31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Complexity Analysis of Pathological Voices by means of Hidden Markov Entropy measurements

Julián D. Arias-Londoño, Juan I. Godino-Llorente, Germán Castellanos-Domínguez, Nicolás Sáenz-Lechón, Víctor Osma-Ruiz

***Abstract*—In this work an entropy based nonlinear analysis of pathological voices is presented. The complexity analysis is carried out by means of six different entropies, including three measures derived from the entropy rate of Markov chains. The aim is to characterize the divergence of the trajectories and theirs directions into the state space of Markov Chains. By employing these measures in conjunction with conventional entropy features, it is possible to improve the discrimination capabilities of the nonlinear analysis in the automatic detection of pathological voices.**

I. Introduction

The systems for automatic assessment of voice disorders have received considerable attention in the last years due to its objectivity and non invasive nature. Much of the work done in this area is based on the use of acoustic parameters, noise measurements and cepstral coefficients [1]. However, several researchers have shown that there is a physical phenomenon involved in the voice production process that cannot be characterized by the above measures, termed *Nonlinear Behaviour*. In the context of speech, such a behaviour is due to the following mechanisms: nonlinear pressure-flow relation in the glottis, delayed feedback from a mucosal wave, the nonlinear stress-strain curves of vocal fold tissues, and the nonlinearities associated with vocal fold collision [2]. In order to overcome this problem, some researchers applied tools from nonlinear time series analysis to disordered speech signals to characterize these nonlinear phenomena (see [3] and cites therein). The most common nonlinear analysis from time series is derived from the theory of dynamical systems, and, most of the cases, it is performed by using two statistics: *Largest Lyapunov Exponent (*LLE*)* and *Correlation Dimension (*CD*)*. LLE is a measure that attempts to quantify the sensibility on the initial conditions of the underlying system [4]. CD is a measure designed for quantifying the geometry (self-similarity) in the state space of the underlying system [4]. Different researches have shown that changes in nonlinear dynamic measures may indicate states of pathophysiological dysfunction [3]. In [3] the CD was used to describe the complexity of sustained vowels produced by normal speakers and by patients with vocal polyps. The database contained 79 normal samples and 68 pathological samples of the sustained vowel /a/. The authors concluded that the CD is higher in pathological than in normal speech, and that the nonlinear analysis can be used as a supplementary method in speech signal processing to clinically evaluate and detect laryngeal pathologies. In [5] as a continuation of the above work, the authors used the CD to discriminate between three types of speech signals according to the definition by Titze [2]. In this case, different types of pathologies were included on the database, but speech signals with strong glottal turbulences were excluded. Although LLE and CD have shown certain discrimination capabilities between normal and pathological voices, such nonlinear statistics require the dynamics of speech to be purely deterministic, and this assumption is inadequate, since randomness due to turbulence is an inherent part of speech production [6]. There are also numerical, theoretical and algorithmic problems associated with the calculation of LLE and CD for real speech signals, casting doubts over the reliability of such tools [6].

To overcome this constriction, a set of features based on information theory have been published in the literature. Such measures attempt to quantify the signal complexity as an alternative way for measuring the nonlinear behaviour, without making assumptions about the nature of the signal (deterministic or stochastic). The most common measures used in that context are: *Approximate Entropy* [7], *Sample Entropy* (A_E) [7] and a modification of the A_E called *Gaussian Kernel Approximate Entropy* (GA_E) [8]. This measures proportionate a better parameterization of the nonlinear behaviour, but some of them present a problem with bias [9]. Additionally, from a pattern recognition point of view, they make a non parametric estimate of the probability mass function of the embedding attractor using a Parzen-window method with a Gaussian or Square kernel [10]. They only attempt to quantify the divergence of the trajectories of the attractor but do not take into account the directions of divergence.

In this work, we use a Discrete Hidden Markov Model (DHMM) to estimate a nonparametric density function of the attractor in order to determine the divergence of the trajectories and its directions into the state space in terms of the transitions between regions provided by the DHMM, and then we use an empirical entropy measure in conjunction with above pointed out entropy measures to characterize the complexity of the signals.

J. Arias-Londoño is with the Digital Signal Processing Group, Universidad Nacional de Colombia sede Manizales, Colombia (corresponding author; e-mail: jdariasl@unal.edu.co).

J. Godino-Llorente, N. Sáenz-Lechón, and V. Osma-Ruiz are with the Department of Circuits & Systems Engineering, Universidad Politécnica de Madrid, Spain (e-mail: igodino@ics.upm.es, nicolas.saenz@upm.es, vosma@ics.upm.)

G. Castellanos-Domínguez is with the Department IEEyC, Universidad Nacional de Colombia sede Manizales, Colombia (e-mail: cgcastellanosd@unal.edu.co).

II. METHODS

A. Embedding

The state space reconstruction is based on the *Time-Delay Embedding Theorem* [4], which can be written as follows: Given a dynamic system with a m-dimensional solution space and an evolving solution $\boldsymbol{h}(t)$, let x be some observation $x(\boldsymbol{h}(t))$. Let us also define the lag vector (with dimension m and common time lag τ) as $x(t)\equiv(x_t, x_{t-\tau}, \ldots, x_{t-(m-1)\tau})$. Then, under very general conditions, the space of vectors $\boldsymbol{x}(t)$ generated by the dynamics contains all the information of the space of solution vectors $\boldsymbol{h}(t)$. The mapping between them is smooth and invertible. This property is referred to as diffeomorphism and this kind of mapping is referred to as an embedding. The embedding theorem establishes that, when there is only a single sampled quantity from a dynamical system, it is possible to reconstruct a state space (embedding attractor) that is equivalent to the original (but unknown) state space composed of all the dynamical variables [4]. In this work the embedding dimension m was chosen using the false neighbours method and time–delay τ by using the first minimum of the auto mutual information function [4].

B. Parameterization

The entropy is a measure of the uncertainty of a random variable [11]. Let X be a discrete random variable with alphabet $\mathcal{X}$ and probability mass function $p(x)=\Pr\{X=x\}$, $x\in\mathcal{X}$. The Shannon entropy $H(X)$ is defined by:

$$H(X) = -\sum_{x\in\mathcal{X}} p(x)\log p(x) \tag{1}$$

If instead of one random variable we have a sequence of n random variables (i.e. a stochastic process), the process can be characterized by a joint probability mass function: $\Pr\{X_1=x_1,\ldots,X_n=x_n\}=p(x_1,x_2,\ldots,x_n)$. Under the assumption of existence of the limit, the rate at which the joint entropy grows with n is defined by [11]:

$$H(X) = \lim_{n\to\infty}\frac{1}{n}H(X_1,X_2,\ldots,X_n) = \lim_{n\to\infty}\frac{1}{n}H_n \tag{2}$$

If the set of random variables are independent but not identically distributed, the entropy rate is given by:

$$H(X) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}H(X_i) \tag{3}$$

On the other hand, let the state space be partitioned into hypercubes of content ε^d, and the state of the system measured at intervals of time δ. Moreover, let $p(k_1, \ldots, k_n)$ denote the joint probability that the state of the system is in the hypercube k_1 at $t=\delta$, k_2 at $t=2\delta$. The Kolmogorov-Sinai (KS) entropy is defined as [9]:

$$H_{KS} = -\lim_{\substack{\delta\to 0\\ \varepsilon\to 0\\ n\to\infty}}\frac{1}{n\delta}\sum_{k_1,\ldots,k_n} p(k_1,\ldots,k_n)\log p(k_1,\ldots,k_n) \tag{4}$$

The KS entropy measures the mean rate of creation of information [9]. For stationary processes it can be shown that [9]:

$$H_{KS} = \lim_{\delta\to 0}\lim_{\varepsilon\to 0}\lim_{n\to 0}\left(H_{n+1}-H_n\right) \tag{5}$$

Numerically, only entropies of finite order n can be computed. However, some methods have been proposed in an attempt to estimate the KS entropy. One of them is the *Approximate Entropy* (A_E). A_E is a measure of the average conditional information generated by diverging points on the trajectory [7;9]. A_E is defined as a function of the *Correlation Sum* (CS) given by:

$$C_i^m(r) = \frac{2}{N(N-1)}\sum_{j=i+1}^{N}\Theta\left(r-\mathrm{norm}\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right)\right) \tag{6}$$

where Θ is the Heaviside function, N is the number of points in the state space and the norm can be defined in any consistent metric space. The CS is the fraction of all possible pairs of points in the state space which are closer than a given distance r in a particular norm. For a fixed m and r, A_E is given by:

$$A_E(m,r) = \lim_{N\to\infty}\left[\Phi^{m+1}(r)-\Phi^m(r)\right] \tag{7}$$

where

$$\Phi^m(r) = \frac{1}{N-m+1}\sum_{i=1}^{N-m+1}\ln C_i^m(r) \tag{8}$$

A first modification of A_E presented in [7] called *Sample Entropy* (S_E), was developed with the aim of obtaining a more independent measure of the signal length than A_E. S_E is given by:

$$S_E(m,r) = \lim_{N\to\infty} -\ln\frac{\Gamma^{m+1}(r)}{\Gamma^m(r)} \tag{9}$$

The difference between Γ and Φ is that the first one does not compare the embedding vectors with themselves (excludes self-matches). The advantage is that the estimator is unbiased [9].

Another modification of A_E presented in [8], called *Gaussian Kernel Approximate Entropy* (GA_E), changes the Heaviside function by a Gaussian Kernel based function with the aim of suppressing the discontinuity of the auxiliary function over the correlation sum (square kernel). In this case, the Heaviside function is replaced by:

$$d_G\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right) = \exp\left(-\left(\left\|\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right)\right\|_1\right)^2/10r^2\right) \tag{10}$$

By using eq. (10), the estimation of GA_E is done in the same way than for A_E (see eqs. (7) and (8)).

C. Hidden Markov Entropy

A Markov chain is a random process $\{X(t)\}$ which can take a finite number of k values at certain moments of time ($t_0 < t_1 < t_2 < \cdots$). The values of the stochastic process change with known probabilities called *transition probabilities*. The particularity of this stochastic process is that the probability of change to other state depends only on the current state of the process; this is known as the *Markov condition*. When such probabilities do not change with time and the initial probability of each state is also constant, the Markov chain is stationary. Let $\{X(t)\}$ be a stationary Markov chain with initial distribution $\boldsymbol{\pi}$ and transition matrix $\mathbf{A}$. Then the entropy rate is given by [11]:

$$H(\mathcal{X}) = -\sum_{ij} \pi_i A_{ij} \log A_{ij} \quad (11)$$

In view of eq. (11), it is possible to observe that the entropy measure is a sum of the individual Shannon entropy measures for the transition probability distribution of each state, weighted with respect to the initial probability of its corresponding state. There exist some processes that can be seen like a Markov chain whose outputs are random variables generated from probability functions associated to each state. Such processes are called Hidden Markov Processes (HMP), because the states of the Markov process cannot be identified from its output (the states are "hidden"). In this case, it is not possible to obtain a close form for the entropy rate [11;12]. A HMP can also be understood as a Markov process with noisy observations [12]. Therefore, in the same way as in eq. (11), it is possible to establish an entropy measure of the HMP as the entropy of the Markov process plus the entropy generated by the noise in each state of the process. We called to this measure *Empirical Entropy*. If we use a DHMM for modeling the stochastic process, the noise is modeled by means of discrete distributions and finally it is possible to obtain a probability mass function for the noise in each state.

Denoting the actual state of the process in time t as S_t, a DHMM can be characterized by the following parameters [13]:

- $\boldsymbol{\pi}=\{\pi_i\}$, i=1,2,…,k: the initial state distribution, where $\pi_i = p(S_0=i)$ is the probability of starting at the *i-th* state.
- $\mathbf{A}=\{A_{ij}\}$, $1\leq i,j\leq k$: the set of transition probabilities among states, where $A_{ij}=p(S_{t+1}=j|S_t=i)$ is the probability of reaching the *jth* state at time t+1, coming from the *ith* sate at time *t.*
- $\mathbf{B}=\{B_{ij}\}$, i=1,2,…,k, j=1,2,…,φ: the probability distribution of the observation symbol, being $B_{ij}=p(o_t=\upsilon_j|S_t=i)$, where o_t is the output at time t, υ_j are the different symbols that can be associated to the output, and φ is the total number of symbols. All parameters are subject to standard stochastic constrains [13].

From this definition, the empirical entropy can be defined as:

$$H_{ES} = H_{MC} + H_g \quad (12)$$

where H_{MC} is the entropy due to the Markov process, as it was defined in eq. (11), and H_g is the entropy due to the noise. By replacing both entropies, H_{ES} can be written as:

$$H_{ES} = -\left(\sum_{ij}^{k} \pi_i A_{ij} \log A_{ij} + \sum_{i=1}^{k} \sum_{j=1}^{\varphi} B_{ij} \log B_{ij} \right) \quad (13)$$

If instead of using Shannon Entropy in the definition of the empiric entropy we use the *Renyi Entropy* [11], the empiric entropy becomes:

$$H_{ER} = \sum_{i=1}^{k} \frac{\pi_i}{1-\alpha} \log \sum_{j=1}^{k} A_{ij}^{\alpha} + \sum_{i=1}^{k} \frac{1}{1-\alpha} \log \sum_{j=1}^{\varphi} B_{ij}^{\alpha} \quad (14)$$

where $\alpha>0$; $\alpha\neq 1$ is the entropy order. In this work we are using α=2, since it is the most common Renyi Entropy [11].

III. Experimental Setup

A. Database

The database used was developed by *The Massachusetts Eye and Ear Infirmary Voice Laboratory* (MEEIVL) [14]. The recordings were resampled to 25 *k*Hz and 16 *bits* of resolution. The registers contain the sustained phonation of the /*ah*/ vowel from patients with a variety of voice pathologies: organic, neurological, and traumatic disorders. A subset of 173 registers of pathological and 53 normal speakers has been taken according to those enumerated in [15]. All the files were chosen to have a diagnosis and similar age distributions between both groups.

B. Results

In this work the speech signal is divided into frames in order to be parameterized by means of short-time analysis with the aim of taking into account as much dynamic information as possible. For each frame, the entropy based features described in sections II.B and II.C are estimated. The size of the each window has been chosen as 55 ms and the parameter r=0.35 according to the results presented in [16]. The classification is based on a DHMM with k=3 states and φ=128 symbols. These values were determined as the best after a set of experiments using different values of k and φ. The validation is performed by using a cross-validation method with 11 folds, using 70% of the speech recordings in the training step and 30% in the validation step. Table I shows the discrimination capacities of each feature, it is possible to observe that the empirical entropies (H_{ES}, H_{ER}) reach a higher accuracy rate than the rest. This fact shows that the estimation of the probability density function made by the DHMM allows not only characterize the state space but the transitions between different regions one from the trajectories of the attractor.

TABLE I.
DISCRIMINATION CAPABILITIES OF THE COMPLEXITY MEASURES IN THE DETECTION OF PATHOLOGICAL VOICES.

Feature	*Sensitivity*	*Specificity*	*Accuracy*
A_E	85.74%	56.97%	79.20%
S_E	93.4%	63.03%	86.50%
GA_E	88.95%	52.12%	80.58%
H_{MC}	89.30%	76.36%	86.36%
H_{ES}	91.09%	75.76%	87.60%
H_{ER}	92.87%	75.15%	88.84%

By using the whole set of features it is possible to reach a classification accuracy of 92.70%. Without the empirical entropies the accuracy reached 87.39%.

This suggest that the empirical entropies contain additional and complementary information related to the system complexity. Thus, an improvement can be obtained using this feature in conjunction with other complexity measures. Figure 1 shows the ROC curves for each of the features in a comparative way. It is possible to observe that H_{ER} covers a larger area in the graph.

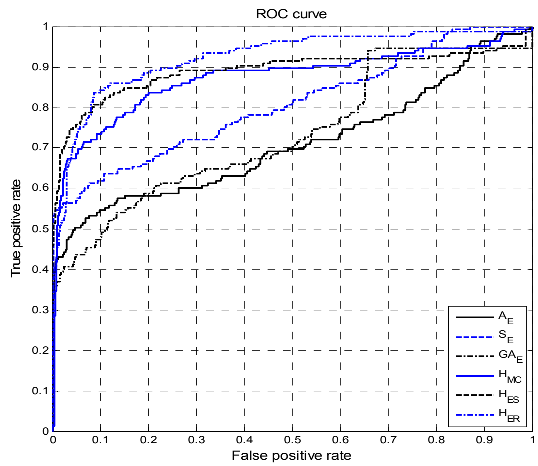

Fig 1. ROC curves for the system using the features in the Table I.

Figure 2 shows the DET plot for the best three features in Table I, which correspond to Markov chains based entropies. In this case, for H_{ER} is possible to define an operating point closer to lower left corner than the rest.

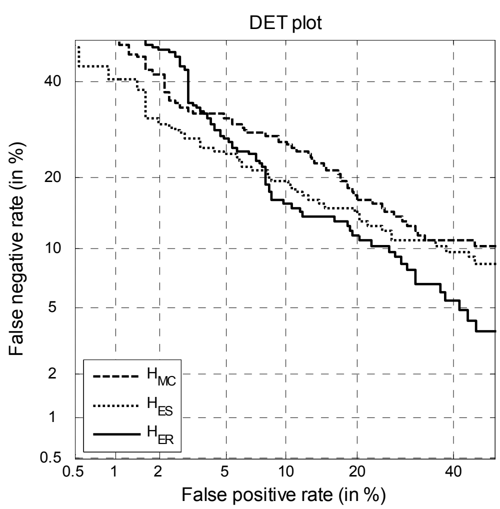

Fig 2. DET plot for the system using Markov chains based entropies

IV. Conclusions

The characterization of the embedding space carried out by DHMM takes into account additional information related to transitions between different regions of the state space from the trajectories of the attractor, which is reflected in the discrimination capabilities of the empirical entropy. The use of the empirical entropy in conjunction with the approximate entropy, sample entropy, and Gaussian kernel approximate entropy, allows obtaining accuracies above 92%, only analyzing the nonlinear behavior of the speech signals.

This methodology does not attempt to replace the analysis based on classical acoustic parameters, but to show a different alternative for the nonlinear analysis of voice signals, complementing the traditional multidimensional studies that use to be carried out for the objective characterization of voice pathologies.

Acknowledgments

This work was supported by: "Convocatoria de apoyo a doctorados nacionales del Instituto Colombiano para el Desarrollo de la Ciencia y la Tecnología-Colciencias, 2007"; and TEC2006-12887-C02, from the Ministry of Education of Spain.

References

[1] N. Sáenz-Lechón, J.I. Godino-Llorente, V. Osma-Ruiz, and P. Gómez-Vilda, "Methodological issues in the development of automatic systems for voice pathology detection," *Biomedical Signal Processing and Control*, vol. 1, no. 2, pp. 120-128, 2006.

[2] I. R. Titze, *The Myoelastic Aerodynamic Theory of Phonation*, Iowa, IA: National Center for Voice and Speech, 2006.

[3] J. J. Jiang, Y, Zhang, and C. McGilligan, "Chaos in voice, from modeling to measurement," *Journal of Voice*, vol. 20, no. 1, pp. 2-17, 2006.

[4] H. Kantz, and T. Schreiber, *Nonlinear time series analysis*, Second ed., CAMBRIDGE University Press, 2004.

[5] Y. Zhang, and J.J. Jiang, "Nonlinear dynamic analysis in signals typing of pathological human voices," *Electronics Letters*, vol. 39, no. 13, pp. 1021-1023, 2003.

[6] M.A. Little, P.E. McSharry, S.J. Roberts, D.A. Costello, and I.M. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *Biomedical Engineering Online*, vol. 6, no. 23, 2007.

[7] J.S. Richman, and J.R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Am J Physiol HeartCirc Physiol*, vol. 278, pp. H2039-H2049, 2000.

[8] L.-S. Xu, K.-Q. Wang, and L. Wang, "Gaussian kernel approximate entropy algorithm for analyzing irregularity of time series," in *Proceedings of the Fourth International Conference on Machine Learning and Cybernetics*, pp. 5605-5608, 2005.

[9] M. Costa, A. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of biological signals," *Physical Review E*, vol. 71, pp. 021906-1-021906-18, 2005.

[10] D. Woodcock, and I.T Nabney, A new measure based on the Renyi entropy rate using Gaussian kernels. 2006. Aston University, UK.

[11] T. M. Cover, and J.A. Thomas, *Elements of information theory*, Second ed., Wiley Interscience, 2006.

[12] M. Rezaeian, "Hidden Markov Process: A New Representation, Entropy Rate and Estimation Entropy," preprint, 2006.

[13] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications on Speech Recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257-286, 1989.

[14] Massachusetts Eye and Ear Infirmary. Voice disorders database. version 1.03. [CD-ROM]. 1994. Lincoln Park, NJ: Kay Elemetrics Corp.

[15] V. Parsa, and D. Jamieson, "Identification of pathological voices using glottal noise measures," *Journal of Speech, Language and Hearing Research*, vol. 43, no. 2, pp. 469-485, 2000.

[16] J.D. Arias-Londoño, J.I. Godino-Llorente, and G. Castellanos-Domínguez, "Short time analysis of pathological voices using complexity measures," in *Proceedings of 3rd Advanced Voice Function Assessment International Workshop*, 2009.