Pathological Voice Detection by Cepstral Analysis Using Multiple Classifiers

Raissa Tavares[1], Nathália Brunet[2], Suzete Correia[3], Silvana Cunha Costa[4]
Electrical Engineering Coordination
Federal Institute of Technological Education - IFPB
João Pessoa, Brazil
[1]raissa@ieee.org, [2]nathaliabrunet@gmail.com, [3]suzete@ifpb.edu.br, [4]silvana@ifpb.edu.br

Benedito G. Aguiar Neto[5], Joseana Macêdo Fechine[6]
Computer and Electrical Engineering Center
Federal University of Campina Grande - UFCG
Campina Grande, Brazil
[5]bganeto@dee.ufcg.edu.br,[6]joseana@dsc.ufcg.edu.br

Abstract—This paper evaluates cepstral classifiers applied to pathological voice detection problem. The goal is analyze the individual and combined performance of classifiers based on cepstral, weighted cepstral, delta cepstral, and weighted delta cepstral parameters. They are evaluated considering two different combination strategies yielding a multiple classifier that is more efficient than either individual technique. The efficiency rates obtained vary from 87% using stand alone weighted delta cepstral to 98% considering the classifiers combination.

Keywords-Acoustic signal analysis, pathological voices, cepstral analysis, multiple classifiers.

I. INTRODUCTION

The diagnosis of laryngeal pathologies is usually made by laryngoscopical exams, which are considered invasive, causing discomfort to patients. Digital signal processing techniques performing an acoustic analysis for vocal quality assessment is a simple and noninvasive measurement procedure. These techniques provide an objective diagnosis of pathological voices, and may be used as complementary tool in laryngoscopical exams [1].

Some researchers have dedicated efforts for obtaining efficient methods for discriminating normal and pathological voices using acoustic analysis [1]-[6]. However, the research for a more detailed and representative acoustic analysis of pathological voice signals is still a promising area.

In this work, a parametric analysis based on cepstral analysis is employed to discriminate pathological voices of speakers affected by vocal fold edema. Cepstral (CEP), weighted cepstral (WCEP) delta cepstral (DCEP), and weighted delta cepstral (WDCEP) parameters are used as features to detect the irregularities of the pathological voices in comparison with the normal voice. A vector quantization technique (VQ) was used associated with a distortion measurement to classify the speech signal by each parameter. The VQ was trained with voices affected by the considered pathology and the results will be used to build an effective method basis for detecting pathological voices.

To improve the performance of the cepstral classifiers, an approach based on multiple classifiers is evaluated. For that, two combination rules are considered: the combination by average and the combination by product, which are modifications of the strategies used in [7].

II. AN OVERVIEW ON THE CEPSTRAL ANALYSIS

Considering that speech signal is the result of convolving excitation with vocal tract sample response, by cepstral analysis, it is possible to separate these two components. All the cepstrum-related features described are obtained after Linear Predictive Coding (LPC) analysis [8],[14].

The linear prediction method estimates each speech sample based on a linear combination of the p previous samples; a larger p enables a more accurate model. It provides a set of speech parameters that represent the vocal tract [8]. It is expected that any change in the anatomical structure of the vocal tract, because of pathology, affects the LPC coefficients and also the cepstral and its derivatives. A linear predictor with prediction coefficients, $\alpha(k)$, is defined as a system whose output is

$$\tilde{s}(n) = \sum_{k=1}^{p} \alpha(k) s(n-k), \qquad (1)$$

where p is the predictor order. The autocorrelation method and the covariance method are standard for computing the predictor coefficients. The prediction coefficients are computed using the Levison-Durbin algorithm [8].

Cepstral derivatives can improve the representation of the spectral speech properties. Pathological speech presents significant spectral differences of normal voices. The cepstral coefficients can be calculated recursively from the linear predictor coefficients, $\alpha(k)$, by means of [9]:

$$\begin{cases} c(1) = -\alpha(1) \\ c_i(n) = -\alpha(n) - \sum_{j=1}^{n-1} (1-\frac{j}{n})\alpha(j)c(n-j) & 1<n\leq p \end{cases} . \qquad (2)$$

Cepstral coefficients obtained by (2) provide a good measure of the difference in the spectral envelope of the speech frames. These coefficients are used in order to observe the information of voice transitions in pathological speech signal versus normal speech [10].

The first derivative of the cepstral coefficients (Delta Cepstral Coefficients) is given by [10]:

$$\frac{\Delta c(n,t)}{\Delta t} = \Delta c_i(n) \approx \phi \sum_{k=-K}^{K} kc(n,\ t+k), \tag{3}$$

where $c(n,t)$ is the n-th LP coefficient at time t, ϕ is a normalization constant and $2K+1$ is the number of frames over which the computation is performed.

The delta cepstral coefficients are obtained as a simplified version of (3), as it was proposed by [10]:

$$\Delta c_i(n) = [\sum_{q=-K}^{K} kc_{i-q}(n)]G, \quad 1 \le n \le p, \tag{4}$$

where G is a gain term (for example, 0.375), p is the number of delta cepstral coefficients, $K=2$, n the coefficient index and i the frame of analysis [12].

In order to account for the sensitivity of the low-order cepstral coefficients to overall spectral slope and the sensitivity of the high-order cepstral coefficients to noise, cepstral weighting (liftering) is employed .

The weighted cepstral coefficients (WCEP), $cw_i(n)$, are obtained by [10]-[12]:

$$cw_i(n) = c_i(n) \cdot w(n). \tag{5}$$

The type of window used in this work was the band pass liftering (BPL), given by [10]:

$$w(n) = \begin{cases} 1 + \frac{L}{2} sin(\frac{n\pi}{L}), & n = 1,2,...,L \\ 0, & \text{otherwise.} \end{cases}, \tag{6}$$

where L is the size of the window. The BPL weights a cepstral sequence by (6) so that the lower- and higher-order components are de-emphasized.

Weighted Delta Cepstral coefficients (WDCEP) are obtained replacing (4) in (5), resulting on

$$\Delta cw_i(n) = \Delta c_i(n) \cdot w(n). \tag{7}$$

The characteristics of weighted cepstral and delta cepstral are associated by using (7).

III. DATABASE AND METHODS

The database was recorded by the Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab [14]. The following cases were selected: 44 patients presenting vocal fold edema - 33 women (17 to 85 years old) and 11 men (23 to 63 years old), most of them (32) with bilateral edema and 53 patients with normal voices which is composed of 21 male (26 to 59 years old), and 32 female (22 to 52 years old).

In the pre-processing stage speech signals are multiplied by a 20 (ms) Hamming window with an overlap of 50%. A filter of pre-emphasis (0.95) is also used. Then each parameter is calculated after LP coefficients (p=12).

To dimensionality reduction of data, a Vector Quantization (VQ) technique [15] is used that is associated with a distortion measurement. The quantization is carried out individually for each feature using just voices under vocal fold edema in the training phase. Thus, different VQ-trained distance classifiers [12] are obtained by the discrimination process. The VQ-classifiers are applied to static feature vectors, which are computed for every analysis frame of the speech samples over a dynamic input sustained vowel /a/. It is used 50% of vocal fold edema cases in the training phase. To the test phase, the other 50% of voices signals under vocal fold edema, and all the normal (53) voices are used. After the feature extraction, a codebook is generated using the Euclidean distortion measurement and the nearest neighbour rule is used to find the codevector. LBG algorithm to quantization and the least mean square distance for classification process are used [16].

IV. RESULTS AND DISCUSSION

The measurements used to evaluate the performance of the methods are the following: Correct acceptance (CA) rate; Correct rejection (CR) rate, False acceptance (FA) rate; False rejection (FR) rate; and the Efficiency representing the correct classification of a given class when that is present, given by E(%)= (CR+CA)/(CR+CA+FA+FR) x 100 [1].

Table I shows the results obtained for each parameter individually. It can be seen that Delta Cepstral (DCEP) method gives the best Efficiency and False Acceptance rates. However, this method presents a higher False Rejection rate compared to cepstral (CEP) method.

TABLE I. PERFORMANCE EVALUATION OF THE INDIVIDUAL FEATURES.

Classifiers	CR (%)	CA (%)	FA (%)	FR (%)	E (%)
CEP	89	91	11	9	90
WCEP	94	86	6	14	90
DCEP	98	86	2	14	92
WDCEP	91	82	9	18	87

To evaluate the combined features, makes the assumption that a speech signal must be assigned to one of the K possible classes and assume that L classifiers are available. The distortion measurement used by the ith QV-classifier is denoted as d_i. Two combination rules have been employed:

- Combination by average:

$$D = \frac{1}{K} \sum_{i=1}^{L} d_i, \tag{8}$$

- Combination by Product:

$$D = \prod_{i=1}^{L} \mathrm{d}_i, \tag{9}$$

where D denotes the distortion obtained after combination.

In order to guarantee an standardization of the that the classifier outputs, the distortion values of each VQ-classifier were normalized (values between 0 and 1). A threshold of D is chosen such as the best separation between the classes is obtained.

The results for average and product combination are presented in Tables II and III, respectively. The results show that, for almost all combinations, the False Rejection rates decrease significantly. In the average rule (Table II), the probability in rejecting the presence of vocal fold edema (14%) in the individual case (DCEP) that gives the best efficiency (Table I) is about 2% when combining the four parameters.

TABLE II. PERFORMANCE EVALUATION – THE AVERAGE RULE

Classifiers	CR (%)	CA (%)	FA (%)	FR (%)	E (%)
CEP and DCEP	92	95	8	5	94
CEP and WDCEP	94	95	6	5	95
CEP and WCEP	96	95	4	5	96
DCEP and WDCEP	98	82	2	18	90
DCEP and WCEP	94	93	6	7	94
WDCEP and WCEP	92	95	8	5	94
CEP, DCEP and WDCEP	94	91	6	9	93
CEP, DCEP and WCEP	96	95	4	5	96
CEP, WDCEP and WCEP	96	98	4	2	97
DCEP, WDCEP and WCEP	94	93	6	7	94
CEP, DCEP, WDCEP and WCEP	94	98	6	2	96

The best result is obtained using combination by product of CEP and WDCEP classifiers (Table III). It can be observed an improvement of at least 6% in efficiency rate, comparing with the DCEP individual classifier. For this case, the probability in detecting the presence of the edema pathology when, in real, it is not present (FA), is null.

TABLE III PERFORMANCE EVALUATION THE PRODUCT RULE

Classifiers	CR (%)	CA (%)	FA (%)	FR (%)	E (%)
CEP, DCEP	100	93	0	7	97
CEP, WDCEP	100	95	0	5	98
CEP, WCEP	92	98	8	2	95
DCEP, WDCEP	96	86	4	14	91
DCEP, WCEP	94	95	6	5	95
WDCEP, WCEP	92	98	8	2	95
CEP, DCEP and WDCEP	94	95	6	5	95
CEP, DCEP and WCEP	94	95	6	5	95
CEP, WDCEP and WCEP	94	98	6	2	96
DCEP, WDCEP and WCEP	94	93	6	7	94
CEP, DCEP, WDCEP and WCEP	74	98	26	2	86

V. CONCLUSIONS

In this paper the individual and combined performance of classifiers based on cepstral, weighted cepstral, delta cepstral, and weighted delta cepstral parameters were evaluated for the pathological voice detection problem. The results show that combination of these classifiers can yield a significant performance improvement related to individual ones. The best efficiency rate in the individual case was 92% and after the combinations, about 98%. This mean that the parameters employed are complementary and can be used to detect vocal disorders caused by the presence of vocal fold pathologies. Future work will focus in the use of others combination rules, such as an weighted average, for example, and in the use of other classifiers, such as Neural Network and/or Hidden Markov Models. Furthermore, the system performance can be tested with other laryngeal pathologies.

REFERENCES

[1] J. I. Godino-Llorente, P. Gomes-Vilda and M. Blanco-Velasco, “Dimensionality Reduction of a Pathological Voice Quality Assessment System Based on Gaussian Mixture Models and Short-Term Cepstral Parameters”, IEEE Trans. on Biom. Engineering, Vol. 53, No. 10, pp. 1943-1953, October, 2006.

[2] K. Shama, A. Krishna, and N. U. 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, Vol. 2007, 2007.

[3] P. J. Murphy and Olatunji O. Akande, “Noise Estimation in Voice Signals Using Short-term Cepstral”, J. of the Acoust. Society of America, pp. 1679-1690, Vol. 121, No. 3, March, 2007.

[4] A. Dibazar, T.W. Berger, and S. S. Narayanan, “Pathological Voice Assessment”. Proc. of the 28th IEEE EMBS Annual International Conference, NY, USA, Aug., 2006.

[5] K. Umapathy, S. Krishnan, V. Parsa, and D. G. Jamieson, “Discrimination of Pathological Voices Using a Time-Frequency Approach”. IEEE Trans. on Biomedical Engineering, Vol. 52, No. 3, March, 2005.

[6] M. Bahoura and C. Pelletier, “Respiratory Sounds Classification using Analysis and Gaussian Mixture Models”, Proceedings of the 26th Annual Conference of the IEEE EMBS, September, 2004.

[7] J. J. de Oliveira Júnior, M. N. Kapp, C. O. A. Freitas, J. M. Carvalho, R. Sabourin. Handwritten Recognition with Multiple Classifiers for Restricted Lexicon. Proceedings of the 17th Brazilian Symposium on Computer Graphics and Image Processing, v. 1. p. 82-89, 2004..

[8] L. R. Rabiner and R. W Schafer., Digital Processing of Speech Signals. New Jersey: Prentice-Hall, 1978.

[9] Yoh’ichi, “A Weighted Cepstral Distance Measure for Speech Recognition”. IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 35, No. 10, pp.1414-1422, October, 1987.

[10] J. R. Mammone, X., Zhang, and R. P. Ramachandran, “Robust Speaker Recognition - A Feature-Based Approach”, IEEE Signal Processing Magazine, Vol. 13, No. 5, pages 58-71, September, 1996.

[11] S. Furui, “Cepstral Analysis Technique for Automatic Speaker Verification”, IEEE Trans. on ASSP, Vol. 29, No. 2, pp 254-272, April, 1981.

[12] J. M. Fechine, “Reconhecimento Automático de Identidade Vocal Utilizando Modelagem Híbrida: Paramétrica e Estatística”, Doctor’s Thesis, Electrical Engineering, Federal University of Paraíba, Brazil, 2000.

[13] Douglas O’Shaughnessy, Speech Communications: Human and Machine, 2nd Edition, NY, IEEE Press, 2000.

[14] Kay Elemetrics, Kay Elemetrics Corp. Disordered Voice Database, Model 4337, 03 Ed, 1994.

[15] J. Makhoul, S. Roucos, and H. Gish, “Vector Quantization in Speech Coding”, Proceedings of the IEEE, Vol. 73, No. 11, November, pp. 1551-1588, 1985.

[16] Y. Linde, A. Buzo, and R. M. Gray, “An Algorithm for Vector Quantizer Design”, IEEE Transaction on Communications, Vol. COM-28, No. 1, pp 84-95, January, 1980.