

Detecting patients with Parkinson's disease using PLP and VQ

Achraf Benba

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

Abdelilah Jilbab

Laboratoire de Recherche en Génie Electrique, Ecole Normale Supérieure de l'Enseignement Technique,
Mohammed V University
Rabat, Morocco

Ahmed Hammouch

Laboratoire de Recherche en Génie Electrique, Ecole Normale Supérieure de l'Enseignement Technique,
Mohammed V University
Rabat, Morocco

Abstract— Parkinson's disease (PD) is a neurological disorder of unknown etiology. PD causes several symptoms during its course, and this includes voice disorders of 90% of patients. In order to improve the evaluation of these disorders, we have used 34 voice samples of sustained vowel /a/, from a set of 34 people including 17 patients with PD. We subsequently extracted from each person, from 1 to 20 coefficients of the Perceptual linear prediction (PLP). The frames of the PLP were compressed using vector quantization, with six codebook sizes namely; 1, 2, 4, 8, 16 and 32. We used the technique of Leave One Person Out (LOPO) and the Support Vector Machines (SVMs) classifier with two types of kernels; RBF and Linear. The obtained results using the codebook size of 1 were not stable. Therefore, we proceeded to a bench of 100 trials. The best average accuracy obtained was 75.8%, and the maximum classification accuracy obtained was 91.17% using the codebook size of 1.

Keywords— Parkinson's disease, Perceptual Linear Prediction, Vector quantization, Leave One Person Out, Support Vector Machines.

I. INTRODUCTION

THE evaluation of the quality of voice, and the identification of the causes of its degradation based on phonological and acoustic traits have become major concerns of clinicians and voice pathologists. They have become more attentive to any external techniques to their domain, which might provide them additional information for the evaluation of PD. During its course, PD causes different symptoms and affects the system which controls the execution of learned motor plans such as walking, talking or completing other simple tasks [1] [2] [3]. PD is generally seen in people whose age is over 50 years and causes voice weakening in approximately 90% of patients [4]. For these patients, physical visits for diagnosis, monitoring and treatment are too hard [5] [6].

In the case of the evaluation of voice disorders caused by PD, clinicians and the voice pathologists have adopted subjective techniques based on acoustic traits to distinguish different disease levels. In order to develop more objective

evaluations, recent studies use measurements of voice quality in time, frequency and cepstral domains [7] to detect voice disorders in the context of PD. Such as; fundamental frequency of the oscillation of vocal folds, absolute sound pressure level, jitter which represents pitch perturbations, shimmer which represents amplitude perturbations, and harmonicity which represents the degree of acoustic periodicity [1] [8] [9].

In this research we focused on the measurements in cepstral domain by applying PLP cepstral coefficients (CC) which have been usually used in speaker identification applications and were first proposed by H. Hermansky [10]. We have extracted PLP coefficients from the voice signals provided in a dataset and used VQ for data compression. We subsequently used the LOPO validation scheme with SVMs for data classification in order to discriminate PD patients from healthy people.

This paper is organized as follows: the voice samples dataset is described in section II. The PLP processes and VQ are presented successively in section III and IV. The

methodology of this research and the results are presented in section V and conclusion in section VI.

II. DATASET

The data collected in the context of this research belongs to 17 patients with PD (6 women, 11 men) and 17 healthy people (8 women, 9 men). Voice samples were recorded 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 15 cm distant from people and they were requested to say sustained vowel /a/ at a comfortable level. All the recordings of voice samples were made in mono channel mode and saved in WAV format; acoustic analyses were applied on these voice samples. All the recordings were sent by M. Erdem Isenkul from the Department of Computer Engineering at Istanbul University, Istanbul, Turkey.

III. PLP PROCESSES

Our first purpose was to transform the voice signal to some type of parametric representation for more analysis and processing [13]. The voice signal is a slow time varying signal which is called quasi-stationary [13]. When it is observed over a short period of time, it appears fairly stable [13]. However, over a long period of time, the voice signal changes its shape.

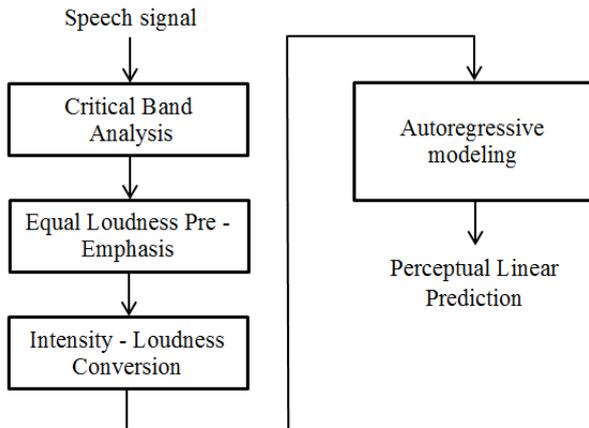


Fig 1: Block diagram of PLP coefficients process

Therefore, it should be characterized by doing short-time spectral analysis [13]. The process of calculating the PLP is shown in Figure 1 and described in the following paragraphs.

A. Spectral Analysis

Since the voice signal is a real signal, it is finite in time; thus, a processing is only possible on finite number of samples [14]. Therefore, the first phase of PLP process is to weight the voice segment by Hamming window [10]. The aim is to reduce signal discontinuities, and make the ends smooth enough to connect with the beginnings [14]. This was

achieved 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 samples [10]:

$$W(n) = \left\{ 0,54 - 0,46 \cdot \cos\left(\frac{2\pi n}{N-1}\right) \right\} \quad (1)$$

where N is the length of the Hamming window, with a length about 20 ms.

The next processing phase consists on converting each frame of N samples from time domain into frequency domain by applying the Fast Fourier Transform (FFT) [13]. We used the FFT for the reason that it is a fast algorithm to implement the Discrete Fourier Transform (DFT) [13]. As known, the DFT is defined on the set of N samples S_n as follow [13]:

$$S_n = \sum_{k=0}^{N-1} s_k e^{-2\pi jkn/N}, n = 0,1,2,\dots, N-1 \quad (2)$$

The short-term power spectrum is calculated by adding the square of the real and imaginary components of short-term voice spectrum, as follow [10]:

$$P(\omega) = \text{Re}[S(\omega)]^2 + \text{Im}[S(\omega)]^2 \quad (3)$$

B. Critical Band Analysis

The short-term power spectrum $P(\omega)$ is warped along its frequency axis ω where $\omega=2\pi f$, into Bark frequency Ω by using the following equation [10]:

$$\Omega(\omega) = 6 \ln \left\{ \frac{\omega}{1200\pi} + \sqrt{\left[\left(\frac{\omega}{1200\pi} \right)^2 + 1 \right]} \right\} \quad (4)$$

$$\Omega(f) = 6 \ln \left\{ \frac{f}{600} + \sqrt{\left[\left(\frac{f}{600} \right)^2 + 1 \right]} \right\} \quad (5)$$

$$\Omega(f) = 6 \sinh^{-1} \left(\frac{f}{600} \right) \quad (6)$$

where ω is the angular frequency in [rad/s], and f is the frequency in [Hz]. The aim of the next phase, is to convolve the resulting warped power with the power spectrum of the simulated critical-band masking curve $\Psi(\Omega)$ approximated by H. Hermansky [10] as follow:

$$\Psi(\Omega) = \begin{cases} 0 & \text{for } \Omega < -1,3 \\ 10^{2,5(\Omega+0,5)} & \text{for } -1,3 \leq \Omega \leq -0,5 \\ 1 & \text{for } -0,5 < \Omega < 0,5 \\ 10^{-1,0(\Omega-0,5)} & \text{for } 0,5 \leq \Omega \leq 2,5 \\ 0 & \text{for } \Omega > 2,5 \end{cases} \quad (7)$$

It is a rather curd approximation of the shape of auditory filters.

The samples of the critical-band power spectrum are produced by doing the discrete convolution of $\Psi(\Omega)$ with $P(\omega)$ by applying the following equation [10]:

$$\theta(\Omega_i) = \sum_{\Omega=-1,3}^{2,5} P(\Omega - \Omega_i) \Psi(\Omega) \quad (8)$$

The convolution between the relatively broad critical-band masking curve $\Psi(\Omega)$ and the short-term power spectrum $P(\omega)$, reduces the spectral resolution of $\theta(\Omega)$ in comparison with the original $P(\omega)$ [10].

C. Equal-loudness Pre-emphasis

The next phase in this process is to pre-emphasize the samples $\Theta[\Omega(\omega)]$ using the simulated equal-loudness curve, by applying the following equation [10]:

$$\Xi[\Omega(\omega)] = E(\omega) \times \Theta[\Omega(\omega)] \quad (9)$$

where, $E(\omega)$ is an approximation to the non-equal sensitivity of human ear perception at different frequencies. The practical approximation used in this research was adopted by H. Hermansky [10] and was first proposed by Makhol *et al* [15] and given by the following equation:

$$E(\omega) = \frac{(\omega^2 + 56,8 \times 10^6) \omega^4}{(\omega^2 + 6,3 \times 10^6)^2 \times (\omega^2 + 0,38 \times 10^9)} \quad (10)$$

$$E(f) = \left[\frac{f^2}{f^2 + 1,6 \times 10^5} \right]^2 \times \left[\frac{f^2 + 1,44 \times 10^6}{f^2 + 9,6 \times 10^6} \right] \quad (11)$$

D. Intensity-loudness Power Law

The last operation before the all-pole modeling is the cubic-root amplitude compression. The following equation approximates the power law of human hearing and simulates the non-linear relation between the intensity of sound and its perceived loudness [10]:

$$\Phi(\Omega) = \Xi(\Omega)^{0,33} \quad (12)$$

E. Autoregressive Modeling

In the final phase of the PLP process, $\Phi(\Omega)$ is approximated by the spectrum of an all-pole model using the autocorrelation

technique of all-pole spectral modeling, this technique is called Linear Prediction (LP) [10] [16], in which the signal spectrum is modeled by an all-pole spectrum. In this research we used the Linear Predictive Coefficient (LPC) analysis to calculate the autoregressive model from spectral magnitude samples. The autoregressive coefficients are transformed to CC of the all-pole model; this was achieved by converting the LPC of n coefficients into frames of CC [10].

F. Liftering

The principal advantage of CC is that they are uncorrelated [14]. However, the problem with them is that the higher orders CC are quite small [14]. Therefore, it is essential to re-scale the CC in order to have quite similar magnitudes [14]. This is achieved by liftering the CC according to the following equation [14]:

$$c'_n = \left(1 + \frac{L}{2} \cdot \sin\left(\frac{\pi \cdot n}{L}\right) \right) \cdot c_n \quad (13)$$

where L is the cepstral sine lifter parameter. In this research, we used $L=0.6$.

IV. VECTOR QUANTIZATION

VQ is a compression technique with data-loss [17]. The basic idea of this technique is to take a large number of data vectors and minimize it to a smaller group of data vectors, which represent the centers of gravity of the distribution.

The VQ technique consists of extracting a small number of the most representative data to characterize different people. Here VQ is used to minimize the number of frames of the coefficients of the PLP in order to have only the most significant vectors which represent the center of gravity of the distribution of other frames of the PLP coefficients. In this research, we have made tests using codebook sizes of 1, 2, 4, 8, 16 and 32 [18].

V. METHODOLOGY & RESULTS

The first phase in this research was to build a dataset containing voice samples of patients with PD and healthy people. Ultimately, we were able to collect 17 voice samples from both groups. This gave us 34 voice samples [19]. These recordings were made through a standard microphone at a sampling rate of 44100 Hz. All participants were asked to pronounce the sustained vowel /a/ at a comfortable level.

All the algorithms were executed on a desktop computer with a Core (TM) i3-2120 CPU and a processing speed of 3.30 GHz. We subsequently extracted from each voice sample, CC of the PLP. The number of PLP coefficients extracted ranged from 1 to 20. We proceeded in this way to get the optimal coefficient number needed for the best classification accuracy.

The PLP coefficients extracted from each sample contains a large number of frames which require extensive processing time for classification and prevents making the correct diagnostic decision. To overcome this problem, and reduce the

processing time, we used a technique of compression with data-loss known as VQ. The detailed description of this technique has been made in section IV. As we know, VQ compresses the frames according to the number of codebooks. In this paper we have used six codebooks of size 1, 2, 4, 8, 16 and 32. We applied this technique over 20 PLP coefficients which have already been extracted from each voice sample, and which contains from 1 to 20 coefficients per person. This makes a total of 120 (6 * 20) extraction operations per person.

To train and validate our classifier, we used a technique of classification called LOPO, that is, we left out all the compressed frames of the PLP coefficients of one person to be used for validation as if it were an unseen person, and trained a classifier on the rest of the compressed frames of other people [6]. We used the LOPO technique of classification iteratively for each coefficient per person until all 20 coefficients per person for the six different codebooks size. In this paper, we used the SVMs classifier with its different types of kernels, i.e.; RBF, and Linear.

During the test section, we noticed that the obtained results when using a codebook size of 1 are not stable. Unlike the other codebook sizes, namely 2, 4, 8, 16 and 32, the compression of the PLP frames using a codebook size of 1, did not always give the same location of the centroids of the clusters forming the compressed PLP coefficients. Therefore, every time we redid the same test, we will not get the same classification results. To evaluate on how the results change, we used a test bed of 100 times. This test bed allows us to obtain the minimum, maximum and average value of the classification results for PD [19] [20] [21].

We made a test bed of 100 times, for the codebook size of 1, and 5 times for the codebook size of 2, 4, 8 and 16, and only one time for the codebook size of 32. As already mentioned, the obtained results using a codebook size higher than 1 are stable, nonetheless we did the test bed 5 times on the others to get an idea of the variation in execution time and to be sure that the results remained the same.

Based on our results, it is clear that using higher codebook size decreases the accuracy of classification as it is mentioned in Figure 3. It is clear from Table I that the time required for processing becomes longer.

The extracted PLP coefficients from each person contains in addition to the number of coefficients used, many frames with different values. The use of a large number of frames leads us to a diversity of results, often very close to the extracted values from other people (PD and healthy) [19]. This similarity of results between different people prevents making the correct diagnostic decisions [19]. By way to explanation; assuming that the frames are in the form of points distributed in space, increasing the number of frames leads to interference between these points [19]. Therefore, the task of the classifier, to find a hyper plane able to separate perfectly the two groups of people (namely patients with PD and healthy people), will be very difficult, if not impossible [19].

As can be seen in Table I, for a single test using a codebook size of 1, we need 78.71 seconds. Each time we increase the

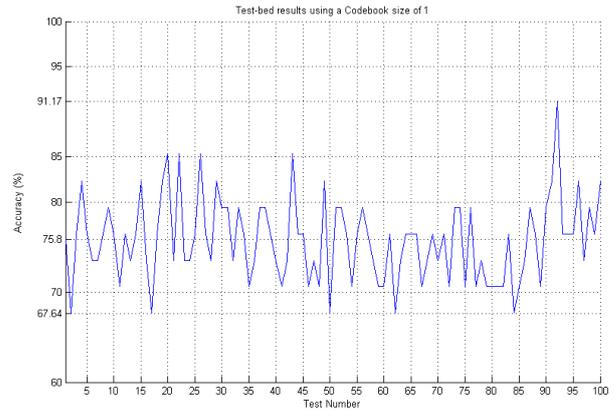


Fig 2: The test bed results using the codebook size of 1 with minimum, maximum and average classification accuracies

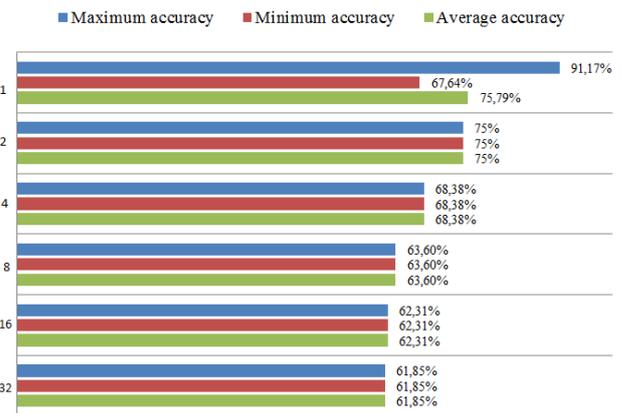


Fig 3: Classification results for different Codebook sizes (i.e. 1, 2, 4, 8, 16, and 32).

TABLE I
 EXECUTION TIME OF THE CLASSIFICATION PROGRAM FOR DIFFERENT SIZES OF THE CODEBOOK

Codebook sizes	Max time (second)	Min time (second)	Average time (second)
1	94.16	75.54	78.71
2	82.06	81.56	81.78
4	130.69	130.20	130.46
8	315.41	312.15	314.17
16	3.15e+03	3.13e+03	3.14e+03
32	4.94e+04	4.94e+04	4.94e+04

size of the codebook, the processing time for classification becomes longer. For a single test using a codebook size of 16 we need about 53 minutes and with a codebook size of 32 we need about 14 hours. For this size, it is not practical to apply a test bed if we already know that the results will remain the same even after 100 trials.

The test bed accuracy results using the codebook size of 1 are represented in Figure 2. A maximum classification accuracy of 91.17% was achieved using a codebook size of 1 as shown in Figure 3, by linear kernel SVMs. As seen from Figure 3, the best average classification accuracy of 75.8% was achieved using a codebook size of 1.

VI. CONCLUSION

A Dysarthria symptom associated with PD is a slow process whose early stages may go unobserved. To improve the evaluation of PD we collected a variety of voice samples from different people during the pronunciation of sustained vowel /a/. The extracted PLP coefficients from different people contain many frames which take maximum processing time in the classification section, and prevent making correct diagnosis. Therefore, we have compressed the extracted PLP coefficients using VQ with different codebook sizes.

After doing the tests we noticed that the obtained results using a codebook size of 1 were not stable. To evaluate on how the results change, we proceeded to a bench of 100 trials. The compression of the frames of the PLP coefficients using VQ with the codebook size of 1 has shown to be a good parameter for the detection of voice disorder in PD, showing an average classification accuracy of 75.8% and a maximum classification accuracy of 91.17%.

ACKNOWLEDGEMENT

The authors would like to thank M. Erdem Isenkul, Istanbul University. Thomas R. Przybeck and Daniel Wood, US Peace Corps Volunteers (Morocco 2013-2015). Dr. Robert Michael Hutchman, USA.

REFERENCES

- [1] Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J., & Ramig, L. O. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *Biomedical Engineering, IEEE Transactions on*, 56(4), 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, Betül 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] Uma Rani, K., and Mallikarjun S. Holi. "Automatic detection of neurological disordered voices using mel cepstral coefficients and neural networks." *Point-of-Care Healthcare Technologies (PHT)*, 2013 IEEE. Bangalore, Indi, pp. 76-79.
- [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] Hermansky, Hynek. "Perceptual linear predictive (PLP) analysis of speech." *the Journal of the Acoustical Society of America* 87.4 (1990): 1738-1752.
- [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," *Proc Biosignals*, Porto, 2009.
- [12] Jafari, A., "Classification of Parkinson's disease patients using nonlinear phonetic features and Mel-frequency cepstral analysis," *Biomedical Engineering: Applications, Basis and Communications* 25.04 (2013).
- [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] Makhoul, John, and Lynn Cosell. "LPCW: An LPC vocoder with linear predictive spectral warping." *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'76*. Vol. 1. IEEE, 1976.
- [16] Makhoul, John. "Spectral linear prediction: properties and applications." *Acoustics, Speech and Signal Processing, IEEE Transactions on* 23.3 (1975): 283-296.
- [17] Kapoor, Tripti, and R. K. Sharma. "Parkinson's disease Diagnosis using Mel-frequency Cepstral Coefficients and Vector Quantization." *International Journal of Computer Applications* 14.3 (2011), pp. 43-46.
- [18] J. Martinez, H. Perez, E. Escamilla, M. M. Suzuki, "Speaker recognition using mel frequency cepstral coefficients (MFCC) and Vector Quantization (VQ) techniques," *IEEE Electrical Communications and Computers*, Cholula, Puebla, Feb 2012, pp. 248-251.
- [19] 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, September 23-25 2014, Saint Petersburg, Russia*. pp. 96-100.
- [20] 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 (WCCS'14, November 10-12 2014, Agadir, Morocco, IEEE, 2014*.pp. 622-625.
- [21] Achraf Benba, Abdelilah Jilbab, 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*, pp. 84-90.