Speech analysis for the detection of Parkinson's disease by combined use of empirical mode decomposition, Mel frequency cepstral coefficients, and the K-nearest neighbor classifier

N. BOUALOULOU[1,2,*], B. NSIRI[2], T. BELHOUSSINE DRISSI[1] and S. ZAYRIT[1,2]
[1]Laboratory Electrical and Industrial Engineering, Information Processing, Informatics, and Logistics (GEITIIL). Faculty of Science Ain Chock. University Hassan II, Casablanca, Morocco
[2]Research Center STIS, M2CS, National Higher School of Arts and Craft, Rabat (ENSAM). Mohammed V University in Rabat, Morocco
[*]boualoulounouha@gmail.com

***Abstract*—Parkinson's disease (PD) is one of the neurodegenerative diseases. The neuronal loss caused by this disease leads to symptoms such as lack of initiative, depressive states, psychological disorders, and impairment of cognitive functions as well as voice dysfunctions. This paper aims to propose a system of automatic recognition of Parkinson's disease by voice analysis. In this system, we are based on a database of 38 recordings, 20 people with Parkinson's disease and 18 healthy people pronounce the vowel /a/.at first, we have decomposed the vocal signal of each patient by the Empirical Mode Decomposition (EMD), then, we extract from 1 to 12 coefficients of the Mel Frequency Cepstral Coefficients (MFCC), to obtain the voiceprint from each voice sample, we compressed the frames by computing their average value. At the end of the classification, we have used the validation scheme "holdout" as well as the K-nearest neighbor (KNN) classifier, the performance of this classification gives accuracy up to 86,67% when applied to 80% of the database as training data.**

Keywords—Parkinson's disease; MFCC; EMD; KNN

I. INTRODUCTION

Parkinson's disease (PD) is a neurodegenerative disorder characterized by the overexpression of a protein called α-synuclein.

Indeed, the diagnosis of this disease is made according to the observation of characteristic symptoms, according to James Parkinson the first visible symptoms are generally resting tremors, rigidity, and bradykinesia [1], there are other clinical signs just as disabling as those we have just mentioned. The non-motor clinical signs are sensory disorders, sleep and vigilance disorders, as well as the motor clinical signs which are speech disorders, in particular dysarthria, which is a difficulty in articulating words, and dysphonia, which corresponds to a decrease in the intensity of the voice, the expression of these symptoms will make it possible to diagnose the disease and to institute a treatment that will conceal them for several years.

Diagnosis of PD from voice signals has been studied by many researchers [2,5]; the most widely explored characteristics to detect PD are phonation features, which include jitter variants, shimmer variants, noise variants [6,8]. The Mel Frequency Cepstral Coefficients (MFCC), Perceptual Linear Prediction (PLP), Linear Predictive Coefficients (LPC), and Linear Predictive Cepstral Coefficients (LPCC) are some of the other features studied by researchers [9,10]. One study used for the detection of PD on a speech database containing 17 patients with Parkinson's and 17 healthy patients are treated by extracting the first 12 cepstral MFCC coefficients as well as the support vector machine (SVM) classifier [11], the same authors but this time use the 11 cepstral PLP coefficients on a database of 50 speech samples [12], Another study based their diagnosis of PD by the joint use of the wavelet transform and the MFCC which allows the extraction of cepstral features, as well as the support vector machine (SVM) classifier, applied on a database of 18 healthy patients and 20 patients with Parkinson's disease [13].

The initial procedure for obtaining the MFCC coefficients begins with pre-emphasis and segmentation of each voice frame, then a Hamming window is applied to each frame to make the ends of the time frame close to zero, after the calculation of the Fast Fourier Transform to convert these frames from the time domain to frequency domain., thus MFCCs are a representation defined as the discrete cosine transform of the logarithm of the speech segment energy spectrum. The spectral energy is computed by applying a bank of uniformly spaced filters on a modified frequency scale, called the Mel scale. The Mel scale redistributes the frequencies according to a non-linear scale that simulates human perception of sounds.

Several researchers used acoustic analysis to extract time-frequency characteristics, such as jitter, shimmer, pitch, harmonicity, etc. [14,16]. In our work, we focus on the evaluation of voice disorders in the cepstral domain by using Mel Frequency Cepstral Coefficients (MFCC), which have been used for the first time by [17], as well as its uses in various applications, this MFCC is used for the recognition of Alzheimer's disease [18,19], and they have been used extensively in language and speaker identifications [20,23], they have also used for emotion recognition [24,26]. In this study, we sought to distinguish two categories of patients; 20 patients with Parkinson's disease and 18 healthy patients; each person was

asked to pronounce the sustained vowel /a/. Firstly, using the EMD method, we adaptively decomposed the speech signal into the sum of the oscillating components that are the Intrinsic Mode Function (IMF), secondly, we extracted the first 12 MFCC coefficients from each IMF, and then we used the "Holdout" method with K-nearest neighbor (KNN) for classification to distinguish between healthy people and people with Parkinson's disease.

This paper is organized as follows: the database topics are detailed in Section II; Sections III, IV, and V give a summary explanation of the EMD method and the MFCC process, as well as the KNN classifier used in this study. The methodology and results are presented in section VI, and finally a conclusion in section VII.

II. DATA ACQUISITION

This paper uses a database gathered in [27], which contains 38 voice recordings including 20 PD patients with PD and 18 healthy people. These records were made via a standard microphone at a sampling frequency of 44100 Hz using a 16-bit sound card on a desktop computer.

This microphone was placed at a distance of 15 m from the person pronouncing the vowel /a/. All voice recordings were made in stereo mode and saved in WAV format.

III. EMPIRICAL MODE DECOMPOSITION (EMD)

Empirical Mode Decomposition (EMD) is a self-adaptive signal processing method, which has been applied to non-stationary and non-linear signals. It was first proposed in 1998 by [28]. who provided a powerful tool for multi-scale adaptive analysis of non-stationary signals. This tool decomposes the signals into components by satisfying the following two conditions.1) The number of extrema (maximum + minima) in the signal and the number of zero crossings do not differ by more than one; 2) The average of the envelopes defined by the local maxima and the local minima must be zero at any point. The component functions that have verified these two conditions are called Intrinsic Mode Function (IMF).

The original signal x(t) is decomposed using the EMD method according to the following equation:

$$x(t) = \sum_{i}^{n} imf_i(t) + r_n(t) \quad (1)$$

In which $imf_i(t)$ are intrinsic mode functions and $r_n(t)$ is the residue.

The EMD has the following proprieties:

- The objective of EMD is to decompose the multi-component signal into some single component signals. Thus, each IMF obtained by EMD has only one frequency at a time.
- The EMD method decomposes the signal into the sum of multiple IMFs respecting integrity and orthogonality.
- EMD decomposes the signal into a function of the signal itself and the number of IMFs is finite.

IV. K-NEAREST NEIGHBOR (KNN)

The K-nearest neighbor (KNN) method is a supervised classification method offering very interesting performances in the search for new biomarkers for diagnosis.

The principle of this classification algorithm is very simple. It is provided with a training data set D, a Euclidean distance d as shown in equation (2), and an integer k. For each new test point x for which it must make a decision, the algorithm searches in D for the k points closest to x in the sense of the Euclidian distance d and assigns x to the class that is the most frequent among these k neighbors.

The Euclidean distance between points X1 and X2 is obtained using the following equation with X1=(x11, x12… x1n), X2=(x21, x22… x2n):

$$d(X1,X2) = \sqrt{\sum_{i=1}^{n} (x_{i1} - x_{i2})^2} \quad (2)$$

V. MEL FREQUENCY CEPSTRAL COEFFICIENTS (MFCC)

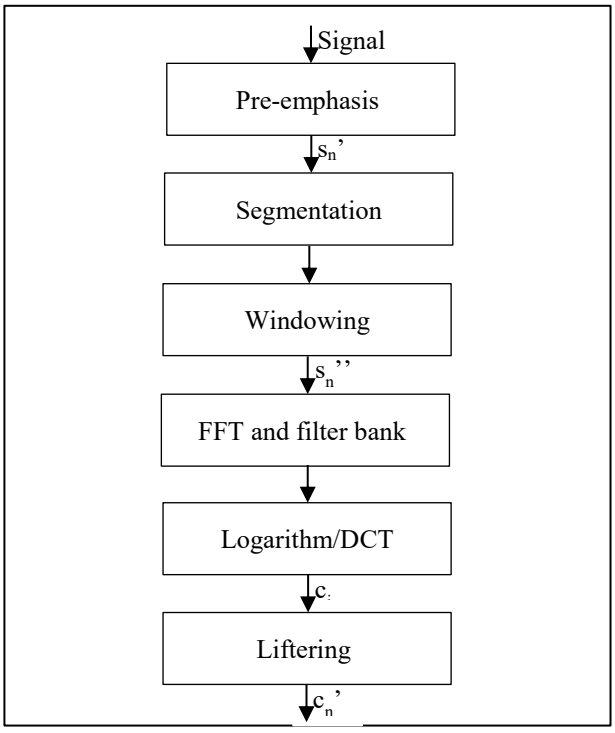

FIGURE I. SCHEMATIC DIAGRAM OF THE MEL FREQUENCY CEPSTRAL COEFFICIENTS

The vocal signal carries several pieces of information such as the linguistic message, the identity of the speaker as well as his emotions and the adopted language, etc. Its great redundancy and variability do not allow it to be directly exploited in its initial state. It is therefore essential to convert this signal into acoustic parameters that are dependent on linguistic information.

In our study, we will work in the cepstral domain by computing the Mel Frequency Cepstral Coefficients (MFCC). These coefficients are the most commonly used parameters in speech recognition systems. The MFCC analysis consists in exploiting the properties of the human auditory system by transforming the linear frequency scale into a Mel scale. In the following, we will describe the MFCC calculation technique as shown in FIGURE I and explain each step in the next paragraphs.

A. Pre-emphasis

The voice samples $\{s_n, n=1...N\}$ are pre-accented to remove the high frequencies which are less energetic than the low frequencies. This step consists in passing the s_n signal through a first-order finite impulse response digital filter given as follows:

$$H(z) = 1 - kz^{-1} \tag{3}$$

In which k is the pre-emphasis coefficient and it must be between $0 \leq k < 1$, in our work we used k=0.97. Thus, the pre-emphasized signal s_n' is related to the signal s_n by the following formula:

$$s_n' = s_n - ks_{n-1} \tag{4}$$

B. Segmentation

The signal processing methods used in speech signal analysis operate on stationary signals, whereas the speech signal is a non-stationary signal. To remedy this problem, the analysis of this signal is performed on successive speech frames, of relatively short duration, over which the signal can generally be considered as quasi-stationary. In this segmentation step, the pre-emphasized signal s_n' is divided into frames of N speech samples. In general, N is set so that each frame corresponds to about 20 to 30 ms of speech. In our study, we used N=25 ms.

C. Windowing

The segmentation of the signal into frames produces discontinuities at the frame boundaries. In the spectral domain, these discontinuities are manifested by side lobes. These effects are reduced by multiplying the samples $\{s_n', n=1...N\}$ by the Hamming window:

$$s_n'' = \left(0{,}54 - 0{,}46 \times \cos\left(\frac{2\pi(n-1)}{N-1}\right)\right) \times s_n' \tag{5}$$

D. The fast Fourier transform (FFT)

During this step, each of the frames of N values is converted from the temporal domain to the frequency domain. The FFT is a fast algorithm for the calculation of the Discrete Fourier Transform (DFT) is defined by the formula (6). The values obtained are called the spectrum.

$$s_n = \sum_{k=0}^{N-1} s_k e^{\frac{-j2\pi kn}{N}} \tag{6}$$

E. Filter bank analysis

The simulation of the human ear, which is based on a specific frequency scale, requires the use of the Mel scale. The latter is composed of a bank of triangular filters FIGURE II spaced linearly up to 1000 Hz, then logarithmically beyond. Each filter provides a coefficient that gives the energy of the signal in the band covered by the filter. The formula for converting the linear scale to Mel scale is as follows:

$$\text{Mel}(f) = 2595 \times \log_{10}\left(1 + \frac{f}{700}\right) \tag{7}$$

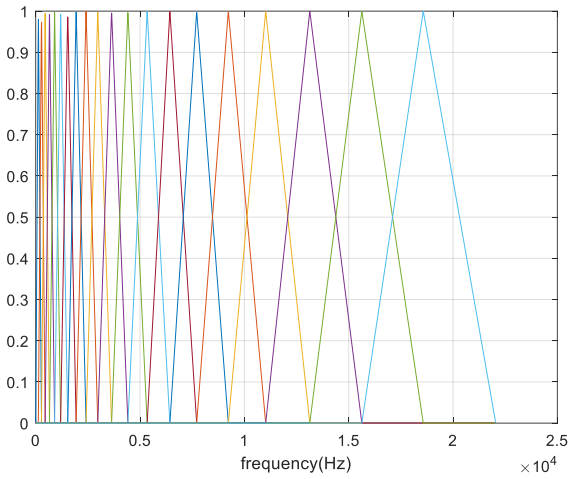

FIGURE II. MEL-SCALE FILTER BANK

F. Logarithm/Discrete Cosine Transform (DCT)

The cepstral coefficients MFCC are obtained by calculating the Discrete Cosine Transform (DCT) applied to the logarithm of the energies leaving the filter bank.

$$c_i = \sum_{j=1}^{M} m_j \times \cos\left(\frac{\pi i}{M}(j - 0{,}5)\right) \tag{8}$$

With mj is the logarithm of the energy obtained with the triangular filter j, M is the number of filter banks, in our work M has been fixed at 20 as shown in FIGURE II.

G. Liftering

Liftering is performed to increase the robustness of the cepstral coefficients. This lifting consists in multiplying these cepstral coefficients by a window of weight W(k) to be less sensitive to the transmission channel and the speaker. As shown in FIGURE III, the higher-order cepstral coefficients are quite small. It is therefore essential to rescale these cepstral coefficients to have similar magnitudes (FIGURE VI). This has been done by lifting the cepstral coefficients according to the following equation

$$c_n' = w(k) \times c_n \tag{9}$$

In which

$$w(k) = \left(1 + \frac{L}{2} \times \sin\left(\frac{\pi n}{L}\right)\right) \quad (10)$$

Where L is the number of coefficients, in our study we used L=22.

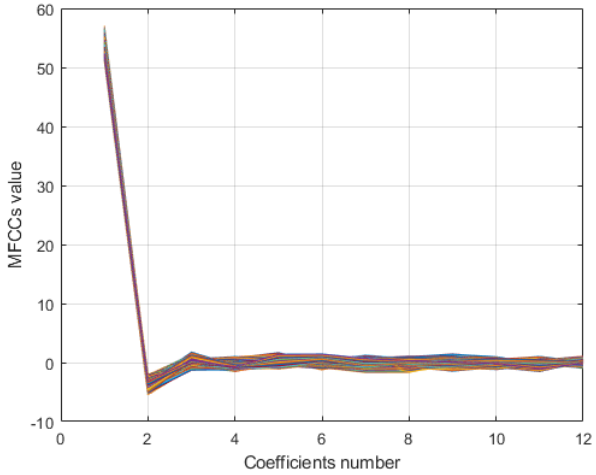

FIGURE III. THE FIRST 12 MEL FREQUENCY CEPSTRAL COEFFICIENTS BEFORE LIFTERING WERE EXTRACTED FROM IMF 2.

VI. METHODOLOGY AND RESULTS

The first step of this work is to construct a dataset containing recordings of speech signals from normal people and people with Parkinson's disease, we apply the database collected in [27] which consists of 20 patients with Parkinson's disease and 18 healthy people, each person (normal and with Parkinson's disease) was asked to pronounce the vowel /a/.

The various calculations are carried out on a PC hp with Windows 10 equipped with a CPU of clock 2,40 GHz and a RAM of 4 Go with the calculation software MATLAB R2019a.

We then used EMD, which is a remarkable method for analyzing nonlinear and nonstationary data. EMD decomposes each patient's speech signal into intrinsic mode function (IMF), which can effectively represent natural signals; therefore, we obtained 8 IMFs for each recording, FIGURE IV shows the signal before and after using the EMD method, and FIGURE V shows the IMF 2 which gave a better result for the classification.

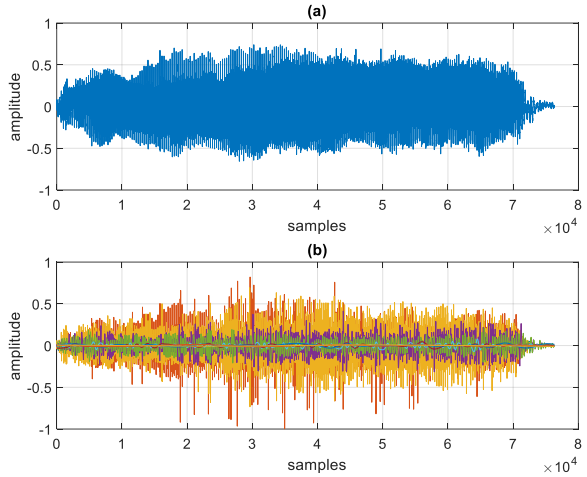

FIGURE IV. (A) SPEECH BEFORE THE TRANSFORMATION. (B) SPEECH AFTER BEING TRANSFORMED BY THE USE OF EMD.

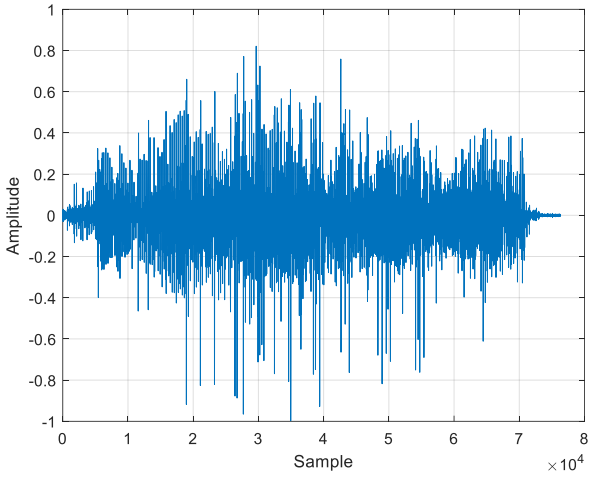

FIGURE V. THE IMF 2 RESULTING FROM THE DECOMPOSITION BY THE EMD METHOD

Subsequently, we used the program "htk mfcc MatLab" [29] to extract from one to 12 MFCC coefficients from each IMF; we work with the first 12 coefficients, which are the most efficient ones. This can be explained by the fact that most of the information useful for the discrimination between patients with PD and healthy patients is found in the low-frequency part of the Mel scale. We proceeded in this way to obtain the optimal number of coefficients needed for the best classification accuracy. The MFCC contains a great number of frames that necessitate a significant processing time for classification and impede making the right diagnostic decision. To surmount this problem, we computed the average value of these frames to obtain the voiceprint of each person; the 12 MFCC coefficients for IMF 2 are shown in FIGURE VI, while FIGURE VII represents the voiceprint of IMF 2.

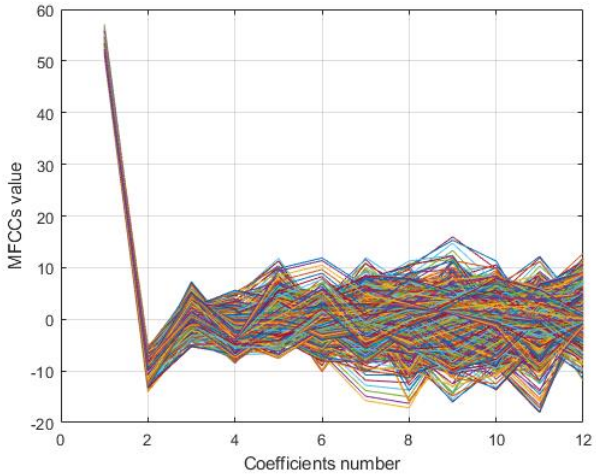

FIGURE VI. THE FIRST 12 MEL FREQUENCY CEPSTRAL COEFFICIENTS AFTER LIFTERING EXTRACTED FROM IMF 2

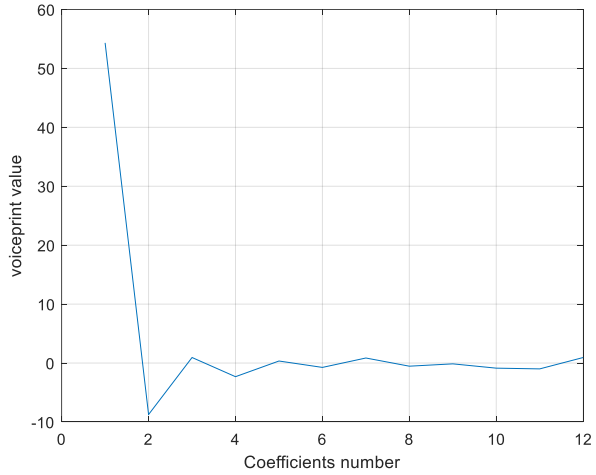

FIGURE VII. VOICEPRINT OF THE FIRST 12 MEL FREQUENCY CEPSTRAL COEFFICIENTS EXTRACTED FROM IMF 2

A final phase concerning classification, in which we make a decision focused on the categorization of patients, in this way we used the "holdout" method in which the data was divided into 80% as training data and we used KNN classifier with its K=1.

To evaluate the performance of our classifier, we used an evaluation metric, which is accuracy, sensitivity, and specificity, by applying the following formulas:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

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

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (13)$$

Where TP means true positive (correctly classified healthy patients), TN means true negative (correctly classified patients), FP means false positive (incorrectly classified patients), and FN means false negative (incorrectly classified healthy patients).

The percentage of the test was expressed by the accuracy, sensitivity, and specificity of all voice signal recordings as shown in Table I, these results are obtained using 80% as the basis of training, so from Table I, it is noticeable that maximum classification accuracy is obtained in IMF 2 by 86.67%.

TABLE I. DIAGNOSTIC RESULTS USING THE KNN CLASSIFIER

IMFs	Accuracy (%)	Sensitivity (%)	Specificity (%)
IMF1	80	38	60
IMF2	86.67	44.44	50
IMF3	80	66.66	65
IMF4	83.33	50	80
IMF5	80	38.88	90
IMF6	83.33	38.88	80
IMF7	60.53	55.55	60
IMF8	76.67	50	50

VII. CONCLUSION

In this paper, we presented a model of diagnosis of Parkinson's disease based on the processing of the speech signal, which we employed the EMD method combined with MFCC applied in a database containing PD patients and healthy patients, they pronounce the vowel /a/. The decomposition of the speech signals is performed by the EMD method, which allows us to obtain 8 IMFs, and then we extract from 1 to 12 MFCC coefficients for each IMF to determine whether the patient is sick or healthy. The classification of Parkinson's disease is performed using a "holdout" method and a KNN classifier, which gives an accuracy of 86,67% when applied to 80% of the database as training data.

REFERENCES

[1] J.PARKINSON," An essay on the shaking palsy. London: Whittingham and Rowland, 1817". Classics in neurology. Huntington, NY: Robert E. Krieger Publishing Co lnc, 1971, p. 158-191.

[2] M.Hireš, M.Gazda, P.Drotár, N. D .Pah, M. A Motin and D. K. Kumar. "Convolutional neural network ensemble for Parkinson's disease detection from voice recordings". Computers in biology and medicine, 2021, p. 105021.

[3] J. L. Manes, E. Herschel, K. Aveni, K. Tjaden, T. Parrish, T. Simuni, ... and A. C. Roberts. "The effects of a simulated fMRI environment on voice intensity in individuals with Parkinson's disease hypophonia and older healthy adults". Journal of Communication Disorders, 2021, vol. 94, p. 106149.

[4] K. Wrobel. "Diagnosing Parkinson's disease by means of ensemble classification of patients' voice samples". Procedia Computer Science, 2021, vol. 192, p. 3905-3914.

[5] D. Meghraoui, B. Boudraa, T. Merazi and P. G. Vilda. "A novel pre-processing technique in pathologic voice detection: Application to Parkinson's disease phonation". Biomedical Signal Processing and Control, 2021, vol. 68, p. 102604.

[6] S. S. Upadhya and A. N Cheeran. "Discriminating Parkinson and healthy people using phonation and cepstral features of speech". Procedia computer science, 2018, vol. 143, p. 197-202.

[7] R. Chiaramonte and M. Bonfiglio. "Acoustic analysis of voice in Parkinson's disease: a systematic review of voice disability and meta-analysis of studies". Revista de neurologia, 2020, vol. 70, no 11, p. 393-405.

[8] Y. E. Huh, J. Park, M. K. Suh, S. E Lee, J. Kim, Y Jeong, ... and J.W. Cho. "Differences in early speech patterns between Parkinson variant of

multiple system atrophy and Parkinson's disease". Brain and language, 2015, vol. 147, p. 14-20.

[9] S. S.Upadhya, A. N. Cheeran, and J. H. Nirmal. "Thomson Multitaper MFCC and PLP voice features for early detection of Parkinson disease". Biomedical Signal Processing and Control, 2018, vol. 46, p. 293-301.

[10] Q. W.Oung, S. N. Basah, H. Muthusamy, V. Vijean and H. Lee. "Evaluation of short-term cepstral based features for detection of Parkinson's Disease severity levels through speech signals". In: IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2018. p. 012039.

[11] A. Benba, A Jilbab and A. Hammouch. "Detecting patients with Parkinson's disease using Mel frequency cepstral coefficients and support vector machines". International Journal on Electrical Engineering and Informatics, 2015, vol. 7, no 2, p. 297.

[12] A.Benba, A. Jilbab, and A. Hammouch." Discriminating between patients with Parkinson's and neurological diseases using cepstral analysis". IEEE transactions on neural systems and rehabilitation engineering, 2016, vol. 24, no 10, p. 1100-1108.

[13] T. B. Drissi, S. Zayrit, B Nsiri and A. Ammoummou. "Diagnosis of Parkinson's disease based on wavelet transform and Mel frequency cepstral coefficients". Int. J. Adv. Comput. Sci. Appl, 2019, vol. 10, p. 125-132.

[14] M. Farrús and J. Codina-Filbà. "Combining prosodic, voice quality and lexical features to automatically detect Alzheimer's disease". arXiv preprint arXiv:2011.09272, 2020.

[15] P. Vizza, P, and G. Tradigo." On the analysis of biomedical signals for disease classification". ACM SIGBioinformatics Record, 2019, vol. 8, no 3, p. 7-10.

[16] S. Mirzaei, M. El Yacoubi, S. Garcia-Salicetti, J. Boudy, C. Kahindo, V. Cristancho-Lacroix, ... and A. S. Rigaud. "Two-stage feature selection of voice parameters for early Alzheimer's disease prediction". Irbm, 2018, vol. 39, no 6, p. 430-435.

[17] R. Frail, J.I. Godino-Llorente, N. Saenz-Lechon, V. Osma-Ruiz and C. Fredouille. "MFCC-based remote pathology detection on speech transmitted through the telephone channel". Proc Biosignals, 2009.

[18] A. Meghanani, C. S Anoop, A. G. Ramakrishnan. "An exploration of log-Mel spectrogram and MFCC features for Alzheimer's dementia recognition from spontaneous speech". In: 2021 IEEE Spoken Language Technology Workshop (SLT). IEEE, 2021. p. 670-677.

[19] J. V. E. López, L. Tóth, I. Hoffmann, J. Kálmán, M. Pákáski and G. Gosztolya. "Assessing Alzheimer's disease from speech using the i-vector approach". In: International Conference on Speech and Computer. Springer, Cham, 2019. p. 289-298.

[20] H. Mukherjee,S. M. Obaidullah, K. C. Santosh, S. Phadikar and K. Roy. "A lazy learning-based language identification from speech using MFCC-2 features". International Journal of Machine Learning and Cybernetics, 2020, vol. 11, no 1, p. 1-14.

[21] K. Sarmah and U. Bhattacharjee. "GMM based Language Identification using MFCC and SDC Features". International Journal of Computer Applications, 2014, vol. 85, no 5.

[22] J. C. Liu, F. Y. Leu, G. L. Lin, H. Susanto. "An MFCC-based text-independent speaker identification system for access control". Concurrency and Computation: Practice and Experience, 2018, vol. 30, no 2, p. e4255.

[23] F. Y. Leu and G. L. Lin." An MFCC-based speaker identification system". In: 2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA). IEEE, 2017. p. 1055-1062.

[24] M. S. Likitha, S. R. R. Gupta, K. Hasitha, and A. U. Raju. "Speech-based human emotion recognition using MFCC". In: 2017 international conference on wireless communications, signal processing, and networking (WiSPNET). IEEE, 2017. p. 2257-2260.

[25] P. P. Dahake, K. Shaw, P. Malathi. "Speaker dependent speech emotion recognition using MFCC and Support Vector Machine". In: 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT). IEEE, 2016. p. 1080-1084.

[26] N. J. Nalini, S. Palanivel." Music emotion recognition: The combined evidence of MFCC and residual phase". Egyptian Informatics Journal, 2016, vol. 17, no 1, p. 1-10.

[27] B. E. Sakar, M. E. Isenkul, C. O. Sakar, A. Sertbas, F. Gurgen, S. Delil, ... and O. Kursun. "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings". IEEE Journal of Biomedical and Health Informatics, 2013, vol. 17, no 4, p. 828-834.

[28] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, ... and H. H. Liu. "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis". Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences, 1998, vol. 454, no 1971, p. 903-995.

[29] Kamil Wojcicki (2021). HTK MFCC MATLAB (https://www.mathworks.com/matlabcentral/fileexchange/32849-htk-mfcc-matlab), MATLAB Central File Exchange. Retrieved December 13, 2021.