Influence of Delay Time on Regularity Estimation for Voice Pathology Detection

J. A. Gómez García , J. I. Godino Llorente
G. Castellanos Dominguez

Abstract— **The employment of nonlinear analysis techniques for automatic voice pathology detection systems has gained popularity due to the ability of such techniques for dealing with the underlying nonlinear phenomena. On this respect, characterization using nonlinear analysis typically employs the classical Correlation Dimension and the largest Lyapunov Exponent, as well as some regularity quantifiers computing the system predictability. Mostly, regularity features highly depend on a correct choosing of some parameters. One of those, the delay time τ, is usually fixed to be 1. Nonetheless, it has been stated that a unity τ can not avoid linear correlation of the time series and hence, may not correctly capture system nonlinearities. Therefore, present work studies the influence of the τ parameter on the estimation of regularity features. Three τ estimations are considered: the baseline value 1; a τ based on the Average Automutual Information criterion; and τ chosen from the embedding window. Testing results obtained for pathological voice suggest that an improved accuracy might be obtained by using a τ value different from 1, as it accounts for the underlying nonlinearities of the voice signal.**

I. Introduction

The automatic detection of voice pathologies is an increasingly important issue, whose main aim is to develop computer-aided diagnostic systems, enabling an objective assessment on the presence of pathologies, reducing the evaluation time and subsequently improving the diagnosis and clinical treatment given to each patient [1]. On this regard, the nonlinear behaviour involved in the voice production process should be taken in account, since it is product of multiple physical phenomena, such as the nonlinear pressure-flow relation in the glottis, the delayed feedback from mucosal wave, the nonlinear stress-strain curves of vocal fold tissues, nonlinearities associated with vocal fold collision [2], or asymmetries between the right and left vocal folds [3].

However, the nonlinear analysis of time series, requires the reconstruction of the underlying dynamical behaviour of the system, so that its states and its evolution, are represented on a m-dimensional space. The most common technique for this purpose, is based on the *Time-Delay Embedding Theorem* [4], which depends on the computation of two parameters: the embedding dimension m and the time delay (or time lag) τ. Then, by having this reconstruction, a feature extraction process is possible, on which the Correlation dimension (d_2) and the Largest Lyapunov Exponent (Λ) emerge as the most typically used characteristics for this purposes. In addition, a set of features called *regularity* features, are employed to quantify the regularity or predictability of the system. Nonetheless, this features depend on the tuning of some parameters, such as the pattern length m, time delay τ, and tolerance r, and which might heavily affect the estimation of regularity. On this regard, r is tuned as a constant α (which varies between 0 and 1) multiplied by the standard deviation of the time series, thus allowing consistency on comparisons between samples having different amplitude [5]. Moreover, and despite m and τ are essentially the same parameters as in the reconstruction process, the embedding is not a part of the regularity computation and, except for τ, the parameter m has a different interpretation on how it should be chosen [6]. With that in mind, it is recommended to fix $m = 2$, as its aim is not to produce a good reconstruction of the time series, but rather to provide an improved estimation of predictability [6]. Finally, the parameter τ is typically fixed to 1 with no apparent reason but simplicity. However, it has been pointed out that using a unity time delay might mask the underlying nonlinearity of the time series which is mainly obscured by the linear autocorrelation of the signal [6]. For dealing with such drawback, it is proposed in [6], to use the same time delay τ of the reconstruction process, computed using the Average Mutual Information (AMI) criterion [4]. This might minimize irrelevance by avoiding temporal correlation, and therefore might improve the estimation of regularity. Nonetheless, the AMI criterion suffers from some drawbacks which might affect the quality of the regularity estimator. Among other, a probability computation is necessary. As it employs histograms, the criterion is dependant on the number of bins chosen to build the histogram. Moreover, AMI uses the first zero of a mutual information function, having no obvious reason to choose this over other minima on the function [7]. Above shortcomings suggest that neither the unity delay nor the value obtained with the AMI criterion, might be optimal for obtaining the time delay τ for the regularity estimation.

With that in mind, this work proposes the employment of a τ obtained from the embedding window, and which might provide a more robust approach that the AMI criterion. The aim is to explore if an improved performance on automatic pathological voice detection labours, is obtained by using such approach. Experiments include testing on a voice disorder database, using three τ estimations: τ using the baseline 1, τ computed using the AMI criterion, and τ computed from the embedding window.

II. Theoretical Background

A. Embedding and embedding parameters

The nonlinear analysis of time series mostly employs a procedure called *embedding*, to represent the dynamical evolution of the system on a m-dimensional space, called *phase* or *state space*, and where all states of the system and its evolution are described.

Let $\vec{\boldsymbol{s}} = \{x[1], x[2], \cdots, x[N]\}$ be a time series of length N, such that the reconstructed states are of the form:

$$\vec{\boldsymbol{x}}[t] = \{x[t], x[t + (m-1)\tau], \cdots, x[t - d_w]\} \tag{1}$$

where m is the embedding dimension and τ is the time lag, and where both parameters are chosen no to confuse dynamics in phase space.

Despite the usual embedding procedure usually focuses on choosing τ and m separately, some authors have stated on the importance of considering directly a quantity termed embedding window $d_w = (m-1)\tau$, as which relates them both [8]. Following that line of thought, a procedure for estimating the embedding window is presented in [8], such that it provides the optimal reconstruction of the underlying dynamics for an observed time series by combining modelling and embedding into a single procedure. To achieve this, they assume that the optimal model that describes the data, is the one which minimises an information criterion called *Description Length* (DL). By using local constant models, the description length of the data is then computed for increasing values of d_w. The value which presents the *Minimum Description Length* (MDL) is then chosen as the optimal.

The procedure is as follows: Let $x[t]$ be a point in the time series, whose reconstructed state vector is $\vec{\boldsymbol{x}}[t]$. Its successor would be $x[s+1]$, where s is chosen for being the nearest neighbour, $\vec{\boldsymbol{x}}[s]$ of $\vec{\boldsymbol{x}}[t]$. Then $x[t+1] = x[s+1]$, and therefore, the prediction error is then calculated as the difference between the successor to that point and the successor to its nearest neighbour:

$$e[t+1] = x[t+1] - x[s+1] \tag{2}$$

That modelling scheme is termed *local constant modelling* and provides a estimation of the DL of the time series as [8]:

$$DL(\vec{\boldsymbol{s}}) \approx \frac{N - d_w}{2} \ln\left[\frac{1}{N - d_w}\sum_{i=d_w+1}^{m} e_i^2\right] + \frac{d_w}{2}\left[\frac{1}{d_w}\sum_{t=1}^{d_w}(\vec{\boldsymbol{s}} - \overline{\boldsymbol{s}})^2\right] + d_w + DL(d_w) \tag{3}$$

where $\overline{\boldsymbol{s}}$ is the mean of the time series, and $d_w^{max} = (m-1)\tau$ is an upper limit on the minimisation procedure.

The procedure could be summarized as follows: Minimize equation (3), by estimating the model prediction error of equation (2), for increasing values of d_w. The minimum for certain d_w will be the optimal embedding window [8].

B. Characterization

Having the time series represented on phase space, it is then possible to extract features. Two types of characteristics are considered in this work, the classical nonlinear dynamics features, and a set of regularity features.

1) Nonlinear dynamic features: The two most classical features on this context are the Correlation Dimension and the Largest Lyapunov Exponent.

The *Correlation dimension* (d_2) quantifies with a dimension the autosimilarity of an embedded time series [4]. For a time series of length N, a quantity, termed correlation sum, is firstly defined as:

$$C(r) = \lim_{n\to\infty}\frac{1}{N^2}\sum_{i,j=1}^{n}\Theta\left(r - \|\vec{\boldsymbol{x}}[i] - \vec{\boldsymbol{x}}[j]\|\right) \tag{4}$$

where Θ is the Heaviside function, r is a tolerance measure, and $\vec{\boldsymbol{x}}[\cdot]$ are reconstructed state vectors as in (1).

It is expected that as $r \to 0$, then $C(r) \to \kappa^{d_2}$, where κ is constant, and d_2 is thus the estimation of the correlation dimension.

On the other hand, the *Largest Lyapunov Exponent* (Λ), is a measure of the divergence of nearby orbits in phase space, thus representing one of the basic attributes of the nonlinear dynamic systems: sensitivity to initial conditions.

Let $\vec{\boldsymbol{x}}[i]$ and $\vec{\boldsymbol{x}}[j]$ be two states in the phase space, with distance defined as $\delta_0 = \|\vec{\boldsymbol{x}}[i] - \vec{\boldsymbol{x}}[j]\| \ll 1$; and let $\delta_{\triangle n} = \|\vec{\boldsymbol{x}}[i + \triangle n] - \vec{\boldsymbol{x}}[j + \triangle n]\|$ be the distance some time later $\triangle n$. Then, Λ will be determined by:

$$\Lambda(\delta_0) = \lim_{\Delta n\to\infty}\lim_{\|\delta_0\to 0\|}\frac{1}{\Delta n}\log\frac{\|\delta_{\Delta n}\|}{\|\delta_0\|} \tag{5}$$

2) Entropy-based quantifiers: Several complexity measures have been developed to measure system regularity. One of the most popular, termed *Approximate Entropy* (ApEn), is proposed in [9].

ApEn is intended to quantify the reproducibility of temporal patterns in a time series through calculation of the "logarithmic likelihood" that in a data set of length N, patterns of length $m+1$ are within tolerance r of each other, given that patterns of length m are within tolerance r of each other [6]. ApEn is defined as follows:

$$\text{ApEn} = \phi^m(r) - \phi^{m+1}(r) \tag{6a}$$

$$\phi^m(r) = \frac{1}{N - (m-1)\tau}\sum_{i=1}^{N-(m-1)\tau}\log C_i^m(r), \tag{6b}$$

$$C_i^m(r) = \frac{B_i^m(r)}{N - (m-1)\tau} \tag{6c}$$

where r is a tolerance measure, and $B_i^m(r)$ is the number of j such that $|\vec{\boldsymbol{x}}[i] - \vec{\boldsymbol{x}}[j]| < r$.

Since ApEn is biased due to a phenomena called self-matching, the *Sample Entropy* (SampEn) is proposed in [5].

SampEn is defined as follows:

$$\text{SampEn} = -\log\left(\frac{A^m(r)}{A^{m+1}(r)}\right) \tag{7a}$$

$$\phi^m(r) = \frac{1}{N-1-(m-1)\tau}\sum_{i=1,i\neq j}^{N-(m-1)\tau} C_i^m(r) \tag{7b}$$

III. Experimental Setup

A. Database

Testing has been carried out with the Massachusetts Eye and Ear Infirmary [10]. Voice disorders database. The registers contain the sustained phonation of the /*ah*/ vowel from patients with a variety of voice pathologies disorders, of organic, neurological and traumatic nature. The registers were previously edited to remove the beginning and ending of each utterance, removing the onset and offset effects in these parts of each utterance. Database is composed by 173 registers of pathological speakers and 53 of normal speakers as those selected by [11].

B. Experiments

Figure 1 presents an outline of the employed experimentation procedure for experimentation, whilst their stages are explained next.

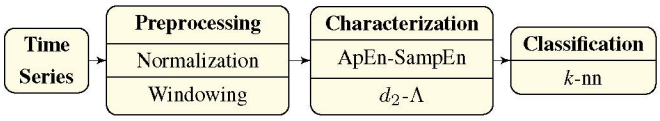

Fig. 1. Outline of the automatic voice pathological system, based on regularity and nonlinear dynamics features, presented on this work

On the *Preprocessing* stage all voice recordings are *z-score* normalized, thus the mean of the time series becomes one and the standard deviation becomes zero. The z-score is as follows:

$$z-score : \frac{(\vec{s}-\bar{s})}{std(\vec{s})} \tag{8}$$

In order to employ a short time analysis, 50% overlapped square windows of 55 ms of duration are used as suggested in [12], therefore, splitting each single recording into frames.

On the *Characterization* stage two experiments are to be considered:

1) Each voice frame is characterized by means of ApEn and SampEn. The parameter m is fixed to 2, whilst to test out the validity of the methodology at different tolerance levels, the α parameter is varied from 0.1 to 0.35, where α is such that $r = \alpha\,\text{std}(\cdot)$. Furthermore, three τ values are considered: First, with τ equal to 1, as its typical on nonlinear characterization using regularity features. Second, with τ tuned using the AMI criterion. Third, with τ tuned as the value given by $\tau = (d_w)/(m_e - 1)$, where d_w is chosen with the criterion suggested in last section, and m_e is chosen using the False Nearest neighbour criterion (FNN) [4].
2) Each voice frame is characterized by means of d_2 and Λ, by using the m_e parameter of the FNN criterion and τ chosen with the AMI and MDL criterion.

Finally, on the *Classification* stage, a *k*-nn classifier is used whose neighbours are varied from 3 to 11. For validation of results a leave-one out procedure is employed, such that a single recording is chosen to validate results, while the remaining are utilized for training the classifier. That same procedure is repeated until having used each single recording for validation. Also, the decision on the class membership of a given test signal, is taken based on majority voting of the class membership of each one of the frames which compose the signal.

C. Results

Figure 2 presents the accuracy for the both ApEn and SampEn, by using a *k*-nn classifier and varying the number of neighbours from 3 to 11. Each graphic depicts the accuracy for a particular α value starting from 0.1, shown on the upper left corner, and finishing on 0.35 on the bottom left corner. The three considered τ parameters are also depicted in each plot.

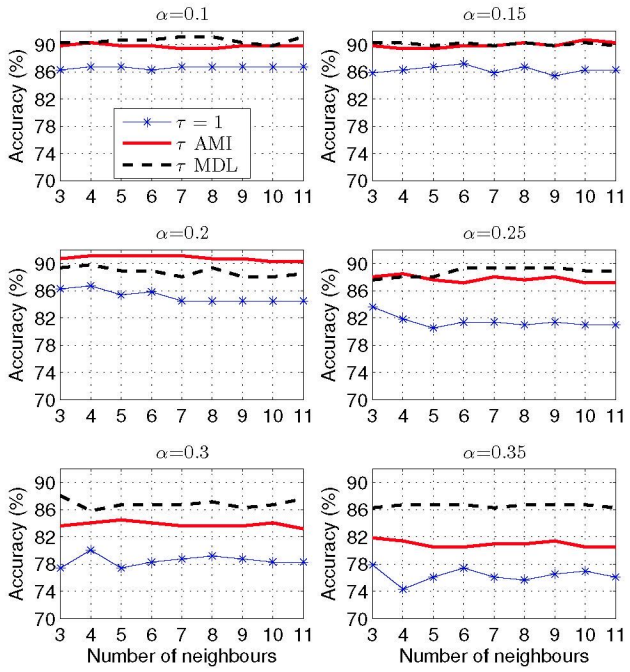

Fig. 2. Accuracy to variations on the number of neighbours, on a *k*-nn classifier, using both ApEn and SampEn as features. Each figure indicates the accuracy obtained using a different α parameter, varied from 0.1 to 0.35 on 0.05 steps.

Figure 3 presents the classification accuracy, for the *k*-nn classifier as explained before, and by using d_2 and Λ features in conjunction. Only two τ parameters are considered: using the AMI criterion and the MDL criterion.

IV. Discussions and Conclusion

As presented in Figure 2, the classification accuracy of ApEn and SampEn is higher using the τ chosen with the

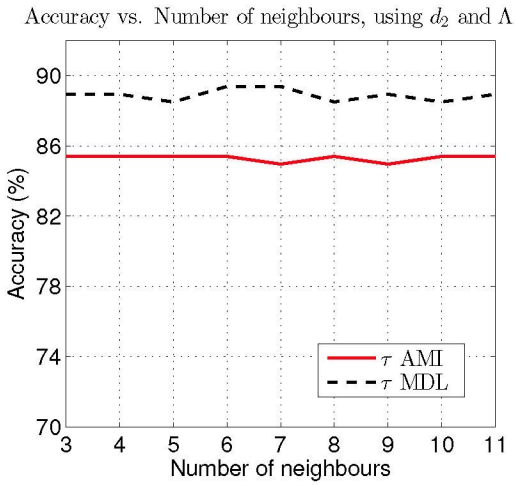

Fig. 3. Accuracy to variation on the number of neighbours, on a k-nn classifier. Both d_2 and Λ were used as features

AMI and the MDL criterion compared to the unity τ of the baseline. This result might indicate that the employment of a delay time, chosen with some criterion which minimizes the influence of linear autocorrelations, might improve the performance on pattern recognition labours. As indicated in [6], when the autocorrelation of the signal decays rapidly, a unity delay time will be sufficient to provide an accurate measure of signal regularity resulting from the nonlinearities of the signal. However, as the results suggest, this is not the case for the voice disorder database.

Results also suggest that the τ chosen with the MDL criterion might provide an improved performance on pathological voice detection labours, as it is evidenced by the higher classification accuracy using all α values but $\alpha = 0.2$. Moreover, and despite in most of the cases a α value between 0.1 and 0.2 might be sufficient for tuning the regularity parameters, the wider range we present serves to provide insight in the higher robustness at different tolerance levels, by using the MDL criterion rather than the AMI criterion.

Figure 3 present the classification accuracy for d_2 and Λ. As in the previous case, the results are better using the MDL criterion, obtaining results up to 4% classification points compared to the delay time chosen with AMI. That might be explained because when using a τ chosen with the MDL, a better reconstruction on phase space might be produced, and hence a higher classification accuracy.

Nonetheless, and despite the increased performance obtained using AMI or MDL, compared to the unity delay time, both criteria provide information about how to minimize the irrelevance of the signal produced by the linear temporal correlation. However, they lack to provide information about redundancy of the parameter. Having that in mind, it reasonable to conclude that both criteria might be suboptimal for choosing the τ parameter for regularity estimation.

Finally, and based on the results presented in this work we can conclude:

- The τ parameter is a critical value which should be tuned to improve classification accuracy of the regularity estimators, since the unity τ can not capture the underlying nonlinearities of time series.
- The MDL criterion for choosing τ might produce an improved classification performance on both classical and regularity features, compared to the unity τ and the one chosen with the AMI criterion
- MDL or AMI criterion might be suboptimal since the found τ only minimizes the irrelevance of signal, having no information about the redundancy of the choosing.

As future work we will study different criteria for choosing the τ parameter, including one that accounts for minimizing irrelevance and redundancy. Moreover, additional testing will be performed on other databases in order to validate the results found in this work.

ACKNOWLEDGEMENT

This research was carried out under grants: TEC2009-14123-C04 from the Spanish Ministry of Education; AL11-P(I+D)-022, and *Ayudas para la realizacion del doctorado en las Escuelas, Facultades, Centros de I+D e Institutos Universitarios*, from Universidad Politécnica de Madrid.

REFERENCES

[1] J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz, S. Aguilera-Navarro, and P. Gómez-Vilda, "An integrated tool for the diagnosis of voice disorders." *Medical engineering & physics*, vol. 28, no. 3, pp. 276–89, May 2006.

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

[3] I. Steinecke and H. Herzel, "Bifurcations in an asymmetric vocal-fold model," *The Journal of the Acoustical Society of America*, vol. 97, p. 1874, 1995.

[4] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, 2nd ed. Cambridge University Press, 1 2004.

[5] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy." *American journal of physiology. Heart and circulatory physiology*, vol. 278, no. 6, pp. H2039–49, Jun. 2000.

[6] F. Kaffashi, R. Foglyano, C. Wilson, and K. Loparo, "The effect of time delay on Approximate & Sample Entropy calculations," *Physica D: Nonlinear Phenomena*, vol. 237, no. 23, pp. 3069–3074, Dec. 2008.

[7] S. Garcia and J. Almeida, "Multivariate phase space reconstruction by nearest neighbor embedding with different time delays," *Physical Review E*, vol. 72, no. 2, pp. 1–4, Aug. 2005.

[8] M. Small and C. Tse, "Optimal embedding parameters: a modelling paradigm," *Physica D: Nonlinear Phenomena*, vol. 194, no. 3-4, pp. 283–296, 2004.

[9] S. M. Pincus, "Approximate entropy as a measure of system complexity." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88, no. 6, pp. 2297–301, Mar. 1991.

[10] M. Eye and E. Infirmary, "Voice disorders database, version 1.03 [cd-rom]," *Lincoln Park, NJ: Kay Elemetrics Corporation*, 1994.

[11] V. Parsa and D. Jamieson, "Identification of pathological voices using glottal noise measures," *Journal of speech, language, and hearing research*, vol. 43, no. 2, p. 469, 2000.

[12] J. D. Arias-Londoño, J. I. Godino-Llorente, and C. G. Castellanos-Domínguez, "Short time analysis of pathological voices using complexity measures," in *3rd Advanced Voice Function Assessment International Workshop, AVFA2009*, 2009.