Review

Alzheimer's disease and automatic speech analysis: A review

María Luisa Barragán Pulido [a,*], Jesús Bernardino Alonso Hernández [a],
Miguel Ángel Ferrer Ballester [a], Carlos Manuel Travieso González [a], Jiří Mekyska [b],
Zdeněk Smékal [b]

[a] Instituto para el Desarrollo Tecnológico y la Innovación en Comunicaciones (IDeTIC), Universidad de Las Palmas de Gran Canaria, Despacho D-102,
Pabellón B, Ed. de Electrónica y Comunicaciones, Campus de Tafira, 35017, Las Palmas, Spain
[b] Department of Telecommunications, Brno University of Technology, Technicka 10, 61600 Brno, Czech Republic

ABSTRACT

The objective of this paper is to present the state of-the-art relating to automatic speech and voice analysis techniques as applied to the monitoring of patients suffering from Alzheimer's disease as well as to shed light on possible future research topics. This work reviews more than 90 papers in the existing literature and focuses on the main feature extraction techniques and classification methods used. In order to guide researchers interested in working in this area, the most frequently used data repositories are also given. Likewise, it identifies the most clinically relevant results and the current lines developed in the field. Automatic speech analysis, within the Health 4.0 framework, offers the possibility of assessing these patients, without the need for a specific infrastructure, by means of non-invasive, fast and inexpensive techniques as a complement to the current diagnostic methods.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

Alzheimer's disease (AD) is, currently, the most common cause of neurodegenerative dementia in the world. Some 70–76% of the cases of dementia are in developed countries, with increasingly more long-lived people (Valle & Cortés, 2015). Although the etiology of AD is unknown and it is likely to be a disease of multifactorial causes, it is clear that its beginning is insidious and appears at an adult age, mainly generating cognitive and behavioural impairment of the patient (Guix, 2011). The damage produced is progressive and irreversible; there is no cure and it involves, in all cases, impairment and neural death (Alzheimer's Association, 2016c). Memory loss appears as one of the first symptoms, which is added to difficulties with language usage, the capacity to conduct daily activities or even, in more advanced stages, difficulties in conducting basic bodily functions, such as walking or swallowing (Escobar & et al., 2010). Finally, the patient completely loses his autonomy and requires the constant attention of a carer or relative. This fact, in addition to the slow development of the disease (usually around 8–10 years from the beginning to death (Escobar & et al., 2010; Alzheimer's Association, 2016a), which is generally caused by another intercurrent illness) turns AD into a fatal disease. Some studies consider AD as one of the diseases with the greatest of social and economic consequences, both for the social burden (including on public health) and the damage produced on patients and their families (Alzheimer's Association, 2016b, 2017). In 2050 the number of people with dementia, the clinical syndrome of AD, will triple the figures from 2010 (Prince et al., 2013).

In relation to the causes of the rise of the prevalence AD, it should be remarked that the inversion of the generational pyramid and life expectancy increase are the dominating factors (Alzheimer's Association, 2016a). One of the main risk factors is age and, due to a greater life expectancy, prevalence is higher in women (Hebert, Scherr, McCann, Beckett, & Evans, 2001; Seshadri, Wolf, & et al., 1997). From 65 years old, the risk of suffering AD doubles every 5 years and it is estimated that one in every two people older than 80 years suffers incipient AD or it is already declared (Ferri et al., 2005). Because of medical assistance, the advances and the social and environmental conditions, the number of older people between 80–90 years in western countries is expected to grow substantially.

AD has become one of the most expensive chronic diseases in society (Hurd, 2013). In comparison with other diseases, the costs entailed for people suffering from AD and other forms of dementia are more than double those of patients of the same age and suffering from cancer, and 74 % higher than patients

* Corresponding author.
E-mail addresses: mbarragabl@idetic.eu (M.L.B. Pulido), jesus.alonso@ulpgc.es (J.B.A. Hernández), miguelangel.ferrer@ulpgc.es (M.Á.F. Ballester), carlos.travieso@ulpgc.es (C.M.T. González), mekyska@feec.vutbr.cz (J. Mekyska), smekal@feec.vutbr.cz (Z. Smékal).

Glossary

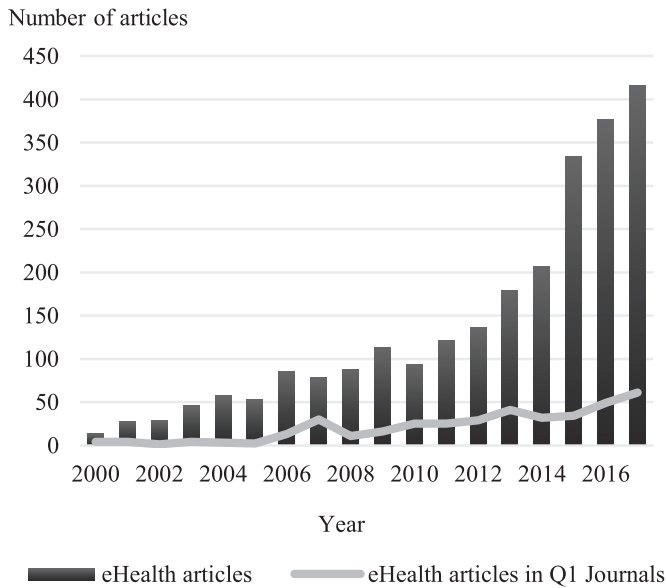|	
AD	Alzheimer's Disease
AB	AdaBoost M1
AE	Approximate Entropy
ALS	Amyotrophic Lateral Sclerosis
ANN	Artificial Neural Network
APOE	Apolipoprotein
ASSA	Automatic Spontaneous Speech Analysis
BN	Bayesian Network
CD	Correlation Dimension
CE	Correlation Entropy
CLAN	Computer Language Analysis
CNN	Convolutional Neural Network
CSF	Cerebrospinal Fluid
CT	Computerized Tomography
cVF	Verbal Fluency categories
DAT	Dementia of Alzheimer Type
DFA	Detrended Fluctuation Analysis
DL	Deep Learning
EEG	Electroencephalography
ERA	Emotional Response Analysis
ESA	Emotional Speech Analysis
ET	Emotional Temperature
EU	European Union
F0	Fundamental Frequency
FD	Fractal Dimension
FMMI	First Minimum of Mutual Information
GCNN	Gated Convolutional Neural Network
GDP	Gross Domestic Product
HC	Healthy Control
HE	Hurst Exponent
HNR	Harmonic to Noise Rate
ICU	Information Content Unit
kNN	k Nearest Neighbours
LDA	Linear Discriminant Analysis
LLE	Largest Lyapunov Exponent
LPC	Linear Prediction Coefficient
MB	Meta Bagging
MCI	Mild Cognitive Impairment
MEG	Magnetoencefalography
MFCC	Mel Frequency Cepstral Component
MFDFA	Multifractal Disactivated Fluctuation Analysis
ML	Machine Learning
MLP	Multilayer Perceptron
MMSE	Mini Mental State Examination
MRI	Magnetic Resonance Image
MSE	Mean Square Error
NB	Naive Bayes
NHR	Noise to Harmonic Rate
NLP	Natural Language Production
NN	Neural Network
PD	Parkinson's Disease
PET	Positron Emission Tomography
PID	Propositional Idea Density
PPA	Primary Progressive Aphasia
RBE1	First order Rényi Block Entropy
RBE2	Second order Rényi Block Entropy
RE	Second order Rényi Entropy
RF	Random Forest
RNN	Recurrent Neural Network
RPDE	Recurrence Probability Density Entropy
RST	Rethoric Structure Theory
SD	Standard Deviation
SE	Sample Entropy
SEO	Square Energy Operator
SHE	Shannon Entropy
SMO	Sequential Minimal Optimization
SPECT	Simple Photon Emission Computed Tomography
SS	Spontaneous Speech
SVM	Support Vector Machine
TEO	Teager Kaiser Energy Operator
UK	United Kingdom
USA	United State of America
VAD	Voice Activity Detection
VR	Variation Range
ZL	Ziv Lempel Complexity

with cardiovascular diseases (Friedman, Shih, Langa, & Hurd, 2015) (Plassman, Langa, & et al., 2007). Besides the high costs of diagnosis and pharmacological costs (approximately 1 % of the total costs of the disease (Winblad, Amouyel, Andrieu, & Ballard, 2016)), costs of care for people with dementia are high and constitute the major expense. In most cases, it is a significant financial burden for both families and patients (Atance, Yusta, & Grupeli, 2004). Today, taking care of an AD patient or another dementia patient presents difficult challenges, especially when the personality and behaviour of the person are affected (Schulz & Beach, 1999). According to an American study, the care of a person suffering from dementia requires, on average, 27 h/week more than the care of people without it (Freedman & Spillman, 2014; Caregiving in the U.S., 2015).

Technological developments provide some means of support to people with dementia and their carers. These mainly aim to improve the safety aspects of these people by means of devices such as smoke detectors or panic buttons, enhancing capacities of memory or the developing of global positioning systems (Lorenz, Freddolino, Comas-herrera, Knapp, & Damant, 2017). Other tools are focused on the treatment and therapeutic issues, for example, through videophones and online assistance (Alzheimer's Association, 2016c; Boots, Vugt, Knippenberg, Kempen, & Verhey, 2014; Czaja, Loewenstein, Schulz, Nair, & Perdomo, 2013) or the usage of video-games in order to physically exercise patients (exergames) (Lorenz et al., 2017; Stanmore, Stubbs, Vancampfort, Bruin, & Firth, 2017). Major companies such as IBM are already working on this kind of solution and have introduced tools like SimpleC (IBM Watson, 2016). Other examples are ComputerLink, AlzOnline and Caring-for-Others (Powell, Chiu, & Eysenbach, 2008).

Using tablets or other tools and apps as AD therapy seems to be an effective resource with several practical implications (Lancioni et al., 2017). According to numerous studies, which have performed a systematic review of their effectiveness, some interventions have positive results, thus decreasing the stress and pressure on the carer. They also show that the results varied with their individual characteristics (ethnic groups, formal support received or basic workload). Nevertheless, more studies are required to precisely define what interventions are the most effective in specific situations (Boots et al., 2014; Powell et al., 2008). Using new technologies could help to find a solution not only for patients and carers but also in searching for non-invasive, economic, fast and easily-applied methods (Laske et al., 2015).

The number of publications produced in recent years on eHealth and Telecare, Health 4.0, has increased remarkably. In Fig. 1, we illustrate the number of publications in this area, distinguishing those published in Q1 journals.

Currently, some large-scale international initiatives have been established, including the Joint Programme of the European Union for Neurodegenerative Disease Research (JPND), the G8 Summit on Dementia (G8 Health and Science Ministers, 2013), the European

Number of articles

Fig. 1. Q1 articles on eHealth and Telecare published from 2000 to 2017. This is based on publications identified in this paper.

Union Innovative Medicines Initiatives (IMI) and the Organization for Economic Co-operation and Development for the collecting of Big Data on AD research (OECD, 2015). Several countries have already initiated large trials of multifactorial intervention based on lifestyle. Some of these completed or ongoing trials are FINGER (Kivipelto, Solomon, & Ahtiluoto, 2013), MAPT (Carrié, van Kan GA, Gillette-Guyonnet, 2012), PreDIVA (Richard, Van den Heuvel E, & Moll van Charante, 2009) and HATICE (Richard, Jongstra, & Soininen, 2015).

Today there is a clear rise in preventive cohorts and an important shift in AD research, which more than ever is focused on identifying the preclinical stages of the disease. The importance of having sufficiently sensitive tools to detect very subtle changes in the preclinical population is crucial to being able to react before damage becomes worse, since current treatments are more effective at these stages rather than later. The European Prevention of Alzheimer's Dementia Consortium (EPAD | European Prevention of Alzheimer's Dementia Consortium," 2015) is an interdisciplinary research programme spanning public and private sector organisations across Europe. EPAD, as with other researchers in the field such as the Alzheimer's Prevention Initiative (API) (Reiman et al., 2011), is working towards understanding more about AD and especially what is possible to do to prevent and treat it. Developing strategies in order to delay AD appearance or decelerate its progression and identifying effective preventive and therapeutic measures able to be applied in different contexts, are the primary objectives (Winblad et al., 2016).

2. Alzheimer's disease

Alzheimer's disease covers the complete process of the initial pathological changes in the brain before its symptoms appear through dementia (Alzheimer's Association, 2016c). This means that AD includes, not only those patients with dementia caused by the illness but also those suffering Mild Cognitive Impairment (MCI) and asymptomatic people who have demonstrated positive AD biomarkers. Brain changes occur when neurons, in specific parts of the brain involving the cognitive functions, are damaged or destroyed. The speed at which symptoms advance varies with the person (Alzheimer's Association, 2016a), and it is faster when AD is diagnosed at early ages.

The most common, earliest symptom of dementia is difficulty in remembering recent events. As the disorder develops, other symptoms appear such as disorientation or mood swings. When AD advances, those changes become more substantial. Difficulties in speaking and communicating appear and motor skills decrease. The progressive accumulation of disability with impairment in a multitude of cognitive domains significantly affects the everyday functioning of the patient, including his/her social and familial development (Alzheimer's Association, 2016a; Winblad et al., 2016). Initially, cognitive changes produced are very subtle, unclear and difficult to detect and differentiate from other pathologies (Alzheimer's Association, 2016a). With AD progression other kinds of underlying illnesses appear and, as a general rule, cause the patient's death (Alzheimer's Association, 2016a). Although currently the exact cause of person developing AD is unknown, it is reported that some of the risk factors include advancing age and the transport of one or two alleles APOE ε4 (Winblad et al., 2016).

Until now, the diagnosis of AD patients (McKhann G, Drachman D, Folstein M, Katzman R & EM., 1984) has been, almost entirely, based on clinical examination, including post-mortem neuropathological confirmation, which is only occasionally performed (Khachaturian, 1985). Currently, physicians are limited to diagnosing likely AD and to ruling out other potential causes of cognitive impairment (Snyder et al., 2011). Nowadays, thanks to the advent of advanced technologies of neuroimaging, such as Magnetic Resonance Imaging (MRI) or Positron Emission Tomography (PET), a wider understanding of the neuropathological processes of patients has been achieved. In addition to neuroimaging and medical examination, there are different neuropsychological tests, analysis of Cerebrospinal Fluid (CSF) and blood analysis (Laske et al., 2015). However, the diagnostic process continues to be complex and it is inevitably performed at advanced phases of the disease (Guix, 2011), is prolonged in time and also requires a specialist physician. In this sense, some works (Connolly, Gaehl, Martin, Morris, & Purandare, 2011) show a lack of specialist doctors and detect in non-specialists some limitations in the ability to identify precisely early AD and MCI (Laske et al., 2015; Connolly et al., 2011). Currently, CSF analysis and imaging methods are the main and most relevant clinical diagnostics (Guix, 2011). The fact that these methods are expensive and invasive limits their use as a screening tool (Laske et al., 2015) and keeps alive the interest in searching for biomarkers located in more accessible parts of the body and AD sensitive before the clinical onset of dementia (Morris, 2005).

Using biomarkers as a screening method presupposes an economic solution for the early diagnosis of preclinical AD (Bateman et al., 2012) and the identification and monitoring of the specific AD stage (Hane, Robinson, Lee, Bai, & Leonenko, 2017). Today, this field constitutes an active research area (Hane et al., 2017) and some biomarkers which occur in the blood are being especially studied. Unfortunately, no blood biomarker has yet attained sufficient specificity or sensitivity (Hane et al., 2017). Especially interesting are clinical tests based on biomarkers obtained from the subjective assessment of memory, late-life depression or speech, olfactory and gait analyses. Currently, different neurophysiological tests based on electroencephalographies (EEG) and magnetoencephalographies (MEG) are also under-study (Laske et al., 2015), although until now, there are no conclusive results.

Nowadays, an AD patient must present appreciable cognitive impairment or concomitant functional deficits to be diagnosed and, as a consequence, to be considered for pharmacological intervention (Snyder et al., 2011). By the time of diagnosis, damage in the cerebral tissue and irreversible losses have inevitably occurred (Jack et al., 2010; Jack &Wiste et al., 2010). Although some medications are marketed (Lopez et al., 2010), they do not cure the disease (Jack and Petersen et al., 2008; Lopez et al., 2010; Petersen et al., 2005). The majority of the clinical trials with

new drugs seem to confirm that neuroprotective therapies applied at the first stages lead to better clinical results (Cummings & Doody, 2007). Today, drugs used for AD treatment improve temporarily the symptoms although their effectiveness varies with the person. Some of the factors that inhibit the development of more effective treatments are the high cost of developing and the relatively long time needed to observe whether a treatment under study genuinely affects disease progression (Alzheimer's Association, 2016c).

3. Speech problems in Alzheimer's disease

Language problems are considered one of the most characteristic symptoms of AD, which appear as a direct and inevitable consequence of cognitive impairment (Gonnerman, Aronoff, Almor, Kempler, & Andersen, 2004). Progressive Primary Aphasia (PPA), the clinical syndrome characterised by a progressive impairment of the language whose aetiology is neurodegenerative, could be the presentation mode of different neurodegenerative disorders such as AD (Matias-Guiu & García-Ramos, 2013).

As demonstrated, language and communication problems are evident during interactions between patients and neurologists, and the interactional comments could be used to differentiate cognitive difficulties due to neurodegenerative disorders or those disorders of functional memory (Mirheidari & Blackburn, 2017). Studies demonstrating the deficient performance of AD patients in different linguistic tests are many (Weintraub, Wicklund, & Salmon, 2012) and it has been possible to prove that people with AD present more difficulties with denomination tests, they repeat more frequently the same ideas, use longer and empty pauses and use simpler language structures. Moreover, they also present a monotonous prosody and a less informative, coherent and cohesive speech than other control groups (López, 2016).

The specific communicative problems, such as aphasia and anomy, depend on the stage of the disease (Laske et al., 2015) and increase with AD progression (Meilan, Martinez-Sanchez, Carro, Carcavilla, & Ivanova, 2018; Nebes, Brady, & Huff, 1989). Years before establishing clinical diagnosis, language often shows significant cognitive impairment in preclinical patients (Bäckman, Jones, Berger, Laukka, & Small, 2005; Deramecourt et al., 2010). Specifically, some resources state that the first year after the onset of the disease, different language skills appear obscured by a loss of interest and spontaneity, spatial disorientation, and memory disorders (Allison, 1962; Lenguaje, La, & Alzheimer, 1988; Sjogren, Sjogren, & Lindgren, 1952). Although the onset of AD affects verbal fluency, this is usually not detected. The emotional response capacity is also affected and there are often social and behavioural changes (McKhann et al., 2011), which could be due to memory loss. Likewise, the alteration of perception abilities could also magnify some emotional responses (Laske et al., 2015).

On the basis of the above, some studies claim that AD can be more sensitively detected by means of a linguistic analysis than through other cognitive assessments. However, the question that arises is what language and speech characteristics could identify the disease in its initial stage? (Szatloczki, 2015). Emotions are the main characteristics of human beings. They affect the perception and daily life, for example, the communicative process and decision-making. Emotions are expressed through speech, facial expressions, gestures and other non-verbal clues. Differences between emotional states can also be considered as one of the most important evaluation criteria to measure the performance of cognition processes (Frijda, 1987; Spence, 1995).

In order to identify and quantify the cognitive impairment degree of an AD patient, a multitude of linguistic tests has been used (Maresova, Valis, Hort, & Kuca, 2015). Some of these have evaluated identification, comprehension, repetition and reading skills.

Spontaneous Speech (SS) has been widely assessed through describing images, videos or conversations. According to some studies (Forbes-McKay & Venneri, 2005), solely from oral description tasks, it is possible to distinguish between AD patients and healthy controls as well as being able to identify other cognitive aspects.

Standardized tests such as the Token Test (Renzi & Neurology, 1962, 1962) (semantic processing), the Boston Test (Kaplan, Goodglass, & Weintraub, 2001) (semantic processing and naming tests) and the Verbal/Naming Fluency Test (grouping and changes of categories) have been defined. There are also batteries including different linguistic and non-linguistic tests (spatial and temporal orientation, attention, memory, drawing and writing) (Cuetos-Vega et al., 2007), such as the Mini Mental State Examination of Folstein (MMSE), the most extended and used in clinical practice (Cuetos-Vega et al., 2007), or the Montreal Cognitive Assessment (MoCA).

The most traditional and extended lines on language and AD have been for years focused on these neuropsychological examinations, manually identifying and assessing the most outstanding and early cognitive, behavioural and communicative symptoms (Barbarotto, Capitani, Jori, Laiacona, & Molinari, 1998; Hoffmann, Nemeth, Dye & Pákáski, Irinyi, & Kálmán, 2010). The fact that they are easily applied, fast and sensitive has caused the development of numerous tests and the loss of any kind of consensus establishing general clinical examinations (Valle & Cortés, 2015).
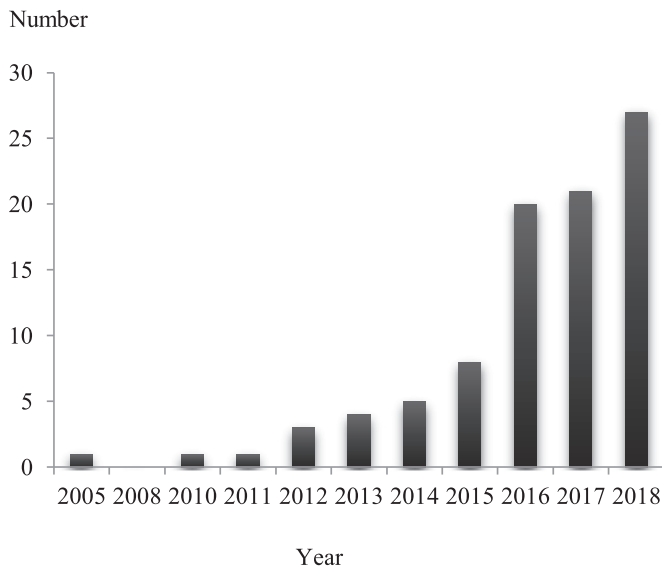
Currently, it seems necessary to define more automatic and objective methods (Wesnes, 2002). These would be directed towards not only automating the methods previously mentioned, which is certainly useful for clinical practice, but also advancing towards a "relatively" new concept of linguistic assessment by using signal processing. Based on previous methods, but from an engineering approach, signal processing offers the possibility to statistically, mathematically and automatically analyse linguistic properties and provide information that it would otherwise be impossible to be manually retrieved and quantified by a human operator.

4. Alzheimer's disease and voice processing

In recent years, techniques based on automatic processing of the voice signal from its record have found an important place in language assessment, as applied to detection of neurodegenerative disorders. These offer the possibility of quantifying signal properties, which are relevant for the description of specific pathologies. Supported by Machine Learning methods, the process of sample classification according to the previous results obtained is subsequently carried out.

The most used and conventional features of AD detection from the voice signal are linear, which are clinically the most easily interpretable. For their part, other current and innovative methods have applied non-linear features. Both kinds of features are important indicators of the expressive structures of the language and depend on the performed tasks (Brabenec, Mekyska, Galaz, & Rektorova, 2017). Apart from the feature extraction process, where there are statistical analysis and/or mathematical models combined (Brabenec et al., 2017) to quantify speech properties, there is a classification process based on the use of Machine Learning techniques. The classification is divided in turn into three main phases: training, testing and validation. Until now, different kinds of classifiers have been designed to perform this function. Some of these are Support Vector Machines (SVM), k-Nearest Neighbours (kNN) and Artificial Neural Networks (ANN), each with its specific design characteristics and operation. In the same context, the Deep Learning concept appears as an automatic learning method, more complex than the previous ones and able to make headway on automatic AD detection from the voice signal.

Number

Fig. 2. Published articles on voice and speech automatic processing applied to Alzheimer's detection from 2005 to early 2018. This is based on the publications found in this paper.

Due to the development of these automatic methods, which process the recording of the voice, and which might facilitate the evolving control of a wide population suffering AD, in recent years, the number of researches in this regard has increased. Their objective is to introduce speech or voice analysis as a non-invasive AD biomarker (Lopez-de-Ipiña et al., 2012) and although by now they have not been included in the clinical practice, the trend shows changes since some companies such as IBM Watson (IBM Watson, 2016), Cantab (Cambridge Cognition, 2019), Winterlights Lab ("Winterlight Labs," 2018) and lately 'ki-elements' (ki-elements, 2018) already offer solutions for cognitive assessment via automatic speech analysis. Some like Winterlights Lab have even developed a two-foot tall robot to this end (Veritone, 2019).

4.1. Methodology

For a more in-depth analysis of the state-of-the-art, a systematic search was carried out through the bibliographical search tools SpringerLink, Scopus, ScienceDirect and Google Academic. The initial criteria used were designed to narrow the search from wider concepts to those more specific, by using "dementia" and "Alzheimer" in a first step; and from here, to delimit the result by successively filtering by "Alzheimer and processing", "Alzheimer and processing and automatic" and finally by "Alzheimer and processing and automatic and voice or speech". These concepts, both in the title and/or in the abstract of the papers, should appear, at least. Apart from making different combinations of them all, to avoid losing any paper in the process or adding some other out of interest, all outputs were manually examined.

As a first approach, the 91 publications found were classified by the year of publication (Fig. 2) and by the kind of feature extraction used in the voice processing (see Sections 4.2 and 4.3).

In spite of the studies located being very heterogeneous, both in terms of the populations that are considered and the tasks they perform to automatically assess impairments, the present review, specifically focused on AD, has not taken into account a more in-depth examination, in, e.g., a detailed classification of MCI subtypes. In this sense, as a first approach, different subtypes, which might have specific impairments (across Healthy Control (HC), AD and MCI groups), will not be analysed in this paper. Regarding the tasks and features obtained from them, this review will not go

deeper to analyse the different meanings of the same feature obtained from two oral tasks (in, e.g., the meaning as a pause made in a SS task has not necessarily the same meaning that a pause made in a picture description task).

In order to analyse the different feature extraction methods, we decided to use a conventional and non-conventional classification instead of the linear and non-linear one. This classification could be confusing since there are techniques, which are mathematically non-linear, but describe linear phenomena. So, in this paper, we will refer to the latter as conventional features, regardless of the mathematical technique used, and, for its part, non-conventional features will be in general those characterising non-linear voice phenomena.

4.2. Conventional feature extraction

Some of the conventional features employed to describe acoustic characteristics of the voice applied to AD detection are shown in Table 1.

In addition to these features, there is also a range of elements properly accompanying linguistic emissions and which constitute signs and clues but are not verbal. These characteristics contextualise and suggest particular interpretations of the proper linguistic information. They are called paralinguistic features and have been used to obtain information from the patient by means of the statistics of the first formants, the concentration of energy in different frequency bands, the Linear Prediction Coefficients (LPC), the Linear Prediction Cepstral Coefficients (LPCC), Log Frequency Power Coefficients (LFPC) or Cepstral Mel-Frequency Components (MFCC), among others (Alonso, Cabrera, Medina, & Travieso, 2015).

It is possible to observe that the majority of the published studies are made on the basis of simple linear features, especially on vocal and prosodic characteristics (Ahangar, Morteza, Fadaki, & Sehhati, 2018; Baldas, Lampiris, Capsalis, & Koutsouris, 2011; Beltrami et al., 2016; Kato, Homma, & Sakuma, 2018; Khodabakhsh & Demiroglu, 2015; König, Satt, David, & Robert, 2016; König et al., 2015; Martínez-Sánchez et al., 2017; Meilan et al., 2018; Meilán, Martínez-Sánchez, Carro, Sánchez, & Pérez, 2012; Mirzaei et al., 2017; Nasrolahzadeh, Mohammadpoory, & Haddadnia, 2016; Rentoumi et al., 2017; Satt, Hoory, König, Aalten, & Robert, 2014; Satt et al., 2013; Tanaka, Adachi, Ukita, Kudo, & Nakamura, 2016a, 2016b; Toth et al., 2018; Tröger, Linz, Alexandersson, König, & Robert, 2017). In general, these studies all use individual features to evaluate the speech signal. Their results have been very varied, rising in accuracy from 70 to 90 %. They have provided clear evidence on the potential of using simple spoken tasks to automatically assess very early dementia and its progression as well as to demonstrate that technology allows not only automatic detection but also conducting it in an economic and remote way and on a large-scale (König et al., 2015; Satt et al., 2014). Some of these studies have even included computerised avatars (Tanaka et al., 2016b, 2016a).

Each study has tried to shed light on different aspects. That is the case of a study based on a stepwise regression linear technique which by using variables such as amplitude, frequency and periodicity, demonstrated that the percentage of unvoiced segments explains a significant part of the variance of the total scores obtained from various neuropsychological tests (Meilán et al., 2012). There are others focused on demonstrating that results change depending on the specific task being developed. One of them (König et al., 2016) implements a mobile application which analyses vocal features of the subjects and demonstrates better results on spontaneous speech and fluency tasks to discriminate between HC, MCI and AD subjects (König et al., 2016). On the same topic, but performing combined tasks (SS and counting), other works

Table 1

Main conventional features used in Alzheimer's disease research (Brabenec et al., 2017; Gómez-Vilda, On, & 2014, 2014; López-de-Ipiña, Alonso, et al., 2015; López-de-Ipiña et al., 2012, 2015; Meilán et al., 2014).

	Interruptions	Percentage and number of pauses of voice$$$Percentage without voice
		Number and percentage of voice breaks
		(MEAN, MAX, MIN, %)
	Voice periods	Number of pulses analysed as voice
		Period (MEAN, SD, Mean number of periods of voice)
Frequential Aspects	Fundamental frequency	F0 (MEAN, MAX, MIN, SD, relF0SD, VR, relF0VR)
		Short time energy or spectral centroid
		High and low global pitch
		Autocorrelation
	Fluctuations	Jitter (Short term, cycle-to-cycle, perturbations in the F0): local jitter, local absolute jitter, relative average perturbation jitter, PPQ.
Intensity	Amplitude	Intensity of voice and unvoiced signals (SD, MAX, MIN)
		SEO (SD, relSEOVR)
		TEO (SD, relTEOVR)
		RMSA
	Phonatory stability	Period perturbation, shimmer (short term, cycle-to-cycle, perturbations in the amplitude of the voice): local shimmer, amplitude perturbation quotient (APQ).
Voice Quality	Noise	HNR
		NHR
Biomechanical aspects	Vocal fold body Movement	Stiffness
	Tongue movement	Using frequency information, as the location of formants, and bandwidths. Although widely known, including this features in speech analysis of Alzheimer's disease if it isn't still extended.

SD: Standard Deviation, VR: Variation Range, SEO: Square Energy Operator, TEO: Teager-Kaiser Energy Operator, HNR: Harmonic-to-Noise Ratio, NHR: Noise-to Harmonic Ratio, F0: Fundamental Frequency, RMSA: Root Mean Square Amplitude.

based on vocal feature analysis reach an accuracy of around 89 % (Tröger et al., 2017).

There are several studies jointly assessing the extraction feature process from the voice signal by means of different vocal parameters and the application of different classifiers (Beltrami et al., 2016; König et al., 2015; Mirzaei et al., 2017; Satt et al., 2014; Toth et al., 2018). Techniques of regularisation to overcome the data shortage have also been used (König et al., 2015; Satt et al., 2014). Through using the Mann–Whitney *U* Test for the selection of vocal characteristics, and by means of a SVM classifier, this study obtained results which discriminated between HC and MCI with accuracy around 79 % $\pm$ 5 %; between HC and AD, 87 % $\pm$ 3 %; and between MCI and AD, 80 % $\pm$ 5 % (AD:26, MCI:23, HC:15) (König et al., 2015).

In another study, the efficiency of different conventional characteristics and the better selection of them has been evaluated from reading tasks in French (AD:14, MCI:14, HC:14) (Mirzaei et al., 2017). The analysed features were 4 related to pitch, 2 related to speech segments and 9 to jitter and shimmer. Two multiclass classifiers, kNN and SVM, were also used. The kNN classifier, by using 10 characteristics, performed better than the SVM classifier, when applied to distinguish between HC and MCI but not with the AD group. Overall, after feature selection, the performance of kNN was relatively improved by 76% with respect to the a priori class probabilities of 33%. We observed that the AD and HC groups were mostly confused with the MCI group. In later works, three classification methods were used: kNN, SVM and decision Tree. To assess the effectiveness of the set of temporal and acoustical voice features (e.g. Jitter and Harmonics-to-Noise Ratio) from read speech of patients, they compare them with two sets of feature parameters that are widely used in speech and speaker recognition applications. A two-stage feature selection process was conducted to optimize classification performance. The proposed method improved the classification accuracy for Early Stage-AD (ES-AD), MCI and HC groups by a relative gain of more than 30% for all classifiers. For these experiments, the data samples of HC, ES-AD and MCI groups were collected at AP-HP Broca Hospital, in Paris (Mirzaei, Yacoubi, & et al., 2018).

Along similar lines, a study (Al-Hameed, Benaissa, & Christensen, 2016), using the DementiaBank database, carried out the function selection (fraction of locally voiceless frames, MFCC2, Kurtosis-MFCC30, Mean-MFCC30, Skewness-MFCC2, Mean-MFCC16, number of voiceless segments, among others) and applied 4 classifiers: Bayesian Networks (BN), Trees-Random Forest (RF), AdaboostM1 (AB) and Meta-Bagging (MB). They used the Weka software (Witten, Frank, Hall, & Pal, 2016) for running the experiments, with k-fold cross-validation, in which they randomly divided the data into K equalized parts. By using a BN classifier, it achieved a classification accuracy of around 94 %, with sensitivity and specificity 97 % and 91 %, respectively. Moreover, pre-processing to mitigate background noise and the 20 best features were used. Other studies, in addition to performing conventional analysis based on the statistical significance of the features selected (Beltrami et al., 2016) (value $p < 0.05$), have applied it to the performed tasks (non-parametric Kolmogorov-Smirnov test: p-value < 0.10, in each task). The automatic classifiers used were kNN 3-neighbours, Logistic Regression and NN. The best results are obtained using NN with an accuracy 79.7 %.

Analysis just from a purely syllabic prospect has been also performed in other studies (AD:45, HC:82) (Martínez-Sánchez et al., 2017) by applying ranges of rhythmic, clinical and metric measures and finally one based on the standard deviation of the syllabic gap duration. This work managed to discriminate AD patients and HC subjects with an accuracy 87 % and both with specificity and sensitivity of 82 %.

On the other hand, although they are less frequent, it is also possible to find some studies whose system is based only on paralinguistic features (Warnita, Inoue, & Shinoda, 2018). What is interesting in this kind of study is that the features used are independent of the specific language and, particularly in the case of this study, it has reached an accuracy of around 73 %. Those works which combine both paralinguistic and linguistic characteristics are more extended (Mueller, Koscik, Hermann, Johnson, & Turkstra, 2018).

Apart from the use of individual features, wider concepts have been defined which combine the parameters mentioned in Section 4.1. This is the case with Automatic Spontaneous Speech Analysis (ASSA) (López-de-Ipiña, Alonso, et al., 2015; López-de-Ipiña et al., 2012) which uses three parameter families: duration (voiced and unvoiced segments), short-time energy in the time domain and the spectral centroid in the frequency domain. It offers very useful information for spontaneous speech analysis and it
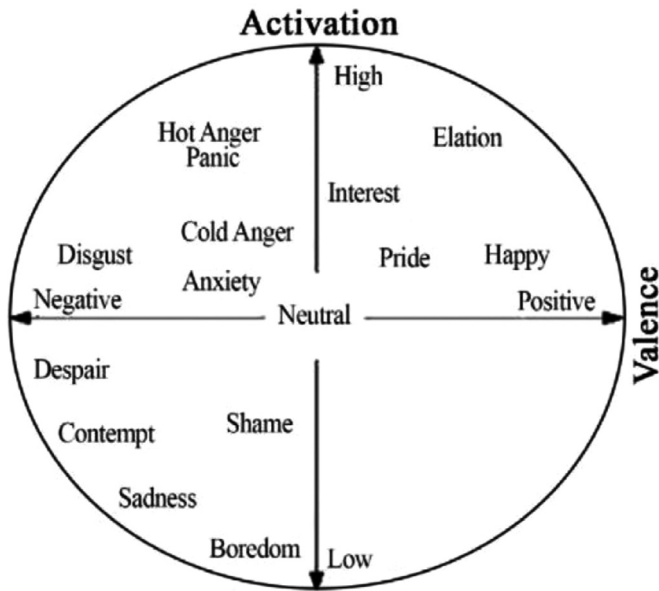
Fig. 3. Activation and valence planes (Alonso et al., 2015).

achieves accuracies around 75.2 % using an SVM classifier and the database AZTIAHO (Proyectos Fundación CITA Alzheimer, 2017).

Automatic speech emotion recognition, for its part, is especially interesting, as previously stated in Section 3, because of the changes produced in AD patients, which disturb their way of expressing emotions when compared with to healthy cognitive subjects. On emotion detection from speech, currently there is no agreement on the number of emotions to be analysed (Cowie & Cornelius, 2003). Thus, most researches focus on four basic emotions: anger, fear, sadness and happiness (Chavhan, Yelure, & Tayade, 2015) or, in some cases, also include surprise and disgust (Balti & Elmaghraby, 2014). Other researches focus on the development of real-time applications and emphasise the usefulness of representing emotions in an evaluation plane in terms of two or more levels or continuous dimensions (Laukka, 2004). In practice, activation and valence levels are the two most commonly used dimensions (Alonso et al., 2015). The activation is related to the intensity perceived from the emotion and the valence level has to do with the pleasant sensation perceived from that stimulus (Alonso et al., 2015). In Fig. 3, both planes and their emotions associated are represented.

Previous researches in automatic speech emotion recognition present feature extraction based on the characterisation of prosodic aspects, such as pitch contour, energy contour, the durations of phonations (Goudbeek & Scherer, 2010; Kwon, Chan, & et al., 2003; Lee & Narayanan, 2003) or Tear Energy Operator (Harimi, Shahzadi, & Ahmadyfard, 2014) which are related to the grammatical structure and the lexical stress. The paralinguistic aspects are reflected in features such as the first formant (Kwon et al., 2003; Lee & Narayanan, 2003) or the concentration of energy in different energy bands, e.g. in Altun and Polat (2009). The classical approach in speech emotion recognition uses a wide feature set, between 13 and 1941 paralinguistic and prosodic features (Altun & Polat, 2009). However, this feature range is not suitable for applications in real-time, due to the complexity of some of the features and the processing time.

Emotional Speech Analysis (ESA) (López-de-Ipiña & Alonso, 2015; López-de-Ipiña et al., 2012), is focused on three parameter families: acoustic characteristics (pitch, SD, maximum and minimum of pitch, SD, maximum and minimum of intensity and SD of period, among others), voice quality characteristics (shimmer, local jitter, HNR, NHR, autocorrelation) and duration characteristics (voiceless segments, among others). Short-term energy is the main feature analysed. Again, these authors used the database AZTHIAO and an SVM classifier. The study reaches an accuracy around 90.7 % in discriminating AD patients from healthy subjects.

Similarly, the concept of Emotional Temperature is based on various prosodic and paralinguistic features derived from the temporal segmentation of the speech signal. ET presents the advantage of quantifying activation using a method with low processing time and is capable of detecting changes in the speaker's affective state. Because of these characteristics, it has future possible use in real-time applications. From each signal fragment, 2 prosodic and 4 paralinguistic features are extracted, related to pitch and energy respectively (Alonso et al., 2015; Alonso, Cabrera, Travieso, López-de-Ipiña, & Sánchez-Medina, 2017). Through ET and the using of SVM, it is possible to classify samples as pathological and non-pathological, thus obtaining positive results on specific AD detection (Alonso et al., 2015). ET and ASSA analysis working jointly have demonstrated a significant fluency loss in AD patients with respect to the percentage and duration of voiced and unvoiced segments (López-de-Ipiña et al., 2012; Lopez-de-Ipiña et al., 2012) and obtain accuracies of around 92.24 %. Likewise, combining ET, ASSA and ESA, the accuracy can reach values around 94 % using SVM classifiers (López-de-Ipiña, Alonso, et al., 2015).

Other interesting trends in voice processing applied to AD detection are Voice Activity Detection (VAD) techniques, in which speech fragments are distinguished from noise by a spectral comparison of the input signal with a stored noise estimation (Barret, 2000). More complex techniques such as Word Spotting are based in grouping word images into similar word clusters by using the coincidence of the image (Rath & Manmatha, 2007). As previously mentioned (see Section 4.1), the widely analysed features continue to be those related to voiced and unvoiced segments (Weiner & Schultz, 2016) and prosodic characteristics in general (Sadeghian, Schaffer, & Zahorian, 2017; Weiner, Herff, & Schultz, 2016). These techniques applied to disease detection like AD permit, once the voice signal is transcribed into text, the carrying out of a linguistic evaluation in addition to an acoustic one. Also, they offer the possibility of performing automatic examinations focused on syntactic, lexical (Hernández-Domíngue, García-Canó, Ratt, & Sierra-Martínez, 2016), semantic (Vincze et al., 2016) and phonetic structure (Thomas, Keselj, Cercone, Rockwood, & Asp, 2005; Tóth, Gosztolya, & et al., 2015; Wankerl, Nöth, & Evert, 2017) or a combination of all of these factors (Hernández-Domínguez, Ratté, Sierra-Martínez, & Roche-Bergua, 2018; Mirheidari, Blackburn, & et al., 2016).

Some researches (Khodabakhsh, Yesil, Guner, & Demiroglu, 2015) have exposed the high discrimination capacity just from prosodic characteristics (where the silence relation is particularly useful), for example, in comparison with other linguistic features (noun-pronoun relations). Other systems studied (Vincze et al., 2016) concluded that features helping to distinguish between MCI and HC are mainly based on morphological aspects (number of words, punctuation signs or nouns). In this work linear features were extracted (such as spoken time and silences or hesitations), morphological and semantic features (such as those related to the memory). It used an SVM classifier. Given the small size of the data set, it applied cross-validation and majority tagging. By using all the features, the system achieved a 69.1 % accuracy.

One of these studies (Fraser, Meltzer, & Rudzicz, 2016) computed a large number of linguistic variables (370) to capture a wide range of linguistic phenomena. By using the DementaBank database, transcripts and acoustic variables from the associated audio files were used to train a classifier and distinguish between AD and HC groups. The authors also added an exploratory factor

analysis to these measures and provided interpretation for the resulting factors used to examine the degree of heterogeneity of linguistic impairment in AD. Short samples of subjects' language on a picture description task were analysed from the DementiaBank TalkBank database. The maximum average accuracy (81.92%) in distinguishing between AD and HC was achieved with the 35 top-ranked features. The accuracy remained relatively constant until the authors choose a feature set of size 50 (accuracy = 78.72%). This fact reinforces the need to carry out feature selection, given the high dimensional feature space used. Semantic impairment (analysed with the pronoun ratio or word length), acoustic abnormality (using the Mel-frequency cepstral coefficients or phonation ratio), syntactic impairment (such as fewer verbs produced), and information impairment (measuring the key words and information units) were clearly detected in AD participants.

Several studies have demonstrated worse results when a large number of features are used in the analysis (Asgari, Kaye, & Dodge, 2017). In a work performed in 2016, by means of German spontaneous speech tasks (HC:80, AACD:13, AD:5), the majority of prosodic features are created from silent segments and voiced segments obtained in turn from VAD, which uses a Hidden-Markov Model recogniser. LDA classifier accuracy is 80.4 % using 14 linear functions (Weiner & Schultz, 2016), although in a later publication (Weiner, Herff, et al., 2016) the authors achieved an overall accuracy of 85.7 %, by using just 10 features.

Works based on n-gram techniques (Thomas et al., 2005) have been developed, which study the frequency correlation of different part of speech and Dementia of Alzheimer Type (DAT) on the database ACADIE (Folstein, Robins, & Helzer, 1983; Rockwood, Graham, & Fay, 2002). The accuracy achieved is high (95 % distinguishing between AD and HC subjects and 70 % discriminating dementia in 2 classes). Also using n-gram models (Wankerl et al., 2017), other research has studied the correlation between statistical calculations and MMSE. It used the DementiaBank and an LDA classifier with cross-validation. In this case, 80.4 % of the samples were correctly classified.

Various studies have highlighted the efficacy of automatic systems applied to voice transcriptions in comparison with those which are non-automatic (Hernández-Domíngue et al., 2016; Tóth et al., 2015). These authors achieved quite close scores to those manual characteristics which were set. The best results were found by using SVM classifier with an accuracy of 82.4 %. Along the same lines, to demonstrate the efficacy of automatic methods (Toth et al., 2018), (MCI: 48, HC: 38, SS), this study was able to separate HC from MCI subjects with F1 punctuation at 78.8 %. In this case, a Random Forest (RF) classifier worked slightly better than that SVM when using the automatic features set. The most significant features between the 2 groups were found in the duration of the speech, for delayed recall tasks and in the number of pauses for answering questions. No final conclusion could be extracted with respect to what machine learning algorithms - SVM or RF - are more suitable for a specific classification task. Other recent research (Abdalla, Rudzicz, & Hirst, 2018) is based on the application of Rhetoric Structure Theory (RST) (Carlson & Marcu, 2001; Mann & Thompson, 1988). Using the DementiaBank ("DementiaBank | TalkBank," 2007) and Carolina Conversations Collection (Carolinas Conversations Collection - About - Who We Are, 2008) databases, relationships to speech based on RST from SS are automatically extracted. To extract these RST relationships the original text is segmented, trees from these segments are built and, subsequently, the most relevant characteristics are selected. The speech relationships, especially those involving elaboration and attribution, are the most significant indicators of AD in speech. On the other hand, transcriptions, involve relationships of logical contingency. Significant differences can be observed in the speech structure between AD and HC subjects, specifically in

SS transcriptions. Likewise, in the DementiaBank subtle differences exist, even between specific dementia subtypes.

A study (Yancheva, 2016), from the denominated Information Content Units (ICU), has demonstrated that AD groups have an idea density significantly lower than HC subjects in any scenario, especially in image description tasks. Another study (Sirts, Piguet, & Johnson, 2017) is based on the concept of Propositional Idea Density (PID), which counts the expressed ideas and can be applied to any text. Also, semantic idea density (SID) is developed and counts pre-defined information content units. This is naturally more applicable to normative domains, such as picture description tasks. These authors also present DEPID, a novel dependency-based method for computing PID, and its version DEPID-R that enables the exclusion of repeated ideas.

Unified computational environments have also been developed to automatically analyse language production in clinical environments such as Coh-Metrix-Dementia (Aluísio, Cunha, & Scarton, 2016; Aluisio, Cunha, Toledo, & Scarton, 2016; Toledo et al., 2018). This system was built on various Natural Language Processing tools. From a transcribed-text sample performed by a subject in Portuguese, 73 text measurements were extracted. In the first phase, it was demonstrated to have a high classification accuracy (Aluisio et al., 2016). In a later study in 2016 (Aluísio et al., 2016), it was concluded that to train regression and classification models it is necessary to reduce the large feature set in Coh-Metrix-Dementia. The researchers also compare three feature selection methods. In 2017 (Toledo et al., 2018), others analysed linear features such as informativity, number of propositions of each text, overall coherence, empty emissions, total idea density, semantic characteristics and quantity of modalisations. This work used non-parametric Kruskal–Wallis tests to compare the performance between three groups of 60 people (HC, MCI and mild AD) with respect to the variables. Mild AD showed a lower informative speech, higher impairment in overall coherence and modalisation, and a lower narrative structure. Despite this, it was not possible to discriminate between MCI and HC.

Other studies (Mueller et al., 2018) have combined models of linear mixed effects by using the WRAP database (Johnson et al., 2018). Transcriptions are coded for automatic analysis by means of Computer Language Analysis (CLAN) (Macwhinney, Fromm, Forbes, & Holland, 2011), which includes codes for filled and empty pauses, repetitions, reviews, semantic units, mistakes (semantic, phonological, lexical) and non-verbal behaviour. Semantic units, parts of speech, grammatical relations and other quantifiers are automatically extracted by CLAN using MOR and MEGRASP tools (Pye & MacWhinney, 1994). Early MCI (eMCI) participants decrease faster in fluent speech features and semantic content than those HC. Lexical diversity and grammatical complexity measurements are not associated with the eMCI stage.

4.3. Non-conventional feature extraction

In more advanced AD stages, voice becomes aperiodic, noisy, irregular and chaotic. Sometimes these features limit the capacity of the traditional parameters to obtain useful information about a specific pathology. On this basis, researchers have developed more complex and robust parameters (see Table 2). Compared with those traditional, non-linear features, these offer more precise information, in spite of being clinically less interpretable (Brabenec et al., 2017). With this regard, for example, it has been possible to demonstrate that a pathological voice signal can be described by non-linear dynamic analysis (Alonso, De Leon, Alonso, & Ferrer, 2001; Mekyska et al., 2015).

There are several works on voice analysis using non-linear features of the signal, although they are relatively recent. More precisely, superior order spectra and quadratic phase coupling of the

Table 2
Main non-conventional features used in Alzheimer's disease research (Mekyska et al., 2015)

Feature	Description
Correlation Dimension (CD)	Attractor geometry in the phase space (Little, McSharry, & Hunter, 2009; Shao, MacCallum, Zhang, & Sprecher, 2010; Vaziri, Almasganj, & Behroozmand, 2010).
Fractal Dimension (FD)	Number of basic building blocks forming a pattern (Esteller & et al., 2001; Vaziri et al., 2010).
Ziv-Lempel Complexity (ZL)	Quantifies the regularity of a temporal series (Aboy, Hornero, & Abásolo, 2006).
Hurst Exponent (HE)	Describes possible long-term dependencies in the speech signal (Orozco-Arroyave & Murillo-Rendón, 2011).
Entropy	
Shannon Entropy (SHE)	Measure of uncertainty and quantifies the complexity of a system
Second-order Rényi Entropy (RE)	RE quantifies the loss of information in time in a dynamic system (Henríquez et al., 2009; Jayawardena, Xu, & Li, 2010).
Correlation Entropy (CE)	CE is the indication of the predictability of the non-linear time series (Henríquez et al., 2009; Jayawardena et al., 2010).
First-order And Second-order Rényi Block Entropy (RBE1 and RBE2) (Henríquez et al., 2009)	
Approximate Entropy (AE) Sample Entropy (SE)	The only difference between AE and SE (Yentes et al., 2013) is that SE does not evaluate a comparison between embedding vectors with themselves (Herisa Khadivi, Seyed Aghazadehb, & Nikkhah-Bahramic, 2009; Vaziri et al., 2010; Yentes et al., 2013).
Recurrence Probability Density Entropy (RPDE)	Uncertainty in the measurement of the pitch period (Little, McSharry, Roberts, Costello, & Moroz, 2007).
Multiscale Permutation Entropy (MPE)	Assesses recording complexity (Lopez-de-Ipina et al., 2018)
First Minimum of Mutual Information (FMMI)	Discriminates between the different voice qualities of the multi-quality database (Henríquez et al., 2009).
Largest Lyapunov Exponent (LLE).	Includes a measure of sensitivity to an initial condition (Orozco-Arroyave & Murillo-Rendón, 2011; Vaziri et al., 2010).
Detrended Fluctuation Analysis (DFA)	Characterises the self-similarity of the graph of a signal from a stochastic process (Little et al., 2007; Tsanas, …, & 2011, 2011).

SS signal for HC and AD groups have been explored by using bi-spectrum and bi-coherence methods (Nasrolahzadeh et al., 2016). It has been possible to prove that the quadratic phase coupling of the SS signal of AD subjects decreases in comparison with HC subjects and the coupled in-phase harmonics of speech move to high frequencies for AD subjects. In 2016, one of these studies, on the frequency variability and dynamics of the SS signal of AD and HC groups (Nasrolahzadeh, Mohammadpoori, & Haddadnia, 2016), examined the surface of the Mean Square Error (MSE) and the quantification of level curves for both groups. It concluded that the voice signal moves from a high dimension chaotic stage in HC subjects to a low dimensional one in AD patients.

In general, these non-conventional feature studies base their perspective on a combination of linear and non-linear. A clear example is the application of ASSA and ERA techniques with the Higuchi Fractal Dimension and a set of acoustic and stationary features (López-de-Ipiña et al., 2013). These were applied by implementing a classifier of a Multilayer Perceptron (MLP) neural network, thus obtaining very promising results (Lopez-de-Ipiña et al., 2014). Quasi-static fluctuation of the vocal folds and the setting provided by the vocal tract are non-linear processes, mainly produced when consonants are pronounced and can be described by means of fractals. According to Mandelbrot (1982), a fractal is a geometric pattern which iterates through smaller or larger scales for performing on auto-similar forms or surfaces, i.e. irregularities that Euclidean geometry cannot represent. The speech signal waveform shows both periodicity and auto-similarity when vowels and consonants jointly form syllables. Different studies have already carried out fractal analysis on the voice signal (Huang et al., 2011).

Other works confirm that additional information provided by non-linear features from FD could be especially useful when datasets are limited. These studies explored the unbalanced dataset issue by means of one-class classifiers and concluded that FD improves the system performance in these cases (Lopez-de-Ipiña et al., 2014; López-de-Ipiña et al., 2015). In general, using FD characteristics with multi-class classifiers (one and two classes) has shown very promising results. In later works, these authors (López-de-Ipiña, Alonso-Hernández, et al., 2015; López-de-Ipiña et al., 2013) went deeper into the utility of FD by using modelling techniques based on Emotional Response features. For this purpose, using the AZTIAHO database, it was proposed

to use Higuchi, Katz and Castiglioni algorithms and add these new non-linear features to the set feeding the training process (López-de-Ipiña, Alonso-Hernández, et al., 2015). The best results were obtained with a combination of quality, duration and acoustic features, with ET and a set involving FD (Castiglioni, 2010; Higuchi, 1988; Katz, 1988) and its statistical variations for the complete signal. This work also analysed different classifiers and obtained the best results by using MLP classifiers. The final system accuracy exceeded 90% in all groups.

Recent researches, using complex systems, show that geometries and naturally evolving events cannot be characterised by just a scale relation (monofractal system), because different parts of the system have different scales. Some works showed that only multifractal analysis can measure its internal dynamic with higher accuracy (Bhaduri, Das, & Ghosh, 2016; Kantelhardt, Koscielny-Bunde, Rego, Havlin, & Bunde, 2001; Stanley et al., 1993). Multifractals are more complex auto-similar objects consisting of fractals weighted in different ways with different non-integrated dimensions. Their scaling properties vary in the different regions of the systems (Chen, Ivanov, Hu, & Stanley, 2002). In various works, the multifractal nature of speech has been analysed for the representation and characterisation of different kinds of signals (González, Luan Ling, & Violaro, 2012).

Detrended Fluctuation Analysis (DFA) is an escalation analysis method where the scaling exponent (similar to Hurst exponent of single scale or FD) is used for quantifying a large range correlation of the stationary and non-stationary signal (Hausdorff et al., 1996). Another method, MultiFractal Detrended Fluctuation Analysis (MFDFA) has been applied to study the multifractal escalation behaviour of diverse invariant time series of a non-stationary scale (Kantelhardt et al., 2002). Some sources (Bhaduri et al., 2016) confirm that results obtained by this method were more reliable in comparison with techniques like Wavelet Analysis, Discrete Wavelet Transform, Wavelet Transform Modulus Maxima, Detrending Moving Average, Band Moving Average, Modified Detrended Fluctuation Analysis, etc. (Huang et al., 2011; Kantelhardt et al., 2002; Oświęcimka, Kwapień, & Drożdż, 2006; Serrano & Figliola, 2009).

In the same study (Bhaduri et al., 2016) a new feature was defined for detecting emotions in AD from speech and MFDFA. It proposed a quantitative parameter to categorize various emotions

through analysis of non-stationary details of the speech signal dynamic. For this purpose, the TESS database was used (Toronto emotional speech set (TESS) | TSpace Repository, 2018). The Hurst exponent and spectral bandwidth were calculated for 1200 clips (200 for each one of 6 emotions) and according to the method described by Kantelhardt et al. (2002). From Hurst exponent values and multifractal spectral wide trends, the results showed that it is possible to clearly distinguish between main emotions and, in turn, to apply this to AD detection (Bhaduri et al., 2016).

4.4. Deep learning

The trend in favour of data classification models based on Deep Learning (DL) has been increasing (Kim, Lee, & Provost, 2013). These models are used to collect complex interactions of non-linear features in multimodal data such as audio and video characteristics. Nevertheless, it should be always taken into account that the same feature vector can offer very different classification results by using different algorithms (Kim et al., 2013).

Until now, the most popular and extended classifiers which use conventional characteristics have been the linear discrimination classifier (Roy & Pentland, 1996) and the k-NN (Ekman & Davidson, 1994), in addition to SVM which also has reached a promising classification performance (You, Chen, Bu, Liu, & Tao, 2006) and is the most used one. Discriminative non-linear classifiers such as Artificial Neural Networks (ANN) and Decision Trees are frequently used because of their robust performance in non-linear classification (Martínez & Cruz, 2005; Schuller et al., 2007). Other DL architectures are Convolutional Neural Networks (CNN) or the Multilayer Perceptron (MLP). In general, different architectures are interesting in resolving different kinds of problems.

On Machine Learning (ML), Deep Learning allows the implementation of computational models consisting of numerous processing layers in order to learn data representations with various levels of abstraction. These methods have drastically improved the state of art, especially in recognition and voice processing (LeCun, Bengio, & Hinton, 2015). DL permits the discovery of complex structures in large datasets by means of retro-propagation algorithms. It shows how a machine should change the internal parameters it uses to calculate representations in each layer from the representation of the previous one. Deep Convolutional Networks have achieved advances in imaging, video, voice and audio processing while Recurrent Networks have enlightened sequential data such as text and speech (LeCun et al., 2015). Studies previously mentioned (see Section 4.2) have applied DL techniques to classification processes. MLP neural networks have been commonly used to classify by using linear and non-linear features (López-de-Ipiña, Alonso-Hernández, et al., 2015; Lopez-de-Ipiña et al., 2014; López-de-Ipiña et al., 2013).

In 2017, an automatic system (Sadeghian et al., 2017) based on linear features processed each sample of separate pauses and voice segments through using VAD techniques. With those speech samples suitably separated, it calculated 22 metrics, including total speech duration, number of pauses and fraction of paused conversation time, pitch and energy, among others. An RNN with 2 hidden layers was set, each one with 256 neurons and reached accuracy 94 % in discriminating AD from HC groups. It included a combination of speech acoustic features, linguistic features extracted from the automatic transcription of speech, including punctuation signs, and MMSE results.

Most recent studies performed with ANN and non-linear parameters obtain very promising results, although this is still a little explored area. Also in 2017 ALZUMERIC was developed (de Lizarduy, Salomón, Vilda, Torres, & de Ipiña, 2017), where specialist physicians could collect voice samples. The system automatically extracted features from them to offer a beginning

of diagnosis based on articulation, speech quality, emotional response analysis, language perception and the complex dynamics of the system. This performed a selection of conventional features, perceptual and non-linear, by means of the Mann-Whitney U-test with value p < 0.1. It compares four classifiers: k-NN, SVM, MLP, and CNN. The software set WEKA (Witten et al., 2016) was used to carry out the experiments. For training and validation, it used k-fold cross-validation with k = 10 (Picard & Cook, 1984). The task with the best accuracy was SS with 95 %. The best results, according to the most relevant feature selection by statistical tests (under medical criteria) and classification methods, were obtained with the Mann-Whitney U-test and SVM (de Lizarduy et al., 2017). In the continuation of the work, it again uses deep learning through CNN and non-linear multi real modelling. It includes conventional features (HNR, pitch, jitter), perceptual features (MFCC...) and non-conventional features (fractals, Shannon Entropy, among others). The authors conclude that automatic integration of the most relevant features by using CNN (2 layers and 20 neurons) provides useful information, not available in other statistical tests. In 2018, the most recent of these studies (Lopez-de-Ipina et al., 2018) analysed three tasks with different levels of language complexity. The automatic classification was carried out using MLP and CNN. Again, the work used conventional linear and perceptual features, Castiglioni fractal dimension and Multiscale Permutation Entropy. It concluded that the most relevant features are selected by means of the non-parametric Mann-Whitney U-test.

Another study (Warnita et al., 2018) proposed an automatic AD detection method using a Gated Convolutional Neural Network (GCNN), which could be trained with a relatively small amount of data and collect temporal information on audio paralinguistic functions. Moreover, given that it used none linguistic features, it could be easily applied to any language. OpenSMILE (Eyben, Weninger, Gross, & Schuller, 2013) and INTERSPEECH 2009-2012 (Schuller et al., 2010; Schuller, Steidl, Batliner, Schiel, & Krajewski, 2011; Schuller et al., 2012) were used. The proposed method reached an accuracy 73.6 %, which was better than conventional Sequential Minimum Optimization (SMO) by 7.6 percentage points. The method is so far unique for evaluating symptoms in different languages.

4.5. Databases

The available databases are a fundamental tool for research and, for this reason, those mentioned in Section 4 are now described in Table 3. This includes some information intended to simply orient the reader and the studies related are just labelled for information purposes rather than being a classification in itself. The state-of-the-art is very heterogeneous regarding the tasks used in recordings. This is a key issue since the features that are able to measure impairment and their meaning obviously vary widely depending on what task is performed. We have considered that SS tasks involve every task where the subject is asked about a specific topic and answers freely during a certain relatively long time. This can include a large range from interviews to picture descriptions. Although pauses made during verbal fluency tasks may be indicative of semantic memory problems, it is not necessarily true for pauses made during a picture description task and it is necessary to qualify that this review does not develop or analyse this point. In the same way, we have described those databases without a clear main task as a mixed group. This would include tasks from SS to reading or even repetitive tasks.

- Atlantic Canada Alzheimer's Disease Investigation of Expectations (ACADIE): includes 2 interviews for each patient, separated by 12 weeks. There are 79 subjects (both AD and HC): mild (50), moderate (53), severe (20) and HC (35). The study

Table 3
DataBases located.

DataBase	Language	Task	HC M/F	MCI M/F	AD M/F	References
SAIOTEK	Multilingual	SS	5	-	3	(López-de-Ipiña et al., 2012; Lopez-de-Ipiña et al., 2012)
AZTIAHO	Multilingual	SS	50	-	20	(López-de-Ipiña, Alonso, et al., 2015; Lopez-de-Ipiña et al., 2014; López-de-Ipiña et al., 2013; López-de-Ipiña et al., 2015; López-de-Ipiña, Alonso-Hernández, et al., 2015; Lopez-de-Ipina et al., 2018)
PGA-OREKA	Multilingual	SS	26/36	-	17/21	de Lizarduy et al., 2017
-	Greek	SS	30	-	30	Rentoumi et al., 2017
Dem@care	Greek	Mixed	4/15	12/31	3/24	Satt et al., 2013
	French	Mixed	6/9	11/12	13/13	Tröger et al., 2017; Satt et al., 2014; König et al., 2015
-	French	Reading	14	14	14	Mirzaei et al., 2017
-	French	SS	5	-	5	Boyé, Tran, & Grabar, 2014
*TRANSC	English	SS	184	-	214	Luz, 2017
-	English	SS	27	14	-	Asgari et al., 2017
-	English	SS	46	-	26	Sadeghian et al., 2017
WRAP	English	SS	200	-	64	Mueller et al., 2018
PC (DB)	English	SS	74	19	169	Al-Hameed et al., 2016; Zhou, Fraser, & Rudzicz, 2016; Hernández-Domínguez et al., 2018; Warnita et al., 2018; Wankerl et al., 2017; Yancheva, 2016; Abdalla et al., 2018; Sirts et al., 2017
ACADIE	English	SS	-	-	95	Thomas et al., 2005
AMI	English	SS	20	-	20	Sirts et al., 2017
CCC	English	SS	-		125	("Carolinas Conversations Collection - About - Who We Are," 2008), (Abdalla et al., 2018)
	English	-	10	-	55	
ILSE	English	SS		23		(Weiner & Schultz, 2016)
	German	-	80	13	5	Weiner, Herff, et al., 2016
-	Spanish	-	-	-	21	Meilán et al., 2012
-	Spanish	Reading	82	-	45	(Martínez-Sánchez et al., 2017)
UNED	Spanish	SS	29	-	34	Hernández-Domingue et al., 2016
Cinderella	Portuguese	SS	20	20	20	Toledo et al., 2018
OPLON	Italian	SS	48	48		(Beltrami et al., 2016)
-	Iranian	SS	15/15	-	16/14	Nasrolahzadeh et al., 2016
-	Japanese	Mixed	7/3	-	9/1	Tanaka et al., 2016a; Tanaka et al., 2016b
-	Japanese	Mixed		73/200		(Kato et al., 2018)
BELBI	Serbian	SS	-	-	12	Graovac, Kovacevic, & Lazetic, 2016
*BEA	Hungarian	SS	13/23	16/32	-	Gosztolya et al., 2016; Toth et al., 2018; Vincze et al., 2016; Toth et al., 2018; Tóth et al., 2015
-	Turkish	SS	31/20	-	18/10	Khodabakhsh et al., 2015
-	French	-	29	-	29	St-Pierre, Ska, & Béland, 2005
-	Persian		0/6	-	0/6	Ahangar et al., 2018; Malekzadeh, Arsalan, & Shahabi, 2009

Others studies using unknown databases

Khodabakhsh & Demiroglu, 2015; König et al., 2016; Aluisio et al., 2016; Aluísio et al., 2016; Nasrolahzadeh et al., 2016.

DB: DementiaBank, PC: Pitt Corpus, CCC: Carolina Conversation Collection, ILSE: Interdisciplinary Longitudinal Study of Adult Development and Aging, ACADIE: Atlantic Canada Alzheimer's Disease Investigation of Expectations, WRAP: Wisconsin Registry for Alzheimer's Prevention, BEA: Hungarian Spoken Language Database, UNED: Universidad Nacional de Educación a Distancia.

was carried out at various sites in Canada (Folstein et al., 1983; Rockwood et al., 2002).

- Interdisciplinary Longitudinal Study of Adult Development and Aging (ILSE): this is a study performed in Germany. It counts more than 8000 h of biographic interviews and cognitive diagnostics (AD, Aging-associated cognitive decline (AACD) and HC groups) from more than 1000 different participants during 20 years. There are only 380 h transcribed (Weiner, Frankenberg, & Telaar, 2016).
- TALKBANK: a set of databases classified on each communication subfield and publicly available. Those focused on clinical aspects are DementiaBank, RHDBank, TBIBank, AphasiaBank, ASDBank, FluencyBank. DementiaBank (DementiaBank | TalkBank, 2007) includes Pitt Corpus and is a database performed by subjects suffering dementia and HC. DementiaBank consists of transcribed interviews and uses imaging descriptions as oral tasks, specifically "Theft of cookies" task (Kaplan et al., 2001) of the Boston Diagnostic Aphasia Examination (Kaplan et al., 2001).
- AMI: contains biographic interviews of AD patients and HC. Each interview covers 4 stories about events of a specific period of the subject's life. It contains multimodal data and has 100 h of records ("AMI Corpus," 2006).
- Dem@care: the European project Dem@Care (Dem@Care Project - Project, 2011) contains different aspects of the patient's speech analysis, among others. It contains different databases such as those related in Table IV. Some of them, in turn, cover different corpuses in different languages.
- Carolina Conversations Collection (CCC): a property of the Medical University of South Carolina. It consists of a password protected digital collection of audio and video records transcribed from spontaneous speech (Carolinas Conversations Collection - About - Who We Are, 2008). It has two cohorts: 125 older multi-ethnic speakers mostly from North Carolina, without disability and with any of 12 chronic conditions. There are 70.45 h of records of 125 people with dementia (Corpus Linguistics and Linguistic Theory, 2011).
- Proyecto Gipuzkoa-Alzheimer (PGA): PGA Project includes various databases, which have in turn defined several subsets:

- ○ AZTIAHO: multicultural and multilingual spontaneous speech database in English, French, Spanish, Basque, Catalan, Chinese, Arabic and Portuguese. It consists of video records, HC: 50 (12 h), AD: 20 (8 h) (Lopez-de-Ipina et al., 2015).
  - ○ AZTIAHORE Subset (Proyectos Fundación CITA Alzheimer, 2017) covers HC: 20 (9 h), AD: 20 (60 minutes).
  - ○ PGA-OREKA: cVF (Animal naming, AN). HC: 187, MCI: 38 forms a PGA subset (Lopez-de-Ipina et al., 2015; Proyectos Fundación CITA Alzheimer, 2017).
  - ○ MINI-PGA: (imaging description). HC: 12, AD: 6 (Proyectos Fundación CITA Alzheimer, 2017).
- Wisconsin Registry for Alzheimer's Prevention (WRAP): this is a study of longitudinal observation with people with an AD parental history. Since the end of 2001, 1561 people have been incorporated with an initial average age of 54 years. Participants perform a second visit four years after the beginning of the study, and subsequent visits are carried out every two years (Johnson et al., 2018).
- Hungarian Spoken Language Database (BEA): Hungarian multipurpose database that covers spontaneous speech tasks, sentence repetition and reading (Gósy, 2013). It consists of 260 h performed by 280 speakers from Budapest (ages between 20 and 90 years, 168 women and 112 men), also providing comments for different researches and practical applications, MCI among them.

Fig. 4. Publications identified on dementia and voice automatic processing applied to AD detection since 1995. This is based on the publications identified in this paper.

5. Discussion

Given the growing interest, in recent decades, and because of the progressive aging of the population, there has been a significant increase in the number of resources invested and publications written about diseases such as AD. AD is one of the great challenges of our society and currently the research lines developed are diverse. In this work, a systematic search has been carried out in order to relate the voice and speech processing techniques applied to AD detection. The trend of the number of publications based on dementia in general and those based on AD and automatic voice processing is represented in Fig. 4.

In order to discuss the trajectory of the AD detection studies in the area of voice automatic analysis, we have carefully divided these publications, depending on the kind of characteristics used: conventional, without any doubt the most used and extensive kind, and the non-conventional, which refer to no such extensive knowledge and which are still to be explored. A third area concerns Deep Learning techniques. While they do not provide knowledge as such, they offer interesting solutions that suitably could detect AD and, in the not too distant future, distinguish it from other impairments, dementias and pathologies.

The present review does not go into clinical factors or into a detailed classification of populations. Thus, the two populations labelled as MCI might have different impairments, due to subtypes or progressions, and therefore differences in their voice patterns. Within the scope of this review it is not possible to explore these populations further. Searching for more detailed subdivision is clearly a very interesting and necessary task, where not only different kinds of MCI could be classified but also different pathologies and their corresponding voice patterns.

Since the first publications emerged, in approximately 2005, 78 % of the researches have focused on using conventional parameters, mainly on the duration of voiced and unvoiced segments, pitch, amplitude and periodicity, as well as other features obtained from the temporal, frequency and cepstral analysis. These variables, as we have seen, have provided data concerning cognitive processes and their values have been directly related to the specific sta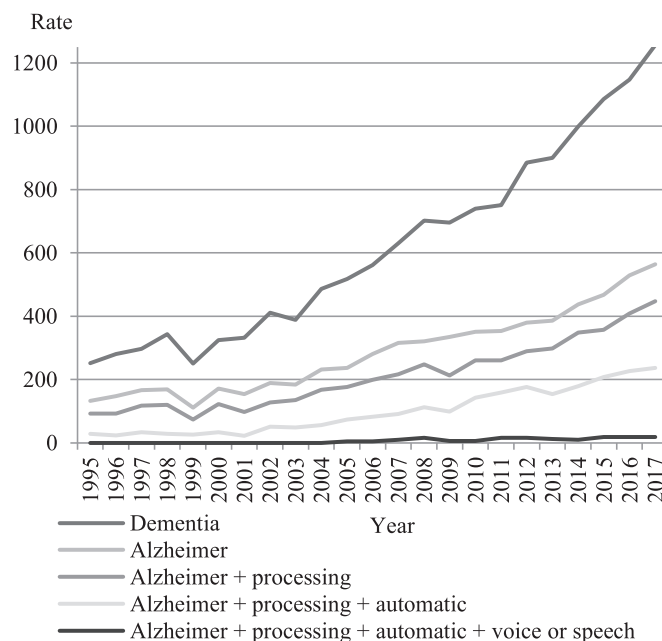te of the AD. Each author has performed different interpretations of these parameters and in this sense has related the cognitive deficits with communicative phenomena and features such as the prosody, phonation, articulation and vocal and paralinguistic features, among others.
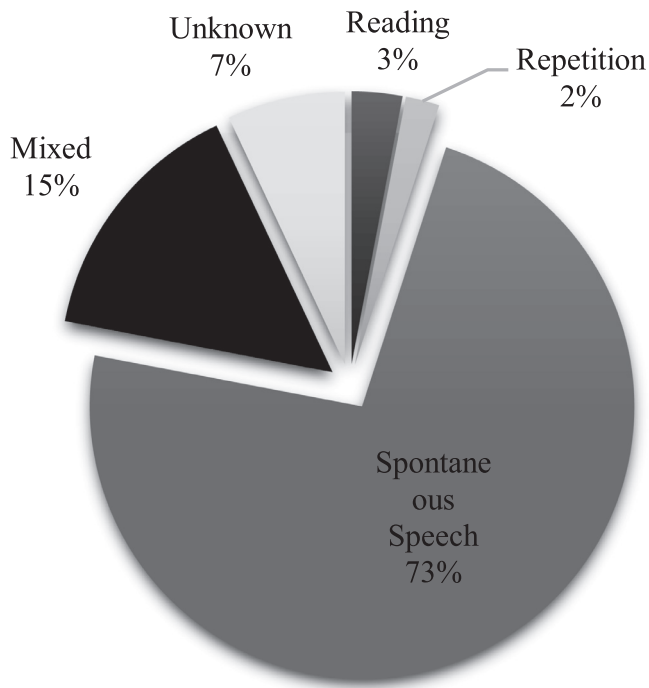
Likewise, through combinations of several parameters, various concepts have been defined such as voice or speech quality or Emotional Temperature. These concepts, correctly applied, have been demonstrated to be good indicators of AD. More elaborate techniques such as Automatic Spontaneous Speech Analysis (ASSA) are presented as methods involving different combined attributes of the voice (ASSA e.g., durations, short time energy and spectral centroid), and they have proved to offer extremely relevant information.

If there is something that most studies point to, it is the importance of the number of characteristics obtained in the feature extraction process. Different studies have already worked on this aspect and, in the light of the outcomes, it is clear that a more reduced set of features leads to better results in terms, in e.g., of accuracy. A feature selection phase is needed and desirable before the classification process.

Regarding the oral tasks carried out by the subjects, they have mainly consisted of a series of spontaneous speech samples. According to our research, this is the most extended task and, perhaps, that which provides the most relevant results because of clearly deficient performance of the AD patients. Although less prevalent, others studies have developed reading, speech repetition or counting samples, among others (see Fig. 5).

The tasks used in recordings deserve special attention. The features that are able to measure impairment in a patient group and their meaning vary widely depending on what task is performed. A deeper analysis would be desirable, in e.g., to answer what is the meaning of a pause made during a verbal fluency task. This might be indicative of semantic memory problems, but it is not necessarily true for pauses made during a counting task. The studies located in this review are not classified to take account of this fact. Further research is required if this ambiguity is to be clarified.
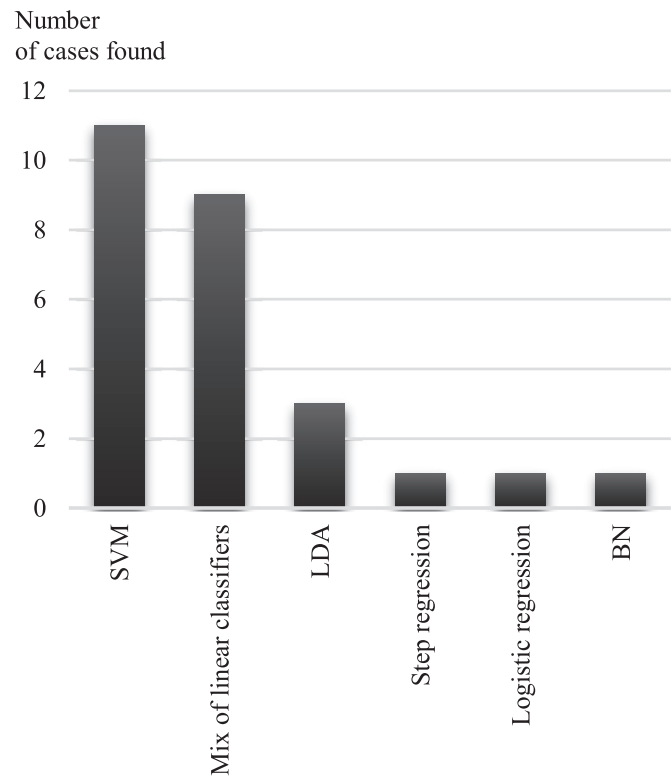
Fig. 5. The most popular tasks carried out in order to perform the voice or speech analysis applied to AD detection. This is based on the publications found in this paper.

Fig. 6. The most popular classifiers on voice analysis applied to AD detection. This is based on the publications discussed in this paper.

However, once the statistical process of feature extraction is performed, data classification should be implemented. In the majority of the cases we have analysed, this is carried out through SVM, kNN, RF or BN classifiers, to a lesser extent, and LDA classifiers (see Fig. 6). The extraction and selection processes of the different features, as well as the previously mentioned classifiers (see Section 4) have demonstrated the achievement of the objective assessment of the AD state. Moreover, they offer the possibility of applying these methods to other neurodegenerative disorders like Parkinson's disease.

It is clear that several linguistic differences can be found, not only in speakers of different languages but also in the various dialects in the same language. This fact could be a problem in respect of implementing a large-scale tool. Not everyone expresses in the same way, speaks at the same speed or uses the same number of words to communicate. At this point, it is interesting to contemplate whether it would be possible to create a valid tool for assisting diagnosis at a global level. Some researchers have already found an answer within the emotional analysis. As demonstrated, the differences of emotional states can be considered as one of the most important assessment criteria for measuring cognitive performance. Since emotions are universal and intrinsic human characteristics, recognising them in the communicative process has caused a great deal of interest in the scientific community.

So, once this limitation is recognised, several studies include classic features such as pitch, intensity and variation of frequency components and, more recently, emotional temperature (a combination of prosodic and paralinguistic features) in the emotional analysis. These studies apply ERA, which uses different linear characteristics and, combined with ET and SS tasks, have achieved discrimination between AD patients and HC subjects with 97 % accuracy. Similarly, several studies have been performed which combine ASSA with these other techniques and they also offer promising results. These methods are based on acoustic and stationary features of the speech audio signal. The majority of them use stationary methods in the frequency domain, such as the Fourier power spectrum.
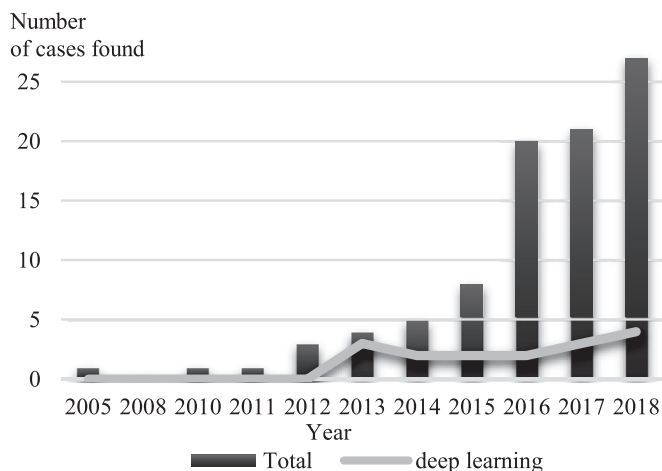
Several studies have also been published using transcriptions and the application of VAD. In addition to the acoustic analysis of the voice or speech signal, they also offer lexical, semantic, punctuation signs and syntactic analysis of the communication process.

In recent years, non-linear and non-stationary aspects related to the dynamic changes on the speech signal affected by cognitive impairment, seems to have awoken a large interest. Since 2012 approximately, more studies in the field increasingly underline the need for relying on non-linear features. Currently, these represent about 22 % of the relevant publications.
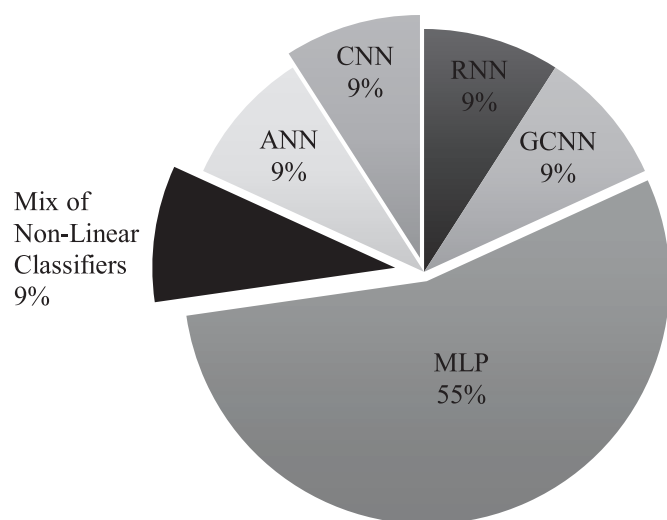
Some researchers have proposed that subtle, cognitive changes in early and pre-clinical states could be detected by means of fractals. For this reason, techniques such as ASSA and ERA have already been included in the set of acoustic and stationary features and also with MLP neural network classifiers, in particular, the Higuchi Fractal Dimension. The results thus obtained are promising and increase the prospects even more of using only linear techniques.

Other recent researches on this complex system have proved that it is not viable to characterise with just a single scale relationship those geometries and events which are naturally evolving. These workers also maintain that a system such as the voice is better characterised by means of multifractals. Among these analysis methods, MDFA is been successfully applied to speech analysis and it results are more reliable in comparison with others such as Wavelet Analysis, Discrete Wavelet Transform, Wavelet Transform Modulus Maxima or the Detrending Moving Average, among others. This has also been applied in order to analyse events such as the dynamics of the cardiac frequency, neuron spiking or economic time series.

The growth of Artificial Intelligence and, specifically, Machine Learning has revolutionised the overall prospects for the success

Fig. 7. Deep Learning applied to automatic detection of voice-speech processing. This is based on the publications found on this paper.

Fig. 8. The most popular DL classifiers on articles detecting AD from the voice signal. This is based on the publications found on this paper.

of speech and voice analysis, by endowing them with more complex systems. ML offers statistical and mathematical analysis and a multitude of feature extraction techniques as well as automatic classification. Seemingly, it consistently meets the new challenges posed. Currently, it is field under considerable study and development. Fig. 7 represents a comparative graphic where the growing trend in the number of published articles on using of DL techniques applied to AD detection from the voice signal can be observed.

According to our researches, this increase has mainly occurred since 2012. Our evidence shows that the most commonly used methods in this regard have been MLP, applied to ASSA, ERA, ET and Fractal Dimension, GCNN, applied to a set of paralinguistic features and CNN, both to conventional features and non-conventional (see Fig. 8).

The innovative techniques of DL seem to offer the drivers required by automatic signal processing. They have placed the signal processing among the most promising methods endowed with solutions to such complex system as the communicative process, the voice and speech production and also the human cognitive processes.

6. Conclusions

For some decades, the number of sufferers around the world in the current century's epidemic of AD, has fast increased. The figures are alarming since they are expected to triple by 2050. Currently, there is no cure and when the first symptoms are evident, the damage caused is already irreparable and chronic. We still do not know exactly how, when or why a person develops AD. Due to the current diagnostic methods and long-term care and treatments, the increase in expected lifespan, as well as the reversal of the generational pyramid, a clear need for more efficient and rational health systems around the entire world will be necessary to deal with AD.

However, although there is still a lot to learn around this disease, there is and have been many important researches in the field as well as on prevention strategies. Today, we surely know more than ever about AD and this will improve with the progress of technology. It is clear that earlier detection would help to better understand the disease. Due to the long time needed to diagnose AD currently, when specialists treat patients, it is late and the effectiveness of the treatment decreases significantly. Knowing more exactly when the disease starts, even in its preclinical phases, could be critical for the effectiveness of the treatment and, for further longitudinal studies, very interesting and important. Based on this, finding accessible biomarkers that serve as a screening method is proposed as a key task for the evolving applicability of AD diagnosis and control.

Recently, several studies have proved that voice or speech analysis is a powerful indicator of the cognitive state of the AD patients. This allows detecting the first symptoms years before the probable clinical diagnosis is established. This fact is well known to all the experts in the field who have manually administered different neuropsychological tests in order to analyse the speech voice information. In this sense, the spearhead for voice processing applied to AD detection would be to detect which MCI patients evolve to early AD. To that end, apart from having access to databases used in the field, it would be especially interesting to collect longitudinal information about patients, over time and in the different AD stages.

Most databases located in this review are cross-sectional. One of the possible reasons why there are fewer longitudinal studies may be because it is more difficult to work with the subjects and ensuring their acceptance when the process is long in time and, even more so, for extended groups. The resources and time required are larger and, until today, there is an important lack of a common criterion for processing the samples. Regarding the amount of data and its availability, due to the fact that collecting this information is not an easy task, databases are not extensive in most cases and, if large, are not usually shared. Developing and working on database challenges publicly would jointly increase knowledge for that purpose and help to achieve common feature extraction and sample processing criteria. For that purpose, a suitable budget is strongly needed, as well as the necessary staff, for engaging subjects to participate in substantial longitudinal studies.

Progress in technology in recent decades, shows a turning point for the linguistic analysis. The new automatic systems offer faster results and with less effort than their existing manual counterparts for language or voice measurements. Beyond automating the analysis of some specific tests, current research has also made important advances in speech acoustic analysis, both in improving accessibility and in reducing the computational burden of data processing. In a broad way, these researches also have opened new and important areas around the concept of Health 4.0, where the trend is to enhance the security aspects of people with dementia by means of sensing devices, as well as developing tools based on the treatment and therapeutic issues for the patients and their caregivers.

Being able to develop web applications, based on biomarkers, which might globalise the evolving and pharmacological control through the voice signal in an easy, fast, non-invasive and scalable way, underlines the importance of this kind of analysis. These applications offer objective values and complement specialist diagnostic work. None of the presented techniques requires an extensive infrastructure or availability of medical equipment: only verbal tests and interviews with the patient are required. Making the pre-screening and control process easier for the patients, by means of transparent techniques, even from home, will help to reduce the stress caused by the current manual, diagnostic procedures.

In this review, the studies analysed have been classified by the feature extraction process they used: conventional or non-conventional. In spite of this, the researches are particularly varied and delving deeper into analysing populations and tasks performed in recording is still pending. Although the most researches in the field have focused on the conventional aspects of the voice signal, non-conventional ones are becoming increasingly more significant. Due to the linear and non-linear aspects of the voice signal, the application of both kinds of combined features offers more complete results. Some researches have pointed to subtle cognitive changes in early and pre-clinical states which could be detected by means of fractals. Thus more conventional techniques such as ASSA and ERA have already included fractals into their set of acoustic and stationary features. For its part, emotional response disturbances suffered by the patient are susceptible to further analysis and quantifications by using the voice signal. This could improve significantly the results of the detection process and might provide a universal criterion to classify, regardless of the language spoken. On the classification process, the experimental and statistical assessments reveal that using ML algorithms to classify linguistic biomarkers through the verbal statements of elder people will facilitate the clinical diagnosis of AD. Seemingly, the future will include the current Deep Learning techniques, including MLP or CNN classifiers, which open an interesting route via more complex classification systems on voice and speech signals. Although the studies performed in this research field have produced encouraging results, there is a need to train the evolving control models on larger data sets.

As mentioned, the state-of-the-art is also very heterogeneous, among others, regarding the tasks used in recordings since each study has to follow its own criteria. This is a key issue because the features, which are able to measure impairment and its meaning, vary widely depending to what task is performed. It is necessary to qualify that this review has not analysed this point and going deeper might be another research in itself. Although we have performed a classification of the studies relevant to the feature extraction process, a new detailed classification regarding the tasks performed or, in e.g., regarding the different AD subtypes could be quite interesting and important. Since the present review is a first attempt at such a review, searching for these new classifications is clearly a very necessary task. Not only different kinds of MCI should be classified, but also different pathologies and their corresponding voice patterns analysed.

Multimodal analysis might be another means to assess the early onset of AD, whether by behavioural patterns analysis (as for writing, speaking or video stimulation) or others kind of biomarkers such as in blood plasma. In order to make advances in this area, it is strongly recommended, at least, to start defining what biomarker is related to a specific pathology, as has previously occurred in many other biometric fields. So, the primary aim is to make possible a differential diagnosis from the voice or speech signal and, once we know how to differentiate one pathology from another, these multimodal assessments might be valid when used as a complement. From the specific acoustic measures used to differentiate, e.g., between AD and PD, it would be also necessary to determine whether the problem is articulatory, cognitive, phonatory or caused by the vocal folds, among others. This means that, although for each pathology there are suitable characteristics for measuring the specific problem, it is necessary to identify what features describe what specific impairment. Even so, on AD research, there are studies underway that are intended to throw light on this point and a differential characterisation regarding other neurodegenerative pathologies, such as Parkinson's Disease or Amyotrophic Lateral Sclerosis.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Credit authorship contribution statement

María Luisa Barragán Pulido: Methodology, Investigation, Data curation, Writing - original draft, Writing - review & editing. **Jesús Bernardino Alonso Hernández:** Conceptualization, Methodology, Resources, Supervision, Project administration, Funding acquisition. **Miguel Ángel Ferrer Ballester:** Resources, Supervision, Project administration, Funding acquisition. **Carlos Manuel Travieso González:** Methodology, Resources, Supervision. **Jiří Mekyska:** Resources, Supervision, Funding acquisition. **Zdeněk Smékal:** Methodology, Supervision, Funding acquisition.

Acknowledgements

References

Abdalla, M., Rudzicz, F., & Hirst, G. (2018). Rhetorical structure and Alzheimer's disease. *Aphasiology, 32*(1), 41–60. doi:10.1080/02687038.2017.1355439.

Aboy, M., Hornero, R., & Abásolo, D. (2006). Interpretation of the Lempel-Ziv complexity measure in the context of biomedical signal analysis. *IEEE Transactions on Biomedical Engineering, 53*(11), 2282–2288. doi:10.1109/TBME.2006.883696.

Ahangar, A., Morteza, S., Fadaki, J., & Sehhati, A. (2018). The comparison of Morpho-Syntactic Patterns Device Comprehension in Speech of Alzheimer and Normal Elderly People. *Zahedan J Res Med Sci.* doi:10.5812/zjrms.9535.

Al-Hameed, S., Benaissa, M., & Christensen, H. (2016). Simple and robust audio-based detection of biomarkers for Alzheimer's disease. In *7th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT)* (pp. 32–36). doi:10.21437/SLPAT.2016-6.

Allison, R. S. (1962). The senile brain: A clinical study. *Posgraduate Medical Journal,* 656.

Alonso, J. B., Cabrera, J., Medina, M., & Travieso, C. M. (2015). New approach in quantification of emotional intensity from the speech signal: emotional temperature. *Expert Systems with Applications, 42*(24), 9554–9564. doi:10.1016/j.eswa.2015.07.062.

Alonso, J. B., Cabrera, J., Travieso, C. M., López-de-Ipiña, K., & Sánchez-Medina, A. (2017). Continuous tracking of the emotion temperature. *Neurocomputing, 255,* 17–25.

Alonso, J. B., De Leon, J., Alonso, I., & Ferrer, M. A. (2001). Automatic detection of pathologies in the voice by HOS based parameters. *EURASIP Journal on Applied Signal Processing, 4,* 275–284.

Altun, H., & Polat, G. (2009). Boosting selection of speech related features to improve performance of multi-class SVMs in emotion detection. *Expert Systems with Applications, 36*(4), 8197–8203. doi:10.1016/J.ESWA.2008.10.005.

Aluísio, S., Cunha, A., & Scarton, C. (2016). Evaluating progression of Alzheimer's disease by regression and classification methods in a narrative language test in Portuguese. In *International conference on computational processing of the Portugues* (pp. 109–114). Springer. doi:10.1007/978-3-319-41552-9_10.

Aluisio, S. M., Cunha, A., Toledo, C., & Scarton, C. (2016). A computational tool for automated language production analysis aimed at dementia diagnosis. *International conference on computational processing of the Portuguese language, XII; demonstration session.* University of Lisbon Retrieved from http://www.producao.usp.br/handle/BDPI/51014.

Alzheimer's Association. (2016a). 2016 Alzheimer's disease facts and figures. *Alzheimer's & Dementia, 12*(4), 459–509. doi:10.1016/j.jalz.2016.03.001.

Alzheimer's Association. (2016b). 2016 Alzheimer's disease facts and figures. *Alzheimer's & Dementia, 12*(4), 459–509. doi:10.1016/j.jalz.2016.03.001.

Alzheimer's Association. (2016c). 2017 Alzheimer's disease facts and figures. *Alzheimer's & Dementia, 13*(4), 459–509. doi:10.1016/j.jalz.2017.02.001.

Alzheimer's Association. (2017). 2017 Alzheimer's disease facts and figures. *Alzheimer's & Dementia, 13*(4), 325–373. doi:10.1016/j.jalz.2017.02.001.

AMI Corpus. (2006). Retrieved May 26, 2018, from http://groups.inf.ed.ac.uk/ami/corpus/.

Asgari, M., Kaye, J., & Dodge, H. (2017). Predicting mild cognitive impairment from spontaneous spoken utterances. *Alzheimer's & Dementia: Translational Research & Clinical Interventions, 3*(2), 219–228. doi:10.1016/j.trci.2017.01.006.

Atance, J. M., Yusta, A. I., & Grupeli, B. G. (2004). Costs study in Alzheimer's disease. *Revista Clínica Española, 204*(2), 64–69. doi:10.1157/13058799.

Bäckman, L., Jones, S., Berger, A.-K., Laukka, J. E., & Small, B. J. (2005). Cognitive impairment in preclinical Alzheimer's disease: a meta-analysis. *Neuropsychology, 19*(4), 520–531.

Baldas, V., Lampiris, C., Capsalis, C., & Koutsouris, D. (2011). Early diagnosis of Alzheimer's type dementia using continuous speech recognition. In *Wireless mobile communication and healthcare. MobiHealth 2010. Lecture notes of the institute for computer sciences, social informatics and telecommunications engineering: 55* (pp. 105–110). Springer. doi:10.1007/978-3-642-20865-2_14.

Balti, H., & Elmaghraby, A. S. (2014). Emotion analysis from speech using temporal contextual trajectories. In *2014 IEEE symposium on computers and communications (ISCC)* (pp. 1–7). IEEE. doi:10.1109/ISCC.2014.6912616.

Barbarotto, R., Capitani, E., Jori, T., Laiacona, M., & Molinari, S. (1998). Picture naming and progression of Alzheimer's disease: an analysis of error types. *Neuropsychologia, 36*(5), 397–405. doi:10.1016/S0028-3932(97)00124-3.

Barret, P. A. (2000). Voice activity detector.

Bateman, R. J., Xiong, C., Benzinger, T. L., Fagan, A. M., Goate, A., Fox, N. C., & Holtzman, D. M. (2012). Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *New England Journal of Medicine, 367*(9), 795–804. doi:10.1056/NEJMoa1202753.

Beltrami, D., Calzà, L., Gagliardi, G., Ghidoni, E., Marcello, N., Favretti, R. R., & Tamburini, F. (2016). Automatic Identification of mild cognitive impairment through the analysis of italian spontaneous speech productions. In *LREC* (pp. 2089–2093). Retrieved from. http://corpora.dslo.unibo.it/People/Tamburini/Pubs/LREC2016_OPLON.pdf.

Bhaduri, S., Das, R., & Ghosh, D. (2016). Non-invasive detection of Alzheimer's disease-multifractality of emotional speech. *Journal of Neurology and Neuroscience, 7*(2). doi:10.21767/2171-6625.100084.

Boots, L., Vugt, M., Knippenberg, R., Kempen, G., & Verhey, F. (2014). A systematic review of Internet-based supportive interventions for caregivers of patients with dementia. *International Journal of Geriatric Psychiatry, 29*(4), 331–344. doi:10.1002/gps.4016.

Boyé, M., Tran, T. M., & Grabar, N. (2014). NLP-oriented contrastive study of linguistic productions of Alzheimer's and control people. In *International Conference on Natural Language Processing. Advances in Natural Language Processing* (pp. 412–424). Springer. doi:10.1007/978-3-319-10888-9_41.

Brabenec, L., Mekyska, J., Galaz, Z., & Rektorova, I. (2017). Speech disorders in Parkinson's disease: early diagnostics and effects of medication and brain stimulation. *Journal of Neural Transmission, 124*(3), 303–334. doi:10.1007/s00702-017-1676-0.

Cambridge Cognition. (2019). CANTAB Mobile | Cambridge Cognition. Retrieved May 15, 2019, from https://www.cambridgecognition.com/products/digital-healthcare-technology/cantab-mobile/.

Caregiving in the U.S. (2015). Retrieved from http://www.caregiving.org/wp-content/uploads/, (2015). Available at:%0AAccessed, 2015/05/2015_CaregivingintheUS_Final-Report-June-4_WEB.pdf.

Carlson, L., & Marcu, D. (2001). *Discourse tagging reference manual. Univ. of Southern California. Information Sciences Institute Tech. Rep.*

Carolinas Conversations Collection - About - Who We Are. (2008). Retrieved May 26, 2018, from http://carolinaconversations.musc.edu/about/who.

Carrié, I, van Kan, GA, Gillette-Guyonnet, S, et al. (2012). Recruitment strategies for preventive trials. The MAPT study (MultiDomain Alzheimer Preventive Trial). *The Journal of Nutrition, Health & Aging, 16*(4), 355–359. doi:10.1007/s12603-012-0046-8.

Castiglioni, P. (2010). What is wrong in Katz's method? Comments on: "a note on fractal dimensions of biomedical waveforms. *Computers in Biology and Medicine, 40*(11–12), 950–952.

Chavhan, Y. D., Yelure, B. S., & Tayade, K. N. (2015). Speech emotion recognition using RBF kernel of LIBSVM. In *2015 2nd international conference on electronics and communication systems (ICECS)* (pp. 1132–1135). IEEE. doi:10.1109/ECS.2015.7124760.

Chen, Z., Ivanov, P. C., Hu, K., & Stanley, H. E. (2002). Effect of nonstationarities on detrended fluctuation analysis. *Physical Review E, 65*(4), 041107. doi:10.1103/PhysRevE.65.041107.

Connolly, A., Gaehl, E., Martin, H., Morris, J., & Purandare, N. (2011). Underdiagnosis of dementia in primary care: variations in the observed prevalence and comparisons to the expected prevalence. *Aging & Mental Health, 15*(8), 978–984. doi:10.1080/13607863.2011.596805.

Corpus Linguistics and Linguistic Theory. (2011). 7 (1), 143–161. doi:10.1515/cllt.2011.007.

Cowie, R., & Cornelius, R. R. (2003). Describing the emotional states that are expressed in speech. *Speech Communication, 40*(1–2), 5–32. doi:10.1016/S0167-6393(02)00071-7.

Cuetos-Vega, F., Menéndez-González, M., Calatayud-Noguera, T., Cuetos-Vega, F., Menéndez-González, M., & Calatayud-Noguera, T. (2007). Descripción de un nuevo test para la detección precoz de la enfermedad de Alzheimer. *Revue Neurologique, 44*(8), 469–474 Retrieved from https://s3.amazonaws.com/academia.edu.documents/33107025/Descripcion_de_un_nuevo_test_para_la_deteccion_precoz_de_la_EA.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1524320921&Signature=vQAOVzwLFkwEhAki9NIr%2BSPLtYQ%3D&response-content-disposition=inli.

Cummings, JL, & Doody, R, C. C. (2007). Disease-modifying therapies for Alzheimer disease: challenges to early intervention. *Neurology, 69*(16), 1622–1634. doi:10.1212/01.wnl.0000295996.54210.69.

Czaja, S. J., Loewenstein, D., Schulz, R., Nair, S. N., & Perdomo, D. (2013). A videophone psychosocial intervention for dementia caregivers. *The American Journal of Geriatric Psychiatry, 21*(13), 1071–1081. doi:10.1016/j.jagp.2013.02.019.

de Lizarduy, U. M., Salomón, P. C., Vilda, P. G., Torres, M. E., & de Ipiña, K. L. (2017). ALZUMERIC: A decision support system for diagnosis and monitoring of cognitive impairment. *Loquens, 4*(1). doi:10.3989/loquens.2017.037.

Dem@Care Project - Project. (2011). Retrieved May 26, 2018, from http://www.demcare.eu/.

DementiaBank | TalkBank. (2007). Retrieved May 26, 2018, from https://dementia.talkbank.org/access/.

Deramecourt, V., Lebert, F., Debachy, B., Mackowiak-Cordoliani, M. A., Bombois, S., Kerdraon, O., & Pasquier, F. (2010). Prediction of pathology in primary progressive language and speech disorders. *Neurology, 74*(1), 42–49 http://doi.org/10.1212/WNL. 0b013e3181c7198e.

Ekman, P. E., & Davidson, R. J. (1994). In O. U. Press (Ed.). Oxford University Press Retrieved from http://psycnet.apa.org/record/1995-97541-000.

EPAD | European Prevention of Alzheimer's Dementia Consortium. (2015). Retrieved May 13, 2019, from http://ep-ad.org/.

Escobar, L., et al. (2010). Calidad de vida del cuidador familiar y dependencia del paciente con Alzheimer. *Avances En Enfermería, 28*(1), 116–128.

Esteller, R., et al. (2001). A comparison of waveform fractal dimension algorithms. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, 48*(2), 177–183. doi:10.1109/81.904882.

Eyben, F., Weninger, F., Gross, F., & Schuller, B. (2013). Recent developments in opensmile, the Munich open-source multimedia feature extractor. In *Proceedings of the 21st ACM international conference on multimedia* (pp. 835–838). doi:10.1145/2502081.2502224.

Ferri, C., Prince, M., Brayne, C., Brodaty, H., Fratiglioni, L., & Ganguli, M. (2005). Global prevalence of dementia: a Delphi consensus study. *Lancet, 366*(9503), 2112–2117. doi:10.1016/S0140-6736(05)67889-0.

Folstein, M. F., Robins, L. N., & Helzer, J. E. (1983). The mini-mental state examination. *Archives of General Psychiatry, 40*(7), 812. doi:10.1001/archpsyc.1983.01790060110016.

Forbes-McKay, K. E., & Venneri, A. (2005). Detecting subtle spontaneous language decline in early Alzheimer's disease with a picture description task. *Neurological Sciences, 26*(4), 243–254. doi:10.1007/s10072-005-0467-9.

Fraser, K., Meltzer, J., & Rudzicz, F. (2016). Linguistic features identify Alzheimer's disease in narrative speech. *Alzheimer's Disease, 49*(2), 407–422. doi:10.3233/JAD-150520.

Freedman, V. A., & Spillman, B. C. (2014). Disability and care needs of older americans by dementia status: An analysis of the 2011 national health and aging trends study. U.S. Retrieved from https://aspe.hhs.gov/report/disability-and-care-needs-older-americans-analysis-2011-national-health-and-aging-trends-study.

Friedman, E. M., Shih, R. A., Langa, K. M., & Hurd, M. D. (2015). US prevalence and predictors of informal caregiving for dementia. *Health Affairs, 34*(10), 1637–1641. doi:10.1377/hlthaff.2015.0510.

Frijda, N. H. (1987). Emotion, cognitive structure, and action tendency. *Cognition & Emotion, 1*(2), 115–143. doi:10.1080/02699938708408043.

G8 Health and Science Ministers. (2013). G8 Dementia summit declaration. Retrieved August 28, 2017, from https://www.gov.uk/government/publications/g8-dementiasummit-%0Aagreements/g8-dementia-summit-declaration.

Gómez-Vilda, P. (2014). Biomechanical characterization of phonation in Alzheimer's Disease. *3rd IEEE International work-conference on bioinspired intelligence*. doi:10.1109/IWOBI.2014.6913931.

Gonnerman, L. M., Aronoff, J. M., Almor, A., Kempler, D., & Andersen, E. S. (2004). From beetle to bug: progression of error types in naming in Alzheimer's disease. In *Proceedings of the annual meeting of the cognitive science society: 26* (p. 26).

González, D. C., Luan Ling, L., & Violaro, F. (2012). Analysis of the Multifractal Nature of Speech Signals. In *IberoAmerican congress on pattern recognition* (pp. 740–748). doi:10.1007/978-3-642-33275-3_91.

Gósy, M. (2013). BEA – A multifunctional Hungarian spoken language database. *Phonetician, 105*, 50–61. Retrieved from http://real.mtak.hu/8410/.

Gosztolya, G., Tóth, L., Grósz, T., Vincze, V., Hoffmann, I., Szatlóczki, G., … Kálmán, J. (2016). Detecting mild cognitive impairment from spontaneous speech by correlation-based phonetic feature selection. In *INTERSPEECH 2016* (pp. 107–111). doi:10.21437/Interspeech.2016-384.

Goudbeek, M., & Scherer, K. (2010). Beyond arousal: Valence and potency/control cues in the vocal expression of emotion. *The Journal of the Acoustical Society of America, 128*(3), 1322. doi:10.1121/1.3466853.

Graovac, J., Kovacevic, J., & Lazetic, G. P. (2016). Machine learning-based approach to help diagnosing Alzheimer's disease through spontaneous

speech analysis. In *Belgrade BioInformatics Conference 2016* (p. 111). Retrieved from. https://www.researchgate.net/profile/Jovana_Kovacevic3/publication/304659180_Machine_learning-based_approach_to_help_diagnosing_Alzheimer%27s_disease_through_spontaneous_speech_analysis/links/57764e6908aeb9427e2772d1/Machine-learning-based-approach-to-hel.

Guix, J. L. M. (2011). Papel de los biomarcadores en el diagnóstico precoz de la enfermedad de Alzheimer. *Revista Española de Geriatría y Gerontología, 46*, 39–41. doi:10.1016/j.regg.2011.10.008.

Hane, F. T., Robinson, M., Lee, B. Y., Bai, O., & Leonenko, Z. (2017). Recent progress in Alzheimer's disease research, Part 3 : diagnosis and treatment. *Journal of Alzheimer's Disease, 57*(3), 645–665. doi:10.3233/JAD-160907.

Harimi, A., Shahzadi, A., & Ahmadyfard, A. (2014). Recognition of emotion using non-linear dynamics of speech. In *7'th International Symposium on Telecommunications (IST'2014)* (pp. 446–451). IEEE. doi:10.1109/ISTEL.2014.7000745.

Hausdorff, J. M., Purdon, P. L., Peng, C. K., Ladin, Z., Wei, J. Y., & Goldberger, A. L. (1996). Fractal dynamics of human gait: stability of long-range correlations in stride interval fluctuations. *Journal of Applied Physiology, 80*(5), 1448–1457. doi:10.1152/jappl.1996.80.5.1448.

Hebert, L. E., Scherr, P. A., McCann, J. J., Beckett, L. A., & Evans, D. A. (2001). Is the risk of developing Alzheimer's disease greater for women than for men. *Hebert, L. E., Scherr, P. A., McCann, J. J., Beckett, L. A., & Evans, D. A.. American Journal of Epidemiology, 153*(2), 132–136. doi:10.1093/aje/153.2.132.

Henríquez, P., Alonso, J. B., Ferrer, M. A., Travieso, C. M., Godino-Llorente, J. I., & Díaz-De-María, F. (2009). Characterization of healthy and pathological voice through measures based on nonlinear dynamics. *IEEE Transactions on Audio, Speech, and Language Processing, 17*(6), 1186–1195. doi:10.1109/TASL.2009.2016734.

Herisa Khadivi, H., Seyed Aghazadehb, B., & Nikkhah-Bahramic, M. (2009). Optimal feature selection for the assessment of vocal fold disorders. *Computers in Biology and Medicine, 39*(10), 860–868. doi:10.1016/j.compbiomed.2009.06.014.

Hernández-Domíngue, L., García-Canó, E., Ratt, S., & Sierra-Martínez, G. (2016). Detection of Alzheimer's disease based on automatic analysis of common objects descriptions. In *Proceedings of the 7th Workshop on Cognitive Aspects of Computational Language Learning* (pp. 10–15). Retrieved from. http://www.aclweb.org/anthology/W16-1902.

Hernández-Domínguez, L., Ratté, S., Sierra-Martínez, G., & Roche-Bergua, A. (2018). Computer-based evaluation of Alzheimer's disease and mild cognitive impairment patients during a picture description task. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 10*, 260–268. doi:10.1016/j.dadm.2018.02.004.

Higuchi, T. (1988). Approach to an irregular time series on the basis of the fractal theory. *Physica D: Nonlinear Phenomena, 31*(2), 277–283. Retrieved from https://www.sciencedirect.com/science/article/pii/0167278988900814.

Hoffmann, I., Nemeth, D., Dye, C. D., Pákáski, M., Irinyi, T., & Kálmán, J. (2010). Temporal parameters of spontaneous speech in Alzheimer's disease. *International Journal of Speech-Language Pathology, 12*(1), 29–34 http://www.ncbi.nlm.nih.gov/pubmed/20380247.

Huang, Y. X., Schmitt, F. G., Hermand, J.-P., Gagne, Y., Lu, Z. M., & Liu, Y. L. (2011). Arbitrary-order Hilbert spectral analysis for time series possessing scaling statistics: Comparison study with detrended fluctuation analysis and wavelet leaders. *Physical Review E, 84*(1), 016208. doi:10.1103/PhysRevE.84.016208.

Hurd, M. D. (2013). Monetary Costs of Dementia in the United States. *The New England Journal of Medicine, 368*(14), 1326–1334. doi:10.1056/NEJMsa1204629.

IBM Watson, I. C. solutions. (2016). SimpleC | IBM. Retrieved May 15, 2019, from https://www.ibm.com/case-studies/w796019n50088s93.

Jack, C. R., Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W., & Trojanowski, J. Q. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology, 9*(1), 119–128. doi:10.1016/S1474-4422(09)70299-6.

Jack, C. R., Jr Petersen, R. C., Grundman, M., Jin, S., Gamst, A., & Ward, C. P. (2008). Longitudinal MRI findings from the vitamin E and donepezil treatment study for MCI. *Neurobiology of Aging, 29*(9), 1285–1295. doi:10.1016/j.neurobiolaging.2007.03.004.

Jack, C. R., Jr Wiste, H. J., Vemuri, P., Weigand, S. D., Senjem, M. L, Zeng, G., & Weiner, M. W. (2010). Brain beta-amyloid measures and magnetic resonance imaging atrophy both predict time-to-progression from mild cognitive impairment to Alzheimer's disease. *Brain, 133*(11), 3336–3348. doi:10.1093/brain/awq277.

Jayawardena, A. W., Xu, P., & Li, W. K. (2010). Modified correlation entropy estimation for a noisy chaotic time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science, 20*(2), 023104. doi:10.1063/1.3382013.

Mekyska, Jiri, Janousova, Eva, Gomez-Vilda, Pedro, Smekal, Zdenek, Rektorova, Irena, Eliasova, Ilona, … López-de-Ipiña, Karmele (2015). Robust and complex approach of pathological speech signal analysis. *Neurocomputing, 167*, 94–111. doi:10.1016/j.neucom.2015.02.085.

Johnson, S. C., Koscik, R. L., Jonaitis, E. M., Clark, L. R., Mueller, K. D., Berman, S. E., … Sager, M. A. (2018). The Wisconsin registry for Alzheimer's prevention: A review of findings and current directions. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 10*, 130–142. doi:10.1016/j.dadm.2017.11.007.

Kantelhardt, J. W., Koscielny-Bunde, E., Rego, H. H., Havlin, S., & Bunde, A. (2001). Detecting long-range correlations with detrended fluctuation analysis. *Physica A: Statistical Mechanics and Its Applications, 295*, 441–454. doi:10.1016/S0378-4371(01)00144-3.

Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Havlin, S., Bunde, A., & Stanley, H. E. (2002). Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and Its Applications, 316*

(1–4), 87–114. Retrieved from https://www.sciencedirect.com/science/article/pii/S0378437102013833.

Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *Boston naming test.* Pro-Ed.

Kato, S., Homma, A., & Sakuma, T. (2018). Easy screening for Mild Alzheimer's disease and mild cognitive impairment from elderly speech. *Current Alzheimer Research, 15*(2), 104–110. doi:10.2174/1567205014666171120144343.

Katz, M. (1988). Fractals and the analysis of waveforms. *Computers in Biology and Medicine, 18*(3), 145–156.

Khachaturian, Z. (1985). Diagnosis of Alzheimer's disease. *Archives of Neurology, 42*(11), 1097–1105. doi:10.1001/archneur.1985.04060100083029.

Khodabakhsh, A., & Demiroglu, C. (2015). Analysis of speech-based measures for detecting and monitoring Alzheimer's disease. In *Data Mining in Clinical Medicine* (pp. 159–173). Humana Press. doi:10.1007/978-1-4939-1985-7_11.

Khodabakhsh, Ali, Yesil, F., Guner, E., & Demiroglu, C. (2015). Evaluation of linguistic and prosodic features for detection of Alzheimer's disease in Turkish conversational speech. *EURASIP Journal on Audio, Speech, and Music Processing, 2015*(1), 9. doi:10.1186/s13636-015-0052-y.

ki-elements. (2018). Δelta - Digital neurocognitive testing. Retrieved May 15, 2019, from https://ki-elements.de/.

Kim, Y., Lee, H., & Provost, E. M. (2013). Deep learning for robust feature generation in audiovisual emotion recognition. In *Acoustics, speech and signal processing (ICASSP), 2013 IEEE international conference* (pp. 3687–3691). doi:10.1109/ICASSP.2013.6638346.

Kivipelto, M., Solomon, A., & Ahtiluoto, S. (2013). The finnish geriatric intervention study to prevent cognitive impairment and disability (FINGER): study design and progress. *Alzheimers Dement, 9*(6), 657–665. doi:10.1016/j.jalz.2012.09.012.

König, A., Satt, A., David, R., & Robert, P. (2016). Innovative voice analytics for the assessment and monitoring of cognitive decline in people with dementia and mild cognitive impairment. *Alzheimer's & Dementia, 12*(7), 363. doi:10.1016/j.jalz.2016.06.673.

König, A., Satt, A., Sorin, A., Hoory, R., Toledo-Ronen, O., Derreumaux, A., & David, R. (2015). Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 1*(1), 112–124. doi:10.1016/J.DADM.2014.11.012.

Kwon, O., Chan, K., et al. (2003). Emotion recognition by speech signals. *Eighth European Conference on Speech Communication and Technology.*

Lancioni, G., Singh, N., O'Reilly, M., Sigafoos, J., D'Amico, F., Pinto, K., & Caffò, A. (2017). A technology-aided program for helping persons with Alzheimer's disease perform daily activities. *Journal of Enabling Technologies, 11*(3), 85–91. doi:10.1108/JET-03-2017-0011.

Laske, C., Sohrabi, H. R., Frost, S. M., López-de-Ipiña, K., Garrard, P., & Buscema, M. (2015). Innovative diagnostic tools for early detection of Alzheimer's disease. *Alzheimer's & Dementia, 11*(5), 561–578. doi:10.1016/j.jalz.2014.06.004.

Laukka, P. (2004). *Vocal expression of emotion : discrete-emotions and dimensional accounts.* Acta Universitatis Upsaliensis, Uppsala universitet.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436. Retrieved from https://www.nature.com/articles/nature14539.

Lee, C., & Narayanan, S. (2003). Emotion recognition using a data-driven fuzzy inference system. *Eighth European conference on speech communication and technology.*

Lenguaje, E. L., La, E. N., & Alzheimer, E. D. E. (1988). El lenguaje en la enfermedad de alzheimer. *Revista de Logopedia, Foniatría y Audiología, 8*(4), 199–205. doi:10.1016/S0214-4603(88)75441-8.

Little, M. A., McSharry, P. E., Roberts, S. J., Costello, D. A., & Moroz, I. M. (2007). Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *BioMedical Engineering OnLine, 6*(1), 23. doi:10.1186/1475-925X-6-23.

Little, M., McSharry, P., & Hunter, E. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Transactions on Biomedical Engineering, 56*(4), 1015–1022. doi:10.1109/TBME.2008.2005954.

López-de-Ipiña, K., Alonso-Hernández, J. B., Solé-Casals, J., Travieso-González, C. M., Ezeiza, A., Faúndez-Zanuy, M., … Beitia, B. (2015). Feature selection for automatic analysis of emotional response based on nonlinear speech modeling suitable for diagnosis of Alzheimer's disease. *Neurocomputing, 150*, 392–401. doi:10.1016/J.NEUCOM.2014.05.083.

Lopez-de-Ipiña, K., Alonso, J. B., Barroso, N., Solé-Casals, J., Ecay-Torres, M., Martinez-Lage, P., … Travieso, C. M. (2014). Spontaneous speech and emotional response modeling based on one-class classifier oriented to alzheimer disease diagnosis. In *XIII Mediterranean conference on medical and biological engineering and computing* (pp. 571–574). Springer. doi:10.1007/978-3-319-00846-2_141.

López-de-Ipiña, K., Alonso, J. B., Solé-Casals, J., Barroso, N., Henriquez, P., Faundez-Zanuy, M., & Eguiraun, H. (2015). On automatic diagnosis of Alzheimer's disease based on spontaneous speech analysis and emotional temperature. *Cognitive Computation, 7*(1), 44–55. doi:10.1007/s12559-013-9229-9.

Lopez-de-Ipina, K., Martinez-de-Lizarduy, U., Barroso, N., Ecay-Torres, M., Martinez-Lage, P., Torres, F., & Faundez-Zanuy, M. (2015). Automatic analysis of Categorical Verbal Fluency for Mild Cognitive imparment detection: A non-linear language independent approach. In *Bioinspired intelligence (IWOBI), 2015 4th international work conference IEEE* (pp. 101–104). San Sebastian. doi:10.1109/IWOBI.2015.7160151.

Lopez-de-Ipina, K., Martinez-de-Lizarduy, U., Calvo, P. M., Mekyska, J., Beitia, B., & Barroso, N. (2018). Advances on automatic speech analysis for early detection of Alzheimer disease: a non-linear multi-task approach. *Current Alzheimer Research, 15*(2), 139–148. doi:10.2174/1567205014666171120143800.

López-de-Ipiña, Karmele, Alonso, J.-B., Travieso, C., Solé-Casals, J., Egiraun, H., Faundez-Zanuy, M., … Lizardui, U. (2013). On the selection of non-invasive

methods based on speech analysis oriented to automatic Alzheimer disease diagnosis. *Sensors, 13*(5), 6730–6745. doi:10.3390/s130506730.

López-de-Ipiña, Karmele, Alonso, J. B., Barroso, N., Faundez-Zanuy, M., Ecay, M., Solé-Casals, J., … Ezeiza, A. (2012). New approaches for Alzheimer's disease diagnosis based on automatic spontaneous speech analysis and emotional temperature. In *Ambient Assisted Living and Home Care. IWAAL 2012. Lecture Notes in Computer Science: 7657* (pp. 407–414). Springer. Springer, Berlin, Heidelberg. doi:10.1007/978-3-642-35395-6_55.

Lopez-de-Ipiña, Karmele, Alonso, J. B., Solé-Casals, J., Barroso, N., Faundez-Zanuy, M., Ecay-Torres, M., … Estanga, A. (2012). Alzheimer disease diagnosis based on automatic spontaneous speech analysis. In *International Joint Conference on Computational Intelligence*. In *IJCCI 2012: proceedings of the 4th international joint conference on computational intelligence: Barcelona, Spain: 5-7 October, 2012* (pp. 698–705). SciTePress. Retrieved from. http://dspace.uvic.cat/handle/10854/2482.

López-de-Ipiña, Karmele, Solé-Casals, J., Eguiraun, H., Alonso, J. B., Travieso, C. M., Ezeiza, A., … Beitia, B. (2015). Feature selection for spontaneous speech analysis to aid in Alzheimer's disease diagnosis: A fractal dimension approach. *Computer Speech & Language, 30*(1), 43–60. doi:10.1016/J.CSL.2014.08.002.

Lopez, O. L., Schwam, E., Cummings, J., Gauthier, S., Jones, R., Wilkinson, D., & Schindler, R. (2010). Predicting cognitive decline in Alzheimer's disease: an integrated analysis. *Alzheimer's & Dementia, 6*(6), 431–439. doi:10.1016/j.jalz.2010.04.003.

López, V. N. (2016). Evaluación del discurso de las personas con Enfermedad de Alzheimer: una revisión. Retrieved from https://ruidera.uclm.es/xmlui/handle/10578/14079.

Lorenz, K., Freddolino, P. P., Comas-herrera, A., Knapp, M., & Damant, J. (2017). Technology-based tools and services for people with dementia and carers : Mapping technology onto the dementia care pathway. *Dementia*, 1471301217691617. doi:10.1177/1471301217691617.

Luz, S. (2017). Longitudinal monitoring and detection of Alzheimer's type dementia from spontaneous speech data. In *Computer-Based Medical Systems (CBMS), 2017 IEEE 30th International Symposium on IEEE* (pp. 45–46). doi:10.1109/CBMS.2017.41.

Macwhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology, 25*(11), 1286–1307. doi:10.1080/02687038.2011.589893.

Malekzadeh, G., Arsalan, G., & Shahabi, M. (2009). A comparative study on the use of cohesion devices by normal age persian natives and those suffering from Alzheimer's disease. *Journal of Medical Science of Islamic Azad University of Mashhad, 5*(3), 153–161.

Mandelbrot, B. (1982). *Fractal geometry of nature.* New York: WH freeman.: Times Books.

Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse, 8*(3), 243–281. Retrieved from https://www.degruyter.com/view/j/text.1.1988.8.issue-3/text.1.1988.8.3.243/text.1.1988.8.3.243.xml.

Maresova, P., Valis, M., Hort, J., & Kuca, K. (2015). Alzheimer's disease and language impairments: social intervention and medical treatment. *Clinical Interventions in Aging, 10*, 1401–1408. doi:10.2147/CIA.S89714.

Martínez-Sánchez, F., Meilán, J. J. G., Vera-Ferrandiz, J. A., Carro, J., Pujante-Valverde, I. M., Ivanova, O., & Carcavilla, N. (2017). Speech rhythm alterations in Spanish-speaking individuals with Alzheimer's disease. *Aging, Neuropsychology, and Cognition, 24*(4), 418–434. doi:10.1080/13825585.2016.1220487.

Martínez, C. A., & Cruz, A. B. (2005). Emotion recognition in non-structured utterances for human-robot interaction. *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication*, 19–23. doi:10.1109/ROMAN.2005.1513750.

Matias-Guiu, J., & García-Ramos, R. (2013). Primary progressive aphasia: from syndrome to disease. *Neurología (English Edition), 28*(6), 366–374. doi:10.1016/j.nrleng.2012.04.018.

McKhann, G, Drachman, D, Folstein, M, Katzman, R, P. D., & EM. , S. (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. *Neurology, 34*(7), 939–944. doi:10.1212/WNL.34.7.939.

McKhann, G., Knopman, D., Chertkow, H., Hyman, B., Jack, C., Kawas, C., … Phelps, C. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia, 7*(3), 263–269. doi:10.1016/j.jalz.2011.03.005.

Meilan, J. J., Martinez-Sanchez, F., Carro, J., Carcavilla, N., & Ivanova, O. (2018). Voice markers of lexical access in mild cognitive impairment and Alzheimer's Disease. *Current Alzheimer Research, 15*(2), 111–119. doi:10.2174/1567205014666170829112439.

Meilán, Juan J. G., Martínez-Sánchez, F., Carro, J., Sánchez, J. A., & Pérez, E. (2012). Acoustic markers associated with impairment in language processing in Alzheimer's Disease. *The Spanish Journal of Psychology, 15*(02), 487–494. doi:10.5209/rev_SJOP.2012.v15.n2.38859.

Meilán, Juan José G, Martínez-Sánchez, F., Carro, J., López, D. E., Millian-Morell, L., & Arana, J. M. (2014). Speech in Alzheimer's disease: can temporal and acoustic parameters discriminate dementia. *Dementia and Geriatric Cognitive Disorders, 37*(5–6), 327–334. doi:10.1159/000356726.

Mirheidari, B., Blackburn, D., et al. (2016). Diagnosing people with dementia using automatic conversation analysis. In *Proceedings of Interspeech, ISCA* (pp. 1220–1224). doi:10.21437/Interspeech.2016-857.

Mirheidari, B., Blackburn, D., et al. (2017). Toward the automation of diagnostic conversation analysis in patients with memory complaints. *Journal of Alzheimer's Disease, 58*(2), 373–387. Retrieved from https://content.iospress.com/articles/journal-of-alzheimers-disease/jad160507.

Mirzaei, S., El Yacoubi, M., Garcia-Salicetti, S., Boudy, J., Muvingi, C. K. S., & Cristancho-Lacroix, V. (2017). Automatic speech analysis for early Alzheimer's disease diagnosis. In *JETSAN 2017: 6e Journées d'Etudes sur la Télésant* (pp. 114–116). Retrieved from. https://hal.archives-ouvertes.fr/hal-01618834/document.

Mirzaei, S., Yacoubi, M., et al. (2018). Two-stage feature selection of voice parameters for early Alzheimer's disease prediction. *IRBM, 39*(6), 430–435. Retrieved from https://www.sciencedirect.com/science/article/pii/S1959031818302847.

Morris, J. C. (2005). Early-stage and preclinical Alzheimer disease. *Alzheimer Disease and Associated Disorders, 19*(3), 163.

Mueller, K. D., Koscik, R. L., Hermann, B. P., Johnson, S. C., & Turkstra, L. S. (2018). Declines in connected language are associated with very early mild cognitive impairment: results from the wisconsin registry for Alzheimer's prevention. *Frontiers in Aging Neuroscience, 9*, 437. doi:10.3389/fnagi.2017.00437.

Nasrolahzadeh, M, Mohammadpoori, Z., & Haddadnia, J. (2016). Analysis of mean square error surface and its corresponding contour plots of spontaneous speech signals in Alzheimer's disease with adaptive wiener filter. *Computers in Human Behavior, 61*, 364–371. doi:10.1016/J.CHB.2016.03.031.

Nasrolahzadeh, Mahda, Mohammadpoory, Z., & Haddadnia, J. (2016). A novel method for early diagnosis of Alzheimer's disease based on higher-order spectral estimation of spontaneous speech signals. *Cognitive Neurodynamics, 10*(6), 495–503. doi:10.1007/s11571-016-9406-0.

Nebes, R. D., Brady, C. B., & Huff, F. J. (1989). Automatic and attentional mechanisms of semantic priming in Alzheimer's disease. *J. Clin. Exp. Neuropsychol., 11*(2), 219–230. doi:10.1080/01688638908400884.

OECD. (2015). Global action to drive innovation in Alzheimer's disease and other dementias: connecting research, regulation and access. *2nd Lausanne Workshop* Retrieved from http://www.oecd.org/sti/%0Abiotech/alzheimers-dementia-research-workshop-2015.htm.

Orozco-Arroyave, J., & Murillo-Rendón, S. (2011). Automatic selection of acoustic and non-linear dynamic features in voice signals for hypernasality detection. In *Twelfth annual conference of the international speech communication association INTERSPEECH* (pp. 529–532). Retrieved from. http://www.isca-speech.org/archive/interspeech_2011/i11_0529.html.

Oświęcimka, P., Kwapień, J., & Drożdż, S. (2006). Wavelet versus detrended fluctuation analysis of multifractal structures. *Physical Review E, 74*(1), 016103. doi:10.1103/PhysRevE.74.016103.

Petersen, R. C., Thomas, R. G., Grundman, M., Bennett, D., Doody, R., Ferris, S., & Pfeiffer, E. (2005). Vitamin E and donepezil for the treatment of mild cognitive impairment. *New England Journal of Medicine, 352*(23), 2379–2388. doi:10.1056/NEJMoa050151.

Picard, R. R., & Cook, R. D. (1984). Cross-validation of regression models. *Journal of the American Statistical Association, 79*(387), 575–583. doi:10.1080/01621459.1984.10478083.

Plassman, B., Langa, K., et al. (2007). Prevalence of dementia in the United States: The aging, demographics, and memory study. *Neuroepidemiology, 29*(1–2), 125–132. doi:10.1159/000109998.

Powell, J., Chiu, T., & Eysenbach, G. (2008). A systematic review of networked technologies supporting carers of people with dementia. *Journal of Telemedicine and Telecare, 14*(3), 154–156. doi:10.1258/jtt.2008.003018.

Prince, M., Bryce, R., Albanese, E., Wimo, A., Ribeiro, W., & Ferri, C. P. (2013). The global prevalence of dementia : A systematic review and metaanalysis. *Alzheimer's & Dementia, 9*(1), 63–75 e2. doi:10.1016/j.jalz.2012.11.007.

Proyectos Fundación CITA Alzheimer. (2017). Retrieved May 26, 2018, from http://www.cita-alzheimer.org/investigacion/proyectos.

Pye, C., & MacWhinney, B. (1994). The CHILDES project: tools for analyzing talk. *Language, 70*(1), 156. doi:10.2307/416745.

Rath, T. M., & Manmatha, R. (2007). Word spotting for historical documents. *International Journal of Document Analysis and Recognition, 9*(2), 139–152. Retrieved from http://ciir.cs.umass.edu/pubfiles/mm-409.pdf.

Reiman, E. M., Langbaum, J. B. S., Fleisher, A. S., Caselli, R. J., Chen, K., Ayutyanont, N., … Tariot, P. N. (2011). Alzheimer's prevention initiative: A plan to accelerate the evaluation of presymptomatic treatments. *Journal of Alzheimer's Disease, 26*(s3), 321–329. doi:10.3233/JAD-2011-0059.

Rentoumi, V., Paliouras, G., Danasi, E., Arfani, D., Fragkopoulou, K., Varlokosta, S., & Papadatos, S. (2017). Automatic detection of linguistic indicators as a means of early detection of Alzheimer's disease and of related dementias: A computational linguistics analysis. In *Cognitive infocommunications (CogInfoCom), 8th IEEE international conference* (pp. 33–38). doi:10.1109/CogInfoCom.2017.8268212.

Renzi, A. De, & Neurology, L. V.-B. (1962). Token test: A sensitive test to detect receptive disturbances in aphasics. *Brain: A Journal of Neurology, 85*, 665–678. doi:10.1093/brain/85.4.665.

Richard, E, Van den Heuvel, E, Moll van Charante, EP, et al. (2009). Prevention of dementia by intensive vascular care (PreDIVA): a cluster-randomized trial in progress. *Alzheimer's Dissease Association Disorders, 23*(3), 198–204. doi:10.1097/WAD.0b013e31819783a4.

Richard, E., Jongstra, S., & Soininen, H. (2015). Healthy ageing through internet counselling in the elderly: Project. *HATICE. BMJ Publishing Group Limited, 6*(6). doi:10.1136/bmjopen-2015-010806.

Rockwood, K., Graham, J. E., & Fay, S. (2002). Goal setting and attainment in Alzheimer's disease patients treated with donepezil. *Journal of Neurology, Neurosurgery & Psychiatry, 73*(5), 500–507. doi:10.1136/jnnp.73.5.500.

Roy, D., & Pentland, A. (1996). Automatic spoken affect classification and analysis. In *Automatic face and gesture recognition, proceedings of the second international conference* (pp. 363–367). doi:10.1109/AFGR.1996.557292.

Sadeghian, R., Schaffer, J., & Zahorian, S. A. (2017). Speech processing approach for diagnosing dementia in an early stage. In *Proceedings Interspeech 2017* (pp. 2705–2709). Retrieved from. http://www.isca-speech.org/archive/Interspeech_2017/pdfs/1712.PDF.

Satt, A., Hoory, R., König, A., Aalten, P., & Robert, P. H. (2014). Speech-based automatic and robust detection of very early dementia. In *Fifteenth annual conference of the international speech communication association. INTERSPEECH-2014* (pp. 2538–2542). Retrieved from. https://mazsola.iit.uni-miskolc.hu/~czap/letoltes/IS14/IS2014/PDF/AUTHOR/IS140749.PDF.

Satt, Aharon, Sorin, A., Toledo-Ronen, O., Barkan, O., Kompatsiaris, I., Kokonozi, A., & Tsolaki, M. (2013). Evaluation of speech-based protocol for detection of early-stage dementia. In *INTERSPEECH 2013* (pp. 1692–1696). Retrieved from. https://www.isca-speech.org/archive/archive_papers/interspeech_2013/i13_1692.pdf.

Schuller, B., Batliner, A., Seppi, D., Steidl, S., Vogt, T., & Wagner, J. (2007). The relevance of feature type for the automatic classification of emotional user states: Low level descriptors and functionals. *Eighth annual conference of the international speech communication association INTERSPEECH* Retrieved from http://www.isca-speech.org/archive/interspeech_2007/i07_2253.html.

Schuller, B., Steidl, S., Batliner, A., Burkhardt, F., Devillers, L., Müller, C., & Narayanan, S. (2010). The INTERSPEECH 2010 paralinguistic challenge. In *Proceedings INTERSPEECH 2010* (pp. 2794–2797). Retrieved from. https://mediatum.ub.tum.de/doc/1082470/file.pdf.

Schuller, B., Steidl, S., Batliner, A., Nöth, E., Vinciarelli, A., Burkhardt, F., & Mohammadi, G. (2012). The interspeech 2012 speaker trait challenge. *Thirteenth annual conference of the international speech communication association* Retrieved from http://www.isca-speech.org/archive/archive_papers/interspeech_2012/i12_0254.pdf.

Schuller, B., Steidl, S., Batliner, A., Schiel, F., & Krajewski, J. (2011). The INTERSPEECH 2011 speaker state challenge. *Twelfth annual conference of the international speech communication association* Retrieved from http://www.isca-speech.org/archive/interspeech_2011/i11_3201.html.

Schulz, R., & Beach, S. R. (1999). Caregiving as a risk factor for mortality the caregiver health effects study. *JAMA, 282*(23), 2215–2219. doi:10.1001/jama.282.23.2215.

Serrano, E., & Figliola, A. (2009). Wavelet leaders: a new method to estimate the multifractal singularity spectra. *Physica A: Statistical Mechanics and Its Applications, 388*(14), 2793–2805. doi:10.1016/j.physa.2009.03.043.

Seshadri, S., Wolf, P., et al. (1997). Lifetime risk of dementia and Alzheimer's disease the impact of mortality on risk estimates in the Framingham study. *Neurology, 49*(6), 1498–1504. doi:10.1212/WNL.49.6.1498.

Shao, J., MacCallum, J., Zhang, Y., & Sprecher, A. (2010). Acoustic analysis of the tremulous voice: assessing the utility of the correlation dimension and perturbation parameters. *Journal of Communication Disorders, 43*(1) 35.44. doi:10.1016/j.jcomdis.2009.09.001.

Sirts, K., Piguet, O., & Johnson, M. (2017). Idea density for predicting Alzheimer's disease from transcribed speech. Retrieved from ArXiv:1706.04473.

Sjogren, T., Sjogren, H., & Lindgren, A. G. (1952). Morbus Alzheimer and morbus Pick; a genetic, clinical and patho-anatomical study. *Acta Psychiatrica et Neurologica Scandinavica. Supplementum, 82*, 1–152. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/13171126.

Snyder, P. J., Jackson, C. E., Petersen, R. C., Khachaturian, A. S., Kaye, J., Albert, M. S., & Weintraub, S. (2011). Assessment of cognition in mild cognitive impairment: A comparative study. *Alzheimer's & Dementia, 7*(3), 338–355. doi:10.1016/j.jalz.2011.03.009.

Spence, S. (1995). Descartes' error: Emotion, reason and the human brain. *BMJ, 310*(6988), 1213. doi:10.1136/bmj.310.6988.1213.

St-Pierre, M.-C., Ska, B., & Béland, R. (2005). Lack of coherence in the narrative discourse of patients with dementia of the Alzheimer's type. *Journal of Multilingual Communication Disorders, 3*(3), 211–215. doi:10.1080/14769670500065984.

Stanley, H. E., Buldyrev, S. V., Goldberger, A. L., Havlin, S., Ossadnik, S. M., Peng, C. K., & Simons, M. (1993). Fractal landscapes in biological systems. *Fractals, 01*(03), 283–301. doi:10.1142/S0218348X93000319.

Stanmore, E., Stubbs, B., Vancampfort, D., Bruin, E. D. De, & Firth, J. (2017). Neuroscience and Biobehavioral Reviews The e ff ect of active video games on cognitive functioning in clinical and non-clinical populations: A meta-analysis of randomized controlled trials. *Neuroscience and Biobehavioral Reviews, 78*(April), 34–43. doi:10.1016/j.neubiorev.2017.04.011.

Szatloczki, G. (2015). Speaking in Alzheimer ' s disease, is that an early sign ? importance of changes in language abilities in Alzheimer ' s disease. *Frontiers in Aging Neuroscience, 7*(October), 1–7. doi:10.3389/fnagi.2015.00195.

Tanaka, H., Adachi, H., Ukita, N., Kudo, T., & Nakamura, S. (2016a). Automatic detection of very early stage of dementia through multimodal interaction with computer avatars. In *Proceedings of the 18th ACM international conference on multimodal interaction - ICMI 2016* (pp. 261–265). ACM Press. doi:10.1145/2993148.2993193.

Tanaka, H., Adachi, H., Ukita, N., Kudo, T., & Nakamura, S. (2016b). Automatic detection of very early stage of dementia through spoken dialog with computer avatars. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction* (pp. 261–265). doi:10.1145/2993148.2993193.

Thomas, C., Keselj, V., Cercone, N., Rockwood, K., & Asp, E. (2005). Automatic detection and rating of dementia of Alzheimer type through lexical analysis of spontaneous speech. In *IEEE international conference mechatronics and automation, 2005: 3* (pp. 1569–1574). IEEE. doi:10.1109/ICMA.2005.1626789.

Toledo, C. M., Aluísio, S. M., Santos, L. B., Brucki, S. M. D., Trés, E. S., de Oliveira, M. O., & Mansur, L. L. (2018). Analysis of macrolinguistic aspects of narratives from individuals with Alzheimer's disease, mild cognitive impairment, and no cognitive impairment. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 10*, 31–40. doi:10.1016/j.dadm.2017.08.005.

Toronto emotional speech set (TESS) | TSpace Repository. (2018). Retrieved May 26, 2018, from https://tspace.library.utoronto.ca/handle/1807/24487.

Tóth, L., Gosztolya, G., et al. (2015). Automatic detection of mild cognitive impairment from spontaneous speech using ASR. In *Sixteenth annual conference of the international speech communication association. INTERSPEECH-2015* (pp. 2694–2698). Retrieved from. https://www.isca-speech.org/archive/interspeech_2015/i15_2694.html.

Toth, L., Hoffmann, I., Gosztolya, G., Vincze, V., Szatloczki, G., & Banreti, Z. (2018). A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech. *Current Alzheimer Research, 15*(2), 130–138. doi:10.2174/1567205014666171121114930.

Tröger, J., Linz, N., Alexandersson, J., König, A., & Robert, P. H. (2017). Automated speech-based screening for alzheimer's disease in a care service scenario. In *Proceedings of the 11th EAI international conference on pervasive computing technologies for healthcare* (pp. 292–297). ACM. doi:10.1145/3154862.3154915.

Tsanas, A. (2011). Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. *Journal of the Royal Society Interface, 8*(59), 842–855. doi:10.1098/rsif.2010.0456.

Valle, A. V, & Cortés, A. D. L. (2015). Correlatos neuroanatómicos y el déficit lingüístico en la enfermedad de alzheimer: diagnóstico temprano. *Revista Grafías, 28*, 21–38.

Vaziri, G., Almasganj, F., & Behroozmand, R. (2010). Pathological assessment of patients' speech signals using nonlinear dynamical analysis. *Computers in Biology and Medicine, 40*(1), 54–63. doi:10.1016/j.compbiomed.2009.10.011.

Veritone. (2019). *Voice analysis to detect Alzheimer's disease within seconds* Retrieved May 15, 2019, from https://www.veritone.com/blog/voice-analysis-detects-alzheimers-disease/.

Vincze, V., Gosztolya, G., Tóth, L., Hoffmann, I., Szatlóczki, G., Bánréti, Z., & Kálmán, J. (2016). Detecting mild cognitive impairment by exploiting linguistic information from transcripts. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (pp. 181–187). Retrieved from. http://www.aclweb.org/anthology/P16-2030.

Wankerl, S., Nöth, E., & Evert, S. (2017). An N-gram based approach to the automatic diagnosis of Alzheimer's disease from spoken language. In *Proceedings of the annual conference of the international speech communication association, INTERSPEECH* Retrieved from https://pdfs.semanticscholar.org/99bd/8d495cf48c9913940ceec1c2008c19a7ac2e.pdf.

Warnita, T., Inoue, N., & Shinoda, K. (2018, March). *Detecting Alzheimer's Disease Using Gated Convolutional Neural Network from Audio Data* Retrieved May 1, 2018, from http://arxiv.org/abs/1803.11344.

Weiner, J., Frankenberg, C., & Telaar, D. (2016). Towards automatic transcription of ILSE–An interdisciplinary longitudinal study of adult development and aging. *LREC*. Retrieved from https://pdfs.semanticscholar.org/cd70/03f0674f5153e1cb539d7b28a9ca5048091f.pdf.

Weiner, J., Herff, C., & Schultz, T. (2016). Speech-based detection of Alzheimer's disease in conversational German. In *INTERSPEECH* (pp. 1938–1942). Retrieved from. https://pdfs.semanticscholar.org/2d9d/fb9211bc9b955707b5e0d4e5258dd1fb8698.pdf.

Weiner, J., & Schultz, T. (2016). Detection of intra-personal development of cognitive impairment from conversational speech. In *Speech Communication; 12th ITG Symposium; Proceedings of VDE* (pp. 1–5). Retrieved from. http://ieeexplore.ieee.org/abstract/document/7776183/.

Weintraub, S., Wicklund, A. H., & Salmon, P. D. (2012). The neuropsychological profile of Alzheimer disease. *Cold Spring Harbor Perspectives in Medicine, 2*(4). doi:10.1101/cshperspect.a006171.

Wesnes, K. (2002). Assessing cognitive function in clinical trials: latest developments and future directions. *Drug Discovery Today, 7*(1), 29–35. doi:10.1016/S1359-6446(01)02068-2.

Winblad, B., Amouyel, P., Andrieu, S., & Ballard, C. (2016). Defeating Alzheimer's disease and other dementias: a priority for European science and society. *Lancet Neurology, 15*(5), 455–532. doi:10.1016/S1474-4422(16)00062-4.

Winterlight Labs. (2018). Retrieved May 15, 2019, from https://winterlightlabs.com/.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data mining: practical machine learning tools and techniques*. Retrieved from https://books.google.es/books?hl=es&lr=&id=1SylCgAAQBAJ&oi=fnd&pg=PP1&dq=Data+mining:+Practical+machine+learning+tools+and+techniques&ots=8IBPufjxxd&sig=DzRoPRLpYX7aIkRiXi7Oo0jr-vM#v=onepage&q=Data mining%3A Practical machine learning tools and techniques.

Yancheva, M. (2016). *Automatic assessment of information content in speech for detection of dementia of the Alzheimer type*. University of Toronto.

Yentes, J. M., Hunt, N., Schmid, K. K., Kaipust, J. P., McGrath, D., & Stergiou, N. (2013). The appropriate use of approximate entropy and sample entropy with short data sets. *Annals of Biomedical Engineering, 41*(2), 349–365. doi:10.1007/s10439-012-0668-3.

You, M., Chen, C., Bu, J., Liu, J., & Tao, J. (2006). Emotion recognition from noisy speech. In *Multimedia and Expo, 2006 IEEE international conference* (pp. 1653–1656). doi:10.1109/ICME.2006.262865.

Zhou, L., Fraser, K. C., & Rudzicz, F. (2016). Speech recognition in Alzheimer's disease and in its assessment. In *INTERSPEECH 2016* (pp. 1948–1952). doi:10.21437/Interspeech.2016-1228.