International Journal of Computer Applications (0975 – 8887)
Volume 14– No.3, January 2011

Parkinson's disease Diagnosis using Mel-frequency Cepstral Coefficients and Vector Quantization

Tripti Kapoor
Dept. Of ECE,
National Institute of Technology
Kurukshetra, Haryana, India

R.K. Sharma
Dept. Of ECE,
National Institute of Technology,
Kurukshetra, Haryana, India

ABSTRACT:
This paper investigates the adaptation of MFCCs to the diagnosis of Parkinson's disease (PD). The aim of this study is to provide a novel method, suitable for keeping track of the evolution of the patient's pathology: easy-to-use, fast, non-invasive for the patient, and affordable for the clinicians. This method will be complementary to the existing ones - the perceptual judgment and the usual objective measurement (jitter, airflows...) which remain time and human resource consuming. The system designed for this particular task relies on the Mel-Frequency Cepstral coefficients (MFCC) for feature extraction and Vector Quantization (VQ) for feature analysis which is the state-of-the-art for speaker recognition.

Keywords
Parkinson disease, disease diagnosis, Mel frequency Cepstral coefficients, Vector Quantization.

I. INTRODUCTION
Parkinson's disease is caused by dopaminergic deficiencies in the basal nuclei that generate motor actions. According to Robbins, 70 to 92% of these patients progress with diseases of the tongue, larynx and pharynx. Oral communication is the main complaint in 30% of patients. Studies describe that vocal alterations, such as imprecise articulation, decreased speech speed, reduced vocal intensity and lower variation of fundamental frequency, are commonly alterations presented by people suffering from Parkinson. Parkinson's dysarthria is characterized by abnormal voice modulation, hoarseness, voice tremor, decrease loudness and monotone voice. There is no medical treatment of PD, although medication is available offering significant alleviation of symptoms, especially at the early stages of the disease (Singh, Pillay, & Choonara, 2007). Many of people with Parkinson's disease will therefore be substantially dependent on clinical intervention. The requisite physical visits to the clinic for monitoring and treatment are difficult for many people with Parkinson's disease. Research has shown that the most important symptoms of PD are dysphonia [1]. In literature, it is claimed that approximately 90% of people with Parkinson's disease exhibit some form of vocal impairment [2]. The people with Parkinson's typically display a constellation of vocal symptom that includes impairment in the normal production of vocal sounds, which is dysphonia. The dysphonia is a broad term that refers to disorders of voice, and it includes any pathological or functional problem with one's voice [3]. The voice will sound hoarse, strained or effortful. The voice may be difficult to understand because their voice sounds very quiet or is so distinctive that it distracts from the content of what the person is saying. Most of the current methods used for evaluating Parkinson's disease (PD) rely heavily on human expertise . In diagnosis of PD literature, there have been extensive studies of speech measurement for general voice disorders [5,6] and PD in particular [7,8].In these studies, the speech sounds produced during standard speech tests are recorded using a microphone, and the recorded speech signals are subsequently analyzed using measurement methods implemented in software algorithms) designed to detect certain properties of these signals.
Mel-frequency cepstral coefficients (MFCC) have traditionally been used in speaker identification applications. Their use has been extended to speech quality assessment for clinical applications during the last few years. In this paper we have used MFCC for the feature extraction from the speech signals provided in the database and vector quantization is used for feature matching in order to discriminate healthy persons from the Parkinson's patients.
This paper is organized as follows: the speech and subject database and methods for feature extraction and feature analysis are described in section II. The results are presented in Section III and conclusion in Section IV.

II. MATERIALS AND METHODS

A. Speech and subject database

For this analysis, the speech record database consisted of a sustained phonation (2 s long) of the vowel /a/ .The PD group consisted of 30 patients with Parkinson's disease, sixteen males (aged between 58 and 72 years, mean of 63.6 years) and fourteen females (aged between 48 and 72 years, mean of 60 years) and duration of the disease from 1 to 18 years, with an average of 8 years. The control group was composed of 20 individuals without neurological disorders, three males (aged from 54 to 71 years, mean of 58.6 years) and two females (aged from 54 to 65 years, mean of 59.5 years).The audio recordings correspond to the phonation of sustained vowels, since the whole paradigm relies on the assumption that a healthy person is able to produce sustained sound that is stationary from the signal processing point of view, while the presence of some kind of dysphonia implies the appearance of nonstationarities (e.g. pitch changes, noise related to air turbulences). Bearing this in mind, the calculated acoustic parameters act as perturbation measures.
The application of mel-frequency cepstral coefficients (MFCC) to the automatic assessment of voice was first proposed by [9] and [10]. The use of MFCC for voice analysis stems from the field of speaker identification and it was first supported more by empirical evidence than by theoretical reasoning. Two additional practical reasons supported this choice: calculation of MFCC does not require pitch detection and these parameters have been shown to be fairly robust against some kinds of voice distortion [11]. In addition, it has been argued that analysis in the cepstral domain for this application is justified by the presence of noise level information in the cepstrum [12] and that MFCC give a compression of this information on the first part of the cepstrum, hence providing some dimensionality reduction and easing the task of pattern classifiers [13].

B. Mel-Frequency Cepstral Coefficients

A block diagram of the structure of an MFCC processor is given in figure 1. The speech input is recorded at a sampling rate of 22050Hz. This sampling frequency is chosen to minimize the effects of aliasing in the analog-to-digital conversion process [14].
The first step in processing has been to split each voice record in 20-ms frames with 50% overlapping between consecutive frames. This short-term processing is justified since detecting the nonstationary nature on pathological voices, as mentioned before, requires avoiding long-term measurements. After splitting the voice signal in frames, each frame undergoes a Finite Fourier transform (FFT), hence computing a short-term FFT of each voice record. This transforms the signal into frequency domain. The left graph of figure 2 shows (in gray) the modulus of the FFT corresponding to one speech frame.

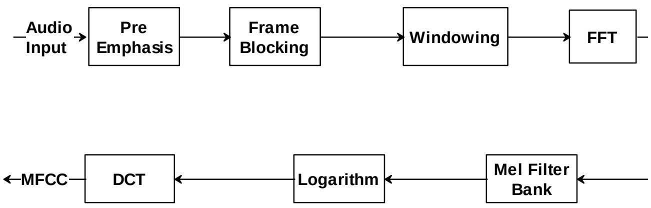

Figure.1 Block diagram of MFCC processor

It is well known that this function results from the superposition of two effects: the rapid variations of the graph, coming from the quasi-periodic vibrations of the vocal folds, and the slow variations or spectral envelope, mainly related with the specific form of the glottal wave and the resonances induced by the vocal tract.

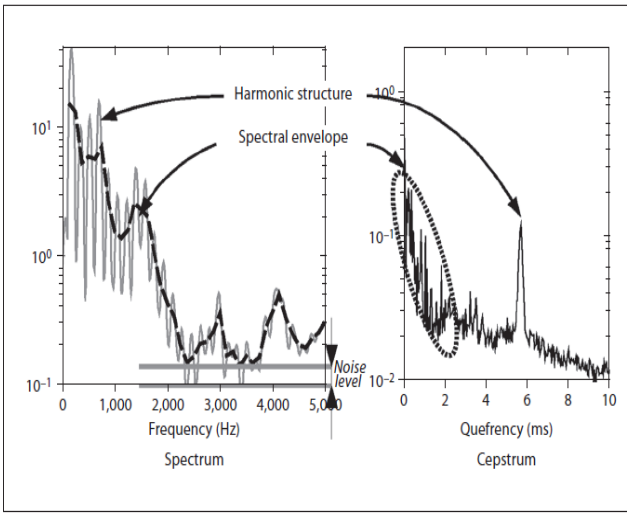

Figure.2 Graphical interpretation of the cepstrum

The second step of MFCC calculation is spectrum smoothing.
This smoothing is carried out using a set of filters that mimic the perceptual behavior of the human auditory system. This choice of filters results in a higher spectral resolution at lower frequency bands, where the most significant information for dysphonia detection is contained [15], and it performs better than a set of regularly spaced filters [11]. The main effect of spectrum smoothing is removing the harmonics of the spectrum while keeping an estimate of the spectral envelope (this is the black dashed line in the same graph). The minimum values of this envelope are related to the aperiodicities or nonstationarities of voice, since for a perfectly periodic signal these values should be zero.
The logarithmic operation that comes afterwards and DCT that transforms the signal from spectral domain into the cepstral domain allow representing the form of the spectral envelope in the first part of the cepstrum (figure.2). For this reason, the first coefficients of the cepstrum keep information on glottal noise [12]. Moreover, the specific spectrum smoothing carried out in MFCC computation further concentrates information on the first cepstral coefficients. Implementation of the above-described cepstral analysis allows converting each 20-ms voice frame, which at a sample frequency of 22050Hz is formed by 500 samples, into a feature vector having an optimal length of 15–20 MFCC [11]. For the experiments reported in this paper, the length of the feature vector has been chosen to be 16.

C. Vector Quantization

Vector quantization (VQ) is a lossy data compression method based on principle of blockcoding [16]. It is a fixed-to-fixed length algorithm. VQ may be thought as an aproximator. Figure. 3 shows an example of a 2-dimensional VQ.

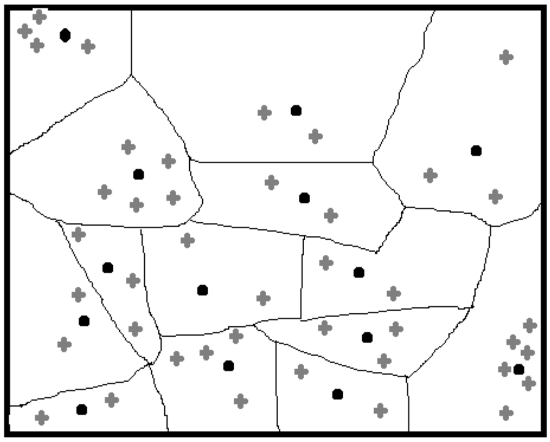
Figure.3 An example of 2-dimensional VQ

Here, every pair of numbers falling in a particular region are approximated by a star associated with that region. In Figure 2, the stars are called *codevectors* and the regions defined by the borders are called *encoding regions*. The set of all codevectors is called the *codebook* and the set of all encoding regions is called the *partition* of the space [16].

D. LBG design algorithm

The LBG VQ design algorithm is an iterative algorithm which alternatively solves optimality criteria [17]. The algorithm requires an initial codebook. The initial codebook is obtained by the splitting method. In this method, an initial code vector is set as the average of the entire training sequence. This code vector is then split into two. The iterative algorithm is run with these two vectors as the initial codebook. The final two code vectors are split into four and the process is repeated until the desired number of code vectors is obtained. Figure 4 shows the LBG Design algorithm.
In the training phase, a speaker-specific VQ codebook is generated for each known speaker by clustering his/her training acoustic vectors. The distance from a vector to the closest codeword of a codebook is called a VQ distortion. In the recognition phase, an input utterance of an unknown voice is "vector-quantized" using each trained codebook and the *total VQ*

distortion is computed. The speaker corresponding to the VQ codebook with the smallest total distortion is identified.

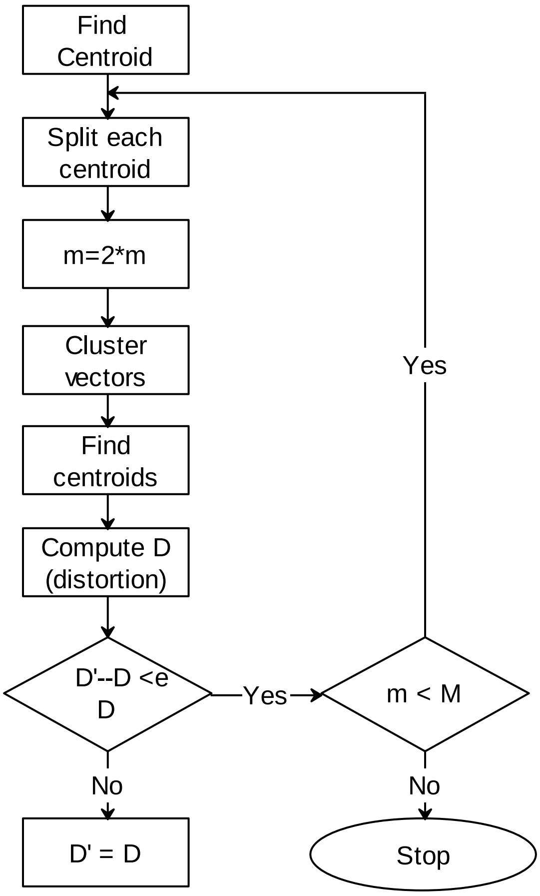

Figure.4 LBG Algorithm Flowchart

III. RESULTS

We use 50 phonations in total, pronounced by different speakers, 20 of whom are normal and the other present pathologies of Parkinson's disease. For the training, we use 20 phonations (10 normal and 10 PD patients). After training, the system is tested with 30 phonations different from those used for the training (10 normal and 20 PD Patients).
Table.1 presents results of classification of the normal voices and voices with the PD.

Table 1
VQ Results with Codebook Size of 16

Phonation	Normal	Patient with PD
Training Number	10	10
Test Number	10	20
Correct Classification	9	19
Rate Of Classification	90%	95%

IV. CONCLUSION

Within this paper, the application of cepstral analysis for the clinical evaluation of voice function has been qualitatively reviewed. The mathematical transformations involved in the analysis have been described as well as the suitability of the analysis for this application. Cepstral analysis, or more specifically MFCC, is a robust procedure that does not required previous pitch detection. At the same time, it keeps information on the spectral envelope of the voice signal, which is related to the form of the glottal signal and the presence of turbulence noise. Moreover, it achieves that with a reduced set of parameters, hence easing subsequent processing by pattern classifiers. VQ is used to minimize the data of the extracted feature. The study reveals that as number of centroids increases, identification rate of the system increases. It also suggests that in order to obtain satisfactory result, the number of centroids has to be increased as the dataset increases.

V. REFERENCES

[1] Little, M.A., McSharry, P.E., Hunter, E.J., Spielman, J., Ramig, L.O., "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease", IEEE Transactions Biomedical Engineering (September 11, 2008), doi: 10101/npre.2008.2298.1

[2] Ho, A., Iansek, R., Marigliani, C., Bradshaw, J., Gates, S., "Speech impairment in a large sample of patients with Parkinson's disease",.Behavioral Neurology 11, 131-37 (1998)

[3]Godino-Llorente, J. I., & Gomez-Vilda, P., "Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors". IEEE Transactions on Biomedical Engineering, 51, 2004, pp 380–384.

[4] Little, M., McSharry, P., Moroz, I., & Roberts, S. (2006), "Nonlinear biophysicallyinformed speech pathology detection", In Proceedings of the ICASSP 2006. New York: IEEE Publishers.

[5] Little, M.A., McSharry, P.E., Roberts, S.J., Costello, D., Moroz, I.M., "Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection", Biomedical Engineering Online **6:23** (2007)

[6] Cnockaert, L., Schoentgen, J., Auzou, P., Ozsancak, C., Defebvre, L., & Grenez, F.(2008), "Low-frequency vocal modulations in vowels produced by Parkinsonian subjects", Speech Communication, 50, pp 288–300.

[7] Rahn, D. A., Chou, M., Jiang, J. J., & Zhang, Y., "Phonatory impairment in Parkinson's disease: Evidence from nonlinear dynamic analysis and perturbation analysis", Journal of Voice, 21, (2007), pp 64–71.

[8] A. A. Dibazar, S. Narayanan, T. W. Berger, "Feature analysis for automatic detection of pathological speech", in: Engineering Medicine and Biology Symposium02, Vol. 1, 2002, pp. 182–183.

[10] Fraile R, Godino-Llorente JI, Sáenz-Lechón N, Osma-Ruiz V, Fredouille C, "MFCC-based remote pathology detection on speech transmitted through the telephone channel", Proc Biosignals, Porto, 2009.

[11] Murphy PJ, Akande OO, "Quantification of glottal and voiced speech harmonics-to-noise ratios using cepstral-based estimation", Proc 3rd Int Conf Non-Linear Speech Processing (NOLISP'05), Barcelona, 2005, pp 224–232.

[12] Fraile R, Sáenz-Lechón N, Godino-Llorente JI, Osma-Ruiz V, Gómez-Vilda P, "Use of melfrequency cepstral coefficients for automatic pathology detection on sustained vowel phonations:

mathematical and statistical justification”, Proc 4th Int Symp Image/Video Commun Fixed and Mobile Networks, Bilbao, 2008.

[13] Md. Rashidul Hasan, Mustafa Jamil, Md. Golam Rabbani Md. Saifur Rahman, “Speaker Identification Using Mel Frequency Cepstral Coefficients”, 3rd Proceedings of International Conference on Electrical & Computer Engineering , ICECE 2004, 28-30 December 2004, Dhaka, Bangladesh,pp 565-568

[14] Pouchoulin G, Fredouille C, Bonastre JF, Ghio A, Giovanni A: “Frequency study for the characterisation of the dysphonic voices”. Interspeech 2007, 8th Annu Conf Int Speech Commun Assoc, Antwerp, 2007, pp 1198–1201.

[15] R. M. Gray, ``Vector Quantization,'' IEEE ASSP Magazine, pp. 4--29, April 1984.

[16] Y. Linde, A. Buzo & R. Gray, “An algorithm for vector quantizer design”, IEEE Transactions on Communications, Vol. 28, pp.84-95, 1980.

[17] R. Fraile a N. Sáenz-Lechón a J.I. Godino-Llorente a V. Osma-Ruiz a C. Fredouille, “Automatic Detection of Laryngeal Pathologies by MFCC”, Folia Phoniatr Logop 2009;61:146–152

[18] Karine Rigaldie, Jean Luc Nespoulous, Nadine Vigouroux, “Dysprosody in Parkinson’s disease : Musical scale production and intonation patterns analysis” , Speech Prosody 2006, Dresden, Germany, May2-5, 2006.