

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

DIPLOMOVÁ PRÁCE

Brno, 2021

Bc. Marek Bezůšek



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

OBJEKTIVIZACE TESTU 3F - DYSARTRICKÝ PROFIL POMOCÍ AKUSTICKÉ ANALÝZY

OBJECTIFICATION OF THE TEST 3F - DYSARTHIC PROFILE BASED ON ACOUSTIC ANALYSIS

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. Marek Bezůšek

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Jiří Mekyska, Ph.D.

BRNO 2021

Diplomová práce

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

Student: Bc. Marek Bezůšek

ID: 195797

Ročník: 2

Akademický rok: 2020/21

NÁZEV TÉMATU:

Objektivizace Testu 3F - dysartrický profil pomocí akustické analýzy

POKYNY PRO VYPRACOVÁNÍ:

V rámci práce budou navrženy a implementovány parametrizační metody umožňující objektivní hodnocení fonace, artikulace, prozodie a respirace při vyšetření Testem 3F - dysartrický profil. Na základě statistické analýzy, popř. strojového učení, budou nalazeny parametry (nebo jejich kombinace), které nejvíce korelují s percepčním hodnocením. K těmto parametrům/kombinacím budou vytvořeny normy.

DOPORUČENÁ LITERATURA:

[1] KOŠTÁLOVÁ, Milena, Martina MRAČKOVÁ, Radek MAREČEK, et al. Test 3F Dysartrický profil – normativní hodnoty řeči v češtině. *Cesk Slov Neurol N.* 2013, 76/109(5), 614-618.

[2] BRABENEC, Luboš, Jiří MEKYSKA, Zoltán GALÁŽ a Irena REKTOROVÁ. Speech disorders in Parkinson's disease: early diagnostics and effects of medication and brain stimulation. *Journal of Neural Transmission.* 2017, 124(3), 303-334. DOI: 10.1007/s00702-017-1676-0. ISSN 0300-9564.

Termín zadání: 1.2.2021

Termín odevzdání: 24.5.2021

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

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

UPOZORNĚNÍ:

Autor diplomové práce nesmí při vytváření diplomové 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

Test 3F je používán pro hodnocení motorické poruchy řeči (dysartrie) u česky mluvících řečníků. Hodnocení dysartrické řeči je zkresleno subjektivním posouzením. Motivací pro napsání této práce je fakt, že neexistuje mnoho automatických analytických nástrojů, které lze použít pro objektivní hodnocení fonace, artikulace, prozodie a respirace dysartrické řeči. Cílem diplomové práce je identifikovat, implementovat a otestovat akustické parametry, které lze použít pro objektivizaci a automatizaci hodnocení Testu 3F. Akustické parametry by měli být klinicky interpretovatelné. Předpokládá se, že akustická analýza by mohla být přesnější než klinické zhodnocení vyšetřujícím. Akustické parametry byly testovány na databázi 151 řečníků (51 zdravých řečníků, 100 pacientů). Pro identifikaci korelace parametrů se subjektivním hodnocením byly použity metody statistické analýzy a strojového učení. Bylo zjištěno, že 27 testovacích úloh ze 30 lze použít v rámci automatického hodnocení. V rámci této diplomové práce bylo analyzováno pouze 10 úloh Testu 3F, zbývající úlohy nebyly analyzovány, protože pouze část databáze byla předzpracována. Výsledkem statistické analýzy je 14 parametrů, které byly identifikovány jako nejvíce významné pro hodnocení. Tři nejvýznamnější parametry jsou: MET (respirace), relF0SD (intonace), relSEOVR (intenzita hlasu – prozodie). Nejmenší chyba predikce regresních modelů strojového učení byla 7.14 %. Výsledky ukazují, že většinu úloh Testu 3F lze automatizovat. Z výsledků analýzy deseti úloh Testu 3F vyplývá, že nejdůležitějším faktorem při hodnocení dysartrie je nedostatečná expirace, monotónnost a nízká variabilita intenzity řeči.

KLÍČOVÁ SLOVA

Akustická analýza řeči, akustické parametry řeči, automatická diagnóza poruchy řeči, dysartrie, objektivizace hodnocení, parametrizační metody, statistická analýza, strojové učení, Test 3F.

ABSTRACT

Test 3F is used to diagnose the extent of motor speech disorder – dysarthria for czech speakers. The evaluation of dysarthric speech is distorted by subjective assessment. The motivation behind this thesis is that there are not many automatic and objective analysis tools that can be used to evaluate phonation, articulation, prosody and respiration of speech disorder. The aim of this diploma thesis is to identify, implement and test acoustic features of speech that could be used to objectify and automate the evaluation. These features should be easily interpretable by the clinician. It is assumed that the evaluation could be more precise because of the detailed analysis that acoustic features provide. The performance of these features was tested on database of 151 czech speakers that consists of 51 healthy speakers and 100 patients. Statistical analysis and methods of machine learning were used to identify the correlation between features and subjective assesment. 27 of total 30 speech tasks of Test 3F were identified as suitable for automatic evaluation. Within the scope of this thesis only 10 tasks of Test 3F were tested because only a limited part of the database could be preprocessed. The result of statistical analysis is 14 features that were most useful for the test evaluation. The most significant features are: MET (respiration), relF0SD (intonation), relSEOVR (voice intensity – prosody). The lowest prediction error of the machine learning regression models was 7.14 %. The conclusion is that the evaluation of most of the tasks of Test 3F can be automated. The results of analysis of 10 tasks shows that the most significant factor in dysarthria evaluation is limited expiration, monotone voice and low variability of speech intensity.

KEYWORDS

Acoustic analysis of speech, acoustic speech features, automatic diagnosis of speech disorder, dysarthria, machine learning, objectification of evaluation, parameterization methods, speech features, statistical analysis, Test 3F.

BEZŮŠEK, Marek. *Objektivizace Testu 3F - dysartrický profil pomocí akustické analýzy*. Brno, 2020, 47 s. Diplomová práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací. Vedoucí práce: Ing. Jiří Mekyska, Ph.D.

PROHLÁŠENÍ

Prohlašuji, že svou diplomovou práci na téma „Objektivizace Testu 3F - dysartrický profil pomocí akustické analýzy“ jsem vypracoval samostatně pod vedením vedoucího diplomové 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 diplomové práce dále prohlašuji, že v souvislosti s vytvořením této diplomové 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 diplomové práce panu Ing. Jíří Mekyskovi, Ph.D. za odborné vedení, konzultace, trpělivost a podnětné návrhy k práci.

Obsah

Úvod	11
1 Úvod do neurologických poruch řeči	12
1.1 Typy neurologických poruch řeči	12
2 Dysartrie	13
2.1 Druhy dysartrií podle percepčních symptomů a jejich lokalizace poškození	13
3 Test 3F	15
4 Akustické parametry pro kvantifikaci dysartrií	17
4.1 Fonační a prozodické parametry	17
4.1.1 Kmitočet základního tónu F0	17
4.1.2 Intenzita řeči	18
4.2 Další fonační parametry	18
4.2.1 Kvalita hlasu	18
4.3 Respirační a fonační parametry	19
4.4 Další prozodické parametry	20
4.4.1 Tempo řeči	20
4.4.2 Pauzy v řeči	20
4.4.3 Rytmus řeči	21
4.5 Artikulační parametry	21
4.5.1 Artikulace samohlásek (Formantové kmitočty)	21
4.5.2 Tempo artikulace hlásek a slabik	22
4.6 Respirační parametry	22
5 Návrh objektivizace Testu 3F	24
6 Databáze a metodologie	26
6.1 Databáze	26
6.2 Metody zpracování dat a parametrizace	26
6.2.1 Implementace akustických parametrů	26
6.2.2 Parametrizace	27
6.3 Statistická analýza	27
6.3.1 Korelační analýza	28
6.3.2 Regresní analýza	29
6.4 Vytvoření norem z akustických dat	30

7	Výsledky	31
7.1	Korelační analýza	31
7.2	Regresní analýza	33
7.3	Normy z akustických dat	34
8	Diskuse	37
	Literatura	41
	Seznam symbolů, veličin a zkratk	43
9	Obsah přiloženého DVD	46
10	Příručka pro použití vytvořeného systému	47

Seznam obrázků

7.1	Multikolinearita parametrů	31
7.2	Významnost parametrů (Lasso, $\alpha = 0.2$)	34
7.3	Rozdíl normy mužů oproti ženám	34

Seznam tabulek

3.1	Dysartrický index	15
3.2	Struktura Testu 3F	16
4.1	Akustické parametry pro kvantifikaci dysartrií	22
5.1	Přiřazení akustických parametrů k položkám Testu 3F	25
6.1	Databáze	26
6.2	Parametrizované úlohy	28
6.3	Seznam hodnocených regresních modelů	29
6.4	Normativní data Testu 3F	30
7.1	Spearmanova korelace parametrů s subjektivním skóre	32
7.2	Deset nejlepších skóre regresní analýzy	33
7.3	Metriky modelu s nejmenší chybou predikce (Lasso, $\alpha = 0.2$)	33
7.4	Norma ženy	35
7.5	Norma muži	36

Úvod

Tato práce se zabývá možnostmi objektivizace a automatizace Testu 3F akustickou analýzou řeči. Test 3F je používán pro kvantifikaci motorické poruchy řeči, nazývané dysartrie, u česky mluvících řečníků. Zhodnocení míry dysartrie vyšetřujícím je ovlivněno subjektivním posouzením. Test 3F tento problém částečně řeší zavedením instrukcí pro vyšetřujícího a testovacích úloh. Přesto jsou ale jednotlivé úlohy hodnoceny subjektivně a celý proces vyšetření a vyhodnocení je zdouhavý. Pokud by bylo možné jednotlivé řečové úlohy analyzovat pomocí metod číslicového zpracování řeči a následně použít metody statistické analýzy a strojového učení pro zhodnocení poruchy řeči, tak by se mohl proces hodnocení objektivizovat a automatizovat. Akustickými parametry lze zaznamenat detailní změny řeči, a proto by mohla být akustická analýza přesnější než klinické zhodnocení vyšetřujícím.

Akustickou analýzou poruch řeči a identifikováním vhodných parametrů se zabývají mnozí výzkumníci v zahraničí i v České republice. Existuje elektronická verze Testu 3F, která zjednodušuje, zpřesňuje a zrychluje vyhodnocení. Automatizovaných testů poruch řeči však mnoho není. Cílem této práce je využití struktury, databáze nahrávek řečových úloh a normativních dat Testu 3F k nalezení akustických parametrů, které lze efektivně použít pro automatické hodnocení testu. Následně tyto parametry implementovat a otestovat jejich korelaci se subjektivním hodnocením Testu 3F. Z akustických dat budou dále vytvořeny normy podle normativních dat Testu 3F. Výsledky automatického hodnocení by měli být klinicky interpretovatelné, proto i akustické parametry by měli reprezentovat vlastnosti řeči, které jsou používány při klinickém vyšetření.

Struktura práce je rozdělena na teoretickou část, která uvádí do problematiky neurologických poruch řeči, je zde definována dysartrie, její typy a symptomy. Dále je popsána struktura Testu 3F, metody hodnocení a vymezení části testu, kterou lze akusticky analyzovat. V dalších kapitolách jsou identifikovány akustické parametry, které lze potencionálně k analýze využít. Následně jsou selektovány parametry, které lze použít pro automatickou analýzu. V poslední kapitole teoretické části je proveden návrh objektivizace Testu 3F. V praktické části práce je popsáno zpracování databáze, implementace parametrů, parametrizace databáze a statistická analýza, pomocí které je zjištěna korelace parametrů se subjektivním hodnocením. V závěru praktické části je popsán proces vytvoření norem z akustických dat.

1 Úvod do neurologických poruch řeči

Na tvorbě řeči se podílí několik neurologických procesů, které lze vyjádřit ve zjednodušeném modelu [9, s. 276]: neuro-kognitivní, neuro-motorický, neuromuskulární [11, 5]. V první fázi tohoto procesu dochází k vyjádření myšlenek a pocitů subjektu ve formě přirozeného jazyka (sémantika, syntax, slovní zásoba, gramatika) a nakonec ve formě fonologické (fonémy, slabiky, prozódie). Nazývá se také jako **kognitivně-lingvistická fáze**. Ve druhé fázi dojde k naplánování pohybů řečových svalů (podle fonémů, slabik a prozódie) a k tzv. programování impulzů pro svaly [9, s. 290–291]. Tento proces je ještě modulován sensorickou zpětnou vazbou [7][5, s. 101], která umožňuje monitorovat pohyby vykonané řečovými svaly. Pokud se v řeči vyskytne chyba, tak prostřednictvím senzomotorických procesů může být opravena. Tato fáze se nazývá **motorické plánování a programování řeči**. Ve třetí fázi dochází k realizaci pohybového plánu a programů prostřednictvím motorických neuronů, nervů a nervových drah, neuromuskulárního spojení a řečových svalů.

1.1 Typy neurologických poruch řeči

V závislosti na tom, která fáze procesu tvorby řeči je narušena, můžeme rozdělit poruchy řeči na jazykové, motorické a sensorické.

Jazykové poruchy řeči (fáze kognitivně-lingvistická) se projevují jazykovými a fonologickými chybami. Mezi jazykové poruchy řeči patří **afázie**. Dle lokalizace poškození mozku se dělí na několik typů a projevuje se nesprávnou formulací a pochopením jazyka (gramatika, slovní zásoba, tvarosloví) a dále neplýnulou řečí.

Motorické poruchy řeči: Apraxie, Dysartrie

U motorický poruch řeči dochází k narušení oblasti centrálního nervového systému (CNS) nebo periferního nervového systému (PNS), která se podílí na druhé nebo třetí fázi procesu tvorby řeči. Tzn., že je narušen proces motorického plánování a programování nebo neuromuskulární realizace řeči. V prvním případě jde o **apraxii** a v řeči se projevuje nesprávnou artikulací a prozodií. Případ druhý, **dysartrie**, se projevuje špatnou kontrolou nad řečovými svaly a tím dochází k problémům s fonací, artikulací, prozodií, respirační a rezonancí.

Senzorické poruchy řeči: Pro monitorování a korekci chyb vlastní řeči využívá nervový systém sensorickou zpětnou vazbu [7], která ho informuje o pohybech řečových svalů (tzv. propriocepce¹) a sluchovou zpětnou vazbu. Podle toho, ve které části nervového systému je zpětná vazba poškozena, tak ovlivňuje funkci některé ze tří fází tvorby řeči. [9, s. 276, 278]

¹Propriocepce je proces nervového systému (NS), který prostřednictvím sítě nervových drah získává sensorické informace o činnosti svalů v těle.

2 Dysartrie

Dysartrie je motorická porucha řeči, která je způsobena poškozením centrálního nebo periferního nervového systému a dochází při ní k poruše neuromuskulární realizace řeči. Projevuje se zhoršenou kontrolou pohybů řečových svalů, tedy zhoršenou kontrolou síly, rychlosti, rozsahu a přesnosti pohybu a také neobvyklým svalovým napětím. V řeči tak vznikají problémy s **respirací** (dýchací svaly), **fonací** (hrtanové svaly), **artikulací** (svaly artikulačního ústrojí), **rezonancí** (svaly artikulačního ústrojí) a také **prozodií**.

Dysartrii můžeme rozlišit na několik typů podle

1. percepčních (řečových) symptomů, nebo podle
2. lokalizace poškození nervového systému.

Identifikace typu dysartrie, podle percepčních symptomů, může pomoci při snaze lokalizovat poškození nervové soustavy [5, s. 193].

2.1 Druhy dysartrií podle percepčních symptomů a jejich lokalizace poškození

Flacidní: Typickými příznaky flacidní dysartrie je slabost svalů, snížené reflexy a snížené svalové napětí, což má vliv na rychlost, rozsah a přesnost pohybů. U tohoto typu dysartrie dochází k poškození nervových drah dolního motorického neuronu. Podle řečových příznaků a také podle zasažených svalů, které tyto řečové symptomy způsobují, lze lokalizaci poškození upřesnit [5, s. 193]. Při flacidní dysartrii dochází k poruše **fonace/respirace** (dýchavičnost, krátké věty, hlasitý nádech), dále **rezonance** (hypernazalita), **artikulace**, **fonace/prozodie** (monotónnost, konstantní hlasitost, drsný hlas). Řečové symptomy se mohou projevovat jednotlivě i dohromady v závislosti na rozsáhlosti poškození nervové soustavy. Flacidní dysartrie může být způsobena onemocněním demyelinizačním, neurodegenerativním, metabolickým [5] a také onemocněním zánětlivým, nádorovým, vaskulárním a traumatickým poraněním hlavy [4, 5].

Spastická: Charakteristickými znaky spastické dysartrie jsou: zvýšené svalové napětí, zvýšené reflexy, svalová slabost a pomalost artikulačních svalů. Slabost a pomalost se projevuje převážně při pohybech jazyka a rtů (**pomalá artikulace, pomalé tempo řeči**). Zvýšené svalové napětí postihuje především hrdelní svaly a ovlivňuje tak funkčnost hlasivek (**porucha fonace**). Zvýšené svalové napětí ovlivňuje také měkké patro, to se v řeči projeví hypernazalitou (**porucha rezonance**). Spasticita svalů snižuje jejich přirozený rozsah pohybu a zapříčiňuje omezený dynamický rozsah hlasu a monotónnost (**porucha prozodie**). Spastická dysartrie je

důsledek bilaterálního poškození horního motorického neuronu tzn. poškození pyramidové a extrapyramidové dráhy.

Dysartrie typu unilaterálního poškození horního motorického neuronu: Příčina a symptomy se podobají spastické dysartrii. Projevuje se slabostí dolní části tváře a jazyka.

Ataktická: Tento typ dysartrie je způsoben lézí v mozečku (cerebellum) a poškozením jeho nervových drah. Dochází především k poruše synchronizace vykonávaných pohybů. Příznakem je také snížené svalové napětí. V řeči se projevuje hlavně poruchou **artikulace hlásek** a poruchou **prozodie** (omezený dynamický rozsah hlasu, omezený tónový rozsah, porucha akcentu: stejný nebo výrazný důraz slabik [1], pomalé tempo řeči [9, s. 281]).

Hypokinetická: Příčinou hypokinetické dysartrie je léze bazálních ganglií a je asociována s Parkinsonovou chorobou [4, s. 9]. Způsobuje ji rigidita, slabost, snížený rozsah pohybu a pomalost pohybu. Řečovými příznaky jsou: **nepřesná artikulace**, **dýchavičná dysfonie** [4] (slabá intenzita hlasu, šepot) a **poruchy prozodie**: monotónnost, konstantní hlasitost, snížený důraz, krátké věty, nepravidelné tempo řeči a krátké zrychlení tempa řeči. Hypokinetická dysartrie může být způsobena neurodegenerativním onemocněním (Parkinsonova nemoc), vaskulárním (cévní mozková příhoda), úrazem hlavy, infekcí, některými metabolickými onemocněními nebo toxickým poškozením (těžké kovy) [5].

Hyperkinetická: Příčinou hyperkinetické dysartrie je léze bazálních ganglií. Projevuje se rychlými nebo pomalými, nekontrolovanými a nepravidelnými pohyby. Příkladem jsou tiky, svalový třes a chorea. V řeči se projevuje **nepřirozenou prozodií**, především poruchou tempa řeči, tzn. pauzy v řeči, krátké věty a také omezený dynamický rozsah hlasu a omezený tónový rozsah. Léze bazálních ganglií může být zapříčiněna například Huntingtonovou nemocí [5, s. 409]

Smíšená dysartrie se objevuje při poškození více oblastí centrálního a periferního nervového systému. Příznaky více typů dysartrií mohou mít tyto onemocnění: roztroušená skleróza, amyotrofická laterální skleróza (ALS), toxické a metabolické poškození NS, poranění hlavy, nádorová onemocnění, autoimunitní onemocnění (např. Lupus), opakovaná cévní mozková příhoda [5].

3 Test 3F

Test 3F je využíván především klinickými logopedy a foniatry ke zhodnocení míry poruchy řeči u dysartrie (dysartrický index) a určení jejího charakteru (dysartrický profil). Při klinickém vyšetření je přesnost zhodnocení ovlivněna subjektivním posouzením vyšetřujícího [10]. Proto při stejných podmínkách analýzy tak nelze dosáhnout vždy stejných výsledků. Tuto rozdílnost výsledků lze do určité míry snížit zavedením testovacích instrukcí a normativních hodnot v podobě Testu 3F. Jelikož je ale testování prováděno vyšetřujícím, budou výsledky vždy zkresleny subjektivním posouzením. V případě, že by byl vyšetřující nahrazen automatickým systémem, který by prováděl analýzu pacienta, tak by mohl být test více objektivní.

Test 3F umožňuje vyšetřujícímu posoudit poruchu řeči **kvalitativně** (což vede k subjektivnímu hodnocení), tzn. podle závažnosti jednotlivých příznaků a dále **kvantitativně** určit tíži dysartrie. Kvalitativní část testování je rozdělena na tři oblasti: **faciokineze**, **fonorespirace** a **fonetika**. První oblast je zaměřena na testování motorických funkcí rtů, jazyka a čelisti. Druhá oblast testuje funkci hlasivek a dýchacích svalů. Třetí oblast testuje artikulaci, prozódii a srozumitelnost. Každá z těchto tří oblastí je rozdělena do třech podkategorií po pěti testovacích úlohách (viz tabulka 3.2). Provedení testovacích úloh je hodnoceno body 0–2 (0 – těžká porucha, 1 – narušený výkon, 2 – bezchybný výkon). Následuje kvantitativní hodnocení, při kterém jsou sečteny body získané v testovacích úlohách. Výsledkem kvantitativního hodnocení je tzv. **dysartrický index**, který je na základě normativních dat rozdělen na několik úrovní vyjadřující míru dysartrie (viz tabulka 3.1).

Tab. 3.1: Přehled úrovní dysartrického indexu [10]

Dysartrický index	míra dysartrie
85–90	bez poruchy
74–85	velmi lehká dysartrie
57–73	lehká dysartrie
36–56	středně těžká dysartrie
17–35	těžká dysartrie
0–17	velmi těžká dysartrie (anartrie)

Dysartrický profil ukazuje úroveň poruchy dle jednotlivých podkategorií (rty, čelist, jazyk, respirace atd.). Srovnává hodnocení těchto podkategorií za sebou v tabulce a přibližně tak popisuje charakter dysartrie. Jak je ale uvedeno v [13, s. 14], podle dysartrického profilu nelze přesně určit typ dysartrie. Je to kvůli podobnosti

symptomů u jednotlivých dysartrií a také kvůli malému vzorku dosud testovaných pacientů.

V rámci **akustické analýzy** je důležitá oblast Fonorespirace a Fonetiky, pro které se používá několik typů úloh vyšetření: prodloužená fonace samohlásek, opakování souhlásek, opakování slabik, čtení slov, vět a textu, monolog (spontánní řeč).

Tab. 3.2: Struktura Testu 3F [10]

FACIOKINEZE	FONORESPIRACE	FONETIKA
Rty	Respirace	Artikulace
Protruze a retrakce rtů mezi zuby	Klidová prodloužená expirace	Přesnost opakování samohlásek
Pevnost sevření rtů v klidu	Síla výdechového proudu	Přesnost opakování souhlásek
Pevnost retního uzávěru při nafouknutých tvářích	Výdrž expirace při syčení (sss...)	Přesnost artikulace při čtení TEXT
Zaostření koutků do úsměvu	Opakování sérií (ss-ss-ss...)	Diadochokineze s fonací (p-t-k) (o-e)
Diadochokineze bez fonace	Plynulé zesilování a zeslabování sykotu	Spontánní řeč
Čelist	Respirace při fonaci	Prozódie
Otevření a zavření úst (volné)	Výdrž expirace při prodloužené fonaci hlásky m	Udržení rytmu v rytmickém celku TEXT
Otevření a zavření úst proti odporu	Výdrž expirace při prodloužené fonaci hlásky íí	Základní větné intonace TEXT
Posouvání mandibuly doprava-doleva	Synchronizovanost respirace s fonací fffííí	Přemístování kontrastního důrazu TEXT
Kroužení mandibulou do stran	Délka výdechové mluvní fráze	Intonační variabilita TEXT
Kontrakce žvýkacích svalů	Mluvní respirace	Komplexní prozodické faktory
Jazyk	Fonace	Srozumitelnost
Vysunutí jazyka z úst a zasunutí	Kvalita hlasu	Srozumitelnost předříkávaných slov TEXT
Zvednutí špičky jazyka vzhůru a spuštění zpět	Rezonance	Srozumitelnost předříkávaných vět TEXT
Vysunutí, obrácení špičky jazyka před ústy vzhůru	Přiměřená a ovládaná hlasitost	Srozumitelnost četby slov TEXT
Přesunutí jazyka z koutku do koutku	Přiměřená a ovládaná výška	Srozumitelnost četby textu TEXT
Kruhovitě olíznutí horního a spodního rtu	Hlasový rozsah	Srozumitelnost spontánní mluvy

4 Akustické parametry pro kvantifikaci dysartrií

V této kapitole jsou popsány akustické parametry používané pro kvantifikaci dysartrií a zařazeny dle podkategorií Testu 3F. Dále jsou k těmto parametrům přiřazeny používané řečové úlohy.

Akustické parametry reprezentují určitou vlastnost řečového signálu a lze je ze signálu vypočítat procesem parametrizace. Parametry lze počítat z krátkých časových úseků řeči (segmentální parametry (délka 20–30 ms), např. formanty) a delších časových úseků (suprasegmentální parametry (délka stovky ms), např. tempo řeči). Některé akustické parametry jsou dále statisticky zpracovány. Například vypočítáním střední hodnoty, směrodatné odchylky (SD – standard deviation), mediánu a statistických momentů.

Akustické parametry, které lze použít pro automatickou kvantifikaci dysartrií jsou uvedeny v tabulce 4.1.

4.1 Fonační a prozodické parametry

Funkci hlasivek, která ovlivňuje fonaci i prozódii řeči, lze analyzovat níže uvedenými parametry. Analýza těchto podsystémů řeči se liší použitými řečovými úlohami a metodami statistického zpracování parametrů.

4.1.1 Kmitočet základního tónu F0

Parametr, kterým lze zjistit poruchu funkce hlasivek je **kmitočet základního tónu hlasu – F0**. Jde o kmitočet, kterým kmitají hlasivky při fonaci. Vlivem zvýšeného nebo sníženého svalového napětí hrdelních svalů se snižuje schopnost hlasivek udržet tón a dochází tak ke kolísání F0, což lze změřit **směrodatnou odchylkou F0 (F0SD)** nebo parametrem **jitter**. Parametr jitter je definován jako míra kolísání rozsahu hlasu. Jitter má několik druhů: local, RAP (Relative Average Perturbation), PPQ (Pitch Perturbation Quotient), DDP – definován jako průměr absolutních hodnot rozdílů z po sobě následujících rozdílů délek (po sobě následujících) period, normalizován průměrnou délkou periody (average absolute Difference between consecutive Differences between consecutive Periods, divided by the average period) [15].

Pro analýzu **fonace** se výše uvedené parametry počítají z prodloužené fonace samohlásek [4, 15, 14], ty lze nalézt v Testu 3F v podkategorii Respirace při fonaci.

Mezi **fonační parametry** lze podle Testu 3F zařadit také změnu přirozené výšky hlasu, kterou je možné vypočítat z delších řečových promluv (čtení vět a textu, monolog) **střední hodnotou F0 (mean F0)**, dále také kolísání tónového rozsahu hlasu **F0VR** (F0 Variation Range). Při analýze delších promluv lze také zjistit poruchu **prozódie**, např. monotónnost a dysprozodie (nízké kolísání F0 – F0SD), nesprávná intonace při čtení textu (F0SD). Používají se i normalizované parametry např. relativní kolísání F0 (**relF0SD**), relativní kolísání tónového rozsahu (**relF0VR**) [3, 4].

4.1.2 Intenzita řeči

Pro výpočet intenzity řeči se používají následující parametry [4]:

- střední krátkodobá energie signálu (v anglické literatuře často značeno jako Mean Intensity, Mean SEO [3, 14]),
- Teagerův-Keiserův energetický operátor (TKEO).

Pomocí těchto parametrů lze určit celkový pokles hlasitosti hlasu a z toho lze dále určit např. míru hypofonie. Vyjádřením směrodatné odchylky těchto parametrů zjistíme kolísání intenzity řeči (tzv. Monoloudness), nebo případně parametrem **shimmer**, který se ale počítá z úloh s prodlouženou fonací samohlásek podobným principem jako jitter. Parametr shimmer se tak často řadí do **fonačních parametrů**. Ostatní výše uvedené parametry se počítají z úloh obsahující monolog nebo čtení vět a textu, takže se řadí mezi **prozodické parametry**. Shimmer má několik variant: APQ (Amplitude Perturbation Quotient), local, DDA – definován jako průměr absolutních hodnot rozdílů z po sobě následujících rozdílů amplitud po sobě následujících period (average absolute Difference between consecutive Differences between the Amplitudes of consecutive period) [15]. Lze použít i normalizované parametry intenzity např. **relSEOSD** – směrodatná odchylka intenzity normalizovaná střední hodnotou, **relSEOVR** – rozsah intenzity normalizován střední hodnotou. Intenzita normalizovaných parametrů je počítána po odstranění pauz delších než 50 ms.

4.2 Další fonační parametry

4.2.1 Kvalita hlasu

Kvalita hlasu je další vlastností řeči, která může být narušena nedostatečným uzavěrem hlasivek a vznikem aperiodicity při fonaci. Akustická analýza kvality hlasu je založena především na detekci úrovně šumu v řeči tj. chraptění, drsnost hlasu, šepot

a další příznaky vyskytující se u dysfonie. Jeden z nejpoužívanějších parametrů popisující poměr harmonického signálu k šumu je **HNR** (Harmonic-to-Noise Ratio). Dalšími jsou [4]:

- normalizovaná energie šumu **NNE** (Normalized Noise Energy),
- poměr glotálního signálu a šumu **GNE** (Glottal-to-Noise Excitation ratio),
- počet průchodů signálu nulovou úrovní **ZCR** (Zero Crossing Rate),
- parametry založené na keprální analýze **CPP** (Cepstral Peak Prominence).

Další parametry kvantifikující kvalitu hlasu analyzují periodicitu prodloužených samohlásek. Pokud periodičita kolísá jde o tzv. dysperiodicitu (nebo aperiodicitu), kterou popisují tyto parametry:

- **SSD** (Segmental Signal-to-Dysperiodicity Ratio) je definován poměrem signálu k dysperiodicitě [4],
- **DUV** (degree of unvoiced segments), který udává míru neznělých segmentů v hlasu [14, 3],
- **VSQ** (Vowel Similarity Quotient) měří stabilitu periodicity tónu hlasu [12].

Dalším parametrem kvality hlasu je **VVQ** (Vowel Variability Quotient), který je definován směrodatnou odchylkou doby fonace samohlásky. Tímto parametrem lze určit stabilitu načasování uzavření a otevření hlasivek [12].

4.3 Respirační a fonační parametry

Následující parametry se často řadí mezi fonační nebo případně artikulační. Podle Testu 3F je lze zařadit spíše mezi respiračně fonační parametry, protože jimi lze sledovat např. dobu expirace při fonaci a synchronizaci respirace s fonací.

- Maximální doba fonace **MPT** (Maximum Phonation Time). Jako řečovou úlohu lze použít prodlouženou fonaci samohlásek.
- Doba navození řeči **VOT** (Voice Onset Time). Parametr VOT měří dobu od zákmitu plozivní souhlásky do začátku fonace samohlásky. Tato doba má jiné typické hodnoty pro souhlásky bilabiální ([p], [b]), alveolární ([t], [d]) a velární ([k], [g]), které se obvykle používají pro úlohy diadochokineze a hodnocení artikulace. Tento parametr by tak mohl být zařazen do části artikulace, nicméně jím lze sledovat i synchronizaci respirace s fonací. V případě Testu 3F je tato synchronizace testována fonací labiodentální souhlásky [f] a samohlásky [i].
- **VOT ratio** – je definován jako poměr délky dob souhlásky a samohlásky [12].

4.4 Další prozodické parametry

Analýza prozodie zahrnuje kromě kmitočtu základního tónu a intenzity řeči také rytmus, tempo a pauzy v řeči. Pro testování prozodie jsou v Testu 3F použity rytmické texty, věty vyjadřující pocity pro analýzu intonace a věty se změnou důrazu. Bylo zjištěno, že tyto řečové úlohy jsou efektivní i při akustické analýze řeči [6].

4.4.1 Tempo řeči

Parametry popisující tempo řeči jsou vyjádřené ve slabikách za sekundu. Jsou odvozeny z následujících třech parametrů popisujících dobu trvání řeči a pauzy v řeči:

1. celková doba trvání řeči TST (Total Speech Time),
2. celková doba trvání pauz TPT (Total Pause Time) (pauza je počítána jako tichý úsek delší než 10 ms),
3. trvání řečového signálu bez pauz NST (Net Speech Time).

Po detekci hlásek a slabik v řeči a pomocí výše uvedených parametrů už lze tempo vyjádřit ve slabikách za sekundu:

- **Absolutní tempo: TSR** (Total Speech Rate) je počítán jako poměr počtu slabik v řeči a doby TST.
- **Tempo řeči bez pauz: NSR** (Net Speech Rate) je počítán jako poměr počtu slabik v řeči a doby NST.
- **Artikulační tempo: AR** (Articulation Rate) je počítán podobně jako parametr NSR ale pauzy jsou delší než 50 ms.

4.4.2 Pauzy v řeči

Netypické pauzy v řeči jsou jedním z příznaků dysartrie. V řeči je lze analyzovat následujícími parametry:

- **Procentuální poměr pauz: PPT** (Percent Pause Time) se počítá z parametrů trvání řeči a trvání pauz: TPT/TST , je vyjádřen v procentech. Délka pauz TPT je větší než 60 ms [15].
- **Počet pauz: NoP** (Number of Pauses) je definován jako počet pauz vztažené k trvání řeči TST.
- **Procentuální poměr mezislovních pauz: RIWP** (Ratio of Itra-Word Pauses) je vyjádřen poměrem celkové délky mezislovních pauz (TIPT) (Total Intra-word Pause Time) a času TPT.

4.4.3 Rytmus řeči

- **Index rytmu: SPIR** (Speech Index of Rhythmicity) vyjadřuje počet mezi-slovních pauz za minutu [6].
- **Procentuální poměr dysfluentních slov: PDW** (Percent Dysfluent Word) je definován poměrem celkového počtu dysfluentních slov a počtu všech slov [17].
- Analýza rytmu pomocí krátkodobé Fourierovy transformace (STFT) a DTW (Dynamic Time Warping): Algoritmus porovnává rytmus analyzované řečové nahrávky s referenční a zjišťuje odchylku v rytmu [15, s. 357].

4.5 Artikulační parametry

Při analýze artikulace se testuje výslovnost samohlásek a souhlásek. Testuje se přesnost i tempo artikulace. Pro testování se používají úlohy čtení vět a textu, monolog a opakování hlásek (Diadochokineze s fonací).

4.5.1 Artikulace samohlásek (Formantové kmitočty)

Parametry používané pro analýzu artikulace samohlásek jsou založené na identifikaci prvních dvou formantů. Lze tak měřit změny rezonancí v dutinách vokálního traktu a z toho nepřímo odvodit hybnost jazyka [12]. Toto odvození je založeno na principu Hellwagova trojúhelníku, který zobrazuje vertikální a horizontální pozici jazyka při výslovnosti jednotlivých samohlásek. Byl popsán i vliv pohybu jazyka na formanty F1 a F2 a z toho odvozen graf **VSA** (Vowel Space Area – vokální oblast hlasového traktu) znázorňující závislost samohlásek na prvních dvou formantech. Z toho vyplývá, že lze samohlásky rozpoznat podle spektrálního charakteru formantů a podle normativních dat určit přesnost artikulace samohlásek.

Podle grafu VSA byl zaveden stejnojmenný akustický parametr, který je počítán z tzv. rohových samohlásek (vrcholy polygonu VSA – samohlásky [a], [u], [i]). Charakter formantů se mění s věkem a je také závislý na pohlaví, což může zkreslovat výsledky analýzy. Proto byly zavedeny parametry, které se tuto závislost snaží eliminovat: **F2i/F2u** (poměr druhých formantů samohlásek [i] a [u]) a další parametry odvozené z VSA: centralizační poměr formantů **FCR** (Formant Centralization Ratio), artikulační index samohlásek **VAI** (Vowel Articulation Index), směrodatná odchylka formantů F1 a F2 relativní ke střední hodnotě: **relF1SD**, **relF2SD**.

Při analýze samohlásek u pacientů s Parkinsonovou nemocí bylo zjištěno, že přesnost automatizované diagnózy se liší v závislosti na použitých řečových úlohách a parametrech [16]. Největší přesnosti bylo dosaženo při analýze monologu pomocí parametrů VSA (80.4 %) a F2i/F2u (80 %).

4.5.2 Tempo artikulace hlásek a slabik

Další parametry popisují tempo výslovnosti hlásek a slabik. Pro testování se používají úlohy čtení textu a diadochokineze (opakování slabik: [pa], [ta], [ka], nebo hlásek: [p], [t], [k], [o], [e]). Parametr **DDK rate** počítá tempo řeči vyjádřené ve slabikách za sekundu a je odvozen z parametru AR (artikulační tempo). Další parametr **DDK regularity** měří délku jednoho cyklu diadochokineze a počítá jejich směrodatnou odchylku. Parametr **RFPC** (Robust Formant Periodicity Correlation) podle druhého formantu analyzuje rychlost artikulace slabik při přechodu ze souhlásky na samohlásku (tzv. F2 slope) [15].

4.6 Respirační parametry

Test 3F hodnotí respiraci podle úrovně intenzity a doby výdechu (nebo syčení), rovnoměrnosti a kolísání intenzity, rozsahu intenzity, plynulosti zeslabování a zesilování sykotu. Následujícími parametry lze některé úlohy analyzovat:

- maximální doba expirace: **MET** (Maximum Expiration Time),
- relativní pokles intenzity: **RRIS** (Robust Relative Intensity Slope). Tento parametr počítá celkový pokles intenzity normalizovaný časem TST [15]. Analýza řeči probíhá na úlohách diadochokineze. Pro analýzu řečových úloh monolog a čtení textu je potřeba odstranit pauzy delší než 50 ms.

Tab. 4.1: Akustické parametry pro kvantifikaci dysartrií

Parametr	Porucha řeči	Řečová úloha	Popis
Fonační			
F2i/F2u	Nestabilita pozice artikulačních orgánů	Prodloužená fonace samohlásek	Poměr druhých formantů samohlásek [i] a [u]
relFOSD	Nepravidelné kolísání výšky	Prodloužená fonace samohlásek	Směrodatná odchylka F0 normalizována ke střední hodnotě F0
Jitter	Mikroperturbace kmitočtu	Prodloužená fonace samohlásek	Míra kolísání rozsahu základního tónu hlasu
Shimmer	Mikroperturbace amplitudy	Prodloužená fonace samohlásek	Míra kolísání intenzity
HNR	Vyšší úroveň šumu	Prodloužená fonace samohlásek	Poměr harmonického signálu k šumu (Harmonic-to-Noise Ratio)
SSD	Vyšší úroveň šumu	Prodloužená fonace samohlásek	Poměr signálu k dysperiodicitě (Segmental Signal-to-Dysperiodicity Ratio)
DUV	Aperiodicita	Prodloužená fonace samohlásek	Míra neznělých segmentů v řeči (degree of unvoiced segments)

Pokračování tabulky 4.1

Parametr	Porucha řeči	Řečová úloha	Popis
Prozodické			
relF0SD	Monotónnost	Čtení vět a textu, monolog	Směrodatná odchylka F0 normalizována ke střední hodnotě F0
relF0VR	Monotónnost	Čtení vět a textu, monolog	Kolísání tónového rozsahu hlasu, normalizováno střední hodnotou F0
relSEOSD	Nízká variabilita hlasitosti	Čtení vět a textu, monolog	Směrodatná odchylka intenzity normalizována střední hodnotou
relSEOVR	Nízká variabilita hlasitosti	Prodloužená expirace/syčení	Rozsah intenzity normalizován střední hodnotou
AR	Nepřirozené tempo řeči	Čtení vět a textu, monolog	Artikulační tempo, počet slabik za sekundu (Articulation Rate)
PPT	Dlouhé pauzy	Čtení vět a textu, monolog	Procentuální poměr pauz (Percent Pause Time)
NoP	Nevhodné pauzy	Čtení vět a textu, monolog	Počet pauz v celkové době řeči (Number of Pauses)
RIWP	Nevhodné pauzy	Čtení vět a textu, monolog	Procentuální poměr mezislovních pauz (Ratio of IntraWord Pauses)
SPIR	Nevhodné pauzy	Čtení vět a textu, monolog	Počet mezislovních pauz za minutu (Speech Index of Rhythmicity)
Artikulační			
VSA	Omezená hybnost artikulačních orgánů	Monolog, čtení textu	Vokální oblast hlasového traktu (Vowel Space Area)
VAI	Omezená hybnost artikulačních orgánů	Monolog, čtení textu	Artikulační index samohlásek (Vowel Articulation Index)
F2i/F2u	Omezená hybnost artikulačních orgánů	Monolog, čtení textu	Poměr druhých formantů samohlásek [i] a [u]
relF1SD, relF2SD	Omezená hybnost artikulačních orgánů	Monolog, čtení textu	Směrodatná odchylka prvního (F1) a druhého (F2) formantu normalizována střední hodnotou F1 nebo F2
Respirační			
MET	Nedostatečná expirace	Prodloužená expirace/syčení	Maximální doba expirace (Maximum Expiration Time)
RRIS	Pokles intenzity	Čtení textu, monolog	Relativní pokles intenzity v čase NST (Robust Relative Intensity Slope)

5 Návrh objektivizace Testu 3F

Test 3F lze objektivizovat pomocí akustické analýzy řečových úloh. Nejdříve je databáze parametrizována. Následně lze výsledky parametrizace zpracovat statistickou analýzou a strojovým učením – regresní analýza. Dále se natrénuje regresní model na známých datech. Natrénovaný regresní model lze následně použít pro odhad celkového skóre testu (dysartrický index) u nových řečníků. Tento proces lze rozdělit na několik kroků:

1. **Záznam řeči:** Řeč je nahrávána mikrofonom do paměti nahrávacího zařízení (PC/smartphone). Kvalita nahrávky závisí na rušení v nahrávacím prostředí (akustické i elektromagnetické) a také na kvalitě nahrávací techniky (mikrofon, kabely, zesilovač, AD převodník). Vzdálenost mezi řečníkem a mikrofonom musí být stejná po celou dobu měření. Je sestavena databáze řečových nahrávek, která je rozdělena podle řečníků.
2. **Předzpracování dat:** Pro záznam každého řečníka jsou poslechem nalezeny časové pozice úloh a jsou zapsány pro pozdější načítání úloh z databáze.
3. **Akustická analýza:** Řečový signál je nejdříve předzpracován (ustřednění, preemfázová filtrace, segmentace). Následně jsou z řeči vypočítány akustické parametry – databáze je parametrizována. Výsledkem je tabulka vypočítaných parametrů pro všechny řečníky – dataset. Návrh parametrů, které lze použít pro automatizaci Testu 3F, je uveden v tabulce 5.1.
4. **Trénování modelu:** V tomto kroku je na datasetu (parametrizovaná databáze) natrénován regresní model. K automatizovanému odhadu skóre Testu 3F lze využít regresní modely lineární, nelineární a ensemble modely. Např. lineární modely: lineární regrese, regulované regrese (Lasso, Elastic Net, Ridge). Nelineární: k nearest neighbors, decision trees. Ensemble modely: náhodné lesy RF (Random Forests), XGBoost (eXtreme Gradient Boosting). Po trénování je model validován na testovacích datech – je zjištěna chyba predikce.
5. **Automatizované hodnocení:** Natrénovaným modelem je podle vstupních parametrů (parametrizovaná řeč testovaného řečníka) odhadnuta výstupní hodnota, tzn. celkové skóre testu.

Tab. 5.1: Přiřazení akustických parametrů k položkám Testu 3F

Položka Testu 3F	Akustický parametr
Respirace	
4.1 Klidová prodloužená expirace	MET, relSEOSD
4.2 Síla výdechového proudu	Tato položka nebude hodnocena kvůli nutné kalibraci mikrofonu
4.3 Výdrž expirace při syčení (sss...)	MET
4.4 Opakování sérií (ss-ss-ss...)	Směrodatná odchylka doby trvání jedné expirace
4.5 Plynulé zesilování a zeslabování sykotu	relSEOVR, koeficient šikmosti/špičatosti SEO profilu
Respirace při fonaci	
5.1 Výdrž expirace při prodloužené fonaci hlásky m	MET
5.2 Výdrž expirace při prodloužené fonaci hlásky ííí	MET
5.3 Synchronizovanost respirace s fonací fffííí	Tato položka nebude hodnocena z důvodu náročnosti automatizovaného zpracování
5.4 Délka výdechové mluvní fráze	AR
5.5 Mluvní respirace	Rytmus: SPIR. Pauzy: PPT, NoP, RIWP. Tempo: AR.
Fonace	
6.1 Kvalita hlasu	HNR, SSD, DUV, Jitter, Shimmer
6.2 Rezonance	HNR, SSD
6.3 Přiměřená a ovládaná hlasitost	relSEOSD, relSEOVR
6.4 Přiměřená a ovládaná výška	relF0VR, relF0SD
6.5 Hlasový rozsah	Rozsah F0 vyjádřen v semitónech
Artikulace	
7.1 Přesnost opakování samohlásek	F2i/F2u, VSA, VAI, FCR
7.2 Přesnost opakování souhlásek	relF1SD, relF2SD
7.3 Přesnost artikulace při čtení TEXT	relF1SD, relF2SD
7.4 Diadochokineze s fonací (p-t-k) (o-e)	Tato položka nebude hodnocena z důvodu náročnosti automatizovaného zpracování
7.5 Spontánní řeč	relF1SD, relF2SD
Prozódie	
8.1 Udržení rytmu v rytmickém celku TEXT	SPIR, PPT, NoP, RIWP
8.2 Základní větné intonace TEXT	relF0VR, relF0SD
8.3 Přemístování kontrastního důrazu TEXT	Porovnání řečové nahrávky se vzorem pomocí algoritmu DTW
8.4 Intonační variabilita TEXT	relF0VR, relF0SD
8.5 Komplexní prozodické faktory	Všechny prozodické parametry
Srozumitelnost	
9.1 Srozumitelnost předříkávaných slov TEXT	relF1SD, relF2SD a všechny prozodické parametry
9.2 Srozumitelnost předříkávaných vět TEXT	relF1SD, relF2SD a všechny prozodické parametry
9.3 Srozumitelnost četby slov TEXT	relF1SD, relF2SD a všechny prozodické parametry
9.4 Srozumitelnost četby textu TEXT	relF1SD, relF2SD a všechny prozodické parametry
9.5 Srozumitelnost spontánní mluvy	relF1SD, relF2SD a všechny prozodické parametry

6 Databáze a metodologie

Následující kapitola popisuje databázi řečových nahrávek, implementaci parametrů, parametrizaci řečových úloh, korelaci parametrů, nalezení nejvhodnějšího regresního algoritmu, důležitost parametrů pro analýzu a vytvoření norem z parametrizovaných dat.

6.1 Databáze

Databáze obsahuje celkem 163 česky mluvících řečníků, z toho 72 žen a 91 mužů. Pro analýzu řeči jsou k dispozici všechny řečové úlohy Testu 3F (viz tabulka 5.1). Řečníci v databázi jsou subjektivním hodnocením rozděleni na zdravé (54) a pacienty (109).

Předzpracování dat: Řečové nahrávky v databázi jsou rozděleny podle řečníka, tzn., že všechny úlohy pro daného řečníka jsou nahrány v jednom záznamu. Pro každý záznam je potřeba manuálně zapsat začátek a konec všech úloh. V rámci diplomové práce byla použita část databáze s již označenými nahrávkami, tj. 12 ze 30 úloh od každého řečníka. Zbývající úlohy nebyly v rámci diplomové práce označeny. V případě, že nahrávka byla nepoužitelná, tak byla v procesu označování vyřazena. Následně byli z databáze vyloučeni řečníci jejichž subjektivní skóre nebylo uvedeno (tj. řečníci: K1053, K1056, P1056). Velikost databáze se tak snížila na 151 řečníků.

Tab. 6.1: Databáze

databáze	počet řečníků	ženy	muži	zdraví	pacienti
původní	163	72	91	54	109
předzpracovaná	151	65	86	51	100

6.2 Metody zpracování dat a parametrizace

6.2.1 Implementace akustických parametrů

Pro implementaci akustických parametrů byla použita python knihovna Parselmouth [8], prostřednictvím které lze spouštět algoritmy programu Praat napsané v C/C++ [2]. Praat obsahuje funkce na čtení audio souborů a funkce pro analýzu řeči, např:

harmonic-to-noise ratio, jitter, shimmer, formanty, kmitočet základního tónu, intenzita a další. Většina těchto algoritmů nejdříve signál segmentuje. Načtené audio soubory byly nejdříve filtrovány preemfázovou filtrací. Kromě Praatu bylo použito Speech-to-text API od Google pro implementaci parametru AR (artikulační tempo) a pro identifikaci začátků a konce slov, díky kterým lze vypočítat počet pauz v průběhu slov a další parametry (RIWP, SPIR). Při implementaci se ale ukázalo, že identifikace časů slov je nepřesná, API nerozpozná pauzy v řeči. Časy slov byly kontinuální, konec jednoho slova byl začátkem dalšího. Přesnější identifikace časů slov bylo dosaženo sledováním intenzity řeči pomocí funkcí Praatu. Následně byla nastavena limitní úroveň intenzity, která rozděluje řeč na slova a pauzy. Délka okna pro identifikaci byla volena spíše experimentálně, ale tak, aby slabiky nebyly započítány jako slova. Získáním přesnějších časových značek bylo možné implementovat parametry počítající pauzy v průběhu slov.

6.2.2 Parametrizace

V následujícím kroku byly parametrizovány úlohy, pro které byla k dispozici předzpracovaná data (označené časy úloh v nahrávkách). Jak je uvedeno v návrhu objektivizace Testu 3F (kapitola 5), některé úlohy nebyly parametrizovány z důvodů uvedených v tabulce 5.1. Seznam parametrizovaných úloh je uveden v tabulce 6.2. V případě, že u některých z řečníků v databázi chyběl záznam úlohy (při předzpracování byl vyřazen), tak byl automaticky nahrazen přehraným záznamem. Pokud i ten chyběl, tak výsledek této úlohy byl nahrazen hodnotou NaN.

Funkci parametrizace lze snadno rozšířit pro výpočet dalších úloh. První krok je přiřazení funkce pro parametrizaci dané úlohy. Název funkce musí respektovat konvenci názvu ostatních funkcí. Další krok je přidání parametru do seznamu parametrů (`list_of_features`). Parametrizace si pak najde příslušnou nahrávku v adresáři, pokud jsou zachovány konvence názvu funkcí, parametrů a nahrávek.

6.3 Statistická analýza

V rámci statistické analýzy byla provedena korelace parametrů a následně regresní analýza. Před analýzou je nutné dataset parametrizovaných úloh zpracovat tak, aby neobsahoval chybějící hodnoty. Jak bylo popsáno v předešlé kapitole, důsledkem chybějících nahrávek v databázi je, že tyto parametrizované úlohy mají hodnotu NaN. Pokud by se chybějící hodnoty nezpracovali, mohlo by dojít ke zkreslení výsledků korelace a regresní analýzy. Chybějící hodnoty lze zpracovat dvěma způsoby. Prvním je vyřadit z datasetu řečníky, jejichž parametrizované úlohy obsahují chybějící hodnoty. Tím by se ovšem zredukovala velikost datasetu ze 151 řečníků na 89, dataset

Tab. 6.2: Parametrizované úlohy

Položka Testu 3F	Akustický parametr
Respirace při fonaci	
5.1 Výdrž exspirace při prodloužené fonaci hlásky m	MET
5.2 Výdrž exspirace při prodloužené fonaci hlásky ííí	MET
Artikulace	
7.1 Přesnost opakování samohlásek	F2i/F2u, VSA, VAI, FCR
Prozódie	
8.1 Udržení rytmu v rytmickém celku TEXT	SPIR, PPT, NoP, RIWP
8.2 Základní větné intonace TEXT	relF0VR, relF0SD
8.3 Přemístování kontrastního důrazu TEXT	Porovnání řečové nahrávky se vzorem pomocí algoritmu DTW
8.4 Intonační variabilita TEXT	relF0VR, relF0SD
Srozumitelnost	
9.1 Srozumitelnost předříkávaných slov TEXT	relF1SD, relF2SD a všechny prozodické parametry
9.2 Srozumitelnost předříkávaných vět TEXT	relF1SD, relF2SD a všechny prozodické parametry
9.4 Srozumitelnost četby textu TEXT	relF1SD, relF2SD a všechny prozodické parametry

by se zmenšil o 41 %. Počet chybějících hodnot u jednoho řečníka byl maximálně 5 z celkových 43. Nejčastěji to byly parametry v úloze 7.1. Vyřazením řečníků z datasetu by došlo k vyřazení velkého množství relevantních dat. Proto byla použita druhá možnost: nahrazení chybějících dat statistickou hodnotou vypočítanou z dat všech řečníků pro daný parametr (metoda: data imputation). V tomto případě byl vypočítán medián hodnot pro daný parametr.

6.3.1 Korelační analýza

Dále byla zjištěna korelace parametrů se subjektivním skóre příslušné úlohy. Subjektivní hodnocení je bodováním rozděleno do tří úrovní: 0, 1, 2 (viz kapitola 3). Proto byla použita Spearmanova korelace, která počítá pořadovou korelaci (rank correlation). Významnost korelace byla hodnocena podle p hodnoty, která vyjadřuje pravděpodobnost s jakou je korelace způsobená náhodnou veličinou. Korelace byla považována za významnou pokud p hodnota byla menší než 0.05. Dále byla zjištěna multikolinearita parametrů.

6.3.2 Regresní analýza

V rámci regresní analýzy byla zjištěna závislost subjektivně hodnoceného dysartrického indexu na jednotlivých parametrech. Krosvalidací bylo provedeno hodnocení několika regresních modelů, z nichž byl vybrán ten s nejlepším skóre. Pro hodnocení modelů byla použita metrika: střední kvadratická chyba (mean squared error MSE). Následně byly pro nejlepší model zobrazeny i metriky: střední absolutní chyba (mean absolute error MAE), root-mean-square error (RMSE). Krosvalidace byla typu k -fold, kde parametr k byl roven 10. Dataset byl tedy rozdělen na 10 částí, z toho 9 částí bylo použito pro trénování modelu a jedna část pro validaci. Tento proces se opakoval 10krát tak, aby každá část byla použita jednou pro validaci a zbytek pro trénování modelu. Protože parametry mají jiný rozsah hodnot a jinou jednotku, tak byly před krosvalidací normalizovány. Hyperparametry lineárních, nelineárních i ensemble modelů byly voleny metodou *Grid Search* – byl stanoven seznam hodnot hyperparametru a následně byla provedena krosvalidace a zjištěno skóre pro kombinace hyperparametrů pro daný model. Pro lineární a nelineární modely byl hyperparametr α v rozsahu 0 až 1 s krokem 0.1. Ensemble modely byly hodnoceny s hyperparametrem *počet stromů* pro hodnoty: 50, 100, 500, 1000, 5000. Celkem bylo hodnoceno 215 kombinací hyperparametrů a modelů. V rámci regresní analýzy byla pro nejlepší model vypočítána významnost parametrů pro regresi (feature importance).

Tab. 6.3: Seznam hodnocených regresních modelů

Lineární modely	Nelineární modely	Ensemble modely
Lineární regrese	k-nearest neighbors	Ada Boost
Lasso	Decision Tree	Bagging
Ridge	Extra Tree	Random Forests
Elastic Net	Support Vector	Extra Trees
Huber regressor		Gradient Boosting
Lasso Lars		
Passive Aggressive regressor		
RANSAC regressor		
SGD regressor		
TheilSen regressor		

6.4 Vytvoření norem z akustických dat

Na základě výsledků v článku [10], byly vytvořeny normy z parametrizovaných dat. Dataset byl nejprve rozdělen na dvě skupiny podle pohlaví. Jak je uvedeno v článku zmíněném výše, normy pro ženy a muže se mírně liší ve všech oblastech Testu 3F a také v celkovém skóre testu (dysartrický index). Pro vytvoření norem z akustických dat jsou použity normy z oblasti fonorespirace, fonetiky a celkové skóre. Norma z akustických dat je vyjádřena statistickými hodnotami: průměr, směrodatná odchylka, minimální hodnota, 5. percentil, 1. kvartil, medián, 3. kvartil, 95. percentil a maximální hodnota.

Tab. 6.4: Normativní data Testu 3F (podle [10])

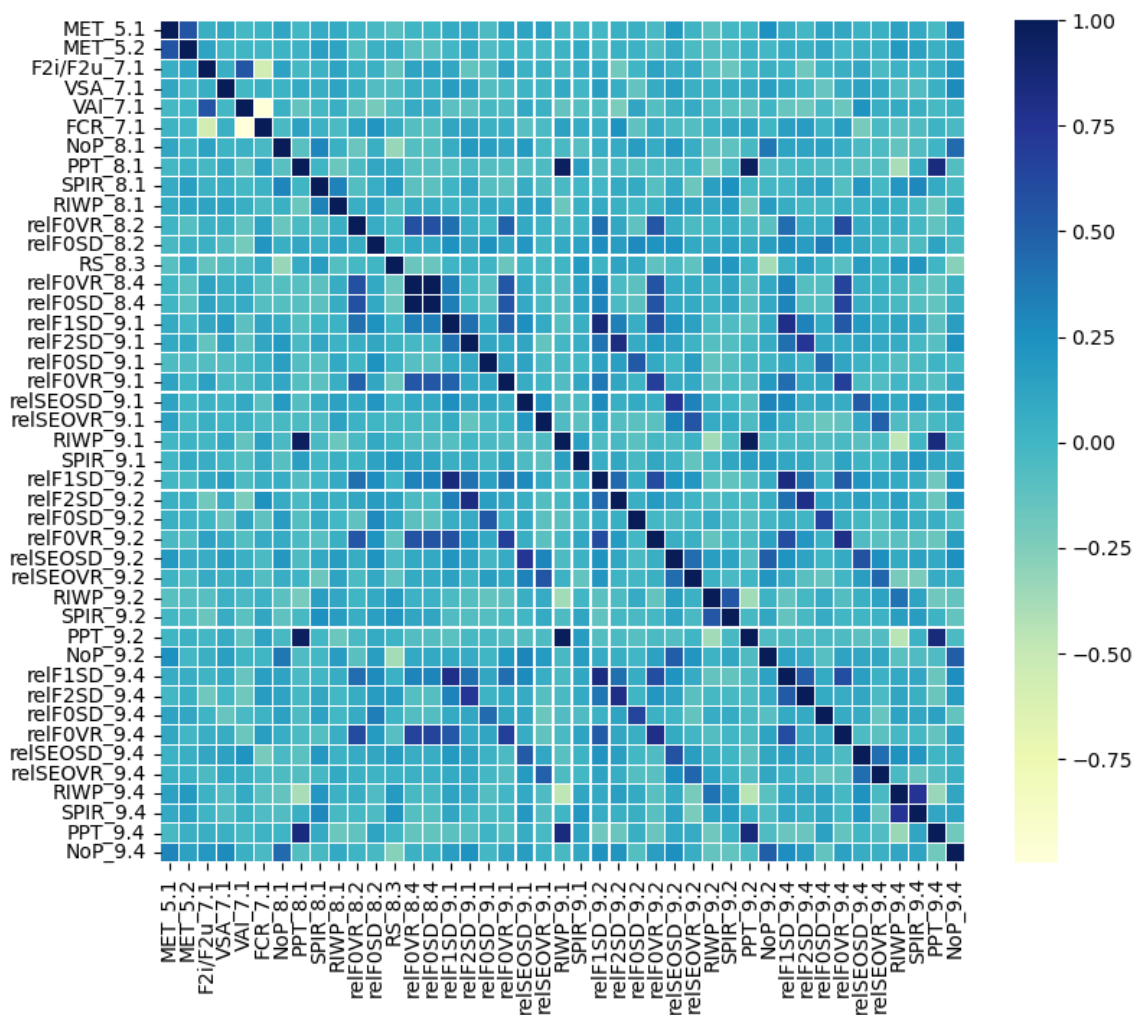
Oblast hodnocení	Ženy	Muži
Fonorespirace	≥ 25	≥ 27
Fonetika	≥ 28	≥ 26
Celkové skóre	≥ 80	≥ 79

7 Výsledky

V této kapitole jsou shrnuty výsledky korelační analýzy, regresní analýzy, významnost parametrů a jsou zde uvedeny normy akustických dat.

7.1 Korelační analýza

Na následujícím obrázku 7.1 je znázorněna multikolinearita parametrů. V korelační matici lze snadno identifikovat přímkou korelace a lze vidět korelaci mezi parametry stejného typu a z podobných úloh. Spearmanovu korelaci parametrů se subjektivním hodnocením úloh lze vidět v tabulce 7.1. Parametry, jejichž p hodnota je menší než 0.05 jsou vyznačeny tučně.



Obr. 7.1: Multikolinearita parametrů

Tab. 7.1: Spearmanova korelace parametrů se subjektivním skóre

parametr	rho	p
MET_5.1	0.536316	1.272533e-12
MET_5.2	0.280576	4.839478e-04
F2i/F2u_7.1	0.082876	3.116946e-01
VSA_7.1	-0.067961	4.070272e-01
VAI_7.1	-0.167133	4.025202e-02
FCR_7.1	0.173678	3.295028e-02
NoP_8.1	0.146607	7.244825e-02
PPT_8.1	-0.045093	5.824626e-01
SPIR_8.1	0.057662	4.818901e-01
RIWP_8.1	0.075477	3.570069e-01
relF0VR_8.2	-0.033644	6.817251e-01
relF0SD_8.2	0.138095	9.084664e-02
RS_8.3	-0.065362	4.252450e-01
relF0VR_8.4	0.098275	2.299460e-01
relF0SD_8.4	0.098275	2.299460e-01
relF1SD_9.1	0.149493	6.694466e-02
relF2SD_9.1	0.069330	3.976187e-01
relF0SD_9.1	0.070414	3.902664e-01
relF0VR_9.1	0.035207	6.678019e-01
relSEOSD_9.1	0.106162	1.945051e-01
relSEOVR_9.1	-0.127286	1.193618e-01
RIWP_9.1	-0.118620	1.468815e-01
SPIR_9.1	0.143535	7.870549e-02
relF1SD_9.2	0.166845	4.060285e-02
relF2SD_9.2	-0.017770	8.285437e-01
relF0SD_9.2	0.044426	5.880604e-01
relF0VR_9.2	0.141176	8.379827e-02
relSEOSD_9.2	0.117483	1.508223e-01
relSEOVR_9.2	-0.118470	1.473960e-01
RIWP_9.2	-0.166351	4.120920e-02
SPIR_9.2	-0.165858	4.182312e-02
PPT_9.2	-0.058741	4.737163e-01
NoP_9.2	-0.025736	7.537692e-01
relF1SD_9.4	0.152820	6.102715e-02
relF2SD_9.4	0.154742	5.780939e-02
relF0SD_9.4	-0.081852	3.177282e-01
relF0VR_9.4	0.070591	3.890684e-01
relSEOSD_9.4	-0.051949	5.264209e-01
relSEOVR_9.4	-0.056609	4.899424e-01
RIWP_9.4	0.036343	6.577512e-01
SPIR_9.4	0.028488	7.284246e-01
PPT_9.4	-0.038232	6.411668e-01
NoP_9.4	0.182975	2.452705e-02

7.2 Regresní analýza

V tabulce 7.2 je uvedeno skóre deseti nejlepších modelů. Modely jsou řazeny podle metriky MSE (mean squared error). Nejvíce se zde vyskytují modely s regulovanou lineární regresí (Lasso, ElasticNet).

Tab. 7.2: Deset nejlepších skóre regresní analýzy

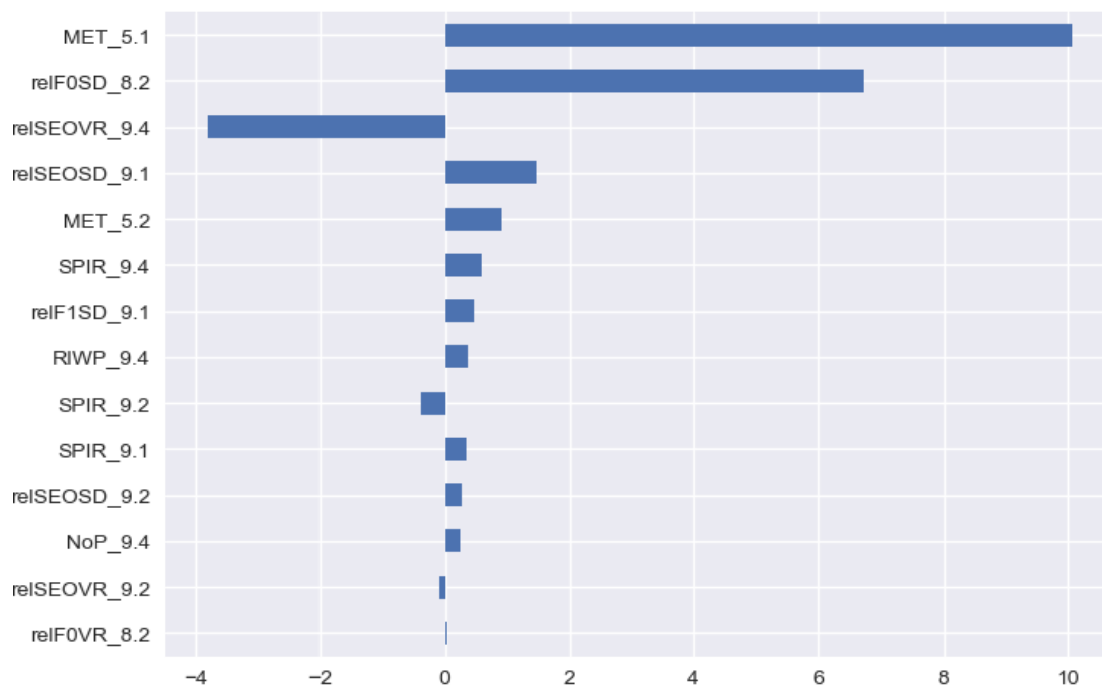
Pořadí	Model	Hyperparametr	MSE
1	Lasso	0.2	82.872
2	ElasticNet	0.2, 1.0	82.872
3	Lasso	0.3	84.776
4	ElasticNet	0.3, 1.0	84.776
5	AdaBoost	100	84.784
6	ExtraTrees	100	85.200
7	ElasticNet	0.1, 0.8	85.253
8	ElasticNet	0.1, 0.9	85.330
9	ElasticNet	0.2, 0.9	85.433
10	ElasticNet	0.1, 0.7	85.511

V tabulce 7.3 jsou uvedeny další metriky pro model s nejmenší chybou predikce. Maximální skóre dysartrického indexu je 90 bodů tzn., že procentuální chyba predikce vypočítána z MAE je 7.84 %.

Tab. 7.3: Metriky modelu s nejmenší chybou predikce (Lasso, $\alpha = 0.2$)

MAE	MSE	RMSE
7.0544	82.8723	9.1034

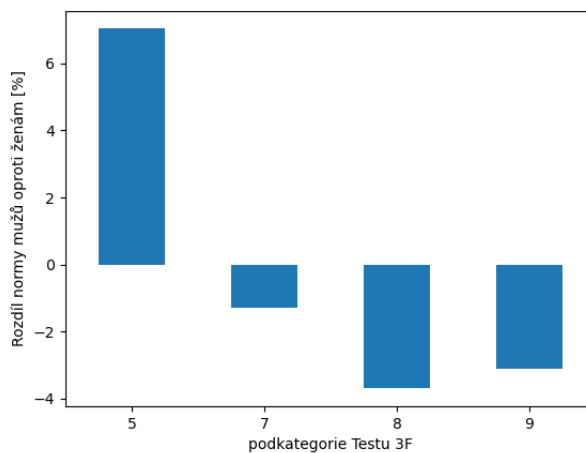
Na obrázku 7.2 jsou seřazeny významnosti parametrů podle jejich absolutní hodnoty. Záporná hodnota koeficientu indikuje nepřímou lineární závislost. Parametry s nulovou významností nejsou zobrazeny. Výsledky jsou pro regulovanou regresi Lasso, která zároveň provádí selekci parametrů tím, že dojde k vynulování těch koeficientů regrese, které jsou pro predikci nedůležité.



Obr. 7.2: Významnost parametrů (Lasso, $\alpha = 0.2$)

7.3 Normy z akustických dat

Na dalších dvou stranách jsou v tabulkách 7.4 a 7.5 pro všechny parametry uvedeny normativní data zvláště pro ženy a muže. Rozdíl normy mužů oproti ženám, vyjádřený v procentech je znázorněn na obrázku 7.3. Rozdíl je uveden pro podkategorie Testu 3F. Muži měli o 7.03 % vyšší skóre v úlohách 5 než ženy. V ostatních podkategoriích měli vyšší skóre ženy.



Obr. 7.3: Rozdíl normy mužů oproti ženám

Tab. 7.4: Norma ženy

	mean	std	min	5%	25%	50%	75%	95%	max
MET_5.1	16.68	4.43	7.73	9.16	13.70	17.18	19.83	23.15	23.38
MET_5.2	12.38	3.87	5.92	7.55	9.74	11.74	14.55	20.11	21.76
F2i/F2u_7.1	1.05	0.08	0.86	0.93	1.00	1.05	1.10	1.15	1.21
VSA_7.1	285829	133611	120110	137628	157066	277286	390103	526267	548373
VAI_7.1	0.42	0.03	0.37	0.38	0.41	0.42	0.44	0.46	0.47
FCR_7.1	2.38	0.15	2.14	2.17	2.26	2.37	2.43	2.62	2.73
NoP_8.1	17.63	2.46	13.00	13.90	16.00	18.00	18.75	22.00	23.00
PPT_8.1	114.33	104.52	42.81	52.94	88.96	100.95	108.75	118.21	659.03
SPIR_8.1	54.82	16.52	28.56	31.66	41.47	55.79	64.88	82.03	95.37
RIWP_8.1	42.52	22.96	1.31	9.41	17.10	53.89	58.94	68.48	69.18
relF0VR_8.2	0.96	0.17	0.61	0.76	0.85	0.92	1.05	1.21	1.47
relF0SD_8.2	0.28	0.11	0.11	0.18	0.21	0.24	0.37	0.49	0.54
RS_8.3	4.21	0.91	0.00	3.75	4.07	4.31	4.45	4.88	6.18
relF0VR_8.4	0.90	0.13	0.62	0.70	0.82	0.88	0.98	1.08	1.22
relF0SD_8.4	0.90	0.13	0.62	0.70	0.82	0.88	0.98	1.08	1.22
relF1SD_9.1	0.28	0.05	0.18	0.21	0.25	0.29	0.32	0.35	0.35
relF2SD_9.1	0.13	0.03	0.09	0.09	0.12	0.13	0.15	0.17	0.20
relF0SD_9.1	0.13	0.06	0.04	0.06	0.08	0.12	0.15	0.21	0.32
relF0VR_9.1	0.95	0.15	0.60	0.78	0.84	0.91	1.06	1.21	1.23
relSEOSD_9.1	0.10	0.01	0.08	0.08	0.09	0.10	0.11	0.12	0.13
relSEOVR_9.1	0.78	0.05	0.69	0.71	0.74	0.77	0.81	0.86	0.87
RIWP_9.1	943.00	780.28	177.14	447.91	779.91	863.52	911.46	1037.96	5020.44
SPIR_9.1	148.08	16.32	102.30	123.23	140.68	152.76	158.91	170.46	175.14
relF1SD_9.2	0.24	0.04	0.18	0.19	0.21	0.23	0.27	0.31	0.35
relF2SD_9.2	0.12	0.02	0.07	0.09	0.10	0.12	0.13	0.14	0.18
relF0SD_9.2	0.18	0.06	0.09	0.11	0.15	0.17	0.21	0.29	0.35
relF0VR_9.2	0.95	0.15	0.66	0.74	0.84	0.97	1.07	1.14	1.35
relSEOSD_9.2	0.12	0.01	0.08	0.10	0.11	0.12	0.13	0.14	0.15
relSEOVR_9.2	0.76	0.06	0.67	0.69	0.72	0.76	0.79	0.85	0.95
RIWP_9.2	1.85	0.67	0.43	0.97	1.54	1.76	2.18	2.69	4.29
SPIR_9.2	37.40	9.98	20.78	25.07	30.72	36.28	40.41	58.90	59.99
PPT_9.2	631.52	385.09	347.39	381.59	528.14	580.52	623.81	749.88	2638.17
NoP_9.2	6.46	0.71	5.00	5.30	6.00	6.40	6.80	7.40	8.00
relF1SD_9.4	0.27	0.04	0.19	0.21	0.25	0.27	0.30	0.33	0.36
relF2SD_9.4	0.13	0.02	0.08	0.10	0.12	0.13	0.14	0.17	0.17
relF0SD_9.4	0.30	0.10	0.15	0.18	0.23	0.30	0.36	0.49	0.52
relF0VR_9.4	0.98	0.14	0.76	0.79	0.90	0.97	1.02	1.18	1.46
relSEOSD_9.4	0.14	0.02	0.11	0.11	0.13	0.14	0.16	0.16	0.17
relSEOVR_9.4	0.85	0.04	0.78	0.80	0.82	0.85	0.87	0.91	0.93
RIWP_9.4	16.54	3.08	9.74	12.19	14.69	16.33	18.37	20.69	24.40
SPIR_9.4	57.81	9.73	36.45	41.52	52.88	57.29	65.26	73.08	78.04
PPT_9.4	90.65	20.33	62.90	72.22	80.85	87.78	94.31	109.29	183.89
NoP_9.4	142.13	24.99	98.00	113.50	133.00	139.00	149.00	164.00	252.00

Tab. 7.5: Norma muži

	mean	std	min	5%	25%	50%	75%	95%	max
MET_5.1	17.13	5.34	6.90	9.29	13.59	16.12	20.74	25.34	27.39
MET_5.2	14.12	6.20	3.82	7.25	11.60	12.67	14.55	26.36	34.52
F2i/F2u_7.1	0.99	0.09	0.80	0.85	0.96	1.00	1.02	1.08	1.25
VSA_7.1	282214	173306	117538	122401	171901	220435	351494	618260	738609
VAI_7.1	0.42	0.03	0.37	0.37	0.40	0.42	0.44	0.47	0.48
FCR_7.1	2.39	0.19	2.10	2.12	2.26	2.39	2.53	2.71	2.75
NoP_8.1	16.30	2.86	11.00	11.60	15.00	16.00	18.00	20.40	22.00
PPT_8.1	108.27	12.40	65.93	98.63	100.87	107.99	114.31	126.81	141.45
SPIR_8.1	58.63	17.11	29.22	34.72	46.27	55.84	65.74	90.35	96.26
RIWP_8.1	37.62	28.29	7.68	8.68	12.07	45.21	55.86	66.79	134.60
relF0VR_8.2	0.90	0.20	0.61	0.63	0.77	0.84	1.00	1.28	1.46
relF0SD_8.2	0.33	0.16	0.07	0.12	0.18	0.37	0.41	0.56	0.69
RS_8.3	4.45	0.41	3.78	3.95	4.10	4.39	4.76	5.13	5.40
relF0VR_8.4	0.80	0.17	0.44	0.60	0.67	0.78	0.93	1.05	1.18
relF0SD_8.4	0.80	0.17	0.44	0.60	0.67	0.78	0.93	1.05	1.18
relF1SD_9.1	0.28	0.06	0.18	0.21	0.24	0.27	0.31	0.37	0.45
relF2SD_9.1	0.14	0.03	0.08	0.09	0.13	0.14	0.16	0.18	0.18
relF0SD_9.1	0.14	0.09	0.03	0.04	0.07	0.13	0.18	0.30	0.37
relF0VR_9.1	0.81	0.18	0.36	0.51	0.73	0.81	0.88	1.08	1.19
relSEOSD_9.1	0.10	0.01	0.07	0.08	0.09	0.10	0.10	0.11	0.12
relSEOVR_9.1	0.76	0.08	0.63	0.67	0.70	0.76	0.80	0.87	1.00
RIWP_9.1	895.90	230.14	107.85	499.03	862.83	943.29	1008.52	1098.74	1324.92
SPIR_9.1	142.37	14.50	117.91	124.06	132.67	139.15	154.38	168.18	176.87
relF1SD_9.2	0.24	0.05	0.16	0.17	0.21	0.23	0.26	0.31	0.41
relF2SD_9.2	0.13	0.03	0.07	0.09	0.12	0.13	0.15	0.17	0.17
relF0SD_9.2	0.26	0.11	0.08	0.11	0.18	0.25	0.33	0.41	0.54
relF0VR_9.2	0.83	0.19	0.47	0.55	0.71	0.83	0.94	1.14	1.22
relSEOSD_9.2	0.11	0.01	0.09	0.09	0.10	0.11	0.12	0.13	0.13
relSEOVR_9.2	0.76	0.06	0.65	0.67	0.71	0.76	0.81	0.85	0.89
RIWP_9.2	2.27	1.10	0.89	1.28	1.62	2.09	2.57	4.02	6.52
SPIR_9.2	44.31	13.24	22.15	28.47	33.15	41.04	55.12	69.60	72.44
PPT_9.2	620.84	109.35	218.79	440.25	590.33	638.34	683.25	727.76	749.28
NoP_9.2	5.87	0.71	4.20	4.52	5.40	6.00	6.40	6.80	7.20
relF1SD_9.4	0.25	0.05	0.17	0.18	0.21	0.25	0.29	0.33	0.38
relF2SD_9.4	0.13	0.03	0.07	0.09	0.11	0.13	0.16	0.17	0.18
relF0SD_9.4	0.41	0.21	0.08	0.11	0.25	0.38	0.60	0.67	0.86
relF0VR_9.4	0.86	0.17	0.58	0.62	0.72	0.86	0.97	1.12	1.27
relSEOSD_9.4	0.15	0.02	0.11	0.12	0.14	0.15	0.16	0.16	0.20
relSEOVR_9.4	0.86	0.06	0.76	0.78	0.82	0.86	0.91	0.96	0.98
RIWP_9.4	18.28	2.85	12.93	13.29	16.70	18.15	19.54	22.52	26.83
SPIR_9.4	59.27	11.53	38.42	41.13	50.44	61.69	65.85	75.72	78.22
PPT_9.4	92.26	9.23	70.62	80.22	86.18	91.41	98.68	109.56	110.59
NoP_9.4	131.27	21.82	76.00	99.40	119.00	130.00	144.00	163.80	178.00

8 Diskuse

Z korelační matice (7.1) je patrné, že parametry, které hodnotí podobné vlastnosti řeči spolu korelují. Jsou to například parametry z úlohy 7.1.: F2i/F2u, VSA, VAI, FCR. Všechny tyto parametry jsou počítány z formantů. Konkrétně mezi parametrem VAI a FCR je nepřímá závislost, což je dáno právě jejich výpočtem: $VAI = 1/FCR$. Dalším příkladem je parametr MET, který je počítán pro úlohy 5.1 a 5.2 (viz tabulka 6.2). Jedná se o téměř stejné úlohy prodloužené fonace, jediný rozdíl je změna fonované hlásky. Z výsledků multikolinearity lze vyvodit, že některé parametry jsou pro hodnocení nadbytečné a predikce by mohla být přesnější, pokud budou některé korelující parametry vyřazeny.

Výsledkem Spearmanovy korelace parametrů se subjektivním skóre příslušné úlohy je osm parametrů, jejichž p hodnota je menší než 0.05. Nejvýznamnější korelaci má parametr MET, počítaný pro úlohu 5.1 a 5.2. Nejvýznamnější korelaci s úlohou 7.1 měli parametry VAI, FCR. Středně silnou korelaci (0.53) má pouze parametr MET (5.1), ostatní korelační koeficienty z těchto osmi parametrů jsou v rozsahu od 0.16 do 0.28.

Výsledky regresní analýzy ukazují, že nejpřesnější predikce bylo dosaženo při selekci 14 parametrů. Zbývající parametry byly vyřazeny nulovými koeficienty regrese Lasso. Stejně jako u Spearmanovy korelace je parametr MET (úloha 5.1) hodnocen jako nejvýznamnější. Další v pořadí jsou převážně parametry hodnotící prozódii řeči: relF0SD (8.2) – intonace, relSEOVR (9.4) – variace rozsahu intenzity, dále srozumitelnost slov: relSEOSD (9.1) – variace intenzity. Významnost dalších parametrů je výrazně nižší, jsou to především prozodické parametry hodnotící pauzy v řeči (RIWP, NoP, SPIR).

Ve srovnání se Spearmanovou korelací si lze všimnout, že úloha 7.1 (Přesnost opakování samohlásek) a její parametry F2i/F2u, VSA, VAI, FCR byly regresní analýzou vyřazeny. Z toho lze odvodit, že artikulace má na celkové hodnocení minimální vliv, nicméně z celé podkategorie artikulace byla hodnocena jen jedna úloha z celkových pěti. Pokud by byly označeny časové pozice i pro ostatní úlohy, tak by se výsledky mohly lišit.

V rámci regresní analýzy a selekce parametrů byla otestována metoda RFE (Recursive Feature Elimination) v kombinaci s krosvalidací – RFECV. Metoda provádí výběr parametrů podle skóre krosvalidace na jejím interním modelu. Výstupem této metody je kombinace parametrů s nimiž bylo dosaženo nejmenší chyby predikce (na interním modelu). Výsledkem tohoto testu byl seznam parametrů stejný jako u regrese Lasso. Významnost parametrů byla mírně odlišná. S těmito parametry byla následně provedena krosvalidace všech modelů a výsledkem byla chyba predikce 7.14 % (model: Passive aggressive regression).

Ve výsledcích regresní analýzy bez selekce parametrů (tab. 7.2), lze vidět, že většina modelů s nejmenší chybou jsou typu regulované regrese, která penalizuje model v případě, že je použito velké množství parametrů nebo korelujících parametrů. U těchto modelů dochází přímo k selekci parametrů tak, aby bylo dosaženo nejmenší chyby predikce.

Z předchozích analýz a testů vyplývá, že selekce parametrů má na predikci vliv. Regresní analýzou byly zjištěny parametry, které mají největší vliv na celkové skóre testu (dysartrický index) (obr. 7.2). Z korelace parametrů se subjektivním skóre pro příslušnou úlohu a také z regresní analýzy vyplývá, že nejdůležitější parametr pro predikci je MET (maximum expiration time) vypočítaný z úlohy 5.1: Výdrž expirace při prodloužené fonaci hlásky m. Další významné parametry hodnotí převážně prozódii řeči – intonace, intenzita, pauzy.

Analýza parametrů byla provedena na nekompletním datasetu. Bylo použito 12 z 30 úloh dostupných v databázi. Oblast Fonace a Respirace nebyla v analýze z tohoto důvodu zahrnuta. V datasetu byly obsaženy převážně delší řečové úlohy (čtení textu, vět a slov), které testují prozódii. Kvůli nekompletnosti datasetu nelze přímo určit, která z oblastí Testu 3F je pro hodnocení dysartrie v rámci akustické analýzy nejvýznamnější. Ale v rámci analýzy provedené na limitovaném datasetu, lze konstatovat, že úroveň dysartrie lze určit nejlépe parametrem MET z doby prodloužené fonace hlásky m a dále z prozodických úloh (relF0SD 8.2, relSEOVR 9.4). Z toho vyplývá, že u pacientů s dysartrií se nejvíce projevuje porucha respirace: nedostatečná expirace, monotónnost a nízká variabilita intenzity řeči.

Závěr

Tato diplomová práce se zabývá nalezením vhodných akustických parametrů pro objektivizaci Testu 3F. Testem 3F lze hodnotit míru dysartrie. Dysartrie je komplexní řečová porucha, která postihuje několik oblastí řeči (fonace, prozodie, artikulace a respirace). Cílem práce bylo navrhnout a implementovat parametry, které lze klinicky interpretovat a efektivně je použít pro automatickou analýzu řečových úloh Testu 3F.

V prvních kapitolách práce byla popsána problematika neurologických procesů podílejících se na tvorbě řeči. Následně byla definována dysartrie, její příznaky a diagnóza pomocí Testu 3F. V dalších kapitolách byly prozkoumány možnosti, jak řeč analyzovat pomocí akustických parametrů. Pro každou oblast řeči byly vybrány parametry, kterými lze analyzovat vlastnosti řeči pro tyto oblasti specifické. U každého parametru byla definována řečová úloha a porucha řeči, kterou jím lze analyzovat. Z těchto parametrů byly následně vybrány ty, které lze použít pro analýzu úloh Testu 3F. V rámci návrhu bylo zjištěno, že automaticky lze analyzovat 27 úloh z 30.

V kapitole č. 6 byla popsána databáze a dále metody předzpracování databáze, implementace akustických parametrů a parametrizace. V rámci diplomové práce byly implementovány všechny parametry uvedené v návrhu, nicméně bylo parametrizováno pouze 10 úloh Testu 3F. Zbývající úlohy nebyly v rámci práce předzpracovány, a tudíž nebyly parametrizovány. Byla provedena analýza multikolinearity parametrů a bylo vyhodnoceno, že některé parametry a úlohy mohou být pro regresi přebytečné. Následně byla zjištěna korelace akustických parametrů se subjektivním skóre pro příslušnou úlohu (korelační analýza) a dále korelace parametrů s dysartrickým indeksem (regresní analýza). V rámci regresní analýzy bylo krosvalidací otestováno 215 kombinací modelů s jejich hyperparametry. Následně byl vybrán model s nejmenší chybou predikce (Lasso, $\alpha = 0.2$) a byla zjištěna významnost parametrů pro predikci (koeficienty regrese). Významnost parametrů byla následně ověřena selekcí parametrů metodou RFE (Recursive Feature Elimination). S modelem Lasso bylo dosaženo chyby predikce 7.84 %. Přestože model Lasso provádí i selekci parametrů, tak v případě kdy byla použita metoda RFE a následná regresní analýza, chyba predikce se snížila na 7.14 % (model: Passive aggressive regression). Celkem bylo pro analýzu použito 43 kombinací parametrů s úlohami a 14 z nich bylo vyselektováno jako významné pro predikci. Nejvýznamnější korelaci měl parametr MET (úloha 5.1). Mezi vybranými parametry jsou převážně prozodické parametry, což je dáno především tím, že pro analýzu byla použita pouze část dostupné databáze. Proto nebyly analyzovány následující podkategorie Testu 3F: Respirace, Fonace. Výsledky selekce parametrů by se mohly lišit, pokud by byly použity i další úlohy. V rámci současných výsledků s limitovanou databází lze uvést, že úroveň dysartrie lze nej-

lépe určit na úloze prodloužené fonace (5.1) analyzované parametrem MET a dále na prozodických úlohách (relF0SD 8.2, relSEOVR 9.4).

Nakonec byly z parametrizovaných dat vytvořeny normy zvlášt pro ženy a muže.

Diplomová práce by měla navazovat zpracováním zbývajících dat – označením časových pozic úloh v záznamu a následně implementací parametrizačních funkcí pro tyto úlohy. Jak bylo uvedeno výše, všechny akustické parametry uvedené v návrhu (tab. 5.1) jsou implementovány. Následně provedením statistické analýzy pro všechny oblasti Testu 3F by mohlo být docíleno zajímavějších výsledků.

Literatura

- [1] ACKERMANN, H.; HERTRICH, I.: Speech rate and rhythm in cerebellar dysarthria: An acoustic analysis of syllabic timing. *Folia Phoniatrica et Logopaedica*, 46(2), 1994: s. 70–78, doi:<https://doi.org/10.1159/000266295>. URL http://www.academia.edu/download/40390911/folia-phoniat_46_70.pdf
- [2] BOERSMA, P.; WEENINK, D.: Praat: doing phonetics by computer [Computer program]. 2021. URL <http://www.praat.org/>
- [3] BRABENEC, L.; MEKYSKA, J.; GALAZ, Z.; AJ.: Speech disorders in Parkinson's disease: early diagnostics and effects of medication and brain stimulation. *Journal of neural transmission*, 124(3), 2017: s. 303–334.
- [4] CSÉFALVAY, Z.; MEKYSKA, J.; KOŠŤÁLOVÁ, M.: Diagnostika dysartrie. In *Diagnostika narušené komunikační schopnosti u dospělých*, Z. Cséfalvay; V. Lechta, Portál, vyd. 1, 2013, ISBN 978-80-262-0364-3, str. 117–143.
- [5] DUFFY, J. R.: *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*. St. Louis, Mo: Elsevier Mosby, 2012, ISBN 978-0-323-07200-7.
- [6] GALAZ, Z.; MEKYSKA, J.; MZOUREK, Z.; AJ.: Prosodic analysis of neutral, stress-modified and rhymed speech in patients with Parkinson's disease. *Computer methods and programs in biomedicine*, 127, 2016: s. 301–317.
- [7] HICKOK, G.; HOUDE, J.; RONG, F.: Sensorimotor Integration in Speech Processing: Computational Basis and Neural Organization. *Neuron*, 69(3), 2011: s. 407–422, ISSN 0896-6273, doi:<https://doi.org/10.1016/j.neuron.2011.01.019>. URL <http://www.sciencedirect.com/science/article/pii/S0896627311000675>
- [8] JADOUL, Y.; THOMPSON, B.; DE BOER, B.: Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, 71, 2018: s. 1–15, doi:<https://doi.org/10.1016/j.wocn.2018.07.001>.
- [9] KENT, R. D.; KENT, J. F.; WEISMER, G.; AJ.: What dysarthrias can tell us about the neural control of speech. *Journal of Phonetics*, 28(3), 2000: s. 273–302, ISSN 0095-4470, doi:<https://doi.org/10.1006/jpho.2000.0122>. URL <http://www.sciencedirect.com/science/article/pii/S0095447000901220>
- [10] KOŠŤÁLOVÁ, M.; MRAČKOVÁ, M.; MAREČEK, R.; AJ.: Test 3F Dysarthrický profil – normativní hodnoty řeči v češtině. *Cesk Slov Neurol N*, 76/109(5), 2013: s. 614–618, ISSN 1210-7859.

- [11] LEVELT, W. J.: Models of word production. *Trends in cognitive sciences*, 3(6), 1999: s. 223–232, doi:[https://doi.org/10.1016/S1364-6613\(99\)01319-4](https://doi.org/10.1016/S1364-6613(99)01319-4).
URL <http://www.sciencedirect.com/science/article/pii/S1364661399013194>
- [12] NOVOTNÝ, M.; RUSZ, J.; ČMEJLA, R.; AJ.: Automatic Evaluation of Articulatory Disorders in Parkinson’s Disease. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(9), 2014: s. 1366–1378, doi:<https://doi.org/10.1109/TASLP.2014.2329734>.
- [13] ROUBÍČKOVÁ, J.; ET AL.: *Dysartrický profil test 3F*. Praha: Galén, 2011, ISBN 978-80-7262-714-1.
- [14] RUSZ, J.; HLAVNIČKA, J.; TYKALOVÁ, T.; AJ.: Quantitative assessment of motor speech abnormalities in idiopathic rapid eye movement sleep behaviour disorder. *Sleep medicine*, 19, 2016: s. 141–147.
- [15] RUSZ, J.; ČMEJLA, R.; RŮŽIČKOVÁ, H.; AJ.: Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson’s disease. *The Journal of the Acoustical Society of America*, 129(1), 2011: s. 350–367, doi:<https://doi.org/10.1121/1.3514381>.
- [16] RUSZ, J.; ČMEJLA, R.; TYKALOVÁ, T.; AJ.: Imprecise vowel articulation as a potential early marker of Parkinson’s disease: Effect of speaking task. *The Journal of the Acoustical Society of America*, 134(3), 2013: s. 2171–2181, doi:<https://doi.org/10.1121/1.4816541>.
- [17] TYKALOVÁ, T.; RUSZ, J.; ČMEJLA, R.; AJ.: Effect of dopaminergic medication on speech dysfluency in Parkinson’s disease: a longitudinal study. *Journal of neural transmission*, 122(8), 2015: s. 1135–1142, doi:<https://doi.org/10.1007/s00702-015-1363-y>.

Seznam symbolů, veličin a zkratek

AD převodník	analogově digitální převodník
APQ	(Amplitude Perturbation Quotient) kvocient pertrubace amplitudy
AR	(Articulation Rate) artikulační tempo
CCP	(Cepstral Peak Prominence) výskyt špičkových hodnot v kepstru
CNS	centrální nervový systém
DDK rate	(Diadochokinetic rate) tempo diadochokineze
DDK regularity	(Diadochokinetic regularity) pravidelnost délky jednoho cyklu diadochokineze
DDA	(average absolute Difference between consecutive Differences between the Amplitudes of consecutive period) průměr absolutních hodnot rozdílů z po sobě následujících rozdílů amplitud po sobě následujících period
DDP	(average absolute Difference between consecutive Differences between consecutive Periods, divided by the average period) průměr absolutních hodnot rozdílů z po sobě následujících rozdílů délek (po sobě následujících) period, normalizován průměrnou délkou periody
DTW	(Dynamic Time Warping) dynamické borcení časové osy
DUV	(Degree of Unvoiced segments) míra neznělých segmentů
F0	kmitočet základního tónu hlasu
F2i	druhý formant samohlásky [i]
F2u	druhý formant samohlásky [u]
FCR	(Formant Centralization Ratio) centralizační poměr formantů
GNE	(Glottal-to-Noise Excitation ratio) poměr glotálního signálu k šumu
HNR	(Harmonic-to-Noise Ratio) poměr harmonického signálu k šumu

MET	(Maximum Expiration Time) maximální doba expirace
MPT	(Maximum Phonation Time) maximální doba fonace
NNE	(Normalized Noise Energy) normalizovaná energie šumu
NoP	(Number of Pauses) počet pauz
NS	nervový systém
NST	(Net Speech Time) celková doba trvání řeči bez pauz
NSR	(Net Speech Rate) tempo řeči bez pauz
PC	(Personal Computer) osobní počítač
PDW	(Percent Dysfluent Word) procentuální poměr dysfluentních slov
PNS	periferní nervový systém
PPQ	(Pitch Pertrubation Quotient) kvocient pertrubace kmitočtu
PPT	(Percent Pause Time) procentuální poměr pauz
RAP	(Relative Average Pertrubation) normalizovaná průměrná pertrubace kmitočtu
RFPC	(Robust Formant Periodicity Correlation) robustní formantová korelace periodicity
RF	(Random Forests) náhodné lesy – klasifikační algoritmus
RIWP	(Ratio of Itra-Word Pauses) procentuální poměr mezislovních pauz
RRIS	(Robust Relative Intensity Slope) robustní relativní pokles intenzity
SD	(Standard Deviation) směrodatná odchylka
SPIR	(Speech Index of Rhythmicity) index rytmu řeči
SSD	(Segmental Signal-to-Dysperiodicity Ratio) poměr signálu k dysperiodicitě
STFT	(Short-Time Fourier Transform) krátkodobá Fourierova transformace

TKEO	(Teager–Kaiser energy operator) Teagerův-Keiserův energetický operátor
TPT	(Total Pause Time) celková doba trvání pauz
TSR	(Total Speech Rate) absolutní tempo
TST	(Total Speech Time) celková doba trvání řeči
VAI	(Vowel Articulation Index) artikulační index samohlásek
VOT	(Voice Onset Time) doba od zákmitu plozivní souhlásky do začátku fonace samohlásky
VOT ratio	(Voice Onset Time ratio) poměr délky dob souhlásky a samohlásky
VR	(Variation Range) kolísání rozsahu
VSA	(Vowel Space Area) vokální oblast hlasového traktu
VSQ	(Vowel Similarity Quotient) kvocient podobnosti samohlásek
VVQ	(Vowel Variability Quotient) kvocient variability samohlásek
XGBoost	(eXtreme Gradient Boosting) – klasifikační algoritmus
ZCR	(Zero Crossing Rate) počet průchodů signálu nulovou úrovní

9 Obsah přiloženého DVD

Přiložené DVD obsahuje elektronickou verzi diplomové práce a adresář *Test_3F*, ve kterém jsou všechny skripty, které byly v rámci diplomové práce naprogramovány. Dále je zde příručka k těmto skriptům. Databáze řečových nahrávek není přiložena.

```
Test_3F
├── data
│   ├── dataset
│   ├── labeled
│   ├── reference
│   ├── results
│   ├── speech_to_text
│   └── test3f_tasks
├── test_3F
│   ├── __init__.py
│   ├── audio_preprocessing.py
│   ├── build_features.py
│   ├── correlation_analysis.py
│   ├── features.py
│   ├── regression_analysis.py
│   ├── score_norms.py
│   ├── speech_to_text.py
│   └── utils.py
├── environment.yml
├── setup.py
├── Diplomova_prace_Bezusek.pdf
└── Prirucka_systemu.pdf
```

10 Příručka pro použití vytvořeného systému

Pro spuštění skriptů je potřeba mít správně nadefinované virtuální prostředí Pythonu, které je definováno v *environment.yml*. Pomocí *package manageru Conda* to lze udělat následujícím příkazem v terminálu:

```
conda env create -f environment.yml
```

Adresář *test_3F* je Python balíček, jsou zde skripty pro předzpracování audio signálu (*audio_preprocessing.py*), akustické parametry (*features.py*), převod řeči na text (*speech_to_text.py*) – výstup je ukládán do adresáře: *speech_to_text*.

V rámci této práce byla na text převedena pouze úloha 8.1. Ostatní úlohy nebylo potřeba v rámci práce převádět. Pro převod řeči na text je potřeba mít vlastní přístupový klíč ke službě *Google Cloud*. Soubor *.json* s tímto klíčem je potřeba uložit do adresáře *speech_to_text* s názvem *key.json*. Nicméně všechny transkripce úlohy, která byla v rámci diplomové práce převáděna, jsou příloze práce.

Dále skript *utils.py* obsahuje pomocné funkce pro načítání řečových úloh a také obsahuje definice cest k adresáři *data*. Skript *build_features.py* parametrizuje řečové úlohy – výstup je tabulka ve formátu *.csv*, která je uložena do adresáře *dataset*. Korelační analýza je naprogramována v *correlation_analysis.py*, její výstup je tabulka korelačních koeficientů, která je uložena do adresáře *results* ve formátu pdf. Výsledky regresní analýzy (*regression_analysis.py*) a vytvořené normy (*score_norms.py*) jsou ukládány také do adresáře *results*. Adresář *test3f_tasks* obsahuje textový soubor se seznamem slabik úlohy 5.4, který je potřeba pro vypočítání parametru AR z této úlohy. Dále je zde definován seznam úloh, které budou parametrizovány v rámci skriptu *build_features.py*. Pokud seznam neexistuje, je vytvořen nový podle parametrů definovaných ve skriptu parametrizace → *list_of_features*. Adresář *labeled* slouží k umístění předzpracovaných dat. Adresář *reference* slouží k umístění referenčních nahrávek z úlohy 8.3 pro porovnání rytmu dvou nahrávek – parametr *rhythm_similarity* (RS).