Detection of Persons with Parkinson's Disease by Acoustic, Vocal, and Prosodic Analysis

Tobias Bocklet [+,#1], Elmar Nöth [#2], Georg Stemmer [#3], Hana Ruzickova [%5], Jan Rusz [%,*4]

[+] *Phoniatric and Paedaudiologic Department, University Clinics Erlangen, Germany*
[#] *Chair of Pattern Recognition, University Erlangen-Nuremberg, Germany*

[*] *Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague*
[4] ruszjan@fel.cvut.cz

[%] *Department of Neurology and Centre of Clinical Neuroscience, Charles University in Prague, Czech Republic*
[5] hruzi@email.cz

***Abstract*—70 % to 90 % of patients with Parkinson's disease (PD) show an affected voice. Various studies revealed, that voice and prosody is one of the earliest indicators of PD. The issue of this study is to automatically detect whether the speech/voice of a person is affected by PD. We employ acoustic features, prosodic features and features derived from a two-mass model of the vocal folds on different kinds of speech tests: sustained phonations, syllable repetitions, read texts and monologues. Classification is performed in either case by SVMs. A correlation-based feature selection was performed, in order to identify the most important features for each of these systems. We report recognition results of 91 % when trying to differentiate between normal speaking persons and speakers with PD in early stages with prosodic modeling. With acoustic modeling we achieved a recognition rate of 88 % and with vocal modeling we achieved 79 %. After feature selection these results could greatly be improved. But we expect those results to be too optimistic. We show that read texts and monologues are the most meaningful texts when it comes to the automatic detection of PD based on articulation, voice, and prosodic evaluations. The most important prosodic features were based on energy, pauses and F0. The masses and the compliances of spring were found to be the most important parameters of the two-mass vocal fold model.**

I. Introduction

Parkinson's disease (PD) is a degenerative disorder of the central nervous system. It results from the death of dopamine-containing cells in the substantia nigra, a region of the mid-brain. The cause of cell-death is unknown. PD accounts for a variety of motor and non-motor deficits and is the second most common neurodegenerative disorder after Alzheimer's disease [1]. The most obvious motor symptoms are shaking, rigidity, slowness of movement, difficulty with walking and gait, and communication. Non-motor symptoms affect the sensory system, sleep, and emotion. Medical treatment alleviates certain symptoms, but there is no causal cure now available, and early diagnosis is critical for maximizing the effect of treatment and improving the quality of the patient's life [2].

The alteration of speech in PD, known as hypokinetic dysarthria, is present in between 70 % and 90 % of all PD patients [3]. Vocal impairment is most likely one of the earliest indicators of the disease [4]. Phonation is the most affected part of speech production [5], followed by articulation [6] and prosody [7]. Different acoustic studies commonly revealed a reduced loudness, monoloudness and monopitch, breathiness and harshness, highly variable speech rate, speech disfluencies, reduced stress, and reduced range of articulation movements [6] with deficits in prosody being one of the most notable signs in early stages of PD [8]. When PD advances, these parameters worsen to a point where the voice is often neither audible nor intelligible [9].

There are several speaking tasks that could be used to evaluate the extent of speech and voice disorders in PD. The most traditional of them including sustained phonation, rapid syllable repetition, and variable reading of short sentences, longer passages or freely spoken spontaneous speech [10].

In this work we focus on an automatic detection of PD speakers based on different systems that use phonation, articulation and prosody features. Phonation is modeled by a glottal excitation system based on two-mass vocal fold modeling, articulation is modeled by spectral features followed by statistical modeling, and the prosody of a speaker is evaluated by prosodic analysis.

A glottal excitation system allows a (distinct) analysis of voice and phonation. The approach is based on the source-filter model of the speech generation process. Voiced speech sounds are generated by the excitation signal, i.e., the source signal of the glottis, which is modeled by a two-mass-spring model. This signal is filtered by the vocal tract, where different frequencies are amplified or softened. The influence of the vocal tract has to be omitted in order to allow meaningful voice evaluations. As an approximation of the excitation signal, the residue of the Linear Predictive Coding (LPC), an inverse filtering of the speech signal with the LPC filter is calculated in a data-driven optimization procedure. The model parameters are now optimized to match the synthetic excitation signal as close as possible to the LPC residue and the estimated pitch. The final parameters are then analyzed in order to differentiate between healthy voices and voices affected by PD. Articulation modeling is achieved by a statistical modeling of acoustic features (Mel Frequency Cepstrum Coefficients (MFCCs)). The statistical modeling is achieved by Gaussian Mixture

Models (GMMs). The mean vectors of these GMMs act as an acoustic representation of this speaker. Prosody is evaluated by a prosodic analysis based on a voiced/unvoiced (VUV) decision. Different structured prosodic features are calculated for the voiced segments. Our prosody module analyzes F_0, energy, duration, pauses, jitter and shimmer and different functionals of them. Support Vector Machines (SVM) are used for classification. We evaluate the different modeling approaches with respect to the different kinds of speech tests (sustained phonation, syllable repetitions, read texts and monologues) and to determine the most expressive features by feature selection.

The outline of this paper is as follows: The data and the different speech tasks are described in Sec. II. The articulation modeling is presented in Sec, III. The prosodic system is described in Sec. IV and the glottal excitation system is described in Sec. V, Details about classification and feature selection are given in Sec. VI. Results are presented and discussed in Sec. VII. The paper is finished by a short summary (Sec. VIII).

II. Data

A. Patients

Data was used from the original study of Rusz et al. [8]. A total of 46 Czech native speakers participated in this study. Twenty-three individuals were diagnosed with an early stage of idiopathic PD. All PD patients were examined in the drug-naive state, immediately after the diagnosis was made and before the symptomatic treatment was started. Their mean age was 61.7 years (SD = 12.6), mean duration of PD symptoms prior to examination was 30.2 months (SD = 22.1), mean global motor score according to the Unified Parkinson's Disease Rating Scale III (motor rating scaled from 0 to 108, where 108 represents severe motor impairment) was 17.5 (SD = 7.3), and the stage of disease according to the Hoehn and Yahr scale ranged from 1-2 (disability scale comprised of stages 1 through 5, where 5 is most severe). In addition, 23 healthy control speakers with no history of neurological or communication disorders were included. Their mean age was 58.1 years (SD = 12.9). Age distribution showed no significant differences between both groups.

The speech data was recorded in a quiet room with a low ambient noise level using an external condenser microphone placed at approximately 15 cm from the mouth. The voice signals were sampled at 48 kHz, with 16-bit resolution.

B. Speech Tasks

Table I details the speech data used in this study. The speaking tasks ranged from producing isolated vowels to reading short sentences and producing a spontaneous monologue on a given subject. In each speaking task, the best speech performances for every subject were retained. See [8] for a comprehensive description of the data and recording procedure.

task	description
T1	Sustained phonation of /i/ on one breath at a comfortable pitch and loudness as constant and long as possible, at least 5-sec.
T2	Rapid steady /pa/-/ta/-/ka/ syllables repetition on one breath as constant and long as possible, repeated at least 5-times.
T3	Reading the same standard text of 136 word.
T4	Monologue, at least approx. 90-sec.
T5	Reading the same text containing 8 variable sentences of 71 words with varied stress patterns on 10 indicated words.
T6	Reading 10 sentences according specific emotions in a comfortable voice in response to an emotionally neutral sentence.
T7	Rhythmically read text containing 8 rhymes of 34 words. following the example set by the examiner.

TABLE I
DESCRIPTION OF THE DIFFERENT SPEECH TASKS USED IN THE EXPERIMENTS.

III. Articulation Modeling

Gaussian Mixture Models (GMMs) model acoustic features, namely Mel Frequency Cepstrum Coefficients (MFCCs) in a statistical way. For acoustic feature extraction a Hamming window with a size of 25 ms and a time shift of 10 ms is applied to the speech signal. Afterwards the Mel-spectrum with 26 triangular filters is calculated and processed by Discrete Cosine Transform (DCT). We take the first 13 Mel-frequency Cepstral coefficients including C_0. Cepstral mean subtraction (CMS) is applied and first- and second order derivatives of these features are calculated over a context of 5 and 9 consecutive frames. In the end a 39-dimensional feature vector is created. This feature vector is then modeled by GMMs. For each speaker one GMM is created by GMM-UBM modeling. After extraction of the spectral features a Universal Background Model (UBM), i.e., a class-independent GMM with 128 Gaussians, is trained on the whole training set using the Expectation-Maximization (EM) algorithm. The means of the UBM are adapted by relevance *Maximum A Posteriori* (MAP) adaptation in order to get speaker specific GMMs. The means of each speaker are then used as speaker-specific features, which forms 4992-dimensional (128×39) feature vectors.

IV. Prosodic Modeling

The prosodic system is not based on any speech recognition output or forced time alignments. Thus, the prosodic features are calculated whenever a voiced speech segment is found. The voiced-unvoiced (VUV) decision is based on the zero crossing rate, the normalized energy of the signal and the maximum energy.

Prosodic base features are calculated on the whole utterance. These are, fundamental frequency (F_0), energy, VUV segments and pitch periods. The structured prosodic features are calculated on the voiced segments. Adjacent segments are merged, when they are separated by less than 50 ms; the corresponding F_0 contour is interpolated to make the segmentation more robust. Context segments, that merge two adjacent segments together, are used additionally. All in all 73 features are calculated for each segment. They model F_0, energy, duration, pauses, jitter and shimmer. Note that the F_0

features are normalized w.r.t. the mean F_0 and transformed to semitones in order to be comparable across gender. A detailed description of the whole feature set is given in [11]. Finally, we compute mean, minimum, maximum and standard deviation of these 73 segment features. This forms our 292-dimensional prosodic feature vector.

V. Glottal Excitation Modeling

A. Two-Mass Model

The approach estimates the parameters of a physical glottis model. The goal is to find pathology-related changes in the model parameters that reflect voice-related parameters in order to detect speakers suffering from PD. Therefore, the used glottis model should ideally have physically meaningful parameters, in contrast to just describing the shape of the excitation signal. The model should be flexible enough to adequately represent pathology-related changes of the voice.

Considering these requirements we employed the two-mass vocal fold model introduced by Stevens [12] and illustrated in Fig. 1. The model consists of two pairs of masses, larger

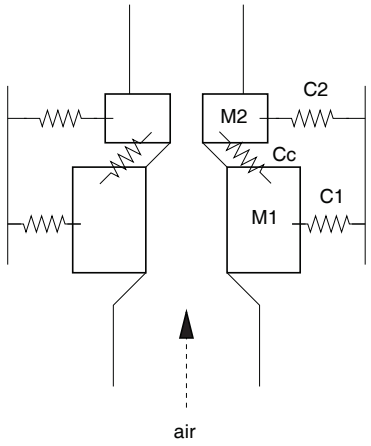

Fig. 1. Two-mass vocal fold model by Stevens [12].

ones (M_1) representing the inferior part of the vocal folds, and smaller ones (M_2) representing the superior part of the vocal folds. The model is symmetric, i.e., there is no differentiation between the masses of the left and right side. The mechanism depends on the fact, that the inferior and superior part of the vocal folds do not move together as a rigid body. There is a certain degree of freedom to move relatively to each other [13]. This freedom is modeled with a coupling compliance by springs. Each mass moves on a spring that is connected with the lateral wall. The masses are connected among themselves by an additional spring. The compliances of the springs are described by the parameters C_1, C_2 and C_c (for the spring that connects M_1 with M_2). Note that parameters for the masses and compliances are given as *mass per unit length* and *compliance per unit length*, i.e., they may change when the vocal folds are stretched. Air flows from bottom to top through the glottis when both M_1 and M_2 have a positive displacement, as shown in Fig. 1.

The excitation function of the two-mass vocal fold model by Stevens is obtained in three steps. First, the displacements $x_1(t)$ and $x_2(t)$ of the inferior and superior part of the vocal folds over time t are computed. The width of the glottal opening $d(t)$ is defined to be $\min(x_1(t), x_2(t))$. Second, from the width of the opening, the airflow $U_g(t)$ through the glottis is determined. In the third step, taking the derivative of $U_g(t)$ results in the excitation function.

The whole process of the excitation function computation is described in Chapter 2 of [12]. However, some details cannot be found in the book. In [14] a detailed derivation of all model formulas is given. The initial and fixed values for all parameters are taken from [12] and summarized in [14].

B. Model Optimization

Our hypothesis is that glottis model parameters contain information about the pathology of PD speakers. To test this hypothesis, we find the optimal model parameters that fit the speech data and observe how they change between healthy speakers and speakers with PD.

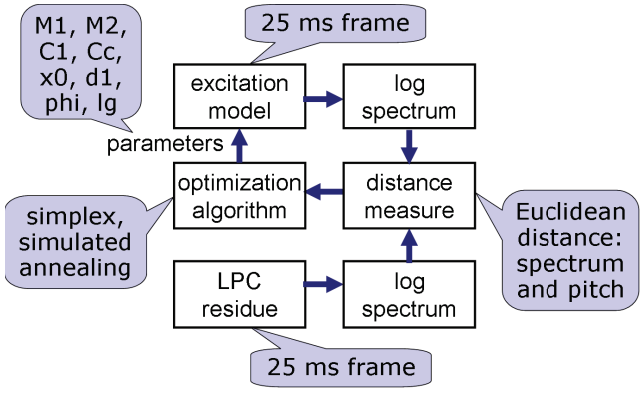

Fig. 2. Optimization of the parameters of the glottal excitation model

Figure 2 depicts a block diagram of the optimization loop. A set of initial parameters (M_1, M_2, C_1, C_c, x_0, d_1, ϕ, l) is the input of the glottis excitation model. The model generates an excitation signal for a 10 ms speech frame. At the same time, the LPC residue of the original speech signal is calculated and the log spectrum transform is applied to both of these excitation signals. The similarity of the generated excitation signal is compared to the original signal using two Euclidean distances. The distance between the log spectrum of the two signals is compared in a first step. In a second step, the distance between the generated and the original pitch for the frame are compared. The combined distance measure is passed to the optimization algorithm, which modifies the parameter set, passing the new parameter set to the excitation model. Thus, an optimization loop is formed, modifying the parameters, generating a new candidate excitation signal, and testing it against the original signal. The simplex algorithm [15] and simulated annealing [16] are used for optimization.

param	description
M_1	mass of inferior part of the vocal fold
M_2	mass of superior part of the vocal fold
C_1	compliance of spring between M_1 and lateral wall
C_c	compliance of spring between M_1 and M_2
d_1	average vertical length of the lower portion of the vocal fold
x_0	resting position of M_1 in the absence of any force
ϕ	skewness factor; constriction of the vocal tract
l	length of glottis (assuming rectangular shape)
D	optimization distance measure (see Eq. 2)

TABLE II
DESCRIPTION OF THE PARAMETERS OF THE GLOTTAL EXCITATION MODEL

task	GMM		PROS		GLOTTAL	
	% REC	AUC	% REC	AUC	% REC	AUC
T1	85.7	0.90	57.1	0.69	59.5	0.58
T2	83.3	0.88	78.6	0.89	61.9	0.57
T3	85.7	0.88	**90.5**	0.94	**78.6**	**0.87**
T4	78.6	0.82	88.1	**0.97**	71.4	0.77
T5	**88.1**	0.88	66.7	0.86	69.1	0.75
T6	85.7	0.90	76.2	0.86	**78.6**	0.86
T7	83.3	**0.94**	85.7	0.91	52.4	0.53

TABLE III
RECOGNITION RESULTS (% REC) AND AREA UNDER THE ROC-CURVE (AUC) OF THE THREE DIFFERENT SYSTEMS (GMM,PROSODIC,GLOTTAL EXCITATION) ON THE DIFFERENT TASKS (T1-T7)

The optimization is formulated as:

$$\begin{aligned} \hat{\theta} &= \underset{\theta}{\operatorname{argmin}}[D(s_m(\theta), s_{org})] \\ \theta &= \{M_1, M_2, C_1, C_c, x_0, d_1, \phi, l\} \end{aligned} \quad (1)$$

where $D(s_m(\theta), s_{\mathrm{org}})$ is the combined distance between the model excitation signal s_m and the original excitation signal s_{org}.

The distance measure combines distances between both the respective log spectra and the respective pitches p_m, p_{org}, and is defined as:

$$\begin{aligned} D(s_m(\theta), s_{\mathrm{org}}) &= D(\mathrm{logspec}(s_m(\theta)), \mathrm{logspec}(s_{\mathrm{org}})) \\ &\quad +\lambda \cdot D(p_m, p_{\mathrm{org}}) \end{aligned} \quad (2)$$

where $D(\cdot,\cdot)$ is the Euclidean distance between two vectors and the constant λ scales the influence of the pitch distance. Note that the optimization is only performed on voiced speech segments. TABLE II contains a description of all parameters of the glottal excitation model. The exact derivation of the formulas and parameters is omitted here. For a description see [12] and [14].

VI. CLASSIFICATION AND FEATURE SELECTION

Classification is done for either way of modeling by Support Vector Machines (SVMs). In order to account for the low number of participants of this study, experiments were performed by cross-validation in a leave-one-speaker-out (LOO) manner. SVM was used from the weka toolkit [17] with standard parameter settings to avoid overfitting.

Feature selection was performed in order to determine the features of the three different approaches that discriminate best between speakers with PD and healthy speakers. We decided to use a correlation based approach that evaluates the worth of features by considering the individual predictive ability of each feature along with the degree of redundancy between them. The correlation-based feature selection (CFS) approach prefers subsets of features that are highly correlated with the class while having low inter-correlation [18]. The evaluation function CFS_S describes the heuristic "merit" of a feature subset S containing k features. r_{cf} accounts for the mean feature-class correlation ($f \in S$), and r_{ff} is the average feature-feature inter-correlation with r being the Pearson correlation coefficient:

$$CFS_S = \frac{k r_{cf}}{\sqrt{k + k(k-1) r_{ff}}} \quad (3)$$

CFS has to be combined with a search strategy. We decided to use a simple forward selection. Forward selection begins with no feature and adds a single feature at a time. This step is repeated until no possible single feature addition results in a higher CFS_S.

VII. RESULTS AND DISCUSSION

We present the results of the three different systems in Sec. VII-A. Results after feature selection are presented in Sec. VII-B.

A. Results of the different systems

TABLE III shows the results of the three different systems. One can see that the GMM system achieves quite balanced recognition rates for the different tasks. None of the differences achieved with the GMM system are significant. However, the best recognition result (88.1 %) is achieved when evaluating the recordings of task T5, which is a reading task. Note that in T5 the participants were instructed to produce an unnatural intonation on a given word. This could be a problem for persons with PD because they have problems with imitating a given stress/intonation pattern. Task T7, a reading task with repeating heard rhymes achieves the highest "area under the receiver operator characteristics (ROC)-curve" (AUC, 0.94). Again and also for T6, the problems of people with PD to imitate a given pattern could be an explanation for the good results. The GMM system achieves the lowest recognition rate and AUC with task T4, where the patients are asked to give a short monologue. GMMs rely on acoustic information in form of MFCCs. Task T4 is the only task, where the patients are not asked to repeat a text. Thus, this task has a higher degree of freedom, since speakers will utter different words. This is problematic when comparing the acoustics/articulation of different speakers.

The prosodic system achieves a recognition rate of 90.5 % when using the recordings of task T3 (reading task). The highest AUC (0.97) is achieved with task T4, a monologue of approx. 90 sec. Fig. 3 contains the ROC curve for this task. Task T1 contains an isolated vowel, so prosodic features

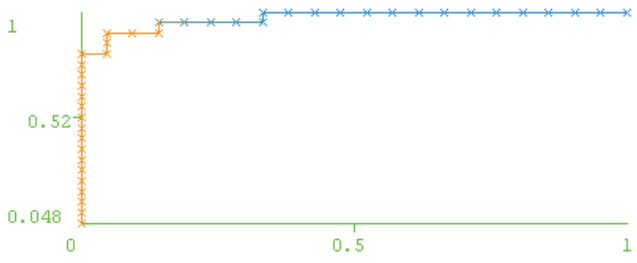

Fig. 3. ROC curve of the prosodic system on task T4 (monologue); AUC=0.97

are expected to produce a low recognition rate (57.1 %). The recognition results of prosody achieved on task T5 are unexpectingly low, especially because of the good results with the GMM system on T5 and the good results of both systems on T7. One might expect that task T2 (repeating steady syllables) should achieve higher recognition results. However, generally speaking the tasks containing read speech and monologues (T3-T7) achieve higher recognition results because they contain more data than isolated vowels or repeated syllables.

The glottal excitation system achieves the best recognition and AUC results on the reading tasks T3 and T6. In task T6 the persons had to repeat sentences with acted emotions. We believe, that the results for the glottal excitation system are higher for this task, because the vocal folds play a strong role in expressing emotions and the capability to express emotions is strongly affected by PD. Task T7, where the speakers are asked to repeat different rhymes, seems not to contain any discriminative glottal parameters.

It is too early to decide, what task would be best for a screening scenario for early detection of PD. The database is too small and the results should be verified on a larger database and across languages. Screening tasks should be short and cheap, i.e., doable without the need of a supervising expert to be present. The best results across all systems were achieved with the standard reading task T3. This is also the easiest and most natural task and could be realized in an unsupervised recording procedure, i.e., people are asked to read a passage over the phone or via an internet-based client/server-system. The spontaneous task requires more supervision, because one has to make sure, that the monologue is long enough and it might be necessary to encourage the tested person to continue. For the imitation tasks (T5-T7) the results are mixed. It is not clear at this moment, why the voice and prosody analysis is worse and less consistent compared to both voice and prosody analysis of T3 and T4 and acoustic analysis of T5-T7. Again, the supervision requirements are higher, because it is an unnatural task to repeat given intonation, stress, and emotion patterns. Tasks T1 and T2 are probably too short to provide enough speaker specific data.

Almost all task/classification-system constellations produce results which are significantly above chance, so no task can be ruled out for a detailed diagnosis or detailed screening procedure, but for a fast and cost-efficient screening task T3 (reading a standard text) seems to be the best choice.

B. Results after feature selection

In order to obtain the parameters of the prosodic and glottal excitation system that are the most meaningful parameters for discriminating between healthy persons and PD patients, we conducted feature selection experiments. For the sake of completeness we also performed feature selection on the features of the GMM system, i.e., the mean vectors of the Gaussian densities. Please keep in mind, that we performed the feature selection on the same dataset, that we used for the LOO experiments. Thus, recognition results after feature selection might be too optimistic and we do not report on any achieved improvement after feature selection. In theory, adding useless features to a classification system should not degrade the performance of a recognizer if the dataset is big enough. Thus, the kind of selected features might be of higher interest.

Feature selection on the GMM system achieved a higher recognition rate than before and is even more balanced over the different tasks than without feature selection. One interesting fact is, that the feature selection approach selects only approx. 50-60 out of 4992 features. After analyzing the selected features with respect to the different tasks, we found that the feature selection mostly selected features of the second derivative, i.e., the mean vectors of coefficients between 27 and 39, and some lower coefficients, i.e., the mean vectors of MFCC1-MFCC4.

The feature vector of the prosodic system contains 12-17 out of 292 features after feature selection. Improvements have been achieved for tasks with a low performance. When analyzing the types of selected features, one sees a clear trend: For the reading and monologue tasks (T3 - T6) various energy features and some F0 features have been selected. Note, that our prosodic feature vector contains 73 different features calculated on voiced segments and we calculated min, max, mean, and stddev features for each speaker. The F0 features were selected from the mean values, energy features were selected from min, max and mean in similar portions. Various studies found out, that PD patients speak in a diminished loudness and monopitch. That perfectly explains the kind of selected features. Task T1 (sustained phonation of /i/) relies on energy features and voiced/unvoiced proportion features. The explanation for this also is associated with the diminished loudness and with the difficulty in producing a sustained vowel, which leads to unvoiced decisions within the vowel. For task T2 (rapid repetition of steady syllables) only minimum and maximum energy features where selected. For this task it is known, that the loudness of PD patients diminishes much faster when repeating steady syllables than the loudness of normal speaking persons.

Only a few glottal features are selected automatically, with the correlation-based feature selection approach for the different tasks. Note, that the feature set contained minimum,

task	GMM		PROS		GLOTTAL	
	% REC	# feat.	% REC	# feat.	% REC	# feat.
T1	90.5	61	76.2	12	66.7	1
T2	97.6	58	85.7	12	71.4	4
T3	92.9	51	**88.1**	12	**83.3**	5
T4	97.6	47	85.7	17	66.7	1
T5	97.6	58	81.0	14	73.8	3
T6	97.6	63	**88.1**	17	76.2	3
T7	**100.0**	59	85.7	16	59.5	2

TABLE IV
RECOGNITION RESULTS (% REC) AND NUMBER OF SELECTED FEATURES (# FEAT.) OF THE THREE DIFFERENT SYSTEMS (GMM,PROS,GLOTTIS) ON THE DIFFERENT TASKS (T1-T7) AFTER FEATURE SELECTION

maximum, mean and stddev functionals of the basic features of the glottal excitation system In each of the reading tasks (T3,T5,T6 and T7) minimum and maximum values for at least one of the masses and one of the different spring compliances are selected. (see Fig. II for details). We expect these features to contain information about the harshness and the breathiness of the voice and about the reduced loudness.

VIII. SUMMARY

In this work we focused on the automatic discrimination between healthy speakers and speakers within early stages of PD. We tried to identify the speech tasks with the most meaningful acoustic, prosodic, and vocal information to achieve this discrimination. We found out, that read speech and monologues contain the most important acoustic, prosodic and vocal information, when it comes to an automatic detection/discrimination of PD speech. With an acoustic system we achieved a recognition rate of 88 %, with n voice modeling system we achieved 79 % recognition rate. The best results (90.5% recognition rate and 0.97 AUC) was achieved with a prosodic system for the detection of PD speakers in early stages. For a fast and cost-efficient screening procedure reading a standard text seems to be most promising. Feature selection experiments revealed, that the most important prosodic features rely on energy pauses and F0. The masses and compliances of the springs where identified to be the most important features of the glottal excitation system. Especially the results achieved by the glottal excitation system are quite promising. However, the reported results on this system are at an early stage, since this system was newly developed and not applied to pathologic speech before. In future work we will focus on an improvement of the glottal excitation modeling and on a combination of the three approaches, since the different ways of modeling should add complementary information when combined.

ACKNOWLEDGEMENT

We are obliged to doctors Jan Roth, Jiri Klempir, Veronika Majerova, and Jana Picmausova for provision of clinical data, and Evzen Ruzicka for the concept and overall coordination of the clinical study. The work was partially funded by the German Research Foundation under grant EY 15/18-2 and by the Czech Ministry of Education, research program MSM 0021620849. The algorithm for the parameter estimation of the glottis model has been developed and implemented by Georg Stemmer and Andrew Cassidy at the Johns-Hopkins-Summerworkshop 2008.

REFERENCES

[1] A. E. Lang and A. M. Lozano, "Parkinson's disease," *New England Journal of Medicine*, vol. 339, no. 15, pp. 1044–1053, 1998.

[2] N. Singh, V. Pillay, and Y. E. Choonara, "Advances in the treatment of parkinsons disease," *Progress in Neurobiology*, vol. 81, pp. 29–44, 2007.

[3] J. A. Logemann, H. B. Fishe, B. Boshes, and E. R. Blonsky, "Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of parkinson patients," *Journal of Speech and Hearing Disorders*, vol. 43, no. 1, pp. 47–57, 1978.

[4] B. Harel, M. Cannizzaro, and P. J. Snyder, "Variability in fundamental frequency during speech in prodromal and incipient parkinson's disease: A longitudinal case study," *Brain and Cognition*, vol. 56, no. 1, pp. 24–29, 2004.

[5] M. K. MacPherson, J. E. Huber, and D. P. Snow, "The intonation-syntax interface in the speech of individuals with parkinson's disease," *J Speech Lang Hear Res*, vol. 54, no. 1, pp. 19–32, 2011.

[6] R. D. Kent and Y. Kim, "Toward an acoustic typology of motor speech disorders*," *Clinical Linguistics and Phonetics*, vol. 17, no. 6, pp. 427–445, 2003. [Online]. Available: http://informahealthcare.com/doi/abs/10.1080/0269920031000086248

[7] A. W. Darkins, V. A. Fromkin, and D. F. Benson, "A characterization of the prosodic loss in parkinson's disease," *Brain and Language*, vol. 34, no. 2, pp. 315 – 327, 1988. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0093934X88901423

[8] J. Rusz, R. Cmejla, H. Ruzickova, and E. Ruzicka, "Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated parkinsons disease," *Journal of the Acoustical Society of America*, vol. 129, pp. 350–367, 2011.

[9] R. J. Holmes, J. M. Oates, D. J. Phyland, and A. J. Hughes, "Voice characteristics in the progression of parkinson's disease," *International Journal of Language and Communication Disorders*, vol. 35, no. 3, pp. 407–418, 2000.

[10] A. M. Goberman and C. Coelho, "Acoustic analysis of parkinsonian speech i: Speech characteristics and l-dopa therapy," *Neurorehabilitation*, vol. 17, pp. 237–246, 2002.

[11] A. Maier, F. Hönig, V. Zeissler, A. Batliner, E. Körner, N. Yamanaka, P. D. Ackermann, and E. Nöth, "A language-independent feature set for the automatic evaluation of prosody," in *Proc. Interspeech 2009*, Brighton, England, 2009, pp. 600–603.

[12] K. N. Stevens, *Acoustic Phonetics*. Cambridge, MA 02141: The MIT Press, 1998.

[13] G. Fant, *Acoustic Theory of Speech Production*. Netherlands: Mouton, 1960.

[14] P. Beyerlein, A. Cassidy, V. Kholhatkar, E. Lasarcyk, E. Nöth, B. Potard, S. Shum, Y. C. Song, W. Spiegl, G. Stemmer, and P. Xu, "Vocal aging explained by vocal tract modelling: 2008 JHU summer workshop final report," Tech. Rep., 2008.

[15] J. A. Nelder and R. Mead, "A Simplex Method for Function Minimization," *The Computer Journal*, vol. 7, no. 4, pp. 308–313, 1965.

[16] S. Kirkpatrick, C. Gelatt, and M. Vecchi, "Optimization by Simulated Annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[17] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, 2009.

[18] M. A. Hall, "Correlation-based feature subset selection for machine learning," Ph.D. dissertation, University of Waikato, Hamilton, New Zealand, 1998.