Speech markers of cognitive impairment in Parkinson's disease

SPEECH MARKERS OF COGNITIVE IMPAIRMENT IN PARKINSON'S DISEASE

A Masters Thesis Presented

By

Kara M. Smith, MD

Submitted to the Faculty of the Morningside Graduate School of Biomedical Sciences at the UMass Chan Medical School in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

NOVEMBER 21, 2022

NEUROLOGY

SPEECH MARKERS OF COGNITIVE IMPAIRMENT IN PARKINSON'S DISEASE

A Masters Thesis Presented

By

Kara M. Smith, MD

The signatures of the Master's Thesis Committee signify completion and approval as to style and content of the Thesis

_____

Robert H. Brown, DPhil, Chair of Committee

_____

Robert Goldberg, Member of Committee

_____

Bruce Barton, Member of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that the student has met all master's degree graduation requirements of the school.

_____

Mary Ellen Lane, Ph.D., Dean of the Morningside Graduate School of Biomedical Sciences

Master's of Clinical Investigation, November 21, 2022

Table of contents

List of Tables

ABSTRACT

Background: Cognitive impairment is a common non-motor symptom of Parkinson's disease (PD), but there are currently inadequate tools available to detect and monitor this complication. Speech undergoes an array of changes in individuals with PD, and speech may become more impaired as cognition declines. Speech markers may serve as useful, easy-to-access proxies of cognitive function. We evaluated differences in speech acoustic features on a reading task and a picture description task, as well as accuracy and pausing on a Stroop sentence task in persons with and without PD. We also assessed whether speech markers from these tasks were associated with mild cognitive impairment in persons with PD (PD-MCI).

Methods: We enrolled participants with PD (n=44) and older adult controls (n=8) at the University of Massachusetts Chan Medical School between January 2020 and October 2022. PD participants underwent cognitive testing in order to categorize cognitive status as mild cognitive impairment (PD-MCI) or normal cognition (PD-NC). All participants were audio recorded while completing a protocol of speech and language assessment, and speech data were processed and analyzed to obtain several acoustic features.

Results: Standard acoustic measures did not differ significantly between PD-MCI and PD-NC. Performance on reading and picture description tasks worsened over the course of speaking in both groups. Variability in fundamental frequency

declined over the course of speaking in the PD-MCI group compared to the PD-NC group during a picture description task. In a Stroop sentence task, accuracy and pausing measures were similar between PD-MCI and PD-NC groups.

Conclusions: Novel speech markers may be able to detect PD-MCI, but optimal speech task selection and analytic approach requires further development. Speech markers hold promise in monitoring PD symptoms, since data may be obtained frequently, remotely, and processed using automated algorithms.

ACKNOWLEDGEMENTS/PREFACE:

This work would not have been possible without the contributions and support from many individuals.

I would like to thank my mentoring team on the NIH/NIDCD K23 award that funded this work and supported my time to complete the MSCI program and this research. Dr. Robert Brown (primary mentor) encouraged and inspired me to complete this work and has been an unwavering source of support. Dr. Brown helped develop the funding proposal and study design, reviewed and critiqued preliminary results, and offered invaluable research career and grant writing guidance. Dr. Bruce Barton (secondary mentor) helped develop and plan the statistical approach for my K23 proposal and provided statistical support for the analyses presented here. Dr. David Caplan (Massachusetts General Hospital, secondary mentor) generously contributed time and expert guidance on linguistics and syntax. Dr. Cara Stepp (Boston University, secondary mentor) and her laboratory members trained me extensively on speech acoustic processing and analysis techniques, contributed analytic tools and support, and provided expert conceptual input in interpretation and dissemination of results. Dr. Thomas Quattieri and Dr. James Williamson, along with Tanya Talkar (Massachusetts

Institute of Technology Lincoln Laboratories, collaborators) provided analytical support on advanced signal processing and machine learning techniques.

I have been honored to have several students and residents contribute to this work. Victoria Wei (UMass Chan Medical School Class of 2022) performed the analysis of creaky voice presented in this work. She assisted with background research, implementation of the creak detector in Matlab, and contributed to interpretation of results on the use of creaky voice in our PD cohort. Anam Furrukh (UMass Chan Medical School Class of 2025) analyzed the Stroop sentence task. She performed the calculations on raw speech data to obtain accuracy and pausing outcomes, and contributed to the analysis and interpretation of results from the Stroop sentence task.

I would also like to thank the study team who supported this research. Meaghan Demers- Peel and Christina Manxhari (Clinical Research Assistants) assisted in recruiting and enrolling participants, collecting the cognitive and speech batteries from participants, pre-processing, organizing and helping to analyze speech files, creating and managing the study database, communicating with collaborators and the UMass Chan Medical School IRB, and contributing to abstract and manuscript writing.

Last but not least, I would like to thank my wonderful and supportive family for making it possible for me to complete the MSCI program and this work.

CHAPTER I: Parkinson's disease clinical features, speech dysfunction, and the role of speech markers

Parkinson's disease (PD) is a neurodegenerative disease characterized by motor symptoms including bradykinesia, rigidity, and tremor, as well as non-motor symptoms such as cognitive impairment. [1] This condition is caused by accumulation of alpha-synuclein protein adopting conformational changes and aggregating into Lewy bodies. Alpha-synuclein accumulation occurs predominantly in particular neural structures, such as the substantia nigra pars compacta (SNpc), which houses dopaminergic neurons in the brain. This neuropathology then spreads via cell to cell transmission to involve more widespread subcortical and cortical structures, with parallel progression of disease related symptoms and complications. [1] One of the critical points in the pathophysiology of PD occurs when SNpc neurodegeneration leads to sufficient dopamine depletion such that dopaminergic inflow to the basal ganglia becomes reduced. The basal ganglia are deep brain structures that control automated and learned motor actions. The hypodopaminergic state impairs basal ganglia motor circuits, leading to the cardinal motor symptoms of PD, including muscle rigidity, bradykinesia, tremor, and changes in posture, gait, and balance. [2] The basal ganglia also contain cognitive and associative circuits, which become impaired in individuals with PD leading to symptoms beyond the motor system.

PD may cause a wide range of non-motor symptoms, including autonomic dysfunction, neuropsychiatric symptoms, olfactory and gastrointestinal symptoms, and cognitive impairment.

Cognitive impairment is one of the most refractory and disabling complications of PD. [3] Cognitive impairment is common in PD, affecting approximately 20% of individuals at the time of diagnosis [3,4] and over 80% of individuals with PD over a typical 20-year disease course. [5] Severity ranges from subjective cognitive complaint, to mild cognitive impairment (MCI) to dementia (PDD). Individuals with PD-MCI are at high risk of converting to PDD, [6] and may be a key target population for better understanding the earlier stages of cognitive decline. While there are no available treatments for PD-MCI, there is promise that intervening in this earlier stage before development of dementia could be effective. Therefore, there is a great need for tools able to easily detect and monitor MCI in PD

Current approaches to identifying and diagnosing MCI in PD are, however, fraught with limitations. Brief cognitive assessments like the Montreal Cognitive Assessment (MoCA) are intended to be screening, not diagnostic tests, and are insensitive to early, subtle deficits and changes over time. [7]

The gold standard approach to diagnosing PD-MCI is extensive neuropsychological testing. The Movement disorders society updated diagnostic criteria for PD-MCI in 2012 to make the construct more concrete and improve the consistency of reporting. [8] The recommended Level II criteria requires

neuropsychological testing including 2 tests in each of 5 cognitive domains. PD-MCI criteria are met by impairment on two tests in one domain, or one test in each of two domains, with scores in the range of 1-2 standard deviations below age-adjusted norms.

While PD-MCI Level II criteria are valuable in research settings, this approach is not feasible in routine clinical care. The neuropsychological tests are time consuming, require the expertise of a neuropsychologist, and cannot be repeated frequently to monitor serial changes in various parameters. Patients with PD may not have access to neuropsychological testing in many geographic locations, wait lists are long, and the testing itself is often inconvenient and stressful. Easy-to-use tools to detect and monitor MCI in PD are needed to improve clinical care and fuel development of effective therapeutics.

Speech markers hold great promise as an innovative, scalable, and ecological approach to identifying and monitoring neurological and cognitive disorders. Speech markers have been studied in Alzheimer's disease, frontotemporal lobar degeneration dementias, amyotrophic lateral sclerosis (ALS), multiple sclerosis (MS), migraine, and others.

Speech markers are a highly promising area of development in persons with PD since communication difficulties occur in the majority of these patients. [9] The speech acoustic changes that occur in PD are typically summarized as "hypokinetic dysarthria" characterized by harsh breathy voice quality, decreased

volume, monopitch, imprecise articulation, and impaired speech rhythm. [10] There may be changes in pausing patterns, related to both cognitive linguistic processing[11,12] and respiratory coordination. [13] Linguistic changes may also occur, such as difficulties with action verb use and complex syntax. [14] Since these speech and language deficits are a well-known complication in PD, researchers have attempted to construct speech-based approaches to identifying and monitoring PD.

Among patients with PD, the study of speech markers has focused mainly on the detection of motor symptoms. It has been proposed that speech tests could supplement the clinical diagnosis of PD, which would enhance detection of PD in populations with poor access to Movement Disorders specialists. In one of the largest studies in this field, over 1,000 PD and 5,000 control participants used their own phones to call in and submit voice recordings. Remote assessment of individuals saying “ah” in this sustained phonation task was able to discriminate those diagnosed with PD from controls with over 60% accuracy.[15] Further work has expanded on this strategy, showing promise in detecting speech markers of early, prodromal PD. [16] By combining several acoustic parameters, reasonable accuracy has been reported for differentiating between individuals with REM behavior disorder (who are at risk for developing PD), early stage PD, and age-matched controls. [17] Speech markers may, therefore, become a powerful tool to identify those at-risk and with a current diagnosis of PD. However, it is less clear

that speech markers can be used to predict or monitor rates of disease progression once an individual is diagnosed.

The optimal models of combined acoustic features fail to be correlated with clinical change over time. [17] There are likely several reasons contributing to this gap in the field. First, it has been clinically established that speech and communication changes are heterogeneous in PD and may occur early or later in an individual's disease course with a highly variable rate of progression. Second, there may be a disconnect between the rate of change in limb motor actions and the speech motor system which involves axial muscles of the vocal apparatus. Third, the role of cognitive function in speech has been overlooked in most studies. This may be problematic since PD affects both speech motor and cognitive abilities, and both systems impact speech. Although speech changes often occur in parallel with cognitive decline, the causal relationship between these systems is not well understood. Some of the changes in speech may seem intuitively more related to motor impairment, such as rigidity of the vocal musculature and vocal tremor. On the other hand, cognitive decline may be more likely related to pausing patterns and the content or linguistics of speech output.

Further work is needed to characterize the patterns and underlying mechanisms of motor-related vs. cognitive-related speech markers in PD and there is emerging evidence that disentangling these elements is possible. A small

number of research groups have recently published work highlighting certain acoustic signatures sensitive to MCI in PD participants. [18-20] However, it will be important for researchers to replicate, validate, and generalize from these preliminary findings. For example, research tends to focus on a few selected acoustic features depending on the authors' prior experience with these measures and not compare novel algorithm performance to what has been published by other groups. It is also unclear how speech markers translate across languages and geographical regions, and how individual characteristics such as age, sex, and race/ethnicity impact algorithm accuracy.

In the current work, we aim to identify speech markers of cognitive impairment in participants with PD using novel experimental tasks and analytical approaches. Due to the concurrent influence of motor and cognitive deficits, speech function in PD may vary depending on the speech task and context. For instance, speech performance on a simple repetition task in a quiet lab environment may differ from "real-life" spontaneous speech performance in a situation with multiple speakers and more complex topics. However, research involving longer, more ecological, and more cognitively challenging spontaneous speech tasks are lacking. Our speech protocol includes standard lab tasks such as reading passages, as well as cognitively demanding tasks. We hypothesize that by comparing speech performance between tasks with varying degrees of cognitive difficulty, we can better capture speech signals associated with cognitive strain in the PD-MCI population.

Speech motor instability and automaticity in PD

Another consideration in PD speech research is that acoustic characteristics are typically averaged over an entire speech task. However, one hallmark feature of PD is difficulty sustaining motor actions with performance worsening over time during the task, or "motor instability".[21] This feature is clinically synonymous with bradykinesia and can be observed in limb movements as well as speech in PD. Motor instability is thought to arise from basal ganglia dysfunction. The globus pallidus is a main output structure of the basal ganglia, controlling initiation and termination of each submovement in a sequence of movements that makes up an automated motor program such as speaking. With globus pallidus dysfunction, the amplitude of submovements cannot be maintained as the sequence progresses. [22] In gait, stride length decreases as walking time and distance increase. In speech, this may manifest as progressive impairment in acoustic features as speaking time increases. [23] We hypothesized that assessment of acoustic performance at earlier and later time points during a speech task might reveal markers of motor instability. This analytic approach has been used only rarely in the field, and our findings will characterize these speech patterns in PD-MCI for the first time.

Another key feature of motor control impairment in PD is dual-task interference. In the general population and increasingly in older persons, individuals perform better when attempting a single task at a time compared to two tasks simultaneously. This dual-task interference has several proposed mechanisms,

including limited attentional resources or limited capacity of cognitive and motor systems. Dual-task interference has been described in older adults as a potential marker of MCI. [24]

In PD, dual-task interference is more severe than in non-PD populations because the basal ganglia dysfunction significantly impairs motor automaticity and increases the need for attention and higher-order cognitive control of previously automatic motor tasks. Compensatory processes and attention become strained when attempting to perform more than one task at a time.[22]

Dual-task interference in individuals with PD has been studied most extensively in gait research. Various posture and gait measurements decline when participants with PD are required to simultaneously perform cognitive or upper extremity motor tasks. [25] From clinical experience, it seems likely that similar dual-task interference mechanisms also occur during speech in individuals with PD. Patients will commonly report that if they focus on producing a loud, clear voice, they are able to do so in isolation, but during conversations and other daily speech situations their voice becomes softer and more difficult to understand. This observation may relate to the limited capacity hypothesis, such that with cognitive demand or multi-tasking (speaking while walking, doing housework,

driving) there is insufficient ability to simultaneously maintain optimal speech performance.

Despite this common clinical phenomenon, only one group has evaluated dual-task interference in speech among persons with PD. Whitfield and colleagues evaluated 12 participants with PD and 11 controls performing two speech tasks (a reading passage and extemporaneous speaking about a topic prompt) in isolation (single-task) and while drawing continuous circles (dual-task). Participants with PD had more pausing in speech on the extemporaneous task in dual-task condition compared to controls, but there was no significant difference in speech acoustic performance. [26,27] These study results showed the feasibility of measuring speech acoustic features during a dual-task paradigm in PD, however the drawing task required little attentional resources and speech motor automaticity may have been relatively preserved in this experiment. The study sample size was small, making it difficult to assess the impact of participant characteristics, including cognitive function, on dual-task interference.

We hypothesized that dual-task interference would be more apparent in more challenging experimental paradigms. We designed a speech protocol containing simple tasks such as single word sentence and paragraph reading, as well as cognitively challenging tasks including picture description, story recall, and a Stroop sentence task. We hypothesized that participants with normal cognitive

function will be less impacted by these cognitive demands and their speech performance will be more consistent and more preserved, whereas those with PD-MCI will have greater speech impairment on cognitively challenging tasks compared with easier tasks.

Through applying detailed and innovative analytic approaches to this speech protocol, we aim to identify novel speech markers of PD-MCI. Our protocols represent challenging dual-tasking paradigms which are more likely to reflect real-life daily experiences of patients living with PD. Understanding how dual-tasking interferes with functional performance in daily life could help shape therapeutic approaches to better address these disabling symptoms.

Specific Aims

The speech and communication deficits that occur in persons with PD are multifaceted and research is needed to develop speech assessment procedures sensitive to specific features of this condition. In this work, we describe a detailed approach to assessing speech, acoustic, and linguistic performance in persons with PD. We propose that administration of more varied and cognitively challenging tasks will better allow for assessment of motor and cognitive processes compared with standard approaches. We will pair our comprehensive assessment of speech with advanced analytical approaches to acoustic features. These novel experimental paradigms may lead to better success in identifying

markers of cognitive impairment than prior work. Our overall objective is to identify optimal methods for detecting MCI in PD using speech markers.

We therefore propose the following specific aims in order to identify novel speech markers of PD-MCI.

Aim 1: Acoustic speech markers in PD and controls

We will present our protocol for assessment of speech acoustic performance during a variety of speech tasks that vary in cognitive difficulty. We will describe speech acoustic performance on these tasks in a cohort of participants with PD-MCI, PD-NC, and controls. For this thesis, analysis will focus on acoustic features related to pitch, prosody, and voice quality. We will describe characteristic differences present in the PD-MCI group that may be further developed into detection and prognostication markers for cognitive decline in PD.

Aim 2: Acoustic markers of speech motor instability in PD

We will explore the phenomenon of motor instability in PD during speech, and evaluate likely different patterns in this endpoint according to cognitive status. We will focus on fundamental frequency, an acoustic feature associated with pitch. We will measure fundamental frequency during a reading passage and a picture description task in the overall PD cohort, and then compare performance in PD-MCI and PD-NC. We will describe patterns of fundamental frequency

decline over the course of speaking from the first to the last sentence on each task. We will draw connections between fundamental frequency instability and cognitive status.

Aim 3: Stroop sentences as a novel test for PD-MCI

We will describe a novel experimental task designed to increase cognitive demand during speaking. The "Stroop sentence task" involves reading sentences with a Stroop test embedded. We will evaluate acoustic performance on the simple (congruent) condition compared with the cognitively demanding (non-congruent) condition in the overall PD cohort and compare those with PD-MCI vs. PD-NC. We will identify potential future uses for this novel task as a cognitive assessment in patients with PD.

CHAPTER II: METHODS

Participants: Participants with idiopathic PD based on U.K. Brain Bank criteria[28] were enrolled at University of Massachusetts Chan Medical School from 2019-2022. PD participants had been diagnosed at least 2 years prior to participation, and were excluded if English was not their primary language. Other exclusion criteria were: previous deep brain stimulation surgery or history of other voice or laryngeal disorders. Older adult control participants were recruited from the Neurology Clinic and regional PD outreach programs, and were the relatives or

care partners of a person living with PD. Controls self-reported no history of neurological or voice disorders.

Ethics/consent: All subjects completed an informed consent form in accordance with the Declaration of Helsinki and approved by the institutional review board of the University of Massachusetts Chan Medical School.

Assessments: Clinical assessments included the Movement disorders society Unified PD rating scale (MDS-UPDRS, total Part III and axial scores), Montreal cognitive assessment (MoCA), and a battery of neuropsychological tests recommended by the MDS Task Force to determine a diagnosis of mild cognitive impairment by level II criteria. [29]

The cognitive battery includes the following tasks: Trail-making test A & B; Symbol digit modalities test; Boston Naming Test (30 item odd); Animal naming; Letter-guided verbal fluency; Judgement of line orientation (15 item odd); Boston Diagnostic Aphasia Examination; Hopkins verbal learning test (HVLT) – R immediate and HVLT-R Delayed and Recognition; Letter number sequencing; Brief visual memory test (BVMT) – R and BVMT-R Delayed and Recognition; Logical memory I (WMS-R, Anna Thompson story) and Logical memory II. Each case was reviewed by a panel of at least 3 clinicians to determine the consensus cognitive diagnosis of either mild cognitive impairment (MCI) or normal cognition (NC). Participants with dementia (n=1) and those with indeterminate diagnosis (n=1) were excluded.

All PD participants were taking PD medications, and were in the “on” state during assessment. All participants enrolled after March 2020 were wearing face coverings during the assessment. While face coverings may impact certain acoustic features of speech [30,31] the main acoustic variables of interest in our study have not been shown to be altered by masking. Additionally, our main analytical methods involve comparing performance within an individual between different tasks or conditions, thus they serve as their own baseline control and modulation of voice by masking would be less of a concern.

Speech assessment included a standard battery of tasks which were audio recorded. This protocol included sustained phonation of vowels, consonant-vowel-consonant patterns, and a series of reading sentences and passages including the Rainbow passage. [32] One set of sentences included an embedded Stroop test (see below). Participants were then asked to describe the Cookie Theft Picture from the Boston Diagnostic Aphasia examination. [33] For this task, the participant was instructed that they would be given 60 seconds to describe the picture presented to them to the best of their ability, aiming to fill the entirety of the 60 seconds with their description.

Speech analysis: Recording was performed using a hand-held digital recorder with a built-in unidirectional head-mounted microphone (Zoom H4n Pro Handy Recorder and Shure WH20 headset) at a sampling rate of 44,100 Hz and 16 bits. The audio recordings were saved as .WAV files and analyzed using the software Praat. [34] The participant’s description of the Cookie Theft picture was then

manually transcribed and segmented into utterances. For purposes of this study, utterances were defined as statements that, at the minimum, included a subject and a verb, and at the maximum included a subject, verb, and restrictive clause. We identified the first and last utterance of the Rainbow passage and the picture description tasks and cropped these into separate wav files.

Pauses were manually identified and measured using Praat. We measured each pause duration and rounded to the nearest 500 milliseconds. For instance, pauses between 500 and 1000 milliseconds were rounded to 500 milliseconds, and pauses 1000-1500 milliseconds were rounded to 1000 milliseconds. We counted pauses within utterances longer than 500 milliseconds and pauses between utterances longer than 2000 milliseconds as clinically relevant. We manually assessed the location of pauses within the transcripts and summed the number and duration of pauses proceeding key word types of interest (see below). The summed duration of pauses before each linguistic component was used to calculate a pause percent, corrected to account for differences in speech output and content between participants. To control for variability in speaking rate and speech motor function, a modified words per minute variable was used (modified wpm = words divided by time after subtracting the pausing time).

Speech acoustic analysis was also performed in Praat. The fundamental frequency (f_0 , mean and standard deviation, f_0 SD) was calculated for the first and the last utterance of the Rainbow Passage and the Cookie Theft picture description tasks. f_0 SDs were then adjusted by conversion into semitones

(STSDs). [35] STSD estimates prosodic variation, is less likely to be confounded by the effects of mean f_o (and thus speaker sex), and is robust to single instances of f_o deviations. [35] Glottalization and other non-modal phonation was excluded from analysis. Two independent technicians (KMS and MDP or CM) independently calculated f_o parameters for all participants. KMS was the primary rater for all results as the principal investigator of this study, and a trained clinical research assistant (either MDP or CM) performed a second set of calculations.

We analyzed several acoustic features that are standard in the vocal analysis of PD across the literature including: harmonic-to-noise ratio (HNR, a measure of periodicity of speech signal that is impaired by laryngeal pathology), cepstral peak prominence-smoothed form (CPPS, a measure of voice quality and breathiness), low to high ratio (L/H ratio, a measure of the strength of the energy across different frequencies). These measurements were performed in Praat using standard settings. We also evaluated novel acoustic features that have not been described in PD including vocal fry/creaky voice. Vocal fry or creaky voice has a low pitch that lies below the model register for both males and females, and has a rough quality with irregular pitch periods in the speech signal.[36] In order to measure creak, wav files were pre-processed with performed with Audacity® software. Data was downsampled to 16 kHz and converted from stereo to mono. Voiced frames were identified with MATLAB Audio Toolbox (Mathworks, Natick MA) and an algorithm by Drugman and Kane was applied[36](Matlab script provided by Daryush Mehta (MIT). Creak percentages were

calculated by dividing the total number of creaky frames over the total number of voiced frames x 100. [37] Our approaches to the analysis of collected data are described in each relevant section. All statistical tests were two sided. Statistical significance was set *a priori* at $p \leq .05$. All statistical analyses were conducted with STATA (Version 17).

CHAPTER III: Aim 1 Results- Baseline acoustic performance in PD

Background: Baseline acoustic performance in PD

In this section, acoustic performance will be described in our cohort of PD and control participants on select speech tasks. We will describe standard acoustic measures known to be affected in voice impairments in general as well as in PD. L/H ratio measures the mean ratio of energy below vs. above 4 hz. Normal voice signal tends to have greater low-frequency energy, so in vocal dysfunction there may be a decrease in this ratio. CPPS reflects the energy spectrum with cepstral transformation and measures the magnitude of the dominant peak relative to a smoothed baseline. CPPS is typically reduced in speakers with PD. Measures of fundamental frequency (f_0) reflect what is perceived as pitch of voice. The variability of f_0 (f_0 SD) is another standard measure, and is decreased in a more monotone voice or those lacking appropriate prosody. STSD is an f_0 -based measure of variation that is controlled for the speaker's baseline f_0 to account for

confounding by speaker sex and inherent differences that may be present between speakers with higher and lower baseline f_0. [35,38]

While analyzing f_0 in our participant sample, we noted frequent instances of very low frequency voice episodes. During these instances, the frequency would suddenly drop below the range of a typical speaker's pitch, and the waveform would appear more irregular. This phenomenon has been described before in the literature and has been classified under several different terms (glottalization, creaky voice, vocal fry). Terminology has varied depending on the speaker population being studied and the location of the occurrences within speech. [36] We will refer to these occurrences using the broadest term, "creaky voice". Creaky voice is acoustically characterized by low frequency below the modal range of male or female speakers with irregular f_0. It can be sporadically found in the general population, associated with environmental factors like dialect, sex, and social group.[39] We noted that most of the literature describing f_0 patterns in PD are based on speech waveform analysis that has excluded these episodes from the waveform prior to averaging (personal communication, Dr. Cara Stepp).

We hypothesized that the dysfunction of the vocal apparatus musculature in PD may increase creaky voice episodes. Mechanistically, creaky voice is produced when there is a low rate of vocal fold vibration and constricted glottis, and the rigidity of the vocal tract in PD may make such episodes more likely. In PD, more severe global motor symptoms may be associated with increased occurrence of creaky voice. We decided to utilize a creaky voice detector algorithm that has

been applied and validated in several speaker populations, but never before to our knowledge in speakers with PD. We evaluated the characteristic patterns of creaky voice in PD in different types of speech tasks. In the current work, we will assess whether motor or cognitive symptoms are associated with the occurrence of this phenomenon .

We report on these acoustic measures in two types of speaking tasks elicited in our protocol, a reading passage and a picture description task. Other speech tasks included in our protocol will be presented in subsequent publications.

Methods

Statistical approach: We calculated mean and SD of baseline clinical and demographic characteristics and compared these by group (PD vs. control) and cognitive status (PD-MCI or PD-NC) using t-tests for continuous variables and chi-squared tests for categorical variables. We characterized the speech performance in our cohort by calculating mean and standard deviation (SD) of major standard acoustic features in select speech tasks (f_o mean, SD and STSD as well as L/H ratio and CPPS). These variables were found to be normally distributed using visualization of quantile plots and thus parametric tests were applied. Vocal creak variables were not normally distributed, so the natural log was taken to achieve normality for the subsequent analyses. T-tests were performed to explore differences in creak percent between gender and cognitive diagnosis in the PD group. Pearson's correlations were calculated to assess the

relationship between creak percent and other clinical characteristics as well as cognitive and motor symptoms. Linear regression models controlling for sex, age, and motor severity were created to determine factors associated with creak percentage in each task.

Results:

Study feasibility and baseline characteristics

Our first aim was to establish a study population of PD and control participants at UMass Chan Medical School. This cohort built off earlier pilot studies by the investigator, adding a more comprehensive speech and language protocol as well as a more rigorous approach to cognitive status determination. Mentorship from external collaborators was also utilized to optimize the technical aspects of voice recording and pre-processing (Dr. Cara Stepp, Boston University and Dr. Thomas Quatieri, MIT Lincoln Labs). Despite the COVID-19 pandemic, we were able to enroll 44 new PD participants and 8 control older adult participants. Participants were recruited from the UMass Memorial Health Neurology Clinic, through newsletters and email announcements from regional PD organizations and support groups, and from the Michael J. Fox Foundation Trial Finder website. All participants were able to complete the entire speech and cognitive battery, and there was no missing data for these key components of the project. One PD participant was excluded due to dementia diagnosis and one due to not meeting exclusion criteria.

Study Population Characteristics

The clinical characteristics of our study cohort are provided in Table 1. Compared with the PD participants, controls were significantly older and included a greater proportion of men. In the PD group, 22 participants were categorized as PD-MCI and 20 as PD-NC and 6 are pending cognitive categorization at the time of this writing.

Table 1. Clinical characteristics of Parkinson's disease and control participants

Variable	PD (n=42)	Control (n=8)	P value
Age (yrs, mean, SD)	66.8 (7.9)	72.6 (3.8)	0.02
Sex (#M, #F)	15 M, 27 F	6 M, 2 F	0.02
Education (yrs, mean, SD)	16.4 (3.0)	15.5 (1.8)	0.45
MoCA (mean, SD)	27.2 (2.3)	27.8 (2.1)	0.54
PD duration (yrs, mean, SD)	4.8 (3.6)	n/a	
MDS-UPDRS Part III	29.8 (11.1)	n/a	
MDS-UPDRS Part III Axial score	4.6 (2.3)	n/a	

MoCA= Montreal cognitive assessment, MDS-UPDRS Part III=Movement disorders society Unified PD rating scale, Part III Motor Examination

Standard measures of speech acoustic performance

On a standardized reading passage task (Rainbow passage), participants with PD demonstrated similar mean f_o when averaged over the entire task compared with controls, and this measure is associated with the lay definition of voice pitch (Table 2). Mean f_o was also similar between PD and control participants on a picture description task (Cookie theft picture) in which the content varied by speaker. Variability in f_o was similar between PD and controls in the Rainbow passage, indicating similar degree of monotonicity or prosody. However, f_o variability was lower in those with PD during the picture description task, indicating greater degree of monotonicity when these participants were asked to generate spontaneous speech content. PD participants demonstrated lower CPPS, a sensitive measure of dysphonia, in both tasks compared with controls. The L/H ratio was similar between PD and control participants in both tasks.

In the PD participants, comparing acoustic performance between the reading and picture description tasks, there was no significant difference in L/H ratio (mean difference = 1.44 (SE 0.86), p=0.10)), mean f_o (mean difference = 2.41 (SE 7.53), p=0.75)), f_o variability (mean difference = 0.90 (SE 2.2), p=0.68)), or CPPS (mean difference = 0.37 (SE 0.28), p=0.18)).

Among PD participants, the PD-MCI group and PD-NC group did not differ significantly in mean f_o or f_o variability (SD or STSD), L/H ratio, or CPPS in either task (data not shown).

Table 2. Acoustic performance on reading and picture description tasks in patients with Parkinson Disease (PD) and controls

	Cookie theft picture			Rainbow passage		
Variable (mean, SD)	PD	Controls	p	PD	Controls	P value
L/H ratio	24.94 (3.91)	23.17 (4.12)	0.25	26.38 (4.15)	24.07 (5.13)	0.17
f_o mean	135.92 (35.16)	158.50 (30.86)	0.10	138.33 (35.50)	159.09 (26.43)	0.12
f_o SD	19.31 (10.54)	29.81 (8.44)	**0.01**	20.22 (9.85)	25.60 (4.69)	0.14
CPPS	5.01 (1.28)	6.24 (1.06)	**0.01**	5.38 (1.31)	6.51 (1.24)	**0.03**

L/H ratio=low to high ratio, f_o = fundamental frequency, CPPS= cepstral peak prominence (smoothed)

Vocal creak was calculated in the Rainbow passage and Cookie theft picture description in the first 26 PD participants in the cohort as an exploratory analysis, since this acoustic feature has not been applied in PD before to our knowledge. Mean creak percentage was 4.23 (SD 4.40) in the Rainbow passage and 4.22 (SD 4.00) in the Cookie theft picture description task. There were no significant

differences observed in creak percent between males and females, or between PD-MCI and PD-NC. There were no statistically significant correlations between the creak percentage variables and either global cognitive function (MoCA) or motor severity (MDS-UPDRS III). In linear regression models, neither cognitive function or motor severity was significantly associated with creak percent on either task, after adjusting for sex and age. We did not observe any factors that were significantly associated with cookie creak percentage after including the covariates of age and sex. Age and sex were the only significant factors associated with Rainbow passage creak percent (data not shown).

Discussion:

Our cohort of participants with PD demonstrated more impaired CPPS in both speech tasks compared to controls. This is consistent with the literature, as leading researchers in this field have found CPPS to be a sensitive marker of PD in both early and later disease stages.[40] We also found a significant impairment in f_0 variability in the picture description task in PD compared with controls, but this deficit was not significantly different from the reading passage among those with PD. Thus, monotonicity was likely a persistent feature of the speakers with PD. In a reading task, the grammatical markers and standardized sentence content serve as guidepoints for prosody, whereas in the picture description task these are absent. In summary, the first set of standard acoustic features we have assessed in our PD cohort revealed expected, well-described speech patterns in both reading and spontaneous speech tasks.

On the more novel acoustic analysis of creaky voice, we found that this feature was highly variable between PD participants. It seems to be an individualistic acoustic property, with some speakers manifesting more creak in both tasks than others. We were unable to identify any predictable patterns of creak depending on motor or cognitive symptoms. We are limited in our conclusions regarding creaky voice as a speech marker in PD, since this is a small sample size and creak percent exhibited a high degree of variability. It is unclear whether automated creaky voice algorithms could be developed as a useful marker of PD status or PD-related symptoms in future work.

CHAPTER IV: Aim 2 Results- Fundamental frequency as a marker of speech motor instability in PD-MCI

Background on fundamental frequency

We examined fundamental frequency (f_o) as a marker of speech motor instability in PD. f_o is a major feature of voice, derived from vocal fold vibrations. Decreased f_o variability is related to the perception of monopitch and is a characteristic of hypokinetic dysarthria in PD. [23,41]Most studies evaluating f_o in PD use values averaged over an entire speech task. Given that phonation requires sustained muscle activation and coordination of the larynx, f_o parameters may decline over the course of speaking in PD. The f_o parameters that are standard to assess

include mean f_o and f_o variability. In the general population, mean f_o increases with prolonged speaking, a phenomenon that has been suggested as an indicator of vocal fatigue. [42]

However, few studies have evaluated how f_o mean and variability change over the course of speaking tasks in participants with PD. Skodda and colleagues examined f_o variability during reading of four complex sentences in a cohort of 138 PD and 50 age-matched control participants. [43] Their results indicated a statistically significant decline in f_o variability from sentence one to sentence four. Bowen and colleagues further explored motor instability of f_o features in PD during the first paragraph of rainbow passage. In contrast to the findings of Skodda et al., they found no significant difference in f_o variability from the first to the last sentence of the paragraph between PD (n=32) and controls (n=32). [35] These inconsistent results may relate to differences in the speech tasks (a set of sentences vs. reading passage) as well as in the size and composition of the sample.

We aimed to address gaps in the literature by examining how f_o mean and variability change over the course of two speaking tasks in a well-characterized cohort of participants with PD. We employed a standard reading task and a picture description task which requires lexical retrieval, planning and organization of content, and generation of grammar and syntax. We also carefully examined participant-level characteristics, focusing on cognitive status (PD-MCI or PD-NC). We hypothesized that participants with PD-MCI would have greater decline in

their f_o performance during the picture description task than those with PD-NC. Specifically, we hypothesized that mean f_o would increase and f_o variability would decrease over the course of speaking time. We hypothesized that when performance during the reading and the picture description task is compared, participants with normal cognition would exhibit a smaller difference in these f_o declines between the tasks, and that participants with impaired cognition would exhibit a larger discrepancy in their f_o declines between the tasks.

Methods

This analysis includes 28 participants with PD from the above-described cohort who had f_o processing complete at the time of this writing. All participants performed the speech protocol described above, and audio files were processed as already described. We obtained f_o mean and variability as well as STSD for the first and the last utterance in each of the speech tasks. Normality of fundamental frequency variables was first assured visually using quantile-quantile plots. Variables of interest had plausibly normal distributions and we therefore used parametric tests. To assess inter-rater consistency, we calculated Pearson's correlations between the two technicians' scores and also re-ran all analyses with the second rater's scores to ensure overall results were consistent. The primary rater's (KMS) values were used for the reported results.

To test the hypothesis that changes in mean f_o and f_o variability over the course of speaking would be greater during picture description compared to reading, we

calculated the difference between the first and the later utterance f_o mean and STSD and compared these differences for the reading vs. the picture description tasks using two paired Student's t-tests. When appropriate, Cohen's *d* effect sizes were calculated to characterize statistically significant differences. Some participants spoke for a shorter duration of time during the picture description task, compared to the reading task in which the amount of content and speaking length was externally determined. We therefore performed a sensitivity analysis to determine if differing speaking time impacted the results, by using the Rainbow Passage sentence that occurred at the same time point as the last utterance in the picture description task (within 5 seconds). To test the hypothesis that participants with normal cognition would exhibit a smaller difference in f_o declines between the tasks, and that participants with MCI would exhibit a larger difference in f_o declines between the tasks, we first calculated the following values:

Between-task Δ mean f_o *= (picture task mean* f_o^{FIRST} *– picture task mean* f_o^{LATER}*) – (reading task mean* f_o^{FIRST} *– reading task mean* f_o^{LATER}*)*

Between-task Δ STSD f_o *= (picture task STSD* f_o^{FIRST} *– picture task STSD* f_o^{LATER}*) – (reading task STSD* f_o^{FIRST} *– reading task STSD* f_o^{LATER}*)*

We ran independent sample t-tests comparing the PD-MCI and PD-NC groups on these values, as well as on clinical features (age, sex, motor severity). We then ran logistic regression models to predict cognitive diagnosis, with f_o

measures as independent variables, controlling for age and sex. The logistic regression model including STSD was also adjusted for mean baseline f_0 to account for excess variability between participants in this measure which can have a wide range among speakers.

Results*:*

In the 28 PD participants used for these analyses, the PD-MCI PD-NC groups did not differ significantly in age, sex, PD duration, or motor severity score on the MDS-UPDRS (Table 3).

Table 3. Clinical characteristics by cognitive diagnosis in participants assessed for fundamental frequency measures

	Mild cognitive impairment (n=16)	Normal cognition (n=12)	P value
Age, years (mean, SD)	68.3 (2.3)	66.5 (2.2)	0.58
Sex (# M/F)	10 M, 6 F	3 M, 9 F	0.05
Disease duration, years (mean, SD)	4.4 (0.72)	6.1 (1.0)	0.21
MDS-UPDRS Part III total (mean, SD)	28.2 (3.7)	31.2 (2.8)	0.51
MDS-UPDRS Part II Speech item	0.64 (0.27)	0.82 (0.18)	0.62

Decline in mean f_o and f_o variability during speaking tasks

The f_o parameters changed over the course of speaking in both the reading and the picture description tasks. In the picture description task, mean f_o decreased significantly by an average of 12.3 Hz (SD 2.30)(Table 4). In the reading task, mean f_o decreased by an average of 11.08 (SD 1.4). There was no statistically significant difference in the degree of mean f_o decline between the two tasks (mean difference 1.22 Hz, SD 3.33, p=0.72). The adjusted f_o variability measure,

STSD, also declined over both speaking tasks (Table 3). However, f_0 variability declined significantly more in the reading task compared to the picture description task (change in STSD f_0 mean difference= -0.52 ST (SD 0.19), $p<0.001$). This finding had a medium effect size (Cohen's d= 0.60, 95% CI=-0.17-1.37).

A sensitivity analysis to control for the longer duration of the full reading task as compared with the picture description task did not change the direction or magnitude of these findings. When these analyses were re-run using a second independent rater's measurements, results were similar in direction of the changes, as well as magnitude and statistical significance. Additionally, there were statistically significant correlations (Spearman's rho values all ≥0.80) between the two raters' acoustic calculations.

Table 4. Fundamental frequency (f_0) measures in first and last utterances of speech tasks

Utterance	Mean f_0	p	STSD	P value
Reading, first (mean, SD)	151.59 (6.37)		2.54 (0.14)	
Reading, last	140.51 (5.99)	<.0001	1.81 (0.13)	<.001
Picture description, first	148.10 (6.55)		2.05 (0.17)	
Picture description, last	135.79 (5.64)	<.001	1.84 (0.23)	0.31

Association of f_0 measures with cognitive diagnosis

Comparing the PD-MCI and PD-NC groups, there were no statistically significant differences in the decline in mean f_0 or STSD in either task.

In a logistic regression model with cognitive diagnosis as the dependent variable, STSD decline during the picture description task was significantly associated with PD-MCI, after adjusting for age, sex, and baseline mean f_0 (OR=.16, 95% CI=0.029-0.923,p=0.04). The mean f_0 decline in the picture description task was not significantly associated with cognitive diagnosis. In the reading task, neither mean f_0 or STSD declines were significantly associated with cognitive diagnosis (data not shown).

Discussion

We characterized f_0 performance in participants with PD over the course of two speaking tasks of differing cognitive demand. Mean f_0 and f_0 variability measures declined over the course of speaking in both reading and picture description tasks. This decline was greater and more consistent in the reading task in which speech content was standard, and more variable between individuals in the picture description task in which content varied. STSD decline, an adjusted marker of f_0 variability, during the course of the picture description task was significantly predictive of PD-MCI status.

The first implication of these results is that temporal changes in f_0 parameters can be identified over the course of relatively brief reading and picture description tasks in participants in PD. The declines we observed in mean f_0 and f_0 variability suggest that the phenomenon of motor instability is evident in just a few minutes of speech in PD. These declines translate clinically into listener perception of lower pitch and worse monopitch, which may impact intelligibility and engagement with a speaker.

Our findings also demonstrate that the cognitive status of an individual speaker with PD affects f_0 performance, especially for speaking tasks with greater cognitive demand. This phenomenon was apparent even with relatively early, mild cognitive deficits as is seen in MCI. Our findings diverge from some recent reports that acoustic measures related to pitch or prosody (such as our f_0

measures) are less useful in discriminating PD-MCI. In two studies with similar approaches, acoustic measures of phonation timing and coordination such as phonemic identifiability were better able to predict PD-MCI status from a story retelling task[19,20]. There are many potential reasons for the differences in our results. The other studies were conducted in Spanish and Czech, did not assess changes in f_0 measures over the course of the task but rather used overall averages. Importantly, cognitive categorization was done using a less rigorous approach (MDS Level I MoCA-based approach). We propose that changes in STSD during speaking may be a useful marker of cognitive function in PD. Analysis of f_0 measures is relatively simple and widely available. Future work should explore whether measurement of STSD decline over longer speaking tasks or use of alternative speaking prompts increases the sensitivity of this measure to predict MCI.

This work is the first to our knowledge to evaluate how speech acoustics change over the course of spontaneous speech in PD. In the general population, differences in acoustic and prosodic measures in different types of speaking tasks have been evaluated. In contrast to our results in PD participants, mean f_0 increases with prolonged reading in the general population. [44,45] One proposed physiological explanation of increased mean f_0 with vocal fatigue is increased laryngeal tension with vocal loading. Of note, studies of vocal fatigue in healthy controls or those with dysphonic disorders have been conducted with reading tasks substantially longer than ours in this study, typically 60-120 minutes.

Experiments with longer duration of speech would be needed to better understand the interplay between vocal fatigue, motor instability, and f_o in PD. If the patterns we identified are also present in longer speech tasks, it may suggest that speakers with PD have a different response to vocal fatigue. One potential explanation is that PD speakers have impaired feedback and feedforward motor speech systems, such that they do not perceive speech changes with prolonged speaking and make fewer compensatory adjustments. [46,47] Abur and colleagues proposed that cognition and attention play a key role in sensorimotor control systems impacting f_o variability and compensatory reflexive changes. [48] Another possible theory is that speakers with PD are less able to compensate for vocal fatigue due to disease-related effects on the vocal apparatus. Further studies are needed to enhance our understanding of motor instability more broadly in PD speech.

There are several limitations to this work. The order of the speaking tasks was not randomized, because they were both part of a longer standardized protocol. The fact that the picture description task occurred later in the protocol than the reading task could have impacted the data in various ways. For example, the participants may have been more fatigued during this task which occurred after about 10 minutes of continuous speech. Another limitation to this study is that we did not assess other standard speech acoustic measures, but rather focused on f_o parameters. Although this is a limitation, this choice allowed us to compare the standardized reading passage with the picture description task; this would not

have been possible for articulatory speech measures, which would have been highly affected by the differences in speech content between the two tasks. We also did not account for respiratory function during speaking, which is known to be impaired in PD.[49] It is possible that the later utterances were more affected by decreased lung volume capacity compared to earlier utterances, and that this impacted the f_0 results differentially between the two speech task types. Skodda and colleagues measured mean phonation time as a marker of respiratory capacity and did not find this measure correlated with f_0 variability. However, the PD participants in the cohort had similar mean phonation time compared to controls, which suggests this may not be a sufficient marker to assess this relationship. [43]

Strengths of this study include the analysis of a spontaneous, naturalistic speech task and rigorous approach to diagnosing cognitive status. Spontaneous speech better reflects daily communication than repetition or reading tasks. Although the picture description task was relatively brief and may be limited in evoking certain linguistic characteristics, it requires the speaker to plan, initiate, and organize semantic and syntactic content under time constraints and to utilize naming and lexical retrieval processes. In comparison, when reading aloud, one is able to use the text and punctuation to plan their pitch structure. By demonstrating different f_0 patterns between picture description and reading, we highlight the importance of validating speech markers for specific types of speech tasks in PD. Furthermore, since f_0 patterns differed by cognitive status, we recommend that comparison of

acoustic markers between speech tasks be performed to better capture both motor and non-motor features of PD.

Understanding how speech declines over the course of speaking in various contexts has important clinical implications for individuals living with PD. Improved speech therapy interventions and novel technology might assist by improving and compensating for these declines. Speech marker development also holds great potential to improve assessment of both motor and cognitive status in PD. If a panel of cognition-specific acoustic markers could be implemented, these assessments could be obtained easily and frequently in patients with PD using remote technology, leading to improved detection and monitoring of PD-MCI.

In summary, mean f_o and f_o variability are unstable over the course of speaking in PD, and patterns of changes in these f_o parameters vary by type of speech task as well as by cognitive status. Future work involving f_o-based speech biomarker development in PD should focus on these time-based and task-dependent characteristics, as they could lead to improved markers of motor and cognitive symptoms in PD.

CHAPTER V: Aim 3 Results- A novel Stroop sentence task

Background*:*

Different types and degrees of cognitive demand may differentially impact speech characteristics in individuals with PD. We have previously described speech tasks with low to moderate cognitive difficulty (reading and picture description), but more complex tasks might yield informative results about the intersection of cognitive impairment and speech performance.

The theory of limited capacity and dual-task interference posits that if one task requires more cognitive control, a second simultaneous task will be de-prioritized and performance will decline. Our main objective has been to establish how this decline in speech performance can be quantified depending on the extent of cognitive impairment of the speaker, and the cognitive demand of the task. The central hypothesis of our work is that individuals with PD-MCI will find tasks with at least some cognitive component more challenging and their speech performance will reflect this “overload.” In this section, we will describe a complementary approach using a task that is more cognitively difficult, the Stroop sentence task.

We know from clinical experience that individuals with PD tend to display greater speech deficits when engaging in cognitively or socially complex speaking situations. Difficulty with lexical retrieval, planning at the word and sentence level, attention, and working memory may impact speech performance. It is important

to understand how specific cognitive abilities impact speech motor and cognitive effects on speech in daily activities in PD. Understanding these deficits may enable improved interventions such as speech therapy programs and communication assistance technologies.

We selected a novel speech task aimed at increasing cognitive load, the Stroop sentence task, to assess the interaction of executive function and speech in individuals with PD. The Stroop task is a widely used neuropsychological test of executive function and cognitive interference. [50] The standard Stroop task consists of a list of printed names of colors. The first condition is the congruent section, in which the list contains names of colors printed in a matching color (blue, red). The second condition is the non-congruent section, in which the list contains names of colors printed in a non-matching color (blue, red), and the patient is instructed to name the color of the ink instead of reading the word.

The non-congruent task adds cognitive demand, as the patient must inhibit the interference of the more natural, automated process of reading the word. The task is usually scored by assessing the number of items correctly named in 45 seconds in each condition, and a calculation is performed to obtain the interference score. [51]

A systematic review of various scoring procedures found that the typical calculations used to score the Stroop test did not adequately account for sources of individual variability such as reading skills, psychomotor speed, and attention.

[52] Additionally, the Stroop task may not reflect a patient's daily cognitive experiences accurately, and a more naturalistic version of the task may be more clinically relevant. Finally, the Stroop task may be less sensitive in persons with PD compared to healthy controls or other disease populations. In PD, slowness in both conditions may result in a ceiling/floor effect, making it difficult to detect differences between the conditions. [53]

To address these limitations, we utilized a Stroop sentence task that has been recently developed but never before applied to patients with PD.[54] By embedding Stroop words in a sentence reading task, acoustic and linguistic measures, such as pitch and prosody-based features, measures of dysphonia, and pausing patterns can be assessed which would not be possible when using only a list of color words. The Stroop sentence task may also be more naturalistic and reflective of cognitive experiences of daily living. In healthy older adults, the non-congruent condition of the Stroop sentence task was associated with increased autonomic nervous system activity which confirms that this condition increases cognitive load compared to the congruent condition. [54]

While the standard Stroop task is not recommended to be used in isolation to screen for, or diagnose, cognitive impairment in PD, we propose that the detailed, quantitative speech measures obtained from the Stroop sentence task may be sensitive to PD-MCI. We evaluated performance on this task extensively, including the number of errors, pausing time between words, and acoustic measures. We also explored correlations between Stroop sentence performance

and cognitive domain-specific neuropsychological tests to better understand which cognitive functions are involved in performing this task.

Methods*:*

Data from 26 participants that had PD-MCI or PD-NC cognitive diagnoses available and Stroop sentence task analyzed were used. Participants were given the Stroop sentence task instructions just before they were shown the sentences, and then audio recording was begun per protocol described above. Study participants completed the congruent condition first and the non-congruent condition second, both of which contained six sentences. The sentences were identical except for the color words and text color, such that linguistic complexity did not differ between the conditions.

Audio files were saved in .wav format and processed using Praat. The congruent sentences were cropped into one audio file, and the non-congruent sentences cropped into another audio file. Errors were scored manually by listening to the audio files. A separate category of errors was created for errors that may have related to difficulty with color discrimination. For example, text printed in red vs. pink vs. orange was challenging for some participants to discriminate. Thus, in the non-congruent condition, if a participant correctly inhibited the impulse to read the word but gave the wrong color (e.g., red for pink) this was counted in the total errors and also in the color discrimination error counts. Silent pauses

between words were manually identified and summed in each condition. Pauses >500 milliseconds were considered clinically relevant.

The sentences for each condition were then subjected to acoustic analysis as described previously.

Statistical analysis: Primary outcomes of interest were the total number of errors in the non-congruent condition and the total duration of pauses during the non-congruent task. Since these variables were not normally disturbed, and given the small sample size, we used non-parametric approaches. Secondary outcome variables included non-congruent condition color discrimination errors, mean f_0, SD f_0 , STSD, L/H ratio, and CPPS. We assessed Spearman's rho correlations between primary outcome variables and cognitive (MoCA) and motor (MDS-UPDRS Part III) scores. We utilized the Mann-Whitney U-test to assess differences in the primary and second outcome variables by cognitive diagnosis group (PD-MCI or PD-NC).

Results:

Performance on the Stroop sentence task in the overall group

In the congruent condition, the majority of participants made no errors (n=21), 3 participants made 1 color discrimination error, 1 participant made 2 color discrimination errors, and 1 participant made 3 non-color discrimination errors.

The median duration of pausing in the congruent condition was 3.8 seconds (IQR 1.5-7.0 seconds).

In the non-congruent condition, the median number of errors was 5.5 (IQR 4-7) and the median number of color discrimination errors was 4 (IQR 3-5). The median duration of pausing in the non-congruent condition was 15.9 seconds (IQR 9.5-21.1 seconds). Acoustic measures differed significantly between the conditions, with CPPS and STSD significantly lower and L/H ratio higher in the non-congruent task compared to the congruent task (Table 1). There was no significant difference in mean f_0 or SD between the conditions.

Table 5. Acoustic measures in congruent vs. non-congruent Stroop sentence conditions

Variable	Congruent condition (mean, SD)	Non-congruent condition (mean, SD)	P value
Mean f_0	147.33 (32.91)	148.49 (33.25)	0.09
f_0 SD	16.58 (7.02)	15.29 (6.01)	0.48
STSD	1.82 (SD 0.49)	1.67 (0.40)	0.01
L/H ratio	30.46 (3.99)	31.38 (4.13)	0.01
CPPS	5.35 (1.23)	4.96 (1.06)	<.001
Sum duration of pauses	4.83 (4.71)	15.91 (8.12)	<.001

Correlations between Stroop sentence variables and clinical and neuropsychological variables

There were no significant associations between the Stroop sentence variables and either motor severity (MDS-UPDRS Part III) or global cognitive function (MoCA) (data not shown). To explore which cognitive domains impacted Stroop sentence performance, we assessed correlations between Stroop sentence variables and each neuropsychological test score. There were significant Spearman's rho correlations between measures of Stroop performance on the non-congruent condition and certain neuropsychological tests. Total errors were significantly correlated with visual and verbal memory (HVLT, rho= -0.434, p=0.027 and BVMT, rho=-0.474, p=0.014), executive function tasks including Trails B-A (rho=0.430, p=0.028) and Digit symbol modalities (rho=-0.568, p=0.0025) and category-based verbal fluency (vegetables, rho=-0.403, p=0.041). Color discrimination errors were significantly associated with visuospatial function on the JoLo (rho=-0.458, p=0.019).

Differences in Stroop sentence performance by cognitive diagnosis

There were no statistically significant differences by cognitive diagnosis group (MCI vs. normal cognition) in any of the Stroop sentence performance variables (congruent, non-congruent, and difference between conditions).

Discussion:

In our administration of this novel Stroop sentence task in the PD population, we found that participants had difficulty with the non-congruent condition of the Stroop sentences. The increased cognitive demand of the non-congruent condition led to increased number of errors, and increased pausing between words in the Stroop sentences, and significant alterations in voice quality and pitch. Color-discrimination errors made up the majority of errors, suggesting that participants had difficulty with color hue discrimination.

While we had hypothesized that the non-congruent condition performance relative to congruent condition would be associated with cognitive status, our results did not support this. Performance in the PD-MCI group compared to the PD-NC group was not significantly different using any of our acoustic, linguistic, or accuracy measures. Using a continuous screening measure of global cognition function (MoCA) there were also no significant correlations with any of the Stroop sentence measures. Our interpretation is that all PD participants struggled with the non-congruent condition and a floor effect limited our power to detect between-group differences. Future studies using larger sample size may show such differences. Alternatively, subtle executive function impairments may be present in individuals with early PD who test within the normal cognition range, [55] and thus inhibition of interfering stimuli may be already impaired well before MCI develops.

Consistent with our findings, other novel quantitative approaches to modifying traditional neuropsychological tasks have shown that deficits can be demonstrated in very early PD. Recently, a digitized version of the clock drawing test showed impairments in participants with early PD and normal global cognitive function compared with healthy controls.[56] While we did not include healthy controls yet in these analyses, this is an aim of future work. Another potential explanation for our results is that color discrimination impairments underlie the errors in the non-congruent Stroop task more than cognitive status. We found that a substantial proportion of the errors made in the non-congruent task could be potentially related to impaired color discrimination, although we did not separately assess hue identification abilities to enable us to make this determination with certainty.

We noted a large proportion of color discrimination errors. These errors may relate to inadequate clarify between hues with the color printer used. However, individuals with PD are known to have impaired color discrimination, which may result from either cortical/subcortical visual processing pathways, retinal pathology, or both. [57,58] If all PD participants regardless of MCI status had color discrimination difficulties, this would make the non-congruent condition more challenging independent of cognitive load and may explain why no between-group differences could be identified. The significance of our findings, in context of other approaches explored in the literature, suggest that automated,

quantitative cognitive tests may be powerful in identifying early PD status, but not necessarily assessing cognitive function or distinguishing those with PD-MCI. Further work is needed to develop tasks that are easy to administer and can be scored automatically to detect PD-MCI.

CHAPTER VI: Conclusions and Future Directions

Development of speech markers for PD must take both motor and cognitive systems into account. Our work has shown that certain acoustic features are variable between individual speaking styles, but not associated with PD symptoms (e.g. creak), while others are related to PD status overall but do not distinguish those with MCI (e.g. CPPS). It may be useful to further develop analytical approaches specifically tailored to speech in PD, taking advantage of known physiologic properties such as motor instability and dual-task interference. We present some evidence that fundamental frequency variability may decline over the course of speaking in PD-MCI, but there was substantial between-participant variability that needs to be addressed by future analyses. As a next

step to better understand these speech patterns, we plan to analyze the results of additional experimental tasks. We will analyze fundamental frequency decline during a story recall task and in additional dual-task experiments. We will also develop models including fundamental frequency decline along with other acoustic features. We will assess additional standard acoustic properties (e.g. vowel articulatory index, relative fundamental frequency) and additional advanced acoustic features with the assistance of collaborators. Combining multiple acoustic markers, for instance using machine learning algorithms, may improve detection of subtle states such as PD-MCI. It should be noted that machine learning algorithms need to be carefully designed and not over-fitted to a particular PD cohort. It remains to be shown that these algorithms can be translationally applied across groups with a range of age, sex, ethnicity and primary language. We will improve the rigor of our approach by assessing the accuracy performance of our models separately in males and females with PD.

My goal is to identify the speech and language markers with the highest signal for PD-MCI association in this cohort. I will then develop an R01 proposal to evaluate the selected markers longitudinally in a larger cohort of PD-NC, PD-MCI and PDD patients.

1. Jankovic J, Tan EK. Parkinson’s disease: etiopathogenesis and treatment. *Journal of Neurology, Neurosurgery & Psychiatry.* 2020;91(8):795-808.
2. Wichmann T, Bergman H, DeLong MR. Basal ganglia, movement disorders and deep brain stimulation: advances made through non-human primate research. *J Neural Transm (Vienna).* 2018;125(3):419-430.
3. Aarsland D, Batzu L, Halliday GM, et al. Parkinson disease-associated cognitive impairment. *Nat Rev Dis Primers.* 2021;7(1):47.
4. Weintraub D, Simuni T, Caspell-Garcia C, et al. Cognitive performance and neuropsychiatric symptoms in early, untreated Parkinson's disease. *Mov Disord.* 2015;30(7):919-927.
5. Hely MA, Reid WG, Adena MA, Halliday GM, Morris JG. The Sydney multicenter study of Parkinson's disease: the inevitability of dementia at 20 years. *Mov Disord.* 2008;23(6):837-844.
6. Pedersen KF, Larsen JP, Tysnes OB, Alves G. Natural course of mild cognitive impairment in Parkinson disease: A 5-year population-based study. *Neurology.* 2017;88(8):767-774.
7. Geurtsen GJ, Hoogland J, Goldman JG, et al. Parkinson's disease mild cognitive impairment: application and validation of the criteria. *J Parkinsons Dis.* 2014;4(2):131-137.
8. Litvan I, Goldman JG, Troster AI, et al. Diagnostic criteria for mild cognitive impairment in Parkinson's disease: Movement Disorder Society Task Force guidelines. *Mov Disord.* 2012;27(3):349-356.
9. Miller N, Allcock L, Jones D, Noble E, Hildreth AJ, Burn DJ. Prevalence and pattern of perceived intelligibility changes in Parkinson's disease. *J Neurol Neurosurg Psychiatry.* 2007;78(11):1188-1190.
10. Canter GJ. SPEECH CHARACTERISTICS OF PATIENTS WITH PARKINSON'S DISEASE. II. PHYSIOLOGICAL SUPPORT FOR SPEECH. *J Speech Hear Disord.* 1965;30:44-49.
11. Lee J, Huber J, Jenkins J, Fredrick J. Language planning and pauses in story retell: Evidence from aging and Parkinson's disease. *J Commun Disord.* 2019;79:1-10.

12. Smith KM, Ash S, Xie SX, Grossman M. Evaluation of Linguistic Markers of Word-Finding Difficulty and Cognition in Parkinson's Disease. *J Speech Lang Hear Res.* 2018;61(7):1691-1699.
13. Darling-White M, Huber JE. The Impact of Parkinson's Disease on Breath Pauses and Their Relationship to Speech Impairment: A Longitudinal Study. *Am J Speech Lang Pathol.* 2020;29(4):1910-1922.
14. Smith KM, Caplan DN. Communication impairment in Parkinson's disease: Impact of motor and cognitive symptoms on speech and language. *Brain Lang.* 2018;185:38-46.
15. Arora S, Tsanas A. Assessing Parkinson's Disease at Scale Using Telephone-Recorded Speech: Insights from the Parkinson's Voice Initiative. *Diagnostics.* 2021;11(10):1892.
16. Hlavnička J, Čmejla R, Tykalová T, Šonka K, Růžička E, Rusz J. Automated analysis of connected speech reveals early biomarkers of Parkinson's disease in patients with rapid eye movement sleep behaviour disorder. *Sci Rep.* 2017;7(1):12.
17. Rusz J, Hlavnička J, Novotný M, et al. Speech Biomarkers in Rapid Eye Movement Sleep Behavior Disorder and Parkinson Disease. *Ann Neurol.* 2021;90(1):62-75.
18. Rektorova I, Mekyska J, Janousova E, et al. Speech prosody impairment predicts cognitive decline in Parkinson's disease. *Parkinsonism Relat Disord.* 2016;29:90-95.
19. Rusz J, Tykalova T. Does Cognitive Impairment Influence Motor Speech Performance in De Novo Parkinson's Disease? *Mov Disord.* 2021;36(12):2980-2982.
20. Garcia AM, Arias-Vergara T, J CV-C, et al. Cognitive Determinants of Dysarthria in Parkinson's Disease: An Automated Machine Learning Approach. *Mov Disord.* 2021;36(12):2862-2873.
21. Berardelli A, Rothwell JC, Thompson PD, Hallett M. Pathophysiology of bradykinesia in Parkinson's disease. *Brain.* 2001;124(Pt 11):2131-2146.
22. Wu T, Hallett M, Chan P. Motor automaticity in Parkinson's disease. *Neurobiol Dis.* 2015;82:226-234.
23. Goberman AM, Coelho C. Acoustic analysis of parkinsonian speech I: speech characteristics and L-Dopa therapy. *NeuroRehabilitation.* 2002;17(3):237-246.
24. Montero-Odasso MM, Sarquis-Adamson Y, Speechley M, et al. Association of Dual-Task Gait With Incident Dementia in Mild Cognitive Impairment: Results From the Gait and Brain Study. *JAMA Neurol.* 2017;74(7):857-865.
25. Raffegeau TE, Krehbiel LM, Kang N, et al. A meta-analysis: Parkinson's disease and dual-task walking. *Parkinsonism Relat Disord.* 2019;62:28-35.
26. Whitfield JA, Kriegel Z, Fullenkamp AM, Mehta DD. Effects of Concurrent Manual Task Performance on Connected Speech Acoustics in Individuals With Parkinson Disease. *Journal of Speech, Language, and Hearing Research.* 2019;62(7):2099-2117.
27. Whitfield JA, Kriegel Z, Fullenkamp AM, Mehta DD. Effects of Concurrent Manual Task Performance on Connected Speech Acoustics in Individuals With Parkinson Disease. *J Speech Lang Hear Res.* 2019;62(7):2099-2117.
28. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry.* 1992;55(3):181-184.
29. Litvan I, Goldman JG, Tröster AI, et al. Diagnostic criteria for mild cognitive impairment in Parkinson's disease: Movement Disorder Society Task Force guidelines. *Movement Disorders.* 2012;27(3):349-356.

30. Georgiou GP. Acoustic markers of vowels produced with different types of face masks. *Appl Acoust.* 2022;191:108691.
31. Zhou P, Zong S, Xi X, Xiao H. Effect of wearing personal protective equipment on acoustic characteristics and speech perception during COVID-19. *Appl Acoust.* 2022;197:108940.
32. Fairbanks G. *Voice and articulation drillbook.* 2 ed. New York: Harper & Row; 1960.
33. Goodglass H, Kaplan, E., Barresi, B. *Boston Diagnostic Aphasia Examination.* 3 ed. San Antonio: Pearson; 2000.
34. *Praat: doing phonetics by computer* [computer program]. Version 6.0.562009.
35. Bowen LK, Hands GL, Pradhan S, Stepp CE. Effects of Parkinson's Disease on Fundamental Frequency Variability in Running Speech. *J Med Speech Lang Pathol.* 2013;21(3):235-244.
36. Abdelli-Beruh NB, Drugman T, Red Owl RH. Occurrence Frequencies of Acoustic Patterns of Vocal Fry in American English Speakers. *J Voice.* 2016;30(6):759 e711-759 e720.
37. Murton O, Shattuck-Hufnagel S, Choi JY, Mehta DD. Identifying a creak probability threshold for an irregular pitch period detection algorithm. *J Acoust Soc Am.* 2019;145(5):EL379.
38. Alharbi GG, Cannito MP, Buder EH, Awan SN. Spectral/Cepstral Analyses of Phonation in Parkinson's Disease before and after Voice Treatment: A Preliminary Study. *Folia Phoniatrica et Logopaedica.* 2019;71(5-6):275-285.
39. Ligon C, Rountrey C, Rank NV, Hull M, Khidr A. Perceived Desirability of Vocal Fry Among Female Speech Communication Disorders Graduate Students. *Journal of Voice.* 2019;33(5):805.e821-805.e835.
40. Šimek M, Rusz J. Validation of cepstral peak prominence in assessing early voice changes of Parkinson's disease: Effect of speaking task and ambient noise. *J Acoust Soc Am.* 2021;150(6):4522.
41. Jiménez-Jiménez FJ, Gamboa J, Nieto A, et al. Acoustic voice analysis in untreated patients with Parkinson's disease. *Parkinsonism Relat Disord.* 1997;3(2):111-116.
42. Laukkanen AM, Ilomäki I, Leppänen K, Vilkman E. Acoustic measures and self-reports of vocal fatigue by female teachers. *J Voice.* 2008;22(3):283-289.
43. Skodda S, Grönheit W, Schlegel U. Intonation and speech rate in Parkinson's disease: general and dynamic aspects and responsiveness to levodopa admission. *J Voice.* 2011;25(4):e199-205.
44. Remacle A, Morsomme D, Berrué E, Finck C. Vocal impact of a prolonged reading task in dysphonic versus normophonic female teachers. *J Voice.* 2012;26(6):820.e821-813.
45. Xue C, Kang J, Hedberg C, Zhang Y, Jiang JJ. Dynamically Monitoring Vocal Fatigue and Recovery Using Aerodynamic, Acoustic, and Subjective Self-Rating Measurements. *J Voice.* 2019;33(5):809.e811-809.e818.
46. Liu H, Wang EQ, Metman LV, Larson CR. Vocal responses to perturbations in voice auditory feedback in individuals with Parkinson's disease. *PLoS One.* 2012;7(3):e33629.
47. Mollaei F, Shiller DM, Gracco VL. Sensorimotor adaptation of speech in Parkinson's disease. *Mov Disord.* 2013;28(12):1668-1674.
48. Abur D, Lester-Smith RA, Daliri A, Lupiani AA, Guenther FH, Stepp CE. Sensorimotor adaptation of voice fundamental frequency in Parkinson's disease. *PLoS One.* 2018;13(1):e0191839.

49. Darling-White M, Huber JE. The Impact of Parkinson's Disease on Breath Pauses and Their Relationship to Speech Impairment: A Longitudinal Study. *American Journal of Speech-Language Pathology.* 2020;29(4):1910-1922.
50. Stroop J. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology.* 1935;18:643-662.
51. Golden C. *Stroop Color and Word Test: A Manual for Clinical and Experimental Uses.* Chicago, IL: Stoelting Co.; 1978.
52. Scarpina F, Tagini S. The Stroop Color and Word Test. *Front Psychol.* 2017;8:557.
53. Hsieh YH, Chen KJ, Wang CC, Lai CL. Cognitive and motor components of response speed in the stroop test in Parkinson's disease patients. *Kaohsiung J Med Sci.* 2008;24(4):197-203.
54. Dahl KL, Stepp CE. Changes in Relative Fundamental Frequency Under Increased Cognitive Load in Individuals With Healthy Voices. *J Speech Lang Hear Res.* 2021;64(4):1189-1196.
55. Kudlicka A, Clare L, Hindle JV. Executive functions in Parkinson's disease: Systematic review and meta-analysis. *Movement Disorders.* 2011;26(13):2305-2315.
56. Schejter-Margalit T, Kizony R, Shirvan J, et al. Quantitative digital clock drawing test as a sensitive tool to detect subtle cognitive impairments in early stage Parkinson's disease. *Parkinsonism Relat Disord.* 2021;90:84-89.
57. Bertrand JA, Bedetti C, Postuma RB, et al. Color discrimination deficits in Parkinson's disease are related to cognitive impairment and white-matter alterations. *Mov Disord.* 2012;27(14):1781-1788.
58. Zhang Y, Zhang X, Yue Y, Tian T. Retinal Degeneration: A Window to Understand the Origin and Progression of Parkinson's Disease? *Front Neurosci.* 2021;15:799526.