BRNO UNIVERSITY OF TECHNOLOGY

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

FAKULTA ELEKTROTECHNIKY
A KOMUNIKAČNÍCH TECHNOLOGIÍ

DEPARTMENT OF TELECOMMUNICATIONS

ÚSTAV TELEKOMUNIKACÍ

DIFFERENTIAL ANALYSIS OF MILD COGNITIVE IMPAIRMENT, DEMENTIA WITH LEWY BODIES, AND PARKINSON'S DISEASE

DIFERENČNÍ ANALÝZA MÍRNÉ KOGNITIVNÍ PORUCHY, DEMENCE S LEWYHO TĚLÍSKY A PARKINSONOVY NEMOCI

MASTER'S THESIS

DIPLOMOVÁ PRÁCE

AUTHOR — Bc. Tomáš Malcher

AUTOR PRÁCE

SUPERVISOR — doc. Ing. Jiří Mekyska, Ph.D.

VEDOUCÍ PRÁCE

BRNO 2025

Master's Thesis

Master's study program **Audio Engineering**
specialization Zvuková produkce a nahrávání
Department of Telecommunications

Student: Bc. Tomáš Malcher ***ID:*** 218400
Year of study: 2 ***Academic year:*** 2024/25

TITLE OF THESIS:

Differential analysis of mild cognitive impairment, dementia with Lewy bodies, and Parkinson's disease

INSTRUCTION:

The aim of this thesis is to propose a methodology that, based on acoustic analysis of speech, would support the differential diagnosis of three neurological diseases: mild cognitive impairment, dementia with Lewy bodies, and Parkinson's disease. This main objective is divided into the following sub-goals: 1) Conduct a literature review on speech and voice disorders in the aforementioned diseases; 2) Familiarize oneself with the provided database of speech recordings; 3) Based on the literature review, design and implement the parametrization of recordings using appropriately selected digital speech biomarkers; 4) Perform an exploratory analysis (identify which acoustic parameters differentiate the mentioned diseases and how well they correlate with clinical scales); 5) Perform a classification analysis (train and evaluate machine learning models that will automatically classify these diseases); 6) Discuss the results in relation to the available scientific literature.

RECOMMENDED LITERATURE:

[1] MARTÍNEZ-NICOLÁS, Israel, et al. Ten Years of Research on Automatic Voice and Speech Analysis of People With Alzheimer's Disease and Mild Cognitive Impairment: A Systematic Review Article. Online. Frontiers in Psychology. 2021, roč. 12. ISSN 1664-1078. Dostupné z: https://doi.org/10.3389/fpsyg.2021.620251

[2] RUSZ, Jan, et al.. Guidelines for Speech Recording and Acoustic Analyses in Dysarthrias of Movement Disorders. Movement Disorders [online]. 2021, 36(4), 803-814 [cit. 2023-06-09]. ISSN 0885-3185. Dostupné z: doi:10.1002/mds.28465

Date of project specification: 10.2.2025 ***Deadline for submission:*** 27.5.2025

Supervisor: doc. Ing. Jiří Mekyska, Ph.D.

doc. Ing. Jiří Schimmel, Ph.D.
Chair of study program board

ABSTRACT

This master's thesis investigates the potential of speech analysis as a tool for differential diagnosis of three neurological disorders: mild cognitive impairment (MCI), Lewy body dementia (DLB), and Parkinson's disease (PD). The study employs a comprehensive approach by extracting both acoustic and linguistic parameters from speech recordings and evaluating their effectiveness in distinguishing between clinical groups and healthy controls (HC). Acoustic parameters, particularly those related to temporal organization and spectral characteristics, demonstrated superior classification performance compared to linguistic features, achieving an accuracy of 0.64 in differentiating MCI from DLB. The HC and PD classification showed high specificity (0.90) but low sensitivity (0.26), reflecting the challenges in detecting speech alterations in early-stage PD.

Exploratory analysys revealed distinct association patterns between speech parameters and neuropsychological measures: temporal features (pause time, average pause duration) correlated primarily with depression scores, while acoustic parameters (especially the parameter focused on the standard deviation of the intensity contour relative to its mean) showed stronger associations with cognitive measures, particularly attention. An unexpected correlation between complex sentence ratio and visuospatial processing suggested potential shared neural substrates.

Despite moderate classification performance limiting immediate clinical application, this research provides valuable insights into the relationship between speech characteristics and neurological disorders, establishing a foundation for the development of objective, non-invasive diagnostic tools. Future directions include longitudinal studies tracking speech changes over disease progression and the integration of advanced acoustic analyses.

KEYWORDS

Neurodegenerative disorders, speech analysis, prosody, articulation, machine learning

ABSTRAKT

Tato diplomová práce zkoumá potenciál analýzy řeči jako nástroje pro diferenciální diagnostiku tří neurologických onemocnění: mírné kognitivní poruchy (MCI), demence s Lewyho tělísky (DLB) a Parkinsonovy nemoci (PD). Studie využívá komplexní přístup, kdy z nahrávek řeči extrahuje akustické a lingvistické parametry a hodnotí jejich účinnost při rozlišování mezi klinickými skupinami a zdravými kontrolami (HC). Akustické parametry, zejména ty související s časovou organizací a spektrálními charakteristikami, vykázaly lepší výsledky v klasifikaci než lingvistické rysy a dosáhly přesnosti 0,64 při rozlišení MCI od DLB. Klasifikace mezi HC a PD vykázala vysokou specificitu (0,90), ale nízkou sensitivitu (0,26), což odráží obtížnost detekce změn v řeči u časných stádií Parkinsonovy nemoci.

Explorativní analýza odhalila odlišné vzorce souvislostí mezi řečovými parametry a neuropsychologickými měřeními: časové parametry (délka pauz, průměrné trvání pauz) korelovaly především se skóre deprese, zatímco akustické parametry (zejména parametr zaměřený na směrodatnou odchylku kontury intenzity vzhledem k jejímu průměru) vykazovaly silnější vztahy s kognitivními funkcemi, především s pozorností. Nečekaný vztah mezi podílem složitých vět a vizuálně-prostorovým zpracováním naznačil možnou existenci sdílených nervových struktur.

Přestože výsledky klasifikace zatím nedosahují úrovně klinického využití, práce přináší cenné poznatky o vztahu mezi charakteristikami řeči a neurologickými poruchami a vytváří základ pro vývoj objektivních, neinvazivních diagnostických nástrojů. Budoucí výzkum by se měl zaměřit na longitudinální sledování změn řeči v průběhu onemocnění a integraci pokročilých akustických analýz.

KLÍČOVÁ SLOVA

Neurodegenerativní onemocnění, analýza řeči, prozódie, artikulace, strojové učení

ROZŠÍŘENÝ ABSTRAKT

Neurodegenerativní onemocnění představují významnou skupinu neurologických poruch, které postupně ovlivňují kognitivní a motorické funkce. Mezi tyto choroby patří mimo jiné Parkinsonova nemoc (PD), demence s Lewyho tělísky (DLB) a mírná kognitivní porucha (MCI), která může být předstupněm rozvoje demence. Jedním z často opomíjených, avšak klinicky cenných příznaků těchto onemocnění jsou poruchy řeči. Řeč je komplexní funkce, která odráží stav motorických i kognitivních mechanismů a její změny mohou signalizovat rané neurologické alterace. Vzhledem k pokroku v oblasti digitálního zpracování signálu a metod strojového učení se analýza řeči stává perspektivním nástrojem pro vývoj objektivních, neinvazivních biomarkerů využitelných v klinické diagnostice a monitoraci průběhu neurodegenerativních onemocnění.

Cílem této diplomové práce bylo vyvinout metodiku pro diferenciální analýzu tří vybraných neurodegenerativních onemocnění, MCI, DLB a PD, na základě analýzy akustických parametrů řeči. Doplňkově byla provedena explorativní analýza vybraných lingvistických parametrů s cílem posoudit jejich přínos pro rozlišení mezi klinickými skupinami. Práce si kladla za cíl nejen ověřit potenciál jednotlivých řečových parametrů pro diferenciální diagnostiku, ale také prozkoumat jejich vztah ke kognitivním a afektivním klinickým měřením.

V rámci rešeršní části byly zpracovány dosavadní poznatky o poruchách řeči u MCI, DLB a PD se zaměřením na akustické a jazykové změny popsané v odborné literatuře. Byly identifikovány relevantní parametry ovlivněné těmito onemocněními, především v oblasti časové organizace řeči, variabilitě základní frekvence, spektrálních charakteristikách a jazykové struktuře produkovaného projevu. Následně byla využita databáze spontánní řeči pacientů a zdravých kontrol, která zahrnovala nahrávky různých řečových úloh a přidružená klinická data. Byla provedena parametrizace řečového signálu s využitím akustických parametrů zaměřených zejména na délku a počet pauz, průměrnou délku pauzy, variabilitu intenzity (směrodatnoá odchylka kontury intenzity vzhledem k jejímu průměru – relSEOSD) a variabilitu základní frekvence. Lingvistické parametry zahrnovaly například podíl komplexních vět v řečovém projevu.

V analytické části byly nejprve provedeny explorativní statistické analýzy za účelem identifikace vztahů mezi řečovými parametry a klinickými měřeními kognitivních a afektivních funkcí. Výsledky korelační analýzy odhalily, že parametry časové organizace řeči, zejména délka pauz a průměrná délka pauzy, silněji korelují s depresivní symptomatikou. Tento nález je v souladu s předchozími studiemi, které poukázaly na psychomotorické zpomalení a jeho odraz v řečové produkci u depresivních pacientů. Naopak akustické parametry spektrální povahy, konkrétně relSEOSD, vykazovaly významné vztahy s výkonem v oblasti pozornosti, což podporuje hypotézu, že určité aspekty řečové produkce mohou být senzitivními indikátory kognitivní dysfunkce.

Zajímavým zjištěním bylo rovněž zjištění korelace mezi poměrem komplexních vět a skóre ve visuospaciálních testech, což není v dostupné literatuře často popisováno. Tento vztah může naznačovat existenci sdílených neuronálních mechanismů mezi syntaktickým zpracováním a visuospaciálním usuzováním, pravděpodobně na úrovni exekutivních kontrolních sítí.

V další fázi byly provedeny binární klasifikační analýzy s využitím strojového učení, kde jednotlivé řečové parametry sloužily jako prediktory rozlišení mezi klinickými skupinami a zdravými kontrolami. Nejvyšší klasifikační přesnosti bylo dosaženo při rozlišení MCI od DLB s využitím akustických parametrů (přesnost 0,64), zatímco jazykové parametry samostatně dosáhly nižší úrovně (přesnost 0,50). Výrazný rozdíl mezi specifitou a senzitivitou se projevil u rozlišení PD a zdravých kontrol, kdy specifita dosahovala vysokých hodnot (0,90), avšak senzitivita zůstala nízká (0,26). Tento jev lze vysvětlit faktem, že pacienti s PD v časném stadiu často nevykazují konzistentní změny v řečovém projevu, což je v souladu s výsledky předchozích studií.

Kombinace akustických a jazykových parametrů přinesla smíšené výsledky, u některých skupin došlo ke zlepšení klasifikace, například při rozlišení DLB a zdravých kontrol, což naznačuje, že synergie mezi těmito dvěma typy řečových ukazatelů může být přínosná. Nicméně celkový efekt této kombinace byl proměnlivý a vyžaduje další optimalizaci, ideálně s využitím pokročilejších metod zpracování.

Práce rovněž diskutuje limitace realizované studie, mezi něž patří především omezená velikost vzorku, což mohlo negativně ovlivnit statistickou sílu a robustnost klasifikačních modelů. Dalším omezením bylo technické zázemí, výpočetní výkon systému neumožnil použití vysoce výkonných modelů pro automatický přepis řeči, což mohlo snížit kvalitu lingvistických parametrů založených na transkripcích. Výsledky je proto třeba interpretovat s opatrností a doporučuje se jejich ověření na větších, multicentrických souborech.

Přestože dosažené klasifikační výsledky zatím neumožňují okamžité klinické využití, diplomová práce přinesla řadu důležitých poznatků o vztazích mezi řečovými parametry a neurodegenerativními onemocněními. Výsledky naznačují, že analýza řeči může sloužit jako cenný doplněk ke stávajícím klinickým metodám, především pro ranou detekci a diferenciální diagnostiku MCI, DLB a PD. Do budoucna by bylo vhodné realizovat longitudinální studie sledující vývoj řečových parametrů v průběhu onemocnění a jejich případné změny v reakci na terapeutické intervence. Další perspektivní oblastí je rozšíření o integraci s neurozobrazovacími daty za účelem objasnění neuronálních korelátů specifických řečových ukazatelů.

Tato práce tak přispívá nejen k rozvoji metodiky automatizované analýzy řeči ve zdravotnictví, ale také k lepšímu pochopení neurobiologických mechanismů, které se podílejí na produkci řeči u neurodegenerativních onemocnění.

MALCHER, Tomáš. *Diferenční analýza mírné kognitivní poruchy, demence s Lewyho tělísky a Parkinsonovy nemoci*. Master's Thesis. Brno: Brno University of Technology, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2024. Advised by doc. Ing. Jiří Mekyska, Ph.D.

Author's Declaration

Author:	Bc. Tomáš Malcher
Author's ID:	218400
Paper type:	Master's Thesis
Academic year:	2024/25
Topic:	Diferenční analýza mírné kognitivní poruchy, demence s Lewyho tělísky a Parkinsonovy nemoci

I declare that I have written this paper independently, under the guidance of the advisor and using exclusively the technical references and other sources of information cited in the paper and listed in the comprehensive bibliography at the end of the paper. AI language models ChatGPT (developed by OpenAI) and Claude 3.5 (developed by Anthropic) were also used for this thesis.

As the author, I furthermore declare that, with respect to the creation of this paper, I have not infringed any copyright or violated anyone's personal and/or ownership rights. In this context, I am fully aware of the consequences of breaking Regulation § 11 of the Copyright Act No. 121/2000 Coll. of the Czech Republic, as amended, and of any breach of rights related to intellectual property or introduced within amendments to relevant Acts such as the Intellectual Property Act or the Criminal Code, Act No. 40/2009 Coll. of the Czech Republic, Section 2, Head VI, Part 4.

Brno ....

author's signature*

*The author signs only in the printed version.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to the supervisor of this master's thesis, doc. Ing. Jiří Mekyska, Ph.D., for his expert guidance, patience, and friendly yet very professional approach.
Také bych rád poděkoval rodičům a babičce s dědou za to, že mě po celou dobu studia podporovali, jak jen mohli.

Contents

List of Figures

List of Tables

Introduction

Speech disorders are a common but often overlooked symptom of neurodegenerative diseases such as Parkinson's disease (PD), Lewy body dementia (DLB) and mild cognitive impairment (MCI). Subtle changes in voice can reflect early disruptions in motor and cognitive processes, offering a unique, non-invasive window into brain function. With advances in digital signal processing and machine learning, speech analysis is becoming an increasingly valuable tool for identifying objective biomarkers to support clinical diagnosis and monitor disease progression.

Most existing research in this area has focused on individual diseases or isolated speech characteristics, often within small, homogeneous samples. While acoustic parameters have shown particular promise in differentiating neurodegenerative conditions, few studies have systematically compared multiple disorders or examined how speech parameters relate to clinical cognitive and affective measures. Additionally, the potential of combining acoustic analyses with linguistic features has received limited attention, leaving room for exploratory investigation.

Aims of this thesis

The primary aim of this thesis is to develop a methodology for the differential analysis of MCI, DLB, and PD using acoustic speech parameters. To extend the scope of this work, a complementary exploratory analysis of selected linguistic features was also included. The specific goals were:

1. Conduct a literature review on speech and voice disorders in MCI, DLB, and PD.
2. Familiarize with a provided database of speech recordings and associated clinical data.
3. Design and implement a parameterization pipeline based on selected acoustic biomarkers, with an exploratory set of linguistic parameters.
4. Perform exploratory and classification analyses to identify discriminative features and their relationships with clinical measures.

Structure

This thesis is organized into several chapters. The Introduction outlines the context of the study, current limitations in the field, its aims, and the overall structure. The Neurodegenerative disorders chapter presents existing research on speech and voice disorders in MCI, DLB, and PD, identifying relevant speech parameters for analysis. The Methodology chapter describes the provided speech database, the parameterization process, and the selection of acoustic and linguistic features and their calculation method. The Implementation chapter presents the process behind the programming, tools, and scripts developed for this thesis. The Results chapter reports findings from exploratory analyses and machine learning classification models, including correlations between speech parameters and clinical measures. The Discussion contextualizes these findings within existing literature, highlighting consistencies, discrepancies, and novel contributions. Finally, the Conclusion and Future Directions summarize the key outcomes of the work and propose avenues for further research in this field.

1 Neurodegenerative disorders

MCI, PD and DLB are three disorders that share similar pathological features, mainly the impairment of cognitive and executive functions of patients. The decline of those functions has an effect on various parts of patients' lives. [1, 2]

Even after receiving a diagnosis, patients with neurodegenerative disorders can continue to live for many years or even decades as the disease gradually progresses. During this time, however, their quality of life often progressively declines. In the later stages of certain conditions, such as PD and various forms of dementia, the deterioration of cognitive and executive functions can become particularly severe. This decline may result in a range of complications, including respiratory and circulatory issues, which may contribute to a decline in overall health status, potentially leading to death. Consequently, managing these disorders becomes increasingly challenging, emphasizing the need for complex care and support for both patients and their families. [3, 4]

1.1 Parkinson's disease

PD is a neurodegenerative disease, mostly prominent among people above 60 years of age. Some may experience early onset of the disease before the age of 50. Symptoms usually start slowly and are barely noticeable, such as a light tremor in the hands of the patient. Age is one of the more important risk factors related to developing PD. However, environmental factors (e.g. water pollution and exposure to pesticides) and other factors (e.g. smoking, exercising) have been found to play a role in the development of PD. The disease is also more prevalent in men, with roughly a 3:2 ratio to women. [5, 6]

While many parts of the brain and nervous system are affected by PD, the disease starts from loss of dopamine–producing neurons in the area near the base of the brain called the *substantia nigra*. The main cause behind neuronal death is the accumulation of misfolded proteins, e.g. α-synuclein. The buildup of these misfolded proteins overwhelms the cell's clearance systems, leading to neurodegeneration. Since dopamine acts as a neurotransmitter and plays a role in communication between the brain and muscles, with its shortage, the motor functions decay and that leads to symptoms characteristic for PD. [7]

- **Bradykinesia**: a noticeable slowness and lack of spontaneous movement, including diminished arm swing, fewer facial expressions, reduced gestures, small handwriting (micrographia), challenges with turning in bed. Additionally, there is a gradual decline in the speed and range of voluntary repetitive movements, such as finger tapping and hand gripping.

- **Resting tremor**: (4–6 Hz) is frequently observed in the limbs, in the hands it is often described as pill-rolling tremor, due to its resemblance to the motion of rolling a small object between the fingers. Also present in the lips, chin, or jaw, though it is uncommon in the head. The amplitude of the tremor decreases or disappears during purposeful movements.
- **Rigidity**: stiffness of muscles characterized by resistance to passive movements of similar intensity in opposing muscle groups, known as lead-pipe rigidity. When this increased muscle tone is accompanied by a tremor, the resistance exhibits a ratchet-like quality, referred to as cogwheel rigidity.

The disease is hard to detect early, because the first impairments to motor functions usually show after about 70–80 % of dopaminergic neurons have died. There are non-motor function symptoms that can precede the onset of motor features, such as: hyposmia (loss of smell, 90 % of patients), sleep disorders (REM sleep behavior disorder, insomnia, vivid and usually unpleasant dreams), neuropsychiatric features (apathy, anxiety, panic attracts and depression) and mild cognitive impairment. At later stages of the disease, roughly 30 % of patients develop dementia. [6, 8]

Medication

Levodopa (L-dopa) is one of the most effective treatments for PD. It serves as a precursor to dopamine, helping to replenish dopamine levels in the brain. To prevent it from converting into dopamine outside the brain, levodopa is often combined with carbidopa or benserazide, which block this conversion in the rest of the body. This combination allows more levodopa to reach the brain, where it can be converted into dopamine, effectively helping to reduce PD symptoms. [8]

1.1.1 Lewy bodies

Frederick Lewy first described what we now know as Lewy bodies in 1912. They are abnormal protein inclusions formed inside nerve cells. When α-synuclein clumps together, it forms a Lewy body, which disrupts normal cellular functions and contributes to the death of neurons. [9, 10]

1.1.2 Diagnosis

Diagnostic accuracy has improved in recent years due to new validated clinical diagnostic criteria. However, with overlapping symptoms and features, the diagnosis is still very challenging, even in fully manifested disease cases. [6]

The first step in the diagnosis process is establishing whether the patient has "parkinsonism". To declare that the patient has parkinsonism, bradykinesia must be present, with at least one of the other two elements. The second step is to distinguish whether the cause of parkinsonism is PD or another neurodegenerative disorder such as a type of dementia. [11, 12]

1.2 Mild cognitive impairment

Mild cognitive impairment is a disorder in which a person's cognitive functions degrade faster than expected for their age and educational level. However, it does not notably influence their day-to-day activities. In some cases, it can act as a state between natural degradation of cognitive and executive function related to aging and more serious impairment caused by another neurodegenerative disease, such as dementia. Not everyone who has developed MCI will progress to the state of dementia. In some cases, patients' cognitive and executive functions may become stable or even improve back to their normal levels. [13, 14]

The brain of MCI patients undergoes similar changes to the brains of patients with other neurodegenerative disorders, e.g. AD, PD or another type of dementia, although to a much lesser extent. These changes include, among other things, Lewy bodies. The main risk factors for MCI are: increasing age and having an APOE e4 gene, which is linked to AD. [14, 15]

1.2.1 Diagnosis

There have been many attempts in history to describe the regression of cognitive abilities in relation to aging. In 1982, two clinical staging systems were published, one of them is called **g**lobal **d**eterioration **s**cale for aging and dementia (GDS). This scale differentiates between seven clinical stages. GDS Stage 1 describes healthy people without any kind of cognitive impairment, while Stage 2 consists of those with self-perceived impairments, such as trouble remembering names. Stage 3 is where the term mild cognitive impairment was first used, and it describes people with subtle impairments in cognition that also may have deficits in their executive functions. Higher stages describe more severe progression of MCI. [13]

Nowadays, the classification of MCI can vary depending on the cognitive tests used and how "impairment" in day-to-day activities is defined. Identifying MCI relies on the sensitivity and specificity of these assessments, along with normative data and estimates of an individual's cognitive function prior to any decline. In cases where clear collateral history is not sufficient, cognitive decline may be inferred from an individual's previous achievements or performance levels. To diagnose MCI, neuropsychological assessments often use a threshold of 1.5 standard deviations below the average performance of the population. While this approach is sensitive to detecting decline, it may reduce specificity. Over the years, the criteria for diagnosing MCI have evolved, reflecting a growing understanding that neurodegenerative diseases can develop long before clinical symptoms appear. Moreover, factors such as age, socioeconomic status, genetic predisposition, education, and mid-life health risks have all been linked to cognitive function in older adults. [1]

1.3 Dementia with Lewy bodies

Dementia as a whole is a progressive neurodegenerative disease that affects mainly people above 65 years old; the prognosis can vary between individual patients. Generally, it causes shorter lifespans, differing levels of disability, and problems with memory and speech. Dementia with Lewy bodies is one of the most frequent types of dementia. It is characterized by the presence of Lewy bodies, mainly inside the brain and other parts of the nervous system. DLB mainly affects cognitive and executive functions. [16]

1.3.1 Symptoms and diagnosis

For patients with DLB, the progressive decline of their cognitive functions starts early, usually after the age of 55. In some cases, symptoms and features of PD, e.g. parkinsonism, can simultaneously develop alongside DLB. [9]

Criteria for clinical diagnosis of DLB are as follows:

- **Central feature** (essential for DLB diagnosis): dementia has to be present. In the early stages, significant memory deficits may not occur. However, executive functions, attention and visuoperceptual ability impairments may be prominent.
- **Core clinical features**:
  - Fluctuating attention, alertness and cognition,
  - Recurrent detailed visual hallucinations,
  - Rapid eye movement sleep behavior disorder, may precede cognitive decline by many years,
  - Features of parkinsonism. While for diagnosis of parkinsonism in PD bradykinesia and either resting tremor or rigidity must be present, in the case of parkinsonism in DLB only one or more features have to be present.

- **Supportive clinical features**: sensitivity to antipsychotic agents, repeated falls, hypersomnia (excessive sleepiness), hyposmia (decreased sense of smell), anxiety, depression, apathy and many more.
- **Biomarkers**: Two categories of biomarkers have been declared; indicative and supportive biomarkers. In patients' bodies, they can be discovered by various types of medical scans, such as SPECT (single-photon emission computed tomography) or PET (positron emission tomography) imaging or CT (computed tomography) and MRI (magnetic resonance imaging) scans.

After evaluation of the above-mentioned parameters: DLB is **probable** in the presence of two or more clinical features or, with only one clinical feature and one or more indicative biomarkers, and **possible** in presence of only one clinical feature, or one or more indicative biomarkers, however, without any clinical features. [17, 18]

1.4 Parkinson's disease dementia

Parkinson's disease dementia (PDD) is closely related to DLB, but distinguished by the timing of its onset. In PDD, cognitive decline occurs at least one year after PD has been diagnosed, whereas in DLB, cognitive symptoms appear before or around the same time as motor symptoms. Both conditions share similar clinical features, such as fluctuating cognition, recurrent visual hallucinations, REM sleep behavior disorder, and parkinsonism. The underlying cause for PDD, just like in DLB, is the accumulation of α-synuclein protein in the brain. [19]

2 Motor speech disorders

Speech is a result of complex interactions of neurocognitive, neuromuscular, neuromotor, and musculoskeletal processes. As the nervous system deteriorates, the mechanisms responsible for speech production also decline. Different neurodegenerative disorders may cause different speech disorders, however, symptoms may overlap. Motor speech disorders refer to a category of speech conditions caused by neurological damage that impacts the processes involved in the execution of speech movements. These disorders typically manifest as either dysarthria (difficulties with the physical production of speech), or apraxia of speech (affects the coordination and planing of speech movement without muscular impairment). [20]

2.1 Speech impairments in PD

Hypokinetic dysarthria (HD) is a distinct motor speech disorder associated with dysfunction in the *basal ganglia* control circuits. It can affect various levels of speech production, including respiration, phonation, resonance, and articulation, though its most noticeable impacts are on voice, articulation, and prosody. The disorder is primarily characterized by rigidity, reduced movement force, limited movement range, and slow individual movements, although repetitive movements may sometimes appear rapid. The reduced range of motion is a major feature, which is why the condition is termed "hypokinetic". Identifying this dysarthria is useful for diagnosing neurological conditions, especially since it is strongly linked to *basal ganglia* pathology, with PD being the most common associated condition. However, other diseases can also lead to HD. [20]

2.1.1 Respiration

Individuals with HD may experience various respiratory issues. These can include an increased breathing rate and uncoordinated movements between the chest and diaphragm, where muscles involved in inhalation and exhalation do not work in harmony. Additionally, there is often a reduced range of motion in respiratory muscles, leading to shallow breaths and poorly controlled exhalations during speech. These respiratory challenges can result in short breathing cycles, which may contribute to speech errors such as rushed or rapid speech, as well as soft, breathy vocal quality due to inadequate breath support. [21]

2.1.2 Phonation

Acoustic and physiological studies have provided valuable insights into laryngeal function in speakers with HD. These studies generally support the findings from perceptual analyses, offering deeper understanding of the mechanisms behind the abnormal speech and voice characteristics observed in this condition. [20]

Fundamental frequency ($F0$) and intensity and their variablility:

While abnormal pitch is typically not a significant perceptual feature of HD, studies have shown a tendency for increased $F0$ in individuals with PD. This increase, however, does not always reach statistical significance when compared to age-matched controls and is often more evident in men than women; in fact, some research indicates that women may exhibit a decrease in $F0$. [20]

In terms of intensity, individuals with PD often display reduced loudness during various speech tasks. However, they can still adjust their loudness based on the distance of their listeners, suggesting that while their physiological ability to regulate loudness remains intact, their motor control is compromised. Additionally, many patients tend to misjudge their loudness, frequently perceiving themselves as louder than they actually are. This misperception, along with further challenges when multitasking (e.g. speaking while performing visual tasks) increases their speech difficulties. [20]

Acoustic and physiological studies of HD reveal that variability in $F0$ and intensity is often task-specific. Findings from vowel prolongation, spontaneous speech, and pitch glide exercises show both increased and decreased variability, reflecting the condition's characteristic features like monopitch and monoloudness. Notably, during tasks such as vowel prolongation, speakers often exhibit greater variability in both $F0$ and amplitude, indicating impaired laryngeal control. Long-term average spectrum (LTAS) analysis has proven effective in identifying these speech abnormalities, providing more nuanced insights than simpler acoustic measures. [20]

Voice tremor

Voice tremor is generally not a key perceptual feature of HD, it resembles tremors found in individuals without speech disorders. However, visual evidence of laryngeal or arytenoid tremors has been noted in patients with Parkinson's disease during endoscopy. While tremor frequencies of 4 to 7 Hz have been observed, they are not consistently identified in assessments and are not necessary for diagnosis. Such tremors may be more common in later stages of PD, often showing variations in amplitude and frequency. Notably, one study reported a high-frequency flutter in a female patient that was more evident in amplitude modulation than in frequency modulation, which is significant for differentiating hypokinetic from flaccid dysarthria. [20]

Maximum phonation time

The maximum phonation time for vowel prolongation may not significantly differ from normal phonation levels, perhaps due to testing methods and inherent variability within and across individuals. However, notable declines have been observed in both the maximum and average durations of sustained vowel prolongations in speakers with PD, occurring over periods ranging from 3 to 36 months. Thus, a reliably obtained maximum phonation time can provide insights into changes over time in individuals with HD, but it may not be sensitive enough to detect the disorder itself. [20]

Jitter, shimmer, motor control and airflow

Speakers with PD often show elevated levels of jitter and shimmer, which may indicate neuromuscular control issues during voice production. Increased shimmer values are linked to perceptions of breathiness due to vocal fold bowing and airflow turbulence. However, these abnormalities do not consistently appear in all individuals with HD. While some studies indicate differences in shimmer for females with PD, jitter and shimmer measures are less effective in distinguishing PD males from healthy controls, suggesting limited reliability in these acoustic indicators. [20]

Furthermore, individuals with PD demonstrate significant impairments in laryngeal control, including delays in initiating phonation, leading to perceived silences and voiceless transitions between vowels and consonants. Continuous voicing in voiceless contexts also indicates challenges in speech production. Videolaryngoscopy studies reveal vocal fold bowing, which correlates with breathiness and reduced intensity. Asymmetries in vocal fold movement and signs of muscle rigidity suggest that these laryngeal issues are tied to neuromuscular deficits in HD, emphasizing the complexity of speech impairments associated with PD. [20]

2.1.3 Resonance

In HD, resonance issues, such as hypernasality, are generally mild but can be present in some cases. Studies show increased nasal airflow, reduced velar movement, and nasalization across syllables, suggesting velopharyngeal dysfunction. This may be due to slow or rigid muscle movement. While perceptual ratings don't always correlate with acoustic measures, some speakers exhibit mild to moderate hypernasality. Instrumental measures like nasal accelerometry can be more sensitive in detecting these abnormalities, even when not perceptually obvious. Overall, velopharyngeal dysfunction can contribute to weak intraoral pressure during speech. [21, 20]

2.1.4 Articulation

In HD, articulation errors are common, with imprecise consonants being one of the most frequently observed. This imprecision often stems from limited movement in the articulators, which leads to distorted and inaccurate speech sounds. For example, stop consonants may resemble fricatives because the articulators are unable to fully block airflow, while fricatives can take on a "mushy" quality due to a larger-than-normal point of constriction. Affricate consonants may reflect both of these errors, as they involve both a stop and a fricative-like release phase. These types of articulation errors are sometimes referred to as "articulatory undershoot". In addition to these articulation challenges, individuals with HD may experience two unusual dysfluencies: repeated phonemes and palilalia. Repeated phonemes typically occur at the beginning of speech or after a pause, characterized by rapid repetitions that may sound like a prolonged vowel due to minimal movement of the articulators. Palilalia, on the other hand, involves compulsive, progressively faster repetition of a word or phrase, often fading into a blurred mumble. This phenomenon is similar to the festinating gait seen in PD, where individuals struggle to stop movement once it begins, and some hypokinetic speakers may have similar difficulty halting speech once initiated. [21]

2.1.5 Prosody

Monopitch, reduced stress, and monoloudness are common speech features in people with hypokinetic dysarthria. These issues arise from a limited range of motion and weak contractions in the muscles that control the larynx. The impact is most noticeable during tasks that require changes in pitch and loudness, like having a conversation, reading aloud, or trying to convey emotions verbally. Inappropriate silences can also occur, often due to difficulty initiating motor responses. These pauses, typically happening at the start or between sentences, can last a few seconds and may give the impression that the speaker hasn't heard or has lost their train of thought. Speech rate can also vary significantly. Some individuals may speak faster than usual, likely because they struggle to stop voluntary movements once they've started. This can result in slurred or blurred speech sounds, making consonants hard to understand. More commonly, people experience short bursts of rapid speech, where a pause is followed by a quick string of words. While some speak faster, others may actually speak more slowly, showing that the speech rate in hypokinetic dysarthria can differ from person to person. [21]

2.2 Speech impairments in MCI

Speech impairments associated with MCI prominently involve deficits in lexical, semantic, and executive functions. Individuals with MCI often experience word-finding difficulties, which lead to reduced semantic and phonemic fluency, as well as diminished lexical richness. These language challenges can manifest as decreased syntactic complexity and lower topic coherence, frequently becoming noticeable before a formal diagnosis. While articulatory abilities are generally maintained in the early stages, the progression of MCI typically exacerbates these language deficits. Conversational speech patterns further reveal cognitive difficulties in MCI patients. Key features include a slower speech rate, a higher frequency of silent pauses, and shorter segments of speech. These temporal characteristics suggest impairments in memory, attention, and executive functions, all of which are essential for effective communication. [24]

2.3 Speech impairments in DLB

DLB typically presents a blend of symptoms associated with AD and PD. From AD, it shares cognitive issues like reduced judgment, confusion, and changes in behavior, while from PD, it borrows motor symptoms such as tremors, difficulty with balance, and muscle rigidity. This overlap is not limited to general symptoms but extends to speech disorders as well. Speech impairments in individuals with DLB display traits from both conditions, mirroring the cognitive and motor disruptions caused by each. Interestingly, among the speech characteristics, DLB patients' speech most closely aligns with that of people suffering from PDD, highlighting the significant connection between these conditions in terms of motor-related speech dysfunction. [22]

A substantial proportion of DLB patients (48.8%) experience freezing of speech (FOS), marked by frequent pauses and communication difficulties, which is notably more common in DLB than in other neurodegenerative disorders. Those in the FOS-positive group were older and demonstrated more advanced dementia stages according to the Clinical Dementia Rating Scale (CDR). [23]

2.3.1 Study comparing speech in DLB and PD

A significant investigation in this area focused on a task where participants were asked to read a story and then retell it as if explaining it to a young child.

Words per minute

The results indicated that the speech rate, quantified as words per minute (WPM), for individuals with Dementia with Lewy Bodies (DLB) and PDD was roughly half that of healthy controls (HC). Notably, this reduced speech rate was also significantly lower than that of those with PD. [25]

Silent pauses

In their free retelling, patients with DLB/PDD spent more than one-third of their speaking time in silent pauses between segments of speech, which was significantly longer than the pauses noted in PD speakers. Additionally, an analysis using modified WPM—calculated by excluding segments of speech that included pauses longer than two seconds—showed that the speech patterns of DLB patients could be distinctly identified from those of other groups. This finding suggested an overall slowdown in their speech. [25]

Articulation errors

The study also evaluated articulation by counting the number of articulation errors per 100 words spoken. It was found that individuals with DLB and PDD made more articulation errors compared to both healthy individuals and those with PD. This increase in errors was correlated with the motor scores measured by the Unified Parkinson's Disease Rating Scale (UPDRS), implying a relationship between motor ability and speech performance. [25]

Grammar

Another critical aspect of the research was the examination of grammatical accuracy. This assessment included analyzing the correctness of utterances, the complexity of sentence structures, and the production of necessary determiners. The most pronounced deficits were found in patients with DLB, underscoring the fact that fluent speech relies heavily on the ability to quickly formulate phrases and sentences while considering the grammatical connections between words. The findings suggest that grammatical challenges play a significant role in the diminished fluency observed in DLB and PDD patients. Furthermore, neuropsychological testing revealed more considerable differences when compared to patients with PD, identifying impairments in executive functions, memory, semantics, and comprehension. These cognitive deficits are thought to exacerbate the reduced speech rates noted in individuals with DLB. Consequently, speech issues associated with DLB resemble those seen in AD, particularly in terms of difficulties with word retrieval and linking perceived words to their appropriate meanings. [25]

2.4 Summary and comparison of features of speech impairment in PD, MCI, and DLB

Speech Rate

In PD, speech is typically slow and may accelerate unexpectedly. In DLB, the speech rate fluctuates and may include occasional disfluencies. MCI usually presents with a normal speech rate.

Prosody

In PD, speech often sounds monotone, with reduced pitch and loudness. In DLB, prosody varies, with sudden changes in volume and pitch. MCI generally features normal prosody.

Articulation

PD is often marked by imprecise, mumbled, and blurred articulation. DLB may involve mild dysarthria, though less severe than in PD. Articulation in MCI is generally normal.

Voice Quality

PD is characterized by a breathy and hypophonic (quiet) voice. In DLB, loudness is reduced and may fluctuate with attention. MCI typically shows preserved voice quality, though mild vocal strain is possible.

Language

In PD, patients tend to use simpler language structures and show reduced spontaneity. DLB often involves language comprehension difficulties and frequent FOS. MCI is associated with frequent pauses due to word-finding difficulties.

3 Methodology

3.1 Database

The database used in this thesis consists of recordings of patients categorized into groups according to their medical condition: PD, MCI, DLB and healthy control (HC) group. The participants in this study range in age from 52 to 84 years, with an average age of 68. The average age along with the standard deviation for participants, categorized by different groups and genders, is illustrated in figure 3.1. The overall age distribution across genders is depicted in figure 3.2. A more detailed demographic distribution based on age can be found in table 3.1. The database also contains various information about individual participants, such as educational level (1 = primary school, 2 = vocational school, 3 = high school, 4 = university), years of education, and z–scores for memory, visuo–spatial abilities, attention, and executive functions. More detailed distribution based on these parameters can be found in table 3.2 It is also important to note that participants with PD were examined approximately one hour after receiving a dose of L-Dopa. For these patients, the L-Dopa equivalent daily dose is also recorded in the database.

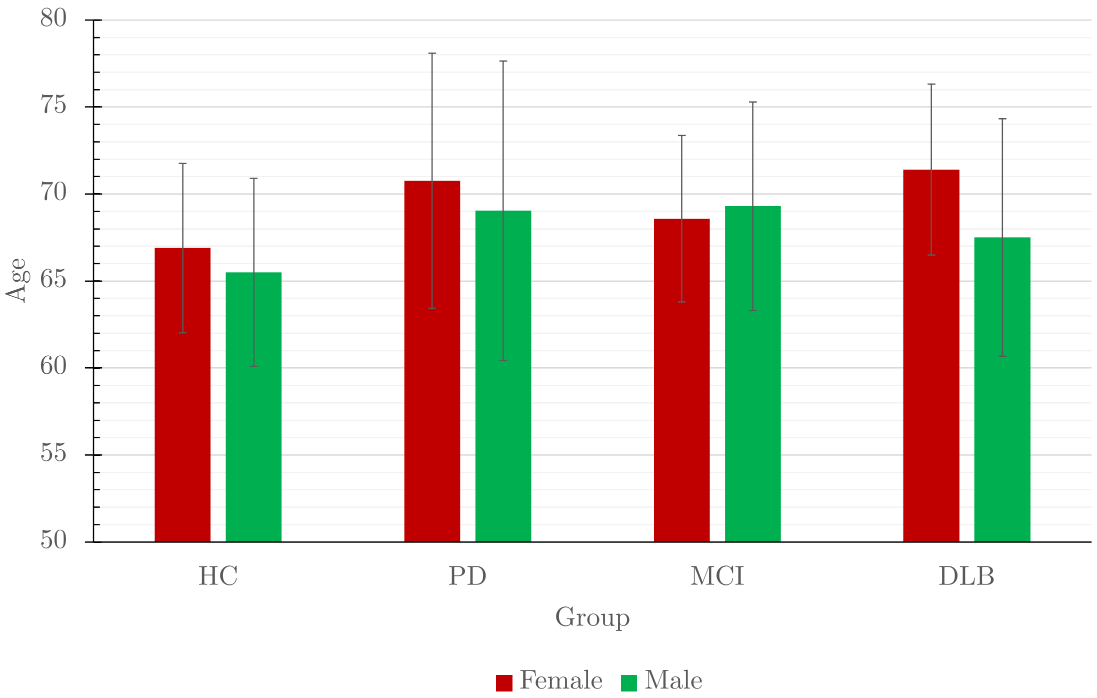

Fig. 3.1: Average age of participants.

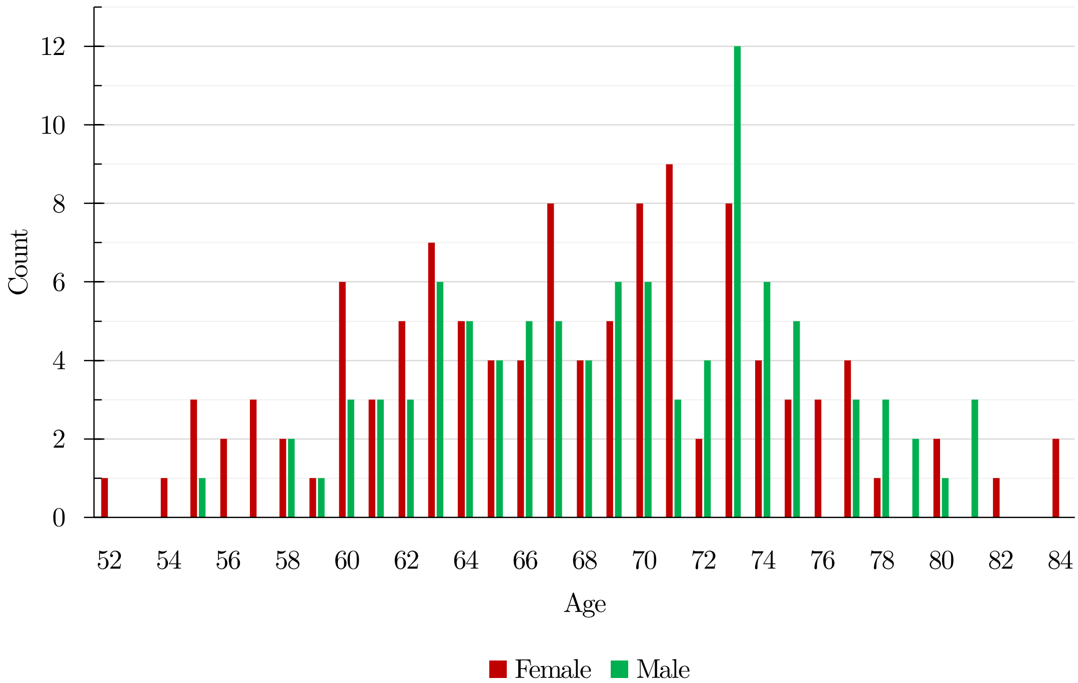

Fig. 3.2: Age distribution across genders.

Individuals with prodromal DLB were diagnosed using the criteria outlined by McKeith et al. (2020). [26] Participants were enrolled from the St. Anne's University Hospital in Brno, or through the Applied Neuroscience at the Central European Institute of Technology, as part of the CoBeN project (EU Horizon 2020 research and innovation program, Marie Skłodowska-Curie grant no. 734718). This research was also supported by project no. LX22NPO5107 (Ministry of Education, Youth and Sports) funded by the European Union's Next Generation EU initiative, as well as Czech Ministry of Health projects no. NU20-04-00294 and no. NV16-30805A. All participants provided informed consent, and the study was approved by the local ethics committee.

3.2 Acoustic analysis

The full database includes a variety of speaking tasks labeled as TSKn, where n represents the task number. TSK1 involves a monologue; during this task, patients were asked to deliver an uninterrupted monologue lasting at least 90 seconds on topics such as their hobbies, family, job, daily activities, and similar subjects. TSK2 consists of reading a short text, which patients could read for themselves in advance. TSK3–5 focus on sustained phonation (no longer than 5 seconds) of vowels (/a/,

Table 3.1: Age distribution by group and gender (F – female, M – male).

Group	HC			PD			MCI			DLB		
	All	**F**	**M**	**All**	**F**	**M**	**All**	**F**	**M**	**All**	**F**	**M**
Count	68.0	38.0	30.0	60.0	24.0	36.0	42.0	23.0	19.0	53.0	29.0	24.0
Average	66.1	66.9	65.5	69.0	69.3	68.6	69.0	68.6	69.3	69.3	71.2	67.8
std	5.2	4.9	5.4	8.0	7.5	8.7	5.5	4.8	6.0	6.3	5.1	6.8
Min	55.0	58.0	55.0	52.0	53.0	52.0	55.0	60.0	55.0	54.0	59.0	54.0
Q1	62.0	63.0	62.0	62.8	63.0	61.0	65.3	65.5	65.5	66.0	68.0	64.0
Median	66.0	66.0	65.5	70.0	70.0	70.0	69.5	70.0	70.0	70.0	71.5	68.0
Q3	70.0	71.8	69.5	75.0	75.0	74.8	73.0	72.5	73.0	74.0	75.0	73.0
Max	77.0	75.0	77.0	84.0	81.0	84.0	80.0	80.0	80.0	84.0	81.0	84.0

/i/, /u/), while TSK6 requires sustained phonation of /a/ at a constant comfortable pitch for as long as possible. TSK7 is a diadochokinetic task involving a rapid steady repetition of the /pa/–/ta/–/ka/ syllables. Finally, TSK8–17 involve the repetition of polysyllabic words according to clinicians. For the purpose of this thesis, the focus will be solely on TSK1.

While there are many acoustic parameters and features, not all of them can be reliably identified from a patient's monologue. For example, tasks focused on prolonged phonation of phonemes are better suited for detecting features such as airflow insufficiency, jitter, shimmer, and jaw tremor. Diadochokinetic tasks, on the other hand, focus on alternating motion rate, acceleration of diadochokinetic rhythm, and its stability. Although these features may also be present in patients' monologues, they are less distinct compared to tasks specifically designed to highlight them.

Vowel Articulation Index (VAI)

From samples of continuous speech, decreased tongue movement can be assessed using the VAI parameter, which manifests as imprecisely articulated vowels. This measure is based on formant centralization, defined as:

$$\mathrm{VAI} = \frac{(F1_a + F2_i)}{(F1_i + F1_u + F2_a + F2_u)}, \tag{3.1}$$

where $F1$ represents the first formant and $F2$ the second, the subscript of each formant represents a vowel.

Table 3.2: Summary statistics for various parameters and scores based on groups.

Variable	Group	Average	std	Min	Q1	Median	Q3	Max
MoCA	HC	27.0	1.7	22.0	26.0	27.0	28.0	30.0
	PD	24.5	2.7	18.0	22.0	24.0	27.0	29.0
	MCI	23.2	4.2	7.0	21.3	24.5	25.8	29.0
	DLB	24.1	3.2	15.0	22.3	25.0	26.0	29.0
Education type	HC	3.5	0.6	2.0	3.0	4.0	4.0	4.0
	PD	3.3	0.7	1.0	3.0	3.0	4.0	4.0
	MCI	2.9	0.9	1.0	2.0	3.0	4.0	4.0
	DLB	2.8	0.8	1.0	2.0	3.0	3.0	4.0
Education length	HC	16.1	2.8	12.0	13.0	17.0	18.0	26.5
	PD	15.2	2.8	9.0	13.0	15.5	18.0	22.0
	MCI	14.4	3.4	9.0	12.0	13.0	17.0	23.0
	DLB	13.9	3.3	9.0	12.0	13.0	15.0	25.0
Memory z-score	HC	0.7	0.8	-1.1	0.3	0.8	1.2	2.8
	PD	0.1	0.9	-1.9	-0.5	0.2	0.6	2.2
	MCI	-0.9	0.8	-3.0	-1.2	-0.9	-0.3	0.2
	DLB	-0.4	0.8	-2.1	-0.9	-0.4	0.1	1.1
Visuospatial z-score	HC	0.3	0.6	-1.3	0.2	0.6	0.6	1.1
	PD	-0.2	1.0	-2.2	-0.7	-0.3	0.4	1.2
	MCI	-0.2	1.0	-2.2	-0.8	-0.3	0.6	1.1
	DLB	-0.5	1.1	-2.2	-1.3	-0.4	0.6	1.1
Attention z-score	HC	0.2	0.5	-1.0	0.0	0.3	0.5	1.6
	PD	-0.7	0.6	-2.2	-1.1	-0.7	-0.2	0.7
	MCI	-0.4	0.6	-2.2	-0.9	-0.4	0.0	0.8
	DLB	-0.3	0.8	-2.7	-0.8	-0.3	0.2	1.2
Executive function z-score	HC	0.3	0.7	-1.6	-0.1	0.3	0.6	1.9
	PD	-0.5	0.7	-1.8	-1.1	-0.4	0.0	1.0
	MCI	-0.7	0.7	-2.0	-1.1	-0.7	-0.3	1.1
	DLB	-0.4	0.7	-1.9	-0.8	-0.4	0.1	1.1
GDS-corr	HC	4.2	3.5	0.0	2.0	4.0	6.0	16.0
	PD	7.8	6.4	0.0	2.0	6.0	10.0	24.0
	MCI	4.3	4.8	0.0	1.0	3.0	6.0	24.0
	DLB	7.9	5.7	1.0	4.0	7.0	10.0	24.0

MoCA: Montreal Cognitive Assessment; GDS–corr: Geriatric Depression Scale

Vowel Space Area (VSA)

The VSA parameter, just like VAI, assesses articulatory precision. However, by a different way to VAI, by measuring the area of the vowel space formed by plotting the first and second formant frequencies ($F1$, $F2$) of corner vowels, typically /a/, /i/, and /u/. The area is calculated using the formant values of these vowels in the $F1$–$F2$ plane.

relF1SD, relF2SD

Rigidity of tongue and jaw can be assessed by parameters relF1SD, relF2SD, they describe standard deviation of first and second formant relative to its mean.

relSEOSD

The parameters outlined above specifically address the articulatory components of speech production. In examining parameters that focus on the prosodic aspects of speech, monoloudness is measured using relSEOSD, which is defined as the standard deviation of the intensity contour relative to its mean, after silences longer than 50 milliseconds have been excluded.

EEVOL

To assess unstable mean loudness, we use EEVOL (Energy Evolution). This parameter measures the slope of the intensity contour, calculated from overlapping 200 ms segments (with a 100 ms overlap), while also excluding silences that exceed 50 milliseconds. The results are then normalized by the total duration of the speech.

relF0SD

Finally, monopitch is evaluated using relF0SD parameter, which quantifies the standard deviation of the $F0$ contour relative to its mean.

When looking at the parameters focused on the rythm of speech, the following are being used:

Speech Time

Speech Time measures the total duration of spoken segments in the speech sample, excluding pauses and silences. It reflects how much time the speaker actively produces speech during the recording.

Pause Time

Pause Time quantifies the total duration of silent intervals or pauses in the speech sample. This parameter captures speech fluency and the speaker's pausing behavior, which can be important for detecting speech impairments or cognitive load.

Words Per Minute (WPM)

WPM is a measure of speech rate calculated by dividing the total number of words spoken by the total speech time (in minutes). It indicates how quickly the speaker articulates words and can reflect speech motor control or conversational tempo.

Total Pauses

Total Pauses counts the number of distinct silent intervals or pauses longer than a predefined threshold within the speech sample. This metric complements Pause Time by indicating the frequency of interruptions in speech flow.

Average Pause

Average Pause computes the mean duration of all pauses detected in the speech sample. It reflects typical pause length and can be informative about speech rhythm and fluency.

3.3 Linguistic analysis

The following parameters can be used to analyse and assess various aspects of lexical diversity, syntactic complexity, and speech fluency from the transcribed speech data.

Total Sentences

The total number of sentences in each sample is determined by segmenting the text using standard punctuation delimiters (periods, exclamation marks, and question marks).

Type-Token Ratio (TTR)

Type-Token Ratio is calculated as the ratio of the number of unique word types (tokens) to the total number of word tokens in the text:

$$\mathrm{TTR} = \frac{\text{Number of unique words}}{\text{Total number of words}}. \tag{3.2}$$

This metric reflects lexical diversity but is sensitive to text length.

Moving-Average Type-Token Ratio (MATTR)

To mitigate the sensitivity of TTR to text length, a Moving-Average Type-Token Ratio (MATTR) is computed by sliding a fixed-size window (e.g., 50 words) across the text. Within each window, the TTR is calculated, and the average across all windows is taken as the MATTR value, providing a more stable measure of lexical diversity.

Unique Word Ratio (Hapax Legomena Ratio)

The ratio of words occurring exactly once in the text, known as hapax legomena, to the total word count is calculated:

$$\text{Unique Word Ratio} = \frac{\text{Number of hapax legomena}}{\text{Total number of words}}. \tag{3.3}$$

This ratio serves as an additional indicator of lexical variety.

Noun-Verb Ratio

The ratio of the total count of nouns to the total count of verbs in the sample is calculated after part-of-speech tagging. This ratio provides insight into the syntactic composition and balance between nominal and verbal expressions.

$$\text{Noun-Verb Ratio} = \frac{\text{Count of nouns}}{\text{Count of verbs}}. \tag{3.4}$$

Average Sentence Length

Average sentence length is calculated by dividing the total number of words by the total number of sentences:

$$\text{Average Sentence Length} = \frac{\text{Total words}}{\text{Total sentences}}. \tag{3.5}$$

This metric captures overall sentence complexity and fluency.

Sentence Length Variance

The variance of sentence lengths across the sample can be computed to quantify variability in sentence structure, providing a measure of syntactic diversity and complexity.

Complex Sentence Ratio

The proportion of complex sentences, defined as sentences containing one or more subordinate clauses, is calculated using syntactic parsing techniques. This ratio highlights the degree of syntactic complexity present in the text.

$$\text{Complex Sentence Ratio} = \frac{\text{Number of complex sentences}}{\text{Total sentences}}. \tag{3.6}$$

Hesitation Count

In spoken data, hesitation markers (e.g. "uh", "um", "ehm") are identified and counted to quantify disfluencies and speech hesitations, which are indicative of cognitive processing difficulty or uncertainty.

Propositional Density

Propositional density can be calculated as the ratio of the number of propositions (basic units of meaning or content clauses) to the total word count. This measure reflects the information density and semantic richness of the sample.

$$\text{Propositional Density} = \frac{\text{Number of propositions}}{\text{Total words}}. \tag{3.7}$$

Filler Words Count

Common filler words (e.g., "vlastně", "jakože", "jako", "prostě") are detected and counted to assess the degree of non-lexical verbal fillers present in the speech or text, which can indicate hesitation or conversational style.

3.4 Statistical analysis and machine learning

3.4.1 Preprocessing

The first step, before computing any acoustic parameters, involved normalizing the audio files using the `Librosa` Python library. The files were normalized with the aim of achieving more consistent audio levels across recordings, providing a better foundation for parameter computation—particularly due to fixed thresholds in various functions. The normalized files were then renamed and saved to a separate folder. For each file, the suffix `_CZ-AZV-TSK1_1` (indicating it is a recording related to TSK1) was removed, so that each filename directly corresponded to a patient ID found in the database. Normalization and renaming of the files is contained to script `normalize.py`.

In the next step, it was necessary to identify covariates, such as age and gender, from the calculated parameters using Spearman's correlation. Once identified, these covariates were removed through linear regression to minimize their influence on the analysis.

3.4.2 Exploratory analysis

The goal of exploratory analysis is to describe the characteristics of the parameters in each group, assess differences between groups. Statistics, including median, first and third quartile, interquartile range, and minimum and maximum values, were calculated for all studied parameters within each group. These results are summarized in tables and figures in chapter 5 to provide a clear overview of the data statisticks and distribution.

Mann-Whitney U–Test

To evaluate which parameters were significant in distinguishing between individual groups, the Mann-Whitney U–Test (MW) was employed. This non-parametric method compares two groups without requiring normal distribution of the data. The test generates p-values that indicate the presence of statistically significant differences between the groups based on specific parameters. However, conducting multiple tests increases the likelihood of false positives. To mitigate this, the p-values were adjusted using a False Discovery Rate (FDR) correction. [27] Six group combinations were tested using MW:

1. HC vs. PD
2. HC vs. MCI
3. HC vs. DLB
4. PD vs. MCI
5. PD vs. DLB
6. MCI vs. DLB

Spearman's Correlation

Relationships between the studied parameters and cognitive and depressive scores were analyzed using Spearman's correlation. This step aimed to uncover potential associations between these variables and highlight patterns of interest. Just like in the case of MW, FDR correction is used to adjust calculated values.

3.4.3 Binary classification

XGBoost

For the classification, the XGBoost (Extreme Gradient Boosting) algorithm was used. This method falls under supervised machine learning and serves as an optimized implementation of the Gradient Boosting framework. By combining multiple simpler models, primarily decision trees, XGBoost enhances overall prediction accuracy. In each iteration, a new decision tree is added to minimize the errors of the previous trees, with the final prediction being the sum of the outputs from all trees. Unlike traditional Gradient Boosting, XGBoost incorporates parallelization during tree creation, significantly accelerating the training process for large datasets. [28]

Cross validation

It is important to ensure, the model does not overfit, in this regard a k-fold cross validation will play a critical role. It is a method used in machine learning to evaluate a model's performance on unseen data. It achieves this by dividing the dataset into k subsets, or folds, which are used for training and testing in a systematic way. During the process, the model is trained k times, each time using $k-1$ folds as the training set and the remaining fold as the test set. This ensures that every data point is used for validation once and for training $k-1$ times. The results from all iterations are then averaged to provide a comprehensive evaluation of the model's effectiveness. [29]

3.4.4 Hyperparameter tuning

To optimize the performance of the machine learning models, hyperparameter tuning will be employed. This process involves selecting the most effective combination of hyperparameters to maximize the model's accuracy and generalization ability. Two key methods will most likely be utilized: random search and Bayesian hyperparameter optimization. [30]

Random Search

In this method, hyperparameter values are sampled randomly from predefined ranges or distributions. Unlike grid search, which tests all combinations exhaustively, random search explores a broader space with fewer iterations, making it computationally efficient. This randomness increases the likelihood of discovering high-performing hyperparameter configurations, particularly when some parameters have more significant effects than others. [30]

Bayesian Optimization

The Bayesian algorithm is more flexible than grid search, consisting of two main elements: a probabilistic surrogate model and an acquisition function. In each iteration, the surrogate model is updated with all the previous data from the target function. The acquisition function then identifies the most promising hyperparameter sets and focuses on those that are likely to improve the search. The algorithm predicts how untested combinations will perform based on the surrogate model. [31]

3.4.5 Evaluation metrics

Evaluation metrics are used to assess a model's performance after training, offering valuable feedback that helps guide improvements until the desired accuracy is reached. To achieve the best results, it's crucial to select the most appropriate metric, or ideally, combine multiple metrics for a more balanced evaluation. An essential tool for evaluating model performance is the Confusion Matrix:

- TP – true positive: the model correctly predicts the positive class.
- TN – true negative: the model correctly predicts the negative class.
- FP – false positive: the model incorrectly predicts the positive class.
- FN – false negative: the model incorrectly predicts the negative class.

The Confusion Matrix, while not a metric in itself, offers an insights that allows the interpretation of various performance metrics. [32]

Accuracy

Accuracy is a metric that measures the overall performance of a classification task by calculating the proportion of correctly predicted data points out of the total number of data points. This metric is based on the predicted class labels shown in the confusion matrix, rather than the individual scores assigned to each data point. [33]

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}. \tag{3.8}$$

Precision

This metric measures the accuracy of the positive predictions made by the model. It tells us how many of the instances predicted as positive are actually positive. [32]

$$Precision = \frac{TP}{TP + FP}. \tag{3.9}$$

Specificity

Specificity measures the proportion of actual negative cases that were correctly identified as negative. It indicates how well the model avoids false positive classifications. [33]

$$Specificity = \frac{TN}{TN + FP}. \tag{3.10}$$

Sensitivity (Recall)

Sensitivity, also referred to as recall or the true positive rate, measures the proportion of actual positive cases that were correctly identified as positive. It reflects the model's ability to detect positive cases. [33]

$$Sensitivity = \frac{TP}{TP + FN}. \tag{3.11}$$

Balanced Accuracy

Balanced accuracy is a metric that measures the overall performance of a classification task by calculating the average of sensitivity (recall) obtained on each class. Unlike standard accuracy, which can be misleading in the presence of imbalanced class distributions, balanced accuracy takes into account both sensitivity (true positive rate) and specificity (true negative rate), giving equal weight to the performance on each class regardless of their prevalence in the dataset. This makes it a more appropriate metric for classification problems with unequal class sizes, as is the case in this study. [33]

The formula for balanced accuracy is:

$$Balanced\ Accuracy = \frac{Sensitivity + Specificity}{2} = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right). \tag{3.12}$$

Throughout this thesis, figures and appendixes, unless stated otherwise, the term "accuracy" refers to "balanced accuracy" as defined above.

F1–score

The F1–score is a metric that represents the harmonic mean of precision and recall, providing a balanced measure of a model's performance, particularly when dealing with imbalanced classes. [33]

$$F1\_score = 2 \cdot \frac{2 \cdot TP}{2 \cdot TP + FP + FN} = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}. \tag{3.13}$$

Matthews correlation coefficient (MCC)

The MCC is a metric used for binary classification problems. It takes into account all four values in the confusion matrix and combines them into a single number that reflects the quality of a binary classification model. [33]

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}. \tag{3.14}$$

3.4.6 Interpretation

To interpret the data calculated in previous steps, the results will be visually represented by the Python SHAP (SHapley Additive exPlanations) library. This library should allow the comparison of the influence of individual parameters across different groups, helping to identify which features are most critical for distinguishing between the conditions.

4 Implementation

The implementation of acoustic parameters and other signal processing tasks was carried out using Python (version 3.11). Figure 4.1 presents the pipeline and relationships between the core scripts and data files used in this project. The primary script, `main.py`, begins by reading an `.xlsx` file containing the full database of subjects, excluding patients unsuitable for analysis (e.g., those with missing neuropsychological data). In contrast, the database of `.wav` files remains intact, with no recordings removed. Patient IDs are stored in an array, which is then used to selectively load only the `.wav` files corresponding to those IDs. The computation of individual acoustic parameters is handled by the script `parameters.py` and the computation of linguistic parameters is handled by `lexical.py`, which is invoked from the main script.

All calculated parameters are saved to a `csv` file which is afterwards used by `covariates.py` to calculate and remove covariates. Corrected parameters are saved to another `csv` file, which acts as a shared output file for subsequent stages. Both `exploratory.py` and `xgboost.py` depend on this corrected dataset, the former for exploratory analysis and the latter for training and evaluating predictive models.

The pipeline has been structured to ensure modularity, allowing each task to be handled by a dedicated script while maintaining clear and traceable data flow throughout the process. This setup simplifies debugging, modification, and scaling of individual components without affecting unrelated parts of the pipeline.

4.1 Calculation of Acoustical Parameters

Pitch and Formant Feature Extraction

Fundamental frequency (F_0) and formant frequencies (F_1, F_2) were extracted using the `disvoice 0.1.8` prosody module and `parselmouth 0.4.5`, respectively. Formant values were sampled every 10 ms. Relative standard deviations were computed as the ratio of standard deviation to mean for each parameter.

Intensity-based Feature Extraction

Speech regions were detected using a pretrained `pyannote.audio 3.3.2` pipeline. Intensity values were extracted via `parselmouth`. The slope of the energy envelope was estimated by applying linear regression to intensity values from overlapping 200 ms segments (100 ms overlap). Relative standard deviation of intensity was computed for each sample.

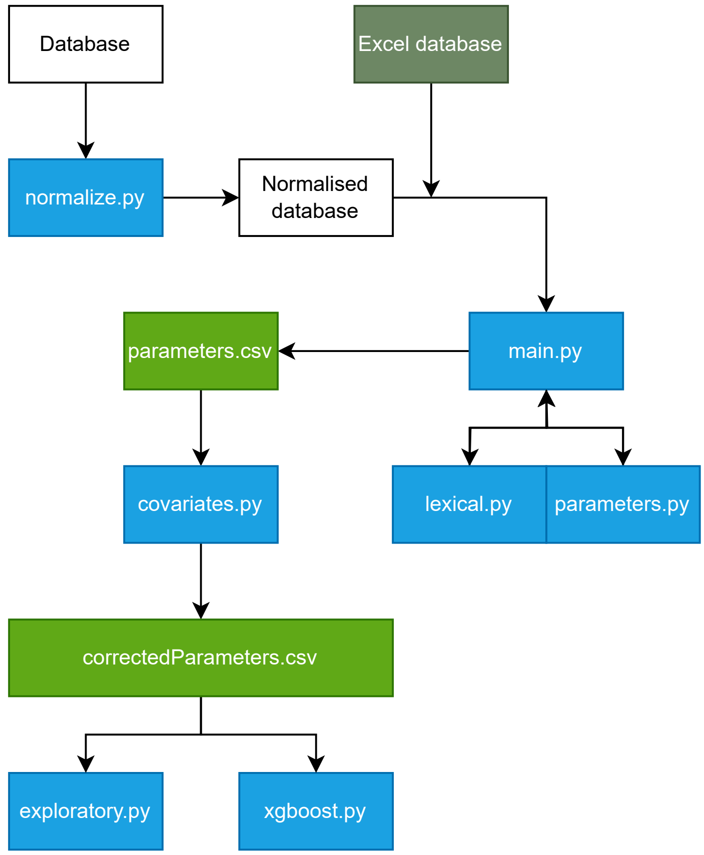

Fig. 4.1: Visualization of the implemented pipeline.

Vowel Classification and Articulatory Feature Extraction

A custom vowel classifier categorized vowels into `/a/`, `/i/`, and `/u/` based on formant frequency ranges. Mean formant values were obtained from stable-pitch, high-intensity frames. VAI and VSA were then calculated from these mean values.

Temporal Feature Extraction

Speech and pause intervals were segmented with `pyannote.audio`. From these, speech time, pause time, average pause duration, and total number of pauses were computed.

Speech Rate and Linguistic Feature Extraction

Audio recordings were transcribed using OpenAI's Whisper model. Word counts and speech time were used to compute WPM.

Together, these acoustical parameters provide a comprehensive multidimensional profile of speech characteristics, encompassing pitch variability, spectral dynamics, intensity modulation, articulatory clarity and temporal structure.

4.2 Calculation of Linguistic Parameters

Linguistic features were derived from speech transcriptions generated by the Whisper. For the case of this analysis, these transcriptions were properly tokenized to individual words or sentences and further analysed to quantify multiple linguistic dimensions: lexical properties such as vocabulary richness, syntactic complexity including sentence length and structure, semantic coherence, and fluency indicators such as pauses, hesitations, and speech rate. The calculation of individual parameters was based on equations (3.2, 3.3, 3.4, 3.5, 3.6, 3.7) and the theoretical framework described in Section 3.3. For differentiating fillers and hesitations, appropriate list of words were manually defined, later individual tokens were compared to those lists. For the differentiation of parts of speech (noun, verb, adverb, etc.) the library `UDPipe 1.3.1.1` was used, more specifically its parts-of-speech tagging system, where each token upon analysis would get tagged with a proper class.

This linguistic analysis captures cognitive-linguistic deficits that might be reflected in spontaneous speech, providing complementary information to acoustic measurements.

5 Results

5.1 Exploratory analysis

5.1.1 MW Test

Group differences in speech parameters were assessed using the Mann–Whitney U test, with the results summarised in Table 5.1. Several comparisons showed uncorrected p-values below the 0.05 threshold.

For all parameters that reached statistical significance prior to the application of the FDR correction, violin plots were generated to provide a more comprehensive interpretation of the results. While p-values obtained from the MW test indicate whether a significant difference exists between groups, they do not convey information about the underlying distribution of the data. Violin plots, by contrast, offer a visual representation of the distribution and central tendency within each group, thereby facilitating a deeper understanding of the observed differences.

For example, in the case of relF1SD, a significant difference was identified between the MCI and DLB groups before FDR correction. The corresponding violin plot (5.1) clearly illustrates how the mean value of relF1SD, represented by a white bar within each plot, is substantially lower in MCI patients compared to DLB patients. This visual representation not only confirms the outcome of the MW test but also highlights the distributional characteristics contributing to this difference. Similar comparisons and interpretations were performed for all other parameters (5.2, 5.3, 5.4, 5.5) that demonstrated significance before FDR adjustment, allowing for a more nuanced and informative analysis of group differences. It should be noted, however, that none of these differences remained statistically significant following FDR correction.

Table 5.1: Mann–Whitney U Test Results Comparing Groups on Speech Parameters

Parameter	Group 1	Group 2	U Statistic	p-value	FDR corrected p-value
relF1SD	MCI	DLB	686	**0.0044**	0.6046
Pause Time	MCI	DLB	1338	**0.0245**	0.6842
Total Pauses	MCI	HC	1791	**0.0257**	0.6842
Total Pauses	MCI	DLB	1322	**0.0333**	0.6842
relF0SD	HC	PD	1793	**0.0454**	0.6842
VAI	MCI	PD	1129	**0.0475**	0.6842

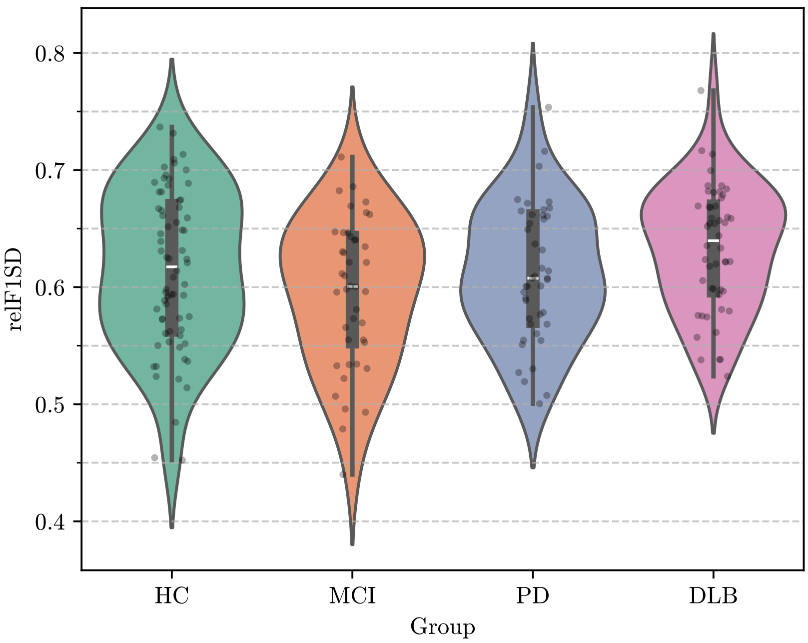

Fig. 5.1: Distribution of relF1SD parameter across all groups.

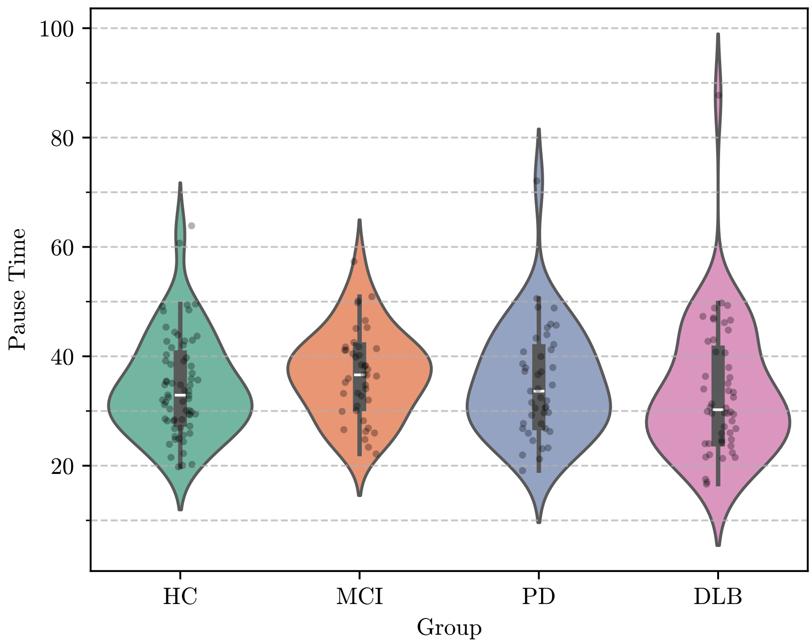

Fig. 5.2: Distribution of Pause Time parameter across all groups.

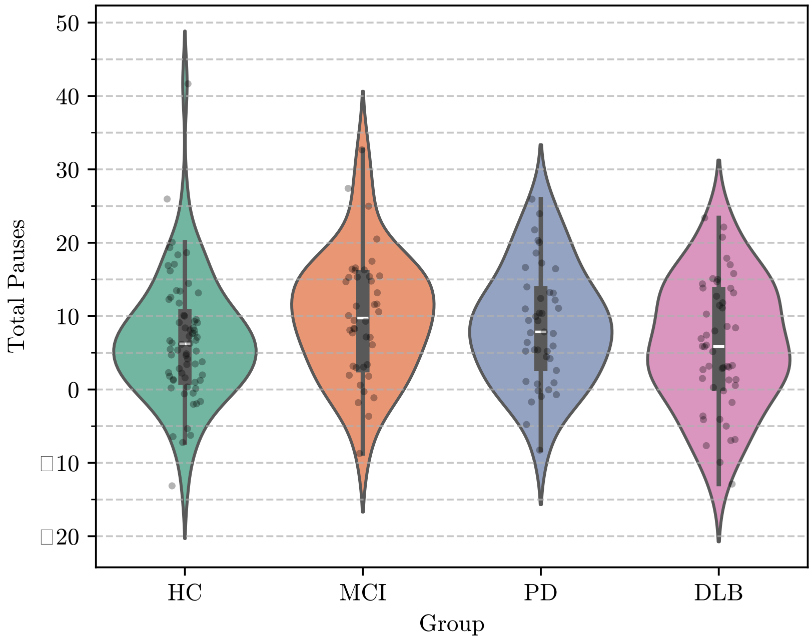

Fig. 5.3: Distribution of Total Pauses parameter across all groups.

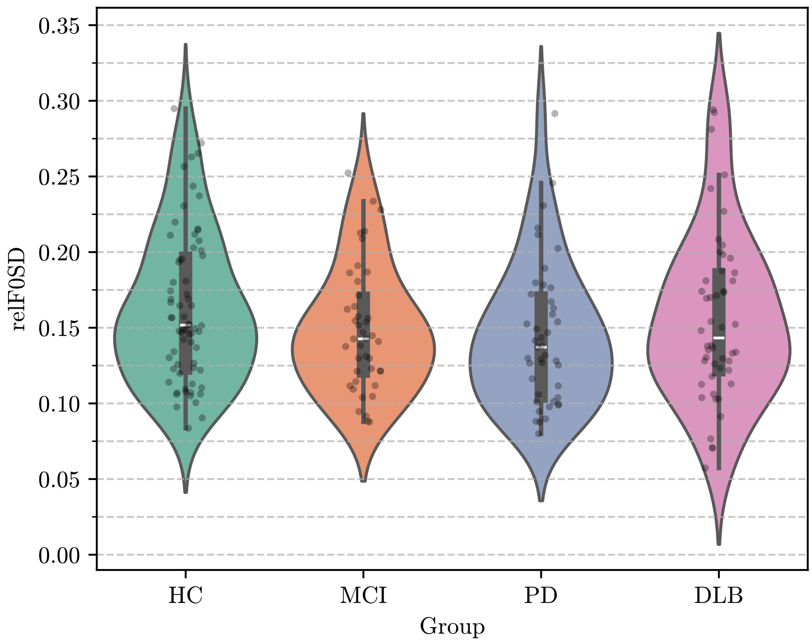

Fig. 5.4: Distribution of relF0SD parameter across all groups.

5.1.2 Correlation of cognitive and depressive scores

Table 5.2 presents the correlations between speech parameters and cognitive and depressive measures. Four correlations remained significant after correction for

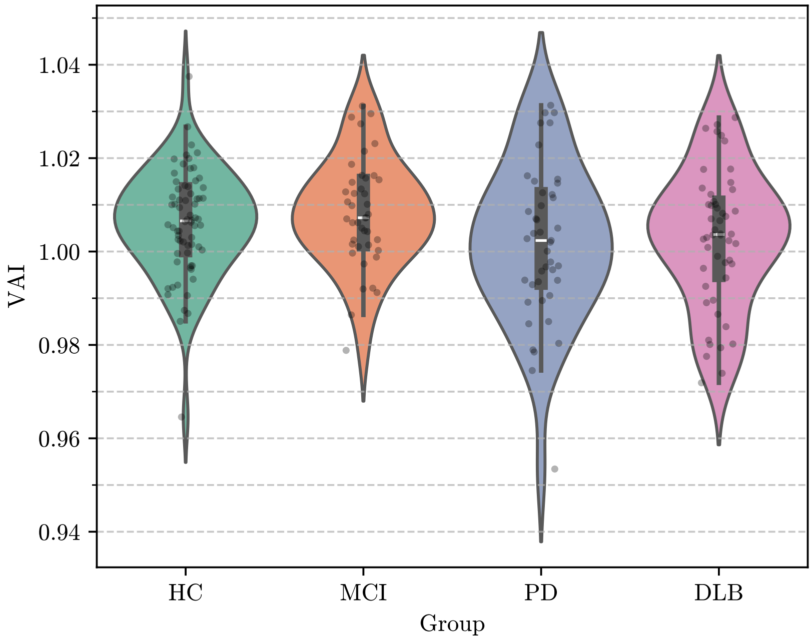

Fig. 5.5: Distribution of VAI parameter across all groups.

multiple comparisons, with several others showing significance before correction.

Temporal speech parameters and depression

Three of the four most robust correlations involved relationships between temporal speech parameters and depressive symptoms measured by GDS-corr. Pause Time showed a significant positive correlation with GDS-corr, indicating that individuals with higher depression scores tend to spend more time pausing during speech production. This finding is complemented by the significant positive correlation between Average Pause duration and GDS-corr, suggesting that not only do more depressed individuals pause more frequently, but their individual pauses tend to be longer. The significant negative correlation between Speech Time and GDS-corr further supports this relationship between temporal speech patterns and depression. As Speech Time represents the proportion of time spent actively speaking (rather than pausing), this negative correlation is conceptually consistent with the positive correlations observed for pause-related parameters. Together, these three significant correlations suggest that depression may particularly affect the temporal organization of speech, possibly reflecting psychomotor changes and cognitive processing alterations associated with depressive states.

Acoustic parameters and cognitive function

The fourth significant correlation after correction was between relSEOSD and attention z-score. This negative correlation indicates that higher variability in this acoustic parameter is associated with poorer performance on attention-related cognitive tests. This relationship suggests potential links between certain acoustic properties of speech and attentional processing. Though not surviving correction for multiple comparisons, several other acoustic parameters showed potentially meaningful relationships with cognitive measures. RelF1SD demonstrated correlations with multiple cognitive domains, including attention z-score, MMSE, memory z-score, and executive function z-score. While these correlations were more modest, the consistency across different cognitive measures suggests that frequency-related variation in speech may have broader relationships with cognitive functioning. RelF2SD showed a positive correlation with executive function z-score before correction, potentially indicating a relationship between certain aspects of articulation and executive processing. These acoustic parameter correlations collectively suggest that subtle characteristics of voice quality and articulation may relate to underlying cognitive processes.

Linguistic features and cognition

Several linguistic parameters showed correlations with cognitive measures before correction. Complex sentence ratio correlated positively with visuospatial z-score and MOCA scores, suggesting that the ability to produce syntactically complex language may relate to overall cognitive status and specifically to visuospatial abilities. Similarly, average sentence length showed a positive correlation with visuospatial z-score. The hesitation count parameter correlated positively with memory z-score before correction. While this relationship might seem counterintuitive, it's possible that hesitations reflect active engagement with memory retrieval processes rather than simply indicating impairment.

Speech rate and cognitive function

WPM showed positive correlations with multiple cognitive domains before correction, including executive function, memory, and attention z-scores. This pattern suggests that faster speech production may be associated with better cognitive performance across various domains, potentially reflecting overall processing efficiency.

Implications of correlation patterns

The pattern of correlations reveals a notable distinction between speech parameters associated with depression and those associated with cognitive function. Temporal organization parameters (pauses, speech time) showed the strongest relationships with depression scores, while acoustic parameters and linguistic complexity measures demonstrated more associations with various cognitive domains. This distinction suggests potential specificity in how different aspects of speech production may reflect different underlying neuropsychological processes. The varying strength of these correlations, with only four surviving correction for multiple comparisons, highlights the complex relationship between speech and neuropsychological status. Nevertheless, the consistency of certain patterns, particularly regarding temporal parameters and depression, suggests promising avenues for speech-based biomarkers in clinical assessment.

Table 5.2: Correlation of cognitive and depressive scores

Parameter	Score	Correlation	p-value	Corrected p-value
Pause Time	GDS-corr	0.27	**0.0001**	**0.0107**
relSEOsd	attention z-score	-0.27	**0.0001**	**0.0107**
Speech Time	GDS-corr	-0.24	**0.0007**	**0.0350**
Average Pause	GDS-corr	0.23	**0.0011**	**0.0433**
relF1SD	attention z-score	0.20	**0.0040**	0.1293
complex_sentence_ratio	visuo-spatial z-score	0.20	**0.0057**	0.1518
relF1SD	GDS-corr	-0.18	**0.0089**	0.2044
relSEOsd	MOCA	-0.17	**0.0131**	0.2554
relF2SD	executive function z-score	0.17	**0.0143**	0.2554
complex_sentence_ratio	MOCA	0.17	**0.0185**	0.2935
hesitation_count	memory z-score	0.16	**0.0222**	0.2935
Total Pauses	GDS-corr	0.16	**0.0222**	0.2935
WPM	executive function z-score	0.16	**0.0263**	0.2935
relF1SD	MMSE	0.24	**0.0267**	0.2935
avg_sentence_length	visuo-spatial z-score	0.16	**0.0273**	0.2935
Total Pauses	memory z-score	-0.15	**0.0331**	0.3308
WPM	memory z-score	0.15	**0.0349**	0.3308
WPM	attention z-score	0.14	**0.0434**	0.3522
Pause Time	MMSE	-0.22	**0.0434**	0.3522
relF1SD	memory z-score	0.14	**0.0477**	0.3522
relF1SD	executive function z-score	0.14	**0.0483**	0.3522

5.2 Classification

This section presents the results of the classification experiments conducted using the XGBoost algorithm. The primary objective was to evaluate the effectiveness of various speech-derived features, specifically prosodic and lexical parameters, in differentiating between four participant groups (HC, PD, MCI, DLB). Two stages of classification were performed. The first involved multiclass classification, where the model attempted to assign each speech sample to one of the four groups. The second stage involved a series of binary classification tasks, where each possible pair of groups was compared individually. In both cases, three feature configurations were tested: prosodic features alone, lexical features alone, and a combination of both.

Table 5.3: Initial hyperparameter search space for Bayesian optimization.

Parameter	Search Range
`n_estimators`	50 to 500
`max_depth`	3 to 10
`learning_rate`	0.01 to 0.3 (log-uniform prior)
`subsample`	0.6 to 1.0
`colsample_bytree`	0.6 to 1.0
`min_child_weight`	1 to 10
`gamma`	0 to 0.5
`reg_alpha`	1×10^{-6} to 1 (log-uniform prior)
`reg_lambda`	1×10^{-6} to 1 (log-uniform prior)

The hyperparameter values listed in Tables 5.4, 5.5, 5.6 were obtained through prior optimization, using Bayesian optimization method (with initial searching parameters described in Table 5.3, and were subsequently used for the final training and evaluation of the model across the respective classifications.

5.2.1 Multiclass classification

A detailed summary of the multiclass classification results is provided in Table 5.7.

The results demonstrate limited overall classification performance across all parameter sets. The highest mean accuracy was achieved using the acoustic feature set, with the combined feature set performing comparably. The linguistic features yielded the lowest classification accuracy, indicating comparatively lower discriminative capability within this domain.

Table 5.4: Optimized hyperparameter values for the **Acoustic** feature set.

Parameter	Multiclass	HC_PD	HC_MCI	HC_DLB	PD_MCI	PD_DLB	MCI_DLB
`n_estimators`	412	116	325	179	50	203	73
`max_depth`	7	3	5	7	4	10	9
`learning_rate`	0.061	0.021	0.231	0.300	0.019	0.013	0.054
`subsample`	0.600	0.966	0.657	0.691	0.734	0.909	0.754
`colsample_bytree`	0.800	0.994	0.752	0.996	0.849	1.000	0.817
`min_child_weight`	10	7	8	1	6	9	4
`gamma`	0.500	0.342	0.000	0.307	0.500	0.437	0.460
`reg_alpha`	0.00667	0.02199	5.43×10^{-6}	0.62013	0.00081	9.23×10^{-5}	0.00266
`reg_lambda`	0.00012	2.23×10^{-5}	0.44562	0.00143	0.00020	0.25714	1.43×10^{-6}

Table 5.5: Optimized hyperparameter values for the **Linguistic** feature set.

Parameter	Multiclass	HC_PD	HC_MCI	HC_DLB	PD_MCI	PD_DLB	MCI_DLB
`n_estimators`	50	50	500	301	420	147	373
`max_depth`	5	3	3	9	5	8	8
`learning_rate`	0.030	0.140	0.139	0.010	0.257	0.010	0.060
`subsample`	0.925	0.940	0.715	0.998	0.874	0.887	0.741
`colsample_bytree`	0.653	0.600	0.600	0.651	0.738	0.999	0.920
`min_child_weight`	3	2	7	4	2	2	9
`gamma`	0.260	0.193	0.484	0.228	0.000	0.466	0.219
`reg_alpha`	0.00352	0.16455	3.12×10^{-6}	0.00093	0.01111	1.00×10^{-6}	0.00035
`reg_lambda`	0.00163	0.02378	6.52×10^{-5}	0.00029	0.212	0.00173	0.00797

Table 5.6: Optimized hyperparameter values for the **Combined** feature set.

Parameter	Multiclass	HC_PD	HC_MCI	HC_DLB	PD_MCI	PD_DLB	MCI_DLB
`n_estimators`	387	493	93	100	254	50	445
`max_depth`	3	5	9	3	6	9	9
`learning_rate`	0.044	0.010	0.076	0.035	0.014	0.014	0.011
`subsample`	0.659	0.646	0.965	0.866	0.823	0.753	0.909
`colsample_bytree`	0.857	0.880	0.925	0.856	0.778	0.822	0.953
`min_child_weight`	5	1	6	1	3	4	5
`gamma`	0.477	0.396	0.086	0.233	0.459	0.031	0.045
`reg_alpha`	0.79545	1.00×10^{-6}	0.03426	2.19×10^{-6}	8.56×10^{-6}	0.00541	2.53×10^{-6}
`reg_lambda`	5.78×10^{-5}	0.00542	0.17210	0.00017	0.03231	1.55×10^{-6}	0.00387

Table 5.7: Multiclass classification results (mean ± standard deviation) across all folds.

Parameter Set	Accuracy	Precision	Recall	F1	MCC
Acoustic	0.311 ± 0.115	0.354 ± 0.172	0.311 ± 0.115	0.307 ± 0.125	0.087 ± 0.154
Linguistic	0.271 ± 0.112	0.261 ± 0.142	0.271 ± 0.112	0.250 ± 0.114	0.046 ± 0.160
Combined	0.302 ± 0.106	0.303 ± 0.112	0.302 ± 0.106	0.283 ± 0.098	0.086 ± 0.158

Similar patterns were observed across the additional evaluation metrics, including precision, recall, F1-score, and MCC. Acoustic features consistently achieved superior scores relative to both the linguistic and combined feature sets. Notably, the MCC values remained low in all cases, underscoring the inherent difficulty of the multiclass classification task in this context and suggesting limited agreement between predicted and true class labels beyond chance level.

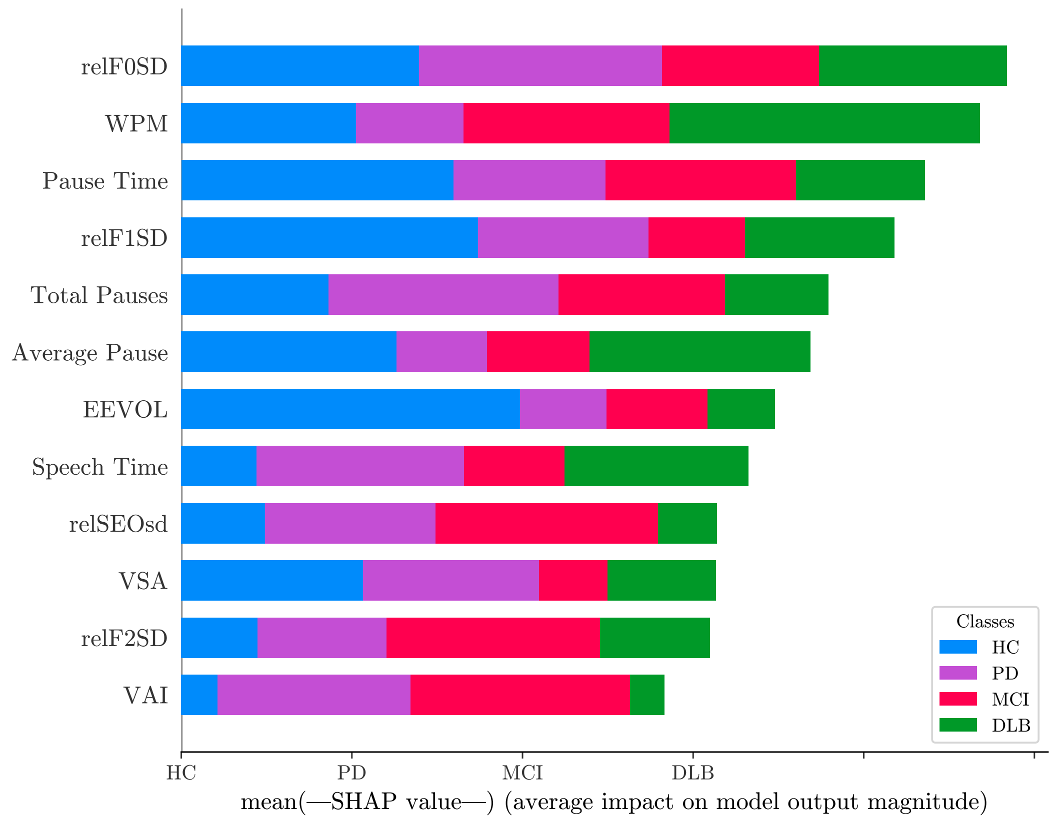

Fig. 5.6: SHAP Feature importance - Acoustic set (Multiclass)

The findings indicate that acoustic parameters convey more robust and reliable class-discriminative information than the linguistic features employed in this study. The lack of substantial improvement when combining both feature sets suggests potential redundancy or suboptimal integration of the two modalities within the current modeling framework. It is possible that the linguistic features, as extracted, offered limited additional information or that their discriminative potential may require alternative feature engineering strategies or more complex, modality-aware classification models to be effectively utilized.

Overall, while acoustic features appear to hold comparatively greater promise for supporting multiclass classification in this dataset, the modest performance metrics highlight the need for further methodological refinement, potentially including larger datasets, advanced feature selection, or more sophisticated multimodal modeling approaches.

Figure 5.6 presents the results of the SHAP analysis, illustrating the relative importance of individual features when utilizing the acoustic parameter set. This analysis reveals which acoustic features most substantially influenced the model's predictions.

5.2.2 Binary classification

This section presents the results of binary classification analyses performed between pairs of participant groups (HC, PD, MCI, and DLB) using XGBoost classifier. Classifications were, just like in the case of multiclass classification, conducted using three feature sets: linguistic parameters only, acoustic parameters only, and combined linguistic and acoustic parameters. For each classification task, model performance was assessed using accuracy, specificity, sensitivity, and MCC.

Linguistic features

The binary classification results using linguistic parameters (Table 5.8) show varying performance across different clinical group comparisons. Most classifications exhibited an imbalance between specificity and sensitivity, suggesting the model's tendency to favor one class prediction over another. For instance, the HC vs PD comparison showed higher specificity than sensitivity, indicating the model more frequently identified healthy controls correctly while misclassifying many PD cases. A notable pattern emerged in the MCI vs DLB classification, where the model achieved perfect specificity but zero sensitivity, suggesting it classified all samples as belonging to one group. This extreme imbalance indicates that the linguistic features used may not capture meaningful differences between MCI and DLB speech patterns in this dataset. The MCC values in the linguistic feature classifications remained relatively low across all comparisons.

Acoustic features

The acoustic parameters (Table 5.9) demonstrated different classification patterns compared to linguistic features. The HC vs PD classification maintained a similar pattern of higher specificity than sensitivity, but with slightly improved overall metrics compared to linguistic features. Interestingly, the MCI vs DLB classification, which showed extreme imbalance with linguistic features, displayed

Table 5.8: Binary classification results using linguistic parameters (mean ± standard deviation across 10 folds).

Group 1	Group 2	Accuracy (mean ± std)	Specificity (mean ± std)	Sensitivity (mean ± std)	MCC (mean ± std)
HC	PD	0.5392 ± 0.1222	0.7333 ± 0.1338	0.3450 ± 0.1753	0.0918 ± 0.2779
HC	MCI	0.5769 ± 0.0707	0.8238 ± 0.1238	0.3300 ± 0.2088	0.1642 ± 0.1465
HC	DLB	0.5495 ± 0.1436	0.3800 ± 0.1887	0.7190 ± 0.2025	0.1112 ± 0.3280
PD	MCI	0.5625 ± 0.1352	0.5850 ± 0.2237	0.5400 ± 0.2059	0.1121 ± 0.2973
PD	DLB	0.5775 ± 0.1339	0.6200 ± 0.1077	0.5350 ± 0.2480	0.1589 ± 0.2801
MCI	DLB	0.5000 ± 0.0000	1.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000

more balanced performance with acoustic features. This suggests that acoustic speech characteristics may better capture the differences between MCI and DLB than linguistic parameters. The PD vs DLB comparison also showed relatively better performance with acoustic features than with linguistic features.

Table 5.9: Binary classification results using acoustic parameters (mean ± standard deviation across 10 folds).

Group 1	Group 2	Accuracy	Specificity	Sensitivity	MCC
HC	PD	0.5788 ± 0.1099	0.8976 ± 0.1127	0.2600 ± 0.1786	0.2109 ± 0.3031
HC	MCI	0.5300 ± 0.1396	0.7500 ± 0.1577	0.3100 ± 0.2289	0.0772 ± 0.3267
HC	DLB	0.5640 ± 0.1037	0.4400 ± 0.1497	0.6881 ± 0.1430	0.1373 ± 0.2117
PD	MCI	0.5750 ± 0.1401	0.5200 ± 0.2909	0.6300 ± 0.2100	0.1614 ± 0.3152
PD	DLB	0.6125 ± 0.1286	0.7600 ± 0.2154	0.4650 ± 0.1689	0.2701 ± 0.2860
MCI	DLB	0.6425 ± 0.1768	0.6400 ± 0.2154	0.6450 ± 0.2650	0.2994 ± 0.3561

Combined features

The use of combined linguistic and acoustic parameters (Table 5.10) resulted in varied performance across the different classification pairs. The HC vs DLB comparison showed relatively better balanced performance between specificity and sensitivity with combined features than with either feature set alone. However, for the PD vs DLB comparison, the combined approach did not perform as well as acoustic features alone. The MCI vs DLB classification showed improvement with combined features compared to the completely imbalanced performance observed with linguistic features alone. This suggests that for certain clinical comparisons, such as MCI vs DLB, combining modalities may help overcome the limitations of individual feature types.

Table 5.10: Binary classification results using combined parameters (mean $\pm$ standard deviation across 10 folds).

Group 1	Group 2	Accuracy	Specificity	Sensitivity	MCC
HC	PD	0.5506 ± 0.1071	0.7762 ± 0.1811	0.3250 ± 0.1662	0.1337 ± 0.2557
HC	MCI	0.5123 ± 0.1367	0.7595 ± 0.2204	0.2650 ± 0.2281	0.0120 ± 0.3248
HC	DLB	0.5955 ± 0.1261	0.4600 ± 0.2375	0.7310 ± 0.2538	0.2211 ± 0.2761
PD	MCI	0.5750 ± 0.2225	0.5700 ± 0.2600	0.5800 ± 0.2431	0.1492 ± 0.4520
PD	DLB	0.5075 ± 0.0814	0.7800 ± 0.1661	0.2350 ± 0.1831	0.0179 ± 0.2096
MCI	DLB	0.6300 ± 0.1900	0.7000 ± 0.2408	0.5600 ± 0.3262	0.2698 ± 0.3942

6 Discussion

The results of this study reveal important insights into the relationship between specific speech parameters and neurological conditions, offering potential for both clinical assessment and theoretical understanding of these disorders.

6.1 Speech Parameters and Clinical Group Differentiation

The binary classification results demonstrate varying effectiveness of speech parameters in differentiating between clinical groups. Acoustic parameters generally outperformed linguistic features, particularly in distinguishing MCI from DLB (accuracy of 0.64 with acoustic vs. 0.50 with linguistic parameters). This finding is similar to findings of Yamada et al. [34], who reported that acoustic voice analysis could effectively distinguish between different neurological diseases. The observed superiority of acoustic parameters suggests that the underlying neuropathology in these conditions may more consistently affect the motor aspects of speech rather than linguistic processing.

Interestingly, HC and PD differentiation showed high specificity (0.90) but low sensitivity (0.26) with acoustic parameters, indicating the model's tendency to correctly identify healthy controls while misclassifying many PD cases. The greater effectiveness of acoustic parameters in this comparison likely reflects the well-documented impact of PD on phonatory and articulatory control systems, particularly affecting fundamental frequency variation and speech timing which si similar to findings by Skodda et al. [35].

The combined feature set yielded mixed results, with improvement in HC vs. DLB classification (accuracy of 0.60) compared to either feature set alone. This observation suggests synergistic information between acoustic and linguistic modalities for certain clinical comparisons, similar to findings by Fraser et al. [36], who reported improved classification of cognitive impairment when combining lexical and acoustic features. However, the inconsistent advantage of the combined approach indicates that feature integration requires further refinement, possibly employing more sophisticated techniques as suggested by Roger et al. [37].

6.2 Relationship Between Speech Parameters and Cognitive-Affective Status

The correlation analysis revealed compelling relationships between specific speech parameters and neuropsychological measures. Temporal organization parameters, particularly pause time and average pause duration, showed the strongest correlations with depression scores, surviving correction for multiple comparisons. This finding corresponds with research by Cummins et al. [38], who identified similar associations between temporal speech features and depression severity. The temporal disruptions may reflect psychomotor slowing characteristic of depression, supporting the potential use of pause-related measures as depression biomarkers.

In contrast, acoustic parameters (especially relSEOSD) demonstrated stronger associations with cognitive measures, particularly attention. This pattern aligns with findings by Toth et al. [39], who observed correlations between acoustic irregularities and cognitive decline in MCI patients. The different patterns of association between speech parameters and neuropsychological domains suggest specificity in how various aspects of speech production reflect distinct underlying neural processes.

The correlation between complex sentence ratio and visuospatial scores is particularly intriguing and less reported in existing literature. This unexpected relationship may indicate shared neural substrates between syntactic complexity and visuospatial processing, potentially involving executive control networks as proposed by Ash et al. [40] in their work on neurodegenerative conditions affecting language.

6.3 Clinical Implications and Future Directions

The moderate classification performance across conditions suggests that speech analysis, while promising, requires further development before clinical implementation. Nevertheless, the identification of specific speech markers associated with different conditions provides foundation for targeted assessment approaches. For instance, the relationship between relSEOSD and attention performance could inform the development of speech-based screening tools for attentional deficits in neurodegenerative populations.

The distinct patterns of association between temporal parameters and depression versus acoustic parameters and cognitive function suggest potentially different neural mechanisms underlying these aspects of speech production. This distinction warrants further investigation using functional neuroimaging to explore the neural correlates of these speech parameters in different clinical populations.

A promising direction for future research would be longitudinal studies tracking changes in these speech parameters over disease progression. Such investigations could determine whether certain speech markers predict clinical outcomes or respond to therapeutic interventions. Additionally, examining speech parameters in prodromal or high-risk populations might reveal early biomarkers before clinical diagnosis.

6.4 Limitations

Several limitations must be acknowledged when interpreting these findings. The classification performance, while above chance level, remains modest, limiting immediate clinical application. This may reflect the heterogeneous nature of these neurological conditions or insufficient sample size to capture the full spectrum of speech variations within each diagnostic group.

The cross-sectional design limits causal interpretations regarding the relationship between speech parameters and neuropsychological status. Moreover, medication effects, particularly in the PD group, may have influenced speech production, potentially masking some disease-specific features as noted in similar studies by Tykalova et al. [41].

A further limitation of this study was hardware performance, specifically limited GPU resources for running the OpenAI Whisper transcription model. While higher-performing models were available, their use was restricted by impractical processing times on the available system. With more precise transcription, perhaps linguistic features that are based on said transcriptions, could have contributed better to overall classification.

Finally, the modest sample size, particularly for some clinical subgroups, may have limited statistical power and the ability to identify more subtle speech markers. Expanding sample sizes and incorporating multicenter data would enhance the generalizability of findings and potentially improve classification performance through more robust model training.

Despite these limitations, this study contributes valuable insights into the relationship between speech parameters and neurological conditions, highlighting the potential of automated speech analysis as an objective, non-invasive assessment tool in clinical neurology and psychiatry.

Conclusion

This thesis focused on developing a methodology for the differential analysis of three neurological disorders – MCI, DLB and PD – through acoustic and linguistic speech analysis. The research explored how specific speech parameters could distinguish between these conditions and HC, while also investigating correlations between speech features and cognitive-affective measures. By utilizing digital signal processing and machine learning approaches, this work aimed to identify objective, non-invasive biomarkers that could support clinical diagnosis and enhance our understanding of the underlying neurological mechanisms in these disorders.

1: Develop a literature review on speech and voice disorders in the aforementioned diseases

This goal was achieved through a comprehensive literature review examining speech and voice disorders associated with MCI, DLB, and PD. The review synthesized current research on how these neurological conditions affect different aspects of speech production, including phonatory and articulatory control systems, temporal organization, and linguistic processing. This foundation was essential for identifying potentially relevant speech parameters and contextualizing the study's findings within the broader scientific literature.

2: Familiarize with the provided database of speech recordings

This goal was successfully accomplished through thorough exploration and analysis of the speech recording database. The process involved understanding the demographic characteristics of participants across different clinical groups (MCI, DLB, PD, and HC) and the specific speech tasks performed. This familiarization was crucial for subsequent analyses and ensured methodological validity throughout the study.

3: Based on the literature review, design and implement the parameterization of recordings using appropriately selected digital speech biomarkers

Based on the literature review, a comprehensive set of both acoustic and linguistic parameters was designed and implemented. The acoustic parameters included features related to temporal organization (pause time, average pause duration), fundamental frequency variation, and spectral characteristics (particularly relSEOsd). Linguistic features included measures of syntactic complexity such as complex sentence ratio. This dual approach to parameterization allowed for a more comprehensive analysis of speech alterations across different neurological conditions.

4: Perform exploratory analysis (identify which acoustic parameters differentiate the mentioned diseases and how well they correlate with clinical scales)

The exploratory analysis revealed important relationships between speech parameters and clinical measures. Temporal organization parameters, particularly pause time and average pause duration, showed the strongest correlations with depression scores. In contrast, acoustic parameters (especially relSEOsd) demonstrated stronger associations with cognitive measures, particularly attention. An intriguing correlation was also found between complex sentence ratio and visuospatial scores. These distinct patterns suggest specificity in how various aspects of speech production reflect different underlying neural processes, providing valuable insights into the relationship between speech and neuropsychological status in these conditions.

5: Perform classification analysis (train and evaluate machine learning models that will automatically classify these diseases)

Classification analysis demonstrated varying effectiveness of speech parameters in differentiating between clinical groups. Acoustic parameters generally outperformed linguistic features, particularly in distinguishing MCI from DLB (accuracy of 0.64 with acoustic vs. 0.50 with linguistic parameters). The HC and PD differentiation showed high specificity (0.90) but low sensitivity (0.26) with acoustic parameters, indicating the model's tendency to correctly identify healthy controls while misclassifying many PD cases. The combined feature set yielded mixed results, with improvement in HC vs. DLB classification (accuracy of 0.60) compared to either feature set alone, suggesting potential synergistic information between acoustic and linguistic modalities for certain clinical comparisons.

6: Discuss the results in relation to available scientific literature

The findings were thoroughly contextualized within existing literature, highlighting both consistencies and novel contributions. The observed superiority of acoustic parameters in differentiating clinical groups aligns with previous research suggesting that the underlying neuropathology in these conditions may more consistently affect the motor aspects of speech rather than linguistic processing. The correlation patterns between specific speech parameters and neuropsychological measures correspond with previous studies identifying associations between temporal speech features and depression severity, and between acoustic irregularities and cognitive decline. The unexpected relationship between syntactic complexity and visuospatial processing offers a new perspective on potential shared neural substrates, contributing to the theoretical understanding of language processing in neurodegenerative conditions.

Future Directions

Future research should focus on longitudinal studies tracking changes in speech parameters over the course of disease progression to determine whether certain speech markers predict clinical outcomes or respond to therapeutic interventions. Additionally, examining speech parameters in prodromal or high-risk populations might reveal early biomarkers before clinical diagnosis. Functional neuroimaging studies exploring the neural correlates of specific speech parameters could also enhance our understanding of the distinct neural mechanisms underlying different aspects of speech production in these clinical populations. Finally, expanding sample sizes and incorporating multicenter data would address current limitations, enhance generalizability, and potentially improve classification performance through more robust model training. While the current classification performance remains modest, limiting immediate clinical application, this work provides a foundation for the development of speech-based screening tools for cognitive and affective assessment in neurodegenerative disorders.

Bibliography

[1] DUNNE, Ross A.; AARSLAND, Dag; O'BRIEN, John T.; BALLARD, Clive; BANERJEE, Sube; FOX, Nick C.; ISAACS, Jeremy D.; UNDERWOOD, Benjamin R.; PERRY, Richard J.; CHAN, Dennis; DENING, Tom; THOMAS, Alan J.; SCHRYER, Jeffrey; JONES, Anne-Marie; EVANS, Alison R.; ALESSI, Charles; COULTHARD, Elizabeth J.; PICKETT, James; ELTON, Peter; JONES, Roy W.; MITCHELL, Susan; HOOPER, Nigel; KALAFATIS, Chris; RASMUSSEN, Jill G. C.; MARTIN, Helen; SCHOTT, Jonathan M.; BURNS, Alistair. *Mild Cognitive Impairment: the Manchester Consensus.* Age and Ageing, 2021, vol. 50, no. 1, p. 72–80. https://doi.org/10.1093/ageing/afaa228.

[2] LAMPTEY, R.N.L.; CHAULAGAIN, B.; TRIVEDI, R.; GOTHWAL, A.; LAYEK, B.; SINGH, J. *A Review of the Common Neurodegenerative Disorders: Current Therapeutic Approaches and the Potential Role of Nanotherapeutics.* International Journal of Molecular Sciences, 2022, vol. 23, p. 1851. https://doi.org/10.3390/ijms23031851.

[3] DONG-WOO, Ryu; KYUNGDO, Han and A-HYUN, Cho. *Mortality in Parkinson's Disease: A Systematic Review and Meta-Analysis.* [online]. 2020. ISSN=1664-2295. Available from: https://www.frontiersin.org/journals/neurology/articles/10.3389/fneur.2023.1236296. [accessed 2024-10-06]

[4] SMALL, Scott A. *Understanding How Dementia Causes Death.* [online]. Columbia University, 2023. Available from: https://www.neurology.columbia.edu/news/understanding-how-dementia-causes-death-0. [accessed 2024-10-06]

[5] UNKNOWN Author. *Parkinson's Disease: Causes, Symptoms, and Treatments.* Online. Available at: https://www.nia.nih.gov/health/parkinsons-disease/parkinsons-disease-causes-symptoms-and-treatments. [accessed 2024-09-27].

[6] TOLOSA, Eduardo, Alicia GARRIDO, Sonja W SCHOLZ and Werner POEWE. *Challenges in the diagnosis of Parkinson's disease.* The Lancet Neurology [online]. 2021(Volume 20, 5), 385-397 [accessed 2024-10-1]. ISSN 1474-4422. Available at: https://www.sciencedirect.com/science/article/pii/S1474442221000302.

[7] MORO-VELAZQUEZ, Laureano; GOMEZ-GARCIA, Jorge A.; ARIAS-LONDOÑO, Julian D.; DEHAK, Najim and GODINO-LLORENTE, Juan I.,

2021. *Advances in Parkinson's Disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects.* Online. Biomedical Signal Processing and Control. Volume 66. ISSN 1746-8094. Available at: https://www.sciencedirect.com/science/article/pii/S174680942100015X. [accessed 2024-10-06].

[8] Emamzadeh, F. N., Surguchov, A. *Parkinson's Disease: Biomarkers, Treatment, and Risk Factors.* Online. Available at: https://www.frontiersin.org/articles/10.3389/fnins.2018.00612/full [accessed 2024-09-27]

[9] Gomperts, S. *Lewy Body Dementias: Dementia With Lewy Bodies and Parkinson Disease Dementia.* CONTINUUM: Lifelong Learning in Neurology, 2016 Apr;22(2 Dementia):435-63. https://doi.org/10.1212/CON.0000000000000309.

[10] Picca, Anna, Guerra, Francesco, Calvani, Riccardo, Romano, Raffaele, Coelho-Júnior, Heráclito J., Bucci, Cecilia, Marzetti, Emanuele. *Mitochondrial Dysfunction, Protein Misfolding and Neuroinflammation in Parkinson's Disease: Roads to Biomarker Discovery.* Biomolecules [online]. 2021, vol. 11, no. 10, p. 1508 [accessed 2024-10-07]. DOI: https://doi.org/10.3390/biom11101508.

[11] STOKER, Thomas B. and GREENLAND, Julia C., 2018. *Parkinson's Disease: Pathogenesis and Clinical Aspects.* Brisbane (AU): Codon Publications. ISBN 978-0-9944381-6-4.

[12] BLOEM, Bastiaan R., OKUN, Michael S., KLEIN, Christine. *Parkinson's disease.* The Lancet [online]. 2021, vol. 397, no. 10291, pp. 2284–2303 [accessed 2024-10-07]. ISSN 0140-6736. Available from: https://doi.org/10.1016/S0140-6736(21)00218-X.

[13] GAUTHIER, Serge. 2006. *Mild cognitive impairment.* The Lancet [online]. (Volume 367, 9518), 1262-1270. ISSN 0140-6736. Available at: https://www.sciencedirect.com/science/article/pii/S0140673606685425 [accessed 2024-9-27]

[14] Mayo Clinic. *Mild cognitive impairment (MCI): Symptoms and causes.* Mayo Clinic [online]. 2023 . Available at: https://www.mayoclinic.org/diseases-conditions/mild-cognitive-impairment/symptoms-causes/syc-20354578 [accessed 2024-10-06]

[15] National Institute on Aging. *What Is Mild Cognitive Impairment?* [online]. National Institute on Aging, 2023 . Available from:

https://www.nia.nih.gov/health/memory-loss-and-forgetfulness/what-mild-cognitive-impairment. [accessed 2024-10-06]

[16] MUELLER, Christoph, Clive BALLARD, Anne CORBETT and Dag AARSLAND. 2017. *The prognosis of dementia with Lewy bodies.* The Lancet Neurology [online]. (Volume 16, 5), 390-398 [accessed 2024-9-27]. ISSN 1474-4422. Available at: https://www.sciencedirect.com/science/article/pii/S1474442217300741.

[17] MCKEITH, Ian; BOEVE, Bradley F.; DICKSON, Dennis W.; HALLIDAY, Glenda; TAYLOR, John-Paul; WEINTRAUB, Daniel; AARSLAND, Dag; GALVIN, James; ATTEMS, Johannes; BALLARD, Clive; BAYSTON, Amber; BEACH, Thomas G.; BLANC, Frederic; BOHNEN, Nico; BONANNI, Laura; BRAS, Jose; BRUNDIN, Patrik; BURN, David; CHEN-PLOTKIN, Alice; KOSAKA, Kenji. *Diagnosis and management of dementia with Lewy bodies: Fourth consensus report of the DLB Consortium.* Neurology [online]. 2017, vol. 89. Available from: https://doi.org/10.17863/CAM.10787.

[18] YAMADA, Masahito; KOMATSU, Junji; NAKAMURA, Keiko; SAKAI, Kenji; SAMURAKI-YOKOHAMA, Miharu; NAKAJIMA, Kenichi; YOSHITA, Mitsuhiro. *Diagnostic Criteria for Dementia with Lewy Bodies: Updates and Future Directions.* Journal of Movement Disorders [online]. 2020, vol. 13, no. 1, pp. 1–10 [accessed 2024-10-07]. Available from: https://doi.org/10.14802/jmd.19052.

[19] HIJAZI, Zina; YASSI, Newaf; O'BRIEN, John T. and WATSON, Rosie. *The influence of cerebrovascular disease in dementia with Lewy bodies and Parkinson's disease dementia.* Online. Year 2022, č. Volume 29, 4, pp. 1254-1265. Available at: https://doi.org/https://doi.org/10.1111/ene.15211. [accessed 2024-10-15].

[20] Duffy, Joseph R. *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management.* Third Edition St. Louis, Mo.: Elsevier Mosby. 2012. ISBN 978- 0323072007.

[21] FREED, Donald B., 2023. *Motor speech disorders: diagnosis and treatment.* Fourth Edition. Plural Publishing. ISBN 978-1635506075.

[22] KLIMOVA, B.; KUCA, K.: *Speech and language impairments in dementia.* Journal of Applied Biomedicine, year 14, 02 2016, https://doi.org/10.1016/j.jab.2016.02.002.

[23] CHIU, P.; TSAI, C.: *P1-302: FREEZING OF SPEECH SINGLE QUESTIONNAIRE AS A SCREENING TOOL FOR DEMENTIA WITH LEWY BODIES.* Alzheimer's & Dementia, , 15: P359-P359. 2019 https://doi.org/10.1016/j.jalz.2019.06.857.

[24] LOOZE, Céline De; DEHSARVI, Amir; CROSBY, Lisa; VOURDANOU, Aisling; COEN, Robert F. et al. *Cognitive and Structural Correlates of Conversational Speech Timing in Mild Cognitive Impairment and Mild-to-Moderate Alzheimer's Disease: Relevance for Early Detection Approaches.* Online. Frontiers in Aging Neuroscience. Year 2021, Volume 13. Available at: https://doi.org/https://doi.org/10.3389/fnagi.2021.637404. [accessed 2024-10-15].

[25] ASH, S.; McMILLAN, C.; GROSS, R.G.; COOK, P.; GUNAWARDENA, D.; MORGAN, B.; BOLLER, A.; SIDEROWF, A.; GROSSMAN, M. *Impairments of speech fluency in Lewy body spectrum disorder.* Brain and Language, 2012, vol. 120, no. 3, p. 290–302. Available at: https://doi.org/10.1016/j.bandl.2011.09.004 [accessed 2024-10-13].

[26] MCKEITH, Ian G.; FERMAN, Tanis J.; THOMAS, Alan J.; BLANC, Frédéric and BOEVE, Bradley F. et al. *Research criteria for the diagnosis of prodromal dementia with Lewy bodies.* Online. Neurology. year 2020, issue 94(17), pp. 743-755. Available at: https://doi.org/10.1212/WNL.0000000000009323. [accessed 2024-10-20].

[27] MCClENAGHAN, E. *Mann-Whitney U Test: Assumptions and Example.* [online] https://www.technologynetworks.com/informatics/articles/mann-whitney-u-test-assumptions-and-example-363425 [accessed 2024-11-24]

[28] Unknown author. *XGBoost.* [online] https://www.geeksforgeeks.org/xgboost/ [accessed 2024-11-29]

[29] Brownlee, J. *A Gentle Introduction to k-fold Cross-Validation.* Machine Learning Mastery: Making Developers Awesome at Machine Learning, 2020. [online] https://machinelearningmastery.com/k-fold-cross-validation/ [accessed 2024-11-29]

[30] ZAHEDI, Leila; GHAREH MOHAMMADI, Farid; REZAPOUR, Shabnam; OHLAND, Matthew W.; AMINI, M. Hadi. *Search Algorithms for Automated Hyper-Parameter Tuning.* 2021. arXiv:2104.14677 [cs.LG]. Available at: https://arxiv.org/abs/2104.14677. [accessed 2024-12-10]

[31] ALIBRAHIM, Hussain; LUDWIG, Simone A. *Hyperparameter Optimization: Comparing Genetic Algorithm against Grid Search and Bayesian Optimization.* In: 2021 IEEE Congress on Evolutionary Computation (CEC). 2021, pp. 1551–1559. DOI: https://doi.org/10.1109/CEC45853.2021.9504761. [accessed 2024-12-10]

[32] OLUGBENGA, M. *Balanced Accuracy: When Should You Use It?* Neptune.ai [online]. Available at: https://neptune.ai/blog/balanced-accuracy [accessed 2024-12-10].

[33] CHICCO, D.; JURMAN, G. *The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification.* BioData Mining. 2023 Feb 17;16(1):4. DOI: https://doi.org/10.1186/s13040-023-00322-4. PMID: 36800973; PMCID: PMC9938573.

[34] YAMADA, Yuma; SHINKAWA, Kazuki; NEMOTO, Masashi; OTA, Masafumi; NEMOTO, Koji; ARAI, Tetsuaki. *Speech and language characteristics differentiate Alzheimer's disease and dementia with Lewy bodies.* Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 2022, vol. 14, no. 1, p. e12364. DOI: https://doi.org/10.1002/dad2.12364

[35] SKODDA, Sabine. *Impaired motor speech performance in Parkinson's disease.* Journal of Neural Transmission, 2013, vol. 120, no. 5, p. 693–700. https://doi.org/10.1007/s00702-013-1115-9

[36] FRASER, Kathleen C.; MELTZER, Jed A.; RUDZICZ, Frank. *Linguistic features identify Alzheimer's disease in narrative speech.* Journal of Alzheimer's Disease, 2016, vol. 49, no. 2, p. 407–422. https://doi.org/10.3233/JAD-150520

[37] ROGER, Vincent; FARINAS, Julien; PINQUIER, Julien. *Deep neural networks for automatic speech processing: a survey from large corpora to limited data.* Journal of Audio, Speech, and Music Processing, 2022, vol. 2022, no. 19. https://doi.org/10.1186/s13636-022-00251-w

[38] CUMMINS, Nicholas; SCHERER, Stefan; KRAJEWSKI, Jarek; et al. *A review of depression and suicide risk assessment using speech analysis.* Speech Communication, 2015, vol. 71, p. 10–49. https://doi.org/10.1016/j.specom.2015.03.004

[39] TÓTH, László; GOSZTOLYA, Gábor; VINCZE, Veronika; et al. *Automatic detection of mild cognitive impairment from spontaneous speech using ASR.* Interspeech, 2018, p. 2994–2998. https://doi.org/10.21437/Interspeech.2015-568

[40] ASH, Sharon; McMILLAN, Corey; GROSS, Rachel G.; et al. *The organization of narrative discourse in Lewy body spectrum disorder.* Brain and Language, 2012, vol. 120, no. 3, p. 290–302. https://doi.org/10.1016/j.bandl.2011.05.006

[41] TYKALOVA, Tereza; NOVOTNY, Martin; RUZICKA, Evzen; DUSEK, Petr; RUSZ, Jan. *Short-term effect of dopaminergic medication on speech in early-stage Parkinson's disease.* NPJ Parkinson's Disease, 2022, vol. 8, no. 1, p. 22. https://10.1038/s41531-022-00286-y

Symbols and abbreviations

AD	Alzheimer's disease
CDR	Clinical Dementia Rating Scale
CT	computed tomography
DLB	dementia with Lewy bodies
EEVOL	energy evolution
FDR	false discovery rate
FN	false negative
FOS	freezing of speech
FP	false positive
GDS–corr	Geriatric Depression Scale
HC	healthy control
HD	Hypokinetic dysarthria
LTAS	long-term average spectrum
MATTR	moving-average type-token ratio
MCC	Matthews correlation coefficient
MCI	mild cognitive impairment
MoCA	Montreal Cognitive Assessment
MRI	magnetic resonance imaging
MW	Mann-Whitney U–Test
PD	Parkinson's disease
PDD	Parkinson's disease dementia
PET	positron emission tomography
relF0SD	standard deviation of the $F0$ contour relative to its mean
relSEOS	standard deviation of the voice intensity

SHAP	SHapley Additive exPlanations
SPECT	single-photon emission computed tomography
TN	true negative
TP	true positive
TSKn	task n
TTR	type-token ratio
VAD	voice activity detection
VAI	vowel articulation index
VSA	vowel space area
WPM	words per minute

A Electronic Appendix Contents

- /xmalch02_priloha .... root directory of the attached archive
  - /Results .... directory containing various results (tables, figures, etc.)
    - /Classification .... folder containing files related to classification
      - /Multiclass - acoustic ....
        - best_xgboost_model.joblib file containing the best-performing model for this scenario and feature set
        - acoustic_average_confusion_matrix.png .. average confusion matrix across 10 folds for this scenario and feature set
        - acoustic_cross_validation_results.csv ..... crossvalidation results across 10 folds for this scenario and feature set
        - shap_bar_plots_acoustic.svg. visualization of feature importance for this scenario and feature set
      - /Multiclass - combined comparable files as the directory above, different feature set.
      - /Multiclass - linguistic ...... comparable files as the directory above, different feature set
      - /Binary - acoustic ....
        - /HC_vs_DLB ...... comparable contents to those described for multiclass directories
        - /HC_vs_MCI ...... comparable contents to those described for multiclass directories
        - /⋮ .... all possible group combinations
        - acoustic_summary.csv average performance metrics across 10 folds for each pair
        - binary_performance_comparison_acoustic.png figure comparing the performance of all pairs
      - /Binary - combined .. similar contents to the Binary - acoustic folder, different feature set
      - /Binary - linguistic similar contents to the Binary - acoustic folder, different feature set
    - /Exploratory Analysis ..... folder containing files related to the exploratory analysis
      - all_cognitive_correlations.svg ...... figure showing the correlation of individual parameters with medical scores
      - cognitive_depression_correlations.csv .. table containing data of the correlation of individual parameters with medical scores
      - descriptive_stats_all_groups.xlsx. descriptive statistics for all groups
      - mann_whitney_results.csv .... results of MW test
      - significant_spearman_cognitive_depression.svg .... figure showing significant correlation of parameters with medical scores after FDR correction

- parameters.csv ...... file containing all computed speech parameters for each subject
- corrected_speech_parameters.csv .... corected data after covariate removal

/Scripts .... directory containing Python scripts, as described in Figure 4.1

- covariates.py .. script that identifies and removes covariates from calculated speech data
- exploratory.py .... script ensuring exploratory analysis
- lexical.py .... script calculating linguistic parameters
- main.py .... main script
- normalize.py .... script that normalises and renames input audio files
- parameters.py .... script calculating acoustic parameters
- xgboost.py .... script containing machine learning algorithm XGBoost