



Voice analysis for detecting patients with Parkinson's disease using the hybridization of the best acoustic features

Achraf BENBA¹, Abdelilah JILBAB² and Ahmed HAMMOUCH³

^{1,2,3}Laboratoire de Recherche en Génie Electrique, Ecole Normale Supérieure de l'Enseignement Technique, Mohammed V University, Rabat, Morocco
achraf.benba@um5s.net.ma

Abstract: Parkinson's disease (PD) is a neurodegenerative disorder of unknown etiology. It causes, during its course, vocal impairment in approximately 90% of patients. PD patients suffer from hypokinetic dysarthria, which manifests in all aspects of speech production: respiration, phonation, articulation, nasality and prosody. To evaluate these, clinicians have adopted perceptual methods, based on acoustic cues, to distinguish different disease states. In order to improve these evaluations, we used a variety of voice samples comprising the numbers from 1 to 10, four rhymed sentences, nine Turkish words plus the sustained vowels "a", "o", and "u". Samples were collected from 40 people, 20 with PD. We used the method of Leave-One-Subject-Out (LOSO) validation with a K Nearest Neighbor (k-NN) classifier with its different types of kernels, (i.e.; RBF, Linear, polynomial and MLP). The best result obtained was 82.5% using two different voice samples; 1- the 4th acoustic features along with the 17th voice samples; 2- The 3th and the 5th acoustic features along with 20th voice sample.

Keywords: Voice analysis, Parkinson's disease, acoustic features, UCI Machine learning, Leave One Subject Out, Support Vector Machines.

1. Introduction

Parkinson's disease (PD) is the second most common neurological disorder after Alzheimer's disease. It causes, during its course, a variety of symptoms. These include difficulty walking, talking, thinking or completing other simple tasks [1] [2] [3]. Such neurological diseases profoundly affect the lives of patients and their families [1]. PD is generally seen in people over the age of 50. For most elderly people who are suffering from the disease, physical visits to the clinic for diagnosis, monitoring, and treatment are difficult and complicated [4] [5]. Parkinson's disease causes vocal impairment for approximately 90% of patients [6]. Vocal disorders do not appear abruptly. They are the result of a slow process whose early stages may be unnoticed. For this reason, the development of early diagnosis and tele-monitoring systems with accurate, reliable, and unbiased predictive models, are crucial for patients and research [1] [7]. These will allow practitioners and patients to act faster and better understand the disease. There are recent studies for the detection of voice disorders with machine learning tools using acoustic measurements (features) of dysphonia. These include fundamental frequency or pitch of vocal oscillation (F0); absolute sound pressure level (indicating the relative loudness of speech); jitter which represents the cycle-to-cycle variation of fundamental frequency; shimmer which represents the extent of variation in speech amplitude from cycle to cycle; and harmonicity which represents the degree of acoustic periodicity [1] [8] [9]. Studies have shown variations in all these measurements for distinguish person with PD from healthy controls [10], which shows that these acoustic parameters could be useful to evaluate speech disorders [1]. Voice impairments can be detected by using acoustic features extracted from people with PD. Other measurements, such as complex nonlinear aperiodicity, turbulent, aero-acoustic, and non-Gaussian randomness of the sound could be useful to increase the efficacy of voice disorder diagnosis systems [8].

Received: October 29th, 2014. Accepted: February 12nd, 2016

DOI: 10.15676/ijeei.2016.8.1.8

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). Betul Erdogdu Sakar *et al* [5] 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] [5]. Success of the diagnostic system is measured with true positive (TP), true negative (TN), false positive (FP) and false negative (FN) rates.

In this work, we used a dataset that was published in the UCI machine learning archive on June 2104 and which was used by Betul Erdogdu Sakar et al [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 this paper we used the same database as [5]. The main idea of our work is to show the effectiveness of using, not multiple types of voice recording as [5], but each type of voice recording independently in order to improve the classification accuracy. For classification, we used k-NN, and to evaluate the success of the models in discriminating healthy subjects from people with Parkinson’s disease, we calculated accuracy, specificity, sensitivity scores [5].

2. Data acquisition

The database of this study (Table I) was downloaded from the UCI machine learning archive, and was used in [5]. It consists of 20 Parkinsonian patients (6 female, 14 male) and 20 healthy subjects (10 female, 10 male) (Figureure 1) who visited the Department of Neurology in the Cerrahpasa Faculty of Medicine, Istanbul University. The test group consisted of patients suffering from Parkinson’s disease for 0 to 6 years. The age of PD patients varied between 43 and 77 years (mean: 64.86, standard deviation: 8.97). The age of healthy subjects ranged between 45 and 83 years (mean: 62.55, standard deviation: 10.79). For each individual, 26 voice samples including sustained vowels, numbers, words, and short sentences were recorded. The voice samples were selected by a group of neurologists from a set of speaking exercises that aim to improve voice performance. All samples were recorded by a Trust MC-1500 microphone with a frequency range between 50 Hz and 13 kHz. The microphone was set to 96 kHz, 30 dB and placed at a distance of 10 cm from the subject, who then read or repeated the specified words or texts. In the UCI machine learning archive dataset [11], there are 26 voice samples, with multiple types of voice recordings. The dataset is organized in a way that the columns represent the 26 features, and the rows represent the 26 types of voice samples for each individual, for instance the first three samples represent the sustained vowels /a/, /o/ and /u/, respectively. The samples from 4 to 13 represent numbers from 1 to 10. The 14th to the 17th voice sample represent short sentences, while the rest of samples represent individual words. This makes a metrics of 1040x26 (40x26 =1040).

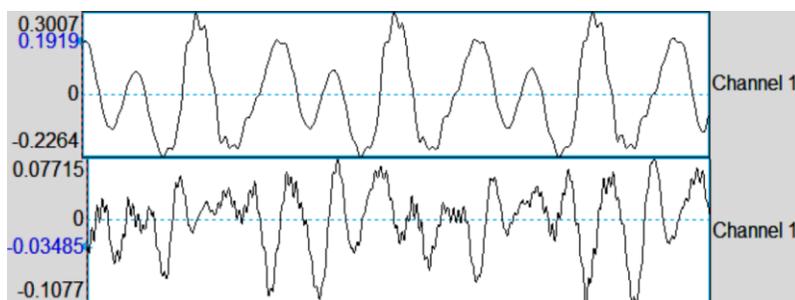


Figure 1. Waveform of a voice sample belonging to a healthy individual (top) and a PWP (bottom).

Table 1. Structure of the database

Database (1040x26)		26 Acoustic features (see Table I) which contains (F0, Jitter, Shimmer, HNR...)					
Subjects	Samples	1	2	3	...	25	26
1	1	-	-	-	-	-	-
	2	-	-	-	-	-	-
	⋮	-	-	-	-	-	-
	26	-	-	-	-	-	-
2	1	-	-	-	-	-	-
	2	-	-	-	-	-	-
	⋮	-	-	-	-	-	-
	26	-	-	-	-	-	-
⋮	1	-	-	-	-	-	-
	2	-	-	-	-	-	-
	⋮	-	-	-	-	-	-
	26	-	-	-	-	-	-
40	1	-	-	-	-	-	-
	2	-	-	-	-	-	-
	⋮	-	-	-	-	-	-
	26	-	-	-	-	-	-

Table 2. Time-Frequency-Based Features given by Pratt acoustic analysis software

Groups	Features	Number of feature
Jitter Parameters	Jitter (local) (%)	1
	Jitter (local, absolute) (s)	2
	Jitter (rap) (%)	3
	Jitter (ppq5) (%)	4
	Jitter (ddp) (%)	5
Shimmer Parameters	Shimmer (local) (%)	6
	Shimmer (local, dB) (dB)	7
	Shimmer (apq3) (%)	8
	Shimmer (apq5) (%)	9
	Shimmer (apq11) (%)	10
	Shimmer (dda) (%)	11
Harmonicity	Mean autocorrelation (AC)	12
	Mean NHR	13
	Mean HNR	14
Pitch Parameters	Median pitch (Hz)	15
	Mean pitch (Hz)	16
	Standard deviation (Hz)	17
	Minimum pitch (Hz)	18
	Maximum pitch (Hz)	19
Pulses Parameters	Number of pulses	20
	Number of periods	21
	Mean period (s)	22
	Standard deviation of period (s)	23
Voicing Parameters	Fraction of locally unvoiced pitch frames (%)	24
	Number of voice breaks	25
	Degree of voice breaks (%)	26

3. Methodology

In their study, Betül Erdogdu Sakar et al [5] used a classification with Leave-One-Subject-Out (LOSO) validation scheme, in which all the 26 voice samples of one individual were left out to be used for validation as if it was an unseen individual, and the rest of the samples are used for training [5]. According to their method, if the majority of the voice samples of a test individual are classified as unhealthy, then the individual is classified as positive [5]. In their study, they presented another classification with Summarized Leave-One-Out (s-LOO). The aim of using this method was to compare the success of conventional Leave-One-Subject-Out validation with an unbiased Leave-One-Out (LOO) [5]. In this method, the feature values of the 26 voice samples of each individual are summarized using central tendency and dispersion metrics such as mean, median, trimmed mean (10% and 25% removed), standard deviation, interquartile range, mean absolute deviation, and a novel form of dataset consisting of N samples is formed where N is the number of individuals [5]. The purpose of summarizing the voice samples of individuals is to minimize the effect of variations between different voice samples of a subject [5]. The best classification accuracy achieved in their research was 77.50% using s-LOO with linear kernel of SVM and the best results of 1000 runs of selecting a random voice samples from each individual was 85% with the same SVM kernel [5] which is not a steady results, because it needs 1000 runs maybe less maybe more and also they used random voice samples and not the same sample for all subjects. The methodology of our method is described in the next paragraphs:

A. Feature Extraction

Dysarthria is the set of voice illnesses related with turbulences of muscular control of the speech organs. Dysarthria includes all malfunctions related to breathing, phonation, articulation, nasalization and prosody. These deficits can be measured and detected by analyzing various features of voice. In this study each subject was asked to read or say predetermined 26 items comprising numbers from 1 to 10, four rhymed sentences, nine words in Turkish language plus sustained vowels “a”, “o”, and “u”. A total of 1040 recordings each represented with by a 26 dimensional feature vector was calculated (Table II) along with a binary PD-score. A PD-score of zero indicates that the feature vector belongs to a person with PD and a score of one indicates that it belongs to a healthy subject. To extract features from voice samples, Praat acoustic analysis software [12] was used. A group of 26 linear and time-frequency based features were extracted from each voice sample.

1. Fundamental Frequency

The fundamental frequency is the number of opening and closing cycles of the vocal folds per second.

2. Voice breaks

Normal voices can easily maintain phonation when saying sustained vowel /a/. Some pathological voices have difficulty with it. This can be measured in different ways [12].

Fraction of locally unvoiced pitch frames (FLUPF) represents the fraction of pitch frames that are considered as unvoiced (the MDVP calls it DUV) [12].

Number of voice breaks is defined as the number of times distances between consecutive pulses that are longer than 1.25 divided by the pitch floor [12].

Degree of voice breaks is computed as the total duration of the breaks between the voiced parts of the signal, divided by the total duration of the analyzed part of the signal (the MDVP calls it DVB). [12].

3. Jitter measurement

Jitter (absolute) represents the cycle-to-cycle variation of fundamental frequency. It is computed as the average absolute difference between consecutive periods, expressed as [15] [16]:

$$Jitter (absolute) = \frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i+1}| \quad (1)$$

Where T_i is the extracted F_0 (fundamental frequency) period lengths and N is the number of extracted F_0 periods.

Jitter (relative) is defined as the average absolute difference between consecutive periods, divided by the average period. It is expressed as a percentage [15] [16]:

$$Jitter (relative) = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i+1}|}{\frac{1}{N} \sum_{i=1}^{N-1} T_i} \quad (2)$$

Jitter (rap) represents the Relative Average Perturbation, computed as the average absolute difference between a period and the average of it and its two neighbors, divided by the average period [15] [16].

Jitter (ppq5) represents the five-point Period Perturbation Quotient, defined as the average absolute difference between a period and the average of it and its four closest neighbors, divided by the average period [15] [16].

4. Shimmer measurement

Shimmer (dB) represents the variability of the peak-to-peak amplitude in decibels, computed as the average absolute base-10 logarithm of the difference between the amplitudes of consecutive periods, multiplied by 20 [15] [16]:

$$Shimmer(dB) = \frac{1}{N-1} \sum_{i=1}^{N-1} \left| 20 \log \left(\frac{A_{i+1}}{A_i} \right) \right| \quad (3)$$

where A_i is the extracted peak-to-peak amplitude data, and N is the number of extracted fundamental frequency periods F_0 .

Shimmer (relative) is expressed as the average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude, expressed as a percentage [15] [16]:

$$Shimmer(relative) = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A_i - A_{i+1}|}{\frac{1}{N} \sum_{i=1}^N A_i} \quad (4)$$

Shimmer (apq11) represents the 11-point Amplitude Perturbation Quotient, defined as the average absolute difference between the amplitude of a period, and the average of the amplitudes of it, and its ten closest neighbors, divided by the average amplitude [15].

5. Harmonicity

Harmonics-to-Noise Ratio (HNR) is expressed as the degree of acoustic periodicity [12].

B. Classification with Leave-One-Subject-Out (LOSO)

Instead of using conventional bootstrapping or LOO validation methods [13] [14] (which consist of sparing some samples of an individual in the training phase and some for the testing phase, creating an artificial overlap between the training and test sets) or s-LOO (which summarizes the feature values of N voice samples of each subject by using central tendency and dispersion metrics [5]), we use a LOSO validation scheme. That is, we left out all the samples of one individual to be used for validation as if it were an unseen individual, and trained a classifier on the rest of the samples. In addition, we compared two methods of using LOSO validation: first with all the 26 voice samples of each individual; then, using LOSO for each voice sample independently.

Evaluation Metrics

In order to measure the success of our classifiers and select the best acoustic features, we used the evaluation metrics accuracy, sensitivity and specificity. Accuracy is the ratio of correctly classified instances to total instances [5]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

Where TP is true positive (Healthy subjects who was correctly classified), TN is true negative (Patients with PD who was correctly classified), FP is false positive (Patients with PD who was incorrectly classified) and FN is false negative (Healthy subjects who were incorrectly classified). Accuracy represents the success of the classifier to discriminate between the two groups, Sensitivity represents the accuracy of detecting Healthy subjects and Specificity represents the accuracy of detecting the other patients with PD [5] [17] [18] [19] [20]:

$$Sensitivity = \frac{TP}{TP + FN} \quad (6)$$

$$Specificity = \frac{TN}{TN + FP} \quad (7)$$

4. Experimental result

In this study, we replicated the work in [5]. Firstly, we used LOSO method with all the 26 voice samples at the same time of each individual as can be seen in Table I. These 26 voice samples contain sustained vowels, numbers, words, and short sentences. For example; in the first step of LOSO method, we left out the 26 voice samples of the first subject to be used for validation, and we trained our classifier on the other voice samples of the other 39 subjects (all subjects except the first one). And then we decide if it is TP, TN, FP or FN. In the second step, we left out the 26 voice samples of the second subject to be used for validation, and we trained our classifier on the other voice samples of the other 39 subjects (all subjects except the second one) then we decide if it is TP, TN, FP or FN. This method is repeated for all until the 40th subject. Then we calculated the accuracy sensitivity and specificity. The best obtained classification accuracy was 48.077% using k-NN with k=3 as shown in the Table III. This is somewhat lower than that what was obtained in [5] using the entire 26 voice sample at the same time. By using s-LOO, they got 85% as maximum classification accuracy after 1000 runs which is not a steady result, because it needs 1000 runs maybe less maybe more and also they used random voice samples and not the same sample for all subjects. In order to build a real system to diagnosis the disease we should use the same recordings for all subjects and not different ones. The best obtained results using same recording was 77.50% using s-LOO with linear kernel of SVM and 82.14% using s-LOO with RBF kernel of SVM for the test dataset which is not use in this study.

Table 3. Classification results using LOSO with all 26 voice samples

k-NN	Accuracy %	Sensitivity %	Specificity %
1	47.1154	49.4231	44.8077
3	48.0769	50.3846	45.7692
5	43.9423	46.1538	41.7308
7	43.75	46.1538	41.3462

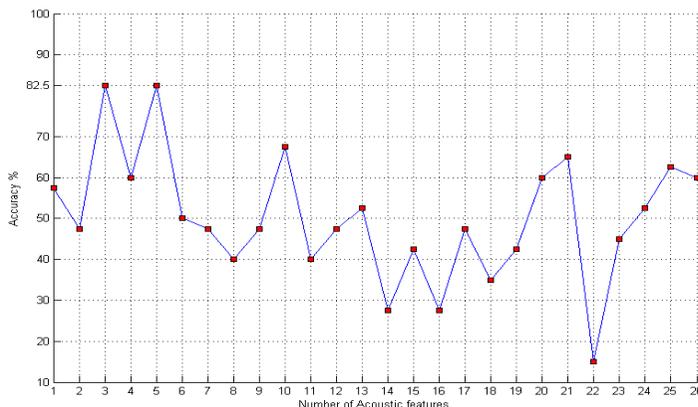


Figure 2. The classification accuracies using Leave-One-Subject-Out (LOSO) validation scheme for the 20th voice sample. The best results was achieved for the 3th and the 5th voice samples (which correspond to Jitter (rap) and Jitter (ddp) successively) using LOSO with Linear k-NN (k = 3)

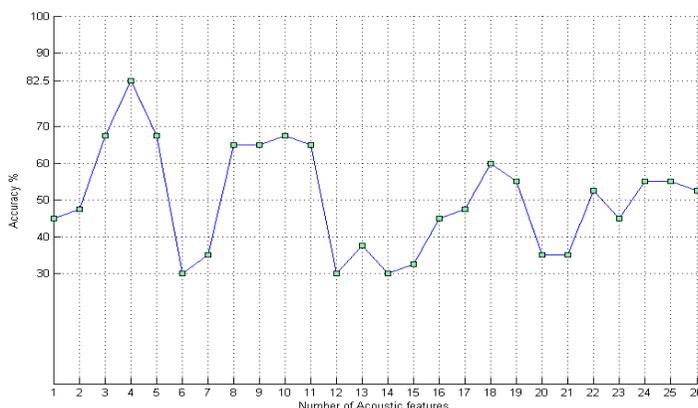


Figure 3. The classification accuracies using Leave-One-Subject-Out (LOSO) validation scheme for the 17th voice sample. The best results was achieved for the 4th (which correspond to Jitter (ppq5)) using LOSO with Linear k-NN (k = 3)

In our study, instead of using all the 26 voice samples at the same time, we used LOSO for every voice sample independently. That is we used LOSO for the first voice sample of each subject and we calculated the accuracy, and then for the second voice sample and so on until the 26th voice sample. As can be seen from Figure 2 and Figure 3 and Table II, the best result obtained was 82.5% using two different voice samples; 1- the 4th acoustic features (Jitter (ppq5)) along with the 17th voice samples (short sentence); 2- The 3th (Jitter (rap)) and the 5th (Jitter (ddp)) acoustic features along with 20th voice sample (word). These samples correspond

to “sentences” and “words”. From these results, it is clear that Jitter measurements and the running speech contain more information about the state of a subject as pathological or healthy.

5. Conclusion

The characteristic disorders of Parkinson’s disease are the result of a slow process whose early stages may go unnoticed. For this reason, we used a variety of voice samples per subject with multiple types of voice recordings, in order to develop method for early diagnosis and to build predictive tele-diagnosis and tele-monitoring models.

The representations of acoustic signals have allowed us to examine the effect of Parkinsonism on the phonological system. However, during the various stages of processing, the results show that some voice samples are not reliable indicator of the state of a subject. Indeed, for Parkinson’s patients some parameters do not confirm the presence of the disease and do not distinguish them from healthy subjects.

The purpose of this study is to show the effectiveness of using each type of voice recording and each acoustic feature independently. The best result obtained was 82.5% using running speech samples along with jitter measurement by k-NN and LOSO validation scheme with k=3.

6. Acknowledgment

The authors would like to thank Thomas R. Przybeck and Daniel Wood, United States Peace Corps Volunteers (Morocco 2013-2015), and all of the participants involved in this study.

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]. Huse, Daniel M., et al. "Burden of illness in Parkinson's disease." *Movement disorders* 20.11 (2005): 1449-1454.
- [5]. 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.
- [6]. S. B. O’Sullivan and T. J. Schmitz, “Parkinson disease,” in *Physical Rehabilitation*, 5th ed. Philadelphia, PA, USA: F. A. Davis Company, 2007, pp. 856–894.2007, pp. 856–894.
- [7]. Ruggiero, C., R. Sacile, and M. Giacomini. "Home telecare." *Journal of Telemedicine and Telecare* 5.1 (1999): 11-17.
- [8]. Little, Max A., et al. "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection." *BioMedical Engineering OnLine* 6.1 (2007): 23.
- [9]. Rahn III, Douglas A., et al. "Phonatory impairment in Parkinson's disease: evidence from nonlinear dynamic analysis and perturbation analysis." *Journal of Voice* 21.1 (2007): 64-71.
- [10]. Zwirner, Petra, Thomas Murry, and Gayle E. Woodson. "Phonatory function of neurologically impaired patients." *Journal of communication disorders* 24.4 (1991): 287-300.
- [11]. UCI Machine Learning Repository, Parkinson speech dataset with multiple types of sound recording data set. Available : [http://archive.ics.uci.edu/ml/datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings]
- [12]. Boersma, Paul. "Praat, a system for doing phonetics by computer." *Glott international* 5.9/10 (2002): 341-345.

- [13]. Efron, Bradley. "Bootstrap methods: another look at the jackknife." *The annals of Statistics* (1979): 1-26.
- [14]. Reunanen, Juha. "Overfitting in making comparisons between variable selection methods." *The Journal of Machine Learning Research* 3 (2003): 1371-1382.
- [15]. Farrús, Mireia, Javier Hernando, and Pascual Ejarque. "Jitter and shimmer measurements for speaker recognition." *INTERSPEECH*. 2007.
- [16]. Shirvan, R. Arefi, and E. Tahami. "Voice analysis for detecting Parkinson's disease using genetic algorithm and KNN classification method." *Biomedical Engineering (ICBME), 2011 18th Iranian Conference of. IEEE*, 2011.
- [17]. BENBA, Achraf, 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.
- [18]. Benba, Achraf, Abdelilah Jilbab, and Ahmed Hammouch. "Hybridization of best acoustic cues for detecting persons with Parkinson's disease." *Complex Systems (WCCS), 2014 Second World Conference on. IEEE*, 2014.
- [19]. Benba, Achraf, 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 (HSBS'14), November 22-24 2014, Florence, Italie*.
- [20]. 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).



BENBA Achraf, received the Master 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).