Available online at www.sciencedirect.com
SCIENCE DIRECT®

ELSEVIER

Physica A 361 (2006) 337–354

PHYSICA A

www.elsevier.com/locate/physa

sinc(*i*) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
H. M. Torres, J. Gurlekian, H. L. Rufiner & M. E. Torres; "Self-organizing map clustering based on continuous multiresolution entropy"
Physica A: Statistical Mechanics and its Applications. Vol. 361, No. 1, pp. 337--354, 2006.

Self-organizing map clustering based on continuous multiresolution entropy

H.M. Torres[a,1], J.A. Gurlekian[a,1], H.L. Rufiner[b,2], M.E. Torres[c,2,*]

[a] *Laboratorio de Investigaciones Sensoriales, Consejo Nacional de Investigaciones Científicas y Técnicas, Instituto de Neurociencias Aplicadas, Hospital de Clínicas, Buenos Aires, Argentina*
[b] *Laboratorio de Cibernética, Fac. Ingeniería, Universidad Nacional de Entre Ríos, Oro Verde, Entre Ríos, Argentina*
[c] *Laboratorio de Señales y Dinámicas no Lineales, Fac. Ingeniería, Universidad Nacional de Entre Ríos, Oro Verde, Entre Ríos, Argentina*

Received 1 April 2005; received in revised form 4 May 2005
Available online 22 June 2005

Abstract

The detection of changes in the parameter values of a nonlinear dynamic system is a branch of study with multiple applications. In this paper, we explore a variant of an automatic detector and clustering of slight parameter variations in nonlinear dynamic systems proposed by Torres et al. [Automatic detection of slight changes in nonlinear dynamical systems using multiresolution entropy tools, Int. J. Bifurc. Chaos 11(4) (2001) 967–981]. The new method takes the advantages of the continuous multiresolution entropy to localize slight changes in the parameters, and uses self-organizing maps to quantify and cluster these changes. We discuss the performance of this method while applied to automatic segmentation of natural

*Corresponding author. Tel.: +54 343 4975078/77x122; fax: +54 343 4975100/01x105.
E-mail address: metorres@ceride.gov.ar (M.E. Torres).

[1] This research has been carried out with the support of the Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina.

[2] Supported by Universidad Nacional de Entre Ríos (UNER) and Agencia Nacional de Promoción Científica y Tecnológica, Argentina, under Project PICT-2002 11-12700, and PID-UNER 07F/6060.

0378-4371/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2005.05.073

sinc(*i*) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
H. M. Torres, J. Gurlekian, H. L. Rufiner & M. E. Torres; "Self-organizing map clustering based on continuous multiresolution entropy"
Physica A: Statistical Mechanics and its Applications. Vol. 361, No. 1, pp. 337--354, 2006.

and synthetic diphthongs in the presence of additive noise. Our results show the potentiality of the proposed method.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Wavelet transform; Continuous multiresolution entropy; Self-organizing maps; Nonlinear systems; Speech segmentation

1. Introduction

It is known that slight parameter variations in nonlinear dynamic systems can produce important changes in the system's behavior. If the system dynamics can be mathematically described, there is a wide range of tools that can be used to characterize its performance [1,2]. When the model equations are not known, and only temporal data sequences acquired by the system under study are available, techniques such as Lyapunov exponents [3,4], phase space reconstruction [4] and correlation dimension [5] can be used. Unfortunately, they require a great amount of data for their estimation, and signal stationarity is commonly assumed, which is not always true in real applications [6,7].

A technique that allows the temporal localization of slight parameter changes in the law governing the subjacent dynamic of a signal, coming from a nonlinear dynamic system, has been proposed in Ref. [8] and is known as the continuous multiresolution entropy (CME). It computes the entropy evolution by means of sliding windows at each scale of the continuous wavelet transform (CWT) of the given signal. The CME has been applied to non-stationary time series, with the advantage that it does not require a large amount of data and it has low computational cost.

CME is sensitive to changes in the dynamic complexity, displaying statistical variations in the multiresolution entropy. In Ref. [8], an automatic detection algorithm for the temporal localization of slight parameter changes at the governing complex system was presented. This algorithm was based on the time series CME analysis.

In the present work we propose a new technique that combines the CME with Kohonen's self-organizing maps (SOM), which shows to be useful not only to detect the above-mentioned parameter changes, but also to perform a sort of clustering of the given data. This is motivated by the SOM capacity to cluster and visualize high-dimensional data, looking for a subjacent structure in the input data. SOMs have been widely used for clustering in different applications, including speech signal processing [9]. The performance and robustness of the new method proposed here is tested in the presence of noise in numerical simulations and applied to vowel segmentation in diphthongs, which are the phoneme sequences that present more difficulties in continuous speech [10,11].

This paper is organized as follows: In Section 2, we present a brief summary of basic concepts of entropy, continuous wavelet transform, entropy temporal evolution, continuous multiresolution entropy and self-organizing maps. In Section

sinc(*i*) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
H. M. Torres, J. Gurlekian, H. L. Rufiner & M. E. Torres; "Self-organizing map clustering based on continuous multiresolution entropy"
Physica A: Statistical Mechanics and its Applications. Vol. 361, No. 1, pp. 337--354, 2006.

3 we describe how we combine these tools in order to perform the desired clustering. Experiments and results are presented and discussed in Section 4. Conclusions are in Section 5.

2. Definitions

2.1. Entropy evolution

The Shannon entropy of a random variable x is defined as [12]

$$H_y = -\int_{\mathbb{R}} f(x)\,\ln(f(x))\,\mathrm{d}x\,, \tag{1}$$

where $f(x)$ is the density function of x, and the integration spreads only over the region where $f(x)\neq 0$, with the agreement that $u\ln(u)=0$ if $u=0$.

In order to define the *temporal entropy evolution* (TEE), we consider the Shannon entropy of the random variable x as seen through a sliding temporal window W_t of a given length $L\in\mathbb{R}$. For $t\in\mathbb{R}$, let $x^{W_t}(u)$ be the function x restricted to interval $W_t=[t-L/2, t+L/2]\subset\mathbb{R}$ and we define $H_x^L(t)$ as [13,14]

$$H_x^L(t) = -\int_{\mathbb{R}} f(x^{W_t})\,\ln(f(x^{W_t}))\,\mathrm{d}x^{W_t}\,. \tag{2}$$

In a similar way, the parametric entropy or q-entropy, which depends on a real parameter $q\neq 1$, known as Tsallis entropy [15], is given by [16,17]

$$H_x^q = (q-1)^{-1}\int_{\mathbb{R}} f(x)[1-f(x)^{q-1}]\,\mathrm{d}x\,. \tag{3}$$

As before, we define the *temporal q-entropy evolution* (TqEE) $H_x^{q,L}(t)$ as [18]

$$H_x^{q,L}(t) = (q-1)^{-1}\int_{\mathbb{R}} f(x^{W_t})[1-f(x^{W_t})^{q-1}]\,\mathrm{d}x^{W_t}\,. \tag{4}$$

2.2. Continuous wavelet transform

In order to analyze a signal $x(t)\in L^2(R)$, the continuous wavelet transform (CWT) [19,20] splits it up making inner products with a collection of functions $\varphi_{a,b}(t)=|a|^{-1/2}\varphi([t-b]/a)$, which are dilated and translated versions of a given *mother wavelet* $\varphi(t)$:

$$d_x(a,b) = |a|^{-1/2}\int_{\mathbb{R}} x(t)\,\varphi^*\!\left(\frac{t-b}{a}\right)\mathrm{d}t\,, \tag{5}$$

where φ^* indicates the complex conjugated of φ and $\varphi(t)$ is an oscillatory function whose Fourier transform $\Phi(\omega)$ satisfies

$$C_\varphi = 2\pi\int_{\mathbb{R}} |\omega|^{-1}\,|\Phi(\omega)|^2\,\mathrm{d}\omega < \infty\,. \tag{6}$$

sinc(*i*) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
H. M. Torres, J. Gurlekian, H. L. Rufiner & M. E. Torres; "Self-organizing map clustering based on continuous multiresolution entropy"
Physica A: Statistical Mechanics and its Applications. Vol. 361, No. 1, pp. 337--354, 2006.

The CWT provides a natural tool for the time-frequency analysis because each $\varphi_{a,b}$ is predominantly located in a certain region of the time-frequency plane with a central frequency that is inversely proportional to the scale a.

2.3. Continuous multiresolution entropy

Let $D = \{d_x(a,b),\, (a,b) \in \mathbb{R}^2\}$ be the set of the CWT of $x(t)$ as defined in Eq. (5). At each fixed scale a, and for $t \in \mathbb{R}$, we consider the temporal window $W_{a,t}$ of length $L \in \mathbb{R}$, defined by $W_{a,t} = [t - L/2, t + L/2]$.

We define the continuous multiresolution entropy (CME), corresponding to the Shannon entropy, as [14]

$$CME_x^L(a,t) = -\int_{\mathbb{R}} f(d_x^{a,t}) \ln(f(d_x^{a,t}))\, \mathrm{d}d_x^{a,t}\ , \tag{7}$$

and the one corresponding to the Tsallis entropy, the continuous multiresolution q-entropy (CMqE), as [14]

$$CMqE_x^L(a,t) = (1-q)^{-1} \int_{\mathbb{R}} f(d_x^{a,t})[1 - (f(d_x^{a,t}))^{q-1}]\, \mathrm{d}d_x^{a,t}\ , \tag{8}$$

where $d_x^{a,t}$ is $d_x(a,b)$ restricted to $b \in W_{a,t}$ for each scale a.

In this way a multiresolution measure is obtained. As the temporal variable t evolves in $\mathbb{R}$, at a fixed scale a, the $CME_x^L(a,t)$ represents the temporal evolution of the wavelet coefficients entropy in the sliding temporal window $W_{a,t}$.

2.4. Kohonen self-organizing maps

An SOM [9] is a non-supervised neural net model that allows to group and visualize high-dimensional data. The SOM aim is to find a subjacent structure in the input data.

An SOM defines a map from an N-dimension input space to a D-dimensional neural arrangement, where, for $j = 1, \ldots, J$, we note as $\mathbf{z}_j$ the input vector and each neuron i has associated a reference vector $\mathbf{w}_i = [w_1^i, \ldots, w_N^i]$, with $i = 1, \ldots, M$, M being the number of neurons, which has to be specified at the design stage. An example for $D = 1$ is shown in Fig. 1.

During an SOM training process, two stages can be distinguished: the first one is the ordering of the reference vectors $\mathbf{w}_i$ and the second one is the convergence of such vectors. Both stages are iterative. At the ordering stage, for each iteration $v = 0, \ldots, V$, which will be indicated as a super index in $\mathbf{w}_i$ ($\mathbf{w}_i^v$), can be formulated as:

(1) Initialize the reference vectors $\mathbf{w}_i^0$ in a random way.

(2) At each iteration v, a $\mathbf{z}_j$ vector is randomly chosen from the input data and its similarity with the reference vectors $\mathbf{w}_i^{v-1}$ is evaluated using, for example, the Euclidean distance. The neuron c which best approximates $\mathbf{z}_j$ is called the ‘winner

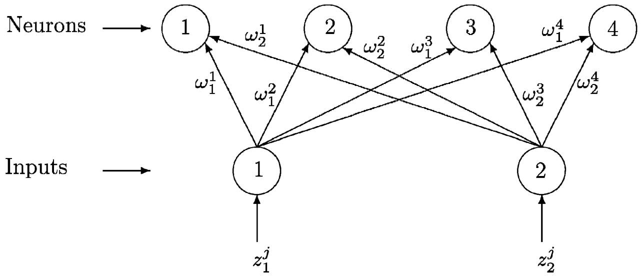

Fig. 1. Scheme of a one-dimensional SOM array, with input $z_j = [z_1^j z_2^j]$, and four elements with reference vector $w_1 = [w_1^1 w_2^1]$, $w_2 = [w_1^2 w_2^2]$, $w_3 = [w_1^3 w_2^3]$ and $w_4 = [w_1^4 w_2^4]$.

neuron' and it is given by

$$c = \arg\min_i \|\boldsymbol{z}_j - \boldsymbol{w}_i^{v-1}\| . \tag{9}$$

(3) The reference vectors are updated to $\boldsymbol{w}_i^v$:

$$\boldsymbol{w}_i^v = \boldsymbol{w}_i^{v-1} + h_i^{v,c}[\boldsymbol{z}_j - \boldsymbol{w}_i^{v-1}] . \tag{10}$$

The function $h_i^{v,c}$, called 'neighborhood function', defines the map topology, and connects the neurons in the arrangement. Usually $h_i^{v,c} = h(\|r_c - r_i\|, v)$, where $r_c \in \mathbb{R}^D$ and $r_i \in \mathbb{R}^D$ are the location vectors of nodes c and i, respectively, in the arrangement. Here we have considered a Gaussian neighborhood function, $h_i^{v,c}$, as

$$h_i^{v,c} = \alpha(v) \exp\left[-\frac{\|r_c - r_i\|^2}{2\,\sigma^2(v)}\right] , \tag{11}$$

where $\sigma(v)$ corresponds to the neighborhood width and $\alpha(v) \in [0, 1)$ is the learning coefficient. Both $\alpha(v)$ and $\sigma(v)$ are monotonically decreasing functions of algorithm iterations. In our case

$$\sigma(v) = 1 + [\sigma(0) - 1]\frac{[V - v]}{V} , \tag{12}$$

and

$$\alpha(v) = \alpha(0)\left[1 - \frac{v}{V}\right] , \tag{13}$$

where $\sigma(0)$ and $\alpha(0)$ must be defined in the design stage. This adaptation rule makes the reference vectors in the neighborhood of c move toward the input vector. The approximation degree is given by the neighborhood function $h_i^{v,c}$, which gradually decreases as the patterns are presented.

(4) Repeat steps (2)–(3) a fixed number of times V.

For the convergence stage, the steps (2)–(3) are repeated for the new values of V, $\sigma(0)$ and $\alpha(0)$.

sinc(*i*) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
H. M. Torres, J. Gurlekian, H. L. Rufiner & M. E. Torres; "Self-organizing map clustering based on continuous multiresolution entropy"
Physica A: Statistical Mechanics and its Applications. Vol. 361, No. 1, pp. 337--354, 2006.

sinc(*i*) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
H. M. Torres, J. Gurlekian, H. L. Rufiner & M. E. Torres; "Self-organizing map clustering based on continuous multiresolution entropy"
Physica A: Statistical Mechanics and its Applications. Vol. 361, No. 1, pp. 337--354, 2006.

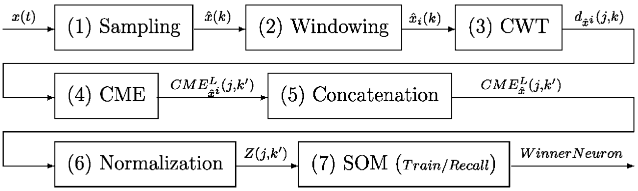

Fig. 2. Schemes of the stages of the proposed method, which are explained in the text.

When the training is finished, the vectors $\boldsymbol{w}^V_i$ approximate the probability density function of the input data. Another important characteristic of the SOM is its generalization capacity; it means that if a new input is presented to the SOM it can be identified by its closest neuron based on the training achieved from inputs previously seen.

3. Self-organizing map clustering based on continuous multiresolution entropy

In this section, we summarize the algorithm proposed in this work for automatic detection and clustering of parameter changes in the underlying dynamics of signal $x(t)$. Let us assume that either one signal or a set of signals, $x(t)$, holds the parameter changes we want to detect. Below we describe the steps corresponding to the *CME*-based algorithm (see Fig. 2). In a similar way it can be performed using the *CMqE*. It involves the following steps:

(1) Let $x(t)$ be the temporal evolution of a given signal and let us consider its discrete evolution $\hat{x}(k)$, obtained by a regular sampling, $\hat{x}(k) = x(k\Delta)$, where $k \in \mathbb{Z}$ and Δ is the sampling rate.
(2) For $i = K/2, (K/2) + (K - 2L), (K/2) + 2(K - 2L), \ldots, i_{max}$ we consider the temporal window $W_{i,k}$ of length $K \in \mathbb{Z}$, defined by $W_{i,k} = [i - (K/2), i + (K/2)]$. Let $\hat{x}^i(k)$ stand for $\hat{x}(k)$ at the temporal window $W_{i,k}$ centered i.
(3) For each $\hat{x}^i(k)$, the corresponding wavelet coefficient sets D^i are obtained, with $D^i = \{d_{\hat{x}^i}(j,k)\}$. At this step a suitable wavelet has to be chosen.
(4) For each scale j the entropy is evaluated by sliding windows of length L and a shift of m. In this way the CME is calculated, obtaining $CME^L_{\hat{x}^i}(j,k')$ by using the discrete version of Eq. (7). In order to see the computational details, please refer to Ref. [21].
(5) With the matrices $CME^L_{\hat{x}^i}$,the matrix $CME^L_{\hat{x}}$ is formed by concatenation: $CME^L_{\hat{x}} = [CME^L_{\hat{x}^0} CME^L_{\hat{x}^1} \ldots CME^L_{\hat{x}^i} \ldots CME^L_{\hat{x}^{imax}}]$.
(6) For each scale j, the corresponding $CME^L_{\hat{x}}(j,k')$ is statistically normalized to zero mean and unit standard deviation, obtaining $Z(j,k')$, the statistically normalized matrix associated with $CME^L_{\hat{x}}$. The matrix Z qualitatively reveals the occurrence

of a slight parameter change in the underlying nonlinear dynamical system by mean of a jump (up or down) at all the scales. However, this information that appears to be redundant does not show the same intensity in the different scales.

(7) Each column of Z is used like an input vector to train an SOM.[3] At this point, we have to select the topology, dimensions and training parameters of the map. In general, no a priori rule exists to fix these parameters, and for each case there exists an optimal set of them. The dimensions of the map and neuron disposition in the arrangement are the more critical items. Once the SOM is trained, temporal evolution of the winner neuron will be considered as the system output. In such a way, changes in the winner neuron represent variations in the system parameters, and for parameters of similar values, neighbor winner neurons will be obtained. We have to emphasize that for different training data set and/or SOM parameters different output configurations will be obtained, since these depend on the training, although the data represent the same configurations of the system.

In this work, a linear arrangement ($D = 1$) of neurons was defined, since it allows an easy visualization and analysis of results. The number of neurons was empirically determined for each of the test signals in order to assure a minimum of neurons equal to the number of steady parameter sets that we want to identify. Extra neurons were added to capture at least one transitional parameter set between steady parameter sets. An excessive number of neurons would allow the network to capture more information of the system, resulting in more than one neuron for each parameter set. The training parameters are chosen to make the SOM to converge. In general, it is accepted that this is obtained with small values of learning coefficients and great number of iterations.

A possible criterion to select the wavelet could be to select one with the best localization properties. For example, Mexican Hat is acknowledged to have the best temporal localization properties [19] or Morlet wavelet is useful in locating singularities [20]. Nevertheless, even if these properties give good hints about which wavelet would be the best suited, the ultimate wavelet is, in most cases, empirically selected and *the best one* might change from signal to signal. The signal's frequency range to be analyzed and the number of voices N (i.e., the length of the logarithmic partition of the frequency range) must be established at this point. This frequency is normalized, as is usual in wavelet analysis. If the frequency range is established as $[f_{min}, f_{max}]$, then $a_0^{N-1} = f_{max}/f_{min}$. As usual in the "quasi-continuous" wavelet decomposition, in order to compute the CWT (Eq. (5)), we consider a dyadic partition of the scales ($a = 2^j$, $j \in \mathbb{Z}$) and a uniform partition of the temporal variable ($t = k\Delta$, $k \in \mathbb{Z}$), which gives rise to a *quasi-continuous* time-scale plane representation $d_{\hat{x}^i}(a = 2^j, t = k\Delta) = d_{\hat{x}^i}(j, k)$ of the sampled data $\hat{x}^i(k)$. If a complex wavelet is used, for example the Morlet wavelet of order 5, the coefficients $d_x(j, k)$ are complex and then $(|d_{\hat{x}}(j, k)|^2)$ has to be considered as matrix D, i.e., the time-scale

[3] In the present work, the SOM-PAK [22] software was used in order to carry out the experiments described in Section 2.4.

sinc(*i*) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
H. M. Torres, J. Gurlekian, H. L. Rufiner & M. E. Torres; "Self-organizing map clustering based on continuous multiresolution entropy"
Physica A: Statistical Mechanics and its Applications. Vol. 361, No. 1, pp. 337--354, 2006.

sinc(i) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
H. M. Torres, J. Gurlekian, H. L. Rufiner & M. E. Torres; "Self-organizing map clustering based on continuous multiresolution entropy"
Physica A: Statistical Mechanics and its Applications. Vol. 361, No. 1, pp. 337--354, 2006.

matrix containing the squared magnitude of the CWT. In this paper we have computed the CWT using the time-frequency toolbox (TFTB).[4]for Matlab.[5]

4. Results and discussion

In this section, some preliminary applications of the automatic detector based on the CME–CMqE and SOM described in the previous section are first illustrated through a simulated example corresponding to logistic map. Then we present the results with synthesized and real speech diphthong signals.

4.1. Logistic equation

In this section we introduce the results obtained with our method applied to a well-known nonlinear system, the logistic map, given by

$$x(k+1) = a(k)x(k)[1 - x(k)]\,. \tag{14}$$

In order to perform the experiments, a signal has been generated according to Eq. (14). Parameter $a(k)$ varies slightly from $a_0 = 3.545$ to $a_f = 3.562$ by steps of $\Delta_a = 0.001$. This variation is smoothed by dividing the sampling region $k \in [1, 9 \times 10^4]$ in subintervals $I_n = [2501 + (n-1)\,10^4/2\,2501 + n\,10^4/2]$ for $n = 1, 2, \ldots, 18$, and defining the parameter variation $a(k)$ in each interval according to

$$a(k) = \frac{a_{1,n} + a_{2,n}}{2} + \frac{a_{2,n} - a_{1,n}}{\pi} \arctan\left[\frac{k - k_{c,n}}{r}\right], \tag{15}$$

where $a_{1,n}$ and $a_{2,n}$ are the initial and final parameter values in the nth interval, $k_{c,n}$ is the central point of the change in I_n, and r is a fixed change radius. Setting $a_{1,1} = a_0$, we obtain $a_{2,1} = a_0 + \Delta_a = a_{1,2}$ and $a_{2,18} = a_f$. Observe that this means that the major parameter changes in each interval is concentrated around $k_{c,n}$ with a radius r, even if the parameter itself changes very slowly in each interval I_n, as can be seen in Fig. 3 for the initial interval ($n = 1$), where the parameter a varies, according to Eq. (15), for $a_{1,1} = 3.545$, $a_{2,1} = 3.546$. This slight variation of the parameter allows to test the segmenting and the clustering capacity of the method. We fixed $r = 100$ and $k_{c,n} = 5000 + (n-1) \times 10^4/2]$ for all n (see Fig. 4(a)).

For the range of frequencies, from $f_{min} = 0.1$ to $f_{max} = 0.45$, the corresponding 40 CWT scales using a Mexican Hat wavelet have been evaluated, with window length $K = 5000$. Then, the CME corresponding to a sliding window, with $L = 500$ samples and shifted every $m = 100$ samples, has been obtained. To estimate the probabilities, 20 bins were used. With the purpose of avoiding border effects, the first window of

[4] TFTB has been developed by Francois Auger, Olivier Lemoine, Paulo Gonçalves and Patrick Flandrin, with the support of the CNRS (France) and Rice University (1995–1996). It is freely available on the WWW at http://perso.wanadoo.fr/francois.auger/tftb.html or at http://gdr-isis.org/Applications/tftb/iutsn.univ-nantes.fr/auger/tftb.html. It includes a reference manual and a tutorial.

[5] Matlab® *The MathWorks, Inc.*, v. 7. http://www.mathworks.com.

sinc(*i*) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
H. M. Torres, J. Gurlekian, H. L. Rufiner & M. E. Torres; "Self-organizing map clustering based on continuous multiresolution entropy"
Physica A: Statistical Mechanics and its Applications. Vol. 361, No. 1, pp. 337--354, 2006.

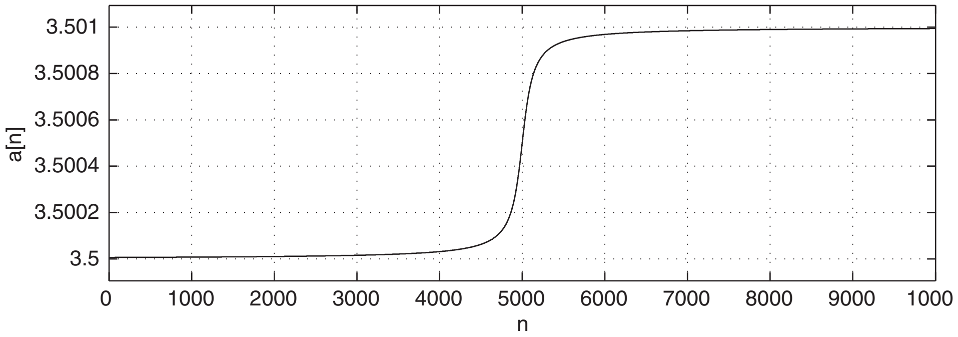

Fig. 3. Parameter evolution a according to Eq. (15), with $a_{1,1} = 3.545$, $a_{2,1} = 3.546$, $r = 100$ and $k_{c,1} = 5000$.

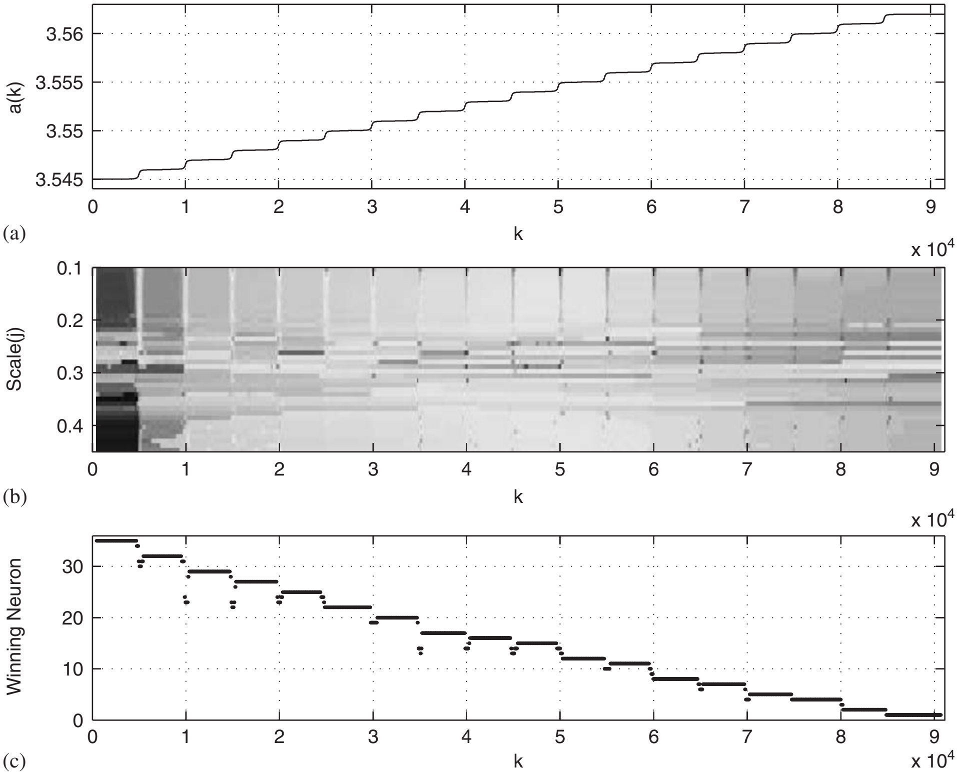

Fig. 4. (a) Time evolution of test signal parameter a, generated with Eq. (15); (b) CME and (c) the winner neuron. Clearly, each parameter value a corresponds to a neuron.

the CME was not taken into account. Finally, the matrix CME has been normalized. With these data, a linear SOM with $M = 35$ neurons was trained using a Gaussian neighborhood. For the ordering stage, we used an initial learning coefficient

sinc(i) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
H. M. Torres, J. Gurlekian, H. L. Rufiner & M. E. Torres; "Self-organizing map clustering based on continuous multiresolution entropy"
Physica A: Statistical Mechanics and its Applications. Vol. 361, No. 1, pp. 337--354, 2006.

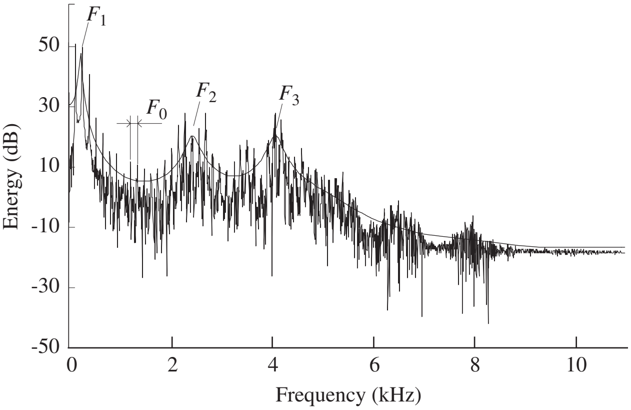

Fig. 5. Power spectrum, and its envelope of a Spanish vowel /i/ pronounced in sustained form, where the formant frequencies have been highlighted.

$\alpha(0) = 0.005$, an initial neighborhood $\sigma(0) = 35$ and a number of iterations $V = 500.000$. For the convergence stage, we used $\alpha(0) = 0.0015$, $\sigma(0) = 2$ neurons and $V = 700.000$ iterations.

Results are presented in Fig. 4. Comparing the first and third plots, we can see that the method not only detects the change in the parameter but also clusters them, assigning a neuron in particular to the parameter value. The previous experiments have been run also using CMqE, for a value of $q = 1.8$, obtaining similar results.

4.2. *Speech signals—diphthongs*

The phoneme segmentation consists of the division of a speech emission in different phonetic units, that is to say phonemes that form part of it [23,24]. The objective here is to insert a temporal mark indicating the beginning and the end of each phoneme, together with the label of the identified phoneme. Different techniques were used in order to carry out this task [25–29,2]. A preliminary attempt to segment speech using entropy was presented in Ref. [30]. In spite of this, there are situations in which the problem has not yet been solved, such as in the presence of noise or in real-time applications.

Automatic segmentation of vowels in diphthongs is one of the most difficult tasks due to the lack of a clear acoustic boundary during transitions between vowels. In Fig. 5 main resonances[6] of the vocal tract can be observed as peaks—called formants—superimposed over the Fourier power spectrum of a vowel /i/. The first three steady state formants over time constitute a way to characterize isolated vowels. As diphthongs are essentially a combination of two vowels, a corresponding set of formant transitions or glides appears between them (see Fig. 8). Fluctuations on continuous speaking rate and adjacent phoneme context produce nonlinear

[6] Estimated from an autoregressive model applied to a time windowed speech signal.

sinc(i) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
H. M. Torres, J. Gurlekian, H. L. Rufiner & M. E. Torres; "Self-organizing map clustering based on continuous multiresolution entropy"
Physica A: Statistical Mechanics and its Applications. Vol. 361, No. 1, pp. 337--354, 2006.

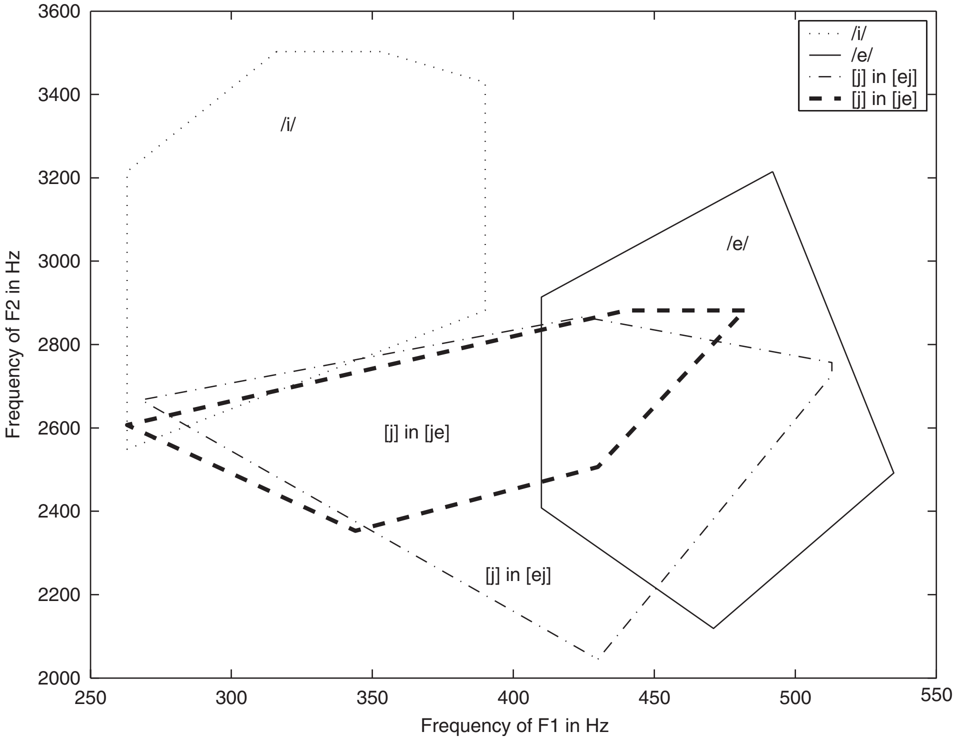

Fig. 6. Comparison between isolated vowels and the positional variants of /i/, in [je] and [ej] diphthongs (adapted from [32]).

variations on both durations of steady state segments and transitional segments. In fast speech, steady state segments of closed vowels /i/ and /u/ tend to disappear and only transitions and the open vowel formants remain available. Considering the 14 vocalic combination pairs, those composed by the pairs /ei/ and /ou/ are the most difficult ones to separate, due to the proximity of their formant frequencies. In running speech, vowels /i/ and /u/ are produced as semi-consonants [j] and [w] where these vowels reduce their duration, intensity and move their formant frequencies toward those corresponding to /e/ and /o/, respectively. Fig. 6 shows the movement of vowel /i/ produced as a semi-consonant [j] in both [je] and [ej] diphthongs. As a reference the F_1–F_2 area of the isolated vowels /i/ and /e/ can be observed in the background. Clearly, F_2 of /i/ is lowered and overlaps with F_2 of /e/. F_1 of /i/ tends to increase and overlaps with F_1 of /e/. We considered both diphthongs within a syllable in a word, and those diphthongs are produced by word fusion extracted from continuous speech under a variety of prosodic conditions.

4.2.1. Artificially generated diphthongs

In order to verify the capacity of our method to reveal the parameter change of a system, we synthesized the speech emissions corresponding to diphthong [je]. We

sinc(*i*) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
H. M. Torres, J. Gurlekian, H. L. Rufiner & M. E. Torres; "Self-organizing map clustering based on continuous multiresolution entropy"
Physica A: Statistical Mechanics and its Applications. Vol. 361, No. 1, pp. 337--354, 2006.

Table 1
Parameters entered into the synthesizer to generate the test diphthongs

Signal	Initial/final time (s)	F_1 (KHz)	F_2 (KHz)	F_3 (KHz)
1	0.070/0.140	.250/.400	3.0/2.0	3.5/2.6
2	0.065/0.140	.330/.400	2.5/2.0	3.0/2.6
3	0.065/0.125	.215/.400	3.0/1.7	3.5/2.3
4	0.075/0.140	.266/.530	3.0/2.0	3.5/2.6
5	0.070/0.135	.250/.530	3.0/2.0	3.5/2.6
6	0.060/0.095	.250/.550	3.0/2.0	3.5/2.6
7	0.075/0.115	.280/.500	3.0/2.0	3.5/2.6
8	0.060/0.085	.280/.500	3.0/2.0	3.5/2.5
9	0.055/0.105	.200/.370	3.0/2.0	3.5/2.6
10	0.055/0.115	.200/.400	3.0/2.0	3.5/2.6

used a Klatt [31] synthesizer in order to generate 10 signals, using 3 formants and frequency parameters, time and amplitude taken at random, with the premise that the generated waves would correspond to the Spanish diphthong [je]. For all the signals the energy was fixed at 60 dB and the fundamental frequency was kept at 150 Hz. In Table 1 the parameters used are shown and Fig. 7(a) is a diagram of the parameters given to the synthesizer in order to obtain the diphthong numbered 1. Steady state formants were added before initial transition time and after final transitional time, to complete a total duration of 200 ms. Furthermore, the transitions are smoothed by the synthesizer.

In these experiments our purpose was to detect the beginning and the end of the transition from one vowel to another. With this in mind, 50 CWT scales have been calculated using a Mexican Hat wavelet, from $f_{min} = 0.1$ to $f_{max} = 0.5$. For the CME calculation, a sliding window with $L = 500$ (0.05 s) samples was used, with a shift $m = 50$ (0.005 s) samples, and in the probability estimation, 10 bins were used. After normalizing the data obtained with the CME, a linear SOM with $M = 5$ neurons was trained using a Gaussian neighborhood. For the ordering stage, we used an initial learning coefficient $\alpha(0) = 0.005$, an initial neighborhood $\sigma(0) = 5$ and a number of iterations $V = 500.000$. For the convergence stage, we used $\alpha(0) = 0.0015$, $\sigma(0) = 1$ neurons and $V = 700.000$ iterations. Here we used all the signals to train and test the SOM.

The detection time of the beginning of the parameter change has a root mean square error (RMSE) of 0.0076 s, with respect to the real ones, with a standard deviation (SD) of 0.0065 s. At the end of the transition an RMSE of 0.011 s, and an SD of 0.011 s were obtained. See Fig. 7(c) for stimuli number 1 output. Taking into account the fixed fundamental frequency of 150 Hz, which is a cycle duration of 0.0066 s, we can say that the results obtained in average do not go beyond two cycles.

4.2.2. *Diphthongs extracted from natural speech*

Going ahead with our "leit motiv", we proceed with the segmentation of diphthongs extracted from continuous speech emissions. Samples were extracted, as

sinc(i) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
H. M. Torres, J. Gurlekian, H. L. Rufiner & M. E. Torres; "Self-organizing map clustering based on continuous multiresolution entropy"
Physica A: Statistical Mechanics and its Applications. Vol. 361, No. 1, pp. 337--354, 2006.

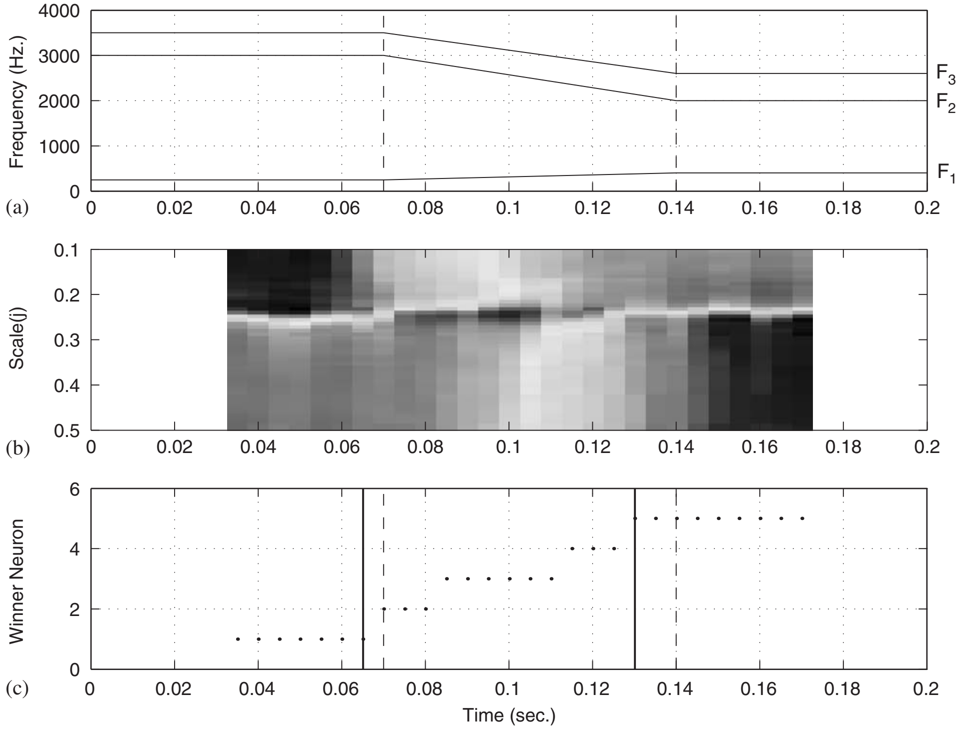

Fig. 7. (a) Diagram of the formant evolution used for the synthesis of the stimuli number 1; (b) CME and (c) the winner neuron. The dashed lines correspond to the beginning and end of the change of the parameters. The solid line corresponds to the segmentation performed by our method.

in the previous example, from a prosodic database [33] containing most (97%) of the Spanish syllables in all allowed positions within a word and in both stressed and unstressed conditions. Diphthongs [ej] appeared 74 times in total, [je] appeared 199 times, [ow] appeared 32 times and [wo] appeared 9 times.

In Fig. 8 we present a portion of the speech waveform excerpted from the sentence [kaDa kamjon karGa entre kinse j Bejnte mil pesos][7] ("Each truck loads between fifteen and twenty thousand pesos"). Phoneme labels, wide band FFT spectrogram and the first three formants are also displayed. We can appreciate that the variation in the formants F_2 and F_3 in between 3.3 and 3.4 s, i.e., when the diphthong [ej] is pronounced, is very smooth.

The corresponding CME was obtained. CWT was calculated using a Mexican Hat wavelet, with 40 scales, $f_{min} = 0.1$ and $f_{max} = 0.5$. In CME analysis, we used a sliding window of $L = 512$ samples (0.032 s) width and an $m = 100$ samples (3.1 ms) shift. For the probability estimation, 10 bins were used. With the obtained CME matrices,

[7] [B;D;G] are representing the approximate allophones of phonemes /b,d,g/.

sinc(*i*) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
H. M. Torres, J. Gurlekian, H. L. Rufiner & M. E. Torres; "Self-organizing map clustering based on continuous multiresolution entropy"
Physica A: Statistical Mechanics and its Applications. Vol. 361, No. 1, pp. 337--354, 2006.

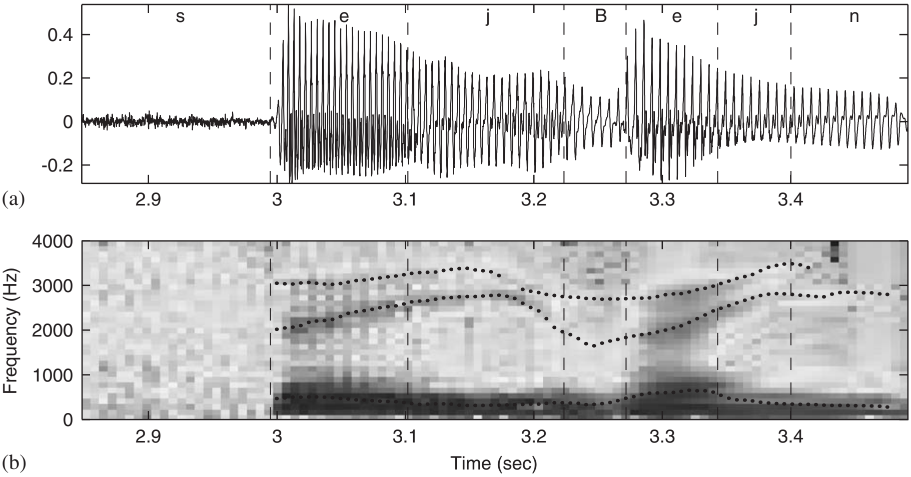

Fig. 8. (a) Speech waveform portion and phoneme labels (on top), and (b) wide band FFT spectrogram, and the first three formants, corresponding to the phone sequences [se j Bejn], showing diphtongs [ej].

an SOM was trained in a similar way as the previous example, but with only three neurons.

In Fig. 9 an example of typical outputs for the pair [ej] is shown. The net activates the neuron number 1 before the presence of /e/, and the neuron number 3 for [j]. In the transition, the winner neuron is number 2. The dashed lines correspond to the segmentation found in the database, i.e., it corresponds to a human labeler mark, and the solid lines correspond to the detection found with our method. With the aim of testing the performance of the method proposed here, we have calculated the RMSE and the SD between the segmentation marks in the database and the time instant given by the middle point in the transition from neuron number 1 to number 3 (that is to say, the middle point of the neuron number 2). The values obtained are shown in Table 2. Taking into account that the considered speech signals had a fundamental frequency of 200 Hz, in average the RMSE does not exceed four cycles, i.e., 20 ms, relative to manual labels.

4.2.3. *Robustness to additive noise*

In order to explore the robustness of our method in the presence of additive noise, we have performed the same experiments as in Section 4.2.2 but adding noise to the signals corresponding to diphthong [ej]. Noisy signals were used both for training and testing stages. Two kinds of noises have been used: white and babble.[8] The signal noise ratio (SNR) is a measure of the strength of a wanted signal relative to the

[8] The Signal Processing Information Base (SPIB), of Rice University, Houston, USA, http://spib.rice.edu/spib.html, has been used.

sinc(*i*) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
H. M. Torres, J. Gurlekian, H. L. Rufiner & M. E. Torres; "Self-organizing map clustering based on continuous multiresolution entropy"
Physica A: Statistical Mechanics and its Applications. Vol. 361, No. 1, pp. 337--354, 2006.

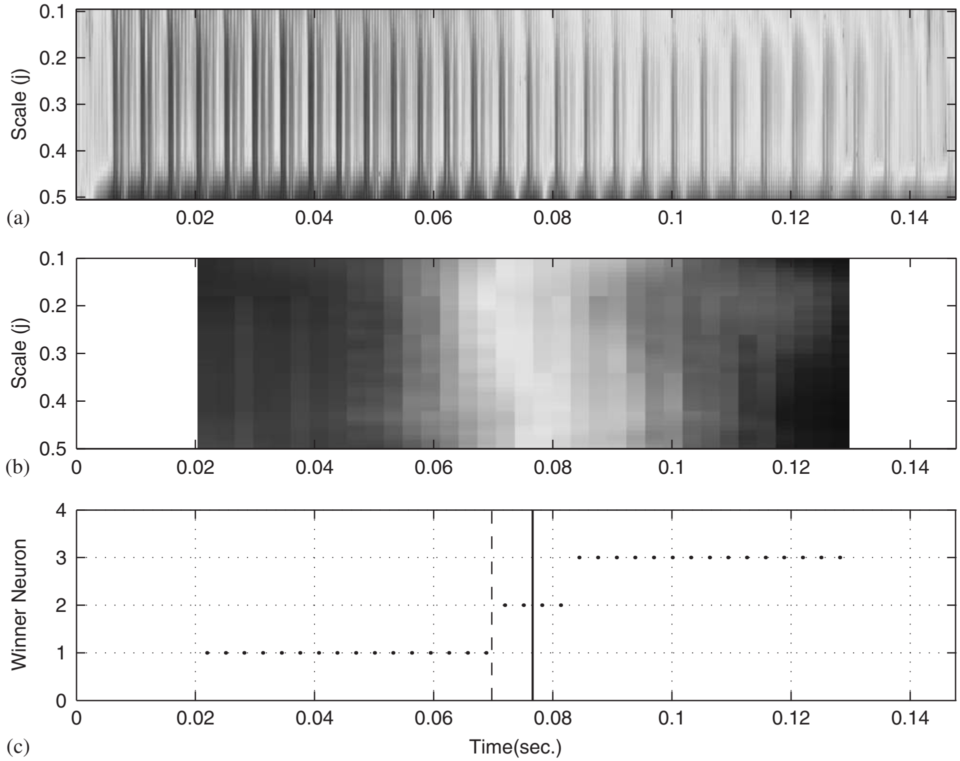

Fig. 9. (a) Scalogram, (b) CME and (c) winner neurons of diphthong [ej] in the context “intenté influir” (“I tried to influence”). The dashed line corresponds to the segmentation found in the database, and the solid line corresponds to the detection found with our method.

Table 2
RMSE and SD obtained by comparing the label marks of the database and the ones generated by our method

	[ej]	[je]	[ow]	[wo]
RMSE (in s)	0.020	0.023	0.017	0.016
SD (in s)	0.017	0.019	0.017	0.017

amount of background noise and it is defined as

$$SNR = 10\log_{10}\frac{P(SignalWithoutNoise)}{P(Noise)}\ , \tag{16}$$

where $P(.)$ indicates the power of the corresponding signal, and it is expressed in decibels (dB). Observe that $SNR = 0\,\text{dB}$ if $P(SignalWithoutNoise) = P(Noise)$ means as much noise power as signal, which means less SNR value than the regular

sinc(*i*) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
H. M. Torres, J. Gurlekian, H. L. Rufiner & M. E. Torres; "Self-organizing map clustering based on continuous multiresolution entropy"
Physica A: Statistical Mechanics and its Applications. Vol. 361, No. 1, pp. 337--354, 2006.

Table 3
RMSE and SD obtained for the signals with additive noise

Performance with white noise										
SNR	−20	−15	−10	−5	0	5	10	15	100	∞
RMSE	0.038	0.036	0.028	0.021	0.018	0.015	0.016	0.017	0.020	0.020
SD	0.029	0.028	0.023	0.019	0.018	0.015	0.015	0.016	0.017	0.017
Performance with babble noise										
SNR	−20	−15	−10	−5	0	5	10	15	100	∞
RMSE	0.064	0.059	0.048	0.030	0.022	0.017	0.018	0.019	0.020	0.020
SD	0.035	0.053	0.046	0.030	0.022	0.016	0.016	0.017	0.017	0.017

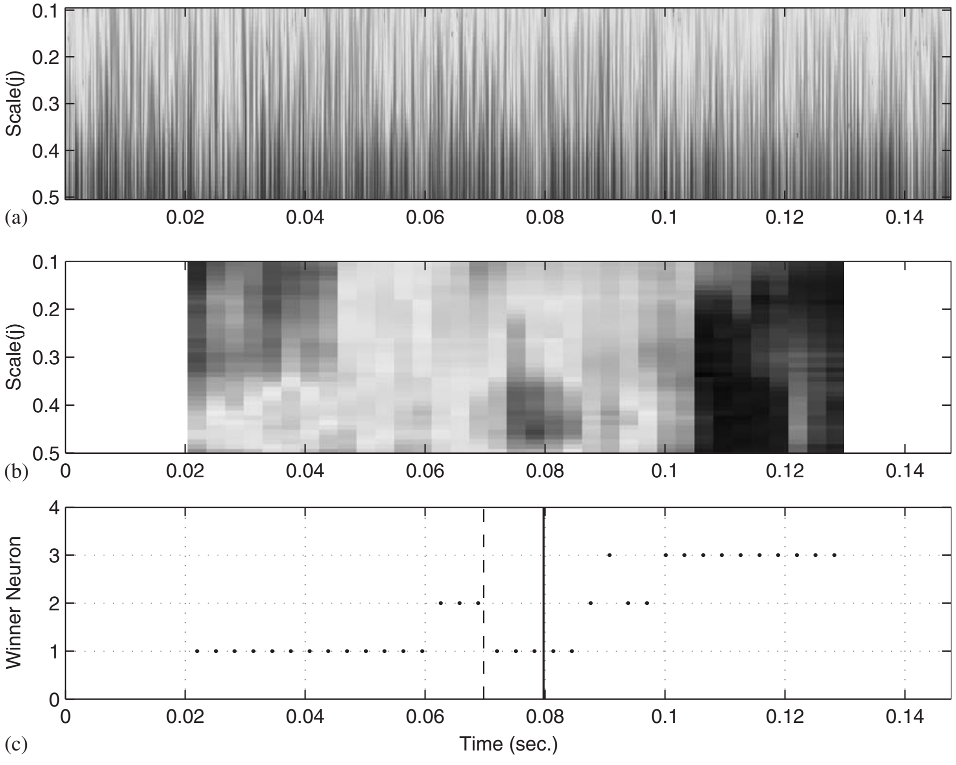

Fig. 10. (a) Scalogram, (b) CME and (c) winner neurons of diphthong [ej] in the context “intenté influir” (“I tried to influence”) with 0 dB of SNR (white noise). The dashed line corresponds to the segmentation found in the database, and the solid line corresponds to the detection found with our method.

observed in speech signals. Table 3 shows that for both kinds of noise, even for $SNR = 0\,\mathrm{dB}$, the method did not suffer degradation in its performance, obtaining RMSE and SD values equivalent to those obtained with the signal without noise, i.e., for $SNR = \infty$.

sinc(*i*) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
H. M. Torres, J. Gurlekian, H. L. Rufiner & M. E. Torres; "Self-organizing map clustering based on continuous multiresolution entropy"
Physica A: Statistical Mechanics and its Applications. Vol. 361, No. 1, pp. 337--354, 2006.

As can be seen in Fig. 10, for SNR equal or lower than 0, the net outputs have variations between the expected output (neuron 1 or 3, whatever corresponds) and intermediate neuron (neuron number 2). These variations in the SOM output increase as the SNR diminishes.

5. Conclusions

In this paper, a variant of the automatic detector proposed in Ref. [8] has been presented. This new tool combines the capability of continuous multiresolution entropy to highlight slight parameter changes in nonlinear dynamic systems with the clustering abilities of a Kohonen self-organizing map. In addition, this method not only detects variations in the parameters, but also allows to identify steady parameter sets. Its ability to detect slight parameter changes has been tested on simulated signals, provided by a toy model, and on both synthetic and real speech signals. The capacity of the method proposed here to detect smooth variations in the parameters in both natural and noisy conditions has been established.

The results obtained for natural diphthongs are successfully compared with manual labeling performed by a phonetician, where the 20 ms error obtained is equivalent to the human phoneme detection threshold. As a promissory result for further studies in real situations, we can mention the robustness of the method in the presence of additive noise with SNR starting at 0 dB. This SNR value represents a very high masking condition, even for human perception.

References

[1] J.P. Eckmann, D. Ruelle, Ergodic theory of chaos and strange attractors, Rev. Mod. Phys. 57 (57) (1985) 617–656.

[2] T. Li, J. Gibson, Speech analysis and segmentation by parametric filtering, IEEE Trans. Speech Audio Process. 4 (3) (1996) 207–220.

[3] A. Wolf, J.B. Swift, H.L. Swinney, J.A. Vastano, Determining Lyapunov exponents from a time series, Physica D 16 (1985) 285–317.

[4] L.D. Iasemidis, J.C. Sackellares, H.P. Zaveri, W.J. Williams, Phase space topography and the Lyapunov exponent of electrocorticograms in partial seizures, Brain Topogr. 2 (3) (1990) 187–201.

[5] J. Van Neerven, Determination of the correlation dimension from a time series, applications to rat EEGs: sleep, theta rhythm and epilepsy, Ph.D. Thesis, Department of Exp. Zoology, University of Amsterdam, 1988.

[6] H. Abarbanel, R. Brown, J. Sidorowich, L. Tsimring, The analysis of observed chaotic data in physical systems, Rev. Modern Phys. 65 (4) (1993) 1331–1392.

[7] P. Abry, Ondelettes et Turbulences. Multirésolutions, algorithmes de décomposition, invariance d'échelle et signaux de pression, Diderot Multimedia, France, 1997.

[8] M.E. Torres, M.M. Añino, L.G. Gamero, M.A. Gemignani, Automatic detection of slight changes in nonlinear dynamical systems using multiresolution entropy tools, Int. J. Bifurc. Chaos 11 (4) (2001) 967–981.

[9] T. Kohonen, Self-organizing maps, Springer Series in Information Sciences, vol. 30, Springer, Berlin, Heidelberg, 1995 (second extended edition 1997).

[10] A.M. Borzone Manrique, Manual de Fonética Acústica, Hachette, Buenos Aires, 1980.

[11] P. Delattre, Comparing the Phonetic Features of English, French, German, and Spanish, Julius Groos Verlag Heidelberg, Santa Barbara, California, USA, 1965.

[12] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (3) (1948) 379–423.

[13] M. Torres, L. Gamero, E. D'Attellis, Pattern detection in eeg using multiresolution entropy, Latin Am. Appl. Res. 53 (1995) 53–57.

[14] M.E. Torres, L. Gamero, P. Flandrin, P. Abry, On a multiresolution entropy measure, in: A.F. Laine Akram Aldroubi, M. Unser (Eds.), SPIE'97 Wavelet Applications in Signal and Image Processing V, vol. 3169, SPIE International Society for Optical Engineering, Washington, 1997, pp. 400–407.

[15] C. Tsallis, Somm comments on Boltzmann–Gibbs statistical mechanics, Chaos Solitons Fractals 6 (1995) 539–559 (and references therein).

[16] Z. Daróczy, Generalized information functions, Inf. Control 16 (1970) 36–51.

[17] J. Havrda, F. Charvat, Quantification method of classification process: concept of structural α-entropy, Kybernetica 3 (1967) 30–35.

[18] M.E. Torres, El procesamiento de señales ligadas a problemas no lineales, Ph.D. Thesis, Universidad Nacional de Rosario—Argentine, 1999 (Math.D. Thesis).

[19] I. Daubechies, Ten Lectures on Wavelets, Society for Industrial and Applied Mathematics, Philadelphia, USA, May 1992.

[20] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, San Diego, California, USA, 1999.

[21] M.E. Torres, L.G. Gamero, Relative complexity changes in time series using information measures, Physica A 286 (3–4) (2000) 457–473.

[22] T. Kohonen, J. Hynninen, J. Kangas, J. Laaksonen, SOM_PAK: The self-organizing map program package, Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science, January 1996.

[23] D. Reddy, Segmentation of speech sound, JASA (1966) 307–312.

[24] J. Hemeret, Automatic segmentation of speech, IEEE Trans. Signal Process. 39 (4) (1991) 1008–1012.

[25] J. Deller, J. Proakis, J. Hansen, Discrete Time Processing of Speech Signals, Macmillan Publishing, New York, 1993.

[26] A. Vortersman, J. Martens, B. Coile, Automatic segmentation and labeling of multi-lingual speech data, Speech Comunication 19 (1996) 271–293.

[27] C. Jeong, H. Jeong, Automatic phone segmentation and labeling of continuous speech, Speech Comunication 20 (1996) 291–311.

[28] P. Grassberger, T. Schreiber, C. Schaffrath, Nonlinear time sequence analysis, Int. J. Bifurc. Chaos 1 (3) (1995) 521–547.

[29] D.H. Milone, J.J. Merelo, H.L. Rufiner, Evolutionary algorithm for speech segmentation, in: Proceedings of the 2002 IEEE World Congress on Evolutionary Computation, Paper No. 7270, IEEE Press, 2002.

[30] W. Wokurek, Corpus based evaluation of entropy rate speech segmentation, in: Proceedings of the International Conference on Phonetic Sciences, ICPhS International Conference on Phonetic Sciences, San Francisco, 1999, pp. 1217–1220.

[31] D. Klatt, Software for a cascade/parallel formant synthesizer, JASA 67 (1980) 971–995.

[32] A.M. Borzone de Aronson, An acoustic study of /i/, /u/ in the Spanish diphtongs, Language and Speech 19 (2) (1976) 121–128.

[33] J. Gurlekian, H. Rodriguez, L. Colantoni, H. Torres, Development of a prosodic database for an argentine spanish text to speech system, in: B. Bird, Liberman (Eds.), Proceedings of the IRCS Workshop on Linguistic Databases, SIAM, University of Pennsylvania, Philadelphia, USA, December 2001, pp. 99–104.

sinc(*i*) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
H. M. Torres, J. Gurlekian, H. L. Rufiner & M. E. Torres; "Self-organizing map clustering based on continuous multiresolution entropy"
Physica A: Statistical Mechanics and its Applications. Vol. 361, No. 1, pp. 337--354, 2006.