

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

BAKALÁŘSKÁ PRÁCE



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV TELEKOMUNIKACÍ

DEPARTMENT OF TELECOMMUNICATIONS

HODNOCENÍ HYBNOSTI MLUVIDEL NA ZÁKLADĚ AKUSTICKÉ ANALÝZY ŘEČI

ASSESSING MOVEMENT OF ARTICULATORY ORGANS BASED ON ACOUSTIC ANALYSIS OF SPEECH

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Kryštof Novotný

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Jiří Mekyska, Ph.D.

BRNO 2021

Bakalářská práce

bakalářský studijní program **Audio inženýrství**
specializace Zvuková produkce a nahrávání
Ústav telekomunikací

Student: Kryštof Novotný

ID: 212566

Ročník: 3

Akademický rok: 2020/21

NÁZEV TÉMATU:

Hodnocení hybnosti mluvidel na základě akustické analýzy řeči

POKYNY PRO VYPRACOVÁNÍ:

V rámci práce bude vytvořena sada funkcí, která na základě akustické analýzy řeči umožní popis hybnosti mluvidel (především jazyka a brady). Funkce budou z řečových nahrávek odhadovat kmitočty formantů a následně modelovat tzv. Hellwagův trojúhelník. Dále budou z tohoto trojúhelníku počítány parametry, jako např. vokální oblast (vowel space area), artikulační index samohlásek (vowel articulation index), artikulačně-akustická vokální oblast (articulatory-acoustic vowel space) atd. Na databázi řečových nahrávek pacientů s Parkinsonovou nemocí bude nakonec otestována síla těchto parametrů odlišit zdravou řeč a řeč ovlivněnou hypokinetickou dysartrií.

DOPORUČENÁ LITERATURA:

[1] SANDOVAL, Steven, Visar BERISHA, Rene L. UTIANSKI, Julie M. LISS a Andreas SPANIAS. Automatic assessment of vowel space area. The Journal of the Acoustical Society of America. 2013, 134(5). DOI: 10.1121/1.4826150. ISSN 0001-4966.

[2] WHITFIELD, Jason A. Exploration of Metrics for Quantifying Formant Space: Implications for Clinical Assessment of Parkinson Disease. Perspectives of the ASHA Special Interest Groups. 2019, 4(2), 402-410. DOI: 10.1044/2019_PERS-SIG19-2018-0004. ISSN 2381-4764.

Termín zadání: 1.2.2021

Termín odevzdání: 31.5.2021

Vedoucí práce: Ing. Jiří Mekyska, Ph.D.

doc. Ing. Jiří Schimmel, Ph.D.
předseda rady studijního programu

UPOZORNĚNÍ:

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

ABSTRAKT

Hypokinetická dysartrie je motorická porucha řeči, často přítomná při průběhu Parkinsonovi nemoci. Postihuje řečové ústrojí včetně artikulačních schopností. Existuje více řečových parametrů popisujících tuto oblast, proto se nabízí zabývat se jejich vzájemným srovnáním. Práce si klade za cíl navrhnout a popsat algoritmus pro výpočet parametrů artikulace, přizpůsobený českému jazyku, a následně porovnat jejich diskriminační sílu. Akustickou analýzu řeči v algoritmu zajišťuje program Praat a pro následné zpracování dat jsou použity základní algoritmy strojového učení jako Expectation-Maximization, K-means nebo lineární regrese. Pro vyhodnocení slouží Mann-Whitneyho U test a zástupci lineárních, nelineárních i souhrnných metod strojového učení s využitím křížové validace a vyvážené přesnosti. Výsledkem jsou skripty pro automatické nalezení hran Hellwagova vokálního trojúhelníku, pro výpočet artikulačních parametrů a pro jejich vyhodnocení. Výstupy analýzy dvou různých databází (PARCZ a CoBeN) dokazují, že mezi běžnou a dysartrickou řečí lze skutečně pozorovat rozdíly v artikulaci. Na základě vzájemného srovnávání výsledků je proto v práci navrženo, kterými parametry a modely strojového učení je vhodné se dále v souvislosti s touto problematikou zabývat.

KLÍČOVÁ SLOVA

Hypokinetická dysartrie, Parkinsonova nemoc, akustická analýza, zpracování řečových signálů, parametrizace řeči, formantové kmitočty, artikulace, strojové učení

ABSTRACT

Hypokinetic dysarthria is a motor speech disorder often present during Parkinson's disease. It affects the speech system, including articulatory abilities. There are several speech parameters describing this domain, so it is suggested to deal with their mutual comparison. This work aims to design and describe an algorithm for calculating the parameters of articulation, adapted for the Czech language, and then compare their discriminative power. The acoustic analysis of speech included in it is done via the Praat program and basic machine learning algorithms such as Expectation-Maximization, K-means and linear regression are used for the subsequent data processing. The Mann-Whitney U test and representatives of linear, nonlinear and ensemble machine learning models using cross-validation and balanced accuracy are used for evaluation. The results are scripts for automatic assessment of vowel space area, for calculating articulation parameters and for their evaluation. The outputs of the analysis of two different databases (PARCZ and CoBeN) prove that differences in articulation can indeed be observed between normal and dysarthric speech. Based on the mutual comparison of results, it is therefore proposed in the work which parameters and models of machine learning are being appropriate for further dealing with this issue.

KEYWORDS

Hypokinetic dysarthria, Parkinson's disease, acoustic analysis, speech signal processing, speech parametrization, formant frequencies, articulation, machine learning

NOVOTNÝ, Kryštof. *Hodnocení hybnosti mluvidel na základě akustické analýzy řeči*. Brno, 2021, 59 s. Bakalářská práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací. Vedoucí práce: Ing. Jiří Mekyska, Ph.D.

PROHLÁŠENÍ

Prohlašuji, že svou bakalářskou práci na téma „Hodnocení hybnosti mluvidel na základě akustické analýzy řeči“ jsem vypracoval samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené bakalářské práce dále prohlašuji, že v souvislosti s vytvořením této bakalářské práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Brno

.....

podpis autora

PODĚKOVÁNÍ

Rád bych poděkoval vedoucímu bakalářské práce panu Ing. Jiřímu Mekyskovi, Ph.D. za odborné vedení, konzultace, trpělivost a podnětné návrhy k práci.

Obsah

Úvod	11
1 Formanty	13
1.1 Využití formantů	14
1.2 Metody analýzy formantů	14
1.2.1 Krátkodobá Fourierova transformace diskrétního signálu	14
1.2.2 Nelineární homomorfická metoda	14
1.2.3 Lineární predikční analýza	15
2 Hellwagův vokalický trojúhelník	17
3 Algoritmus automatického výpočtu Hellwagova trojúhelníku	19
3.1 Extrakce formantů	19
3.2 Filtrování	20
3.3 Shluky	20
4 Parametry založené na formantových kmitočtech	21
4.1 VSA	21
4.2 LnVSA	21
4.3 Formant Centralization Ratio (FCR)	22
4.4 Vowel Articulation Index (VAI)	22
4.5 Working Vowel Space Metrics (WVSM)	23
4.5.1 Vowel Space Hull Area (VSHA)	23
4.5.2 Articulatory-Acoustic Vowel Space (AAVS)	23
4.5.3 Vowel Space Density (VSD)	24
4.5.4 Density Percentage (DP)	24
5 Implementace algoritmu odhadu hran trojúhelníku	25
5.1 Vykreslování grafů	25
5.2 Práce s maticemi	25
5.3 Matematické operace	26
5.4 Využití funkcí prostředí Praat	26
5.5 GMM	28
5.6 K-means	28
5.7 Konvexní obal	29
6 Databáze	30

7 Metodologie	33
7.1 Statistická analýza	33
7.2 Strojové učení	34
7.3 Metody použité v evaluaci systému	35
7.3.1 Mann-Whitneyův U test	35
7.3.2 Lineární metody strojového učení	35
7.3.3 Nelineární metody strojového učení	35
7.3.4 Souhrnné metody strojového učení	36
7.3.5 Vyvážená přesnost	36
7.3.6 Křížová validace	37
8 Výsledky	38
8.1 Samostatné výsledky jednotlivých parametrů	38
8.2 Výsledky modelování pomocí strojového učení	43
9 Diskuze	48
Závěr	50
Literatura	52
Seznam symbolů, veličin a zkratk	56
Obsah příložených souborů	59

Seznam obrázků

1.1	Průřez mluvním ústrojím	13
1.2	Vokál [a]	15
1.3	Vokál [e]	15
1.4	Vokál [i]	16
1.5	Vokál [o]	16
1.6	Vokál [u]	16
2.1	Hellwagův vokalický trojúhelník	17
3.1	Algoritmus automatického výpočtu Hellwagova trojúhelníku	20
5.1	Analyzovaný řečový signál	26
5.2	Extrahované dvojice F1 – F2	27
5.3	Zvolené centroidy	28
5.4	Odebrané dvojice F1 – F2	28
5.5	Nalezená centra shluků	29
5.6	Vykreslení konvexního obalu	29
6.1	Demografické zastoupení v použitých databázích	30
7.1	Blokové schéma zpracování dat v této práci	37
8.1	Porovnání parametrů na základě výsledků z PARCZ	42
8.2	Porovnání parametrů na základě výsledků z CoBeN	42
8.3	Srovnání nejlepších výsledků obou databází	43
8.4	Různé modely trénované na PARCZ	44
8.5	Nejlepší modely trénované na PARCZ	45
8.6	Různé modely trénované na CoBeN	46
8.7	Nejlepší modely trénované na CoBeN	47
9.1	Srovnání filtrace pomocí GMM u PN a HC	49
9.2	Srovnání VSD u PN a HC	49

Seznam tabulek

2.1	Kmitočty formantů	17
4.1	Porovnání VSA a FCR	22
6.1	PARCZ – Věk	31
6.2	PARCZ PN – LED	31
6.3	PARCZ PN – UPDRS III	31
6.4	PARCZ PN – délka trvání PN	31
6.5	CoBeN – Věk	32
6.6	CoBeN PN – LED	32
6.7	CoBeN PN – UPDRS III	32
6.8	CoBeN PN – délka trvání PN	32
8.1	Deskriptivní statistiky - VSA	38
8.2	Deskriptivní statistiky - LnVSA	39
8.3	Deskriptivní statistiky - VAI	39
8.4	Deskriptivní statistiky - VSHA	40
8.5	Deskriptivní statistiky - AAVS	40
8.6	Deskriptivní statistiky - RSDF1	40
8.7	Deskriptivní statistiky - RSDF2	41
8.8	Deskriptivní statistiky - DP	41
8.9	Různé modely trénované na PARCZ	44
8.10	Nejlepší modely trénované na PARCZ	45
8.11	Různé modely trénované na CoBeN	46
8.12	Nejlepší modely trénované na CoBeN	47

Úvod

Tvorba řeči je složitý motorický proces zahrnující fonaci (kmitání hlasivek), respiraci (dýchání, zde především jeho rytmus), rezonanci (utváření barvy hlasu), prozodii (zvukové vlastnosti jazyka) a v neposlední řadě také artikulaci. Při motorických poruchách řeči (MSD – Motor Speech Disorders) dochází k omezení právě těchto faktorů. Podle [8] představuje až 9,3% ze všech MSD porucha jménem hypokinetická dysartrie (HD). Původem rodiny neurogenních poruch nazývaných souhrnně jako dysartrie je nejčastěji dysfunkce centrálního nervového systému na úrovni mozku, může se však jednat i o postižení některého z hlavových nervů, postižení nervosvalového přenosu nebo dokonce poškození svalu samotného. Až 10% ze všech výskytů dysartrií tvoří již zmíněná HD, vznikající při postižení extrapyramidového systému, potažmo funkčního systému propojení bazálních ganglií. Důvodem zabývat se touto poruchou je fakt, že HD má tendenci vyskytnout se až u 90% případů průběhu obecně známého onemocnění – Parkinsonovy nemoci (PN). [8]

U samotné HD se nejedná o problém tvořit řeč z pohledu jazyka jako systému. Pacient nemá problém rozumět tomu, co slyší nebo čte, není oslabena schopnost psát ani vymýšlet věty. Vědomí narušeno není, a tedy se neztrácí informace, jako například významy slov nebo pravidla gramatiky. Jak je zmíněno v prvním odstavci, jedná se o motorický problém. Dýchání může být přerývané, nekoordinované s řečí; hlas vyšší, přiškrcený, tichý nebo naopak příliš hlasitý, nestálý, s nepřírozenou intonací; řeč monotónní nebo trhaná, s abnormálním tempem, nesprávným frázováním a rytmem. Z důvodu snížené pohyblivosti mluvidel může být řeč setřelá nebo huhňavá. Tato práce se zaměřuje na akustickou analýzu právě těchto artikulačních problémů.

Popisem výše zmíněné problematiky z akustického hlediska se již několik zahraničních autorů zabývalo. Jelikož jsou ale parametry řeči vždy ovlivněny jazykem mluvčího, nelze jejich přístupy bez ověření generalizovat a aplikovat u česky mluvících pacientů.

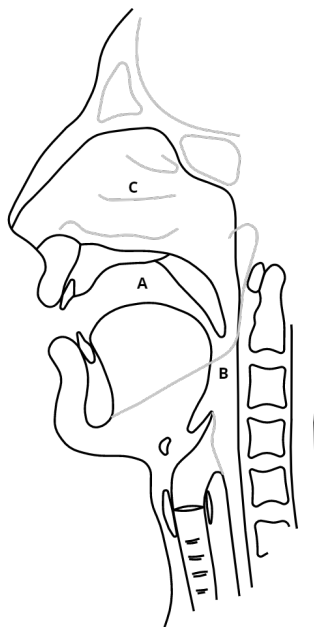
Cílem této bakalářské práce je uvést doposud používané postupy při zkoumání pohyblivosti mluvidel, popsat jejich princip a najít silné i slabé stránky. Dále na jejich základě navrhnout, popsat a zhodnotit podobný algoritmus, uzpůsobený našim podmínkám. Jeho výstupem by mělo být několik parametrů charakterizujících pohyblivost mluvidel. Následně by měla být zhodnocena síla jednotlivých parametrů rozlišit řeč pacientů s PN od zdravé kontrolní skupiny (HC). Posledním cílem je na získaných parametrech natrénovat modely strojového učení a porovnat rozlišovací schopnosti všech metrik jako celku.

V práci je postupně rozebrána problematika formantů lidského hlasu, jejich vznik a obecné využití. Jsou uvedeny metody jejich analýzy a popsán princip těchto metod. Na téma formantů navazuje kapitola zabývající se jejich uspořádáním do Hellwagova vokálního trojúhelníku. Nejprve mu je věnována pozornost z pohledu fonetiky. Následně je popsána jeho akustická analogie, pro kterou je rovněž vysvětlen algoritmus výpočtu. Další část shrnuje parametry získatelné z objektů rozebraných v předchozích kapitolách. Každý parametr je vysvětlen a je uveden způsob jeho výpočtu. Rozsáhlá kapitola se poté věnuje návrhu algoritmu pro výpočet Hellwagova vokálního trojúhelníku. Pozornost je zaměřena jak na popis implementace, tak i na použité knihovny třetích stran. Navazuje kapitola přibližující použité databáze řečových nahrávek a také celý proces zpracování a vyhodnocování dat. Ve zkratce jsou zde poté uvedeny základní informace k jednotlivým použitým metodám strojového učení a metrikám evaluace. Následující dvě kapitoly shrnují naměřená a vypočítaná data, uvádějí je do souvislostí a vyvozují závěry z nich vyplývající.

1 Formanty

Hlavní oblasti koncentrace energie ve zvukovém spektru akustického signálu se nazývají formanty. V makroskopickém měřítku od sebe lze díky formantům rozeznávat jednotlivé řečníky, v mikroskopickém měřítku potom jednotlivé vokály – v českém jazyce: [a], [e], [i], [o], [u]. Druhý pohled na věc je pro tuto práci důležitější.

Tón lidského hlasu je buzen v hlasivkách. Při porovnání hlasového traktu s modelem oscilátor-rezonnátor-radiátor-excitátor, jsou tedy hlasivky oscilátorem. Rezonátorem jsou pak dutiny tohoto ústrojí. Ty ve výsledném tónu určují jeho formanty, tedy kmitočty, na kterých dochází k rezonanci. Dutiny lze rozdělit podle toho, jestli je možné je při řeči měnit. Za ty statické, které se při řeči nemění, lze považovat dutinu hrudní, dutinu nosní (obr. 1.1 – C) a částečně dutinu hrdelní (obr. 1.1 – B). Tyto rezonátory vytvářejí v hlase nepohyblivé formanty, jejichž střední kmitočet je závislý na velikosti těchto dutin. Zástupcem dynamických kavit je naopak dutina ústní (obr. 1.1 – A). U této dutiny je možné silně měnit její objem a tím ve spektru měnit kmitočet formantu. To je demonstrováno pro pět českých vokálů na obrázcích 1.2, 1.3, 1.4, 1.5 a 1.6. Někdy se hovoří ještě o takzvaném retním rezonátoru. Ten má význam například v němčině, kde existuje více samohlásek, tvořených při stejném objemu ústní dutiny i se stejnou pozicí jazyka v ní. Příkladný rozdíl mezi [y] a [ü] potom ovlivňuje právě retní rezonátor. [15, 23]



Obr. 1.1: Průřez mluvním ústrojím podle [15]

1.1 Využití formantů

Mimo využití formantů za účelem rozeznání jednotlivých vokálů a jejich parametrů, mají řečové formanty i mnoho dalších aplikací. Jedním z častých použití je syntéza řeči. Jedná se o opačný proces, než je ten výše zmíněný. Zde se pomocí kmitočtového generátoru, zesilovačů a filtrů emuluje hlasový trakt. Reálnou funkci kavit potom zastávají rezonanční obvody nebo lineárně predikční analýza. Toto se využívá například v tzv. TTS (Text-to-Speech) programech. [23, 35]

Dalším příkladem využití formantů je kódování řeči. Zde se opět využívá lineární predikce. Tento model využívají téměř všechny kodeky pro veřejné komunikace díky jeho nízké přenosové rychlosti. Přenáší se bitový datový tok, který je po přenosu rekonstruován na řečový signál na základě předem přijatých parametrů. Ukázkovým zástupcem je například kodek Global System for Mobile Communications označovaný zkratkou GSM. [35]

Na vzájemně si podobném principu funguje určování šepotu, emocí řečníka nebo hladiny alkoholu v krvi. V každém z těchto případů dochází k zesílení určitých formantů nebo jejich posunu na kmitočtové ose. Těchto vlastností bude využito i v této práci, a sice k akustické analýze artikulace u pacientů s HD. [16]

1.2 Metody analýzy formantů

1.2.1 Krátkodobá Fourierova transformace diskrétního signálu

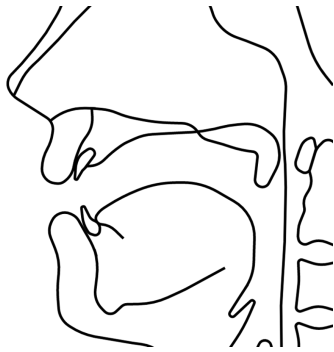
Obraz získaný diskrétní Fourierovou transformací (DFT) obsahuje jak buzení hlasivkami, tak i vliv modifikačního hlasového ústrojí. Pro analýzu hodnocení mluvidel není budící signál tolik důležitý. Samotné spektrum DFT proto není nejvhodnějším způsobem analýzy formantů.

1.2.2 Nelineární homomorfická metoda

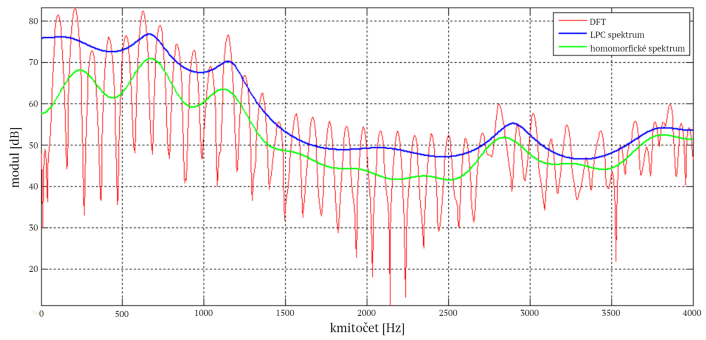
Každá kavita v hlasovém traktu představuje v kmitočtové oblasti filtr, konkrétně rezonátor. Je možné vytvořit matematický model tohoto ústrojí, aniž bychom uvažovali vliv budícího signálu. Jednou z možností modelování tohoto filtru je využití reálného kepstra. Na modulové spektrum DFT je po aplikaci přirozeného logaritmu použita zpětná DFT. Ve výsledném reálném kepstru je pak pomocí obdélníkového okna oddělena impulzní odezva hlasového traktu. Dopřednou DFT a po aplikaci exponenciální funkce pak získáme modulovou kmitočtovou charakteristiku filtru hlasového ústrojí.

1.2.3 Lineární predikční analýza

Další metodou, která se velmi často používá, je lineární predikce (LPC). Každý vzorek signálu je chápán jako lineární kombinace vzorků předcházejících. Vzorky jsou násobeny lineárními predikčními koeficienty a odpovídají koeficientům jmenovatele přenosové funkce hlasového traktu. Pomocí těchto koeficientů je tedy opět možné získat spektrální obálku řeči. Oproti nelineární metodě však nedokáže tato modelovat tzv. antiformanty, tj. lokální údolí v obálce. [23, 35]



(a) Průřez mluvním ústrojím

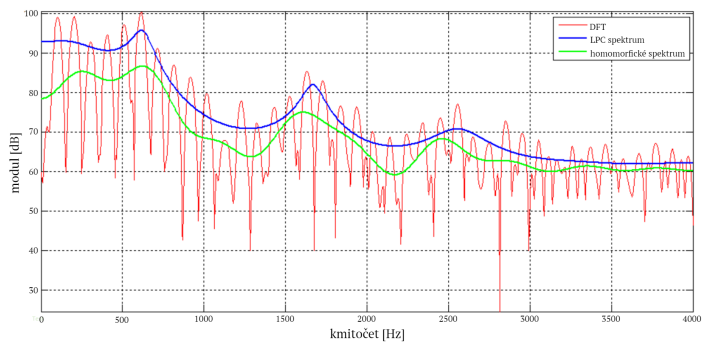


(b) Spektrum pomocí DFT, LPC a nelineární metody

Obr. 1.2: Vokál [a]

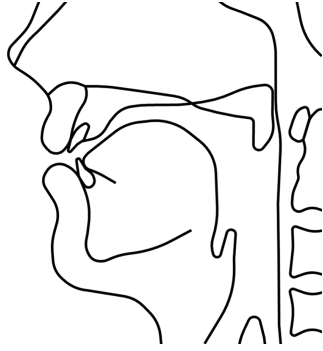


(a) Průřez mluvním ústrojím

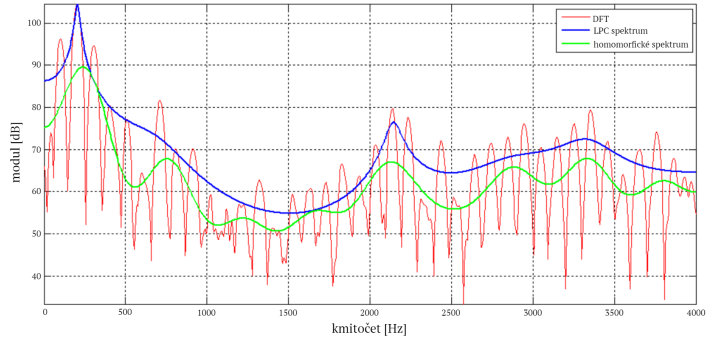


(b) Spektrum pomocí DFT, LPC a nelineární metody

Obr. 1.3: Vokál [e]



(a) Průřez mluvním ústrojím

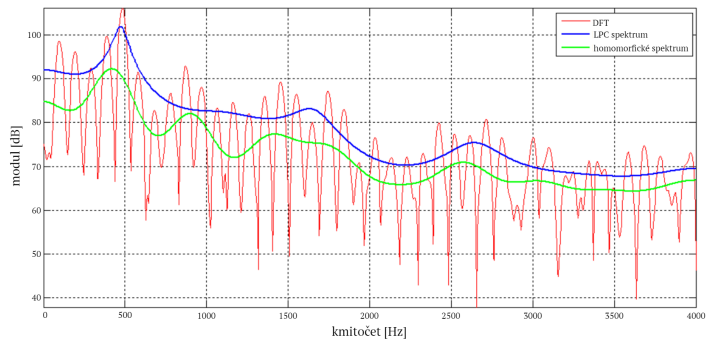


(b) Spektrum pomocí DFT, LPC a nelineární metody

Obr. 1.4: Vokál [i]

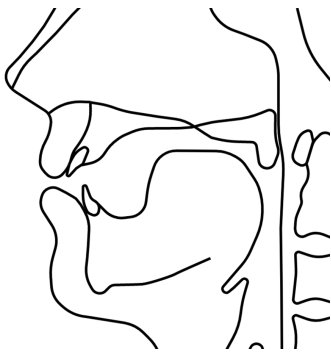


(a) Průřez mluvním ústrojím

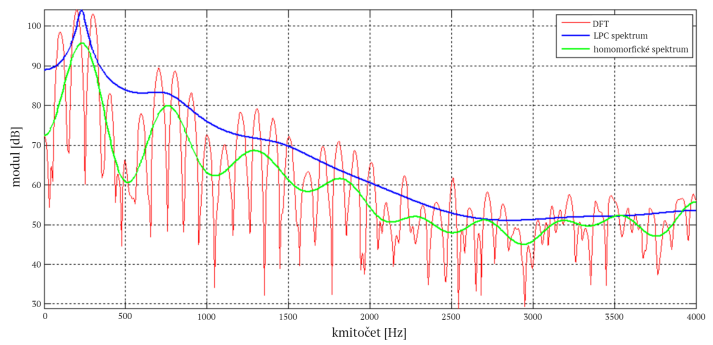


(b) Spektrum pomocí DFT, LPC a nelineární metody

Obr. 1.5: Vokál [o]



(a) Průřez mluvním ústrojím



(b) Spektrum pomocí DFT, LPC a nelineární metody

Obr. 1.6: Vokál [u]

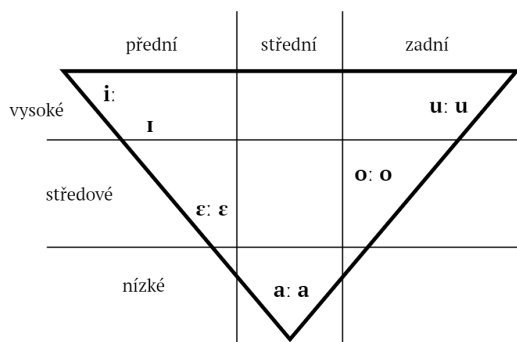
2 Hellwagův vokalický trojúhelník

Pořadí formantů řeči se určuje od těch na nižších kmitočtech po ty vyšší. Nejpohyblivější je formant druhý – F2 (viz tab. 2.1). Tento formant odpovídá dutině ústní a je silně ovlivňován především polohou jazyka a velikostí otevření čelistí. Druhým nejpohyblivějším je formant dutiny hrdelní – F1. Je také ovlivňován jazykem, i když už ne v takové míře. Jde o polohu kořene jazyka. Pohybem jazyka se tedy mění poměr objemů těchto dvou kavit. [16, 18]

Postavením jazyka v dutině ústní se zabýval Ch. F. Hellwag, který v roce 1781 vytvořil obecné schéma vokálů, známé jako vokalický trojúhelník (obr. 2.1). Je pozorován posun jazyka ve vertikální a horizontální ose. Mírné vylepšení přinesl D. Jones v roce 1914 začleněním více vokálů. Z původního trojúhelníka se stal lichoběžník. Tato historická schémata jsou natolik univerzální, že je lze aplikovat i na nově objevené vokály (například nuance v nářečí), a tedy se používají dodnes. Pro práci s řečovými signály v českém jazyce je postačující původní Hellwagovo schéma, protože vokály tvořící čtvrtý vrchol Jonesova lichoběžníku se ve spisovné české řeči nevyskytují. [15, 36]

Tab. 2.1: Kmitočty první tři formantů českých vokálů podle [18]

Vokál	F1 [Hz]	F2 [Hz]	F3 [Hz]
[i]	300–500	2000–2800	2600–3500
[e]	480–700	1560–2100	2500–3000
[a]	700–1100	1100–1500	2500–3000
[o]	500–700	850–1200	2500–3000
[u]	300–500	600–1000	2400–2900



Obr. 2.1: Schéma Hellwagova vokalického trojúhelníku pro české vokály podle [15]

Analogií Hellwagova vokalického trojúhelníka ve spektrální oblasti je vokální oblast samohlásek (VSA – Vowel Space Area). Základní způsob identifikace jednotlivých vokálů na základě jejich spektra je založen na poměru kmitočtů prvních dvou formantů. Ten je pro každý vokál různý. Propojení bodů, získaných zanesením kmitočtů F1 a F2 jednotlivých vokálů do grafu, vytváří opět trojúhelník. Vrcholy tohoto trojúhelníka odpovídají stejným vokálům, které definují i Hellwagův trojúhelník – [i], [u] a [a]. [15, 31]

Díky úzkému spojení prvních dvou formantů s velikostí kavit hlasového traktu, je VSA využíváno k analýze pohyblivosti mluvidel. Některé poruchy řeči se mohou projevat zhoršenou hybností artikulačních orgánů (např. jazyka) a tím pádem centralizací formantů jednotlivých vokálů. Proto má VSA uplatnění při výzkumu vývoje řeči a řečových poruch. Je prokázáno, že u řeči zřetelné a srozumitelné nabývá VSA větších hodnot. [8]

VSA má ovšem i své limitace. Jednou z nich je například fakt, že výpočty vycházející z této metody jsou založené pouze na vokálech nacházejících se ve vrcholech trojúhelníka. Případné nesrozumitelnosti ostatních vokálů tedy nemusí být věnována taková pozornost. Výzkumy také dokázaly, že přestože je pozorována změna ve VSA u řečníků trpících např. PN, tento rozdíl nemusí dosáhnout statistické významnosti. Proto se uplatňuje rozšiřující metoda, kdy se zkoumá schopnost mluvčího modulovat svůj hlas. Bylo dokázáno, že u pacientů s PN, dochází k větším rozdílům ve VSA při různých stylech řeči. Další výraznou limitací je individualita formantů pro každého řečníka. Jejich přesná poloha závisí na pohlaví, věku... Tím je znemožněn univerzální přístup k VSA, který by byl nezávislý na řečníkovi. [31]

3 Algoritmus automatického výpočtu Hellwagova trojúhelníku

Tradiční způsob výpočtu VSA spočívá v určení F1 a F2 pro každý vokál ve vrcholu trojúhelníka. Klasická metoda vyžaduje fonetickou segmentaci řečového signálu, tak aby bylo možné určit formanty pro každou hlásku zvlášť. Nabízelo by se pořádit jen nahrávky jednotlivých vokálů, ale pro umocnění efektu poruchy hybnosti mluvidel na výsledný signál je vhodnější určovat formanty ze spojitě řeči. Kvůli fonetické segmentaci je tato metoda pro určování rozdílů mezi řečí zdravou a tou s poruchami poměrně složitá.

Proto byl v roce 2013 vyvinut algoritmus umožňující automatický výpočet VSA. Jeho předností je právě automatizace, dále schopnost zpracovávat řečový signál o libovolné délce (za předpokladu, že obsahuje všechny potřebné vokály) a také uvažování všech vokálů obsažených v signálu (ne tedy pouze těch ve vrcholech trojúhelníka podle klasické metody). F1 a F2 jsou vypočítávány rámeček po rámečku, přičemž každý rámeček je ještě před extrakcí formantů váhován oknem. Následně jsou z časových řad formantů odstraněny odlehlé hodnoty. Ty zbývající tvoří shluky, jejichž centra určují vrcholy VSA. Kroky algoritmu znázorňuje schéma na obrázku 3.1.

Výsledky Sandovala (2013) ukazují, že automatizovaný odhad VSA vykazuje silnou korelaci s odhadem ručně segmentovaným. Z principu algoritmu vyplývá, že analyzovaný obsah musí být mezi porovnávanými řečníky foneticky vyvážený (ideálně totožný). [20, 24]

3.1 Extrakce formantů

K určení dvojice F1 – F2 v každém rámečku je možné použít program Praat. Ten na základě autokorelační metody odhadující periodicitu posuzuje výskyt znělé řeči v každém rámečku. Rámce mají vzájemný překryv 1 ms. Praat při výpočtu formantů převzorkuje signál na dvojnásobek uvažované horní hranice a poté aplikuje preemfázový filtr. Kvůli poklesu energie 6 dB na oktávu v řečovém signálu jsou vyšší formanty několikanásobně převyšovány F1. Preemfázový filtr se stará o vyrovnání kmitočtové charakteristiky a jako horní propust energii na nižších kmitočtech potlačuje. Poté je v časové doméně signál váhován Gaussovým oknem. Následně jsou odhadnuty lineární predikční koeficienty LPC podle Burgovy metody. Výsledné dvojice F1 – F2 Praat ukládá do matice o velikosti $N \times 2$, kde N je počet zkoumaných rámečků v řečovém signálu. [20, 24]

3.2 Filtrování

Mezi výsledky automatického určování formantů mohou být i odlehlé hodnoty. K jejich odhalení se používá Gaussův smíšený model (Gaussian Mixture Model – GMM). Modeluje odhad distribuce F1 – F2 pro každého řečníka. Díky tomu jsou odebrány nepravděpodobné hodnoty. Parametry k výpočtu GMM jsou získány pomocí EM algoritmu (Expectation-Maximization).

Ten funguje ve dvou krocích. V prvním (Expectation) se provede odhad GMM parametrů a výpočet funkce věrohodnosti (likelihood function). V druhém kroku (Maximization) jsou aktualizovány GMM parametry s účelem maximalizovat funkci věrohodnosti. Pro zvýraznění rozdílů je možné výsledky zlogaritmovat.

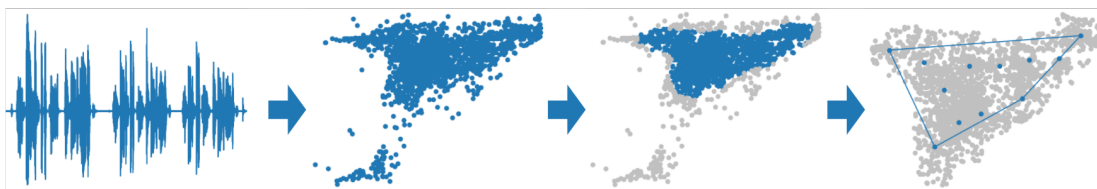
Pro vyřazení odlehlých hodnot Sandoval (2013) používá hranici 0,3 průměru věrohodností všech pozorování. Dvojice F1 – F2 s menší hodnotou filtruje pryč. To je přibližně 15 % z celkového počtu hodnot. [11, 20, 24, 34]

3.3 Shluky

Ze zbývajících dvojic formantů jsou určeny shluky náležící jednotlivým vokálům. K tomu se používá algoritmus k-means, založený na předem známém počtu výsledných shluků (k).

Každému shluku náleží jeden centroid. Ty jsou na počátku zvoleny náhodně. Všechny dvojice F1 – F2 jsou pak přiřazeny vždy k tomu nejbližšímu. Následně je přepočítána poloha všech centroidů tak, aby odpovídala poloze výsledného těžiště bodů F1 – F2 v jednom shluku. Proces se opakuje, dokud nedojde k ustálení. Výhodou tohoto algoritmu je jeho rychlost a spolehlivost. Konvergence je totiž vždy zaručena. Naopak nevýhodou je fakt, že výsledkem algoritmu mohou být různá řešení, odvislá od volby počátečních centroidů.

Sandoval (2013) používá $k = 12$, tedy počet anglických vokálů. Výsledné dvojice F1 – F2 odpovídající centřům shluků ukládá do matice. [14, 20, 24]



Obr. 3.1: Schéma algoritmu automatického výpočtu Hellwagova trojúhelníku podle [20]

4 Parametry založené na formantových kmitočtech

4.1 VSA

VSA je běžným parametrem analýzy vokálů. Blíže byla tato oblast popsána v kapitole 2. Kvůli limitacím této metriky se dnes hledají přesnější parametry, které by ji mohly nahradit. Hodnota VSA jako číselného údaje je pro trojúhelníkový prostor určena z následující rovnice: [33]

$$VSA = \left| \frac{F1_i(F2_a - F2_u) + F1_a(F2_u - F2_i) + F1_u(F2_i - F2_a)}{2} \right| \quad (4.1)$$

4.2 LnVSA

Mírné zlepšení oproti VSA přináší parametr LnVSA. Jeho hodnota je stejně jako u VSA vypočítána z prvních dvou formantů vokálů ve vrcholech trojúhelníka, s tím rozdílem, že u LnVSA jsou osy logaritmické. Použití logaritmu redukuje vliv rozdílů řečníků, i když ne zcela. Výpočet se provádí z následujících rovnic: [22]

$$LnVSA = \sqrt{LnS(LnS - LnED_{iu})(LnS - LnED_{ia})(LnS - LnED_{au})} \quad (4.2)$$

kde:

$$LnED_{iu} = \sqrt{\left(\ln \frac{F1_i}{F1_u}\right)^2 + \left(\ln \frac{F2_i}{F2_u}\right)^2} \quad (4.3)$$

$$LnED_{ia} = \sqrt{\left(\ln \frac{F1_i}{F1_a}\right)^2 + \left(\ln \frac{F2_i}{F2_a}\right)^2} \quad (4.4)$$

$$LnED_{au} = \sqrt{\left(\ln \frac{F1_a}{F1_u}\right)^2 + \left(\ln \frac{F2_a}{F2_u}\right)^2} \quad (4.5)$$

$$LnS = \frac{LnED_{iu} + LnED_{ia} + LnED_{au}}{2} \quad (4.6)$$

4.3 Formant Centralization Ratio (FCR)

Hlavním využitím FCR je rozlišení dysartrické řeči. Výpočet opět vychází z prvních dvou formantů samohlásek určujících vrcholy VSA. Rovnice pro výpočet vypadá následovně:

$$FCR = \frac{F2_u + F2_a + F1_i + F1_u}{F2_i + F1_a} \quad (4.7)$$

Jedná se o poměr, kde se v čitateli nacházejí formantové kmitočty s tendencí růst při vyšší centralizaci a ve jmenovateli naopak klesat. Se zvýšenou centralizací vokálů FCR roste. Z úpravy rovnice, kdy $F2_u = F2_a = F2_i$; a $F1_u = F1_a = F1_i$; vyplývá, že maximální hodnota FCR je 2. Parametr byl navrhnut pro zvýšení citlivosti pozorování centralizace vokálů a minimalizování vlivu variability řečníků. Díky tomu, že jde o poměr, se snižuje právě vliv věku a pohlaví. Tabulka 4.1 publikovaná Sapirem (2010) zachycuje snížení vlivu různých skupin na parametr FCR. [21]

Tab. 4.1: Porovnání vlivu různých skupin na parametry VSA a FCR podle [21]

	F1 _i [Hz]	F2 _i [Hz]	F1 _a [Hz]	F2 _a [Hz]	F1 _u [Hz]	F2 _u [Hz]	VSA [Hz ²]	FCR [-]
Muži	342	2332	768	1333	378	997	264423	0,99
Ženy	437	2761	936	1551	459	1105	399862	0,96
Děti	452	3081	1002	1688	494	1345	448147	0,97
Průměr	410	2721	902	1524	444	1149	370811	0,97
σ	60	381	121	179	60	178	95245	0,01
v_x	15 %	14 %	13 %	12 %	13 %	16 %	26 %	1 %
<i>Pozn.</i> σ = směrodatná odchylka, v_x = variační koeficient								

4.4 Vowel Articulation Index (VAI)

VAI je reciproční hodnota pro FCR. Počítá se podle vzorce:

$$VAI = \frac{F2_i + F1_a}{F2_u + F2_a + F1_i + F1_u} \quad (4.8)$$

Zatímco FCR lépe popisuje míru centralizace formantů, VAI je vhodnější pro porovnávání s VSA a LnVSA, protože stejně jako u těchto parametrů s rostoucí centralizací hodnota VAI klesá. Normální hodnota je blízká 1,0; protože součty kmitočtů v čitateli i jmenovateli jsou si velmi podobné. [22]

4.5 Working Vowel Space Metrics (WVSM)

Na základě VSA je odvozena rodina metod WVSM. Ta nemodeluje trojúhelník podle formantů vzorků vokálů v jeho vrcholech, ale průběžně v čase formanty odhaduje a vytváří spojitý obrazec drah jejich změn. Tato technika bývá aplikována na řečový signál o délce přibližně jedné věty. Výsledkem může být více parametrů. [31]

4.5.1 Vowel Space Hull Area (VSHA)

VSHA je obdobou klasického VSA. Rozdíl je v použité metodě získání vrcholů prostoru. Zatímco VSA je obsah vypočítaný pouze ze tří nebo pěti krajních vokálů Hellwagova trojúhelníku, VSHA uvažuje všechny analyzované vokály díky algoritmu z kapitoly 3. Tím pádem je výsledek daleko přesnější. Parametr je ale náročnější na požadavky. Kritické je obsažení všech typů vokálů v analyzované větě a její shodnost mezi porovnávanými řečníky. [31]

4.5.2 Articulatory-Acoustic Vowel Space (AAVS)

V roce 2014 představili Whitfield a Goberman nový parametr vokálního prostoru – AAVS. Je založen na metodách WVSM, díky čemu je citlivější než klasické VSA. AAVS je standardizovaný všeobecný rozptyl (Standardized General Variance – SGV) dvojrozměrných vzorků F1 – F2. Zobecněný rozptyl (Generalized Variance – GV) je míra zavedená Wilksem (1932) jako skalární míra celkového vícedimensionálního rozptylu (scatter). Pro jakýkoliv p -rozměrný náhodný vektor X , je GV definováno jako determinant jeho variančně-kovarianční matice Σ . SGV je potom kladná p -tá odmocnina GV:

$$SGV = \sqrt[p]{\det \Sigma} \quad (4.9)$$

SGV poskytuje míru formantové variability v prostoru F1 – F2, která je ekvivalentní dvojrozměrné směrodatné odchylce. Zvýšení disperze formantů proto povede ke zvýšení SGV. Díky tomu je to vhodný parametr k analýze rozsahu artikulačního pohybu nebo pracovní oblasti formantů.

Whitfield a Goberman (2014) zkoumali schopnost AAVS sledovat změny srozumitelnosti řeči u pacientů s PN, o které je známo, že je spojena s motorickými poruchami řeči. Řeč pacientů s PN má oproti zdravé řeči významně nižší AAVS. Tento parametr u nich navíc vzrostl po vyzvání k lepšímu vyslovování. To znamená, že jím lze sledovat změny jak v rámci jednoho jedince, tak i v rámci skupiny. [24, 31, 32, 33]

4.5.3 Vowel Space Density (VSD)

Fox & Jacewicz (2017) prokázali, že VSA nedokáže charakterizovat rozdíly v regionálních dialektech americké angličtiny. Navrhli, že ačkoli dva mluvčí z různých oblastí mohou využívat stejnou pracovní vokální oblast, rozdíly ve stylu mluvení lze lépe zachytit vnitřním rozložením bodů $F1 - F2$ (nebo oblastmi spektrální hustoty). Místo označení „vowel space“ používají pro rozložení formantů pojem „formant space“ (prostor formantů), aby zdůraznili, že se jedná o absolutní kmitočtovou distribuci bodů $F1 - F2$ bez ohledu na to, z jaké kategorie vokálů pocházejí. Použitím metod WVSM tento přístup zohledňuje dynamiku formantů, protože lze tvrdit, že pokud se dva mluvčí liší dynamikou formantů, objeví se to také v jejich formantové oblasti jako překrývání regionů různých vokálů.

Aby Fox & Jacewicz (2017) odvodili oblasti spektrální hustoty, začali vykreslováním bodového grafu hodnot $F1 - F2$. Poté z něj vytvořili 3D histogram přidáním třetího rozměru zohledňujícího hustotu bodů dvojic formantů. Dále prostorově vyhladili histogram interpolací založenou na mřížce. Výsledkem je obrysová mapa vyhlazeného histogramu umožňující snadnou představu o vnitřním rozložení formantů ve vokální oblasti.

Tento přístup představuje neparametrickou flexibilní formu odhadu hustoty z neznámého rozložení bodů. Postup, který používali Fox & Jacewicz (2017), je pravděpodobně nejstarší a nejméně sofistikovanou formou odhadu hustoty. Lepší je použití jádrového odhadu hustoty. [9, 24, 25, 31]

4.5.4 Density Percentage (DP)

Density Percentage (procento hustoty) je vlastní parametr vycházející z VSD. Důvodem použití tohoto parametru je fakt, že samotné VSD je vyjádřené pouze graficky (viz obrázek 9.2), a jeho porovnávání s ostatními parametry proto není možné.

Jedná se o procentuální vyjádření plochy, kterou zabírají oblasti s nejhustším rozložením bodů, z celkové zkoumané plochy. Konkrétně se jedná o vrchní pětinu z celkového rozpětí škály hustoty. Předpoklad je, že u řečníků s PN by DP mělo být nižší v důsledku centralizace vokálů.

5 Implementace algoritmu odhadu hran trojúhelníku

Tato část bakalářské práce popisuje skript napsaný v programovacím jazyce Python určený k extrakci formantů a určení hran Hellwagova vokalického trojúhelníku. Jazyk Python byl zvolen pro své všeobecné rozšíření, jednoduché ovládání, a především širokou podporu knihoven a modulů třetích stran. Skript byl napsaný ve vývojovém prostředí Visual Studio Code (verze 1.51.1). Zdrojový kód byl okomentován, aby byl co nejpřehlednější. [28]

Základní kroky principu odhadu hran trojúhelníku jsou založeny na algoritmu Sandovala (2013) popsáném v kapitole 3. Ten k výpočtům používá skriptovací programovací jazyk Matlab. Další rozdíly budou popsány níže. Dalším významným zdrojem inspirace pro tento skript je práce od Stehra (2018) vytvořená v programovacím jazyce R určeném pro statistickou analýzu dat a jejich grafické zobrazení. Při ní sice nebylo dosaženo přesných výsledků (výsledné odhady formantů ve vrcholech neodpovídají skutečným hodnotám), ale postup v ní je téměř shodný s algoritmem Sandovala a v určitých ohledech nabízí i detailnější vysvětlení. [19, 20, 24]

5.1 Vykreslování grafů

K vykreslování grafů je použita sada funkcí `matplotlib.pyplot`. Vychází z Matlabu a přenáší jeho funkční grafické schopnosti do prostředí Pythonu. Základní možnosti využití v tomto skriptu jsou například vynesení bodů do grafu, vykreslení křivek, popis os, zrcadlové převrácení os nebo nastavení barvy výsledné grafiky. [12]

5.2 Práce s maticemi

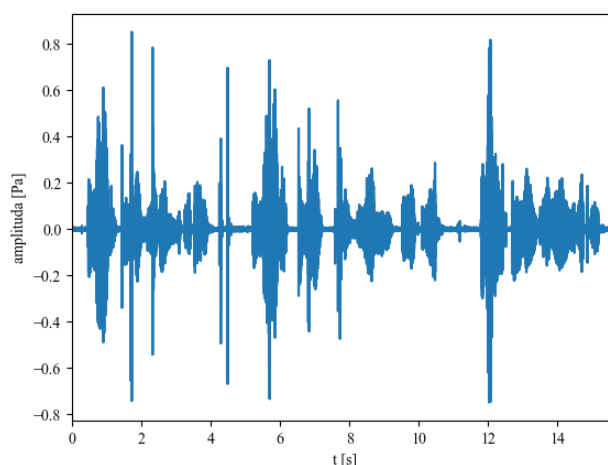
Základní knihovnou (nejen pro tento skript) pro práci s maticemi je knihovna NumPy. Je na ní vystavěna i široká škála dalších knihoven, jako například vědeckotechnicky zaměřená SciPy, nebo výše zmíněná matplotlib. Jejím hlavním přínosem pro tuto práci je zjednodušení práce s poli dat. Umožňuje rychlou úpravu pole jako celku bez nutnosti procházet matici prvek po prvku pomocí cyklů. [10]

5.3 Matematické operace

Přístup k základním matematickým funkcím a konstantám umožňuje modul `math`. Je vystavěn na základech matematické knihovny jazyka C. Obsahuje i funkce pro kontrolu obsahu dat, například kladných/záporných čísel, konečných/nekonečných hodnot nebo hodnot definovaných/nedefinovaných. Toho posledního bylo využito právě v tomto skriptu. [28]

5.4 Využití funkcí prostředí Praat

Praat je volně dostupný software používaný k fonetické analýze řeči. Aby bylo možné využít jeho možnosti přímo z prostředí Pythonu, byla použita jeho Cython verze jménem `Parselmouth`. Cython je nástroj sloužící k převodu zdrojových kódů Pythonu do zdrojového kódu určeného pro překlad standardními překladači jazyka C (ve kterém je napsán právě Praat). Díky tomu `Parselmouth` nabízí kompletní Python knihovnu. Běžným řešením, jak použít možnosti Praatu v prostředí Pythonu, je buď knihovna přímo reimplementující jeho funkce nebo knihovna umožňující komunikaci se skriptovacím jazykem Praatu. Nevýhodou prvního řešení je možná neshoda výsledků funkcí knihovny a funkcí Praatu, protože se nejedná o shodný kód, a tudíž nemusí poskytovat identické výstupy. Nevýhodou druhého řešení je pomalá rychlost zpracování, kvůli probíhající komunikaci mezi dvěma programy a nutnost znalosti skriptovacího jazyka Praatu. `Parselmouth` tedy díky svému řešení pomocí nástroje Cython postrádá všechny tyto nevýhody. [2, 3, 13]

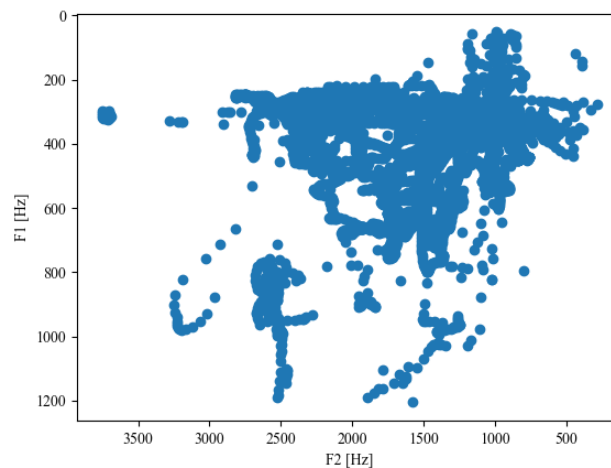


Obr. 5.1: Testovací řečový signál pro analýzu

Hlavním účelem použití Parselmouthu je extrakce formantů. Parametry funkce `.to_formant_burg` převádějící objekt třídy `Sound` na objekt třídy `Formant` jsou nastaveny tak, jak je doporučeno od vývojářů Praatu – maximální počet formantů = 5, maximální hodnota formantu = 5500 Hz, velikost okna = 25 ms, mezní kmitočet preemfázového filtru = 50 Hz. V rámci získaných při zjišťování formantů je určován také základní tón řeči. Ten je ale přítomný jen ve znělých částech analyzované nahrávky. To jsou stejné části, ze kterých mají být načítané hodnoty formantů. Ostatní úseky odpovídají buď tichu nebo neznělým souhláskám. Cyklus `for` tedy prochází data rámeček po rámečku a načítá velikosti získaných formantů jen z úseků se znělou řečí. [3, 13]

Načítaná data musí projít ještě druhou podmínkou. Pokud se pro daný rámeček nepodařilo získat alespoň hodnoty prvních čtyř formantů, výsledky jsou rovněž filtrovány pryč. Tato podmínka zamezuje ukládání dat z rámečků, u kterých nebylo možné určit správné hodnoty nejčastěji kvůli posunutí celého číslování formantů, kdy nedošlo k nalezení prvního formantu. Čtvrtý formant pak leží až v pásmu nad 5500 Hz, ale tam už funkce nehledá. Nedefinované hodnoty jak základního tónu, tak i čtvrtého formantu, jsou zjišťovány pomocí funkce `isnan` knihovny `math`.

Výsledkem těchto kroků je matice `f1f2_mat` obsahující kmitočty prvních dvou formantů všech rámečků, které prošly oběma podmínkami. Jejich rozložení znázorňuje obrázek 5.2.

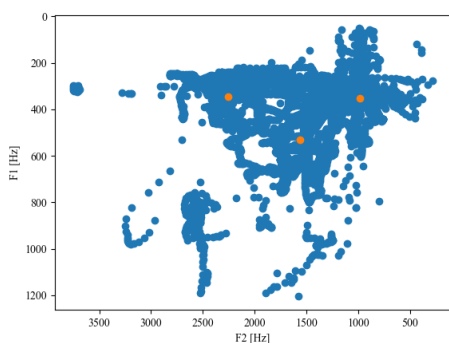


Obr. 5.2: Extrahované dvojice $F1 - F2$

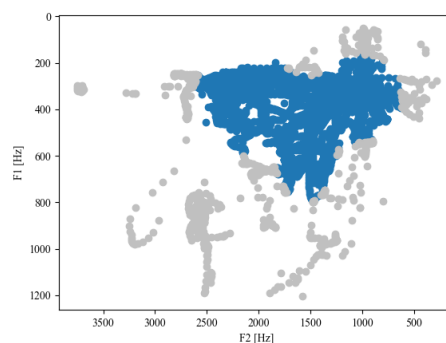
5.5 GMM

Na již zmíněných knihovnách NumPy, SciPy a matplotlib stojí i scikit-learn – knihovna zabývající se strojovým učení. Její balíček `sklearn.mixture` přináší nástroje důležité pro další krok v Sandovalově algoritmu – GMM. Objekt třídy `GaussianMixture` je vytvořen se třemi centroidy (viz obrázek 5.3) reprezentujícími hledaný trojúhelník. To je rozdíl oproti [20], ke se pracuje se 4 centroidy, jelikož pro vokalickou oblast anglického jazyka je charakterističtější lichoběžník. Následně je algoritmu GMM, popsanému v kapitole 3.2, podrobena matice `f1f2_mat`.

Pomocí metody třídy `score.samples` jsou získány logaritmické věrohodnosti dvojic F1 – F2 a díky NumPy je celé pole těchto výsledků pro všechny dvojice formantů převedeno exponenciální funkcí na samotné věrohodnosti. Odfiltrovány jsou pak ty dvojice, jejichž věrohodnost je nižší než 0,3 násobku průměru všech věrohodností v tomto poli. Dvojice F1 – F2 splňující podmínku jsou uloženy do matice `f1f2_filt`. Plošné rozložení odfiltrovaných dvojic a dvojic splňujících podmínku ilustruje obrázek 5.4. [17]



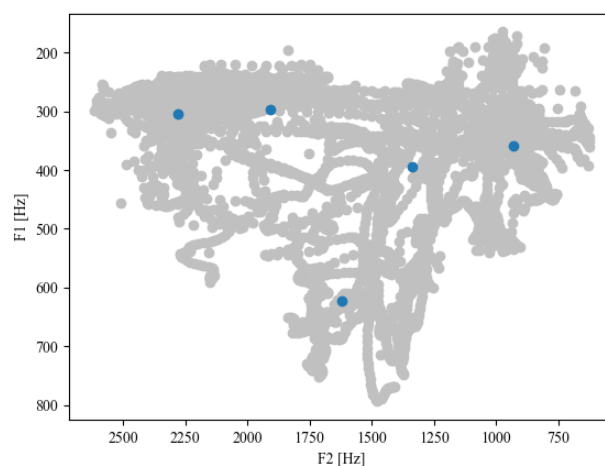
Obr. 5.3: Zvolené centroidy



Obr. 5.4: Odebrané dvojice F1 – F2

5.6 K-means

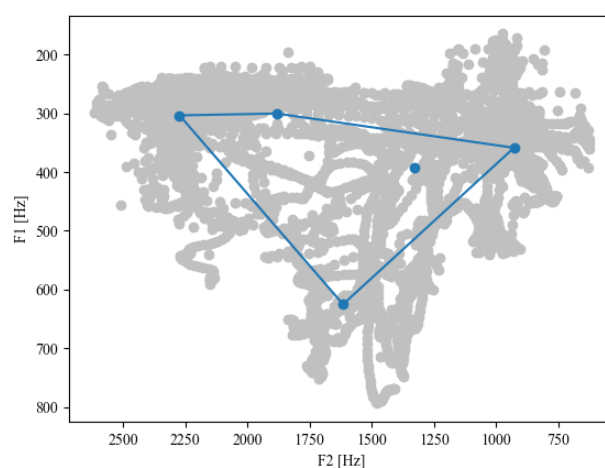
Dalším krokem ve skriptu pro automatický výpočet hran Hellwagova trojúhelníku je algoritmus pro shlukovou analýzu K-means. Ten poskytuje knihovna scikit-learn ve svém modulu `sklearn.cluster`. Jako parametr pro vytvoření instance třídy `KMeans` je předáván odhadovaný počet shluků, tedy počet vokálů v českém jazyce. Díky tomu ve skriptu opět dochází k logickému odklonu od [20]. Následně je tomuto algoritmu předána matice `f1f2_filt`. Výsledkem jsou dvojice formantů F1 – F2 center nalezených shluků uložené do pole `centers`. Nalezená centra lze vidět na obrázku 5.5. [17]



Obr. 5.5: Nalezená centra shluků algoritmem k-means

5.7 Konvexní obal

Aby bylo možné vykreslit vokalickou oblast, je potřeba nalézt nejmenší možný prostor ohraničený dvojicemi F1 – F2 z matice `centers`, který splňuje podmínku, že žádný jeho vnitřní úhel nepřesahuje úhel 180° . Tento prostor se nazývá konvexní obal. K jeho nalezení lze použít třídu `ConvexHull` modulu `scipy.spatial` knihovny SciPy. Jako parametr pro vytvoření instance je předávána matice `centers`. Z atributů vytvořené instance je pak možné graficky ztvárnit konvexní obal znázorňující Hellwagův vokalický trojúhelník jako například na obrázku 8.2. [29]



Obr. 5.6: Vykreslení konvexního obalu

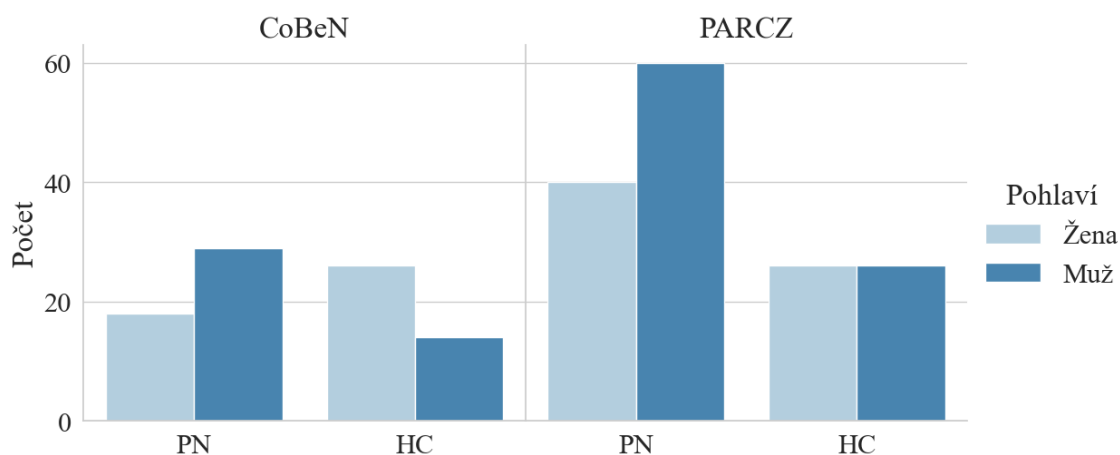
6 Databáze

Jako výchozí data pro tuto práci posloužily zvukové nahrávky ze dvou databází – PARCZ a CoBeN.

PARCZ je databáze, která obsahuje 100 pacientů s PN (40 žen/60 mužů; průměrný věk $67,36 \pm 8,35$) a 52 HC (26 žen/26 mužů; průměrný věk $63,69 \pm 9,09$). Byla zaznamenána v I. Neurologické klinice Fakultní nemocnice u sv. Anny v Brně v rámci projektu Ministerstva zdravotnictví ČR č. NT13499 (Řeč, její poruchy a kognitivní funkce u Parkinsonovy nemoci). Demografické a klinické informace subjektů je možné vidět v tabulkách 6.1, 6.2, 6.3 a 6.4. Zkratka LED reprezentuje denní levodopa-ekvivalentní dávku (Levodopa Equivalent Dose), tedy dopaminergní medikaci. UPDRS III zastupuje Unified Parkinson's Disease Rating Scale, což by se dalo přeložit jako jednotná škála hodnocení PN. Jedná se o její třetí část, která je konkrétně věnována vyšetření motorických příznaků PN.

CoBeN databáze, která obsahuje 47 pacientů s PN (18 žen/29 mužů; průměrný věk $69,89 \pm 7,89$) a 40 HC (26 žen/18 mužů; průměrný věk $67,50 \pm 5,63$), byla zaznamenána ve Skupině aplikovaných neurověd Středoevropského technologického institutu v rámci projektu Ministerstva zdravotnictví ČR č. NV16-30805A (Efekt neinvazivní stimulace mozku na hypokinetickou dysartrii, mikrografii a mozkovou plasticitu u pacientů s Parkinsonovou nemocí) a projektu EU Horizon 2020 č. 734718 - CoBeN (Novel Network-Based Approaches for Studying Cognitive Dysfunction in Behavioral Neurology). Demografické a klinické informace subjektů je možné vidět v tabulkách 6.5, 6.6, 6.7 a 6.8.

Všichni pacienti byli zaznamenáni v tzv. ON stavu, tj. přibližně 1 hodinu po podání dopaminergní léčby. Studie byla schválena lokální Etickou komisí a všichni účastníci podepsali informovaný souhlas.



Obr. 6.1: Demografické zastoupení v použitých databázích

Tab. 6.1: PARCZ – Věk

Skupina	Počet	Průměr	σ	Min	Q1	Medián	Q3	Max
PN	100	67,36	8,3503	46	62,75	67,5	72	87
Muži	60	66,2833	8,6261	46	62	67	71,25	87
Ženy	40	68,9750	7,7443	49	64	70	73	86
HC	52	63,6923	9,0890	45	59	63,5	68,25	87
Muži	26	65,5769	8,9002	49	59	65	73	83
Ženy	26	61,8077	9,0511	45	57,5	63	65	87

Tab. 6.2: PARCZ PN – LED

Skupina	Počet	Průměr	σ	Min	Q1	Medián	Q3	Max
Celek	100	1597,91	987,87	0	868,75	1356,5	2102,75	4954
Muži	60	1816,73	996,02	56	1043,50	1715,0	2490,50	4103
Ženy	40	1269,68	890,12	0	704,25	1119,5	1662,50	4954

Tab. 6.3: PARCZ PN – UPDRS III

Skupina	Počet	Průměr	σ	Min	Q1	Medián	Q3	Max
Celek	100	24,91	11,97	3	14,75	25,5	33,00	55
Muži	60	26,83	10,47	5	20,00	27,0	33,25	55
Ženy	40	22,03	13,56	3	10,75	21,5	31,25	55

Tab. 6.4: PARCZ PN – délka trvání PN

Skupina	Počet	Průměr	σ	Min	Q1	Medián	Q3	Max
Celek	100	7,80	4,4153	1	4	7	11	22
Muži	60	7,87	4,2724	1	5	7,5	11	21
Ženy	40	7,70	4,6751	2	4	7	11	22

Tab. 6.5: CoBeN – Věk

Skupina	Počet	Průměr	σ	Min	Q1	Medián	Q3	Max
PN	47	69,89	7,8940	53	64	71	77,00	84
Muži	29	69,45	8,1868	53	63,25	71	75,50	84
Ženy	18	69,78	8,4753	57	62	71	76,50	84
HC	40	67,50	5,6250	60	63	67	71,25	83
Muži	14	66,71	5,4973	60	62	66	71,75	76
Ženy	26	67,92	5,7545	60	64	67	70,75	83

Tab. 6.6: CoBeN PN – LED

Skupina	Počet	Průměr	σ	Min	Q1	Medián	Q3	Max
Celek	47	900,000	487,50	0	500	900	1274,5	1850
Muži	29	897,375	463,52	80	500	905	1246	1850
Ženy	18	783,778	466,14	80	455	782,5	1069,5	1557

Tab. 6.7: CoBeN PN – UPDRS III

Skupina	Počet	Průměr	σ	Min	Q1	Medián	Q3	Max
Celek	47	12,596	7,4386	0	8	13	16,50	29
Muži	29	13,525	7,3832	0	9	14	17,25	29
Ženy	18	12,556	8,7530	0	6,25	15	17,75	28

Tab. 6.8: CoBeN PN – délka trvání PN

Skupina	Počet	Průměr	σ	Min	Q1	Medián	Q3	Max
Celek	47	6,574	5,5315	1	2	4	9,50	20
Muži	29	6,725	5,5423	1	2	5	9,25	20
Ženy	18	5,778	4,9056	1	2,25	4	7,75	20

7 Metodologie

Vzhledem k odlišnostem v nahrávkách jednotlivých databází musí být výsledky z každé databáze zpracovávány zvlášť.

Nejprve jsou z nahrávky každého řečníka extrahovány vrcholy konvexního obalu, tak jak bylo popsáno v kapitole 5. Následně jsou z těchto dvojic prvního a druhého formantu vybrány ty, které reprezentují hlásky [i], [a] a [u]. Jako [i] je vybrána dvojice s nejvyšším kmitočtem druhého formantu, jako [u] naopak nejnižším. Hlásku [a] reprezentuje dvojice kmitočtů s nejvyšším prvním kmitočtem. Díky těmto třem vrcholům pomyslného trojúhelníku lze vypočítat základní parametry charakterizující pohyblivost mluvidel – VSA, LnVSA a VAI. Parametr VSHA je vypočítán ze všech vrcholů konvexního obalu. Pro parametr AAVS je použita matice všech hodnot, které prošly filtrací a pro DP a relativní směrodatné odchylky prvního a druhého formantu matice všech hodnot, včetně těch odfiltrovaných. Výsledné hodnoty všech 8 parametrů jsou uloženy společně s ID řečníka do souboru .csv. Pro databázi PARCZ je výstupním souborem těchto kroků soubor PAR_3_03_13.csv a pro databázi CoBeN soubor COBEN_3_03_13.csv.

7.1 Statistická analýza

V dalším kroku jsou hromadně zpracovávány vypočítané parametry pro všechny řečníky. Nejprve je ze všech výsledků pomocí lineární regrese odstraněn vliv věku a pohlaví. Následně je lineární regrese použita znovu, ale už jen na řečníky s PN a je díky ní odstraněn vliv medikace (LED). Podle zdrojové databáze je použit vhodný soubor s kovariáty, buďto PAR_info.csv pro PARCZ nebo COBEN_info.csv pro CoBeN.

Pro každou skupinu zvlášť jsou určeny deskriptivní statistiky jednotlivých parametrů. Poté všechny hodnoty vždy jednoho parametru od obou skupin řečníků podstoupí neparametrický Mann-Whitneyův U test (M-W test). Ten posuzuje statistickou významnost vlivu PN na daný parametr. Jako hladina významnosti byla zvolena hodnota $p = 0,05$. Pro každý parametr je také vykreslen houslový graf porovnávací rozložení dat řečníků obou skupin.

7.2 Strokové učení

V závěrečném kroku jsou na vypočítaných parametrech pro hodnocení hybnosti mluvidel trénovány modely strojového učení. Nejdříve byla provedena explorativní analýza, v rámci které byly testovány různé klasifikační algoritmy. Díky tomu můžeme poznat, kterými algoritmy by se bylo vhodné zabývat do budoucna. Kód pro explorativní analýzu a její vyhodnocení byl převzat z [6] a mírně upraven pro účely této práce. V rámci algoritmu jsou zahrnuty lineární, nelineární i souhrnné modely strojového učení. Z těch lineárních jsou použity: logistická regrese (Logistic), hřebenová regrese (Ridge, pro regularizační síly od 0,1 do 1,0), gradientní metody prvního řádu (SGD, stochastic gradient descent, s maximálním počtem průchodů trénovacími daty rovným 1000) a pasivně agresivní klasifikátor (PA, passive-aggressive, opět s maximálním počtem průchodů trénovacími daty rovným 1000). Z nelineárních metod jsou vybrány: algoritmus k-nejbližších sousedů (k-NN, k-nearest neighbors, pro k od 1 do 21), klasifikátor rozhodovacího stromu (CART, classification and regression tree), klasifikátor extrémně náhodného stromu (ET, extra tree), metoda podpůrných vektorů (SVM, support vector machines – pro lineární a poly typ jádra a regularizační parametr od 0,1 do 1,0) a naivní Bayesův klasifikátor (Bayes). Souhrnné metody zastupují: adaptivní posilování (ADA, adaptive boosting), tzv. bagging metoda (BAG), náhodný les (RF, random forest), extrémně náhodné stromy (ETS, extra trees) a gradientní posilování (GBM, gradient boosting). Pro všechny souhrnné metody je navíc použito vyhledávání v mřížce pro vyladění optimálních hyperparametrů. Zvolené hodnoty pro rychlost učení u GBM jsou 0,01 a 0,1, u ADA navíc ještě 1,0. Hodnoty pro počet odhadů/stromů jsou 10 a 100 pro BAG, 100 a 200 pro ADA a GBM, 50 a 100 pro ETS a pro RF stejné, ale navíc ještě 150. Počty dílčích vzorků jsou stanoveny na 0,1; 0,5; 0,7 a 1,0 pro metody GBM a BAG. Poslední hyperparametr maximální hloubka má zvolené hodnoty 3, 7, 9 a None pro libovolnou hloubku. Tyto hodnoty využívají metody RF a ETS. GBM využívá pro maximální hloubku stejné hyperparametry, akorát místo None používá 1.

Veškeré modely strojového učení jsou srovnávány po třikrát opakované desateronásobné křížové validaci (CV, Cross-validation) na základě porovnávací metriky – vyvážené přesnosti (BA, Balanced accuracy).

Pro manipulaci s daty i jejich statistické zpracování je ve většině případů použita knihovna pandas a následné grafické zobrazení výsledků zprostředkovává knihovna seaborn. Na trénování jednotlivých modelů strojového učení jsou využity balíčky knihovny scikit-learn. Výslednými skripty jazyka Python jsou `calculation.py` provádějící extrakci a zpracování dvojic formantů (viz kapitola 5), `functions.py` počítající řečové parametry, `machinelearningmastery.py` trénující modely strojového učení a `processing.py` vykonávající statistickou analýzu získaných výsledků.

7.3 Metody použité v evaluaci systému

7.3.1 Mann-Whitneyův U test

Tento neparametrický test nulové hypotézy pracuje s daty dvou nespárovaných skupin reprezentujících nezávislá měření. Zjišťuje se, zda mají výběry shodný populační medián. Nejprve se vytvoří směsný výběr seřazením prvků obou pokusných zásahů vzestupně do jedné variační řady. Následně jsou všem prvkům přiřazena pořadová čísla a určeny jejich součty pro oba pokusné zásahy. Z těchto součtů jsou pak vypočítány testovací statistiky a určeno testovací kritérium. Podle zvolené hladiny významnosti je pak vypočtené testovací kritérium porovnáváno s tabulkovou kritickou hodnotou. Na základě jejich vztahu je určena statistická významnost testované veličiny. Způsob výpočtu není zcela jednoduchý, a test by se proto neměl používat pro porovnávání dvou výběrů s výrazně odlišným rozdělením. [1]

7.3.2 Lineární metody strojového učení

U lineárních modelů se vystupuje z předpokladu, že cílová hodnota je lineární kombinací jednotlivých vstupních prvků. Tyto metody mohou být považovány za zastaralé, ale to jim nemůže upřít jejich nespornou výhodu – vysokou rychlost zpracování dat, právě díky jejich jednoduchosti. Patří sem také regularizační algoritmy, tedy rozšíření aplikovaná na jinou metodu. Ty penalizují modely na základě jejich složitosti a upřednostňují jednodušší metody. [7]

7.3.3 Nelineární metody strojového učení

Nelineárních metod je více druhů. Použité modely se tedy více liší, protože zastupují jednotlivé kategorie nelineárních metod.

Algoritmy založené na instancích si typicky vytváří databázi ukázkových dat a následně s touto databází porovnávají nová data. Cílem je na základě podobnosti najít nejlepší shodu a díky té vytvořit odhad. Z tohoto důvodu se tyto algoritmy také často nazývají jako učení založené na paměti (memory-based learning). Zde tuto rodinu reprezentuje algoritmus k-nejbližších sousedů a nebo SVM (pracující na odlišném principu).

Metody rozhodovacího stromu vytvářejí model rozhodnutí učiněných na základě hodnot parametrů vstupních dat. Rozhodování se rozvětňuje ve stromových strukturách, dokud se pro danou položku neudělá finální odhad. Zde je reprezentují CART a ET.

Bayesovské metody jsou modely, které výslovně aplikují Bayesovu větu na problémy, jako je klasifikace a regrese. Ta popisuje vzájemný vztah pravděpodobnosti určitého jevu a pravděpodobnosti jevu opačně podmíněného. Zde tyto metody zastupuje naivní Bayesův klasifikátor. [7]

7.3.4 Souhrnné metody strojového učení

Souhrnné modely jsou složené z několika slabších modelů, které jsou samostatně trénovány a jejichž předpovědi jsou určitým způsobem kombinovány pro dosažení celkového odhadu. Velký důraz je kladen právě na výběr typu primitivnějších modelů a na způsob jejich kombinace. Tato třída technik je velmi silná a v dnešní době se těší velké popularitě. Nevýhodou je časová náročnost. [7]

7.3.5 Vyvážená přesnost

Jedná se o metriku, která se využívá při hodnocení kvality binárního klasifikátoru. Obzvláště užitečná je v případě, kdy jsou obě třídy významově stejně důležité, ale co do počtu prvků nevyvážené, tedy když jsou prvky jedné třídy zastoupeny ve výrazně větším počtu než prvky druhé třídy. K takovému jevu dochází v mnoha prostředích, příkladem může být detekce anomálií nebo výskyt nemoci.

BA je založená na dvou dalších běžně používaných metrikách. Jedná se o senzitivitu (SEN) a specificitu (SPE). SEN zjednodušeně odpovídá na otázku: „Kolik pozitivních případů bylo odhaleno?“. SPE pak odpovídá na totéž, ale pro negativní případy. BA uvádí tyto dvě metriky do vzájemného vztahu. [26]

V případě kdy:

- Počet pozitivních prvků určených jako pozitivní = SP (skutečný pozitivní)
- Počet pozitivních prvků určených jako negativní = FN (falešný negativní)
- Počet negativních prvků určených jako pozitivní = FP (falešný pozitivní)
- Počet negativních prvků určených jako negativní = SN (skutečný negativní)

Tak

$$SPE = \frac{SN}{FP + SN} \cdot 100 \quad (7.1)$$

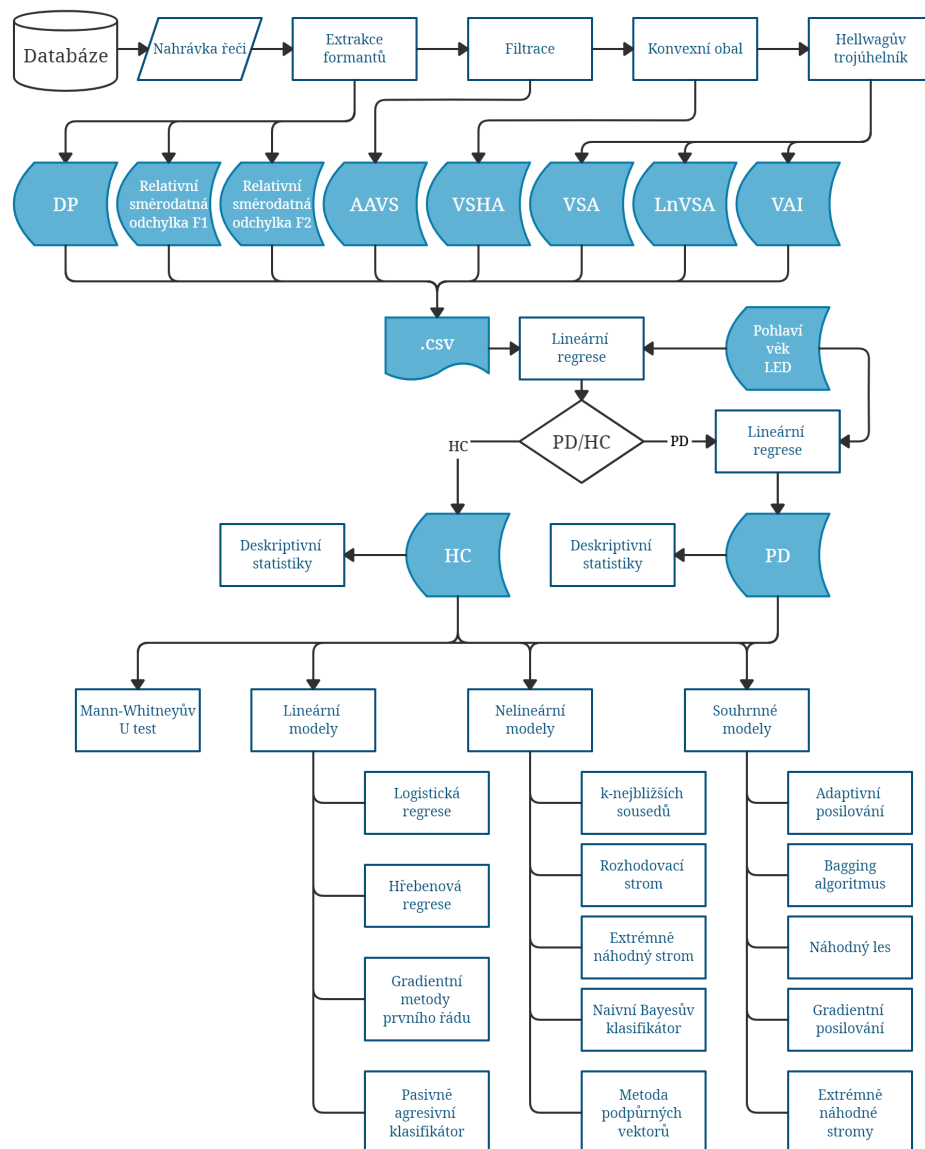
$$SPE = \frac{SP}{SP + FN} \cdot 100 \quad (7.2)$$

A tedy:

$$BA = \frac{SEN + SPE}{2} \quad (7.3)$$

7.3.6 Křížová validace

Křížová validace je přerozdělovací metoda používaná k vyhodnocení modelů strojového učení z omezeného vzorku dat. Primárně se používá v aplikovaném strojovém učení k odhadu dovednosti modelu strojového učení na dosud neviděných datech. To znamená použití omezeného vzorku k odhadu toho, jak se obecně očekává, že model bude fungovat, když se použije k předpovědi na datech nepoužívaných během trénování modelu. Zde je použita k-fold varianta křížové validace. Ta pracuje s množinou prvků, která se rozdělí na k podmnožin. K -krát opakovaný proces vždy použije jednu množinu jako testovací vzorek a zbylé množiny jako trénovací data. [5]



Obr. 7.1: Blokové schéma zpracování dat v této práci

8 Výsledky

Jak již bylo řečeno v kapitole Metodologie, pro každého řečníka v databázi bylo vypočítáno 8 parametrů. Tyto parametry se následně hodnotily a porovnávaly v rámci celé databáze. Vzájemné porovnání mezi databázemi PARCZ a CoBeN není možné, jelikož data obou databází prošla úpravou pomocí lineární regrese. Jelikož hodnoty parametrů pro každou databázi procházely očištěním od kovariátů zvláště, může se lišit jejich výsledná hodnota tak, že srovnání nelze uskutečnit. Na druhou stranu díky této úpravě není v této kapitole nutné rozlišovat například pohlaví řečníků a lze tedy posuzovat sílu jednotlivých parametrů v rámci celé jejich množiny.

8.1 Samostatné výsledky jednotlivých parametrů

Parametr VSA dosahuje lepších výsledků u databáze CoBeN. Hodnota pravděpodobnosti p se zde řádově blíží hladině významnosti. Hodnota vypočítaná pro PARCZ zamítá statisticky významný rozdíl. U obou databází lze sledovat velmi široké rozpětí mezi minimální a maximální získanou hodnotou parametru jak u PN, tak i u HC. Pro CoBeN PN se jedná o překročení dvou řádů. Výsledky shodně vykazují, že hodnota VSA vychází vyšší pro HC. Veškeré vypočítané deskriptivní statistiky pro tento parametr je možné vidět v tabulce 8.1.

Tab. 8.1: Deskriptivní statistiky - VSA

Skupina	Průměr	σ	Min	Q1	Medián	Q3	Max
PAR-PN	148265	65152	18893	111194	140219	177494	387076
PAR-HC	159281	73340	40066	109763	145948	186904	412358
CBN-PN	199019	77269	8862	143725	198790	253508	369579
CBN-HC	235553	103173	46765	174393	231225	287054	519603
$p_{PAR} = 0.25892$ VSA podle M-W testu nedosahuje statistické významnosti							
$p_{CBN} = 0.07206$ VSA podle M-W testu nedosahuje statistické významnosti							

Taktéž parametr LnVSA přináší lepší výsledky u databáze CoBeN. Zde již hodnota p rovnající se 0,042320 dosahuje hladiny statistické významnosti. Pro PARCZ Mann-Whitneyův U test shledává rozdělení PN a HC shodná. Opět lze pozorovat rozdíl mezi minimální a maximální hodnotou parametru v rozpětí řádu. Výsledky obou databází potvrzují teorii, že LnVSA je vyšší u HC. Podrobnější data je možné získat z tabulky 8.2.

Tab. 8.2: Deskriptivní statistiky - LnVSA

Skupina	Průměr	σ	Min	Q1	Medián	Q3	Max
PAR-PN	0,20340	0,07854	0,02666	0,15903	0,19770	0,25159	0,5136
PAR-HC	0,21309	0,09135	0,07304	0,14774	0,20914	0,25285	0,62184
CBN-PN	0,27419	0,09000	0,08794	0,19902	0,27231	0,35493	0,45599
CBN-HC	0,31127	0,10198	0,07270	0,25811	0,30298	0,39153	0,50972
$p_{PAR} = 0.38000 \dots LnVSA$ podle M-W testu nedosahuje statistické významnosti							
$p_{CBN} = 0.04230 \dots LnVSA$ je podle M-W testu statisticky významnou veličinou							

Parametr VAI přináší velmi podobné výsledky pro obě testované databáze. Hodnoty pravděpodobnosti získané při Mann-Whitneyho U testu zde činí 0,05671 pro PARCZ a 0,06530 pro CoBeN a oba výsledky se tedy pohybují na hranici statistické významnosti. Zároveň z tabulky 8.3 vyplývá, že hodnota VAI je opět vyšší pro HC, přestože u PARCZ je maximální hodnoty dosaženo u řečníka s PN.

Tab. 8.3: Deskriptivní statistiky - VAI

Skupina	Průměr	σ	Min	Q1	Medián	Q3	Max
PAR-PN	0,93451	0,09246	0,70300	0,88126	0,94373	0,99830	1,15914
PAR-HC	0,96061	0,06205	0,83466	0,92349	0,97066	1,00773	1,09008
CBN-PN	1,01108	0,09634	0,74278	0,95333	1,00830	1,08449	1,23707
CBN-HC	1,04057	0,09384	0,77341	1,00035	1,02949	1,10561	1,24154
$p_{PAR} = 0.05671 \dots VAI$ je podle M-W testu na hranici statistické významnosti							
$p_{CBN} = 0.06530 \dots VAI$ je podle M-W testu na hranici statistické významnosti							

VSHA se jeví jako silný parametr dosahující statistické významnosti u obou databází, přičemž pro CoBeN je hodnota pravděpodobnosti p rovna 0,01789 a velmi se blíží hranici 0,01, od kdy by se o VSHA mohlo mluvit jako o statisticky velmi významné veličině. I tento parametr vykazuje nepřirozeně velké rozpětí mezi minimem a maximem. Hodnoty VSHA opět vycházejí vyšší pro HC.

Parametr AAVS dosahuje velmi dobrých výsledků u obou databází. U PARCZ je dosaženo nejmenší hodnoty pravděpodobnosti p vůbec a to 0,0038. Vyšší hodnota vychází stejně jako doposud u všech parametrů pro HC. Maximální získanou hodnotu tohoto parametru u PARCZ lze ale pozorovat u řečníka s PN.

Tab. 8.4: Deskriptivní statistiky - VSHA

Skupina	Průměr	σ	Min	Q1	Medián	Q3	Max
PAR-PN	241060	94469	43932	187235	216021	263294	555428
PAR-HC	261327	83820	130359	203853	249430	305312	591942
CBN-PN	272066	84801	66901	214970	262544	332653	449313
CBN-HC	321522	108300	96085	257188	322119	380460	621856

$p_{PAR} = 0.03775$ VSHA je podle M-W testu statisticky významnou veličinou
 $p_{CBN} = 0.01789$ VSHA je podle M-W testu statisticky velmi významnou veličinou

Tab. 8.5: Deskriptivní statistiky - AAVS

Skupina	Průměr	σ	Min	Q1	Medián	Q3	Max
PAR-PN	40622,8	16382,9	11222,6	32251,3	37319,7	46085,0	127944
PAR-HC	44761,7	11946,4	25568,3	36089,5	43268,4	51963,2	92466
CBN-PN	47926,8	15143,1	15531,2	35994,8	48519,8	58358,6	81469
CBN-HC	56703,3	19186,0	23272,4	46400,6	53863,2	63347,3	119191

$p_{PAR} = 0.00384$ AAVS je podle M-W testu statisticky velmi významnou veličinou
 $p_{CBN} = 0.03386$ AAVS je podle M-W testu statisticky významnou veličinou

Mann-Whitneyho U test nedokáže na základě relativní směrodatné odchylky prvního formantu rozlišit PN a HC u PARCZ ani u CoBeN. Rozložení hodnot PN a HC se překrývá tak, že pro CoBeN vychází medián vyšší u PN a u PARCZ naopak u HC.

Tab. 8.6: Deskriptivní statistiky - RSDF1

Skupina	Průměr	σ	Min	Q1	Medián	Q3	Max
PAR-PN	0,64704	0,15987	0,23285	0,54439	0,64860	0,75931	0,96287
PAR-HC	0,63513	0,09522	0,38888	0,55618	0,65275	0,69731	0,81675
CBN-PN	0,62392	0,11090	0,43302	0,55999	0,61504	0,69279	0,86857
CBN-HC	0,59713	0,12205	0,33720	0,52125	0,59080	0,68050	0,85139

$p_{PAR} = 0.22668$... RSDF1 podle M-W testu nedosahuje statistické významnosti
 $p_{CBN} = 0.18673$... RSDF1 podle M-W testu nedosahuje statistické významnosti

Relativní směrodatná odchylka druhého formantu vykazuje lepší výsledky než relativní směrodatná odchylka prvního formantu. U PARCZ sice Mann-Whitneyho U test stále nemůže zamítnout hypotézu o shodném rozdělení, ale u CoBeN již výsledky vypovídají o rozdílném rozložení HC a PN. Pro obě databáze mimo třetí kvartil a maximální vypočítanou hodnotu PARCZ platí, že HC dosahují vyšších hodnot.

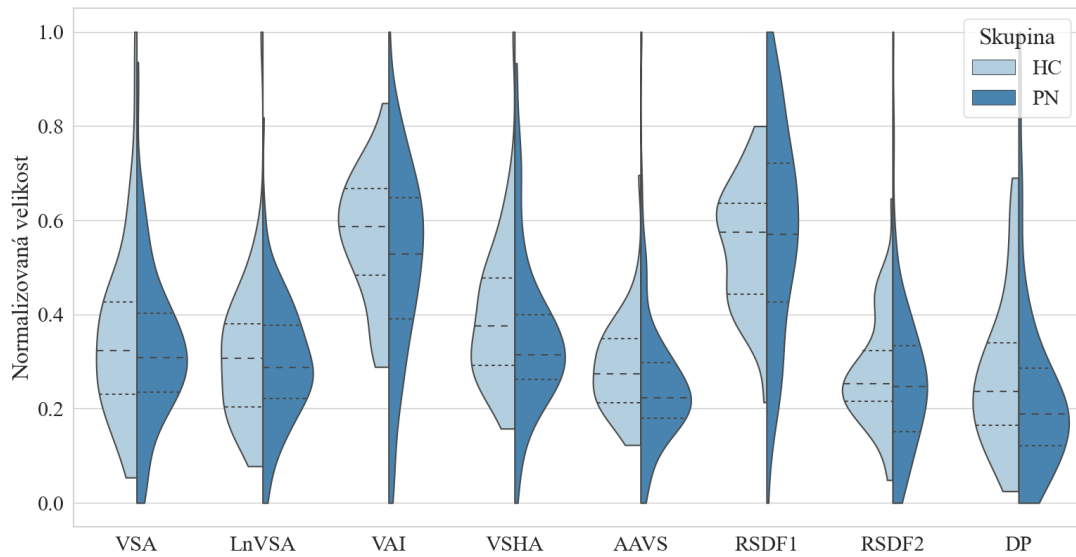
Tab. 8.7: Deskriptivní statistiky - RSDF2

Skupina	Průměr	σ	Min	Q1	Medián	Q3	Max
PAR-PN	0,28005	0,03453	0,22511	0,25709	0,27719	0,29540	0,43558
PAR-HC	0,28397	0,02438	0,23530	0,27042	0,27846	0,29306	0,36109
CBN-PN	0,30247	0,03514	0,22800	0,28119	0,29880	0,33267	0,36287
CBN-HC	0,31949	0,03084	0,25523	0,29471	0,32030	0,33839	0,38230
$p_{PAR} = 0.12093 \dots RSDF2$ podle $M-W$ testu nedosahuje statistické významnosti							
$p_{CBN} = 0.02068 \dots RSDF2$ je podle $M-W$ testu statisticky významnou veličinou							

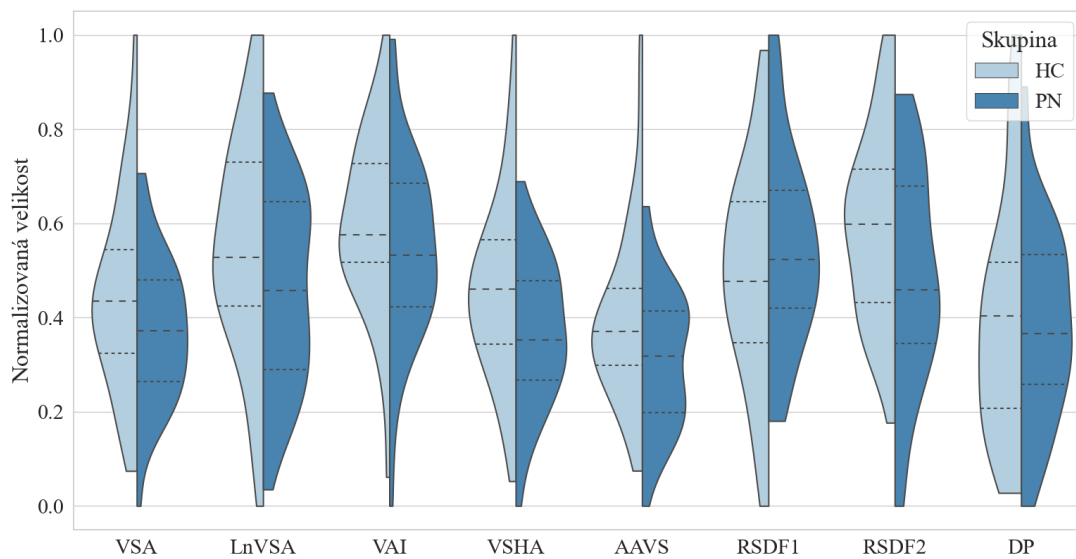
U parametru DP dosahuje hodnota pravděpodobnosti p podstatně lepších hodnot u databáze PARCZ, kde se rovná 0,01384, blíží se tak hranici statistické významnosti. U CoBeN DP nedosahuje statistické významnosti. Mimo maximum u PARCZ dosahuje tento parametr vyšších hodnot u HC.

Tab. 8.8: Deskriptivní statistiky - DP

Skupina	Průměr	σ	Min	Q1	Medián	Q3	Max
PAR-PN	0,36678	0,24306	0,07136	0,22728	0,31363	0,43752	1,35331
PAR-HC	0,42292	0,20918	0,10311	0,28201	0,37386	0,50625	0,95607
CBN-PN	0,46245	0,21124	0,05998	0,32605	0,43771	0,61171	0,98022
CBN-HC	0,47851	0,27265	0,08853	0,27391	0,47735	0,59383	1,09381
$p_{PAR} = 0.01384 \dots DP$ je podle $M-W$ testu statisticky velmi významnou veličinou							
$p_{CBN} = 0.42068 \dots DP$ podle $M-W$ testu nedosahuje statistické významnosti							



Obr. 8.1: Porovnání parametrů na základě výsledků z PARCZ

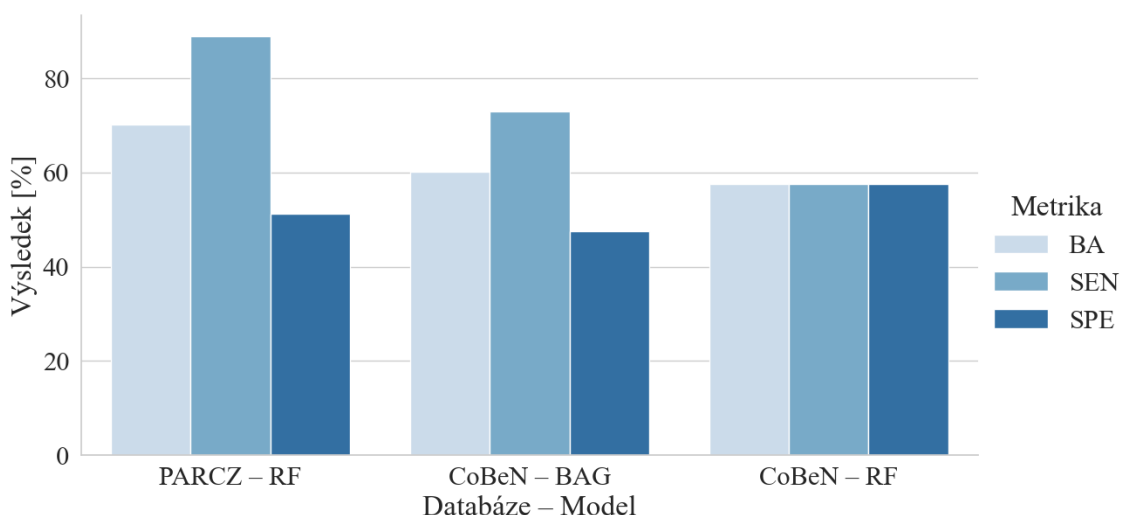


Obr. 8.2: Porovnání parametrů na základě výsledků z CoBeN

8.2 Výsledky modelování pomocí strojového učení

Pro databázi PARCZ dosahují obecně nejvyšších hodnot BA souhrnné metody strojového učení. Vůbec nejlepší výsledek má model RF, u kterého po třikrát opakované desateronásobné křížové validaci vyšla průměrná BA 70,2 %. V tabulce 8.9 lze vidět deset nejlepších výsledků různých modelů strojového učení po optimalizaci hyperparametrů u každého z nich. U všech modelů dosahuje SEN vyšších hodnot než SPE. Nejvyšší získaná SPE rovna 56,7 % byla vypočítána u GBM. Na obrázku 8.4 je možné vidět, že výsledky pro ADA a GBM jsou rovněž konzistentnější než u ostatních modelů strojového učení. Tabulka 8.10 uvádí nejlepší získané výsledky vůbec. Nejvíce zastoupeny jsou modely ADA a RF. Graficky jsou tyto výsledky zobrazeny na obrázku 8.5.

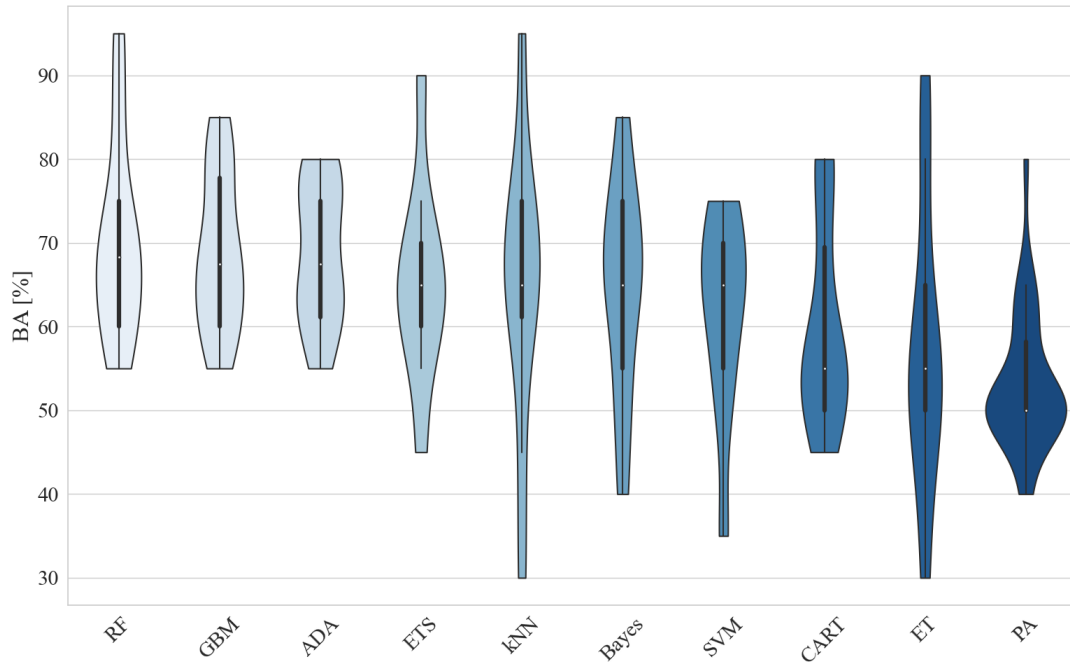
Modely strojového učení trénované na databázi CoBeN dosahují celkově horších výsledků než modely trénované na databázi PARCZ. BA větší než 60 % dosahuje i po vyladěné optimálních hyperparametrů pouze jediný model a tím je BAG s výslednou hodnotou 60,2 %. Hodnoty SPE vycházejí podobné jako u databáze PARCZ. Výjimkou je model kNN u kterého SPE dosáhla 72,5 % a přesáhla i hodnotu SEN = 42,0 %. Pořadí nejlepších výsledků různých metod lze vidět v tabulce 8.11 a na obrázku 8.6. V tabulce 8.12 je uvedeno deset natrénovaných modelů s nejlepšími výsledky vůbec. Ty jsou od výsledků získaných z databáze PARCZ nižší přibližně o 10 %. Nejvíce zastoupeným modelem je BAG s RF, které bylo u PARCZ první, na druhém místě (viz obrázek 8.3). Model ADA dosahuje v několika případech i BA = 100 %, což lze vyčíst i z obrázků 8.6 a 8.7. Tento model však vykazuje také velmi vysokou směrodatnou odchylku, na rozdíl například od BAG s nejlepší průměrnou BA.



Obr. 8.3: Srovnání nejlepších výsledků obou databází

Tab. 8.9: Různé modely trénované na PARCZ

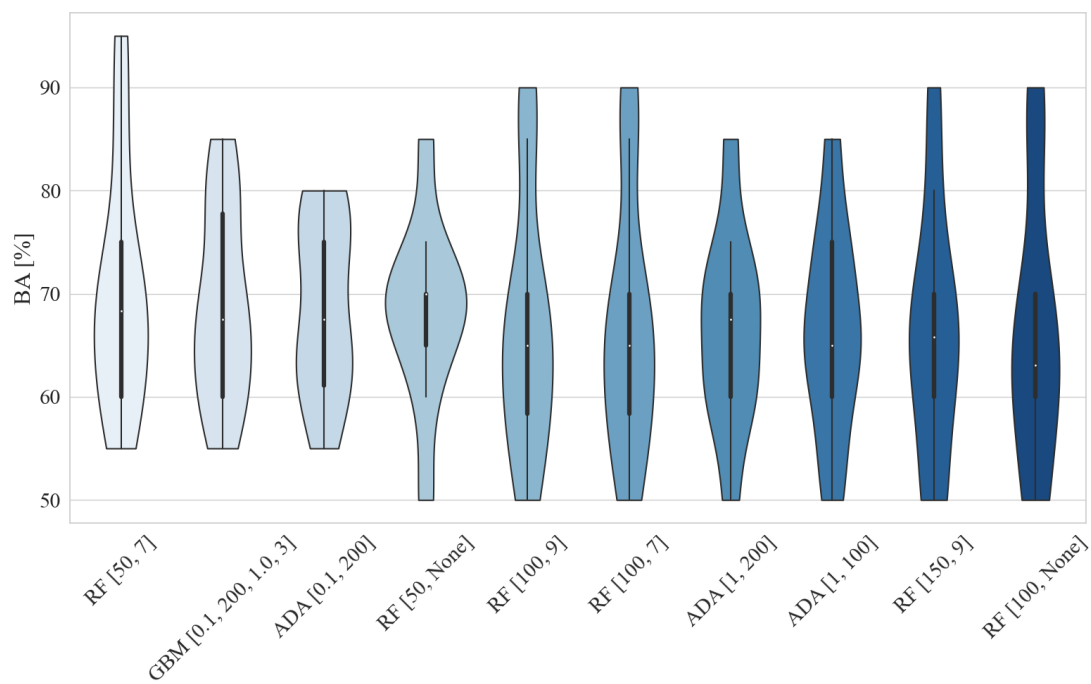
Model	BA [%]	SEN [%]	SPE [%]
RF	70,2	89,0	51,3
GBM	68,8	80,9	56,7
ADA	68,6	83,9	53,3
ETS	65,2	85,0	45,3
kNN	64,6	75,9	53,3
Bayes	64,5	75,7	53,3
SVM	61,7	86,0	37,3
CART	59,9	70,9	49,0
ET	57,2	70,7	43,7
PA	53,1	68,3	33,3



Obr. 8.4: Různé modely trénované na PARCZ

Tab. 8.10: Nejlepší modely trénované na PARCZ

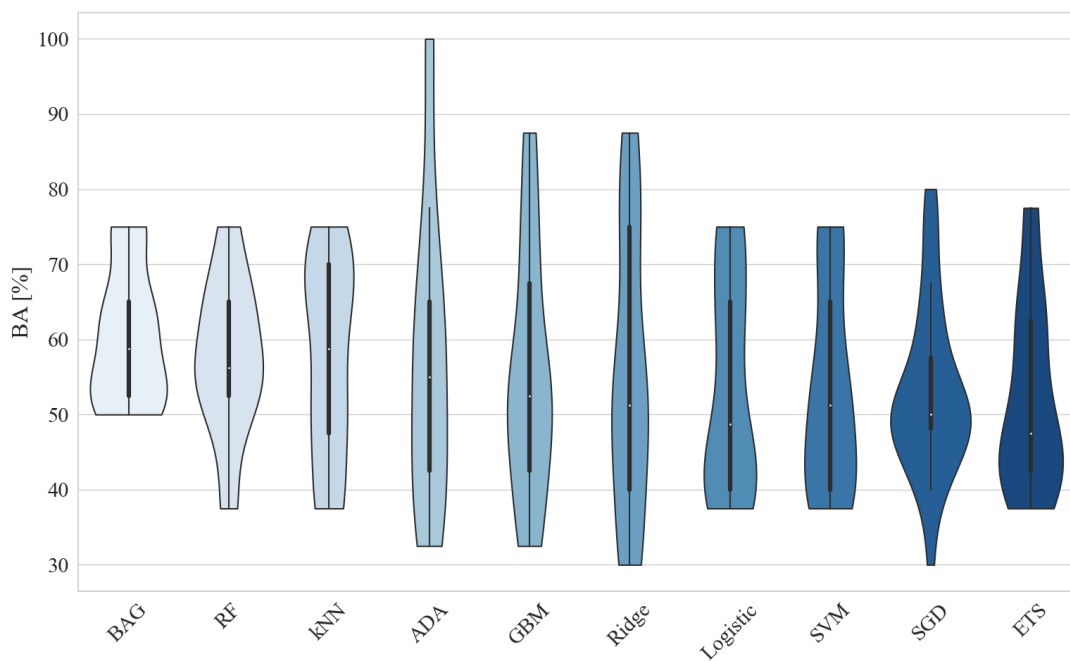
Model [hyperparametry]	BA [%]	SEN [%]	SPE [%]
RF [50, 7]	70,2	89,0	51,3
GBM [0.1, 200, 1.0, 3]	68,8	80,9	56,7
ADA [0.1, 200]	68,6	83,9	53,3
RF [50, None]	68,2	85,0	51,3
RF [100, 9]	66,8	90,0	43,7
RF [100, 7]	66,8	90,0	43,7
ADA [1, 200]	66,5	78,0	55,0
ADA [1, 100]	66,4	77,9	55,0
RF [150, 9]	66,2	87,0	45,3
RF [100, None]	66,1	86,9	45,3



Obr. 8.5: Nejlepší modely trénované na PARCZ

Tab. 8.11: Různé modely trénované na CoBeN

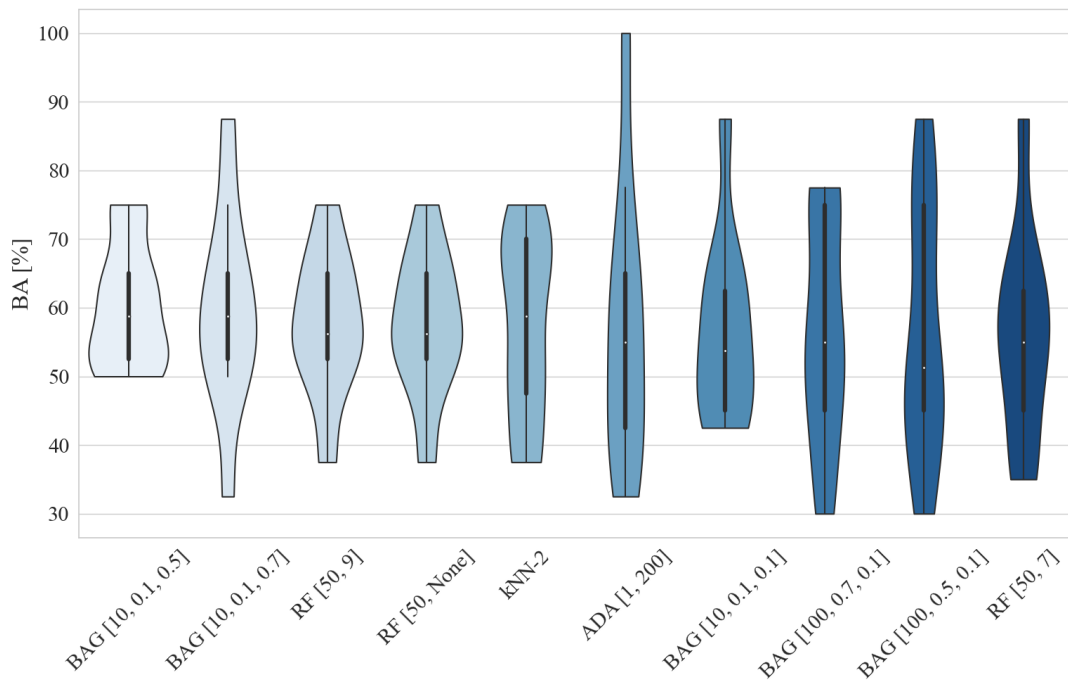
Model	BA [%]	SEN [%]	SPE [%]
BAG	60,2	73,0	47,5
RF	57,5	57,5	57,5
kNN	57,3	42,0	72,5
ADA	57,0	64,0	50,0
GBM	55,3	68,0	42,5
Ridge	55,2	68,0	42,5
Logistic	53,5	69,5	37,5
SVM	53,3	74,0	32,5
SGD	52,8	53,0	60,0
ETS	52,0	64,0	40,0



Obr. 8.6: Různé modely trénované na CoBeN

Tab. 8.12: Nejlepší modely trénované na CoBeN

Model [hyperparametry]	BA [%]	SEN [%]	SPE [%]
BAG [10, 0.1, 0.5]	60,2	73,0	47,5
BAG [10, 0.1, 0.7]	59,5	66,5	52,5
RF [50, 9]	57,5	57,5	57,5
RF [50, None]	57,5	57,5	57,5
kNN-2	57,3	42,0	72,5
ADA [1, 200]	57,0	64,0	50,0
BAG [10, 0.1, 0.1]	56,8	66,0	47,5
BAG [100, 0.7, 0.1]	56,5	68,0	45,0
BAG [100, 0.5, 0.1]	56,5	68,0	45,0
RF [50, 7]	56,0	62,0	50,0



Obr. 8.7: Nejlepší modely trénované na CoBeN

9 Diskuze

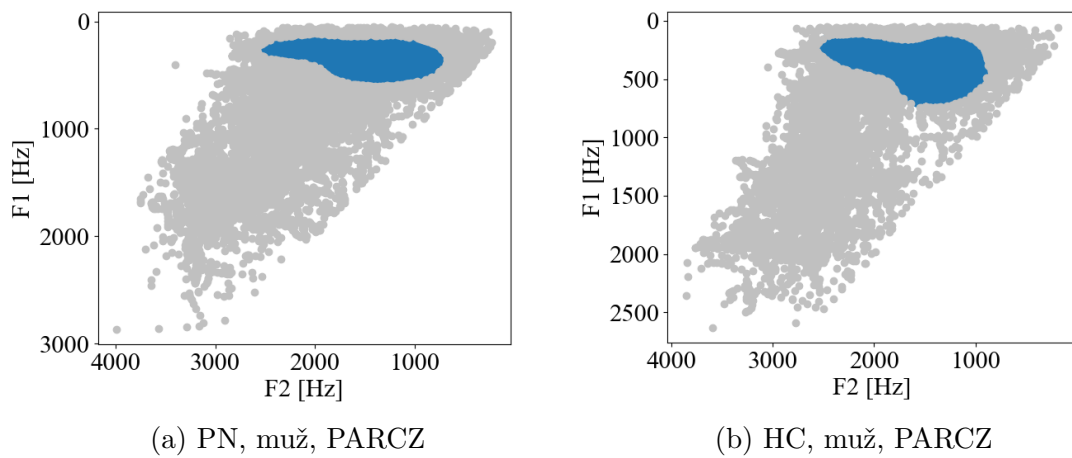
Jako parametry s nejvyšší diskriminační silou se podle výsledků deskriptivní statistiky jeví VSHA a AAVS. To, že se parametry VSA a LnVSA neprojevily tak dobře jako VSHA, může být důkazem významnosti velikosti celé vokální oblasti samohlásek včetně oblastí „oříznutých“ při vykreslování Hellwagova vokalického trojúhelníku. Rozhodující tedy nemusí být pouze poloha formantů v jeho vrcholech, ale celkové rozložení všech dvojic $F1 - F2$. Podobné výsledky uvádí i studie [31] z roku 2019. Ta dosahuje nejvyšších rozlišovacích schopností právě u druhého zmíněného parametru AAVS. Velmi kladně tato studie hodnotí i rozlišování PN a HC za pomoci pozorování rozdílů velikostí oblastí ve VSD. Lze říct, že to v této práci reprezentuje parametr DP. Jeho rozlišovací schopnosti byly ale dostatečně silné pouze u jedné z testovaných databází. Stejně jako ve studiích [32, 33] nevykazují VSA a LnVSA dobré výsledky. Nejshodnější výstup z obou testovaných databází přináší VAI. Oba výsledky potvrzují teorii z kapitoly 4. Hodnoty se skutečně blíží jedné a u řečníků s PN jsou nižší než u HC. Tento rozdíl ale není příliš výrazný. Nižší VAI ve vokální oblasti indikuje vyšší centralizaci dvojic formantů $F1 - F2$. V řeči to může být zaviněno ztuhlostí jazyka nebo mluvidel obecně. Ze zkoumaných relativních odchylek formantů vychází lepší rozlišovací schopnost u relativní směrodatné odchylky druhého formantu. Ten souvisí s předozadností vokálu a je tedy ovlivněn především pohyblivostí jazyka. Tento výstup by tedy mohl podpořit hypotézu o jeho ztuhlosti u pacientů s PN.

Velké rozpětí mezi minimální a maximální vypočítanou hodnotou u parametrů VSA, LnVSA a VSHA je velmi pravděpodobně způsobeno nevhodnou filtrací odlehklých hodnot. Obecně se tento blok v algoritmu zpracování dat jeví jako jedno z jeho slabších míst. Parametry filtrace silně ovlivňují vypočítané hodnoty parametrů pro hodnocení hybnosti mluvidel, které se v závislosti na tom ukazují jako velmi citlivé. Nabízí se možnost vytvořit pro filtrování algoritmus založený na parametrech VSHA, potažmo DP, který by propouštěl pouze dvojice formantů $F1 - F2$ z oblastí s normalizovanou hustotou vnitřního rozložení formantů vyšší, než by byla stanovená hranice. Pomyslné vyhlazené hranice jednotlivých oblastí by mohly poskytnout spravedlivější síto než Gaussův smíšený model. Na obrázku 9.2 lze porovnat rozdíl mezi VSD řečníka s PN a HC a na obrázku 9.1 rozdíl mezi filtracemi pomocí GMM stejných dvou řečníků. Výsledkem filtrace pomocí VSD by například mohly být všechny dvojice formantů obsažené v oblastech s hustotou indikovanou zeleně a výše (větší než 0,4).

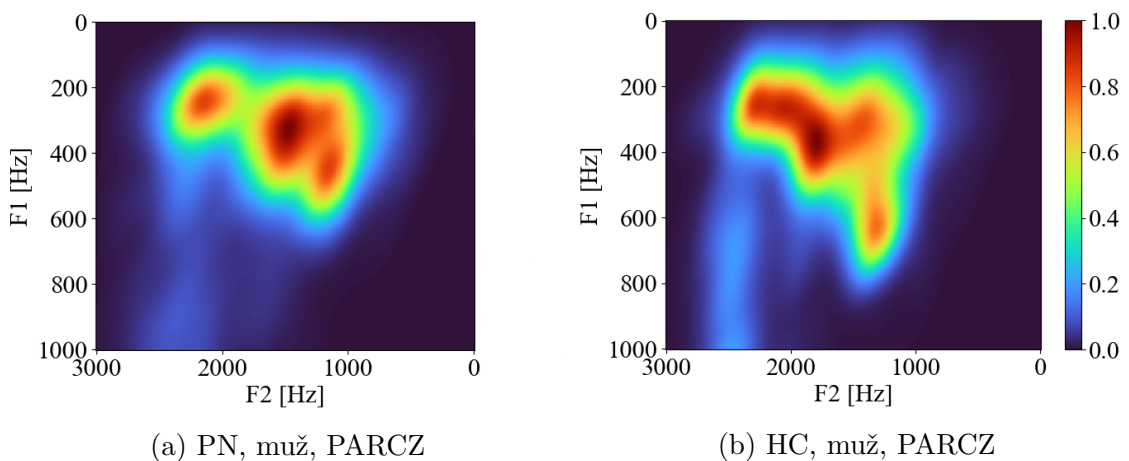
Z trénovaných modelů strojového učení nejlépe vyšly souhrnné metody. Z výsledků získaných z obou databází lze říct, že nejlepší rozlišovací schopnost má algoritmus RF. Platí, že SEN je v drtivé většině případů vyšší než SPE. To může být zaviněno faktem, že obě databáze obsahují více řečníků s PN než HC a nabízejí tedy

lepší podmínky pro natrénování modelu na jejich poznávání. Tomu by nasvědčovaly i lepší výsledky u databáze PARCZ, která obsahuje téměř dvojnásobný počet řečníků s PN oproti HC, viz obrázek 6.1. U databáze CoBeN je počet řečníků s PN vyšší pouze o 17,5 % oproti HC. Početní zastoupení jednotlivých skupin řečníků patrně z tabulek 6.1 a 6.5 by odpovídalo i výsledkům SPE a SEN. Podobný počet HC v obou databázích by se odrazil na podobné SPE a rozdílná SEN by byla způsobená vyšším počtem pacientů s PN u PARCZ.

Mimo způsob filtrování odlehlých hodnot jsou tedy silnou limitací této práce i malé databáze nahrávek řeči. Jejich vyšší počet by poskytl jak objektivnější zhodnocení samostatných parametrů, tak i lepší podmínky pro trénování modelů strojového učení.



Obr. 9.1: Srovnání filtrace pomocí GMM u PN a HC



Obr. 9.2: Srovnání VSD u PN a HC

Závěr

Tato práce se zabývá popisem formantů, Hellwagova vokálního trojúhelníku, parametrů artikulace lidské řeči a algoritmů pro jejich výpočty. V první kapitole je popsán vliv dutin řečového ústrojí člověka na jeho hlas. Je představen pojem formant a vysvětlena jeho funkce, využití i metody určení. Dále jsou uvedeny rozdíly mezi jednotlivými vokály českého jazyka, jak po stránce jejich tvorby, tak i odlišnosti ve formantech. Následující kapitola zavádí pojem Hellwagův vokální trojúhelník. Je zde vysvětlena jeho podstata a původ. Dále je představena jeho akustická paralela v kmitočtovém prostoru a její využití včetně limitací, které přináší. Navazující kapitola je věnována algoritmu pro její automatický výpočet. Pozornost je zaměřena na dílčí úkony tohoto algoritmu, jako je extrakce formantů z řečového signálu, odfiltrování odlehlých hodnot a nalezení center shluků dvojic $F1 - F2$. V kapitole „Parametry založené na formantových kmitočtech“ je shrnuta problematika metrik jako například VSA, VAI nebo metod rodiny WVSM. Důraz je kladen především na způsoby výpočtu a využití. Ze všech těchto poznatků vystupuje část týkající se návrhu skriptu v jazyce Python pro určení hran Hellwagova vokálního trojúhelníku. Podrobně jsou vysvětleny jednotlivé kroky, včetně přiblížení použitých funkcí, modulů a knihoven. Další kapitoly jsou věnovány praktickému výzkumu a přibližují použité zdrojové databáze nahrávek řeči pacientů s PN a HC, metody analýzy parametrů z nich vypočítaných a celkově kompletní proces zpracování dat od začátku do konce. Získané výsledky poté prezentuje další kapitola, kde je vše přehledně shrnuto především formou tabulek a grafů. Závěrečná část se zabývá polemikou nad výstupními daty a pokouší se je zhodnotit a odůvodnit.

Prvním cílem této bakalářské práce bylo vyhledat a popsat doposud používané postupy při zkoumání pohyblivosti mluvidel. Díky řadě vědeckých článků se podařilo sepsat a zhodnotit několik více či méně používaných parametrů. Především se povedlo vyzdvihnout rozdíly v jejich využití, poukázat na vhodnou implementaci a najít limitace každého z nich.

Druhým cílem bylo navrhnout a implementovat způsob určení Hellwagova vokálního trojúhelníku z řečové nahrávky. I tento cíl se podařilo splnit. Výsledkem je skript `calculation.py`. Jeho podrobný popis lze nalézt v kapitole 5, která mu je věnovaná. Navržený algoritmus je obdobou zahraničních vědeckých postupů, přizpůsobený českému jazyku (došlo k optimalizaci). Podmínkou pro vstupní řečový signál je vyváženost zastoupení všech vokálů. Pro porovnávání více řečníků je nezbytná i shodnost vět nebo textových úseků, které čtou. S tímto cílem se nabízí pracovat i do budoucna. Bylo by možné kompletně přepracovat způsob filtrování odlehlých hodnot, který ve své stávající podobě silně ovlivňuje citlivé výsledky a vnáší do procesu mnoho nepřesností. Hypotetický návrh jeho náhrady je uveden v kapitole 9.

Třetím cílem bylo pomocí navrženého algoritmu vypočítat parametry hodnotící pohyblivost mluvidel. Zde bylo díky skriptu `functions.py` získáno osm parametrů. Ty mohou být zatíženy již uvedenou chybou při předchozí filtraci nebo také nesprávnou volbou dvojic formantů pro vrcholy Hellwagova trojúhelníku, a tedy i jejich výpočet (taktéž chyba v algoritmu z předchozího cíle). To se týká jednodušších parametrů jako VSA, LnVSA a VAI. Způsobem výběru dvojic formantů F1 – F2 se tedy také nabízí zabývat i v budoucnu.

Čtvrtý cíl, tedy vzájemné porovnání a vyhodnocení diskriminační síly jednotlivých parametrů se rovněž podařilo splnit. Na základě deskriptivní statistiky a Mann-Whitneyova U testu jsou v kapitole 8 jednotlivé parametry ohodnoceny a v kapitole 9 poté okomentovány. V rámci tohoto cíle byl vytvořen skript `processing.py`.

Pátým cílem bylo vyzkoušet diskriminační sílu kombinace těchto parametrů. K tomu byly využity různé modely strojového učení podrobně vypsané v kapitole 7. K účelům tohoto cíle posloužil vytvořený skript `machinelearning.py`. Pro databázi PARCZ dosáhl nejlepšího výsledku algoritmus RF s BA = 70,2% a pro databázi CoBeN algoritmus BAG s BA = 60,2%. Podrobné výstupy skriptu jsou shrnuty a okomentovány ve stejných kapitolách jako čtvrtý cíl.

Jak již bylo naznačeno, nabízí se přepracovat způsob filtrace odlehlých hodnot, popřípadě i volbu dvojic formantů F1 – F2 pro vrcholy Hellwagova trojúhelníku. Také by bylo vhodné otestovat skripty na větší databázi, která by mohla poskytnout příhodnější podmínky pro trénování modelů strojového učení i objektivnější posouzení jednotlivých parametrů.

Literatura

- [1] Bedáňová, I.: *Neparametrické testy*. Brno: Váternární a farmaceutická univerzita Brno, 2015.
URL <https://cit.vfu.cz/statpotr/POTR/Teorie/Predn4/MannWhit.htm>
- [2] Behnel, S.; Bradshaw, R.; Citro, C.; Dalcin, L.; Seljebotn, D. S.; aj.: Cython: The best of both worlds. *Computing in Science & Engineering*, ročník 13, č. 2, 2011: s. 31–39.
- [3] Boersma, P.; Weenink, D.: PRAAT, a system for doing phonetics by computer. *Glott international*, ročník 5, 01 2001: s. 341–345.
- [4] Boersma, P.; Weenink, D.: Praat: doing phonetics by computer [Computer program]. Version 6.1.35, retrieved 29 November 2020 <http://www.praat.org/>, 2020.
- [5] Brownlee, J.: A Gentle Introduction to k-fold Cross-Validation. *Machine Learning Mastery: Making Developers Awesome at Machine Learning*, 2020.
URL <https://machinelearningmastery.com/k-fold-cross-validation/>
- [6] Brownlee, J.: How to Develop a Framework to Spot-Check Machine Learning Algorithms in Python. *Machine Learning Mastery: Making Developers Awesome at Machine Learning*, 2020.
URL <https://machinelearningmastery.com/spot-check-machine-learning-algorithms-in-python/>
- [7] Brownlee, J.: A Tour of Machine Learning Algorithms. *Machine Learning Mastery: Making Developers Awesome at Machine Learning*, 2020.
URL <https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/>
- [8] Duffy, J. R.: *Motor Speech Disorders E-Book: Substrates, Differential Diagnosis, and Management*. Mayo Clinic College of Medicine, Rochester, Minnesota: Elsevier Health Sciences, třetí vydání, 2019, ISBN 9780323550512.
URL <https://books.google.cz/books?id=08q1DwAAQBAJ>
- [9] Fox, R.; Jacewicz, E.: Reconceptualizing the vowel space in analyzing regional dialect variation and sound change in American English. *The Journal of the Acoustical Society of America*, ročník 142, č. 07, 2017: s. 444–459, doi:10.1121/1.4991021.

- [10] Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; et al.: Array programming with NumPy. *Nature*, ročník 585, č. 7825, Zář 2020: s. 357–362, doi:10.1038/s41586-020-2649-2.
URL <https://doi.org/10.1038/s41586-020-2649-2>
- [11] Holčík, J.; Komenda, M.: *Matematická biologie: e-learningová učebnice [online]*. Brno: Masarykova univerzita, první vydání, 2015, ISBN 978-80-210-8095-9.
URL <http://portal.matematickabiologie.cz/>
- [12] Hunter, J. D.: Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, ročník 9, č. 3, 2007: s. 90–95, doi:10.1109/MCSE.2007.55.
- [13] Jadoul, Y.; Thompson, B.; de Boer, B.: Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, ročník 71, 2018: s. 1–15, doi:<https://doi.org/10.1016/j.wocn.2018.07.001>.
- [14] Kalyanakrishnan, S.: K-means Clustering. *CS 344 and CS 386: Artificial Intelligence*, 2017.
URL <https://www.cse.iitb.ac.in/~shivaram/teaching/old/cs344+386-s2017/resources/classnote-2.pdf>
- [15] Krčmová, M.: *Fonetika a fonologie*. Brno: Masarykova univerzita, třetí vydání, 2009, ISBN 80-210-0137-2.
URL <http://is.muni.cz/elportal/?id=852835>
- [16] Pavlíček, T.: *Zpracování řečových signálů ve spektrální oblasti*. Bakalářská práce, Vysoké učení technické v Brně. Fakulta elektrotechniky a komunikačních technologií. Ústav radioelektroniky, Brno, 2012.
URL <http://hdl.handle.net/11012/10116>
- [17] Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; et al.: Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, ročník 12, 2011: s. 2825–2830.
- [18] Psutka, J.; Müller, L.; Matoušek, J.; Radová, V.: *Mluvíme s počítačem česky*. Praha: Academia, první vydání, 2006, ISBN 80-200-1309-1.
- [19] R Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.
URL <http://www.R-project.org/>
- [20] Sandoval, S.; Berisha, V.; Utianski, R.; Liss, J.; Spanias, A.: Automatic assessment of vowel space area. *The Journal of the Acoustical Society of America*, ročník 134, č. 11, 2013: s. EL477–83, doi:10.1121/1.4826150.

- [21] Sapir, S.; Ramig, L. O.; Spielman, J. L.; Fox, C.: Formant Centralization Ratio: A Proposal for a New Acoustic Measure of Dysarthric Speech. *Journal of Speech, Language, and Hearing Research*, ročník 53, č. 01, 2010: s. 114–125, doi:10.1044/1092-4388(2009/08-0184).
- [22] Sapir, S.; Ramig, L. O.; Spielman, J. L.; Fox, C.: Acoustic metrics of vowel articulation in Parkinson’s disease: Vowel space area (VSA) vs. Vowel articulation index (VAI). *Models and Analysis of Vocal Emissions for Biomedical Applications - 7th International Workshop, MAVEDA 2011*, ročník 7, č. 01, 2011: s. 173–175.
- [23] Smékal, Z.: Číslicové zpracování řeči. *Elektronická skripta pro magisterská studia. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií. Ústav telekomunikací. Brno*, 2009.
- [24] Stehr, D.: *A Tutorial on Recent Methods for Estimating Working Vowel Space from Connected Speech*. Irvine: University of California, 2018.
URL https://rstudio-pubs-static.s3.amazonaws.com/348434_89d7208104214419b406a6f15d0d8bf3.html#sandoval-et-al.-2013-approach
- [25] Story, B.; Bunton, K.: Vowel Space Density as an Indicator of Speech Performance. *The Journal of the Acoustical Society of America*, ročník 141, č. 05, 2017: s. EL458–EL464, doi:10.1121/1.4983342.
- [26] Tay, K.: What is balanced accuracy? *Statistical Odds & Ends*, 2020.
URL <https://statisticaloddsandends.wordpress.com/2020/01/23/what-is-balanced-accuracy/>
- [27] Team, T. P. D.: pandas-dev/pandas: Pandas. Únor 2020, doi:10.5281/zenodo.3509134.
URL <https://doi.org/10.5281/zenodo.3509134>
- [28] Van Rossum, G.; Drake Jr, F. L.: *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995.
- [29] Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; aj.: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 2020.
- [30] Waskom, M. L.: seaborn: statistical data visualization. *Journal of Open Source Software*, ročník 6, č. 60, 2021: str. 3021, doi:10.21105/joss.03021.
URL <https://doi.org/10.21105/joss.03021>

- [31] Whitfield, J.: Exploration of Metrics for Quantifying Formant Space: Implications for Clinical Assessment of Parkinson Disease. *Perspectives of the ASHA Special Interest Groups*, ročník 4, č. 04, 2019: s. 1–9, doi:10.1044/2019_PERS-SIG19-2018-0004.
- [32] Whitfield, J.; Goberman, A.: Articulatory-Acoustic Vowel Space: Application to Clear Speech in Individuals with Parkinson disease. *Journal of Communication Disorders*, ročník 51, 09 2014, doi:10.1016/j.jcomdis.2014.06.005.
- [33] Whitfield, J.; Goberman, A.: Articulatory-acoustic vowel space: Associations between acoustic and perceptual measures of clear speech. *International journal of speech-language pathology*, ročník 19, č. 04, 2017: s. 184–194, doi:10.1080/17549507.2016.1193897.
- [34] Čermák, J.: *Rozpoznávání emočních stavů na základě analýzy řečového signálu*. Diplomová práce, Vysoké učení technické v Brně. Fakulta elektrotechniky a komunikačních technologií. Ústav telekomunikací, Brno, 2009.
URL <https://dspace.vutbr.cz/xmlui/handle/11012/10954>
- [35] Černocký, J.: Zpracování řečových signálů – studijní opora. *FIT VUT v Brně*, 2006.
- [36] Šrámková, K.: *Czech Expats' English Pronunciation*. Diplomová práce, Masarykova univerzita, Department of English and American Studies, Brno, 2019.
URL https://is.muni.cz/th/wo9xg/Sramkova_Czech_Expats__English_Pronunciation.pdf

Seznam symbolů, veličin a zkratek

AAVS	Articulatory-Acoustic Vowel Space
ADA	adaptivní posilování – Adaptive Boosting
BA	vyvážená přesnost – Balanced Accuracy
BAG	bagging metoda
Bayes	naivní Bayesův klasifikátor
CART	klasifikátor rozhodovacího stromu – Classification and Regression Tree
CoBeN	databáze ze skupiny aplikovaných neurověd Středoevropského technologického institutu
CV	křížová validace – Cross-validation
DP	parametr odvozený s VSD – Density Percentage
EM	Expectation-Maximization
ET	klasifikátor extrémně náhodného stromu – Extra Tree
ETS	extrémně náhodné stromy – Extra Trees
F1	první formant
F2	druhý formant
FCR	Formant Centralization Ratio
GBM	gradientní posilování – Gradient Boosting
GMM	Gaussův smíšený model – Gaussian Mixture Model
GV	zobecněný rozptyl – Generalized Variance
HC	zdravá kontrolní skupina – Healthy Control
HD	hypokinetická dysartrie
kNN	algoritmus k-nejbližších sousedů – k-Nearest Neighbors
LED	denní levodopa-ekvivalentní dávka – Levodopa Equivalent Dose
LnVSA	logaritmické vyjádření vokální oblasti samohlásek

Logistic	logistická regrese
LPC	lineární predikce – Linear Predictive Coding
MSD	morické poruchy řeči – Motor Speech Disorders
M-W test	Mann-Whitneyho U test
<i>p</i>	pravděpodobnost
<i>p_{PAR}</i>	pravděpodobnost zjištěná na databázi PARCZ
<i>p_{CBN}</i>	pravděpodobnost zjištěná na databázi CoBeN
PA	pasivně agresivní klasifikátor – Passive-Aggressive
PARCZ	databáze z I. Neurologické kliniky Fakultní nemocnice u sv. Anny v Brně
PN	Parkinsonova nemoc
RF	náhodný les – Random Forest
Ridge	hřebenová regrese
RSDF1	relativní směrodatná odchylka prvního formantu
RSDF2	relativní směrodatná odchylka druhého formantu
SEN	senzitivita
SGD	gradientní metody prvního řádu – Stochastic Gradient Descent
SGV	standardizovaný všeobecný rozptyl – Standardized General Variance
SPE	specifická
SVM	metoda podpůrných vektorů – Support Vector Machines
UPDRS III	jednotná škála hodnocení Parkinsonovi nemoci (vyšetření motoriky) – Unified Parkinson's Disease Rating Scale
VAI	Vowel Articulation Index
VSA	vokální oblast samohlásek – Vowel Space Area
VSD	Vowel Space Density

VSHA	Vowel Space Hull Area
WVSM	Workin Vowel Space Metrics
σ	směrodatná odchylka
Σ	variančně-kovarianční matice

Obsah příložených souborů

Následující struktura popisuje organizaci souborů v příloženém archivu. Vytvořené skripty jsou určeny pro jazyk Python a byly vytvořeny v jeho verzi 3.9.0. Veškeré kódy jsou okomentovány pro vyšší přehlednost. Důvodem použití anglických komentářů je zvýšení přístupnosti širší veřejnosti. Soubory `.csv` jsou určeny primárně pro uložení získaných dat a jejich předávání mezi jednotlivými skripty, nikoliv pro jejich přehledné zobrazení.

```
/. .....kořenový adresář příloženého archivu
├── skripty ..... skripty jazyka Python
│   ├── calculation.py ..... výpočet formantů a vrcholů Hellwagova trojúhelníku
│   ├── functions.py ..... funkce pro výpočet artikulačních parametrů
│   ├── processing.py ..... statistická analýza vypočítaných parametrů
│   └── machinelearningmastery.py ..... funkce pro explorativní analýzu
├── tabulky ..... zdrojové a výstupní tabulky pro jednotlivé skripty
│   ├── COBEN_3_03_13.csv ..... artikulační parametry jednotlivých řečníků
│   ├── COBEN_all.csv ..... výsledky různých modelů strojového učení
│   ├── COBEN_best.csv ..... výsledky modelů s optimalizovanými hyperparametry
│   ├── COBEN_PN_stats.csv ..... deskriptivní statistika pro PN
│   ├── COBEN_HC_stats.csv ..... deskriptivní statistika pro HC
│   ├── COBEN_info.csv ..... kovariáty jednotlivých řečníků
│   ├── PAR_3_03_13.csv ..... artikulační parametry jednotlivých řečníků
│   ├── PAR_all.csv ..... výsledky různých modelů strojového učení
│   ├── PAR_best.csv ..... výsledky modelů s optimalizovanými hyperparametry
│   ├── PAR_PN_stats.csv ..... deskriptivní statistika pro PN
│   ├── PAR_HC_stats.csv ..... deskriptivní statistika pro HC
│   └── PAR_info.csv ..... kovariáty jednotlivých řečníků
```