

# A New E-Health Tool for Early Identification of Voice and Neurological Pathologies by Speech Processing

Bouafif Lamia

ISTMT, LITI laboratory, Tunis Manar University,  
2092 Tunis, Tunisia

Ellouze Noureddine

LITI laboratory, Tunis Manar University,  
ENIT, BP 37, Tunis belvedere 1002, Tunisia

**Abstract**—The objective of this study is to develop a non-invasive method of early identification and classification of voice pathologies and neurological diseases by speech processing. We will present a new automatic medical diagnosis tool which can assist specialists in their medical diagnosis. The developed strategy is based on speech acquisition of the patient followed by audio features extraction, training and recognition by using the HTK toolkit. The computed parameters are compared to standard values from a codebook database. The experiments and tests are conducted by using the MEEI pathological database of KEY Pentax. The obtained results give good discrimination with a mean pathology recognition ratio about 95%. Finally, this E-Health application is helpful for the prevention of specific diseases and improving the quality of patient care as well as reducing the costs of healthcare.

**Keywords**—E-Health; voice disorder; HMM classification; feature extraction; MFCC; pathology recognition rate

## I. INTRODUCTION

Developments of non-invasive methods for voice pathology diagnosis were developed in order to assist medical staff and otolaryngologists to conduct objective and efficient diagnosis. At present a number of classic diagnostic tools are available on the market which were based on speech measurements and imaging analysis.

Many studies which used the speech features extraction succeeded to obtain acceptable discrimination ratio between normal and pathologic speakers. Some of them have achieved classification accuracies between 70% and 90% [1]. In fact, acoustic analysis allows estimating a large amount of long-term acoustic parameters such pitch, formants, jitter, shimmer, Amplitude Perturbation Quotient, Harmonics to Noise Ratio and Normalized Noise energy [2]. These features are very useful for characterizing speaker disorders especially if they are associated with MFCC or RASTA-PLP coefficients.

In other references, advances in speech processing have contributed to the identification of some neurological diseases (Parkinson, dyslexia, scleroses) by voice parameters analysis. The developed method is based on the determination of speech parameters of a speaker from a hardware interface and software for digital acquisition and processing of the speech signal [3]. In this research, we will develop a speech processing tool for clinical observation and detection of pathological. This interface is intended, not only for patients

but also for people currently using voice (singers, teachers). The methodology is very easy and is based on a voice recording. Then, the extracted speech parameters are applied as inputs to the famous HTK toolkit (HMM classifier). The results are compared with normal and pathological values for a detection and classification disease by using the famous MEII database [4].

## II. RELATED WORKS

During the last decade, several digital methods of pathological identification from speech processing have been used for the classification and early identification of some diseases. These strategies can be classified into three categories:

- The first method is based on the extraction of acoustic parameters and the search of new descriptors and metrics of quality, distortion and voice irregularities such as MFCC, RASTA, LPC coefficients, Jitter, Shimmer and Harmonic ratio. The MFCC parameters were considered in other numerous studies, such as [5] and [6]. In [5] subjects with nodules, edema and unilateral vocal fold paralysis were analyzed with not encouraging accuracy results (78%), while in [6] patients suffering from spasmodic dysphonia were selected.
- The second one is based on machine learning techniques such as SVM and LDA. Among several machine learning techniques existing in literature, Support Vector Machine (SVM) has been widely used in voice signal processing such as the work of L.Godino [7] and S.N. [8] with accuracy ratio of 86%.
- The third is a statistical method which uses Hidden Markov Models (HMM) or Gaussian mixtures (GMM). It is based on learning and testing procedures for voice recognition and classification. The learning procedure constitutes the codebook (database of speech models and parameters), hence the testing procedure consists the audio real time acquisition and recognition step.

For example, Emary [9] uses GMM algorithm on a very small subset of the SVD database containing 38 pathological and 63 healthy voices in order to identify neurological disorders.

A. The Studied Voice and Neurology Pathologies

Vocal fold pathologies can be classified as physical, neuromuscular, traumatic and psychogenic diseases. They affect the voice quality. Several voices, neurological, organic or genetic diseases are associated with speech disorders and dysfunctioning [10]. In fact, a voice disorder can generate a language disorder causing degradation of the voice and its intelligibility. These disorders can be divided into next classes:

- *Dysphonia*: It can be considered as an abnormality of the speech production and quality or a paralysis or a kind of laryngitis.
- *Dysarthria*: Is a speech disorder related to paralysis or to poor coordination of the muscles involved in the articulation. This disease has a neurological origin and conducts to dyslexia disease for children.
- *Aphasia*: Is a language disorder due to a lesion of the cerebral cortex. The patient no longer includes the meaning of words or can no longer be expressed [10].
- Sclerosis is a neurological disease in which affects the brain of a part of the brain and spinal cord [11]. It causes muscle weakness, trouble with sensation, coordination and speaking [12].
- Parkinson and Alzheimer are neurological diseases which affects the brain controls then body mechanisms and articulations.

B. Speech Aspects

The speech signal is full of physiological and acoustic parameters. It can inform as about the identity of the speaker, its health and even its emotional state.

The speech is characterized by its variability in amplitude and phase and its non-stationary behaviour. It is the result of a convolution between a phonation (glottis source) and an articulation (vocal tract). The source is characterized by the pitch  $F_0$ , yet the vocal tract is characterized by a formant structure which reflects the resonance of the vocal tract given by the formants ( $F_1, F_2, F_3, \dots$ ) [12]. For example, Fig. 1 and 2 represent an illustration of the waveform, spectrogram, pitch and formants parameters of speech and music signals.

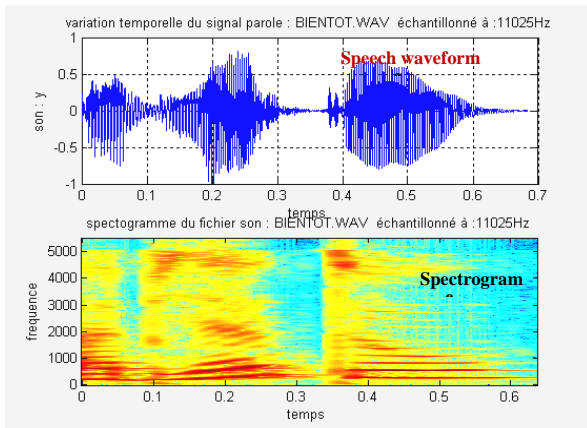


Fig. 1. Speech Waveform & Spectrogram.

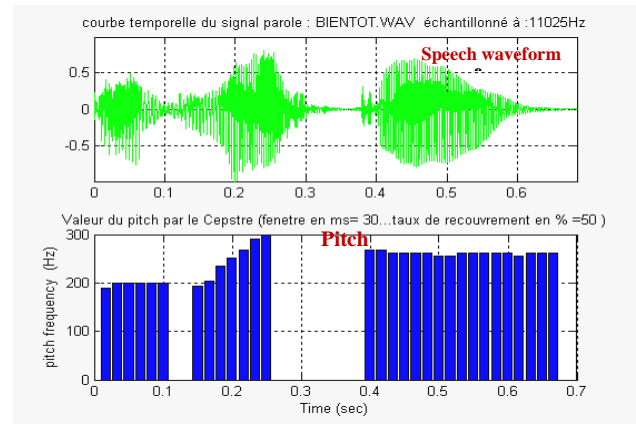


Fig. 2. Pitch Evolution of a Female Speech “Bientot.Wav” Sampled at  $F_s=11025\text{Hz}$ .

Fig. 1 and 2 represent the waveform, spectrogram and pitch profile of a female speech signal “bientot.wav” sampled at 11025Hz. According to Fig. 2, we can observe that mean pitch frequency is about 245Hz with a silent zone between 0.25 and 0.4 seconds.

The wide band spectrogram of Fig. 1 shows the formantic character of the speech illustrated by the red curves.

III. MATERIALS AND METHODS

In this work, we used the statistical HMM method, because it is very famous for speech recognition and synthesis and gives high accuracy for classification especially for high databases and noised environments. Other references used SVM, LDA and GMM classifiers [13, 14].

A. Speech Pathology Database

We have used the MEEI database of disordered voice (Kay Elemetrics Corporation) which was produced by the Kay Pentax [4]. The database is composed of many data dealing with the assessment of voice pathologies. It is considered as the most widely used dataset for research in pathological voice classification. The KAY database includes recordings of vowels pronounced by 53 normal subjects and 657 pathological voices coming from several diseases. The technical sheets are provided with the Recording and data files, containing information on the subjects (age, sex, language, smoking or not) and the results of the analysis calculated by the software MDVP.

TABLE I. CONTENT OF THE MEEI DATABASE

Pathology	Number of patients
Dysarthria	17
Gastric Refux	12
Ventricular compression	17
Nodule	6
Odeme	15
Disphonia	22
Normal	17

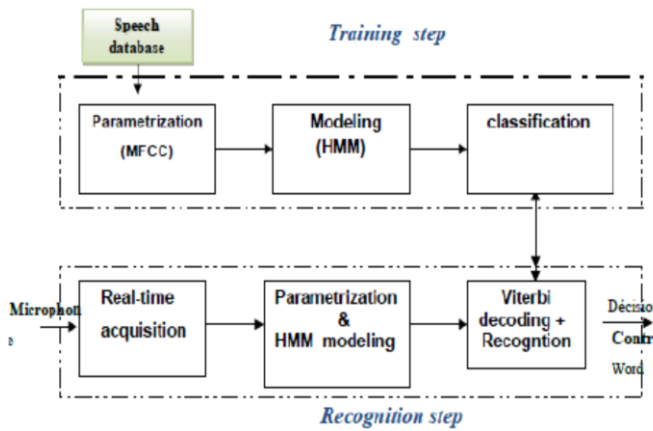


Fig. 3. The Training and Classification Method.

This software is also exclusively produced by Kay Pentax Corporation and is widely used in the clinical field as a tool for the recording and analysis of patient’s voice. The available pathologies are: the dysphonia, nodules, paralysis, polypoïde degeneration, and vocal cords disorder. These pathologies are recorded up to 10 seconds by men and women. Table I gives more details about the content of this database.

**B. Pathology Identification with HMM**

We have used the famous HTK platform based on the Hidden Markov Models (HMM) in order to recognize the pathological voices and a further disease classification. This

tool is a set of libraries and programs in C language developed at Cambridge University under the direction of Young in 1989 [15] in order to develop a performing technical Automatic Speech Recognition System. This toolkit is composed of:

- a speech database
- a training procedure by using the Baum-Walch and K-means algorithm for speech modeling. This step is applied on the speech database to constitute the reference codebook.
- a recognition procedure which is based on a real time acquisition and analysis , then a comparison with the training words by using Viterbi algorithm.

This procedure is illustrated by Fig. 3 where we can observe the different steps of parameterization (feature extraction), training, recognition and classification. In this step, the test audio model is compared with the codebook in order to find any similarity or coincidence with pathological models.

**C. Speech Features Extraction**

The first step of the speech analysis before modeling and coding is the parameterization of the speech frames into MFCC, LPC, PLP or RASTA coefficients. The Mel Frequency Cepstral Coefficients (MFCC) are the most famous method in speech processing, recognition and synthesis. Its principle is illustrated by Fig. 4.

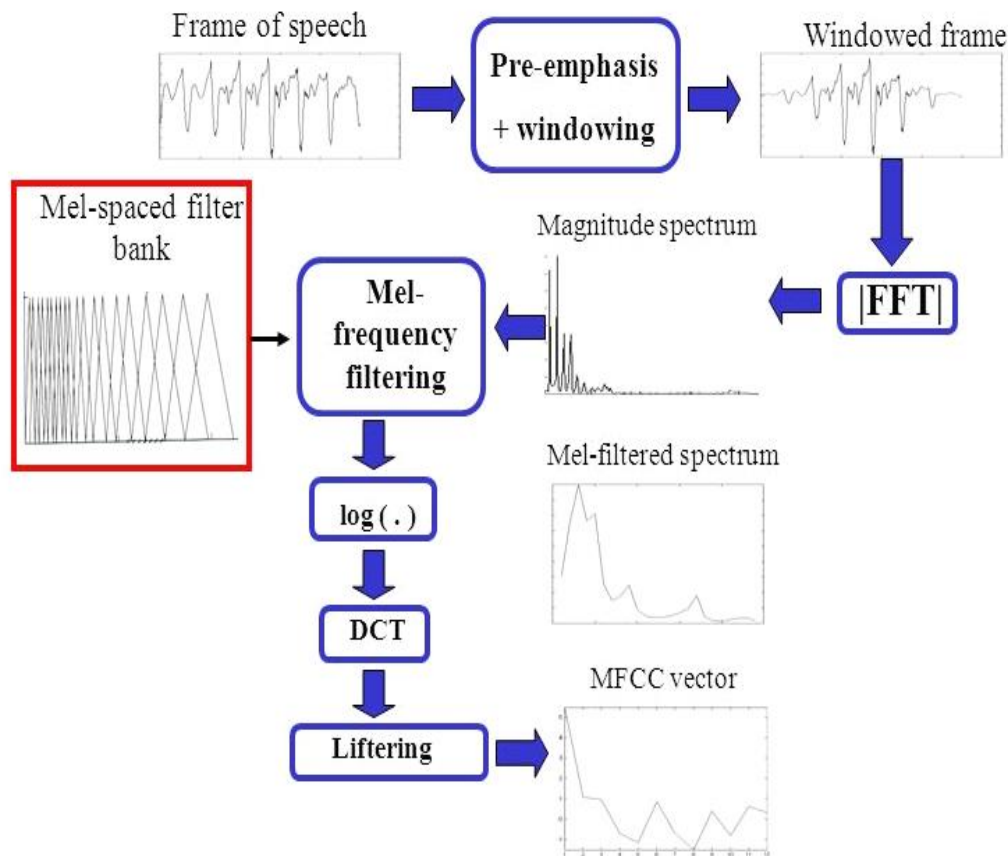


Fig. 4. MFCC Feature Algorithm.

In fact, MFCC is the most used for speech feature extraction and parameterization. MFCC algorithm which is represented by Fig. 2 can be expressed as [16]. The MFCC formula can be expressed as:

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

Where:  $E_k$ : is the energy of the  $k^{th}$  filter

$N$ : is the number of band-pass filters

Two others parameters are very useful in voice disorders analysis which is Jitter and shimmer. These indicators represent the irregularities and perturbations respectively in frequency and intensity. The expressions are given by next equations [16, 17]:

$$Jitter = \frac{\frac{1}{N-1} \sum_{k=1}^N |T_k - T_{k+1}|}{\frac{1}{N} \sum_{k=1}^N |T_k|} \quad (2)$$

$$Shimmer = \frac{\frac{1}{N-1} \sum_{k=1}^N |A_k - A_{k+1}|}{\frac{1}{N} \sum_{k=1}^N |A_k|} \quad (3)$$

With:

$T$ : the pitch period

$A$ : the amplitude of the pitch

$N$ : the number of samples.

$k$ : indices of the frame

#### D. HMM Training and Recognition

Fig. 5 represents the principle of the training-recognition-classification procedure. The training step uses a database or a codebook constituted of audio parameters and Baum-Walch and K-means algorithms. The recognition procedure uses HMM modeling and classification by using Viterbi decoder [18].

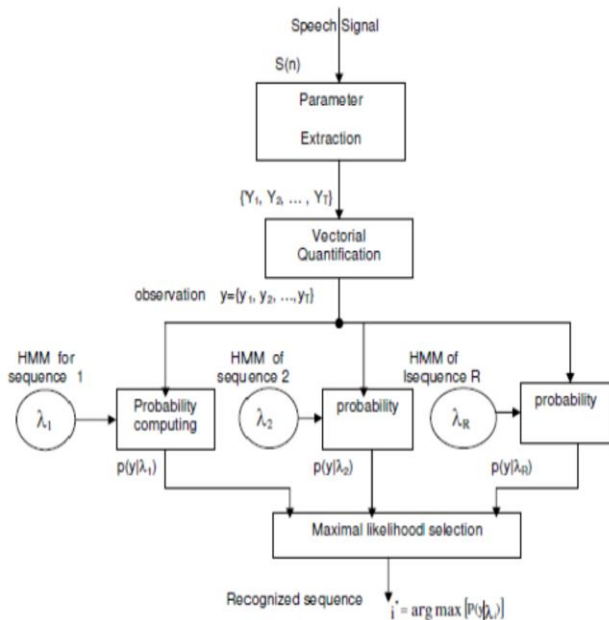


Fig. 5. HMM Algorithm.

Hidden Markov Models is a useful for data statistical modeling and classification.

The implementation of the HMM system requires three phases:

- Describe a network whose topology reflects the sentences, vocabulary words or basic units
- Make the training mode settings:  $\lambda = (\pi, A, B)$
- Carry out the actual recognition occurrence by calculating the maximum likelihood [14].

#### IV. SIMULATION RESULTS

Several platforms and software are used in speech processing such as, Praat, Vocalab, EDVP, Speech Analyser Matlab and HTK toolkit. These tools offer many parameters and indicators form speech evaluation such as, pitch, formants, Jitter, shimmer and SNR.

##### A. Effect on the Pitch

Pitch is the first indicator of the speech production as represents the period of the glottis signal. Fig. 6 shows the variations of the speech waveform, the zero-crossing, the pitch, the spectrogram, the spectrum and the formants of a normal speaker (without any disease). We can observe that the pitch (Fig. 6) is characterized by a continuous and constant profile with of value  $F1 = 210$  Hz (male speaker).

However, in the case of a pathological voice (organic or neurological origin), the speech profile presents distortions and dynamic variations around the pitch nominal value as illustrated in Fig. 7 to 10 which will be discussed later in details.

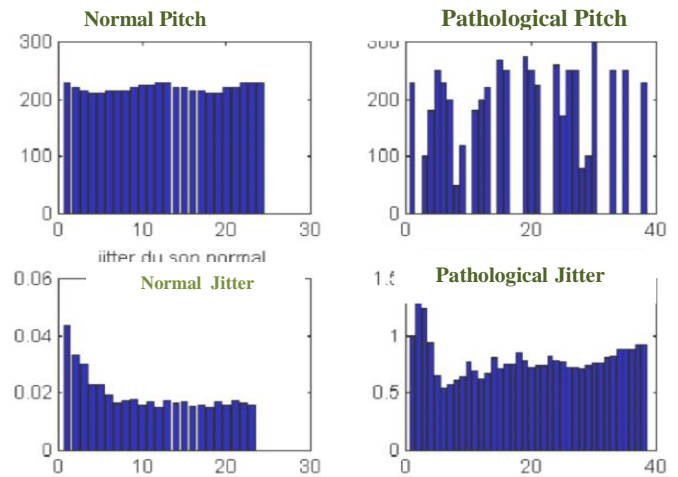


Fig. 6. Jitter Variation of Normal and Pathological Voice.

##### B. Effect on the Jitter and Shimmer

Disturbances of the durations of glottal cycles (Jitter) are irregularities in the period glottal signal. These disturbances are a basic phenomenon that is present in the voice and are therefore a feature of vocal timbre. This disturbance can be used to characterize spectrally hoarse voices, neurological, emotional or normal. The method is based on a study of the spectral effects of the glottal cycle's variance and the

evolution of the jitter values. On the other side, the study of perturbations of the amplitude (Shimmer) shows that they are a consequence of disturbances durations and energy. These mechanisms are asymmetries in the movement of both vocal cords and acoustic propagation of the glottal signal through the vocal tract [17]. Fig. 6 demonstrates that the normal jitter value is 0.02 (2%), hence the pathological value is over 0.8 (80%).

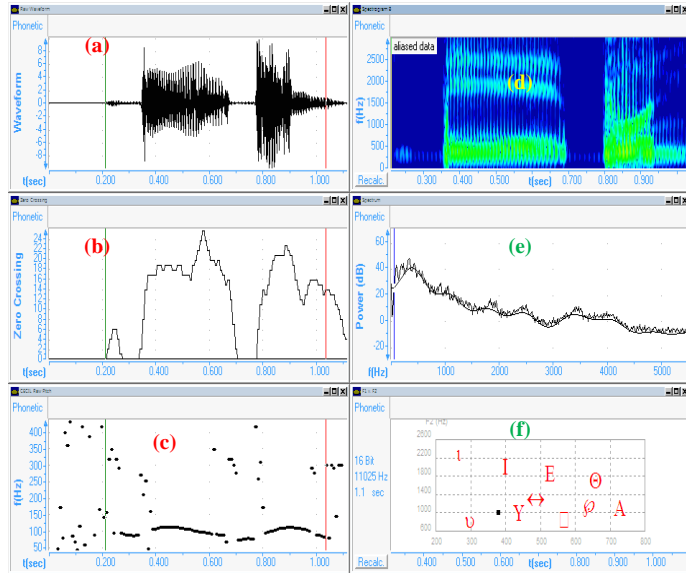


Fig. 7. (a) Speech, (b) Zero-Crossing, (c) Pitch, (d) Spectrogram, (e) Spectrum and (f) Formants Profiles (Case of a Normal Voice).

## V. DISEASE IDENTIFICATION AND CLASSIFICATION

### A. Multiple Sclerosis Disease

In this case of disease, the most common deficits affect recent memory, attention, processing speed, speech, visual-spatial abilities and executive function.

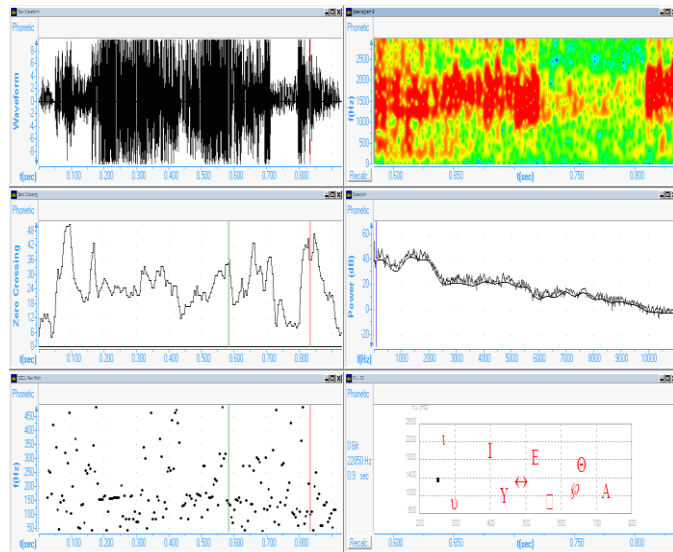


Fig. 8. Speech, Zero-Crossing, Pitch, Spectrogram, Spectra and Formants Profiles. (Case of pathological voice: Multiple sclerosis).

Symptoms related to cognition include emotion, instability and fatigue including neurological fatigue [18]. The following speech profiles and parameters of Fig. 8 (pitch, formants) demonstrated a correlation between the speech features and this pathology.

In fact, we can observe in Fig. 8 that in this case, the pitch profile becomes very disturbed with a high standard deviation contrarily with normal and safety speaker of Fig. 7. This state indicates a dysfunctioning of the speech production system which is monitored by the brain.

### B. Dyslexia

It is considered a cognitive disorder but it does not affect intelligence. Problems may include difficulties in spelling words, reading quickly, writing words, "sounding out" words in the head [19], [20]. The examination of the speech of figure 9 shows disturbances of the pitch curve at the level of the co-articulations and the changes of vowel consonants in the pronounced word. These variations remain around 20% of the nominal value, but the standard deviation remains almost 12%. Also, the wide band and narrow band spectrograms are affected by this variation and disturbance.

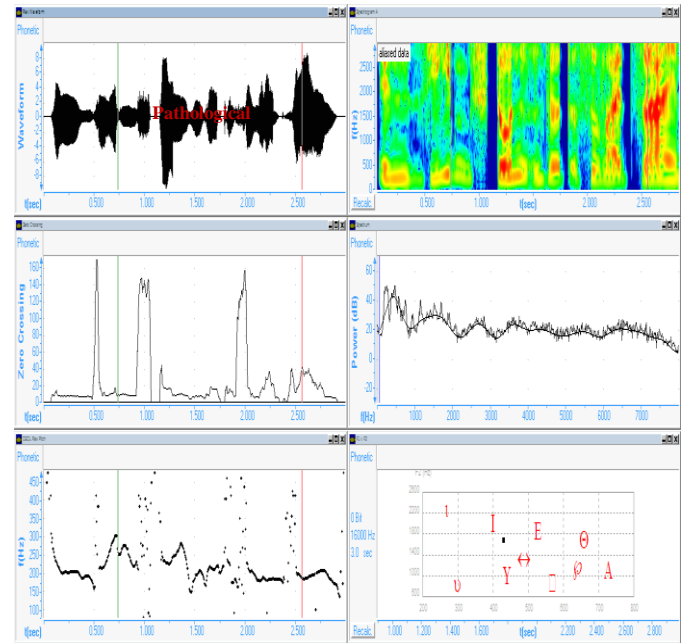


Fig. 9. Speech, Zero-Crossing, Pitch, Spectrogram, Spectra and Formants Profiles. (Case of pathology: Dyslexia).

### C. Alzheimer

It is a chronic neurodegenerative disease that usually starts slowly and worsens over time. It is the cause of 60% to 70% of cases of dementia. The most common early symptom is short-term memory loss), problems with language, speech, disorientation, mood swings, loss of motivation and behavioral issues [21]. Recent research studies demonstrate relations between speech production and Alzheimer disease. Fig. 10 illustrates the speech parameters of a speaker suffering from Alzheimer (age: 72).

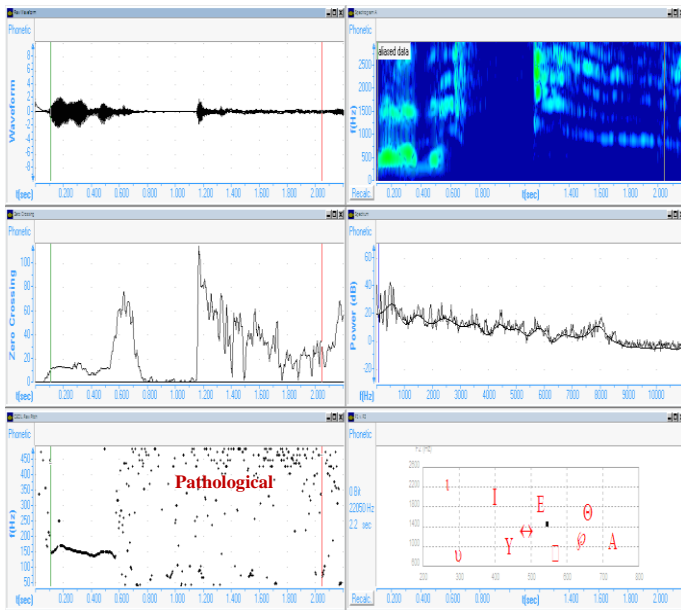


Fig. 10. Speech, Zero-Crossing, Pitch, Spectrogram, Spectra and Formants Profiles.

(Case of pathology: Alzheimer)

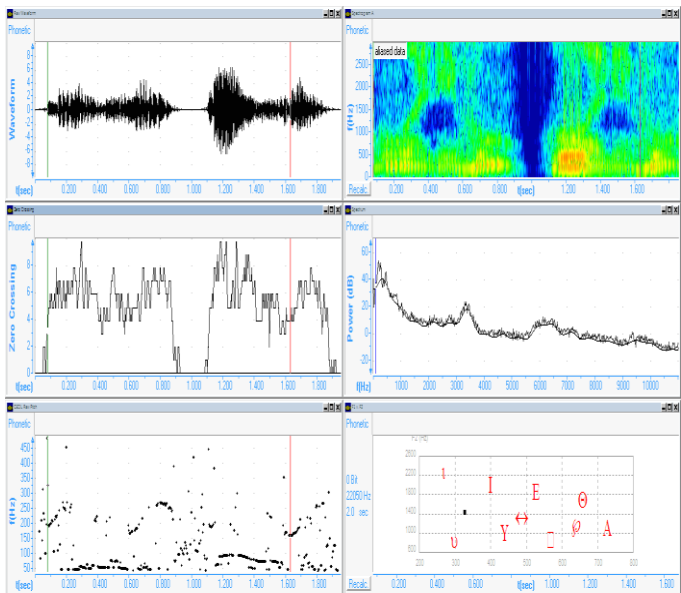


Fig. 11. Speech, Zero-Crossing, Pitch, Spectrogram, Spectra and Formants Profiles. (Case of pathology: Parkinson).

The examination of speech shows a loss of controls of the glottis after the pronunciation of the first word. No value of the pitch can be extracted due to the complete dispersion of the acquired values. The spectrogram of the same figure confirms these results and dot give us any information about the value of neither pitch nor the formants.

#### D. Parkinson

This neurology disease has a long-term degenerative disorder of the central nervous system that mainly affects the motor system. The symptoms generally come on slowly over time. At a first step, the most obvious are shaking, rigidity,

slowness of movement, and difficulty with walking. Thinking and behavioral problems may also occur. Dementia becomes common in the advanced stages of the disease. Depression and anxiety are also common occurring, including sensory, difficulty of speaking, sleep, and emotional problems [22]. The examination of the speech through Fig. 11 shows high deviations and variations of the pitch (glottis signal) rather large of 100% which alters the language understanding and loses the speech recognition of the speaker.

#### E. Pathology Recognition Ratio

All the described procedures and steps (acquisition, training, feature extraction, recognition and pathology classification) are embedded and inserted in a smart interface illustrated in Fig. 12.

Our tests are compared to the MEEI of KAY Pentax speech database, described in the last paragraph, on the HTK toolkit. According to Table II, we obtained a pathology recognition ratio (RR) between 86% and 100%.

These values are very interesting because they can discriminate several diseases from the normal voices characterized by a 100% value.

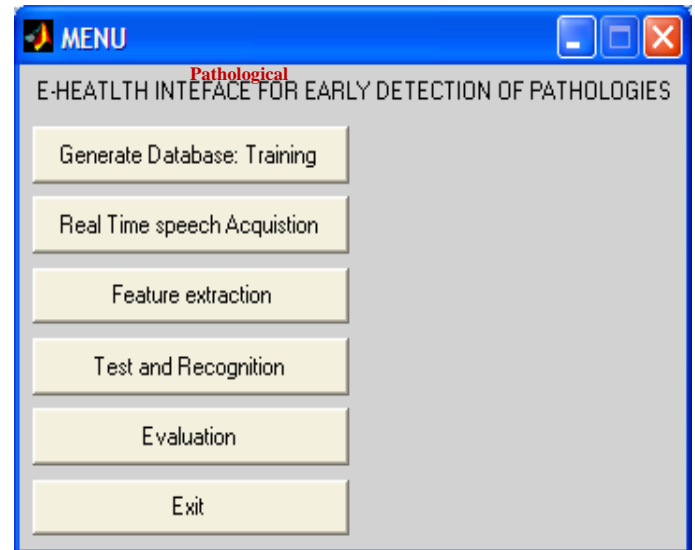


Fig. 12. Illustration of the E-Health Interface.

TABLE II. PATHOLOGY RECOGNITION RATIO OF SEVERAL SPEECH DISEASES

Disease	Recognition Ratio (RR) in %
Dysarthria	94
Gastric Refux	97
Ventricular compression	86
Nodule	92
Odeme	90
Disphonia	98
Normal	100

## VI. DISCUSSION

The obtained Pathology identification ratios demonstrate that we obtain high (RR) values for both hyper-function (dysarthria) and paralysis diseases (dysphonia) respectively 94% and 98%. Besides, we compared our results with other studies using similar and different databases.

Table III compares our proposed algorithm with previous significant works [5, 6, 7, 8, 23, 24]. Although in these works the databases are different, it is observed that the proposed algorithm with HMM appears competitive and has a high accuracy to identify pathological and normal voices. We succeeded at the first step of our work to identify more than 8 kinds of organic and neurological diseases.

TABLE III. COMPARISON OF PERFORMANCES OF SEVERAL PATHOLOGY STUDIES

Reference	Database	Code book	Feature extraction	Classifier	Pathology identification
Godino	MEII	226	MFCC	SVM	94
Fonseca	MEII	154	MFCC	GMM	78
Souissi	Sharbrok	110	MFCC+D+DD	SVM	86
Boyanov	Private	400	Energy	LDA	93
<b>Our method</b>	<b>MEII</b>	<b>657</b>	<b>MFCC+ Jitter + Shimmer</b>	<b>HMM</b>	<b>86 to 100</b>

## VII. CONCLUSION

In this paper, we developed a new tool dedicated to identification and diagnosis of vocal and neurological diseases. The method is based on analysis of acoustic parameters of a patient after a real time speech acquisition and processing. The modeling and classification procedures are automated by using HMM, training and recognition procedures. The validation was carried out thanks to the pathological famous database MEEI of Pentax. The obtained recognition ratio of the pathology is around 95%. The most significant indicators of the pathological speech are disturbances in amplitude (Shimmer) and frequency distortion and irregularities of pitch (Jitter) and finally the loss of glottis control (high standard value of the pitch). Besides, this application allows us to follow changes in the physiological state (heartbeat, blood pressure, ECG) and acoustic parameters (pitch, formants, timbre) and then we can compare them with normal and standard values. This is very interesting because it helps us to follow the disease evolution, to predict and to avoid patient complication and to improve his re-education therapy.

The following step of this study is to extend this application to other critical diseases such as cancer and Hepatics C and then to evaluate it through a large number of patients.

## ACKNOWLEDGMENTS

This work was supported by the National Institute of Biomedical Studies of Tunis. The authors declare no conflict of interest.

## ABBREVIATIONS

HMM : Hidden Markov Models

GMM : Gaussian Mixture Model

MFCC : Mel Frequency Cepstral Coefficients

LPC : Linear Predictive Coding

RR: Pathology recognition rate:

HTK : HMM Toolkit

RASTA-PLP : Perception Linear coefficients

SVM : Support Vector Machines

LDA : Linear Discriminate Algorithm

## REFERENCES

- [1] Maguire C., de Chazal P., Reilly R.B., and Lacy P. Automatic Classification of voice pathology using speech analysis. In *Proceedings of the World Congress on Biomedical Engineering and Medical Physics*, Sydney, 2003.
- [2] Boyanov, B. and Hadjitodorov, S. Acoustic analysis of pathological voices. A voice analysis system for the screening of laryngeal diseases. *IEEE Engineering in Medicine & Biology Magazine* 16:74-82, 1997.
- [3] Manfredi, C., D'Aniello, M., Brusciagioni, P. and A. Ismaelli. A comparative analysis of fundamental frequency estimation methods with application to pathological voices. *Medical Engineering and Physics* 22:135-147, 2000.
- [4] Kay Elemetrics Corp. *Disordered Voice Database and Program*, Model 4337. Lincoln Park, NJ, Kay Elemetrics Corp. Canada 2010.
- [5] H. Cordeiro, J. Fonseca, I. Guimarães, and C. Meneses, "Voice pathologies identification speech signals, features and classifiers evaluation," in *Proc. Signal Process., Algorithms, Architect., Arrangements, Appl. (SPA)*, Sep. 2015, pp. 81–86.
- [6] F. Amara, M. Fezari, and H. Bourouba, "An improved GMM-SVM system based on distance metric for voice pathology detection," *Appl. Math*, vol. 10, no. 3, pp. 1061–1070, 2016.
- [7] J. I. Godino-Llorente, P. Gómez-Vilda, N. S-Lechón, M. Blanco-Velasco, F. Cruz-Roldán, and M. A. Ferrer-Ballester, "Support vector machines applied to the detection of voice disorders," in *Proc. Int. Conf. Nonlinear Anal. Algorithms Speech Process.*, vol. 3817. 2005, pp. 219–230.
- [8] N. Souissi, C. Adnen, "Dimensionality reduction for voice disorders identification system based on mel frequency cepstral coefficients and support vector machine," in *Proc. 7th Int. Conf. Modelling, Identificat. Control (ICMIC)*, Dec. 2015, pp. 1–6.
- [9] I. M. M. El Emary, M. Fezari, and F. Amara, "Towards developing a voice pathologies detection system," *J. Commun. Technol. Electron.*, vol. 59, no. 11, pp. 1280–1288, 2014.
- [10] L. Verde, G.D. Pietro, G. Sannino. *Voice Disorder Identification by Using Machine Learning Techniques*. *IEEE Access journal*, Vol 6, April 2018. DOI: 10.1109/ACCESS.2018.2816338
- [11] NINDS Multiple Sclerosis Information Page". *National Institute of Neurological Disorders and Stroke*. November 2015.
- [12] Nakahara J, Maeda M, Aiso S, Suzuki N. "Current concepts in multiple sclerosis: autoimmunity versus oligodendroglipathy." *Clinical reviews in allergy & immunology*. 42 (1):26-34. 2012. DOI :10.1007/s12016-011-8287-6.
- [13] J. Wang and C. Jo, "Vocal folds disorder detection using pattern recognition methods," in *Proc. 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBS)*, Aug. 2007, pp. 3253–3256.
- [14] R. T. Ritchings, M. McGillion, and C. J. Moore, "Pathological voice quality assessment using artificial neural networks," *Med. Eng. Phys.*, vol. 24, nos. 7–8, pp. 561–564, Sep. 2002.
- [15] Young S. J., Woodland P. C., Byrne W. J., "HTK Reference Manual for HTK Version 3.1", 2001.

- [16] L. Bailly. Interaction between the vocal cords and ventricular strips in phonation: exploration in vivo, physical modeling, validation in-vitro. Doctoral thesis, University of Nantes, France, 2009.
- [17] B. Teston. The objective assessment of dysphonia: current state and prospects of evolution.. Report of the Laboratory speech and language, University of Provence, CNRS France, 2011.
- [18] Bobholz J, Rao S. "Cognitive dysfunction in multiple sclerosis: a review of recent developments". *Current Opinion in Neurology*. **16** (3), 2003. doi:10.1097/00019052-200306000-00006.
- [19] Phillips, Sylvia; Kelly, Kathleen; Symes, Liz . *Assessment of Learners with Dyslexic-Type Difficulties*. SAGE. p. 7, 2013. ISBN 978-1-4462-8704-0.
- [20] Campbell, Robert Jean. *Campbell's Psychiatric Dictionary*. Oxford University Press, 2009, pp.310–312. ISBN 978-0-19-534159-1.
- [21] Querfurth HW, LaFerla FM (28). "Alzheimer's disease". *The New England Journal of Medicine* **362** (4):329–44. January 2010. doi:10.1056/NEJMra0909142.
- [22] Sveinbjornsdottir, S. "The clinical symptoms of Parkinson's disease.". *Journal of Neurochemistry*. **139**: 318–324. July 2016. doi:10.1111/jnc.13691.
- [23] B. Boyanov and S. Hadjitodorov, "Acoustic analysis of pathological voices. A voice analysis system for the screening of laryngeal diseases," *IEEE Eng. Med. Biol. Mag.*, vol. 16, no. 4, pp. 74–82, Jul. 1997.
- [24] S. Jothilakshmi, "Automatic system to detect the type of voice pathology," *Appl. Soft Comput.*, vol. 21, pp. 244–249, Aug. 2014.