Detecting Patients with Parkinson's disease using Mel Frequency Cepstral Coefficients and Support Vector Machines

Achraf Benba[1], Abdelilah Jilbab[2], and Ahmed Hammouch[3]

[1, 2, 3]Laboratoire de recherche en génie électrique, Ecole Normale Supérieure de l'Enseignement Technique (ENSET), Mohammed V University, Rabat, Morocco
[1]achraf.benba@um5s.net.ma

Abstract: In order to develop the assessment of speech disorders for detecting patients with Parkinson's disease (PD), we have collected 34 sustained vowel / a /, from 34 subjects including 17 PD patients. We subsequently extracted from 1 to 20 coefficients of the Mel Frequency Cepstral Coefficients (MFCCs) from each individual. To extract the voiceprint from each individual, we compressed the frames by calculating their average value. For classification, we used the Leave-One-Subject-Out (LOSO) validation scheme and the Support Vector Machines (SVMs) with its different types of kernels, (i.e.; RBF, Linear and polynomial). The best classification accuracy achieved was 91.18% using the first 12 coefficients of the MFCCs by Linear kernels SVMs.

Keywords: Voice analysis, Parkinson's disease, Mel Frequency Cepstral Coefficients, Voiceprint. Leave One Subject Out, Support Vector Machines.

1. Introduction

Parkinson's disease (PD) is actually the second most common neurological syndrome after Alzheimer's disease. During its course, PD causes diverse symptoms and it influences the system which controls the execution of learned motor plans such as walking, talking or completing other simple tasks [1] [2] [3]. For this purpose, the assessment of the quality of speech, and the identification of the causes of its degradation in the context of Parkinson's disease based on phonological and acoustic cues have become main anxieties of clinicians and speech pathologists. They have become more attentive to techniques or methods external to their domain, which might offer them extra information for the diagnosis and the assessment of Parkinson's diseases. As is known, PD generally causes voice weakening in approximately 90% of patients [4] and affects people whose age is over 50 years, making the physical visits for diagnosis, monitoring and treatment extremely difficult [5] [6]. Clinicians and the speech pathologists have adopted subjective methods based on acoustic cues to distinguish different disease states in PD patients. Recent studies use measurements of voice quality in time, spectral and cepstral domains [7] in order to develop more objective assessments to detect voice disorders. These measurements contain fundamental frequency of vocal oscillation (F0), absolute sound pressure level, jitter, shimmer, and harmonicity [1] [8] [9].

As for disorders, Little et al. [1] aimed to discriminate healthy people from people with PD by detecting dysphonia. In their study, sustained vowel "a" phonations were recorded from 31 subjects, of whom 23 were diagnosed with PD. They then selected ten highly uncorrelated measures, and found four that, in combination, lead to overall correct classification performance of 91.4%, using a kernel Support Vector Machine (SVM). BetulErdogdu Sakar et al [6] analyzed multiple types of sound recordings collected from people with Parkinson's disease. The extracted features were fed into SVM and k-Nearest Neighbor (k-NN) classifiers for PD diagnosis by using a leave-one-subject-out (LOSO) cross-validation scheme and summarized Leave-One-Out. To distinguish healthy subjects from PWP, most studies use SVM classification [1] [6]. Success of the diagnostic system is measured with true positive (TP), true negative (TN), false positive (FP) and false negative (FN) rates.

Received: December 25th, 2014. Accepted: April 20th, 2015

In this study we focused on the measurements and the assessments of speech disorders in cepstral domain by applying Mel-Frequency Cepstral Coefficients (MFCCs) which have been traditionally used in speaker recognition and identification applications [10]. The automatic assessment of speech disorders in the context of Parkinson's disease using the Mel-frequency cepstral coefficient (MFCCs) was first proposed by Fraile et al [11] [12]. In the last few years, the usage of the MFCCs has been extended to the assessment of speech quality for clinical applications [10]. We have extracted MFCCs from the speech signals provided in a database and calculated the average value of the frames to get the voiceprint of each individual. We then used a Leave One Subject Out (LOSO) validation scheme with Support Vector Machines for feature classification in order to discriminate patients with Parkinson's disease from healthy subjects.

This paper is organized as follows: the subject database is described in section II. The MFCCs processes are presented in section III. The methodology of this research is presented in section IV. The obtained results are presented in Section V and conclusion in Section VI.

2. Data acquisition

The indications of speech disorders associated with disturbances of muscular control of the speech organs can be measured and detected by analyzing various features of speech. The dataset collected in this study belong to 17 patients with PD (6 female and 11 male) and 17 healthy individuals (8 female and 9 male). Voice recordings were done through a standard microphone at a sampling frequency of 44,100 Hz using a 16-bit sound card in a desktop computer. The microphone was placed at a distant of 15 cm from subjects and then, they were asked to pronounce the sustained vowel /a/ at a comfortable level.

All the recordings were made in stereo-channel mode and saved in WAVE format; analyses were done on these recordings. All the voice samples used in this study were collected by Mr. M. Erdem Isenkul from the Department of Computer Engineering at Istanbul University, Istanbul, Turkey.

3. MFCCs Processes

Our first purpose is this section, was to transform the speech signal to some type of parametric representation for more analysis and processing [13]. The speech waveform is a slow time varying signal which is called quasi-stationary [13]. When it is perceived over a short period of time, it seems fairly stable [13]. Nonetheless, over a long period of time, the speech signal changes its waveform. Thus, it should be characterized by doing short-time spectral analysis [13].

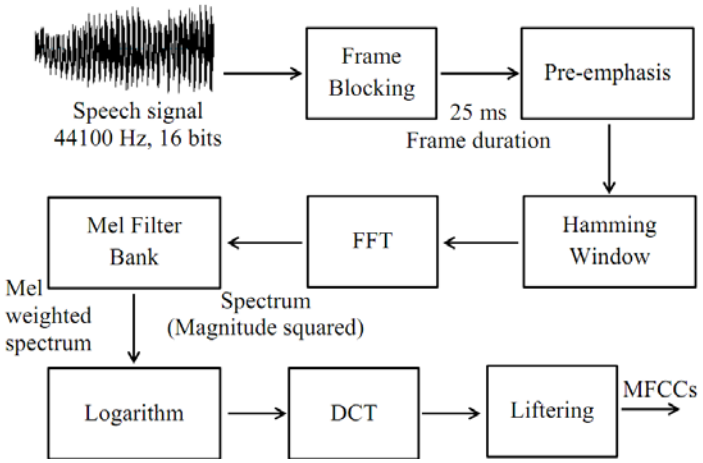

Figure 1. Block diagram of Mel Frequency Cepstral Coefficients (MFCCs) extraction

The calculation of the MFCCs is based on a Mel scale. This scale estimates the frequency perception of the human ear [14]. It was calculated in such a way that 1000 Hz corresponds to 1000 Mel. The Mel scale is approached by a bank of (15 to 30) triangular filters spaced linearly up to 1 kHz and logarithmic above 1 kHz [15]. The technique of calculating the MFCCs is shown in Figure 1 and described in the next paragraphs.

A. Framing

The examination of the speech signal over a long time periods shows that the speech waveform is not stationary [13]. For this purpose, it is essential to proceed with the technique of short time analysis. Generally, within the interval of 10 ms to 30 ms, the speech waveform can be considered stable [13]. The rate of movement of the speech articulators is limited by physiological limitations [13]. For this reason, the analysis of speech signal is done within uniformly frames of typical duration (from 10 to 30 ms) [13]. In frame blocking, the speech waveform is divided into frames of N samples. Neighboring frames should be separated by M (M < N) [13] [15].

B. Pre-emphasis

In this section, we increase the energy in the speech waveform, by accentuating the higher frequencies [15]. For that we apply the first order difference equation to the voice samples $\{s_n, n = 1,..., N\}$ [14]:

$$s'_n = s_n - k.s_{n-1} \tag{1}$$

here k is the pre-emphasis coefficient and it should be within the range of $0 \le k < 1$ [14]. In this work we used a pre-emphasis coefficient of $k = 0{,}97$.

C. Hamming windowing

The speech signal is a real signal, so it is finite in time; thus, a processing is only possible on limited number of samples [14]. To this end, the succeeding step of MFCCs process is to window each frame. The purpose of this step is to reduce signal discontinuities, and make the ends smooth enough to connect with the beginnings [14]. This was realized by using Hamming window to taper the signal to zero in the beginning and in the end of each frame, by applying the following equation to the voice samples $\{s_n, n = 1,..., N\}$ [14]:

$$s'_n = \left\{0{,}54 - 0{,}46 \cdot \cos\left(\frac{2\pi(n-1)}{N-1}\right)\right\} \cdot s_n \tag{2}$$

D. Fast Fourier Transform (FFT)

The aim of this step is to transform each frame of N samples from time domain into frequency domain using the Fast Fourier Transform (FFT) [13]. We used the FFT since it is a fast algorithm the implement the Discrete Fourier Transform (DFT) [13]. The DFT is defined on the set of N samples (S_n) as follow [13]:

$$S_n = \sum_{k=0}^{N-1} s_k e^{-2\pi jkn/N} \ , n = 0,1,2,..., N-1 \tag{3}$$

E. Filter bank analysis

Psychophysical research has revealed that human ear resolution of frequencies does not follow a linear scale across the audio spectrum [14]. Consequently, for each frequency measured in Hertz (Hz), a subjective pitch is measured on the Mel scale.

The general form of the filter bank is represented in Figure 2. As can be seen, the Mel-frequency scale is linearly spaced less than 1000 Hz and logarithmic above 1000 Hz and the filters have a triangular form [15].

To compute a Mel for a given frequency, we use the following approximate equation [14]:

$$Mel(f) = 2595 \cdot \log_{10}\left(1 + \frac{f}{700}\right) \tag{4}$$

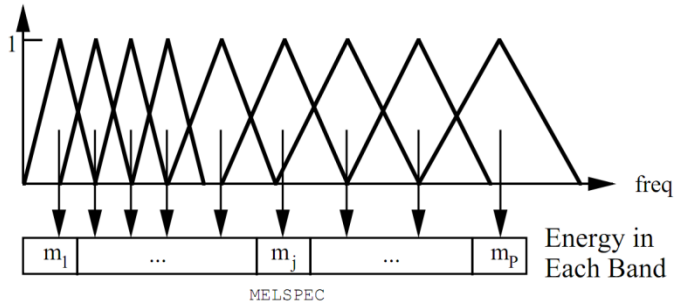

Figure 2. Mel-Scale Filter Bank [14]

F. Logarithm / DCT

In this phase, the Mel-Frequency Cepstral Coefficients (MFCCs) are calculated from the log filter bank amplitudes (m_j) through the Discrete Cosine Transform (DCT) [14]:

$$c_i = \sqrt{\frac{2}{N}} \sum_{j=1}^{N} m_j \cdot \cos\left(\frac{\pi i}{N}(j - 0{,}5)\right) \tag{5}$$

where$N$ is the number of filter bank channels.

G. Liftering

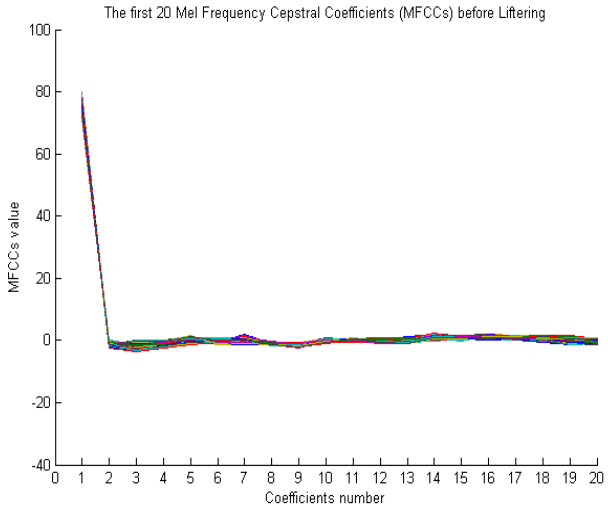

Figure. 3. The first 20 Mel Frequency Cepstral Coefficients (MFCCs) before Liftering extracted from PD patient

The main advantage of cepstral coefficients is that they are uncorrelated [14]. However, the problem with them is that the cepstral coefficients of higher order are fairly small [14], as shown in Figure 3. For this purpose, it is essential to rescale these cepstral coefficients to have quite similar magnitudes (Figure 4) [14]. This was realized by Liftering the cepstral coefficients according to the following equation [14]:

$$c'_n = \left(1 + \frac{L}{2} \cdot \sin\left(\frac{\pi \cdot n}{L}\right)\right) \cdot c_n \tag{6}$$

where$L$ is the Cepstral sine lifter parameter. In this work, we used L=22.

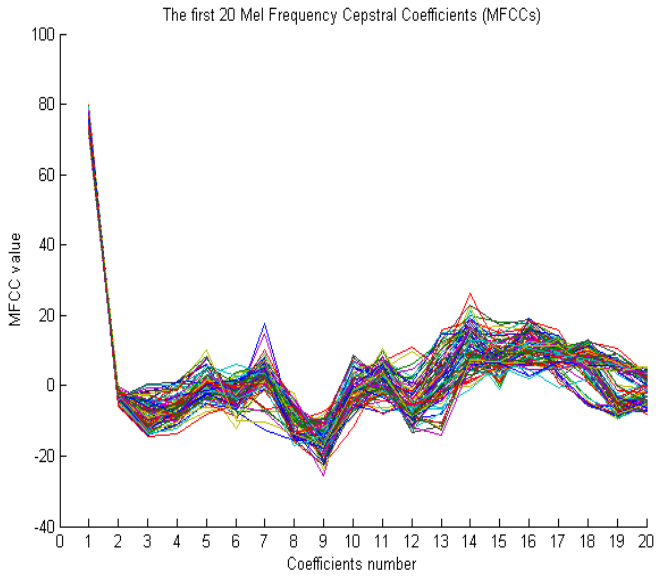

Figure 4. The first 20 MeL Frequency Cepstral Coefficients (MFCCs) after Liftering extracted from PD patient

4. Methodology and results

In our previous work [16], weextracted from each voice sample using the same database used in this study, cepstral coefficients of the MFCCs. The number of coefficients extracted ranged from 1 to 20. We proceeded in this way to get the exact number of coefficients required for achieving the best diagnostic accuracy. The MFCCs extracted from each sample contains a large number of frames which require extensive processing time for classification and this prevents making the accurate diagnostic decision. To overcome this problem, we used a method of lossy data compression known as vector quantization (VQ). We applied this method over 20 MFCCs that have already been extracted from each voice sample. The best average result obtained was 82%.

In another work [17], we have used Perceptual linear prediction (PLP) technique instead of MFCCs. The frames of the PLP were compressed using VQ, with six codebook sizes. We used the technique LOSO and SVMs classifier with two types of kernels; RBF and Linear. The obtained results using the codebook size of 1 were no stable. Therefore, we proceeded to a bench of 100 trials. The best average accuracy obtained was 75.79%.

In our precious work [18], we extracted from 1 to 20 coefficients of the Perceptual Linear Prediction (PLP) from each individual in the same database. We then extracted the voiceprint from each individual; we compressed the frames by calculating their average value. We used for classification LOSO and SVMs with its different types of kernels, (i.e.; RBF, Linear and polynomial). The best classification accuracy achieved was 82.35% using the first 13 and 14 coefficients of the PLP by Linear kernels SVMs.

The first phase in this study was to build a dataset containing voice samples recordings of normal individuals and patients with Parkinson's disease. Ultimately, 17 voices were collected from both groups which gave us 34 records. All individuals (Normal and PD) were invited to pronounce the sustained vowel / a / at a comfortable level. The database used in the context of this study was collected in [5]. In their study, multiple voice samples per subject were collected during the pronunciation of numbers from 1 to 10, four rhymed sentences, nine words in Turkish language along with sustained vowels "a", "o", and "u" from 40 people, 20 with Parkinson's disease. In their work they were able to detect PD patients using multiple types of voice recording with a best classification accuracy of 85%.

In this study, we extracted from each voice sample, multi cepstral coefficients of the MFCCs. The extracted number of coefficients ranged from 1 to 20. We proceeded in this way to get the optimal number of coefficients needed for the best classification accuracy.

The MFCCs extracted from each voice sample contains a large number of frames which demand an extensive processing time for classification and prevents making the correct diagnostic decision.

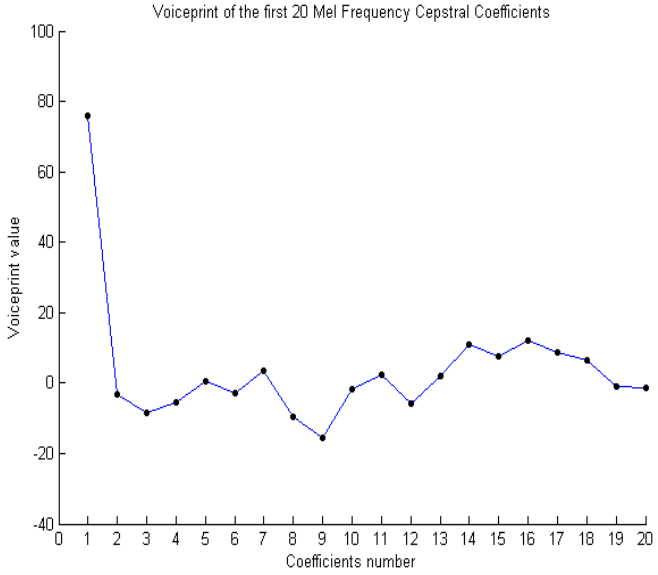

Figure. 5. Voiceprint of the first 20 Mel Frequency Cepstral Coefficients (MFCCs) extracted from PD patient

To overcome this problem, we calculated the average value of these frames to get the voiceprint of each individual [16].

To train and validate our classifier, we used a method of classification called LOSO, i.e., we left out all the compressed frames of the MFCCs of one individual to be used for validation as if it were an unobserved individual, and trained a classifier on the rest of the compressed frames of other individuals [6]. We used the LOSO classification scheme iteratively for each

coefficient per subject until all 20 coefficients per subject. In this work, we used the SVM classifier with its different types of kernels, i.e.; RBF, Linear and polynomial.

To measure the success of our classifier and select the best coefficient needed for the best diagnosis accuracy, we used an evaluation metrics which contain accuracy, sensitivity and specificity. Accuracy is the ratio of correctly classified instances divided to the whole instances [6] [19]:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{7}$$

where TP is the number of true positives (healthy), TN true negatives (pathological), FP false positives (pathological but it shown as healthy), and FN false negatives (healthy but it shown as pathological). Sensitivity is a statistical measure of correctly classified positive and Specificity is a statistical measure of negative instances [6] [19].

$$Sensitivity = \frac{TP}{TP+FN} \tag{8}$$

$$Specifisity = \frac{TN}{TN+FP} \tag{9}$$

It is clear from figure 6 that the best classification accuracy of 91.18% was achieved using linear kernel of SVM with the first 12nd coefficients of the MFCCs. This means that 31 individuals were correctly classified and 3 individuals were wrongly classified. The classification accuracy using linear kernel SVM sreaches its maximum values between the 10th and the 13th first coefficients. From the same figure, the maximum accuracy of 73.53% was achieved by the two other kernels. For the Polynomial kernel it was achieved using only the first coefficient while with the RBF kernel it was achieved using the first and the second coefficient. This means that 25 individuals were correctly classified and 9 individuals were wrongly classified. It is also clear that when we use a larger coefficient number, the accuracy of diagnosis decreases until reaching 2.94% using a classification with RBF kernels, and for polynomial kernel the accuracy results are almost stable and varies between 73.53% and 52.94%.

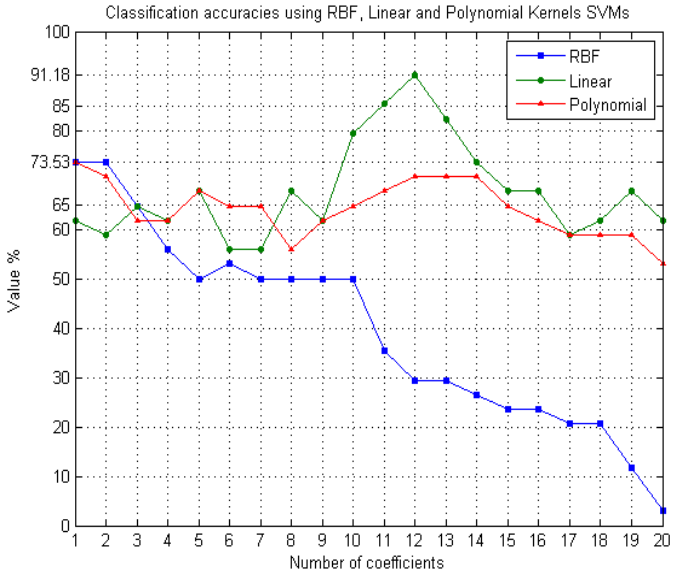

Figure. 6. Accuracy results using RBF, Linear and Polynomial Kernels SVMs

In this study, sensitivity results represent the classification metrics of healthy individuals. Based on the results of figure 7, it is clear that the maximum sensitivity of 100% was achieved using linear kernel SVM with the first 12nd coefficients of the MFCCs. This means that all the 17 normal voice samples used in this study were correctly classified which means that by using this method we can perfectly detect healthy individual. From the same figure, the maximum sensitivity of 94.12% was achieved by the two other kernels. For the Polynomial kernel it was achieved using only the first coefficient while with the RBF kernel it was achieved using the first and the second coefficient. This means that 16 individuals were correctly classified and only 1 individual was wrongly classified. It is also clear that when we use a larger coefficient number, the sensitivity of diagnosis decreases until reaching 5.88% using a classification with RBF kernels, and for polynomial kernel the sensitivity decreases between the 2nd and the 10th coefficients until reaching 58.82% and becomes almost stable between the 10th and the 20th coefficients.

In this study, specificity results represent the classification metrics of patients with Parkinson's disease. Based on the results of figure 7, it is clear that the maximum sensitivity of 82.35% was achieved using linear kernel SVM with the first 12nd coefficients of the MFCCs. This means that from the 17 PD patients, 14 were correctly classified and 3 patients were classified as healthy. Those 3 classifications are the wrong classifications found in the accuracy result of 91.18% represented in figure 6. From the same figure, the maximum specificity of 76.47% was achieved by the polynomial kernel with the first 5th coefficient of the MFCCs. This means that from the 17 PD patients, 13 were correctly classified and 4 patients were classified as healthy this is much close to the obtained result using linear kernel with the 12nd coefficient, and the same as the obtained results using the 10th and the 11th coefficients of MFCCs. The maximum specificity result of 58.82% was achieved using RBF kernel with the 3th coefficient of the MFCCs and decrease until reaching 0%.

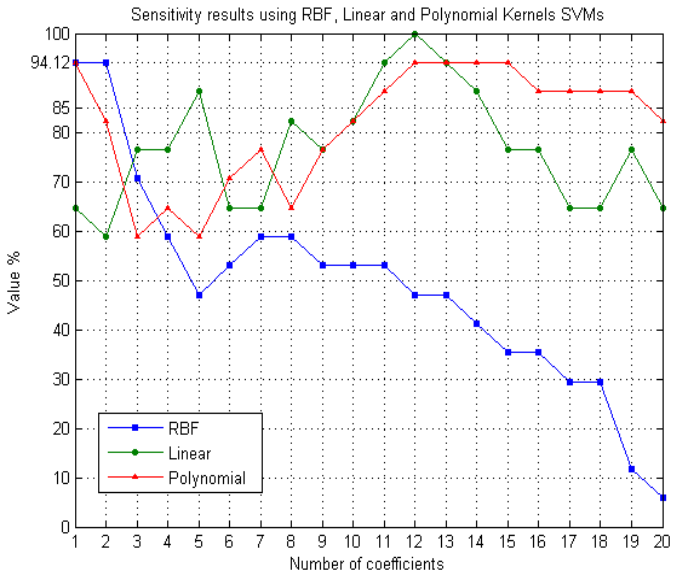

Figure. 7. Sensitivity results using RBF, Linear and Polynomial Kernels SVMs

From all these results we can conclude that the best model to discriminate PD patients from healthy subject is the linear model using the first 12 coefficients of the MFCCs. This model remains the best model created with only 3 incorrect diagnoses of PD patients, but it is perfect for detecting healthy individuals. The obtained results outclassed the results of all previous studies which have been achieved using the same database used in the context of this study.

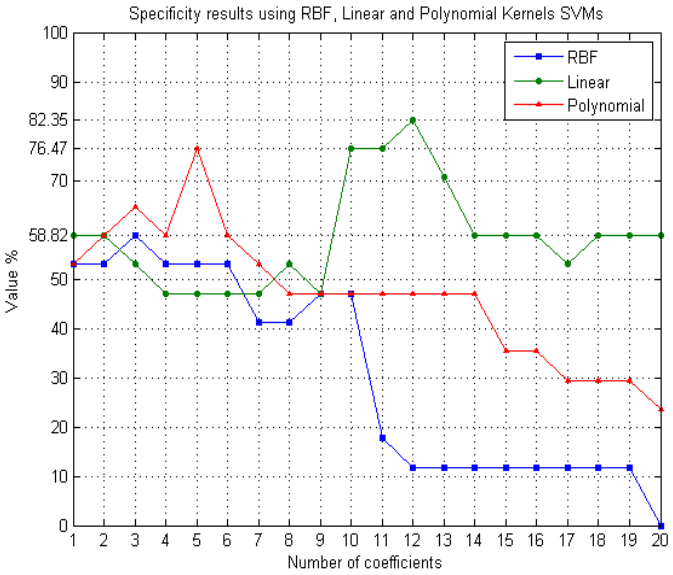

Figure. 8. Specificity results using RBF, Linear and Polynomial Kernels SVMs

5. Conclusion

Dysarthria symptoms associated with Parkinson's are a slow process whose first stages may go unnoticed. To enhance the assessment of Parkinson's disease we collected a variety of voice recordings from different individuals during the pronunciation of sustained vowel /a/. The extracted MFCCs from different participants contain many frames which take maximum processing time in the classification process, and prevent making correct diagnosis.

The compression of the MFCCs frames using their average value to extract the voiceprints from individuals, has shown to be a good parameter for the detection of voice disorder in the context of Parkinson's disease, showing a maximum classification accuracy of 91.18% using the first 12 coefficients of the MFCCs by Linear Kernels SVMs.

6. Acknowledgment

The authors would like to thank Mr. Erdem Isenkul from Department of Computer Engineering at Istanbul University, Mr Thomas R. Przybeck and Mr. Daniel Wood from United States Peace Corps Volunteers (Morocco 2013-2015), and all of the participants involved in the dataset collection process.

7. References

[1]. Little, Max A., et al. "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease." *Biomedical Engineering, IEEE Transactions on* 56.4 (2009): 1015-1022.

[2]. Ishihara, L., and C. Brayne. "A systematic review of depression and mental illness preceding Parkinson's disease." *Acta Neurologica Scandinavica* 113.4 (2006): 211-220.

[3]. Jankovic, Joseph. "Parkinson's disease: clinical features and diagnosis."*Journal of Neurology, Neurosurgery & Psychiatry* 79.4 (2008): 368-376.

[4]. S. B. O'Sullivan, T. J. Schmitz, "Parkinson disease," *Physical Rehabilitation, 5th ed.* Philadelphia, PA, USA: F. A. Davis Company, 2007, pp. 856–894.2007, pp. 856–894.

[5]. Huse, Daniel M., et al. "Burden of illness in Parkinson's disease." *Movement disorders* 20.11 (2005): 1449-1454.

[6]. Sakar, Betul Erdogdu, et al. "Collection and Analysis of a Parkinson Speech Dataset With Multiple Types of Sound Recordings." *Biomedical and Health Informatics, IEEE Journal of* 17.4 (2013): 828-834.
[7]. U. K. Rani, M.S. Holi, "Automatic Detection of Neurological Disordered Voices Using Mel Cepstral Coefficients and Neural Networks," *2013 IEEE Point-of-Care Healthcare Technologies (PHT)*, Bangalore, India, 16 - 18 January, 2013.
[8]. M. A. Little, P. E. McSharry, S. J. Roberts, D. A. Costello, I. M. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection." *Biomed. Eng. Online*, 2007.
[9]. D. A. Rahn, M. Chou, J. J. Jiang, Y. Zhang, "Phonatory impairment in Parkinson's disease: Evidence from nonlinear dynamic analysis and perturbation analysis." *J. Voice.* 21:64-71, 2007.
[10]. T. Kapoor, R. K. Sharma, "Parkinson's disease diagnosis using Mel-frequency cepstral coefficients and vector quantization," *International Journal of Computer Application*, Vol. 14, no. 3, Jan. 2011.
[11]. R. Frail, JI. Godino-Llorente, N. Saenz-Lechon, V. Osma-Ruiz, C. Fredouille, "MFCC-based remote pathology detection on speech transmitted through the telephone channel," *ProcBiosignals*, Porto, 2009.
[12]. A. Jafari, "Classification of Parkinson's disease patients using nonlinear phonetic features and Mel-frequency cepstral analysis," *Biomed. Eng. Appl. Basis Commun*, Vol. 52, no. 4, 2013. Available: http://www.worldscientific.com
[13]. Ch. S. Kumar, P. R. Mallikarjuna, "Design of an automatic speaker recognition system using MFCC, Vector Quantization and LBG algorithm," *International Journal on Computer Science and Engineering*, Vol. 3, no. 8, 2011.
[14]. S. Young, G. Evermann, T. Hain, D. Kershaw, X. Liu, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, P. Woodland, "The HTK Book (for HTK Version 3.4)," Copyright. 2001-2006, Cambridge University Engineering Department.
[15]. J. Martinez, H. Perez, E. Escamilla, M. M. Suzuki, "Speaker recognition using mel frequency cepstral coefficients (MFCC) and Vactor Quantization (VQ) techniques," *IEEE Electrical Communications and Computers*, Cholula, Puebla, Feb 2012, pp. 248-251.
[16]. Achraf BENBA, Abdelilah JILBAB and Ahmed HAMMOUCH. "Voice analysis for detecting persons with Parkinson's disease using MFCC and VQ." *The 2014 International Conference on Circuits, Systems and Signal Processing, 2014.*
[17]. BENBA, ACHRAF, ABDELILAH JILBAB, and AHMED HAMMOUCH. "VOICE ANALYSIS FOR DETECTING PERSONS WITH PARKINSON'S DISEASE USING PLP AND VQ." *Journal of Theoretical & Applied Information Technology*70.3 (2014).
[18]. Achraf BENBA, Abdelilah JILBAB and Ahmed HAMMOUCH " Voiceprint analysis using Perceptual Linear Prediction and Support Vector Machines for detecting persons with Parkinson's disease", *the 3rd International Conference on Health Science and Biomedical Systems*, Florence, Italy, November 22-24 2014.
[19]. Achraf BENBA, Abdelilah JILBAB and Ahmed HAMMOUCH. "Hybridization of best acoustic cues for detecting persons with Parkinson's disease," *2nd World conference on complex system,*Agadir, Morocco, November 10-12 2014.

Achraf BENBA received his Master's degree in Electrical Engineering from "Ecole Normale Supérieure de l'Enseignement Technique" ENSET, Rabat Mohammed V University, Morocco, in 2013 he is a research student of Sciences and Technology of the Engineer in Ecole Nationale Supérieure d'Informatique et d'Analyse des Systèmes ENSIAS, Research Laboratory in Electrical Engineering LRGE, Research Team in Computer and Telecommunication ERIT at ENSET, Mohammed V University, Rabat, Morocco. His interests are in Signal processing for detection neurological disorders.

Abdelilah JILBAB is a teacher at the Ecole Normale Supérieure de l'Enseignement Technique de Rabat, Morocco; He acquired his PhD in Computer and Telecommunication from Mohammed V Agdal University, Rabat, Morocco in February 2009. His thesis is concerned with the Filtering illegal sites on the Internet: Contribution to the type of image recognition based on the Principle of Maximum Entropy. Since 2003 he is a member of the laboratory LRIT (Unit associated with the CNRST, FSR, Mohammed V University, Rabat, Morocco).

Ahmed HAMMOUCH received the master degree and the PhD in Automatic, Electrical, Electronic by the Haute Alsace University of Mulhouse (France) in 1993 and the PhD in Signal and Image Processing by the Mohammed V Agdal University of Rabat (Morocco) in 2004. From 1993 to 2013 he was professor in the Mohammed V-Souissi University in Morocco. Since 2009 he manages the Research Laboratory in Electronic Engineering. He is an author of several papers in international journals and conferences. His domains of interest include multimedia data processing and telecommunications. He is currently head of Department for Scientific and Technical Affairs in National Center for Scientific and Technical Research in Rabat (Morocco).