On Automatic Diagnosis of Alzheimer's Disease based on Spontaneous Speech Analysis and Emotional Temperature

K. Lopez-de-Ipiña[1], J.B. Alonso[2], J. Solé-Casals[3], N.Barroso[1], P. Henriquez[2], M. Faundez-Zanuy[4], C. Travieso[2], M. Ecay-Torres[5], P. Martinez-Lage[5], H. Egiraun[16]

[1]Systems Engineering and Automation Department, University of the Basque Country, Donostia 20008, Spain, {karmele.ipina, nora.barroso, harkaiz.egiraun}@ehu.es
[2] Universidad de Las Palmas de Gran Canaria, IDeTIC, {jalonso,ctravieso}@dsc.ulpgc.es, phenriquez@gi.ulpgc.es
[3] Digital Technologies Group. University of Vic, jordi.sole@uvic.cat
[4] Escola Universitaria Politècnica de Mataró (UPC), Tecnocampus , faundez@tecnocampus.cat
[5] CITA-Alzheimer Foundation; mecay,pmlage}@cita-alzheimer.org
[6] Research Center for Experimental Marine Biology and Biotechnology, Plentzia Marine Station, University of the Basque Country, Plentzia, Spain

Keywords: Alzheimer disease diagnosis, Spontaneous Speech, Emotion Recognition.

Abstract: Alzheimer's disease is the most prevalent form of progressive degenerative dementia; it has a high socio-economic impact in Western countries. Therefore it is one of the most active research areas today. Alzheimer's is sometimes diagnosed by excluding other dementias, and definitive confirmation is only obtained through a post-mortem study of the brain tissue of the patient. The work presented here is part of a larger study that aims to identify novel technologies and biomarkers for early Alzheimer's disease detection, and it focuses on evaluating the suitability of a new approach for early diagnosis of Alzheimer's disease by non-invasive methods. The purpose is to examine, in a pilot study, the potential of applying Machine Learning algorithms to speech features obtained from suspected Alzheimer sufferers in order help diagnose this disease and determine its degree of severity. Two human capabilities relevant in communication have been analyzed for feature selection: Spontaneous Speech and Emotional Response. The experimental results obtained were very satisfactory and promising for the early diagnosis and classification of Alzheimer's disease patients.

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 plan for the future and offers the best opportunity to treat the symptoms of the disease. According to current criteria, 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 typical AD pathological changes at autopsy [1-3]. This paper presents a new approach for early AD diagnosis based on two non-invasive and low-cost automatic methods: Automatic Spontaneous Speech Analysis (ASSA) and Emotional Temperature (ET).

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. Associated behavioral and psychological symptoms include apathy, irritability, depression, anxiety, delusions, hallucinations, disinhibition, aggression, aberrant motor behavior, as well as eating or sleep behavior changes [4,5]. All these symptoms lead to impaired performance in everyday family, social, and professional activities as the disease progresses from mild to moderate to severe dementia.

The diagnosis of AD is made on clinical grounds and requires the confirmation of a progressive dementia syndrome as well as the exclusion of other potential causes on the basis of clinical history and examination, complete blood workup and a brain-imaging test such as computer tomography (CT) or magnetic resonance imaging (MRI). This diagnosis of "exclusion" has changed in the past few years, as the interpretation of neuroimaging tests, including functional imaging with Single Photon Emission Computed Tomography (SPECT) and Photon Emission computed Tomography (PET), has focused on the "positive" findings of typical AD changes (medial temporal atrophy on CT or MRI, temporoparietal hypometabolism in PET) [6-9]. Nonetheless, the diagnosis of the early stages of not only mild cognitive impairment but also mild dementia remains problematic. On one hand, patients and relatives tend to ignore the first clinical manifestations, or ascribe them to the expectable cognitive changes related to age. Patients usually seek medical advice only 2 to 3 years after the onset of symptoms. On the other hand, physicians may feel uncertain about and uncomfortable with establishing a diagnosis until the whole picture of dementia is fully present, as they might need to apply long neuropsychological batteries, expensive neuroimaging techniques or invasive tests such as a lumbar puncture to reach a diagnosis. It is, then, not surprising that most patients are diagnosed only once they have reached the moderate stage of the disease and become substantially dependent. At that stage, unfortunately, no treatment strategy is very likely to be significantly efficacious in stopping or even in delaying the disease process [4,5]. Approaches to the early diagnosis of AD have in the past few years made significant advances in the development of reliable clinical biomarkers [10]. However, the cost and technology requirements make it impossible to apply such biomarkers to all patients with memory complaints; rather, patients must be clinically selected, so that an invasive lumbar puncture or an expensive PET is performed only on those patients for whom there is a strong suspicion of an underlying AD pathology. 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 and Emotional Temperature, 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 [11-15].

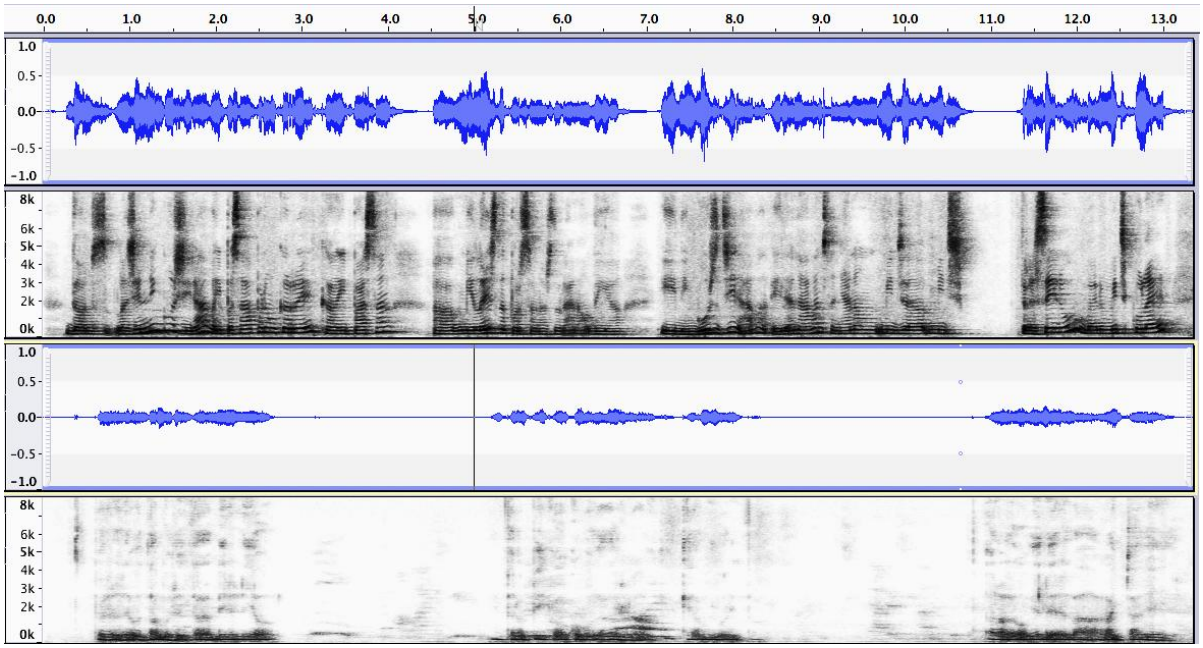

Figure 1. Signal and spectrogram of a control subject (top) and a subject with AD (bottom) during Spontaneous Speech

In addition to the loss of memory, one of the major problems caused by AD is the loss of language skills, which is reflected in difficulty with both speaking and comprehension and complicates natural communication with the environment (Figure 1). This inability to communicate is already present in the early phases of the diseases. There are various communication deficits in the area of language, including [16-18] 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 [3,4]:

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.

Besides language ability, the emotional response in Alzheimer's patients becomes impaired; this, too, seems to go through different stages. In the early stages, social and even sexual disinhibition appears, along with other behavioral changes (for example, being angry or being unable to remember, perform common tasks, or express oneself) [19-22]. However, patients retain their emotional memory, often crying easily and gratefully acknowledging caresses, smiles and hugs. The Alzheimer's patient reacts aggressively to things that for healthy persons are harmless and perceives threats or dangers where they do not exist. In more advanced stages of Alzheimer's disease, patients may often seem shy and apathetic, symptoms often attributed to memory problems or difficulty in finding the right words.

Some responses are likely to be magnified due to an alteration in perception. Research also suggests that patients in advanced stages of AD may display a reduced ability to feel emotions due to loss of memory; this may in turn induce apathy and depression [22,23].

The final objective of the full project 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 pilot experiment to define thresholds for a number of biomarkers related to spontaneous speech and emotional response. The data obtained will complement the biomarkers of each person. Other biomarkers can be extracted from multiple invasive tests (marrow extraction) as well as from neuroimaging and analytical or neuropsychological tests and from non-invasive automatic analysis of handwritten materials and drawings.

We have evaluated non-invasive tests in order to identify the correct features to characterize the four groups (control group (CR) and the three Alzheimer levels (ES, IS and AS)). Two such tests are Automatic Spontaneous Speech Analysis and Emotional Response Analysis (ERA) [15].

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), which is the freest and more natural expression of communication, possibly combined with other methodologies, has the potential to become a useful non-invasive method for early AD diagnosis [11-15].

Emotional Response Analysis on speech also has that potential: emotions are cognitive processes related to the architecture of the human mind, such as decision-making, memory, or attention. These processes, in turn, are closely linked to the learning and understanding that arise in intelligent natural or artificial systems in response to the necessity of surviving in a changing and partially unpredictable world [24-26]. Human interaction includes emotional information about communication partners that is transmitted through language explicitly and implicitly through nonverbal communication. Human emotions are affected by the environment and by direct interaction with the outside world, but also by the emotional memory emerging from the experience of the individual and cultural environment, the so-called socialized emotion [24]. Emotions use the same components – subjective, cultural, physiological and behavioral – that the individual's perception registers with regard to the individual's mental and physical state and the body's interaction with the environment [25]. In this work ERA has been performed with regard to classical features and to Emotional Temperature, described in section 3. This feature is based on the analysis of a number of prosodic and paralinguistic feature sets obtained from a temporal segmentation of the speech signal.

2. MATERIALS

2.1 Main Database of individuals

Trying to develop a new methodology applicable to a wide range of individuals of different sex, age, language and cultural and social background, we have constructed a multicultural and multilingual (English, French, Spanish, Catalan, Basque, Chinese, Arabian and Portuguese) database with video recordings of 50 healthy and 20 AD patients (with a prior diagnosis of Alzheimer) recorded for 12 hours 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 efforts to find 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.

2.2 Pre-processing

The video was processed and the audio extracted in wav format (16 bits and 16 Khz). The first step was to remove non-analyzable events: laughter, coughing, short hard noises and segments where speakers overlapped. Next, background noise was removed manually 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.

2.3 Individuals selected for the study

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 (ES=4, IS=10, AS=6). The patients speak English, Spanish, Basque, and Portuguese. The control group (CR) was made up of 20 subjects (10 male and 10 female, aged 20-98 years) representing a wide range of speech responses. Selected languages are: English, French, Spanish, Catalan, Basque, Chinese and Arabian. This subset of the database is called AZTIAHORE. The preliminary experiment whose results are reported here was carried out on a subset of AZTIAHORE called AZTITXIKI and consisting of 5 members of the control group and 5 AD-diagnosed patients at different stages of the disease (ES=1, SS=2, AS=2). The control-group members (2 females and 3 males) were of middle age (M) or elderly (E); the AD-diagnosed persons (2 females and 3 males) were all elderly.

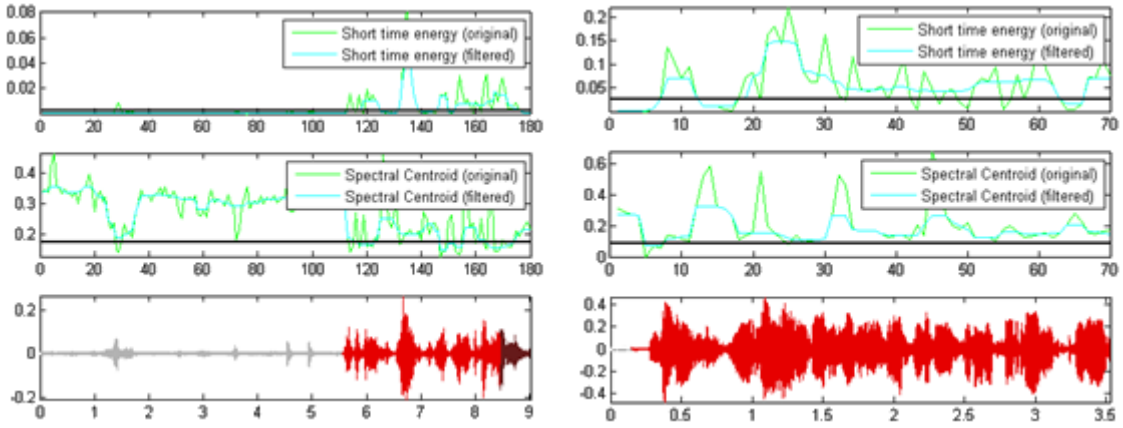

Figure 2. Plots of Speech Signal, Short Time Energy and Spectral Centroid, filtered by a median filter, for a control subject and a subject with AD

3 METHODS

3.1 Feature extraction

Feature extraction was performed using the Praat [27] software package and software developed by us in MATLAB.

3.1.1 Automatic Spontaneous Speech Analysis (ASSA)

The analysis of Spontaneous Speech fluency is based on three families of features (SSF set), obtained using Praat software. For that purpose, an automatic Voice Activity Detector (VAD) [28,29] has extracted voiced/voiceless segments as parts of an acoustic signal. These three families of features include:

1. *Duration:* this includes the histogram calculated over the most relevant voiced and voiceless segments, the average of the most relevant voiced/voiceless, voiced/voiceless percentage and spontaneous speech evolution along the time dimension, and the voiced and voiceless 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 (see Figure 2). These features form an SSF set.

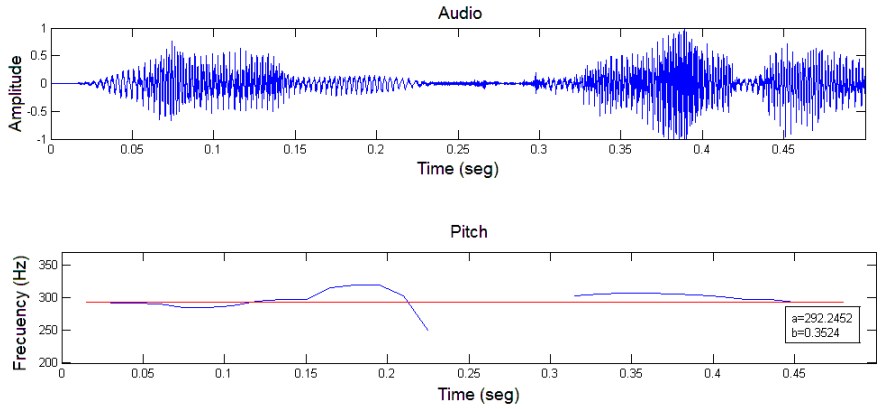

Figure 3. Example of a model of a pitch curve with a polynomial function of first order

3.1.2 Emotional Speech Analysis

In this study, we aim to accomplish the automatic selection of emotional speech by analyzing three families of features in speech:

1. *Acoustic features:* pitch, standard deviation of pitch, max and min pitch, intensity, standard deviation of 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 voiceless frames, degree of voice breaks.

Short-term energy is the principal and most natural feature that we analyzed. Physically, energy is a measure of how strong a signal is present at any one time. Energy is measured in continuous speech to discover voiced sounds, which have a higher energy than silence or un-voiced sounds, as shown in Figure 2. These features form an EF set. The energy of the signal is calculated as in section 3.1.1 on a short- time basis, by windowing the signal at a particular time, squaring the samples and taking the average [30]. The square root of this result is the engineering quantity known as the root-mean square (RMS) value.

3.1.3 Emotional Temperature

This method proposes a new 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% [31]. 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 (see Figure 3).

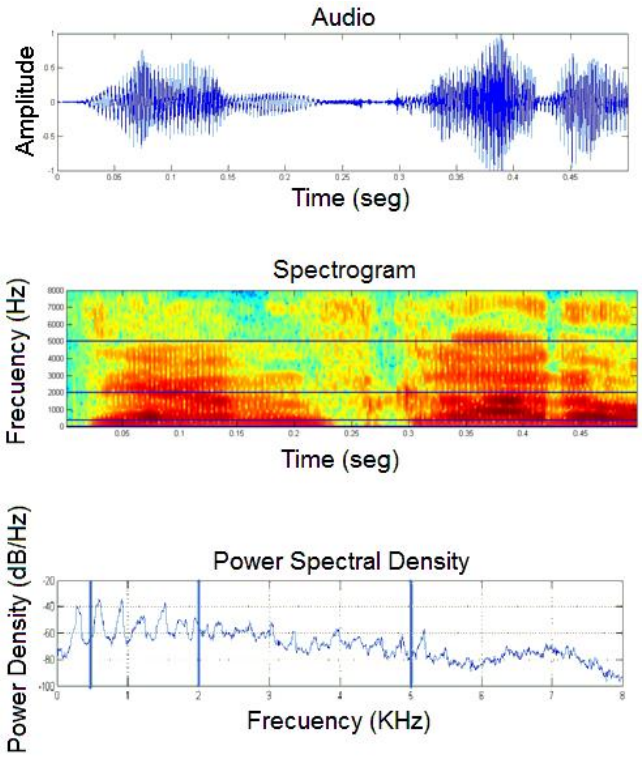

Figure 4. Example of the spectral distribution in different frequency bands

For prosodic features, a voiced/voiceless decision is made for each frame and two linear regression coefficients of the pitch contour $p(n)$ [32-34] are obtained:

$$MIN(a,b) = \sum_{I=1}^{n} \left(p_i(n) - a - bx_i(n)\right)^2 \quad (3.1)$$

where the coefficients a and b are computed using the method of least squares. In our implementation we use the pitch estimation algorithm called YIN [35].

For paralinguistic features, voice spectral energy balances [32] are calculated from each frame, quantified using 4 percentages of energy concentration in 4 frequency bands Bi. For a sampling frequency greater than 16KHz, the frequency bands are divided into the following ranges: B0=[0Hz, 400Hz], B1=[400Hz, 2KHz], B2=[2KHz, 5KHz], B3=[5KHz, 8KHz], estimated in our previous research related to the phonatory system [36] (see Figure 4).

The percentage of energy in each frequency band is obtained using the following expression:

$$E_{B_i} = \frac{\sum_{f=B_i}|X(f)|^2}{\sum_{f=0}^{8KHz}|X(f)|^2} 1 \leq i \leq 5 \quad (3.2)$$

Where $|X(f)|^2$is a periodogram of the temporal frame $\{x(n)\}$.

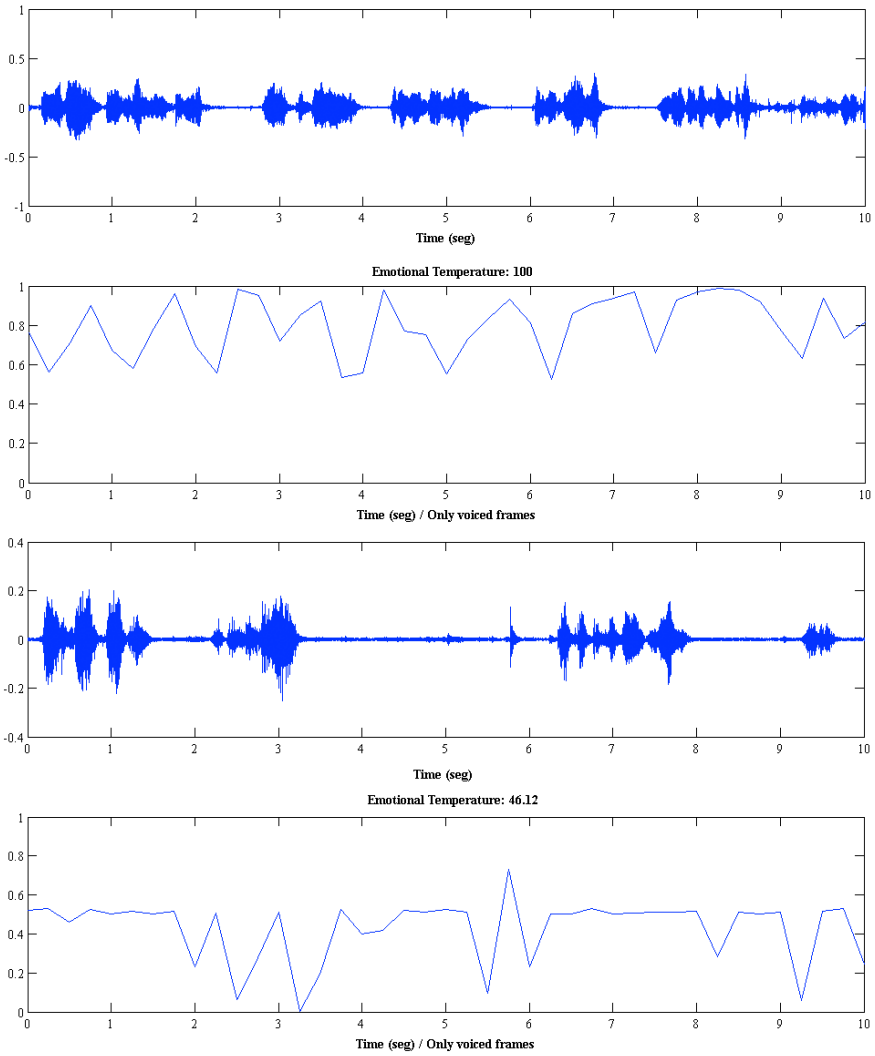

Figure 5. Emotional Temperature for healthy persons and persons with AD

Then Emotional Temperature is calculated as follows:

1. A Support Vector Machine (SVM) is trained with a balanced segment set extracted from the database (AZTIAHORE).
2. For each speech segment, each temporal frame is classified by the SVM as "pathological or "non- pathological".
3. The percentage of temporal frames classified as "non-pathological" is calculated. This value, i.e., the number of non-pathological frames, is the "emotional temperature".
4. The "Emotional Temperature" is finally normalized in order to have ET=50 as the threshold obtained from the training database; that threshold indicates the limit between pathological and non-pathological frames. This normalization will make it substantially easier for medical specialists to interpret the data.

Figure 5 shows an example of ET values for a healthy subject (ET=94.9) and for an AD subject (ET=44.7). We have used a freely available implementation called LIBSVM [37] with a radial basis kernel function.

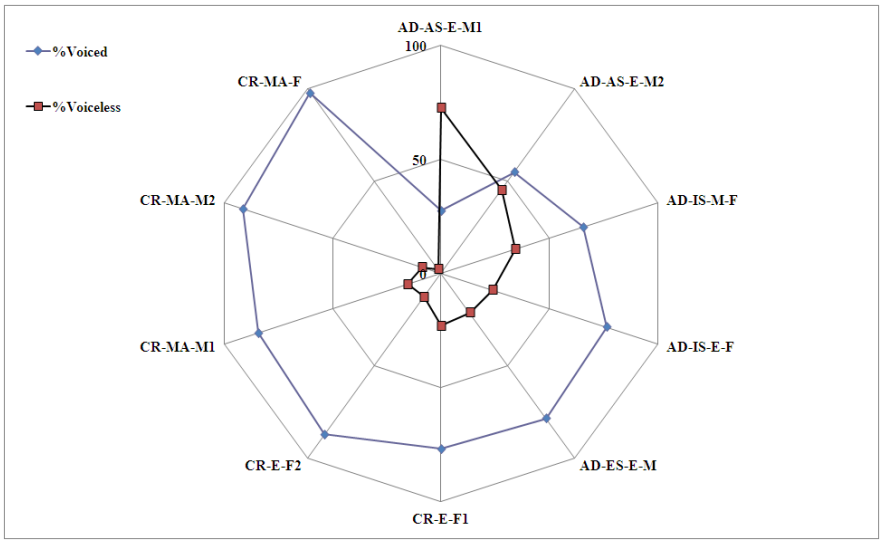

Figure 6. Voiced and voiceless (%) for AZTITXIKI set. There is an increment of voiceless (%) and a decrement of voiced (%) as the disease progresses from ES to AS.

3.1.4 Feature sets

Based on these presented characteristics, four feature sets have been created for experimentation:

1. SSF, set described in 3.1.1
2. EF, set described in 3.1.2
3. EF+ET: EF and Emotional Temperature (ET) for characterization of Emotional Response
4. SSF+EF: speech features set for integral characterization, including SS features and Emotional Response on Speech features
5. SSF+EF+ET: previous feature set combination for analyzing Integral Response on speech (speech fluency and emotional response).

3.2 Automatic Classification

The goal of these experiments was to examine the potential of the selected features to help in the automatic measurement of the degradation of Spontaneous Speech, Emotional Response and their manifestation in persons with AD as compared to the control group. The experiments have analyzed the automatic measurement and definition of appropriate features in Spontaneous Speech, Emotional Response Analysis and Integral Response in the speech of persons with AD over AZTIAHORE subset (200 segments). Four groups/classes (CR, ES, IS, AS) and four different feature sets were evaluated (section 3.1.4). Three tests were designed for the experimentation: 1) *Spontaneous Speech Fluency*; 2) *Analysis of Emotional Response*; 3) *INTEGRAL*, which analyzes the integral degradation on speech (speech fluency and emotional response). WEKA [38] software was used in carrying out the experiments.

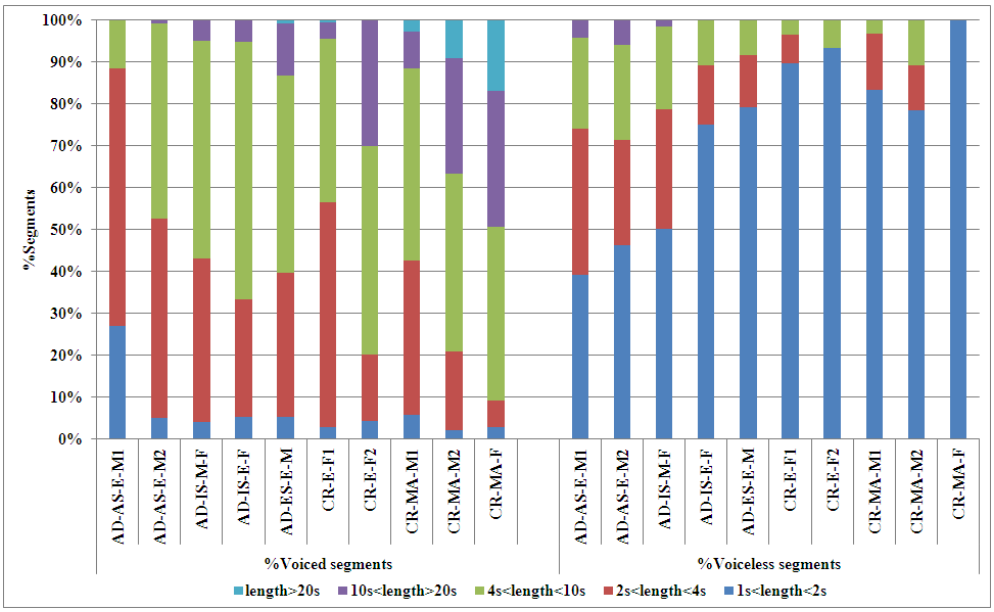

Figure 7. Example of the preliminary experimentation**.** Voiced and voiceless segment (%) analysis with regard to voiceless segment length.

The automatic classification of speech is based on several classifiers. We used five different classifiers: (1) a SVM with polynomial kernel, (2) a Multi Layer Perceptron (MLP) with one hidden layer of 100 neurons with 1000 training steps, (3) a K-Nearest Neighbor Algorithm (K-NN), (4) Decision Trees (DT) and (5) a Naive Bayes net. The WEKA software toolkit was used in carrying out the experiments.

The results were evaluated using Accuracy (Acc) and Classification Error Rate (CER) measurements with regard to segments. For the training and validation steps, we used *k*-fold cross-validation with *k*=10. Cross validation is a robust validation for variable selection [39]. These features will define the CR group and the three AD levels.

4 RESULTS AND DISCUSSION

4.1 Preliminary Experimentation

In the first stage, we carried out preliminary experimentation on the AZTITXIKI subset, analyzing the direct characteristics of Spontaneous Speech involved in AD symptoms. The experiment was designed to detect changes and features in SS characterizing the control group, on the one hand, and each of the different groups of AD levels, on the other. The first set of tests consisted of ASSA experiments. Persons suffering from AD manifested a lower voiced percentage and a higher voiceless percentage (see Figure 6) in their spontaneous speech than healthy subjects. This indicates a significant loss of fluency in the speech of AD-suffering subjects. The analysis shows that persons with AD tend to decrease the number and fluency of voiced segments by increasing the voiceless segment length and decreasing the length of voiced segments (see Figure 7).

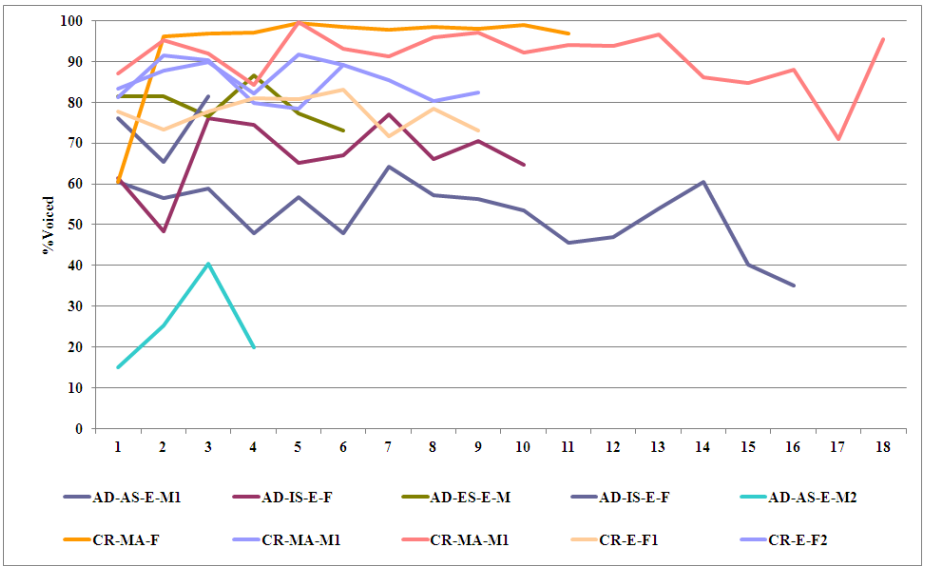

Figure 8. Spontaneous Speech Evolution with regard to the Speech Percentage along the time axis, for the consecutive segments (S1:SN).

Persons with AD also displayed a decreasing slope in the evolution of their Spontaneous Speech, and a decreasing slope is evident in the analysis of their voiced and voiceless segments with regard to length segments. Figure 7 displays the tendency to use an increased number of short voiced segments and long voiceless segments. The speech of AD sufferers is fluent only for short periods of time, and segments longer than 20 seconds seldom appear in their spontaneous speech. Figure 8 shows spontaneous speech evolution with regard to the speech percentage along the time axis for consecutive segments (S1:SN): it can be observed that persons with AD decrease the voiced percentage along the time axis. The results also show a higher Short Time Energy for members of the control group than for AD sufferers and a higher Spectral Centroid for the AD group (see Figure 2).

4.2 Automatic Classification

In a preliminary step, we calculated the Emotional Temperature for each segment, using the method described in section 3.2. Automatic classification was then performed over the speech features sets described in section 3.1.4 using the Machine Learning paradigms described in 4.2. The aim was to use the pilot study to analyze the efficacy of the following tests: Automatic Spontaneous Speech fluency, Emotional Response in speech, and a combination of both in Integral Speech.

In the first test, *Spontaneous Speech Fluency*, only the SSF set was used. The performance of SSP is fairly good for most of the paradigms, but it could be improved for MLP and Naive Bayes (see Figure 9), which tend to confuse ES group members with CR members and also confuse the three stages of AD. KNN, one of the paradigms with good performance, also has the lowest computational cost.

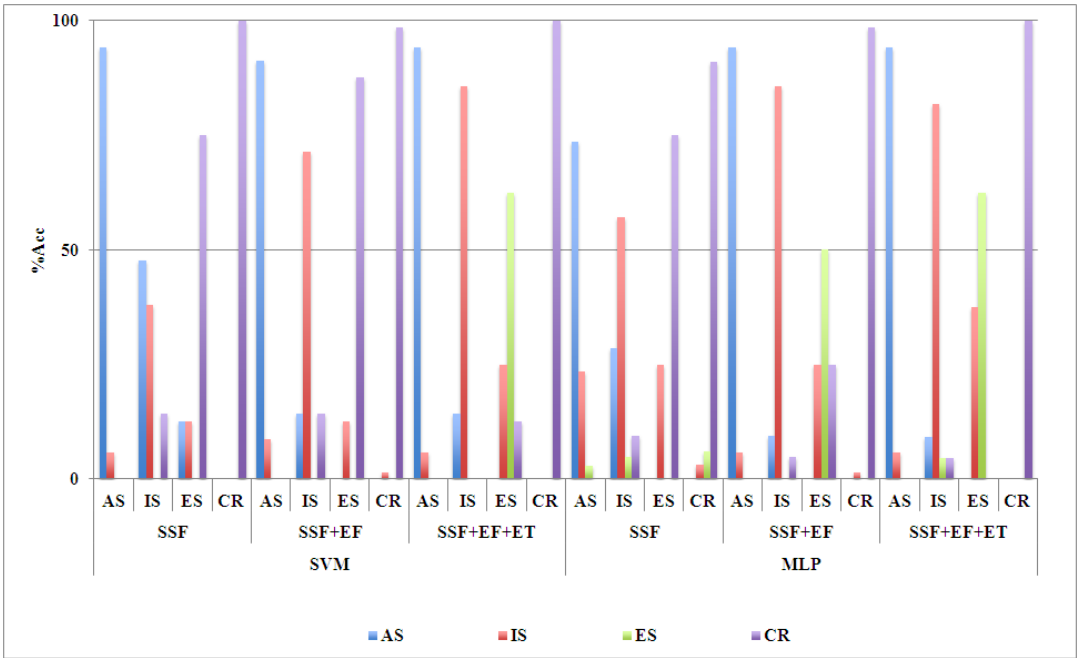

Figure 9 Classification Error Rate in % for the three defined tests in the pilot study and for each corresponding feature set.

In the second test, Emotional Response, EF and EF+T were used. The best results were obtained when the ET feature was included, with Acc near to the optimum value (see Figure 9). All the paradigms obtained very good results, the best one being MLP – though KNN also came near the optimum, and with less computational cost. The main mistakes the system makes is confusing the ES and IS groups and confusing ES group members with CR members when the ES patients are in the pre-disease stage. The presence of ET in the sets points towards the optimum for all paradigms without adding computational cost.

With regard to the *INTEGRAL* test, this includes information on both Spontaneous Speech and Emotional Response, the results also show an improvement for all paradigms when ET is included. The best results are obtained for MLP and SVM, but KNN also achieves a fairly good result without increase in computational cost.

A detailed analysis of the *INTEGRAL* test with regard to Accuracy (%) for all classes is shown in Figure 10, while Table 1 summarizes the Accuracy (%) global results with regard to the *INTEGRAL* test for all paradigms. Figure 10 shows the results obtained for the control group (CR) and for the three levels of AD (ES, IS and AS) for classes with MLP and SVM paradigms. In the results for these classes, the inclusion of ET obtains the best results for all classes. This set also improves the classification with regard to early detection (ES class). IS is also better at identifying mid-level AD. The model is able also to discriminate between pathological and non-pathological segments in each patient for the three tests of the pilot study.

Table 1. Accuracy (%), global results with regard to test and feature sets for *INTEGRAL* test.

Feature set	MLP	SVM	DT	KNN	Naive Bayes
SSF+EF	92,24	86,04	83,72	84,49	87,59
SSF+EF+ET	93,02	93,79	91,47	87,59	87,59

We analyzed the results with regard to the three tests (Emotional response, Spontaneous Speech fluency and *INTEGRAL* analysis). The analysis has taken into account the global system results, the results for each patient, and also AD level results.

1. Global system results The results are satisfactory for this preliminary pilot study. The objectives of the project have been achieved. The obtained results will be used to adjust the tool so that new and more accurate analyses can be carried out with a new set of volunteers in the early detection stage.
2. Specific test results. The system helps analyze a group's condition not only with regard to global features but also with regard to the specific test. This opens a promising research avenue in mixing biomarkers from different sources.
3. Results for the groups. The system obtains good results for the most advanced disease stages, IS and AS, and an encouraging rate – 60% – for the ES stage. The rate should be adjusted for the ES stage by analyzing a larger set of currently available data.
4. AD level results: Regarding the IS and AS classification errors, these are detected for segments belonging to patients at the border between two stages. The segments of ES patients mainly appear mixed with the CR group and sometimes with the IS group. This also points to the existence of disease progression.

On the whole, then, these non-invasive tests could form a very useful tool in the future for medical specialists working to define early diagnosis grounds for clinical AD.

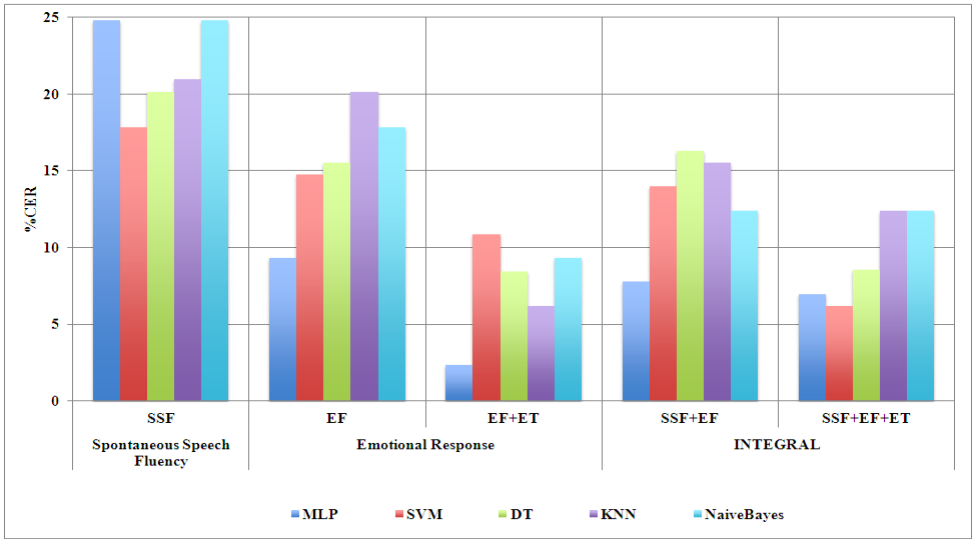

Figure 10 Accuracy (%) of classes for *INTEGRAL* test and each corresponding feature set.

7 CONCLUSIONS

The main goal of the present pilot study was to analyze features in Spontaneous Speech and Emotional Response in the pre-clinical stage of AD in order to design appropriate tests for the early diagnosis of the disease. These features significantly aid health specialists in distinguishing AD sufferers from healthy subjects as well as in discriminating between the three AD levels. We have analyzed features relating to speech duration, time domain, spectral domain, as well as a newly defined feature called Emotional Temperature. The performance of the approach is very satisfactory, and the results are promising for the early diagnosis and classification of AD patient groups. In future work, we will evaluate this approach with an early diagnosis database and new tests oriented toward semantic and memory tasks. We will also introduce new features related to a non-linear dynamic. Further, we will integrate the methodologies described here with methods of to automatically analyze drawing and handwriting, as well as with automatic analysis of facial features. We will also extend the study by examining a larger number of test subjects and a greater variety of pathologies.The first step will be developed with new recording material from the CITA Alzheimer Foundation and local hospitals.

ACKNOWLEDGEMENTS

This work has been partially supported by a SAIOTEK grant from the Basque Government, the University of Vic under the research grant R0904, and the Spanish Ministerio de Ciencia e Innovación TEC2012-38630-C04-03. Professor Iciar Martinez (Research Center for Experimental Marine Biology and Biotechnology-Plentziako Itsas Estazioa (PIE), University of the Basque Country & IKERBASQUE, Basque Foundation for Science), is gratefully acknowledged for helpful discussions and for her contribution to the preparation of the manuscript.

REFERENCES

1. Mc Kahn G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Workgroup on Alzheimer's disease. 1984; 24:939-944.
2. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, Mohs RC, Morris JC, Rossor MN, Scheltens P, Carrillo MC, Thies B, Weintraub S, Phelps CH. 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 (PDF). Alzheimers Dement. 2011 May;7(3):263-269.
3. Van de Pole LA, Van der Flier WM, Hensel A, Gertz HJ, Scheltens P. The effects of age and Alzheimer's disease on hippocampal volumes, a MRI study. Alzheimer's and Dementia: The Journal of the Alzheimer's Association. 2005; 1(1,Supplement 1):51. doi: 10.1016/j.jalz.2005.06.205
4. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology. 1993; 43(11): 2412b-2414b.
5. Diagnostic and Statistical Manual of Mental disorders, 4th Edition Text Revision: DSM-IV-TR; Publisher: American Psychiatric Association, Washington DC, USA, 2000.
6. Petrella JR, Coleman R, Doraiswamy P. Neuroimaging and early diagnosis of Alzheimer's disease: A look to the future. Radiology. 2003 ; 226:315–336.
7. Wernickand MN, Aarsvold JN. EmissionTomography: TheFundamentals of PET and SPECT. Publisher: Elsevier New York, USA. 2004.
8. Pareto D, Aguiar P, Pavia J, Gispert J, Cot A, Falcon C, Benabarre A, Lomena F, Vieta E, Ros D. Assessment of SPM in per-fusion brain SPECT studies. A numerical simulation study using boot-strap resampling methods. IEEE Trans. Biomed. Engineeing (TBE). 2008; 55(7):1849-1853.
9. Álvarez I, Górriz JM, Ramírez J, Salas-Gonzalez D, López M, Segovia F, Padilla P, Gracía C. Projecting independent components of SPECT images for computer aided diagnosis of Alzheimer's disease. Pattern Recognition Letters. 2010; 31(11):1342-1347.
10. Alzheimer’s Association. Available online: http://www.alz.org/research/funding/global_biomarker_consortium.asp
11. Faundez-Zanuy M, Hussain A, Mekyska J, Sesa-Nogueras E, Monte-Moreno E, Esposito A, Chetouani M, Garre-Olmo J, Abel A, Smekal Z, Lopez-de-Ipiña K. Biometric Applications Related to Human Beings: There Is Life beyond Security. Cognitive Computation. 2012 ; 5(1):136-151. doi: 10.1007/s12559-012-9169-9.
12. Sesa-Nogueras E, Faundez-Zanuy M, Mekyska J. An information analysis of in-air and on-surface trajectories in online handwriting. Cognitive Computation. 2013; 4(1):195-205. doi: 10.1007/s12559-011-9119-y.
13. Gómez-Vilda P, Rodellar-Biarge V, Nieto-Lluis V, Muñoz-Mulas C, Mazaira-Fernández L.M, Martínez-Olalla R, Álvarez-Marquina A, Ramírez-Calvo C, Fernández-Fernández M. Characterizing Neurological Disease from Voice Quality Biomechanical Analysis. Cognitive Computation. 2013; doi: 10.1007/s12559-013-9207-2.
14. Henríquez P., Alonso-Hernández J.B, Ferrer-Ballester M.A, Travieso-González C.M, Orozco-Arroyave J.R. Global Selection of Features for Nonlinear Dynamics Characterization of Emotional Speech. Cognitive Computation. 2012; doi: 10.1007/s12559-013-9157-0.
15. López de Ipiña K, Alonso JB, Solé-Casals J, Barroso N, Faundez M, Ecay M, Travieso C, Ezeiza A, Estanga A. Alzheimer's Disease Diagnosis based on Automatic Spontaneous Speech Analysis. IWAAL Special Session in Challenges in Neuroengineering, Proceeedings of International Conference on Neural Computation Theory and Applications (NCTA). Barcelona. 2012.
16. Buiza C. Evaluación y tratamiento de los trastornos del lenguaje; Matia Fundazioa; Donostia. 2010.

17. Martinez F, Garcia J, Perez E, Carro J, Anara JM. Patrones de Prosodia expresiva en pacientes con enfermedad de Alzheimer. Psicothema. 2012 ; 24(1):16-21.
18. Hu WT, McMillan C, Libon D, Leight S, Forman M, Lee VMY, Trojanowski JQ, Grossman M., Multimodal predictors for Alzheimer's disease in non fluent primary progressive aphasia, Neurology. 2010; 75(7):595-602.
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 MS, Gyurak A, McCarthy M, Miller BL, Levenson RW. Emotion regulation deficits in frontotemporal lobar degeneration and Alzheimer's disease. Psychol Aging. 2010; 25(1):30-37. 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(5):1132-1146. doi:10.1044/1092-4388(2010/09-0030).
23. Henry JD, Rendell PG, Scicluna A, Jackson M, Phillips LH. Emotion experience, expression, and regulation in Alzheimer's disease. Psychology and Aging. 2009; 24(1):252-257.
24. Knapp, M.L. Essentials of nonverbal communication. Publisher: Holt, Rinehart & Winston, NY, USA. 1980.
25. 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.
26. Plutchnik R. Emotion: A psychoevolutionary synthesis. Publisher: Harper and Row New York, USA. 1980.
27. Praat: doing Phonetics by Computer. Available online: www.fon.hum.uva.nl/praat
28. Voice Activity Detector algorithm (VAD). Available online: www.mathwork.com
29. Solé J, Zaiats V., A Non-Linear VAD for Noisy Environment. Cognitive Computation. 2010; 2(3):191-198.
30. M.M. Rahman, M.A. Bhuiyan , Continuous Bangla Speech Segmentation using Short-term Speech Features Extraction Approaches, International Journal of Advanced Computer Sciences and Applications, 2012 vol.3-11, pp 131-138
31. Pao TL, Chien CS, Yen JH, Chen YT, Cheng YM. Continuous tracking of user emotion in mandarin emotional speech. Proceedings of 3th International Conference on International Information Hiding and Multimedia Signal Processing (IIH-MSP'07), Splendor Kaohsiung, Taiwan. 2007 November 26-28; 1:47-52.
32. Petrushin VA. Emotion in speech: recognition and application to call centers. Proceedings, Conference on Artificial Neural Networks in Engineering (ANNIE'99), St. Louis, Missouri, USA. 1999 November 7-10; 7-10
33. Lee CM, Narayanan S. Emotion recognition using a data-driven fuzzy interference system. Proceedings of 8th European Conference on Speech Communication and Technology (ECSCT'03), Geneva, Switzerland. 2003 September 1-4; 157-160.
34. Kwon, O.W.; Chan, K.; Hao J.; Lee, T.W. Emotion recognition by speech signals. Proceedings of 8th European Conference on Speech Communication and Technology

(ECSCT'03), Geneva, Switzerland. 2003 September 1-4; 125-128
35. De Cheveigné A, Kawahara H. YIN, a fundamental frequency estimator for speech and music. Journal of the Acoustical Society of America. 2002; 111(4):1917-1930.
36. 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.
37. Chang CC, Lin CJ. LIBSVM: a library for support vector machines; 2001. Available online: http://www.csie.ntu.edu.tw/~cjlin/libsvm
38. WEKA. Available online: http://www.cs.waikato.ac.nz/ml/weka/
39. Picard R, Cook D. Cross-Validation of Regression Models. Journal of the American Statistical Association. 1984; 79(387):575–583.