**Abstract**

A novel methodology, based on the estimation of nonlinear dynamics features, is presented for automatic detection of pathologies in the phonatory system considering continuous speech recordings (text-dependent). The proposed automatic segmentation and characterization of the voice registers does not require the estimation of the pitch period, therefore it is independent on the gender and the patients intonation. A robust methodology for finding the features that better discriminate between healthy and pathological voices and also for analyzing the affinity among them is also presented. An average success rate of 95.35% in the automatic detection of voice pathologies is achieved considering only six features. The results indicate that nonlinear dynamics is a good alternative for automatic detection of abnormal phonations in continuous speech.

**Methodology**

Database → Segmentation → Characterization → Feat. Sel. → Classification

**Database**

- MEEI (Kay Elemetrics): “The rainbow passage”
- 360 recordings from people with different voice impairments
- 36 recordings from healthy people
- Fs = 85000 kHz, 16 resolution bits

**Characterization**

**Correlation Dimension ($D_r$)**

The Correlation Dimension is a measure of the space dimensionality occupied by the attractor. It is able to quantify the auto-similarity of an embedded signal.

**Largest Lyapunov Exponent ($\lambda_1$)**

It indicates the average divergence rate of the neighbor trajectories in the state space, thus it can give an idea of the complexity of the reconstructed attractor.

**Hurst Exponent ($H$)**

It can describe possible long-term relationships in the time series (voice signal).

**Lempel-Ziv Complexity ($LZ$)**

Classically used to quantify the complexity of a computer algorithm, but here is implemented to quantify the complexity of the voice signal. The number of different bit patterns required to describe the signal is taken as LZ.

Since the mean value ($m$), standard deviation ($std$), skewness ($sk$) and kurtosis ($k$) are taken for each feature, the assigned indexes for each one is as follows:

- $m(D_r, \lambda_1, H, LZ)$: indexes = [1, 2, 3, 4]
- $std(D_r, \lambda_1, H, LZ)$: indexes = [5, 6, 7, 8]
- $sk(D_r, \lambda_1, H, LZ)$: indexes = [9, 10, 11, 12]
- $k(D_r, \lambda_1, H, LZ)$: indexes = [13, 14, 15, 16]

**Feature selection**

The 70% of the database is taken for the feature selection stage. A 10-fold cross validation is applied 10 times over this subset to find the optimal feature set. The procedure is repeated 50 times and the final feature vector is obtained by selecting the most selected features.

**Most relevant features:**

$(sk(H), std(LZ), std(\lambda_1)_r, m(\lambda_1)_r, k(LZ), m(LZ))$

**Classification and validation**

The parameters of the classifiers were tuned considering the same 70% of the database that was used in the feature selection stage, and the remaining 30% is used for validation purposes. The process is repeated 200 times to build the confidence intervals. Note that in these experiments only the most relevant features were considered. Three different classifiers were implemented: Soft Margin – Support Vector Machine (SM-SVM), K-Nearest Neighbors (K-NN) and Neural Nets (NN).

**Results**

The overall accuracy of the system is showed in the following table:

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>SM – SVM</td>
<td>95% ± 3.94%</td>
</tr>
<tr>
<td>K – NN</td>
<td>89.18% ± 6.9%</td>
</tr>
<tr>
<td>NN</td>
<td>87.83% ± 7.74%</td>
</tr>
</tbody>
</table>

**Conclusions**

- It is possible to perform the automatic detection of pathological voices considering nonlinear dynamics features.
- The results presented here are near to the state of the art (96.3%) which considers a set of 36 features (acoustics, Cepstral and noise).
- The presented methodology is robust to voice signals with high level of pathology.
- The most relevant feature is the Skewness of the Hurst Exponent ($H$).

**Acknowledgments**

This work was funded by ARTICA project Nº1115-470-22065, and COLCIENTCIAS through the program “becas de doctorado nacional 2011 – Generación del bicentenario” (Colombia), by CODI at Universidad de Antioquia, project MC11-1-03 and by Spanish government MCINN TEC2009-14123-C04 research project and a research training grant from the AGHSL of the Canary Autonomous Government (Spain).