A Hybrid Method for the Diagnosis and Classifying Parkinson’s Patients based on Time–frequency Domain Properties and K-nearest Neighbor

Abstract
The vibrations of hands and arms are the main symptoms of Parkinson’s ailment. Nevertheless, the affection of the vocal cords leads to troubles and defects in the speech, which is another obvious symptom of the disease. This article presents a diagnostic model of Parkinson’s disease (PD) and proposes the time–frequency transform (wavelet WT) and Mel-frequency cepstral coefficients (MFCC) treatment for this disease. The proposed treatment is centered on the vocal signal transformation by a method based on the WT and to extract the coefficients of the MFCC and eventually the categorization of the sick and healthy patients by the use of the classifier K-nearest neighbor (KNN). The analysis used in this article uses a database that contains 18 healthy patients and twenty patients. The Daubechies mother WT is used in treatments to compress the vocal signal and extract the MFCC cepstral coefficients. As far as, the diagnosis of Parkinson’s ailment is concerned the KNN classifying performance gives 89% accuracy when applied to 52% of the database as training data, whereas when we increase this percentage from 52% to 73%, we reach 98.68% accuracy which is higher than using the support-vector machine classifier. The KNN is conclusive in the determination of the PD. Moreover, the higher the training data is, the more precise the results are.

Keywords: K-nearest neighbor, Mel-frequency cepstral coefficient, Parkinson’s disease, wavelet

Introduction
In 1817, James Parkinson described Parkinson’s ailment,[1] which is a neurodegenerative disease of unknown cause, characterized by the progressive destruction of a specific population of neurons.

The loss of dopamine in the midbrain induces the slowness of movement, difficulty with walking, communication trembling, and rigidity which are the most obvious motor symptoms.[2]

The performing neuron analyses and magnetic resonance imaging examination of the brain are employed in the Parkinson’s disease (PD) detection. The phonation and articulation means of speech extraction and analysis can give the needed guidance in the spotting of PD.

The PD has a lot of indicators; among them, the vocal impairment which is one of the earliest.[3] Exactly, the phonation is the main part of speech production affected.[4]

Several methods are used for the diagnosis of PD (prediction cepstral coefficient,[2] perceptual linear predictive [PLP],[3] and Mel-frequency cepstral coefficient [MFCC][3,5]). Focusing on the vocal signal, we are interested in the most used in recognition systems which are the MFCC method. To exploit the human auditory system characteristics through the change of frequencies linear scale into Mel scale that allows to make cepstral analysis by passage in the log-spectral domain,[4] the cepstral analysis had been used by Shourie[6] of the electroencephalogram signals in the process of perception observance, and mental imagery proves the impact of artistic expertise and also in the appraising of hypernasality for children affected by cleft palate centered on cepstrum analysis by Akafi et al.[7]

The diagnostic of patients affected by PD who undergo a categorization process for appraising and home monitoring of tremor in those patients elaborated by Bazgir

Zayrit Soumaya1, Belhoussine Drissi Taoufiq1, Nsiri Benayad2, Bena Achraf3, Abdelkrim Ammoumou1
1Laboratory Industrial Engineering, Information Processing and Logistics (GITIL), Faculty of Science Ain Chok, University Hassan II - Casablanca, 2Laboratory Research Center STIS, MZCS, Higher School of Technical Education of Rabat (ENSET), 3Electronic Systems Sensors and Nanobiotechnologies (E2SN), ENSET, Mohammed V University in Rabat, Morocco

Address for correspondence:
Dr. Zayrit Soumaya,
Faculty of Science,
Ain Chok University,
Hassan II – Casablanca,
Morocco.
E-mail: zayritsoumaya@gmail.com

Access this article online
Website: www.jmssjournal.net
DOI: 10.4103/jmss.JMSS_61_18

et al.\cite{1} are also some other ways to reach the diagnosis of the disease, for example, handwriting.

Filter banks that are devised accordingly to the perceptual criteria of the human ear will be required for the computation of MFCC and PLP features. It is important that this spectrum obtained by computing the discrete Fourier transform (DFT) of the windowed speech frames should be estimated accurately. The procedure involves the estimation of a short-time spectrum.

Whereas, our interest is focused on the diagnosis of PD from vocal disorders detection. The spotting of PD centered on the extraction of the cepstral coefficients of the MFCC from the speech was first proposed by Frail et al.\cite{2,3} Shahbakhi et al. diagnosed PD\cite{4} by the measures of fundamental frequency disturbance (Jitter), amplitude disturbance (Shimmer), and fundamental frequency disturbance (Jitter), amplitude disturbance (Shimmer), and fundamental frequency (\(F_0\)).

Recent researches using the cepstral coefficients of the MFCC\cite{5,6} and PLP\cite{7} performed by Banba et al. and Upadhya et al.\cite{8} also conducted a study on the spotting of PD by extracting MFCC and PLP by the use of the Thomson multitaper window technique. Recent studies are based on the works performed by Taoufiq BELHOUSSEINE DRISSI et al. that deal with wavelet transform and MFCC and the support-vector machine (SVM) classifier.\cite{9,10} In this work, they transform the speech signals through the sorts of DWT which were tested; then, they extracted MFCC coefficients from the signals and applying the SVM as a classifier.

Among the simplest machine learning algorithms, we find the K-nearest neighbor (KNN) algorithm, which is a robust classification method. This method is widely applied in real-time applications. The SVM principal based on the use of hyper planes to separate the classes. The shape of the decision will change if a different kernel provides, so choosing the kernel is necessary. The choice of good kernel needs to have some knowledge about the data that is not always available. Besides that, more the size of the dataset used for training is big, more the computational time for training grows nonlinearly with it. Whereas, the KNN being based on vector distance concept, so errors are bound to be less. Hence, in this article, we will choose the KNN as classifier in the aim of having higher accuracy.

In this work, we come up with a diagnosis model of the PD based on a time–frequency treatments of speech signals of a database\cite{11} that consists of 18 sound patients and twenty affected by “PD,” then extracting the cepstral coefficient of the MFCC, and in the end, a classification will be performed by the KNN classifier. We will create two training bases when the first accounts for 52% of the database and the other 73% and apply the suggested treatment (wavelet, MFCC, and KNN) on the totality of the database.

### Continuous time–frequency transform

The continuous time–frequency transform (CWT) was devised by the French geophysicist Morlet in 1980 to study earth tremor signals.\cite{12} Then, Grossmann, Meyer, Mallat, and Daubechies laid their mathematical basis for wavelets.\cite{16} Since that time, WT is more and more used in signal processing.

A wavelet uses two coefficients: a coefficient of scale “\(a\)” which permitted to obtain various versions, which were compressed or dilated of windows stemming from the same mother wavelet, this coefficient represents the inverse of the frequency and a coefficient of translation “\(b\)” that characterizes the displacement of the window along the axis of time.

The CWT of signal \(s(t)\) is defined by:\cite{17}

\[
W_s(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(t) \psi^*(\frac{t - b}{a}) dt
\]  

(1)

here \(\psi(t)\) is the mother wavelet, and \(\psi^*(t)\) is the conjugate complex \(\psi(t)\).

It should be noted that the wavelet transform gives adequate temporal resolution at high frequencies and adequate frequency resolution at low ones.

### Discrete wavelet transform

The discrete wavelet transform (DWT) is the discrete version of the continuous time–frequency transform (CWT). It is achieved by the use of the Mallat algorithm\cite{18} that is regarded as a multiresolution analysis. This algorithm is based on the definition of a pair of filters \(H\) (low-pass filter) and \(G\) (high-pass filter) and whose impulse responses \(h\) and \(g\). Several sorts of wavelets are used in literature: Haar, Beylkin, Coiflet, Daubechies, Symmlet, Vaidyanathan, Battle, etc.

In this work, we will only use the wavelets of Daubechies.

### Mel-frequency Cepstral Coefficient

MFCCs refer to the parameters that used the most in speech recognition systems. MFCC analysis consists of the adaption of the linear scale of frequencies into the Mel scale to exploit the properties of the human auditory system\cite{19} that give the most effective illustration of the speech signal. The process of extracting the coefficients is shown in Figure 1.\cite{13,20}

**Preemphasis**

This is a voice signal filtering process \((s_n, n = 1,\ldots, N)\) with a first-order finite impulse response numeric filter \(s_n\) given as follows:\cite{19,21}

\[
H(z) = 1 - k z^{-1}
\]

(2)

Where, \(k\) is the coefficient of the preemphasis that must be comprised between 0.9\(\leq k \leq 1\). In this study, we fixed the
In this way, the pre-emphasized signal is related to the signal \( s_n \) by the formula below:

\[
s'_n = s_n - ks_{n-1}
\]  (3)

**Segmentation**

The vocal signal is nonstationary, but the signal processing ways are stationary signals. To solve this problem, we will proceed to the segmentation of the signal into \( N \) speech samples of frames in the lapse of 10–30 ms where the voice signal is regarded as stable. To dodge unexpected transitions from frame to frame, the two adjacent frames are overlapped.\(^{[3]}\)

**Windowing**

As a result of the segmentation, some discontinuities are shown at the borders of the frames; in the aim of reducing the revealed discontinuities, we multiply the samples \( s'_n, n = 1, \ldots, N \) of the frame by a Hamming window.\(^{[20,22]}\)

\[
s''_n = \left(0.54 - 0.46 \cdot \cos \left(\frac{2\pi (n - 1)}{N - 1}\right)\right) \cdot s'_n
\]  (4)

Where, \( N \) is the number of samples in the frame.

**The fast Fourier transform**

The fast Fourier transform (FFT) application consists of converting every single frame of \( N \) samples to the frequency domain instead of the time domain. The FFT is a fast algorithm for implementing DFT.

The definition of the DFT is as follows:\(^{[13,22]}\)

\[
S_n = \sum_{k=0}^{N-1} s_k e^{-j2\pi ka/N}
\]  (5)

**Mel filtering with a filter bank**

The human ear follows a nonlinear scale through an audible spectrum.\(^{[20]}\) Consequently, we will use the transformation of the linear scale of frequencies to the Mel scale. The latter is in a linear space under 1000 Hz (low frequencies) and logarithmic above 1000 Hz (high frequencies).

The conversion from the linear scale to the Mel scale\(^{[13,20,22]}\) is given as follows:

\[
\text{Mel}(f) = 2595 \cdot \log_{10} \left(1 + \frac{f}{700}\right)
\]  (6)

**Logarithm/discrete cosine transform**

The MFCC coefficients may be worked out first hand by employing the discrete cosine transform (DCT) on the logarithm of energies coming out of a bank of \( M \) triangular filters, apart from according to the Mel scale\(^{[13]}\) by the following equation.

\[
c_i = \sum_{j=1}^{M} m_j \cos \left(\frac{\pi j (i - 0.5)}{M}\right)
\]  (7)

Where, \( m_j \) is the logarithm of the energy obtained with the triangular filter \( j \), \( M \) is the number of filters bank, in our article \( M \) was set to 20,\(^{[13]}\) and \( i \) is the number of coefficients to be extracted.

**Lifting**

As the higher order of the MFCC coefficients is so small, we have to apply the lifter to lift the cepstrum. Consequently, it is important to increase these amplitudes so that they become quite similar.\(^{[20,22]}\) To achieve that, we liftered the cepstral coefficient so that the following equation can be applied:

\[
c'_i = \left(1 + \frac{L}{2} \sin \left(\frac{\pi i}{L}\right)\right) c_i
\]  (8)

Here, \( L \) is the parameter of the lifter. In this article, we set \( L = 22\).\(^{[13]}\)

**K-nearest Neighbor**

KNN classifier is of a simple principle based on the theory of statistical training. First, we give a database that contains the two classes with a label vector is the training phase where the feature space is reached so that the database become separable. At the test phase, the database classified seeks the nearest neighbor given by training database, and according to this, it is classified either in class 1 or class 2. The Euclidean distance was applied to spot the nearest neighbor in the KNN algorithm.\(^{[23]}\)

Between the two points \( x \) and \( y \), we calculate the Euclidean distance \( d(x, y) \) using Eq. 9. Here, \( N \) is the number of characteristics such that \( x = (x_1, x_2, x_3, \ldots, x_N) \) and \( y = (y_1, y_2, y_3, \ldots, y_N) \).

\[
d(x, y) = \sum_{i=1}^{N} \sqrt{x_i^2 - y_i^2}
\]  (9)

**Results**

The goal of this study is to determine the KNN performance. Before the bloc of extraction of the MFCC coefficients, DWT block will be injected to achieve a
correct diagnosis of PD as shown in Figure 2.

We apply the database\(^{[14]}\) that consists of 18 sound ones and twenty recordings of patients suffering from PD. They all utter the vowel “a.”

The algorithm of DWT is centered on the definition of a pair of filters H (low-pass filter) and G (high-pass filter). The filter outputs are subsampled by a factor of 2. The high-pass filter provides DWT coefficients or signal details at a given scale. The low-pass filter gives the coefficients of the approximation of the signal at the same scale. The same operation is again applied to the approximation, thus generating another detail and a new approximation.\(^{[24]}\)

A process of PD diagnosis which is similar to our process is applied in the article,\(^{[13]}\) the difference between them is the classifier. The DWT gives the higher accuracy at level 2 and the 3\(^{rd}\) scale. Hence, in our study, we will work with the Daubechies db2 wavelet at scale 3, and we are interested only in the approximation \(a_3\) [Figure 3].

In the first phase, we transform the vocal recordings by the use of Daubechies wavelet. The vocal signal of PD patient before and after using the Daubechies wavelet is shown in Figure 4. Figure 5 shows a zoom at the two representations of the signal.

In the second phase, we will execute an input of \(a_3\) approximation to the MFCC block to obtain from every single patient, the first 12 MFCC coefficients employing the program “Htk mfc matlb.”\(^{[25]}\) These coefficients will be the characteristics that be relied on to get a classification to reach an exact diagnosis. The MFCC is composed of numerous frames that need significant processing time to classify. However, such operation hinders a precise result.\(^{[19]}\) To cope with this problem, we had recourse to the calculation of the average value of these images to obtain the voiceprint. The 12 MFCC and voiceprint for a sound patient are featured in Figure 6, as for Figure 7, it illustrates the MFCC and voiceprint for a patient suffered from PD.

In the third phase, in which we take a decision based on the categorization of the patients. In this aims, we create two
training bases one of 52% and the other 73% of the database. At the first step, we will carry out a test on our database by using the first training base (of 52%) and another test during the use of the second training base (of 73%) classifier.

In a categorization problem, the labels are part of the following possible identities:

Moreover, the task consists of assigning a test example to one of the C classes. KNN classifier is the most used procedure. Moreover, the widely used method is setting $K = 1$ yields the nearest neighbor classification rule.

In spite of it is simplicity, KNN so often gives a good performance mainly for large data sets.

We calculated measures such as accuracy, sensitivity, and specificity by applying the following formula to determine the performance of the classifier:[13,26,27]

$$\text{Accuracy} = \frac{TN + TP}{TP + TN + FP + FN} \quad (10)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (11)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (12)$$

With:
- TP stands for true positive (correctly classified healthy patients)
• TN stands for true negative (correctly classified patients)
• FP stands for false positive (incorrectly classified patients)
• FN stands for false negative (incorrectly classified sound patients).

The percentage of the test is accuracy, sensitivity as well as specificity of all the recordings by the use of the 52% training is shown in Table 1, and then their percentage by using the training base of 73% in the test of all the recordings (including 6 sick and 4 healthy patients) is shown in Table 2.

**Conclusion**

We have presented in this article, a sample of diagnosis based on PD that is centered on the signal treatment, in which we will employ the wavelet transform and the MFCC using a database of recordings of sick patients and healthy ones while they pronounce the vowel “a.” The change of speech signals is treated by Daubechies wavelet by the third-scale approximation and then, we will recover the 12 cepstral coefficients after injecting the approximation into the MFCC bloc. To make a decision on which one is sick or healthy, we work with the KNN classifier by using two-learning bases, one is 52% and the other 73% of the database. When you work with the database of 52%, one obtains an accuracy of 89% which is higher than the accuracy obtained by using the classifier SVM with the database of 73%, and when we increase the percentage of the database to 73%, we get an accuracy of 98.68% and from that one can conclude that the increase of the base of data gives us better results by increasing the accuracy of the classifier and that the KNN is more accurate than the SVM classifier.

**Financial support and sponsorship**

None.

**References**

Soumaya, et al.: Parkinson diagnosis by using DWT, MFCC and SVM


