Voice Pathology Recognition and Classification using Noise Related Features

HAMDI Rabeh[1], HAJJI Salah[2], CHERIF Adnane[3]

Faculty of Mathematical, Physical and Natural Sciences of Tunis, University of Tunis El Manar, Tunisia
Laboratory of Analysis and Processing of Signals And Electric and Energy System, 1060 Tunis, Tunisia

Abstract—**Nowadays, the diseases of the voice increase because of bad social habits and the misuse of voice. These pathologies should be treated from the beginning. Indeed, it is no longer necessary that the diseases of the voice lead to affect the quality of the voice as heard by a listener. The most useful tool for diagnosing such diseases is the Acoustic analysis. We present in this work, new expression parameters in order to clarify the description of the vocal signal. These parameters help to classify the unhealthy voices. They describe essentially the fundamental frequency F0, the Harmonics-to-Noise report (HNR), the report Noise to Harmonics Ratio (NHR) and Detrended Fluctuation Analysis (DFA). The classification is performed on two Saarbruecken Voice and MEEI pathological databases using HTK classifiers. We can classify them into two different types: the first classification is binary which is used for the normal and pathological voices; the second one is called a four-category classification used in spasmodic, polyp, nodule and normal female voices and male speakers. And we studied the effects of these new parameters when combined with the MFCC, Delta, Delta second and Energy coefficients.**

Keywords— HTK; MFCC; MEEI; SVD; pathological voices

I. INTRODUCTION

Many pathologies, may affect the voice as nodules, spasmodic folds, polypoid causing irregular vibrations due to the malfunction of many factors that contribute to vocal vibrations. Beside, pathologies of the voice may affect differently the vibration of the vocal field, first of all it depends on the type of disorder, but also the location of the disease in the folds of the voice, so it allows them to produce different shades of base [15].

To tackle those problems of the voice, digital processing on voice signals is a found tool that helps with nonvasive analysis for doctors. It allows identification of vocal disorders especially from the beginning [16].

The disease affecting more people is the dysphonia because of the disruption of the speech. There are various types of dysphonia. First, the dysfunctional dysphonia is characterized in some obstacles of pronunciation but without changing the organic composition of the vocal cords. The dysfunctional dysphonia can lead to organic dysphonia because of the application of compensation by the patient. Second the organic dysphonia is a pathological change in the vocal cords. Third, we note neurological dysphonia. To evaluate and determine the therapy, the evaluation of the voices is very relevant. The quality of voice could be assessed by diagnosis or by the laryngostroboscopy testing as. Two different approaches are involved: the perception and the objective approach.

On the other hand, to establish the subjective measurement of voice quality it should be based on the individual experience. The subjective measurement may vary. The method detection of automatic voice-pathology can be accomplished by various types of signal analysis which can be long term or short-term. These parameters can be determined using cepstral coefficients with Mel frequency [13] [14], linear predictive cepstral coefficients (LPCC) [12], and so on. The old research presented different tools to establish an evaluation. Obviously, in the related work many methods of acoustic diagnosis of pathological voices have been proposed. Between them, a big attention was given to the automatic classification of the troubled voice. For classification of pathological voices there are very important classifiers such as: hidden Markov model (HMM) [19] as well as the neural networks [17], the support vector machines (SVM) [20] and finally the Gaussian mixing model (GMM) [18].

A normal binary / pathological classification of vocal samples [1, 2] has been proposed in the literature, the best performances are obtained by using specific parameters of the HMM classification. However, few studies that have classified the pathologies [3] and the obtained results were not effective.

In the present work, the classification of pathological voices is studied using the method of extraction of the parameters MFCC with energy, derivative and acceleration combined with the prosodic parameters, noise-to-harmonic ratio (NHR), harmonic-to-noise ratio (HNR), Detrended Fluctuation Analysis (DFA) and fundamental frequency (F0) which are calculated for each frame. To validate this work we used two bases to give MEEI Database and Saarbruecken Voice Database. The aim of this work is to show the ability of these parameters to detect and classify pathologies of the voice, using a scenario where these parameters are used alone with MFCC and hybrid.

II. METHODS AND MATERIALS

The classical characteristics are derived from the benchmarks used in the domain of acoustic recognition. These parameters are essentially the analysis of trend fluctuations (DFA), the harmonic/noise ratio (HNR), the fundamental frequency F0 and the Cepstral coefficient of the frequency Mel (MFCC) combined with the energy and the first and second derivatives.

The characteristics involved in the pathological voice which are the most common are described in the section below.

A. Fundamental Frequency

The fundamental frequency (F0) is one of the essential parameters in acoustic measurement. This frequency expresses the vibration rate of the voice fold. This setting describes the voice state. It is sometimes used with the Mel-Frequency (MFCC) Cepstral coefficients in the form of conjunction.

B. Mel-Frequency Cepstral Coefficients

The parameter MFCCs is used to decrease the voice signal redundancy. It is also used in other areas such as voice recognition [4]. The calculation of these coefficients is done by using the method of weighting the signal Fourier transform through a bank of filters distributed on a "Mel" scale, then from this weighted spectrum by calculating the cepstrum and at the end calculate the discrete cosine transformation for this cepstrum.

MFCC belongs to a family of parameters that are used in speech processing. Based on the human knowledge of the sounds, MFCC does a frequency analysis of the signal. By listening to the signal an experienced therapist can detect the presence of a speech disorder [2]. For each frame, the extraction procedure is done after a 16 kHz interpolation, with a bank of 29 Mel filters and a 25 ms with a 10 ms step, to get 12 MFCC plus log-energy, Delta and Delta seconds.

C. Noise to Harmonics Ratio and Harmonics to Noise Ratio

The harmonic-noise ratio HNR measures objectively the feeling of perception in a hoarse voice [5]. The calculation of the harmonic-noise ratio, the signal must be dropped sampled at 16 KHz, and divided into 25 ms length of the frames, with a step of 10 Ms. The comb filter is applied to the signal in each frame, in order to calculate the energy in the components Harmonics. For the logarithm of this quantity, the logarithmic energy of the noise is inferred to obtain the HNR.

D. Detrended Fluctuation Analysis (DFA)

Detrended Fluctuation Analysis characterizes the extent of turbulent noise in the speech signal, quantifying the stochastic self-similarity of the noise caused by turbulent air flow in the vocal tract, e.g. incomplete vocal fold closure can increase the DFA value. It is applied to parole signals, shows the ability to detect voice disorders in general. [6]

III. THE RESULTS

A. Databases

In all this work we use two different databases MEEI data base and Saarbruecken database. In the first data base the voices samples are based mainly on the phonation of vowels [a] whose duration is about 3 or 4 s by men and women. And in the second data base we used the recording of the phrase "Guten Morgen, wie geht es Ihnen?" (" Good morning, how are you?").

Table 2 gives the number of samples of the pathological voices of each base.

TABLE I.   AN OVERVIEW OF THE MEASURES OF DYSPHONIA USED IN THIS STUDY

Measure	Motivation	Number of features
Mel Frequency Cepstral Coefficients (MFCC)	The vocal pathologies cause a drop of the control of the articulators (vocal music channels). and MFCCs are trying to analyze the voice conduit regardless of the vocal cords	13-26-39, depends on extracted components and the use of delta and delta-delta coefficients
Harmonics-to-Noise Ratio (HNR) and Noise to Harmonics Ratio (NHR)	In the pathologies of the voice, the incomplete closure of the vocal fold cause an increase of the noise due to turbulent air flow. MST and NHR measure information report real signal to noise.	2
Fundamental Frequency (F0)	Average vocal fundamental frequency	1
Detrended Fluctuation Analysis (DFA).	Quantify the stochastic self-similarity of the noise caused by turbulent airflow	1

1) MEEI Database

MEEI-KayPENTAX is the database that was invented by the Massachusetts Eye and Ear Infirmary Voice and Speech Labs Corp and was published in 1994. The recordings are manifested in a sustained phonation of the vowel/Ah/(53 normal and 657 pathological) and the statement of the first sentence of the Rainbow passage (53 normal and 662 pathological). They are pronounced by patients who have these types of diseases like neurological, organic, traumatic and psychogenic speech disorders from the beginning of the disease to the complete elaboration. The recording environment of speech samples has the following characteristics 16 bit of resolution and the sampling frequency is about 25 khz or 50 khz. We chose a subset of voices which comprise 53 normal and 60 pathological [7].

2) Saarbruecken Voice Database

Recently the Saarbruecken Voice Database was published online [8]. This database is a collection of voice recordings of more than 2000 people; a recording session contains the following recordings:

- recordings of vowels /a/, /i/, /u/ produced at normal, high, low and low-high-low pitch.

- recording of sentence" Guten Morgen, wie geht es Ihnen?" (" Good morning, how are you?").

Each session contains 13 registration files. Moreover, for each case of the electroglottogram signal (EGG) is saved in a separate folder. These files contains vowels whose length is 1-3 seconds. All recordings are made in a controlled environment at 50 kHz and their resolution is 16-bit. These recordings contain 71 different pathologies, including organic and functional. 1320 pathological voices are divided into 609 males

and 711 females. In other hand there are 650 normal voices (400 males and 250 females). We worked with a subset of voices which comprise 133 normal and 134 pathological.

3) Hidden Markov Model Toolkit

The Hidden Markov Model Toolkit (HTK 3.4.1) is a portable toolkit for building and manipulating hidden Markov models. HTK is primarily used for speech recognition research although it has been used for numerous other applications including research into speech synthesis, pathological voice recognition [9]. A hidden Markov model with a Gaussian mixing density (HMM-GM), five observation states (a simple model from left to right) and four diagonal state mixtures were formed for each pathological voice. [10]

TABLE II.   THE DIFFERENT PATHOLOGICAL VOICES FOR THE TWO BASES

Diseases	MEEI Database		Saarbruecken Voice Database	
	Test	Train	Test	Train
Nodules	6	14	5	20
Spasmodic	6	12	21	43
Polyploid	7	15	15	30
Normal	17	36	45	88

We have developed a parametrization method that extracts the MFCC coefficients with energy, derivative and concatenated acceleration with the parameters that measure the disturbance of the vocal signal (prosodic parameters such as: F0, HNR, NHR, DFA). These parameters are calculated for each frame.

B. Global Rate Recognition for MEEI Database

Table 3 below gives the results of the rate of recognition of pathological voices for the MEEI database. For the first database, acoustic modeling is refined, estimating four-Gaussian probability densities. The recognition having the best rates are respectively obtained MFCC with all the parameters (94.44%), MFCC_NHR_DFA (91.67%) and MFCC_DFA (88.89%).

TABLE III.   GLOBAL RECOGNITION RATE OF PATHOLOGIES FOR FOUR GAUSSIANS OF MEEI DATABASE

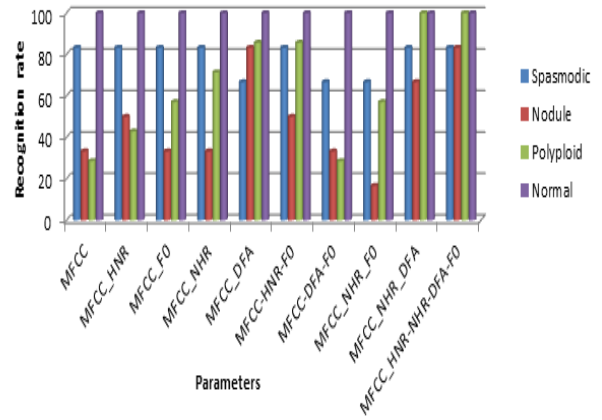| Parameters	Recognition rate
MFCC	72.22
MFCC_HNR	77.78
MFCC_F0	77.78
MFCC_NHR	80.56
MFCC_DFA	**88.89**
MFCC-HNR-F0	86.11
MFCC-DFA-F0	77.78
MFCC_NHR_F0	72.22
MFCC_NHR_DFA	**91.67**
MFCC_HNR-NHR-DFA-F0	**94.44**

Fig. 1.   Recognition rate by pathology of different techniques for 4 Gaussian of the MEEI Database.

The combination of Harmonics to Noise Ratio, Noise to Harmonics Ratio, Detrended Fluctuation Analysis and the fundamental frequency parameters has the ability to recognize the voice disease while, the Noise to Harmonics Ratio combined with the fundamental frequency shows the disability to recognize the diseases. Moreover, it appears that with the MFCC coefficients the recognition of normal voice is with high rate.

C. Global Rate Recognition for Saarbruecken Voice Database

Table 4 below gives the results of the rate of recognition of the pathological voices for the Saarbruecken Voice Database. For the second database, acoustic modeling is refined, estimating two-Gaussian probability densities. The best results are obtained when using the parameters MFCC_NHR_DFA (94.19%) and MFCC_NHR (91.86%).

The Noise to Harmonics Ratio combined with Detrended Fluctuation Analysis appear that this combination was the most able of knowing and distinguishing the types of pathologies. While, the MFCC is not able to distinguish pathological voices.

TABLE IV.   GLOBAL RECOGNITION RATE OF PATHOLOGIES FOR TWO GAUSSIANS OF SAARBRUECKEN VOICE DATABASE

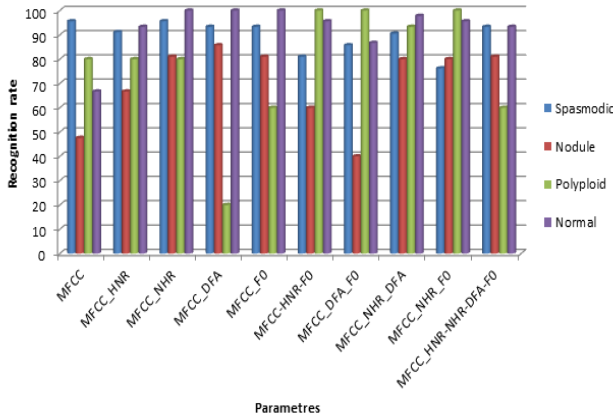| Parameters	Recognition rate
MFCC	77.91
MFCC_HNR	84.88
MFCC_NHR	**91.86**
MFCC_DFA	88.37
MFCC_F0	89.53
MFCC-HNR-F0	90.70
MFCC_NHR_F0	90.70
MFCC_DFA_F0	86.05
MFCC_NHR_DFA	**94.19**
MFCC_HNR-NHR-DFA-F0	88.37

Fig. 2. Recognition rate by pathology of different techniques for 4 Gaussian of Saarbruecken Voice Database.

D. Pathologic/Normal Classification.

The results of the different experiments used for the classification and the detection of pathology are expressed by

These terms:

- Accuracy(ACC): It is the ratio of the correctly detected samples by the total number of samples used.

- Sensitivity(SN): It represents the proportion of pathological samples correctly identified.

- Specificity (SP): It is the proportion of normal samples that are negatively identified.

The following distinct equations show how to calculate these terms:

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \qquad (1)$$

$$SN = \frac{TP}{TP+FN} \qquad (2)$$

$$SP = \frac{TN}{TN+FP} \qquad (3)$$

The expression where true negative (TN) can be explained as follows: the system detects a normal subject as a normal subject, while the true positive (TP) means that the system detects a pathological subject as a pathological subject, besides the false negative (FN) means that the system detects a pathological issue as a normal subject and ultimately false positives (FP) means that the system detects a normal subject matter as a pathological subject. [11]

The extracted parameters from the two different databases must be checked in the detection and classification processes.

The experimental analysis shows that the data obtained varied between the databases and varied according to the types of parameters (HNR, NHR, DFA, F0 and their combination) in the same database.

The two tables 5 and 7 represent the Confusing matrix for the normal / pathological classification respectively of the MEEI and SVD databases.

Tables 6 and 8 give the results of the sensitivity, specificity and accuracy calculations of the different combinations of parameters MFCC, HNR, NHR, F0 and DFA; for normal / pathological classification of MEEI database and Saarbruecken Voice Database, respectively. These results are deduced using tables 5 and 7 and equations (1), (2) and (3).

These tables indicate the best precisions obtained for each database with the different types of parameters. Its show that the accuracy varied from one database to another for the same used characteristic, in other hand the accuracies obtained also varied for the same database as a function of parameters used to carry out the experiment.

Generally, the highest achieved accuracies are 100% for MEEI Voice Database, in the case of using the MFCC_HNR, MFCC_F0, MFCC_DFA and MFCC_F0. While using the MFCC_HNR we get the highest acquired accuracies which is 100% for Saarbruecken Voice Database.

TABLE V.   CONFUSING MATRIX OF THE NORMAL / PATHOLOGIC CLASSIFICATION USING ALL VOICES OF THE MEEI DATABASE

Parameters		Normal	Pathologic
MFCC	Normal	**100**	0
	Pathologic	5.3	**94.7**
MFCC_HNR	Normal	**100**	0
	Pathologic	0	**100**
MFCC_F0	Normal	**100**	0
	Pathologic	0	**100**
MFCC_NHR	Normal	94.1	5.9
	Pathologic	5.3	94.7
MFCC_DFA	Normal	**100**	0
	Pathologic	0	**100**
MFCC-HNR-F0	Normal	94.1	5.9
	Pathologic	5.3	94.7
MFCC-DFA-F0	Normal	**100**	0
	Pathologic	0	**100**
MFCC_NHR_F0	Normal	**100**	0
	Pathologic	5.3	94.7
MFCC_NHR_DFA	Normal	94.1	5.9
	Pathologic	5.3	94.7
MFCC_HNR-NHR-DFA-F0	Normal	94.1	5.9
	Pathologic	5.3	94.7

TABLE VI.   BEST DETECTION ACCURACIES IN THE MEEI DATABASE USING VARIOUS PARAMETERS

Parameters	Sensitivity	Specificity	Accuracy
MFCC	**100**	94.97	95.20
MFCC_HNR	**100**	**100**	**100**
MFCC_F0	**100**	**100**	**100**
MFCC_NHR	94.14	94.67	95.17
MFCC_DFA	**100**	**100**	**100**
MFCC-HNR-F0	94.14	94.67	95.17
MFCC-DFA-F0	**100**	**100**	**100**
MFCC_NHR_F0	**100**	94.97	95.20
MFCC_NHR_DFA	94.14	94.67	95.17
MFCC_HNR-NHR-DFA-F0	94.14	94.67	95.17

TABLE VIII.   BEST DETECTION ACCURACIES IN THE SAARBRUECKEN VOICE DATABASE USING VARIOUS PARAMETERS

Parameters	Sensitivity	Specificity	Accuracy
MFCC	97.82	98.89	99.39
MFCC_HNR	97.85	**100**	**100**
MFCC_F0	97.82	98.89	99.39
MFCC_NHR	97.57	89.48	88.89
MFCC_DFA	**100**	98.91	99.40
MFCC-HNR-F0	97.82	98.89	99.39
MFCC-DFA-F0	97.82	98.89	99.39
MFCC_NHR_F0	**100**	89.69	89
MFCC_NHR_DFA	97.51	87.63	86.69
MFCC_HNR-NHR-DFA-F0	95.08	86.52	85.58

For the MEEI database, we obtained a total recognition of normal and pathological samples in each case (TN = 100% and TP = 100%) for the combinations MFCC_HNR, MFCC_F0, MFCC_DFA and MFCC_DFA_F0.

For the other combinations the recognition is not perfect in each case for the different types of voice samples.

While for Saarbruecken Voice Database we did not get total recognition of normal and pathological samples in each case, but all normal type samples are recognized as normal for MFCC_DFA and MFCC_NHR_F0. While pathological voices are recognized as pathological voices for MFCC_HNR only.

TABLE VII.   CONFUSING MATRIX OF THE NORMAL / PATHOLOGIC CLASSIFICATION USING ALL VOICES OF THE SAARBRUECKEN VOICE DATABASE

Parameters			Normal	Pathologic
MFCC		Normal	**97.8**	2.2
		Pathologic	1.1	**98.9**
MFCC_HNR		Normal	97.8	2.2
		Pathologic	0	**100**
MFCC_F0		Normal	97.8	2.2
		Pathologic	1.1	98.9
MFCC_NHR		Normal	97.8	2.2
		Pathologic	11.5	88.5
MFCC_DFA		Normal	**100**	0
		Pathologic	1.1	98.9
MFCC-HNR-F0		Normal	97.8	2.2
		Pathologic	1.1	98.9
MFCC-DFA-F0		Normal	97.8	2.2
		Pathologic	1.1	98.9
MFCC_NHR_F0		Normal	**100**	0
		Pathologic	11.5	88.5
MFCC_NHR_DFA		Normal	97.8	2.2
		Pathologic	13.8	86.2
MFCC_HNR-NHR-DFA-F0		Normal	95.6	4.4
		Pathologic	14.9	85.1

IV.   DISCUSSIONS AND VALIDATION

In this study we used the parameters that measure the disturbance of the vocal signal, in two databases for the detection and classification of vocal pathologies.

Indeed, the obtained results are better or comparable than the other results reported using the MEEI and SVD databases.

TABLE IX.   COMPARISON OF ACCURACIES BETWEEN METHODS (PATHOLOGY DETECTION)

Method	*MEEI*	*SVD*
In this paper	100%	100%
Method [11]	99.96%	92.79%
Method [15]	88.21%	99.68%
Method [21]	94.07%	-
Method [22]	94.80%	81%

For example, in al-Nasheri et al. [15] used the MEEI and SVD databases with the SVM classifier and used the MDVP parameters to detect pathologies, obtaining accuracies of 99.68% and 88.21% respectively for SVD and MEEI.

In addition, al-Nasheri et al. [11] used the SVD and MEEI databases and used the autocorrelation and entropy parameters for the detection and classification of pathologies, obtaining respectively 99.96% and 92.79% accuracy for MEEI and SVD.

Thus Godino_Liorente et al. [21] used the MEEI database and reported an accuracy of 94.07%. While Marinez et al. [22] used the SVD database and the SVM classifier and achieved an accuracy of 81%. The authors also used the MEEI database and the accuracy gained was 94.80%.

In our study, the accuracy obtained in the case of SVD and MEEI is better at the accuracy obtained in other cases.

Table 9 illustrates the comparison between our contribution and other contributions mentioned in the related work using both bases MEEI and SVD. In our work, we obtained a high 100% accuracy for detection.

V. CONCLUSION

In this study, we presented our approach which is manifested in the addition of new and classical parameters to each other .Also we presented the study of the effect of the classical parameters formed by the MFCC coefficients, the energy, their first and second derivatives in the classification performances. In addition, we classified all speakers who have pathological and normal voices in binary classification.

Our contribution is tested on two pathological voice databases: SVD and MEEI only

The acoustic modeling is refined, estimating the probability densities respectively at four Gaussian for the first database MEEI and at two Gaussian for the second database SVD. The best recognition rates of the MEEI database are respectively obtained MFCC with all the parameters (94.44%), MFCC_NHR_DFA (91.67%) and MFCC_DFA (88.89%) while for the SVD base using the parameters MFCC_NHR_DFA (94.19%) and MFCC_NHR (91.86%) we have obtained the best result of recognition rates.

For the normal / pathological classification of MEEI database and Saarbruecken Voice Database, respectively.

Generally, the highest achieved accuracies are 100% for MEEI Voice Database, in the case of using the MFCC_HNR, MFCC_F0, MFCC_DFA and MFCC_F0. While the values found are the highest in the case of using MFCC_HNR is 100% for Saarbruecken Voice Database.

REFERENCES

[1] 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, pp.219-230, 2005.

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

[3] 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 5, 2006.

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

[5] Teixeira, João Paulo, Carla Oliveira, and Carla Lopes. "Vocal acoustic analysis–jitter, shimmer and HNR parameters." Procedia Technology 9: 1112-1122. 2013

[6] Little MA, McSharry PE, Roberts SJ, Costello DA, Moroz IM. "Exploiting Nonlinear recurrence and Fractal scaling properties for voice disorder detection". Biomed Eng Online. 2007 Jun 26

[7] MEEI Database, Disorder database model 4337. Massachusetts Eye and Ear Infirmary Voice and Speech Lab, Boston, MA. 2002.

[8] W. J. Barry and M. Putzer, "Saarbruecken Voice Database", Institute of Phonetics, Univ. of Saarland, http://www.stimmdatenbank.coli.uni-saarland.de/.

[9] http://htk.eng.cam.ac.uk/

[10] Young, S., G. Evermann, M. Gales, T. Hain and D. Kershaw et al., 2009. The HTK book version 3.4.1. Cambridge University Engineering Department, Cambridge, U.K.

[11] A. Al-Nasheri et al. "Voice Pathology Detection and Classification Using Auto-Correlation and Entropy Features in Different Frequency Regions". IEEE volume 6, March 9, 2018.

[12] Anusuya MA, Katti SK. "Front end analysis of speech recognition": a review. Int J Speech Technol,14:99–145. 2010

[13] Muda, Lindasalwa, Mumtaj Begam, and Irraivan Elamvazuthi. "Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques." arXiv preprint arXiv:1003.4083 2010.

[14] Ali Z, Aslam M, Ana María ME. "A speaker identification system using MFCC features with VQ technique". Proceedings of 3rd IEEE International Symposium on Intelligent Information Technology Application, pp. 115–119, 2009.

[15] Al-nasheri, A., Muhammad, G., Alsulaiman, M., Ali, Z., Mesallam, T. A., Farahat, M., ... & Bencherif, M. A. "An investigation of multidimensional voice program parameters in three different databases for voice pathology detection and classification". Journal of Voice, 31(1), 113-e9. 2017.

[16] Al-Nasheri, A., Muhammad, G., Alsulaiman, M., Ali, Z., Malki, K. H., Mesallam, T. A., & Ibrahim, M. F. Voice "pathology detection and classification using auto-correlation and entropy features in different frequency regions". IEEE Access, 6, 6961-6974. 2018.

[17] Ritchings, R. T., McGillion, M., & Moore, C. J. "Pathological voice quality assessment using artificial neural networks". Medical engineering & physics, 24(7-8), 561-564. 2002.

[18] Godino-Llorente, J. I., Gomez-Vilda, P., & Blanco-Velasco, M. "Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters". IEEE transactions on biomedical engineering, 53(10), 1943-1953. 2006.

[19] Daniel Jurafsky & James H. Martin. "Hidden Markov Models" Speech and Language Processing chapter 9. Draft of August 7, 2017.

[20] Abe S. "Support Vector Machines for Pattern Classification". Berlin, Heidelberg, New York, NY: Springer-Verlag; 2005.

[21] Godino-Llorente JI, Gomes-Vilda P, Blanco-Velasco M. "Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters". IEEE Trans Biomed Eng. 53:1943–1953. 2006.

[22] D. Martínez, E. Lleida, A. Ortega, A. Miguel, and J. Villalba, ''Voice pathology detection on the Saarbrücken voice database with calibration and fusion of scores using multifocal toolkit, in Advances in Speech and Language Technologies for Iberian Languages". Berlin, Germany: Springer, pp. 99–109 , 2012