

**ČESKÉ VYSOKÉ  
UČENÍ TECHNICKÉ  
V PRAZE**

**FAKULTA**



**BAKALÁŘSKÁ PRÁCE**

**2022**

**SVITLANA  
HRYNCHENKO**

České vysoké učení technické v Praze

Fakulta elektrotechnická

Katedra teorie obvodů



Bakalářská práce

Analýza hlasu a řeči u pacientů se schizofrenií

Voice and speech analysis in patients with schizophrenia

Autor: Svitlana Hrynchenko

Vedoucí práce: prof. Ing. Roman Čmejla, CSc.

Vedoucí práce – specialista: Mgr. Bc. Juraj Jonáš

Studijní program: Lékařská elektronika a bioinformatika

Praha 2022

## I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Hrynchenko** Jméno: **Svitlana** Osobní číslo: **492081**  
Fakulta/ústav: **Fakulta elektrotechnická**  
Zadávací katedra/ústav: **Katedra teorie obvodů**  
Studijní program: **Lékařská elektronika a bioinformatika**

## II. ÚDAJE K BAKALÁŘSKÉ PRÁCI

Název bakalářské práce:

**Analýza hlasu a řeči u pacientů se schizofrenií**

Název bakalářské práce anglicky:

**Voice and speech analysis in patients with schizophrenia**

Pokyny pro vypracování:

1. Proveďte rešerši literatury zabývající se akustickými charakteristikami schizofrenie v řečovém signálu.
2. Analýzu proveďte v prostředích Praat a MATLAB. Zvažte také možnost využití parametrů získaných automaticky [4] v toolboxu DYSPAN.
3. Vybrané akustické parametry ověřte na řečových signálech kontrolních osob a pacientů pořízených v NÚDZ v Klecanech.
4. Vyhodnoťte možnosti detekce schizofrenie z akustických charakteristik hlasu a řeči.

Seznam doporučené literatury:

- [1] Low, D.M.; Bentley, K.H.; Ghosh, S.S. "Automated assessment of psychiatric disorders using speech: A systematic review." *Laryngoscope Investigative Otolaryngology*, 2020, 5.1: 96-116.
- [2] Fu, J., et al. Sch-net: a deep learning architecture for automatic detection of schizophrenia. *BioMedical Engineering OnLine*, 2021, 20.1: 1-21.
- [3] Rapcan, V., et al. Acoustic and temporal analysis of speech: A potential biomarker for schizophrenia. *Medical engineering & physics*, 2010, 32.9: 1074-1079.
- [4] Hlavníčka, J. "Automated analysis of speech disorders in neurodegenerative diseases". Ph.D. thesis, CTU FEE, 2019, <https://dspace.cvut.cz/handle/10467/79223?show=full>

Jméno a pracoviště vedoucí(ho) bakalářské práce:

**prof. Ing. Roman Čmejla, CSc. katedra teorie obvodů FEL**

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

**Mgr. Bc. Juraj Jonáš Národní ústav duševního zdraví**

Datum zadání bakalářské práce: **26.01.2022**

Termín odevzdání bakalářské práce: **20.05.2022**

Platnost zadání bakalářské práce: **30.09.2023**

prof. Ing. Roman Čmejla, CSc.  
podpis vedoucí(ho) práce

doc. Ing. Radoslav Bortel, Ph.D.  
podpis vedoucí(ho) ústavu/katedry

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

## III. PŘEVZETÍ ZADÁNÍ

Studentka bere na vědomí, že je povinna vypracovat bakalářskou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v bakalářské práci.

\_\_\_\_\_  
Datum převzetí zadání

\_\_\_\_\_  
Podpis studentky



## **Prohlášení**

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

V Praze dne 20.05.2022

.....

Podpis autora práce

## **Poděkování**

Chtěla bych především poděkovat svému vedoucímu bakalářské práce prof. Ing. Romanu Čmejlovi, CSc. za skvělé vedení, podporu a poskytnuté odborné znalosti. Dále bych ráda poděkovala Mgr. Bc. Juraji Jonášovi za poskytnutí studijních materiálů z oblasti výzkumu schizofrenie. Dále bych chtěla poděkovat své rodině, která mi pomáhala při psaní práce i v celém bakalářském studiu.

## Abstrakt

Od doby, kdy byla schizofrenie poprvé identifikována jako duševní porucha, se na téma poruch řeči provedlo značné množství studií. Tyto studie uvedly důležité atypické hlasové projevy, jako je alogie, plochý afekt a anhedonie v moderní klasifikaci onemocnění. Cílem této práce bylo navrhnout vhodné akustické markery pro hodnocení řečových abnormalit u pacientů se schizofrenií.

Za účelem prozkoumání současného stavu problematiky byly nejprve analyzovány publikace o výzkumu poruch řeči u pacientů se schizofrenií. Dále, pro vyhodnocení jednotlivých parametrů a jejich výběr pro automatickou akustickou analýzu, byly pomocí softwaru PRAAT a MATLAB a také prostřednictvím toolboxu DYSAN zkoumány prozodické, artikulační a časové akustické charakteristiky. Testování proběhlo na řečových signálech od 3 kontrolních osob a 3 pacientů pořízených v NÚDZ v Klecanech.

Ze statistické analýzy akustických charakteristik byly jako signifikantní parametry pro detekci schizofrenie vyhodnoceny zejména snížený rozsah a průměrná hodnota základní frekvence v semitónech, zvýšený počet pauz a jejich délky, snížená řečová rychlost, snížený poměr doby artikulace k trvání nahrávky ve spontánní promluvě, zmenšených rozsah prvního formantu u samohlásky /a/  $F_{1a}$  a velikost plochy samohláskového prostoru VSA. Porovnání jednotlivých charakteristik dále naznačuje, že úlohy spontánní promluvy, tedy úlohy s pravděpodobně vyšší kognitivní zátěží, mají schopnost lépe odhalovat vliv schizofrenie na poruchy řeči než úloha prostého čtení.

Tato práce dochází k závěru, že analýza řeči je perspektivním nástrojem pro diagnostiku schizofrenie, protože je neinvazivní, levná a rychlá.

**Klíčová slova:** akustické charakteristiky hlasu, schizofrenie, základní frekvence, rychlost řeči, formanty

## Abstract

There has been a significant volume of studies conducted on the subject of speech disorders ever since schizophrenia was first identified as a mental disorder. These studies indicated important atypical voice patterns, such as alogia, flat affect, and anhedonia in modern disease classifications. This thesis aimed to propose appropriate acoustic markers for the evaluation of speech abnormalities among the patients with schizophrenia.

In order to investigate the current state of the art primarily analysis of publications on linguistic research on speech disorders among patients with schizophrenia was conducted. Next, by using PRAAT and MATLAB software and via the DYSAN toolbox prosodic, articulatory, and temporal acoustic characteristics were investigated to evaluate each parameter and reveal which ones are applicable for automatic acoustic analysis. Testing was performed on speech signals of 3 control subjects and 3 patients taken at the National Institute of Health in Klecany.

From the statistical analysis, the acoustic characteristics evaluated as significant parameters for schizophrenia detection were reduced range and the average value of fundamental frequency in semitones, increased number and pause duration, decreased speech rate, the ratio of articulation time to the recording duration for spontaneous speech, reduced range of the first formant of the vowel /a/  $F_{1a}$  and vowel space area size VSA. Further comparison of individual characteristics suggested that spontaneous speaking tasks, i.e., tasks with probably higher cognitive load, tend to involve larger effect sizes than the reading task.

This thesis concludes that, speech analysis is a prospective tool for diagnosing schizophrenia because it is non-invasive, low-cost, and speedy.

**Keywords:** acoustic features of the voice, schizophrenia, fundamental frequency, speaking rate, formants



# Obsah

1.	Teoretický úvod do problematiky .....	11
1.1.	Cíle práce.....	11
1.2.	Definice schizofrenie.....	11
1.3.	Současné diagnostické metody schizofrenie .....	12
1.4.	Objektivní markery schizofrenie .....	13
1.4.1.	Biologické a fyziologické indikátory .....	14
1.4.2.	Behaviorální indikátory.....	14
1.5.	Řeč jako objektivní indikátor.....	15
1.5.1.	Produkce řeči.....	15
2.	Související práce a popis anomálií řeči pacientů se schizofrenií.....	17
2.1.	Prozodické vlastnosti.....	17
2.1.1.	Základní hlasivková frekvence $f_0$ .....	18
2.1.2.	Intenzita hlasu.....	20
2.1.3.	Artikulační tempo řeči .....	21
2.2.	Formanty .....	23
3.	Výběr metodiky.....	24
3.1.	Databáze promluv .....	24
3.1.1.	Řečové nahrávky .....	24
3.2.	Návrh vlastního algoritmu .....	25
3.2.1.	PRAAT a Parselmouth-praat .....	26
3.2.2.	MATLAB.....	26
3.3.	Analýza jednotlivých parametrů .....	26
3.3.1.	Řečová a artikulační rychlost .....	26
3.3.2.	Variabilita artikulační rychlosti .....	28
3.3.3.	Střední rychlost produkce pauz, poměr trvání řeči k trvání nahrávky .....	29
3.3.4.	Základní frekvence $f_0$ v Hz a semitónech .....	30
3.3.5.	Rozsah intenzity hlasu .....	31
3.3.6.	Analýza formantů F1 a F2.....	31
3.4.	Parametry získané automaticky v toolboxu DYSAN .....	33
3.4.1.	Prozodické charakteristiky .....	34
3.4.2.	Časové charakteristiky .....	34
3.4.3.	Artikulační charakteristiky .....	35
3.4.4.	Fonační charakteristiky .....	35
3.4.5.	Respirační charakteristiky .....	36
4.	Výsledky.....	37
4.1.	Prozodické charakteristiky.....	37
4.2.	Časové a artikulační charakteristiky.....	38
4.3.	Fonační charakteristiky .....	39
4.4.	Parametry získané automaticky v toolboxu DYSAN .....	41
5.	Diskuze.....	43
6.	Závěr .....	48
	Použitá literatura.....	50

Příloha A – Zkratky a symboly: .....	57
Příloha B – Doplnující tabulky a obrázky: .....	58
Příloha C – Protokol k výzkumu pacientů se schizofrenií:.....	61

# 1. Teoretický úvod do problematiky

---

## 1.1. Cíle práce

Řeč je nositelem mnoha významných informací týkajících se sociálního vnímání. Patří sem informace o identifikaci mluvčího (např. fyzické charakteristiky, jako je pohlaví, věk) a také emocionální stav mluvčího. Psychické a neurologické poruchy jsou dvě hlavní kategorie lidských poruch, které ovlivňují myšlení, řeč a chování člověka [1]. V současné době je celosvětová prevalence psychických a neurologických poruch více než 12 %, resp. 10 % [2].

Automatizované řečové biomarkery jsou stále více žádoucí pro odborníky v oblasti řečových patologií a neurology, ale složitost interpretace jednotlivých řečových charakteristik je hlavní překážkou mezi technickými a klinickými aplikacemi.

Tato bakalářská práce má dva základní cíle, které vyplývají z uvedených pokynů pro vypracování. Prvním cílem je provést rozsáhlou rešerši literatury a stanovit nejvýznamnější řečové biomarkery, které je možné použít pro vývoj budoucího automatického rozpoznání přítomnosti schizofrenie u pacientů. Ve stručnosti budeme hledat řečový projev schizofrenie na základě akustických, prozodických a formantových charakteristik řeči, které vycházejí z teoreticky popsaných projevů v řeči pacientů se schizofrenií. Druhým cílem je tyto navržené markery prozkoumat na základě řečových signálů kontrolních osob a pacientů pořízených v NÚDZ v Klecanech. Velká část této práce je věnována návržení užitečných algoritmů pomocí softwarů PRAAT a MATLAB pro hodnocení možnosti detekce schizofrenie z akustických charakteristik hlasu a řeči.

## 1.2. Definice schizofrenie

Schizofrenie je chronické a závažné duševní onemocnění, které postihuje zhruba 1 % populace na celém světě [3] a které do značné míry ovlivňuje každodenní život a práci pacientů. Psychotické příznaky (jako jsou halucinace a bludy) se u mužů obvykle projevují na konci dospívání a před 25. rokem života a u žen mezi 25. a 35. rokem života. Zřídka se objevují po 45. roce života a velmi zřídka před pubertou, ačkoli existují popsané případy schizofrenie u dětí ve věku 5 let. Prvními příznaky u dospívajících může být změna kamarádů, zhoršení

prospěchu, potíže se spánkem, dráždivost. Vzhledem k tomu, že stejné modely chování jsou běžné u mnoha psychicky zdravých jedinců, je obtížné v této době stanovit diagnózu. Výzkum ukázal, že schizofrenie postihuje stejně muže i ženy a že výskyt je ve všech etnických skupinách na světě přibližně stejný.

Klinická diagnóza schizofrenie, o které bude uvedeno více v *Podkapitole 1.2*, je obvykle založena na úplném psychiatrickém zhodnocení řeči a chování pozorovaném prostřednictvím klinických rozhovorů. Příznaky schizofrenie lze rozdělit na dva typy, pozitivní příznaky a negativní příznaky. Mezi pozitivní příznaky patří halucinace a bludy a mezi negativní příznaky patří plochý afekt, alogie (*chudost řeči*), ztráta zájmu a postižení v aktivitách.

Schizofrenie se vyznačuje poruchami myšlení, řeči a abnormálním chováním, je spojena s výraznými deficity v oblasti řečových a emočních funkcí [4]. V této práci se zaměříme na negativní příznaky schizofrenie, které se projevují v řeči pacientů zploštěním intonace, monotónností, zvýšeným počtem řečových pauz, sníženou řečovou aktivitou, dále snížením řečové a artikulační rychlosti. U pacientů je porucha řeči spojena s poruchami pozornosti a pracovní paměti, která se také zhoršuje vlivem stresu a s přechodem onemocnění do chronické fáze. Studie naznačují, že celoživotní prevalence výrazných primárních negativních příznaků je 15-20 % [5]; nicméně mnoho, ne-li většina osob se schizofrenií má určitou úroveň negativní symptomatologie.

Navzdory současným znalostem v oblasti genetiky rizika schizofrenie zůstává etiologie a patogeneze tohoto onemocnění stále nedostatečně vysvětlena [6]. Klinický obraz jednotlivých pacientů se může výrazně lišit, což je způsobené velkou inter-individuální variabilitou v projevech symptomů u pacientů se schizofrenií. Klinická diagnóza navíc závisí na zkušenostech klinických odborníků, proto je nezbytné navrhnout metodu, která by umožnila objektivně a efektivně diagnostikovat pacienty s negativními příznaky. V této práci, jak již bylo uvedeno v cílech, se zaměříme především na navržení metody definice schizofrenie z řečových záznamů.

### **1.3. Současné diagnostické metody schizofrenie**

Definice schizofrenie se vyvíjela v průběhu šesti vydání Diagnostického a statistického manuálu duševních poruch (DSM-I, DSM-II, DSM-III, DSM-III-R, DSM-IV a DSM-IV-TR; Americká psychiatrická asociace, 1952, 1968, 1980, 1987, 1994, 2000). DSM-V zahrnuje celou řadu diagnóz, mezi které patří depresivní poruchy, stravování, poruchy osobnosti a poruchy související s traumaty a stresory.

Současné strategie hodnocení schizofrenií jsou však závislé na subjektivních stupnicích hodnocení příznaků, které mají často psychometrická a praktická omezení. [8]. Škály diagnostiky pozitivních a negativních příznaků zůstávají hlavním způsobem hodnocení klinickými lékaři a výzkumníky.

Přestože jsou tyto škály široce používány, kvalita i obecná užitečnost jednotlivých škál se liší. Existuje relativně méně článků o užitečnosti novějších škál, jako jsou CAINS (*Clinical Assessment Interview for Negative Symptoms*) a BNSS (*Brief Negative Symptom Scale*), které je srovnávají se staršími škálami PANSS (*Positive and Negative Symptoms Scale*), SAPS (*Scale for the Assessment of Positive Symptoms*), SANS (*the Scale for the Assessment of Negative Symptoms*), NSA-16 (*Negative Symptom Assessment-16*) a CGI-SCH (*Clinical Global Impression Schizophrenia*) [9]. Přehled těchto škál, jejich výhody a nevýhody použití, které vyplývají z literární studie, jsou uvedeny v příloze v *Tabulce B1*.

Ačkoliv metody hodnocení kliniků vykazují akceptovatelnou spolehlivost, jsou v mnohém omezené. Především jsou založeny na poměrně nejednoznačných definicích pokrývajících širokou škálu projevů chování, nedostatečně zohledňují kulturní a individuální odlišnosti, které výrazně ovlivňují řečový a jiný projev, a nabízejí omezené možnosti odpovědi, které nejsou příliš citlivé [10, 11]. Technologický pokrok však umožňuje spolehlivější a přesnější hodnocení na základě počítačové akustické analýzy [12]. Tato měření nabízejí potenciálně větší spolehlivost, platnost a citlivost, což vede k větší přesnosti hodnocení [13]. Žádná jednoznačná metoda počítačového hodnocení dosud nebyla vymyšlena a přijata k širšímu praktickému použití. V konečném výsledku má výzkum objektivních markerů potenciál pomoci snížit velké socioekonomické náklady spojené se schizofrenií a zlepšit přesnost psychiatrické diagnostiky a účinnost léčby mnoha duševních onemocnění [14].

## **1.4. Objektivní markery schizofrenie**

Objektivní markery schizofrenie mají široké možnosti využití v oblasti psychiatrie. Mohly by být použity jako objektivní diagnostická podpora v zařízeních primární péče a na specializovaných klinikách, poskytnout nástroj s nízkými nároky na automatický sběr a analýzu informací u velkých populací a poskytnout okamžitou zpětnou vazbu a léčebné poradenství ohroženým lidem. Správné snímání neverbálních signálů založených na chování by mohlo být využito například v platformě pro smartphony pro diagnostiku na dálku, pomáhat v urgentních situacích [15].

Vzhledem k širokému klinickému profilu schizofrenie je však pravděpodobné, že k nalezení skutečně objektivních markerů bude nezbytný víceúrovňový přístup. Proto je důležité stručně uvést výsledky současného výzkumu souvisejících biologických, fyziologických a behaviorálních indikátorů.

#### **1.4.1. Biologické a fyziologické indikátory**

Schizofrenie je běžné duševní onemocnění, které je důsledkem složité souhry genetických a s životním prostředím souvisejících rizikových faktorů. Stanovení její primární molekulární a buněčné fyziologie se ukázalo jako problematické. Biologické markery lze definovat jako biochemické, fyziologické nebo anatomické charakteristiky, které jsou specifické pro určité podmínky. Příklady potenciálních biologických markerů schizofrenie zahrnují: zvětšené objemy postranních komor, ukazatele narušené neurochemie dopaminu a abnormální oční pohyby.

Pomocí zobrazování magnetickou rezonancí bylo prokázáno, že v mozku nemocných schizofrenií lze nalézt strukturální změny [16] a že kognitivní poruchy postihují až 73 % pacientů se schizofrenií [17]. Ve studii [18] se uvádí, že u mladých lidí se schizofrenií dochází k 7 % zmenšení objemu thalamu za rok. Bylo zjištěno, že zmenšení objemu těchto hlubokých struktur šedé hmoty lineárně koreluje se zhoršeným kognitivním výkonem. Ve studii [19] bylo uvedeno významné zmenšení šedé hmoty ve frontálních oblastech u dospívajících pacientů se schizofrenií v průběhu 3-5 let.

Hledání platných markerů u schizofrenie je omezeno neexistencí systému diagnostických kritérií, která by vedla k určení diagnózy široce uznávaných patofyziologií. Podrobný popis biologie schizofrenie by byl nesmírně cenný.

#### **1.4.2. Behaviorální indikátory**

Přestože velká část lidské komunikace je neverbální a velká část tohoto komunikačního kanálu je mimo vědomou kontrolu, současné diagnostické metody schizofrenie tyto dodatečné informace nevyužívají. Včetně mnoha markerů založených na řeči, o nichž bude podrobně pojednáno v kapitole 1.4, existuje celá řada behaviorálních indikátorů, které lze využít k rozlišení pacienta se schizofrenií od zdravých jedinců.

### **Tyto indikátory jsou:**

- halucinace. Mnoho pacientů slyší hlasy, které komentují jejich chování, přikazují jim, aby něco udělali, varují je před hrozícím nebezpečím nebo spolu mluví (obvykle o pacientovi).
- abnormální bludy. Mohou se u nich také objevit bludy o vlastní velikosti a přesvědčení, že jsou slavnými historickými postavami.
- zploštělý afekt, který se například projevuje strnulým výrazem obličeje.
- neohrabané, nekoordinované a mimovolné pohyby nebo zvláštní grimasy
- expresivní deficity vztahující se k spontánním emočním projevům, úsměvům a gestikulaci.
- snížená aktivita obličeje, zejména v horní části

## **1.5. Řeč jako objektivní indikátor**

V důsledku složitosti produkce řeči je řeč citlivý systém, ve kterém drobné fyziologické a kognitivní změny mohou potenciálně vyvolat výrazné akustické změny. Současné úspěchy v oblasti počítačové lingvistiky umožňují klinickým lékařům obrátit se na řečový výstup jako nový biomarker, který je nízkonákladový a časově efektivní. Porucha řeči je časným a výrazným projevem neurologické poruchy. Jelikož je známo, že schizofrenie vyvolává kognitivní a fyziologické změny, které ovlivňují proces produkce řeči a ovlivňují akustickou kvalitu produkované řeči způsobem, který je možné objektivně ohodnotit, má pro klinickou praxi cenný význam klasifikace poruch řeči.

Jedním z cílů této bakalářské práce je navrhnout řečové indikátory vhodné pro diagnostiku u schizofrenie a poskytnout poznatky o specifičnosti některých poruch pro dané onemocnění, které by mohly být užitečné při včasné diferenciaci diagnostice. Před tím, než se pustíme do specifikace jednotlivých akustických charakteristik, je nutné se seznámit s procesem tvorby řeči.

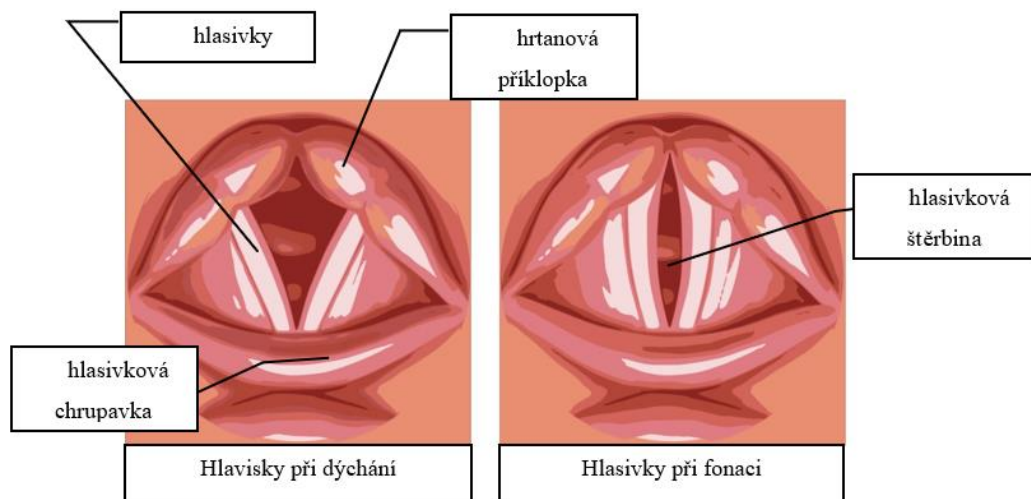
### **1.5.1. Produkce řeči**

Lidský hlas je periodickým zhušťováním a zředováním vzduchu. Fonační ústrojí se nachází v hrtanu, který se podílí na vytváření základního hlasu. Hrtan je široká, krátká trubice z chrupavek

a měkkých tkání. Hlasivky, dva hlasové valy pokryté sliznicí, jsou napjaty mezi chrupavkou štítnou a chrupavkami hlasivkovými. Délka hlasivek u žen je v průměru 18-20 mm, u mužů se pohybuje od 20 do 24 mm, přičemž čím jsou hlasivky kratší, tím rychleji kmitají a běžný hlas je vyšší.

Hlasivky jsou při fonaci zavřené (viz *Obrázek 1.5.1*). Podle myoelastické teorie [16] je vznik hlasu popsán takto: když jsou hlasivky přitahovacími svaly uvedeny do uzavřené fonační polohy, nastává koordinované výdechové úsilí. Vzduch v plicích, stlačený výdechovým úsilím, je hnán průdušnicí nahoru proti spodnímu okraji hlasivek. Jakmile se podhrtanový tlak zvýší natolik, že překoná uzavírací úsilí hlasivek, glottis se roztrhne, unikne obláček vzduchu, podhrtanový tlak se sníží. Díky své pružnosti a působení hrtanových svalů zužujících hlasovou štěrbinu se hlasivkové vazy vracejí do původní, tj. střední polohy, aby se pokračujícím tlakem vydechaného proudu vzduchu opět rozestoupily. Subglotický tlak se opět zvýší a celý cyklus se opakuje. Hlasivky vibrují čistě pasivně v proudu vanoucího vzduchu a jsou pouze udržovány ve fonační poloze přitahovacími svaly, které jsou aktivovány hrtanovými nervy. Při fonaci tedy dochází k rozkmitání hlasivek. Tyto vibrace jsou spíše příčné než podélné, tj. hlasivky se pohybují spíše dovnitř a ven než nahoru a dolů.

V důsledku vibrací hlasivek se pohyb vydechaného proudu vzduchu nad hlasivkami mění na kmitání částic vzduchu. Tyto vibrace se přenášejí do okolí a my je vnímáme jako hlasové zvuky.



**Obrázek 1.5.1.** Schéma změny polohy hlasivek při dýchání a fonaci.



## 2. Související práce a popis anomálií řeči pacientů se schizofrenií

---

Potíže s ovládním hlasu při vyjadřování emočních stavů nebo při označování relevantní informace mohou dramaticky snížit schopnost nemocných jedinců efektivně komunikovat ve společnosti. Hlasové atypičnosti byly charakteristickými rysy schizofrenie již od jejích prvních definic. Negativní projevy patří mezi základní diagnostická kritéria schizofrenie [21-23], a tak mohou představovat důležitý marker, který v průběhu času paralelně charakterizuje jak klinické vlastnosti, tak sociální kognitivní fungování jedinců se schizofrenií [12, 24].

Jedinci se schizofrenií vykazují atypické hlasové projevy, kvalitativně popsané jako chudost řeči (*alogie*), plochý afekt (*flat affect*), zvýšený počet pauz apod.[13, 25]. Nicméně se také uvádí studie, které neprokázaly rozdíl v akustických vlastnostech mezi pacienty se schizofrenií a zdravými jedinci [26,27]. Některé příznaky, jako je plochý nebo otupený afekt, anhedonie, jsou považovány za primární charakteristiky schizofrenního syndromu [28], tento "plochý" afektivní stav je přítomen nejméně u 66 % pacientů [29]. Rozdíl mezi plochým afektem (*flat affect*) a otupeným afektem (*blunted affect*) je v intenzitě redukce emocí. Zatímco plochý afekt znamená chybění emoční reaktivity, otupený afekt je vyjádřen jejím snížením. Pojem anhedonie popsal Theodule Ribot v roce 1896 jako neschopnost prožívat potěšení.

V této kapitole je uveden stručný přehled toho, jak jsou běžně používané prozodické, artikulační, časové a formantové charakteristiky řeči ovlivněny schizofrenií. Hlavní zdroje informací prezentovaných v této kapitole jsou [8, 24, 30, 31].

### 2.1. Prozodické vlastnosti

Prozodie byla dosud prozkoumána v menší míře než schopnost identifikovat a vyjádřit emoce obličejem. Prozodické vlastnosti představují dlouhodobé (na úrovni fonémů) změny ve vnímání rytmu a intonace řeči. Prozodie je melodická linie řeči, která vzniká změnami výšky tónu, rytmu a důrazu ve výslovnosti [32]; aprozodie je tedy neschopnost (nebo snížená schopnost) takový tón v řeči vytvořit. Mezi nejznámější příklady patří tempo řeči, výška (sluchové vnímání tónu), hlasitost a dynamika energie. V praxi jsou základní frekvence  $f_0$  a tempo řeči nejběžněji používané prozodické vlastnosti, protože pro ně existuje dost velký počet měřících metod.

Ve většině výzkumů řeči schizofreniků se odhalilo, že pacienti trvale vykazují prozodické abnormality řeči, jako je snížená hodnota základní frekvence  $f_0$ , snížený rozsah  $f_0$ , pomalejší tempo řeči a artikulační chyby. Každou z těchto abnormalit teď detailněji popíšeme a uvedeme výsledky předchozích studií.

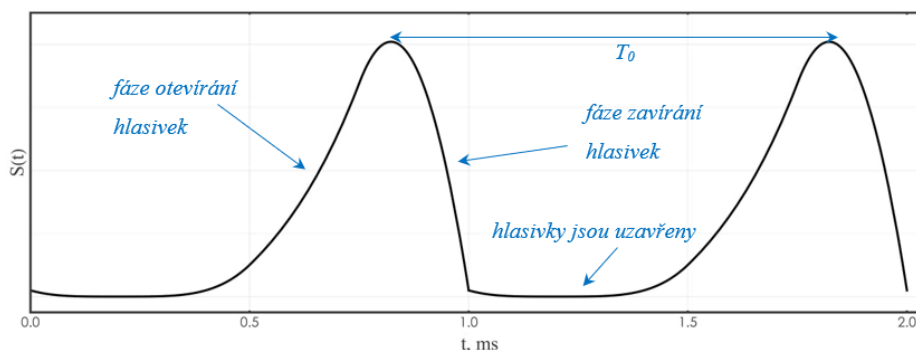
### 2.1.1. Základní hlasivková frekvence $f_0$

#### Definice základní frekvence $f_0$

Základní frekvence ( $f_0$ ), měřená v hertzech, je jednou z nejčastěji analyzovaných prozodických charakteristik řeči při rozpoznávání schizofrenie. Tato charakteristika je definována jako počet opakování zvukové vlny, která je vytvářena hlasivkami během určitého časového úseku, což odpovídá počtu cyklů otevření a zavření glottis.

Kdybychom považovali průchod proudu vzduchu hlasivkovou štěrbinou  $S(t)$  za periodický (viz *Obrázek 2.1.1*), dala by se tedy vyjádřit základní frekvence pomocí časové vzdálenosti dvou sousedních cyklů kmitání hlasivkové štěrby:

$$f_0 = \frac{1}{T_0} \text{ [Hz]}. \quad (2.1)$$



**Obrázek 2.1.1.** Schéma průchodu proudu vzduchu hlasivkovou štěrbinou.

Určení  $f_0$  z akustického signálu je z několika důvodů problematické [33] :

- $f_0$  se v řečovém signálu proměňuje v čase, protože kmitání hlasivek je kvaziperiodické. Čím rychleji mění svoji hodnotu, tím obtížnější je pak její správné určení;

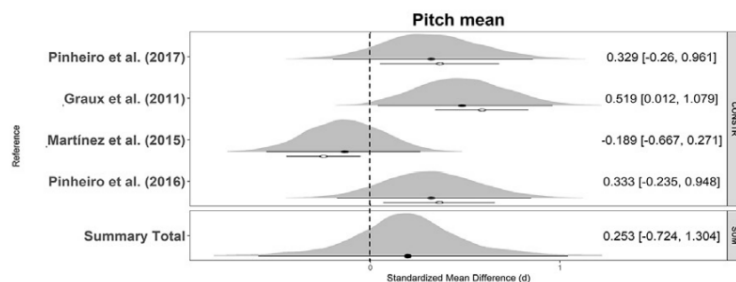
- Akustický signál není vždy znělý, takže každý algoritmus pro automatické určení  $f_0$  musí nejprve rozhodnout o tom, zda je interval znělý nebo neznělý;
- Řečový signál může ovlivnit šum, rozlišování znělých a neznělých intervalů tedy začíná být mnohem složitější úlohou;
- Vzhledem k rezonanci v dutinách, subharmonické násobky  $f_0$  mohou ovlivňovat proces určení hodnoty  $f_0$ .

Hodnoty základní frekvence však nejsou stacionární, existuje typické rozmezí hodnot této frekvence pro různá pohlaví a věkové kategorie. Jak jsme již uvedli v *Podkapitole 1.4.1*, mužské hlasivky jsou mohutnější a delší než hlasivky žen, takže základní frekvence mužského hlasu se obvykle pohybuje v rozmezí od 80 Hz do 240 Hz, zatímco u žen od 140 Hz do 500 Hz. Naopak hlasivky žen jsou delší a mohutnější než hlasivky dětí: základní výška dětského hlasu se proto pohybuje od 170 Hz do 600 Hz.

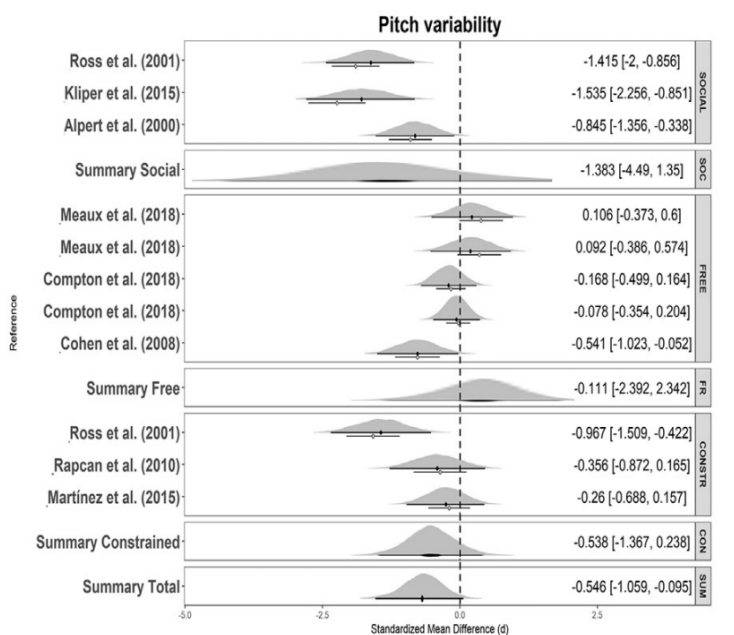
### **Výsledky rešerše literatury**

Nižší průměrná základní frekvence ( $f_0$ ) a snížená variabilita často souvisí s negativním příznakem plošného efektu (*flat effect*), který se projevuje u pacientů se schizofrenií. Ukazuje se, že poruchy  $f_0$  se zvyšují při patologii hrtanu, a mohou tedy indikovat poruchy hlasu [24]. Na *Obrázku 2.1.2* je uveden výsledek metaanalýzy odlišnosti průměrné hodnoty  $f_0$  pro kontrolní osoby a pacienty se schizofrenií, převzatý z [30]. V této studii bylo použito 46 článků, celkem 1254 pacientů se schizofrenií a 699 zdravých kontrol. Souhrnné velikosti účinků (*Hedgesův a Pearsonův r*) byly vypočteny pomocí víceúrovňového Bayesovského modelování. Grafy velikosti účinku (*Hedges'g*) pro všechna akustická měření budou uvedeny níže pro příslušný parametr. Na ose  $x$  jsou uvedeny velikosti efektů (černá tečka, kladné hodnoty znamenají, že jedinci se schizofrenií mají v daném akustickém měření vyšší hodnoty, zatímco záporné hodnoty naopak). Osa  $y$  označuje studie, u nichž byly uvedeny statistické odhady. Tečkovaná svislá čára označuje nulovou hypotézu (žádný rozdíl mezi skupinami). Studie jsou seskupeny podle řečového úkolu použitého ke sběru hlasových záznamů (*Constr* = omezené monologické, *Free* = volné monologické, *Social* = sociální interakce).

Navíc řeč pacientů se schizofrenií se vyznačuje monotónností (viz *Obrázek 2.1.2*), takže hodnoty variability výšky tónu  $\sigma_{f_0}$  pro nás budou také význačnými hodnotami. Studie [34] ukazují, že schizofrenní řeč se vyznačuje menší variabilitou výšky hlasu než normální řeč.



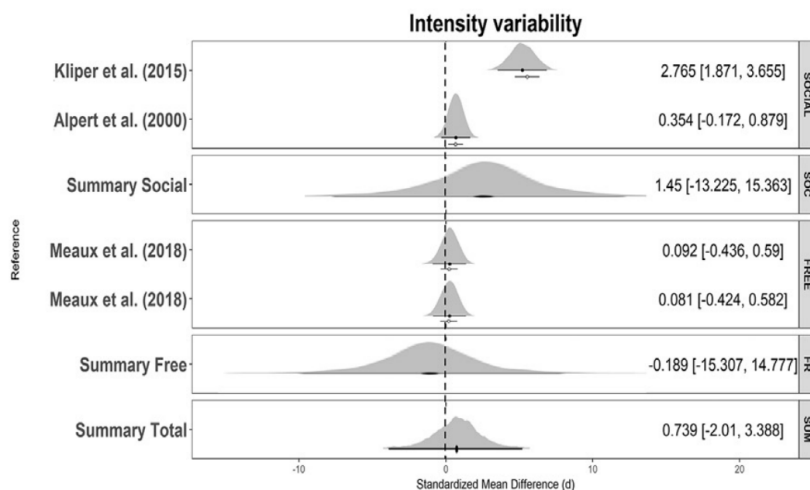
**Obrázek 2.1.2.** Metaanalýza výsledků rešerše literatury průměrné hodnoty  $f_0$ . Převzato z [30].



**Obrázek 2.1.2.** Metaanalýza výsledků rešerše literatury variability výšky tónu. Celkem 351 účastníků se schizofrenií a 232 zdravých kontrol. Převzato z [30].

## 2.1.2. Intenzita hlasu

Intenzita hlasu je intuitivním ukazatelem pro sdělování emocionálních informací v lidské komunikaci. Předchozí studie [35] vypočítávají rysy související s intenzitou na základě variability energie na sekundu nebo slabiku a výsledky experimentů ukazují, že intenzita hlasu pacientů se schizofrenií má menší variabilitu než u kontrolních osob [30].



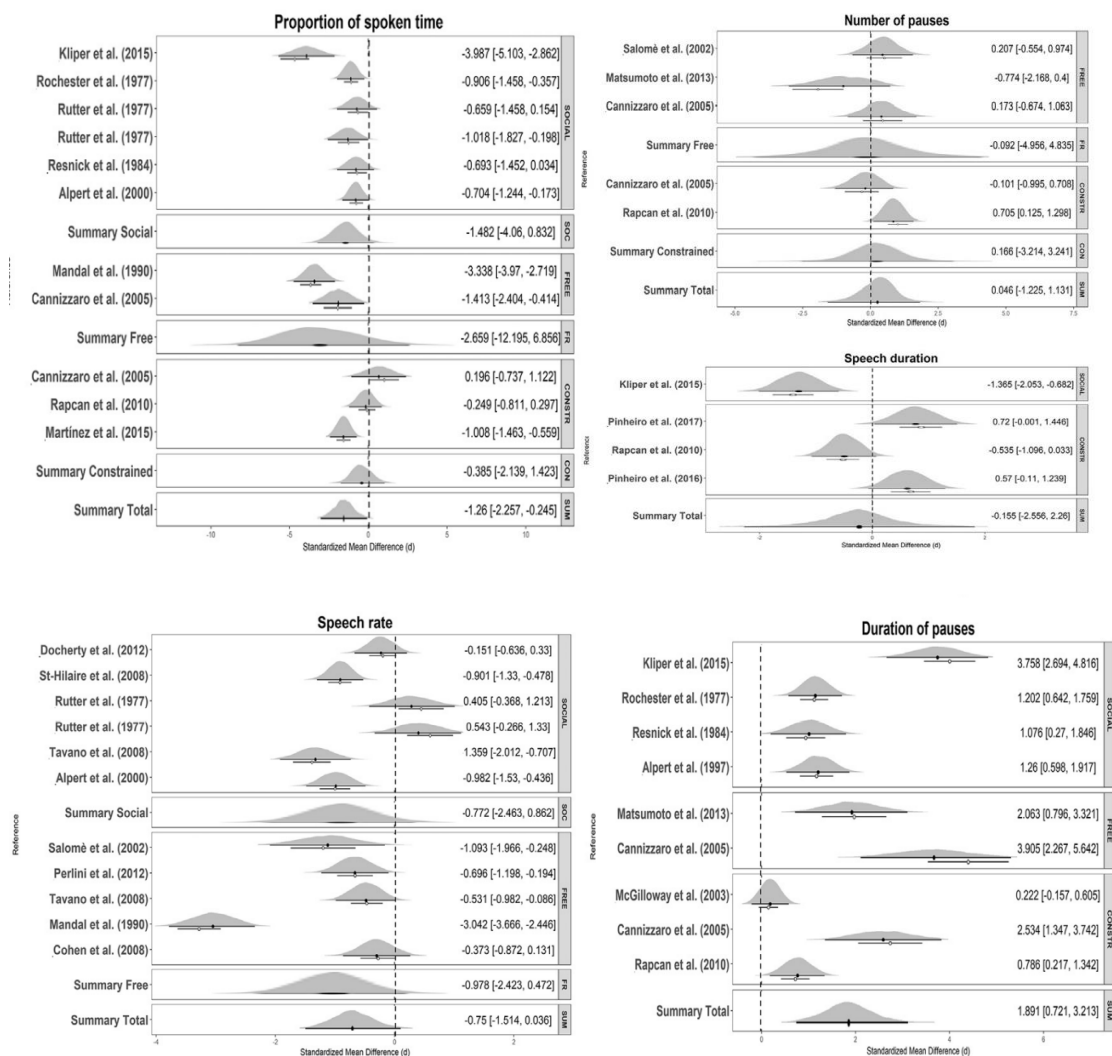
**Obrázek 2.1.3.** Metaanalýza výsledků rešerše literatury variability intenzity. Převzato z [30].

### 2.1.3. Artikulační tempo řeči

Téměř všechny typy neurologických dysartrií jsou charakterizovány poruchami řeči. Proto rychlost řeči představuje významnou roli v klinické diagnostice a v terapii. Rovněž snížená produktivita řeči je příznakem, který má zásadní význam pro studium negativního syndromu schizofrenie, protože jak sociální aktivita, tak mezilidské vyjadřování vyžadují alespoň minimální úroveň produktivity řeči. V důsledku toho jsou pacienti často hodnoceni jako asociální nebo s plochým afektem. Během literární rešerše vztahů mezi hodnocením negativního syndromu klinikem a charakteristikami řeči pacienta bylo stanoveno, že aspekty produktivity řeči, které nejsilněji souvisely s hodnocením alogie, byly vlastnosti časování; vlastnosti tempa řeči pacienta po odstranění doby pauz téměř nesouvisely [36].

Pacienti se schizofrenií s poruchou myšlení mají tendenci mít delší pauzy mezi jednotlivými větami než osoby bez poruchy myšlení [36]. Ve studii [36] porovnávali pauzy mezi větami u pacientů a kontrolní skupiny a došli k závěru, že prodloužené pauzy mezi větami se nezdají být jednoznačně spojeny s poruchami myšlení nebo poruchami řeči toho typu, který souvisí se schizofrenií.

Ve studii [38] bylo použito emočně neutrální, 2 až 3 minuty dlouhé čtecí úlohy k vyvolání řeči u zdravých kontrolních subjektů i u pacientů se schizofrenií. Pacienti se schizofrenií se významně lišili od kontrolní skupiny, pokud jde o celkový počet pauz, trvání pauz za sekundu a celkovou délku jejich pauz. Větší počet pauz také dobře koreloval s vyšším skóre negativních



**Obrázek 2.1.4.** Metaanalýza výsledků rešerši literatury artikulačních a časových charakteristik. Převzato z [30].

symptomů. Tyto výsledky naznačují, že délka jednotlivých pauz v řeči zkoumaných jedinců je hlavním faktorem, který přispívá k celkovým odlišnostem mezi skupinami pacientů a kontroly.

Rychlost řeči můžeme vyhodnotit jak manuálně (tento postup je výhodný tím, že má téměř 100 % přesnost, ale pro úlohy datové analýzy je časově náročný), tak i automaticky. Při automatickém hodnocení, které bude použité v této práci, je rychlost řeči definována jako počet fonetických jednotek (slabik nebo slov) vyslovených za časovou jednotku.

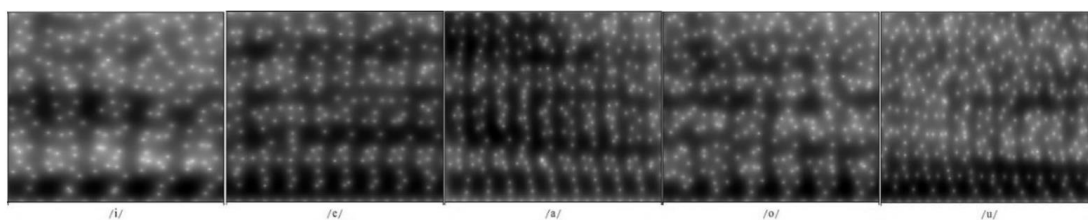
Podle definice *Roache* [40] se slabika "sestává z jádra, které umožňuje detekci slabik v řeči". Tato definice umožňuje věrohodný způsob detekce slabik v řeči [41].

Budeme rozlišovat pojmy *rychlost řeči* a *artikulační rychlost*. Rychlost řeči lze vyjádřit jako počet slabik za sekundu, čas zahrnuje i pauzy. Zatímco artikulační rychlost se vyjadřuje jako počet slabik za sekundu bez pauz. Algoritmus vyhodnocení těchto prozodických charakteristik bude uveden v *Podkapitole 3.3.4*.

## 2.2. Formanty

Artikulaci a další faktory podílející se na produkci řeči lze měřit v akustickém signálu, například rozkladem zvukových vln na formanty, tj. akustické rezonanční frekvence, které udávají polohu a pohyb artikulačních orgánů při mluvení.

Český hláskový inventář obsahuje 10 monoftongů a tři diftongy. Monoftongy se skládají z pěti různých samohláskových kvalit /a/, /ε/, /ɪ/, /o/ a /u/, které se vyskytují ve dvou formách – jako krátké a dlouhé. Dlouhé samohlásky jsou přibližně 1,7krát delší než jejich krátké protějšky. Na akustický výzkum formantových frekvencí v češtině se však zaměřilo jen několik předchozích studií [42].



**Obrázek 2.2.1.** Spektrogram českých vokálů *ɪ–ε–a–o–u*.

Kvalitu a srozumitelnost každé samohlásky lze určit především podle výrazného vrcholu akustické energie prvního ( $F_1$ ) a druhého ( $F_2$ ) formantu [43]. Formant je deskriptor, který odráží rezonanční frekvenci hlasového traktu.  $F_1$  označuje otevření čelistí/úst a výšku jazyka, přičemž frekvence  $F_1$  se mění nepřímo úměrně s výškou jazyka, zatímco  $F_2$  odpovídá poloze jazyka (vpředu/vzadu) a zaokrouhlení rtů a mění přímo s vysunutím jazyka [44].

Ve studii [31] se prokazuje, že rozsah druhého formantu  $F_2$  u schizofrenní řeči je menší než u kontrol. Snížené formantové frekvence zaznamenané u schizofrenní řeči jsou důkazem poklesu artikulačního úsilí s rostoucí úrovní onemocnění mluvčího.

## 3. Výběr metodiky

---

V této kapitole je představen způsob zpracování dat a následující postup analýzy jednotlivých akustických parametrů.

### 3.1. Databáze promluv

Ověření akustických parametrů bylo provedeno na databázi řečových signálů kontrolních osob a pacientů pořizovaných v NÚDZ v Klecanech, která vzniká od roku 2022. Sběr dat byl schválen Etickou komisí NÚDZ a každý účastník podepsal informovaný souhlas s účastí na výzkumu. Celkem byli zkoumáni 3 pacienti se schizofrenií (3 muži) a 3 kontrolní osoby (2 ženy a 1 muž). Kvůli uvedenému pohlavnímu zastoupení pacientů (PT) a kontrolních osob (KT) musejí být zkoumané akustické charakteristiky nezávislé na pohlaví. Mateřským jazykem všech osob byla čeština.

Nahrávání probíhalo v tiché místnosti, kde byli mluvčí zkoumáni na základě dále uvedeného protokolu pro akustickou analýzu. Při nahrávání byl použit kondenzátorový hlavový mikrofon Bayerdynamic Opus 55, zapojený přímo do kapesního záznamníku nastaveného do 16bitového režimu nekomprimovaných 48 kHz.

#### 3.1.1. Řečové nahrávky

Během rešerše literatury [31] bylo stanoveno, že různé typy požadavků na získání zvukových nahrávek řeči mohou ve skutečnosti způsobit určité rozdíly ve fonetických parametrech, ačkoli hodnoty jednotlivých parametrů spolu zpravidla vysoce korelují. Například ve stejné práci [31] se ukázalo, že snížená variabilita ve výšce tónu ( $f_0$ ) byla pozorovaná při čtení emočně zabarveného fragmentu a snížená variabilita  $F_2$  byla patrná i při emočně neutrálních úlohách čtení. Při budoucím vývoji počítačových lingvistických nástrojů se bude tedy patrně nutné opřít o několik způsobů získávání řeči, aby bylo možné plně zachytit různou podstatu řečových indikátorů schizofrenie.

V předchozí kapitole popsané akustické charakteristiky budou v tomto výzkumu zkoumány na řečových úlohách čtyř kategorií (detailnější popsání úloh je uvedeno v protokolu výzkumu (viz příloha C):



<b>Úloha 1</b>	Hlasité čtení emočně neutrálního textu (79 slov). Tento fragment byl vybrán speciálně pro svou slovní a sémantickou jednoduchost.
<b>Úloha 2</b>	Spontánní promluva – převyprávění pohádky o Popelce (cca 90s).
<b>Úloha 3</b>	Spontánní promluva – odpověď na emočně neutrální otázku “Popište svůj být/dům” (cca 90s).
<b>Úloha 4</b>	Spontánní promluva – vytvoření průběhu o obrázku, který vyvolává emoční reakci (cca 90s). Účastníkům byl ukázán obrázek, který pochází ze standardizovaného projektivního psychologického testu TAT (Tematický apercepční test) a jejich úkolem bylo co nejpodrobněji obraz popsat.

**Tabulka 3.1.1.** Přehled řečových úloh v této práci. Podrobněji viz v příloze C.

## 3.2. Návrh vlastního algoritmu

Všechny řečové signály byly před analýzou dat vizuálně překontrolovány. Na začátku a na konci každého záznamu byly odstraněny pauzy za účelem zajištění analýzy od okamžiku, kdy osoba začne mluvit, až do okamžiku, kdy osoba přestane mluvit (jinak by byl počet pauz chybně zvýšen o dvě, což by mělo za důsledek odpovídající nesprávné zvýšení délky trvání pauz).

### Pro každý soubor byly extrahovány následující vlastnosti:

1. Základní hlasivková frekvence v Hz a v semitónech	$f_0\_Hz$ , $f_0\_semi$
2. Rozsah výšky tónu v Hz a v semitónech	$f_0\_range\_Hz$ , $f_0\_range\_semi$
4. Průměrná hodnota výšky tónu v Hz a v semitónech	$f_0\_mean\_Hz$ , $f_0\_mean\_semi$
5. Variabilita výšky tónu v Hz a v semitónech	$\sigma_{f_0\_Hz}$ , $\sigma_{f_0\_semi}$
6. Průměrná hodnota intenzity hlasu	$I_{mean}$
7. Rozsah intenzity hlasu	$I_{range}$
8. Variabilita výšky tónu	$\sigma_{intens}$
9. Řečová rychlost	$v_{speech\_rate}$
10. Artikulační rychlost	$v_{articulation\_rate}$
11. Variabilita artikulační rychlosti	$\sigma_{articulation\_rate}$
12. Střední rychlost produkce pauz	$v_{pause\_rate}$
13. Střední doba trvání pauz	$t_{pause}$
14. Poměr doby artikulace k trvání nahrávky	$proportion\_of\_spoken\_time$
15. VSA (z anglického <i>vowel space area</i> )	VSA
16. VAI (z anglického <i>vowel articulation index</i> )	VAI
17. Relativní vzdálenost mezi druhými formanty samohlásek <i>i</i> a <i>u</i>	$F_{2i}/F_{2u}$

**Tabulka 3.2.1.** Seznam analyzovaných akustických charakteristik.

### 3.2.1. PRAAT a Parselmouth-praat

Datová analýza ve fonetických oborech se běžně opírá o PRAAT [45], rozsáhlý softwarový balík, který slouží fonetikům ke každodenní práci již více než dvě desetiletí. PRAAT pracuje s formáty .wav a mezi nejvhodnější funkce pro analýzu řeči patří spektrální analýza, analýza průběhu  $f_0$  a intenzity hlasu, tempa řeči a formantová analýza. PRAAT má také svůj vlastní skriptovací jazyk, ale bohužel je tento jazyk obecně málo zdokumentovaný a neintuitivní.

Parselmouth je open-source knihovna jazyka Python, která zpřístupňuje hlavní funkce jazyka PRAAT v jazyce Python. V této práci tedy můžeme využívat výhod mnoha nástrojů a utilit pro sofistikovanou akustickou analýzu, které poskytuje PRAAT, a na druhé straně využít schopnosti rozsáhlého jazyka Python. Funkce a metody Parselmouth mají stejné argumenty jako příkazy PRAAT, takže většina implementační části práce byla prováděna právě pomocí této knihovny.

### 3.2.2. MATLAB

Studium prozodických charakteristik v implementační části byl proveden v MATLAB [46]. MATLAB má řadu vestavěných funkcí pro numerickou analýzu a vizualizaci dat, což jej činí užitečným pro úlohy, jako je statistické modelování a analýza dat.

## 3.3. Analýza jednotlivých parametrů

Jak bylo uvedeno v předchozí kapitole, anomálie řeči při schizofrenii mohou souviset s několika aspekty, stručně charakterizujeme algoritmy použité pro analýzu jednotlivých charakteristik, včetně fonace, artikulace a prozodie.

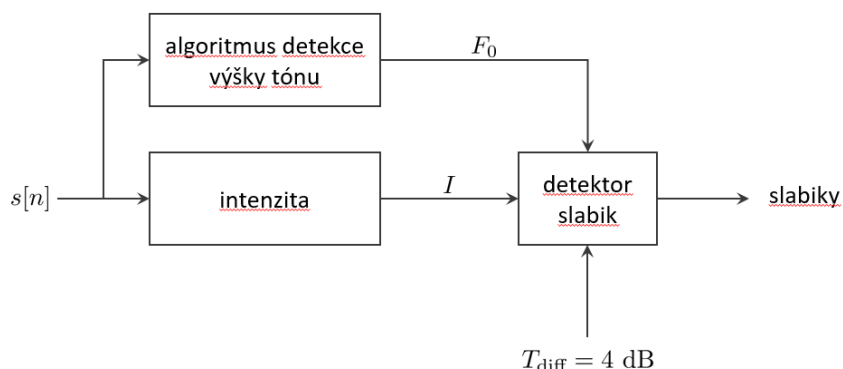
### 3.3.1. Řečová a artikulační rychlost

Skript PRAAT pro automatickou detekci jader slabik a měření rychlosti řeči byl vyvinut de Jongem a Wempem<sup>1</sup> a metoda byla detailně popsána v následujícím článku. [47].

---

<sup>1</sup> <https://osf.io/r8jau/?ref=499aefb361abec341bcebd133699270d3d66f0d5>

Krátce popíše me posloupnost jednotlivých kroků, které skript provede, aby na základě *intenzity* (dB) algoritmu detekce *výšky tónu* (Hz) vyhledal jádra slabik.



**Obrázek 3.3.1.** Principiální schéma metody PRAAT skript pro detekci jader slabik.

**Krok 1.** Extrahujeme intenzitu  $I[n]$ , právě z její trajektorie maxima nad mediánem intenzity naměřeným v celém zvukovém souboru považujeme za potenciální slabiky. Pro výpočet prahu používáme *medián*, nikoli průměr, abychom se vyhnuli zahrnutí extrémních maxim do výpočtu prahu.

**Krok 2.** Pokud rozdíl mezi maximem a mediánem intenzity  $T_{diff}$  celé promluvy nepřesahuje alespoň 4 dB, tento kandidát je zahozen.

**Krok 3.** Extrahujeme průběh  $f_0$  (s parametrem "*minimal pitch*" nastaveným na hodnotu  $f_{min} = 50\text{Hz}$ ), která je upravena pomocí parametru časového Gaussovského okna  $w[n]$  (délka časového okna  $L = 100$  ms) s časovým krokem 20 ms, a to pomocí autokorelace:

$$f_0[n] = \sum_{i=-L}^L s^2[n] w[n - i + 1], \quad (3.1)$$

kde  $s[n]$  je vstupní signál.

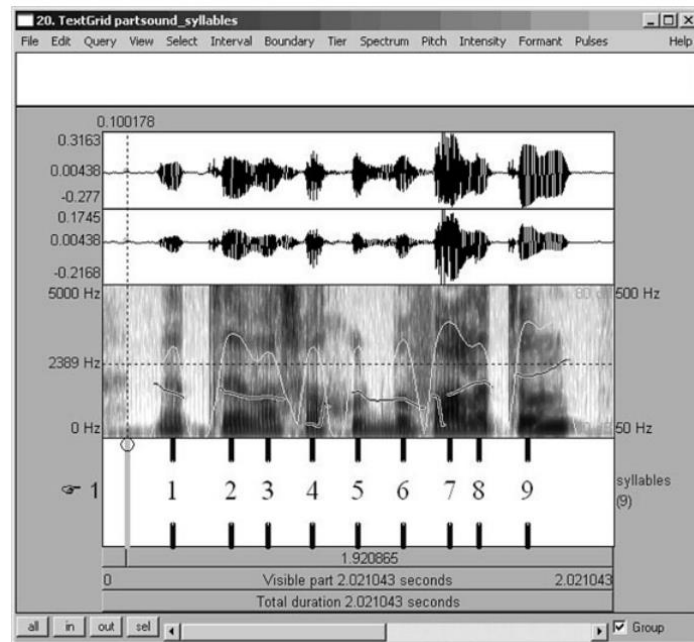
V Parselmouth-praat je již implementován výše popsany algoritmus, ze kterého zjistíme počet všech slabik  $N_{\text{syll}}$  a následně řečovou rychlost  $v_{\text{speech\_rate}}$  a artikulační rychlost (*articulation rate*)  $v_{\text{articulation\_rate}}$ .

Celkovou řečovou rychlost  $v_{\text{speech\_rate}}$ , která je ovlivněna celým obsahem řeči, tedy i počtem a délkou pauz, spočítáme jako počet všech slabik  $N_{\text{syll}}$  v promluvě, který vydělíme její časovou délkou  $t_{\text{speech}}$  v sekundách.

$$v_{\text{speech\_rate}} = \frac{N_{\text{syll}}}{t_{\text{speech}}} \text{ [slabik/s]}. \quad (3.2)$$

Pro analýzu řeči pacientů se schizofrenií budeme analyzovat i artikulační rychlost  $v_{\text{articulation\_rate}}$ , která je také závislá na počtu slabik v promluvě a která se liší od řečové rychlosti  $v_{\text{speech\_rate}}$  tím, že od délky promluvy  $t_{\text{speech}}$  odečteme časová trvání pauz  $t_{\text{pauses}}$ :

$$v_{\text{articulation\_rate}} = \frac{N_{\text{syll}}}{t_{\text{speech}} - t_{\text{pauses}}} \text{ [slabik/s]}. \quad (3.3)$$



**Obrázek 3.3.1.** Část řečového záznamu v programu PRAAT s intenzitou (bílá čára ve spektru) a výškou tónu (tmavá čára ve spektru). Body ve vrstevnici jsou slabiky jader, jak je detekoval skript. Převzato z [47].

### 3.3.2. Variabilita artikulační rychlosti

Jak bylo popsáno v podkapitole 2.1.1, řeč pacientů se schizofrenií se vyznačuje monotónností, tedy málo měněnou rychlostí v promluvě. V podkapitole 3.3.3 již byla uvedena charakteristika

variability výšky hlasu  $\sigma_{f0\_Hz}$ , kterou můžeme tento řečový indikátor zkoumat. Jiný přístup je přes variabilitu artikulační rychlosti  $\sigma_{articulation\_rate}$ , kterou vypočítáme jako *standardní odchylku průběhu artikulační rychlosti* v řeči  $v_{articulation\_rate}[n]$ , spočítanou po časových oknech délky 2 s ( $N$  je počet oken,  $n$  je index okna):

$$\sigma_{articulation\_rate} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (v_{articulation\_rate}[i] - \overline{v_{articulation\_rate}})^2}, \quad (3.4)$$

kde  $\overline{v_{articulation\_rate}}$  je střední hodnota artikulační rychlosti za celou promluvu.

Implementace časové segmentaci rychlosti byla provedena pomocí Parselmouth-praat, hodnota  $\sigma_{articulation\_rate}$  následně vypočítaná v MATLAB.

### 3.3.3. Střední rychlost produkce pauz, poměr trvání řeči k trvání nahrávky

Parselmouth-praat nám umožňuje provádět měření počtu pauz, jejich délky a poměru trvání řeči k trvání nahrávky pro každého účastníka. Použitý skript [47] je založen na tom, že doba trvání pauz se vypočítá odečtením doby fonace od celkové doby mluvení účastníka. Za minimální dobu trvání pauz bylo považováno 30 ms.

Kromě toho v této práci získáme také jednu z nejvíce patrných charakteristik řeči pacientů se schizofrenií, a to průměrnou délku trvání pauz  $t_{\text{pause}}$ , kterou dostaneme vydělením celkové doby trvání pauz  $t_{\text{pauses}}$  počtem pauz  $N_{\text{pauses}}$ :

$$t_{\text{pause}} = \frac{t_{\text{pauses}}}{N_{\text{pauses}}} \text{ [s]}. \quad (3.5)$$

Celkový počet pauz  $N_{\text{pauses}}$  během promluvy není objektivní charakteristika, protože nahrávky mají různou časovou délku, zavedeme tedy veličinu střední rychlost produkce pauz  $v_{\text{pause\_rate}}$ . Střední rychlost produkce pauz je definována jako poměr celkového počtu pauz  $N_{\text{pauses}}$ , získané pomocí Parselmouth-praat, a doby trvání nahrávky  $t_{\text{all}}$ :

$$v_{\text{pause\_rate}} = \frac{N_{\text{pauses}}}{t_{\text{all}}} \text{ [s}^{-1}\text{]}. \quad (3.6)$$

Doba trvání pauz  $t_{\text{pauses}}$  se měří v sekundách a podíl mluvené doby se vyjadřuje v procentech z celkové doby trvání nahrávky:

$$\text{proportion\_of\_spoken\_time} = \frac{t_{\text{pauses}}}{t_{\text{all}}} \cdot 100 \text{ [\%]}. \quad (3.7)$$

### 3.3.4. Základní frekvence $f_0$ v Hz a semitónech

Existuje několik způsobů odhadu základní frekvence, například cross – a autokorelace. V této práci jsme použili klasickou metodu auto-korelace prostřednictvím softwaru PRAAT, ve kterém byl vyhotoven kontur základní frekvence pro celou příslušnou promluvu. Časový průběh  $f_0$  byl následně extrahován do MATLAB a převeden z Hz do logaritmické tónové škály, tj. do semitónů. Tento krok byl uskutečněn v souvislosti s potřebou porovnání frekvenčních charakteristik pro muže a ženy, které jak bylo popsáno v *Podkapitole 2.1.1* mají různé standardní rozmezí výšky hlasu. Jak vyplývá z rešerše literatury [48], porovnání mužských a ženských profilů ukázalo, podobné hodnoty frekvenčního rozsahu v semitónech

Počet semitónů  $n$  v rozmezí frekvencí  $f_1$  a  $f_2$  vyjádříme z rovnice:

$$n = \frac{\log_2 \left( \frac{f_1}{f_2} \right)}{\log_2 \sqrt[12]{2}}, \quad (3.8)$$

kde se ve jmenovateli nachází logaritmus poměru dvou frekvencí, vyjadřující jeden semitón, který určíme vzhledem ke znalosti, že semitón je definován jako  $\frac{1}{12}$  oktávy, reprezentující frekvenční poměr 2.

Pro převádění základní frekvence z Hz ( $f_{0\_Hz}$ ) do semitónů ( $f_{0\_semi}$ ) zvolili jsme  $f_2 = 50$  Hz abychom dostávali jen kladné hodnoty frekvence v semitónech. Vzorec pro převod bude mít tedy tvar:

$$f_{0\_semi} = 12 \log_2 \left( \frac{f_{0\_Hz}}{50} \right) \text{ [ST]}. \quad (3.9)$$

Tradiční statistické funkce, jako je maximum  $f_{0\_max}$ , minimum  $f_{0\_min}$ , a průměrná hodnota  $f_{0\_mean}$  základní frekvence, byly vypočteny pro celou příslušnou promluvu a implementovány v MATLAB. Směrodatnou odchylku výšky hlasu, kterou také počítáme v Hz  $\sigma_{f_{0\_Hz}}$  a semitónech  $\sigma_{f_{0\_semi}}$  počítáme pomocí vzorců:

$$\sigma_{f0\_Hz} = \sqrt{\frac{1}{n} \sum_{i=1}^n (f0\_mean\_Hz - f0\_Hz[i])^2}. \quad (3.10)$$

$$\sigma_{f0\_semi} = 12 \log_2 \left( \frac{\sigma_{f0\_Hz}}{50} \right). \quad (3.11)$$

### 3.3.5. Rozsah intenzity hlasu

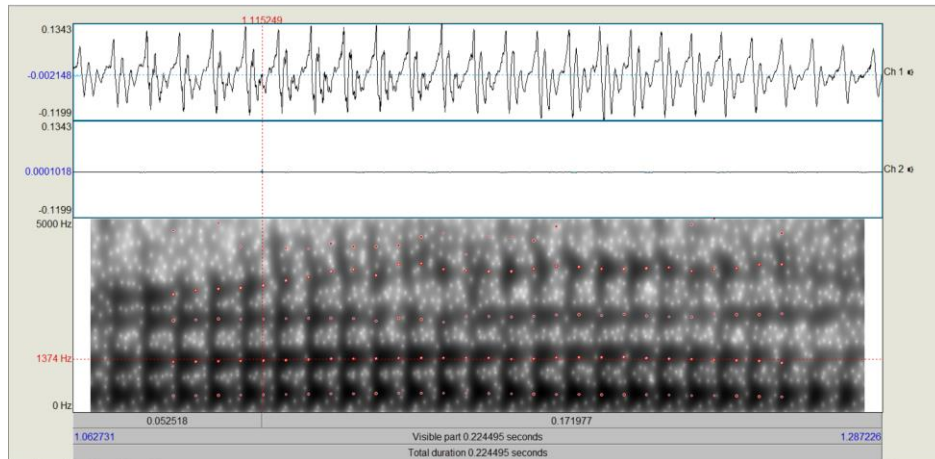
Analýza charakteristik intenzity, a to hodnot rozsahu  $I_{range}$ , průměrné hodnoty  $I_{mean}$  a variability intenzity  $\sigma_{intens}$ , byly vypočteny podobně jako hodnoty  $f_0$ . Hodnota rozsahu intenzity je pro naše cíle důležitá v souvislosti s tím, že ji lze použít k hodnocení monotónnosti řeči u pacientů se schizofrenií.

Časový průběh  $I[n]$  byl získán v PRAAT a následně extrahován do MATLAB. Pro odstranění úseků pauz a šumu při zpracování byly akceptovány takové hodnoty, kdy rozdíl mezi maximem a mediánem intenzity  $T_{diff}$  celé promluvy přesahoval alespoň 4 dB.

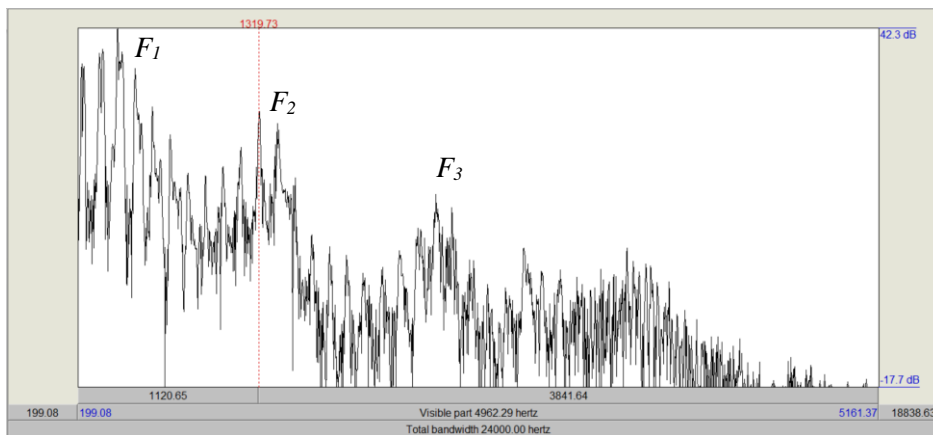
### 3.3.6. Analýza formantů F1 a F2

Kvalitu a srozumitelnost každé samohlásky lze určit především podle hodnot prvního ( $F_1$ ) a druhého ( $F_2$ ) formantu. V podkapitole 2.2 jsme uvedli, že u pacientů se schizofrenií je však ovlivněn i další typ akustické charakteristiky: snížená variabilita formantu  $F_2$ . Variability formantu  $F_1$  a jiných formantových charakteristik odlišnosti mezi výsledky pro PT a KT na základě již známých studií nebyly získány.

Hodnocení artikulace samohlásek provedeme ruční analýzou formantových frekvencí v PRAAT. Pro řečové záznamy úlohy čtení (*úloha I*) bylo extrahováno ze stejných slov (viz přílohu C, kde jsou podtrženy zkoumané samohlásky) a následně analyzováno celkem 30 samohlásek – 10 *a*, 10 *i*, 10 *u*, ze kterých jsou na základě frekvenčního spektra určeny  $F_1$  a  $F_2$  (viz Obrázek 3.3.3). Byl zvolen přístup přes frekvenční spektrum (viz Obrázek 3.3.2) místo odhadnutí ze spektrogramu vzhledem k tomu, že ze spektrogramu nelze přesně určit hodnotu. Špičky odpovídající rezonančním frekvencím jsou ve frekvenčním spektru zřetelnější, proto je tento algoritmus přesnější. Samohláska musí být alespoň 30 ms dlouhá a vyslovena samostatně nebo po neznělé souhlásce. Následně byly pro každou osobu hodnoty formantů zprůměrovány.



**Obrázek 3.3.2.** Spektrogram vybrané části signálu (/a/ ve slově ‚sazenička‘) s vyznačenými formantovými frekvencemi.



**Obrázek 3.3.3.** Frekvenční spektrum vybrané části signálu (/a/ ve slově ‚sazenička‘) s vyznačenými formantovými frekvencemi.

Kromě průměrných hodnot formantu  $F_1$  a  $F_2$  rohových samohlásek /a/, /u/, /i/ a jejich variability bude zjištěna velikost oblasti samohláskového prostoru VSA (z anglického *vowel space area*), která je konstruována euklidovskými vzdálenostmi mezi souřadnicemi  $F_1$  a  $F_2$  rohových samohlásek /a/, /i/ a /u/ v trojúhelníkovém samohláskovém prostoru  $F_1 - F_2$  [50] v ve srovnání s řečí kontrolních osob. Hodnota VSA je vyjádřena v  $\text{Hz}^2$  a lze ji snadno vypočítat podle následujícího vzorce [51] :

$$\text{VSA} = \frac{1}{2} |F_{1i}(F_{2a} - F_{2u}) + F_{1a}(F_{2u} - F_{2i}) + F_{1u}(F_{2i} - F_{2a})| [\text{Hz}^2]. \quad (3.12)$$



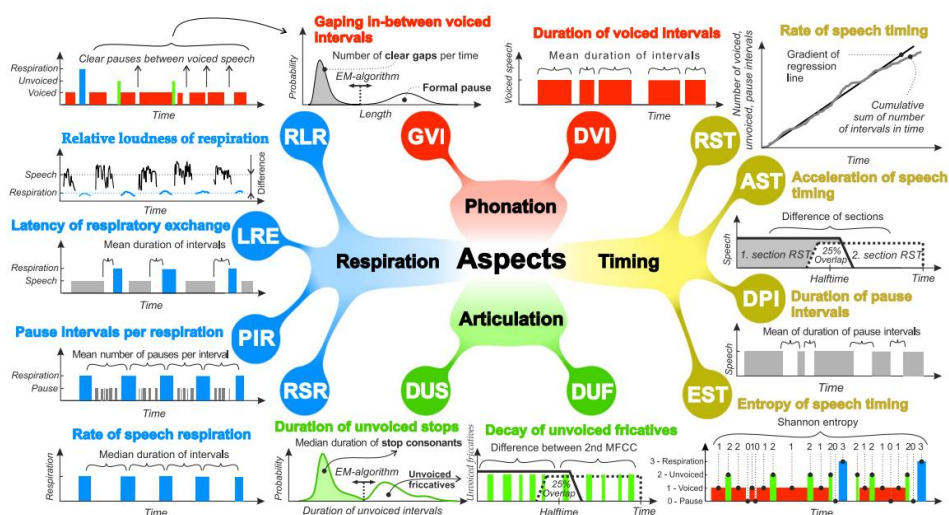
Současná studie se dále zabývá otázkou, zda jsou některá měření založená na formantech citlivější než jiná měření při zachycování poruch artikulace samohlásek [43]. Kromě hodnoty VSA v této práci provedeme porovnání hodnoty samohláskového artikulčního indexu VAI (z anglického *vowel articulation index*), který odráží centralizaci samohlásek, a poměru  $F_{2i}/F_{2u}$  vyjadřující vzdálenost mezi rohovými samohláskami /i/ a /u/. Pro hodnotu VAI platí vztah[51] :

$$VAI = \frac{F_{1a} + F_{2i}}{F_{1i} + F_{1u} + F_{2a} + F_{2u}} . \quad (3.13)$$

### 3.4. Parametry získané automaticky v toolboxu DYSAN

Při hodnocení časových, artikulčních, fonačních a dechových charakteristik řeči byla použita sada 12 charakteristik vypočítaných na základě segmentace časových intervalů pomocí toolboxu DYSAN [52]. Tato disertační práce zkoumá rozsáhlou databázi pacientů postižených neurodegenerací a představuje metodiku pro získávání vysoce interpretovatelných řečových charakteristik s využitím nového přístupu v oblasti digitálního zpracování signálu, strojového učení a modelování pohlavního dimorfismu a závislosti na věku.

V této práci se pokusíme použít toolbox na naši databázi a provést diskuzi ohledně klinické použitelnosti pro účely řečové diagnostiky schizofrenie. Popis níže uvedených charakteristik je převzat z původní publikace Hlavnička *et al* [53] .



**Obrázek 3.4.1.** Myšlenková mapa znázorňující základní principy jednotlivých akustických vlastností řeči. Převzato v [53] .

### 3.4.1. Prozodické charakteristiky

- **Směrodatná odchylka základní frekvence ( $stdF0$ )**

Charakteristika  $stdF0$  byla implementována jako *směrodatná odchylka zjištěného  $f_0$*  v semitónech.

- **Směrodatná odchylka výkonové obálky ( $stdPWR$ )**

Signál byl decimován na 8 kHz, pak filtrován klouzavým průměrem o délce 20 milisekund a vyjádřen pomocí logaritmické stupnice. Charakteristika  $stdPWR$  byla stanovena jako *směrodatná odchylka výsledné výkonové obálky* vypočtená na všech znělých intervalech.

### 3.4.2. Časové charakteristiky

- **Rychlost průběhu řeči (RST)**

RST poskytuje lepší odhad tempa řeči než prosté měření délek pauz, neboť bere v úvahu nejen pauzu, ale i znělé a neznělé intervaly. RST tedy aproximuje rychlost řeči komplexně, protože řečové poruchy rychlosti souvisí s deficitem ve všech aspektech řeči. Znělé intervaly poskytují dodatečnou informaci o narušení fonační regulace, zatímco neznělé intervaly o nepravidelné artikulaci. Každý interval je popsán dobou trvání, která byla určena jako průměrná doba mezi začátkem a koncem řečového intervalu. RST může sloužit jako náhradní měření slabikové rychlosti, kdy tato rychlost zahrnuje znělé, neznělé intervaly a pauzy. RST se počítá jako *sklon regresní přímky v časovém průběhu*.

- **Zrychlení průběhu řeči (AST)**

Akcelerace řeči může být odhadnuta pomocí RST. Pro výpočet AST se průběh řeči rozdělí na dvě poloviny s 25 % překryvem, kde AST se stanoví jako *rozdíl jednotlivých RST dělený celkovou dobou trvání řečového záznamu*.

- **Doba trvání pauz (DPI)**

DPI se počítá jako medián pauz. Pauzy zahrnují všechny intervaly delší než 30 milisekund, které nejsou hlasovou promluvou nebo neznělou řečí. Dýchání je podmnožinou pauz.

- **Entropie časového průběhu řeči (EST)**

EST popisuje uspořádanost řeči včetně pauz  $np$ , znělých  $nv$ , neznělých  $nu$  a dechových intervalů  $nr$  následovně:

$$EST = -\frac{nv}{nt} \cdot \log_2 \left( \frac{nv}{nt} \right) - \frac{nu}{nt} \cdot \log_2 \left( \frac{nu}{nt} \right) - \frac{np}{nt} \cdot \log_2 \left( \frac{np}{nt} \right) - \frac{nr}{nt}, \quad (3.14)$$

kde  $nt$  je celkový počet intervalů a snížená entropie EST se pak projevuje ve zhoršené kvalitě řeči.

### 3.4.3. Artikulační charakteristiky

- **Doba trvání neznělých exploziv (DUS)**

DUS se počítá jako medián s okluzí explozivních souhlásek. Hodnotí nepřesnou artikulaci, která se projevuje zvýšeným šumem doprovázejícím explozivy.

- **Útlum neznělých frikativ (DUF)**

DUF měří kvalitu artikulace. Neznělé frikativy se vyznačují energií koncentrovanou na vysokých frekvencích. Informace o kvalitě artikulace se měří jako časový útlum vysokých frekvencí. Neznělé frikativy jsou parametrizovány pomocí *24 mel-frekvenčních cepstrálních koeficientů* (MFCC) a z nichž je počítán DUF.

### 3.4.4. Fonační charakteristiky

- **Mezery mezi znělými intervaly (GVI)**

GVI hodnotí schopnosti mluvčího rozdělit znělé úseky pomocí pauz. Při poruchách fonace dochází ke slučování jednotlivých slabik nebo slov. Významné *pauzy a přestávky* ve znělé řeči jsou detekované EM-algorem a parametr GVI pak hodnotí jejich *počet v řečové promluvě*.

- **Doba trvání znělých intervalů (DVI)**

DVI hodnotí fonaci v prodloužených znělých intervalech a počítá se jako *průměrná délka znělých intervalů*.

### 3.4.5. Respirační charakteristiky

- **Respirační výdechová frekvence (RSR)**

RSR odhaduje *dechovou frekvenci* z detekovaných intervalů dýchání a vyjadřuje ji jako *počet nádechů za minutu*.

- **Respirační pauzy (PIR)**

PIR se počítá jako *medián pauz mezi detekovanými dechovými intervaly*.

- **Relativní úroveň hlasitosti respirace (RLR)**

RLR hodnotí *rozdíl výkonových obálek respiračního šumu a řeči*, tudíž je tento parametr nezávislý na zesílení mikrofonu.

- **Latence mezi nádechem a výdechem (LRE)**

LRE hodnotí dobu mezi výdechem a nádechem, který se při respiračních obtížích se prodlužuje a je počítán jako *průměrná hodnota doby latence všech detekovaných nádechů a výdechů*.

## 4. Výsledky

---

V této kapitole se budeme věnovat porovnání výsledků analýzy prozodických, časových, artikulačních a fonačních charakteristik pro kontrolní osoby (KT) a pacienty se schizofrenií (PT) dle uvedené metodiky v Kapitole 3. Každou charakteristiku počítáme pro všechny úlohy popsané v podkapitole 3.1.1 pro rozpoznání nejvíce významných příznaků v různých typech záznamů.

Statistickým chybám bylo zabráněno použitím vhodné metody analýzy výsledků a také definicí příslušné nulové hypotézy. Statistické hodnocení jednotlivých rozdílů mezi skupinami KT a PT bude provedeno za předpokladu, že tyto skupiny jsou nezávislými náhodnými vektory podléhajícími normálnímu rozdělení s různými středními hodnotami.

Nulovou hypotézu  $H_0$  definujeme za předpokladu, že i střední hodnoty  $\mu_{KT}$  pro kontrolní osoby a  $\mu_{PT}$  pro pacienty se rovnají:

$$H_0 : \mu_{KT} = \mu_{PT} \quad (4.1)$$

V případě zamítnutí  $H_0$  bude zvolená alternativní hypotéza  $H_1$  :

$$H_1 : \mu_{KT} \neq \mu_{PT} . \quad (4.2)$$

Testování statistiky provedeme pomocí *nepárového dvojnýběrového t-testu*. *Nepárový dvojnýběrový t-test* je příkladem parametrického testu, který funguje na základě dat s normálním rozdělením a porovnává jejich průměrné hodnoty a rozptyl. *P-value* neboli nejmenší hladina významnosti, při které ještě zamítneme  $H_0$ , je statistická hodnota vyjádřená číslem mezi 0 a 1, která se používá k testování hypotézy. Tato hodnota se používá k určení, zda je výsledek získaný v rámci klinické studie náhodný. Jako statisticky významný (a umožňující zamítnutí nulové hypotézy) je definován výsledek, jehož p-hodnota je rovna nebo menší než hladina významnosti. V této práci se označí jako  $p \leq 0.05$ .

### 4.1. Prozodické charakteristiky

V tabulkách (viz *Tabulka 4.1.1* a *Tabulka 4.1.2*) vidíme průměrné hodnoty prozodických charakteristik pro skupiny kontrolních osob (KT) a pacientů se schizofrenií (PT).

**Úloha 1. Hlasité čtení****Úloha 2. Převyprávění pohádky**

	<b>KT</b>	<b>PT</b>	<b>KT</b>	<b>PT</b>
$f_{0\_range\_Hz}$	397.70	409.20	423.53	209.84
$f_{0\_mean\_Hz}$	173.51	117.70	166.25	107.61
$\sigma_{f0\_Hz}$	42.33	36.40	39.87	15.51
$f_{0\_range\_semi}$	28.94	32.26	32.72	20.15
$f_{0\_mean\_semi}$	20.86	14.20	20.08	12.84
$\sigma_{f0\_semi}$	3.27	3.07	3.60	1.90
$I_{range}$	21.15	16.67	25.39	22.36
$I_{mean}$	55.79	59.33	51.93	56.12
$\sigma_{intens}$	4.21	3.85	4.71	4.95

**Tabulka 4.1.1.** Hodnoty prozodických charakteristik řeči pro skupiny KT a PT pro úlohu 1.

**Úloha 3. Odpověď na otázku****Úloha 4. Popis obrázku**

	<b>KT</b>	<b>PT</b>	<b>KT</b>	<b>PT</b>
$f_{0\_range\_Hz}$	419.06	208.06	422.49	219.90
$f_{0\_mean\_Hz}$	161.48	108.14	164.52	106.52
$\sigma_{f0\_Hz}$	39.48	18.86	34.49	15.66
$f_{0\_range\_semi}$	32.55	19.97	32.59	20.04
$f_{0\_mean\_semi}$	19.59	12.91	20.01	12.68
$\sigma_{f0\_semi}$	3.73	2.05	3.06	1.78
$I_{range}$	26.37	19.03	23.36	25.14
$I_{mean}$	51.82	57.33	54.14	53.85
$\sigma_{intens}$	4.69	3.89	4.60	6.02

**Tabulka 4.1.2.** Hodnoty prozodických charakteristik řeči pro skupiny KT a PT pro úlohy 2,3,4.

## 4.2. Časové a artikulační charakteristiky

Tabulky 4.2.1 zobrazují hodnoty časových charakteristik, spojených s parametry pauz a hodnotami fonační doby, a také artikulačních charakteristik, ve kterých jsou pro nás nejvíce patrné řečová a artikulační rychlost.

**Úloha 1. Hlasité čtení****Úloha 2. Převyprávění pohádky**

	<b>KT</b>	<b>PT</b>	<b>KT</b>	<b>PT</b>
$t_{all}$	36.71	40.37	207.76	106.89
$N_{syll}$	151	151.00	667.67	304.67
$v_{pause\_rate}$	0.98	0.91	1.08	0.86
$t_{speech}$	28.96	32.22	162.04	71.96
$v_{speech\_rate}$	4.14	3.87	3.25	2.80
$v_{articulation\_rate}$	5.25	4.76	4.00	4.23
$\sigma_{articulation\_rate}$	0.75	0.46	1.017	1.002
$t_{pause}$	0.22	0.21	0.19	0.43
<i>proportion_of_spoken_time</i>	0.79	0.80	0.82	0.66

**Tabulky 4.2.1.** Hodnoty artikulačních a časových charakteristik pro skupiny KT a PT pro úlohu 1.**Úloha 3. Odpověď na otázku****Úloha 4. Popis obrázku**

	<b>KT</b>	<b>PT</b>	<b>KT</b>	<b>PT</b>
$t_{all}$	116.83	109.95	136.43	108.83
$N_{syll}$	396.00	342.00	469.67	262.00
$v_{pause\_rate}$	1.08	0.77	0.90	0.58
$t_{speech}$	91.06	80.51	101.45	64.50
$v_{speech\_rate}$	3.39	3.09	3.41	2.41
$v_{articulation\_rate}$	4.37	4.27	4.63	4.09
$\sigma_{articulation\_rate}$	1.025	0.940	1.127	0.967
$t_{pause}$	0.20	0.35	0.29	0.71
<i>proportion_of_spoken_time</i>	0.78	0.73	0.74	0.60

**Tabulky 4.2.2.** Hodnoty artikulačních a časových charakteristik pro úlohy 2,3,4.**4.3. Fonační charakteristiky**

Tabulky (viz *Tabulky 4.3.1*) a diagramy (viz *Obrázek 4.3.2*) ukazují průměrné hodnoty a rozptyl (chybové úsečky) formantů  $F_1$  a  $F_2$  mezi skupinami KT a PT v úloze 1 (hlasité čtení emočně neutrálního textu). U měření jednotlivých formantů se mezi mluvčími KT a PT

nejvýrazněji lišily  $F_{1a}$ ,  $F_{2a}$  a  $F_{2i}$ . Bohužel však musíme zdůraznit, že hodnoty formantů jsou závislé na pohlaví jednotlivců, o čemž ještě bude diskutováno v následující kapitole.

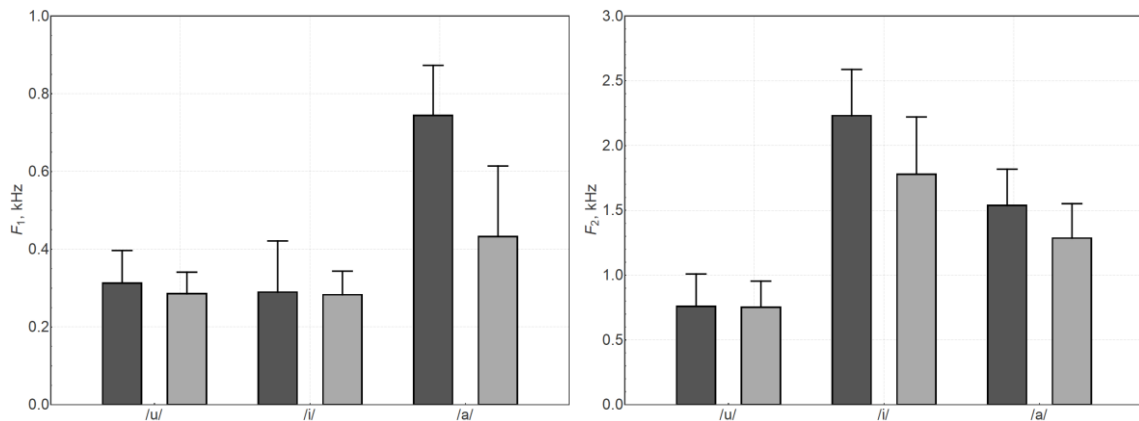
### Formant $F_1$

	KT	PT
$F_{1a}$	808.58	523.58
$F_{1u}$	355.00	313.33
$F_{1i}$	355.58	313.42

### Formant $F_2$

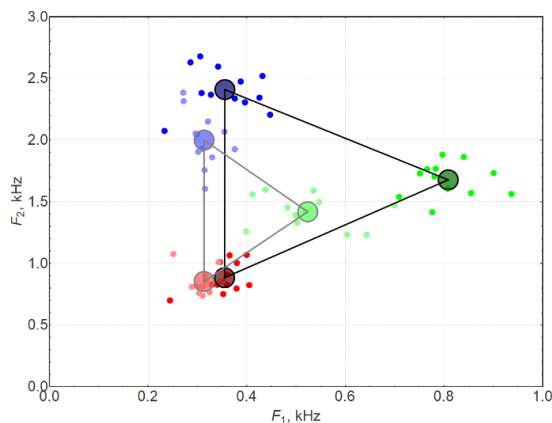
	KT	PT
$F_{2a}$	1677.75	1418.92
$F_{2u}$	884.25	853.75
$F_{2i}$	2409.33	2000.17

**Tabulky 4.3.1.** Hodnoty formantů  $F_1$  a  $F_2$  pro skupiny KT a PT pro různé řečové úlohy.



**Obrázek 4.3.2.** Porovnání průměrných hodnot a rozptylu  $F_1$  a  $F_2$  pro KT a PT pro úlohu 1.

Vzhledem ke sníženému samohláskovému prostoru byly významné rozdíly mezi skupinami PT a KT zachycené pomocí VSA (Obrázek 4.3.3). Co se týče samohláskové centralizace, byly zjištěny rozdíly mezi oběma skupinami, hodnota VAI je menší pro PT. V souladu s naší hypotézou lze tedy nepřesnou artikulaci samohlásek považovat za časný marker pro detekci schizofrenie.



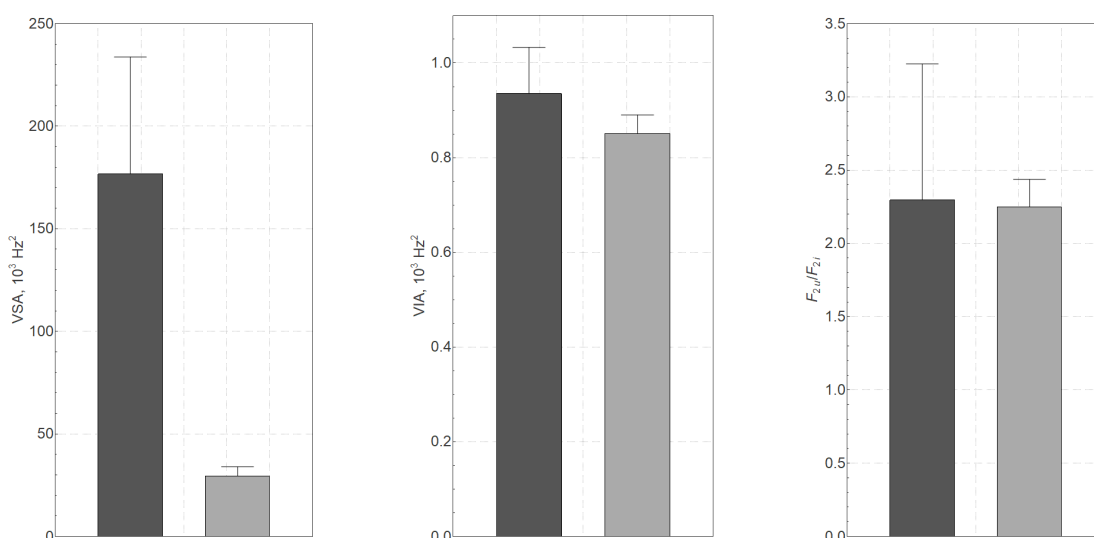
**Obrázek 4.3.3.** Centralizace rohových samohlásek u PT ve srovnání se zdravými jedinci.



### VSA, VAI, $F_{2i}/F_{2u}$

	<b>KT</b>	<b>PT</b>
VSA	345643	120492
VAI	0.98	0.87
$F_{2i}/F_{2u}$	2.72	2.34

**Obrázek 4.3.4.** Porovnání průměrných hodnot a rozptylu VSA, VAI a  $F_{2u}/F_{2i}$  pro KT a PT.



**Obrázek 4.3.5.** Porovnání průměrných hodnot a rozptylu VSA, VAI a  $F_{2u}/F_{2i}$  pro KT a PT.

## 4.4. Parametry získané automaticky v toolboxu DYSAN

Navržený algoritmus [52] je použitelný pro hodnocení komplexních hlasových deficitů a bude pravděpodobně schopen sledovat průběh řeči s dobou léčby ve všech řečových charakteristikách včetně časování, artikulace, fonace a respirace. Výsledky studie skutečně naznačují, že většina řečových charakteristik byla ovlivněna typem mluvního úkolu a všechny úkoly mohou poskytnout užitečné informace. Většina charakteristik s výjimkou EST, *stdPWR*, RSR vykazovaly rozdíly mezi pacienty se schizofrenií a kontrolními osobami (viz *Tabulka 4.4.1* a *Tabulka 4.4.2*).

### Úloha 1.

	<b>KT</b>	<b>PT</b>
EST	1.53	1.53
RST	450.85	445.85
AST	22.04	28.21
DPI	172.37	170.79
DVI	177.24	196.50
GVI	46.48	57.91
DUS	23.88	26.88
DUF	-0.19	3.38
RLR	-23.56	-25.29
PIR	4.67	5.67
RSR	20.77	17.19
LRE	159.15	120.13
<i>stdPWR</i>	3.61	3.72
<i>stdF0</i>	1.91	1.82

**Tabulky 4.4.1.** Hodnoty charakteristik řeči z toolboxu DYSAN pro úlohu 1.

### Úloha 2

### Úloha 3

### Úloha 4

	<b>KT</b>	<b>PT</b>	<b>KT</b>	<b>PT</b>	<b>KT</b>	<b>PT</b>
EST	1.55	1.52	1.54	1.54	1.53	1.53
RST	380.90	275.90	367.60	307.87	316.43	142.48
DPI	194.91	260.80	215.09	228.81	231.43	435.47
DVI	234.68	323.55	235.57	280.42	283.36	550.06
GVI	41.69	29.89	39.64	40.13	31.95	14.20
DUS	20.88	29.87	27.62	32.87	25.38	32.87
RLR	-18.84	-26.96	-18.94	-27.92	-24.65	-27.68
PIR	4.00	3.00	4.00	4.00	3.67	1.33
RSR	21.78	17.76	22.91	17.51	23.12	18.95
LRE	162.34	496.95	156.45	388.15	298.47	800.07
<i>stdPWR</i>	5.31	4.67	4.26	3.32	5.14	5.37
<i>stdF0</i>	1.88	1.41	1.30	1.46	1.46	1.32

**Tabulky 4.4.2.** Hodnoty charakteristik řeči z toolboxu DYSAN pro úlohy 2,3,4.

---

**Zkratky:** EST = entropie časového průběhu řeči, RST = rychlost průběhu řeči, AST = zrychlení průběhu řeči, DPI = doba trvání pauz, DVI = doba trvání znělých intervalů, GVI = mezery mezi znělými intervaly, DUS = doba trvání neznělých exploziv, DUF = útlum neznělých frikativ, RLR = relativní úroveň hlasitosti respirace, PIR = Respirační pauzy, RSR = respirační výdechová frekvence, LRE = latence mezi nádechem a výdechem, *stdPWR* = směrodatná odchylka výkonové obálky, *stdF0* = směrodatná odchylka základní frekvence

## 5. Diskuze

---

V této kapitole se zaměříme na porovnání a vysvětlení výsledků uvedených v předchozí kapitole. Jako první začneme prozodickými charakteristikami spojenými se základní frekvencí  $f_0$  a intenzitou hlasu  $I$ , jejichž výsledky jsou uvedeny v podkapitole 4.1.

### Prozodické charakteristiky

Protože při analýzách byly kontrolními osoby 2 ženy a 1 muž, zatímco všichni nemocní byli muži, nelze porovnávat hodnoty průměrné základní frekvence v Hz, protože tato prozodická charakteristika je silně ovlivněna odlišnou artikulací různých pohlaví a je nutné převést  $f_0$  z Hz do semitónů [54].

Vyjdeme-li z průměrné hodnoty  $f_{0\_mean\_semi}$ , vidíme, že pro všechny úkoly (*Tabulka 4.1.1*) se průměrná hodnota u kontrolních osob pohybuje od  $19.59 \pm 3.73$  ST pro monolog do  $20.86 \pm 3.27$  ST pro, zatímco u pacientů se pohybuje od  $12.68 \pm 1.78$  ST pro popis obrázku do  $14.20 \pm 3.07$  ST pro čtení. Při výpočtu  $p$ -value pro různé řečové úkoly vidíme, že nepřesahují 5 % ( $p = 0.04$  pro čtení,  $p = 0.03$  pro převyprávění pohádky,  $p = 0.04$  pro monologickou úlohu a  $p = 0.03$  pro popis obrázku). To nám umožňuje konstatovat, že hodnota základní frekvence v semitónech je reprezentativní pro akustickou analýzu řeči pacientů se schizofrenií. Zatímco analýza rozsahu  $f_{0\_range\_semi}$  ukazuje, že tento parametr není tak reprezentativní. U kontrolních osob se  $f_{0\_range\_semi}$  pohybuje od 28.94 do 32.72 ST, zatímco u pacientů se pohybuje od 20.04 do 32.26 ST, přičemž  $p$ -value se pohybuje od 15 % pro popis obrázku do 15.7 % pro monolog.

Naopak nevhodnými parametry jsou parametry spojené s intenzitou. Při analýze průměrné hodnoty intenzity v dB vidíme, že  $I_{mean}$  se pohybuje od  $51.82 \pm 4.69$  do  $55.79 \pm 4.21$  dB u kontrolních osob, zatímco u pacientů se pohybuje od  $57.33 \pm 3.89$  do  $59.33 \pm 3.85$  dB u nemocných s  $p$ -value od 22 % pro čtení do 36 % pro převyprávění pohádky. Toto kritérium, zda se že není dostatečně vhodné pro rozlišení mezi kontrolou a nemocním pacientem. Ohledně rozsahu intenzity platí, že pro tuto charakteristiku jsme nezamítli hypotézu o rovnosti středních hodnot předpokládaných normálních rozdělení pro obě skupiny mluvčích.

Lze tedy konstatovat, že ze zkoumaných prozodických charakteristik nejvíce ukazuje možnost detekce schizofrenie průměrná hodnota  $f_{0\_mean\_semi}$ .

### Artikulační a časové charakteristiky

Při vyhodnocení artikulačních a časových charakteristik vidíme, že absolutní hodnoty doby trvání nahrávek, celkového počtu slabik, počtu pauz a doby fonačního času u různých úloh nelze použít jako hodnotící parametr. Je to zřejmě způsobeno tím, že v závislosti na druhu úkolu může být doba trvání nahrávky pro pacienty kratší nebo delší než odpovídající doba u kontrolních osob, například u čtení je delší u pacientů, zatímco pro úlohu převyprávění pohádky je naopak dvakrát kratší. Podíváme-li se na poměr počtu pauz k celkové době nahrávky  $v_{\text{pause\_rate}}$ , vidíme, že u všech úloh, s výjimkou čtení, je tento poměr u nemocných výrazně nižší: 1.08 u KT a 0.86 u PT pro převyprávění pohádky ( $p = 0.09$ ); 1.08 pro KT a 0.77 pro PT u monologu ( $p = 0.003$ ); 0.9 pro KT a 0.58 pro PT u popisu obrázku ( $p = 0.007$ ), zatímco u monologu 0.98 pro KT a 0.91 pro PT u čtení ( $p = 0.5$ ). Tato charakteristika je tedy skutečně reprezentativní pro úlohy spontánní promluvy, nikoliv pro čtený text.

Dalším podstatným parametrem, který je příznačný pro všechny úlohy (opět s výjimkou čtení), je rychlost řeči  $v_{\text{speech\_rate}}$ . Pro úlohu převyprávění pohádky u KT je rychlost řeči 3.25 slabik/s, u PT 2.8 slabik/s ( $p = 0.03$ ), pro monolog: 3.39 slabik/s a 3.09 slabik/s ( $p = 0.04$ ), zatímco pro popis obrázku: 3.41 slabik/s a 2.41 slabik/s ( $p = 0.01$ ). Pro úlohu čtení hodnota  $p$ -value přesáhla hladinu významnosti 5 %. Vidíme tedy, že i přesto, že u všech úloh je rychlost řeči  $v_{\text{speech\_rate}}$  u PT statisticky nižší než u KT, lze ji spolehlivě používat jako kvalitní parametr pouze u pohádky, monologu a popisu obrázku, zatímco úloha čtení není tak reprezentativní, rychlost řeči v ní není relevantním ukazatelem.

Obdobně se projevuje i průměrná délka pauz  $t_{\text{pause}}$ : u čtení je to 0.22 s pro KT a 0.21s pro PT ( $p = 0.46$ ), zatímco u úlohy převyprávění pohádky pro KT je 0.19 s, pro PT 0.43s ( $p = 0.06$ ), u monologu pro KT 0.2s, pro PT 0.35s ( $p = 0.03$ ), pro úlohu popisu obrázku 0.29s pro KT a 0.71s pro PT ( $p = 0.06$ ). Důvodem takového výrazného rozdílu je to, že při čtení obvykle člověk nedělá dlouhé pauzy, protože text je předem definovaný. Z výsledků je zřejmé, že právě charakteristiky související s pauzami byly při rozlišování mezi schizofreniky a kontrolními osobami nejvýznamnější. Pacienti se schizofrenií mají v průměru tendenci vkládat do řeči více pauz, což způsobuje, že poměr doby artikulace k trvání nahrávky je u pacientů se schizofrenií vyšší než u kontrolních subjektů pro každou řečovou úlohu. Hodnota  $proportion\_of\_spoken\_time$  je u KT 0.79 a u PT 0.8 u čtení ( $p = 0.37$ ), zatímco pro převyprávění pohádky je rozdíl významnější vzhledem k delšímu trvání pauz (0.82 pro KT a 0.66 pro PT,  $p = 0.01$ ).

## Formanty

Analyzujeme-li formanty, pozorujeme, že jejich různé charakteristiky jsou nestejně reprezentativní. Především na základě analýzy absolutních hodnot formantů  $F_1$  a  $F_2$  vidíme, že hodnoty  $F_1$  se u hlásek /i/ a /u/ se liší o 12 % až 17 %, zatímco u hlásek /a/ o 45 % až 56 %. Ještě jednou zopakujeme, že databáze použitá v této práci se skládala ze 2 žen a 1 muže (KT) a 3 mužů (PT), pozorované rozdíly tedy také souvisí s pohlavním zastoupením. Porovnáme-li hodnoty získané v implementační části s genderovým rozdílem pro muže a ženy (viz v příloze *Obrázek B2*), vidíme, že rozdíl pro hlásku /i/ je od 14 % do 20 %, pro hlásku /u/ od 11 % do 15 %, proto veličina  $F_{1u}$  a  $F_{1i}$  není při řečovém vyšetřování schizofrenie demonstrativní.

Pokud se však zaměříme na diferenci  $F_1$  pro hlásku /a/, zjistíme, že běžný rozdíl mezi jednotlivými pohlavími se pohybuje od 17 % do 28 %. To nám dovoluje tvrdit, že rozdíl v  $F_1$  pro hlásku /a/ je reprezentativní a lze jej využít i pro analýzu schizofrenie, avšak toto tvrzení je třeba ověřit na větším počtu pacientů, abychom mohli s jednoznačnou platností.

Při analýze parametrů  $F_2$  vidíme, že  $F_2$  pro hlásku /a/ se liší o 15-21 %, pro hlásku /u/ o 1-5 %, pro hlásku /i/ o 17-23 %, přičemž obvyklý rozdíl mezi pohlavími je 11-14 % pro hlásku /a/, 3-7 % pro hlásku /u/ a 18-23 % pro hlásku /i/. Poznamenejme, že druhé formanty se v našem výzkumu u PT a KT výrazně neliší, takže tyto charakteristiky nepředstavují možnost posouzení vlivu schizofrenie.

Přesuneme-li se k analýze dalších parametrů, které byly popsány v *Podkapitole 3.3.6*, začneme hodnotou VSA. Obvyklý rozdíl mezi pohlavími je 40-150 %, zatímco v našem šetření je to 400-600 %. Je to způsobeno především rozdílem ve frekvenci  $F_1$  pro hlásku /a/, který jsme popsali výše. Všimněte si, že rozdíl v těchto hodnotách nám umožňuje tvrdit, že pro analýzu schizofrenie je uvedený parametr dokonce vhodnější kandidát než  $F_1$  pro hlásku /a/, a to z důvodu vyššího absolutního rozdílu ve výsledcích, avšak, jak jsme poznamenali, je k tomu třeba provést lepší statistickou analýzu.

Při analýze dalších parametrů si všimneme, že VIA, parametr, který není závislý na pohlaví, se v naší analýze liší o 3-5 % (což je každopádně méně než statistický rozptyl, který nedovoluje jeho použití pro analýzu). Stejně se chová i hodnota parametru  $F_{2i}/F_{2u}$ , která se v naší analýze liší o 7-12 %, což je však méně než statistická odchylka. Oba tyto parametry tedy nejsou v rámci naší analýzy dostatečně reprezentativní.

Shrnutím jsme dospěli k závěru, že analýza umožňuje uvést jako nejvýznamnější charakteristiky, které lze použít pro řečovou diagnostiku schizofrenie, tyto ukazatele: průměrná

hodnota základní frekvenci v semitónech  $f_{0\_mean\_semi}$ , řečová rychlost  $v_{speech\_rate}$ , poměr doby artikulace k trvání nahrávky *proportion\_of\_spoken\_time* pro spontánní promluvu, formant  $F_{1a}$  a velikost oblasti samohláskového prostoru VSA.

### **Analýza pomocí toolboxu DYSAN**

Nejvýznamnějšími rozdíly v hodnotách získaných pomocí toolboxu DYSAN byly rozdíly v oblasti časových charakteristik, jak se i očekávalo na základě teoretického výzkumu a výsledků implementační části studie. Průměrné hodnoty DPI se lišily nejvíce u úlohy s popisem obrázku vyvolávajícím emoce. Hodnota DPI pro PT byla o 11-88 % vyšší, což podle uvedeného popisu [52] znamená obtíže při zahájení řeči a chybění krátkých pauz a může být spolehlivým markerem pro detekci schizofrenie. Z oblasti fonačních charakteristik byly odhaleny odlišnosti u hodnot GVI a DVI. Hodnota GVI snižená o 29-43 % při monologických úlohách ukazuje na omezenou schopnost hlasivek ukončit vyslovování prostřednictvím addukcí, avšak tuto charakteristiku nemůžeme jednoznačně posoudit vzhledem k malému počtu analyzovaných nahrávek. Velice zajímavých výsledků jsme dosáhli v oblasti respiračních charakteristik, pro které žádný rozsáhlý výzkum v souvislosti se schizofrenií nebyl dosud proveden. Hodnota LRE se pro úlohy spontánní promluvy lišila překvapivě o 60-168 %. Z těchto hodnot se dá usoudit snižená schopnost přechodu z výdechu do nádechu, obzvláště problémy se zahájením nádechu.

Uvedený proces automatizované akustické analýzy je klíčovým nástrojem pro získání hlubokého vhledu do řečových poruch, protože poskytuje výsledky nezkreslené pohlavní dvojitvárností a vlivem stárnutí [55] a může sloužit k rozpoznání schizofrenie. V tuto chvíli však nelze vyvodit jednoznačné závěry vzhledem k nedostatku poskytnutých nahrávek řeči, ale zajisté je třeba tuto metodu počítačové analýzy dále zkoumat, až bude k dispozici rozsáhlejší databáze.

### **Porovnání s podobnými studiemi**

Pokusíme se porovnat získané výsledky s podobnými studiemi, především s články [8, 24, 31]. Tyto práce, jak bylo popsáno detailněji v *Kapitole 2*, ukazují na zřetelné účinky snížené produkce řeči (poměru doby artikulace k trvání nahrávky, střední rychlosti produkce pauz a řečové rychlosti), variability výšky tónu, průměrnou hodnotu výšky tónu a střední doby trvání pauz pro pacienty se schizofrenií. Ovšem důkladného průzkumu formantů a formantových charakteristik

VSA, VAI,  $F_{2i}/F_{2u}$ , závislosti rozdílu akustických hodnot na definice akustických úloh, srovnání významnosti řečové a artikulační rychlosti, variability řečové rychlosti nebylo dosud provedeno.

Srovnání hodnot této práce a literatury shrneme do tabulek. Začneme výsledky porovnání artikulačních a časových charakteristik pro úlohu čtení. Ve studiu [24] se ukazuje, že pacienti se schizofrenií mají v průměru tendenci vkládat do čtené řeči více pauz (+27 %), což vede k tomu, že poměr doby pauz k trvání nahrávky a jejich celková délka jsou u pacientů se schizofrenií vyšší než u kontrolních osob (viz *Tabulka 5.1*). Průměrná délka pauzy je u pacientů se schizofrenií také mírně vyšší (+10 %).

	<u>Výsledky z práce [24]</u>		<u>Výsledky této práce</u>	
	<b>KT</b>	<b>PT</b>	<b>KT</b>	<b>PT</b>
$t_{\text{pause}}$	0.57	0.63	0.22	0.21
<i>proportion_of_spoken_time</i>	0.84	0.81	0.79	0.80

**Tabulka 5.1.** Porovnání artikulačních a časových charakteristik řeči pro skupiny KT a PT pro čtení.

Dále pro úlohu čtení vidíme, že získané v této práci snížená průměrná hodnota a variabilita základní frekvence se potvrzují na základě studia [57].

	<u>Výsledky z práce [57]</u>		<u>Výsledky této práce</u>	
	<b>KT</b>	<b>PT</b>	<b>KT</b>	<b>PT</b>
$f_{0\_mean\_Hz}$	156.57	143.02	173.51	117.70
$\sigma_{f0\_Hz}$	29.67	27.50	42.33	36.40

**Tabulka 5.2.** Porovnání prozodických charakteristik řeči pro skupiny KT a PT pro čtení.

Ve studii [31] se prokazuje, že rozsah druhého formantu  $F_2$  u schizofrenní řeči je menší než u kontrol. Druhé formanty se v našem výzkumu u PT a KT výrazně neliší, což také může být způsobené nedostatkem dat v této práci. Nicméně pozitivní a perspektivní je získaná v této práci souvislost mezi parametrem VSA, který nebyl v jiné literatuře zkoumán.

## 6. Závěr

---

### Výsledky práce.

Základním cílem této práce bylo vyhodnotit možnost měření poruch řeči pomocí počítačové akustické analýzy. Při zkoumání objektivních příznaků schizofrenie má měření akustických charakteristik své výhody: postup není pro pacienta stresující, může jej provádět vyškolený pracovník a nahrávky jsou vždy k dispozici pro opakovanou analýzu. Nevýhodou jsou však technické a metodické komplikace při získávání spolehlivých údajů z nahrávek.

Při hodnocení jednotlivých akustických charakteristik získaných ze spontánní a čtené řeči považujeme naše předchozí hypotézy, stanovené z rešerše literatury v *Kapitole 2*, za potvrzené. Tabulky a diagramy z *Kapitoly 4* ukazují výsledky klasifikace ve spontánní řeči a čtení textu.

Většina studií o akustické analýze pacientů se schizofrenií využívá úkoly založené na spontánní promluvě pro vyvolání emocí. V naší práci jsme se zaměřili na víceúrovňový přístup s použitím jak úlohy čteného textu, odpovědi na emočně neutrální a emoce vyvolávající otázky, tak i a převyprávění pohádky. Ačkoli hodnoty jednotlivých parametrů ve čtyřech úlohách spolu zpravidla vysoce korelují, ukázalo se, že různé typy úloh k získání zvukového záznamu řeči mohou ve skutečnosti způsobit určité rozdíly v prozodických, artikulačních a časových parametrech. Jak je uvedeno v *Tabulce 4.2.1*, byla celková míra detekci schizofrenie z řečových úloh při použití spontánní řeči vyšší než u čtené řeči, což naznačuje, že spontánní řeč obsahuje více relevantních informací ohledně akustických změn. Proto při budoucím vývoji bude nejspíše nutné vycházet z několika způsobů vyvolání řeči, aby bylo možné plně zachytit rozdíly v akustických charakteristikách. Tyto odlišnosti v akustických charakteristikách řeči mohou souviset s kognitivními poruchami, které postihují osoby se schizofrenií [56] .

### Navazující výzkum.

Využití řeči pro hodnocení schizofrenie může ve vývoji umožnit vznik vzdáleného a plně automatizovaného kontrolního systému. S ohledem na objektivní nedostatek řečových záznamů by tato práce měla být brána jako předstupeň rozsáhlejšího výzkumu dané problematiky, který rozsáhlý počet prozodických, artikulačních a časových akustických charakteristik vyžaduje. Další studie by se měly zaměřit na ověření dosažených vztahů mezi různými řečovými vzorci a různými formami schizofrenie na výrazně větším počtu dat, jakožto i na následnou statistickou analýzu trendů. Obzvláště zajímavé by bylo rozšířit výzkum formantů, hodnot VSA pro potvrzení



dosažených závislostí a vyloučení možné podstaty těchto výsledků ve vztahu k závislosti na pohlaví. Důležitým aspektem by bylo také provést srovnání použitých metod v implementační části s jinými metodami, které jsou k dispozici v akustické počítačové analýze, a porovnat jejich robustnost. Jedním z intervenujících faktorů, který může výsledky, především spojených s rychlostí řeči a délkou pauz ovlivnit, je medikace pacientů, kdyby byl zajímavý vliv medikace pacientů na řečové projevy pacientů. Dále by bylo možné použitou metodu detekce schizofrenie rozšířit pro poruchy artikulace u cizích jazyků.

## Použitá literatura

---

- [1] KAUR P., SHARMA M., “Diagnosis of Human Psychological Disorders using Supervised Learning and Nature-Inspired Computing Techniques: A Meta-Analysis”. *Journal of Medical Systems*. 2019, 43(7). ISSN 0148-5598. doi:10.1007/s10916-019-1341-2
- [2] GAUTAM R., SHARMA M., “Prevalence and Diagnosis of Neurological Disorders Using Different Deep Learning Techniques: A Meta-Analysis”. *Journal of Medical Systems*. 2020, 44(2). ISSN 0148-5598. doi:10.1007/s10916-019-1519-7
- [3] LAVRETSKY H., “History of schizophrenia as a psychiatric disorder”. In K. T. Mueser & D. V. Jeste (Eds.), *Clinical handbook of schizophrenia* (pp. 3–13), The Guilford Press, (2008).
- [4] DOCHERTY N.M., “Affective reactivity of symptoms as a process discriminator in schizophrenia”. *Journal of Nervous and Mental Disease*, 1996: 184(9), 535–541, doi: 10.1097/00005053-199609000-00004
- [5] BUCHANAN R.W., “Persistent Negative Symptoms in Schizophrenia: An Overview”. *Schizophrenia Bulletin*. 2007, 33(4), 1013-1022. ISSN 0586-7614. doi:10.1093/schbul/sbl057
- [6] KESHAVAN M.S., NASRALLAH H.A, TANDON R., “Schizophrenia, “Just the Facts” 6. Moving ahead with the schizophrenia concept: From the elephant to the mouse”. *Schizophrenia Research*. 2011, 127(1-3), 3-13. ISSN 09209964. doi:10.1016/j.schres.2011.01.011
- [7] SEKAR A., BIALAS A., DE RIVERA H. *et al.* “Schizophrenia risk from complex variation of complement component 4”. *Nature*. 2016, 530(7589), 177-183. ISSN 0028-0836. doi:10.1038/nature16549
- [8] COHEN A.S., ALPERT M., NIENOW T.M., DINZEO T.J., DOCHERTY N.M.. “Computerized measurement of negative symptoms in schizophrenia”. *Journal of Psychiatric Research*. 2008, 42(10), 827-836. ISSN 00223956. doi:10.1016/j.jpsychires.2007.08.008

- [9] KUMARI S., MALIK M., FLORIVAL C., MANALAI P., “An Assessment of Five (PANSS, SAPS, SANS, NSA-16, CGI-SCH) commonly used Symptoms Rating Scales in Schizophrenia and Comparison to Newer Scales (CAINS, BNSS)”. *Journal of Addiction Research & Therapy*. 2017, 08(03). ISSN 21556105. doi:10.4172/2155-6105.1000324
- [10] ALPERT M., SHAW RJ., POUGET ER., LIM KO., “A comparison of clinical ratings with vocal acoustic measures of flat affect and alogia”. *Journal of Psychiatric Research*. 2002, 36(5), 347-353 [cit. 2022-05-19]. ISSN 00223956. doi:10.1016/S0022-3956(02)00016-X
- [11] TAWARI A., TRIVEDI MM., “Speech Emotion Analysis: Exploring the Role of Context”. *IEEE Transactions on Multimedia*. 2010, 12(6), 502-509. ISSN 1520-9210. doi:10.1109/TMM.2010.2058095
- [12] COHEN AS., RENSHAW TL., MITCHELL KR., KIM Y., “A psychometric investigation of “macroscopic” speech measures for clinical and psychological science”. *Behavior Research Methods*. 2016, 48(2), 475-486. ISSN 1554-3528. doi:10.3758/s13428-015-0584-1
- [13] COHEN AS., NAJOLIA GM., KIM Y., DINZEO TJ., “On the boundaries of blunt affect/alogia across severe mental illness: Implications for Research Domain Criteria”. *Schizophrenia Research*. 2012, 140(1-3), 41-45. ISSN 09209964. doi:10.1016/j.schres.2012.07.001
- [14] BALSTERS MJN., KRAHMER EJ., SWERTS MGJ., VINGERHOETS AJJM.. “Verbal and Nonverbal Correlates for Depression: A Review”. *Current Psychiatry Reviews*. 2012, 8(3), 227-234. ISSN 15734005. doi:10.2174/157340012800792966
- [15] DEVAULT D., GEORGILIA K., ARTSTEIN R., MORBINI F., TRAUM D., SCHERER S., RIZZO A., MORENCY LP., “Verbal indicators of psychological distress in interactive dialogue with a virtual human”. *Proceedings of SigDial*. 2013. Association for Computational Linguistics, pp. 193–202.
- [16] SHENTON ME., DICKEY CC., FRUMIN M., MCCARLEY RW., “A review of MRI findings in schizophrenia”. *Schizophrenia Research*. 2001, 49(1-2), 1-52. ISSN 09209964. doi:10.1016/S0920-9964(01)00163-3

- [17] PALMER BW., HEATON RK., KUCK J., BRAFF D. *et al.*, “Is it possible to be schizophrenic yet neuropsychologically normal? ” *Neuropsychology*. 1997, 11(3), 437-446. ISSN 0894-4105. doi:10.1037//0894-4105.11.3.437
- [18] RAPOPORT JL., “Childhood-Onset Schizophrenia”. *Archives of General Psychiatry*. 1997, 54(10). ISSN 0003-990X. doi:10.1001/archpsyc.1997.01830220013002
- [19] RAPOPORT JL., GIEDD JN., BLUMENTHAL J. *et al.*, “Progressive Cortical Change During Adolescence in Childhood-Onset Schizophrenia”. *Archives of General Psychiatry*. 1999, 56(7). ISSN 0003-990X. doi:10.1001/archpsyc.56.7.649
- [20] TITZE IR., “Comments on the Myoelastic - Aerodynamic Theory of Phonation”. *Journal of Speech, Language, and Hearing Research*. 1980, 23(3), 495-510. ISSN 1092-4388. doi:10.1044/jshr.2303.495
- [21] COUTURE SM., GRANHOLM EL., FISH SC., “A path model investigation of neurocognition, theory of mind, social competence, negative symptoms and real-world functioning in schizophrenia”. *Schizophrenia Research*. 2011, 125(2-3), 152-160. ISSN 09209964. doi:10.1016/j.schres.2010.09.020
- [22] HÄFNER H., LÖFFLER W., MAURER K., HAMBRECHT M., HEIDEN W., “Depression, negative symptoms, social stagnation and social decline in the early course of schizophrenia”. *Acta Psychiatrica Scandinavica*. 1999, 100(2), 105-118. ISSN 0001-690X. doi:10.1111/j.1600-0447.1999.tb10831.x
- [23] KIROLA C., BERARDO G., KAPUR S., “Negative symptoms have greater impact on functioning than positive symptoms in schizophrenia: Analysis of CATIE data”. *Schizophrenia Research*. 2012, 137(1-3), 147-150. ISSN 09209964. doi:10.1016/j.schres.2012.01.015
- [24] RAPCAN V., D’ARCY S., YEAP S., AFZAL N., THAKORE J., REILLY RB., “Acoustic and temporal analysis of speech: A potential biomarker for schizophrenia”. *Medical Engineering & Physics*. 2010, 32(9), 1074-1079. ISSN 13504533. doi:10.1016/j.medengphy.2010.07.013
- [25] ALPERT M., ROSENBERG SD., POUGET ER., SHAW RJ., “Prosody and lexical accuracy in flat affect schizophrenia”. *Psychiatry Research*. 2000, 97(2-3), 107-118 [cit. 2022-05-19]. ISSN 01651781. doi:10.1016/S0165-1781(00)00231-6

- [26] COHEN AS., MITCHELL KR., DOCHERTY NM., HORAN WP., “Vocal expression in schizophrenia: Less than meets the ear”. *Journal of Abnormal Psychology*. 2016, 125(2), 299-309. ISSN 1939-1846. doi:10.1037/abn0000136
- [27] MEAUX LT., MITCHELL KR., COHEN AS., “Blunted vocal affect and expression is not associated with schizophrenia: A computerized acoustic analysis of speech under ambiguous conditions”. *Comprehensive Psychiatry*. 2018, 83, 84-88. ISSN 0010440X. doi:10.1016/j.comppsy.2018.03.009
- [28] MEEHL PE., “Schizotaxia, schizotypy, schizophrenia”. *American Psychologist*. 1962, 17(12), 827-838. ISSN 0003-066X. doi:10.1037/h0041029
- [29] TRÉMEAU F., MALASPINA D., DUVAL F., CORRÊA H., HAGER-BUDNY M., COIN-BARIOU L., MACHER J-P., GORMAN JM., “Facial Expressiveness in Patients With Schizophrenia Compared to Depressed Patients and Nonpatient Comparison Subjects”. *American Journal of Psychiatry*. 2005, 162(1), 92-101. ISSN 0002-953X. doi:10.1176/appi.ajp.162.1.92
- [30] PAROLA A., SIMONSEN A., BLIKSTED V., FUSAROLI R., “Voice patterns in schizophrenia: A systematic review and Bayesian meta-analysis”. *Schizophrenia Research*. 2020, 216, 24-40. ISSN 09209964. doi:10.1016/j.schres.2019.11.031
- [31] COMPTON MT., LUNDEN A., CLEARY SD. *et al.*, “The aprosody of schizophrenia: Computationally derived acoustic phonetic underpinnings of monotone speech”. *Schizophrenia Research*. 2018, 197, 392-399. ISSN 09209964. doi:10.1016/j.schres.2018.01.007
- [32] WYMER JH., LINDMAN SL., BOOKSH RL., “A Neuropsychological Perspective of Aprosody: Features, Function, Assessment, and Treatment”. *Applied Neuropsychology*. 2010, 9(1), 37-47. ISSN 0908-4282. Dostupné z: doi:10.1207/S15324826AN0901\_5
- [33] TALKIN D., “A Robust Algorithm for Pitch Tracking (RAPT) ”. *Speech Coding and Synthesis*, W. B. Kleijn and K. K. Palatal (eds), Elsevier Science B.V., 1995, 497-518
- [34] MEI X-D., PAN J., SUN S-H., “Efficient algorithms for speech pitch estimation”. *In: Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing. ISIMP 2001 (IEEE Cat. No.01EX489)*. IEEE, 2001, s. 421-424. ISBN 962-85766-2-3. doi:10.1109/ISIMP.2001.925423

- [35] KOHAVI R., “A study of cross-validation and bootstrap for accuracy estimation and model selection”. In: *Proceedings of Fourteenth International Joint Conference on Artificial Intelligence, IJCAI*, Montreal, 1995, CA, pp. 1137–1143
- [36] ALPERT M., POUGET ER., WELKOWITZ J., COHEN J., “Mapping schizophrenic negative symptoms onto measures of the patient's speech: Set correlational analysis”. *Psychiatry Research*. 1993, 48(3), 181-190. ISSN 01651781. doi:10.1016/0165-1781(93)90070-W
- [37] RESNICK HS., OLTMANNS TF., “Hesitation patterns in the speech of thought-disordered schizophrenic and manic patients”. *Journal of Abnormal Psychology*. 1984, 93(1), 80-86. ISSN 1939-1846. doi:10.1037/0021-843X.93.1.80
- [38] PÜSCHEL J., STASSEN HH., BOMBEN G., SCHARFETTER C., HELL D., “Speaking behavior and speech sound characteristics in acute schizophrenia”. *Journal of Psychiatric Research*. 1998, 32(2), 89-97. ISSN 00223956. Dostupné z: doi:10.1016/S0022-3956(98)00046-6
- [39] STASSEN HH., ALBERS M., PÜSCHEL J., SCHARFETTER C., TEWESMEIER M., WOGGON B. “Speaking behavior and voice sound characteristics associated with negative schizophrenia”. *Journal of Psychiatric Research*. 1995, 29(4), 277-296. ISSN 00223956. doi:10.1016/0022-3956(95)00004-O
- [40] ROACH P., “English Phonetics and Phonology: A Practical Course”, *Cambridge University Press*, Cambridge, 2000
- [41] WANG D., NARAYANAN SS., “Robust Speech Rate Estimation for Spontaneous Speech”. *IEEE Transactions on Audio, Speech and Language Processing*. 2007, 15(8), 2190-2201. ISSN 1558-7916. doi:10.1109/TASL.2007.905178
- [42] TYKALOVA T., SKRABAL D., BORIL T., CMEJLA R., VOLIN J. a RUSZ J. “Effect of Ageing on Acoustic Characteristics of Voice Pitch and Formants in Czech Vowels”. *Journal of Voice*. 2021, 35(6), 931.e21-931.e33. ISSN 08921997. doi:10.1016/j.jvoice.2020.02.022
- [43] RUSZ J., CMEJLA R., TYKALOVA T. *et al.*, “Imprecise vowel articulation as a potential early marker of Parkinson's disease: Effect of speaking task”. *The Journal of the Acoustical Society of America*. 2013, 134(3), 2171-2181. ISSN 0001-4966. doi:10.1121/1.4816541

- [44] KENT RD., WEISMER G., KENT JF., VORPERIAN HK., DUFFY JR., “Acoustic studies of dysarthric speech”. *Journal of Communication Disorders* . 1999, 32(3), 141-186. ISSN 00219924. doi:10.1016/S0021-9924(99)00004-0
- [45] BOERSMA P., WEENINK D., “Praat: doing phonetics by computer”. [Computer program]. Version 6.1.10, 2020. <http://www.praat.org/>.
- [46] MATLAB. version 9.10.0 (R2021a). The MathWorks Inc., Natick, Massachusetts, 2018.
- [47] DE JONG NH., WEMPE T., “Praat script to detect syllable nuclei and measure speech rate automatically”. *Behavior Research Methods*. 2009, 41(2), 385-390. ISSN 1554-351X. doi:10.3758/BRM.41.2.385
- [48] COLEMAN RF., MABIS JH., HINSON JK., “Fundamental Frequency-Sound Pressure Level Profiles of Adult Male and Female Voices”. *Journal of Speech and Hearing Research*. 1977, 20(2), 197-204. ISSN 0022-4685. doi:10.1044/jshr.2002.197
- [49] MURRY T. “Speaking Fundamental Frequency Characteristics Associated with Voice Pathologies”. *Journal of Speech and Hearing Disorders*. 1978, 43(3), 374-379. ISSN 0022-4677. doi:10.1044/jshd.4303.374
- [50] KENT RD., KIM Y.-J. “Toward an acoustic typology of motor speech disorders”. *Clinical Linguistics & Phonetics*. 2009, 17(6), 427-445. ISSN 0269-9206. doi:10.1080/0269920031000086248
- [51] DE MAESSCHALCK R., JOUAN-RIMBAUD D., MASSART DL., “The Mahalanobis distance”. *Chemometrics and Intelligent Laboratory Systems*. 2000, 50(1), 1-18. ISSN 01697439. doi:10.1016/S0169-7439(99)00047-7
- [52] HLAVNIČKA J., TYKALOVÁ T., ČMEJLA R., KLEMPÍŘ J., RŮŽIČKA J., RUSZ J., “Dysprosody Differentiate Between Parkinson’s Disease, Progressive Supranuclear Palsy, and Multiple System Atrophy”. *In: Interspeech 2017*. ISCA: ISCA, 2017, 2017-8-20, s. 1844-1848. doi:10.21437/Interspeech.2017-762
- [53] HLAVNIČKA J., ČMEJLA R., TYKALOVÁ T., ŠONKA K., RŮŽIČKA E., RUSZ J., “Automated analysis of connected speech reveals early biomarkers of Parkinson’s disease in patients with rapid eye movement sleep behaviour disorder”. *Scientific Reports*. 2017, 7(1). ISSN 2045-2322. doi:10.1038/s41598-017-00047-5

- [54] COLEMAN RF., MABIS JH., HINSON JK. "Fundamental Frequency-Sound Pressure Level Profiles of Adult Male and Female Voices". *Journal of Speech and Hearing Research*. 1977, 20(2), 197-204. ISSN 0022-4685. 10.1044/jshr.2002.197
- [55] BUKOVSKÝ P., Diplomová práce "Automatické hodnocení poruch artikulace samohlásek u Parkinsonovy nemoci", 2017. České vysoké učení technické v Praze, Fakulta elektrotechnická, Katedra teorie obvodů. Vedoucí práce: Ing. Jan Ruzs, Ph.D.
- [56] PALMER BW., HEATON RK., KUCK J., BRAFF D. et al., "Is it possible to be schizophrenic yet neuropsychologically normal?" *Neuropsychology*. 1997, 11(3), 437-446. ISSN 0894-4105. doi:10.1037//0894-4105.11.3.437
- [57] MARTÍNEZ-SÁNCHEZ F., MUELA-MARTÍNEZ JA., CORTÉS-SOTO P., GARCÍA MEILÁN JJ., VERA FERRÁNDIZ JA., EGEA CAPARRÓS A., a PUJANTE VALVERDE IA., "Can the Acoustic Analysis of Expressive Prosody Discriminate Schizophrenia? ". *The Spanish Journal of Psychology*. 2015, 18. ISSN 1138-7416. doi:10.1017/sjp.2015.85



## Příloha A – Zkratky a symboly:

---

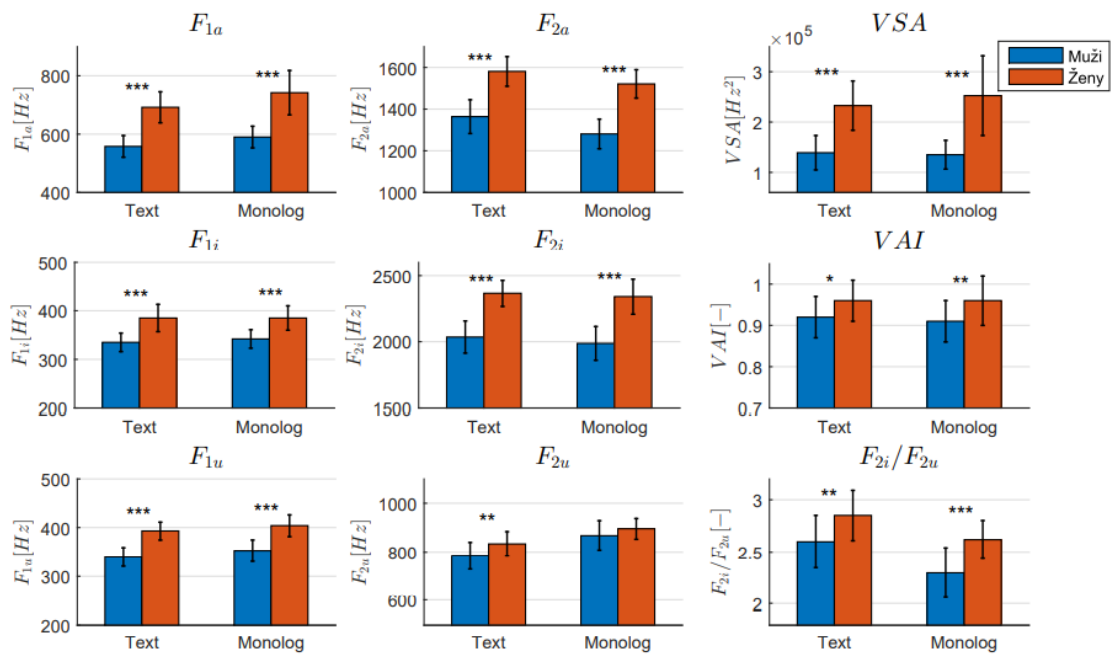
PT	Pacienty se schizofrenií
KT	Kontrolní osoby
$F_0$	Základní formantová frekvence
$F_1$	První formantová frekvence
$F_2$	Druhá formantová frekvence
$F_3$	Třetí formantová frekvence
VAI	Artikulační index (z anglického vowel articulation index)
VSA	Obsah vokální plochy (z anglického vowel space area)
EST	Entropie časového průběhu řeči (z anglického Entropy of speech timing)
RST	Rychlost průběhu řeči (z anglického The rate of speech timing)
AST	Zrychlení průběhu řeči (z anglického Acceleration of speech timing)
DPI	Doba trvání pauz (z anglického Duration of pause intervals)
DVI	Doba trvání znělých intervalů (z anglického Duration of voiced intervals)
GVI	Mezery mezi znělými intervaly (z anglického Gaping in between voiced intervals)
DUS	Doba trvání neznělých exploziv (z anglického Duration of unvoiced stops)
DUF	Útlum neznělých frikativ (z anglického Decay of unvoiced fricatives)
RLR	Relativní úroveň hlasitosti respirace (z anglického Relative loudness of respiration)
PIR	Respirační pauzy (z anglického Pause intervals per respiration)
RSR	Respirační výdechová frekvence (z anglického Rate of speech respiration)
LRE	Latence mezi nádechem a výdechem (z anglického Latency in respiratory Exchange)
<i>stdF0</i>	Směrodatná odchylka základní frekvence (z anglického standard deviation of fundamental frequency)
<i>stdPWR</i>	Směrodatná odchylka výkonové obálky (z anglického Standard deviation of power)

## Příloha B – Doplňující tabulky a obrázky:

Škála (Autor,rok)	Počet příznaků	Přednosti	Slabiny
Positive and Negative Syndrome Scale. (PANSS; Kay et al., 1987)	Celkem 30 bodů. Obsahuje pozitivní (7), negativní (7) příznaky a obecnou psychopatologickou škálu (16)	Nejčastěji používaná škála hodnocení. Díky tomu, že je citlivý na změny, je " gold standardem" ve studiích zabývajících se léčbou. PANSS není určen výhradně k hodnocení negativních příznaků, je to spíše komplexní škála pro hodnocení patologie	Zastaralé, dlouhotrvající. PANSS a SANS byly kritizovány, protože obsahují otázky, které měří kognitivní funkci (zkreslení pozornosti nebo abstraktní myšlení), které byly nyní uznávány jako odlišná kategorie od negativních příznaků .
Scale for Assessment of Positive Symptoms. SAPS- (Andreasen, 1984)	Celkem 34 bodů, měření halucinace, bludy, bizarní chování a myšlení poruchy.	Rozpoznává pozitivní příznaky. Má dobré platnost pro pozitivní příznaky.	Nelze použít samostatně. Používá se v ve spojení s SANS.
The Scale for Assessment of Negative Symptoms (SANS), (Andreasen, 1983)	SANS v původním znění obsahuje 25 položek. V současné době se SANS skládá z 19 položek představujících 5 škál: 1) Afektivní zploštění nebo otupení 2) Alogia 3) Apatie 4) Anhedonie- Asociálnost 5) Nepozornost.	Odděluje negativní příznaky od pozitivních příznaků a deprese.	Nelze použít samostatně; vyžaduje SAPS.
Clinical Assessment Interview for	Hodnocení klinického pracovníka. Skládá se z	CAINS je stručná, ale komplexní škála, kterou lze použít v širokém spektru	Škála CAINS není silně propojena s depresí, neklidem nebo

<p>Negative Symptoms (CAINS).</p>	<p>dvou stupnic, které se hodnotí samostatně. Všechny body jsou hodnoceny na základě pětibodové škály od 0 (žádné postižení) do 4 (závažný deficit). Škála motivace a potěšení (devět bodů) a škála exprese (čtyři body).</p> <ol style="list-style-type: none"> <li>1) Výraz obličeje</li> <li>2) Hlasový projev</li> <li>3) Expresivní projevy</li> <li>4) Kvalita řeči.</li> </ol>	<p>klinických a výzkumných oblastí. Jedná se o dobře vyvinutou a vyhodnocenou škálu pro měření negativních příznaků. Prokazuje dobrou vnitřní konzistenci a stabilitu při opakovaném testování, a spolehlivost a shodu mezi jednotlivými hodnotiteli. Prokázala také větší konvergentnost validitu než BPRS a SANS, pokud jde o negativní příznaky (Kring, 2010)</p>	<p>pozitivními příznaky (Kring, 2010)</p>
<p>Brief Negative Symptom Scale (BNSS) (Kirkpatrick, 2011)</p>	<p>13 bodů rozdělených do 6 subškál.</p> <ol style="list-style-type: none"> <li>1) Otupělý afekt</li> <li>2) Alogie</li> <li>3) Asocialita</li> <li>4) Anhedonie</li> <li>5) Distres</li> </ol>	<p>Její návrh umožňuje výzkumníkům zvážit mnoho aspektů negativních příznaků. Výsledky BNSS vysoce korelují s SANS a PANSS. Má silné interater, test-retest a vnitřní konzistenci. (Kirkpatrick, 2011)</p>	<p>Potřebují se zjistit, zda BNNS je citlivá na změnách (není známo, zda může být použita v klinických studiích)</p>

**Tabulka B1.** Škály pro hodnocení schizofrenie. Přepřacováno z [9].



**Obrázek B2.** Porovnání středních hodnot parametrů mužů a žen pro text a monolog. Hvězdičky indikují signifikantní rozdíly mezi ženami a muži: \*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$ . Převzato z [55] .

# Příloha C – Protokol k výzkumu pacientů se schizofrenií:

---

## I. Čtený text – akustická analýza

Když člověk poprvé vsadí do země sazeničku, chodí se na ni dívat třikrát denně: tak co, povyrostla už nebo ne? I tají dech, naklání se nad ní, přitlačí trochu půdu u jejich kořínků, načechrává jí lístky a vůbec ji obtěžuje různým konáním, které považuje za užitečnou péči. A když se sazenička přesto ujme a roste jako z vody, tu člověk žasne nad tímto divem přírody, má pocit čehosi jako zázraku a považuje to za jeden ze svých největších úspěchů.

## II. Spontánní promluva – cca 90s – převyprávění pohádky – prosodická informace

Popelka žije v domečku se svými dvěma línými sestrami Kateřinou a Dorou. Ony s jejich zlá macecha nutí Popelku k neustálé práci. Jednou se Popelčin hodný tatínek vydá do města a Kateřina s Dorou po něm chtějí drahé šaty a šperky, jen Popelce stačí to, co mu cestou zavadí o kabát. Když se tedy otec vrátí, předá Popelce větvičku se třemi oříšky, která mu cestou zpět zavádila o kabát. Svě ženě také řekne novinu, že prince se chystá oženit. A skutečně ze zámku se ozve hlas, který oznamuje, že se na zámku budou po tři dny konat plesy, kde si princ

Mojmír vybere svoji budoucí ženu. Na ples jsou zvány všechny dívky z království – chudé i bohaté, urozené i neurozené. V den, kdy se koná ples je macecha se svými dcerami v plné parádě. Ale aby se ujistila, že Popelka se smutně pustí do práce, když najednou se na okně objeví holoubek, který Popelce poradí, aby rozlouskla oříšek a o práci, že se postará on a jeho bratři. Hloubek jí zároveň varuje, že do půlnoci musí být zpátky. Popelka rozlouskne oříšek, obleče si modré šaty a odjede na ples.

Princ Mojmír se ženit nechce a z hledání nevěsty je otráven. Vše se ale změní, když do sálu vstoupí Popelka. Princ se do ní ihned zamiluje a protancuje s ní celý večer. Jenomže ještě, než odbije půlnoc. Popelka se z plesu vytratí. Druhý den musí Popelka opět přebírat popel od čocky. Ale i tentokrát jí přispěchají na pomoc holoubci. Popelka rozlouskla druhý oříšek a najde stříbrné šaty.

Princ už na ní na plesě čeká a znovu s ní stráví celý večer. Na třetí den je princ celý nedočkavý, a tak jeho otec a jeho rádce vymyslí malou lest. Natře schody lepidlem, aby se dívka, až bude

odcházet, přilepila. Popelka tentokrát najde v oříšku zlaté šaty a zase spěchá na ples. Princ na ni již čeká a požádá ji o ruku, nicméně když odbije půlnoc, Popelka uteče pryč. Na schodech se jí přilepí střevíček, který tam musí nechat. Princovi tak po jeho neznámé dívce zbude střevíček. Nakonec dojede až do chaloupky, kde bydlí Popelka. Její sestry se zkusí do střevíčku nasoukat, ale střevíček je moc malý. Princ už skoro ztrácí naději, když do chaloupky vejde Popelčin otec a řekne princovi, že má ještě jednu dceru – Popelku. Popelka střevíček zkusí a padne jí jako ulitý. Princ pozná, že je to dívka, do které se zamiloval a odveze si ji i s jejím otcem na zámek.

### III. Spontánní promluva – 90s – odpověď na otázku (emočně neutrální)

- Popište byt nebo dům, ve kterém bydlíte.

### IV. Spontánní promluva – 90s – popis obrázku (vyvolání emocí)



**Obrázky C1.** Obrázky 1, 4, a 6BM z Tematického apercepčního testu:

<https://www.psychestudy.com/general/personality/detailed-procedure-thematic-procedure-test>.