CZECH TECHNICAL UNIVERSITY

FACULTY OF ELECTRICAL ENGINEERING

Acoustic Investigation of Stress Patterns Using Read Passages in Patients with Degeneration of CNS

Diploma Thesis

Study programme: Communication, Multimedia and Electronics

Study branch: Multimedia Technology

Supervisor: Ing. Tereza Tykalová, PhD.

Student: Bc. Lada Kohoutová May 2017

Acknowledgements

First and foremost, I would like thank my supervisor Tereza Tykalová for her guidance, patience, and invested time. I am also grateful to Iva Novotná, Hana Růžičková for providing the theroretical word stress pattern of the passage used in this study, and to Jan Volín for providing the word stress pattern and marking syllabic boundaries in the text. Furthermore, I am obliged to Barbora Bénová, Dana Horáková, Jiří Klempíř, Michal Novotný, Hana Růžičková, and Evžen Růžička for recruiting participants, managing recording procedures, and providing of clinical data. Finally, my deepest gratitude goes to my parents, little sister, and the rest of our family for believing in me, encouraging me in tough times, and sharing joy every single day.

Declaration

I declare that I have written the presented thesis independently and named all used sources of information in accordance with Methodical instructions about ethical principles for writing academic theses.

Date:…………………………… Signature: ……………………………

ČESKÉ
VYSOKÉ
UČENÍ
TECHNICKÉ
V PRAZE

ZADÁNÍ DIPLOMOVÉ PRÁCE

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Kohoutová** Jméno: **Lada** Osobní číslo: **406080**

Fakulta/ústav: **Fakulta elektrotechnická**

Zadávající katedra/ústav: **Katedra radioelektroniky**

Studijní program: **Komunikace, multimédia a elektronika**

Studijní obor: **Multimediální technika**

II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

Akustická analýza slovního důrazu s využitím čtených promluv u pacientů s poruchou CNS

Název diplomové práce anglicky:

Acoustic Investigation of Stress Patterns Using Read Passages in Patients with Degeneration of CNS

Pokyny pro vypracování:

1. Seznamte se s problematikou analýzy poruch hlasu a řeči u neurodegenerativních onemocnění. Konkrétně se pak zaměřte na charakteristiku a porovnání hypokinetické a ataxické dysarthrie v souvislosti s Parkinsonovou nemocí a roztroušenou sklerózou.
2. Navrhněte jednoduchý algoritmus, který ze zadaného úseku řeči automaticky vypočte akustické parametry charakterizující slovní důraz. Pro návrh algoritmu využijte výpočetní prostředí MATLAB. Pro detekci základní hlasivkové frekvence software Praat. Navržený algoritmus otestujte na vybraném vzorku řečových promluv zdravé populace a pacientů s výskytem Parkinsonovy nemoci a roztroušené sklerózy.
3. Na bázi výstupu algoritmu navrhněte akustické parametry vhodné pro charakteristiku slovního důrazu ze čtených promluv. Na dostupných datech proveďte experiment a s pomocí jednoduchých statistických testů vyhodnoťte, zda narušená schopnost vyjádřit důraz u pacientů s CNS je detekovatelná ze čtených promluv a zda se mezi skupinami pacientů liší.

Seznam doporučené literatury:

[1] Duffy JR. Motor speech disorders: substrates, differential diagnosis and management. 3nd ed. New York: Mosby; 2013.
[2] Tykalova T, Rusz J, Cmejla R, Ruzickova H, Ruzicka E. Acoustic Investigation of Stress Patterns in Parkinson's Disease. J Voice. 2014;28: 129.e1.-129.e8.

Jméno a pracoviště vedoucí(ho) diplomové práce:

Ing. Tereza Tykalová Ph.D., katedra teorie obvodů FEL

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

Datum zadání diplomové práce: **13.02.2017** Termín odevzdání diplomové práce: _____________

Platnost zadání diplomové práce: **31.08.2018**

Podpis vedoucí(ho) práce Podpis vedoucí(ho) ústavu/katedry Podpis děkana(ky)

III. PŘEVZETÍ ZADÁNÍ

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

Datum převzetí zadání Podpis studentky

Abstract

Deviations in production of stress patterns in speech are characteristic for hypokinetic and ataxic dysarthria associated with Parkinson's disease (PD) and multiple sclerosis (MS), respectively. In this study, we aimed to quantify deficits in stress patterns in speech using appropriate acoustic variables. In particular, we analysed stress patterns in 30 PD and 30 MS patients, compared to 30 healthy controls (HC) using a speaking task of passage reading. We performed acoustic analyses of the measurement of fundamental vocal frequency (F0), intensity, and duration. Furthermore, we employed a parameter termed the stress pattern index (SPI) which reflects effects of all basic acoustic quantities related to stress, i.e. intensity, duration and F0. We also designed an algorithm that calculated SPI values in syllables in the passage. The acoustic analyses were performed on a set of selected stressed and unstressed syllables separately, and on the ratio of the values for stressed and unstressed syllables. Although we did not find any significant differences between HC subjects and PD patients, our results of the SPI analysis showed that our MS patients tended to stress syllables excessively and equally. The analysis of F0 suggested that MS patients extended their F0 range in order to convey stress on syllables. We also found excess variation in intensity in MS patients. Finally, the analysis of duration indicated presence of scanning speech in our MS patients.

Keywords: Dysarthria; Acoustic analysis; Stress patterns; Parkinson's disease; Multiple Sclerosis

Abstrakt

Narušená schopnost vyjadřovat důraz v řeči je charakteristická pro hypokinetickou a ataxickou dysartrii spojenou s Parkinsonovou nemocí (PN) a roztroušenou sklerózou (RS). V této studii jsme se zaměřili na kvantifikování důrazu v řeči s použitím vhodných akustických veličin. Konkrétně jsme analyzovali slovní důraz z čtených promluv 30 PN a 30 RS pacientů a 30 zdravých kontrol. Provedli jsme akustické analýzy měření základní hlasivkové frekvence (F0), intenzity a délky trvání slabik. Dále jsme využili parametr nazývaný stress pattern index (SPI), který uvažuje všechny základní akustické veličiny spojené s důrazem – intenzitu, délku trvání a F0. V rámci studie jsme též navrhli algoritmus pro výpočet hodnoty SPI z každé slabiky z textu. Akustické analýzy byly provedené na sadách zdůrazněných a nezdůrazněných slabik a poměru hodnot pro zdůrazněné a nezdůrazněné slabiky. Přestože jsme neodhalili žádný významný rozdíl mezi zdravými kontrolami a PN pacienty, výsledky analýzy SPI ukázaly, že RS pacienti měli sklon zdůrazňovat slabiky nadměrně a do stejné míry. Analýza F0 poukázala na to, že RS pacienti rozšířili svůj rozsah F0 za účelem vyjádření důrazu na slabikách. V případě RS pacientů jsme dále zjistili, že i jejich rozsah intenzity byl nadměrný v porovnání se zdravými kontrolami. Konečně, analýza délky trvání slabik ukazovala na výskyt tzv. scanning speech u pacientů s RS.

Klíčová slova: dysartrie; akustická analýza; slovní přízvuk; Parkinsonova nemoc; roztroušená skleróza

Abbreviations

ANOVA	analysis of variance
CNS	central nervous system
EDSS	Expanded Disability Status Scale
F0	fundamental vocal frequency
HC	healthy control
MS	multiple sclerosis
MSD	motor speech disorder
PD	Parkinson’s disease
SPI	stress pattern index
UPDRS	Unified Parkinson’s Disease Rating Scale

Contents

1 Introduction

Language is a complex ability in humans to connect arbitrary symbols with their meanings. It provides an important means to express our feelings and thoughts to ourselves or to others (Purves et al., 2012). Language is present in cultures in various forms, such as manual, or written symbols, or the spoken language - speech.

Speech is a unique form of language. It is a powerful tool possessed by human beings that enabled many achievements of human culture (Purves et al., 2012). Despite the complexity of the underlying processes of speech production, a healthy individual is capable of smooth communication. However, a neurologic disturbance can negatively affect these processes (Duffy, 2013).

1.1 Speech Production

From the aspect of neurophysiology, language and speech are complex processes involving both central and peripheral nervous systems. In majority of people, the main functions related to language are located within the left hemisphere of the brain. Major brain regions involved in speech and language production and comprehension include Broca’s and Wernicke’s areas. However, they are supported by the primary motor, sensory, auditory, and visual cortices (Purves et al., 2012). Figure 1.1 depicts the particular locations of these areas and cortices important for speech production.

For smooth speech production, the language centres in the cortex are supported by subcortical parts. These parts include the thalamus, basal ganglia, and cerebellum. The thalamus is an important sensory processor and influences the emotional content of speech (Duffy, 2013; Speech 2017). The basic motor function of the basal ganglia relevant to speech production is to plan and program postural and supportive elements of the motor activity. The cerebellum helps coordinate the timing of movements and scale their size, thus the temporal and prosodic properties of speech are refined (Diener & Dichgans, 1992; Duffy, 2013).

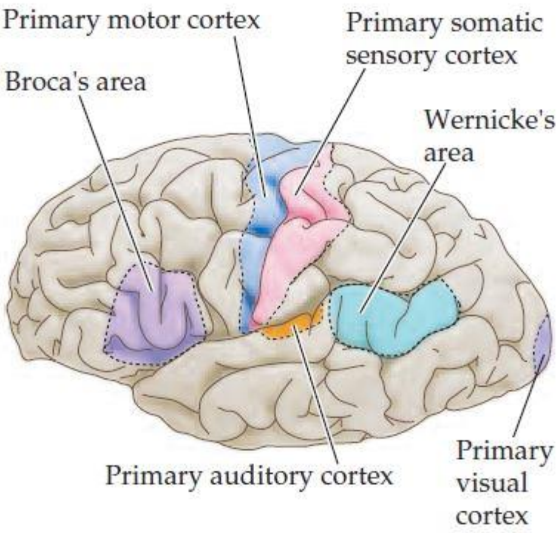


Figure 1.1 Major regions of the brain involved in language and speech production and comprehension. Adapted from Purves et al. (2012).

At the physiological level, speech production is dependent on the proper function of the speech apparatus. The speech apparatus can be divided into three subsystems – the respiratory, phonatory, and articulatory subsystem – which work closely together and are highly interactive (Kent & Read, 2002).

The respiratory subsystem comprises of the trachea, lungs, rib cage, and various muscles. It generates most of the aerodynamic energy of speech (Kent & Read, 2002). The larynx is at the top of trachea. It opens into the pharynx and forms the phonatory subsystem. It is composed of cartilages and muscles. Particularly important muscles are the vocal folds (Figure 1.2). Vocal folds adduct or abduct to close or open the laryngeal airway, respectively. The opening between them is termed the glottis. In Figure 1.2 only open and closed vocal folds are depicted. However, the glottis can open in various degrees. The degree of the glottal constriction determines voicing of the generated sound. When the vocal folds are largely abducted, voiceless sounds are generated. On the other hand, some sounds, such as stop consonants, require tight adduction (Kent & Read, 2002). When the vocal folds are adducted with less resistance, they vibrate and cause oscillations of the sound pressure. The frequency of these oscillations is evident in voiced speech sounds and determines a speaker’s vocal pitch (Purves et al., 2012).

The articulatory subsystem extends from the larynx through the nasal and oral cavities up to the lips and nose. Figure 1.2 shows the whole articulatory subsystem and its components called articulators. Articulators, which include the tongue, lips, jaw, and velum, constitute an important part of the system. Their movements shape the vocal tract, thus the resonance properties are determined (Kent & Read, 2002). The resonant frequencies are called formants. As the shape of the vocal tract changes, different patterns of formants are produced (Purves et al., 2012).

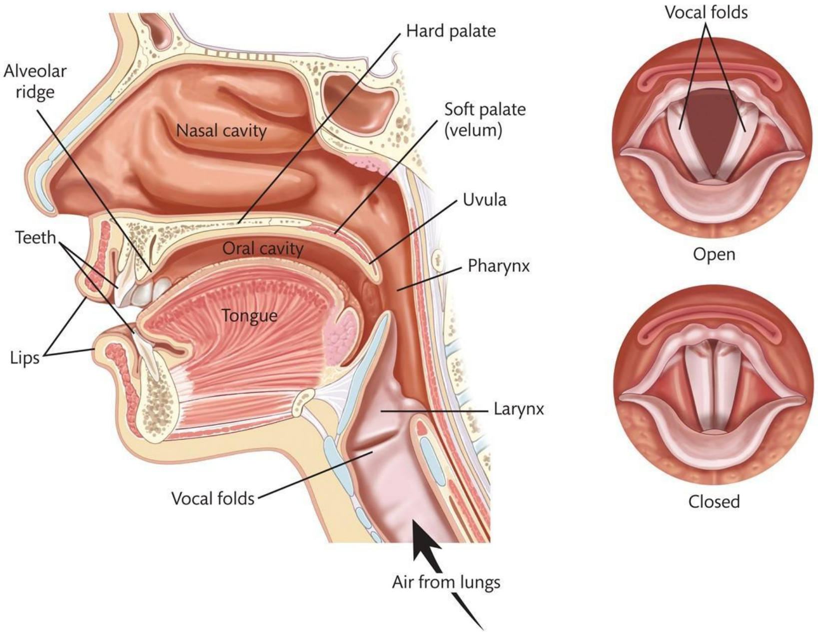


Figure 1.2 Vocal tract and vocal folds. Adapted from Kang (2017).

Speech is composed of dimensions that overlap and can be characterised as cooperating processes that result from functioning of the subsystems described above. These dimensions include prosody, phonation, articulation, resonance, and respiration. Prosody is the combination of the appropriate loudness, intonation, emphasis, and rhythm so that the speech sounds natural. Phonation is a process of production of speech sounds using the larynx. Articulation is the result of precise coordination of the articulators. Resonance is described as the process of changing the shape, size, or number of cavities to direct voice to resonate in them. Finally, respiration is the muscle activity used in breathing for speech.

1.2 Phonetic Aspects of Speech

Phonetics is a study that investigates speech sounds from the physiological level of their production to their acoustic properties. In phonetics, the speech sounds are principally divided into two categories. These categories are consonants and vowels. Consonants and vowels constitute larger units – syllables or words – that can be described by suprasegmental features. These features, among others, include variations in vocal pitch, loudness, duration, and stress, and constitute the speech dimension of prosody. To be of any importance in linguistics, values of these features have to be expressed relatively to the others in the same utterance (Phonetics, 2017).

1.2.1 Vocal Pitch

Modulations of the vocal pitch enhance expression of the emotional and prosodic content, thus a listener's comprehension. Speakers with higher vocal pitch are also perceived as those with higher voices, and vice versa. The vocal pitch is a result of vibrating vocal folds. Its acoustic correlate is the fundamental vocal frequency. The vocal pitch depends on the size, elasticity, and mass of the vocal folds and, therefore, it is not surprising that it also depends on speaker's gender and age. Due to the anatomical differences, the vocal pitch in women is generally higher than in men. Despite some ambiguous findings in previous literature (Nishio & Niimi, 2008; Ramig et al., 2001), the vocal pitch is generally thought to decrease with age in women and increase with age in men (Nishio & Niimi, 2008; Torre III & Barlow, 2009). Researchers also agree that the vocal pitch is less stable in older adults than in younger ones (Gorham-Rowan & Laures-Gore, 2006; Torre III & Barlow, 2009).

1.2.2 Loudness

Loudness is a perceived property of voice affected by the nature of speech sounds, emotional state of a speaker, or position in an utterance (Ververidis & Kotropoulos, 2006). It has also been shown that speakers tend to increase or decrease loudness of their voice at the beginning or end of an utterance, respectively (Lieberman, 1967). The important mechanism for increasing loudness is to increase amplitude of vocal fold vibration and the degree of adduction of the vocal folds (Scherer, 1991). However, it has also been found that perceived loudness is associated with precise articulation, or modulation of the resonant properties in speech production (Myers & Finnegan, 2015). In acoustics, loudness is predominantly related to acoustic intensity. Loudness is thought to be a gender-independent variable (Baker et al., 2001). Effects of age on loudness were also investigated, for example, by Baker et al. (2001) who suggested that with increasing age loudness of voice decreases in both genders.

1.2.3 Duration

Duration of an element in an utterance can be associated with accent, emotions, or a position of the token in an utterance. For example, it has been reported that a lengthened syllable can mark the end of an utterance (Cooper & Danly, 1981). Duration in an utterance is also closely related to breathing as speech is limited to periods of slow expiration which are followed by a phase of quick inspiration (Kent & Read, 2002). In an individual speaker, duration can be measured directly in milliseconds as duration of a particular token, or expressed as the speech rate which is a number of syllables per second. Studies of effects of age on speech rate showed that young adults tend to speak faster than older adults (Jacewicz et al., 2009; Quené, 2008). With respect to gender, researchers suggested that men speak slightly faster than women (Jacewicz et al., 2009; Yuan, Liber, & Cieri, 2006).

1.2.4 Stress

Stress in an utterance may serve various ways. There are three elements of an utterance on which stress can occur – a sentence, phrase, or word. Sentence stress, or prosodic stress, is used to emphasise certain part of a sentence. Phrasal stress helps differentiate phrases when their sequences of words are the same at the segmental level. Word stress describes the stress pattern on each syllable in a word (Accent, 2017). The stress patterns vary across languages. In Czech language, the word stress pattern is fixed and occurs usually on the first syllable of a word (Halle, 1997).

A stressed part of an utterance is usually perceived and distinguished by a listener as a louder part compared to the unstressed part. On the other hand, from the standpoint of a speaker stress can be described as larger amount of effort given to speech production (Lehiste, 1976). In the domain of acoustics, stress in speech is usually described by three parameters: duration, vocal pitch, and intensity (Bolinger, 1961; Fry, 1955; Lehiste, 1976).

The fundamental vocal frequency that a speaker modulates to convey stress is considered to be the prominent parameter (Atkinson, 1978; Morton & Jassem, 1965). However, intensity and duration have been also showed to describe stress (Cooper, Eady, & Mueller, 1985; Fry, 1955; Huss, 1978; Sluijter & Van Heuven, 1996). In addition, Tykalová et al. (2014) suggested a parameter to quantify stress termed the stress pattern index that is based on all three acoustic correlates of stress. To the best of our knowledge, age and gender differences have never been thoroughly investigated. Nonetheless, stress can be expected to be gender- or age-dependent since the vocal pitch, loudness, and duration are.

1.3 Neurological Disorders Affecting Speech Production

Neurological disorders can affect any part of the central or peripheral nervous system, thus speech production may be affected as well. The impairment in speech can appear at any point of the development of a neurological disease. In addition, defects in speech may be the first signs of a neurological disorder (Rusz et al., 2011). Since the symptoms and signs of a disorder generally reflect the location of a lesion, speech impairment can indicate the underlying neuropathology (Duffy, 2013). The damage in the brain can be caused by inflammatory, neoplastic, vascular, or degenerative diseases, or traumatic injuries. Vascular diseases are probably the most common cause of a speech impairment. The most frequent vascular disease in the nervous system is stroke (Duffy, 2013).

As our focus in this work is on people with Parkinson's disease and multiple sclerosis, only these two diseases are described in detail in the following subsections.

1.3.1 Parkinson's Disease

In 1817, James Parkinson described a disease called "Shaking Palsy" and provided six case studies of the disease (Parkinson, 1817). Later in the 19th century, Jean Martin Charcot suggested to name this disease Parkinson's disease (Lees, Hardy & Revesz, 2009).

Parkinson's disease (PD) is the second most frequent degenerative disease of the nervous system after Alzheimer's disease (De Lau & Breteler, 2006). The age of onset is usually between the ages of 50 and 70, with a mean age being 55. The disease leads to death in 10 to 20 year after diagnosis (Purves et al. 2012; Dauer & Przedborski, 2003). It affects 2 – 3% of the population over the age of 65 (Williams-Gray & Worth, 2016). Aging appears to be the major risk factor, although 10% of people with PD are under 45 years of age (Lees, Hard & Revesz, 2009; Lang & Lozano, 1998). About 90% to 95% of PD cases are referred to as sporadic (Dauer & Przedborski, 2003).

PD is characterised by four major features – tremor at rest, rigidity, bradykinesia, and postural instability (Jankovic, 2008). Furthermore, patients show diminished facial expressions and lack of associated movements, such as arm-swinging during walking. They have difficulties to initiate some movements, and to terminate them (Purves et al., 2012). A speech disorder develops in about 90% of PD cases (Duffy, 2013). Apart from the motor characteristics, non-motor symptoms including cognitive impairment, neuropsychiatric symptoms (such as depression, hallucinations, and dementia), sleep disorders and autonomic symptoms (such as sexual dysfunction, sweating, and bladder disturbances) also appear (Chaudhuri, Healy & Schapira, 2006).

The defects in motor function are caused by the progressive loss of dopaminergic neurons in the part of the brain called substantia nigra. In the normal state, dopaminergic effects decrease the inhibitory outflow of the basal ganglia. Therefore, the excitability of upper motor neurons is increased. When dopaminergic cells in the substantia nigra pars compacta are destroyed, the inhibitory flow becomes abnormally high. Thus, the probability of timely thalamic activation of upper motor neurons in the motor cortex lessens (Purves et al., 2012).

The most effective drug for PD treatment is levodopa. After initiation of treatment, in about 80% of patients, bradykinesia and rigidity improves significantly, whereas improvement in tremor is unpredictable (Macphee & Stewart, 2012). Another possibility of treatment is dopamine agonists. Both of the therapies are usually accompanied by adverse side effects (Williams-Gray & Worth, 2016).

1.3.2 Multiple Sclerosis

Multiple sclerosis (MS) was first described by Jean Martin Charcot in 1868. He observed patients with intermittent neurologic problems and found that the inflammatory cells accumulate in a perivascular distribution in white matter in the brain and spinal cord (Hafler, 2004).

MS is a disease of the central nervous system (CNS) characterised by a variety of symptoms that result from demyelination and inflammation along axons in multiple regions in the brain (Purves et al., 2012). It is the most common CNS disorder in young and middle-aged adults (Duffy, 2013). The age of onset is usually between 20 and 40 years (Goldenberg, 2012). The disease affects women more than men, in a ratio of 2:1 (Milo & Kahana, 2010).

MS cases are usually categorised into four groups according to the course of the disease. About 85% of patients experience exacerbation of symptoms followed by periods when symptoms remit or disappear (Goldenberg, 2012). This form of MS is called relapsing-remitting MS. The other categories include primary and secondary progressive MS and progressive-relapsing MS. Primary progressive MS develops in around 10% of MS patients (Goldenberg, 2012). It is characterised by gradually worsening symptoms with no periods of remission. Secondary progressive MS can develop in patients with relapsing-remitting MS. Patients with secondary progressive MS never fully recover from relapses and the disabilities continue to worsen. Progressive-relapsing MS is a rare form of MS affecting less than 5% of MS patients (Goldenberg, 2012). In this form, MS is progressive from the onset with periods of superimposed relapses (Milo & Kahana, 2010).

Symptoms and signs of MS are dependent on the affected regions (Purves et al., 2012). Therefore, any symptoms of a CNS disease can appear (Duffy, 2013). Very common are fatigue, depression, motor weakness or paralysis, bladder dysfunction, monocular blindness, and difficulties with speech (Compston & Coles, 2008; Hauser & Oksenberg, 2006). Speech impairment occurs in 25% to 50% of people with MS (Duffy, 2013).

MS affects diverse areas in the brain, especially white matter and periventricular areas, the spinal cord, brainstem, and optic nerves. The damaged regions are characterised by demyelination associated with inflammation, and death of oligodendrocytes (cells that synthesise myelin) (Duffy, 2013). In some cases, the axons themselves are destroyed. The loss of myelin sheaths impairs action potential conduction. The result is aberrant patterns of nerve conduction that give rise to most of the clinical problems of the disease (Purves et al., 2012).

The cause of MS remains unknown (Duffy, 2013; Milo & Kahana, 2010; Purves et al, 2012). The immune system contributes to the damage in the brain but it is not clear what activates it to cause the impairment. The most popular hypothesis claims that MS is an autoimmune disease occurring in a genetically susceptible individual triggered by environmental factors (Compston & Coles, 2008).

With respect to disease management, existing therapies are aimed at reducing the disease activity, shortening periods of exacerbations and improving quality of life. Severe relapses of MS are usually treated with a short course of corticosteroids (Goldenberg, 2012; National Multiple Sclerosis Society, 2017).

1.4 Motor Speech Disorders

Speech disorders which result from neurological impairments affecting motor speech processes are referred to as motor speech disorders (MSDs). Motor speech processes are combined processes of planning, programming, control, and execution of speech. MSDs include two distinct groups of disorders – apraxia of speech, and dysarthria (Duffy, 2013).

1.4.1 Dysarthria

Dysarthria is a name for a whole group of neurologic speech disorders. It reflects abnormalities in the strength, speed, range, steadiness, tone, or accuracy of movements required for speech production. The neuropathophysiologic disturbances occur due to weakness, spasticity, incoordination, involuntary movements, or excessive, reduced or variable muscle tone (Duffy, 2013; Rampello et al., 2016). There are a few categories of dysarthria that can be distinguished by their perceptual characteristics, and underlying neuropathophysiology. The main dysarthria subtypes are flaccid, spastic, ataxic, hypokinetic, hyperkinetic, unilateral upper motor neuron, and mixed. Accurate determination of the subtype of dysarthria can enhance localisation of the causal disorder (Duffy, 2013). The subtypes of dysarthria and their localisation are listed in Table 1.1.

Subtypes of dysarthria of particular interest in this work are hypokinetic and ataxic dysarthria. Therefore, detailed information only on these two dysarthria subtypes will be given in the following text.

Table 1.1 Localisation of the dysarthria subtypes (Duffy, 2013).

Dysarthria subtype	**Localisation**
Flaccid	Lower motor neuron
Spastic	Bilateral upper motor neuron
Ataxic	Cerebellum
Hypokinetic	Basal ganglia (Substantia nigra)
Hyperkinetic	Basal ganglia (Putamen and caudate)
Unilateral upper motor neuron	Unilateral upper motor neuron
Mixed	More than one

Hypokinetic Dysarthria

Hypokinetic dysarthria is a motor speech disorder linked with basal ganglia control circuit deterioration. One of the functions of the basal ganglia control circuit is to regulate execution of movements. Thus, malfunction of the circuit can reduce movement. In such a case, speech deficits that are characteristic for hypokinetic dysarthria reflect the effects of rigidity, reduced force and range of movement (Duffy, 2013).

Features of hypokinetic dysarthria are most apparent in the processes of phonation, articulation, and prosody. Hypokinetic dysarthria is mainly characterised by monopitch, reduced stress, monoloudness, breathiness, short phrases, variable rate, short rushes of speech, and imprecise consonants (Darley, Aronson, & Brown, 1975). From a listener's perspective, these deficits may lead to flat, attenuated, or accelerated speech (Duffy, 2013).

Hypokinetic dysarthria can be caused by any process that affects the basal ganglia control circuit, e.g. cerebral hypoxia, or traumatic brain injury. Due to this association with basal ganglia pathology, PD is the prototypic disease related with hypokinetic dysarthria. It is estimated that about 90 % of PD patients suffer from hypokinetic dysarthria (Duffy, 2013). Speech in PD patients have been investigated in a number of studies. Cheang and Pell (2007) reported reduced loudness in production of phonemic and contrastive stress, and emotion. They also found the vocal pitch to be aberrant only in production of contrastive stress. Harel, Canizzaro, and Snyder (2004) suggested that the vocal pitch variability in PD decreases but can be normalised after initiating pharmacological treatment. These findings were also proposed by Harel et al. (2004).

Some researchers have also suggested that speech may reflect early symptomatic changes in PD (Holmes et al., 2000; King et al., 1993).

Ataxic Dysarthria

Ataxic dysarthria is caused by damage to the cerebellar control circuit. The function of the cerebellum is to coordinate movements of muscles and muscle groups (Ito, 1984). The cerebellum is responsible for precise timing of movements and scaling their size. Thus, a lesion in the cerebellum leads to incoordination of muscles. Consequently, ataxic dysarthria may occur. Characteristics of ataxic dysarthria are most evident at the levels of articulation and prosody (Duffy, 2013). The deficits in the articulatory level of speech are imprecise consonants, irregular articulatory breakdowns, and distorted vowels. The abnormalities in prosody include excess and equal stress, prolonged phonemes and intervals, and slowness. Such impaired speech is referred to as scanning speech (Hartelius et al., 2000). Scanning speech reflects the prolongation of speech elements, and equal stress on syllables (Duffy, 2013).

Ataxic dysarthria in its clear form is present only in spinocerebellar ataxia 6 whereas in other neurologic disorders it is usually accompanied by other dysarthria subtypes. Such disorders are those that affect the cerebellum including among others multiple system atrophy, Friedreich’s ataxia, or MS. Cerebellar malfunction, and subsequent ataxic dysarthria, can be also caused by vascular lesions or tumours (Duffy, 2013).

Speech impairment in MS occurs in 25% to 50% of people with MS (Duffy, 2013). The characteristic feature of dysarthria in MS is scanning speech. Scanning speech, as mentioned above, is also a feature of ataxic dysarthria (Hartelius et al., 2000). However, the speech disorder occurring in MS can show signs of almost any dysarthria type. When signs of two or more dysarthria subtypes appear, the dysarthria is referred to as mixed dysarthria. The most frequent mixed dysarthria in MS is spastic-ataxic dysarthria (Duffy, 2013).

Comparison of hypokinetic and ataxic dysarthria

Hypokinetic and ataxic dysarthria are two distinct dysarthria subtypes with different underlying neuropathophysiology. It is, therefore, not surprising that they also differ in characteristic speech deficits. One of the main differences in speech production between these two dysarthria subtypes is in the ability to convey stress properly as well as to speak at a natural level of loudness. While hypokinetic dysarthria is characterised by reduced stress and monoloudness, the characteristic features of ataxic dysarthria are excess or equal stress and excess loudness variation. Another specific feature of hypokinetic dysarthria is monopitch. In ataxic dysarthria, scanning speech and irregular breakdowns in articulation are considered to be prominent characteristics. Table 1.2 compares selected features of hypokinetic and ataxic dysarthria subtypes.

To the best of our knowledge, there are only a few studies that investigated speech performances of both PD and MS patient groups. For example, Kuo & Tjaden (2016) studied 12 PD and 15 MS patients in comparison to 14 healthy individuals using a task of passage reading. Speakers with MS in this study were diagnosed with spastic, ataxic, or spastic-ataxic dysarthria. Speakers with PD were diagnosed with hypokinetic or hyperkinetic dysarthria. The researchers reported significantly slower articulation rate in MS subjects compared to both, healthy speakers and speakers with PD. They also found loudness to be greater in individuals with MS than in those with PD. Another study conducted by Kuo, Tjaden, & Sussman (2014) was focused on acoustic correlates of intelligibility in faster-than-habitual speech of 30 MS and 16 PD speakers, and 14 healthy controls. Based on perceptual judgements, intelligibility in faster-than-habitual rate was found to vary for individual speakers with PD and MS. Subsequent acoustic analysis performed on selected PD and MS speakers showed that to maintain intelligibility in faster-than-habitual rate speakers increase articulation and speech rates, loudness and modulate fundamental vocal frequency.

1.5 Motivation

Dysarthria is a motor speech disorder accompanying a number of neurological disorders. In clinical practice, perceptual methods are widely used to assess the impairment of speech. These methods, however, are subjective in nature, difficult to quantify, and may significantly vary between two evaluators. In contrast, acoustic methods can provide a precise and objective method for assessment of speech.

Deviations in prosody, particularly in stress patterns, are often characteristic for dysarthria subtypes. However, stress production is a complex process involving several acoustic variables. Thus, a single parameter describing stress production could enhance the acoustic assessment of stress and be a step towards fully automatic speech evaluation, as well. In addition, speaking tasks that are commonly used nowadays are specialised tasks requiring thorough instructions. Therefore, it would be suitable to find a descriptor of stress that could be employed in unspecialised speaking tasks such as passage reading. Finally, deficits in stress production vary across dysarthria subtypes. Thus, a precise assessment of stress production in speech can also reveal the underlying neuropathophysiology and enhance the diagnosis of a disease.

Table 1.2 Selected features occurring in hypokinetic and ataxic dysarthria; -: never occurs; +: may occur but is not distinguishing; ++: prominent or distinguishing, or both (Duffy, 2013)

Feature	**Dysarthria subtype**	
	Hypokinetic	**Ataxic**
Slow rate	-	+
Variable rate	++	-
Excess and equal stress	-	++
Monopitch	++	-
Reduced stress	++	-
Monoloudness	++	-
Excess loudness variation	-	++
Irregular articulatory breakdowns	-	++
Breathiness	+	-
Distorted vowels	-	++
Prolonged phonemes	-	+
Short rushes of speech	++	-
Scanning speech	-	++

1.6 Objectives

In this study, we aimed to analyse word stress patterns in patients with hypokinetic and ataxic dysarthria compared to healthy speakers. For these purposes, we chose a dataset of PD patients who suffer from hypokinetic dysarthria and MS patients with predominant cerebellar damage who are expected to demonstrate mainly ataxic dysarthria.

Our main objective was to test well-known acoustic measurements of stress patterns as well as to design a novel methodology that would be suitable for evaluation of stress patterns in patients with both ataxic and hypokinetic dysarthria. Furthermore, this methodology was expected to be applicable to a speaking task of passage reading.

2 Methods

2.1 Participants

The participants of the study comprised three groups – the PD group, MS group, and healthy control (HC) group. The PD and MS patients were recruited in collaboration with Department of Neurology and Centre of Clinical Neuroscience, First Faculty of Medicine, Charles University in Prague.

The PD group consisted of 30 participants diagnosed with idiopathic PD, 13 of whom were males and 17 were females. Their age ranged between 41 and 77 years. The Unified Parkinson's Disease Rating Scale part III[1] (UPDRS III) score of the patients ranged from 6 to 38. The clinical diagnoses of all PD patients were established by a specialist in movement disorders according to the UK Parkinson's Disease Society Bank Criteria for PD (Hughes, Kilford, & Lees, 1992). At the time of the examination, all patients were treated pharmacologically and were on stable medication for at least 4 weeks, consisting of various doses of levodopa alone or in combination with different dopamine agonists. Disease duration in the PD group was estimated based on the self-reported occurrence of first motor symptoms.

The MS group was comprised of 30 patients including 11 males and 19 females. The age of these individuals ranged from 27 to 62 years. The overall Expanded Disability Status Scale[2] (EDSS) score ranged between 2.5 and 6.5. For the purpose of this study, patients with the cerebellar EDSS subscore higher or equal to 3 were selected from all MS patients originally recruited. To comprise a group of 30 MS subjects, three patients with cerebellar EDSS subscore of 2 were also added. Finally, the cerebellar EDSS subscores of all selected patients ranged from 2 to 4. Table 2.1 provides a detailed summary of the clinical characteristics of MS and PD patient groups. The clinical diagnoses of all MS patients were established by a physician according to the Revised MacDonald Criteria for diagnosis of multiple sclerosis (Polman et al., 2011). At the time of the examination, each patient treated pharmacologically was on stable medication for at least 4 weeks.

---

[1] UPDRS III is an objective measure of severity of parkinsonian motor symptoms; the range of values is from 0 to 56, the higher the value the more severe motor impairment (Goetz et. al, 2007).

[2] EDSS quantifies disability in multiple sclerosis. The scale ranges from 0 to 10 by 0.5 increments, 10 indicates the terminal stage of MS (Kurtzke, 1983).

Disease duration in the MS group was assigned as the number of years since establishing of the MS diagnosis.

The HC group consisted of 13 male and 17 female individuals aged between 41 and 77 years, with a mean of 60.5 and a standard deviation of 8.3. None of them reported any previous neurologic problems or disorders affecting hearing, speech, or language.

There were no significant differences in age distributions detected among the PD and HC groups (t-test, $p = 0.91$). The age of onset in PD is typically around 60 years whereas the age of onset in MS is typically around 30 years. Therefore, the mean age of the MS group was lower than the one of the PD group (t-test, $p < 0.01$).

The study was approved by the local ethics committee. Every participant gave written, informed consent to the participation in the project and recording procedure. All participants were Czech native speakers.

Table 2.1 Clinical characteristics of patient groups

	Parkinson's Disease *n* = 30; 13 men, 17 women			**Multiple Sclerosis** *n* = 30; 11 men, 19 women		
	Mean	**SD**	**Range**	**Mean**	**SD**	**Range**
Age (years)	60.8	8.1	41-77	44.2	9.5	27-62
Disease duration (years)	7.2	5.1	1-24	16.7	7.7	5-31
UPDRS III (-)	16.6	7.8	6-38	-	-	-
EDSS (-)	-	-	-	4.9	1	2.5-6.5
EDSS – cerebellar subscore (-)	-	-	-	3	0.5	2-4
Levodopa (mg/day)	719	463.2	0-2080	-	-	-

2.2 Recording Procedure and the Speaking Task

The recordings of all participants were obtained in a quiet room with low level of ambient noise. Each recording was performed during one session with a speech therapist. A head-mounted condenser microphone (Bayerdynamic Opus 55, Heilbronn, Germany) was used and placed approximately 5 cm from the speaker’s lips. The recorded speech was sampled at 48 kHz with 16-bit quantisation.

Because the aim of the current study was to investigate word stress patterns, we used passage reading as the speaking task. It provides good approximation of spontaneous speech while the prosodic and linguistic aspects of speech are controlled. The participants were instructed to read a given text aloud in a normal, comfortable loudness and pitch. The text used was a passage of 79 words written by Czech author Karel Čapek. The particular text of the passage can be found in Figure 2.1. Each participant was asked to read the text twice.

2.3 Manual Time Labelling

In order to investigate word stress patterns, the onset and closure of each syllable in every recording had to be manually labelled. For this purpose, we divided the whole text into particular syllables. A Czech phonetician (Jan Volín) provided a pattern for definition of syllabic boundaries related to each word in the text. The final number of syllables within the text was 153 (see Figure 2.1).

„|Když| |člo|věk| |po|prvé| |vsa|dí| |do| |ze|mě| |sa|ze|nič|ku|, |cho|dí| |se| |na| |ni| |dí|vat| |tři|krát| |de|nně|: |tak| |co|, |po|vy|ros|tla| |už| |ne|bo| |ne|?

|I| |ta|jí| |dech|, |na|klá|ní| |se| |nad| |ní|, |při|tla|čí| |tro|chu| |pů|du| |u| |je|jích| |ko|řín|ků|, |na|čech|rá|vá| |jí| |líst|ky| |a| |vů|bec| |ji| |ob|tě|žu|je| |růz|ným| |ko|ná|ním|, |kte|ré| |po|va|žu|je| |za| |u|ži|teč|nou| |pé|či|. |A| |když| |se| |sa|ze|nič|ka| |pře|sto| |u|jme| |a| |ros|te| |ja|ko| |zvo|dy|, |tu| |člo|věk| |žas|ne| |nad| |tím|to| |di|vem| |pří|ro|dy|, |má| |po|cit| |če|ho|si| |ja|ko| |zá|zra|ku| |a| |po|va|žu|je| |to| |za| |je|den| |ze| |svých| |nej|vět|ších| |o|sob|ních| |ú|spě|chů|.“

Figure 2.1 Definition of syllabic boundaries. The syllabic boundaries here are marked by the straight lines.

Subsequently, time labelling was performed according to the given pattern in PRAAT® (Boersma, 2002), a software specialised in speech analysis. The software creates an object for labelling called TextGrid. The TextGrid object is composed of a number of tiers defined by the user. The tiers can be either interval tiers which consist of labelled intervals or point tiers which are sequences of labelled points.

The boundaries of syllables were determined by the graphical representation of the speech signal in the time domain, the spectrogram, and contours of the fundamental vocal frequency, formant frequencies, and intensity supported by listening. Figure 2.2 shows an example of syllabic boundaries within a word. During the labelling procedure, several basic rules were taken into account. Firstly, when a stop consonant appeared at the onset of the syllable, we omitted the silent stop gap (the interval when the vocal tract is blocked) and labelled the onset at the burst of the consonant (see Figure 2.3). Secondly, if a multiple burst occurred, we labelled the first burst. Furthermore, the closure of a syllable ending with voiced sound followed by a pause was determined by abrupt weakening of formants and marked decreasing of intensity. A similar rule was applied on syllables within words followed by a stop consonant (see syllables ‘při’ and ‘tla’ in Figure 2.2). If a syllable ending with voiced sound was followed by another voiced segment, the boundaries between syllables were estimated based on the maximal changes in the first and second formant frequency contours.

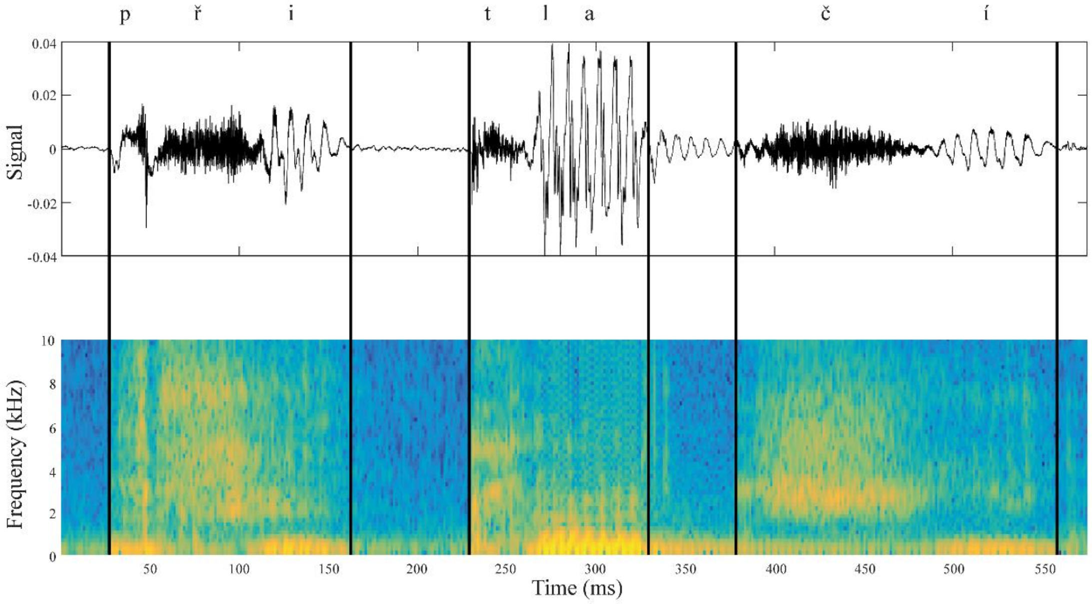


Figure 2.2 Syllabic boundaries within the word 'přitlačí' in the time and frequency domain.

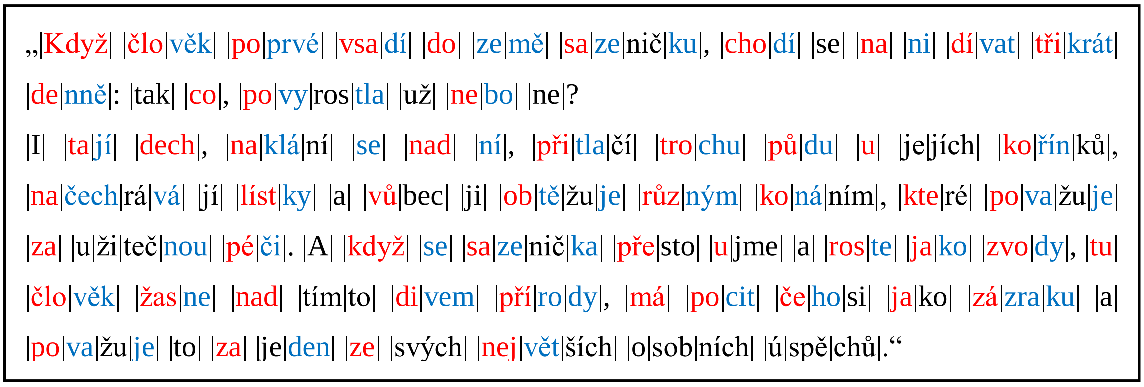
In case of a missing syllable (e.g. when a speaker misread or omitted a syllable), we assigned two labels within an interval of approximately 10 ms in an appropriate tier. Nevertheless, the missing data formed an insignificant portion of all data. In particular, the missing syllables in recordings of PD patients composed 0.5 % of all syllables, while in the HC and MS groups 0.3 % of the syllables were missing. For detailed information see Table 2.2.

Table 2.2 Numbers of recordings in groups with no, one, or more than one missing syllable. The maximum of missing syllables in an HC and PD recording was four, in the MS the maximum was seven.

No. of missing syllables	**HC**	**PD**	**MS**
0	36 (60 %)	28 (46.7 %)	43 (71.7 %)
1	19 (31.7 %)	22 (36.7 %)	14 (23.3 %)
> 1	5 (8.3 %)	10 (16.6 %)	3 (5 %)

2.4 Definition of Stressed and Unstressed Syllables

Theoretical word stress patterns were provided by 3 evaluators including a phonetician (Jan Volín), Czech language teacher (Iva Novotná), and speech therapist (Hana Růžičková). However, due to occasional ambiguities in some syllables, more than one pattern was given. Therefore, we created one final theoretical stress pattern based on congruence of all 3 suggested stress patterns (red and blue syllables in Figure 2.3). When only 2 of the 3 suggested patterns agreed on a syllable, we excluded the syllable from the further evaluation (black syllables in Figure 2.3). Furthermore, last two words were because the quality of reading along with loudness usually declines in speakers at the end of a read passage.



„|Když| |člo|věk| |po|prvé| |vsa|dí| |do| |ze|mě| |sa|ze|nič|ku|, |cho|dí| |se| |na| |ni| |dí|vat| |tři|krát| |de|nně|: |tak| |co|, |po|vy|ros|tla| |už| |ne|bo| |ne|?
|I| |ta|jí| |dech|, |na|klá|ní| |se| |nad| |ní|, |při|tla|čí| |tro|chu| |pů|du| |u| |je|jích| |ko|řín|ků|, |na|čech|rá|vá| |jí| |líst|ky| |a| |vů|bec| |ji| |ob|tě|žu|je| |růz|ným| |ko|ná|ním|, |kte|ré| |po|va|žu|je| |za| |u|ži|teč|nou| |pé|či|. |A| |když| |se| |sa|ze|nič|ka| |pře|sto| |u|jme| |a| |ros|te| |ja|ko| |zvo|dy|, |tu| |člo|věk| |žas|ne| |nad| |tím|to| |di|vem| |pří|ro|dy|, |má| |po|cit| |če|ho|si| |ja|ko| |zá|zra|ku| |a| |po|va|žu|je| |to| |za| |je|den| |ze| |svých| |nej|vět|ších| |o|sob|ních| |ú|spě|chů|.“

Figure 2.3 The stress pattern. Stressed syllables are given in red, the unstressed syllables are in blue. The syllables in black are those with a high level of ambiguity in emphasis or at the end of the passage. The straight lines depict the syllabic boundaries.

2.5 Acoustic Analysis

2.5.1 Extraction of Fundamental Vocal Frequency and Intensity Contours

The fundamental vocal frequency (F0) curve of each speaker was automatically detected and calculated using the autocorrelation method in software PRAAT®. The voicing threshold and other available detection settings were kept at default values of PRAAT®. However, it was necessary to consider the various F0 ranges of the speakers. Thus, we set the F0 boundaries (minimum and maximum eligible F0 value) individually for every speaker to ensure correct detection of F0. Since the F0 detection was automatic, we checked and modified the data manually if a correction was necessary. The resulting values of F0 were expressed in hertz and listed in a text file. As the F0 curve was obtained from the whole signal, the file contained samples with undetected F0, e.g. voiceless parts of speech, as well as the actual values of F0.

We also obtained the values of the speech intensity of the whole signal from PRAAT®. The default settings of the intensity range were adjusted when necessary. The values of intensity contours were expressed in decibels and listed in a text file.

2.5.2 Algorithm for the Stress Pattern Index Calculation

As the descriptor of stress we employed stress pattern index (SPI) defined by Tykalová et al. (2014) as

$$SPI = \left[1 + \log \frac{\mathrm{F0}_{max}}{F0_{min}}\right] \sum_{n} E_n \tag{2.1}$$

$\mathrm{F0}_{\mathrm{max}}$ and $\mathrm{F0}_{\mathrm{min}}$ are the maximum and minimum values of F0 in a syllable, respectively. $\sum_n E_n$ is the cumulative sum of energy of the syllable. We adapted the processes done in MATLAB for calculation of SPI. The scheme of the algorithm is given in Figure 2.4.

The initial step of the algorithm was to load the TextGrid file with time labels, and text files with the values of intensity and F0. Each file contained a different type of data, thus the processes of loading were adapted to them.

Since the objective of the algorithm was to assign a value to each syllable, the whole process required information on the timing, i.e. the time labels of the syllabic boundaries. Therefore, the next procedure was extracting of the time labels from a TextGrid file.

Subsequently, we calculated the cumulative sum of the energy in each syllable employing the time labels. Thereafter, the algorithm processed the information on F0. Since F0 is a quantity that cannot be always detected in speech, we had to eliminate the samples with undetected frequency to enable further processing. In the next step, we applied the time labels on to the remaining F0 values and found the maximum and minimum F0 values in every syllable. In this step, we utilised the corrections made to the automatically detected F0 curves. Eventually, we

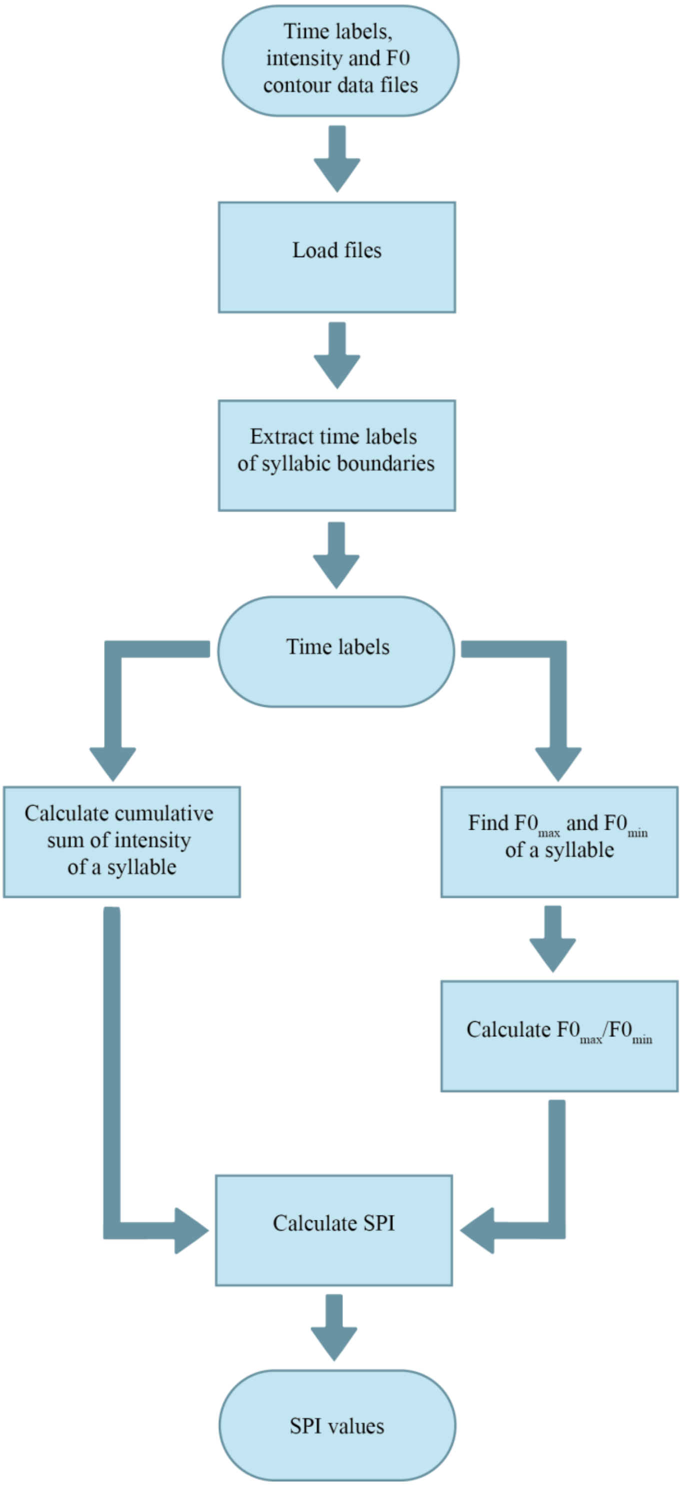


Figure 2.4 Flow chart of the algorithm. After loading all required data files, time labels are extracted and used in calculation of the cumulative sum of energy and finding the minimum and maximum F0 values in each syllable. The ratio of the two F0 values is calculated, and used along with the energy in the SPI calculation.

obtained a matrix of the maximum and minimum values of F0 for every syllable. It was necessary to take into consideration that ranges of F0 are gender-dependent. Therefore, we converted the values of the matrix, which were originally in hertz, to semitones (st) to eliminate this difference. The conversion was done using following formula:

$$f_{\mathrm{st}} = 69 + 12\ \log_2\left(\frac{f_{\mathrm{Hz}}}{440}\right) \tag{2.2}$$

where f_{Hz} denotes the frequency in hertz (Jang, 2005).

As the final step in F0 processing, we calculated the ratio of $F0_{\mathrm{max}}$ to $F0_{\mathrm{min}}$.

Finally, we employed the formula 2.1 to calculate the SPI values. As a result, we obtained a vector of 153 SPI values corresponding to each syllable in a recording.

2.6 Data Analysis

As we mentioned earlier in Section 2.3, some of the data were missing, usually due to misreading. Therefore, before the actual data analysis we eliminated the values of each analysed acoustic variable corresponding to the missing syllables.

In the next step, we took into consideration that every speaker was recorded twice. Thus, we calculated the arithmetic mean of the values of the acoustic variables describing the two signals. If one the values was eliminated earlier, the other value was assigned as the resulting one. Consequently, each speaker was described by vectors of acoustic variables corresponding to each syllable in the text that was read by the speaker.

In Section 2.4 we showed the theoretical stress pattern in the read passage. However, final set of selected syllables had to be further reduced because of potential effects of the syllable duration on the result. Vowels in Czech language are of three types – long vowels, short vowels, and diphthongs. Therefore, if a long vowel or diphthong appears in a syllable, it increases duration of the syllable although the syllable may not be stressed. A similar case may be caused by the presence of a fricative consonant in a syllable. An opposite problem could occur when a stop consonant was present on the onset of a syllable. In such case, duration of the syllable could be decreased. Therefore, the final selection of syllables was composed of syllables of approximately similar duration in order to minimise these effects.

Thus, as a result, we selected a set of 50 syllables for each speaker and recording for further analysis – 25 of them were supposed to be stressed, while other 25 unstressed. The final set of syllables used in the analysis is depicted in Figure 2.5.

Analysed features

Each syllable was described by four acoustic variables including the SPI, fundamental vocal frequency ratio (F0 ratio; ratio of the maximum and minimum F0 value), intensity ratio (ratio of the maximum and minimum intensity value), and duration. We calculated the mean of these variables in three sets of syllables that included the stressed syllables, unstressed syllables, and the whole syllable set in each speaker. Furthermore, we analysed the ratio of the resulting mean values of the stressed and unstressed syllables to seek differences between the sets.

„|**Když**| |**člo**|věk| |po|prvé| |**vsa**|dí| |do| |ze|mě| |**sa**|ze|nič|ku|, |**cho**|dí| |se| |na| |ni| |**dí**|vat| |tři|krát| |de|nně|: |tak| |**co**|, |po|vy|ros|tla| |už| |ne|bo| |ne|?

|I| |ta|jí| |dech|, |na|klá|ní| |se| |**nad**| |ní|, |**při**|tla|čí| |tro|chu| |**pů**|du| |u| |je|jích| |ko|řín|ků|, |na|čech|rá|vá| |jí| |líst|ky| |a| |vů|bec| |ji| |**ob**|tě|žu|je| |**růz**|ným| |ko|ná|ním|, |**kte**|ré| |po|va|žu|je| |**za**| |u|ži|teč|nou| |**pé**|či|. |A| |**když**| |se| |**sa**|ze|nič|ka| |pře|sto| |u|jme| |a| |**ros**|te| |ja|ko| |**zvo**|dy|, |tu| |**člo**|věk| |žas|ne| |**nad**| |tím|to| |di|vem| |pří|ro|dy|, |**má**| |po|cit| |**če**|ho|si| |ja|ko| |zá|zra|ku| |a| |po|va|žu|je| |to| |**za**| |je|den| |ze| |svých| |**nej**|vět|ších| |o|sob|ních| |ú|spě|chů|.“

Figure 2.5 Syllables selected for the final analysis. The final set of stressed syllables is shown in bold, the set of unstressed syllables is underlined.

2.7 Statistical Analysis

First, we confirmed the normal distribution of the data by the Kolmogorov-Smirnov test. Since the data were normally distributed, the analysis of variance (ANOVA) with post hoc Bonferroni test was performed on the acoustic features to search for differences among the HC, PD, and MS groups. The nominal alpha level was set to 0.05.

The analyses were graphically depicted in box plots with boxes extending from the first to the third quartile, the median visualised as a band inside the box, and whiskers extending from the first quartile to the lowest value within 1.5 IQR and from the third quartile to the highest value within 1.5 IQR.

3 Results

3.1 Stress Pattern Index

Statistically significant group differences were found in both mean SPI values of stressed [$F(2,87) = 3.8$, $p < 0.05$] and mean SPI values of unstressed [$F(2,87) = 6.95$, $p < 0.01$] syllable sets. In the stressed syllable set, a post hoc Bonferroni test showed that MS speakers reached higher SPI values than HC speakers ($p < 0.05$). In case of the unstressed syllable set, the MS group also showed significantly higher SPI values than the HC group ($p < 0.01$). We also found significant group differences in the ratio of the mean SPI value of stressed and unstressed syllables [$F(2,87) = 7.27$, $p < 0.01$]. A post hoc test revealed that HC speakers had significantly higher stressed-to-unstressed mean SPI ratio than MS speakers ($p < 0.05$). Statistically significant group differences were also found in mean SPI of the whole syllable set [$F(2,87) = 5.64$, $p < 0.01$]. A post hoc test showed that the mean SPI of the whole set was significantly higher in the MS group than the one of the HC group ($p < 0.01$). The results of statistical analyses of the mean SPI are depicted in Figure 3.1.

3.2 Fundamental Vocal Frequency Ratio

Statistically significant group differences in the mean F0 ratio were found in the separate sets of the stressed [$F(2, 87) = 6.67$, $p < 0.01$] and unstressed [$F(2, 87) = 3.87$, $p < 0.05$] syllables. In the stressed syllable set, post hoc comparisons revealed that MS group had higher values than both the HC ($p < 0.05$) and PD ($p < 0.01$) group. In the unstressed syllable set, the HC group was found to have higher values than the PD group ($p < 0.05$). We also found significant differences in the ratio of the values of stressed and unstressed syllables [$F(2, 87) = 6.54$, $p < 0.01$]. A post hoc test revealed that the HC group had lower ratio of the values than both the PD ($p < 0.05$) and MS ($p < 0.01$) groups. Group differences in the mean F0 ratio of the whole syllable set were also found to be significant [$F(2, 87) = 4.84$, $p < 0.05$]. In this case, a post hoc test showed that the MS group had significantly higher values than the PD group ($p < 0.01$). Results of analyses of the mean F0 ratio are shown in Figure 3.2.

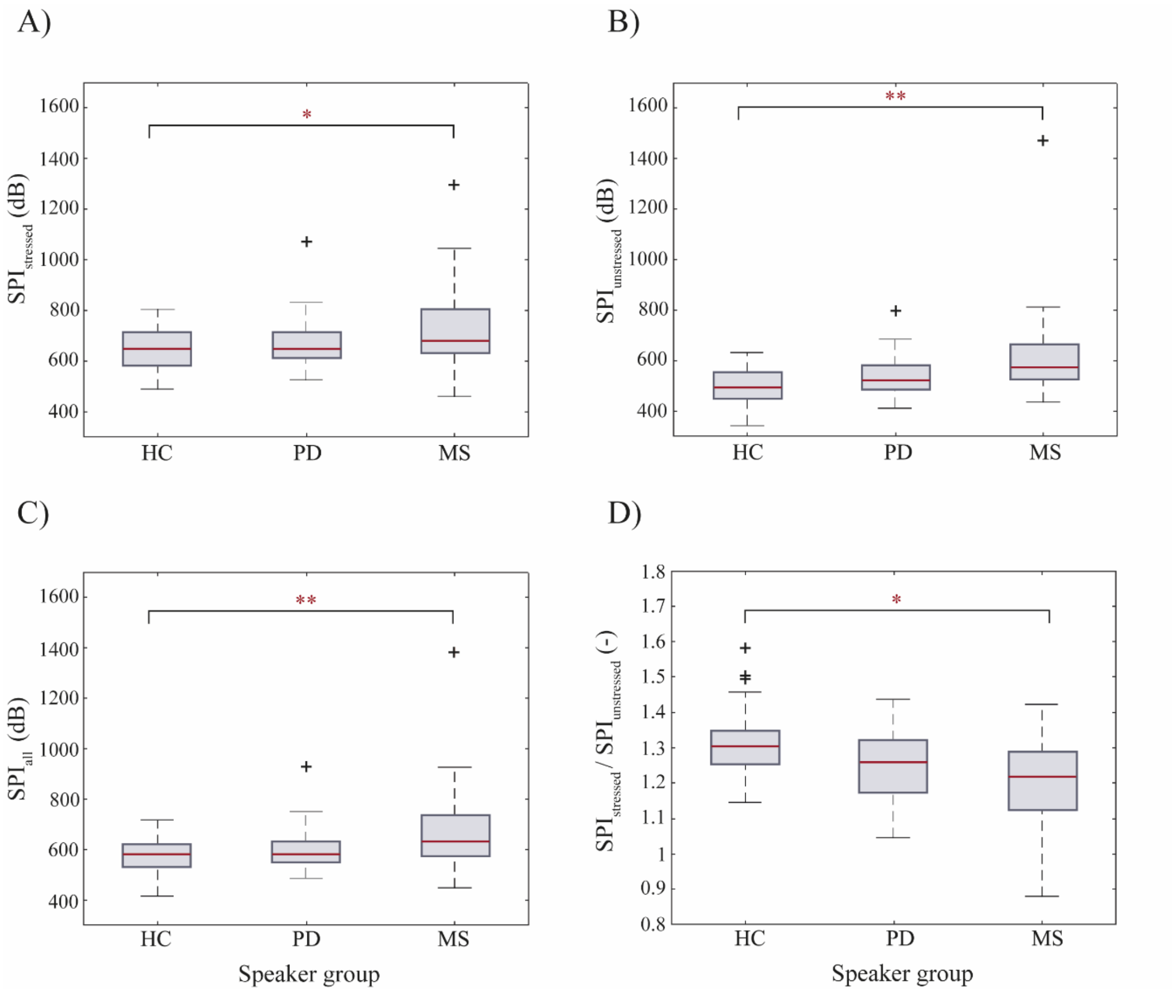


*Figure 3.1 Results of analyses of the mean SPI. A) SPI of stressed syllables. B) SPI of unstressed syllables. C) SPI of the whole syllable set. D) Ratio of the SPI of stressed syllables and SPI of unstressed syllables. * $p < 0.05$, ** $p < 0.01$*

3.3 Intensity Ratio

Statistically significant group differences in were found the mean intensity ratio of both stressed [$F(2, 87) = 21.45$, $p < 0.001$] and unstressed [$F(2, 87) = 9.8$, $p < 0.001$] syllable sets. In case of the stressed syllable set, a post hoc test revealed that MS speakers produced intensity in a higher ratio than HC ($p < 0.001$) and PD ($p < 0.001$) speakers. In the unstressed syllable set, MS speakers were again found to reach higher intensity values than HC ($p < 0.01$) and PD ($p < 0.001$) speakers. Group differences in the ratio of the values of stressed and unstressed syllables were also found

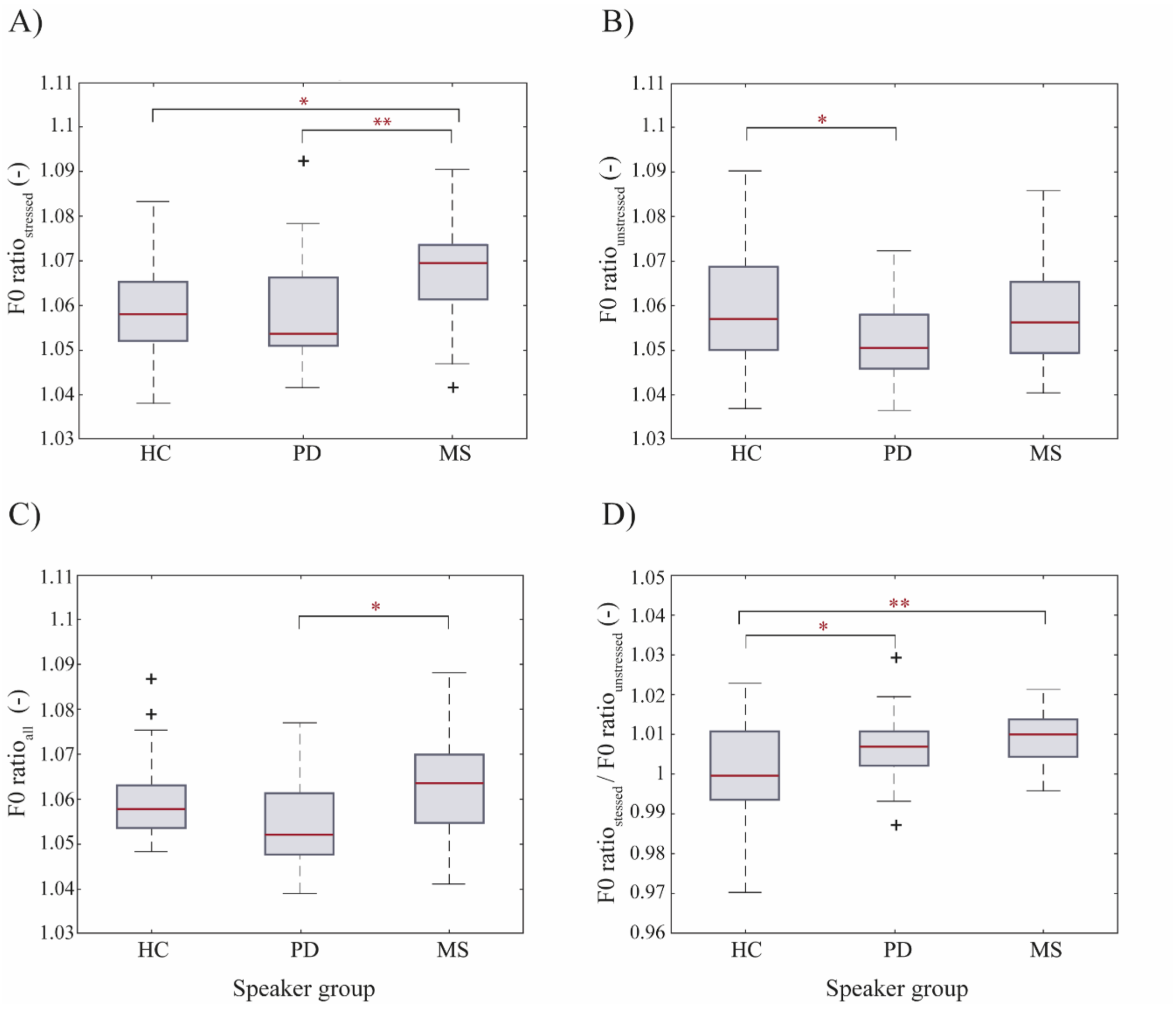


*Figure 3.2 Results of analyses of the mean F0 ratio. A) F0 ratio of stressed syllables. B) F0 ratio of unstressed syllables. C) F0 ratio of the whole syllable set. D) Ratio of the F0 ratio of stressed syllables and F0 ratio of unstressed syllables. * $p < 0.05$, ** $p < 0.01$*

to be statistically significant [$F(2, 87) = 7.49$, $p < 0.01$]. Similarly as in the previous cases, post hoc comparisons revealed that MS speakers reached higher values than HC ($p < 0.01$) and PD ($p < 0.05$) speakers. Statistically significant group differences were also found in the mean intensity ratio of the whole syllable set [$F(2, 87) = 18.21$, $p < 0.001$]. Here, a post hoc test showed again that MS speakers reached higher values of the intensity ratio than HC ($p < 0.001$) and PD ($p < 0.001$) speakers. Results of analyses of the mean intensity ratio are depicted in Figure 3.3.

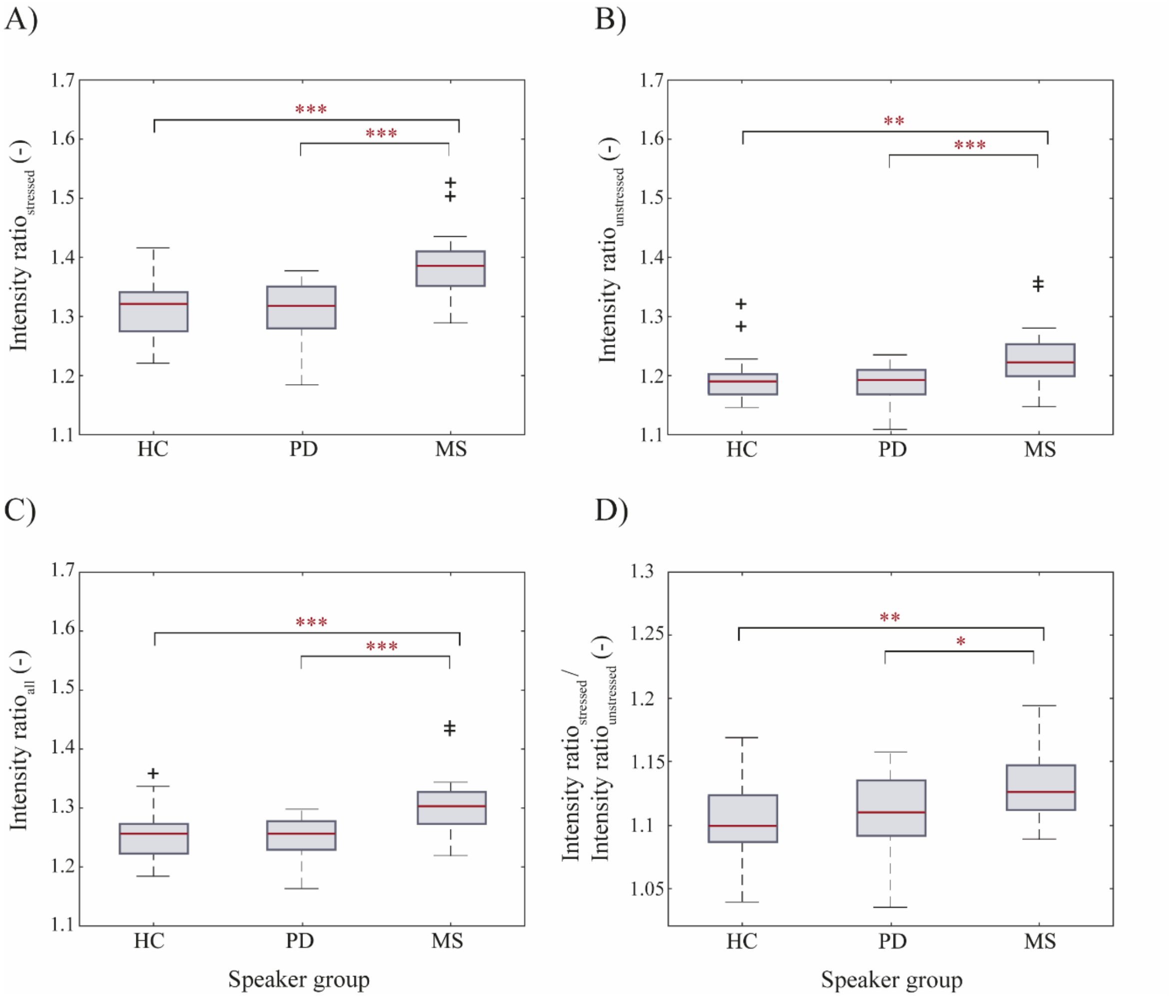


*Figure 3.3 Results of analyses of the mean intensity ratio. A) Intensity ratio of stressed syllables. B) Intensity ratio of the unstressed syllables. C) Intensity ratio of the whole syllable set. D) Ratio of the intensity ratio of stressed syllables and intensity ratio of unstressed syllables. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$*

3.4 Duration

Statistically significant group differences were found in mean duration of the stressed syllables [$F(2, 87) = 11.08$, $p < 0.001$]. A post hoc test further revealed that mean duration of the stressed syllables in MS speakers was significantly longer than the one of HC ($p < 0.001$) and PD ($p < 0.001$) speakers. Similarly, mean duration of the unstressed syllables was found to be significantly different among the groups [$F(2, 87) = 13.15$, $p < 0.001$]. Post hoc comparisons revealed that mean unstressed syllable duration of MS speakers was significantly longer than the one of HC ($p < 0.001$) and PD ($p < 0.001$) speakers. Furthermore, significant group differences were found in the ratio of mean stressed and unstressed syllable duration [$F(2, 87) = 7.39$, $p < 0.01$]. A post hoc test showed that the ratio was significantly higher in HC speakers than in both PD ($p < 0.05$) and MS ($p < 0.001$) speaker groups. Significant differences among the groups were also found in the mean duration of the whole syllable set [$F(2, 87) = 12.85$, $p < 0.001$]. A post hoc test showed that mean duration values of the set in MS speakers were higher than in both HC ($p < 0.001$) and PD ($p < 0.001$) speaker groups. Results of analyses of mean duration are shown in Figure 3.4.

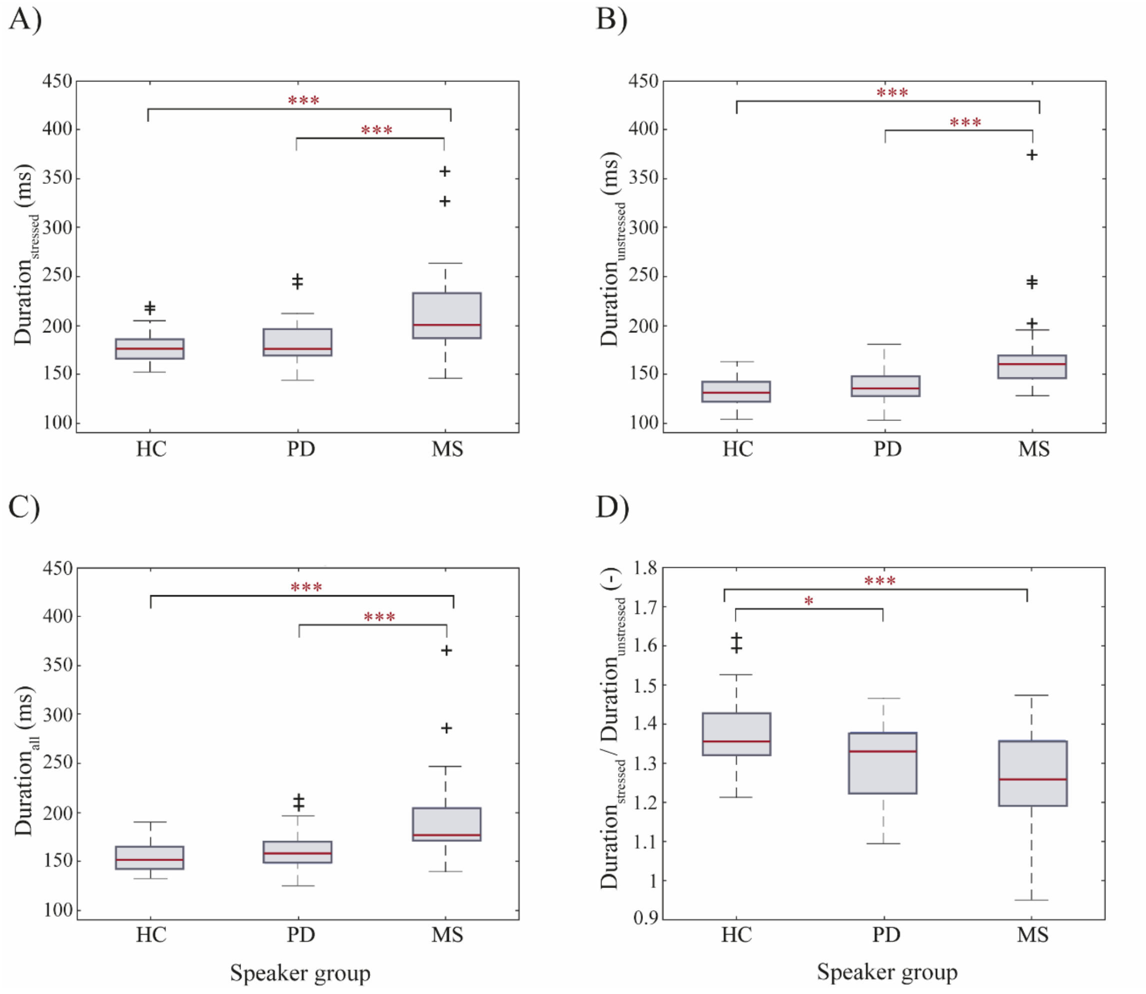


*Figure 3.4 Results of analyses of the mean duration. A) Duration of stressed syllables. B) Duration of unstressed syllables. C) Duration of the whole syllable set. D) Ratio of duration of stressed syllables and duration of unstressed syllables. * $p < 0.05$, *** $p < 0.001$*

4 Discussion

In this study, we focused on quantifying and analysing word stress in PD and MS patients compared to HC using a speaking task of passage reading. In order to quantify stress, we employed the SPI, a parameter that takes into account all three basic acoustic variables associated with word stress production. Results of SPI analyses showed that deviations in word stress exploitation were detectable in MS patients but not in PD patients. Furthermore, analyses of individual acoustic variables indicated a variety of characteristic features, such as excess and equal stress, scanning speech, and excess loudness variation, that are frequently found in dysarthria associated with MS.

With respect to SPI analyses, we found deviations in word stress in MS patients that are commonly described in literature as features of ataxic dysarthria. In comparison with HC, MS patients reached higher SPI values in both stressed and unstressed syllables. This finding indicates a finding consistent with a feature of ataxic dysarthria described by Duffy (2013), that is our MS patients with predominant cerebellar damage tended to stress syllables excessively. In addition, low difference in SPI between stressed and unstressed syllables suggests that our MS patients stressed all syllables to the same extent or equally. Equally stressed syllables are characteristic for scanning speech, which is perceptually well defined speech deficit often associated with ataxic dysarthria (Hartelius et al., 2000). In summary, our observations are in agreement with features of ataxic dysarthria described by Duffy (2013) and Hartelius et al. (2000). On the other hand, our results did not indicate any difference in SPI between HC and PD speakers. This finding is in contrast to a study published by Tykalová et al. (2014) who reported that the SPI revealed the difference in contrastive stress production between PD patients and HC speakers. However, it should be mentioned, that in this study (Tykalová et al, 2014) authors used a specialised speaking task during which participants were asked to unnaturally emphasise five key words while reading a short block of text. In contrary, we applied the measurement of SPI to a commonly read passage where lower level of word stress is presented. In other words, the SPI parameter appears not to be easily applicable to all speaking tasks. Another issue is the difference in the database used. While Tykalová et al. (2014) investigated a group of untreated de-novo diagnosed PD patients, we surveyed PD patients with average duration of PD equal to 7.2 years and average doses of levodopa equal to 719 mg per day. Therefore, the discrepancies in results might also be partially attributed to the positive beneficial effect of levodopa treatment on speech production. Indeed, a previous report (Rusz et al., 2016) found a slight improvement of speech performances primarily associated with the improvement of stop consonant articulation and loudness variability in originally de-novo PD patients after 3–6 years of levodopa treatment.

Considering analyses of individual acoustic variables, F0 ratio was found to be higher in MS than in PD patients and HC in the stressed syllable set suggesting that MS speakers extended their F0 range to convey stress on syllables. On the other hand, there was not a difference in F0 ratio between MS and any other group in the unstressed syllables. In addition, MS speakers differed substantially in F0 ratio in stressed and unstressed syllables compared to HC. These findings suggest that F0 modulation in MS patients tended to lead to its excess variation. Despite the fact that excess pitch variation is not commonly listed among characteristics of ataxic dysarthria, our observation agrees with Duranovic et al. (2011) who found higher deviation in F0 in MS than in HC subjects, and partially with Feijó et al. (2004) who reported high F0 deviation only in men with MS. Nevertheless, since we converted F0 values to semitones in this study, a speaker's gender was not expected to significantly affect the results of F0 analyses. Low F0 ratio in PD patients compared to HC indicating monopitch was found only in case of unstressed syllables. However, this finding was not supported by analyses of other syllable sets.

Analyses of the intensity ratio revealed that MS patients extended their intensity range in both stressed and unstressed syllables significantly more than PD patients and HC speakers. Furthermore, we also found that the difference between intensity ratio of the stressed and unstressed syllables was larger compared to PD patients and HC speakers. These findings suggest that MS patients tended to vary in loudness excessively which is consistent with a feature of excessive loudness variation of ataxic dysarthria described by Duffy (2013). On the other hand, analyses did not show any significant differences between HC and PD patients. Thus, we did not observe signs of monoloudness in PD patients.

Syllable duration was found to be longest in MS patients. Both stressed and unstressed syllables were significantly longer in MS patients than in PD patients and HC subjects. Long syllable duration is closely related to slow articulation rate which is one the characteristic features of ataxic dysarthria according to Duffy (2013). Moreover, this finding is consistent with Kuo & Tjaden (2016) who reported significantly slower articulation rate in MS patients compared to PD patients and HC speakers. In addition, the analysis of the ratio of stressed and unstressed syllable duration indicated that MS patients pronounced all syllables with approximately same duration. As described by Duffy (2013) and Hartelius et al. (2000), equal syllable duration is a sign of scanning speech which is characteristic for ataxic dysarthria.

There are several limitations to the current study. Since previous studies (Hertrich & Ackermann, 1995; Skodda, Visser, & Schlegel, 2011) suggested that a speaker's gender may have an effect on prosodic aspects of an utterance and we did not study gender groups separately, we cannot assess the impact of gender on stress production. Furthermore, participants of this study were not completely age-matched. In particular, due to the lower age of the disease onset, MS patients were generally younger than PD patients. However, individuals of the HC group were age-matched to PD patients. Therefore, we could expect that their characteristics of voice were worse than those of younger healthy adults. This could finally lead to smaller differences between the MS and HC group.

5 Conclusion and Future Work

Although the methodology of stress evaluation designed in this study did not fulfil our expectations completely, the study made a contribution to our understanding of word stress production mainly in MS patients with predominant cerebellar damage. In spite of the fact that we detected deviations in stress production in MS patients using the SPI parameter, we were not able to do so in case of PD patients. Additionally, the SPI parameter appears not to be easily applicable to unspecialised speaking task such as passage reading, because of its high dependence on syllable duration. Thus, one can conclude that a modification of SPI parameter which is not so strongly dependent on syllable duration could be a solution for evaluation of stress patterns in passage reading.

References

Accent. (2017). *Encyclopædia Britannica*. [Online] Available from http://academic.eb.com.ezproxy.lib.cas.cz/levels/collegiate/article/accent/3474 [Accessed: 19th February 2017]

Atkinson, J. E. (1978). Correlation analysis of the physiological factors controlling fundamental voice frequency. *The Journal of the Acoustical Society of America*, *63*(1), 211-222.

Baker, K. K., Ramig, L. O., Sapir, S., Luschei, E. S., & Smith, M. E. (2001). Control of vocal loudness in young and old adults. *Journal of Speech, Language, and Hearing Research*, *44*(2), 297-305.

Boersma, P. P. G. (2002). Praat, a system for doing phonetics by computer. *Glot international*, *5*, 341–345.

Bolinger, D. L. (1961). Contrastive accent and contrastive stress. *Language*, *37*(1), 83-96.

Chaudhuri, K. R., Healy, D. G., & Schapira, A. H. (2006). Non-motor symptoms of Parkinson's disease: diagnosis and management. *The Lancet Neurology*, *5*(3), 235-245.

Cheang, H. S., & Pell, M. D. (2007). An acoustic investigation of Parkinsonian speech in linguistic and emotional contexts. *Journal of Neurolinguistics*, *20*(3), 221-241.

Compston, A., & Coles, A. (2008) Multiple Sclerosis. *The Lancet*, 372(9648), 1502-1517.

Cooper, W. E., & Danly, M. (1981). Segmental and temporal aspects of utterance-final lengthening. *Phonetica*, *38*(1-3), 106-115.

Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question–answer contexts. *The Journal of the Acoustical Society of America*, *77*(6), 2142-2156.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1975). *Motor speech disorders*. Philadelphia: Saunders.

Dauer, W., & Przedborski, S. (2003) Parkinson's disease: mechanisms and models. *Neuron* 39(6), 889-909.

De Lau, L. M. L., & Breteler, M. B. M. (2006) Epidemiology of Parkinson's disease. *The Lancet Neurology* 5(6), 525-535.

Diener, H. C., & Dichgans, J. (1992). Pathophysiology of cerebellar ataxia. *Movement Disorders*, *7*(2), 95-109.

Duffy, J. R. (2013) *Motor speech disorders: substrates, differential diagnosis, and management*, 3rd ed. St. Louis, Missouri: Mosby.

Duranovic, M., Salihovic, N., Ibrahimagic, A., & Toromanovic, N. (2011). Characteristics of voice in individuals with multiple sclerosis. *Materia Socio-Medica*, *23*(1), 23.

Feijó, A. V., Parente, M. A., Behlau, M., Haussen, S., De Veccino, M. C., & de Faria Martignago, B. C. (2004). Acoustic analysis of voice in multiple sclerosis patients. *Journal of Voice*, *18*(3), 341-347.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America*, *27*(4), 765-768.

Goetz, C. G., Fahn, S., Martinez-Martin, P., Poewe, W., Sampaio, C., Stebbins, G. T., Stern, M. B., Tilley et al. (2007). Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Process, format, and clinimetric testing plan. *Movement Disorders*, *22*(1), 41-47.

Goldenberg, M. M. (2012). Multiple sclerosis review. *Pharmacy and Therapeutics*, *37*(3), 175-184.

Gorham-Rowan, M. M., & Laures-Gore, J. (2006). Acoustic-perceptual correlates of voice quality in elderly men and women. *Journal of communication disorders*, *39*(3), 171-184.

Hafler, D. (2004). Multiple sclerosis. *Journal of Clinical Investigation*, 113(6), 788-794.

Halle, M. (1997). On Stress and Accent in Indo-European. *Language, 73*(2), 275-313.

Harel, B., Cannizzaro, M., Cohen, H., Reilly, N., & Snyder, P. (2004). Acoustic characteristics of Parkinsonian speech: a potential biomarker of early disease progression and treatment. *Journal of Neurolinguistics*, *17*(6), 439-453.

Harel, B., Cannizzaro, M., & Snyder, P. J. (2004). Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: A longitudinal case study. *Brain and cognition, 56*(1), 24-29.

Hartelius, L., Runmarker, B. O. R., Andersen, O., & Nord, L. (2000). Temporal speech characteristics of individuals with multiple sclerosis and ataxic dysarthria: ‘Scanning speech’ revisited. *Folia phoniatrica et logopaedica*, *52*(5), 228-238.

Hauser, S. L., & Oksenberg, J. R. (2006) The neurobiology of multiple sclerosis: genes, inflammation, and neurodegeneration. *Neuron*, 52(1), 61-76.

Hertrich, I., & Ackermann, H. (1995). Gender-specific vocal dysfunctions in Parkinson's disease: electroglottographic and acoustic analyses. *Annals of Otology, Rhinology & Laryngology*, *104*(3), 197-202.

Holmes, R. J., Oates, J. M., Phyland, D. J., & Hughes, A. J. (2000). Voice characteristics in the progression of Parkinson's disease. *International Journal of Language & Communication Disorders*, *35*(3), 407-418.

Hughes, A. J., Daniel, S. E., Kilford, L., & Lees, A. J. (1992). Accuracy of clinical diagnosis of idiopathic Parkinson’s disease – A clinicopathological study of 100 cases. *Journal of Neurology, Neurosurgery & Psychiatry*, 55(3), 181–184.

Huss, V. (1978). English word stress in the post-nuclear position. *Phonetica*, *35*(2), 86-105.

Ito, M. (1984) *The cerebellum and neural control*. New York: Raven Press.

Jacewicz, E., Fox, R. A., O'Neill, C., & Salmons, J. (2009). Articulation rate across dialect, age, and gender. *Language variation and change*, *21*(02), 233-256.

Jang, R. (2005). *5-4 Pitch*. [Online] Mirlab.org. Available at: http://mirlab.org/jang/books/audioSignalProcessing/basicFeaturePitch.asp?title=5-4%20Pitch%20(%AD%B5%B0%AA) [Accessed 6th Mar. 2017]

Jankovic, J. (2008). Parkinson’s disease: clinical features and diagnosis. *Journal of Neurology, Neurosurgery & Psychiatry*, *79*(4), 368–376.

Kang, M. (2017). Source-filter model of speech production. *Studyblue*. Retrieved 26 March 2017, from https://www.studyblue.com/#flashcard/view/14380451

Kent, R. D., & Read, C. (2002). *The acoustic analysis of speech*, 2nd. ed. San Diego: Singular/Thomson Learning.

King, J. B., Ramig, L. O., Lemke, J. H., & Horii, Y. (1993). Parkinson's disease: longitudinal changes in acoustic parameters of phonation. *N CVS Status and Progress Report*, 135–149.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, *59*(5), 1208-1221.

Kuo, C., & Tjaden, K. (2016). Acoustic variation during passage reading for speakers with dysarthria and healthy controls. *Journal of communication disorders*, *62*, 30–44.

Kuo, C., Tjaden, K., & Sussman, J. E. (2014). Acoustic and perceptual correlates of faster-than-habitual speech produced by speakers with Parkinson's disease and multiple sclerosis. *Journal of communication disorders*, *52*, 156-169.

Kurtzke, J. F. (1983). Rating neurologic impairment in multiple sclerosis an expanded disability status scale (EDSS). *Neurology*, *33*(11), 1444-1452.

Lang, A. E., & Lozano, A. M. (1998). Parkinson's disease. *New England Journal of Medicine*, *339*(15), 1044–1053.

Lees, A., Hardy, J. & Revesz, T. (2009). Parkinson's disease. *The Lancet*, 373(9680), 2055-2066.

Lehiste, I. (1976). Suprasegmental features of speech. *Contemporary issues in experimental phonetics*, 225–239.

Lieberman, P. (1967). Intonation, perception, and language. *MIT Research Monograph*.

Macphee, G. J., & Stewart, D. A. (2012). Parkinson's disease: treatment and non-motor features. *Reviews in Clinical Gerontology*, *22*(04), 243–260.

Milo, R., & Kahana, E. (2010). Multiple sclerosis: geoepidemiology, genetics and the environment. *Autoimmunity reviews*, *9*(5), A387-A394.

Morton, J., & Jassem, W. (1965). Acoustic correlates of stress. *Language and Speech*, *8*(3), 159-181.

Myers, B. R., & Finnegan, E. M. (2015). The effects of articulation on the perceived loudness of the projected voice. *Journal of Voice*, *29*(3), 390.e9-390e15.

National Multiple Sclerosis Society. (2017). *Medications*. [Online] Available from http://www.nationalmssociety.org/Treating-MS/Medications [Accessed: 30th March 2017]

Nishio, M., & Niimi, S. (2008). Changes in speaking fundamental frequency characteristics with aging. *Folia phoniatrica et logopaedica*, *60*(3), 120-127.

Parkinson, J. (1817). *An Essay on the Shaking Palsy*. London: Wittingham and Rowland

Patel, R., & Campellone, P. (2009). Acoustic and perceptual cues to contrastive stress in dysarthria. *Journal of Speech, Language, and Hearing Research*, *52*(1), 206-222.

Patel, R., Connaghan, K., Franco, D., Edsall, E., Forgit, D., Olsen, L., Ramage, L., Tyler, E., & Russell, S. (2013). “The Caterpillar”: A Novel Reading Passage for Assessment of Motor Speech Disorders. *American Journal of Speech-Language Pathology*, *22*(1), 1-9.

Phonetics. (2017). *Encyclopædia Britannica*. [Online] Available from http://academic.eb.com.ezproxy.lib.cas.cz/levels/collegiate/article/phonetics/108587 [Accessed: 26th March 2017]

Polman, C., Reingold, S., Banwell, B., Clanet, M., Cohen, J., Filippi, M., Fujihara, K., Havrdova, E., et al. (2011). Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria. *Annals of Neurology*, 69(2), 292-302.

Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., LaMantia, A. S., McNamara, J. O., & White, L. E. (2012). *Neuroscience*, 5th ed. Sunderland, Mass.: Sinauer Associates

Quené, H. (2008). Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *The Journal of the Acoustical Society of America*, *123*(2), 1104-1113.

Ramig, L. O., Gray, S., Baker, K., Corbin-Lewis, K., Buder, E., Luschei, E., et al. (2001). The aging voice: a review, treatment data and familial and genetic perspectives. *Folia phoniatrica et logopaedica*, *53*(5), 252-265.

Rampello, L., Rampello, L., Patti, F., & Zappia, M. (2016). When the word doesn't come out: A synthetic overview of dysarthria. *Journal of the Neurological Sciences*, *369*, 354-360.

Rusz, J., Cmejla, R., Ruzickova, H., & Ruzicka, E. (2011). Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson’s disease. *The journal of the Acoustical Society of America*, *129*(1), 350-367.

Rusz, J., Tykalová, T., Klempíř, J., Čmejla, R., & Růžička, E. (2016). Effects of dopaminergic replacement therapy on motor speech disorders in Parkinson’s disease: longitudinal follow-up study on previously untreated patients. *Journal of Neural Transmission*, *123*(4), 379-387.

Scherer, R. C. (1991). Physiology of phonation: A review of basic mechanics. *Phonosurgery: Assessment and surgical management of voice disorders*, 77-93.

Skodda, S., Visser, W., & Schlegel, U. (2011). Gender-related patterns of dysprosody in Parkinson disease and correlation between speech variables and motor symptoms. *Journal of Voice*, *25*(1), 76-82.

Sluijter, A. M., & Van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustical society of America*, *100*(4), 2471-2485.

Speech. (2017). *Encyclopædia Britannica*. [Online] Available from http://academic.eb.com.ezproxy.lib.cas.cz/levels/collegiate/article/speech/108588 [Accessed: 30th March 2017]

Torre III, P., & Barlow, J. A. (2009). Age-related changes in acoustic characteristics of adult speech. *Journal of communication disorders*, *42*(5), 324-333.

Tykalová, T., Rusz, J., Čmejla, R., Růžičková, H., & Růžička, E. (2014). Acoustic Investigation of Stress Patterns in Parkinson's Disease, Journal of Voice, 28(1), 129e1-129e8.

Ververidis, D., & Kotropoulos, C. (2006). Emotional speech recognition: Resources, features, and methods. *Speech communication*, *48*(9), 1162-1181.

Williams-Gray, C. and Worth, P. (2016). Parkinson's disease. *Medicine*, 44(9), 542-546.

Yuan, J., Liberman, M., & Cieri, C. (2006, September). *Towards an integrated understanding of speaking rate in conversation*. Paper presented at the International Conference on Spoken Language Processing (Interspeech 2006), Pittsburgh, PA.

Appendix A

Enclosed CD

Contents:

- Electronic copy of this thesis (Diploma thesis.pdf)
- File spi_main.m which is the main script
- Functions required to run the script (gettimelabels.m; intpreprocessing.m; cumsumInt.m; openF0.m; f0preprocessing.m; minmax.m; getF0ratio.m; getSPI.m)