Speech analysis for the differential diagnosis between Parkinson's disease, progressive supranuclear palsy and multiple system atrophy

Gongfeng Li

Université Paris Saclay

Multimedia Networking

Master thesis
September 2017

Speech analysis for the differential diagnosis between Parkinson's disease, progressive supranuclear palsy and multiple system atrophy

LI Gongfeng

Supervisor: Khalid Daoudi

Contents

1 Abstract **3**

2 Introduction **3**

3 State of the art **4**

3.1 Perceptual evaluation . . . 4

3.2 Objective evaluation . . . 6

3.3 Discussion . . . 8

I Comparison study on voicing and articulation features **9**

4 Comparison of acoustic speech analyse results **9**

4.1 Acoustic features . . . 9

4.2 Statistical analyses . . . 10

4.3 Estimation procedure . . . 12

4.3.1 Airflow insufficiency, Jitter, Shimmer, HNR, Strained-strangled voice, Excess pitch fluctuations . . . 12

4.3.2 Vocal Tremor . . . 13

4.3.3 Rapid AMR, Slow AMR, Irregular AMR . . . 13

5 Comparison of deviant speech dimensions characteristics **14**

5.1 Introduction . . . 14

5.2 Table . . . 14

5.3 Analysis . . . 15

6 Classification **16**

6.1 Introduction . . . 16

6.2 Preprocessing . . . 17

6.3 Feature selection and parameter tuning . . . 17

6.4 Result . . . 17

6.5 Discussion . . . 19

II Deep analysis **22**

7 Prosody features **22**

7.1 Intraword pause ratio . . . 22

7.2 Percentage of pause time and Number of pauses . . . 22

7.3 Monopitch . . . 23

8 Principal component analysis **23**

8.1 Introduction . . . 23

8.2 Variables factor map . . . 23

9 Linear discriminant analysis **24**
9.1 Introduction . . . 24
9.2 1d LDA projection . . . 24
9.3 Classification on 1d LDA projection values . . . 26
9.4 2d LDA projection and classification . . . 26

10 3-layers model **27**
10.1 Logistic regression . . . 27
10.2 Model structure . . . 28
10.3 1d Result . . . 28
10.4 2d Result . . . 29
10.5 Discussion . . . 29

11 Conclusion **30**

12 References **31**

1 Abstract

Acoustic speech analysis has been shown to have a good potential in differentiation between Parkinson's disease and atypical Parkinsonian syndromes (APS) such as progressive supranuclear palsy (PSP) and multiple system atrophy (MSA). Objective speech features were able to discriminate between PD and APS with 95% accuracy and between PSP and MSA with 75% accuracy in [7]. However, accuracy between PSP and MSA still has a large space to be improved and the more important aim is to provide more explicit information for differential diagnosis. In [7], 75% accuracy was achieved using support vector machine classifier based on radial basis function kernel. This means it's difficult to interpret the relation between selected features and decision hyperplane. In this internship, for discrimination between PSP and MSA, 9% higher accuracy (i.e, 84% accuracy) was attained by using support vector machine classifier based on radial basis function kernel and 80% accuracy was attained using linear dimension reduction methods and linear classifier. More importantly, with this strategy, we obtain a better understanding of feature discriminative power. This can be indeed very useful in clinical application.

2 Introduction

Parkinson's disease (PD) and atypical Parkinsonian syndromes (APS) are neurodegenerative diseases. In the early periods of the disease, the symptoms of PD and APS are very similar. The differential diagnosis may be very difficult in the early stages of the diseases, while that certainty of early diagnosis is important for the patient due to divergent prognosis.

Speech disorders, commonly known as dysarthria, are an early symptom common to both diseases from different origins. Speech assessment is an inexpensive, quick and simple technique that could potentially used in evaluation of subjects even in early periods of diseases.

Most studies of speech disorder analysis focused on the description of the dysarthria profile, some of them are based on perceptual dysarthria assessment[1-3], while in recent years several studies tend to perform objective acoustic analysis to provide the description of dysarthria [4-7]. The dysarthria may provide important clues to the discrimination between these three diseases since [7] was already able to discriminate between PD and APS with 95% accuracy and between PSP and MSA with 75% accuracy.

The objective of this internship is to use different voice impairment measurements (features) to perform a preliminary experimental study on discriminating power of these different measurements. Our approach is to use impairment speech dimensions, obtained by digital processing of voice recordings of patients provided by [7], as a means for discrimination between PD, PSP and MSA.

Our study is divided into two parts, the first part is the comparison study with results obtained by [7] using only voicing (sustained vowels) and articulation (syllable repetitions) features. We can achieve 84% accuracy with compared to 75% in [7] only with voicing and articulation features. But we can't rely on this result since we may have overfitting problem. Moreover, with such nonlinear SVM classification, we can't get insight about the discriminating power of features. Given these observations, we carry out in Part 2. The goal of the second part is to get a deeper understanding of relationship among speech dimensions through exploration of linear dimension reduction methods which could aid in classification procedure. With linear dimension reduction, linear classifier could provide decent accuracy which is also higher than [7]. This lower complexity model will provide greater insight into the relationship between speech dimensions and differential diagnosis result.

3 State of the art

3.1 Perceptual evaluation

Studies of speech disorder analysis before 2000 are mainly based on perceptual estimation of dysarthria type[1-3].

Study[1] recorded two-minute conversational speech samples of 200 PD patients and evaluated them in three dimensions (voice(e.g. harsh quality, reduced volume, disturbed intonation), articulation (undershooting of articulatory movement resulting in imprecise articulation), fluency (e.g. motor initiation difficulties, inappropriate pauses, syllable repetition, rushes of speech)) subjectively. It reported that the hypokinetic dysarthria exist in majority (74%) of PD patients.

Meanwhile, study[2,3] mainly focused on the characteristics of dysarthria of MSA and PSP. In general, APS patients typically develop mixed dysarthria with various combinations of hypokinetic, spastic and ataxic components.

To be more clear, here are some descriptions of these three types of dysarthrias:

Hypokinetic dysarthria can be described as the dysarthria associated with disorders of the extra pyramidal motor system resulting in reduction and rigidity of movement, causing monotony of pitch and loudness, reduced stress, and imprecise enunciation of consonants[11]. Spastic can be described as the disturbance associated with upper motor neuron disorders causing excess tone and limited range in muscle movements, characterized by imprecise consonants, monotony of pitch and reduced stress, and a labored voice quality[11]. Ataxic can be described as the dysarthria associated with damage to the cerebellar system, characterized by imprecise consonants, excess and equal stress, inconsistent articulatory errors, and monotony of pitch and volume[11].

Study[2] investigated 44 PSP patients using oral examinations and oral agility assessment as well as perceptual speech analysis to identify the deviant speech dimensions and types of dysarthria in PSP. Perceptual speech analysis included identifying and rating severity of the deviant speech dimensions during the examination and from videotaped or audiotaped samples of spontaneous speech and oral reading of the "Grandfather Passage". The definitions of deviant speech dimensions are given by Darley et al[10]. And they used the University of Michigan classification of the deviant speech dimensions for ataxic and spastic, Mayo clinic hypokinetic list for hypokinetic(table 1).

Study[2] has reported that all PSP patients developed two or three types mixed dysarthrias, 50% of patients had prominent spastic dysarthria and hypokinetic dysarthria was greater than the the spastic and ataxic components for 34% of patients. In general, the spastic components were present in all cases and were the most severe components and the ataxic components were least severe.

Study [3] evaluated 46 MSA patients with oral motor function, oral agility and perceptual speech analysis. The perceptual speech analysis consisted of quantitative evaluation of spontaneous speech, expository speech and oral reading of "The Grandfather passage". They used the definitions of deviant speech dimensions as given by Darley et al[10] and the University of Michigan classification of ataxic, spastic, and hypokinetic dysarthrias(table 1).

For MSA patients in study[3], 48% of them had prominent hypokinetic components and 35% of patients had ataxic components that were greater than other two type dysarthrias. In the view of mean total score of severity, the hypokinetic components were the most severe (total score, 12.7/9.0, mean / SD) followed by ataxic (total score, 11.1/8.2, mean / SD) and spastic components (total score, 9.1/5.7, mean / SD).

In addition, study [3] have reported that the severity of some dysarthria components have relations with certain abnormalities found during the oral motor examination. The severity of hypokinetic components correlated significantly with the severity of masked face($r = 0.57$, $p < 0.001$) and was significantly associated with the presence of lip tremors($t = 5.1$; $p = 0.001$) and tongue tremors ($t = 2.9$; $p = 0.005$). A significant inverse correlation was found between the severity of ataxic components and the severity of masked face($r = -0.44$; $p = 0.003$). And ataxic components were significantly less prominent in patients with lip tremors($t = 4.0$; $p = 0.003$).

Table 1. Deviant Speech Dimensions for Hypokinetic, Ataxic, and Spastic Dysarthrias:

University of Michigan Classification			Mayo Clinic classification
Hypokinetic	Ataxic	Spastic	Hypokinetic
Low volume, able to increase on command	Excess and equal stress	Strained-strangled sound	Monopitch
Monopitch	Irregular articulation breakdown	Reduced stress	Reduced stress
Loudness decay	Alternating loudness variation	Harsh voice, continuous	Monoloudness
Short rushes of speech	Fluctuating pitch levels	Slow rate	Imprecise consonants
Increased speaking rate over time	Variable rate	Low pitch	Inappropriate silences
Imprecise phonemes over time	Harsh voice, transient	Imprecise phoneme	Short rushes
Decreased stress	Breathy voice, transient	Monoloudness	Harsh voice quality
Repetition of sounds, words and phrases	Altered nasality, transient	Hypernasality, continuous	Breathy voice, continuous
Inappropriate silences, difficulty initiating phonation	Voice tremors	Monopitch	Pitch level
Breathy voice, continuous	Audible inspiration	Prolonged phoneme and/or intervals	Variable rate

From study[2] and [3], it can be concluded that spastic components were most severe in PSP and ataxic components were the least severe and only presented in 68% of patients(compared to 100% for spastic and 95% for hypokinetic). While for MSA patients, the spastic components were the least severe and ataxic components presented in 89% of MSA patients. Considering individual speech aspects, stuttering occurred only in PSP.

For future study, especially for discrimination between MSA and PSP, speech dimensions of spastic and ataxic dysarthrias can be helpful since spastic components were the most severe in PSP and least severe in MSA and ataxic components were extremely less severe than the other two type dysarthrias (spastic : 10.7, hypokinetic : 9.6, ataxic: 2.6) in PSP while were quite severe in MSA patients (hypokinetic: 12.7, ataxic: 11.1, spastic: 9.7). The maximum score for each type of dysarthria is 30 for PSP and is 40 for MSA. Thus it can be expected to get some discriminative speech dimensions from spastic and ataxic components.

For example, there are some speech dimensions that haven't been used in recent acoustic objective analysis:

Spastic: Reduced stress: Speech shows reduction of proper stress or emphasis patterns. Hypernasality: Voice sounds excessively nasal. Excessive amount of air is resonated by nasal cavities.

Ataxic: Irregular articulatory breakdown: Intermittent nonsystematic breakdown in accuracy of articulation. Alternating loudness variations: There are alternating changes in loudness. Audible inspiration: Audible, breathy inspiration.

However, a recent study claimed that there is no consistent and significant differences found when using perceptual evaluations for discrimination between PSP and MSA[12]. Thus, we still mainly focus on acoustical measures.

3.2 Objective evaluation

In last ten years, several studies focused on an objective analysis of the dysarthria profile[4-7]. Most of them provided some characteristics in speech dimensions compared between PD and APS. Generally, the impairment of some specific speech dimensions is more severe in APS than in PD.

Study[4] investigated 22 PD patients, 18 PSP patients and 20 MSA patients to evaluate the presence and characteristics of dysarthria by quantitative assessment of three parameters: maximum phonation time (MPT), semantic fluency and reading speed.

For MPT, patients were instructed to take a deep breath and then sustain phonation [a] for as long as possible. Three samples were obtained and the longest response was taken. Semantic fluency was defined as the number of names of animals a person was able to spontaneously report in one minute. Reading speed was obtained by asking the patient to read aloud a standard paragraph in Hindi language. The number of words read by the subject in one minute was considered as his / her reading speed.

Significant overall difference was only seen for MPT ($p = 0.015$), the reading speed was affected most in PSP group but not significant and the semantic fluency is comparable between groups.

Study[5] focused on measuring quantitatively different speech parameters in PSP as compared with PD by acoustic analysis including mean F0, standard deviation of fundamental frequency, net speech rate (syllables per second related to net speech time), pause ratio (percentage of pause time related to total speech time), ratio of intraword pauses (percentage of pauses within polysyllabic words in relation to overall speech pauses) and Vowel articulation index. Twenty-six PSP patients and 30 age- and gender-matched PD patients were tested by performing a speech task consisting of a standard reading passage composed of four complex sentences.

In PSP group, the net speech rate was significantly reduced ($p = 0.001$) in while pause ratio was significantly increased ($p = 0.017$) and ratio of intraword pauses was reduced($p < 0.001$). Vowel articulation index (VAI) is a gender dependent measure, in the male patients' subgroup, VAI was significantly reduced ($p = 0.002$) whereas no such difference seen in the female patients' subgroup. The intonation variability measured by F0 SD was lower in PSP group ($p < 0.001$) and no significant difference was seen for the mean fundamental frequency.

For study[6], speech samples were acquired from 29 PD patients and 26 MSA patients by performing speech tasks consisted of sustained vowel phonation and reading a syntactically balanced text composed of 9 sentences.

To assess the pitch and quality of voice, mean fundamental frequency (F0), jitter, shimmer and noise-to-harmonic ratio (NHR) were measured from a relatively stable 1.5-s period of sustained phonation. In addition, the total speech rate (syllables per second based on total speech time of sentence 1 or 9), articulatory acceleration (total speech rate of sentence 9 minus that of sentence 1), total pause duration(of sentence 1 or 9), pause ratio within polysyllabic words (defined as in study [5] and of sentence 1 or 9) were estimated.

For the comparison between MSA and PD, study[6] has reported that among male patients' subgroups, F0 was significantly increased in MSA group($p = 0.017$), total speech rate was markedly decreased ($p = 0.048$ for sentence 1 and

$p = 0.008$ for sentence 9). In addition, total pause time for sentence 9 was longer in men with MSA than in those with PD ($p = 0.047$). None of speech variables showed significant differences between female patient groups.

Previous studies[4-6] provided us some characteristics in speech dimensions related to PD, PSP and MSA rather than discriminate them with these characteristics. However, study[7] was able to discriminate between APS and PD with 95 % accuracy and between PSP and MSA with 75 % accuracy which is really interesting.

Speech samples were acquired from 77 subjects including 15 PD, 12 PSP, 13 MSA and 37 healthy controls. None of the patients received antipsychotic therapy. 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 and monologue on a given topic for approximately 90s.

They evaluated sixteen dimensions in total, eight dimensions in hypokinetic dysarthria of PD, including airflow insufficiency,harsh voice, rapid AMR, inappropriate silences, reduceloudness, monopitch, imprecise vowels and dysfluency. While strained-strangled voice quality, slow AMR and slow rate were assessed as elements of spastic dysarthria. Meanwhile, excess pitch fluctuations, vocal tremor, irregular AMR, prolonged phonemes and excess intensity variations were related to ataxic dysarthria.

A support vector machine with a Gaussian radial basis kernel was applied in a classification experiment to determine the best combination of acoustic features to differentiate between PD, PSP and MSA groups. The combination of six acoustic features related to five deviant speech dimensions including harsh voice (jitter),inappropriate silences (percent pause time and number of pauses), slow AMR (diadochokinetic rate), excess intensity variation (intensity variation) and excess pitch fluctuation (pitch variation) were used to separate PD from APS 95% accuracy. Considering discrimination between PSP and MSA,the four deviant speech dimensions including harsh voice (harmonics-to-noise ratio), fluency (percent dysfluent word), slow rate (articulation rate) and vocal tremor (frequency tremor intensity index) were able to discriminate PSP from MSA with 75% accuracy.

Another study[8] focused on the different latencies of dysarthria and dysphagia for PD, PSP and MSA. Median dysarthria latencies were short in PSP and MSA (24 months each), and long in PD (84 months)[8]. Median dysphagia latencies were intermediate in PSP (42 months), MSA (67 months), and long in PD (130 months). Dysarthria or dysphagia within 1 year of disease onset was a distinguishing feature for APS (specificity, 100%).

The most recent study investigated the patterns and degree of consonant articulation deficits also focused on discrimination between PD, PSP and MSA [9]. Speech samples were acquired from 16 PD, 16 PSP, 16 MSA and 16 healthy control speakers by completing a series of speaking tasks lasting approximately 20 min. During the task, the participants were instructed to read the words presented by the examiner on paper cards which are tokens designed as "CVtka" used for the assessment of consonant articulation, where C represented a consonant and V corresponded to a corner vowel.

Three acoustic variables including VOT (Voice onset time), VOT ratio and vowel duration were investigated in this study where VOT was determined as the interval between the articulatory release of stop and the onset of vocal fold vibration. The acoustic variables were assessed for a subset of voiceless and voiced consonants separately.

For the subsets of voiceless plosives, the VOT is significantly longer in both PSP and MSA compared to PD or HC(both $p < 0.001$). The HC group manifested significantly smaller VOT ratio than all patient groups including PD ($p < 0.05$), PSP ($p < 0.01$) and MSA ($p < 0.001$).Meanwhile the vowel length is significantly longer in PSP group compared to both HC ($p < 0.05$) and PD ($p < 0.01$). For the subsets of voiced plosives, MSA group manifested significantly shorter negative VOT as compared to all groups including HC ($p < 0.001$), PD ($p < 0.01$) and PSP ($p < 0.001$).

3.3 Discussion

In general, quite a few studies investigated the relation between the characteristics of speech disorder especially in the aspect of dysarthria. The study[7] is the first one who did the classification between PD, PSP and MSA using these dysarthria features and achieved high accuracy (95 %) in differentiation between PD and APS patients. Meanwhile, the accuracy of discrimination between PSP and MSA is 75%, which may can be improved with considering VOT variable in the subsets of voiced plosives since the MSA group manifested significantly shorter negative VOT comparing to PSP ($p<0.001$).

In addition, there is one point should be noticed is that the difference of MPT was significant between the PSP and MSA groups ($p = 0.014$) in study[4] which is contrary to the result in [7].

The ratio of intraword pauses in study [5] and study [6], which hasn't been used in study [7] can be helpful for discrimination between MSA and PSP. According to the study[5], the ratio of intraword pauses is significantly decreased in PSP group compared to PD group while according to the study[6], the ratio of intraword pauses is not significantly different between MSA group and PD group.

Study [6] compared the speech acoustic parameters between MSA and PD with gender separated. It concluded that some speech acoustic parameters were discriminative between MSA and PD, not for all the patients, but only for male or female subgroup. The disadvantage for this method is that the database would be even smaller. This idea will bring us more accurate results for our future study if there are enough patient samples.

Part I

Comparison study on voicing and articulation features

This part mainly includes acoustic features extracting, comparison and analysis of obtained results and classification experiment. In this part, all experiments were performed based on the voicing and articulation recordings for [7].

4 Comparison of acoustic speech analyse results

A comparison study was performed based on the part of database of [7], including the voicing and articulation recordings, aiming to reproduce the results in the [7] for 10 speech dimensions .

4.1 Acoustic features

We evaluated 10 deviant speech dimensions including airflow insufficiency, Harsh voice, strained-strangled voice, excess pitch fluctuations, slow AMR, rapid AMR, irregular AMR and vocal tremor. See table as following for more comprehensive details. This table is provided by [7].

Deviant speech dimension	Vocal task	Acoustic measure	Description
Airflow insufficiency	Sustained phonation	Maximum phonation time(MPT)	Insufficient breath support for speech production;
Jitter	Sustained phonation	Random period variability	Harsh, rough and raspy voice
Shimmer	Sustained phonation	Amplitude perturbation	
HNR	Sustained phonation	ratio between harmonic signal power and noise signal power	
Strained-strangled voice	Sustained phonation	Degree of voicelessness(DUV)	Voice(phonation) sounds strained or strangled (effortful squeezing of voice through glottis)
Excess pitch fluctuations	Sustained phonation	Pitch variability(F0 SD)	Uncontrolled alterations in voice pitch
Slow AMR	Syllable repetition	Diadochokinetic (DDK) rate	Abnormally slow motion rate of articulators
Rapid AMR	Syllable repetition	DDK acceleration	Abnormally slow rate of
Irregular AMR	Syllable repetition	DDK regularity	Rate alternates from slow to fast
Vocal tremor	Sustained phonation	Frequency tremor intensity index (FTRI)	Tremulous phonation

4.2 Statistical analyses

Values used of speech dimensions for statistical analyses are measured as the average value of two voice samples for each patient except that for maximal phonation time we considered the maximum one of two recordings. To assess group differences, each speech dimension was compared across all three groups using a Kruskal-Wallis test. Effect sizes were measured with Cohen's d, with d > 0.5 indicating a medium effect and d > 0.8 indicating a large effect.

The statistical result of features is presented in the following table:

	Source	Groups				Group differences	Effect size		
Speech Dimension		HC Mean/SD (range)	PSP Mean/SD (range)	MSA Mean/SD (range)	PD Mean/SD (range)	p	PSP vs PD	MSA vs PD	PSP vs MSA
Hypokinetic									
1.Airflow insufficiency MPT(s)	(1)	(11.7-27.1)	13.2/5.0 (7.6-23.5)	13.5/6.8 (6.4-33.6)	17.1/8.5 (7.6-43.0)	0.18	-0.56	-0.47	-0.05
	(2)	19.88/5.16 (9.24-30.58)	15.11/4.98 (8.28-25.17)	15.03/7.19 (7.67-36.90)	19.09/8.53 (8.94-45.63)	0.12	-0.53	-0.49	0.01
2.jitter(%)	(1)	(0.26-1.77)	1.60/1.27 (0.43-3.30)	1.62/1.21 (0.29-4.27)	0.73/0.36 (0.35-1.83)	0.22	0.93	1.00	0.02
	(2)	0.51/0.28 (0.25-1.76)	0.83/0.39 (0.44-1.60)	1.29/0.98 (0.27-3.89)	0.60/0.18 (0.36-1.00)	0.16	0.74	0.97	-0.58
3.shimmer(%)	(1)	(2.05-9.83)	8.48/3.12 (2.33-12.65)	8.58/3.92 (2.31-16.18)	5.40/2.76 (2.59-11.74)	0.03	1.05	0.94	-0.02
	(2)	4.23/1.98 (1.71-9.90)	7.37-2.59 (2.34-10.84)	8.02/3.51 (2.35-14.69)	5.28/2.27 (2.61-11.22)	0.08	0.83	0.90	-0.20
4.HNR(dB)	(1)	(15.3-25.2)	15.0/3.9 (10.0-23.6)	16.4/5.3 (11.0-25.2)	20.4/2.6 (15.0-24.4)	0.008	-1.62	-0.95	-0.29
	(2)	19.69/3.22 (14.01-25.73)	14.85/3.76 (10.43-21.66)	15.34/5.71 (5.54-24.60)	19.21/2.52 (13.68-23.73)	0.03	-1.34	-0.87	-0.10
5.Rapid AMR DDK acceleration	(1)	(0.87-1.02)	1.06/0.47 (0.75-2.53)	1.00/0.16 (0.84-1.41)	0.95/0.13 (0.77-1.36)	0.85	0.32	0.32	0.18
	(2)	1.05/0.06 (0.86-1.18)	1.14/0.16 (0.78-1.36)	1.12/0.16 (0.83-1.48)	1.09/0.13 (0.89-1.45)	0.36	0.33	0.15	0.16
Spastic									
6.Strained-strangled voice DUV(%)	(1)	(0-2.23)	3.57/5.75 (0-18.05)	11.21/22.09 (0-81.35)	0.20/0.52 (0-1.55)	0.005	0.83	0.71	-0.47

	(2)	0.82/1.12 (0.05-6.31)	2.47/2.86 (0.14-8.34)	5.76/7.57 (0.05-29.05)	0.88/0.63 (0.00-2.34)	0.05	0.78	0.91	-0.54
7.Slow AMR DDK rate(syll/s)	(1)	(5.49-8.03)	5.72/1.32 (3.60-8.03)	5.45/1.32 (3.42-7.61)	6.82/1.12 (5.51-9.69)	0.03	-0.90	-1.12	0.20
	(2)	6.82/0.76 (4.25-8.52)	5.30/1.56 (2.04-7.99)	5.35/1.18 (2.88-7.38)	6.53/1.07 (5.15-9.19)	0.02	-0.90	-1.02	-0.03
Ataxic									
8.Excess pitch fluctuations F0 SD(st)	(1)	(0.16-0.80)	0.71/0.41 (0.28-1.43)	1.02/0.61 (0.16-2.32)	0.34/0.14 (0.18-0.68)	<0.001	1.19	1.54	-0.61
	(2)	1.39/1.30 (0.27-5.42)	3.91/1.87 (0.44-6.72)	3.88/2.44 (0.17-7.96)	1.80/1.34 (0.24-5.02)	0.01	1.27	1.04	0.01
9.Vocaltremor FTRI(%)	(1)	(0.16-1.11)	0.86/0.53 (0.29-2.19)	1.81/1.58 (0.22-5.39)	0.51/0.23 (0.22-1.02)	0.02	0.86	1.15	-0.81
	(2)	3.63/4.24 (0.00-15.62)	6.29/4.31 (0.00-12.97)	9.31/7.64 (1.02-21.94)	3.21/3.32 (0.00-13.31)	0.06	0.78	1.02	-0.46
10.Irregular AMR DDK regularity(ms)	(1)	(9.7-35.4)	51.1/40.3 (9.8-131.0)	43.7/27.1 (12.9-94.9)	18.6/8.8 (5.4-35.4)	0.009	1.12	1.25	0.22
	(2)	29.04/12.03 (15.96-79.03)	73.10/70.23 (16.99-287.05)	65.72/46.57 (29.25-216.24)	39.04/20.87 (18.90-88.56)	0.04	0.66	0.73	0.12

(1)results in the paper
(2)results computed by ourselves

Airflow insufficiency

For Maximum phonation time, we count the maximum one of two voice records of each patients. The tendency is similar, this symptom is more widely exists in APS patients than PD patients. However, it isn't a speech dimension discriminative compared to others.

Harsh Voice (Jitter, Shimmer, HNR)

We can conclude that we get similar results for these three speech dimensions in absolute values and tendency, both results confirm that APS patients manifest harsh voice more severe than PD patients. These speech dimensions are quite discriminative and jitter is more discriminative between MSA and PSP compared to (1), which maybe useful for the classification procedure.

Rapid AMR

The results are similar with the paper, indicating that this speech dimension is the least discriminative one compared to others.

Strained-strangled voice

The strained-stragled voice was determined using the degree of voicelessness(DUV). We obtained similar results that confirm APS patients manifest the strained-strangled voice more severe than PD patients and this speech dimension is very discriminative between PSP and MSA patients. In addition, it's also a discriminative speech dimension between MSA and PSP.

Slow AMR

We obtained similar results and can conclude that this is a discriminative speech dimension between APS patients and PD patients, the slow AMR DDK rate exists more widely in APS patients than PD patients.

Excess pitch fluctuations

The results obtained by us are quite different with the results in the paper, the scale of values are larger while the discrimination between MSA and PSP is disappear in our result.

Vocal Tremor

Our results confirm that the Vocal tremor is more severe in APS patients than PD patients which is similar to the paper, however, discrimination between PSP and MSA patients is less in our result compared to the paper.

Irregular AMR

Our result is a bit different compared to the results in paper, the discrimination is less for our result even though the tendency is similar that APS patients manifest the symptom of irregular AMR more severe than PD patients.

4.3 Estimation procedure

A free program called PRAAT for the analysis and reconstruction of acoustic speech signals is used to calculate these 10 speech dimensions. The calculation procedures of these 10 speech dimensions can be divided into 3 groups according to the scripts used.

4.3.1 Airflow insufficiency, Jitter, Shimmer, HNR, Strained-strangled voice, Excess pitch fluctuations

These speech dimensions are calculated based on standard autocorrelation to get pitch with default parameters, then get pulses with command (pulses = To PointProcess (cc)). By selecting sound file, pitch and pulses, we get voice report with default parameters in PRAAT, then we extract values as following in voice report.

Airflow insufficiency - MPT(Maximum phonation time): duration

Jitter: Jitter (local)

Shimmer: Shimmer (local)

HNR: Mean harmonics-to-noise ratio
Strained-strangled voice - DUV(degree of voicelessness): Fraction of locally unvoiced pitch frames
Excess pitch fluctuation - F0 SD: Standard deviation of pitch

4.3.2 Vocal Tremor

Measure of vocal tremor was based on frequency termor intensity index (FTRI) defined as the intensity/magnitude of the strongest low-frequency modulation of F0.
In script,the tremor frequencies are determined by auto-correlating the contours. If the highest autocorrelation coefficient that can be detected in the contour is smaller than the threshold (that can be set individually; standard value: 0.15), it is assumed that there is no tremor and therefore no tremor frequency nor intensity nor power – and the output will be 'undefined'. FTRI is computed by the average of local maximas and local minimas of contours after subtracting their linear fit in order to compensate for natural declinations.

The algorithm and the attached script is provided in [13], but there are some modifications since too many 'undefined' FTRI results if we use the default version, we change a little bit the parameters compared to default version for better results:
Minimal_pitch_(Hz) 60 -> 50
Voicing threshold 0.3 -> 0.25

4.3.3 Rapid AMR, Slow AMR, Irregular AMR

Rapid AMR was defined as the ratio for DDK rate of first half of /pa/-ta/-/ka/ utterance compared to the DDK rate of second half of /pa/-ta/-/ka/ utterance. Slow AMR was defined as the DDK rate measured as the number of syllables per second based on the first seven repetitions of the /pa/-/ta/- /ka/ syllables. Irregular AMR is based on the first seven repetitions of /pa/-/ta/-/ka/ syllables and was defined as the standard deviation of distances between following local maxima, representing the greatest energy during the performed /pa/, /ta/, or /ka/ syllable.

To compute these values, the key is to identify the positions of syllables, once we have the positions of syllables, the remaining calculation procedures are simple, just following the definitions. The syllables are considered as the local maximas of intensity contours of sound files. While not all local maximas are considered as syllables, there are two mainly filtering conditions:

1. Differences between two intensity peaks should be larger than 2dB.

2. The local maximas of intensity should be large enough to be considered as voicing part.

Then the positions of these filtered local intensity maximas are considered as the positions of syllables.

The algorithm and the script is modified based on [14], the positions of syllables are obtained by default version, then Rapid AMR, Slow AMR and Irregular AMR are computed according to the definitions[15].

5 Comparison of deviant speech dimensions characteristics

5.1 Introduction

In this section we provide and compare results computed by ourselves using PRAAT and by Prof. Etienne SICARD using VOCALAB and results in the [7]. Since each patient has two voice recordings, we computed each voice file and counting the average value for one patient except that we count the maximum value for MPT, while Prof. Etienne SICARD only computed the value of the first file of each patient except that the maximum value for MPT.

The standards provide ranges that if the speech dimension value falls in this range, this patient is considered as manifesting this symptom.

5.2 Table

NO.	Deviant speech dimension	Standards	source	PSP	MSA	PD
Hypokinetic						
1.	Airflow insufficiency		(1)	Common(42%)	Common(31%)	Common(27%)
		MPT $<$ 10s	(2)	Occasional(25%) 3/12	Occasional(23%) 3/13	Occasional(13%) 2/15
		MPT $<$ 10s	(3)	Common(33%) 4/12	Common(38%) 5/13	Occasional(15%) 2/13
2.	Harsh voice		(1)	Abundant(75%)	Frequent(69%)	Occasional(13%)
	Jitter	$>$ 1%	(2)	Common(33%)	Frequent(46%)	Rare(0%)
	Shimmer	$>$ 6%	(2)	Frequent(67%)	Frequent(69%)	Common(27%)
	HNR	$<$ 17dB	(2)	Frequent(67%)	Frequent(62%)	Occasional(13%)
3.	Rapid AMR		(1)	Occasional(25%)	Common(31%)	Occasional(13%)
		DDK acceleration ratio $<=$ 0.95	(2)	Rare(8%) 1/12	Occasional(15%) 2/13	Rare(7%) 1/15
		DDK rate $>$ 7 syll/s	(3)	Occasional(17%) 2/12	Rare(8%) 1/13	Common(33%) 5/15
Spastic						
4.	Strained-strangled voice		(1)	Common(42%)	Frequent(62%)	Rare(0%)
		Fraction of locally unvoiced pitch frames $>$5%	(2)	Occasional(17%) 2/12	Common(38%) 5/13	Rare(0%) 0/15

		Unvoiced/ Voiced > 5 %	(3)	Rare(17%) 2/12	Common(38%) 5/13	Rare(0%) 0/13
5.	Slow AMR		(1)	Frequent(50%)	Frequent(54%)	Rare(0%)
		DDK rate < 5syllables/s	(2)	Frequent(58%) 7/12	Common(38%) 5/13	Rare(0%) 0/15
		DDK rate < 5syllables/s	(3)	Common(42%) 5/12	Frequent(46%) 6/13	Rare(0%) 0/15
Ataxic						
6.	Excess pitch fluctuations		(1)	Common(33%)	Frequent(69%)	Rare(0%)
		SD F0 > 2 semitones	(2)	Abundant(83%) 10/12	Frequent(69%) 9/13	Common(40%) 6/15
		Prosody > 2 notes /a/	(3)	Frequent(50%) 6/12	Abundant(77%) 10/13	Rare(8%) 1/13
7.	Vocal tremor		(1)	Common(33%)	Frequent(54%)	Rare(0%)
		FTRI > 7%	(2)	Frequent(50%)	Frequent(46%)	Rare(7%)
8.	Irregular AMR		(1)	Common(33%)	Common(31%)	Rare(0%)
		Regularity > 50ms	(2)	Frequent(50%) 6/12	Frequent(62%) 8/13	Occasional(20%) 3/15
			(3)	Common(33%) 4/12	Frequent(46%) 6/13	Common(27%) 4/15

(1)results in the paper
(2)results computed by ourselves
(3)results computed by VOCALAB

5.3 Analysis

Airflow insufficiency

The result (2) and (3) are a bit different with (1) but could generally confirm that the airflow insufficiency is more severe for APS patients.

Harsh voice

We can't directly compare the results since we don't know the standards for paper but can confirm that the harsh voice is slightly more severe for PSP patients than MSA and much more severe than PD patients.

Rapid AMR

Results (2) and (3) are quite different with the result in paper, (2) shows even less difference between APS and PD patients and PD patients even manifest more severe rapid AMR symptom than APS patients in result (3). In general, it's not a speech dimension discriminative for APS and PD patients.

Strained-strangled voice

The results show similar tendency with the paper. Result (2) and (3) are less discriminative between APS and PD patients but still confirm that the Strained-strangled voice exists only in APS patients. In addition, our result show the difference between the MSA and PSP patients in this speech dimension.

Slow AMR

The results are quite similar and all confirm that slow AMR exists only in APS patients.

Excess pitch fluctuations

Result(2) is different with the (1) and (3), our result has higher values and lost the difference between MSA and PSP patients. All results confirm that the APS patients manifest more severe excess pitch fluctuations symptom than PD patients.

Vocal tremor

Both results confirm the tendency that APS patients manifest more severe Vocal tremor than PD patients but the difference between MSA and PSP patient is less compared to the paper.

Irregular AMR

Results(1) and (2) confirm that APS patients manifest more severe Irregular AMR than PD patients but according to the paper, this speech dimension never exists in PD patients which is different with (2). However, result (3) lost the discrimination between APS and PD patients.

6 Classification

6.1 Introduction

In following sections, we focused on differentiating between MSA and PSP since the accuracy obtained in [7] was 75% which could be improved. A classification experiment was performed using python with scikit-learn library to determine the best combination of acoustic features and the highest accuracy that can be achieved in differentiating between MSA and PSP. A support vector machine with a Gaussian radial basis kernel was applied and a leave one out cross-validation scheme was used, where the original data was separated into a training set contains all subjects exclude 1 (24 subjects in our case) and the test set contains only one subject. This process was repeated 25 times and each time test set contains different subject. The average percentage of correctly classified subjects into correspond group was considered as the performance of the model.

6.2 Preprocessing

Scaling the original data is highly recommended when applying a support vector machine since the support vector machine algorithms are not scale invariant and if a feature has a variance in much larger scale compared to others, it might dominate the objective function. Here the scaled data has zero mean and unit variance for each feature.

6.3 Feature selection and parameter tuning

During the cross-validation process, the automatic feature selection and the parameter tuning were performed at the same time. To determine the best combination of features, a naive method is to generate all the combinations of features for each cross-validation process. However, even our dataset is quite small, it still took much time since that parameter tuning was also necessary to be performed at the same time.

Thus, greedy search policy was implemented. Greedy search policy serves for searching in a large range of parameter tuning area, since it has much lower time complexity. Exhaustive search policy which means generate all the combinations of features used for searching in a relative small range of parameter tuning area and to decide the best combination of features.

Greedy search policy: In the beginning, all the two-features combinations were generated, and the one which has highest accuracy was chosen as the temporary optimal one. Then considering all the reset feature, each time one feature is added to the temporary optimal combination and then the combination which has the highest accuracy is chosen. Continue this process until the accuracy is decreased or unchanged when a new feature is added. This approach has much lower time complexity than the exhaustive search and can get approximate accuracy which helps us perform the parameter tuning with much smaller step.

Exhaustive search policy: It generates all the subset combinations of features so that can get the optimal result but has very high time complexity.

With greedy search, parameter tuning was performed in a large range primitively, then we focused on the smaller range which has higher accuracy. This process was repeated until the range is relatively small. Then the exhaustive search is used to determine the best combination of features performing fine-tuning.

6.4 Result

There were two combinations of three acoustic features able to discriminate PSP from MSA with an accuracy of 84% which is better compared to the 75% – the best accuracy obtained in [7]. The combination of features change with the parameters of SVM model.

Fig 1 describes the relation between the accuracy and the parameters.

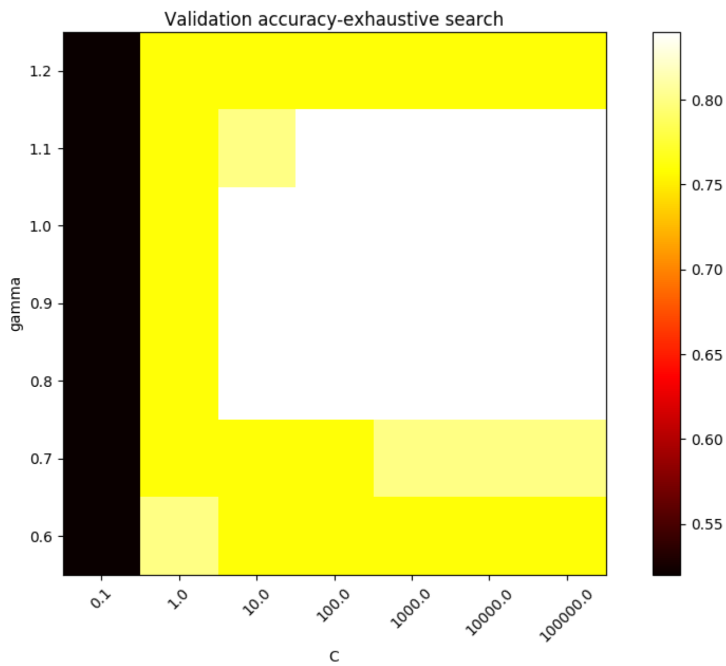

Fig.1

When gamma = 0.8, 0.9 and 1.0, C = 10, the three acoustic features including shimmer, HNR and vocal tremor were able to achieve an accuracy of 84% for discrimination between PSP and MSA.

When gamma = 0.8, 0.9, 1.0 and 1.1, C ranges from 100 to 100000, the combination of three features including jitter, strained-strangled voice (DUV) and vocal tremor were also able to discriminate PSP from MSA with an accuracy of 84%.

In this model, C trades off the misclassification of training samples against the simplicity of the decision surface, a large C makes the model try to classify all the the training correctly rather than have a big margin. Thus, keeping C in small value means that the distance from sample points to the decision surface would be large.

Since there are too few samples(12 PSP patients and 13 patients), it's very difficult to avoid the overfitting problem. As consequence C was set to a low value(<10) to keep smoother decision hyperplane to avoid overfitting problem as much as we can.

Thus, we focus on the C in low value range(from 1 to 10) as figure 2:

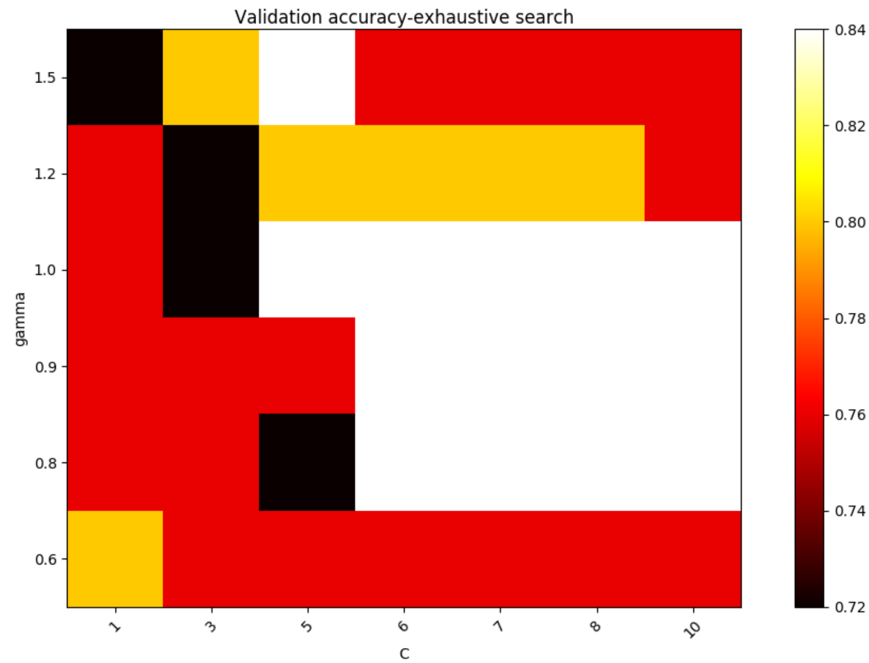

Fig.2

It can be observed that even C is smaller than 10, in the white area of the figure, with the three acoustic features including shimmer, HNR and vocal tremor, the 84% accuracy still can be achieved which means a relative simple decision hyperplane can possibly be obtained.

6.5 Discussion

Figure 3 is a 3d visualization figure plotted to observe more clearly the relation between the classified label and the combination of features. The misclassified patients were marked in the figure. Each time there were 4 misclassified patients of these 5 patients, the misclassified patients change when applying different values of sigma and C.

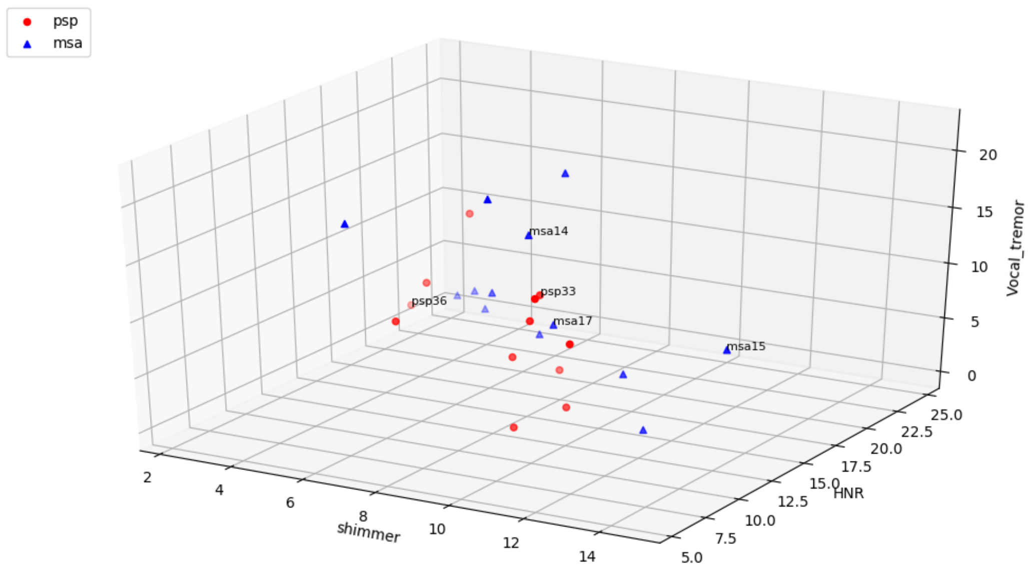

Fig.3

From the figure 3, it's hard to find a obvious decision surface but it can be noticed that the misclassified patient MSA17a is inside the PSP patients which is not normal and may have a great impact on decision surface. Thus MSA17a can be considered as a noise sample and a new classification process was implemented with the dataset excluding the MSA17a, the new 3d visualization figure is figure 4:

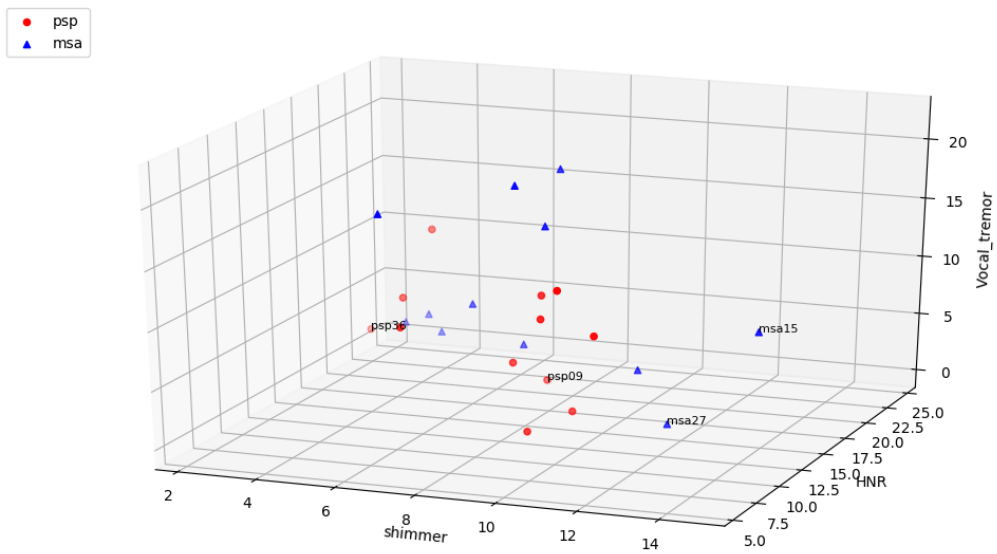

Fig.4

The best accuracy for this new classification process is 87.5% which is normal since one misclassified sample was excluded and the combination of features to achieve this accuracy stays at the same and the range of C and gamma values is almost the same as before. Each time there were 3 misclassified patients of these marked 4 patients, the misclassified patients change when applying different values of sigma and C. Generally, the classified samples were more separate compared to the figure 3, however the expecting result is that much more accuracy improvement can be gained since msa17 was considered as a noisy sample. And it's still hard to find a obvious decision surface since SVM may have a high dimension decision surface in high dimension feature space and there were too few patients so that any patient can't be ignored to get a rough decision surface.

As consequence, some linear analysis experiments were performed in the following to try to get a explicit linear decision boundary even the accuracy may not as high as 84%.

Part II

Deep analysis

In the previous experiment, only features from voicing and articulation recordings were used, here several new features were implemented based on monologue and reading recordings of database of paper[7]. And the maximum phonation time was excluded thus there are 13 acoustic measurements in total as followings:

categories of features	features
voicing	Jitter, Shimmer, HNR, Unvoiced degree(DUV), F0 SD, Vocal Tremor
articulation	Slow AMR, Rapid AMR, Irregular AMR
Prosody	Intraword pause ratio, number of pauses(No.Pauses), Percentage of pause time (PPT), Monopitch

In this part, we mainly focused on relations between features then try some dimension reduction methods as projection to observed data and linear classifier to separate MSA and PSP patients thus the result could be much easier to interpret.

7 Prosody features

7.1 Intraword pause ratio

The intraword pause ratio was computed based on the reading recordings and defined as the ratio of total pause time within the polysllabic words relative to the total pause time for all speech. The definition of "pauses" is the silence period lasts more than 10ms [4].

From the state of the art, the intraword pause ratio would be a discriminative feature between PSP and MSA, since according to the study[5], the ratio of intraword pauses is significantly decreased in PSP group compared to PD group while according to the study[6], the ratio of intraword pauses is not significantly different between MSA group and PD group.

To distinguish the inter-word pauses and intraword pauses, it's better to have a speech recognition system to detect word boundaries. However, it will go far away from our topic of internship and take too much time, I decided to distinguish the inter-word pauses and intraword pauses simply by the silence length since the study[4-6] didn't provide specific information about it. The intraword pause is mainly the voice onset time(VOT), which is a feature of the production of stop consonants. The typical length of VOT is less than 100ms in English, but may vary with different languages. Our voice recordings come from Czech native speakers, however no study really focused on the VOT for Czech. The threshold is set at 120ms which means the length of silence larger than 120ms is considered as the inter-word pause, otherwise the intraword pause.

7.2 Percentage of pause time and Number of pauses

These two features were examined using reading passage recordings. Percentage of pause time was defined as the ratio of total pause time of recordings relative to total speech time and number of pauses as the number of all pauses, where pauses are defined as silence longer than 60ms.[15]

7.3 Monopitch

Monopitch was calculated as the standard deviation of voice fundamental frequency(F0 SD). representing the variations of vibration rate of vocal folds. The computing procedure and algorithm to compute Monopitch was the same as the F0 SD in previous part except the recordings are monologue recordings rather than vowel recordings.

8 Principal component analysis

8.1 Introduction

Principal component analysis is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables to a set of values of linearly uncorrelated variables which are called principal components. This transformation makes the first principal component has the largest variance which means representing the variance of data as much as possible. In this part, the implementation of PCA was based on the FactoMineR library of software R.

8.2 Variables factor map

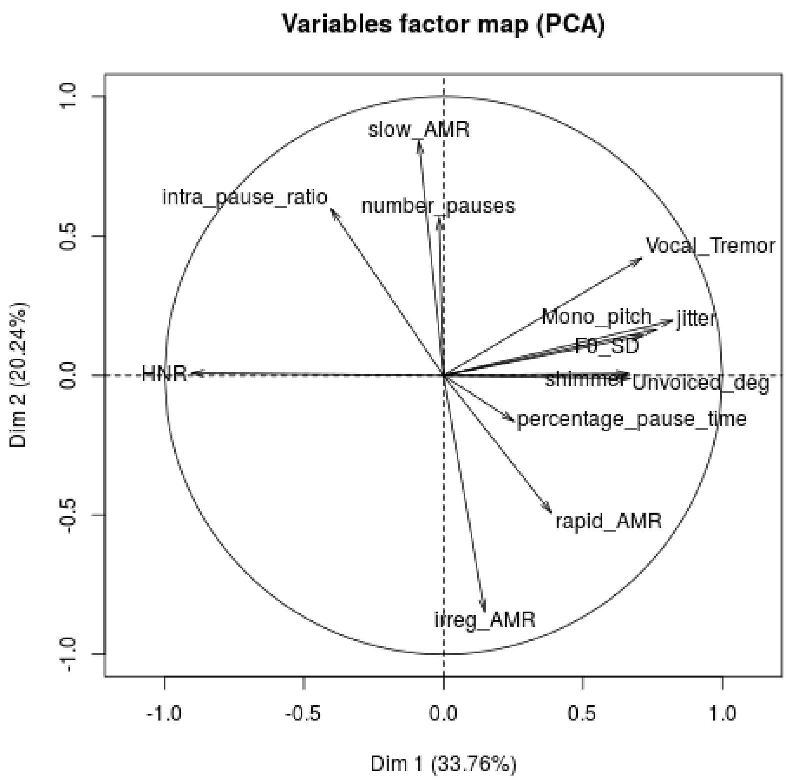

Fig.5

Fig.5 presents the relations between features and the first and second principal components. Dim 1 is the first principal component which accounts 33.76% variability of data and Dim 2 is the second principal component which accounts 20.24%

variability of data. From Fig.5, approximately, vowel features are almost in the first principal component while syllable and prosody features stay in the second principal component except the percentage pause time has very small variance in the first and second principal component and Monopitch is closer to the first dimension.

According to the linearly uncorrelated property of first and second principal component, it can be concluded that vowel features and the reset features are approximately linearly uncorrelated. Thus these 13 features can be divided into two parts: 1. Vowel features plus Monopitch. 2.Syllable features plus intraword pause ratio and number of pauses.

9 Linear discriminant analysis

9.1 Introduction

Linear discriminant analysis(LDA) is a feature reduction method used to find a linear combination of features so that the variance within the group is minimal while the variance between the group is maximal. Unlike PCA, LDA is a supervised method, the linear combination can be seen as a linear classifier or used for dimensionality reduction before later classification. The implementation of LDA experiment was based on LinearDiscriminantAnalysis function of scikit-learn library.

9.2 1d LDA projection

Firstly LDA was applied on three different categories of features (i.e., voicing, articulation and prosody) separately and on all features.

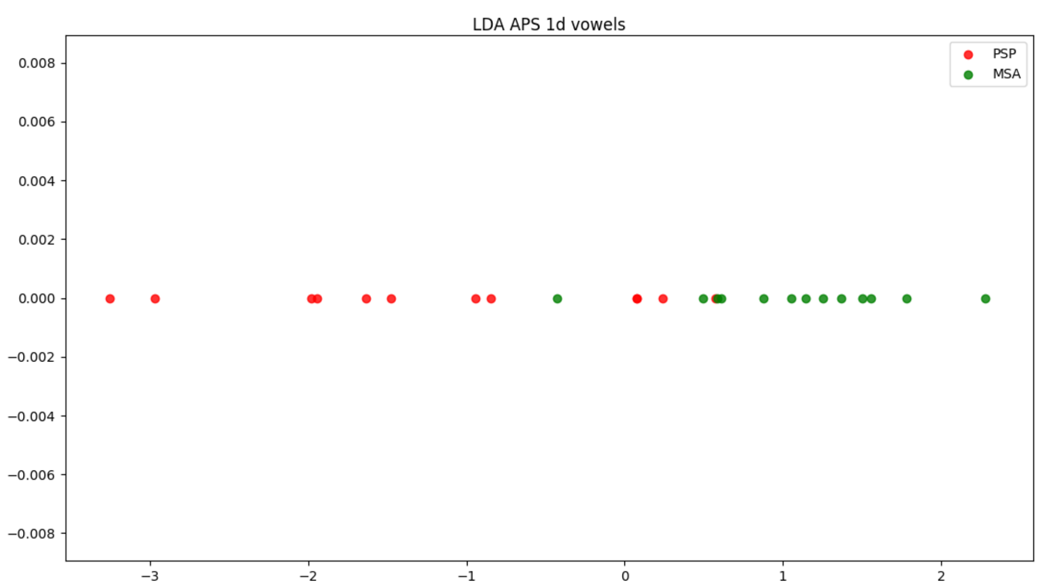

Fig.6

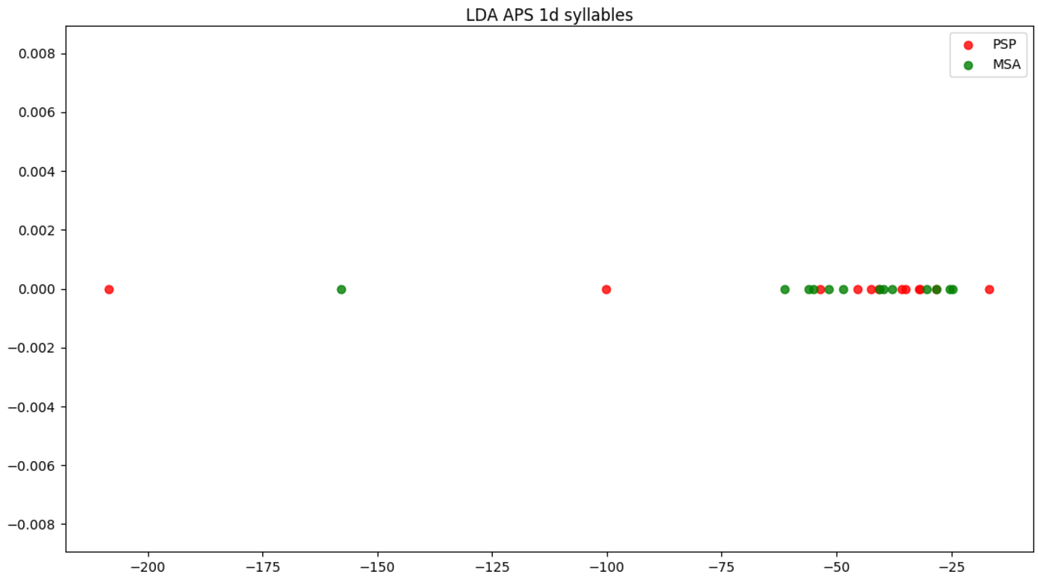

Fig.7

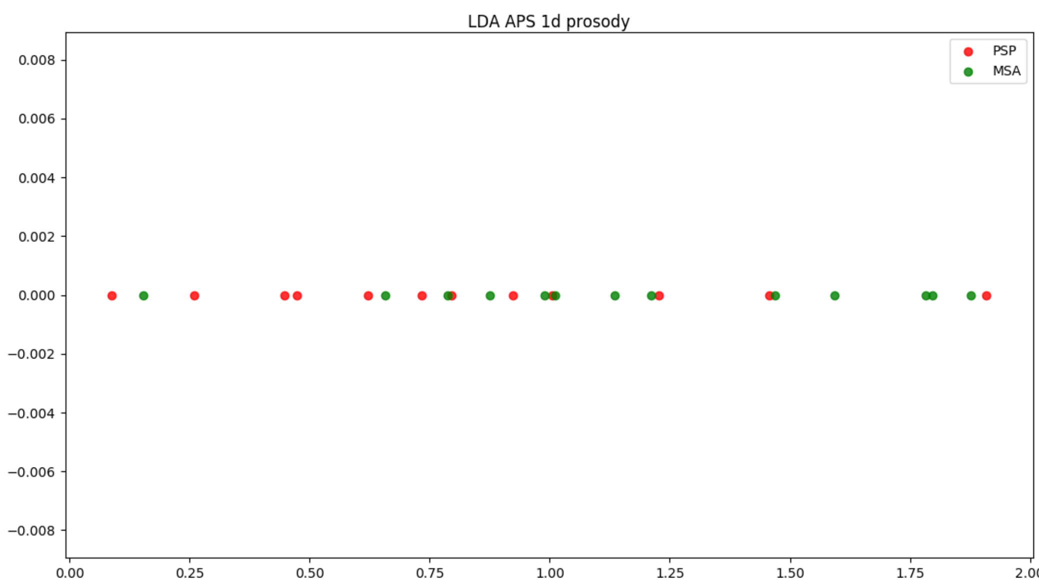

Fig.8

Figure 6, 7 and 8 present the projection values of LDA using voicing, articulation and prosody features separately and respectively. It can be concluded that in these three categories, only vowel features can differentiate PSP from MSA well which confirm Fig.5 that features close to first principal component are almost vowel features.

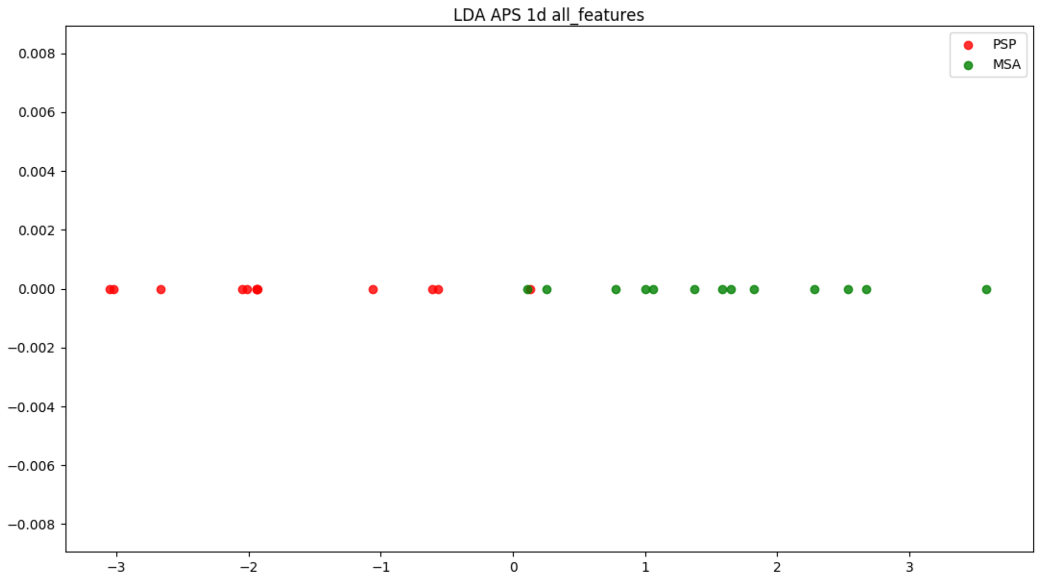

Fig.9

Figure 9 represents the projection values of LDA using all 13 features, the result is almost prefect. However, the curse of dimensionality should be considered since there are only 25 samples.

9.3 Classification on 1d LDA projection values

Consequently, a classification experiment was performed taken the all features projection values as input using SVM with linear kernel in a leave one out cross validation scheme which is the same with the previous classification experiment in the section 5. The implementation of SVM was based on scikit-learn library with svm.svc() class. In the following several classification experiments in section 8 and 9, same classifier and cross validation scheme were used and to expect a good generalization ability, C is always less than 10 in linear SVM.

As expected, overfitting problem occurs and an accuracy of 72% was achieved when C =1, which decreased greatly compared to the Fig.9.

9.4 2d LDA projection and classification

Inspired by the Fig.5, features were divided into two groups to perform LDA projection separately:

group	features
1	Jitter, Shimmer, HNR, DUV, F0 SD, Vocal Tremor, Monopitch
2	Slow AMR, Rapid AMR, Irregular AMR, Intraword pause ratio, No. Pauses

The Percentage of pause time was excluded since the mode of this vector was small both in first and second principal component. Then these two group's projection values are represented as two axis of figure which are reasonable since features in first group are approximately linearly uncorrelated with features in second group.

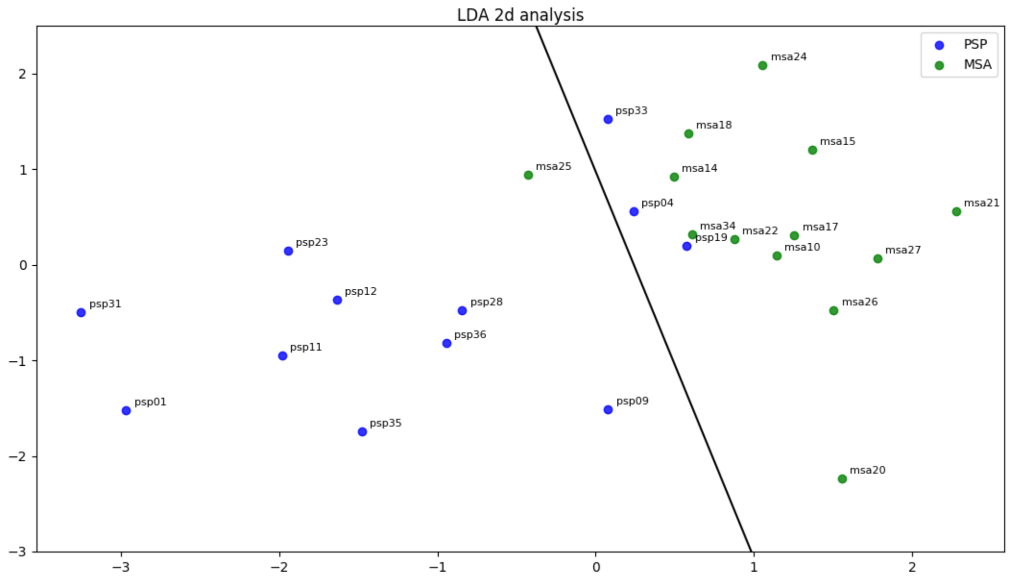

Fig.10

A good separation between MSA and PSP patients can be found in the figure 10. The black line represents the decision boundary which decided by a linear SVM with C=1.

Similarly, a classification experiment(same conditions as the one in 1d LDA classification) was performed on 2d LDA projection values. An accuracy of 68% was attained which was even worse compared to classification on 1d LDA projection values. Since less features were used for each dimension in LDA projection process, a better generalization ability was expected, however, the accuracy was even worse.

A possible reason is that the margin is not large enough especially for MSA samples. As consequence, one more learning layer was added to make our model has more learning ability to push the samples far away from decision boundary.

10 3-layers model

10.1 Logistic regression

Logistic regression is a regression model where the dependent variable is categorical. In our case, this binary logistic regression process was used to estimate the probability of MSA patient based on LDA projection values.

10.2 Model structure

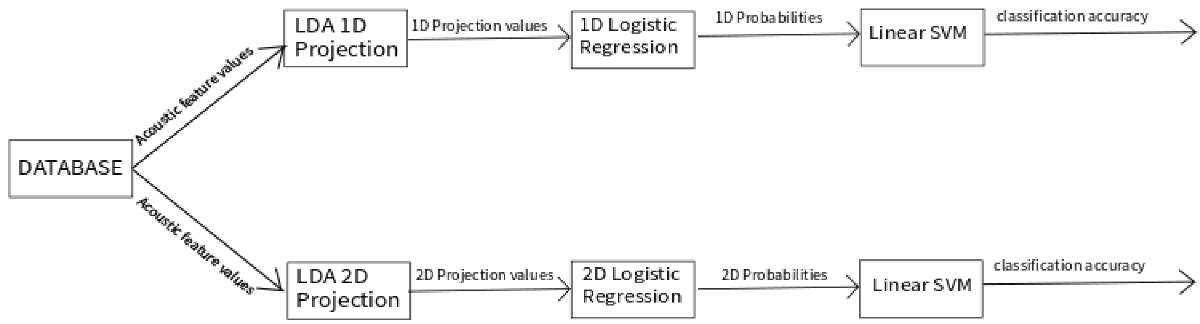

Fig.11

As Fig.11 describes above, logistic regression learning process was added between the LDA projection and Linear SVM classification process. The output of logistic regression layer is the probability of MSA which should close to 0 for PSP group and close to 1 for MSA group. Here, logistic regression regularization term C was set to 1 which means a good generalization ability can be expected.

10.3 1d Result

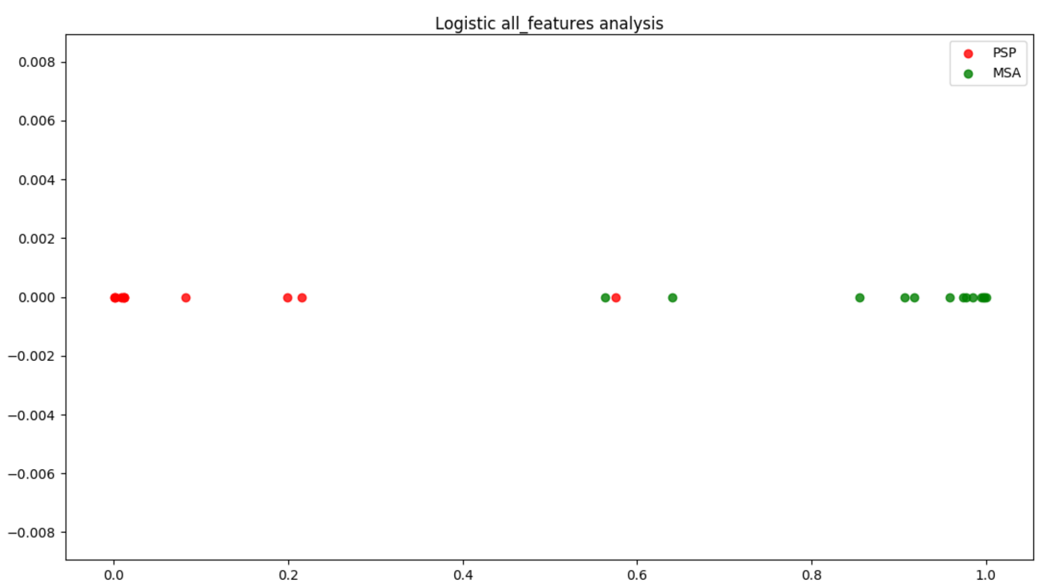

Fig.12

Comparing figure 9 and 12, it can be found that logistic regression layer 'push' the data into two sides which makes a better separation in probabilities compared to projection values. As described in upper path of figure 11, same classification process with section 8 was applied on 1d probability values. An accuracy of 72% was attained which is the same as the classification process on 1d projection values. For 1d(i.e., all features used), the adding logistic regression didn't make a contribution during classification process even though it makes training data more separate.

10.4 2d Result

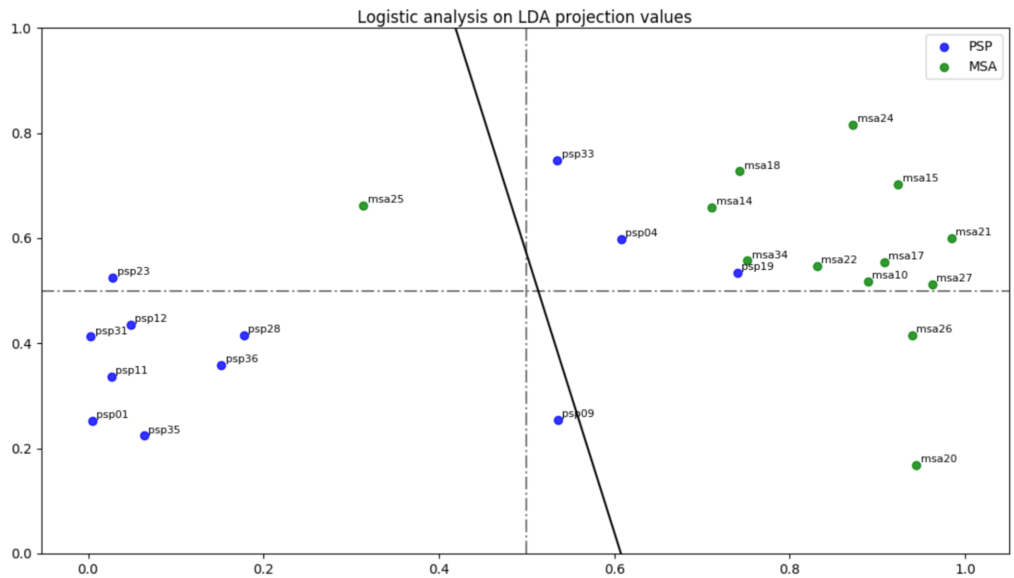

Fig.13

Compared by figure 10, the adding logistic regression layer has similar effect with 1d result, 'pushing' the data into left-down and right-up corner in figure 13. Thus 2d probabilities are more separated compared to 2d LDA projection values. Similarly, the black line represents the decision boundary which decided by a linear SVM with C=1.

Following the lower path of figure 11, same classification process was applied on 2d probability values. Here, an interesting result was achieved. An accuracy of 80% was attained with SVM regularization term C = 1, which improves 8% compared to perform classification process directly on LDA 2d projection values.

10.5 Discussion

Even though 80% accuracy is a little bit lower than the best accuracy 84% that achieved by using SVM based on rbf kernel, this structure only contains linear projection and linear classifier which greatly reduces the complexity of model and makes the interpretation easier.

2D 3-layers model achieved good accuracy proving that the method dividing features into two parts is reasonable. Here is the weights of features for 2D LDA projection.

2D LDA projection weights		
dimension	features	weights
1	Jitter, Shimmer, HNR, DUV, F0 SD, Vocal Tremor, Monopitch	1.42, 4.72, 6.30, 2.84, -0.95, 2.13, -1.18
2	Slow AMR, Rapid AMR, Irregular AMR, Intraword pause ratio, No.Pauses	-0.68, 0.18, -0.27, 0.58, 0.78

For first dimension, Shimmer, HNR, DUV and Vocal Tremor are more important than other features which partially confirm the PCA result and the optimal feature combination to achieve 84% accuracy that obtained by nonlinear SVM classifier.

For second dimension, weights of Slow AMR, Intraword pause ratio and NO.Pauses are larger than others.

Since combination of features found by LDA is a linear combination and logistic regression also is a linear projection, these combinations found in these two dimensions are more meaningful and can give much more information for the relations between features and diseases compared to the combination found by nonlinear SVM.

11 Conclusion

In first part, this study performed a comparison study which basically confirmed the conclusion in [7] and gained an 9% accuracy improvement(84% accuracy) on discriminating between PSP and MSA. However, this 84% accuracy was obtained by using a support vector machine with a Gaussian radial basis kernel. Since only 25 patients were acquired for our study, which are too few for this high complexity model, it's difficult to avoid overfitting problem. We thus seek obtaining an comparable accuracy with simpler classifier, which is the motivation of second part. In second part, 80% accuracy was achieved which still better than [7] by only using linear dimension reduction and linear classifier which proves a decent accuracy can be achieved by linear combination of certain features. Considering we never have too much data on this kind of problem, the second part shows a good path for future study. In addition, this result also provides us a deeper understanding of relationships among features. It proves that features can be divided into two approximately uncorrelated dimensions which is also an important conclusion for future objective acoustic study. The results of this internship will be submitted to IEEE-ICASSP'2018 (HTTP://2018.ieeeicassp.org/).

12 References

1. Ho AK, Iansek R, Marigliani C, Bradshaw J, Gates S (1998) Speech impairment in large sample of patients with Parkinson's disease. Behav Neurol 11:131–137.

2. Kluin KJ, Foster NL, Berent S, Gilman S (1993) Perceptual analysis of speech disorders in progressive supranuclear palsy.Neurology 43:563–566.

3. Kluin KJ, Gilman S, Lohman M, Junck L (1996) Characteristics of the dysarthria in multiple system atrophy. Arch Neurol 53:545–548.

4. Sachin, S., Shukla, G., Goyal, V., Singh, S., Aggarwal, V., Gureshkumar & Behari, M.(2008). Clinical speech impairment in Parkinson's disease, progressive supranuclear palsy, and multiple system atrophy. Neurology India, 56(2),122–126.

5. Skodda, S., Visser, W., & Schlegel, U. (2011). Acoustical analysis of speech in progressive supranuclear palsy. Journal of Voice, 25(6), 725–731.

6. Huh, Y. E., Park, J., Suh, M. K., Lee, S. E., Kim, J., Jeong, Y., ... Cho, J. W. (2015).Differences in early speech patterns between Parkinson variant of multiple system atrophy and Parkinson's disease. Brain and Language, 147, 14–20.

7. Rusz, J., Bonnet, C., Klempir, J., Tykalova, T., Baborova, E., Novotny, M., ... Ruzicka, E.(2015). Speech disorders reflect differing pathophysiology in Parkinson'sdisease, progressive supranuclear palsy and multiple system atrophy. Journal of Neurology, 262(4), 992–1001.

8. Muller, J., Wenning, G. K., Verny, M., McKee, A., Chaudhuri, K. R., Jellinger, K., ...Litvan, I. (2001). Progression of dysarthria and dysphagia in postmortem-confirmed Parkinsonian disorders. Archives of Neurology, 58(2), 259–264.

9. Tereza Tykalova , Jan Rusz , Jiri Klempir , Roman Cmejla , Evzen Ruzicka.(2017).Distinct patterns of imprecise consonant articulation between Parkinson's disease, progressive supranuclear palsy and multiple system atrophy. Brain and Language, 165, 1-9.

10. Darley FL, Aronson AE, Brown JR. Motor Speech Disorders. Philadelphia, Pa: WB Saunders Co; 1975.

11. Medical Dictionary for the Health Professions and Nursing © Farlex 2012

12. Nick Miller, Uma Nath, Emma Noble & David Burn. (2017). Utility and accuracy of perceptual voice and speech distinctions in the diagnosis of Parkinson's disease, PSP and MSA-P , Neurodegenerative Disease ManagementVol.7,No.3

13. Brückl, M. (2012): Vocal Tremor Measurement Based on Autocorrelation of Contours. In: Proceedings of the ISCA Conference Interspeech '12, Portland (OR).

14. Syllable Nuclei Copyright (C) 2008 Nivja de Jong and Ton Wempe

15. Rusz, J., Bonnet, C., Klempir, J., Tykalova, T., Baborova, E., Novotny, M., ... Ruzicka, E.(2015). Electronic supplementary material: Acoustic measurements for objective evaluation of motor speech disorders.