Tampere University

Miia Ketolainen

THE BEST SPEECH TASK FOR DIFFERENTIATING BETWEEN PARKINSONIAN AND HEALTHY SPEECH

Integrative literature review

Faculty of Social Sciences
Bachelor's thesis
April 2026

ABSTRACT

Miia Ketolainen: The best speech task for differentiating between Parkinsonian and healthy speech
Bachelor's thesis
Tampere University
Bachelor's Programme in Logopedics
April 2026

The aim of this study was to determine which speech task is the best for automatically identifying Parkinson's disease based on speech signal. The second aim was to examine whether different languages differ in which speech task is best used to identify Parkinson's disease. The study was conducted as an integrative literature review. Seven peer-reviewed scientific research articles were selected for the data based on predefined inclusion and exclusion criteria. The speech tasks, languages, classification algorithms, and results were extracted from the articles selected for the review. The results were examined both collectively for all languages studied and separately classified by language.

The articles selected for the review examined a total of six different speech tasks: prolonged vowel, rapid syllable repetition, reading single words, reading single sentences, reading text, and spontaneous speech. The results suggest that Parkinson's disease can best be identified based on spontaneous speech. The articles examined speech signals recorded from speakers of six languages: Spanish, Finnish, German, Czech, Lithuanian and English. Six of the seven articles examined Spanish speakers. The other languages were examined in one article each. Based on the results, the best speech task for Spanish and English speakers to identify Parkinson's disease was spontaneous speech. In contrast, the best results were obtained with prolonged vowel for Lithuanian speakers, with reading text for Finnish speakers, and for German and Czech speakers with rapid syllable repetition. It is noteworthy, however, that spontaneous speech was not examined in the data for Lithuanian, German and Czech speakers.

The results of the study can be utilized for the development of new cost-effective methods for diagnosing Parkinson's disease at an earlier stage.

Keywords: Parkinson's disease, hypokinetic dysarthria, speech task, automatic detection

TIIVISTELMÄ

Miia Ketolainen: Paras puhetehtävä Parkinsonin taudin erottamiseen puheesta
Kandidaatin tutkielma
Tampereen yliopisto
Logopedian tutkinto-ohjelma
Huhtikuu 2026

Tämän tutkielman tavoitteena oli selvittää, mikä on paras puhetehtävä Parkinsonin taudin tunnistamiseen automaattisesti puhesignaalin perusteella. Toisena tavoitteena oli tarkastella, onko eroavatko eri kielet siinä, millä puhetehtävällä Parkinsonin tauti tunnistetaan parhaiten. Tutkielma toteutettiin integroivana kirjallisuuskatsauksena. Aineistoon valikoitui seitsemän vertaisarvioitua tieteellistä tutkimusartikkelia ennalta määritettyjen sisäänotto- ja poissulkukriteerien perusteella. Katsaukseen valituista artikkeleista poimittiin niissä käytetyt puhetehtävät, kielet, luokittelualgoritmit sekä tulokset. Tuloksia tarkasteltiin sekä yhteisesti kaikkien tutkittujen kielten osalta sekä luokiteltuna kielen mukaan.

Katsaukseen valikoituneissa artikkeleissa tarkasteltiin yhteensä kuutta erilaista puhetehtävää: pitkää vokaaliääntöä, nopeaa tavujen toistamista, yksittäisten sanojen lukemista, yksittäisten lauseiden lukemista, tekstin lukemista sekä spontaania puhetta. Tulokset viittaavat siihen, että Parkinsonin tauti voidaan parhaiten tunnistaa spontaanin puheen perusteella. Artikkeleissa tarkasteltiin kuuden eri kielen, espanjan, suomen, saksan, tšekin, liettuan ja englannin kielten puhujilta nauhoitettua puhesignaalia. Kuudessa seitsemästä artikkelista tutkittiin espanjankielisiä puhujia. Muita kieliä tarkasteltiin kutakin yhdessä artikkelissa. Tulosten perusteella espanjan- ja englanninkielisillä puhujilla paras puhetehtävä Parkinsonin taudin tunnistamiseen oli spontaani puhe. Sen sijaan liettuankielisillä puhujilla paras tulos saatiin pitkällä vokaaliäännöllä, suomenkielisillä puhujilla tekstin lukemisella, saksan- ja tšekinkielisillä puhujilla nopealla tavujen toistamisella. Huomattavaa kuitenkin on, että liettuan-, saksan- ja tšekinkielisillä puhujilla aineistossa ei tarkasteltu spontaania puhetta.

Tutkielman tuloksia voidaan hyödyntää uusien kustannustehokkaiden menetelmien kehittämiseen Parkinsonin taudin diagnosoimiseksi varhaisemmassa vaiheessa.

Avainsanat: Parkinsonin tauti, hypokineettinen dysartria, puhetehtävä, automaattinen tunnistaminen

USE OF AI IN THESIS

I have utilised AI tools in my thesis:

☐ No
☒ Yes

The AI tools utilised in my thesis and their purposes are described below:

Names and versions of AI tools:
ChatGPT, GPT-4
Microsoft Copilot

Purpose of using AI tools:
ChatGPT was used for finding synonyms for search terms and for formatting the search phrase.

ChatGPT was used for translating unfamiliar words, explaining unfamiliar terms and making sure unclear sentences in articles were understood correctly.

Sections where AI tools were used:
ChatGPT was used in preparation of literature search process (section 4.2).
ChatGPT was used in analysis of the data by translating words and sentences and explaining terms (section 4.3).
Microsoft Copilot was used in formatting table 3 (section 4.3)

I acknowledge that I am fully responsible for the entire content of my thesis, including the parts generated by AI, and accept accountability for any violations of ethical standards in publications.

CONTENTS

1 INTRODUCTION

Parkinson's disease (PD) is a progressive neurologic disease with symptoms such as tremor, muscle stiffness and slowing down of movement. About 1 % of over 60-year-old people have PD (Atula, 2023).

PD is most commonly diagnosed in the age of 50-70 years (Atula, 2023). It is important to diagnose it in the early phase and speech can be utilized in diagnosing. Several different speech tasks have been investigated in differentiating between Parkinsonian and healthy voice, such as utterances of long vowels, spontaneous monologue and passage reading (Liu et al., 2023). Kodali and colleagues (2024) compared five speech tasks, four features of speech and four automatic classifiers. Vásquez-Correa and colleagues (2018) used sustained phonation of a vowel, sentence reading, diadochokinesia, text reading and monologue task.

PD affects voice by making movements of vocal tract slower and weaker and speech in PD is classified as hypokinetic dysarthria. According to Miller (2012), changes in PD can be detected in several features of speech and voice. First, due to changes in muscles controlling respiration, voice becomes more silent and monotonous, weaker, breathier and hoarser. Second, articulation is affected by PD. Production of speech sounds is impaired and different speech sounds may have only a little difference, making speech slurred and indistinct. Third, prosodic features, such as tone, stress and rhythm are affected, and monopitch, monoloudness and irregularities are typical of Parkinsonian speech.

Speech in PD has also been researched in several languages. Liu and colleagues (2023) analyzed speech of Finnish speakers and Orozco-Arroyave and colleagues (2016) used Spanish, German and Czech speech recordings in their analysis. Different languages have differences, for example in phonemes and prosodic features, so it is possible that the same speech task is not suitable in differentiating Parkinsonian and healthy speech universally.

2 THEORY

2.1 Parkinson's disease and diagnosing

PD is a neurodegenerative disease characterized by a progressive loss of dopaminergic neurons in the substantia nigra (Albin, 2023). The cause of degeneration is unknown (Atula, 2023). The degeneration results in lack of neurotransmitter dopamine which leads to difficulties in regulating voluntary muscle movements. Typical symptoms include bradykinesia, rigidity and tremor.

Rizzo and colleagues (2016) conducted a systematic review and reported that the accuracy of clinical diagnosis of PD varied between 73,8% and 83,9% depending on diagnostic criteria and the group of experts performing the diagnosis. Accuracy did not significantly improve during the 25 years that the articles selected for their review covered. Therefore, new methods for diagnosing PD are needed. Tolosa and others (2021) present the current status of available and potential methods for diagnosing PD. The potential methods include magnetic resonance imaging, radiotracer imaging, tissue biopsies, biomarkers in cerebrospinal fluid, defecation, and digital biomarkers.

There are several tools for assessing and evaluating PD. United Kingdom PD Society Brain Research Center's diagnostic clinical criteria are described by Hughes and others (1992). First, a patient has to have bradykinesia and at least one of the following symptoms: muscle rigidity, 4-6 Hz rest tremor or postural instability which is not explained by primary visual, vestibular, cerebellar, or proprioceptive deficits. Additionally, other potential causes must be excluded.

Movement Disorder Society's Unified Parkinson's Disease Rating Scale (MDS-UPDRS) is an updated revision of UPDRS originally published in 1980s (Goetz et al., 2008). MDS-UPDRS is a tool used for evaluating the severity of symptoms of PD. It has four parts, which are targeted to assess non-motor experiences of daily living, motor experiences of daily living, motor examination and motor complications.

The Hoehn & Yahr scale (H&Y) is a five-point scale originally published in 1967 (Hoehn & Yahr, 2001). The H&Y scale is a widely used clinical tool for describing the progression and severity of PD across five stages. In Stage 1, symptoms are typically unilateral and associated with minimal or no functional disability. Stage 2 is characterized by bilateral involvement, although balance remains largely unaffected. In Stage 3, the degree of disability ranges from mild to moderate, with patients potentially exhibiting impaired reflexes and increased instability during movement. Stage 4 denotes severe disability;

however, individuals are generally still able to move independently. By Stage 5, patients are typically bedridden or dependent on a wheelchair for mobility. It was later updated to contain also stages 1.5 and 2.5 (Goetz et el., 2004).

2.2 Speech in Parkinson's disease

Symptoms of PD may occur in the muscles involved in speech production. The motor speech disorder associated with PD is called hypokinetic dysarthria (Duffy, 2020). Typical speech symptoms include reduced range and force of movements of articulatory muscles and increased muscle tone. These symptoms affect phonation, articulation, prosody and intelligibility of speech. Phonatory insufficiencies observed in hypokinetic dysarthria are low pitch, breathiness, harsh voice quality and monoloudness. Articulatory deficiencies include imprecise consonants, short rushes of speech, variable rate and repeated phonemes. Prosodic features in hypokinetic dysarthria are monopitch, reduced stress, inappropriate silences, increased rate in segments and increased rate overall.

Frenchay Dysarthria Assessment (FDA) is a tool developed for evaluating dysarthria and defining the type of dysarthria (Enderby, 1980). It is divided in nine sections that evaluate reflexes, respiration, movements of lips, jaws, palate, larynx and tongue, intelligibility of speech and speech rate. The modified version of FDA, m-FDA, is introduced by Vásquez-Correa and colleagues (2018). It does not require a visit to the examiner and can be administered based on speech recordings.

Impact of PD in different speech features has been studied by Moro-Velazquez and colleagues (2021), who reviewed papers focusing on phonatory and articulatory aspects of speech. They present a comprehensive review on the current status of research on automatic assessment of PD. Based on their findings, earlier studies suggest that PD influences patients' phonation more than articulation, but more recent studies show that articulation provides more information for differentiating between healthy and Parkinsonian speakers. According to Moro-Velazquez and others (2021), the accuracy in phonatory approaches varies between 75 % and 90 %, while in articulatory approaches the range is from 80 % to 95 %.

According to Ho and others (1998) voice impairment occurs earlier than fluency and articulatory impairment within PD patients. They recorded speech samples of 200 PD patients and classified them into five severity levels. Their findings show that 94,5 % of mildly impaired patients had voice deficits. Fluency and articulatory deficits occurred less frequently in mildly impaired patients but became more frequent in moderately and

severely impaired patients. The result suggests that voice-based biomarkers can help detect PD in early stage.

Fumel and colleagues (2024) studies which prosodics features were impaired in PD. They found that lower fundamental frequency (F0) variability and longer pauses occurred frequently in Parkinsonian speech. There was some evidence that speech rate and intensity variability are also reduced. Articulation rate and pause ratio were not found to be affected by PD.

Penttilä and others (2022) investigated prosodic features in Finnish adults with PD. According to their findings prosody of PD patients' speech differs from that of healthy controls (HC) in multiple ways. First, male PD patients had higher pitch and syllabic prosody index than male HC speakers. Second, female PD patients had less variability in pitch than HC. Third, energy distribution was shifted towards lower frequencies in PD speakers than in HC speakers.

2.3 Speech tasks

According to Kempler and Lancker (2002) speech task has effect on the intelligibility of Parkinsonian speech. They recorded five speech tasks from one patient with PD using the same phrases and 64 listeners listened to recordings once and wrote down what they heard. The study showed that utterances extracted from spontaneous speech were less intelligible than utterances extracted from reading, repeated spoken, repeated sung and spontaneous sung recordings, which did not have significant differences in intelligibility.

Different speech tasks are used in evaluating different aspects of speech. Sustained vowels, sentence reading, passage reading, spontaneous speech, diadochokinesia (rapid repetition of syllables) task and isolated words are common speech tasks used in research. In sustained vowel task a participant is typically instructed to pronounce a vowel as long as they can. The vowel /a/ is most typically used in the vowel task, such as in Almeida et al (2019), but also /e/, /i/, /o/ and /u/ are used, e.g. in Kodali et al. (2024). The sustained vowel task is used to extract phonation-related parameters from speech, such as fundamental frequency, jitter, shimmer, Harmonic to Noise ratio (HNR) and Noise to Harmonics ratio (NHR) (Moro-Velazquez et al, 2021). The sustained vowel task is referred as VOWEL in this literature review.

Passage reading task typically consists of a predefined paragraph of text and in some research articles it is referred to as text reading or paragraph reading task. Different texts are used in different research articles and therefore they are not fully comparable. For

example, in Liu et al (2023), the text known as ‘The Northwind and The Sun’ is used for Finnish speakers. The passage reading task is referred as TEXT in this literature review.

Sentence reading task includes different sentences and similar to TEXT task, sentences differ between languages that the studies are conducted in. For example, in Orozco-Arroyave et al. (2016) German, Czech and Spanish speakers read different numbers of sentences, and all three sentences in Czech were questions while none of the German and Spanish were questions. The sentence reading task is referred to as SENTENCE in the rest of the literature review.

Spontaneous speech task is in some research articles referred as monologue task (Kodali et al., 2023; Liu et al., 2023; Vásquez-Correa et al., 2018; Vásquez-Correa et al., 2020). In spontaneous speech task a participant is instructed to tell about something with their own words. Questions or topics are often used to help participants to complete the task and produce long enough speech recording. The spontaneous speech task or monologue task is referred to as SPON in the coming sections of this thesis.

Diadochokinesia task is referred to as DDK in most research articles and the abbreviation is used in this literature review as well. The DDK task is often referred to as syllable repetition task as a participant is instructed to repeat syllables /pa/, /ta/ and /ka/ and possibly some variations and combinations of them. The DDK task is used for measuring articulatory aspects of speech and how a speaker is able to move their articulatory organs.

In isolated words task participants are typically instructed to pronounce predefined words. Similar to TEXT and SENTENCE tasks, words differ from language to language. Typically words are chosen to demonstrate transitions from one phoneme to another. The task is referred to as WORD in the following chapters.

TEXT, SENTENCE and SPON tasks are used for recording connected speech and can be used for extracting articulatory parameters, such as VAI, tVSA, Mel-Frequency Cepstral Coefficents (MFCC) and Voice Onset Time (VOT).

2.4 Automatic detection of Parkinson’s disease

Jiménez-Monsalve and others (2017) studied phonation and articulation measures in three different pathologies, one of which was PD. They considered sustained vowel, sentence reading and text reading tasks from the Spanish dataset PC-GITA. Articulation and phonation features were extracted from speech recordings and they were classified using support vector machine (SVM) algorithm. The results suggested that articulation features extracted from sentence and text reading tasks were more accurate at detecting

PD than sustained vowels. The combination of ten sentences and text reading gave even better results.

Moro-Velazquez and colleagues (2019) investigated articulatory and phonatory aspects of speech in two different corpora, one of which consisted of Colombian Spanish and the other of Castilian Spanish speech recordings. They used sustained vowel /a/ and six text dependent utterances from both datasets. Phonatory features were extracted from sustained vowel and articulatory features were extracted from the text dependent utterances. The results of the analysis showed that articulatory features provided better accuracy than phonatory features in detecting PD.

Pompili and others (2017) used Portuguese dataset to investigate detection of PD. The dataset contained recordings of sustained vowel /a/, maximum phonation of /a/, DDK, word, sentence and text reading tasks, story telling speech guided by visual stimuli, and reading prosodic sentences. According to the results, the highest accuracy in detecting PD was obtained with reading of prosodic sentences, followed by storytelling guided by visual stimuli. The lowest accuracy was achieved with the vowel tasks. This suggests that natural speech like tasks provide the most accurate information for detecting PD.

3 RESEARCH QUESTIONS

The purpose of this thesis is to find out what is the best speech task for differentiating between parkinsonian and healthy voice. There is no single definitive diagnostic procedure or laboratory test for establishing a diagnosis of PD. Therefore, clinicians rely on a combination of diagnostic approaches. These typically include clinical examinations using standardized assessment scales such as mentioned before the MDS-UPDRS and the Hoehn and Yahr (H&Y) staging scale, as well as laboratory tests and neuroimaging techniques. However, these methods can be time-consuming and expensive and it is important to develop new methods for diagnosing PD in early stage and voice- and speech-based biomarkers have recently attracted the attention of researchers. Speech-based biomarkers may provide a cost-efficient and easy-to-use alternative for current ways to diagnose PD. Therefore, it is important to investigate what speech task provides the most accurate and sensitive information for diagnosis. However, languages differ from each other, for example in articulatory and prosodic features, so the same speech task may not be suitable for speakers of different languages. Thus, it is also important to consider if the different speech tasks provide different information in different languages.

The research questions are:

1. What is the best speech task for distinguishing PD based on speech?
2. Are there differences between languages in what is the best speech task?

4 RESEARCH METHODS

4.1 Literature review as a research method

Literature review is a research method in which the research material consists of previous research (Vilkka, 2023). The aim of the literature review is to describe the current status of the research of the topic, what is already known and what could be investigated further. Main types of literature review are descriptive literature review, systematic literature review and meta-analysis. Descriptive literature reviews can be further divided into narrative and integrative literature reviews. Regardless of the type of review, the method can be described as disciplined, critical and systematic. Discipline means that the process of material search is executed based on the predefined rules consistently and transparently and is also described so that the research can be repeated. Criticality means that all the phases of the research process are reflected by the researcher and all the decisions during the process are considered based on the research questions. Evaluation of the quality of the articles and inclusion or exclusion is also executed critically. Systematicity means that the literature review follows a certain predefined protocol, which defines the phases of the process.

The literature review as a method consists of four phases, search, appraisal, synthesis and analysis (Booth et al., 2021). In the search phase the researcher forms a search phrase and chooses the databases in which the search is then executed (Kangasniemi et al., 2013). In the appraisal phase the quality and suitability of articles are evaluated and the articles for review are chosen based on the predefined inclusion and exclusion criteria. In the synthesis phase the information provided in the chosen articles is gathered together for constructing an overall picture of the topic, results of the articles can be compared and classified as groups, differences and conflict between them can be described. The analysis phase consists of observation of the results. It includes critical discussion of the methods and results and provides further research questions.

This literature review is implemented as an integrative literature review. The integrative literature review is a subtype of the descriptive literature review and it differs from the narrative literature review by aiming to combine existing information to create new approaches and practices, while the aim of the narrative literature review is to gradually broaden the knowledge (Vilkka, 2023). The integrative review resembles the systematic review in the systematic search process but the aim of the systematic review is to find an answer to a specific research question.

4.2 Implementation of the literature review

Seven articles were selected for this literature review. The literature search was executed in February 2025. Only peer-reviewed academic articles published between 2015-2025 were included in the search. The search phrase was:

("parkinson's disease" OR parkinsonism OR "hypokinetic dysarthria") AND ("speech task*" OR "speaking task*") AND automatic AND (assessment OR classification OR evaluation OR detection).

The databases used in this literature review were EBSCO (Academic Search Ultimate, CINAHL Complete, Communication & Mass Media Complete, MEDLINE, MLA International Bibliography with Full Text), IEEE Xplore – IEEE Electronic Library, ProQuest (Linguistics Collection), PubMed and Scopus. The number or articles found in each database is reported in Table 1.

Table 1.

Number of articles found in each database.

Database	Search results, title and abstract read
EBSCO (Academic Search Ultimate, CINAHL Complete, Communication & Mass Media Complete, MEDLINE, MLA International Bibliography with Full Text)	15
IEEE Xplore – IEEE Electronic Library	13
ProQuest (Linguistics Collection)	46
PubMed	11
Scopus	313
Total	**398**

The search results were imported to Covidence in RIS format. Covidence is a web-based collaboration software platform that streamlines the production of systematic and other literature reviews. Covidence automatically removed 41 duplicates from the search results. In total 357 articles were screened in Title and abstract screening phase, of which 285 were deemed irrelevant. In Full text review phase 72 articles were read, and seven articles were selected for the literature review. The inclusion criteria were: Participants included both PD patients and HC speakers, at least two speech tasks were used and the results for speech tasks were reported separately. Articles published in academic journals with JuFo classification 1 or lower and articles with no full text available were excluded.

4.3 Analysis of the data

The selected articles were read several times, and the following information was gathered for each: databases used in the article, languages, speech recording method, number of PD patients, number of HC speakers, speech tasks, data preprocessing methods, algorithms used for classification, measures, relevant results. The speech tasks used in the articles and a short description of each are presented in Table 2.

Table 2.

Overview of the speech tasks.

Speech task	Description
VOWEL	Prolonged vowel
DDK	Syllable repetition.
WORD	Reading a single word
SENTENCE	Reading a single sentence
READ	Reading a text or paragraph
SPON	Spontaneous speech or monologue

The gathered information was organized in a table and analyzed based on the research questions. The gathered information is presented in Table 3.

Table 3.

Information gathered from the articles.

Article	Database	Information about databases	Data preprocessing	Participants	Algorithms	Languages	Measures	Speech tasks	Results
Almeida et al., 2019	Lithuanian database with no name mentioned in the article	Recorded using two channels from acoustic cardioid (AKG Perception 220, frequency range 20-20000 Hz) and a smartphone (Samsung Galaxy Note 3). About 10 cm from participant's mouth. Audio format: mono PCM wav (16 bits at 44,1 kHz sampling rate)	Audio speech was split using Praat software toolbox, generating voiced and unvoiced parts. Features were extracted, generating in total 144 files	HC: mean age 41,8 years, 11 male, 24 female. PD patients: mean age 61,5 years, 30 male, 34 female. Hoehn and Yahr stage 1-2,5 (early-to-mild stage of the disease)	kNN, Multi-Layer Perceptron (MLP), Optimum-path Forest (OPF), SVM	Lithuanian	Accuracy + AUC	VOWEL (a)	AC + YA + kNN: 94,55 ± 3,33 % + 0,87 SP + KT + kNN: 92,94 ± 2,55 % + 0,92
								SENTENCE	AC + IS4 + SVM: 72,65 ± 4,76 % + 0,84 SP + KT + SVM: 71,18 ± 4,52 % + 0,87
Kodali et al., 2024	PC-GITA	Original sampling frequency was 44,1 kHz, down-sampled to16 kHz for this study. Speakers were divided to three severity levels, healthy, mild, severe based on MDS-UPDRS-III and -S scales. The speech samples of patients were recorded less than 3 hours after their morning medication.	Phonation, articulation, prosody and their combination features were extracted. Phonation features (vocal fold vibration's temporal and amplitudinal variation) were extracted from voiced segments. Articulation features were derived using spectral measurements and speech energy at onset/offset transitions. Fundamental frequency contours and energy were used to derive	50 PD, 50 HC.	SVM, RF, MLP, AdaBoost	Spanish	Accuracy	VOWEL (a,e,i,o, u)	MDS-UPDRS-III + Fusion + RF: 48 ± 1 % MDS-UPDRS-III + Articulation + MLP: 48 ± 4 % MDS-UPDRS-S + Fusion + RF: 55 ± 3 % MDS-UPDRS-S + Fusion + AdaBoost: 55 ± 2 %
								SENTENCE	MDS-UPDRS-III + Fusion + SVM: 47 ± 4 %. MDS-UPDRS-III + Articulation + RF: 47 ± 5 %. MDS-UPDRS-III + Articulation + AdaBoost: 47 ± 5 % MDS-UPDRS-S + Fusion + MLP: 49 ± 2 %.
								DDK	MDS-UPDRS-III + Articulation + SVM: 50 ± 5 % MDS-UPDRS-S + Articulation + RF: 51 ± 7 %
								READ	MDS-UPDRS-III + Prosody + MLP: 53 ± 8 % MDS-UPDRS-S + Phonation + SVM: 50 ± 4 % MDS-UPDRS-S + Phonation + MLP: 50 ± 3 %
								SPON	MDS-UPDRS-III + Articulation + MLP: 58 ± 6 % MDS-UPDRS-S + Fusion + MLP: 56 ± 4 %

Article	Database	Information about databases	Data preprocessing	Participants	Algorithms	Languages	Measures	Speech tasks	Results
			prosody features. DisVoice framework was used to extract all features.						
Liu et al., 2023	PDSTU, PC-GITA	PDSTU: Speech recorded in mono channel with 32 bits, sampling rate 44,1 kHz with a close-talking microphone. Spanish recordings from PC-GITA.	Speech attribute scores (SAS), Glottal features, MFCCs and low-level descriptors in eGeMaps were extracted	PDSTU: HC: 9 female, 6 male, mean age 57,7 years. PD: 21 female 14 male, mean age 65,6 years. PC-GITA: 50 PD, 50 HC.	CNN	Finnish	PCC	VOWEL (a, e, i, o, u)	Intelligibility + SAS+Glottal: 0,72. Voice impairment + eGeMAPS: 0,63. Overall severity + SAS+Glottal: 0,62.
								READ	Intelligibility + SAS+Glottal: 0,73. Voice impairment + SAS/SAS+Glottal: 0,64. Overall severity + SAS+Glottal: 0,72.
								SPON	Intelligibility + SAS+Glottal: 0,71. Voice impairment + eGeMAPS: 0,57. Overall severity + SAS: 0,69.
						Spanish	PCC	VOWEL (a, e, i, o, u)	UPDRS + MFCC+eGeMAPS: 0,18
								READ	UPDRS + MFCC: 0,33
								SPON	UPDRS + SAS: 0,41
Reddy & Alku, 2023	MDVR-KCL, PC-GITA	English: MDVR-KCL. Recorded with Motorola Moto G4 smartphone. Sample rate 44,1 kHz, 16 bits. Spanish: PC-GITA	Recordings were downsampled to 16 kHz. The speech recordings in MDVR-KCL were preprocessed by removing long pauses and unwanted sounds and by segmenting the recordings into three-second segments.	PC-GITA: 50 PD speakers, HC speakers. MDVR-KCL: 16 PD speakers, 21 HC speakers.	SVM,RF, LSRC, NSRC, Proposed-LSRC, Proposed-NSRC	Spanish	Recall + Precision + F1-score + Accuracy + MCC	VOWEL (a, e, i, o, u)	IS10 + Proposed-NSRC: 77,07 + 70,42 + 73,59 + 73,52 ± 3,79 + 0,45 ± 0,09. IS10+gMFS + Proposed-NSRC: 74,00 + 77,08 + 75,51 + 76,00 ± 4,09 + 0,52 ± 0,10.
								SENTENCE	IS10 + Proposed-NSRC: 80,00 + 82,76 + 81,36 + 81,67 ± 4,22 + 0,63 ± 0,10. IS10+gMFS + Proposed-NSRC: 81,67 + 84,51 + 83,17 + 82,84 ± 3,81 + 0,64 ± 0,08.
								DDK	IS10 + Proposed-NSRC: 83,33 + 78,13 + 80,65 + 80,00 ± 4,19 + 0,60 ± 0,10. IS10+gMFS + Proposed-NSRC: 86,67 + 80,00 + 83,20 + 82,50 ± 4,23 +0,63 ± 0,06.

Article	Database	Information about databases	Data preprocessing	Participants	Algorithms	Languages	Measures	Speech tasks	Results
			INTERSPEECH 2010 (IS10) feature set was extracted. IS10 was combined with glottal MFCC features (IS10+gMFS).			English	Recall + Precision + F1-score + Accuracy + MCC	READ	IS10 + Proposed-NSRC: 88,41 + 80,60 + 83,71 + 78,88 ± 7,33 + 0,47 ± 0,08. IS10+gMFS + Proposed-NSRC: 89,84 + 82,73 + 86,14 + 82,46 ± 7,06 + 0,50 ± 0,09.
								SPON	IS10 + Proposed-NSRC: 84,83 + 81,00 + 82,63 + 81,94 ± 4,63 + 0,55 ± 0,07. IS10+gMFS + Proposed-NSRC: 82,46 + 79,66 + 81,03 + 83,08 ± 4,58 + 0,57 ± 0,09.
Orozco-Arroyave et al., 2016	PC-GITA	Spanish: PC-GITA. Sampled at 44.1 kHz with 16 resolution bits. Captured in noise-controlled conditions. German: Recorded on-state. Sampled at 16 kHz with 16 resolution bits. Czech: Sampled at 48 kHz with 16 resolution bits. Patients newly diagnosed with PD, none of them had been medicated before or during the recordings.	Silence was manually removed from beginning and end of utterances, voiced and unvoiced segments were separated. Speech frames were windowed using Hamming windows of 40 ms with 20 ms of overlap. MFCC-based features, prosodic features and features based on voiced and unvoiced frames were extracted.	Spanish: HC: 25 male, 25 female. PD: 25 male, 25 female. German: HC: 44 male, 44 female. PD: 47 male, 41 female., Czech: HC: 16, 16 male, mean age 61,8 years. PD: 20, 20 male, mean age 61 years.	SVM	Spanish	Accuracy + AUC	READ	Unvoiced: 97 ± 4,8 % + 0,99
								SENTENCE	Unvoiced + Sentence 2: 97 ± 4,80 + 0,97
								DDK	Unvoiced: 99 ± 3,2 % + 0,99
								WORD	Unvoiced + Atleta/Campana: 99 ± 3,2 % + 0,99
						German	Accuracy + AUC	READ	Unvoiced: 94,3 ± 3,9 % + 0,93
								SENTENCE	Unvoiced + Sentence 4: 96,7 ± 5,9 % + 0,97
								DDK	Unvoiced: 97,8 ± 2,9 % + 0,98
								WORD	Unvoiced + Bahnhofsvorsteher: 96,6 ± 2,9 % + 0,97
						Czech	Accuracy + AUC	READ	Unvoiced: 85,0 ± 23,4 % + 0,85
								SENTENCE	Prosody + Sentence 3: 94,4 ± 16,2 % + 0,93
								DDK	Unvoiced: 93,6 ± 16 % + 0,96
								WORD	Unvoiced + kuká: 96,6 ± 12,9 % + 0,96
Vásquez-Correa et al., 2018	PC-GITA.	Extended version of PC-GITA	Phonatory, articulatory, prosodic and intelligibility features were extracted, and multidimensional vectors were formed from them.	68 PD speakers, 33 female, 35 male. 50 HC speakers, 25 male, 25 female.	linear regression with regularization (LASSO), linear support vector regression (Linear-SVR), non-linear support vector regression (rbf-	Spanish	Spearman's correlation coefficient	DDK	Phonation + LASSO + DDK2: 0,56. Articulation + rbf-SVR + DDKS: 0,63. Prosody + Linear-SVR + DDK1: 0,52. Combination + LASSO + DDKS: 0,64
								SENTENCE	Phonation + Linear-SVR + Sentence10: 0,43. Articulation + rbf-SVR + Sentences: 0,61. Prosody + KRR + Sentences: 0,48. Intelligibility + KRR/BRR + Sentences: 0,44. Combination + KRR + Sentences: 0,61.

Article	Database	Information about databases	Data preprocessing	Participants	Algorithms	Languages	Measures	Speech tasks	Results
					SVR), linear ridge regression (LRR), kernel ridge regression (KRR) and Bayesian ridge regression (BRR)			READ	Phonation + LASSO: 0,36. Articulation + rbf-SVR: 0,55. Prosody + rbf-SVR: 0,49. Intelligibility + LRR: 0,44. Combination + LASSO: 0,57.
								SPON	Phonation + BRR: 0,45. Articulation + LASSO: 0,52. Prosody + rbf-SVR: 0,48. Combination + Linear-SVR: 0,53.
								VOWEL (a)	Phonation + Linear-SVR: 0,37. Articulation + KRR: 0,50.
Vásquez-Correa et al., 2020	PC-GITA	PC-GITA	A spectrogram was formed from 500 ms segments from speech signal and used as an input to trained convolutional and recurrent autoencoders which produced the extracted features.	50 HC speakers, 50 PD speakers	deep neural network (DNN) and SVM	Spanish	AUC	DDK	CAE + bottleneck + KA: 0,841
								SENTENCE	CAE + bottleneck&error: 0,749
								READ	CAE + bottleneck: 0,859
								SPON	CAE + bottleneck: 0,874

For the first research question, each language-specific analysis was treated as a separate unit of analysis. Thus, if an article examined multiple languages, each language was considered separately. For instance, an article investigating three languages contributed three language-specific analyses, each producing an independent result. The most effective speech tasks were identified by examining and comparing the measures reported by the authors. For the second research question the results were classified by the languages examined in the research articles and conclusions were made by comparing the results by language. Thus, if a language was examined in two articles, the best speech task for the language was determined based on two units of analysis.

4.4 Research ethics

The Finnish Advisory Board on Research Integrity (2023) has published the guidelines for ethical and responsible research. The guidelines are based on the principles of reliability, honesty, appreciation and responsibility. These guidelines are followed in every phase of this bachelor's thesis. The process of the research is described accurately and openly and the work of other researchers is appreciated by referring to it appropriately.

5 RESULTS

Total of seven articles were included in this review. The speech tasks in the articles were VOWEL, DDK, WORD, SENTENCE, READ and SPON. The speech tasks differ slightly from each other from article to article. For example, Liu and colleagues' (2023) VOWEL task was implemented as prolonging a vowel in a real word while in other studies participants pronounced vowel as long as they could. Additionally, different vowels were used. Almeida and others (2019) used only sustained /a/, while other studies had several vowels. Naturally also the texts in READ tasks varied from language to another.

In total six languages were investigated in the articles included in the review. Lithuanian, Finnish, English, German and Czech were studied in one article each, and Spanish was studied in six articles. Four articles studied only one language while two articles studied two languages and one article studied three languages. The speech tasks, languages and the best speech task of each analysis are reported in Table 3.

Table 3.

Overview of languages, speech tasks and the most suitable speech task across reviewed articles.

Article	Languages	Speech tasks used	Best speech task
Almeida et al., 2019	Lithuanian	VOWEL (a) SENTENCE	VOWEL
Kodali et al., 2023	Spanish	VOWEL (a, e, i, o, u) SENTENCE DDK READ SPON	SPON
Liu et al., 2023	Finnish	VOWEL(a, e, i, o, u) READ SPON	READ
	Spanish	VOWEL (a, e, i, o, u) READ SPON	SPON
Reddy & Alku, 2023	Spanish	VOWEL (a, e, i, o, u) SENTENCE DDK	SENTENCE
	English	READ SPON	SPON
Orozco-Arroyave et al., 2016	Spanish	READ SENTENCE DDK WORD	READ, DDK
	German	READ SENTENCE DDK WORD	DDK
	Czech	READ SENTENCE DDK WORD	DDK
Vásquez-Correa et al., 2018	Spanish	VOWEL (a) SENTENCE DDK READ SPON	DDK
Vásquez-Correa et al., 2020	Spanish	DDK SENTENCE READ SPON	SPON

Additionally, the measures used for comparisons between speech task had variation between studies. Liu and colleagues (2023) used Pearson correlation coefficient (PCC) which is used for measuring linear dependency between two variables (Heikkilä, 2014). Accuracy was used by Kodali and others (2023) and Reddy and Alku (2023), and it

describes the percentage of correct predictions. Accuracy was used by Almeida and others (2019) in combination with area under curve (AUC), which was used by Orozco-Arroyave and colleagues (2016) as well. AUC describes the probability that a model assigns a randomly chosen positive observation a higher score than to a randomly chosen negative observation. Vasquez-Correa and others (2018) calculated Spearman's correlation coefficient, which describes the order of observations (Heikkilä, 2014).

5.1 VOWEL task

VOWEL was used in six language-specific analysis units (Almeida et al., 2019; Kodali et al., 2023; Liu et al., 2023; Reddy & Alku, 2023; Vásquez-Correa et al., 2018). The task differed slightly between articles. In two articles only vowel /a/ was used in analysis (Almeida et al., 2019; Vásquez-Correa et al., 2018). Liu and colleagues (2023), Reddy and Alku (2023) and Kodali and colleagues (2023) investigated /a/, /e/, /i/, /o/ and /u/. In the speech corpus used by Liu and colleagues (2023) the Finnish speakers were instructed to prolong vowels in words, such as /si:ka/. Almeida and colleagues (2019) compared VOWEL with SENTENCE and the results showed that for both acoustic cardioid (AC) and smartphone (SP) recordings the best accuracy and AUC was achieved using VOWEL task and kNN algorithm. For AC, the accuracy was 94,55 ± 3,33 % and AUC was 0,87 and for SP the 92,94 ± 2,55 % and 0,92, respectively. However, the extracted features used for the best result differed.

5.2 DDK task

DDK was used in seven language-specific analysis units (Kodali et al., 2023; Reddy & Alku, 2023; Orozco-Arroyave et al., 2016; Vásquez-Correa et al., 2018; Vásquez-Correa et al., 2020). PC-GITA speech corpus includes six different DDK exercises: repetitions of nonwords pataka, pakata, petaka and syllables pa, ta and ka (Vásquez-Correa et al., 2018). The best results were obtained with DDK in four units of analysis (Orozco-Arroyave et al., 2016; Vásquez-Correa et al., 2018). In Orozco-Arroyave and others' (2016) study DDK and READ performed equally well in Spanish. Vásquez-Correa and others (2018) examined every DDK task in Spanish separately and combination of them. The results showed that when only articulation features where considered, the combination of all DDK tasks performed better than individual DDKs, but when only prosody and only phonatory features were considered, individual DDK tasks performed better than combination.

5.3 WORD task

WORD was used in three language-specific analysis units (Orozco-Arroyave et al., 2016). Orozco-Arroyave and others (2016) investigated 13 isolated words in Spanish, 12 in Czech and six in German. In all languages investigated, the best results for isolated words were obtained using the features extracted from unvoiced segments of speech signals, but other speech tasks outperformed WORD. Isolated words were examined separately. In Spanish two words, atleta and campana, performed equally well with DDK task, but DDK outperformed other isolated words. In Czech the word kuká was superior to the other speech tasks, but there was a lot of variation between words. Therefore, WORD was considered unstable for the purpose.

5.4 SENTENCE task

SENTENCE was used in eight language-specific analysis units (Almeida et al., 2019; Kodali et al., 2023; Reddy & Alku, 2023; Orozco-Arroyave et al., 2016; Vásquez-Correa et al., 2018; Vásquez-Correa et al., 2020). PC-GITA database used in six articles includes ten different sentences (Vásquez-Correa et al., 2018). Almeida and colleagues (2019) examined only one Lithuanian phrase. Number of sentences varied between languages in Orozco-Arroyave and colleagues' (2016) study. They investigated six Spanish sentences, five German sentences and three Czech sentences. In Spanish and German the best result for each sentence was obtained using features derived from unvoiced segments, but in Czech prosodic features provided the most information for two of three sentences. However, other speech tasks performed better in the study. Also Reddy and Alku (2023) used sentences from PC-GITA speech corpus. Their analysis of Spanish was the only one where SENTENCE outperformed other speech tasks in accuracy.

5.5 READ task

READ was used in nine language-specific analysis units (Kodali et al., 2024; Liu et al., 2023; Reddy & Alku, 2023; Orozco-Arroyave et al., 2016; Vásquez-Correa et al., 2018; Vásquez-Correa et al., 2020). Liu and others (2023) investigated reading passage "Northwind and the Sun". In PC-GITA database used by Kodali and others (2024) the text read by participants was a dialogue between a doctor and a patient.

READ was superior to others speech tasks in one analysis (Liu et al., 2023). Speech intelligibility assessment together with combination of SAS and glottal features achieved the highest PCC (0,73). READ also performed equally well with DDK considering AUC

in one analysis (Orozco-Arroyave et al., 2016). In Kodali and others' (2024) study READ achieved accuracy of 53 ± 8 % when classified using MDS-UPDRS-III scale. Reddy and Alku (2023) compared READ and SPON in English and while two measures, recall and F1-score, where higher for READ, the highest accuracy was achieved with SPON and IS10+gMFS features (83,08 ± 4,58 %).

5.6 SPON task

SPON was used in nine language-specific analysis units (Kodali et al., 2024; Liu et al., 2023; Reddy & Alku, 2023; Orozco-Arroyave et al., 2016; Vásquez-Correa et al., 2018; Vásquez-Correa et al., 2020). Implementation of SPON differed between articles. In PC-GITA database (Kodali et al., 2024) the participants were asked to talk about their normal daily life. Liu and colleagues (2023) investigated different spontaneous monologue for HC and PD speakers in the Finnish PDSTU speech corpus. The HC speakers described a one-page cartoon, and the PD speakers were asked to talk about what they did last summer.

In Kodali and colleagues' (2024) study the highest accuracy was achieved with SPON when classified by both MDS-UPDRS-III and MDS-UPDRS-S (58 ± 6 % and 56 ± 4 %, respectively). Liu and others (2023) achieved PCC 0,71 with SPON in Finnish and 0,41 in Spanish. In Spanish SPON was superior to other speech tasks, but in Finnish READ performed better. In Reddy and Alku's (2023) study SPON achieved better accuracy than READ in English. In Vásquez-Correa and colleagues' (2018) study the best Spearman' correlation coefficient for SPON was achieved with combination of phonatory, articulatory and prosodic features (0,53). However, other speech tasks performed better. Vásquez-Correa and others (2020) achieved the highest accuracy with SPON (0,874).

6 DISCUSSION

The purpose of this study was to investigate what is the best speech task to differentiate between Parkinsonian and healthy speech according to the existing research and whether there are differences between languages. The research method was integrative literature review. In total seven articles were included in the review, and six different languages were investigated in the reviewed articles. Spanish was used in six articles, all of which used the same speech corpus, PC-GITA. Lithuanian, Finnish, English, German and Czech were all investigated in one article each. Three articles investigated more than one language. Six different speech tasks were used, number of tasks in one article varying from two to five. Therefore, none of the six speech tasks was present in all of the analyses. Across seven studies, a total of 11 language-specific analyses were conducted to determine PD from healthy speech.

6.1 Speech tasks

Spontaneous speech task was used in six analyses. Out of the six speech tasks that were investigated in the articles, SPON is the most similar to every-day speech. SPON task was superior to the others speech tasks in four analyses (Kodali et al., 2023; Liu et al., 2023; Reddy & Alku, 2023; Vasquez-Correa et al., 2020). However, the speech recordings were made in different conditions and, thus, the quality, length and contents varied.

Syllable repetition task was used in seven analyses and it outperformed other tasks in four of them. In one article of them (Orozco-Arroyave et al., 2016) DDK and READ tasks performed equally well in Spanish dataset. DDK is a well-repeatable and simple task which provides information about movements of articulation organs, which are often reduced in PD. Therefore it would possible be a good indicator in diagnosing PD.

Along with SPON, READ task is also one of the most similar to every-day speech. READ task was used in eight analyses and it was the best speech task in one of them (Liu et al., 2023), in addition to performing equally well with DDK in one analysis.

VOWEL was the best speech task in one of six analysis (Almeida et al., 2019). However, in that case there was only one other speech task it was compared to. Benefits of VOWEL task are its simplicity and language-independency. It is standardizable and can be easily repeated in separate measuring occasions. Therefore it would be a good candidate to be utilized in automatic PD detections, but the results do not support it.

SENTENCE task outperformed others tasks in one of eight analyses (Reddy & Alku, 2023). Sentence reading would be a simple task to perform for patients but detecting the best sentence or a set of them to provide suitable information in diagnosing PD might become a problem.

According to the results, SPON seems to provide the best information for differentiating between Parkinsonian and healthy voice, followed by syllable repetition task. The more complex the task is, the more information it seems to provide for differentiating PD from healthy voice. The results are consistent with the previous research findings (Jiménez-Monsalve et al., 2017; Moro-Velazquez et al., 2019; Pompili et al., 2017). However, the datasets in the articles included in this review were relatively small and not all language-specific analyses included all speech tasks.

6.2 Languages

For Lithuanian the best speech task was found to be sustained vowel /a/ (Almeida et al., 2019). However, Lithuanian was represented only in one article and only two speech tasks were investigated. The vowel /a/ was also the only vowel included in the speech corpus. Additionally, in other reviewed articles, only six of eleven language-specific analyses included the VOWEL task and the other five analyses included other vowels as well.

For Finnish the best speech task was the READ task (Liu et al., 2023). Three speech tasks were included in the Finnish analysis. However, only one article had a Finnish dataset. In the Finnish dataset, the VOWEL task was implemented by prolonging vowels in words. Therefore, the task was somewhat different from the VOWEL tasks in other language-specific analyses.

An English dataset was studied only in one article (Reddy & Alku, 2023). The analysis was performed for two speech tasks only, READ and SPON. Spontaneous speech was found to be more differentiating than reading text.

A German dataset was used in one of the articles (Orozco-Arroyave et al., 2016). The DDK task outperformed READ, SENTENCE and WORD tasks. VOWEL and SPON tasks were not used in this analysis.

For Czech dataset the DDK task was the best speech task for differentiating Parkinsonian and healthy speech (Orozco-Arroyave et al., 2016). Only one of eleven language-specific analyses was conducted for Czech speech recordings. The speech tasks included in the database were READ, DDK, WORD and SENTENCE. SPON and VOWEL tasks were missing.

Spanish dataset, PC-GITA was used in six of eleven language-specific analyses, three of which resulted in SPON task being the most accurate to differentiate between Parkinsonian and healthy speech (Kodali et al., 2023; Liu et al., 2023; Vasquez-Correa et al., 2020). One analysis found that SENTENCE was the best speech task (Reddy & Alku, 2023). SPON task was not present in that analysis. Vasquez-Correa and colleagues (2018) found that DDK outperformed the VOWEL, SENTENCE, READ and SPON tasks. Also Orozco-Arroyave and colleagues (2016) found DDK to be superior to SENTENCE and WORD tasks but perform equally well with the READ task. SPON task was missing in their study. Spanish was the most common language used in the article in this review. However, all the articles studying Spanish used the same dataset.

The results suggest that spontaneous speech is a good candidate for a speech task used for differentiating Parkinsonian and healthy speech. Of all speech tasks studied in the articles included in the review, the SPON task is the most similar to the situations were we speak in every day life. However, the methods used in the articles differed from each other in many ways. Various algorithms were used and preprocessing and feature extraction processes were heterogeneous from one another. The speech recordings were made with non-identical devices and in different environments. The measures used for analyses were variable. All of this may have affected the results and therefore the results are not fully comparable.

Even though the results cannot be generalized, they indicate that the most effective way to distinguish between Parkinsonian and healthy voice is to investigate SPON. When looking at the investigations across different languages, if SPON was used then it was shown as the most consistent assessment for PD. The result can be used for developing tools for diagnosing PD using speech signals as biomarker. This could possibly help people who might not have easy access to health care services.

6.3 The research method

The integrative literature review was chosen as a research method, because it is more critical and systematic in gathering material than narrative literature review but not as strict as systematic literature review. The method was suitable for this research because the topic has already been studied widely, but there is a lack of consensus about the best method for differentiating Parkinsonian and healthy speech based on speech signals. The integrative literature review as method also allows both quantitative and qualitative studies to be included. However, in this review only quantitative studies were included. The nature of the research questions was a limiting factor, because not many qualitative studies answer a specific question, if any. The systematic literature review

would also have been suitable method for answering the research questions, but the method would have limited the data gathering more.

6.4 Clinical relevance and suggestions for future research

The results of this study can be utilized in developing tools and applications for diagnosing PD based on speech signals. Currently there is no single method for diagnosing PD so it would be beneficial and cost-efficient to offer clinical professionals a new non-invasive method was detecting PD. Although the results of this literature review indicate that some speech tasks may be better that others in differentiating between Parkinsonian and healthy speech, the topic must be researched further. In the future studies the material should be broader and more heterogenous, more speech recordings from more people in more languages should be included to ensure that results can be utilized more globally. Also cross-linguistic studies should be made.

The combinations of multiple speech tasks should be investigated, since the features extracted from speech signals vary between speech tasks. By combining for example SPON, READ and DDK it might be possible to achieve better diagnostic differentiation ability than with a single task, because speech tasks provide different information.

In the future further research could focus on systematically comparing different speech tasks and their combinations with larger datasets and evaluating their suitability for early and clinically useful identification of PD.

7 REFERENCES

Albin, R. L. (2023). *Parkinson disease*. Oxford University Press.

Almeida, J. S., Rebouças Filho, P. P., Carneiro, T., Wei, W., Damaševičius, R., Maskeliūnas, R., & de Albuquerque, V. H. C. (2019). Detecting Parkinson's disease with sustained phonation and speech signals using machine learning techniques. *Pattern Recognition Letters*, *125*, 55–62. https://doi.org/10.1016/j.patrec.2019.04.005

Atula, S. (2023). Parkinsonin tauti. Lääkärikirja Duodecim. Available in https://www.terveyskirjasto.fi/dlk00055. Referenced 19.1.2025.

Booth, A., Sutton, A., Clowes, M. & Martyn-St James, M. (2021). *Systematic Approaches to a Successful Literature Review, 3rd Edition* (third edition). Sage Publications Ltd UK.

Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia. Available at www.covidence.org.

Duffy, J. R. (2020). *Motor speech disorders : substrates, differential diagnosis, and management* (Fourth edition.). Elsevier.

Enderby, P. (1980). Frenchay Dysarthria Assessment. *International Journal of Language & Communication Disorders*, *15*(3), 165–173. https://doi.org/10.3109/13682828009112541

Fumel, J., Bahuaud, D., Weed, E., Fusaroli, R., & Basirat, A. (2024). A Systematic Review and Bayesian Meta-Analysis of Acoustic Measures of Prosody in Parkinson's Disease. *Journal of Speech, Language, and Hearing Research*, *67*(8), 2548–2564. https://doi.org/10.1044/2024_JSLHR-23-00588

Goetz, C. G., Poewe, W., Rascol, O., Sampaio, C., Stebbins, G. T., Counsell, C., Giladi, N., Holloway, R. G., Moore, C. G., Wenning, G. K., Yahr, M. D., & Seidl, L. (2004). Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: Status and recommendations. *Movement Disorders*, *19*(9), 1020–1028. https://doi.org/10.1002/mds.20213

Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martinez-Martin, P., Poewe, W., Sampaio, C., Stern, M. B., Dodel, R., Dubois, B., Holloway, R., Jankovic, J., Kulisevsky, J., Lang, A. E., Lees, A., Leurgans, S., LeWitt, P. A., Nyenhuis, D., Olanow, C. W., Rascol, O., Schrag, A., Teresi, J. A., van Hilten, J.J. & LaPelle, N. (2008). Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and

clinimetric testing results. *Movement Disorders*, *23*(15), 2129–2170. https://doi.org/10.1002/mds.22340

Heikkilä, T. (2014). *Tilastollinen tutkimus* ([9. uudistettu painos]). Edita.

Ho, A. K., Iansek, R., Marigliani, C., Bradshaw, J. L., & Gates, S. (1998). 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

Hoehn, M. M., & Yahr, M. D. (2001). Parkinsonism: onset, progression, and mortality. 1967. *Neurology*, *57*(10), S11-26.

Hughes, A. J., Daniel, S. E., Kilford, L., & Lees, A. J. (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *Journal of Neurology, Neurosurgery and Psychiatry*, *55*(3), 181–184. https://doi.org/10.1136/jnnp.55.3.181

Jiménez-Monsalve, J. C., Vásquez-Correa, J. C., Orozco-Arroyave, J. R., & Gomez-Vilda, P. (2017). Phonation and Articulation Analyses in Laryngeal Pathologies, Cleft Lip and Palate, and Parkinson's Disease. In H. Adeli, J. M. Ferrández Vicente, J. Toledo Moreo, J. R. Álvarez-Sánchez, & F. de la Paz López (Eds.), *Biomedical Applications Based on Natural and Artificial Computing* (Vol. 10338, pp. 424–434). Springer International Publishing. https://doi.org/10.1007/978-3-319-59773-7_43

Kangasniemi, M., Utriainen, K., Ahonen, S.-M., Pietilä, A.-M., Jääskeläinen, P., & Liikanen, E. (2013). Kuvaileva kirjallisuuskatsaus: eteneminen tutkimuskysymyksestä jäsennettyyn tietoon/Narrative literature review: from a research question to structured knowledge. *Hoitotiede*, *25*(4), 291-.

Kempler, D., & Lancker, D. V. (2002). Effect of Speech Task on Intelligibility in Dysarthria: A Case Study of Parkinson's Disease. *Brain and Language*, *80*(3), 449–464. https://doi.org/10.1006/brln.2001.2602

Kodali, M., Kadiri, S. R., & Alku, P. (2024). Automatic classification of the severity level of Parkinson's disease: A comparison of speaking tasks, features, and classifiers. *Computer Speech & Language*, *83*, 101548-. https://doi.org/10.1016/j.csl.2023.101548

Liu, Y., Reddy, M. K., Penttila, N., Ihalainen, T., Alku, P., & Rasanen, O. (2023). Automatic Assessment of Parkinson's Disease Using Speech Representations of Phonation and Articulation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *31*, 242–255. https://doi.org/10.1109/TASLP.2022.3212829

Miller, N. (2012). Speech, Voice and Language in Parkinson's Disease: Changes and Interventions. *Neurodegenerative Disease Management*, *2*(3), 279–289. https://doi.org/10.2217/nmt.12.15

Moro-Velázquez, L., Gómez-García, J. A., Dehak, N., & Godino-Llorente, J. I. (2019). Analysis of phonatory features for the automatic detection of Parkinson's disease in two different corpora. *Models and Analysis of Vocal Emissions for Biomedical Applications 11th International Workshop, December, 17-19, 2019*, 33–36.

Moro-Velazquez, L., Gomez-Garcia, J. A., Arias-Londoño, J. D., Dehak, N., & Godino-Llorente, J. I. (2021). Advances in Parkinson's Disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects. *Biomedical Signal Processing and Control*, *66*, 102418-. https://doi.org/10.1016/j.bspc.2021.102418

Orozco-Arroyave, J. R., Hönig, F., Arias-Londoño, J. D., Vargas-Bonilla, J. F., Daqrouq, K., Skodda, S., Rusz, J., & Nöth, E. (2016). Automatic detection of Parkinson's disease in running speech spoken in three different languages. *The Journal of the Acoustical Society of America*, *139*(1), 481–500. https://doi.org/10.1121/1.4939739

Penttilä, N., Tavi, L., Hyppönen, M., Rontu, K., Rantala, L., & Werner, S. (2022). *Prosodic features in Finnish-speaking adults with Parkinson´s disease*.

Pompili, A., Abad, A., Romano, P., Martins, I. P., Cardoso, R., Santos, H., Carvalho, J., Guimarães, I., & Ferreira, J. J. (2017). Automatic Detection of Parkinson's Disease: An Experimental Analysis of Common Speech Production Tasks Used for Diagnosis. In K. Ekstein & V. Matousek (Eds.), *Text, Speech, and Dialogue* (Vol. 10415, pp. 411–419). Springer International Publishing AG. https://doi.org/10.1007/978-3-319-64206-2_46

Reddy, M. K., & Alku, P. (2023). Exemplar-based Sparse Representations for Detection of Parkinson's Disease from Speech. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *31*, 1–11. https://doi.org/10.1109/TASLP.2023.3260709

Rizzo, G., Copetti, M., Arcuti, S., Martino, D., Fontana, A., & Logroscino, G. (2016). Accuracy of clinical diagnosis of Parkinson disease: A systematic review and meta-analysis. *Neurology*, *86*(6), 566–576. https://doi.org/10.1212/WNL.0000000000002350

The Finnish Advisory Board on Research Integrity (2023). Responsible conduct of research and procedures for handling allegations of misconduct in Finland. Guidelines of the Finnish Advisory Board on Research Integrity 2023. Helsinki:

The Finnish Advisory Board on Research Integrity. Available in https://tenk.fi/sites/default/files/2023-03/HTK-ohje_2023.pdf

Tolosa, E., Garrido, A., Scholz, S. W., & Poewe, W. (2021). Challenges in the diagnosis of Parkinson's disease. *Lancet Neurology*, *20*(5), 385–397. https://doi.org/10.1016/S1474-4422(21)00030-2

Vásquez-Correa, J. C., Orozco-Arroyave, J. R., Bocklet, T., & Nöth, E. (2018). Towards an automatic evaluation of the dysarthria level of patients with Parkinson's disease. *Journal of Communication Disorders*, *76*, 21–36. https://doi.org/10.1016/j.jcomdis.2018.08.002

Vilkka, H. (2023). *Kirjallisuuskatsaus metodina, opinnäytetyön osana ja tekstilajina*. Art House.