Journal of Computer Science

Original Research Paper

Recognition of Pathological Voices by Human Factor Cepstral Coefficients (HFCC)

Rabeh Hamdi, Salah Hajji and Adnene Cherif

Signal Processing and Electrical Systems Laboratory, Science Faculty of Tunis, University of Tunis El Manar, Tunisia

Article history
Received: 11-05-2020
Revised: 23-07-2020
Accepted: 04-08-2020

Corresponding Author:
Rabeh Hamdi
Signal Processing and
Electrical Systems Laboratory,
Science Faculty of Tunis,
University of Tunis El Manar,
Tunisia
Email: hamdirabeh@gmai.com

Abstract: Human speech is a means of communication that is very important in our daily lives. It is characterized by its great ability to transmit our ideas, our emotions, our personality etc. So, any alteration of the voice can prevent the person from exercising his professional and daily life naturally. It is for these reasons that it is very necessary to implement systems for detecting and classifying vocal pathologies. These automatic systems can help clinicians customize and detect the existence of any vocal pathology. In this context, several tools have been introduced to achieve early detection of voice disorders. Among these tools are the Human Factor Cepstral Coefficients (HFCC) combined with prosodic parameters, the Noise-Harmonic Ratio (NHR), the Harmonic-Noise Ratio (HNR), analysis of trend Fluctuations (DFA) and Fundamental frequency (F0). These parameters are introduced and calculated in every frame. In this study, we used a variation of HFCC called Equivalent Rectangular Bandwidth (ERB) to study the effects of HFCC on the classification of pathological voices. Using the HTK classifiers, the classification is carried out on two pathological databases, Massachusetts Eye and Ear Infirmary (MEEI) and Saarbruecken Voice Database (SVD). To assess the performance of the system, we used sensitivity and specificity.

Keywords: Pathological Voices, Sensibility, Specificity, ERB, HFCC, HTK, MEEI, SVD

Introduction

In biomedical applications of speech technology, the diagnosis of pathological voice is an important matter. The human voice may be affected by several diseases that appear in the vocal cords. Thus, the vocal treatment of the pathological voice presents some favors, such as its non-invasive and quantitative nature. These benefits allow the identification and observation of diseases of the vocal system and reduce the cost and time required for its treatment. The main objective in the classification of pathological voices is to predict whether the patient's voice is normal or pathological. Proper grading will allow automatic diagnosis and treatment of the disease (Wang and Jo, 2007).

For several years until now, the detection of vocal pathology can be evaluated in a subjective or objective way (Mehta and Hillman, 2008). Indeed, the objective evaluation of acoustic signals is done through computer tools. This assessment identifies and quantifies the underlying vocal pathology that humans cannot hear (Mekyska *et al*., 2015).

Thanks to the technological revolution, the voice can be easily manipulated, so smart devices are used for recording and cloud technologies help with remote processing.

In these works (Al-nasheri *et al*., 2017; Eskidere and Gürhanlı, 2015; Hemmerling *et al*., 2016; Muhammad *et al*., 2017a) the authors used signal processing techniques and machine learning algorithms to build a reliable system to distinguish precisely between healthy voices and pathological ones.

In this same context, we have developed in this study an automatic recognition system for pathological voices. This system is composed of two basic modules which are the parametrization module which extracts the relevant parameters from pathological voices. This module is based on the Cepstral coefficient of the Human Factor (HFCC) proposed by (Skowronski and Harris, 2004). The second module is the classifier used to classify vocal pathologies. We used a Hidden Markov

Model with a Gaussian Mixture density (HMM-GM) (Ali *et al*., 2017), through The Hidden Markov Model Toolkit (HTK) (HTK 3.4.1) (Young *et al*., 2009).

Researchers in this field have frequently used objective assessment of vocal pathology using several databases. We note here the most used databases such as the database (MEEI) (KAYPENTAX; Mekyska *et al*., 2015), Saarbruecken Voice Database (SVD) (Al-nasheri *et al*., 2017; Muhammad *et al*., 2017a; Barry and Pützer, 2016) and Arabic Voice Pathology Database (AVPD) (Mesallam *et al*., 2017; Muhammad *et al*., 2017a). The research carried out on these bases is generally based on the analysis of the phonation of the vowel /a/, for example in the works (Al-Nasheri *et al*., 2014; Amami and Smiti, 2017; Dahmani and Guerti, 2017; Muhammad *et al*., 2017b). While in other works some researchers have combined vowels to do the analysis, for example (Eskidere and Gürhanlı, 2015; Hemmerling *et al*., 2016; Martínez *et al*., 2012).

In our work, we used two databases MEEI Database and Saarbruecken Voice Database for the classification of pathological voices. By comparing our work to the previous ones, we did not analyze vowels but some sentences. In the first database, acoustic samples are recordings of up to 12 sec of readings of the sentence "Rainbow Passage" by men and women and in the other database, we used the recording of the sentence "Guten Morgen, wie geht es Ihnen?" ("Good Morning, how are you?"). Thus, our study is based on the HFCC method combined with the prosodic parameters, the Harmonic Noise Ratio (NHR), the Harmonic-Noise Ratio (HNR), the relaxed analysis of fluctuations (DFA) and the Fundamental frequency (F0) which are calculated for each image.

There are various measures of the performance of a diagnostic test that include different indices such as sensitivity, specificity, accuracy, etc. (Grenier, 1999) and the use of ROC curves (operating characteristic of the receiver).

The probability that the test is positive corresponds to the sensitivity, given that the subject is sick. So, it, measures the ability of a test to detect patients. The closer the sensitivity is to the unit, the fewer errors in the detection of sick subjects (false negatives). The probability that the test is negative corresponds to the specificity, considering that the subject is healthy. So, it measures the ability of a test to detect healthy individuals. The closer the specificity is to unity, the less false positives there are (Bertrand *et al*., 2010).

The relation between the sensitivity and the specificity of a test is represented graphically by the ROC curve, calculated for all possible threshold values. The Area Under the ROC Curve (AUC) is one of the most used overall measures of test performance. It varies between 0.5 in the case of a non-informative test to 1 in the case of perfect execution (Bertrand *et al*., 2010).

The aim of this work is to determine the capacity of these parameters to detect and classify voice pathologies. Another scenario has been used for the parameters alone with HFCC and hybrid. To validate the performance of the recognition system, we used the ROC curve and its under area (AUC).

Materials and Methods

Fundamental frequency F0, Human Factor Cepstral Coefficient (HFCC), the Harmonic to Noise Ratio (HNR) and Detrended Fluctuation Analysis (DFA) are essentially the classical characteristics used for the classification of pathological voices. These classic features are inspired by the cues used in the field of voice recognition. This section provides an overview of the most common features involved in the pathological voice.

Fundamental Frequency

The Fundamental frequency (F0) For a speech signal corresponds to the frequency of vibration of a speaker's vocal cords. This parameter is used in most studies, sometimes in conjunction with the Human Factor Cepstral Coefficient (HFCC) (Hamdi *et al*., 2018). The estimation of F0 has been widely dealt with in the literature and many methods have been proposed, including autocorrelation, instantaneous frequency, cross-correlation, etc. In this study, we used the Sawtooth Waveform Inspired Pitch Estimator (SWIPE) according to (Tsanas *et al*., 2014) and (Camacho and Harris, 2008). This algorithm makes it possible to estimate the pitch in the frequency domain. SWIPE builds on information that is found across the spectrum using kernels. It identifies the harmonics in the square root of the spectrum. Then it imposes kernels with decaying weights on the detected harmonic locations.

Human Factor Cepstral Coefficient (HFCC)

There are many methods of extracting robust functionality available; one of the efficient methods of feature extraction is the Human Factor Cepstral Coefficient (HFCC). This is a new approach to extracting speech characteristics that have been proposed and described in detail in (Skowronski and Harris, 2004).

Skowronski and Harris (2004) introduced the HFCC variant, which is the most recent implementation of the mel band. HFCC is based on a measurement of the width of the filters called "Equivalent Rectangular Bandwidth, or Equivalent Rectangular Bandwidth (ERB)" proposed by Moore and Glasberg in 1983. For each filter, the ERB value is defined as the width of an ideal bandpass filter of the same central frequency, the measurement of these ERBs illustrates the frequency resolution of the hearing system, it is given by the following formula:

$$ERB = 6.23\ 10^{-6} f_c^2 + 93.39\ 10^{-3} f_c + 28.52 \text{ Hz} \qquad (1)$$

With the central frequency (f_c) of the filter expressed in Hz. The bandpass filter calculated in (1) is weighted by a constant called (according to Skowronski and Harris, 2004) ERBscaleFactor (Larsson Alm, 2019).

Figure 1 represents the 32 filters, which cover the frequency range of [115, 8000] Hz, With n = 1, 2,.. (L-1) and L equal to 512 samples. A filter can overlap its closest neighbors as well as its most distant neighbors as shown in the HFCC diagram.

The block diagram of the extraction of the HFCC characteristics is shown in Fig. 2.

First, the speech signal is pre-emphasized and then windowing and weighted by the Hamming window. 25 ms with a frame offset of 10 ms and we apply, The DFT is applied for each frame to obtain the spectrum X (j). Then, the X (j) obtained is used to calculate the amplitude spectrum |X(j)|. Subsequently, the result is filtered by applying a Human Factor Filter Bank. The outputs of the filter bank are compressed by the logarithm function. Finally, the Discrete Cosine Transform (DCT) is used to decorrelate the obtained outputs, yielding the HFCC Coefficients (Ganchev, 2011):

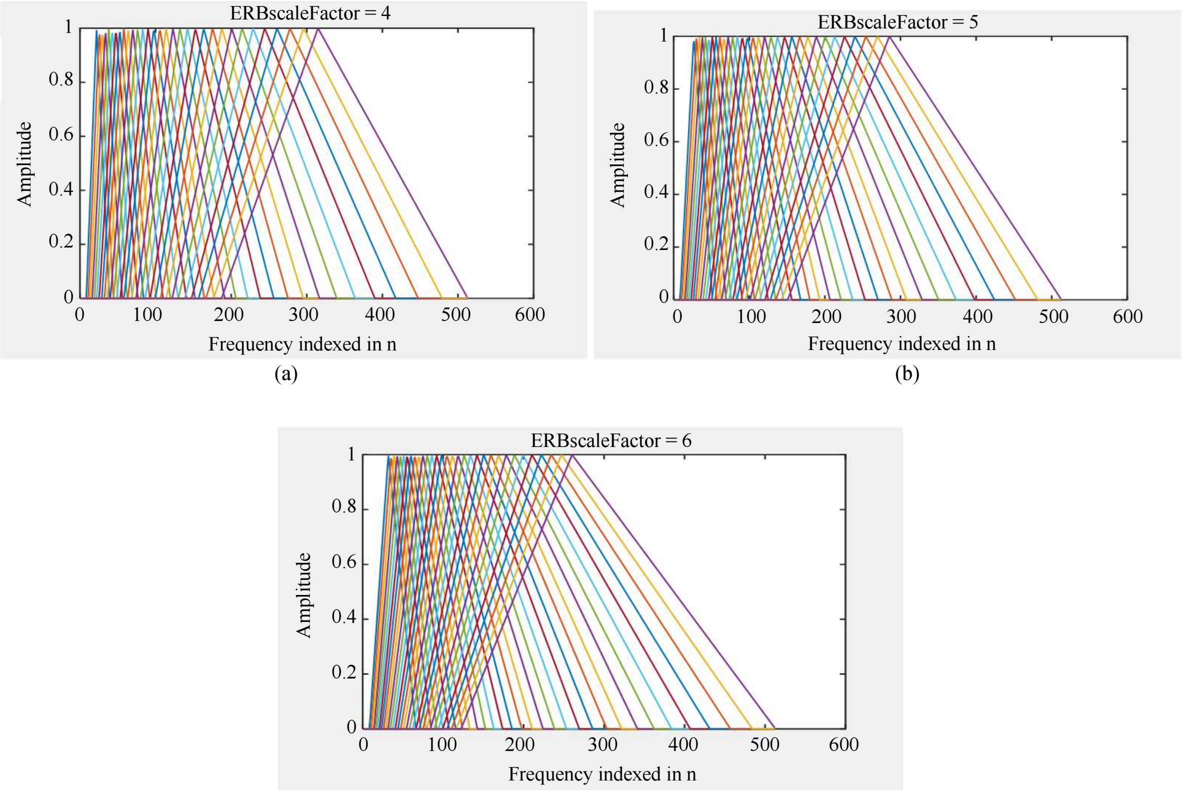
$$HFCC_i = \sqrt{\frac{2}{N}} \sum_{K=1}^{N} \log\left(S_k\right)\ Cos\left[\frac{\pi i}{N}\left(K - \frac{1}{2}\right)\right]$$

$$\text{with} \qquad i = 1,2,3,.......,M \qquad (2)$$

$$\text{and} \qquad S_k = \sum_{j=0}^{511} \left|X(j)\right| H_K(j)$$

where, N is the number of filters in the filter-bank, M is the number of HFCC coefficients and represents the logarithmic energy output of the k^{th} filter (k = 1, 2… N). N and M are chosen as the following: N = 32 and M = 12 for the HFCC computations and $H_k(j)$, k = 1, 2,., N, represents the filter bank in the frequency space.

The sampling frequency is 16000 Hz, Skowronski and Harris have proposed this implementation of the 32-filter HFCC filter bench, which covers the frequency range [115 8000] Hz.



(a)

(b)

(c)

Fig. 1: The HFCC filter bank shown in figures (a), (b) and (c) are respectively for ERB = 4, 5 and 6

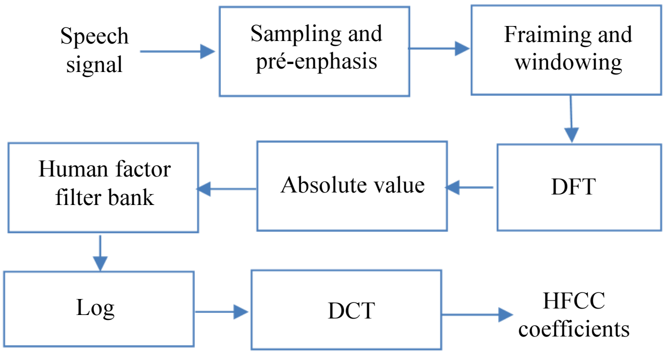


Fig. 2: HFCC implementation

The design of the HFCC filter bank is described as follows. First, we choose the number of filters M as well as the minimum frequency f_{low} and maximum f_{high} of the entire filter bank, the central frequencies f_{c1} and f_{cM} are calculated as follows:

$$f_{ci} = \frac{1}{2}\left(-\bar{b} + \sqrt{\bar{b}^2 - 4a\bar{c}}\right) \quad (3)$$

With i is the index of the center frequency 1 or M, the coefficients $\bar{b}$ and $\bar{c}$ are defined by:

$$\bar{b} = \frac{b - \hat{b}}{a - \hat{a}} \ \ and\ \bar{c} = \frac{c - \hat{c}}{a - \hat{a}} \quad (4)$$

The constants a, b and c mentioned in (1) are expressed by the following values: 6.23*10^{-6}, 93.39*10^{-3} and 28.52 respectively and they vary in both cases, for the first filter, the coefficients $\hat{a}$, $\hat{b}$ and $\hat{c}$ are calculated as follows:

$$\hat{a} = \frac{1}{2\left(700 + f_{low}\right)} \quad (5)$$

$$\hat{b} = \frac{700}{700 + f_{low}} \quad (6)$$

$$\hat{c} = -\frac{f_{low}}{2}\left(1 + \frac{1}{700 + f_{low}}\right) \quad (7)$$

For the last filter, these coefficients are given by:

$$\hat{a} = -\frac{1}{2\left(700 + f_{high}\right)} \quad (8)$$

$$\hat{b} - \frac{700}{700 + f_{high}} \quad (9)$$

$$\hat{c} = \frac{f_{high}}{2}\left(1 + \frac{700}{700 + f_{high}}\right) \quad (10)$$

Once the center frequencies of the first and last filter are calculated, the generation of the center frequencies of the filters in the middle is easy because they are equidistant on the mel scale, the step $\widehat{\Delta f}$ between the center frequencies of the filters adjacent is calculated by:

$$\widehat{\Delta f} = \frac{\widehat{f_{cM}} - \widehat{f_{c1}}}{M - 1} \quad (11)$$

The passage of $f_{c1} \to \widehat{f_{c1}}$ and $f_{cM} \to \widehat{f_{cM}}$ is given by the formula:

$$\widehat{f_{mel}} = 2595\ \log_{10}\left(1 + \frac{f_{ci}}{700}\right) \quad (12)$$

the center frequencies of adjacent filters are calculated by:

$$\widehat{f_{ci}} = \widehat{f_{c1}} + \left(i - 1\right)\widehat{\Delta f} \quad \text{for } i = 2,3,....,\left(M - 1\right) \quad (13)$$

Finally, the maximum and minimum frequencies of each i filter are expressed by:

$$f_{low\ i} = -\left(700 + ERB_i\right) + \sqrt{\left(700 + ERB_i\right)^2 + f_{ci}\left(f_{ci} + 1400\right)} \quad (14)$$

$$f_{high\ i} = f_{low\ \mathrm{i}} + 2ERB_i \quad (15)$$

$$with\ \ ERB_i = \frac{1}{2}\left(f_{high\ \mathrm{i}} - f_{low\ \mathrm{i}}\right) \quad (16)$$

Harmonic to Noise Ratio (HNR) and Noise to Harmonic Ratio (NHR)

The NHR measures the amount of noise in the voice signal and assesses vocal quality. When the signal-to-noise ratio is high, there will be good voice signal quality (Grueber, 2011). Thus, the HNR is a measure examining the presence of noise during phonation. To calculate it, the signal is firstly down sampled to 16 kHz and split into 25 ms length frames, with 10 ms shift. In each frame, a comb filter is applied to the signal to compute the energy in the harmonic components (Teixeira *et al*., 2013).

Detrended Fluctuation Analysis (DFA)

This acoustic parameter characterizes the extent of the troubled noise in the voice signal. It determines the value of random automatic similarity to noise caused by turbulent airflow in the audio channel (Tsanas, 2012). For example, you may have an increase in the DFA value when the voice fold is incompletely closed (Little *et al*., 2007).

Databases

The experimental study was developed on two pathological voices databases, the MEEI database and the Saarbruecken Voice Database (SVD). In the first database, acoustic samples are recordings of up to 12 seconds of readings of the sentence "Rainbow Passage" by men and women and in the other database, we used the recording of the sentence "Guten Morgen, wie geht es Ihnen?" ("Good Morning, how are you?").

For the MEEI database, we chose a subset comprising 53 healthy voices (33 female voices and 20 male voices) and 96 pathological voices (47 female voices and 49 male voices). As well as for the second base Saarbruecken Voice Database (SVD), we selected a subset comprising 211 healthy voices (127 female voices and 84 male voices) and 154 pathological voices (95 female voices and 59 male voices). Tables 1 and 2 summarize the number of samples of pathological voices from each base.

MEEI Database

The database (MEEI) was registered at the Massachusetts Eyes and Ears Infirmary and marketed by Kay Elemetrics. It contains records of sustained vowel phonations [ah] (3 to 4 s long) and the first 12 sec of rainbow passage spoken by normalophonic subjects and patients with psychogenic, neurological, organic and traumatic, the voice at different stages (from beginning to full development). The environment of recording speech samples is controlled at 25 or 50 kHz and 16 bits resolution (KAYPENTAX).

Saarbruecken Voice Database

The Saarbruecken Voice Database (SVD) is available for free and was registered by the Institute of Phonetics at the University of Saarland. This is a collection of voice recordings of more than 2000 people. This database also contains the recording of the sentence "Guten Morgen, wie geht es Ihnen?" ("Good morning, how are you?") recorded by healthy subjects and those with pathologies. A total of 71 larynx pathologies are identified for this database. 1320 (609 males and 711 females) sessions belong to pathological speakers and 650 (400 males and 250 females) to normal speakers (Hamdi *et al*., 2018; Barry and Pützer, 2016).

Table 1: The number of samples of the pathological voices of MEEI database

	Female		Male	
Pathology	TEST	TRAIN	TEST	TRAIN
Ventricular	10	19	7	14
Gastric	5	10	7	15
Edema	11	22	4	7
Paralysis	11	22	11	20
Hyperfunction	10	21	12	25
Normal	11	22	7	13
Total	58	174	47	94

Table 2: The number of samples of the pathological voices of SVD database

	Female		Male	
Pathology	TEST	TRAIN	TEST	TRAIN
Hyperfunction	55	111	15	30
laryngitis	19	38	27	54
Polyp	7	12	9	16
Spasmodic	14	28	8	14
Normal	127	255	84	168
Total	222	444	143	282

Hidden Markov Model Toolkit

The Hidden Markov Model Toolbox (HTK 3.4.1) is a portable tool used to build and manipulate hidden Markov models. HTK is used primarily in the field of search for voice recognition as well for speech synthesis [htk link]. For each pathological voice we associate it with a Hidden Markov Model with a density of Gaussian Mixture (HMM-GM), four mixtures of diagonal state and five states of observation (Young *et al*., 2009).

An HMM is a probabilistic automaton. It is controlled by the first hidden stochastic process that is internal to HMM, this process begins on the initial state and then moves from state to state respecting the topology of the HMM. The second stochastic process that controls HMM generates the language units that correspond to each state traversed by the first process. A GMM is defined as a mixture of probability distributions that allows you to follow a multivariate Gaussian law. It is the basis of the HMMs recognition systems that are most used. A probability density function is estimated by the finite sum of the Gaussian components of the GMM.

Our developed parametrization method extracts HFCC coefficients with the prosodic parameters of the voice signal (such as F0, HNR, NHR, DFA). These parameters are calculated for each frame.

Evaluation Measures

When we refer to the performance of a classification model, we are concerned with the model's ability to predict or separate diseases correctly. When examining the errors made by a classification model, the confusion matrix gives an overview.

Performance

Several terms are commonly used with the description of sensitivity and specificity. They are True Positive (TP), True Negative (TN), False Negative (FN) and False-Positive (FP). A true positive test is when such a disease is proven on a patient by the diagnostic test. Similarly, the test result is true negative if the disease is proven to be absent in a patient. Besides, if the diagnostic test indicates the presence of the disease to a person who does not suffer from any disease, the result in this case of the test is falsely positive. Also, if a patient is 100% affected by the disease and the result of the diagnostic test suggests that the disease is absent from that patient, we talk about a false-negative test result (Zhu *et al*., 2010).

Accuracy

Shows us how accurate the model is to detect the negative and positive class. It is calculated as the sum of correct classifications divided by the total number of classifications.

Precision

It indicates the probability of a successful classification of a predicted positive class. It is defined by:

$$precision = \frac{TP}{TP + FP} \tag{17}$$

Sensitivity

It corresponds to how positive the test is, given that the subject is ill. It measures the ability of a test to detect diseases. It is written by the formula:

$$sensitivity = \frac{TP}{TP + FN} \tag{18}$$

Specificity

Specificity is the probability that the test will be negative, knowing that the subject is healthy. It, therefore, measures the ability of a test to detect healthy individuals. It is defined by the formula:

$$specificity = \frac{TN}{TN + FP} \tag{19}$$

The Receiver Operating Characteristic (ROC)

Receiver Operating Characteristics (ROC) analysis is a useful method of measuring the ability of a voice recognition model to distinguish between people with illness and those without. Its use in speech processing was born as a method to synthesize the specificity and sensitivity of diagnostic tests across a range of possible cutting points. The area under the ROC curve can be interpreted as a probability of correct classification or prediction (Hajian-Tilaki, 2013).

We discuss in this study the use of the Area Under the ROC Curve (AUC) as a measure of the performance of a classifier.

The Area Under the ROC Curve

The Area Under the ROC Curve (AUC) is one of the most popular summary indices that are associated with the ROC curve. It is an overall measure that indicates the performance of the diagnostic test. AUC's value is between 0 and 1. The overall diagnostic performance of the test is precise when the value of AUC is close to 1 (Obuchowski, 2003; Zhou *et al*., 2009).

Results and Discussion

The results of the experiments carried out for the detection and classification of pathologies are expressed in different terms.

These terms are accuracy (the ratio of correctly detected samples to the total number of samples), sensitivity (proportion of pathological samples identified positively), specificity (proportion of normal samples identified negatively) and the area below the Receiver Operating Characteristic curve (ROC), called the area under the curve.

The functionality extracted from the two different databases must be checked in the detection and classification processes. Therefore, many of experiments have been carried out to verify their reliability and accuracy in both processes. To ensure accuracy, different detection and classification experiments were carried out individually for each combination of parameters (HFCC, F0, HNR, NHR and DFA) and for each value of ERB.

Overall Recognition Rate for the MEEI Database

Figure 3 below illustrates the results of the pathological voice recognition rate for the MEEI database. The acoustic modeling of this database is refined, estimating the probability densities of four-Gaussian. The best recognition rates are obtained for HFCC-NHR with ERB = 6 and HFCC-F0-NHR with ERB = 4 (99.07%), HFCC-F0-NHR with ERB = 5 and HFCC-HNR-DFA with ERB = 6 (98.13%) respectively.

We notice that if we increase the ERB, the recognition rate will be better. In this case, the best rate is obtained for ERB = 4, 5 and 6.

Overall Recognition Rate for the Saarbruecken Voice Database

Figure 4 below illustrates the results of the pathological voice recognition rate for the SVD database. The acoustic modeling of this database is refined, estimating the probability densities of four-Gaussian. The best recognition rates are obtained for HFCC-NHR with ERB = 6 (87.40%) and HFCC-HNR-NHR with ERB = 6 (86.03%) respectively.

We notice that if we increase the ERB, the recognition rate will be better. In this case, the best rate is obtained for ERB = 6.

Pathology Recognition Rate of Male and Female Voices for the MEEI Database

Combining HFCC-NHR Parameters

According to Fig. 3, the combination of the HFCC-NHR parameters gives the highest overall recognition rate for ERB = 6. This result is detailed in Table 3, which gives the recognition rate by pathology for female and male voices for each ERB value. We notice in this case that the best recognition was for the female voices.

Tables 3 and 4 present the classification results for each type of pathology and for different ERB values for female and male voices.

Table 3 gives different performance results of the recognition system for the type of female voices and for the combination of HFCC-NHR parameters. We notice that the precision varies between 83 and 100% for a variation of ERB between 1 and 6. Furthermore, we conclude that the increase in the value of ERB improves the general performance of pathology recognition.

Figure 5 shows a combined measurement of sensitivity and specificity and we see that the pathological voice recognition system, in this case, is efficient. The Area Under the Curve (AUC) varies between 0.99 and 1 for the different ERB values. It shows that the best performance could be obtained in the case of the Equivalent Rectangular Bandwidth (ERB) equal to 5 and equal to 6. In, this case, the distinction between people with the disease and those who do not have the disease is perfect.

Similarly, for male voices, the performance results of the recognition system for the combination of HFCC-NHR parameters are mentioned in Table 4. It is noted that the system is precise; the precision varies between 84 and 100% for a variation ERB between 1 to 6. Besides, it is noted that the increase in the value of the ERB improves the general performance of the recognition of pathologies. (For ERB = 5 and ERB = 6 all diseases are predicted).

Figure 6 shows a combined measure of sensitivity and specificity and we see that the pathological voice recognition system, in this case, is efficient. The Area Under the Curve (AUC) varies between 0.99 and 1 for the different ERB values. It shows that the best performance could be obtained in the case of the Equivalent Rectangular Bandwidth (ERB) equal respectively to 5 and 6. In, this case, the distinction between the different types of pathology is perfect.

Table 3: Evaluation measures for the combination of HFCC-NHR parameters and for different ERB values for each pathology of the MEEI base of female voices

		Ventricular	Gastric	Edema	Paralysis	Hyperfunction	Normal
ERB = 1	ACC (%)	100.00	100.0	100.00	100.00	100.00	100.00
	Precision	0.90	1.0	1.00	0.84	0.90	1.00
	Sensitivity	1.00	1.0	1.00	1.00	1.00	1.00
	Specificity	0.98	1.0	1.00	0.97	0.98	1.00
ERB = 2	ACC (%)	100.00	100.0	100.00	100.00	100.00	100.00
	Precision	1.00	1.0	0.91	1.00	0.83	0.91
	Sensitivity	1.00	1.0	1.00	1.00	1.00	1.00
	Specificity	1.00	1.0	0.98	1.00	0.97	0.98
ERB = 3	ACC (%)	90.00	100.0	100.00	100.00	100.00	100.00
	Precision	0.90	1.0	1.00	0.91	1.00	0.91
	Sensitivity	0.90	1.0	1.00	1.00	1.00	1.00
	Specificity	0.98	1.0	1.00	0.98	1.00	0.98
ERB = 4	ACC (%)	100.00	80.0	100.00	100.00	100.00	100.00
	Precision	0.90	1.0	1.00	1.00	1.00	0.91
	Sensitivity	1.00	0.8	1.00	1.00	1.00	1.00
	Specificity	0.98	1.0	1.00	1.00	1.00	0.98
ERB = 5	ACC (%)	100.00	10.0	100.00	100.00	100.00	100.00
	Precision	1.00	1.0	0.91	0.91	1.00	0.91
	Sensitivity	1.00	1.0	1.00	1.00	1.00	1.00
	Specificity	1.00	1.0	0.98	0.98	1.00	0.98
ERB = 6	ACC (%)	100.00	100.0	100.00	100.00	100.00	100.00
	Precision	1.00	1.0	1.00	1.00	1.00	0.91
	Sensitivity	1.00	1.0	1.00	1.00	1.00	1.00
	Specificity	1.00	1.0	1.00	1.00	1.00	0.98

Table 4: Evaluation measures for the combination of HFCC-NHR parameters and for different ERB values for each pathology of the MEEI base of male voices

		Ventricular	Gastric	Edema	Paralysis	Hyperfunction	Normal
ERB = 1	ACC (%)	85.70	62.500	50.00	100.00	100	100
	Precision	1.00	1.000	1.00	0.84	1	1
	Sensitivity	0.85	0.625	0.50	1.00	1	1
	Specificity	1.00	1.000	1.00	0.97	1	1
ERB = 2	ACC (%)	85.70	62.500	75.00	100.00	100	100
	Precision	1.00	1.000	1.00	0.91	1	1
	Sensitivity	0.85	0.625	0.75	1.00	1	1
	Specificity	1.00	1.000	1.00	0.99	1	1
ERB = 3	ACC (%)	100.00	87.500	50.00	100.00	100	100
	Precision	0.87	1.000	1.00	1.00	1	1
	Sensitivity	1.00	0.875	0.50	1.00	1	1
	Specificity	0.99	1.000	1.00	1.00	1	1
ERB = 4	ACC (%)	100.00	87.500	50.00	100.00	100	100
	Precision	0.87	1.000	1.00	0.91	1	1
	Sensitivity	1.00	0.875	0.50	1.00	1	1
	Specificity	0.99	1.000	1.00	0.98	1	1
ERB = 5	ACC (%)	100.00	87.500	50.00	100.00	100	100
	Precision	1.00	1.000	1.00	1.00	1	1
	Sensitivity	1.00	0.875	0.50	1.00	1	1
	Specificity	1.00	1.000	1.00	1.00	1	1
ERB = 6	ACC (%)	100.00	87.500	100.00	100.00	100	100
	Precision	1.00	1.000	1.00	1.00	1	1
	Sensitivity	1.00	0.875	1.00	1.00	1	1
	Specificity	1.00	1.000	1.00	1.00	1	1

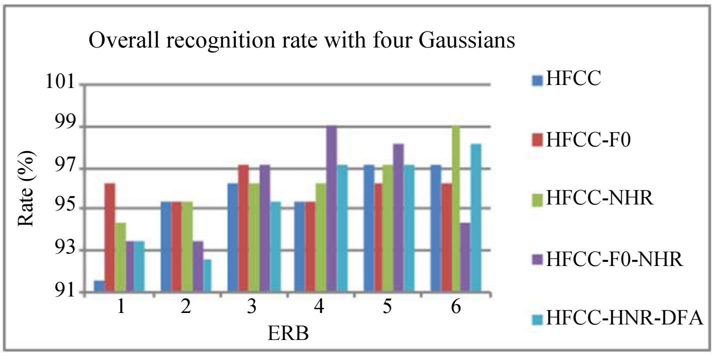


Fig. 3: Performance of the extracted features on each ERB with Four Gaussians for the MEEI Database

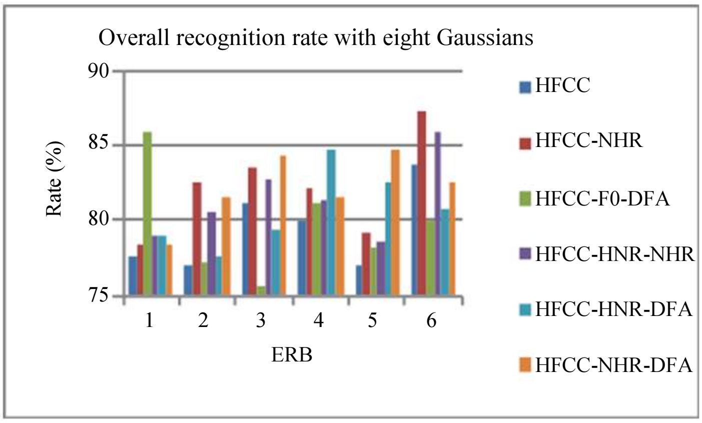


Fig. 4: Performance of the extracted features on each ERB with Eight Gaussians for the SVD Database

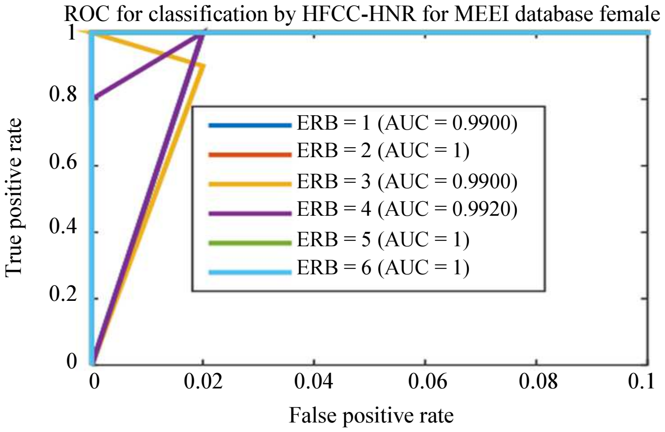


Fig. 5: ROC for classification by HFCC-HNR for the MEEI database for female voices

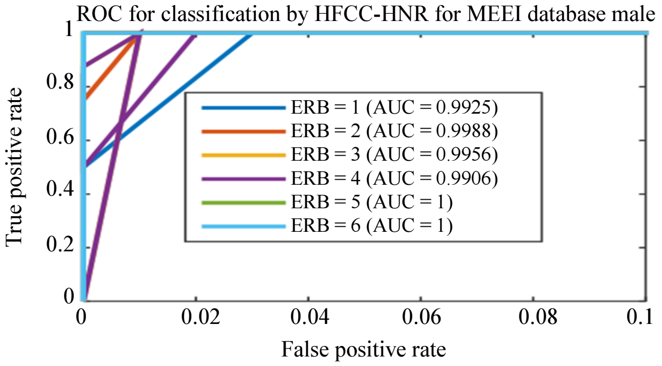


Fig. 6: ROC for classification by HFCC-HNR for MEEI database for male voices

Combining HFCC-F0-NHR Parameters

In this study for the combination of HFCC-F0-NHR parameters, we can see that the recognition system is more precise concerning type pathologies (ventricular, paralysis and hyperfunction) for male voices and for all ERB values. On the other hand, we note that for all types of pathologies and at different ERB values, the recognition rate of female voices is better. Tables 5 and 6 give an overview of the different results. Figure 7 shows that the test is perfectly discriminating when ERB equal to 4 and 5 for the HFCC-F0-NHR combination for female voices in the MEEI database. While for male voices and in the same base, Fig. 8 shows that the test is perfectly discriminating for ERB equal to 3, 4 and 5.

Pathology Recognition rate of Male and Female Voices for the SVD Database for Combining HFCC-NHR Parameters

According to Fig. 3, the best pathology recognition rate for the combination of HFCC-NHR parameters is obtained for female voices. Table 7 shows that laryngitis and spasmodic pathologies have a recognition rate of 100% for an ERB value, respectively equal to 3, 4 and 5. While for male type voices, the best recognition rate is obtained for healthy voices of 98.8%.

We note that the recognition rate of pathologies is improved when the value of ERB increases.

Table 7 illustrates the different performance values of the recognition system for the SVD database for female voices and for the combination of HFCC-NHR parameters. We note that the best recognition rates for the different pathologies are 100% (ERB = 3, 4 et 5), 96.1% (ERB = 6) and 90.9% (ERB = 6) corresponds respectively to pathologies of laryngitis, Spasmodic type, Normal and Hyperfunction. However, the accuracy is not enough to give a conclusion on the effectiveness of the recognition system and therefore these measures should be supplemented by the ROC curve as shown in Fig. 9. The latter shows that the best performance is obtained with the values of the highest Equivalent Rectangular Bandwidth (ERB) (AUC = 0.9704).

Table 8 illustrates the different performance values of the recognition system for the SVD database for male's voices and for the combination of HFCC-NHR parameters. We find that the best recognition rates for the different pathologies are 98.8% (ERB = 5), 87.5% (ERB = 1,2,4), 77.8% (ERB = 3, 6) and 76.7% (ERB = 6) corresponds respectively to Normal, Spasmodic, laryngitis, polyp and Hyperfunction pathologies, However, the accuracy is not sufficient to give a final conclusion on the effectiveness of the recognition system and therefore these measures should be supplemented by the Curve ROC as shown in Fig. 10. This last shows that the best performance is obtained with the values of the Equivalent Rectangular Bandwidth (ERB) equal respectively 5, 4 and 3 with the values of the Area Under the Curve (AUC) equal respectively AUC = 0.9752, 0.9734 and 0.9671.

Table 5: Evaluation measures for the combination of HFCC-F0-NHR parameters and for different ERB values for each pathology of the MEEI base of female voices

		Ventricular	Gastric	Edema	Paralysis	Hyperfunction	Normal
ERB = 1	ACC (%)	100.00	80.00	100.00	90.9	100.0	100.00
	Precision	0.90	0.80	1.00	0.90	1.0	0.91
	Sensitivity	1.00	0.80	1.00	0.90	1.0	1.00
	Specificity	0.98	0.99	1.00	0.98	1.0	0.98
ERB = 2	ACC (%)	100.00	80.00	100.00	100.00	100.0	100.00
	Precision	0.83	1.00	0.91	0.91	1.0	0.91
	Sensitivity	1.00	0.8	1.00	1.00	0.9	1.00
	Specificity	0.97	1.00	0.98	0.98	1.0	0.98
ERB = 3	ACC (%)	100.00	100.00	100.00	100.00	100.0	100.00
	Precision	0.83	1.00	1.00	1.00	1.0	0.91
	Sensitivity	1.00	1.00	1.00	1.00	1.0	1.00
	Specificity	0.97	1.00	1.00	1.00	1.0	0.98
ERB = 4	ACC (%)	100.00	100.00	100.00	100.00	100.0	100.00
	Precision	1.00	1.00	1.00	1.00	1.0	0.91
	Sensitivity	1.00	1.00	1.00	1.00	1.0	1.00
	Specificity	1.00	1.00	1.00	1.00	1.0	0.98
ERB = 5	ACC (%)	100.00	100.00	100.00	100.00	100.0	100.00
	Precision	1.00	1.00	0.91	1.00	1.0	0.91
	Sensitivity	1.00	1.00	1.00	1.00	1.0	1.00
	Specificity	1.00	1.00	0.98	1.00	1.0	0.98
ERB = 6	ACC (%)	90.00	100.00	100.00	100.00	100.0	100.00
	Precision	0.81	1.00	0.91	0.91	1.0	0.91
	Sensitivity	0.90	1.00	1.00	1.00	1.0	1.00
	Specificity	0.97	1.00	0.98	0.98	1.0	0.98

Table 6: Evaluation measures for the combination of HFCC-F0-NHR parameters and for different ERB values for each pathology of the MEEI base of male voices

		Ventricular	Gastric	Edema	Paralysis	Hyperfunction	Normal
ERB = 1	ACC (%)	100	75.000	25.00	100.00	100.00	100
	Precision	1	1.000	1.00	0.84	0.92	1
	Sensitivity	1	0.750	0.25	1.00	1.00	1
	Specificity	1	1.000	1.00	0.97	0.98	1
ERB = 2	ACC (%)	100	75.000	25.00	100.00	100.00	100
	Precision	1	1.000	1.00	0.91	0.92	1
	Sensitivity	1	0.750	0.25	1.00	1.00	1
	Specificity	1	1.000	1.00	0.98	0.98	1
ERB = 3	ACC (%)	100	75.000	75.00	100.00	100.00	100
	Precision	1	1.000	1.00	1.00	1.00	1
	Sensitivity	1	0.750	0.75	1.00	1.00	1
	Specificity	1	1.000	1.00	1.00	1.00	1
ERB = 4	ACC (%)	100	87.500	100.00	100.00	100.00	100
	Precision	1	1.000	1.00	1.00	1.00	1
	Sensitivity	1	0.875	1.00	1.00	1.00	1
	Specificity	1	1.000	1.00	1.00	1.00	1
ERB = 5	ACC (%)	100	75.000	100.00	100.00	100.00	100
	Precision	1	1.000	1.00	1.00	1.00	1
	Sensitivity	1	0.750	1.00	1.00	1.00	1
	Specificity	1	1.000	1.00	1.00	1.00	1
ERB = 6	ACC (%)	100	75.000	25.00	100.00	100.00	100
	Precision	1	1.000	1.00	0.91	1.00	1
	Sensitivity	1	0.750	0.25	1.00	1.00	1
	Specificity	1	1.000	1.00	0.98	1.00	1

Table 7: Evaluation measures for the combination of HFCC-NHR parameters and for different ERB values for each pathology of the SVD database of female voices

		Hyperfunction	laryngitis	Polyp	Spasmodic	Normal
ERB = 1	ACC (%)	67.30	36.80	28.60	78.60	92.90
	Precision	0.75	0.77	0.66	0.91	0.83
	Sensitivity	0.67	0.58	0.40	0.78	0.92
	Specificity	0.96	0.97	0.99	0.99	0.90
ERB = 2	ACC (%)	67.30	73.70	28.60	92.90	92.10
	Precision	0.82	0.63	1.00	0.61	0.91
	Sensitivity	0.67	0.73	0.28	0.92	0.92
	Specificity	0.97	0.97	1.00	0.97	0.95
ERB = 3	ACC (%)	69.10	**100**.00	14.30	85.70	92.90
	Precision	0.80	0.67	0.50	0.52	0.96
	Sensitivity	0.69	1.00	0.14	0.85	0.92
	Specificity	0.95	0.97	0.99	0.96	0.98
ERB = 4	ACC (%)	56.40	68.40	28.60	**100**.00	92.90
	Precision	0.83	0.76	1.00	0.50	0.90
	Sensitivity	0.56	0.68	0.28	1.00	0.92
	Specificity	0.98	0.98	1.00	0.96	0.94
ERB = 5	ACC (%)	69.10	**100**.00	28.60	85.70	83.50
	Precision	0.86	0.70	1.00	0.32	0.96
	Sensitivity	0.69	1.00	0.28	0.85	0.84
	Specificity	0.98	0.97	1.00	0.93	0.98
ERB = 6	ACC (%)	**90.90**	78.90	**42.90**	92.90	**96.10**
	Precision	1.00	1.00	0.75	0.61	0.97
	Sensitivity	0.90	0.78	0.42	0.92	0.96
	Specificity	1.00	1.00	0.99	0.97	0.98

Table 8: Evaluation measures for the combination of HFCC-NHR parameters and for different ERB values for each pathology of the SVD database of male voices

		Hyperfunction	laryngitis	Polyp	Spasmodic	Normal
ERB = 1	ACC (%)	26.700	70.40	33.30	**87.500**	92.90
	Precision	0.260	0.57	1.00	1.000	0.89
	Sensitivity	0.260	0.70	0.33	0.875	0.92
	Specificity	0.960	0.95	1.00	1.000	0.96
ERB = 2	ACC (%)	53.300	74.10	66.70	**87.500**	91.70
	Precision	0.570	0.64	0.46	1.000	0.93
	Sensitivity	0.530	0.74	0.66	0.875	0.91
	Specificity	0.980	0.96	0.98	1.000	0.98
ERB = 3	ACC (%)	40.000	59.30	**77.80**	75.000	97.60
	Precision	0.350	0.76	0.58	1.000	0.92
	Sensitivity	0.400	0.59	0.77	0.750	0.97
	Specificity	0.970	0.98	0.98	1.000	0.97
ERB = 4	ACC (%)	60.000	74.10	55.60	**87.500**	96.40
	Precision	0.750	0.64	0.55	1.000	0.88
	Sensitivity	0.600	0.74	0.55	0.875	0.96
	Specificity	0.990	0.96	0.98	1.000	0.96
ERB = 5	ACC (%)	40.000	55.60	22.20	75.000	**98.80**
	Precision	0.850	0.68	0.66	0.750	0.79
	Sensitivity	0.400	0.55	0.22	0.750	0.98
	Specificity	0.990	0.97	0.99	0.990	0.92
ERB = 6	ACC (%)	**76.700**	**77.80**	44.40	62.500	94.00
	Precision	0.875	0.51	1.00	0.550	0.89
	Sensitivity	0.460	0.77	0.44	0.625	0.94
	Specificity	0.990	0.94	1.00	0.980	0.96

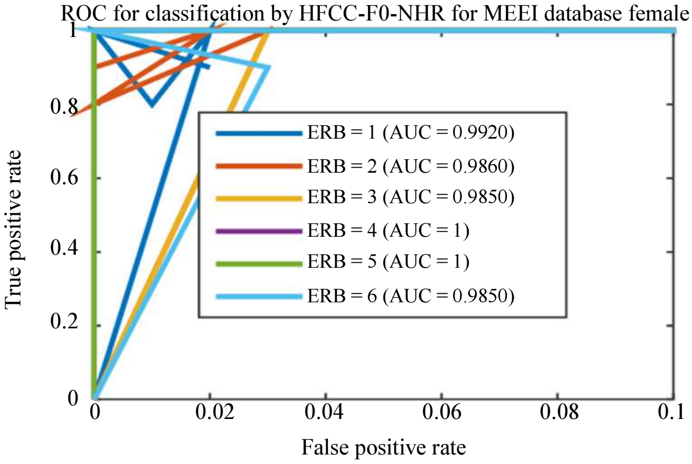


Fig. 7: ROC for classification by HFCC-F0-NHR for MEEI database for female voices

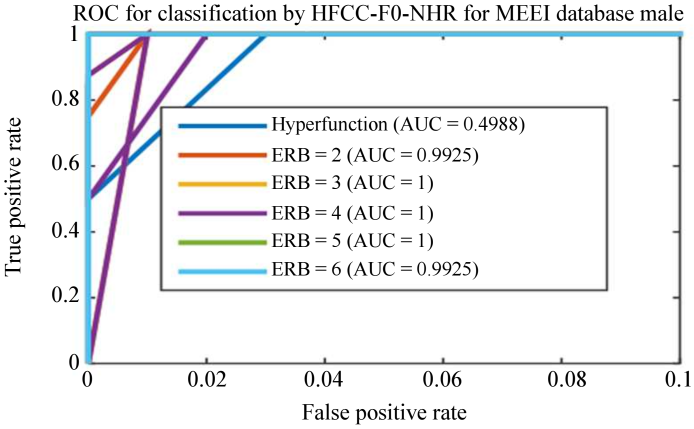


Fig. 8: ROC for classification by HFCC-F0-NHR for MEEI database for male voices

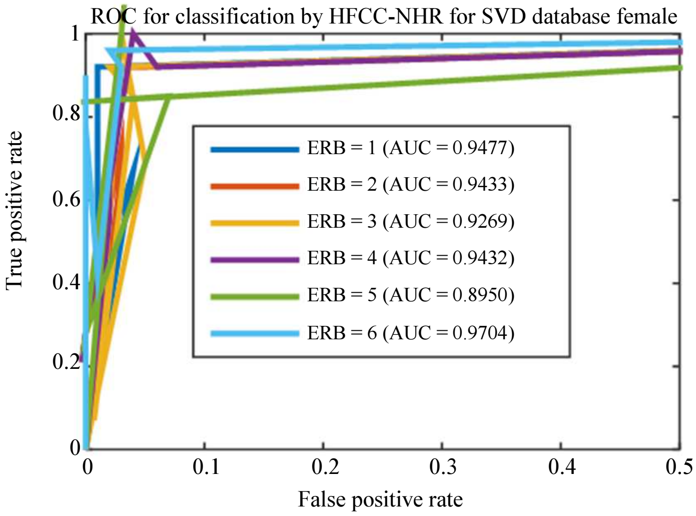


Fig. 9: ROC for classification by HFCC-NHR for SVD database for female voices

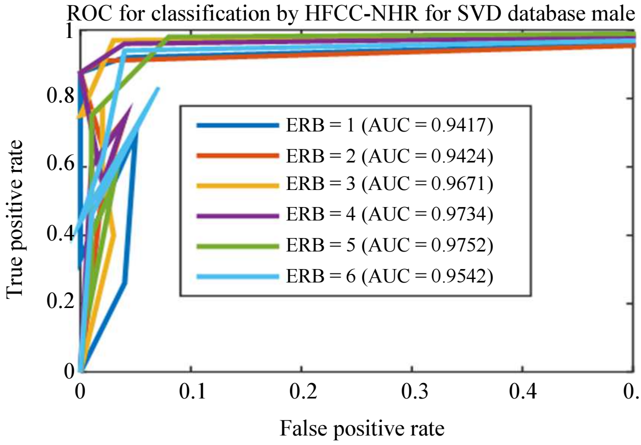


Fig. 10: ROC for classification by HFCC-NHR for SVD database for male voice

Conclusion

As part of this work, we improved the assessment of voice disorder using prosodic parameters using two different types of MEEI and SVD databases. Recognition rates varied from the database to database with the same combination of prosodic parameters. The best overall recognition rates are 99.07% for the samples taken in MEEI and 87.40% for the samples taken in SVD.

The recognition rates obtained, as well as the sensitivities in this study, are essential to detect and classify vocal pathologies. For example, certain combinations of parameters have an excellent indication that they can contribute to the detection and classification of voice pathologies such as HFCC-NHR and HFCC-F0-NHR.

For the MEEI database, the recognition of pathological voices is better for female type voices for these two combinations of parameters. The recognition system is more precise for the value of the Equivalent Rectangular Bandwidth (ERB) equal to 6. HFCC-F0-NHR.

For the SVD database for specific pathologies, we conclude that the recognition of female type voices is better compared to those of male type such as (spasmodic, laryngitis and hyperfunction). On the other hand, for the male polyp pathology is well recognized compared to that of the female.

Author's Contributions

All authors equally contributed in this work.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

Abbreviations

ERB: Equivalent Rectangular Bandwidth; HFCC: Human Factor Cepstral Coefficients; HTK: Hidden Markov Model Toolkit; MEEI: Massachusetts Eye and Ear Infirmary; SVD: Saarbruecken Voice Database; NHR: Noise to Harmonic Ratio; HNR: Harmonic to Noise Ratio; DFA: Detrended Fluctuation Analysis; HMM-GM: Hidden Markov Model Gaussian Mixture; AVPD: Arabic Voice Pathology Database; F0: Fundamental Frequency; ROC: Receiver Operating Characteristics; AUC: Area Under the ROC Curve; DFT: Discrete Fourier Transform; DCT: Discrete Cosine Transform; TP: True Positive; TN: True Negative; FP: False Positive; FN: False Negative.

Availability of Data and Materials

We have used two pathological voices databases to consolidate the experimental results. The MEEI database is from KAYPENTAX, voice and speech laboratory, Massachusetts Eye and Ear Infirmary, Boston, MA (Disordered Voice Database and Program, Model 4337, version 1.03) and Saarbruecken Voice Database is available on the following website (http://www.stimmdatenbank.coli.uni-saarland.de/help_en.php4). These pathological voices databases are frequently used in the literature.

References

Ali, Z., Alsulaiman, M., Muhammad, G., Elamvazuthi, I., Al-Nasheri, A., Mesallam, T. A., ... & Malki, K. H. (2017). Intra-and inter-database study for Arabic, English and German databases: do conventional speech features detect voice pathology? Journal of Voice, 31(3), 386-e1.

Al-Nasheri, A., Ali, Z., Muhammad, G., & Alsulaiman, M. (2014, November). Voice pathology detection using auto-correlation of different filters bank. In 2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA) (pp. 50-55). IEEE.

Al-Nasheri, A., Muhammad, G., Alsulaiman, M., & Ali, Z. (2017). Investigation of voice pathology detection and classification on different frequency regions using correlation functions. Journal of Voice, 31(1), 3-15.

Amami, R., & Smiti, A. (2017). An incremental method combining density clustering and support vector machines for voice pathology detection. Computers & Electrical Engineering, 57, 257-265.

Barry, W. J., & Pützer, M. (2016). Saarbrücken Voice Database, Institute of Phonetics, Univ. of Saarland.

Bertrand, D., Fluss, J., Billard, C., & Ziegler, J. C. (2010). Efficacité, sensibilité, spécificité: Comparaison de différents tests de lecture. LAnnee psychologique, 110(2), 299-320.

Camacho, A., & Harris, J. G. (2008). A sawtooth waveform inspired pitch estimator for speech and music. The Journal of the Acoustical Society of America, 124(3), 1638-1652.

Dahmani, M., & Guerti, M. (2017, May). Vocal folds pathologies classification using Naïve Bayes Networks. In 2017 6th international conference on systems and control (ICSC) (pp. 426-432). IEEE.

Eskidere, Ö., & Gürhanlı, A. (2015). Voice disorder classification based on multitaper mel frequency cepstral coefficients features. Computational and mathematical methods in medicine, 2015.

Ganchev, T. (2011). Contemporary methods for speech parameterization. In Contemporary Methods for Speech Parameterization (pp. 1-106). Springer, New York, NY.

Grenier, B. (1999). Évaluation de la décision médicale : Introduction à l'analyse médico-économique ». Paris: Masson.

Grueber, M. (2011). Effets de la stimulation sousthalamique bilatérale sur la voix et la parole de patients parkinsoniens (Doctoral dissertation, Université de Lausanne, Faculté de biologie et médecine).

Hajian-Tilaki, K. (2013). Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian journal of internal medicine, 4(2), 627.

Hamdi, R., Hajji, S., & Cherif, A. (2018). Voice Pathology Recognition and Classification using Noise Related Features. INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS, 9(11), 82-87.

Hemmerling, D., Skalski, A., & Gajda, J. (2016). Voice data mining for laryngeal pathology assessment. Computers in biology and medicine, 69, 270-276.

KAYPENTAX, Disordered Voice Database, Version 1.03 (CD-ROM), MEEI, Voice and Speech Lab, Kay Elemetrics Corp., Boston, MA, USA, Oct. 1994. http://read.pudn.com/downloads603/ebook/2459102/Disordered%20Voice%20Database%20flyer%202007%20opt%20(1).pdf

Larsson Alm, K. (2019). Automatic Speech Quality Assessment in Unified Communication: A Case Study.

Little, M. A., McSharry, P. E., Roberts, S. J., Costello, D. A., & Moroz, I. M. (2007). Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. Biomedical engineering online, 6(1), 23.

Martínez, D., Lleida, E., Ortega, A., Miguel, A., & Villalba, J. (2012). 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 (pp. 99-109). Springer, Berlin, Heidelberg.

Mehta, D. D., & Hillman, R. E. (2008). Voice assessment: updates on perceptual, acoustic, aerodynamic and endoscopic imaging methods. Current opinion in otolaryngology & head and neck surgery, 16(3), 211.

Mekyska, J., Janousova, E., Gomez-Vilda, P., Smekal, Z., Rektorova, I., Eliasova, I., ... & López-de-Ipiña, K. (2015). Robust and complex approach of pathological speech signal analysis. Neurocomputing, 167, 94-111.

Mesallam, T. A., Farahat, M., Malki, K. H., Alsulaiman, M., Ali, Z., Al-Nasheri, A., & Muhammad, G. (2017). Development of the arabic voice pathology database and its evaluation by using speech features and machine learning algorithms. Journal of healthcare engineering, 2017.

Muhammad, G., Alsulaiman, M., Ali, Z., Mesallam, T. A., Farahat, M., Malki, K. H., ... & Bencherif, M. A. (2017a). Voice pathology detection using interlaced derivative pattern on glottal source excitation. Biomedical signal processing and control, 31, 156-164.

Muhammad, G., Alhamid, M. F., Hossain, M. S., Almogren, A. S., & Vasilakos, A. V. (2017b). Enhanced living by assessing voice pathology using a co-occurrence matrix. Sensors, 17(2), 267.

Obuchowski, N. A. (2003). Receiver operating characteristic curves and their use in radiology. Radiology, 229(1), 3-8.

Skowronski, M. D., & Harris, J. G. (2004). Exploiting independent filter bandwidth of human factor cepstral coefficients in automatic speech recognition. The Journal of the Acoustical Society of America, 116(3), 1774-1780.

Teixeira, J. P., Oliveira, C., & Lopes, C. (2013). Vocal acoustic analysis–jitter, shimmer and hnr parameters. Procedia Technology, 9, 1112-1122.

Tsanas, A. (2012). Accurate telemonitoring of Parkinson's disease symptom severity using nonlinear speech signal processing and statistical machine learning (Doctoral dissertation, Oxford University, UK).

Tsanas, A., Zañartu, M., Little, M. A., Fox, C., Ramig, L. O., & Clifford, G. D. (2014). Robust fundamental frequency estimation in sustained vowels: Detailed algorithmic comparisons and information fusion with adaptive Kalman filtering. The Journal of the Acoustical Society of America, 135(5), 2885-2901.

Wang, J., & Jo, C. (2007, August). Vocal folds disorder detection using pattern recognition methods. In 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 3253-3256). IEEE.

Young, S. J., Evermann, G., Gales, M. J. F., Hain, T., Kershaw, D., Liu, X., ... & Valtchev, V. (2009). The HTK Book (for HTK version 3.4. 1), Cambridge University.

Zhou, X. H., McClish, D. K., & Obuchowski, N. A. (2009). Statistical methods in diagnostic medicine (Vol. 569). John Wiley & Sons.

Zhu, W., Zeng, N., & Wang, N. (2010). Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. NESUG proceedings: Health care and life sciences, Baltimore, Maryland, 19, 67.