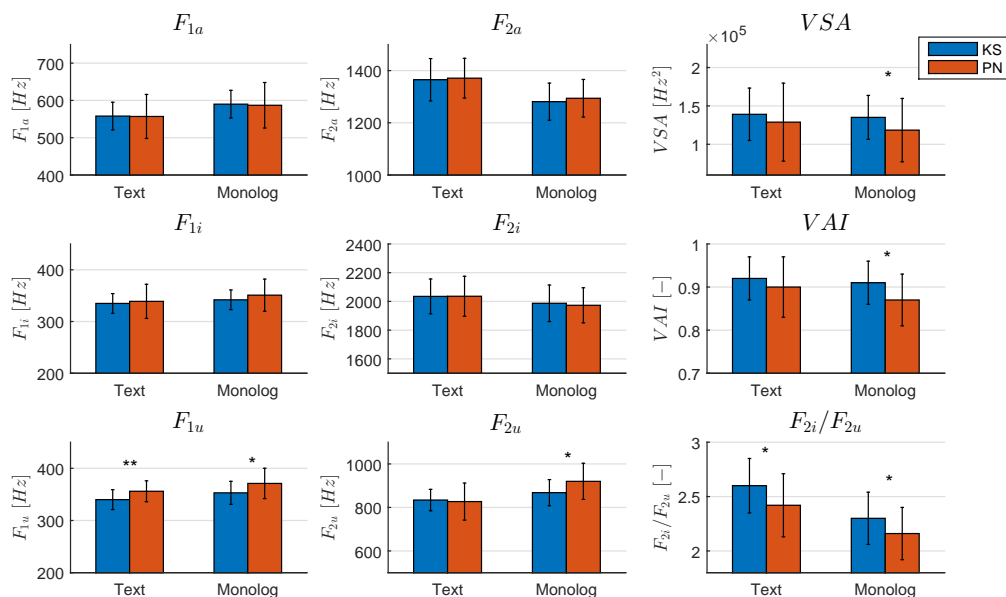


Obrázek 24. Porovnání středních hodnot parametrů mezi KS a PN pro text a monolog u mužů a žen dohromady. Hvězdičky indikují signifikantní rozdíly mezi KS a PN: * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$.



Obrázek 25. Porovnání středních hodnot parametrů mezi KS a PN pro text a monolog u žen. Hvězdičky indikují signifikantní rozdíly mezi KS a PN: * $p < 0,05$, ** $p < 0,01$.

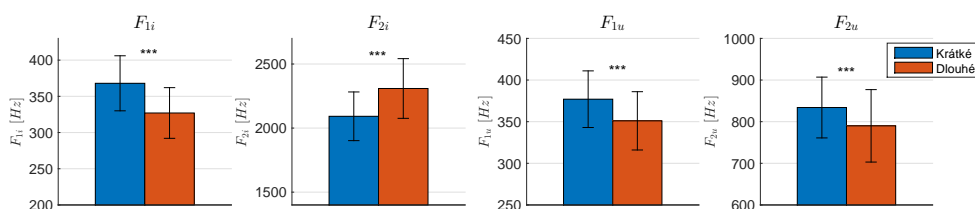
3. Výsledky



Obrázek 26. Porovnání středních hodnot parametrů mezi KS a PN pro text a monolog u mužů. Hvězdičky indikují signifikantní rozdíly mezi KS a PN: * $p < 0,05$, ** $p < 0,01$.

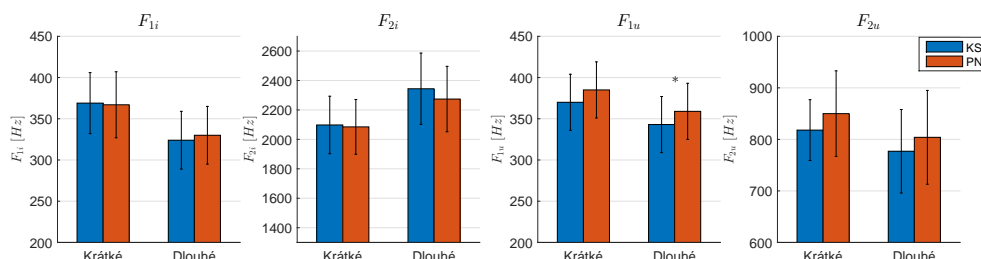
3.1.4. Porovnání formantů dlouhých a krátkých samohlásek

Porovnání středních hodnot mezi formantovými frekvencemi krátkého fonému i a dlouhého fonému i : a mezi formantovými frekvencemi krátkého fonému u a dlouhého fonému u : jsou uvedeny na obr. 27. Párový test středních hodnot ukázal u obou formantových frekvencí mezi i a i : a mezi u a u : statisticky významné rozdíly ($p < 0,001$). Nejvyšší rozdíl je patrný u formantu F_{2i} , u kterého převyšuje průměrná hodnota frekvence dlouhých fonémů frekvenci krátkých fonémů průměrně o více než 210 Hz. První formantová frekvence dlouhého fonému i : má v průměru o 40 Hz menší hodnotu než frekvence krátkého fonému. Vzhledem k tomu, že se průměrná formantová frekvence obou formantů u zdravých jedinců pohybuje kolem 340 – 390 Hz u F_{1i} , respektive 2050 – 2350 Hz u F_{2i} (v závislosti na pohlaví), je patrné, že rozdíl formantu i a i : může dosahovat i více než 10 % průměrné hodnoty obou formantových frekvencí. Podobně to platí také pro samohlásku u , kde rozdíly mezi formanty krátkých a dlouhých fonémů dosahují přes 5 % průměrné hodnoty formantů. Ve vokalickém prostoru se tento jev projevuje tak, že formantové frekvence dlouhých fonémů se budou vzdalovat od středu, zatímco formanty krátkých fonémů se budou více centralizovat.



Obrázek 27. Porovnání formantů krátkých a dlouhých samohláskových fonémů u úlohy čteného textu. Hvězdičky indikují rozdíly mezi formanty krátkých a dlouhých fonémů: *** $p < 0,001$.

Pokud se ze čteného textu extrahují formanty buď pouze pro krátké, nebo pouze pro dlouhé samohlásky, budou se lišit rozdíly mezi skupinami KS a PN (viz obr. 28). Oproti výsledkům, kde jsou uvažovány dlouhé i krátké samohlásky dohromady (viz obr. 24) se liší zejména rozdíly u druhého formantu dlouhého fonému i :, kde namísto 30 Hz je průměrný rozdíl mezi KS a PN více než 70 Hz. Ani tento rozdíl však není statisticky signifikantní. Nicméně například vypočítaný parametr F_{2i}/F_{2u} z formantů dlouhého fonému i : a krátkého fonému u již statisticky významný je ($p < 0,05$). Jako signifikantní parametr byl vyhodnocen také první formant dlouhého fonému u : ($p < 0,05$), což je další patrný rozdíl oproti tomu, kdy jsou uvažovány dlouhé a krátké fonémy dohromady. Bohužel foném u : není v běžné mluvě příliš zastoupen (viz obr. 4) a pro jeho detekci by bylo potřeba dlouhého řečového signálu.



Obrázek 28. Porovnání středních hodnot mezi KS a PN u formantů F_{1i} , F_{2i} , F_{1u} a F_{2u} při extrakci pouze krátkých, respektive pouze dlouhých fonémů z textu. Hvězdičky indikují rozdíly mezi formanty krátkých a dlouhých formantů: $*p < 0,05$.

V úloze čtení, kde má každý mluvčí stejný text a navíc jsou vybírány stejné samohlásky, nemusí být rozdíly mezi vybíráním všech samohlásek či pouze krátkých nebo dlouhých samohlásek tak markantní. Tyto rozdíly se však mohou projevit u vedeného monologu. Každý mluvčí vede různý monolog, proto se budou lišit i počty krátkých a dlouhých fonémů v řeči. Při vybírání všech samohlásek tak může nastat nežádoucí jev, kdy rozdílné zastoupení dlouhých a krátkých fonémů v řeči zastíní skutečné rozdíly mezi KS a PN. Z tohoto důvodu bude algoritmus s rozpoznávačem řeči testován na základě různých výběrů samohláskových fonémů.

3.2. Automatické metody

3.2.1. Test spolehlivosti algoritmů pro výpočet formantů

Výsledky testu spolehlivosti algoritmů pro výpočet formantových frekvencí udává tab. 16. Ze tří srovnávaných algoritmů vykazuje nejvyšší spolehlivost algoritmus implementovaný v softwaru *Praat*. Naopak nejhůře dopadl algoritmus *Ghosh*, u kterého nebyla ani u jednoho formantu překročena hodnota Pearsonova korelačního koeficientu 0,3. Frekvence F_{2a} má dokonce záporný koeficient korelace a většina průměrných hodnot formantů vypočítaných tímto algoritmem se liší v řádu stovek Hz od výsledků ruční analýzy. To indikuje nevhodnost použití tohoto algoritmu u úloh jako je čtený text a vedený monolog.

Metoda *Mustafa & Bruce* počítá spolehlivě pouze formanty F_{1a} ($r = 0,82$, $p < 0,001$) a F_{1u} ($r = 0,92$, $p < 0,001$). Nejhůře jsou počítány formanty fonému i . Zde se liší hodnoty průměrné formantové frekvence oproti ruční analýze o 223 Hz u F_{1i} a o 493 Hz u F_{2i} .

Metoda *Praat* počítá s vysokou spolehlivostí všechny formantové frekvence, zejména F_{1a} ($r = 0,95$, $p < 0,001$) a F_{2a} ($r = 0,93$, $p < 0,001$). Výjimkou je formant F_{1i} ($r = 0,47$, $p < 0,05$), kde bylo dosaženo hodnoty korelace s manuální metodou pod 0,5, přesto se však průměrná frekvence lišila ve srovnání s ruční analýzou pouze o 8 Hz. Nicméně i tato hodnota korelačního koeficientu se statisticky liší od nuly. Navíc formantová frekvence F_{1i} není nej-

3. Výsledky

stěžejnější parametr pro určování rozdílů mezi KS a PN a nemá ani zásadní vliv na hodnoty parametrů *VSA* a *VAI*.

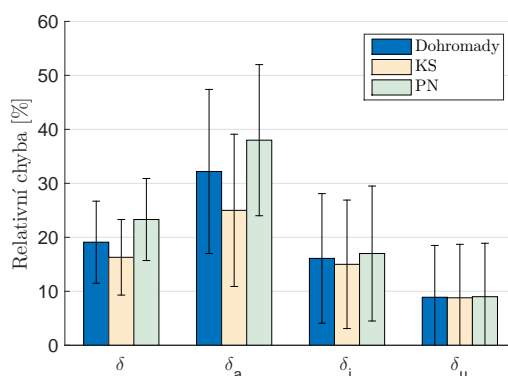
Na základě výsledků testu reliability bude pro implementaci automatických metod použit algoritmus *Praat*, jenž s výjimkou F_{1u} dosahoval nejvyšší spolehlivosti výpočtu všech formantů ze všech testovaných algoritmů.

	<i>Ghosh</i>			<i>Mustafa & Bruce</i>			<i>Praat</i>		
	$\bar{\Delta}$	<i>r</i>	<i>p</i>	$\bar{\Delta}$	<i>r</i>	<i>p</i>	$\bar{\Delta}$	<i>r</i>	<i>p</i>
F_{1a}	60	-0,09	0,73	-16	0,82	< 0,001	5	0,95	< 0,001
F_{2a}	-139	-0,44	0,06	86	0,40	0,097	4	0,93	< 0,001
F_{1i}	-220	0,27	0,29	-223	0,14	0,57	8	0,47	< 0,05
F_{2i}	397	-0,09	0,73	493	-0,19	0,45	31	0,83	< 0,001
F_{1u}	-100	0,30	0,23	-62	0,92	< 0,001	-13	0,80	< 0,001
F_{2u}	-636	0,30	0,22	-74	0,44	0,069	-84	0,79	< 0,001

Tabulka 16. Výsledky testu spolehlivosti algoritmů pro výpočet formantových frekvencí. $\bar{\Delta}$ [Hz] - průměrný rozdíl mezi výsledky ruční analýzy a výsledky algoritmu pro výpočet formantových frekvencí; *r* - Pearsonův korelační koeficient; *p* - *p*-hodnota.

3.2.2. Test rozpoznávače řeči

Na obr. 29 jsou uvedeny chyby rozpoznávače řeči pro samohlásky vyznačené v textu. Celková průměrná chyba rozpoznávače (δ), tedy v kolika procentech případů není správně rozpoznán hledaný foném, na skupinách KS a PN dohromady je 19 %. Hodnota průměrné chyby se pohybuje pod hranicí 24 %, která je uvedena u vybraného rozpoznávače řeči pro český jazyk. Při testování skupin KS a PN zvláště byla naměřena větší chyba u PN, která převyšovala 21 %. Tento fakt je způsoben především tím, že u skupiny PN byla změřena relativní chyba fonému *a* (δ_a) 38 % oproti 25 % u KS. U fonémů *i* a *u* byly rozdíly mezi oběma skupinami nepatrné. Ze samohláskových fonémů *a*, *i* a *u* byla změřena největší průměrná relativní chyba u fonému *a*, která dosahovala 32 %. U fonému *i* byla zjištěna chyba 16 % a u fonému *u* 9 %.

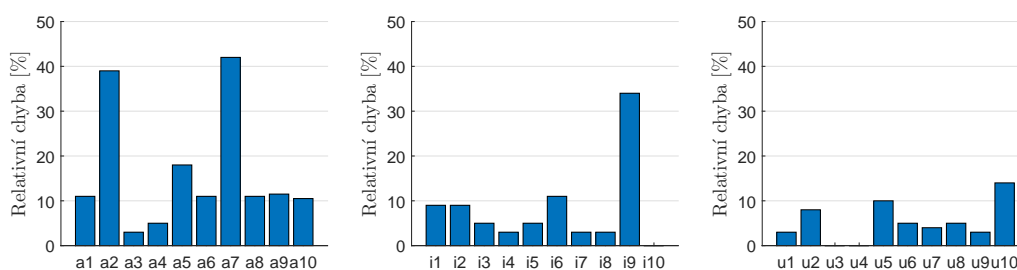


Obrázek 29. Relativní chyba rozpoznávače řeči při rozpoznávání samohlásek. δ - průměrná chyba pro samohlásky *a*, *i* a *u*; δ_a - průměrná chyba pro samohlásku *a*; δ_i - průměrná chyba pro samohlásku *i*; δ_u - průměrná chyba pro samohlásku *u*

Obr. 30 zobrazuje relativní chyby všech deseti samohlásek *a*, *i* a *u* vyznačených v textu. Je patrné, že rozložení chyb není rovnoměrné a některé samohlásky vyznačené v textu jsou

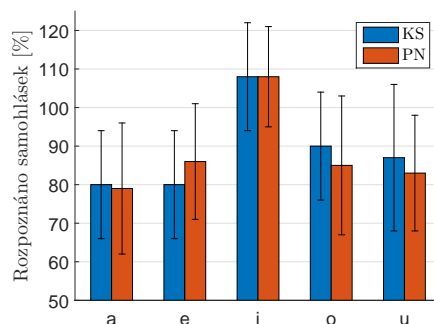
na chyby více náchylné. Nejvíce chyb bylo pozorováno u samohlásek *a2* a *a7*, které jsou vyznačeny ve slově „sazeničku“, respektive „sazenička“ po souhlásce *s*. Rozpoznávač ve většině případů místo fonému *a* označil úsek řeči jako *e* nebo jako *o*. Tato chyba je pravděpodobně způsobena přechodem mezi *s* a následným *z*, kde je důraz ve slově dán na slabiku *ze*. Kvůli tomu nemusí být slabika *sa* vyslovena dostatečně zřetelně. Nejvíce chyb u fonému *i* bylo změřeno na vyznačené samohlásce *i9*. Tato samohláska se nachází ve slovním spojení „čehosi jako“. Fonémy *i* a *j* mají podobné formantové frekvence a při čtení textu se z fonému *i* do fonému *j* plynule přechází. Rozpoznávač tak nemusí rozpoznat prvotní foném *i*, ale rozpozná až následný foném *j*. U fonému *u* je nejvyšší chyba u *u10*, které je vyznačeno v posledním slově textu. Někteří mluvčí už u posledního slova mohou mít tendenci snižovat intenzitu hlasu nebo slovo zadrmolit, což může vést k následným chybám rozpoznávače.

Další chyba rozpoznávače nastávala u samohlásky *u5*, která se nachází ve slovním spojení „půdu u jejich“. Rozpoznávač v několika případech považuje dvě po sobě jdoucí *u* za dlouhý foném *u*. Chyby také vznikaly ve slovních spojení, ve kterém je na konci prvního slova samohláska *a* a na začátku druhého slova samohláska *u* (např. „za užitečnou“). V těchto případech se stávalo, že rozpoznávač označil přechod mezi slovy difonem *au*.



Obrázek 30. Relativní chyba rozpoznávače v detekci vyznačených samohlásek v úloze čteného textu.

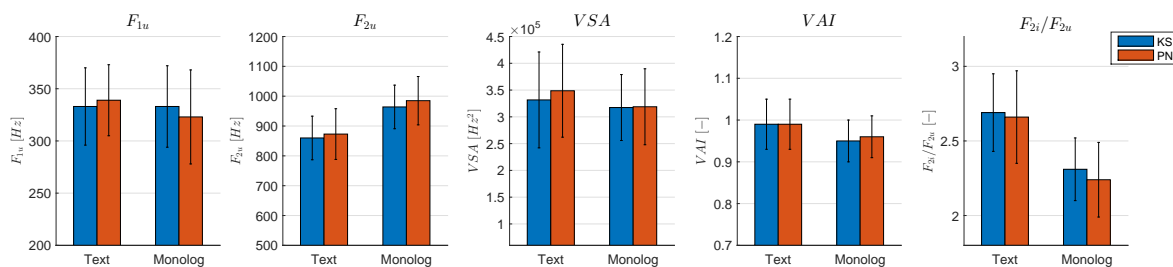
Obr. 31 znázorňuje poměr počtu jednotlivých fonémů detekovaných rozpoznávačem řeči vůči počtu fonémů, které se v textu skutečně vyskytují. Zajímavé je, že rozpoznávač řeči detekuje více fonému *i* než se v textu skutečně vyskytuje. Tento fakt je způsoben zejména chybou rozpoznávače, kdy je některá ze samohlásek zaměněna za samohlásku *i*. Ostatních fonémů je rozpoznáno 80 až 90 % z celkového počtu vyskytujících se v textu. Rozpoznávač detekuje podobné procentuální zastoupení samohlásek u zdravých a nemocných jedinců. Nejvyšší rozdíl je patrný u fonému *e*, kde je u skupiny PN detekováno o 5 % více fonémů *e* než u KS. Naopak u skupiny KS je rozpoznáno o 4 % více fonémů *o* a o 3 % více fonémů *u*.



Obrázek 31. Poměr zachycených samohláskových fonémů vůči celkovému počtu fonémů v textu.

3.2.3. Automatická metoda bez rozpoznávače řeči

Porovnání středních hodnot získané automatickou metodou bez rozpoznávače řeči pro četný text a vedený monolog je zahrnuto v obr. 32. Z výsledků je patrné, že ani jeden z parametrů F_{1u} , F_{2u} , VSA , VAI a F_{2i}/F_{2u} neprokázal statisticky významnou odlišnost mezi skupinami KS a PN. Ani následná analýza mužů a žen zvlášť neshledala ani jeden parametr statisticky signifikantním. V porovnání s ruční analýzou, jejíž výsledky jsou uvedeny v kap. 3.1.3, je patrné, že zejména parametr VSA je velmi nadhodnocen při použití automatické metody bez rozpoznávače. Navíc je průměrná hodnota VSA u metody bez rozpoznávače u skupiny PN vyšší než u KS, což nekoresponduje s ruční analýzou a s předpokladem centralizace formantů. Pro tento typ úloh tedy není metoda bez rozpoznávače řeči vhodnou metodou.



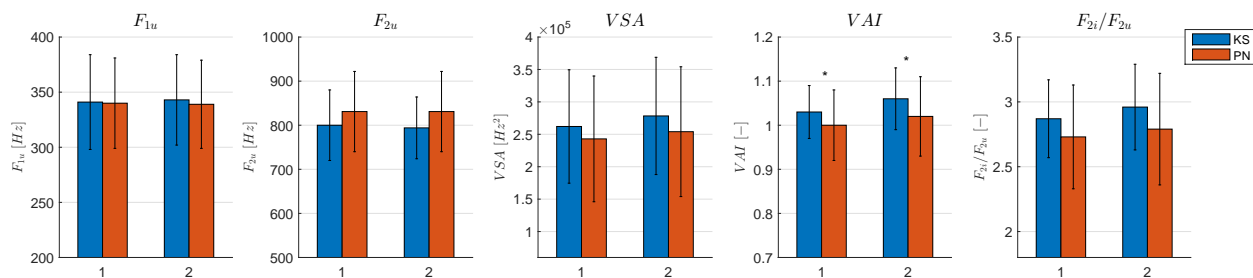
Obrázek 32. Porovnání středních hodnot parametrů mezi KS a PN z výsledků automatické metody bez rozpoznávače řeči pro text a monolog u mužů a žen dohromady.

3.2.4. Automatická metoda s rozpoznávačem řeči

U automatické metody s rozpoznávačem řeči byly testovány různé způsoby výběru samohlásek s ohledem na délku řečového signálu (viz tab. 13). U úlohy textu byly vybírány všechny samohlásky detekovány rozpoznávačem řeči a pak také všechny samohlásky s tím, že byly vyloučeny krátké fonémy i . Vzhledem k malé délce řečového signálu nedetekoval rozpoznávač dostatečný počet krátkých fonémů u , proto nebylo možné vyloučit dlouhé fonémy u . Řečový signál u úlohy monologu již nabýval větší délky řečového signálu, proto mohlo být vybírání samohlásek specifitější. Celkem proběhly 4 způsoby výběru samohlásek: všechny samohlásky, všechny samohlásky s vyloučením fonémů u ; a i , samohlásky po neznělých souhláskách a samohlásky po neznělých souhláskách s vyloučením fonému i . U některých řečových signálů se nemuselo podařit detekovat dostatečný počet požadovaných samohlásek. Pro tyto signály nebyly počítány formantové frekvence a nebyly tedy do porovnání mezi KS a PN zahrnuty.

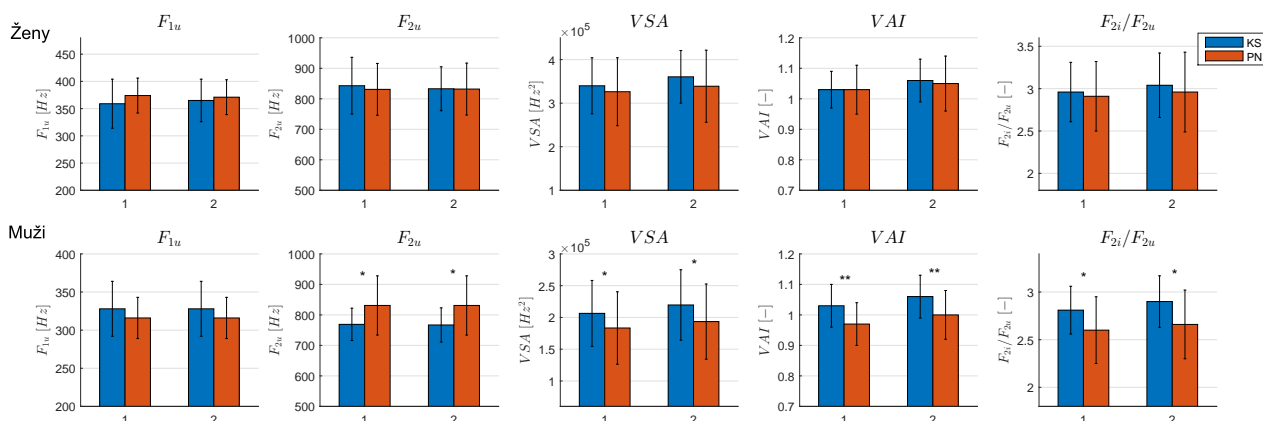
Výsledky pro oba uvedené způsoby výběru samohlásek u úlohy textu jsou zobrazeny na obr. 33. U obou způsobů výběru samohlásek byl shledán jako signifikantní parametr pro rozlišení KS a PN parametr VAI ($p < 0,05$). U druhého způsobu vybírání samohlásek bylo dosaženo o něco lepších výsledků. Například u parametru F_{2i}/F_{2u} se zvětšil rozdíl mezi oběma skupinami o 0,03. Tento rozdíl je zřejmě způsoben neovlivňováním výsledků krátkým fonémem i , kde jsou hodnoty F_{1i} a F_{2i} vypočítány pouze z dlouhého fonému i . Nicméně vzhledem ke shodě textů nejsou rozdíly natolik zřejmé, jako tomu je u úlohy monologu. U obou způsobů výběru bylo dosaženo obdobných výsledků jako u ruční analýzy, jejíž výsledky jsou uvedeny v kap. 3.1.3. V porovnání s ruční analýzou je patrné, že podobně jako metoda bez rozpoznávače řeči i metoda s rozpoznávačem nadhodnocuje parametr VSA . Nadhodnoceny jsou i parametry VAI a F_{2i}/F_{2u} . Nicméně ve srovnání s metodou bez rozpoznávače jsou rozdíly mezi skupinami KS a PN zachovány.

Analýza mužů a žen zvlášť pro úlohu textu, jejíž výsledky jsou zahrnuty v obr. 34, ukázala významné rozdíly jen mezi zdravými muži a muži s PN. U žen nebyly shledány stejně jako



Obrázek 33. Porovnání středních hodnot parametrů mezi KS a PN z výsledků automatické metody s rozpoznávačem řeči pro text u mužů a žen dohromady. Osa x označuje jednotlivé výběry: výběr všech fonémů (1), výběr všech fonémů kromě i (2). Hvězdičky indikují rozdíly mezi KS a PN: $*p < 0,05$.

v ruční analýze žádné statisticky významné parametry. U mužů byly vyhodnoceny jako signifikantní parametry pro odlišení KS a PN parametry F_{2u} ($p < 0,05$), VSA ($p < 0,05$), VAI ($p < 0,01$) a F_{2i}/F_{2u} ($p < 0,05$). Při výběru samohlásek bez fonému i jsou patrnější vyšší rozdíly mezi oběma skupinami oproti výběru všech samohlásek. Například rozdíl parametru F_{2i}/F_{2u} je u mužů i u žen vyšší o 0,03 a rozdíl u parametru VSA je u mužů vyšší o více než 3000 Hz^2 a u žen dokonce o necelých 8000 Hz^2 při vyloučení krátkého fonému i z výběru.

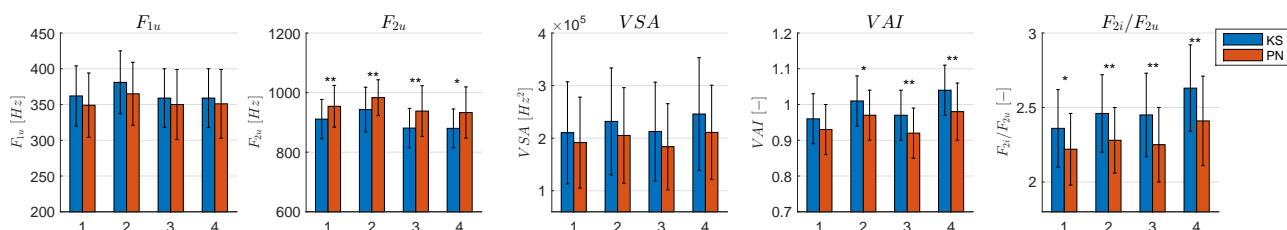


Obrázek 34. Porovnání středních hodnot parametrů mezi KS a PN z výsledků automatické metody s rozpoznávačem řeči pro text zvlášť u žen (horní obrázky) a u mužů (spodní obrázky). Osa x označuje jednotlivé výběry: výběr všech fonémů (1), výběr všech fonémů kromě i (2). Hvězdičky indikují rozdíly mezi KS a PN: $*p < 0,05$, $**p < 0,01$.

U úlohy monologu byly testovány 4 způsoby výběru samohlásek a výsledky shrnuje obr. 35. Pro všechny 4 výběry byly jako statisticky významné parametry vyhodnoceny parametry F_{2u} a F_{2i}/F_{2u} . Ze všech 4 způsobů výběru udával nejhorší výsledky výběr všech samohlásek, kde byly jako statisticky významné parametry shledány pouze zmíněné dva. U ostatních výběrů byl jako signifikantní parametr určen navíc parametr VAI . Při porovnání výběrů, kde byly vybírány všechny samohlásky a kde byly ze všech vybraných samohlásek vyloučeny fonémy i a u ., je patrné, jaký vliv na výsledek má rozlišování toho, zda jsou vybírány dlouhé a krátké fonémy dohromady či pouze krátké nebo dlouhé fonémy. U všech uvedených parametrů, kde byly vyloučeny fonémy i a u : došlo ke zvýšení průměrného rozdílu mezi KS a PN. Nejvyšší rozdíl byl patrný zejména u parametru VAI , který se stal signifikantním pro rozlišení obou skupin a parametr F_{2i}/F_{2u} , jehož rozdíl se mezi KS a PN navýšil o 0,04. Lepšího výsledku však bylo dosaženo pro výběr samohlásek po neznělých souhláskách, kde byly parametry F_{2u} , VAI

3. Výsledky

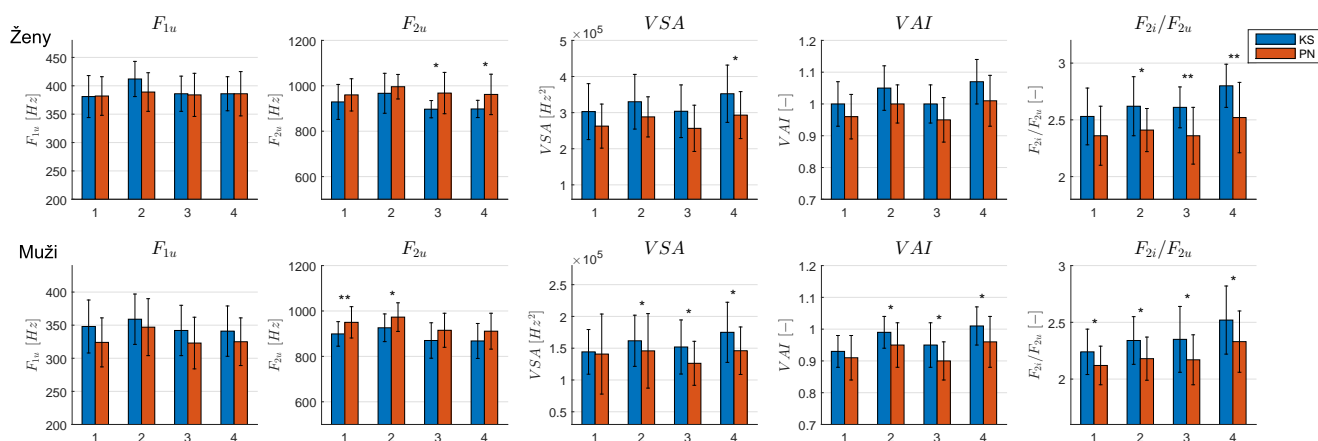
a F_{2i}/F_{2u} shledány signifikantními s $p < 0,01$. Podobně dobrých výsledků bylo dosaženo i při výběru samohlásek po neznělých souhláskách, ze kterého byly vyloučeny krátké fonémy i . Zde bylo dokonce u parametrů VAI a F_{2i}/F_{2u} dosaženo vyšších rozdílů mezi KS a PN. Vzhledem k nedostatečné délce některých monologů nebyl u některých nahrávek zachycen dostatečný počet požadovaných samohlásek, proto byly některé subjekty vyřazeny.



Obrázek 35. Porovnání středních hodnot parametrů mezi KS a PN z výsledků automatické metody s rozpoznávačem řeči pro monolog u mužů a žen dohromady. Osa x označuje jednotlivé výběry: výběr všech fonémů (1), výběr všech fonémů kromě i a u : (2), výběr všech fonémů po neznělých fonémech (3), výběr fonémů po neznělých fonémech s vyloučením i (4). Hvězdičky indikují rozdíly mezi KS a PN: * $p < 0,05$, ** $p < 0,01$.

Analýza rozdílů mezi KS a PN u úlohy monologu pro muže a ženy zvláště uvádí obr. 36. Při výběru všech samohlásek jsou statisticky významné rozdíly mezi KS a PN evidovány pouze u mužů a to u parametrů F_{2u} ($p < 0,01$) a F_{2i}/F_{2u} ($p < 0,05$). Při vyloučení fonémů i a u : je signifikantní parametr pro odlišení KS a PN u žen F_{2i}/F_{2u} ($p < 0,05$). U mužů jsou při stejném výběru samohlásek shledány statisticky významnými parametry F_{2u} ($p < 0,05$), VSA ($p < 0,05$), VAI ($p < 0,05$) a F_{2i}/F_{2u} ($p < 0,05$). Tato skutečnost také dokládá, jak důležité je rozlišovat mezi krátkými a dlouhými fonémy. Při výběru všech samohlásek po neznělých souhláskách byly u žen vyhodnoceny za signifikantní parametry F_{2u} ($p < 0,05$) a F_{2i}/F_{2u} ($p < 0,01$) a u mužů parametry VSA ($p < 0,05$), VAI ($p < 0,05$) a F_{2i}/F_{2u} ($p < 0,05$). Při vyloučení fonému i po neznělých souhláskách bylo dosaženo obdobných výsledků jako v předchozím případě s tím rozdílem, že za statisticky signifikantní parametr byl u žen shledán i parametr VSA ($p < 0,05$). Navíc při vyloučení fonému i byly evidovány u některých parametrů vyšší rozdíly oproti výběru všech samohlásek po neznělých souhláskách. Například u žen se zvýšil rozdíl mezi KS a PN u parametru VSA o téměř $17000 Hz^2$, u VAI o $0,02$ a u F_{2i}/F_{2u} o $0,03$. U mužů byly evidovány rozdíly u parametru VSA o $3300 Hz^2$ a u F_{2i}/F_{2u} o $0,01$, což značí, že se rozdíly projeví zejména mezi zdravými a nemocnými ženami.

Vzhledem k výsledkům obou automatických je patrné, že mnohem lepší výsledky udává metoda s rozpoznávačem řeči, proto budou v další analýze v této práci použity pouze výsledky dosažené metodou s rozpoznávačem. V práci budou analyzovány výsledky s výběrem všech fonémů kromě fonému i u úlohy textu, kde byly evidovány vyšší rozdíly mezi KS a PN oproti výsledkům při výběru všech samohlásek, a výsledky s výběrem fonémů po neznělých souhláskách s vyloučením fonému i u úlohy monologu, kde byly hlavně pro analýzu mužů a žen zvláště evidovány vyšší rozdíly mezi KS a PN oproti ostatním výběrům samohlásek.



Obrázek 36. Porovnání středních hodnot parametrů mezi KS a PN z výsledků automatické metody s rozpoznávačem řeči pro monolog zvlášť u žen (horní obrázky) a u mužů (spodní obrázky). Osa x označuje jednotlivé výběry: výběr všech fonémů (1), výběr všech fonémů kromě i a u : (2), výběr všech fonémů po neznělých fonémech (3), výběr fonémů po neznělých fonémech s vyloučením i (4). Hvězdičky indikují rozdíly mezi KS a PN: * $p < 0,05$, ** $p < 0,01$.

3.2.5. Porovnání ruční analýzy s automatickou metodou

Parametry, které byly za statisticky významné shledány v ruční analýze se do jisté míry shodují se signifikantními parametry, které vzešly z automatické metody s rozpoznávačem řeči. Tab. 17 shrnuje signifikantní parametry, kterých bylo dosaženo u ruční analýzy a u automatické metody s rozpoznávačem řeči. Jak u ruční analýzy, tak i u automatické metody byl nejčastěji vyhodnocen za statisticky významný parametr poměr F_{2i}/F_{2u} . U úlohy textu u žen nebyly shledány žádné parametry signifikantními. Naopak u stejné úlohy u mužů vzešly z ruční analýzy statisticky signifikantní parametry F_{1u} a F_{2i}/F_{2u} a z automatické metody parametry F_{2u} , VSA , VAI a F_{2i}/F_{2u} . Signifikantní parametry u úlohy monologu při analýze mužů a žen dohromady se u ruční analýzy a automatické metody shodovaly. Rozdíly byly patrné u analýzy mužů, kde v automatické metodě nebyl shledán signifikantním ani jeden z formantů fonému u , ale pouze odvozené parametry VSA , VAI a F_{2i}/F_{2u} . U žen pak byl vyhodnocen statisticky signifikantním parametrem VSA místo VAI v porovnání automatické metody s ruční analýzou, jinak se zbývající parametry shodovaly. Signifikantní parametry vypočítané automatickou metodou se navíc poměrně dobře shodují s ruční analýzou i z hlediska hladiny významnosti. Jediná odlišnost je patrná u úlohy monologu u analýzy obou pohlaví dohromady, kde bylo dosaženo vyšší statistické významnosti u formantu F_{2u} u ruční analýzy.

Výsledky automatických metod byly porovnány s ruční analýzou na základě testu spolehlivosti. Výsledky, které nekorespondují s ruční analýzou, nemusí být automaticky špatně. U ruční analýzy byly vybírány specifické samohlásky podle daných pravidel, které mohly být u manuálních metod aplikovány pouze z části. U automatických metod nebylo například možné vyloučit slova, která nebyla vjemově normální. Dále mohlo být u úlohy monologu z jednoho slova vybráno více různých samohlásek. Navíc byly u automatické metody vyloučeny krátké fonémy i . Z těchto důvodů nemusí výsledky s ruční analýzou plně korelovat.

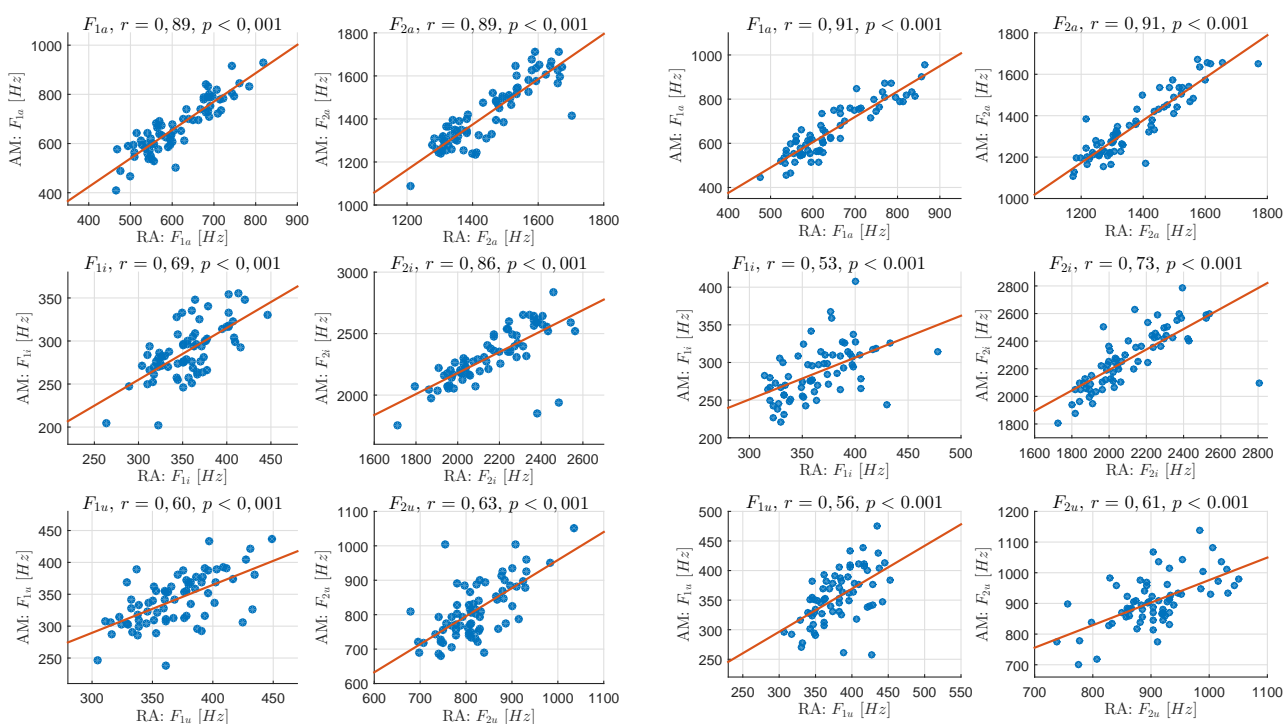
Porovnání automatické metody používající rozpoznávač řeči s ruční analýzou je uvedeno na obr. 37 (a) pro úlohu textu a na obr. 37 (b) pro úlohu monologu. Pro obě úlohy korespondují s ruční analýzou nejlépe formantové frekvence samohlásky a . Nejhůře pak s ruční analýzou korelují formanty F_{1i} , F_{1u} a F_{2u} . U všech frekvencí se koeficient korelace pohybuje nad hodnotou 0,5, který se statisticky významně liší od nuly ($p < 0,001$). Tyto výsledky mohou do jisté míry souviset i s přesností vybraného algoritmu pro výpočet formantových frekvencí, jehož po-

3. Výsledky

	Ruční analýza			Automatická metoda		
	Dohromady	Ženy	Muži	Dohromady	Ženy	Muži
Text	F_{2i}/F_{2u} *		F_{1u} ** F_{2i}/F_{2u} *	VAI *		F_{2u} * VSA * VAI ** F_{2i}/F_{2u} *
Monolog	F_{2u} *** VAI ** F_{2i}/F_{2u} **	F_{2u} * VAI * F_{2i}/F_{2u} **	F_{1u} * F_{2u} * VSA * VAI * F_{2i}/F_{2u} *	F_{2u} * VAI ** F_{2i}/F_{2u} **	F_{2u} * VSA * F_{2i}/F_{2u} **	VSA * VAI * F_{2i}/F_{2u} *

Tabulka 17. Statisticky signifikantní parametry vycházející z ruční analýzy a z automatické metody u úlohy textu a monologu. Hvězdičky indikují rozdíly mezi KS a PN: * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$.

rovnání s ruční analýzou bylo popsáno v kap. 3.2.1. Nejvyšší rozdíly průměrných hodnot mezi automatickou metodou a ruční analýzou byly zjištěny především u formantů F_{1i} a F_{2i} . Tyto rozdíly jsou způsobené tím, že jsou u automatické metody z výběru samohlásek vyloučeny krátké fonémy i , což vede ke snížení frekvence F_{1i} a zvýšení frekvence F_{2i} .

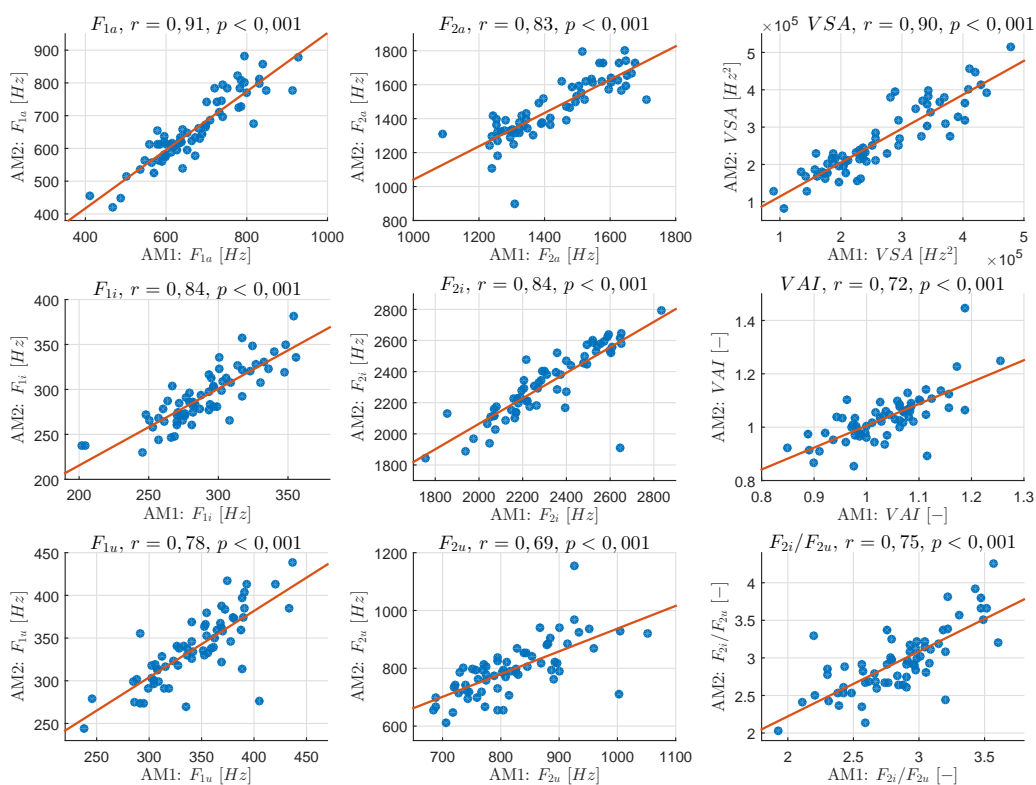


Obrázek 37. Porovnání ruční analýzy a automatické metody s rozpoznávacím řeči. Osa x odpovídá výsledkům ruční analýzy (RA), osa y odpovídá výsledkům automatické metody (AM). r - Pearsonův korelační koeficient; p - p -hodnota.

3.2.6. Opakovatelnost automatické metody

Opakovatelnost automatické metody byla vyhodnocena pro úlohu čteného textu, kde pro většinu měřených subjektů byly pořizeny 2 řečové záznamy. Výsledek testu opakovatelnosti automatické metody je uveden na obr. 38. Koeficienty korelace se u všech formantových frekvencí významně liší od nuly ($p < 0,001$). To značí vysokou spolehlivost opakovatelnosti automatické metody. Nejvyšší hodnoty korelace bylo dosaženo u formantových frekvencí fonémů *a* a *i*, u kterých koeficient korelace dosahoval hodnoty nad 0,8. U fonému *u* bylo dosaženo koeficientu korelace 0,78 pro první formant a 0,69 pro druhý formant. Z odvozených parametrů bylo nejvyšší spolehlivosti dosaženo u *VSA* ($r = 0,90$, $p < 0,001$). U parametrů *VAI* a F_{2i}/F_{2u} byl koeficient korelace vyšší než 0,7, což také značí poměrně vysokou spolehlivost.

Podobně jako pro text by mohla být testována opakovatelnost pro spontánní monolog. Test by probíhal tak, že by monolog byl rozdělen na 2 stejně dlouhé části a pro každou část by byly spočítány formantové frekvence. Bohužel většina nahrávek monologů neměla dostatečnou délku na to, aby mohl být rozdělen a mohl tak proběhnout test opakovatelnosti.

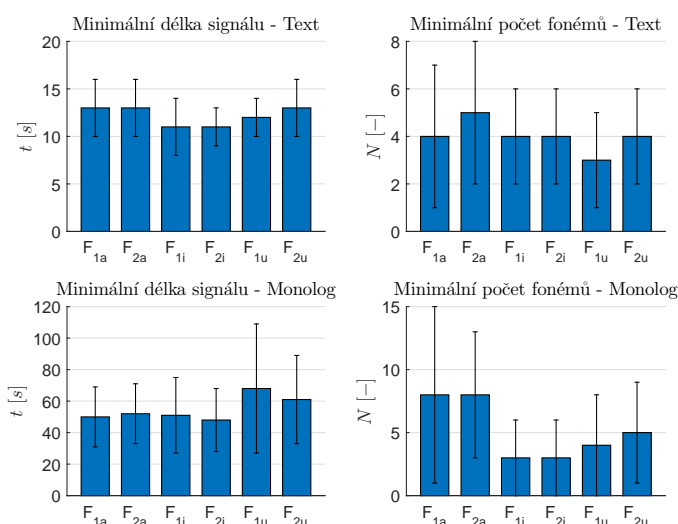


Obrázek 38. Test opakovatelnosti automatické metody na úloze textu. Osa x odpovídá výsledkům automatické hodnoty prvního čtení (AM1), osa y odpovídá výsledkům automatické hodnoty druhého čtení (AM2). r - Pearsonův korelační koeficient; p - p-hodnota.

3.2.7. Minimální délka řečového signálu

Obr. 39 uvádí minimální délku řečového signálu a minimální počet fonémů, který je potřeba k přesnému výpočtu formantových frekvencí. U úlohy čteného textu, kde jsou vybírány všechny fonémy s výjimkou fonému i , je pro spolehlivý výpočet jednotlivých formantů potřeba minimální délka řečového signálu 11 – 13 s, což odpovídá 3 – 5 jednotlivým detekovaným fonémům.

U úlohy monologu je potřeba delšího řečového signálu z toho důvodu, že jsou vybírány pouze fonémy vyslovené po neznělých samohláskách. Průměrná minimální délka pro spolehlivý výpočet formantových frekvencí fonémů a a i je kolem 50 s. U fonému u je to 60 – 70 s, což je dáno menší relativní četností fonému u v mluvené řeči. V úloze monologu byla spočítána vyšší hodnota směrodatné odchylky oproti textu. Je to z důvodu, že každý mluvčí mluví jinak rychle a také volí jiná slova, proto může být dostatečný počet fonémů detekován u některých mluvčích dříve a u některých po delším čase. Pro spolehlivý výpočet formantů fonému a je potřeba průměrně 8 detekovaných fonémů. U formantových frekvencí samohlásek i a u je potřeba průměrně 3 – 5 rozpoznaných fonémů.



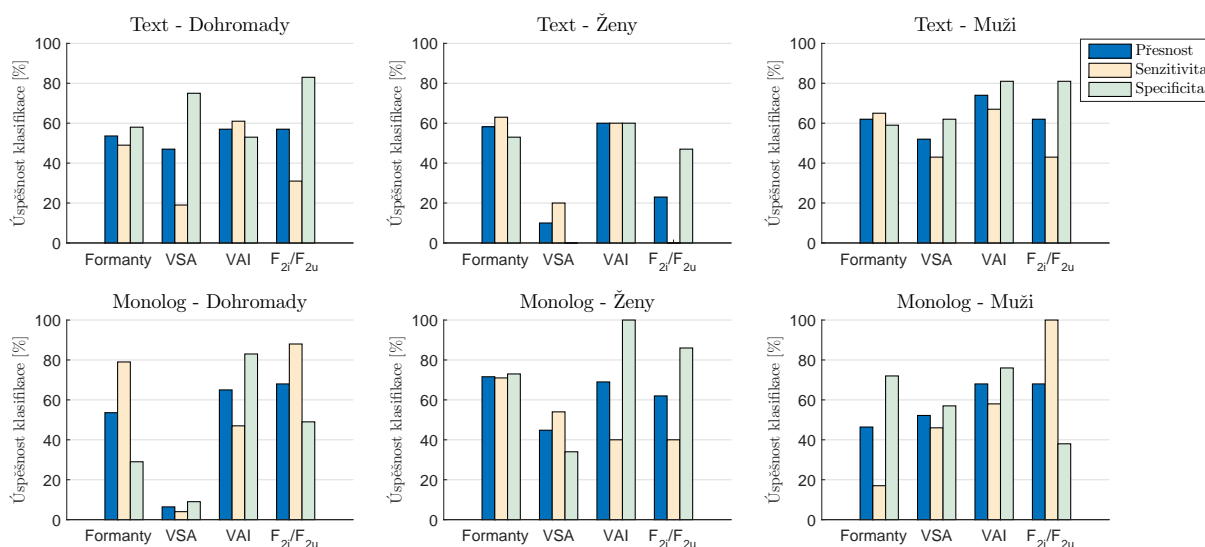
Obrázek 39. Minimální délka signálu a minimální počet zachycených fonémů pro přesný výpočet formantových frekvencí jednotlivých samohlásek u úlohy textu a monologu.

3.3. Klasifikace

Úspěšnost klasifikace pro všechny formanty dohromady a pro parametry VSA , VAI a F_{2i}/F_{2u} vypočítané automatickou metodou s rozpoznávačem u úlohy textu a monologu je zahrnuta v obr. 40. V obrázku jsou uvedeny úspěšnosti klasifikace pro muže a ženy dohromady a také zvlášť pro obě pohlaví. U úlohy textu při nerozlišování pohlaví bylo dosaženo maximální přesnosti klasifikace necelých 60 % a byla pozorována u parametrů VAI ($ACC = 57 \%$, $SE = 61 \%$, $SP = 53 \%$) a F_{2i}/F_{2u} ($ACC = 57 \%$, $SE = 83 \%$, $SP = 31 \%$). Při klasifikaci mužů a žen zvlášť bylo patrné zlepšení zejména u mužů, kde přesnost klasifikace u VAI dosahovala 74 % ($SE = 67 \%$, $SP = 81 \%$). Nad hranici 60 % celkové přesnosti klasifikace se dostaly také formanty ($ACC = 62 \%$, $SE = 65 \%$, $SP = 59 \%$) a parametr F_{2i}/F_{2u} ($ACC = 62 \%$, $SE = 43 \%$, $SP = 81 \%$). Vzhledem k velmi nízké hodnotě senzitivity se však parametr F_{2i}/F_{2u} nejvíce jeví jako nejvhodnější pro odlišení obou skupin. U úlohy textu u žen bylo nejvyšší úspěšnosti stejné

jako u mužů dosaženo u parametru VAI ($ACC = 65 \%$, $SE = 47 \%$, $SP = 83 \%$) a u formantů ($ACC = 58 \%$, $SE = 63 \%$, $SP = 53 \%$).

U úlohy monologu při klasifikaci obou pohlaví dohromady byla nejvyšší úspěšnost podobně jako v úloze textu pozorována u parametru VAI ($ACC = 65 \%$, $SE = 47 \%$, $SP = 83 \%$) a F_{2i}/F_{2u} ($ACC = 68 \%$, $SE = 88 \%$, $SP = 49 \%$). U obou parametrů však dosahovala hodnota senzitivity či specifity pod hranici 50 %. Klasifikace žen a mužů zvlášť přinesla zlepšení především u žen. Při klasifikaci všech formantů dohromady se dostala přesnost klasifikace, senzitivita i specifita nad hodnotu 70 % ($ACC = 72 \%$, $SE = 71 \%$, $SP = 73 \%$). Vzhledem k vyrovnanosti senzitivity a specifity se formanty jeví jako vhodné pro rozlišování zdravých a nemocných žen. U mužů byly nejvyšší přesnosti klasifikace patrné u parametrů VAI ($ACC = 68 \%$, $SE = 58 \%$, $SP = 76 \%$) a F_{2i}/F_{2u} ($ACC = 68 \%$, $SE = 100 \%$, $SP = 38 \%$). U parametru F_{2i}/F_{2u} je opět patrná silná nevyváženost hodnot senzitivity a specifity, proto se jeví jako vhodnější parametr pro rozlišení zdravých a nemocných mužů VAI .



Obrázek 40. Úspěšnost klasifikace na výsledcích automatické metody s rozpoznávačem řeči pro úlohu textu a monologu.

4. Diskuze

Cílem této práce bylo navrhnout algoritmus pro automatické hodnocení poruch artikulace samohlásek u Parkinsonovy nemoci, která měla být porovnána s výsledky ruční analýzy. Pro hodnocení poruch artikulace samohlásek se vycházelo z hodnot první a druhé formantové frekvence rohových samohlásek a od nich odvozených parametrů VSA , VAI a F_{2i}/F_{2u} . Testování proběhlo na databázi 72 subjektů. Databáze obsahovala data od 36 zdravých jedinců tvořící kontrolní skupinu a 36 pacientů s PN. U subjektů byly pořízeny zvukové záznamy čteného textu a spontánního monologu, z kterých se pro hodnocení poruch artikulace vycházelo.

Na databázi nejprve proběhla statistická analýza výsledků ručně naměřených dat. Nejdříve proběhla analýza mužů a žen, která prokázala signifikantní rozdíly mezi hodnotami u většiny parametrů u obou řečových úloh. Z toho bylo usouzeno, že nelze nahlížet na zdravé a nemocné lidi dohromady, ale některé rozdíly mezi zdravými a nemocnými lidmi mohou být patrné až po analýze mužů a žen zvlášť. Proto proběhla analýza pro posouzení rozdílů mezi zdravými lidmi a pacienty jednak s muži a ženami dohromady a jednak zvlášť pro obě pohlaví. Při analýze mužů a žen dohromady byly zjištěny celkem 3 statisticky významné parametry: F_{2i}/F_{2u} u obou řečových úloh a dále F_{2u} a VAI u úlohy monologu. U analýzy mužů zvlášť byly signifikantní rozdíly pozorovány také u parametrů F_{1u} a VSA . U žen byly shledány signifikantní parametry pouze u úlohy monologu. Celkově se jevila jako citlivější úloha pro rozlišení zdravých a nemocných mluvčích úloha monologu, u které bylo zjištěno více statisticky významných parametrů pro odlišení obou skupin. Výsledky ruční analýzy korespondují s výsledky studií, které byly v práci zmíněny, z hlediska rozdílů formantů u mužů a žen, z hlediska signifikantních parametrů a také z hlediska citlivosti úloh pro odlišení zdravých lidí a pacientů s PN. (viz např. [1, 6, 49, 52, 53])

U výsledků ruční analýzy byl dále zkoumán vliv délky fonémů na formantovou frekvenci u samohlásek i a u . Pro formantové frekvence dlouhých a krátkých samohlásek, které jsou vyznačené v zadaném textu, proběhl test středních hodnot, který prokázal statisticky významný rozdíl mezi formanty dlouhých a krátkých fonémů. Krátké fonémy mají tendenci se centralizovat, zatímco dlouhé se decentralizují. U krátkého fonému i byla zjištěna v průměru o 41 Hz vyšší hodnota prvního formantu a o 217 Hz nižší hodnota druhého formantu než u dlouhého fonému i . To odpovídá zhruba 10 % průměrné hodnoty formantů samohlásky i . Rozdíly mezi krátkými a dlouhými formanty samohlásky u odpovídají 5 % průměrné hodnoty. U úlohy čtení, kde mají mluvčí zadaný stejný text a jsou vybírány stejné samohlásky nemusí nastat problém ve vybírání dlouhých a krátkých samohlásek dohromady. Problém však může nastat u úlohy monologu, kde může být pro různé jedince vybrán různý počet krátkých a dlouhých fonémů. Tyto rozdíly mohou zastínit nebo naopak zvýraznit celkové rozdíly mezi skupinami zdravých a nemocných lidí. Proto se jeví jako vhodné zejména u samohlásky i vybírat fonémy pouze jednoho typu.

Ještě před samotnou implementací automatické metody bylo nutné vybrat algoritmus pro automatický výpočet formantových frekvencí. Ze všech tří testovaných algoritmů udával nejvyšší spolehlivost algoritmus implementovaný v softwaru Praat. Kromě první formantové frekvence fonému i , dosahoval koeficient korelace hodnot vyšších než 0,75. U zbývajících dvou testovaných algoritmů byly patrné velké nedostatky při výpočtu formantů.

Při návrhu automatické metody bylo využito dvou přístupů: bez rozpoznávače řeči a s využitím rozpoznávače řeči pro segmentaci fonémů. Přístup bez rozpoznávače řeči pracuje s celým

řečovým signálem tedy i se souhláskami či pauzami. Pro úlohy čteného textu a spontánního monologu se tato metoda neosvědčila. Vzhledem k vysoké koncentraci souhlásek nevynikly shluky odpovídající formantovým frekvencím jednotlivých samohlásek. To se projevilo ve výsledcích, které neodpovídaly výsledkům ruční analýzy ani předpokladu centralizace fonémů u osob s PN. Například u parametru VSA , který by měl vzhledem k centralizaci fonémů u PN nabývat vyšší hodnoty u zdravých mluvčích oproti nemocným, byla vypočítána vyšší průměrná hodnota právě u pacientů s PN. Ani jeden z parametrů u metody bez rozpoznávače nebyl shledán statisticky významným pro odlišení zdravých lidí od nemocných. Tato metoda však může mít své využití v jiných typech úloh. Metoda se může uplatnit například v úkolu, kde jsou opakovaně artikulovány pouze samohlásky. Díky tomu, že v metodě nedochází ke zprůměrování žádných časových úseků řeči a pracuje se s celým průběhem formantových frekvencí v čase, mohou vznikat důležité shluky samohlásek v trochu jiných částech vokálního prostoru než je tomu u úlohy monologu či čtení, kde jsou pro jednotlivé časové úseky, v nichž se nachází samohlásky, hodnoty formantů zprůměrovány.

Metoda s využitím rozpoznávače řeči bere v úvahu pouze části řečového signálu, v nichž se vyskytují požadované samohlásky. U rozpoznávače řeči byly otestovány různé způsoby výběru samohlásek. U úlohy čtení bylo nejlepších výsledků dosaženo pro výběr všech samohlásek kromě krátkého fonému i . Ze statistické analýzy mužů a žen dohromady vzešel signifikantní parametr VAI . Stejně jako u ruční analýzy ani u automatické metody nebyly u úlohy čtení shledány žádné signifikantní parametry zvláště u žen. U mužů byly naopak vyhodnoceny za statisticky významné hned 4 parametry: F_{2u} , VSA , VAI a F_{2i}/F_{2u} . To naznačuje, že se nemocné ženy dokáží svojí řečí u úlohy čtení přiblížit zdravým ženám. Naopak deficity nemocných mužů jsou u této úlohy zřejmé. U úlohy spontánního monologu, kde byly vybírány samohlásky po neznelých souhláskách s vyloučením krátkého i , jsou patrné rozdíly mezi zdravými a nemocnými u obou pohlaví. U žen byly shledány signifikantními parametry F_{2u} , VSA a F_{2i}/F_{2u} a u mužů VAI , VSA a F_{2i}/F_{2u} . Z výsledků je patrné, že největší rozdíly mezi parametry zdravých a nemocných osob byly patrné zejména u odvozených parametrů VAI a F_{2i}/F_{2u} a dále u druhé formantové frekvence samohlásky u . U parametr VSA , jehož hodnota je velmi závislá na pohlaví, se projeví rozdíly mezi zdravými a nemocnými pouze u analýzy mužů a žen zvláště. Výsledky automatické metody se co do výčtu signifikantních parametrů i co do statistické významnosti jednotlivých parametrů velmi dobře shodují s ruční analýzou, ve které byly vyhodnoceny jako nejcitlivější pro odlišení zdravých a nemocných také především parametry F_{2u} , VAI a F_{2i}/F_{2u} .

Výsledky automatické metody s rozpoznávačem řeči byly srovnávány s manuální metodou pomocí testu spolehlivosti. Pearsonův korelační koeficient se u všech formantů u obou úloh statisticky lišil od nuly. Nejvyšší hodnota korelace byla pozorována u F_{1a} a F_{2a} , která se pohybovala kolem 0,9. Naopak nejnižší spolehlivost s ruční analýzou byla evidována u formantů samohlásky u , kde koeficient korelace dosahoval hodnot kolem 0,6. Vzhledem k tomu, že u automatické metody byly samohlásky vybírány podle jiných podmínek než u ruční analýzy, slouží výsledky testu reliability pouze pro orientaci, jak moc jsou výsledky obou metod podobné. Nelze z nich však usuzovat celkovou spolehlivost automatické metody.

Automatická metoda byla na úloze čteného textu testována z hlediska opakovatelnosti. Test reliability prokázal nejvyšší spolehlivost výpočtu formantů samohlásek a a i s koeficientem korelace nad 0,8. Formantové frekvence fonému u nabývají koeficientu korelace 0,69 u prvního formantu a 0,78 u druhého formantu a také se signifikantně liší od nuly. U parametrů VAI a F_{2i}/F_{2u} bylo dosaženo koeficientu korelace přes 0,7. U parametr VSA dosáhla hodnota korelace dokonce 0,9. Výsledky testu značí vysokou spolehlivost opakovatelnosti automatické metody na úloze čteného textu.

U automatické metody s rozpoznávačem řeči byla zjištěna minimální délka řečového signálu a počet fonémů pro spolehlivý výpočet jednotlivých formantových frekvencí. Pro automatickou metodu nastavenou na čtený text (tedy při vybírání všech fonémů s výjimkou krátkého i)

byla změřena minimální délka řečového signálu pro přesný výpočet formantů všech fonémů od 11 do 13 s. Celkem bylo potřeba detekovat 3 – 5 jednotlivých fonémů pro přesný výpočet formantů. U úlohy monologu, kde byly vybírány samohlásky po neznělých souhláskách s vyloučením fonému *i*, byla zjištěna různá délka signálu a různý počet potřebných fonémů pro různé samohlásky. Pro samohlásky *a* a *i* je potřeba signál o minimální délce 50 s, ale zatímco u samohlásky *i* je nutné detekovat pouze 4 fonémy u samohlásky *a* je to 8 fonémů. Pro dostatečně přesný výpočet formantů samohlásky *u* je potřeba minimální signál o délce 60 – 70 s, což odpovídá 4 – 5 formantům. Tedy pro přesný výpočet všech formantových frekvencí s nastavením metody pro úlohu monologu je potřebný signál o délce alespoň 70 s. Nicméně vzhledem k tomu, že každý mluvčí vede jiný monolog a tedy volí i jiná slova a navíc mluví každý jinak rychle, může se u jednotlivých osob lišit minimální délka signálu k přesnému výpočtu formantů.

Cílem klasifikace bylo zjistit, s jakou přesností lze od sebe odlišit skupinu zdravých a nemocných lidí jednak na základě všech formantů a jednak na základě odvozených parametrů. Z výsledků mělo vyplynout, které parametry jsou pro odlišení obou skupin nejvhodnější. Ze statistické analýzy bylo vyhodnoceno, že je vhodné ženy a muže posuzovat zvlášť. To se projevilo i u výsledků klasifikace, kde klasifikace na mužích a ženách zvlášť vedla k vyšší přesnosti než při nerozlišování pohlaví. Nejvyšší přesnost klasifikace u úlohy textu u žen bylo dosaženo u parametru VAI ($ACC = 60 \%$). Pro muže bylo u stejné úlohy nejvyšší přesnosti dosaženo taktéž pro parametr VAI ($ACC = 74 \%$). Pro úlohu monologu bylo nejvyšší přesnosti u žen dosaženo u klasifikace všech formantů ($ACC = 72 \%$). U monologu u mužů bylo dosaženo nejvyšší přesnosti pro VAI ($ACC = 68 \%$). Pro obě úlohy se u mužů jevil jako nejvhodnější parametr pro odlišení obou skupin VAI a celkově vyšší přesnosti klasifikace bylo dosaženo u úlohy čtení. U žen bylo naopak vyšší úspěšnosti dosaženo u úlohy monologu a jako vhodnými parametry se jevily formantové frekvence. Nicméně pro klasifikaci bylo k dispozici málo dat, proto nemusí být tyto výsledky úplně vypovídající. Navíc vzhledem k věkové variabilitě formantových frekvencí by mohlo být vyšší přesnosti klasifikace dosaženo pro testování určitých věkových skupin zvlášť.

Navržená automatická metoda byla testována pouze na lidech v mírném až mírně pokročilém stádiu PN. Nicméně výsledky, které automatická metoda s rozpoznávačem přinesla se zdají být příslibem pro zautomatizování celého procesu hodnocení poruch artikulace. Kromě zautomatizování procesu je další přínos metody s rozpoznávačem řeči i v jednoduchém vybírání samohláskových fonémů (krátké či dlouhé) a jejich analýzou po specifických souhláskách. V práci byly testovány různé výběry samohlásek, nicméně vzhledem k nedostatečné délce řečových záznamů nemohl proběhnout specifičtější výběr jednotlivých fonémů.

Do budoucna by mohla být metoda otestována na neléčených pacientech v brzkém stádiu PN. Pokud by výsledky byly dostatečně citlivé pro rozlišení zdravých a nemocných lidí, mohla by metoda najít své využití v brzké diagnostice PN. Další přínos automatické metody by mohl být ve vyhodnocování úspěšnosti léčby a progresi onemocnění. Vzhledem k nynějšímu subjektivnímu hodnocení poruchy řeči u PN by mohla automatická metoda sloužit k objektivnímu hodnocení a vyhodnocovat tak například logopedickou terapii. Pokračující studie by se mohly týkat také hodnocení poruch artikulace samohlásek u pacientů ve vyšších stádiích PN či u dalších neurologických onemocněních, která jsou spojena s poruchou artikulace samohlásek. Kromě detekce poruch by mohla metoda sloužit také v jiných oblastech výzkumu, například v biometrii pro identifikaci nebo verifikaci mluvčích nebo ve zkoumání věkových závislostí artikulačních parametrů.

V navazujících pracích by bylo vhodné vylepšit stávající metodu pro výpočet formantových frekvencí zejména u samohlásek *i* a *u* a dále také vyzkoušet jiný rozpoznávač řeči s menší chybovostí. Dále by bylo vhodné automatickou metodu rozšířit pro detekci poruch artikulace u cizích jazyků, s čímž také souvisí použití jiných rozpoznávačů řeči.

5. Závěr

V rámci této práce byla navržena automatická metoda pro hodnocení poruch artikulace u PN. Metoda byla otestována a porovnána s výsledky ruční analýzy. Automatická metoda využívající rozpoznávač řeči přinesla srovnatelné výsledky s ruční analýzou co do výčtu i statistické významnosti signifikantních parametrů. Automatická metoda s rozpoznávačem se jeví jako stejně citlivá pro rozlišení zdravých lidí od pacientů s PN, což se zdá být příslibem ve zautomatizování celého procesu detekce poruch artikulace. Pro návrh metody byl využit algoritmus pro výpočet formantových frekvencí dostupný ze softwaru Praat, u kterého byla naměřena nejvyšší spolehlivost výpočtu jednotlivých formantů ze všech testovaných metod. Klasifikační experiment ukázal, že nejvhodnější parametr pro odlišení zdravých lidí a pacientů s PN je *VAI* a také jednotlivé formanty. Naopak nejméně vhodným se ukázal být parametr *VSA*. Z obou testovaných úloh se jevila senzitivnější úloha čteného textu pro muže a úloha monologu pro ženy. Automatická metoda bez rozpoznávače řeči byla pro tento typ úloh nevhodnou metodou. Vzhledem k velkému výskytu souhlásek v řeči je nutné provést segmentaci, aby byly zpracovány pouze relevantní části signálu. Automatická metoda s rozpoznávačem řeči by mohla své potenciální využití najít v pomoci při diagnóze PN, v objektivním hodnocení progresu nemoci a úspěšnosti léčby a v hodnocení logopedické terapie. Případným dalším užitím by mohlo být hodnocení poruch artikulace samohlásek u jiných neurologických nemocí, které souvisejí s tímto deficitem. Uplatnění této metody by mohlo být také v jiných oblastech výzkumu.

Literatura

- [1] Shimon Sapir et al. “Formant centralization ratio: a proposal for a new acoustic measure of dysarthric speech”. In: *Journal of Speech, Language, and Hearing Research* 53.1 (2010), s. 114–125.
- [2] Milan Sigmund. *Analýza řečových signálů: přednášky*. FEI VUT, Brno, 2000.
- [3] Harish Chander Mahendru. “Quick review of human speech production mechanism”. In: *International Journal of Engineering Research and Development* 9 (2014), s. 48–54.
- [4] Václav Cvrček. *Mluvnice současné češtiny: Jak se píše a jak se mluví. I*. Karolinum, 2015.
- [5] Ray D Kent et al. “Acoustic studies of dysarthric speech: Methods, progress, and potential”. In: *Journal of communication disorders* 32.3 (1999), s. 141–186.
- [6] Erwan Pépiot. “Voice, speech and gender: Male-female acoustic differences and cross-language variation in English and French speakers”. In: *Corela. Cognition, representation, language* HS-16 (2015).
- [7] Biswajit Das et al. “Effect of aging on speech features and phoneme recognition: a study on Bengali voicing vowels”. In: *International journal of speech technology* 16.1 (2013), s. 19–31.
- [8] Josef Psutka. *Komunikace s počítačem mluvenou řečí*. Academia, 1995.
- [9] Monica Colamartino et al. “Recent advances in pharmacological therapy of Parkinson’s disease: Levodopa and carbidopa protective effects against DNA oxidative damage”. In: *Health* 4 (2012), s. 1191–1199.
- [10] Per Brodal. *The central nervous system: structure and function*. Oxford University Press, 2004.
- [11] Ray D Kent et al. “The dysarthrias: speech-voice profiles, related dysfunctions, and neuropathology.” In: *Journal of Medical Speech-language pathology* (1998), s. 165–211.
- [12] Anthony E Lang a Andres M Lozano. “Parkinson’s disease”. In: *New England Journal of Medicine* 339.15 (1998), s. 1044–1053.
- [13] Jan Roth, Marcela Sekyrová a Evžen Růžička. *Parkinsonova nemoc*. Maxdorf, 1999.
- [14] Margaret M Hoehn a Melvin D Yahr. “Parkinsonism onset, progression, and mortality”. In: *Neurology* 17.5 (1967), s. 427–427.
- [15] S Fahn. “Levodopa in the treatment of Parkinson’s disease”. In: *Oxidative Stress and Neuroprotection*. Springer, 2006, s. 1–15.
- [16] Roongroj Bhidayasiri a Daniel Tarsy. *Movement disorders: a video atlas*. Springer Science & Business Media, 2012.
- [17] Pablo Martinez-Martin et al. *Assesment Scales in Parkinson’s Disease*. Springer, 2014.
- [18] Jeri A Logemann et al. “Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients”. In: *Journal of Speech and Hearing Disorders* 43.1 (1978), s. 47–57.

- [19] Frederic L Darley, Arnold E Aronson a Joe R Brown. “Differential diagnostic patterns of dysarthria”. In: *Journal of Speech, Language, and Hearing Research* 12.2 (1969), s. 246–269.
- [20] Frederic L Darley, Arnold E Aronson a Joe R Brown. “Clusters of deviant speech dimensions in the dysarthrias”. In: *Journal of Speech, Language, and Hearing Research* 12.3 (1969), s. 462–496.
- [21] Gerald J Canter. “Speech characteristics of patients with Parkinson’s disease: I. Intensity, pitch, and duration.” In: *Journal of Speech & Hearing Disorders* (1963), s. 221–229.
- [22] Gerald J Canter. “Speech characteristics of patients with Parkinson’s disease: II. Physiological support for speech.” In: *Journal of Speech & Hearing Disorders* (1965), s. 44–49.
- [23] Lorraine Olson Ramig et al. “Comparison of two forms of intensive speech treatment for Parkinson disease”. In: *Journal of Speech, Language, and Hearing Research* 38.6 (1995), s. 1232–1251.
- [24] Nancy Pearl Solomon a Thomas J Hixon. “Speech breathing in Parkinson’s disease”. In: *Journal of Speech, Language, and Hearing Research* 36.2 (1993), s. 294–310.
- [25] JB King et al. “Variability in acoustic and perceptual parameters of phonation in patients with Parkinson’s disease”. In: *Journal of Medical Speech-Language Pathology* 2 (1994), s. 29–42.
- [26] Rhonda J Holmes et al. “Voice characteristics in the progression of Parkinson’s disease”. In: *International Journal of Language & Communication Disorders* 35.3 (2000), s. 407–418.
- [27] Lorraine O Ramig et al. “Changes in vocal loudness following intensive voice treatment (LSVT®) in individuals with Parkinson’s disease: A comparison with untreated patients and normal age-matched controls”. In: *Movement Disorders* 16.1 (2001), s. 79–83.
- [28] Aileen K Ho, Robert Ianseck a John L Bradshaw. “Motor instability in parkinsonian speech intensity”. In: *Cognitive and Behavioral Neurology* 14.2 (2001), s. 109–116.
- [29] Aileen K Ho et al. “Speech volume regulation in Parkinson’s disease: Effects of implicit cues and explicit instructions”. In: *Neuropsychologia* 37.13 (1999), s. 1453–1460.
- [30] Félix Javier Jiménez-Jiménez et al. “Acoustic voice analysis in untreated patients with Parkinson’s disease”. In: *Parkinsonism & Related Disorders* 3.2 (1997), s. 111–116.
- [31] P Doyle et al. “Fundamental frequency and acoustic variability associated with production of sustained vowels by speakers with hypokinetic dysarthria”. In: *Journal of Medical Speech-Language Pathology* 3.1 (1995), s. 41–50.
- [32] Ingo Hertrich a Hermann Ackermann. “Gender-specific vocal dysfunctions in Parkinson’s disease: electroglottographic and acoustic analyses”. In: *Annals of Otology, Rhinology & Laryngology* 104.3 (1995), s. 197–202.
- [33] Leigh Lisker a Arthur S Abramson. “A cross-language study of voicing in initial stops: Acoustical measurements”. In: *Word* 20.3 (1964), s. 384–422.
- [34] Karen Forrest, Gary Weismer a Greg S Turner. “Kinematic, acoustic, and perceptual analyses of connected speech produced by Parkinsonian and normal geriatric adults”. In: *The Journal of the Acoustical Society of America* 85.6 (1989), s. 2608–2622.
- [35] Gerald J Canter. “Speech characteristics of patients with Parkinson’s disease: III. Articulation, diadochokinesis, and overall speech adequacy.” In: *Journal of Speech & Hearing Disorders* (1965b), s. 217–224.

- [36] Jeri A Logemann a Hilda B Fisher. “Vocal tract control in Parkinson’s disease”. In: *Journal of Speech and Hearing Disorders* 46.4 (1981), s. 348–352.
- [37] Gary Weismer et al. “Articulatory characteristics of Parkinsonian dysarthria: Segmental and phrase-level timing, spirantization, and glottal-supraglottal coordination”. In: *The dysarthrias: Physiology, acoustics, perception, management* (1984), s. 101–130.
- [38] Samuel G Fletcher. “Time-by-count measurement of diadochokinetic syllable rate”. In: *Journal of Speech, Language, and Hearing Research* 15.4 (1972), s. 763–770.
- [39] Hermann Ackermann, Ingo Hertrich a Thomas Hehr. “Oral diadochokinesis in neurological dysarthrias”. In: *Folia Phoniatrica et Logopaedica* 47.1 (1995), s. 15–23.
- [40] Guylaine Le Dorze et al. “A comparison of the prosodic characteristics of the speech of people with Parkinson’s disease and Friedreich’s ataxia with neurologically normal speakers”. In: *Folia Phoniatrica et Logopaedica* 50.1 (1998), s. 1–9.
- [41] JF Caekebeke et al. “The interpretation of dysprosody in patients with Parkinson’s disease.” In: *Journal of Neurology, Neurosurgery & Psychiatry* 54.2 (1991), s. 145–148.
- [42] E Jeffrey Metter a Wayne R Hanson. “Clinical and acoustical variability in hypokinetic dysarthria”. In: *Journal of Communication disorders* 19.5 (1986), s. 347–366.
- [43] C Mawdsley a CV Gamsu. “Periodicity of speech in Parkinsonism”. In: *Nature* 231.5301 (1971), s. 315–316.
- [44] Kenneth K Nakano, Howard Zubick a H Richard Tyler. “Speech defects of parkinsonian patients: Effects of levodopa therapy on speech intelligibility.” In: *Neurology* (1973).
- [45] MP Caligiuri. “Short-term fluctuations in orofacial motor control in Parkinson’s disease”. In: *Recent advances in clinical dysarthria* (1989), s. 199–212.
- [46] Jan Ruzs et al. “Evaluation of speech impairment in early stages of Parkinson’s disease: a prospective study with the role of pharmacotherapy”. In: *Journal of Neural Transmission* 120.2 (2013), s. 319–329.
- [47] Heejin Kim, Mark Hasegawa-Johnson a Adrienne Perlman. “Vowel contrast and speech intelligibility in dysarthria”. In: *Folia Phoniatrica et Logopaedica* 63.4 (2011), s. 187–194.
- [48] Ray D Kent a Y-J Kim. “Toward an acoustic typology of motor speech disorders”. In: *Clinical linguistics & phonetics* 17.6 (2003), s. 427–445.
- [49] Shimon Sapir et al. “Effects of intensive voice treatment (the Lee Silverman Voice Treatment [LSVT]) on vowel articulation in dysarthric individuals with idiopathic Parkinson disease: acoustic and perceptual findings”. In: *Journal of Speech, Language, and Hearing Research* 50.4 (2007), s. 899–912.
- [50] Gary Weismer et al. “Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders”. In: *Folia Phoniatrica et Logopaedica* 53.1 (2001), s. 1–18.
- [51] Michiko Hashi, John R Westbury a Kiyoshi Honda. “Vowel posture normalization”. In: *The Journal of the Acoustical Society of America* 104.4 (1998), s. 2426–2437.
- [52] Sabine Skodda, Wenke Visser a Uwe Schlegel. “Vowel articulation in Parkinson’s disease”. In: *Journal of Voice* 25.4 (2011), s. 467–472.
- [53] Jan Ruzs et al. “Imprecise vowel articulation as a potential early marker of Parkinson’s disease: Effect of speaking task”. In: *The Journal of the Acoustical Society of America* 134.3 (2013), s. 2171–2181.

- [54] Daniel Kempler a Diana Van Lancker. “Effect of speech task on intelligibility in dysarthria: A case study of Parkinson’s disease”. In: *Brain and language* 80.3 (2002), s. 449–464.
- [55] Alexander M Goberman a Carl Coelho. “Acoustic analysis of Parkinsonian speech I: Speech characteristics and L-Dopa therapy”. In: *NeuroRehabilitation* 17.3 (2002), s. 237–246.
- [56] Geralyn M Schulz a Megan K Grant. “Effects of speech therapy and pharmacologic and surgical treatments on voice and speech in Parkinson’s disease: a review of the literature”. In: *Journal of communication disorders* 33.1 (2000), s. 59–88.
- [57] Paulus Petrus Gerardus Boersma et al. “Praat, a system for doing phonetics by computer”. In: *Glott international* 5 (2002).
- [58] Stephen W Metz et al. “Auditory modeling applied to formant tracking of noise-corrupted speech”. In: *Industrial Electronics, Control and Instrumentation, 1991. Proceedings. IECON’91., 1991 International Conference on.* IEEE. 1991, s. 2120–2124.
- [59] Ashwin Rao a Ramdas Kumaresan. “On decomposing speech into modulated components”. In: *IEEE Transactions on Speech and Audio Processing* 8.3 (2000), s. 240–254.
- [60] Ian C Bruce et al. “Robust formant tracking in noise”. In: *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on.* Sv. 1. IEEE. 2002, s. 281–284.
- [61] Abu Shafin Mohammad Mahdee Jameel et al. “Noise Robust Formant Frequency Estimation Method Based on Spectral Model of Repeated Autocorrelation of Speech”. In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* (2016).
- [62] Satrajit Ghosh. *Formant Tracker for Matlab*. URL: <http://cns.bu.edu/~speech/ftrack.php> (cit. 25.04.2017).
- [63] Kamran Mustafa a Ian C Bruce. “Robust formant tracking for continuous speech with speaker variability”. In: *IEEE Transactions on Audio, Speech, and Language Processing* 14.2 (2006), s. 435–444.
- [64] Lawrence R Rabiner a Ronald W Schafer. *Digital processing of speech signals*. Prentice Hall, 1978.
- [65] Steven Sandoval et al. “Automatic assessment of vowel space area”. In: *The Journal of the Acoustical Society of America* 134.5 (2013), EL477–EL483.
- [66] Petr Schwarz. *Phoneme recognition based on long temporal context*. Disertační práce, Fakulta informačních technologií, VUT v Brně. 2009.
- [67] Prasanta Chandra Mahalanobis. “On the generalized distance in statistics”. In: *Proceedings of the National Institute of Sciences (Calcutta)* 2 (1936), s. 49–55.
- [68] Roy De Maesschalck, Delphine Jouan-Rimbaud a Désiré L Massart. “The mahalanobis distance”. In: *Chemometrics and intelligent laboratory systems* 50.1 (2000), s. 1–18.
- [69] Yoav Freund, Robert E Schapire et al. “Experiments with a new boosting algorithm”. In: *icml*. Sv. 96. 1996, s. 148–156.

Příloha A.

Text

Když člověk poprvé vsa¹dí do země sa²zeničku¹, chodí se na ni dívat tři¹krát denně: ta³ko, povyrosla u²ž nebo ne? I² ta⁴jí dech, naklání se nad ní, při³tlačí⁴ trochu³ pů⁴du u⁵ jejích kořínků⁶, načechrává jí lístky a⁵ vůbec ji obtěžuje různým konáním, které považuje za u⁷ž⁵tečnou péči⁶. A⁶ když se sa⁷zenička⁸ přesto u⁸jme a roste jako z vody, tu člověk žasne nad tímto divem pří⁷rody, má poci⁸t čehosi⁹ jako zázra⁹ku⁹ a¹⁰ považuje to za jeden ze svých největší¹⁰ch osobních ú¹⁰spěchů.

Příloha B.

Obsah příloženého CD

K této práci je přiloženo CD, na kterém je uložena práce ve formátu PDF a zdrojové kódy. Struktura adresářů na CD je uvedena níže.

