

Diplomová práce



České  
vysoké  
učení technické  
v Praze

**F3**

Fakulta elektrotechnická  
Katedra radioelektroniky

## Automatická akustická analýza artikulační rychlosti u pacientů s parkinsonismem

**Bc. Vojtěch Illner**

Vedoucí: Doc. Ing. Jan Ruzs, PhD.  
Obor: Otevřené elektronické systémy  
Květen 2020



## I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Illner** Jméno: **Vojtěch** Osobní číslo: **457432**  
Fakulta/ústav: **Fakulta elektrotechnická**  
Zadávající katedra/ústav: **Katedra radioelektroniky**  
Studijní program: **Otevřené elektronické systémy**  
Studijní obor: **Komunikace a zpracování signálu**

## II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

**Automatická akustická analýza artikulační rychlosti u pacientů s parkinsonismem**

Název diplomové práce anglicky:

**Automated Acoustic Analysis of Articulation Rate in Patients with Parkinsonism**

Pokyny pro vypracování:

1. Seznamte se s motorickými poruchami řeči u Parkinsonovy nemoci a atypických parkinsonských syndromů. Na základě dostupné literatury dále nastudujte možnosti detekce řečové aktivity a artikulační rychlosti.
2. Navrhněte algoritmus pro automatickou detekci artikulační rychlosti ve spojitě řeči. Pro návrh algoritmu využijte výpočetní prostředí MATLAB. Navržený algoritmus otestujte na vybraném vzorku řečových promluv zdravé populace, pacientů Parkinsonovy nemoci a pacientů multisystémové atrofie. Výsledky porovnejte s konvenčním algoritmem pro detekci rychlosti řeči.
3. S pomocí vyvinutého algoritmu navrhněte sadu řečových příznaků pro popis artikulační rychlosti a vhodnými statistickými testy vyhodnoťte nejvhodnější parametry pro odlišení zdravých jedinců a jednotlivých neurologických onemocnění.

Seznam doporučené literatury:

- [1] Skodda S, Schlegel U. Speech rate and rhythm in Parkinson's disease. *Mov Disord* 2008;23, 985-992.
- [2] Jiao Y, Berisha V, Tu M, Liss J. Convex weighting criteria for speaking rate estimation. *IEEE ACM T Audio Speech Lang Proces* 2015;23:1421-1430.
- [3] Wang D, Narayanan SS. Robust speech rate estimation for spontaneous speech. *IEEE T Audio Speech Lang Proces* 2007;15:2190-2201

Jméno a pracoviště vedoucí(ho) diplomové práce:

**doc. Ing. Jan Rusz, Ph.D., katedra teorie obvodů FEL**

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

Datum zadání diplomové práce: **31.01.2020**

Termín odevzdání diplomové práce: **22.05.2020**

Platnost zadání diplomové práce: **30.09.2021**

\_\_\_\_\_  
doc. Ing. Jan Rusz, Ph.D.  
podpis vedoucí(ho) práce

\_\_\_\_\_  
doc. Ing. Josef Dobeš, CSc.  
podpis vedoucí(ho) ústavu/katedry

\_\_\_\_\_  
prof. Mgr. Petr Páta, Ph.D.  
podpis děkana(ky)

## III. PŘEVZETÍ ZADÁNÍ

Diplomant bere na vědomí, že je povinen vypracovat diplomovou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v diplomové práci.

\_\_\_\_\_  
Datum převzetí zadání

\_\_\_\_\_  
Podpis studenta



## Poděkování

Chtěl bych poděkovat Doc. Ing. Janu Ruszovi, PhD. za skvělé vedení této práce, bezproblémovou spolupráci a poskytnuté odborné znalosti. Dále bych rád poděkoval všem kteří mě podporovali při psaní této práce i v celém studiu na ČVUT FEL. Velký dík patří mým spolužákům, bez jejichž vzájemné morální podpory by bylo mé studium těžší a znatelně méně zajímavé.

## Prohlášení

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

V Praze, 22. května 2020

## Abstrakt

Poruchy řečové rychlosti představují častý příznak dysfunkcí řeči pro pacienty s parkinsonismem. Téma této práce bylo navrhnout automatickou akustickou metodu měření parametrů spojených s řečovou rychlostí.

Navržená nová metoda byla porovnána se čtyřmi konvenčními algoritmy odhadu rychlosti řeči. Byla použita nahraná data 95ti improvizovaných monologů od zdravých lidí, pacientů s Parkinsonovou nemocí (PN), poruchou chování v REM spánku (RBD) a multisystémovou atrofií (MSA). Sada příznaků řečové rychlosti byla určena jako celková řečová rychlost, celková artikulační rychlost a variabilita artikulační rychlosti.

Mezi všemi metodami poskytuje nejlepší výsledky v odhadu daných parametrů algoritmus Praat skript, s vysokou korelací k manuálně přepsaným referenčním hodnotám ( $r = 0.84$ ,  $p < 0.001$ ). Příznaky celkové řečové a artikulační rychlosti byly podle provedených statistických testů schopné výrazně rozlišit mezi zdravými lidmi a pacienty s MSA ( $p < 0.001$ ). Trend celkového zpomalení mluvy byl zpozorován u PN a RBD pacientů, avšak bez významných meziskupinových statistických rozdílů.

Robustní a přesné metody odhadu defektů řečové rychlosti mohou sloužit ke zvýšení efektivity většího systému na hodnocení řečových poruch u neurodegenerativních nemocí.

**Klíčová slova:** Synukleinopatie, rychlost řeči, artikulační rychlost

**Vedoucí:** Doc. Ing. Jan Rusz, PhD.

## Abstract

Speech rate abnormalities represent a common symptom of speech disorder in parkinsonism. The aim of this thesis was to design an automatic acoustic method allowing to assess speech rate related parameters.

One novel method was designed and compared to four conventional algorithms for speech rate estimation. Monologues were collected from 95 subjects, including healthy controls, Parkinson's disease (PD), rapid eye movement sleep behaviour disorder (RBD), and multisystem atrophy (MSA). The speech rate features including overall speech rate, articulation rate, and articulation rate variability were proposed.

Compared to all tested methods, the Praat script reached the best results in estimating speech rate with a strong correlation to the transcribed reference values ( $r = 0.84$ ,  $p < 0.001$ ). Statistical tests showed that the overall speech and articulation rate were able to significantly differentiate between healthy controls and MSA patients ( $p < 0.001$ ). Also, a trend of overall speech slowness was seen in PD and RBD; however, no significant statistical differences were demonstrated between the groups.

Reliable automatic measurement methods of speech rate abnormalities might serve as a vital part of a greater system for speech disorders evaluation in parkinsonism.

**Keywords:** Synucleinopathy, speech rate, articulation rate

**Title translation:** Automated acoustic analysis of articulation rate in patients with parkinsonism

## Obsah

<b>1 Úvod</b>	<b>1</b>	<b>2 Metodika</b>	<b>9</b>
1.1 Lidská řeč	2	2.1 Databáze promluv	9
1.1.1 Fyziologie	2	2.1.1 Účastníci studie	9
1.1.2 Základní hlasivková frekvence	2	2.1.2 Řečové nahrávky	10
1.1.3 Řečová rychlost	3	2.2 Návrh vlastního algoritmu	11
1.2 Synukleinopatie	4	2.2.1 Detektor spektrálních změn	11
1.2.1 PN	4	2.2.2 Klasifikace intervalů	17
1.2.2 RBD	4	2.2.3 Celkový popis algoritmu	21
1.2.3 MSA	5	2.2.4 Číselné hodnoty parametrů algoritmu	22
1.2.4 Poruchy řeči u synukleinopatií	5	2.3 Porovnávací metody	22
1.3 Postupy odhadu řečové rychlosti	6	2.3.1 Rozpoznávač	24
1.3.1 Problémy automatického odhadu řečové rychlosti	7	2.3.2 Theta Oscilátor	25
1.3.2 Dostupné akustické detektory pro odhad tempa řeči u dysartrie	8	2.3.3 LFME	27
1.4 Cíle práce	8	2.3.4 Praat skript	31
		2.4 Testování metod	32
		2.4.1 Kritéria testování a cílové veličiny	32
		2.4.2 Referenční hodnoty	34

2.5	Statistické zhodnocení . . . . .	34
2.5.1	Navržené příznaky . . . . .	34
2.5.2	Statistické nástroje . . . . .	35
<b>3</b>	<b>Výsledky</b>	<b>37</b>
3.1	Výsledky testování algoritmů . . .	37
3.2	Výsledky statistického zhodnocení	40
3.2.1	Výsledky testů předpokladů .	40
3.2.2	Celkové výsledky . . . . .	41
<b>4</b>	<b>Závěrečné zhodnocení</b>	<b>43</b>
4.1	Výsledky práce . . . . .	43
4.2	Navazující výzkum . . . . .	44
	<b>Literatura</b>	<b>47</b>



## Obrázky

2.1 Ilustrační obrázek průběhu výstupu BACD algoritmu pro promluvu „Dej to sem“ . . . . .	17	2.11 Principiální schéma fungování LFME algoritmu. . . . .	30
2.2 Ilustrační obrázek vyhlazení filtrovaného BACD výstupu pomocí kubické spline interpolace. . . . .	18	2.12 Ilustrační obrázek průběhu intenzity $I[n]$ pro promluvu „Dej to sem“ . . . . .	32
2.3 Ilustrační obrázek podoby intervalů na charakteristice BACD $[n]$ spolu se segmenty obsahující $F_0[n]$ . . . . .	21	2.13 Principiální schéma metody Praat skript. . . . .	33
2.4 Principiální schéma navrženého algoritmu. . . . .	22	3.1 Závislost referenčních a odhadnutých hodnot $v_{sr}$ a $v_{ar}$ algoritmem Praat skript. Černá čára představuje perfektní shodu. . . . .	40
2.5 Principiální schéma fonémového rozpoznávače. . . . .	25	3.2 Závislost referenčních a odhadnutých hodnot $v_{sr}$ a $v_{ar}$ algoritmem BACD detektor. Černá čára představuje perfektní shodu. . . . .	41
2.6 Schéma detektoru slabik na základě fonémového rozpoznávače, natrénovaného na české databázi. . . . .	25	3.3 Výsledky post-hoc testů pro příznaky $v_{sr}$ , $v_{ar}$ a $\sigma_{ar}$ odhadnuté pomocí metod Praat skript a BACD detektor. . . . .	42
2.7 Ilustrační obrázek podoby trajektorie zvučnosti (sonority) pro promluvu „Dej to sem“ . . . . .	27		
2.8 Schéma detektoru slabik theta Oscillator. . . . .	27		
2.9 Ilustrační obrázek LFME trajektorie pro promluvu „Dej to sem“ . . . . .	29		
2.10 Celkové schéma prahování u detektoru slabik LFME algoritmu. . . . .	30		

## Tabulky

2.1 Základní klinické charakteristiky skupin pacientů.....	10
2.2 Popis Bayesovského autoregresního rekurzivního algoritmu spektrálních změn (BACD).....	16
2.3 Seznam všech použitých parametrů navrženého algoritmu, spolu s číselnými hodnotami a popisem. .	23
3.1 Výsledky testování metod podle kritérií NRMSE a Spearmanova koeficientu $r$ pro HC, PN, RBD a MSA skupinu u cílových veličin $v_{sr}$ a $v_{ar}$ . . . . .	39
3.2 Výsledky testů jednocestné ANOVA a Kruskal-Wallis pro příznaky $v_{sr}$ , $v_{ar}$ a $\sigma_{ar}$ .....	41

# Kapitola 1

## Úvod

Parkinsonova nemoc (PN) představuje druhé nejčastější neurodegenerativní onemocnění na světě, po Alzheimerově chorobě [1]. Je odhadováno, že v lidské populaci starší 65ti let je prevalence nemoci 1.8% [1] a do budoucna se očekává pozvolný růst, díky stárnutí obyvatelstva a prodlužující se délky života [2]. I přes intenzivní vývoj neuroprotektivních léčebných metod v dnešní době neexistuje žádný postup, který by probíhající onemocnění plně potlačil. Dostupné léčby jsou pouze schopny ulevit od motorických příznaků, ale nemají vliv na samotný průběh nemoci.

Hlavní důvod, proč je úspěšné neuroprotektivní léčby těžké dosáhnout, spočívá v již rozvinutém stádiu nemoci, která může působit i mnoho let před prvními motorickými příznaky. Brzká diagnóza choroby, v jejím prodromálním stádiu, je tedy zásadní pro rozvoj případných neuroprotektivních postupů [3].

Porucha řeči patří mezi jeden z nejtypičtějších ukazatelů přítomnosti nemoci, detekovatelný již ve velmi brzkých stádiích [4]. Monitorování řečové kvality v čase může přispět k brzké diagnostice onemocnění, a to až několik let před prvními motorickými příznaky [5].

Sledování průběhu řečových příznaků představuje logisticky náročnou úlohu, jelikož je zapotřebí veliké množství vzorků (promluv) pro dané sledovací období. Vyhodnocování každé promluvy řečovým expertem je tak prakticky nemožné. Proto je veden intenzivní výzkum v oblasti automatických metod, vyhodnocujících příznaky popisující určité řečové vady.

Mezi méně zkoumané patří příznaky spojené s *rychlostí řeči* (řečovým tempem). Ukazuje se, že pacienti mohou mít vlivem nemoci odlišné charakteristiky v této oblasti než zdraví lidé [6]. Rychlost řeči a odvozené příznaky by tak mohly sloužit jako užitečný biomarker na klasifikaci závažnosti Parkinsonovy nemoci a jejích syndromů.

V této práci se budeme zabývat možnostmi automatické estimace řečové rychlosti u pacientů s parkinsonismem a případného klinického využití na

zhodnocení vývoje a stádia nemoci.

## 1.1 Lidská řeč

Lidská řeč je nejpřirozenější formou mezilidské komunikace a představuje objekt zájmu mnoha výzkumných odvětví. Pomocí řečové analýzy lze například rozpoznávat řeč, jazyk, a emocionální rozpoložení mluvčího. Pomocí vhodně zvolených příznaků lze také monitorovat průběh a závažnost některých onemocnění [7]. Převážně se jedná o neurodegenerativní typy, ale výsledky lze aplikovat například i v oblastech psychopatologie [8]. V této práci se budeme věnovat příznakům souvisejícím s řečovou rychlostí.

### 1.1.1 Fyziologie

Řeč je zvuková vlna tvořená soustavou vokálního traktu lidského těla. Tento proces se skládá ze tří hlavních částí: tvorba tlaku vzduchu v plicích, regulace zvukových vibrací a kontrola rezonancí. Z plic přichází vytvořený proud vzduchu a prochází hlasivkami, skládající se z hlasivkové štěrbinou a obklopujících hlasivkových svalů. Průchodem vzduchu štěrbinou dochází k rozkmitání hlasivek s danou frekvencí, regulovanou citlivou soustavou svalů okolo. Vzduchový proud, takto přeměněný na přerušované pulzy, dále prochází zúžením traktu a dostává se do jeho horní části. Zde se nachází rezonátory, dutina hrdelní, nosní a ústní. Ty zesilují přicházející zvuk a mění ho na konkrétní podobu. Výsledný tvar pak dotváří artikulátory, jako jazyk, rty a čelist [9].

Při tvorbě neznělých hlásek vzduch prochází traktem s větší rychlostí a dochází ke vzduchovým turbulencím, deformující původní tvar. Tento výsledný zvuk již přestává být funkcí frekvence hlasivek.

### 1.1.2 Základní hlasivková frekvence

Při průchodu vzduchu hlasivkovou štěrbinou dochází k jejímu rozkmitání, kdy se opakovaně otevírá a zavírá. Proces je ovlivněn zejména soustavou hlasivkových svalů a samotným proudem vzduchu z plic. V ideálním případě vzniká působením oscilací periodický impulzový signál  $U(t)$ . Pokud je  $T_0$  perioda mezi dvěma následujícími pulzy, základní frekvence hlasivek  $F_0$  je

dána jako inverze této hodnoty,

$$F_0 = \frac{1}{T_0}. \quad (1.1)$$

Z důvodu mnoha fyziologických aspektů vokálního traktu a samotných hlasivek nemá  $U(t)$  ideální periodický tvar. Signál se stává náhodným, nestacionárním a kvaziperiodickým, měnícím své vlastnosti v čase. Definice periody  $T_0$  tak přestane být zřejmá. Literatura se různí v názorech na  $F_0$  a samotné definici  $F_0$  [10]. Definice by v případě perfektně periodického signálu vedly na stejné výsledky. V praxi se výsledné odhady  $F_0$  mohou měnit v závislosti na definici, kterou daná metoda používá.

### 1.1.3 Řečová rychlost

Řečová rychlost, nebo také řečové tempo, představuje jednu ze základních charakteristik s velkým vlivem na výslednou podobu mluvy. Je dána podobou a vlastnostmi vokálního traktu člověka a schopnostmi artikulace. Je úzce spojená se *srozumitelností* a také s emocionálním vnímáním mluvy. Při zpomalování a zrychlování řeči dochází k protahování a zkracování jednotlivých hlásek. Světové jazyky mají každý vlastní průměrné řečové tempo, např. mezi rychlé patří italština.

Charakteristiky řečové rychlosti se používají mimo jiné ke zpřesnění metod rozpoznávání řeči [13, 14] i řečníka, testování srozumitelnosti řeči [17] a klasifikaci neurologických a jiných onemocnění [6]. Je ovlivněna mimo jiné délkou promluvy, variabilitou délek mluvených slov, pauz mezi nimi [11] i emocionálním rozpoložením [8].

Rychlost řeči se dá popisovat termíny jako "rychlá", "normální", "pomalá", což je ze sluchového hlediska přirozené nicméně vyžaduje lidské hodnocení a kategorizace nemá jasné hranice [15]. Alternativa je rychlost kvantifikovat jako počet fonetických prvků za danou časovou délku. Nejčastější a nejvíce blízká lidskému vnímání je volba počet slabik za vteřinu [12, 13].

### Slabika

Při volbě měřit řečovou rychlost jako slabiky za jednotku času spočívá úloha v jejich detekci. Nastává problém samotné definice slabiky, jelikož ta se napříč jazyky liší a není pevně daná. Většina autorů automatických metod detekce se přiklání k fonetické definici [18], popisující slabiku jako útvar „mající poměrně hlasité jádro, kde dochází k malé nebo žádné zábraně proudu vzduchu ve vokálním traktu, ohraničené méně hlasitými úseky s větší zábranou proudu

vzduchu“. Takový popis dává návod jak detekovat jednotlivé slabiky v řeči. Také splňuje předpoklad, že se zrychleným tempem řeči také úměrně roste i následný počet slabik.

## 1.2 Synukleinopatie

*Synukleinopatie* značí kategorii neurodegenerativních onemocnění spojené stejnou patofyziologií. V této práci se budeme zabývat Parkinsonovou nemocí, RBD (*Rapid eye movement sleep behavior disorder*) a MSA (*Multiple system atrophy*). RBD představuje tzv. *prodromální* stádium nemoci, kde se očekává vývoj do *klinického* stádia, kam patří PN i MSA.

### 1.2.1 PN

Při Parkinsonově chorobě dochází k hromadnému úbytku neuronových populací v mozku i tzv. *dopaminergních neuronů* nacházející se v bazálních gangliích. Následkem se tak projevují příznaky jako bradykineze, klidový třes, postojová nestabilita, snížená schopnost začít pohyb a další [19].

Hlavní motorické příznaky nemoci, které většinou vedou k sestavení diagnózy, se objevují až v pozdějších stádiích nemoci, kdy je neurodegenerace v pokročilé fázi a až 50% dopaminergních neuronů již mohlo být nenávratně poškozeno [20]. V tomto stádiu je prakticky nemožné sestavit neuroprotektivní léčbu, která by dokázala nemoc významně zpomalit či dokonce zastavit [3].

### 1.2.2 RBD

RBD je porucha spánku, při které mají pacienti narušenou jeho REM fázi. Dochází k motorické aktivitě a vokalizaci, typicky související s popisovanými živými sny. Při vážnějších projevech může docházet až ke zranění pacienta nebo člověka nacházejícího se blízko němu [21].

Přítomnost RBD představuje velmi silný prediktor vývoje nemoci do neurodegenerativní choroby, často Parkinsonovy. Je odhadováno, že po 4-9 letech po diagnóze RBD tato nemoc propuká a 11-16 let trvá do projevů prvních příznaků [22].

### ■ 1.2.3 MSA

MSA představuje méně častou neurodegenerativní chorobu. Je způsobena úbytkem neuronů v několika částech mozku, bazálních gangliích, malém mozku a dalších [23]. Prvotní příznaky nemoci jsou velmi podobné PN, časem se však objevují závažnější disfunkce, mj. poruchy autonomního nervového systému. Na rozdíl od PN jsou příznaky MSA velice těžko potlačitelné a nemoc je hůře léčitelná [23].

### ■ 1.2.4 Poruchy řeči u synukleinopatií

Řeč je nejkomplexnější motorická dovednost zapojující více než 100 svalů lidského těla a je velice náchylná na zhoubné změny v neuronových strukturách mozku, které zodpovídají za motorické funkce [26].

Až u 90% pacientů s PN se vyvinou poruchy řeči, kolektivně nazývané *hypokinetická dysartrie* [27]. Mezi tyto patří poruchy *respirace* (disfunkce dýchání, špatné frázování, obtížná změna hlasitosti), *fonace* (špatná produkce znělých hlásek), *artikulace* (problém zřetelně vyslovovat slabiky jdoucí za sebou, zadrhávání) a *prozódie* (souhrnné změny v hlasitosti, tónu a rytmu řeči, omezená variabilita) [27].

U RBD, vzhledem k tomu že se jedná o prodromální stádium, nemusí být projevy hypokinetické dysartrie tak značné. Primárním příznakem nemoci je porucha spánku a k řečovým disfunkcím může docházet až po jejím delším průběhu. Mezi známé příznaky patří konkrétně změny v základním tónu a rytmu (pomalejší střídání znělých a neznělých intervalů a prodloužená délka pauz) [52].

MSA představuje velmi závažné klinické stádium neurodegenerační nemoci. U prakticky všech pacientů se projevují aspekty poruchy řeči, často v závažné míře [24]. Kromě příznaků společných s PN dochází k výskytu výrazných variací intenzity hlasu, podstatnému vokálnímu třesu a díky svalovým křečím tzv. *strained-strangled* hlasu, vypjatému a zaškrcenému, který vypadá, že působí mluvčímu enormní námahu vyslovovat [25].

### ■ Poruchy rychlosti řeči

Při disfunkcích správné respirace a artikulace dochází k poruchám v oblasti řečové rychlosti. To se může zásadně podepsat na podobě řečového projevu. Mezi typické příznaky patří celkové zpomalování mluvy, zadrhávání, nenadálé

zrychlování v průběhu, ústící v nesrozumitelné úseky a nedokončování slov [28].

Pacienti s PN mají obecně velmi širokou škálu možné závažnosti v tomto aspektu poruchy řeči [28]. Jsou případy, kdy pacient nevykazuje žádnou abnormalitu v řečovém tempu a nemoc se projevuje jinak. Stejně tak se u některých případů jedná o jeden z primárních příznaků nemoci. Z postupnou progresí choroby se očekává, že se postupně poruchy řečové rychlosti projeví u většiny diagnostikovaných. V počátečních fázích však nemusí být patrné. Podle [52] je možné, že příznak vlastní PN je zvýšená řečová variabilita, nenadálé zrychlování a zpomalování v průběhu.

U RBD není oblast poruch řečové rychlosti významněji prozkoumána. Očekává se, že jako prodromální stádium nebudou příznaky tak zřetelné jako u těch klinických. Ukazuje se [52], že RBD pacienti vykazují obdobné dysfunkce řečového tempa jako PN, avšak v nižší míře.

MSA pacienti obvykle trpí podobnými příznaky jako PN, často ještě v závažnější míře [25]. Navíc se objevuje efekt velkého prodlužování fonémů. Obecně je u MSA z hlediska řečové rychlosti nejvíce pozorováno celkové zpomalení mluvy [25].

Zmíněné poruchy řečové rychlosti, lze pozorovat a hodnotit pomocí vhodně zvolených příznaků [4, 6, 24, 25, 52]. Jejich analýzou je možné stanovit přítomnost a závažnost dané dysfunkce a zhodnotit tak například stádium nemoci. Postupů, jak příznaky měřit je několik a podstatně se liší ve svých principech. Jsou popsány v následující části.

### 1.3 Postupy odhadu řečové rychlosti

Jedno ze zásadních dělení mezi metodami estimace řečové rychlosti spočívá ve formě jak veličinu udávat. První skupinu představují metody klasifikující řečové tempo do diskretních skupin, např. "rychlé", "pomalé"[15]. Druhý přístup mu dává rozměr počtu řečových útvarů za jednotku času. Jako nejčastěji používaný útvar se v literatuře používá *slabika*.

Další základní dělení postupů je mezi automatickými a s manuálním vyhodnocením. Metody s manuálním zhodnocením využívají řečové experty, stanovující úroveň rychlosti řeči, případně ručně vyhodnocující počty slabik a jejich umístění. Častá je také úloha, kdy pacient čte text s předem známým počtem slabik a manuální přepis pak není nutný.

Ruční zhodnocení má výhodu téměř 100% přesnosti stanovení příznaků. Velice významnou nevýhodou je však časová náročnost, zvláště při delších promluvách, případně fixace na předem daný text. Úloha zjištění rychlosti řeči tak bývá na klinikách často vynechávána, především právě z časových důvodů [31].



U automatických metod bývá klasifikace do diskretních skupin méně početná a převažuje odhad ve slabikách za sekundu.

Při volbě tohoto přístupu spočívá celá úloha v jejich detekci na daném řečovém signálu. Principiální schéma je podobné pro všechny tyto metody. Vstupní signál je nejdříve uzpůsoben vnitřním specifikům daného detektoru, např. podvzorkováním nebo filtrováním dolní propustí. Následně je ze signálu po segmentech stanovena určená charakteristika, či vícero charakteristik. Poté nastává jejich analýza, stanovující lokace významných bodů a úseků, označující možná umístění slabik.

Výsledný vektor s lokacemi se pak někdy porovnává s výstupem např. VAD detektoru (voice activity detector) a jsou vyřazeni kandidáti ve fyziologicky nemožných úsecích utvoření slabiky. Výstupem celého algoritmu je pak vektor s lokacemi jednotlivých slabik.

Mnoho metod využívá předtrénování na daném počtu testovacích promluv a známých referenčních hodnot k nastavení svých vnitřních parametrů [32].

Mezi nejznámější veřejně dostupné databáze na testování a trénování algoritmů patří ICSI Switchboard [35] a TIMIT [36]. Obě poskytují velké množství promluv anglicky mluvících zdravých lidí s manuálně vyznačenými slabikami. Metody natrénované pomocí Switchboard databáze mají tendenci být robustnější, jelikož nabízí větší různorodost nahrávek z hlediska tempa, hlasitosti aj. než TIMIT, kde jsou nahrávky více homogenní a zřetelně vyslovované [14]. Switchboard bývá považován za poměrně dobře reprezentující spontánní zdravou lidskou řeč.

### 1.3.1 Problémy automatického odhadu řečové rychlosti

Jelikož automatická detekce slabik není příliš zavedené ani prozkoumané odvětví, zatím žádný vyvinutý algoritmus není vyloženě robustní. Často velice záleží na podmínkách, hlavně délce textu, typu řečové úlohy (recitace slabik, čtení textu, improvizovaný monolog), způsobu nahrání promluvy, samotném mluvčím a v neposlední řadě také *jazyce* [31]. Často daná metoda funguje přesně a efektivně jen v omezeném počtu aplikací a jinde selhává.

Typicky jsou takto některé algoritmy dimenzovány na velmi krátké nahrávky, někdy i s fixním tematickým obsahem [33]. Tyto pak mají velký problém s např. improvizovanými monology o délce až několik minut. Také se ukazuje, že dostatečně robustní nejsou algoritmy postavené na základě dat od malého počtu mluvčích [34].

Všechny popsané problémy jsou pro naši úlohu stěžejní. Data budou pocházet od mnoha různých mluvčích, často s velice odlišným stylem řeči, který může být do velké míry ovlivněn probíhající nemocí. Nahrávky jsou improvizované monology, lišící se ve své délce. Celkově tak úloha představuje poměrně komplikovanou úlohu pro daný algoritmus odhadu řečové rychlosti.

### ■ 1.3.2 Dostupné akustické detektory pro odhad tempa řeči u dysartrie

Z důvodů uvedených v předchozí části jsou dosavadní vyvinuté automatické metody z velké části určené na konkrétní použití v poměrně zjednodušujících podmínkách. Opravdu robustních metod, se širokým využitím je zatím poměrně málo [32].

Stejně tak nebylo v nedávné době provedeno žádné porovnání většího množství algoritmů v různých podmínkách. Srovnání přináší možnosti nadhledu a orientace nad celým odvětvím. Dokáže tak případně poukázat na přesnou a robustní metodu, nebo naopak naznačit směr, ve kterém se zatím nedosáhlo žádných uspokojivých výsledků a je potřeba dalšího výzkumu. Poslední nalezené srovnání algoritmů na odhad rychlosti řeči je z roku 2007 [29].

Zároveň jsou dostupné existující metody vyvinuté pouze pro zdravé lidi a jen málo i pro pacienty s dysartrií [30, 32]. Z těchto prací má [32] velmi zjednodušující podmínky, úlohu krátkých čtených vět. První práce, [30], zahrnuje i úlohu řečového monologu, ale pouze délky pár větových celků.

Nebylo zatím provedeno žádné testování metod na širší spektrum stádií dysartrie, od nízkých po vysoké, objevující se např. u MSA.

## ■ 1.4 Cíle práce

Cílem této práce bude návrh nové automatické metody odhadu řečové rychlosti s motivací využití u dysartrických pacientů nižších a vyšších stádií jako typického projevu synukleinopatií.

Zároveň bude tato metoda porovnána s dalšími vybranými přístupy existujícími v literatuře, slibujícími dobré výsledky. Porovnáním algoritmů bude odhadnuta úroveň výzkumu v tomto směru, zjištěny nevhodné, nefunkční přístupy a naopak poukázáno na ty, které se za daných podmínek dobře uplatní.

Využitím vhodných automatických metod budou navrženy příznaky popisující aspekty řečové rychlosti. Statistickými metodami pak bude vyhodnoceno, které parametry se ukazují jako nejvhodnější na popis poruch řečové rychlosti a na odlišení zdravých jedinců a jednotlivých neurologických onemocnění.

Tyto příznaky by mohly sloužit jako vhodný popis řečové rychlosti, využitý jako jeden z prvků na stanovení diagnózy, vyhodnocení závažnosti poruchy řeči i celého onemocnění. Určení příznaků by probíhalo automaticky, bez velké zátěže na klinický personál.

# Kapitola 2

## Metodika

Organizace této kapitoly je následující: Nejdříve je popis využité databáze promluv, 2.1, následuje část s návrhem vlastního algoritmu, 2.2, poté jsou uvedeny porovnávací metody, 2.3, dále způsoby testování metod, 2.4 a nakonec využité nástroje statistického zhodnocení, 2.5.

### 2.1 Databáze promluv

#### 2.1.1 Účastníci studie

V průběhu let 2015 - 2019 byla v češtině pořízena data od 100 mluvčích. Studie byla schválena etickou komisí Všeobecné fakultní nemocnice v Praze. Databáze obsahuje záznamy monologů od 25ti zdravých lidí (HC), 25ti pacientů s nově diagnostikovanou PN, 25ti s RBD a 20ti s MSA. U MSA bylo původně také 25 promluv, 5 však bylo vyřazeno z důvodu veliké nesrozumitelnosti a nemožnosti stanovit referenční hodnoty cílových veličin.

U PN a RBD bylo provedeno zhodnocení severity onemocnění podle UPDRS (Unified Parkinson's Disease Rating Scale) škály [53]. Zhodnocení motorických příznaků bylo určeno podle MDS-UPDRS III v rozsahu hodnot 0 (žádné motorické příznaky) až 132 (zásadní poruchy pohybu). MDS-UPDRS III *speech item* škála byla použita na ohodnocení poruch řeči v rozsahu 0 (normální řeč) až 4 (nesrozumitelná řeč). U žádného PN ani RBD pacienta dosud nebyla nasazena antiparkinsonská léčba.

U MSA pacientů bylo provedeno zhodnocení podle škály NNIPPS (Neuroprotection and Natural History in Parkinson Plus Syndromes) [54] pro motorické příznaky. Pro řečové příznaky bylo opět použito ohodnocení MDS-UPDRS III speech item.

Délka trvání symptomů byla stanovena od první návštěvy klinického pracoviště po úvodních motorických příznacích. Skupinové klinické detaily jsou uvedeny v tabulce 2.1.

<b>HC skupina</b>	
Průměrný věk [roky]	62.0 ( $\sigma$ 8.5, min 40, max 72)
Genderový poměr	10 žen, 15 mužů
<b>PN skupina</b>	
Průměrný věk [roky]	60.8 ( $\sigma$ 11.5, min 41, max 75)
Genderový poměr	10 žen, 15 mužů
Průměrná doba trvání symptomů [roky]	3.0 ( $\sigma$ 2.4, min 1, max 11.3)
průměrné skóre MDS-UPDRS III	31.1 ( $\sigma$ 12.0, min 10, max 56)
průměrné skóre MDS-UPDRS III speech item	0.8 ( $\sigma$ 0.4, min 0, max 1)
<b>RBD skupina</b>	
Průměrný věk [roky]	61.8 ( $\sigma$ 8.3, min 40, max 73)
Genderový poměr	9 žen, 16 mužů
Průměrná doba trvání symptomů [roky]	5.4 ( $\sigma$ 4.2, min 1, max 20)
průměrné skóre MDS-UPDRS III	7.6 ( $\sigma$ 5.8, min 0, max 24)
průměrné skóre MDS-UPDRS III speech item	0 ( $\sigma$ 0.2, min 0, max 1)
<b>MSA skupina</b>	
Průměrný věk [roky]	61.2 ( $\sigma$ 7.4, min 45, max 72)
Genderový poměr	10 žen, 10 mužů
Průměrná doba trvání symptomů [roky]	4.0 ( $\sigma$ 1.8, min 2, max 7)
Průměrné skóre NNIPPS	76.9 ( $\sigma$ 33.9, min 35, max 123)
průměrné skóre MDS-UPDRS III speech item	1.5 ( $\sigma$ 0.7, min 0, max 3)

**Tabulka 2.1:** Základní klinické charakteristiky skupin pacientů.  $\sigma$  značí standardní směrodatnou odchylku.

### 2.1.2 Řečové nahrávky

Nahrávky byly pořízeny v uzavřené místnosti s velice nízkou hladinou šumu na pozadí (< 50 dB). Byl použit vysoce kvalitní kondenzátorový mikrofon, vzorkovací frekvence nastavena na 48 kHz a bitová hloubka 16 bitů.

Každý účastník byl nahráván v rámci jedné návštěvy kliniky za přítomnosti řečového specialisty provádějícího standardním protokolem. Účastníci byli požádáni o improvizovaný monolog o délce zhruba 90 sekund na libovolné téma.

Nahrávky nebyly dále upravovány, kromě vystřížení případných instrukcí od přítomného specialisty a velice dlouhých pauz (například když pacient přemýšlel, co ještě řekne). U MSA pak byly z monologu vypuštěny úseky s naprosto nesrozumitelným kontextem, stěžující stanovení referenčních hodnot, viz pozdější sekce 2.4.2.

## 2.2 Návrh vlastního algoritmu

### Požadavky algoritmu

Algoritmus musí být celkově robustní. Vzhledem k formě řečové promluvy, improvizovaného monologu, je nutné pracovat se širokou škálou časových délek, od jednotek a desítek sekund po jednotky minut. Algoritmus musí být připraven na dysartrickou řeč, kde se mohou objevovat závažnější řečové poruchy a nízká srozumitelnost. Stejně tak může být vstupní signál místy velice hlasitý a jindy tišší. Toto se může měnit v rámci samotné promluvy i jiného mluvčího.

Metoda by měla být ideálně *jazykově nezávislá*. Cíl výzkumného směru není zhodnotit automatickou analýzu pouze a jen pro češtinu, z hlediska omezeného dopadu.

Nedostatek velkého počtu řečových dat od dysartrických pacientů naznačuje směr k úlohám klasického zpracování signálu, jelikož velikost využitých databází není dostatečná na natrénování případných neuronových sítí a podobných přístupů.

Vzhledem k tomu, že předchozí požadavky již představují poměrně problematickou úlohu, nebude dále v této práci hodnocena šumová odolnost. Tento aspekt bude nutné provést v pozdějších studiích.

Po prvotních zkušebních testech a také z respektu k výše zmíněným požadavkům bylo zcela upuštěno od využití výkonových charakteristik signálu v jednotlivých částech algoritmu. Finální přístup je založen na spektrálních změnách a použití estimátoru základní hlasivkové frekvence.

### 2.2.1 Detektor spektrálních změn

Motivace využití podstatných spektrálních změn je založena na faktu, že jednotlivé fonémy mají rozdílné spektrální rozložení [45]. Jádro slabiky může

vykazovat velice odlišné charakteristiky než její okraje, čehož by se velmi dobře dalo využít na její detekci. Velkou výhodou tohoto přístupu je vyhnutí se závislosti na charakteristikách založených na výkonu, které jsou podstatně více ovlivnitelné různými podmínkami nahrávky (např. střídání nízké a vysoké hlasitosti). Ačkoliv mají jednotlivé jazyky odlišné spektrální vlastnosti [47], rozdíly mezi jádrem slabiky a jejich okraji jsou zachovány.

Ukazuje se, že hledanou spektrální změnu dobře podchytí Bayesovské detektory [45, 46]. Zde konkrétně využijeme Bayesovský autoregresní detektor změn, podle [46].

### ■ Bayesovský autoregresní detektor změn (BACD)

Detekce změn je zde založena na autoregresním (AR) modelování. Aktuální vzorek signálu je predikován z lineární kombinace vzorků předchozích a aditivního bílého šumu s nulovou střední hodnotou.

Za předpokladu, že se vstupní signál  $x[n]$  skládá ze dvou částí, rozdělených pozicí změny  $m$ , může být každá část popsána příslušným AR modelem. Dohromady pak udávající velikost dané změny v  $m$ . Uvažujme levý AR model s řádu  $M_1$  a koeficienty  $a_k$  a pravý AR model řádu  $M_2$  a koeficienty  $b_k$ . Potom je možné signál  $x[n]$  vyjádřit jako

$$x[n] = \begin{cases} \sum_{k=1}^{M_1} a_k x[n-k] + e[n], & n \leq m \\ \sum_{k=1}^{M_2} b_k x[n-k] + e[n], & n > m \end{cases} \quad n = 1, \dots, N, \quad (2.1)$$

kde  $e[n]$  je bílý šum s nulovou střední hodnotou a  $N$  je délka signálu. Výraz (2.1) se dá přepsat do maticové podoby jako

$$\mathbf{x} = \mathbf{G}_A \mathbf{b}_A + \mathbf{e}, \quad (2.2)$$

kde  $\mathbf{b}_A$  představuje matici koeficientů a  $\mathbf{e}$  šumu.  $\mathbf{G}_A$  obsahuje data a její tvar závisí na neznámé pozici změny  $m$  a má rozměry  $N \times M$  pro  $M = M_1 + M_2$ .

Za předpokladu, že všechny parametry AR modelu jsou stejně pravděpodobné a za použití Bayesovské marginalizace, je možno zjistit aposteriorní pravděpodobnost přítomnosti signálové změny  $m$  jako [48]

$$p(\{m\}|\mathbf{x}, M) \approx \frac{\left( \mathbf{x}^T \mathbf{x} - \mathbf{x}^T \mathbf{G}_A (\mathbf{G}_A^T \mathbf{G}_A)^{-1} \mathbf{G}_A^T \mathbf{x} \right)^{\frac{-(N-M)}{2}}}{\sqrt{\det(\mathbf{G}_A^T \mathbf{G}_A)}}, \quad (2.3)$$

kde  $(\mathbf{G}_A^T \mathbf{G}_A)^{-1} = \Phi_A$  je inverzní autokorelační matice signálu,  $\mathbf{x}^T \mathbf{x} = D$  je energie signálu a  $\mathbf{x}^T \mathbf{G}_A = \mathbf{g}_A$  je tzv. autokorelační vektor. Při zavedení  $\Delta_A = \det(\mathbf{G}_A^T \mathbf{G}_A)$  se výraz (2.3) zjednoduší na

$$p(\{m\}|\mathbf{x}, M) \approx \frac{\left( D - \mathbf{g}_A \Phi_A \mathbf{g}_A^T \right)^{\frac{-(N-M)}{2}}}{\sqrt{\Delta_A}}. \quad (2.4)$$

Nevýhoda tohoto přístupu spočívá ve faktu, že nejde použít pokud signál obsahuje více než jednu změnu, což představuje velmi restriktivní limitaci. Tento problém je vyřešen počítáním  $p(\{m\}|\mathbf{x}, M)$  v krátkých oknech, posouváných po délce signálu.

Aby bylo možné porovnávat jednotlivé aposteriorní pravděpodobnosti v jednotlivých oknech, je nutné provést její normalizaci. To je provedeno pomocí tzv. *Bayesovské evidence*, využívající matici  $\mathbf{G}_E$ , která obsahuje prvky celého signálu (v daném okně), nerozlišující na pravou a levou část, na rozdíl od  $\mathbf{G}_A$ . Normalizace je pak provedena jako

$$p(\{m\}|\mathbf{x}, M) \approx \frac{\left(D - \mathbf{g}_A \Phi_A \mathbf{g}_A^T\right)^{\frac{-(N-M)}{2}}}{\sqrt{\Delta_A}} \frac{\sqrt{\Delta_E}}{\left(D - \mathbf{g}_E \Phi_E \mathbf{g}_E^T\right)^{\frac{-(N-M_1)}{2}}}, \quad (2.5)$$

kde jsou příslušné  $\Phi_E$ ,  $\mathbf{g}_E$  a  $\Delta_E$  přepočítány pro  $\mathbf{G}_E$ .

Využitím logaritmu aplikovaného na výraz (2.5) se dá dosáhnout větší numerické stability a dalšího zjednodušení. Zanedbán je také vliv obou členů  $\Delta_A$ ,  $\Delta_E$  z důvodu jejich malého dopadu na výslednou pravděpodobnost. V posledním kroku jsou také zanedbány multiplikatивní konstanty, které vzniknou aplikací logaritmu. Pomocí těchto změn získáme z (2.5) výraz

$$p(\{m\}|\mathbf{x}, M) \approx \log_{10} \left(D - \mathbf{g}_E \Phi_E \mathbf{g}_E^T\right) - \log_{10} \left(D - \mathbf{g}_A \Phi_A \mathbf{g}_A^T\right). \quad (2.6)$$

Nejzásadnější změny v signálu jsou reprezentovány maximální hodnotou aposteriorní pravděpodobnosti  $p(\{m\}|\mathbf{x}, M)$  v místě dané změny. Výsledek záleží čistě jen na vstupním signálu.

Z důvodu vysoké numerické náročnosti je celý algoritmus počítán rekurzivně.  $p(\{m\}|\mathbf{x}, M)$  je počítána vždy pro změnu  $m$ , nacházející se v polovině aktuálního okna, které je posouváno po jednotlivých vzorcích. Popis celého rekurzivního algoritmu je v tabulce 2.2.

---

**BACD** (Bayesian autoregressive changepoint detector) rekurzivní algoritmus

---

**Vstup:**  $\mathbf{x}$  (vstupní signál),  $M_1, M_2$  (řád AR modelů),  $N$  (délka okna).

**Výstup:** Hustota aposteriorní pravděpodobnosti změn  $p(\{m\}|\mathbf{x}, M)$  v dané pozici  $m$ .

---

### I. Inicializace

Sestavení inicializačních matic  $\mathbf{G}_A$  (s respektem místa změny  $m = N/2$ ) a Bayesovské evidence  $\mathbf{G}_E$  určené k normalizaci. Z nich pak následně spočtené počáteční tvary  $D = \mathbf{x}^T \mathbf{x}$ ,  $\Phi_A = (\mathbf{G}_A^T \mathbf{G}_A)^{-1}$ ,  $\Phi_E = (\mathbf{G}_E^T \mathbf{G}_E)^{-1}$ ,  $\mathbf{g}_A = \mathbf{x}^T \mathbf{G}_A$  a  $\mathbf{g}_E = \mathbf{x}^T \mathbf{G}_E$ . Inicializační matice  $\mathbf{G}_A, \mathbf{G}_E$  mají tvar

$$\mathbf{G}_A = \begin{pmatrix} x[0] & x[-1] & \cdots & 0 & 0 & \cdots \\ x[1] & x[0] & \cdots & 0 & 0 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \\ x[N/2-1] & x[N/2-2] & \cdots & 0 & 0 & \cdots \\ 0 & 0 & \cdots & x[N/2] & x[N/2-1] & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \\ 0 & 0 & \cdots & x[N-1] & x[N-2] & \cdots \end{pmatrix},$$

$$\mathbf{G}_E = \begin{pmatrix} x[0] & x[-1] & \cdots \\ x[1] & x[0] & \cdots \\ \vdots & \vdots & \cdots \\ x[N/2-1] & x[N/2-2] & \cdots \\ x[N/2] & x[N/2-1] & \cdots \\ \vdots & \vdots & \cdots \\ x[N-1] & x[N-2] & \cdots \end{pmatrix}.$$

---

### II. Aktualizace dat

Pozorován nový vzorek  $x_{\text{new}}$  na konci aktuálního okna.

**II.A** Jsou vytvořeny prvky, zahrnující jeho vliv:

$$\mathbf{G}_{A_{\text{new}}} = \left( \overbrace{0 \quad 0 \quad \cdots}^{M_1} \quad \overbrace{x_{\text{new}} \quad x_{\text{new}-1} \quad \cdots}^{M_2} \right),$$

$$D_{\text{new}} = D + x_{\text{new}}^2,$$

$$\mathbf{w}_1 = \Phi_A \mathbf{G}_{A_{\text{new}}}^T.$$

**II.B** Aktualizace klíčových matic těmito prvky

$$\mathbf{g}_{A_{\text{new}}} = \mathbf{g}_A + x_{\text{new}} \mathbf{G}_{A_{\text{new}}},$$

$$\lambda = 1 + \mathbf{G}_{A_{\text{new}}} \mathbf{w}_1,$$

$$\Phi_{A_{\text{new}}} = \Phi_A - \mathbf{w}_1 \lambda^{-1} \mathbf{w}_1^T.$$

**II.C** Odstranění starého vzorku na začátku okna. Prvky zahrnující jeho vyřazení jsou



$$\mathbf{Z} = \left( \overbrace{x_{\text{new}-N-1} \quad x_{\text{new}-N-2} \quad \cdots}^{M_1} \quad \overbrace{0 \quad 0 \quad \cdots}^{M_2} \right),$$

$$D_{\text{new2}} = D_{\text{new}} + x_{\text{new}-N}^2,$$

$$\mathbf{w}_2 = \Phi_{\text{A}_{\text{new}}} \mathbf{Z}^T.$$

**II.D** Aktualizace klíčových matic těmito prvky

$$\mathbf{g}_{\text{A}_{\text{new2}}} = \mathbf{g}_{\text{A}_{\text{new}}} - x_{\text{new}-N} \mathbf{Z},$$

$$\lambda = 1 - \mathbf{Z} \mathbf{w}_2,$$

$$\Phi_{\text{A}_{\text{new2}}} = \Phi_{\text{A}_{\text{new}}} - \mathbf{w}_2 \lambda^{-1} \mathbf{w}_2^T.$$


---

### III. Aktualizace pozice

Je zapotřebí aktualizovat pozici uprostřed okna (očekávána změna) pro nový vzorek. To je provedeno ve dvou fázích. Nejprve je  $m$ -tý řádek matice  $\mathbf{G}_A$  nahrazen nulami.

**III.A** Prvky nahrazující  $m$ -tý řádek nulami

$$\mathbf{R} = \left( \overbrace{0 \quad 0 \quad \cdots}^{M_1} \quad \overbrace{x_{\text{new}-N/2} \quad x_{\text{new}-N/2-1} \quad \cdots} \right),$$

$$\mathbf{w}_3 = \Phi_{\text{A}_{\text{new2}}} \mathbf{R}^T.$$

**III.B** Aktualizace klíčových matic těmito prvky

$$\mathbf{g}_{\text{A}_{\text{new3}}} = \mathbf{g}_{\text{A}_{\text{new2}}} - x_{\text{new}-N/2} \mathbf{R},$$

$$\lambda = 1 - \mathbf{R} \mathbf{w}_3,$$

$$\Phi_{\text{A}_{\text{new3}}} = \Phi_{\text{A}_{\text{new2}}} - \mathbf{w}_3 \lambda^{-1} \mathbf{w}_3^T.$$

**III.C** V druhé fázi je nulový řádek nahrazen aktualizovanou pozicí středu okna. Prvky popisující tuto operaci jsou

$$\mathbf{Q} = \left( \overbrace{x_{\text{new}-N/2} \quad x_{\text{new}-N/2-1} \quad \cdots}^{M_1} \quad \overbrace{0 \quad 0 \quad \cdots}^{M_2} \right),$$

$$\mathbf{w}_4 = \Phi_{\text{A}_{\text{new3}}} \mathbf{Q}^T.$$

**III.D** Nahrazení aktualizovanou pozicí středu okna u klíčových matic

$$\mathbf{g}_{\text{A}_{\text{new4}}} = \mathbf{g}_{\text{A}_{\text{new3}}} - x_{\text{new}-N/2} \mathbf{Q},$$

$$\lambda = 1 - \mathbf{R} \mathbf{w}_4,$$

$$\Phi_{\text{A}_{\text{new4}}} = \Phi_{\text{A}_{\text{new3}}} - \mathbf{w}_4 \lambda^{-1} \mathbf{w}_4^T.$$


---

### IV. Aktualizace Bayesovské evidence

V tomto kroku musí být do Bayesovské evidence  $\mathbf{G}_E$  zaevídován nový vzorek a zahozen starý.

**IV.A** Prvky na přidání nového vzorku

$$\mathbf{G}_{\text{E}_{\text{new}}} = \left( \overbrace{x_{\text{new}} \quad x_{\text{new}-1} \quad \cdots}^{(M_1+M_2)/2} \right),$$

$$\mathbf{w}_E = \Phi_E \mathbf{G}_{\text{E}_{\text{new}}}^T.$$

**IV.B** Aktualizace klíčových matic

$$\mathbf{g}_{\text{E}_{\text{new}}} = \mathbf{g}_E + x_{\text{new}} \mathbf{G}_{\text{E}_{\text{new}}},$$

$$\lambda = 1 - \mathbf{G}_{\text{E}_{\text{new}}} \mathbf{w}_E,$$

$$\Phi_{\text{E}_{\text{new}}} = \Phi_E - \mathbf{w}_E \lambda^{-1} \mathbf{w}_E^T.$$

**IV.C** Prvky na zahození starého vzorku na konci okna

$$\mathbf{Z}_E = \overbrace{(x_{\text{new}-N} \quad x_{\text{new}-N-1} \quad \cdots)}^{(M_1+M_2)/2},$$

$$\mathbf{w}_{E2} = \Phi_{E_{\text{new}}} \mathbf{Z}_E^T.$$

**IV.D** Aktualizace klíčových matic

$$\mathbf{g}_{E_{\text{new}2}} = \mathbf{g}_{E_{\text{new}}} + x_{\text{new}} \mathbf{Z}_E,$$

$$\lambda = 1 - \mathbf{Z}_E \mathbf{w}_{E2},$$

$$\Phi_{E_{\text{new}2}} = \Phi_{E_{\text{new}}} + \mathbf{w}_{E2} \lambda^{-1} \mathbf{w}_{E2}^T.$$

## V. Spočítání aposteriorní pravděpodobnosti

Změna  $m$  je očekávána ve středu okna a  $p(\{m\}|\mathbf{x}, M)$  je určena podle (2.6) s ohledem na aktualizované klíčové matice jako

$$p(\{m\}|\mathbf{x}, M) \approx \log_{10} \left( D_{\text{new}2} - \mathbf{g}_{E_{\text{new}2}} \Phi_{E_{\text{new}2}} \mathbf{g}_{E_{\text{new}2}}^T \right) - \log_{10} \left( D_{\text{new}2} - \mathbf{g}_{A_{\text{new}4}} \Phi_{A_{\text{new}4}} \mathbf{g}_{A_{\text{new}4}}^T \right).$$

## VI.

Vrácení na bod II. pro zakomponování nového vzorku a aktualizovat pozici změny  $m = m + 1$ .

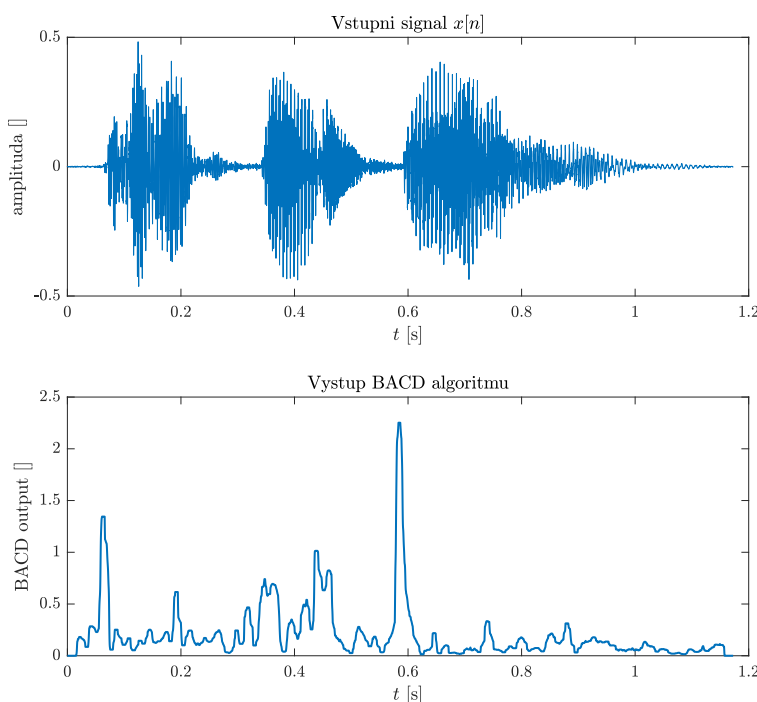
**Tabulka 2.2:** Popis Bayesovského autoregresního rekurzivního algoritmu spektrálních změn (BACD). V kroku I. dochází k inicializaci všech potřebných prvků algoritmu podle vstupních dat. Pokud časový index u signálu  $x$  ukazuje na neexistující hodnotu (signál ještě nezačal), je tento prvek nahrazen nulou. V kroku II. je zahrnut nově pozorovaný prvek na samém konci okna a zároveň vyřazen starý prvek na jeho začátku. Krok III. aktualizuje pozici  $m$  očekávané změny, která je přesně ve středu posouvaného okna. V kroku IV. je aktualizována Bayesovská evidence s ohledem na aktuální podobu okna, určená k následné normalizaci. V kroku V. je spočtena hodnota aposteriorní pravděpodobnosti změny v místě  $m$  pomocí aktualizovaných matic. Následně je v kroku VI. okno posunuto o jeden prvek signálu dál a algoritmus se vrací do bodu II.

Z důvodů vysoké detailnosti a nevyhlazenosti výstupu, je na doporučení autorů vstupní signál podvzorkován na frekvenci  $f_{s\_re} = 16$  kHz. Samotný výstup je pak filtrován dolní propustí s mezní frekvencí  $f_0 = 20$  Hz.

Ilustrační průběh filtrovaného výstupu BACD pro podvzorkovaný vstupní signál, se nachází na obrázku 2.1.

## ■ Vyhlazení BACD průběhu

Na obrázku 2.1 je patrné, že ani po doporučených úpravách není charakteristika dostatečně vyhlazená na účely následné analýzy. Proto je provedena ještě



**Obrázek 2.1:** Ilustrační obrázek průběhu výstupu BACD algoritmu pro promluvu „Dej to sem“.

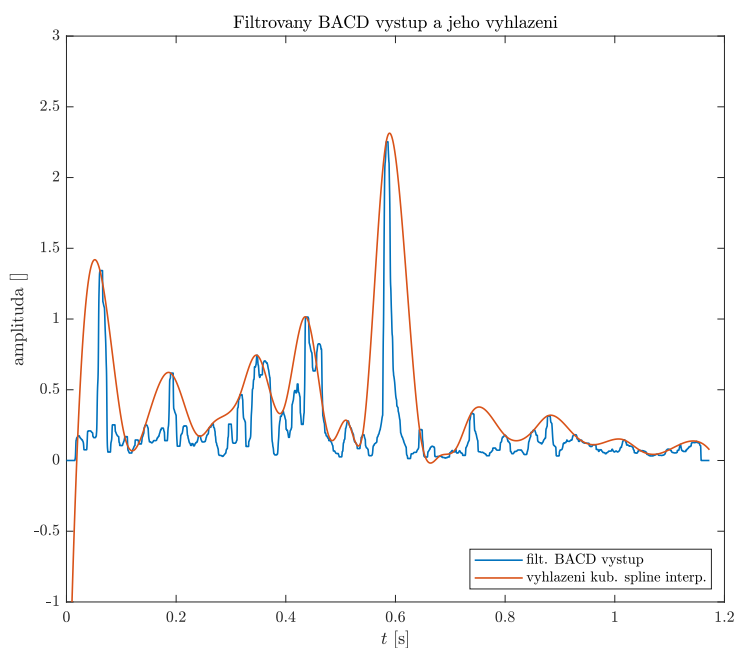
jedna operace před dalším zpracováním. Jsou zjištěny pozice a hodnoty všech lokálních maxim a ty které od sebe dělí alespoň  $N_p$  vzorků jsou následně interpolovány pomocí kubické *spline* metody [49].

Ilustrační obrázek vyhlazení charakteristiky je zobrazen na 2.2. Tato charakteristika bude předmětem následné analýzy a bude nadále označena jako  $BACD[n]$ .

### 2.2.2 Klasifikace intervalů

Výše popsaná charakteristika  $BACD[n]$  popisuje signifikantní spektrální změny v řečovém signálu. Jednotlivá slabika je pak v ideálním případě určena úsekem mezi dvěma zásadními změnami, z důvodu odlišných vlastností svého jádra a okrajů.

V praxi však detekovaný interval nemusí značit přítomnost slabiky. Spektrální změny mohou vyvolat například některé nerečové artefakty, poměrně často se vyskytující úkaz v dysartrické řeči. Dále rovněž dvě rychle po sobě jdoucí slabiky nemusí vykazovat zásadní spektrální změnu mezi sebou, mohou být vyslovené jen z části a neúplně. Tento efekt se nazývá *smearing*, rozmazání rozdílů.



**Obrázek 2.2:** Ilustrační obrázek vyhlazení filtrovaného BACD výstupu pomocí kubické spline interpolace.

Z těchto důvodů je zapotřebí analýza každého určeného intervalu, klasifikující zda se jedná o slabiku či nikoliv. Při kladné odpovědi zároveň posuzující zda daný úsek obsahuje právě jednu, či více slabik.

Byly testovány různé vlastnosti řečového signálu, které by se daly vhodně použít. Výkonové charakteristiky a energie nízkých frekvencí fungovaly pro případy některých neřečových artefaktů, jelikož typicky mají výkon poměrně nižší než mluvené části. Ve výsledku se tento přístup ale ukázal jako nedostatečně robustní a byla zvolena analýza podle základní hlasivkové frekvence  $F_0$ . Na její odhad byla použita metoda SWIPE [51], která vykazuje přesné výsledky pro podmínky naší úlohy [50].

Výhoda využití průběhu  $F_0$  spočívá ve vlastnostech odhadu zbavit se vlivu šumu na pozadí a občasného výrazného vlivu souhlásek, kdy jsou zaměněny za jádro slabiky [14]. Také se dá použít na podchycení *smearing* efektu [14].

SWIPE disponuje účinnou sadou parametrů, z nichž nejdůležitější je tzv. *pitch strength*,  $T_F\_strength$ , které klasifikují a prahují odhadované hodnoty  $F_0$  [51]. Princip je podobný jako u algoritmů detekce znělých a neznělých úseků (VAD), které jsou některými metodami detekce slabik také využívány.

## ■ SWIPE, metoda odhadu základní hlasivkové frekvence

Sawtooth Waveform inspired Pitch Estimator (SWIPE) [51] představuje metodu fungující ve spektrální oblasti, využívající principu maximální shody upraveného spektra signálu s uměle vytvořenou harmonickou funkcí.

Ze vstupního signálu  $s[n]$  je po segmentech vypočítáno jeho spektrum  $X[f, m]$  podle krátkodobé Fourierovy transformace (STFT), kde  $f$  je daná frekvence a  $m$  index okna. Je zvoleno Hannovo váhovací okno z důvodu jeho jednoduchosti a vhodných spektrálních vlastností.

Ze spektra  $X[f, m]$  je spočten jeho modul,  $|X[f, m]|$ , který je následně *odmocněn* na druhou, jako náhrada aplikace logaritmu pro vnitřní účely estimátoru.

Takto upravené spektrum následně prochází procesem borcení lineární frekvenční osy podle tzv. *ERB* škály (equivalent rectangular bandwidth scale), definované jako zobrazení  $ERBs(f): f \mapsto \varepsilon$

$$ERBs(f) = 21.4 \log_{10} \left( 1 + \frac{f}{229} \right) \quad [\varepsilon], \quad (2.7)$$

kde  $f$  odpovídá lineárnímu frekvenčnímu měřítku (Hz) a  $\varepsilon$  je odpovídající nelineární frekvence v jednotkách ERBs.

Výsledné upravené spektrum  $\sqrt{|X[\varepsilon, m]|}$  bude porovnáno s uměle vytvořenou harmonickou funkcí  $K(\varepsilon, \check{\varepsilon}_0)$  s danou základní frekvencí  $\check{\varepsilon}_0$ . Z důvodu případného problému odhadu subharmonických frekvencí, je funkce  $K(\varepsilon, \check{\varepsilon}_0)$  násobena klesajícími vahami v hodnotách  $\varepsilon^{-1/2}$ , postupně potlačující jejich vliv.

Algoritmus určuje odhad základní frekvence pro daný interval jako takovou hodnotu  $\check{\varepsilon}_0$ , která maximalizuje vzájemnou shodu mezi spektrem a umělou harmonickou funkcí v podobě normalizovaného skalárního součinu, tedy

$$\hat{\varepsilon}_0[m] = \arg \max_{\check{\varepsilon}_0} \langle \sqrt{|X[\varepsilon, m]|}, \varepsilon^{-1/2} K(\varepsilon, \check{\varepsilon}_0) \rangle. \quad (2.8)$$

Výsledný odhad  $\hat{\varepsilon}_0[m]$  je pak následně převeden zpět do lineární škály v Hertzích pomocí zpětné transformace. Výsledný vektor odhadů  $F_0[m]$  resp.  $F_0[n]$  při zopakování hodnot pro dané okno představuje průběh základní frekvence řečového signálu.

Zároveň je v každém okně určen tzv. *pitch strenght* atribut daného odhadu, charakterizující míru podobnosti spektra vstupního signálu s ideálním čistým tónem. Tento parametr pak funguje jako účinná prahová hodnota pro následné hodnocení průběhu základní hlasivkové frekvence v promluvě za daných podmínek.

Místa, která jsou pak klasifikována jako úseky bez znělosti, kde se základní frekvence neurčuje, jsou v kontuře  $F_0[n]$  označena příslušným symbolem, který značí absenci znělosti.

## ■ Detekce pozice intervalů

Jednotlivý interval značí možnou přítomnost řečené slabiky. Začátek je dán signifikantní spektrální změnou v řečovém signále, představující vrchol ve vyhlazené charakteristice  $BACD[n]$ . Daný interval pak trvá až do dalšího vrcholu, značícího novou změnu. Jako vrchol, ohraničující hranice intervalů, je prohlášen ten, který je vyšší než předchozí údolí minimálně o  $T_{p\_height}$ . Zároveň vrcholy nesmí být u sebe blíže než  $T_{p\_dist}$  vzorků. K detekovaným lokálním maximům je nakonec přiřazen i konec aktuálního analyzovaného okna. To se při testování ukázalo jako užitečné, jelikož algoritmus měl problémy s detekcí intervalů právě na samém konci okna.

Pokud je počet všech nalezených vrcholů  $N_{peaks}$ , potom je počet intervalů v daném okně roven  $N_{int} = N_{peaks} - 1$ .

## ■ Klasifikace intervalu podle $F_0$

Jakmile je zjištěno umístění daného intervalu, zbývá určit, zdali je tvořen slabikou, vícero slabikami, nebo nechtěným artefaktem. Jak bylo zmíněno, toto rozhodování provedeme pomocí průběhu  $F_0[n]$ .

Tak jako je použito u mnoha dalších metod [31, 14, 33, 34], jestliže interval neobsahuje žádný výskyt  $F_0[n]$  ve svém průběhu, zřejmě se nejedná o slabiku. Zároveň, pokud součástí intervalu je více souvislých průběhů  $F_0[n]$  než jeden, nejspíše je slabik vícero. Jejich počet v intervalu je tak po tomto prvním kroku analýzy stanoven na *počet souvislých segmentů*  $F_0[n]$ , delších než  $T_{F\_len}$ . Rozhodnutí prahovat délku segmentu byla zvolena z důvodu možných chyb SWIPE v kontuře  $F_0[n]$  a možného krátkého přesahu průběhu z vedlejších intervalů.

Při rychlém sledu více slabik za sebou však může nastat scénář výše zmíněného *smearing* efektu. V takových případech je v intervalu několik slabik, avšak průběh  $F_0[n]$  může být souvislý po celou dobu jejich trvání. Z tohoto důvodu je provedeno hledání signifikantních lokálních extrémů v kontuře  $F_0[n]$  uvnitř intervalu. Nalezený extrém, kdy mluvčí pozměnil způsob tvorby svého hlasu, může značit hranici dvou slabik, kterou  $BACD[n]$  nepostihl.

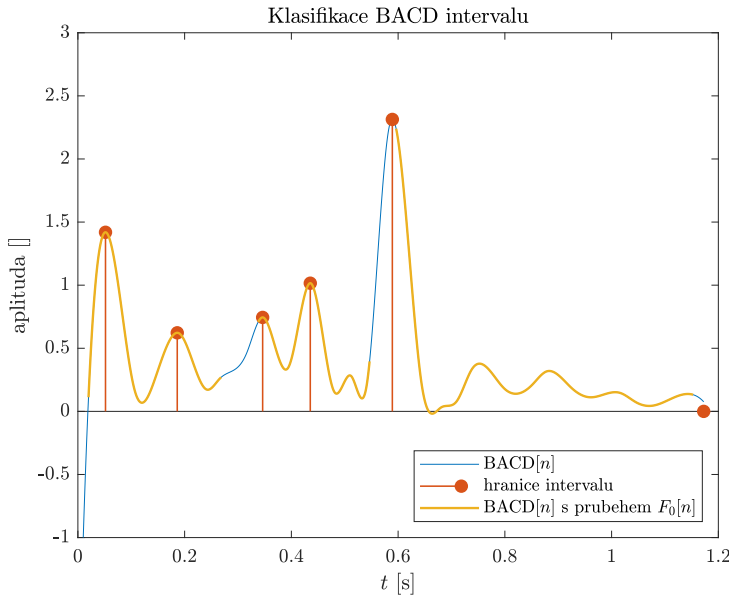
V tomto druhém kroku je tedy k již existujícímu počtu slabik přičten počet všech lokálních extrémů  $F_0[n]$  uvnitř intervalu, které se od předchozího údolí/vrcholu liší aspoň o  $T_{F\_diff}$  Hz a jsou vzdálené od nejbližší hranice intervalu aspoň  $T_{F\_dist}$  vzorků. Prahování vzdálenosti od hranice je použito proto, aby se nedetekovala slabika v místě, kde už jí obsáhl  $BACD[n]$ .

V posledním kroku jsou analyzovány pozice jednotlivých detekovaných slabik. Pokud jsou některé dvě u sebe blíže než  $T_{s\_dist}$  vzorků, je druhá z nich považována za fyziologicky nemožnou a je zahozena.

Finální počet slabik  $N_{\text{syll}}$  v daném intervalu je tedy roven

$$N_{\text{syll}} = N_{\text{syll1}} + N_{\text{syll2}} - N_{\text{syll3}}, \quad (2.9)$$

kde  $N_{\text{syll1}}$  je počet slabik ze souvislých segmentů  $F_0[n]$  v prvním kroku,  $N_{\text{syll2}}$  je počet při zhodnocení lokálních extrémů  $F_0[n]$  v druhém kroku a  $N_{\text{syll3}}$  je množství slabik zahozených při kontrole vzájemné vzdálenosti ve třetím kroku analýzy intervalu. Ilustrační obrázek hranic intervalů a segmentů obsahující  $F_0[n]$  je na 2.3.



**Obrázek 2.3:** Ilustrační obrázek podoby intervalů na charakteristice BACD[n] spolu se segmenty obsahující  $F_0[n]$ .

### 2.2.3 Celkový popis algoritmu

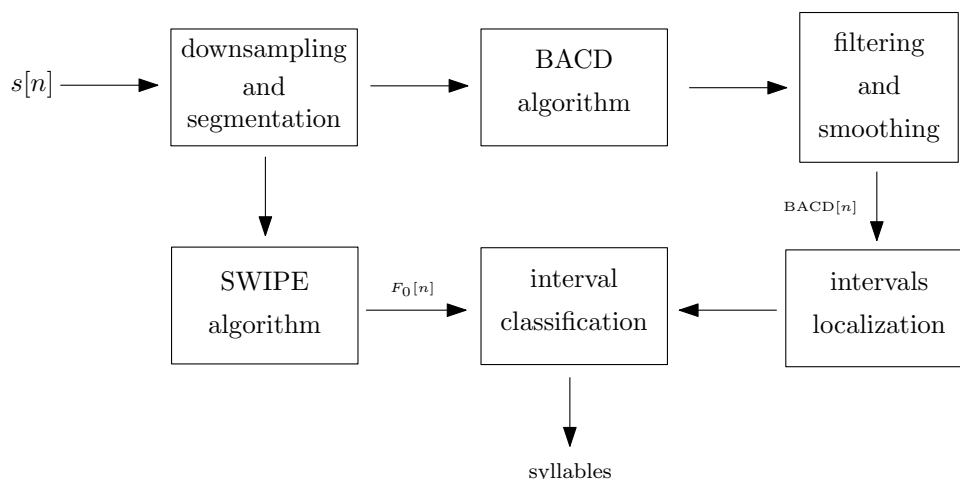
Celý algoritmus lze tedy shrnout následovně. Vstupní řečový signál  $s[n]$  je podvzorkován s frekvencí  $f_{s\_re}$  a zpracováván po oknech délky  $N_{\text{win}}$ , se vzájemných překryvem  $N_{\text{overlap}}$ . Data v každém segmentu slouží jako vstup do BACD algoritmu, spolu s řády AR modelů  $M_1$ ,  $M_2$  a délkou okna  $N_{\text{BACD\_win}}$ . Výstup algoritmu je nejprve filtrován dolní propustí s mezní frekvencí  $f_0$  a následně vyhlazen pomocí kubické spline interpolace při využití parametru  $N_p$ . Výsledná charakteristika, BACD[n] je posléze po intervalech analyzována.

Paralelně je na datech v daném okně odhadována trajektorie  $F_0$  pomocí metody SWIPE, za použití parametrů  $F_{\text{min}}$ ,  $F_{\text{max}}$ ,  $N_{\text{F\_step}}$  a  $T_{\text{F\_strength}}$ . Výsledný průběh je v celém okně je označen  $F_0[n]$ .

Hranice intervalů jsou stanoveny jako pozice lokálních maxim BACD[n], při prahových hodnotách  $T_{p\_height}$  a  $T_{p\_dist}$ . V každém je spočten celkový

počet slabik  $N_{\text{syll}}$  podle (2.9), vzhledem k počtu souvislých úseků a lokálních extrémů  $F_0[n]$  vně intervalu a za použití prahových hodnot  $T_{F\_len}$ ,  $T_{F\_diff}$ ,  $T_{F\_dist}$  a  $T_{s\_dist}$ .

Na konci je spočten celkový počet slabik jako součet  $N_{\text{syll}}$  přes všechny intervaly a okna, spolu s jejich pozicemi. Principiální schéma celého navrženého algoritmu je na obrázku 2.4.



**Obrázek 2.4:** Principiální schéma navrženého algoritmu.

## 2.2.4 Číselné hodnoty parametrů algoritmu

V předchozí části bylo popsáno celkové schéma navrženého algoritmu. Zbývá uvést numerické hodnoty využitých parametrů. Ty byly určeny na základě analýzy několika řečových nahrávek ze skupin zdravých lidí, pacientů s PN, RBD a MSA.

Parametry se váží hlavně k průběhům  $BACD[n]$  a  $F_0[n]$ , které se ukázaly jako poměrně dosti robustní s ohledem na různorodost řečové nahrávky. Celkový seznam všech hodnot spolu s jejich popisem je uveden v tabulce 2.3.

## 2.3 Porovnávací metody

### Výběr metod

Z dostupné literatury bylo vybráno několik metod na otestování v podmínkách našeho scénáře. K hledání bylo použito webů IEEE Xplore a Google Scholar za



Název	Hodnota	Popis
$f_{s\_re}$	16 kHz	Vzorkovací frekvence se kterou je vstupní signál $s[n]$ podvzorkován.
$N_{win}$	$2 \cdot f_{s\_re}$	Délka analyzovaného okna.
$N_{overlap}$	$0.05 \cdot N_{win}$	Délka překryvu jednotlivých oken.
$M_1, M_2$	6	Řády AR modelů BACD algoritmu.
$N_{BACD\_win}$	$0.03 \cdot f_{s\_re}$	Délka posuvného okna BACD algoritmu.
$f_0$	20 Hz	Mezní frekvence dolní propusti k filtrování BACD výstupu.
$N_p$	$0.025 \cdot f_{s\_re}$	Minimální vzdálenost lokálních maxim při spline interpolaci BACD výstupu.
$F_{min}$	60 Hz	Minimální možná odhadovaná $F_0$ .
$F_{max}$	400 Hz	Maximální možná odhadovaná $F_0$ .
$N_{F\_step}$	$0.01 \cdot f_{s\_re}$	Posun analyzovaného okna u SWIPE algoritmu.
$T_{F\_strength}$	0.22	Tzv. <i>pitch strength</i> , parametr SWIPE algoritmu.
$T_{p\_height}$	0.4	Minimální výška vrcholu BACD $[n]$ , vzhledem k předchozímu údolí.
$T_{p\_dist}$	$0.075 \cdot f_{s\_re}$	Minimální vzdálenost vrcholů BACD $[n]$ .
$T_{F\_len}$	$0.025 \cdot f_{s\_re}$	Prahová hodnota délky pro souvislé úseky $F_0[n]$ .
$T_{F\_diff}$	5 Hz	Prahová hodnota pro rozdíl výšky lokálního extrému $F_0[n]$ a předchozího údolí/vrcholu.
$T_{F\_dist}$	$0.125 \cdot f_{s\_re}$	Prahová hodnota vzdálenosti lokálního extrému $F_0[n]$ od hranice intervalu.
$T_{s\_dist}$	$0.090 \cdot f_{s\_re}$	Minimální vzdálenost dvou detekovaných slabik od sebe.

**Tabulka 2.3:** Seznam všech použitých parametrů navrženého algoritmu, spolu s číselnými hodnotami a popisem.

pomocí klíčových slov speech rate, articulation rate, syllable detection, speech rhythm. Z nalezených pak byly vyřazeny metody s neuspokojivými výsledky v porovnání s ostatními, metody explicitně určené na odlišnou aplikaci (např čtení velmi krátkého textu) a také algoritmy jejichž reprodukce by byla složitá a zjevně by se lišila od té navrhované, díky neúplnému a nejasnému popisu.

Celkem byly otestovány 4 estimátory řečové rychlosti: (i) metoda založená na automatickém rozpoznávací fonémů (za použití [37]), (ii) přístup využívající charakteristiky harmonických oscilátorů [38], (iii) detektor pracující s energetickou obálkou signálu modulovanou výkonem nízkých frekvencí [39] a (iv) algoritmus detekce na charakteristice intenzity signálu [31]. Pro přehlednost budou v práci nadále označovány jako *Rozpoznávač*, *theta Oscilátor*, *LFME* (low frequency modulated energy) a *Praat skript*, s respektem k názvům daných od samotných autorů. K těmto doplníme vlastní navrženou metodu

k celkovému porovnání.

### ■ 2.3.1 Rozpoznávač

Využití rozpoznávače představuje jednu z nejpřímějších variant detekce slabik. Ve zkratce je analyzován jeho výstup, tedy automatický přepis daného signálu, a pomocí jednoduchých lingvistických pravidel se z přepisu určí jednotlivé slabiky a jejich pozice. Problémy může způsobovat to, že rozpoznávače často nejsou univerzální pro všechny jazyky a uzpůsobené na jinou než zdravou řeč. Většinou také bývá situace naopak, totiž že estimátor řečové rychlosti je součástí automatických rozpoznávačů a pomáhá jim zpřesňovat výsledky [39].

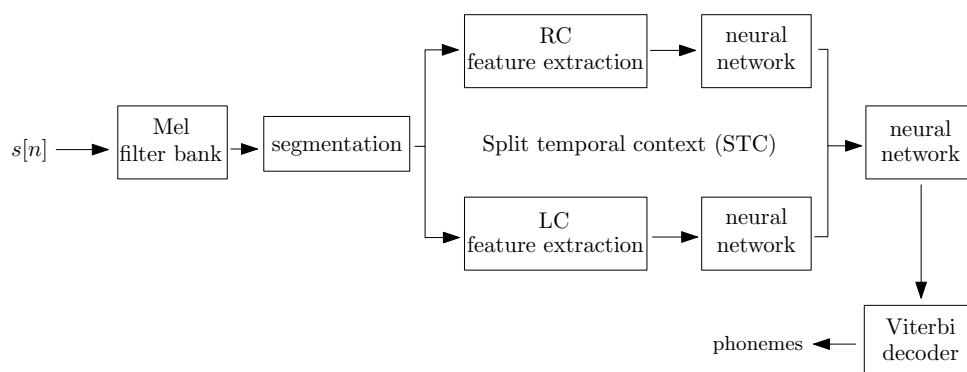
Vysoce sofistikované nástroje jako například Google rozpoznávač [40], natrénované na velkém počtu dat, mohou mít problémy se silnou dysartrickou řečí, jelikož je snaha, aby výstup měl "smysl" (tedy minimálně neobsahoval např. neexistující slova). Což u pacientů s většími poruchami řeči je velmi častý úkaz. Využití těchto metod pro estimaci řečové rychlosti však nebylo dosud podrobně zkoumáno.

V této práci využijeme automatický rozpoznávač fonémů vyvinutý na VUT v Brně [37], do této chvíle schopný pracovat s češtinou, maďarštinou, ruštinou a angličtinou. Funguje na základě systému umělých neuronových sítí (ANN) a Viterbiho dekodéru.

Principiální schéma rozpoznávače je naznačeno na obrázku 2.5. Vstupní řečový signál  $s[n]$  prochází přes banku filtrů na bázi Melovské stupnice a následně je po segmentech spočítána energie. Dále je signál po úsecích zpracováván pomocí principu STC (split temporal context), který rozděluje foném časové oblasti na dvě části, které jsou nezávisle zpracovány. Celý blok je před rozdělením váhován Hammingovým oknem a dochází ke zvýraznění centrální části. Odpovídající průběhy Melovské energie jsou pomocí diskrétní kosinové transformace redukovány na fixní počet koeficientů. Následně je blok rozdělen na levou (LC) a pravou (RC) část. V každé části slouží příslušné koeficienty jako vstup do umělé neuronové sítě. Výsledné posteriorní pravděpodobnosti fonémů z každé sítě jsou pak dohromady zpracovány třetí sítí, jejímž výstupem jsou finální posteriorní pravděpodobnosti. Dekodér, v podobě Viterbiho algoritmu, pak z tohoto výstupu určuje výsledné fonémy.

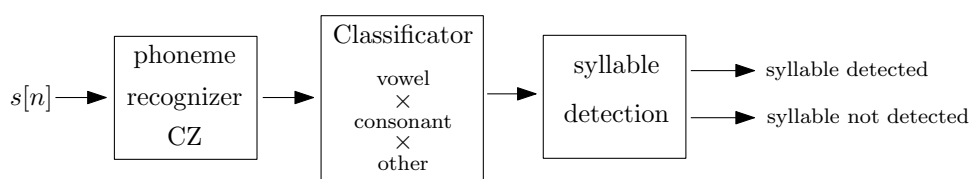
Neuronové sítě byly natrénovány na základě databáze SpeechDat-E obsahující přes 10 hodin konverzací pro každou z řečí čeština, maďarština, ruština a angličtina. Byl využit zpětno-propagační algoritmus s chybovou funkcí v podobě křížové entropie. Výsledný počet skrytých neuronových vrstev byl okolo 500 [41].

Na základě výstupu z výše popsaného rozpoznávače, obsahujícího jed-



**Obrázek 2.5:** Principiální schéma fonémového rozpoznávače.

notlivé fonému spolu s jejich časovými značkami, byl na základě jednoduchých pravidel postaven detektor slabik. Algoritmus postupně prochází fonémy a klasifikuje samohlásky, souhlásky a zbylé zvuky. Pokud nastane jedna ze situací (i) souhláska a samohláska po sobě, (ii) tři souhlásky po sobě, (iii) čtyři souhlásky po sobě a (iv) dvě souhlásky a samohláska, je tento útvar označen za slabiku s danou časovou značkou a index algoritmu se posune na jeho konec. Příklady popsaných situací mohou být např. (i) pa, (ii) krk, (iii) vzkvět a (iv) pro. Principiální schéma celého detektoru slabik postaveného na fonémovém rozpoznávači je znázorněno na obrázku 2.6.



**Obrázek 2.6:** Schéma detektoru slabik na základě fonémového rozpoznávače, natrénovaného na české databázi.

Tento systém budeme nadále používat v této práci na estimaci řečové rychlosti, pod názvem *Rozpoznávač*.

### 2.3.2 Theta Oscilátor

Algoritmus nazvaný *theta Oscilátor* [38], využívá na detekci slabik veličinu zvanou *zvučnost* (sonority) a byl vyvinut pro detekci útvarů připomínající slabiky u dětí které ještě nemají rozvinutou schopnost řeči. Tzv. sekvenční princip zvučnosti (sonority sequencing principle) udává, že lidský zvukový trakt používá amplitudových modulací řečové obálky k segmentování řeči do útvarů podobných slabikám [42]. Centrum slabiky značí vrchol v trajektorii zvučnosti, zatímco hluboké údolí znamená předěl mezi nimi. Tento přístup

je nezávislý na jakékoliv fonologické interpretaci a poskytuje návod, jak jednotlivé slabiky v trajektorii hledat. Trajektorie se dá vypočítat z veličin velice blízkých zvučnosti, z amplitudy nebo intenzity, ale i za pomoci komplexnějších přístupů, které modelují principy lidského slyšení. Popisovaný algoritmus estimuje trajektorii sonority z modelu, založeného na harmonických oscilátorech a rytmické segmentaci.

Za účelem simulace frekvenční selektivity lidského sluchového systému je vstupní řečový signál  $s[n]$  filtrován do 20ti logaritmicky umístěných frekvenčních pásů mezi 100 - 7500 Hz. Filtrace je provedena Gammatone bankou filtrů, často používanou v modelech sluchových systémů. Výsledné amplitudové obálky každého pásu  $e_c[n]$ , kde  $c = 1, 2, \dots, 20$ , jsou podvzorkovány na frekvenci  $f_{s\_re} = 1000$  Hz a následně zpracovány harmonickým oscilátorem na zhodnocení rytmických charakteristik dané obálky.

Každý oscilátor má danou centrální frekvenci  $f_0$  a šířku pásma  $\Delta f$ .  $f_0$  představuje nejcitlivější frekvenci oscilátoru a  $\Delta f$  je rozsah hodnot okolo  $f_0$ , které ovlivňují samotný výstup. Tyto parametry jsou nastaveny tak, aby co nejlépe reflektovaly slabikovou strukturu, standardně je  $f_0 = 5$  Hz a  $\Delta f = 6$  Hz.

Oscilátor v diskrétním čase s obálkou  $e_c[n]$  jako vstupem je popsán následovně

$$F_c[n] = e_c[n] - kx_c[n-1] - dv_c[n-1], \quad (2.10)$$

$$v_c[n] = v_c[n-1] + F_c[n]/mf_{s\_re}, \quad (2.11)$$

$$x_c[n] = x_c[n-1] + v_c[n]/f_{s\_re}, \quad (2.12)$$

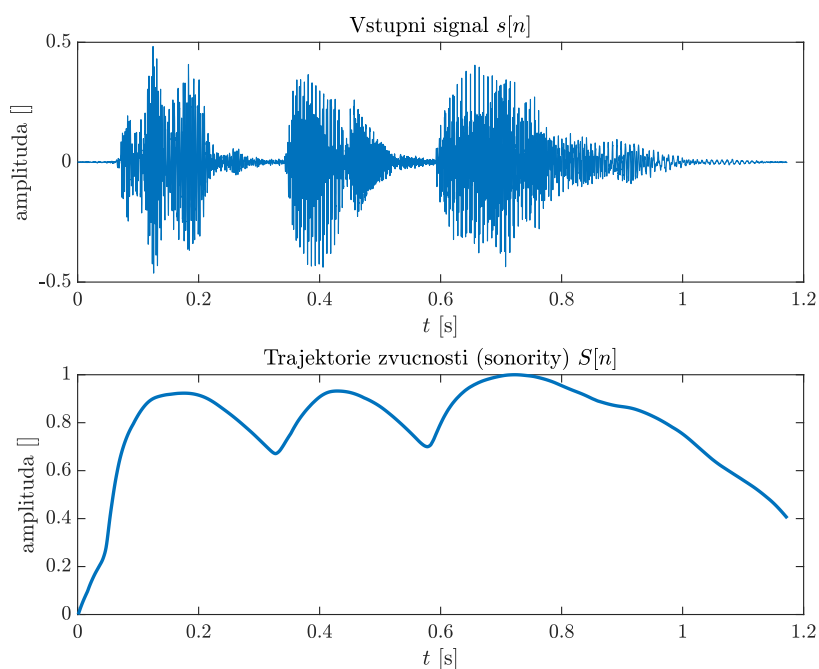
kde  $F_c[n]$ ,  $v_c[n]$  a  $x_c[n]$  představují sílu, rychlost a amplitudu daného oscilátoru frekvenčního pásu  $c$  v čase  $n$ . Parametry  $m$ ,  $k$  a  $d$  jsou ekvivalentem hmotnosti, konstanty tuhosti a útlumu. Jejich nastavením se kontroluje hodnota centrální frekvence  $f_0$  a šířky pásma  $\Delta f$ . V algoritmu je nastavena fixní hodnota  $k = 1$  a zbylé dva parametry se nastaví podle požadovaných hodnot  $f_0$  a  $\Delta f$ .

V dalším kroku je spočtena hodnota trajektorie sonority  $S[n]$  v daném čase. Je vzato  $N$  amplitud oscilátoru  $x_c[n]$ , spočtených podle (2.12), s největším celkovým výkonem a je proveden jejich logaritmický součet,

$$S[n] = \sum_{i=1}^N \log_{10}(\mathbf{x}_i[n]), \quad (2.13)$$

kde  $\mathbf{x}_i[n]$  představuje vektor amplitud  $x_c[n]$  pro každý frekvenční pás  $c$ , seřazených sestupně podle výkonu v čase  $n$ . Fixní konstanta je přičtena ke všem  $\mathbf{x}_i$  aby logaritmus nepracoval se zápornými čísly a výsledek by reálný. Předběžnými testy se hodnota  $N$  nastavila na danou hodnotu  $N = 8$  pásů. Ilustrační ukázka podoby trajektorie sonority je na obrázku 2.7.

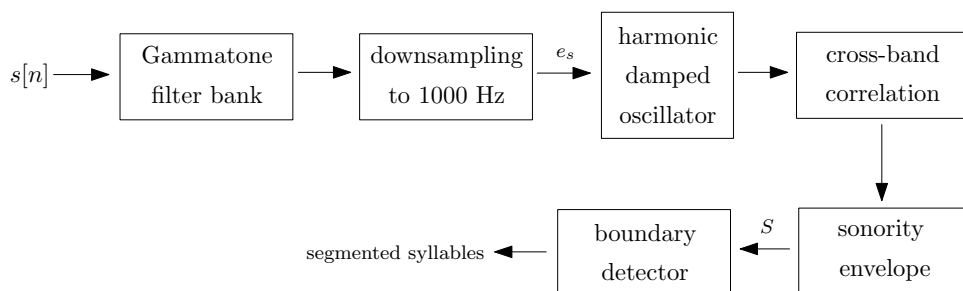
Tento postup představuje analogii k celkové kros-korelační funkci mezi  $N$  nejvíce výkonnými frekvenčními pásy, je však numericky více stabilní a výpočetně nenáročnější.



**Obrázek 2.7:** Ilustrační obrázek podoby trajektorie zvučnosti (sonority) pro promluvu „Dej to sem“.

V posledním kroku je trajektorie zvučnosti  $S[n]$  normována na hodnoty v rozsahu  $[0, 1]$ . Hranice mezi slabikami je určena v místech lokálního minima, kterému předchází lokální maximum, vyšší aspoň o fixní danou prahovou hodnotu.

Principiální schéma fungování algoritmu je na obrázku 2.12.



**Obrázek 2.8:** Schéma detektoru slabik theta Oscillator.

### 2.3.3 LFME

Low frequency modulated energy (LFME) algoritmus [39] funguje na základě energetických charakteristik a předpokladu, že podstatná informace

o slabikovém složení je obsažena v nízkofrekvenčních pásmech. Metoda byla vyvinuta speciálně pro práci s dysartickou řečí na kratších promluvách. Princip fungování je podobný jako u předchozího přístupu. Algoritmem je stanovena charakteristická LFME trajektorie, ze které se posléze určují možné pozice slabik.

Ze vstupního řečového signálu  $s[n]$  jsou vypočteny energetické obálky  $e_i[m]$  příslušných frekvenčních pásů podle

$$e_i[m] = \frac{1}{N_{\text{dft}}N_{\text{win}}} \sum_{f=F_{i,0}}^{F_{i,1}} w[f]|X[f, m]|^2, \quad (2.14)$$

kde  $X[f, m]$  je krátkodobá Fourierova transformace (STFT) signálu  $s[n]$ , kde  $m$  značí příslušný časový úsek (okno) a  $f$  danou frekvenci. Je implementována pomocí rychlého DFT algoritmu a  $N_{\text{dft}}$  značí jeho užitou délku.  $N_{\text{win}}$  značí délku okna a  $w[f]$  jsou zvyšující se váhy působící proti energetickému poklesu u vyšších frekvencí. Hranice energetických pásů jsou obsaženy v proměnné  $F_{i,j}$ , kde  $i$  značí index řádku a  $j$  sloupce, definované jako

$$F_{i,j} = \begin{pmatrix} 50 & 800 \\ 800 & 1500 \\ 1500 & 2500 \\ 2500 & 4000 \end{pmatrix} \quad [\text{Hz}]. \quad (2.15)$$

Z vypočtených energetických pásem  $e_i[m]$  je následně spočtena LFME trajektorie jako

$$\text{LFME}[m] = e_0^2[m]e_s[m], \quad (2.16)$$

kde

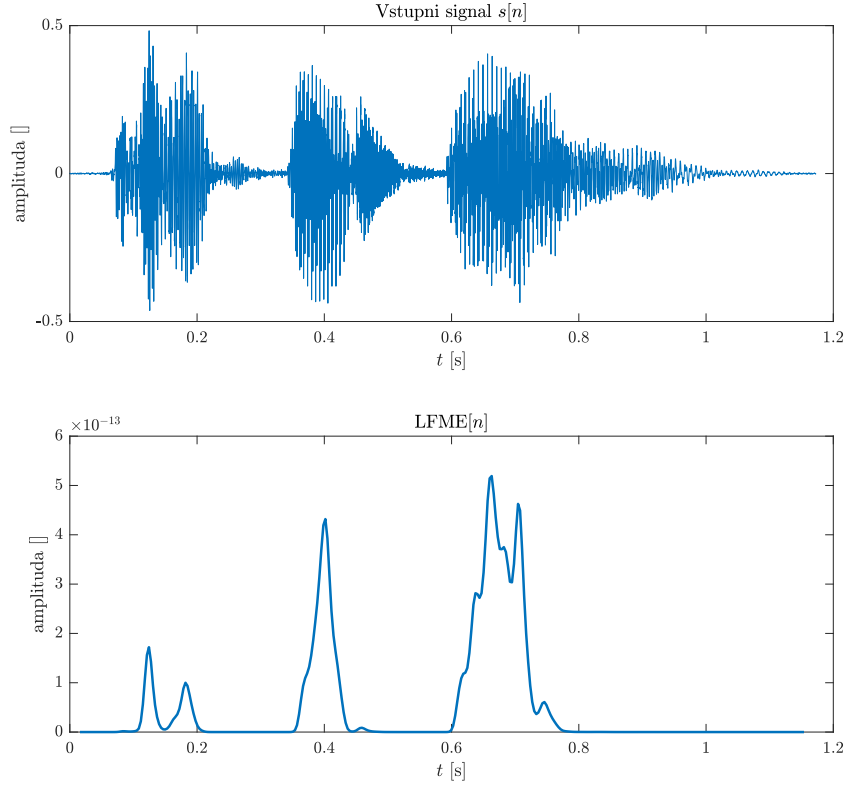
$$e_s[m] = \sum_{i=1}^3 e_i[m]. \quad (2.17)$$

Jedná se o násobení dvou energetických obálek.  $e_0$  nese informaci o výkonové charakteristice v nejnižším frekvenčním pásmu a  $e_s$  v ostatních, vyšších pásmech. Podle předpokladu, že podstatná část informace o rozložení slabik se nachází na nízkých energiích, je efekt  $e_0$  zvýrazněn kvadrátem. Ilustrační obrázek LFME trajektorie je znázorněn na obrázku 2.9.

Jednotlivé slabiky způsobí vrcholy v této trajektorii. Ne všechny vrcholy však mohou být způsobeny právě slabikou, může se jednat i o jiné útvary a zvuky. Také může nastat situace, že jedna slabika vyprodukuje více vrcholů. Detektor slabik na LFME charakteristice má několik kroků, ve kterých jsou tyto případy ošetřeny.

Pokud je výška vrcholu, měřená jako rozdíl mezi vrcholem a údolím před ním, menší než hodnota  $E_{\text{height}}$ , jedná se pravděpodobně o vrchol způsobený jinými zvuky a je zahozen.

Lokální maximum může být také z důvodu výskytu plozivní souhlásky. To je ošetřeno provedením energetického poměru mezi dvěma frekvenčními pásy,



**Obrázek 2.9:** Ilustrační obrázek LFME trajektorie pro promluvu „Dej to sem“.

$E_{r2b}[m_p]$ , spočítaného jako  $E_{r2b}[m_p] = e_0[m_p]/e_s[m_p]$  v místě vrcholu  $m_p$ . Pokud je nižší než prahová hodnota, je druhý vrchol vyhozen.

Na ošetření případu, kdy jedna slabika způsobí několik vrcholů, je nejprve spočtena jejich vzájemná vzdálenost. Pokud jsou moc blízko sebe, je zavedena vektorová veličina  $E_n$ , popisující spektrální rozložení energií v daném místě jako

$$E_n[m] = \frac{1}{\sum_{i=0}^3 e_i[m]} (e_0[m], e_1[m], \dots, e_3[m]). \quad (2.18)$$

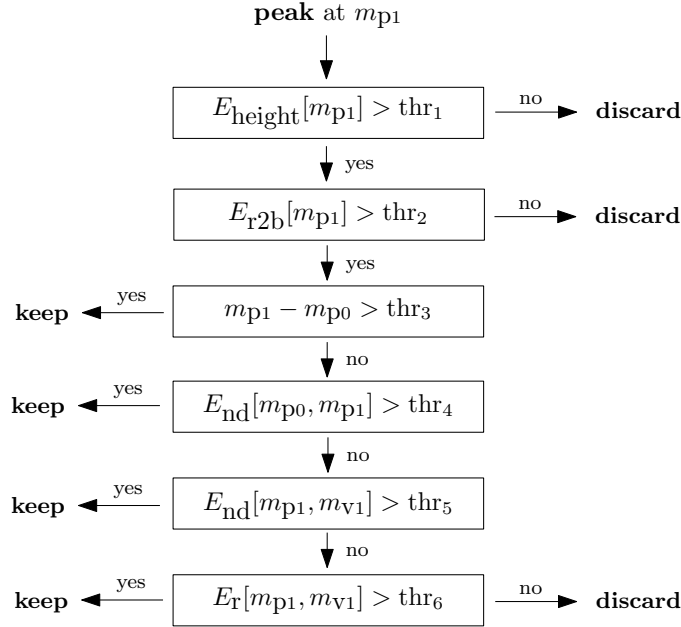
Rozdíl ve spektrálním rozložení dvou pozic vrcholů  $m_{p0}$  a  $m_{p1}$  je pak spočten jako Euklidovská vzdálenost

$$E_{nd}[m_{p0}, m_{p1}] = \left| \left| 10 \log_{10} E_n[m_{p1}] - 10 \log_{10} E_n[m_{p0}] \right| \right|. \quad (2.19)$$

Pokud je hodnota  $E_{nd}[m_{p0}, m_{p1}]$  nižší než daný práh, nejspíše se jedná o stejnou slabiku. Stále se však může jednat o situaci dvou slabik, oddělených jiným fonémem. To je ošetřeno spočtením hodnoty  $E_{nd}[m_{v1}, m_{p1}]$ , kde  $m_{v1}$  je pozice údolí mezi danými dvěma vrcholy. Pokud je tato hodnota velká, dochází k podstatné změně spektrální distribuce mezi údolím a druhým vrcholem a měl by být zachován.

Pokud je tato hodnota malá, je provedeno poslední prahování pro případ, kdy je velký energetický rozdíl mezi údolím a vrcholem, pravděpodobně

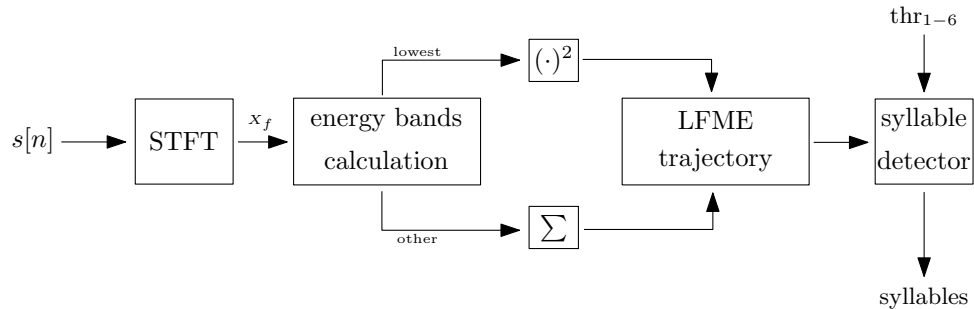
značící úmysl mluvího vyslovit dvě slabiky za sebou. To je zhodnoceno pomocí energetického poměru dané pozice vrcholu a údolí,  $E_r[m_{p1}, m_{v1}]$ . Celkové schéma prahování vrcholů je na obrázku 2.10.



**Obrázek 2.10:** Celkové schéma prahování u detektoru slabik LFME algoritmu.

V původním článku autoři stanovili hodnotu potřebných šesti prahovacích koeficientů pomocí vyhledávacího algoritmu na mřížce fixních hodnot parametrů. Byla zvolena ta kombinace, která minimalizuje RMS chybu mezi referenčními a spočtenými počty slabik pro celou promluvu. Tato data bohužel nebyla dána k dispozici, tedy pro tuto práci byly hodnoty nastaveny podle čtyř promluv. Vždy po jedné ze skupiny zdravých, pacientů s PN, RBD a MSA. Hodnota vah  $w_f$  byla určena tak, aby kompenzovala pokles -6 dB na oktávu (dvojnásobek aktuální frekvence), podle doporučení autorů.

Principiální schéma celého LFME algoritmu je na obrázku 2.11.



**Obrázek 2.11:** Principiální schéma fungování LFME algoritmu.



### ■ 2.3.4 Praat skript

Metoda [31] byla vyvinuta v softwarovém systému PRAAT [43], standardním nástrojem na řečovou analýzu. Obdobně jako u předchozích přístupů je sestavena charakteristická trajektorie ze které jsou následně detekovány pozice jednotlivých slabik. Zde tuto funkci plní tzv. *intenzita*. Následný detektor pak využívá estimátoru základní hlasivkové frekvence pro klasifikaci pozic slabik.

Motivací pro sestavení této metody bylo automatizovat úlohu měření rychlosti řeči, jako jedné z částí posouzení řečové plynulosti. Podle autorů je tato úloha v praxi často přeskakována z důvodu značné časové náročnosti při manuálním vyhodnocování.

V prvním kroku je spočtena trajektorie intenzity  $I[n]$ . Ta je získána konvolucí kvadrátu vstupního signálu  $s[n]$  a Gaussovského okna  $w[n]$  jako

$$I[n] = \sum_{i=-L}^L s^2[n]w[n-i+1], \quad (2.20)$$

kde  $L$  je délka okna  $w[n]$  definovaného jako

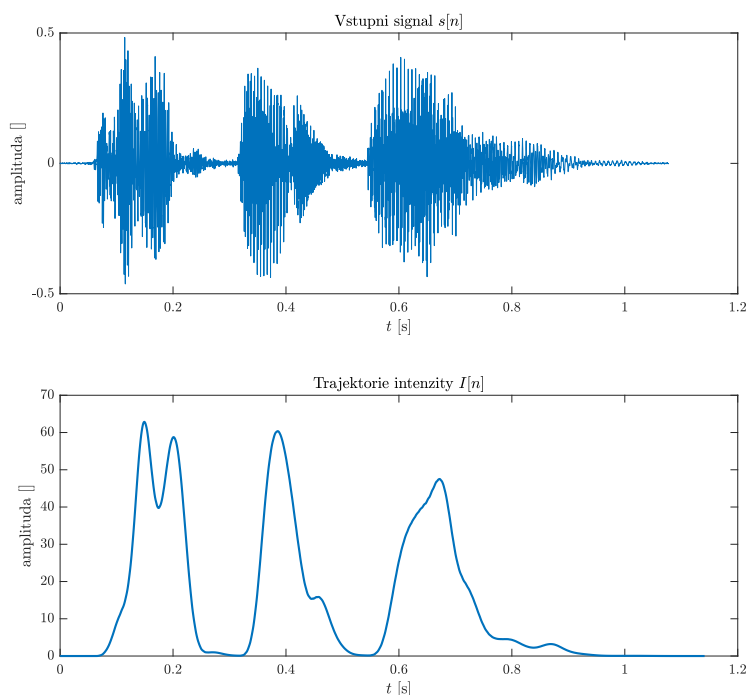
$$w[n] = \exp\left(-\frac{n^2}{2\sigma^2}\right), \quad -\frac{L-1}{2} \leq n \leq \frac{L-1}{2}. \quad (2.21)$$

$\sigma$  představuje standardní odchylku Gaussovské náhodné proměnné. Délka okna  $L$  byla stanovena na  $3.2f_s/f_{\min}$ , kde  $f_{\min}$ , minimální frekvence možné periodicity signálu, je určena jako  $f_{\min} = 50$  Hz a  $f_s$  je příslušná vzorkovací frekvence. Ilustrační obrázek průběhu  $I[n]$  je na obrázku 2.12.

Z trajektorie  $I[n]$  jsou následně určeni kandidáti na pozice potenciálních slabik. Jako kandidát je určen ten vrchol, který přesáhne svou výškou danou hodnotu  $T_{\text{height}}$  nad *mediánem* intenzity celé promluvy. Volba mediánu místo střední hodnoty je z důvodu možné přítomnosti extrémních maxim a jejich nežádoucího vlivu.

V dalším kroku je určen výškový rozdíl mezi vrcholem a předcházejícím údolím. Pokud je menší než fixní hodnota  $T_{\text{diff}}$  je kandidát zahozen.

V posledním kroku je využit estimátor základní hlasivkové frekvence a je určen průběh  $F_0$  přes celou promluvu. Vrcholy, které se nenacházejí v daném průběhu jsou považovány za neznělé a jsou také vyhozeny. Je použit autokorelační estimátor [44], součást systému PRAAT. Principiální schéma celé metody je na obrázku 2.13.



**Obrázek 2.12:** Ilustrační obrázek průběhu intenzity  $I[n]$  pro promluvu „Dej to sem“.

## 2.4 Testování metod

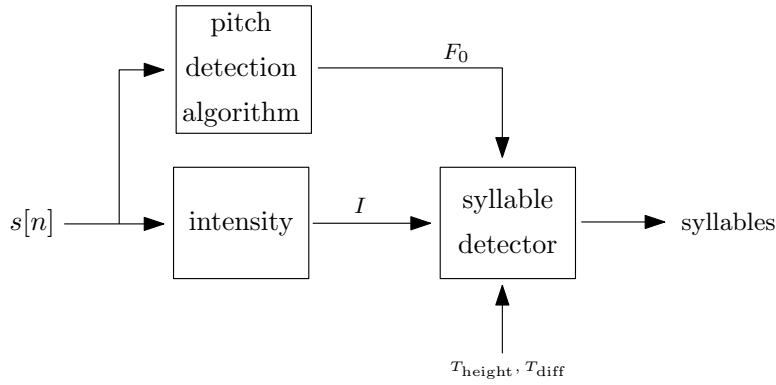
V následující části je uvedena metodika testování přesnosti všech zmíněných algoritmů detekce slabik v úloze měření řečové rychlosti. Jedná se tedy o čtyři metody převzaté z literatury, *Rozpoznávač*, *theta Oscilátor*, *LFME* a *Praat skript*, jejichž princip fungování byl popsán v části 1.3.

K těm byl přidán nově navržený algoritmus, popsán v části 2.2.3. Metodu zde pojmenujeme *BACD detektor*, podle hlavní charakteristiky na které je celý přístup postaven.

### 2.4.1 Kritéria testování a cílové veličiny

Nejprve musíme ustanovit cílové veličiny, podle kterých budou algoritmy testovány. Vzhledem k existující literatuře a praktičnosti stanovení referenčních hodnot byly zvoleny celková *řečová rychlost* a *artikulační rychlost*.

Celková řečová rychlost (speech rate)  $v_{SR}$ , bude spočítána jako počet všech slabik  $N_{syll}$  v celkové promluvě  $s[n]$ , dělený její časovou délkou  $t_s$  v sekundách,



**Obrázek 2.13:** Principiální schéma metody Praat skript.

neboli

$$v_{\text{sr}} = \frac{N_{\text{syll}}}{t_{\text{s}}}, \quad [\text{slabik/s}]. \quad (2.22)$$

$v_{\text{sr}}$  je ovlivněna veškerým obsahem promluvy. Má na ní například vysoký vliv počet a délka řečových pauz v promluvě. Proto bývá  $v_{\text{sr}}$  nižší pro improvizované monology, než předčítaný text pro stejného mluvčího. Celková řečová rychlost bude obdobná pro řečníka mluvčího velice rychle ale vytvářejícího dlouhé pauzy a člověka s normálním tempem řeči bez delších pauz.

Veličina, nezávisající na řečových pauzách je celková artikulační rychlost (articulation rate)  $v_{\text{ar}}$ . Je spočítána obdobně jako  $v_{\text{sr}}$ , ale od celkové délky promluvy  $t_{\text{s}}$  jsou odečteny všechna časová trvání pauz,  $t_{\text{p}}$ , neboli

$$v_{\text{ar}} = \frac{N_{\text{syll}}}{t_{\text{s}} - t_{\text{p}}}, \quad [\text{slabik/s}]. \quad (2.23)$$

Tyto veličiny,  $v_{\text{sr}}$  a  $v_{\text{ar}}$ , jsou přímo závislé na počtu slabik v promluvě  $N_{\text{syll}}$ . Tuto hodnotu získáme jako výstup dané automatické metody. Časovou délku pauz v promluvě,  $t_{\text{p}}$ , získáme pomocí segmentačního algoritmu [52], který klasifikuje úseky promluvy mezi řečové, neřečové a pauzy.

Kritéria testování budou dvě. První je chybová hodnota NRMSE (normalised root mean square error), definovaná jako

$$\text{NRMSE} = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{x}_i - x_i)^2}}{\max(\hat{x}) - \min(\hat{x})}, \quad (2.24)$$

kde  $N$  je počet promluv,  $\hat{x}_i$  je odhadnutá hodnota veličiny  $x$  pro promluvu  $i$  a  $x_i$  je související referenční hodnota.

Druhé kritérium je Spearmanův korelační koeficient  $r$  jako metrika nelineární míry korelace mezi dvěma náhodnými veličinami, přičemž ty nemusí splňovat podmínku normality rozložení.

## 2.4.2 Referenční hodnoty

Vzhledem k podobě cílových veličin  $v_{sr}$  a  $v_{ar}$ , definované jako (2.22) resp. (2.23), stačí ke stanovení referenčních hodnot počet slabik v promluvě,  $N_{syll}$ . Ten byl stanoven na základě ručního přepisu každé řečové nahrávky.

Manuální přepis byl stanoven podle poslechu promluvy příslušným přepisovatelem. Byla preferována spisovná slova, dodržovalo se značení nedokončených slov a koktavosti, zachovaly se významné zvukové vsuvky (jako eee apod.). U nesrozumitelných slov silně dysartických pacientů byl jejich význam tipován, případně přepsáno doslova, pokud to bylo možné. V případě, že některé úseky bylo nemožné smysluplně přepsat, byly tyto úseky z promluvy vynechány.

Z výsledné přepsané nahrávky pak bylo spočítáno celkové množství slabik za pomoci rozdělování slov podle slabikotvorných pravidel a následného použití funkce počtu slov nástroje MS Word.

## 2.5 Statistické zhodnocení

Statistické vyhodnocení má za cíl stanovit, zdali odhadnuté charakteristiky spojené s tempem řeči mohou sloužit jako *biomarker*, indikátor přítomnosti nemoci. Zde konkrétně, jestli nějaký příznak není tak silný, že dokáže rozlišit mezi klinickými skupinami, HC, PN, RBD a MSA. Hlavní hledaný rozdíl bude spočívat v odlišnosti příznaku u zdravých lidí a parkinsonických skupin, tj. zdali nemoc příznak výrazně ovlivňuje, či nikoliv.

V této kapitole budou navrženy odpovídající řečové příznaky a uvedeny nástroje statistického zhodnocení.

### 2.5.1 Navržené příznaky

V této části jsou navrženy řečové příznaky popisující řeč z hlediska její rychlosti, tempa, které charakterizují tento aspekt a projeví se na nich případná porucha správného fungování. V literatuře jsou nejčastěji používané *řečová* a *artikulační rychlost*. K těmto dvěma parametrům bude předmětem zkoumání i nově navržený příznak *variabilita artikulační rychlosti*.

- Celková řečová rychlost  $v_{sr}$ . Spočítaná jako celkový počet slabik dělený délkou promluvy, viz (2.22), udávaná v jednotkách slabik za sekundu.

$v_{sr}$  zahrnuje i podstatný vliv pauz na charakteristiky řečového tempa. Pauzy a jejich pozice a délka mohou mj. reflektovat artikulační přesnost tvorby řeči [6] a jsou závislé na rychlosti, jakou se člověk snaží mluvit.

- Celková artikulační rychlost  $v_{ar}$ . Daná jako celkový počet slabik, dělený celkovou délkou promluvy po odečtení celkové doby trvání všech pauz, viz. výraz (2.23). Měřená též ve slabikách za sekundu. Na rozdíl od  $v_{sr}$  není ovlivněna pauzami v promluvě. Délka všech pauz je určena pomocí řečového segmentačního algoritmu [52].
- Variabilita artikulační rychlosti  $\sigma_{ar}$ . Jedná se o dosud nezkoumaný příznak v oblasti řečového tempa. Je vypočítán jako standardní odchylka průběhu artikulační rychlosti v celé promluvě, počítané po časových oknech, tedy

$$\sigma_{ar} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (v_{ar}[i] - \bar{v}_{ar})^2}, \quad (2.25)$$

kde  $v_{ar}[n]$  představuje průběh artikulační rychlosti,  $n$  je index okna,  $N$  je počet oken a  $\bar{v}_{ar}$  představuje průměrnou hodnotu za celou promluvu. Délka okna byla nakonec zvolena 2 sekundy jako přijatelný kompromis mezi detailností a názorností.

$\sigma_{ar}$  reflektuje míru změn v řečovém tempu. Nízká hodnota znamená konstantní, nebo málo měněnou rychlost v promluvě. Vyšší hodnoty naopak značí výskyt zrychlování a zpomalování řeči.

## 2.5.2 Statistické nástroje

Na zhodnocení významných rozdílů v uvedených příznacích mezi klinickými skupinami byla jako standardní nástroj zvolena jednocestná analýza variance, ANOVA. Byl vybrán Bonferroniho typ na post-hoc testy z důvodu jeho širší uplatnitelnosti než u některých ostatních testů a dále vhodných podmínkách poměrně malého počtu porovnání. Nejprve je však nutné prozkoumat platnost požadovaných předpokladů.

Pro použití nástroje jednocestné ANOVA je třeba ověřit, zda veličiny pochází z normálního rozdělení a mají stejný rozptyl. Podmínka nezávislosti je splněna automaticky, jedná se o data od různých pacientů a jejich nahrávání se nemohlo vzájemně ovlivnit.

Test normality byl proveden pomocí Shapiro-Wilcoxon testu, jako standardního postupu s větší věrohodností než ostatní dostupné testy [55]. Analýza rovnosti rozptylů byla provedena Bartlettovým testem.



## Kapitola 3

### Výsledky

V této kapitole jsou uvedeny výsledky provedených experimentů. Nejprve jsou popsány výsledky testování a porovnání algoritmů, 3.1, a následně výsledky statistického zhodnocení signifikantních rozdílů v určených příznacích mezi klinickými skupinami, 3.2.

#### 3.1 Výsledky testování algoritmů

Výsledky testování metod, z hlediska cílových veličin  $v_{sr}$  a  $v_{ar}$  a testovacích kritérií NRMSE a Spearmanova korelačního koeficientu  $r$  jsou uvedeny v tabulce 3.1.

Z dat je patrné, že metoda *LFME* selhává pro tuto aplikaci u všech skupin (průměrné NRMSE 1.00 pro  $v_{sr}$  a 1.24 u  $v_{ar}$ ,  $0.06 \leq r \leq 0.56$ ,  $p$  od hodnoty menší než 0.01 až po 0.61). Tyto výsledky jsou o mnoho horší než u zbylých metod.

Algoritmy *theta Oscilátor* a *BACD detektor* vykazují podobnou přesnost. Druhý zmíněný má obvykle nižší NRMSE (v průměru 0.33 pro  $v_{sr}$  a 0.49 u  $v_{ar}$ ) oproti 0.56 pro  $v_{sr}$  a 0.61 pro  $v_{ar}$  u prvního. Hodnoty Spearmanova koeficientu jsou obdobné pro oba,  $0.58 \leq r \leq 0.83$  s  $p < 0.001$  nebo  $p < 0.01$  pro BACD detektor a  $0.48 \leq r \leq 0.85$  s  $p < 0.001$  nebo  $p < 0.1$ .

*Rozpoznávač* má zpravidla velmi vysoký korelační koeficient,  $0.64 \leq r \leq 0.90$ , vždy s  $p < 0.001$ . Výsledky z hlediska NRMSE má však horší než poslední dvě zmíněné metody, průměrně 0.64 pro  $v_{sr}$  a 0.85 pro  $v_{ar}$ .

Nejlépejší výsledky ze všech přístupů pro obě kritéria má metoda *Praat*

*skript*, s průměrným NRMSE 0.28 pro  $v_{sr}$  a 0.37 u  $v_{ar}$ . Hodnoty  $r$  se nachází v mezích  $0.65 \leq r \leq 0.91$  s hodnotou  $p$  vždy menší než 0.001.

Celkově se výsledky metod zásadně neliší v závislosti na dané skupině. Některé algoritmy mají případný propad nebo naopak vylepšení u konkrétní skupiny, nejedná se ale o výrazný skok. Například pro BACD detektor je rozdíl mezi PN a RBD skupinou v NRMSE z 0.36 na 0.28 (v hodnotě  $r$  z 0.59 na 0.73,  $p < 0.01$  resp.  $p < 0.001$ ) pro  $v_{sr}$ . Pro  $v_{ar}$  se jedná o změnu NRMSE z 0.52 na 0.40 (pro  $r$  z 0.58 na 0.61,  $p < 0.01$  u obou skupin).



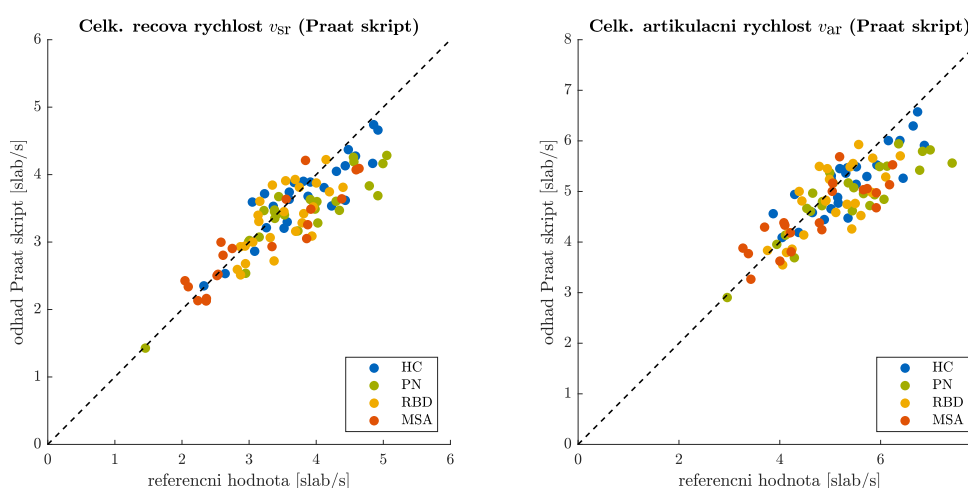
HC skupina	řečová rychlost $v_{sr}$			artikulační rychlost $v_{ar}$		
	NRMSE	$r$	$p$	NRMSE	$r$	$p$
Praat skript	0.22	0.85	<0.001	0.30	0.82	<0.001
theta Oscilátor	0.54	0.75	<0.001	0.55	0.64	<0.001
Rozpoznávač	0.60	0.86	<0.001	0.81	0.79	<0.001
LFME	0.81	0.17	0.42	0.98	0.11	0.61
BACD detektor	0.32	0.67	<0.001	0.45	0.59	<0.01
<b>PN skupina</b>						
Praat skript	0.31	0.85	<0.001	0.43	0.84	<0.001
theta Oscilátor	0.71	0.69	<0.001	0.85	0.48	<0.1
Rozpoznávač	0.60	0.87	<0.001	0.83	0.90	<0.001
LFME	0.90	0.24	0.25	1.10	0.12	0.55
BACD detektor	0.36	0.59	<0.01	0.52	0.58	<0.01
<b>RBD skupina</b>						
Praat skript	0.29	0.69	<0.001	0.36	0.65	<0.001
theta Oscilátor	0.49	0.56	<0.001	0.52	0.57	<0.01
Rozpoznávač	0.68	0.71	<0.001	0.86	0.64	<0.001
LFME	1.24	0.14	0.50	1.46	0.06	0.79
BACD detektor	0.28	0.73	<0.001	0.40	0.61	<0.01
<b>MSA skupina</b>						
Praat skript	0.28	0.91	<0.001	0.38	0.82	<0.001
theta Oscilátor	0.48	0.85	<0.001	0.53	0.64	<0.01
Rozpoznávač	0.68	0.87	<0.001	0.88	0.79	<0.001
LFME	1.06	0.56	<0.1	1.40	0.16	0.49
BACD detektor	0.36	0.83	<0.001	0.58	0.66	<0.01
<b>všechny skup.</b>						
Praat skript	0.23	0.84	<0.001	0.31	0.79	<0.001
theta Oscilátor	0.49	0.77	<0.001	0.55	0.56	<0.001
Rozpoznávač	0.55	0.85	<0.001	0.68	0.79	<0.001
LFME	0.96	0.20	<0.1	1.17	0.08	0.44
BACD detektor	0.33	0.69	<0.001	0.45	0.52	<0.001

**Tabulka 3.1:** Výsledky testování metod podle kritérií NRMSE a Spearmanova koeficientu  $r$  pro HC, PN, RBD a MSA skupinu u cílových veličin  $v_{sr}$  a  $v_{ar}$ .

## 3.2 Výsledky statistického zhodnocení

Na základě výsledků testování algoritmů, uvedených v části 3.1, budeme příznaky odhadovat metodou *Praat skript*, jelikož se ukázala jako celkově nejpřesnější z hlediska daných testovacích kritérií. Doplnkově budou také příznaky odhadovány algoritmem *BACD detektor*, který vykazoval lepší výsledky než zbylé metody v rámci NRMSE kritéria, nepočítaje Praat skript. Takto bude možné porovnat, jak velkou roli má přesnost detektoru na případné odlišení skupin.

Grafy závislosti reference a odhadnutých hodnot  $v_{sr}$  a  $v_{ar}$  algoritmy Praat skript a BACD detektor jsou na obrázcích 3.1, resp. 3.2, pro lepší představu přesnosti výstupu obou metod.

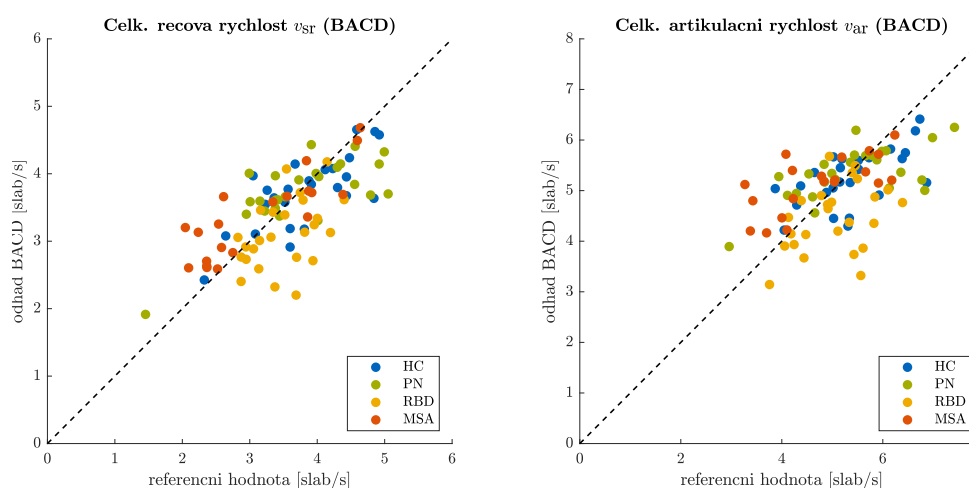


**Obrázek 3.1:** Závislost referenčních a odhadnutých hodnot  $v_{sr}$  a  $v_{ar}$  algoritmem Praat skript. Černá čára představuje perfektní shodu.

### 3.2.1 Výsledky testů předpokladů

Shapiro-Wilcoxon test normality ukázal, že řečovou a artikulační rychlost lze brát jako náhodné veličiny s normálním rozdělením. Na hranici významnosti  $\alpha = 0.05$ , byla nulová hypotéza o normalitě dat zamítnuta pouze u jednoho případu ze 16ti. U příznaku variabilita artikulační rychlosti  $\sigma_{ar}$  se však ukazuje, že předpoklad normality zde neplatí. Nulová hypotéza byla zamítnuta na stejné hladině významnosti  $\alpha$  pro 5 případů z 8mi.

Bartlettův test nezamítl nulovou hypotézu rovnosti rozptylů ve skupinách u žádného případu, pro  $\alpha = 0.05$ .



**Obrázek 3.2:** Závislost referenčních a odhadnutých hodnot  $v_{sr}$  a  $v_{ar}$  algoritmem BACD detektor. Černá čára představuje perfektní shodu.

### 3.2.2 Celkové výsledky

Na základě ověření platnosti předpokladů v předchozí sekci bude zhodnocení významných rozdílů mezi skupinami nástrojem jednocestné ANOVA provedeno pro veličiny řečové a artikulační rychlosti,  $v_{sr}$  a  $v_{ar}$ . Jelikož příznak variabilita artikulační rychlosti  $\sigma_{ar}$  nesplňuje podmínku normality dat aspoň do postačující míry, bude v tomto případě zhodnocení provedeno pomocí Kruskal-Wallis testu, jako forma neparametrické ANOVA.

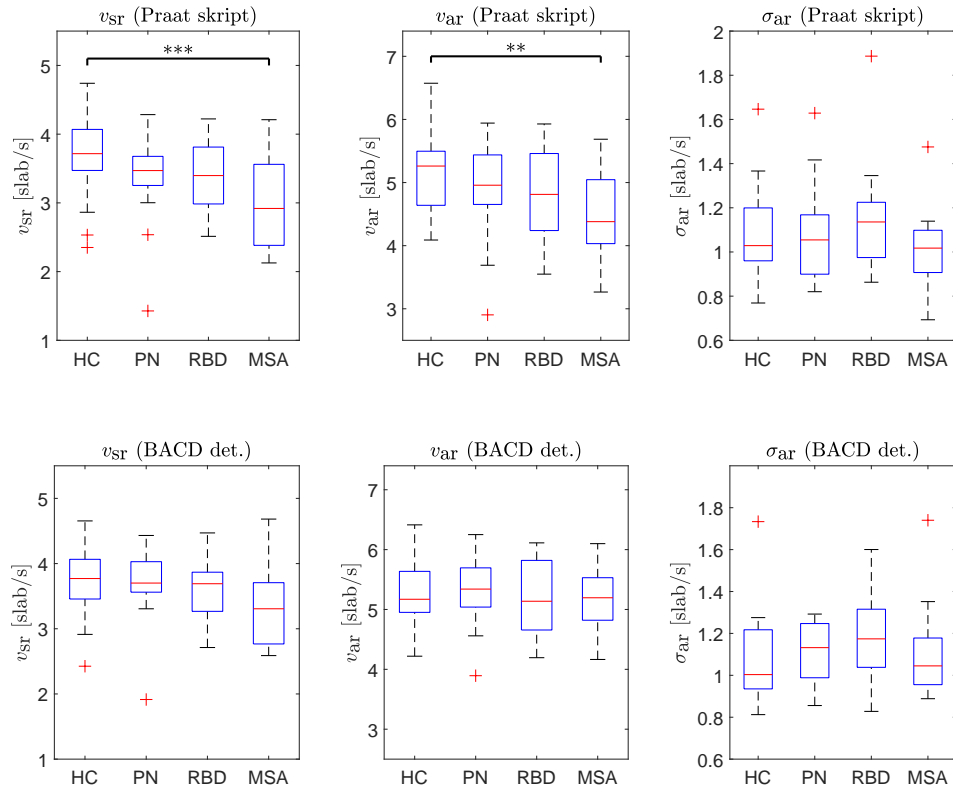
Výsledky, uvedené pro příznaky odhadnuté pomocí metod *Praat skript* a *BACD detektor*, jsou uvedeny v tabulce 3.2. Lze si povšimnout především

$[F(3, 91) = 5.81, p < 0.01, \eta^2 = 0.16]$ ,	pro $v_{sr}$ (Praat skript)
$[F(3, 91) = 2.40, p < 0.1, \eta^2 = 0.07]$ ,	pro $v_{sr}$ (BACD detektor)
$[F(3, 91) = 3.84, p < 0.1, \eta^2 = 0.11]$ ,	pro $v_{ar}$ (Praat skript)
$[F(3, 91) = 0.64, p = 0.59, \eta^2 = 0.02]$ ,	pro $v_{ar}$ (BACD detektor)
$[\chi^2(3, 91) = 4.86, p = 0.18, \eta^2 = 0.05]$ ,	pro $\sigma_{ar}$ (Praat skript)
$[\chi^2(3, 91) = 6.81, p < 0.1, \eta^2 = 0.07]$ .	pro $\sigma_{ar}$ (BACD detektor)

**Tabulka 3.2:** Výsledky testů jednocestné ANOVA a Kruskal-Wallis pro příznaky  $v_{sr}$ ,  $v_{ar}$  a  $\sigma_{ar}$ .

velkého rozdílu mezi algoritmy, kdy pro  $v_{sr}$  a  $v_{ar}$  mají významně větší sílu příznaky odhadnuté metodou Praat skript než BACD detektor ( $F$ -size 5.81 a 3.84 oproti 2.40 a 0.64,  $\eta^2$  0.16 a 0.11 oproti 0.07 a 0.02). Naopak u  $\sigma_{ar}$  je trend opačný, větší sílu má  $\sigma_{ar}$  odhadnutý pomocí BACD detektor ( $\chi^2$ -size 6.81 oproti 4.86,  $\eta^2$  0.07 proti 0.05).

Výsledky post-hoc testů podle Bonferroniho metody jsou znázorněné na obrázku 3.3. Pro příznaky odhadnuté metodou BACD detektor nebyl nalezen žádný významný rozdíl mezi skupinami. U celkové řečové rychlosti  $v_{sr}$ , odhadnuté pomocí Praat skript, byl nalezen signifikantní rozdíl mezi HC a MSA ( $p < 0.001$ ). Pro  $v_{ar}$  byla tato odlišnost také detekována, z lehce nižší významností ( $p < 0.01$ ). Pro  $\sigma_{ar}$  nebyl nalezen žádný výrazný rozdíl.



**Obrázek 3.3:** Výsledky post-hoc testů pro příznaky  $v_{sr}$ ,  $v_{ar}$  a  $\sigma_{ar}$  odhadnuté pomocí metod Praat skript a BACD detektor. Červená linka představuje medián, modrý obdélník mezikvartilové rozpětí, čárkovaná linka je rozsah hodnot, stále nepovažovaných za odlehlé a červené křížky jsou odlehlé hodnoty. \*\*\* značí signifikantní rozdíl  $p < 0.001$ , \*\* značí rozdíl  $p < 0.01$ .

## Kapitola 4

### Závěrečné zhodnocení

#### 4.1 Výsledky práce

V této práci byl na základě dostupné odborné literatury navrhnout algoritmus detekce slabik, určený k automatickému stanovení příznaků rychlosti řeči, pojmenovaný BACD detektor. Zároveň byly vybrány čtyři vhodné již existující metody a bylo provedeno vzájemné porovnání a otestování.

Bylo ukázáno, že nejlepší výsledky z hlediska NRMSE a nelineární korelace s referenčními hodnotami poskytuje metoda *Praat skript*. Dále byly navrženy příznaky, popisující řečovou rychlost - celková řečová rychlost  $v_{sr}$ , celková artikulační rychlost  $v_{ar}$  a variabilita artikulační rychlosti  $\sigma_{ar}$ .

Příznaky  $v_{sr}$  a  $v_{ar}$ , odhadnuté pomocí algoritmu *Praat skript*, vykazovaly zásadní odlišnost pro skupinu zdravých jedinců (HC) a MSA pacientů na významné statistické hladině. Žádný signifikantní rozdíl nebyl pozorován mezi ostatními skupinami a také pro příznak  $\sigma_{ar}$ . Pro srovnání, žádný z parametrů odhadnutých metodou BACD detektor též nevykazoval výrazné meziskupinové rozdíly.

Na základě provedených experimentů a jejich výsledků lze tvrdit, že v oblasti automatických metod odhadu řečové rychlosti a s tím spojené detekce slabik je zapotřebí dalšího intenzivního výzkumu. Žádná z uvedených metod není vyloženě robustní s vynikajícími výsledky. Algoritmus *Praat skript* nabízí uspokojivé výkony, do případného použití v praxi by však bylo potřeba dalšího zlepšení. Navíc byly algoritmy testovány na nahrávkách bez šumu, jehož přítomnost by patrně výsledky ještě více zhoršila.

Ukázalo se, že výkonové charakteristiky nemusejí poskytovat tak špatný a

nerobustní odhad, jak bylo předpokládáno. Zásadní je mít dobře dynamicky ošetřený prahovací mechanismus. Spoléhání na větší počet fixních parametrů nevedl k požadovanému cíli, viz. metoda LFME.

Dále se ukazuje, že pro dostatečně robustní algoritmus není dysartická řeč i ve vyšších stádiích velkou překážkou. Pro všechny čtyři metody, BACD detektor, theta Oscilátor, Rozpoznávač a Praat skript, jsou výsledky obdobné u všech testovaných skupin. Ukazuje se, že pro využití v klinické praxi by mohla sloužit i metoda s původně velmi odlišnou motivací vývoje a rozšiřují se tak hranice, které přístupy dále testovat.

Navržené příznaky řečové rychlosti jsou ve své aktuální podobě a úrovni odhadu schopny významně odlišit pouze pacienty s MSA od zdravé populace. Podle výsledků v sekci 3.2.2 a obrázku 3.3, se tak potvrzuje předpoklad celkového zpomalování řeči u MSA, což je v souladu s literaturou. Ukazuje se také trend zpomalování u RBD, což zatím není potvrzeno žádnou podobnou studií. Trend zpomalování u PN se naopak objevuje v mnohem menší míře, než jaký předpokládá literatura [28, 52].

Ostatní poruchy řečové rychlosti, jako například nenadálé změny rytmu, původně zachycené v parametru  $\sigma_{ar}$ , se zatím ukazují jako nepostihnuté, či nevýrazné s ohledem na větší populaci. Což je opět v rozporu s předpokladem narušení tohoto parametru u PN [52].

Možné vysvětlení proč v této studii nemají PN pacienti tak významné příznaky může být způsoben faktem, že se jedná o de novo pacienty s nízkým stupněm motorického postižení a dysartrie, dosud neovlivněné kognicí. Ukazuje se, že k významným změnám plynulosti, zrychlování řeči a koktavosti dochází až cca 5 let po pravidelné léčbě [56].

## 4.2 Navazující výzkum

S ohledem na úroveň dosažených výsledků lze tuto práci brát jako předstupeň dalšího vývoje v dané oblasti, poukazující na směry, u kterých je zapotřebí navazující výzkum.

První závěr, který se ukazuje jako potřebný, je vylepšit přesnost a robustnost algoritmů detekce slabik. Pro případné použití v praxi by měla daná metoda splňovat alespoň podmínky NRMSE okolo 0.10 a menší s odpovídajícím korelačním koeficientem zhruba 0.90 a vyšším. Tyto výsledky by měly být stabilní bez ohledu na různé typy nahrávek, synukleinopatií a podobně.

Zároveň by u slibných metod mělo být provedeno vyhodnocení odolnosti proti šumu, který v nahrávkách může být přítomen, a stanovení SNR (signal to noise ratio) hranice, při které algoritmy přestávají úlohu zvládat.

Je možné, že hledaná robustní metoda na detekci slabik již byla vyvinuta, avšak z důvodů absence rozsáhlejší aktuální porovnávací studie a nezavedené tradice zveřejňovat zdrojové kódy algoritmů jako open-source (jako například

u metod odhadu  $F_0$ ) nebyla dosud odhalena. I když detekce slabik nepatří mezi vysoce rozvinuté oblasti zpracování signálu, dostupných metod není málo a bez poskytnutých zdrojových dat by celkové větší srovnání nebylo z časových důvodů možné.

Jako nejvhodnější přístup tedy nejspíše bude investování úsilí do navržení nového algoritmu, plně vyhovujícího potřebné úloze s případným srovnáním s několika dalšími vytipovanými metodami.

Při existenci přesného a stabilního algoritmu bude nutné přezkoumat efekty, které jeho přítomnost přinese. Je možné, že se odhadované příznaky stanou přesnějšími a budou schopny odlišit mezi více klinickými skupinami. Přesnost by také umožňovala zavedení případných dalších parametrů, popisujících řečovou rychlost.

Na základě závěrů z výše uvedených možných výzkumných přístupů by se následně definitivně stanovilo, jak velkou roli mají parametry řečové rychlosti v úloze zhodnocení a diagnostiky neurodegeneračních onemocnění.

Zjištění, že vybrané parametry rychlosti řeči fungují jako vhodné markery, v rámci rozsáhlejší studie mapující více řečových příznaků, by vedlo ke zvýšené účinnosti a věrohodnosti zhodnocení dané nemoci. Právě věrohodnost je jeden z nesmírně důležitých aspektů celé úlohy.

Zároveň by byla poskytnuta metoda, jak tyto markery automaticky měřit, která by vedla k poměrně snadné integraci do celkového systému.

Takový systém, sestávající se z podrobné klasifikace a hodnocení řečových poruch, by pak sloužil jako velice efektivní, časově nenáročný a neinvazivní nástroj pro účely hodnocení závažnosti probíhající choroby, monitoringu jejího průběhu, sledování efektů zvolené léčby i případné diagnostiky nemoci.







## Literatura

- [1] M. C. de Rijk, L. J. Launer, K. Berger, M. M. Breteler, J. F. Dartigues, M. Baldereschi, L. Fratiglioni, A. Lobo, J. Martiney-Lage, J. Trenkwalder, A. Hofman, Prevalence of Parkinson's disease in Europe: A collaborative study of population-based cohorts. *Neurologic Diseases in the Elderly Research Group, Neurology* 54 (11) (2000), 21–23.
- [2] L. J. Findley, The Economic Impact of Parkinson's Disease, *Park. Relat. Disord.* 13 (2007), 8–12.
- [3] C. H. Schenck, J. Y. Montplaisir, B. Frauscher, B. Hogl, J. F. Gagnon, R. Postuma, K. Sonka, P. Jennum, M. Partinen, I. Arnulf, V. Cochen de Cock, Y. Dauvilliers, P. H. Luppi, A. Heidebreder, G. Mayer, F. Sixel-Döring, C. Trenkwalder, M. Unger, P. Young, Y. K. Wing, L. Ferini-Strambi, R. Ferri, G. Plazzi, M. Zucconi, Y. Inoue, A. Iranzo, J. Santamaria, C. Bassetti, J. C. Möller, B. F. Boeve, Y. Y. Lai, M. Pavlova, C. Saper, P. Schmidt, J. M. Siegel, C. Singer and E. St Louis, A. Videnovic, W. Oertel, Rapid eye movement sleep behavior disorder: devising controlled active treatment studies for symptomatic and neuroprotective therapy a consensus statement from the International Rapid Eye Movement Sleep Behavior Disorder Study Group, *Sleep Med* 14 (2013), 795–806.
- [4] J. Rusz, J. Hlavnička, T. Tykalová, J. Bušková, O. Ulmanová, E. Růžička, K. Šonka, Quantitative assessment of motor speech abnormalities in idiopathic rapid eye movement sleep behaviour disorder, *Sleep Med.* 19 (2016), 141-147.
- [5] R. B. Postuma, A. E. Lang, J. F. Gagnon, A. Pelletier, J. Y. Montplaisir, How does parkinsonism start? Prodromal parkinsonism motor

- changes in idiopathic REM sleep behaviour disorder, *Brain* 135 (6) (2012), 1860–1870.
- [6] Skodda S, Schlegel U. Speech rate and rhythm in Parkinson’s disease. *Mov Disord* 2008;23, 985-992.
- [7] R. D. Kent, G. Weismer, J. F. Kent, H. K. Vorperian, J. R. Duffy. Acoustic studies of dysarthric speech: methods, progress, and potential. *J. Commun. Disord.* 32 (1999), 141–186.
- [8] E. Szabadi, C. M. Bradshaw, J. A. O. Besson, Elongation of pause-time in speech: a simple, objective measure of motor retardation in depression. *Br J Psychiatry*.129 (1976), 592-597.
- [9] J. Benesty, M. M. Sondhi, Y. Huang. *Handbook of Speech Processing*. Springer, 2008, 7-25.
- [10] J. Benesty, M. M. Sondhi, Y. Huang. *Handbook of Speech Processing*. Springer, 2008, 183-184.
- [11] F. Goldman-Eisler, *Psycholinguistics: experiments in spontaneous speech*. New York: Academic Press, 1968.
- [12] S. Nooteboom. The prosody of speech: melody and rhythm, W.J. Hardcastle, J. Laver (eds.) *The handbook of phonetic sciences*. Blackwell, Oxford, 1997, 640-673.
- [13] S. Narayanan, D. Wang, Speech Rate Estimation via Temporal Correlation and Selected Sub-Band Correlation. *ICASSP-88* (1988), 413-416.
- [14] D. Wang, S. Narayanan, Robust Speech Rate Estimation for Spontaneous Speech. *IEEE Trans. Aud. Spee. Lang. Proc.* 15 (2007), 2190-2201.
- [15] B. Zellner, Fast and slow speech rate: A characterisation for french. *Proc. Int. Conf. Spoken Lang. Proc.*, Sydney, Australia, 7 (1998), 3159–3163.
- [16] T. Pfau, R. Faltlhauser, G. Ruske, A combination of speaker normalization and speech rate normalization for automatic speech recognition, *Proc. Int. Conf. Spoken Lang. Process.*, Beijing, China, 4 (2000), 362–365.
- [17] J. Benesty, M. M. Sondhi, Y. Huang. *Handbook of Speech Processing*. Springer, 2008, 70-73.
- [18] P. Roach, *English Phonetics and Phonology. A Practical Course*, Third ed. Cambridge Univ. Press, 2000.
- [19] W. Poewe, K. Seppi, C. M. Tanner, G. M. Halliday, P. Brundin, J. Volkmann, A. E. Schrag, A. E. Lang, Parkinson disease, *Nat. Rev. Dis. Primers* 23 (3) (2017) 17013.

- [20] M. C. Rodriguez-Oroz, M. Jahanshahi, P. Krack, I. Litvan, R. Macias, E. Bezard, J. A. Obeso, Initial clinical manifestations of Parkinson's disease: Features and pathophysiological mechanisms, *Lancet Neurol.* 8 (2009) 1128–1139.
- [21] E. K. Louis, A. R. Boeve, B. F. Boeve BF, REM Sleep Behavior Disorder in Parkinson's Disease and Other Synucleinopathies. *Mov. Disord.* 32(5) (2007), 645–658.
- [22] D. Arnaldi, E. Antelmi, E. K. Louis, R. B. Postuma, I. Arnulf I, Idiopathic REM sleep behavior disorder and neurodegenerative risk: To tell or not to tell to the patient? How to minimize the risk?. *Sleep Med. Rev.* 36 (2017), 82–95.
- [23] A. Fanciulli, G. K. Wenning, Multiple-System Atrophy. *N. Engl. J. Med.* 372 (2015), 249-263.
- [24] K. Kluin, S. Gilman, M. Lohman, L. Junck, Characteristics of the Dysarthria of Multiple System Atrophy. *Arch Neurol.* 53(6) (1996), 545–548.
- [25] J. Rusz, C. Bonnet, J. Klempir, T. Tykalová, E. Baborová, M. Novotný, A. Rulseh, E. Růžička, Speech disorders reflect differing pathophysiology in Parkinson's disease, progressive supranuclear palsy and multiple system atrophy. *J. neurol.* 262 (2015).
- [26] J. R. Duffy, Motor speech disorders: Substrates, differential diagnosis and management, 3rd ed. Elsevier Health Sciences, (2013).
- [27] A. K. Ho, R. Ianseck, C. Marigliani, J. L. Bradshaw, S. Gates, Speech Impairment in a Large Sample of Patients with Parkinson's Disease. *Behav. Neurol.* 11 (1999) 131–137.
- [28] G. van Nuffelen, M. De Bodt, F. Wuyts, P. Van de Heyning, The Effect of Rate Control on Speech Rate and Intelligibility of Dysarthric Speech. *Folia. Phoniatr. Logop.* 61 (2009), 69-75.
- [29] T. Dekens, M. Demol, W. Verhelst, P. Verhoeve, A comparative study of speech rate estimation techniques. *Interspeech* (2007), 510-513.
- [30] H. Martens, T. Dekens, G. Nuffelen, L. Latacz, W. Verhelst, M. Bodt, Automated Speech Rate Measurement in Dysarthria. *J. sp., lan., hear. res.* 58 (2015).
- [31] N. de Jong, T. Wempe, Praat script to detect syllable nuclei and measure speech rate automatically. *Behav. res. met.* 41 (2009), 385-90.
- [32] Y. Jiao, V. Berisha, M. Tu, J. Liss, Convex weighting criteria for speaking rate estimation. *IEEE ACM T Audio Speech Lang Proces* 23 (2015), 1421-1430.

- [33] H. M. Huici, H. Kairuz, H. Martens, G. Nuffelen, M. Bodt, Speech rate estimation in disordered speech based on spectral landmark detection. *Biomed. Sig. Proc. Cont.* 27 (2016), 1-6.
- [34] R. Villing, J. Timoney, T. Ward, J. Costello, Automatic Blind Syllable Segmentation for Continuous Speech. *ISSC* (2004).
- [35] S. Greenberg, The switchboard transcription project, Tech. Rep., 1996 Johns Hopkins CLSP Work.on Innov. Techn. for Large Voc. Cont. Sp. Rec., Baltimore, MD, (1997).
- [36] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren, DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus. Vienna, Austria: Springer, (1993).
- [37] P. Schwarz, Phoneme Recognition based on Long Temporal Context, PhD Thesis, Brno University of Technology, (2009).
- [38] O. Räsänen, G. Doyle, M. Frank, Pre-linguistic segmentation of speech into syllable-like units. *Cognition* 171 (2018), 130–150.
- [39] T. Dekens, H. Martens, G. Nuffelen, M. Bodt, W. Verhelst, Speech rate determination by vowel detection on the modulated energy envelope. *Europ. Sig. Proc. Conf.* (2014), 1252-1256.
- [40] Google Inc., Google text-to-Speech, November 2013, <https://cloud.google.com/text-to-speech>.
- [41] P. Schwarz, P. Matejka, J. Cernocky, Hierarchical Structures of Neural Networks for Phoneme Recognition, *IEEE Int. Conf. Ac. Spee. Sig. Proc.*, Toulouse (2006), I-I.
- [42] J. Bertoncini, J. Mehler, Syllables as units in infant speech perception. *Infant Behav. Dev.*, 4 (1981), 247–260.
- [43] P. Boersma, Praat, a system for doing phonetics by computer, *Glott. International* 5 (2002) 341–345.
- [44] Paul Boersma, Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proc. Inst. Phon. Sci.* 17 (1993), 97–110.
- [45] R. Cmejla, P. Sovka, Pavel. (2001). Estimation of Boundaries between Speech Units Using Bayesian Change-point Detectors. *Text Sp. Dial*, 2166 (2001), 291-298.
- [46] R. Cmejla, J. Ruzs, P. Bergl, J. Vokral, Bayesian change-point detection for the automatic assessment of fluency and articulatory disorders. *Spee. Comm.*, 55(1) (2013), 178-189.
- [47] L. Kung-Pu, Automatic language identification using syllabic spectral features, *IEEE Int. Conf. Ac., Spee. Sig. Proc.*, Adelaide, (1994), 297-300.

- [48] O. Ruanaidh, W. J. Fitzgerald, Numerical Bayesian methods applied to signal processing, Series on Statistics and Computing, Springer-Verlag (1996).
- [49] G. Birkhoff, H. L. Garabedian, Smooth Surface Interpolation, *J. Math. Phy.*, 39 (1960).
- [50] V. Illner, P. Sovka, J. Ruzs, Validation of freely-available pitch detection algorithms across various noise levels in assessing speech captured by smartphone in Parkinson's disease, *Biomed. Sig. Proc. Cont.* 58 (2020), 101831.
- [51] A. Camacho, G. Harris, A sawtooth waveform inspired pitch estimator for speech and music. *J. Ac. Soc. Am.* 124 (2008), 1638.
- [52] J. Hlavnička, R. Čmejla, T. Tykalová, K. Sonka, E. Růžička, J. Ruzs, Automated analysis of connected speech reveals early biomarkers of Parkinson's disease in patients with rapid eye movement sleep behaviour disorder. *Sci. Rep.* 7 (2017).
- [53] C. Goetz, B. Tilley, S. Shaftman, G. Stebbins, S. Fahn, P. Martinez-Martin, et al, Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov. Disord.* 23 (2008), 2129–2170.
- [54] C. Payan, F. Viallet, B. Landwehrmeyer, A. Bonnet, M. Borg, F Durif, et al, Disease severity and progression in progressive supranuclear palsy and multiple system atrophy: validation of the NNIPPS–Parkinson Plus Scale. *PLoS ONE* 6 (2011), e22293.
- [55] N. Razali, Y. Wah, Power comparisons of Shapiro–Wilk, Kolmogorov–Smirnov, Lilliefors and Anderson–Darling tests. *J. Stat. Mod. Analyt.* 2(1) (2011), 21–33.
- [56] T. Tykalova, J. Ruzs, R. Cmejla, J. Klempir, H. Ruzickova, J. Roth, E. Ruzicka, Effect of dopaminergic medication on speech dysfluency in Parkinson's disease: a longitudinal study. *J Neural Transm* 122 (2015), 1135-1142.