More than words: Recognizing speech of people with Parkinson's disease

Verkhodanova, Vass

DOI:
10.33612/diss.183425053

CHAPTER 1
GENERAL INTRODUCTION

Parkinson's disease (PD) is now the fastest-growing neurological disorder in the world. The number of people diagnosed with PD is estimated to be nearly 13 million by 2040 (Dorsey et al., 2020). PD stems from a progressive loss of dopaminergic nerve cells in the substantia nigra (Kalia and Lang, 2015). While the cause of PD remains unknown, research points to the complex interplay of environmental and genetic factors (Kalia and Lang, 2015). As these factors need time to manifest, the likelihood that a person develops PD increases with age. However, it is not the ageing itself that contributes to the development of the disease, but rather living longer gives time for the nerve cell loss to occur (Levy, 2007; Dorsey et al., 2020). As the world's older adult population steadily grows as the result of increased life expectancy (Roser et al., 2019), so too does the number of people facing PD. The risk of developing PD roughly triples with each passing decade after one's forties (Dorsey et al., 2020; Van Den Eeden et al., 2003). Moreover, a recent global survey of neurological diseases revealed that even after correction for age-related factors, the number of people with PD (hereafter, PwPD) is projected to continue to rise (Feigin et al., 2019; Bloem et al., 2021). Studies also demonstrated that PD is associated both with a burden for caregivers who experience excessive strain (Macchi et al., 2019; Bloem et al., 2021), and with the mounting socioeconomic burden for society (Dorsey et al., 2018; Bloem et al., 2021).

Unlike other degenerative diseases, the duration of PD can span decades with a slow progression and accumulating disability for affected individuals (Bloem et al., 2021). The typical picture of PD includes reduced movement (hypokinesia), along with other symptoms such as muscle rigidity, bradykinesia, and mild cognitive impairment that can appear at the earliest disease stages (Gelb et al., 1999; Weil et al., 2018). The disease is commonly characterised by a combination of both motor manifestations (such as bradykinesia and tremor) and a range of non-motor manifestations (such as depression, constipation, and disturbed sleep) that can appear before the motor syndrome (Bloem et al., 2021). Among symptoms that can manifest in early and even prodromal stages of the disease progression is a specific speech disorder - *hypokinetic dysarthria* (HD).

Even though HD is commonly said to originate from the limited muscular control over mechanisms involved in speech production (Darkins et al., 1988), many researchers suggested that HD is related to non-dopaminergic deficits and is associated with non-motor symptoms (Rousseaux et al., 2004; Tykalová et al., 2015; Brabenec et al., 2017; Miller, 2017). People affected by HD experience a number of speech and voice deficits such as soft and monotonous voice or slurred articulation (Darley et al., 1969b) that cause

progressive loss of communication and lead to social isolation (Miller et al., 2006; Miller, 2017). Even James Parkinson noted in his substantive and historical text "An Essay on the Shaking Palsy" that in the course of the disease progression PwPD often became "scarcely intelligible" (Parkinson, 1817).

Today, research about dysarthria in PD has strongly developed, but a major part of research on speech of PwPD still targets intelligibility issues (Brabenec et al., 2017; Smith et al., 2019; Carvalho et al., 2020), as intelligibility is a relevant factor contributing to the clinical assessment of dysarthria. However, studies have demonstrated that PwPD experience speech changes and communication difficulties long before impairment of intelligibility becomes apparent (Miller et al., 2006). Miller (2017) claims that the "absence of obvious changes to the naked ear does not mean the person with PD and/or their family are not affected by communication changes"(p. 272). Yet, it is still rare for studies on speech of PwPD to target the "naked ear" of listeners and explore aspects of speech other than intelligibility. More researchers now investigate and advocate for more global ratings of overall speech competence or dysarthria severity (Kreiman et al., 1993; Weismer et al., 2001; Sussman and Tjaden, 2012), as well as address more abstract components of speech that may impede communication such as speech naturalness (Anand and Stepp, 2015; Klopfenstein, 2016).

Recently, studies on speech of PwPD have gained a new perspective. Technological progress in the last decades propelled research into developing better measures of PD, monitoring PD progression and searching for early markers of PD. Due to its non-invasive and objective measurements, acoustic analysis of dysarthric speech became a promising direction for such searches. Even though existing acoustic-data-driven approaches still lack specificity and sensitivity to serve as clinical diagnostic tools, there is evidence suggesting that automatic acoustic analysis can already be an asset for perceptual assessment of PD (Brabenec et al., 2017; Huh et al., 2015), or be a valuable help for evaluating disease progression and treatment efficacy in PD (Rusz et al., 2013a).

Besides the continuous search for a more holistic approach to the assessment of dysarthria as well as for acoustic markers and ways of detecting PD at the early stages, the topic of communication difficulties of PwPD remains under-researched. Furthermore, the increasing multilingual society and migration pose additional challenges for PwPD for whom the main language differs from the dominant language of their environment. For example, a recent study by Näsström and Schalling (2020) highlighted the growing need for assessment of acquired dysarthria in people speaking a language not familiar to a speech and language therapist (SLT). On top of that, the communication difficulties of PwPD (Miller, 2017) combined with a language barrier for older migrants (Pot et al., 2018b) present an alarming mixture of challenges for migrant PwPD. Therefore, the lack of cross-linguistic studies into changes in speech and communication of PwPD calls for attention (Pinto et al., 2017).

The main aim of this dissertation is to gain insights into listeners' impressions of dysarthric speech and to investigate the acoustic correlates of such impressions. This is done by exploring two sides of communication: acoustics of speech production of PwPD, and aspects of listeners' recognition[1] of speech of PwPD. In this way, the subjective listeners' impressions may be mapped onto the objective acoustic measurements of

[1] for terminological clarification see section 1.1

speech providing a perspective on speech changes in PwPD rooted in the context of communication.

1.1. TERMINOLOGICAL COMMENTS

This dissertation investigates speech of PwPD from both speakers' and listeners' perspectives. This section briefly outlines the terminological framework for the dissertation, because there is a lack of consensus in the literature regarding the use of such terms as *speech perception* or *speech recognition*. To avoid any ambiguity, this dissertation uses *speech recognition* throughout. Consequently, all experimental studies comprising the dissertation use the term *speech recognition*, operationalizing it as follows: "the recognition of spoken language involves the extraction of acoustic-phonetic information from the speech signal, and the mapping of this information onto cognitive representations" (McQueen and Cutler, 2010, p. 489). However, since it is conventional to contrast speech production with speech perception while defining perception similar to the earlier quoted definition of speech recognition, chapter 3 exploits the term *speech perception* for the systematic literature review on linguistic prosody recognition of speech of PwPD.

Other related terms used in the dissertation are *speech comprehension* and *speech intelligibility*. Speech comprehension, as a cognitive process, relates to speech recognition and refers to the ability of listeners to understand and interpret spoken language (Crystal, 2011). The second term, speech intelligibility, is operationalized as a measurable characteristic of speech representing the degree to which speech is comprehensible to a listener. The measure of speech intelligibility is commonly used in speech therapy to evaluate speech severity of dysarthria in PD (Sussman and Tjaden, 2012; Smith et al., 2019).

This dissertation also uses two terms with regards to knowledge and experience of listeners: training and expertise. In the literature, studies often do not distinguish between these two, using different descriptions to refer to trained and untrained populations (e.g., it is common to make a distinction between "expert" and "naïve" listeners). chapter 3 differentiates among three groups of listeners: untrained non-experts/trained non-experts and trained experts, while experimental chapters differentiate between trained and untrained listeners without introducing this three-way comparison.

Another terminological distinction needs to be introduced for the term *prosody*. This dissertation adopts a broad interpretation of prosody and uses the term to refer to the linguistic structure or syntagmatic relationships which determine various suprasegmental properties of syllables, words, and utterances (Cutler et al., 1997; Barth-Weingarten et al., 2010). Here, emotional prosody, which helps individuals communicate and discriminate between emotional states, is distinguished from linguistic prosody, which conveys linguistic structure (e.g. questions and statements) through markers like intonation and stress.

1.2. DISSERTATION FOCUS

The principal topic of interest in the dissertation stems from the under-researched aspects of communication difficulties experienced by PwPD. The studies discussed in the

preceding section show that communication troubles extend beyond mere intelligibility problems. The results of various studies that explore more global and abstract ratings for dysarthria in PD, as well as the relationship between PwPD's speech production and listener's dysarthric speech recognition may become an asset for the dysarthria treatments in PD, as it will allow communication problems that effect listeners' impressions to be targetted via specific speech characteristics.

This dissertation is concerned with Dutch speakers with PD in the Netherlands, where around 57,000 people are currently living with the diagnosis of PD, and the numbers are projected to rise (ParkinsonNet, https://www.parkinsonnet.nl/parkinson/). The studies in this dissertation approach the topic of speech changes in PD from both the speakers' side - via acoustic analysis of speech, and the listeners' side - via experiments exploring the influence of expertise and language background on recognition of speech of PwPD. Moreover, to obtain a more comprehensive picture of these perspectives, the studies of this dissertation are multifaceted, explore cross-linguistic aspects of dysarthric speech recognition and include both cross-sectional and longitudinal designs.

1.3. OUTLINE OF THE CHAPTERS

The dissertation is divided into four parts. The first three parts have their own set of research questions and can each be read as a separate study. The fourth part comprises a general discussion and the conclusions of the dissertation. The schematic overview of the four parts and experimental studies is presented in Table 1.1.

Part I of the dissertation aims to expand the understanding of different aspects of speech changes in HD. **Chapter 2** provides a general introduction to the acoustic changes caused by HD and reports on two pilot studies contributing to the evidence on acoustic changes in German and Dutch speakers with HD. One pilot study explores the effect of mild cognitive impairment on acoustic correlates of vowel articulation in German speakers with PD, the other investigates acoustic correlates of both vowel articulation and speech prosody in Dutch speakers with PD. **Chapter 3** systematically reviews the literature on linguistic prosody perception and recognition by different groups of listeners building upon the classification of dysarthria based on clusters of "deviant dimensions" which was introduced by Darley et al. (1969b) and laid the foundation for the contemporary dysarthria classification. **Chapter 4** elaborates on linguistic prosody in different types of dysarthria relying upon the evidence from the same classification by Darley et al. (1969b). It investigates the acoustics of prosody production in speakers with different types of dysarthria and explores how accurately a group of untrained listeners is able to recognize the intended linguistic prosody produced by the speakers.

Accordingly, the main focus of the first part of the dissertation is on the listeners' recognition of linguistic prosody produced by speakers with HD. This part embeds questions of prosodic deviancy caused by HD in the broader context of different types of dysarthria alongside the acoustic analysis of these deviances. The exploration of linguistic prosody in several types of dysarthria that involves both acoustic analysis and listeners' impressions allows to better understand the role of speech changes that are specific to dysarthria in PD. Moreover, the first part highlights the fact that the effect of listener's expertise and training on recognition of speech affected with HD is often overlooked

Table 1.1 | Overview of the dissertation parts and experimental studies. MCI - mild cognitive impairment, HC - neurologically healthy controls.

Part/ chapter		Design	Research focus
I	2	Cross-sectional group study. Speaker groups: MCI+PD, PD, HC. Listeners: untrained German group.	Vowel articulation imprecision.
	2	Cross-sectional group study. Speaker groups: PD, HC.	Vowel articulation imprecision and prosody.
	3	Systematic review.	Effect of listeners' training on perception of linguistic prosody in PD.
	4	Cross-sectional group study. Speaker groups: PD, SCA, MS, HC. Listeners: untrained Dutch group.	Production and recognition of linguistic prosody in dysarthria.
II	5	Cross-sectional group study. Speaker groups: PD, HC. Listeners: trained and untrained Dutch and non-Dutch groups.	Effect of expertise and language background on recognition of speech healthiness and prosody.
	6	Cross-sectional group study. Speaker groups: PD, HC. Listeners: trained and untrained Dutch groups, untrained non-Dutch group.	Acoustic correlates of listeners' responses in speech healthiness recognition task.
III	7	Longitudinal single PD case study, one year of dense measurements. Listeners: trained and untrained Dutch groups.	Reflections of longitudinal speech changes in the acoustic measurements and in listeners' perceptual assessments.
	8	Longitudinal single PD case study, 3.6 years of dense measurements in two working languages of one bilingual speaker with PD.	Longitudinal speech changes in two working languages of the bilingual speaker with PD and speech therapy effects.
IV	9		Discussion and conclusions.

in the literature, which provides additional rationale for the studies in this dissertation to explore this under-researched topic. The factor of listener's expertise and training may shape the experiences of listeners and affect their recognition and understanding of speech produced by PwPD. This is all the more reason to investigate whether speech changes appearing due to HD would trigger different recognition strategies in listeners with different expertise and training.

Part II deals with two listener-related factors that may influence speech recognition: listeners' expertise and, training and language background. As research on intelligibility has demonstrated, these factors alongside listener's biases (e.g., fatigue) and types of stimuli may affect speech recognition and comprehension (Kreiman et al., 1990; Hustad, 2006; Carvalho et al., 2020). The studies included in **chapter 5** investigate both listener-related factors of expertise and language background together with the input-related factor of task types (recognition of speech healthiness and of sentence type intonations). **Chapter 5** reports on two experimental studies with different groups of listeners and aims to provide more insight into how accuracy of dysarthric speech recognition depends on listener-related factors. One study described in **chapter 5** explores the influence of training in phonetics and speech sciences, while the other provides evidence about influence of training in speech and language therapy. Both studies also report on differences between Dutch and non-Dutch speaking listeners. **Chapter 5**, therefore, provides the foundation for the exploration of the acoustic correlates of listeners' responses to the recognition task in relation to the factors of expertise and training as well as language background.

Chapter 6 tests the possibility of predicting listener responses in a speech healthiness recognition task from the set of acoustic features. With the input from the previous chapter, we trained Random Forest models exploring the predictive power of acoustic and demographic feature sets. The chapter demonstrates that certain features in dysarthric speech are more or less representative of universal aspects of acoustic change in speech of PwPD, as they appear to be prominent for listeners independently of their first language or expertise. It also becomes clear that certain demographic features of speakers have varying effects on the recognition of listeners from different groups. Alongside the insights into the relationship between the acoustics of the dysarthric speech change and listeners-related factors, the presence of such variance in responses may also provide insights into the formation of the negative attitudes towards people whose speech is affected by HD (Miller et al., 2006; Miller, 2017).

Part III focuses on a longitudinal case study of one bilingual speaker with PD. Compared to many longitudinal studies on speech of PwPD with pretest-posttest designs (Skodda et al., 2009; Tykalová et al., 2015; Skodda et al., 2013), chapters in Part III provide a unique exploration of longitudinal monitoring of speech of the bilingual individual with PD. The single case studies described in (**chapters 7** and **8**) report on the results of dense observations over time. **Chapter 7** examines speech of the individual over one year of recordings, while **chapter 8** is concerned with the period of over 3.6 years of speech recordings of the same individual. The first one-year study explores the same aspects that are discussed in the first two parts: acoustics of speech change and listeners' responses in a speech recognition task. Through a set of acoustic measurements and results of experiments on speech healthiness recognition, **chapter 7** provides evidence of changes

in acoustics of speaker's phonation and prosody in Dutch. **Chapter 8** provides a closer acoustic investigation of speech changes appearing in both working languages of the participant, Dutch and English. From the findings reported in the chapter, a differential pattern of speech changes in both languages of the speaker becomes clear. Moreover, the evidence from both chapters demonstrates that studies with longitudinal designs and dense observations over time could be very valuable for speech therapists, as they can provide a detailed picture of speech disorder progression and help monitor speech therapy effects.

The last chapter of this dissertation, **chapter 9**, provides an overall discussion of the findings reported by the different studies. This chapter ties the parts together, uncovers common topics and generalizes the implications. Together, the three parts demonstrate the diversity of the ways in which the recognition of dysarthric speech may be affected. It adds to the existing evidence that multiple listener-related factors are of importance for successful communication, as it becomes clear that successful recognition of dysarthric speech is largely affected by listeners' backgrounds. As both factors of listeners' training and language background introduce additional variables into speech recognition processes, it brings out additional communication challenges for PwPD and calls for a reflection on how we approach the concept of expertise. In essence, considering both listeners' impressions and acoustics of speech changes in relation to HD increases our understanding of communication losses for PwPD. Moreover, translating subjective listeners' impressions into objective acoustic measurements may facilitate innovations in speech therapy and possibly contribute to early PD diagnosis.

I

REFLECTIONS OF PARKINSON'S DISEASE IN SPEECH

Not knowing the name of the tree,
I stood in the flood
of its sweet scent.

Matsuo Bashō