ACOUSTIC SPEECH MARKERS FOR TRACKING CHANGES IN HYPOKINETIC DYSARTHRIA ASSOCIATED WITH PARKINSON'S DISEASE

MRIDHULA MURALI

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

QUEEN MARGARET UNIVERSITY

2023

Abstract

Previous research has identified certain overarching features of hypokinetic dysarthria associated with Parkinson's Disease and found it manifests differently between individuals. Acoustic analysis has often been used to find correlates of perceptual features for differential diagnosis. However, acoustic parameters that are robust for differential diagnosis may not be sensitive to tracking speech changes. Previous longitudinal studies have had limited sample sizes or variable lengths between data collection. This study focused on using acoustic correlates of perceptual features to identify acoustic markers able to track speech changes in people with Parkinson's Disease (PwPD) over six months. The thesis presents how this study has addressed limitations of previous studies to make a novel contribution to current knowledge.

Speech data was collected from 63 PwPD and 47 control speakers using an online podcast software at two time points, six months apart (T1 and T2). Recordings of a standard reading passage, minimal pairs, sustained phonation, and spontaneous speech were collected. Perceptual severity ratings were given by two speech and language therapists for T1 and T2, and acoustic parameters of voice, articulation and prosody were investigated. Two analyses were conducted: a) to identify which acoustic parameters can track perceptual speech changes over time and b) to identify which acoustic parameters can track changes in speech intelligibility over time. An additional attempt was made to identify if these parameters showed group differences for differential diagnosis between PwPD and control speakers at T1 and T2.

Results showed that specific acoustic parameters in voice quality, articulation and prosody could differentiate between PwPD and controls, or detect speech changes between T1 and T2, but not both factors. However, specific acoustic parameters within articulation could detect significant group and speech change differences across T1 and T2. The thesis discusses these results, their implications, and the potential for future studies.

Keywords: acoustic speech markers; Parkinson's disease; hypokinetic dysarthria; tracking speech changes.

Acknowledgements

Completing a doctorate during the pandemic has been no small feat and would not have been possible without the combined support and effort of many over the past three years.

First, I would like to extend my sincere gratitude to Joan Ma for her unwavering support and guidance throughout my doctoral journey. Thank you for your constant trust in me through this process and for pushing me when I needed it the most. I would also like to thank my second supervisor Robin Lickley who has been instrumental in shaping my research and without whom I would not have succeeded. Thank you to Queen Margaret University for the generous bursary which allowed me to pursue this degree.

The CASL research centre has helped my development as an independent researcher, and the support of the staff has been central to my doctoral experience. Thank you to my friends from the speech lab – Maria, Susie, Erin, Anna, and Eleanor. I have taken comfort in being able to discuss my struggles or share our ideas with each other. Thank you to Steve Cowen for being my sounding board whenever I needed to vent or flesh out my thoughts. This group made completing a doctorate remotely enriching, despite the exceptional circumstances.

I want to thank Parkinson's UK for granting me access to their research network so I could recruit participants for my project. Thank you to my participants, whose enthusiasm and interest in the project was deeply appreciated.

To my friends in Scotland and abroad, I cannot express how much your presence and support has meant to me. To Annemiek, thank you for being my touchstone these past few years. Your friendship means the world to me. Lastly, to my parents, my brother, and my sister-in-law, I could not have asked for a more supportive family. Your faith in me in all my pursuits is astounding.

Table of Contents

List of Publications

Portions of the methodology and data collection procedure presented in the thesis have been previously published:

Murali, M., (2022). Using a podcast application to collect high-quality speech data online for acoustic analysis in people with parkinson’s disease. In *SAGE Research Methods: Doing Research Online*. SAGE Publications, Ltd., https://dx.doi.org/10.4135/9781529600575.

List of Tables

List of Figures

1. Introduction

Motor speech disorders can often present with a broad set of speech impairments that limit the ability to communicate with others effectively, resulting in socially challenging situations for the individual and their communication partners (Ansel & Kent, 1992). Diagnoses of motor speech disorders like dysarthria largely relies on perceptual evaluation by a speech and language therapist (SLT), which is a subjective assessment (Selouani, Dahmani, Amami & Hamam, 2012). While the subjectivity of perceptual assessments can call to question the reliability of this method (Kent, 1996), it is still an essential aspect of diagnosing speech and voice disorders as it provides value in understanding the impact of speech disorders on individuals' communication and intelligibility. The use of perceptual assessments to SLTs is also valuable for intervention (Kent, 1996; Oates, 2009).

There has been an increasing effort to understand the physiological and acoustic properties of dysarthria and their links to perceptual characteristics in order to support diagnosis and develop more effective intervention plans (Ackermann & Ziegler, 1991; Allison & Hustad, 2018; Bunton, Kent, Kent, & Rosenbek, 2000; Grosset et al., 2009; Liss, White, Mattys, Lansford, Lotto, Spitzer, & Caviness, 2009; Theodoros & Ramig, 2011). Perceptual assessments can be combined with acoustic analysis to quantify speech characteristics in dysarthria, providing acoustic correlates to perceptual features (Weismer, 2006).

Finding the same overarching acoustic parameters present in each case can be challenging because of the degree of variance in how dysarthria manifests between its different types and between individuals (Murdoch, 1998; 2009; Theodoros & Ramig, 2011). For example, one of the most prominent characteristics identified in hypokinetic dysarthria is the presence of articulatory undershoot during closure, which is the tendency to display less precise articulation during the production of plosives (Karlsson, Olofsson, Blomstedt, Linder, Nordh, & van Doorn, 2014). Although people with PD (PwPD) seem to display this more than controls (Ackermann & Ziegler, 1991;

Weismer, 1984), the variability in occurrence often lies within a range that is observed among controls as well, making it difficult to isolate as a marker unless observed consistently or in combination with other distinguishing features (Karlsson et al., 2014).

This challenge can make it difficult to find robust acoustic parameters useful for differential diagnosis of dysarthria. Researchers have attempted to identify these acoustic parameters with success in those robust for differential diagnosis between dysarthria and control speech and the various subtypes of dysarthria (Bunton et al., 2000; Forrest et al., 1989; Kent et al., 1999; Liss et al., 2009; Weismer, 1984). However, there have been several shortcomings in these studies, such as the methods employed or the limited size of datasets that still limit the applicability of these acoustic parameters. In addition, acoustic parameters that are useful for differential diagnosis do not provide information on disease progression. Therefore, acoustic analysis may have an added purpose in identifying acoustic parameters for tracking dysarthria progression (Harel, Cannizzaro, Cohen, Reilly & Snyder, 2004; Skodda, Rinsche & Schlegel, 2009; Skodda, Vissel & Schlegel, 2010; Grönheit, Mancinelli & Schlegel, 2013). These parameters can be clinically relevant to SLTs if acoustic values can capture perceptual changes that can be useful for intervention.

This thesis will focus on hypokinetic dysarthria, a motor speech disorder most commonly associated with Parkinson's disease (PD), and explore the acoustic speech markers unique to hypokinetic dysarthria and PD. The present study attempted to find acoustic markers to track perceptual changes in speech and overall intelligibility in hypokinetic dysarthria associated with PD over six months and differentiate PwPD and control speech. The following sections of this chapter will delve into dysarthria and its subtypes and provide background on PD and hypokinetic dysarthria associated with PD. The chapter will also introduce the connection between perceptual and acoustic analysis methods and highlight the thesis's central theme. This central theme will focus on finding acoustic parameters that can track speech changes in PwPD and how they may be common to those for differential diagnosis or different.

The rest of the structure of this thesis is as follows: chapter two will present the salient literature pertinent to this study, identifying gaps in the literature and justifying the basis of this study which will lead to chapter three presenting the research questions and the method employed to answer them. Chapters four to six will report and discuss the results of this study, and chapter seven will detail the interpretations of the results further in the general discussion, present the study's limitations, propose future investigations, and make conclusions.

1.1. Dysarthria

Dysarthria is the most common motor speech disorder and can affect the speech subsystems of respiration, phonation, articulation, prosody, and resonance. It results from a neurological disorder that disrupts muscular control and causes reduced speech intelligibility, timing and accuracy disturbances (Darley, Aronson & Brown, 1975). Dysarthria affects the control of muscles required for speech production, which causes changes to the speech signal.

According to Darley et al. (1975), six salient neuromuscular features are most influential on motor speech production: muscle strength, speed of movement, range of excursion, the accuracy of movement, motor steadiness, and tone. Abnormalities in any of these aspects can affect several systems of speech production. Based on their work in the Mayo Clinic Study, Darley et al.'s (1969a, b) dysarthria classification system is the norm in the literature. They suggested that the aetiology would allow perceptual assessments to help identify the disorder's unique perceptual features. One of the types of dysarthria, called hypokinetic dysarthria, is most associated with PD and is the focus of this thesis.

1.1.1. Prevalence of dysarthria

The overall incidence of dysarthria is challenging to quantify, but it is thought to be one of the most prevalent acquired communication disorders (Duffy, 2020). PD is cited as a clinical aetiology in which dysarthria occurs as a frequent and prominent symptom

(Yorkston, 1996). Dysarthria is commonly associated with ageing—and the ageing population means that the incidence of dysarthria is increasing. Within the United Kingdom, individuals aged over 65 years account for approximately 15.5% of the national population and form one of the fastest-growing population sectors (*State of Ageing*, 2022). The most common dysarthria aetiologies are stroke — which affects approximately 5-7% of people aged 65 and older (Feigin, Lawes, Bennett, & Anderson, 2003)— and PD — which is estimated to affect 1% of the population over 60 (Tysnes & Storstein, 2017). Hence, an ageing population contributes to the increasing prevalence of dysarthria.

1.1.2. Treatment of dysarthria

There is a wide range of treatment options available for speakers with dysarthria. Some impairments associated with the disorder can be treated medically (Duffy, 2020). However, while these options can significantly help speakers with dysarthria, their application remains limited (Fletcher, 2016). Pharmacological and surgical interventions usually cannot cure or completely halt the progression of dysarthria. Implants and assistive devices only address impairments within certain speech subsystems and only aid specific types of muscle impairment. For these reasons, there are many speakers with dysarthria for which none of the above treatment options is appropriate (Yorkston et al., 2001).

In contrast, behavioural intervention can be utilised by speakers with a range of dysarthrias—and is the primary focus of speech and language therapy. In its broadest sense, speech and language therapy aims to improve speakers' quality of life with dysarthria by enhancing their communication ability in everyday situations. Various approaches and strategies are used to try and achieve these improvements. For example, there is evidence that specific behavioural alterations, like changes to posture and breath control (Pennington, Smallman, & Farrier, 2006) or practising specific articulatory targets (Marchant, McAuliffe, & Huckabee, 2008; Robertson, 2001) may aid speakers in producing more intelligible speech. These exercises are regularly incorporated into ‘traditional dysarthria therapy’ programmes (Palmer, Enderby, & Hawley, 2007). Additionally, there has been consideration of how the listener and communicative environment might contribute to a person’s disability (Howe, 2008). As a result, strategies for the communication partner to help reduce communicative breakdowns (Yorkston, 1996) are also being utilised in speech and language therapy.

Now that we understand what dysarthria is and its various types, the following sections will focus on hypokinetic dysarthria and how it manifests in Parkinsonism.

1.2. Parkinsonism

Parkinsonism is a degenerative neurological syndrome resulting from damage to the extrapyramidal. Although it has been assumed that the various motor symptoms in PD were caused solely by dopamine depletion (Kalia, Brotchie, & Fox, 2013, 2013; Xia & Mao, 2012), many studies have shown that additional neural structures and neurochemical systems were also responsible for the occurrence of motor symptoms in PD; these include prefrontal cortical areas and the cerebellum, as well as serotonergic, glutamatergic, and cholinergic systems (Bohnen et al., 2013; S. H. Fox, 2013). Parkinsonism is a term used to refer to different types of PD, such as idiopathic PD, secondary (or symptomatic) PD and Parkinson-plus syndromes, of which idiopathic PD is the most common (Darley et al., 1975; Murdoch, 1998; Theodoros & Ramig, 2011).

The basal ganglia (a collection of subcortical grey matter structures) include the caudate nucleus, globus pallidus, and putamen. The caudate nucleus and the putamen together form the striatum. Pathways that initiate motor function run through the striatum and globus pallidus. This highly complex system of internal circuits modulates normal motor movement. The basal ganglia rely on neurotransmitters such as dopamine and acetylcholine to control muscular movement. Some dopamine receptors are inhibitory in specific pathways, and some are excitatory in others. A shortage of dopamine can therefore cause inhibition in pathways that are typically excitatory and vice versa. This leads to the typical symptomatology presented in Parkinsonism (Grosset, Fernandez, Grosset & Okun, 2009); however, treatment has shown that compensating dopamine level alone does not restore normal motor function (although it aids it), suggesting that other transmitters and pathways are also impacted to varying degrees.

Speech and swallowing disorders are typical consequences of PD. These are often accompanied by cognitive impairments such as memory deficits, reduction in visual-spatial function, and loss of executive function (i.e., difficulty in problem-solving or abstract reasoning). The most common communication problem individuals face with PD is hypokinetic dysarthria, which can occur in over 50% of PwPD (Grosset, Fernandez, Grosset & Okun, 2009). Hypokinetic dysarthria is predominantly the result of PD but can be secondary to other conditions. Hypokinetic dysarthria associated with PD is the focus of this thesis.

1.2.1. Motor Symptoms

PD is associated with primary (akinesia, bradykinesia, tremor, rigidity, and postural instability) and secondary motor symptoms (e.g., gait disturbance, grip impairment, and speech problems; Grosset et al., 2009; Jankovic & Tolosa, 2003; Lees, Hardy, & Revesz, 2009). Akinesia, which can be understood as a loss of moving a muscle voluntarily, difficulty initiating movements, and bradykinesia (slow movements) are considered among the primary motor features in PD (Moustafa et al., 2016). Rigidity is associated with a feeling of stiffness, and clinicians often assess it by the examinations of resistance of a muscle against a passive stretching (Moustafa et al., 2016). Tremor can be distinguished by delineating between the symptom's resting, postural, and kinetic manifestations. Resting tremor is the most common form of tremor in PD, while action/kinetic tremor (which is a tremor that occurs during voluntary movements) and postural tremor (inability to maintain stable posture; Toth, Rajput, & Rajput, 2004) are more common in essential tremor (Bhidayasiri, 2005).

A PD diagnosis requires a subset of motor symptoms (tremor, rigidity, akinesia, bradykinesia, and postural imbalance). However, these motor subtypes might have unique clinical profiles and outcomes for each individual (Moustafa et al., 2016). PwPD usually develops speech and voice disorders in PD (Ho, Iansek, Marigliani, Bradshaw, & Gates, 1999). Speech production was found to correlate with other global motor symptoms in PD, including akinesia (Skodda, Visser, et al., 2011b). It was also found that bradykinesia was related to speech disorders in PD (Robbins, Logemann, & Kirshner, 1986). Speech disturbances have also been shown to be more common in PD patients who had a high occurrence of freezing of gait (Park et al., 2014).

The most common outcome on which studies have focused was speech production. However, motor function also exerts a significant influence on speech function. One study investigated the sequencing of lip and jaw movements while speaking and showed a decreased coordination across these articulators in PD patients (Connor, Abbs, Cole, & Gracco, 1989). Another study found altered movements of lips and jaws in PD patients compared to the controls (Forrest, Weismer, & Turner, 1989). While investigating motor function is beyond the scope of the present study, future studies should investigate whether and how such processes impact speech production in PD.

1.3. Perceptual features and intelligibility in hypokinetic dysarthria

When hypokinetic dysarthria is sufficiently severe to reduce intelligibility, the processes of parsing the signal by the listener may be challenged in ways that require increased cognitive effort and can impact communicative success. Although hypokinetic dysarthria varies in presentation and severity across PwPD, there are speech features commonly exhibited (Darley et al., 1975; Theodoros & Ramig, 2011). Therefore, it has been proposed that the degradation of certain acoustic cues or constellations of cues will have a predictable impact on the intelligibility of PwPD speech (Lansford et al., 2011). For example, articulatory imprecision can lead to phonemic uncertainty, resulting in speech sound distortions or poor speech signal audibility from weak breathy phonation. This phonemic uncertainty can hinder the listener's ability to use lexically guided speech segmentation strategies (Lansford et al., 2011).

Speech rate in hypokinetic dysarthria often is judged to be different from controls (Ackermann, Konczak, & Hertrich, 1997; Caligiuri, 1989; Flint et al., 1992; Metter & Hanson, 1986). The range of movement in PwPD is restricted, despite preserved velocity of movement, and this, along with articulatory imprecision, is thought to give the impression of rapid, mumbled speech (Caligiuri, 1989; Yorkston et al., 1996). Findings suggest that some dysarthric speakers are more intelligible when they slow their rate of speech (Van Nuffelen et al., 2010; Yorkston, 1996), which usually results in rate control intervention strategies. This is evidenced by kinematic evidence that people with hypokinetic dysarthria have limited labial movement during their typical speaking rate. However, the movement is similar to controls when slower speech rate techniques are employed. (Caligiuri, 1989). Speaking slowly has been shown to improve phonemic distinctiveness (i.e., articulatory precision) in healthy control speakers (Mefferd & Green, 2010). In addition, Tjaden & Wilding (2004) found that speaking slowly expanded the vowel working space in people with dysarthria. Such phonemic distinctiveness has been demonstrated to be a predictor of intelligibility in the speech of controls (Neel, 2008) and people with dysarthria (McRae et al., 2002; Weismer et al., 2001). Thus, speech rate impacts articulatory precision, which impacts speech intelligibility in PwPD. Improved articulatory precision decreases phonemic uncertainty and improves speech intelligibility (Lansford et al., 2011).

Reduced loudness or hypophonia is another common manifestation of hypokinetic dysarthria and is a primary contributor to reduced intelligibility (Lansford et al., 2011). Reduced vital lung capacity, chest wall rigidity, and glottal incompetence are a few examples of the physiological presentations of respiratory and phonatory insufficiency, which is the presumed cause of hypophonia observed in the PD population (Solomon & Hixon, 1993; Tjaden, 2008). These physiological findings largely have been attributed to the overall muscle rigidity caused by PD (Darley et al., 1969b, 1969a; Duffy, 2020; Theodoros & Ramig, 2011). An impairment of internal cues in PwPD results in diminished speech movement initiation, amplitude, and timing (L. O. Ramig et al., 2008). Hypokinetic speech produced louder is perceived to have greater intelligibility than listeners of digitally amplified hypokinetic speech (Neel, 2009). This is likely due to the acoustic changes associated with producing louder speech, such as increased vocal intensity and improved use of pitch (Neel, 2009; Tjaden & Wilding, 2011), change in vowel formant values and ratios (Sapir et al., 2007), and alter articulatory displacements (Schulman, 1989). The production of loud speech improves the speech signal's overall audibility and syllabic stress cues (e.g., pitch and vowel production). Acoustic cues to syllabic stress in hypokinetic dysarthria are reduced (Liss et al., 1998). Thus, treatments that improve the contrast between stressed and unstressed syllables should promote lexical segmentation in hypokinetic dysarthria. The other acoustic/articulatory changes associated with loud speech, such as increases in vowel space area and articulatory displacements that approximate those of healthy control speakers, result in greater articulatory precision (Lansford et al., 2011). Thus, greater articulatory precision reduces phonemic uncertainty and improves speech intelligibility in PwPD speech.

1.4. Perceptual features and acoustic correlates in hypokinetic dysarthria.

Prosody is the most significantly affected speech dimension in hypokinetic dysarthria (Murdoch, 2009). Predominant features include monopitch, monoloudness, and reduced stress. Acoustically, a reduction in fundamental frequency (F_0) variation and reduced F_0 range have been cited as the most noticeable features in hypokinetic dysarthria speech by Harel et al. (2004). They conducted a case study on two individuals where F_0 variation was markedly lower than controls and showed a declining trend during initial diagnosis and disease progression. F_0 and F_0 range were proposed as potential markers for early disease progression in PD speech (Harel et al., 2004). Variable speech rate has also been noted in multiple studies, with some reporting increased speech rate, including short rushes of speech in segments (Darley et al., 1975) decreased speech rate, and some reporting normal speech rate compared to a control group (Grosset et al., 2009). It has been noted that listeners may also misperceive speech rate, claiming a faster rate but finding no significant deviation from normal upon acoustic examination. This may be due to articulatory errors and lack of pausing, making it harder for listeners to pick up on cues that indicate acoustic contrasts in the speech signal (Darley et al., 1975; Grosset et al., 2009; Murdoch, 2009).

Disordered phonation typically presents as a voice disturbance perceived as harsh or breathy with reduced loudness and a perceptible vocal tremor during prolonged vowels (Theodoros & Ramig, 2011). Reduced vocal loudness is considered one of the most debilitating features of hypokinetic dysarthria and the highest contributor to low intelligibility. Studies (Darley et al., 1975; Grosset et al., 2009; Murdoch, 2009) have shown that vocal loudness is consistently lower than control groups in various contexts but may sometimes be undetectable, based on the stage of progression of the disease, the type of task being performed during an assessment, and the individual's ability to compensate for the deficiency during assessment (Theodoros & Ramig, 2011; Murdoch, 2009).

Despite the variations in loudness found in the above studies, voice quality has been reported as consistently being impacted in PwPD. Dysphonia is presented as an expected outcome of PD speech, with breathiness, hoarseness, roughness, and vocal tremor among the most prominent features (Grosset et al., 2009; Theodoros & Ramig, 2011). Research has also shown that breathiness and hoarseness co-occur in PwPD (Baumgartner, Sapir & Ramig, 2001). Acoustic analysis has revealed high jitter and shimmer values in PwPD compared to controls related to irregular vocal fold vibration (Holmes et al., 2000; Kent, Vorperian, Kent & Duffy, 2003). Higher shimmer values are suggested as being correlated to breathiness in PwPD and likely related to the vocal fold bowing (Theodoros & Ramig, 2011).

Imprecise articulation of consonants is reported as another significant feature of PD speech (Darley et al., 1975), with stops such as /p/ being perceived as fricatives due to incomplete closure of the vocal tract, and fricatives such as /s/ or /z/ are reduced in sharpness due to poor constriction (Grosset et al., 2009; Murdoch, 2009; Theodoros & Ramig, 2011; Weismer, 1984). Studies investigating articulatory imprecision in hypokinetic dysarthria associated with PD suggest that speakers often do not reach articulatory targets and cannot maintain sufficient contact required to ensure precise articulation of a speech sound, referred to as articulatory undershoot (Theodoros & Ramig, 2011). Kent and Rosenbek (1982) found that articulatory undershoot was observed in the acoustic speech signal as a lack of distinction between sound and syllables. Other support for undershooting has been seen as reduced articulatory constriction during stop production (Ackermann & Ziegler, 1991), restricted vowel space (Weismer, Jeng, Laures, Kent & Kent, 2001), and reduced vowel articulation index (inverse of formant centralisation ratio; (Skodda, Visser & Schlegel, 2011)). In addition to evidence supporting articulatory undershooting in PD, articulatory imprecision has been attributed to continuous or inappropriate voicing of consonants, a prominent feature of the speech of people with PD (Theodoros & Ramig, 2011; Weismer, 1984).

Although evidence supports the claim that there is impairment to the velopharyngeal tract in PD (Grosset et al., 2009), there appears to be a disparity between instrumental and perceptual findings. Studies have shown variable results in the observation of hypernasality, with some showing mild occurrence, severe observation, and others claiming no presence of hypernasality (Darley et al., 1975; Freed, 2020; Grosset et al., 2009; Theodoros & Ramig, 2011). However, instrumental studies show consistent hypernasality (through the nasalance acoustic measure) in hypokinetic dysarthria associated with PD (Darley et al., 1975).

Studies investigating hypokinetic dysarthria have focused on perceptual assessment to find perceptual features that can describe the speech characteristics of the subtype (Chenery, Murdoch, & Ingram, 1988; Darley et al., 1969a, 1969b; Murdoch, 1998; Theodoros & Ramig, 2011). Other studies have tried to find acoustic parameters that correlate with perceptual features to quantify them (Bunton et al., 2000; Forrest et al., 1989; Kent et al., 1999; Liss et al., 2009; Weismer, 1984). The relationship between perceptual features and intelligibility was established in section 1.3, highlighting how individual perceptual features can impact overall speech intelligibility in PwPD. In addition, salient perceptual features and their acoustic correlates were also mentioned in 1.4 to establish further the use of instrumental measures to understand hypokinetic dysarthria.

1.5. The need for acoustic markers to track speech changes in hypokinetic dysarthria in PwPD

Listeners often describe people with dysarthria as having some combination of the following: imprecise articulation, slow speaking rate, voice disturbances, reduced prosodic variation (Mackenzie, 2011), or described as rough, effortful and mumbled (Liss et al., 2009). Adopting acoustic analysis can be a meaningful way of quantifying perceptual features and comparing units of measurement between different speakers. These acoustic parameters correlated with perceptual features can, in turn, provide an understanding of how acoustic parameters contribute to overall intelligibility in PwPD. However, it should be noted that no one acoustic parameter can detect the extent of impairment in dysarthria as effectively as a listener can (Liss et al., 2009, 2010; Sapir et al., 2010). In addition, research has shown that some processes, such as vocal fold spasticity or increased nasal emission, are still difficult to capture by any one acoustic parameter (Kent et al., 1999; Maryn et al., 2010). Using acoustic analysis to find correlates based on perceptual assessments, as in the present study, can help SLTs interpret the extent to which speech characteristics are impacted or change over time (Kent, 1996) and utilise the advantages of both perceptual assessments and acoustic analysis. This is not a perfect method and not all acoustic parameters can always be highly correlated to their perceptual counterparts. However, it does provide a closer link between the two and can help clinicians deduce their relationship.

Reduced speech intelligibility is another common consequence of dysarthria, and Ansel and Kent (1992) described this reduction in intelligibility as “the most clinically and socially important aspect of dysarthria” (p. 297). A study on individuals with PD resulting in dysarthria showed that reduced speech intelligibility impacted daily living even in mild cases (Miller, Noble, Jones & Burn, 2006). Participants in Miller et al.’s (2006) study expressed that some of the problems they encountered were making themselves understood in a normal conversation and dealing with the reactions of others. Finding acoustic correlates based on perceptual assessments can also be used to provide an understanding of why PwPD has poor intelligibility. Speech intelligibility is often rated during perceptual assessments to help understand the overall impact of speech impairment on an individual's intelligibility (Yorkston, 1996).

The prosodic features of hypokinetic speech (e.g., accelerated and variable speaking rate, short rushes of speech, dysfluency, monopitch and monoloudness) may result in reduced cues to syllabic stress. Since syllabic stress cues become important for identifying word boundaries, particularly when the acoustic-phonetic information is degraded, it can lead to the listener's perception of reduced speech intelligibility. How listeners parse the speech signal of hypokinetic dysarthric speech in PD has been the focus of a series of studies (Liss et al., 1998, 2000, 2002), which found that listeners were generally able to use the available acoustic cues in moderate to severe hypokinetic dysarthria to identify word boundaries. The listener error patterns revealed a significant tendency to treat strong syllables as word onsets. However, the tendency was less consistent than for normal speech presented at low listening levels. This supports the interpretation that part of the intelligibility reduction in hypokinetic dysarthria is linked to the reduced acoustic-perceptual contrast between strong and weak syllables (Lansford et al., 2011; Liss et al., 2002).

Since speech results from a complex process, intelligibility measures include aspects of all speech dimensions (De Bodt et al., 2002). In a group of head-injured patients with overall intelligibility problems, patients showed more than 90% of the dimensions related to resonance, articulation, prosody and voice (De Bodt et al., 2002). Therefore, it can be beneficial to understand how some perceptual features contribute to speech intelligibility in hypokinetic dysarthria associated with PD.

Speech impairment in individual perceptual features may not have a severe overall impact on speech intelligibility but still indicate an individual may have dysarthria (Kent, Weismer, Kent, & Rosenbek, 1989). Individuals can still display speech impairment in various speech subsystems but remain intelligible. Understanding the degree to which individual perceptual features contribute to overall intelligibility is crucial to investigate. Using acoustic analysis to investigate individual perceptual features and overall speech intelligibility can help delineate the relationship between these perceptual areas and ascertain whether acoustic parameters can track speech changes in PwPD. If acoustic parameters can track speech changes in individual perceptual features and overall intelligibility, then acoustic parameters can provide a greater understanding of speech in PwPD and help quantify how PD speech may change naturally in the long term. Investigating the natural change in PwPD speech over time can be used as a comparison against PwPD speech undergoing speech and language therapy.

Some findings from various studies have helped identify perceptual and acoustic characteristics of hypokinetic dysarthria (Ackermann & Ziegler, 1991; Allison & Hustad, 2018; Bunton et al., 2000; Grosset et al., 2009; Liss et al., 2009; Theodoros & Ramig, 2011). These allowed researchers and SLTs to understand how dysarthria may impact individuals' overall intelligibility and how they can plan for interventions to help improve intelligibility. For example, reduced loudness is a common characteristic of hypokinetic dysarthria (Theodoros & Ramig, 2011). The purposeful production of loud speech improves the speech signal's overall audibility and syllabic stress cues (e.g., pitch and vowel production). Acoustic cues to syllabic stress in hypokinetic dysarthria are reduced (Liss et al., 1998). Thus, treatments that improve the contrast between stressed and unstressed syllables should promote speech intelligibility in PwPD. However, the individual variability in the presentation of perceptual features makes isolating the acoustic markers of hypokinetic dysarthria challenging (Harel et al., 2004).

In addition, the acoustic parameters identified are primarily valuable for differential diagnosis between PwPD and controls and do not imply that they can track speech progression. As perceptual features in PwPD may not present with the same level of severity at each stage of PD and therefore change and influence speech intelligibility as PD progresses (Skodda et al., 2013), acoustic parameters that can differentiate between PD and control speech may not be sensitive to speech changes over time and requires further investigation. Tracking speech change in PwPD is valuable as it provides a better understanding of how changes in speech characteristics may impact speech intelligibility over time. This influences the management strategies SLTs may employ to improve speech intelligibility in PwPD over time. Identifying acoustic parameters that are effective in tracking speech changes in PwPD can be used to understand individual variations that may exist. If there are acoustic parameters that are reliable for both differential diagnosis and tracking speech change, long-term research can be advantageous to test whether the acoustic parameters are effective for diagnosis and attempt to create a database for tracking speech changes over time. Understanding PD speech change can provide further insight into aiding SLTs to provide PwPD with a possible prognosis. The present study used acoustic correlates of perceptual features to investigate PD speech. The thesis took a first step toward distinguishing between acoustic markers for tracking speech changes in PwPD and differential diagnosis between PwPD and controls.

The following chapter will delve into previous literature, highlighting the relationship between speech intelligibility ratings, perceptual assessments, and acoustic analysis. In addition, previous literature pertaining to finding acoustic parameters for differential diagnosis and tracking speech changes in dysarthria will be explored, and their gaps identified. These will lead to the aims of the present study. The acoustic markers identified in the study may be used to supplement existing methods of hypokinetic dysarthria evaluation and pave the way for future research that could implement these markers in an automatic detection system. This would further ease the role of the SLT during assessment and give patients access to a more quantifiable way of assessing how disease progression impacts their speech.

2. Perceptual and Acoustic Analysis in Dysarthria

This chapter will further explore hypokinetic dysarthria associated with PD and present relevant literature highlighting prominent aspects of speech in PD, identifying gaps in the literature and showing how this study shall address these gaps and contribute uniquely to PD research.

2.1. Speech intelligibility in dysarthric speech

According to Yorkston et al. (1996), intelligibility can be defined as what listeners understand of the phonetic realisation of speech. Intelligibility rating can indicate dysarthria's impact on a speaker's communicative performance. It is the result of how the five speech subsystems interact (De Bodt, Hernández-Díaz Huici, & Van De Heyning, 2002). Investigating speech intelligibility in dysarthria, especially in connected speech, has ecological validity as it is a closer estimation of the functional communication level of a speaker (Weismer et al., 2001).

Intelligibility impairment can negatively impact people's everyday activities and is considered one of the most prominent features of PD speech, resulting in a negative quality of life (Ansel & Kent, 1992). Since speech production can be highly individual and people with dysarthria can sound quite different from one another, it is essential to acknowledge that the sources of these differences are due to their aetiologies and the individual differences in the features of their speech (Duffy, 2020). Therefore, even if two individuals with dysarthria have the exact aetiology, their speech features may be present with unique variations in characteristics such as voicing, articulation of vowels and consonants, and in speech rate (Kim, Kent, & Weismer, 2011), which will all impact how intelligible they may sound to a listener.

However, even when there can be a significant influence of speech impairment in the daily communication of people with dysarthria, the extent to which it may impact their participation in daily activities is variable. This variability is due to several precursors, such as individual personalities, social and occupational demands, and individual coping mechanisms (Dykstra, Hakel, & Adams, 2007).

Understanding and assessing speech intelligibility in PwPD serves several purposes. It can provide a method of quantifying how speech production deficits might influence the listener's perception of PD speech. Measuring intelligibility can also be used to quantify the change in speech production resulting from intervention, disease progression, the impact of medication, or any recovery. Intelligibility is often also used to judge the severity of speech impairment in PD, which provides a reasonable estimate of the deterioration in speech production. Intelligibility assessment can further provide information about different speakers and can be used to compare speech performance (Stipancic & Kris Tjaden, 2022; van Brenk et al., 2022). However, a general measure of speech intelligibility is insufficient in targeting specific speech deficits contributing to lower intelligibility. Therefore, specific speech features that impact listeners' perceptions of intelligibility need to be identified.

Identifying speech properties that verify speech intelligibility could be a methodologically advanced task. However, it has utility for implementing effective remedial ways of improving the intelligibility of individuals with dysarthria. Within the attempt to create effective remediation for dysarthric speech, it is crucial to see that certain aspects of the speech signal will indicate the changes that need to be made to make significant gains in intelligibility (Ansel & Kent, 1992; Fletcher, 2016). For example, lower than control speech intensity can indicate less loudness or monoloudness, a common feature of hypokinetic dysarthria. Quantifying and observing the loudness level in the speech signal can indicate that exercises to improve loudness can positively impact speech intelligibility (Lansford et al., 2011). Two people with precisely the same score or rating on an intelligibility test could have completely different errors underlying that score (Kent, Weismer, Kent, & Rosenbek, 1989). Phonetic and acoustic analyses facilitate the attempt to identify the variations in speech errors. The ability to elucidate intelligibility deficits in terms of specific acoustic correlates has clear clinical implications as a diagnostic tool to document deficits, develop economic rehabilitation programs, focus treatment, and substantially enhance speech proficiency (Ansel & Kent, 1992).

Previous studies have focused primarily on two objectives: a) in identifying the properties underlying speech intelligibility deficits in dysarthria within the speech signal and b) in identifying acoustic characteristics that pertain to specific types of dysarthria. The studies that have identified acoustic parameters that can predict speech intelligibility in people with dysarthria have identified the following parameters: voice onset time (VOT; Liu, Tseng, & Tsao, 2000), acoustic vowel space (Liu et al., 2000; McRae, Tjaden, & Schoonings, 2002; Weismer, Jeng, Laures, Kent, & Kent, 2001), and second formant frequency slope (F2) (Kent et al., 1989; Kim, Weismer, Kent, & Duffy, 2009; Mulligan, Carpenter, Riddel, Delaney, Badger, Kursinski, & Tandan, 1994; Weismer, Martin, Kent, & Kent, 1992). Prominent acoustic parameters identified in hypokinetic dysarthria, specifically those that can predict speech intelligibility scores, have included either normal or faster-than-normal speaking rates, high mean fundamental frequency (F0), decreased F2 extents and slopes, and decreased F0 variability (Canter, 1963; Forrest et al., 1989; Goberman, Coelho, & Robb, 2005; Solomon & Hixon, 1993; Weismer, 1984). This shows that the objective has been to find acoustic parameters that predict speech intelligibility scores or to investigate the acoustic correlates of perceptual features in certain dysarthrias. However, there has yet to be a focus on investigating tracking changes in speech intelligibility using acoustic parameters. While it is assumed that greater severity of dysarthria leads to lower speech intelligibility (Fletcher, 2016; Kim et al., 2011), it is unclear whether speech intelligibility changes linearly as severity increases.

Studies that have tried to find acoustic parameters that can predict speech intelligibility have been limited but note that it is an essential part of understanding how changes in speech production can be aided by finding parameters that correlate with changes in speech intelligibility (Ansel & Kent, 1992; Fletcher, 2016; Kim et al., 2011). Kim et al. (2011) examined acoustic predictors of speech intelligibility in speakers with several types of dysarthria secondary to different diseases. They conducted classification analysis using acoustic measures according to disease, speech severity, and dysarthria type. Ansel and Kent (1992) examined the relationship between word intelligibility and acoustic measures. Fletcher (2016) investigated acoustic measures that could predict intelligibility gains in dysarthria resulting from intervention. These studies note that such research is fundamental since the Mayo clinic studies did not control for severity. Instead, severity was allowed to vary within each disease group investigated (Darley et al., 1975; Kim et al., 2011). Therefore, the impact of different levels of severity on speech intelligibility and their acoustic correlates remains unknown within the Mayo clinic studies.

However, understanding acoustic correlates of speech intelligibility relies on the accurate perceptual assessment of dysarthric speech, and there are some limitations of perceptual assessment that may prevent finding reliable correlates. It has been suggested that perceptual labels such as “imprecise consonants” and “fast rate” may not be related in a straightforward way to aspects of the speech signal (Ansel & Kent, 1992; Kent, 1996). The perceptual fast rate in hypokinetic dysarthria associated with PD may be related to spirantisation and compressed vowel space observed in the speech signal and not just related to the speaking rate of a PwPD speaker (Kent & Rosenbek, 1982).

Acoustic analysis can be a valuable complement to perceptual assessments, as illustrated in previous research that has provided various measures, including VOT, vowel formant frequencies and vowel and consonant durations. For example, research in hypokinetic dysarthria resulting from PD and ataxic dysarthria (Kent & Netsell, 1975; Kent, Netsell, & Abbs, 1979; Kent & Rosenbek, 1982) reported results of speech timing discrepancies and segmental abnormalities citing physiological explanations for the perceptual characteristics usually associated with those dysarthrias. An alternate explanation could be that the perceptual feature “short rushes of speech” generally associated with hypokinetic dysarthria may be the result of a reduction in the range of articulatory movements (articulatory undershoot) rather than the result of an increase in the rate of articulation (Kent & Rosenbek, 1982).

Despite the advantages, acoustic analysis provides to interpreting perceptual features of dysarthric speech, acoustic analysis alone may not be able to assess aspects of speech production that relate to an individual's ability to communicate effectively. Therefore, both perceptual and acoustic analyses are inextricable in understanding speech intelligibility and other aspects of dysarthric speech. The section has outlined some acoustic parameters that correlate with the perceptual assessment of speech intelligibility in dysarthria. However, it is imperative to delve into some of the limits of both acoustic and perceptual methods of assessment that may influence investigating PwPD speech effectively. This is detailed in the following section.

2.2. Perceptual versus acoustic analysis

Studies that evaluate speech and voice in PwPD adopt perceptual or instrumental assessments, depending on the focus of the evaluation. In perceptual assessments, experienced listeners use some protocol or set of agreed criteria to make judgments on the speakers' quality of speech and voice. Instrumental assessments use the speech signal to employ algorithms and signal processing techniques, often through acoustic analysis, to evaluate the speech characteristics of a speaker. From the studies found in literature performing an objective assessment of the speech of PwPD, some propose measurements or features that can be used for diagnosis or that correlate with the severity of the disease (Ackermann & Ziegler, 1991; Allison & Hustad, 2018; Bunton et al., 2000; Grosset et al., 2009; Liss et al., 2009; Theodoros & Ramig, 2011). Others propose automatic detectors to differentially diagnose between speakers with and without PD or to predict the disease stage automatically (Moro-Velazquez & Dehak, 2020; Novotný et al., 2014). Phonatory studies usually employ sustained vowels to measure acoustic parameters such as noise, frequency, amplitude perturbation, fundamental frequency, or formants frequency that can also be measured on running speech under certain circumstances. The studies framed within the prosodic and articulatory aspects are based on the feature extraction from connected speech or the processing of specific segments of the speech (Moro-Velazquez & Dehak, 2020).

There are limitations to the reliability and validity of using instrumental speech analysis methods, and there needs to be more agreement on which measures are the most sensitive. However, perceptual assessments have equally been criticised for their subjectivity (Oates, 2009).

Voice is a perceptual response to an acoustic stimulus (Eadie & Baylor, 2006; Shrivastav, 2003; Shrivastav, Sapienza, & Nandur, 2005) and therefore using perceptual assessment seems the logical approach to evaluate voice. Since voice is perceptual in nature, listeners tend to have a shared understanding of various perceptual features that pertain to its evaluation which can result in a number of these perceptual features being intuitive to listeners to discern (Wuyts, Bodt, & Heyning, 1999). For example, it is more intuitive to describe the voice quality of a speaker as breathy and rough and would be largely understood in meaning by a listener rather than providing the harmonic-to-noise ratio of that speaker. However, while these perceptual features may seem intuitive, they may not be as simple to delineate, and listeners cannot rely on a shared intuition of what each perceptual feature refers to or means to make perceptual assessments (Oates, 2009). Even though perceptual feature definitions may be inconsistent, reducing the reliability and validity of perceptual assessments, the shared intuitive meaning behind the perceptual features among listeners may contribute to the popularity of using perceptual assessments of disordered speech.

While it may be challenging to demonstrate how much clinicians rely on perceptual assessments, it has been reported that there are surveys which have indicated that perceptual methods are employed frequently and valued highly (Kent, 1996). In order for perceptual judgements to be valuable clinically, the need to fulfil specific criteria: there must exist a common understanding of the definitions of perceptual features such as hoarse, breathy, monoloudness, or rough; there needs to be an agreement on how these features are evaluated, what scale they are judged on, and what a rating implies; there is potential to isolate one perceptual dimension for other co-occurring dimensions; there are differences between several listener’s judgement are minor which allows clinically relevant changes in speech to be meaningful. This is a challenging undertaking as listeners often may not have the same definitions for perceptual features even when using a rating system that attempts to delineate each perceptual feature within speech subsystems, and experienced listeners seem to disagree on which perceptual features need to be rated for a particular disorder (Kent, 1996; Oates, 2009).

Since perceptual ratings of various dimensions are demonstrated to be intercorrelated (Kent, 1996; Oates, 2009; Sheard, Adams, & Davis, 1991), the values that may be extracted for any one speech dimension will likely be influenced by other dimensions within a speech disorder. In addition, various perceptual dimensions may not be rated with the same reliability, and differences between listeners may impact this reliability. For example, a disadvantage to perceptual assessments in voice quality is that listeners may disagree within themselves on which aspects of voice quality are most relevant in rating normative and disordered speech. In fact, Kreiman, Gerratt, and Precoda (1990) reported that clinical training resulted in more significant discrepancies among listeners in the judgement of voice quality, which is counterintuitive to the commonly held assumption that more training results in better inter-rater agreement.

Zeplin and Kent (1996) (as cited by Kent, 1996) replicated the studies of Darley et al. (1969a, b) of the ratings conducted using the recorded materials used and resulted in only partial agreement between the most deviant perceptual features found in Zeplin and Kent's study and Darley et al.'s study. Some perceptual features rated as most deviant across the different types of dysarthria in Zeplin and Kent's study had higher inter-rater reliability. These perceptual features were imprecise consonants, loudness, pitch level, and fast rate. However, monopitch and monoloudness had large standard deviations in their ratings for some types of dysarthria, which indicates that these perceptual features may be difficult to rate with a high agreement between listeners. Therefore, from the work of Zeplin and Kent, it can be concluded that the perceptual ratings of some features may be more reliable than others.

Sheard, Adams, and Davis (1991) examined the performance of 15 SLTs in rating the speech of 15 individuals presenting with ataxic dysarthria on five perceptual features across different dimensions: imprecise consonants, excess and equal stress, irregular articulatory breakdown, distorted vowels, and harsh voice. They concluded that due to high correlations between the perceptual features that were rated, listeners must consider that judgements made on individual perceptual features may be an overall judgment made on a cluster of salient speech features instead.

A study investigating acoustic and perceptual underpinnings of the speech characteristic monotony in PwPD speakers found that listeners' ratings could not distinguish between monopitch, monoloudness, and monoduration (Kim, 1994). Furthermore, it was reported that the perceptual features monopitch and monoloudness were strongly correlated and therefore posed an important question about how listeners could reliably judge these two features independently.

Further studies investigating perceptual judgement of voice quality show that some perceptual features, such as breathiness and roughness, display high reliability in judgement (De Bodt, Wuyts, Van de Hyning, & Croux, 1997; Dejonckere, Obbens, Moor, & Wieneke, 1993; Hammarberg, Fritzell, Gauffin, Sundberg, &Wedin, 1980; Webb, Carding, Deary, Mackenzie, Steen, & Wilson, 2004), other perceptual features such as vocal strain have poor listener reliability and have very little control over the elements that influence the perceptual judgement of low-reliability perceptual features. In addition, there is a reduced agreement and reliability between listeners when asked to make judgments by isolating specific perceptual features in speech that present with a complex cluster of perceptual characteristics, especially isolating perceptual features in mild or moderate speech impairment (Kreiman & Gerratt, 2000).

Another concern is using complex systems to derive detailed information from perceptual judgments, which may not be beneficial to accurate judgements. The complexity typically involves increasing the number of response categories or making fine distinctions within individual response categories within the perceptual assessment system. The goal of such added complexity is to increase information from perceptual assessments, but the added information often increases the unreliability of the judgements (Kent, 1996). Adding too much detail and complexity into a rating system does not guarantee that the listener's auditory-perceptual decisions will fulfil the demands of the analysis task.

Despite the various limitations of perceptual assessments and listener judgements presented above, it is undeniable that experienced listeners can make distinctions between different dysarthrias with some reliability. Therefore, there is relevant perceptual information that is correlated with relevant acoustic signal properties that can provide valuable information about dysarthric speech. Some evidence shows that speakers with PD (who present with hypokinetic dysarthria) are distinguished with greater reliability than other aetiologies of dysarthria (Zyski & Weisiger, 1987). The study suggests that listeners may cue into certain perceptual features more than others to distinguish between different types of dysarthria rather than focusing on the entire acoustic speech signal. When the most salient perceptual features are not present or are subdued, then only co-occurring perceptual features in the various dysarthria types are available to listeners, which results in less reliable judgements.

One way of overcoming the limitations of the perceptual analysis of speech or voice is to supplement perceptual ratings with instrumental (acoustic) analyses. Despite the extensive research done over the years on finding instrumental measures in disorders of speech, an ideal set is yet not agreed upon (Hillman, Montgomery, & Zeitzels, 1997; Ma & Yiu, 2006). Some parameters have shown some promise (see section 2.3 below for details), but much has yet to be researched on the sensitivity of these parameters. A high degree of inter- and intra-individual variation makes it difficult to find normative values. Furthermore, it can be challenging to control for multiple confounding variables that may impact evaluation. These variables include the recording environment conditions, the types of hardware and software systems used, protocols and analyses adopted, and the speech's severity level. In addition, judgements are usually made on voice samples, such as sustained phonation and reading passages. It is unclear whether judgements made from these samples can be generalised to spontaneous speech (Oates, 2009). It is also unclear whether inconsistencies in finding acoustic correlates in perceptual features are due to methodology issues in investigating the acoustic-perceptual relationship or inherent limitations in both perceptual assessments and acoustic analysis.

In a seminal theoretical paper, Weismer (2006) outlines the shortcomings of using only oromotor, nonverbal tasks for dysarthric diagnosis rather than focusing on speech production (which employs acoustic and perceptual methods). The author discusses the prominence of the Mayo Clinic Study (Darley et al., 1969a, b), which resulted in a surge of research that focused on oromotor, nonverbal characteristics to diagnose neurological diseases such as dysarthria and, in turn, reduced the prominence of acoustic and perceptual investigations.

Speech production studies that began to emerge after that used Darley et al.'s (1969a, b) system of differential diagnosis, which held perceptual analysis as the primary method of investigating speech in dysarthria, often with studies aiming to find disease-specific markers (Weismer, 2006). This supported the Mayo Clinic Study, which advocated that dysarthria results from neuropathological symptoms that would filter down to speech and interrupt speech naturalness.

Perceptual analysis methods are easy to access and low cost, but they can only be accurate if the clinician is experienced, or the method of judgement is appropriate. In addition, it is challenging to standardise results because assessments are patient and environment dependent (Selouani et al., 2012). Acoustic analysis would balance this potential pitfall as it would provide a more quantifiable way of looking at speech production by isolating specific patterns of speech that can be seen in the speech signal. Weismer (2006) supports finding specific features of MSDs but contends that the most efficacious way to achieve this is through acoustic and perceptual analysis, which rests on investigating the speech signal itself.

Therefore, using acoustic methods to supplement already present methods of perceptual analysis used by SLTs would create a more robust assessment method for dysarthria.

Since perceptual evaluation methods have always been popular but bear the risk of being unreliable, there has been an interest in focusing on acoustic analysis to conduct a differential diagnosis of dysarthria, including hypokinetic dysarthria in PD. This often involved looking at spectrograms and other aspects of the speech signal to find anything out of the norm (Weismer, 1984). This allowed researchers to conclude with some certainty that perceptual features such as monoloudness and monopitch are consistently present in hypokinetic dysarthria because they can be isolated in the speech signal acoustically through vocal intensity, F_0 variability, F_0 SD, Mean F_0, and F_0 range (Galaz et al., 2016).

It is essential to note the limitations of perceptual and acoustic analysis to maximise the insight they can provide in PD speech research and how they might be combined in the present study to assess speech changes in PwPD speech. The following section outlines vital literature on acoustic parameters used to differentially diagnose dysarthric speech and how these studies may inform the present study's objectives.

2.3. Acoustic analysis for differential diagnosis and distinguishing between controls

While dysarthria is generally considered a movement disorder, it rarely occurs in isolation. Associated neurological etiologies frequently cause co-occurring impairments in language and cognitive skills. As a result, people with dysarthria often struggle to manage their attention, memory and mood and sometimes lack complete self-awareness of how their speech may impact their everyday lives (Duffy, 2013; Fox, Morrison, Ramig, & Sapir, 2002). Since it is generally assumed that specific speech characteristics are homogeneous for each type of dysarthria, there can often be a need for more investigation into a baseline of speech patterns using perceptual or acoustic evaluations.

Although the classification system developed from the Mayo Clinic studies (Darley et al., 1969a, b) is widely held as the most prominent dysarthria classification system, studies have elucidated certain limitations within the classification system (Kim et al., 2011; Weismer, 1984). A central assumption of the Mayo classification system is the relative homogeneity within each classification group rather than between groups. There are clusters of certain perceptual features that have been suggested to co-occur within each classification group, making each group distinct from another group perceptually (Duffy, 2020).

However, there is a concern that there is poor reliability between clinical diagnosis given to dysarthric speakers and the classifications given by experienced listeners when they are blinded to the aetiology of speech (Fonville, Worp, Maat, Aldenhoven, Algra, & Gijn 2008; Van der Graaff et al., 2009). These findings challenge the central proposition that each classification group consists of distinct and recognisable perceptual speech symptoms. Furthermore, Liss et al. (2009) and (Liss, LeGendre, and Lotto (2010) also demonstrate that the Mayo classification system has been only minimally validated with large-scale studies of acoustic speech characteristics.

There are similar issues as those identified above. There have been limitations highlighted in classification based on neurogenic aetiology. This posits that classification is based on injury in similar brain regions resulting in predictable and similar patterns of underlying motor disorders in each speech subsystem (Duffy, 2020). These motor disorders may include reduced strength, spasticity, rigidity, or coordination. This type of classification was also the basis of the Mayo classification system. Although aetiology can give insight into how dysarthria types may be grouped alternatively, it has been suggested that this may not be any more valid technique (Weismer, 1984, 2006).

It has been strongly assumed in the past that particular patterns of muscular disorder will contribute in a direct way to similar speech symptoms in dysarthria. This would make it easier to generalise speech symptoms and treatments to speakers with the same underlying type of neurological impairment. However, this rationale that speech symptoms result from differences in the strength or the steadiness of muscles in non-speech-related tasks is unfounded (Weismer, 2006). By focusing on the neurologic aetiology basis of classification, there can incorrectly be an emphasis placed on training muscles within severely impaired speech subsystems which may not translate to improving speech (Weismer, 2006).

Acoustic analysis can offer an advantage in mitigating some of the limitations above. The acoustic analysis method can also benefit when the baseline variability in speech symptoms causes significant differences between the measures within a particular group, often resulting in statistically insignificant findings in dysarthria research of treatment approaches (Fletcher, 2016). Acoustic analysis provides a method to investigate features in a quantifiable way by measuring characteristics that may not be directly impacted due to the presence of others, even if there may be within-speaker commonalities in these features. However, it is essential to note that acoustic analysis is limited in how much it can quantify how different a dysarthric speaker may “sound” compared to another. No single acoustic measure can detect dysarthria as effectively as a listener (Liss et al., 2010; Sapir, Ramig, Spielman, & Fox, 2010). Furthermore, some physiological processes, such as increased nasal emission and vocal fold spasticity, remain challenging to isolate via any single acoustic measurement (Kent, Weismer, Kent, Vorperian, & Duffy, 1999; Maryn, Corthals, Van Cauwenberge, Roy, & De Bodt, 2010).

Some acoustic parameters that have been investigated and shown promise in identifying speech differences across different types of dysarthria and controls are speech rate and rhythm (Liss et al., 2009), vowel articulation (Sapir et al., 2010), and pitch and intensity variation (Bunton et al., 2000; Rosen, Kent, Delaney, & Duffy, 2006). There have been reported differences in the VOT between different types of dysarthria as well (Duffy, 2020). Vowel space area (VSA) has also been shown to be a suitable parameter for the differential diagnosis of dysarthria. However, VSA measurements have been shown to often present with inter-speaker variability and also have different success rates in distinguishing between dysarthria and control speech (Sapir et al., 2010). However, there are significant limitations in the information that acoustic parameters can give on the listener’s perception of the disorder, which is why an understanding of speech intelligibility (as detailed at the start of the chapter) is also valuable towards establishing a stronger link between perceptual assessments and acoustic parameters to aid investigating variation in dysarthric speech.

Research has indicated a significant discrepancy between perceptual and acoustic findings concerning speech rate. A listener making a judgement of PD speech may perceive a fast rate of speech compared to normative speech even though speaking rates may be revealed to quantifiably be similar (Theodoros & Ramig, 2011). This difference between the perception of speech rate and the actual speaking rate has been posited to be due to potential articulatory imprecision and continuous voicing, which can blur acoustic contrast within the speech signal (Kent & Rosenbek, 1982; Weismer, 1984). Considering this, it becomes crucial for listeners to pay attention to their perceptual assessment of speech rate and compare these results with other raters through acoustic analysis.

Hypokinetic dysarthria also commonly presents with dysfluent speech, which can manifest as inappropriate silences, difficulty initiating speech, phoneme or syllable repetitions, or palilalia (Chenery et al., 1988; Darley et al., 1969b, 1969a; Duffy, 2020). Phoneme repetition has been reported to be presented in 16 to 44% of PwPD but only in a mild form (Darley et al., 1975; Theodoros & Ramig, 2011). Palilalia is the compulsive repetition of words or phrases and occurs along with the presentation of increasing rate and decrease in loudness of PD speech. These word and phrase repetitions are likely to occur at the end of utterances compared to the repetition of phonemes at the beginning of words (Duffy, 2020; Lapointe & Horner, 1981). It has been reported that moderate to severe dysfluency is not common in PwPD (Theodoros & Ramig, 2011). Individuals with a history of childhood stuttering may experience greater stuttering severity upon PD onset (Shahed & Jankovic, 2001).

Respiratory impairment in PD is cited as contributing to deviant speech characteristics such as reduced loudness, variable speech rate, and short rushes of speech (Murdoch, 2009; Theodoros & Ramig, 2011). Physiological consequences of respiratory dysfunction include muscular weakness, reduced endurance, reduced vital capacity, and shortness of breath. Ewanowski's work (as cited in Theodoros & Ramig, 2011) found that Parkinsonism displayed differences in breathing-phonation patterns, where there was longer latency before beginning expiration after forceful inspiration compared to a control group and longer latency before beginning phonation compared to a control group. Darley et al. (1975) suggest that respiratory dysfunction in PD results from muscular rigidity and reduced muscular activation, which limits the movement of the chest wall and coordination of the rib cage with the abdomen. These weaknesses will often manifest in speech through reduced vowel length & duration and a lower number of syllable utterances per breath.

Although a focus on differential diagnosis has isolated certain acoustic features unique to hypokinetic dysarthria in PD, these features are not consistently observed in all studies (Murdoch, 1998; 2009; Theodoros & Ramig, 2011). For example, one of the most prominent characteristics identified in hypokinetic dysarthria is the presence of articulatory undershoot during closure, which is the tendency to display less precise articulation during the production of plosives (Karlsson et al., 2014). This can be observed in the speech signal as spirantisation, representing the presence of aperiodic noise during closure by people with PD. Although patients with PD seem to display this more than controls (Ackermann & Ziegler, 1991; Weismer, 1984), the variability in occurrence often lies within a range that is observed among controls as well, making it difficult to isolate as a marker unless observed consistently or in combination with other distinguishing features (Karlsson et al., 2014).

Loudness tends to lower during conversational speech compared to reading tasks or tasks that require practising the material (Skodda et al., 2009) and suggested that this may be due to some demand on attention by PwPD (Ho, Iansek & Bradshaw, 2001). In addition, PwPD finds it challenging to maintain loudness through the course of a reading task or sustained phonation of vowels (A. Ho et al., 2001), resulting in a more significant loudness decay in these tasks, and during diadochokinetic tasks (Rosen, Kent & Duffy, 2005) when compared to normative speech. Based on the studies above, it is essential that loudness is measured reliably and should use perceptual and instrumental assessment (such as acoustic analysis) while being cautious of how results are interpreted in light of the type of tasks that are analysed.

One way to mitigate the above problem is by using alternative systems that quantify perceptual and acoustic features so they can be easily extracted from the speech signal and compared against control speech to identify the more distinguishing features. This allows for more data to be analysed, making it likely that the features used will be reliable (Kent et al., 1999). Kent et al. (2003) reviewed the literature of studies that used the Multi-Dimensional Voice Program (MDVP; Elemetrics, 1993) to analyse voice dysfunction in various types of dysarthria, including hypokinetic dysarthria associated with PD. MDVP can extract many features that provide information about the articulatory, phonatory, and resonance problems associated with dysarthria. Based on studies that have used MDVP and their study (Kent et al., 2003), overarching features of the sample group with hypokinetic dysarthria in PD were identified: F_0 variation, peak-amplitude variation, and soft phonation index. However, the dataset of individuals with PD was only male and in-group variation was noted but not investigated further. This poses a problem for using systems that ignore in-group variation since the issue of finding acoustic markers that are consistently observed is still a challenge.

In an attempt to find more efficient diagnostic systems, studies have attempted to use machine learning techniques to see if dysarthric speech can be accurately recognised and categorised (Gemmeke, Sehgal, Cunningham, & van Hamme, 2014; Hawley et al., 2007; Middag, 2012; Rosen et al., 2006; Selouani et al., 2012). Hypokinetic dysarthria can consistently be distinguished from the control (Rosen et al., 2006; Van Son, Middag & Demuynck, 2018). The system shows that some measures can lend themselves to automatic detection but also shows that speaker-specific markers may be needed to ensure robust systems and interventions (Middag, 2012). Researchers have proposed that their future studies will focus on identifying speaker-specific markers and then compare them with group markers to assess which features are the strongest indicators of dysarthria (including hypokinetic dysarthria in PD; Middag, 2012; Van Son et al., 2018).

Over the last two decades, scientists have developed several acoustic signal analysis methods aimed at assessing parkinsonian speech (Benba, Jilbab, & Hammouch, 2016; Eliasova, Mekyska, Kostalova, Marecek, Smekal, & Rektorova, 2013; Rusz, Cmejla, Ruzickova, Klempir, Majerova, Picmausova, Roth, & Růžička, 2011; Tsanas, Little, McSharry, & Ramig, 2011). Despite extensive research, some issues (for example, early-stage detection or accurate progress estimation) remain unresolved. New, robust, and sophisticated speech parametrisation methods have emerged over time. However, this investigative evolution frequently creates a barrier between research and clinicians, referred to as "the issue of clinical interpretability" (Mekyska et al., 2022). A feature with high discrimination power or an excellent ability to monitor disease progression can be proposed. However, it becomes useless once we try to find relationships between its value and clinical signs of hypokinetic dysarthria. Clinicians require transparent parametrisation in order to make an accurate diagnosis. When the value of a feature changes, clinicians must know what the outcome will be in terms of clinical signs. In other words, clinically interpretable features will directly quantify clinical signs, but clinically uninterpretable features may only provide a correlation between values and clinical signs. However, they will not provide clarity on the relationship between the two (Mekyska et al., 2022). Therefore, using acoustic correlates of perceptual features, as in the present study, can help interpretability (Kent, 1996). These issues are essential to bear in mind as research into dysarthric speech continues.

2.4. Acoustic analysis for tracking dysarthria progression

The previous section outlined research highlighting issues with dysarthria classification systems, acoustic parameters suitable for differential diagnosis of dysarthria, and limitations in perceptual and acoustic approaches that impact how interpretable results are. Acoustic parameters that are robust for diagnosis may not necessarily be present in speech over time. Some longitudinal studies have been conducted to identify features in hypokinetic dysarthria over time but have suggested that the same problem of high variability in speech makes it difficult to identify markers that can be tracked (Harel et al., 2004; Skodda et al., 2009, 2010, 2013; Skodda, Flasskamp, et al., 2011). However, methodological limitations also made the acoustic parameters found hard to generalise and reliable acoustic markers of PD speech change.

A study that has most closely tried to find trackable acoustic markers is Harel et al. (2004), which conducted a retrospective case study analysis on two individuals diagnosed with PD. Participants' speech was extracted from publicly accessible databases and selected from at least five years prior to diagnosis and up to three years after. They found that F_0 variation and voice onset time (VOT) were consistently less than normal speech and showed deterioration over time. Both measures have been suggested as potential biomarkers to be identified as pre-cursors to PD onset and early disease progression. Percent pause time and diadochokinesis (DDK) were not significant markers for hypokinetic dysarthria in PD. However, the authors (Harel et al., 2004) suggest that the former result may be due to the use of free speech for acoustic analysis, which prevented control over the type of utterances and length of segment durations. A follow-up analysis (Harel et al., 2004) was conducted on another four speakers with PD and a control group to find corroboration for the above results. F_0 variability was once again found to be a marker, though VOT did not show the same effect, and instead, percent pause time was a factor.

In another study (Skodda et al., 2009), the acoustic parameters of speech rate and pitch variation were analysed to monitor PwPD speech over time and compare them to control speakers. The total speech rate (syllables per second related to the total speech time) and net speech rate in the PD group decreased from the first data collection to the second data collection, especially in male participants. However, these values did not show a significant difference from the values of the control group. Pitch variation also showed gender disparities where the female participants with PD had decreased pitch variability over time, while the male participants' values remained relatively stable.

Finally, the F0 variation in male and female participants was significantly reduced compared to the control group in the first and second data collections.

Their results indicated that changes in speech between 7-79 months after the first data collection were independent of dopaminergic medication. Throughout the study, global motor impairment in PD remained largely stable in PwPD speakers. Further limitations of the study were that data was not collected from the control group at the second data collection, which means that all PD values from the first and second data collection were compared to only the first data collection of the control group. This would not account for any natural ageing effects that may have occurred during the study. In addition, as the PD disease duration and the time interval between both data collection periods were not controlled and were between a large range, the study cannot provide evidence on whether the speech changes in PwPD occurred concurrent to speech motor deterioration. Finally, all the participants in the study were at an advanced stage of PD, and there can be no inference on how speech change occurs in the early stages of PwPD. Therefore, further investigations are required that analyse speech changes in PwPD at a set interval between data collection and include early stages of PD as well.

The hypothesis that the speech changes observed in PwPD speech may be independent of dopaminergic medication was further substantiated in a later study (Skodda et al., 2013), where the same authors regulated the dopaminergic medication to maintain similar levels of global motor function in PwPD speakers throughout their study and once again found that speech changes occurred while global motor impairment remained stable. This would substantiate the necessity to further explore acoustic parameters able to track PwPD speech change as they may provide insight into speech changes that have an underlying nondopaminergic mechanism, and this will form the basis for the development of therapeutic approaches. However, since the majority of PD participants in their study were in an advanced stage of disease at the first data collection (mean disease duration about six years) and, like the previous study, the follow-up interval of analysis lay within a wide range, the results cannot answer the question if speech deterioration occurs continuously or instead stepwise and if the different speech modalities show a similar pattern of decline over time (Skodda et al., 2013).

The authors suggest that since global motor impairment was controlled through dopaminergic medication, the findings of the study would suggest that voice and speech impairment in PwPD may be the result of an escalation in dysfunction that may be too subtle to be mirrored in global motor function but could be captured in the Hoehn-Yahr score (Hoehn & Yahr, 1967). The Hoehn-Yahr scale helps describes the progression of PD through its various stages indicating the severity of each case. This could explain the scale's close correlation with the perceptual ratings and some of the acoustic measures (pause ratio and percentage of pauses in polysyllabic words) in the study (Skodda et al., 2013). Alternatively, they posit that the changes observed in the acoustic parameters could be completely independent of motor performance that may be based upon nondopaminergic mechanisms. However, their study cannot be used to infer the natural progression of PwPD speech changes in speakers not on dopaminergic medication.

The studies detailed in this section provide insight into the longitudinal studies conducted that specifically tried to use acoustic parameters to assess speech changes in PwPD. They provide an excellent starting point for further investigation in this domain. However, as stated above, the various limitations of the studies substantiate further investigation, which will be conducted in the present study. The present study will attempt to control for these limitations and find acoustic parameters that quantify perceptual features to take advantage of the clinical value of perceptual assessments, as presented earlier in this chapter.

2.5. Aims and rationale for the present study

The literature explored in this chapter's previous sections has shown that several studies have arrived at certain overarching features of hypokinetic dysarthria associated with PD. However, they often focus on only one dimension of perceptual features or cannot generalise results because of individual variation. Even though there appears to be a pooling effort to find acoustic markers for differential diagnosis of PD from control, only a small amount of research exists that focuses on trackable markers that capture PwPD speech change. Harel et al.’s (2004) study provide some evidence that F0 variation and VOT manifest in individuals consistently over some time and support group studies that show these features are present in the group as a whole (Galaz et al., 2016; Grosset et al., 2009; Theodoros & Ramig, 2011). The overarching acoustic parameters found in PwPD must be effective for differential diagnosis and tracking speech changes in PwPD. In addition, investigating whether acoustic parameters can track changes in speech intelligibility will help elucidate the acoustic correlates that contribute to speech intelligibility in PwPD over time and the relationship between perceptual features that contribute to speech intelligibility over time.

The present study focused on identifying trackable acoustic speech markers that could indicate speech changes in PwPD. Although independent longitudinal studies of PD speech have been conducted, the variable manifestation of speech, as well as the variable intervals between data collection of longitudinal studies, makes identifying acoustic markers for tracking PD speech challenging (Grosset et al., 2009; Harel et al., 2004; Skodda et al., 2009, 2013; Skodda, Flasskamp, et al., 2011; Theodoros & Ramig, 2011). The research presented in this thesis is at the forefront of finding acoustic markers that can track PwPD speech change. The study investigated not only acoustic markers that can track PwPD speech change but disentangled trackable acoustic markers from those that are effective for differential diagnosis between PwPD and control speakers. How this was achieved will be detailed in the following chapter.

The present study, as far as the author is aware, is a first attempt to use both perceptual assessments and acoustic analysis in identifying acoustic markers to track PwPD speech changes across a fixed time interval of six months. Six months allows the study to fix the interval between analyses (which the previous longitudinal studies failed to do) and identify any acoustic parameters that can track PwPD speech change in less than a year. This study aims to contribute to research in achieving the larger objective of creating a reliable and standardised system that adds to existing perceptual evaluation and intervention methods. The findings in this study will allow more efficient systems to supplement current ones in the future.

The acoustic markers found in this study could be used to create preliminary person-specific 'voice profiles' as an efficient and cost-effective method of observing speech change in the future. Creating a voice profile per person will have the added benefit of being compared to a control speaker to verify deviations and quickly identify areas where compensatory mechanisms in speech production can be employed. In addition, this study creates the groundwork needed for future research that could employ automatic detection processes (e.g., machine learning) to extract these established markers from the speech signal of PwPD and assist SLTs in creating intervention plans.

3. Method

3.1. Research Questions

This study aimed to identify acoustic speech markers that are able to track speech changes in PD over time. This project will be exploratory in nature and attempt to answer the following global research question:

Which acoustic parameters are able to capture speech changes in hypokinetic dysarthria associated with PD within a year?

Specific research questions:

1. Which acoustic parameters can track perceptual changes in PwPD speech over time?
2. Which acoustic parameters can track changes in PwPD speech intelligibility over time?

This study will first test the null hypothesis that the acoustic features are not able to track speech changes in PwPD.

3.2. Methodology

3.2.1. Participant Recruitment

After ethical approval was obtained from the Queen Margaret University Ethics Committee on 18th September 2020. Speakers with PD were recruited from Parkinson's UK facilities using their recruitment portal. The control group was recruited through channels associated with Parkinson's UK and Queen Margaret University, Edinburgh.

There were two data collection points for the study: February 2021 (T1) and August 2021 (T2). At the T1 collection point, 120 participants were recruited. However, before the T2 collection point, 10 participants dropped out. This study's final recruitment was 110 participants (PwPD = 63; Control = 47).

Within the PwPD group, the age range of the participants was between 45-93 years (Mean = 69; SD = 8.4), with 20 females and 43 males. The time since PD onset ranged from 6 months to 40 years before T1 for this study, with a mean of 8.4 years (SD = 6.7) post PD onset. All PwPD speakers self-reported they did not have any cognitive or mental health condition (such as dementia or depression) associated with PD symptomatology or otherwise or a diagnosis of a speech or voice disorder that is not dysarthria.

Within the control group, the age range of the participants was between 35-86 years (Mean = 64; SD = 12.2), with 30 females and 17 males. All control speakers self-reported they did not have any cognitive or mental health condition or ever had a speech or voice disorder diagnosis.

3.2.2. Design

The design of the study used quantitative research to answer the research questions. Since the objective involved comparing speech over two different time points, data was collected from the same group of participants six months apart.

The analysis involved a perceptual experiment that involved two SLTs rating PwPD speech across the five subsystems for speech severity. Based on the results of this perceptual experiment, acoustic analysis of PwPD and control speech was conducted by extracting acoustic features from the speech signal to assess the patterns and markers across speakers specific to hypokinetic dysarthria. Statistical analysis involved linear mixed effects model analyses to compare changes both between groups and over time. Portions of this methods section and procedure have been published in Murali (2022).

3.2.3. Materials

Participants were given a participant information sheet prior to recruitment (see Appendix A). Those interested then filled out a participant consent form and demographic sheet (Appendix B and C).

Having mental health issues can have a negative impact on speech (Cohen et al., 2014). It would make it difficult to distinguish between speech deficits resulting from low mental wellbeing and PD. Therefore, it was important to try and control this confound for this study. In order to assess whether deterioration of mental wellbeing was observed between both time points, both the PwPD and control participants were asked to complete the self-reporting Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS; Tennant et al., 2007; see Appendix D) before T1 and T2, to mitigate the confound of secondary conditions developing between data collection points. In addition, PwPD participants were asked prior to the T2 about any changes in their PD diagnosis. Both PwPD and control participants were asked if any significant life changes had occurred in the previous six months. All participants reported none of the above.

Speech materials were selected based on previous clinical research and studies involving acoustic analysis (Harel et al., 2004; Skodda et al., 2009; Theodoros & Ramig, 2011). The type of materials selected for recording was intended to collect a wide variety of speech data to ensure a thorough investigation of the research questions could be conducted. Speech samples (see Appendix E) from each participant included the following: a) recordings of minimal pairs (20 pairs, e.g., sip-ship, sat-chat, key-tea), b) sustained phonation (extended production of the vowel /a/), c) reading from a standard passage used in clinical settings (the Grandfather Passage; Darley et al., 1975), and d) two minutes of spontaneous speech (based on cues). The order of speech tasks was as follows: minimal pairs list, the grandfather passage, and the sustained phonation, which were produced twice. The final task was the spontaneous speech which was produced once per cue.

3.2.4. Procedure

3.2.4.1. Pilot 1

An initial, small pilot was conducted to ensure the entire study procedure was clear and feasible. It was also used to verify the sound quality of the data and ensure the acoustic analysis could be conducted. Only acoustic analysis was done to check that the data collected was amenable to feature extraction. This pilot recruited people (N = 3) through channels within Queen Margaret University, Edinburgh. The participants in the pilot group were not included as part of the main study.

The initial remote data collection method involved the participants using their smartphones to download a specific voice recording application (*AVR*, 2021) and recording their own speech (by reading speech material sent to them via post). They were given instructions on how to set up before the recording, i.e., the distance they would sit from a table, the distance at which their smartphone would be placed, their posture, etc. They would then be asked to upload their recordings to a secure folder dedicated to each participant.

The pilot revealed that the recording/data collection process was too elaborate and put too much burden on the participant to follow lengthy instructions to record their speech. It also assumed that participants would follow instructions correctly. The researcher could not monitor the process unless a video call were conducted. Therefore, another method of data collection was deemed necessary.

3.2.4.2. Pilot study 2

Podcasts often rely on the remote recording of high-quality audio signals in interviews. So, it was decided to investigate the feasibility of using a Podcast programme to capture data for the study. After looking at various options, Squadcast (2021, Version 4.5) was selected as it was easy to setup (i.e., no installation required), the audio quality was as required for acoustic analysis (44kHz, 16-bit in.wav format) and the video call feature which ensured that the researcher and participant could see each other and feel comfortable prior to recording. The caveat of this method was that it relied on a strong and stable internet connection. However, this was not unique to Squadcast but a feature of most remote data collection options. Squadcast (2021, Version 4.5) also records using the local device's microphone before saving them to a cloud with no file compression occurring. In addition, cloud backups were uploaded simultaneously and saved in case technical issues were experienced.

The new means of data collection using Squadcast was tested with a second pilot study (N = 2) using the same recruitment channels. As in the previous pilot study, speech data sheets were sent to participants via the local postal service before recording. Participants from the pilot reported finding Squadcast easy to use, and the researcher found no issues with audio quality, connectivity, or data backups. Based on this pilot, Squadcast was selected as the data collection program used for this study, with no further changes to the program setting needed.

3.2.4.3. Data collection procedure

After the consent form, participant demographic forms and WEMWBS were filled out, participants were contacted to arrange a suitable time to record. They were then sent an electronic link which allowed them to join a video call hosted on the podcast program Squadcast (2021, Version 4.5) at the time they selected. Participants could join either through an Android smartphone, a laptop, a desktop computer, or a tablet with Chrome, Firefox, or Microsoft Edge web browsers. All participants joined the following laptop or desktop computers: PwPD (32 = Dell laptop; 28 = Macbook Air; 3 = iMac desktop computer), and control (28 = Macbook Air; 19 = Dell laptop). Model information for laptops were either unknown or not provided. No program installation was required, which eased the process and greatly reduced the potential for participant dropout. The recordings were made short to avoid physical discomfort or strain experienced by participants, and the study allowed participants to record at a time that was convenient for them.

Each participant was given instructions on setting up for the video call prior to beginning the recording and was requested to sit in a quiet environment with windows closed. In addition, participants would ideally sit at a table with a comfortable seat and place the device used for the call approximately 30 cm from them, trying not to lean to or from the device during the recording. The researcher began the audio recording once the participant was ready. Speech data was read from the sheets previously posted to participants. Only the participant's audio containing the speech data was recorded to protect the identity of the participant and retain confidentiality. The entire video call took 15 minutes on average, excluding any technical issues that needed to be fixed.

The above data collection procedure was repeated six months later with the same group of participants.

3.3. Data Management

The recordings for each participant were coded using a unique participant number and code, as well as a group code, to quickly identify the PwPD and the control group. This was done to collate and link-anonymise the speech data. All data codes and participant demographic data were recorded on an encrypted excel sheet for reference and were only accessed by the research team. Personally identifiable data was coded and not disclosed to anyone outside the research team or in any other form (publication/conferences, etc.).

In the cases of dropouts, participants personal information was destroyed, but participants were informed that any recordings already collected would be retained and stored anonymously. This information was recorded next to the participant's code and not included in the data analysis. Incomplete or missing data were only included for analysis if they could be justified and met the needs of the study.

3.4. WEMWBS rating scale results

The WEMWBS (Tennant et al., 2007) was used to assess the baseline mental wellbeing of the participants before T1 data collection and reassessed before the T2 data collection point to ensure that there was no deterioration of mental wellbeing. The WEMWBS Scale is a self-reporting14-item measure of wellbeing, first validated in the UK (Tennant et al., 2007). Each item is rated on a five-point Likert scale ranging from ‘none of the time’ to ‘all of the time’ and scored by adding the total for each time, giving a result ranging from 14-70.

The mean and standard deviation of scores for all participants across the PwPD group and the control group were very similar to the UK population norms (mean = 51.6; SD = 8.7; Tennant et al., 2007). According to the scoring norms, a change is considered meaningful if a participant’s score changes by +/- 3 between both response times. However, a meaningful change does not indicate a significant change.

The PwPD group scores prior to T1 were: (mean = 49.07; SD = 8.36), and T2: (mean = 51.03; SD = 8.52). Based on responses from 63 participants, there was a 17.5 % (N = 11) meaningful positive change in scores and a 33.3 % (N = 21) meaningful negative change in scores from T1 to T2. The mean change in PwPD group scores from T1 to T2 was only -1.00 showing individual participants did not have extreme changes in their scores between both data collection points. A Wilcoxon signed rank test used to assess any significance of change between PwPD group scores, showed the change was not significant ($p > 0.05$; see table 1).

The control group scores prior to T1 were: (mean = 55.10; SD = 6.92), and T2: (mean = 54.67; SD = 7.33). Based on responses from 47 participants, there was a 15.2 % (N = 7) meaningful positive change in scores and a 23.9 % (N = 11) meaningful negative change in scores from the first to second time point. The mean change in scores from T1 to T2 was -0.98, indicating individual participants did not have extreme changes in their scores between both data collection points. The results of the Wilcoxon signed rank test (see table 2) used to assess any significant change between control group scores, showed the change was not significant ($p > 0.05$).

Figure 1 shows the PwPD and control group participant scores across T1 and T2, including the percentage of participants ranked as having either low, medium, or high wellbeing based on the following: 'low' if the total score <43, 'medium' if the total score <61, 'high' if total score >61. It is important to note that while some participants were categorised into low mental wellbeing, none of the participants reported being diagnosed with depression or other significant mood disorders at either data collection point. Several participants also reported that responses to the WEMWBS scale were highly influenced by feelings of isolation and general low arousal due to the COVID-19 lockdown. This could have led to the resulting 'low' wellbeing scores.

Based on the results of the WEMWBS scores, it was concluded that all participants could be included for the purposes of the study.

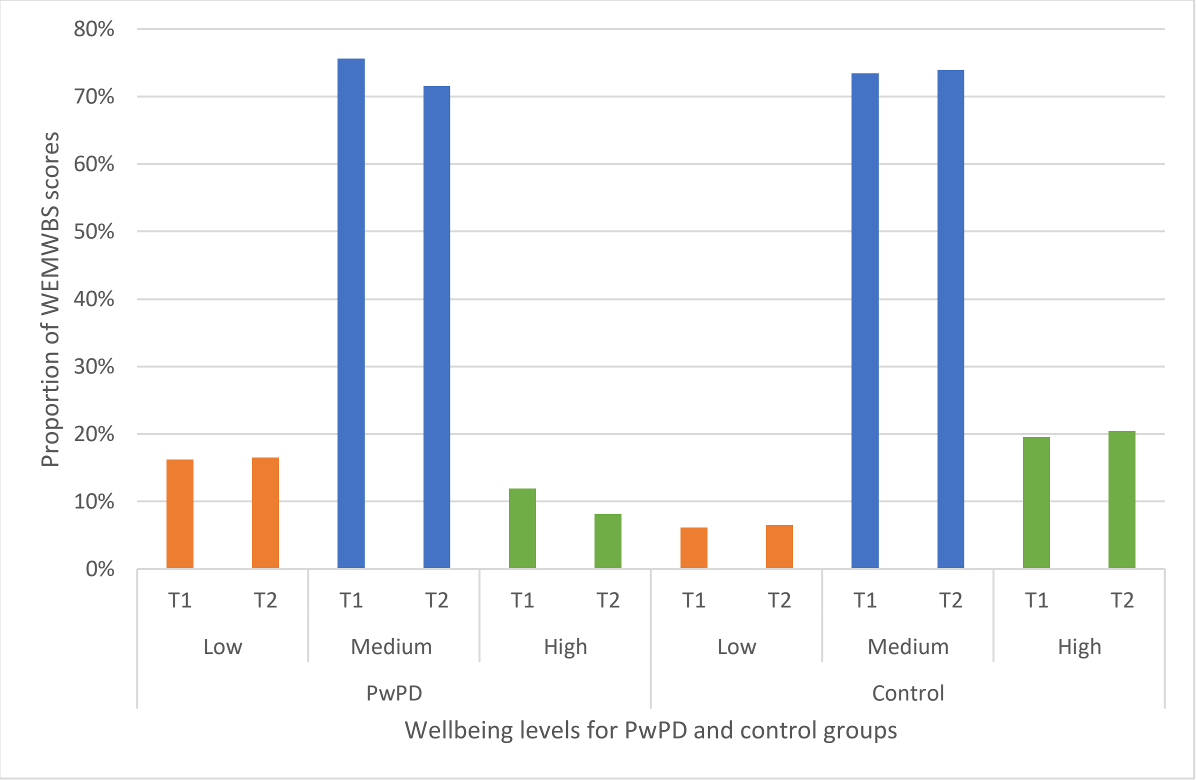
Table 1. Results of the Wilcoxon signed rank test of PwPD speakers WEMWBS scores of both data collection points.

	n	$\mathbf{U_R}$	**SD**	**z**	**p**
Positive ranks	15	1008	146.07	-2.66	>0.05

Table 2. Results of the Wilcoxon signed rank test of control speakers WEMWBS scores of both data collection points.

	n	$\mathbf{U_R}$	**SD**	**z**	**p**
Positive ranks	13	540.5	91.53	-1.74	>0.05

Figure 1. Percentage of WEMWBS scores for PwPD and Control speakers showing low, medium, or high wellbeing over the two data collection points.



3.5. Perceptual feature ratings

3.5.1. Setting up the SLT experiment

Two SLTs were asked to conduct the ratings and were recruited through Queen Margaret University, Edinburgh. Both raters were experienced in perceptual ratings of disordered speech, and were familiar with dysarthric speech. Rater 1 had over fifteen years of experience and rater 2, one year. Given the large difference in experience between both raters, an inter-rater reliability analysis was conducted (see 3.5.2) to determine the agreement of the perceptual ratings between both SLTs.

The two SLTs participated in a perceptual experiment on Gorilla Experiment Builder (Anwyl-Irvine et al., 2020) (accessed from www.gorilla.sc) to establish the overall perceptual speech severity rating and profiles of the deviant speech characteristics of each speaker. The perceptual rating was conducted by setting up an experiment for the five speech subsystems, respiration, phonation, articulation, resonance, and prosody. The five subsystems were regrouped into seven categories based on Darley et al., (1975)'s system of classification and adapted by Duffy (2020). The seven categories were articulation, prosody, pitch, respiration, resonance, voice quality and loudness. All references will be made to these seven categories henceforth.

Speech samples of 30 second duration were extracted from the standard reading passage, The Grandfather Passage, starting from the second sentence for each PwPD speaker. The first sentence was not included in the perceptual rating to avoid an inaccurate assessment of speech rate while participants acclimatized to the reading passage, or any errors limited to starting speech production, unrelated to PD. These samples were uploaded to Gorilla Experiment Builder and used to set up a perceptual experiment that facilitated ratings across all seven categories. The experiment was set up to limit rating to only one category per day (to avoid fatigue and inaccuracies). The speech samples were presented in a randomised order and allowed to be re-played up to 3 times. Perceptual speech characteristics were rated in each category for one speaker before moving on to the next. Ratings were made on a scale of 0-4 (0= normal; 1= mild; 2= moderate; 3= marked; 4= severely deviant and perceptual features in each category were used as presented in Duffy, 2020; (see Appendix F). The rating system is based on Darley et al.'s (1969a, b, 1975) studies and has since been used in other perceptual assessments (see Duffy, 2020 for further details). Each SLT rated all the categories for T1 first, and the second round of ratings was completed following T2 data collection. Since ratings were done after each data collection point and were six months apart, there was no priming or bias between raters from one set of data to the other.

Once ratings were complete, each category was compared between both T1/T2 and between both raters to assess which categories were rated as most deviant across speakers. Once the categories rated as most deviant were identified, speech features within those categories rated as “marked” or “severe” were isolated to select the acoustic parameters that were used during acoustic analysis.

3.5.2. Perceptual ratings reliability analysis

Intra-class correlation (ICC) was used to evaluate the inter-rater reliability of the perceptual ratings by the two SLT raters prior to data analysis. Cohen’s k was used to evaluate intra-rater reliability for each rater at the two data collection points. Results for inter-rater reliability shown in Table 3 confirmed that raters had fair agreement T1 data collection point (0.415) and fair agreement at T2 (0.480) based on the mean ratings both raters gave for each data collection time. Both raters’ agreement at both time points is significantly higher than chance.

Table 3. ICC showing inter-rater reliability between two SLT raters for data collection 1 and 2.

		95% Confidence Interval		**F Test with True Value 0**	
	ICC	**Lower Bound**	**Upper Bound**	**Value**	**df**
Single ratings per rater, time 1	.261	.207	.313	1.767	2708
Mean ratings, time 1	.415	.343	.476	1.767	2708
Single ratings per rater, 2	.316	.235	.388	2.044	2708
Mean ratings, time 2	.480	.381	.559	2.044	2708

3.6. Speech subsystems and acoustic parameter selection

Based on the perceptual features, three categories were rated as most severely impacted within the subset and were chosen for acoustic analysis: articulation, prosody, and voice quality. Within each category, perceptual characteristics rated as deviant were identified, and acoustic parameters commonly associated with these speech features were identified for subsequent analysis. Below are the acoustic parameters that were selected for each of the three categories, accompanied by a description of each acoustic parameter. Details of how each acoustic parameter was selected, annotated, and extracted from the speech recordings are provided in Chapters 4-6, corresponding to each category.

3.6.1. Voice quality

Within voice quality, perceptual ratings showed that breathy, hoarse (wet), and strained were the features rated as being severely impacting PwPD speech. The acoustic parameters used to assess voice quality were: 1. Jitter, which is a measure of frequency variation between consecutive periods of the sound. Jitter measures the variability of F0 from one cycle to the next; 2. Shimmer, which is a measure of amplitude perturbation and measures the maximum amplitude of each vocal fold vibration from one cycle to the next; 3. Harmonics-to-noise ratio (HNR), which is a measure of increased noise. It measures the periodic and non-periodic components of a speech sound; 4. Cepstral peak prominence (CPP) was chosen as an additional measure to capture decreased voice quality (Rusz et al., 2021b) in the reading passage. CPP can be defined as the measure of cepstral peak amplitude normalised for overall amplitude. The acoustic parameters selected are predicted to show both trackable and global changes, based on previous literature (Harel et al., 2004; Rusz et al., 2021a).

Some of the speech motor symptoms in PD that impact voice quality include weakness in the vocal fold resulting in a gap between the vocal fold (known as vocal fold bowing; Hanson et al., 1984; Perez et al., 1996), muscle weakness in the tongue, lower face and velum, involuntary movement of the tongue, tremor in the lips and vocal tremor due to reduced movement of the vocal fold, and abnormal muscle tone at rest (Theodoros & Ramig, 2011). Jitter and shimmer are the acoustic correlates of rough dysphonic speech, which can be due to the diminishing control of the laryngeal muscles, which leads to unstable periods of vocal fold opening. HNR is a correlate of turbulent noise, which occurs when improper control of the vocal folds leads to a reduced rate of airflow. CPP can correlate to the control of laryngeal muscles. Deteriorated control of laryngeal muscles would result in unstable periods of vocal fold opening, causing a breathy voice. A change in motor function could result in a change in CPP, which might be detected during acoustic analysis.

3.6.2. Articulation

The perceptual features in the subsystem articulation that were rated as marked/ severely deviant were: imprecise consonant production and articulatory breakdown. The acoustic parameters selected to investigate changes in these perceptual features were the plosive and fricative mean intensity for each plosive and fricative in syllable initial position in the grandfather passage. In addition, voice onset time (VOT) was also included as another acoustic parameter.

When plosives and fricative production are impacted, it is broadly referred to as inaccurate articulation (Ackermann & Ziegler, 1991; Duffy, 2020; Pawlukowska et al., 2015). Consonant production in PwPD is often inaccurate, resulting from articulatory breakdown (Y. Kim, 2017). Fricatives can show lower intensity, and plosives can look like fricatives or are produced weaker in PwPD**.** Based on this variability in fricative and plosive production in PwPD, acoustic analysis of plosive and fricative intensity was used to detect inaccurate articulation.

Plosives and fricatives in syllable initial position were identified, and the VOT of these consonants was measured. VOT is a measure of the coordination of speech articulation and voicing and defined as the time between the release of a stop consonant to onset of voicing or vocal fold vibration. This can be impacted by the slowing of lip and tongue movement, which leads to increased time required to produce each consonant (Rusz, Tykalova, Novotný, Růžička & Dušek, 2021; Rusz, Tykalova, Ramig & Tripoliti, 2021).

3.6.3. Prosody

Within all the perceptual features in prosody, speech rate was the only perceptual feature that showed a high severity rating. Speech rate was measured using the recording of the grandfather passage by measuring the number of syllables over the total duration of speech after the removal of pauses.

To better understand the severity rating for the subsystem prosody, mean intensity and intensity variation was added as loudness measures extracted from the reading passage. Intensity variation is extracted as the standard deviation of the speech intensity contour of voiced segments and can indicate monoloudness. It could correspond to the reduced functioning of the respiratory and thyroarytenoid muscles (Rusz et al., 2021).

Monopitch was also measured from the reading passage by measuring the fundamental frequency variation (F0SD), converting the contour to a semitone scale. The motor symptom correspondence of monopitch would be reduced vocal cord movement leading to glottal incompetence.

A summary shown in table 5 of all acoustic parameters used for this study, along with a description of each parameter, can be found below.

Table 4. Summary table of acoustic parameters and speech tasks used for acoustic analysis.

Speech subsystem	**Acoustic parameter**	**Speech task**	**Description**
Voice	Jitter	Sustained phonation	Variability of the fundamental frequency from one cycle to the next.
	Shimmer	Sustained phonation	The maximum amplitude of each vocal fold vibration from one cycle to the next.
	HNR	Sustained phonation	the ratio between the periodic and non-periodic components of a speech sound
	CPP	Reading passage	The measure of cepstral peak amplitude normalized for overall amplitude.
Articulation	Plosive and fricative mean intensity	Reading passage	The average sound pressure in the plosives and fricatives.
	VOT	Reading passage	The length of a consonant form initial burst to vowel onset.
Prosody	Speech rate	Reading passage	The number of syllables over the total duration of speech after the removal of pauses.
	Mean intensity	Reading passage	The average sound pressure over the total duration of speech after pauses are removed.
	IntSD	Reading passage	The standard deviation of the speech intensity contour of voiced segments.

F0SD	Reading passage	The standard deviation of the fundamental frequency contour.

3.7. Analyses

Two analyses were conducted to answer the two research questions. The first analysis answered the first specific research question of identifying the acoustic parameters able to track perceptual changes in PwPD speech over time. The second analysis answered the second specific research question of identifying the acoustic parameters able to track speech changes in different intelligibility groups over time. Both analyses were based on the results of the perceptual experiment using the acoustic parameters outlined in table 5 above.

3.7.1. Perceptual feature ratings (PFR) analysis

This analysis used the results of the SLTs perceptual feature ratings. Recordings from a subset of PwPD participants were selected for acoustic analysis based on the perceptual features ratings negatively changing from T1 to T2 data collection. A negative change was interpreted as having occurred if both SLTs rated the perceptual features having increased in severity from T1 to T2 (a rating changed from 2 “moderate” to 3 or 4 “marked” or “severely deviant”). If a positive change occurred in the SLT ratings from T1 to T2 (a reduction in severity in perceptual features), the participant was not included for this analysis.

As stated previously, the categories that displayed the most significant change in PwPD participants’ speech were articulation, prosody, and voice quality. The subset was further narrowed down by identifying participants with the most deviant speech characteristics, rated as 3 or 4 (“marked” or “severely deviant”), and therefore more likely for the differences to be tracked acoustically. Using these criteria, recordings of a subset of five participants were selected for analysis of prosody, six for articulation, and fourteen for voice quality. There were four common participants who had been selected for both prosody and articulation but no other common participants across the categories. This brought the total to 21 PwPD participants with a negative perceptual change between T1 and T2, referred to as the PwPD-change group. 21 PwPD participants who were rated as having no perceptual change in their speech (PwPD-no change group) were selected to control against the PwPD-change group. The demographic information for each PwPD group is as follows: PwPD-change: M = 15; F =6, Age = 50-93 years (mean = 69); PwPD-no change: M =13; F = 8, Age = 56-84 years (mean = 71).

Recordings from 10 control participants (M = 5; F = 5, Age = 51-82 years (mean = 70)) were selected to confirm speech differences between the PwPD groups and control group. The control group was age and gender matched as closely to both the PwPD groups as possible. This brought the subset total to 52 speakers. Further demographic information of each speaker is presented in Appendix G.

Both raters were consistent in their rating of a positive, negative or no change and there were not instances found where the raters disagreed within the speech categories of interest.

3.7.2. Intelligibility groups (IG) analysis

This analysis used the overall intelligibility rating provided by the SLTs. The complete dataset of 63 PwPD participants was used to test the same acoustic parameters previously analysed. Speech recordings of PwPD participants were based *only* on the overall intelligibility rating, which was a rating on how severely intelligibility was impacted. Based on the results of the intelligibility ratings, participants were grouped into either mild (for ratings between 0-1, indicating mildly impacted intelligibility), moderate (for ratings of 2-3, moderately impacted intelligibility), and severe (ratings of 4, severely impacted intelligibility). Within each group, a further separation was made based on whether the intelligibility ratings indicated a negative change or no change between T1 and T2. A negative change was interpreted the same way for this analysis as the PFR analysis. The data split was based on the initial rating from T1 data collection, as some participants' ratings changed from mild to moderate by T2 data collection. The ratings indicated that all PwPD participants were categorised into the mild group. The final PwPD groups for testing were mild-change: N = 34(M = 22; F = 12), age = 52-81years (mean = 66); mild-no change: N = 29; M = 21; F = 8, Age = 50-93 years (avg: 72).

Recordings from the 47 control participants (M = 17; F = 30, age = 35-86 years (mean = 64) were compared against each of the intelligibility groups.

Further demographic information for the participants can be seen in Appendix G. Both raters were consistent in their rating of a positive, negative or no change and there were not instances found where the raters disagreed within the speech categories of interest.

The following chapters will explore the voice, articulation, and prosody analysis conducted during the perceptual feature ratings and intelligibility group analyses, highlighting key findings in the literature and comparing previous findings to the results found for each of the three categories. In addition, the efficacy of using the above acoustic parameters to track speech changes will be discussed, and potential links to motor symptomatology will be described. Thereafter a summary of all findings and their implications will be discussed in a concluding chapter, along with directions for future studies.

4. Voice Quality

Voice quality is dependent on factors such as precise vocal fold control, coordination between muscle contraction/relaxation, and smoothness of vibration (Duffy, 2020). The sounds that emerge from the vocal folds are amplified by resonance which then makes it audible and recognisable as speech. Approximately 90% of PwPD will have some vocal impairment affecting voice quality (Chenausky et al., 2011). Impairment in voice quality can be detected using acoustic parameters extracted from PwPD speech (Benba et al., 2016). As mentioned in chapters one and two, while speech and motor characteristics in PwPD are understood, there is a need to elucidate how speech changes in PwPD might manifest over time and whether acoustic parameters are able to reliably capture these changes. Tracking speech changes in PwPD is important as PD development, and speech changes are often not linear. Speech impairments can be a precursor to PD or one of its first symptoms, while other cases of PwPD will only develop speech impairments years after first receiving a PD diagnosis (Goberman, 2005; Goberman et al., 2002; Ho et al., 2008; Skodda et al., 2009, 2010, 2013). It is necessary to further explore the nature of voice quality in PD, specifically as listeners of PwPD speech often report perceptual characteristics that are the result of an impact on voice quality, with dysphonia being regarded as one of the most prominent features of PD speech (Kent, 1996; Oates, 2009; Sachdeva & Shrivastava, 2018). By isolating specific acoustic parameters that may be able to detect speech changes and attempting to illustrate how these may relate to perceptual assessments of PD speech, a clearer picture of voice quality can be gained. This was explored in this study by combining perceptual assessment and acoustic analysis.

The previous chapter outlined the method and procedure implemented throughout this thesis and the intention to test acoustic parameters pertaining to three different categories (voice, articulation, and prosody) based on the results of the SLT ratings (see chapter 3, section 3.5). In this chapter, the results of the acoustic and statistical analysis of voice quality will be presented. The purpose of these analyses was to check if the acoustic parameters pertaining to voice quality are robust enough to not only distinguish between PwPD speech and controls but also track speech changes over time.

The SLTs rated voice quality at T1 as one of the most impacted speech dimensions, with the perceptual features breathy, hoarse, and strain-strangled voice rated as either 'marked' or 'severe' (rated '3' or '4') on a five-point scale of severity (from '0-4'). Acoustic parameters were selected to quantitatively measure the perceptual features breathy, hoarse, and strain-strangled from the speech signal. The parameters selected were jitter, shimmer, harmonic-to-noise ratio (HNR) and cepstral peak prominence (CPP). The rationales for the selection of these acoustic parameters are included in section 4.1 below.

Two independent data analyses were conducted using the speech recordings of sustained phonation and The Grandfather Passage to investigate the effectiveness of the selected acoustic parameters in answering the research questions. The perceptual feature ratings (PFR) analysis was conducted on a subset of the dataset (n = 52) to answer the first research question: which acoustic parameters can track perceptual changes in PwPD speech over time? This subset was selected based on the SLT ratings. Participants with perceptual features rated as more severe (breathy, hoarse, and strain-strangled) were selected. The recordings of this subset were included in groups based on whether the rating of the severe perceptual features negatively changed from T1 and T2 collection points or did not change at all. Therefore, this analysis focused on grouping the recordings based on detecting a change in the perceptual features over time.

The intelligibility groups (IG) analysis was conducted on the entire participant sample (referred to as the complete dataset; n = 110) and grouped based on the overall intelligibility ratings given by the SLTs and whether those ratings negatively changed from T1 to T2. This analysis would answer the second research question: which acoustic parameters can track changes in PwPD speech intelligibility over time? The overall intelligibility ratings were used to group PwPD recordings as either mild, moderate, or severe and if the ratings increased or decreased between data collection points.

The structure of the rest of the chapter includes details of the selection of acoustic parameters used for both analyses, how each analysis was conducted along with the results and discussion, and an overall discussion of the results of this chapter and its implications for voice quality in PwPD speech.

4.1. Selection of acoustic parameters

Physiological impairments in PD include speech motor symptoms such as weakness in the vocal fold, muscular weakness in the tongue, lower face and velum, involuntary movement of the tongue, tremor in the lips and vocal tremor due to lack of proper vocal fold control, and abnormal muscle tone at rest (Theodoros & Ramig, 2011). PD results in reduced airflow and reduced control of the vocal folds, which increases the turbulent noise observed in dysarthria (Rusz et al., 2021b). Vocal fold bowing (a gap between vocal folds due to atrophy or weakness), as well as incomplete vocal fold closure during phonation, have been observed in the laryngeal mechanism of PwPD with dysarthria (Hanson et al., 1984; Perez et al., 1996). Acoustic studies also indicate a tendency in PwPD speech toward reduced sound pressure level, reduced voice pitch variability, and phonatory instability, as suggested by increased noise and cycle to cycle variability during phonation.(Tjaden, 2008).

Based on the physiological impairments outlined above, speech in PwPD can be perceived as breathy, hoarse, or strain-strangled. These perceptual speech features were rated as most severe by the SLTs (detailed in the previous chapter). The acoustic parameters of jitter, shimmer and HNR were selected based on previous studies (Jiménez-Jiménez et al., 1997; Rusz et al., 2011a,b; Vizza et al., 2019) that used these acoustic parameters as reliable measures of the perceptual features breathy, hoarse, and strained-strangled voice quality.

Jitter is a measure of frequency variation between consecutive periods of the sound. Jitter measures the variability of F0 from one cycle to the next. Shimmer is a measure of amplitude perturbation and measures the maximum amplitude of each vocal fold vibration from one cycle to the next. (Rusz et al., 2021a,b). Jitter and shimmer can be heard as rough dysphonic speech, which can be due to the diminishing control of the laryngeal muscles, which leads to unstable periods of vocal fold opening (Teixeira & Gonçalves, 2016). HNR can be heard as turbulent noise, which occurs when improper control of the vocal folds leads to a reduced rate of airflow. Jitter and shimmer are measures of voice perturbation assessing the micro-instability of vocal fold vibrations, known to increase with laryngeal pathology and effectively help discriminate between voice pathology types (Brockmann et al., 2011; Teixeira & Gonçalves, 2016). Shimmer and Jitter are important parameters for clinical voice assessment (Maryn & Weenink, 2015).

Jitter values for speech of those with laryngeal pathology have shown to be outside the typical range of 0.5-1%, and a shimmer threshold of 3% is for pathological speech (Teixeira et al., 2013; Vizza et al., 2019). Further, studies have also indicated that when investigating perceptual features such as breathy and hoarse voices post-thyroid surgery, perturbations measures of jitter and shimmer may be able to track subtle changes in voice quality that may be undetectable through perceptual methods of assessment (Gelzinis et al., 2008; Ortega et al., 2009). Therefore, jitter and shimmer, as measures of breathiness, hoarseness, and strain-strangled voice, are not only effective at distinguishing between normal and pathological speech but can potentially track changes in speech.

Jitter uses the timing interval between consecutive peaks (the period) to measure the stability of F0 over time, whereas shimmer uses the differences in consecutive peak amplitudes to measure stability in vocal intensity over time (Schultz & Vogel, 2022). These metrics are typically applied to recordings of sustained vowels to achieve reliable estimates of articulator function (Teixeira & Fernandes, 2014). Increased jitter is hypothesized to reflect a lack of motor control over vocal cord vibrations, and increased shimmer is hypothesized to represent a decrease in glottal resistance and/or the presence of vocal cord lesions (Zwetsch et al., 2006).

HNR is a measure of increased noise in the speech signal which measures the periodic and non-periodic components of a speech sound. It has been used as a measure for distinguishing between types of dysarthria and normal speech (Vizza et al., 2019). Since reduced airflow and control of the vocal folds increase the turbulent noise, pathologic speech usually displays reduced HNR indicating dysphonia, with an HNR of lower than 20 dB (Teixeira et al., 2013). A healthy speaker can produce a sustained /a/ or /i/ with a harmonicity (acoustic periodicity) of around 20 dB, while hoarse speakers will have an /a/ harmonicity lower than 20 dB (Boersma & Weenink, 2020). Therefore, HNR was selected to measure hoarseness in PwPD.

HNR is based on the principle that, for a perfectly periodic signal, the maximum possible autocorrelation coefficient is 1, so the noise can be estimated by subtracting the periodic part of the signal from 1. The harmonic component measures the periodic part of the signal (rx), and the inharmonic component (i.e., noise) is the difference between the theoretical maximum and the observed harmonic component ($1 - \mathrm{rx}$). HNR is measured by obtaining the ratio between the harmonic and inharmonic components, expressed in dB (Schultz & Vogel, 2022). Measures of jitter, shimmer, and HNR rely on the assumption of stationarity, that is, that the periodic signal and its mean are relatively stable over time (Schultz & Vogel, 2022).

To understand the overall impact on voice quality in PwPD speech, CPP was added as a measure of dysphonia to account for any changes that may be picked up through acoustic analysis that may not have been encompassed in the individual perceptual features rated by the SLTs. CPP can be defined as the measure of cepstral peak amplitude normalised for overall amplitude. CPP is an acoustic measure linked to overall dysphonia in the speech signal (Hillenbrand et al., 1994; Hillenbrand & Houde, 1996; Kent, 1996), likely because it compares the voiced component of speech with noise within the speech signal. CPP has been measured using a few variations, but all measures compare the cepstral peak within the F0 range relative to an approximation of the noise within the signal (Schultz & Vogel, 2022).

It has been chosen as a global measure of dysphonia, showing the periodic and aperiodic energy in a speech signal (Awan et al., 2016) and in distinguishing between dysphonic and control speakers (Murton et al., 2020; Patel et al., 2018). CPP detects the perceptual characteristic of reduced voice quality (Rusz et al., 2021a,b) and has been shown to be an effective individual parameter for the assessment of voice quality (Patel et al., 2018). It has also been suggested that CPP might be effective in detecting changes in dysphonia in the early stages of PD (Šimek & Rusz, 2021). As cepstral measures have shown links with perceived and physiological voice changes, they show promise as objective measures of dysphonia and voice disorders (Lowell et al., 2012; Peterson et al., 2013). CPP has generally shown good performance in studies detecting change in PD speech (Moya-Galé et al., 2022).

Based on the above, three acoustic parameters were selected against the perceptual features that were rated as most deviant in the SLT ratings and CPP as a general measure of voice quality to see if they are effective in tracking perceptual changes in PwPD speech over time.

4.2. Perceptual feature ratings (PFR) analysis

4.2.1. Participant demographics

The subset consisted of 52 speakers with speech recordings divided into three groups: PwPD-change, PwPD-no change, and control. The demographic information for each group can be seen in Table 5.

Table 5. Participant demographics for the PFR analysis.

Groups	Number of participants	Age range (years)	Mean age (years)
PwPD-change	N= 21; M = 15; F = 6	50-93	69 (SD = 10.46)
PwPD-no change	N= 21; M = 13; F = 8	56-84	71 (SD = 6.84)
Control	N= 10; M = 5; F = 5	51-82	70 (SD = 9.94)

4.2.2. Segmentation of recordings

The type of data recordings used for this analysis were the sustained phonation recordings to analyse jitter, shimmer and HNR. The reading passage recordings were used to analyse CPP.

The boundaries marking the beginning and end of phonation were marked for the sustained phonation recordings using Praat (Boersma & Weenink, 2012) for the parameters jitter, shimmer and HNR. If participants inhaled again and continued phonation after loss of breath, the boundary for the end of phonation was marked before this point.

All recordings of the reading passage needed to be segmented by syllable (for calculation CPP). All boundaries were created according to standard segmentation criteria based on guidelines in Kent and Read (2001).

4.2.3. Extracting parameters

Values for each acoustic parameter were manually extracted from Praat (Boersma & Weenink, 2020). This involved selecting the 'cross-correlation' method under 'Pitch' settings to optimise for voice analysis. Jitter and shimmer values were extracted based on Rusz et al. (2021b) as relative forms with the local values expressed as a percentage as they are shown not to be affected by sex-specific factors (Brockmann et al., 2011). The algorithms for jitter (local), and shimmer (local) were based on Praat's in-built algorithms for calculating the local values of jitter and shimmer as percentages. Praat's local values were selected because Praat uses waveform-matching, which uses a cross-correlation maximum by looking for the best match wave shape to determine the duration of a period. According to Boersma, (2009), the waveform-matching method is considered a good method for extracting jitter and shimmer values as it averages out the influence of additive noise which could be present in recordings of sustained phonation. The algorithm for HNR performs an acoustic periodicity detection on the basis of an accurate autocorrelation method (Boersma, 1993).

4.2.4. Reliability of acoustic annotations

To determine the reliability of the annotations, 10% of text grids were manually re-examined by the primary researcher and an external researcher trained in acoustic analysis. The text grids were randomly selected from either data collection point. Boundaries for the start and end of sustained phonation production were manually rechecked. The external researcher did not find any discrepancies in the annotations made.

Syllable boundaries for the reading passage (needed to exclude boundaries for the calculation of CPP) and beginning and end boundaries of sustained phonation production were manually rechecked.

4.2.5. Descriptive statistics

4.2.5.1. Jitter

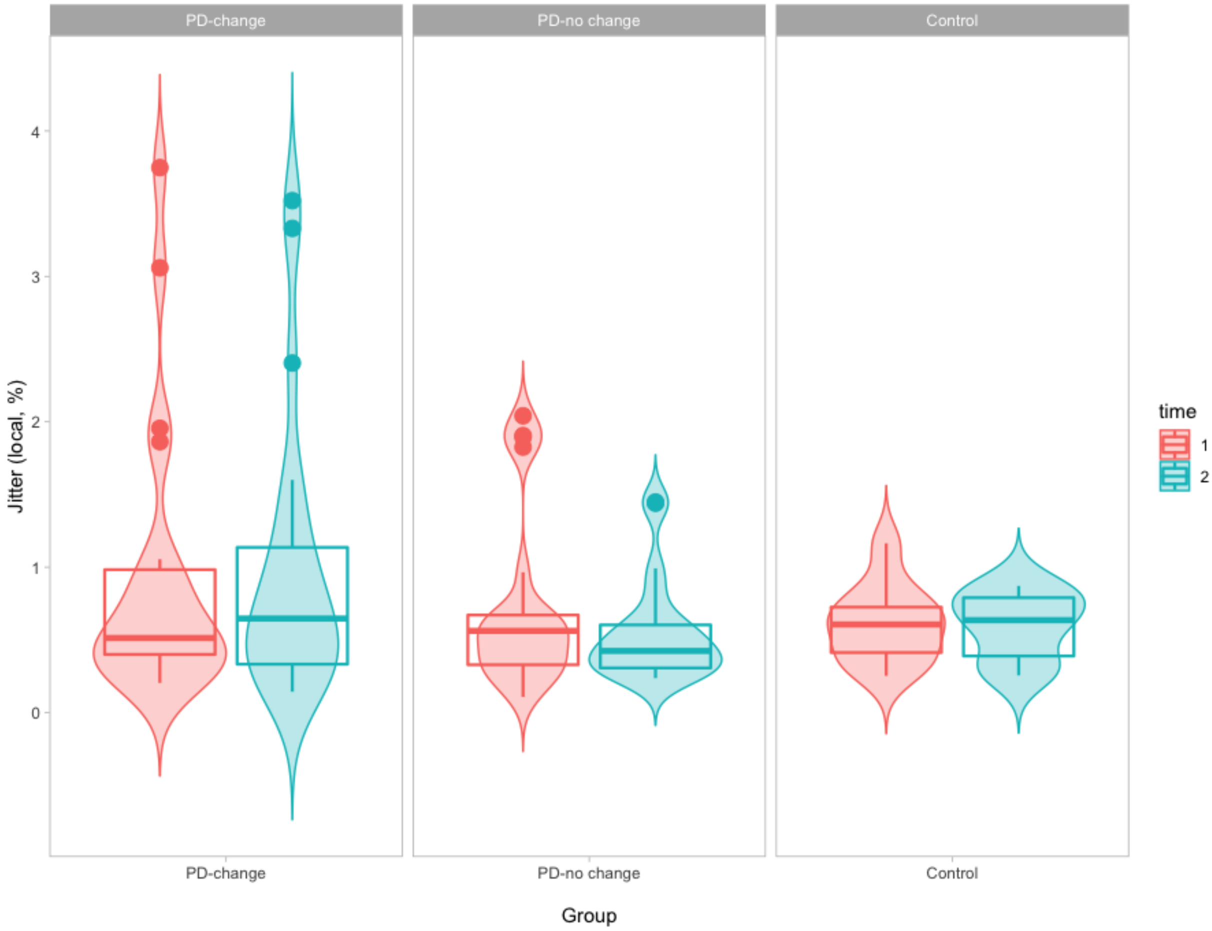
Descriptive statistics indicated that there were no extreme values among any of the groups or between T1 and T2 (see Table 6 below). The PwPD-change group's mean jitter value increased by 0.09 from T1 to T2, while the standard deviation (SD) increased by 0.02. There is a slight decrease in mean jitter from T1 to T2 (0.19) for the PwPD-no change as well, but the control group stays pretty stable.

Table 6. Summary table of descriptive statistics of jitter for each group at T1 and T2. N = 21 (PwPD-change); 21 (PwPD-no change); 10 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	0.61	0.27	0.61	0.25	1.16
Control (T2)	0.60	0.23	0.64	0.25	0.87
PwPD-change (T1)	0.93	0.96	0.51	0.20	3.75
PwPD-change (T2)	1.01	0.97	0.65	0.14	3.52
PwPD-no change (T1)	0.74	0.62	0.56	0.11	2.04
PwPD-no change (T2)	0.56	0.35	0.42	0.22	1.45

To get a holistic view of the descriptive statistics, as well as capture any extreme values, jitter was plotted against each group at both T1 and T2 compared in Figure 2 below. Visual inspection of the combined boxplot within violin plot shows that each group's jitter value distribution does not appear to change markedly between data collection T1 and T2. There is a greater change in distribution of jitter values within the PwPD-change group compared to the other groups. The potential outliers seen in the plot for the PwPD-change group explains the slightly higher mean jitter values compared to the other groups.

Figure 2. The distribution of jitter of each group (PwPD-change, PwPD-no change, Control) at T1 and T2.



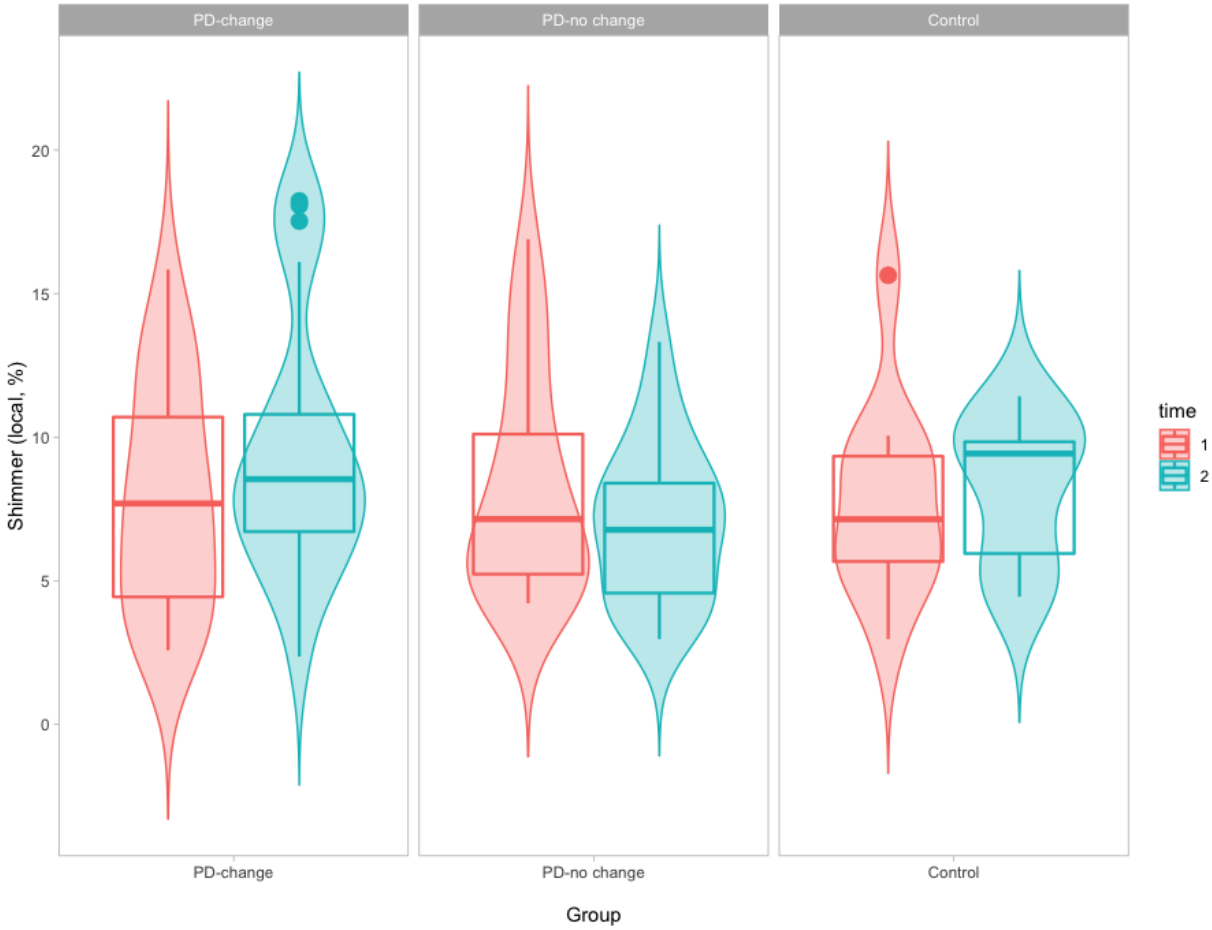
4.2.5.2. *Shimmer*

The summary table of descriptive statistics in Table 7 below shows that the distribution of shimmer is similar among all groups. The distribution of shimmer plotted against each group from T1 and T2 can be seen in Figure 3 below. The general distribution between groups is quite similar for T1. However, there is a greater change in mean shimmer values for the PwPD-change group from T1 to T2 (increased by 1.62) compared to the other groups. This can be seen in the higher distribution of the box plot, compared to other groups, despite the median of the control group being higher in T2.

Table 7. Summary table of descriptive statistics of shimmer for each group at T1 and T2. N = 21 (PwPD-change); 21 (PwPD-no change); 10 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	7.86	3.52	7.14	2.97	15.64
Control (T2)	8.26	2.57	9.43	4.45	11.43
PwPD-change (T1)	7.88	4.00	7.69	2.58	15.84
PwPD-change (T2)	9.50	4.54	8.54	2.37	18.23
PwPD-no change (T1)	8.25	3.89	7.15	4.22	16.89
PwPD-no change (T2)	7.03	2.76	6.77	2.97	13.32

Figure 3. The distribution of shimmer of each group (PwPD-change, PwPD-no change, Control) at T1 and T2.



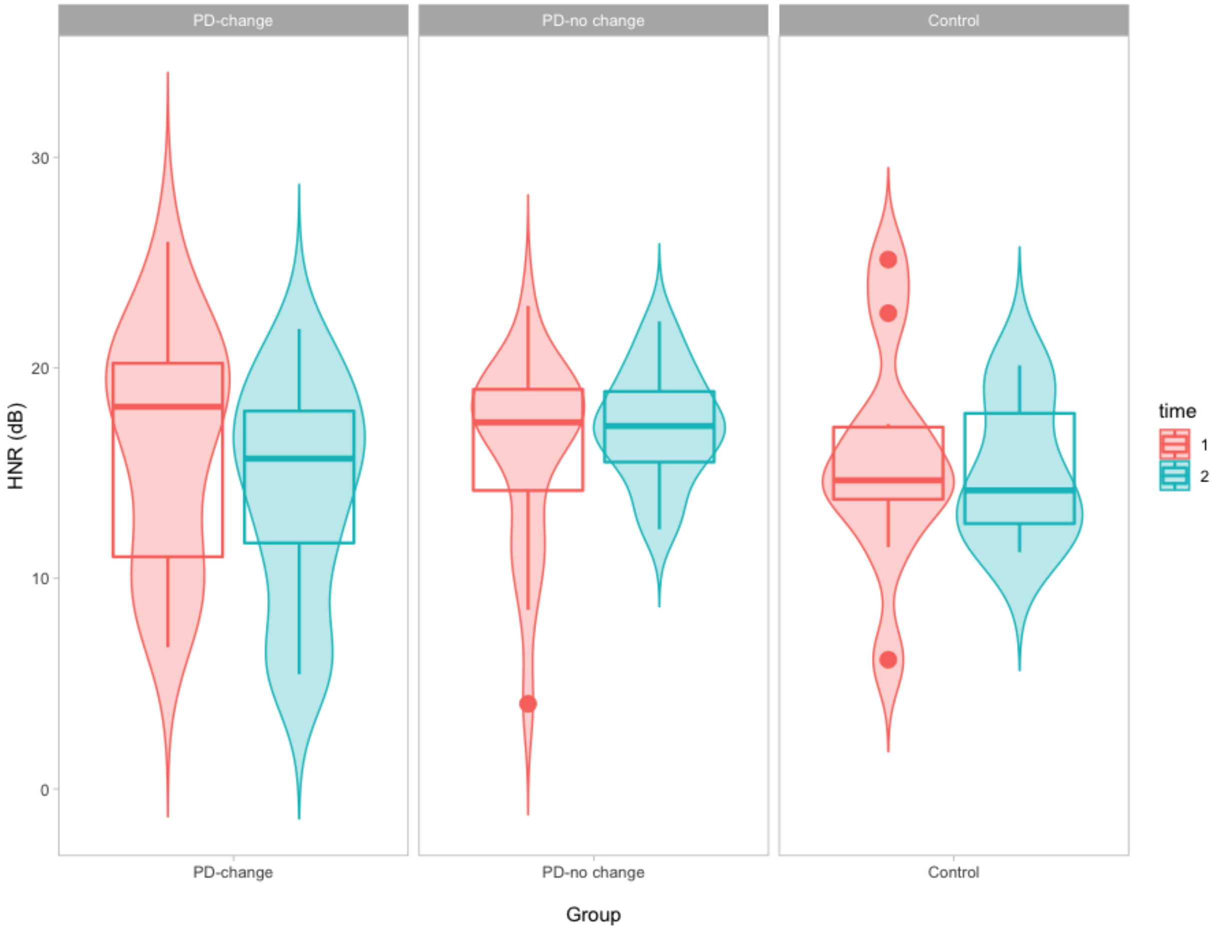
4.2.5.3. HNR

HNR values appear to change slightly from T1 toT2 with slight differences between both the PwPD-groups and the control group, with higher HNR values for the former at T1 (as seen in Table 8 below). There is a decrease in HNR values for the PwPD-change group form T1 to T2 and a slight increase in HNR for the PwPD-no change group. The control group doesn’t show much change between both data collection time points. This can be seen clearly in the combined boxplot within violin plot in Figure 4 below. Th distribution of HNR values is more spread for the PwPD-change group compared to the other groups and could indicate why the change in mean values from T1 to T2 is greater.

Table 8. Summary table of descriptive statistics of HNR for each group at T1 and T2. N = 21 (PwPD-change); 21 (PwPD-no change); 10 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	15.67	5.36	14.66	6.14	25.12
Control (T2)	15.07	3.30	14.18	11.25	20.12
PwPD-change (T1)	16.47	5.47	18.15	6.73	25.98
PwPD-change (T2)	14.01	5.18	15.69	5.45	21.84
PwPD-no change (T1)	16.12	4.47	17.41	4.04	22.94
PwPD-no change (T2)	17.25	2.65	17.24	12.33	22.20

Figure 4. The distribution of HNR of each group (PwPD-change, PwPD-no change, Control) at T1 and T2.



4.2.5.4. CPP

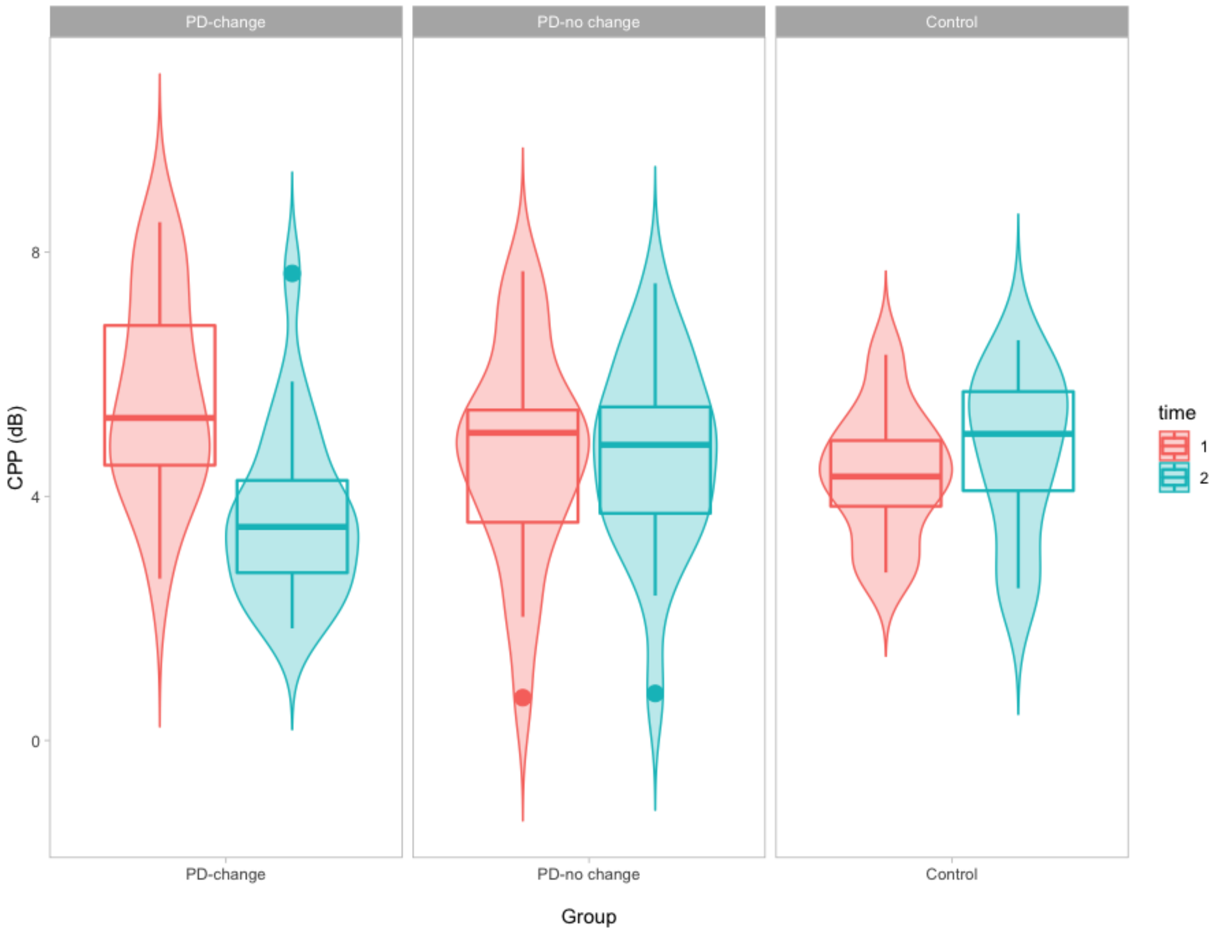
Based on the descriptive statistics in Table 9 below, the control and PwPD-no change groups have similar mean CPP values, while the PwPD-change group has a greater variance in values compared to the two groups. In addition, the PwPD-group has a greater change in mean CPP values from T1 to T2 compared to the other groups.

Table 9. Summary table of descriptive statistics of CPP for each group at T1 and T2. N = 21 (PwPD-change); 21 (PwPD-no change); 10 (control).

	Mean	SD	Median	Min	Max
Control (T1)	4.38	1.08	4.32	2.75	6.32
Control (T2)	4.74	1.37	5.02	2.49	6.56
PwPD-change (T1)	5.64	1.65	5.28	2.65	8.49
PwPD-change (T2)	3.67	1.43	3.50	1.84	7.65
PwPD-no change (T1)	4.67	1.74	5.04	0.70	7.68
PwPD-no change (T2)	4.63	1.57	4.84	0.77	7.49

In Figure 5 below, there appears to be a larger difference in CPP distribution for both the PwPD-groups compared to the control group which can be confirmed by the Min and Max values in the table above. In addition, CPP values plotted against group suggests that there is a change from T1 and T2 data collection points for the groups. However, the change appears to be greater with the PwPD-change group than others.

Figure 5. The distribution of CPP of each group (PwPD-change, PwPD-no change, Control) at T1 and T2.



4.2.6. Results of LMMs

To test the significance of each fixed effect on each acoustic parameter (jitter, shimmer, HNR, and CPP), a linear mixed effects model (LMM) analysis was conducted using R (R Core Team, 2019). The models were run with the lme4 package (Bates et al., 2015) using lmer.

The goodness-of-fit for each model was checked by adding any possible interactions between fixed effects and any slopes between all factors. When creating exploratory models, it is considered good practice to build in complexity variable-by-variable from a simple intercept-only model (Field et al., 2012). Convention dictates that each variable is added in its own model, and subsequent models are compared on observed decreases in the Akaike information criterion (AIC). When a variable is added, its contributions to the model are examined and this process continues until the addition of a variable causes the AIC to increase. In this way, the variables' ability to predict the outcome variable (in the present study, the various acoustic parameters) is used as a criterion about whether it should be included in the final model. The best fit was found to only contain each acoustic parameter tested against time and group as predictors (or fixed effects) and participant as the random factor. Optimisers were used to ensure that over-fitting did not occur.

A model was run with each acoustic parameter (outcome variable) against the predictors time (data collection time points T1 and T2) and group (PwPD-change, PwPD-no change, and control) with the random factor of participant including a random intercept and random slope. A random intercept means that the final model takes into account that each individual *participant* can show higher or lower values for each acoustic parameter regardless of *group* or the data collection *time*. A random slope for *participant* means that the model considers and accounts for some between-participant variation.

Based on the assumptions of an LMM, the outcome variable does not need to be normally distributed if the residuals of each model are normally distributed. The residuals for each model run on each acoustic parameter (outcome variable) were checked and did not violate the assumption of normality. Therefore, none of the data required any transformations. There were no significant correlations between any of the predictors, and therefore was not reported.

LMMs create an intercept in each model representing each acoustic parameter against T1 data collection and the control group. The intercept represents how different the outcome variable and predictors are from chance. The other predictors are compared against this intercept model, and results are reported based on how different other predictors are from the intercept.

A summary table of the results of the LMM for each acoustic parameter of voice quality can be seen in Table 10, based on 104 observations from 52 participants.

Results of the LMMs on jitter, shimmer, and HNR indicated that the effects of the predictors *time* (T1 and T2) and *group* (PwPD-change, PwPD-no change, and control) on these acoustic parameters were not significant. This implies that jitter, shimmer, and HNR values were not significantly different between any of the groups and did not significantly change between T1 and T2.

The results of the LMM on CPP indicated that the effects of the predictor *time* had a significant impact (β= -0.743, 95% CI [-1.160, -0.325], $p < 0.001$), implying there was a significant change between T1 and T2, with CPP an estimated 0.74dB lower at T2 than at T1 which indicates overall dysphonia increased from T1 to T2. However, the predictor *group* did not have a significant impact on CPP, implying that CPP did not change in a meaningful way between groups. The results show that there was a change in CPP between T1 and T2 across all groups. Revisiting the plot in Figure 5 above shows that the most change in CPP from T1 to T2 is observed in the PwPD-change group, compared to the control and PwPD-no change groups. This could be the cause for skewing the overall change in CPP values from T1 to T2, and it is important to note where the most change in CPP lies. It does indicate that since the PwPD-change group may have had the most changes in CPP from T1 and T2 that it agrees with the results of the SLT ratings.

Table 10. Linear mixed effects model results of the PFR analysis of jitter, shimmer, HNR, and CPP against each of the levels of the predictors time and group.

Fixed Effects:						
Acoustic parameter	**Predictor**	**Estimate**	**Std. Error**	**df**	**t-value**	**p**
Jitter	Intercept (Control, T1)	0.63	0.21	51.66	2.97	<0.01*
	T2	-0.04	0.07	51.00	-0.63	0.54
	PwPD-change	0.36	0.25	49.00	1.43	0.16
	PwPD-no change	0.04	0.25	49.00	0.16	0.87
Shimmer	Intercept (Control, T1)	7.94	1.03	57.61	7.73	<0.01*
	T2	0.24	0.59	51.00	0.41	0.68
	PwPD-change	0.63	1.19	49.00	0.53	0.59
	PwPD-no change	-0.42	1.19	49.00	-0.35	0.73
HNR	Intercept (Control, T1)	15.69	1.27	56.67	12.32	<0.01*
	T2	-0.65	0.69	51.00	-0.95	0.35
	PwPD-change	-0.13	1.49	49.00	-0.09	0.93
	PwPD-no change	1.32	1.49	49.00	0.89	0.38
CPP	Intercept (Control, T1)	4.93	0.46	54.46	10.65	<0.01*
	T2	-0.74	0.21	51.00	-3.49	0.001**
	PwPD-change	0.09	0.55	49.00	0.17	0.87
	PwPD-no change	0.09	0.55	49.00	0.16	0.87

4.2.7. Discussion

The PFR analysis was conducted to investigate the effectiveness of the acoustic parameters jitter, shimmer, HNR, and CPP, at capturing perceptual changes in PwPD speech over time. This was done by conducting acoustic and statistical analyses (LMMs) on speech data collected over two time points, six months apart, from the same group of participants. The models for the LMMs were run with each acoustic parameter (the outcome variable) modelled against the predictors *time* (indicating the data collection time) and *group* (indicating the grouping of recordings into either PwPD-change, PwPD-no change, and control).

The results of statistical analysis revealed that the acoustic parameters jitter, shimmer and HNR were unable to significantly track perceptual changes from T1 to T2 and were unable to significantly distinguish between groups either. A potential explanation for the lack of significant results for these acoustic parameters could be: if vocal fold weakness in both the PwPD groups is only minor, it would suggest only mild speech impairment, which would not be severe or distinct enough from control speech, explaining the lack of significant group differences in jitter, shimmer, or HNR values. Therefore, perceptual changes in voice quality would also be minor, explaining a lack of significant change over time. As mentioned in section 4.1 of this chapter, incomplete vocal fold closure is commonly observed in PwPD (Rusz, Tykalová, et al., 2021; Theodoros & Ramig, 2011), and the lack of significant results of the effect of both the predictor variables of *group* and *time* on each of the acoustic parameters may suggest that the laryngeal mechanism among the PwPD participants in this analysis may not be as severely impacted and vocal fold weakness not severe enough to cause increased turbulence in the vocal tract. However, this does not coincide with the average rating of 3 and 4 ('marked' and severe') given by the two SLTs for the perceptual features of breathy, hoarse, and strain-strangled, which correspond to the acoustic parameters jitter, shimmer and HNR. This could suggest that these acoustic parameters may not be sufficient for capturing the perceptual changes indicated in the SLT ratings.

A potential lack of significant group differences in jitter, shimmer, HNR and CPP valuers in the PFR analysis is the possibility that tracking specific perceptual features of voice quality is challenging in the early stages of impairment and that jitter, shimmer, HNR and CPP are better suited for detecting group differences in studies involving PwPD with greater speech impairments. A related impact of this could be that specific acoustic parameters may only be able to effectively capture mild speech impairments in specific speech categories rather than others.

Another explanation for the lack of significant effect of the predictor *time* on the parameters jitter, shimmer and HNR could be the way the PwPD recordings were selected for analysis. As detailed in chapter 3, SLTs rated each speech category, and the perceptual features rated higher for severity (either ‘marked’ or ‘severe’) were selected for the first analysis. Recordings were included in the PwPD-change group if the rating of the highly rated perceptual features had changed (given a higher or lower rating) for the second data collection point compared to the first. For example, if the perceptual features breathy and hoarse were rated 3 (‘marked’) for a PwPD speaker for data collection one and rated 4 (‘severe’) for data collection two, then this participant’s recordings were included in the PwPD-change group for analysis. If ratings stayed the same between both data collection points, then recordings for participants were included in the PwPD-no change group. This grouping, based on a change detected in the SLT ratings, may be quite subjective, given the psychometric properties of the rating scale. While each speech category was rated on different days, all perceptual features within each category (such as all the perceptual features within voice quality) were rated at the same time and since these perceptual features are correlated (Kent, 1996; Oates, 2009) it may have had a bearing on the rating given to each perceptual feature.

A final explanation for the lack of significant group differences and significant perceptual change over time in jitter, shimmer, and HNR values could be due to the size of the dataset. While there were 42 PwPD in total, they were divided into two groups and compared against only 10 controls. This reduced the overall statistical power of the models and may also explain the lack of significant results for *time*.

The acoustic parameter CPP showed a significant effect of the predictor *time*, but no significant effect of the predictor *group*. The significant effect of *time* indicates that CPP values significantly decreased from T1 to T2, which is interpreted as an increase in dysphonia in PwPD (Murton et al., 2020; Šimek & Rusz, 2021). This is consistent with previous work suggesting CPP may be able to pick up on speech changes (Šimek & Rusz, 2021). The lack of group differences in CPP values is consistent with the jitter, shimmer, and HNR results as well and might be due to the size of the dataset as well.

The significant change in only CPP values over time may indicate that acoustic parameters that attempt to capture perceptual changes over time are better at doing so if they capture more general measures of voice quality, such as dysphonia, rather than specific perceptual features, such as hoarse, breathy, and strain-strangled that are often correlated features (Kent, 1996; Oates, 2009) and therefore may be hard to capture individually using jitter, shimmer, and HNR. In addition, jitter, shimmer, and HNR are extracted from sustained phonation, which relies on fundamental frequency computations which may not be able to capture perceptual changes in voice quality that is has more has moderate dysphonia (Murton et al., 2020). This is an important observation as even though SLT ratings individually rated certain speech features as ‘marked’ or ‘severe’, the perceptual change may be better captured by CPP, which was extracted from the grandfather passage and does not rely on fundamental frequency computation. It is important to bear in mind that not all acoustic parameters are able to capture perceptual changes as holistically as others, necessitating the need for this study and others that create a distinction between global acoustic speech markers of PD and those that are able to track speech change.

The results of this analysis may also suggest that two time points may not have been sufficient for jitter, shimmer, and HNR to capture perceptual changes. However, it must also be noted that some individual participants may have had greater perceptual changes than others and could have averaged out the results. This study did not focus on individual differences in perceptual changes in voice quality and, therefore cannot conclude with certainty if the results are due to the inability of the acoustic parameters jitter, shimmer, and HNR to capture perceptual changes in PwPD speech over time. It does suggest that further investigation would be warranted and should attempt to include a larger dataset, more data collection time points or a more extended period between data collection times, and a look at whether there are distinct individual differences in the results.

4.3. Intelligibility group (IG) analysis

4.3.1. Participant demographics

Analysis was conducted on the complete dataset (n =110) to test for a change in PwPD speech intelligibility. The recordings were grouped based on the overall intelligibility rating based on the SLT ratings and grouped PwPD speech into mild (rated 0 or 1), moderate (rated 2 or 3), and severe (rated 4). All the PwPD recordings were found to be rated as mild, indicating only mild impairment to speech intelligibility. In addition, the overall intelligibility rating was checked for both T1 and T2 data collection time points and grouped based on whether the SLT rating had changed (either increased or decreased) in T2. The final groups for this analysis can be seen in Table 11 below:

Table 11. Participant demographics for the IG analysis.

Group	**Number of participants**	**Age range (years)**	**Mean age (years)**
Mild-change	N= 34; M = 22; F = 12	52-81	66 (SD = 7.09)
Mild-no change	N= 29; M = 21; F = 8	50-93	72 (SD = 9.16)
Control	N= 47; M = 17; F = 30	35-86	64 (SD = 12.23)

4.3.2. Segmentation of recordings

The type of data recordings used for this analysis was the same as the first analysis. The sustained phonation recordings were used to analyse jitter, shimmer and HNR. The reading passage recordings were used to analyse CPP.

Boundaries for the sustained phonation were marked using the same method employed for the perceptual feature ratings analysis. All recordings of the reading passage needed to be segmented for by syllable (for calculation CPP) in Praat (Boersma & Weenink, 2012). All boundaries were created according to standard segmentation criteria based on guidelines in Kent and Read (2001). Extraction of jitter, shimmer HNR and CPP for the rest of the dataset was as previously outlined. CPP was extracted using the 'To PowerCepstrogram' option in 'Pitch analyses'. CPP was extracted with a parabolic interpolation and the fit method 'robust'.

4.3.3. Reliability of annotations

In order to determine the reliability of the annotations, 10% of textgrids were manually re-examined by the primary researcher and the same external researcher was used for the subset. Boundaries for the start and end of sustained phonation production were manually rechecked. The researcher did not find any discrepancies in the annotations made.

Syllable boundaries for the reading passage (needed to exclude boundaries for the calculation of CPP) and beginning and end boundaries of sustained phonation production were manually rechecked as done for the subset. Both researchers agreed with most of the syllable boundaries made, and any discrepancies were manually corrected.

4.3.4. Descriptive statistics

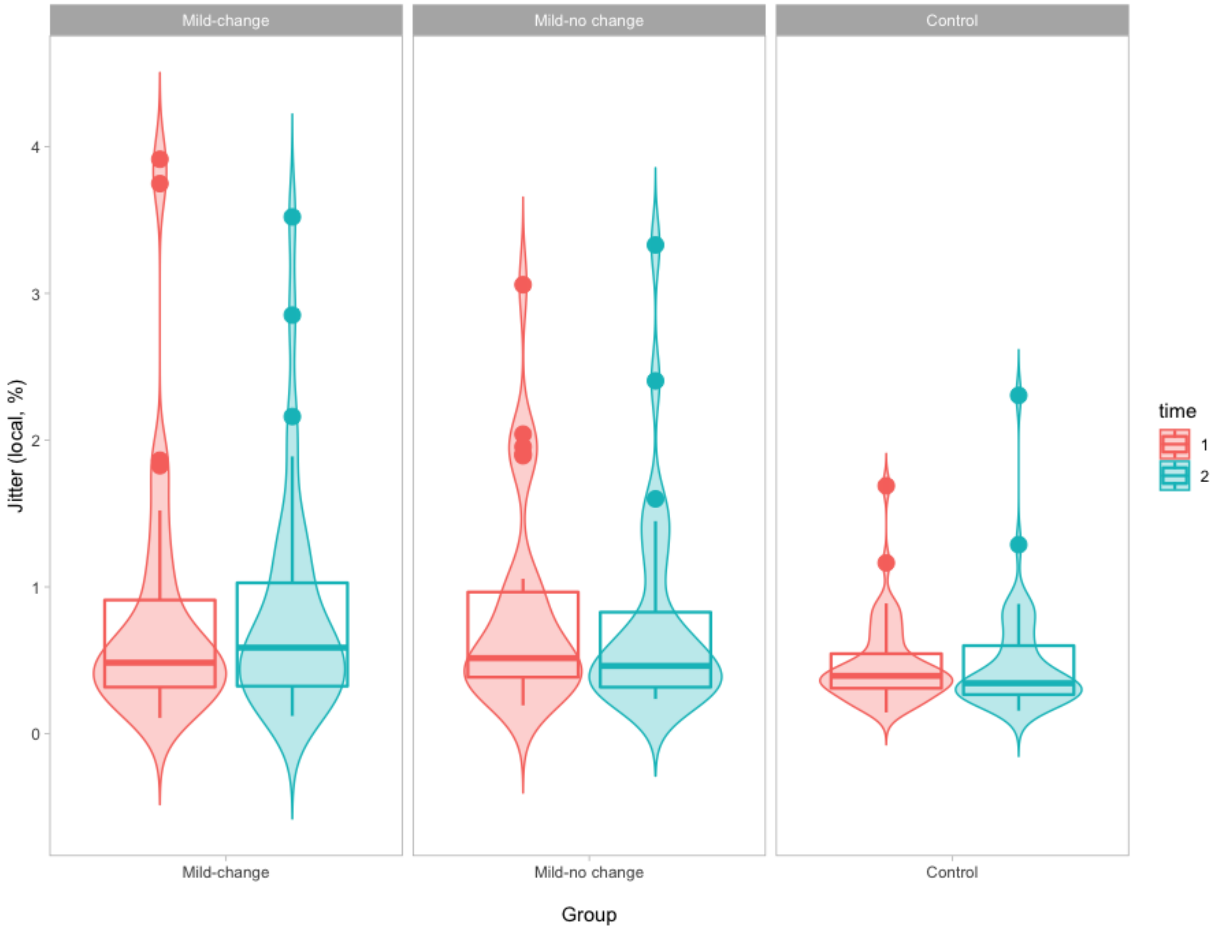
4.3.4.1. Jitter

The summary of the descriptive statistics in Table 12 below shows that there is a distinct difference in the mean jitter values between with the control group and the two PwPD mild groups. However, there are similar jitter values between the two mild PwPD groups. There is a decrease in the mean jitter values for the mild-no change group from T1 to T2, while the mild-change and control groups' jitter values remain similar at both data collection time points. This can be seen in the combined boxplot within violin plot in Figure 6 of just the jitter values plotted against each group from T1 to T2 below.

Table 12. Summary table of descriptive statistics of jitter for each group at T1 and T2. N = 34 (mild-change); 29 (mild-no change); 47 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	0.47	0.28	0.39	0.14	1.69
Control (T2)	0.48	0.37	0.34	0.16	2.31
Mild-change (T1)	0.82	0.89	0.48	0.11	3.91
Mild-change (T2)	0.85	0.78	0.58	0.12	3.52
Mild-no change (T1)	0.82	0.69	0.52	0.19	3.06
Mild-no change (T2)	0.76	0.71	0.46	0.24	3.33

Figure 6. The distribution of jitter of each intelligibility group (Mild-change, Mild-no change, Control) at T1 and T2.



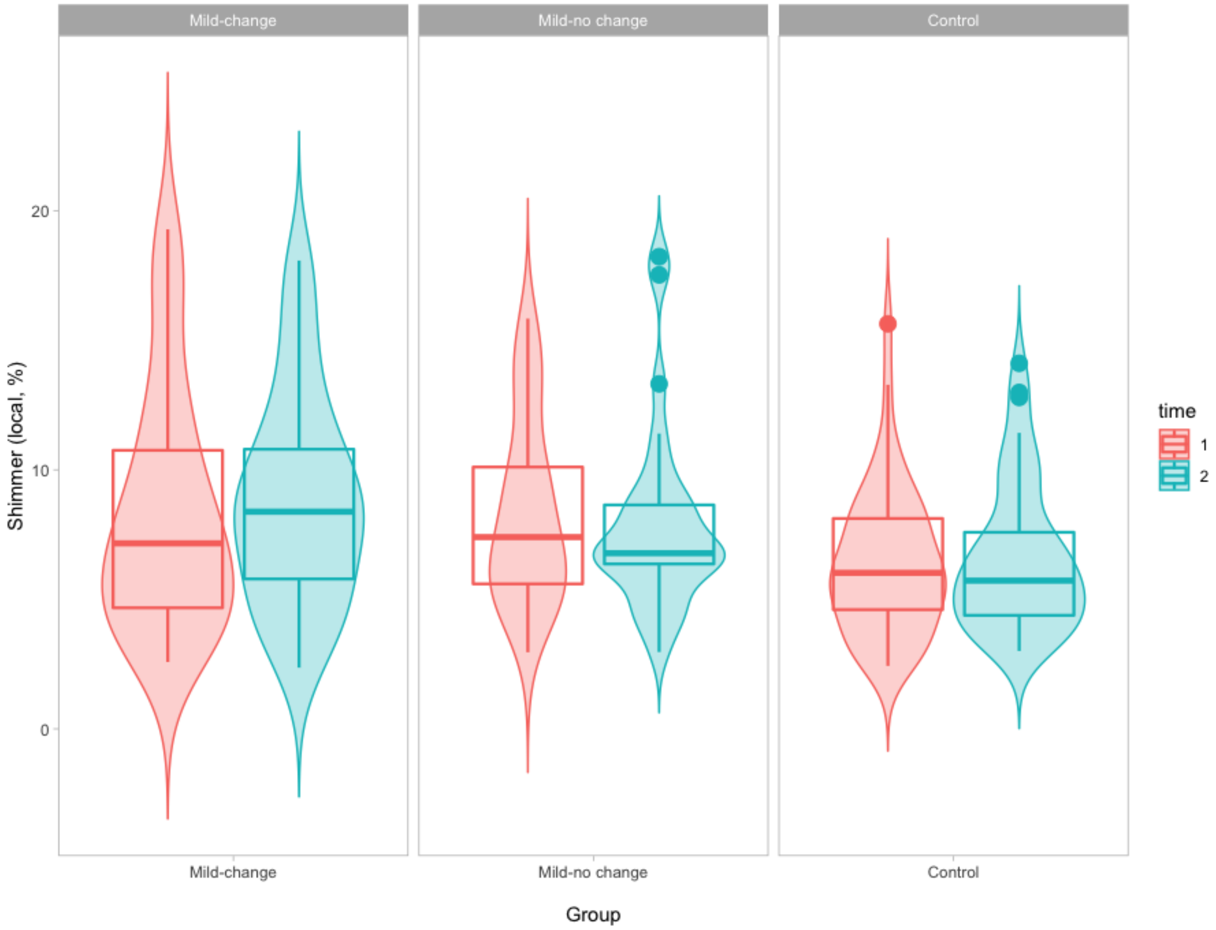
4.3.4.2. Shimmer

Descriptive statistics of shimmer values seen in Table 13 below shows a group difference between the control and both PwPD mild groups, similar to jitter. In addition, there is a similar decrease in the mean shimmer for the mild-no change group from T1 to T2 (-0.42) compared to the other groups. The distribution of shimmer seen in Figure 7 shows a wider distribution and less extreme values than jitter, confirmed by the mean shimmer values being closer to each other for all three groups.

Table 13. Summary table of descriptive statistics of shimmer for each group at T1 and T2. N = 34 (mild-change); 29 (mild-no change); 47 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	6.38	2.81	6.02	2.43	15.64
Control (T2)	6.57	2.83	5.72	3.01	14.11
Mild-change (T1)	8.32	4.68	7.16	2.58	19.28
Mild-change (T2)	8.78	4.14	8.39	2.37	18.08
Mild-no change (T1)	8.30	3.57	7.40	2.96	15.84
Mild-no change (T2)	7.88	3.54	6.78	2.97	18.23

Figure 7. The distribution of shimmer of each intelligibility group (Mild-change, Mild-no change, Control) at T1 and T2.



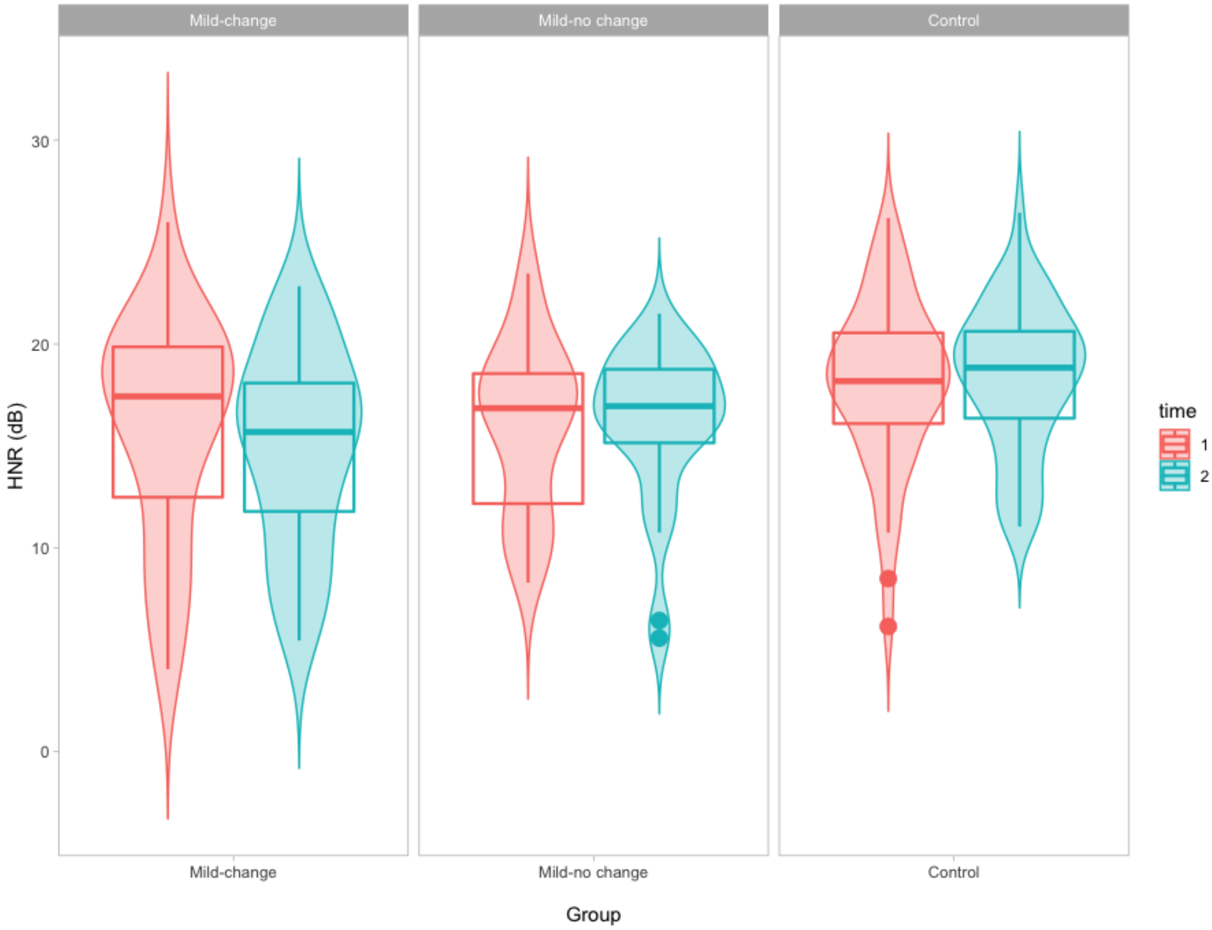
4.3.4.3. HNR

In Table 14 below, showing the descriptive statistics for HNR, the mean values for the mild-change group are distinct from both the control and the mild-no change group. The former appears to change more from T1 to T2 (decreases by 0.98) than the latter two groups as well. The boxplot within the violin plot in Figure 8 shows this decrease in HNR values in the mild-change group, but the plot appears to suggest that the HNR values for all three groups are fairly similar from T1 to T2. The distribution of HNR values for the mild-group is more spread than the other groups, which have a more central distribution.

Table 14. Summary table of descriptive statistics of HNR for each group at T1 and T2. N = 34 (mild-change); 29 (mild-no change); 47 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	18.06	4.21	18.18	6.14	26.18
Control (T2)	18.39	3.64	18.84	11.04	26.45
Mild-change (T1)	15.89	5.52	17.43	4.04	25.98
Mild-change (T2)	14.90	4.92	15.68	5.45	22.84
Mild-no change (T1)	16.03	4.14	16.85	8.28	23.45
Mild-no change (T2)	16.13	3.81	16.95	5.57	21.49

Figure 8. The distribution of HNR of each intelligibility group (Mild-change, Mild-no change, Control) at T1 and T2.



4.3.4.4. CPP

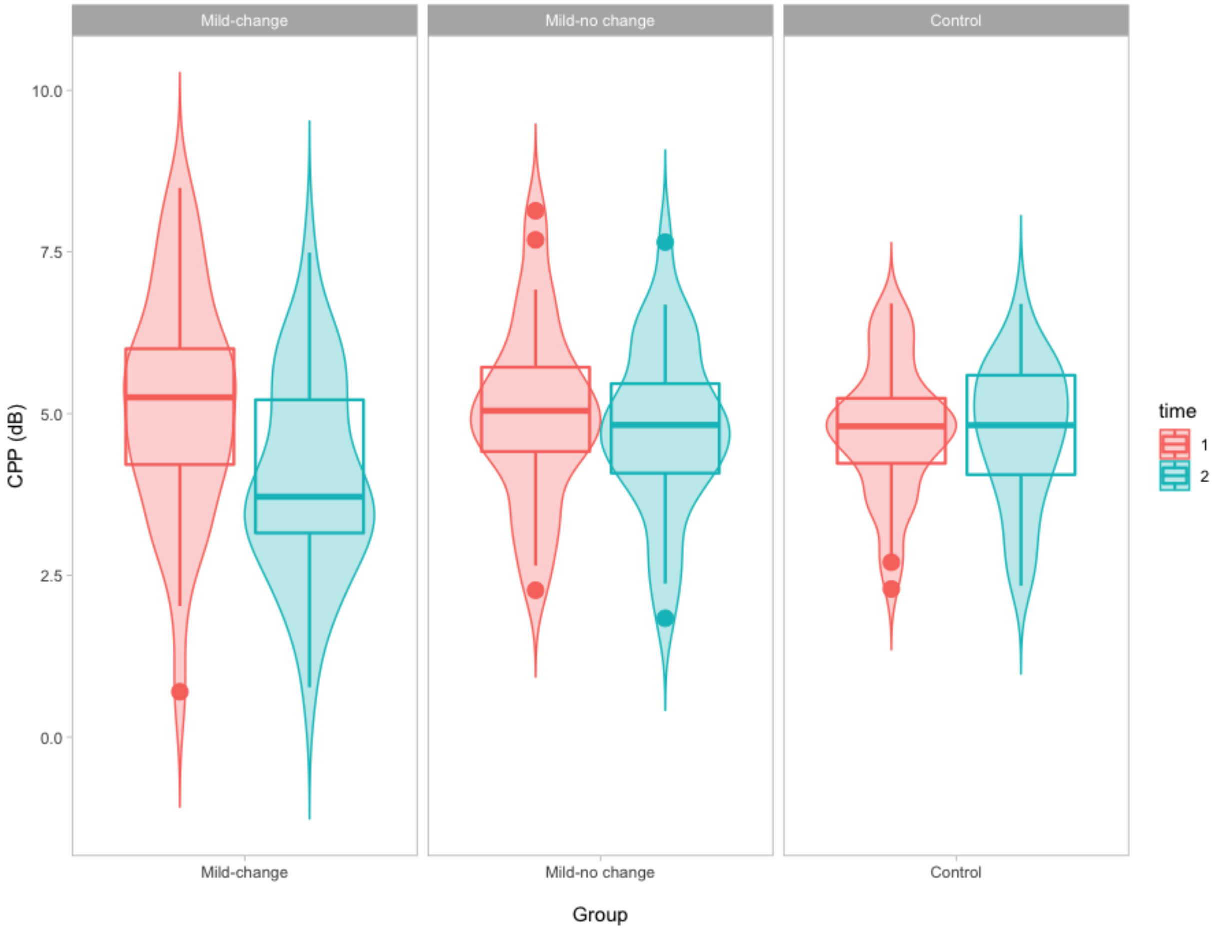
The descriptive statistics in

Table 15 for CPP below shows that the mean values for the mild-change group are distinct from both the control and the mild-no change group and appear to change more from T1 to T2 than the other two groups as well. The boxplot within the violin plot in Figure 9 illustrates the change in CPP values from T1 to T2 in the mild-change group well. However, the group difference and change over time may not be statistically significant.

Table 15. Summary table of descriptive statistics of CPP for each group at T1 and T2. N = 34 (mild-change); 29 (mild-no change); 47 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	4.74	1.04	4.81	2.29	6.70
Control (T2)	4.69	1.09	4.82	2.34	6.70
Mild-change (T1)	5.12	1.75	5.25	0.70	8.49
Mild-change (T2)	3.99	1.52	3.72	0.77	7.49
Mild-no change (T1)	4.99	1.40	5.05	2.27	8.14
Mild-no change (T2)	4.76	1.32	4.83	1.84	7.65

Figure 9. The distribution of CPP of each intelligibility group (Mild-change, Mild-no change, Control) at T1 and T2.



4.3.5. Results of LMMs

A linear mixed effects model analysis was run through R (R Core Team, 2019) to test the significance of each predictor on each acoustic parameter. The models were run with the lme4 package (Bates et al., 2015) using lmer. The model structures were the same as those used for the PFR analysis.

A model was run with each acoustic parameter (outcome variable) against the predictors *time* (data collection time points T1 and T2) and *intelligibility group* (mild-change, mild-no change, and control) and random factor of *participant* with a random intercept and random slope. A random intercept means that the final model takes into account that each individual *participant* can show higher or lower values for each acoustic parameter regardless of their *intelligibility group* or the data collection *time*. A random slope for *participant* means that the model considers that some between-participant variation exists.

Based on the assumptions of a linear mixed effects model, the residuals for each model run for each acoustic parameter were checked and found to be normally distributed. Therefore, no transformations of the parameters data were required. There were no significant correlations between any of the predictors.

A summary table of the results of the LMM for all the acoustic parameters can be seen in Table 16 below, based on 220 observations from 110 participants.

Results of the LMMs on jitter, shimmer, and HNR indicated that the effects of the predictor *time* (T1 and T2) on these acoustic parameters were not significant. This implies that jitter, shimmer, and HNR values were not significantly different for any of the groups from T1 and T2. However, jitter, shimmer, and HNR did have a significant impact on the effect of *group* (mild-change, mild-no change, and control), indicating that the jitter, shimmer, and HNR values for each group were distinct from each other regardless of the data collection time point. Jitter and shimmer values for both the PwPD groups being higher than the control group indicates higher perturbation and amplitude, and HNR values for both the PwPD groups were lower than the control group indicating more breathiness, hoarseness and strain-strangled quality to voice in PwPD (Rusz et al., 2011; Teixeira et al., 2013).

The results of the LMM on CPP indicated that the predictor time had a significant impact (β= -0.434, 95% CI [-0.671, -0.197], SE= 0.121, $p < 0.05$), implying there was a significant change between T1 and T2, with CPP an estimated 0.43dB lower at T2 than at T1 indicating that overall dysphonia increased from T1 to T2. However, the predictor group did not have a significant impact on CPP, implying that CPP did not change in a meaningful way between groups. The results show that there was a change in CPP between T1 and T2 occurred across all groups. Revisiting the plot in

Table 15 above shows that the most change in CPP from T1 to T2 is observed in the mild-change group, compared to the mild-no change group and control group, similar to the CPP results in the PFR analysis. This could be the cause for skewing the overall change in CPP values from T1 to T2, and it is important to note where the most change in CPP lies.

Table 16. Linear mixed effects model results of the IG analysis of jitter, shimmer, HNR, and CPP against each of the levels of the predictors time and group.

Fixed Effects:						
Acoustic parameter	**Predictor**	**Estimate**	**Std. Error**	**df**	**t-value**	**p**
Jitter	Intercept (Control, T1)	0.48	0.09	130.04	5.56	<0.01*
	T2	-0.002	0.05	109.00	-0.03	0.97
	mild-change	0.36	0.13	107.00	2.85	<0.01**
	mild-no change	0.32	0.13	107.00	2.41	0.02*
Shimmer	Intercept (Control, T1)	6.42	0.47	146.59	13.61	< 0.01*
	T2	0.11	0.38	109.00	0.28	0.78
	mild-change	2.08	0.67	107.00	3.12	<0.01*
	mild-no change	1.62	0.70	107.00	2.32	0.02*
HNR	Intercept (Control, T1)	18.29	0.59	140.68	31.15	< 0.01*
	T2	-0.14	0.44	109.00	-0.31	0.76
	mild-change	-2.83	0.84	107.00	-3.37	0.001**
	mild-no change	-2.15	0.88	107.00	-2.43	0.02*
CPP	Intercept (Control, T1)	4.93	0.19	132.12	26.57	< 0.01*
	T2	-0.43	0.12	109.00	-3.58	0.001**
	mild-change	-0.16	0.27	107.00	-0.59	0.56
	mild-no change	0.16	0.28	107.00	0.58	0.57

4.3.6. Post-Hoc Testing

Estimated marginal means (EMMs) were obtained using the "emmeans" package (Lenth, Russel V., 2022) in R (R Core Team, 2019). The package used Kenward-roger method to calculate degrees of freedom. EMMs were run to conduct a pairwise comparison of all three *groups* and the results automatically averaged (collapsed) over the two levels of the predictor *time* with a confidence level of 95%, giving the mean response value of each level of the predictor *group*, and contrasting it with the other levels. This would help identify how different each group is from the other. A pairwise comparison of the two levels of *time* was also done, averaged over the predictor *group.*

Results of the EMMs for the relevant acoustic parameters are discussed below.

4.3.6.1. Jitter

Since the significant predictor for jitter was *group*, the EMMs were run to conduct a pairwise comparison of all three groups, and the results automatically averaged over the two levels of the predictor *time* with a confidence level of 95%.

As seen from Table 17 below, there is a significant difference between the control and mild-change group when averaged over the predictor *time* (T1 and T2) (E= -0.347, df=107, p <0.05) and between the control and mild-no change group (E= -0.316, df=107, p <0.05). However, there is no significant group difference between the mild-change and mild no-change group (E=0.041, df=107, p >0.05).

Table 17. Pairwise differences of the predictors group and time for jitter.

Group comparisons	**Estimate**	**SE**	**df**	**t-ratio**	**p**
Control – Mild-change	-0.36	0.13	107	-2.85	0.01*
Control – Mild-no change	-0.32	0.13	107	-2.41	0.05*

Mild-change – Mild-no change	0.04	0.14	107	0.29	0.95

Note: SE = standard error; df = degrees of freedom.

4.3.6.2. Shimmer

EMMs for shimmer were also run for the predictor *group* in using the same method as for jitter.

As seen from Table 18 below, there is a significant difference between the control and mild-change group at both T1 and T2 (E= -2.078, df=107, p <0.01). However, there no significant group difference between the control and mild-no change group (E= -1.620, df=107, p >0.05), and the mild-change and mild no-change group (E=0.458, df=107, p >0.05).

Table 18. Pairwise differences of the predictors group and time for shimmer.

Group comparisons	**Estimate**	**SE**	**df**	**t-ratio**	**p**
Control – Mild-change	-2.08	0.67	107	-3.12	<0.01*
Control – Mild-no change	-1.62	0.70	107	-2.32	0.06
Mild-change – Mild-no change	0.46	0.75	107	0.62	0.81

Note: SE = standard error; df = degrees of freedom.

4.3.6.3. HNR

HNR EMMs were obtained for the predictor *group* average over the two levels of the predictor *time*.

As seen from Table 19 below, there is a significant difference between the control and mild-change group averaged over the predictor *time* (both T1 and T2) (E= 2.834, df=107, $p < 0.01$), and between the control and mild-no change group (E= 2.146, df=107, $p < 0.05$). However, there is no significant group difference between the mild-change and mild no-change group (E=-0.687, df=107, $p > 0.05$).

Table 19. Pairwise differences of the predictors group and time for HNR.

Group comparisons	**Estimate**	**SE**	**df**	**t-ratio**	**p**
Control – Mild-change	2.834	0.841	107	3.369	0.003**
Control – Mild-no change	2.146	0.882	107	2.432	0.044*
Mild-change – Mild-no change	-0.687	0.944	107	-0.728	0.7475

Note: SE = standard error; df = degrees of freedom.

4.3.7. Discussion

This section presented the results of the IG analysis conducted on the complete dataset of participants, analysing the recordings of sustained phonation and the grandfather passage. The acoustic parameters jitter, shimmer, HNR and CPP were all used to investigate this, similar to the PFR analysis. Since the grouping of the recordings was based on the overall intelligibility ratings given by the SLTs, all participant recordings were included.

The models for the LMMs were run with each acoustic parameter modelled against the predictor *time*, indicating the data collection time points T1 and T2, and *group*, indicating the grouping of recordings into either mild-change, mild-no change, and control. Results indicated that for the acoustic parameters jitter, shimmer, and HNR, there was a significant effect of the predictor *group* but no significant effect for the predictor *time*. However, for CPP, there was a significant effect for the predictor *time*, but no significant effect for the predictor *group*. These results indicate that the acoustic parameters jitter, shimmer, and HNR were able to successfully distinguish between different intelligibility groups regardless of the data collection time point (T1 or T2).

Although there was a significant impact of the predictor *group* on jitter, shimmer and HNR, post-hoc testing revealed that the differences lied between the control group and the two PwPD groups but were not significant between the mild-change and mild-no change groups. This could be due to both PwPD groups having mild speech symptoms (based on the overall intelligibility ratings) and, therefore any differences between them not being distinct enough from each other. This is not surprising as the objective was to investigate if each acoustic parameter could distinguish between different intelligibility groups. Since both the PwPD groups in the dataset used for the IG analysis were categorised with the same intelligibility (mild), it would be unexpected for a significant group difference to be found between the mild-change and mild-no change group. Voice characteristics in dysarthria can be hard to distinguish, especially when symptoms are mild or when PwPD are in the early stages of dysarthria/ PD (Rusz et al., 2021). However, the result of a significant effect of *group* confirms findings in previous literature that the acoustic parameters jitter, shimmer and HNR can distinguish between control and PwPD speakers (Awan et al., 2016; Jiménez-Jiménez et al., 1997; Patel et al., 2018; Rusz et al., 2021a,b; Vizza et al., 2019).

The lack of significant effect for the predictor *time* on jitter, shimmer and HNR can be interpreted in a few ways: a) jitter, shimmer and HNR are simply not sensitive enough to detect an acoustic change that might otherwise be audibly perceivable. While this interpretation is a possibility, it is more likely that the change in SLT ratings between both data collection time points was minor. That is, the overall intelligibility ratings only went up or down by one point and therefore were not a significant enough change to be acoustically picked up; b) Since the groups were based on overall intelligibility rather than a specific perceptual feature, it is possible jitter, shimmer and HNR did not capture the kind of speech change that had occurred, i.e. another speech feature pertaining to another speech dimension had a more distinct change, better acoustically captured by other acoustic parameters. This interpretation is explored in the following chapters with the speech dimensions articulation and prosody; c) these parameters can detect speech changes effectively only in the cases of moderate to severe dysarthria.

Contrastingly, the significant effect of the predictor *time* on CPP confirms findings that it can track speech changes over time (Šimek & Rusz, 2021). However, it did not have a significant of the predictor *group,* which can also be attributed to the PwPD groups only having mild speech symptoms and, therefore not significantly different from the control group. Studies have shown that decreased voice quality may be more prevalent in more atypical types of PD that are the results of more severe brain damage (Rusz et al., 2015; Tykalova et al., 2017), which may explain these results. As mentioned previously, CPP is linked with representing decreased voice quality, and while it has been able to pick up on speech changes over time, the lack of group differences may also be due to a general pattern of age-related decline in voice quality (Shih et al., 2007; Wong et al., 1984).

The slightly increased jitter and shimmer values compared to the control group are in line with other studies reporting higher values as well and maybe the results of excessive laryngeal muscle activity causes irregular periods of vocal fold opening. The HNR values are only marginally lower than the control group (HNR tends to be lower than control in pathological speech). However, due to the lack of significance, it is possible that airflow was not severely diminished in PwPD. The CPP results indicate that there is some change in vocal fold vibration efficiency across the two data collection times. While the lack of group differences for CPP indicate that these changes are likely within the normal range, further testing is required to confirm this or whether a true age/gender matched control group might result in a group significance.

Whether the findings above are attributable to reduced effort, weakness, and reduced muscle activation or excessive laryngeal muscle activity, as well as muscular rigidity in the larynx, is unclear but may be correlated.

Despite the lack of strong evidence for jitter, shimmer and HNR parameters in being able to track speech changes in different levels of PwPD speech intelligibility over time, CPP still provides future promise as an acoustic marker, as it can capture speech changes even in mild cases of PwPD where speech intelligibility is not severely impacted. It is clear from the results of this analysis that jitter, shimmer, and HNR continue to be good markers for distinguishing between PwPD speech and controls but may not be good indicators of changes in PwPD speech intelligibility over time. This result is limited in its generalisation due to only having a dataset with one PwPD intelligibility group and requires further investigating with other levels of intelligibility.

Similarly, it is unclear if CPP may be able to capture significant group differences in PwPD speech intelligibility is moderately or severely impacted. However, CPP's ability to track changes in PwPD speech intelligibility within a period of six months is promising. This should be investigated further in longitudinal studies to ascertain its validity as an acoustic marker for tracking PwPD speech intelligibility. CPP should also be investigated further to check whether it is possible to identify a distinct pattern in speech over time and how much individual variation drives these changes.

4.4. Chapter conclusions

While the subset in the PFR analysis was selected based only on those participants who had perceptual features rated as 'marked' and 'severe' by SLTs, overall intelligibility ratings were used for the IG to group participants based on how severely intelligibility was rated as being impacted.

The results of the PFR analysis indicated that CPP was the only acoustic parameter able to capture perceptual changes in PwPD speech over time. However, the results also indicated that the change in CPP values was not significantly different between groups. This may have been the result of averaging out of more extreme values from some PwPD speakers than others, but it cannot be concluded from the analysis. The descriptive statistics indicate that the largest change may have been with the PwPD-change group, which matches with the results of the SLT ratings. The results suggest that jitter, shimmer, and HNR were unable to capture perceptual changes in PwPD speech over time.

The results of the IG analysis indicated that CPP was also able to capture changes in different levels of PwPD speech intelligibility over time but once again that CPP values changed between all groups from T1 to T2. It is possible that CPP values were not significantly different between groups because speech intelligibility was only mildly impacted, and CPP is unable to capture significant differences between PwPD and control speech in such cases. Jitter, shimmer, and HNR were able to capture these group differences but were unable to capture changes in PwPD speech intelligibility over time.

The results of both analyses suggest that CPP is better at capturing perceptual speech changes and changes in speech intelligibility over time. However, the lack of significant group differences in CPP values in the PFR analysis would suggest that CPP is not a robust acoustic parameter for differential diagnosis. This interpretation is limited since all PwPD participants had similar levels of speech severity and intelligibility, which seems to have been relatively close to the control group. CPP may better capture group differences in PwPD speakers with greater speech severity. Further studies are required to confirm this.

The results for voice quality suggest that CPP may be a better acoustic parameter to capture speech changes in PwPD as an overall measure of dysphonia and that PwPD speakers in this study had increased impairment in voice quality over time, with the PwPD-change group showing a greater impairment that the PwPD-no change group (based on the descriptive statistics), which agrees with the results of the SLT ratings. However, as previously stated, this interpretation is inconclusive as a statistical significance was not found between groups and may have been due to individual variation being averaged out during analysis.

A potential correlation between the findings reported in this chapter and PD motor symptomology may be the effect of increased vocal fold activity but not a lot of rigidity (Rusz, Cmejla, et al., 2011; Skodda et al., 2009), which resulted in only slight detectable differences between the control group and PwPD and minor changes over time. Further studies can combine the use of acoustic analysis and vocal fold imaging, such as laryngoscopy, to quantifiably explore the correlation between vocal fold activity and changes in acoustic parameters of voice quality.

The following chapter will delve into the acoustic and statistical analysis of acoustic parameters pertaining to the speech category articulation.

5. Articulation

Studies investigating articulatory imprecision in PwPD suggest that speakers often do not reach articulatory targets and are unable to maintain sufficient contact required to ensure precise articulation of a speech sound (articulatory undershoot; Theodoros & Ramig, 2011). Kent and Rosenbek (1982) found that articulatory undershoot was observed in the acoustic speech signal as a lack of distinction between sound and syllables. Other support for undershoot has been seen as reduced articulatory constriction during stop production (Ackermann & Ziegler, 1991), restricted vowel space (Weismer et al., 2001), and reduced vowel articulation index (inverse of formant centralization ratio; (Skodda, Visser & Schlegel, 2011)). In addition to evidence supporting articulatory undershooting in PD, articulatory imprecision has been attributed to continuous or inappropriate voicing of consonants, a prominent feature of the speech of people with PD (Theodoros & Ramig, 2011; Weismer, 1984). These findings, along with the results of the SLT ratings in the present study suggest that articulatory imprecision in PwPD is largely variable with multiple acoustic parameters able to capture this speech characteristic of PwPD speech. However, the present chapter will report results pertaining to acoustic parameters that may be able to track PwPD speech change within the speech characteristic.

Articulation was rated by the SLTs at T1 as one of the most impacted speech categories with the perceptual features imprecise consonant production and articulatory breakdown rated as either ‘marked’ or ‘severe’ (rated ‘3’ or ‘4’) on a five-point scale of severity (from ‘0-4’). Acoustic parameters were selected to measure the perceptual features imprecise consonant production and articulatory breakdown quantitatively from the speech signal. The parameters selected were the plosives and fricatives intensities and VOT extracted from The Grandfather Passage. The rationales for the selection of these acoustic parameters were included in section 5.1 below.

Two independent data analyses were conducted using the speech recordings of The Grandfather Passage to investigate the effectiveness of the selected acoustic parameters in in answering the research questions. The perceptual feature ratings (PFR) analysis was conducted on a subset of the dataset (n = 52) to answer the first research question: which acoustic parameters can track perceptual changes in PwPD speech over time? This subset was selected based on the SLT ratings. Participants with the perceptual features rated as more severe (imprecise consonant production and articulatory breakdown) were selected. The recordings of this subset were included in groups based on whether the rating of the severe perceptual features negatively changed from T1 and T2 collection points or did not change at all. Therefore, this analysis focused on grouping the recordings based on detecting a change in the perceptual features over time.

The intelligibility groups (IG) analysis was conducted on the entire participant sample (referred to as the complete dataset; n = 110) and grouped based on the overall intelligibility ratings given by the SLTs and whether those ratings negatively changed from T1 to T2. This analysis would answer the second research question: which acoustic parameters can track changes in PwPD speech intelligibility over time? The overall intelligibility ratings were used to group PwPD recordings as either mild, moderate, or severe and if the ratings increased or decreased between data collection points.

The structure of the rest of the chapter includes details of the selection of acoustic parameters used for both analyses, how each analysis was conducted along with the results and discussion, and an overall discussion of the results of this chapter and its implications for voice quality in PwPD speech.

5.1. Selection of acoustic parameters

Speech in PwPD can result in inaccurate articulation (Duffy, 2020), which encompasses imprecise consonant production and articulatory breakdown. These two perceptual features can involve articulatory undershoot which is due to rigidity resulting in the reduction in the range of movements of the lips, tongue, and the jaw

(Ackermann & Ziegler, 1991; Duffy, 2020; Pawlukowska et al., 2015). Articulatory breakdown can include increased syllable stress, loudness, pitch outbursts, prolongations of phonemes and prolonged intervals between sounds and words.

Investigating intensity of articulatory production can give insight into the strength of articulators as intensity can often be reduced when articulators are unable to maintain the necessary position required to execute correct articulatory postures (Dromey et al., 1995; Kim, 2017). Imprecise articulation has specifically been argued to be a central feature in nearly all subtypes of dysarthria (Kent & Kim, 2003). Plosives are reported to be the most commonly affected consonants and may have reduced perceptual quality as a consequence of failures in reaching the expected complete articulatory closure (Ackermann & Ziegler, 1991), possibly to the point where the plosive is perceived as a fricative (Chenausky et al., 2011; Eklund et al., 2015; Forrest et al., 1989; Kim, 2017).

It has been shown that reduced articulatory precision, especially in consonant production, can lead to reduced acoustic intensity in PwPD (Kim, 2017; Rusz et al., 2021a) due to the complexity of forcing air through a narrow constriction for an extended period of time and maintaining the correct articulatory posture (which is difficult with reduced lip, tongue, and jaw mobility).

Fricatives are reported as the next most commonly affected consonants in PD. This is because the realisation of fricatives in PwPD is sensitive to errors in place of articulation and in voicing contrast, with consequent adverse effects on speech intelligibility (Eklund et al., 2015). It has also previously been reported that fricatives can be produced with lower intensity in PwPD (Kim, 2017).

Interestingly, there is a link between articulatory precision and perception of loudness i.e., the more precise articulation is, the more speech is perceived as louder and having greater projection during production (Myers & Finnegan, 2015). Increasing vocal intensity in speech treatment can lead to overarticulation which results in positive changes in articulation (Dromey et al., 1995; Lansford et al., 2011). Therefore, measuring intensity might indicate the relationship between loudness and articulation during analyses. Fricative intensity was included in both the analyses of the present study in consideration of two speech characteristics of PD, reduced loudness and reduced phonetic contrast (often described as “blurred”, or “mumbling” speech), which can also result from imprecise consonant production and articulatory breakdown (Kim, 2017).

Based on the above, the intensity of plosives and fricatives and were selected as the parameters to investigate articulatory breakdown and imprecise consonant production in PwPD. VOT was also selected and justified below.

VOT determined for plosives is perhaps the most frequently used parameter and a relatively large amount of data has been published on VOT in PwPD (Argüello-Vélez et al., 2020; Kim et al., 2011; Klatt, 1975; Novotný et al., 2014; Rusz, Tykalová, et al., 2021). VOT is a measure of the coordination of speech articulation and voicing and defined as the length of a consonant from initial burst to vowel onset. In PwPD, the coordination of the articulators can be impacted by slowing of lip and tongue movement which leads to increased time required to produce each consonant (Rusz et al., 2021a,b). Therefore, it can be used as a reliable measure of imprecise consonant production. However, previous studies have provided rather contradictory findings. While some researchers have reported increased VOT duration (Forrest et al., 1989; Novotný, Rusz, Cmejla, & Ruzicka 2014; Tykalova, Rusz, Klempir, Cmejla, & Ruzicka, 2017) others have observed unchanged (Fischer & Goberman, 2010; Ravizza, 2003) or even decreased VOT (Flint, Black, Campbelltaylor, Gailey, & Levinton, 1992) in PwPD compared to controls. It has been suggested that these discrepancies may be due to the fact that the measurement of VOT is dependent on speaking rate but, VOT ratio, a rate-independent variation of VOT, did not clarify these ambiguous findings (Fischer & Goberman, 2010; Novotný et al., 2014). Despite these contradictory findings, which may be due to a number of factors including the speech severity of PwPD, VOT in the present study can be used to assess its ability to track PwPD speech change and compare results against previous literature.

5.2. Perceptual feature ratings (PFR) analysis

5.2.1. Participant demographics

The subset consisted of 52 speakers with speech recording divided into three groups: PwPD-change, PwPD-no change, and control. The demographic information for each group is presented in Table 20 below:

Table 20. Participant demographics for the PFR analysis.

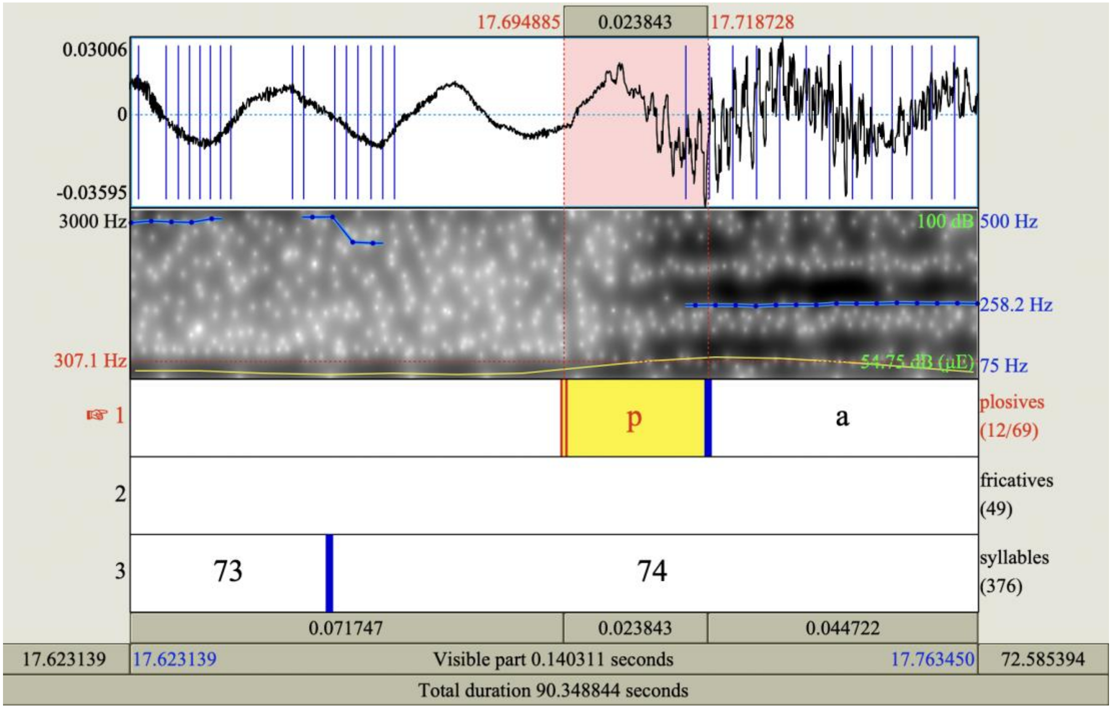
Groups	**Number of participants**	**Age range (years)**	**Mean age (years)**
PwPD-change	N= 21; M = 15; F = 6	50-93	69 (SD = 10.46)
PwPD-no change	N= 21; M = 13; F = 8	56-84	71 (SD = 6.84)
Control	N= 10; M = 5; F = 5	51-82	70 (SD = 9.94)

5.2.2. Segmentation of recordings

The type of data recordings used for this analysis were the reading passage recordings. All plosives and fricatives from The Grandfather Passage that were at syllable initial position but not part of a consonant cluster were annotated using Praat (Boersma & Weenink, 2012). This resulted in annotating four occurrences for /p/, /b/, /t/, /d/, /g/, /f/, /v/, and /s/. However, there were only two occurrences of /z/, and one occurrence of /k/ and /ʃ/. Since recordings of The Grandfather Passage were collected twice, both recordings were used to increase the number of plosives and fricatives available for analysis. The plosives and fricatives were then annotated from the rest of the signal, with the onset being defined as the point in a digital spectrographic record where aperiodic noise first appeared in the high frequency range. Fricative offset was defined as the point of intensity minimum immediately prior to the onset of the periodic portion of the vocalic nucleus following the fricative.

Plosive, stop consonants (/p/ /t/ and /k/) were also annotated for VOT measurement and therefore the closure, the release and the burst of the consonant was inspected in the total plosive duration window until the start of vowel production (Rusz et al., 2021). Inspection of time wave displays, and narrow-band spectrograms served, to validate the presence of a stop burst. VOT measures were taken from the first evidence of stop release to the onset of voicing. An example of the annotation of the plosive /p/ can be seen in Figure 10 below. It is an example of one of the less clear boundaries of the stop consonant as verified by an external researcher.

Figure 10. The annotation in praat for the plosive /p/ from The Grandfather Passage based on visual inspection of the spectrogram. The presence of closure and release was marked with the boundaries.



5.2.3. Extracting parameters

All the plosive and fricative absolute intensity values were extracted from Praat (Boersma & Weenink, 2020) by extracting all the annotations for each plosive and fricative and obtaining the average mean intensity in decibels (dB) for each plosive and fricative. For example, all annotations for /p/ were extracted and the average mean intensity for all the occurrences of /p/ included in the analysis. In addition, a combined mean intensity value was extracted for all plosives together and all fricatives together. This resulted in the following extracted values for analysis: plosives – average mean plosive intensity, /p/, /b/, /t/, /d/, /k/, /g/; fricatives – average mean fricative intensity, /f/, /v/, /s/, /z/, /ʃ/.

Differences in VOT have been shown to distinguish voiced and voiceless stop consonants. For voiceless consonants to be perceived, voicing must be delayed by more than 25 milliseconds (ms) relative to plosive release. If VOT is less than 20 ms, a voiced plosive is perceived (Klatt, 1975). VOT was extracted based on (Rusz et al., 2021) by using the annotations to calculate the interval between the initial articulatory release of a stop consonant and the onset of vocal fold vibration in ms.

5.2.4. Reliability of acoustic annotations

To determine reliability of the annotations, 10% of textgrids were manually re-examined by the primary researcher and an external researcher trained in acoustic analysis. The textgrids were randomly selected from either data collection point. Plosive and fricative annotations for the reading passage were manually rechecked. The external researcher did not find any discrepancies in the annotations made, with all the annotations between the primary and external research being within 1.5ms of each other.

5.2.5. Descriptive statistics of plosives

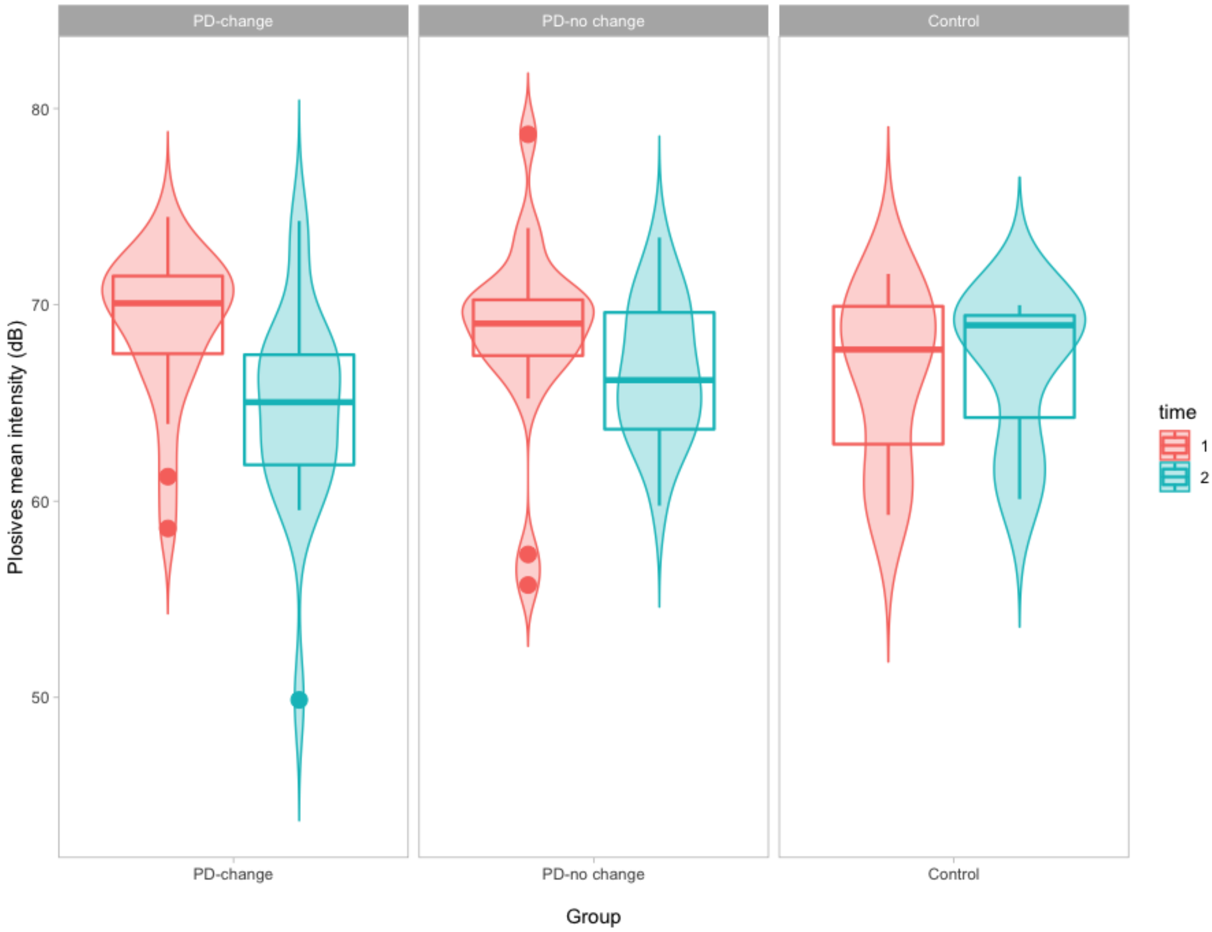
5.2.5.1. Average plosives mean intensity

The descriptive statistics of the average plosives mean intensity can been seen in Table 21 below. It indicates that mean values for the PwPD groups were higher than the control group at T1, but relatively similar at T2 indicating a decrease in average plosive intensity values in both PwPD groups over time. In addition, the greater SD for the PwPD-change group at T2 (5.16) suggests a wider distribution of values compared to the other groups. This can be better visualised in Figure 11 below where there appears to be a general decrease in average plosive mean intensity values in the PwPD groups while the control group remains stable over time.

Table 21. Summary table of descriptive statistics of average plosives mean intensity for each group at T1 and T2. N = 21 (PwPD-change); 21 (PwPD-no change); 10 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	66.44	4.40	67.72	59.30	71.57
Control (T2)	66.94	3.82	68.96	60.10	69.98
PwPD-change (T1)	68.93	3.89	70.08	58.61	74.48
PwPD-change (T2)	64.63	5.16	65.04	49.87	74.27
PwPD-no change (T1)	68.52	4.91	69.05	55.73	78.69
PwPD-no change (T2)	66.66	3.52	66.16	59.77	73.43

Figure 11. The distribution of average plosives mean intensity of each group (PwPD-change, PwPD-no change, Control) at T1 and T2.



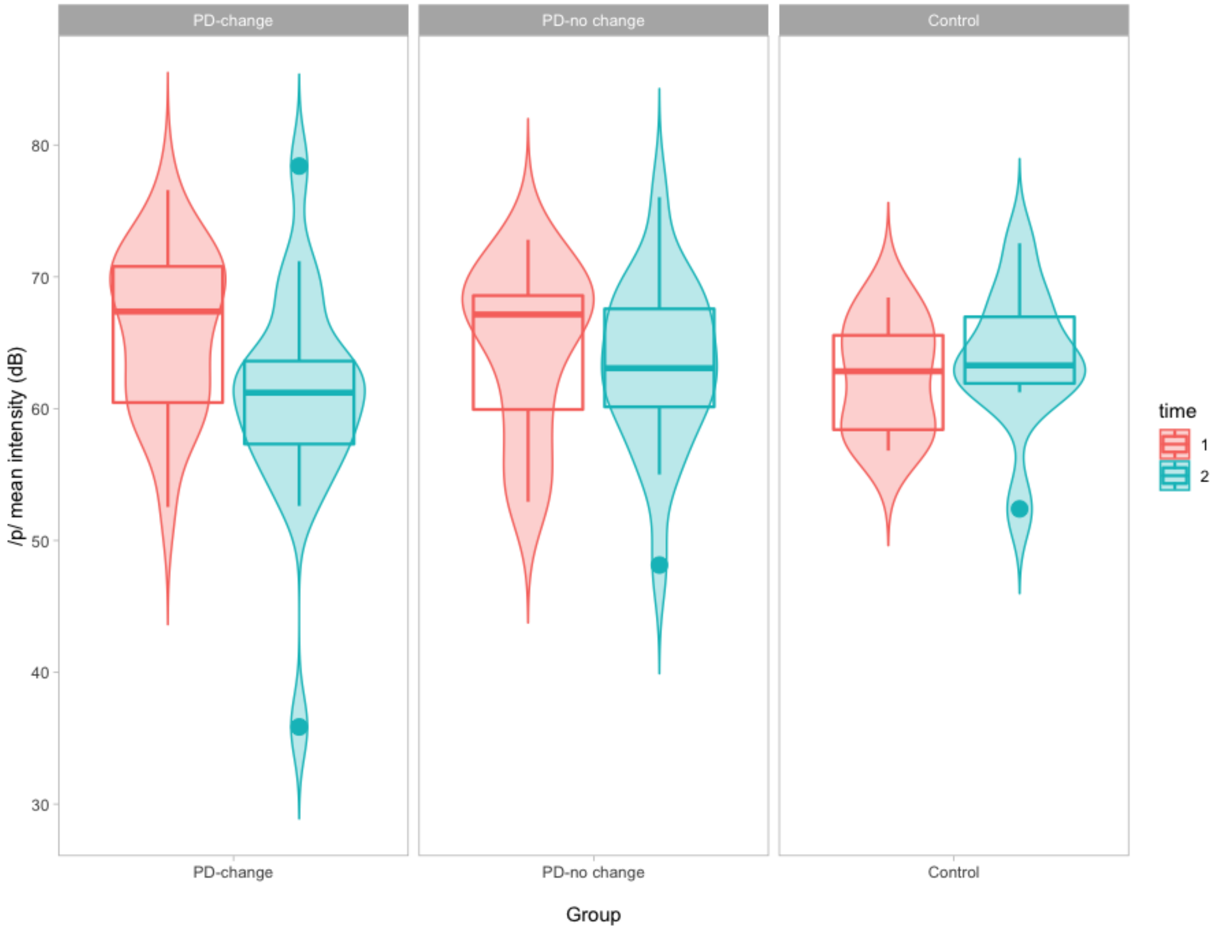
5.2.5.2. /p/ mean intensity

Descriptive statistics of /p/ mean intensity (see Table 22) suggests that values were higher for both PwPD groups compared to the control group at T1, and decreased over time. There is a greater decrease of the mean of values (5.28dB) in the PwPD-change group compared to the PwPD-no change group which decreased by a mean of 1.46dB. However, there is a larger variance in distribution of values in the PwPD-change group compared to other groups, as indicated by the SDs at T1 (6.07) and T2 (8.44). This may be the cause for the larger decrease in mean values from T1 to T2 compared to the other groups. The distribution of values can be seen at a glance in Figure 12 which also suggest a greater change in /p/ mean intensity in the PwPD-change group. This will be further explored during the statistical analysis.

Table 22. Summary table of descriptive statistics of /p/ mean intensity for each group at T1 and T2. N = 21 (PwPD-change); 21 (PwPD-no change); 10 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	62.39	4.22	62.84	56.81	68.43
Control (T2)	63.80	5.35	63.28	52.40	72.54
PwPD-change (T1)	66.01	6.07	67.37	52.54	76.58
PwPD-change (T2)	60.73	8.44	61.21	35.85	78.40
PwPD-no change (T1)	64.75	6.25	67.14	52.94	72.81
PwPD-no change (T2)	63.29	6.28	63.07	48.13	76.04

Figure 12. The distribution of /p/ mean intensity of each group (PwPD-change, PwPD-no change, Control) at T1 and T2.



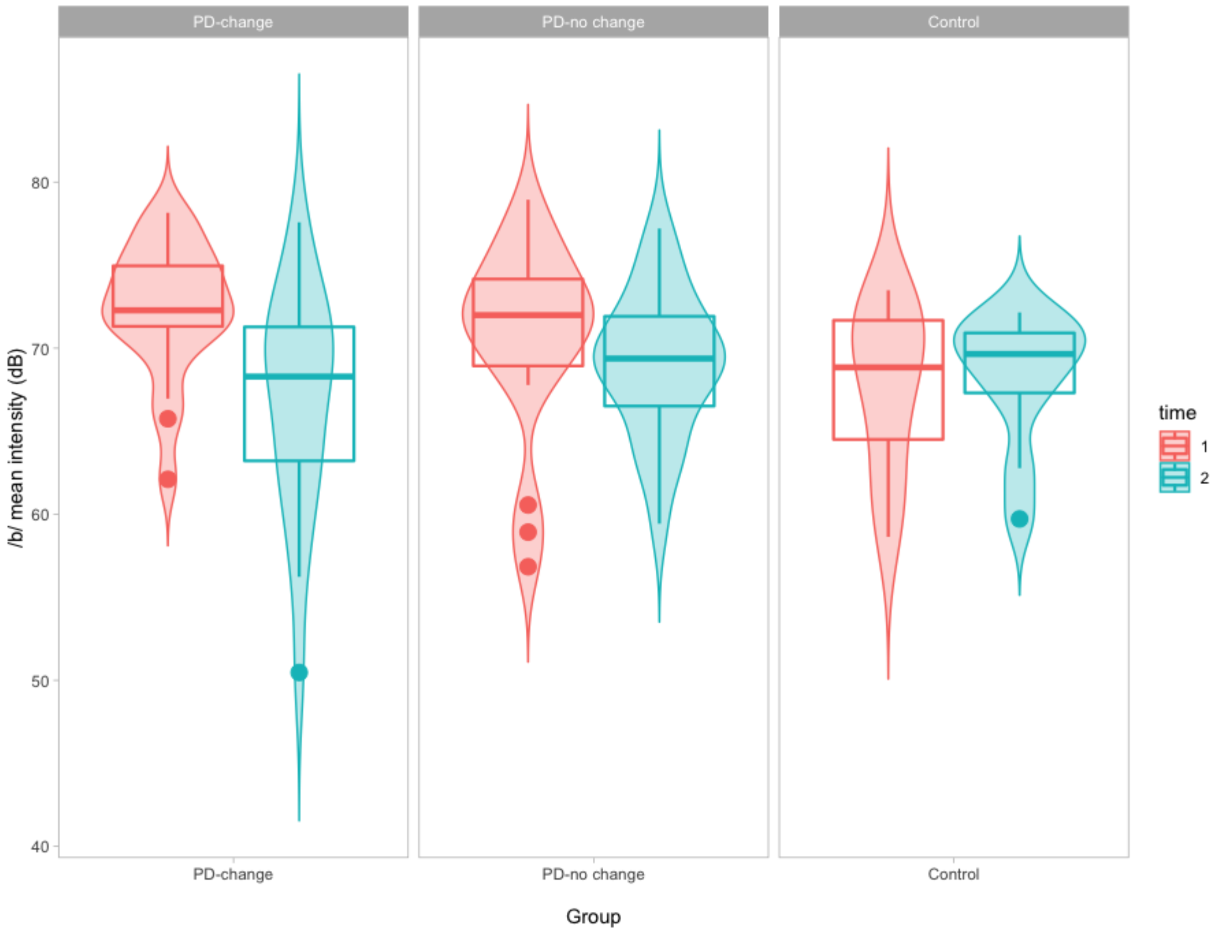
5.2.5.3. /b/ mean intensity

A summary table of the descriptive statistics pertaining to /b/ mean intensity is presented in Table 23. It shows that mean values were higher for both PwPD groups compared to the control group at T1 and decreased at T2, similar to the /p/ mean intensity values. In addition, a similar indication as the /p/ mean intensity values is that there is a greater variation in distribution of /b/ mean intensity values in the PwPD-change group at T2 based on the SD which might be driving the larger decrease in mean values (5.62dB) in the PwPD-change group from T1 to T2. The distribution can be better compared by looking at Figure 13 below where although there is a slight decrease in PwPD-no change group the values for this and the control group appear to be quite similar and stable over time.

Table 23. Summary table of descriptive statistics of /b/ mean intensity for each group at T1 and T2. N = 21 (PwPD-change); 21 (PwPD-no change); 10 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	67.71	5.02	68.84	58.64	73.49
Control (T2)	68.23	4.05	69.64	59.72	72.15
PwPD-change (T1)	72.38	3.97	72.28	62.10	78.15
PwPD-change (T2)	66.86	6.77	68.29	50.47	77.58
PwPD-no change (T1)	70.72	5.76	71.99	56.84	78.94
PwPD-no change (T2)	69.09	4.35	69.37	59.44	77.20

Figure 13. The distribution of /b/ mean intensity of each group (PwPD-change, PwPD-no change, Control) at T1 and T2.



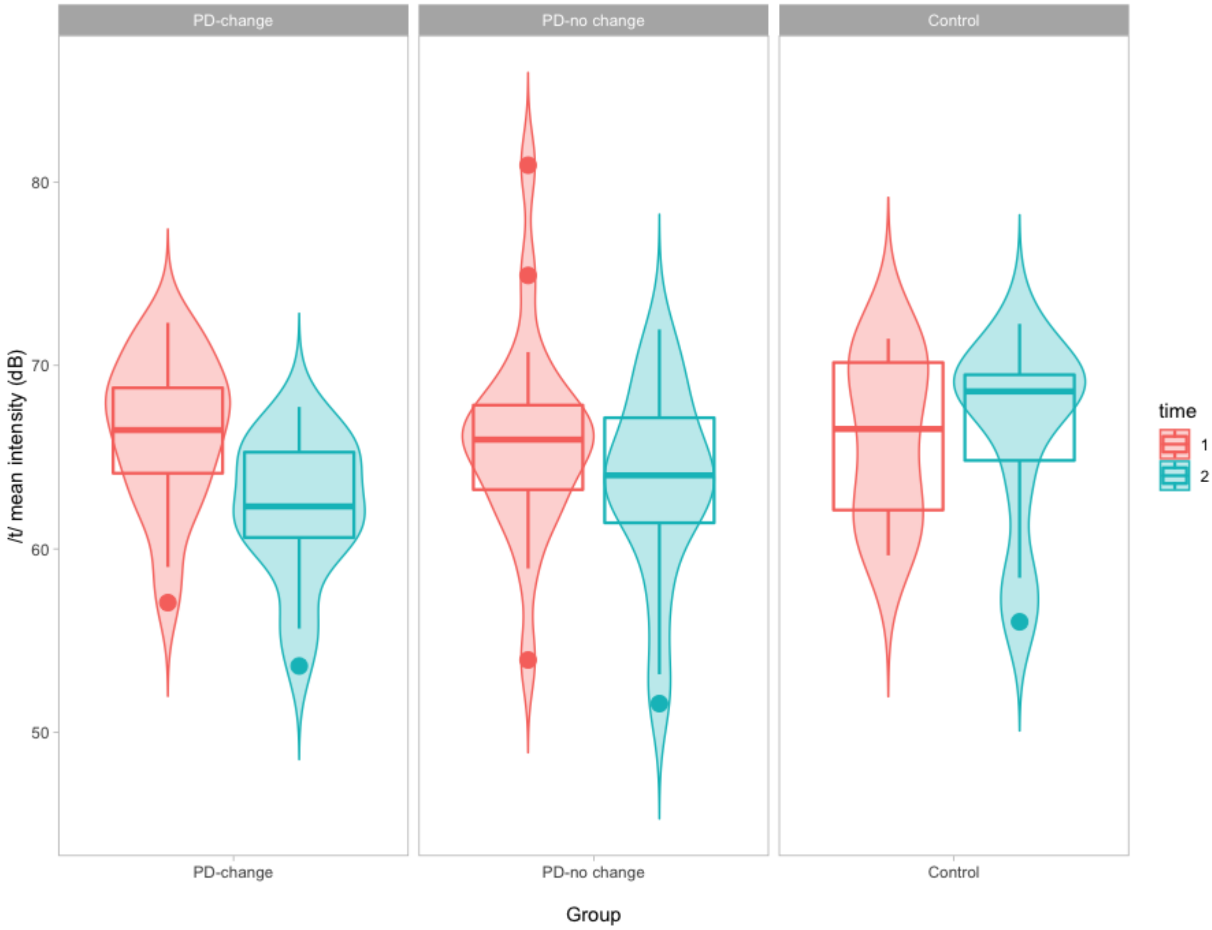
5.2.5.4. /t/ mean intensity

The results of the descriptive statistics for /t/ mean intensity (see Table 24) shows that mean values are relatively similar for all groups at T1, but the PwPD groups decrease in mean values over time. The PwPD-change group's mean values decrease from T1 (66.35; SD = 4.09) to T2 (62.18; SD = 3.77) more than the PwPD-no change group from T1 (66.06; SD = 5.49) to T2 (63.48; SD = 5.21). The control group's values remain relatively stables over time. There is an indication that the greater variance in values for the PwPD-no change group clearly in Figure 14 may be contributing to the change in mean values over time.

Table 24. Summary table of descriptive statistics of /t/ mean intensity for each group at T1 and T2. N = 21 (PwPD-change); 21 (PwPD-no change); 10 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	66.08	4.52	66.54	59.65	71.46
Control (T2)	66.41	5.31	68.58	56.03	72.26
PwPD-change (T1)	66.35	4.09	66.48	57.07	72.33
PwPD-change (T2)	62.18	3.77	62.32	53.69	67.74
PwPD-no change (T1)	66.06	5.49	65.96	53.95	80.91
PwPD-no change (T2)	63.49	5.21	64.01	51.57	71.96

Figure 14. The distribution of /t/ mean intensity of each group (PwPD-change, PwPD-no change, Control) at T1 and T2.



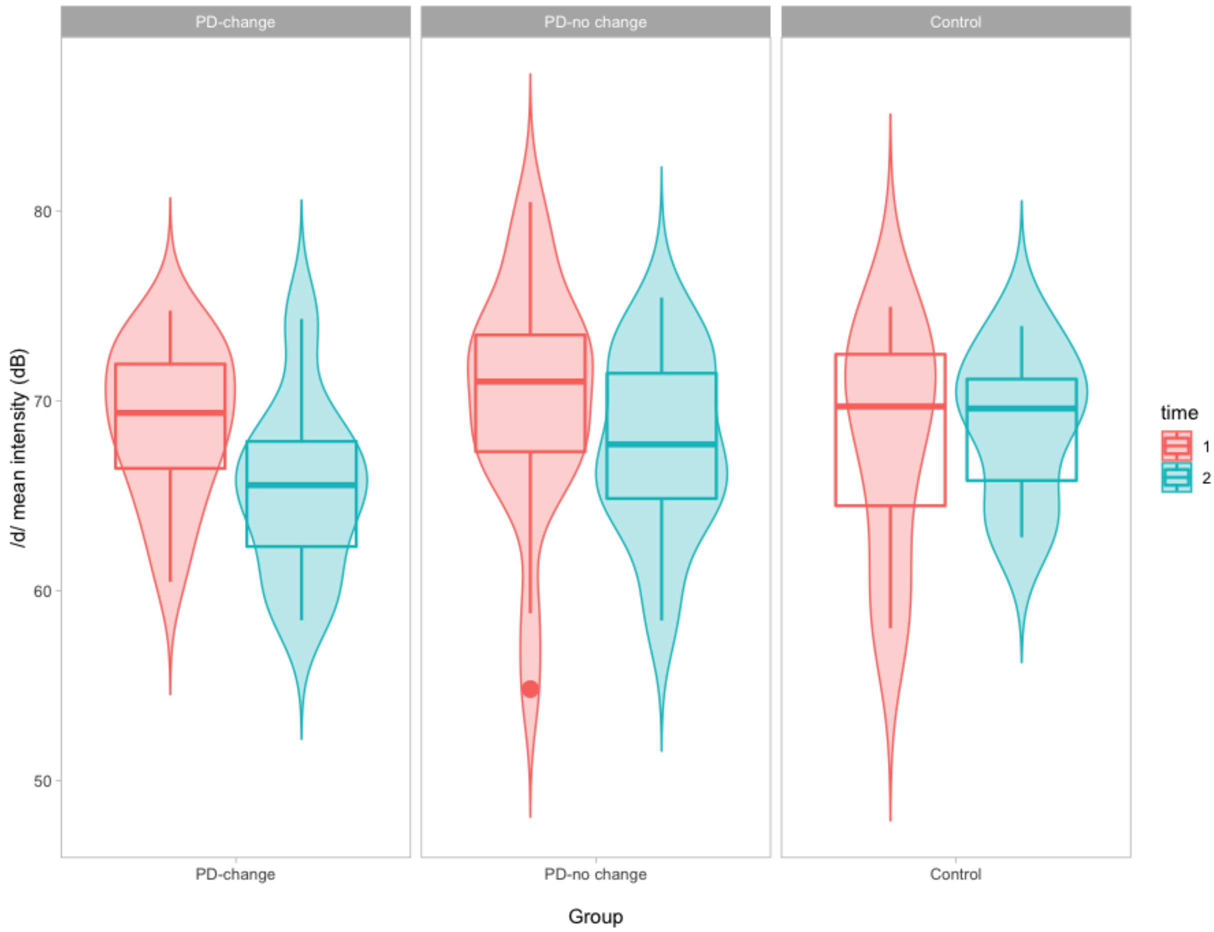
5.2.5.5. /d/ mean intensity

Descriptive statistics of /d/ mean intensity seen below in Table 25 that both PwPD groups mean values are higher than the control group similar to the /p/ and /b/ mean intensity values. The PwPD-change group's mean values decrease from T1 to T2 (-3.53dB) slightly more than the PwPD-no change group (-2.4dB), but the higher variance in distribution of the PwPD-no change group (see Figure 15) is likely contributing to the mean values seen in the table. The control group's mean values remained relatively stable from T1 to T2 with a slight increase (by 0.82dB) showing a common trend based on the descriptive statistics of the previous plosives reported where control values stayed relatively stable over time. Further statistical analysis may indicate if the decrease in /d/ mean intensity in any of the groups is significant.

Table 25. Summary table of descriptive statistics of /d/ mean intensity for each group at T1 and T2. N = 21 (PwPD-change); 21 (PwPD-no change); 10 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	67.99	5.96	69.70	58.03	74.95
Control (T2)	68.81	3.87	69.59	62.81	73.94
PwPD-change (T1)	68.90	4.00	69.36	60.48	74.75
PwPD-change (T2)	65.37	4.54	65.56	58.45	74.31
PwPD-no change (T1)	70.08	6.02	71.01	54.82	80.47
PwPD-no change (T2)	67.68	4.68	67.71	58.43	75.44

Figure 15. The distribution of /d/ mean intensity of each group (PwPD-change, PwPD-no change, Control) at T1 and T2.



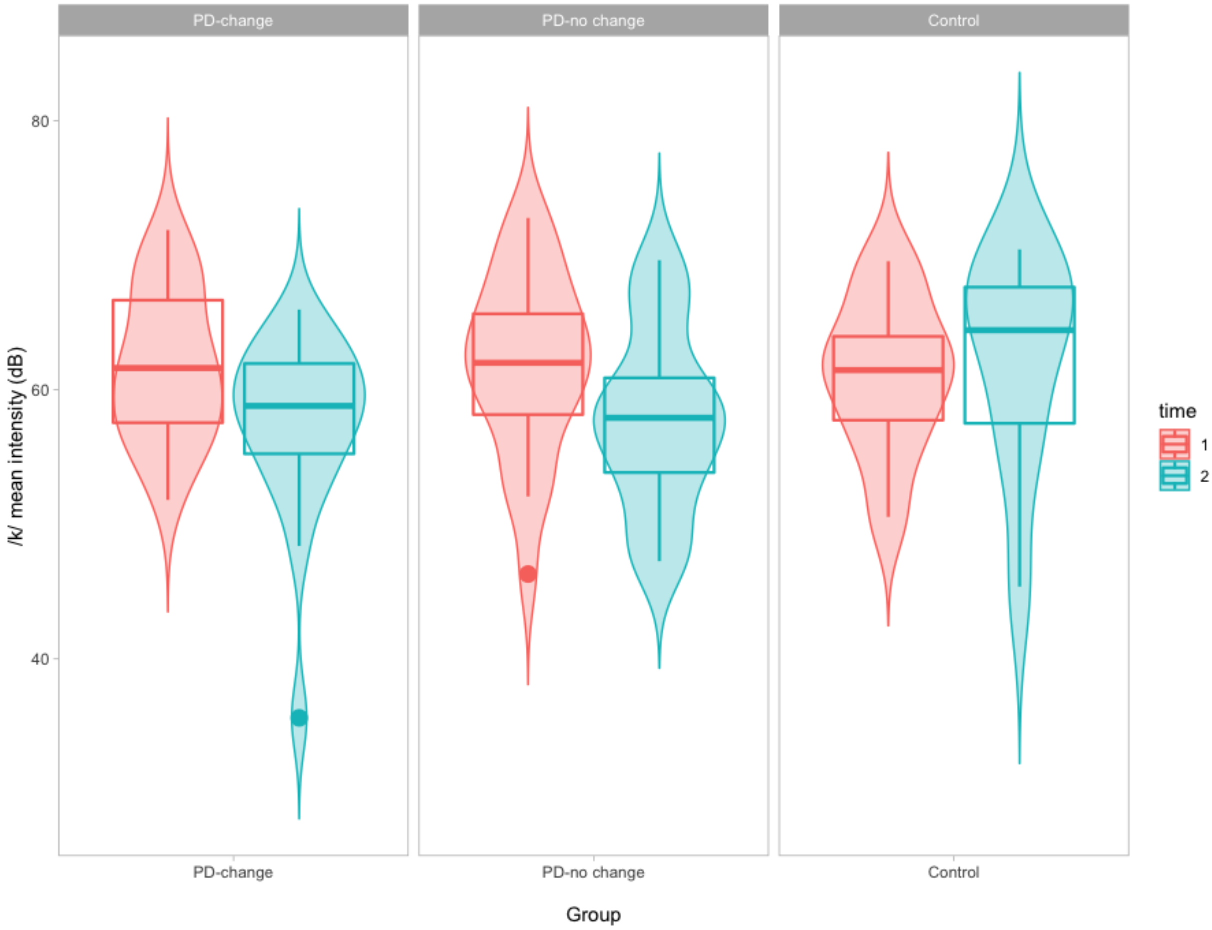
5.2.5.6. /k/ mean intensity

The results of the descriptive statistics on /k/ mean intensity in Table 26 shows that both the PwPD groups and the control groups mean values were all similar at T1 and while PwPD groups shows a decrease over time, the control group's mean values increased slightly over time. However, there is a larger SD for the control group at T2 (8.09) compared to T1 which may be influencing the slight increase seen over time. Both the PwPD-change and PwPD-no change group show a relatively similar decrease in mean values over time. The PwPD-change group's mean values decrease by only 0.28dB more than the PwPD-no change group. However, this may be influenced by the distribution of values which is plotted in Figure 16. The plot suggests some change occurs between all groups over time but whether this could be significant is unclear.

Table 26. Summary table of descriptive statistics of /k/ mean intensity for each group at T1 and T2. N = 21 (PwPD-change); 21 (PwPD-no change); 10 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	61.10	5.87	61.47	50.55	69.56
Control (T2)	62.33	8.09	64.44	45.37	70.44
PwPD-change (T1)	61.73	5.66	61.62	51.82	71.88
PwPD-change (T2)	57.37	6.89	58.80	35.61	65.96
PwPD-no change (T1)	61.92	6.82	62.01	46.31	72.78
PwPD-no change (T2)	57.82	6.75	57.92	47.27	69.64

Figure 16. The distribution of /k/ mean intensity of each group (PwPD-change, PwPD-no change, Control) at T1 and T2.



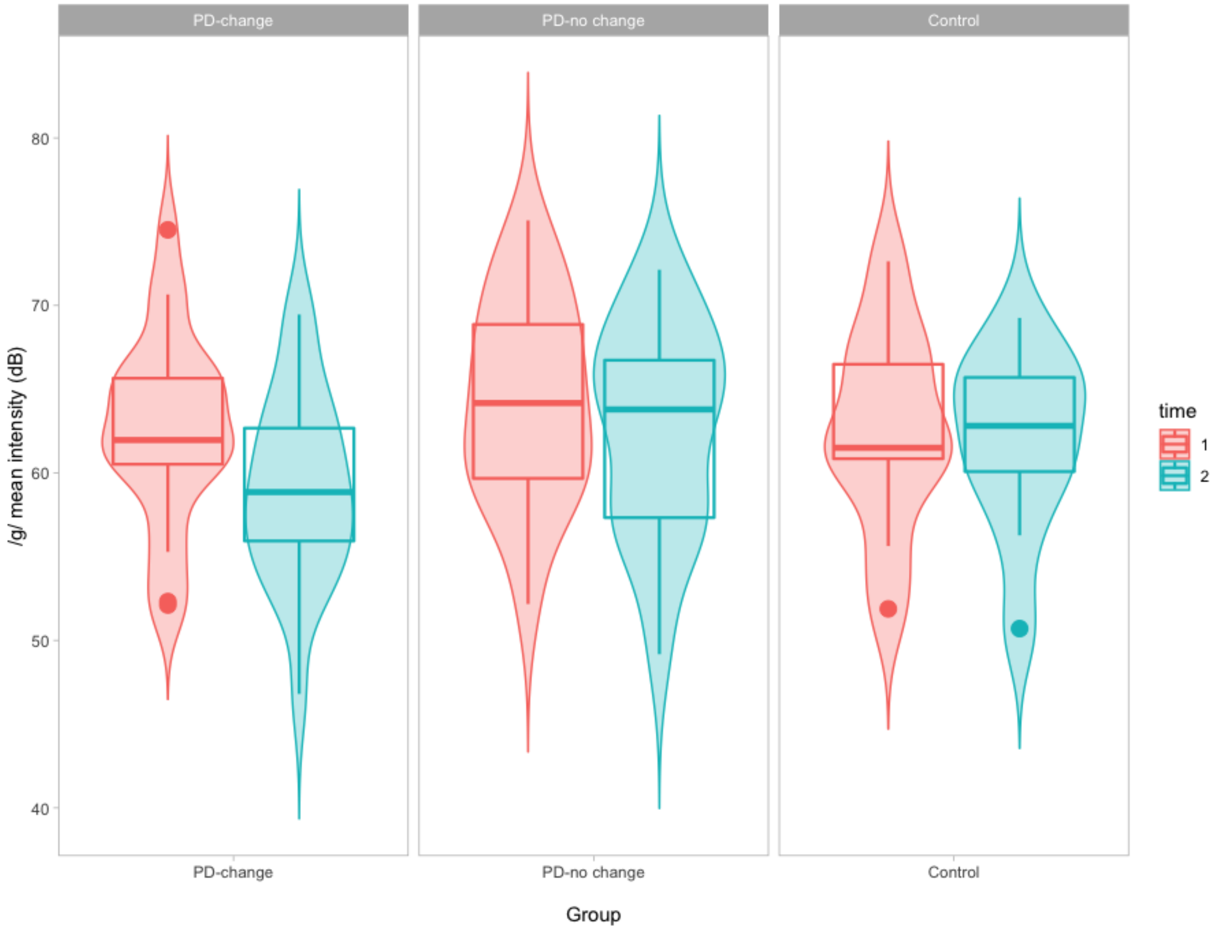
5.2.5.7. */g/ mean intensity*

In Table 27, the descriptive statistics of /g/ mean intensity are presented. The table shows that while the PwPD-no change and control group values remain relatively stable over time, there is a larger difference in the PwPD-change group mean values from T1 to T2. This suggest a larger negative change over time compared to the PwPD-no change group, which shows similar values to the control group at T2, although a slightly higher value than the control group at T1. The statistical significance of the negative change in the PwPD groups is unclear but the distribution of values seen in Figure 17 follows a similar pattern to all the previous plosives distributions plotted above, where both PwPD groups indicate some reduction in plosive intensity over time.

Table 27. Summary table of descriptive statistics of /g/ mean intensity for each group at T1 and T2. N = 21 (PwPD-change); 21 (PwPD-no change); 10 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	62.58	6.08	61.50	51.88	72.63
Control (T2)	61.91	5.45	62.81	50.70	69.26
PwPD-change (T1)	62.68	5.65	61.96	52.11	74.51
PwPD-change (T2)	59.04	5.64	58.86	46.81	69.46
PwPD-no change (T1)	64.25	6.01	64.17	52.17	75.08
PwPD-no change (T2)	62.62	6.27	63.79	49.18	72.13

Figure 17. The distribution of /g/ mean intensity of each group (PwPD-change, PwPD-no change, Control) at T1 and T2.



5.2.6. Results of LMMs of plosives

To test the significance of each fixed effect on each acoustic parameter (all plosive intensities), a linear mixed effects model (LMM) analysis was conducted using R (R Core Team, 2019). The models were run with the lme4 package (Bates et al., 2015) using lmer. The fit for each model was checked by adding any possible interactions between fixed effects and any slopes between all factors. The best fit was found to only contain each acoustic parameter tested against *time* and *group* as predictors and *participant* as the random factor. Optimizers were used to ensure that over-fitting did not occur.

A model was run with each acoustic parameter (outcome variable) against the predictors *time* (data collection time points T1 and T2) and *group* (PwPD-change, PwPD-no change, and control) with the random factor of *participant* including a random intercept and random slope. A random intercept means that the final model takes into account that each individual *participant* can show higher or lower values for each acoustic parameter regardless of *group* or the data collection *time*. A random slope for *participant* means that the model considers and accounts for some between-participant variation.

Based on the assumptions of an LMM, the outcome variable does not need to be normally distributed if the residuals of each model are normally distributed. The residuals for each model run on each acoustic parameter (outcome variable) was checked and did not violate the assumption of normality. Therefore, none of the data required any transformations. There were no significant correlations between any of the predictors and therefore was not reported.

LMMs create an intercept in each model representing each acoustic parameter against T1 data collection and the control group. The intercept represents how different the outcome variable and predictors are from chance. The other predictors are compared against this intercept model and results are reported based on how different other predictors are from the intercept.

A summary table of the results for each of the plosive acoustic parameters based on 104 observations from al 52 participants can be seen in

Table 28 below.

The LMMs results indicated that the effect of the predictor *time* on all the plosives mean intensities, except /p/ mean intensity was significant which suggests that for the acoustic parameters average plosive mean intensity, /b/, /t/, /k/, and /g/ mean intensity, there was a significant negative change over time. The largest negative change was seen in the /k/ mean intensity values by an estimated 3.31dB from T1 to T2. However, the results also indicated that the effect of the predictor *group* was not significant which implies that the negative change in the significant plosive acoustic parameters was observed across all groups over time. While the statistical analysis suggests an overall negative change in intensity in the plosive acoustic parameters above regardless of group, the descriptive statistics reported in the previous sections suggested that negative changes were only observed in the PwPD groups and the control group either remained stable or slightly increased. This can help interpret the results, suggesting that the significant negative change may be driven by the PwPD groups more than the control group.

Table 28. Linear mixed effects model results of the PFR analysis of all plosives against each of the levels of the predictors time and group.

Fixed Effects:						
Acoustic parameter	**Predictor**	**Estimate**	**Std. Error**	**df**	**t-value**	**p**
Avg. plosives intensity	Intercept (Control, T1)	67.89	1.14	62.17	59.50	< 0.01*
	T2	-2.39	0.79	51.00	-3.02	0.004**
	PwPD-change	0.09	1.30	49.00	0.07	0.95
	PwPD-no change	0.90	1.30	49.00	0.69	0.49
/p/ intensity	Intercept (Control, T1)	64.32	1.66	61.25	38.78	< 0.01*
	T2	-2.44	1.23	48.93	-1.98	0.05
	PwPD-change	0.33	1.88	46.58	0.18	0.86
	PwPD-no change	0.88	1.88	46.58	0.47	0.64
/b/ intensity	Intercept (Control, T1)	69.35	1.34	63.11	51.67	< 0.01*
	T2	-2.76	0.97	51.14	-2.84	0.007**
	PwPD-change	1.68	1.53	49.18	1.10	0.28
	PwPD-no change	1.94	1.52	48.80	1.28	0.21
/t/ intensity	Intercept (Control, T1)	67.58	1.31	57.17	51.42	< 0.01*
	T2	-2.66	0.73	51.00	-3.64	0.006**
	PwPD-change	-1.98	1.53	49.00	-1.29	0.20
	PwPD-no change	-1.47	1.53	49.00	-0.96	0.34
/d/ intensity	Intercept (Control, T1)	69.52	1.36	55.93	51.29	< 0.01*
	T2	-2.24	0.78	49.45	-2.85	0.006**
	PwPD-change	-1.30	1.60	48.14	-0.81	0.42
	PwPD-no change	0.48	1.58	47.42	0.30	0.76

/k/ intensity	Intercept (Control, T1)	63.41	1.85	56.55	34.24	< 0.01*
	T2	-3.31	1.16	44.46	-2.85	0.007**
	PwPD-change	-2.20	2.12	45.43	-1.03	0.31
	PwPD-no change	-1.83	2.13	45.82	-0.86	0.40
/g/ intensity	Intercept (Control, T1)	63.40	1.61	56.05	39.31	<0.01*
	T2	-2.32	0.91	49.68	-2.54	0.014*
	PwPD-change	-1.46	1.88	47.99	-0.78	0.44
	PwPD-no change	1.19	1.88	47.66	0.64	0.53

5.2.7. Descriptive statistics of fricatives

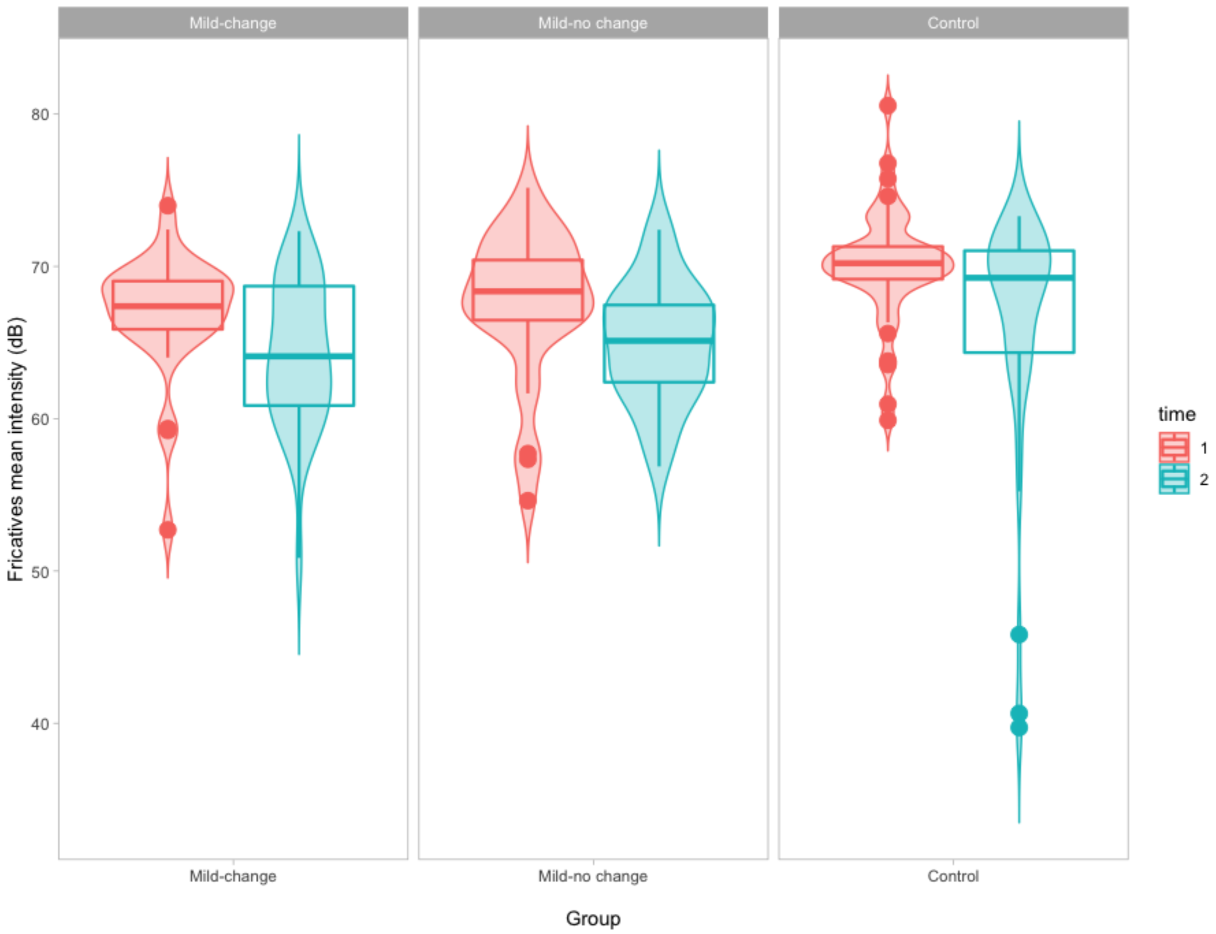
5.2.7.1. Average fricatives mean intensity

The results of the descriptive statistics of the average fricatives mean intensity is reported in Table 29 below. The table indicates that the control group had slightly higher mean values compared to the two PwPD groups and remained relatively stable over time. The PwPD change group (T1: 67.73dB, SD = 3.01; T2: 64.64dB, SD = 4.68) shows a similar negative change in mean intensity values to the PwPD-no change group (T1: 67.00dB, SD = 5.23; T2: 64.35, SD = 5.46) over time. However, a look at the plotted distribution of values in Figure 18 below suggests that the greater distribution of values in the control group may be influencing the means and therefore it is unclear how different the control group’s average fricative mean intensity was from either of the PwPD groups over time.

Table 29. Summary table of descriptive statistics of average fricatives mean intensity for each group at T1 and T2. N = 21 (PwPD-change); 21 (PwPD-no change); 10 (control).

	Mean	SD	Median	Min	Max
Control (T1)	68.97	5.03	68.50	60.96	75.76
Control (T2)	68.19	5.58	70.64	57.90	73.21
PwPD-change (T1)	67.73	3.01	68.37	59.35	71.63
PwPD-change (T2)	64.64	4.68	65.01	50.86	71.22
PwPD-no change (T1)	67.00	5.23	66.74	54.62	75.16
PwPD-no change (T2)	64.35	4.56	65.11	56.87	72.41

Figure 18. The distribution of average fricatives mean intensity of each group (PwPD-change, PwPD-no change, Control) at T1 and T2.



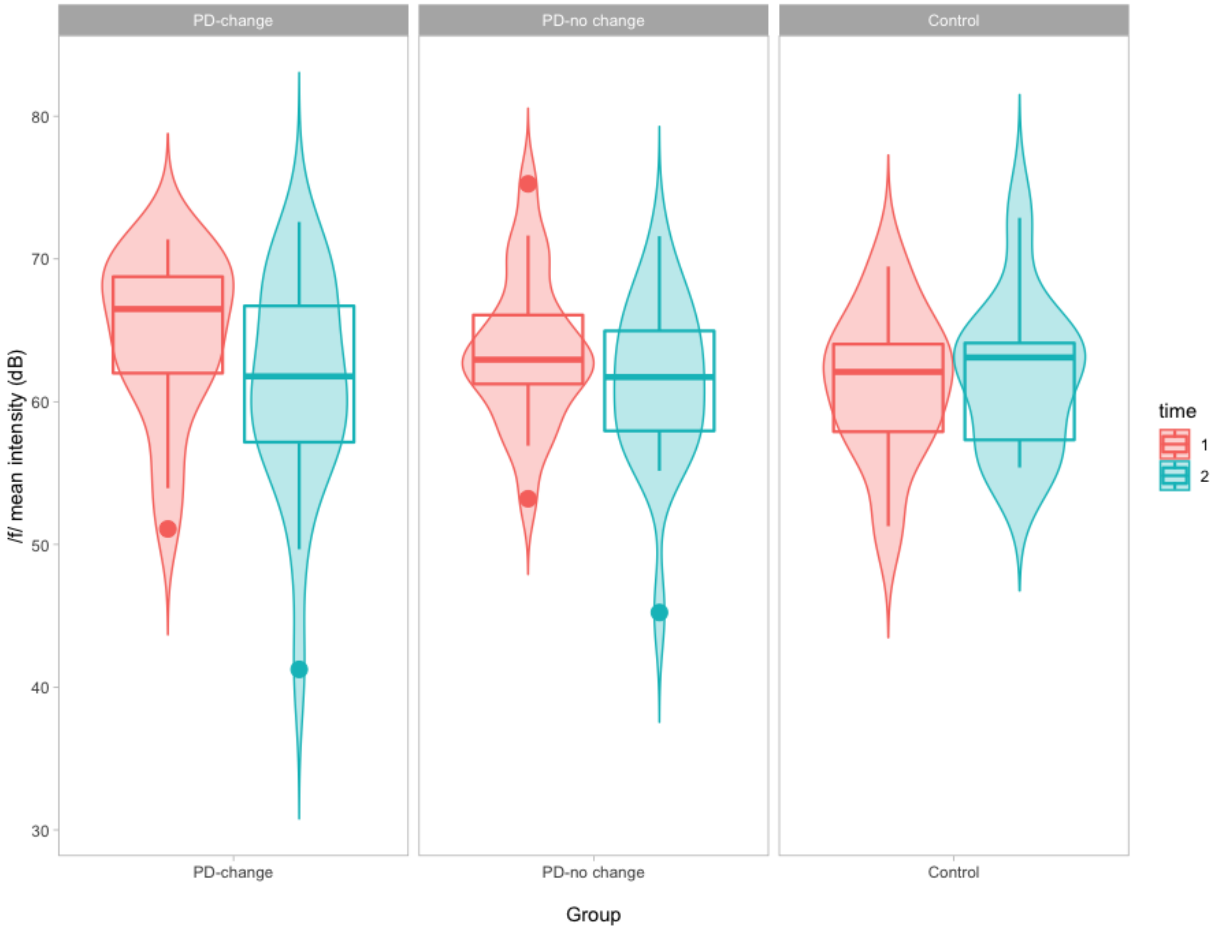
5.2.7.2. /f/ mean intensity

The /f/ mean intensity descriptive statistics presented below inn Table 30 shows that both PwPD groups had higher mean values compared to the control group at T1 and negatively changed at T2. However, the PwPD-change group had a larger negative change in means of 3.41dB compared to the PwPD-no change group which had a negative change in means of 2.23dB. However, the PwPD-change group had the highest SD of 7.45 at T2 which indicates a greater variation in distribution which likely influence the inferences made above. The distribution of values is plotted in Figure 19, but as just stated, the greater negative change suggested in the PwPD-change group may be more influenced by a larger variance in values and its relation to the other groups will need to be confirmed with statistical analysis.

Table 30. Summary table of descriptive statistics of /f/ mean intensity for each group at T1 and T2. N = 21 (PwPD-change); 21 (PwPD-no change); 10 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	61.25	5.16	62.09	51.28	69.47
Control (T2)	61.90	5.29	63.10	55.39	72.88
PwPD-change (T1)	64.59	5.70	66.49	51.09	71.37
PwPD-change (T2)	61.18	7.45	61.78	41.25	72.60
PwPD-no change (T1)	63.50	5.23	62.94	53.20	75.26
PwPD-no change (T2)	61.27	5.83	61.72	45.23	71.59

Figure 19. The distribution of /f/ mean intensity of each group (PwPD-change, PwPD-no change, Control) at T1 and T2.



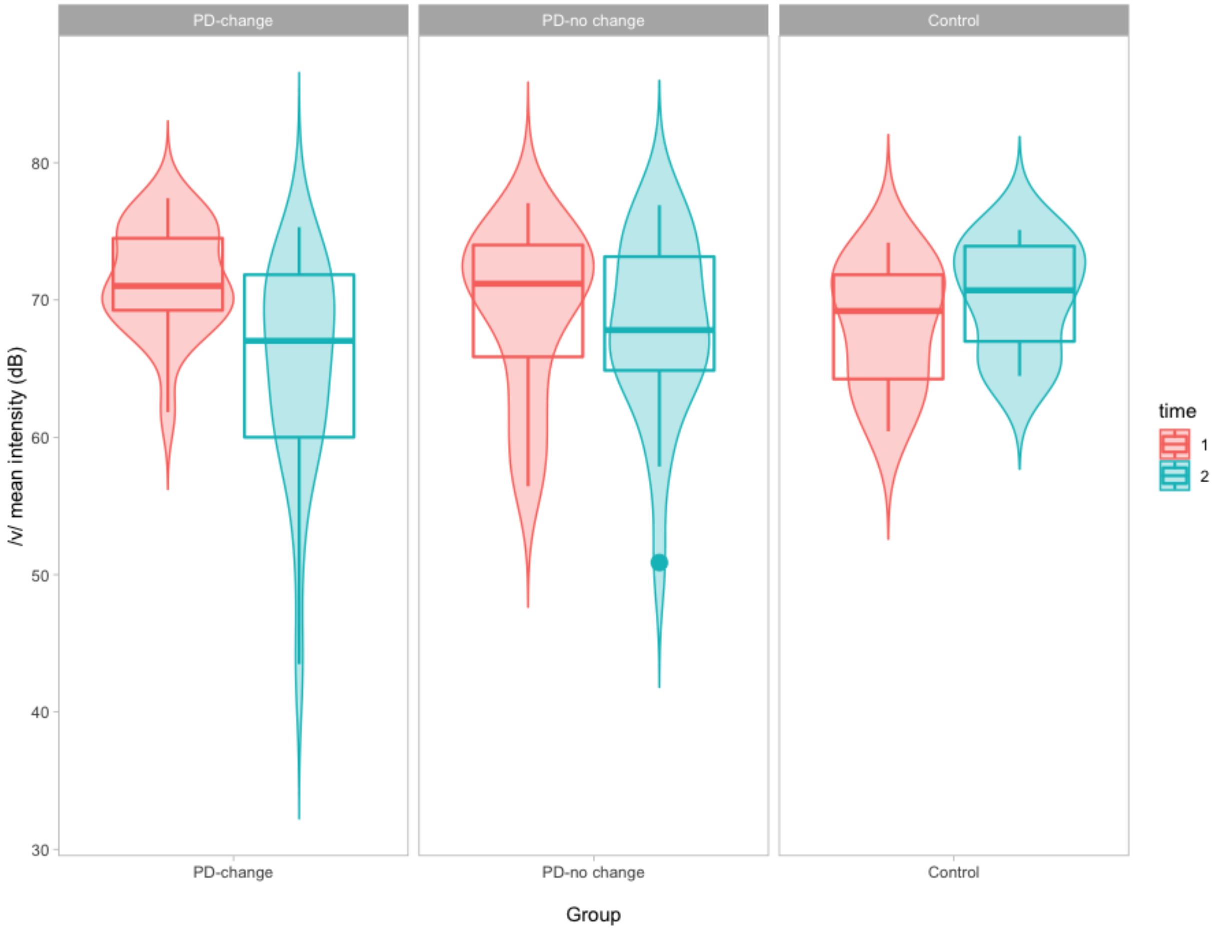
5.2.7.3. /v/ mean intensity

In Table 31, of the descriptive statistics of /v/ mean intensity, the PwPD groups have higher means compared to the control group, but the PwPD-change group has the highest means (71.52dB) which leads to a greater negative change in T2 to 65.12dB (a change of -6.40dB) compared to the PwPD-no change group which has a negative change in means of 1.37dB. The distributions plotted in Figure 20 display this negative change but also indicates that there is a greater variance of distribution in the PwPD-change group at T2 which would influence the mean values reported in the descriptive statistics table. However, the negative change in intensity is consistent with the previous fricatives.

Table 31. Summary table of descriptive statistics of /v/ mean intensity for each group at T1 and T2. N = 21 (PwPD-change); 21 (PwPD-no change); 10 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	68.22	4.61	69.19	60.45	74.18
Control (T2)	70.33	3.98	70.70	64.47	75.10
PwPD-change (T1)	71.52	3.83	71.02	61.84	77.41
PwPD-change (T2)	65.12	7.67	67.03	43.48	75.30
PwPD-no change (T1)	69.35	5.99	71.19	56.44	77.06
PwPD-no change (T2)	67.98	6.56	67.81	50.88	76.90

Figure 20. The distribution of /v/ mean intensity of each group (PwPD-change, PwPD-no change, Control) at T1 and T2.



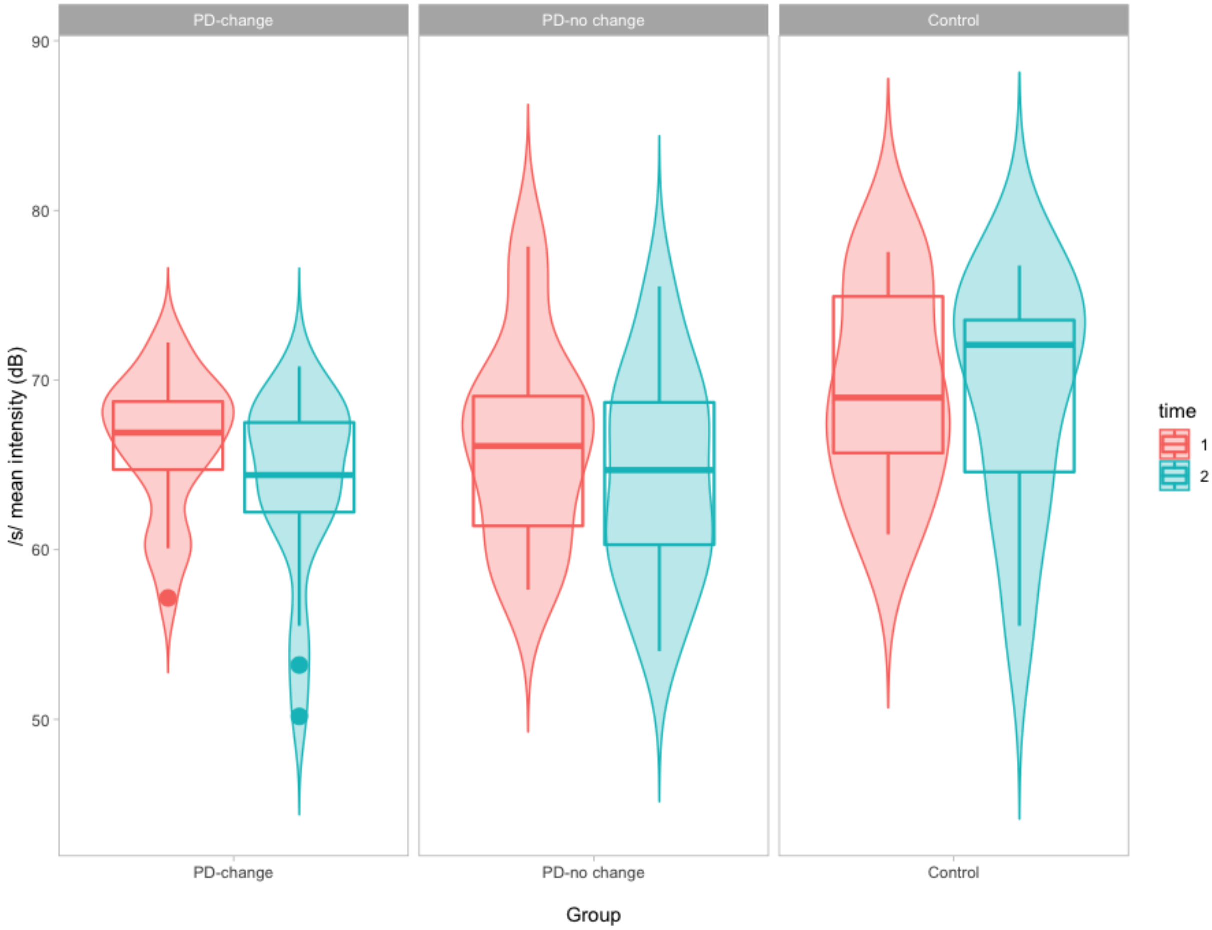
5.2.7.4. /s/ mean intensity

Descriptive statistics of /s/ mean intensity, reported in Table 32, shows that both PwPD groups have lower means compared to the control group at both T1 and T2, and show a negative change in both PwPD groups while the control group remains relatively stable over time. The PwPD-change group has a slightly greater negative change in means of 2.39dB compared to the PwPD-no change group (2.25dB), but this may not be statistically significant. In addition, The PwPD-no change group has a higher variance of distribution at both T1 (SD = 5.92) and T2 (SD = 6.04) compared to the PwPD-change group at T1 (SD = 4.03) and T2 (SD = 5.42). The plotted distributions for all three groups in Figure 21 suggest a large change in the control group but this is mased on the median which might be influenced by the variance in distribution in the control group. The distribution of values seems smaller with the PwPD-change group compared to the to the other groups, but it unclear from the plot or the descriptive statistics whether these changes are suggesting any significance.

Table 32. Summary table of descriptive statistics of /s/ mean intensity for each group at T1 and T2. N = 21 (PwPD-change); 21 (PwPD-no change); 10 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	69.46	6.00	68.97	60.91	77.55
Control (T2)	69.38	6.94	72.08	55.52	76.75
PwPD-change (T1)	66.32	4.03	66.90	57.16	72.22
PwPD-change (T2)	63.93	5.42	64.40	50.18	70.81
PwPD-no change (T1)	66.48	5.92	66.11	57.65	77.87
PwPD-no change (T2)	64.23	6.04	64.69	54.03	75.52

Figure 21. The distribution of /s/ mean intensity of each group (PwPD-change, PwPD-no change, Control) at T1 and T2.



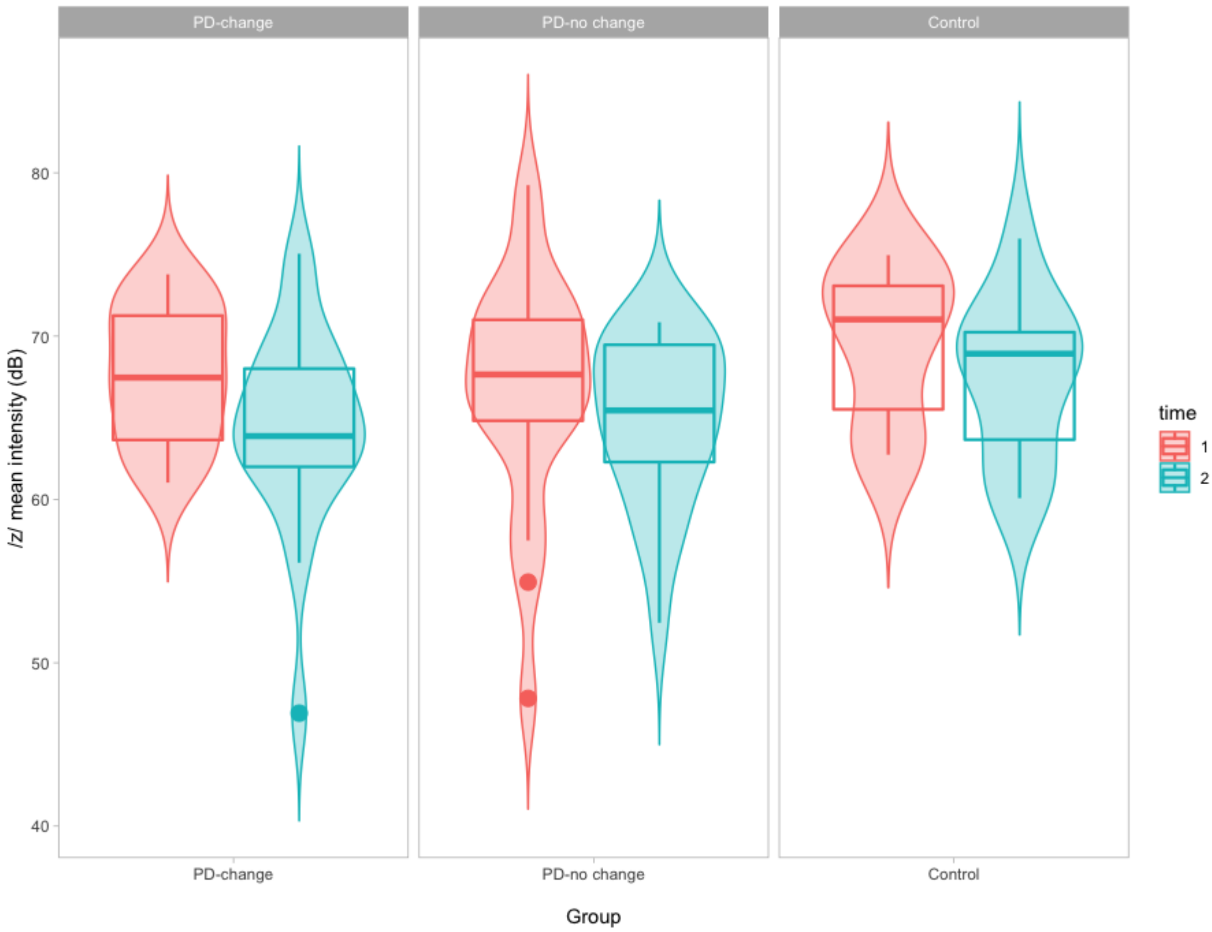
5.2.7.5. /z/ mean intensity

The /z/ mean intensity descriptive statistics presented in Table 33 indicates that for this fricative, the means in both the PwPD groups are lower at both T1 and T2 compared to the control group. However, the PwPD-change groups has a higher variance in distribution at T2 (SD = 6.09), and the PwPD-no change group has a higher variance of distribution at T1 (SD = 7.28). This can be seen with more clarity in Figure 22 below, showing a greater negative change in the PwPD-change group compared to the other groups.

Table 33. Summary table of descriptive statistics of /z/ mean intensity for each group at T1 and T2. N = 21 (PwPD-change); 21 (PwPD-no change); 10 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	69.70	4.78	71.02	62.74	74.96
Control (T2)	67.52	4.99	68.92	60.08	75.97
PwPD-change (T1)	67.70	4.13	67.46	61.03	73.78
PwPD-change (T2)	64.44	6.09	63.88	46.91	75.05
PwPD-no change (T1)	66.72	7.28	67.65	47.82	79.24
PwPD-no change (T2)	64.80	5.08	65.45	52.45	70.85

Figure 22. The distribution of /z/ mean intensity of each group (PwPD-change, PwPD-no change, Control) at T1 and T2.



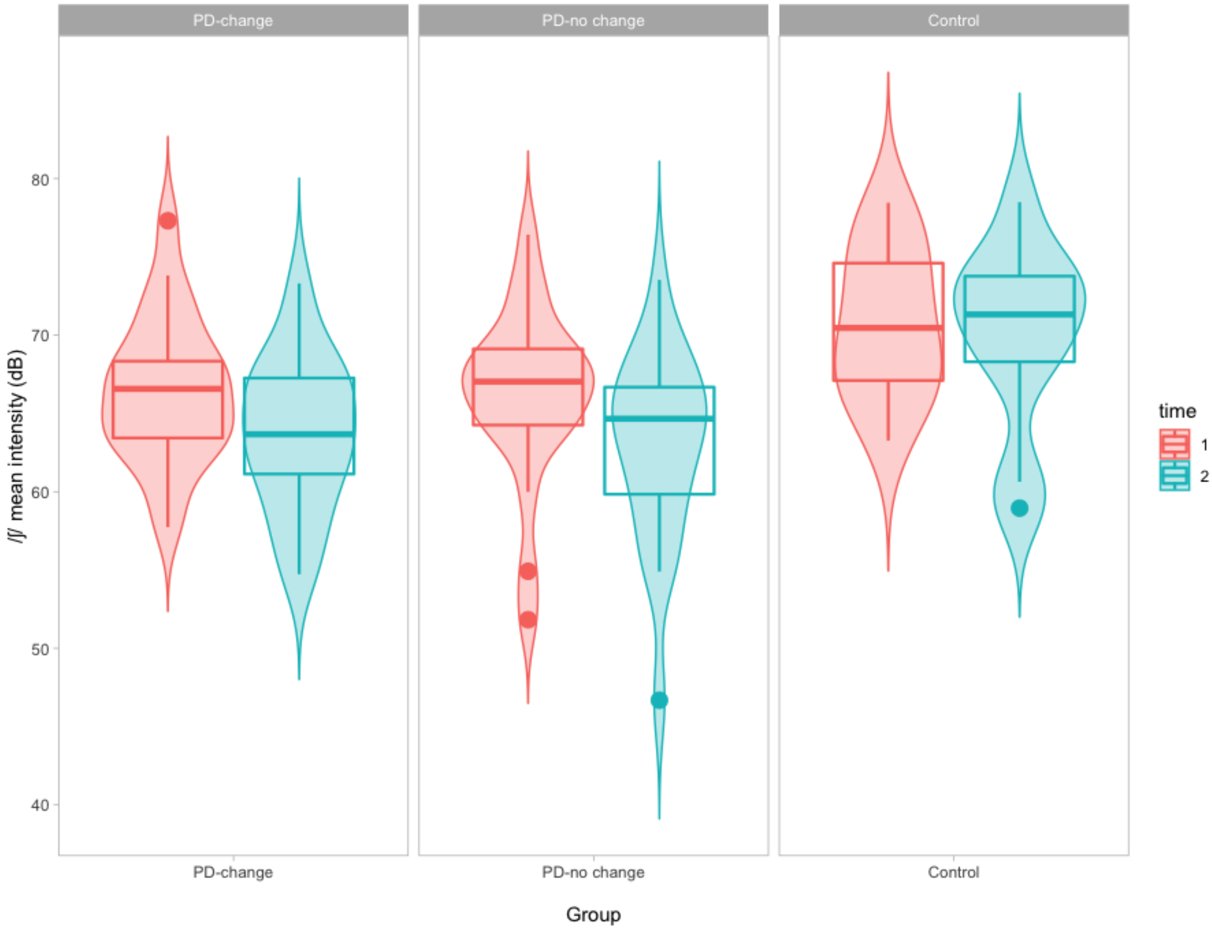
5.2.7.6. /ʃ/ mean intensity

Results of the descriptive statistics on /ʃ/ mean intensity presented in Table 34 below shows that mean values for both the PwPD groups is lower than the control group at both T1 and T2. This is consistent with the means of /s/ mean intensity which was lower for both PwPD groups compared to the control as well. There is also a greater negative change in both the PwPD groups compared to the control group over time and is illustrated in Figure 23 below.

Table 34. Summary table of descriptive statistics of /ʃ/ mean intensity for each group at T1 and T2. N = 21 (PwPD-change); 21 (PwPD-no change); 10 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	70.74	4.88	70.47	63.27	78.46
Control (T2)	69.93	6.10	71.33	58.95	78.50
PwPD-change (T1)	66.47	4.70	66.57	57.74	77.31
PwPD-change (T2)	63.69	4.92	63.67	54.73	73.30
PwPD-no change (T1)	66.19	5.62	67.04	51.83	76.42
PwPD-no change (T2)	62.88	6.13	64.67	46.69	73.53

Figure 23. The distribution of /ʃ/ mean intensity of each group (PwPD-change, PwPD-no change, Control) at T1 and T2.



5.2.8. Results of LMMs of fricatives

Results of the LMMs run on all fricative intensities were conducted similarly to those of the plosives reported above, using the same model components. A summary table of the LMM results for all the fricative acoustic parameters can been seen in Table 35 below, based on 104 observations from 52 participants.

The results of the LMMs for average fricatives mean intensity, /f/, /v/, and /z/ mean intensity indicated that the effect of the predictor *time* on each of these acoustic parameters was significant. All these acoustic parameters showed a significant negative change in intensity over time. However, the effect of the predictor *group* on these parameters was not significant, suggesting that the negative change occurred across all groups from T1 to T2. The descriptive statistics reported in the previous section suggests that the negative change in all the above acoustic parameters might have been driven by the PwPD groups, and likely a greater negative change in the PwPD-change group compared to the other groups.

The results of the LMMs also indicated that the effect of both the predictors *time* and *group* had a significant impact on the acoustic parameters /s/ and /ʃ/ mean intensity indicating that the values in these acoustic parameters had a significant negative change. The negative change in /s/ intensity was an estimated 1.89dB over time, and the negative change in /ʃ/ intensity was an estimated 2.53dB over time. Both these acoustic parameters also differed significantly from the control group with the /s/ mean intensity in the PwPD-change group being an estimated 4.29dB lower than the control group and the PwPD-no change group an estimated 4.06dB lower than the control group. The /ʃ/ mean intensity in the PwPD-change group was an estimated 5.26dB lower than the control group and the PwPD-no change group was an estimated 5.69dB lower than the control group. However, it is unclear if the PwPD groups differed significantly from each other and post-hoc analysis in section 5.2.11 will provide more insight.

Table 35. Linear mixed effects model results of the PFR analysis of all fricatives against each of the levels of the predictors time and group.

Fixed Effects:						
Acoustic parameter	**Predictor**	**Estimate**	**Std. Error**	**df**	**t-value**	**p**
Avg. fricative intensity	Intercept (Control, T1)	69.82	1.29	56.19	54.32	< 0.01*
	T2	-2.47	0.67	51.00	-3.67	0.001**
	PwPD-change	-2.40	1.51	49.00	-1.59	0.12
	PwPD-no change	-2.91	1.51	49.00	-1.92	0.06
/f/ intensity	Intercept (Control, T1)	62.66	1.72	54.12	36.42	< 0.01*
	T2	-2.16	0.77	51.00	-2.81	0.007**
	PwPD-change	1.31	2.04	49.00	0.64	0.52
	PwPD-no change	0.81	2.04	49.00	0.40	0.69
/v/ intensity	Intercept (Control, T1)	70.64	1.61	59.72	43.86	< 0.01*
	T2	-2.74	1.02	51.00	-2.70	0.009**
	PwPD-change	-0.95	1.86	49.00	-0.51	0.61
	PwPD-no change	-0.61	1.86	49.00	-0.33	0.75
/s/ intensity	Intercept (Control, T1)	70.36	1.59	55.14	44.28	<0.01*
	T2	-1.89	0.77	51.00	-2.45	0.02*
	PwPD-change	-4.29	1.87	49.00	-2.29	0.03*
	PwPD-no change	-4.06	1.87	49.00	-2.17	0.04*
/z/ intensity	Intercept (Control, T1)	69.86	1.42	63.50	49.38	<0.01*
	T2	-2.50	1.02	51.00	-2.44	0.02*
	PwPD-change	-2.56	1.60	49.00	-1.60	0.12
	PwPD-no change	-2.85	1.60	49.00	-1.78	0.08

/ʃ/ intensity	Intercept (Control, T1)	71.60	1.54	53.90	46.45	< 0.01**
	T2	-2.53	0.72	50.09	-3.51	0.001**
	PwPD-change	-5.26	1.82	48.30	-2.89	0.01*
	PwPD-no change	-5.69	1.82	48.57	-3.12	0.003**

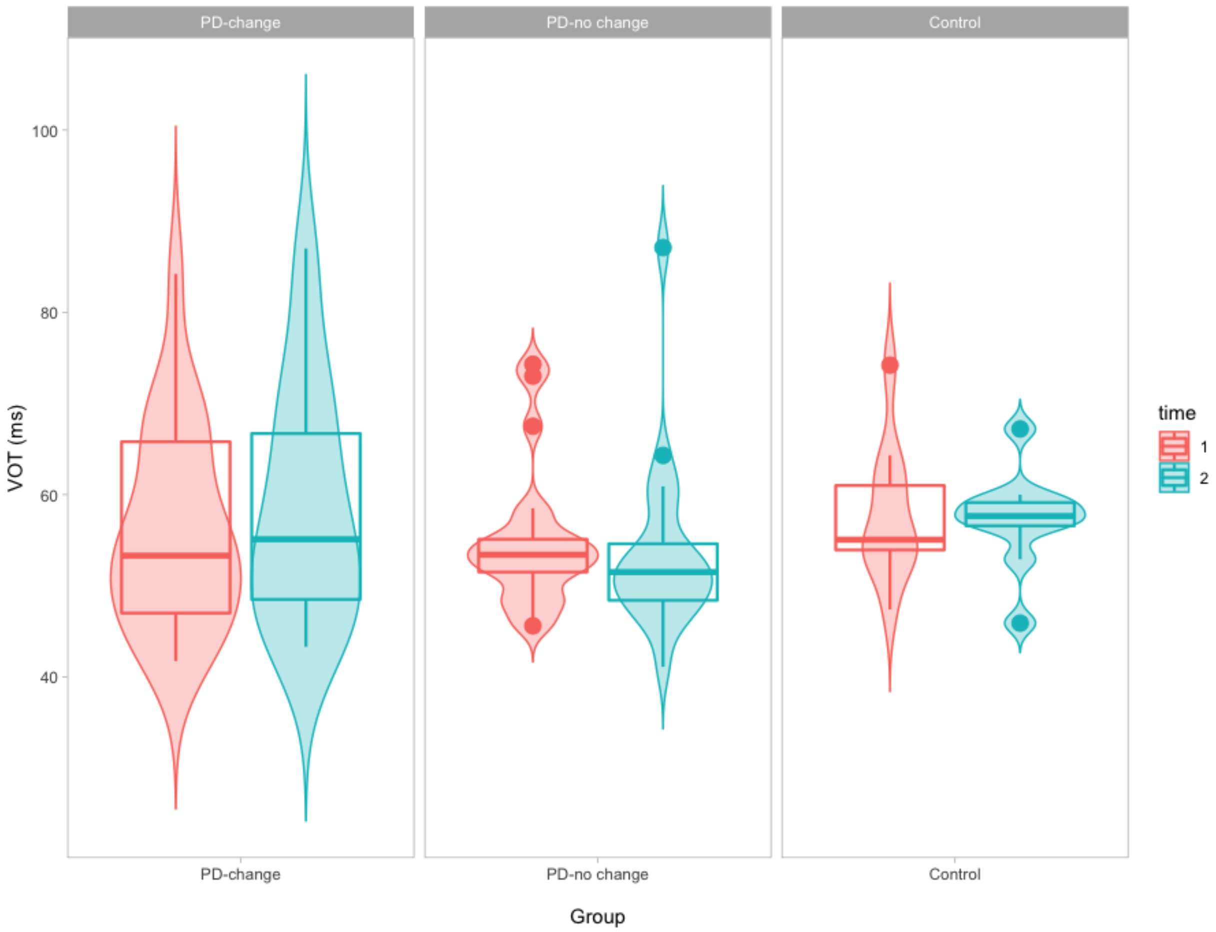
5.2.9. Descriptive statistics of VOT

Since the fricatives and plosive results have been reported, the descriptive statistics of the final acoustic parameter in the PFR analysis, VOT, is presented in Table 36 below. The VOT means for the control group is higher than those of both the PwPD groups at T1, but the PwPD-change group has a higher VOT at T2 (58.66ms, SD = 13) compared to the PwPD-no change group which had a negative change in VOT at T2 (53.64ms, SD = 9.47). However, looking at the plot of VOT distributions of all three groups in Figure 24, there is a greater variance in distribution in the PwPD-change group compared to the other two groups which suggests it may influence the degree to which VOT changes over time and between groups. This will be further reported on in the following section on the statistical analysis of VOT.

Table 36. Summary table of descriptive statistics of VOT for each group at T1 and T2. N = 21 (PwPD-change); 21 (PwPD-no change); 10 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	57.84	7.56	55.05	47.40	74.20
Control (T2)	57.27	5.40	57.65	45.90	67.20
PwPD-change (T1)	56.05	11.04	53.30	41.74	84.20
PwPD-change (T2)	58.66	13.00	55.10	43.30	87.00
PwPD-no change (T1)	55.00	7.80	53.40	45.60	74.30
PwPD-no change (T2)	53.64	9.47	51.50	41.10	87.10

Figure 24. The distribution of VOT of each group (PwPD-change, PwPD-no change, Control) at T1 and T2.



5.2.10. Results of LMM on VOT

Results of the LMM of VOT is presented in Table 37 below based on 104 observations of 52 participants. The results indicate that the effect of both the predictors *time* and *group* was not significant, implying that VOT did not change significantly over time and was not significantly different between groups. The descriptive statistics presented in the previous section suggested that the higher variance in distribution of VOT in the PwPD-change group may skew the degree to which VOT may differ between groups, or whether the change over time may be significant. The results of the LMM suggest no significance but it is possible the PwPD-change group may have influenced the results, though this cannot be confirmed. The implications of this result and the previous plosive and fricative results will be discussed in section 5.2.12.

Table 37. Linear mixed effects model results of the PFR analysis of VOT against each of the levels of the predictors time and group.

Fixed Effects:						
Acoustic parameter	**Predictor**	**Estimate**	**Std. Error**	**df**	**t-value**	**p**
VOT	Intercept (Control, T1)	57.36	2.92	52.67	19.65	<0.01**
	T2	0.340	1.11	51.00	0.36	0.72
	PwPD-change	-0.20	3.48	49.00	-0.06	0.96
	PwPD-no change	-3.24	3.48	49.00	-0.93	0.36

5.2.11. Post-hoc tests

Estimated marginal means (EMMs) were obtained using the "emmeans" package (Lenth, Russel V., 2022) in R (R Core Team, 2019). The package used Kenward-roger method to calculate degrees of freedom. This was used to compare the predictor *group* and *time* in the relevant articulation models where both predictors were significant to pinpoint where exactly the significance lies. These were the models for the fricatives /s/ and /ʃ/ mean intensity.

EMMs were run to conduct a pairwise comparison of all three *groups* and the results automatically averaged (collapsed) over the two levels of the predictor *time* with a confidence level of 95%, giving the mean response value of each level of the predictor *group,* and contrasting it with the other levels. This would help identify how different each group is from the other. A pairwise comparison of the two levels of *time* was also done averaged over the predictor *group.*

Results of the EMMs for each significant acoustic parameter are discussed below.

5.2.11.1. /s/ mean intensity

Table 38 shows the results of the EMMs and indicate there is no significant difference between the Control and PwPD-change group between the two levels of *time* (E= 4.29, df=49, p >0.05), the Control and PwPD-no change group (E= 4.06, df=49, p >0.05), or between the PwPD-change and PwPD-no change group (E=-0.23, df=49, p >0.05). There was a significant difference between the two data collection *time* points (E= 1.89, df=51, p <0.05). These results partially confirm the results in the LMM indicating that /s/ mean intensity changed significantly over time. However, the EMMs indicate that the significant difference between groups in /s/ mean intensity reported in the LMM is not present here. The implications of this will be discussed in the discussion (section 5.2.12).

Table 38. Pairwise differences of the predictors group and time for /s/ mean intensity.

Predictor comparisons	Estimate	SE	df	t-ratio	p
Control – PwPD-change	4.29	1.87	49	2.29	0.07
Control – PwPD-no change	4.06	1.87	49	2.17	0.09
PwPD-change – PwPD-no change	-0.23	1.50	49	-0.15	0.99
T1- T2	1.89	0.773	51	2.45	0.02*

Note: SE = standard error; df = degrees of freedom.

5.2.11.2. /ʃ/ mean intensity

As seen from Table 39 of the results of EMMs below, there is a significant difference between the Control and PwPD-change group between the two levels of time (E= 5.26, df= 48.6, p <0.05), and the Control and PwPD-no change group (E= 5.70, df= 48.9, p <0.05). However, there isn't a group difference between the PwPD-change and PwPD-no change group (E=0.44, df=49.0, p >0.05). There was a significant difference between the two data collection *time* points (E= 2.53, df=50.4, p <0.01). These results partly confirm the results of the LMM indicating that /ʃ/ mean intensity had a significant change over time and was significantly different between the PwPD groups and the control group. However, /ʃ/ mean intensity was not significantly different between the PwPD-change group and the PwPD-no change group. These results will be discussed in the next section below.

Table 39. Pairwise differences of the predictors group and time of /ʃ/ mean intensity.

Predictor comparisons	**Estimate**	**SE**	**df**	**t-ratio**	**p**
Control – PwPD-change	5.26	1.82	48.6	2.89	0.02*
Control – PwPD-no change	5.69	1.82	48.9	3.12	0.01*
PwPD-change – PwPD-no change	0.44	1.47	49.0	0.30	0.95
T1- T2	2.53	0.72	50.4	3.51	<0.01**

Note: SE = standard error; df = degrees of freedom.

5.2.12. Discussion

The PFR analysis was conducted to investigate the effectiveness of the acoustic parameters of plosives (average plosives intensity, /p/, /b/, /t/, /d/, /k/, /g/ intensity), fricatives (average fricatives intensity, /f/, /v/, /s/, /z/, /ʃ/ intensity), and VOT, at capturing perceptual changes in PwPD speech over time. This was done by conducting acoustic and statistical analyses (LMMs) on speech data collected over two time points, six months apart, from the same group of participants. The models for the LMMs were run with each acoustic parameter (the outcome variable) modelled against the predictors *time* (indicating the data collection time) and *group* (indicating the grouping of recordings into either PwPD-change, PwPD-no change, and control).

The results of statistical analysis revealed that a number of acoustic parameters showed a negative change in PwPD perceptual features over time. These acoustic parameters were average plosives intensity, /b/, /t/, /d/, /k/, and /g/ intensity, average fricatives in intensity, /f/, /v/, /s/, /z/, and /ʃ/ intensity. Of these acoustic parameters only /s/ and /ʃ/ intensity also showed a significant difference in intensity between PwPD speech and control speech. The acoustic parameters /p/ intensity and VOT were unable to track PwPD perceptual changes over time. The interpretations of these results are discussed below.

The acoustic parameters /s/ and /ʃ/ intensity showed a statistically significant negative change over time and was statistically significant between groups. The statistically significant for lower intensity in PwPD speech compared to control speech concurs with previous research (Duffy, 2020; Y. Kim, 2017) showing that sibilants in particular are largely impacted in the speech of PwPD. Post-hoc tests confirmed that both /s/ and /ʃ/ intensity changed significantly over time, and /ʃ/ intensity was also significantly different between groups. This results indicates that /ʃ/ intensity is able to track PwPD perceptual changes over time suggesting that there may be some articulatory undershoot (Ackermann & Ziegler, 1991; Duffy, 2020; Pawlukowska et al., 2015) in this fricative more than others preventing PwPD from maintain the correct postures to execute this consonant. However, it should be noted that /ʃ/ had the least number of occurrences in The Grandfather Passage which may influence the legitimacy of these results. It is possible that will more occurrences of /ʃ/ in the recording and used for analysis, the results may vary. However, it can be stated, that based on the occurrence of /ʃ/ in both repetitions of The Grandfather Passage, that there is a significant negative change in PwPD performance over time.

However, the post-hoc results for /s/ intensity indicated that the estimated means of both PwPD and control groups were not significantly different from one another. This suggests that the significant effect of the predictor group on /s/ intensity was not found when averaged over the factor time which implies that /s/ intensity was significantly different between groups at each of the time points, but not when averaged over time points, indicating that all groups showed some change in /s/ intensity over time. This result is consistent with the changes in the means of the descriptive statistics where both PwPD groups showed a negative change over time, and the control group showed a positive change over time. The acoustic parameters average plosives intensity, /b/, /t/, /d/, /k/, and /g/ intensity, average fricatives in intensity, /f/, /v/, /s/, /z/, and /ʃ/ intensity also indicated only a negative change over time and no significant differences in intensity values between groups. These findings contradict the SLT ratings which suggested that only the PwPD-change group should show increased imprecise consonant production over time. It might indicate that the aforementioned acoustic parameters were not severely impacted in the PwPD groups and that they do not contribute to the perceived increase imprecise consonant production.

However, intensity in the acoustic parameters largely had negative changes in both PwPD groups but with the control group largely remaining stable or showing a slight increase. This suggests some deterioration in PwPD groups though not significant. This insignificant change over time may also indicate that changes in intensity across all groups may be the result of a natural variation in speech in all participants, unrelated to the perceptual feature imprecise consonant production and articulatory breakdown. The present study cannot substantiate this claim and a baseline of natural speech variation in the control group could be investigated in the future.

The lack of a significant change in VOT over time contradicts previous studies which showed that VOT is significantly different between PwPD and control speakers (Argüello-Vélez et al., 2020; Fischer & Goberman, 2010; Rusz et al., 2021). However, since VOT can often be rate-dependent, which is linked to PD severity, the lack of these findings may be due to mild severity within the PwPD speakers. VOT tends to be more distinct from controls in more severe cases of PD (Rusz et al., 2021); if speech severity among PwPD is mild, then VOT could be similar between groups. This claim will be detailed in the IG analysis in the next section. In addition, it is possible that VOT is unable to capture the perceptual feature imprecise consonant production.

Finally, a potential reason for the trend in the acoustic parameters to show a change in intensity over time with an insignificant group difference is that all parameters were extracted across the reading passage and previous literature (Kuo & Tjaden, 2016) suggests that dysarthric speech (including PwPD) changes over the course of a reading passage, often with acoustic parameters indicating more deterioration at the end compared to the beginning of the task. This deterioration in parameters over the course of the reading task could be related to articulatory fatigue (Solomon, 2000; Vayra & Fowler, 1992) and which could impact the ability of participants motor function throughout the task, though this has not been substantiated in PwPD.

5.3. Intelligibility group (IG) analysis

5.3.1. Participant demographics

Analysis was conducted on the complete dataset (n =110) to test for a change in PwPD speech intelligibility. The recordings were grouped based on the overall intelligibility rating based on the SLT ratings and grouping PwPD speech into mild (rated 0 or 1), moderate (rated 2 or 3), and severe (rated 4). All the PwPD recordings were found to be rated as mild indicating only mild impairment to speech intelligibility. In addition, the overall intelligibility rating was checked for both T1 and T2 data collection time points and grouped based on whether the SLT rating had changed (either increased or decreased) in T2. The final groups for this analysis are presented in Table 40 below:

Table 40. Participant demographics for the IG analysis.

Group	**Number of participants**	**Age range (years)**	**Mean age (years)**
Mild-change	N= 34; M = 22; F = 12	52-81	66 (SD = 7.09)
Mild-no change	N= 29; M = 21; F = 8	50-93	72 (SD = 9.16)
Control	N= 47; M = 17; F = 30	35-86	64 (SD = 12.23)

5.3.2. Segmentation of and extraction of parameters

All plosives and fricatives were annotated from the grandfather passage using Praat (Boersma & Weenink, 2012) using the same method employed for the analysis based on testing changes in perceptual feature ratings. Extraction of the parameters remained the same as well.

5.3.3. Reliability of annotations

The reliability of the annotations was check for the IG analysis similarly to the PFR analysis. This in involved manually re-examining 10% of textgrids by the primary researcher and the same external researcher used for the PFR analysis. Annotations for the plosives and fricatives were manually rechecked and any discrepancies were corrected.

5.3.4. Descriptive statistics of plosives

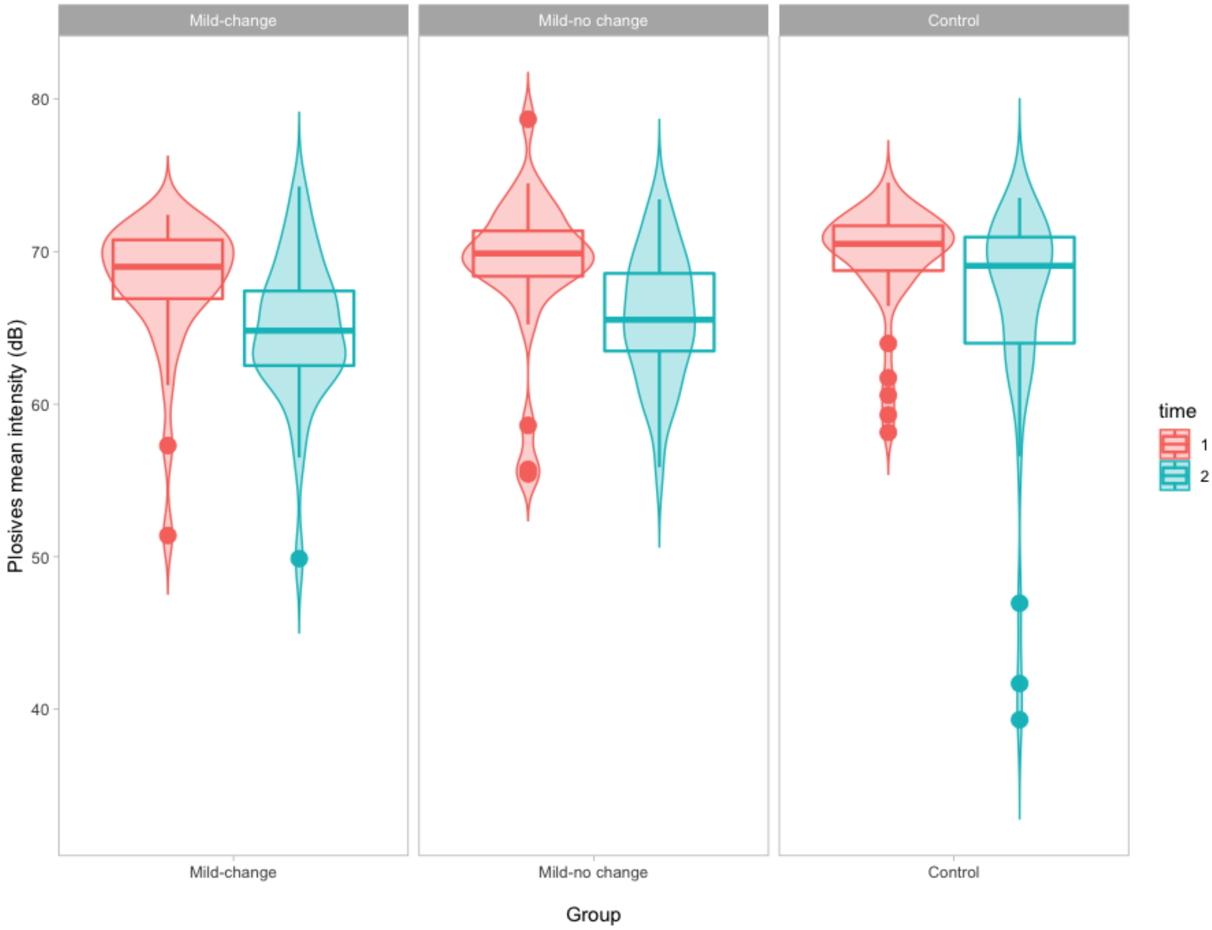
5.3.4.1. Average plosives mean intensity

The descriptive statistics for the average plosives mean intensity in Table 41 below suggests that all three groups show a negative change in means over time. The means also suggest that the there is a greater negative change over time in the PwPD groups compared to the control group. However, the table suggests that the difference in means between groups may not be significant. This is illustrated better in the plot of the distribution for all groups in Figure 25 which also shows that the control group may have greater variance in values at T2 which may be causing the negative change.

Table 41. Summary table of descriptive statistics of average plosives mean intensity for each group at T1 and T2. N = 34 (mild-change); 29 (mild-no change); 47 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	69.60	3.57	70.51	58.15	74.53
Control (T2)	66.36	7.40	69.07	39.31	73.53
Mild-change (T1)	67.92	4.37	69.01	51.39	72.41
Mild-change (T2)	64.89	4.77	64.82	49.87	74.27
Mild-no change (T1)	69.00	5.03	69.88	55.43	78.69
Mild-no change (T2)	65.74	4.24	65.54	55.87	73.43

Figure 25. The distribution of average plosives mean intensity of each group (Mild-change, Mild-no change, Control) at T1 and T2.



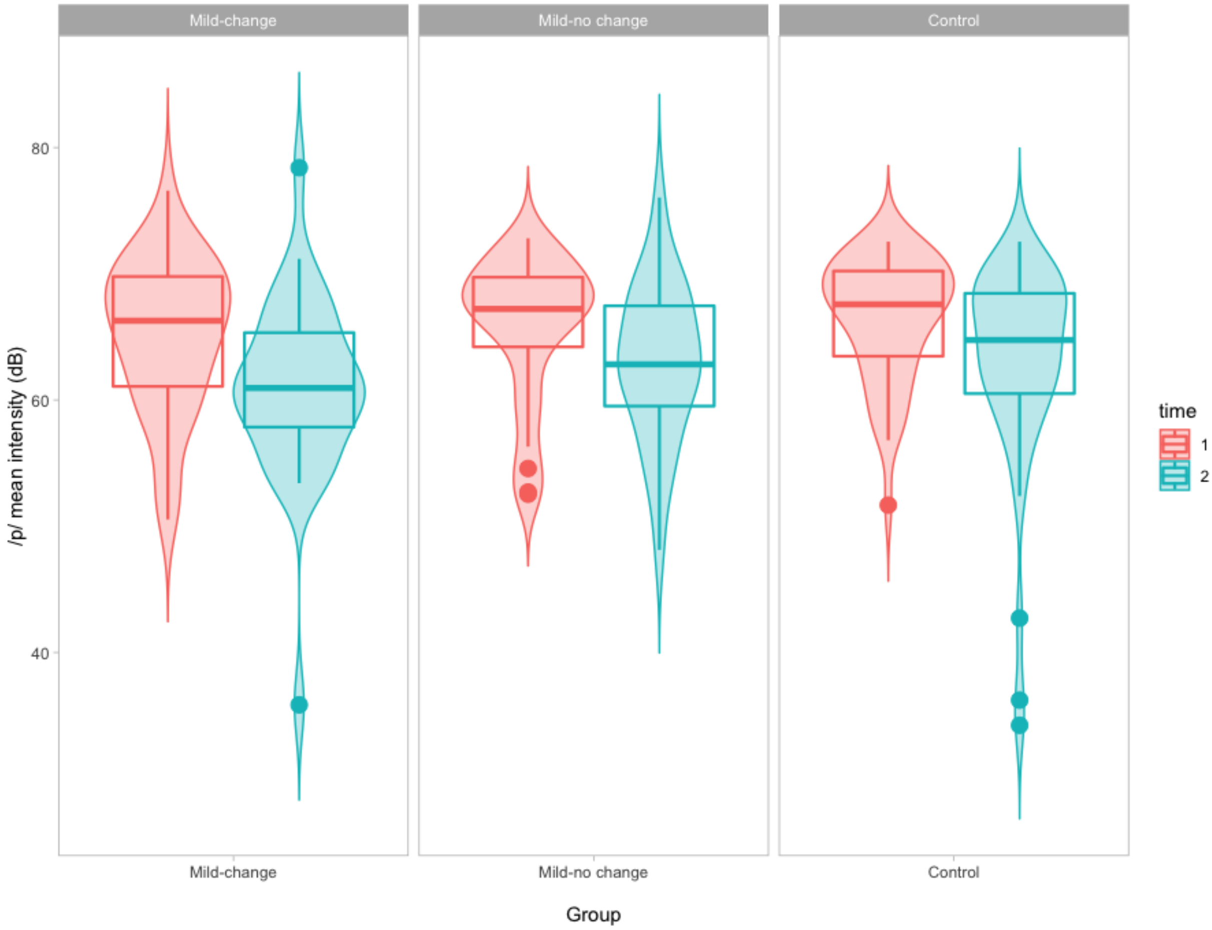
5.3.4.2. */p/ mean intensity*

In Table 42 below, the descriptive statistics of the acoustic parameter /p/ intensity is presented. The means of all three groups show a negative change over time and is similar to the descriptive statistics above of the average plosives intensity. These two acoustic parameters may behave similarly. The plotted distribution for all three groups in Figure 26 also shows a greater variance at T2 for the control group, similar to the average plosives intensity which may be causing the negative change.

Table 42. Summary table of descriptive statistics of /p/ mean intensity for each group at T1 and T2. N = 34 (mild-change); 29 (mild-no change); 47 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	66.55	4.82	67.58	51.68	72.55
Control (T2)	62.95	8.24	64.76	34.24	72.54
Mild-change (T1)	65.19	6.07	66.28	50.54	76.58
Mild-change (T2)	60.92	7.23	60.95	35.85	78.40
Mild-no change (T1)	65.79	5.787	67.22	52.54	72.81
Mild-no change (T2)	63.01	6.25	62.83	48.13	76.04

Figure 26. The distribution of /p/ mean intensity of each group (Mild-change, Mild-no change, Control) at T1 and T2.



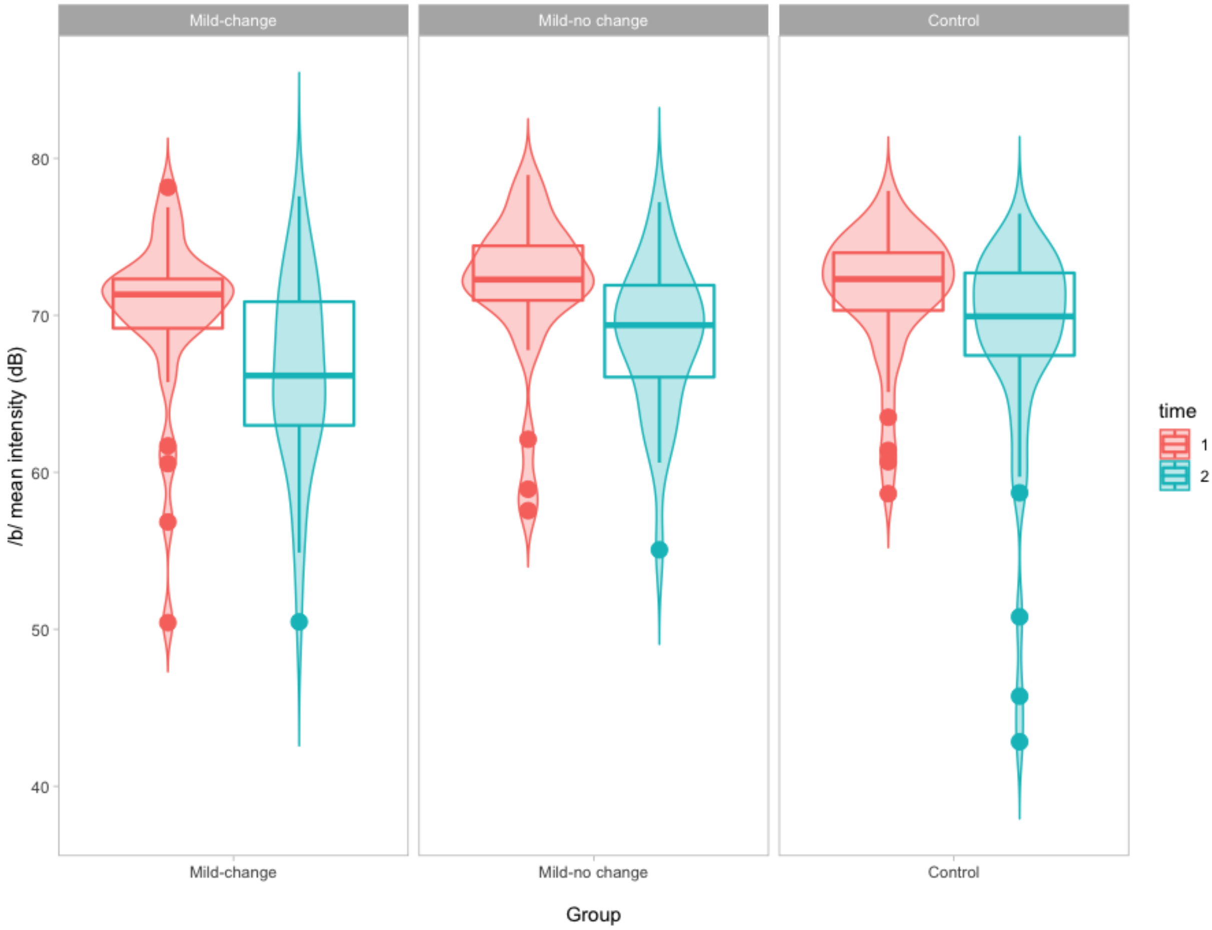
5.3.4.3. /b/ mean intensity

The descriptive statistics of /b/ intensity in Table 43 seems to follow the trend of a negative change observed in all three groups. However there appears to be a greater variance of distribution from T1 to T2 for both the mild-change group (T1: 69.97dB, SD = 5.53; T2: 66.25, SD = 6.12) and the control group (T1: 71.50dB, SD = 4.15; T2: 68.23, SD = 7.04) at T2 compared to T1 which may be influencing the negative change. This can be seen in the distributions plotted in Figure 27, which also shows that the mild-no change group has less variance in values that the other groups. Whether there might be a significant different between groups is unclear.

Table 43. Summary table of descriptive statistics of /b/ mean intensity for each group at T1 and T2. N = 34 (mild-change); 29 (mild-no change); 47 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	71.50	4.15	72.31	58.64	77.93
Control (T2)	68.23	7.04	69.93	42.84	76.48
Mild-change (T1)	69.97	5.53	71.33	50.43	78.15
Mild-change (time 2)	66.25	6.12	66.16	50.47	77.58
Mild-no change (T1)	71.70	4.96	72.28	57.56	78.94
Mild-no change (T2)	68.55	4.76	69.37	55.06	77.20

Figure 27. The distribution of /b/ mean intensity of each group (Mild-change, Mild-no change, Control) at T1 and T2.



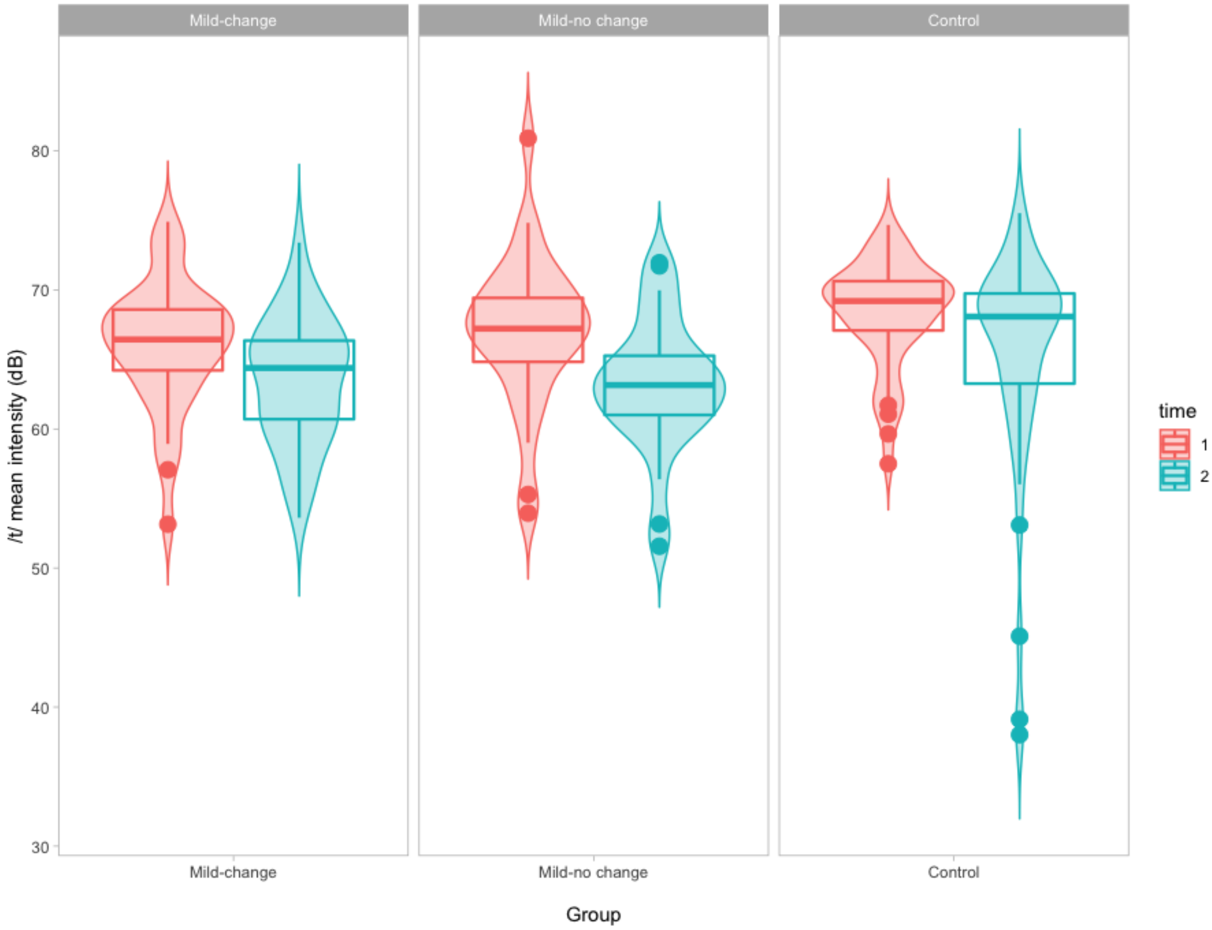
5.3.4.4. /t/ mean intensity

The descriptive statistics of /t/ intensity in Table 44 below shows that both PwPD groups means are lower than the control group means T1. The mild-change group has a lower mean (66.08dB, SD = 4.63) at T1 compared to the mild-no change group (67.07dB, SD = 5.50). However, both the mild-change group and mild-no change group seem to have a relatively similar decline in means at T2. The control group also shows a negative change over time but also shows a greater variance of distribution seen in Figure 28. It is difficult to say whether the three groups could be significantly different from each other.

Table 44. Summary table of descriptive statistics of /t/ mean intensity for each group at T1 and T2. N = 34 (mild-change); 29 (mild-no change); 47 (control).

	Mean	SD	Median	Min	Max
Control (T1)	68.37	3.80	69.20	57.50	74.68
Control (T2)	65.31	8.05	68.09	38.01	75.52
Mild-change (T1)	66.08	4.63	66.44	53.15	74.91
Mild-change (T2)	63.47	4.60	64.39	53.62	73.40
Mild-no change (T1)	67.07	5.50	67.22	53.95	80.91
Mild-no change (T2)	63.14	4.74	63.16	51.57	71.96

Figure 28. The distribution of /t/ mean intensity of each group (Mild-change, Mild-no change, Control) at T1 and T2.



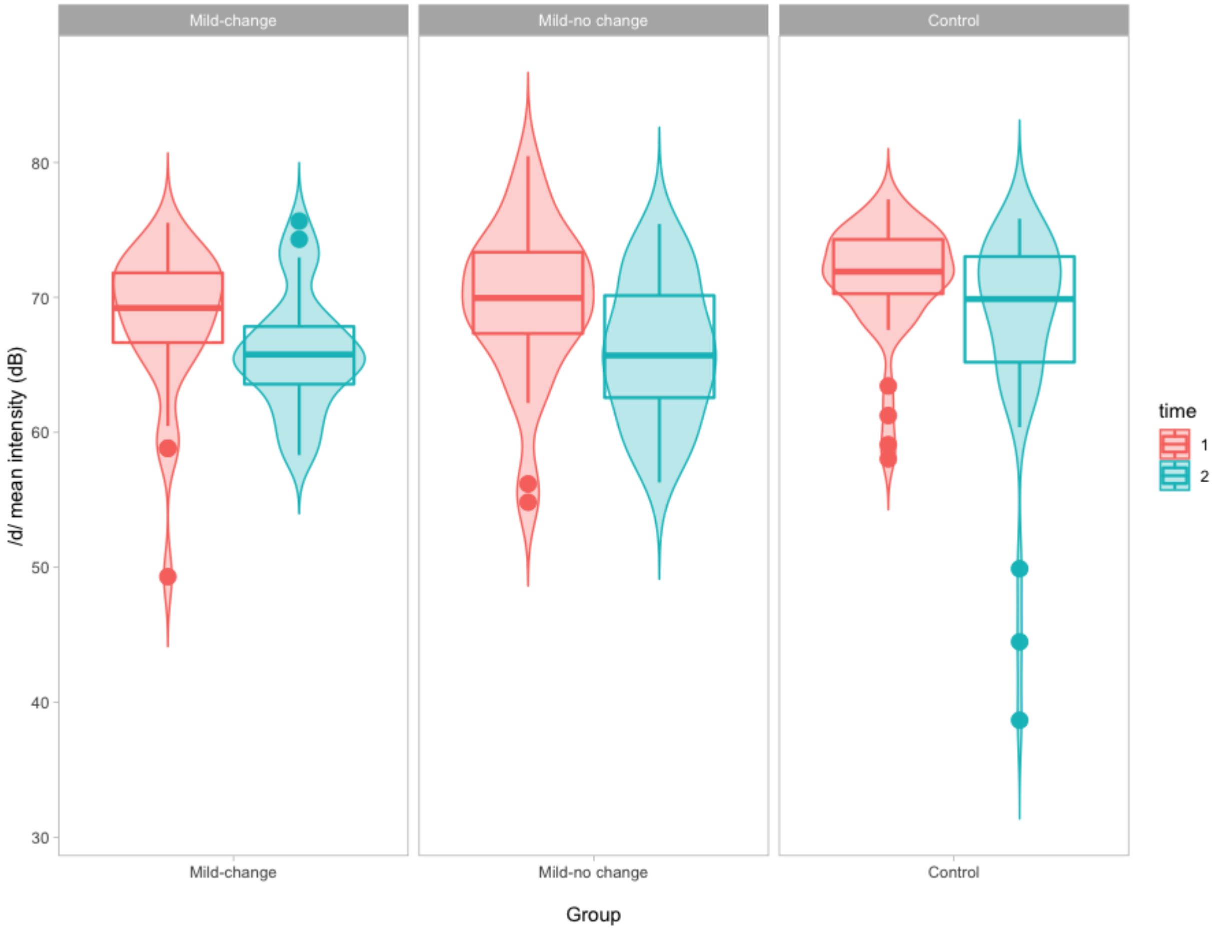
5.3.4.5. /d/ mean intensity

The /d/ intensity descriptive statistics shown in Table 45 follows a similar pattern to /t/ intensity where both PwPD groups have lower mean values than the control group. In addition, the mild-change group means (T1: 68.43dB, SD = 5.41; T2: 65.95dB, SD = 4.43) for /d/ intensity are also lower than the mild-no change group (T1: 69.63, SD = 5.82; T2: 66.04dB, SD = 5.20) at both T1 and T2. Similar to /t/ intensity descriptive statistics, the control group also shows a negative change in /d/ intensity, and based on Figure 29 of the plotted distributions, there is also a greater variance in distribution in the control group at T2.

Table 45. Summary table of descriptive statistics of /d/ mean intensity for each group at T1 and T2. N = 34 (mild-change); 29 (mild-no change); 47 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	71.39	4.18	71.91	58.03	77.27
Control (T2)	68.12	7.58	69.88	38.67	75.83
Mild-change (T1)	68.43	5.41	69.21	49.31	75.53
Mild-change (T2)	65.95	4.43	65.77	58.32	75.65
Mild-no change (T1)	69.63	5.82	69.96	54.82	80.47
Mild-no change (T2)	66.04	5.20	65.70	56.29	75.44

Figure 29. The distribution of /d/ mean intensity of each group (Mild-change, Mild-no change, Control) at T1 and T2.



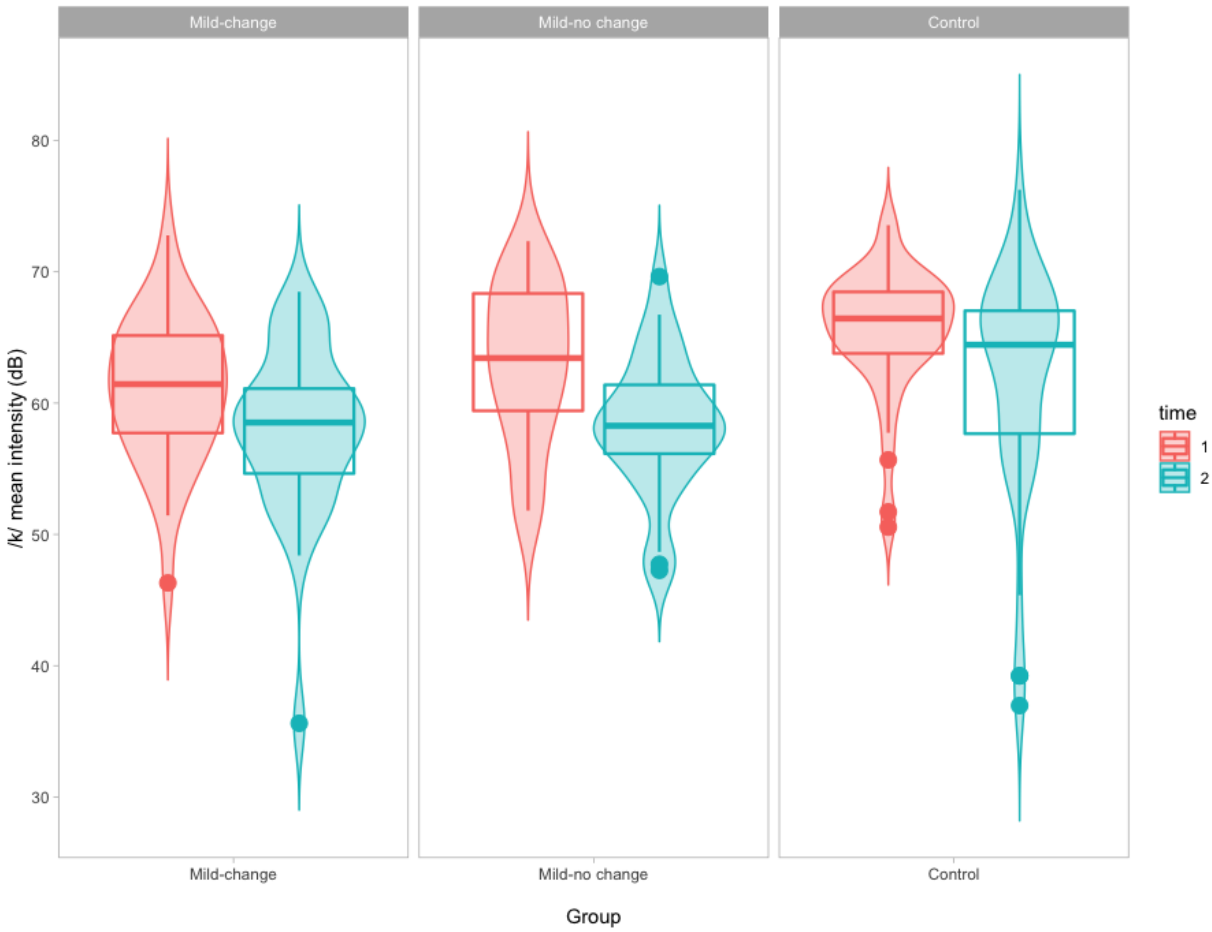
5.3.4.6. /k/ mean intensity

The /k/ intensity descriptive statistics shown in Table 46 follows a similar pattern to /d/, and /t/ intensity where both PwPD groups have lower mean values than the control group. Once again, the mild-change group means (T1: 61.02dB, SD = 5.53; T2: 57.88dB, SD = 6.57) for /k/ intensity are also lower than the mild-no change group (T1: 63.16, SD = 6.05; T2: 58.27dB, SD = 5.31) at both T1 and T2. Similar to /t/ and /d/ intensity descriptive statistics, the control group also shows a negative change in /k/ intensity and based on Figure 30 of with a greater variance in distribution in the control group at T2 which may influence the negative change.

Table 46. Summary table of descriptive statistics of /k/ mean intensity for each group at T1 and T2. N = 34 (mild-change); 29 (mild-no change); 47 (control).

	Mean	SD	Median	Min	Max
Control (T1)	65.57	4.75	66.44	50.55	73.55
Control (T2)	62.01	8.64	64.45	36.96	76.25
Mild-change (T1)	61.02	5.53	61.44	46.31	72.78
Mild-change (T2)	57.88	6.57	58.52	35.61	68.48
Mild-no change (T1)	63.16	6.05	63.43	51.82	72.33
Mild-no change (T2)	58.27	5.31	58.28	47.27	69.64

Figure 30. The distribution of /k/ mean intensity of each group (Mild-change, Mild-no change, Control) at T1 and T2.



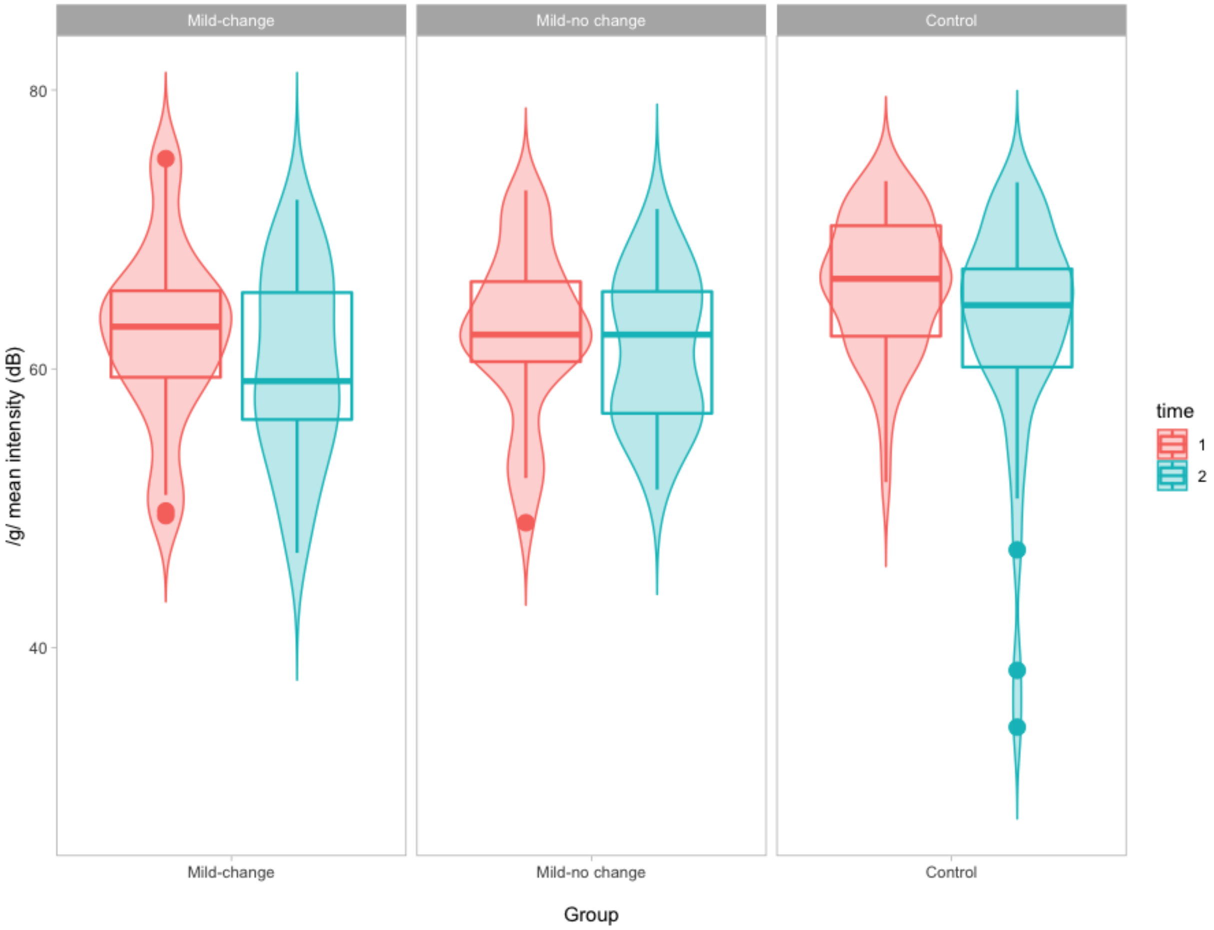
5.3.4.7. /g/ mean intensity

The /g/ intensity descriptive statistics presented below in Table 47 follows a similar pattern to the previously reported plosive intensities. Both PwPD groups have lower mean values than the control group. However, the mild-change group means at T1 (62.51dB, SD = 6.45) for /g/ intensity are relatively similar to the mild-no change group (69.63, SD = 5.82), and lower in the mild-change group at T2 (60.24dB, SD = 6.80) compared to the mild-no change group (66.04dB, SD = 5.20). Similar to the previous plosive intensities descriptive statistics, the control group also shows a negative change in /g/ intensity, and based on Figure 31 of the plotted distributions, there is also a greater variance in distribution in the control group at T2. The common trend in the plosive acoustic parameters makes it unclear whether group differences will be found but this is explored through the statistical analysis in the next section.

Table 47. Summary table of descriptive statistics of /g/ mean intensity for each group at T1 and T2. N = 34 (mild-change); 29 (mild-no change); 47 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	66.10	4.83	66.46	51.88	73.47
Control (T2)	63.12	8.06	64.57	34.31	73.38
Mild-change (T1)	62.51	6.45	63.03	49.48	75.08
Mild-change (T2)	60.24	6.80	59.13	46.81	72.13
Mild-no change (T1)	62.69	5.99	62.46	48.97	72.80
Mild-no change (T2)	61.44	5.45	62.46	51.35	71.47

Figure 31. The distribution of /g/ mean intensity of each group (Mild-change, Mild-no change, Control) at T1 and T2.



5.3.5. Results of LMMs of plosives

A linear mixed effects model analysis was run through R (R Core Team, 2019) to test the significance of each predictor on each acoustic parameter. The models were run with the lme4 package (Bates et al., 2015) using lmer. The model structures were the same as those using for the PFR analysis.

A model was run with each acoustic parameter (outcome variable) against the predictors *time* (data collection time points T1 and T2) and *intelligibility group* (mild-change, mild-no change, and control) and random factor of *participant* with a random intercept and random slope. A random intercept means that the final model takes into account that each individual *participant* can show higher or lower values for each acoustic parameter regardless of their *intelligibility group* or the data collection *time*. A random slope for *participant* means that the model considers that some between-participant variation exists.

Based on the assumptions of a linear mixed effects model, the residuals for each model run for each acoustic parameter was checked and found to be normally distributed, therefore no transformations of the parameters data were required. There were no significant correlations between any of the predictors.

A summary table of LMM results of all the plosive acoustic parameters can be seen in Table 48 below, based on 220 observations from 110 participants.

Results of the LLMs indicated that the effect of the predictor *time* on the acoustic parameters average plosives intensity, /p/, and /b/ mean intensity was significant, which shows that for these acoustic parameters there was a significant negative change over time. However, the effect of the predictor *group* on the same three acoustic parameters was not significant which indicates that the negative change in intensity occurred regardless of group. The descriptive statistics suggested the same trend across all three groups.

The results also indicated that the effect of both the predictors *time* and *group* on the acoustic parameters /t/, /d/, /k/, ang /g/ mean intensity was significant which means that for these acoustic parameters there was a significant negative change in intensity over time and the intensity was significantly different between groups. However, for the acoustic parameter /d/ intensity, the intensity values were not significantly different between the mild-no change group and the control group.

Table 48. Linear mixed effects model results of the IG analysis of all plosives against each of the levels of the predictors time and group.

Fixed Effects:						
Acoustic Parameter	**Predictor**	**Estimate**	**Std. Error**	**df**	**t-value**	**p**
Avg. plosives intensity	(Control, T1)	69.57	0.65	169.50	106.84	< 0.01*
	T2	-3.18	0.64	109.00	-4.96	<0.001**
	mild-change	-1.58	0.88	107.00	-1.80	0.08
	mild-no change	-0.61	0.92	107.00	-0.67	0.51
/p/ intensity	(Control, T1)	66.54	0.83	166.59	80.41	< 0.01*
	T2	-3.59	0.82	107.10	-4.40	<0.001**
	mild-change	-1.70	1.12	106.31	-1.51	0.13
	mild-no change	-0.35	1.17	104.45	-0.30	0.76
/b/ intensity	(Control, T1)	71.55	0.71	165.29	101.02	< 0.01*
	T2	-3.37	0.68	109.05	-4.97	<0.001**
	mild-change	-1.75	0.96	107.67	-1.82	0.07
	mild-no change	0.26	1.01	106.77	0.26	0.80
/t/ intensity	(Control, T1)	68.42	0.71	163.12	96.92	< 0.01*
	T2	-3.15	0.66	109.00	-4.76	<0.001**
	mild-change	-2.06	0.96	107.00	-2.14	0.03*
	mild-no change	-1.74	1.01	107.00	-1.72	0.09*

/d/ intensity	(Control, T1)	71.33	0.74	151.57	96.92	< 0.01*
	T2	-3.15	0.64	107.60	-4.93	<0.001**
	mild-change	-2.61	1.04	107.13	-2.51	0.01*
	mild-no change	-1.92	1.07	105.35	-1.79	0.08
/k/ intensity	(Control, T1)	65.70	0.83	152.84	78.90	< 0.01*
	T2	-3.83	0.76	102.36	-5.05	<0.001**
	mild-change	-4.38	1.15	103.50	-3.80	<0.001**
	mild-no change	-3.03	1.20	100.84	-2.53	0.01*
/g/ intensity	(Control, T1)	65.77	0.83	154.23	78.97	< 0.01*
	T2	-2.33	0.73	107.80	-3.18	0.002**
	mild-change	-3.28	1.16	106.49	-2.83	0.006**
	mild-no change	-2.55	1.21	105.64	-2.10	0.02*

5.3.6. Descriptive statistics of fricatives

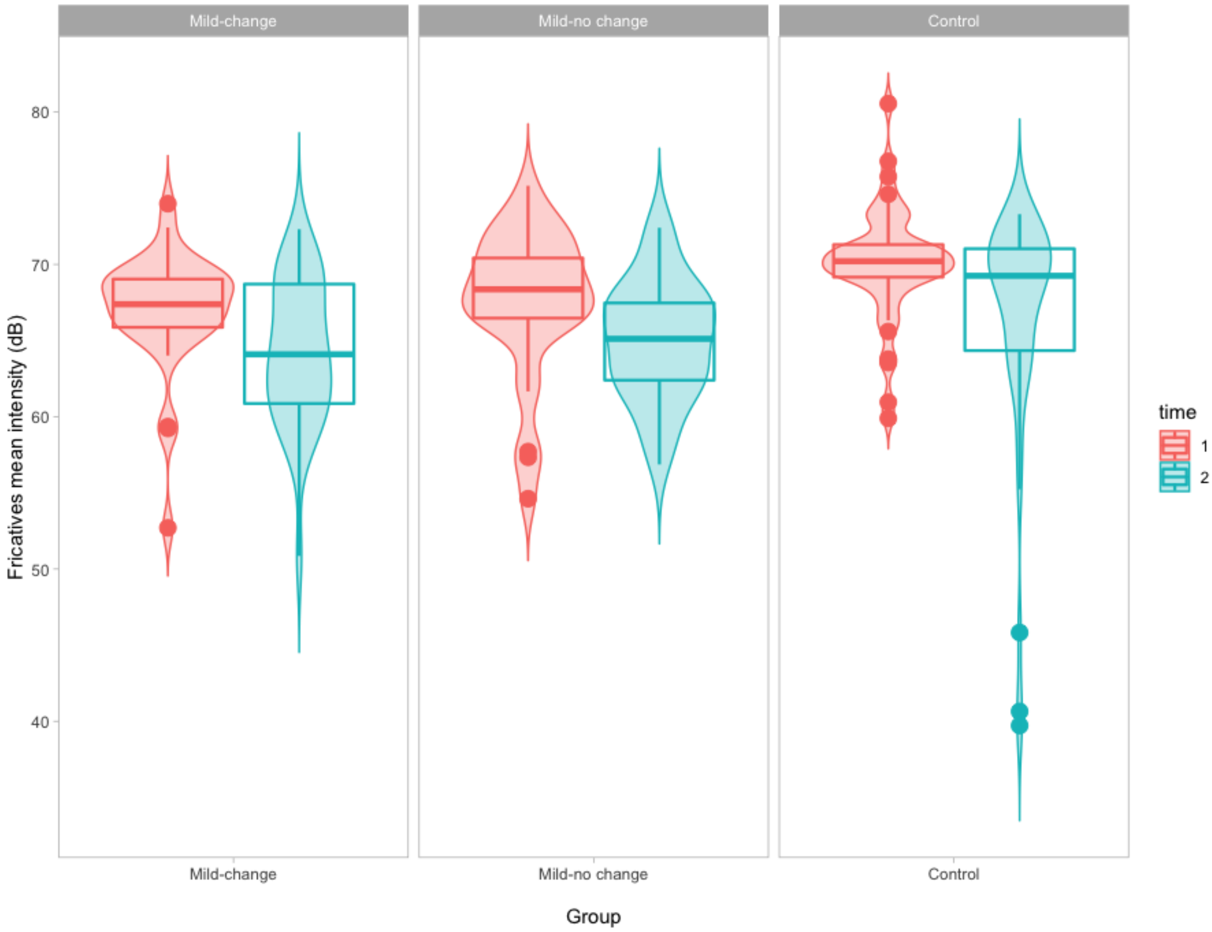
5.3.6.1. Average fricatives mean intensity

The average fricatives intensity descriptive statistics is presented in Table 49 below. The means indicate that both PwPD groups have lower mean values than the control group at both T1 and T2. The mild-change group shows a relatively similar negative change (by 2.83dB) compared to the mild-no change group (by 2.52dB), and there is a greater negative change in the control group by 3.53dB compared to the other groups. This may be due to the larger variance in distribution of control group values, illustrated in Figure 32.

Table 49 Summary table of descriptive statistics of average fricatives mean intensity for each group at T1 and T2. N = 34 (mild-change); 29 (mild-no change); 47 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	70.09	3.66	70.21	59.89	80.56
Control (T2)	66.56	7.68	69.25	39.72	73.30
Mild-change (T1)	67.06	3.89	67.39	52.70	73.99
Mild-change (T2)	64.23	4.74	64.10	50.86	72.31
Mild-no change (T1)	67.56	4.87	68.37	54.62	75.16
Mild-no change (T2)	65.04	3.91	65.11	56.87	72.41

Figure 32. The distribution of average fricatives mean intensity of each group (Mild-change, Mild-no change, Control) at T1 and T2.



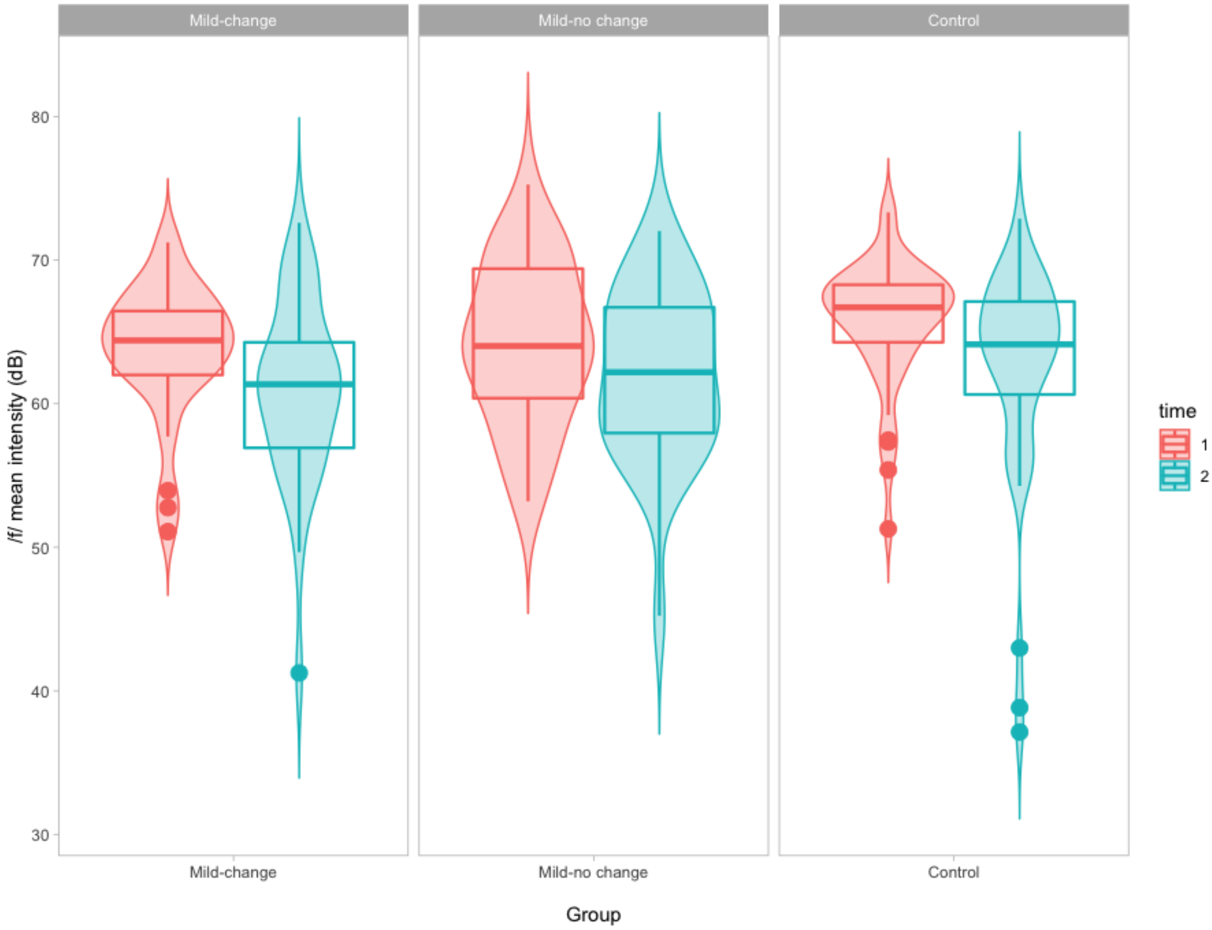
5.3.6.2. /f/ mean intensity

The descriptive statistics of /f/ intensity in Table 50 shows that both PwPD groups have lower mean values than the control group, similar to the means of average fricatives intensity. The control group shows the greatest negative change (by 3.33dB) compared to the mild-no change (2.19dB) and mild-change group (2.89dB), but the control group also shows the greater variance of values. However, whether these values may indicate a significant difference in unclear when viewing the distributions in Figure 33.

Table 50. Summary table of descriptive statistics of /f/ mean intensity for each group at T1 and T2. N = 34 (mild-change); 29 (mild-no change); 47 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	65.86	4.31	66.71	51.28	73.33
Control (T2)	62.53	7.51	64.13	37.13	72.88
Mild-change (T1)	63.79	4.64	64.41	51.09	71.23
Mild-change (T2)	60.90	6.35	61.35	41.25	72.60
Mild-no change (T1)	64.19	5.67	64.01	53.20	75.26
Mild-no change (T2)	62.00	5.98	62.19	45.23	72.02

Figure 33. The distribution of /f/ mean intensity of each group (Mild-change, Mild-no change, Control) at T1 and T2.



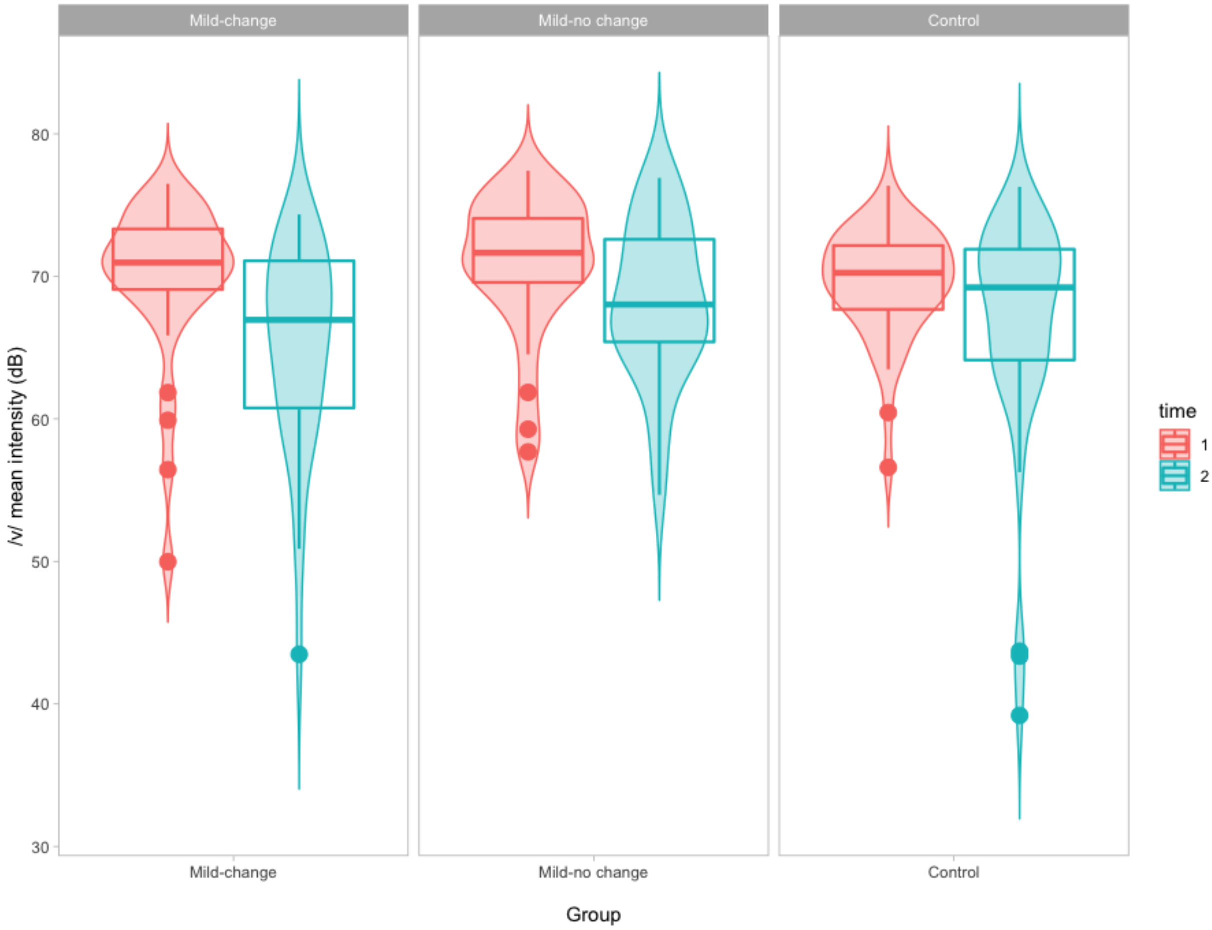
5.3.6.3. /v/ mean intensity

The descriptive statistics of /v/ intensity shown in Table 51 indicates that the mild-no change group had higher values at T1 (mean = 70.82, SD = 4.97) and at T2 (mean = 68.22, SD = 5.54) compared to the mild-change and control group. However, all three groups follow the now common negative trend in intensity over time. There seems to be a similar greater variation of values observed in Figure 34 for the mild-change and control group at both T1 and T2, and the plot does not clearly illustrate if a group difference will be present.

Table 51. Summary table of descriptive statistics of /v/ mean intensity for each group at T1 and T2. N = 34 (mild-change); 29 (mild-no change); 47 (control).

	Mean	SD	Median	Min	Max
Control (T1)	69.61	3.80	70.25	56.60	76.36
Control (T2)	66.76	7.99	69.22	39.20	76.28
Mild-change (T1)	69.88	5.61	70.97	49.98	76.49
Mild-change (T2)	65.10	7.09	66.95	43.48	74.35
Mild-no change (T1)	70.82	4.97	71.65	57.69	77.41
Mild-no change (T2)	68.22	5.54	68.02	54.67	76.90

Figure 34. The distribution of /v/ mean intensity of each group (Mild-change, Mild-no change, Control) at T1 and T2.



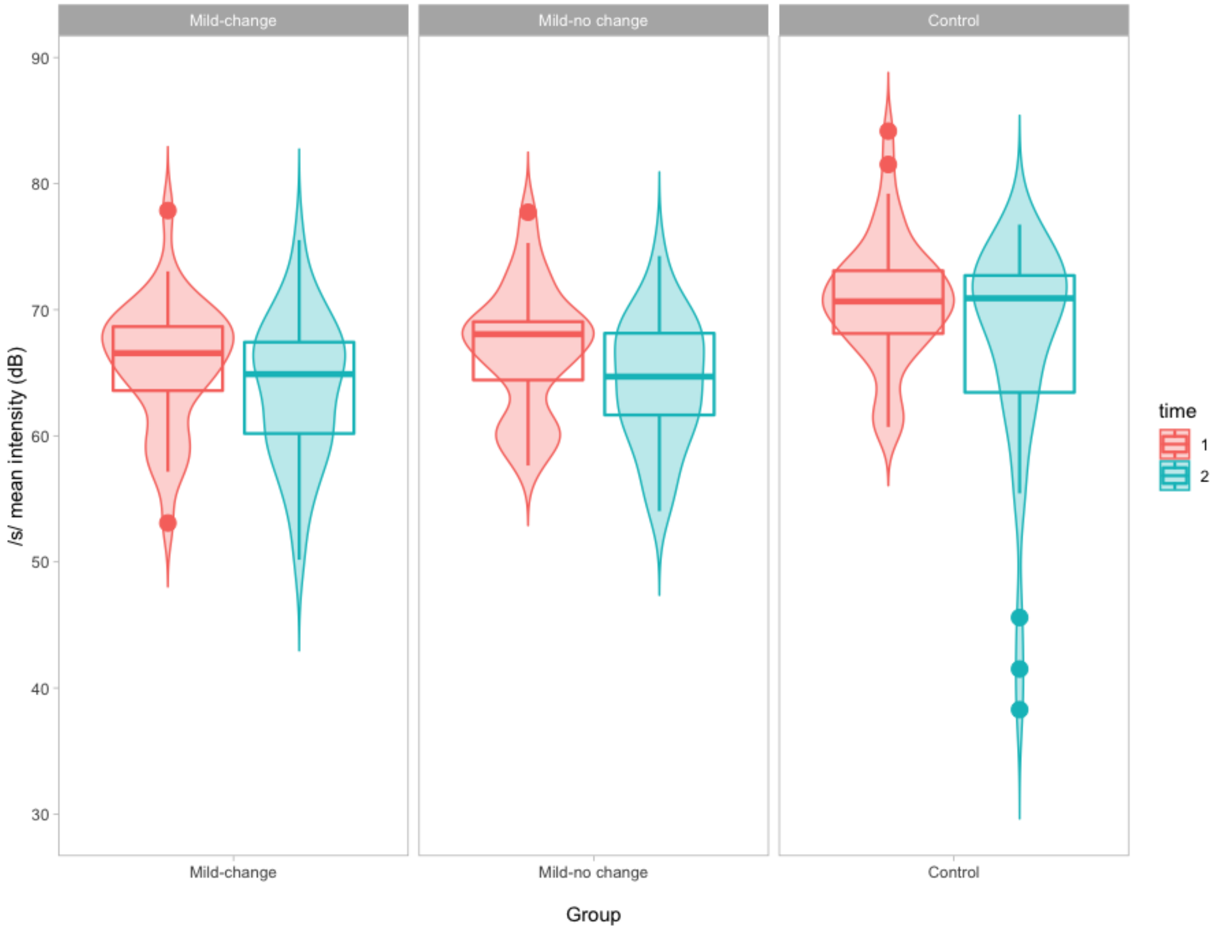
5.3.6.4. /s/ mean intensity

The /s/ intensity descriptive statistics shown in Table 52 follows a similar pattern to /f/ intensity where both PwPD groups have lower mean values than the control group. Once again, the mild-change group means (T1: 65.79dB, SD = 5.06; T2: 63.66dB, SD = 5.67) for /s/ intensity are lower than the mild-no change group (T1: 66.90, SD = 5.01; T2: 64.23dB, SD = 5.03) at both T1 and T2. Similar to /f/ intensity descriptive statistics, the control group also shows a negative change in /s/ intensity and based on Figure 35, a greater variance in distribution in the control group at T2 is observed which may influence the negative change.

Table 52. Summary table of descriptive statistics of /s/ mean intensity for each group at T1 and T2. N = 34 (mild-change); 29 (mild-no change); 47 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	70.75	4.95	70.66	60.71	84.16
Control (T2)	67.46	8.61	70.91	38.29	76.75
Mild-change (T1)	65.79	5.06	66.55	53.09	77.87
Mild-change (T2)	63.66	5.67	64.89	50.18	75.52
Mild-no change (T1)	66.90	5.01	68.06	57.65	77.73
Mild-no change (T2)	64.23	5.03	64.69	54.03	74.26

Figure 35. The distribution of /s/ mean intensity of each group (Mild-change, Mild-no change, Control) at T1 and T2.



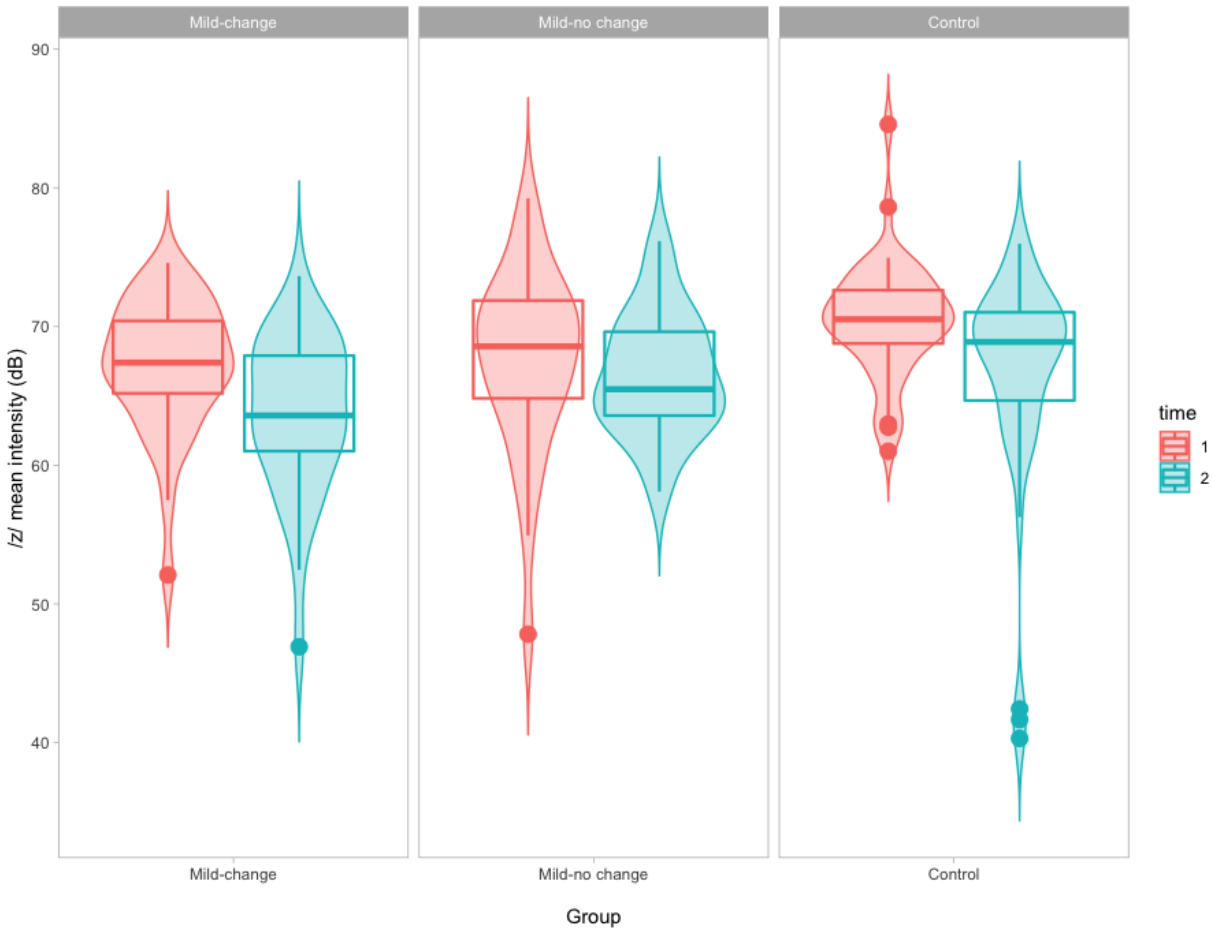
5.3.6.5. /z/ mean intensity

The results of the descriptive statistics on /z/ intensity (Table 53) also follows the pattern of both PwPD groups having lower mean values compared to the control group. The mild-change group shows the greater negative change by 3.51dB compared to the mild-no change group of 0.95dB. The distributions plotted for all groups in Figure 36 shows that the median values in the control group a relatively similar at both time points, but some extreme values may influence the means. The mild-change and mild-no change group seem to have a similar distribution but a slightly greater variance in the mild-change group at T2.

Table 53. Summary table of descriptive statistics of /z/ mean intensity for each group at T1 and T2. N = 34 (mild-change); 29 (mild-no change); 47 (control).

	Mean	SD	Median	Min	Max
Control (T1)	70.49	4.14	70.52	61.01	84.57
Control (T2)	66.61	7.88	68.89	40.31	75.97
Mild-change (T1)	67.17	4.78	67.40	52.09	74.59
Mild-change (T2)	63.66	5.74	63.59	46.91	73.64
Mild-no change (T1)	67.51	6.72	68.57	47.82	79.24
Mild-no change (T2)	66.56	4.40	65.48	58.10	76.15

Figure 36. The distribution of /z/ mean intensity of each group (PwPD-change, PwPD-no change, Control) at T1 and T2.



5.3.6.6. */ʃ/ mean intensity*

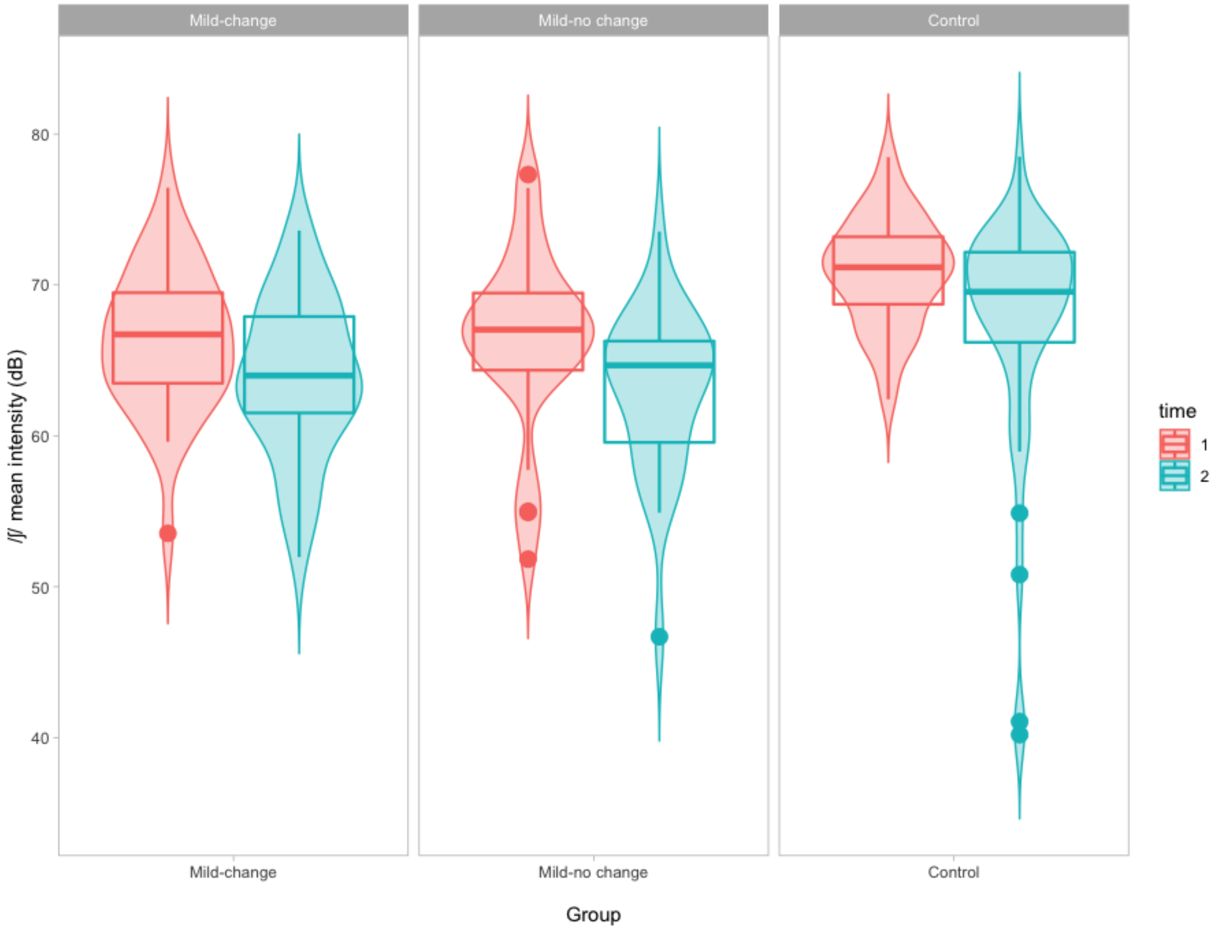
The /ʃ/ intensity descriptive statistics in Table 54 follows the above observed pattern of negative changes in all three groups over time with the PwPD groups having lower mean values compared to the control group at both T1 and T2. The mild-change and mild-no change group have a similar decline in mean values over time but Figure 37 of the plotted distributions of the group suggests a difference in the variance of values between the mild-change and mild-no change group may be influencing the means. Based on the plot, the PwPD groups seem clearly lower than the control group, but whether this difference is significant is unclear.

The general negative trend observed in the fricative values presented in the descriptive statistics will be tested for significance in the following section.

Table 54. Summary table of descriptive statistics of /ʃ/ mean intensity for each group at T1 and T2. N = 34 (mild-change); 29 (mild-no change); 47 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	70.89	3.60	71.16	62.42	78.46
Control (T2)	67.49	7.79	69.54	40.20	78.50
Mild-change (T1)	66.32	4.72	66.72	53.54	76.43
Mild-change (T2)	63.84	5.30	63.99	51.97	73.60
Mild-no change (T1)	66.50	5.88	67.04	51.83	77.31
Mild-no change (T2)	63.07	5.38	64.67	46.69	73.53

Figure 37. The distribution of /ʃ/ mean intensity of each group (Mild-change, Mild-no change, Control) at T1 and T2.



5.3.7. Results of LMMs of fricatives

A summary table of LMM results of all the fricative intensities can be seen in Table 55 below, based on 220 observations from 110 participants. The LMM were run using the same model components as for the plosives reported in the previous section.

The LMM results indicated that the effect of the predictor *time* on the acoustic parameters /f/ and /v/ intensity was significant which mean that both these parameters had a significant negative change in intensity over time. However, the effect of the predictor *group* on /f/ and /v/ intensity was not significant implying that the intensity values for these acoustic parameters did not significantly vary between groups and that the change over time was across groups.

The results indicated that the effects of both the predictors *time* and *group* had a significant impact on the acoustic parameters average fricative intensity, /s/, /z/, and /ʃ/ intensity. This result shows that for these acoustic parameters there was a significant negative change over time that differed between groups. However, for the acoustic parameter /z/ intensity, the intensity values were not significantly different between the mild-no change group and the control group.

The greatest change in intensity was overserved in /ʃ/ intensity with an estimated 3.11dB negative change over time. /ʃ/ intensity also seems to differ the most between groups with the mild-change group as estimated 4.09dB lower than the intensity in the control group and the mild-no change group an estimated 4.40dB lower than the control group. Post-hoc testing in section 5.3.10 may indicate whether there is a statistically significant difference between the mild-change and mild-no change group for any of the significant acoustic parameters.

Table 55. Linear mixed effects model results of the IG analysis of all fricatives against each of the levels of the predictors time and group.

Fixed Effects:

Acoustic parameter	Predictor	Estimate	Std. Error	df	t-value	p
Avg. fricative intensity	Intercept (Control, T1)	69.85	0.66	163.98	106.42	< 0.01*
	T2	-3.05	0.62	109.00	-4.91	<0.001**
	mild-change	-2.68	0.89	107.00	-3.00	0.003**
	mild-no change	-2.03	0.94	107.00	-2.17	0.03*
/f/ intensity	Intercept (Control, T1)	65.64	0.77	149.57	85.10	< 0.01*
	T2	-2.90	0.64	109.00	-4.53	<0.001**
	mild-change	-1.85	1.08	107.00	-1.71	0.09
	mild-no change	-1.10	1.14	107.00	-0.97	0.33
/v/ intensity	Intercept (Control, T1)	69.87	0.78	163.64	90.18	<0.01*
	T2	-3.38	0.73	109.00	-4.63	<0.001**
	mild-change	-0.69	1.06	107.00	-0.66	0.51
	mild-no change	1.34	1.11	107.00	1.21	0.23
/s/ intensity	Intercept (Control, T1)	70.49	0.79	155.76	89.75	<0.01*
	T2	-2.77	0.69	109.00	-4.01	<0.001**
	mild-change	-4.38	1.09	107.00	-4.02	<0.001**
	mild-no change	-3.54	1.14	107.00	-3.10	0.002**
/z/ intensity	Intercept (Control, T1)	70.04	0.74	169.40	94.57	<0.01*
	T2	-2.99	0.73	109.00	-4.10	<0.001**

	mild-change	-3.13	0.10	107.00	-3.14	0.002**
	mild-no change	-1.51	1.04	107.00	-1.45	0.15
/ʃ/ intensity	Intercept (Control, T1)	70.75	0.73	155.88	96.72	<0.01*
	T2	-3.11	0.65	108.43	-4.79	<0.001**
	mild-change	-4.09	1.02	107.11	-4.03	0.001**
	mild-no change	-4.40	1.06	106.25	-4.15	<0.001**

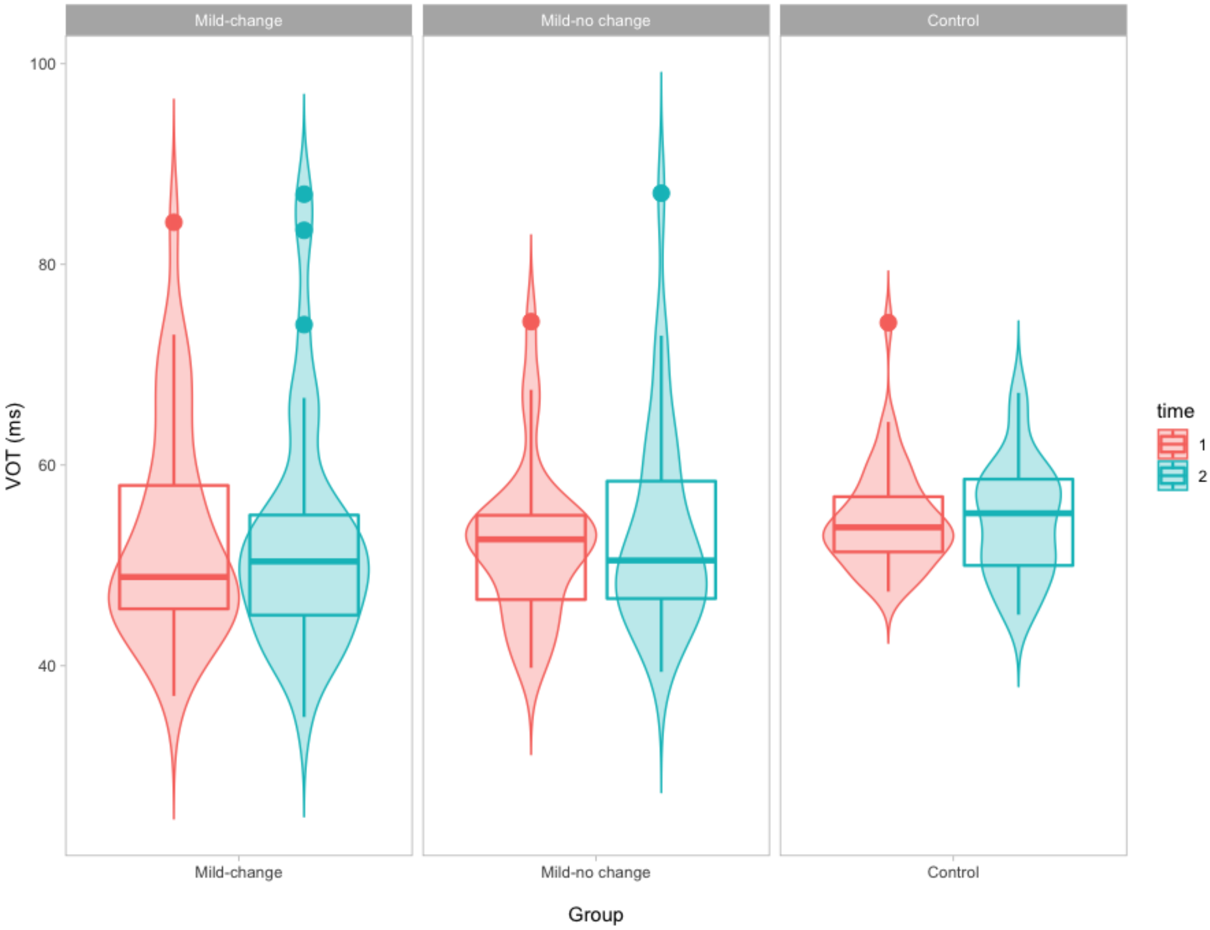
5.3.8. Descriptive statistics of VOT

The results of the VOT descriptive statistics in Table 56 shows that both PwPD groups had lower VOT mean values compared to the control group. The mild-change group and the control groups mean remained relatively stable over time, and the mild-no change group's mean increased by 1.10dB from T1 to T2. However, it is not clear whether this change will be significant. The plot in Figure 38 suggests that there is a greater variance in distribution between the PwPD group VOT values and the control group. This may influence the extent to which it may impact the changes observed over time, but statistical significance is tested in the following section to provide further insight into the noted observations.

Table 56. Summary table of descriptive statistics of VOT for each group at T1 and T2. N = 34 (mild-change); 29 (mild-no change); 47 (control).

	Mean	SD	Median	Min	Max
Control (T1)	54.47	5.04	53.80	47.40	74.20
Control (T2)	54.46	5.72	55.20	45.10	67.20
Mild-change (T1)	52.75	10.84	48.85	37.00	84.20
Mild-change (T2)	52.64	11.79	50.40	34.90	87.00
Mild-no change (T1)	52.26	7.64	52.60	39.80	74.30
Mild-no change (T2)	53.36	10.45	50.50	39.40	87.10

Figure 38. The distribution of VOT of each group (Mild-change, Mild-no change, Control) at T1 and T2.



5.3.9. Results of LMM of VOT

The results of the LMM on VOT is reported in Table 57 below, based on 220 observations from 110 participants. The results indicated that the effects of the predictors *time* and *group* on VOT was not significant implying that VOT neither changed significantly over time, nor differed significantly between groups. The result is further discussed in section 5.3.11, after the results of post-hoc tests are reported.

Table 57. Linear mixed effects model results of the IG analysis of VOT against each of the levels of the predictors time and group.

Fixed effects:

Acoustic parameter	**Predictor**	**Estimate**	**Std. Error**	**df**	**t-value**	**p**
VOT	Intercept (Control, T1)	54.34	1.20	121.78	45.21	<0.01*
	T2	0.25	0.61	109.00	0.41	0.68
	mild-change	-1.77	1.79	107.00	-0.99	0.33
	mild-no change	-1.66	1.88	107.00	-0.88	0.38

5.3.10. Post-hoc tests

Estimated marginal means (EMMs) were obtained using the “emmeans” package (Lenth, Russel V., 2022) in R (R Core Team, 2019). The package used Kenward-roger method to calculate degrees of freedom. This was used to compare factors in the relevant significant voice models to pinpoint where exactly the significance lies. The significant models were for the acoustic parameters /t/, /d/, /k/, /g/ mean intensity, average fricatives mean intensity, /s/, /z/, and /ʃ/ mean intensity.

EMMs were run to conduct a pairwise comparison of all three *groups* and the results automatically averaged (collapsed) over the two levels of the predictor *time* with a confidence level of 95%, giving the mean response value of each level of the predictor *group*, and contrasting it with the other levels. This would help identify how different each group is from the other. A pairwise comparison of the two levels of *time* was also done averaged over the predictor *group*.

Results of the EMMs for each acoustic parameter are discussed below.

5.3.10.1. /t/ mean intensity

In the results of the EMMs presented in Table 58 below, there is no group difference between the Control and Mild-change group between the two levels of *time* (E= 2.06, df= 107, p >0.05), the Control and Mild-no change group (E= 1.74, df= 107, p >0.05), and the Mild-change and Mild no-change group (E= -0.33, df=107, p >0.05). There was a significant difference between the two data collection *time* points (E= 3.15, df=109, p <0.001).

This result partially confirms the LMM in support that /t/ intensity changed significantly over time. However, it does not confirm the results that values were significantly different between groups.

Table 58. Pairwise differences of the predictors group and time for /t/ mean intensity.

Predictor comparisons	**Estimate**	**SE**	**df**	**t-ratio**	**p**
Control – Mild-change	2.06	0.96	107	2.14	0.09
Control – Mild-no change	1.74	1.01	107	1.72	0.20
Mild-change – Mild-no change	-0.33	1.08	107	-0.30	0.95
T1 – T2	3.15	0.66	109	4.76	<0.001**

Note: SE = standard error; df = degrees of freedom.

5.3.10.2. /d/ mean intensity

The results of the EMMs of /d/ mean intensity in Table 59 below, show there is a group difference between the Control and Mild-change group between the two levels of *time* (E= 2.61, df= 107, p <0.05), but no significance between the Control and Mild-no change group (E= 1.92, df= 105, p >0.05), and the Mild-change and Mild no-change group (E= -0.69, df=106, p >0.05). There was a significant difference between the two data collection *time* points (E= 3.15, df=107, p <0.001).

This result confirms the LMM in support that /d/ intensity changed significantly over time. It also confirms the result that the mild-no change group did not differ significantly from the control group, and that that values were significantly different between the mild-change and the control group. It also suggests that /d/ intensity was not significantly different between the mild-change and mild-no change groups.

Table 59. Pairwise differences of the predictors group and time for /d/ mean intensity.

Predictor comparisons	**Estimate**	**SE**	**df**	**t-ratio**	**p**
Control – Mild-change	2.61	1.04	107	2.51	0.04*
Control – Mild-no change	1.92	1.07	105	1.79	0.18
Mild-change – Mild-no change	-0.69	1.16	106	-0.60	0.82
T1 – T2	3.15	0.64	107	4.93	<0.001**

Note: SE = standard error; df = degrees of freedom.

5.3.10.3. /k/ mean intensity

Based on the results of EMMs of /k/ mean intensity (see

This result confirms the LMM in support that /k/ intensity changed significantly over time. It also confirms the result that the mild-no change group and the mild-no change group differ significantly from the control group. However, it suggests that the mild-change and mild-no change groups did not differ from each other significantly.

Table 60 below), there is a significant difference between the Control and Mild-change group between the two levels of *time* (E= 4.38, df= 107, $p < 0.01$), and the Control and Mild-no change group (E= 3.03, df= 105, $p < 0.05$). However, there isn't a group difference between the Mild-change and Mild no-change group (E= -1.35, df=107, $p > 0.05$). There was a significant difference between the two data collection *time* points (E= 3.83, df=106, $p < 0.01$).

This result confirms the LMM in support that /k/ intensity changed significantly over time. It also confirms the result that the mild-no change group and the mild-no change group differ significantly from the control group. However, it suggests that the mild-change and mild-no change groups did not differ from each other significantly.

Table 60. Pairwise differences of the predictors group and time for /k/ mean intensity.

Predictor comparisons	**Estimate**	**SE**	**df**	**t-ratio**	**p**
Control – Mild-change	4.38	1.15	107	3.80	0.001**
Control – Mild-no change	3.03	1.20	105	2.53	0.03*
Mild-change – Mild-no change	-1.35	1.29	107	-1.04	0.55
T1 – T2	3.83	0.76	106	5.04	<0.001**

Note: SE = standard error; df = degrees of freedom.

5.3.10.4. /g/ mean intensity

In Table 61 below, the EMM results of /g/ mean intensity are presented. There is a significant difference between the Control and Mild-change group between the two levels of *time* (E= 3.28, df= 107, p <0.05). However, there isn't a group difference between the Control and Mild-no change group (E= 2.55, df= 106, p >0.05), and the Mild-change and Mild no-change group (E=-0.73, df=107, p >0.05). There was a significant difference between the two data collection *time* points (E= 2.33, df=109, p <0.01).

This result confirms the LMM in support that /g/ intensity changed significantly over time. It also confirms the result that the mild-no change group and the mild-no change group differ significantly from the control group. However, it suggests that the mild-change and mild-no change groups did not differ from each other significantly.

Table 61. Pairwise differences of the predictors group and time for /g/ mean intensity.

Predictor comparisons	**Estimate**	**SE**	**df**	**t-ratio**	**p**
Control – Mild-change	3.28	1.16	107	2.83	0.02*
Control – Mild-no change	2.55	1.21	106	2.10	0.09*
Mild-change – Mild-no change	-0.73	1.30	107	-0.56	0.84
T1 – T2	2.33	0.73	109	3.18	<0.01**

Note: SE = standard error; df = degrees of freedom.

5.3.10.5. Average fricatives mean intensity

Referring to Table 62 below, the results of the EMMs of average fricatives mean intensity indicates there is a significant difference between the Control and Mild-change group between the two levels of *time* (E= 2.68, df= 107, p <0.01). However, there isn't a group difference between the Control and Mild-no change group (E= 2.03, df= 107, p >0.05), and the Mild-change and Mild no-change group (E= -0.65, df=107, p >0.05). There was a significant difference between the two data collection *time* points (E= 2.33, df=109, p <0.01).

This result partially confirms the LMM in support that average fricative intensity changed significantly over time. It also confirms the result that the mild-change group differed significantly from the control group. However, it does not confirm the results that the mild-no change group was significantly different from the control group and suggests values were not significantly different between the mild-change and the control group.

Table 62. Pairwise differences of the predictors group and time for average fricatives mean intensity.

Predictor comparisons	**Estimate**	**SE**	**df**	**t-ratio**	**p**
Control – Mild-change	2.68	0.89	107	3.00	<0.01*
Control – Mild-no change	2.03	0.94	107	2.17	0.08
Mild-change – Mild-no change	-0.65	1.00	107	-0.65	0.79
T1 – T2	3.04	0.62	109	4.91	<0.001**

Note: SE = standard error; df = degrees of freedom.

5.3.10.6. /s/ mean intensity

As seen from Table 63 below of the EMMs of /s/ mean intensity, there is a significant difference between the Control and Mild-change group between the two levels of *time* (E= 4.38, df= 107, p <0.01), and the Control and Mild-no change group (E= 3.54, df= 107, p <0.01). However, there isn't a group difference between the Mild-change and

Mild no-change group (E= -0.84, df=107, p >0.05). There was a significant difference between the two data collection *time* points (E= 2.77, df=109, p <0.01).

This result confirms the LMM in support that /s/ intensity changed significantly over time. It also confirms the result that the mild-no change group and the mild-no change group differ significantly from the control group. However, it suggests that the mild-change and mild-no change groups did not differ from each other significantly.

Table 63. Pairwise differences of the predictors group and time for /s/ mean intensity.

Predictor comparisons	**Estimate**	**SE**	**df**	**t-ratio**	**p**
Control – Mild-change	4.38	1.09	107	4.02	<0.001**
Control – Mild-no change	3.54	1.14	107	3.10	<0.01*
Mild-change – Mild-no change	-0.84	1.22	107	-0.69	0.77
T1 – T2	2.77	0.69	109	4.01	<0.001**

Note: SE = standard error; df = degrees of freedom.

5.3.10.7. /z/ mean intensity

Results of the EMMs of /z/ mean intensity in Table 64 below shows there is a significant difference between the Control and Mild-change group between the two levels of *time* (E= 3.13, df= 107, p <0.01). However, there isn't a group difference between the Control and Mild-no change group (E= 1.51, df= 107, p >0.05), and the Mild-change and Mild no-change group (E= -1.62, df=107, p >0.05). There was a significant difference between the two data collection *time* points (E= 2.33, df=109, p <0.01).

This result confirms the LMM in support that /z/ intensity changed significantly over time. It also confirms the result that the mild-no change group was significantly different from the control group, and that that values were not significantly different between the mild-no change and the control group. The EMMs also suggests that /z/

intensity was not significantly different between the mild-change and mild-no change groups.

Table 64. Pairwise differences of the predictors group and time for /z/ mean intensity.

Predictor comparisons	**Estimate**	**SE**	**df**	**t-ratio**	**p**
Control – Mild-change	3.13	0.10	107	3.14	<0.01*
Control – Mild-no change	1.51	1.04	107	1.45	0.32
Mild-change – Mild-no change	-1.62	1.12	107	-1.47	0.32
T1 – T2	2.99	0.73	109	4.10	<0.001**

Note: SE = standard error; df = degrees of freedom.

5.3.10.8. /ʃ/ mean intensity

Finally, the results of the EMMs on /ʃ/ mean intensity are presented in Table 65 below. There is a significant difference between the Control and Mild-change group between the two levels of *time* (E= 4.09, df= 107, p <0.01), and the Control and Mild-no change group (E= 4.40, df= 106, p <0.01). However, there isn't a group difference between the Mild-change and Mild no-change group (E= 0.31, df=107, p >0.05). There was a significant difference between the two data collection *time* points (E= 3.11, df=109, p <0.01).

This result confirms the LMM in support that /ʃ/ intensity changed significantly over time. It also confirms the result that the mild-no change group and the mild-no change group differ significantly from the control group. However, it suggests that the mild-change and mild-no change groups did not differ from each other significantly.

Table 65. Pairwise differences of the predictors group and time for /ʃ/ mean intensity.

Predictor comparisons	**Estimate**	**SE**	**df**	**t-ratio**	**p**
Control – Mild-change	4.09	1.01	107	4.03	<0.001**
Control – Mild-no change	4.40	1.06	106	4.15	<0.001**
Mild-change – Mild-no change	0.31	1.14	107	0.28	0.96
T1 – T2	3.11	0.65	109	4.79	<0.001**

Note: SE = standard error; df = degrees of freedom.

5.3.11. Discussion

This section presented the results of the IG analysis conducted on all participants, analysing the recordings of the grandfather passage. The acoustic parameters on plosives (average plosives intensity, /p/, /b/, /t/, /d/, /k/, /g/ intensity), fricatives (average fricatives intensity, /f/, /v/, /s/, /z/, /ʃ/ intensity), and VOT were all used to investigate this, similar to the PFR analysis, to test if the acoustic parameters could track changes in PwPD speech intelligibility over time.

Results indicated that among plosives, /t/, /d/, /k/ and /g/ intensity had a significant negative change over time and was significantly different from the control group. Regarding fricatives, average fricatives mean intensity, /s/, /z/ and /ʃ/ intensity had a significant negative change over time and was significantly different from the control group. These results suggest that the mentioned acoustic parameters are able to track changes in PwPD speech intelligibly over time.

Post-hoc testing on the acoustic parameters /d/, /k/, /g/, /s/, /z/, and /ʃ/ intensity revealed that the differences lied between the control group and the two PwPD groups but was not significant between the mild-change and mild-no change groups. This could be due to both PwPD groups having mild speech symptoms (based on the overall intelligibility ratings) and therefore any differences between them not being distinct enough from each other. This is not surprising as the objective was to investigate if each acoustic parameter could distinguish between different intelligibility groups. Since both the PwPD groups in the dataset used for the IG analysis were categorised with the same intelligibility (mild), it would be unexpected for a significant group difference to be found between the mild-change and mild-no change group.

Post-hoc results for /t/ intensity and average fricative intensity indicated that the estimated means of both PwPD and control groups were not significantly different from one another. This suggests that /t/ intensity and average fricative intensity showed some change in all groups over time. This result is consistent with the changes in the means of the descriptive statistics where all the groups showed a negative change over time. The acoustic parameters average plosives intensity, /p/, /b/, /f/, and /v/ intensity also indicated only a negative change over time and no significant differences in intensity values between groups, but the negative change in these parameters was larger in the mild-change group compared to the other groups (based on the descriptive statistics). This may indicate that the mild-change group may have been driving the change.

However, a negative change in all groups for the aforementioned acoustic parameters contradict the SLT ratings which suggested that only the mild-change group should show a change in speech intelligibility over time. This could suggest that these acoustic parameters independently did not contribute to the perceived change in PwPD speech intelligibility over time as rated by the SLTs. Other parameters may have had a more substantial impact, lying outside of the variation seen in control speech. These results also suggest that the claim made in the PFR analysis discussion that there is a natural degree of variation in control speech may warrant further investigation. Finally, all the acoustic parameters of plosive and fricative intensity could be impacted by within task speech variation (Kuo & Tjaden, 2016) suggested in the PFR analysis.

As in the PFR analysis, VOT was unable to track changes in PwPD speech intelligibility over time as VOT values did not significantly change over time and were not significantly different between groups. This corroborates the claim made in the PFR analysis that VOT may not have been able to track PwPD speech changes in mild impairment. Further investigation is warranted to assess if VOT can track PwPD speech changes over time in moderate to severe speech impairment. In addition, Flint et al. (1992) found that VOT duration was shorter in speakers with PD as compared to healthy controls. Fischer & Goberman (2010) stated that the inconsistent findings for individuals with PD may be due to the lack of examination independent of speech rate. Fischer and Goberman's (2010) study found that PD speakers presented with articulatory undershoot and this warranted no difference in VOT of PD speakers as compared to healthy controls which may have contributed to the lack of a significant group difference in VOT in the present study.

5.4. Chapter conclusions

While the subset in the PFR analysis was selected based only those participants who had perceptual features rated as 'marked' and 'severe' by SLTs, overall intelligibility ratings were used for the IG to group participants based on how severely intelligibility was rated as being impacted.

The results of the PFR analysis indicated that the fricative mean /ʃ/ intensity was the only acoustic parameter able to capture PwPD perceptual changes over time and distinguish between the control group. This suggests that the acoustic parameter mean /ʃ/ intensity could be used for differential diagnosis and capturing changes in perceptual features in PwPD. Further testing with other levels of severity, or replication studies could help ascertain the result of the present study. The IG analysis showed that the acoustic parameters /d/, /k/, /g/, /s/, /z/, and /ʃ/ intensity were able to capture changes in PwPD speech intelligibility over time and distinguish between the control group. Acoustic parameters seem to be better at capturing change in PwPD speech intelligibility than in PwPD perceptual changes. This could be due in part because individual consonants do not contribute maximally to the perceptual feature "imprecise consonant production” but acoustic parameters pertaining to consonant intensity seem to contribute to speech intelligibility.

Individual variation should be investigated in future studies as there may have been individual speech changes which influenced the overall group averages. This was indicated by the large variation in distribution of all parameters reported in the descriptive statistics. Participant was controlled for during LMM analysis as a random factor accounting for some variation, but the extent of this influence is unclear. In addition, there might be some influences of sex on results that may have influenced results and cannot be accounted for. Further, the present study did not look at pitch glides or contrasts which may have provided more information on imprecise consonant production and articulatory breakdown therefore the extent to which these parameters may track PwPD speech change is unclear.

Despite some contradictions to the findings in the acoustic parameters pertaining to plosive and consonant intensity, there is evidence that some of these parameters are able to track PwPD perceptual changes and changes in PwPD speech intelligibility and further investigation may have delineate some of the interpretations made. In addition, further data collection at regular intervals may illustrated whether the changes observed in the PwPD group will become more distinct from the control group over time. This has been evidenced in studies where a larger range in the interval between data collection time points have shown that PwPD speech change is distinct from control speech (Harel et al., 2004; Skodda et al., 2009, 2012, 2013).

The next chapter will delve into the results of the analysis conducted on the acoustic parameters of the final speech category, prosody before proceeding to the general discussion chapter.

6. Prosody

Prosody is cited as the most significantly affected speech dimension in hypokinetic dysarthria (Murdoch, 2009). Prosody consists of distinct subdimensions, namely, speech rhythm and velocity, speech rate and speech-to-pause ratio, speech intensity, and pitch variation (Skodda, Visser, et al., 2011a). Predominant deviant perceptual features include monopitch, monoloudness, and reduced stress. Acoustically, reduction in F0 variation and reduced F0 range has been identified as the most noticeable acoustic features in hypokinetic dysarthria speech by Harel et al. (2004). They carried out a case study on two individuals with PD assessing changes in acoustic parameters at two time points over five years where F0 variation was found to be markedly reduced than controls and showed a declining trend during initial diagnosis and disease progression over five years. Based on the results of Harel et al.'s (2004) study, F0 and F0 range were proposed as potential markers for early disease progression in PD speech. Another acoustic parameter variable speech rate has also been noted in multiple studies, with some reporting increased speech rate, including short rushes of speech in segments (Darley et al., 1975) decreased speech rate, and some reporting normal speech rate compared to a control group (Grosset et al., 2009).

The previous chapter outlined the acoustic and statistical analyses conducted on the speech category articulation based on the results of the SLT ratings (see chapter 3, section 3.5). In this chapter, the results of the acoustic and statistical analysis of prosody will be presented using the same analyses conducted in the voice and articulation chapters.

The SLTs rated prosody at T1 as one of the most impacted speech categories of the seven categories rated. Acoustic parameters were selected to measure the perceptual feature speech rate and investigate the general impact on prosody in PwPD by quantifying this perceptual category from the speech signal. The parameters selected were speech rate, mean intensity, F0 variation (F0SD), and intensity variation (IntSD) extracted from The Grandfather Passage. The rationales for selecting these acoustic parameters are included in section 6.1 below.

Two independent data analyses were conducted using The Grandfather Passage's speech recordings to investigate the selected acoustic parameters' effectiveness in answering the research questions. The perceptual feature ratings (PFR) analysis was conducted on a subset of the dataset (n = 52) to identify acoustic parameters that could reflect a negative change in the perceptual rating of speech prosody in PwPD between T1 and T2. A subset of data was selected based on the SLT's perceptual ratings on the prosodic aspect. Participants with prosodic perceptual features rated as 'marked' or 'severe' by SLTs were selected. The recordings of this subset were included in groups based on whether the SLTs rating of the severe perceptual features negatively changed from T1 and T2 collection points or did not change at all. Therefore, this analysis focused on grouping the recordings based on detecting a change in the perceptual features over time.

The intelligibility groups (IG) analysis was conducted to address whether changes in the acoustic properties of the prosodic markers captured changes in intelligibility in PwPD speech on the entire participant sample (n = 110) and grouped based on the overall intelligibility ratings given by the SLTs and whether those ratings negatively changed from T1 to T2. The overall intelligibility ratings were used to group PwPD recordings as either mild, moderate, or severe and if the ratings increased or decreased between data collection points.

The structure of the rest of the chapter includes details of the selection of acoustic parameters used for both analyses, how each analysis was conducted along with the results and discussion, and an overall discussion of the results of this chapter and its implications for voice quality in PwPD speech.

6.1. Selection of Acoustic parameters

Within all the perceptual features in prosody, speech rate was the only perceptual feature that showed a high severity rating. As the basal ganglia is supposed to regulate the temporospatial aspects in the motor cortex, abnormalities in PwPD speech rate is assumed (Brown & Marsden, 1998; A. M. Goberman & McMillan, 2005). An altered speech rate in PwPD compared to control could result from increased rigidity and hypokinesia in the speech production system (Solomon & Hixon, 1993). In a longitudinal study (Skodda et al., 2009), the acoustic parameters of speech rate and F0 variation were analysed to monitor PwPD speech over time and compare them to control speakers. The total speech rate (syllables per second related to the total speech time) and net speech rate in the PD group decreased from the first data collection to the second data collection. However, previous studies on speech rate in PwPD present inconsistent results, probably due to differences in methodology and sample sizes (Ackermann, Konczak, & Hertrich, 1997; Caligiuri, 1989; Flint et al., 1992; Metter & Hanson, 1986). Despite the inconsistencies in the results of previous studies, speech rate is a prominent feature of PwPD speech and rated as ‘marked’ or ‘severely’ impacted in participants in the present study, warranting further investigation.

In addition to investigating speech rate, the present study will analyse other acoustic parameters that quantify perceptual features of prosody that can often co-occur in PwPD speech along with altered speech rate. These acoustic parameters are mean intensity, intensity variation (IntSD), and F0 variation (F0SD). Previous research on prosody unequivocally indicates a significantly reduced F0 variability in patients with PD compared with healthy controls and often co-occurring with altered speech rate (Flint et al., 1992; A. M. Goberman & McMillan, 2005; Metter & Hanson, 1986; Skodda et al., 2009).

Mean intensity and intensity variation, which quantify the perceptual feature monoloudness, could correspond to the reduced functioning of the respiratory and thyroarytenoid muscles. F0SD quantifies the perceptual feature monopitch. The motor symptom correspondence of monopitch would be reduced vocal cord movement leading to glottal incompetence (Rusz et al., 2021). Monopitch and monoloudness are often cooccurring perceptual features (Kent, 1996; Kim, 1994) so analysing the acoustic parameters mean intensity, IntSD and F0SD would help track if PwPD speech change in speech rate co-occurred with other measures of prosody.

6.2. Perceptual feature ratings (PFR) analysis

6.2.1. Participant demographics

The type of data recordings used for this analysis were the entire reading passage recordings. The subset consisted of 52 speakers with speech recordings divided into three groups: PwPD-change, PwPD-no change, and control. The demographic information for each group is presented in Table 66 below:

Table 66. Participant demographics for the PFR analysis.

Groups	**Number of participants**	**Age range (years)**	**Mean age (years)**
PwPD-change	N= 21; M = 15; F = 6	50-93	69 (SD = 10.46)
PwPD-no change	N= 21; M = 13; F = 8	56-84	71 (SD = 6.84)
Control	N= 10; M = 5; F = 5	51-82	70 (SD = 9.94)

6.2.2. Segmentation of recordings

Syllable and pause boundaries were manually annotated using the program Praat (Boersma & Weenink, 2020). Pauses were defined as a period of silence lasting for a minimum of 10 milliseconds (Skodda & Schlegel, 2008).

6.2.3. Extracting parameters

Acoustic parameters of prosody were extracted from recordings of the grandfather passage. Prosodic parameters were extracted as proposed by (Lowit-Leuschel & Docherty, 2001).

Speech rate (syllable per second) was calculated by dividing the total number of syllables in the passage by the overall duration (in seconds) of the passage. Speech rate measured the duration of each syllable and each pause (in milliseconds/ms), respectively, based on the oscillographic sound pressure signal.

Mean intensity is calculated by extracting the mean (in dB) of the intensity values of the frames within the specified frame (the grandfather passage. Since the averaging method is dB, the mean intensity between the times t1 and t2 is defined as

$$1/(t2 - t1) \int_{t1}^{t2} x(t)\, dt$$

where x(t) is the intensity as a function of time in dB (Boersma & Weenink, 2020).

Intensity variation was extracted as the standard deviation of the speech intensity contour of voiced segments. The standard deviation between the times t1 and t2 (the start and end times within the specified frame) is defined as

$$\sqrt{\{1/(t2 - t1) \int_{t1}^{t2} dt\, (x(t) - \mu)^2\}}$$

where x(t) is the intensity (in dB) as a function of time, and μ its mean (Boersma & Weenink, 2020).

F0SD involved extracting fundamental frequency from the speech sample as the lowest audio frequency with the highest intensity less harmonic contents (using an acoustic periodicity detection first described by (Boersma, 1993). F0SD was not extracted in Hertz but in semitones by converting the F0 contour into a semitone scale as it is supposed to capture F0 variation more accurately and control for gender differences (Rusz, Tykalova, et al., 2021).

6.2.4. Reliability of acoustic annotations

To determine the reliability of the annotations, 10% of text grids, randomly selected from either data collection point, were manually re-examined by the primary researcher and an external researcher trained in acoustic analysis. The text grids were randomly selected from either data collection point. Syllable boundaries were manually rechecked. The researcher did not find any discrepancies in the annotations made, with all the annotations between the primary and external research being within 1.5ms of each other.

6.2.5. Descriptive statistics

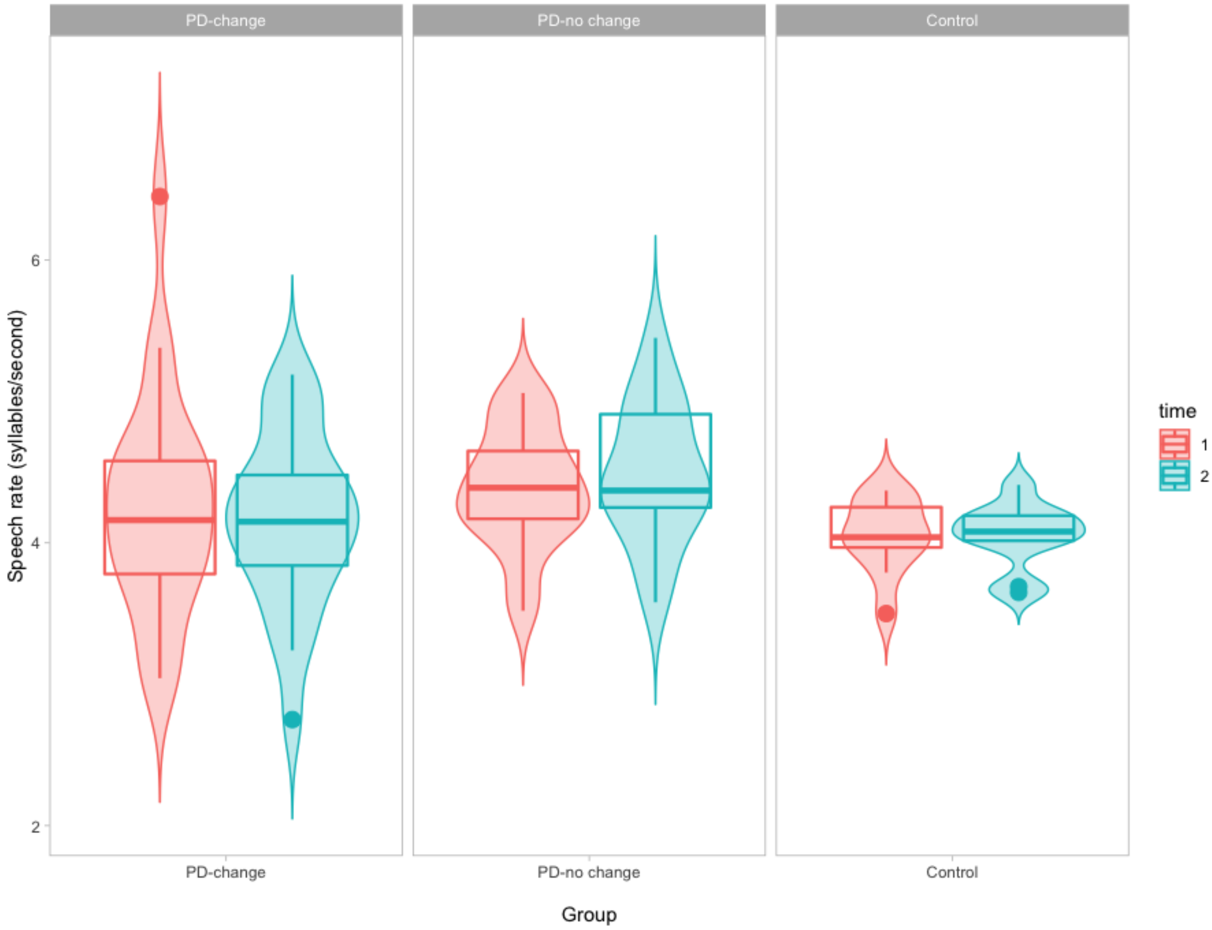
6.2.5.1. Speech rate

The descriptive statistics of speech rate in Table 67 indicate similar means for the control group, the PwPD-change group and the PwPD-no change group, and the mean remains relatively stable across the two time points. Both PwPD groups have higher mean values for speech rate compared to the control group. The PwPD-change group has a decline in speech rate from T1 (mean = 4.27 syll/sec; SD = 0.80) to T2 (mean = 4.15; SD = 0.62), and the PwPD-no change group has an increase in mean values from T1 (mean = 4.38 syll/sec; SD = 0.42) to T2 (mean = 4.49 syll/sec; SD = 0.49). However, there appears to be greater variance in the distribution of values in the PwPD-change group, as evidenced by the higher SDs at both time points. This is illustrated clearly in Figure 39 below. Given the small difference in values between groups and over time, it is not expected for these differences to be significant, but this will be investigated during statistical analysis.

Table 67. Summary table of descriptive statistics of speech rate for each group at T1 and T2. N = 21 (PwPD-change); 21 (PwPD-no change); 10 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	4.05	0.27	4.04	3.50	4.37
Control (T2)	4.05	0.23	4.08	3.65	4.41
PwPD-change (T1)	4.27	0.80	4.16	3.04	6.45
PwPD-change (T2)	4.15	0.62	4.15	2.75	5.19
PwPD-no change (T1)	4.38	0.42	4.39	3.52	5.06
PwPD-no change (T2)	4.49	0.49	4.37	3.58	5.45

Figure 39. The distribution of speech rate (syllable per second) of each group (PwPD-change, PwPD-no change, Control) at T1 and T2.



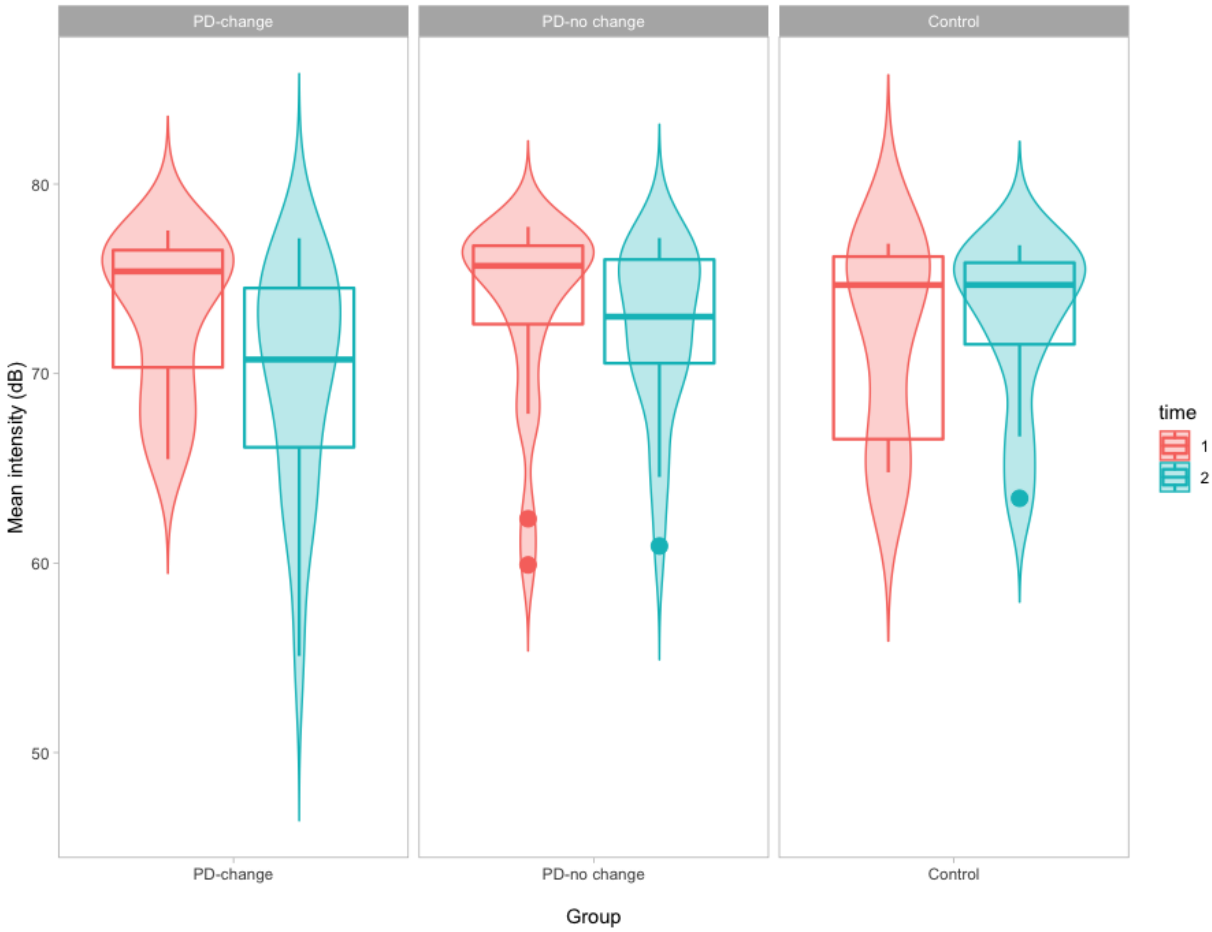
6.2.5.2. *Mean intensity*

Mean intensity descriptive statistics presented in Table 68 show that PwPD groups have higher mean values compared to the control group but also have a negative change over time, while the control group remain relatively stable over time. The PwPD-change group shows a greater change in mean values with a negative decrease of 3.16dB mean compared to the PwPD-no change group of 1.32dB mean decrease. The plot of the distribution of values in Figure 40 shows that the PwPD-no change group and the control group have a similar variance in distribution. In contrast, the PwPD-change group has a greater variance which would have influenced the mean values.

Table 68. Summary table of descriptive statistics of mean intensity for each group at T1 and T2. N = 21 (PwPD-change); 21 (PwPD-no change); 10 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	72.03	5.23	74.68	64.79	76.85
Control (T2)	72.82	4.51	74.68	63.42	76.77
PwPD-change (T1)	73.16	4.10	75.39	65.48	77.55
PwPD-change (T2)	69.94	5.92	70.74	55.10	77.14
PwPD-no change (T1)	73.56	5.02	75.69	59.90	77.74
PwPD-no change (T2)	72.24	4.47	73.00	60.90	77.15

Figure 40. The distribution of mean intensity of each group (PwPD-change, PwPD-no change, Control) at T1 and T2.



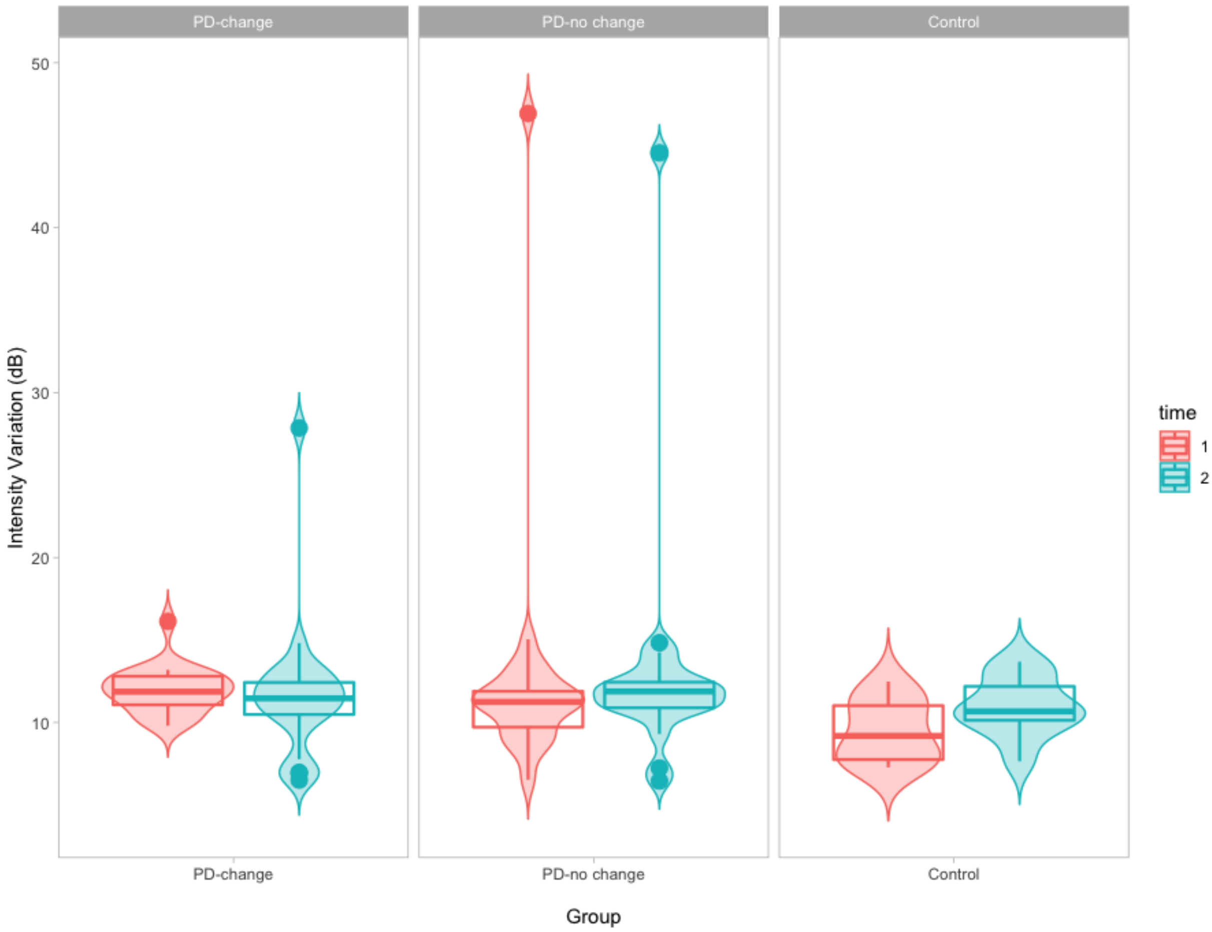
6.2.5.3. IntSD

IntSD descriptive statistics in Table 69 below show that both PwPD groups have higher mean values than the control group at both time points. In addition, the PwPD-no change group and control group values increase in mean over time. However, visual inspection of the plot in Figure 41 shows that some outliers in the PwPD-no change and control group likely impact the mean values. Based on the figure, not including the outliers present, the general distribution of the three groups looks similar and seems relatively stable over time. However, whether the outliers impact the statistical analysis will be explored further and tested for significance.

Table 69. Summary table of descriptive statistics of intensity variation (IntSD) for each group at T1 and T2. N = 21 (PwPD-change); 21 (PwPD-no change); 10 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	9.48	1.90	9.19	7.28	12.48
Control (T2)	10.97	1.83	10.67	7.65	13.69
PwPD-change (T1)	11.97	1.42	11.87	9.81	16.13
PwPD-change (T2)	11.74	4.33	11.46	6.51	27.86
PwPD-no change (T1)	12.72	8.07	11.25	6.53	46.92
PwPD-no change (T2)	13.02	7.49	11.89	6.45	44.54

Figure 41. The distribution of intensity variation (IntSD) of each group (PwPD-change, PwPD-no change, Control) at T1 and T2.



6.2.5.4. *F0SD*

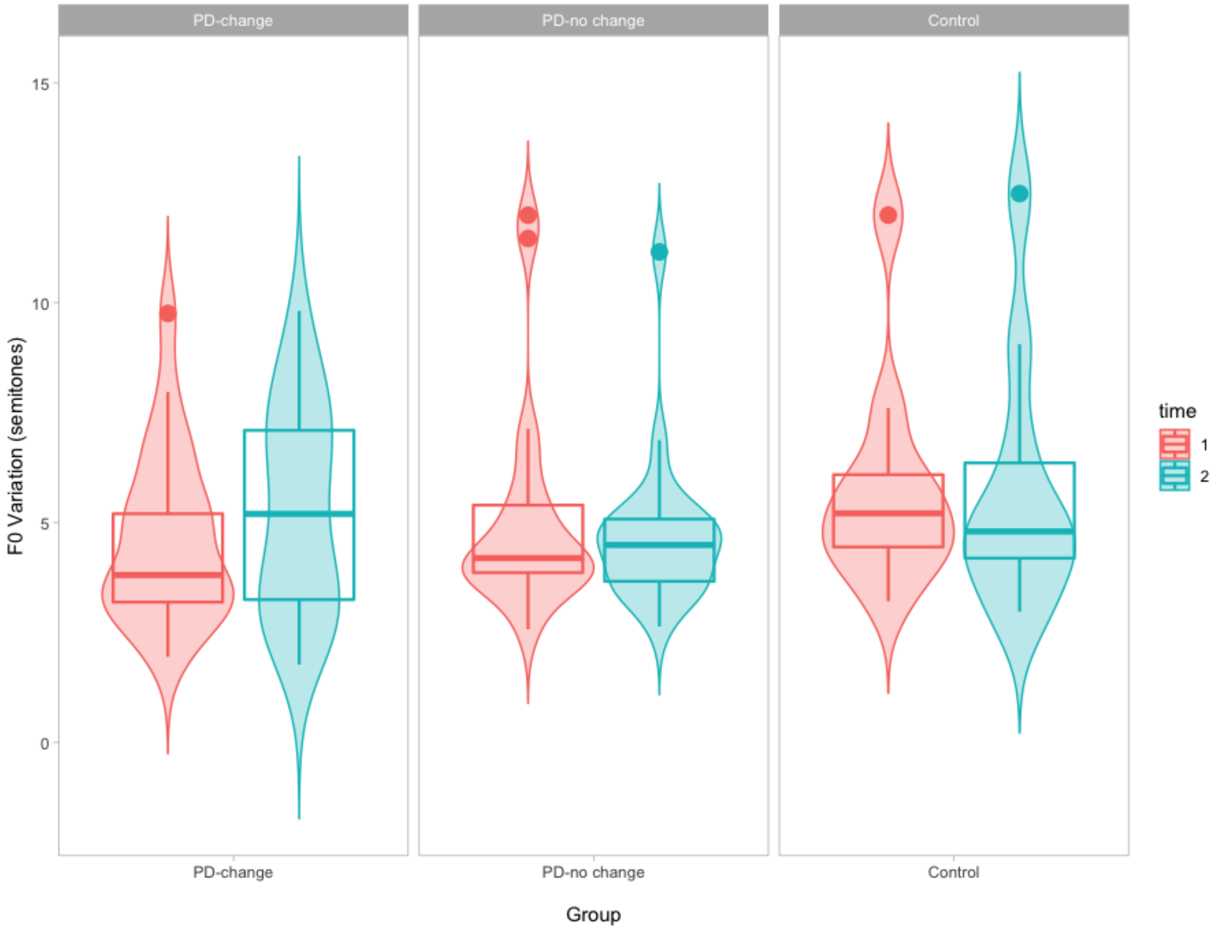
Descriptive statistics of F0SD are presented in Table 70. The table shows that the mean values for the PwPD groups change over time while the control group's means remain relatively stable over time. The PwPD-change group has an increase in mean values from T1 (mean = 4.65 semitones, SD = 1.97) to T2 (mean = 5.25 semitones, SD = 2.39), and the PwPD-no change group has decreased in mean from T1 (mean = 5.10 semitones, SD = 2.45) to T2 (mean = 4.74 semitones, SD = 1.76). This suggests that the change in F0SD diverges over time between the PwPD-change and PwPD-no change groups. However, whether this is statistically significant will need to be tested. The plotted distribution of F0SD for all groups over time can be seen in Figure 42.

Based on the plot, the distribution for the PwPD-change group seems to change over time more than the other two groups.

Table 70. Summary table of descriptive statistics of fundamental frequency variation (F0SD) for each group at T1 and T2. N = 21 (PwPD-change); 21 (PwPD-no change); 10 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	5.81	2.50	5.21	3.21	12.00
Control (T2)	5.76	2.96	4.80	2.97	12.48
PwPD-change (T1)	4.65	1.97	3.81	1.95	9.76
PwPD-change (T2)	5.25	2.39	5.20	1.77	9.82
PwPD-no change (T1)	5.10	2.45	4.19	2.58	11.99
PwPD-no change (T2)	4.74	1.76	4.49	2.64	11.16

Figure 42. The distribution of fundamental frequency variation (F0SD) of each group (PwPD-change, PwPD-no change, Control) at T1 and T2.



6.2.6. Results of LMMs

To test the significance of each predictor (time and group) on each acoustic parameter, a linear mixed effects model (LMM) analysis was conducted using R (R Core Team, 2019). The models were run with the lme4 package (Bates et al., 2015) using lmer. The fit for each model was checked by adding any possible interactions between fixed effects and any slopes between all factors. The best fit was found only to contain each acoustic parameter tested against *time* and *group* as predictors and *participant* as the random factor. Optimisers were used to ensure that over-fitting did not occur.

A model was run with each acoustic parameter (outcome variable, i.e., speech rate, mean intensity, IntSD, and F0SD) against the predictors *time* (data collection time points T1 and T2) and *group* (PwPD-change, PwPD-no change, and control) with the random factor of *participant* including a random intercept and random slope. A random intercept means that the final model takes into account that each individual *participant* can show higher or lower values for each acoustic parameter regardless of *group* or the data collection *time*. A random slope for *participant* means that the model considers and accounts for some between-participant variation.

Based on the assumptions of an LMM, the outcome variable does not need to be normally distributed if the residuals of each model are normally distributed. The residuals for each model run on each acoustic parameter (outcome variable) were checked, and the residuals of the IntSD model violated the normality assumption. Transformations were applied to the data but did not resolve normality, so outliers were checked. Two participants' data from the PwPD group (whose values fell outside of two standard deviations) was dropped from the IntSD variable. The model was rerun, and residual normality was checked and did not violate the assumption. There were no significant correlations between any of the predictors and therefore were not reported.

LMMs create an intercept in each model representing each acoustic parameter against T1 data collection and the control group. The intercept represents how different the outcome variable and predictors are from chance. The other predictors are compared against this intercept model, and results are reported based on how different other predictors are from the intercept.

A summary table of the results for the acoustic parameters speech rate, mean intensity, and F0SD is seen in Table 71 based on 104 observations from all 52 participants. The results of the acoustic parameter IntSD are also presented in Table 71 below, with the outliers removed and based on 100 observations from 50 participants.

Results of the LMMs indicated that the effects of both the predictors of time and group did not have a significant impact on the acoustic parameters speech rate, mean intensity, F0SD, or IntSD. This means that none of the groups showed a significant change in the acoustic parameters over time and was not significantly different from one another. These results are discussed in the following section.

Table 71. Linear mixed effects model results of the PFR analysis of speech rate, mean intensity, IntSD, and F0SD against each of the levels of the predictors time and group.

Fixed Effects:

Acoustic parameter	**Predictor**	**Estimate**	**Std. Error**	**df**	**t-value**	**p**
Speech rate	Intercept (Control, T1)	4.05	0.17	50.72	23.931	< 0.001**
	T2	-0.00	0.04	51.00	-0.073	0.94
	PwPD-change	0.16	0.20	49.00	0.771	0.44
	PwPD-no change	0.38	0.20	49.00	1.885	0.07
Mean intensity	Intercept (Control, T1)	73.27	1.29	61.14	56.873	< 0.001**
	T2	-1.68	0.86	51.00	-1.955	0.06
	PwPD-change	-0.87	1.48	49.00	-0.592	0.56
	PwPD-no change	0.48	1.48	49.00	0.322	0.75
IntSD	Intercept (Control, T1)	10.49	0.49	54.03	21.59	< 0.001**
	T2	0.12	0.32	45.89	0.37	0.71
	PwPD-change	1.00	0.56	45.44	1.79	0.08
	PwPD-no change	1.12	0.56	45.12	1.99	0.06
F0SD	Intercept (Control, T1)	5.74	0.68	52.23	8.470	< 0.001**
	T2	0.09	0.24	51.00	0.369	0.71
	PwPD-change	-0.83	0.81	49.00	-1.029	0.31
	PwPD-no change	-0.86	0.81	49.000	-1.066	0.29

6.2.7. Discussion

The PFR analysis was conducted to investigate the effectiveness of the acoustic parameters speech rate, mean intensity, F0SD, and IntSD at capturing perceptual changes in PwPD speech over time. This was done by conducting acoustic and statistical analyses (LMMs) on speech data collected from the same group of participants six months apart over two time points. The models for the LMMs were run with each acoustic parameter (the outcome variable) modelled against the predictors *time* (indicating the data collection time) and *group* (indicating the grouping of recordings into either PwPD-change, PwPD-no change, and control).

In this analysis, LMMs revealed that none of the prosody acoustic parameters were able to track PwPD perceptual features over time. These results partially contradict previous studies indicating some impact of speech rate and F0SD in PwPD (Ackermann et al., 1997; Caligiuri, 1989; Flint et al., 1992; Metter & Hanson, 1986; Skodda et al., 2009). However, as stated previously in section 6.1, results on speech rate are inconsistent with a number of studies reporting no change or a reduced speech rate over time, either no group difference between PwPD and control speakers, a faster speech rate than controls, or a slower speech rate than controls (Ackermann, Konczak, & Hertrich, 1997; Caligiuri, 1989; Flint et al., 1992; Metter & Hanson, 1986).

It has been noted that listeners may also perceive a faster speech rate but find no significant deviation from control upon acoustic examination. This may be due to articulatory errors and lack of pausing, which makes it harder for listeners to pick up on cues that indicate acoustic contrasts in the speech signal (Darley et al., 1975; Grosset et al., 2009; Murdoch, 2009). Research has also indicated that there is a significant discrepancy between perceptual and acoustic findings with respect to speech rate. A listener making a judgement of PD speech may perceive a fast rate of speech compared to normative speech even though speaking rates may be revealed to quantifiably be similar (Theodoros & Ramig, 2011). This difference between the perception of speech rate and the actual speaking rate has been posited to be due to potential articulatory imprecision and continuous voicing, which can blur acoustic contrast within the speech signal (Kent & Rosenbek, 1982; Weismer, 1984). Considering this, it becomes particularly important for listeners to pay attention to their perceptual assessment of speech rate. This discrepancy between perceived speech rate and acoustic speech rate influenced the results in the present study.

Reduction in F0 variation has been cited as one of the most noticeable acoustic features in hypokinetic dysarthria speech by Harel et al. (2004). They carried out a longitudinal study on two individuals where F0 variation was found to be markedly lower than controls and showed a declining trend during initial diagnosis and disease progression. However, this was based on conversational speech only. Therefore, the lack of significance between PwPD and control and speech changes in F0SD in this study was likely influenced by the task. The perceptual characteristic of monotonous speech in PwPD is not always reflected in acoustic analysis studies. Harel and colleagues reported differences in F0 measures in read speech and free speech between PD speakers and control speakers were observed when OFF dopaminergic medication but not when ON medication. In addition to the on/off state of the speaker, Metter and Hanson reported a measurable reduction in F0 variability in PD speakers with severe dysarthria. In contrast, no consistent F0 variability was found between PD speakers with mild dysarthria and control speakers (Metter & Hanson, 1986). The effects of on/off state and severity of dysarthria could have contributed to the lack of F0SD differences between PwPD and controls in the present study, as all speakers ON dopaminergic medication at both T1 and T2, and all the participants were rated as mild severity of dysarthria.

In addition, it is possible that PwPD speech in this study did contribute significantly to the perception of the perceptual feature monopitch. Given that monopitch often co-occurs with monoloudness (Kent, 1996; Kim, 1994), it would be logical to surmise the lack of statistical changes over time and group differences in the acoustic parameters mean intensity and IntSD as well.

A potential reason for not having differences between the PwPD and control group could be due to the size of the dataset. While there were 42 PwPD in total, they were divided into two groups and compared against only 10 controls. These controls were age-matched as close as possible but unable to match against the PwPD groups exactly. Therefore, it is possible that a statistical group difference was not found due to being unable to capture inherent age-related differences in speech rate between PwPD groups and the control group. The cause of this variation in speech in older adults may be attributed to several factors, including the presence of a longer vocal tract, an increase in self-monitoring skills, or the modification of speaking rate (Amerman & Parnell, 1992).

6.3. Intelligibility group (IG) analysis

6.3.1. Participant demographics

Analysis was conducted on the complete dataset (n =110) to test for a change in PwPD speech intelligibility. The recordings were grouped based on the overall intelligibility rating based on the SLT ratings and grouped PwPD speech into mild (rated 0 or 1), moderate (rated 2 or 3), and severe (rated 4). All the PwPD recordings were rated as mild, indicating only mild impairment to speech intelligibility. In addition, the overall intelligibility rating was checked for both T1 and T2 data collection time points and grouped based on whether the SLT rating had changed (either increased or decreased) in T2. The final groups for this analysis are presented in Table 72 below:

Table 72. Participant demographics for the IG analysis.

Group	**Number of participants**	**Age range (years)**	**Mean age (years)**
Mild-change	N= 34; M = 22; F = 12	52-81	66 (SD = 7.09)
Mild-no change	N= 29; M = 21; F = 8	50-93	72 (SD = 9.16)
Control	N= 47; M = 17; F = 30	35-86	64 (SD = 12.23)

6.3.2. Segmentation of and extraction of parameters

Segmentation and extraction of the acoustic parameters for all participant recordings were conducted on The Grandfather Passage speech data. The same procedure used for the PFR analysis was also adopted for the IG analysis.

6.3.3. Reliability of annotations

The reliability of the annotations was checked for the IG analysis, similarly to the PFR analysis. This involved manually re-examining 10% of text grids by the primary researcher and the same external researcher used for the PFR analysis. Annotations for the syllable boundaries were manually rechecked, and 98% of annotations between the primary researcher and the external researcher were within 1.5ms of each other. Any discrepancies were corrected manually.

6.3.4. Descriptive statistics

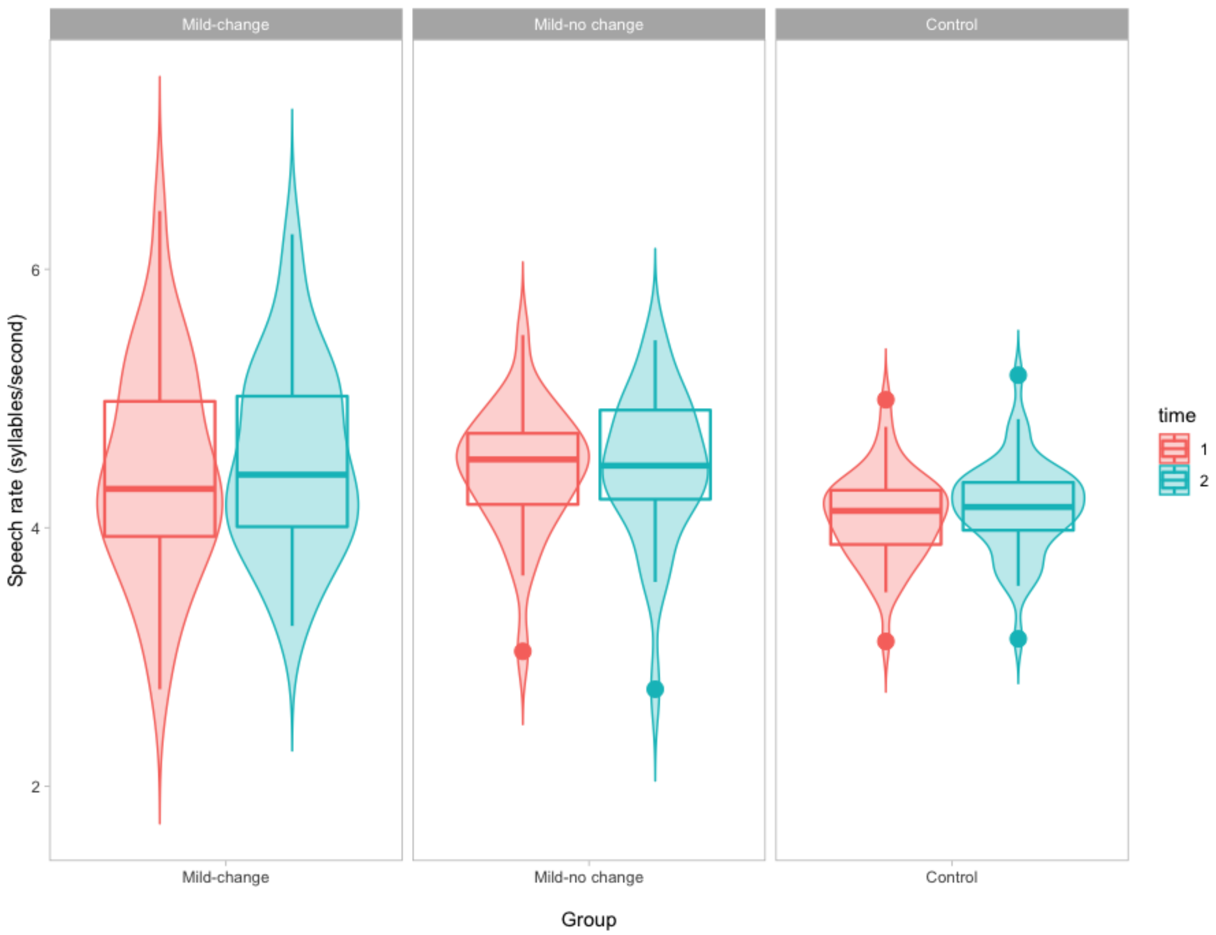
6.3.4.1. Speech rate

The descriptive statistics for speech rate is reported in Table 73 below. The results show that speech rate means increased in the control group and the mild-change group, while the PwPD-no change group remained relatively stable over time. In addition, the means for both PwPD groups was higher than the control group at both time points. There was a higher increase in mean values in the mild-change group from T1 to T2 by 0.10, compared to the control group from T1 to T2 by 0.08. However, it is unclear whether this might be statistically significant for either group based on the means alone. Based on the plotted distribution in Figure 43 of speech rate values for all groups over time, there is a greater variance in distribution within the mild-change group compared to the other groups.

Table 73. Summary table of descriptive statistics of speech rate for each group at T1 and T2. N = 34 (mild-change); 29 (mild-no change); 47 (control).

	Mean	SD	Median	Min	Max
Control (T1)	4.08	0.35	4.13	3.12	4.99
Control (T2)	4.16	0.36	4.16	3.14	5.18
Mild-change (T1)	4.42	0.82	4.30	2.75	6.45
Mild-change (T2)	4.52	0.72	4.41	3.24	6.27
Mild-no change (T1)	4.45	0.48	4.53	3.04	5.49
Mild-no change (T2)	4.46	0.58	4.48	2.75	5.45

Figure 43. The distribution of speech rate of each group (Mild-change, Mild-no change, Control) at T1 and T2.



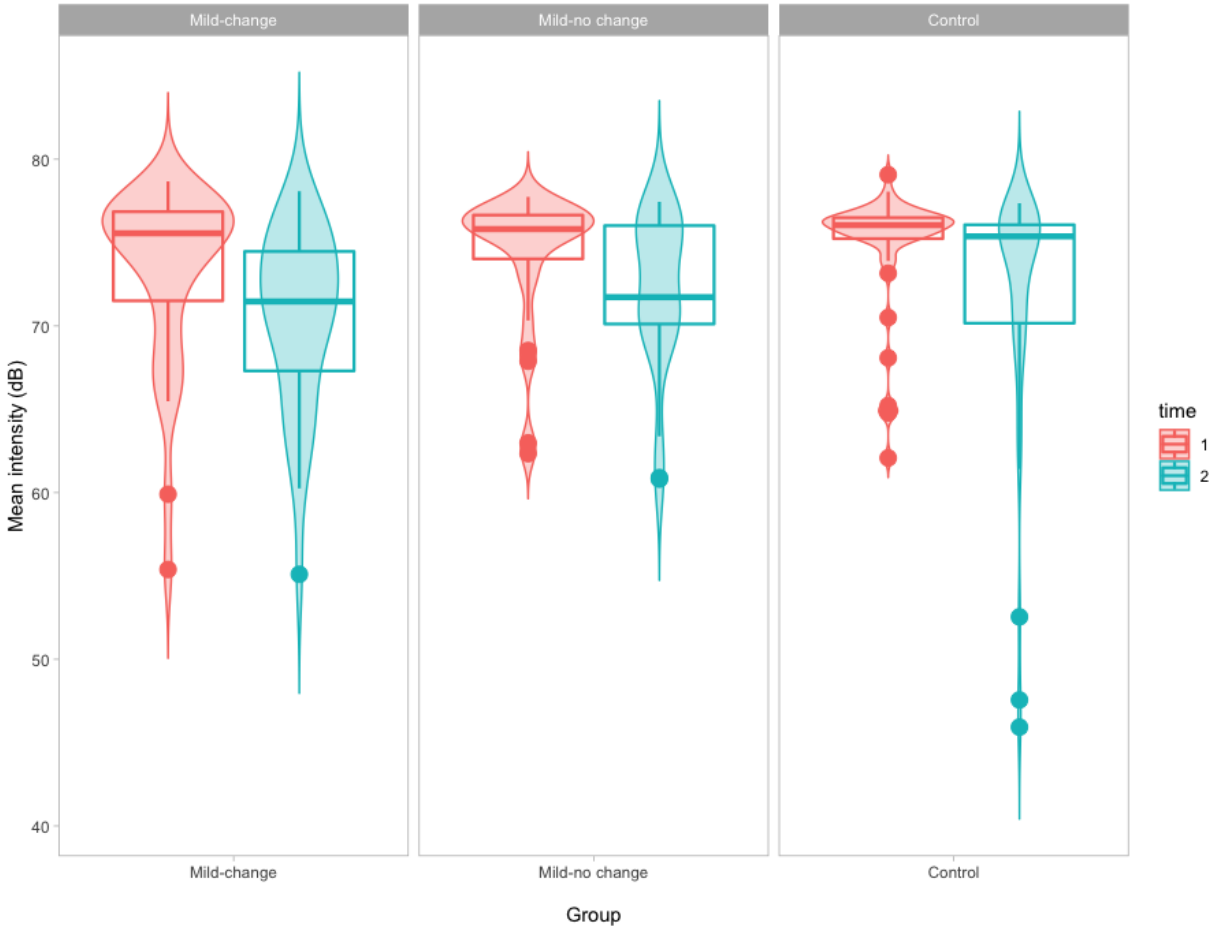
6.3.4.2. *Mean intensity*

Mean intensity descriptive statistics are presented in Table 74 and indicate that the means for all three groups decrease over time. However, mean values between groups seem relatively similar at T1. There is a relatively similar decrease in mean values from T1 to T2 in all three groups: The mild-change group (by 2.81dB), mild-no change group (by 2.1dB), and the control group (by 2.87dB). The plotted distribution in Figure 44 indicates that there may be some potential outliers in the data that will need to be further investigated during statistical analysis. These values likely influence the mean values reported in the descriptive statistics. However, a general look at the distribution suggests a larger variance of values in the mild-change group compared to the other groups. The extent to which mean intensity might be statistically significant is unclear without first investigating if outliers may be skewing the data.

Table 74. Summary table of descriptive statistics of mean intensity for each group at T1 and T2. N = 34 (mild-change); 29 (mild-no change); 47 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	74.78	3.70	76.05	62.08	79.07
Control (T2)	71.91	7.28	75.37	45.93	77.36
Mild-change (T1)	73.30	5.41	75.56	55.39	78.67
Mild-change (T2)	70.49	5.39	71.47	55.10	78.08
Mild-no change (T1)	74.31	4.06	75.81	62.34	77.74
Mild-no change (T2)	72.21	4.62	71.73	60.81	77.44

Figure 44. The distribution of mean intensity of each group (Mild-change, Mild-no change, Control) at T1 and T2.



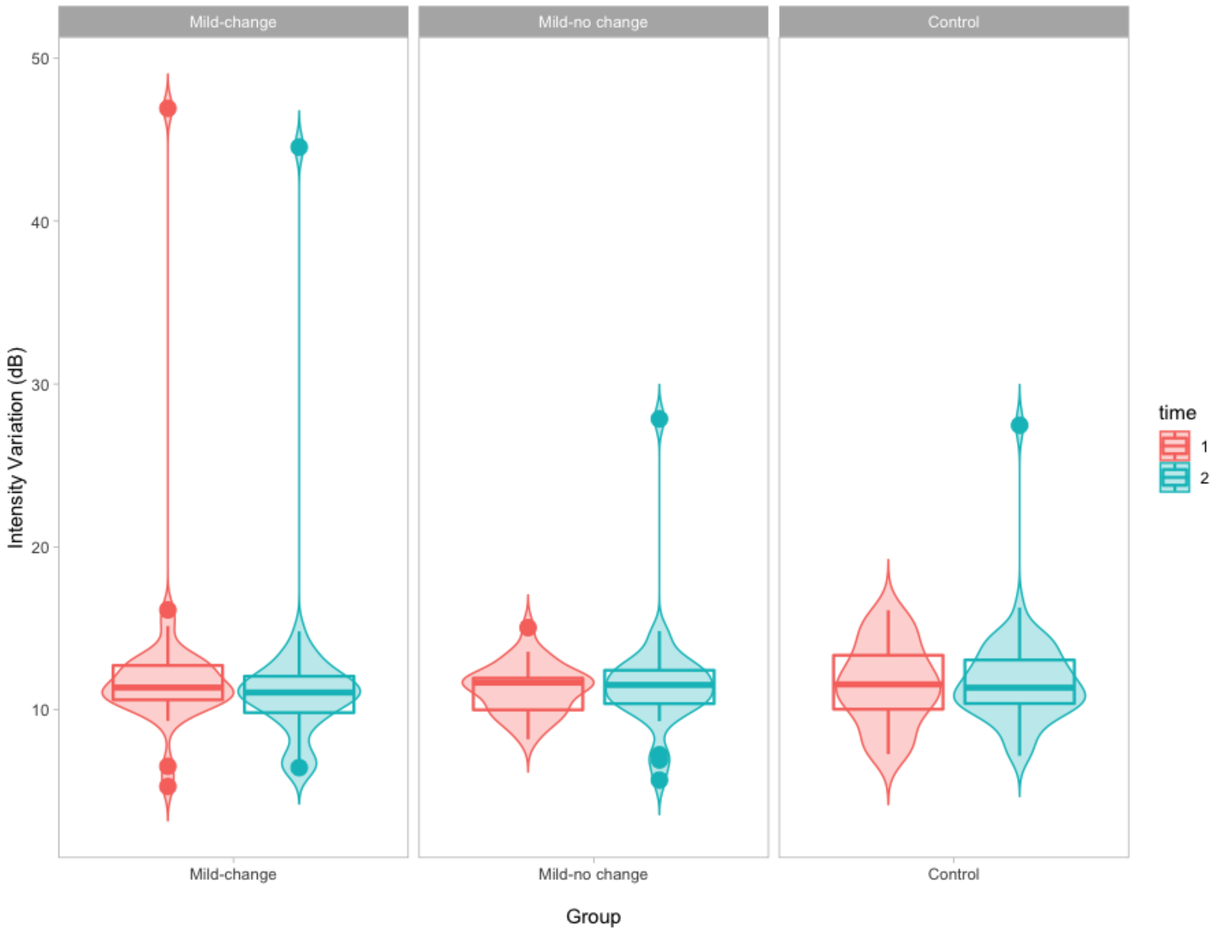
6.3.4.3. IntSD

In Table 75, the IntSD descriptive statistics are reported. These results indicate that the mild-change group had a higher mean value at T1 compared to the other groups and declined at T2. The other groups' means seem to show a slight increase over time, but it is unclear whether these will be statistically significant. Furthermore, much like the mean intensity values, the plot of the distribution of IntSD values in Figure 45 indicates that outliers might exist in the data, which may be skewing the means in the descriptive. Therefore, further analysis should reveal significance once any potential outliers are removed.

Table 75. Summary table of descriptive statistics of intensity variation (IntSD) for each group at T1 and T2. N = 34 (mild-change); 29 (mild-no change); 47 (control).

	Mean	SD	Median	Min	Max
Control (T1)	11.58	2.49	11.56	7.28	16.12
Control (T2)	11.87	3.07	11.36	7.18	27.47
Mild-change (T1)	12.44	6.41	11.37	5.29	46.92
Mild-change (T2)	11.71	6.15	11.06	6.45	44.54
Mild-no change (T1)	11.30	1.53	11.67	8.19	15.05
Mild-no change (T2)	11.70	3.71	11.53	5.68	27.86

Figure 45. The distribution of intensity variation (IntSD) of each group (Mild-change, Mild-no change, Control) at T1 and T2.



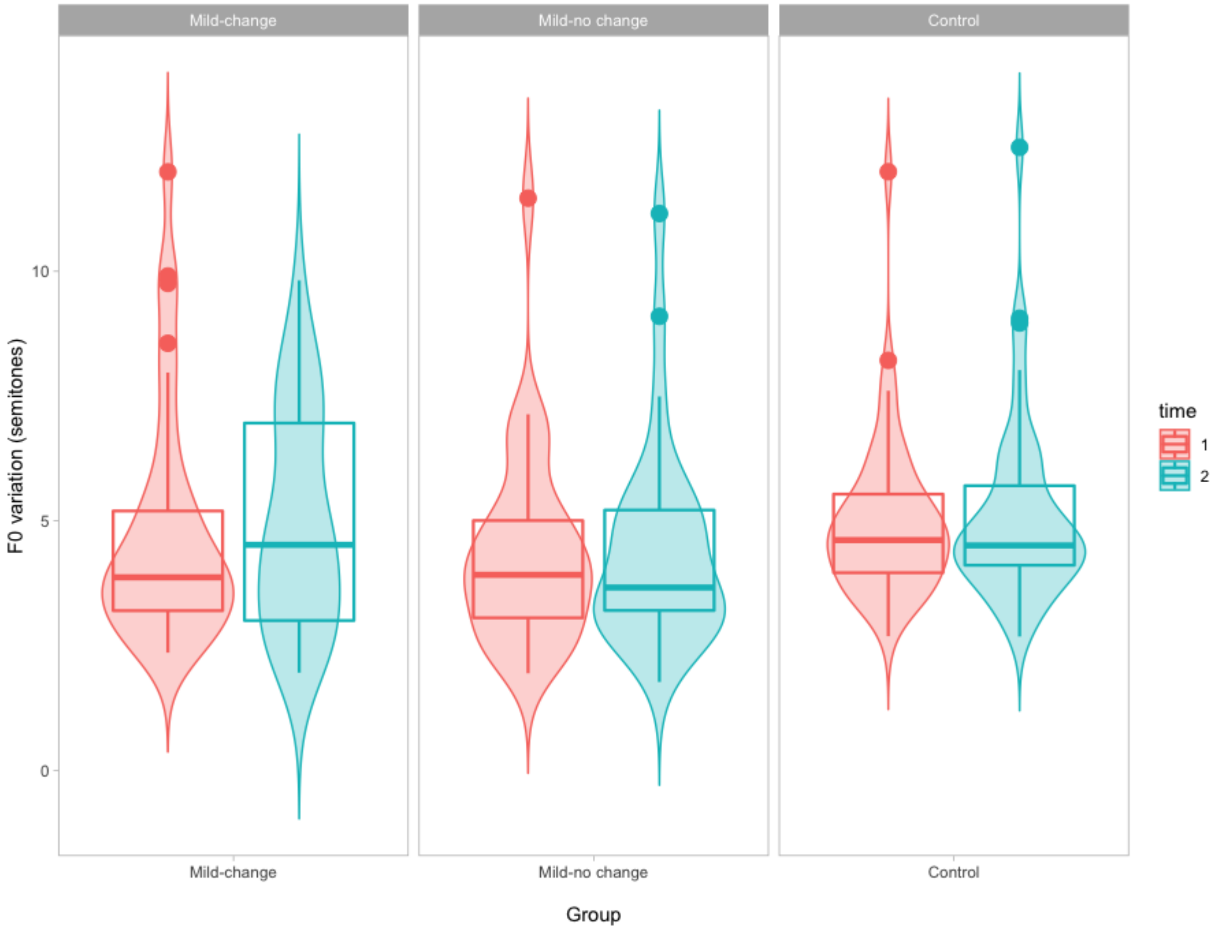
6.3.4.4. F0SD

Finally, the descriptive statistics for F0SD are presented in Table 76. These results suggest that there is an increase in mean values over time for all groups. The mild-change group has a greater increase in mean values from T1 to T2 (by 0.20) compared to the mild-no change group (increased by 0.06) and the control group (increased by 0.11). However, this difference may not be statistically significant. The plotted distribution of all the F0SD for the groups can be seen in Figure 46. The figure indicates that the distribution variance is relatively similar in the mild-no change and the control group over time. However, there is a greater variance of distribution in the mild-change group. This suggests that values may have changed more in the mild-change group compared to the other groups and will be checked for significance in the next section.

Table 76. Summary table of descriptive statistics of fundamental frequency variation (F0SD) for each group at T1 and T2. N = 34 (mild-change); 29 (mild-no change); 47 (control).

	Mean	**SD**	**Median**	**Min**	**Max**
Control (T1)	4.98	1.59	4.62	2.69	12.00
Control (T2)	5.09	1.81	4.51	2.68	12.48
Mild-change (T1)	4.78	2.37	3.87	2.36	11.99
Mild-change (T2)	4.98	2.19	4.52	1.96	9.82
Mild-no change (T1)	4.37	1.96	3.92	1.95	11.47
Mild-no change (T2)	4.43	2.06	3.67	1.77	11.16

Figure 46. The distribution of fundamental frequency variation (F0SD) of each group (Mild-change, Mild-no change, Control) at T1 and T2.



6.3.5. Results of LMMs

A linear mixed effects model analysis was run through R (R Core Team, 2019) to test the significance of each predictor on each acoustic parameter. The models were run with the lme4 package (Bates et al., 2015) using lmer. The model structures were the same as those used for the PFR analysis.

A model was run with each acoustic parameter (outcome variable, i.e., speech rate, mean intensity, IntSD, and F0SD) against the predictors *time* (data collection time points T1 and T2) and *intelligibility group* (mild-change, mild-no change, and control) and random factor of *participant* with a random intercept and random slope. A random intercept means that the final model takes into account that each individual *participant* can show higher or lower values for each acoustic parameter regardless of their *intelligibility group* or the data collection *time*. A random slope for *participant* means that the model considers that some between-participant variation exists.

Based on the assumptions of a linear mixed effects model, the residuals for each model run for each acoustic parameter were checked. The residuals of the IntSD model violated the normality assumption. Transformations were applied to the data but did not resolve normality, so outliers were checked, and four participants' data from the PwPD group was dropped from the variable. The model was rerun, and residual normality was checked and did not violate the assumption. There were no significant correlations between any of the predictors.

A summary table of LMM results of the acoustic parameters speech rate, mean intensity, and F0SD can be seen in Table 77 below, based on 220 observations from 110 participants. The results of the acoustic parameter IntSD are also presented in Table 77 below, with the outliers removed and based on 212 observations from 106 participants.

Results from the LMMs indicated that the effect of both the predictors *time* and *group* had a significant impact on the acoustic parameter speech rate, indicating that speech rate had a significant change over time and was significantly different between groups. Speech rate had a positive change over time of an estimated 0.07 syll/sec. The PwPD-change group had a speech rate an estimated 0.35 syll/sec higher than the control group, and the PwPD-no change group had an estimated 0.33 syll/sec higher speech rate compared to the control group.

The effect of the predictor of *time* had a significant impact on the acoustic parameter mean intensity, with an estimated 2.65dB decrease over time. However, the effect of the predictor *group* did not have a significant impact on mean intensity, indicating that the values were not significantly different between groups. Results do indicate that the PwPD group had a lower mean intensity compared to the control group, though not significant. In addition, the results suggest that all groups varied in mean intensity over time.

Finally, the effect of the predictors *time* and *group* did not significantly impact the acoustic parameters F0SD and IntSD. This indicates that both these acoustic parameters did not significantly change over time and were not significantly different between the PwPD groups and the control group.

Table 77. Linear mixed effects model results of the IG analysis of speech rate, mean intensity, IntSD, and F0SD against each of the levels of the predictors time and group.

Fixed Effects:

Acoustic parameter	Predictor	Estimate	Std. Error	df	t-value	p
Speech rate	Intercept (Control, T1)	4.09	0.08	113.33	51.06	< 0.001**
	T2	0.07	0.03	109.00	2.49	0.01*
	mild-change	0.35	0.12	107.00	2.87	<0.01*
	mild-no change	0.33	0.13	107.00	2.61	0.01*
Mean intensity	Intercept (Control, T1)	74.67	0.67	170.00	111.34	< 0.001**
	T2	-2.65	0.66	109.00	-3.99	<0.001**
	mild-change	-1.45	0.90	107.00	-1.61	0.11
	mild-no change	-0.09	0.94	107.00	-0.10	0.92
IntSD	Intercept (Control, T1)	11.76	0.26	126.69	46.06	< 0.001**
	T2	-0.21	0.19	95.80	-1.13	0.26
	mild-change	-0.30	0.38	104.02	-0.79	0.43
	mild-no change	-0.24	0.39	100.26	-0.60	0.55
F0SD	Intercept (Control, T1)	4.97	0.27	126.44	18.21	< 0.001**
	T2	0.13	0.16	109.00	0.79	0.43
	mild-change	-0.15	0.41	107.00	-0.38	0.71
	mild-no change	-0.63	0.42	107.000	-1.49	0.14

6.3.6. Post-hoc testing

Estimated marginal means (EMMs) were obtained using the "emmeans" package (Lenth, Russel V., 2022) in R (R Core Team, 2019). The package used the Kenward-roger method to calculate degrees of freedom. This was used to compare factors in the relevant significant voice models to pinpoint where exactly the significance lies. The significant models were for the acoustic parameters of speech rate and mean intensity.

EMMs were run to conduct a pairwise comparison of all three groups, and the results automatically averaged (collapsed) over the two levels of the predictor *time* with a confidence level of 95%, giving the mean response value of each level of the predictor *group,* and contrasting it with the other levels. This would help identify how different each group is from the other. A pairwise comparison of the two levels of *time* was also made and averaged over the predictor *group.*

The results of the EMMs for the acoustic parameter speech rate are presented below.

As seen from Table 78 below, there is a group difference between the Control and Mild-change group between the two levels of *time* (E= -0.350, df= 107, $p < 0.05$), the Control and Mild-no change group (E= -0.333, df= 107, $p < 0.05$). However, there is no group significance between the Mild-change and Mild no-change groups (E= 0.016, df=107, $p > 0.05$). There was a significant difference between the two data collection *time* points (E= -0.068, df=109, $p < 0.05$).

The EMMs confirm the results presented for the LMM and indicates that speech rate changed significantly over time and was significantly different between the PwPD groups and the control group. However, the table also indicates that speech rate was not significantly different between the mild-change and the mild-no change group. This result is further discussed in the next section.

Table 78. Pairwise differences of the predictors group and time for speech rate.

Predictor comparisons	**Estimate**	**SE**	**df**	**t-ratio**	**p**
Control – Mild-change	-0.35	0.12	107	-2.87	<0.05*
Control – Mild-no change	-0.33	0.13	106	-2.61	<0.05*
Mild-change – Mild-no change	0.02	0.14	107	0.12	0.99
T1 – T2	-0.07	0.03	109	-2.49	<0.05*

Note: SE = standard error; df = degrees of freedom.

6.3.7. Discussion

This section presented the results of the IG analysis conducted on all participants, analysing the recordings of the grandfather passage. The acoustic parameters speech rate, mean intensity, F0SD, and IntSD were all used to investigate this, similar to the PFR analysis, to test if the acoustic parameters could track changes in PwPD speech intelligibility over time.

Speech rate showed a significant positive change over time and was significantly different between the PwPD groups and the control groups. Post-hoc testing on speech rate confirmed this finding but showed that speech rate was not significantly different between the mild-change and mild-no change groups. This result is unsurprising as both PwPD groups present with the same level of impairment in speech intelligibility. Therefore, acoustic parameters would not be expected to be significant differences between the two PwPD groups. Based on the results, speech rate can track changes in PwPD speech intelligibility over time. This is contrary to the PFR analysis, where a significant effect was not found. However, speech rate contributed more toward the perception of overall speech intelligibility, and the larger sample size revealed overall time and group-related changes more clearly.

Mean intensity showed a significant negative change over time, showing that this acoustic parameter can track these changes. However, there was no significant difference between groups, indicating that all groups had some negative change in mean intensity over time. This would suggest that both PwPD groups change in mean intensity over time within a normal range of variation displayed by the control group. This would coincide with the results of IntSD and F0SD, which did not change significantly over time and was not significantly different between groups. Therefore, PwPD speech remained within a normal range of loudness and pitch compared to the control group. However, further investigation is required to establish the range of normal variation in a control group over time, especially since the control group in this study was not age-matched to PwPD speakers.

In Skodda et al. (2009), speech rate declined from the first to the second data collection time, especially in male participants. However, these values did not show a significant difference from the values of the control group. F0 variation also showed some gender disparities where the female participants with PD had decreased pitch variability over time, while the male participants' values remained relatively stable. Finally, the F0 variation in both male and female participants was significantly reduced compared to the control group in the first and second data collections. The present study could not get age and sex-matched controls to compare against PwPD participants and, therefore, cannot discount any sex-specific variation in the acoustic parameters. It is possible that the F0SD may have been significantly impacted by one sex compared to the other, but further investigation with a balanced sample is needed to confirm this.

The tendency for speech rate reduction over the course of PD disease progression might serve as an explanation for previous inconsistent findings in speech rate compared to control speech. The previous research often did not account for sex-related influences and the stage of disease (Ackermann et al., 1997; Caligiuri, 1989; Flint et al., 1992; Metter & Hanson, 1986). This is supported by a previous longitudinal study which found that articulatory rate declined as PD disease progression occurred (Skodda & Schlegel, 2008). Sex-related disparities might be related to the differences in laryngeal size between sexes and have a different impact on disease-specific changes in the vocal apparatus in PwPD males and females (Hertrich & Ackermann, 1995; Skodda et al., 2009).

6.4. Chapter conclusions

The results of the PFR analysis indicated that none of the acoustic parameters could track PwPD perceptual changes over time. The results of the IG analysis showed that speech rate and mean intensity was the only acoustic parameters able to track changes in PwPD speech intelligibility over time. However, only speech rate could distinguish between PwPD and control speech. The results of the prosodic acoustic parameters investigated show that none of them can differentially diagnose between groups (based on the PFR analysis) or track changes in perceptual features. However, there is evidence that speech rate can successfully track changes in PwPD speech intelligibility, a valuable outcome measure that can be used during speech and language therapy.

The differences in the analysis indicate that speech rate can better track changes in PwPD speech intelligibility than changes in PwPD perceptual features. As stated in the discussion of the IG analysis, previous studies have often ignored the progression or stage of disease when investigating speech rate, which has resulted in inconsistent findings (Ackermann et al., 1997; Caligiuri, 1989; Flint et al., 1992; Metter & Hanson, 1986). Therefore, in this study, speech rate has changed concurrently with any change in PwPD speech intelligibility. Since speech rate has been posited to decline as PD disease progression occurs (Skodda et al., 2009), it explains why PwPD participants in this study show slightly higher speech rate than the control group, as they all displayed only mild impairment to speech intelligibility. Further investigation can reveal whether speech rate declines consistently as PwPD speech intelligibility declines or it occurs in a step-wise manner throughout PD progression.

To understand the causes of the differences in prosody between PwPD and controls, a study (Hammer & Barlow, 2010) suggested an abnormal somatosensory laryngeal function in PwPD (19 participants) in comparison to controls (18 participants) associated with the timing of phonatory onset, respiratory driving pressure or lung volume employed per syllable. The study suggested that PD results in somatosensory laryngeal deficits that can lead to phonatory, articulatory and prosodic abnormalities. These somatosensory and auditory mechanisms provide the feedback to produce the accurate adjustments necessary for correct phonation, articulation, and prosody. The authors further suggest some deficits in the basal ganglia's integration of the stimuli received in the sensory cells while speaking, resulting in a malfunction of voice and speech mechanisms. The abnormalities in the laryngeal somatosensory function produce deficits in prosodic acoustic parameters, including speech rate (Moro-Velazquez & Dehak, 2020).

The results of three speech categories of voice quality, articulation, and prosody have been presented and discussed. The following chapter will give a general discussion of the findings of this study, linking it to the central focus, some limitations, and the impact and future research directions on this topic.

7. General Discussion

This project explored the use of acoustic speech markers to track PwPD speech changes over time by answering two research questions. The first research question aimed to identify which acoustic parameters could track perceptual changes in PwPD speech over time. The second research question attempted to identify which acoustic parameters could track changes in PwPD speech intelligibility over time. In order to answer the research questions, data were collected at two time points, six months apart, from 63 PwPD and 47 control speakers. Recordings of PwPD speakers were rated across seven categories within all five speech subsystems by SLTs. Ratings by SLTs indicated that voice quality, articulation, and prosody were the most severely impacted categories in PwPD speakers.

A subset of PwPD speakers, who showed deviant speech characteristics (rated as ‘marked’ or ‘severe’ by SLTs during perceptual assessment) in voice quality, articulation, or prosody, were selected for perceptual features rating (PFR) analysis. The PFR analysis investigated whether acoustic parameters that correlate to the perceptual features could track speech changes in PwPD speech across the two data collection time points (T1 and T2). The PwPD speakers selected were split into PwPD-change (N = 21) and PwPD-no change (N = 21) to distinguish between speakers who were rated by SLTs as having had a negative change in speech from T1-T2, or no change from T1-T2. These two groups were compared against a subset of control speakers (N = 10). The analysis itself involved conducting a Linear Mixed Model analysis (LMM) with models run for each acoustic parameter against the predictor variables *time* (T1, T2), and *group* (PwPD-change, PwPD-no change, control), and the random factor *participant*.

An intelligibility group (IG) analysis was conducted to answer the second research question by including the complete dataset of speakers and using SLT ratings of overall speech intelligibility. All PwPD speakers were rated by SLTs on speech intelligibility as being only mildly affected. The IG analysis investigated whether acoustic parameters that quantified specific perceptual features in each of the three categories could track changes in PwPD speech intelligibility between T1 and T2. To conduct the analysis, PwPD speakers were split into Mild-change (N = 34) and Mild-no change (N = 29) to distinguish between speakers who were rated by SLTs as having had a negative change in speech intelligibility from T1-T2, or no change from T1-T2. These two groups were compared against all the control speakers (N = 47). The analysis itself involved conducting a Linear Mixed Model test with models run for each acoustic parameter against the predictor variables *time* (T1, T2), and *group* (Mild-change, Mild-no change, control), and the random factor *participant*.

This chapter will discuss the results of the PFR analysis and IG analysis and contextualises them using the literature presented in chapter 1 and 2.

7.1. Acoustic markers for tracking perceptual features and intelligibility in PwPD speech over time

Based on the results of the PFR analysis across voice quality, articulation, and prosody, the articulatory acoustic parameters of mean intensity of /s/ and /ʃ/ were the only measures able to both differentially diagnose between PwPD and control speech and track changes in perceptual features over time.

The acoustic parameters unable to distinguish between PwPD and control speakers were CPP (voice quality), average plosives intensity, mean intensity of these plosives: /b/, /t/, /d/, /k/, /g/, average fricatives intensity, and mean intensity of these fricatives: /f/, /v/, and /z/(articulation). These acoustic parameters did detect significant speech changes between T1-T2. Significant changes in the mentioned acoustic parameters indicate that all speakers (regardless of group) had some change from T1-T2 suggesting these acoustic parameters may be able to capture normal speech changes over time.

Descriptive statistics suggested the largest acoustic change lay within the PwPD-change group compared to the PwPD-no change and control group. The acoustic parameters showing some change in all speakers, including those in the PwPD-no change and control group, could be an indication of acoustic changes that do not result in crossing a threshold required for a detectable perceptual change for listeners. Since speakers in the PwPD-no change group were rated by SLTs as not having a perceptual change from T1-T2, and control speakers should also show no significant change, the interpretation that the acoustic changes found may not have resulted in a detectable perceptual change is plausible. However, this study did not investigate at which point an acoustic change might become detectable perceptually and, therefore cannot confirm what the minimal detectable perceptual difference might be.

The acoustic parameters of mean intensity of /s/ and /ʃ/ (quantifying imprecise consonant production) were able to capture perceptual changes in PwPD speech over time and indicated a negative change, which matches with the results of the SLT ratings implying a deterioration in PwPD speech from T1 to T2. This is consistent with previous findings (Harel et al., 2004; Holmes et al., 2000; Skodda et al., 2009, 2012, 2013). The other acoustic parameters found significant in voice quality and articulation were able to capture general speech changes across all groups. The results of this analysis have answered the first research question in a preliminary way and act as a first-pass investigation into which acoustic parameters may be better than others in tracking perceptual changes in PwPD speech.

There are several studies that show acoustic parameters are able to quantify perceptual characteristics of speech in PwPD (Karlsson & Hartelius, 2019; Kent et al., 1999, 2003; Rusz et al., 2011). However, most of these studies have focused on the use of acoustic analysis to identify acoustic parameters effective for differential diagnosis of PD (Bunton et al., 2000; Darley et al., 1969b, 1969a; Liss et al., 2009; Rosen et al., 2006; Rusz, Rusz, et al., 2011), and a limited number of studies have investigated whether specific perceptual features can be tracked over time (Harel et al., 2004; Skodda et al., 2009, 2013). This poses a challenge since perceptual judgements are subjective, and multiple features are often correlated (Kent, 1996; Oates, 2009), which can mean individual acoustic parameters may not be able to capture an individual perceptual feature if it changes relative to other perceptual features. This means that the psychometric properties of a perceptual rating system may be inadequate in quantifying how much each individual perceptual feature may have changed independent of the other perceptual features since each feature is hard to disentangle when listening to PwPD speech (Kent, 1996). This may be where acoustic parameters can be more accurate in quantifying speech change, but a change in an acoustic parameter may be related to more than one perceptual feature which makes interpretation tricky.

Previous studies have demonstrated that there are either only low or moderate associations between instrumental measures such as acoustic analysis and perceptual ratings of dysphonia (Kempster et al., 2009; Ma & Yiu, 2006; Yiu, 1999). Shrivastav (2003) suggests that studies trying to correlate acoustic parameters and perceptual ratings have been inconsistent, with limited interpretability and contradictions in which acoustic parameters have a stronger correlation with specific perceptual features more than others. The extent to which inconsistencies in previous findings may be the result of methodological differences and complexities involved in the study of the relationship between acoustic parameters and perceptual ratings or the result of limitations that may be inherent in the relationship is unclear (Oates, 2009).

A difficulty in correlating acoustic parameters and perceptual features can be seen in the results of the PFR analysis on voice quality, where jitter, shimmer, and HNR did not show a statistically significant change over time even though the perceptual features they quantify (hoarse, breathy, strain-strangled) indicated a negative change over time in descriptive statistics. This result may indicate that either the acoustic parameters jitter, shimmer, and HNR are unable to capture perceptual changes effectively or are unable to disentangle changes in individual perceptual features.

The significant change in only CPP values over time may indicate that acoustic parameters that attempt to capture perceptual changes over time are better at doing so if they capture more general measures of voice quality, such as dysphonia, rather than specific perceptual features, such as hoarse, breathy, and strain-strangled that are often correlated features (Kent, 1996; Oates, 2009) and therefore may be hard to capture individually using jitter, shimmer, and HNR. In addition, jitter, shimmer, and HNR are extracted from sustained phonation, which relies on fundamental frequency computations which may not be able to capture perceptual changes in voice quality that has more moderate dysphonia (Murton et al., 2020). This is an important observation as even though SLT ratings individually rated certain speech features as 'marked' or 'severe', the perceptual change may be better captured by CPP, which was extracted from the grandfather passage and doesn't rely on fundamental frequency computation.

The argument that perceptual changes may be better looked at using acoustic parameters that are able to encompass more perceptual features would be supported by the fact that CPP was able to track speech changes over time, however, CPP was not effective at distinguishing between PwPD and control speech. Perceptual features of voice will not be useful in making judgments on overall speech performance on their own. Therefore, combining perceptual judgements methods that measure vocal tract function, or tracking acoustic parameters that evaluate voice change over time and following intervention are a good approach for future studies (Oates, 2009).

The PFR analysis results for articulation showed that the acoustic parameters were able to capture changes in PwPD speech over time. Most of the plosive mean intensities (except /p/) and all the fricative intensities indicated a significant change from T1 to T2 in line with the SLT ratings indicating that perceptual changes can be tracked. However, only mean intensity of /s/ and /ʃ/ were able to both differentially diagnose between PwPD and control speech and track speech changes, indicating that the other acoustic parameters were either unable to capture a significant group difference, or that PwPD speech and control speech was not significantly different except in /s/ and /ʃ/ mean intensities. Once again, the SLT ratings of the perceptual features may need more than one acoustic parameter to capture change and since the ratings did not isolate which feature of articulation was impacted the most, the acoustic parameters in this case can be used to determine that it lies in fricative production, specifically /s/ and /ʃ/. Since rigidity is the one of the main motor symptoms of PD which is associated with reduced range of movement, and therefore reduced articulatory precision consonant production is consequently impacted (Argüello-Vélez et al., 2020; Y. Kim, 2017; Tykalova et al., 2017). The results of the thesis showed the fricatives /s/ and /ʃ/ as effective at tracking perceptual feature changes in PwPD over time, potentially narrowing down the area of articulatory imprecision. However, this result is limited in its interpretation as only intensities were investigated, and pitch glides or consonant contrasts were not investigated. Researching this may yield further insight into fricative production in the PwPD sample. In addition, SLT ratings were based on judgements on perceptual features rather than on transcriptions or identification of speech errors in specific phonemes. A deeper perceptual analysis could lead to a better comparison between perceptual ratings and the significant articulation acoustic parameters found in this study.

It is also possible that there was more individual variation that influences the results for each group. However, not all PwPD participants were rated as having 'marked' and 'severe' impairments in voice quality, articulation, and prosody and therefore there may not have been a statistical significance even if the entire participant dataset were included for analysis. This study cannot infer any influence of individual variation as it was not the focus of the study, but research (Skodda et al., 2009) does indicate that speech deterioration may be driven by individual variation since there was no evidence that the time interval between data collection influenced the changes found in acoustic parameters. This may also be the case in the present study and a future study can focus on assessing this.

The other conclusion to be made is that acoustic parameters may be unsuitable for tracking perceptual changes in PwPD speech when overall speech impairment is mild (Kent, 1996). While this may seem counterintuitive, it is possible that greater impairment in only a few perceptual features, while having mild to no impairment in others means acoustic parameters are unable to capture perceptual changes unless multiple perceptual features are severely affected or an entire speech dimension is severely impacted (De Bodt et al., 2002; Kent, 1996; Lansford et al., 2011). A perceptual rating of overall speech intelligibility may be easier to track as it encompasses all perceptual features, and some acoustic parameters may be more capable of tracking changes in PwPD intelligibility even if they are unable to isolate individual changes in specific perceptual features. It suggests that there may be a threshold of severity that each perceptual feature may have to cross for certain acoustic parameters to reliably track perceptual changes.

The results from the IG analysis indicate that some acoustic parameters in voice quality, articulation and prosody were able to track changes in PwPD speech intelligibility over time. These included CPP in voice quality, average plosives intensity, mean intensity of /t/, /d/, /k/, /g/, average fricatives intensity, and mean intensity of /s/, /z/ and /ʃ/ in articulation, and speech rate in prosody.

In the IG analysis, jitter, shimmer, and HNR were able to distinguish between PwPD and control speech but were unable to track changes in PwPD speech intelligibility. There is some agreement in the results between the PFR analysis and the IG analysis which indicates two common acoustic parameters (mean intensity of /s/ and /ʃ/) are effective at capturing PwPD speech changes over time. However, more acoustic parameters were able to capture changes in PwPD speech intelligibility. It is logical to infer then that acoustic parameters seem to easily capture changes in PwPD speech intelligibility more effectively than perceptual changes.

It is possible that acoustic parameters can better capture changes in PwPD speech intelligibility because the SLT ratings were based on specific perceptual features. Acoustic parameters were unable to reliably track changes in specific perceptual features in this study (PFR analysis), and overall intelligibility was better captured because intelligibility encompasses multiple perceptual features that pertain to a number of speech subsystems. Individual perceptual features contribute to speech intelligibility in PwPD to varying degrees with articulatory perceptual features contributing more to speech intelligibility compared to others (De Bodt et al., 2002). This would suggest that tracking PwPD speech intelligibility can encompass multiple acoustic parameters that may be able to provide insight into changes in specific speech subsystems. In evaluating the link between acoustic parameters and perceptual features in relation to speech deterioration in PD, Skodda and colleagues (2013) found that no one acoustic parameter was able to capture overall speech intelligibility, but rather the combination of multiple acoustic parameters was linked to intelligibility.

As previously stated, due to the subjective aspect of perceptual speech ratings, it is possible that rating overall intelligibility was more accurate than rating individual perceptual features (which may be perceptually similar to each other (Kent, 1996)). Zeplin and Kent (1996) (as cited by Kent, 1996) argue that the ratings of certain perceptual features (such as imprecise consonants, loudness, pitch level, and fast rate) may be more reliable than others (monopitch and monoloudness) as they are perceived as more distinct and have higher inter-rater reliability. This may suggest that ratings of the individual perceptual features may not have been very accurate or tend toward a certain bias. For example, one obstacle in voice ratings is that raters may not agree with each other on which aspects of voice quality are more important when making judgements on disordered speech (Oates, 2009). Therefore, it is hard to agree on which perceptual features to optimize for. Kreiman et al. (1990, 1992) concluded that counter to generally held assumptions that greater clinical training leads to better agreement between raters in audio-perceptual judgements, those with more clinical training resulted in rater agreements to be lower rather than higher.

A study focused on how speech subsystems contribute to intelligibility in dysarthric speech and the extent that each of the subsystems affects overall intelligibility (De Bodt et al., 2002). The study found that articulation was judged by experts as contributing the most to overall intelligibility, followed by voice and prosody. The present study also showed that articulation, voice quality, and prosody were judged by SLTs as most impacted in PwPD speakers. Results presented in this thesis also suggest that there were more acoustic parameters in the category articulation able to track changes in PwPD speech intelligibility compared to voice quality or prosody. Acoustic parameters in articulation may contribute to speech intelligibility more than voice quality or prosody and could be explored in the future. The acoustic parameters found to be significant in this study provide an understanding of change in PwPD speech intelligibility over time and could serve as potential acoustic markers in future longitudinal studies. This can be valuable if applied in investigating the predictive value of acoustic markers in speech intelligibility. Previous studies have investigated the effectiveness of certain acoustic parameters in making intelligibility predictions in disordered speech (Ansel & Kent, 1992; Y. Kim et al., 2011), and in predicting intelligibility gains resulting from speech and language therapy (Fletcher, 2016). The acoustic markers in this study could be investigated in a similar vein to see if they also contribute to speech intelligibility predictions in PwPD speech.

7.2. Speech changes and their relation to PD

The results of this study show a trend for some PwPD speech to have deteriorated over time across all three categories of voice quality, articulation, and prosody. Statistically significant negative changes were summarised in the general discussions in sections 7.1 and generally agree with the findings in previous literature (Harel et al., 2004; Holmes et al., 2000; Skodda et al., 2009, 2012, 2013) and with the results of the SLT ratings in this study. Of the statistically significant findings, a decrease in CPP, indicating an increase in dysprosody over time suggests an increase in rigidity resulting in reduced control of vocal cords (Jiménez-Jiménez et al., 1997; Rusz, Rusz, et al., 2011; Rusz, Tykalová, et al., 2021). In addition, a reduction in the various articulation acoustic parameters of plosive and fricative intensities indicates lowered precision in consonant production could indicate reduced lip, tongue, and jaw mobility (Y. Kim, 2017) which can lead to strain in forcing air through a narrow constriction while maintaining the correct articulatory postures. Lastly, reduced speech rate over time indicates increased rigidity and hypokinesia in the articulatory muscles (Goberman & McMillan, 2005; Solomon & Hixon, 1993). However, these correlations between deterioration in acoustic parameters and their speech-motor symptom progression are only an indication and future studies will need to be conducted to confirm this.

In this study, the results suggest that perceptual changes in PwPD speech was likely due to symptoms associated with PD progression, as all PwPD participants were on dopaminergic medication, were not undergoing speech therapy during the course of the study and did not self-report any significant changes in their PD symptoms. This does indicate that speech changes may occur independent of motor symptom progression and dopaminergic deficits as seen in Goberman, (2005); Goberman et al., (2002); Ho et al., (2008); Skodda et al., (2009, 2010, 2013), but this theory is speculative for this study as the Unified Parkinson's Disease Rating Scale (UPDRS) assessment were not made available for the study. However, if speech change in PwPD speech is independent of motor symptom progression it confirms previous findings from the studies cited showing that speech progression and stage of PD are non-linear and strengthens the need for future studies focusing on tracking speech changes and how it may be different at different stages of PD. In fact, a study (Skodda et al., 2013) showed that deterioration in PwPD speech did not have any correlation between the time interval between data collection indicating that progression of time and global PD motor symptoms does not directly correlate to speech change. Therefore, the use of acoustic analysis, as in the present study, to assess how speech in PwPD changes should be a continued effort if speech changes independent of motor symptomatology and therefore may not predictably change in parallel to changes in motor symptoms.

In a study on voice and speech performance in PD using acoustic analysis to test changes in PD speakers on dopaminergic mediation and comparing it to their UPDRS scores. Skodda et al. (2013) found that while speech deteriorated over the two time points, motor impairment was generally stable, and they proposed that this result indicated that nondopaminergic mechanisms influenced the decline in speech performance. They tested this by maintaining the same levels of overall motor performance in PwPD by regulating dopaminergic medication as needed to ascertain if speech performance deteriorated regardless. Results suggested that nondopaminergic mechanism may indeed be driving changes in speech performance. This could suggest that the results of the present study may have also been driven my nondopaminergic mechanism since all PwPD participants maintained the same medication throughout the course of the study and self-reported no significant changes in PD symptoms. However, since the dopaminergic levels were not monitored in this study, and motor scores were not available or tested, any potential influence of medication on global motor symptoms cannot be ruled out. While the present study controlled for the general time of day (morning or afternoon) data collection was done at both T1 and T2, it is not clear when participants had taken medication and there may have been an influence on speech. Further, any negative influence of dopaminergic medication on speech performance cannot be discounted or disentangled from the results.

An early cross-sectional study (Ho et al., 1999) showed that speech abnormalities were found even in cases with mild overall global motor impairment in PD, and that articulation and fluency declined in advanced stages of the disease. It may suggest that speech symptoms deteriorate as global motor symptoms decline but as seen from findings in other studies (Goberman, 2005; Goberman et al., 2002; Ho et al., 2008; Skodda et al., 2009, 2010, 2013) where global motor symptoms stay largely stable despite speech deterioration, it is possible that speech changes may follow a certain trajectory with global motor impairment and then behave independently at later stages. Findings from these previous studies also suggest that PD disease duration has far less influence on speech than the impact of global motor impairment (at least up to a point). The present study recruited participants regardless of disease duration to assess speech change in as wide a sample as possible and given that disease duration and speech change do not appear to be closely linked, as evidenced from the studies above, this provided insight into which acoustic markers could track PwPD speech independent of when PD onset occurred. In fact, some PwPD participants in the present study only developed speech symptoms years after their initial PD diagnosis while others reported developing speech symptoms within the first six months. However, an even distribution of PwPD participants at different disease durations was not obtained and the study cannot dismiss any influence of disease duration on speech performance.

Studies (Hilker et al., 2005; J. Jankovic & Kapadia, 2001) have stated that most marked PD progression occurs during the early stages, continuing as PD advances (Baker et al., 1998; A. K. Ho et al., 1998; Luschei et al., 1999), but plateaus after a point. This would be valuable to track as these studies have largely performed analyses on two time points years apart and it was inconclusive if speech change follows motor symptoms progression linearly or not. In addition, it doesn't account for cases (including some participants in the present study) where PwPD only developed speech symptoms at an advance stage of PD (5-10 years later). The attempt to find acoustic speech markers that are able to capture PwPD speech changes seems valuable in providing a better picture of how PD impacts speech and may promote further research in finding what underlying mechanisms, independent of global motor impairment, could be driving these changes.

7.3. Limitations and future research

There are several limitations to this study that influence how the results of this study can be generalised. With regard to participants, the data collection was based on a convenience sampling and therefore the control group was not age and gender matched to the PwPD participants. Effects of sex were also not investigated in the results and there may be a greater change observed in curtained acoustic parameters for one sex over the other. The gender split for each group, since they were not age matched, would not have allowed for such an investigation, but future studies should investigate how the acoustic markers in this study may be influence by sex.

PwPD participants in this study were all on dopaminergic medication and therefore can conclude that speech changes that have occurred are independent of the impact of the medication, but how much of an influence the medication has had on their speech prior to and during the study is not possible to estimate. In addition, the study cannot make assumptions about the natural development of speech performance in PwPD who are not undergoing dopaminergic treatment.

The PwPD participants in this study were rated as having only mild impairment in intelligibility and while it is promising that the statistically significant acoustic parameters were able to track changes in PwPD speech intelligibility with only mild impairment, it does not imply they will be effective in tracking changes in moderately and severely impaired speech intelligibility. There also can be no claim made on the naturalness of speech compared to intelligibility in this study. A related limitation of the study is data collection was conducted at two time points, which limits how effective the acoustic parameters may be at tracking PwPD speech change over a longer period. In addition, the results of the present study do not provide enough information to establish a trend for how the speech of PwPD participants may change in the future. Further investigation will be required with a larger participant sample, more diverse severity groups, and attempt to disentangle any individual variation that may drive changes in particular groups potentially trying to isolate if these changes are related to PD progression or independent of them.

While this study compared acoustic parameters in PwPD speech to controls in order to gauge significant speech changes against normative data, it is important to acknowledge that there isn’t a clear understanding the level of natural variation that occurs amongst control speakers, especially within older age groups. Information about the speech production of older speakers is a central aspect to gaining insight into how speech changes with age and to delineate what acoustic variation occurs naturally from the changes that occur resulting from acquired neurological diseases (Fletcher, 2016). Previous studies have shown an abundance of evidence that speech rate in older speakers slows down with age (Harnsberger et al., 2008; Jacewicz et al., 2009; L. A. Ramig, 1983; Shewan & Henderson, 1988; Smith et al., 1987). In addition, parameters such as increased segment duration and vowel centralization have also been reported to occur in healthy aging older speakers as well as features of dysarthria which showcases the importance of further investigating normative speech changes in older speakers to clarify to what extent these features co-occur (Fletcher, 2016). The results of the present study are still valid as control speakers were compared at both data collection time points indicating that any naturally occurring speech changes have been accounted for. However, since control speakers were not age and gender matched, it is unclear how different acoustic speech changes in PwPD were from the control group.

Future studies would benefit from establishing a baseline with control speech and further investigations into changes in acoustic parameters in older speaker over time would strengthen longitudinal studies in disordered speech.

Finally, the WEMWBS could not account for the impact the COVID lockdown may have had on the well-being of participants prior to the study and during the course of the study. While a significant decline in well-being was not found in the results of the WEBWBS analysis (see chapter 3 for results), participants self-reported general lower arousal at both time points due to lockdown and the scale may not have captured a potential effect of this.

7.3.1. Limitations of some acoustic measures

There are acoustic measures that were used for analysis in this study that present some reliability issues due to the nature of collecting speech data online. While collecting speech data online provided a convenient and replicable method, there are certain acoustic measures such as jitter and shimmer that are more susceptible to greater random errors. This increase in random errors in the jitter and shimmer values are due to the frequency response of the device microphone used and the effects of background noise, or reverberance and has been previously tested in studies comparing studio microphones against various smartphones (Jannetts et al., 2019; McLaren et al., 2021; Uloza et al., 2015). This study was unable to control for these factors as remote data collection relied on the devices that each participant had available to them, largely laptops whose microphones could not be compared against studio microphones. However, participants were asked to use the same device at both data collection points to ensure that there was no variation in the device microphone for each participant.

The values of the voice quality measures jitter, shimmer, and HNR, may have been impacted by the method of data collection used which may have skewed the results in this study. However, CPP is robust against a number of the limitations listed in this section as it measures a trend over time which avoids obscuring any impact of noise

(Schultz & Vogel, 2022). Therefore, CPP provides a better indication of the speech change in voice quality in PwPD within this study. It may be possible to run a subset analysis in the future controlling for device type, if further information on the devices for each participant are acquired. This can then be compared to the CPP results to see if jitter, shimmer and HNR show similar changes from T1 to T2 in PwPD.

Absolute intensity is also impacted by the frequency response of device microphones (Schultz & Vogel, 2022) which may have resulted in an error in the values due to differences in participant devices. One way to investigate whether the results presented in this study are robust would be to conduct a secondary analysis by measuring the relative intensity of the plosives and fricatives and their following vowel, in The Grandfather Passage of the present study. It has been evidenced that measures of relative intensity can predict speech intelligibility well (Oxenham et al., 2017). Studies have also investigated measuring the relative intensity of plosives and fricatives while manipulating the burst level of plosives, or manipulating the sound pressure level of the burst using masking (Hazan & Simpson, 1998; Kapoor & Allen, 2012; Ohde et al., 1995; Ohde & Stevens, 1983). The results indicated that the relative energy of plosive-burst has a direct one-to-one relationship to the consonants' intelligibility. A secondary analysis of the data in the present study in the future using relative intensity can be compared against those of the absolute intensity measures. A similar result would indicate that the absolute intensity measures were able to track perceptual changes and changes in intelligibility in PwPD speech, and differing results may reveal relative intensity to be a more robust acoustic parameter. In addition, further information on devices used by participants can also be factored in the LMM analysis to control for the effect of different devices on the acoustic parameters.

7.3.2. Further analyses with other speech data

Speech data that was analysed was limited to sustained phonation and grandfather passage and therefore cannot make any judgements on the extent to which the results of this study will apply to conversational speech. Investigation is required to see if these results can generalise to spontaneous speech or whether other acoustic parameters are better at capturing perceptual and speech intelligibility changes in PwPD in extemporaneous speech.

The acoustic analysis conducted was based on the sustained phonation and grandfather passage in order to achieve comparable data but it is well known from the literature that speech performance is influenced by the kind of speech task that is analysed (Skodda et al., 2009). Indeed, previous studies have shown that structured speech such as word lists, sentences, and reading passages may show the influence of external cues on prosodic parameters (Möbes et al., 2008; van Brenk et al., 2022), and indicates a 'forced' elicitation to an external cue (Siegert et al., 2002) which would not capture the entire range of speech dysfluencies and impairments as spontaneous speech. Previous studies conducted on spontaneous speech (Goberman, 2005; Harel et al., 2004; Picheny et al., 1985) showed that speech in PwPD was markedly poorer than control speech and acoustic parameters measured at lower values than in structured speech tasks. The structured nature of reading itself constrains speakers and will not provide enough insight into how PD may impact day-to-day interactions. A future study can use the data collected during this project to investigate these claims.

The present study collected a number of speech data that was not analysed including a list of minimal pairs which would provide insight to whether PwPD participants showed any impairments in producing sound contrasts and in articulatory precision. In addition, two recordings (of roughly one minute each) of spontaneous speech were collected from participants based on two different cues (see Appendix E). The acoustic parameters found significant in this study could be used to analyse the spontaneous speech collected in the future and assess whether results can be generalised to this type of speech data, and whether there is a greater speech change over time in PwPD speech as suggested in previous literature (Goberman, 2005; Harel et al., 2004; Picheny et al., 1985).

7.3.3. Potential for online data collection in future studies

Online data collection is more resource efficient and with more access to options that still provide high quality audio recording. Online data collection can increase the number of participants being recruited in future studies, as well as provide a sustainable option for larger scale, longitudinal projects in the future. The method of data collection employed for the present study can be replicated (Murali, 2022) and therefore compared in the future. In addition, it allows this project to be expanded in the future by adopting the same method and gathering a larger dataset of the potential for following up on the same participants. The large sample size in this study provides a database of PwPD speech and strengthens the benefit of using an online method of collecting speech data.

This method of data collection increases the ecological validity of studies as speech data would be collected in ‘natural’ environments that PwPD speakers usually interact in. While it becomes imperative to place some controls to limit the influence of noise masking speech production, or other artifacts influencing acoustic analysis, an argument can be made for promoting more remote forms of data collection to make results more generalisable. Further, it allows research to study speech in a cost-effective manner while optimizing speech data collection for speech monitoring purposes.

7.3.4. Other applications resulting from the present study

This study could feed into current perceptual methods of evaluation and aid in the assessment stage of speech therapy. By continuing research into acoustic markers of HD associated with PD, it could provide a reliable method that allows early intervention by SLTs, as well as provide patients with a concrete way to observe changes in their speech as a result of PD progression. One of the self-reported impacts of PD that participants of this study shared, was that most participants were unable to isolate how or how much their speech had changed, and only being aware of changes over longer periods or from what friends and family members reported. Results from the present stud provides some clarity to PwPD and can be delivered remotely, and taken further through the development of monitoring applications in the future.

In addition, while some studies have investigated changes in PwPD speech longitudinally (Harel et al., 2004; Holmes et al., 2000; Skodda et al., 2009, 2012, 2013) over two time points, the interval between data collection times were large and variable. To the best of the author's knowledge, there have not been longitudinal studies investigating changes in PwPD speech using acoustic analysis at fixed intervals. This study forms the first of potential future studies that would be able to accurately track the nature of PwPD speech change using acoustic markers identified and provide insight into weather speech subsystems in PwPD speech decline continuously, or at a stepwise rate. The added advantage of employing an online method of data collection may have a positive impact on sample size in the future. Future longitudinal studies on voice and speech are required, ideally with a baseline established as soon as the first motor signs of PD are noticeable and with defined and fixed intervals of data collection conducted regularly in the course of disease progression.

Future research can look into employing automatic detection processes (e.g., machine learning) to form a low-cost, efficient method of identifying speech changes resulting from disease progression. Further, automatic learning systems would become more reliable as more data is fed into them, lessening the load of examinations by SLTs and moving toward a standardised system of diagnosis and prognosis that adds to current perceptual methods. Future research can delve into finding an appropriate automatic process that could be built to adopt the markers found in this study and test it on a larger sample of PwPD speakers. If found to be robust, the same could be systematically researched and employed for other types of dysarthria as well. In the long-term, such an approach would be a more sustainable system than those currently employed, while creating a database of dysarthric speech.

The present study focused on acoustic parameters being used to track speech changes in PwPD speakers in relation to speech ratings given by SLTs in order to contextualize the acoustic correlates of perceptual features. There may have been acoustic parameters that were effective at tracking speech changes in PwPD that have not been perceptible to SLTs (Skodda et al., 2013). This is rather hard to quantify and would involve analysing a much larger group of acoustic parameters without being able to hypothesize where an underlying impairment may be. Therefore, any acoustic parameter that is found to effectively track speech changes would have little bearing on its usefulness in tracking perceptible changes over time, unless the acoustic parameter is shown to detect early signs of future deterioration. This may be a worthwhile pursuit but would be harder to conduct on a large sample without looking at individual speakers.

7.4. Conclusions

Not one person with dysarthria will sound exactly like another person. Even though two people's dysarthria may be caused by similar neurogenic origins, their speech characteristics can be significantly different in speech rate, lexical pressure, vowel and consonant pronunciation, and vocal characteristics can also vary widely (Fletcher, 2016; Y. Kim et al., 2011). There are significant differences in the presentation of speech symptoms, and how the speech of a person with dysarthria will be perceived depends on the listener’s perceptual assessment of speech, impairment-based factors (e.g., multiple possible lesion site, severity of injury) and individual differences in indexed characteristics of their speech (Duffy, 2020).

Since homogeneity of linguistic features is generally assumed in any classification of dysarthria, detailed studies of the perceptual or auditory characteristics between participants' speech patterns are often not included. The Mayo clinic system of classification (Darley et al., 1969b, 1969a, 1975), was developed over 50 years ago and since that time has been the basis of the only widely accepted dysarthria classification framework (Duffy, 2020). However, studies have provided some evidence that there are limitations to this method of classification (Y. Kim et al., 2011; Weismer, 1984, 2006). The main claim of the Mayo clinic approach is that there is a certain degree of homogeneity within groups compared to between groups of dysarthria. There are supposed to be clusters of perceptual features that are unique and are common within each subtype of dysarthria which makes each subtype perceptually different from the others to an experienced listener (Duffy, 2020). However, it has been suggested that some baseline of natural speech variability will exist in the presentation of speech deficits between speakers and over time. This variability would underlie significant differences in the acoustic measures within any subtype or group investigated, and it has been proposed that the common occurrence of statistically insignificant results in research of treatment protocols of dysarthria are related to this baseline variability (Fletcher, 2016). Therefore, investigating variability between speakers can be valuable as well as assessing variability that may exist within groups over time.

Acoustic analysis can be an advantage to curbing some of the limitations of the Mayo system presented above. Acoustic analysis allows the opportunity to examine speech features using a systematic approach. Even though some perceptual features may co-occur within a given speaker the presence of one acoustic measure will not be directly impacted by the presence of others. Adopting acoustic analysis can be a meaningful way of comparing units of measurement between different speakers and tracking speech over time can help elucidate how variable dysarthric speech might be. However, it should be noted that there still is no one acoustic parameter than is able to detect dysarthria as effectively as a listener can (Liss et al., 2009, 2010; Sapir et al., 2010). In addition, research has shown that some processes such as vocal fold spasticity or increased nasal emission are still difficult to capture by any one acoustic parameter (Kent et al., 1999; Maryn et al., 2010). Regardless, acoustic analysis can provide a way to quantify perceptual features and through a combination of acoustic parameters, encompass an understanding of PwPD speech across speech dimensions.

Over the last two decades, scientists have developed a number of acoustic signal analysis methods aimed at assessing parkinsonian speech (Benba et al., 2016; Eliasova et al., 2013; Rusz, Rusz, et al., 2011; Tsanas et al., 2011). Despite extensive research, issues such as acoustic parameters for early-stage detection or accurate progress estimation remain unresolved. New, robust, and sophisticated speech parametrization methods emerge over time (Hawley et al., 2007; Mekyska et al., 2022; Moro-Velazquez & Dehak, 2020; Novotný et al., 2014; Rosen et al., 2006). A feature with high discrimination power or good ability to monitor disease progression can be proposed and be of value to clinicians as long as they can be quantified and therefore interpreted in relation to PD progression. When the value of a feature changes, clinicians must know what the outcome will be in terms of clinical signs. In other words, features that are clinically interpretable will directly quantify clinical signs but clinically uninterpretable features may only provide correlation between values and clinical signs but will not provide clarity on the relationship between the two (Mekyska et al., 2022). Therefore, quantifying perceptual features using acoustic analysis as in the present study can help interpretability especially when trying to help clinicians identify when a perceptual feature changes (Kent, 1996). This is not a perfect method and not all acoustic parameters can always be highly correlated to their perceptual counterparts, but it does provide a closer link between the two and can help clinicians deduce their relationship.

The present study attempted to use the acoustic correlates of perceptual features to investigate whether acoustic markers could be found that can track PwPD speech over time, specifically answering the following research questions: Which acoustic parameters can track perceptual changes in PwPD speech over time?; Which acoustic parameters can track changes in PwPD speech intelligibility over time?

To investigate these research questions and take advantage of both perceptual and acoustic methods in providing insight into PwPD speech change, the present study recruited two SLTs to perform perceptual feature ratings on PwPD speech (N = 63) collected online at two data collection points, six months apart. Results from the SLT ratings were used to identify speech categories within the five main speech subsystems that were rated as being most impacted in PwPD participants. The speech categories voice quality, articulation, and prosody were investigated by selecting acoustic parameters that corresponded to the perceptual features rated as either ‘marked’ and ‘severely’ deviant by the SLTs. These acoustic parameters were then used to investigate whether the parameters could differentially diagnose (between PwPD and control speech) and track perceptual changes in PwPD speech in the PFR analysis. The acoustic parameters were also used to investigate whether they could track changes in PwPD speech intelligibility in the IG analysis.

Both analyses showed that common acoustic parameters (/s/ and /ʃ/ mean intensities) were able to track both perceptual feature changes and changes in PwPD speech intelligibility over the two data collection time points, indicating that the acoustic parameters could confirm negative changes indicated in the SLT ratings. However, results also showed there were more acoustic parameters able to track changes in PwPD speech intelligibility over time compared to tracking individual perceptual features in PwPD speech.

As stated in the general discussion there is evidence that suggests rating overall speech intelligibility seems to be more accurate than rating individual perceptual features as some perceptual features are easier to delineate from the speech signal than others (Kent, 1996; Oates, 2009; Yorkston & Beukelman, 1978). In addition, there doesn't seem to be enough evidence of a significant difference between naïve and experienced listeners of dysarthria rating speech intelligibility (Oates, 2009; Webb et al., 2004; Yorkston & Beukelman, 1978, 1980). The relationship between speech intelligibility ratings and acoustic measures would need to be explored further, but the acoustic markers in this study are a good first step. There are clear benefits to continuing investigation in the speech intelligibility-acoustic parameter relationship as it has clinical implications as a diagnostic tool to document and track speech deficits, to develop relevant rehabilitation schemes, for treatment, and to verify any improvements in speech. In a related proposal, it has been shown that two people with the same speech intelligibility scores need not have the same speech errors underlying that score (Kent et al., 1989). Acoustic analysis can help highlight these differences in their speech errors and a better understanding of speech intelligibility ratings would strengthen this objective. In other words, exploring the speech intelligibility-acoustic parameters relationship will also benefit investigations of individual speech variations in speech and voice disorders.

While the results of the present study are promising, this chapter presented some limitations to its interpretability including that speech data was restricted to sustained phonation and a standardized reading passage which cannot account for how PwPD may vary with spontaneous speech. In addition, all PwPD participants were rated as having only mild speech impairment and further investigations are required to establish the ability of the significant acoustic markers found in this study on tracking PwPD speech change in other levels of impairment.

Finally, Stephens and Daniloff (1977) as cited by (Kent, 1996) reported that there are lower reliability values and a greater ranges of scores given during perceptual assessments of speech and voice disorders when speech was audio-recorded versus given in-person. The visual cues available during in-person perceptual assessments and the quality of the acoustic signal play important roles in the judgements made. This would be a common limitation of most speech analysis research and ensuring high audio-quality must be controlled in order to combat this. Speech data in the present study was collected online with high audio-quality maintained. An added measure to strengthen the reliability of future perceptual assessments given to speech and voice disorders can be to collect short video recordings of speech prior to acoustic analysis. This will help combine the advantages of online speech data collection with and acoustic analysis in order to track speech changes in PwPD.

Regardless of these limitations, the study presented in this thesis has provided a foundation for acoustic markers that can track PwPD speech change using an online method of data collection that can easily be expanded for future longitudinal studies and collect varied speech data that is high quality, cost effective, and convenient.

8. References

Ackermann, H., Konczak, J., & Hertrich, I. (1997). The Temporal Control of Repetitive Articulatory Movements in Parkinson's Disease. *Brain and Language*, *56*(2), 312–319. https://doi.org/10.1006/brln.1997.1851

Ackermann, H., & Ziegler, W. (1991). Articulatory deficits in parkinsonian dysarthria: An acoustic analysis. *Journal of Neurology, Neurosurgery, and Psychiatry*, *54*(12), 1093–1098. https://doi.org/10.1136/jnnp.54.12.1093

Aileen K. Ho, Ho, A. K., Iansek, R., Marigliani, C., Bradshaw, J. L., & Gates, S. (1999). Speech impairment in a large sample of patients with Parkinson's disease. *Behavioural Neurology*, *11*(3), 131–137. https://doi.org/10.1155/1999/327643

Alexander M. Goberman. (2005). Correlation between acoustic speech characteristics and non-speech motor performance in Parkinson Disease. *Medical Science Monitor*, *11*(3), 109–116. www.medscimonit.com/abstract/index/idArt/15386

Allison, K. M., & Hustad, K. C. (2018). Acoustic Predictors of Pediatric Dysarthria in Cerebral Palsy. *Journal of Speech, Language, and Hearing Research*, *61*(3), 462–478. https://doi.org/10.1044/2017_JSLHR-S-16-0414

Amerman, J. D., & Parnell, M. M. (1992). Speech timing strategies in elderly adults. *Journal of Phonetics*, *20*(1), 65–76. https://doi.org/10.1016/S0095-4470(19)30254-2

Ansel, B. M., & Kent, R. D. (1992). Acoustic-phonetic contrasts and intelligibility in the dysarthria associated with mixed cerebral palsy. *Journal of Speech and Hearing Research*, *35*(2), 296–308. http://search.ebscohost.com/login.aspx?direct=true&db=cmedm&AN=1573870&site=eds-live&custid=s1234290&authtype=ip,shib

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Argüello-Vélez, P., Arias-Vergara, T., González-Rátiva, M. C., Orozco-Arroyave, J. R., Nöth, E., & Schuster, M. E. (2020). Acoustic Characteristics of VOT in Plosive Consonants Produced by Parkinson's Patients. In P. Sojka, I. Kopeček, K. Pala, & A. Horák (Eds.), *Text, Speech, and Dialogue* (pp. 303–311). Springer International Publishing.

AVR (8.1.0). (2021).

Awan, S. N., Roy, N., Zhang, D., & Cohen, S. M. (2016). Validation of the Cepstral Spectral Index of Dysphonia (CSID) as a Screening Tool for Voice Disorders: Development of Clinical Cutoff Scores. *Journal of Voice*, *30*(2), 130–144. https://doi.org/10.1016/j.jvoice.2015.04.009

Baker, K. K., Ramig, L. O., Luschei, E. S., & Smith, M. E. (1998). Thyroarytenoid muscle activity associated with hypophonia in Parkinson disease and aging. *Neurology*, *51*(6), 1592–1598. https://doi.org/10.1212/WNL.51.6.1592

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baumgartner, C. A., Sapir, S., & Ramig, L. O. (2001). Voice Quality Changes Following Phonatory-Respiratory Effort Treatment (LSVT®) Versus Respiratory Effort Treatment for Individuals with Parkinson Disease. *Journal of Voice*, *15*(1), 105–114. https://doi.org/10.1016/S0892-1997(01)00010-8

Benba, A., Jilbab, A., & Hammouch, A. (2016). Voice analysis for detecting patients with Parkinson's disease using the hybridization of the best acoustic features. *International Journal on Electrical Engineering and Informatics*, *8*(1), 108–116. https://doi.org/10.15676/ijeei.2016.8.1.8

Bhidayasiri, R. (2005). Differential diagnosis of common tremor syndromes. *Postgraduate Medical Journal*, *81*(962), 756–762. https://doi.org/10.1136/pgmj.2005.032979

Boersma, P. (2009). Should Jitter Be Measured by Peak Picking or by Waveform Matching? *Folia Phoniatrica et Logopaedica*, *61*(5), 305–308. https://doi.org/10.1159/000245159

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences*.

Boersma, P., & Weenink, D. (2020). *Praat: Doing phonetics by computer* (6.1.09). http://www.praat.org/

Bohnen, N. I., Frey, K. A., Studenski, S., Kotagal, V., Koeppe, R. A., Scott, P. J. H., Albin, R. L., & Muller, M. L. T. M. (2013). Gait speed in Parkinson disease correlates with cholinergic degeneration. *Neurology*, *81*(18), 1611–1616. https://doi.org/10.1212/WNL.0b013e3182a9f558

Brockmann, M., Drinnan, M. J., Storck, C., & Carding, P. N. (2011). Reliable Jitter and Shimmer Measurements in Voice Clinics: The Relevance of Vowel, Gender, Vocal Intensity, and Fundamental Frequency Effects in a Typical Clinical Task. *Journal of Voice*, *25*(1), 44–53. https://doi.org/10.1016/j.jvoice.2009.07.002

Brown, P., & Marsden, C. (1998). What do the basal ganglia do? *The Lancet*, *351*(9118), 1801–1804. https://doi.org/10.1016/S0140-6736(97)11225-9

Bunton, K., Kent, R. D., Kent, J. F., & Rosenbek, J. C. (2000). Perceptuo-acoustic assessment of prosodic impairment in dysarthria. *Clinical Linguistics & Phonetics*, *14*(1), 13–24. https://doi.org/10.1080/026992000298922

Caligiuri, M. (1989). The influence of speaking rate on articulatory hypokinesia in parkinsonian dysarthria*1. *Brain and Language*, *36*(3), 493–502. https://doi.org/10.1016/0093-934X(89)90080-1

Canter, G. J. (1963). Speech Characteristics of Patients with Parkinson's Disease: I. Intensity, Pitch, and Duration. *Journal of Speech and Hearing Disorders*, *28*(3), 221–229. https://doi.org/10.1044/jshd.2803.221

Chenausky, K., MacAuslan, J., & Goldhor, R. (2011). Acoustic Analysis of PD Speech. *Parkinson's Disease*, *2011*, 1–13. https://doi.org/10.4061/2011/435232

Chenery, H. J., Murdoch, B. E., & Ingram, J. C. L. (1988). Studies in Parkinson's Disease: I. Perceptual Speech Analyses. *Australian Journal of Human Communication Disorders*, *16*(2), 17–29. https://doi.org/10.3109/asl2.1988.16.issue-2.02

Cohen, A. S., McGovern, J. E., Dinzeo, T. J., & Covington, M. A. (2014). Speech deficits in serious mental illness: A cognitive resource issue? *Schizophrenia Research*, *160*(1–3), 173–179. https://doi.org/10.1016/j.schres.2014.10.032

Connor, N. P., Abbs, J. H., Cole, K. J., & Gracco, V. L. (1989). PARKINSONIAN DEFICITS IN SERIAL MULTIARTICULATE MOVEMENTS FOR SPEECH. *Brain*, *112*(4), 997–1009. https://doi.org/10.1093/brain/112.4.997

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969a). Clusters of Deviant Speech Dimensions in the Dysarthrias. *Journal of Speech and Hearing Research*, *12*(3), 462–496. https://doi.org/10.1044/jshr.1203.462

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969b). Differential Diagnostic Patterns of Dysarthria. *Journal of Speech and Hearing Research*, *12*(2), 246–269. https://doi.org/10.1044/jshr.1202.246

Darley, F. L., Aronson, A. E., & Brown, J. R. (1975). *Motor speech disorders* (Main Stock 616.855 DAR). Saunders. http://search.ebscohost.com/login.aspx?direct=true&db=cat02513a&AN=qmu.b1047717&site=eds-live&custid=s1234290&authtype=ip,shib

De Bodt, M. S., Hernández-Díaz Huici, M. E., & Van De Heyning, P. H. (2002). Intelligibility as a linear combination of dimensions in dysarthric speech. *Journal of Communication Disorders*, *35*(3), 283–292. https://doi.org/10.1016/S0021-9924(02)00065-5

De Bodt, M. S., Wuyts, F. L., Van de Heyning, P. H., & Croux, C. (1997). Test-retest study of the GRBAS scale: Influence of experience and professional background on perceptual rating of voice quality. *Journal of Voice*, *11*(1), 74–80. https://doi.org/10.1016/S0892-1997(97)80026-4

Dejonckere, P. H., Obbens, C., de Moor, G. M., & Wieneke, G. H. (1993). Perceptual Evaluation of Dysphonia: Reliability and Relevance. *Folia Phoniatrica et Logopaedica*, *45*(2), 76–83. https://doi.org/10.1159/000266220

Dromey, C., Ramig, L. O., & Johnson, A. B. (1995). Phonatory and Articulatory Changes Associated With Increased Vocal Intensity in Parkinson Disease: A Case Study. *Journal of Speech, Language, and Hearing Research*, *38*(4), 751–764. https://doi.org/10.1044/jshr.3804.751

Duffy, J. R. (2020). *Motor speech disorders. [Electronic resource]: Substrates, differential diagnosis, and management* (Forth edition.). Elsevier.

Dykstra, A., Hakel, M., & Adams, S. (2007). Application of the ICF in Reduced Speech Intelligibility in Dysarthria. *Seminars in Speech and Language*, *28*(4), 301–311. https://doi.org/10.1055/s-2007-986527

Eadie, T. L., & Baylor, C. R. (2006). The Effect of Perceptual Training on Inexperienced Listeners' Judgments of Dysphonic Voice. *Journal of Voice*, *20*(4), 527–544. https://doi.org/10.1016/j.jvoice.2005.08.007

Eklund, E., Qvist, J., Sandström, L., Viklund, F., Van Doorn, J., & Karlsson, F. (2015). Perceived articulatory precision in patients with Parkinson's disease after deep brain stimulation of subthalamic nucleus and caudal zona incerta. *Clinical Linguistics & Phonetics*, *29*(2), 150–166. https://doi.org/10.3109/02699206.2014.971192

Elemetrics, K. (1993). Multi-dimensional voice program (MDVP).[Computer program.]. *Pine Brook, NJ: Author*.

Eliasova, I., Mekyska, J., Kostalova, M., Marecek, R., Smekal, Z., & Rektorova, I. (2013). Acoustic evaluation of short-term effects of repetitive transcranial magnetic stimulation on motor aspects of speech in Parkinson's disease. *Journal of Neural Transmission*, *120*(4), 597–605. https://doi.org/10.1007/s00702-012-0953-1

Feigin, V. L., Lawes, C. M., Bennett, D. A., & Anderson, C. S. (2003). Stroke epidemiology: A review of population-based studies of incidence, prevalence, and case-fatality in the late 20th century. *The Lancet Neurology*, *2*(1), 43–53. https://doi.org/10.1016/S1474-4422(03)00266-7

Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. SAGE Publications, Ltd.

Fischer, E., & Goberman, A. M. (2010). Voice onset time in Parkinson disease. *Journal of Communication Disorders*, *43*(1), 21–34. https://doi.org/10.1016/j.jcomdis.2009.07.004

Fletcher, A. R. (2016). *Predicting treatment outcomes in dysarthria through speech feature analysis.*

Flint, A. J., Black, S. E., Campbell-Taylor, I., Gailey, G. F., & Levinton, C. (1992). Acoustic analysis in the differentiation of Parkinson's disease and major depression. *Journal of Psycholinguistic Research*, *21*(5), 383–399. https://doi.org/10.1007/BF01067922

Fonville, S., Worp, H. B., Maat, P., Aldenhoven, M., Algra, A., & Gijn, J. (2008). Accuracy and inter-observer variation in the classification of dysarthria from speech recordings. *Journal of Neurology*, *255*(10), 1545–1548. https://doi.org/10.1007/s00415-008-0978-4

Forrest, K., Weismer, G., & Turner, G. S. (1989). Kinematic, acoustic, and perceptual analyses of connected speech produced by Parkinsonian and normal geriatric adults. *The Journal of the Acoustical Society of America*, *85*(6), 2608–2622. https://doi.org/10.1121/1.397755

Fox, C. M., Morrison, C. E., Ramig, L. O., & Sapir, S. (2002). Current Perspectives on the Lee Silverman Voice Treatment (LSVT) for Individuals With Idiopathic Parkinson Disease. *American Journal of Speech-Language Pathology*, *11*(2), 111–123. https://doi.org/10.1044/1058-0360(2002/012)

Fox, S. H. (2013). Non-dopaminergic Treatments for Motor Control in Parkinson's Disease. *Drugs*, *73*(13), 1405–1415. https://doi.org/10.1007/s40265-013-0105-4

Galaz, Z., Mekyska, J., Mzourek, Z., Smekal, Z., Rektorova, I., Eliasova, I., Kostalova, M., Mrackova, M., & Berankova, D. (2016). Prosodic analysis of neutral, stress-modified and rhymed speech in patients with Parkinson's disease. *Computer Methods and Programs in Biomedicine*, *127*(Journal Article), 301–317. https://doi.org/10.1016/j.cmpb.2015.12.011

Gelzinis, A., Verikas, A., & Bacauskiene, M. (2008). Automated speech analysis applied to laryngeal disease categorization. *Computer Methods and Programs in Biomedicine*, *91*(1), 36–47. https://doi.org/10.1016/j.cmpb.2008.01.008

Gemmeke, J. F., Sehgal, S., Cunningham, S., & Van hamme, H. (2014). *Dysarthric vocal interfaces with minimal training data*. 2014 IEEE Spoken Language Technology Workshop (SLT), South Lake Tahoe, NV, USA. https://doi.org/10.1109/SLT.2014.7078582

Goberman, A., Coelho, C., & Robb, M. (2002). Phonatory characteristics of Parkinsonian speech before and after morning medication: The ON and OFF states. *Journal of Communication Disorders*, *35*(3), 217–239. https://doi.org/10.1016/S0021-9924(01)00072-7

Goberman, A., Coelho, C., & Robb, M. (2005). Prosodic characteristics of Parkinsonian speech: The effect of levadopa-based medication. *Journal of Medical Speech-Language Pathology*, *13*, 51–68.

Goberman, A. M., & McMillan, J. (2005). Relative Speech Timing in Parkinson Disease. *Contemporary Issues in Communication Science and Disorders*, *32*(Spring), 22–29. https://doi.org/10.1044/cicsd_32_S_22

Grosset, D., Fernandez, H., Grosset, K., & Okun, M. (2009). *Parkinson's Disease: Clinican's Desk Reference*. CRC Press.

Hammarberg, B., Fritzell, B., Gaufin, J., Sundberg, J., & Wedin, L. (1980). Perceptual and Acoustic Correlates of Abnormal Voice Qualities. *Acta Oto-Laryngologica*, *90*(1–6), 441–451. https://doi.org/10.3109/00016488009131746

Hammer, M. J., & Barlow, S. M. (2010). Laryngeal somatosensory deficits in Parkinson's disease: Implications for speech respiratory and phonatory control. *Experimental Brain Research*, *201*(3), 401–409. https://doi.org/10.1007/s00221-009-2048-2

Hanson, D. G., Gerratt, B. R., & Ward, P. H. (1984). CINEGRAPHIC OBSERVATIONS OF LARYNGEAL FUNCTION IN PARKINSON??S DISEASE: *The Laryngoscope*, *94*(3), 348???353. https://doi.org/10.1288/00005537-198403000-00011

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

Harnsberger, J. D., Shrivastav, R., Brown, W. S., Rothman, H., & Hollien, H. (2008). Speaking Rate and Fundamental Frequency as Speech Cues to Perceived Age. *Journal of Voice*, *22*(1), 58–69. https://doi.org/10.1016/j.jvoice.2006.07.004

Hawley, M. S., Enderby, P., Green, P., Cunningham, S., Brownsell, S., Carmichael, J., Parker, M., Hatzis, A., O'Neill, P., & Palmer, R. (2007). A speech-controlled environmental control system for people with severe dysarthria. *Medical Engineering and Physics*, *29*(5), 586–593. https://doi.org/10.1016/j.medengphy.2006.06.009

Hazan, V., & Simpson, A. (1998). The effect of cue-enhancement on the intelligibility of nonsense word and sentence materials presented in noise. *Speech Communication*, *24*(3), 211–226. https://doi.org/10.1016/S0167-6393(98)00011-9

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. https://doi.org/10.1177/000348949510400304

Hilker, R., Schweitzer, K., Coburger, S., Ghaemi, M., Weisenbach, S., Jacobs, A. H., Rudolf, J., Herholz, K., & Heiss, W.-D. (2005). Nonlinear Progression of Parkinson Disease as Determined by Serial Positron Emission Tomographic Imaging of Striatal Fluorodopa F 18 Activity. *Archives of Neurology*, *62*(3), 378. https://doi.org/10.1001/archneur.62.3.378

Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic Correlates of Breathy Vocal Quality. *Journal of Speech, Language, and Hearing Research*, *37*(4), 769–778. https://doi.org/10.1044/jshr.3704.769

Hillenbrand, J., & Houde, R. A. (1996). Acoustic Correlates of Breathy Vocal Quality: Dysphonic Voices and Continuous Speech. *Journal of Speech, Language, and Hearing Research*, *39*(2), 311–321. https://doi.org/10.1044/jshr.3902.311

Hillman, R. E., Montgomery, W. W., & Zeitels, S. M. (1997). CURRENT DIAGNOSTICS AND OFFICE PRACTICE: Appropriate use of objective measures of vocal function in the multidisciplinary management of voice disorders. *Current Opinion in Otolaryngology & Head and Neck Surgery*, *5*(3), 172–175. https://doi.org/10.1097/00020840-199706000-00005

Ho, A., Iansek, R., & Bradshaw, J. (2001). Motor instability in parkinsonian speech intensity. *Neuropsychiatry Neuropsychol Behav Neurol*, *14*(2), 109–116. PubMed. http://europepmc.org/abstract/MED/11417664

Ho, A. K., Bradshaw, J. L., Cunnington, R., Phillips, J. G., & Iansek, R. (1998). Sequence Heterogeneity in Parkinsonian Speech. *Brain and Language*, *64*(1), 122–145. https://doi.org/10.1006/brln.1998.1959

Ho, A. K., Bradshaw, J. L., & Iansek, R. (2008). For better or worse: The effect of levodopa on speech in Parkinson's disease. *Movement Disorders*, *23*(4), 574–580. https://doi.org/10.1002/mds.21899

Hoehn, M. M., & Yahr, M. D. (1967). Parkinsonism: Onset, progression, and mortality. *Neurology*, *17*(5), 427–427. https://doi.org/10.1212/WNL.17.5.427

Holmes, R. J., Oates, J., Phyland, D. J., Hughes, A. J., & Andrew J. Hughes. (2000). Voice characteristics in the progression of Parkinson's disease. *International Journal of Language & Communication Disorders*, *35*(3), 407–418. https://doi.org/10.1080/136828200410654

Howe, T. J. (2008). The ICF Contextual Factors related to speech-language pathology. *International Journal of Speech-Language Pathology, 10*(1–2), 27–37. https://doi.org/10.1080/14417040701774824

Huei-Mei Liu, Chin-Hsing Tseng, Fen. (2000). Perceptual and acoustic analysis of speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *Clinical Linguistics & Phonetics, 14*(6), 447–464. https://doi.org/10.1080/026992000415877

Jacewicz, E., Fox, R. A., O'Neill, C., & Salmons, J. (2009). Articulation rate across dialect, age, and gender. *Language Variation and Change, 21*(2), 233–256. https://doi.org/10.1017/S0954394509990093

Jankovic, J. J., & Tolosa, E. (2003). Parkinson's Disease and Movement Disorders: Book Reviews. *European Journal of Neurology*, *10*(5), 603–604. https://doi.org/10.1046/j.1468-1331.2003.00637.x

Jankovic, J., & Kapadia, A. S. (2001). Functional Decline in Parkinson Disease. *Archives of Neurology, 58*(10), 1611. https://doi.org/10.1001/archneur.58.10.1611

Jannetts, S., Schaeffler, F., Beck, J., & Cowen, S. (2019). Assessing voice health using smartphones: Bias and random error of acoustic voice parameters captured by different smartphone types. *International Journal of Language & Communication Disorders*, *54*(2), 292–305. https://doi.org/10.1111/1460-6984.12457

Jiménez-Jiménez, F. J., Gamboa, J., Nieto, A., Guerrero, J., Orti-Pareja, M., Molina, J. A., García-Albea, E., & Cobeta, I. (1997). Acoustic voice analysis in untreated patients with Parkinson's disease. *Parkinsonism & Related Disorders*, *3*(2), 111–116. https://doi.org/10.1016/S1353-8020(97)00007-2

Kalia, L. V., Brotchie, J. M., & Fox, S. H. (2013). Novel nondopaminergic targets for motor features of Parkinson's disease: Review of recent trials: Nondopaminergic Targets for Motor Features of PD. *Movement Disorders*, *28*(2), 131–144. https://doi.org/10.1002/mds.25273

Kapoor, A., & Allen, J. B. (2012). Perceptual effects of plosive feature modification. *The Journal of the Acoustical Society of America*, *131*(1), 478–491. https://doi.org/10.1121/1.3665991

Karlsson, F., & Hartelius, L. (2019). How Well Does Diadochokinetic Task Performance Predict Articulatory Imprecision? Differentiating Individuals with Parkinson's Disease from Control Subjects. *FOLIA PHONIATRICA ET LOGOPAEDICA*, *71*(5–6), 251–260. https://doi.org/10.1159/000498851

Karlsson, F., Olofsson, K., Blomstedt, P., Linder, J., Nordh, E., & van Doorn, J. (2014). Articulatory closure proficiency in patients with Parkinson's disease following deep brain stimulation of the subthalamic nucleus and caudal zona incerta. *Journal of Speech, Language, and Hearing Research*, *57*(4), 1178–1190.

Kempster, G. B., Gerratt, B. R., Verdolini Abbott, K., Barkmeier-Kraemer, J., & Hillman, R. E. (2009). Consensus Auditory-Perceptual Evaluation of Voice: Development of a Standardized Clinical Protocol. *American Journal of Speech-Language Pathology*, *18*(2), 124–132. https://doi.org/10.1044/1058-0360(2008/08-0017)

Kent, R. (1996). Hearing and Believing: Some Limits to the Auditory-Perceptual Assessment of Speech and Voice Disorders. *American Journal of Speech-Language Pathology, 5*, 7–23.

Kent, R. A., & Read, C. (2001). *The acoustic analysis of speech* (Main Stock 414 KEN; 2nd ed.). Singular/Thompson Learning. http://search.ebscohost.com/login.aspx?direct=true&db=cat02513a&AN=qmu.b1018442&site=eds-live&custid=s1234290&authtype=ip,shib

Kent, R. D., & Kim, Y. (2003). Toward an acoustic typology of motor speech disorders. *Clinical Linguistics & Phonetics*, *17*(6), 427–445. http://search.ebscohost.com/login.aspx?direct=true&db=rzh&AN=106703939&site=eds-live&custid=s1234290&authtype=ip,shib

Kent, R. D., Netsell, R., & Abbs, J. H. (1979). Acoustic Characteristics of Dysarthria Associated with Cerebellar Disease. *Journal of Speech, Language, and Hearing Research*, *22*(3), 627–648. https://doi.org/10.1044/jshr.2203.627

Kent, R. D., & Rosenbek, J. C. (1982). Prosodic disturbance and neurologic lesion. *Brain and Language*, *15*(2), 259–291. https://doi.org/10.1016/0093-934X(82)90060-8

Kent, R. D., Vorperian, H. K., Kent, J. F., & Duffy, J. R. (2003). Voice dysfunction in dysarthria: Application of the Multi-Dimensional Voice Program™. *Journal of Communication Disorders*, *36*(4), 281–306. https://doi.org/10.1016/S0021-9924(03)00016-9

Kent, R. D., Weismer, G., Kent, J. F., & Rosenbek, J. C. (1989). Toward Phonetic Intelligibility Testing in Dysarthria. *Journal of Speech and Hearing Disorders*, *54*(4), 482–499. https://doi.org/10.1044/jshd.5404.482

Kent, R. D., Weismer, G., Kent, J. F., Vorperian, H. K., & Duffy, J. R. (1999). Acoustic Studies of Dysarthric Speech: Methods, Progress, and Potential. *Journal of Communication Disorders*, *32*(3), 141–186. http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ593127&site=eds-live&custid=s1234290&authtype=ip,shib

Kent, R., & Netsell, R. (1975). A Case Study of an Ataxic Dysarthric: Cineradiographic and Spectrographic Observations. *Journal of Speech and Hearing Disorders*, *40*(1), 115–134. https://doi.org/10.1044/jshd.4001.115

Kim, H.-H. (1994). *Monotony of speech production in Parkinson's disease: Acoustic characteristics and their perceptual relations*.

Kim, Y. (2017). Acoustic Characteristics of Fricatives /s/ and /ʃ/ Produced by Speakers with Parkinson's Disease. *Clinical Archives of Communication Disorders*, *2*(1), 7–14. https://doi.org/10.21849/cacd.2016.00080

Kim, Y., Kent, R. D., & Weismer, G. (2011). An Acoustic Study of the Relationships Among Neurologic Disease, Dysarthria Type, and Severity of Dysarthria. *Journal of Speech, Language & Hearing Research*, *54*(2), 417–429. https://doi.org/10/10-0020)

Kim, Y., Weismer, G., Kent, R. D., & Duffy, J. R. (2009). Statistical Models of F2 Slope in Relation to Severity of Dysarthria. *Folia Phoniatrica et Logopaedica*, *61*(6), 329–335. https://doi.org/10.1159/000252849

Klatt, D. H. (1975). Voice Onset Time, Frication, and Aspiration in Word-Initial Consonant Clusters. *Journal of Speech and Hearing Research*, *18*(4), 686–706. https://doi.org/10.1044/jshr.1804.686

Kreiman, J., & Gerratt, B. R. (2000). Sources of listener disagreement in voice quality assessment. *The Journal of the Acoustical Society of America*, *108*(4), 1867–1876. https://doi.org/10.1121/1.1289362

Kreiman, J., Gerratt, B. R., & Precoda, K. (1990). Listener Experience and Perception of Voice Quality. *Journal of Speech, Language, and Hearing Research*, *33*(1), 103–115. https://doi.org/10.1044/jshr.3301.103

Kreiman, J., Gerratt, B. R., Precoda, K., & Berke, G. S. (1992). Individual Differences in Voice Quality Perception. *Journal of Speech, Language, and Hearing Research*, *35*(3), 512–520. https://doi.org/10.1044/jshr.3503.512

Kuo, C., & Tjaden, K. (2016). Acoustic variation during passage reading for speakers with dysarthria and healthy controls. *Journal of Communication Disorders*, *62*(Generic), 30–44. https://doi.org/10.1016/j.jcomdis.2016.05.003

Lansford, K. L., Liss, J. M., Caviness, J. N., & Utianski, R. L. (2011). A Cognitive-Perceptual Approach to Conceptualizing Speech Intelligibility Deficits and Remediation Practice in Hypokinetic Dysarthria. *Parkinson's Disease*, *2011*, 1–9. https://doi.org/10.4061/2011/150962

Lapointe, L. L., & Horner, J. (1981). Palilalia: A Descriptive Study of Pathological Reiterative Utterances. *Journal of Speech and Hearing Disorders*, *46*(1), 34–38. https://doi.org/10.1044/jshd.4601.34

Lees, A. J., Hardy, J., & Revesz, T. (2009). Parkinson's disease. *The Lancet*, *373*(9680), 2055–2066. https://doi.org/10.1016/S0140-6736(09)60492-X

Liss, J. M., LeGendre, S., & Lotto, A. J. (2010). Discriminating Dysarthria Type From Envelope Modulation Spectra. *Journal of Speech, Language & Hearing Research*, *53*(5), 1246–1255. https://doi.org/10/09-0121)

Liss, J. M., Spitzer, S., Caviness, J. N., Adler, C., & Edwards, B. (1998). Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech. *Journal of the Acoustical Society of America*, *104*(4), 2457–2566. https://doi.org/10.1121/1.423753

Liss, J. M., Spitzer, S. M., Caviness, J. N., & Adler, C. (2002). The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. *The Journal of the Acoustical Society of America*, *112*(6), 3022–3030. https://doi.org/10.1121/1.1515793

Liss, J. M., Spitzer, S. M., Caviness, J. N., Adler, C., & Edwards, B. W. (2000). Lexical boundary error analysis in hypokinetic and ataxic dysarthria. *Journal of the Acoustical Society of America*, *107*(6), 3415. http://search.ebscohost.com/login.aspx?direct=true&db=edb&AN=74416032&site=eds-live&custid=s1234290&authtype=ip,shib

Liss, J. M., White, L., Mattys, S. L., Lansford, K., Lotto, A. J., Spitzer, S. M., & Caviness, J. N. (2009). Quantifying speech rhythm abnormalities in the dysarthrias. *Journal of Speech, Language & Hearing Research*, *52*(5), 1334–1352. http://search.ebscohost.com/login.aspx?direct=true&db=rzh&AN=105328036&site=eds-live&custid=s1234290&authtype=ip,shib

Lowell, S. Y., Kelley, R. T., Awan, S. N., Colton, R. H., & Chan, N. H. (2012). Spectral- and Cepstral-Based Acoustic Features of Dysphonic, Strained Voice Quality. *Annals of Otology, Rhinology & Laryngology*, *121*(8), 539–548. https://doi.org/10.1177/000348941212100808

Lowit-Leuschel, A., & Docherty, G. J. (2001). Prosodic variation across sampling tasks in normal and dysarthric speakers. *Logopedics Phoniatrics Vocology*, *26*(4), 151–164. https://doi.org/10.1080/14015430127772

Luschei, E. S., Ramig, L. O., Baker, K. L., & Smith, M. E. (1999). Discharge Characteristics of Laryngeal Single Motor Units During Phonation in Young and Older Adults and in Persons With Parkinson Disease. *Journal of Neurophysiology*, *81*(5), 2131–2139. https://doi.org/10.1152/jn.1999.81.5.2131

Ma, E. P.-M., & Yiu, E. M.-L. (2006). Multiparametric Evaluation of Dysphonic Severity. *Journal of Voice*, *20*(3), 380–390. https://doi.org/10.1016/j.jvoice.2005.04.007

Mackenzie, C. (2011). Dysarthria in stroke: A narrative review of its description and the outcome of intervention. *International Journal of Speech-Language Pathology*, *13*(2), 125–136.

Marchant, J., McAuliffe, M. J., & Huckabee, M.-L. (2008). Treatment of articulatory impairment in a child with spastic dysarthria associated with cerebral palsy. *Developmental Neurorehabilitation*, *11*(1), 81–90. https://doi.org/10.1080/17518420701622697

Maryn, Y., Corthals, P., Van Cauwenberge, P., Roy, N., & De Bodt, M. (2010). Toward Improved Ecological Validity in the Acoustic Measurement of

Overall Voice Quality: Combining Continuous Speech and Sustained Vowels. *Journal of Voice, 24*(5), 540–555. https://doi.org/10.1016/j.jvoice.2008.12.014

Maryn, Y., & Weenink, D. (2015). Objective Dysphonia Measures in the Program Praat: Smoothed Cepstral Peak Prominence and Acoustic Voice Quality Index. *Journal of Voice, 29*(1), 35–43. https://doi.org/10.1016/j.jvoice.2014.06.015

McLaren, B., Andrews, S. C., Glikmann-Johnston, Y., Mercieca, E.-C., Murray, N. W. G., Loy, C., Bellgrove, M. A., & Stout, J. C. (2021). Feasibility and initial validation of 'HD-Mobile', a smartphone application for remote self-administration of performance-based cognitive measures in Huntington's disease. *Journal of Neurology*, *268*(2), 590–601. https://doi.org/10.1007/s00415-020-10169-y

McRae, P. A., Tjaden, K., & Schoonings, B. (2002). Acoustic and Perceptual Consequences of Articulatory Rate Change in Parkinson Disease. *Journal of Speech, Language, and Hearing Research, 45*(1), 35–50. https://doi.org/10.1044/1092-4388(2002/003)

Mefferd, A. S., & Green, J. R. (2010). Articulatory-to-Acoustic Relations in Response to Speaking Rate and Loudness Manipulations. *Journal of Speech, Language, and Hearing Research*, *53*(5), 1206–1219. https://doi.org/10.1044/1092-4388(2010/09-0083)

Mekyska, J., Smekal, Z., Galaz, Z., Mzourek, Z., Rektorova, I., Faundez-Zanuy, M., & Lopez-De-Ipina, K. (2022). *Perceptual Features as Markers of Parkinson's Disease: The Issue of Clinical Interpretability*. https://doi.org/10.48550/ARXIV.2203.10830

Metter, E. J., & Hanson, W. R. (1986). Clinical and acoustical variability in hypokinetic dysarthria. *Journal of Communication Disorders*, *19*(5), 347–366. https://doi.org/10.1016/0021-9924(86)90026-2

Middag, C. (2012). *Automatic Analysis of Pathological Speech*, *Journal Article*.

Miller, N., Noble, E., Jones, D., & Burn, D. (2006). Life with communication changes in Parkinson's disease. *Age and Ageing*, *35*(3), 235–239.

Möbes, J., Joppich, G., Stiebritz, F., Dengler, R., & Schröder, C. (2008). Emotional speech in Parkinson's disease: Emotional Speech in Parkinson's Disease. *Movement Disorders*, *23*(6), 824–829. https://doi.org/10.1002/mds.21940

Moro-Velazquez, L., & Dehak, N. (2020). A Review of the Use of Prosodic Aspects of Speech for the Automatic Detection and Assessment of Parkinson's Disease. In J. I. Godino-Llorente (Ed.), *Automatic Assessment of Parkinsonian Speech* (Vol. 1295, pp. 42–59). Springer International Publishing. https://doi.org/10.1007/978-3-030-65654-6_3

Moustafa, A. A., Chakravarthy, S., Phillips, J. R., Gupta, A., Keri, S., Polner, B., Frank, M. J., & Jahanshahi, M. (2016). Motor symptoms in Parkinson's disease: A unified framework. *Neuroscience & Biobehavioral Reviews, 68*, 727–740. https://doi.org/10.1016/j.neubiorev.2016.07.010

Moya-Galé, G., Spielman, J., Ramig, L. A., Campanelli, L., & Maryn, Y. (2022). The Acoustic Voice Quality Index (AVQI) in People with Parkinson's Disease Before and After Intensive Voice and Articulation Therapies: Secondary Outcome of a Randomized Controlled Trial. *Journal of Voice*, S0892199722000807. https://doi.org/10.1016/j.jvoice.2022.03.014

Mulligan, M., Carpenter, J., Riddel, J., Delaney, M. K., Badger, G., Krusinski, P., & Tandan, R. (1994). Intelligibility and the Acoustic Characteristics of Speech in Amyotrophic Lateral Sclerosis (ALS). *Journal of Speech, Language, and Hearing Research*, *37*(3), 496–503. https://doi.org/10.1044/jshr.3703.496

Murali, M. (2022). *Using a Podcast Application to Collect High-Quality Speech Data Online for Acoustic Analysis in People With Parkinson's Disease*. SAGE Publications, Ltd. https://doi.org/10.4135/9781529600575

Murdoch, B. E. (1998). *Dysarthria: A physiological approach* (Main Stock 616.8552 MUR). Stanley Thornes. http://search.ebscohost.com/login.aspx?direct=true&db=cat02513a&AN=qmu.b1065589&site=eds-live&custid=s1234290&authtype=ip,shib

Murdoch, B. E. (2009). *Acquired Speech and Language Disorders*. John Wiley & Sons, Incorporated. http://ebookcentral.proquest.com/lib/qmu/detail.action?docID=698388

Murton, O., Hillman, R., & Mehta, D. (2020). Cepstral Peak Prominence Values for Clinical Voice Evaluation. *American Journal of Speech-Language Pathology, 29*(3), 1596–1607. https://doi.org/10.1044/2020_AJSLP-20-00001

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.e9-390.e15. https://doi.org/10.1016/j.jvoice.2014.07.022

Neel, A. T. (2008). Vowel Space Characteristics and Vowel Identification Accuracy. *Journal of Speech, Language, and Hearing Research*, *51*(3), 574–585. https://doi.org/10.1044/1092-4388(2008/041)

Neel, A. T. (2009). Effects of Loud and Amplified Speech on Sentence and Word Intelligibility in Parkinson Disease. *Journal of Speech, Language, and Hearing Research*, *52*(4), 1021–1033. https://doi.org/10.1044/1092-4388(2008/08-0119)

Novotný, M., Rusz, J., Čmejla, R., & Růžička, E. (2014). Automatic Evaluation of Articulatory Disorders in Parkinson's Disease. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *22*(9), 1366–1378. https://doi.org/10.1109/TASLP.2014.2329734

Oates, J. (2009). Auditory-Perceptual Evaluation of Disordered Voice Quality. *Folia Phoniatrica et Logopaedica*, *61*(1), 49–56. https://doi.org/10.1159/000200768

Ohde, R. N., Haley, K. L., Vorperian, H. K., & McMahon, C. W. (1995). A developmental study of the perception of onset spectra for stop consonants in different vowel environments. *The Journal of the Acoustical Society of America*, *97*(6), 3800–3812. https://doi.org/10.1121/1.412395

Ohde, R. N., & Stevens, K. N. (1983). Effect of burst amplitude on the perception of stop consonant place of articulation. *The Journal of the Acoustical Society of America*, *74*(3), 706–714. https://doi.org/10.1121/1.389856

Ortega, J., Cassinello, N., Dorcaratto, D., & Leopaldi, E. (2009). Computerized acoustic voice analysis and subjective scaled evaluation of the voice can avoid the need for laryngoscopy after thyroid surgery. *Surgery*, *145*(3), 265–271. https://doi.org/10.1016/j.surg.2008.11.002

Oxenham, A. J., Boucher, J. E., & Kreft, H. A. (2017). Speech intelligibility is best predicted by intensity, not cochlea-scaled entropy. *The Journal of the Acoustical Society of America*, *142*(3), EL264–EL269. https://doi.org/10.1121/1.5002149

Palmer, R., Enderby, P., & Hawley, M. (2007). Addressing the needs of speakers with longstanding dysarthria: Computerized and traditional therapy compared. *International Journal of Language & Communication Disorders*, *42*(s1), 61–79. https://doi.org/10.1080/13682820601173296

Park, H. K., Yoo, J. Y., Kwon, M., Lee, J.-H., Lee, S. J., Kim, S. R., Kim, M. J., Lee, M. C., Lee, S. M., & Chung, S. J. (2014). Gait freezing and speech disturbance in Parkinson's disease. *Neurological Sciences*, *35*(3), 357–363. https://doi.org/10.1007/s10072-013-1519-1

Patel, R. R., Awan, S. N., Barkmeier-Kraemer, J., Courey, M., Deliyski, D., Eadie, T., Paul, D., Švec, J. G., & Hillman, R. (2018). Recommended Protocols for Instrumental Assessment of Voice: American Speech-Language-Hearing Association Expert Panel to Develop a Protocol for Instrumental Assessment of Vocal Function. *American Journal of Speech-Language Pathology*, *27*(3), 887–905. https://doi.org/10.1044/2018_AJSLP-17-0009

Pawlukowska, W., Gołąb-Janowska, M., Safranow, K., Rotter, I., Amernik, K., Honczarenko, K., & Nowacki, P. (2015). Articulation disorders and duration, severity and l-dopa dosage in idiopathic Parkinson's disease. *Neurologia i Neurochirurgia Polska*, *49*(5), 302–306. https://doi.org/10.1016/j.pjnns.2015.07.002

Pennington, L., Smallman, C., & Farrier, F. (2006). Intensive dysarthria therapy for older children with cerebral palsy: Findings from six cases. *Child Language Teaching and Therapy*, *22*(3), 255–273. https://doi.org/10.1191/0265659006ct307xx

Perez, K. S., Ramig, L. O., Smith, M. E., & Dromey, C. (1996). The Parkinson larynx: Tremor and videostroboscopic findings. *Journal of Voice*, *10*(4), 354–361. https://doi.org/10.1016/S0892-1997(96)80027-0

Peterson, E. A., Roy, N., Awan, S. N., Merrill, R. M., Banks, R., & Tanner, K. (2013). Toward Validation of the Cepstral Spectral Index of Dysphonia (CSID) as an Objective Treatment Outcomes Measure. *Journal of Voice*, *27*(4), 401–410. https://doi.org/10.1016/j.jvoice.2013.04.002

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking Clearly for the Hard of Hearing I: Intelligibility Differences between Clear and Conversational Speech. *Journal of Speech, Language, and Hearing Research*, *28*(1), 96–103. https://doi.org/10.1044/jshr.2801.96

R Core Team. (2019). R: A Language and Environment for Statistical Computing. *Vienna, Austria*. https://www.R-project.org/

Ramig, L. A. (1983). Effects of physiological aging on speaking and reading rates. *Journal of Communication Disorders*, *16*(3), 217–226. https://doi.org/10.1016/0021-9924(83)90035-7

Ramig, L. O., Fox, C., & Sapir, S. (2008). Speech treatment for Parkinson's disease. *Expert Review of Neurotherapeutics*, *8*(2), 297–309. https://doi.org/10.1586/14737175.8.2.297

Ravizza, S. M. (2003). Dissociating the performance of cortical and subcortical patients on phonemic tasks. *Brain and Cognition*, *53*(2), 301–310. https://doi.org/10.1016/S0278-2626(03)00131-3

Robbins, J. A., Logemann, J. A., & Kirshner, H. S. (1986). Swallowing and speech production in Parkinson's disease. *Annals of Neurology*, *19*(3), 283–287. https://doi.org/10.1002/ana.410190310

Robertson, S. (2001). The Efficacy of Oro-Facial and Articulation Exercises in Dysarthria following Stroke. *International Journal of Language &*

Communication Disorders, *36*(s1), 292–297. https://doi.org/10.3109/13682820109177900

Rosen, K. M., Kent, R. D., Delaney, A. L., & Duffy, J. R. (2006). Parametric quantitative acoustic analysis of conversation produced by speakers with dysarthria and healthy speakers. *Journal of Speech, Language & Hearing Research*, *49*(2), 395–411. http://search.ebscohost.com/login.aspx?direct=true&db=rzh&AN=106314846&site=eds-live&custid=s1234290&authtype=ip,shib

Rosen, K. M., Kent, R. D., & Duffy, J. R. (2005). Task-Based Profile of Vocal Intensity Decline in Parkinson's Disease. *Folia Phoniatrica et Logopaedica*, *57*(1), 28–37. https://doi.org/10.1159/000081959

Rusz, J., Bonnet, C., Klempíř, J., Tykalová, T., Baborová, E., Novotný, M., Rulseh, A., & Růžička, E. (2015). Speech disorders reflect differing pathophysiology in Parkinson's disease, progressive supranuclear palsy and multiple system atrophy. *Journal of Neurology*, *262*(4), 992–1001. https://doi.org/10.1007/s00415-015-7671-1

Rusz, J., Cmejla, R., Ruzickova, H., Klempir, J., Majerova, V., Picmausova, J., Roth, J., & Růžička, E. (2011). Acoustic analysis of voice and speech characteristics in early untreated Parkinson's disease. *Models and Analysis of Vocal Emissions for Biomedical Applications - 7th International Workshop, MAVEBA 2011*, 181–184.

Rusz, J., 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. *Journal of the Acoustical Society of America*, *129*(1), 350–367. https://doi.org/10.1121/1.3514381

Rusz, J., Tykalová, T., Novotný, M., Růžička, E., & Dusek, P. (2021). Distinct patterns of speech disorder in early-onset and late-onset de-novo Parkinson's disease. *Npj Parkinson's Disease*, *7*(1), 98. https://doi.org/10.1038/s41531-021-00243-1

Rusz, J., Tykalova, T., Ramig, L. O., & Tripoliti, E. (2021). Guidelines for Speech Recording and Acoustic Analyses in Dysarthrias of Movement Disorders. *Movement Disorders*, *36*(4), 803–814. https://doi.org/10.1002/mds.28465

Sachdeva, K., & Shrivastava, T. (2018). Dysphonia and its Correlation with Acoustic Voice Parameters. *International Journal of Phonosurgery & Laryngology*, *8*(1), 6–12. https://doi.org/10.5005/jp-journals-10023-1151

Sapir, S., Ramig, L. O., Spielman, J. L., & Fox, C. (2010). Formant Centralization Ratio: A Proposal for a New Acoustic Measure of Dysarthric Speech. *Journal of Speech, Language, and Hearing Research*, *53*(1), 114–125. https://doi.org/10.1044/1092-4388(2009/08-0184)

Sapir, S., Spielman, J. L., Ramig, L. O., Story, B. H., & Fox, C. (2007). Effects of Intensive Voice Treatment (the Lee Silverman Voice Treatment [LSVT]) on Vowel Articulation in Dysarthric Individuals With Idiopathic Parkinson

Disease: Acoustic and Perceptual Findings. *Journal of Speech, Language, and Hearing Research*, *50*(4), 899–912. https://doi.org/10.1044/1092-4388(2007/064)

Schulman, R. (1989). Articulatory dynamics of loud and normal speech. *The Journal of the Acoustical Society of America*, *85*(1), 295–312. https://doi.org/10.1121/1.397737

Schultz, B. G., & Vogel, A. P. (2022). A Tutorial Review on Clinical Acoustic Markers in Speech Science. *Journal of Speech, Language, and Hearing Research*, *65*(9), 3239–3263. https://doi.org/10.1044/2022_JSLHR-21-00647

Selouani, S. -A., Dahmani, H., Amami, R., & Hamam, H. (2012). Using speech rhythm knowledge to improve dysarthric speech recognition. *International Journal of Speech Technology*, *15*(1), 57.

Shahed, J., & Jankovic, J. (2001). Re-emergence of childhood stuttering in Parkinson's disease: A hypothesis. *Movement Disorders*, *16*(1), 114–118. https://doi.org/10.1002/1531-8257(200101)16:1<114::AID-MDS1004>3.0.CO;2-2

Sheard, C., Adams, R. D., & Davis, P. J. (1991). Reliability and Agreement of Ratings of Ataxic Dysarthric Speech Samples With Varying Intelligibility. *Journal of Speech, Language, and Hearing Research*, *34*(2), 285–293. https://doi.org/10.1044/jshr.3402.285

Shewan, C. M., & Henderson, V. L. (1988). Analysis of spontaneous language in the older normal population. *Journal of Communication Disorders*, *21*(2), 139–154. https://doi.org/10.1016/0021-9924(88)90002-0

Shih, M. C., Franco de Andrade, L. A., Amaro, E., Felicio, A. C., Ferraz, H. B., Wagner, J., Hoexter, M. Q., Lin, L. F., Fu, Y. K., Mari, J. J., Tufik, S., & Bressan, R. A. (2007). Higher nigrostriatal dopamine neuron loss in early than late onset Parkinson's disease?—A [99mTc]-TRODAT-1 SPECT study. *Movement Disorders*, *22*(6), 863–866. https://doi.org/10.1002/mds.21315

Shrivastav, R. (2003). The use of an auditory model in predicting perceptual ratings of breathy voice quality. *Journal of Voice*, *17*(4), 502–512. https://doi.org/10.1067/S0892-1997(03)00077-8

Shrivastav, R., Sapienza, C. M., & Nandur, V. (2005). Application of Psychometric Theory to the Measurement of Voice Quality Using Rating Scales. *Journal of Speech, Language, and Hearing Research*, *48*(2), 323–335. https://doi.org/10.1044/1092-4388(2005/022)

Siegert, R. J., Harper, D. N., Cameron, F. B., & Abernethy, D. (2002). Self-Initiated Versus Externally Cued Reaction Times in Parkinson's Disease. *Journal of Clinical and Experimental Neuropsychology*, *24*(2), 146–153. https://doi.org/10.1076/jcen.24.2.146.991

Šimek, M., & Rusz, J. (2021). Validation of cepstral peak prominence in assessing early voice changes of Parkinson's disease: Effect of speaking task and ambient noise. *The Journal of the Acoustical Society of America*, *150*(6), 4522–4533. https://doi.org/10.1121/10.0009063

Skodda, S., Flasskamp, A., & Schlegel, U. (2011). Instability of Syllable Repetition as a Marker of Disease Progression in Parkinson's Disease: A Longitudinal Study. *MOVEMENT DISORDERS*, *26*(1), 59–64. https://doi.org/10.1002/mds.23382

Skodda, S., Groenheit, W., & Schlegel, U. (2012). Impairment of Vowel Articulation as a Possible Marker of Disease Progression in Parkinson's Disease. *PLOS ONE*, *7*(2). https://doi.org/10.1371/journal.pone.0032132

Skodda, S., Grönheit, W., Mancinelli, N., & Schlegel, U. (2013). Progression of Voice and Speech Impairment in the Course of Parkinson's Disease: A Longitudinal Study. *Parkinson's Disease*, *2013*, 1–8. https://doi.org/10.1155/2013/389195

Skodda, S., Rinsche, H., & Schlegel, U. (2009). Progression of dysprosody in Parkinson's disease over time-A longitudinal study. *Movement Disorders*, *24*(5), 716. http://search.ebscohost.com/login.aspx?direct=true&db=edb&AN=64243544&site=eds-live&custid=s1234290&authtype=ip,shib

Skodda, S., & Schlegel, U. (2008). Speech rate and rhythm in Parkinson's disease: Speech Rate and Rhythm in Parkinson's Disease. *Movement Disorders*, *23*(7), 985–992. https://doi.org/10.1002/mds.21996

Skodda, S., Visser, W., & Schlegel, U. (2010). Short- and long-term dopaminergic effects on dysarthria in early Parkinson's disease. *Journal of Neural Transmission*, *117*(2), 197–205. https://doi.org/10.1007/s00702-009-0351-5

Skodda, S., Visser, W., & Schlegel, U. (2011a). Gender-Related Patterns of Dysprosody in Parkinson Disease and Correlation Between Speech Variables and Motor Symptoms. *Journal of Voice*, *25*(1), 76–82. https://doi.org/10.1016/j.jvoice.2009.07.005

Skodda, S., Visser, W., & Schlegel, U. (2011b). Vowel Articulation in Parkinson's Disease. *Journal of Voice*, *25*(4), 467–472. https://doi.org/10.1016/j.jvoice.2010.01.009

Smith, B. L., Wasowicz, J., & Preston, J. (1987). Temporal Characteristics of the Speech of Normal Elderly Adults. *Journal of Speech, Language, and Hearing Research*, *30*(4), 522–529. https://doi.org/10.1044/jshr.3004.522

Solomon, N. P. (2000). Changes in Normal Speech After Fatiguing the Tongue. *Journal of Speech, Language, and Hearing Research*, *43*(6), 1416–1428. https://doi.org/10.1044/jslhr.4306.1416

Solomon, N. P., & Hixon, T. J. (1993). Speech Breathing in Parkinson's Disease. *Journal of Speech, Language, and Hearing Research*, *36*(2), 294–310. https://doi.org/10.1044/jshr.3602.294

Squadcast (4.5). (2021). https://squadcast.fm/

State of Ageing. (2022). ageing-better.uk.org

Stipancic, K. L. & Kris Tjaden. (2022). Minimally Detectable Change of Speech Intelligibility in Speakers With Multiple Sclerosis and Parkinson's Disease. *Journal of Speech Language and Hearing Research*. https://doi.org/10.1044/2022_jslhr-21-00648

Teixeira, J. P., & Fernandes, P. O. (2014). Jitter, Shimmer and HNR Classification within Gender, Tones and Vowels in Healthy Voices. *Procedia Technology*, *16*, 1228–1237. https://doi.org/10.1016/j.protcy.2014.10.138

Teixeira, J. P., & Gonçalves, A. (2016). Algorithm for Jitter and Shimmer Measurement in Pathologic Voices. *Procedia Computer Science*, *100*, 271–279. https://doi.org/10.1016/j.procs.2016.09.155

Teixeira, J. P., Oliveira, C., & Lopes, C. (2013). Vocal Acoustic Analysis – Jitter, Shimmer and HNR Parameters. *Procedia Technology*, *9*, 1112–1122. https://doi.org/10.1016/j.protcy.2013.12.124

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., Parkinson, J., Secker, J., & Stewart-Brown, S. (2007). The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): Development and UK validation. *Health and Quality of Life Outcomes*, *5*(1), 63. https://doi.org/10.1186/1477-7525-5-63

Theodoros, D. G., & Ramig, L. O. (2011). *Communication and Swallowing in Parkinson Disease: Design and Implementation of the Patient Experience*.

Plural Publishing, Inc. http://ebookcentral.proquest.com/lib/qmu/detail.action?docID=1891333

Tjaden, K. (2008). Speech and Swallowing in Parkinson's Disease. *Topics in Geriatric Rehabilitation*, *24*(2), 115–126. https://doi.org/10.1097/01.TGR.0000318899.87690.44

Tjaden, K., & Wilding, G. (2011). The Impact of Rate Reduction and Increased Loudness on Fundamental Frequency Characteristics in Dysarthria. *Folia Phoniatrica et Logopaedica*, *63*(4), 178–186. https://doi.org/10.1159/000316315

Tjaden, K., & Wilding, G. E. (2004). Rate and Loudness Manipulations in Dysarthria: Acoustic and Perceptual Findings. *Journal of Speech, Language, and Hearing Research*, *47*(4), 766–783. https://doi.org/10.1044/1092-4388(2004/058)

Toth, C., Rajput, M., & Rajput, A. H. (2004). Anomalies of asymmetry of clinical signs in parkinsonism. *Movement Disorders*, *19*(2), 151–157. https://doi.org/10.1002/mds.10685

Tsanas, A., Little, M. A., McSharry, P. E., & Ramig, L. O. (2011). Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. *Journal of The Royal Society Interface*, *8*(59), 842–855. https://doi.org/10.1098/rsif.2010.0456

Tykalova, T., Rusz, J., Klempir, J., Cmejla, R., & Ruzicka, E. (2017). Distinct patterns of imprecise consonant articulation among Parkinson's disease, progressive supranuclear palsy and multiple system atrophy. *Brain & Language, Journal Article*, 1–9. https://doi.org/10.1016/j.bandl.2016.11.005

Tysnes, O.-B., & Storstein, A. (2017). Epidemiology of Parkinson's disease. *Journal of Neural Transmission*, *124*(8), 901–905. https://doi.org/10.1007/s00702-017-1686-y

Uloza, V., Padervinskis, E., Vegiene, A., Pribuisiene, R., Saferis, V., Vaiciukynas, E., Gelzinis, A., & Verikas, A. (2015). Exploring the feasibility of smart phone microphone for measurement of acoustic voice parameters and voice pathology screening. *European Archives of Oto-Rhino-Laryngology*, *272*(11), 3391–3399. https://doi.org/10.1007/s00405-015-3708-4

van Brenk, F., Stipancic, K. L., Kain, A., & Kris Tjaden. (2022). Intelligibility Across a Reading Passage: The Effect of Dysarthria and Cued Speaking Styles. *American Journal of Speech-Language Pathology*. https://doi.org/10.1044/2021_ajslp-21-00151

Van der Graaff, M., Kuiper, T., Zwinderman, A., Van de Warrenburg, B., Poels, P., Offeringa, A., Van der Kooi, A., Speelman, H., & De Visser, M. (2009). Clinical Identification of Dysarthria Types among Neurologists, Residents in Neurology and Speech Therapists. *European Neurology*, *61*(5), 295–300. https://doi.org/10.1159/000206855

Van Nuffelen, G., De Bodt, M., Vanderwegen, J., Van de Heyning, P., & Wuyts, F. (2010). Effect of Rate Control on Speech Production and Intelligibility in Dysarthria. *Folia Phoniatrica et Logopaedica*, *62*(3), 110–119. https://doi.org/10.1159/000287209

van Son, R., Middag, C., & Demuynck, K. (2018). *Vowel Space as a Tool to Evaluate Articulation Problems*. 357–361.

Vayra, M., & Fowler, C. A. (1992). Declination of Supralaryngeal Gestures in Spoken Italian. *Phonetica*, *49*(1), 48–60. https://doi.org/10.1159/000261902

Vizza, P., Tradigo, G., Mirarchi, D., Bossio, R. B., Lombardo, N., Arabia, G., Quattrone, A., & Veltri, P. (2019). Methodologies of speech analysis for neurodegenerative diseases evaluation. *International Journal of Medical Informatics*, *122*(Generic), 45–54. https://doi.org/10.1016/j.ijmedinf.2018.11.008

Webb, A. L., Carding, P. N., Deary, I. J., MacKenzie, K., Steen, N., & Wilson, J. A. (2004). The reliability of three perceptual evaluation scales for dysphonia. *European Archives of Oto-Rhino-Laryngology*, *261*(8). https://doi.org/10.1007/s00405-003-0707-7

Weismer, G. (1984). *Acoustic descriptions of dysarthric speech: Perceptual correlates and physiological inferences*. *5*(04), 293–314.

Weismer, G. (2006). Philosophy of research in motor speech disorders. *Clinical Linguistics & Phonetics*, *20*(5), 315–349. https://doi.org/10.1080/02699200400024806

Weismer, G., Jeng, J.-Y., Laures, J. S., Kent, R. D., & Kent, J. F. (2001). Acoustic and Intelligibility Characteristics of Sentence Production in Neurogenic Speech Disorders. *Folia Phoniatrica et Logopaedica*, *53*(1), 1–18. https://doi.org/10.1159/000052649

Weismer, G., Martin, R., Kent, R. D., & Kent, J. F. (1992). Formant trajectory characteristics of males with amyotrophic lateral sclerosis. *The Journal of the Acoustical Society of America*, *91*(2), 1085–1098. https://doi.org/10.1121/1.402635

Wong, D. F., Wagner, H. N., Dannals, R. F., Links, J. M., Frost, J. J., Ravert, H. T., Wilson, A. A., Rosenbaum, A. E., Gjedde, A., Douglass, K. H., Petronis, J. D., Folstein, M. F., Toung, J. K. T., Bums, H. D., & Kuhar, M. J. (1984). Effects of Age on Dopamine and Serotonin Receptors Measured by Positron Tomography in the Living Human Brain. *Science*, *226*(4681), 1393–1396. https://doi.org/10.1126/science.6334363

Wuyts, F. L., De Bodt, M. S., & Van de Heyning, P. H. (1999). Is the reliability of a visual analog scale higher than an ordinal scale? An experiment with the GRBAS scale for the perceptual evaluation of dysphonia. *Journal of Voice*, *13*(4), 508–517. https://doi.org/10.1016/S0892-1997(99)80006-X

Xia, R., & Mao, Z.-H. (2012). Progression of motor symptoms in Parkinson's disease. *Neuroscience Bulletin*, *28*(1), 39–48. https://doi.org/10.1007/s12264-012-1050-z

Yiu, E. M.-L. (1999). Limitations of perturbation measures in clinical acoustic voice analysis. *Asia Pacific Journal of Speech, Language and Hearing*, *4*(3), 155–166. https://doi.org/10.1179/136132899807557475

Yorkston, K. M. (1996). Treatment efficacy: Dysarthria. *Journal of Speech, Language, and Hearing Research*, *39*(5), S46–S57.

Yorkston, K. M., & Beukelman, D. R. (1978). A comparison of techniques for measuring intelligibility of dysarthric speech. *Journal of Communication Disorders*, *11*(6), 499–512. https://doi.org/10.1016/0021-9924(78)90024-2

Yorkston, K. M., & Beukelman, D. R. (1980). An Analysis of Connected Speech Samples of Aphasic and Normal Speakers. *Journal of Speech and Hearing Disorders*, *45*(1), 27–36. https://doi.org/10.1044/jshd.4501.27

Yorkston, K. M., Spencer, K., Duffy, J., Beukelman, D., Golper, L. A., Miller, R., Strand, E., & Sullivan, M. (2001). Evidence-Based Practice Guidelines for Dysarthria: Management of Velopharyngeal Function—Academy of Neurologic Communication Disorders and Sciences: Writing Committee for Practice Guidelines in Dysarthria: *Journal of Medical Speech-Language Pathology*, *9*, 257–274.

Yorkston, K. M., Strand, E. A., & Kennedy, M. R. T. (1996). Comprehensibility of Dysarthric Speech: Implications for Assessment and Treatment Planning. *American Journal of Speech-Language Pathology*, *5*(1), 55–66. https://doi.org/10.1044/1058-0360.0501.55

Zwetsch, I. C., Fagundes, R. D. R., Russomano, T., & Scolari, D. (2006). *Digital signal processing in the differential diagnosis of benign larynx diseases [Abstract in English]*.

Zyski, B. J., & Weisiger, B. E. (1987). Identification of dysarthria types based on perceptual analysis. *Journal of Communication Disorders*, *20*(5), 367–378. https://doi.org/10.1016/0021-9924(87)90025-6

9. Appendices

9.1. Appendix A

Queen Margaret University
EDINBURGH

Title of Study: Acoustic markers in speakers with Parkinson's Disease

You are being invited to take part in a research study titled Acoustic Markers in speakers with Parkinson's Disease. Before you decide to participate, it is important for you to understand why this research is being carried out and what it will involve. Please take your time to read the following information sheet and feel free to ask any questions if there is anything that is not explained clearly. If you would like more information, please contact the primary researcher (contact details are provided at the end of this information sheet).

What is the purpose of this study?

I am a PhD candidate from the School of Health Sciences at Queen Margaret University in Edinburgh. I am undertaking a research project for my PhD dissertation.

The aim of this study is to identify distinct features in the speech of speakers with Parkinson's disease and see if these features can be followed over the course of one year. This will help assess how speech in speakers changes over a year alongside Parkinson's disease progression. Although research has been conducted in this area, there is not enough information on which features are constant and whether they can reflect disease progression over time in speech. Individuals with Parkinson's disease, who often face communicative difficulties as a secondary symptom may find improvements or decline in their speech based on changes in their physiological symptoms, and how medication may impact speech. This project could help you understand how your speech may change over the course of a year, alongside disease progression, and will help speech language therapists during assessment and therapy.

Your participation in this study could help find any speech features that are robust enough to reflect disease progression over time and provide further insight into speech production in Parkinson's disease. The project will also record unimpaired speech in order to compare it against the speakers with Parkinson's disease. Therefore, I will also be looking for volunteers for a control group.

Can you volunteer for this study?

<u>If you would like to volunteer:</u>

1. You will need to have been diagnosed with Parkinson's disease, however it does not matter when you were diagnosed.
2. You are at least 35 years old and are a fluent speaker of English.
3. You have some communicative or speech problems. You do not need to have been formally diagnosed with a speech or voice disorder and may or may not be in speech therapy for it.
4. You do not have any other cognitive or mental health condition.
5. You are willing to record your speech now and after six months.
6. You have access to a smart phone (android only) or a laptop to join a live call for recording your data.
7. You are willing to provide your address in order to be sent a copy of the speech data you will be recording. This information will be confidential and used only as a means to send you relevant documentation. The details will be removed from our records once the study has concluded.

If you do decide to participate, it would help the purposes of the research if you are able to provide any diagnostic test results you may have undergone that are related to your initial Parkinson's disease diagnosis, your disease progression, and any information about your speech. This would be your Hoehn and Yahr Scale scores, if available to you, or related test results. This information is strictly to be used for the research and to record your disease progression and will be dealt with confidentiality. The data will be anonymised, and you will not be identified from it.

<u>If you would like to volunteer for the control group:</u>

1. You are at least 35 years old and are a fluent speaker of English.
2. You have not been diagnosed with any cognitive or mental health condition.

3. You have not been diagnosed with a speech/ voice/ or learning disorder.
4. You are willing to record your speech now and after six months.
5. You have access to a smart phone (android only) or a laptop to join a live call for recording your data.
6. You are willing to provide your address in order to be sent a copy of the speech data you will be recording. This information will be confidential and used only as a means to send you relevant documentation. The details will be removed from our records once the study has concluded.

Participation in this study is entirely voluntary and you are free to decline participation or to withdraw from the study at any time. You do not need to give any reasons if you decide to leave the study.

What will happen if you decide to participate?

This study will be done remotely and can be done at your convenience. If you agree to participate in the study, you will be asked to complete a mental wellbeing scale before starting the study. It is a simple self-reporting scale and will only take a few minutes to complete.

You will be sent the speech data that you will be reading and will be recorded data collection. You will be sent an electronic link to join a private video call with the primary researcher, at your convenience. During this call, you will be briefed on the study and how you will set up before recording takes place. You will also have the time to have any questions you have answered. Once you are ready, the researcher will begin recording the call for data collection. Only your audio will be recorded so no video or audio aside from the speech data will be stored.

The speech data includes a number of words, sentences, a passage, and free speech (based on a prompt given).

You will also be asked to record your speech again after six months, using the same process you will go through this time.

Are there any risks/ disadvantages and any benefits?

There is a risk that you will experience fatigue or strain if the recording session is too long. The researcher has tried to limit speech data to only what is necessary for the study in order to avoid increasing discomfort or strain. You will be allowed to begin recordings when you are ready and comfortable. The researcher is not aware of any other risks associated with this study. The whole procedure should take no longer than 20 minutes, but the length of the session may vary per individual as each recording needs to be of high quality and will require repetitions.

There are no immediate benefits to participating in this study. However, the findings of the project could provide more information on communicative problems in Parkinson’s disease and if acoustic markers are found, could help early intervention before speech disorders occur and help follow changes in speech during speech therapy, alongside disease progression. It would also help future research that could try to use these markers an automatic detection system for a cost-effective diagnosis.

Will your information be kept confidential?

All of your personal information will be treated in accordance with the terms of the UK Data Protection Act 2018 and the General Data Protection Regulation (GDPR). Appropriate security measures including anonymisation will be put in place to protect your data at all times.

All data will be anonymised as much as possible, but you may be identifiable from recordings of your voice. Your name will be replaced with a participant number, and it will not be possible for you to be identified in any reporting of the data gathered. Any other personal information such as your age, sex, any diagnostic information you provide, or other related information will be stored on a secure, password-protected server that will only be accessed by the research team during the study. Any information shared outside of the research team will be anonymised and you will not be identified from it.

Your personal information will only be retained for as long as is necessary, during the course of the study. Thereafter, personal information will be destroyed. The data gathered from the study and your consent form will be stored on a secure server indefinitely to aid future research but will be stored separately not attached to any indefinable information after the study has concluded.

The findings may also be written and published in medical/scientific journals to aid other clinicians and patients elsewhere. Neither you nor your data will be identifiable in these publications.

You have the right to withdraw your consent to us processing your personal data at any time. In order to do so, please contact the primary researcher: Mridhula Murali; MMurali@qmu.ac.uk. Please note that your data may be used in the production of formal research outputs before you withdraw consent, therefore it is advisable to contact us as soon as possible if you wish to withdraw your consent. We will destroy your identifiable data upon request, where possible, however in some situations we will require to use the data collected up until your withdrawal of consent. If you have any questions relating to the processing of your data which are not resolved by contacting Mridhula Murali please contact the QMU Data Protection Officer Lorraine Kerr -LKerr2@qmu.ac.uk

If you would like to contact an independent person, who knows about this project but is not involved in it, you are welcome to contact Dr. Sara Wood. Her contact details are given below.

If you have read and understood this information sheet, any questions you had have been answered, and you would like to be a participant in the study, please now see the consent form.

Contact details of the research team

Name of researcher: Mridhula Murali, Primary Researcher

Address: PhD Candidate, CASL Research centre,
Speech and Hearing Sciences Division,

School of Health Sciences,
Queen Margaret University, Edinburgh
Queen Margaret University Drive
Musselburgh
East Lothian EH21 6UU
Email / Telephone: MMurali@qmu.ac.uk / 07716516201

Name of independent academic: Dr. Sara Wood
Address: Reader, Speech and Hearing Sciences Division,
School of Health Sciences,
Queen Margaret University, Edinburgh
Queen Margaret University Drive
Musselburgh
East Lothian EH21 6UU
Email / Telephone: swood@qmu.ac.uk / 0131 474 0000

9.2. Appendix B

Queen Margaret University
EDINBURGH

Consent Form

Participant Identification Number for this project:	

Title of Study: Acoustic markers in speakers with Parkinson's Disease
Name of Primary Researcher: Mridhula Murali

	Please initial box
I have read the information sheet for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	
I understand that my participation is voluntary and that I can withdraw at any time without giving any reason.	
I agree to be audio recorded for this project and have my recordings stored over a secure server.	
I give permission to individuals from the research team and a qualified speech language therapist to access my speech recordings, and any diagnostic records I share.	
I understand that the data collected in this study and my consent form will be stored securely for an indefinite time to aid future research. I am aware that my consent form will not be linked in any way to my data after the end of this study.	
I agree to take part in the above study.	

Name of Participant	
Email and Phone Number	
Date	

Signature	

Name of Researcher	Mridhula Murali
Date	
Signature	

Contact details of the researcher

Name of researcher: Mridhula Murali
Address: PhD Candidate, CASL Research Centre,
Speech and Hearing Sciences Division,
School of Health Sciences,
Queen Margaret University, Edinburgh
Queen Margaret University Drive
Musselburgh
East Lothian EH21 6UU
Email / Telephone: MMurali@qmu.ac.uk / 07716516201

9.3. Appendix C

Demographic Information Sheet

Please fill out the following information below. It will ensure you fit the criteria of this study and will allow the primary researcher to contact you and share information regarding the study. The information will not be shared with anyone outside of the primary research team. At the end of the study, your name and contact information will be removed and will be made anonymous. Information such as your age, information regarding your diagnosis, and speech data will be retained and anonymously linked to help future research.

1. Name:

2. Date of birth (dd/mm/yyyy):

3. Contact details:

Your email will be your primary contact, but if you cannot be reached then will you be called by phone. Your home address is <u>only</u> used to send you a sheet with the speech data and a hard copy of your consent form, if you would prefer – It will be not used for any other reason and will be removed from records at the end of the study. Please provide an email you are comfortable with and being sent information about the study. Do <u>not</u> share this information with anyone unless they are helping you for this study.

4. Email:

5. Phone Number:

6. Home Address:

7. Do you have access to a laptop (Windows/Mac are both okay) or Android phone (iPhone/ iPad is not compatible with the recording programme)?

8. Are you a native speaker of English?

9. What is your Nationality?

10. Have you ever been diagnosed with any speech/ voice disorder? If so, please state the type of speech disorder and when you were diagnosed.

11. Have you ever been diagnosed with any cognitive or mental health condition such as dementia or depression? If so, please state the type of condition and when you were diagnosed.

12. If you have been diagnosed with Parkinson's Disease, please answer the following questions, otherwise skip these questions.
    a. What year were you diagnosed with Parkinson's disease?

    b. Have you ever been on dopaminergic medication? If so, please state when.

    c. Are you currently on dopaminergic medication?

    d. Do you have any communicative problems you have noticed since your diagnosis? If so, please describe them as best you can.

    e. Have you been diagnosed with any speech/ voice disorder after your diagnosis of Parkinson's disease? If so, please state the type of speech disorder and when you were diagnosed.

    f. Have you been diagnosed with any cognitive/ mental health condition such as dementia or depression since your diagnosis of Parkinson's disease? If so, please state the type of mental health and when you were diagnosed.

9.4. Appendix D

Warwick-Edinburgh Mental Wellbeing Scale

Warwick Edinburgh Mental Wellbeing Scale (WEMWBS)

Below are some statements about feelings and thoughts.

Please circle the box that best describes your experience of each over the last 2 weeks.

	None of the Time	Rarely	Some of the Time	Often	All of the Time
I've been feeling optimistic about the future	1	2	3	4	5
I've been feeling useful	1	2	3	4	5
I've been feeling relaxed	1	2	3	4	5
I've been feeling interested in other people	1	2	3	4	5
I've had energy to spare	1	2	3	4	5
I've been dealing with problems well	1	2	3	4	5
I've been thinking clearly	1	2	3	4	5
I've been feeling good about myself	1	2	3	4	5
I've been feeling close to other people	1	2	3	4	5
I've been feeling confident	1	2	3	4	5
I've been able to make up my own mind about things	1	2	3	4	5
I've been feeling loved	1	2	3	4	5
I've been interested in new things	1	2	3	4	5
I've been feeling cheerful	1	2	3	4	5

9.5. Appendix E

Speech data to be recorded

Minimal Pairs

1. Tease – Cheese
2. Sat – Chat
3. Head – Hedge
4. Doll – Toll
5. Joke – Choke
6. Key – Tea
7. Grip – Drip
8. Coat – Goat
9. Bead – Beat
10. Peck – Peg
11. Peas – Bees
12. High – Tie
13. Hairy – Fairy
14. Heap – Sheep
15. Seat – Heat
16. Vest – West
17. Cell – Fell
18. Sign – Fine
19. File – Dial
20. Sip – Ship

Grandfather Passage

You wished to know all about my grandfather. Well, he is nearly ninety-three years old. He dresses himself in an ancient black frock coat, usually minus several buttons; yet he still thinks as swiftly as ever. A long, flowing beard clings to his chin, giving those who observe him a pronounced feeling of the utmost respect. When he speaks his voice is just a bit cracked and quivers a trifle. Twice each day he plays skilfully and with zest upon our small organ. Except in the winter when the ooze or snow or ice prevents, he slowly takes a short walk in the open air each day. We have often urged him to walk more and smoke less, but he always answers, “Banana Oil!” Grandfather likes to be modern in his language.

Sustained Phonation

You will need to say the vowel /a/ at a normal loudness and hold it for **as long and as steadily as you can, until you run out of air.** Take a deep breath and say /a/. Repeat this again.

Cues for recording free speech

1. Describe what you enjoy doing in your free time.
2. Describe a recent holiday you took, where you went and what you did.

9.6. Appendix F

Label from Mayo Clinic Study Perceptual Features	**Description as in Duffy (2020)**
Abnormal pitch	Pitch is consistently too high or too low for age and sex.
Pitch breaks	Pitch shows sudden and uncontrolled variation (e.g., falsetto breaks).
Monopitch	Voice is characterized by monopitch or monotone. Voice lacks normal pitch variation.
Voice tremor	Voice shows fairly regular tremor, usually in 4-7 Hz range.
Monoloudness	Voice shows monotony of loudness. It lacks normal variations in loudness.
Excess loudness variation	Voice shows sudden, uncontrolled alterations in loudness, sometimes becoming too loud, sometimes too quiet.
Loudness decay	Progressive diminution or decay of loudness within an utterance.
Alternating loudness	Alternating changes in loudness within an utterance.
Loudness level (overall)	Voice is insufficiently or excessively loud.
Harsh voice	Voice is harsh, rough, and raspy.
Hoarse (wet) voice	There is wet, “liquid-sounding” hoarseness.
Breathy voice, or breathiness (continuous)	Voice is continuously breathy.
Breath voice, or breathiness (transient)	Breathiness is transient or intermittent.

Strained (strained-strangled) voice	Voice quality sounds strained or strangled (an apparently effortful squeezing od voice through glottis).
Voice stoppages (interruptions/arrests)	There are sudden stoppages of voice, as if airflow has been impeded.
Hypernasality	Resonance is excessively nasal.
Hyponasality	Resonance is hyponasal/ denasal
Nasal emission	There is nasal emission of air during speech, sometimes audible.
Forced inspiration-expiration	Speech is interrupted by sudden inspiration or expiration.
Audible inspiration	Audible, breathy inspiration.
Grunt at end of expiration	There is a grunt at the end of expiration during speech.
Rate, slow or fast	Rate of speech is abnormally slow or rapid.
Short phrases	Phrases are short, possibly because inspirations occur more often than normal. It sounds as if the speaker has run out of air. Often associated with reduced maximum vowel duration.
Increased rate in segments (accelerated rate)	Rate increases progressively from beginning to end of sample.
Increased rate overall (rapid rate)	Rate increases progressively from beginning to end of sample.
Reduced stress	Rate varies within or across utterances.
Variable rate	Rate varies within or across utterances.
Prolonged intervals	There is a prolongation of inter-word or inter-syllable intervals.
Inappropriate silences	There are inappropriate silent intervals.

Short rushes of speech	There are short, rapid rushes of speech separated by pauses.
Excess and equal stress	There is excess stress on usually unstressed syllables or parts of speech.
Imprecise consonants/ articulation	Consonants lack precision, the show inadequate sharpness, distortions, and lack crispness.
Prolonged phonemes	Phonemes are prolonged.
Irregular articulatory breakdowns	There are intermittent, non-systematic breakdowns in articulatory precision.
Distorted vowels	Vowels are distorted in their phonetic accuracy.
Palilalia	Compulsive repetition of words or phrases, usually in a context of accelerating rate and decreasing loudness.

9.7. Appendix G

Table 79. Participant demographic information for each group of the subset in the perceptual feature ratings analysis.

Groups	Participant code	Sex	Age	Time since PD onset (years)
PwPD-change	P11	F	52	5
PwPD-change	P19	M	81	3
PwPD-change	P21	F	64	4
PwPD-change	P22	M	93	8
PwPD-change	P23	M	71	21
PwPD-change	P25	M	50	24
PwPD-change	P28	M	77	11
PwPD-change	P38	M	64	7
PwPD-change	P46	M	78	11
PwPD-change	P48	M	66	1
PwPD-change	P56	F	78	4
PwPD-change	P59	M	59	5
PwPD-change	P62	F	73	8
PwPD-change	P66	F	68	13
PwPD-change	P70	F	67	7
PwPD-change	P72	M	75	8
PwPD-change	P76	M	67	6
PwPD-change	P80	M	70	3
PwPD-change	P90	M	59	5
PwPD-change	P95	M	82	7
PwPD-change	P99	M	59	6
PwPD-no change	P100	M	75	13
PwPD-no change	P24	M	72	7

PwPD-no change	P34	M	67	8
PwPD-no change	P35	M	76	10
PwPD-no change	P39	M	77	40
PwPD-no change	P42	M	70	2
PwPD-no change	P44	F	77	5
PwPD-no change	P50	F	66	12
PwPD-no change	P51	M	81	1
PwPD-no change	P52	M	73	10
PwPD-no change	P60	F	56	3
PwPD-no change	P61	M	77	9
PwPD-no change	P63	F	66	9
PwPD-no change	P67	F	67	15
PwPD-no change	P7	F	62	12
PwPD-no change	P73	M	66	7
PwPD-no change	P74	M	84	5
PwPD-no change	P84	M	79	4
PwPD-no change	P93	F	67	9
PwPD-no change	P94	M	67	4
PwPD-no change	P96	F	73	10
Control	Control_16	F	64	
Control	Control_20	M	78	
Control	Control_26	F	72	
Control	Control_29	F	51	
Control	Control_4	F	67	
Control	Control_45	F	82	
Control	Control_50	M	59	
Control	Control_55	M	72	
Control	Control_56	M	75	
Control	Control_58	M	81	

Table 80. Participant demographic information for each group for the intelligibility group analysis.

Group	Participant code	Sex	Age	Time since PD onset (years)
Mild-no change	P22	M	93	8
Mild-no change	P25	M	50	24
Mild-no change	P28	M	77	11
Mild-no change	P46	M	78	11
Mild-no change	P59	M	59	5
Mild-no change	P76	M	67	6
Mild-no change	P95	M	82	7
Mild-no change	P100	M	75	13
Mild-no change	P24	M	72	7
Mild-no change	P34	M	67	8
Mild-no change	P35	M	76	10
Mild-no change	P39	M	77	40
Mild-no change	P42	M	70	2
Mild-no change	P44	F	77	5
Mild-no change	P50	F	66	12
Mild-no change	P51	M	81	1
Mild-no change	P52	M	73	10
Mild-no change	P60	F	56	3
Mild-no change	P67	F	67	15
Mild-no change	P7	F	62	12
Mild-no change	P74	M	84	5
Mild-no change	P96	F	73	10
Mild-no change	P27	M	77	12
Mild-no change	P30	M	75	10
Mild-no change	P32	M	73	11

Mild-no change	P40	F	74	18
Mild-no change	P64	M	82	9
Mild-no change	P88	F	63	19
Mild-no change	P97	M	62	4
Mild-change	P11	F	52	5
Mild-change	P19	M	81	3
Mild-change	P21	F	64	4
Mild-change	P23	M	71	21
Mild-change	P38	M	64	7
Mild-change	P48	M	66	1
Mild-change	P56	F	78	4
Mild-change	P62	F	73	8
Mild-change	P66	F	68	13
Mild-change	P70	F	67	7
Mild-change	P72	M	75	8
Mild-change	P80	M	70	3
Mild-change	P90	M	59	5
Mild-change	P99	M	59	6
Mild-change	P61	M	77	9
Mild-change	P63	F	66	9
Mild-change	P73	M	66	7
Mild-change	P84	M	79	4
Mild-change	P93	F	67	9
Mild-change	P94	M	67	4
Mild-change	P1	M	62	5
Mild-change	P17	F	60	4
Mild-change	P18	M	59	4
Mild-change	P37	M	72	1
Mild-change	P53	M	64	5

Mild-change	P57	M	69	6
Mild-change	P58	M	68	6
Mild-change	P6	M	59	18
Mild-change	P69	F	63	21
Mild-change	P79	M	67	4
Mild-change	P8	F	68	26
Mild-change	P83	M	60	6
Mild-change	P87	M	58	2
Mild-change	P98	F	59	7
Control	Control_16	F	64	
Control	Control_20	M	78	
Control	Control_26	F	72	
Control	Control_29	F	51	
Control	Control_4	F	67	
Control	Control_45	F	82	
Control	Control_50	M	59	
Control	Control_55	M	72	
Control	Control_56	M	75	
Control	Control_58	M	81	
Control	Control_2	M	60	
Control	Control_3	F	70	
Control	Control_5	M	58	
Control	Control_6	F	69	
Control	Control_7	F	62	
Control	Control_8	F	49	
Control	Control_9	F	59	
Control	Control_10	F	57	
Control	Control_11	F	57	
Control	Control_12	F	82	

Control	Control_14	F	56
Control	Control_15	F	70
Control	Control_21	F	45
Control	Control_22	F	64
Control	Control_23	F	67
Control	Control_24	M	76
Control	Control_27	F	62
Control	Control_28	F	55
Control	Control_30	F	67
Control	Control_31	F	37
Control	Control_34	M	73
Control	Control_35	M	58
Control	Control_36	M	73
Control	Control_37	F	74
Control	Control_38	F	37
Control	Control_40	F	35
Control	Control_41	F	86
Control	Control_43	F	72
Control	Control_44	M	61
Control	Control_46	F	55
Control	Control_48	M	57
Control	Control_51	M	86
Control	Control_52	F	68
Control	Control_54	F	48
Control	Control_57	M	71
Control	Control_59	M	70
Control	Control_60	M	63