Effects of a New Voicing Parameter on Pathological Voice Discrimination by SVM

Asma Belhaj*, Aicha Bouzid, Noureddine Ellouze
Laboratory of Signal, Images and Information Technology, Tunis El Manar University, National Engineering School of Tunis, TUNISIA

*E-mail: belhajasma {at} yahoo.fr

Abstract— **In this paper, we introduce new voicing parameters to describe the speech signal and we study their effects on the classification of disordered voices. These parameters concern the fundamental frequency and the open quotient. The open quotient is defined as the ratio of the open phase by the pitch period. These parameters are calculated using the multiscale product method (MPM). The classification is operated on two pathological databases MAPACI and MEII using SVM classifiers multi-class one-against-all. We consider two types of classifications: a binary classification into normal and pathological voices for female and male speakers and a three-category classification into edema, nodule and normal voices for the female speakers only. The effects of these new parameters are studied when added to the MFCC coefficients, delta, delta second, and the energy.**

Keywords- Pathological Voices, SVM, MFCC, Open Quotient.

I. INTRODUCTION

Speech is the most important form of a direct communication and any pathology that affects the speaking capabilities will have a large impact on the both professional and social activities. In recent years, researchers in such fields as laryngology and speech science have become increasingly interested in the acoustic characteristics of normal and pathological voices [1-3]. Nowadays, dysphonia is a disease affecting more and more subjects due to the disturbance of the produced speech, while the larynx is involved in the phonation. We can distinguish some types of dysphonia; organic dysphonias that are caused by pathological changes in the vocal folds and dysfunctional dysphonias characterised by difficulties in phonation without obvious organic alteration of the vocal folds. But due to the application of compensation mechanisms by the patient, a dysfunctional dysphonia can induce an organic dysphonia. Besides, we note neurological dysphonias caused by neurological damages.

The assessment of pathological voices can be very relevant for both diagnosis and therapy evaluation. The assessment of the voice quality can be made by a diagnostician or by a direct examination as the laryngostroboscopy.

The diagnosis may be performed following two different approaches: the perceptive and the objective ones. On the one hand, the perceptive assessment consists in qualifying the voice pathologies by listening the production of a patient. This evaluation is performed by trained professionals who rate the speech samples on a GRBAS grade scale [4] according to their perception of voice disorder. This subjective evaluation suffers of the drawbacks to be highly dependent on the experience of the listener and on its inconsistency on judging pathological voice quality. On the other hand, the objective analysis consists in qualifying and quantifying the voice pathologies by acoustical, aerodynamic, and physiological measurements. It offers the advantages to be quantitative, cheaper, faster, and more comfortable for the patient than methods like the electroglottography (EGG) [5] or the imaging of the vocal folds by stroboscopy [6] or more recently by high-speed camera [7].

Specialists achieve a diagnosis after a clinical observation and the learning evaluation of the patient's voice quality. Given the complex and subjective nature of the personal listening, researchers have developed various tools for establishing a diagnosis.

In fact, many methods of acoustic evaluation of pathological voices have been proposed in the literature. Among them, the automatic classification of pathological voice has received a considerable attention.

The most important classifiers used in the speech recognition are considered in the classification of pathological voices as: neural networks, the Gaussian mixture model (GMM), the hidden Markov model (HMM), and the support vector machines (SVM). In the literature, these studies have proposed a binary normal/pathological classification of voice samples [8,10]. Besides the SVM classifier using specific parameters, have achieved the best performance. However, the classification between the pathologies is operated in few works [9] and the results are not sufficiently efficient.

In this work, we consider the two types of classification using the classic features MFCC, Δ, $\Delta\Delta$, the energy, and the

fundamental frequency added to a new parameter proposed in this work: the open quotient. The open quotient is defined as the ratio of the open phase by the pitch period. The open phase is the time interval separating the glottal opening instant (GOI) and the following glottal closure instant (GCI).

This paper is organized as follows. In Section 2, we give a brief description of the classic features used in the pathologic voice classification. Section 3 presents the most important classifiers. In section 4, we describe the two databases used in this work: MEII database and MAPACI database. Section 5 present the features extracted from the speech signal and essentially the fundamental frequency and the open quotient. In Section 6, we give the principles of the binary and the multi-category classifications using SVM. The results are presented in section 7. Finally conclusion and future work are drawn in section 8.

II. Classic Features in Pathologic Voice Discrimination

The classic features used in the classification of pathologic voices are inspired from the cues used in the field of the speech recognition that are essentially the fundamental frequency F0, the mel frequency cepstral coefficients (MFCC), their first and second derivatives and the energy, the harmonic to noise ratio (HNR).
The subject of this section is the overview of the most common features involved in the pathological voice assessment.

A. Fundamental Frequency

It's an obvious parameter describing the speech voicing state. This parameter is used in most of the studies, sometimes in conjunction with the Mel-Frequency Cepstral Coefficients (MFCC).

B. Mel -Frequency Cepstral Coefficients

MFCCs are one of the most widely-used features to reduce the redundancy of the speech signal to be used in domains like recognition or coding [11]. These coefficients are computed by weighting the Fourier Transform of the signal by a MEL filterbank, then computing the cepstrum from this weighted spectrum and finally the Discrete Cosine Transform (DCT) of this cepstrum.

C. Harmonic to Noise Ratio (HNR)

This parameter is defined as the log ratio of the energy of the periodic and aperiodic components [12]. It can be computed with different approaches. In fact, some methods are based on a model in which speech is assumed to be composed of a periodic component and an aperiodic component [13, 14] while other use the short-time autocorrelation function [15]. All these approaches are based on the estimation of the fundamental frequency. It is used in [16] for the discrimination between normal and pathological voices using the MEEI database. The same measure is used in [17] to show that HMM is able to classify different voice qualities and in [18, 19] to discriminate normal and pathological voices through a telephone channel.

D. Acoustic Features from MDVP Software

The Multi-Dimensional Voice Program (MDVP) is software produced by KayPentax Corp [20]. This software provides some acoustic descriptors defined in [21] and stored with speech samples in the MEEI database. The parameters concern the perturbation of the fundamental frequency and the amplitude of the signal.

Some proposed classification systems use these acoustic descriptors computed directly from the MDVP software [22, 23]. Some other systems use features inspired from those computed by MDVP software [18, 19], meaning that their definitions is taken or inspired from [20].

III. Classifiers used in Normal/Pathological Voices Discrimination

The aim of this section is to describe the different types of classifiers used in the voice pathology assessment. Their structure and behavior are briefly presented.

A. Gaussian Mixture Model

The Gaussian Mixture Modeling (*GMM*) is widely used in Automatic Speaker Recognition, where it acts as a supervised classification system. It is adapted from speaker identification to a classification in one grade of *GRBAS* scale (from 0 (normal) to 3). The GMM classification system operates following three steps [24].
(i) Parametrization, (ii) Model training and (iii) Classification: when a speech sample has to be classified, the likelihood between this sample and each GMM is estimated and the decision relies on the maximum between these likelihoods.
For the normal/pathological classification, 95% of normal subjects and 81.7% of pathological ones are correctly classified. For the grade classification, 95% is obtained for the grade 0 corresponding to the normal subjects while a loss of performance is observed for the pathological ones, specially between adjacent grades. The same system is used in [25] to determine which kind of information is better suited to the classification of the four grades.

B. Support Vector Machines

Support Vector Machines (SVMs) [26] are a well-known classifier used in problems of classification, regression, and novelty detection. Recent researches use this classifier in discrimination between normal and pathological samples. For example, [2] proposes to use a set of features consisting of 11 MFCC coefficients, HNR (Harmonic to Noise Ratio), NNE

(Normalized Noise Energy), GNE (Glottal to Noise Excitation), Energy, and their first derivatives.
The classifier is trained on the vowels /a/ from the pathological corpus of MEII Database (53 normal samples and 77 pathological samples) and the average correct classification rate is 95.12%. The SVM classifier using features extracted from wavelet transform of speech samples to discriminate between normal and pathological voices [27]. The correct classification rate is this time 97.5% for normal voices and 100% for pathological ones.

C. *Neural Networks*

The Artificial Neural Networks are ones of the widely used classifier in various domains, as pattern classification and recognition and particularly speech recognition. Basically this type of classifier can be viewed as an interconnexion between simple small units, the neurons, designed to model to some extent the behaviour of human brain. In [28], this type of classifier is applied on MEII database to distinguish between normal and pathological samples. The input layer is composed of 26 neurons corresponding to 26 acoustic descriptors given by the MDVP software. Besides, the classifier is composed of 1-hidden layer and 1-neuron output layer for normal or pathologic decision. The average correct classification rate is 94% when HNR, VTI, and ShdB are used as input features. The discrimination between normal and pathological samples is also operated on a database of 5 spanish sustained vowels (100 normal samples and 68 pathological samples) [29]. Each vowel is treated by a neural network which takes as input classic parameters and others extracted from the bicoherence. The decisions from the 5 networks are then combined to decide if the input sample is healthy or not. The correct classification rate is 94.4% for the classic parameters and is increased of 4% when the others ones are added.

IV. DATABASES

A. *MEII corpus*

The Kay Elemetrics Voice Disorder Database was developed by the Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Labs (Kay Elemetrics Corp., 1994). The acoustic samples are sustained phonations of vowel [a] (3 – 4 s long) and the first 12 seconds of the Rainbow Passage spoken by normophonic subjects and patients with organic, neurological, traumatic, and psychogenic voice disorders at different stages (from early to fully developed). The speech samples have been recorded in a controlled environment at 25 kHz or 50 kHz and 16 bits of resolution. We have considered a subset comprising 53 normal and 169 pathological voices omitting recordings devoid of a diagnosis and balancing samples with regard to sex and chronological age [30] as shown in table 1.

B. *MAPACI corpus*

In the MAPACI speech pathology database, all voice samples were recorded using a Senheiser headset microphone at 44,1 Hz during the life time project of MAPACI (2003). This database consists of 24 male voice samples (12 normal and 12 pathological) and 24 female voice samples (12 normal and 12 pathological). The recordings consist in three utterances of the vowel /a/ of about 3s record [31].The details of recordings (Range, average and standard deviation in years) are given in table 1.

	Subjets		Range(yaers)		Average (years)		Standard deviation (years)	
	Male	Female	Male	Female	Male	Female	Male	Female
MEII								
Normal	21	32	26/58	22/52	38.8	34.2	8.49	7.87
Patho-logic	70	103	26/58	21/51	41.7	37.6	9.38	8.19
MAPACI								
Normal	12	12	20/68	24/39	24.7	32.1	5.6	5.5
Patho-logic	12	12	27/68	20/63	49.5	40.8	13.9	15.7

Table 1. Details of the recordings used in this study

V. FEATURE EXTRACTION

The feature extraction constitutes the first step in a classification system whose scheme is summarized in the figure 1.

This step consists in:

1- Determining the MFCCC coefficients, Δ, ΔΔ, and the energy,

2- Estimating the fundamental frequency, the open quotient and their variations.

The first coefficients are computed using the melcepst function provided by the voicebox toolbox [32].

The speech signals are divided into frames of 46.44 ms and with a half recovery.
i) For the MAPACI database:

The sampling frequency is 44.1 kHz, so each window contains 2048 samples with an overlap of 1024 samples.

ii) For the MEII database :

The sampling frequency is even 25 kHz, so the window contains 1161 samples with an overlap of 581 samples, or 50 kHz, and the window contains 2322 samples with an overlap of 1161 samples.

The open quotient and the fundamental frequency are determined by the glottal closure instant (GCI) and the glottal opening instant (GOI) detected by the multiscale product (MP) of the speech signal.

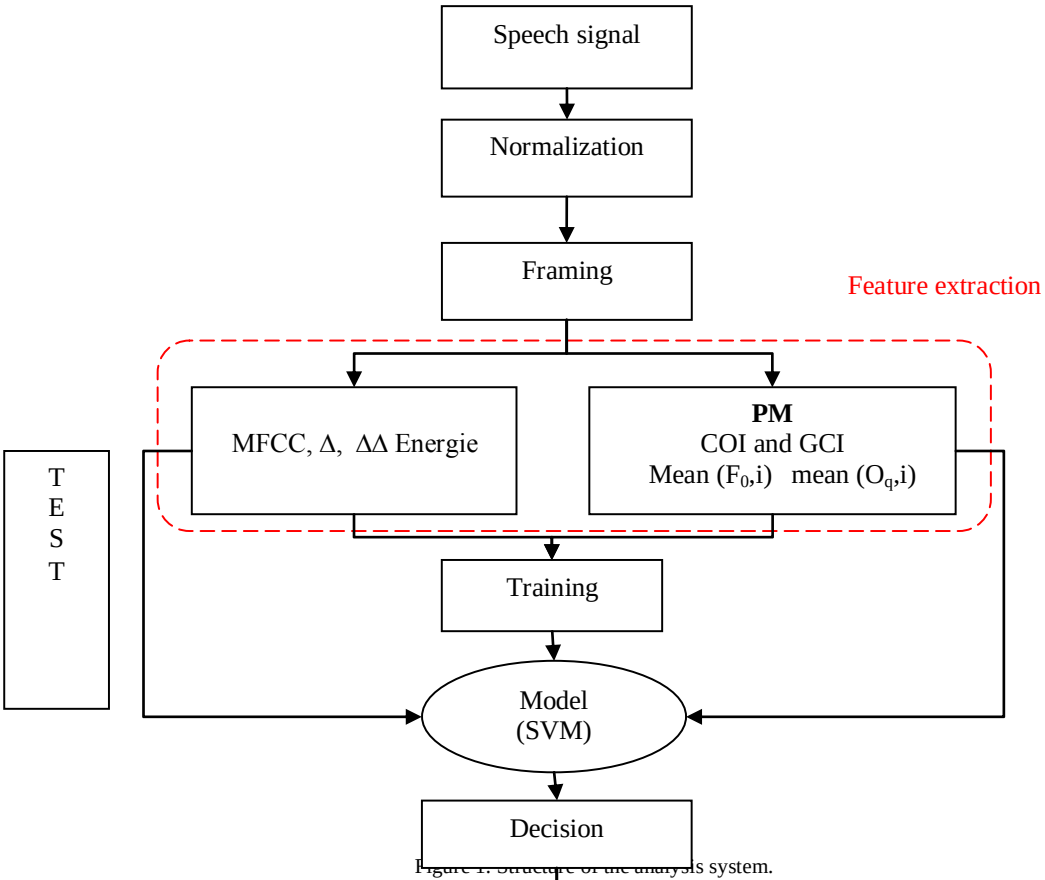

Figure 1. Structure of the analysis system.

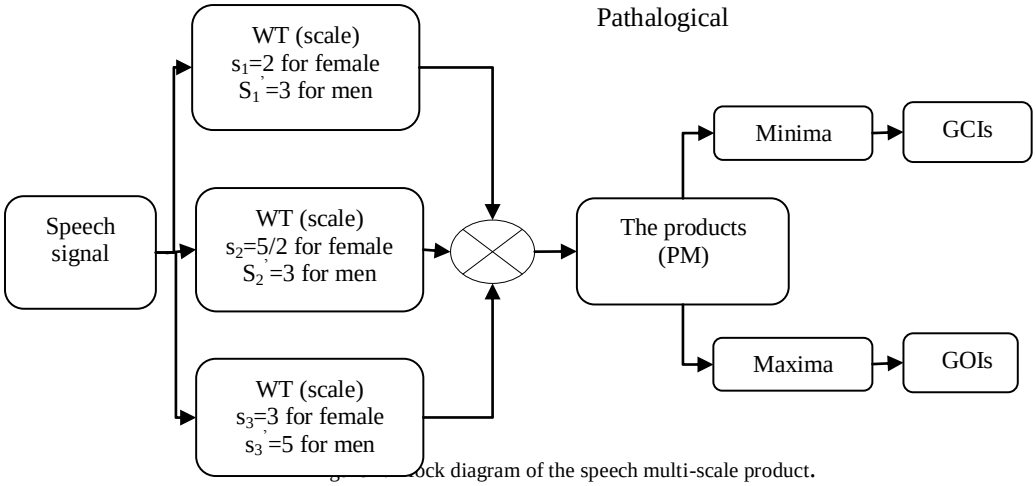

Figure 2. Block diagram of the speech multi-scale product.

A. MP for GCI and GOI detection

The algorithm shall calculate the product of the coefficients of the wavelet transform for different successive scales as shown in figure 2:

$$p(n) = \prod_{j=1}^{3} \psi_{2^j}(f(n)) \quad (1)$$

Where $\psi_{2^j}(f(n))$ is the wavelet transform of function $f(n)$ at scale 2^j.

We note in the cross scale product two types of peaks, minima corresponding to GCI are the most distinguishable, and maxima related to GOI are considerably weaker but discernible. The GOI is the maximum detected between two GCI, as shown in figure 3 and 4.

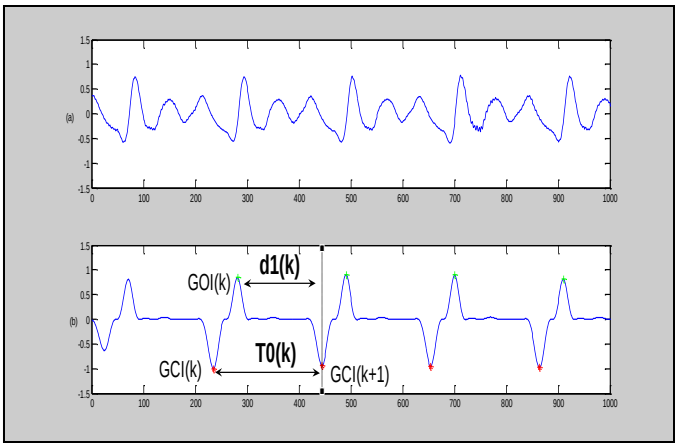

Figure 3. Speech normal voice corresponding to a sustained vowel /a/ extracted from AXH1 pronounced by a female speaker and its MP of MEEI database.

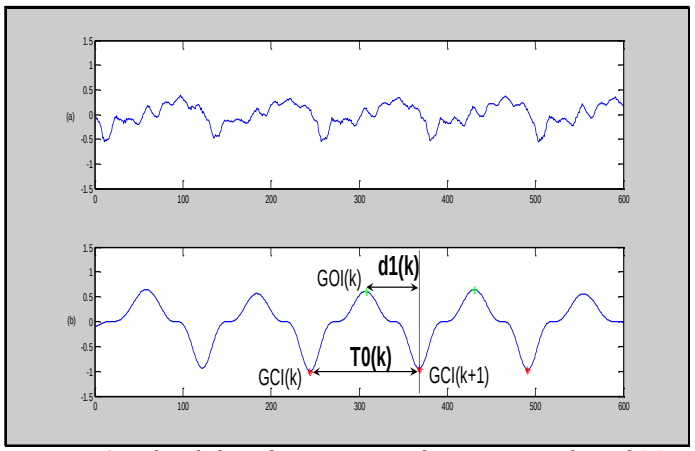

Figure 4. Speech pathological voice corresponding to a sustained vowel /a/ extracted from paralysis AXT13 pronounced by a male speaker and its MP of MEEI database.

B. Fundamental frequency and open quotient estimation

For our databases, we use two combinations of scales s1 = 2, s2 = 5/2, s3= 3 for women and a second s1'= 3, s2' = 4, s3 '= 5 for men. This difference in the choice of scales is that the discontinuities at the GCI and GOI had more in men than in women as shown in figure 3 and 4. The wavelet used in this work is the quadratic spline function. For each analysis window, we locate the GCIs and GOIs, then we calculate the instantaneous values of the fundamental frequency and the open quotient from the following equations.

The local pitch period is given by the following formula:

$$T_0(k) = GCI(k+1) - GCI(k) \quad (2)$$

The local fundamental frequency F_0 (k) is given by the inverse of the pitch period:

$$F_0(k) = \frac{1}{T_0(k)} \quad (3)$$

The open quotient is defined as the ratio of the duration of the open phase by the fundamental period.

$$Q_\phi = \frac{GOI(k+1) - GCI(k)}{T_0(k)} \quad (4)$$

The mean values of the fundamental frequency and the open quotient for the i[th] window are calculated according to the following relationships:

$$F_0^i = \frac{1}{N}\sum_{k=1}^{N} F_0^i(k) \quad (5)$$

$$Q_\phi^i = \frac{1}{N}\sum_{k=1}^{N} Q_\phi^i(k) \quad (6)$$

$$\sigma_{F_0}^i = \frac{1}{N}\sum_{k=1}^{N} \left(F_0^i(k) - F_0^i\right)^2 \quad (7)$$

$$\sigma_{Q_\phi}^i = \frac{1}{N}\sum_{k=1}^{N} \left(Q_\phi^i(k) - Q_\phi^i\right)^2 \quad (8)$$

With:

i : Is the index of the window.

k : Is the index of the period in the window.

N : Is number of periods in the window i.

$F_0^i(k)$: Is the instantaneous fundamental frequency of the k[th] period in the i[th] window.

$Oq_0^i(k)$: Is the instantaneous open quotient of the k[th] period in the i[th] window.

Motivated by the efficiency and the robustness of the multi-scale product for edge detection, we apply it on pathological signal [33-36].

VI. SVM CLASSIFIER PRINCIPLE

Support vector machines have been used for the automatic classification of normal/pathological voices [37]. In the linearly separable case, the SVM optimization algorithm maximizes the margin between the two classes as shown in figure 5. In non-linearly separable cases, a mapping of the input data to some higher-dimensional space, where the data are linearly separable, is carried out by means of a Kernel function *K* that has to satisfy some properties called the Mercer conditions. The margin maximization algorithm leads to the following classifier, with *x* a data point to be assigned to one of two classes, according to the sign of the function in equation (9).

$$f(x)=\sum_{i=1}^{N}\alpha_i y_i K(x_i,x)+b \qquad (9)$$

Where $y_i \in \{-1,1\}$ are the class labels, b the bias term and

$$\sum_{i=1}^{N}\alpha_i y_i = 0, \quad \alpha_i > C$$

In this paper, a Gaussian kernel has been used.

$$K(x,z)=\exp\left(-\frac{\|x-z\|^2}{2\sigma^2}\right) \qquad (10)$$

Prior to training, inverse kernel width γ and a penalty parameter *C* that is part of the cost function must be fixed. A larger *C* value corresponds to assigning a higher penalty to classification errors. A grid-search within intervals defined by the user is carried out to identify (*C*, γ) pairs that enable the classifier to predict unknown data as accurately as possible $C = [10^0, 10^3]$ and $\gamma = [10^{-4}, 10^{-2}]$. The SVM toolbox *Rouen software* has been used for SVM training and classification [39].

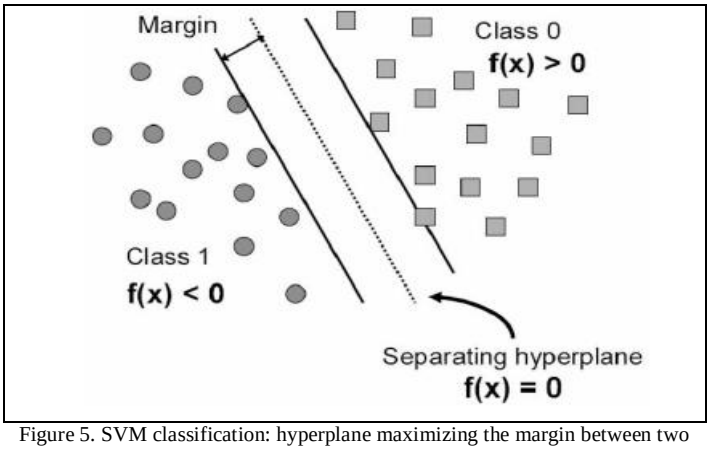

Figure 5. SVM classification: hyperplane maximizing the margin between two classes.

A. *Multi-category SVM classifier*

The one-against-all method has been used for multi-category classification in the framework of which one classifier is constructed for every pair of different classes [38, 39]. The total number of binary classifications is *K* with *K* the number of categories. The final decision is made using a majority rule. For each binary classification, the vote of the category in which the unknown sample has been classified is incremented by one. The sample is assigned to the class with the largest vote [40].

VII. RESULTS

The results of our work will be presented for the MAPACI and MEII databases. Besides, we operate on each database two types of classification: a binary classification for all voices to provide a preliminary diagnosis concerning normal or pathologic voice and a classification into three classes normal/edema/nodule for female voices only. For the second classification, we consider only female voices because the number of male voices presenting edema or nodule pathologies in the MAPACI database is very reduced.

A. *Normal / pathologic voice classification using MAPACI database*

The normal / pathologic classification is performed on the entire MAPACI database men and women using various combinations of parameters.

The classification rates of normal and pathological voice are reported in the table 2 and illustrated in plotted on the figure 6. Using the MFCC parameters only permits to obtain good results, expressed by 91.84% for normal voices and 87.03% for pathological voices. The introduction of the new parameter Oq with MFCC coefficients improves the results of normal / pathological classification, as the coefficients Δ and ΔΔ. Indeed, the open quotient improves the rate to 92.6% for normal voices and the Δ and ΔΔ coefficients can improve the rate to 94.85% for normal voices and 87.12% for the pathological voices. These are the best rates.

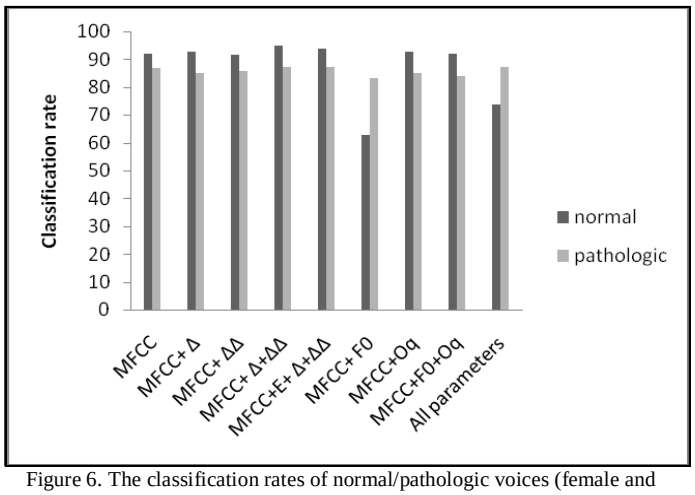

Figure 6. The classification rates of normal/pathologic voices (female and male) using MAPACI database

The classification rates of normal voices drop to 62.75% with the fundamental frequency and to 73.89% with all parameters.

Table 2. Confusing matrix of the normal / pathologic classification using all voices of the MAPACI database

		Normal	Pathologic
MFCC	Normal	91.84	8.16
	Pathologic	12.97	87.03
MFCC+ Δ	Normal	92.74	7.26
	Pathologic	14.84	85.16
MFCC+ ΔΔ	Normal	91.64	8.36
	Pathologic	14.27	85.73
MFCC+ Δ+ΔΔ	**Normal**	**94.85**	**5.15**
	Pathologic	**12.88**	**87.12**
MFCC+E+ Δ+ΔΔ	Normal	93.74	6.26
	Pathologic	12.85	87.15
MFCC+ F_0	Normal	62.75	37.25
	Pathologic	16.69	83.31
MFCC+ O_q	**Normal**	**92.6**	**7.4**
	Pathologic	**14.88**	**85.12**
MFCC+ F_0+ O_q	Normal	91.84	8.16
	Pathologic	15.88	84.12
All parameters	Normal	73.89	27.11
	Pathologic	12.85	87.15

B. Triple edema/nodule/normal classification for all women of the MAPACI database

In this classification, we use 6 edema, 6 nodule and 12 normal voices.

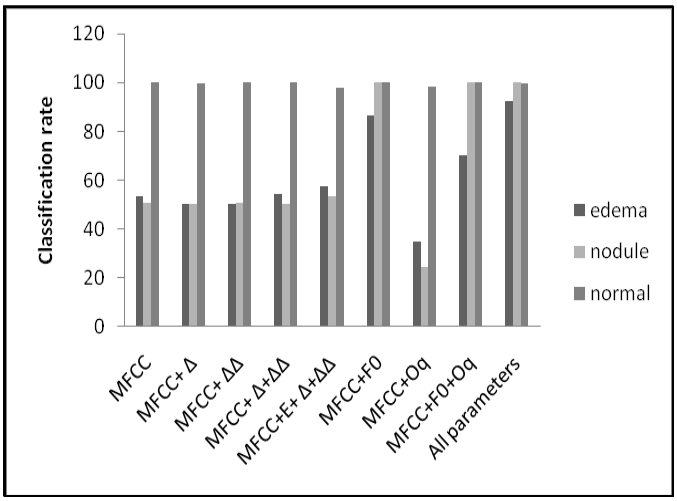

Figure 7. Classification rates in the MAPACI database of edema, nodule, and normal female voices using a 3-class SVM.

The classification rates are reported on the figure 7 and table 3.

Table 3. Confusing matrix for the classification in the MAPACI database of edema, nodule, and normal female voices using a 3-class SVM

		Edema	Nodule	Normal
MFCC	Edema	53.45	43.36	3.19
	Nodule	25.5	50.71	23.79
	Normal	0	0	**100**
MFCC+ Δ	Edema	50.28	48.03	1.69
	Nodule	14.04	50.42	5.54
	Normal	0.54	0	99.46
MFCC+ ΔΔ	Edema	50.46	46.91	2.62
	Nodule	17.76	50.71	31.52
	Normal	0	0	100
MFCC+ Δ+ΔΔ	Edema	54.39	43.36	2.25
	Nodule	15.47	50.42	34.10
	Normal	0	0	100
MFCC+E+Δ+ΔΔ	Edema	57.57	36.07	6.36
	Nodule	20.92	53.29	25.79
	Normal	2.32	0	97.68
MFCC+ F_0	Edema	**86.35**	4.3	9.35
	Nodule	0	**100**	0
	Normal	0	0	**100**
MFCC+ O_q	Edema	**34.95**	61.12	3.93
	Nodule	23.20	**24.64**	52.16
	Normal	1.94	0	**98.06**
MFCC+ F_0+ O_q	Edema	70.09	28.78	1.13
	Nodule	0	100	0

	Normal	0	0	100
All parameters	**Edema**	**92.33**	**2.61**	**5.06**
	Nodule	**0**	**100**	**0**
	Normal	**0**	**0.23**	**99.77**

The best recognition rates are obtained with all parameters or MFCC + F_0. The fundamental frequency appears as the most discriminating parameter for the recognition of diseases. However, the open quotient introduces confusion between the two diseases. In addition, it seems that the MFCC coefficients are efficient to recognize normal voices with a high rate.

C. Normal / pathologic voice classification using MEII database

The normal / pathologic classification is also performed on the entire MEII database men and women using various combinations of parameters.

As reported in table 4 and figure 8, we can see that all parameter combinations provide good results. However, the best rates are given with the MFCC coefficients only or associated with their first and second derivatives or with the open quotient O_q.

The fundamental frequency drops the recognition of pathological voices from 88% to 83%. With all the parameters, the results are lightly worse than with F0.

These results are broadly in agreement with those obtained with the MAPACI database.

Table 4. Confusing matrix of the normal / pathologic classification using all voices of the MEII database

		Normal	Pathologic
MFCC	Normal	**99.65**	**0.35**
	Pathologic	**11.26**	**88.74**
MFCC+ Δ	Normal	99.82	0.67
	Pathologic	11.40	88.60
MFCC+ ΔΔ	Normal	99.65	0.35
	Pathologic	11.26	88.74
MFCC+ Δ+ΔΔ	Normal	99.82	0.18
	Pathologic	11.26	88.74
MFCC+E+ Δ+ΔΔ	Normal	99.34	0.66
	Pathologic	11.40	88.60
MFCC+ F_0	Normal	96.16	3.84
	Pathologic	16.69	83.31
MFCC+ O_q	Normal	**99.82**	**0.18**
	Pathologic	**11.26**	**88.74**
MFCC+ F_0+ O_q	Normal	96.16	3.84
	Pathologic	16.69	83.31
All parameters	Normal	94.21	5.79
	Pathologic	16.69	83.31

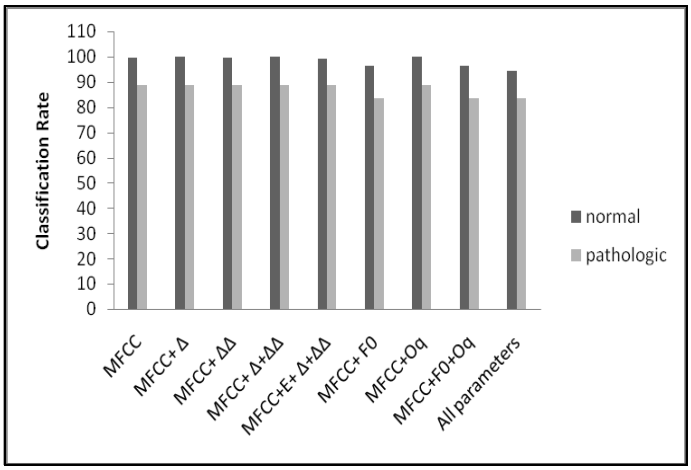

Figure 8. The classification rates of normal/pathological voices (female and male) using MEII database.

D. Triple edema/nodule/normal classification for all women of the MEII database

In this classification we use 14 edema, 14 nodule, and 14 normal voices. The classification rates are reported on the figure 9 and table 5.

Table 5. Confusing matrix for the classification in the MEII database of edema, nodule, and normal female voices using a 3-class SVM

		Edema	Nodule	Normal
MFCC	Edema	**73.38**	26.62	0
	Nodule	12.57	**74.85**	12.58
	Normal	0	0	**100**
MFCC+ Δ	Edema	73.38	26.62	0
	Nodule	12.57	74.85	12.58
	Normal	0	0	100
MFCC+ ΔΔ	Edema	73.38	26.62	0
	Nodule	12.57	74.85	12.58
	Normal	0	0	100
MFCC+ Δ+ΔΔ	Edema	73.38	26.62	0
	Nodule	12.57	74.85	12.58
	Normal	0	0	100
MFCC+E+ Δ+ΔΔ	Edema	78.42	21.58	0
	Nodule	12.57	74.25	13.18
	Normal	0	0	100
MFCC+ F_0	Edema	64.74	31.65	3.60
	Nodule	31.14	40.71	28.15
	Normal	0	0	100
MFCC+ O_q	Edema	**40.95**	55.12	3.93

	Nodule	52.15	**24.64**	23.21
	Normal	1.94	0	**98.06**
MFCC+ F_0+ O_q	Edema	67.62	32.38	0
	Nodule	33.53	61.08	5.39
	Normal	0	1.24	98.76
All parameters	**Edema**	**100**	**0**	**0**
	Nodule	**20.96**	**79.04**	**0**
	Normal	**13.65**	**0**	**86.35**

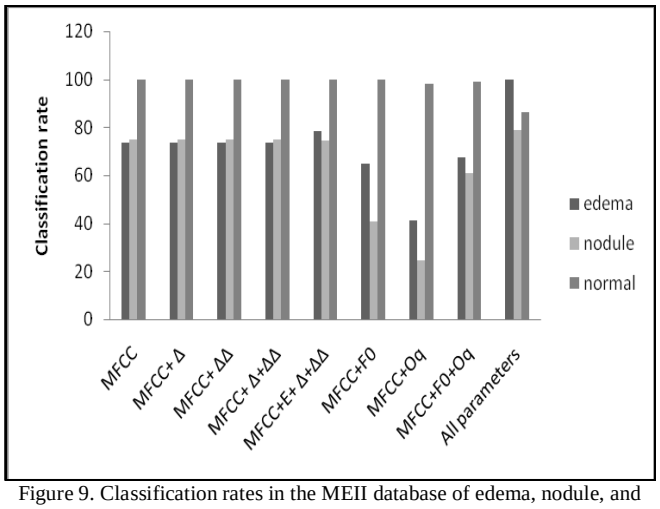

Figure 9. Classification rates in the MEII database of edema, nodule, and normal female voices using a 3-class SVM.

We note that the best recognition is obtained with all parameters. The MFCC coefficients used alone recognize 100% of the normal speakers, the edema with 73.39% and the nodule with 74.85%. The open quotient Oq influence on edema and nodule classes by lowering the recognition of edema to 64.74% and the nodule to 40.71%, against the recognition of normal voices remains effective.

VIII. PERFORMANCE OF THE CLASSIFICATION SYSTEM

In order to quantify the classifier performance, we consider three measures: sensitivity, specificity, and the overall accuracy. These measures are calculated from:

- the true positive (TP: the classifier classified as pathology when pathological samples are present),
- true negative (TN: the classifier classified as normal when normal samples are present),
- false positive (FP: the classifier classified as pathological when normal samples are present),
- false negative (FN: the classifier classified as normal when pathological samples are present).

These measures are calculated using the following equations

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

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

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

These parameters are extracted from the confusion matrix as follows in table 6.

Table 6. The structure of a confusion matrix

		True Class	
		Normal	*Pathological*
Class predict	*Normal*	*TN*	*FN*
	Pathological	*FP*	*TP*

A. Performance of our normal / pathologic classification system using MAPACI database

Table 7 gives the values of sensitivity, specificity and accuracy for different settings for normal / pathological classification of MAPACI database.

In the MAPACI database, the classification using MFCC coefficients and their first and second derivatives presents the best performance in terms of specificity, sensitivity and accuracy. The classification using MFCC coefficients + O_q has high of these parameters but are just below the best case. The classification using MFCC coefficients only remains efficient. The classifications using MFCC coefficients + F_0 or all parameters has a bit far better performance.

Table7. Performance of normal / pathological classification of MAPACI database

	Sensibility %	*Specificity %*	*Accuracy %*
MFCC	**91.42**	**87.62**	**89.43**
MFCC+ Δ+ΔΔ	**94.42**	**88.04**	**95.28**
MFCC+ F_0	69.1	78.99	73.03
MFCC+ O_q	**92**	**86.15**	**88.86**
All parameters	76.26	85.18	80.52

B. Performance of our edema/nodule/normal female voice classification using MAPACI database

Table 8 gives the values of the classification accuracy by a 3-SVM for selected parameters for all the female voices of the MAPACI database according to edema, nodule or normal.

The classification using all parameters has the best accuracy. It is followed by MFCC + F_0 far beyond the MFCC alone or with $\Delta + \Delta\Delta$ or O_q.

Table 8. Accuracy rates of the edema/nodule/normal classification for female voices of the MAPACI database

	Accuracy %
MFCC	75.75
MFCC+ Δ+ΔΔ	76
MFCC+ F_0	96.5
MFCC+ O_q	63.75
All parameters	**97.5**

C. Comparison of the Performance of our normal / pathologic classification system with other studies using the MAPACI database

Our approach for the normal/pathologic classification using the most efficient parameters is compared to those proposed by Hariharan, Polatb, Sindhuc, Yaacoba in [41] and Arias-Londono, Godino-Llorente, Markaki, Stylianou in[42].

The sensitivity, specificity and accuracy values of each approach are given in table 9. We notice that our SVM classifier with MFCC + Δ + ΔΔ parameters is more performing in terms of sensitivity, specificity and accuracy than the approach in [42]. The approach given in [41] and using the PCA / FCM (fuzzy c-means clustering) parameters is the best.

Table 9. Comparison of the Performance of our normal / pathologic classification system with other studies using the MAPACI database

Author	***Method***	***Parameters Years***		***SE %***	***SP %***	***ACC %***
[42]	*GMM*	*12MFCC*	*2010*	**77**	**83.04**	**80.01**
	SVM	*MS*	*2010*	**80.50**	**82.91**	**81.70**
[41]	*SVM*	*PCA*	*2013*	**90.73**	**85.59**	**87.96**
	SVM	*LDA*	*2013*	**82.27**	**73.24**	**77.02**
	SVM	*PCA/FCM*	*2013*	**99.85**	**100**	**99.43**
Our app-roach	***SVM***	***MFCC+ Δ+ΔΔ***	***2014***	**94.42**	**88.04**	**95.28**

D. Performance of our normal / pathologic classification system using MEII database

Table 10 illustrates the sensitivity, specificity and accuracy values for selected parameters for the classification of normal and pathologic voices in the MEII database.

The classification using the MFCC coefficients alone or with their first and second derivatives and the combination MFCC + O_q give the highest rates.

The classifications using the MFCC coefficients + F_0 or all parameters give the worst results.

Table 10. Performance of normal / pathological classification in MEII database

	Sensibility %	*Specificity %*	*Accuracy %*
MFCC	**99.6**	**89.84**	**94.19**
MFCC+Δ+ΔΔ	**99.8**	**89.86**	**94.28**
MFCC+ F_0	95.59	85.21	89.73
MFCC+ O_q	**99.8**	**89.86**	**94.28**
All parameters	95 .51	84.95	88.76

E. Performance of our edema/nodule/normal female voice classification using MEII database

The triple classification edema / nodule / normal for women voices by a 3-class SVM using MEII database is presented in table 11. The combination of all parameters has the highest accuracy rate. The combinations MFCC + F_0 and MFCC + O_q have the less accuracy.

Table 11. Accuracy rate of the classification results (edema, nodule, and normal) for female speakers by a 3-class SVM using MEII database

	Accuracy %
MFCC	82.33
MFCC+ Δ+ΔΔ	82.33
MFCC+F_0	68
MFCC+O_q	54
All parameters	**88.33**

F. Comparison of the Performance of our normal / pathologic classification system with other works using MEII database

We compare the best performance of our approach obtained with MFCC + O_q to those proposed by Hariharan, Polatb, Sindhuc, Yaacoba in [41], Arias-Londono, Godino-Llorente, Markaki, Stylianou in [42] and Alpan, Schoentgen, Maryn, Grenez in [43] as shown in table 12.

Table 12. Comparison of the performance of our approach with other works

Author ACC %	***Method***	***Parameters***	***Years***	***SE %***	***SP%***
[42]	*GMM*	*12MFCC*	*2010*	**95.20**	**91.04**	**94.22**
	SVM	*MS*	*2010*	**97.38**	**79.72**	**93.22**
[43]	*SVM*	*SDR*	*2010*	**98.2**	**97.9**	**98.1**
	SVM	*MFCC+E+ Δ+ΔΔ*	*2010*	**98.7**	**94**	**97.6**
[41]	*SVM*	*PCA*	*2013*	**92.01**	**90.29**	**91.12**
	SVM	*LDA*	*2013*	**90.22**	**89.05**	**89.61**
	SVM	*PCA/FCM*	*2013*	**100**	**99.95**	**99.98**
Our App-roach	***SVM***	***MFCC+O_q***	***2013***	***99.8***	***89.86***	***94.28***

using MEII database for normal / pathological classification

In terms of sensitivity, our approach and the one using SVM with the PCA / FCM parameters are the best.

In terms of specificity, our approach is better than the SVM +MS and SVM +LDA systems, but it lags behind the rest.

In terms of accuracy, our approach is more efficient than GMM+12 MFCC, SVM + MS and SVM+ LDA only.

CONCLUSION

This paper presents the evaluation of the behavior of our proposed classification system applied on pathological voices and depending on the parameterisation. Our contribution concerns the addition of a new parameter to the classical parameters formed by MFCC coefficients, the energy, their first and second derivatives and F_0 and the study of the effect of this parameter in the classification performances. Besides, we consider a binary classification between normal and pathological voices for all speakers and a triple classification between normal, edema and nodule voices for female speakers only.

The Open quotient and the fundamental frequency are computed from the glottal opening instant GOI and the glottal closing instant GCI localised by the multi-scale product (MP). The classification is performed by an SVM multiclass system according to one against all approach using the gaussian kernel. The proposed approach is tested on two databases of pathological voices : MAPACI and MEII.

For all these classifications, we vary the set of parameters to investigate the relative effect of the fundamental frequency and the open quotient on the classification rates .

In the MAPACI database, the classification by an SVM two classes normal / pathologic for all patients women and men, using the open quotient with MFCC coefficients provides better rates of 92.6% for normal voices and 85.12% for pathological voices Than MFCC +F_0. These rates are close to the best respective rates of 94.85% and 87.12% obtained with the MFCC coefficients with Δ + ΔΔ.

For the 3-class SVM classification into edema, nodule and normal for women, the best recognition rate is obtained by using all parameters. In addition, the fundamental frequency F_0 appears to be the most discriminant parameter for pathological recognition. The open quotient O_q limits the discrimination between the two diseases edema and nodule. The MFCC coefficients alone are used to recognize normal speakers.

In the MEII database , the classification by SVM two classes normal / pathological for all speakers, the MFCC coefficients significantly improve the classification rate. When adding the O_q or Δ + ΔΔ to MFCC coefficients, the recognition rate remains high. As against, the parameter F0 drops the recognition of patients from 88% to 83%.

For the 3-class SVM classification into edema, nodule and normal for women, the best recognition is obtained with all parameters. The MFCC coefficients allow to recognize 100% of normal speakers, the edema disease to 73.39% and nodule disease to74.85%. The open quotient O_q deteriorates the recognition of the 2 diseases.

Future works concern the classification into between pathologies in the MEII database and testing other parameters extracted from the speech multi-scale product.

REFERENCES

[1] J.I. Godino-Llorente, P. Gomez-Vilda, and T. Lee, "Analysis and Signal Processing of Oesophageal and Pathological Voices", EURASIP Journal on Advances in Signal Processing, Special Issue on Analysis and Signal Processing of Oesophageal and Pathological Voices, 2009.

[2] J.I. Godino-Llorente, P. Gomez-Vilda, N. Saenz-Lechon, M. Blanco-Velasco, F. Cruz Roldan, and M.A. Ferrer, "Discriminative methods for the detection of voice disorders", In: NOLISP 2005 International Conference on Non-Linear Speech Processing, April 2005; Barcelona, Spain.

[3] J.I. Godino-Llorente, R. Fraile, N. Saenz-Lechon, V. Osma-Ruiz, and P. Gomez-Vilda, "Automatic Detection of Voice Impairments from Text-Dependent Running Speech using a Discriminative Approach", In: MAVEBA 2007, pp. 25–28.

[4] M. Hirano, "Psycho-Acoustic Evaluation of Voice: GRBAS Scale for Evaluation the Horse Voice", Springer 1981; Berlin, Germany.

[5] K. Marasek, "An Attempt to classify lx signals", In: EuroSpeech 1995 the 4th European Conference on Speech Communication and Technology; September 1995; Madrid, Spain.

[6] D. Deliyski, "High-speed videoendoscopy: recent progress and clinical prospects", In: AQL 2006 the 7th International Conference on Advances in Quantitative Laryngology Voice and Speech Rearch , 2006; University of Groningen, pp.1-12.

[7] J. Demeyer, and B. Gosslin, "Glottis segmentation with a high speed glottography: a new approach", In Proceedings of Liege Image Days, March 2008; Liege, Belgium .

[8] J.I. Godino-Llorente, P. Gomez Vilda, N. Saenz-Lechon1, M. Blanco-Velasco, F. Cruz-Roldan, and M. Angel Ferrer-Ballester, "Support Vector Machines Applied to the Detection of Voice Disorders", Springer-Verlag, Berlin Heidelberg, 2005. pp.219-230.

[9] J. Iagnacio Godino-Llorente , Member, IEEE, P. Gomez Vilda , Member, IEEE, M. Blanco-Velasco, Member, IEEE, "Dimensionality Reduction of a Pathological Voice Quality Assesment System Based on Gaussian Mixture Models and Short-Term Cepstral Parameters", In: IEEE 2006 Transactions on Biomedical Engeneering, October 2006; 5.

[10] A.A. Dibazar, T.W. Berger, and S.S. Narayanan, "Pathological Voice Assesment", In: IEEE 2006 EMBS 2006; New York.

[11] J. Benesty, M.M. Sondhi, and Y. Huang, "Springer Handbook of Speech Processing", Springer, Berlin, Germany 2008.

[12] F. Servin, B. Bozkurt, and T. Dutoit, "Hnr extraction in voiced speech oriented towards voice quality analysis", In: EUSIPCO 2005 13th European Signal Processing Conference; September 2005; Antalya, Turkey.

[13] G. De Krom, "Spectral correlates of breathiness and roughness for different types of vowel fragments", In:ICSLP 1994 the 3rd International Conference on Spoken Language Processing; September 1994; Yokohama, Japan.

[14]C. D'Alessandro,F. Yegnanarayana, and A. Darsinos, « Decompositions », In: ICASSP 1993 the IEEE International Conference on Acoustics, Speech, and Signal; May 1993; Detroit, Mich, USA: pp.760-763.

[15] P. Boersma, "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound", In: IFA1993 of the Institute of Phonetic Sciences; 1993; Amsterdam, The Netherlands.

[16] K. Shama, A. Krishna, and N.U. Cholayya, "Study of harmonics-to-noise ratio and critical-band energy spectrum of speech as acoustic indicators of laryngeal and voice pathology", EURASIP Journal on Advances in Signal Processing 2007; Article ID 85286, 9.

[17] M. Wester, "Automatic classification of voice quality: comparing regression models and hidden markov models", In: VOICEDATA 1998 Symposium on Databases in Voice Quality Research and Education;1998.

[18]R.B. Reilly, R. Moran, and P. Lacy, "Voice pathology assessment based on a dialogue system and speech analysis", In: AAAI 2004 Symposium on Dialogue Systems for Health Communication; 2004; pp. 104–109.

[19] R.J. Moran, R.B. Reilly, P. de Chazal, and P.D. Lacy, "Telephony-based voice pathology assessment using automated speech analysis", IEEE Transactions on Biomedical Engineering; 2006; 3:468–477.

[20] Corp KE. Multi-dimensional voice program (mdvp) [computer program]. Tech Rep, Kay Elemetrics Corp, 2008.

[21] Corp K. E. Disordered voice database model (version 1.03). Tech Rep, Massachussets Voice Eye and Ear Infirmary Voice and Speech Lab, 1994.

[22] A. Dibazar, S. Narayanan, "A system for automaticdetection of pathological speech", In: 36th Asilomar 2002 Conference on Signals, Systems and Computers; November 2002; Pacific Grove, Calif, USA.

[23] J.I. Godino-Llorente, S. Aguilera-Navarro, C. Hernandez- Espinosa, M. Fernandez-Redondo, and P. Gomez-Vilda, "On the selection of meaningful speech parameters used by a pathologic/non pathologic voice register classifier", In: EUROSPEECH 1999 the 6th European Conference on Speech Communication and Technology; September 1999; Budapest, Hungary.

[24]C. Fredouille, G. Pouchoulin, J.F. Bonastre, M. Azzarello, A. Giovanni, and A. Ghio, "Application of automatic speaker recognition techniques to pathological voice assessment (dysphonia)", In: EuroSpeech 2005, the 9th European Conference on Speech Communication and Technology; September 2005; Lisbon, Portugal, pp. 149–152.

[25]G. Pouchoulin, C. Fredouille, J. Bonastre , A. Ghio, M. Azzarello, and A. Giovanni, «Modélisation statistique et informations pertinentes pour la caractérisation des voix pathologiques (dysphonies) », In : JEP 2006 (Journée d'Etudes sur la Parole); 2006.

[26]C. Bishop, "Pattern Recognition and Machine Learning", Springer, New York, NY, USA, 2006.

[27]P. Kukharchik, I. Kheidorov, E. Bovbel, and D. Ladeev, Image and signal processing, In: Speech Signal Processing Based on Wavelets and SVM for Vocal Tract Pathology Detection, Lecture Notes in Computer Science Springer; Berlin, Germany, pp. 192–199, 2008.

[28]J.I. Godino-Llorente, S. Aguilera-Navarro, C. Hernandez-Espinosa, M. Fernandez-Redondo, and P. Gomez-Vilda, "On the selection of meaningful speech parameters used by a pathologic/non pathologic voice register classifier". In: EUROSPEECH 1999 the 6th European Conference on Speech Communication and Technology; September 1999; Budapest, Hungary.

[29]J.B. Alonso, J. de Leon, I. Alonso, and M.A. Ferrer, "Automatic detection of pathologies in the voice by HOS based parameters". EURASIP Journal on Advances in Signal Processing 2001; 4: pp. 275–284.

[30]Kay Elemetrics Inc. Voice disorders database, version 1.03[CDROM][Online].Available:http://www.kaypentax.com/Product%20Info/CSL%20Options/4337/4337.htm.

[31] MAPACI P. Voice Disorder Database [Online]. Available: http:// www. Mapaci.com/index-ingles.php

[32]http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html

[33]A. Belhaj, A. Bouzid, N. Ellouze, and A. Nait-ali, « Paramétrisation des voix pathologiques à partir du MPM et leurs classifications », Quatrièmes journées de phonétique clinique ;2011 ; Strasbourg, France.

[34]A. Belhaj, A. Bouzid, N. Ellouze, and A. Nait-ali, "Disordered voice parametrisation using the multi-scale product", In : SETIT 2012Sciences of Electronics, Technologies of Information and Telecommunications; Mars 2012 ;Sousse, Tunis.

[35]A. Belhaj, A. Bouzid, N. Ellouze, "Statistical voicing parameter analysis of pathological signals using the Multi-scale Product and SVM classification", In: ATSIP 2014 International Conference on Advanced Technologies for Signals & Image processing; Mars 2014; Sousse,Tunis.

[36]S. Chekili, A. Belhaj, A. Bouzid, N. Ellouze, "Recognition of pathological voices", In: IEEE 2014 International Multi-Conference on Systems, Signals &

Devices, Conference on Communication & Signal Processing; Février 2014; Barcelona, Espagne.

[37]V. Vapnik, "An overview of statistical learning theory", In: IEEE 1999 Transactions on Neural Networks; September 1999; IEEE. pp. 988-1000.

[38] http://asi.insarouen.fr/enseignants/~arakoto/toolbox/index.html.

[39] C.W. Hsu, and C.J. Lin, "A comparison of methods for multi-class support vector machine", In: IEEE 2002Transactions on Neural Networks; 2002; IEEE. pp.415-425.

[40]J. Friedman, "Another Approach to Polychotomous Classification", Technical report, Department of Statistics, Stanford University, 1996.

[41] M. Hariharan, K. Polatb, R. Sindhuc and S. Yaacoba, "A hybrid expert system approach for telemonitoring of vocal fold pathology", Applied Soft Computing 13, pp. 4148–4161, 2013.

[42] J. D. Arias-Londono, J. I. Godino-Llorente, M. Markaki, Y. Stylianou and Y. Stylianou, "On combining information from modulation spectra and mel-frequency cepstral coefficients for automatic detection of pathological voices", Logopedics Phoniatics Vocoly, 2010.

[43] A. Alpan, J. Schoentgen, Y. Maryn and F. Grenez, "Automatic Perceptual Categorization of Disordered Connected Speech", In Proceedings of Interspeech, pp. 2574–257, 2010.

.