

Article

On the Selection of Non-Invasive Methods Based on Speech Analysis Oriented to Automatic Alzheimer Disease Diagnosis

KarmeLe López-de-Ipiña ^{1,*}, Jesus-Bernardino Alonso ², Carlos Manuel Travieso ², Jordi Solé-Casals ³, Harkaitz Egiraun ^{1,4}, Marcos Faundez-Zanuy ⁵, Aitzol Ezeiza ¹, Nora Barroso ¹, Miriam Ecay-Torres ⁶, Pablo Martinez-Lage ⁶ and Unai Martinez de Lizardui ¹

¹ Systems Engineering and Automation Department, University of the Basque Country UPV/EHU, Donostia 20018, Spain; E-Mails: harkaitz.egiraun@ehu.es (H.E.); aitzol.ezeiza@ehu.es (A.E.); nora.barroso@ehu.es (N.B.); unai.martinezdelizardui@ehu.es (U.M.L.)

² Signal and Communication Department (DSC), Institute for Technological Development and Innovation in Communications (IDeTIC), University of Las Palmas de Gran Canaria (ULPGC), Campus of Tafira, Las Palmas de Gran Canaria 35017, Spain; E-Mails: jalonso@dsc.ulpgc.es (J.-B.A.); ctravieso@dsc.ulpgc.es (C.M.T.)

³ Digital Technologies Group, University of Vic, Sagrada família 7, Vic 08500, Spain; E-Mail: jordi.sole@uvic.cat

⁴ Research Center for Experimental Marine Biology and Biotechnology, Plentzia Marine Station, University of the Basque Country, Plentzia 48620, Spain

⁵ Escola Universitaria Politècnica de Mataró (UPC), Tecnocampus, Mataró, Barcelona 08302, Spain; E-Mail: faundez@tecnocampus.cat

⁶ CITA-Alzheimer Foundation, San Sebastian 20009, Spain; E-Mails: mecay@cita-alzheimer.org (M.E.-T.); pmlage@cita-alzheimer.org (P.M.-L.)

* Author to whom correspondence should be addressed; E-Mail: karmeLe.ipina@ehu.es; Tel.: +34-943-018-667; Fax: +34-943-017-230.

Received: 15 April 2013; in revised form: 8 May 2013 / Accepted: 13 May 2013 /

Published: 21 May 2013

Abstract: The work presented here is part of a larger study to identify novel technologies and biomarkers for early Alzheimer disease (AD) detection and it focuses on evaluating the suitability of a new approach for early AD diagnosis by non-invasive methods. The purpose is to examine in a pilot study the potential of applying intelligent algorithms to speech features obtained from suspected patients in order to contribute to the improvement of diagnosis of AD and its degree of severity. In this sense, Artificial Neural Networks (ANN) have been used for the automatic classification of the two classes (AD and control

subjects). Two human issues have been analyzed for feature selection: Spontaneous Speech and Emotional Response. Not only linear features but also non-linear ones, such as Fractal Dimension, have been explored. The approach is non invasive, low cost and without any side effects. Obtained experimental results were very satisfactory and promising for early diagnosis and classification of AD patients.

Keywords: Alzheimer's disease diagnosis; spontaneous speech; emotion recognition; machine learning; non-invasive diagnostic techniques; dementia

1. Introduction

Alzheimer's disease (AD) is the most common type of dementia among elderly people in Western countries and it has a large socioeconomic cost to society which is expected to increase in the near future. It is characterized by progressive and irreversible cognitive deterioration with memory loss, impaired judgment and language and other cognitive deficits and behavioural 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 to finally become completely dependent. An early and accurate diagnosis of AD would be of much help for patients and their families, both to 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 requires the demonstration of the typical AD pathological changes in brain tissue by autopsy (*post-mortem* analysis) [1–3]. The clinical hallmark and earliest manifestation of AD is episodic memory impairment. At the time of clinical presentation other cognitive deficits are usually already present in their language, executive functions, orientation, perceptual abilities and constructional skills. Associated behavioural and psychological symptoms include apathy, irritability, depression, anxiety, delusions, hallucinations, disinhibition, aggression, aberrant motor behaviour, as well as eating or sleep behaviour changes [1–5]. All these symptoms lead to impaired performance in family, social or professional activities of daily life as the disease progresses from mild to moderate and to severe.

As already mentioned above, the diagnosis of AD is made on clinical grounds and requires, on one hand, the confirmation of a progressive dementia syndrome and, on the other, the exclusion of other potential causes of dementia by clinical history and examination, complete blood workup tests and brain-imaging analysis test, such as computer tomography (CT) or magnetic resonance imaging (MRI). The criteria to exclude other potential causes have changed in the last years as the interpretation of neuroimaging tests, including functional imaging with Single-Photon Emission Computed Tomography (SPECT) and Positron Emission Tomography (PET), has focused on the positive findings of typical AD changes, such as medial temporal atrophy detected by CT or MRI and temporoparietal hypometabolism by PET [6–9].

Nonetheless, the diagnosis of the early stages of not only mild cognitive impairment but also mild dementia remains problematic, since patients and relatives tend to either ignore the first clinical manifestations or ascribe them to the expected cognitive changes related to age. It usually takes 2 to 3 years to seek medical advice after the onset of the symptoms [1–3]. In addition, physicians may feel uncertain or uncomfortable to establish a diagnosis when the whole picture of dementia is not yet fully present; therefore, they usually feel the need to apply long neuropsychological tests, expensive neuroimaging techniques or invasive tests such as a lumbar puncture to reach a diagnosis [10]. It is consequently not surprising that most of the patients are diagnosed when they have already reached the moderate stage of the disease and have become substantially dependent. At this stage, it is very difficult for any treatment strategy to show significant efficacy to stop or even delay the disease process [2–5].

Significant advances have taken place during the last years in the early diagnosis of AD using clinical biomarkers [10], but the currently high cost and technology requirements make it unfeasible to use these diagnostic procedures on any patient displaying only memory complaints. As a result, they are usually applied to pre-selected patients based on their being highly suspect of suffering an underlying AD pathology; who are then apt to have an invasive lumbar puncture or a very expensive PET performed [10–13].

In this setting, the development of non-invasive intelligent diagnosis techniques would 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 perform (after proper training) without altering or blocking the patient's abilities [14–19]. Automatic Spontaneous Speech Analysis (ASSA) and Emotional Response Analysis (ERA) on speech are two of them [15].

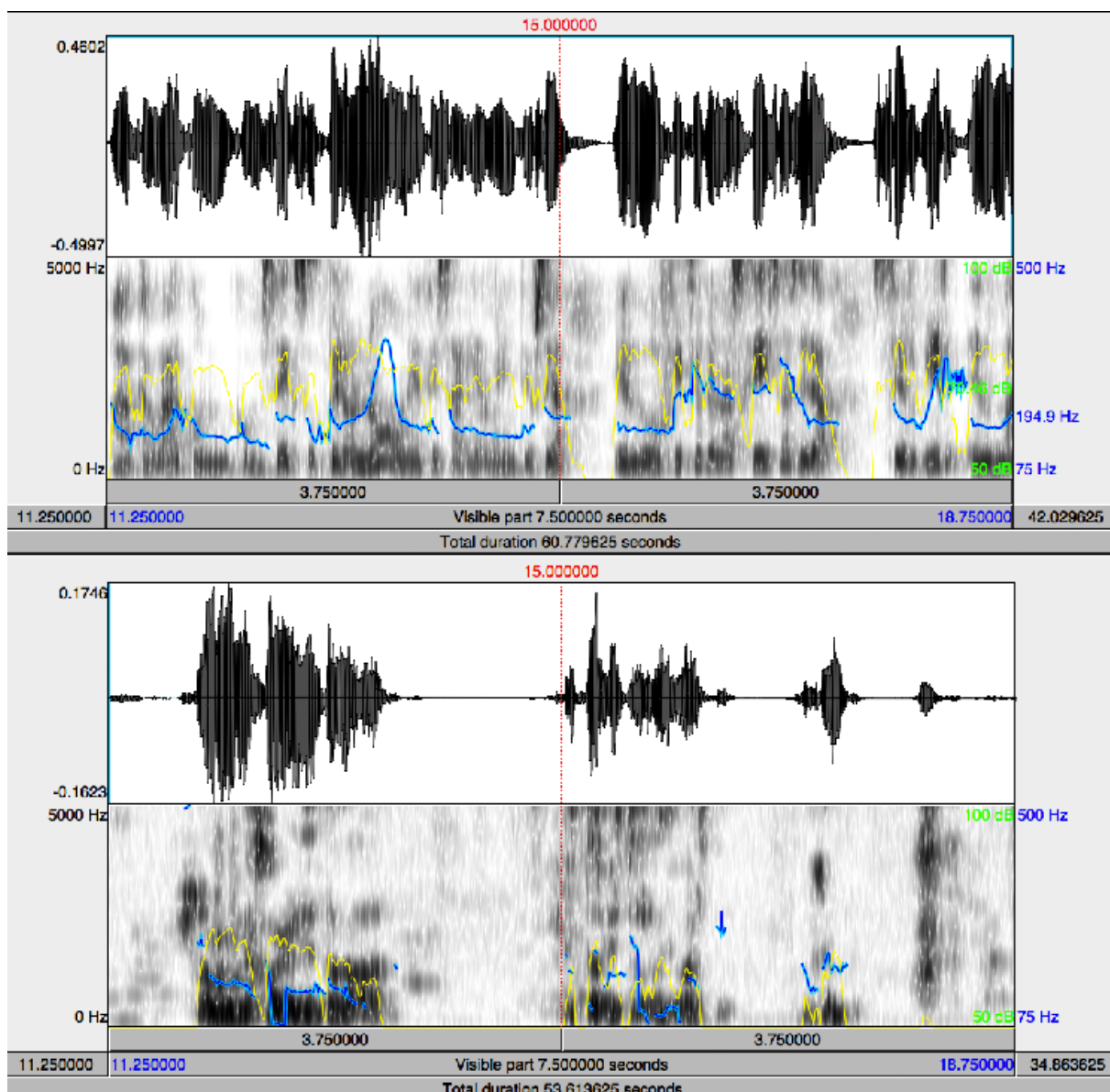
Spoken language is one of the most important elements defining an individual's intellect, his/her social life and personality; it allows us to communicate with each other, share knowledge, and express our cultural and personal identity. Spoken language is the most spontaneous, natural, intuitive and efficient method of communication among people. Therefore, the analysis by automated methods of Spontaneous Speech (SS) or Automatic Speech Analysis (ASSA) which is the freer 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 [15,20–24].

Emotional Response Analysis (ERA) 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, closely linked to learning and understanding that arise in intelligent systems when they become necessary to survive in a changing and partially unpredictable world [25–27]. The nonverbal information, which often includes body-language, attitudes, modulations of voice, facial expressions, etc., is essential in human communication as it has a large effect on the communication provision of the partners and on the intelligibility of speech [25]. Human emotions are affected by the environment, the direct interaction with the outside world but also by the emotional memory emerging from the experience of individual and cultural environment, the so called socialized emotion. Emotions use the same components subjective, cultural, physiological and behavioural that the individual's perception expresses with regard to the mental state, the body and how it interacts with the environment [26,27]. In this work ERA has been analyzed by classical features and by Emotional Temperature, described in

Section 3. This feature is based on the analysis of several prosodic and paralinguistic features sets obtained from a temporal segmentation of the speech signal.

Finally, we wished to apply non-invasive methods to estimate the severity of Alzheimer in the patient. In this sense, analysis of spontaneous speech is not perceived as a stressful test and moreover its cost is lower than for other methods. None of these speech analysis based techniques require extensive infrastructure or the availability of medical equipment, and suggest obtaining by these means is easy, quick and inexpensive [14,15].

Figure 1. Signal and spectrogram of a control subject (**Top**) and an AD subject (**Bottom**) during spontaneous speech (pitch in blue, intensity in yellow).



We have focused our work on non-invasive diagnostic techniques based on the analysis of speech and emotions because after the loss of memory, one of the major problems of AD is the loss of language skills, illustrated by the poorer signal and spectrogram during spontaneous speech of the AD patient shown in Figure 1. This loss is reflected in difficulties both to speak and to understand others,

which makes more difficult the natural communication process with the environment. The inability to communicate appears already in the early phases of the diseases. It is possible to find different communication deficits in the area of language, including [28,29] aphasia (difficulty in speaking and understanding) and anomic aphasia (difficulty for recognizing and naming things). The specific communication problems the patient encounters depend on the stage of the disease [2–5,28,29]:

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.

Not only the language but also the emotional responses in Alzheimer's patients become impaired and seem to go through different stages. In the early stages, social and even sexual disinhibition appears and behavioural changes are also observed (for example, being angry and not being able to perform common tasks, express themselves or remember) [30–33]. However, the emotional memory remains, and they 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 a threat or danger where none exists. In more advanced stages they may often seem shy and apathetic, symptoms often attributed to memory loss and/or difficulty in finding the right words and some 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 [31,33]. The work presented here is part of a larger study to identify novel technologies and biomarkers for early AD detection, and it focuses on evaluating the suitability of a new approach for early AD diagnosis based on non-invasive and low cost methods, namely Automatic Spontaneous Speech Analysis and the Emotional Response Analysis, whose results are susceptible to be used for the automatic classification of tested individuals.

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 built up a multicultural and multilingual (English, French, Spanish, Catalan, Basque, Chinese, Arabian and Portuguese) database with the video recordings of 50 healthy and 20 AD patients (with a previous diagnosis) 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 strictly following the ethical consideration of the organizations involved in the project. The recordings consisted of videos of Spontaneous Speech where people tell pleasant stories or feelings and interact with each other in a friendly conversation. The recording atmosphere is 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

Video has been processed and audio extracted in wav format (16 bits and 16 kHz). Firstly non-analyzable events have been removed: laughter, cough, short hard noises and speaker mixes. Then background noise has been removed by a denoiser adaptive filtering. After the pre-processing, about 80% and 50% of the material from the control and AD groups respectively, is suitable for further analysis. The full database consisted of about 60 minutes for the AD group and of about 9 hours for the control. The speech is divided into consecutive segments of 60 seconds in order to obtain appropriate segments for all speakers. Finally, a database of about 600 segments of spontaneous speech is obtained.

2.3. Individuals Selected for the Study

From the original database, a subset of 20 AD patients (68–96 years of age, 12 women, 8 men, within the three stages of AD: First Stage (ES = 4), Secondary Stage (IS = 10) and Tertiary stage (AS = 6) was chosen. 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.

3. Methods

3.1. Feature Extraction

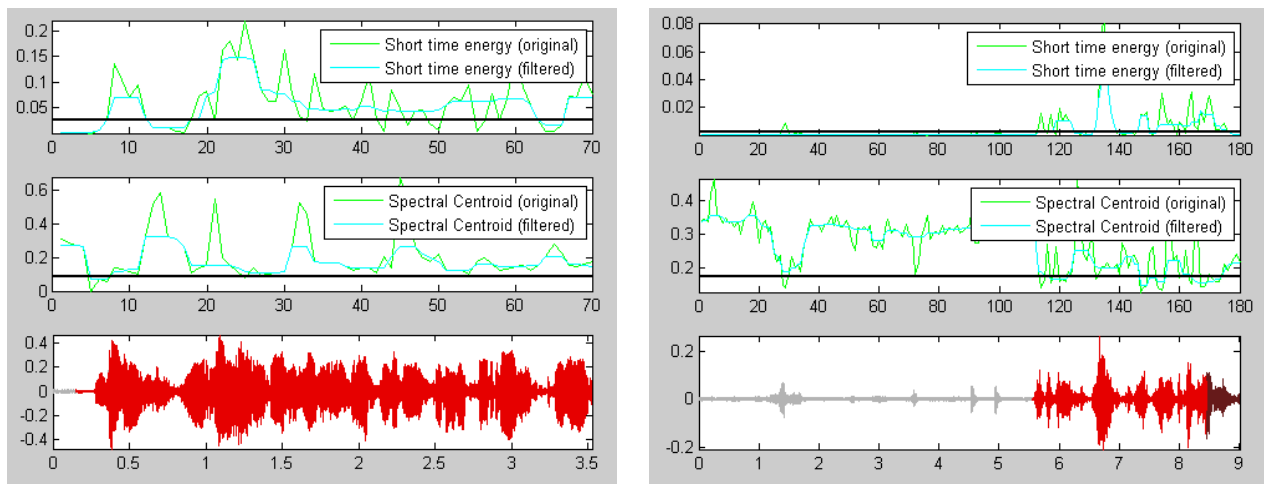
3.1.1. Automatic Spontaneous Speech Analysis (ASSA)

The analysis of spontaneous speech fluency is based on three families of features (SSF set), obtained using Praat software [34]. For that purpose, an automatic Voice Activity Detector (VAD) [35,36] has extracted voiced/unvoiced 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 unvoiced segments, the average of the most relevant voiced/unvoiced, voiced/unvoiced percentage and spontaneous speech evolution along the time, 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 “centre of gravity” using the Fourier transform's frequency and magnitude information (Figure 2).

Figure 2. Plots of speech signal, Short Time Energy and Spectral Centroid, all measures filtered by a median filter, for a control subject (**Left**) and an AD subject (**Right**).



3.1.2. Higuchi Fractal Dimension

When appropriate corpora 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. These latter types of tasks often require experimentation with alternative techniques, which can lead to improved systems. One such alternative technique of particular interest is nonlinear analysis, and some works show that combining nonlinear features with linear ones can produce higher recognition accuracies without substituting the whole linear system with novel nonlinear approaches (see [37,38] for examples on nonlinear speech processing). This is especially promising for solving non-typical tasks, since it would be very demanding to design a complete nonlinear system from scratch for solving a task already made difficult by the scarcity of resources.

The fractal dimension is one of the most popular features, which describe the complexity of a system. Most if not all of the 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. This current work focus on the alternatives, which do not need previous modelling of the system. Higuchi proposed an algorithm for measuring the fractal dimension of discrete time sequences directly from time series, so in our experiments we use the method described in [39] (see Figure 3).

3.1.3. Emotional Speech Analysis (ESA)

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

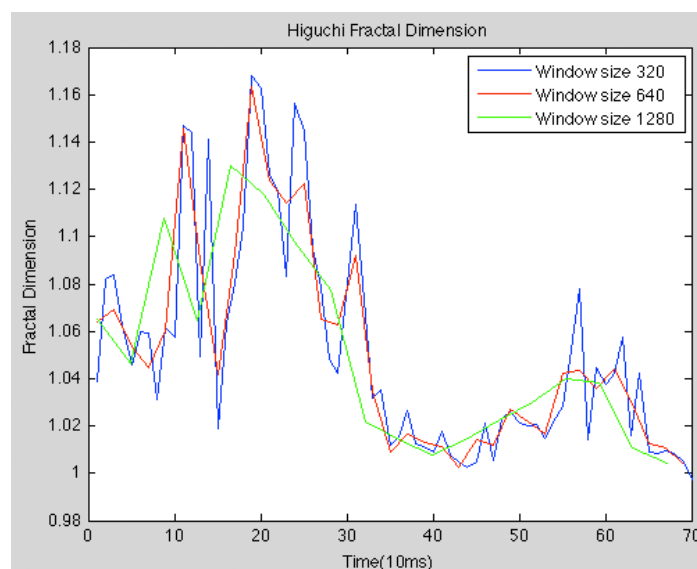
1. *Acoustic features:* pitch, standard deviation of pitch, max and min pitch, intensity, standard deviation of intensity, max and min intensity, period mean, period standard deviation, and Root Mean Square amplitude (RMS);

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 unvoiced frames, degree of voice breaks.

Short-term energy is the principal and most natural feature that has been used. Physically, energy is a measure of how much signal exists at any one time. Energy is used in a continuous speech to discover voiced sounds, which have higher energy than silence/un-voiced, as shown in Figure 2.

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 [36]. The square root of this result is the engineering quantity, known as the root-mean square (RMS) value.

Figure 3. Higuchi Fractal Dimension c of speech signal for an AD subject and different window length.



3.1.4. Emotional Temperature

The Emotional Temperature (ET) is based on the analysis of a few prosodic and paralinguistic features sets obtained from a temporal segmentation of the speech signal [40–42].

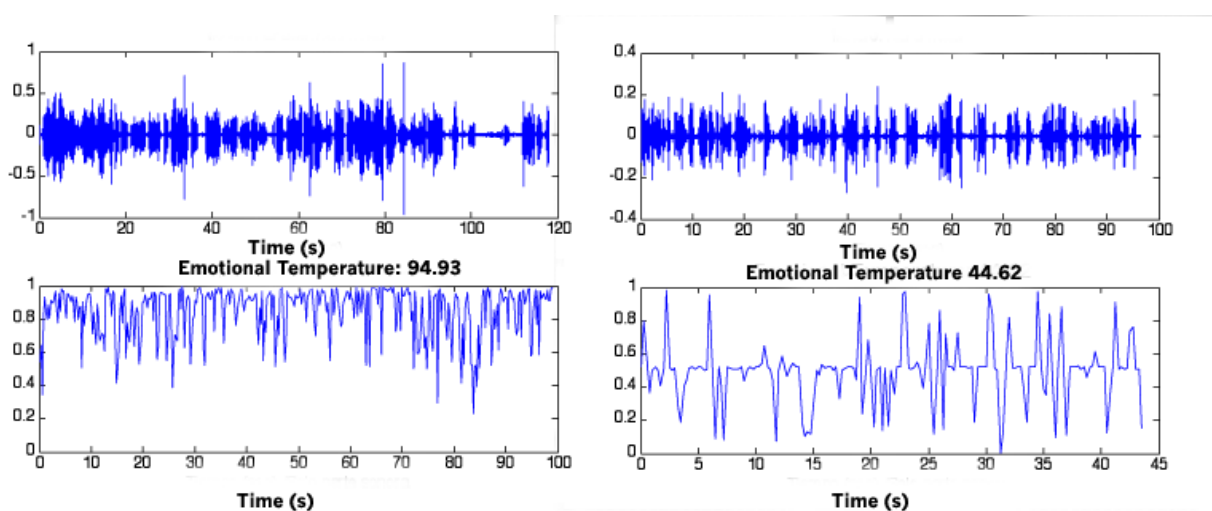
Two prosodic and four paralinguistic features related to the pitch and energy, respectively, were estimated from each frame. These features were chosen because their robustness in emotion recognition has been proven [43–47], they are quickly and easily calculated, and they are independent of linguistic segmentation, which helps us to avoid problems in real time applications on real environments. For prosodic features, a voiced/unvoiced decision is made to each frame and two linear regression coefficients of the pitch contour [43–46] are obtained. For paralinguistic features, voice spectral energy balances [43,47] are calculated from each frame, and quantified using 4 percentages of energy concentration in 4 frequency bands. Then Emotional Temperature is calculated as follows:

1. A Support Vector Machine (SVM) is trained with a balanced segment set extracted from the database (SVMs have been used to quantify the discriminative ability of the proposed measures [45]. We have used a freely available implementation called LIBSVM [48] with a radial basis kernel function).

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 threshold obtained from the training database, which indicates the limit between pathological and non-pathological frames. This normalization will help medical specialists to easily interpret the data.

Figure 4 shows an example of ET values for a healthy subject ($ET = 94.93$) and for an AD subject ($ET = 44.62$).

Figure 4. Emotional Temperature for a healthy subject (**Left**) and an AD subject (**Right**).



3.1.5. Feature Sets

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

1. SSF.
2. EF, set described in Section 3.1.3.
3. FD1: Higuchi Fractal Dimension (HFD).
4. FD2: HFD, maximum HFD, minimum HFD, variance HFD and standard deviation HFD.

3.2. Automatic Classification

The automatic classification of emotional speech is based on the Multi Layer Perceptron (MLP) neural network with one hidden layer of 100 neurons and 1,000 training steps. WEKA [49] software has been used in carrying out the experiments. The results are evaluated using Accuracy (Acc) and Classification Error Rate (CER) measurements. For the training and validation steps, we used k-fold cross-validation in order to ensure solid results. Cross validation is a robust validation for variable selection [50]. These features will define CR group and the three AD levels. The original sample set was randomly divided into k subsets. Then, a single subset was retained as the validation data set for testing the model, and the remaining k-1 subsets were used as training data. The cross-validation

process was repeated k times, with each of the k subsets used exactly once as the validation data set. The k obtained values from the folds, were then averaged to obtain the final result. The advantage of this method is that all observations are used for both training and validation, and each observation is used for validation exactly once. In our experiments we use $k = 10$. These features will discriminate among control group (CR) and the three AD levels.

4. Results and Discussion

The experimentation has been carried out with the balanced subset AZTIAHORE. The goal of these experiments was to examine the potential of selected features for automatic measurement of the degradation of Spontaneous Speech, Emotional Response and their integration in people with AD. Thus, previously defined feature sets have been evaluated in order to properly define control and AD level groups.

In a first stage Emotional Temperature is calculated for each segment by the method described in Section 3.1.4. Automatic classification by MLP was performed over the speech features sets described in Section 3.1.5 in order to analyze for the pilot study the tests: Automatic Spontaneous Speech fluency, Emotional Response in speech and both in Integral Speech. Table 1 summarized Accuracy (%) global results with regard to pre-clinical test and feature sets.

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

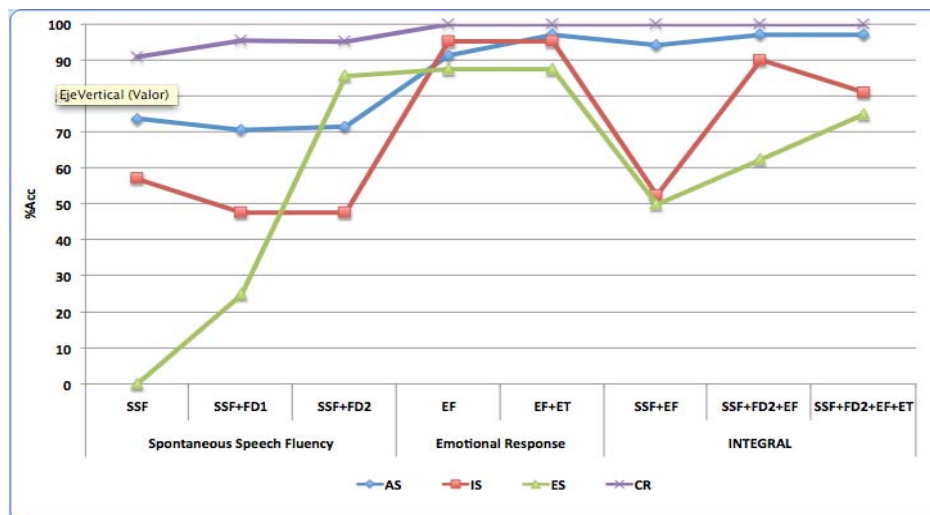
Test	Feature Set	%Acc
Spontaneous Speech Fluency	SSF	75.2
	SSF + FD1	76.7
	SSF + FD2	86.1
Emotional Response	EF	90.7
	EF + TE	97.7
INTEGRAL	SSF + EF	92.2
	SSF + FD2 + EF	94.6
	SSF + FD2 + EF + TE	94.6

- In the first test, *Spontaneous Speech Fluency* test, SSF set alone has been used and also integrated with two different Fractal Dimension (FD) sets. The use of FD features outperforms the results, and the system obtains an improvement of %10 with FD2 set, which includes several FD features.
- In the second test, *Emotional Response* test, the best result is obtained when ET feature is included being Acc near to the optimum value.
- With regard to the *INTEGRAL* test, (which includes information relative to both Spontaneous Speech and Emotional Response) results show also an improvement of non-linear features and better results when ET is included.

Figure 5 shows the obtained results for control group (CR) and the three levels of AD (ES, IS and AS) for classes. In these classes' results the inclusion of non-linear features FD2 set obtains the best results for all classes (Figure 5). This set improves also the classification with regard to early detection (ES class). IS has also better rate to discriminate middle AD level. The model is able also to

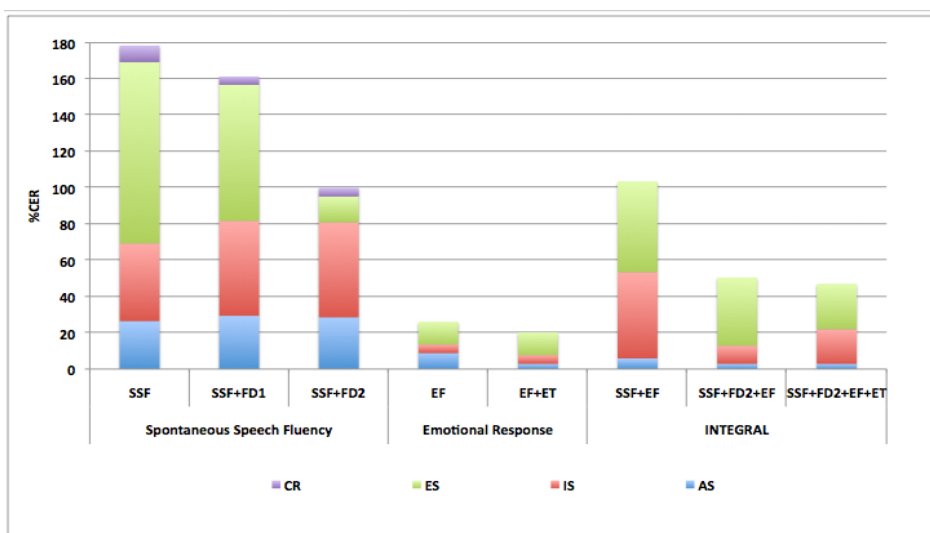
discriminate pathological and non-pathological segments in each patient for the three tests of the pilot study.

Figure 5. Accuracy in % for the three defined tests in the pilot study and each corresponding feature sets.



Accumulative Classification Error Rate (%CER) is detailed in Figure 6. Less accumulative error rate is obtained for experiments which include FD2 non-linear feature set and ET. This is relevant for early diagnosis because in these cases better results are obtained with smaller classification error for ED class.

Figure 6. Accumulative Classification Error Rate (%) for the three defined tests and each corresponding feature sets.

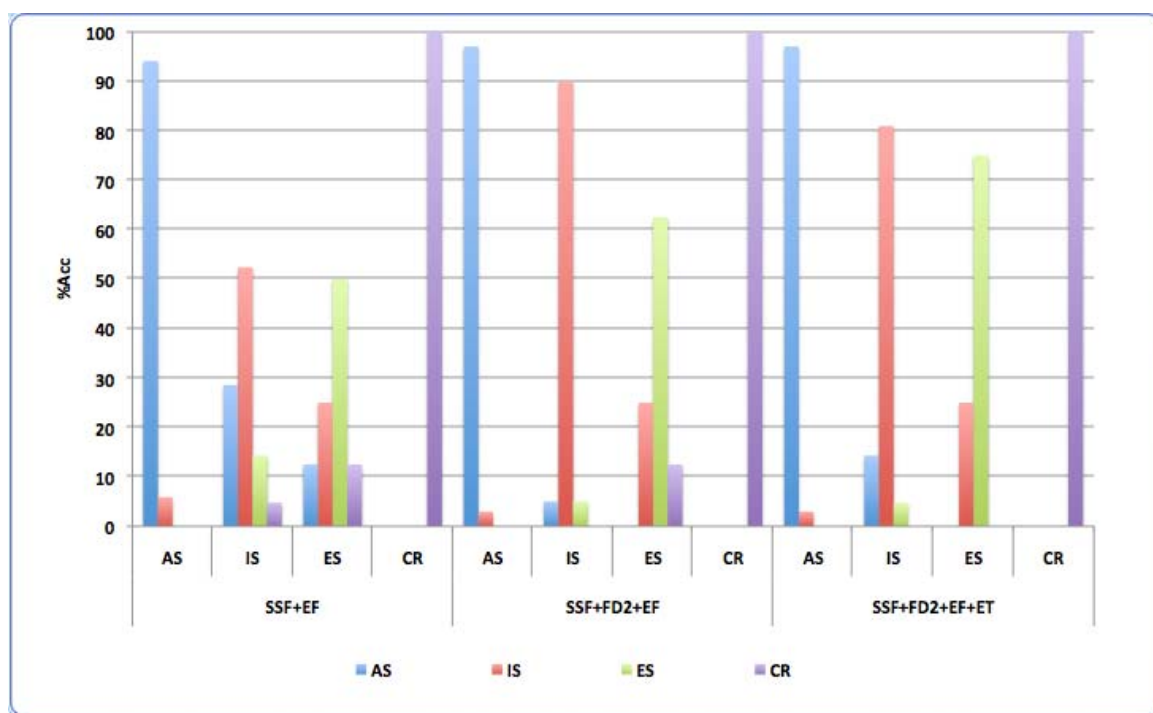


A detailed analysis of *INTEGRAL* test, with regard to Acc (%) for all classes is showed in Figure 7.

- When the results with simple speech features are analyzed, it is observed optimum performance for control group but very mixed classification in AD levels mainly for IS and ES.

- The global system obtains very good results when FD2 set is added because not only outperforms Acc in middle stage IS, but also there is a great improvement in Acc for ES.
- In the third experiment, ET is included and Acc obtains very good results for all classes, and ES outperforms because confusion disappears with regard to the segments of control subjects.

Figure 7. Accuracy (%) of classes for *INTEGRAL* test and each corresponding feature sets.



Finally, the health specialists notice the relevance of the system's ability to carry out both the analysis of independent biomarkers as Spontaneous Speech and Emotional Response features, and/or the integral analysis of several biomarkers. The final confusion among segments could be due, in some cases, to the possible occurrence of segments with different pathological levels in the same individual. This possibility will be explored with new tests in future works. Therefore these non-invasive tests could be a very useful tool for medical specialists' in future clinical AD early diagnosis grounds.

5. Conclusions

The main goal of the present work is the analysis of features in Spontaneous Speech and Emotional Response oriented to pre-clinical evaluation for the definition of appropriate tests for early AD diagnosis. These features are of great relevance for health specialists to define health people and the three AD levels. Features relative to speech duration, time domain, spectral domain and fractal dimension have been analyzed. In this work a first approach including nonlinear features is described. More precisely, an implementation of Higuchi's algorithm in order to add this new feature to the set that feeds the training process of the model. The approach's performance is very satisfactory and promising results for early diagnosis and classification of AD patient groups. Moreover, new features (Higuchi Fractal Dimension and Emotional Temperature) significantly outperform previous results. In

future work we will evaluate this approach with an early diagnosis database and new tests oriented to semantic and memory tasks, and we will also introduce new features relatives to non-linear dynamic.

Acknowledgments

This work has been partially supported by a SAIOTEK from the Basque Government, University of Vic under the research grant R0904, and the Spanish Ministerio de Ciencia e Innovación TEC2012-38630-C04-03. 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.

Conflict of Interest

The authors declare no conflict of interest.

References

1. McKhann, G.; Drachman, D.; Folstein, M.; Katzman, R.; Price, D.; Stadlan, E.M. Clinical diagnosis of Alzheimer's disease: Report of the National Institute on Aging-Alzheimer's Association workgroup on Alzheimer's disease. *Neurology* **1984**, *24*, 939–944.
2. McKhann, G.; Knopman, D.S.; Chertkow, H.; Hyman, B.T.; Jack, C.R., Jr.; Kawas, C.H.; Klunk, W.E.; Koroshetz, W.J.; Manly, J.J.; Mayeux, R.; *et al.* 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. *Alzheimers Dement.* **2011**, *7*, 263–269.
3. Alzheimer's Association. Available online: <http://www.alz.org> (accessed on 15 May 2013).
4. *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed.; American Psychiatric Association: Washington, DC, USA, 2000.
5. Family Caregiver Alliance (FGA), National Center on Caregiving. Alzheimer's Disease. Available online: <http://www.caregiver.org/caregiver/jsp/home.jsp> (accessed on 15 May 2013).
6. Morris, J.C. The Clinical dementia rating (CDR): Current version and scoring rules. *Neurology* **1993**, *43*, 2412b–2414b.
7. Van de Pole, L.A.; Van der Flier, W.M.; Hensel, A.; Gertz, H.J.; Scheltens, P. The effects of age and Alzheimer's disease on hippocampal volumes, a MRI study. *Alzheimer's and Dementia. J. Alzheimer's Assoc.* **2005**, *1*(Supplement 1), doi:10.1016/j.jalz.2005.06.205.
8. Laakso, M.P.; Lehtovirta, M.; Partanen, K.; Riekkinen, P.J., Sr.; Soininen, H. Hippocampus in Alzheimer's disease: A 3-year follow-up MRI study. *Biol. Psychiat.* **2000**, *47*, 557–561.
9. Petrella, J.R.; Coleman, R.; Doraiswamy, P. Neuroimaging and early diagnosis of Alzheimer disease: A look to the future. *Radiology* **2003**, *226*, 315–336.
10. Alzheimer's Association/Biomarkers. Available online: http://www.alz.org/research/funding/global_biomarker_consortium.asp (accessed on 15 May 2013).

11. Wernickand, M.N.; Aarsvold, J.N. *Emission Tomography: The Fundamentals of PET and SPECT*; Elsevier: New York, NY, USA, 2004.
12. 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. Eng. TBE* **2008**, *55*, 1849–1853.
13. Álvarez, I.; Górriz, J.M.; Ramírez, J.; Salas-Gonzalez, D.; López, M.; Segovia, F.; Padilla, P.; Gracia, C. Projecting independent components of SPECT images for computer aided diagnosis of Alzheimer's disease. *Pattern Recognit. Lett.* **2010**, *31*, 1342–1347.
14. Faundez-Zanuy, M.; Hussain, A.; Mekyska, J.; Sesa-Nogueras, E.; Monte-Moreno, E.; Esposito, A.; Chetouani, M.; Garre-Olmo, J.; Abel, A.; Smekal, Z.; *et al.* Biometric applications related to human beings: There is life beyond security. *Cognitive Computation* **2012**, *5*, 136–151.
15. López-de-Ipiña, K.; Alonso, J.B.; Solé-Casals, J.; Barroso, N.; Faundez, M.; Ecay-Torres, M.; Travieso, C.; Ezeiza, A.; Estanga, A. New approaches for Alzheimer's disease diagnosis based on automatic spontaneous speech analysis and emotional temperature. *Ambient Assist. Living Home Care* **2012**, *7657*, 407–414.
16. Hu, W.T.; McMillan, C.; Libon, D.; Leight, S.; Forman, M.; Lee, V.M.Y.; Trojanowski, J.Q.; Grossman, M. Multimodal predictors for Alzheimer disease in nonfluent primary progressive aphasia. *Neurology* **2010**, *75*, 595–602.
17. Grimaldi, G.; Manto, M. Neurological tremor: Sensors, signal processing and emerging applications. *Sensors* **2010**, *10*, 1399–1422.
18. Almeida, A.; López-de-Ipiña, D. Assessing ambiguity of context data in intelligent environments: Towards a more reliable context managing system. *Sensors* **2012**, *12*, 4934–4951.
19. Bravo, J.; Villarreal, V.; Hervás, R.; Urzaiz, G. Using a communication model to collect measurement data through mobile devices. *Sensors* **2012**, *12*, 9253–9272.
20. Horley, K.; Reid, A.; Burnham, D. Emotional prosody perception and production in dementia of the Alzheimer's type. *J. Speech Lang. Hear. Res.* **2010**, *53*, 1132–1146.
21. Pakhomov, S.V.S.; Hemmy, L.S.; Lim, K.O. Automated semantic indices related to cognitive function and rate of cognitive decline. *Neuropsychologia* **2012**, *50*, 2165–2175.
22. Haugrud, N.; Crossley, M.; Vrbancic, M. Clustering and switching strategies during verbal fluency performance differentiate Alzheimer's disease and healthy aging. *J. Speech, Lang. Hear. Res.* **2010**, *53*, 1132–1146.
23. Henry, J.D.; Crawford, J.R.; Pshillips, L.H. Verbal fluency performance in dementia of the Alzheimer's type: A meta-analysis. *Neuropsychologia* **2004**, *42*, 1212–1222.
24. Meilán, J.J.; Martínez-Sánchez, F.; Carro, J.; Sánchez, J.A.; Pérez, E. Acoustic markers associated with impairment in language processing in Alzheimer's Disease. *Span. J. Psychol.* **2012**, *15*, 487–494.
25. Knapp, M.L. *Essentials of Nonverbal Communication*; Holt, Rinehart & Winston: New York, NY, USA, 1980.
26. Cowie, R.; Douglas-Cowie, E.; Tsapatsoulis, N.; Votsis, G.; Kollias, S.; Fellenz, W.; Taylor, J.G. Emotion recognition in human-computer interaction. *IEEE Signal Process. Mag.* **2001**, *18*, 32–80.
27. Plutchik, R. *Emotion: A Psychoevolutionary Synthesis*; Harper and Row: New York, NY, USA, 1980.

28. Buiza, C. *Evaluación y Tratamiento de los Trastornos del Lenguaje*; Matia Fundazioa: Donostia, Spain, 2010.
29. Martinez, F.; Garcia, J.; Perez, E.; Carro, J.; Anara, J.M. Patrones de Prosodia expresiva en pacientes con enfermedad de Alzheimer. *Psicothema* **2012**, *24*, 16–21.
30. 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 Cognit.* **2001**, *47*, 423–433.
31. 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*, 30–37.
32. Cadieux, N.; Greeve, K. Emotion processing in Alzheimer's disease. *J. Int. Neuropsychol. Soc.* **1997**, *3*, 411–419.
33. Henry, J.D.; Rendell, P.G.; Scicluna, A.; Jackson, M.; Phillips, L.H. Emotion experience, expression, and regulation in Alzheimer's disease. *Psychol. Aging* **2009**, *24*, 252–257.
34. Praat: Doing Phonetics by Computer. Available online: www.fon.hum.uva.nl/praat (accessed on 15 May 2013).
35. Voice Activity Detector Algorithm (VAD). Available online: www.mathwork.com (accessed on 15 May 2013).
36. Solé-Casals, J.; Zaiats, P.V. A non-linear VAD for noisy environments. *Cogn. Comput.* **2010**, *2*, 191–198.
37. Solé-Casals, J.; Zaiats, P.V.; Monte-Moreno, E. Non-linear and non-conventional speech processing: Alternative techniques. *Cogn. Comput.* **2010**, *2*, 133–134.
38. Ezeiza, A.; López-de-Ipiña, K.; Hernández, C. Enhancing the feature Extraction Process for Automatic Speech Recognition with Fractal Dimensions. Available online: <http://link.springer.com/content/pdf/10.1007/s12559-012-9165-0.pdf> (accessed on 15 May 2013).
39. Higuchi, T. Approach to an irregular time series on the basis of the fractal theory. *Phys. D* **1988**, *31*, 277–283.
40. Schuller, B.; Batliner, A.; Seppi, D.; Steidl, S.; Vogt, T.; Wagner, J.; Devillers, L.; Vidrascu, L.; Amir, N.; Kessous, L.; *et al.* The Relevance of Feature Type for the Automatic Classification of Emotional User States: Low Level Descriptors and Functionals. In Proceedings of 8th Annual Conference of the International Speech Communication Association (Interspeech'07), Antwerp, Belgium, 27–31 August 2007; pp. 2253–2256.
41. Vogt, T., André, E. Comparing Feature Sets for Acted and Spontaneous Speech in View of Automatic Emotion Recognition. In Proceeding of IEEE International Conference on Multimedia and Expo (ICME'05), Amsterdam, The Netherlands, 5–8 July 2005; pp. 474–477.
42. Pao, T.L.; Chien, C.S.; Yen, J.H.; Chen, Y.T.; Cheng, Y.M. Continuous Tracking of User Emotion in Mandarin Emotional Speech. In Proceedings of 3th International Conference on International Information Hiding and Multimedia Signal Processing (IIH-MSP 2007), Splendor Kaohsiung, Taiwan, 26–28 November 2007; Volume 1, pp. 47–52.
43. Petrushin, V.A. Emotion in Speech: Recognition and Application to Call Centers. In Proceedings of Conference on Artificial Neural Networks in Engineering (ANNIE 1999), St. Louis, MO, USA, 7–10 November 1999; pp. 7–10.

44. Lee, C.M.; Narayanan, S. Emotion Recognition Using a Data-Driven Fuzzy Interference System. In Proceedings of 8th European Conference on Speech Communication and Technology (ECSCT 2003), Geneva, Switzerland, 1–4 September 2003; pp. 157–160.
45. Kwon, O.W.; Chan, K.; Hao, J.; Lee, T.W. Emotion Recognition by Speech Signals. In Proceedings of 8th European Conference on Speech Communication and Technology (ECSCT 2003), Geneva, Switzerland, 1–4 September 2003; pp. 125–128.
46. De Cheveigné, A.; Kawahara, H. YIN, a fundamental frequency estimator for speech and music. *J. Acoust. Soc. Am.* **2002**, *111*, 1917–1930.
47. Alonso, J.; de León, J.; Alonso, I.; Ferrer, M.A. Automatic detection of pathologies in the voice by HOS base parameters. *J. Appl. Signal Process.* **2001**, *4*, 275–284.
48. Chang, C.C.; Lin, C.J. LIBSVM: A Library for Support Vector Machines. Available online: <http://www.csie.ntu.edu.tw/~cjlin/libsvm> (accessed on 15 May 2013).
49. WEKA. Available online: <http://www.cs.waikato.ac.nz/ml/weka/> (accessed on 15 May 2013).
50. Picard, R.; Cook, D. Cross-validation of regression models. *J. Am. Statist. Assoc.* **1984**, *79*, 575–583.

© 2013 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).