Spectral Envelope and Periodic Component in Classification Trees for Pathological Voice Diagnostic

Cordeiro H. *IEEE Student Member*, Fonseca J. and Meneses C.

Abstract—**This work investigates the effectiveness of features from the spectral envelope such as the frequency and bandwidth of the first peak obtained from a 30th order Linear Predictive Coefficients (LPC) to identify pathological voices. Other spectral features are also investigated and tested to improve the recognition rate. The value of the Relative Power of the Periodic Component is combined with spectral features, to diagnose pathological voices. Healthy voices and five vocal folds pathologies are tested. Decision Tree classifiers are used to evaluate which features have pathological voice information. Based on those results a simple Decision Tree was implemented and 94% of all the subjects in the database are correctly diagnosed.**

I. Introduction

Voice pathologies appear due to various circumstances, such as the extensive or incorrect use of voice, stress, inhalation of tobacco smoke, gastric reflux or hormonal problems. These pathologies typically affect the vocal folds and are detectable by direct laryngoscopy, which is the visualization of the vocal folds using a camera. This method is invasive, uncomfortable for the patient and may, depending on the equipment used, require a local anesthetic. Alternatively, the test can be performed by indirect laryngoscopy, using a mirror which, although less invasive, often involves the use of small amounts of local sedatives. This equipment is also expensive and has maintenance costs as it needs to be sterilized between diagnoses, consuming time and resources.

The use of an efficient, non-invasive, easy and quick method for pathological voices identification may be useful as an initial evaluation or as a complementary method for the diagnosis of voice pathologies. A method of pathological voices diagnosis based in speech signals is useful in screening situations because it dos not involve specialized equipment. In this context the use of speech processing techniques can be applied and exploited for pathological voices identification.

Speech production is associated to the source-filter model [1]. The source is related with the vocal folds that are responsible for the production of a periodic signal in voiced regions and the filter that is related with the vocal tract. This work deals with vocal folds pathologies, which affect naturally the source and the periodicity.

Several features such as pitch jitter, normalized autocorrelation of the residual signal in pitch lag, shimmer and noise measures like HNR (Harmonic to Noise Ratio), are used to identify pathological voices [2-4]. In [5], a summary of several studies about the identification of pathological voices is presented, using source related features but also filter related feature such as MFFC (Mel-Frequency Cepstrum Coefficients) and LP (Linear Predictive) coefficients. In [6], critical band energies from spectra are analyzed to recognize differences between normal and pathological voices.

In [7,8], pathological voice recognition is accomplished using MFCC. This feature is related with the vocal tract and perceptual information and applies state-of-the-art classifiers such as SVM (support vector machines). These studies have 98% accuracy but the use of complex features and high level classifiers makes difficult to understand how the pathologies affect the speech signal.

In previous work [9], the LPC spectrum was inspected to find features that can be used in voice pathology detection, using nodules and Reinke's Edema as pathologies, from a database, already used in other works [10,11]. For unhealthy subjects, a first spectral peak was observed before the typical first vowel formant. To observe this feature a high order LPC must be applied. The no using of pre-emphasis filter also contributes to the highlight of this feature. This first peak can be also observed in some healthy subjects, but with higher bandwidth. Only based on that feature it was possible to discriminate all the unhealthy subjects of the database.

The main objective of this work is to verify if the first peak frequency is also present in a larger database and if it is also discriminant of other vocal folds pathologies. Other features are also searched and tested to improve the recognition rate. The database used is the Massachusetts Ear and Ear Infirmary (MEEI) voice disorder database [12]. Five vocal folds pathologies are tested. A Decision Tree is applied to diagnose the pathologies. Decision Tree are easy to understand and are constructed using the importance of the features, based on the information gain [13]. In this context it is possible to understand which features have more pathological voice information and create an improved version, based on rules, of the decision tree.

The next section describes the database used in this work. Section 3 describes the features used. Section 4 describes the implemented systems and the results obtained. This paper ends with conclusions and future work.

H. Cordeiro is with the Department of Electrical Engineering, Faculty of Sciences and Technology of New University of Lisbon (FTC-UNL), Caparica, Portugal, and also with the Department of Electronics and Telecommunications and Computers, High Institute of Engineering of Lisbon (ISEL), Lisbon, Portugal(Tel.: +351 218 317 224; fax: +351 218 317 114; *E-mail address:* hcordeiro@deetc.isel.ipl.pt.)

J. Fonseca is with the Department of Electrical Engineering, Faculty of Sciences and Technology of the New University of Lisbon (FTC-UNL), Caparica, Portugal, (e-mail: jmf@uninova.pt).

C. Meneses Ribeiro is with the Department of Electronics and Telecommunications and Computers, High Institute of Engineering of Lisbon (ISEL), Lisbon, Portugal (e-mail: cmeneses@deetc.isel.ipl.pt).

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

II. Database Description

The MEEI voice disorder database [12] is composed by 53 healthy subjects and 700 unhealthy subjects, most of them not labeled. This work uses the 53 healthy subjects and 153 unhealthy subjects diagnosed with vocal folds pathologies, as showed in Table I.

TABLE I. DATABASE DESCRIPTION

Pathology	Number of Subjects
Healthy	53
Nodules	15
Edema	37
Paralasys	65
Polyps	15
Keratosis\Leukoplakia	21

Almost all the unhealthy subjects in this set have more than one pathology, such as hiperfunction, gastric reflux and various types of A-P squeezing.

The healthy subjects recorded the vowel /a/ for 3 seconds and the unhealthy subjects for about 1 second. These audio files have sampling rates of 50 kHz and 25 kHz. Files sampled with 50 kHz were downsampled to 25 kHz.

The database was divided in two sets, one for train and other for test. Each set has about 50% of the healthy subjects and 50% of each pathology subjects.

III. Features

A. LPC Spectrum Features

The spectral envelope has a visual representation of the frequencies and bandwidths of the formants and is estimated with the LPC filter frequency response, also known as LPC spectrum. The formant central frequency (1) is given by the LPC poles angle w_k and the bandwidth (2) is given by the pole distance r_k to the unitary circle [14].

$$F_k = \frac{w_k}{2\pi T_S} \quad (1)$$

$$B_k = -\frac{\ln(r_k)}{\pi T_S} \quad (2)$$

The LPC spectrum provides a spectral envelope of the signal spectrum. This spectral envelope has information of the resonant frequencies of the vocal tract represent by peaks in the spectral envelope. Typically an order of 10-16 is used for frequency samples of 8 kHz in telephone bandwidth speech applications. This assumes that the vocal tract have four resonant frequencies or four formants. In this work the order of the LPC has 30th in order to characterize other peaks in the spectral envelope.

As illustrated in Fig. 1, a first spectral peak was found before the normal first formant when a high order LPC was used. From the LPC spectrum the frequency and bandwidth of the first and second peak was evaluated. The first peak in healthy subjects is typically the first formant. However, in the unhealthy subjects the typical first formant of the vowel is the second peak in the 30th LPC spectrum.

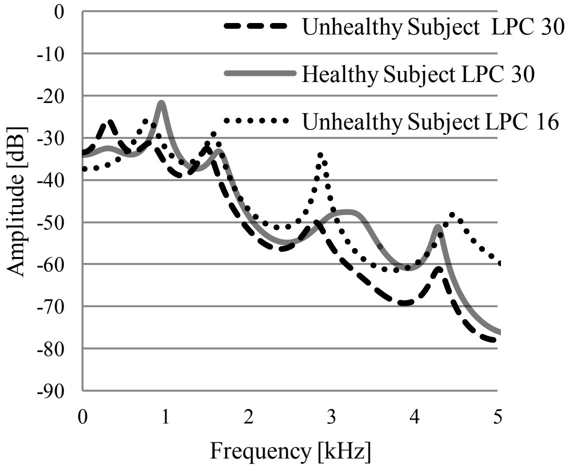

Figure 1. Healthy and Unheathly subjects LPC spectrum [9].

Fig. 2 shows that the frequency of the first peak, *F1*, of the LPC spectrum for unhealthy subjects is typically lower than for healthy subjects. The value of the first formant of the vowel /a/ is about 650 Hz and it can be seen that most of the unhealthy subjects have a lower peak than 500 Hz. Although classes were not completely separated it is possible to verify the existence of two patterns. Some of the healthy subjects have as well a lower first peak frequency, but normally with larger bandwidth, *BW1*. This accomplishes one of the main objectives of this work that was, to confirm if that peak can also be found in this database containing more subjects and pathologies than in [9].

However, some of the unhealthy subjects do not have the first peak in lower frequency. In almost all these cases, it was observed that the difference between the second and first frequency peak is bigger than for healthy subjects allowing discrimination. Fig. 3 presents this new feature, the difference between the second and the first peak of the spectrum, *F2-F1*, relative to the first peak.

B. Relative Power of the Periodic Component.

The Relative Power of the Periodic Component (RPPC), described in [15], taken from the normalized correlation for the pitch lag, is also evaluated. This feature has information about the HNR and the jitter, showing a higher value in more periodical speech signals such as healthy voices and smaller value in the presence of harmonic noise and/or jitter as in the pathological voices.

To estimate the value of the periodic component it is important to have the same number of pitch periods for all the subjects. Thus, the autocorrelation is estimated twice. The first estimation using a 30 ms window is used to compute the period of the fundamental frequency. Then six times the pitch period window is used to compute the *RPPC*.

Fig. 4 presents the mean frequency of the first peak of LPC relative to mean Relative Power of the Periodic Component. It can be seen that no healthy subjects have a value smaller than 0.965.

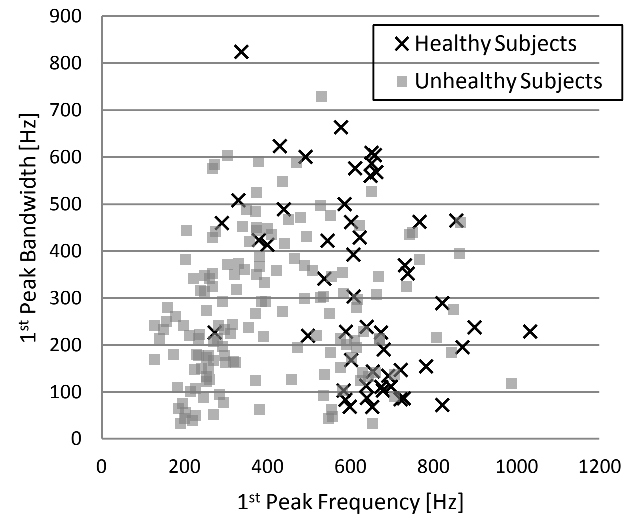

Figure 2. Mean LPC spectrum first peak frequency vs. mean bandwidth for all the subjects in the corpus.

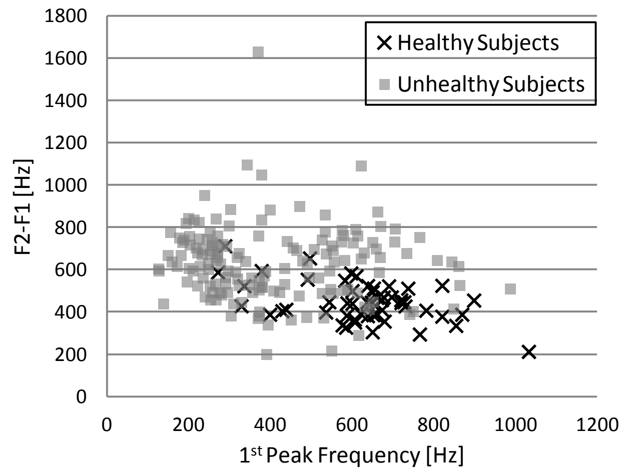

Figure 3. Mean LPC spectrum first peak frequency vs. mean diference of the second and the first peak frequency for all the subjects in the corpus.

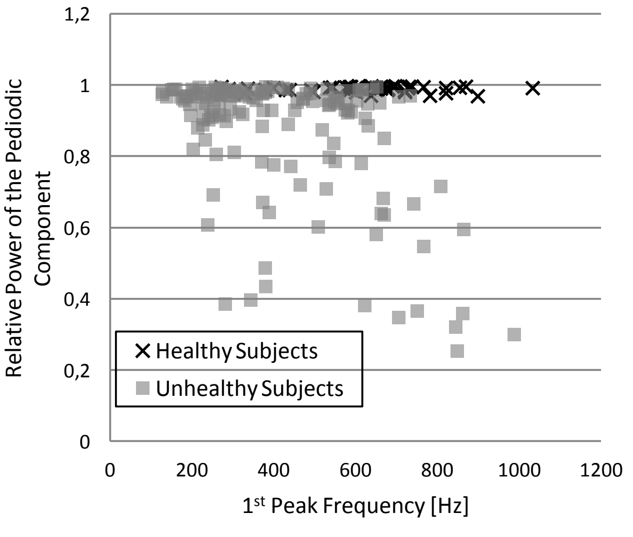

Figure 4. Mean LPC spectrum first peak frequency vs. mean relative periodic component power for all the subjects in the corpus.

IV. Implemented System and Results

For each subject the entire file is computed using 30 ms frames with 10 ms overlap. The files have about three seconds for healthy subjects and one second for unhealthy subjects. LPC coefficients are computed with 30th order.

The classification system is implemented with a decision tree. The Matlab Statistics Toolbox was used for the tree training using the Gini's Diversity Index method. For the tree training, the average of all frames for each subject was used. For testing, each frame is evaluated by the classifier and the subject is classified in the class assigned with more frames.

Table II shows the tree classification results for each pair mentioned in Fig. 2-4 for the train and test sets. The last row shows the result with all the features, achieving an accuracy of 86.4%, smaller than the 89.3% obtained with just *F1* and *RPPC*, as shown in Fig. 4.

The main idea of these tests was to have an approximation of the features thresholds to construct a rule based tree, inspired on Fig. 2-4. As shown in Fig. 4, all the subjects with *RPPC* below 0.96 are unhealthy subjects. Therefore, this value can be used on the first node, with 76% of the unhealthy subjects correctly identified. The remaining 24% unhealthy subjects and all the healthy subjects can now be separated measuring the spectral envelope features: Subjects with *F1* smaller than 500 Hz are considered unhealthy, unless the bandwidth is bigger than 700 Hz. In this case the subject is considered healthy (Fig. 1); Subjects with *F1* bigger than 500 Hz are considered healthy, unless *F2-F1* is bigger than 600 Hz, in which the subject is considered unhealthy (Fig. 2). The resulting tree, that achieves an accuracy of 94.2% (see also the first row of Table III), is presented in Fig. 5. To compare with the results of the training and test set used in the automatic tree, presented in Table II, the same subsets were evaluated in the Improved Tree.

If the *RPPC* node is removed from the tree, the overall accuracy rate will be 87.4% using only the features of the spectral envelope. *F1* alone can achieve an overall accuracy rate of 77.2%. The use of the first peak bandwidth allows an impact of more five healthy subjects correctly classified and the *F2-F1* adds more eight unhealthy subjects correctly classified, demonstrating the impact of the proposed feature.

As an initial evaluation, or as a complementary method for the diagnosis of voice pathologies, it is desirable to have less unhealthy subject errors even if the overall accuracy decrease by increasing the healthy subjects errors. With the *F1* threshold set to 600 Hz maintaining the remaining tree, no errors in the test set in the unhealthy subjects were found. However the overall accuracy decreases to 90.6%, as shown in the second row of Table III. In the tests set, the accuracy for healthy voices decreases to 70%.

Another important figure when evaluate a two classes recognizer is the value correspondent to equal diagnostic rate between the classes. In this case the value of *RPPC* changed to 0.97 and the value of *F1* to 350 Hz. The equal error rate is 9.3% and the overall accuracy rate is, in this case, 90.7%.

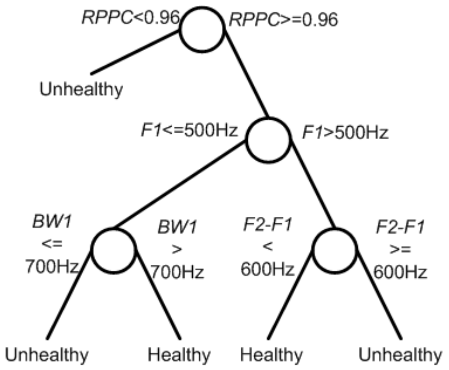

Figure 5. Improved Tree.

TABLE II. DIAGNOSTIC RATE FOR AUTOMATIC TREE

Features	Train Set		Test Set		Test Accy	Overall Accy
	H	UH	H	UH		
F1 and *BW1*	9/27	74/77	8/26	72/74	80%	79,1%
F1 and *RPPC*	24/27	73/77	19/26	68/74	87%	89,3%
F1 and *F2-F1*	19/27	66/79	18/26	71/74	89%	84,4%
All	18/27	74/77	18/26	68/74	86%	86,4%

TABLE III. DIAGNOSTIC RATE OF THE IMPROVED TREE

Features Values	Train Set		Test Set		Test Accy	Overall Accy
	H	UH	H	UH		
Fig. 5	23/27	73/77	24/26	72/74	96%	94,2%
F1 600 Hz	20/27	73/77	18/26	74/74	92%	90,6%
RPPC=0,97 F1=350 Hz	24/27	68/77	24/26	69/74	93%	90,8%

H stands for Healthy subjects and HU for Unhealthy subjects

V. CONCLUSION

This work studies the impact of features taken from the spectral envelope, when computed with a high order LPC. The use of a 30thLPC showed, in previous work of the same authors, the presence of a peak before the first formant in subjects with nodules and Reinke's Edema. This work applied the same method to the MEEI voice disorder database and confirmed the existence of this characteristic in three other pathologies of the vocal folds.

It was also found in this work that the difference between the second and the fist peak, have impact in the diagnosis of unhealthy subjects. With the Relative Power of the Periodic Component and the features extracted from the spectral envelope (first peak frequency, first peak bandwidth and the difference between the frequencies of second and first peak) a decision tree was constructed. This classifier obtained a recognition rate of 94.2% on the overall results.

In future work other high level classifiers will be tested to discriminate unhealthy subjects and pathologies.

ACKNOWLEDGMENT

The authors would like to thank Prof. Ana Mendes from Polytechnic Institute of Setúbal for sharing the database used in this work.

REFERENCES

[1] G. Fant. "Acoustic Theory of Speech Production". *Mouton The Haugue*, 1970.

[2] S. Iwata, "Periodicities of pitch perturbations in normal and pathological larynges", *J. Acoust. Soc. Amer*. Vol. 45, pp. 344-353, 1972.

[3] S. B. Davis, "Acoustic characteristics of normal and pathological voices". *Speech and language: advances in basic research and practice*,1, pp. 271-335, 1979

[4] V. Parsa and D. G Jamieson, "Identification of pathological voices using glottal noise measures." *Journal of Speech, Language & Hearing Research*,*43*.2, 2000

[5] N. Sáenz-Lechón, J. I. Godino-Llorente, V. Osma-Ruiz, P. Gómez-Vilda, "Methodological issues in the development of automatic systems for voice pathology detection", *Biomedical Signal Processing and Control*, Volume 1, Issue 2, pp. 120-128, April 2006

[6] K. Shama, A. Krishna and N. U. Niranjan Cholayya. "Study of harmonics-to-noise ratio and critical-band energy spectrum of speech as acoustic indicators of laryngeal and voice pathology." *EURASIP Journal on Advances in Signal Processing*, Volume 2007, 9 pages, 2007

[7] J.D Arias-Londoño.; J.I. Godino-Llorente; N. Sáenz-Lechón; V. Osma-Ruiz; G. Castellanos-Domínguez; "Automatic Detection of Pathological Voices Using Complexity Measures, Noise Parameters, and Mel-Cepstral Coefficients," *Biomedical Engineering, IEEE Transactions on*, vol.58, no.2, pp.370-379, Feb. 2011

[8] X. Wang, J. Zhang and Y. Yan; , "Automatic Detection of Pathological Voices Using GMM-SVM Method," *2nd International Conference Biomedical Engineering and Informatics*, Oct. 2009

[9] H. Cordeiro, J. Fonseca and C. Meneses, "LPC Spectrum First Peak Analysis for Voice Pathology Detection", *HCIST 2013 - International Conference on Health and Social Care Information Systems and Technologies*, Lisbon , Oct. 2013

[10] E. S. Fonseca , R. C. Guido, P. R. Scalassara, C. D. Maciel and J. C. Pereira , ''Wavelet time-frequency analysis and least squares support vector machines for the identification of voice disorders,'' *Comput. Biol. Med.*, vol. 37, pp. 571–578, 2006.

[11] P. R. Scalassara, M. E. Dajer, Maciel C. D., R. C. Guido, J. C. Pereira, "Relative entropy measures applied to healthy and pathological voice characterization", *Applied Mathematics and Computation*, Volume 207, Issue 1, pp. 95-108, Jan. 2009.

[12] Kay Elemetrics, *Elemetrics Disordered Voice Database (Version 1.03)*, 1994

[13] L. Rokach and O. Maimon, "Top-down induction of decision trees classifiers-a survey". *IEEE Transactions on Systems, Man, and Cybernetics*, Part C 35 (4): pp. 476–487, 2005

[14] B.S. Atal, and S.L. Hanauer, "Speech synthesis by linear prediction of the speech wave", *Journal of the Acoustical Society of America*, vol. 50, pp. 637-655, 1971.

[15] P. Boersma, "Accurate short-term analysis of the fundamental frequency and the harmonic-to-noise ratio of a sample sound." IFA Proceedings 1993; 17, pp. 97-110.