Available online at www.sciencedirect.com
ScienceDirect
journal homepage: www.elsevier.com/locate/bbe

Original Research Article

Classification of speech intelligibility in Parkinson's disease

Taha Khan [a,b,*], Jerker Westin [b], Mark Dougherty [b]

[a] School of Innovation, Design and Technology, Malardalen University, Vasteras, Sweden
[b] School of Technology and Business Studies, Computer Engineering, Dalarna University, Falun, Sweden

ARTICLE INFO

Article history:
Received 7 August 2013
Received in revised form
20 September 2013
Accepted 4 October 2013

Keywords:
Parkinson's disease
Speech processing
Dysarthria
Support vector machine
Tele-monitoring

ABSTRACT

A problem in the clinical assessment of running speech in Parkinson's disease (PD) is to track underlying deficits in a number of speech components including respiration, phonation, articulation and prosody, each of which disturbs the speech intelligibility. A set of 13 features, including the cepstral separation difference and Mel-frequency cepstral coefficients were computed to represent deficits in each individual speech component. These features were then used in training a support vector machine (SVM) using *n*-fold cross validation. The dataset used for method development and evaluation consisted of 240 running speech samples recorded from 60 PD patients and 20 healthy controls. These speech samples were clinically rated using the Unified Parkinson's Disease Rating Scale Motor Examination of Speech (UPDRS-S). The classification accuracy of SVM was 85% in 3 levels of UPDRS-S scale and 92% in 2 levels with the average area under the ROC (receiver operating characteristic) curves of around 91%. The strong classification ability of selected features and the SVM model supports suitability of this scheme to monitor speech symptoms in PD.

1. Introduction

Parkinson's disease (PD) is caused by the progressive deterioration of dopamine producing nerve cells in the mid-brain [1]. The dopamine serves as a messenger that allows communication between the mid-brain and other parts of the brain that are responsible for producing smooth and controlled body movements. A lack of dopamine causes a number of motor symptoms including reduced muscular movement, tremor and speech dysfunctions. These symptoms advance with the disease progression and degrade the quality of life of people with PD. Medication and surgical intervention can alleviate some of these symptoms but there is no cure available. PD treatments are optimized by following up the patients at regular intervals; this is problematic given the physical restrictions of patients and the established assessment procedures. Tele-monitoring of symptoms through internet or mobile devices have potential to complement traditional clinical practices and may relieve the workload of clinicians as well as reduce treatment cost [2]. In this aspect objective assessment algorithms are developed that record biometric signals associated with PD symptom severity and quantified on standard clinical scale such as the UPDRS (Unified

* *Corresponding author at*: School of Technology and Business Studies, Computer Engineering, Dalarna University, 79188 Falun, Sweden.
E-mail addresses: tkh@du.se (T. Khan), jwe@du.se (J. Westin), mdo@du.se (M. Dougherty).

0208-5216/$ – see front matter
http://dx.doi.org/10.1016/j.bbe.2013.10.003

Parkinson's Disease Rating Scale) [3]. Speech is particularly suitable in this regard as it is convenient to self-record without supervision and expensive equipment.

Speech disturbance is an early indicator of PD and previous investigations revealed that speech degradation and general PD symptom severity are strongly interlinked [4]. Several methods on PD speech classification are reported to have analyzed speech signals to discriminate between PD patients and healthy controls [5–8]. Traditional investigations involved voice signal analysis to estimate dysphonic symptoms manifested in sustained-vowel phonation. In the recent methods [6–8], the running speech is analyzed to demonstrate deficits in motor speech, suggesting that PD can affect all different subsystems of speech including respiration, phonation, articulation and prosody.

The speech item utility in motor UPDRS was previously examined by Zraick et al. [9]. According to them, a standard speech protocol to identify symptom severity should include reading of an unfamiliar passage containing different linguistic structures and a description to assess the reading ability. A strong inter-rater reliability coefficient was produced between symptom severity ratings, performed separately by the medical (neurologists) and non-medical (speech pathologists) experts, when a standard speech protocol was utilized in the motor examination. It was inferred that the running speech with standard formulation has potential to exploit capacious symptoms in PD speech, providing a broader perspective of evaluation.

The structural analysis of running speech is complex due to linguistic confounds and annotations at different levels of processing e.g. separating syllables, phonemes and prosodic units. Instead of processing individual speech units for symptom analysis, acoustic features such as variation in fundamental frequency, sound pressure level, speech rate, pause intervals and signal-to-noise ratio have been relied upon to identify PD speech impairment [6–8,10]. In a recent method for evaluating spastic dysarthria, the Mel-frequency cepstral coefficients (MFCC), glottal-to-noise energy and harmonic-to-noise ratio were evaluated in running speech samples [11] and indicated high correlation. Llorente et al. [12] proposed a scheme for detection of voice impairment from text-dependent running speech. They parameterized MFCCs from 140 recorded running speech samples. These MFCCs were then used for classification between117 dysarthric and 23 normal speech samples with an accuracy of 96%.

For an accurate monitoring of speech symptom status in PD, statistical mapping between the computed features and clinical ratings of speech symptom severity is an important step. A difficulty in the clinical assessment of running speech is to track underlying deficits in individual speech components which as a whole disturb the speech intelligibility. The aim of this work is to extract signal features from running-speech samples computing deficits in individual speech components, and to utilize these features for classification between speech symptom severity levels in accordance with the UPDRS-S using support vector machines (SVM) [13]. A recently introduced speech measure, cepstral separation difference (CSD) [14] has been explored in pursuit to categorize the level of speech impairment.

2. Patients and data

The data were obtained from a feasibility study of an at-home testing device [2] conducted at the University of California, San Francisco (UCSF) in collaboration with Parkinson's Institute. A total of 80 subjects (48 males and 32 females) with an average age of 63.8 years, participated in this study over a course of a year (i.e. from June 2009 to June 2010). 60 participants (40 males and 20 females) had a mean PD duration of 75.4 weeks and 20 other participants were normal controls. Speech samples were recorded during examinations of speech by a clinician. The recording equipment consisted of a microphone connected to a computer-based test-battery called QMAT. Subjects were asked to recite static paragraphs displayed on the QMAT screen in 3 standard running speech tests (RST). The paragraphs [15], ''The North Wind and the Sun'', ''The Rainbow Passage'' and ''The Grandfather Passage'' were recited by the subjects in RST type 1, 2 and 3 respectively. These paragraphs were devised in a way such that the level of textual difficulty increases from RST 1 to 3, demanding a greater stress in reading [14,16].

Each subject was rated by a clinician based on his/her reading performance in each RST using the UPDRS examination of speech. The speech examination is item 18 in UPDRS part III and is abbreviated as UPDRS-S [3]. The UPDRS-S is ranged from 0 to 4 where '0' represents normal speech, '1' represents mildly impaired speech, '2' represents moderately impaired speech, '3' represents severely impaired speech and '4' represents unintelligible speech. Out of the 80 subjects, 24 subjects were rated '0', 25 subjects were rated '1', 28 subjects were rated '2' and 3 subjects were rated '3'. The speech signals were sampled at 48 kHz with 16 bit resolution. In total, 240 speech samples (80 subjects $\times$ 3 RST types) have been utilized for classification between the symptom severities.

3. Methods

The intelligibility of speech can be disturbed by a number of PD symptoms. Pinto et al. [4] identified the relation between PD symptoms and anatomical substrates of speech components. According to them, vocal impairment in PD is associated with pathological changes to mainly three components of speech: respiration, phonation and articulation, attributed to the dysfunction of musculatures at subglottis (lungs, trachea, windpipes etc.), glottis (larynx) and supraglottis (jaw, lips, tongue, velum, pharynx etc.) respectively. The collective dysfunction in these components gives rise to the dysfunction in the fourth speech component called prosody.

In this work, several acoustic features were extracted from running speech signals to estimate symptoms in each speech component. For the sake of description, these features were organized into groups as: (1) measures relating to the phonatory symptoms, (2) measures relating to the articulatory symptoms, and (3) measures relating to the prosodic symptoms. The respiratory symptoms (e.g. reduced loudness) are manifested in speech prosody. The phonatory measures represent symptoms which emerge due to the in-coordination between phonation and respiration and cause harshness and

Please cite this article in press as: Khan T, et al. Classification of speech intelligibility in Parkinson's disease. Biocybern Biomed Eng (2013), http://dx.doi.org/10.1016/j.bbe.2013.10.003

hoarseness in speech [4]. The articulatory measures represent symptoms that emerge by subtle changes in the motion of articulators and cause imprecise articulation and short rushes of speech. The prosodic measures represent symptoms in rhythm, stress, loudness and intonation in speech.

3.1. *Measures of phonatory symptoms*

Lungs are the primary source of speech production [17]. Voice is produced when an airflow generated by the lungs passes through the glottis, modulated by the vocal-fold vibration and filtered by the vocal-tract resonances. According to the source-filter speech model [17], the vibration of vocal folds generates a source excitation signal holding the properties of pressure wave expelled from the lungs. This source signal is filtered by the spectral envelope of vocal tract resonances to form a speech signal. In order to estimate the phonatory symptoms, disturbance in the pressure wave can be estimated using pressure magnitude of source and filter log-spectrums derived from the speech signal [18]. Note that the source excitation signal in running speech is constituted by a series of pitch pulses.

Harshness in phonation is the PD symptom related to the laryngeal hyper-function [19]. A harsh voice quality is produced when an involuntary glottal constriction restricts the airflow and generates turbulence at the constriction point. The noise in speech spectrum generated as a result of this turbulence is termed aspiration noise [19]. Further, the irregular vibration of vocal folds disturbs the modulation of source excitation signal that affects frequency distribution of harmonics throughout the spectrum [18]. The auditory impression in this case is of a breathy voice quality due to an audible escape of air on phonation.

An uncontrolled glottal closure pattern is the most frequently manifested symptom in PD speech [20]. At an event of an unintended glottal constriction on phonation, the subglottal pressure increases at the glottis and is not delivered to the supraglottis (vocal tract). The increased subglottal pressure raises the energy level of aspiration noise in the source excitation. Moreover, the deficient pressure waves propagate to the supraglottis and weaken the energy level of resonances. This phenomenon can be observed if source and filter log-spectrums derived from the speech signal of a severely impaired PD patient are compared to that of a normal control (Fig. 1).

The cepstral separation difference (CSD) [14] was utilized to estimate the pressure wave disturbance caused by the uncontrolled glottal closures in speech. CSD computes the log-magnitude ratio between source and filter log-spectrums to estimate the energy difference caused by the raised aspiration noise in the source. The source and filter log-spectrums were computed by performing the cepstral analysis of speech signal. In the first step, the real cepstrum c[*n*], of cepstral coefficients *n*, was computed by applying an inverse discrete Fourier transform on the real log of the discrete Fourier transform of speech signal S[*i*]. In the cepstral domain, the source cepstrum $c_e[n]$ and the filter cepstrum $c_h[n]$ were separated by liftering the cepstrum c[*n*] by applying a high-time and a low-time lifter on c[*n*] respectively. A cutoff value of 20 cepstral coefficients was used for liftering which is generally used in identifying the speech source in speech recognition systems [21]. In the third step, the source and filter log-spectrums were computed by applying the real discrete Fourier transform on source cepstrum $c_e[n]$ and filter cepstrum $c_h[n]$ respectively.

The residual log-spectrum $R[\omega]$ was computed using Eq. (1), where $E[\omega]$ and $H[\omega]$ represent the log-magnitude frequency spectrums of source and filter respectively, and ω represents the log-power coefficient. $R[\omega]$ was computed between the frequency bands 0–1000 Hz incorporating the range of human voice fundamental frequencies:

$$R[\omega] = \log|E[\omega]| - \log|H[\omega]| \tag{1}$$

Our experiments on PD running speech samples have shown that the elevated aspiration energy in source log-spectrum in conjunction with energy depression in filter log-spectrum results in higher residual values in log-spectrum *R*, compared to that of speech samples from healthy controls. Moreover, an increasing irregularity in the modulation of log-spectrum *R* relative to the increasing symptom severity was

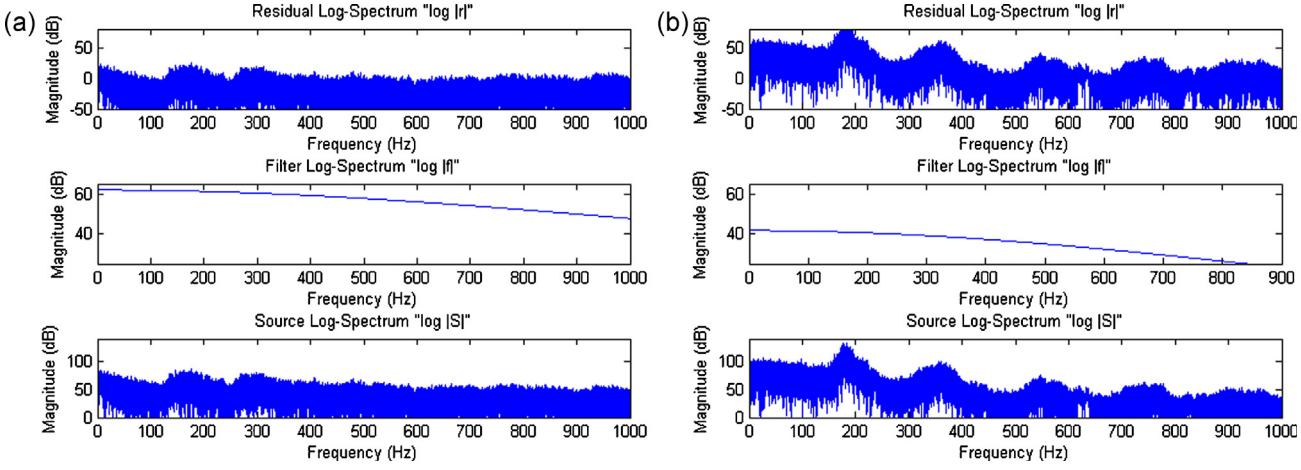


Fig. 1 – Cepstral separation difference 'R' in running speech test-3 is shown for two subjects: (a) normal speech sample (rated '0'), dCSD = 21.7 dB; and (b) severely impaired speech sample (rated '3'), dCSD = 108.4 dB. The increase in source magnitude (integrating aspiration) and reduction in filter magnitude accompanied by irregular modulation shifts in 'R' can be noticed in the sample rated '3'. The δ_{CSD} is markedly increased in the sample rated '3'.

Please cite this article in press as: Khan T, et al. Classification of speech intelligibility in Parkinson's disease. Biocybern Biomed Eng (2013), http://dx.doi.org/10.1016/j.bbe.2013.10.003

observed. The mean-absolute deviation (represented as δ_{CSD}) was used to compute dispersion in the modulation of $R[\omega]$ (Eq. (2)), where $\bar{R}$ is the overall mean of $R[\omega]$. δ_{CSD} values increase relative to the increasing symptom severity in speech [14]:

$$\delta_{CSD} = \frac{1}{1000}\sum_{\omega=1}^{1000}|R[\omega] - \bar{R}| \quad (2)$$

Hoarseness in speech is another symptom related to impaired function of the larynx. Hoarseness is produced by an interference with optimum vocal fold adduction characterized by a breathy escape of air on phonation [19]. The vocal fold adduction increases the subglottal pressure at the glottis, resulting in increased aspiration level, followed by a meager propagation of pressure waves in the vocal tract. This phenomenon results in speech depression which can be measured by the CSD by comparing the energy levels between source and filter log-spectrums. In order to investigate the depression in speech frequency through CSD, a peak-detector was applied on R to locate peaks that represent the level of residual energy at each frequency. The average peaks' magnitude (AP_{CSD}) was found to be elevated in PD speech samples and was rising with increasing symptom severity. The δ_{CSD} along with AP_{CSD} were selected as the representative measures of phonatory symptoms for classification of speech symptom severity.

3.2. Measures of articulatory symptoms

The PD symptoms in articulation involve short rushes of speech and articulation blurring (e.g. imprecise consonant articulation) that arise as a consequence of hypokinetic movement of articulators (tongue, velum, pharynx, lips etc.) [22]. The Mel-frequency cepstral coefficients (MFCC) are considered as effective measures to identify articulatory symptoms [5–7,11,12]. MFCC are aimed at detecting subtle changes in the motion of articulators that interfere with speech intelligibility [23]. For instance, the placement of the tongue has a key role in creating resonances (formants) in mouth, and slight misplacement of the tongue can alter the energy between the frequency bands. MFCC compute energy differences between frequency bands of speech signal, which can be used to discriminate varying energy levels of impaired resonances.

The MFCC are computed by partitioning the speech frequency into overlapping Mel-frequency filter bands followed by the application of cepstral and cosine transformations on each band [23]. The Mel-frequency filter bands are triangular in shape and compute the energy spectrum around the center frequency in each individual band of speech frequency. The boundary frequencies of filter bands are uniformly spaced using the Mel-scale given in Eq. (3):

$$m = 1127\ln\left(1 + \frac{f}{700}\right), \quad 0 \le f \le F_s \quad (3)$$

where, f and F_s are speech frequency and sampling rate in hertz respectively. In the next step, the log-energy at the output of each filter is computed. The MFCC is the discrete cosine transform of the filter energy outputs, given in Eq. (4):

$$MFCC_n = \sum_{k=1}^{K} E_k \cos\left[\frac{n(k-0.5)\pi}{K}\right], \quad n = 0, \ldots, L \quad (4)$$

where L is the number of MFCC coefficients. Typically, a value of L between 10 and 16 is used. n is the order of MFC coefficient. The 0th MFC coefficient represents the original signal energy and is ignored. E_k is the log energy of the kth filter. K is the number of filter bands and is chosen between 20 and 40.

In order to compute MFCC, the speech signal was divided into frames of 50 ms each. A Mel-frequency filter bank of $K = 24$ was applied to extract up to 10th order MFCC from each frame. This choice of filters results in a higher spectral resolution at lower frequency bands where the most significant information regarding articulatory impairment is contained. The mean of MFCC between each frame are chosen as the representative measures of articulatory symptoms for classification of symptom severity.

3.3. Measures of prosodic symptoms

The prosodic symptoms in PD are categorized by reduced vocal stress, monopitch intonation, monoloudness and abnormality in speech rate [24]. Loudness and speech rate can be estimated using the short-term dynamics of speech signal. Pitch can be estimated by computing the fundamental frequency (F0).

3.3.1. Short-term spectral dynamics

The prosodic changes in speech are reflected in the dynamics of short-term spectral components of speech signal. Number of pauses, voice intensity levels, speech/pause intervals and articulation rate can be estimated from spectral envelopes of a speech signal. Rosen et al. [25] found 'Pause Time' and 'Spectral Range' as the most specific (95%) and accurate (95%) differentiators of speech prosody. Inspired by this, we developed a pause detection algorithm that locates pause occurrences and pause intervals in a speech signal. Acoustic features such as zero-crossing rate, short-term energy and spectral centroid were derived simultaneously during this process.

In this approach, the number of pauses (N_p) in recorded speech is computed by segmenting between the voicing and unvoicing regions in speech spectrum. Voice segmentation is based on thresholding the short-term energy (STE) and spectral centroid (SC) in each signal frame. First, the speech time-series $x(n)$ of length N is broken into i short frames of size 50 ms. Then, the STE and SC are computed in each ith frame using the formulae given in Eqs. (5) and (6) respectively, where $X_i(k)$, for $k = 1, \ldots, N$, represent the discrete Fourier coefficients of the ith frame:

$$STE_i = \frac{\sum_{n=1}^{N}|x_i(n)|^2}{N} \quad (5)$$

$$SC_i = \frac{\sum_{k=1}^{N} kX_i(k)}{\sum_{k=1}^{N} X_i(k)} \quad (6)$$

Generally, the STE and SC sequences depict higher magnitude in the voiced frames and are relatively weaker in the unvoiced frames [26]. To qualify a segment as a pause, two thresholds i.e. T_1 for SC and T_2 for STE respectively are

Please cite this article in press as: Khan T, et al. Classification of speech intelligibility in Parkinson's disease. Biocybern Biomed Eng (2013), http://dx.doi.org/10.1016/j.bbe.2013.10.003

computed. In order to compute T_1 and T_2, two histograms are produced for each SC and STE sequences respectively. The positions of first and second local maxima (represented as M_1 and M_2 respectively) are computed in each histogram. The thresholds are calculated using Eq. (7), where W is a user-defined constant:

$$T = \frac{WM_1 + M_2}{W + 1} \tag{7}$$

Larger values of W lead to a strict threshold value closer to M_1 which may result in voice information loss in case of impaired speech signals that may possibly have weak STE and SC. The value of W was relaxed and was set as 0.1. The formula generates thresholds T_1 and T_2 for SC and STE histograms respectively. For the ith frame, if the SC_i and STE_i values are larger than the thresholds T_1 and T_2 respectively, then the ith frame is declared as the voiced frame. Whereas, if the SC_i and STE_i values are less than the thresholds T_1 and T_2 respectively, then the ith frame is declared as a pause.

Our experiments showed that the N_p and the pause intervals elevate in the higher symptom severity levels. The STE sequence was ignored because it is affected by the amplitude and sound pressure changes caused by the varying distance between mouth and mic. The SC sequence was selected as a measure to detect depression in voice intensity as it is invariant to temporal changes in the speech signal. In our experiments, the mean magnitude of SC_i (for $i = 1...N$) was found higher in normal speech signals and found relatively lower in impaired speech.

Further, the zero-crossing rate (ZCR) in voiced regions of speech signal was investigated. Prosodic changes can be measured using ZCR because the ZCR is lower in voiced regions in presence of fundamental frequency which is low-frequency in nature [18]. The ZCR sequences were computed for each ith voiced region of speech signal $x(n)$ using Eq. (8):

$$\mathrm{ZCR(i)} = \frac{f_s}{2n}\left(\sum_{l=1}^{k} |sign(x_i(l)) - sign(x_i(l-1))|\right) \tag{8}$$

where, f_s is the sampling rate, n is the length of speech signal $x(n)$, x_i is the voiced region in speech signal $x(n)$, k is the length of the voiced region x_i and ZCR(i) is the zero-crossing rate in the voiced region x_i. The energy (variance) of ZCR sequences (represented as ε_{ZCR}) given in Eq. (9), was used for symptom level classification:

$$\varepsilon_{ZCR} = \frac{\sum_N^{i=1}(ZCR_i - \overline{ZCR})^2}{N} \tag{9}$$

where, N is the total number of voiced regions and $\overline{ZCR}$ is the average between ZCR values in these regions. A negative correlation between ε_{ZCR} and symptom scores would indicate increasing monopitch intonation in speech.

3.3.2. F0 variation

Reduced F0 variability is a noticeable feature of prosody that leads to the audible impression of monopitch intonation in PD speakers [6–8]. A problem in the F0 estimation of PD speech is to find periodic patterns (pitch periods) in a speech signal. A normal voice region exhibits a periodic pattern in the speech signal. By contrast, an impaired voice region exhibits a noise-like non-periodic pattern. In order to find correct pitch periods in an impaired speech signal, an algorithm must be able to discriminate between the voiced regions of impaired speech and the unvoiced regions holding the additive noise.

To cope with this problem, a modified cepstrum-based pitch detector [27] was utilized to evaluate F0 variation in PD speech. This method utilizes ZCR and STE to remove unvoiced regions in speech signal based on the assumption that the unvoiced regions exhibit low STE and high ZCR. This type of filtering preserves the impaired voiced regions having higher STE and generally low ZCR and removes the unvoiced regions holding the additive noise. Once a signal is translated into the cepstral domain, the cepstral peaks representing the 'false' pitch periods are readily eliminated by the prior removal of unvoiced frames. The remaining cepstral peaks provide the locations for the correct pitch periods. A pitch period T0 is estimated subsequently by converting back the cepstral peak position into the time domain. F0 is estimated using $F0 = F_s/T0$, where F_s is the sampling frequency (i.e. 48 kHz).

Apart from the reduced F0 variation in impaired speech signals, our experiments have shown that the pitch-period locations were spread in small chunks across the speech spectrogram. Distortion in pitch periods and interval between the periods increased relatively to increasing symptom severity. In addition to F0 standard deviation ($F0_{std}$), the entropy between T0 intervals (I_{ent}; Eq. (10)) and jitter in T0 (J_{PPQ}: pitch perturbation quotient [28]; Eq. (11)) were computed and were used in symptom level classification:

$$I_{ent} = -\sum_{i=1}^{N-1} P_i \ln P_i \tag{10}$$

$$J_{PPQ} = \frac{(1/N-2)\sum_{i=2}^{N-1} |(T0^{i-1} + T0^{i} + T0^{i+1})/3 - T0^{i}|}{(1/N)\sum_{i=1}^{N} T0^{i}} \tag{11}$$

where, P_i is the probability that the interval lengths of two adjacent pitch periods (T0) in speech spectrogram are equal. N is the total number of pitch periods in speech spectrogram.

4. Feature analysis

A correlation analysis between the speech features and the clinical ratings of speech is complicated. First, due to the fact that the clinical ratings are based on a set of multiple symptoms in which each symptom is represented by a different feature. Secondly, the UPDRS-S ratings follow a monotonic rank order from '0' to '3' (normal-to-severe level). Under these conditions, the one-to-one mapping between a calculated feature (representing an individual symptom) and its corresponding subjective rating (based on multiple symptoms) is not possible through one-dimensional scales (e.g. rank-order or Likert scales).

Louis Guttman proposed the 'Guttman Scale' [29] to perform systematic correlation between qualitative rank-order variables. The model suits the qualitative ranked nature of speech dataset where a human rater examines proportions of different speech symptoms to choose between the severity levels. In the Guttman scale, a variable y (i.e. a human rater) with h distinct ordered values (i.e. UPDRS-S classes) is said to

Please cite this article in press as: Khan T, et al. Classification of speech intelligibility in Parkinson's disease. Biocybern Biomed Eng (2013), http://dx.doi.org/10.1016/j.bbe.2013.10.003

be a simple function of variable x (i.e. a speech feature) with i distinct ordered values, if for each value of x there is only one value of y. The converse needs not to hold and for the same value of y, there may be one or multiple values of x. The Guttman monotonicity coefficient (μ_2) expresses an increase in variable x relative to an increase in variable y without assuming that the intervals between each value of y are perfectly scaled. Accordingly, μ_2 relies on that ties between x and y can be untied in the same order without penalty which suits the situation where subjective rating is based on a set of multiple symptoms in which proportion of one symptom may vary from another. A μ_2 equal to '+1' depicts perfect correlation between x and y. The formula to compute the Guttmann monotonicity coefficient between x and y is given in Eq. (12):

$$\mu_2 = \frac{\sum_{h=1}^{n}\sum_{i=1}^{n}(x_h - x_i)(y_h - y_i)}{\sum_{h=1}^{n}\sum_{i=1}^{n}|x_h - x_i||y_h - y_i|} \quad (12)$$

where h is the order of UPDRS-S levels '0' to '3' and i is the corresponding order of feature values relative to h.

We performed two correlation tests. In the first test, μ_2 was utilized to correlate between the features computed from speech samples in RST 1, 2 and 3 (each consisting of 80 samples) and the UPDRS-S ratings. The features were evaluated for each RST separately so that changes in correlation can be observed with respect to the increasing textual difficulty in RST 1, 2 and 3 respectively. In the second test, the features computed from the total speech samples (consisting of 240 samples) were correlated with the UPDRS-S ratings. As there were few samples in symptom group '3', they were merged into symptom group '2'. The dataset was stratified through Jackknifing [30] to estimate the precision of μ_2 by leaving out one or more samples at a time from the dataset. The Jackknife estimates of μ_2 are listed in Table 1.

The measures of phonatory symptoms (CSD features) were strongly correlated with the clinical ratings and this correlation improved with the increasing textual difficulty. The δ_{CSD} produced higher correlation than any other acoustic feature in this study i.e. 0.70, 0.74 and 0.78 in RST 1, 2 and 3 respectively, and 0.73 in the total dataset. These correlations were statistically significant ($p < 0.05$). The δ_{CSD} estimates dispersion in the glottal energy difference that arise by the uncontrolled closures of glottis which is according to one study [20] the most frequently manifested symptom in PD speech. Also, the CSD measure of speech depression (AP_{CSD}) indicated strong correlation ($\mu_2 > 0.5$) with subjective ratings.

The measures of articulatory symptoms (MFCC) showed strong statistically significant ($p < 0.05$) correlation with the clinical ratings. However, only the 4th MFCC showed improving correlation with the increasing textual difficulty i.e. 0.58, 0.65 and 0.68 in RST 1, 2 and 3 respectively. The MFCCs in the order 7th-to-10th indicated strong negative correlation with the subjective ratings. The negative correlation is due to the cosine sign change in the MFCC order 7th-to-10th amid discrete cosine transformation.

From the measures of prosodic symptoms, the number of pauses N_p showed high statistically significant ($p < 0.05$) correlation with the clinical ratings e.g. 0.77 in RST-3 and 0.64 in the total dataset. The correlation between pause intervals P_I and subjective ratings were also strong in RST-3. The high statistically significant correlation of N_p and P_I, specifically in RST-3, suggests that the respiratory symptoms are discernible if the textual difficulty in reading is strong enough to stress the patients forcing them to rest at several occasions during the recitation. Among other prosodic measures, the ε_{ZCR} was moderately correlated. The SC_{AVG} and the F0 features were weakly correlated with the ratings.

Further analysis revealed that the 7th and 8th MFCCs and the 9th and 10th MFCCs were strongly correlated between each other. Moreover, the 1st, 2nd, 5th and 6th MFCC were not correlated with the subjective ratings in all the speech tests. For these reasons the 8th and 10th MFCC as well as the 1st, 2nd, 5th and 6th MFCC were excluded from the final list of features used in symptom level classification. Nevertheless, a majority of suspected symptoms in PD speech were covered by a total of 13 distinct acoustic features associated with each component of speech. Importantly, none of these features showed very weak correlation ($\mu_2 < 0.2$) with the clinical ratings that justifies the higher classification accuracy obtained in this study.

5. Classification

The support vector machine (SVM) is widely relied on in biomedical decision support systems [31] for its ability to regularize global optimality in the training algorithm and for having excellent data-dependent generalization bounds to model non-linear relationships. However, the classification success of SVM depends on the properties of the given dataset and accordingly the choice of an appropriate kernel function. Training a linear SVM is equivalent to finding a hyper plane with maximum separation. In case of a high-dimensional feature space with low input data size, instances may scatter in groups and classification with a linear SVM may lead to imperfect separation between the hyper planes. The solution is then to utilize a nonlinear SVM that maps these features into a 'higher-dimensional' space by incorporating slack variables. This leads to a very large quadratic programming (QP) optimization problem but it can be solved using the sequential minimal optimization (SMO) algorithm [13]. SMO decomposes the overall QP problem into QP sub-problems. This decomposition is performed by solving the smallest possible QP optimization problem at every step involving two Lagrange multipliers satisfying the linear equality constraint to find local optima. At each decomposition step, SMO finds the optimal values for these multipliers and updates the SVM cost function to reflect new optimal marginal separations between the hyper planes.

Another important consideration in SVM is the choice of kernel function for transforming the non-linear feature space into a straight linear classification solution. Kernel functions can be linear, polynomial or radial basis and the choice of function is based on the nature of feature space. In our case, the underlying specificity regarding the qualitative nature of data could not be determined. To circumvent this limitation, a universal kernel function based on the Pearson VII function (PUK) [32] is utilized. PUK is generally used for curve fitting purposes and has the general form given in Eq. (13). Here H is the peak height at the center x_0 of the peak and x is an independent variable. The variables σ and ω control the

Please cite this article in press as: Khan T, et al. Classification of speech intelligibility in Parkinson's disease. Biocybern Biomed Eng (2013), http://dx.doi.org/10.1016/j.bbe.2013.10.003

Table 1 – Jackknife estimates of Guttman Correlation Coefficient (μ_2) between the symptom measures and UPDRS speech ratings. Estimates in bold represent very high (>0.7) statistically significant ($p < 0.05$) correlation. Estimates in *italic* represent improving correlation relative to the increasing textual difficulty in RST 1, 2 and 3 respectively.

Measurement type	Feature	Symbol	μ_2			
			RST-1[a]	RST-2[b]	RST-3[c]	In total samples
		Measures of phonation				
Cepstral separation difference	(1) Mean absolute deviation in R	δ_{CSD}	**0.70**	**0.74**	**0.78**	**0.73**
	(2) Average between peaks' magnitude in R	AP_{CSD}	0.57	0.56	0.62	0.57
		Measures of articulation				
Mel-frequency cepstral coefficients	(3) 1st Mel-frequency cepstral coefficient	MFCC1	0.08	0.12	−0.01	0.06
	(4) 2nd Mel-frequency cepstral coefficient	MFCC2	0.16	0.15	0.15	0.13
	(5) 3rd Mel-frequency cepstral coefficient	MFCC3	0.38	0.55	0.52	0.48
	(6) 4th Mel-frequency cepstral coefficient	MFCC4	0.58	0.65	0.68	0.63
	(7) 5th Mel-frequency cepstral coefficient	MFCC5	−0.01	0.08	0.14	0.13
	(8) 6th Mel-frequency cepstral coefficient	MFCC6	−0.02	−0.12	−0.12	−0.17
	(9) 7th Mel-frequency cepstral coefficient	MFCC7	−0.47	−0.64	−0.58	−0.56
	(10) 8th Mel-frequency cepstral coefficient	MFCC8	−0.58	−0.65	−0.60	−0.60
	(11) 9th Mel-frequency cepstral coefficient	MFCC9	−0.49	−0.51	−0.52	−0.50
	(12) 10th Mel-frequency cepstral coefficient	MFCC10	−0.52	−0.54	−0.59	−0.55
		Measures of prosody				
Spectral dynamics	(13) Number of pauses	N_p	0.67	0.62	**0.77**	0.64
	(14) Pause intervals	P_I	0.46	0.22	**0.71**	0.45
	(15) Energy (variance) in ZCR sequence	ε_{ZCR}	−0.46	−0.48	−0.58	−0.50
	(16) Mean between SC_i sequence	SC_{AVG}	−0.27	−0.33	−0.32	−0.29
F0 variability	(17) Fundamental frequency F0 standard deviation	$F0_{std}$	−0.41	−0.42	−0.51	−0.26
	(18) Interval entropy between T0	I_{ent}	0.23	0.28	0.37	0.28
	(19) Jitter (pitch perturbation quotient)	J_{PPQ}	0.54	0.21	0.59	0.38

[a] *Running Speech Test-1*: ''The north wind and the sun were disputing which one is the stronger when a traveler came along wrapped in a warm cloak. They agreed that the one who first succeeded in making the traveler's take his cloak off should be considered the stronger. Then the north wind blew as hard as it could but the more he blew the more closely the traveler pulled his cloak around him and at last the north wind gave up the attempt. Then the sun shined out and immediately the traveler took off his cloak and so the north wind was agreed that the sun was the stronger of the two.''

[b] *Running Speech Test-2*: ''When the sunlight strikes rain drops in the air, they act like a prism and form a rainbow. The rainbow is a division of white light into many beautiful colors. These take the shape of a long round arch with its path high above and its two ends apparently beyond the horizon. There is according to a legend a boiling part of gold at one end. People look but no one ever finds it. When a man looks for something beyond his reach, his friends say he is looking for the part of gold at the end of the rainbow.''

[c] *Running Speech Test-3*: ''Do you wish to know all about my grandfather; well he is nearly 93 years old. He dresses himself in an ancient black frock coat usually minus several buttons. Yeah he still thinks he is swiftly as ever. A long flowing beard clings to his chin giving those who observe him a pronounced feeling of an outmost respect. When he speaks, his voice is just a bit cracked and covers the trifle. Twice each day, he plays skillfully and with a zest upon a small organ except in the winter when the ooze or snow or ice prevents he slowly takes a short walk in the open air each day. We have often urged him to walk more and smoke less but he always answers 'banana oil'. Grandfather likes to be modern in his language.''

half-width and the tailing factor of peak respectively. Importantly a curve with ω equals to 3 and σ equals to 1, is comparable to a sigmoid function used in the neural network modeling [32]:

$$f(x) = \frac{H}{[1 + ((2(x - x_0)\sqrt{2^{1/\omega} - 1})/\sigma)^2]^{\omega}} \quad (13)$$

The SMO configured with PUK kernel function (of $\sigma = 1$ and $\omega = 3$) was used for classification of speech data. Originally, there were 72 samples in class '0' (24 subjects × 3 RST), 75 samples in class '1' (25 subjects × 3 RST), 84 samples in class '2' (28 subjects × 3 RST) and only 9 samples in class '3' (3 subjects × 3 RST). In order to avoid high standard error in class '3' pertaining to a low sample size as well as to balance the class distribution, samples in class '3' were merged into class '2' leaving behind 3 levels for symptom classification where class '0' consisted of 72 samples, '1' consisted of 75 samples and '2' consisted of 93 samples.

There were two important investigations performed: First, to analyze the influence of reading stress on symptom level classification with a hypothesis that classification rate should improve relative to the increasing textual difficulty demanding a greater stress in reading. Three different classification tests were performed each on the samples of a different RST as the level of textual difficulty increases from RST 1 to 3. The second investigation was to estimate the classification performance of SVM on classifying the complete speech dataset. A 10-fold cross validation strategy [33] was adopted to obtain unbiased generalization estimates. For optimal results, the SVM regularization constant was tuned between 1 and 10. A regularization constant of 7 produced the best generalization performance in this dataset and was maintained in all classification tests.

Please cite this article in press as: Khan T, et al. Classification of speech intelligibility in Parkinson's disease. Biocybern Biomed Eng (2013), http://dx.doi.org/10.1016/j.bbe.2013.10.003

5.1. *Investigation-1*

In the first investigation, three classification matrices of dimensions 13 (features) × 80 (samples) for RST 1, 2 and 3 respectively were prepared for classification between the 3-level UPDRS-S targets. The confusion matrices (Table 2.1a, 2.2a and 2.3a) were used to portray the prediction performance in each classification test. Each row in the confusion matrix represents the actual class instances while each column represents instances in the predicted class. The matrix diagonal presents the correctly predicted samples or the true positives in each class. In the first classification test taking RST-1 into account, the SVM classified the samples into 3 symptom levels with a classification accuracy of 63% and true positive rates (TPR) of 63%, 40% and 81% in class 0, 1 and 2 respectively.

The low TPR of class 0 and 1 indicates the difficulty of discriminating between the normal and mildly impaired speech samples. One reason could be that the textual difficulty in RST-1 paragraph was not strong enough to stress the mildly impaired subjects, who comfortably read the passage without displaying any symptom and were thus classified as being normal. Additionally, the high TPR of class 2 indicates that the more severely impaired subjects exhibited reading difficulty even in this low-stress setting and revealed symptoms which were effectively quantified.

Table 2 – *Investigation 1*: Textual difficulty vs. classification rate: Classification rate increases proportionally to the increasing level of reading stress. (a) Confusion matrix. (b) Receiver operating characteristic curves.

1. Classification Test on RST-1
(Stressful)

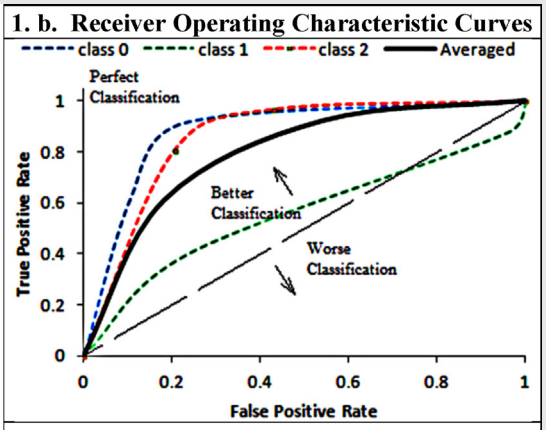
Total Instances = 80
Classification Accuracy = 63%
Correctly Classified Instances = 50
Incorrectly Classified Instances = 30

True Positive Rate
Class '0' = 63%
Class '1' = 40%
Class '2' = 81%

Averaged ROC Area = 76%

1.a. Confusion Matrix

System Prediction \ Clinical Ratings	Class '0'	Class '1'	Class '2'
Class '0'	**15**	8	1
Class '1'	6	**10**	9
Class '2'	1	5	**25**



2. Classification Test on RST-2
(More Stressful)

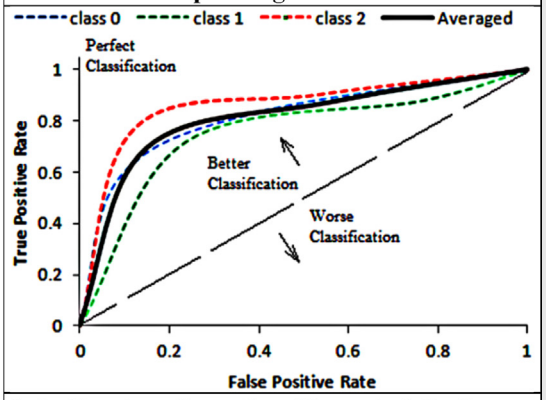
Total Instances = 80
Classification Accuracy = 70%
Correctly Classified Instances = 56
Incorrectly Classified Instances = 24

True Positive Rate
Class '0' = 60%
Class '1' = 72%
Class '2' = 78%

Averaged ROC Area = 80%

2.a. Confusion Matrix

System Prediction \ Clinical Ratings	Class '0'	Class '1'	Class '2'
Class '0'	**14**	8	2
Class '1'	3	**18**	4
Class '2'	2	5	**24**



3. Classification Test on RST-3
(Most Stressful)

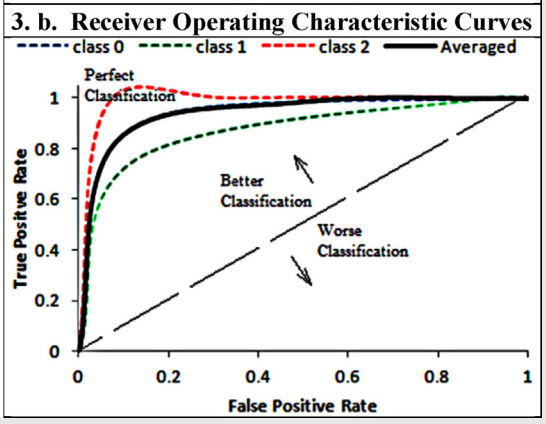
Total Instances = 80
Classification Accuracy = 84%
Correctly Classified Instances = 67
Incorrectly Classified Instances = 13

True Positive Rate
Class '0' = 75%
Class '1' = 76%
Class '2' = 97%

Averaged ROC Area = 91%

3.a. Confusion Matrix

System Prediction \ Clinical Ratings	Class '0'	Class '1'	Class '2'
Class '0'	**18**	6	0
Class '1'	3	**19**	3
Class '2'	0	1	**30**



Please cite this article in press as: Khan T, et al. Classification of speech intelligibility in Parkinson's disease. Biocybern Biomed Eng (2013), http://dx.doi.org/10.1016/j.bbe.2013.10.003

Table 3 – *Investigation 2*: SVM classification performance. (1) On overall speech dataset. (2) On training and testing set. In both cases, the averaged ROC curves are protruded toward perfect classification. Class '2' ROC curves (in 1b and 2b) are parallel to the level of perfect classification.

1. Classification Test on Overall Speech dataset
(including RST 1, 2 & 3)

Total Instances = 240
Classification Accuracy = 83%
Correctly Classified Instances = 198
Incorrectly Classified Instances = 42

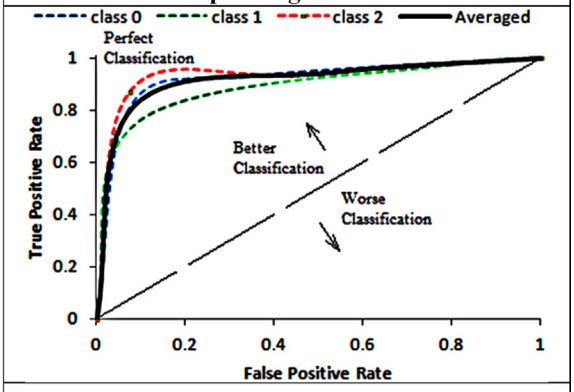
True Positive Rate
Class '0' = 84%
Class '1' = 76%
Class '2' = 87%

Averaged ROC Area = 89%

1.a. Confusion Matrix

System Prediction \ Clinical Ratings	Class '0'	Class '1'	Class '2'
Class '0'	**60**	9	3
Class '1'	10	**57**	8
Class '2'	5	7	**81**

1.b. Receiver Operating Characteristic Curves



2. Classification Test on Training and Test set

Training Set (RST 1 & 2):
Total Instances = 160
Classification Accuracy = 99%

Test Set (RST 3):
Total Instances = 80
Classification Accuracy = 82%
Correctly Classified Instances = 65
Incorrectly Classified Instances = 15

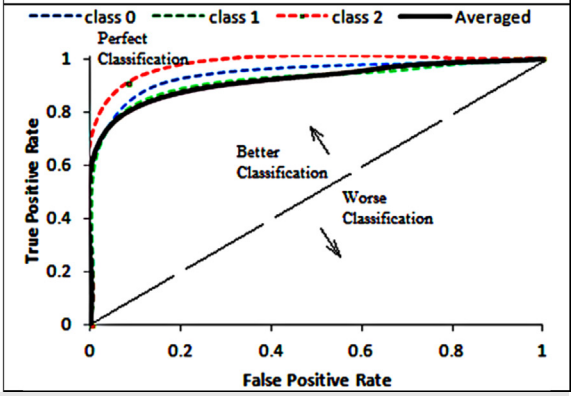
True Positive Rate
Class '0' = 59%
Class '1' = 80%
Class '2' = 100%

Averaged ROC Area = 90%

2.a. Confusion Matrix (Test Set)

System Prediction \ Clinical Ratings	Class '0'	Class '1'	Class '2'
Class '0'	**14**	6	4
Class '1'	0	**20**	5
Class '2'	0	0	**31**

2.b. Receiver Operating Characteristic Curves (Test Set)



In the second classification test, the increased textual difficulty in RST-2 improved the classification accuracy to 70% and TPRs to 60%, 72% and 78% in class 0, 1 and 2 respectively. Upgraded TPR in class 0 and 1 indicate the improved ability of features in discriminating between the samples in class 0 and 1. This finding supports that the mild symptoms which remained hidden in RST-1 were detected by the more demanding reading difficulty level in RST-2. The finding was confirmed in the third classification test when the subjects were exposed to the highest textual difficulty. The speech samples were classified with a marked improvement in the classification accuracy (84%) and the TPRs in class 0 (75%), 1 (76%) and 2(97%) respectively.

The receiver operating characteristic (ROC) curve is generally used to analyze the feasibility of a classification model independent of the class distribution [34]. A ROC curve can be plotted by taking false-positive rate of a symptom class on x-axis against the true-positive rate of that class on y-axis. An area under the ROC curve of 100% represents a 'perfect model' and an area near 50% corresponds to a 'worthless model'. In all the three classification tests, the mean ROC curves were protruded upwards from the diagonal threshold indicating the indubious distinction of samples in each symptom class (Table 2.1b, 2.2b and 2.3b). The area under the mean ROC curve in all the three tests remained above 75% ('good' model) and improved with the level of textual difficulty i.e. 76% ('good model'), 80% ('very good model') and 91% ('excellent model') in RST 1, 2 and 3 respectively.

5.2. *Investigation-2*

In order to analyze the performance of computed features and SVM model in classifying the total speech data set, a new matrix with dimensions 13 (features) × 240 (total speech samples; i.e. 3 RST × 80 samples = 240 samples) was formed for separation between the 3 levels of speech symptom severity. Data stratification with 10-fold cross-validation on the input vector produced an overall classification rate of 83% with the TPR of 84%, 76% and 87% in class '0', '1' and '2' respectively (Table 3.1). Noticeably, the area under the ROC curves in each symptom level was larger than the previous three classification tests, specifically in class '2' whose curve was parallel to the level of perfect classification. Class '2' is the combined representative class of moderate and severe speech symptoms. The marked distinction of samples belonging specifically to this group supports that the selected measures

Please cite this article in press as: Khan T, et al. Classification of speech intelligibility in Parkinson's disease. Biocybern Biomed Eng (2013), http://dx.doi.org/10.1016/j.bbe.2013.10.003

are representative features of speech symptoms in PD. Further tuning of the SVM suggests that the classification rate can be improved (to 85%) if the samples are stratified by 20-fold cross validation, keeping in mind that this type of experimentation with stratification parameters could possibly have introduced a bias. The TPRs were improved to 84%, 82% and 89% in class '0', '1' and '2' respectively with the improved averaged ROC area of 91%.

Another unbiased approach to validate the generalization performance of selected features, is to introduce novel unseen data to the classification model, with a statistical assumption that the new data will have a similar distribution to the data used in training the classifier. The selected 13 features were computed from 80 samples of RST 1 and 2 respectively, and were used to form a training set matrix of 13 (features) $\times$ 160 (samples; i.e. 80 RST-1 samples + 80 RST-2 samples = 160 samples). This matrix was then used to train the SVM classifier against the UPDRS-S ratings 0, 1 and 2. Another set of same features computed from 80 samples in RST-3 was used to form a test set matrix of 13 (features) $\times$ 80 (samples). This test set matrix was used for testing the trained classifier. A high accuracy (82%) was achieved by this scheme in classifying the test set between the 3 levels of UPDRS-S (Table 3.2) with an averaged ROC area of 90%. Specifically the samples in class '2' were predicted again with a very high true positive rate (100%).

6. Discussion

The results were compared with two other methods on objectification of running-speech in PD [6,8]. Both these methods utilized running speech samples together with sustained vowel phonation and diadochokinetic test samples to derive estimates of prosody, phonation and articulation respectively. The first method [6] used SVM on 11 different acoustic features to classify between 23 PD patients and 23 healthy controls and achieved a classification rate of 85%. The second method [8] applied simple naïve Bayes rule on 19 different acoustic features derived from the same dataset and produced a classification rate of 91.30%. The two methods employed similar prosodic measures as used in this paper (i.e. $F0_{STD}$, number of pauses, pause time and voice intensity) but different articulatory and phonatory measures. Both these methods performed two level classification of speech.

In order to compare the two methods with our approach, we merged the samples rated '1' (mild symptom) in our speech database with the samples rated '0' (normal) so that the new classification scheme was two level i.e. 'normal-mild' and 'moderate-severe'. The application of the configured SVM on a total of 13 features derived from 240 running speech samples produced a two-level classification accuracy of 92%. Although different methods can't be conclusively compared when tested on different data sets, these results suggest that the features from the running speech are enough to identify PD speech symptoms if they are able to track deficits in individual speech components.

Our experiments further suggest that the improvement in classification accuracy of speech symptoms is proportional to the increasing level of textual difficulty in our data set from mild PD stage. It was observed that the mild speech symptoms were undetected in the recitation of easy-to-read text. Even in this situation, the high values of Guttmann's μ_2 suggest that the CSD and MFCC were robust in characterizing between the speech symptom severity levels. In particular, the δ_{CSD} indicated very strong correlation with the clinical speech ratings and this correlation increased with increasing level of textual difficulty. The strong correlation between 4th MFCC and speech ratings directs further clinical research to explore the articulators responsible for frequency disturbances in this sub-band.

The implication of pitch features limits this classification model to English speakers only due to the fact that the sounding of phonemes in other languages may involve different fundamental tones. Besides, since the MFCC and CSD features do not incorporate the exclusive computation of fundamental frequency, the strong Guttman correlation between these features and clinical ratings suggests that these features have the potential to detect PD speech anomalies in languages other than English. In general, the high classification performance by the SVM supports this model and the selected pool of features as a suitable tool to categorize speech symptom severity levels in early stage PD.

Conflict of interest

The authors have filed a patent application for scoring speech using CSD.

Funding support

This research is a module of project 'PAULINA' which has been running in Dalarna University in collaboration with Animech and Nordforce Technology, and is funded by a grant from the Swedish Knowledge Foundation.

REFERENCES

[1] Olanow CW, Stern MB, Sethi K. The scientific and clinical basis for the treatment of Parkinson's disease. Neurology 2009;72:s1–36.
[2] Goetz CG, Stebbins GT, Wolff D, DeLeeuw W, Bronte-Stewart H, Elble R, et al. Testing objective measures of motor impairment in early Parkinson's disease: feasibility study of an at-home testing device. Mov Disord 2009;24:551–6.
[3] Fahn S, Elton R, The UPDRS Development Committee. Unified Parkinson's disease rating scale. Recent Dev Parkinson's Dis 1987;2:153–63.
[4] Pinto S, Ozsancak C, Tripoliti E, Thobois S, Dowsey PL, Auzou P. Treatments for dysarthria in Parkinson's disease. Lancet Neurol 2004;3(9):547–56.
[5] Londono JDA, Llorente JIG, Lechon NS, Ruiz VO, Dominguez GC. Automatic detection of pathological voices using complexity measures, noise parameters, and Mel-cepstral coefficients. IEEE Trans Bio-Med Eng 2011;58 (2):370–8.
[6] Rusz J, Cmejla R, Ruzickova H, Ruzicka E. Acoustic analysis of voice and speech characteristics in early untreated Parkinson's disease. In: Proc. 7th Intl. Workshop on MAVEBA. Firenze University Press; 2011. p. 181–4. 77.

Please cite this article in press as: Khan T, et al. Classification of speech intelligibility in Parkinson's disease. Biocybern Biomed Eng (2013), http://dx.doi.org/10.1016/j.bbe.2013.10.003

[7] Gelzinis A, Verikas A, Bacauskiene M. Automated speech analysis applied to laryngeal disease categorization. Comput Methods Programs Biomed 2008;91(1):36–47.
[8] Rusz J, Cmejla R, Ruzickova H, Ruzicka E. Objectification of dysarthria in Parkinson's disease using Bayes Theorem. In: Proc. 10th WSEAS. Vouliagmeni, Athens 2011. pp. 165–9.
[9] Zraick R, Dennie TM, Tabbal SM, Hutton TJ, Hicks GM, Sullivan PS. Reliability of speech intelligibility ratings using the Unified Parkinson Disease Rating Scale. J Med Speech Lang Pathol 2003;11(4):227–40.
[10] Looze CD, Ghio A, Scherer S, Pouchoulin G, Viallet F. Automatic analysis of the prosodic variations in Parkinsonian read and semi-spontaneous speech. In: Proc. Int. Conf. Speech Prosody. China: Tongji University Press; 2012.
[11] Paja MS, Falk TH. Automated dysarthria severity classification for improved objective intelligibility assessment of spastic dysarthric speech. In: Proc. 13th Annu. Conf. ISCA. Portland, Oregon 2012.
[12] Llorente JIG, Fraile I, Lechón RS, Ruiz NO, Vilda VGP. Automatic detection of voice impairments from text-dependent running speech. Biomed Signal Process Control 2009;4(3):176–82.
[13] Scholkopf B, Platt JC, Shawe-Taylor J, Smola AJ, Williamson RC. Estimating the support of a high-dimensional distribution. Neural Comput 2001;13(7):1443–71.
[14] Khan T, Westin J, Dougherty M. Cepstral separation difference: a novel approach for speech impairment quantification in Parkinson's disease. Biocybern Biomed Eng 2013. http://dx.doi.org/10.1016/j.bbe.2013.06.001.
[15] No C. International Phonetic Association, editor. Handbook of the international phonetic association: a guide to the use of the international phonetic alphabet. Cambridge: Cambridge University Press; 1999.
[16] Silbert N, Jong KD. Focus, prosodic context, and phonological feature specification: patterns of variation in fricative production. J Acoust Soc Am 2008;5:2769–79.
[17] Flanagan JL, Ishizaka K, Shipley KL. Synthesis of speech from a dynamic model of the vocal cords and vocal tract. Bell Syst Technol J 1975;54:485–506.
[18] Murphy P. Source-filter comparison of measurements of fundamental frequency perturbation and amplitude perturbation for synthesized voice signals. J Voice 2008;22:125–37.
[19] Murdoch BE, editor. Dysarthria: a physiological approach to assessment and treatment. Cheltenham, UK: Stanley Thornes; 1998.
[20] Midi I, Dogan M, Koseoglu M, Can G, Sehitoglu MA, Gunal DI. Voice abnormalities and their relation with motor dysfunction in Parkinson's disease. Acta Neurol Scand 2008;117:26–34.
[21] Kim S, Eriksson T, Kang HG. On the time variability of vocal tract for speaker recognition. In: 8th ICSLP. Jeju Island, Korea 2004.
[22] Freed D. Motor speech disorders: diagnosis and treatment, 2nd ed., New York: Delmar; 2012.
[23] Stevens S, Volkman J. The relation of pitch to frequency: a revised scale. Am J Psychol 1940;53(3):329–53.
[24] Jones HN. Prosody in Parkinson's disease. Perspect Neurophysiol Neurogenic Speech Lang Disord 2009;19 (3):71–6.
[25] Rosen KM, Kent RD, Delaney AL. Parametric quantitative acoustic analysis of conversation produced by speakers with dysarthria and healthy speakers. J Speech Lang Hear Res 2006;49(2):395–411.
[26] Le PN, Ambikairajah E, Epps J, Sethu V, Choi EH. Investigation of spectral centroid features for cognitive load classification. Speech Commun 2011;53(4):540–51.
[27] Ahmad S, Spanias AS. Cepstrum-based pitch detection using a new statistical V/UV classification algorithm. IEEE Trans Speech Audio Process 1999;7(3):333–8.
[28] Ma EM, Yiu EL. Suitability of acoustic perturbation measures in analyzing periodic and nearly periodic voice signals. Folia Phoniatr Logop 2005;57(1):38–47.
[29] Guttman L. A basis for scaling qualitative data. Am Sociol Rev 1944;9(2):139–50.
[30] Berger YG. A jackknife variance estimator for uni-stage stratified samples with unequal probabilities. Biometrika 2007;94:953–64.
[31] Guan W. New support vector machine formulations and algorithms with application to biomedical data analysis. PhD thesis. Georgia Institute of Technology; 2011.
[32] Ustun B, Melssen WJ, Buydens LMC. Facilitating the application of support vector regression by using a universal Pearson VII function based kernel. Chemom Intell Lab Syst 2006;81(1):29–40.
[33] Stone M. Cross-validatory choice and assessment of statistical predictions. J Roy Statist Soc Ser B (Methodological) 1974;36:111–47.
[34] Metz CE. Basic principles of ROC analysis. Semin Nucl Med 1978;8(4):283–98.

Please cite this article in press as: Khan T, et al. Classification of speech intelligibility in Parkinson's disease. Biocybern Biomed Eng (2013), http://dx.doi.org/10.1016/j.bbe.2013.10.003