Recognition of pathological voices by Human Factor Cepstral Coefficients (HFCC)

Rabeh Hamdi (✉ hamdirabeh@gmail.com)

hamdirabeh <https://orcid.org/0000-0003-0894-1983>

Salah HAJJI

energy research and technology center CRTEn Borj-cedria technopark

Adnene Cherif

Universite de Tunis El Manar Faculte des Sciences de Tunis

Research

Keywords: pathological voices, sensibility, specificity, ERB, HFCC, HTK, MEEI, SVD

Posted Date: April 22nd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-23108/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Journal of Computer Science on August 1st, 2020. See the published version at <https://doi.org/10.3844/jcssp.2020.1085.1099>.

Recognition of pathological voices by Human Factor Cepstral Coefficients (HFCC)

HAMDI Rabeh¹, HAJJI Salah, and CHERIF Adnane

Abstract. 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 work, 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*

1. 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. [1]

For several years until now, the detection of vocal pathology can be evaluated in a subjective or objective way [2]. 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 [3].

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 [4, 5, 6, 7] 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 [8]. The second module is the classifier used to classify vocal pathologies. we used a hidden Markov model with a Gaussian mixture density (HMM-GM) [9], through The Hidden Markov Model Toolkit (HTK) (HTK 3.4.1) [10]

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) [11, 3], Saarbruecken Voice Database (SVD) [4, 7, 12] and Arabic Voice Pathology Database (AVPD) [13, 7]. 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 [14, 15, 16, 17]. While in other works some researchers have combined vowels to do the analysis, for example [5, 6, 18].

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 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

H. Rabeh (✉) · H. Salah · C. Adnane

Laboratory: Analysis and Processing of Signals and Electric and Energy System, 1060 Tunis, Tunisia

e-mail: hamdirabeh@gmail.com

H. Salah

e-mail: msahli.salah@yahoo.fr

C. Adnane

e-mail: adnen2fr@yahoo.fr

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. [19] 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. [20]

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. [20]

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).

2. Methods and Materials

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.

2.1. 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).[21]

2.2. 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 [8].

In 2004 Skowronski and Harris 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:

$$ERB = 6.2310^{-6} f_c^2 + 93.3910^{-3} f_c + 28.52 \text{ HZ} \quad (1)$$

With the central frequency (Fc) of the filter expressed in Hz. The bandpass filter calculated in (1) is weighted by a constant called (according to Skowronski and Harris) ERBscaleFactor. [22].

(a)

(b)

(c)

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

Figure. 1 represents the 32 filters, which cover the frequency range of [115, 8000] Hz, With $n = 1.2$, $N-1$ and N 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 Figure.2.

Fig. 2 HFCC Implementation.

First, the speech signal is pre-emphasized and weighted by the Hamming window. 25 ms with a frame offset of 10 ms. On apply thereafter, The DFT 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. [23]

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

$i = 1, 2, 3, \dots, M$

$$S_k = \sum_{j=0}^{511} |X(j)| H_k(j)$$

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 ($k=1, 2, \dots, N$). N and M are chosen as the following: $N = 32$ and $M = 12$ for the HFCC computations. And $H_k(j)$, $k=1, 2, \dots, N$, 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.

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:

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

The constants a , b and c mentioned in (1) are expressed by the following values: $6.23 \cdot 10^{-6}$, $93.39 \cdot 10^{-3}$ 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:

$$\hat{a} = \frac{1}{2(700 + f_{low})}; \quad (5) \quad \hat{b} = \frac{700}{700 + f_{low}} \quad (6)$$

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

For the last filter, these coefficients are given by:

$$\hat{a} = -\frac{1}{2(700 + f_{high})}; \quad (8) \quad \hat{b} = -\frac{700}{700 + f_{high}} \quad (9)$$

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

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 Δf between the center frequencies of the filters adjacent is calculated by:

$$\Delta f = \frac{f_{CM} - f_{C1}}{M - 1} \quad (11)$$

The passage of $f_{C1} \rightarrow f_{C1}$ and $f_{CM} \rightarrow f_{CM}$ is given by the formula:

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

the center frequencies of adjacent filters are calculated by:

$$\hat{f}_{ci} = \hat{f}_{c1} + (i - 1) \cdot \Delta f \quad \text{for } i = 2, \dots, (M - 1) \quad (13)$$

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

$$f_{lowi} = -(700 + ERB_i) + \sqrt{(700 + ERB_i)^2 + f_{ci} \cdot (f_{ci} + 1400)} \quad (14)$$

$$f_{highi} = f_{lowi} + 2ERB_i \quad (15)$$

$$\text{With } ERB_i = \frac{1}{2} \cdot (f_{highi} - f_{lowi}) \quad (16)$$

2.3. 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 [24]. Thus the HNR is a measure examining the presence of noise during phonation. To calculate it, the signal is firstly downsampled 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.[25]

2.4. 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. [26] For example, you may have an increase in the DFA value when the voice fold is incompletely closed. [27]

3. 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). Table 1 summarizes the number of samples of pathological voices from each base.

3.1. 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 seconds 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 kHz or 50 kHz and 16 bits resolution. [11].

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

<i>Pathology</i>	<i>Female</i>		<i>Male</i>	
	<i>TEST</i>	<i>TRAIN</i>	<i>TEST</i>	<i>TRAIN</i>
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

3.2. 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.[21] [12]

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

3.3. Hidden Markov Model Toolkit

<i>Pathology</i>	<i>Female</i>		<i>Male</i>	
	<i>TEST</i>	<i>TRAIN</i>	<i>TEST</i>	<i>TRAIN</i>
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

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 [28]. 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. [10] 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.

4. Multiple class confusion matrix.

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.

4.1. Performance measures

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 actually 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. [29]

➤ **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:

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (14)$$

➤ **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:

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN}) \quad (15)$$

➤ **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:

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP}) \quad (16)$$

➤ **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. [30]

we discuss in this paper 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. [31] [32]

5. RESULTS AND DISCUSSIONS

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.

5.1. Overall recognition rate for the MEEI database

Table 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. Figure 3 below shows this.

Table 3: Overall Recognition Rate of pathologies with Four Gaussians for the MEEI Database

Parameters	ERB=1	ERB=2	ERB=3	ERB=4	ERB=5	ERB=6
HFCC	91,59	95,33	96,26	95,33	97,20	97,20
HFCC-F0	96,26	95,33	97,20	95,33	96,26	96,26
HFCC-NHR	94,39	95,33	96,26	96,26	97,20	99,07
HFCC-F0-NHR	93,46	93,46	97,20	99,07	98,13	94,39
HFCC-HNR-DFA	93,46	92,52	95,33	97,20	97,20	98,13

Fig 3. Performance of the extracted features on each ERB for MEEI database

5.2. Overall recognition rate for the Saarbruecken Voice Database

Table 4 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 $ERB=6$ (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 $ERB=6$. Figure 4 below shows this.

Table 4: Overall Recognition Rate with Eight Gaussians for the SVD Database

Parameters	ERB=1	ERB=2	ERB=3	ERB=4	ERB=5	ERB=6
HFCC	77.53	76.99	81.10	80	76.99	83.84
HFCC-NHR	78.36	82.47	83.56	82.19	79.18	87.40
HFCC-F0-DFA	85.86	77.30	75.66	81.25	78.29	79.93
HFCC-HNR-NHR	78.90	80.55	82.74	81.37	78.63	86.03
HFCC-HNR-DFA	78.90	77.53	79.45	84.66	82.47	80.82
HFCC-NHR-DFA	78.36	81.64	84.38	81.64	84.66	82.47

Fig 4. Performance of the extracted features on each ERB for the SVD database.

5.3. Pathology recognition rate of male and female voices for the MEEI database

5.3.1 Combining HFCC-NHR parameters

According to Table 3, the combination of the HFCC-NHR parameters gives the highest overall recognition rate for $ERB = 6$. This result is detailed in Table 5, 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 6 and 7 present the classification results for each type of pathology and for different ERB values for female and male voices.

Table 6 gives different performance results of the recognition system for the type of female voices and for the

combination of HFCC-NHR parameters. It can be seen that the accuracy varies between 83% and 100% for a variation in ERB between 1 to 6. Besides, it can be seen that the increase in the value of ERB improves the general performance of the recognition of pathologies.

Table 5: Pathology recognition rate for combining HFCC-NHR parameters with four Gaussians

		Pathology					
		Ventricular	Gastric	Edema	Paralysis	Hyperfunction	Normal
Female Voices	ERB=1	100	100	100	100	100	100
	ERB=2	100	100	100	100	100	100
	ERB=3	90	100	100	100	100	100
	ERB=4	100	80	100	100	100	100
	ERB=5	100	100	100	100	100	100
	ERB=6	100	100	100	100	100	100
Male voices	ERB=1	85.7	62.5	50	100	100	100
	ERB=2	85.7	62.5	75	100	100	100
	ERB=3	100	87.5	50	100	100	100
	ERB=4	100	87.5	50	100	100	100
	ERB=5	100	87.5	50	100	100	100
	ERB=6	100	87.5	100	100	100	100

Table 6: 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	100	100	100	100	100
	Precision	0.90	1	1	0.84	0.90	1
	Sensitivity	1	1	1	1	1	1
	Specificity	0.98	1	1	0.97	0.98	1
ERB=2	ACC (%)	100	100	100	100	100	100
	Precision	1	1	0.91	1	0.83	0.91
	Sensitivity	1	1	1	1	1	1
	Specificity	1	1	0.98	1	0.97	0.98
ERB=3	ACC (%)	90	100	100	100	100	100
	Precision	0.9	1	1	0.91	1	0.91
	Sensitivity	0.9	1	1	1	1	1
	Specificity	0.98	1	1	0.98	1	0.98
ERB=4	ACC (%)	100	80	100	100	100	100
	Precision	0.90	1	1	1	1	0.91
	Sensitivity	1	0.8	1	1	1	1
	Specificity	0.98	1	1	1	1	0.98
ERB=5	ACC (%)	100	100	100	100	100	100
	Precision	1	1	0.91	0.91	1	0.91
	Sensitivity	1	1	1	1	1	1
	Specificity	1	1	0.98	0.98	1	0.98
ERB=6	ACC (%)	100	100	100	100	100	100
	Precision	1	1	1	1	1	0.91
	Sensitivity	1	1	1	1	1	1
	Specificity	1	1	1	1	1	0.98

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.

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

Similarly, for male voices, the performance results of the recognition system for the combination of HFCC-NHR parameters are mentioned in table 7. 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).

Table 7: 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 (%)	85.7	62.5	50	100	100	100
	Precision	1	1	1	0.84	1	1
	Sensitivity	0.85	0.625	0.5	1	1	1
	Specificity	1	1	1	0.97	1	1
ERB=2	ACC (%)	85.7	62.5	75	100	100	100
	Precision	1	1	1	0.91	1	1
	Sensitivity	0.85	0.625	0.75	1	1	1
	Specificity	1	1	1	0.99	1	1
ERB=3	ACC (%)	100	87.5	50	100	100	100
	Precision	0.87	1	1	1	1	1
	Sensitivity	1	0.875	0.5	1	1	1
	Specificity	0.99	1	1	1	1	1
ERB=4	ACC (%)	100	87.5	50	100	100	100
	Precision	0.87	1	1	0.91	1	1
	Sensitivity	1	0.875	0.5	1	1	1
	Specificity	0.99	1	1	0.98	1	1
ERB=5	ACC (%)	100	87.5	50	100	100	100
	Precision	1	1	1	1	1	1
	Sensitivity	1	0.875	0.5	1	1	1
	Specificity	1	1	1	1	1	1
ERB=6	ACC (%)	100	87.5	100	100	100	100
	Precision	1	1	1	1	1	1
	Sensitivity	1	0.875	1	1	1	1
	Specificity	1	1	1	1	1	1

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.

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

5.3.2 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. Table 8 gives an overview of the different results.

Table 8: Pathology recognition rate for combining HFCC-F0-NHR parameters with four Gaussians

		Pathology					
		Ventricular	Gastric	Edema	Paralysis	Hyperfunction	Normal
Female voices	ERB=1	100	80	100	90.9	100	100
	ERB=2	100	80	100	100	100	100
	ERB=3	100	100	100	100	100	100
	ERB=4	100	100	100	100	100	100
	ERB=5	100	100	100	100	100	100
	ERB=6	90	100	100	100	100	100
Male voices	ERB=1	100	75	25	100	100	100
	ERB=2	100	75	25	100	100	100
	ERB=3	100	75	75	100	100	100
	ERB=4	100	87.5	100	100	100	100
	ERB=5	100	75	100	100	100	100
	ERB=6	100	75	25	100	100	100

Table 9: 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	80	100	90.9	100	100
	Precision	0.90	0.8	1	0.90	1	0.91
	Sensitivity	1	0.8	1	0.90	1	1
	Specificity	0.98	0.99	1	0.98	1	0.98
ERB=2	ACC (%)	100	80	100	100	100	100
	Precision	0.83	1	0.91	0.91	1	0.91
	Sensitivity	1	0.8	1	1	0.9	1
	Specificity	0.97	1	0.98	0.98	1	0.98
ERB=3	ACC (%)	100	100	100	100	100	100
	Precision	0.83	1	1	1	1	0.91
	Sensitivity	1	1	1	1	1	1
	Specificity	0.97	1	1	1	1	0.98
ERB=4	ACC (%)	100	100	100	100	100	100
	Precision	1	1	1	1	1	0.91
	Sensitivity	1	1	1	1	1	1
	Specificity	1	1	1	1	1	0.98
ERB=5	ACC (%)	100	100	100	100	100	100
	Precision	1	1	0.91	1	1	0.91
	Sensitivity	1	1	1	1	1	1
	Specificity	1	1	0.98	1	1	0.98
ERB=6	ACC (%)	90	100	100	100	100	100
	Precision	0.81	1	0.91	0.91	1	0.91
	Sensitivity	0.9	1	1	1	1	1
	Specificity	0.97	1	0.98	0.98	1	0.98

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

Table 10: 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	25	100	100	100
	Precision	1	1	1	0.84	0.92	1
	Sensitivity	1	0.75	0.25	1	1	1
	Specificity	1	1	1	0.97	0.98	1
ERB=2	ACC (%)	100	75	25	100	100	100
	Precision	1	1	1	0.91	0.92	1
	Sensitivity	1	0.75	0.25	1	1	1
	Specificity	1	1	1	0.98	0.98	1
ERB=3	ACC (%)	100	75	75	100	100	100
	Precision	1	1	1	1	1	1
	Sensitivity	1	0.75	0.75	1	1	1
	Specificity	1	1	1	1	1	1
ERB=4	ACC (%)	100	87.5	100	100	100	100
	Precision	1	1	1	1	1	1
	Sensitivity	1	0.875	1	1	1	1
	Specificity	1	1	1	1	1	1
ERB=5	ACC (%)	100	75	100	100	100	100
	Precision	1	1	1	1	1	1
	Sensitivity	1	0.75	1	1	1	1
	Specificity	1	1	1	1	1	1
ERB=6	ACC (%)	100	75	25	100	100	100
	Precision	1	1	1	0.91	1	1
	Sensitivity	1	0.75	0.25	1	1	1
	Specificity	1	1	1	0.98	1	1

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

5.4. Pathology recognition rate of male and female voices for the SVD database for Combining HFCC-NHR parameters

According to Table 3, the best pathology recognition rate for the combination of HFCC-NHR parameters is obtained for female voices. Table 11 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 11: Pathology recognition rate for combining HFCC-NHR parameters with eight Gaussians for the SVD database

		Pathologies				
		Hyperfunction	laryngitis	Polyp	Spasmodic	Normal
Female voices	ERB=1	67.3	36.8	28.6	78.6	92.9
	ERB=2	67.3	73.7	28.6	92.9	92.1
	ERB=3	69.1	100	14.3	85.7	92.9
	ERB=4	56.4	68.4	28.6	100	92.9
	ERB=5	69.1	100	28.6	85.7	83.5
	ERB=6	90.9	78.9	42.9	92.9	96.1
Male voices	ERB=1	26.7	70.4	33.3	87.5	92.9
	ERB=2	53.3	74.1	66.7	87.5	91.7
	ERB=3	40	59.3	77.8	75	97.6
	ERB=4	60	74.1	55.6	87.5	96.4
	ERB=5	40	55.6	22.2	75	98.8
	ERB=6	46.7	77.8	44.4	62.5	94

Table 12: 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 (%)	67.3	36.8	28.6	78.6	92.9
	Precision	0.75	0.77	0.66	0.91	0.83
	Sensitivity	0.67	0.58	0.4	0.78	0.92
	Specificity	0.96	0.97	0.99	0.99	0.90
ERB=2	ACC (%)	67.3	73.7	28.6	92.9	92.1
	Precision	0.82	0.63	1	0.61	0.91
	Sensitivity	0.67	0.73	0.28	0.92	0.92
	Specificity	0.97	0.97	1	0.97	0.95
ERB=3	ACC (%)	69.1	100	14.3	85.7	92.9
	Precision	0.80	0.67	0.5	0.52	0.96
	Sensitivity	0.69	1	0.14	0.85	0.92
	Specificity	0.95	0.97	0.99	0.96	0.98
ERB=4	ACC (%)	56.4	68.4	28.6	100	92.9
	Precision	0.83	0.76	1	0.5	0.90
	Sensitivity	0.56	0.68	0.28	1	0.92
	Specificity	0.98	0.98	1	0.96	0.94
ERB=5	ACC (%)	69.1	100	28.6	85.7	83.5
	Precision	0.86	0.70	1	0.32	0.96
	Sensitivity	0.69	1	0.28	0.85	0.84
	Specificity	0.98	0.97	1	0.93	0.98
ERB=6	ACC (%)	90.9	78.9	42.9	92.9	96.1
	Precision	1	1	0.75	0.61	0.97
	Sensitivity	0.90	0.78	0.42	0.92	0.96
	Specificity	1	1	0.99	0.97	0.98

Table 12 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 sufficient to give a final conclusion on the effectiveness of the recognition system, and therefore these measures should be supplemented by the ROC curve as shown in Figure 9. The latter shows that the best performance is obtained with the values of the highest equivalent rectangular bandwidth (ERB) (AUC = 0.9704).

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

Table 13 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 Figure 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.

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

Table 13: 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 (%)	26.7	70.4	33.3	87.5	92.9
	Precision	0.26	0.57	1	1	0.89
	Sensitivity	0.26	0.70	0.33	0.875	0.92
	Specificity	0.96	0.95	1	1	0.96
ERB=2	ACC (%)	53.3	74.1	66.7	87.5	91.7
	Precision	0.57	0.64	0.46	1	0.93
	Sensitivity	0.53	0.74	0.66	0.875	0.91
	Specificity	0.98	0.96	0.98	1	0.98
ERB=3	ACC (%)	40	59.3	77.8	75	97.6
	Precision	0.35	0.76	0.58	1	0.92
	Sensitivity	0.4	0.59	0.77	0.75	0.97
	Specificity	0.97	0.98	0.98	1	0.97
ERB=4	ACC (%)	60	74.1	55.6	87.5	96.4
	Precision	0.75	0.64	0.55	1	0.88
	Sensitivity	0.6	0.74	0.55	0.875	0.96
	Specificity	0.99	0.96	0.98	1	0.96
ERB=5	ACC (%)	40	55.6	22.2	75	98.8
	Precision	0.85	0.68	0.66	0.75	0.79
	Sensitivity	0.4	0.55	0.22	0.75	0.98
	Specificity	0.99	0.97	0.99	0.99	0.92
ERB=6	ACC (%)	76.7	77.8	44.4	62.5	94
	Precision	0.875	0.51	1	0.55	0.89
	Sensitivity	0.46	0.77	0.44	0.625	0.94
	Specificity	0.99	0.94	1	0.98	0.96

6. 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. The first database consists of a number of test samples of 105 records of the first 12 seconds of the Rainbow Passage spoken by men and women (58 female patients and 47 male patients) whom they represent 5 types of pathologies (ventricular, gastric, edema, paralysis and hyperfunction). The second database consists of 4 types of pathologies (hyperfunction, laryngitis, polyp and spasmodic); these samples are recordings of the phrase "Guten Morgen, wie geht es Ihnen?" ("Good morning, how are you?") by men and women. The second database consists of 4 types of pathologies (hyperfunction, laryngitis, polyp and spasmodic); these samples are recordings of the phrase "Guten Morgen, wie geht es Ihnen?" ("Good morning, how are you?") pronounced by male and female patients. There are 365 records in this database divided into 143 samples from male patients and 222 samples from female patients. Recognition rates varied from the database to database with the same combination of prosodic parameters. The best overall recognition rates we obtained were 99.07% and 87.40% for the samples taken in MEEI, SVD, respectively. 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.FCC-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.

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.

Author's contributions

The three authors contributed equally to the realization of this work. The author(s) read and approved the final manuscript.

Authors' information

Rabeh HAMDI first author, PhD student in the Signal Processing and Electrical systems laboratory, Science Faculty of Tunis, University of Tunis El Manar, Tunisia., Salah HAJJI Assistant professor, Teaching researcher in the Laboratory of Nanomaterials and Systems for Renewable Energies. ENERGY RESEARCH AND TECHNOLOGY CENTER CRTEn, Borj -Cedria Technopark BP 95 2050, BEN AROUS H. LIF Tunisia, Adnene Cherif, Senior Professor and Director of Signal Processing and Electrical systems laboratory, Science Faculty of Tunis, University of Tunis El Manar, Tunisia.

Competing interests

The authors declare that they have no competing interests.

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.

Funding

Not applicable.

Acknowledgment

Not applicable.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

References

- [1] Wang, Jianglin, and Cheolwoo Jo. "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, 2007.
- [2] Mehta, Daryush D., and Robert E. Hillman. "Voice assessment: updates on perceptual, acoustic, aerodynamic, and endoscopic imaging methods." *Current opinion in otolaryngology & head and neck surgery* 16, no. 3 (2008): 211.
- [3] Mekyska, Jiri, Eva Janousova, Pedro Gomez-Vilda, Zdenek Smekal, Irena Rektorova, Ilona Eliasova, Milena Kostalova et al. "Robust and complex approach of pathological speech signal analysis." *Neurocomputing* 167 (2015): 94-111.
- [4] Al-nasheri, Ahmed, Ghulam Muhammad, Mansour Alsulaiman, and Zulfiqar Ali. "Investigation of voice pathology detection and classification on different frequency regions using correlation functions." *Journal of Voice* 31, no. 1 (2017): 3-15.
- [5] Eskidere, Ömer, and Ahmet Gürhanlı. "Voice disorder classification based on multitaper mel frequency cepstral coefficients features." *Computational and mathematical methods in medicine* 2015 (2015).
- [6] Hemmerling, Daria, Andrzej Skalski, and Janusz Gajda. "Voice data mining for laryngeal pathology assessment." *Computers in biology and medicine* 69 (2016): 270-276.
- [7] Muhammad, Ghulam, Mansour Alsulaiman, Zulfiqar Ali, Tamer A. Mesallam, Mohamed Farahat, Khalid H. Malki, Ahmed Al-nasheri, and Mohamed A. Bencherif. "Voice pathology detection using interlaced derivative pattern on glottal source excitation." *Biomedical signal processing and control* 31 (2017): 156-164.
- [8] Skowronski, Mark D., and John G. Harris. "Exploiting independent filter bandwidth of human factor cepstral coefficients in automatic speech recognition." *The Journal of the Acoustical Society of America* 116, no. 3 (2004): 1774-1780.
- [9] Ali, Zulfiqar, Mansour Alsulaiman, Ghulam Muhammad, Irraivan Elamvazuthi, Ahmed Al-nasheri, Tamer A. Mesallam, Mohamed Farahat, and Khalid H. Malki. "Intra-and inter-database study for Arabic, English, and German databases: do conventional speech features detect voice pathology?" *Journal of Voice* 31, no. 3 (2017): 386-e1.
- [10] Young, S. J., G. Evermann, M. J. F. Gales, T. Hain, D. Kershaw, X. Liu, G. Moore et al. "The HTK Book (for HTK version 3.4.1), Cambridge University." (2009).
- [11] MEEI Database, Disorder database model 4337. Massachusetts Eye and Ear Infirmary Voice and Speech Lab, Boston, MA. 2002.
- [12] Barry, W. J., and M. Pützer. "Saarbrücken Voice Database, Institute of Phonetics, Univ. of Saarland." (2016).
- [13] Mesallam, Tamer A., Mohamed Farahat, Khalid H. Malki, Mansour Alsulaiman, Zulfiqar Ali, Ahmed Al-Nasheri, and Ghulam Muhammad. "Development of the arabic voice pathology database and its evaluation by using speech features and machine learning algorithms." *Journal of healthcare engineering* 2017 (2017).
- [14] Al-Nasheri, Ahmed, Zulfiqar Ali, Ghulam Muhammad, and Mansour Alsulaiman. "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, 2014.
- [15] Amami, Rimah, and Abir Smiti. "An incremental method combining density clustering and support vector machines for voice pathology detection." *Computers & Electrical Engineering* 57 (2017): 257-265.
- [16] Dahmani, Mohamed, and Mhania Guerti. "Vocal folds pathologies classification using Naïve Bayes Networks." In 2017 6th international conference on systems and control (ICSC), pp. 426-432. IEEE, 2017.
- [17] Muhammad, Ghulam, Mohammed F. Alhamid, M. Shamim Hossain, Ahmad S. Almogren, and Athanasios V. Vasilakos. "Enhanced living by assessing voice pathology using a co-occurrence matrix." *Sensors* 17, no. 2 (2017): 267.
- [18] Martínez, David, Eduardo Lleida, Alfonso Ortega, Antonio Miguel, and Jesús 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*, pp. 99-109. Springer, Berlin, Heidelberg, 2012.
- [19] Grenier, B. "Évaluation de la décision médicale : Introduction à l'analyse médico-économique ». 3ème édition Broché – 2 Juillet 1999 Paris : Masson.
- [20] Bertrand, Daisy, Joel Fluss, Catherine Billard, and Johannes C. Ziegler. "Efficacité, sensibilité, spécificité : Comparaison de

- différents tests de lecture." *L'Année psychologique* 110, no. 2 (2010): 299-320.
- [21] Hamdi, Rabeh, Salah Hajji, and Adnane Cherif. "Voice Pathology Recognition and Classification using Noise Related Features." *INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS* 9, no. 11 (2018): 82-87.
- [22] Larsson Alm, Kevin. "Automatic Speech Quality Assessment in Unified Communication: A Case Study." (2019).
- [23] Ganchev, Todor. "Contemporary methods for speech parameterization." In *Contemporary Methods for Speech Parameterization*, pp. 1-106. Springer, New York, NY, 2011.
- [24] Grueber, Maude. "Effets de la stimulation sous-thalamique bilatérale sur la voix et la parole de patients parkinsoniens." PhD diss., Université de Lausanne, Faculté de biologie et médecine, 2011.
- [25] Teixeira, João Paulo, Carla Oliveira, and Carla Lopes. "Vocal acoustic analysis-jitter, shimmer and HNR parameters." (2013).
- [26] Tsanas, Athanasios. "Accurate telemonitoring of Parkinson's disease symptom severity using nonlinear speech signal processing and statistical machine learning." Ph.D diss., Oxford University, UK, 2012.
- [27] Little, Max A., Patrick E. McSharry, Stephen J. Roberts, Declan AE Costello, and Irene M. Moroz. "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection." *Biomedical engineering online* 6, no. 1 (2007): 23.
- [28] <http://htk.eng.cam.ac.uk/>
- [29] Zhu, Wen, Nancy Zeng, and Ning Wang. "Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations." *NESUG proceedings: health care and life sciences*, Baltimore, Maryland 19 (2010): 67.
- [30] Hajian-Tilaki, Karimollah. "Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation." *Caspian journal of internal medicine* 4, no. 2 (2013): 627.
- [31] Obuchowski, Nancy A. "Receiver operating characteristic curves and their use in radiology." *Radiology* 229, no. 1 (2003): 3-8.
- [32] Zhou, Xiao-Hua, Donna K. McClish, and Nancy A. Obuchowski. *Statistical methods in diagnostic medicine*. Vol. 569. John Wiley & Sons, 2009.

Figures

Figure 1

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

Figure 2

HFCC Implementation.

Figure 3

Performance of the extracted features on each ERB for MEEI database

Figure 4

Performance of the extracted features on each ERB for the SVD database.

Figure 5

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

Figure 6

ROC for classification by HFCC-HNR for MEEI database for male voices

Figure 7

ROC for classification by HFCC-F0-NHR for MELL database for female voices

Figure 8

ROC for classification by HFCC-F0-NHR for MEEI database for male voices

Figure 9

ROC for classification by HFCC-NHR for SVD database for female voices

Figure 10

ROC for classification by HFCC-NHR for SVD database for male voice