PARKINSON’S DISEASE ASSESSMENT USING SPEECH ANOMALIES: A REVIEW

Taha Khan
Department of Computer Engineering
Dalarna University, 78188
Borlange, Sweden
tkh@du.se

ABSTRACT
This paper presents a detailed systematic literature review on Parkinson’s disease severity assessment methods based on speech impairment. Techniques that cope up the challenges of voice signal processing in real-time have been reviewed. An analysis on finding relevant acoustic parameters that are robustly influenced by Parkinson’s disease symptoms and medication-related changes was done. Time-series analysis of voice acoustics, feature extraction and pattern recognition methods are evaluated on the basis of portability and compatibility to mobile devices. A two-tier review methodology is developed based on the applicability of pathological voice recognition methods in real-time. The goal is to design an algorithm based on the review synthesis that can tackle real-time constraints in pathological voice recognition for the assessment of Parkinson’s disease severity. Pause detection, peak to average power rate clipping and zero thresholding rate calculations are proposed for human voice segmentation. Previous work depicts that these methods produce rich voice features in real-time. We suggest that these features may be further processed using wavelet transforms and used with a neural network for detection and quantification of speech anomalies related to Parkinson’s disease.

General Terms

Keywords
Parkinson Disease, Hypokinetic Dysarthria, Speech Recognition.

1. INTRODUCTION
The need for real world Parkinson’s disease (PD) treatment remains unmet for the vast majority of patients. Since PD is progressive, there is a need of follow up treatments over time. Medicine dosing needs to be adjusted daily with respect to physical exercise, food intake and mood [1]. Only 3-4% of the patients receive timely treatment [2]. Reasons for this are that individuals may have physical limitations which make it difficult for them to come for treatment or they may not have easy access to treatment provided by a therapist.

High-performance personal digital assistants (PDA) with advanced sound processing capabilities have potential to tackle the challenge of treatment accessibility [3]. Clinician monitoring and feedback can be preserved using the PDA since it has the portability to adapt clinician directed treatment to home self training[3]. Portable feedback devices for pathological individuals have been previously investigated. Zicker et al. [3] found that an individual with PD was able to modify her behavior when she received feedback from a device that her speech was too soft. Multisite treatment and the effective data to evaluate the delivery of consistent treatment among clinicians can be acquired using a PDA [3].

1.1 PD effects on Human Voice
PD is characterized by the loss of dopaminergic neurons in brain [4]. This loss results in dysfunction of basal ganglia pathway which is an essential part of the circuitry that mediates motor and cognitive functions. As a result of dopamine loss in basal ganglia, there can be a number of motor symptoms such as rigidity, akinesia (loss of control over voluntary muscle movements), bradykinesia (abnormal slowness in muscle movements), rest tremor, postural abnormalities, and speech dysfunction [4].

Physical symptoms that can occur in the limbs can also occur in the speech system. These symptoms are classified as hypokinetic dysarthria (HKD) [5]. "Dysarthria" refers to a speech disorder due to a change in muscle control. Hypokinetic means reduced movement or lack of coordination of face muscles [7]. Imagine vocal folds vibration during phonation creates pitch of the voice. Vocal folds vibrate quickly during high-pitched sounds and vibrate slowly during low-pitched sounds [6]. Many individuals with PD notice changes in pitch of their voices [7]. Monotone or lack of vocal inflection or melody in voice is also a common complaint[8].

Resonating system determines richness of the voice. Soft palate, located in the back of the mouth roof closes off the nasal cavity while speaking, except when producing nasal sounds such as “...ing”, “m...” or “n...”. The soft palate does not move normally in PD. A nasal quality in voice is produced as the air is leaked into the nose due to the soft palate’s inadequate movement [8].

Articulator system comprises of the face muscles, lips, tongue, and jaw. Imprecise articulation in PD is attributed to the reduced movement or lack of coordination of face muscles [7]. Imagine...
that your face was very cold and it was difficult to move your facial muscles, your speech becomes slurred and unclear [6].

1.3 Research Challenges

Previous research on HKD analysis had constraints due to laboratory controlled settings. Voice classification in real-time environment is challenging. Speech datasets could possibly be collected via phone calls by the PD patients so that speech assessment can be more timely accurate and productive. But due to the background noise or in the medium of transmission, human voice detection becomes difficult. Distance of patient’s mouth from the phone’s mouth-piece may also create problem in recognizing voice amplitude [9]. Male and female voices have different pitch properties [10]. All these issues can lead to incorrect speech segmentation.

Enough data sets are required for training and classifying the data according to clinical rating scale [11]. Due to the medical ethics and patient’s privacy issue this is another challenge.

Effect of speech rate on overall intelligibility in PD is a matter of debate. A comparison of results of previous studies on speech rate in PD is hampered by methodological differences [12]. The research should also contribute to prove the validity of the above hypothesis.

The paper has been sectioned as follows; section 2 reveals the aim of the study. Section 3 elaborates the synthesis of previous research. An approach for the characterization of speech impairment based on the review synthesis has been proposed in section 4. The paper has been concluded in section 5.

2. AIM OF THE STUDY

The purpose of this study is to review methods which can distinguish between pathological and normal human speech based on a synthesis from a literature review. The goal of the review is to find solutions to problems that occur in real time voice signal processing.

Acoustic features that are robust to medication changes in patients need to be determined. A choice of an efficient time-series analysis tool for voice signals needs to be made for better evaluation of segmented speech. A classification tool to classify pathological speech according to clinical speech assessments using the UPDRS (Unified Parkinson’s Disease Rating Scale) [21] for speech items ranked from 0-to-4 (normal to severe pathological state) needs to be chosen. The review methodology has been illustrated in figure 1.

3. SYNTHESIS OF PREVIOUS WORK

In this section, the literature on acoustic analysis of HKD in PD patient is reviewed and synthesized. Choice of publications for review is based on the validation of methodology, credibility of experimental techniques for real-time voice analysis and portability of algorithm for the mobile companion. Synthesis is presented in form of a time-line i.e. starting from the investigation of essential acoustic parameters for HKD recognition to the methods implemented for classification between HKD from the normal voice.

Keywords such as “Parkinson’s disease”, “Hypokinetic Dysarthria”, “Speech Recognition in real time” were used in the search engines. Histogram (in figure 2) displays the search engines used, number of total relevant literature found and the filtered publications based on review methodology. The most useful search hits for the synthesis have been found using the IEEE Explorer search engine. Other search engines used were Google Scholar, ELIN, EBRARY, and LIBRIS. Below is the synthesis of relevant literature evaluated.
Kris Tjaden [14] used vowel co-articulation and vowel syllables to study acoustic variability between 9 HKD males with PD and a group of 10 HC males [14]. Graded speaking rate was used to investigate effect of rate variation on co-articulation. Ratio of F2 formant onset frequency and F2 target frequency was used to infer co-articulation. Fricative F2 was obtained for speech stimuli and compared to F2 onset measures. The result (shown in figure 3a and 3b) was, ratios tended to be smaller for PD speakers than for HC [14].

Izworski et al. [10] worked with continuous sound analysis of PD patients for automatic HKD recognition. Sound emitted for a long period of time allowed for sound power analysis. Power value of each frame is calculated by summing up the values of the signal energy within the respective $x(a)$ and $x(a+m)$ limits of the t frame given in equation 1.1 where $m$ is the frame length.

$$P(t) = \sum_{k=1}^{m} |x(k)|^2$$  \hspace{1cm} (1.1) $$\text{Polynomial } p(x) \text{ of degree 4 that fits the data, } p(x(t)) \text{ to } P(t) \text{ in least squares sense is calculated to represent average values of } P(t) \text{ is given in equation 1.2.}$$ $$p(x) = p_0 x^n + p_1 x^{n-1} + ... + p_n x + p_{n+1}$$  \hspace{1cm} (1.2) $$\text{Sum of differences between vector } P(t) \text{ and } p(x) \text{ can be used as a voice stability parameter which is given as.}$$ $$\text{Stab} = \sum_{k=1}^{m} |P(k) - p(k)|$$  \hspace{1cm} (1.3) $$\text{Values obtained were compared with the values obtained in the control group as shown in figure 4a and 4b. Values of Stab parameter were much greater for patients (i.e. Stab>0.6) than for persons from the control group. With many patients, distinct and varying breaks in phonation were observed as shown in figure 4b.}

3.3 Automatic HKD Recognition

Above studies narrowed down the research to formant frequency of consonants and vowels as it contains most relevant acoustics for HKD analysis. Emily Budkowski [8] calculated PD severity ratings using fundamental frequency of Voice Onset Time (VOT is the interval between the initial articulatory release of a stop consonant and onset of voicing for the subsequent vowel). Participants’ speech ranged from mild to severe. He found out that Levodopa appeared to have greater effect on VOT within PD group and it is a measure of medication related rate change. Results indicated that all participants exhibited HKD in OFF medication state.
With healthy persons, gradual quietening occurred, whereas PD patients ended the emission abruptly. This audio signal is transformed into Fourier series and it is found out that changes are observable especially in the vowels articulation. Results constitute the beginning of tests concentrated on automatic voice classification [10].

Artificial Neural Networks (ANN) can be used as a statistical tool to formulate the boundaries between speech abnormalities using the voice features. Mehmet F, et al. [18] used ANN for feature selection of pathological voice datasets. A fuzzy logic based ANN (called ANFC: Adaptive Neuro-Fuzzy Classifier) gave best recognition results. Datasets for this study consisted of 195 sustained vowel phonation from 31 people, of which 23 were diagnosed with PD. Pathological voice features such as fundamental frequency of voice, shimmer and jitter were used for classification. ANFC produced 94.35% accuracy for classification of PD group dataset.

3.4 Automatic HKD recognition using hybrid approach: Wavelet Analysis and ANN

L. Salhi et al [19] used Linear Predictive coding method on PD patients voice on formants F1, F2 and F3. Comparing with the normal voice, pathological voice showed high variations.

![Image](image1)

**Figure 5. A Hybrid approach for HKD recognition [11].**

Although these methods distinguished HKD with the normal voice but the problem was, these methods do not produce quantification values to make the decision. They introduced a hybrid approach using wavelet analysis and ANN. Normalized energies and entropies of wavelet coefficients were used to formulate feature vector of speech sample. The feature vector was used to identify impaired speech. A 3-layered feed forward ANN with Back Propagation (BP) algorithm was used for classification. The flowchart of this algorithm is shown in fig. 5.

Using Discrete Wavelets Transforms (DWT) a clear difference was noticed between wavelet evolutions of HKD and normal voice [19]. A visual pattern is shown in figure 6a and 6b.

ANN classification with five DWT coefficients produced 90% classification rate for pathological voice and 100% classification rate for normal voice [19]. Experiment proved that the formant analysis methods using Fourier Transform is not a suitable tool for speech analysis [11]. Since speech is a highly non-stationary signal, wavelet transform proved to be a better tool for analysis of non stationary signals [19] as it is useful in localizing a symptom both in time and frequency scales.

![Image](image2)

**Figure 6. Wavelet Analysis [11].**

Above synthesis reveals that the assessment of PD is possible through speech datasets via reduced voice loudness, mono pitch, short rushes, prolonged syllables, long pauses, and reduced phonation time. The synthesis depicts that HKD effected PD patients can be effectively classified on the basis of UPDRS ratings using classification tool like ANN.

4. PROPOSED ALGORITHM

The biggest drawback found in the previous work was the controlled setting of experiments in laboratory environment [9, 20]. Since this study is aimed for real time processing of speech signals in a mobile companion, the biggest challenge is voice segmentation and noise removal in the speech signals achieved during phone calls in home environment. An algorithm design is proposed to tackle this challenge. First pre-processing of voice signals needs to be done succeeded with voice segmentation. The segmented voice may be decomposed using wavelet and later trained in the neural network for classification. Below are the details of each step.

4.1 Voice Pre-processing

Most intelligibility relevant speech segments (consonants and vowels) contains minority of speech energy [9]. Intelligibility can be improved by amplifying if total energy remains constant. Amplitude compression is a dynamic function which provides larger amplification when input has low power over certain time. The effect is that, power of sound varies less in time and hence it is less noisy [9]. Speech enhancements can be done using High Pass filter and clipping [9]. Also the speech signal does not have unique maximum amplitude. Some waveforms are asymmetric than others, that is, they reach higher peaks in positive amplitude than negative or vice versa. More asymmetric waveforms have larger peak power. Human speech has “Peak to Average Peak Power” (PAPR) [9]. In order to remove noise, peaks over or less
than PAPR threshold range can be clipped off in the speech signals. The speech signal can be further “windowed” using a Hamming Window [9].

Once the signal is pre-processed, voice can be segmented based on vocal sounds. Syllable segmentation of speech is important for correct voice classification because it reduces search space effectively.

An automatic processor should succeed at identifying silence, vocalic, fricative and nasal sounds and stop gaps to provide gains in speech intelligibility [19]. In case of HKD, segmentation is difficult because syllable units are spread roughly via intensity changes but exact boundary positions are elusive in successive vowels.

4.2 Decomposition with Wavelet Transforms

Once the voice signal is segmented, speech can be decomposed in transient and tonal components using Discrete Wavelet transforms (DWT) [11].

In disordered speech, the non-stationary behavior of the pitch can be analyzed using wavelets. The criteria for selecting a proper mother wavelet is to have a wavelet function with enough number of vanishing moments in order to represent the salient features of the disturbance. At the same time, this wavelet should provide sharp cut-off frequencies. Selected wavelet should be orthogonal normal. Daubechies 40 (Db40) shows sharper cut off frequency compared with the others. Using Db40, leakage energy between different resolution levels can be reduced [11]. Wavelet transforms outputs energy coefficients that represent the signal in time and frequency. Since wavelet energy coefficients are sensitive to transients, areas of changes can be amplified and detected using Laplace derivative. This can make the speech transients very rich [9].

Table 1. Voice Segmentation based on ZCR analysis [20].

<table>
<thead>
<tr>
<th></th>
<th>With ZCR Rules</th>
<th>Without ZCR Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syllables</td>
<td>856</td>
<td>856</td>
</tr>
<tr>
<td>Correct</td>
<td>819</td>
<td>803</td>
</tr>
<tr>
<td>Deletion</td>
<td>37</td>
<td>53</td>
</tr>
<tr>
<td>Insertion</td>
<td>64</td>
<td>79</td>
</tr>
<tr>
<td>Correct Rate</td>
<td>95.7%</td>
<td>93.8%</td>
</tr>
<tr>
<td>Error Rate</td>
<td>11.8%</td>
<td>15.4%</td>
</tr>
</tbody>
</table>

For continuous speech segmentation, zero crossing rates (ZCR) provide spectral information at low computational cost. ZCR is the rate when a waveform crosses time axis or changes its algebraic sign. Consonants have high ZCR whereas vowels have low ZCR [20]. Therefore an onset of high ZCR means beginning of consonant and it should be a starting boundary. An offset of high ZCR means an endpoint of consonant and it should be an ending boundary [20].

Table 1 depicts the comparison between voice segmentation with ZCR rules and without ZCR rules. Segmentation accuracy is drastically improved when ZCR rules are applied to the voice segmentation algorithm as shown in table 1.

4.3 Classification Using ANN

Resulting DWT coefficients can be used as feature vector for training artificial neural networks using Back Propagation algorithm. In order to obtain optimal results, nature of the input coefficients vector is to be varied for training in ANN.

As suggested by L. Salhi et al.[19], three different input vectors can be used for ANN. They are (1) DWT coefficients (2) DWT energy coefficients and (3) DWT entropy coefficients. The resulted pathological voice can be further classified according to UPDRS ratings [21] Based on the classification results, the level adjustments of medication can be made and other likely
treatments can be opted. The flowchart of the proposed algorithm is shown in figure 8.

5. CONCLUSION

This review evaluated HKD recognition algorithms to develop a tool for PD severity assessment using speech datasets of PD patients. Previous research had constraints due to the controlled settings of the laboratory environment. Possible solutions for voice analysis in real-time have been investigated. Voice acoustic parameters that were robust to medication changes are evaluated. Since voice signals are highly non-stationary, focus was laid on quantization of speech signals. During the study it was found that time-frequency methods fail to quantify the voice signals over the time-series based on the voice frequency. Wavelet transformation is found promising for voice analysis because it quantizes the non-stationary voice signals over the time-series using scale and translation parameters. In this way voice intelligibility in the waveforms can be analyzed in each time frame. During the review, artificial neural network was found to be a promising tool for voice classification.

This review is a basis for future work to conceive a HKD recognition tool in a PDA for self-treatment of PD patients. Also clinicians will be aided from feedback generated by this tool in follow up of patients who suffer from HKD.

6. ACKNOWLEDGEMENT

This research is a module of the project “E-Motions”. This project has been running in Dalarna University in collaboration with Abbott Laboratories and Nordforce Technology and is funded by a grant from Swedish Knowledge Foundation.

7. REFERENCES


