



HAL
open science

Automated objective speech markers for differential diagnosis between parkinson's disease and atypical parkinsonian disorders

Biswajit Das

► **To cite this version:**

Biswajit Das. Automated objective speech markers for differential diagnosis between parkinson's disease and atypical parkinsonian disorders. Modeling and Simulation. Université de Bordeaux, 2021. English. NNT : 2021BORD0225 . tel-03436409

HAL Id: tel-03436409

<https://theses.hal.science/tel-03436409>

Submitted on 19 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE EN COTUTELLE PRÉSENTÉE
POUR OBTENIR LE GRADE DE
DOCTEUR DE
L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE DE MATHÉMATIQUES ET INFORMATIQUE
SPÉCIALITÉ : INFORMATIQUE

Par Biswajit DAS

**MARQUEURS VOCAUX OBJECTIFS AUTOMATISÉS POUR LE
DIAGNOSTIC DIFFÉRENTIEL ENTRE LA MALADIE DE
PARKINSON ET LES TROUBLES PARKINSONIENS
ATYPIQUES**

Sous la direction de : Dr. Khalid Daoudi

Soutenu le : 30th September, 2021

Membres du jury :

Mme ANDRÉ-OBRECHT Régine	Professeure émérite, Institut de Recherche en Informatique de Toulouse	Président
Mme ANDRÉ-OBRECHT Régine	Professeure émérite, Institut de Recherche en Informatique de Toulouse	Rapporteur
M. SICARD Etienne	Professeur, l'INSA Toulouse	Rapporteur
M. BOUGUÏLA Nizar	Professeur, Concordia Institute for Information Systems Engineering	Examineur
M. YAHIA Hussein	Chargé de Recherche, Inria Bordeaux	Co-Directeur
M. DAOUDI Khalid	Chargé de Recherche, Inria Bordeaux	Directeur de Thèse

Résumé

Les troubles de la parole sont une manifestation précoce et fréquente des troubles neurologiques. Par conséquent, l'analyse des troubles de la parole et la détection de la pathophysiologie sous-jacente revêtent une grande importance pour la pratique clinique. Le parkinsonisme est un trouble neurologiques qui fait référence à la maladie de Parkinson idiopathique (MP) et aux syndromes parkinsoniens atypiques (SPA), tels que la paralysie supranucléaire progressive (PSP) et l'atrophie multisystématisée (AMS). Le diagnostic différentiel entre ces groupes de maladies est une tâche difficile en raison de la similitude des symptômes aux premiers stades, alors que la certitude du diagnostic précoce est essentielle pour le patient en raison du pronostic divergent. En effet, malgré des efforts récents, aucun marqueur objectif validé n'est actuellement disponible pour guider le clinicien dans le diagnostic différentiel. Cette thèse vise à concevoir des marqueurs vocaux adaptés au diagnostic différentiel du parkinsonisme et à fournir des indications sur leur spécificité qui pourraient être utiles dans le diagnostic différentiel précoce.

Le premier défi de cette thèse était de concevoir des marqueurs vocaux distinctifs pour le diagnostic différentiel entre la MP et l'AMS-P, le sous-type parkinsonien de l'AMS. Nous avons commencé par analyser la réalisation des consonnes initiales des mots à partir de pseudo-mots en utilisant à la fois l'inspection visuelle du spectrogramme et une méthode objective. Nous avons utilisé une base de données collectée dans le cadre du projet Voice4PD-MSA et qui consiste en des enregistrements vocaux de patients français PD et AMS-P ainsi que de témoins sains. L'analyse a révélé un dévoisement fréquent des obstruents voisés chez les patients AMS-P par rapport aux patients PD. L'occurrence de bursts dans les fricatives non voisées a également été identifiée, en utilisant la détection visuelle et automatique, comme un autre marqueur vocalique distinctif de l'AMS-P.

La réalisation des voyelles a été analysée en utilisant des voyelles tenues et une phrase extraite d'un texte lu. Des anomalies dans la synchronisation de la vibration des plis vocaux et des mouvements involontaires des articulateurs ont été identifiées comme des troubles distinctifs de l'AMS-P. La réduction de l'espace vocalique est également prédominante chez les patients atteints d'AMS-P par rapport à ceux atteints de la maladie de Parkinson. La diadochokinésie (DDK), répétition rapide de mouvements alternés des articulateurs, a été analysée à l'aide de tâches de répétition de syllabes. Elle a révélé un déviance rythmique plus importante chez les patients atteints de l'AMS-P par rapport à ceux atteints de la MP.

Le deuxième défi de cette thèse était de concevoir des marqueurs de parole pour le diagnostic différentiel entre la PSP et l'AMS. Pour effectuer l'analyse, nous avons utilisé une base de données d'enregistrements vocaux de patients tchèques MP, PSP et AMS. Nous avons adopté une procédure semi-supervisée de combinaison linéaire de caractéristiques pour concevoir deux ensembles de nouvelles caractéristiques vocales distinctives. Le premier est lié aux sous-systèmes de production de la parole : respiration, phonation, articulation, prosodie et timing. Le second est lié aux sous-types de dysarthrie : hypokinétique, ataxique et spastique. Les deux ensembles ont permis de discriminer la PSP et l'AMS avec une grande précision.

Dans l'ensemble, cette thèse a fourni un cadre pour détecter et concevoir des marqueurs vocaux interprétables et distinctifs, à partir de différentes tâches vocales, pour le diagnostic différentiel entre les troubles parkinsoniens. Certains des résultats de cette thèse pourraient servir de base et/ou d'inspiration pour de futures recherches visant à atteindre l'objectif final très ambitieux : le diagnostic différentiel **précoce**.

Mots clés: Troubles de la parole, Neurodégénération, Maladie de Parkinson, Atrophie du système multiple, Paralyse supranucléaire progressive, Dysarthrie, Traitement pathologique de la parole, Diagnostic différentiel.

IN PARTIAL FULFILLMENT OF REQUIREMENTS OF THE DEGREE OF

Doctor of Philosophy

UNIVERSITY OF BORDEAUX

DOCTORAL SCHOOL OF MATHEMATICS AND COMPUTER SCIENCE

SPÉCIALITÉ : INFORMATION TECHNOLOGY

By Biswajit DAS

AUTOMATED OBJECTIVE SPEECH MARKERS FOR DIFFERENTIAL DIAGNOSIS BETWEEN PARKINSON'S DISEASE AND ATYPICAL PARKINSONIAN DISORDERS

Under the direction : Dr. Khalid Daoudi

Thesis defense : 30th September, 2021

Membres du jury :

Dr. ANDRÉ-OBRECHT Régine	Professeure émérite, Institut de Recherche en Informatique de Toulouse	Jury president
Dr. ANDRÉ-OBRECHT Régine	Professeure émérite, Institut de Recherche en Informatique de Toulouse	Reviewer
Dr. SICARD Etienne	Professeur, l'INSA Toulouse	Reviewer
Dr. BOUGUÏLA Nizar	Professeur, Concordia Institute for Information Systems Engineering	Examiner
Dr. YAHIA Hussein	Chargé de Recherche, Inria Bordeaux	Co-Advisor
Dr. DAOUDI Khalid	Chargé de Recherche, Inria Bordeaux	Advisor

Abstract

Speech disorder is an early and prominent manifestation of neurological disorders. Therefore, the breakdown of speech disorders and detecting underlying pathophysiology have a valuable importance to clinical practice. Parkinsonism is one of the neurological disorder that refers to idiopathic Parkinson's Disease (PD) and Atypical Parkinsonian Syndromes (APS), such as Progressive Supranuclear Palsy (PSP) and Multiple System Atrophy (MSA). Differential diagnosis between these disease groups is a challenging task due to similar symptoms at the early stages, while early diagnostic certainty is essential for the patient because of the diverging prognosis. Indeed, despite recent efforts, no validated and cost-effective objective marker is currently available to guide the clinician for the differential diagnosis. This thesis aims to design speech markers suitable for differential diagnosis in Parkinsonism and to provide some insights on the disease-specificity of some impairments that could be useful in early differential diagnosis.

The first challenge of this thesis was to design distinctive speech markers for differential diagnosis between PD and MSA-P, the Parkinsonian subtype of MSA. We started by analysing the realization of word-initial consonants from pseudo-words using both visual spectrogram inspection and an objective method. We used a database collected in the framework of the Voice4PD-MSA project and which consists in speech recordings of PD and MSA-P French patients as well as healthy controls. Analysis revealed frequent devoicing in voiced obstruents for MSA-P patients as compared to PD. The occurrence of bursts in unvoiced fricatives was also identified, using visual and automatic detection, as another distinctive speech markers of MSA-P.

Vowel realization was analyzed using sustained vowels and a sentence extracted from a passage reading. Abnormality in vocal folds vibration timing, and involuntary movements of articulators came up as distinctive speech disorder for MSA-P. Reduced vowel space area was also found to be predominant for MSA-P patients compared to PD. Diadochokinesis (DDK), the rapid repetition of alternating movements of articulators, was analyzed using syllable repetition tasks. It revealed a more prominent rhythmic disorder of MSA-P patients compared to PD.

The second challenge of this thesis was to design speech markers for differential diagnosis between PSP and MSA. To conduct the analysis, we used

a database of speech recordings of PD, PSP, and MSA Czech patients. We adopted a semi-supervised feature linear combination procedure to design two sets of new distinctive speech features. The first one is related to subsystems of speech production: respiration, phonation, articulation, prosody and timing. The second one is related to dysarthria subtypes: hypokinetic, ataxic and spastic. Both sets led to discrimination between PSP and MSA with a high accuracy.

Overall, this thesis provided a framework to detect and design interpretable disease specific speech markers from different speech tasks for differential diagnosis between Parkinsonian disorders. Some of the results of the thesis could serve as basis and/or inspiration for future research towards the very challenging ultimate target: **early** differential diagnosis.

Key words: Speech disorders, Parkinson's disease, Multiple system atrophy, Progressive supranuclear palsy, Dysarthria, Pathological speech processing, Differential diagnosis.

Unité de recherche

Equipe GEOSTAT, centre de recherche INRIA Bordeaux Sud-Ouest,
200 Avenue de la Vieille Tour, 33405 Talence, France.

Résumé étendu en Français

Les troubles de la parole sont une manifestation précoce et importante des troubles neurologiques. Par conséquent, la décomposition des troubles de la parole et la détection de la pathophysiologie sous-jacente ont une importance inestimable pour la pratique clinique. Les troubles de la parole sont généralement attribués au vieillissement, mais le modèle de trouble est surtout distinct pour la voix neurogène. Le parkinsonisme est l'un des troubles neurologiques qui fait référence à la maladie de Parkinson idiopathique (MP) et aux syndromes parkinsoniens atypiques (SPA), tels que la paralysie supranucléaire progressive (PSP) et l'atrophie multisystémique (AMS). Le diagnostic différentiel de ces derniers groupes de maladies (MP et SPA) reste une tâche difficile en raison de la similitude des symptômes aux stades précoces, alors que la certitude diagnostique précoce est essentielle pour le patient en raison du pronostic divergent. En effet, malgré des efforts récents, aucun marqueur objectif validé de la parole n'est actuellement disponible pour guider le clinicien dans son diagnostic différentiel. Cette thèse a donc pour but de concevoir et de définir les marqueurs de la parole qui permettraient de mieux comprendre les troubles de la parole causés par des maladies neurologiques, puis de poser un diagnostic différentiel.

L'analyse des troubles de la parole nécessite au moins une base de données vocales permettant d'évaluer le modèle des anomalies de la parole. La base de données vocales des maladies neurologiques MP et AMS-P n'est pas disponible en langue française. Ainsi, le développement d'une base de données vocales (Voice4PD-MSA) pour les groupes PD et MSA-P était l'un des objectifs de cette thèse. Lors du développement de la base de données Voice4PD-MSA, nous avons exploré la base de données Czech-Data qui comprend des échantillons de parole des groupes de maladies MP, PSP et AMS en langue tchèque pour le diagnostic différentiel.

L'algorithme automatique est toujours en demande pour quantifier l'observation perceptive et visuelle afin de capturer des troubles particuliers de la parole. Les composantes de la parole cliniquement interprétables sont considérées comme des anomalies de la respiration, de la production de voyelles, des mouvements de l'articulateur et de la prosodie par des méthodes objectives à partir de voyelles soutenues, de consonnes initiales de mots, de tâches diadochocinétiques (DDK) et de la parole continue. Les voyelles imprécises comprennent les déficits d'ouverture et de fermeture des plis vocaux, les mouvements involontaires de l'articulateur, l'hypernasalité, les tremblements et les modifications de la zone de l'espace vocal sont observés comme étant importants pour le diagnostic différentiel des patients atteints de AMS-P et de MP. Dans les obstructions imprécises, le dévoicing dans les obstructions voisées et l'éclatement

dans les fricatives (anti-spirantisation) sont identifiés comme des marqueurs vocaux distinctifs pour le AMS-P. En outre, les indices vocaux liés au sous-système de production de la parole et à la dysarthrie permettent une différenciation encourageante et une spécificité de la maladie dans les groupes de maladies. Compte tenu de la faible quantité de données, les caractéristiques vocales bidimensionnelles sont conçues de manière à ce que l'un des groupes de maladies prédomine dans une dimension vocale, ce qui permet de distinguer les groupes de maladies avec un bon score de classification.

Le diagnostic différentiel précoce était un autre objectif essentiel de la présente étude. La présente étude a observé des indications encourageantes sur le diagnostic différentiel précoce en explorant la tendance des marqueurs vocaux par rapport aux signes cliniques. Ainsi, nous aspirons à ce que la méthodologie présentée dans cette thèse serve d'outil de diagnostic potentiel dans la pratique clinique et inspire le développement de méthodes automatiques pour étudier les troubles de la parole dans le parkinsonisme. Ainsi, les sections suivantes décriront la contribution au développement de la base de données de la parole, les troubles des voyelles à partir de la voyelle soutenue et de la lecture de texte, les consonnes initiales des mots à partir du logatome, de la tâche diadochocinétique et de la tâche de lecture pour les patients atteints de la MP et de l'AMS-P. De plus, il décrira le diagnostic différentiel de la PSP et du AMS.

0.1 Base de données

Deux bases de données vocales différentes ont été utilisées pour trouver des marqueurs vocaux spécifiques à la maladie pour le diagnostic différentiel. La base de données, Voice4PD-MSA était dédiée à la collecte d'échantillons de parole (voyelles soutenues, /s/ soutenu, tâche diadochocinétique, lecture de texte, monologue et logatomes) en langue française de patients PD et AMS-P. De 2018 au moment de la rédaction de cette thèse, un total de 60 locuteurs français ont été recrutés dans le cadre d'un projet de recherche impliquant les services de neurologie et d'ORL de 2 hôpitaux universitaires français. 27 patients (8 femmes et 19 hommes) ont été diagnostiqués avec une MP idiopathique, avec un âge moyen de 60 ans et une durée moyenne des symptômes de 4 ans. 13 sujets (8 femmes et 5 hommes) ont reçu un diagnostic d'AMS-P, avec un âge moyen de 67 ans et une durée moyenne des symptômes de 3,5 ans. une durée moyenne des symptômes de 3,5 ans. 20 témoins sains (HC) d'un âge moyen de 56 ans (10 femmes et 10 hommes) sans antécédents de troubles neurologiques ou de communication ont été recrutés. D'autres modes d'enregistrement tels que l'électroglottographe (EGG), les données aérodynamiques et la laryngostroboscopie ont également été pris en compte.

La base de données tchèque est constituée d'échantillons de parole provenant des groupes de maladies MP, PSP et AMS. Le groupe sain comprend 150 sujets (95 hommes et 55 femmes). Le groupe MP idiopathique comprend 25 patients (16 hommes et 9 femmes). Le groupe PSP contient 20 patients (13 hommes et 7 femmes), tandis que 19 patients PSP ont diagnostiqué le syndrome de Richardson (PSP-RS). D'autre part, le groupe AMS compte 25 patients (15 hommes et 10 femmes).

0.2 Distorsion des voyelles dans la MP et l'AMS-P

L'altération des voyelles est la plus fréquente dans les troubles neurologiques, et donc largement étudiée [45] pour évaluer les déficits de la vibration des plis vocaux et des mouvements de l'articulateur. Cependant, des études supplémentaires sont nécessaires pour valider les résultats précédents et, en particulier, pour trouver des marqueurs objectifs de la parole liés à la production de voyelles pour le diagnostic différentiel.

La présente étude explore l'altération des voyelles par des déficits laryngés et des mouvements imprécis de l'articulateur. Deux tâches vocales différentes sont utilisées pour cette évaluation. Tout d'abord, la voyelle soutenue /a/ est utilisée pour évaluer le dysfonctionnement laryngé, l'instabilité de l'articulateur et les activités vélopharyngées. Ensuite, la tâche de lecture de texte est utilisée pour l'évaluation de l'espace vocalique (VSA). À notre connaissance, la VSA n'a jamais été utilisée pour le diagnostic différentiel du MP et du AMS-P.

Les caractéristiques acoustiques de la phonation telles que le jitter, le shimmer, le HNR, la moyenne et la déviation standard du Quotient Quasi-Open (QOQ), la dérivée de F0, le degré de voiceless (DUV), l'indice d'intensité du tremblement de fréquence (FTRI) calculé par Disvoice toolkit, et les scripts Praat. Pour évaluer les mouvements volontaires de l'articulateur, les caractéristiques acoustiques telles que la déviation standard de la densité spectrale de puissance dans la bande de fréquence et les caractéristiques nasales sont calculées par l'auteur de la thèse en suivant des études précédentes [211, 130]. Les caractéristiques acoustiques liées à la vibration des plis vocaux permettent une discrimination encourageante entre les patients MP et AMS-P. De plus, les patients atteints d'AMS-P présentent des déficits prédominants dans les dimensions de l'articulation et de la parole nasale. Deux indices de la parole sont développés pour mesurer la sévérité globale des sous-systèmes de phonation et d'articulation, ce qui améliore les différences de groupe entre la MP et l'AMS-P. L'utilisation de deux indices vocaux dans l'arbre de décision donne une précision de 95 %, une spécificité de 96,29 % et une sensibilité de 92,30 %.

Les modifications des fréquences des formants des voyelles sont évaluées par l'aire de l'espace vocalique (VSA), l'indice d'articulation vocalique (VAI), le rapport des formants et la dispersion des formants. Comme le dimorphisme du sexe peut jouer un rôle important dans les fréquences des formants, les caractéristiques acoustiques indépendantes du sexe sont prises en compte pour le diagnostic différentiel du MP et de l'AMS-P. La conversion de l'unité Hz en demi-ton fournit des propriétés indépendantes du sexe dans les mesures de l'VSA et de la dispersion des formants. Dans la plupart des cas, les patients atteints d'AMS-M ont montré une sévérité prédominante par rapport à la MP. Il est important de noter que l'énergie de la bande de la deuxième fréquence des formants est comparativement plus faible chez les patients atteints d'ASA-M que chez les patients atteints de la MP, probablement en raison d'une hypokinésie.

0.3 Distorsion des consonnes dans la MP et l'AMS-P

La production d'une consonne implique toujours une synchronisation et une coordination des fonctions articulatoires et laryngées. L'analyse de la distorsion consonantique se limite principalement aux obstructions (où le flux d'air est partiellement ou totalement obstrué) en raison de sa prévalence dans les troubles du langage. Les obstruents se composent de plosives et de fricatives arrêtées. Les arrêts non voisés sont principalement analysés par les propriétés du burst (réalisation de la fermeture de l'articulateur). D'autre part, le mode de friction est examiné pour les fricatives non voisées. Les propriétés des obstruents ont été explorées dans des études précédentes : Blumstein1979, Kazumi2009.

Selon le mode d'articulation de la consonne, les obstructions sont classées en deux groupes : les obstructions voisées et les obstructions non voisées. Les obstructions voisées sont caractérisées par la présence d'une vibration des plis vocaux et d'un type particulier d'obstruction par l'articulateur. Dans cette étude, nous avons d'abord exploré les anomalies de la production d'obstructions voisées par le MP et le AMS-P. Ensuite, les obstructions non voisées sont étudiées pour trouver des anomalies spécifiques.

Dans cette étude, les distorsions ciblées de la parole sont d'abord détectées par des méthodes visuelles, puis des mesures objectives sont proposées. Par ce processus, nous avons observé un dévoicing (absence de vibration des plis vocaux) significatif dans les obstructions vocales chez les patients atteints d'AMS-P par rapport à ceux atteints de MP. Notamment, le dévoicing est plus fréquent dans les obstructions vélares. Une mesure objective est proposée par l'auteur de la thèse pour quantifier le degré de dévoicing dans les obstructions vocales. De plus, des mesures temporelles comme le temps d'apparition de la voix (VOT) et le rapport du temps d'apparition de la voix (VOTR) sont calculées comme mesures conventionnelles. Dans les deux cas, les patients de l'AMS-P présentent un temps d'apparition de la voix et un rapport de temps d'apparition de la voix plus faibles que ceux de l'MP et de l'HC.

La distorsion des obstructions non voisées est évaluée par le mode d'éclatement des obstructions, comme la présence d'éclatement dans les fricatives, d'éclatement faible dans les plosives stop et d'éclatement multiple dans les plosives stop. Ce dernier trouble a d'abord été évalué par la méthode visuelle. Ensuite, une méthode automatique est proposée pour mesurer ces anomalies. Nous avons observé la présence fréquente de bursts dans la fricative bilabiale /f/ chez les patients AMS-P par rapport aux MP et HC. La méthode automatique correspond également au résultat de l'observation visuelle. Un faible burst est observé dans le stop alvéolaire /t/ pour les patients MP et AMS-P. Bien que l'éclatement faible ne soit pas adapté au diagnostic différentiel, il pourrait servir d'indice de trouble de la parole. Les VOT et VOTR des plosives stop sont plus longs chez les patients AMS-P que chez les patients MP.

0.4 Trouble de la parole dans la production diadochocinétique de la parole

La dysarthrie est une manifestation courante du trouble parkinsonien. La dégénérescence des ganglions de la base peut affecter les aspects temporo-spatiaux de la parole motrice et du rythme de la parole [293, 40, 94]. En revanche, la dégénérescence du cervelet peut affecter le maintien de la précision de l'intervalle de synchronisation [265, 288]. Ainsi, nous pouvons émettre l'hypothèse que toute activité rythmique nécessite une interaction étroite entre les ganglions de la base et les circuits de contrôle cérébelleux. Comme la voyelle soutenue a été considérée pour évaluer la vibration des plis vocaux, la tâche de répétition de syllabes servirait à vérifier les mouvements articulatoires. En outre, les tâches de parole complexes peuvent révéler un large éventail de troubles de la parole par rapport à la simple phonation soutenue. Par conséquent, la tâche diadochocinétique (DDK) a été conçue pour évaluer principalement les déficits dans les mouvements articulatoires et la coordination des sous-systèmes respiratoires, phonatoires et articulatoires pour la parole pathologique. De plus, la tâche diadochocinétique conviendrait pour mesurer les consonnes imprécises, le débit des syllabes, l'irrégularité dans la répétition des syllabes ; [74].

Dans cette étude, nous avons mesuré plusieurs mesures acoustiques telles que la durée des voyelles (VD), la déviation standard de la densité spectrale de puissance (stdPWR), le taux de DDK (DDKR), l'irrégularité de la DDK (DDKI), l'accélération de la DDK (DDKA), le VOT, le Weak Burst (WB) pour évaluer la tâche diadochocinétique /pa-ta-ka/ pour les patients MP et AMS-P. Une méthodologie automatique est développée pour segmenter les consonnes et les voyelles d'arrêt avec une grande précision. Les mesures objectives ont montré une sévérité prédominante chez les patients atteints d'AMS-P par rapport à ceux atteints de la maladie de Parkinson. Notamment, le mode rapide de /pa-ta-ka/ a produit plus de troubles pour les maladies neurologiques. La combinaison du stdPWR et du VD permet une très bonne discrimination entre la MP et l'AMS-P par rapport à la MP, probablement en raison de l'ataxie plus importante chez les patients de l'AMS.

0.5 Trouble de la parole dans la lecture d'un texte

La parole spontanée est le modèle le plus complexe de production de la parole. Il comprend des fonctions cognitives ainsi que l'exécution de fonctions motrices de la parole. Dans l'aspect cognitif, les pensées, les sentiments et les émotions sont d'abord formés en fonction du langage pour la communication verbale. Ensuite, le message verbal prévu doit être organisé pour l'exécution neuromusculaire. Ces activités comprennent la sélection, le séquençage et la régulation de "programmes" sensorimoteurs qui activent les muscles de la parole à des moments coarticulés, des durées et des fréquences appropriés. temps, durées et intensités coarticulés appropriés [69]. Ainsi, les tâches d'élocution spontanée peuvent révéler des déficits dans la coordination des fonctions phonatoires et respiratoires, des mouvements articulatoires précis, et de la coordination laryngée et supra-laryngée (lèvres, mâchoire, langue, etc.), qui reflètent

également l'aspect prosodique et temporel de la parole.

La présente étude se penche sur l'analyse de la lecture de textes en langue française pour les patients MP et l'AMS-P. Par conséquent, cette étude a d'abord tenté de segmenter manuellement les événements de la parole (vocale, non vocale, pause, respiration). Ensuite, nous avons adopté la méthodologie décrite dans [306, 130] pour calculer plusieurs caractéristiques acoustiques afin d'étudier la disparité entre MP et AMS-P. En outre, un autre logiciel open-source est également utilisé pour extraire les paramètres acoustiques.

Les mesures objectives de la segmentation manuelle donnent des troubles de la parole encourageants pour la MP et l'AMS-P. Cependant, une simple moyenne des mesures acoustiques permet d'évaluer la qualité de la parole. Cependant, une simple moyenne des mesures acoustiques, qui représentent la dysarthrie hypokinétique, donne une bonne discrimination entre la MP et l'AMS-P. De plus, les caractéristiques prosodiques ont été calculées par une boîte à outils open-source, Disvoice, qui a également observé des troubles de la parole chez les patients atteints de la MP et de l'AMS-P.

0.6 Diagnostic différentiel entre PSP et AMS

Le syndrome parkinsonien est un terme générique qui désigne la maladie de Parkinson (MP), les syndromes parkinsoniens atypiques (SPA) comme la paralysie supranucléaire progressive (PSP) et l'atrophie multisystémique (AMS). Le SPA diffère de la MP par une atteinte neuronale plus étendue, qui se traduit par des signes cliniques supplémentaires, une progression plus rapide de la maladie et une mauvaise réponse au traitement de substitution de la dopamine [266]. La majorité des patients atteints de PSP et l'AMS développent des caractéristiques cliniques qui chevauchent celles de la MP. Ainsi, le diagnostic correct peut être très difficile à établir dans les premiers stades de la maladie.

La présente étude se concentre sur la définition de nouveaux indices de la parole qui peuvent mesurer objectivement les déficits des sous-systèmes de production de la parole et/ou des sous-types particuliers de dysarthrie pour les groupes de maladies PSP et MSA. Les dimensions acoustiques sont conçues de manière à pouvoir montrer une spécificité de la maladie. Ces caractéristiques auront un comportement (statistique) pour la PSP, qui est significativement différent de celui de l'AMS. De plus, ces caractéristiques seront conçues de manière à pouvoir être interprétées afin d'améliorer la compréhension des troubles de la parole dans la PSP et l'AMS. De toute évidence, le premier avantage d'une telle étude serait une discrimination précise et objective entre la PSP et l'AMS, étant donné que l'évaluation subjective est assez difficile en raison du comportement perceptif similaire [197]. Le second avantage, plus important, est de permettre de formuler des hypothèses concernant le stade précoce des maladies. En outre, le sexe des participants est également pris en compte dans cette étude, ce qui peut permettre de déduire des informations supplémentaires concernant la pathologie. Des études antérieures ont également indiqué un dimorphisme de genre, mais ces études ont surtout utilisé un nombre moins important de composantes

acoustiques : [125, 282]. Inversement, certaines études ont montré que l'influence du sexe était soit indépendante [247], soit ignorée [255, 164] dans les études précédentes sur le diagnostic différentiel. Par conséquent, une analyse détaillée du dimorphisme de genre est justifiée pour trouver l'influence du genre sur les paramètres de la parole. A cette fin, nous proposons une méthodologie pour concevoir des marqueurs vocaux bidimensionnels qui permettraient une bonne discrimination entre PSP et MSA.

Au total, 15 mesures acoustiques sont calculées séparément à partir de monologues et de textes lus en langue tchèque par l'équipe SAMI de Prague. Nous poursuivons l'expérimentation pour étudier le dimorphisme de genre, la spécificité de la maladie dans les mesures acoustiques. Les indices de parole liés au sous-système de production de la parole et à la dysarthrie sont conçus pour le diagnostic différentiel des groupes PSP et MSA.

La plupart des composantes acoustiques individuelles ont montré un dimorphisme de genre. Peu de composantes acoustiques fournissent une différence de groupe entre PSP et MSA, ce qui n'est pas suffisant pour un diagnostic différentiel. Par conséquent, une combinaison de caractéristiques par sous-systèmes de la parole et de la dysarthrie a été tentée, ce qui a donné une discrimination encourageante entre les PSP et les MSA. Les patients atteints de PSP présentent des troubles prédominants dans la respiration et l'articulation par rapport à l'ASM, tandis que les patients atteints d'ASM présentent une plus grande sévérité dans la phonation et un indice d'articulation séparé. En ce qui concerne l'indice dysarthrique, les patients atteints de l'AMS présentent une ataxie prédominante alors que les patients atteints de la maladie de Parkinson présentent une hypokinésie plus importante. Les groupes PSP et MSA présentent tous deux une spasticité. Enfin, des caractéristiques acoustiques bidimensionnelles sont conçues et permettent d'obtenir une précision de classification de 88,63 %, une sensibilité de 89,47 % et une spécificité de 88 %.

La corrélation entre les indices de la parole et les sous-scores de gravité fournit également un modèle encourageant qui indique un diagnostic différentiel précoce. Cependant, des données supplémentaires sont nécessaires pour valider la présente observation.

0.7 Conclusion

Cette thèse a présenté un ensemble de caractéristiques de la parole de manière catégorique afin d'évaluer la production imprécise de voyelles et de consonnes et la parole spontanée pour les troubles parkinsoniens. Comme cette thèse vise à concevoir une méthodologie pour le diagnostic différentiel de la MP, de la PSP et du MSA, les paramètres de la parole sont considérés comme présentant au moins des propriétés distinctives pour les groupes de maladies. Des marqueurs vocaux fiables pour le diagnostic différentiel font en effet défaut à l'heure actuelle. La présente étude a également analysé le dimorphisme de genre de chaque caractéristique de la parole, ce qui faciliterait la conclusion finale sur la spécificité de la maladie. Un autre avantage de cette étude est que les caractéristiques vocales sont conçues par une approche basée sur la connaissance plutôt que par une approche récente basée sur les données,

ce qui serait plus acceptable dans la pratique clinique.

Acknowledgements

I would like to express my sincere gratitude to my thesis supervisors Dr. Khalid Daoudi and Dr. Hussein Yahia for constant guidance and advice which carried me through all the stages of my thesis work. Particularly, I would like thank Dr. Khalid Daoudi for giving me chance to work in Voice4PD-MSA (funded by ANR, France) project which motivated me to continue my reseach in a new and challenging domain.

I would like to express my gratitude to all my jury members for letting my defense be an enjoyable moment, and for your brilliant comments and suggestions: Dr. Régine André-Obrecht, Professeure émérite, Institut de Recherche en Informatique de Toulouse; Dr. Etienne Sicard, Professeur, l'INSA Toulouse.

I am also very much grateful to friends and colleagues who have supported me during my PhD tenure: Anass, Arash, Augusto, Rajkumar, Rupayan (TCS). I would like to give a special thank to Sabrina for all types of administrative supports.

I would also like to give special thanks to my wife and daughter, Anindita and Adrija and mother, Namita for their continuous support and understanding when undertaking my research and writing of my thesis. Finally, I would like to thank God, for letting me through all the difficulties and finish my degree.

Contents

1	Introduction	29
1.1	Organization of the thesis	31
2	State of the art	33
2.1	Parkinsonian disorders	33
2.1.1	Parkinson's disease	34
2.1.2	Multiple system atrophy	34
2.1.3	Progressive supranuclear palsy	35
2.2	Clinical differential diagnosis	35
2.3	Dysarthria	37
2.3.1	Hypokinetic dysarthria	37
2.3.2	Ataxic dysarthria	38
2.3.3	Spastic dysarthria	39
2.4	Imprecise vowel in parkinsonism	39
2.4.1	Laryngeal	40
2.4.2	Articulation	41
2.4.3	Resonance	42
2.5	Imprecise consonants in parkinsonism	43
2.5.1	Stop plosives	43
2.5.2	Fricatives	44
2.6	Imprecise syllables repetition in parkinsonism	45
2.7	Imprecise words in parkinsonism	46
2.8	Imprecise spontaneous speech in parkinsonism	46
2.9	Differential diagnosis by acoustic dimensions	46
2.9.1	Impaired vowels	47
2.9.2	Impaired consonants	48
2.9.3	Impaired diadochokinetic task	48
2.9.4	Impairment in spontaneous speech	48
2.10	Classification	49
2.11	Synthetic review of existing objective measures for Differential diagnosis	49
2.12	Objective	54
3	Acoustic Features	57
3.1	Respiration	58
3.2	Phonation	59

3.3	Articulation	63
3.4	Timing	63
3.5	Prosody	64
3.6	Nasalic	65
4	Database	67
4.1	Speech databases of parkinsonian disorder	67
4.2	Voice4PD-MSA	68
4.2.1	Recording setup	68
4.2.2	Recording protocol	70
4.2.3	Data recording	71
4.2.4	Recording details	72
4.2.5	Clinical details	73
4.3	Czech database (CzechData)	74
5	Vowel distortion in PD and MSA-P	77
5.1	Vowel distortion	77
5.2	Sustained vowels	79
5.2.1	Methodology	79
5.2.2	Results	80
5.3	Vowel space area: Reading text	88
5.3.1	Speech database: Voice4PD-MSA	88
5.3.2	Methodology	88
5.4	Result	90
5.5	Discussion	94
6	Consonant distortion in PD and MSA-P	97
6.1	Introduction: Consonants	97
6.2	Methodology	99
6.2.1	Database: Voice4PD-MSA	99
6.2.2	Data processing of logatomes	99
6.2.3	Methods to evaluate logatomes	100
6.3	Voiced obstruents	102
6.3.1	Devoicing analysis by Visual method	103
6.3.2	Devoicing analysis by objective analysis	106
6.3.3	VOT analysis of voiced plosives	108
6.3.4	Classification of PD and MSA-P	109
6.3.5	Discussion	111
6.4	Unvoiced obstruents	113
6.4.1	Introduction: unvoiced obstruents	113
6.4.2	Methodology	117
6.4.3	Results	119
6.4.4	Voice onset time (VOT)	125
6.4.5	Discussion on unvoiced obstruents	128

7	Speech disorder in diadochokinetic speech production	131
7.1	Introduction	131
7.2	Methodology	132
7.2.1	Database	132
7.2.2	Automatic speech segmentation	132
7.2.3	Acoustic features	133
7.3	Results	135
7.3.1	Phoneme segmentation accuracy	135
7.3.2	Acoustic analysis of DDK features	135
7.3.3	Classification	137
7.4	Discussion	138
8	Speech disorder in reading text	139
8.1	Introduction: Spontaneous speech	139
8.2	Methodology	140
8.2.1	Database	140
8.2.2	Manual segmentation	140
8.2.3	Prosodic features	141
8.3	Results	141
8.3.1	Prosodic features from manual segmentation	141
8.3.2	Prosodic features from Disvoice tool	142
8.4	Discussion	143
9	Differential diagnosis between PSP and MSA	147
9.1	Introduction	147
9.2	Methodology	149
9.2.1	Database	149
9.2.2	Acoustic feature	150
9.2.3	Acoustic feature analysis	150
9.2.4	Classification	152
9.3	Experimental result	152
9.3.1	Univariate analysis	152
9.3.2	Speech features by subsystems of speech	156
9.3.3	Speech features by dysarthria	171
10	Conclusions and future work	181
10.1	Summary and discussion	181
10.2	Limitation and future works	183
	Bibliography	186

List of Figures

2.1	Clustering of parkinsonian disorder according to parkinsonian syndromes and proteinopathy [195]	33
4.1	Recording scenario for database Voice4PD-MSA	69
4.2	Reduced reverberation of click sound after applying sound absorber .	72
4.3	Reduced reverberation of clapping sound after applying sound absorber	73
4.4	Comparison of noise properties for three microphones	73
5.1	Designed phonation feature (X1_a); * p<0.05, ** p<0.01, *** p<0.001, **** p<0.0001	82
5.2	Combination of articulation features from sustained vowel; ** p<0.001, ns: not significant	83
5.3	Designed nasalic feature; "ns" stands for not significant	84
5.4	Combination of articulation and nasalic features, X2_a from sustained vowel; **** p<0.0001, ns: not significant	85
5.5	Biplot of phonation feature (X1_a) w.r.t. articulation feature (X2_a) (dotted line represent decision thresholds)	86
5.6	Time-frequency representation of the sentence; F1 and F2 represent first and second formants	89
5.7	Vowel space using three corner vowels ('o' for /i/, '+' for /u/, '*' for /a/) for HC, PD and MSA-P	92
5.8	VSA, VAI, and FR features computed by Praat method; * p<0.05, ** p<0.01, *** p<0.001, **** p<0.0001, ns: not significant	93
5.9	Power in second formant of three vowels /i/, /u/, and /a/; ; * p<0.05, ** p<0.01, *** p<0.001, **** p<0.0001, ns: not significant	93
6.1	Manually labelling of phonetic unit of logatomes "berdo" and "quina"	100
6.2	Acoustic parameters for consonants evaluation	101
6.3	Acoustic parameters for vowels evaluation	101
6.4	Voiced stop plosives /b/, /d/, and /g/ from "berdo", "dirou", and "guizant" consequently; circled box represents voicing bar	102
6.5	Voiced fricatives /v/, /z/, and /Z/ from "vonia", "zacu", and "jinin" consequently; circled box represents voicing bar	102
6.6	Example of no/partial/total devoicing of /b/, /d/, and /g/ in a HC/PD/MSA-P (top). Example of normal/shorter/vanishing VOT of /g/ for the same HC/PD/MSA-P (bottom)	104

6.7	Example of no/partial/total devoicing of /v/, /z/, and /Z/ in a HC/PD/MSA-P	105
6.8	Biplot of $DVT(\%)$ w.r.t to $VOT_{/g/}$ and $VOTR_{/g/}$ (dotted line represent decision thresholds); (3) means that 3 MSA-P patients have same coordinates (total devoicing)	110
6.9	Decision tree using DVT and $VOT_{/g/}$ or $VOTR_{/g/}$ (in green) dimensions for discrimination between PD and MSA-P	111
6.10	DV, VOT, and VOTR of /g/ from CCV syllable /gR@/	111
6.11	Example of no/burst/multiple burst in /f/ in a HC/PD/MSA-P	119
6.12	Weak burst energy in /t/ for HC, PD and MSA-P subjects	120
6.13	Multiple burst in /k/ for HC, PD and MSA subjects	121
6.14	Example of vowel onset detection by MNCC method	122
6.15	Example of boundary detection between consonant-consonant (CC) combination using log filterbank feature in STM method	123
6.16	Average modified VOTR of unvoiced stop plosive (/p/, /t/, and /k/	127
6.17	Average VOT of unvoiced stop plosive (/p/, /t/, and /k/	129
7.1	Example of manual annotation of /pa-ta-ka/ (part)	133
7.2	An example of automatic vowel detection by MNCC method in /pa-ta-ka/ task	135
7.3	Group difference among HC, PD, and MSA-P using ataxic dimension (A_{ddk} from DDK features	137
8.1	Example of manual annotation of reading text (partial); annotation labels are represented as 0: pause, 1: respiration, 2: unvoiced consonants, 3: voiced speech	141
8.2	Group difference of HC, PD, and MSA-P by X1 designed by AST, EST, GVI, DVI, and DUS	143
9.1	Acoustic features and it's categorization	150
9.2	Plot of individual speech parameters which provide encouraging discrimination for PSP and MSA patients	154
9.3	Respiration index (F_r) for the 4 groups	157
9.4	First phonation index (F_{p1}) for the 4 groups	158
9.5	Second phonation index (F_{p2}) for the 4 groups	159
9.6	Correlation of F_{p2} feature w.r.t. overall severity	160
9.7	Articulation feature (F_a) for the 4 groups	161
9.8	Group differences between groups by prosodic index F_{pr}	162
9.9	Timing feature (F_t) for the 4 groups	163
9.10	Group differences with DDK index (F_{ddk}) for the 4 groups	164
9.11	Feature F1 for the 4 groups	165
9.12	Feature X2 for the 4 groups	166
9.13	Correlation of X2 w.r.t. overall and bradykinesia NNIPPS subscore	166
9.14	Biplot using X1 and X2 dimensions for four groups	168
9.15	Hypokinetic feature H_1 of 4 groups	171

9.16 Hypokinetic feature H_2 of 4 groups	172
9.17 Group difference of HC, PD, PSP and MSA by ataxic features A_1 and A_2	173
9.18 Spastic feature S of 4 groups	174
9.19 Group difference of four groups using feature $D1_{H_1S}$ and D2	176
9.20 Group difference of four groups using the modified feature D1; Biplot by D1 and D2	178

Chapter 1

Introduction

Speech is a unique, complex, dynamic motor activity that requires the integrity and integration of numerous neurocognitive, neuromotor, neuromuscular and musculoskeletal activities [69]. The deficit in any of these activities may be the cause of speech disorder, and disturb day-to-day life to a great extent. Thus analysis of speech disorder has immense advantage to identify the indication of neurologic disease as well as design specific speech therapy to improve quality of life. Naturally, the source of speech disorder is warranted for unmasking the complex underpinning of speech. Identification of underlying neurophysiologic bases is valuable for understanding nervous system organization for speech motor control, localization of neurologic disease and differential diagnosis, prevalence, and management.

Speech disorder due to neurodegenerative disease is commonly known as dysarthria. Dysarthria reflects abnormalities in the strength, speed, range, steadiness, tone, or accuracy of movements required for the respiration, phonatory, resonatory, articulatory, or prosodic aspects of speech production. The responsible neuropathophysiologic disturbances of control or execution are due to one or more sensorimotor abnormalities, which most often include weakness, spasticity, incoordination, involuntary movements, or excessive, reduced, or variable muscle tone [69].

The precise involvement of the brain area for speech production is not well understood. However, several studies were conducted to understand the functions of the brain in speech production. For instance, the study [234] demonstrated the involvement of the motor and premotor cortex, the cerebellum, the supplementary motor area (SMA), the superior temporal gyri, the tempoparietal cortices, and the anterior insula with left-lateralized activation in the putamen for speech production. In speech production strat with motor planning and sequencing, which is mainly accomplished by anterior insula [311, 241] and the SMA area [127]. Speech breathing is primarily attributed to a bilateral region of the sensorimotor cortex [201]. Likewise, the larynx motor cortex region has a relation with human vocalization [41]. Lobule VI of the posterior cerebellum is considered as an orofacial part of the cerebellum, which is activated by lip and tongue movements [97]. Different speech rate is documented by additional activation, like at the time of low speech rate left putamen is activated whereas activation of the cerebellum during higher rates [320]. These somatotopic maps could help elucidate the coupling of particular speech impairment

to its underlying brain area.

Parkinsonian disorder is an umbrella term. Parkinsonism refers primarily two subgroups: Parkinson's Disease (PD) and Atypical Parkinsonian Syndrome (APS). APS group consists of Multiple System Atrophy (MSA), Progressive Supranuclear Palsy (PSP), Corticobasal Degeneration (CBD), Neurodegeneration with brain iron accumulation (NBIA), Dementia with Lewy Bodies (DLB). Differential diagnosis of PD and APS remain a challenging task, particularly in early stage of the disease. In clinical domain, neurologists mainly rely on clinical motor signs and brain imaging techniques. However, imaging techniques are very complex and costly methods, and subjects need to be exposed to radiation. In contrast, speech is a noninvasive and cost effective mode, and in addition speech motor disorder is established as an early manifestation of neurological disorder. Analysis of speech disorders thus have immense scope to design clinical tool to investigate disease specificity.

As speech analysis is an inexpensive, non-invasive medium, it has attracted researchers and clinicians for decades. However, in clinical practice, speech analysis remains limited to perceptual assessment for a long time, a laborious job, and demands skilled professionals. Recent efforts towards acoustic methods to quantify perceptual judgments provided confidence to clinical practice. Acoustic methods can infer speech disorder by the visual display as well as by quantification of speech parameters. In addition, acoustic methods can detect additional speech abnormalities which are perceptually not identified. Automatic algorithms are also being developed to provide end-to-end solutions for assessing speech disorders.

Reduced speech quality naturally comes with aging onset by deformation in the oral cavity. According to the United Nations, the population of older adults is increasing very fast, and the total older adults 65 years or over will reach 1.5 billion by 2050. In the aging population, the parkinsonian disorder is becoming prevalent, which is attributed to unknown etiology. There is no treatment to cure parkinsonian diseases. Notably, early diagnosis of neurological conditions can improve quality of life [107, 22, 243]. As speech disorder is an early and prominent manifestation of neurological disorder, it would serve as a good marker [38] for disease diagnosis. Notably, discrimination of healthy subjects from neurological patients is not sufficient in clinical practice; differential diagnosis is also equally important to identify the particular type of neurological disease for respective medication and therapy. Likewise, getting high accuracy in differential diagnosis using complex speech parameters (unconventional) would not be accepted in clinical practice because speech markers must be interpretable. For instance, speech markers need to be explained by their physiological meaning, reflecting the degree of deficits in any particular part of speech production subsystems. In addition, it is required to check gender dimorphism, age dependency of the speech markers. Moreover, speech samples from different neurological disease groups are invaluable in motor speech disorder analysis and all the aspect speech parameters.

Automatic speech processing methods have been developed by speech researchers to compute speech parameters. Speech features design requires to consider different parameter like quality of speech sound, and mode of speech data acquisition. Acoustic features computed from speech data collected by uniform recording setup increase the

reliability in clinical practice. Variability in recording setup may introduce diversity in acoustic features even in same group of speakers. Furthermore, automatic method need to be accurate for designing acoustic features. As example, speech features computed from multi-syllable word, diadochokinetic task, reading text, and monologue particularly demand accurate segmentation of speech class like vowel, consonants, pause, and respiration. Furthermore, stop consonants analysis require to detect vowel onset and burst onset accurately. Several methods have been proposed for detecting speech targets, however those are mainly for healthy speakers. Thus it is required to evaluate existing methods for pathological speech, and develop additional methods where necessary.

Speech databases for parkinsonian disorders are limited by disease diversity, number of patients, and speech protocol (tasks). Discrete existence of speech databases related to neurological disorder was reported in several languages. Most of the databases only collected speech samples from PD group. However, to investigate differential properties, a database comprising all possible parkinsonian disease groups is essential. Language is another impediment for building global database. Many speech parameters may show disparity by language difference. Thus analysis of speech disorder first need to be restricted to particular language.

1.1 Organization of the thesis

As of now, the thesis is dedicated to abstract and introduction. The next chapter 2 summarized state-of-the-art, and objectives. The state-of-the-art describes clustering of parkinsonian disorder, types of dysarthria, and acoustic metrics for differential diagnosis. Chapter 3 presents acoustic parameters used for differential diagnosis of PD, PSP and MSA.

Speech database development procedure is described in Chapter 4. It describes different mode of data collection, data processing, and data quality assessment.

First part of the thesis is dedicated to the differential diagnosis of PD and MSA-P. Chapter 5 presents the methods to measure several speech parameters to assess vowel impairments for PD and MSA-P. This chapter first consider acoustic features from sustained vowel /a/ and presents differential properties. Next section is dedicated to vowel space area computation by three corner vowels. For vowel space area analysis, reading text speech task is considered. This section also describes manual formant frequencies measurement, failing of automatic approach, and gender disparity.

Description of consonants distortion for PD and MSA-P patients is presented in Chapter 6. The latter chapter presents the analysis of voiced and unvoiced obstruents in two sections. It describes the unique pattern of disorder in stop plosives and fricatives. Both temporal and spectral properties of obstruents are discussed in this chapter.

The Chapter 7 presents the pattern of disorder in syllable repetition task. It also describes automatic evaluation method of abnormality detection in syllable repetition.

The Chapter 8 presents the pattern of disorder in text reading task. It describes different acoustic components and its differential properties. It also discussed speech

segmentation procedure followed by automatic acoustic feature computation.

The second part describes the differential diagnosis of PSP and MSA patients. In Chapter 9, acoustic features from monologue and text reading are described. In addition, it represents the methodology of designing speech indexes for differential diagnosis of PSP and MSA. This chapter first analyzed individual acoustic features to find out group wise statistical difference and gender disparity. Next, individual acoustic features are combined according to subsystem of speech, and followed by dysarthria subtypes. After that classification of PSP and MSA was discussed.

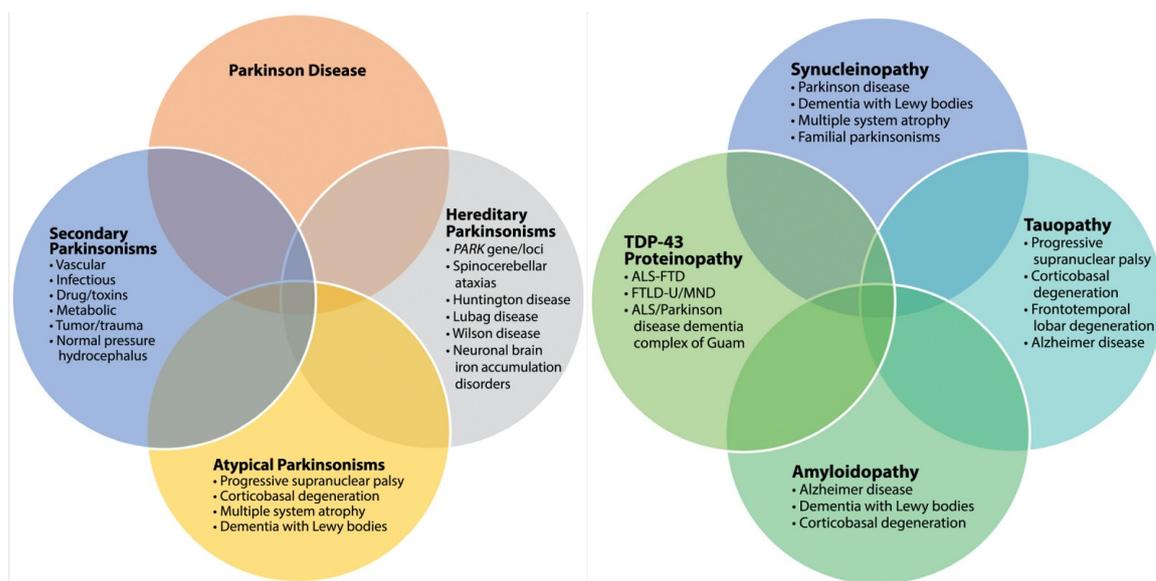
In the following chapter, we discuss state-of-the-art of clinical information about parkinsonism, different dysarthria subtypes related to parkinsonian disorder, clinical differential diagnosis, speech disorder in different speech subsystems, and existing acoustic features in differential diagnosis.

Chapter 2

State of the art

2.1 Parkinsonian disorders

Parkinsonism is a group of neurological diseases. The first clear medical description was written in 1817 by James Parkinson. In the mid-1800s, Jean-Martin Charcot refined and expanded this early description and disseminated information internationally about Parkinson's disease. He separated Parkinson's disease from other disorders characterized by tremor, and he recognized cases that later would likely be classified among the Parkinsonism-plus syndromes or Atypical Parkinsonian Syndromes (APS).



(a) Overlap of parkinsonian syndromes

(b) Clinicopathologic overlap of neurodegenerative proteinopathies

Figure 2.1: Clustering of parkinsonian disorder according to parkinsonian syndromes and proteinopathy [195]

Atypical Parkinsonian Syndromes includes subtypes like Progressive Supranuclear Palsy (PSP), Multiple System Atrophy (MSA), Huntington's disease, Corticobasal

Degeneration (CBD), Neurodegeneration with brain iron accumulation (NBIA), Dementia with Lewy Bodies (DLB), Multiple Sclerosis (MS), Amyotrophic Lateral Sclerosis (ALS) and few more. Detailed categorization of parkinsonian subtypes and proteinopathy is presented in Figure 2.1.

In this study, we mainly focus on neurological diseases PD, MSA, and PSP according to the availability of speech data. Hence, in the following section, three disease groups will be discussed.

2.1.1 Parkinson’s disease

Idiopathic Parkinson’s disease (PD) is the most common neurodegenerative disease after Alzheimer’s disease. The prevalence is 1.5% of the population over 65 years, and around 170,000 French are affected [298]. Given the general aging of the people, the prevalence is likely to increase over the next decade. PD is characterized by the progressive loss of dopaminergic neurons within the substantia nigra pars compacta (SNpc) due to intra-neuronal aggregation of α -synuclein in the form of Lewy bodies and Lewy neurites in the majority of cases [118]. The resulting imbalance of dopamine and acetylcholine disturbs the basal ganglia function, which participates in the planning, timing, control, and execution of muscle movements. There is no reliable biomarkers currently available for diagnosis of PD with acceptable sensitivity and/or specificity. Nonetheless, the clinical diagnosis requires the presence of cardinal motor deficits like bradykinesia (akinesia, hypokinesia), together with additional motor manifestation among rigidity, resting tremor (4-6 Hz) and postural instability [122, 51]. The clinical diagnosis is confirmed by a sustained response to dopamine replacement therapy. Clinical criteria have a sensitivity of 89% and a positive predictive value of 82% for a diagnosis of PD. In contrast, the definitive diagnosis is based on post-mortem confirmation of alpha-synuclein containing Lewy bodies. Autopsy studies showed that one-quarter of PD patients are misdiagnosed [123, 237]. Thus design of a reliable biomarkers is remain an ongoing research.

2.1.2 Multiple system atrophy

Multiple System Atrophy (MSA) is a relentlessly progressing rare neurodegenerative disease of unknown etiology. As per neurodegenerative proteinopathy clustering in the Figure 2.1, MSA is belong to synucleinopathies group. Together with PD, synucleinopathies group characterized by progressive cell loss in the brain due to abnormal aggregation of alpha-synuclein in neurons and glia [319].

MSA usually begins in the sixth decade [266, 299] and has a very poor prognosis; median survival ranges between 5.8 to 9.5 years [25, 14, 318]. Revised consensus criteria allow the clinical diagnosis of MSA with two degrees of certainty, ”possible” and ”probable”, while the diagnosis of ”definite” MSA requires post-mortem confirmation of alpha-synuclein containing glial cytoplasmic inclusions [89]. Another clinical diagnosis is confirmed by poor response to dopaminergic therapy [46]. Revised consensus criteria include brain imaging results as additional features for the diagnosis of ”possible” MSA. However, the sensitivity of these criteria remains relatively low for the

diagnosis of "possible" MSA and requires further improvement [324]. Depending on the leading presentation of the motor impairment, revised consensus diagnosis criteria distinguish between MSA-P where parkinsonism predominates and MSA-C where cerebellar symptoms are most prominent [244, 236]. MSA-P accounts for two-thirds of cases in Western populations [153, 167, 90]. MSA group is characterized by a variable combination of parkinsonism, cerebellar impairment, autonomic failure, and pyramidal tract signs [90]. Non-motor signs and symptoms of MSA are gastrointestinal, cardiovascular, and urogenital abnormalities [244]. Motor sign and symptoms of MSA include bilateral rigid-akinetic form, early falls, ataxic gait, limb ataxia, and cerebellar oculomotor dysfunction [162].

2.1.3 Progressive supranuclear palsy

PSP is also a progressing rare neurodegenerative syndrome characterized by postural instability, axial rigidity, supranuclear gaze palsy, mild dementia, and pseudobulbar palsy with consistent pathological findings defined by an accumulation of tau protein and neuropil threads, mainly in the pallidum, subthalamic nucleus, red nucleus, substantia nigra, pontine tegmentum, striatum, oculomotor nucleus, medulla, and dentate nucleus [202]. Symptoms of PSP are most commonly seen in people in their early 60's, but may begin in some people who are in their 40's with prevalence ranges from 5 to 6.4/100000 [266, 203]. PSP has average survival ranging from 6 to 8.6 years [50, 10, 29]. PSP phenotypes from recent clinical presentation are Richardson's syndrome (PSP-RS), parkinsonian variant (PSP-P), and pure akinesia with gait freezing (PSP-PAGF). The heterogeneity of PSP has recently been examined in detail in post-mortem series to define pathological substitute parameters based on the extent and pattern of tau pathology to distinguish RS from PSP-P [321, 132].

Early symptoms of this disease may be related to a person's increased difficulty with walking and balance, often resulting in frequent falls. It is common for a person in the early stages of PSP to develop other motor-related symptoms like slowed or awkward movements while walking [159]. Signs that help to differentiate PSP from other neurodegenerative diseases, like Parkinson's, are often related to a person's vision and eye movements. People with PSP often experience blurred vision and an inability to control eye movements. Some cannot look downward or cannot open their eyelids [159]. Speech and swallowing complaints are seen at an early stage of the disease. Mental and physical slowness is also observed for this disease but not in the early stage.

There is neither confirm diagnostic test for this disease nor a specific treatment. However, it can be differentiated from PD by the inadequate response of levodopa, lack of tremor, and gait instability leading to falling within the first year of the illness.

2.2 Clinical differential diagnosis

Differential diagnosis is the distinguishing of a particular disease from others that present similar clinical features. In the early stage of disease, PD, PSP, MSA patients

manifest similar clinical signs [324]. Thus discrimination of PD patients from Healthy Control (HC) is not enough in clinical practice. Diagnosed PD patient might be a PSP or MSA patient. Hence, differential diagnosis is very much crucial for prognosis and therapeutic planning. Moreover, early diagnosis is essential because disease progression in the APS group is comparably rapid than PD [218]. Definite diagnosis of PD, PSP, MSA is only based on a post-mortem pathological confirmation. However, in clinical practice, a definite diagnosis cannot be reached for apparent reasons, and clinicians have to rely on the suggested clinical diagnostic criteria, which include clinical features and neuroimaging features [89]. Clinical features characterized by bradykinesia, rigidity, tremor and postural instability are predominant in PD patients. PSP patients differentiate from PD patients by six clinical features, axial rigidity, symmetry, extended posture, backward falls, absence of postural tremors in the upper limbs, and lack of response to levodopa [177, 188, 187]. Clinical features for a diagnosis of MSA consist of autonomic failure in combination with motor symptoms. Autonomic failure in MSA includes cardiovascular dysfunction, genitourinary dysfunction, thermoregulatory and sudomotor dysfunction, fecal incontinence and constipation, and sleep-disordered breathing [289, 52], among which orthostatic hypotension or urinary symptoms are required for the diagnosis. Motor symptoms include poorly levodopa-responsive parkinsonism or cerebellar ataxia. In clinical diagnostic criteria, different rating-based measures were developed to evaluate cardinal disorders. United Parkinson's Disease Rating Scale (UPDRS) tool assesses PD patients and their response to treatment. On the other hand, PSP and MSA patients are rated by the natural history and neuroprotection in Parkinson plus syndromes–Parkinson plus scale (NNIPPS) [223]. A high value in rating score signifies more severity in cardinal motor sign.

Finding reliable biomarkers for the differential diagnosis between PD and APS (PSP and MSA) remains a very challenging task. However, a variety of imaging techniques such as Magnetic Resonance Imaging, Diffusion Tensor Imaging, Positron Emission Tomography, Single-photon Emission Computed Tomography, and Transcranial Sonography may be used in the assessment of various parkinsonian syndromes [28]. In particular, automatic image-based classification based on metabolic patterns is highly accurate in distinguishing between PD, PSP, and MSA patients at early stages of the disease, with more than 84 % sensitivity and 94 % specificity [296]. Magnetic resonance imaging (MRI) of the brain may help the clinician reveal distinct abnormalities in MSA patients [129]. Different patterns of nigro-striatal involvement in PD and MSA by using multimodal MRI techniques were reported in the study [20]. Multimodal MRI technique also observed multi-parametric modifications within the cerebellum and putamen in both MSA-C and MSA-P patients, compared to PD patients [235]. However, brain MRI can also be expected, especially in patients where the differential diagnosis between PD and MSA is difficult. Other imaging techniques such as [18F]-fluorodeoxyglucose positron emission tomography (FDG-PET) allow identifying distinct metabolic patterns in PD, and MSA [206]. However, this technique is very costly and not available in clinical routine. Besides imaging techniques, several studies have compared plasma, and cerebrospinal fluid levels of alpha-synuclein, markers of axonal degeneration and catecholamines between PD and

MSA patients [172]. No significant conclusions can be drawn from these, and further efforts are urgently needed to improve diagnostic accuracy between PD and APS (and within APS). Furthermore, techniques are preferred to be cost-effective and less manual intervention.

Speech disorder is also an early and prodromal marker in neurological diseases [144, 231]. Speech disorders primarily develop in the majority of patients with parkinsonian disorder [116, 158]. Speech disorder due to degeneration in neural structure is called dysarthria. Manifestation of differential dysarthria was observed in previous studies [160, 158, 247]. Description of dysarthria and its subtypes are summarized in the following section.

2.3 Dysarthria

Dysarthria is an umbrella term for a group of neurologic speech disorders that reflect abnormalities in the strength, speed, range, steadiness, tone, or accuracy of movements required for the breathing, phonatory, resonatory, articulatory, or prosodic aspects of speech production. The responsible neuropathophysiologic disturbances of control or execution are due to one or more sensorimotor abnormalities, which most often include weakness, spasticity, incoordination, involuntary movements, or excessive, reduced or variable muscle tone [69]. According to underlying localization, sensory actions (execution, control, coordination), the pattern of physiological symptoms, a total of seven different types of dysarthrias are defined by Darley, Aronson, and Brown (DAB). Table 2.1 shows some differential information about those dysarthria subtypes [69].

Out of the above mentioned dysarthria subtypes, hypokinetic, spastic, and ataxic subtypes are frequently analyzed due to their prevalent manifestation. Individual dysarthria subtypes manifest predominant speech disorder either in speech subsystem or in speech dimensions.

2.3.1 Hypokinetic dysarthria

Hypokinetic dysarthria is a perceptually distinct motor speech disorder (MSD) associated with basal ganglia control circuit pathology. This disorder reflects the effect of rigidity, reduced force and range of movement, and slow individual but sometimes fast repetitive movements on speech. Decreased range of movements is a significant contributor to this disorder. The etiology of hypokinetic dysarthria is related to abnormalities with basal ganglia circuit function. This include degenerative, vascular, traumatic, infectious, inflammatory, neoplastic, and toxic-metabolic diseases. The exact distribution of causes of hypokinetic dysarthria is unknown.

Respiratory abnormalities frequently occur in hypokinetic dysarthria. Abnormalities in the respiratory system reflected as reduced maximum phonation time, reduced airflow volume during vowel prolongation, fewer syllables per breath, shorter utterance length, increased breathing at the time of reading. Some of these characteristics are also related to laryngeal function abnormalities [61, 121]. Other distinctive im-

Type of dysarthria	Origin	Primary speech level abnormality	Evident characteristics	Cause of dysarthria
Flacid	Lower motor neuron	Phonatory and articulatory	Predominantly neuromuscular execution	Weakness
Hypokinetic	Basal ganglia circuit	Manifest in respiratory, phonatory, resonatory, and articulatory	Rigidity, reduced range of movements. It can be characterized by reduced movements	Control of proper background tone and supportive neuromuscular activity, and speech motor control
Spastic	Damage to the direct and indirect activation pathways of central nervous system	Manifest in respiratory, phonatory, resonatory, and articulatory	Weakness and spasticity that slows movement and reduces its range and force	Predominantly neuromuscular execution
Ataxic	Damage to the cerebellar control circuit	Manifest in respiratory, phonatory, resonatory, and articulatory, but its characteristics are most evident in articulatory and prosody	Timing and coordination	Problem of motor control
Hyperkinetic	Basal ganglia circuit	prosody and rate	Predominantly neuromuscular control	Involuntary movements
Unilateral upper motor neuron	Unilateral upper motor neuron	Articulation, phonation, and prosody	Execution/control	Upper motor neuron weakness, incoordination, and spasticity
Mixed	mixed	Any of the subsystem of speech production	Execution and/or control	More than one

Table 2.1: Major types of dysarthria and it’s characteristics

paired speech dimensions are monopitch, reduced stress, monoloudness, inappropriate silences, short rushes of speech, variable rate, and imprecise consonants [58]. Another study [183] stated that out of 200 PD patients, 89% of patients had voice abnormalities characterized by hoarseness, roughness, tremulousness, and breathiness, and 45% had articulation problems. In addition, 20% of patients had rate abnormalities characterized by syllable repetitions, shortened syllables, lengthened syllables, and excessive pauses, and 10% percent were hypernasal. Electromyographic (EMG) study also documented undershooting of articulatory movements [99].

2.3.2 Ataxic dysarthria

Ataxic dysarthria is also a perceptually distinct MSD associated with the cerebellar control circuit. The disorder reflects the effects of incoordination and perhaps reduced muscle tone, resulting in slowness and inaccuracy in the force, range, timing, and direction of speech movements. It is primarily an abnormality of timing and coordination. Hypotonia can occur in cerebellar disease. It is characterized by excessive pendulousness. Again, it is due to reduced muscle tone. Overshooting (Dysmetria) is observed many times in movements. Dysmetria and dysdiadochokinesis are also common symptoms of ataxia due to poor control, timing and coordination.

Ataxic dysarthria is distinguished from other types of dysarthria groups by irregular articulatory breakdowns, telescoping, irregular speech AMRs, excess and equal stress, excess loudness variation, and distorted vowels speech dimensions. The stability of long-term and short-term phonation are found to be abnormal for cerebel-

lar disease. It has been speculated that asymmetrically distributed motor deficits at laryngeal level and altered sensory control of laryngeal and respiratory reflexes could account for impaired control of tension in intrinsic laryngeal muscle, leading to phonatory instability. Physiological investigation of nonspeech respiratory function (spirometry) has shown that some ataxic speakers have reduced vital capacity [3, 200]. Incoordination of respiratory and phonatory section of speech for isolated phonatory task lead to excessive loudness variation and fundamental frequency variation [194].

2.3.3 Spastic dysarthria

Spastic dysarthria is a perceptually distinct MSD produced by bilateral damage to the direct and indirect activation pathway of Central Nervous System (CNS). It is characterized by weakness and spasticity, which slows movements and reduces its range and force. Spasticity, a hallmark of Upper Motor Neuron (UMN), seems to be the major contributor to this dysarthria. Spasticity is primarily a problem of neuromuscular execution, rather than planning, programming, and control. In general, damage in direct and indirect activation pathways bilaterally can be the reason for spastic dysarthria.

Spastic dysarthria is distinguished from other types of dysarthria by strained-harsh voice quality, monopitch and monoloudness, slow speech rate, and slow and regular speech Alternate Motion Rates (AMRs) speech dimensions [58, 186]. Physiological investigation showed abnormality in respiration [56], laryngeal [17, 326], velopharyngeal functions [326].

Above mentioned dysarthrias are commonly observed in any or all of the subsystem components of speech. To assess the particular deficits in subsystem, a particular speech task needs to be designed. In previous decades, several speech tasks were developed, such as sustained vowel for laryngeal activities, dyadochokinetic task (/pa-ta-ka/, /ba-da-ga/, /pa-pa-pa/) for evaluating articulatory movements, monologue, and text reading to assess overall prosody and timing of speech. The following part will discuss the speech task and its related subsystem's impairment and available acoustic dimensions.

2.4 Imprecise vowel in parkinsonism

Analysis of vowels can reveal functionalities in laryngeal, articulatory, and velopharyngeal activities. Vowels are generated by vocal folds vibration followed by a particular articulator's position in the vocal tract. Possible deficits may be found in either vocal folds vibration, articulators stability, or velopharyngeal control according to neurological disease and/or other deficits. In the perceptual investigation, several acoustic features were defined by Mayo clinical dysarthria studies, such as abnormal pitch, pitch break, monopitch, voice tremor, harsh voice, and distorted vowels. Physiological and visual methods also investigated impairment in laryngeal action, and articulatory positioning for better explanation. For example, videolaryngostroboscopy studies have shown that as Parkinson's disease progresses, glottic competence and vocal fold

vibration are compromised, with a bowed closure configuration, phase asymmetry, aperiodicity, voice tremor, and mucosal wave abnormalities [285, 226, 308]. However, this visual method is an invasive procedure that causes discomfort to the patient. Another study showed that laryngeal electromyography (LEMG) can better explain the immobility of vocal folds vibration than laryngostroboscopy. Although, vowel production depends on normal vocal folds vibration, it also depends on sufficient airflow from lungs and precision of articulator's positions. In addition, insufficient airflow may force to compromise vocal folds vibration. On the other hand, imprecise articulator position will change the vowel properties (resonance properties).

2.4.1 Laryngeal

Laryngeal activity (mostly related to pitch) was reported normal for untreated PD patients at early stage [117, 135]. Conversely, the impaired motion of the vocal cord happens early in the disease process, accounting for the early development of laryngeal symptoms such as dysphonia, stridor, and sleep apnea for MSA patients [169]. Latter mentioned, laryngeal symptoms are partly attributed to paralysis and atrophy of the vocal cord abductor [128]. Another study reported that dystonia, rather than paralysis, of laryngeal muscles play a more pertinent role in laryngeal pathology of MSA patients [196] and PSP patients [21]. PD patients manifest laryngeal symptoms at the later stage of disease [117].

Laryngeal impairment is primarily evaluated by harsh voice, creaky voice, excessive pitch variability, reduced pitch variability perceptually and acoustically. In literature, sustained vowels (/a/, /i/, and /u/) and/or spontaneous speech (monologue or reading text) protocols were used to assess deficits in vocal folds vibration. Studies [275, 197, 160, 158] perceptually evaluated voice distortion by hoarseness, tremulousness, reduced loudness, whispery or scratchy voice. From sustained vowels, harsh voice is primarily evaluated by conventional acoustic measures like jitter (perturbation in cycle-to-cycle duration), shimmer (perturbation in cycle-to-cycle amplitude), Harmonic-to-Noise Ratios (HNRs). According to clinical characteristics, harsh voice speech dimension is classified as hypokinetic dysarthria [58]. Variable vocal folds vibration results in increased jitter and shimmer, whereas incomplete vocal folds closure produces reduced HNR. Several studies used harsh voice speech dimension to assess speech disorder [249, 275, 247, 138, 164]. In those studies, Praat [35] software toolkit was mostly used to compute jitter, shimmer, and HNR. Deficits in vocal folds vibration were also evaluated by micro label acoustic measures, Quasi-open quotient (QOQ), and normalised amplitude quotient (NAQ) [101, 217].

Excessive pitch variability computed as the standard deviation of pitch contour from sustained vowels can reveal vocal folds control status. Due to more significant ataxia in patients, excess pitch variability was observed in sustained vowel [247, 255]. On the other hand, reduced intonation is computed by monopitch (reduced pitch variability), which is computed as the standard deviation of pitch contour from reading text or monologue [281, 247].

The vocal tremor was also analyzed from sustained vowel, which may reflect the modulating frequency in pitch and intensity [42, 44]. Latter study proposed six acous-

tic measures to evaluate vocal tremors. The frequency tremor frequency (FTrF) is the frequency of the strongest low-frequency modulation of the fundamental frequency (F0), amplitude tremor frequency (ATrF) is the frequency of the strongest low-frequency modulation of the amplitude (intensity). The frequency tremor intensity index (FTrI) is the magnitude of the strongest low-frequency modulation of F0, the amplitude tremor intensity index (ATrI) is the magnitude of the strongest low-frequency modulation of amplitude (intensity). Similarly, two more acoustic measures, frequency tremor power index (FTrP) and amplitude tremor power index (ATrP), were also introduced to assess vocal tremor. Later on, FTrI was used in the study [247, 88]. Nasolaryngoscopy suggests that Parkinson’s disease voice tremor is not associated with the vocal folds and may involve the palate, the global larynx, and the arytenoids tremor in the vertical larynx on /a/, and tremor in the arytenoid cartilages on /s/ [86]. PD patients did not show significant difference from HC in speech tremor measures [44].

Spasmodic dysphonia causes involuntary spasms in the muscles of the voice box or larynx causes the voice to break, and have a tight, strained, or strangled sound is a predominant characteristic of spastic dysarthria. Strained-strangled voice is evaluated by Degree of voicelessness (DUV) [247] and Subharmonic to harmonic ratio(S2H) [112, 164]. A sustained vowel speech task assesses this particular speech disorder. The capability of vocal folds abduction and adduction was evaluated by gapping in between voiced interval (GVI) and duration of voice duration (VDI) by exploiting monologue and reading text [111]. PD patients did not show group difference from HC in GVI and DVI.

2.4.2 Articulation

Stability and accurate range of articulators movements are essential for particular vowel production. At the time of sustained vowel, involuntary movements of articulators may cause variability in resonance characteristics of the speech over time. The standard deviation of the power spectral density (stdPSD) was designed to capture the increased variability of the spectrum and thus the severity of involuntary movements [130].

In general, vowel characteristics are represented by first formant (F1) and second formant (F2) frequencies. Imprecise movement (reduced range of movement) of articulators (tongue, jaw, and lips) may result in imprecise oral shape, changing the F1 and F2 frequencies. Literature [300] stated that the tongue position mainly defines frequencies of F1 and F2. F1 frequency is inversely related to the height of the tongue, whereas the F2 frequency is directly related to the advancement of the tongue position. For example, F1 increases while the tongue moves forward, and F2 decreases as the tongue moves backward. F1 decreases with the elevation of the tongue and increases as the tongue is lowered or downward movement of the jaw. In addition, F1 and F2 decrease while lips are rounded and increase when the lips are unrounded [146]. The Vowel Space Area (VSA) is a conventional acoustic proxy for the kinematic displacements of the articulators [148, 26]. Two different approaches were reported in literature to measure VSA e.g, triangular VSA (tVSA) [178, 283, 277] and quadrilat-

eral VSA (qVSA) [92, 170, 77]. For both variants, the VSA is calculated as the area formed by connecting the corner vowels (triangular vowels: /a/, /i/, /u/ and quadrilateral: /a/, /i/, /u/, /ae/) using the Euclidean distance between each coordinate in F1-F2 space. Besides traditional VSA measures, alternative acoustic measures were also proposed to analyze imprecise vowel articulation in prior studies. For example, the Vowel Articulation Index (VAI) was proposed in the study [245, 283, 277]. The reciprocal measure of VAI, Formant Centralization Ratio (FCR), was proposed in the study [260]. Those described above acoustic metrics use formant values (F1 and F2) to examine vowel articulation. In another study, an automated VSA assessment from connected speech has been proposed to improve the accuracy of vowel space measurement. This method measures the peripheral vowel space area of formant frequency data using a convex-hull algorithm [259]. Another acoustic measure called $F2_i / F2_u$ which represents ratio of second formant frequencies of vowel /i/ to /u/ was proposed to assess articulators movements [262]. Above mentioned vowel space related acoustic features are belong to hypokinetic dysarthria. Different speaking tasks have great impact on VSA measure [92, 302]. The study [92, 302] showed that clear speech provides larger VSA compared to conversational speech tasks. Sustained vowels were also used for vowel space area measure, which reported contradictory results. The study [18] stated that VSA from sustained vowels could reveal articulatory deficits. In converse, another study [250] showed that sustained phonation was not suitable for VSA analysis. Interestingly, vocal tract length varies with gender, yielding different formant frequencies for male and female groups. Hence, It is required to consider gender dimorphism and speaking tasks while using vowel space related features to assess articulator's deficits.

2.4.3 Resonance

Hypernasality is another manifestation of palatal immobility, slow movement, and incomplete velopharyngeal closure. Velopharyngeal assessment comprised of perceptual speech evaluation and functional imaging, including video nasendoscopy and speech videofluoroscopy. The slow movement of soft palate or incomplete velopharyngeal closure allows air emission through the nasal cavity during the sustained vowel production, leading to nasalic sound. Irregular nasality including both hyponasality and hypernasality is attributed to involuntary movements in soft palette [190]. The perceptual analysis found the least frequent resonance (hypernasality) in PD patients [184]. However, the latter study used text reading tasks and used nasal consonants (/m/, /n/). Therefore, hypernasality detection in sustained vowels would be suitable for measuring deficits in velopharyngeal functions. In objective measure, hypernasality is assessed by two objective measures, voice low tone high tone ratio (VLHR) and the 1/3-octave spectra analysis [142, 174]. Within these two methods, the 1/3-octave spectra analysis has been shown to be more sensitive [309] and successfully used for dysarthria analysis [227, 210, 211]. Average energy over all windows (E_{fn_M}) in 1000 Hz frequency band was used as a marker of nasality [210, 130]. In addition, fluctuation in nasality was estimated as the standard deviation of values measured across all windows (E_{fn_SD}) [85]. High variability of nasality can be a manifestation

of cerebellar deficits.

2.5 Imprecise consonants in parkinsonism

The articulatory subsystem of speech production includes mainly lips, jaw, and tongue, transforming air stream into particular sound by constriction. Consonants production require precise, rapid movements of articulators [32, 2]. Physiological correlation with dysarthria were studied to find disrupted movements of articulator. Lip muscle movement disorder was observed by Electromyographic (EMG) analysis which suggests the presence of increased activity of antagonist muscles in certain motoric activities that would serve to balance the action of agonist muscles during voluntary movements [173]. Deficits in jaw movement as example, decrease in amplitude and velocity during jaw opening and closing, aberrant patterns and low amplitude of EMG activity during clenching, and low vertical amplitude and prolonged durations of occlusion during rhythmic movements was observed for parkinson's disease patients [242]. If force and range of motion are excessive, structures may overshoot targets. If force and range of motion are decreased, target undershooting may occur. If timing is poor, the direction and smoothness of movements may be faulty. In a recent study [292], electromagnetic articulography (EMA) was efficiently used to examine both extra-oral and intra-oral articulatory movements (initiation and coordination) in real time and 3D manner. The latter study showed physiologically that dopamine medication improved the articulator's initiation and coordination. Another similar study [322] observed a reduced range of movement of the tongue at the time of articulation.

According to the manner of articulation, consonants are clustered as stop plosives, fricatives, affricative, glide, lateral, and nasal. The classic study by [182, 58] perceptually found imprecise consonant articulation to be one of the most deviant speech dimensions in PD. It was found that the most affected phoneme class are stop-plosives, affricates, and fricatives. Particularly stop plosives pronunciation require accurate and rapid (impulsive) complete constriction in the vocal tract by lips and tongue [32]. On the other hand, precise movement of articulator and it's steady position is mandatory for fricative sounds [182]. Voiced stop plosives and fricatives are complex in characteristics that involve precise coordination of laryngeal and articulators activity.

2.5.1 Stop plosives

During the production of stops, acoustic pressure is built up behind a closure at a place within the vocal tract, resulting in a silent interval or a low level acoustic signal, with or without voicing. When the pressure is released suddenly, it introduces a relatively high energy burst or transient in the acoustic signal, spanning a short interval followed by aspiration. The instant in the acoustic signal corresponding to the sudden release is referred to as the "burst-onset" or the closure-burst boundary or the Closure Burst Transition (CBT) [151, 290].

Impaired stop consonants articulation in various neurological diseases has been

assessed perceptually in subgroups of dysarthria [47, 57, 182, 106]. Inadequate lip, tongue elevation may realize stop consonant to fricative (generally called spirantization). In addition, stop plosive's place of articulation may also be changed, and voiced (voiceless) stops may be realized as voiceless (voiced) [67]. Incomplete closure of stop consonants was visually analyzed in the study [15]. The presence of multiple transients is observed as another disorder in stop plosives [222].

In the objective measure of imprecise consonant articulation, various measures of duration, formant transitions, spectral moments, or energy-based measures have been proposed [151]. To produce the stop consonant-vowel sequence, the release of vocal tract occlusion and initiation of glottal vibration need to be synchronized. The time lag between stop consonant burst and vowel onset Time (VOT) is defined as Voice Onset Time (VOT). VOT is the durational measure of the timing of orofacial and laryngeal events [143, 6]. Thus, VOT measured from stop plosives perhaps the most frequently used parameter to evaluate imprecise consonant articulation. In previous studies, automatic VOT estimation was attempted in time domain [102, 12] and frequency domain [180, 208, 291, 176, 204]. Plosion Index (PI) and Maximum Normalized Cross-Correlation (MNCC) methods were proposed for CBT detection in continuous speech [12]. Spectral domain approaches were mostly based on sudden band energy transition [180, 208, 291, 176]. Another different approach based on Single Frequency Filter (SFF) followed by phase reconstructed signal was developed to detect the burst [204]. PI method provided better CBT detection compared to TEO based technique [102] and band energy related measures [180, 208, 176]. Very few studies related to automatic CBT detection or VOT measure are available for pathological speech [209]. Otherwise, most of the previous studies [75, 73, 306, 16] computed VOT from manual labeling to analyze neurological disorders. Besides the durational measure, some of the objective measures, spectral tilt (difference of energy between lower and upper frequency band) and intensity difference (difference between maximum energy of stop to a maximum energy of the following vowel), were also proposed in the study [272, 120]. Study [209] proposed consonant spectral trend (CST), consonant spectrum moment (CSM) for consonant articulation precision, and formant trend for tongue movement. The latter study observed significant disorder in PD patients compared to HC.

2.5.2 Fricatives

Fricatives are produced by partial constriction of articulators and present/absence of vocal folds vibration for voiced and voiceless fricatives. Fricatives are clustered by manner of articulation (voice or voiceless) and place of articulation (bilabial, alveolar, palatal) Labio-dental (/f/, /v/) fricatives show relatively flat spectra below 10 kHz with no dominating peaks [189]. Alveolar fricatives (/s/, /z/) are characterized by spectral energy (above 4 kHz) [124] and major peaks (3.5–5 kHz, [23]; 6–8 kHz, [136]) at higher frequencies. Palato-alveolars (/ʃ/, /ʒ/) are characterized by spectral energy in 2 – 4 kHz; [124, 23], which display increased relative amplitudes.

Most frequent impairment in fricatives is reduced sharpness, due to reduced range of movement of articulators. By physiological method, electropalatography (EPG)

reveals imprecise articulators movement for parkinsonian disorder [192] in /s/ production. Peak frequency measures the sharpness of fricatives [305, 79, 268], spectral moments which describe the central tendency, dispersion, tilt, and peakiness of the spectrum [110], and fricative's amplitude to the following vowel [48]. Voicing in voiceless fricatives is another impairment which manifest deficits in fricatives production and laryngeal synchronization [15]. Overshooting of articulators result in presence of burst/transient in fricatives which was visually observed in the study [15].

2.6 Imprecise syllables repetition in parkinsonism

Dysdiadochokinesis is a manifestation of decomposition of movement (dyssynergia), which refers to errors in the timing and speed of components of a movement, with resultant poor coordination can elicit by testing alternating repetitive movements (AMRs). To assess speech AMRs, /pa-ta-ka/ [229, 185] and /ba-da-ga/ syllable repetition tasks were designed, which was recorded in normal and rapid style. Syllable rate, irregularity in syllable repetition, and sudden increase in rate are the most common acoustic measures in syllable repetition. Imprecise articulation was observed in oral DDK [7]. The perceptual analysis yields greater impairment in syllable /ka/ for PD patients [141]. The authors of the study [265] concluded that cerebellar is primarily responsible for maintaining the precision of timing interval, whereas basal ganglia rather serve to maintain rhythm stability over time. The latter study observed variability in DDK pace for PD and spinocerebellar ataxia.

The syllable rate is computed as a number of syllables per second. Spastic dysarthria in speaker results in reduced syllable rate [71]. To compute the syllable rate, the major challenge is to correctly count the number of syllables. In the study [274], oscillographic acoustic pressure signal was used to identify the number and duration of a syllable. Automatic syllable detection method in /pa-ta-ka/ was proposed in the study [209]. The latter method was used to compute rate of DDK (DDKR) in the study [247, 255]. However, several factors can reduce the accuracy for vowel onset and offset, and burst onset detection. The latter method was particularly designed for rhythmic syllable train. It may fail to detect syllable duration in irregular syllable repetition. In addition, for vowel onset detection accuracy was 81.7% by 5ms threshold using Bayesian Step Change point Detector (BSCD) [246]. Burst onset detection accuracy was 79.2%. Both tasks have further scope to improve the accuracy considering short duration of labial stops.

Irregularity in syllable rate was computed as the standard deviation of syllable duration [247, 255]. Another study [273, 252] proposed Coefficient of Variance (COV) of syllable duration to measure instability in syllable repetition. The latter study observed instability even in mild speech motor disorder of PD patients. Instability of power in syllables is also regarded as another speech disorder that characterizes poor respiratory-phonatory coordination and control [82, 69]. The standard deviation of power was automatically computed from DDK task [130]. Besides the PD, other disease groups like MSA, PSP, HD, CA showed power instability [130].

Syllable duration and vowel duration also used to assess the prolongation of syl-

lables of a vowel. Previous studies [278, 130] used vowel duration from DDK task to evaluate the slowness of repetitive movements with excessive vocal emphasis typical of ataxic dysarthria. APS group manifested significant vowel prolongation compared to HC, whereas PD group did not show the trend [112].

2.7 Imprecise words in parkinsonism

The use of single word as recording protocol is very much rare in literature. The study [306] used "CVtka" kind of token to assess impaired consonants employing VOT. In another study [264], simple bi-syllabic (Consonant-Vowel) combination of meaningful words were used to evaluate articulators movements in five place of articulation (velars, palatals, retroflexes, dentals, bilabials). The later study analyzed VOT, formant frequencies, F2 transition, mean intensity of consonants. The latter study did not show concrete trend for disease groups.

2.8 Imprecise spontaneous speech in parkinsonism

Spontaneous speech requires the involvement of all the subsystems of speech production (respiration, phonation, articulation, timing, prosody). In the spontaneous speech task, two types of speech recording were considered in general: reading text and monologue. In perceptual analysis, spontaneous speech recordings were mainly used for analyzing intonation and prosody. Previous studies computed monopitch, monoloudness, inappropriate silence from spontaneous speech task [281, 247, 125]. On the other hand, net speech rate (words per second) is mostly computed from text reading tasks because of prior knowledge about the number of words. In a novel study [114, 306], several acoustic dimensions were proposed to capture deficits in subsystem of speech production. Respiration features include Rate of Speech Respiration (RSR), Pause Intervals per Respiration (PIR), Relative Loudness of Respiration (RLR), and Latency of Respiratory Exchange (LRE). Phonation features include Gaping in-between Voiced intervals (GIV), duration of voiced intervals (DVI). The articulatory subsystem consists Duration of stop consonants (DUS) and decay of unvoiced fricatives (DUF). Timing and prosodic group consist of Rate of Speech Timing (RST), Acceleration of Speech Timing (AST), Duration of Pause Intervals (DPI), and entropy of speech timing (EST). Prosodic parameters also capable to capture overall abnormality in syllable repetition task. The study [278] used total speech time, total pause time, and percentage of pause time within polysyllabic words (PRinw) to assess text reading task. PD patients showed significant abnormalities in net speech rate, and PRinw using compared to HC.

2.9 Differential diagnosis by acoustic dimensions

This section will discuss differential diagnosis of PD, PSP, and MSA by acoustic dimensions. Acoustic features computed from vowel, consonants, syllable repetition,

spontaneous speech will be explored sequentially in this part.

2.9.1 Impaired vowels

Harsh voice parameters (jitter, shimmer, HNR) was used in the differential diagnosis [225, 126, 247, 175]. Later studies used sustained /a/ for evaluating harsh voice. The study [225] used Goettingen Hoarseness analysis software [83] for computing jitter, shimmer, irregularity of voicing, and glottal noise. The latter study observed differential properties only in shimmer. However, getting very low jitter and shimmer for PSP compared to PD and MSA can be questionable. Other three studies used Praat software for the computation of harsh voice parameters. In the early stage of disease, PD and MSA patients did not show the difference in harsh voice speech parameters [126]. The study [247] observed that PSP and MSA patients manifest increased jitter, shimmer, and HNR compared to PD, which was attributed to predominant hypokinnesia. It is important to note that harsh voice speech dimensions were not suitable to discriminate PSP and MSA patients. Vocal folds opening and closing related features Quasi-open quotient (QOQ) and normalized amplitude quotient (NAQ) were not studied for differential diagnosis.

Excessive pitch fluctuation (stdF0) was computed using sustained vowels, which evaluate vocal folds control. PSP and MSA patients manifest increased pitch fluctuation compared to PD [247]. Another study also used stdF0, but did not analyze its differential properties [175]. In the study [255], the stdF0 measure did not exhibit group differences between PD and two phenotypes of MSA (MSA-P and MSA-C). However, MSA-P patients showed a tendency of increased pitch fluctuation. Another feature, monopitch did not show differential characteristics [126, 247, 255].

Vocal tremor (by FTRI) was predominant in MSA patients compared to PD [247]. It was computed from a sustained vowel. Study [87] showed that vocal tremor is not related to vocal folds muscle tremor; rather tremor was found in any of palate, global larynx, and arytenoids. Another study [175] also used FTRI with other acoustic features for differentiating PSP and MSA patients, but did not mention clearly which group manifest more significant vocal tremor.

The degree of voicelessness (DUV) is another manifestation of involuntary squeezing of vocal folds due to spastic dysarthria. Interestingly, MSA patients manifest predominant DUV while producing sustained phonation compared to PD. On the other hand, PSP patients also showed increased DUV compared to PD but not significant. Reliable computation of subharmonics can reveal discriminative characteristics of laryngeal muscle vibration for PD and APS [112]. The latter study showed that APS patients manifest prominent subharmonics compared to PD. In addition, increased modulation by laryngeal muscles appears to be a distinctive symptom of multiple system atrophy. Another study also reported a similar observation of subharmonics by the measure Proportion of Subharmonics Interval (PSI) [255]. The latter study showed that subharmonics are prominent for MSA-P patients compared to HC. Another vocal folds control related feature, GVI did not show differential properties among PD, PSP, and MSA [111].

Articulatory feature stdPSD was not used yet for differential diagnosis. Neverthe-

less, MSA, HD, CA patients showed increased stdPSD compared to HC [130]. Vowel space related feature, VAI was used in the study [247], which reported that PSP patients showed prominent impairment compared to PD. More precisely, male PSP patients manifest greater impairment compared to male PD [281]. Contradictory results also observed while using sustained vowels for vowel space measure [18, 250]. Hence, a detailed study of vowel space for differential diagnosis is required to understand other aspects of articulatory deficits.

Hypernasality-related features are less explored for differential diagnosis. The study [211] showed that MSA patients manifest high fluctuation in nasality (Efn_SD) compared to HC. The latter study used sustained /i/ for nasality analysis. Contradictory result is observed in Efn_M [211, 130]. It requires further investigation to analyze hypernasality in parkinsonism.

2.9.2 Impaired consonants

In previous studies, stop plosives are mostly analyzed by Voice Onset Time (VOT). The study [306, 264] observed prolonged VOT in voiceless plosives for PSP and MSA compared to PD. MSA patients manifest prominent reduction of negative VOT and VOT ratio in voiced plosives compared to PD and PSP. Duration of stop consonants (DUS) computed from monologue also provided encouraging differentiation between PD and APS groups [306]. This observation was attributed to predominant hypokinesia in PSP and MSA patients. Subjective analysis (visual method) of incomplete closure in stop plosives showed that Amyotrophic Lateral Sclerosis (ALS) group is most impaired compared to Cerebellar Ataxia (CA) [15].

Fricatives are less explored in differential diagnosis. Decay in unvoiced fricative (DUF), developed in the study [306], which did not show differential properties. Abnormal closure of articulator in subjective analysis of fricative /s/ did not exhibit its prevalence. Sharpness of friction was not analyzed for differential diagnosis.

2.9.3 Impaired diadochokinetic task

Rhythmic movements of articulators are examined by syllable repetition task. In DDK irregularity (DDKI, computed from /pa-ta-ka/), both PSP and MSA patients manifest increased irregularity in syllable duration compared to PD [247]. The average DDK rate was low (not significant) for PSP and MSA compared to PD [225, 247]. From the syllable repetition task, vowel duration was also computed. Prolonged vowel duration was observed for MSA patients compared to PD, which was attributed to ataxic dysarthria [247].

2.9.4 Impairment in spontaneous speech

Spontaneous speech (monologue and text reading) primarily used for examining a prosodic aspect of speech production. Monopitch and monoloudness are widely used acoustic measures. Later mentioned acoustic measures did not show differential characteristics [247, 255]. Inappropriate pause is another parameter that can reveal

deficits in the initiation of speech. In the number of pauses and Percent Pause Time (PPT), PSP and MSA patients exhibit more significant impairment compared to PD [247]. In total pause duration and Pause ratio within polysyllabic words (PRww) were more impaired for male MSA-P patients compared to PD, but no discrimination in female MSA-P and PD [126].

Respiration features (computed from monologue), PIR, RSR, and LRE provided encouraging discrimination between PD and APS [114]. Overall, PSP patients manifest predominant respiration problems compared to MSA and PD. Particularly in PIR, PSP patients display more impairment compared to MSA. In timing features, MSA and PSP patients showed predominant impairment in DPI and RST compared to PD. Particularly, PSP patients showed prolonged pause duration compared to MSA. In addition, increased vowel duration and voiceless stop plosives duration were observed in PSP and MSA patients compared to PD.

2.10 Classification

Several studies were devoted to identifying PD patients from Healthy Control (HC), whereas studies related to differential diagnosis (quantitative and objective measures) are more minor in numbers. The study [247] achieved 95% accuracy for PD and APS disease groups classification by speech rate, speech fluency, pauses and pitch, and amplitude fluctuations. PSP and MSA classification score was only 75% using voice quality, fluency, rate, and voice tremor. Important to note that the later study used Gaussian radial basis kernel in Support Vector Machine (SVM) classifier, which can introduce model overfitting problems, particularly in small data scenarios. The latter study objectively showed that PD patients manifest pure hypokinetic dysarthria whereas PSP patients manifest combination of predominant hypokinetic and spastic dysarthria. MSA patients manifest combination of hypokinetic and ataxic dysarthria [247].

Above existing acoustic features need to be validated with other speech databases. In addition, design of speech markers by combining homogeneous acoustic components may lead to provide encouraging differential diagnosis.

2.11 Synthetic review of existing objective measures for Differential diagnosis

This section will present available objective acoustic features for differential diagnosis. Acoustic features are organized by hypokinetic, ataxic, and spastic dysarthria types. Table 2.2 presents available objective acoustic measures of hypokinetic dysarthria for differential diagnosis. The Table 2.3 summarized available acoustic measures from ataxic and spastic dysarthria for differential diagnosis.

Deviant speech dimension	Vocal task	Acoustic measure	Description
--------------------------	------------	------------------	-------------

Hypokinetic:

1. Airflow insufficiency [247, 126]	Sustained phonation	Maximum phonation time (MPT)	Insufficient breath support for speech production
2. Harsh voice Jitter (\uparrow), Shimmer (\uparrow), HNR (\downarrow), NHR (\uparrow) [247, 126, 225, 175]	Sustained phonation	Jitter: Frequency perturbation; Shimmer: Amplitude perturbation; Harmonics-to-noise ratio (HNR): Amount of noise in voiced speech; Noise-to-Harmonics ratio (NHR)	Harsh, rough and raspy voice
3. Rapid AMR (\uparrow) [247, 225, 175]	Syllable repetition	Diadochokinetic (DDK) acceleration	Pace acceleration, rapid, blurred speech
4. Inappropriate silences			
i) Percent pause time (PPT) (\uparrow) [281, 247, 175]	Monologue	PPT is measured as the percentage of pause time relative to total speech time	Acceleration
ii) Number of pauses (\downarrow) [247, 126, 253, 175]	Monologue	No. of pauses measured as the average number of pauses per second	Inability to start speech spontaneously
iii) Pause ratio within polysyllabic words (PRww) (\downarrow) [281, 126]	Monologue	Percentual ratio of pauses within polysyllabic words;	Ratio of pause time within polysyllabic words to total pause time; Low PRww values indicate imprecise articulation
iv) Duration of pause interval(DPI) (\uparrow) [114, 253]	Monologue	DPI is defined as the median length of pause interval.	The ability to intermit and initiate speech was characterized by duration of pause intervals
v). Gaping in-between voiced intervals (GVI) (\downarrow) [114, 130]	Monologue	Clear pauses (i.e. pauses in-between voiced speech) were modelled as a bimodal distribution of formal pauses and gaps using EM-algorithm. GIV was computed as the rate of clear gaps recognized by Bayes discriminant.	Phonatory aspects provide information about disabilities to control opening and closing of vocal folds.

5. Reduced loudness (↓) [247]	Monologue	Mean speech intensity (Mean Int)	Insufficiently loud, i.e. hypophonic voice
6. Monopitch (↓) [247, 126]	Monologue	Pitch variability	Monotone voice, lacking normal pitch and inflection changes
7. Pitch range [225]	Sustained phonation	Pitch maximum and minimum	
8. Imprecise vowels (↓) [281, 250, 247]	Monologue	Vowel articulation index (VAI)	Vowel sounds are distorted throughout their total duration
9. Dysfluency (↑) [247]	Monologue	Percent dysfluent words (PDW)	Involuntary repetition of speech movements, prolongation of sounds and vocal blocks
10. Imprecise consonant articulation; Voiceless (↑), Voiced (↓) [306, 264]	Syllable repetition	Voice onset time (VOT)	Though it is kept in hypokinetic group, but it is also evident in ataxic and spastic dysarthria
11. Rate of speech timing (RST) (↓) [253, 114, 113]	Monologue	RST is defined as the rate of voiced, unvoiced and pause intervals measured as the slope of the regression line of total interval counts per time. Each interval was described as mean time between onset and offset of time.	Decreased rate of follow-up speech segment
12. Duration of stop consonants (DUS) (↑) [114, 113]	Monologue	Articulatory aspects were quantified for unvoiced fricatives and stop consonants independently	The stability of supralaryngeal movements and explosion about its preciseness
13. Relative loudness of respiration (RLR) (↑) [114, 113]	Monologue	Computed as the difference between median loudness of respiration and median loudness of speech.	Measure hypokinesia and decreased range of rib cage motion.
14. Rate of speech respiration (RSR) (↑) [114, 113]	Monologue	Estimated as the median duration between respiration	Respiratory aspects were measured on inspirations represented by respiratory intervals and expiration represented by speech intervals

15. pause intervals per respiration (PIR) (↓) [114, 113]	Monologue	Measured as mean number of pauses between respirations	Evaluate breath groups
16. Latency of respiratory exchange (LRE) (↑) [114, 113]	Monologue	Calculated as mean duration between end of speech and start of consequent respiration.	Increased latency of exchange between expiration and inspiration associated with rigidity and bradykinesia of respiratory muscles
17. Acceleration of speech timing (AST) (↑) [114, 113]	Monologue	Computed as the difference of RST between two overlapping halftimes divided by total time.	The tendency to accelerate speech rate
18. Duration of voiced intervals (DVI) (↑) [114, 113]	Monologue	Mean duration of voiced intervals determines DVI.	Prolonged phoneme is an example of ataxic dysarthria, which might be due to cerebellar damage.
19. Entropy of speech timing (EST) (↓)	Monologue	Shannon entropy computed from the occurrence of voiced, unvoiced, pause, and respiratory intervals.	The heterogeneity of speech; It is probably mixed dysarthria
20. Decay of unvoiced fricatives (DUF) (↑) [114, 113]	Monologue	Measured as difference of the second MFCC computed upon unvoiced fricatives of two overlapping halftimes.	The stability of supralaryngeal movements and explosion about its preciseness
21. Rhythm acceleration (RA) (↑) [252]	Syllable repetition	RA is defined as the gradient of the regression line obtained through regression performed on these syllable gaps time.	Measure accelerated speech.

Table 2.2: Available hypokinetic feature list for differential diagnosis; ↑: High value means high severity, ↓: low value means high severity

Deviant speech dimension	Vocal task	Acoustic measure	Description
Spastic:			
1. Strangled voice [247, 175]	Strained-voice (↓)	Sustained phonation	Degree of voicelessness (DUV)
2. Strangled voice [164]	Strained-voice (↓)	Sustained phonation	Subharmonic to harmonic ratio(S2H)
3. Slow AMR [281, 247, 126, 175]	(↓)	Syllable repetition	DDK rate
4. Slow rate [247]	(↓)	Monologue	Words count per second
Ataxic:			
1. Excess pitch fluctuations [281, 247, 175]	(↑)	Sustained phonation	Pitch variability
2. Vocal tremor [247, 175]	(↑)	Sustained phonation	Frequency tremor intensity index (FTRI)
3. Irregular AMR (↑) [225, 247, 175]	(↑)	Syllable repetition	DDK irregularity (DDKI)
4. Prolonged phonemes [247, 175]	(↑)	Syllable repetition	Vowel duration
5. Excess intensity variations (↑) [247]	(↑)	Monologue	Intensity variations (Int SD)
6. Rhythm instability (RI) (↑) [252]	(↑)	Syllable repetition	RI was calculated as the sum of absolute deviations of each observation in terms of gaps duration from the regression line, weighted to the total speech;

Table 2.3: Feature list belongs to ataxic and spastic dysarthria in differential diagnosis

Given speech features analysis mainly restricted to specific type of task and dataset. It requires further validation, particularly for different languages, and different age group, and different disease duration.

2.12 Objective

This thesis targets to design of speech markers for differential diagnosis of parkinsonian diseases. Speech markers are designed to measure the degree of deficits in a particular subsystem of speech production and/or degree of particular dysarthria. The following objectives summarize the scope of the thesis:

- Database: Develop a speech database from healthy people, PD, and MSA-P groups in the French language to analyze speech disorders. Speech samples are recorded by different modes such as microphones for speech, electroglottograph (EGG) for glottal pulses, aerodynamic data by EVA2 sensors, and video data by laryngostroboscopy.
- Imprecise vowels: Measure possible impairments in vowels for PD and MSA-P patients. To do so, speech processing toolkits like Praat, Disvoice, and additional developments by the authors are exploited to capture deficits in phonation, resonance, and articulation part of speech production. Many of these speech parameters have never been used for differential diagnosis. To the end, speech parameters are linearly combined to design an index to measure overall deficits in particular subsystems and further used for the classification of PD and MSA-P.
- Imprecise consonants: Consonants are other sound units found frequently impaired for the parkinsonian disorder. Mainly, misarticulated stop plosives and fricatives were observed in consonants. Stop plosives are mostly evaluated by perceptual analysis or durational measures. Spectral analysis of stop plosives (voiced and unvoiced) can provide additional information regarding stop plosives impairment. In addition, fricatives were mostly ignored in previous studies. Manner of friction and devoicing (resp. voicing) of voiced (resp. unvoiced) fricatives may give a better measure of imprecise fricatives. Notably, stop plosives require synchronization and coordination of laryngeal and supralaryngeal activities which introduce high possibility of misarticulation. Thus, the present thesis targets to analyze word initial obstruents by temporal and spectral methods. In addition, an automatic method is developed to segment word initial consonants, which would lead to assess the consonant automatically. To the end, PD and MSA-P patients are classified by designed speech markers.
- Imprecise syllable repetition: Syllable repetition task was frequently used to assess articulator movements. In this thesis, variability of rhythm in temporal and loudness is automatically investigated. To do so, combination of methods are implemented to segment vowel, stop plosives with high accuracy.
- Imprecise reading text: Reading text protocol is mainly used to investigate prosodic abnormality. In this thesis, reading text is manually annotated by voiced, unvoiced, pause, and respiration segments. Given the four speech segments, 13 acoustic features are computed following the previous study [130]. In addition, an open-source software, Disvoice is also used to compute other

prosodic features. Those acoustic features are analyzed towards differential diagnosis.

- Differential diagnosis of PSP and MSA: Design speech indexes to capture deficits in subsystems of speech production as well as in dysarthria to differentiate PSP and MSA patients. Only statistical analysis of individual speech parameters is not sufficient for differential diagnosis. Moreover, producing high classification accuracy using high dimensional features in linear/non-linear classifiers may raise the possibility of the curse of dimensionality. Thus, only one or 2-dimensional features would be suitable for a small amount of data.

Chapter 3

Acoustic Features

Section 2.11 provided a list of acoustic features used for differential diagnosis, particularly for PD, PSP, and MSA disease groups. This chapter elaborates on all the acoustic dimensions used in this study by subsystems of speech production. Most acoustic features are computed by either open source software such as Disvoice (extended version of Neurospeech toolkit), Praat, or developed tool in Voice4PDMSA project. In addition, a set of features also provided by the research team SAMI.

Acoustic features are computed from different speech recording task such as, sustained vowel (/a/, /i/), diadochokinetic task (/pa-ta-ka/, /ba-da-ga/, /pa-pa-pa/), single word, reading text, monologue, and sustained fricatives.

Features computed from reading text and monologue require segmentation of basic speech clusters like voiced speech, unvoiced speech, pause, and respiration. The collaborator proposed an automatic method to segment four clusters. Segmentation was carried out by systematic and sequential processes. In speech segmentation, four acoustic parameters, such as zero-crossing rate (ZCR), the variance of auto-correlation function (ACR), power (PWR), and Linear-Frequency Cepstral Coefficients (LFCC) were computed from each window frame as discussed in the study [113]. Acoustic parameters were defined as follows:

$$PWR = \frac{1}{N} \sum_{i=1}^N x^2[i].h[i] \quad (3.1)$$

$$R_x[k] = \frac{1}{N.\sigma_x^2} \sum_{i=1}^N (x[i] - \mu_x).(x[i+k] - \mu_x) \quad (3.2)$$

$$ACR = \frac{1}{N-1} \sum_{k=1}^N (R_x[k] - \bar{R}_x)^2 \quad (3.3)$$

$$ZCR = \frac{1}{N-1} \sum_{i=1}^{N-1} |sign(R_x[i+1]) - sign(R_x[i])| \quad (3.4)$$

where x is a signal in window of length N , h is hamming window, R_x represents the autocorrelation function, σ_x is the standard deviation of the signal, and μ_x is the

mean of the signal. Here, ZCR is computed on autocorrelation signal rather than time series data. The LFCC feature is computed from the spectrum of signal x . Log magnitude is computed from the spectrum. Next 24 triangular filters are applied on the log magnitude spectrum and then, Discrete Cosine Transform (DCT) provides the LFCC feature.

At first, voiced and unvoiced speech segments were detected by three acoustic parameters (ZCR, PWR, and ACR) in Gaussian Mixture Modelling (GMM). Next, unvoiced speech segments are clustered as unvoiced consonants and silence by the first five LFCC coefficients. Silence segments are clustered further by pause and respiration by the first five LFCC features. Finally, several temporal criteria were imposed for speech segmentation [113].

After getting segments of voiced, unvoiced consonant, pause, and respiration segments, speech features were computed for respiration, phonation, articulation, and timing speech subsystems.

3.1 Respiration

Respiration features were designed by the research team SAMI, Prague [114, 130]. Total four acoustic dimensions were designed to capture different respiration deficits for speech production. All four acoustic features are clustered as hypokinetic dysarthria according to their characteristics. Respiration features are computed from either monologue or text reading speech tasks. A hierarchical methodology was proposed to detect 4 sound clusters: voiced, unvoiced consonants, pause, and respiration.

Respiration speech rate (RSR)

Decreased air capacity (in lungs) can lead to an increased respiration rate. From the detected respiration events, duration in consecutive respiration was measured. The RSR was estimated as an inversion of the median respiratory period and expressed in respiration per minute [114, 130].

$$RSR = \frac{1}{median(Dur_{respiration})} * 60 \quad (3.5)$$

Increased respiration rate reflects increased severity of respiration.

Pause intervals per respiration (PIR)

Impaired control and synchronization of respiration with other subsystems can disturb normal speech production. Deficits in airflow control can lead to compromise pause production in-between respiration. Decreased number of pauses reflect increased severity in respiration control. PIR was calculated as the median number of pauses over all the in-between respiratory intervals [114, 130].

$$PIR = median(N_{pause}) \quad (3.6)$$

Latency in respiratory exchange (LRE)

Movement disorders become critical at later stages of the disease, leading to delayed initiation of inspiration followed by end of expiration. Median duration of end of speech to starting of inspiration was calculated as LRE.

$$LRE = \text{median}(Dur_{Speech\text{-end-to-starting-of-respiration})} \quad (3.7)$$

Increased LRE reflects impaired initiation of respiration [114, 130].

Relative loudness of respiration (RLR)

Obstruction in respiratory airways and oral cavity results in increased respiration noise compared to speech. Constriction in the laryngeal muscle is characterized as hyperkinetic dysarthria. RLR was computed over a power (PWR) envelope. RLR was defined as the difference of median power of respiratory ($PWR_{respiration}$) intervals and median power of voiced (PWR_{voiced}) speech [114, 130]. Thus, increased RLR reflects more significant obstruction in the airways.

$$RLR = \text{median}(PWR_{respiration}) - \text{median}(PWR_{voiced}) \quad (3.8)$$

Standard deviation of power (stdPWR)

Impairment of respiratory-phonatory coordination and control can lead to variation in loudness of voiced segments. Excessive loudness variation is eminent in ataxic and hyperkinetic dysarthria. Loudness variation is mostly attributed to hyperadduction of vocal folds vibration or dystonia in respiratory system [82, 69]. stdPWR was calculated as standard deviation of power in voiced segments.

$$stdPWR = \sigma(PWR_{voiced}^i) \quad (3.9)$$

Maximum phonation time (MPT)

Decreased air capacity in lungs can lead to shortened sustained vowel duration. MPT is calculated as the duration of sustained vowel. Decreased duration reflects more severity in the respiration system.

$$MPT = Duration_{sustained-vowel} \quad (3.10)$$

3.2 Phonation

Phonation features are related to the larynx. Excessive variation of vocal folds vibration, weakness in laryngeal muscles, tremor in the larynx are evaluated by several acoustic measures. The phonation subsystem includes several acoustic dimensions, which will be discussed in the following part.

Jitter

Jitter is the variation of duration of consecutive periods. In conventional methods, four parameters were proposed for measuring jitter perturbation. Jitter(absolute) is the average difference of consecutive periods.

$$Jitter(absolute) = \frac{1}{K-1} \sum_{i=1}^{K-1} |T_i - T_{i-1}| \quad (3.11)$$

T_i is the glottal period length and K is the number of glottal periods. Jitter(local) is the average difference of consecutive periods, divided by the average period.

$$Jitter(local) = \frac{\frac{1}{K-1} \sum_{i=1}^{K-1} |T_i - T_{i-1}|}{\frac{1}{K} \sum_{i=1}^K T_i} * 100 \quad (3.12)$$

Jitter(rap) is the relative average perturbation defined as the average absolute difference between a period and its average and its two neighbors, divided by the average period.

$$Jitter(rap) = \frac{\frac{1}{K-1} \sum_{i=1}^{K-1} |T_i - (\frac{1}{3} \sum_{n=i-1}^{i+1} T_n)|}{\frac{1}{K} \sum_{i=1}^K T_i} * 100 \quad (3.13)$$

Jitter (ppq5) is the five point period perturbation quotient which is defined as the average absolute difference between a period and the average of it and its four neighbors, divided by the average period.

$$Jitter(ppq5) = \frac{\frac{1}{K-1} \sum_{i=2}^{K-2} |T_i - (\frac{1}{5} \sum_{n=i-2}^{i+2} T_n)|}{\frac{1}{K} \sum_{i=1}^K T_i} * 100 \quad (3.14)$$

In conventional methods, glottal pulses were used for measuring jitter perturbation [35, 297]. However, in the recent study [214], fundamental frequency (F0) contour instead of glottal pulses was used for measuring jitter.

Shimmer

Shimmer is defined as the variation of amplitude in consecutive periods. Shimmer is comprised of four acoustic measures: shimmer (absolute), shimmer (local), shimmer (rap), and shimmer (ppq5) are defined by replacing T_i by amplitude A_i in the Equations 3.11, 3.12, 3.13, 3.14. In the recent study [214], frame amplitude contour instead of glottal pulses amplitude was used for measuring shimmer.

Harmonics-to-noise ratio (HNR)

HNR measures the ratio between periodic and non-periodic components of a speech sound. It has been used mostly in the vocal acoustic analysis to diagnose pathologic

voices. Praat toolkit is frequently used for HNR [36, 35]. Periodic and non-periodic part is computed by the auto-correlation method.

$$HNR(dB) = 10 * \log_{10} \frac{r'_x(\tau_{max})}{1 - r'_x(\tau_{max})} \quad (3.15)$$

where $r'_x(\tau) = \frac{R_x(\tau)}{R_x(0)}$ is the normalized autocorrelation. τ_{max} is the maximum value at lag.

Standard deviation of F0 (StdF0)

Standard deviation of fundamental frequency (F0) is generally computed over pitch contour. Most of the studies [247, 126, 281] used Praat software for computing pitch contour followed by stdF0. In the study [214], a robust algorithm for pitch tracking (RAPT) [295] method is implemented alongside Praat.

$$stdF0 = \sigma(F0) \quad (3.16)$$

Degree of unvoiced (DUV)

This is the total duration of the breaks between the voiced parts of the speech signal, divided by the whole duration of the analyzed part of the speech signal [35]. Beginning and end part of silences are deleted automatically. It is computed by Praat software. Another study mentioned the voice breaks (or subharmonics) by the proportion of subharmonic intervals (PSI) [112, 130].

Duration of voiced intervals (DVI)

Duration of voiced intervals (DVI) was computed from the monologue and text reading task. After automatic segmentation of voiced and unvoiced intervals, DVI was computed by the mean interval of the voiced part.

$$DVI = \mu(Dur_{\text{Voiced intervals}}) \quad (3.17)$$

Gaping in between voiced intervals (GVI)

Gaping in between voiced intervals (GVI) was designed to assess vocal folds' ability to adduction and abduction [114, 130]. The short pauses are attributed to adduction, hereby referred to as gaps. In contrast, vocal folds' abduction naturally produces long pauses between words or sentences, hereby referred to as formal pauses. Bi-modal distribution of pauses within voiced segments assists in detecting gaps with a shorter mean. The GVI was computed as the number of gaps per total speech time.

$$GVI = median(\text{No. of gaps}) \quad (3.18)$$

Average Quasi-open quotient (AvgQOQ)

The average Quasi-open quotient (AvgQOQ) is computed by Disvoice software. AvgQOQ is defined as the average rate of opening phase duration, divided by duration of the glottal cycle for consecutive glottal cycles [24]. To compute this phonation feature, Glottal Closure Instants (GCI) first need to be detected, where residual excitation and a mean-based signal algorithm were used [65].

Standard deviation of Quasi-open quotient (StdQOQ)

The Disvoice toolkit computes the standard deviation of Quasi-open quotient (StdQOQ). StdQOQ is defined as the standard deviation of the rate of opening phase duration, divided by duration of the glottal cycle for consecutive glottal cycles [24].

Average normalized amplitude quotient (AvgNAQ)

The average normalized amplitude quotient (AvgNAQ) is computed by the Disvoice toolkit. AvgNAQ is defined as the average ratio of the amplitude quotient and the duration of the glottal cycle for consecutive glottal cycles [24].

Standard deviation of normalized amplitude quotient (StdNAQ)

The standard deviation of normalized amplitude quotient (StdNAQ) is computed by the Disvoice toolkit. StdNAQ is defined as the average ratio of the amplitude quotient and the duration of the glottal cycle for consecutive glottal cycles [24].

Difference between the first two harmonics (H1H2)

The Disvoice toolkit computes the difference from magnitude of first harmonic (H1) and second harmonic (H2).

$$H1H2 = H1 - H2$$

. Average of H1H2 (AvgH1H2) and the standard deviation of H1H2 (StdH1H2) is computed from H1H2 over time [24].

Average of Harmonic richness factor (AvgHRF)

The average of Harmonic richness factor (HRF) is computed by the Disvoice toolkit. AvgHRF is defined as the average ratio of the sum of the amplitude of the harmonics and the amplitude of the fundamental frequency [24].

Standard deviation of Harmonic richness factor (StdHRF)

The standard deviation of the Harmonic richness factor (HRF) is computed by the Disvoice toolkit. StdHRF is defined as the standard deviation of the ratio of the sum of the amplitude of the harmonics and the amplitude of the fundamental frequency [24].

3.3 Articulation

Voice onset time (VOT)

Voice onset time (VOT) was computed automatically in the study [114, 130]. We'll use acoustic features VOT from latter study. In addition, in this thesis, VOT is computed automatically by a designed method from logatomes (pseudo word) and /pa-ta-ka/. It uses Plosion Index (PI) and Maximum Normalized Cross-Correlation (MNCC) on spectrum to robustly detect the burst and vowel onset [12].

Duration of unvoiced stops (DUS)

Duration of unvoiced stops (DUS) was computed from monologue and text reading speech tasks. This acoustic feature is designed and provided by the research team SAMI [114].

Decay of unvoiced fricatives (DUF)

Imprecise frication in unvoiced fricatives is computed by Decay of Unvoiced Fricatives (DUF). DUF is designed by the collaborator [114]. It is computed as the difference between the mean second MFCCs in the two halftimes divided by the total duration of the fricatives.

Resonant frequency attenuation (RFA)

Resonant frequency attenuation (RFA) was designed in the study [130] to assess articulatory decay. Less prominent resonances are surrounded naturally by shallow valleys. It is computed by the difference of first maxima followed by the first minima from the cepstral liftered spectrum.

Standard deviation of the power spectral density (StdPSD)

The standard deviation of the power spectral density (StdPSD) was designed to evaluate involuntary movement of articulators [130] at the time of sustained vowel. StdPSD was also developed in the Voice4PDMSA work.

3.4 Timing

Vowel duration (VD)

Vowel duration (VD) is computed from the syllable repetition task (/pa-ta-ka/) [130]. This acoustic feature is provided by the research team SAMI. It is also computed in the present study.

The rate of speech timing (RST)

Speech rate can be disturbed by reduced range of movements. On the other hand, acceleration or reduced speech rate may be observed because of imprecise speech production in different subsystems. RST computation includes computation of the total number of voiced, unvoiced, and pause intervals in a time instant followed by measuring gradient from the regression line (modeled by the total number of intervals over time) [114, 130]. This feature was computed from a monologue or text reading task. The SAMI team provided this acoustic feature.

Acceleration of speech timing (AST)

Speech acceleration was computed by acoustic feature RST. AST can provide insights into reduced range movements or increased speech rate. AST was also computed from monologue or text reading. The whole speech was split into two parts by 25% overlap. The AST was determined as the difference between the RST calculated in each halftime divided by the total duration of a speech utterance [114, 130]. The SAMI team provides this acoustic feature.

Duration of pause intervals (DPI)

Duration of pause intervals (DPI) can reflect the speech initiation deficits due to hypokinesia. DPI was computed from the monologue or text reading task. It is defined as the median length of the pause interval [114, 130].

Net speech rate (NSR)

Slow speech movement was evaluated by the conventional acoustic dimension, Net speech rate (NSR). NSR was computed by the number of syllables/words in the text reading task [130].

3.5 Prosody

Monoloudness

Insufficient airflow from the respiratory system and reduced muscle strength of the larynx resulting in reduced loudness variation in speech [69]. It is computed as the standard deviation of Power (stdPWR). This feature was computed by the SAMI team, and it can be computed also by the Disvoice toolkit.

Monopitch

Reduced melody of the voice is a common manifestation of reduced control over laryngeal muscles. Monopitch was computed as the standard deviation of fundamental frequency (stdF0). Monopitch was computed from a monologue or text reading task. Monopitch is computed by the Praat as well as the Disvoice toolkit.

3.6 Nasalic

Degree of hypernasality (Efn_M)

The degree of hypernasality (Efn_M) has been developed (by the author of the thesis) by following the study [211]. Increased energy in 1000 Hz is an important marker for hypernasality. 1/3 octave spectra analysis, based on the multirate filter bank, was used for this purpose. Total 18 frequency bands (from 75 to 4000 Hz) was computed. Average energy around 1000 Hz (890.9 Hz to 1122.5 Hz frequency band) is computed over all the time intervals.

Variability of hypernasality (Efn_SD)

Involuntary movements of the soft palate may result in variations of nasality. The standard deviation of energy around 1000 Hz (890.9 Hz to 1122.5 Hz frequency band) was computed as Efn_SD.

Chapter 4

Database

4.1 Speech databases of parkinsonian disorder

Increased interest of speech disorder analysis in Parkinsonism for building predictive tediagnosis and telemonitoring models has been a new trend. Hence, it is natural to have a speech database for analyzing speech disorders and develop a tool to serve initial diagnosis. In previous studies, several speech databases were used to evaluate speech disorder for PD [258, 270, 276, 248, 249, 91, 11] in different languages such as Czech, English, French, German, Spanish. In comparison, speech database of Atypical Parkinsonian Syndromes (APS) is less in number [281, 247, 126, 255, 164]. The most frequently used speech task is sustained vowel in previous studies, probably due to easy access and effective properties. Other speech tasks are diadochkinetic task, reading text, and monologue.

Sustained vowel was designed to evaluate particularly laryngeal functions. Additionally, it can reveal airflow insufficiency, articulator instability. Diadochokinetic syllable tasks are used to assess a client's ability to make rapidly alternating speech movements. DDK task can infer rhythm and speed of articulators; and also synchronization of laryngeal and supra-laryngeal activities. Text reading and monologue are used to assess overall speech disorder, which may be attributed to any or all of the speech subsystem deficits. Latter mentioned speech tasks can provide information mostly about prosody, timing, and respiration of speech production. Single-word (multi-syllable) task can precisely identify impairment in first phoneme and following vowel. Besides these speech protocols, other speech tasks are also warranted to study specific types of speech deficits. Important note, previous databases are primarily recorded in a very clean sound treated room. In most of the cases, a high quality condenser microphone was employed. However, in real life scenario, the availability of high quality recording room and high quality microphone are not always feasible. Consequently, although speech recording in anechoic chamber provides noise-free speech quality (preserve actual pathological speech properties), telemedicine framework has a high chance to introduce multi-dimensional noise from reverberation or other kind of environment noise.

In this thesis, the project Voice4PD-MSA aims to build a speech database to

analyze speech disorders and consequently study differential diagnosis for PD and MSA-P. At present, building Voice4PD-MSA speech corpus is in progress. Another database of parkinsonian disorders provided by Czech Republic research team is also considered in this thesis. Details of the databases are described in the following sections.

4.2 Voice4PD-MSA

Voice4PD-MSA database development involves the department of neurology and ENT departments of 2 French university hospitals (recruitment is continuing). Each recruitment was first clinically confirmed in its group (HC, PD, MSA-P). Next, the ENT specialist conducts the voice recording session by finalized speech protocol. The author of the thesis was also present at the time of recording to configure all the recording setup. Several challenges popped up at the time of recording configuration, such as hardware (microphones, sound card, peripherals) and recording room. In the following sections, the recording scenario and quality of audio will be discussed.

4.2.1 Recording setup

Recording is conducted in a small room at hospital. The recording room is not acoustically treated. Figure 4.1 displays the overall recording setup. It consists of different hardware for different mode of signal recording. A desktop computer is used to plugin recording peripheral, EVA2 system.

EVA2

EVA2 is a multipurpose workstation that can handle a variety of phonatory function tests. It is designed to measure most data currently used to evaluate speech production, including sound wave, pitch, intensity, airflow, and pressure. It has proven to be a valuable tool to assist physicians not only in the diagnosis of speech disorders but also in follow-up after surgery, medical treatment, and voice therapy. There is two input slot in EVA2. Connect a microphone to Input1-Left slot. Use the knob to select "Micro" option. Connect Electroglottograph (EGG) output (analog) to Input2-Right slot. Use the knob to select EGG. The important task before starting recording is the gain control. It can be accomplished with the gain control knob both for microphone and EGG. Check the peak indicator to adjust the gain. Gain should be set such that the indicator will not be lit up at the time of recording. ENT specialist always tried to maintain a standard level of gain. Another peripheral responsible to record aerodynamic data is also plugged-in into EVA2 system. It is designed to measure intra-oral pressure at the time of speech production [198].

H4n

H4n is usually used as a portable audio recorder. It has two in-built microphone with phase changing facility. Two other external ports can also be used for voice record-

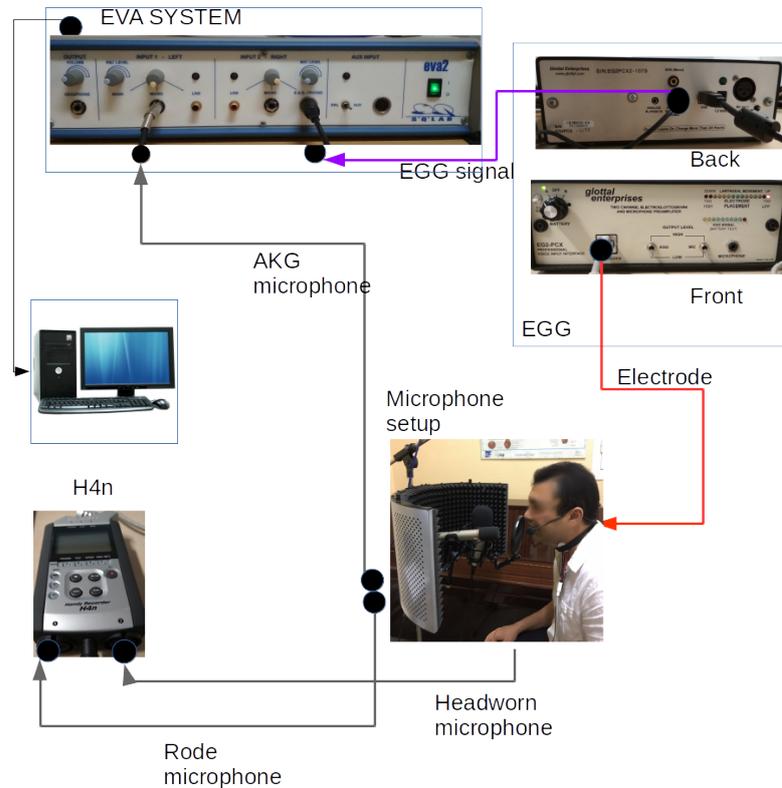


Figure 4.1: Recording scenario for database Voice4PD-MSA

ing. In the present task, we have connected a high quality condenser microphone (Rode NT1) in one of XLR port. In the other XLR port, we connected a head-worn microphone. As stated earlier, the gain should also be adjusted before recording starts to avoid clipping. There is a recording button on H4n device. A single click on this button will help to change the gain. We have to decrease the gain until the microphone indicator blink while speaking with loud volume. We selected 48000Hz sampling frequency and 16-bit quantization for the recording voice samples.

Electroglottograph

The electroglottograph is a device for measuring how much electricity flows across the larynx. The amount of electricity flow between two electrodes depends on the state of the vocal folds. Maximum electricity flows when vocal folds are closed. This instrument can capture the vocal folds vibration dynamics (only source information).

Aerodynamic tool

The aerodynamic tool is designed to capture voice sound by sensors. It can measure subglottic pressure, intra-oral pressure, and laryngeal resistance. It is connected to

EVA2 system.

Laryngostroboscopy

Laryngostroboscopy is an imaging instrument (by video) commonly used for vocal fold examination, and evaluation of patients with voice disorders [285]. It is an invasive method to capture video of vocal folds vibration.

Microphones

Total three microphones are used for capturing voice. Three microphones are NT1 from RØDE (https://cdn1.rose.com/nt_datasheet.pdf), C1000 S from AKG (https://www.ake.com/on/demandware.static/-/Sites-masterCatalog_Harman/default/dwea5c5440/pdfs/AKG_C1000S_Manual.pdf), and Head-worn microphone, hc444 from t.bone (https://www.tbone-mics.com/uploads/tx_ioproducts/Datasheet_HC-444-TWS_de_web_01.pdf). Table 4.1 presents properties of three microphones.

	C1000S AKG	NT1 RODE	Headset HC-444-TWS
Polar pattern	Cardoid, Hyper-cardoid	Cardioid	Super cardioid
Frequency range	50 to 20000 Hz	20 to 20000 Hz	20 to 18000 Hz
Sensitivity	6 mv/pa (-44 dBV)	-31.9dB re 1V/Pa @ 1kHz	-21.2 dB re 1V/Pa
Electrical impedance	200 Ohms	100 Ohms	350 Ohms

Table 4.1: Properties of three microphones

The sound absorber

The t.bone Mic Screen XL - Adjustable 5-panel absorber/diffuser (https://www.thomann.de/gb/the_t.bone_micscreen_xl.htm) which fits behind any microphone to reduce the transmission of unwanted room reflections, echoes and ambient noise. It is recommended for vocals and instruments. It is suitable for live and studio applications. It can be used on desktop or mounted on all stands, depth adjustable mic mount. Variable 5-panel screen design facilitates to widen or narrow wings to adjust acoustic behavior to fit the recording situation.

4.2.2 Recording protocol

In this project several recording protocols are considered to capture different speech production deficits. Recording protocols are listed below:

1. Sustained vowel /a/. It is used in vocal task, aerodynamic task, and laryngostroboscopy task. Participants are asked to repeat sustained /a/ twice for at

least 5 seconds for vocal task. In aerodynamic data recording, participants are asked to record /a/ as long as possible twice.

2. Sustained fricative /s/. Each participant repeats this task twice at least 5 seconds.
3. Different diadochokinetic task like /Pa-Ta-Ka/, /Ba-Da-Ga/, and /Pa-Pa-Pa/ in normal and rapid speed. DDK task also repeated twice by each participants.
4. Reading text [1]:
 ”Monsieur Seguin n’avait jamais eu de bonheur avec ses chèvres. Il les perdait toutes de la même façon. Un beau matin, elles cassaient leur corde, s’en allaient dans la montagne, et là-haut le loup les mangeait. Ni les caresses de leur maître ni la peur du loup rien ne les retenait. C’était, paraît-il, des chèvres indépendantes aimant à tout prix le grand air et la liberté.”
5. Monologue for spontaneous speaking for 90 seconds. At first, 10 different images were selected to describe the scenario in spontaneous fashion. Later on, observing difficulties by participants to describe image, participants are asked to speak daily life incidents for spontaneous speech.
6. Logatomes consist of 25 words listed in Table 4.2.

SL. No.	Words	Phoneme	SL. No.	Words	Phoneme
1)	berdo	b e R d o	14)	nouillo	n u j o
2)	broto	b R o t o	15)	perva	p e R v a
3)	chastu	S a s t y	16)	psegra	p s e g R a
4)	crancto	k R @ k t o	17)	quina	k cinq s a
5)	dirou	d i R u	18)	roursou	R u R s u
6)	feju	f e Z y	19)	sochin	s o S cinq
7)	frambi	f R a m b i	20)	spenzi	s p e g z i
8)	guizant	g I z @	21)	touca	t u k a
9)	granfa	g R @ f a	22)	tunia	t y n j a
10)	jinin	Z i n cinq	23)	vonja	v o n j a
11)	larni	l a R n i	24)	yuni	j y n i
12)	mindou	m cinq d u	25)	zacu	z a k y
13)	nianfin	n j @ f cinq			

Table 4.2: List of logatomes and it’s phonetic representation

4.2.3 Data recording

The recording room is situated in the hospital, which introduces different sources of noise. The recording room itself has different furniture and walls to reflect sound, which may introduce reverberation into signal. There is little scope to alter the room

structure to improve the acoustic quality. Therefore, the portable sound absorber is employed to mitigate the reverberation noise.

Headset microphone and Rode microphone are connected to H4n recorder. H4n recorder provides 48V phantom power to operate two microphones. The speech signals were recorded with 48kHz sampling frequency and 16-bit resolution by a headmount condenser microphone (t.bone HC 444 TWS) and Rode microphone placed at a distance of approximately 5cm and 10cm from the speaker’s mouth respectively. AKG microphone is connected to the EVA2 system. It is also placed with Rode microphone at a distance of 10cm. One microphone pop filter was placed in-front of microphone to filter out air burst. Eva2 system records audio with 25kHz sampling frequency and 16-bit resolution. EVA2 uses Sesane software to record and analyze recordings. Initially, the EGG device was connected to the EVA2 system to record vocal folds vibration in parallel. But, Sesane software faced random difficulty in capturing EGG signals. So, later on, the EGG signal was recorded by separate laptop using another audio adapter.

All the recordings were conducted after getting ethics committee approval before recruitment, and all participants gave written informed consent.

4.2.4 Recording details

As discussed, the recording room was not acoustically treated, and there was little scope to modify the room. Consequently, different kinds of noises, such as room reverberation, people talking outside, noise of vehicle can easily interrupt speech sample recording. To reduce reverberation, a sound absorber is used behind the microphone. It helps to improve the speech quality. Two impulsive sounds (click and clapping) are used to evaluate the effect of the sound absorber. Figure 4.2 showed that after applying sound absorber reverberation is reduced significantly for Rode microphone. Reverberation is measured by reverberation time (RT). Most of the time, RT is measured by the duration at which reduction of magnitude is 60dB from the onset of impulse. It is also referred as "RT60". RT60 without and with sound absorber is 0.4 seconds and 0.25 seconds respectively for click sound.

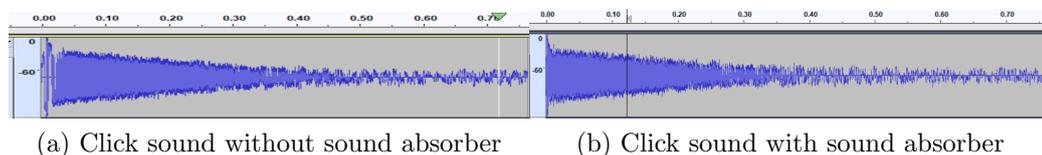


Figure 4.2: Reduced reverberation of click sound after applying sound absorber

Figure 4.3 showed that RT60 of clapping sound is also reduced after applying sound absorber from 0.6 to 0.35 seconds. Lower value of RT60 is preferred. The sound absorption material has mitigated the reverberation effect by an average of 0.2 seconds.

Now, it is time to select one of the microphone out of three microphones for speech analysis. Headset microphone has some extra advantages over the other two

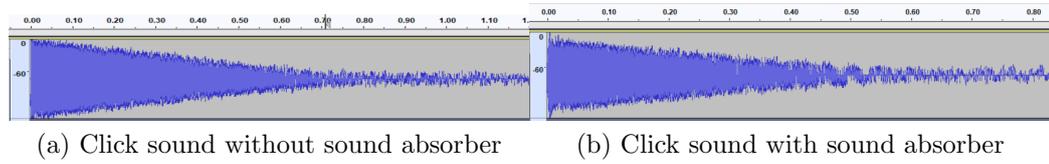


Figure 4.3: Reduced reverberation of clapping sound after applying sound absorber

microphones. The distance of microphone from mouth always remains constant. Conversely, there is a higher chance to vary the distance from mouth to microphone due to involuntary head movements by participants. In addition, the headset is less sensitive to the noise. Due to low sensitivity, headset is more capable to filter out reverberation as well as noise. To measure noise level in three microphones, the silent part of the recording sample are selected from manually labelled data. The average power spectrum over frames can display the noise profile of microphones. Figure 4.4 presents the noise properties of three microphones. It is observed that headset microphone provides lower noise levels than "Rode" and "AKG" microphones. The noise level is higher for "AKG" microphone probably due to EVA2 system.

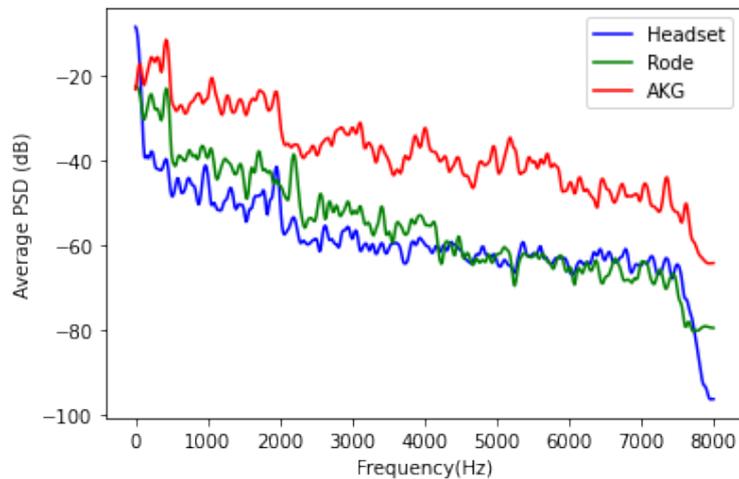


Figure 4.4: Comparison of noise properties for three microphones

Signal-to-Noise Ratio (SNR) can also be used to evaluate the recording quality of microphones. Highest SNR is found for headset microphone, 27 dB, whereas Rode and AKG microphone provides SNR value 26 dB and 23 dB respectively. Considering all above observation, headset microphone speech data will be used for further speech analysis.

4.2.5 Clinical details

From 2018 to the time of writing this thesis, a total of 60 French speakers were recruited in the framework of a research project involving the neurology and ENT departments of 2 French university hospitals (recruitment is continuing). 27 patients

(8 females and 19 males) were diagnosed with idiopathic PD, with a mean age of 60 and a mean symptom duration of 4 years. 13 subjects (8 females and 5 males) were diagnosed with MSA-P, with a mean age of 67 and a mean symptom duration of 3.5 years. 20 healthy controls (HC) with a mean age of 56 (10 female and 10 male) with no history of neurological or communication disorders were recruited.

4.3 Czech database (CzechData)

From 2011 to 2018, we enrolled a total of 65 consecutive patients for the present study, including 20 with a medical diagnosis of probable PSP (13 men and 7 women), 25 with a medical diagnosis of probable MSA (15 men and 10 women) and 20 with a medical diagnosis of idiopathic PD (13 men and 7 women). A movement disorders specialist established the clinical diagnoses of all patients according to the NINDS-PSP clinical diagnostic criteria for PSP [177], the consensus diagnostic criteria for MSA [89] and the Movement Disorder Society clinical diagnostic criteria for PD [230]. The PSP group consisted of 17 subjects diagnosed with PSP-Richardson syndrome, 2 with PSP-parkinsonism, and 1 with PSP-pure akinesia with gait freezing while the MSA group was composed of 19 subjects diagnosed with MSA-parkinsonian subtype and 6 patients with MSA-cerebellar subtype. At the time of examination, each treated PSP or MSA patient was on stable medication for at least 4 weeks consisting of various doses of levodopa alone or combined with different dopamine agonists and/or amantadine. PD patients were investigated immediately after the diagnosis was established before the initiation of dopaminergic treatment. No PD subject manifested dyskinesias at the time of the examination. Disease duration was determined based on the self-reported occurrence of the first motor symptoms. Each PSP and MSA patient underwent neurological examination including scoring according to the Neuroprotection and Natural History in Parkinson Plus Syndromes (NNIPPS) scale [224], while PD patients were rated by the Movement Disorder Society - Unified Parkinson's Disease Rating Scale (MDS-UPDRS) motor subscore. Item 3.1 MDS-UPDRS III was used for the perceptual description of speech severity. None of the patients reported a history of speech-language disorders unrelated to possible neurologic disease manifestations. No statistically significant differences were found between PSP and MSA groups for disease duration, medication doses, cognitive status, motor or speech severity (Mann-Whitney U test: $p=0.110.59$). Patient's clinical and demographic characteristics are summarised in Table 4.3. The control group consisted of 150 healthy subjects (95 men and 55 women) of comparable age (mean age 65.5, SD 7.1, range 45–83). No control subject reported a history of neurological disorders or other disorders that may affect speech, language or hearing. All subjects were Czech native speakers, and none manifested marked depressive or cognitive deficits that would interfere with the recording procedure. The study was approved by the Ethics Committee of the General University Hospital in Prague, Czech Republic and have therefore been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. All participants provided written, informed consent to the neurological examination and recording procedure.

Speech recordings were performed in a quiet room with a low ambient noise level using a head-mounted condenser microphone (Bayerdynamic Opus 55, Heilbronn, Germany) situated approximately 5 cm from each subject's mouth. Speech signals were recorded with 48 kHz sampling frequency and 16-bit resolution. Each participant was instructed to perform sustained phonation of the vowel /a/ per one breath as long and steadily as possible, fast /pa/-/ta/-/ka/ syllable repetition at least seven times per one breath, a reading passage, and a monologue on a given topic for approximately 90 seconds. All participants performed the sustained phonation and syllable repetition tasks twice.

	PD Mean/SD (range)	PSP Mean/SD (range)	MSA Mean/SD (range)
General			
Age	59.1/13.6 (37-81)	67.1/6.2 (54-84)	62.5/6.7 (45-72)
Symptom duration (years)	3/1.7 (0.3-6.7)	4/1.5 (2-7)	3.9/1.6 (2-7.5)
L-dopa equivalent (mg)	0	545/501 (0-1500)	371.4/457 (0-1500)
Amantadine (mg)	0	155/167 (0-500)	82/119.8 (0-400)
UPDRS III	30.3/11 (10-53)		
UPDRS III speech 18 item	0.6/0.5 (0-1)		
Subscores			
Tremor	8.5/4.5 (3.0-21.0) ^b	2/2 (0-6) ^a	7.3/7.5 (0-20) ^a
Rigidity	4.0/2.1 (1.0-10.0) ^b	3/3 (0-7) ^a	7.5/4.6 (0-14) ^a
Bradykinesia	18.0/6.9 (5.0-28.0) ^b	11/5 (2-20) ^a	12.8/3.7 (5-19.5) ^a
Bulbar/pseudobulbar		9/4 (3-17) ^a	7.8/2.6 (3-13) ^a
Pyramidal		1/1 (0-3) ^a	1/1.2 (0-3) ^a
Cerebellar		0.1/0.3 (0-1) ^a	3.7/6.0 (0-22) ^a

Table 4.3: Clinical characteristics of patients; NNIPPS natural history and neuroprotection on Parkinson plus syndromes-Parkinson plus scale, UPDRS unified Parkinson disease rating scale; a-NNIPPS subscore, b-UPDRS III subscore

Chapter 5

Vowel distortion in PD and MSA-P

5.1 Vowel distortion

Studies of vowel impairment are most frequent in assessing speech disorders. As an example, more than 500 studies accounted for vowel quality evaluation [45]. Vowel distortion mainly was investigated in three subsystems of speech production, like phonation (larynx), articulation, and nasal (velopharyngeal). Section 2.9.1 described vowel impairments in neurological disorder and proposed several acoustic measures for evaluation. In the phonation subsystem, the pattern of vocal folds vibration can be investigated by fundamental frequency (variability and range), perturbation measurements, and measurements of the glottal pulse shape. On the other hand, resonance properties of the vowel can reveal articulator (lips, jaw, and tongue) position and velopharyngeal functions. Hypernasality in sustained vowel is described in the Section 2.4.3. In general, vowel characteristics are represented by first (F1) and second (F2) formant frequencies. Imprecise movement (reduced range of movement) of articulators (tongue, jaw, and lips) may result in improper oral shape which in consequence change the F1 and F2 frequencies.

Hypokinetic dysarthria, a subtype of dysarthria, is a well-recognized clinical manifestation of Parkinsonian disorders. It is one of the common manifestations for Parkinson's Disease (PD) [183], Progressive Supranuclear Palsy (PSP) and Multiple System Atrophy (MSA) during the course of the disease [161, 199, 247]. It is mainly attributed to rigidity, reduced force and range of movement, and bradykinesia. It may manifest in any or all respiratory, phonatory, resonatory, and articulatory levels of speech, but its characteristics are more evident in articulation and prosody. Reduced amplitude (possible cause of articulation undershooting) and velocity of articulators (lips, tongue, jaw) in parkinsonian speakers was found in previous studies [80, 147, 69]. Basal ganglia dysfunction resulting from loss of dopaminergic input to the sensorimotor region of the striatum frequently results in movement deficits in parkinsonian disorder [240].

Variable impairment of vowel articulation was observed for PD patients. Imprecise vowel articulation was observed even in mild PD [283]. Improvement of vowel articulation was observed after dopaminergic treatment in PD patients [279, 251]. In

contrast, vowel articulation degraded even after therapeutic treatment over a time period for PD [277]. Studies related to differential diagnosis using vowel articulation measures are very rare. Impaired vowel articulation was partly investigated for PD and MSA in [247]. In another study [303], vowel impairment was studied for PD and Multiple sclerosis (MS). The latter study observed larger vowel space for clear speech and both PD and MS groups exhibited reduced vowel space area compared to HC.

Reduced range of movement may alter resonance properties of the speech signal (particularly for vowels). Each vowel is characterized by unique harmonic frequencies (called “formants”). The first and second formants are roughly related to the size and shape of the cavities created by jaw opening, lip rounding, and tongue position. Literature [301] stated that the tongue position mainly defines frequencies of F1 and F2. F1 frequency is inversely related to the height of tongue, whereas the F2 frequency is directly related to the advancement of the tongue position. For example, F1 increases while the tongue moves forward, and F2 decreases as the tongue moves backward. F1 decreases with elevation of the tongue and increases as the tongue is lowered or downward movement of the jaw. In addition, F1 and F2 decrease while lips are rounded and increase when the lips are unrounded [147].

The Vowel Space Area (VSA) is an conventional acoustic proxy for the kinematic displacements of the articulators [149, 27]. Evidence from acoustic studies also supports the conclusion that the reduced range of articulator movements in PD leads to imprecise vowel articulation caused by impaired and less distinctive “formant” generation [313]. Two different approaches were reported in literature to measure VSA e.g, triangular VSA (tVSA) [179, 277] and quadrilateral VSA (qVSA) [93, 171, 78]. For both variants, the VSA is calculated as the area formed by connecting the corner vowels (triangular vowels: /a/, /i/, /u/ and quadrilateral: /a/, /i/, /u/, /ae/) using the Euclidean distance between each coordinate in F1-F2 space.

Besides traditional VSA measures, alternative acoustic measures were also proposed to analyze imprecise vowel articulation in prior studies. For example, the Vowel Articulation Index (VAI) was proposed in the study [245, 283]. The reciprocal measure of VAI, Formant Centralization Ratio (FCR), was proposed in the study [261]. Both aforesaid acoustic metrics use formant values (F1 and F2) to examine vowel articulation. In another study, an automated VSA assessment from connected speech has been proposed to improve the accuracy of vowel space measurement. This method measures the peripheral vowel space area of formant frequency data using a convex-hull algorithm [259]. Another sensitive acoustic measure called $F2_i/F2_u$ was proposed in [262]. $F2_i/F2_u$ represents the ratio of second formant frequencies of vowel /i/ to /u/.

Effects of different speaking conditions on VSA for PD and MS were also studied in prior studies [93, 303]. The study [303] showed that clear speech provides the largest VSA. In the study [93], VSA was found to be expanded for clear speech compared to conversational speech tasks. Another study [251] showed that spontaneous and non-spontaneous speech was suitable for assessing early changes in vowel articulation associated with PD. However, the latter study considered only male speakers to assess vowel articulation. Vowel articulation impairment in laryngeal and articulatory features were analyzed in sustained vowels /a/, /e/, /i/ and /u/ for PD [18]. A

VSA measure using sustained phonation yielded a group difference between PD and HC [18]. In converse, study [251] showed that sustained phonation is not suitable to discriminate against HC and PD.

In this chapter, vowel impairment by laryngeal deficits and imprecise articulator movements are explored. Two different speech tasks are used for this evaluation. First, sustained vowel /a/ is used to evaluate laryngeal dysfunction, articulator instability, and velopharyngeal activities. Previous study [211] used sustained /i/ for measuring abnormal nasality in MSA, PD, and MS patients. The latter study did not find differential properties in hypernasal speech parameters. In this study, sustained vowel /a/ is used for capturing nasal abnormalities. Next, the reading text speech task is used for the assessment of the vowel space area. To the best of our knowledge, VSA has never been used for differential diagnosis of PD and MSA-P.

5.2 Sustained vowels

Sustained vowel production demands sufficient airflow from the lungs followed by precise vibration by vocal folds and resonance in the oral cavity and nasal cavity. Present study targets finding disorders in laryngeal, articulation, and velopharyngeal functions. In addition, differential diagnosis of PD and MSA-P is given primary importance.

5.2.1 Methodology

This section will discuss types of data and several acoustic features for measuring vowel distortion in PD and MSA-P.

Database

Sustained vowel /a/ is used in this study from Voice4PD-MSA database. Total 60 participants recorded sustained /a/ as discussed in the Section 4.2. Each speaker recorded sustained vowel /a/ twice. In detail, the database consists 20 HC (10 female and 10 male), 27 PD (8 female and 19 male) and 13 MSA-P (8 female and 5 male).

Acoustic features

Acoustic features are computed by different toolkits, such as Praat, Disvoice, and Voice4PDMSA toolkit. Table 5.1 presents a description of acoustic features, which are grouped by phonation, articulation, and nasal. Some of the acoustic features by name are the same but computed by different methods. As an illustration, jitter is calculated by glottal pulse information in Praat, whereas Disvoice uses fundamental frequency (F0) contour. It is also held for computation of shimmer. Extraction of pitch is another most studied field. Several methods are developed for robust pitch tracking. Praat software uses auto-correlation and cross-correlation methods, whereas Disvoice software uses Robust Algorithm for Pitch Tracking (RAPT). Comparison of different pitch extraction methods showed that auto-correlation (Praat) and RAPT

perform better in a clean environment, but RAPT method provided better accuracy in noisy conditions. However, the pitch tracking method specifically for the pathological voice was not found in the literature; rather, existing methods are used for pathological speech.

Disvoice toolkit is used to measure vocal folds opening and closing pattern [101, 217]. Selected features are described in Table 5.1.

Among different vocal tremor indexes, the frequency tremor intensity index (FTrI) is considered in this study which is defined as the magnitude of the strongest low-frequency modulation of F0 [42].

For univariate analysis, each feature is first used for normality tests by the one-sample Kolmogorov-Smirnov test. If data is normally distributed, the student t-test is used for group differences, otherwise Kruskal-Wallis test (suitable while pairwise group difference is computed).

5.2.2 Results

Phonation

Phonation features are designed such that they can investigate laryngeal functions. Table 5.2 presents the group difference of individual phonation features. Among the phonation features, encouraging group difference between PD and MSA-P is found in QOQ, jitter, PPQ, and DF0 by initial investigation. Important to note, irregular rhythm (in duration) in vocal folds vibration is predominant for MSA-P patients in jitter, QOQ, PPQ, DF0, and DDF0 acoustic features.

1-dimensional phonation feature ($X1_a$) is designed by combining 9 acoustic features from phonation subsystem acoustic features. All selected features yield group differences between PD and MSA-P. Other features are ignored for further analysis. $X1_a$ is defined as follows:

$$X1_a = \frac{1}{2} * (std_avg_QOQ + avg_std_QOQ + avg_jitter + std_jitter + std_DDF0) + \frac{1}{4} * (avg_ppq + std_ppq + std_DF0 + FTRI)$$

Dimension $X1_a$ provides encouraging group differences between PD and MSA-P as shown in Figure 5.1. Important note, selected acoustic features primarily represent timing variation of vocal fold vibration. Total 7 MSA-P patients showed distinctly higher severity than PD patients. Furthermore, PD group also yield group difference w.r.t. HC. Thus it would also serve as hypokinetic index.

Articulation

Articulation features mainly measure the involuntary movements of articulators. Table 5.3 presents a group difference of articulation features from sustained vowel. In both acoustic features, MSA-P patients manifest predominant impairment in articulator stability compared to PD and HC.

Subsystem	Acoustic features	Definition	Description
Phonation	Jitter	Variation of consecutive glottal pulse duration or consecutive F0; jitter is represented by local and pitch perturbation quotient (PPQ)	Reduced control of vocal folds vibration
	Shimmer	Variation of consecutive glottal pulse amplitude or consecutive F0 amplitude; shimmer is represented by local and amplitude perturbation quotient (APQ)	Disrupted airflow and reduced control of vocal folds vibration
	HNR	Harmonic to Noise Ratio which measures amount of noise in vowel	Reduced airflow and reduced control of vocal folds cause improper closure
	QOQ	Average and standard deviation of Quasi-open Quotient (QOQ) is measured by the rate of opening phase duration /duration of glottal cycle	Reduced control of vocal folds vibration
	NAQ	Average and standard deviation of Normalized Amplitude Quotient (NAQ) is measured by the ratio of the amplitude quotient and the duration of the glottal cycle	Reduced control of vocal folds vibration
	Derivative of F0	First and second derivative (DF0 and DDF0) of F0 contour	Variable airflow and variation in laryngeal vibration
	DUV	Degree of voiceless; Fraction of locally unvoiced frames	Voice sounds strained or strangled (effortful squeezing of voice through glottis)
	FTRI	Frequency tremor intensity index (FTRI)	Vocal tremor
Articulation	stdPSD	Variation of power spectral density in frequency band	Involuntary movements of articulators
	stdlogE	Variation of energy in frames	Involuntary movements of articulators and airflow variation.
Nasalic	Efn_M	Average energy in 1000Hz	Improper velopharyngeal function; Imprecise closure by soft palate.
	Efn_SD	Variation of energy at 1000Hz in frames	Involuntary movements of soft palate

Table 5.1: Acoustic features related to vowel distortion

Features	HC	PD	MSA	P-Value		
	Mean/SD	Mean/SD	Mean/SD	HC_PD	HC_MSA	PD_MSA
avg_DF0	-0.02 / 0.06	-0.05 / 0.09	-0.08 / 0.15	0.36	0.09	0.34
avg_DDF0	0.024 / 0.04	0.008 / 0.11	-0.0097 / 0.0831	0.69	0.28	0.42
avg_std_QOQ	0.04 / 0.01	0.04 / 0.012	0.06 / 0.015	1	0.03	0.05
std_var_GCI	0.0003 / 0.0003	0.0004 / 0.0004	0.0008 / 0.0009	0.15	0.03	0.23
std_avg_QOQ	0.0284 / 0.0141	0.0309 / 0.0172	0.0611 / 0.0321	0.69	0.0004	0.0002
avg_Jitter	0.53 / 0.20	0.76 / 0.32	1.21 / 0.66	0.003	0.0001	0.006
std_Jitter	0.91 / 0.72	1.27 / 0.65	2.27 / 1.98	0.016	0.0012	0.027
avg_Shimmer	1.59 / 0.54	1.92 / 0.7	2.26 / 0.74	0.081	0.0041	0.16
std_Shimmer	2.11 / 0.62	2.62 / 0.85	2.88 / 0.76	0.008	0.002	0.26
avg_apq	3.31 / 1.45	3.89 / 1.58	4.67 / 1.58	0.17	0.002	0.11
avg_ppq	0.40 / 0.20	0.59 / 0.34	1.10 / 0.85	0.016	0.0002	0.008
std_ppq	0.53 / 0.57	0.74 / 0.54	1.66 / 1.86	0.015	0.0009	0.012
std_DF0	2.0 / 1.86	2.79 / 2.02	5.27 / 3.71	0.05	0.0001	0.006
std_DDF0	2.49 / 2.69	3.47 / 2.59	7.26 / 5.74	0.036	0.0001	0.003
FTRI	1.70 / 1.78	2.37 / 1.78	3.97 / 2.69	0.11	0.005	0.04

Table 5.2: Group difference of phonation features from sustained vowel /a/; blue and red colour represents predominant severity by PD and MSA-P respectively

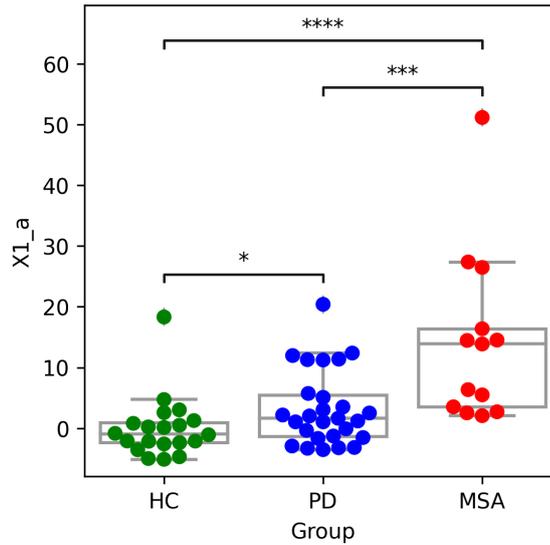


Figure 5.1: Designed phonation feature ($X1_a$); * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$

Features	HC	PD	MSA	P-Value		
	Mean/SD	Mean/SD	Mean/SD	HC_PD	HC_MSA	PD_MSA
stdlogE	2.01 / 0.72	2.01 / 0.68	2.73 / 0.82	0.73	0.025	0.006
stdPSD	0.87 / 0.22	0.91 / 0.19	1.12 / 0.17	0.704	0.004	0.003

Table 5.3: Group difference of articulation features from sustained vowel /a/; blue and red colour represents predominant severity by PD and MSA-P respectively

Combination of stdlogE and stdPSD features also further provide improved group

difference as presented in the Figure 5.2. It can also serve as an index of involuntary articulator movements and defined as:

$$F_art = \frac{stdlogE + stdPSD}{2}$$

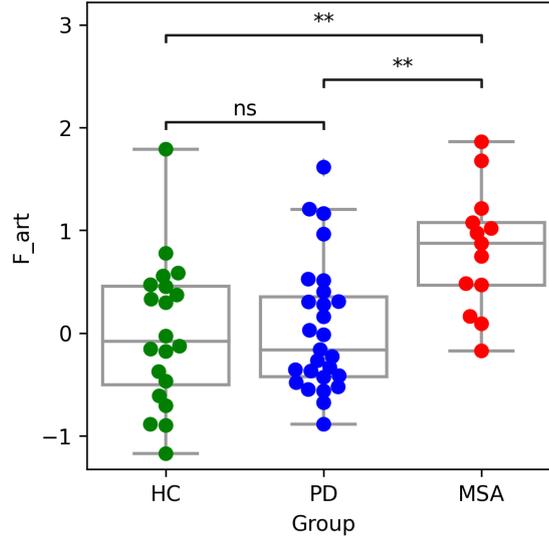


Figure 5.2: Combination of articulation features from sustained vowel; ** $p < 0.001$, ns: not significant

Nasal

Two nasal features (Efn_M and Efn_SD) provided encouraging distinctive characteristics. Table 5.4 presents group differences by nasalic features. MSA-P patients manifest predominant hypernasality and variability of hypernasality in sustained vowel /a/. In accordance with previous study [211], MSA patients manifest greater variability in nasality compared to HC. Notably, the latter study did not find differential properties among MSA, MS, and CA groups. Important to note, previous study computed nasalic features from sustained vowel /i/.

Features	HC	PD	MSA	P-Value		
	Mean/SD	Mean/SD	Mean/SD	HC_PD	HC_MSA	PD_MSA
Efn_M	-28.42 / 1.66	-27.72 / 1.62	-26.24 / 2.55	0.07	0.004	0.04
Efn_SD	0.19 / 0.05	0.22 / 0.07	0.37 / 0.15	0.28	0.0001	0.0003

Table 5.4: Group difference of nasalic features from sustained vowel /a/; blue and red colour represents predominant severity by PD and MSA-P respectively

The 1-dimensional nasality feature is designed by combining two nasalic features. Before feature combination, individual features are transformed by Z-normalization

(0 mean and unit standard deviation). Overall nasality is defined as:

$$F_{nasal} = \frac{Efn_M + Efn_SD}{2}$$

. Designed nasalic feature yield better discrimination compared to individual features. Figure 5.3 displays differentiation within HC, PD, and MSA by individual nasalic feature and combined feature. MSA-P patients showed higher abnormalities compared to PD and HC groups.

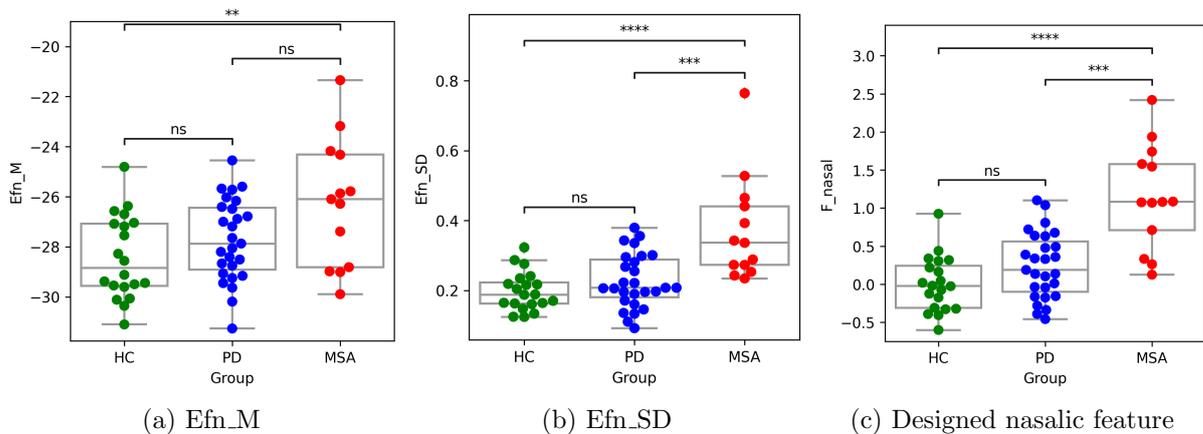


Figure 5.3: Designed nasalic feature; "ns" stands for not significant

Classification

Considering the small size of the database, 1-dimensional or 2-dimensional features would be suitable to avoid overfitting problems. Hence, 2-dimensional features are adopted in this study. Initial analysis of individual acoustic features provided encouraging differentiation between PD and MSA-P, but not enough to yield good accuracy by logistic regression and LOSO cross-validation. Table 5.5 summarized classification scores. Hence, feature combinations may aid in getting high accuracy.

Features	Threshold	Accuracy (%)	Specificity (%)	Sensitivity (%)
X1_a	0.5	75	85.18	53.84
F_art	0.35	75	74	76.92
F_nasal	0.5	85	92.59	69.23

Table 5.5: Classification score by logistic regression using X1_a, F_art, and F_nasal for PD and MSA-P

Articulation and nasalic features are combined to measure an overall articulation (jaw, lips, tongue, and soft palate) impairment. Overall articulation feature is defined as follows:

$$X2_a = stdPSD + 2 * (F_{nasal} + std.logE)$$

Feature $X2_a$ yields encouraging discrimination between PD and MSA-P. Figure 5.4 presents comparably greater impairment for MSA-P patients in $X2_a$ dimension. Total 8 (out of 13) MSA-P patients manifest predominant disorder than PD. Thus $X2_a$ index could be used as an important acoustic marker for differentiating PD and MSA-P.

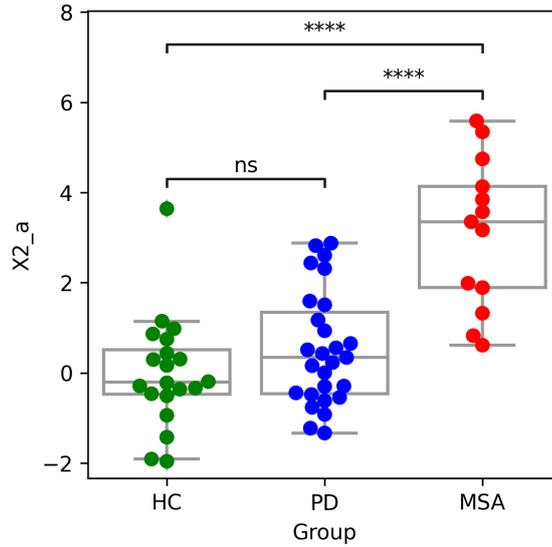


Figure 5.4: Combination of articulation and nasalic features, $X2_a$ from sustained vowel; **** $p < 0.0001$, ns: not significant

Individual designed feature did not yield good accuracy by logistic regression. Hence, we adopted decision tree to classify PD and MSA-P patients. Figure 5.5 showed that 1-dimensional phonation feature ($X1_a$) and articulation feature ($X2_a$) provide high discrimination between PD and MSA-P. In both dimensions, MSA-P patients manifest higher deficits compared to PD. Soft threshold of "7" in $X1_a$ and "3" in $X2_a$ can serve good discrimination between PD and MSA-P. Total 3 MSA-P patients exhibit impairment in both $X1_a$ and $X2_a$ which can be referred as "probable" category. On the other hand, MSA-P patients (total 9) who manifest impairment either in $X1_a$ or $X2_a$ can be regarded as "possible" category.

For the classification of PD and MSA-P, the decision tree is exploited using $X1_a$ and $X2_a$. Considering the small amount of data, Leave-one-subject-out (LOSO) cross-validation method is adopted in this experiment. LOSO classification method yields good accuracy, which is presented in the Table 5.6.

Features	Accuracy (%)	Specificity (%)	Sensitivity (%)
[$X1_a, X2_a$]	95	96.29	92.30

Table 5.6: Classification score by decision tree and LOSO cross-validation using $X1_a$ and $X2_a$ feature dimensions for PD and MSA-P

Thus, two acoustic dimensions can be used as acoustic markers for differential

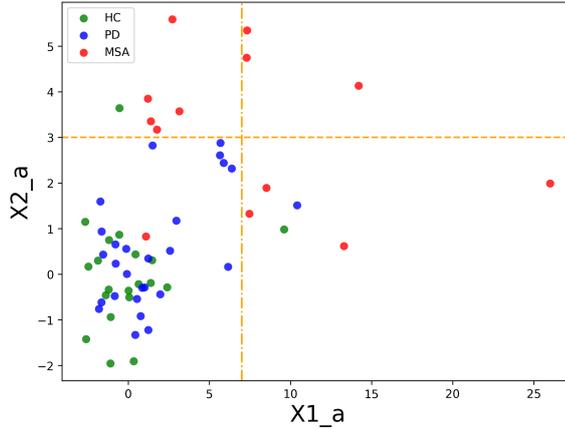


Figure 5.5: Biplot of phonation feature ($X1_a$) w.r.t. articulation feature ($X2_a$) (dotted line represent decision thresholds)

diagnosis of PD and MSA-P patients.

Discussion

Current study presents quantitative analysis of overall phonation and articulation disorder in sustained vowel /a/ for PD and MSA-P. As both disease group manifest parkinsonism, differential diagnosis is an challenging task at early stage. Acoustic features are automatically computed by existing speech processing tool and developed tool by author (inspired by literature). The notable contribution is the designing speech markers to assess vocal folds vibration variability (temporal) and movement deficits of articulators. Indeed, two orthogonal features lead us to discriminate two parkinsonism type disease groups (PD and MSA-P) by good margin. Thus, analysis can be used as a potential diagnostic screening tool. Some of the individual speech components yield significant differentiation between PD and MSA-P patients, which were not analyzed particularly for MSA-P patients. In agreement with previous studies [247], the present study also observed variability in vocal folds vibration ($stdF0$) for MSA patients. It is most probably due to combined deficits in basal ganglia and cerebellar circuits [293, 40, 94, 265, 288]. In agreement with previous study [247], a group of MSA-P patients exhibited high vocal tremor compared to PD. Additionally, in agreement with the study [130], the variability in spectral band power was predominated in MSA-P patient compared to PD. Important to note, previous study include phenotype of MSA (MSA-C and MSA-P), whereas present study only considered MSA-P patients. Both disorders are attributed to ataxic dysarthria due to tremulous arytenoid movements [219] and involuntary movements of articulators primarily the tongue. Notably, predominant involuntary movements of articulators was also observed in cerebral ataxia and huntington disease [130], but not observed for PSP.

Quantitative analysis observed high hypernasality and intermittent nasality for MSA-P patients compared to PD. Previous studies reported controversial results re-

garding nasality in neurological disease groups. The perceptual investigation found rare nasality in PD patients [183], whereas other studies [49, 210] found significant nasality in PD. Previous studies observed nasality in ataxic dysarthria [228] and one study attributed hypernasality to basal ganglia dysfunction [210]. In addition, according to [69] intermittent nasality is attributed to hyperkinetic dysarthria. In agreement with the studies [183, 211], present study also did not observe significant nasality in PD group. According to the study [130], we also observed hypernasality in MSA patients. Additionally, MSA-P patients also manifest significant irregular nasality compared to PD and HC, which was not observed in previous studies.

Designed acoustic indexes was a novel contribution of the present study. Two acoustic dimensions were designed to represent phonation and articulation deficits. In the first dimension, $X1_a$ represents the disorder in timing of vocal folds vibration. MSA patients predominate in dimension $X1_a$. Notably, MSA patients are diagnosed as parkinsonian type. It can be hypothesized that patients may have deficits in basal ganglia control circuit. On the other hand, $X2_a$ measures primarily involuntary movements of articulator and soft palate. It can be hypothesized as the manifestation of deficits in cerebellar circuit. It thus confirms previous statement that MSA patients manifest wide spread clinical symptoms.

Above analysis and findings need to be validated by larger data samples. Moreover, other disease groups are also need to be included in the similar analysis.

5.3 Vowel space area: Reading text

In the present study, French vowel articulation is analyzed for parkinsonian subtypes (PD, MSA-P) using three corner vowels (/a/, /i/, and /u/). To the best of our knowledge, no previous study analyzed vowel articulation in detail for MSA-P. Three corner vowels are extracted from the reading passage speech task, considering spontaneous speech task is more suitable for capturing early changes in parkinsonism [250]. To measure F1 and F2 formant frequencies, two methods were adopted. First, F1 and F2 frequencies were manually measured by analyzing a wide-band spectrogram for each vowel. Second, the Praat speech analysis tool was used to measure F1 and F2. Standard acoustic measures such as VSA, VAI, and FR were used to assess French vowel articulation. Duration of vowels are also assessed to find particular changes in PD and MSA-P.

5.3.1 Speech database: Voice4PD-MSA

In the Voice4PD-MSA speech database, a total of 60 French speakers are recruited for the experimental study. 27 subjects (8 females and 19 males) are diagnosed with idiopathic PD. 13 subjects (8 females and 5 males) are diagnosed with MSA-P (parkinsonian type). Age matched 20 Healthy Control (HC) subjects are recruited (10 female and 10 male) for the purpose of comparison. Description of reading text used in this study is presented in Section 4.2.2.

The paragraph of *La chèvre de Monsieur Seguin* [1] is considered for reading task. Second sentence of the reading task is considered to evaluate the impreciseness of three corner vowels /a/, /i/, and /u/. The second sentence consists of the text, “Il les perdait toutes de la même façon”. Three corner vowels /i/, /u/, and /a/ are extracted from words “Il”, “toutes”, and “façon” respectively. Figure 5.6 shows the time-frequency representation of a sentence.

5.3.2 Methodology

Three corner vowels /a/, /u/, and /i/ were labeled using a wide-band spectrogram in Praat [35]. Next, First (F1) and second (F2) formant frequencies are measured separately from each vowel. Speaker’s gender is a sensitive parameter while computing formant frequencies. Formant frequencies have a direct relation with vocal tract length. Female speakers have shorter vocal tract (average 14.5 cm) compared to male speakers (17 to 18 cm). Hence, gender specific formants estimation is mostly considered in previous studies for evaluation of imprecise vowel articulation. Default setting in PRAAT was considered for formant measurement. The highest formant frequency parameter is a sensitive parameter that needs to be chosen precisely according to gender. In the default setting, the highest formant frequency is fixed to 5500 Hz (for female) and 5000 (for male). The maximum number of formants is set to 5. The input signal is pre-emphasized (from 50 Hz) to enhance high frequency amplitude. The pre-emphasis is that vowel spectra tend to fall by 6 dB per octave; the pre-emphasis

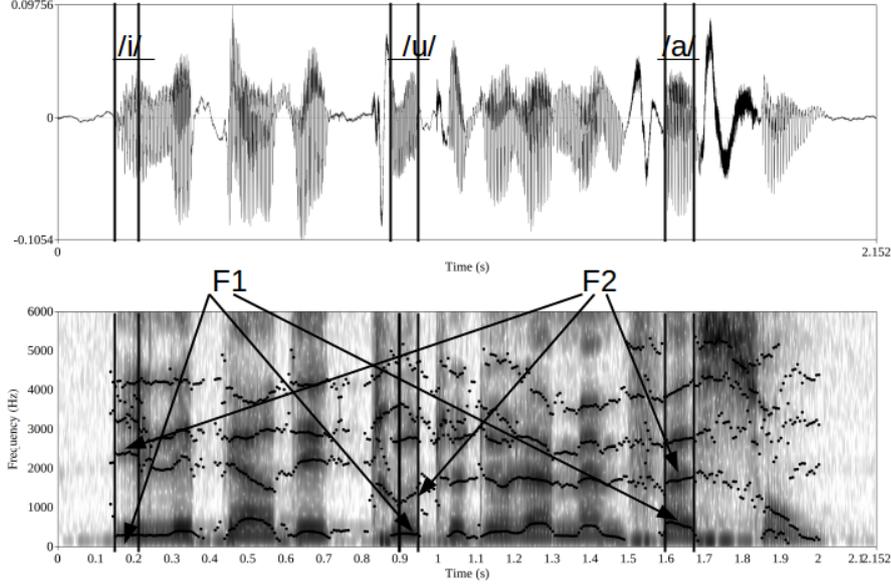


Figure 5.6: Time-frequency representation of the sentence; F1 and F2 represent first and second formants

creates a flatter spectrum, which is better for formant analysis because it matches the local peaks the global spectral slope.

Additionally, F1 and F2 frequencies are manually measured using the steady-state portion of wide-band spectrograms and Linear Predictive Coding (LPC) spectrum as per guidelines [63]. As a result, F1 and F2 are marked at the middle of the concentrated energy over the frequency band.

In Praat, formant values are averaged over a 50% time interval around the temporal midpoint of each vowel. Averaged F1 and F2 are used to calculate VSA, VAI, and $FR = F2_i/F2_u$. VSA is computed as follows [179]:

$$VSA = 0.5 \times |F1_i \times (F2_a - F2_u) + F1_a \times (F2_u - F2_i) + F1_u \times (F2_i - F2_a)|, \quad (5.1)$$

where $F1_i$ and $F2_i$ are first and second formant frequencies of vowel /i/ respectively. $F1_a$ and $F2_a$ are first and second formant frequencies of vowel /a/ respectively. First and second formant frequencies of vowel /u/ are represented by $F1_u$ and $F2_u$. The other measure, VAI is formulated as [245]:

$$VAI = \frac{F2_i + F1_a}{F1_i + F1_u + F2_u + F2_a}. \quad (5.2)$$

Another two measures related to formant spacing are also adopted to evaluate tongue elevation and tongue advancement. Compact-diffuse (C-D) and grave-acute (G-A) are the index of tongue elevation and tongue advancement, respectively [31]. C-D is defined as “ $F2 - F1$ ” of each vowel. G-A is defined as “ $(F2 + F1)/2$ ”.

Above stated acoustic measures are computed in two different units (Hz and Semi-tone). Next, gender independence properties of acoustic measurements are evaluated

for two units. In semitone units, it may be hypothesized that gender difference will be significantly reduced in acoustic measures compared to Hz units. To assess the group difference, kruskal-wallis hypothesis test is used while data is not normally distributed, and Student t-test is used while data is normally distributed.

Weak energy in F2 of vowels /i/, /u/, and /a/ is also computed by average band energy around F2 frequency (100 Hz bandwidth). F2 computed by visual approach of vowels is used to compute band energy.

In addition to conventional acoustic measures, prosodic aspects of vowels (duration) are also investigated as secondary investigations. Study [302] showed that speaking rate (by vowel duration) may affect the vowel space area. Duration of /a/, /u/ and /i/ are analyzed to find it's impact on vowel impairment. Duration of vowels is measured from manually labeled data.

5.4 Result

Manually measured formant frequencies of three corners vowels are first analyzed. Detailed numerical data of participants are listed in Table 5.7 as average and Coefficient of Variation (CV). Numerical values exhibit clear differences between male and female speakers within three groups (HC, PD, MSA-P). Notably, male MSA patients show increased F2 compared to PD and HC in vowel /u/. In subjective analysis (while estimating formants), weak energy was observed in F2 of /i/ and /u/ for 50% of PD and MSA patients. Weak energy may be an indicator of impaired vowel articulation in parkinsonism. To the best of our knowledge, weakness in second formant of vowels is first time observed in the present study. Additionally, an objective measure is proposed to capture weak energy in F2.

	Female						Male					
	HC		PD		MSA		HC		PD		MSA	
	Mean	CV	Mean	CV	Mean	CV	Mean	CV	Mean	CV	Mean	CV
F1_i	327.9	8.65	308.875	7.92	304	14.66	322.3	11.98	284.36	13.08	249	23.86
F2_i	2451.5	4.25	2434.25	5.91	2333.125	11.38	2044.6	6.66	2098.68	7.12	1982.4	11.69
F1_u	342.9	7.31	323.87	8.24	302	7.73	323.6	10.84	287.68	10.69	299.4	16.62
F2_u	1192.2	17.32	1137.75	12.12	1337.62	15.94	1033.1	8.08	1128.05	12.67	1304	14.72
F1_a	622.9	8.24	655.12	9.64	591.12	9.33	615.9	7.63	552.63	10	483.4	25.38
F2_a	1692.7	5.69	1696.87	4.54	1656.5	8.82	1367.1	4.83	1392.78	7.64	1462	6.22

Table 5.7: Manually computed average/CV of F1 and F2 of three groups (HC, PD, MSA-P) by gender

Formants are also computed by separate configuration (as discussed in section 5.3.2) for male and female by Praat. Table 5.8 provides Mean Absolute Difference (MAD) of formants between subjective method and Praat. Second formant frequency of /u/ show maximum deviation. Overall, absolute MAD is more significant for male speakers compared to female subjects. It indicates that formant estimation becomes challenging for male speakers.

Now, gender differences in acoustic measures are assessed for two formant frequency measuring methods. Male and female speakers of only HC groups are consid-

	F1_i (Hz)	F2_i (Hz)	F1_u (Hz)	F2_u (Hz)	F1_a (Hz)	F2_a (Hz)
Female	47.26	56.11	34.42	88.67	54.89	54.34
Male	71.70	72.41	79.38	158.11	111.43	63.78

Table 5.8: MAD of formants measured by manual method and Praat

ered for group difference (using student t-test hypothesis test). Gender discrimination can be ignored if statistical significance more than 0.05 ($p > 0.05$). Individual formant frequency shows clear gender difference ($p < 0.05$) for three vowels in both units (Hz, Semitone). Thus, it is confirmed that a single formant value is not suitable for gender independent scenarios. Table 5.9 presents the group difference (p-value) of male and female groups for all the acoustic measures of vowel articulation. It is evident that VSA measure in Hz is gender sensitive. Conversely, VSA measure in semitone shows gender independence. It confirms that inter-speaker difference (mostly related to gender) significantly reduced while frequency in Hz is converted to the logarithmic domain (semitone). In addition, ratio based acoustic measures (VAI, FR) exhibit gender independence in both units (Hz, Semitone) as mentioned in the study [260]. In the formant spacing feature, G-D feature (a measure of tongue advancement) did not show gender independence in both units (Hz, Semitone). In contrast, C-D features (measure of tongue elevation) show gender independence in vowels /i/ and /u/.

Acoustic Measure	Manual	Praat method
	Male vs Female	Male vs Female
VSA (Hz^2)	0.070	0.021
VSA ($Semitone^2$)	0.641	0.737
VAI (Hz)	0.795	0.285
VAI (Semitone)	0.399	0.446
FR (Hz)	0.305	0.862
FR (Semitone)	0.493	0.870
C-D_/i/ (Hz)	0.000	0.000
C-D_/i/ (Semitone)	0.0010	0.754
C-D_/u/ (Hz)	0.048	0.292
C-D_/u/ (Semitone)	0.157	0.997
C-D_/a/ (Hz)	0.000	0.0005
C-D_/a/ (Semitone)	0.000	0.188
G-A_/i/ (Hz)	0.000	0.000
G-A_/i/ (Semitone)	0.0007	0.000
G-A_/u/ (Hz)	0.037	0.069
G-A_/u/ (Semitone)	0.047	0.0058
G-A_/a/ (Hz)	0.000	0.0005
G-A_/a/ (Semitone)	0.000	0.18

Table 5.9: Gender difference (p-value) in acoustic measures related to formant frequencies; gender independent features are marked as bold ($p > 0.05$)

The above analysis suggests that acoustic measures in Semitone are suitable to examine vowel articulation in gender independent scenarios. Henceforward, we decide

to use acoustic measures with semitone in the following comparative experiment. For individual formant measurement, F2_{/u/} provide group differences between PD and MSA-P (p=0.004) in manually computed formants. MSA-P patients show increased formant in F2_{/u/}. However, this F2_{/u/} formant measure did not show gender independence. Hence, this parameter can not be considered for the present scenario. Conversely, individual formant frequencies computed by Praat did not show group differences between PD and MSA-P patients.

The F1-F2 plot of three corner vowels of three groups may reflect the overall scenario of vowel space. In general hypothesis, F2_{/u/} is increased, and F2_{/i/} is decreased for PD patients. Figure 5.7 shows vowel space for HC, PD and MSA-P by two different formant estimation methods. The vowel space of the three groups show encouraging visual discrimination. In average formant value, MSA patients show dispersion and reduction of vowel space compared to HC.

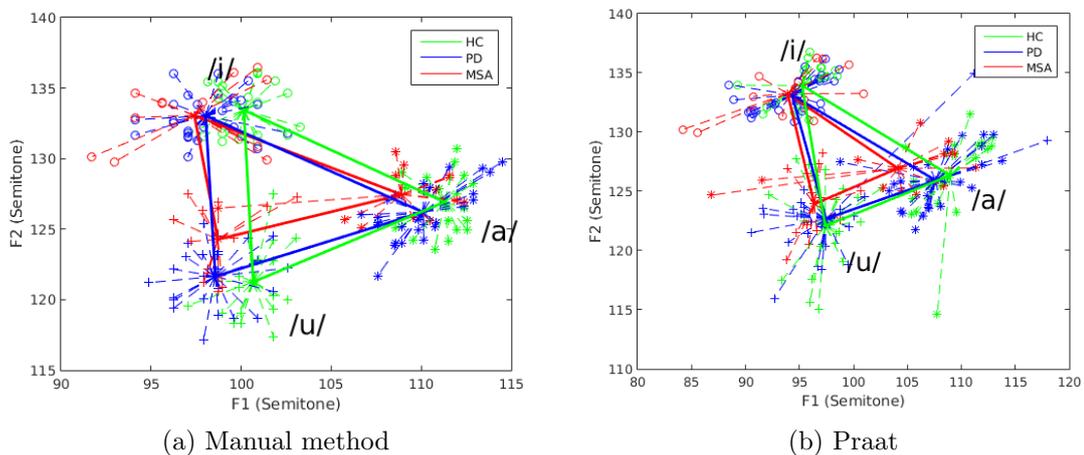


Figure 5.7: Vowel space using three corner vowels ('o' for /i/, '+' for /u/, '*' for /a/) for HC, PD and MSA-P

Now, we proceed to examine other acoustic measures to find specific differences among groups. Table 5.10 provides pairwise group difference (p-value) among HC, PD, and MSA-P. In both formant measuring methods (manual and Praat), MSA-P patients manifest greater impairment than PD. Reduced VSA is observed for MSA-P patients compared to PD (in manual and Praat method) and HC (in manual method). In addition, ratio based features, VAI and FR, also show similar trends as VSA. In FR, Praat method did not provide group difference between PD and MSA-P due to F2 computation error in /u/. On the other hand, tongue elevation deficits is observed in acoustic measure C-D for PD and MSA-P. MSA-P patients predominates in C-D of /u/ and /a/ than PD.

Additionally, we provided analysis of vowel space related features in Figure 5.8. Though we got very good group differences, however visual analysis showed that margin of separation is not very much prominent.

Energy in F2 was found weak in subjective analysis for vowel /i/ and /u/. In

Measure	Manual method			PRAAT method		
	HC vs PD	HC vs MSA	PD vs MSA	HC vs PD	HC vs MSA	PD vs MSA
VSA (<i>Semitone</i> ²)	0.71	0.0072	0.0041	0.59	0.016	0.002
VAI (<i>Semitone</i>)	0.6	0.02	0.0037	0.98	0.016	0.006
FR (<i>Semitone</i>)	0.56	0.0006	0.0019	0.37	0.032	0.06
C-D_/i/ (<i>Semitone</i>)	0.035	0.047	0.49	0.75	0.5	0.63
C-D_/u/ (<i>Semitone</i>)	0.005	1.96e-05	0.0092	0.85	0.047	0.068
C-D_/a/ (<i>Semitone</i>)	0.315	0.007	0.007	0.42	0.0048	0.004

Table 5.10: Group difference (p-value) of acoustic measures for HC, PD, MSA-P; Red color represents higher impairment for MSA-P

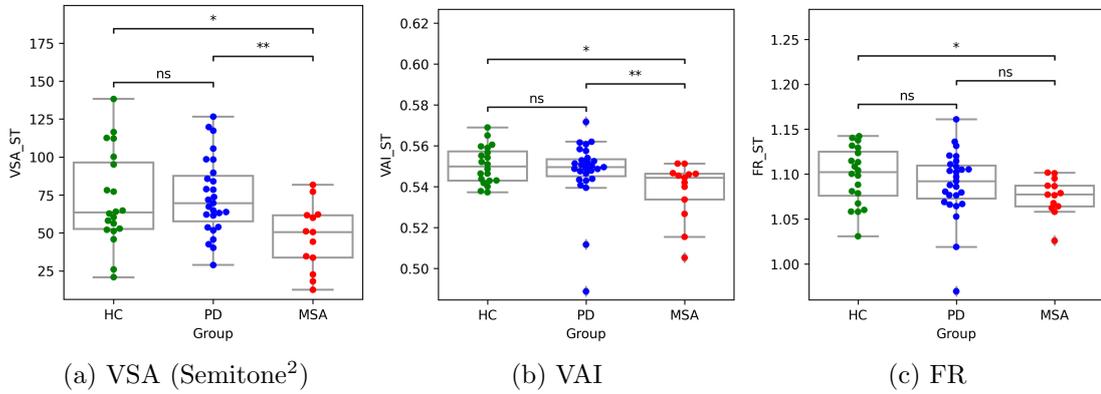


Figure 5.8: VSA, VAI, and FR features computed by Praat method; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, ns: not significant

objective measures, weak energy is observed in all three vowels. Figure 5.9 displays distinctive weak power in /u/ and /a/ for MSA-P patients. Notably, a group of MSA-P patients showed weak energy in around F2 of vowels compared to PD and HC.

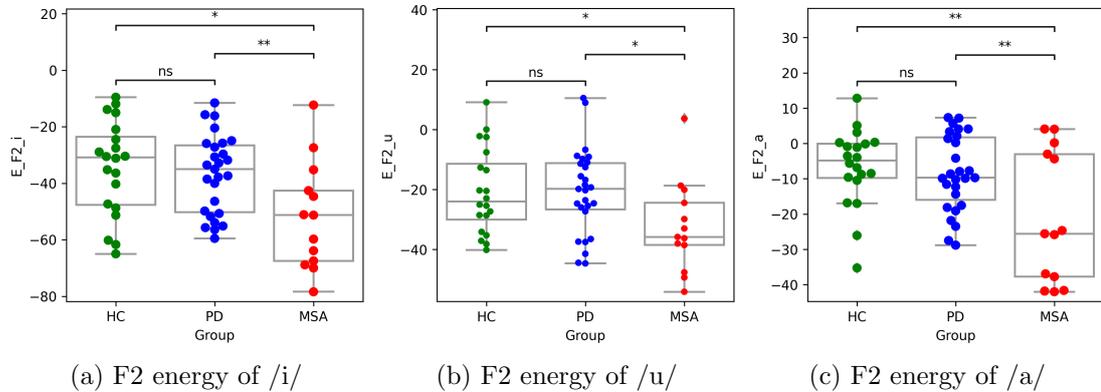


Figure 5.9: Power in second formant of three vowels /i/, /u/, and /a/; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, ns: not significant

Shortening of vowels results in a fast speaking rate which may affect VSA [304]. Table 5.11 details average vowel duration across three groups. Vowel duration of /i/ becomes shorter for PD (p=0.07) and MSA-P (0.03) compared to HC. Vowel duration of /a/ is shorter for only PD patients (p=0.02) compared to HC. Important to note, PD patients manifest decreased vowel duration for PD compared to HC. Vowel /u/ did not show any discrimination among groups. In addition, we did not find group differences between PD and MSA-P. Thus, it is not clear whether vowel duration has a direct relation with VSA.

Vowels	HC		PD		MSA	
	Mean (ms)	CV (%)	Mean (ms)	CV (%)	Mean (ms)	CV (%)
/i/	65.15	11.65	56.85	26.22	56.25	20.72
/u/	59.00	34.22	62.70	33.46	62.88	30.43
/a/	81.52	11.93	69.59	25.92	81.84	29.23

Table 5.11: Mean duration and coefficient of variance (CV) for vowels /i/, /u/, and /a/; Blue and red color represent higher impairment for PD and MSA respectively

5.5 Discussion

This study analyzed vowel articulation disorder for neurological disorders, particularly in parkinsonism (PD and MSA-P). Previous studies provided results about impaired vowel articulation either for general dysarthria or aiming discrimination of PD (early stage or disease progression) from the control group [93, 277, 251]. In addition to neurological disorder, other factors like hearing loss [220], people who stutter [31], hyper-nasality associated with cleft palate [103], and also depression may alter vowel properties. Hence, it is required to be very cautious while taking any conclusion regarding the cause of vowel articulation distortion. Moreover, the manner of articulation (clear, loud, intelligibility) or speaking task (sustained phonation, reading passage, monologue) may change the vowel properties [93, 251].

The present study is probably the first time investigated vowel articulation impairment in the differential diagnosis scenario (PD vs MSA-P). As previous study [251] showed that spontaneous speaking task is more suitable to capture impairment at even early stages of the disease, reading passage task is adopted to analyze vowel articulation in the present study. In the present study, discrimination between HC and PD was not observed (though it is not the primary goal). Important to note, the present study included male and female speakers together in the discrimination task. Encouraging discrimination between PD and MSA-P patients achieved in acoustic measures, VSA, VAI, FR, C-D of /a/ and /u/. Quantitative analysis shows that PD patients exhibit low impairment in vowel articulation. In contrast, MSA-P patients manifest more significant impairment in vowel articulation, probably due to higher deficits in articulators (tongue, jaw, lips) movement. Deficits in tongue elevation are observed using formant spacing feature C-D in vowels /a/ and /u/ for MSA-P patients compared to PD.

Formant frequencies have been estimated through various methods, e.g., visual inspection of a FFT spectrogram and automatic formant detection through LPC or cepstral analysis. Spectrum (FFT spectrum or LPC) or cepstrum based methods generally use a root solving or peak picking procedure [314] to identify formants. Different factors may introduce inaccurate formant frequencies [104, 150]. Not a single method is error-free, and sometimes a combination of all methods may ensure the accurate results. In this study, fusion methods were not employed, but subjective measures provided comparably reliable results. At the time of subjective analysis, weak energy was observed in F2 for vowels /i/ and /u/ for a group of PD and MSA patients. It might be a promising indicator for vowel articulation impairment. The highest difference in F2 of vowel /u/ was observed while comparing the subjective method and Praat method. It is probably due to the pathological factors which lead to inaccurate formant estimation.

The major limitation of this study is the small number of participants in the HC group and gender imbalance within each disease group. It is evident that the number of male and female speakers is not balanced across the disease groups. It may influence all the comparative studies. To avoid this discrepancy, gender independent acoustic measures could be targeted to analyze group specific vowel impairment, but small amounts of data is the main impediment. Present observation need to be validated with additional data.

Chapter 6

Consonant distortion in PD and MSA-P

Analysis of consonant distortion is mainly confined to obstruents (where airflow is partially or completely obstructed) due to its prevalence in speech disorders. Obstruents consist of stop plosives and fricatives. Unvoiced stops are mostly analyzed by burst (realization of articulator's closure) properties. On the other hand, manner of friction is examined for unvoiced fricatives. Properties of obstruents were explored in previous studies [32, 189].

6.1 Introduction: Consonants

Consonant sound production requires precise articulator's position and synchronization of laryngeal and supralaryngeal functions. Deficits in any of these functions yield imprecise consonants. In phonology, consonants are clustered by the manner of articulation and place of articulation. Table 6.1 presents word-initial consonant's grouping and respective logatome example.

Place of Articulation	Manner of articulation						
	Plosive		Fricative		Liquid	Nasal	
	Voiced	Unvoiced	Voiced	Unvoiced	Voiced	Voiced	
Bilabial	/b/- berdo, broto	/p/- perva, pataka				/m/ - mindou	
Labiodental			/v/- vonia	/f/-feju, frambi			
Dental	/d/ - dirou	/t/ - touca			/l/ - larni	/n/ - nouillo, nianfin	
Alveolar		/t/ - tunia	/z/ - zacu	/s/ - sochin, spegzi, psegra			
Palatal			/Z/ - jinin	/S/ - chastu			
Velar	/g/ - guizant, granfa	/k/ - quinsa, crancto			/R/ roursou		

Table 6.1: List of logatomes (pseudo-words) used for imprecise consonant and vowel analysis

Consonant distortion is common in neurological disorder [182, 58] as discussed in the Section 2.5. Evaluation of consonant distortion can reveal underlying deficits of articulators movement and synchronization of laryngeal and supralaryngeal activities. Consonant distortion across various diseases have been typically assessed using perceptual evaluation [182, 106, 67, 257, 155, 15, 197]. Stop-plosives, affricates, and fricatives were found to be the most impaired phoneme classes [182] for PD patients. Particularly, velar stops (/k/, /g/) and alveolar fricatives (/s/, /z/) were mostly impaired for PD. In another study [15], Amyotrophic Lateral Sclerosis (ALS) and PD groups manifest high articulation imprecision and intelligibility deterioration compared to Cerebellar Ataxia (CA) group in perceptual analysis. The latter study also observed predominant impaired closure in stop plosives for ALS, devoicing in voiced stop plosives for CA and PD groups, voicing in voiceless consonants for ALS group in the visual investigation. The presence of abnormal burst/closure in fricative /s/ was rare for ALS, CA, and PD. In study [197], PSP and MSA-P patients also manifest greater articulation impairment compared to PD in perceptual analysis. The latter study did not find a single speech dimension to differentiate PSP and MSA in perceptual analysis.

During the last 2 decades, a considerable effort has been produced to develop objective measures that assess consonant distortions in PD [6, 5, 209, 216, 154, 306, 255, 95]. In these studies, voice onset time (VOT) and Voice Onset Time ratio (VOTR) have been the most analyzed feature but with rather contradictory outcomes [75, 73, 209, 306]. Some of the researchers reported increased VOT [80, 209, 264, 255], and other studies observed no change in VOT [73, 239] for PD. Another study [75] also observed decreased VOT for PD. Most of the previous studies computed VOT from manual segmentation of pathological speech. Automatic VOT measurement for pathological speech is rare [209]. Studies of imprecise fricatives for neurological disorder are also limited [48, 154, 115]. Frication properties of fricatives were studied by mostly spectral moments. On the other hand, only a few studies have addressed consonant distortion in the differential diagnosis between PD and APS [15, 264, 306, 113, 255]. The same statement holds for dysarthria-based differential diagnosis in general. Indeed, while there exists a large amount of work on comparing PD and HC speech, there is only few studies on comparison/discrimination between PD and APS or between APS subgroups [47, 281, 126, 113, 43, 175, 59, 164]. The study [306] observed contrastive characteristics of VOT for voiced and voiceless stop plosives. While VOT and VOTR of voiceless stops prolonged for APS group compared to PD, MSA group distinctively showed reduced VOT (prevoicing) compared to PSP and PD in voiced stops. Another study [264] also observed prolonged VOT particularly for PSP patients compared to HC.

Selection of speech task is an important step for any kind of speech disorder analysis. Previous studies mostly considered reading text (paragraph) speech task to evaluate impaired consonants production [182, 66, 67, 15]. Important to note, in continuous speech, effect of co-articulation and speech rate may alter the consonant characteristics. To avoid effect of speech rate and/or co-articulation effect, single-word speech task would be more appropriate to observe distinctive patterns of imprecise articulation. Hence, bi-syllabic words were considered for unbiased analysis of stop

consonants [264, 306].

In the present study, consonant distortion is evaluated by subjective and objective analysis from word-initial consonants using pseudo-words, called Logatomes [168]. Among all the consonants, obstruents (plosives and fricatives) yield the most exciting results. In particular, we show that voiced obstruents manifest appealing, distinctive impairments in devoicing and VOT duration. In addition, unvoiced obstruents also exhibit distinguishing impairments, such as VOT duration, presence of burst in fricatives, multiple burst in stop plosives, and weak frication in fricatives. For consonants analysis, recorded logatomes were segmented and annotated by manual process.

6.2 Methodology

6.2.1 Database: Voice4PD-MSA

Total 59 speakers are selected from the Voice4PD-MSA database. It consists of 20 HC (10 male and 10 females), 26 PD (19 male and 7 female), and 13 MSA-P (5 male and 8 female) participants. Word-initial phoneme from 25 logatomes are considered for analysis.

6.2.2 Data processing of logatomes

To inspect distortion of particular consonant production, segmentation of that consonant sound need to be done for quantitative analysis. Accurate phoneme segmentation is a challenging but essential task for finding consonants disorders, particularly for pathological speech. Manual phoneme segmentation remains the gold standard, but it is a time consuming process for larger speech data. In the present study, targeted word-initial consonants and other phoneme were manually segmented, which will also be used as reference. This reference will be used for evaluating automatic phoneme segmentation accuracy.

Total 25 logatomes were recorded in continuous fashion in Voice4PD-MSA database. Each logatome represented by it's corresponding phonemes. Table 6.2 presents list of logatomes and it's phonetic units.

Each logatome was first manually segmented and then each phoneme within the logatome. Then, all the phonetic units were labeled in Praat toolkit by the criteria provided in study [73]. This criterion includes fundamental frequency (F0), formant frequencies transition, change in energy. Figure 6.1 shows the examples of the manually segmented phoneme of two logatomes.

Important to note, logatome "plerva" was mainly mispronounced as "klerva" by pathological speakers and healthy speakers. After observing this phenomenon, "plerva" was changed to "perva". To analyze /p/ homogeneously, "pataka" was taken from first part of the normally spoken /pa-ta-ka/ DDK task.

In voiced stop plosives, prevoicing part (e.g. /b_prv/ in Figure 6.1) and burst part are labelled separately. Prevoicing labeling will help to compute negative VOT,

SL. No.	Words	Phoneme	SL. No.	Words	Phoneme
1)	berdo	b e R d o	14)	nouillo	n u j o
2)	broto	b R o t o	15)	perva	p e R v a
3)	chastu	S a s t y	16)	psegra	p s e g R a
4)	crancto	k R @ k t o	17)	quina	k cinq s a
5)	dirou	d i R u	18)	roursou	R u R s u
6)	feju	f e Z y	19)	sochin	s o S cinq
7)	frambi	f R a m b i	20)	spegzi	s p e g z i
8)	guizant	g I z @	21)	touca	t u k a
9)	granfa	g R @ f a	22)	tunia	t y n j a
10)	jinin	Z i n cinq	23)	vonja	v o n j a
11)	larni	l a R n i	24)	yuni	j y n i
12)	mindou	m cinq d u	25)	zacu	z a k y
13)	nianfin	n j @ f cinq			

Table 6.2: List of logatomes and it's phonetic representation

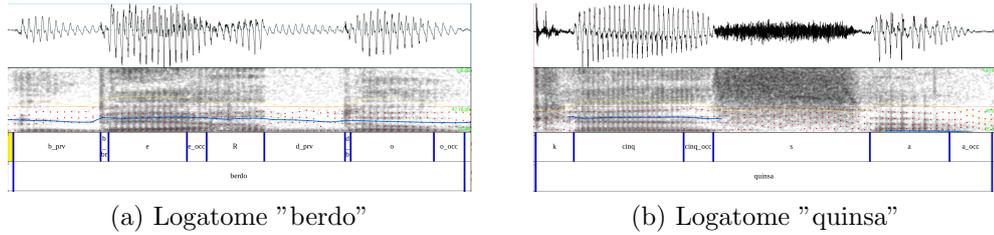


Figure 6.1: Manually labelling of phonetic unit of logatomes "berdo" and "quina"

whereas burst labeling will help to evaluate burst strength. In vowels, the steady part and occlusion part are labeled separately.

6.2.3 Methods to evaluate logatomes

Logatomes are evaluated first by perceptual analysis to get overall distortion in logatomes. Next, distortion in logatomes is assessed by visual acoustic analysis (by spectrogram). Initially, consonants and vowels are evaluated by specific criteria. Consonants are evaluated by six criteria as mentioned in Figure 6.2. Voiced stops are analyzed by prevoicing properties, energy in transient, and energy in high frequency. Unvoiced stops are evaluated by energy in the closure part, energy in transient, energy concentration in low and high frequency. Fricatives are assessed by low frequency and high frequency energy. Nasal consonants are difficult to assess by visual inspection, however, energy concentration in low frequency band (700-1200 Hz) is considered as degree of nasality.

Vowel sounds are also analyzed by five acoustic parameters. Five acoustic parameters are provided in the Figure 6.3. In formant transition, slow and/or flat transition are considered to be impaired formant transition. Vowel impairments are also assessed

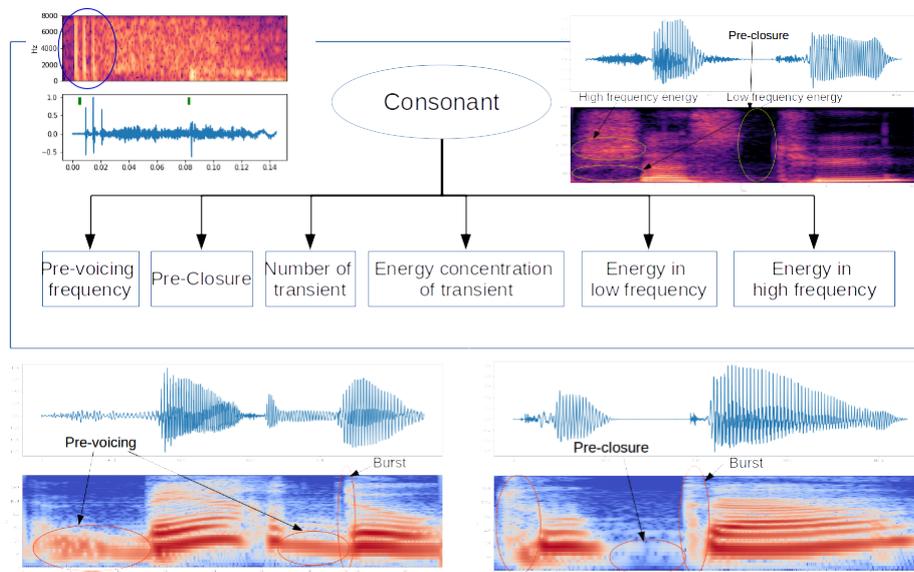


Figure 6.2: Acoustic parameters for consonants evaluation

by harmonic richness and/or breaks, pitch disruption, and occlusion part (energy).

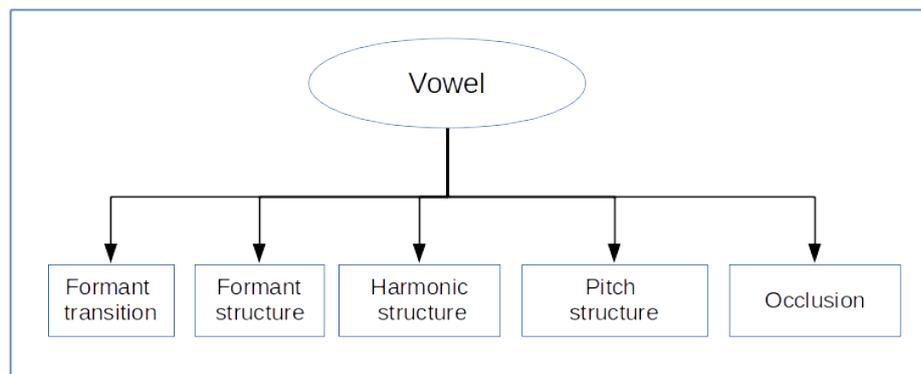


Figure 6.3: Acoustic parameters for vowels evaluation

Overall deficits in consonants, later on, are summarized by two different aspects. The first dimension consists devoicing in voiced obstruents, anti-spirantization in fricatives, multiple burst in stop plosives, and tentative voicing in voiceless obstruents. Latter mentioned deficits primarily attributed to reduced synchronization functionality of laryngeal and supralaryngeal activities. Another dimension represents mainly movements disorder. The second dimension consists weak transient and resonance energy.

In subjective and visual analysis, stop plosives and fricatives provided encouraging distinctive deficits for MSA and PD patient groups. In the following sections, voiced obstruents are first assessed. Next, unvoiced obstruents are investigated for unique distortion.

6.3 Voiced obstruents

Voiced obstruents consist of voiced stop plosives and fricatives. Voiced stop plosives (/b/, /d/, and /g/) are characterized as vocal folds vibration in closure (prevoicing) followed by a burst. In general, burst is weak for voiced stop plosives compared to unvoiced stops. Voiced fricatives are characterized by vocal folds vibration along with frication. Possible impairment in voiced obstruents would be weak voicing, absence of burst, and weak frication. Figure 6.4 and 6.5 represents spectral properties of voiced obstruents from healthy speaker. Six voiced obstruents present a prominent voice bar. In particular, fricatives show voicing bar along with frication in higher frequency.

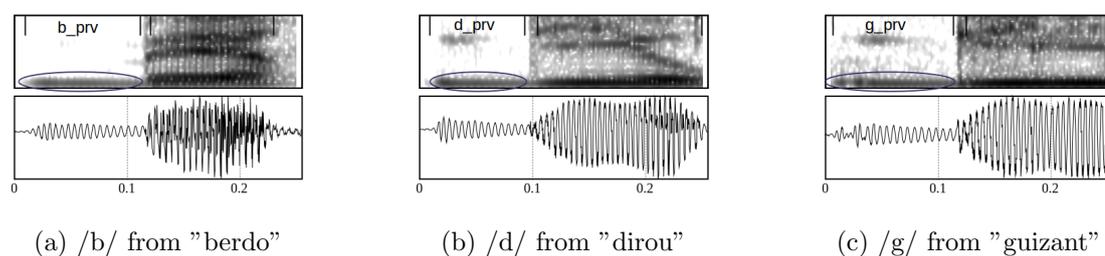


Figure 6.4: Voiced stop plosives /b/, /d/, and /g/ from "berdo", "dirou", and "guizant" consequently; circled box represents voicing bar

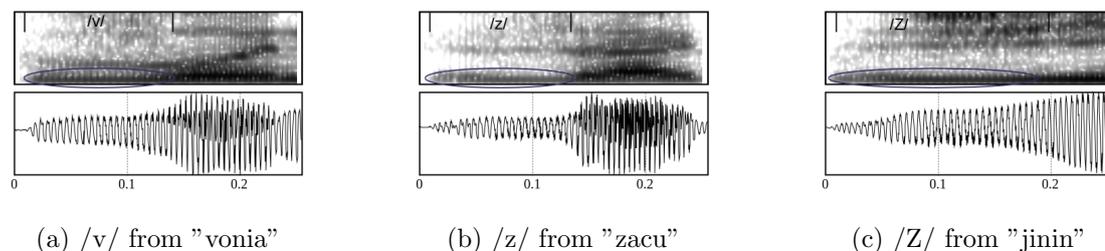


Figure 6.5: Voiced fricatives /v/, /z/, and /Z/ from "vonía", "zacu", and "jinín" consequently; circled box represents voicing bar

According to morphology, word-initial voiced obstruents are extracted from either Consonant-Vowel (CV) syllable or Consonant-Consonant-Vowel (CCV). In this study, at first voiced obstruents from CV syllable were analyzed. As example, /b/ from syllable /be/, /d/ from syllable /di/, and /g/ from syllable /gi/, /v/ from syllable /vo/, /z/ from syllable /za/, /Z/ from syllable /Zi/ were considered for analysis. In the next step, voiced obstruents from CCV syllable are considered for analyzing difficulty of neurological patients. In CCV syllable, quick articulators transition are required in which pathological speaker may manifest prominent deficits. As example, /b/ from /bRo/ (logatome "broto") and /g/ from /gR@/ (logatome "granfa") were considered in this analysis.

As for the French language, to the best of our knowledge, there exists no study comparing consonant production between PD and MSA. In [15], a comparison has been subjectively performed (spectrogram visual inspection) but among PD, Amyotrophic Lateral Sclerosis, and Cerebellar Ataxia. In the present study, French consonant distortion was first analyzed for differential diagnosis between PD and MSA-P. To do so, subjective and objective analyses were carried out of word-initial voiced obstruents. Among all the consonants, obstruents (plosives and fricatives) yield the most encouraging results. In particular, voiced obstruents manifested appealing, distinctive impairments, in term of devoicing and VOT duration. In the following section, the presence of devoicing in voiced obstruents and change of VOT in voiced stops will be studied by visual analysis followed by objective/quantitative analysis. Another possible acoustic cue could be the weak transient/burst in the voiced stop plosives. However, it is challenging to evaluate degree of weakness in bursts, even in subjective analysis. Hence, a weak burst for voice stops is not considered in this study.

6.3.1 Devoicing analysis by Visual method

The visual method is mainly performed by looking into spectrograms of voiced obstruents. To this purpose, Praat software is used to generate wide-band spectrogram. Devoicing is assessed by the total or partial absence of voicing bars (due to vocal folds vibration in closure segment) in the realization of voiced obstruents. In the following section, obstruents from CV syllable will be explored first, followed by CCV syllable.

Devoicing from CV syllable

Visual inspection of devoicing in voiced obstruents reveals encouraging results, particularly for the velar place of articulation. Figure 6.6 shows an example of spectrograms of the consonant /b/, /d/, and /g/ pronounced commonly by an HC, with a partial/total devoicing by a PD and with a total devoicing by an MSA-P patient. Another phenomenon that could be considered as partial devoicing, the occurrence of voicing bars with weak energy. However, for sake of clarity, reproducibility, and to reduce subjectivity effects, we did not use this criterion in our assessment of devoicing.

Figure 6.7 presents example of devoicing in fricative /v/, /z/, and /Z/. Complete devoicing in /v/, /z/, and /Z/ are realized as /f, p/, /s/, /S/ respectively.

The presence of devoicing in individual voiced obstruents by visual analysis is presented in Table 6.3. Devoicing is more frequent in stop plosives compared to fricatives. In particular, devoicing is more frequent in velar (/g/) and bilabial (/b/) compared to alveolar (/d/). MSA-P patients manifest frequent devoicing compared to PD and HC.

In visual analysis 69.23% (9 out of 13) of MSA-P and 11.5% (3 out of 26) of PD presented devoicing in at least one obstruent from CV segment. In particular, devoicing was not observed in any of /b/ nor /d/ in PD, while 30.76% (4 out of 13) of MSA-P showed devoicing in these consonants. This suggests that devoicing of /b/ or/and /d/ could be a signature of MSA-P. However, recent study [15] reported that 37% of PD presented devoicing in /d/ or /g/. In contrast, analyzing present data,

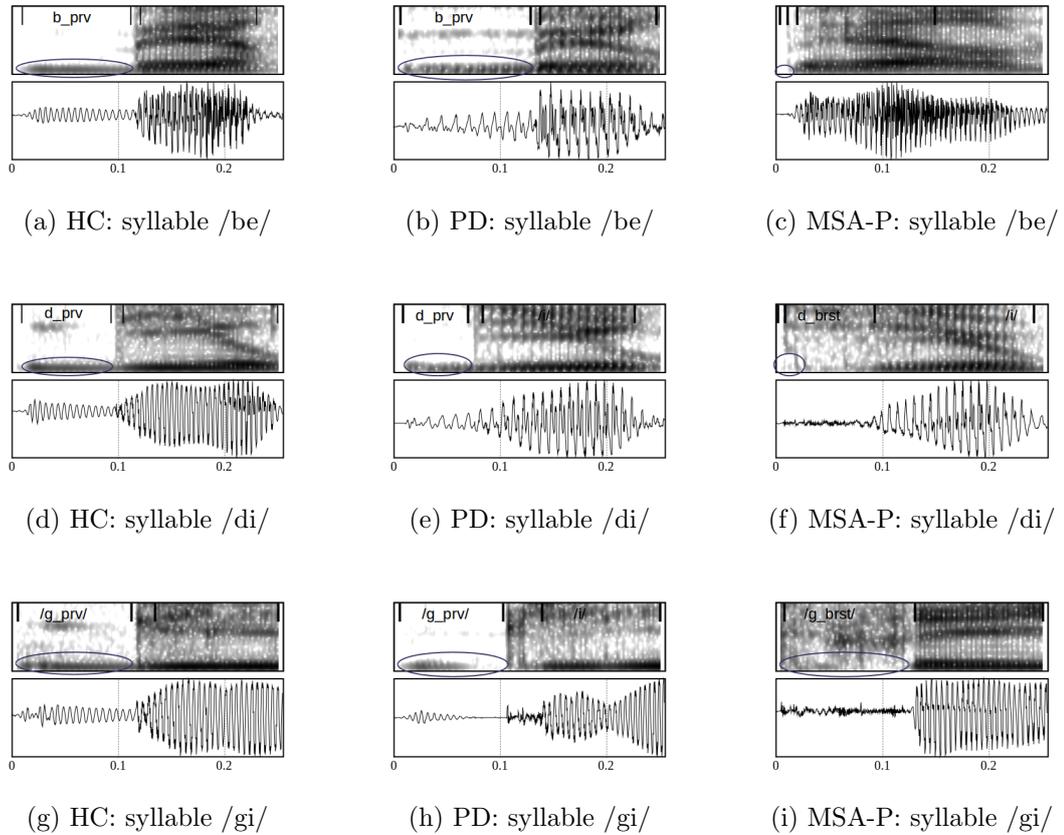


Figure 6.6: Example of no/partial/total devoicing of /b/, /d/, and /g/ in a HC/PD/MSA-P (top). Example of normal/shorter/vanishing VOT of /g/ for the same HC/PD/MSA-P (bottom)

Obstruents	Consonant	Devoicing (%)		
		HC	PD	MSA-P
Plosives	/b/ (/be/)	0	0	23
	/d/ (/di/)	0	0	7.69
	/g/ (/gi/)	0	3.8	30.76
Fricatives	/v/ (/vo/)	0	3.8	15.38
	/z/ (/za/)	0	3.8	0
	/Z/ (/ji/)	0	7.6	23

Table 6.3: Devoicing (%) of individual voiced obstruents in HC, PD, and MSA-P groups

only 3.8% of PD showed devoicing in /g/ (and thus in /d/ or /g/), while 38.46% (5 out of 13) of MSA-P showed devoicing in /d/ or /g/. This difference in PD is might be due to the relatively small size of the dataset as compared to the one of [15] and different speech tasks. The study [15] used a text reading task for consonant distortion. As for voiced fricatives, 38.46% (5 out of 13) of MSA-P and 11.53% (3

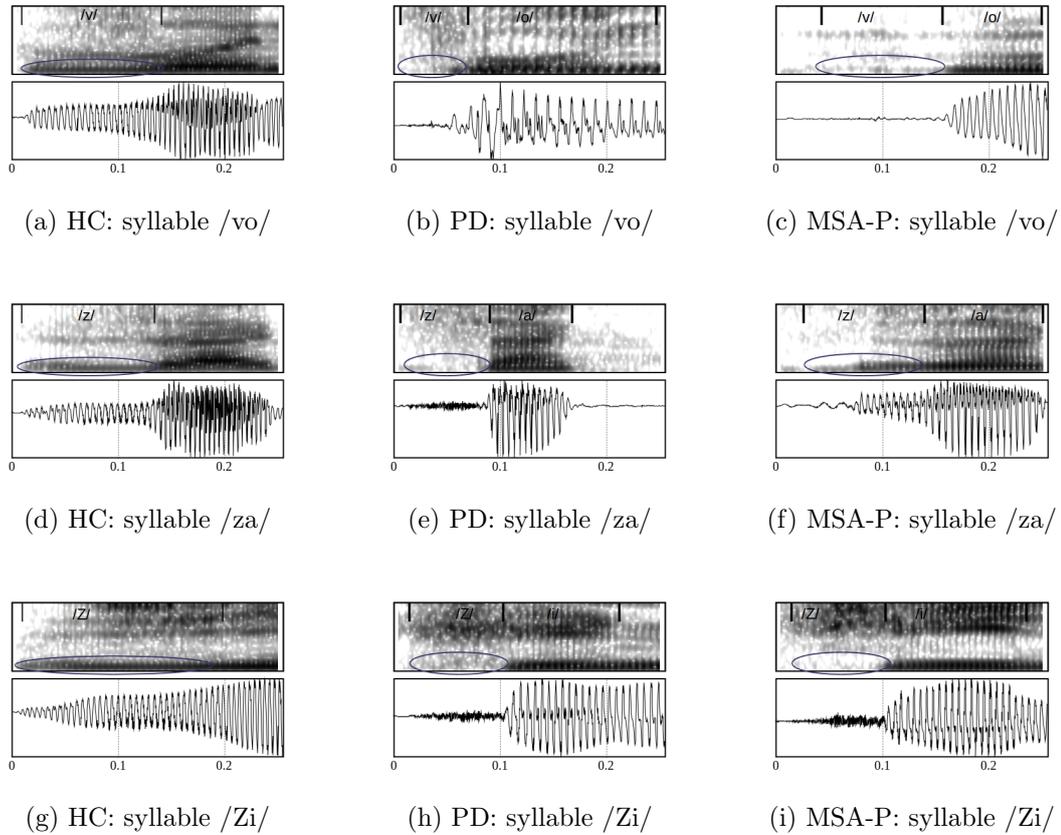


Figure 6.7: Example of no/partial/total devoicing of /v/, /z/, and /Z/ in a HC/PD/MSA-P

out of 26) of PD presented devoicing in at least one of /v/, /z/ and /Z/.

Devoicing from CCV syllable

Table 6.4 summarized presence of devoicing in voiced stop by visual analysis.

POA	Consonant	Devoicing (%)		
		HC	PD	MSA-P
Plosives	/b/ (/bRo/)	0	3.8	7.69
	/g/ (/gR@/)	0	15.38	69.23

Table 6.4: Devoicing (%) of individual voiced obstruents from CCV syllables in HC, PD, and MSA-P groups by visual analysis

Visual analysis revealed greater impairment for MSA-P patients than PD in velar consonant /g/ from the CCV segment (from logatome "granfa"). Total 69.23% (9 out of 13) of MSA-P patients manifest complete devoicing, whereas 15.38% (4 out of 26) of PD manifests mostly partial devoicing in only /g/. This result indicates the importance of appropriate logatome selection. Timing and movement deficits

become predominant in /g-R-@/ syllable. Important to note, MSA-P patients showed co-occurrence of devoicing in /b/ and in /g/, which is not always valid for PD.

6.3.2 Devoicing analysis by objective analysis

An objective measure to detect devoicing was provided in this section. Given a labeled consonant, a simple way to assess the total or partial absence of voicing is to consider the degree of voicing measure:

$$DV = \frac{\text{number of voiced frames}}{\text{total number of frames}} (\%)$$

Using the very soft threshold 50%, we can fairly consider that an obstruent (or a consonant in general) is devoiced if $DV < 50\%$ (Praat software was used to compute DV by exploiting fundamental frequency (F0)). Other objective criteria can also be used to define and detect devoicing, and present measure is, however easy to interpret and reproduce. Given a speaker, let's now define the total degree of voicing DVT as the minimum DV over all devoiced obstruents. DVT is a simple quantification of the global amount of voicing in the obstruents devoiced by a given speaker (if any). DVT will be thus always less (resp. higher) than 50% in the presence (resp. absence) of devoicing. In objective analysis, presence of devoicing mostly matched with visual analysis. In addition, distortion in prevoicing is also detected in some cases.

Devoicing from CV syllable

Voiced obstruents from CV syllables would infer synchronization of the consonant to vowel transition. Table 6.5 presents the degree of voicing (DV) of individual voiced obstruents.

		HC Mean/SD (Range)	PD Mean/SD (Range)	MSA Mean/SD (Range)
DV	/b/	89.26 / 10.38 (62.5 - 100.0)	90.6 / 9.59 (53.85 - 100.0)	78.07 / 30.56 (0.0 - 96.77)
	/d/	89.28 / 8.9 (66.67 - 96.15)	90.28 / 12.94 (40.0 - 100.0)	86.06 / 26.45 (0.0 - 100.0)
	/g/	91.9 / 8.64 (66.67 - 100.0)	85.27 / 14.85 (30.77 - 96.15)	66.19 / 39.32 (0.0 - 92.31)
	/v/	89.58 / 14.77 (53.57 - 100.0)	91.2 / 19.37 (0.0 - 100.0)	80.4 / 33.54 (0.0 - 100.0)
	/z/	91.56 / 11.51 (64.0 - 100.0)	87.75 / 18.07 (11.76 - 100.0)	88.69 / 11.62 (57.14 - 100.0)
	/Z/	93.47 / 8.39 (64.0 - 100.0)	87.45 / 22.41 (4.55 - 100.0)	71.86 / 39.68 (0.0 - 98.04)

Table 6.5: Degree of voicing (Average/standard deviation, range) of six obstruents

Table 6.6 presents percentage of impaired obstruents in groups by devoicing.

Using this measure, the assessment of devoicing matched perfectly with the visual observations, that is, 69.23% of MSA-P and 11.5% of PD presented devoicing. The value of DVT is shown in Figure 6.8 for each participant (see the projection over the DVT dimension). Along this dimension, one can note the large margin between subjects manifesting devoicing and the others. This suggests that devoicing is generally strong when it occurs and thus easy to detect objectively by the standard tool Praat.

Obstruents	Consonant	Devoicing (%)		
		HC	PD	MSA-P
Plosives	/b/ (/be/)	0	0	23
	/d/ (/di/)	0	3.8	7.69
	/g/ (/gi/)	0	3.8	30.76
Fricatives	/v/ (/vo/)	0	3.8	15.38
	/z/ (/za/)	0	3.8	0
	/Z/ (/ji/)	0	7.6	23

Table 6.6: Devoicing (%) of individual voiced obstruents in HC, PD, and MSA-P groups by objective analysis

Devoicing from CCV syllable

Prevoicing part of /g/ from syllable /gr@/ is most disrupted for MSA-P patients. Table 6.7 presents DV of stop plosives from CCV syllable.

DV		HC	PD	MSA
		Mean/SD (Range)	Mean/SD (Range)	Mean/SD (Range)
	/b/ (/br/)	85.72 / 24.91 (0.0 – 100)	90.33 / 9.69 (53.85 - 100)	71.71 / 37.28 (0 – 96)
	/g/ (/gr/)	78.36 / 14.72 (41.67 – 92.31)	74.06 / 26.3 (0 – 100)	26.74 / 41.82 (0 – 90.91)

Table 6.7: Degree of voicing (Average/standard deviation, range) of /b/ and /g/ from CCV syllable

Overall distortion in prevoicing by disease groups is provided in the Table 6.8. Important to note, 9 out of 13 MSA-P patients manifest complete devoicing.

POA	Consonant	Devoicing (%)		
		HC	PD	MSA-P
Plosives	/b/ (/bRo/)	0	3.8	7.69
	/g/ (/gR@/)	0	15.38	69.23

Table 6.8: Devoicing (%) of individual voiced obstruents from CCV syllables in HC, PD, and MSA-P groups by objective analysis

The objective measure of /g/ from CCV syllable also exactly match the visual analysis, that is, 69.23% of MSA-P, and 15.38% of PD patients manifest devoicing. It is an encouraging result where MSA-P patients display predominant devoicing compared to PD in particularly /g/. Another important observation related to devoicing computation is that only pitch frequency does not always reflect the original scenario. Hence, sometimes detected devoicing in the visual analysis is not found in objective analysis. This suggests to use other acoustic parameter like energy in prevoicing could be helpful for complete devoicing measures.

Overall, these results show that devoicing can be a valuable cue for differential diagnosis between PD and MSA-P. However, this cue alone is not sufficient to achieve this diagnosis with a high accuracy.

6.3.3 VOT analysis of voiced plosives

As mentioned earlier, VOT is among the most studied features in consonant distortion. VOT is generally associated with plosives and is defined as the duration between the vocal fold vibration starts relative to the release of the plosive (there exist, however VOT definitions for other consonant types [4]). In the case of voiced plosive, vibration begins before the release; VOT is thus considered as negative. When negative VOT tends to 0, it actually corresponds to a total devoicing. In order to avoid a potential dependency on speaking rate, VOT ratio (VOTR) is sometimes considered. VOTR is defined as VOT divided by the duration of whole syllable [73].

VOT analysis of CV syllable

The purpose of this section is to determine whether VOT analysis of voiced plosives (/b/, /d/, and /g/) from CV syllables can yield another distinctive cue (hopefully complementary to devoicing). Table 6.9 presents VOT of stop plosives and voiced fricatives. In velar stop /g/, VOT analysis provided encouraging differentiation among groups.

		HC Mean/SD (Range)	PD Mean/SD (Range)	MSA Mean/SD (Range)
VOT	/b/	110.18 / 46.81 (0.99 - 195.3)	100.81 / 34.38 (35.52 - 181.59)	95.18 / 46.86 (3.95 - 167.59)
	/d/	103.56 / 25.85 (63.05 - 145.81)	103.29 / 44.44 (11.73 - 227.7)	80.88 / 48.66 (2.23 - 139.42)
	/g/	103.52 / 27.58 (70.14 - 166.33)	80.73 / 34.42 (19.5 - 148.61)	41.96 / 33.49 (1.07 - 107.68)
	/v/	152.1 / 41.99 (81.09 - 243.02)	140.16 / 38.35 (76.76 - 209.57)	99.63 / 52.96 (8.13 - 162.59)
	/z/	130.94 / 35.08 (91.51 - 216.67)	132.77 / 50.13 (45.18 - 282.43)	122.07 / 25.98 (84.29 - 162.06)
	/Z/	168.02 / 44.96 (123.57 - 304.51)	150.11 / 30.05 (94.44 - 223.48)	138.76 / 61.52 (43.57 - 254.22)

Table 6.9: VOT analysis of individual voiced obstruents from CV syllable

Using the manual segmentation, the statistical group difference between HC and PD, HC and MSA-P, PD and MSA-P were computed using VOT and VOTR. Table 6.10 shows the obtained p-value of each group difference. It is observed that, for VOT, statistical significance between MSA-P and the other groups was achieved only for /g/. More interestingly, this impairment was more severe in MSA-P than in PD. The waveforms of Figure 6.6 show an example of such a distortion. This trend was confirmed by VOTR with an additional group difference between PD and HC. Globally, this is in accordance with the findings of [306] which reported shorter VOT and lower VOTR for MSA averaged on all voiced plosives (with Czech patients). The present analysis cannot, however confirm the same statement for /b/ and /d/. On the other hand, VOT/VOTR of /g/ can be confidently considered as a valuable cue for the differential diagnosis. However, as devoicing, this cue alone is insufficient to achieve this diagnosis with high accuracy.

VOT analysis of CCV syllable

Negative VOT of /b/ and /g/ are computed from manual segmentation of logatome "broto" and "granfa". Reduced VOT is more frequent for MSA-P patients compared

Feature	Consonant	HC vs PD	HC vs MSA-P	PD vs MSA-P
		p-value		
VOT	/b/	0.36	0.15	0.41
	/d/	0.8	0.9	0.66
	/g/	0.07	0.004	0.03
VOTR	/b/	0.08	0.001	0.029
	/d/	0.26	0.03	0.161
	/g/	0.002	0.0009	0.014

Table 6.10: Results of acoustic speech analyses for three voiced plosives including /b/,/d/ and /g/. Bold numbers indicate group difference ($p < 0.05$)

to PD in /g/ as presented in the Table 6.11. Both PD and MSA-P patients manifest reduced VOT in /b/ compared to HC.

		HC Mean/SD (Range)	PD Mean/SD (Range)	MSA Mean/SD (Range)
VOT	/b/ (/br/)	90.55 / 29.3 (43.88 - 136.79)	68.91 / 35.22 (16.88 - 159.25)	64.23 / 37.44 (5.6 - 134.97)
	/g/ (/gr/)	59.81 / 18.33 (30.2 - 97.21)	60.24 / 32.0 (3.21 - 131.63)	22.8 / 33.71 (1.43 - 108.14)

Table 6.11: VOT analysis of individual voiced obstruents from CCV syllables

Table 6.12 presents group differences result for voiced stops from CCV syllables.

Feature	Consonant	HC vs PD	HC vs MSA-P	PD vs MSA-P
		p-value		
VOT	/b/	0.047	0.043	0.705
	/g/	0.962	0.0008	0.0019

Table 6.12: VOT analysis of /b/ and /g/ from CCV syllable. Bold numbers indicate group difference ($p < 0.05$)

6.3.4 Classification of PD and MSA-P

Given the findings of the previous sections, it is natural to proceed with an analysis over the 2 deviant speech dimensions, devoicing and VOT of /g/ ($VOT_{/g/}$).

Classification using CV syllable

Figure 6.8a shows the biplot of DVT w.r.t to $VOT_{/g/}$. Using our HC data as summarized in Table 6.9, the mean/standard deviation of the VOT of /g/ is $-103/22(ms)$, which is in accordance with the $-109/32(ms)$ using single word as protocol reported in [306] and [73] (the latter reported the mean only). As we did for devoicing, we can set a very soft threshold at $-60ms$ above, which we confidently consider that a VOT impairment of /g/ is occurring.

It is observed that all but one MSA-P manifested devoicing or/and short VOT. Thus using the simple decision tree of Figure 6.9, with the soft thresholds $DVT = 50\%$

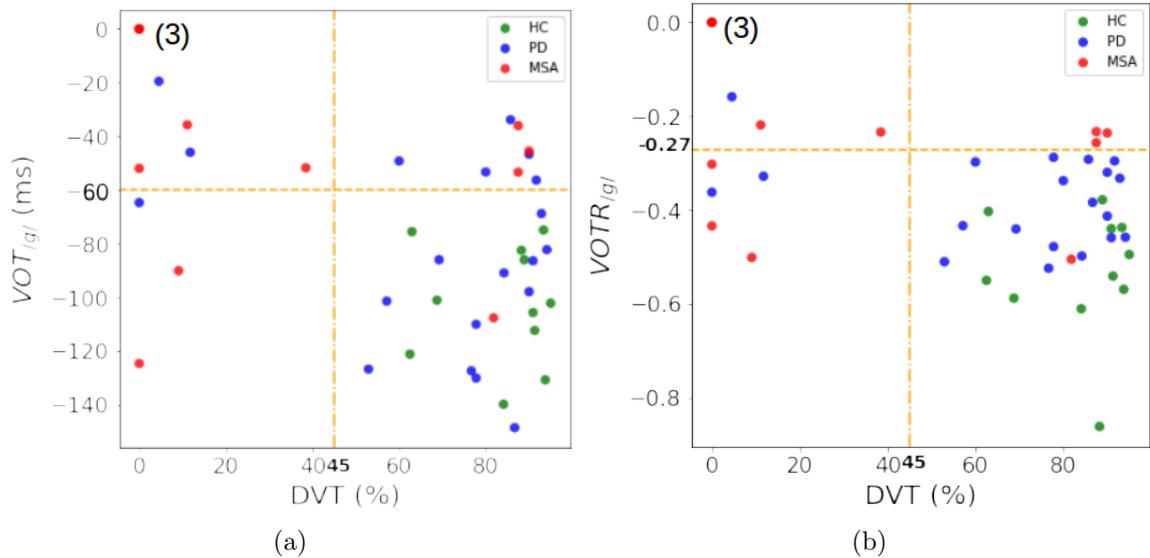


Figure 6.8: Biplot of $DVT(\%)$ w.r.t to $VOT_{/g/}$ and $VOTR_{/g/}$ (dotted line represent decision thresholds); (3) means that 3 MSA-P patients have same coordinates (total devoicing)

and $VOT_{/g/} = -60ms$, we obtain an accuracy of 72%, with a high sensitivity (correctly classified MSA-P) of 92% but a low specificity of 60%. This means that a mis-diagnosis of MSA-P presenting devoicing or short VOT of /g/ is unlikely. However, this statement does not hold for PD.

Figure 6.8b shows the biplot of DVT w.r.t to $VOTR_{/g/}$. We see now that, along the $VOTR_{/g/}$ dimension, a separation appears between the 3 MSA-P and 5 PD, which were confused using $VOT_{/g/}$ (right top rectangle of 6.8b). If we target only classification score and replace the threshold $VOT_{/g/} = -60ms$ by $VOTR_{/g/} = -0.28$ in the decision tree, then the specificity increased to 85% and the accuracy to 87.5%. However, the threshold -0.28 is likely overfitted to our data. Thus, given the small amount of instances of /g/ and its following vowels, we cannot confidently claim that VOTR is a better feature for discrimination than VOT. These results show however that the prevoicing duration of /g/ (and probably all voiced plosives) could be a complementary cue to devoicing of obstruents in order to achieve a high accuracy differential diagnosis between PD and MSA-P.

Overall, the results (along with literature reporting) show that devoicing of voiced obstruents and VOT of /g/ are 2-distinctive and deviant speech dimensions worth considering in the differential diagnosis between PD and MSA-P.

Classification using CCV syllable

It was observed in the previous section that devoicing in /g/ alone provided an encouraging distinction between PD and MSA-P. DV of /g/ from CCV syllable reveals 9 MSA-P patients manifest complete devoicing whereas only 1 PD showed complete

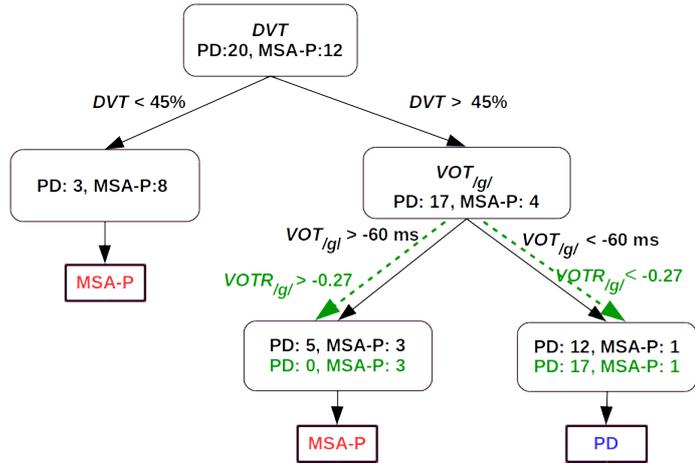


Figure 6.9: Decision tree using DVT and $VOT_{/g/}$ or $VOTR_{/g/}$ (in green) dimensions for discrimination between PD and MSA-P

devoicing and 3 PD manifest partial devoicing as shown in Figure 6.10. Interesting to note, $VOTR$ provided marginally better discrimination than VOT which can be explained as vowel prolongation.

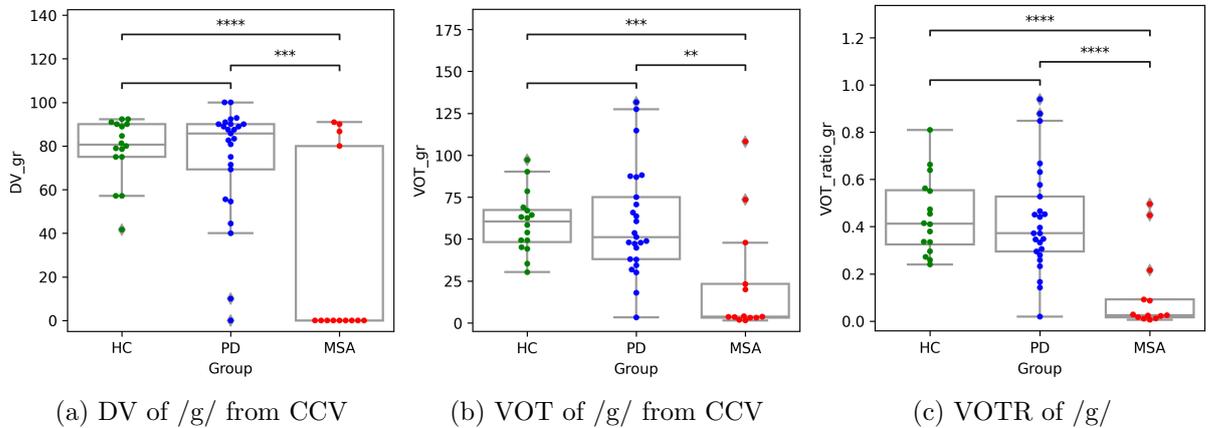


Figure 6.10: DV, VOT, and VOTR of /g/ from CCV syllable /gR@/

However, DV parameter alone is not enough to get high accuracy in classification.

6.3.5 Discussion

This work constitutes the first study that attempts to highlight distinctive cues in the distortion of French voiced obstruents realization in PD and MSA-P. This work showed the importance of particular logatome selection for analyzing devoicing in PD and MSA-P. The present study observed that syllable /gR@/ is more effective than syllable /gi/ regarding devoicing in /g/. The current results partially confirmed previous findings on negative VOT with other languages [306]. Indeed, it has been

found that VOT of the voiced plosive /g/ was significantly reduced in MSA-P while it was customary for most PD. On the other hand, VOT is not the only factor in the distortion of voiced plosives (and obviously fricatives). It showed that the absence of voicing leads was the main factor of voiced obstruents distortions and is the most distinctive cue between PD and MSA-P (in the production of voiced obstruents). Moreover, there was a perfect matching in devoicing assessment between perceptual and objective evaluations. This supports a potential use of devoicing in clinical practice as an additional tool for examining patients with a suspicion of MSA-P. The present analysis also showed that the combination of VOT and devoicing could significantly improve the differential diagnosis accuracy.

VOT impairment can be explained by a difficulty in initiating articulation resulting from a deficit in maintaining the speech motor program [306]. The latter is a characteristic of hypokinetic dysarthria, a known feature in both PD and MSA. Accurate production of word-initial voiced plosives requires precise coordination between glottal opening and articulatory closure. Devoicing is a manifestation of an impairment of such coordination. This is a characteristic of ataxic dysarthria, which is known to manifest in MSA. The present results are thus in accordance with the consensus that PD develops essentially hypokinetic dysarthria while MSA develops mixed type dysarthria. More importantly, since ataxia seems responsible for devoicing, the latter might manifest in early disease stages. If proven, devoicing would thus constitute a valuable deviant speech dimension to consider in the early differential diagnosis.

There are some limitations to our study. The most significant one is the relatively small dataset size due to the difficulty of recruiting patients, particularly with a rare disease such as MSA-P. We are, however, continuing the effort of recruitment. Moreover, the dataset is unbalanced in gender, we cannot thus exclude that gender-specific effects bias our findings. Another limitation is that we used only one consonant instance per speaker. We do not thus know how the results stand to intra-speaker pronunciation variability. We can expect, however, that the restriction reduces the effect of such variability to word-initials. From this perspective, our study should be considered a promising first step in analyzing French voiced obstruents in PD and MSA-P. Our findings need to be confirmed by additional data. This is the purpose of our on going research.

6.4 Unvoiced obstruents

6.4.1 Introduction: unvoiced obstruents

Two types of obstruents (stop plosives and fricatives) are available according to manner of obstruction. Unvoiced stop plosives are produced by complete blocking of air in oral cavity followed by sudden release. In the occlusion, vocal folds do not vibrate. On the other hand, fricatives are produced by partial constriction in oral cavity. Thus manner of constriction is an important landmark to identify obstruents. In this analysis, we consider to analyze word-initial consonants due to presence of frequent disorder observed in subjective analysis. Thus segmentation of word-initial unvoiced obstruents is the major step which can be accomplished by detecting first vowel onset. Next, analysis of constriction will be analyzed from segmented signal.

Vowel onset detection

In automatic VOT measure or abnormal presence of CBT, it is required to detect Vowel Onset Point (VOP) for selecting word initial phoneme (from CV or CCV syllable). Most of the phoneme segmentation task also employed vowel onset detection to separate vowels from other phonetic units. For vowel onset detection, several methods were found, consisting of temporal as well as spectral features in unsupervised and supervised approaches. Periodicity in signal is a very important acoustic cue for vowel onset detection. Hence, autocorrelation or cross correlation methods were widely used for F0 frequency estimation [35] which is mostly used for voicing onset detection. Other pitch tracking methods also proposed in previous studies such as a robust algorithm for pitch tracking (RAPT) [295], Robust Epoch And Pitch Estimator (REAPER) [<https://github.com/google/REAPER>], YIN [60]. Both methods, RAPT and REAPER used Normalized Cross-correlation (NCC). In other studies [64, 12], Maximum NCC (MNCC) on time series data was adopted for vowel onset. The recent study [238] used simultaneously zero frequency filtering (ZFF) and YIN method for vowel onset for phoneme segmentation task. The above methods were primarily evaluated on TIMIT database. The study [114] exploited ZCR, variance of auto-correlation function (ACR), power (PWR) in Gaussian Mixture model to detect voiced regions. The latter method was evaluated on continuous speech tasks from PD patients. Another study [215] also used Praat toolkit [35] for computing pitch to detect vowel onset and offset for HC and PD subjects.

The spectral methods based on the appearance of rapidly increasing resonance peaks in the amplitude spectrum [108]; wavelet transform [134] was also employed for VOP. Instead of a single acoustic feature, combination of spectral peak, excitation source, and modulation spectrum rendered improved VOP accuracy compared to individual feature [233]. Excitation source energy from LP residuals was also exploited in the study [232] for improved vowel onset and offset point. Bessel features was also used for emphasizing vowel region in the study [263].

In the study [133], zero-crossing rate, energy, and pitch information was used in neural network for VOP. Seven acoustic features consisting different spectral band

energies, Wiener entropy, zero crossing, F0 values by RAPT method were used in Recurrent Neural Network (RNN) which also yielded very high accuracy for VOP detection for single word [9], which is implemented in Dr.VOT toolkit.

In this study, we consider unsupervised methods for first vowel onset detection. We also compare the vowel onset detection accuracy for particularly pathological speech.

Consonant-consonant (CC) segmentation

The phoneme boundary detection method can be categorized by unsupervised text-independent and supervised text-dependent methods. Unsupervised context-independent phoneme segmentation remain an interesting field of research because of its degree of freedom. In previous research, several speech features and different methods were adopted to improve the phoneme segmentation or the phoneme boundary detection. The study [96] used short term frequency features over different frequency bands and Bayesian Decision Surface (BDS) to yield 80% accuracy for TIMIT database. Unsupervised phoneme segmentation was implemented using MFCCs features and Rate Distortion (RD) method in the study [325]. The later study yielded 89% recall score and also outperform methods proposed in studies [70, 72]. Auditory attention features was exploited in the study [139] which outperform other text independent phoneme segmentation [70, 325]. Several studies [140, 269] used different acoustic features such as cepstral coefficients (CFCC), perceptual linear prediction cepstral coefficients (PLPCC) and RelAtive SpecTrAl (RASTA)-based PLPCC and Mel frequency cepstral coefficients (MFCC) in Spectral Transition Measure (STM) [] for phonetic segmentation task. Another study [238] reported 95.4% accuracy (by 10 msec threshold) using a rule based approach. The latter study used Zero Frequency Filter, power spectrum of autocorrelation signal and peak counting methods. Another unsupervised training approach was also recently developed in the study [165] which yield a recall score 83.55% for TIMIT database.

Supervised methods were also developed for improving speech segmentation accuracy. In the study [221], Mel Frequency Cepstral Coefficients (MFCC) and it's delta features were used in context-independent HMMs which yield 97% accuracy (20 ms tolerance). MFCCs and Hidden Markov Models (HMMs) yield 94.84% accuracy using 30 ms tolerance [39]. A bidirectional LSTM network was used in the study [81] which provided high phoneme boundary detection accuracy. Montreal Focred Aligner (MFA) also developed using Kaldi speech recognition toolkit, which also can be used for phoneme segmentation [193]. MFA performs well relative to two existing open-source aligners with simpler architecture (Prosodylab-Aligner and FAVE). Above mentioned methods are mostly evaluated by clean, laboratory environment data. We did not find a study which exploits above mentioned methods over pathological speech.

In the present study, segmentation of CC combination was the major challenge particularly for fricative-fricative and stop plosives-fricative segmentation. We use unsupervised STM method for segmentation of CC syllable. We also compare the accuracy of STM and MFA in this study.

Burst definition

During the production of stops, acoustic pressure is built up behind a closure at a place within the vocal tract, resulting in a silent interval or a low level acoustic signal, with or without voicing. When the pressure is released suddenly, it introduces a relatively high energy burst or transient in the acoustic signal, spanning a short interval. The instant in the acoustic signal corresponding to the sudden release is called the “burst-onset” or the closure-burst boundary or the Closure Burst Transition (CBT). The burst onset could be a transient speech landmark [151], lasting for only a few milliseconds (3-5 ms). CBT detection is an essential cue for Voice Onset Time (VOT) which also remains a crucial acoustic measure for detecting the manner of stop plosives. Invariant characteristics of burst in stop plosives were explained in the study [290].

Burst onset detection

The speech landmark detection approach is used for CBT detection in previous studies. In literature, several learning (modeling) based methods are found to detect stop plosive landmarks by CBT detection. The study [180] tracked several sub-band energy trajectories and registered their relative fluctuations at specific locations to detect burst onsets. Degree of abruptness, i.e., energy difference between two appropriately located frames, is used as an acoustic measure in [30]. The study [208, 207] detected stop consonants by tracking total energy, high-band energy, and spectral flatness around closure-burst transitions. The study [119] proposed intensity discrimination applied to bark-scale frequency bands. They combine the separate frequency-band information into a single measurement using Baye’s rule. Then, they use a threshold to select a number of candidate frames for further spectral-domain artificial neural network classification. The study [191] also adopted the latter method for burst features. A technique using Recurrent Neural Networks (RNNs) to detect burst onsets with standard frame-based spectral features was proposed in [157]. A set of spectral and temporal features like energy ratios and zero crossings in neural network classifiers were proposed for burst onset detection in [137]. The study [131] used a Gaussian mixture model (GMM) of smoothed log magnitude spectrum (256 coefficients) and the rate of change of the components of the GMM to detect stop consonants. The study [176] has used a two-dimensional cepstrum as the feature vector (56 dimensional) and a random forest (RF) classifier for detecting burst-onset landmarks. In another similar study [287], a set of temporal and spectral features such as spectral energy, low frequency band energy, Wiener entropy, Zero Crossing Rate (ZCR) was used for Support Vector Machine (SVM) classifier to detect burst onset and vowel onset for Voice Onset Time (VOT). The latter method yielded marginally better VOT accuracy than the method proposed in the study [176]. The study [256] also exploited spectral features in RBF kernel and used a max-margin classifier trained by Stochastic Gradient Descent. Another recent study computed positive and negative voice onset time (VOT) by bidirectional Recurrent Neural Network [271]. The latter method is suitable for only word-initial VOT measurement. Above mentioned

methods are mainly based on the learning process, which requires labeled training data.

On the other hand, threshold based methods are also found where multi-dimensional acoustic features being used for CBT detection. In this category, acoustic features were computed either in the time domain or in spectral domain. In time domain approach, the study [102] used a non-linear energy tracking algorithm, Teager Energy Operator (TEO) for closure-burst-transition (CBT) detection of unvoiced stop plosives [102]. This approach uses the amplitude modulation component (AMC), derived from the TEO, to detect the initial burst and vowel onset in single words. The latter method requires additional knowledge about unvoiced stop plosives like certain band energy. In the study [12], Plosion Index (PI) was proposed to detect burst onset [12] for both voiced and unvoiced stop plosives. PI was defined as the peak amplitude ratio (in Hilbert envelop) in the CBT to the average of absolute values over an appropriate interval excluding the immediate neighborhood. The latter method explicitly described the CBT detection method by providing a threshold. The PI method provided better CBT detection compared to the studies [180, 208, 176]. According to the definition of PI, it may fail to detect multiple CBTs while preceding samples contain high amplitude. The author also mentioned that this method might fail in bilabial fricative /f/, glottal fricative /h/ because of transient-like properties. The PI method also adopted in the study [166] for CBT detection.

In spectral domain method, a comparison of spectral power of time-frequency reassessment was used to detect CBT followed by VOT measure [291]. Studies [209, 212] proposed a rule-based filtering method on spectrogram to enhance transient-like properties. The envelope of the filtered spectrogram by summing all values in each time window enhances the noisy burst. The difference of the envelope highlights and specifies the stop release position. Another approach based on Single Frequency Filter (SFF) followed by phase reconstructed signal was developed to detect the burst [204]. The phase reconstruction method highlights the burst in stop plosives; however, it also highlights noise components. Finally, the euclidean distance over magnitude envelope and Wiener entropy were used for CBT detection by setting a threshold.

In previous studies, very few studies were found related to automatic CBT detection or VOT measure for pathological speech. The study [209] evaluated the developed method for PD patients. Otherwise, most of the previous studies [75, 73, 306, 16] computed VOT from manual labeling to analyze neurological disorders.

In the present study, we thus consider to analyze the burst properties of stop plosives from pathological speech. Existing automatic burst detection methods are implemented for comparison. The present study also developed a methodology for automatic segmentation of first phoneme by detecting first vowel onset. To the best of our knowledge, analysis of burst in detail was not conducted for neurological disorder.

6.4.2 Methodology

Database

This study analyzes word initial unvoiced obstruents taken from respective logatomes. Total 3 unvoiced stop plosives (/p/ from "pataka", /t/ from "touca", and /k/ from "quina/crancto") and 3 unvoiced fricatives (/f/ from "feju", /s/ from "sochin", and /ʃ/ from "chastu") are considered.

Visual method

Imprecise articulation is evaluated first by visual investigation. It is accomplished by wide-band spectrogram. The visual method mainly consider 3 types of disorder in unvoiced obstruents.

- **(Muti)Burst in fricative**

Fricatives are produced by partial obstruction of air with different place of articulators. Overshooting of articulators results in a burst-like realization in fricatives. Detection of burst in fricative may help to elucidate underlying pathology.

- **Weak Burst in stop plosives:** It targets to identify bursts in fricatives, the presence of multiple bursts in stop plosives, weak bursts in stop plosives. Burst/multiple bursts are visually identified as uniform energy distribution in all the frequency bands. In contrast, a weak burst is characterized by relatively lower energy in burst compared to the following vowel.
- **Muti-Burst in plosive:** In general, a single burst is expected for word-initial unvoiced stop plosives. The presence of multiple burst can be regarded as disorder for different underlying pathology or linguistic properties.

Objective method

Objective method starts with first phoneme segmentation which is accomplished by first vowel onset detection followed by segmentation of consonant-consonant combination. Next, burst detection methods are exploited to find best method.

- First vowel onset:

1. **Maximum normalized cross-correlation (MNCC):** First vowel onset is detected by MNCC [295, 64, 12]. Cross correlation method is widely used in periodicity measure.

$$z[k] = \frac{\sum_{l=0}^{|x|-1} x_l y_{l-k+N-1}^*}{\sqrt{\sum x^2 * \sum y^2}} \quad k = 0, 1, \dots, |x| + |y| - 2 \quad (6.1)$$

where $|x|$ and $|y|$ is length of x and y respectively. N is defined as $\max(|x|, |y|)$. We use local mean subtraction from each reference window before cross correlation measure.

Maximum value from NCC is filtered by setting a threshold ($0.5 * MNCC$). Less than the threshold is set to zero for MNCC. Next, greater than the “0” is considered as vowel onset. If the vowel duration is less than 20 msec, it was not considered as vowel onset.

2. **Pitch (F0) method:** The pitch is estimated by REAPER method which out-performed other F0 detection methods.
3. **Montreal forced aligner (MFA):** MFA is a supervised method. It is a phoneme aligner. We have used the trained model (for French language) for vowel detection.
4. **Dr.VOT:** We also used Dr.VOT toolkit for vowel onset detection and compare accuracy.

- Phoneme boundary detection in CC combination:

1. **Spectral transition measure (STM):** STM is implemented [84] using log-filterbank features. This method was previously used in several studies using different acoustic features.
2. **Montreal phoneme alignment (MFA):** MFA [193] is a supervised method and open-source toolkit. We did not change any parameter or adapt the acoustic model with our data.

- Burst detection methods:

Several methods were developed in previous studies to detect burst onset in time domain or spectral domain. Two widely used methods, the Plosion index (PI) and Teager Energy Operator (TEO) were developed to detect burst onset. Intuitively, for a signal with a transient characterized by a significant change in local energy, the ratio of the peak amplitude in the transient to the average of absolute values over an appropriate interval excluding the immediate neighborhood of the peak amplitude may be expected to be high, and it was represented as Plosion Index (PI) [12]. To check the efficacy of PI method, it is implemented in the present study. The signal is high pass filtered with a cut-off frequency of 400 Hz to discard glottal influence. Hilbert transform is used to capture the envelope. PI is measured as mentioned in the study [12].

$$PI(n_0, m_1, m_2) = \frac{|s(n_0)|}{|s_{avg}(m_1, m_2)|}$$

where,

$$s_{avg}(m_1, m_2) = \frac{\sum_{i=n_0-(m_1+1)}^{i=n_0-(m_1+m_2)} |s(i)|}{m_2}$$

where, n_0 is current sample and m_1 and m_2 are previous samples.

The study [102] proposed TEO based burst onset detection. This method performs better than Hilbert Transform (HT), which is another possible method to decompose speech signals into AM-FM components. Amplitude modulated

signal is used for burst detection. The study [102] did not describe the criteria of onset explicitly.

In the spectral domain, we selected two different methods to implement for burst onset. The spectral magnitude with filtering (SMF) was developed in the study [209]. Diadochokinetic speech task was clustered by syllable. Each frequency bin magnitude of the frame was filtered by the conditional filter, which was developed in latter study. If energy distribution is present in every frequency bin, we can consider that frame as burst. Another method developed burst onset detection by instantaneous magnitude and phase, computed by single frequency filter (SFF) [204, 13]. In the latter method, phase only information is used to reconstruct the signal. Phase reconstructed signal highlighted the burst signal. SFF output is used to compute the envelope difference and Weiner entropy of each sample. We also implemented this method for comparison.

In this study, we compare burst detection accuracy by previously discussed four methods.

6.4.3 Results

Results of visual analysis

- **Burst/closure in fricatives:**

By visual inspection of the spectrograms, the presence of burst/multiple burst is assessed in unvoiced fricatives. Figure 6.11 shows an example of spectrograms of the consonant /f/ pronounced commonly by a HC, with a presence of burst by a PD and with a multiple bursts by an MSA-P patient.

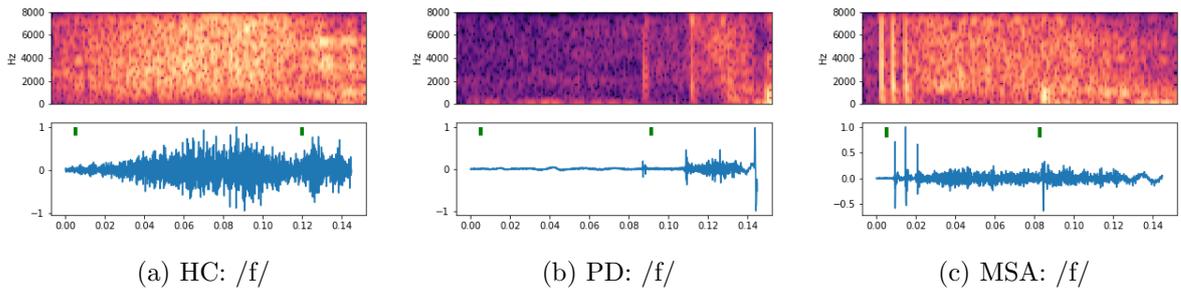


Figure 6.11: Example of no/burst/multiple burst in /f/ in a HC/PD/MSA-P

It is observed that 46.15% (6 out of 13) of MSA-P and 15.38% (4 out of 26) of PD presented burst in fricative /f/ from logatome “frambi”. On the other instance, in /f/ from logatome “feju”, 30.76% (4 out of 13) MSA-P and 3.84% (1 out of 26) PD patients displays burst/closure. In alveolar fricative /s/ from logatome “sochin”, burst is not observed in any of the group. Only 1 MSA-P patient showed burst in /s/ from logatome “spegzi”. In palatal fricative /ʃ/, 2 MSA-P patients showed burst. It may be hypothesized that a particular context

has a greater influence to produce bursts in fricatives. It indicates that the use of /f/ from “frambi” is more appropriate than “feju”. Moreover, manifestation of abnormal burst is most frequent in consonant-consonant cluster. Table 6.13 summarized presence of anti-spirantization in individual fricatives.

Obstruents	Consonant	Anti-spirantization (%)		
		HC	PD	MSA-P
Fricatives	/f/ (/fe/)	0	4	31
	/f/ (/fra/)	0	15	46
	/s/ (/so/)	0	0	8
	/S/ (/Cha/)	0	0	15

Table 6.13: Anti-spirantization (%) of individual unvoiced fricatives in HC, PD, and MSA-P groups

Considering all instances of burst in unvoiced fricatives, it is observed that 19.23% (5 out of 26) PD and 61.53% (8 out of 13) MSA patients exhibit bursts in fricatives (anti-spirantization). Thus, anti-spirantization would serve as a good speech marker for being MSA.

- **Weak burst in stop plosives:**

In /k/, it is observed only 1 PD and 1 MSA patients exhibit minor weak burst. A weak burst is frequently observed in dental stop /t/ for PD and MSA. Figure 6.12 presents example of normal burst energy and weak burst in /t/ from logatome “touca”.

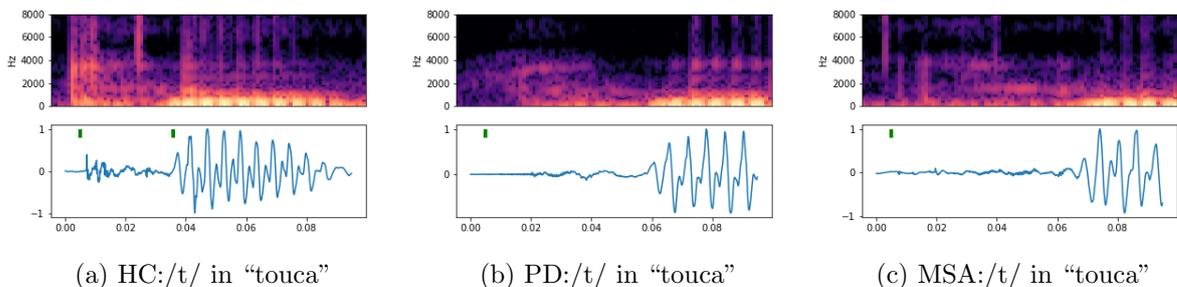


Figure 6.12: Weak burst energy in /t/ for HC, PD and MSA-P subjects

In total, 15% (3 HC) , 35% (9 PD) and 46% (6 MSA) patients produced weak burst in /t/. Thus, it can be hypothesized that both PD and MSA patients manifest a reduced range of movements, particularly in pronunciation of /t/ in ”touca”. As both PD and MSA groups manifest weak burst in /t/, it will not serve as a disease specific speech marker. However, weak burst can be used to measure the overall speech disorder.

- **Multiple burst in stop plosives:**

The subjective analysis found Multiple Burst (MB) mostly in stop plosived /k/

from logatome “quina”. On the other hand, presence of MB is rare in stop plosives /t/ and /p/. In /k/, we observed 6 HC (30%), 13 PD (50%), and 6 MSA (46.15%) exhibit MB. Figure 6.13 displays example of multiple burst in /k/ for PD and MSA patients. As HC, PD and MSA patients exhibit multiple bursts, this measure will not help in differential diagnosis. Notably, the presence of multiple burst in velar stop /k/ probably common.

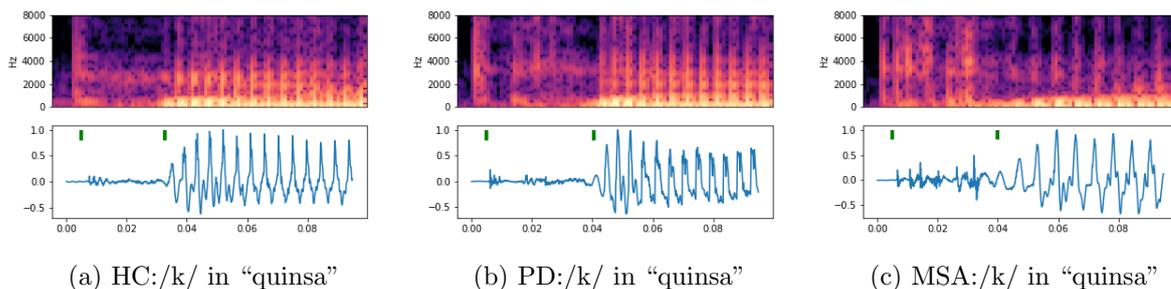


Figure 6.13: Multiple burst in /k/ for HC, PD and MSA subjects

As multiple burst did not exhibit differential properties between disease groups, we did not measure the MB objectively.

Results of objective analysis

Objective methods are first evaluated to detect vowel onset detection followed by segmentation of consonant-consonant boundary. Next, we detect the burst (multi) in unvoiced fricatives.

- **First vowel onset detection:**

10 logatomes (“chastu”, “crancto”, “feju”, “frambi”, “quina”, “spegzi”, “psegra”, “sochin”, “touca”, “tunia”), starting with unvoiced consonant used in the evaluation. In this study, the tolerance is set to 20 msec considering the occurrence of burst in the initial position of obstruents. In stop plosive and vowel cluster, MNCC method yields better vowel onset detection compared to the pitch (F0), MFA, and Dr.VOT methods. It is observed that segmentation accuracy is higher for HC compared to PD and MSA. The latter observation is logical considering signal is prone to be distorted for PD and MSA. In addition, the aging effect is another parameter that may distort signal sometimes. All these parameters may sometimes disturb vowel onset detection.

Figure 6.14 represents vowel onset detection method. MNCC yields better vowel onset compared to other methods. Vowel onset by pitch estimation is also implemented where REAPER method outperformed other pitch computation methods like YIN, RAPT, PRAAT (default). This is the reason to keep REAPER method in comparison.

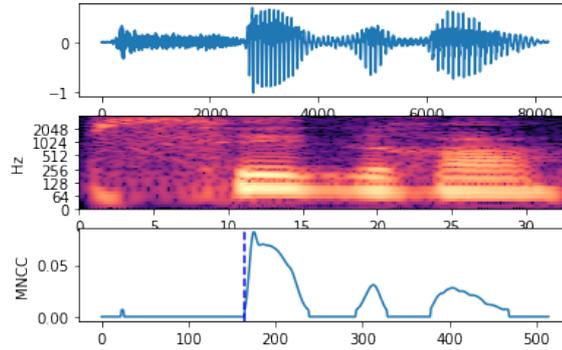


Figure 6.14: Example of vowel onset detection by MNCC method

	Time tolerance	HC	PD	MSA	Overall
MNCC	5ms	78	79.23	64.62	75.59
	10ms	94	94.62	76.92	90.51
	20ms	99	99.23	90.77	97.29
	30ms	99.00	100	93.85	98.31
REAPER (F0)	5ms	69	56.15	61.54	61.69
	10ms	86	85.38	84.62	85.42
	20ms	93	96.15	89.23	93.56
	30ms	97	98.46	90.77	96.27
MFA	5ms	42.50	26.40	16.92	28.89
	10ms	56.25	54.40	40.00	51.48
	20ms	83.75	78.40	72.31	78.52
	30ms	88.75	88.80	80.00	86.67
Dr.VOT	5ms	65.45	56.00	41.67	54.42
	10ms	89.09	79.00	71.67	79.53
	20ms	90.91	82.00	81.67	84.19
	30ms	90.91	83.00	81.67	84.65

Table 6.14: Vowel onset accuracy for HC, PD and MSA in logatomes starting with unvoiced stop plosives

Figure 6.15 represents accuracy of vowel onset detection methods for fricatives. MNCC also provides improved vowel onset detection compared to other methods.

Vowel onset detection accuracy is comparably better for MNCC method compared to other methods such as REAPER, MFA, and Dr.VOT. Hence, we decided to use MNCC method for rest of the experiment. After detecting first vowel onset, unvoiced obstruents from CV syllable are directly segmented, but need to segment CC combination to get word-initial obstruents.

- **Phoneme boundary between CC combination:**

	Time tolerance	HC	PD	MSA	Overall
MNCC	5ms	87	78.21	71.79	76.84
	10ms	95	94.23	88.46	91.81
	20ms	100	98.08	97.44	98.59
	30ms	100	98.72	98.72	99.15
Reaper(F0)	5ms	60	60.26	61.54	60.73
	10ms	89	91.03	82.05	88.14
	20ms	95	95.51	89.74	94.07
	30ms	96	96.15	92.31	95.20
MFA	5ms	31.25	32.67	26.92	29.63
	10ms	58.75	51.33	44.87	50.93
	20ms	87.50	89.33	82.05	87.35
	30ms	96.25	92.67	93.59	93.83
Dr.VOT	5ms	48.48	43.33	31.94	41.47
	10ms	89.39	84.17	76.39	83.33
	20ms	96.97	91.67	84.72	91.09
	30ms	96.97	91.67	86.11	91.47

Table 6.15: Vowel onset accuracy for HC, PD and MSA in logatomes starting with unvoiced fricatives

In this experiment, STM method has been adopted to segment consonant-consonant cluster [84]. In this general method, any kind of acoustic parameter can be used for phoneme discrimination. After exploring several acoustic parameters (MFCC, LFCC) finally, filterbank feature is selected in the present study. Figure 6.15 represents an example to discriminate /f/ and /R/.

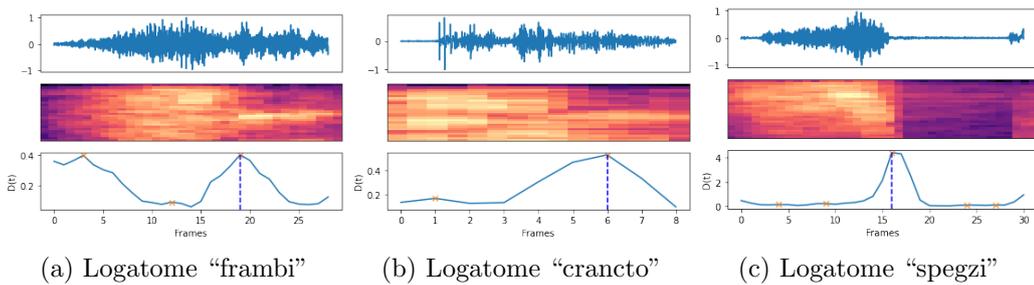


Figure 6.15: Example of boundary detection between consonant-consonant (CC) combination using log filterbank feature in STM method

As the occurrence of burst is observed mainly at the beginning of the consonants, we can permit 20 msec tolerance of the first phoneme ending. Table 6.16 shows that accuracy is not very high for 5ms and 10ms threshold, but accuracy is better for 20ms and 30ms thresholds. The STM method using log filter-bank feature yields first phoneme segmentation accuracy 83.33% by 20 msec tolerance in CC cluster, which is quite good considering signal complexity. Accuracy is

always deteriorated due to MSA groups. The proposed method yield marginally better accuracy compared to MFA method.

	Time tolerance	HC	PD	MSA	Overall
STM	5ms	27.08	26.67	10.26	22.84
	10ms	56.25	62.67	35.90	54.32
	20ms	87.50	89.33	66.67	83.33
	30ms	95.83	96.00	74.36	90.74
MFA	5ms	25.00	32.00	28.20	29.01
	10ms	45.83	54.67	51.28	51.23
	20ms	87.50	82.67	71.79	81.48
	30ms	95.83	90.67	76.92	88.89

Table 6.16: Segmentation of CC for HC, PD and MSA in logatomes

Important to note, burst detection using automatic segmentation method also produce the same result compared to manual segmentation in fricatives.

- **Burst/closure in fricatives:**

Burst is detected in automatically segmented word-initial fricatives. Table 6.17 presents burst detection accuracy in /f/ from "frambi" by previously developed burst detection methods. Previous methods either detected false bursts or could not detect true bursts. However, PI method came up as a good method for burst detection. In visual investigation, 10 subjects (4PD, 6 MSA) showed burst in fricative /f/ from logatome "frambi".

	HC		PD		MSA		Accuracy	
	True	False	True	False	True	False	Precision (%)	Recall(%)
Plosion index (PI)	0	0	4	1	6	1	83.33	100
TEO	0	1	3	2	6	0	75	100
Spectral magnitude	0	0	0	0	2	0	100	8
SFF	0	0	4	2	6	1	76.9	100

Table 6.17: Burst detection result using manual segmentation for logatome "frambi" by different methods

Similarly, Table 6.18 illustrates that PI method exactly match the number of burst in /f/ from "feju" by visual method i.e, 4 subjects (1 PD, 3 MSA). On the other hand, other methods either yield false burst detection or true burst rejection.

Above burst detection analysis showed that PI method yields comparably better accuracy than other methods. Hence, we decide to use PI method for burst detection in fricative as well as for stop plosives. The automatic methodology also able to detect burst in fricatives which showed that MSA patients manifest frequent burst in fricatives than PD.

	HC		PD		MSA		Accuracy	
	True	False	True	False	True	False	Precision (%)	Recall(%)
Plosion index (PI)	0	0	1	0	4	0	100	100
TEO	0	2	1	2	4	1	50	100
Spectral magnitude	0	0	1	0	1	0	100	40
SFF	0	0	1	3	3	2	44.44	100

Table 6.18: Burst detection result using manual segmentation for logatome “feju” by different methods

6.4.4 Voice onset time (VOT)

Voice onset time (VOT) is a frequent measure for analyzing stop plosives analysis towards the speech disorder. Hence, we consider to use it as another additional speech feature for unvoiced stop plosives /p/, /t/, and /k/.

Methodology

We measure acoustic features first by manual segmentation of logatomes (/p/ from ”pataka”, /t/ from ”touca”, and /k/ from ”quina”).

Next, automatic method is developed to measure VOT from stop plosives. In this analysis, Dr.VOT tool [271] and combination of Plosion Index and MNCC are employed to automatically measure the VOT for voiceless stop plosives. For automatic VOT measure, it is required to detect the initial burst and following vowel onset correctly. Detected onset of burst and vowel by different methods (Dr.VOT and PI+MNCC) for stop plosives /p/, /t/, and /k/ are evaluated by manual segmentation.

Two temporal measures are considered in this study as follows:

- **Voice onset time (VOT):**

It is measured as the duration of first burst instance to the following vowel onset.

- **Voice onset time ratio (VOTR):**

In conventional method, VOTR is measured as VOT of stop divided by the syllable duration.

According to presumption, as both VOT and vowel duration increased for pathological voice, VOT Ratio may not reflect actual impairment. Hence, in the present study, we propose a modified formula to calculate the VOT Ratio as follows:

$$ModVOTRatio = \frac{VOT_{stop}}{(Max_{syl} - Dur_{syl} + \epsilon)} \quad (6.2)$$

where $\epsilon = \mu_{syl}$, Max_{syl} is the possible maximum syllable duration for respective stop vowel syllable, and Dur_{syl} is the syllable duration. This relative VOT ratio would be more sensitive towards changes in VOT and vowel duration.

Result

We start with analyzing speech features from manual segmentation. Next, same speech features will be automatically computed by segmentation provided by two methods.

- **Analysis by manual segmentation:**

Table 6.19 presents group-wise VOT of unvoiced stop plosives and group differences. VOT using manual segmentation of /p/ is 14 ms, /t/ is 28 ms, and for /k/ is 32 ms for HC group. The latter observation mostly matches with previous studies. In agreement with the previous study [306], MSA-P patients also show prolonged VOT for unvoiced stops. It is also important to note that PSP patients also showed prolonged VOT for unvoiced stops. Thus, VOT of unvoiced stops would discriminate PD and APS disease groups, but not within APS (PSP and MSA). In contrast, study [264] presented the inconsistent result of VOT across the different places of articulation and disease groups. The latter study found prolonged VOT in velar stops only for PD, retroflex (alveolar) for PD, PSP, MSA, and bilabial for PSP and MSA. Thus it demands to assess imprecise consonants in differential scenarios.

	Plosives	Mean/SD of VOT (msec)			P-Value		
		HC	PD	MSA	HC_PD	HC_MSA	PD_MSA
Manual	/p/	14.86/5.26	15.02/5.17	21.43/11.87	0.87	0.035	0.069
	/t/	28.84/8.42	31.33/10.42	46.21/17.13	0.38	0.0072	0.017
	/k/	32.58/9.12	35.48/11.32	44.12/21.65	0.34	0.055	0.29
	/ptk/	25.43/5.67	27.28/6.89	37.25/13.75	0.21	0.0025	0.012

Table 6.19: Analysis of VOT of /p/, /t/, and /k/ from manual segmentation

Vowel duration, conventional VOT ratio (VOTR), and modified VOT ratio are computed by manual segmentation from stop plosives /p/, /t/, and /k/ and following vowels /a/, /u/ and /iŋ/ respectively. Vowel duration of /a/ in syllable /pa/ get prolonged for MSA-P patients compared to PD and HC. In Conventional VOTR measure, only stop /t/ provided group difference between PD and MSA-P. Interestingly, modified VOTR yields improved group differences between PD and MSA-P in three unvoiced stop plosives.

The average conventional VOTR of three unvoiced stops did not provide discrimination among groups. Average modified VOTR of /p/, /t/, and /k/ yields further improvements in group differences between PD and MSA-P. Figure 6.16 presents boxplot of average VOTR for HC, PD and MSA-P. MSA-P patients manifest increased VOTR compared to PD and HC.

- **Analysis by automatic segmentation:**

Starting and ending of stop plosives are detected by two methods: Dr.VOT and PI+MNCC. Table 6.21 presents boundary detection accuracy of phoneme /p/, /t/, /k/. Plosion index (used for burst starting) and MNCC (used for ending of stop plosive) yields the highest accuracy.

	Units	HC vs PD	HC vs MSA	PD vs MSA
Vowel Duration	a(p)	0.10	0.0008	0.023
	u(t)	0.34	0.42	0.86
	cinq(k)	0.18	0.21	0.45
VOT Ratio	/p/	0.58	0.88	0.86
	/t/	0.84	0.055	0.029
	/k/	0.82	0.63	0.69
Mod VOT Ratio	/p/	0.74	0.007	0.02
	/t/	0.18	0.0025	0.023
	/k/	0.21	0.015	0.073

Table 6.20: Vowel duration, VOTR, and modified VOTR of unvoiced stop plosives

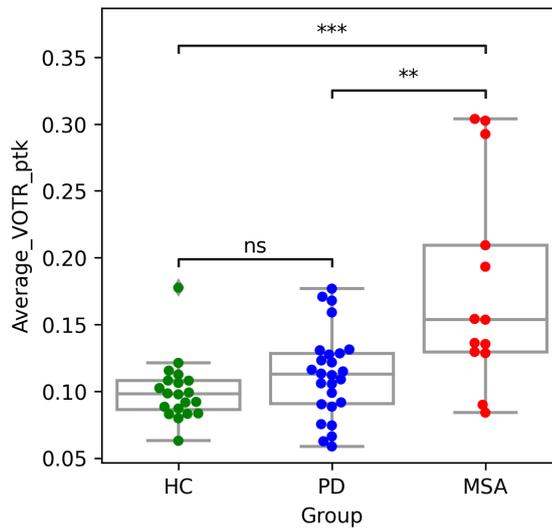


Figure 6.16: Average modified VOTR of unvoiced stop plosive (/p/, /t/, and /k/)

	Tolerance	Starting				Ending			
		HC	PD	MSA	Overall	HC	PD	MSA	Overall
Dr.VOT	5ms	85.00	60.26	64.10	69.49	58.33	46.15	41.03	49.15
	10ms	86.67	62.82	71.79	72.88	76.67	67.95	69.23	71.19
	20ms	86.67	71.79	76.92	77.97	83.33	70.51	74.36	75.71
	30ms	86.67	74.36	76.92	79.10	83.33	70.51	74.36	75.71
PI+MNCC	5ms	98.33	96.15	92.31	96.05	73.33	70.51	64.10	70.06
	10ms	98.33	96.15	97.44	97.18	93.33	93.59	79.49	90.40
	20ms	98.33	98.72	97.44	98.31	98.33	98.72	84.62	95.48
	30ms	98.33	98.72	97.44	98.31	98.33	100.00	87.18	96.61

Table 6.21: Accuracy of burst onset and vowel onset by Dr.VOT and PI+MNCC methods

Dr.VOT method was originally evaluated on PG-WORDS dataset and claimed accuracy 98% (using 10 msec threshold) in previous study [271]. However,

Dr.VOT yields average accuracy of 72.88% for the onset of burst and 78.79% for following vowel onset for the present database, Voice4PDMSA. Reduced accuracy can be explained by distorted speech for aging voice and parkinsonian disorder. Notably, burst onset detection using Dr. VOT was comparably better for velar stop /k/, whereas it mostly failed for /t/ and /p/, probably due to weak burst, particularly for patients. In comparison, Plosion index and MNCC method provide better accuracy than Dr.VOT and previous methods in [102, 209].

VOT analysis using the automatic segmentation is provided in the Table 6.22. We can not rely on the VOT analysis result of Dr.VOT because of erroneous burst detection and vowel onset. In comparison, the PI+MNCC yields a similar result as the manual method. Thus the latter method would be a suitable automatic method for positive VOT measure.

	Plosives	Mean/SD of VOT (msec)			P-Value		
		HC	PD	MSA	HC_PD	HC_MSA	PD_MSA
Dr. VOT	/p/	39.60/51.02	45.08/51.09	54.77/60.28	0.19	0.34	0.50
	/t/	29.20/10.17	33.46/33.62	62.15/61.26	0.77	0.24	0.21
	/k/	32.40/8.60	39.58/24.75	47.15/21.48	0.42	0.015	0.15
	/ptk/	33.73/18.54	39.37/23.76	54.69/30.94	0.39	0.01	0.04
PLMNCC	/p/	14.64/26.03	12.87/7.53	23.86/14.30	0.18	0.002	0.01
	/t/	34.55/34.37	39.67/40.11	109.62/104.80	0.101	0.003	0.018
	/k/	31.02/9.35	36.66/11.91	45.64/23.16	0.05	0.029	0.31
	/ptk/	26.74/15.49	29.73/15.74	59.71/37.92	0.10	0.001	0.002

Table 6.22: Group difference of Average VOT of /p/, /t/, and /k/

To analyze overall VOT, we averaged VOT of /p/, /t/, and /k/. Table 6.22 presents significant prolongation of VOT for MSA patients compared to PD. To visualize the discrimination between PD and MSA, we provided the result as boxplot in the Figure 6.17.

6.4.5 Discussion on unvoiced obstruents

This work constitutes the first study that attempts to highlight distinctive cues in the distortion of French unvoiced obstruents. The present study observed the frequent bursts in fricatives for MSA-P patients compared to PD and HC. In particular, the presence of burst was most frequent in bilabial fricative /f/, indicating comparably higher lip movement deficits. The presence of burst can be attributed to overshooting (dysmetria) of articulators. Dysmetria is a common symptom of ataxia due to poor control, timing, and coordination. Lesions in the lateral and paravermal cerebellar hemispheres are associated with intention tremor and incoordination (errors in timing, direction, and extent of voluntary movements). Incoordination is reflected in dysmetria [69]. Notably, an automatic algorithm was developed to detect burst in fricatives which match the visual finding. In a single study [15], the presence of burst

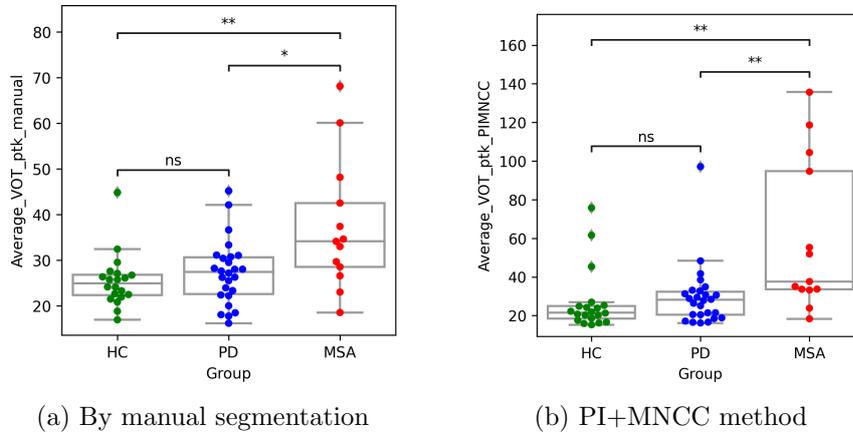


Figure 6.17: Average VOT of unvoiced stop plosive (/p/, /t/, and /k/)

was conducted by visual investigation in alveolar fricative /s/ and rarely observed anti-spirantization in disease groups. In agreement with the latter study, the present study also did not find burst in fricative /s/ for PD and MSA-P.

Presence of multiple bursts is mainly attributed to stuttering or dysarthria [145]. Occasionally, healthy speakers also produce multiple bursts in stop plosives. The presence of multiple bursts in stop plosives was studied for the PD patients [222]. The latter study found that PD patients show significantly higher MB in alveolar stop compared to HC. Surprisingly, HC group also showed frequent MB in velar stop /k/. In agreement with previous study, the present study also observed MB in velar stop /k/ for HC, PD and MSA-P groups. In addition, MB was not observed for bilabial stop /p/ and alveolar stop /t/. Thus it may be hypothesized that multiple burst is not a suitable marker for speech disorder. As a consequence, MB measure could be ignored for dysarthria analysis.

The presence of burst in word initial unvoiced stops is natural [152]. The study [222] observed weak burst (absence of burst) in PD patients. It was justified by inadequate articulatory closure [312, 182] most probably due to reduced range of movement (hypokinesia). Another study [8] also observed reduced intensity at the place of burst onset, which was also attributed to inadequate closure of articulators. In agreement with previous studies, weak burst is observed for word-initial alveolar stop /t/ for PD and MSA-P. This result reflects the observation about the inadequate tongue elevation and subsequent insufficient tongue constriction at alveolar position [182]. In comparison, weak burst is rarely observed in /p/ and /k/.

VOT is frequently used speech parameter to assess the incoordination of laryngeal and supra-laryngeal (jaw, lips, tongue, etc.) to produce stop plosives. Previous studies presented contradictory findings about VOT prolongation. As example, prolonged VOT was reported in studies [306, 80, 254]; no change in VOT [73, 239] and some times decreased VOT [76]. Prolongation of positive VOT is attributed to difficulty of initiating articulation due to a reduced ability to maintain the speech motor program, which is characteristic for speakers with hypokinetic dysarthria [293]. In agreement with previous studies [306], prolongation of VOT is observed for MSA-P patients

compared to PD and HC.

Different speech protocols may provide a different result. For example, in reading text or monologue, one can not ignore co-articulation effect in targeted phoneme. Consequently, word initial targeted unit from single-word material may be more suitable for investigating this kind of speech impairment.

Chapter 7

Speech disorder in diadochokinetic speech production

7.1 Introduction

Dysarthria is a common manifestation of the parkinsonian disorder. Degeneration in basal ganglia may affect temporal-spatial aspects of the motor speech and speech rhythm [293, 40, 94]. In contrast, degeneration in cerebellar can affect maintaining the precision of timing interval [265, 288]. Thus, we can hypothesize that any rhythmic activities require close interaction between the basal ganglia and cerebellar control circuits. As sustained vowel was considered for assessing vocal folds vibration, syllable repetition task would serve to check articulatory movements. In addition, Complex speech tasks can reveal a wide range of speech disorders compared to simply sustained phonation. Therefore, the Diadochkinetic (DDK) task was designed to primarily assess deficits in articulatory movements and coordination of respiratory, phonatory, and articulatory subsystems for pathological speech. Furthermore, the diadochokinetic task would be suitable for measuring imprecise consonants, syllable rate, irregularity in syllable repetition [74].

For DDK tasks, /pa-ta-ka/ [229, 185], /ba-da-ga/, and /pa-pa-pa/ syllable repetition [53] tasks were designed, which was recorded in normal and rapid style. The study [53] observed significant impairment for PD patients compared to HC in the perceptual and visual method. Notably, PD patients had comparably high disease duration. Another study [274] defined several acoustic measures to assess DDK speech for PD patients by oscillograph. Irregularity-related acoustic measures predominated in PD patients compared to HC.

Automatic evaluation of /pa-ta-ka/ was presented in the study [209] by acoustic measures related to articulatory deficits in vowel quality, coordination of laryngeal and supralaryngeal activity, precision of consonant articulation, tongue movement, occlusion weakening, and speech timing were analyzed. PD patients manifest prolongation of Voice Onset Time (VOT) and VOT ratio (VOTR), DDK fluctuation, and reduced DDK rate compared to PD. However, the segmentation accuracy of consonants and vowels must be improved for robust acoustic feature computation. Study

[252] proposed an accurate syllable detection method followed by measuring acoustic features like rhythm instability (RI) and rhythm acceleration (RA) adopted from [274]. Dysrhythmia was pronounced in HD, MSA, and PSP than HC. In the differential aspect, RI yielded a significant difference between PD and HD. Notably, only PD patients manifested RA compared to HC and HD. The latter method also adopted methods from [209] to detect the onset of bursts and vowels. In the thesis [112], four acoustic measures such as DDK Instability (DDKI), DDK Rate (DDKR), Vowel duration (VD), Standard deviation of power (stdPWR) was automatically computed from /pa-ta-ka/. Though the accuracy of burst detection and vowel detection was not provided in the latter study. Notably, the latter method used Gaussian Mixture Model to detect the speech clusters (voiced, unvoiced, pause). A recent study also proposed a Convolutional Neural Network (CNN) approach, which yielded better accuracy compared to the method [209]. However, the data-driven approach always requires a large amount of data plus sometimes annotation.

In the present study, DDK speech tasks like /pa-ta-ka/, and /pa-pa-pa/ with normal and rapid speed are considered to analyze the pattern of syllable repetition for PD and MSA-P patients. We adopted an automatic simple unsupervised method described in Section 6.4.2 for vowel onset and offset detection. Plosion Index (PI) method is also adopted from the study [12] because it can provide burst location by sample number rather than frame number (in other methods). Next, acoustic measures are computed from speech segmentation.

7.2 Methodology

7.2.1 Database

Total 60 participants recorded syllable repetition task. Each participants repeated syllable repetition task twice in normal and rapid mode. Details of the participants are described in Section 4.2.5. In perceptual evaluation, 3 HC participant could not perform the /pa-ta-ka/ task correctly. Therefore, those 3 HC subjects are discarded from the analysis. Finally, 17 HC, 27 PD, and 13 MSAP-P subjects are analyzed.

Some of the recorded samples from /pa-ta-ka/ task are manually annotated by Praat toolkit using standard protocol [73] to use it as reference. An example of manual labelling is presented in the Figure 7.1 In the annotation, stops (/p/,/t/, and /k/), pre-closure segment of stops, and vowel (/a/) are marked with starting and ending time stamp.

7.2.2 Automatic speech segmentation

Automatic segmentation of vowel, and consonants are the essential for analyzing syllable repetition tasks. At first, vowel onset and offset are detected using the MNCC method described in the Section 6.4.2. MNCC method outperformed fundamental frequency method (F0) by REAPER for first vowel detection. In unsupervised (rule based) method, setting a threshold always a challenge. Here, a first level threshold is

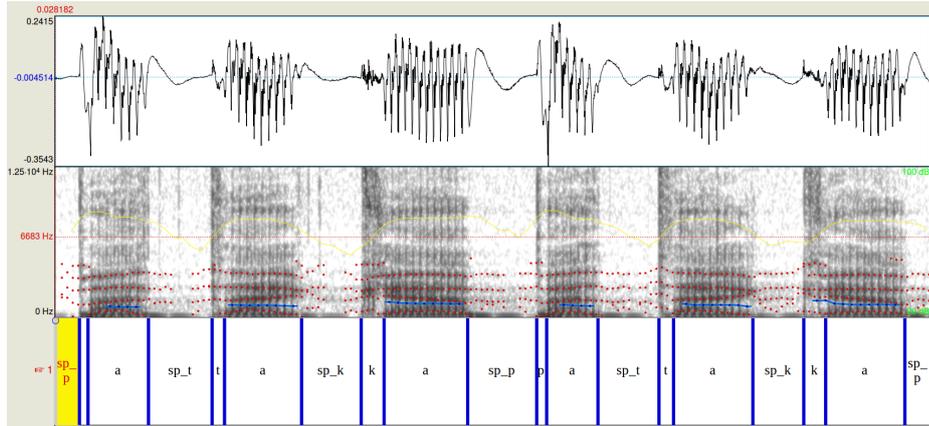


Figure 7.1: Example of manual annotation of /pa-ta-ka/ (part)

computed by following formula:

$$T = 0.5 * \frac{1}{N} \sum_{i=0}^{N-1} MNCC[i]$$

where N is the number of frames and $MNCC[i]$ is the maximum normalized cross-correlation value of i^{th} frame. Next, $MNCC$ values are converted to binary value by the following conditional equation:

$$VUV[i] = \begin{cases} 1 & MNCC[i] \geq T \\ 0 & MNCC[i] < T \end{cases} \quad (7.1)$$

Important to note, we have used same methodology in word-initial unvoiced obstruent segmentation task in Section 6.4.2. In previous section, only the first vowel onset is measured. In the present scenario, we need to detect the vowel onset and offset of all instances. To find the change, simple differentiation of VUV is enough as follows:

$$DT = VUV[i + 1] - VUV[i]$$

where $i = 0 \dots N - 1$. While DT value is countered as "1", it is marked as vowel onset and "-1" is labelled as vowel offset. If pause duration within vowel is less than 30 ms, we consider it as vowel.

Burst detection is carried out on the basis of vowel onset. Bursts are detected in the 50 ms preceding and 10 ms succeeding interval of each voice onset. This interval would help to detect burst of stops, particularly for /p/ because of short VOT. To detect the burst, Plosion Index (PI) method is adopted.

7.2.3 Acoustic features

Irregularity in DDK (DDKI)

Articulatory breakdown and involuntary movements of speech apparatus may disturb to maintain rhythm and timing of syllable repetition. To compute the irregularity,

duration of inter-vowel onsets is measured. DDKI is defined as the standard deviation of duration of inter-vowel onsets.

Rate of DDK (DDKR)

Rate of syllable repetition is frequently used by the speech pathologist. Slow syllable rate refers to slow movements of articulators, most probably due to spasticity, rigidity in articulator's muscle. The DDKR is computed as the inversion of the median duration between consecutive voice onsets. Median was preferred in order to increase robustness against wrong detection [112].

Acceleration of DDK (DDKA)

Deficits in Basal ganglia control circuit result in disrupted rhythm and timing. Sudden acceleration of speech is observed in PD patients. Acceleration of speech production can be attributed to either reduced range of movements articulator or disrupted timing sequence. To compute DDKA, the speech task is splitted into two halftimes by 25% overlap. The DDKA is defined as the ratio of DDKR in two halftimes.

Vowel duration (VD)

Degeneration in cerebellar control circuit result in delay of refining neuromuscular movements by basal ganglia control circuit [69]. As consequence vowel duration is prolonged for pathological speech. Vowel duration is measured by the median duration of vowel segments.

Standard deviation of power (stdPWR)

Variability in syllable to syllable loudness can be attributed to poor respiratory-phonatory coordination and control. Excessive variation in loudness is commonly observed in ataxic and hyperkinetic dysarthria [69]. stdPWR is computed as the standard deviation of maximum power of each vowel segments.

Voice onset time (VOT)

Prolongation of VOT for unvoiced stop plosives is commonly observed in parkinsonism. Prolongation of VOT is generally associated with synchronization problem of laryngeal and supra-laryngeal activities. VOT is defined as the duration between burst onset to vowel onset. Average VOT is computed from individual VOT of /p/, /t/, and /k/.

Weak burst (WB)

Reduced range of movement in hypokinetic dysarthria may result in imprecise burst (weak). It is not always perceptually detectable, but in spectral domain weak burst can be clearly visible. In /pa-ta-ka/, stop plosive is preceded by vowel. Total number

of missing burst by burst detection method (PI) is counted for each sample. WB is computed as the total number of missing burst in each sample.

7.3 Results

7.3.1 Phoneme segmentation accuracy

First, phoneme segmentation accuracy is discussed. Table 7.1 presents vowel onset and offset accuracy by the MNCC and method proposed in [130]. An example of vowel detection by MNCC method is presented in Figure 7.2.

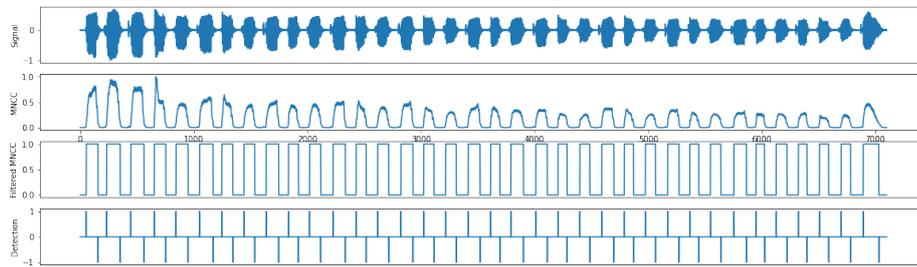


Figure 7.2: An example of automatic vowel detection by MNCC method in /pa-ta-ka/ task

Table 7.1 shows that MNCC method yields better vowel detection accuracy compared to method described in [130]. In 10 ms threshold, MNCC method produced 99.19% vowel onset accuracy. Comparably, vowel offset accuracy is less accurate most probably due to vowel occlusion in end part.

	Vowel onset accuracy (%)				Vowel offset accuracy (%)			
	5 ms	10 ms	20 ms	30 ms	5 ms	10 ms	20 ms	30 ms
MNCC	87.7	99.19	100	100	50.3	91.19	98.38	99.19
Method [130]	15.24	63.64	96.41	97.11	22	58.28	95.54	98.54

Table 7.1: Vowel onset and offset accuracy by thresholds

Therefore, segmentation using MNCC method is used for computing rhythmic features.

7.3.2 Acoustic analysis of DDK features

Given two types of speed mode, both DDK tasks are analyzed separately. We first start with normal speed /pa-ta-ka/ task. Acoustic features are computed from vowel segments and stop plosives segments. Table 7.2 presents group wise acoustic feature's statistics. MSA-P patients manifested prominent speech disorder in stdPWR, VD, DDKR, and WB compared to PD and HC. Surprisingly, PD patients also show group difference compared to HC in stdPWR probably due variable stress in syllables. Notably, PD patients also produced significant prolongation of average VOT than HC.

Present analysis did not observe significant acceleration in normal speed DDK task for PD patients.

	Groups			P-Value		
	HC	PD	MSA-P	HC_PD	HC_MSA	PD_MSA
	Mean/SD	Mean/SD	Mean/SD			
stdPWR (dB)	1.05/0.33	1.30/0.29	1.74/0.56	0.0032	0.0006	0.027
VD (ms)	78.36/13.27	78.94/13.32	96.94/18.72	0.68	0.0089	0.0021
DDKI (ms)	0.030/0.02	0.03/0.01	0.057/0.06	0.6212	0.082	0.075
DDKR (syll/sec)	5.21/0.67	5.33/0.71	4.41/0.70	0.50	0.0054	0.0005
DDKA	1.009/0.04	1.004/0.04	1.001/0.03	0.55	0.36	0.51
VOT	18.68/3.98	23.65/6.49	26.76/6.88	0.01	0.0014	0.16
WB	0.97/1.21	3.07/4.28	6.46/5	0.08	0.0002	0.01

Table 7.2: Group wise mean/standard deviation and group difference (p-value) of acoustic features from normal speed /pa-ta-ka/ task; blue and red colour represents predominant severity by PD and MSA-P respectively

Acoustic features from rapid speed /pa-ta-ka/ also produced encouraging differentiation between PD and MSA-P. Rapid style of DDK task may reflect functionality of timing and quick movements of articulator. MSA-P patients showed predominant disorder in all stdPWR, VD, DDKI, DDKR, and VOT acoustic features compared to HC and PD. In VOT and WB, PD patients also displayed significant disorder than HC probably due to hypokinesia.

	Groups			P-Value		
	HC	PD	MSA-P	HC_PD	HC_MSA	PD_MSA
	Mean/SD	Mean/SD	Mean/SD			
stdPWR	1.14/0.4	1.31/0.34	1.73/0.51	0.11	0.0028	0.013
VD	64.8/8	66.27/8.24	84.94/18.56	0.5	0.0028	0.0015
DDKI	0.024/ 0.01	0.028/0.02	0.04/0.02	0.52	0.031	0.039
DDKR	6.28/0.9	6.3/0.7	5.32/0.79	0.74	0.012	0.0016
DDKA	0.99/0.04	0.98/0.04	0.95/0.07	0.17	0.06	0.23
VOT (all)	15.75/4.14	19.3/4.83	23.09/4.96	0.0089	0.0006	0.027
WB	1.20/1.33	4.88/6.1	8.61/7.75	0.0097	0.0003	0.082

Table 7.3: Group wise mean/standard deviation and group difference (p-value) of acoustic features from rapid speed /pa-ta-ka/ task; blue and red colour represents predominant severity by PD and MSA-P respectively

Average vowel duration is longer for normal mode of speaking compared to rapid mode. It is confirmed that speakers compromised phoneme duration to speed up the syllable repetition. Likewise, VOT duration is longer for comfortable DDK task compared to rapid task.

After analyzing two different mode of /pa-ta-ka/ task, we may hypothesize that rapid mode is more appropriate compared to normal DDK to assess quick transition and timing of articulator.

7.3.3 Classification

For classification, Logistic regression classifier is employed. Given small amount of data, Leave-One-Subject-Out (LOSO) cross-validation method is adopted. First, individual acoustic feature from rapid DDK task is used for classification of PD and MSA-P. VD and stdPWR yield comparably good accuracy but not enough. We can proceed to feature combination to design ataxic dimension.

	Accuracy(%)	Specificity(%)	Sensitivity(%)
stdPWR	82.50	88.89	76.92
VD	80.00	88.89	61.54
DDKI	75.00	92.59	46.15
DDKR	67.50	92.59	23.08
VOT	57.50	85.18	7.69

Table 7.4: Classification accuracy by individual acoustic features from rapid /pa-ta-ka/

Simple average of stdPWR and VD produce an encouraging differentiation between PD and MSA-P. The combined feature can be regarded as ataxic dimension and defined as

$$A_{ddk} = \frac{stdPWR + VD}{2}$$

. Figure 7.3 shows that 9 MSA-P patients manifest predominant ataxia than PD and HC.

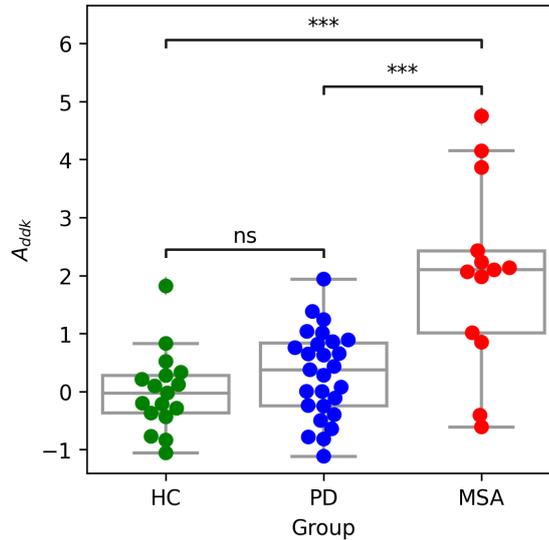


Figure 7.3: Group difference among HC, PD, and MSA-P using ataxic dimension (A_{ddk} from DDK features)

A_{ddk} feature also improved the classification accuracy to **87.5%**, specificity 96.29%, and sensitivity 76.92%.

7.4 Discussion

In the present study, we present a fully unsupervised automatic approach to investigate rhythm in diadochkinetic task for PD and MSA-P patients. A comparably accurate vowel detection method is presented, which outperform previously proposed method [130, 209], particularly for pathological speech. The study [274, 130] observed instability in duration for PD compared to HC, however present PD patients did not show variable duration than HC.

Contrasting result is found for two different mode of DDK task. In normal mode of DDK, PD patients showed group difference w.r.t. HC, which is not hold for rapid task. It can be explained by pattern of speech production by PD in normal mode, where stress label was not maintained for each syllable. Conversely, in rapid mode, PD subjects mostly gave equal stress in each syllable. Thus we may hypothesize that rapid mode DDK task are more suitable to extract speech impairment.

MSA-P patients manifest high variability in syllable to syllable power (stdPWR) and duration (DDKI), and prolongation of vowel (VD). Both acoustic features are commonly associated with ataxic dysarthria. In agreement with previous studies [247, 255], present study also observed that MSA-P patients also manifest ataxia. It was also observed in the study [130]. Important to note, other disease groups like PSP, HD, CA, and MS also manifest increased DDKI compared to HC. Thus only degree of severity can provide disparity among groups. Increased stdPWR variation can be explained by involuntary movements of respiratory muscles or deficits in coordination of phonation and respiration. Likewise, excess DDKI can be explained by impaired timing, planning, or involuntary movements of subsystems. On the other hand, prolonged VD is explained by slow movements and excessive vocal emphasis [130].

Prolonged voice onset time for unvoiced stops is frequent for both PD and MSA-P than HC. Deficits in laryngeal and supra-laryngeal coordination leads to increased VOT. MSA-P patients manifest comparably increased average VOT, probably due to combination and/or individual lesion in cerebellar and basal ganglia control circuit [5]. Notably, prolonged VOT is specific to MSA only, it was also frequently observed in PSP, HD, CA, and MS groups [130]. Thus quantitative disparity analysis may help to use it in differential diagnosis.

In addition, MSA-P patients also manifest significant spasticity by DDKR compared to PD and HC, which was also observed for PSP, HD, CA, and MS. Reduced DDKR is attributed to decreased articulator movements. In overall, present analysis observed significant group difference between PD and MSA-P in VD, DDKI, and DDKR, which was not true for previous studies [247, 255]. Combination of DDK features to represent ataxic index, also yield improved differentiation between PD and MSA-P. This approach was only attempted in the study [255].

Chapter 8

Speech disorder in reading text

8.1 Introduction: Spontaneous speech

Spontaneous speaking is the most complex model of speech production. It includes cognitive as well as execution of speech motor functions. In the cognitive aspect, thoughts, feelings, and emotions are first formed according to language for verbal communication. Next, the intended verbal message must be organized for neuromuscular execution. These activities include the selection, sequencing, and regulation of sensorimotor “programs” that activate speech muscles at appropriate coarticulated times, durations, and intensities [69]. Thus spontaneous speaking tasks may reveal deficits in the coordination of phonatory and respiratory function, precise articulator movements, and coordination of laryngeal and supra-laryngeal (lips, jaw, tongue etc.), which also reflect the prosodic and timing aspect of speech.

The previous study found continuous speech as the most sensitive speech task for assessing deficits in parkinsonian speech [250, 156]. In general, the prosody and timing of the speech were commonly analyzed by perceptual investigation. Prosody demonstrates suprasegmental speech factors like intonation, tone, stress, and rhythm. Dysprosody is commonly associated with hypokinetic dysarthria [58, 278, 7]. Several study observed predominant dysprosody by pitch variation, pause duration, loudness variation, total speech rate for PD patients [181, 105, 249, 33, 275, 34, 113]. Monopitch and monoloudness are common manifestation of PD and several studies found disparity between PD and HC [249, 279]. In differential diagnosis, Studies [247] did not find differences among PD, PSP, and MSA with monopitch and monoloudness acoustic dimensions. In perceptual analysis, the study [126] also did not observe disparity between HC, PD, MSA-P. In contrast, objective measures of monopitch yielded group differences between HC and parkinsonian disorder (PD and MSA-P) but not within PD and MSA. Pause-related features showed encouraging differences between PD and APS [281, 247, 126], but not within APS. Pause duration is prolonged, but the number of pauses decreases for the APS group (PSP, MSA). The gender-specific disparity was also observed in the number of pause and monopitch, particularly male patients displayed greater severity [126]. Speech rate is another parameter, which was frequently used to assess articulatory speed in reading text. PD and APS groups com-

monly displayed reduced overall speech rate [281, 126]. Studies related to early-stage differential diagnosis are very limited. The study [126] observed F0, prolonged pause time, and reduced speech rate might be used for early-stage differential diagnosis of PD and MSA.

Automatic analysis is a challenging task. Most of the previous studies mainly considered voice/unvoice/pause speech segmentation [281, 34, 249]. Automatic segmentation of voiced, unvoiced, pause, and respiration was developed to design several speech components for respiratory, phonatory, articulation, timing subsystem of speech [306, 130]. The latter study achieved pause detection accuracy to 86.2% and respiration detection accuracy to 81.6%. PSP patient showed predominant speech respiration and timing deficits compared to PD, and MSA [306, 112].

The present study considers analyzing reading text in the French language for PD and MSA-P patients. Therefore, this study first attempted to manually segment speech events (voiced, unvoiced, pause, respiration). Next, we adopted the methodology described in [306, 130] to compute several acoustic features to investigate the disparity between PD and MSA-P. In addition, another open-source software is also used to extract acoustic parameters.

8.2 Methodology

8.2.1 Database

The Section 4.2.2 describes the recording protocol for reading text. Total 60 subjects were recruited, which include 20 HC (10 male, 10 female), 27 PD (19 male, 8 female), 13 MSA-P (5 male and 8 female). Each speaker read the text single time in comfortable speed. Speech samples from headphone are considered for this analysis due to better quality than other two microphones.

8.2.2 Manual segmentation

Prosodic features computation requires to segment different speech part like voiced, unvoiced, pause, and optionally respiration. Automatic segmentation of unvoiced consonants, pause, and respiration always remain a difficult task, particularly noisy data or dysarthric speech. In this thesis, reading text is manually annotated by the author following standard procedure described in the study [73, 113]. An example of manual annotation is presented in the Figure 8.1.

Voiced segments are annotated by the presence of harmonicity. Unvoiced consonants are labelled by absence of fundamental frequency (F0), high turbulent noise. Pause is labelled by absence of spectral energy in all frequency bands. Respiration segment is labelled by the presence of energy in 500–2000 Hz frequency band, and longer than 100 ms.

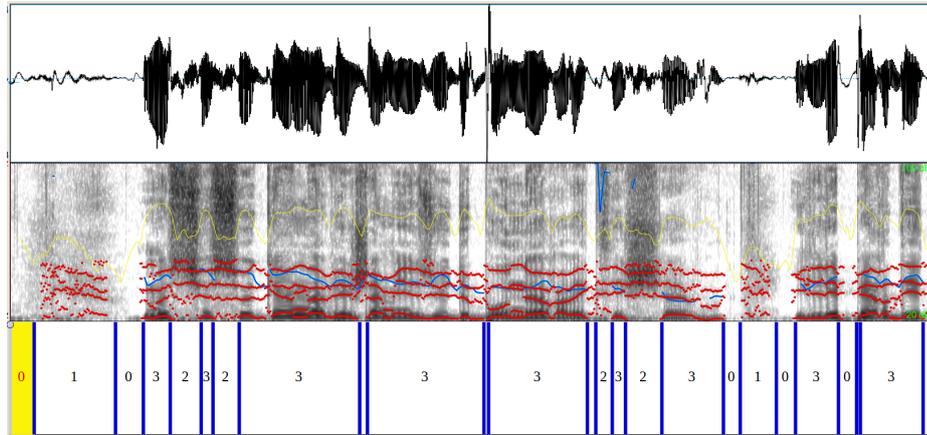


Figure 8.1: Example of manual annotation of reading text (partial); annotation labels are represented as 0: pause, 1: respiration, 2: unvoiced consonants, 3: voiced speech

8.2.3 Prosodic features

Using four segment, 13 acoustic components described in Chapter 3 are computed. Respiration group consists of Respiration Speech Rate (RSR), Pause Intervals per Respiration (PIR), Latency in Respiratory Exchange (LRE), and Relative Loudness of Respiration (RLR) acoustic components. Phonation group consists of Duration of Voiced Intervals (DVI), and Gaping in between Voiced Intervals (GVI) features. Articulation group includes Duration of Unvoiced Stops (DUS), Resonant Frequency Attenuation (RFA), and Decay of unvoiced fricatives (DUF). Timing group consists of four acoustic features such as Rate of Speech Timing (RST), Acceleration of Speech Timing (AST), Duration of pause intervals (DPI), and Entropy of Speech Timing (EST). Next, these acoustic features are used for differential diagnosis.

Another method is also used to extract prosodic features from reading text by Disvoice toolkit [62, 307]. The latter toolkit compute high dimensional acoustic features using different feature statistics (max, min, average, standard deviation, kurtosis, skewness). However, present study only considers average and standard deviation for clear explanation of pitch frequency (F0), loudness, voiced duration, unvoiced duration, and pause duration. In addition, duration ratios related acoustic features like PVU as $\text{Pause}/(\text{Voiced}+\text{Unvoiced})$; PU as $\text{Pause}/\text{Unvoiced}$; UVU as $\text{Unvoiced}/(\text{Voiced}+\text{Unvoiced})$; VVU as $\text{Voiced}/(\text{Voiced}+\text{Unvoiced})$; VP as $\text{Voiced}/\text{Pause}$; and UP as $\text{Unvoiced}/\text{Pause}$ are also considered for computing timing aspect.

8.3 Results

8.3.1 Prosodic features from manual segmentation

Statistical analysis of individual feature is presented in the Table 8.1. In respiration feature only number of pause within respiration is significantly reduced for PD and MSA-P patients than HC. Average rate of respiration also increased for both PD

and MSA-P but not statistically significant. In agreement with the study [112], PD and MSA-P display significant severity than HC in timing feature RST. In contrast to previous studies [306, 112], duration of pause did not provide group difference. Encouraging group difference between PD and MSA-P is observed in vowel duration and DUS. In addition, reduced gap between vowel interval is observed for PD and MSA-P than HC.

	Groups			P-Value		
	HC	PD	MSA-P	HC_PD	HC_MSA	PD_MSA
	Mean/SD	Mean/SD	Mean/SD			
RSR	18.88/6.11	21.68/6.09	21.43/5.72	0.19	0.33	0.86
PIR	6.55/1.71	4.94/1.03	4.50/1.24	0.0008	0.0009	0.30
LRE	164.83/85.78	175.23/60.01	210.7296/109.30	0.18	0.13	0.39
RLR	-52.37/8.50	-54.12/7.36	-51.20/11.09	0.84	0.65	0.29
RST	5.20/0.51	4.85/0.56	4.37/0.89	0.045	0.008	0.075
AST	0.89/1.04	0.51/0.69	1.09/1.64	0.16	0.43	0.87
EST	1.75/0.04	1.77/0.04	1.79/0.04	0.11	0.024	0.29
DPI	81.98/11.28	85.46/16.89	101.13/31.23	0.50	0.31	0.11
DVI	317.49/34.68	324.33/41.50	385.56/93.97	0.36	0.015	0.025
GVI	76.86/20.81	64.80/19.21	52.56/24.97	0.035	0.009	0.091
DUS	38/10.04	37.70/7.18	50.87/12.36	0.45	0.0002	0.001
DUF	-39.28/28.15	-32.52/18.11	-29.80/23.52	0.68	0.43	0.53
RFA	118.91/11.09	113.15/9.31	113.10/9.77	0.08	0.14	0.98

Table 8.1: Group difference of acoustic features computed from manual segmentation of reading text; blue and red colour represents predominant severity by PD and MSA-P respectively

All the above individual acoustic feature belong to hypokinetic dysarthria. By exhaustive search, combination of simple average of AST, EST, GVI, DVI, and DUS improved the discrimination of PD and MSA-P and defined as follows:

$$X1 = \frac{AST + EST + GVI + DVI + DUS}{5}$$

This combined feature is represented as "X1" and plotted in Figure 8.2. In this dimension, MSA-P patients manifest greater severity than PD and HC.

Using Logistic regression with LOSO cross-validation, dimension "X1" yields classification accuracy 87.5%, specificity 92.59%, and sensitivity 76.92%. It can be considered as good accuracy by only 1-dimensional feature.

8.3.2 Prosodic features from Disvoice tool

Group differences between groups are presented in the Table 8.2. Several acoustic features yield group differences between HC and PD, and HC and MSA-P. Monopitch and monoloudness features did not yield group difference between groups, however, a group of MSA-P patients showed monopitch. In contrast, the thesis [130] showed predominant monopitch and monoloudness for PD, PSP, and MSA patients. Notably,

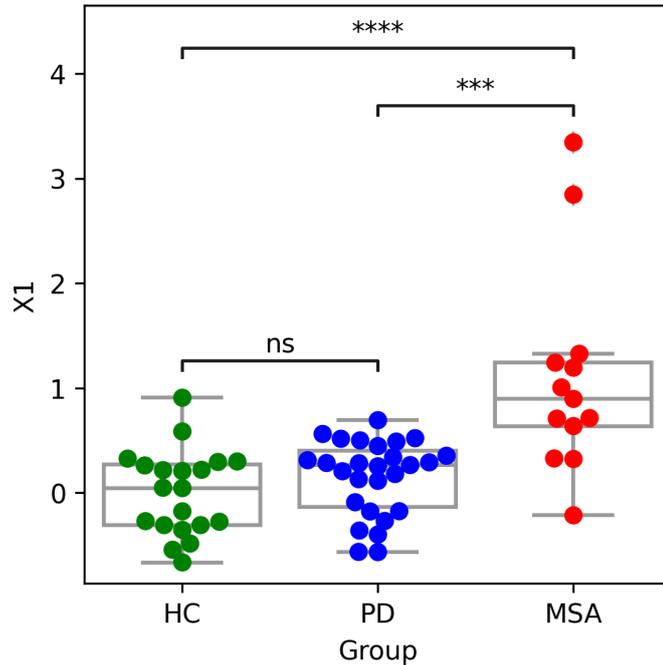


Figure 8.2: Group difference of HC, PD, and MSA-P by X1 designed by AST, EST, GVI, DVI, and DUS

both PD and MSA patients manifest prolongation of pause duration than HC. In duration ratio features, PD and MSA-P patients also yield disorder than HC. Again, those features did not provide differential characteristics between PD and MSA-P. As duration of voiced segment get prolonged, reduced voice rate is observed for PD and MSA-P. Interesting to note, a group of MSA-P patients manifest reduced "Vrate" compared to PD, which was also observed in the thesis [130].

8.4 Discussion

This chapter explored prosodic acoustic features for PD and MSA-P patients. In agreement with previous studies [114, 130], the present study also observed predominant dysprosody for PD and MSA-P patients than HC. Manual segmentation helps to measure the acoustic features accurately compared to the automatic segmentation method; however, it is time-consuming. In agreement with the study [130], PD and MSA patients did not show respiration deficits by RSR, LRE, and RLR; however number of pauses is reduced for both groups. It is explained by high hypokinesia and rigidity in the chest wall, reducing the vital capacity. As a consequence, the number of pauses is compromised, and the rate of respiration is increased. Notably, the study [114, 130] observed predominant respiration deficits in PSP patients.

Deficits in respiration also may affect the vocal folds functions. Reduced air out at the time of exhalation also force to compromise vocal folds abduction and adduction. It is reflected in GVI and DVI. Reduced control over vocal folds functions

Acoustic features	P-Value		
	HC_PD	HC_MSA	PD_MSA
F0avg	0.62	0.94	0.57
F0std	0.42	0.25	0.82
avgEvoiced	0.31	0.26	0.08
stdEvoiced	0.88	0.42	0.36
avgEunvoiced	0.47	0.76	0.49
stdEunvoiced	0.81	0.46	0.39
avgdurvoiced	0.055	0.029	0.23
stddurvoiced	0.11	0.007	0.10
avgdurunvoiced	0.75	0.60	0.70
stddurunvoiced	0.09	0.65	0.058
avgdurpause	0.007	0.018	0.39
stddurpause	0.003	0.022	0.69
PVU	0.102	0.27	0.94
PU	0.016	0.011	0.24
UVU	0.047	0.065	0.42
VVU	0.047	0.065	0.42
VP	0.19	0.65	0.61
UP	0.016	0.011	0.24
Vrate	0.02	0.003	0.071

Table 8.2: Group difference of prosodic acoustic components from reading text; blue and red colour represents predominant severity by PD and MSA-P respectively

may prolong vowel duration in the pause segment. In agreement with the study [130], thus MSA-P patients showed prolonged DVI and reduced number of pauses compared to PD and HC. Notably, it was observed in the study [130] that PSP, HD, and CA patients also manifest prolonged vowel duration. Accurate computation of gap between vowel yields reduced median gap duration in PD and MSA-P compared to HC, which is attributed to decreased ability of vocal folds to stop voicing by adduction [130]. Important to note, the study [130] only observed reduced GVI in MSA, PSP, and HD disease groups.

Unvoiced stop plosives production requires coordinated function between laryngeal and supra-laryngeal subsystems [32, 6]. Therefore, deficits in timing may result in a prolongation of voice onset time. The duration of unvoiced stops gets significantly increased for MSA-P patients, indicating dysfunction of the basal ganglia control circuit. Notably, increased DUS is not unique for only MSA; instead, it was also frequent in PSP, HD, and MS groups [130].

Previous studies [284, 276] showed that PD patients manifest predominant monopitch, but the present study did not observe group differences among HC, PD, and MSA-P in reading text, which in agreement with the study [247]. It may be due to different languages, different disease duration, etc. In addition, both PD and MSA-P exhibit impairment in average and standard deviation of pause duration and the ratio

of voiced, unvoiced, and pause. It is attributed to timing and coordination deficits in subsystems of speech production.

As the reading text and monologue are the most complex speech protocol, a precise analysis may help identify early-stage speech disorders in neurological diseases. Moreover, this speech task can also detect deficits in all the subsystems of speech production like respiration, phonation, articulation, and prosody. Now, it is the only challenge to accurately segment the speech clusters from spontaneous speech.

Chapter 9

Differential diagnosis between PSP and MSA

9.1 Introduction

As discussed in chapter 2, the parkinsonian syndrome is an umbrella term that refers to Parkinson's disease (PD), Atypical Parkinsonian Syndromes (APS) such as Progressive Supranuclear Palsy (PSP) and multiple system atrophy (MSA). APS differs from PD by more widespread neuronal involvement, resulting in additional clinical signs, more rapid disease progression, and poor response to dopamine replacement therapy [266]. The majority of PSP and MSA patients develop clinical features that overlap those of PD. Thus the correct diagnosis can be very challenging in the early stages of the disease. However, an accurate early diagnosis is essential not only in assessing prognosis and making decisions regarding treatment but also for understanding the underlying pathophysiology and for the development of new therapies [316].

Deficits in vocal functions result in speech impairment which is frequently an early and prominent clinical feature of PD and APS. Thus, during the last decades, there has been an increasing interest in PD speech and voice analysis [38]. However, very few attempts have been done in differential diagnosis between PD and APS, or within APS [160, 158, 257, 281, 264, 126, 255, 164]. Perceptual evaluation of speech disorder remains a gold standard in the clinical practice. In perceptual analysis, the study [158] observed that most of the MSA patients had combination of hypokinetic, ataxic, and spastic dysarthria. In contrast, PSP patients manifest prominent hypokinetic, and spastic components and comparably less ataxic components [160]. Indeed, studies related to the differential diagnosis of PSP and MSA are very few. In differential diagnosis perspective, one of the study [106] found eight acoustic dimensions in perceptual investigation which differentiate PSP and MSA. The latter study also observed that both groups manifest a combination of dysarthria. Moreover, as both groups manifest a combination of dysarthria, severity would be more helpful than specific speech dimensions for differential diagnosis. In contrast, differential diagnosis between PSP and MSA found to be very much challenging in the recent study [197].

Although perceptual method remains a gold standard for clinical differential diagnosis, judgments of severity, many decisions about management, and the assessment of meaningful temporal change, it is subject to unreliability among clinicians, difficult to quantify, and cannot directly test hypotheses about the pathophysiology underlying perceived speech abnormalities [69].

On the other hand, acoustic methods can visually display and numerically quantify different speech parameters. In objective analysis, the study [281] showed that PSP patients manifest more significant speech impairment compared to PD by acoustic components, speech velocity, intonation variability, and the fraction of intra-word pauses. In the study [126], male MSA-P exhibited more prominent speech abnormalities compared to PD in acoustic components, increased voice pitch, prolonged pause time, and reduced speech rate. Notably, the latter study considered patients with disease stage (0-3 years). A pioneer work [247] provides a quantitative and objective analysis of speech characteristics for the discrimination between PD and APS and between MSA and PSP. The basic conclusion is that PD speakers manifest pure hypokinetic dysarthria, ataxic components are more affected in MSA whilst PSP subjects demonstrate severe deficits in hypokinetic and spastic elements of dysarthria. Using an SVM with a Gaussian radial basis kernel and an exhaustive search, [247] reported a 95% accuracy in objective discrimination between APS and PD and 75% in discrimination between PSP and MSA. It was emphasized that classification performance was not the main purpose of [247], but rather a way to seek disease-specific dysarthric signs. Though univariate feature analysis provided important information, collective impairment in different subsystems of speech production or in dysarthria by disease groups was missing in the latter study. Individual subsystem of speech production related acoustic components were designed from spontaneous speech task to differentiate PD, PSP, and MSA [115]. The latter study showed that PSP patients predominate deficits in respiration function over speech production. Particularly, PSP patients manifest prolonged duration of stop consonants, and pauses. Although, the latter study devised encouraging acoustic features, it did not examine overall subsystem disorder by any acoustic index which would discriminate PD and APS and within APS. In addition, gender dimorphism was not analyzed in the study [115] whereas some prosodic features showed gender dimorphism in the study [282, 126]. Imprecise stop consonants also found to be a predominate acoustic disorder for PSP and MSA groups [264, 306]. While PSP patients predominate in unvoiced stops by prolonged VOT, MSA patients manifest reduced prevoicing of voiced stops. Notably, VOT, VOT ratio, and following vowel duration were computed by manual segmentation. Moreover, gender dimorphism was not found in speech parameters [306]. In contrast, study [267] showed gender dimorphism in speech parameter VOT. In another recent study [175], standard linear and generalized linear models were explored to address the curse of dimensionality problem, particularly for small amounts of data. The later study grouped acoustic features into three subsystems of speech productions such as phonation, articulation, and prosody. This study led to an 80% accuracy in classification between MSA and PSP. However, the latter study did not provide any subsystem wise impairment. The study [255] attempted to design a dysarthria index to discriminate between phenotypes of MSA (MSA-P and MSA-C). The latter

study established quantitatively that PD patients are characterized by pure hypokinesia whereas MSA patients manifest combination of ataxic, hypokinetic, and spastic dysarthria.

Previous studies mentioned that early differential diagnosis is challenging between PD and APS and within APS. A single study is found [126] where early-stage differential diagnosis was attempted to differentiate PD and MSA-P. Gender dimorphism would be an important parameter to consider while performing differential diagnosis. Gender-related disparity was observed in previous studies [109, 275, 280].

The present study focuses on defining new speech indexes which can objectively measure deficits in subsystems of speech production and/or particular dysarthria subtypes for disease groups PSP and MSA. Acoustic dimensions are designed so that it can show disease specificity. Such features would have a (statistical) behavior for PSP, which is significantly different from MSA. Moreover, those features will be designed in such a way that they can be interpretable in order to improve the understanding of speech impairments in PSP and MSA. Obviously, the first benefit of such investigation would be accurate and objective discrimination between PSP and MSA, given that subjective evaluation is quite challenging due to similar perceptual behavior [197]. The second and more important benefit is to potentially allow drawing hypotheses regarding the early stage of the diseases. Furthermore, participant's gender is also considered in this study which can infer additional information regarding pathology. Previous studies also indicated gender dimorphism, but those studies mostly used less number of acoustic components [126, 282]. Conversely, some studies found the influence of gender was either independent [247] or ignored [255, 164] in previous differential diagnosis studies. Hence, a detailed analysis of gender dimorphism is warranted for finding gender influence over speech parameters. To the end, propose a methodology to devise 2-dimensional speech markers which would yield good discrimination between PSP and MSA.

9.2 Methodology

9.2.1 Database

Acoustic features are computed from different speech tasks. Respiration features are computed from text reading and monologue. Phonation features are computed from sustained vowel /a/, text reading, and monologue. Articulation features are measured from DDK task (/pa-ta-ka/), sustained /a/, reading text, and monologue. Timing and prosodic features are computed from DDK task (/pa-ta-ka/), reading text, and monologue. Nasal features are computed from sustained /i/.

In Section ??, clinical details of participants are summarized. It consist 150 HC, 20 PD, 19 PSP, and 25 MSA.

9.2.2 Acoustic feature

Individual speech parameters were developed by the collaborator [112]. Previous studies [114, 112] analyzed only individual speech parameters for different disease groups. In this thesis, individual features are used to design 1-dimensional speech index to capture particular deficits in subsystems and dysarthria dimensions. To do so, individual speech parameters are first grouped according to related subsystems and dysarthria subtypes. For a better understanding of speech parameters, a brief description of those features is essential. Description of all the selected features is provided in Chapter 3. Detailed categorization of acoustic features according to subsystems of speech production and it's dysarthria type are provided in the Figure 9.1.

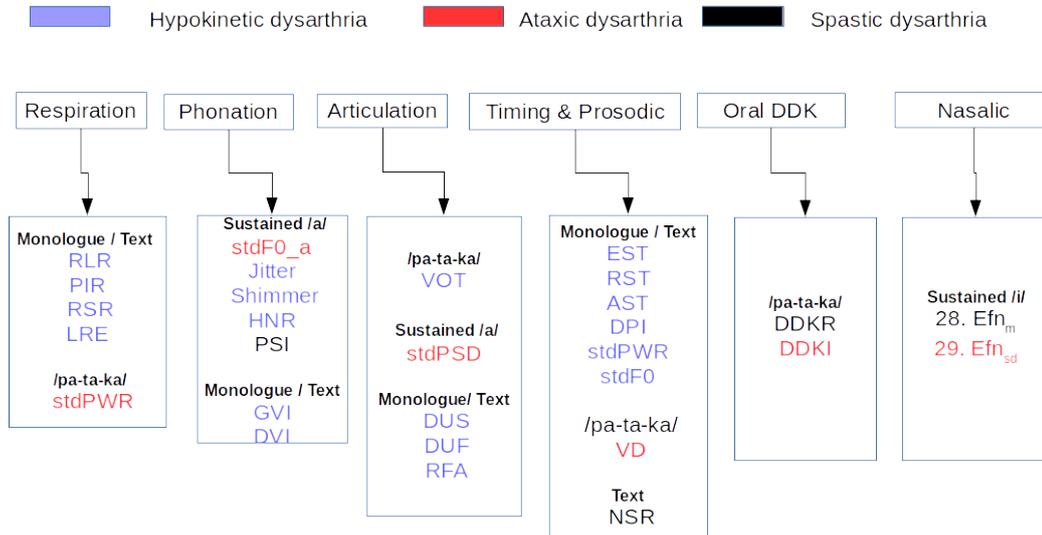


Figure 9.1: Acoustic features and its categorization

9.2.3 Acoustic feature analysis

Univariate analysis

Univariate analysis of acoustic features are first accomplished by data statistics. At first, individual acoustic parameter's distribution (normality test) is checked using the one-sample Kolmogorov-Smirnov test. It is found that several acoustic parameters are not normally distributed. Hence, Mann-Whitney U test is applied for pairwise group differences in non-normally distributed data. On the other hand, Student's t-Test is used when data is normally distributed. In the following experiments, statistical significance is represented as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, ***** $p < 0.00001$.

Multivariate analysis

In individual acoustic features analysis, HC group is used to find the influence of gender dimorphism. Additionally, gender disparity is also tested for particular disease groups. For this purpose, statistical group difference is computed by Student t-test (resp. Mann-Whitney U test) if data is normally (resp. non-normally) distributed. Before acoustic feature combination, individual features are converted to the z-score using the HC mean and standard deviation. For acoustic features in which lower raw scores are associated with greater dysarthria, the z-score are reversed. Thus higher z-scores indicate more speech impairment. List of features reversed (multiply by -1) as follows:

1. Respiration: RLR (monologue, text) , PIR (monologue, text)
2. Phonation: HNR, GVI (monologue, text)
3. Articulation: RFA (monologue, text)
4. Oral diadochokinesis: DDKR
5. Timing: EST (monologue, text), RST(monologue, text), AST (text), NSR
6. Prosodic: stdF0 (monologue, text), stdPWR(text)
7. Nasalic: Efn_m

In the next step, acoustic features from each subsystem (respiration, phonation, articulation, timing, prosody, diadochokinetic, and nasal) are mostly grouped in two clusters to manifest greater PSP and MSA patients speech impairments in either cluster. Then, according to group difference value, acoustic features from each cluster are combined by manual weight (for PSP and MSA). Knowledge based feature combination using manual weight would serve severity by its value. Thus single-dimensional index for each subsystem is designed. To the end, 2-dimensional acoustic features (F1, F2) are computed by combining subsystem indexes. Parallel to this, each respective acoustic index is used to find correlation with NNIPPS severity subscores.

Another approach also adopted to design dysarthria indexes (hypokinetic, ataxic, spastic) by regrouping acoustic features. Again, each dysarthria subtypes may have two groups where either PSP or MSA display prominent severity. Similar to the subsystem of speech production index, dysarthria indexes are also combined by manual weight to end up with 2 dimensions (D1, D2).

Correlation analysis

A correlation study is adopted to find the relationship between acoustic dimensions w.r.t. clinical motor symptoms encountered in parkinsonism. The NNIPPS scale measured clinical symptoms responsible for speech production. In this study, total 15 motor symptoms such as mental, bulbar/pseudobulbar, ADL / mobility, tremor,

rigidity, myoclonus, bradykinesia, bradykinesia (axial and limb), oculomotor, dystonia (axial and limb), pyramidal, cerebellar, orthostatic, and urinary were assessed by neurologist. These motor symptoms like dystonia, urinary, myoclonus, and pyramidal were found discretely present in disease groups. As a consequence, later motor symptoms are ignored. Two bradykinesia subscores were averaged to represent the overall bradykinesia subscore. Finally, 6 NNIPPS subscores (bradykinesia, rigidity, tremor, bulbar/pseudobulbar, oculomotor, and overall) are considered for finding relation with speech disorder. In correlation measure, Pearson(resp. Spearman) method is used to compute correlation while data is normally (resp. non-normally) distributed.

Gender dimorphism

In individual acoustic features analysis, HC group is mainly used to find influence of gender dimorphism. Additionally, gender disparity is also tested for individual disease groups. For this purpose, statistical group difference is computed by Student t-test (resp. Mann-Whitney U test) if data is normally (resp. non-normally) distributed. Each designed acoustic index is tested to find gender dimorphism in 4 groups. Group difference ($p < 0.05$) value is considered to decide manifestation of gender disparity.

9.2.4 Classification

In this experiment, Logistic Regression (LR) is used for the classification and Leave-one-sample-out (LOSO) method for training/test. As LR is frequently used in medical fields, present study adopted it for classification purpose [37].

9.3 Experimental result

9.3.1 Univariate analysis

First step was to analyze individual features statistically to find group difference between PSP and MSA. Table 9.1 summarized group difference of individual feature between PSP and MSA. Subscript “ml” stands for monologue and “txt” stands for text reading. Among all features, only 3 acoustic measures from articulation (RFA_{ml} , RFA_{txt} , $stdPSD$) and 1 acoustic measure from phonation ($stdF0_a$) subsystem show group difference between PSP and MSA. Finding less number of speech parameters in PSP and MSA discrimination is normal as stated in previous study [197] where the perceptual investigation was conducted.

Figure 9.2 displays some encouraging individual acoustic features for discrimination of PSP and MSA groups. MSA patients manifest higher impairment in ataxic component ($stdF0_a$, $stdPSD$) and hypokinetic component (RFA_{ml} , RFA_{ml}) compared to PSP. PSP patients predominantly exhibited in hypokinetic component (RSR_{ml} , DUS_{ml}). Notably, low value in “RFA” implies high impairment, thus MSA patients exhibit greater impairment in “RFA”.

Correlation with severity subscore: It is observed in quantitative analysis that some of the acoustic features strongly correlated with NNIPPS subscore. In over-

Feature	HC		PD		PSP		MSA		HC vs PD		HC vs PSP		HC vs MSA		PSP vs MSA	
	Mean/SD (Range)	P-value	Mean/SD (Range)	P-value	Mean/SD (Range)	P-value	Mean/SD (Range)	P-value	P-value	P-value	P-value	P-value	P-value	P-value	P-value	P-value
Hypokinetic:																
<i>RLR_{end}</i>	-24.66/3.80 (-33.09-14.77)		-25.47/3.46 (-30.63-16.72)		-28.35/5.11 (-37.89-18.49)		-27.21/3.47 (-36.25-22.10)		0.289	0.002	0.005		0.26			
<i>RLR_{int}</i>	-26.48/3.55 (-35.95-16.46)		-26.95/2.85 (-33.16-23.21)		-27.97/5.22 (-38.06-18.53)		-26.69/3.89 (-32.60-16.93)		0.786	0.127	0.901		0.374			
<i>RSR_{end}</i>	17.08/4.36 (8.02-33.13)		16.22/4.58 (9.35-26.99)		21.68/5.51 (9.18-34.54)		18.57/4.12 (9.82-24.74)		0.360	0.0004	0.068		0.069			
<i>RSR_{int}</i>	17.03/4.00 (7.41-27.93)		18.11/4.10 (9.23-26.08)		21.48/6.17 (11.68-32.81)		18.63/5.26 (7.09-27.43)		0.381	0.0040	0.126		0.188			
<i>PIR_{end}</i>	5.13/1.76 (2.00-13.00)		5.47/2.11 (2.00-12.00)		2.76/1.52 (1.00-6.00)		3.18/1.47 (1.00-6.50)		0.535	0	0		0.262			
<i>PIR_{int}</i>	6.43/1.78 (3.00-13.75)		6.44/1.52 (4.00-10.00)		4.24/2.58 (1.50-11.00)		4.55/1.85 (2.25-10.00)		0.789	0	0		0.334			
<i>LRE_{end}</i>	165.94/77.66 (49.45-434.17)		172.07/67.57 (89.00-391.85)		301.17/123.73 (112.62-570.40)		327.92/132.19 (163.85-704.05)		0.469	0	0		0.610			
<i>LRE_{int}</i>	140.96/55.52 (45.88-369.50)		140.57/50.95 (19.41-236.25)		234.49/132.60 (66.69-586.35)		206.00/101.72 (37.59-449.82)		0.625	0.004	0.0005		0.84			
<i>jitter</i>	0.48/0.24 (0.20-2.58)		0.47/0.23 (0.23-1.30)		0.65/0.35 (0.27-1.42)		0.71/0.37 (0.32-1.79)		0.34	0.209	0.0006		0.25			
<i>shimmer</i>	2.68/1.13 (0.84-7.96)		2.26/0.63 (1.52-3.75)		4.02/1.75 (1.30-8.90)		3.83/1.68 (1.41-7.71)		0.125	0.0006	0.001		0.767			
<i>HNR</i>	19.34/3.20 (10.94-26.24)		19.20/2.57 (13.20-25.76)		16.59/2.61 (11.25-22.41)		16.90/3.26 (12.31-23.78)		0.764	0.0004	0.001		0.821			
<i>DV_{I_{end}}</i>	254.59/52.24 (170.06-494.30)		259.24/78.89 (141.32-452.76)		409.66/162.76 (191.78-750.14)		353.88/94.44 (212.56-646.43)		0.666	0.0001	0		0.325			
<i>DV_{I_{int}}</i>	205.23/32.85 (146.68-387.91)		201.02/41.04 (113.08-288.04)		301.20/138.16 (125.49-558.91)		255.82/57.31 (181.75-415.21)		0.424	0.008	0		0.821			
<i>GV_{I_{end}}</i>	43.53/11.22 (11.45-72.14)		40.58/14.28 (15.32-66.33)		24.97/11.42 (8.66-44.47)		24.06/9.06 (7.94-40.72)		0.416	0	0		0.877			
<i>GV_{I_{int}}</i>	56.50/11.00 (24.90-78.96)		55.66/13.54 (25.91-74.56)		39.78/16.46 (13.84-65.76)		40.20/12.22 (18.64-66.77)		0.969	0.00004	0		0.933			
<i>VOT</i>	21.85/4.44 (11.14-35.15)		22.08/6.46 (13.47-42.11)		29.12/6.21 (16.19-40.42)		28.36/4.85 (17.66-35.31)		0.795	0.00004	0		0.767			
<i>DU_{S_{end}}</i>	25.24/9.89 (13.38-85.38)		29.91/16.60 (17.88-74.12)		48.78/17.40 (22.38-85.38)		39.20/13.61 (17.88-67.38)		0.462	0	0		0.062			
<i>DU_{S_{int}}</i>	22.42/7.43 (11.12-66.25)		27.38/14.48 (15.62-73.00)		37.30/13.49 (20.12-62.88)		33.72/16.36 (15.62-92.12)		0.286	0	0.00001		0.239			
<i>DU_{F_{end}}</i>	0.01/0.85 (-2.63-2.28)		0.44/0.77 (-1.55-1.60)		-0.50/1.21 (-3.84-1.48)		0.05/0.94 (-2.55-2.16)		0.02	0.104	0.736		0.196			
<i>DU_{F_{int}}</i>	-0.95/2.42 (-9.84-5.25)		-0.75/2.38 (-5.10-3.81)		-0.12/2.13 (-3.49-5.86)		-0.21/2.37 (-7.43-3.79)		0.88	0.273	0.057		0.530			
<i>RF_{A_{end}}</i>	9.42/1.40 (6.32-13.35)		8.22/1.37 (5.61-10.79)		9.95/1.63 (7.70-14.02)		8.64/1.26 (6.54-10.98)		0.001	0.258	0.013		0.013			
<i>RF_{A_{int}}</i>	10.68/1.54 (7.40-15.76)		9.61/1.22 (7.55-11.18)		11.24/1.78 (8.93-14.97)		10.19/1.25 (7.54-12.57)		0.007	0.252	0.18		0.062			
<i>EST_{end}</i>	1.55/0.01 (1.50-1.58)		1.55/0.01 (1.53-1.57)		1.54/0.01 (1.51-1.56)		1.53/0.02 (1.47-1.56)		0.112	0.012	0.0003		0.627			
<i>EST_{int}</i>	1.55/0.01 (1.50-1.57)		1.55/0.01 (1.53-1.57)		1.54/0.02 (1.49-1.57)		1.54/0.01 (1.52-1.56)		0.456	0.000001	0		0.455			
<i>RST_{end}</i>	364.11/60.58 (170.97-511.27)		356.40/82.75 (198.44-524.13)		241.96/86.53 (128.34-410.73)		247.24/76.25 (127.60-411.75)		0.919	0.00001	0		0.731			
<i>RST_{int}</i>	433.34/32.65 (270.35-595.46)		434.83/79.03 (304.48-653.87)		315.13/113.02 (148.11-498.11)		338.58/69.85 (206.45-474.38)		0.474	0.192	0.473		0.514			
<i>AST_{end}</i>	0.90/6.04 (-17.46-19.89)		0.30/8.04 (-11.47-19.36)		2.03/6.86 (-15.75-14.37)		0.25/4.33 (-7.54-11.82)		0.355	0.002	0.0001		0.499			
<i>AST_{int}</i>	19.37/13.88 (-12.30-56.94)		21.49/14.26 (-8.74-47.81)		8.74/10.47 (-8.23-30.16)		7.15/15.05 (-15.46-40.13)		0.436	0	0		0.84			
<i>DPI_{end}</i>	193.13/60.68 (95.21-493.41)		205.50/65.95 (124.90-373.22)		377.05/136.50 (157.83-653.95)		370.41/144.44 (129.29-691.27)		0.869	0	0		0.313			
<i>DPI_{int}</i>	149.07/25.82 (93.93-225.29)		150.43/30.77 (114.84-248.82)		291.80/150.69 (142.28-620.88)		227.07/72.66 (131.55-379.83)		0.003	0.20	0.044		0.990			
<i>stdPW_{R_{end}}</i>	3.96/0.80 (2.37-6.38)		3.33/0.70 (1.64-4.70)		4.65/1.91 (2.41-10.60)		4.34/0.88 (2.78-5.97)		0.002	0.180	0.044		0.713			
<i>stdPW_{R_{int}}</i>	3.95/0.79 (2.32-8.42)		3.40/0.79 (1.78-5.14)		3.82/1.13 (2.42-6.48)		3.81/0.94 (2.01-6.59)		0.002	0.106	0.258		0.165			
<i>stdF_{0_{end}}</i>	2.00/0.73 (0.84-6.26)		1.47/0.40 (0.81-2.29)		1.72/0.46 (1.09-2.64)		1.48/0.34 (0.88-2.35)		0.0002	0	0.00003		0.065			
<i>stdF_{0_{int}}</i>	2.51/0.75 (0.93-5.80)		1.72/0.60 (0.97-3.16)		1.28/0.27 (0.81-1.91)		1.38/0.51 (0.76-3.23)		0.000004	0	0		0.952			
Ataxic:																
<i>stdF₀</i>	0.35/0.24 (0.11-2.00)		0.36/0.28 (0.13-1.38)		0.50/0.30 (0.16-1.08)		0.69/0.25 (0.28-1.14)		0.857	0.027	0		0.019			
<i>stdPSD</i>	2.22/0.39 (1.37-3.48)		2.12/0.37 (1.41-3.04)		2.22/0.34 (1.65-2.91)		2.56/0.42 (1.85-3.96)		0.360	0.00002	0.0004		0.007			
<i>V_D</i>	49.63/9.82 (28.16-86.72)		50.38/9.45 (37.88-74.56)		95.08/52.53 (44.34-221.91)		73.96/21.67 (38.03-116.88)		0.89	0.000002	0.000001		0.40			
<i>DDKI</i>	22.50/11.82 (8.37-86.02)		26.90/12.05 (10.02-47.57)		85.75/55.11 (11.04-215.26)		63.85/29.47 (22.85-119.87)		0.103	0	0		0.291			
<i>E_{f_{nsD}}</i>	3.88/0.88 (1.40-6.41)		4.01/0.85 (2.29-6.30)		3.59/0.84 (2.47-5.38)		3.73/0.72 (2.23-5.08)		0.63	0.14	0.33		0.47			
Spastic:																
<i>PSI</i>	4.96/10.06 (0.00-71.64)		6.24/9.35 (0.00-35.70)		11.51/15.06 (0.00-55.02)		10.69/13.02 (0.00-55.87)		0.409	0.202	0.003		0.55			
<i>DDKR</i>	6.45/0.72 (3.97-8.51)		6.61/0.78 (5.26-8.19)		4.93/1.44 (2.31-7.78)		5.35/0.94 (3.52-7.23)		0.353	0.000007	0.000001		0.188			
<i>NSR_{end}</i>	2.41/0.27 (1.67-3.17)		2.43/0.28 (1.98-2.99)		2.04/0.60 (1.01-3.61)		2.14/0.39 (1.24-2.82)		0.942	0.0003	0.0009		0.400			
<i>E_{f_{nsM}}</i>	-35.14/1.47 (-38.09-28.84)		-35.10/1.17 (-36.81-31.49)		-33.66/2.63 (-36.65-24.21)		-34.31/1.48 (-37.67-30.95)		0.81	0.01	0.03		0.53			

Table 9.1: Univariate analysis of acoustic features

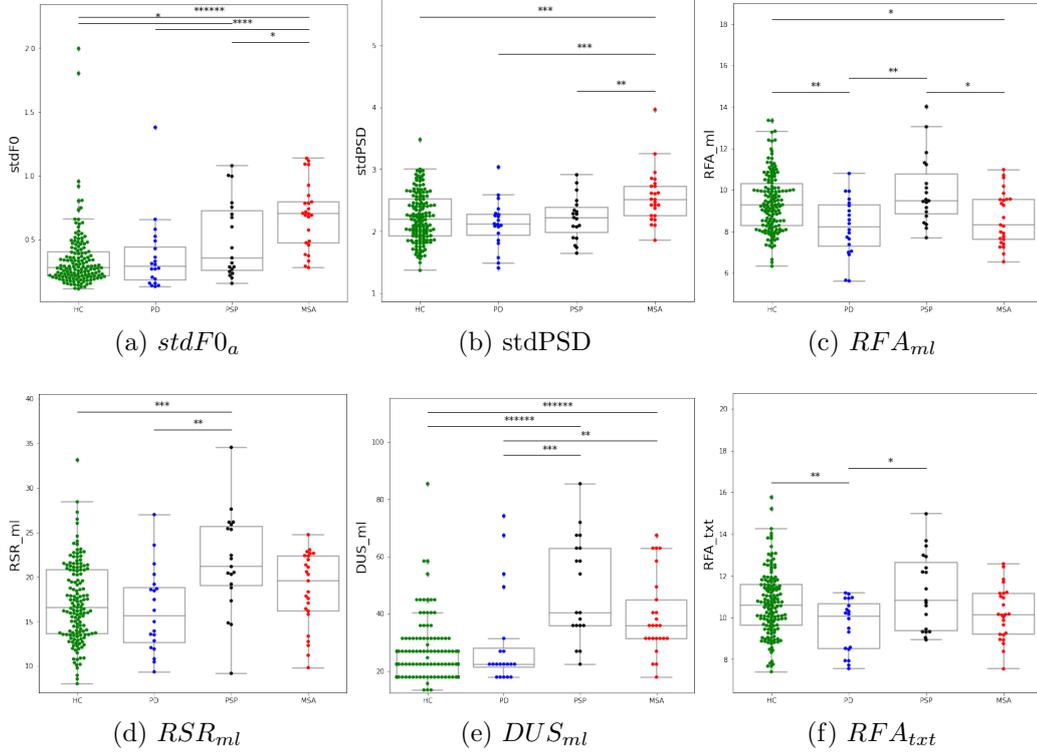


Figure 9.2: Plot of individual speech parameters which provide encouraging discrimination for PSP and MSA patients

all NNIPPS subscore, MSA patients showed significant correlation with $DDKR_{ddk}$ ($r = -0.42$, $p = 0.03$), NSR_{txt} ($r = -0.59$, $p = 0.001$), $stdF0_{ml}$ ($r = -0.48$, $p = 0.01$), $stdF0_{txt}$ ($r = -0.51$, $p = 0.007$), and DUS_{txt} ($r = 0.49$, $p = 0.01$) acoustic features. In contrast, PSP patients did not show significant correlation between individual acoustic features with overall NNIPPS subscore.

In cerebellar subscore, MSA patients exhibited significant correlation with $DDKR_{ddk}$ ($r = -0.51$, $p = 0.008$), VD_{ddk} ($r = 0.39$, $p = 0.05$), $stdPWR_{ml}$ ($r = 0.41$, $p = 0.04$), $stdPWR_{txt}$ ($r = 0.44$, $p = 0.01$) whereas PSP patients did not show correlation in latter mentioned acoustic features.

Gender dimorphism: Analysis of gender dimorphism in individual acoustic features reveals interesting differences between male and female HC subjects as presented in Table 9.2. Keeping aside some respiration, two timing, one prosodic, and one phonation acoustic features, other acoustic measures yield significant group differences between male and female HC subjects. Important to note, primarily male HC subjects are characterized by high severity compared to female HC. Notably, gender dimorphism is less frequent in disease groups (PD, PSP, MSA). Similar to HC subjects, male patients from disease groups also manifest higher impairment compared to female patients. Thus gender dimorphism would be an important parameter while acoustic features are being used for differential diagnosis.

In the classification task, individual features did not yield a good classification

	P-Value			
	HC	PD	PSP	MSA
Hypokinetic:				
<i>RLR_{ml}</i>	0.35696	0.37563	0.38369	0.44485
<i>RLR_{txt}</i>	0.24256	0.50000	0.23626	0.18003
<i>RSR_{ml}</i>	0.29321	0.37563	0.35185	0.44485
<i>RSR_{txt}</i>	0.07397	0.31723	0.06957	0.40144
<i>PIR_{ml}</i>	0.00039	0.27410	0.09691	0.11738
<i>PIR_{txt}</i>	0.23492	0.09460	0.09315	0.02661
<i>LRE_{ml}</i>	0.02980	0.40605	0.41633	0.15240
<i>LRE_{txt}</i>	0.00338	0.34597	0.48315	0.12774
<i>jitter_a</i>	0.01324	0.37563	0.38369	0.33870
<i>shimmer_a</i>	0.00000	0.00872	0.14538	0.16584
<i>HNR_a</i>	0.00000	0.01968	0.48315	0.13970
<i>DVI_{ml}</i>	0.00003	0.50000	0.03460	0.01422
<i>DVI_{txt}</i>	0.00188	0.02860	0.44956	0.02447
<i>GVI_{ml}</i>	0.00000	0.02860	0.09510	0.08707
<i>GVI_{txt}</i>	0.00001	0.01325	0.14538	0.00150
<i>VOT_{ddk}</i>	0.00009	0.43704	0.23626	0.26177
<i>DUS_{ml}</i>	0.00000	0.45072	0.07405	0.02614
<i>DUS_{txt}</i>	0.00025	0.00202	0.19860	0.05032
<i>DUF_{ml}</i>	0.98604	0.46842	0.67013	0.88012
<i>DUF_{txt}</i>	0.05984	0.40605	0.40140	0.32768
<i>RFA_{ml}</i>	0.00309	0.05650	0.01558	0.33870
<i>RFA_{txt}</i>	0.01645	0.43704	0.00800	0.10600
<i>EST_{ml}</i>	0.08133	0.40605	0.18743	0.09619
<i>EST_{txt}</i>	0.40145	0.37563	0.02861	0.07079
<i>RST_{ml}</i>	0.00001	0.31723	0.08158	0.04534
<i>RST_{txt}</i>	0.00015	0.01325	0.38369	0.01233
<i>AST_{ml}</i>	0.30745	0.31723	0.02861	0.31864
<i>AST_{txt}</i>	0.00324	0.40605	0.29138	0.16584
<i>DPI_{ml}</i>	0.00019	0.13363	0.14538	0.06358
<i>DPI_{txt}</i>	0.00023	0.04805	0.12694	0.00791
<i>stdPWR_{ml}</i>	0.00006	0.46842	0.08158	0.08707
<i>stdPWR_{txt}</i>	0.00000	0.07688	0.29138	0.03571
<i>stdF0_{ml}</i>	0.46891	0.07688	0.41633	0.22698
<i>stdF0_{txt}</i>	0.00608	0.08897	0.04967	0.42303
Ataxic:				
<i>stdF0_a</i>	0.37308	0.10242	0.06957	0.26177
<i>stdPSD_a</i>	0.00635	0.02860	0.23626	0.11651
<i>VD_{ddk}</i>	0.13871	0.37563	0.26308	0.07079
<i>DDKI_{ddk}</i>	0.00859	0.21406	0.48315	0.03157
<i>Efn_{-SDI}</i>	0.01197	0.50000	0.21102	0.08707
Spastic:				
<i>PSI_a</i>	0.05554	0.18989	0.08957	0.27048
<i>DDKR_{ddk}</i>	0.00177	0.40605	0.08158	0.11651
<i>NSR_{txt}</i>	0.14797	0.08897	0.08158	0.06358
<i>Efn_{-MI}</i>	0.02702	0.28955	0.23626	0.13970

Table 9.2: Gender diphormism analysis for individual acoustic features in HC, PD, PSP, and MSA groups; $p - value < 0.05$ is marked as bold

score. While individual features are not sufficient to discriminate PSP and MSA patients, the multivariate analysis may help describe particular deficits. In the following section, subsystem related deficits are first investigated for disease groups. Next, disease specific dysarthria manifestation is analyzed for understanding underlying pathophysiology.

9.3.2 Speech features by subsystems of speech

1D Respiration feature

Four features (RLR, PIR, RSR, and LRE) are related to the problem of movement and initiation of inspiration and expiration. Individual respiration speech parameters did not show group differences ($p < 0.05$) between PSP and MSA. Mainly, RSR features (from monologue and reading text) individually did not show group difference, but the average of those two features yield group difference as provided in Table 9.3. PSP patients manifest a more significant respiration rate (RSR) deficit, probably due to reduced vital capacity. PIR seems to have an inverse physical relation with RSR. Reduced inspiratory capacity may lead to compromise in the number of pauses in-between respiration. Thus PIR is kept though it did not find group differences. Obstruction (resp. reduced efficiency) in the airways may increase the loudness of respiration (resp. reduced loudness), which is measured by RLR. RLR is also kept in the overall respiration feature because it may provide complementary information compared RSR and PIR. On the other hand, latency in speech exchange (LRE) did not show any indication to differentiate PSP and MSA. Averaging PIR and RLR (monologue and text) did not yield group differences as summarized in Table 9.3.

	P-Value					
	HC vs PD	HC vs PSP	HC vs MSA	PD vs PSP	PD vs MSA	PSP vs MSA
$0.5*(RSR_{ml}+RSR_{txt})$	0.89	0.00014	0.039	0.003	0.22	0.045
$0.5*(PIR_{ml}+PIR_{txt})$	0.44	3.67e-06	6.16e-08	0.0002	4.75e-05	0.29
$0.5*(RLR_{ml}+RLR_{txt})$	0.38	0.011	0.13	0.10	0.64	0.29
$0.5*(LRE_{ml}+LRE_{txt})$	0.63	7.85e-06	1.86e-08	0.0007	7.76e-05	0.84
$0.25*(PIR_{ml}+PIR_{txt}+RLR_{ml}+RLR_{txt})$	0.85	1.7e-06	1.2e-05	0.0003	0.0017	0.291

Table 9.3: Combination of respiration components

Till now, it is observed that only RSR measure yields group difference and combination of PIR, RLR even did not produce discrimination for PSP and MSA. However, combining PIR, TLR with RSR by giving more importance to RSR (a factor of 2) improves group difference. Respiration feature (F_r) is defined by the Equation 9.1.

$$F_r = \frac{\mu(RLR_{ml}, RLR_{txt}, PIR_{ml}, PIR_{txt})}{2} + \mu(RSR_{ml}, RSR_{txt}) \quad (9.1)$$

The designed respiration index (F_r) will represent the measure of overall respiration deficits. More precisely, the developed 1-dimensional respiration index may capture deficiency in vital capacity. Reduced vital capacity results in increased respiration rate, decreased number of pause, and reduced respiration loudness. Figure 9.3 displays the group difference using respiration index among four groups. Globally, both PSP and MSA patients exhibit respiration deficits compared to HC and PD. However, PD patients did not manifest respiration deficits compared to HC. Important to note, PSP patients manifest predominant respiration deficits compared to MSA and PD in F_r .

Correlation with F_r : Significant correlation of designed respiration index w.r.t. clinical severity (bradykinesia, rigidity, tremor, cerebellar etc.) could provide additional information. Only PSP patients showed significant correlation between respiration index (F_r) and bulbar NNIPPS subscore ($r = 0.52$, $p - value = 0.02$) and

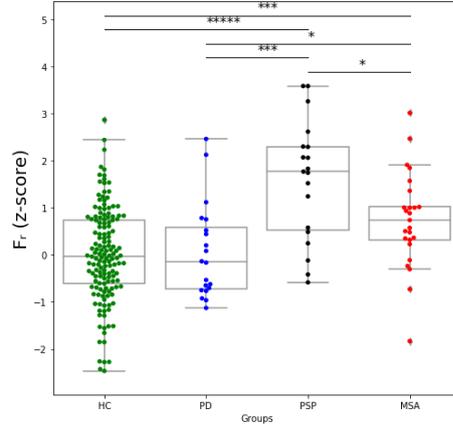


Figure 9.3: Respiration index (F_r) for the 4 groups

with oculomotor NNIPPS subscore ($r = 0.49$, $p = 0.03$). F_r did not show significant correlation with bradykinesia and rigidity for both PSP and MSA while general presumption was respiration deficits are attributed to rigidity and bradykinesia.

Gender dimorphism with F_r : Important to note, respiration index (F_r) did not show gender dimorphism in HC ($p = 0.88$), PD ($p = 0.54$), PSP ($p = 0.14$), and MSA ($p = 0.64$). Thus F_r can be used for differential diagnosis without considering gender differences. Notably, a group of PSP male patients displayed higher respiration disorder compared to female PSP patients. In addition, it is required to consider that each disease group consists of a small number of male and female patients.

1D Phonation feature

Out of 7 phonation features, DVI and GVI are measured from monologue and text recording protocol and rest phonation features are measured from sustained /a/. Only $stdF0_a$ yields a group difference between PSP and MSA. While the intention is to design a phonation index, $stdF0_a$ alone is not sufficient to capture laryngeal deficits. Thus, other phonation features also need to be explored for capturing complementary laryngeal deficits. Table 9.4 summarizes group differences of homogeneous phonation components. Average of GVI yields good discrimination between PD and APS, but not for PSP and MSA. On the other hand, a group of PSP patients (6 out of 19) exhibited prolonged vowel duration by speech parameter DVI compared to MSA. Harsh voice dimension also did not provide differentiation between PSP and MSA. However, median value of individual harsh voice component suggest that MSA predominate in jitter and PSP predominate in shimmer and HNR.

	P-Value					
	HC vs PD	HC vs PSP	HC vs MSA	PD vs PSP	PD vs MSA	PSP vs MSA
$\frac{GVI_{ml} + GVI_{txt}}{2}$	0.94	5.39e-07	7.89e-11	0.001	0.0003	0.95
$\frac{DVI_{ml} + DVI_{txt}}{2}$	0.37	0.0001	8.16e-09	0.003	0.0001	0.46
$\frac{jitter + shimmer + HNR}{3}$	0.61	0.0003	0.0002	0.002	0.002	0.91

Table 9.4: Combination of phonation components

In the next step, phonation components that manifest a higher impairment in MSA are considered to combine with $stdF0_a$ and remaining speech parameters are combined to devise another phonation group where PSP predominate. After analyzing all phonation acoustic features, two phonation groups showed PSP and MSA patients manifest disparity. One phonation group DVI, shimmer, HNR which mainly captures decreased control and coordination of the laryngeal muscles, and reduced range of movement in laryngeal muscles, is severe for PSP patients. Thus, average DVI (monologue and text), shimmer, and HNR are applied to capture the latter mentioned phonation deficits. The first phonation index (F_{p1}) is defined as follows:

$$F_{p1} = \frac{DVI_{ml} + DVI_{txt} + Shimmer + HNR}{4} \quad (9.2)$$

Figure 9.4 displays group differences among four groups by feature F_{p1} . A group of PSP patients manifests relatively high severity compared to MSA (group difference $p=0.1$).

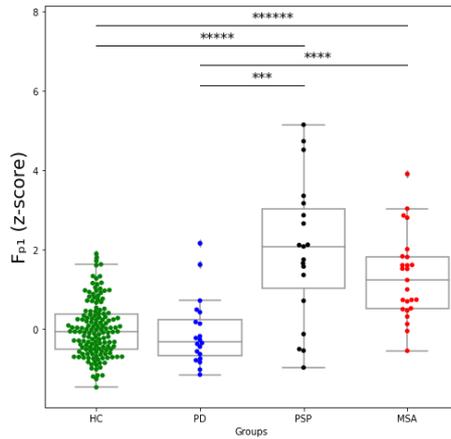


Figure 9.4: First phonation index (F_{p1}) for the 4 groups

Correlation with F_{p1} : In correlation experiment, F_{p1} showed significant correlation with oculomotor NNIPPS subscore for PSP ($r = 0.46$, $p = 0.04$). It is observed that MSA patients hardly show oculomotor severity.

Gender dimorphism with F_{p1} : Strong gender dimorphism is observed in the phonation index, F_{p1} . Distinct gender difference is found for HC ($p = 0$), PD ($p = 0.02$), MSA ($p = 0.01$), but not for PSP ($p = 0.52$). In all groups most of male participants manifest greater disorder compared to female participants. Thus it is important to check influence of gender difference in group difference.

On the other hand, a group of MSA patients shows higher severity in another phonation group $stdF0_a$, jitter, GVI which mainly captures irregularity in glottal cycle. Thus, combination by averaging of features mentioned above, giving more importance on $stdF0_a$ (factor of 2) come up as second phonation dimension (F_{p2}) which is orthogonal to the first phonation impairment (F_{p1}). This phonation dimension is defined by Equation 9.3.

$$F_{p2} = stdF0_a + \frac{\mu(Jitter, PSI, GVI_{ml}, GVI_{txt})}{4} \quad (9.3)$$

A group of MSA patients show comparably higher deficits (irregularity) in vocal fold vibration from other groups in feature F_{p2} . Figure 9.5 presents group differences with F_{p2} .

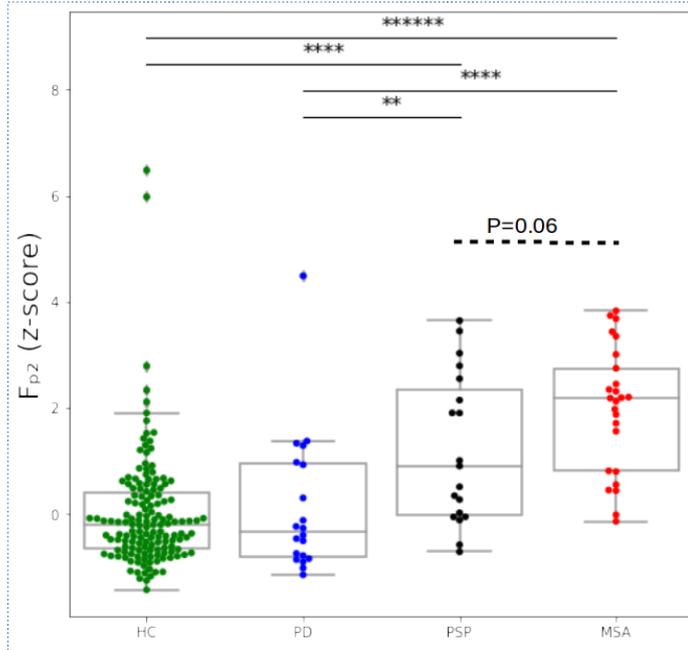


Figure 9.5: Second phonation index (F_{p2}) for the 4 groups

Correlation with F_{p2} : Figure 9.6 shows the significant correlation of F_{p2} w.r.t. overall NNIPPS score. Phonation index (F_{p2}) yields correlation with overall NNIPPS score for PSP patients ($r = 0.46$, $p = 0.04$). It is encouraging to observe that PSP patients present the significant correlation between F_{p2} and overall NNIPPS score. In contrast, irrespective of severity MSA patients mostly manifest high impairment in F_{p2} . It would be an important speech marker for the early stage diagnosis. Notably, it is also required to notice that the MSA group has fewer patients in low severity subscores. Thus it demands more data to validate the initial indication about early differential diagnosis.

Gender dimorphism with F_{p2} : Notably, phonation index, F_{p2} did not show gender dimorphism in HC ($p = 0.16$), PD ($p = 0.27$), PSP ($p = 0.25$), and MSA ($p = 0.48$). However, a group of male subjects from HC group showed comparably greater disorder than female groups.

1D Articulation feature

Total 5 acoustic features are present in the articulation group. Important to note, RFA_{ml} and $stdPSD$ yield group differences ($p < 0.05$) between PSP and MSA. The latter mentioned two features capture reduced range of movements and involuntary

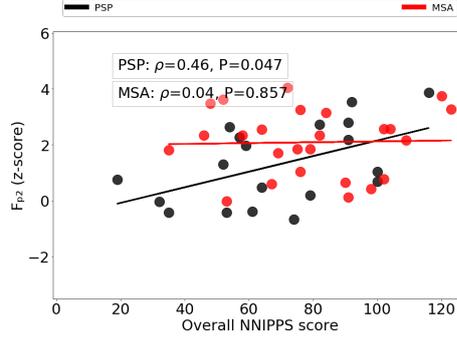


Figure 9.6: Correlation of F_{p2} feature w.r.t. overall severity

movements of articulators respectively. Notably, RFA from monologue provides better discrimination compared to reading text which opens up another possibility of speech disorder, but not considered in this thesis. In both features, MSA patients manifest greater impairment compared to PSP. On the other hand, PSP patients show prolongation of unvoiced stops (DUS_{ml}) compared to MSA. DUS_{ml} indirectly measure the impaired stop consonant articulation or articulatory precision from a spontaneous speech in the time domain. In both DUS_{txt} and VOT , PSP and MSA patients exhibit significant prolongation of unvoiced stop consonants compared to PD and HC. However, DUS_{txt} and VOT are not suitable for discriminating between PSP and MSA. Table 9.5 presents group differences by the combination of homogeneous articulation acoustic features. Only the RFA feature yields a group difference between PSP and MSA. The combination of unvoiced stop consonants duration features even did not provide a group difference between PSP and MSA.

	P-Value					
	HC vs PD	HC vs PSP	HC vs MSA	PD vs PSP	PD vs MSA	PSP vs MSA
$\frac{RFA_{ml}+RFA_{txt}}{2}$	0.001	0.24	0.04	0.003	0.17	0.02
$\frac{DUS_{ml}+DUS_{txt}+VOT}{3}$	0.22	1.77e-09	7.66e-09	0.0001	0.001	0.27
$\frac{DUF_{ml}+DUF_{txt}}{2}$	0.09	0.31	0.18	0.059	0.76	0.1

Table 9.5: Combination of articulation features

Thus, two groups were finally selected from articulation subsystem. In first group only DUS_{ml} has been considered where PSP patients show greater severity in speech control and coordination. Figure 9.2 displays group difference among four groups with the feature DUS_{ml} .

The articulation index (F_a) is defined by taking average of $RFA_{ml,txt}$ and $stdPSD$.

$$F_a = \frac{RFA_{ml} + RFA_{txt} + stdPSD}{3} \quad (9.4)$$

In this articulation index (F_a), MSA patients manifest predominant impairment compared to HC and PSP groups. Figure 9.7 shows group difference analysis with F_a .

Correlation with DUS_{ml} and F_a : Articulation feature, DUS_{ml} did not show significant correlation with any severity subscore for both PSP and MSA. On the

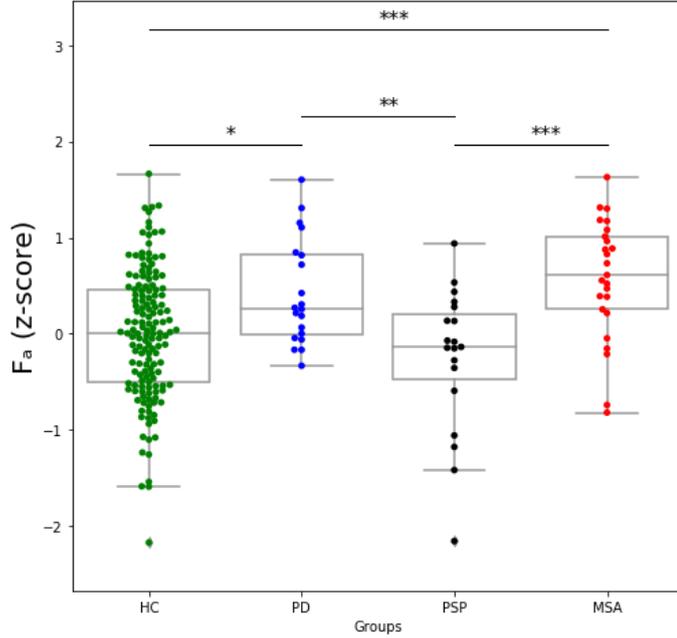


Figure 9.7: Articulation feature (F_a) for the 4 groups

other hand, articulation index (F_a) show encouraging correlation with bradykinesia NNIPPS subscore for PSP patients ($r = 0.4$, $p = 0.09$). Important to note, MSA patients display high impairment in F_a irrespective of severity in bradykinesia NNIPPS subscore, whereas PSP patients show a strong correlation. Thus a preliminary hypothesis can be drawn about the early stage differential diagnosis. PSP patients would show low impairment whereas MSA patients would manifest high impairment at early stage.

Gender dimorphism with F_a : The articulation index, F_a did not show gender demorphism in HC ($p = 0.16$), PD ($p = 1.0$), MSA ($p = 0.93$), except in PSP ($p = 0.03$).

1D Prosodic feature

In the prosodic subsystem group, two acoustic features related to monopitch and monoloudness are present. First, in monopitch features (both from monologue and reading text), PD, PSP, and MSA patients manifest significant disorder compared to HC as provided in Table 9.6.

	P-Value					
	HC vs PD	HC vs PSP	HC vs MSA	PD vs PSP	PD vs MSA	PSP vs MSA
$\frac{stdF0_{ml} + stdF0_{txt}}{2}$	3.76e-06	9.55e-08	1.68e-10	0.8	0.14	0.21
$\frac{stdPWR_{ml} + stdPWR_{txt}}{2}$	0.54	0.008	0.01	0.008	0.02	0.99

Table 9.6: Combination of prosodic features

However, latter acoustic features did not exhibit discrimination between PSP and MSA as both disease groups manifest impairment. Notably, monoloudness did not

show impairment in PSP and MSA compared to HC which is questionable. Conversely, PD, PSP, and MSA patients exhibit predominant impairment in monopitch compared to HC. Hence, prosodic index is designed by only monopitch features (from monologue and reading text) as follows:

$$F_{pr} = \mu(stdF0_{ml}, stdF0_{txt}) + \frac{\mu(stdPWR_{ml}, stdPWR_{txt})}{2} \quad (9.5)$$

The prosodic index yields encouraging group difference in pathological groups compared to HC as illustrated in Figure 9.8. Thus this F_{pr} index could be used for measuring overall speech disorder but not for discriminating PSP and MSA.

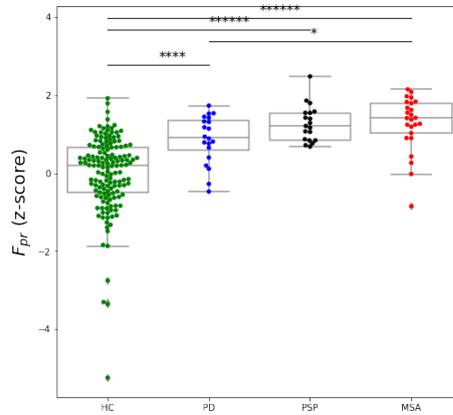


Figure 9.8: Group differences between groups by prosodic index F_{pr}

Correlation with F_{pr} : Monopitch feature show significant correlation w.r.t. overall ($r = 0.51, p = 0.009$), bradykinesia ($r = 0.51, p = 0.009$), rigidity ($r = 0.43, p = 0.032$), bulbal/pseudobulbar ($r = 0.38, p = 0.059$) NNIPPS subscore for MSA patients. Irrespective to severity, PSP patients show greater impairment in monopitch feature. Prosodic index (F_{pr}) yields significant correlation w.r.t. overall ($r = 0.48, p = 0.016$), bradykinesia ($r = 0.58, p = 0.002$), bulbal/pseudobulbar ($r = 0.44, p = 0.02$) NNIPPS subscore for MSA patients whereas PSP patients did not exhibit correlation.

gender dimorphism with F_{pr} : Prosodic index did not show gender difference in HC ($p = 0.16$), PD ($p = 0.47$), PSP ($p = 0.52$), and MSA ($p = 0.84$) groups.

1D Timing feature

In timing subsystem, individual feature analysis did not find discrimination between PSP and MSA. However, those features are considered to keep for further analysis as those are presented in the study [114]. Those timing components should capture timing impairments, particularly in MSA. Thus acoustic features, EST_{ml} , RST_{ml} and AST_{txt} are averaged to capture timing deficits. The timing index is defined by the Equation 9.6.

$$F_t = \frac{EST_{ml} + RST_{ml} + AST_{txt}}{3} \quad (9.6)$$

The proposed timing index yields encouraging discrimination between PD and APS (PSP and MSA). This index also shows marginally higher severity for a group of MSA patients though it did not show group differences between PSP and MSA. Figure 9.9 shows group differences among groups with timing feature (F_t).

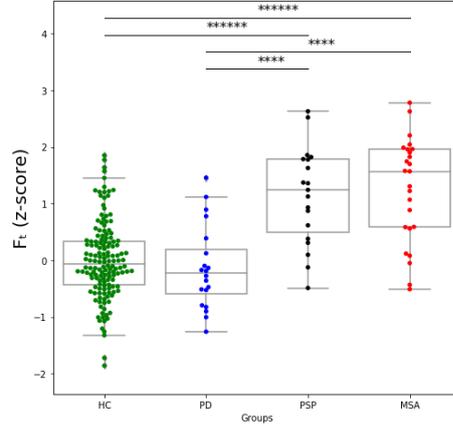


Figure 9.9: Timing feature (F_t) for the 4 groups

Correlation with F_t : Designed timing index, F_t did not show significant correlation in any severity subscore. However, MSA patients show indicative correlation w.r.t. bradykinesia ($r = 0.36$, $p = 0.07$), rigidity ($r = 0.37$, $p = 0.07$), and bulbar/pseudobulbar ($r = 0.36$, $p = 0.08$) NNIPPS subscore. It can be hypothesized that there is great influence of aforesaid severity subscores on timing index.

Gender dimorphism with F_t : Timing index (F_t) also manifest gender dimorphism in HC group ($p = 0.00002$) and in MSA ($p = 0.06$). PD ($p = 0.53$) and PSP ($p = 0.16$) did not show gender difference in timing index. In overall, male participants manifest high disorder in timing index compared to female participants.

Oral diadochokinetic feature

Features from the oral diadochokinetic group did not yield group differences for PSP and MSA. In both DDKR (syllable rate) and DDKI (syllable irregularity), PSP and MSA patients exhibit predominant impairment compared to HC and PD. APS group manifests predominant deficits in DDK features. Average of DDK features also replicate individual feature. F_{ddk} is defined as follows:

$$F_{ddk} = \frac{DDKR + DDKI}{2} \quad (9.7)$$

Figure 9.10 displays group difference result for diadochokinetic index (F_{ddk}). F_{ddk} yield encouraging discrimination between PD and APS, but did not produce group difference between PSP and MSA. However, a group of PSP (6 PSP) patients manifest distinctively high disorder compared to MSA in DDK index.

Correlation with F_{ddk} : F_{ddk} yields significant correlation w.r.t. bradykinesia ($r = 0.43$, $p = 0.032$), cerebellar ($r = 0.45$, $p = 0.024$) for MSA patients. On the

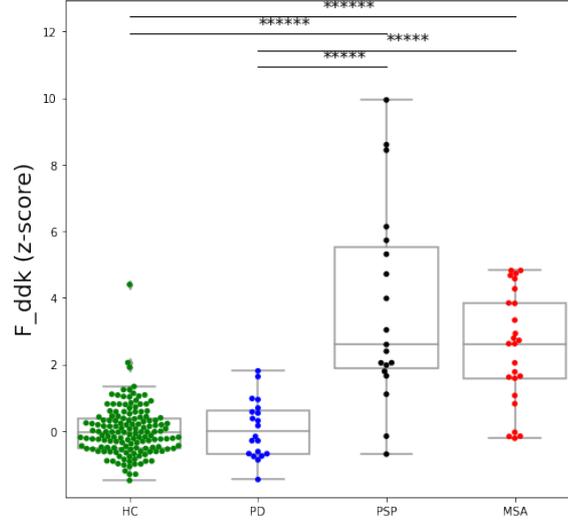


Figure 9.10: Group differences with DDK index (F_{ddk}) for the 4 groups

other hand, PSP patients present significant correlation for F_{ddk} w.r.t. oculomotor NNIPPS subscore ($r = 0.67$, $p = 0.002$).

Gender dimorphism with F_{ddk} : In diadochokinetic index, F_{ddk} also yield gender independence for HC ($p = 0.84$), PD ($p = 0.60$), and PSP ($p = 0.58$). Conversely, marginal gender difference ($p = 0.06$) is found in MSA group. A group of malt MSA patients showed greater disorder than female female patients.

1D Nasalic feature

Two nasalic features did not show group differences for PSP and MSA. However, both PSP and MSA patients manifest hypernasality by Efn_M compared to PD and HC. The combination of nasalic feature even did not help to differentiate PSP and MSA.

Design two speech dimensions

The previous section illustrates that respiration, phoantion, and articulation indexes are important for discrimination of PSP and MSA. However, those individual subsystem indexes are not enough for differential diagnosis of PSP and MSA. Now, it is required to combine homogeneous subsystem indexes to design two speech dimensions. In that direction, respiration index (F_r) and single articulation feature, DUS_{ml} are found where PSP patients show greater impairment compared to MSA. Thus the linear combination of F_r and DUS_{ml} generate the first dimension (X1) where more importance is given to respiration impairment (factor of 2). X1 is defined by the Equation 9.8.

$$X1 = F_r + \frac{DUS_{ml}}{2} \quad (9.8)$$

Figure 9.11 illustrates group differences analysis among four groups with feature X1. This feature improves group difference between PSP and MSA than individ-

ual subsystem indexes. However, this feature alone is not sufficient for differential diagnosis.

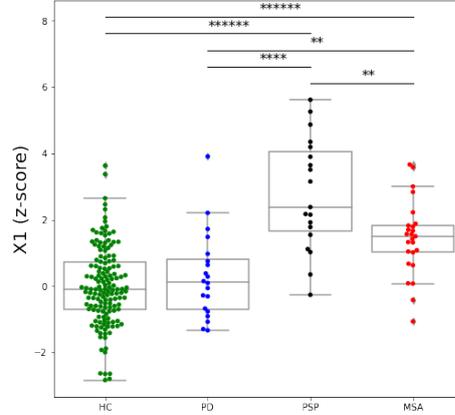


Figure 9.11: Feature F1 for the 4 groups

Correlation with X1 dimension: The first dimension $X1$ shows significant correlation with bulbar/pseudobulbar ($r = 0.52, p = 0.023$), oculomotor ($r = 0.49, p = 0.032$) for PSP patients whereas MSA patients did not show correlation in any of the NNIPPS subscores.

Gender dimorphism with X1: First dimension comprising of respiration index and single articulation speech parameter showed gender difference in HC ($p = 0.02$) and PSP ($p = 0.04$), but not in PD ($p = 0.20$) and MSA ($p = 0.27$). In agreement to previous observation, male patients manifest greater deficits compared to female patients.

To design the second speech dimension $X2$, features F_a , F_{p2} , and F_t are combined as in Equation 9.9. More importance is given to articulation impairment (factor of 2) than phonation and timing impairment for PSP and MSA discrimination task. In this speech dimension, MSA patients manifest greater deficits in articulation, phonation, and timing features.

$$X2 = F_a + \frac{F_{p2}}{2} + \frac{F_t}{2} \quad (9.9)$$

Figure 9.12 presents group differences analysis among four groups with the designed speech dimension, $X2$. Important to note, MSA patients show specificity with regards to the severity, which is an encouraging phenomenon. However, this feature alone did not yield good classification accuracy for PSP and MSA.

Correlation with X2: Second feature dimension, $X2$ shows significant correlation with overall ($r = 0.53, p = 0.02$), and bradykinesia ($r = 0.5, p = 0.028$) for PSP patients. It can be hypothesized that MSA patients manifest high deficits even at low NNIPPS severity subscores whereas PSP patients show low impairment in low severity as presented in Figure 9.13.

Gender dimorphism with X2: Second dimension ($X2$) did not show gender difference in HC ($p = 0.52$), PD ($p = 0.27$), PSP ($p = 0.22$), and MSA ($p = 0.17$).

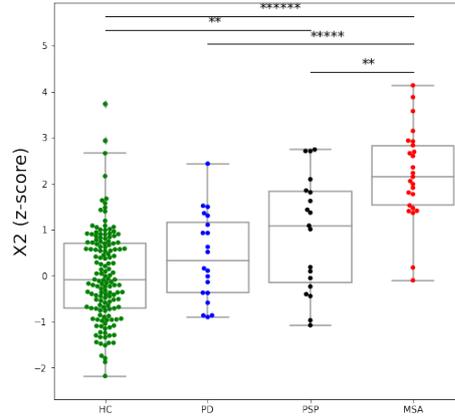
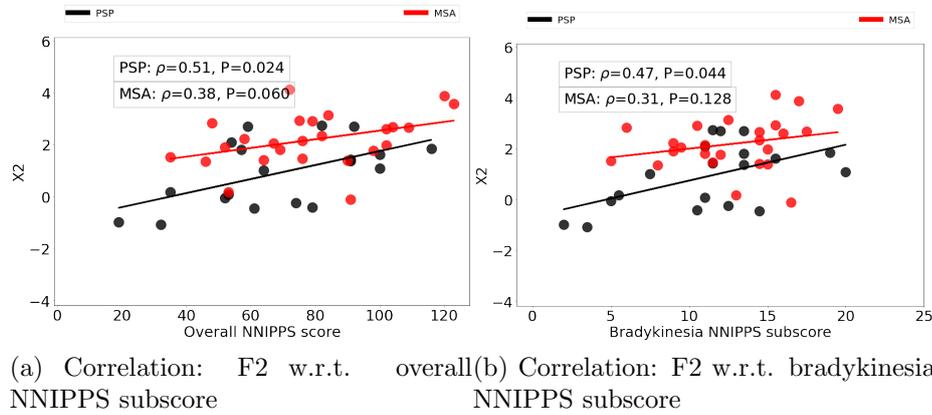


Figure 9.12: Feature X2 for the 4 groups



(a) Correlation: F2 w.r.t. overall NNIPPS subscore (b) Correlation: F2 w.r.t. bradykinesia NNIPPS subscore

Figure 9.13: Correlation of X2 w.r.t. overall and bradykinesia NNIPPS subscore

Thus dimension $X2$ can be considered as a gender independent feature at least at this stage. However, it is important to note that more number of male and female patients are warranted for confirming gender dimorphism.

Table 9.7 presents the percentage of affected patients in individual subsystem index for PD, PSP, and MSA. The finding reflects previous observations about individual subsystem index analysis. PSP patients manifest frequent deficit in respiration index (F_r), one phonation index (F_{p1}) compared to MSA whereas MSA patients showed frequent deficit in one phonation index (F_{p2}), and in articulation index (F_a). Both PSP and MSA patients showed frequent deficits in timing index (F_t), and diadochokinetic index (F_{dk}).

Finally, in 2-dimensional acoustic dimensions, PSP patients manifest abundant deficit in $X1$ whereas MSA patients showed predominant deficits in $X2$ as summarized in Table 9.8. In $X1 = F_r + 0.5 * DUS_{ml}$, the percentage of affected patients in PSP is distinctively high compared to PD. Inclusion of F_{p1} or F_{pr} in first dimension, percentage of affected patients in MSA increased from 32% to 52%. Thus discrimination of PSP and MSA patients could be challenging in modified $X1$.

Till now, two orthogonal speech dimensions are designed, which are easy to in-

Subsystem Index	PD (%)	PSP (%)	MSA (%)
F_r	Rare (10)	Frequent (57.89)	Occasional (16)
F_{p1}	Rare (10)	Abundant (73.68)	Frequent (52)
F_{p2}	Rare (5)	Common (42.11)	Abundant (72)
F_a	Occasional (20)	Rare (0)	Occasional (24)
F_t	Rare (5)	Frequent (52.63)	Frequent (56)
F_{pr}	Common (40)	Common (57.89)	Frequent (72)
F_{ddk}	Rare (10)	Abundant (89.47)	Abundant (76)

Table 9.7: Impairment for PD, PSP, and MSA groups in Subsystem index; The parentheses represent percentage of affected persons according to specific speech dimension: 0–10 % subjects affected are considered rare, 11–25 % occasional, 26–45 % common, 46–70 % frequent, and 71–100 % abundant

Combination of Subsystems	PD (%)	PSP (%)	MSA (%)
$X1 = F_r + 0.5 * DUS_{ml}$	Occasional (10)	Abundant (73.68)	common (32)
$X1 = F_r + 0.5 * DUS_{ml} + 0.5 * F_{p1}$	Occasional (10)	Abundant (78.94)	Frequent (52)
$X1 = F_r + 0.5 * DUS_{ml} + 0.5 * F_{pr}$	Occasional (20)	Abundant (78.94)	Frequent (52)
$X2 = F_a + F_{p2}$	Occasional (15)	Common (31.57)	Abundant (72)
$X2 = F_a + F_{p2} + F_t$	Occasional (15)	Common (36.84)	Abundant (80)

Table 9.8: Impairment for PD, PSP, and MSA groups in combination of subsystem indexes; The parentheses represent percentage of affected persons according to specific speech dimension: 0–10 % subjects affected are considered rare, 11–25 % occasional, 26–45 % common, 46–70 % frequent, and 71–100 % abundant

interpret and have physiological relation. The first dimension ($X1$) shows gender dimorphism in HC and PSP whereas dimension ($X2$) did not show gender difference. Next, these two features are used in 2D linear classifier to discriminate PSP and MSA patients.

Classification

Logistic Regression (LR) method is used for classification task. Given a small number of samples, Leave-one-sample-out (LOSO) method is adopted for training/test. Individual subsystem indexes are first used for classification of PSP and MSA. Table 9.9 presents classification scores with individual subsystem dimensions. Classification experiment shows that respiration, phonation, and articulation dimensions would be appropriate for discrimination of PSP and MSA.

Next, designated two speech dimensions ($X1$, $X2$) are used as input for classification. It yields accuracy 84.84%, specificity 84.21, sensitivity 84% which can be considered good accuracy, while the discrimination of PSP and MSA is even perceptually challenging. Table 9.10 provides classification scores between PSP and MSA.

Figure 9.14 displays the grouping of healthy subjects and patients in 2D space. Figure shows that PD, PSP and MSA patients are clustered in 2D space. While HC

Feature	Threshold	Accuracy (%)	Sensitivity (%)	Specificity (%)
F_r	0.5	70.45	57.89	80
F_{p1}	0.5	61.36	31.58	84
F_{p2}	0.6	61.36	57.89	64
F_a	0.6	72.72	73.68	72
F_t	0.45	56.82	0.0	100
F_{pr}	0.45	56.82	0.0	100
F_{ddk}	0.45	63.63	15.78	100

Table 9.9: Classification score individual subsystem dimension

Feature	Threshold	Accuracy (%)	Sensitivity (%)	Specificity (%)
$X1 = F_r + \frac{DUS_{ml}}{2}$	0.55	72.72	63.15	80
$X1 = F_r + \frac{DUS_{ml}}{2} + \frac{F_{p1}}{2}$	0.4	70.45	36.84	96
$X1 = F_r + \frac{DUS_{ml}}{2} + \frac{F_{pr}}{2}$	0.5	68.18	47.36	84
$X2 = F_a + F_{p2}$	0.45	70.45	42.10	92.0
$X2 = F_a + F_{p2} + F_t$	0.5	70.45	42.10	92.0
$(X1 = F_r + \frac{DUS_{ml}}{2}, X2 = F_a + F_{p2} + F_t)$	0.55	84.04	84.21	84

Table 9.10: Classification accuracy using X1 and X2 as 1-dimension and 2-dimensions for PSP and MSA

subjects present low quantitative value, PSP (resp. MSA) patients show high value in X1 (resp. X2) dimension.

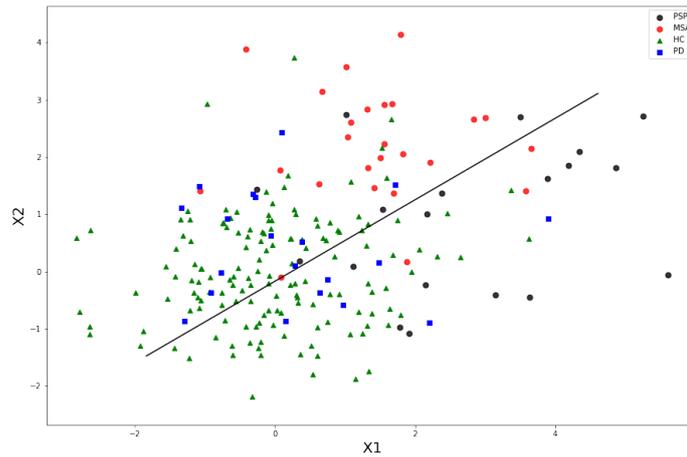


Figure 9.14: Biplot using X1 and X2 dimensions for four groups

In the next step, prosodic index (F_{pr}) is included in the first dimension, X1. The classification score remain same as 84.84%. However, inclusion of first phonation index in first dimension yields reduced accuracy 81.81%. It suggests that F_{p1} did not provide additional discriminative information in dimension X1.

Discussion on speech features by subsystems

The current quantitative study attempted to design a single-dimensional index to measure particular deficits in specific subsystems of speech production (respiration, phonation, articulation, prosody, timing, nasal, and diadochokinetic). Respiration index came up as an encouraging measure where PSP manifests major disorder compared to MSA. It is known that the respiratory system (mainly lungs) serves as the energy source (airflow) for speech production. Amount of air, rhythm, and rate of respiration are directly controlled by several neurons from the Central Nervous System (CNS). Several widely distributed, bilaterally located groups of neurons in the medulla and pons mostly control automatic and rhythmic respiration [98, 69]. In addition, the third, fourth, and fifth cervical segments of the spinal cord, where damage can paralyze the diaphragm bilaterally and seriously affect breathing. Damage to areas mentioned above can produce severe respiratory abnormalities, which also compel to compromise other subsystems accordingly.

Rigidity and bradykinesia in rib cage muscle can reduce the excursion for PD patients [286]. However, In four respiration features, PD patients did not show group differences with HC. Central hypoventilation is more commonly described in the synucleinopathies, particularly MSA, in which degeneration of the pontomedullary autonomic respiratory center is thought to result in diminished response to hypercapnia or hypoxemia. Conversely, hypoventilation is described infrequently in patients with tauopathies [205, 213], but the study found hypoventilation for a PSP patient [163]. Reduced vital capacity can drive to increase the rate of inscription and also to maintain speech rate, speakers can compromise the number of pauses within the breath cycle. PSP patients present predominant respiration deficits by RSR. Both PSP and MSA patients manifest predominant impairment to produce a sufficient number of pauses within respiration. Inspiratory stridor (a strained, high-pitched, harsh respiratory sound) was found to be common (9–69 %) for MSA patients [317, 315, 323]. Laryngeal adduction can be another possible reason for Inspiratory stridor [54]. In contrast to clinical findings, acoustic feature RLR provides decreased inspiration noise in MSA and PSP patients than HC and PD, probably due to the decreased inspiratory effort. In addition, the designed respiration index exhibited significant correlation with bulbar/pseudobulbar NNIPPS subscore. Thus it could be hypothesized that degeneration in bulbar/pseudobulbar may lead to respiratory impairment in speech production [100].

Laryngeal deficits are another most common pathology for parkinsonian disorder. Both PSP and MSA patients showed prominent deficits in phonation acoustic measures (GVI, DVI, $stdF0_a$, shimmer, HNR). Impairment in GVI, jitter, and $stdF0_a$ can be explained by laryngeal deficits mostly attributed to paralysis or atrophy of vocal folds abductor [128], originating from either a selective loss of abductor motor neurons in the nucleus ambiguus [19, 128] or depletion of medullary serotonergic projection, exerting tonic drive to abductor motor neurons [294]. Dystonia can be another possible cause of laryngeal pathology [196]. The current study found that PD, along with PSP and MSA patients manifest insufficient prosody (particularly pitch variation) mostly due to reduced range of movement of laryngeal functions, which is also

a distinct property of hypokinetic dysarthria. A previous study showed that MSA patients exhibit impaired motion of the vocal cord that occurs early in the disease process, accounting for early development of laryngeal symptoms such as dysphonia, stridor, and sleep apnea [169]. Conversely, PD patients manifest high pitch at the later stage of the disease [117]. In phonation features (F_{p2}), both PSP and MSA patients manifest prevalent laryngeal pathology compared to HC and PD, particularly in pitch variation. More precisely, MSA patients exhibit higher deficits than other groups, probably due to cerebellar deficits. In other phonation dimensions, features related to control, coordination, and reduced range of movement of laryngeal muscles for which PSP is expected/supposed to show higher impairment. DVI (monologue, text) captures control of the laryngeal muscles, and coordination of the laryngeal and supra-laryngeal muscles may manifest via voicing that interferes or continues within voiceless intervals, including unvoiced speech or pauses. Shimmer captures the perturbation of amplitude during sustained vowel phonation. A group of PSP patients shows higher shimmer compared to MSA, probably due to in-coordination with respiration. HNR may capture the reduced range of movement of laryngeal muscles. Most PSP patients introduced more noise due to incomplete closure of vocal folds. All these observations help to make the hypothesis that a particular class of laryngeal deficits is prevalent for PD, PSP and MSA patients.

Articulatory impreciseness is perceptually detected by several studies [47, 55, 106, 182]. Objective analysis also detected imprecise articulation in studies [312, 306, 114, 115]. In agreement with previous studies [306], PSP and MSA patients showed prolonged VOT and DUS for unvoiced stop consonants from speech task /pa-ta-ka/ and connected speech respectively. It can be explained by high dysarthria severity in the APS group. Articulatory decay and instability are captured by acoustic feature RFA and stdPSD where MSA patients showed greater severity compared to PD and PSP.

Timing of speech is disrupted by the deficits of the basal ganglia circuit, which lead to hypokinetic dysarthria. Predominant speech-timing disturbances were observed in the APS (PSP and MSA) group compared to PD and HC.

Designed two-dimensional subsystem features improved the discrimination between PSP and MSA. By this approach, subsystem specific impairment reveals encouraging pathophysiology of disease groups. $F1$ captures particularly respiration, articulation, and monopitch impairments. These impairments are more predominant for PSP patients. $X2$ captures a particular impairment in articulation (reduced range of movement and stability), phonation (laryngeal abnormality) and timing subsystems. These impairments are more predominant for MSA patients. Together, these 2 dimensions allow to discriminate between PSP and MSA patients with a good accuracy, sensitivity, and specificity, 84%. The methodology and result are promising for PD vs APS discrimination (particularly PD vs MSA which is the concern of the project, Voice4PD-MSA). In the following section, these results are used to proceed with designing dysarthria-type measures.

9.3.3 Speech features by dysarthria

Dysarthria type analysis may elucidate impairments more efficiently for voice pathologist. To do so, features used in X1 and X2 are regrouped according to their dysarthria type. X1 consists in only hypokinetic features. Considering only monopitch speech parameter from prosodic subsystem in X1 yield group difference between PD and HC. Thus, it can be considered as an hypokinetic measure and rename it H_1 . X2 consists of an hypokinetic feature H_2 and an ataxic one A_2 .

Hypokinetic measure (H_1)

First hypokinetic measure (H_1) is defined by the Equation 9.10. Now onward, H_1 will be used as one of the hypokinetic measures.

$$H_1 = F_r + \frac{DUS_{ml}}{2} + \frac{F_{pr}}{2} \quad (9.10)$$

Figure 9.15 illustrates group wise statistical difference. 7 (out of 19) PSP patients show clear discrimination in this hypokinetic index from MSA.

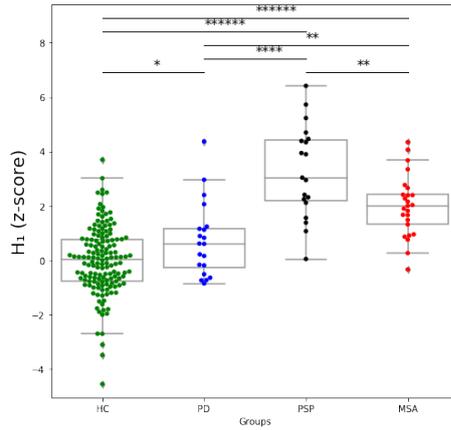


Figure 9.15: Hypokinetic feature H_1 of 4 groups

Hypokinetic measure (H_2)

The articulation, phonation and timing features from hypokinetic dysarthria are grouped as follows:

- Articulation: $H_a = \frac{RFA_{ml} + RFA_{txt}}{2}$
- Phonation: $H_p = \frac{Jitter + GVI_{ml} + GVI_{txt}}{3}$
- Timing: $H_t = \frac{EST_{ml} + RST_{ml} + AST_{txt}}{3}$

Next, second hypokinetic measure is defined as in Equation 9.11.

$$H_2 = H_a + \frac{H_p}{2} + \frac{H_t}{2} \quad (9.11)$$

Figure 9.16 illustrates group wise statistical difference. In this dimension 9 (out of 25) MSA patients show clear discrimination from PSP. In addition, PD patients exhibit statistical group difference w.r.t. HC.

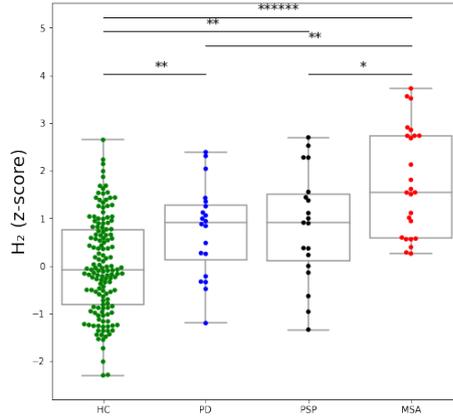


Figure 9.16: Hypokinetic feature H_2 of 4 groups

Correlation with H_1 and H_2 : Hypokinetic index H_1 show correlation with only bulbar NNIPPS subscores for PSP patients ($r = 0.44, p = 0.06$). On the other hand, hypokinetic index, H_2 showed correlation with overall severity ($r = 0.42, p = 0.07$), bradykinesia ($r = 0.4, p = 0.09$) NNIPPS subscore for PSP patients and with rigidity ($r = 0.42, p = 0.03$) NNIPPS subscore for MSA patients.

Gender dimorphism in H_1 and H_2 : First hypokinetic index, H_1 showed gender difference in HC ($p = 0.03$), PSP ($p = 0.05$), but not for MSA ($p = 0.11$) and PD group ($p = 0.18$). On the other hand, second hypokinetic index, H_2 manifest gender independence for HC ($p = 0.38$), PD ($p = 0.75$), PSP ($p = 0.08$), and MSA ($p = 0.25$) groups. Important to note, PSP female patients manifest greater deficits compared to PSP male patients in H_2 index.

Ataxic feature

In the ataxic dysarthria, two groups are found where PSP and MSA manifest orthogonal properties. First group consists features from oral diadochokinetic (DDK) recording protocol. Ataxic feature (A_1) is designed from DDK as:

$$A_1 = \frac{VD + DDKI}{2}$$

. In ataxic feature A_1 , PSP and MSA patients are not statistically separable. However, a small group of PSP patients (7 PSP patients) show distinctively higher severity as in Figure 9.17. Individual ataxic feature analysis from this dimension reveals that

5 PSP patients in VD and 3 PSP patients in DDKI manifest distinctly higher impairment compared to MSA.

Another ataxic feature is designed from phonation task as:

$$A_2 = \frac{stdF0_a + stdPSD}{2}$$

. Figure 9.17 illustrates that MSA patients manifest greater impairment in the ataxic feature A_2 . Interesting to observe that in this dimension MSA patients uniquely exhibit ataxia.

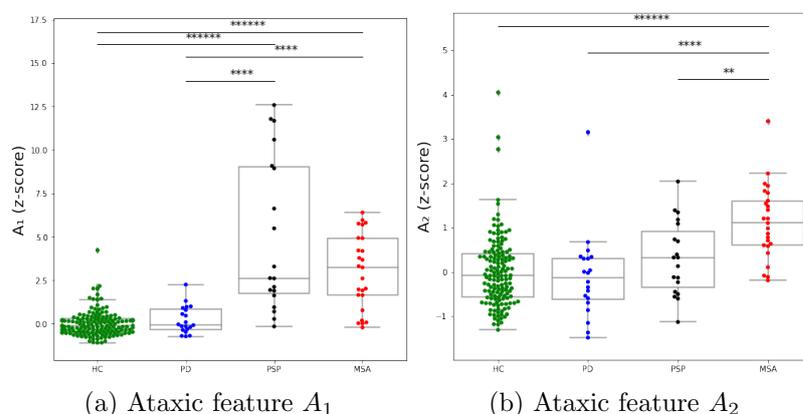


Figure 9.17: Group difference of HC, PD, PSP and MSA by ataxic features A_1 and A_2

Correlation with A_1 and A_2 : The cross-correlation between ataxic indices and NNIPPS subscores yield encouraging result. Ataxic dimension, A_1 showed significant correlation with bradykinesia ($r = 0.47, p = 0.01$), cerebellar ($r = 0.39, p = 0.05$) NNIPPS subscore for MSA patients and with oculomotor NNIPPS subscore ($r = 0.67, p = 0.001$) for PSP patients. Other ataxic index, A_2 showed correlation with overall severity ($r = 0.51, p = 0.02$), bradykinesia ($r = 0.49, p = 0.03$), bulbar/pseudobulbar ($r = 0.47, p = 0.04$) NNIPPS subscore for PSP patients. MSA patients did not show significant correlation for A_2 with NNIPPS subscore rather MSA patients mostly manifest higher disorder in A_2 index.

Gender dimorphism in A_1 and A_2 : Ataxic index, A_1 did not show gender difference in any of the groups such as HC ($p = 0.5$), PD ($p = 0.33$), PSP ($p = 0.83$), and MSA ($p = 0.06$). In another ataxic index, A_2 , did not show gender dimorphism in HC ($p = 0.42$), PD ($p = 0.14$), PSP ($p = 0.76$), and MSA ($p = 0.52$). Notably, male MSA patients manifest greater impairment than female MSA patients in A_1 .

Spastic feature (S)

Four spastic features are available in the feature set. DDKR and NSR measure a similar aspect of speech production, like syllable rate, in two different recording tasks. Inspired by latter observation, two rate-related features are averaged first and

then combined with other spastic feature generate the spastic feature. The spastic feature (S) is defined as:

$$S = \frac{\mu(DDKR,NSR) + PSI + Efn_m}{3}$$

. Figure 9.18 illustrates that both PSP and MSA group manifest predominant spasticity compared to HC and PD. However, 4 PSP patients exhibit distinctively higher impairment compared to MSA.

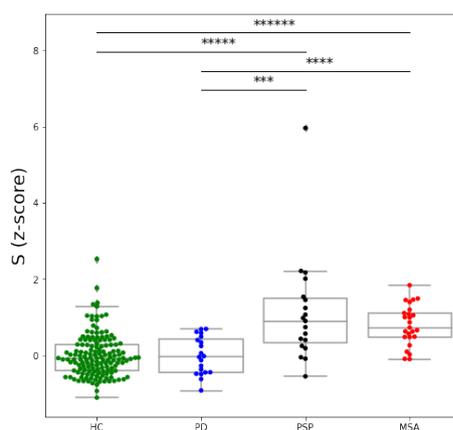


Figure 9.18: Spastic feature S of 4 groups

Correlation with S : Spastic index (S) show correlation with bradykinesia, rigidity, bulbar/pseudobulbar, and overall NNIPPS subscore for MSA patients as provided in Table 9.11. PSP patients showed correlation of index S with only oculomotor NNIPPS subscore ($r = 0.44, p = 0.06$).

Disease group	NNIPPS subscore	Bradykinesia	Rigidity	Cerebellar	Bulbar / Pseudobulbar	Overall
MSA	Spastic Index (S)	0.4*	0.46*	0.25	0.48*	0.59**

Table 9.11: The cross-correlations between dysarthric and clinical motor indices: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Gender dimorphism in S : In spastic index (S), gender dimorphism is observed in HC ($p = 0.0008$) and PSP ($p = 0.049$) disease groups whereas PD and MSA patients did not manifest gender difference. Interestingly, female HC and PSP patients showed comparably higher spasticity compared to males.

Table 9.12 summarizes the percentage of affected patients in individual dysarthria index. In the hypokinetic index (H_1), the PSP group showed abundant deficits, whereas the MSA group manifested frequent deficits. In the hypokinetic index (H_2),

MSA patients showed frequent deficits. Both PSP and MSA groups showed prominent deficits in ataxic index (A_1). MSA patients showed frequent deficits in another ataxic index (A_2). In agreement with previous studies [247], PD patients only manifest occasional hypokinesia.

Dysarthria Index	PD (%)	PSP (%)	MSA (%)
H_1	Occasional (20)	Abundant (78.95)	Frequent (52)
H_2	Occasional (15)	Occasional (21.05)	Frequent (48)
A_1	Rare (5)	Abundant (78.95)	Abundant (76)
A_2	Rare (5)	Occasional (21.05)	Frequent (48)
S	Rare (0)	Common (42.10)	Common (36)

Table 9.12: Impairment for PD, PSP, and MSA groups in dysarthria index; The parentheses represent percentage of affected persons according to specific speech dimension: 0–10 % subjects affected are considered rare, 11–25 % occasional, 26–45 % common, 46–70 % frequent, and 71–100 % abundant

Table 9.13 illustrates percentage of affected patients in combination of dysarthric index. In $D2$, most of the MSA patients manifest speech disorder whereas in $D1$ and $D1_{H_1S}$ both PSP and MSA patients show a combination of dysarthria.

Combined dysarthria Index	PD (%)	PSP (%)	MSA (%)
$D1_{H_1S} = H_1 + \frac{S}{2}$	Occasional (20)	Abundant (84.21)	Abundant (72)
$D1 = H_1 + \frac{S}{2} + \frac{A_1}{2}$	Occasional (15)	Abundant (84.21)	Abundant (76)
$D2 = H_2 + A_2$	Occasional (20)	Common (42.10)	Abundant (84)

Table 9.13: Impairment for PD, PSP, and MSA groups in dysarthria index; The parentheses represent percentage of affected persons according to specific speech dimension: 0–10 % subjects affected are considered rare, 11–25 % occasional, 26–45 % common, 46–70 % frequent, and 71–100 % abundant

Though individual dysarthric indexes provide encouraging group differences between PSP and MSA, it is time to check classification scores. If the classification score is not satisfactory, combination of dysarthric indexes may yield improved discrimination between PSP and MSA as presented in Table 9.12.

Classification with dysarthria features

Naturally, individual dysarthria feature dimensions are used first for classification using logistic regression. It is clear from the classification score provided in Table 9.14 that individual dimensions are not sufficient for good classification. However, H_1 and A_2 yield an indication to use in two separate dimensions.

As decided previously that two speech dimensions may be suitable for classification, two speech dimensions are designed with dysarthria dimensions. First speech

Feature	Threshold	Accuracy (%)	Sensitivity (%)	Specificity (%)
H_1	0.55	68.18	57.89	76
H_2	0.45	63.63	26.32	92
A_1	0.45	68.18	31.58	96
A_2	0.55	72.73	63.16	80
S	0.45	56.82	0.0	100.0

Table 9.14: Classification accuracy using individual dysarthria indexes

dimension (D1) is defined as:

$$D1_{H_1S} = H_1 + \frac{S}{2}$$

. Prevalence of PSP patients in hypokinetic and spastic dysarthria motivates to combine H_1 and S indexes. PSP patients show higher impairment compared to other groups in hypokinetic and spastic dysarthria. Other speech dimension (D2) is defined as:

$$D2 = H_2 + A_2$$

MSA patients show greater severity in hypokinetic and ataxic dysarthria. Figure 9.19 displays boxplot of two speech dimensions.

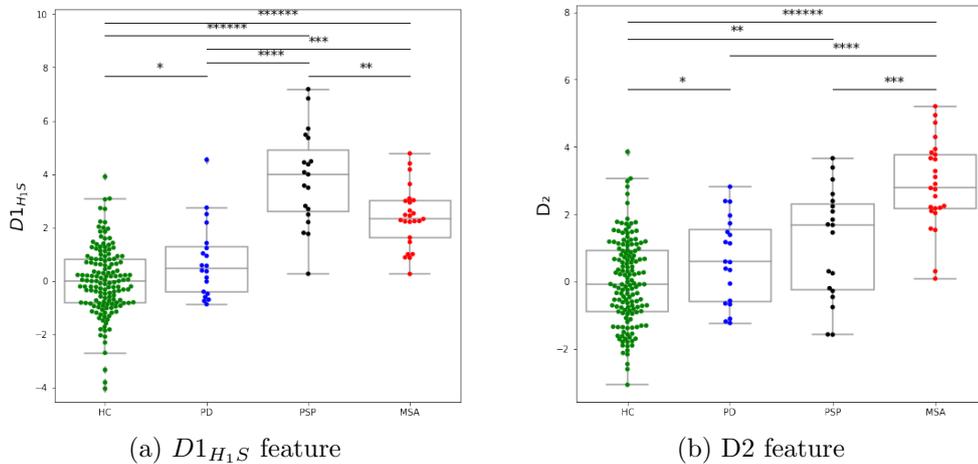


Figure 9.19: Group difference of four groups using feature $D1_{H_1S}$ and D2

Table 9.15 presents classification accuracy by combining dysarthric dimensions $D1_{H_1S}$ and D_2 . Although accuracy is little improved, sensitivity is low as many PSP patients are misclassified as MSA. Next, two dimensions are used in logistic regression, which yields a comparably good classification score.

Classification by 2-dysarthria dimensions yields good accuracy, 86.36%. In agreement with previous studies, PSP patients manifest predominant hypokinetic and

Feature	Threshold	Accuracy (%)	Sensitivity (%)	Specificity (%)
$D1_{H_1S} = H_1 + \frac{S}{2}$	0.5	72.72	57.89	84.0
$D_2 = H_2 + A_2$	0.45	70.45	42.11	92.0
$(D1_{H_1S}, D2)$	0.5	86.36	84.21	88

Table 9.15: Classification accuracy using $D1_{H_1S}$ and $D2$ as 1-dimension or 2-dimensions input

spastic dysarthria whereas MSA patients are characterized by ataxic and hypokinetic dysarthria. Important to note, PD patients only manifest hypokinesia by H_1 and H_2 .

Designed dimension, $D1$ did not show gender difference in HC ($p = 0.12$), PD ($p = 0.15$), and PSP ($p = 0.14$) whereas MSA ($p = 0.04$) disease group manifest gender dimorphism. On the other hand, dimension $D2$ did not show gender dimorphism in HC ($p = 0.23$), PD ($p = 0.10$), PSP ($p = 0.16$), and MSA ($p = 0.09$). Figure ?? illustrates that male MSA patients manifest higher deficits compared to female MSA whereas both male and female PSP exhibit comparably high deficits in $D1$. Thus it can be hypothesized that classification of male PSP and MSA would be more challenging.

Now, it is time to decide about the ataxic dimension A_1 . Interestingly, inclusion of ataxic feature A_1 in the first dimension further improves the classification score for PSP and MSA. Modified $D1$ is defined as:

$$D1 = H_1 + \frac{A_1}{2} + \frac{S}{2}$$

. Classification accuracy improved little in $D1$ but half of the PSP patients are misclassified. $D2$ dimension is kept unchanged with A_2 and H_2 . Table 9.16 shows that 2D features yield very good classification accuracy 88.63 %.

Feature	Threshold	Accuracy (%)	Sensitivity (%)	Specificity (%)
$D_1 = H_1 + \frac{A_1}{2} + \frac{S}{2}$	0.45	75	47.37	96
$(D1, D2)$	0.6	88.63	89.47	88

Table 9.16: Classification accuracy using $D1$ as 1-dimension input and $D1, D2$ as 2-dimensions input

Figure 9.20 presents the boxplot of modified dimension $D1$ and 2D plot of $D1$ and $D2$.

In the modified first dimension ($D1$), gender dimorphism is observed in HC ($p = 0.04$), PD ($p = 0.05$), and MSA ($p = 0.01$) groups whereas PSP ($p = 0.41$) did not show gender difference.

In addition, proposed dysarthria features are used to classify PD and PSP; and PD and MSA. Feature $D1$ consists of H_1, A_1 , and S can discriminate between PD and PSP by 87%. On the other hand, both $D1$ and $D2$ discriminate between PD and MSA

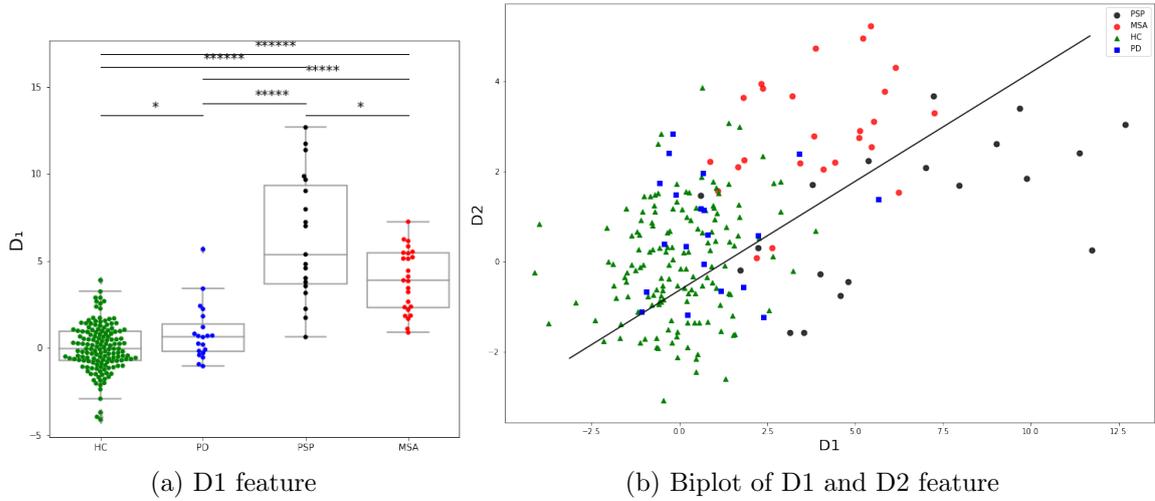


Figure 9.20: Group difference of four groups using the modified feature D1; Biplot by D1 and D2

by 87%. It suggests that designed dimensions would be helpful for discrimination of PD and MSA by good accuracy.

Discussion on speech features by dysarthria

In both hypokinetic indexes (H_1, H_2), PD, PSP, and MSA patients exhibit pronounced hypokinesia. It can be justified by degeneration in basal ganglia circuits. PD patients manifest less severity compared to PSP and MSA in hypokinetic indexes, probably due to slow disease progression of PD [14]. Furthermore, PD patients did not exhibit ataxia and spasticity, which was also stated in previous studies [247]. In overall, PSP and MSA patients manifest hypokinetic, ataxic and spastic dysarthria. However, the present study reveals that PSP and MSA patients predominate in particular types of hypokinesia and ataxia.

In differential aspect, PSP patients mostly presents particular type of hypokinetic dysarthria (H_1), spastic (S) and ataxic (A_1) components. Only a small group of PSP exhibit comparably higher spasticity and ataxia from MSA in S and A_1 . The latter observation is consistent with widespread neurodegeneration involving the midbrain as well as the globus pallidus, striatum, hypothalamic nucleus, pons, superior cerebellar peduncle and cerebellar dentate nucleus [202]. Relation with neuronal loss and gliosis in the substantia nigra with severity of hypokinetic dysarthria was confirmed by the study [159]. Conversely, MSA patients manifest predominant ataxia plus a particular type of hypokinesia compared to PSP. The latter observation can be attributed to degeneration of cerebellum, middle cerebellar peduncle, striatum, substantia nigra, inferior olivary nucleus, and pons [89]. In addition, a relationship with spastic dysarthria and bulbar/pseudobulbar severity was established in the studies [247, 255]. Relation with ataxic dysarthria with cerebellar severity was also confirmed in studies [247, 255]. Present study observed correlation between H_1 w.r.t. bulbar NNIPPS

subscore and between H_2 w.r.t. bradykinesia for PSP patients. MSA patients showed correlation between H_2 w.r.t. rigidity NNIPPS subscore. It confirms that both PSP and MSA exhibit different degrees of hypokinesia and spasticity. The present study also observed significant correlation with ataxic index A_1 w.r.t. cerebellar NNIPPS subscore for MSA patients. Both ataxic index (A_1 and A_2) yield correlation with bradykinesia NNIPPS subscore for MSA patients. As MSA groups consist of both parkinsonian and cerebellar subtypes, correlation with bradykinesia and cerebellar is logical. In addition, it will help to localize underlying pathophysiology.

PSP patients showed mostly respiration deficits (primarily due to bradykinesia, rigidity), prolonged phoneme (probably due to weak movements, degeneration in cerebellar forced to rely on basal ganglia circuit for precise movements). Conversely, MSA patients showed mostly irregularities in vocal folds vibration, repetitive articulators movements and instability/desynchronization of articulatory movements.

First hypokinetic index (H_1) manifested gender dimorphism in HC, but H_2 did not show gender dimorphism. Ataxic indexes did not show gender difference in groups. However, spastic index yield significant gender differences in HC and PSP. Overall male subjects manifest greater severity compared to female subjects. It could be hypothesized that in the patient group male patients would exhibit high speech disorder compared to females. It also suggests devising gender specific speech markers for better diagnosis. However, small data size is still an impediment towards these gender-specific speech indexes.

Chapter 10

Conclusions and future work

This chapter summarizes the contributions towards differential diagnosis of PSP and MSA; and PD and MSA-P groups by voice analysis. Next, future works are described towards accurate speech parameter computation and early differential diagnosis.

10.1 Summary and discussion

This thesis presented a set of speech features categorically to assess imprecise vowel, consonants production and spontaneous speech for parkinsonian disorder. As this thesis aims to design methodology for differential diagnosis of PD, PSP, and MSA, speech parameters are considered which atleast show distinctive properties for disease groups. Reliable speech markers for differential diagnosis is indeed lacking at present time. Present study also analyzed gender dimorphism of each speech feature which would facilitate taking final conclusion about disease specificity. Another advantage of the present study is that speech features are designed by knowledge-driven approach rather than recent data driven approach which would be more acceptable in clinical practice.

Vowel distortion was frequently studied in speech disorder by mostly conventional features. Jitter, shimmer, and HNR computed by Praat tool [35] manifested insensitiveness to mild changes whereas jitter and shimmer computed by fundamental frequency (F0) showed much more sensitivity to mild distortion [214]. Speech features like jitter, Quasi-open Quotient (QOQ), derivative of F0, and vocal tremor individually show a trend of disease specificity. In agreement with previous studies [169, 128, 19], present study also observed variability in vocal folds vibration is predominated for MSA patients. Latter studies also find the laryngeal deficits for MSA patients even in early stages. Notably, variability in vocal folds vibration also in other disease groups like PSP, HD, MS, and CA [247, 112]. Combination of those phonation features using knowledge driven weight yields improved discrimination between PD and MSA-P. Involuntary movements of articulators (tongue, lips, and jaw) in sustained vowel /a/ measured by stdPSD and stdLogE also predominate in MSA-P patients and attributed to hyperkinetic and/or ataxic dysarthria [69]. Deficits in velopharyngeal function also predominate in MSA-P patients. Hypernasality in

MSA-P patients is pronounced compared to PD. Variability in nasality was more frequent in MSA-P patients which could be attributed to involuntary movements of soft palate which was attributed to cerebellar dysfunction [68]. Notably, hypernasality also observed for HD and CA disease groups [112]. Speech index is designed to capture involuntary movements of articulators and soft palate yields encouraging disease specificity for MSA-P. Though MSA patients are parkinsonian type, predominance of ataxic dysarthria could be explained by even mild cerebellar deficits [247]. Two indexes representing variability in vocal folds vibration and involuntary movements of articulators able to discriminate PD and MSA-P by very high accuracy. Reduced range of movement due to hypokinesia manifested in MSA-P patients by vowel space area analysis. To the best of our knowledge, change in formant frequencies was not studied in differential diagnosis [283, 249, 310]. However, only one study [247] mention imprecise vowel by VAI computed from monologue where PSP patients show greater impairment compared to PD. We also developed an automatic method to segment vowel and unvoiced stop plosives from diadochokinetic task (/pa-ta-ka/). Standard acoustic features from these speech segments provided encouraging discrimination between PD and MSA-P. MSA-P patients manifest greater disorder compared to PD which is in accordance with previous studies. Combination of two DDD features, represent ataxic dysarthria yield high classification between PD and MSA-P. In addition, analysis of reading text also provided distinctive disorder in phonation and articulation features. Manual segmentation (voiced, unvoiced, pause, and respiration) of reading text reflects accurate speech disorder. In overall, in prosodic features both PD and MSA-P patients exhibited disorder compared to HC.

This thesis also conducted an exploratory investigation of imprecise consonants from word initial consonants. In subjective analysis, in agreement with previous study [182] frequent distortion is observed in stop plosives and fricatives. The thesis also proposed automatic algorithm to detect abnormalities in stop plosives and fricatives. Present study objectively measured devoicing in voiced obstruents and predominates in MSA-P compared to PD. To the best of our knowledge, devoicing never been investigated for MSA patients. Presence of devoicing mostly attributed to cerebellar dysfunction. Frequent devoicing observed in velar voiced stop /g/. Furthermore, devoicing is frequently observed in consonant-consonant (/gR/) syllable for MSA patients which indicate importance of speech task. In unvoiced consonants, three speech parameters like multiple burst and weak burst in stop plosives and presence of burst in fricatives came up as prominent distortion. Among these speech parameters, burst in unvoiced fricatives yield encouraging discrimination between PD and MSA-P. Burst in fricatives is the manifestation of overshooting (clinical term "dysmetria"). If it is confirmed, it could be attributed to lesions in the lateral and paravermal cerebellar hemispheres [69]. Reduced range of movement also manifested by weak burst in both PD and MSA.

Design of speech markers related to subsystem of speech production and subtypes of dysarthria towards discrimination of PSP and MSA provided encouraging novel result. In subsystem related speech index captures deficits in respiration, phonation, articulation, prosodic and timing aspect of speech production. In all these speech indexes both groups (PSP and MSA) manifest deficits in all subsystems. Speech dis-

order in particular subsystem is not disease specific, however degree of speech disorder can serve disease specificity. As example, degree of impairment in respiration system is severe for PSP patients compared to MSA. Likewise, in a phonation speech index, MSA patients manifest higher severity compared to PSP. The concept of designing speech index to capture overall abnormalities is an unique strategy adopted in this thesis. Moreover, in agreement with previous studies [247, 255], present endeavors to design dysarthria indexes also brought additional information about neurological diseases. Present study finds PSP and MSA patients manifest combination of hypokinetic, spastic, and ataxic dysarthrias which confirms widespread neurological lesions for APS groups. Notably, Subgroups of dysarthrias are observed in hypokinetic and ataxic dysarthria where either PSP or MSA predominates. As example, PSP patients manifest prolongation of duration and rate related hypokinetic components whereas MSA patients predominated in instability and variability related hypokinetic components. Likewise, a group of PSP patients manifest greater severity in prolongation of vowel and articulatory rhythm from ataxic components whereas MSA patients predominated in irregularity and instability related ataxic components.

Insights of gender dimorphism was explored in this study. Several speech components exhibited gender disparity for HC group which indicate to consider gender at the time of differential diagnosis. Notably, male subjects showed higher speech disorder compared to female subjects.

10.2 Limitation and future works

Speech data is an invaluable asset for analyzing speech disorder in neurological disorder. Voice4PD-MSA project faces several difficulties like administrative to pandemic situation to collect speech data from registered patients and control subjects. Therefore, more data is required to be recorded to validate initial findings about differential diagnosis and gender dimorphism. As gender dimorphism can play a crucial role in acoustic feature disparity, it demands gender balance in each group. Participant's age is the another important parameter related to speech disorder. Hence, analysis of these multidimensional investigation requires sufficient data. Till now, we could not accommodate clinical data for Voice4PD-MSA database. Particularly, clinical data is most essential to correlate w.r.t. speech markers which would explain underlying pathophysiology. In addition, inclusion of other parkinsonian diseases is also essential for true differential diagnosis.

Presently available methods to analyze vocal folds vibration were developed for healthy speakers. Thus those methods sometime fails to assess the deficits in vocal folds vibration. As example, there is no established method to measure pitch frequency (F0) in different condition. Likewise, formant frequency computation mostly fails for pathological speech due to imprecise movements of articulator. It is thus demands to at least mention weak formant.

Consonants sound units require further exploration to capture other aspect of abnormalities. Automatic segmentation of voiced consonant and following vowel is required to individually analyze voiced obstruents. Realization of voiced and unvoiced

consonants is thus further demand to explore other abnormalities.

Design accurate speech segmentation from continuous speech task is further essential for developing accurate speech dimensions. Detection of respiration is still not accurate. Thus, speech segmentation of voiced, unvoiced, pause, and respiration require more complex methods.

Publications

Conference

1. **Biswajit Das**, Khalid Daoudi, Jiri Klempir, and Jan Ruzs. "Towards disease-specific speech markers for differential diagnosis in Parkinsonism." In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5846-5850. IEEE, 2019.
2. Khalid Daoudi, **Biswajit Das**, Solange Milhé de Saint Victor, Alexandra Foubert-Samier, Anne Pavy-Le Traon, Olivier Rascol, Wassilios Meissner, and Virginie Woisard. "Distortion of voiced obstruents for differential diagnosis between parkinson's disease and multiple system atrophy." In INTERSPEECH 2021.

Bibliography

- [1] Daudet A. La chèvre de monsieur seguin. *In les Lettres de mon Moulin*, 1869.
- [2] J A Logemann and H B Fisher. Vocal tract control in parkinson's disease: Phonetic feature analysis of misarticulations. *The Journal of speech and hearing disorders*, 46:348–52, 12 1981.
- [3] JH Abbs, CiJ Hunker, and SM Barlow. Differential speech motor subsystem impairments with suprabulbar lesions: Neurophysiological framework and supporting data. *Clinical dysarthria*, pages 21–56, 1983.
- [4] A.S. Abramson and D.H. Whalen. Voice onset time (vot) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics*, 63:75 – 86, 2017.
- [5] Hermann Ackermann, Susanne Gräber, Ingo Hertrich, and Irene Daum. Phonemic vowel length contrasts in cerebellar disorders. *Brain and Language*, 67(2):95 – 109, 1999.
- [6] Hermann Ackermann and Ingo Hertrich. Voice onset time in ataxic dysarthria. *Brain and Language*, 56(3):321 – 333, 1997.
- [7] Hermann Ackermann, Jürgen Konczak, and Ingo Hertrich. The temporal control of repetitive articulatory movements in parkinson's disease. *Brain and language*, 56(2):312–319, 1997.
- [8] Hermann Ackermann and Wolfram Ziegler. Articulatory deficits in parkinsonian dysarthria: an acoustic analysis. *Journal of Neurology, Neurosurgery & Psychiatry*, 54(12):1093–1098, 1991.
- [9] Yossi Adi, Joseph Keshet, Olga Dmitrieva, and Matt Goldrick. Automatic measurement of voice onset time and prevoicing using recurrent neural networks. pages 3152–3155, 09 2016.
- [10] Ikuko Aiba, Yufuko Saito, Akiko Tamakoshi, and Yukihiro Matsuoka. Prognosis of patients with progressive supranuclear palsy. *Rinsho shinkeigaku = Clinical neurology*, 45(8):565—570, August 2005.

- [11] Jefferson S. Almeida, Pedro P. Rebouças Filho, Tiago Carneiro, Wei Wei, Robertas Damaševičius, Rytis Maskeliūnas, and Victor Hugo C. de Albuquerque. Detecting parkinson’s disease with sustained phonation and speech signals using machine learning techniques. *Pattern Recognition Letters*, 125:55–62, 2019.
- [12] T. V. Ananthapadmanabha, A. P. Prathosh, and A. G. Ramakrishnan. Detection of the closure-burst transitions of stops and affricates in continuous speech using the plosion index. *The Journal of the Acoustical Society of America*, 135(1):460–471, 2014.
- [13] G. Aneeja and B. Yegnanarayana. Single frequency filtering approach for discriminating speech and nonspeech. *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, 23(4):705–717, April 2015.
- [14] Schrag Anette, Wenning Gregor K., Quinn Niall, and Ben-Shlomo Yoav. Survival in multiple system atrophy. *Movement Disorders*, 23(2):294–296.
- [15] Tanja Kocjancic Antolík and Cécile Fougeron. Consonant distortions in dysarthria due to parkinson’s disease, amyotrophic lateral sclerosis and cerebellar ataxia. pages 2152–2156, 2013.
- [16] Patricia Argüello-Vélez, Tomás Arias-Vergara, María González-Rátiva, Juan Rafael Orozco, Elmar Noeth, and Maria Schuster. *Acoustic Characteristics of VOT in Plosive Consonants Produced by Parkinson’s Patients*, pages 303–311. 09 2020.
- [17] A.E. Aronson and J.R. Brown. *Motor speech disorders*. WB Saunders Company, 1975.
- [18] Young-Im Bang, Kyunghoon Min, Young Sohn, and Sung-Rae Cho. Acoustic characteristics of vowel sounds in patients with parkinson disease. *NeuroRehabilitation*, 32:649–54, 05 2013.
- [19] ROGER BANNISTER, WILLIAM GIBSON, LESLIE MICHAELS, and DAVID R. OPPENHEIMER. LARYNGEAL ABDUCTOR PARALYSIS IN MULTIPLE SYSTEM ATROPHY: A REPORT ON THREE NECROPSIED CASES, WITH OBSERVATIONS ON THE LARYNGEAL MUSCLES AND THE NUCLEI AMBIGUI. *Brain*, 104(2):351–368, 06 1981.
- [20] Gaetano Barbagallo, Maria Sierra-Pe, Federico Nemmi, Anne Pavy-Le, Wassilios Meissner, Olivier Rascol, and Patrice Péran. Multimodal mri assessment of nigro-striatal pathway in multiple system atrophy and parkinson disease. 31, 10 2015.
- [21] CL Barclay and AE Lang. Dystonia in progressive supranuclear palsy. *Journal of Neurology, Neurosurgery & Psychiatry*, 62(4):352–356, 1997.

- [22] Georg Becker, Antje Müller, Stefan Braune, Thomas Büttner, Reiner Be-
necke, Wolfgang Greulich, Wolfgang Klein, Günter Mark, Jürgen Rieke, and
Reiner Thümler. Early diagnosis of parkinson’s disease. *Journal of neurology*,
249(3):iii40–iii48, 2002.
- [23] Susan J. Behrens and Sheila E. Blumstein. Acoustic characteristics of english
voiceless fricatives: a descriptive analysis. *Journal of Phonetics*, 16(3):295–298,
1988.
- [24] E. Belalcázar-Bolaños, J. R. Orozco-Arroyave, J. Vargas-Bonilla, T. Haderlein,
and E. Nöth. Glottal flow patterns analyses for parkinson’s disease detection:
Acoustic and nonlinear approaches. In *TSD*, 2016.
- [25] Y. Ben-Shlomo, G. K. Wenning, F. Tison, and N. P. Quinn. Survival of patients
with pathologically proven multiple system atrophy. *Neurology*, 48(2):384–393,
1997.
- [26] Visar Berisha, Steven Sandoval, Rene Utianski, Julie Liss, and Andreas Spanias.
Characterizing the distribution of the quadrilateral vowel space area. *The Jour-
nal of the Acoustical Society of America*, 135(1):421–427, 2014.
- [27] Visar Berisha, Steven Sandoval, Rene Utianski, Julie Liss, and Andreas Spanias.
Characterizing the distribution of the quadrilateral vowel space area. *The Jour-
nal of the Acoustical Society of America*, 135(1):421–427, 2014.
- [28] Valentina Berti, Alberto Pupi, and Lisa Mosconi. Pet/ct in diagnosis of move-
ment disorders. *Annals of the New York Academy of Sciences*, 1228:93–108, 06
2011.
- [29] Sam Birdi, Ali H. Rajput, Mark Fenton, Jeffery R. Donat, Bohdan Rozdilsky,
Christopher Robinson, Rob Macaulay, and David George. Progressive supranu-
clear palsy diagnosis and confounding features: Report on 16 autopsied cases.
Movement Disorders, 17(6):1255–1264, 2002.
- [30] N. N. Bitar. *Ph.D. dissertation, Boston University, Massachusetts*, 1998.
- [31] Michael Blomgren, Michael Robb, and Yang Chen. A note on vowel centraliza-
tion in stuttering and nonstuttering individuals. *Journal of speech, language,
and hearing research : JSLHR*, 41:1042–51, 11 1998.
- [32] Sheila E. Blumstein and Kenneth N. Stevens. Acoustic invariance in speech
production: Evidence from measurements of the spectral characteristics of stop
consonants. *The Journal of the Acoustical Society of America*, 66(4):1001–1017,
1979.
- [33] Tobias Bocklet, Elmar Nöth, Georg Stemmer, Hana Ruzickova, and Jan Ruzs.
Detection of persons with parkinson’s disease by acoustic, vocal, and prosodic
analysis. In *2011 IEEE workshop on automatic speech recognition & under-
standing*, pages 478–483. IEEE, 2011.

- [34] Tobias Bocklet, Stefan Steidl, Elmar Nöth, and Sabine Skodda. Automatic evaluation of parkinson’s speech-acoustic, prosodic and voice related cues. In *Interspeech*, pages 1149–1153, 2013.
- [35] P. Boersma and D. Weenink. Praat: doing phonetics by computer [computer program]. *Version 5.3.51*, June 2013.
- [36] Paul Boersma. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *IFA Proceedings 17*, pages 97–110, 1993.
- [37] Carl R Boyd, Mary Ann Tolson, and Wayne S Copes. Evaluating trauma care: the triss method. trauma score and the injury severity score. *The Journal of trauma*, 27(4):370–378, 1987.
- [38] L. Brabenec, J. Mekyska, Z. Galaz, and I. Rektorova. Speech disorders in parkinson’s disease: early diagnostics and effects of medication and brain stimulation. *Journal of Neural Transmission*, March 2017.
- [39] S. Brognaux and T. Drugman. Hmm-based speech segmentation: Improvements of fully automatic approaches. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(1):5–15, 2016.
- [40] P Brown and CD Marsden. What do the basal ganglia do? *The Lancet*, 351(9118):1801–1804, 1998.
- [41] Steven Brown, Elton Ngan, and Mario Liotti. A larynx area in the human motor cortex. *Cerebral cortex*, 18(4):837–845, 2008.
- [42] M. Brückl. Vocal tremor measurement based on autocorrelation of contours. In *INTERSPEECH*, pages 715–718, 2012.
- [43] M. Brückl, A. Ghio, and F. Viallet. Measurement of tremor in the voices of speakers with parkinson’s disease. *Procedia Computer Science*, 128:47 – 54, 2018. 1st International Conference on Natural Language and Speech Processing.
- [44] Markus Brückl, Alain Ghio, and François Viallet. Measurement of tremor in the voices of speakers with parkinson’s disease. *Procedia Computer Science*, 128:47–54, 2018. 1st International Conference on Natural Language and Speech Processing.
- [45] Eugene Buder. *Acoustic Analysis of Voice Quality: A Tabulation of Algorithms 1902-1990*. 01 2000.
- [46] Giovanna Calandra-Buonaura, Andrea Doria, Giovanna Lopane, Pietro Guaraldi, Sabina Capellari, Paolo Martinelli, Pietro Cortelli, and Manuela Contin. Pharmacodynamics of a low subacute levodopa dose helps distinguish between multiple system atrophy with predominant parkinsonism and parkinson’s disease. *Journal of neurology*, 263(2):250–256, 2016.

- [47] N Chakraborty, T Roy, Avijit Hazra, Atanu Biswas, and K Bhattacharya. Dysarthric bengali speech: A neurolinguistic study. *Journal of postgraduate medicine*, 54:268–72, 10 2008.
- [48] Helen Chen and Kenneth N. Stevens. An acoustical study of the fricative /s/ in the speech of individuals with dysarthria. *Journal of Speech, Language, and Hearing Research*, 44(6):1300–1314, 2001.
- [49] Helen J Chenery, Bruce E Murdoch, and John CL Ingram. Studies in parkinson’s disease: I. perceptual speech analyses. *Australian Journal of Human Communication Disorders*, 16(2):17–29, 1988.
- [50] Wang Zheng Chiu, L D Kaat, Harro Seelaar, Sonia M Rosso, Agnita JW Boon, Wouter Kamphorst, and John C van Swieten. Survival in progressive supranuclear palsy and frontotemporal dementia. *Journal of Neurology, Neurosurgery & Psychiatry*, 81(4):441–445, 2010.
- [51] Amy Colcher and Tanya Simuni. Clinical manifestations of parkinson’s disease. *Medical Clinics of North America*, 83(2):327–347, 1999.
- [52] Carlo Colosimo. Nonmotor presentations of multiple system atrophy. *Nature reviews. Neurology*, 7:295–8, 02 2011.
- [53] Nadine P Condor, Christy L Ludlow, and GERALYN M Schulz. Stop consonant production in isolated and repeated syllables in parkinson’s disease. *Neuropsychologia*, 27(6):829–838, 1989.
- [54] Pietro Cortelli, Giovanna Calandra-Buonaura, Eduardo E. Benarroch, Giulia Giannini, Alex Iranzo, Phillip A. Low, Paolo Martinelli, Federica Provini, Niall Quinn, Eduardo Tolosa, Gregor K. Wenning, Giovanni Abbruzzese, Pamela Bower, Enrico Alfonsi, Imad Ghorayeb, Tetsutaro Ozawa, Claudio Pacchetti, Nicolò Gabriele Pozzi, Claudio Vicini, Angelo Antonini, Kailash P. Bhatia, Jacopo Bonavita, Horacio Kaufmann, Maria Teresa Pellecchia, Nicole Pizzorni, Antonio Schindler, François Tison, Luca Vignatelli, and Wassilios G. Meissner. Stridor in multiple system atrophy. *Neurology*, 93(14):630–639, 2019.
- [55] FL Darley, AE Aronson, and JR Brown. Clusters of deviant speech dimensions in the dysarthrias. *Journal of speech and hearing research*, 12(3):462–496, September 1969.
- [56] Frederic L Darley, 1928-(joint author.) Aronson, Arnold E. (Arnold Elvin), and 1911-(joint author.) Brown, Joe Robert. *Motor speech disorders*. Philadelphia : Saunders, 1975. References interspersed.
- [57] Frederic L. Darley, Arnold E. Aronson, and Joe R. Brown. Clusters of deviant speech dimensions in the dysarthrias. *Journal of Speech and Hearing Research*, 12(3):462–496, 1969.

- [58] Frederic L. Darley, Arnold E. Aronson, and Joe R. Brown. Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research*, 12(2):246–269, 1969.
- [59] B. Das, K. Daoudi, J. Klempír, and J. Ruzs. Towards disease-specific speech markers for differential diagnosis in parkinsonism. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2019, Brighton, United Kingdom, May 12-17, 2019*, pages 5846–5850. IEEE, 2019.
- [60] Alain de Cheveigné and Hideki Kawahara. Yin, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, 111(4):1917–1930, 2002.
- [61] Miet de letter, Patrick Santens, Marc Bodt, Georges Maele, John Borsel, and Paul Boon. The effect of levodopa on respiration and word intelligibility in people with advanced parkinson’s disease. *Clinical neurology and neurosurgery*, 109:495–500, 08 2007.
- [62] Najim Dehak, Pierre Dumouchel, and Patrick Kenny. Modeling prosodic features with joint factor analysis for speaker verification. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(7):2095–2103, 2007.
- [63] Ekaterini Derdemezis, Houri K Vorperian, Ray D Kent, Marios Fourakis, Emily L Reinicke, and Daniel M Bolt. Optimizing vowel formant measurements in four acoustic analysis systems for diverse speaker groups. *American Journal of Speech-Language Pathology*, 25(3):335–354, 2016.
- [64] N. Dhananjaya, B. Yegnanarayana, and Peri Bhaskararao. Acoustic analysis of trill sounds. *The Journal of the Acoustical Society of America*, 131(4):3141–3152, 2012.
- [65] Thomas Drugman, Mark Thomas, Jon Gunason, Patrick Naylor, and Thierry Dutoit. Detection of glottal closure instants from speech signals: A quantitative review. *IEEE Transactions on Audio, Speech & Language Processing*, 20:994–1006, 03 2012.
- [66] D. Duez. Syllable structure, syllable duration and final lengthening in parkinsonian french speech. *Journal of Multilingual Communication Disorders*, 4(1):45–57, 2006.
- [67] Danielle Duez. Acoustic analysis of occlusive weakening in parkinsonian french speech. 08 2007.
- [68] Joseph R Duffy. *Motor speech disorders e-book: Substrates, differential diagnosis, and management*. Elsevier Health Sciences, 2019.
- [69] J.R. Duffy. *Motor Speech Disorders Substrates, Differential Diagnosis, and Management*. Elsevier, 2013.

- [70] Sorin Dusan and Lawrence Rabiner. On the relation between maximum spectral transition positions and phone boundaries. volume 2, 01 2006.
- [71] James Paul Dworkin and Arnold E. Aronson. Tongue strength and alternate motion rates in normal and dysarthric subjects. *Journal of Communication Disorders*, 19(2):115–132, 1986.
- [72] Y.G. Estevan, Vincent Wan, and Odette Scharenborg. Finding maximum margin segments in speech. volume 4, pages IV–937, 05 2007.
- [73] Emily Fischer and Alexander Goberman. Voice onset time in parkinson disease. 43:21–34, 09 2009.
- [74] Samuel G Fletcher. Time-by-count measurement of diadochokinetic syllable rate. *Journal of speech and hearing research*, 15(4):763–770, 1972.
- [75] A.J. Flint, S.E. Black, I. Campbell-Taylor, G.F. Gailey, and C. Levinton. Acoustic analysis in the differentiation of parkinson’s disease and major depression. *Journal of Psycholinguistic Research*, 21(5):383–399, Sep 1992.
- [76] Alastair J. Flint, Sandra E. Black, Irene Campbell-Taylor, Gillian F. Gailey, and Carey Levinton. Acoustic analysis in the differentiation of parkinson’s disease and major depression. *Journal of Psycholinguistic Research*, 21(5):383–399, Sep 1992.
- [77] Peter Flipsen and Sungbok Lee. Reference data for the american english acoustic vowel space. *Clinical Linguistics & Phonetics*, 26(11-12):926–933, 2012.
- [78] Peter Flipsen and Sungbok Lee. Reference data for the american english acoustic vowel space. *Clinical Linguistics & Phonetics*, 26(11-12):926–933, 2012.
- [79] K. Forrest, G. Weismer, P. Milenkovic, and R. Dougall. Statistical analysis of word-initial voiceless obstruents: preliminary data. *The Journal of the Acoustical Society of America*, 84 1:115–23, 1988.
- [80] Karen Forrest, Gary Weismer, and Greg S. Turner. Kinematic, acoustic, and perceptual analyses of connected speech produced by parkinsonian and normal geriatric adults. *The Journal of the Acoustical Society of America*, 85(6):2608–2622, 1989.
- [81] J. Franke, M. Mueller, F. Hamlaoui, S. Stueker, and A. Waibel. Phoneme boundary detection using deep bidirectional lstms. In *Speech Communication; 12. ITG Symposium*, pages 1–5, 2016.
- [82] D. Freed. *Motor speech disorders: Diagnosis & treatment*. Boston, Massachusetts: Cengage, 2011.
- [83] Matthias Froehlich, Dirk Michaelis, and Eduardo Kruse. Objektive beschreibung der stimmgüte unter verwendung des heiserkeits-diagramms. *HNO*, 46:684–689, 07 1998.

- [84] Sadaoki Furui. On the role of spectral transition for speech perception. *The Journal of the Acoustical Society of America*, 80:1016–25, 11 1986.
- [85] C. Ghez and S. Fahn. The cerebellum. *Principles of Neural Science 2 nd edition*, New York: Elsevier, page 502–522, 1985.
- [86] Patricia Gillivan-Murphy, Paul Carding, and Nick Miller. Vocal tract characteristics in parkinson’s disease. *Current Opinion in Otolaryngology & Head and Neck Surgery*, 24:1, 02 2016.
- [87] Patricia Gillivan-Murphy, Paul Carding, and Nick Miller. Vocal tract characteristics in parkinson’s disease. *Current Opinion in Otolaryngology & Head and Neck Surgery*, 24:1, 02 2016.
- [88] Patricia Gillivan-Murphy, Nick Miller, and Paul Carding. Voice tremor in parkinson’s disease: An acoustic study. *Journal of Voice*, 33(4):526–535, 2019.
- [89] S. Gilman, G.K. Wenning, P.A. Low, D.J. Brooks, C.J. Mathias, et al. Second consensus statement on the diagnosis of multiple system atrophy. *Neurology*, 71:670–676, August 2008.
- [90] Sid Gilman, S.J. May, C.W. Shults, C.M. Tanner, Walter Kukull, V. Lee, E Masliah, Phillip Low, Paola Sandroni, J.Q. Trojanowski, Laurie Ozelius, and T Foroud. The north american multiple system atrophy study group. *Journal of neural transmission (Vienna, Austria : 1996)*, 112:1687–94, 09 2006.
- [91] Giovanni Dimauro; Francesco Girardi. Italian parkinson’s voice and speech. 2019.
- [92] Alexander Goberman and Lawrence W Elmer. Acoustic analysis of clear versus conversational speech in individuals with parkinson’s disease. *Journal of communication disorders*, 38:215–30, 05 2005.
- [93] Alexander Goberman and Lawrence W Elmer. Acoustic analysis of clear versus conversational speech in individuals with parkinson’s disease. *Journal of communication disorders*, 38:215–30, 05 2005.
- [94] Alexander M. Goberman and Jessica McMillan. Relative speech timing in parkinson disease. *Contemporary Issues in Communication Science and Disorders*, 32(Spring):22–29, 2005.
- [95] U. Goswami, S. Nirmala, Vikram C M, S. Kalita, and S. Prasanna. Analysis of articulation errors in dysarthric speech. *Journal of Psycholinguistic Research*, 49, 10 2019.
- [96] D. B. Grayden and M. S. Scordilis. Phonemic segmentation of fluent speech. i:I/73–I/76 vol.1, 1994.

- [97] Wolfgang Grodd, Ernst Hülsmann, Martin Lotze, Dirk Wildgruber, and Michael Erb. Sensorimotor mapping of the human cerebellum: fmri evidence of somatotopic organization. *Human brain mapping*, 13(2):55–73, 2001.
- [98] A Guz. Brain, breathing and breathlessness. *Respiration Physiology*, 109(3):197–204, 1997.
- [99] Hirose H. Pathophysiology of motor speech disorders (dysarthria). *Folia Phoniatri Logop*, 38:61–88, 1986.
- [100] S Hadjikitou and Charles Wiles. Respiratory complications related to bulbar dysfunction in motor neuron disease. *Acta neurologica Scandinavica*, 103:207–13, 05 2001.
- [101] Jane Hanratty, Catherine Deegan, Mary Walsh, and Barry Kirkpatrick. Analysis of glottal source parameters in parkinsonian speech. volume 2016, pages 3666–3669, 08 2016.
- [102] John H.L. Hansen, Sharmistha S. Gray, and Wooil Kim. Automatic voice onset time detection for unvoiced stops (/p/,/t/,/k/) with application to accent classification. *Speech Communication*, 52(10):777 – 789, 2010.
- [103] Shahina Haque. Variability of acoustic features of hypernasality and it’s assessment. 7, 10 2016.
- [104] Shahina Haque, Md Hanif Ali, AKM Fazlul Haque, et al. Variability of acoustic features of hypernasality and it’s assessment. *Int. J. Adv. Comput. Sci. Appl*, 7(9):195–201, 2016.
- [105] Brian Harel, Michael Cannizzaro, and Peter Snyder. Variability in fundamental frequency during speech in prodromal and incipient parkinson’s disease: A longitudinal case study. 56:24–9, 11 2004.
- [106] L. Hartelius, H. Gustavsson, M. Astrand, and B. Holmberg. Perceptual analysis of speech in multiple system atrophy and progressive supranuclear palsy. *Journal of Medical Speech-Language Pathology*, 14:241–248, 2006.
- [107] Lena Hartelius and Per Svensson. Speech and swallowing symptoms associated with parkinson’s disease and multiple sclerosis: a survey. *Folia phoniatri et logopaedica*, 46(1):9–17, 1994.
- [108] Dik J. Hermes. Vowel-onset detection. *The Journal of the Acoustical Society of America*, 87(2):866–873, 1990.
- [109] Ingo Hertrich and Hermann Ackermann. Gender-specific vocal dysfunctions in parkinson’s disease: Electroglottographic and acoustic analyses. *Annals of Otolaryngology, Rhinology & Laryngology*, 104(3):197–202, 1995.

- [110] Weismer G.-Hoit J Hixon, T. Preclinical speech science: Anatomy, physiology, acoustics, perception. *2nd. San Diego: Plural Publishing*, 2013.
- [111] J. Hlavnicka, R. Cmejla, T. Tykalova, K. Sonka, E. Ruzicka, and J. Rusz. Automated analysis of connected speech reveals early biomarkers of parkinson’s disease in patients with rapid eye movement sleep behaviour disorder. *Sci Rep*, 7(12), 2017.
- [112] Jan Hlavnicka, Roman Cmejla, Jirí Klempír, Evzen Ruzicka, and Jan Rusz. Acoustic tracking of pitch, modal, and subharmonic vibrations of vocal folds in parkinson’s disease and parkinsonism. *IEEE Access*, 7:150339–150354, 2019.
- [113] Jan Hlavnička, Roman Čmejla, Tereza Tykalová, Karel Šonka, Evžen Ržička, and Jan Rusz. Automated analysis of connected speech reveals early biomarkers of parkinson’s disease in patients with rapid eye movement sleep behaviour disorder. *Scientific reports*, 7(1):1–13, 2017.
- [114] Jan Hlavnicka, Tereza Tykalová, Roman Cmejla, Jirí Klempír, Evzen Ruzicka, and Jan Rusz. Dysprosody differentiate between parkinson’s disease, progressive supranuclear palsy, and multiple system atrophy. In *INTERSPEECH*, pages 1844–1848, 2017.
- [115] Jan Hlavnicka, Tereza Tykalová, Roman Cmejla, Jirí Klempír, Evzen Ruzicka, and Jan Rusz. Dysprosody differentiate between parkinson’s disease, progressive supranuclear palsy, and multiple system atrophy. In *INTERSPEECH*, pages 1844–1848, 2017.
- [116] Aileen K. Ho, Robert Ianseck, Caterina Marigliani, John L. Bradshaw, and Sandra Gates. Speech impairment in a large sample of patients with parkinson’s disease. *Behavioural neurology*, 11 3:131–7, 1998.
- [117] R.J. Holmes, Jennifer Oates, Debra Phyland, and A.J. Hughes. Voice characteristics in the progression of parkinson’s disease. 35:407–18, 07 2009.
- [118] O Hornykiewicz. Parkinson’s disease and the adaptive capacity of the nigrostriatal dopamine system: possible neurochemical mechanisms. *Advances in neurology*, 60:140–147, 1993.
- [119] John-Paul Hosom and Ronald Cole. Burst detection based on measurements of intensity discrimination. pages 564–567, 01 2000.
- [120] José Ignacio Hualde, Ryan Shosted, and Daniel Scarpace. Acoustics and articulation of spanish /d/ spirantization. In *ICPhS*, 2011.
- [121] Jessica Huber, E.T. Stathopoulos, L.O. Ramig, and S.L. Lancaster. Respiratory function and variability in individuals with parkinson disease: Pre- and post-lee silverman voice treatment. *Journal of Medical Speech-Language Pathology*, 11:185–201, 12 2003.

- [122] A J Hughes, S E Daniel, L Kilford, and A J Lees. Accuracy of clinical diagnosis of idiopathic parkinson’s disease: a clinico-pathological study of 100 cases. *Journal of Neurology, Neurosurgery & Psychiatry*, 55(3):181–184, 1992.
- [123] Andrew J Hughes, Susan E Daniel, Siobhan Blankson, and Andrew J Lees. A clinicopathologic study of 100 cases of parkinson’s disease. *Archives of neurology*, 50(2):140–148, 1993.
- [124] George W. Hughes and Morris Halle. Spectral properties of fricative consonants. *The Journal of the Acoustical Society of America*, 28(2):303–310, 1956.
- [125] Y.E. Huh, J. Park, M.K Suh, S.E. Lee, J. Kim, Y. Jeong, H. Kim, and J.W. Cho. Differences in early speech patterns between parkinson variant of multiple system atrophy and parkinson’s disease. *Brain and language*, 147:14–20, May 2015.
- [126] Y.E. Huh, J. Park, M.K Suh, S.E. Lee, J. Kim, Y. Jeong, H. Kim, and J.W. Cho. Differences in early speech patterns between parkinson variant of multiple system atrophy and parkinson’s disease. *Brain and language*, 147:14–20, May 2015.
- [127] Peter Indefrey and Willem JM Levelt. The spatial and temporal signatures of word production components. *Cognition*, 92(1-2):101–144, 2004.
- [128] Eiji Isozaki, Toshio Shimizu, Kiyomi Takamoto, Satoshi Horiguchi, Tetsuro Hayashida, Masaya Oda, and Hitoshi Tanabe. Vocal cord abductor paralysis (vcap) in parkinson’s disease: difference from vcap in multiple system atrophy. *Journal of the Neurological Sciences*, 130(2):197–202, 1995.
- [129] Brooks David J. and Seppi Klaus and. Proposed neuroimaging criteria for the diagnosis of multiple system atrophy. *Movement Disorders*, 24(7):949–964, 2009.
- [130] Hlavnicka J. Automated analysis of speech disorders in neurodegenerative diseases. *Ph.D. Thesis, Faculty of Electrical Engineering, Czech Technical University, Prague, Czechia*, 2018.
- [131] A. R. Jayan and P. C. Pandey. Detection of stop landmarks using gaussian mixture modeling of speech spectrum. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4681–4684, 2009.
- [132] Kurt Jellinger. Different tau pathology pattern in two clinical phenotypes of progressive supranuclear palsy. *Neuro-degenerative diseases*, 5:339–46, 04 2008.
- [133] Jhing-Fa Wang, Chung-Hsien Wu, Shih-Hung Chang, and Jau-Yien Lee Lee. A hierarchical neural network model based on a c/v segmentation algorithm for isolated mandarin speech recognition. *IEEE Transactions on Signal Processing*, 39(9):2141–2146, 1991.

- [134] Jhing Fa Wang and Shi-Huang Chen. A c/v segmentation algorithm for mandarin speech signal based on wavelet transforms. In *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258)*, volume 1, pages 417–420 vol.1, 1999.
- [135] Félix Javier Jiménez-Jiménez, Javier Gamboa, Alberto Nieto, Josana Guerrero, Miguel Orti-Pareja, Jose Antonio Molina, Esteban García-Albea, and Ignacio Cobeta. Acoustic voice analysis in untreated patients with parkinson’s disease. *Parkinsonism & Related Disorders*, 3(2):111–116, 1997.
- [136] Allard Jongman, Ratreë Wayland, and Serena Wong. Acoustic characteristics of english fricatives. *The Journal of the Acoustical Society of America*, 108(3):1252–1263, 2000.
- [137] Jun Hou, L. Rabiner, and S. Dusan. Automatic speech attribute transcription (asat) - the front end processor. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, volume 1, pages I–I, 2006.
- [138] Abdellah Kacha, Christophe Mertens, Francis Grenez, Sabine Skodda, and Jean Schoentgen. On the harmonic-to-noise ratio as an acoustic cue of vocal timbre of parkinson speakers. *Biomedical Signal Processing and Control*, 37:32–38, 2017.
- [139] Ozlem Kalinli. Automatic phoneme segmentation using auditory attention features. In *INTERSPEECH*, 2012.
- [140] Ozlem Kalinli. Combination of auditory attention features with phone posteriors for better automatic phoneme segmentation. In *INTERSPEECH*, 2013.
- [141] Fredrik Karlsson, Ellika Schalling, Katja Laakso, Kerstin Johansson, and Lena Hartelius. Assessment of speech impairment in patients with parkinson’s disease from acoustic quantifications of oral diadochokinetic sequences. *The Journal of the Acoustical Society of America*, 147(2):839–851, 2020.
- [142] Dr. Ryuta Kataoka, Dr. Ken-Ichi Michi, Dr. Kaoru Okabe, Dr. Tanetoshi Miura, and Dr. Hiroshi Yoshida. Spectral properties and quantitative evaluation of hypernasality in vowels. *The Cleft Palate-Craniofacial Journal*, 33(1):43–50, 1996.
- [143] E. Keller. *Speech Motor Timing*. (eds) *Speech Production and Speech Modelling*. NATO ASI Series (Series D: Behavioural and Social Sciences), vol 55. Springer, Dordrecht, 1990.
- [144] Ray D. Kent, Jane Finley Kent, Joe R. Duffy, Jack E. Thomas, Gary Weismer, and Sarah Stuntebeck. Ataxic dysarthria. *Journal of Speech, Language, and Hearing Research*, 43(5):1275–1289, 2000.
- [145] Ray D. Kent and Yunjung Kim. *Acoustic Analysis of Speech*, chapter 22, pages 360–380. John Wiley & Sons, Ltd, 2009.

- [146] Ray D Kent, Gary Weismer, Jane F Kent, Houri K Vorperian, and Joseph R Duffy. Acoustic studies of dysarthric speech: Methods, progress, and potential. *Journal of Communication Disorders*, 32(3):141 – 186, 1999.
- [147] Ray D Kent, Gary Weismer, Jane F Kent, Houri K Vorperian, and Joseph R Duffy. Acoustic studies of dysarthric speech: Methods, progress, and potential. *Journal of Communication Disorders*, 32(3):141 – 186, 1999.
- [148] Raymond Kent and Yunjung Kim. Toward an acoustic typology of motor speech disorders. *Clinical linguistics & phonetics*, 17:427–45, 10 2003.
- [149] Raymond Kent and Yunjung Kim. Toward an acoustic typology of motor speech disorders. *Clinical linguistics & phonetics*, 17:427–45, 10 2003.
- [150] Raymond D Kent and Houri K Vorperian. Static measurements of vowel formant frequencies and bandwidths: A review. *Journal of communication disorders*, 74:74–97, 2018.
- [151] R.D. Kent and C. Read. The acoustic analysis of speech. *San Diego, Calif.:Singular Pub. Group*, 1992.
- [152] R.D. Kent and C. Read. The speech sciences. *San Diego, Calif.:Singular Pub. Group*, 1997.
- [153] Han-Joon Kim, Beom Jeon, Jee-Y Lee, and Ji Yun. Survival of korean patients with multiple system atrophy. *Movement disorders : official journal of the Movement Disorder Society*, 26:909–12, 04 2011.
- [154] Y. Kim. Acoustic characteristics of fricatives /s/ and /ʃ/ produced by speakers with parkinson’s disease. *Clinical Archives of Communication Disorders*, 2, 01 2017.
- [155] Y. Kim, R.D. Kent, J. Kent, and J.R. Duffy. Perceptual and acoustic features of dysarthria in multiple system atrophy. *Journal of Medical Speech-Language Pathology*, 18:66–71, October 2010.
- [156] Yunjung Kim, Raymond D Kent, and Gary Weismer. An acoustic study of the relationships among neurologic disease, dysarthria type, and severity of dysarthria. 2011.
- [157] Simon King and Paul Taylor. Detection of phonological features in continuous speech using neural networks. *Computer Speech & Language*, 14:333–353, 10 2000.
- [158] Kluin KJ, Gilman S, Lohman M, and Junck L. Characteristics of the dysarthria of multiple system atrophy. *Archives of Neurology*, 53(6):545–548, 1996.
- [159] Kluin KJ, Gilman S, Foster NL, and et al. Neuropathological correlates of dysarthria in progressive supranuclear palsy. *Archives of Neurology*, 58(2):265–269, 2001.

- [160] Karen J. Kluin, Norman L. Foster, Stanley Berent, and Sid Gilman. Perceptual analysis of speech disorders in progressive supranuclear palsy. *Neurology*, 43:563–563, 1993.
- [161] Karen J. Kluin, Norman L. Foster, Stanley Berent, and Sid Gilman. Perceptual analysis of speech disorders in progressive supranuclear palsy. *Neurology*, 43(3 Part 1):563–563, 1993.
- [162] Martin Köllensperger, Felix Geser, Jean-Pierre Ndayisaba, Sylvia Boesch, Klaus Seppi, Karen Østergaard, Erik Dupont, A Cardozo, Eduardo Tolosa, Michael Abele, et al. Presentation, diagnosis, and management of multiple system atrophy in europe: final analysis of the european multiple system atrophy registry. *Movement Disorders*, 25(15):2604–2612, 2010.
- [163] Nobu Komai, Chiho Ishida, Kazuya Takahashi, and Atsuro Tagami. Hypercapnic respiratory failure is often seen in advanced progressive supranuclear palsy (p5.250). *Neurology*, 82(10 Supplement), 2014.
- [164] Renata Kowalska-Taczanowska, Andrzej Friedman, and Dariusz Kozirowski. Parkinson’s disease or atypical parkinsonism? the importance of acoustic voice analysis in differential diagnosis of speech disorders. *Brain and Behavior*, 10(8):e01700, 2020.
- [165] F. Kreuk, Joseph Keshet, and Y. Adi. Self-supervised contrastive learning for unsupervised phoneme segmentation. In *INTERSPEECH*, 2020.
- [166] S. R. Kuberski, S. J. Tobin, and A. I. Gafos. A landmark-based approach to automatic voice onset time estimation in stop-vowel sequences. In *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 60–64, 2016.
- [167] Martin Köllensperger, Felix Geser, Jean-Pierre Ndayisaba, Sylvia Boesch, Klaus Seppi, Karen Østergaard, Erik Dupont, A Cardozo, Eduard Tolosa, Michael Abele, Thomas Klockgether, Farid Yekhlief, Francois Tison, Christine Danieks, Günther Deuschl, Miguel Coelho, Cristina Sampaio, Maria Bozi, Niall Quinn, and Gregor Wenning. Presentation, diagnosis, and management of multiple system atrophy in europe: Final analysis of the european multiple system atrophy registry. *Movement disorders : official journal of the Movement Disorder Society*, 25:2604–12, 10 2010.
- [168] M. Lalain, A. Ghio, L. Giusti, D. Robert, C. Fredouille, and V. Woisard. Design and development of a speech intelligibility test based on pseudowords in french: Why and how? *Journal of Speech, Language, and Hearing Research*, 63(7):2070–2083, 2020.
- [169] Ian Lalich, Dale Ekbom, Sidney Starkman, Diana Orbelo, and Timothy Morgenthaler. Vocal fold motion impairment in multiple system atrophy. *The Laryngoscope*, 124, 03 2014.

- [170] Jennifer Lam, Kris Tjaden, and Greg Wilding. Acoustics of clear speech: Effect of instruction. *Journal of speech, language, and hearing research : JSLHR*, 55, 03 2012.
- [171] Jennifer Lam, Kris Tjaden, and Greg Wilding. Acoustics of clear speech: Effect of instruction. *Journal of speech, language, and hearing research : JSLHR*, 55, 03 2012.
- [172] Brice Laurens, Radu Constantinescu, Roy Freeman, Alexander Gerhard, Kurt Jellinger, Andreas Jeromin, Florian Krismer, Brit Mollenhauer, Michael G. Schlossmacher, Leslie M. Shaw, Marcel M. Verbeek, Gregor K. Wenning, Kristian Winge, Jing Zhang, and Wassilios G. Meissner. Fluid biomarkers in multiple system atrophy: A review of the msa biomarker initiative. *Neurobiology of Disease*, 80:29 – 41, 2015.
- [173] R Leanderson, A Persson, and S Öhman. Electromyographic studies of facial muscle activity in speech. *Acta oto-laryngologica*, 72(1-6):361–369, 1971.
- [174] Guo-She Lee, Ching-Ping Wang, and Sherry Fu. Evaluation of hypernasality in vowels using voice low tone to high tone ratio. *The Cleft palate-craniofacial journal : official publication of the American Cleft Palate-Craniofacial Association*, 46:47–52, 02 2009.
- [175] Gongfeng Li, Khalid Daoudi, Jiri Klempir, and Jan Ruzs. Linear classification in speech-based objective differential diagnosis of parkinsonism. In *IEEE-ICASSP - 2018 IEEE International Conference on Acoustics, Speech and Signal Processing*, Calgary, Canada, Apr 2018.
- [176] Chi-Yueh Lin and Hsiao-Chuan Wang. Burst onset landmark detection and its application to speech recognition. *Audio, Speech, and Language Processing, IEEE Transactions on*, 19:1253 – 1264, 08 2011.
- [177] I. Litvan, Y. Agid, D. Calne, G. Campbell, B. Dubois, R.C. Duvoisin, C.G. Goetz, L. Golbe, J. Grafman, J.H. Growdon, et al. Clinical research criteria for the diagnosis of progressive supranuclear palsy (steele-richardson-olszewski syndrome) report of the ninds-spap international workshop. *Neurology*, 47:1–9, 1996.
- [178] Huei-Mei Liu, Feng-Ming Tsao, and Patricia Kuhl. The effect of reduced vowel working space on speech intelligibility in mandarin-speaking young adults with cerebral palsy. *The Journal of the Acoustical Society of America*, 117:3879–89, 06 2005.
- [179] Huei-Mei Liu, Feng-Ming Tsao, and Patricia Kuhl. The effect of reduced vowel working space on speech intelligibility in mandarin-speaking young adults with cerebral palsy. *The Journal of the Acoustical Society of America*, 117:3879–89, 06 2005.

- [180] Sharlene A. Liu. Landmark detection for distinctive feature-based speech recognition. *The Journal of the Acoustical Society of America*, 100(5):3417–3430, 1996.
- [181] AJ Lloyd. Comprehension of prosody in parkinson’s disease. *Cortex*, 35(3):389–402, 1999.
- [182] Jeri A. Logemann and Hilda B. Fisher. Vocal tract control in parkinson’s disease. *Journal of Speech and Hearing Disorders*, 46(4):348–352, 1981.
- [183] Jeri A. Logemann, Hilda B. Fisher, Benjamin Boshes, and E. Richard Blonsky. Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of parkinson patients. *Journal of Speech and Hearing Disorders*, 43(1):47–57, 1978.
- [184] Jeri A Logemann, Hilda B Fisher, Benjamin Boshes, and E Richard Blonsky. Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of parkinson patients. *Journal of Speech and hearing Disorders*, 43(1):47–57, 1978.
- [185] Anja Lowit. Quantification of rhythm problems in disordered speech: a re-evaluation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1658):20130404, 2014.
- [186] Donna Lundy, Soham Roy, Jun Xue, Roy Casiano, and Daniel Jassir. Spastic/spasmodic vs. tremulous vocal quality: Motor speech profile analysis. *Journal of voice : official journal of the Voice Foundation*, 18:146–52, 04 2004.
- [187] Frederic Macia, G Ballan, F Yekhlef, O Delmer, Claude Vital, A Lagueny, and F Tison. Progressive supranuclear palsy: A clinical, natural history and disability study. *Revue neurologique*, 159:31–42, 02 2003.
- [188] E. R. Maher and A. J. Lees. The clinical features and natural history of the steele-richardson-olszewski syndrome (progressive supranuclear palsy). *Neurology*, 36(7):1005–1005, 1986.
- [189] Kazumi Maniwa, Allard Jongman, and Travis Wade. Acoustic characteristics of clearly spoken english fricatives. *The Journal of the Acoustical Society of America*, 125:3962–73, 07 2009.
- [190] Peter Mariën, Herman Ackermann, Michael Adamaszek, Caroline Barwood, Alan Beaton, John Desmond, Elke De Witte, Angela Fawcett, Ingo Hertrich, Michael Küper, Maria Leggio, Cherie Marvel, Marco Molinari, Bruce Murdoch, Roderick Nicolson, Jeremy Schmahmann, Catherine Stoodley, Markus Thürling, Dagmar Timmann, and Wolfram Ziegler. Consensus paper: Language and the cerebellum: an ongoing enigma. *Cerebellum (London, England)*, 13, 12 2013.

- [191] Julie Mauclair, Lionel Koenig, M Robert, and Peggy Gatignol. Burst-based features for the classification of pathological voices. pages 2167–2171, 01 2013.
- [192] Megan J. McAuliffe, Elizabeth C. Ward, and Bruce E. Murdoch. Speech production in parkinson’s disease: I. an electropalatographic investigation of tongue-palate contact patterns. *Clinical Linguistics & Phonetics*, 20(1):1–18, 2006.
- [193] Michael McAuliffe, Michaela Socolof, Sarah Mihuc, Michael Wagner, and Morgan Sonderegger. Montreal forced aligner: Trainable text-speech alignment using kald. pages 498–502, 08 2017.
- [194] M. McClean, D. Beukelman, and K. Yorkston. Speech-muscle visuomotor tracking in dysarthric and nonimpaired speakers. *Journal of speech and hearing research*, 30 2:276–82, 1987.
- [195] Nikolaus R McFarland. Diagnostic approach to atypical parkinsonian syndromes. *Continuum: Lifelong Learning in Neurology*, 22(4 Movement Disorders):1117, 2016.
- [196] Ilaria Maria Merlo, Antonio Occhini, Claudio Pacchetti, and Enrico Alfonsi. Not paralysis, but dystonia causes stridor in multiple system atrophy. *Neurology*, 58(4):649–652, 2002.
- [197] N. Miller, U. Nath, E. Noble, and D. Burn. Utility and accuracy of perceptual voice and speech distinctions in the diagnosis of parkinson’s disease, psp and msa-p. *Neurodegenerative Disease Management*, 7:191–203, June 2017.
- [198] Sarr Moustapha, Alain Ghio, Robert Espesser, Teston Bernard, Moustapha Dramé, and François Viallet. *Relevance of Aerodynamic Evaluation in Parkinsonian Dysarthria*. 12 2011.
- [199] Jörg Rudolf Müller, Gregor Karl Wenning, Marc Verny, Ann Mckee, Kallol Ray Chaudhuri, Kurt Jellinger, Werner Poewe, and Irene Litvan. Progression of dysarthria and dysphagia in postmortem-confirmed parkinsonian disorders. *Archives of neurology*, 58 2:259–64, 2001.
- [200] Bruce E. Murdoch, Helen J. Chenery, Peter D. Stokes, and William J. Hardcastle. Respiratory kinematics in speakers with cerebellar disease. *Journal of Speech, Language, and Hearing Research*, 34(4):768–780, 1991.
- [201] K Murphy, DR Corfield, A Guz, GR Fink, RJS Wise, J Harrison, and L Adams. Cerebral areas associated with motor control of speech in humans. *Journal of Applied Physiology*, 83(5):1438–1447, 1997.
- [202] U. Nath, Y. Ben-Shlomo, R.G. Thomson, A.J. Lees, and D.J. Burn. Clinical features and natural history of progressive supranuclear palsy. *Neurology*, 60(6):910–916, 2003.

- [203] Upal Nath, Y Ben-Shlomo, R Thomson, H Morris, Nicholas Wood, AJ Lees, and David Burn. The prevalence of progressive supranuclear palsy (steele-richardson-olszewski syndrome) in the uk. *Brain : a journal of neurology*, 124:1438–49, 07 2001.
- [204] Bhanu Teja Nellore, RaviShankar Prasad, Sudarsana Reddy Kadiri, Suryakanth V. Gangashetty, and B. Yegnanarayana. Locating burst onsets using sff envelope and phase information. *Proc. Interspeech 2017*, pages 3023–3027, 2017.
- [205] David J Nicholl, Michael A Greenstone, Carl E Clarke, Patrizia Rizzu, Daniel Crooks, Alex Crowe, John Q Trojanowski, Virginia M-Y Lee, and Peter Heutink. An english kindred with a novel recessive tauopathy and respiratory failure. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 54(5):682–686, 2003.
- [206] Martin Niethammer and David Eidelberg. Metabolic brain networks in translational neurology: concepts and applications. *Annals of neurology*, 72 5:635–47, 2012.
- [207] P. Niyogi, C. Burges, and P. Ramesh. Distinctive feature detection using support vector machines. 1:425–428 vol.1, 1999.
- [208] P. Niyogi and M. M. Sondhi. Detecting stop consonants in continuous speech. *The Journal of the Acoustical Society of America*, 111(2):1063–1076, 2002.
- [209] M. Novotny, J. Ruzs, R. Cmejla, and E. Ruzicka. Automatic evaluation of articulatory disorders in parkinson’s disease. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 22:1366–1378, September 2014.
- [210] Michal Novotný, Jan Ruzs, Roman Čmejla, Hana Ržičková, Jiří Klempír, and Evžen Ržička. Hypernasality associated with basal ganglia dysfunction: evidence from parkinson’s disease and huntington’s disease. *PeerJ*, 4:e2530, 2016.
- [211] Michal Novotný, Jan Ruzs, K Spálenka, Jirí Klempír, D Horáková, and Evzen Ruzicka. Acoustic evaluation of nasality in cerebellar syndromes. In *INTER-SPEECH*, pages 3132–3136, 2017.
- [212] M. Novotný, J. Pospíšil, R. Čmejla, and J. Ruzs. Automatic detection of voice onset time in dysarthric speech. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4340–4344, 2015.
- [213] M Omoto, S Suzuki, T Ikeuchi, T Ishihara, T Kobayashi, Y Tsuboi, J Ogasawara, M Koga, M Kawai, T Iwaki, et al. Autosomal dominant tauopathy with parkinsonism and central hypoventilation. *Neurology*, 78(10):762–764, 2012.
- [214] Juan Rafael Orozco, Juan Vasquez, J. Vargas-Bonilla, Raman Arora, N. Dehak, Phani Nidadavolu, Heidi Christensen, Frank Rudzicz, M. Yancheva, Hamid R.

- Chinaei, A. Vann, N. Vogler, Tobias Bocklet, Miloš Cerňak, Julius Hannink, and Elmar Noeth. Neurospeech: An open-source software for parkinson’s speech analysis. *Digital Signal Processing*, 07 2017.
- [215] J. R. Orozco-Arroyave, F. Höning, J. D. Arias-Londoño, J. F. Vargas-Bonilla, S. Skodda, J. Ruzs, and E. Nöth. Voiced/unvoiced transitions in speech as a potential bio-marker to detect parkinson’s disease. In *INTERSPEECH*, 2015.
- [216] J. R. Orozco-Arroyave, F. Höning, J. D. Arias-Londoño, J. F. V. Bonilla, S. Skodda, J. Ruzs, and E. Nöth. Voiced/unvoiced transitions in speech as a potential bio-marker to detect parkinson’s disease. pages 95–99, 2015.
- [217] Juan Rafael Orozco-Arroyave, Juan Camilo Vásquez-Correa, Jesús Francisco Vargas-Bonilla, R. Arora, N. Dehak, P.S. Nidadavolu, H. Christensen, F. Rudzicz, M. Yancheva, H. Chinaei, A. Vann, N. Vogler, T. Bocklet, M. Cernak, J. Hannink, and Elmar Nöth. Neurospeech: An open-source software for parkinson’s speech analysis. *Digital Signal Processing*, 77:207–221, 2018.
- [218] Sean O’Sullivan, Luke Massey, D Williams, Laura Silveira-Moriyama, Peter Kempster, J Holton, Tamas Revesz, and A Lees. Clinical outcomes of progressive supranuclear palsy and multiple system atrophy. *Brain : a journal of neurology*, 131:1362–72, 06 2008.
- [219] Tetsutaro Ozawa, Hideo Shinoda, Masahiko Tomita, Takayoshi Shimohata, Hideaki Nakayama, and Masatoyo Nishizawa. Tremulous arytenoid movements predict severity of glottic stenosis in multiple system atrophy. *Movement disorders*, 25(10):1418–1423, 2010.
- [220] Sallyanne Palethorpe, Catherine Watson, and Rosalind Barker. Acoustic analysis of monophthong and diphthong production in acquired severe to profound hearing loss. *The Journal of the Acoustical Society of America*, 114:1055–68, 09 2003.
- [221] S. S. Park and N. S. Kim. On using multiple models for automatic speech segmentation. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(8):2202–2212, 2007.
- [222] Sabiha Parveen and Alexander Goberman. Presence of stop bursts and multiple bursts in individuals with parkinson disease. *International journal of speech-language pathology*, 16, 08 2013.
- [223] Christine A. M. Payan, François Viallet, Bernhard G. Landwehrmeyer, Anne-Marie Bonnet, Michel Borg, Franck Durif, Lucette Lacomblez, Frédéric Bloch, Marc Verny, Jacques Fermanian, Yves Agid, Albert C. Ludolph, Peter N. Leigh, Gilbert Bensimon, and on behalf of the NNIPPS Study Group.
- [224] Christine AM Payan, François Viallet, Bernhard G Landwehrmeyer, Anne-Marie Bonnet, Michel Borg, Franck Durif, Lucette Lacomblez, Frédéric Bloch,

- Marc Verny, Jacques Fermanian, et al. Disease severity and progression in progressive supranuclear palsy and multiple system atrophy: validation of the mnpips–parkinson plus scale. *PloS one*, 6(8):e22293, 2011.
- [225] Heike Penner, Nicholas Miller, and Maria Wolters. Motor speech disorders in three parkinsonian syndromes: a comparative study. 01 2007.
- [226] Kathe S Perez, Lorraine Olson Ramig, Marshall E Smith, and Christopher Dromey. The parkinson larynx: tremor and videostroboscopic findings. *Journal of Voice*, 10(4):354–361, 1996.
- [227] Matthew Poole, Jessica Wee, Joanne Folker, Louise Corben, Martin Delatycki, and Adam Vogel. Nasality in friedreich ataxia. *Clinical linguistics & phonetics*, 29:1–13, 09 2014.
- [228] Matthew L Poole, Jessica S Wee, Joanne E Folker, Louise A Corben, Martin B Delatycki, and Adam P Vogel. Nasality in friedreich ataxia. *Clinical linguistics & phonetics*, 29(1):46–58, 2015.
- [229] Robert A. Portnoy and Arnold E. Aronson. Diadochokinetic syllable rate and regularity in normal and in spastic and ataxic dysarthric subjects. *Journal of Speech and Hearing Disorders*, 47(3):324–328, 1982.
- [230] Ronald B Postuma, Daniela Berg, Matthew Stern, Werner Poewe, C Warren Olanow, Wolfgang Oertel, José Obeso, Kenneth Marek, Irene Litvan, Anthony E Lang, et al. Mds clinical diagnostic criteria for parkinson’s disease. *Movement disorders*, 30(12):1591–1601, 2015.
- [231] Ronald B. Postuma, A. E. Lang, J. F. Gagnon, Andrea J. Pelletier, and Jaques Montplaisir. How does parkinsonism start? prodromal parkinsonism motor changes in idiopathic rem sleep behaviour disorder. *Brain : a journal of neurology*, 135 Pt 6:1860–70, 2012.
- [232] Gayadhar Pradhan, Avinash Kumar, and Syed Shahnawazuddin. Excitation source features for improving the detection of vowel onset and offset points in a speech sequence. pages 1884–1888, 08 2017.
- [233] S. Prasanna, B. Reddy, and Krishnamoorthy Palanisamy. Vowel onset point detection using source, spectral peaks, and modulation spectrum energies. *Audio, Speech, and Language Processing, IEEE Transactions on*, 17:556 – 565, 06 2009.
- [234] C. Price. The anatomy of language: a review of 100 fmri studies published in 2009. *Annals of the New York Academy of Sciences*, 1191, 2010.
- [235] P. Péran, G. Barbagallo, F. Nemmi, M. Sierra, M. Galitzky, A. P. Traon, P. Payoux, W. G. Meissner, and O. Rascol. Mri supervised and unsupervised classification of parkinson’s disease and multiple system atrophy. *Movement Disorders*, 33(4):600–608, 2018.

- [236] Niall Quinn. A short clinical history of multiple system atrophy. *Clinical Autonomic Research*, 25(1):3–7, 2015.
- [237] AH Rajput, BL Rozdilsky, and Alex Rajput. Accuracy of clinical diagnosis in parkinsonism—a prospective study. *Canadian Journal of Neurological Sciences*, 18(3):275–278, 1991.
- [238] Pravin Bhaskar Ramteke and Shashidhar G. Koolagudi. Phoneme boundary detection from speech: A rule based approach. *Speech Communication*, 107:1–17, 2019.
- [239] Susan Ravizza. Dissociating the performance of cortical and subcortical patients on phonemic tasks. *Brain and cognition*, 53:301–10, 12 2003.
- [240] Peter Redgrave, Mercedes Fernández Rodríguez, Jonathan Mark Smith, Maria C. Rodríguez-Oroz, Stéphane LeHéricy, H. David Bergman, Yves Agid, Mahlon R. DeLong, and Jose A. Obeso. Goal-directed and habitual control in the basal ganglia: implications for parkinson’s disease. *Nature Reviews Neuroscience*, 11:760–772, 2010.
- [241] Axel Riecker, Dirk Wildgruber, Wolfgang Grodd, and Hermann Ackermann. Reorganization of speech production at the motor cortex and cerebellum following capsular infarction: a follow-up functional magnetic resonance imaging study. *Neurocase*, 8(6):417–423, 2002.
- [242] L Robertson and John Hammerstad. Jaw movement dysfunction related to parkinson’s disease and partially modified by levodopa. *Journal of neurology, neurosurgery, and psychiatry*, 60:41–50, 02 1996.
- [243] Maria C Rodríguez-Oroz, Marjan Jahanshahi, Paul Krack, Irene Litvan, Raúl Macías, Erwan Bezard, and José A Obeso. Initial clinical manifestations of parkinson’s disease: features and pathophysiological mechanisms. *The Lancet Neurology*, 8(12):1128–1139, 2009.
- [244] Dusan Roncevic, Jose-Alberto Palma, Jose Martinez, Niamh Goulding, Lucy Norcliffe-Kaufmann, and Horacio Kaufmann. Cerebellar and parkinsonian phenotypes in multiple system atrophy: similarities, differences and survival. *Journal of Neural Transmission*, 121(5):507–512, 2014.
- [245] Nelson Roy, Shawn Nissen, Christopher Dromey, and Shimon Sapir. Articulatory changes in muscle tension dysphonia: Evidence of vowel space expansion following manual circumlaryngeal therapy. *Journal of communication disorders*, 42:124–35, 12 2008.
- [246] Joseph JK O Ruanaidh and William J Fitzgerald. *Numerical Bayesian methods applied to signal processing*. Springer Science & Business Media, 2012.

- [247] J. Rusz, C. Bonnet, J. Klempir, T. Tykalova, E. Baborov, M. Novotny, A. Rulseh, and E. Ruzicka. Speech disorders reflect differing pathophysiology in parkinson's disease, progressive supranuclear palsy and multiple system atrophy. *Journal of neurology*, 262:992–1001, 2015.
- [248] J. Rusz, R. Cmejla, H. Ruzickova, J. Klempir, V. Majerova, J. Picmausova, J. Roth, and E. Ruzicka. Acoustic assessment of voice and speech disorders in parkinson's disease through quick vocal test. *Movement Disorders*, 26:1951–1952, April 2011.
- [249] J. Rusz, R. Cmejla, H. Ruzickova, and E. Ruzicka. Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated parkinson's disease. *The journal of the Acoustical Society of America*, 129:350–367, October 2011.
- [250] Jan Rusz, Roman Cmejla, Tereza Tykalova, Hana Ruzickova, Jiri Klempir, Veronika Majerova, Jana Picmausova, Jan Roth, and Evzen Ruzicka. Imprecise vowel articulation as a potential early marker of parkinson's disease: Effect of speaking task. *The Journal of the Acoustical Society of America*, 134(3):2171–2181, 2013.
- [251] Jan Rusz, Roman Cmejla, Tereza Tykalova, Hana Ruzickova, Jiri Klempir, Veronika Majerova, Jana Picmausova, Jan Roth, and Evzen Ruzicka. Imprecise vowel articulation as a potential early marker of parkinson's disease: Effect of speaking task. *The Journal of the Acoustical Society of America*, 134(3):2171–2181, 2013.
- [252] Jan Rusz, Jan Hlavnička, Roman Čmejla, and Evžen Ržička. Automatic evaluation of speech rhythm instability and acceleration in dysarthrias associated with basal ganglia dysfunction. *Frontiers in bioengineering and biotechnology*, 3:104, 2015.
- [253] Jan Rusz, Jan Hlavnička, Tereza Tykalova, Michal Novotný, Petr Dušek, Karel Šonka, and Evžen Ržička. Smartphone allows capture of speech abnormalities associated with high risk of developing parkinson's disease. *IEEE transactions on neural systems and rehabilitation engineering*, 26(8):1495–1507, 2018.
- [254] Jan Rusz, Marika Megrelishvili, Cecilia Bonnet, Michael Okujava, Hana Brožová, Irine Khatiasvili, Madona Sekhniashvili, Marina Janelidze, Eduardo Tolosa, and Evžen Ržička. A distinct variant of mixed dysarthria reflects parkinsonism and dystonia due to ephedrone abuse. *Journal of Neural Transmission*, 121(6):655–664, Jun 2014.
- [255] Jan Rusz, Tereza Tykalová, Giulio Salerno, Serena Bancone, Johara Scarpelli, and Maria Teresa Pellecchia. Distinctive speech signature in cerebellar and parkinsonian subtypes of multiple system atrophy. *Journal of Neurology*, 266(6):1394–1404, Jun 2019.

- [256] Neville Ryant, Jiahong Yuan, and Mark Liberman. Automating phonetic measurement: The case of voice onset time. *The Journal of the Acoustical Society of America*, 133:3608, 05 2013.
- [257] S Sachin, Garima Shukla, Vaibhav Goyal, Shivangi Singh, Vijay Aggarwal, Madhuri Behari, et al. Clinical speech impairment in parkinson’s disease, progressive supranuclear palsy, and multiple system atrophy. *Neurology India*, 56(2):122, 2008.
- [258] C. Okan Sakar and Olcay Kursun. Telediagnosis of parkinson’s disease using measurements of dysphonia. *Journal of medical systems*, 34:591–9, 08 2010.
- [259] Steven Sandoval, Visar Berisha, Rene Utianski, Julie Liss, and Andreas Spanias. Automatic assessment of vowel space area. *The Journal of the Acoustical Society of America*, 134:EL477–83, 11 2013.
- [260] Shimon Sapis, Lorraine O. Ramig, Jennifer L. Spielman, and Cynthia Fox. Formant centralization ratio: a proposal for a new acoustic measure of dysarthric speech. *Journal of speech, language, and hearing research : JSLHR*, 53 1:114–25, 2010.
- [261] Shimon Sapis, Lorraine O. Ramig, Jennifer L. Spielman, and Cynthia Fox. Formant centralization ratio: a proposal for a new acoustic measure of dysarthric speech. *Journal of speech, language, and hearing research : JSLHR*, 53 1:114–25, 2010.
- [262] Shimon Sapis, Jennifer L. Spielman, Lorraine O. Ramig, Brad H. Story, and Cynthia Fox. Effects of intensive voice treatment (the lee silverman voice treatment [lsvt]) on vowel articulation in dysarthric individuals with idiopathic parkinson disease: Acoustic and perceptual findings. *Journal of Speech, Language, and Hearing Research*, 50(4):899–912, 2007.
- [263] B. D. Sarma, S. S. Prajwal, and S. R. M. Prasanna. Improved vowel onset and offset points detection using bessel features. In *2014 International Conference on Signal Processing and Communications (SPCOM)*, pages 1–6, 2014.
- [264] M. Saxena, M. Behari, S. S. Kumaran, V. Goyal, and V. Narang. Assessing speech dysfunction using bold and acoustic analysis in parkinsonism. *Parkinsonism & Related Disorders*, 20(8):855 – 861, 2014.
- [265] Tanja Schmitz-Hübsch, Oleksandr Eckert, Uwe Schlegel, Thomas Klockgether, and Sabine Skodda. Instability of syllable repetition in patients with spinocerebellar ataxia and parkinson’s disease. *Movement disorders*, 27(2):316–319, 2012.
- [266] A. Schrag, Y. Ben-Shlomo, and N.P. Quinn. Prevalence of progressive supranuclear palsy and multiple system atrophy: a cross-sectional study. *The Lancet*, 354:1771–1775, November 1999.

- [267] Kayleen Elizabeth Schreiber. Who cares who’s talking? the influence of talker gender on how listeners hear speech. 2017.
- [268] C. H. Shadle and S. J. Mair. Quantifying spectral characteristics of fricatives. 3:1521–1524, 1996.
- [269] N. J. Shah, B. B. Vachhani, H. B. Sailor, and H. A. Patil. Effectiveness of plp-based phonetic segmentation for speech synthesis. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 270–274, 2014.
- [270] R. Arefi Shirvan and E. Tahami. Voice analysis for detecting parkinson’s disease using genetic algorithm and knn classification method. In *2011 18th Iranian Conference of Biomedical Engineering (ICBME)*, pages 278–283, 2011.
- [271] Yosi Shrem, Matthew Goldrick, and Joseph Keshet. Dr.vot: Measuring positive and negative voice onset time in the wild. pages 629–633, 09 2019.
- [272] Miquel Simonet, José Hualde, and Marianna Nadeu. Lenition of /d/ in spontaneous spanish and catalan. volume 2, 10 2012.
- [273] Sabine Skodda. Steadiness of syllable repetition in early motor stages of parkinson’s disease. *Biomedical Signal Processing and Control*, 17:55–59, 2015.
- [274] Sabine Skodda, Andrea Flasskamp, and Uwe Schlegel. Instability of syllable repetition as a model for impaired motor processing: Is parkinson’s disease a ”rhythm disorder”? *Journal of neural transmission (Vienna, Austria : 1996)*, 117:605–12, 03 2010.
- [275] Sabine Skodda, W Grönheit, N Mancinelli, and U Schlegel. Progression of voice and speech impairment in the course of parkinson’s disease: A longitudinal study. *Parkinson’s disease*, 2013:389195, 12 2013.
- [276] Sabine Skodda, Wenke Grönheit, and Uwe Schlegel. Intonation and speech rate in parkinson’s disease: General and dynamic aspects and responsiveness to levodopa admission. *Journal of voice : official journal of the Voice Foundation*, 25:e199–205, 11 2010.
- [277] Sabine Skodda, Wenke Grönheit, and Uwe Schlegel. Impairment of vowel articulation as a possible marker of disease progression in parkinson’s disease. *PloS one*, 7:e32132, 02 2012.
- [278] Sabine Skodda, Heiko Rinsche, and Uwe Schlegel. Progression of dysprosody in parkinson’s disease over time—a longitudinal study. *Movement Disorders*, 24(5):716–722, 2008.
- [279] Sabine Skodda, Heiko Rinsche, and Uwe Schlegel. Progression of dysprosody in parkinson’s disease over time—a longitudinal study. *Movement disorders : official journal of the Movement Disorder Society*, 24:716–22, 04 2009.

- [280] Sabine Skodda and Uwe Schlegel. Speech rate and rhythm in parkinson’s disease. *Movement Disorders*, 23(7):985–992, 2008.
- [281] Sabine Skodda, Wenke Visser, and Uwe Schlegel. Acoustical analysis of speech in progressive supranuclear palsy. *Journal of Voice*, 25(6):725–731, 2011.
- [282] Sabine Skodda, Wenke Visser, and Uwe Schlegel. Gender-related patterns of dysprosody in parkinson disease and correlation between speech variables and motor symptoms. *Journal of Voice*, 25(1):76–82, 2011.
- [283] Sabine Skodda, Wenke Visser, and Uwe Schlegel. Vowel articulation in parkinson’s disease. *Journal of Voice*, 25(4):467 – 472, 2011.
- [284] Sabine Skodda, Wenke Visser, and Uwe Schlegel. Vowel articulation in parkinson’s disease. *Journal of voice*, 25(4):467–472, 2011.
- [285] Marshall E Smith, Lorraine Olson Ramig, Christopher Dromey, Kathe S Perez, and Ráz Samandari. Intensive voice treatment in parkinson disease: laryngostroboscopic findings. *Journal of voice*, 9(4):453–459, 1995.
- [286] N. Solomon and T. Hixon. Speech breathing in parkinson’s disease. *Journal of speech and hearing research*, 36 2:294–310, 1993.
- [287] Morgan Sonderegger and Joseph Keshet. Automatic measurement of voice onset time using discriminative structured prediction. *The Journal of the Acoustical Society of America*, 132(6):3965–3979, 2012.
- [288] Kristie A. Spencer and Margaret A. Rogers. Speech motor programming in hypokinetic and ataxic dysarthria. *Brain and Language*, 94(3):347 – 366, 2005.
- [289] Nadia Stefanova, Philipp Bücke, Susanne Duerr, and Gregor Wenning. Multiple system atrophy: An update. *Lancet neurology*, 8:1172–8, 12 2009.
- [290] K. N. Stevens and S. E. Blumstein. Invariant cues for place of articulation in stop consonants. *The Journal of the Acoustical Society of America*, 64(5):1358–1368, 1978.
- [291] Veronique Stouten and Hugo Van hamme. Automatic voice onset time estimation from reassignment spectra. *Speech Communication*, 51:1194–1205, 12 2009.
- [292] Y. Sun, M. L. Ng, C. Lian, L. Wang, F. Yang, and N. Yan. Acoustic and kinematic examination of dysarthria in cantonese patients of parkinson’s disease. In *2018 11th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, pages 354–358, 2018.
- [293] Per Svensson, Christer Henningson, and Stig Karlsson. Speech motor control in parkinson’s disease: a comparison between a clinical assessment protocol and a quantitative analysis of mandibular movements. *Folia Phoniatica et Logopaedica*, 45(4):157–164, 1993.

- [294] Mari Tada, Akiyoshi Kakita, Yasuko Toyoshima, Osamu Onodera, Tetsutaro Ozawa, Takashi Morita, Masatoyo Nishizawa, and Hitoshi Takahashi. Depletion of medullary serotonergic neurons in patients with multiple system atrophy who succumbed to sudden death. *Brain*, 132(7):1810–1819, 05 2009.
- [295] David Talkin and W Bastiaan Kleijn. A robust algorithm for pitch tracking (rapt). *Speech coding and synthesis*, 495:518, 1995.
- [296] Chris Tang, Kathleen Poston, Thomas Eckert, Andrew Feigin, Steven Frucht, Mark Gudesblatt, Vijay Dhawan, Martin Lesser, Jean-Paul Vonsattel, Stanley Fahn, and David Eidelberg. Differential diagnosis of parkinsonism: A metabolic imaging study using pattern analysis. *Lancet neurology*, 9:149–58, 02 2010.
- [297] João Teixeira. Algorithm for jitter and shimmer measurement in pathologic voices. *Procedia Computer Science*, 100:271–279, 12 2016.
- [298] F. Tison, J. F. Dartigues, L. Dubes, M. Zuber, A. Alperovitch, and P. Henry. Prevalence of parkinson’s disease in the elderly: a population study in gironde, france. *Acta Neurologica Scandinavica*, 90(2):111–115, 1994.
- [299] F Tison, F Yekhlef, V Chrysostome, and C Sourgen. Prevalence of multiple system atrophy. *The Lancet*, 355(9202):495 – 496, 2000.
- [300] Ingo R. Titze and Daniel W. Martin. Principles of voice production. *The Journal of the Acoustical Society of America*, 104(3):1148–1148, 1998.
- [301] Ingo R. Titze and Daniel W. Martin. Principles of voice production. *The Journal of the Acoustical Society of America*, 104(3):1148–1148, 1998.
- [302] Kris Tjaden, Jennifer Lam, and Greg Wilding. Vowel acoustics in parkinson’s disease and multiple sclerosis: Comparison of clear, loud, and slow speaking conditions. *Journal of speech, language, and hearing research : JSLHR*, 56, 07 2013.
- [303] Kris Tjaden, Jennifer Lam, and Greg Wilding. Vowel acoustics in parkinson’s disease and multiple sclerosis: Comparison of clear, loud, and slow speaking conditions. *Journal of speech, language, and hearing research : JSLHR*, 56, 07 2013.
- [304] Kris Tjaden, Deanna Rivera, Gregory Wilding, and Greg S Turner. Characteristics of the lax vowel space in dysarthria. 2005.
- [305] Kris Tjaden and Greg S. Turner. Spectral properties of fricatives in amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research*, 40(6):1358–1372, 1997.
- [306] T. Tykalova, J. Rusz, J. Klempir, R. Cmejla, and E. Ruzicka. Distinct patterns of imprecise consonant articulation among parkinson’s disease, progressive supranuclear palsy and multiple system atrophy. *Brain and language*, 165:1–9, February 2017.

- [307] Juan Camilo Vásquez-Correa, JR Orozco-Arroyave, T Bocklet, and E Nöth. Towards an automatic evaluation of the dysarthria level of patients with parkinson's disease. *Journal of communication disorders*, 76:21–36, 2018.
- [308] Ali Vefa, Hikmet Yilmaz, M Eđrilmez, and Samet Karaca. Voice analysis and videolaryngostroboscopy in patients with parkinson's disease. *European archives of oto-rhino-laryngology : official journal of the European Federation of Oto-Rhino-Laryngological Societies (EUFOS) : affiliated with the German Society for Oto-Rhino-Laryngology - Head and Neck Surgery*, 259:290–3, 07 2002.
- [309] Adam Vogel, Hasherah Ibrahim, Sheena Reilly, and Nicky Kilpatrick. A comparative study of two acoustic measures of hypernasality. *Journal of speech, language, and hearing research : JSLHR*, 52:1640–51, 12 2009.
- [310] Bridget Walsh and Anne Smith. Basic parameters of articulatory movements and acoustics in individuals with parkinson's disease. *Movement Disorders*, 27(7):843–850, 2012.
- [311] Kate E Watkins, Stephen M Smith, Steve Davis, and Peter Howell. Structural and functional abnormalities of the motor system in developmental stuttering. *Brain*, 131(1):50–59, 2008.
- [312] G. Weismer. Articulatory characteristics of parkinsonian dysarthria: Segmental and phrase-level timing, spirantization, and glottal-supraglottal coordination. In: *McNeil M, Rosenbek J, Aronson A: The Dysarthrias: Physiology, Acoustics, Perception, Management. College-Hill Press, San Diego*, pages 101–130, 1984.
- [313] Gary Weismer, JY Jeng, Jacqueline Laures-Gore, Raymond Kent, and Jane F. Kent. Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. *Folia phoniatrica et logopaedica : official organ of the International Association of Logopedics and Phoniatics (IALP)*, 53:1–18, 01 2001.
- [314] Lutz Welling and Hermann Ney. Formant estimation for speech recognition. *IEEE Transactions on Speech and Audio Processing*, 6(1):36–48, 1998.
- [315] GK Wenning, Y Ben-Shlomo, A Hughes, SE Daniel, A Lees, and NP Quinn. What clinical features are most useful to distinguish definite multiple system atrophy from parkinson's disease? *Journal of Neurology, Neurosurgery & Psychiatry*, 68(4):434–440, 2000.
- [316] G.K. Wenning, I. Litvan, and E. Tolosa. Milestones in atypical and secondary parkinsonisms. *Movement Disorders*, 26:1083–1095, May 2011.
- [317] GK Wenning, F Tison, Y Ben Shlomo, SE Daniel, and NP Quinn. Multiple system atrophy: a review of 203 pathologically proven cases. *Movement disorders*, 12(2):133–147, 1997.

- [318] Gregor Wenning, Y Ben Shlomo, Marina Magalhães, S.E. Daniel, and N.P. Quinn. Clinical features and natural history of multiple system atrophy. an analysis of 100 cases. 117 (Pt 4):835–45, 09 1994.
- [319] Gregor K Wenning, Felix Geser, Florian Krismer, Klaus Seppi, Susanne Duerr, Sylvia Boesch, Martin Köllensperger, Georg Goebel, Karl P Pfeiffer, Paolo Barone, et al. The natural history of multiple system atrophy: a prospective european cohort study. *The lancet neurology*, 12(3):264–274, 2013.
- [320] Dirk Wildgruber, Hermann Ackermann, and Wolfgang Grodd. Differential contributions of motor cortex, basal ganglia, and cerebellum to speech motor control: effects of syllable repetition rate evaluated by fmri. *Neuroimage*, 13(1):101–109, 2001.
- [321] David Williams, Janice Holton, Catherine Strand, Alan Pittman, Rohan Silva, Andrew Lees, and Tamas Revesz. Pathological tau burden and distribution distinguishes progressive supranuclear palsy-parkinsonism from richardson’s syndrome. *Brain : a journal of neurology*, 130:1566–76, 07 2007.
- [322] Min Ney Wong, Bruce Murdoch, and B. Whelan. Lingual kinematics in dysarthric and nondysarthric speakers with parkinson’s disease. *Parkinson’s disease*, 2011:352838, 10 2011.
- [323] Mika Yamaguchi, Kimihito Arai, Masato Asahina, and Takamichi Hattori. Laryngeal stridor in multiple system atrophy. *European neurology*, 49(3):154–159, 2003.
- [324] Osaki Yasushi, Ben-Shlomo Yoav, Lees Andrew J., Wenning Gregor K., and Quinn Niall P. A validation exercise on the new consensus criteria for multiple system atrophy. *Movement Disorders*, 24(15):2272–2276, 2009.
- [325] Yu Qiao, Naoya Shimomura, and Nobuaki Minematsu. Unsupervised optimal phoneme segmentation: Objectives, algorithm and comparisons. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3989–3992, 2008.
- [326] W. Ziegler and D. von Cramon. Spastic dysarthria after acquired brain injury: an acoustic study. *The British journal of disorders of communication*, 21 2:173–87, 1986.