



Feature selection for spontaneous speech analysis to aid in Alzheimer's disease diagnosis: A fractal dimension approach

KarmeLe López-de-Ipiña^{a,*}, Jordi Solé-Casals^b, Harkaitz Eguiraun^c,
J.B. Alonso^d, C.M. Travieso^d, Aitzol Ezeiza^a, Nora Barroso^a,
Miriam Ecay-Torres^e, Pablo Martinez-Lage^e, Blanca Beitia^f

^a Department of Systems Engineering and Automation, University of the Basque Country, Donostia, Spain

^b Data and Signal Processing Research Group, University of Vic – Central University of Catalonia, Vic, Spain

^c Research Center for Experimental Marine Biology and Biotechnology, Plentzia Marine Station,
University of the Basque Country, UPV/EHU, Plentzia, Bizkaia, Spain

^d Signal and Communications Department, The Institute for Technological Development and Innovation on Communications,
University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain

^e Neurology Department CITA-Alzheimer Foundation, Donostia, Spain

^f Department of Mathematics, University of the Basque Country, Vitoria-Gasteiz, Spain

Received 15 October 2013; received in revised form 1 August 2014; accepted 18 August 2014

Available online 27 August 2014

Abstract

Alzheimer's disease (AD) is the most prevalent form of degenerative dementia; it has a high socio-economic impact in Western countries. The purpose of our project is to contribute to earlier diagnosis of AD and allow better estimates of its severity by using automatic analysis performed through new biomarkers extracted through non-invasive intelligent methods. The method selected is based on speech biomarkers derived from the analysis of spontaneous speech (SS). Thus the main goal of the present work is feature search in SS, aiming at pre-clinical evaluation whose results can be used to select appropriate tests for AD diagnosis. The feature set employed in our earlier work offered some hopeful conclusions but failed to capture the nonlinear dynamics of speech that are present in the speech waveforms. The extra information provided by the nonlinear features could be especially useful when training data is limited. In this work, the fractal dimension (FD) of the observed time series is combined with linear parameters in the feature vector in order to enhance the performance of the original system while controlling the computational cost.

© 2014 Elsevier Ltd. All rights reserved.

Keywords: Nonlinear speech processing; Alzheimer's disease diagnosis; Spontaneous speech; Fractal dimensions

* Corresponding author at: Department of Systems Engineering and Automation, University of the Basque Country, Polytechnic School, Europa Plaza 1, 20008 Donostia, Spain. Tel.: +34 653723757.

E-mail addresses: karme.le.ipina@ehu.es (K. López-de-Ipiña), jordi.sole@uvic.cat (J. Solé-Casals), harkaitz.eguirau@ehu.es (H. Eguiraun), jesus.alonso@ulpgc.es (J.B. Alonso), carlos.travieso@ulpgc.es (C.M. Travieso), aitzol.ezeiza@ehu.es (A. Ezeiza), nora.barroso@ehu.es (N. Barroso), mecay@cita-alzheimer.org (M. Ecay-Torres), pmlage@cita-alzheimer.org (P. Martinez-Lage), mariablanca.beitia@ehu.es (B. Beitia).

<http://dx.doi.org/10.1016/j.csl.2014.08.002>

0885-2308/© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Alzheimer's disease (AD) is the most common type of dementia among the elderly. It is characterized by progressive and irreversible cognitive deterioration with memory loss and impairments in judgment and language, together with other cognitive deficits and behavioral symptoms. The cognitive deficits and behavioral symptoms are severe enough to limit the ability of an individual to perform everyday professional, social or family activities. As the disease progresses, patients develop severe disability and full dependence. An early and accurate diagnosis of AD helps patients and their families to plan for the future and offers the best possibilities of treating the symptoms of the disease. According to current criteria, the diagnosis is expressed with different degrees of certainty as possible or probable AD when dementia is present and other possible causes have been ruled out. The diagnosis of definite AD requires the demonstration of the typical AD pathological changes at autopsy (McKhann et al., 1984, 2011; Van de Pole et al., 2005). The clinical hallmark and earliest manifestation of AD is episodic memory impairment. At the time of clinical presentation, other cognitive deficits are present in areas like language, executive functions, orientation, perceptual abilities and constructional skills (Morris, 1993; APA, 2000). All these symptoms lead to impaired performance in everyday activities. Approaches to the early diagnosis of AD have in the past few years made significant advances in the development of reliable clinical biomarkers (AA, 2014).

Despite the usefulness of biomarkers, the cost and technology requirements involved make it impossible to apply such tests to all patients with memory complaints. Given these problems, non-invasive intelligent techniques of diagnosis may become valuable tools for early detection of dementia. Non-technical staff in the habitual environments of the patient could use these methodologies, which include e.g. automatic spontaneous speech analysis (ASSA) (Fig. 1), without altering or blocking the patients' abilities, as the spontaneous speech involved in these techniques is not perceived as a stressful test by the patient. Moreover, these techniques are very low-cost and do not require extensive infrastructure or the availability of medical equipment. They are thus capable of yielding information easily, quickly, and inexpensively (Faundez-Zanuy et al., 2012; López-de-Ipiña et al., 2013a,b).

In addition to the loss of memory, one of the major problems caused by AD is the loss of language skills. We can detect different communication deficits in the area of language, including aphasia (difficulty in speaking and understanding) and anomia (difficulty in recognizing and naming things). The specific communication problems the patient encounters depend on the stage of the disease (McKhann et al., 2011; Van de Pole et al., 2005; Morris, 1993):

1. First stage or early stage (ES): difficulty in finding the right word in spontaneous speech. Often remains undetected.
2. Second stage or intermediate stage (IS): impoverishment of language and vocabulary in everyday use.
3. Third stage or advanced stage (AS): answers sometimes very limited and restricted to very few words.

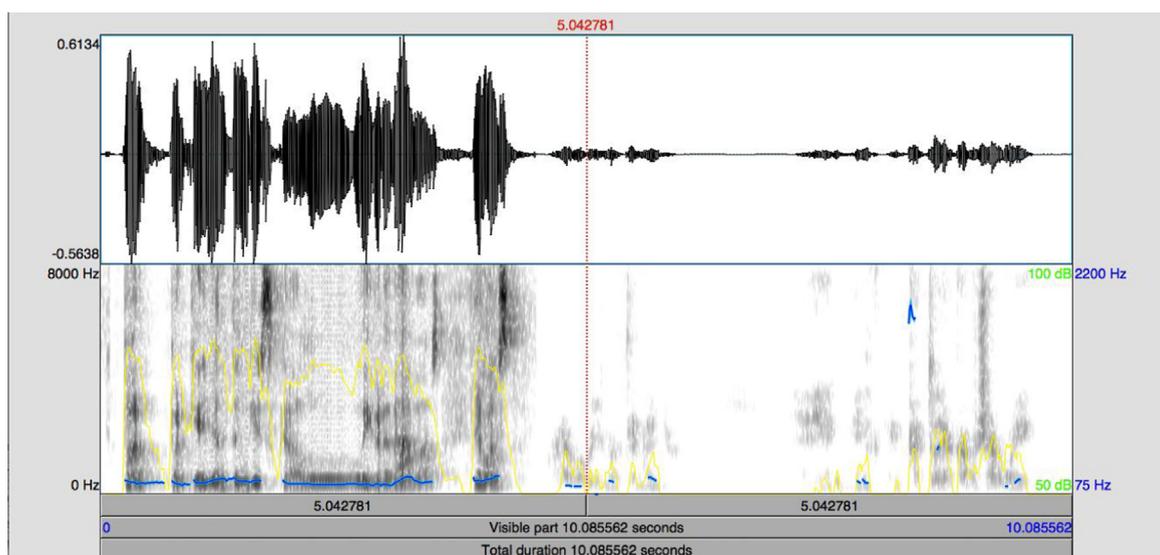


Fig. 1. Signal and spectrogram of a control subject (left) and a subject with AD (right) during spontaneous speech.

The main goal of the present work is feature search in spontaneous speech aiming at pre-clinical evaluation for the definition of test for AD diagnosis. These features will be used to define the group (CR) and the three AD levels. As phonological and articulatory impairments may occur at presentation or early in the course of Alzheimer's disease (Croota et al., 2000), we will use measures that can reflect the nonlinear nature of such changes in the acoustical signals. One of the most relevant nonlinear techniques for automatic speech recognition (ASR) is the consideration of the fractal dimension (FD) of the speech signal as a feature to be used in the training process. In general, fractal dimensions can be utilized to quantify the complexity, concerning the geometry of a dynamical system given its multidimensional phase-space. This quantification is related to the active degrees of freedom of the assumed dynamical system, providing a quantitative characterization of a system's state (Pitsikalis and Maragos, 2009; Maragos and Potamianos, 1999).

Biological systems are regulated by interacting mechanisms that operate across multiple spatial and temporal scales. The output variables of these systems often have complex fluctuations that are not solely due to noise but also contain information about the intrinsic dynamics. Time series generated by biological systems most likely contain deterministic and stochastic components (Costa et al., 2005). Classical methods of signal and noise analysis can quantify the degree of regularity of a time series by evaluating the appearance of repetitive patterns, but most such methods only model linear components without introducing any information about non-linearity, irregularities or stochastic components. This complex information could be essential when subtle changes are analyzed. Thus, when appropriate data are available, linear systems can be implemented fairly rapidly, as they rely on well-known machine learning techniques to achieve their goals, avoiding complex adjustments to the system.

The interest in fractals in speech date back to the mid-1980s (Pickover and Khorasani, 1986), and they have been used for a variety of applications, including consonant/vowel characterization (Martinez et al., 2003; Langi and Kinsner, 1995), speaker identification (Nelwamondo et al., 2006), and end-point detection (Li et al., 2007), even for whispered speech (Chen and Zhao, 2006). Recent research concerns the analysis of pathological voices through a fractal approach (Chouard et al., 2001; Ouayoun et al., 1999; Péan et al., 2000, 2002). The fractal dimension (FD) quantifies the roughness of a temporal signal and estimates its degree of freedom, and is thus a good approach to modeling its fluctuations. Moreover, the fractal approach allows one to quantify the roughness of the voice, between 1 (sinusoidal complex signal) and 2 (white noise). Maragos and Potamianos (1999) provide some motivation and justification from the field of speech aerodynamics for using fractal dimension to quantify the degree of turbulence in speech signals. They also explore differences in articulator trajectories and global intensity. In this research FD outperforms zero-crossings analysis in the case of unvoiced segments. Like in other work on AD diction based on biosignals (EEG and MEG) (Abásolo et al., 2008; Gómez et al., 2009), this methodology could be useful for detecting symptoms of Alzheimer's, such as doubts, breaks or silences in the speech signal. Figs. 2 and 3 show differences for a control subject (CR) and a person with AD in intensity, pitch and FD.

The approach adopted in this work is to improve the system developed in our previous work (López-de-Ipiña et al., 2013a,b), augmenting the features with FD. The FD is a well-known and used feature for describing the complexity of a system, and could help in the detection of subtle changes useful for early diagnosis. Moreover, this feature has the ability to capture the dynamics of the system and thus reveal relevant variations in speech utterances. More precisely, we use implementations of Higuchi's, Katz's and Castiglioni's algorithms (Higuchi, 1988; Katz, 1988; Castiglioni, 2010) in order to add this new feature to the set that feeds the training process of the model.

The remainder of this paper is organized as follows: in Section 2, the materials are presented. Section 3 explains the methodology of the experiments, Section 4 shows the experimental results, and finally, conclusions are presented in Section 5.

2. Materials

This study is focused on early AD detection and its objective is the identification of AD in the pre-clinical (before first symptoms) and prodromic (some very early symptoms but no dementia) stages. The research presented here is a complementary preliminary experiment to define thresholds for a number of biomarkers related to spontaneous speech. The feature search in this work is designed for pre-clinical evaluation to select appropriate tests for AD diagnosis. The data obtained will complement the biomarker set of each person in diagnosing AD.

In an attempt to develop a new methodology applicable to a wide range of individuals differing with regard to sex, age, language and cultural and social background, we have constructed a multicultural and multilingual (English, French, Spanish, Catalan, Basque, Chinese, Arabic and Portuguese) database with video recordings of 50 healthy and

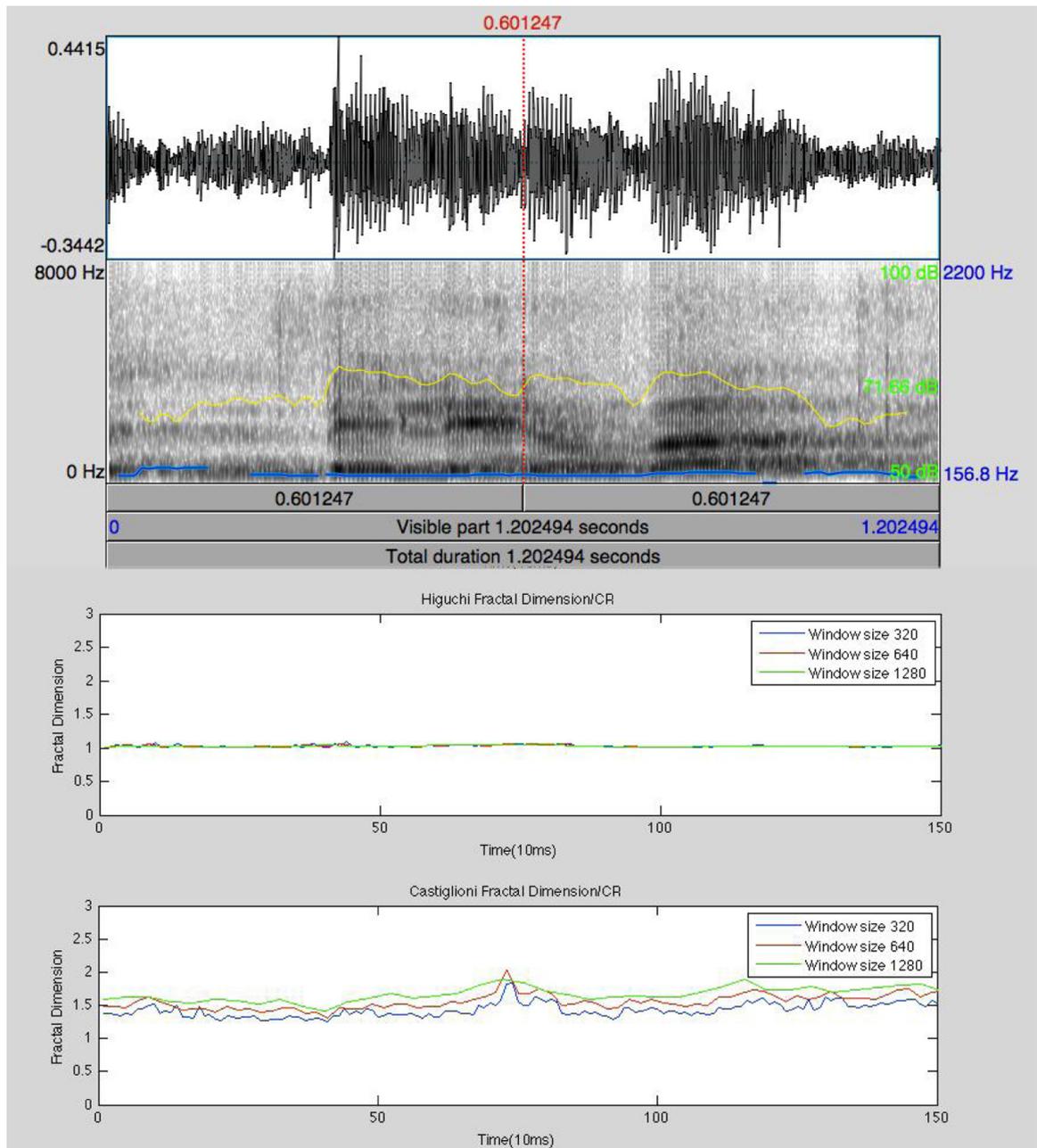


Fig. 2. Analysis of a control person (CR). Utterance in Basque language, “. . . *Belero bat. . .*”: (a) speech signal, pitch (blue) and intensity (yellow), (b) spectrogram, (c) Higuchi fractal dimension (CFD) and (d) Castiglioni fractal dimension (HFD). (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

20 AD patients (i.e., patients with a prior diagnosis of Alzheimer’s) recorded for 12 h and 8 h, respectively. The age span of the individuals in the database was 20–98 years and there were 20 males and 20 females. This database is called AZTIAHO. All the work was performed in strict accordance with the ethical guidelines of the organizations involved in the project.

The recordings consisted of videos of spontaneous speech – people telling pleasant stories or recounting pleasant feelings as well as interacting with each other in friendly conversation. The recording atmosphere was relaxed and non-invasive. The shorter recording times for the AD group are due to the fact that AD patients find speech more of

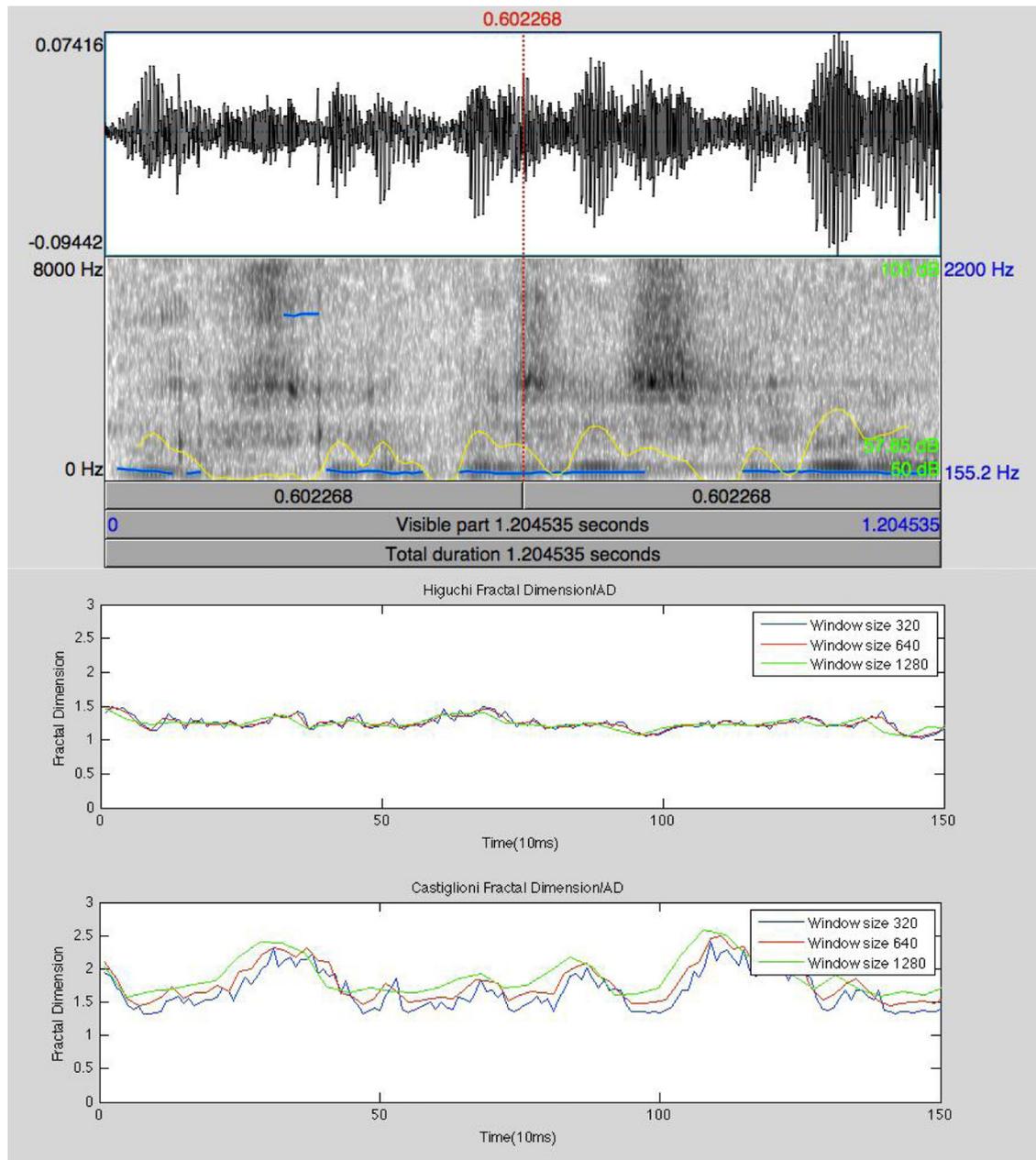


Fig. 3. Analysis of a person with AD (AD), utterance in Spanish, “No se lo que es eso. . .”: (a) speech signal, pitch (blue) and intensity (yellow), (b) spectrogram. (c) Higuchi fractal dimension (CFD) and (d) Castiglioni fractal dimension (HFD). (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

an effort than healthy individuals: they speak more slowly, with longer pauses, and with more time spent on looking for the correct word and uttering speech disfluencies or break messages. In the advanced stage of the disease, they find this effort tiring and often want to stop the recording. We complied with their requests. The video was processed and the audio extracted in wav format (16 bits and 16 kHz). The first step was removing non-analysable events: laughter, coughing, short hard noises and segments where speakers overlapped. Next, background noise was removed using denoiser adaptive filtering. After the pre-processing, about 80% of the material from the control group and 50% of the material from the AD group remained suitable for further analysis. The complete speech database consists of about 60 min of material for the AD group and about 9 h for the control. The speech was next divided into consecutive

segments of 60 s in order to obtain appropriate segments for all speakers, resulting finally in a database of about 600 segments of spontaneous speech.

Finally, in order to perform our experiments a subset of 20 AD patients was selected (68–96 years of age, 12 women, 8 men) with a distribution in the three stages of AD as follows: first stage [ES = 4], secondary stage [IS = 10] and tertiary stage [AS = 6]. The control group (CR) was made up of 20 individuals (10 male and 10 female, aged 20–98 years) and representing a wide range of speech responses. This subset of the database is called AZTIAHORE.

3. Methods

3.1. Feature extraction

3.1.1. Features selected for automatic spontaneous speech analysis (ASSA)

Spoken language is one of the most important elements defining an individual's intellect, social life, and personality; it allows us to communicate with each other, share knowledge, and express our cultural and personal identity. Spoken language is the most spontaneous, natural, intuitive, and efficient method of communication among people. Therefore, the analysis by automated methods of spontaneous speech (SS – free and natural spoken communication), possibly combined with other methodologies, could be a useful non-invasive method for early AD diagnosis (López-de-Ipiña et al., 2013a,c). The analysis of spontaneous speech fluency is based on three families of features (SSF set), obtained by the *Praat* software package (Praat) and software that we ourselves developed in MATLAB. For that purpose, an automatic *Voice Activity Detector* (VAD) (VAD; Solé and Zaiats, 2003) has extracted voiced/unvoiced segments as parts of an acoustic signal (López-de-Ipiña et al., 2013a,b).

The three families of features include (López-de-Ipiña et al., 2013b):

- (1) *Duration*: the histogram calculated over the most relevant voiced and unvoiced segments, the average of the most relevant voiced/unvoiced, voiced/unvoiced percentage and spontaneous speech evolution along the time dimension, and the voiced and unvoiced segments' mean, max and min;
- (2) *Time domain*: short time energy;
- (3) *Frequency domain, quality*: spectral centroid.

The energy of a signal is typically calculated on a short-time basis, by windowing the signal at a particular time, squaring the samples and taking the average. The spectral centroid is commonly associated with the measure of the brightness of a sound. This measure is obtained by evaluating the “center of gravity” using the Fourier transform's frequency and magnitude information.

3.1.2. Fractal dimension

Most fractal systems have a characteristic called self-similarity. An object is self-similar if a close-up examination of the object reveals that it is composed of smaller versions of itself. Self-similarity can be quantified as a relative measure of the number of basic building blocks that form a pattern, and this measure is defined as the fractal dimension.

There is not any precise reference of the FD value a given waveform should have. In addition, speech waveforms are not stationary, so most ASR techniques employ short sections of the signal in order to extract features from the waveform. This means that one plausible technique for extracting features from speech waveforms, for the purpose of recognizing different phonemes, is to divide the signal in short chunks and calculate the features for each chunk. This was the approach we adopted. In other words, we calculated the fractal dimension of short segments of the waveform and observed the evolution of the obtained values along the whole signal, with the aim of finding in it fractal characteristics that could help in identifying different elements of the spoken message.

There are several algorithms for measuring the fractal dimension. In the current work we focus on the alternatives that are especially suited for time series analysis and do not require previous modeling of the system. Three of these algorithms are Higuchi (1988), Katz (1988) and Castiglioni (2010), named for their authors. Higuchi and Castiglioni were chosen because they have been reported to be more accurate in previous work with under-resourced conditions (Ezeiza et al., 2013; López-de-Ipiña et al., 2013b,c). Katz is also reported as a robust algorithm to calculate fractal dimension (Esteller et al., 2001).

Higuchi (1988) proposed an algorithm for measuring the fractal dimension of discrete time sequences directly from the time series $x(1), x(2), \dots, x(N)$. The algorithm is based on a new time series, x_m^k , constructed from the original one, as following:

$$x_m^k = \left\{ x(m), x(m+k), x(m+2k), \dots, x\left(m + \left\lfloor \frac{N-m}{k} \right\rfloor k\right) \right\}, \quad \text{for } m = 1, 2, \dots, k \quad (1)$$

where $\lfloor a \rfloor$ means integer part for a , m indicates the initial time value and k indicates the discrete time interval between points (delay). For each of the time series x_m^k constructed, the average length $L_m(k)$ is computed as:

$$L_m(k) = \frac{\sum_{i=1}^{\lfloor \frac{N-m}{k} \rfloor} |x(m+ik) - x(m+(i-1)k)| (N-1)}{\lfloor \frac{N-m}{k} \rfloor k} \quad (2)$$

Next, the length of the curve for the time interval k is defined as the sum value over k sets of $L_m(k)$ as showed in Eq. (3):

$$L(k) = \sum_{m=1}^k L_m(k) \quad (3)$$

Finally, the slope of the curve $\ln(L(k))/\ln(1/k)$ is estimated using least squares linear best fit, and the result is the Higuchi fractal dimension (HFD).

On the other hand, Katz (1988) proposed a normalized formula of the fractal dimension (see Eq. (3)) because this fractal dimension depends on the particular units of measure.

The FD of a planar curve can be defined as:

$$FD = \frac{\log_{10}(L)}{\log_{10}(d)} \quad (4)$$

In this case L is the length of the curve and d is the diameter (the planar extent) of the curve. For time series, which are ordered sets of (x, y) point pairs, the total length L is nothing but the sum of the distances between successive points, as presented in Eq. (5):

$$L = \sum_{i=1}^n l_{i,i+1} \quad (5)$$

where $l_{i,j}$ means the Euclidean distance between two point pairs (x_i, y_i) and (x_j, y_j) as shown in Eq. (5), and $n+1$ is the total number of point pairs.

$$l_{i,j} = \sqrt{(y_i - y_j)^2 + (x_i - x_j)^2} \quad (6)$$

The diameter of the time series can be calculated as the farthest distance between the starting point (point 1) and any other point (point i) of the time series, as presented in Eq. (6):

$$d = \max\{l_{1,i}\} \quad (7)$$

The FD compares the actual number of units in a curve with the minimum number of units required to reproduce a pattern of the same spatial extent; and as mentioned above, will depend upon the measurement units used because FD is calculated in a discrete space not in continua. Katz's approach solves this problem by normalizing d and L by the length of the average step a , the average distance between successive points, defined as L/n , where $n+1$ is the total number of point pairs (as presented in Eq. (5)). Therefore, normalizing distances in (4) we obtain Katz's approach to calculate the FD of a time series (8):

$$FD = \frac{\log_{10}(n)}{\log_{10}(n) + \log_{10}(d/L)} \quad (8)$$

However, Castiglioni (2010) claims that in the X - Y plane, which describes the waveform, the magnitudes x and y in use are intrinsically different since they correspond to the magnitude of the signal (y) and time points (x). Therefore, given that the input signal is a mono-dimensional waveform, the length and the extension can be rewritten using

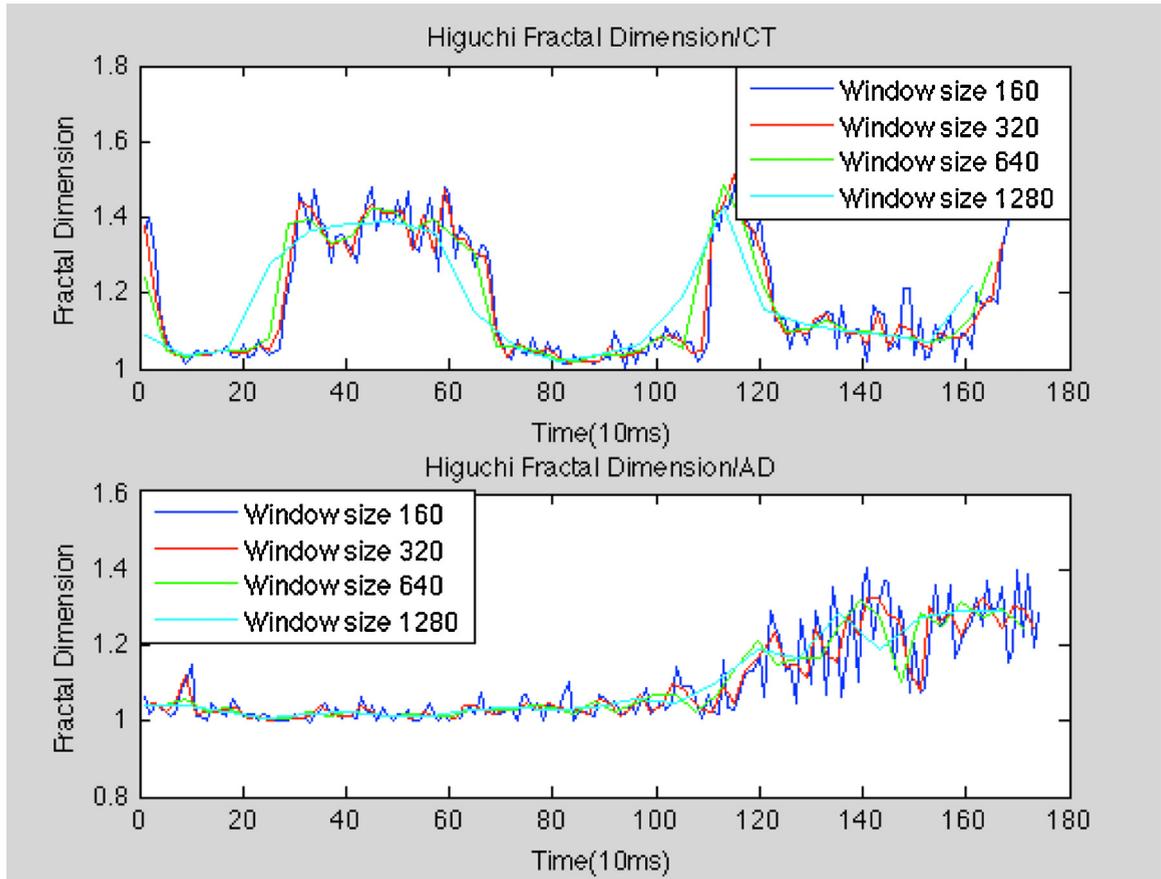


Fig. 4. Higuchi fractal dimension (HFD) for a control subject (above) and an AD subject (below), for different window sizes.

Mandelbrot's approach. A simple and efficient way to do this is to measure these two magnitudes directly in their own dimension. For each, the extension on the y -axis is the range of y_k :

$$d = \max\{y_k\} - \min\{y_k\} \quad (9)$$

and the length L is the sum of all the increments, in modulus:

$$L = \sum_{k=1}^n |y_{k+1} - y_k| \quad (10)$$

Where again, d is the diameter of the curve, L is the length of the curve and $n + 1$ is the total number of point pairs.

3.1.3. Window size during the feature extraction process

The selection of an appropriate window size to be used during the experiments is essential. Fig. 4 illustrates the qualitative effect of window size on the results of the experiments. Broadly speaking, the fractal dimension is a tool for attempting to capture the dynamics of the system. With a short window, the estimation is highly local and adapts fast to the changes in the waveform. When the window is longer, some details are lost but the fractal dimension better anticipates the characteristics of the signal. Additionally, previous studies that take into account the window size of similar dimension estimations (Tsonis, 2007; Jang, 2011; Esteller et al., 2001) suggest that a longer window could be useful in some cases. Consequently, four window-sizes of 160, 320 and 1280 points will be analyzed.

Table 1
Accuracy (%) in the final system with MLP and KNN for CR and AD.

Name of the set	Description
SSF	Spontaneous speech features
SSSF	Automatic selection of spontaneous speech features
SSSF + FD	SSSF set and fractal dimension (FD)
VFD	Maximum, minimum, variance, standard deviation, median and mode average for full signal and voiced signal
SVFD	Automatic selection of maximum, minimum, variance, standard deviation, median and mode average for full signal and voiced signal

Table 2
Accuracy (%) in the final system with MLP (SSSF + CSVFD) and KNN (SSSF + HSVFD) for CR and AD (WS = 320) and confidence interval for 95% and 90%.

Class	KNN			MLP		
	Acc	CI (95%)	CI (90%)	Acc	CI (95%)	CI (90%)
CR	90.90	[84.09, 97.90]	[85.21, 96.79]	92.43	[86.85, 99.15]	[87.83, 98.16]
AD	87.30	[79.07, 95.52]	[80.39, 94.20]	90.9	[84.10, 98.03]	[85.06, 96.93]

3.2. Automatic feature selection

In order to develop applications with low computational cost, feature sets should be optimized. In a first approach, three algorithms will be used for automatic attribute selection. Specifically, the following ones from the WEKA software (WEKA) suite were selected to take into account gain relative to each class, each feature and the relations among them:

1. *SVMAttributeEval*: Evaluates the usefulness of an attribute by using an SVM classifier. Attributes are ranked by the square of the weight assigned by the SVM. Attribute selection for multiclass problems is handled by ranking attributes for each class separately using a one-versus-all method and then dealing from the top of each pile to give a final ranking (WEKA).
2. *CfsSubsetEval*: Evaluates the usefulness of a subset of attributes by considering the individual predictive ability of each feature, along with the degree of redundancy between them. Subsets of features that are highly correlated with the class while features having low intercorrelation are preferred (WEKA).
3. *GainRatioAttributeEval*: Evaluates the usefulness of an attribute by measuring the gain ratio with respect to the class (WEKA).

These algorithms will be used over the initial feature sets to select smaller and appropriate attribute sets for low computational cost performance.

3.3. Feature sets

In the experimentation, three families of feature sets will be used:

1. SSF and SSSF
2. SSSF + FD
3. SSSF + VFD and SSSF + SVFD

From these three families we derive five different sets that will be used in the following experiments, as detailed in Table 1 (see also Table 3 for a complete list of acronyms used in this work).

3.4. Automatic classification

The main goal of the present work is feature search in spontaneous speech oriented to pre-clinical evaluation for the definition of tests for AD diagnosis. These features will define the CR group and the three AD levels. A secondary goal will be the control of computational cost oriented to real-time applications. Thus automatic classification will be modeled with a view toward efficiency. Two different paradigms will be evaluated:

1. Multi layer perceptron (MLP) with neuron number in hidden layer (NNHL) = (attribute/number + classes/number) and training step (TS) $NNHL \times 10$, MLPA500.
2. k -Nearest neighbors (KNN) paradigm.

The WEKA software suite has been used in carrying out the experiments. The results were evaluated using Classification Error Rate (CER), Accuracy (Acc) and Cumulative Error (Arias-Londoño et al., 2010; Godino et al., 2005, 2006). For the training and validation steps, we used k -fold cross-validation with $k=10$. Cross-validation is a robust validation method for variable selection (Picard and Cook, 1984). Repeated cross-validation (as calculated by the WEKA environment) allows robust statistical tests. We also use the measurement provided automatically by WEKA “Coverage of cases” (0.95 level) and confidence interval for percentages (CI) for 95%, 90% and 80%.

4. Experimental results

The experimentation was carried out with AZTIAHORE dataset.

4.1. Analysis of the fractal dimension

In this first stage about 80 attributes were generated with regard to the criteria described above: features selected for automatic spontaneous speech analysis (ASSA) and features based on the fractal dimension. Then, automatic feature selection was carried out using the methodology described in 3.2 for all three fractal dimension algorithms and four different window-sizes (160, 320, 640 and 1280 points).

4.2. Automatic feature selection

Automatic feature selection was carried out using the methodology described in Section 3.2. For all three fractal dimension algorithms and four different window-sizes (160, 320, 640 and 1280 points), 12 similar proposals of feature sets were analyzed. About 50 features were selected from the original sets. The selected attributes are the same for all cases: features relative to ASSA and features based on the fractal dimension. Finally following feature families were used in the next Section 4.3: SSF, selection of SSF (SSSF), SSSF + FD, full sets of features (SSSF + VFD) and sets with the selected features (SSSF + SVFD). SSSF set outperforms the CER rates in % of SSF (MLP, 24.81% and KNN, 29.46%) in about 5%. Specifically, the reference results CER in % for SSSF, without FD, are: MLP, 20.16% and KNN, 27.14% (see Ref blue bar in Fig. 5). Recognition rates for both algorithms are optimal for the CR group but very poor for ES. Thus SSSF was selected for use in the subsequent experiments described below.

4.3. Automatic classification

In the next stage the methodology described in Section 3.4 was used. The task was automatic classification, with the classification targets being: healthy speakers without neurological pathologies and speakers diagnosed with AD. Four kinds of experiments were carried out: selection of window-size; feature analysis; analysis of global results; analysis of classes' results.

Fig. 5 shows the classification error rate in % for fractal dimension algorithms and both paradigms: MLP and KNN. In this task, several window sizes were analyzed: 160, 320, 640 and 1280 points. The SSSF set and fractal dimension sets were used in this stage. The minimum CER is achieved with MLP and windows of middle size: 320 and 640. Both window-sizes will be used in subsequent experiments. Most of the new proposals, which include the fractal dimension, achieved improvements over the reference methods. Fig. 6 analyses the influence of attribute selection with

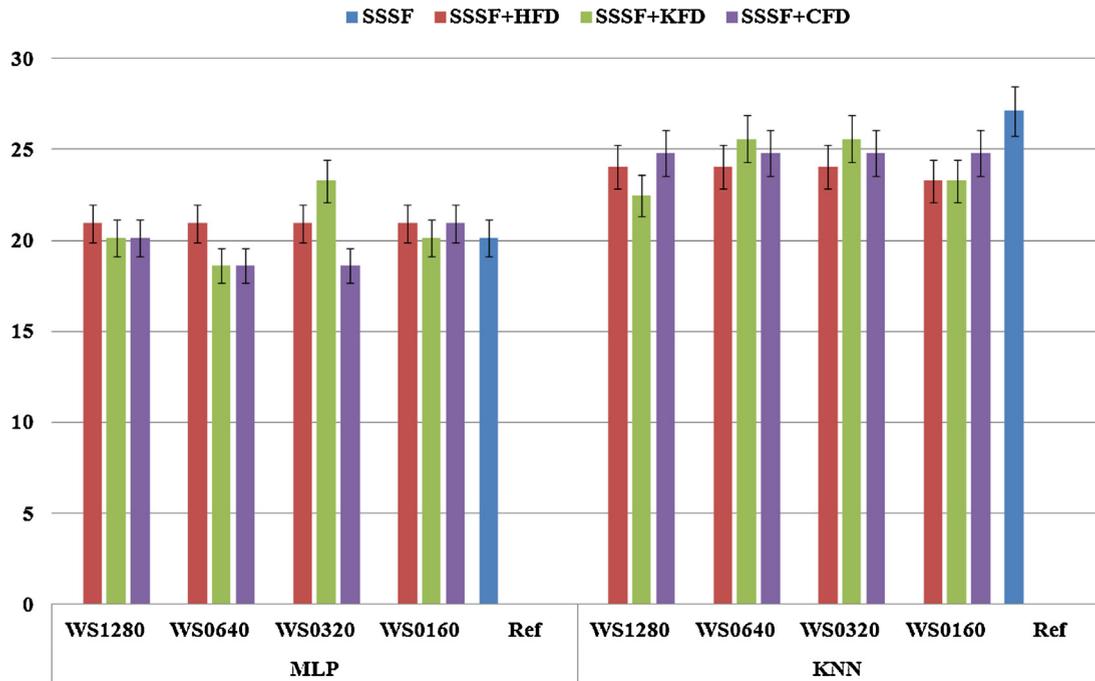


Fig. 5. Classification error rate (%) for fractal dimension algorithms, several window sizes and both paradigms: MLP and KNN. Ref bars were obtained without using fractal dimension features.

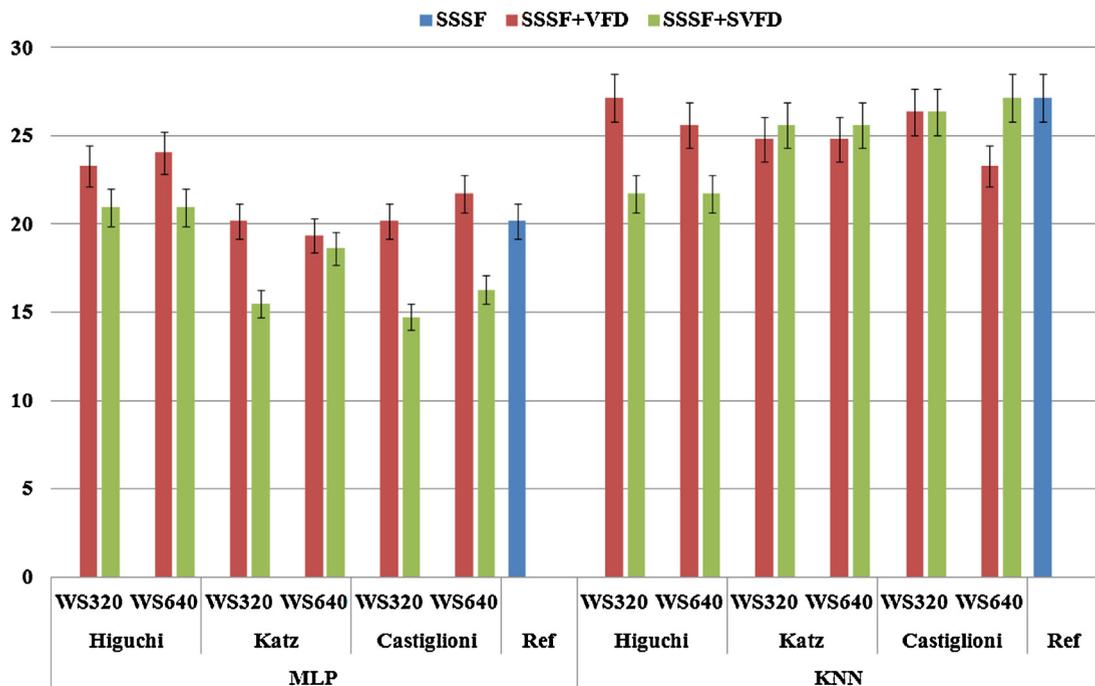


Fig. 6. Classification error rate (%) for several window sizes and both paradigms: MLP and KNN with fractal dimension feature sets (VFD) and attribute sets after selection (SVFD). Ref bars were obtained without using fractal dimension features.

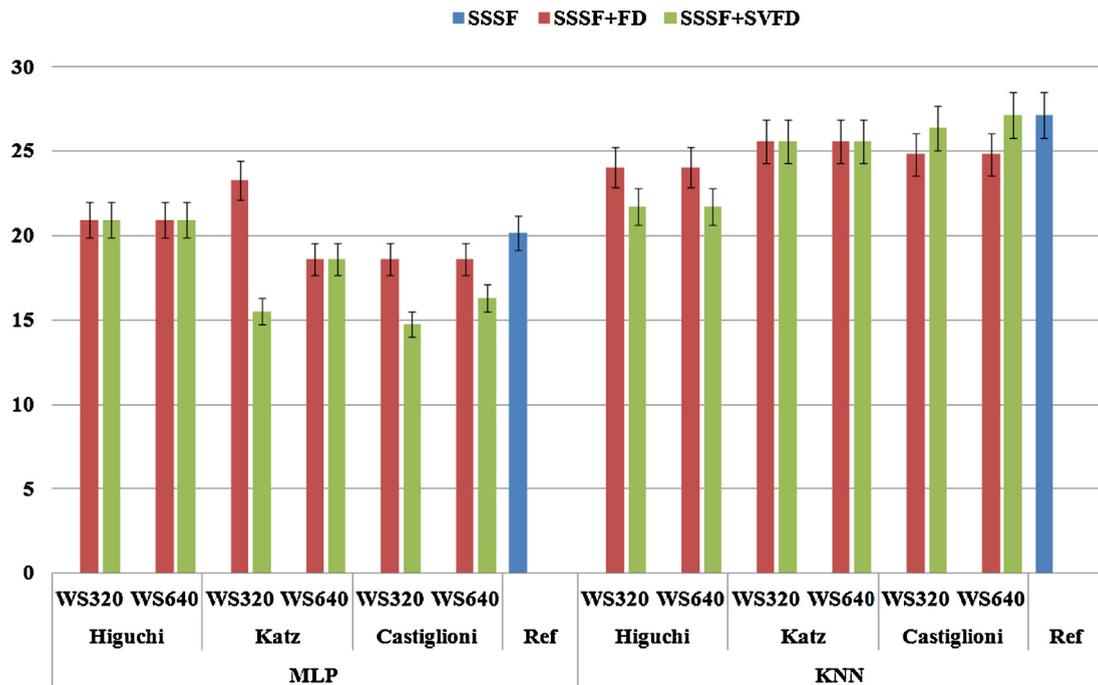


Fig. 7. Classification error rate in % for the three defined fractal dimension algorithms with only FD and with sets after attribute selection (SVFD) and for both paradigms MLP and KNN. Ref bars were obtained without using fractal dimension features.

regard to fractal dimension algorithms, modeling and window size. With MLP the CER in % decreased in all cases. Regarding KNN, the best global results were obtained for the Higuchi algorithm. The computational cost in all cases is significantly reduced with the KNN paradigm (see Fig. 10). Thus, the SSSF+SVFD sets, with features selected by automatic selection, were used in subsequent experiments. The global results of this study were satisfactory. The new fractal features decrease CER in MLP from 20.16% to 14.73% and in KNN from 27.14% to 21.71% (see Fig. 6). This suggests that the sets that include the fractal dimension, its detailed variations and automatic attribute selection, were able to properly model non-linear signal features. In both experiments the coverage of cases was about 94%. The confidence intervals are of about ± 3 for MLP and ± 5 for KNN over 80%. Fig. 7 shows the improvement (CER in %) in most cases when SVFD is included, mainly for MLP. The best results are achieved with MLP for Katz and Castiglioni with a window-size of 320 points. In general, KNN is less stable than MLP. Moreover, MLP presents good rates for computational cost (see Fig. 10). KNN has the lowest computational cost for all cases and with fractal dimension features is able to detect ES segments, but cannot achieve the results obtained by MLP. The best global results are produced by SSSF+CSVFD (14.73%) with MLP and WS = 320. KNN obtains the best results with Higuchi algorithm variants (21.71%).

Fig. 8 shows the results for classes (accuracy in %). SSSF+CSVFD with MLP and WS = 320 is again the best option. This set obtains the best results for all classes and also improves the classification with regard to early detection (ES class). In the case of IS, we also obtained a better rate of identifying the middle AD level. The model is also able to discriminate between pathological and non-pathological segments in each patient. Finally, regarding cumulative error for classes in % (CE), Fig. 9 shows the obtained results. The W320 option, which presents the lowest CE, seems to better capture the dynamics of the signal with the variants of Castiglioni fractal dimension. It should be noted that there is a small percentage of false positives, which may be due to doubts in utterance production. Coverage of cases is about 95% in the global system. The differences between ES and the other stages with regard to rates is due to the fact that the specific communication problems the patient encounters depend on the stage of the disease, as described in Section 1 (aphasia, anomia). Moreover, when patients are in the ES or AS stage, problems related to age – such as dysphonic voice – are often present as well, as aging causes alterations in the voice production system. Table 2 shows accuracy (%) in the final system with MLP and KNN for CR and AD and also confidence interval for 95% and 90%.

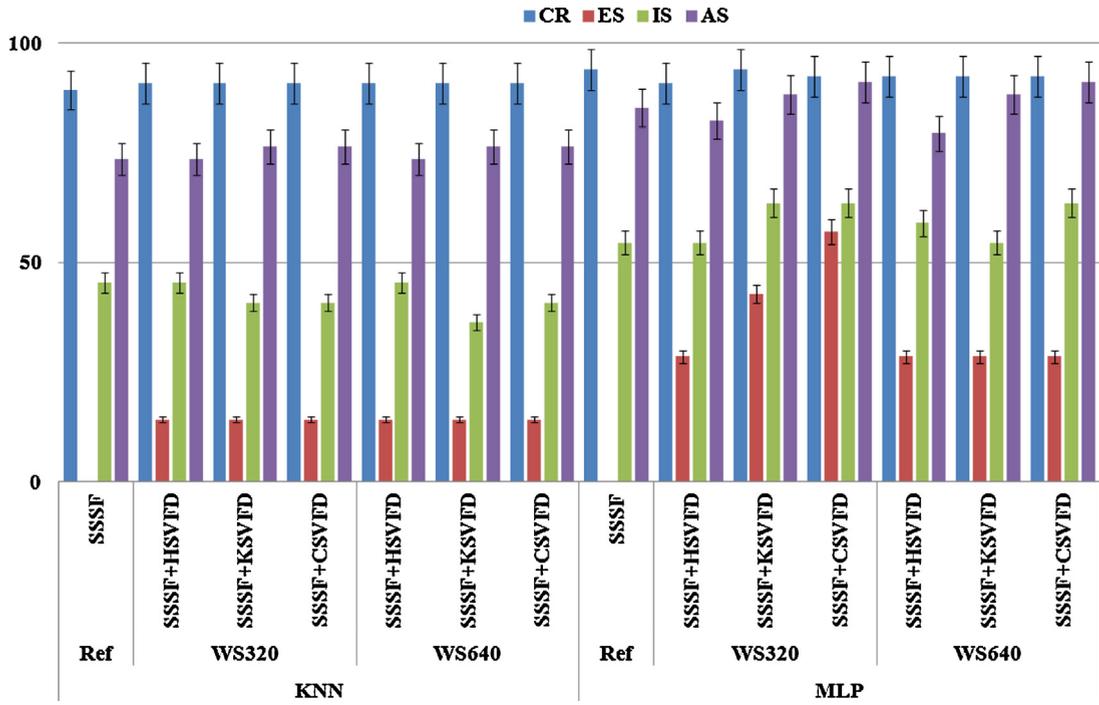


Fig. 8. Accuracy (%) with MLP and KNN for different classes: CR, ES, IS and AS. Ref bars were obtained without using fractal dimension features.

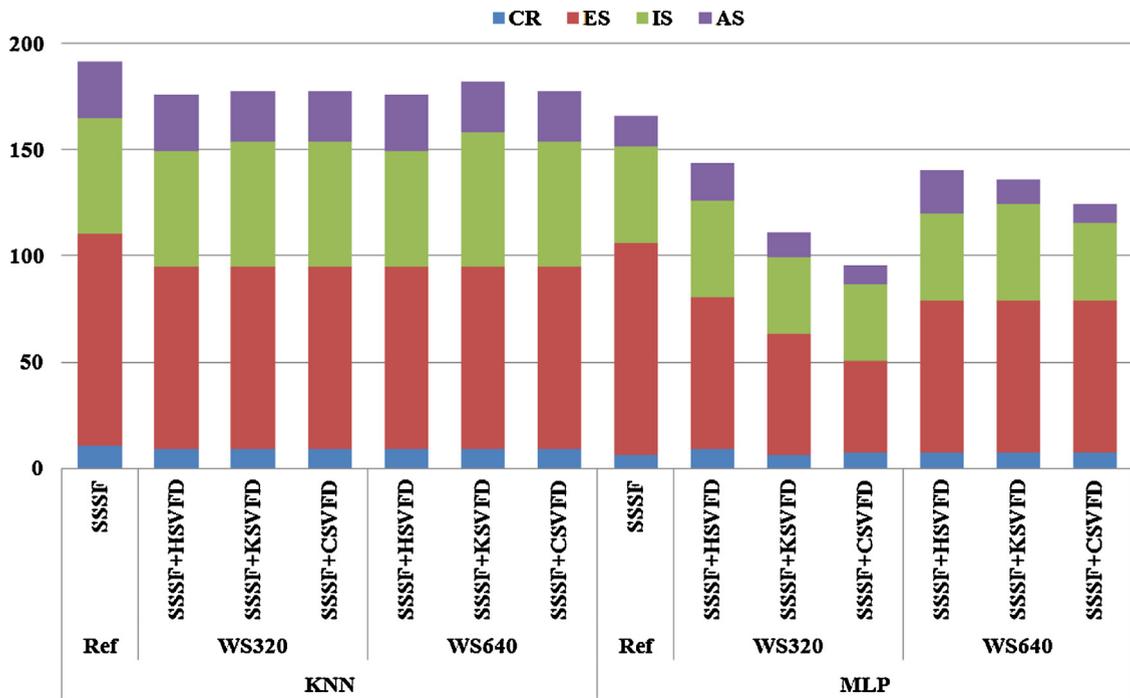


Fig. 9. Cumulative error (%) with MLP and KNN for different classes: CR, ES, IS and AS.

Table 3
Acronyms used.

Acronym	Description
AD	Alzheimer's disease
CR	Control
FD	Fractal dimension
KFD	Katz FD
HFD	Higuchi FD
CFD	Castiglioni FD
SSF	Spontaneous speech features
SSSF	Selection of SSF
SSSF + FD	SSSF set and HFD, KFD or CFD
STD	Standard deviation
VFD	Maximum, minimum, standard deviation, median and mode average for full and voiced signal
SVFD	Automatic selection of VFD
MLP	Multi layer perception
NNHL	Neuron number in hidden layer
TS	Training step
MLP501000	NNHL = 50 and TS = 1000
MLPA500	NNHL = $\max(\text{attribute/number} + \text{classes/number})$ and TS = NNHL \times 10
KNN	k -Nearest neighbors
CER	Classification error rate
Acc	Accuracy
CE	Cumulative error

Fig. 10 analyses the computational cost of the proposed algorithm: (a) with regard to Model Building and Recognition Process for: KNN, MLPA500 and MLP501000 (MLP with NNHL = 500 and TS = 1000); and (b) with regard to Model Building and Recognition Process for: KNN, and MLPA500 with different features set.

We can observe (Fig. 10a) that when building the model, KNN and MLPA500 have similar computational cost, followed by MLP501000; but in classification tasks, clearly KNN is the best one regarding this aspect. What is more interesting is to observe in detail Fig. 10b, where we present also computational cost for building the model and for classification task, but in this case only for KNN and MLPA500, and with different sets of features. Now, even if still KNN outperforms MLP, automatic selection of parameters allows us to clearly diminish the computational time for MLPA500 (about 4 times less than MLP501000), which in turns is the method that gives better performance in terms of Classification Error Rate (see Figs. 6–8). Therefore, MLP with automatic selection of parameters is the best option because it obtains the lowest CER and a reasonable computational time for the recognition process (less than 10 s, which is suitable for real time applications).

Computational cost with regard to Model Building and Recognition Process for: KNN and MLPA500 with different feature sets

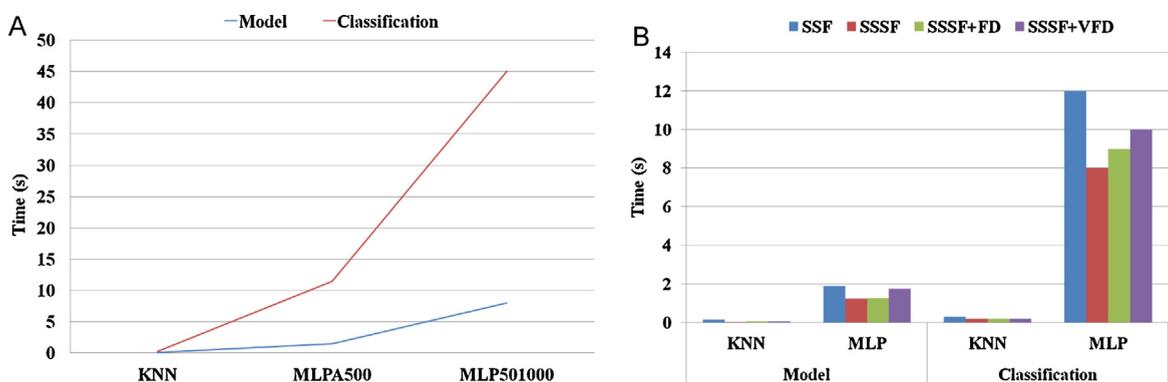


Fig. 10. (a) Computational cost with regard to model building and recognition process for: KNN, MLPA500 and MLP501000 (MLP with NNHL = 500 and TS = 1000), (b).

5. Conclusions and future work

The main goal of the present project is feature search in spontaneous speech to aid in the pre-clinical evaluation and the selection of appropriate tests for AD diagnosis. These features are of great relevance for health specialists to identify AD sufferers and differentiate between the three AD levels. The approach of this work is to improve previous modeling based on spontaneous speech features with fractal dimensions. With regard to previous work, new modeling applications with small computational cost have been evaluated. More precisely, we propose to use Higuchi's, Katz's, and Castiglioni's algorithms in order to add these new features to the set that feeds the training process of the model. In this work, an approach to the inclusion of nonlinear features is described. This straightforward approach appears robust in terms of capturing the dynamics of the whole waveform, as the CER decreases when using FD features, with less computational cost than the previously used ones. In future work we will introduce new features related to speech modeling that can potentially be applied to standard medical tests for AD diagnosis and to emotion response analysis. We will also model the fractal dimension using other algorithms, and entropy features as new nonlinear features. Finally, a new approach based on one-class classifier designed for early detection will be also developed.

Acknowledgments

This work has been partially supported by a SAIOTEK from the Basque Government, University of Vic – Central University of Catalonia under the research grant R0947, and the Spanish Ministerio de Ciencia e Innovación TEC2012-38630-C04-03.

References

- AA, 2014. Alzheimer's Association, Available from: <http://www.alz.org/>
- American Psychiatric Association (APA), 2000. *Diagnostic and Statistical Manual of Mental disorders*, Washington DC, 4th ed., Text Revision.
- Abásolo, D., Escudero, J., Hornero, R., Espino, P., Gómez, C., 2008. Fractal dimension of the EEG in Alzheimer's disease. *En cycl. Healthcare Inform. Syst.*
- Arias-Londoño, J.D., Godino-Llorente, J.I., Sáenz-Lechón, N., Osma-Ruiz, V., Castellanos-Domínguez, G., 2010. An improved method for voice pathology detection by means of a HMM-based feature space transformation. *Pattern Recogn.* 43 (9), 3100–3112, <http://dx.doi.org/10.1016/j.patcog.2010.03.019>.
- Castiglioni, P., 2010. What is wrong in Katz's method? Comments on: "A note on fractal dimensions of biomedical waveforms". *Comput. Biol. Med.* 40, 11–12.
- Chen, X., Zhao, H., 2006. Fractal characteristic-based endpoint detection for whispered speech. In: *Proceedings of the 6th WSEAS International Conference on Signal, Speech and Image Processing*, Lisbon, Portugal.
- Chouard, C.H., Pean, V., Ouayoun, M., Meyer, B., 2001. A fractal approach to the features of speech consonants. *Acta Otolaryngol.* 121 (January (2)), 249–253.
- Croota, K., Hodgesa, J.R., Xuerebc, J., Pattersona, K., 2000. Phonological and articulatory impairment in Alzheimer's disease: a case series. *Brain Lang.* 75 (2), 277–309.
- Esteller, R., Vachtsevanos, G., Echaux, J., Litt, B., 2001. A comparison of waveform fractal dimension algorithms. *IEEE Trans. Circuits Syst. I: Fundam. Theory Appl.* 48 (2), 177–183.
- Ezeiza, A., López de Ipiña, K., Hernández, C., Barroso, N., 2013. Enhancing the feature extraction process for automatic speech recognition with fractal dimensions. *Cogn. Comput.* 5, 545–550, <http://dx.doi.org/10.1007/s12559-12-91-65.29120>.
- Faundez-Zanuy, M., et al., 2012. Biometric applications related to human beings: there is life beyond security. *Cogn. Comput.*, <http://dx.doi.org/10.1007/s12559-012-91699>.
- Godino, J.I., Gómez-Vilda, P., Blanco-Velasco, M., 2006. Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short term cepstral parameters. *IEEE Trans. Biomed. Eng.* 53 (10), 1943–1953.
- Godino, J.I., Gómez-Vilda, P., Sáenz-Lechón, N., Blanco-Velasco, M., Cruz-Roldán, F., Ferrer-Ballester, M.A., 2005. *Support Vector Machines Applied to the Detection of Voice Disorders*. Springer Verlag, Berlin Heidelberg, pp. 219–230.
- Gómez, C., Mediavilla, A., Hornero, R., Abásolo, D., Fernández, A., 2009. Use of the Higuchi's fractal dimension for the analysis of MEG recordings from Alzheimer's disease patients. *Med. Eng. Phys.* 31 (3), 306–313, ISSN 1350-4533.
- Higuchi, T., 1988. Approach to an irregular time series on the basis of the fractal theory. *Phys. D* 31277, 283.
- Jang, J.S.R., 2011. Audio Signal Processing and Recognition. In: Roger Jang's Homepage, Available from: <http://www.cs.nthu.edu.tw/~jang> (accessed 2011).
- Katz, M., 1988. Fractals and the analysis of waveforms. *Comp. Biol. Med.* 18 (3), 145–156.
- Langi, A., Kinsner, W., 1995. Consonant characterization using correlation fractal dimension for speech recognition. In: *Proceedings of Communications, Power, and Computing. Conference Proceedings. IEEE*.
- Li, Y., Fan, Y., Tong, Q., 2007. Endpoint detection in noisy environment using complexity measure. In: *Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition*, Beijing, China.

- López-de-Ipiña, K., Alonso, J.B., Solé-Casals, J., Barroso, N., Faundez-Zanuy, M., Travieso, C., Ecay-Torres, M., Martínez-Lage, P., Egiraun, H., 2013b. On automatic diagnosis of Alzheimer's Disease based on Spontaneous, Speech Analysis and Emotional Temperature, Cognitive Computation. Springer, Berlin, <http://dx.doi.org/10.1007/s12559-013-9229-9>.
- López-de-Ipiña, K., Egiraun, H., Solé-Casals, J., Ecay, M., Ezeiza, A., Barroso, N., Martínez-Lage, P., Martínez-de-Lizardui, U., 2013c. Feature extraction approach based on fractal dimension for spontaneous speech modelling oriented to Alzheimer disease diagnosis. In: NOLISP, 2013, pp. 144–151.
- López-de-Ipiña, K., Alonso, J.B., Travieso, C.M., Solé-Casals, J., Egiraun, H., Faundez-Zanuy, M., Ezeiza, A., Barroso, N., Ecay, M., Martínez-Lage, P., Martínez-de-Lizardui, U., 2013a. On the selection of non-invasive methods based on speech analysis oriented to automatic Alzheimer disease diagnosis. *Sensors* 13 (5), 6730–6745.
- Costa, M., Goldberger, A.L., Peng, C.K., 2005. Multiscale entropy analysis of biological signals. *Phys. Rev. E* 71 (2 pt. 1).
- Maragos, P., Potamianos, A., 1999. Fractal dimensions of speech sounds: Computation and application to automatic speech recognition. *J. Acoust. Soc. Am.* 105 (3), 1925–1932.
- Martínez, F., Guillaumon, A., Martínez, J.J., 2003. Vowel and consonant characterization using fractal dimension in natural speech. In: Proceedings of NOLISP'2003.
- McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M., 1984. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Workgroup on Alzheimer's Disease, 24, pp. 939–944.
- McKhann, G.M., et al., 2011. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dement.* 7 (3), 263–269.
- Morris, J.C., 1993. The clinical dementia rating (CDR): current version and scoring rules. *Neurology* 43, 2412b–2414b.
- Nelwamondo, F.V., Mahola, U., Marwola, T., 2006. Multi-scale fractal dimension for speaker identification systems. In: Proceedings of the 8th WSEAS International Conference on Automatic Control, Modeling & Simulation, pp. 81–86.
- Ouayoun, M., Péan, V., Meyer, B., Chouard, C.H., 1999 Mar. A study of speech fractal dimensions. *Acta Otolaryngol.* 119 (2), 261–266.
- Péan, V., Ouayoun, M., Fugain, C., Meyer, B., Chouard, C.H., 2000. A fractal approach to normal and pathological voices. *Acta Otolaryngol.* 120 (2), 222–224.
- Péan, V.M., Ouayoun, C., Fugain, B., Meyer, C.H., 2002. Chouard a fractal approach of pathological voices, fractals in biology and medicine. *Math. Biosci. Interact.*, 147–151, http://dx.doi.org/10.1007/978-3-0348-8119-7_16.
- Picard, R., Cook, D., 1984. Cross-validation of regression models. *J. Am. Stat. Assoc.* 79 (387), 575–583, <http://dx.doi.org/10.2307/2288403> (JSTOR 2288403).
- Pickover, C.A., Khorasani, A., 1986. Fractal characterization of speech waveform graphs. *Comp. Graph.* 10 (1), 51–61.
- Pitsikalis, V., Maragos, P., 2009. Analysis and classification of speech signals by generalized fractal dimension features. *Speech Commun.* 51 (12), 1206–1223.
- Praat. Doing Phonetics by Computer. Available from: <http://www.fon.hum.uva.nl/praat>
- Solé, J., Zaiats, V., 2003. A non-linear VAD for noisy environment. *Cogn. Comput.* 2 (3), 191–198.
- Tsonis, A., 2007. Reconstructing dynamics from observables: the issue of the delay parameter revisited. *Int. J. Bifurc. Chaos* 17, 4229, <http://dx.doi.org/10.1142/S0218127407019913>.
- VAD, Voice Activity Detector algorithm. Available from: <http://www.mathwork.com>
- Van de Pole, L.A., et al., 2005. The effects of age and Alzheimer's disease on hippocampal volumes, a MRI study. *Alzheimer's Dement.* 1 (1 Supp. 1), 51.
- WEKA. Available from: <http://www.cs.waikato.ac.nz/ml/weka/>



Karmele López-de-Ipiña received the Ph.D. degree in Computer Science in 2003, and a Master degree in Electronics and Automation and the B.Sc. degree in Physics in 1990, at the Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU). She worked for enterprises until 1995 when joined the Department of Systems Engineering and Automation of the University of the Basque Country. He was Director at the UPV/EHU (2004–2009). She is currently head of the Engineering and Society research group, ELEKIN. Her research interests are in Bioengineering and Biomedical Engineering, Pattern Recognition, Signal Processing, Ambient Intelligence and Robotics.



Jordi Solé-Casals received the Ph.D. degree with European label in 2000, and the B.Sc. degree in Telecommunications in 1995, both from the Polytechnic University of Catalonia (UPC), Barcelona. In 1994 he joined the Department of Digital Technologies and Information of the University of Vic, where he was the Director (2010–2012). He is currently Head of the Data and Signal Processing Research Group. He is visiting scientist with the GIPSA-Lab (Grenoble, France), LABSP RIKEN (Tokyo, Japan) and BMU (Cambridge, UK). His research interests are in biomedical signal processing (EEG, fMRI, speech, biometric applications), neural networks, source separation and independent component analysis.



Harkaitz Eguiraun received a B.Sc. degree in Industrial Technical Engineer, specialty in Industrial Electronics in 2004, a M.Sc. degree in Industrial Automation and Electronics Engineering in 2011 and a Master degree in Control Engineering, Automation and Robotics in 2012, all of them by the University of the Basque Country UPV/EHU, Spain. He has been working in corporate world since 2004 until 2013 when he joined the PhD Program Marine Environment and Resources at the Research Center for Experimental Marine Biology & Biotechnology of the University of the Basque Country UPV/EHU, Spain. His fields of expertise, and research interests, include Systems Engineering, Systems Biology, Modeling, Control Strategies and Non-linear Analysis.



Jesús B. Alonso-Hernández received the Telecommunication Engineer degree in 2001 and the Ph.D. degree in 2006 from University of Las Palmas de Gran Canaria (ULPGC-Spain) where he is an Associate Professor in the Department of Signal and Communications from 2002. He has researched in different International and Spanish Research Projects. He has numerous papers published in international journals and conferences. He has been reviewer in different international journals and conferences since 2003. His research interests include signal processing in biocomputing, biometrics, nonlinear processing, recognition systems, audio characterization and data mining. He was a Guest Editor of ISI-JCR Special Issues.



Carlos M. Travieso-González received the M.Sc. degree in 1997 in Telecommunication Engineering at Polytechnic University of Catalonia (UPC), Spain; and Ph.D. degree in 2002 at University of Las Palmas de Gran Canaria (ULPGC-Spain). His research lines are biometrics, biomedical signals, data mining, classification, signal and image processing, and environmental intelligence. He has researched in International and Spanish Research Projects, some of them as head researcher. He is co-author books, book chapters, co-editor of Proceedings Books, and Guest Editor ISI-JCR. He is reviewer in different international JCR journals and conferences since 2001. He was General Chair in different international conferences.



Aitzol Ezeiza received the Ph.D. degree in Robotics and Automation in 2012 and the B.Sc. degree in Computer Science in 2000, at the Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU). He joined the Department of Systems Engineering and Automation of the University of the Basque Country in 2005. He is currently Member of the Engineering and Society research group. His research interests are in Bioengineering and Biomedical Engineering, Pattern Recognition, Signal Processing and Robotics.



Nora Barroso received the Ph.D. degree in Robotics and Automation in 2011 and the B.Sc. degree in Telecommunications in 2005, at the Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU) and the Universidad Pública de Navarra. She worked for enterprises until 2011 when joined the Department of Systems Engineering and Automation of the University of the Basque Country. She is currently Member of the Engineering and Society research group. Her research interests are in Bioengineering and Biomedical Engineering, Pattern Recognition, Signal Processing and Robotics.



Miriam Ecay-Torres obtained his psychology degree in 2007 at the University of Deusto, a master degree in Neuropsychology in 2010 and a master degree in Psychology Research in 2013. She started the PhD Program in 2013. Her project is focused on the study of potential neuropsychology profiles in the preclinical stages of Alzheimer's disease.



Pablo Martínez-Lage obtained his medical degree in 1988 at the Medical School of the University of Navarra and then became a clinical neurologist. After his PhD in 1994 at the University of Navarra he attended a Clinical program in Cerebrovascular Diseases and Dementia at the Department of Clinical Neurological Sciences of the University of Western Ontario (London, Ontario, Canada) until 1996. From 2010 he is director of the Neurology area at the Center for Research and Advanced Therapies of the Fundacion CITA-alzheimer Fundazioa in San Sebastian. He is a member of the Spanish Society of Neurology and a founding member of the International Society on Vascular behavioral and cognitive disorders (Vas-Cog). He has been the Coordinator of the Dementia Study Group of the Spanish Society of Neurology (2009–2012) and is a member of the Dementia Panel of the European Federation of Neurological Societies since November 2008.



Blanca Beitia received the Ph B.Sc. degree in Science in 1982, at the Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU). She joined the Department of Mathematics of the University of the Basque Country in 1983. She is currently Member of the Engineering and Society research group. Her research interests are in Bioengineering and Biomedical Engineering, Pattern Recognition, Data Mining, Machine Learning and Statistics.