

## Improved Algorithm for Pathological and Normal Voices Identification

Brahim Sabir<sup>1</sup>, Fatima Rouda<sup>2</sup>, Yassine Khazri<sup>3</sup>, Bouzekri Touri<sup>4</sup>, Mohamed Moussetad<sup>5</sup>

<sup>1,3,5</sup>Physics Department, Faculty of Science Ben M'Sik –Casablanca, Morocco

<sup>2</sup>Chemistry Department, Faculty of Science Ben M'Sik –Casablanca, Morocco

<sup>4</sup>Language and Communication Department, Faculty of Science Ben M'Sik –Casablanca, Morocco

---

### Article Info

#### Article history:

Received Mar 4, 2016

Revised May 18, 2016

Accepted Jun 4, 2016

---

#### Keyword:

Classification methods

Communication disorders

Neural networks

Voice disorders

---

### ABSTRACT

There are a lot of papers on automatic classification between normal and pathological voices, but they have the lack in the degree of severity estimation of the identified voice disorders. Building a model of pathological and normal voices identification, that can also evaluate the degree of severity of the identified voice disorders among students. In the present work, we present an automatic classifier using acoustical measurements on registered sustained vowels /a/ and pattern recognition tools based on neural networks. The training set was done by classifying students' recorded voices based on threshold from the literature. We retrieve the pitch, jitter, shimmer and harmonic-to-noise ratio values of the speech utterance /a/, which constitute the input vector of the neural network. The degree of severity is estimated to evaluate how the parameters are far from the standard values based on the percent of normal and pathological values. In this work, the base data used for testing the proposed algorithm of the neural network is formed by healthy and pathological voices from German database of voice disorders. The performance of the proposed algorithm is evaluated in a term of the accuracy (97.9%), sensitivity (1.6%), and specificity (95.1%). The classification rate is 90% for normal class and 95% for pathological class.

Copyright © 2017 Institute of Advanced Engineering and Science.  
All rights reserved.

---

### Corresponding Author:

Brahim Sabir,  
Departement of Physics,  
Faculty of Science Ben M'Sik,  
Casablanca, Morocco.  
Email: sabir.brahim@hotmail.com

---

## 1. INTRODUCTION

In academic field, among others, voice is considered the main tool of human communication, and it has been shown to carry much information about the general health and well-being of a student. Thus, voice disorders cause significant changes in speech and impact particularly student academic results and his overall activities. Medical methods to assess these voice disorders are either by inspection of vocal folds or by a physician's direct audition [1], which is subjective, based on perceptual analysis and both depend on physician's experiences.

Acoustic analysis based on instrumental evaluation which comprises acoustic and aerodynamic measure of normal and pathological voices have become increasingly interesting to researchers because of its nonintrusive nature and its potential providing quantitative data with reasonable analysis time. Speech disorder assessment can be made by a comparative analysis between pathological acoustic patterns and the normal acoustic patterns saved in a database [2]. In voice processing we distinguish three principal approaches: acoustic, parametric and nonparametric approach and statistical methods.

The first approach consist to compare acoustics parameters between normal and abnormal voices such as fundamental frequency, jitter, shimmer, harmonic to noise ratio, intensity [3]. The second approach is

the parametric and non-parametric for features selection [4], [5]. The classification of voice pathology can be seen as pattern recognition so statistical methods are an important approach.

This paper is organized as a follow: in second Section is dedicated to relevant acoustic parameters that differentiate pathological from normal voices, the classification methods are presented in Section 3. The degree of severity in Section 4, Section 5 deals with the defined threshold in literature to differentiate pathological and normal voices. The proposed method is presented in Section 5, the results in Section 6, and the last Section is reserved for the conclusion and future work.

### 1.1. Relevant acoustic parameters

Analysis of voice signal is performed by the extraction of acoustic parameters using digital signal processing techniques [6]. However the amount of these parameters is huge to be analyzed, which lead to define the relevant ones. Many papers identify HNR (harmonic-to-noise ratio), Pitch, shimmer as relevant [7], [8] to identify pathological voices among normal ones.

Other papers [9], [10] found that jitter, normalized autocorrelation of the residual signal in pitch lag, shimmer and noise measures like HNR (Harmonic to Noise Ratio), are used to identify pathological voices. Amplitude perturbation, voice break analysis, subharmonic analysis, and noise related analysis are the most relevant ones in [9].

And in [11], pitch, jitter, shimmer, Amplitude Perturbation Quotient (APQ), Pitch Perturbation Quotient (PPQ), Harmonics to Noise Ratio (HNR), pitch perturbation quotient (PPQ), Normalized Noise Energy (NNE), Voice Turbulence Index (VTI), Soft Phonation Index (SPI), Frequency Amplitude Tremor (FATR), Glottal to Noise Excitation (GNE), have been seen as relevant ones. There are also many works that tested the combination of mentioned features [12].

### 1.2. Classification methods

Various pattern classification methods have been used such as : Gaussian Mixture Models(GMM) and artificial neural network (ANN) [13], [14], Bidirectional Neural Network (BNN) [15], multilayer perceptron (MLP) which achieved a classification rate of 96% [16], support vector machine (SVM) [17-19], Genetic Algorithm (GA) [20], [21], method based on hidden markov model [22], [23]. The use of modulation spectra, classification based on multilayer network [8].

The correct classification rate obtained in previous researches to distinguish between pathological and healthy voices varies significantly: 89.1% [24], 91.8% [25], 99.44% [26], 90.1%, 85.3% and 88.2% [27]. However, the comparison among the researches carried out is very complex due to the wide range of measures, data sets and classifiers employed. Authors reported detection accuracy from 80% to 99%. The results depend on efficiency of methods, choice of classifiers and characteristics of databases. The best classification was obtained using nine acoustic measures and achieving an accuracy of 96.5 % [28].

### 1.3. Degree of severity

Actually, the hard part of pathological voice detection is to discriminate light or moderate pathological voices from normal subjects. The degree according to the G parameter of the GRBAS scale proposed by [29]. On this G-based scale, a normal voice is rated as 0, a slight voice disorder as 1, a moderate voice disorder as 2 and finally, a severe voice disorder as 3.

## 2. THRESHOLD IN LITERATURE

Based on the literature, we have considered: pitch, jitter, shimmer and harmonic-to-noise ratio (HNR) as relevant parameters to identify the pathological voices. In order to classify initially the recorded utterances, we are based on threshold defined in the literature as listed in Table 1 and Table 2.

Table 1. Recommended Values of Pitches for Male and Female Signals ([7])

	Mean Pitch	Minimum Pitch	Maximum Pitch
Female signal recommended value	225 Hz for adult females,	155 for adult females,	334Hz for adult ,
Male signal recommended value	adult males 128 Hz ,	adult males 85 Hz ,	adult males 196 Hz,

Table 2. Recommended Values to differentiate Pathological and Normal Voices

		Praat [8]	Teixeira [8]
Jitter ddp%	Female signal recommended value	$\leq 1.04\%$	$\leq 0.66$
	Male signal recommended value	$\leq 1.04\%$	$\leq 0.44$
Shimmer dda%	Female signal recommended value	$\leq 3.810\%$	$\leq 2.43$
	Male signal recommended value	$\leq 3.810\%$	$\leq 2.01\%$
HNR (dB)	Female signal recommended value	$< 20$ dB	15.3dB
	Male signal recommended value	$< 20$ dB	17.3 dB

### 3. PROPOSED METHOD

The corpus used in this paper is composed of 50 voices of male and 50 voices of females aged 19 to 22 (mean: 20.2). The speech material is obtained by sustained vowel /a/ varies from 3 to 5 seconds (mean 3s). Among the 50 voices, 25 are normal and 25 are pathological based on initial classification. The signals are recorded keeping mic 5 cm away from the mouth using a Dictaphone (Sony ICDPX240 4GB).

The record consisted in a 3-4 seconds of sustained sound of the vowel /a/ for each student. The sampling frequency used for recording these signals was 22.05 kHz, with 16 bit resolution and mono. Praat software is used to extract the acoustic parameters after transferring the recorded utterance from Dictaphone to a personal computer Dell (intel ® core ™ i7 CPU M640 @ 2.8 Ghz 2.8 Ghz, 4 Go memory) using audacity software.

Connect the headphone jack to the line input of the PC with a 3.5 jack cable male / male, then activates audio recording with audio processing software audacity. An initial classification based on threshold from the literature, in order to classify the recorded utterances on healthy and pathological. The technic used to identify pathological voices is ANN (artificial neural network): 4 coefficients are extracted (pitch, HNR, jitter and shimmer) from the signal, these coefficients are the vector input of the net.

The net is formed by 3 layers and the used algorithm is back propagation algorithm. The activation function is sigmoid. The proposed algorithm was tested by German database (The sample are from 19 to 22 years old) utterances of /a/. The degree of the severity is evaluated by: Degree of Severity in % = (Measured value – normal value) / (normal value). Figure 1 shows the normal value is related to the threshold defined in the literature. Macro steps of the proposed algorithm and Figure 2 shows training of ANN with input vector (4 acoustic parameters).

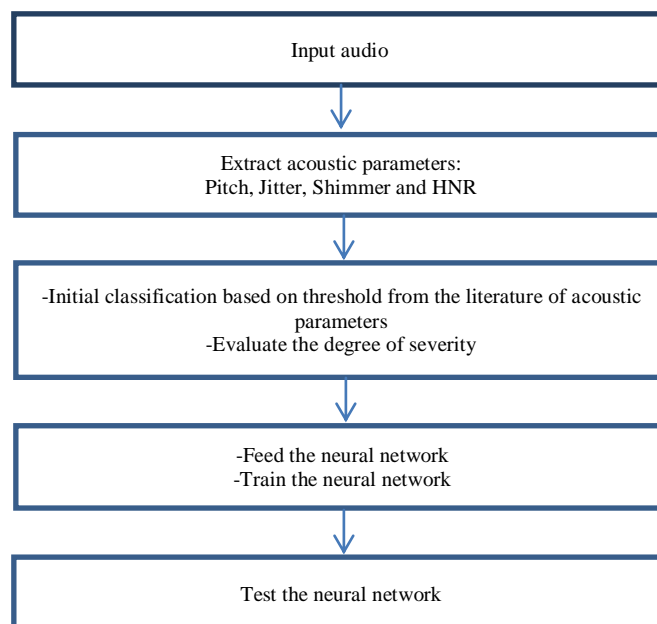


Figure 1. Macro Steps of the Proposed Algorithm

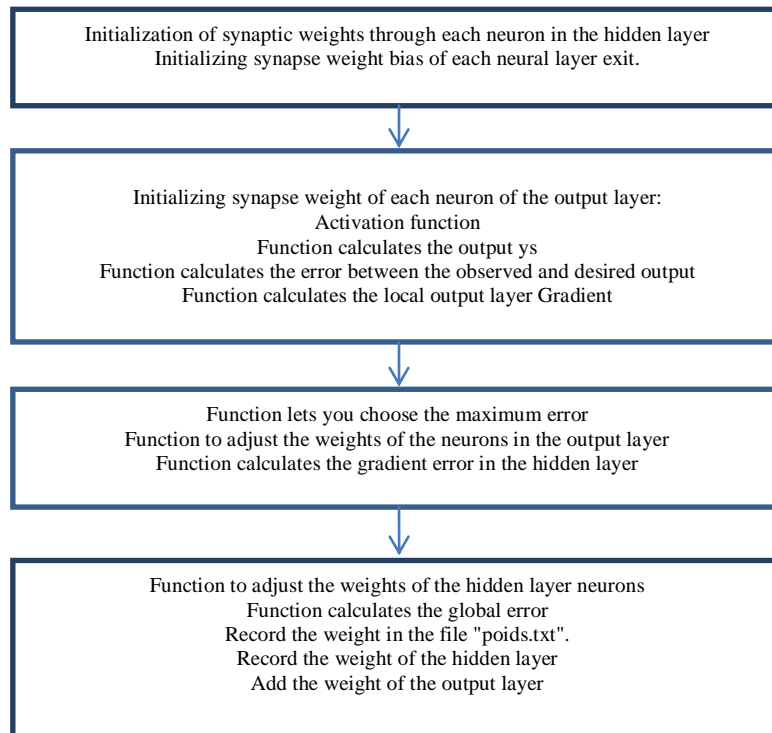


Figure 2. Training of ANN with Input Vector (4 Acoustic Parameters)

#### 4. RESULTS

Description of the dataset as shown in Table 3:

Table 3. Description of the Dataset

Pronounced vowel /a/	Normal	pathological
Training number	50	50
Test number	20	20
Correct classification	18	19
Rate of classification	90%	95%

In this work, the testing set we have choose a German database for voice disorder developed by Putzer in [30] which contain healthy and pathological voice, where each one pronounce vowels [i, a, u] /1-2 s in wav format at different pitch (low, normal, high). The results obtained in that work indicate a percentage of correct classification of 90% for normal voices and 95% for pathological voices.

These results prove that the pitch, HNR, jitter and shimmer can be best input parameters for discrimination and identification of pathological voice using neural network. Also the degree of severity of the identified pathology was estimated in order to give the clinician clear idea on severity of the communication disorder. In this process the accuracy, sensitivity and specificity for each threshold were observed to get the threshold which achieves the best accuracy, and in the same time preserves a high sensitivity and specificity.

True positive (TP): refer to pathological voices that were classified as pathological by proposed algorithm.

True negative (TN): refer to normal voices that were classified as normal by proposed algorithm.

False positive (FP): refer to normal voices that were incorrectly classified as pathological by proposed algorithm.

False negative (FN): refer to pathological voices that were incorrectly classified as normal by proposed algorithm.

1) Specificity =  $TN / (TN + FP) = 95.1\%$

2) Sensitivity =  $TP / (TP + FN) = 1.6\%$

3) Accuracy =  $(TN + TP) / (TN + FP + TP + FN) = 97.9\%$

Table 4 shows evaluation of the proposed algorithm.

Table 4. Evaluation of the Proposed Algorithm

	Classification	
	Pathological	Normal
Pathological	TP:58.3	FN:0.8
Normal	FP:0.7	TN:13.8

## 5. CONCLUSION

The purpose of this work is to conceive a model to assist the clinicians and the professors to follow the evolution of the voice disorders among students, based only on acoustic properties of a student's voice. The results using the ANN (Artificial neural network) classifier gives 90% for normal class and 95% for pathological class.

The performance of the proposed algorithm is evaluated in a term of the accuracy (97.9%), sensitivity(1.6%), and specificity(95.1%). This work can be applied in the field of preventive medicine in order to achieve early detection of voice pathologies. In addition to that, an estimation of the degree of severity was proposed. The major advantage of this type of automatic identification tool is a determinism that is currently lacking in the subjective analysis.

As a possible improvement, system performance can be improved by increasing the corpus Learning. And as a Future work : identify the type of pathology for each voice disorder( stuttering, dysphonia...), the study of the continuous speech is a superior objective and an evident next step and implement the proposed algorithm on a hardware ready to use device as a DSP(Digital signal processing board) or FPGA (Field Programmable Gate Arrays Board).

## ACKNOWLEDGEMENT

The authors fully acknowledged the students Ayoub Ratnane and Abdellah Elhaoui for their valuable contributions.

## REFERENCES

- [1] P. Henriquez, *et al.*, "Characterization of healthy and pathological voice through measures based on nonlinear dynamics," *IEEE Trans on Audio, Speech and Language Processing*, vol/issue: 17(6), 2009.
- [2] J. Camburn, *et al.*, "Parkinson's disease: Speaking out," Denver, CO, The National Parkinson Foundation, 1998.
- [3] M. Vasilakis and Y. Stylianou, "Voice Pathology Detection Based on Short-Term Jitter Estimations in Running Speech," *Folia Phoniatr Logop*, vol. 61, pp. 153–170, 2009.
- [4] N. S. Lechon, *et al.*, "Effect of Audio Compression in Automatic Detection of Voice Pathologies," in *IEEE Transaction on Biomedical Engineering*, vol/issue: 55(12), 2008.
- [5] N. S. Lechon, *et al.*, "Methodological issues in the development of automatic systems for voice pathology detection," in *Biomed. Signal Processing Control*, vol/issue: 1(2), pp. 120-128, 2006.
- [6] L. Salhi, *et al.*, "Voice Disorders Identification Using Multilayer Neural Network," *The International Arab Journal of Information Technology*, vol/issue: 7(2), pp. 177-185, 2010.
- [7] Teixeira J. P., *et al.*, "Vocal Acoustic Analysis M Jitter, Shimmer and HNR Parameters," *Procedia Technology, Elsevier*, vol. 9, pp. 1112-1122, 2013.
- [8] P. Boersma and Weenink D., "Praat: doing phonetics by computer (Version 5.1.17)," 2009. [Computer program] Retrieved October 5, 2009, from <http://www.praat.org/>.
- [9] V. Parsa and D. G Jamieson, "Identification of pathological voices using glottal noise measures," *Journal of Speech, Language & Hearing Research*, vol/issue: 43(2), pp. 469–485, 2000.
- [10] V. Srinivasan, *et al.*, "Classification of Normal and Pathological Voice using GA and SVM," *International Journal of Computer Applications*, vol/issue: 60(3), 2012.
- [11] J. I. G. Llorente, *et al.*, "Support vector machines applied to the detection of Voice disorders," *Nonlinear Analyses and Algorithms for Speech Processing*, pp 219-230, 2005.
- [12] J. D. A. Londono, *et al.*, "Automatic detection of pathological voices using complexity measures, noise parameters, and mel-cepstral coefficients," *IEEE Trans. Biomed. Eng.*, vol/issue: 58(2), 2011.
- [13] B. Boyanov and S. Hadjitodorov, "Acoustic analysis of pathological voices," *IEEE Engineering in Medicine and biology*, pp. 74-82, 1997.
- [14] John H. and L. Hansen, "A non linear operator-based speech feature analysis method with application to vocal fold pathology assessment," *IEEE Transaction on Biomedical Engineering*, vol/issue: 45(3), pp. 300-312, 1998.

- [15] I. Esmaili, *et al.*, “Two-stage feature compensation of clean and telephone speech signals employing bidirectional neural network,” *10th International Conference on Information Science, Signal Processing and their Applications*, Malaysia, 2010.
- [16] N. S. Lechon, *et al.*, “Methodological issues in the development of automatic systems for voice pathology detection,” *Biomedical Signal Processing and Control*, vol. 1, pp. 120–128, 2006.
- [17] M. Arjmandi and M. Pooyan, “An optimum algorithm in pathological voice quality assessment using wavelet-packet-based features, linear discriminant analysis and support vector machine,” *Biomedical signal processing and control*, vol. 7, pp. 3-19, 2012.
- [18] T. Bocklet, *et al.*, “Evaluation and assessment of speech intelligibility on pathologic voices based upon acoustic speaker models,” in *proceeding of 3rd advanced voice function assessment (AVFA '09) international workshop*, pp. 89-92, 2009.
- [19] C. M. Vikram and K. Umarani, “Pathological Voice Analysis To Detect Neurological Disorders Using MFCC & SVM,” *International Journal of Advanced Electrical and Electronics Engineering*, vol/issue: 2(4), 2013.
- [20] D. Pravena, *et al.*, “Pathological Voice Recognition for Vocal Fold Disease,” *International Journal of Computer Applications*, vol/issue: 47(13), 2012.
- [21] R. Behroozmand and F. Almasganj, “Optimal selection of wavelet-packet-based features using genetic algorithm in pathological assessment of patients' speech signal with unilateral vocal fold paralysis,” *Comput. Biol. Med.*, vol. 37, pp. 474–485, 2007.
- [22] V. Majidnezhad and I. Kheidorov, “A HMM-Based Method for Vocal Fold Pathology Diagnosis,” *International Journal of Computer Science*, vol/issue: 9(6), 2012.
- [23] A. A. Dibazar, *et al.*, “Pathological Voice Assessment,” *IEEE EMBS 2006 NEW YORK*, 2006.
- [24] R. B. R. N. Wormald, *et al.*, “Performance of an automated, remote system to detect vocal fold paralysis,” in *Ann Otol Rhinol Laryngol*, vol. 117, pp. 834–838, 2008.
- [25] M. Little, *et al.*, “Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection,” *BioMedical Engineering OnLine*, vol/issue: 6(1), pp. 23, 2007.
- [26] A. Dibazar, *et al.*, “Feature analysis for automatic detection of pathological speech,” *Proceedings of the Second Joint EMBS/BMES Conference*, vol. 1, pp. 182–183, 2002.
- [27] E. Fonseca and J. Pereira, “Normal versus pathological voice signals,” *Engineering in Medicine and Biology Magazine, IEEE*, vol/issue: 28(5), pp. 44–48, 2009.
- [28] V. Parsa and D. G. Jamieson, “Acoustic discrimination of pathological voice: Sustained vowels versus continuous speech,” *Journal of Speech, Language, and Hearing Research*, vol. 44, pp. 327–339, 2001.
- [29] M. Hirano, “Psycho-acoustic evaluation of voice: GRBAS Scale for evaluating the hoarse voice,” *Clinical Examination of voice*, Springer Verlag, 1981.
- [30] <http://stimmdb.coli.uni-saarland.de/>