Feature selection for automatic analysis of emotional response based on non-linear speech modeling suitable for diagnosis of Alzheimer's disease

K. Lopez-de-Ipina[1], J.B. Alonso-Hernández[2], .J. Solé-Casals[3], C.M. Travieso-González[2], A. Ezeiza[1], M. Faúndez-Zanuy[4]

[1]System Engineering and Automation Department, University of the Basque Country, Donostia 20008, Spain, {karmele.ipina, aitzol.ezeiza@ehu.es}@ehu.es

[2]Universidad de Las Palmas de Gran Canaria, IDeTIC, {jalonso,ctravieso}@dsc.ulpgc.es, Spain

[3]Digital Technologies Group. University of Vic, jordi.sole@uvic.cat, Spain

[4]Escola Universitaria Politècnica de Mataró (UPC), Tecnocampus , faundez@tecnocampus.cat, Spain

Abstract—Alzheimer's disease (AD) is the most common type of dementia among the elderly. This work is part of a larger study that aims to identify novel technologies and biomarkers or features for the early detection of AD and its degree of severity. The diagnosis is made by analyzing several biomarkers and conducting a variety of tests, but nowadays only a post-mortem examination of the patients' brain tissue is considered to provide definitive confirmation. Thus, non-invasive intelligent diagnosis techniques would be very valuable for this purpose. Specifically this paper deals with Automatic Analysis of Emotional Response (AAER) in spontaneous speech based on classical and new emotional speech features: Emotional Temperature (ET) and Fractal Dimension (FD). This is a pre-clinical study aiming to validate future diagnosis tests and biomarkers. The method has the great advantages of being non-invasive, low cost, and without any side effects. The AAER showed very promising results for the definition of features useful in early diagnosis of AD.

1. INTRODUCTION

Alzheimer's disease (AD) is the most common type of dementia among the elderly, and its socioeconomic cost to society is sizeable and expected to increase. It is characterized by progressive and irreversible cognitive deterioration with memory loss, impaired judgment and language, and other cognitive deficits and behavioral symptoms that end up becoming severe enough to limit the ability of an individual to perform professional, social or family activities of daily life. As the disease progresses, patients develop increasingly severe disabilities, becoming in the end completely dependent.

An early and accurate diagnosis of AD would be of much help to patients and their families, both so they could more easily plan for the future and to start an early treatment of 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, but an unambiguous diagnosis of AD is considered to require that a post-mortem analysis demonstrate the typical AD pathological changes in brain tissue [1,2,3,4]. Episodic memory impairment is the clinical hallmark of the earliest manifestations of AD. At the time of clinical presentation, other cognitive deficits are usually already present in the patient's language, executive functions, orientation, perceptual abilities and constructional skills. Associated behavioral and psychological symptoms include apathy, irritability, depression, anxiety, delusions, hallucinations, inhibition decrease, aggression, aberrant motor behavior, as well as changing in eating or sleep patterns [4,5,6]. All these symptoms lead to impaired performance in family, social or professional activities of daily life, as the disease progresses from mild to moderate to severe. The diagnosis of AD is made through diagnostic tests and requires, on the one hand, the confirmation of a progressive dementia syndrome and, on the other, the exclusion of other potential causes of dementia as demonstrated by the patient's clinical history, and in some cases expensive and invasive diagnostic tests such as computer tomography (CT), magnetic resonance imaging (MRI) and/or lumbar puncture. The development of non-invasive intelligent diagnosis techniques would therefore be very valuable for the early detection and classification of different types of dementia, particularly because they do not require specialized personnel or laboratory equipment, so that anyone in the habitual environment of the patient could, after proper training, apply such techniques without altering or blocking the patient's abilities [7,8].

Emotional Speech Analysis (ESA) has that potential. Emotions are cognitive processes related to the architecture of the human mind, such as decision-making, memory or attention; they are closely linked to the learning and understanding that arise in intelligent natural or artificial systems when such learning becomes necessary for survival in a changing and partially unpredictable world [9-14]. Human interaction includes emotional information about partners that is transmitted through language explicitly and implicitly through nonverbal communication [15-17].

Figure 1. Speech signal and spectrogram for a control (top) and a person with AD (bottom)

Nonverbal information, which often includes body-language, attitudes, modulations of voice, facial expressions, etc., is essential in human communication: it has a significant impact on the communication provision of the partners and on the intelligibility of speech [15-17]. Human emotions are affected by the environment and by direct interaction with the outside world, but also by the emotional memory that emerges from the experience of the individual and cultural environment, the so-called socialized emotion. Emotions use the same components – subjective, cultural, physiological, and behavioral – that the individual's perception expresses with regard to the mental state, the body, and its interaction with the environment [15-17].

Our work focuses on non-invasive diagnostic techniques based on the analysis of emotional response to speech (Figure 1) [9,18,37]. The emotional response in Alzheimer's patients becomes impaired, and such impairment appears to go through different stages. In the early stages, social and even inhibition decreases and behavioral changes are also observed (for example, anger and inability to perform common tasks, express oneself or remember) [8,18]. However, the emotional memory remains, and Alzheimer's sufferers often cry more easily and gratefully acknowledge caresses, smiles and hugs. The Alzheimer's patient reacts aggressively to things that for healthy people are harmless and perceives threats or dangers where none exist. In more advanced stages, the person affected with Alzheimer's may often seem shy and apathetic – symptoms often attributed to memory loss and/or difficulty in finding the right words. Some of these responses are likely to be magnified due to an alteration in perception. Alternatively, it has been suggested that the reduced ability to feel emotions is due to memory loss, which may in turn induce the appearance of apathy and depression [19-23].

The work presented here is part of a larger study that aims to identify novel technologies and biomarkers or features for the early detection of AD. The purpose of this work is to evaluate the suitability of a new approach in Automatic Analysis of Emotional Response (AAER) for early AD diagnosis based on the analysis of classical parameters, Emotional Temperature and non-linear parameters, whose results are suitable for use in the automatic classification of tested individuals.

This paper is organized as follows: Section 2 explains the materials and the methodology used, Section 3 presents the experimental results, Section 4 offers discussion of the results, and Section 5 lays out the conclusions.

2. MATERIALS AND METHODS

2.1. Materials

The research presented here is in the nature of a preliminary experiment; its aim is to define thresholds for a number of biomarkers related to spontaneous speech. It forms part of a broader study focused on early AD detection. Feature search in this work aims at pre-clinical evaluation so as to formulate useful tests for AD diagnosis [8,18].

In an effort to develop a new methodology applicable to a wide range of individuals of different sex, age, language and cultural and social background, we have built a multicultural and multilingual (English, French, Spanish, Catalan, Basque, Chinese, Arabian and Portuguese) database with video recordings of 50 healthy subjects and 20 AD patients (with a prior diagnosis of Alzheimer) recorded for 12 and 8 hours, 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 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; in such cases, we always 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-analyzable 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 minutes of material for the AD group and about 9 hours for the control. The speech was next divided into consecutive segments of 60 seconds in order to obtain appropriate segments for all speakers, resulting finally in a database of about 600 segments of Spontaneous Speech. Finally, for experimentation from the original database, 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) representing a wide range of speech responses. This subset of the database is called AZTIAHORE.

2.2. Methods

2.2.1 Feature extraction

2.2.1.1 Emotional Speech Analysis

In this study, we aim at being able to automatically distinguish emotional speech by an Alzheimer's patient from that of a healthy subject by analyzing three families of features in speech [24]:

1. Acoustic features: pitch, standard deviation pitch, max and min pitch, intensity, standard deviation intensity, max and min intensity, period mean, period standard deviation, and Root Mean Square amplitude (RMS).
2. Voice quality features: shimmer, local jitter, Noise-to-Harmonics Ratio (NHR), Harmonics-to-Noise Ratio (HNR) and autocorrelation.
3. Duration features: fraction of locally unvoiced frames, degree of voice breaks.

These three feature families compose the Emotional Feature (EF) set in the experiment.

2.2.1.2 Emotional Temperature

Our aim was to apply a non-invasive method to estimating the severity of Alzheimer's in a patient. The method discussed here proposes a strategy based on a few prosodic and paralinguistic feature sets obtained from a temporal segmentation of the speech signal. The speech signal $\{s(n)\}$ is windowed by a hamming window of 0.5 seconds overlapped 50% [8]. In each frame $\{x(n)\}$ the DC component is removed and a z-normalization of

the frame is made. From each frame 2 prosodic features and 4 paralinguistic features related to pitch and energy, respectively, are estimated. These features were chosen for several reasons: first, they are quickly and easily calculated; second, their robustness in emotion recognition has been proven; and, finally, they are independent of linguistic segmentation, which means that problems in real-time applications in real environments can be avoided. Finally the "Emotional Temperature" (ET) is calculated as follows [8,24,25]:

- For prosodic features, a voiced/voiceless decision is made for each frame (with the help of a voice activity detector [26]) and two linear regression coefficients of the pitch contour are obtained.
- For paralinguistic features, voice spectral energy balances are calculated from each frame, quantified using 4 percentages of energy concentration in 4 frequency bands.
- The percentage of temporal frames classified as "non-pathological" is calculated by a SVM [27] with a radial basis kernel function. This value, i.e., the number of non-pathological frames, is the "Emotional Temperature".
- The "Emotional Temperature" is finally normalized in order to have ET=50 as the threshold indicating the limit between pathological and non-pathological frames. This normalization will make it substantially easier for medical specialists to interpret the data.

2.2.1.3 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, a measure called the Fractal Dimension. It should be noted that the Fractal Dimension of natural phenomena is only measurable using statistical approaches. Consequently, there exists no precise reference of the Fractal Dimension value that a given waveform should have. In addition, speech waveforms are not stationary, so most Automatic Speech Processing (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 events in the speech signal 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 present work we focus on the alternatives that are specially suited for time series analysis and that do not require the previous modeling of the system. Three of these algorithms are Higuchi [28], Katz [29] and Castiglioni [30], named from their authors. Higuchi and Castiglioni were chosen because they have been reported to be more accurate in previous works on under-resourced conditions [31]. Katz is also reported as a robust algorithm to calculate the Fractal Dimension.

Higuchi [28] proposed an algorithm for measuring the Fractal Dimension of discrete time sequences directly from the time series x(1),x(2),…,x(n). The algorithm calculates the length *Lm(k)* for each value of m and k covering all the series,

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

where $L(k)$ is the sum of all the lengths *Lm(k)* for each k, determined as,

$$L(k) = \sum_{m=1}^{k} L_m(k) \quad (2)$$

Thus 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).

Katz [28] has proposed a normalized formula of the Fractal Dimension (KFD):

$$FD = \frac{\log(n)}{\log(n) + \log(\frac{d}{L})} \quad (3)$$

In this case L is the length and d is the extension of the curve, and they are normalized using the average step a=L/n,

$$L = \sum_{i=1}^{n} l_{i,i+1} \qquad (4)$$

$$d = \max\left\{l_{i,j}\right\} \qquad (5)$$

where l is the Euclidean distance between two points,

$$l_{ij} = \sqrt{(y_i - y_j)^2 + (x_i - x_j)^2} \qquad (6)$$

Castiglioni [30], however, 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 (x). Therefore, given that the input signal is a mono-dimensional waveform, the length and the extension can be rewritten using Mandelbrot's approach. A simple and efficient way to do this is to measure these two magnitudes directly in their own dimension. Therefore the extension on the Y-axis is the range of yk

$$d = \max\left\{y_k\right\} - \min\left\{y_k\right\} \qquad (7)$$

and the length L is the sum of all the increments in modulus, (Castiglioni Fractal Dimension, CFD):

$$L = \sum_{k=1}^{n} | y_{k+1} - y_k | \qquad (8)$$

2.2.1.4 Window Size during the Feature Extraction process

The selection of an appropriate window size to be used during the experiments is essential. Figure 2 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 detail is 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 [32-34] suggest that a bigger window could be useful in some cases. Consequently, four window-sizes of 160, 320, 640 and 1280 points will be analyzed [31].

Figure 2. Fractal Dimension (HFD) for a control subject (up) and an AD subject (down), for different window sizes.

2.2.2 Feature sets

In the experimentation four feature sets will be used:

1. EF set described in 2.2.1.1
2. EF+VFD: EF set, Fractal Dimension (HFD, KFD and CFD) and variations of Fractal Dimension (maximum, minimum, variance, standard deviation, median and mode) for full signal and voiced signal: HVFD, KVFD, CVFD, .
3. EF+VFD+ET: previous set and Emotional Temperature

2.2.3 Automatic Classification

The main goal of the present work is feature search in Spontaneous Speech aiming at pre-clinical evaluation in order to define tests for AD diagnosis. These features will define the control group (CR) and the three AD levels (ES, IS and AS). A secondary goal is the optimization of computational cost with the aim of making these techniques useful for real-time applications in real environments. Thus automatic classification will be modeled with this in mind. Two different paradigms will be evaluated:

1. Multi Layer Perceptron (MLP) with Neuron Number in Hidden Layer (NNHL) = max(Attribute Number, Classes Number) and Training Step (TS) NNHL*10.
2. k-Nearest Neighbors (k-NN) paradigm.

WEKA software [35] has been used in carrying out the experiments. The results were evaluated using Classification Error Rate (CER), Accuracy (Acc) and Accumulative Error Rate (AER). For the training and validation steps, we used k-fold cross-validation with k=10. Cross validation is a robust validation for variable selection [36].

3.   Experimental results

In the next stage the methodology described in 2.2 will be used. The task was Automatic Classification, with the classification targets being healthy speakers without neurological pathologies and speakers diagnosed with AD. The experimentation is carried out with AZTIAHORE dataset. Four kinds of experiments have been carried out: selection of window-size; feature analysis; analysis of global results; analysis of classes' results. In this first stage references (baselines) for Automatic Analysis of Emotional Response are obtained for k-NN and MLP paradigms. Specifically CER in %, without FD, are: MLP 14.73% and k-NN 13.96%; these will be the reference values from now on. Recognition rates for both algorithms are optimum for the CR and AS groups and poor for ES.

3.1 Selection of window-size

In this task several window sizes have been analyzed over Higuchi algorithm [24]: 160, 320, 640 and 1280 points. During this stage, EF and HVFD features sets have been used.

Figure 3 shows the Classification Error Rate in % for all the classes (CR, ES, IS and AS), with HVFD, several window sizes and both paradigms (MLP and k-NN). Most of the new proposals, which include Fractal Dimension, show improvement over the reference ones. The minimum CER is achieved with MLP and windows of middle size that are able to capture local and global signal evolution (320 and 640). Both window-sizes will therefore be used in the next experiments.

Figure 3 Classification Error Rate (%) for EF and Higuchi Fractal Dimension (HVFD) algorithms, several window sizes and both paradigms: MLP and k-NN.

3.2   Fractal Dimension Analysis

In this stage, Fractal Dimension features, paradigm modeling and window size are analyzed. With regard to global results with MLP, the CER decreases in all cases. On the other hand, Figure 4 shows CER (%) in classes' results, with middle window-size (320 and 640) and both paradigms (MLP and k-NN), with Fractal Dimension Feature sets (VFD) for the three algorithms (HVFD, KVFD and CVFD). Regarding k-NN, the best global results are obtained for Higuchi algorithm; performance is worst with the

Katz algorithm. The computational cost in all cases is significantly reduced with that paradigm. The best results are obtained for MLP algorithm, but oriented to early diagnosis the option of a window size of 640 points and HFD and CFD outperform the others. In the case of the k-NN paradigm these are also the best options. Therefore, a window-size of 640 will be used from now on.

Figure 4 Classification Error Rate (%) in classes for middle window-size (320 and 640) and both paradigms (MLP and k-NN) with Fractal Dimension Feature sets (VFD) for the three options HVFD, KVFD and CVFD.

3.3   Emotional Temperature

Figure 5 shows the improvement (CER in %) in most cases when ET is included, mainly for MLP. MLP achieves similar global results for all algorithms (CR=3.11%). In general, k-NN is less stable than MLP and obtains the best global results for Higuchi (CR=3.18%) and Katz (CR=4.66%). Moreover k-NN has the lowest computational cost for all cases and with Fractal Dimension features is able to detect properly ES segments, but cannot achieve the results obtained by MLP. k-NN obtains the best results with EF+HVFD+ET. On the other hand MLP with the selected configuration for real time applications presents good rates for computational cost. The best one in global result is EF+CVFD+ET, with MLP (CER=3.11%).

Figure 5 Classification Error Rate (%) for the three defined Fractal Dimension algorithm, both paradigms (MLP and k-NN) for Fractal Dimension Feature sets (VFD) and Emotional Temperature.

Finally, regarding Accumulative Error Rate (AER) for classes in %, Figure 6 shows the results obtained. The MLP paradigm outperforms k-NN but not its computational cost. In the case of k-NN, EF+HVFD+ET is the best configuration. However, when the MLP paradigm is used, EF+KVFD+ET and EF+CVFD+ET show improvement over EF+HVFD+ET. These last sets obtain the best results for all classes and the classification with regard to early detection is optimum (ES class, this method distinguishes mid-level AD patients more effectively. The model is also able to discriminate pathological and non-pathological segments of speech for each patient.

Figure 6. Accumulative Error Rate (AER) for classes (%) for both paradigms and window size of 640 points.

The option with a window size of 640 points presents the lowest AEC and seems to better capture the dynamics of the signal with the variants of Katz and Castiglioni Fractal Dimension being Accuracy in % about 96.89.

4. Discussion

Non-linear modeling is appropriate for detecting subtle changes oriented to early detection of AD. The inclusion of Fractal Dimension features adds relevant information with regard to non-linearity in the speech signal and aids in properly analyzing the Emotional Response. This looks useful for modeling subtle differences in order to detect changes in the Emotional Response aiming at early diagnosis.

Emotional Temperature appears to be a good reference for measuring the Emotional Response in AD. This feature provides several advantages: first, it is quickly and easily calculated; second, its robustness in emotion recognition has been proven; and, finally, it is independent of linguistic segmentation, which means that problems in real-time applications and in real environments can be avoided.

Moreover, oriented to real time applications, k-NN and MLP with the configuration proposed in 2.2.3 obtains appropriate Accuracy and computational cost. The use of new features such as ET introduces robustness in the system without increasing the computational cost.

It should be noted that there is a small percentage of false positives. This confusion can be due to doubts in utterance production. The analysis and classification of pathological and non-pathological segments also appear relevant for disease evolution monitoring.

Health specialists note the relevance of the system's ability to carry out both the analysis of independent biomarkers as spontaneous speech and/or the integral analysis of several biomarkers.

5. Conclusions and future work

The main goal of the present project is feature search in Emotional Response oriented to pre-clinical evaluation for the definition of test for AD diagnosis. These features are of

great relevance for health specialists attempting to define healthy subjects and to distinguish the three AD levels from one another. The approach adopted in this work is to use Fractal Dimensions and Emotional Temperature to improve previous modeling techniques based on Emotional Response features. Moreover, new modeling oriented to real-time applications has been evaluated. More precisely, we propose using 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. Our work also describes an approach to the inclusion of nonlinear features. This straightforward approach might be robust in terms of capturing the dynamics of the whole waveform, and offers many advantages in terms of computability. It also facilitates comparing the power of the new features against the previous ones. In future work, we will introduce new features related to speech modeling useful for standard medical tests for AD diagnosis and to emotion response analysis. We will also model Fractal Dimension using other algorithms and we will introduce new non-linear features. Finally, a new approach based on One-Class Classifier oriented to early detection will also be developed.

REFERENCES

[1] Mc Kahn G, et al.. Clinical diagnosis of AD: report of the NINCDS-ADRDA Workgroup on AD. (1984); 24:939-944.

[2] McKhann GM et al.. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the NIAA Association workgroups on diagnostic guidelines for AD. Alzheimers Dement. (2011) May;7(3):263-9.

[3] Van de Pole, L.A., et al., The effects of age and Alzheimer's disease on hippocampal volumes, a MRI study. Alzheimer's and Dementia, 2005. 1(1, Supplement 1): p. 51.

[4]   Prashanth J. & J. Huang, Alzheimer's disease: a review of diagnostic criteria, Meds 2012,2013

[5]   Morris JC, The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology, (1993). 43: p. 2412b-2414b.

[6]   American Psychiatric Association. Diagnostic and Statistical Manual of Mental disorders, 4th Edition Text Revision. Washington DC. (2000).

[7] Marcos Faundez-Zanuy, Amir Hussain, Jiri Mekyska, Enric Sesa-Nogueras, Enric Monte-Moreno, Anna Esposito, Mohamed Chetouani, Josep Garre-Olmo, Andrew Abel, Zdenek Smekal, Karmele Lopez-de-Ipiña, Biometric Applications Related to Human Beings: There Is Life beyond Security, Cognitive Computation, 2012, DOI 10.1007/s12559-012-9169-9

[8]   Lopez-de-Ipiña K., J.B. Alonso, C.M. Travieso, J. Solé-Casals , H. Egiraun, M. Faundez-Zanuy, A. Ezeiza, N. Barroso, M. Ecay , P. Martinez-Lage, and U. Martinez-de-Lizardui, 2013a. On the selection of non-invasive methods based on speech analysis oriented to Automatic Alzheimer Disease Diagnosis, Sensors, vol.: 13 (5) pp, 6730-6745

[9] Ekman, Paul, "Basic Emotions", in Dalgleish, T; Power, M, Handbook of Cognition and Emotion, Sussex, UK: John Wiley & Sons. 1999.

[10] H. Gardner, Multiple intelligences: The theory in practice, New York: Basic Books, 1993.

[11] Taylor, J. G., Scherer, K., & Cowie, R. (2005). Emotion and brain: Understanding emotions and modelling their recognition. Neural Networks, Vol 18. 313-316.

[12] Lang, P.J., Bradley, M., Emotion and the motivational brain. Biological Psycholog 84, 437–450. 2010.

[13] Fragopanagos, N., & Taylor, J. G. Emotion recognition in human–computer interaction. Neural Networks, Vol 18. 389-405, 2005.

[14] D. Sander et al. "A systems approach to appraisal mechanisms in emotion," Neural Networks, vol 18, issue 4, pp 317-352, 2005.

[15] R. W. Picard, Affective Computing, MIT Press, 1997.

[16]   Knapp, M.L. Essentials of nonverbal communication. Publisher: Holt, Rinehart & Winston, NY, USA, 1980.

[17]   Cowie, R.; Douglas-Cowie, E.; Tsapatsoulis, N.; Votsis, G.; Kollias, S.; Fellenz, W.; Taylor, J. G. Emotion Recognition in Human-Computer Interaction. IEEE Signal Processing Magazine 2001, 18(1), 32-80.

[18]   Lopez-de-Ipiña K, J.B. Alonso, J. Solé-Casals, N.Barroso, M. Faundez-Zanuy, C. Travieso, M. Ecay-Torres, P. Martinez-Lage, H. Egiraun,, 2013b On Automatic Diagnosis of Alzheimer's Disease based on Spontaneous, Speech Analysis and Emotional Temperature, Cognitive Computation, Springer, Berlin, 10.1007/s12559-013-9229-9

[19]   Shimokawa, A.; Yatomi, N.; Anamizu, S.; Torii, S.; Isono, H.; Sugai, Y.; Kohno, M. Influence of deteriorating ability of emotional comprehension on interpersonal behaviour in Alzheimer-type dementia. Brain and Cognition 2001, 47, 423-433.

[20]   Goodkind, M.S.; Gyurak, A.; McCarthy, M.; Miller, B.L.; Levenson, R.W. Emotion regulation deficits in frontotemporal lobar degeneration and Alzheimer's disease. Psychol Aging 2010, 25(1), 30-7. doi: 10.1037/a0018519.

[21]   Cadieux, N.; Greeve, K. Emotion processing in Alzheimer's disease. Journal of the International Neuropsychological Society 1997, 3, 411-419.

[22]   Horley, K.; Reid, A.: Burnham, D. Emotional prosody perception and production in Dementia of the Alzheimer's type. Journal of Speech, Language, and Hearing Research 2010, 53, 1132-1146.

[23]   Henry, J.D.; Rendell P.G.; Scicluna A.; Jackson M.; Phillips, L.H. Emotion experience, expression, and regulation in Alzheimer's disease. Psychology and Aging 2009, 24(1), 252-257.

[24] Lopez-de-Ipina, K., Alonso, J.B., Travieso, C.M., Egiraun, H., Ecay, M., Ezeiza, A., Barroso, N., Martinez-Lage, P. Automatic analysis of emotional response based on non-linear speech modeling oriented to Alzheimer disease diagnosis, 2013 IEEE 17th

International Conference on Intelligent Engineering Systems (INES). San Jose Costa Rica. 2013.

[25]   Alonso J., De León J., Alonso I., Ferrer MA. Automatic detection of pathologies in the voice by HOS base parameters. Journal on Applied Signal Processing. 2001; 4:275-284.

[26] J. Solé-Casals, V. Zaiats, "A Non-Linear VAD for Noisy Environments", Cognitive Computation, Vol 2 (3), pp. 191-198, DOI 10.1007/s12559-010-9037-4

[27]   Chang C.C.; Lin C.J. LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST) 2001. Available online: http://www.csie.ntu.edu.tw/~cjlin/libsvm

[28]   Higuchi T. Approach to an irregular time series on the basis of the fractal theory. Physica D (1988). 31277:283.

[29] Katz M., 1988. Fractals and the analysis of waveforms. Comput Biol Med . 18:3:145:156.

[30] Castiglioni P., 2010. What is wrong in Katz's method? Comments on: "A note on fractal dimen-sions of biomedical waveforms". Comput Biol Med, 40:11-12.

[31] Ezeiza A., K. López de Ipiña, C. Hernández and N. Barroso., 2013 Enhancing the feature extraction process for automatic speech recognition with fractal dimensions, Cognitive Computation, pp1.6, Springer-Verlag. (2012), DOI: 10.1007/s12559-12-91-65.0

[32] Tsonis A., 2011. Reconstructing dynamics from observables: the issue of the delay parameter revisited. Int J Bifurcat Chaos 17:4229-4243.

[33] Jang JSR, 2011 Audio Signal Processing and Recognition. In: Roger Jang's Homepage. 2011. http://www.cs.nthu.edu.tw/~jang. Accesed in (2011).

[34] Esteller R, Vachtsevanos G, Echauz J, Litt B. , 2012. A comparison of waveform fractal dimen-sion algorithms. IEEE Trans Circuits Syst I Fundam Theory Appl 2001. 48:2:177-183.

[35] WEKA. Available online: http://www.cs.waikato.ac.nz/ml/weka/

[36] Picard, R.; Cook, D. "Cross-Validation of Regression Models". Journal of the American Statistical Association 1984, 79(387), 575–583.

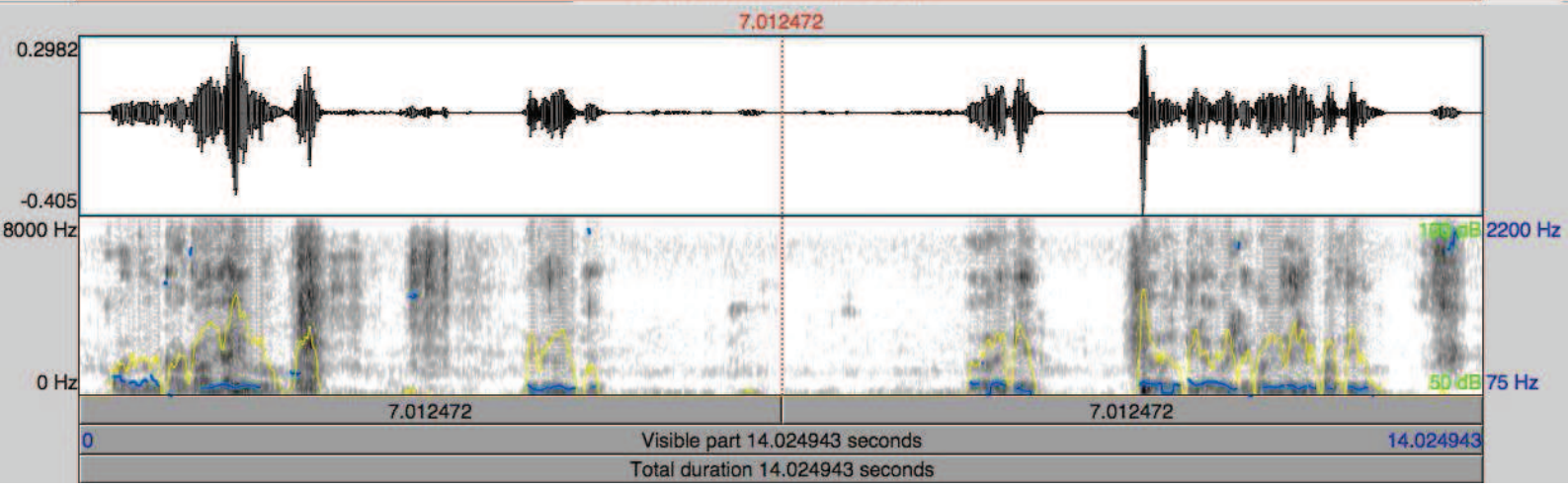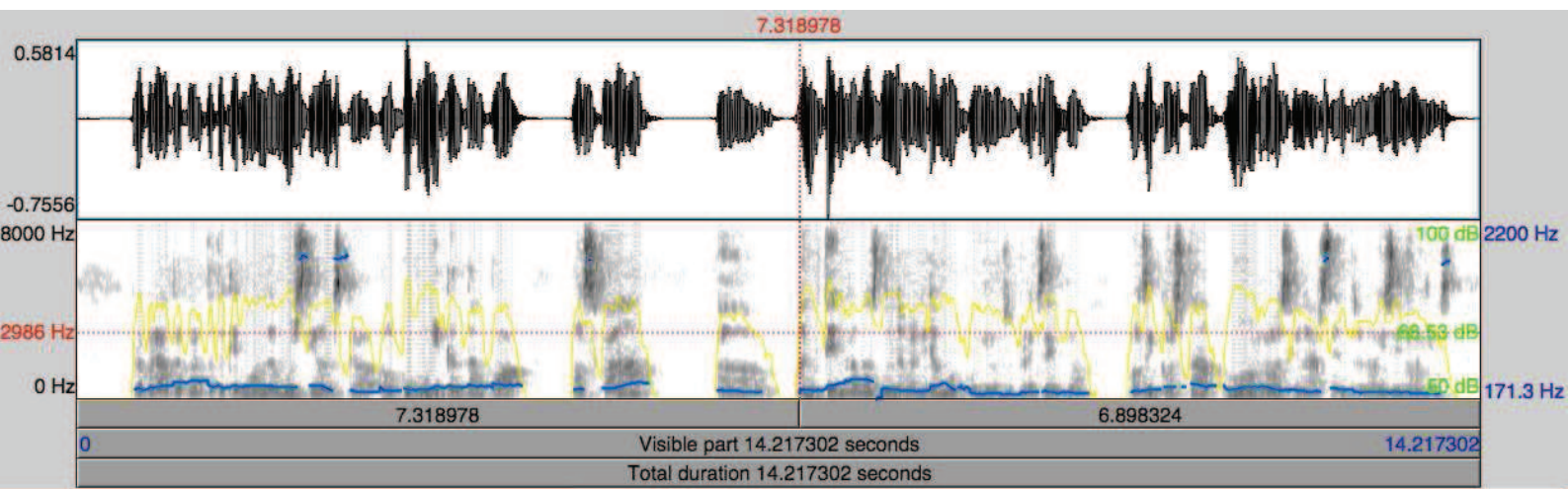[37] Praat: doing Phonetics by Computer. Available online: www.fon.hum.uva.nl/praat

Figure1

Figure2

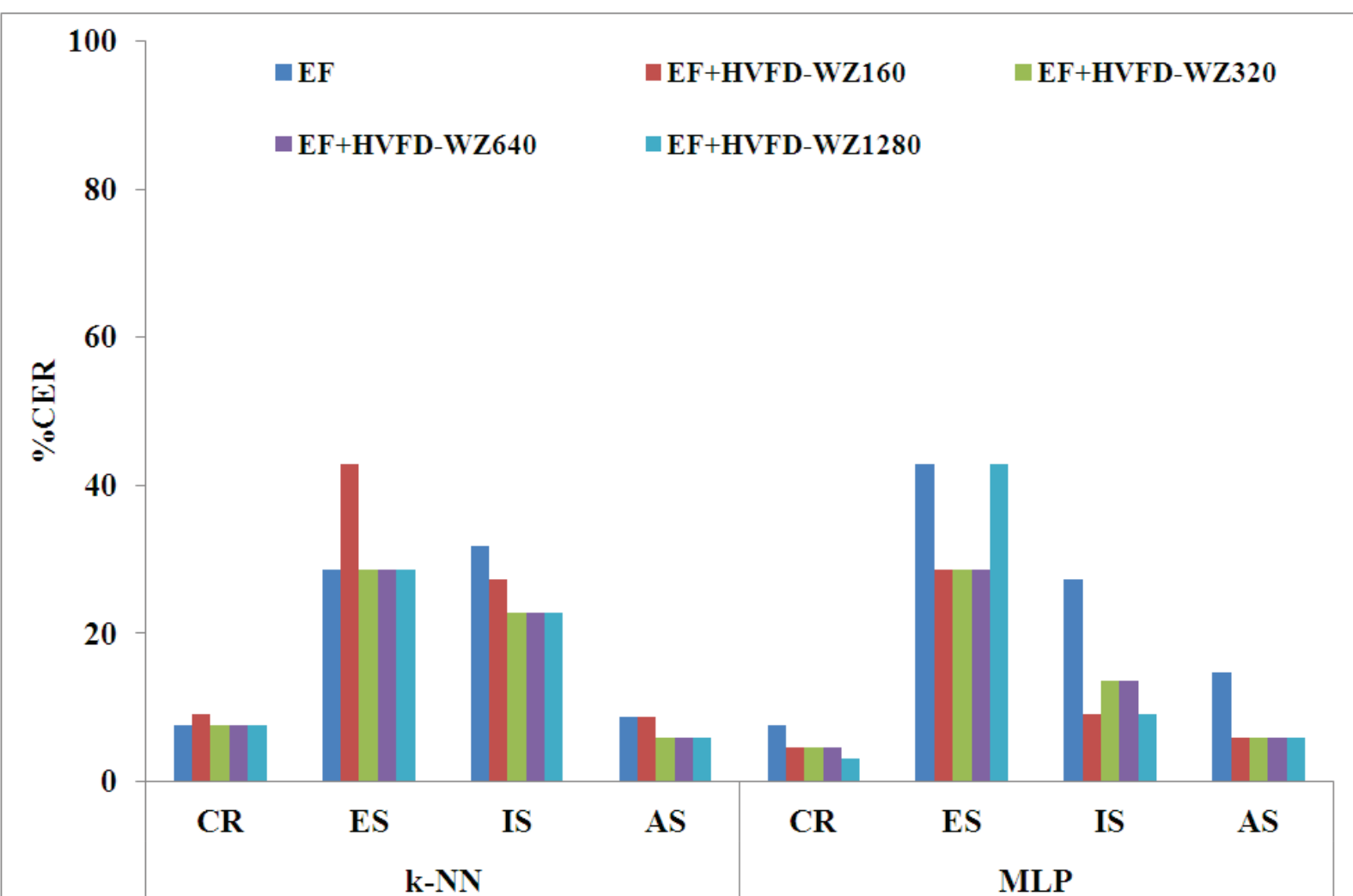Higuchi Fractal Dimension/CT

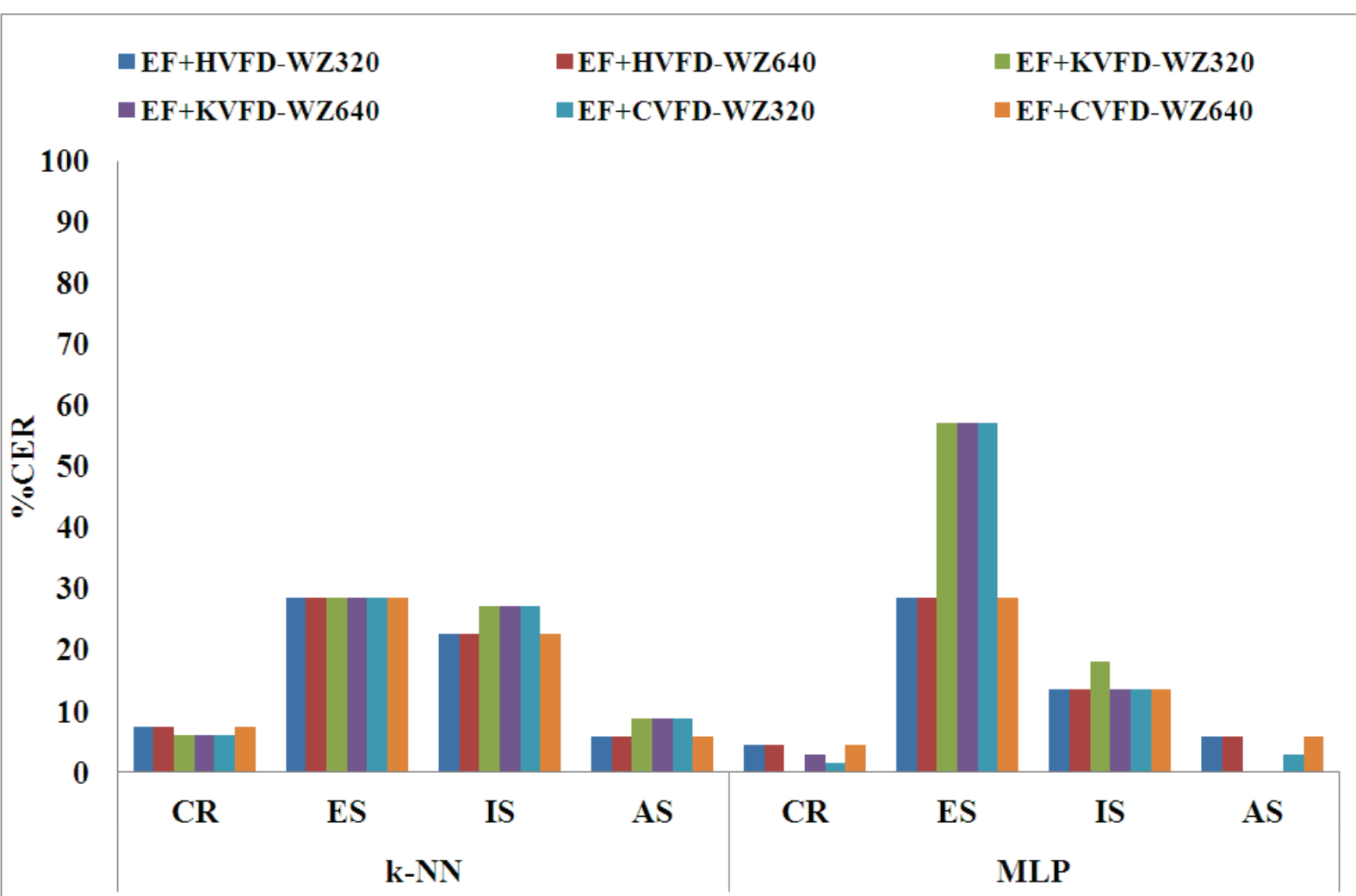Higuchi Fractal Dimension/AD
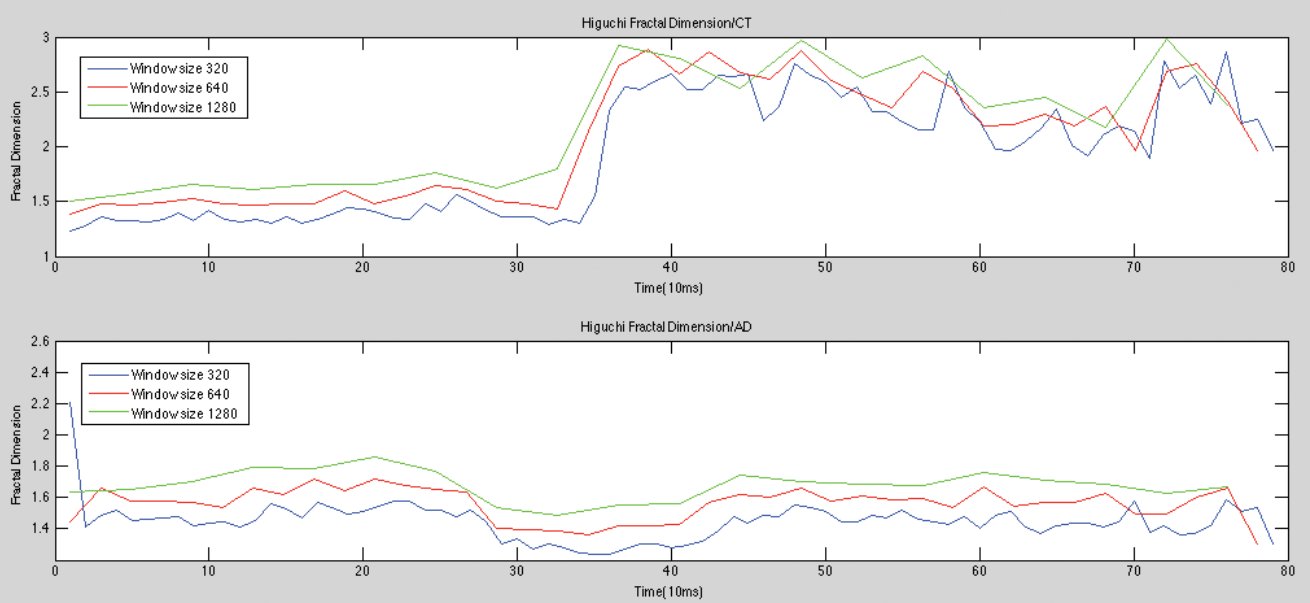
Figure3
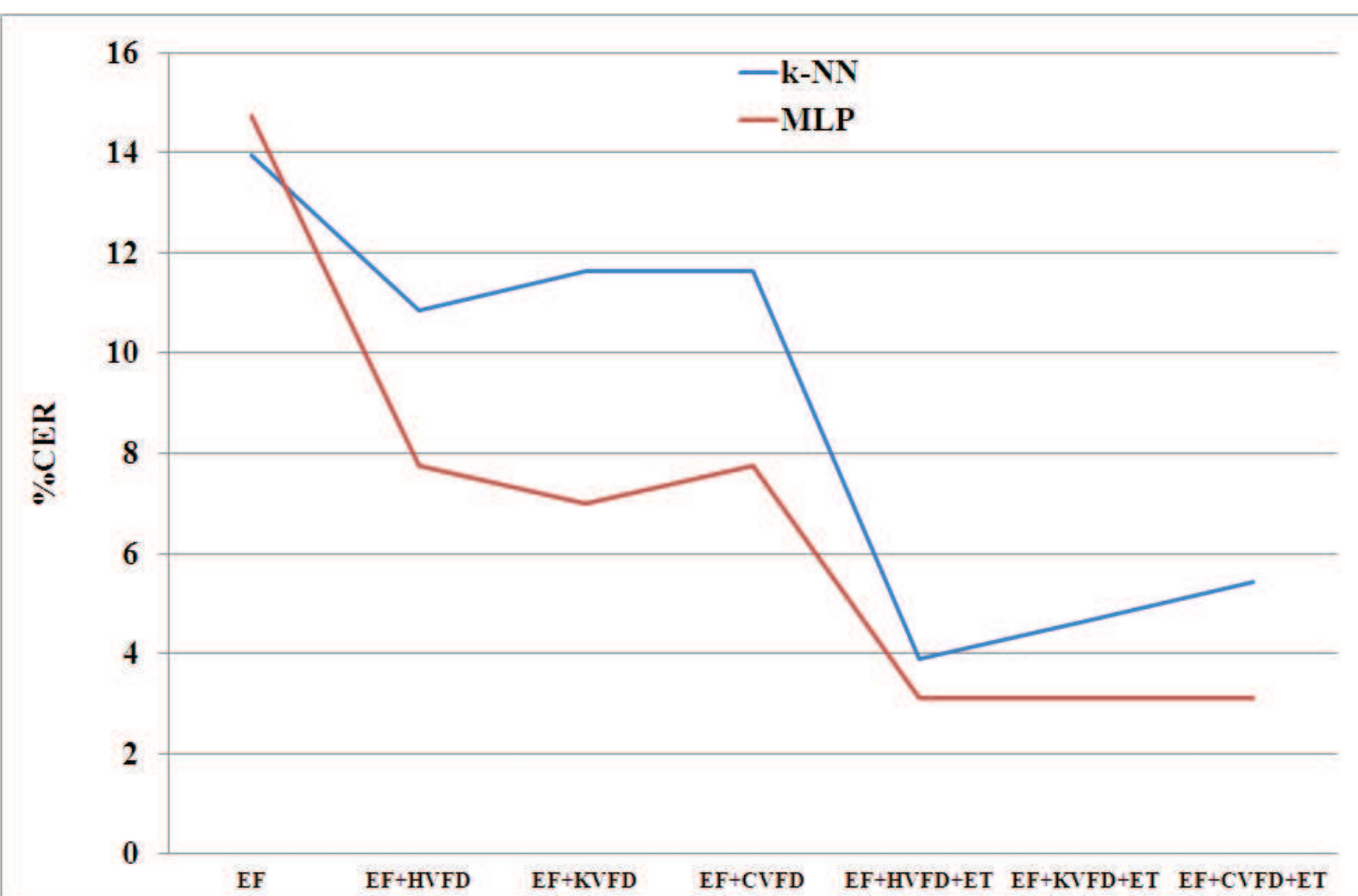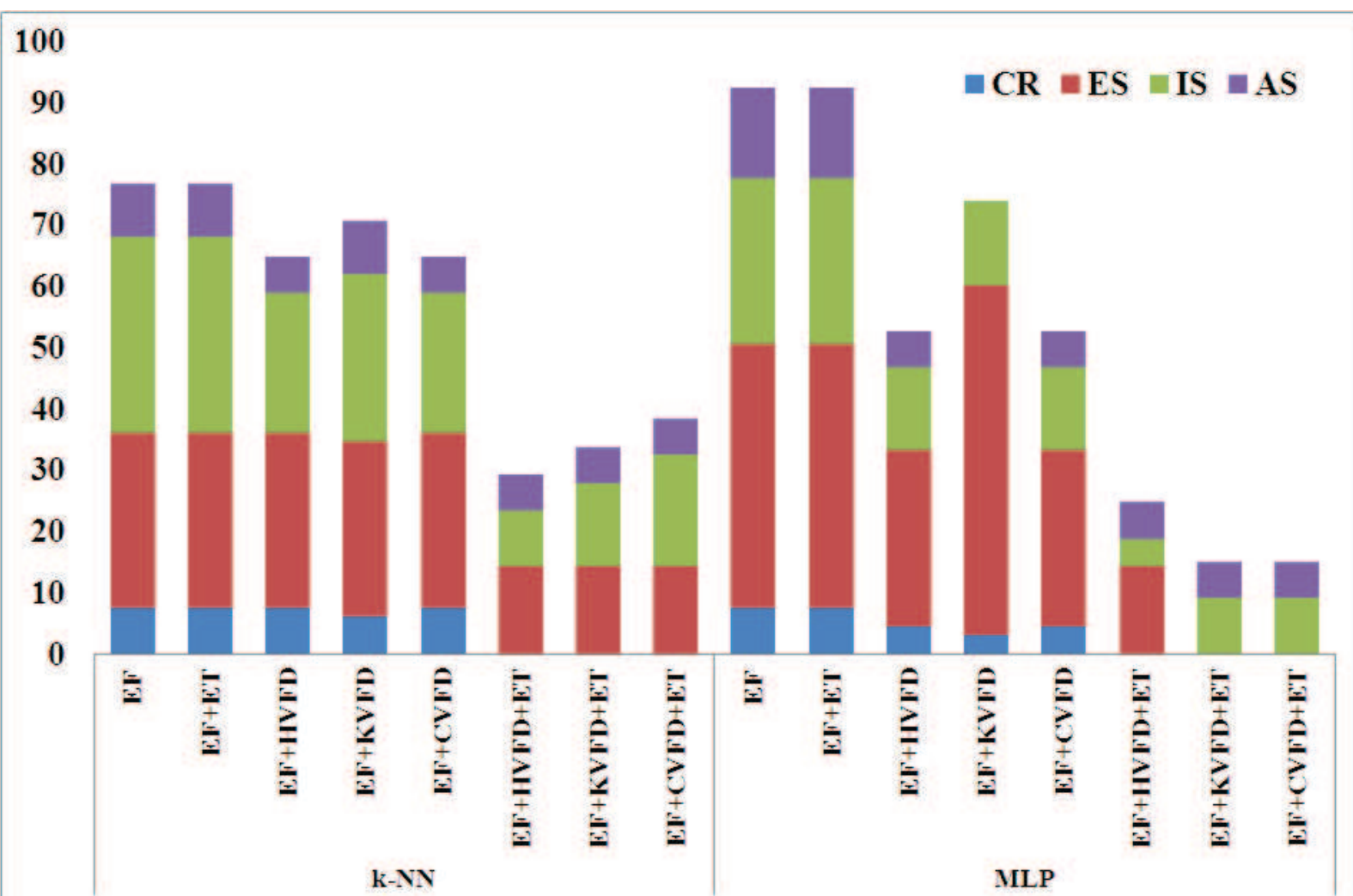
Figure4

Figure5

Figure6

Karmele López-de-Ipiña received the Ph.D. degree in Computer Science in 2003 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 Directot at the UPV/EHU (2004-2009). She is currently Head of the Engineering and Society research group. Her research interests are in Bioengineering and Biomedical Engineering, Pattern Recognition, Signal Processing, Ambient Intelligence and Robotics.

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.

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.

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 Digital Technologies research group. He is visiting scientist with the LIS-INPG Laboratory (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.

Aitzol Ezeiza, University of the Basque Country. A member of Member of the interdisciplinary Engineering and Society research group, he received his PhD degree on Engineering in 2012, but he works on Human Language Technology research since 1998.

Marcos Faundez-Zanuy was born in Barcelona, Spain. He received the B.Sc. degree in Telecommunication in 1993 and the Ph.D. degree in 1998, both from the Polytechnic University of Catalunya. He is now full Professor and Dean at EUP Mataró Tecnocampus and heads the signal processing group there. His research interests lie in the fields of biometrics and speech coding. He was the initiator and chairman of the European COST action 277 "Nonlinear speech processing". Prof. Faundez-Zanuy is a Spanish Liaison for EURASIP (European Association for Signal and Image Processing).