

Article

Using a Human Interviewer or an Automatic Interviewer in the Evaluation of Patients with AD from Speech

Jesús B. Alonso-Hernández *, María Luisa Barragán-Pulido, José Manuel Gil-Bordón, Miguel Ángel Ferrer-Ballester  and Carlos M. Travieso-González 

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; mbarragabl@idetic.eu (M.L.B.-P.); jose.gilbordon@idetic.eu (J.M.G.-B.); miguelangel.ferrer@ulpgc.es (M.Á.F.-B.); carlos.travieso@ulpgc.es (C.M.T.-G.)

* Correspondence: jesus.alonso@ulpgc.es; Tel.: +34-928-452863

Abstract: Currently, there are more and more frequent studies focused on the evaluation of Alzheimer's disease (AD) from the automatic analysis of the speech of patients, in order to detect the presence of the disease in an individual or for the evolutionary control of the disease. However, studies focused on analyzing the effect of the methodology used to generate the spontaneous speech of the speaker who undergoes this type of analysis are rare. The objective of this work is to study two different strategies to facilitate the generation of the spontaneous speech of a speaker for further analysis: the use of a human interviewer that promotes the generation of speech through an interview and the use of an automatic system (an automatic interviewer) that invites the speaker to describe certain visual stimuli. In this study, a database called Cross-Sectional Alzheimer Prognosis R2019 has been created, consisting of speech samples from speakers recorded using both methodologies. The speech recordings have been studied through a feature extraction based on five basic temporal measurements. This study demonstrates the discriminatory capacity between the speakers with AD and the control subjects independent of the strategy used in the generation of spontaneous speech. These results are promising and can serve as a basis for knowing the effectiveness and extension of automated interview processes, especially in telemedicine and telecare scenarios.

Keywords: Alzheimer's disease (AD); automatic voice recognition; automatic interviewer; telecare; telemedicine



Citation: Alonso-Hernández, J.B.; Barragán-Pulido, M.L.; Gil-Bordón, J.M.; Ferrer-Ballester, M.Á.; Travieso-González, C.M. Using a Human Interviewer or an Automatic Interviewer in the Evaluation of Patients with AD from Speech. *Appl. Sci.* **2021**, *11*, 3228. <https://doi.org/10.3390/app11073228>

Academic Editor: Michael Döllinger

Received: 3 March 2021

Accepted: 30 March 2021

Published: 3 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Alzheimer's disease (AD) is, currently, the most common cause of neurodegenerative dementia in the world [1,2]. It implies 70–76% of dementia cases in developed countries, increasingly long-lived [1] where numbers threaten to triple by 2050. Memory loss appears as one of the first symptoms, to which others such as difficulties with language use or spatial and temporal disorientation are added. In more advanced stages, the ability to carry out daily activities or even basic body functions, such as walking or swallowing [3], decreases or disappears. Either way, when the first symptoms are revealed, the damage caused is already irreparable and chronic.

No cure has been found up until now, nor is it possible to have a reliable diagnosis in life. The diagnostic process remains complex, is inevitably carried out in the advanced stages of the disease and is prolonged over time. Its use as a screening method is limited since current methods are expensive and invasive [4]. For this reason, there is great interest in finding biomarkers in more accessible parts of the body that are sensitive to AD before the clinical onset of dementia, and which also allow the specific stages of the disease to be monitored afterwards [5]. In such circumstances, developing eHealth 4.0 solutions based on more accessible particular biomarkers would allow democratizing the evolutionary and pharmacological control in an easy, fast, non-invasive and scalable way. It would provide

objective parameters and facilitate medical specialist's work. None of these techniques would require an extensive infrastructure or availability of medical equipment and could be used, if necessary, even remotely as a Telecare solution.

In this context, many studies have shown that speech analysis is a powerful indicator of a patient's cognitive status and that even the first symptoms can be found years before a probable clinical diagnosis is stated [6,7]. Furthermore, it has been shown that some specific communication problems, such as aphasia or anomie, depend on the stage of the disease [4] and increase with the course of AD [8–10]. Another determinant is the emotional responsiveness of the patients, which is also affected by the course of the disease causing social and behavioral changes [11] that affect the ability to communicate.

In the last few years, techniques based on automatic processing of the voice signal from its recording have found an important niche in language evaluation applied to identify neurodegenerative diseases [12]. These techniques offer the possibility of quantifying signal properties that are relevant to the description of a given pathology. Subsequently, with the support of Automatic Learning techniques, the process of classifying the samples is performed according to the results obtained. In this framework, the concept of Deep Learning [13] appears as a more complex method of automatic learning which is also making its way into the field [14]. The use of automatic grading methods that use linguistic characteristics resulting from verbal emissions also has the advantage of being able to apply automatically, avoiding the possible influence of an intermediary or interviewer.

In recent years, an almost exponential increase has been documented in the number of investigations aimed at shedding light on speech analysis as a non-invasive biomarker of AD [15]. Since the first lines appeared, 78% have been based on the use of conventional parameters; mainly duration of deaf or sound segments, pitch, amplitude and periodicity, as well as others obtained from analysis of the frequency and cepstral domains [16–18]. Also, concepts such as speech quality or Emotional Temperature (ET) have been defined. Other techniques such as Automatic Spontaneous Speech Analysis (ASSA) [19] are presented as methods that encompass different combined voice qualities (durations, short time energy and spectral centroid) and provide relevant information by classifying these data, in most cases using Support Vector Machine, k-Nearest Neighbors, Linear Discriminant Analysis [20] or Multi-Layer Perceptron [21] classifiers. Currently, studies also include classical characteristics in emotional analysis such as pitch, intensity and variation of frequency components and, more recently, ET [22]. They present methods like ASSA and Emotional Speech Analysis (ESA) that, combined with ET and Spontaneous Speech (SS) tasks, have achieved discrimination between patients with AD and healthy controls with 94% accuracy using an SVM classifier [23]. Numerous studies have also been published on the basis of Voice Activity Detector (VAD) transcriptions [24] which, in addition to an acoustic analysis of the voice or speech signal, offer lexical, semantic, punctuation or syntactic analysis of the communicative process [25,26].

Since approximately 2012, more and more works in this line point to the need to seek support in the non-linear and non-stationary aspects [15]. Some researchers have proposed that subtle cognitive changes in the early and preclinical stages could be more accurately detected by fractals. In some works they are presented in combination with other non-conventional measures such as the Hurst exponent [27] or just in combination with linear characteristics [28].

It is also important to outline that, so far, one of the main limitations has been the scarcity and, in turn, diversity of samples available to train models for the evolutionary control of AD.

Although there are databases of the most varied disciplines for the study of AD, ranging from meta-analyses focused on the genetic association with putative Alzheimer's disease [29] to others based on the olfactory system for the precocious in Parkinson's and Alzheimer's [30], the reality is that the most of the located databases regarding the speech recognition in AD lack the required amount of data to carry out a truly consistent analysis and there is also the disadvantage of having been established according to

different guidelines and criteria. Most localized databases focus on spontaneous speech recordings [31–33], although there are others that invite the subject to perform tasks as repetition or reading [20]. The recording of the samples is usually carried out in a timely manner on a number of control subjects and AD patients classified according to different degrees of severity: mild, moderate and severe [34–37]. Although less, some have carried out longitudinal studies on the subjects [38,39] in order to find a relationship between the progression of AD and language impairment.

As a review of the state of the art, localized databases that include recordings for linguistic analysis of AD patients are listed in Table 1. Information is also included on the type of interviewer used in them; human/automatic, and on the type of study over time: longitudinal/transverse. This table includes other information such as the language used by the subjects in the recordings, the classification of the participants (Healthy Control (HC), Mild Cognitive Impairment (MCI), Alzheimer’s Disease (AD)) and the linguistic tasks performed. It is important to clarify how we have classified the tasks used for the recordings of each of the studies. In this way, we define: spontaneous speech (SS), which includes telling stories, experiences, answering questions freely or describing images, inter alia. Reading, repetition, counting or animal denomination tasks (cVF) are others that appear in some studies, although less, and we understand that they are not related to the concept of spontaneous speech, as we define it. We have used the term Mixed when both SS and other tasks have been included in the study.

Table 1. Databases of recordings used for linguistic analysis of localized Alzheimer’s disease (AD) and types of study according to the distribution in time of taking measurements and the type of interviewer used.

DataBase	Interviewer	Long/Transv	Language	Task	HC	MCI	AD	References
					M/F	M/F	M/F	
SAIOTEK	Human	Cross sectional	Multilingual	SS	5	-	3	[19,40]
AZTIAHO				SS	50	-	20	[23,28,41–44]
PGA-OREKA				cVF	26/36	-	17/21	[45]
MINI-PGA				SS	12	-	6	[46]
-	Human	Cross sectional	Greek	SS	30	-	30	[18]
Dem@care	Human	Cross sectional	Greek	Mixed	4/15	12/31	3/24	[47]
		Cross sectional	French	Mixed	6/9	11/12	13/13	[48–50]
-	Human	Cross sectional	French	Reading	14	14	14	[51]
-	Human	Cross sectional	French	SS	5	-	5	[52]
*TRANSC	Human	Longitudinal	-	SS	184 *	-	214 *	[53]
ClinicalTrials.gov	Human	Longitudinal	English	SS	27	14	-	[54]
-	Human	Cross sectional	English	SS	46	-	26	[55]
WRAP	Human	Longitudinal	English	SS	200	-	64	[56]
Pitt (DB)	Human	Longitudinal	English	SS	74	19	169	
Kempler (DB)	Human	Cross sectional	English	SS	-	-	6	
Lu (DB)	Human	Cross sectional	Mandarin	SS	-	-	52	[31–33,39,57–61]
Lu (DB)	Human	Cross sectional	Taiwanese	SS	-	-	16	
PerLA (DB)	Human	Cross sectional	Spanish	SS	-	-	21	
ACADIE	Human	Longitudinal	English	SS	-	-	95	[62]
AMI	Human	Cross sectional	English	SS	20	-	20	[61]
CCC	Human	Cross sectional	English	SS	10	-	55	[60,63]
ILSE	Human	Longitudinal	German	SS	80	13	5	[64]
CREA-IMSERSO	Human	Cross sectional	Spanish	Reading	-	-	21	[65]
-	Human	Cross sectional	Spanish	Reading	82	-	45	[66]
-	Human	Cross sectional	Spanish	Mixed	29	-	34	[25,67]
Cinderella	Human	Cross sectional	Portuguese	SS	20	20	20	[68]
OPLON	Human	Cross sectional	Italian	Mixed	48	48		[69]

Table 1. Cont.

DataBase	Interviewer	Long/Transv	Language	Task	HC	MCI	AD	References
					M/F	M/F	M/F	
-	Human	Cross sectional	Iranian	SS	15/15	-	16/14	[70]
-	Automated	Cross sectional	Japanese	SS	7/3	-	9/1	[17]
-	Human	Cross sectional	Japanese	SS		73/200		[71]
BELBI	Human	Cross sectional	Serbian	SS	-	-	12	[72]
BEA	Human	Cross sectional	Hungarian	SS	13/23	16/32	-	[26,73–75]
-	Human	Cross sectional	Turkish	SS	31/20	-	18/10	[16]
-	Human	Cross sectional	French	SS	29	-	29	[76]
-	Human	Cross sectional	Persian	SS	0/6	-	0/6	[77,78]
Others studies using unknown databases								
[79–83]								

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. * Transcriptions, no participants.

Automatic Interviewers

A common denominator in the collection of voice samples in the AD study is to resort to a person who leads the interview, usually by encouraging the subject to recall past memories or answer questions [32,33], read texts [84] or describe a photo [85], although some have already included computer avatars in their work [71,74] which would make the results more objective and less dependent on human factors such as the ability of the interviewer to direct the interview. In any case, a common and objective method of recording speech for AD patients seems necessary.

Several research efforts have identified relevant advantages in computer-assisted communication compared to human interaction. Among these advantages, we identify that the participant feels more anonymity while being interviewed. These assisted communication systems are being equipped increasingly, among others, with communication skills that facilitate communication by using vision and prosodic analysis to implement active listening behaviours, smiles, head movements and postural mimicry. They use non-verbal strategies and show empathy so that the interviewed subject can feel comfortable and the conversation can go more smoothly. This fact helps to generate feelings of sympathy and trust in the interviewee [86]. In any case, it is necessary for the automatic interviewer to be able to perceive the human behaviour in order to process it, understand it and react to it. It is also essential that these skills do not work in isolation and that they allow integration into other systems, so that they learn from human interaction as they are used.

In general, there are systems that already work to reduce the barriers of interviewers and virtual systems that interact with humans. They are implementing automatic speech recognition systems [87], facial expression recognition [88] and natural language generation [89]. All the standardization efforts have tried to consolidate all the systems. For example, SAIBA [90] is a framework to generate empathic conversations. On the other hand, there are LiteBody and DTAsk that aim at creating social-emotional relationships with users through verbal and non-verbal social behaviours [91]. GRETA [92] is a conversational agent that complies with SAIBA, and focuses less on the interaction of natural language and more on the generation and performance of non-verbal affective behaviours. It applies a complex facial generation model, in particular. The project SEMAINE [93] aims to integrate various research technologies, including some of the above, into the creation of a virtual listener. The focus would be on perception and reverse channeling rather than deep representations of dialogue. There are even tools that allow the creation of virtual humans that interact with people in a conversation [94]. An example of an automatic interviewer that is based on a virtual human and employs multiple resources to make communication as human as possible is SimSensei Kiosk [86].

Either way, few works have focused on applying these strategies to AD. Although there is evidence of some solutions such as SimpleC built by IBM [95] that replace purely human variables (such as the interviewer's empathy with the subject), these remain rare.

It has not yet been demonstrated to what extent, for the same subject, the parameters obtained from a human or automatic interviewer can vary or how this fact could fit into the current reality.

Within the framework of the detection and evolutionary control of AD based on voice recordings and their automatic processing, this work intends to objectively determine the discriminatory capacity (healthy and pathological AD voices) of two different recording strategies, specifically by obtaining the samples through human and automatic interviewers.

2. Materials and Methods

This section describes the materials and methodology we have used to determine the discriminatory capacity of the voice of AD patients and HC subjects under the two proposed scenarios: the human interviewer and the automatic interviewer.

A database has been created to accomplish the study which we have called Cross-Sectional Alzheimer Prognosis R2019, and which has been used to compare both types of samples. In Section 2.2.1 the way the recordings were made is properly explained. Subsequently, voice processing and feature extraction based on speech times have been performed. Using a series of descriptive statistical measures and then applying a statistical analysis based on the Wilcoxon Summation Test [96], two parallel studies were suggested to determine the discriminatory capacity of each interviewer.

2.1. Methods

To find out to what extent the type of interviewer influences the difference of healthy and pathological AD voices, the Cross-Sectional Alzheimer Prognosis R2019 database has been created. In it, on each of the subjects recorded, both AD and HC patients, two types of interviews were done and, therefore, two different types of recordings. The first type of recording is characterized by having been carried out by an automatic interviewer (Prognosis software, see Section 2.2.3). The second type of recording, as is widely done in the field, has been realized by a member of the research team. Hereafter, we will define the first type of recordings as induced speech samples, and those obtained through the human interviewer will be defined as spontaneous speech samples. It is interesting to clarify that all recordings have always been made in the environment of the patients and the HC subjects, so that the result can be as realistic as possible.

After collecting recordings of both types for all subjects, a Voice Activity Detector (VAD) was applied to the samples using Matlab[®] software (Figure 1). As a result of this process, different speech and silence sequences are obtained and stored in a matrix, where the initial and final time of each sound fragment S_i is indicated.

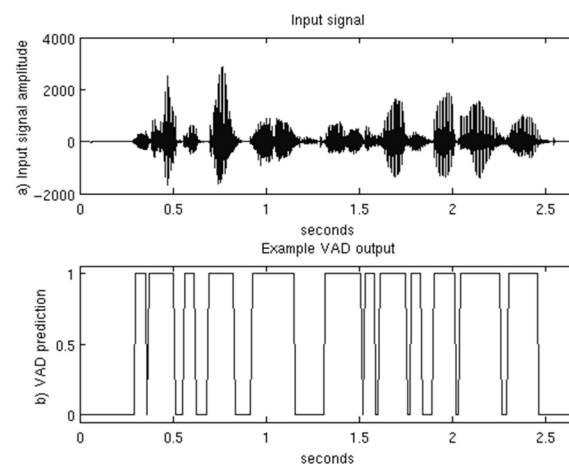


Figure 1. Voice Activity Detector (VAD) applied on voice signal.

The voice signal is characterized using different descriptive statistics estimators of the duration of the different sound fragments in which it is divided the speech of a speaker. The different features are described below.

- Average speech time \bar{t}_S : describes the average speech time of the different sound fragments which can be identifier in a sample of speech of a speaker. It is estimated [97] from the arithmetic mean of the duration of all the sound fragments in a single recording.

$$\bar{t}_S = \frac{\sum_{i=1}^N t_{S_i}}{N} \quad (1)$$

In which t_{S_i} is the time duration of each sound fragments (S_1, S_2, \dots, S_N) is which is divided each speech recording $\{S_i\}$.

- Variance of speech time ($\sigma_{t_S}^2$): describes the variation of the different sound fragments in a recording. It is estimated [98] using the following estimator of the variance:

$$\sigma_{t_S}^2 = \frac{\sum_{i=1}^N (t_{S_i} - \bar{t}_S)^2}{N - 1} \quad (2)$$

- Skewness of speech time $\tilde{\mu}_{t_{S3}}$: This measure allows characterizing the behavior of the probability distribution function of the duration of the different sound fragments. This measure quantifies [99] the lack of symmetry of the average duration of the voice fragments. In this way, when the studied samples follow a normal distribution, the value of $\tilde{\mu}_{t_{S3}}$ will be zero. Positive or negative values of $\tilde{\mu}_{t_{S3}}$ indicate data skewed to the right of their distribution curve or to the left, respectively. Skewness of speech time is calculated using the following estimator:

$$\tilde{\mu}_{t_{S3}} = \frac{\sum_{i=1}^N (t_{S_i} - \bar{t}_S)^3}{N \cdot (\sqrt{\sigma_{t_S}^2})^3} \quad (3)$$

where t_{S_i} is the time duration of each sound fragments, \bar{t}_S is the average speech time, $\sigma_{t_S}^2$ is the variance of speech time, and N is the number of sound fragments in the sample of speech.

- Kurtosis of speech time ($Kurt_{t_S}$): This is other measure which allows characterizing other aspects of the behavior of the probability distribution function of the duration of the different sound fragments. This measure states the quantity of sound fragments in a recording with a duration that is close to the average duration (\bar{h}). The larger $Kurt_{t_S}$ is, the steeper its distribution curve will be. $Kurt_{t_S}$ is calculated [99] using the following estimator:

$$Kurt_{t_S} = \frac{\sum_{i=1}^N (t_{S_i} - \bar{t}_S)^4}{N \cdot (\sqrt{\sigma_{t_S}^2})^4} - 3 \quad (4)$$

- Index of speech time (Ind_{t_S}): describes the relationship between the total time the subject is speaking to the total duration of the recording. It is calculated as the division between the total time of speech sequences by the total recording time of the sample.

$$Ind_{t_S} = \frac{\sum_{i=1}^N t_{S_i}}{T_{TOTAL}} \quad (5)$$

in which t_{S_i} is the time duration of each sound fragments (S_1, S_2, \dots, S_N) is which is divided a speech recording and T_{TOTAL} is the duration of the speech recording.

In order to know exactly which of the above variables are AD-discriminatory and which are not for each interviewer, a descriptive statistical analysis and a non-parametric study based on the Wilcoxon Summation Test [100], has been done, since the samples do not show a normal distribution [101]. The Stata[®] software was used for the statistical study.

2.2. Materials

2.2.1. Databases

The Cross-Sectional Alzheimer Prognosis R2019 database is a database created to test how discriminating a voice sample of AD can be based on the type of interviewer used in the recording. Based on this, for each subject, AD or HC patient, this database has both spontaneous and induced speech samples, all collected in a single session.

It should be noted that for the production of speech in the subjects we have not looked for a pre-designed sentence structure, or balanced sentences, but we have recorded spontaneously generated speech. In this sense, there is no design of sentences, but recollection is used through stimuli in the participants: in the case of the human interviewer, by inviting the subject to speak, and in the case of the automatic interviewer, through videos and images, with the aim of producing reminiscences in the subjects that evoke memories.

The tools used to proceed with the recording have been a laptop, headset with microphone [102] and, in the case of automatic interviews, the Prognosis software (see Section 2.2.3). Specifically, four different recordings are obtained for each subject, three of induced speech and one of spontaneous speech. They were recorded at a sampling rate of 44,100 Hz. The recording file type used was WAV.

The total number of people recorded was 87, of whom 41 were AD patients and 46 HC subjects, all over 65 years of age. The total number of women recorded is 56 compared to 31 men. As for the distribution according to the degree of the disease we have that of the 41 patients with AD, 15 correspond to the moderate degree and 26 to the mild degree. Table 2 as well as Figures 2–4 represent the distribution of the people recorded according to sex, presence or absence of AD and the comparison between the different degrees analyzed (HC subjects, mild or moderate patients).

Table 2. Distribution of the people recorded according to sex, presence or absence of AD and the comparison between the different degrees analyzed (HC subjects, mild or moderate patients).

	Polulation	Women	Men
Healthy control	46	23	23
AD patients	Mild	19	7
	Moderate	3	12

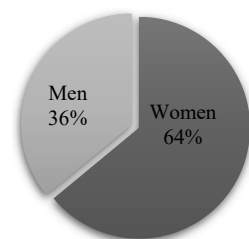


Figure 2. Distribution of the population described in the database by sex.

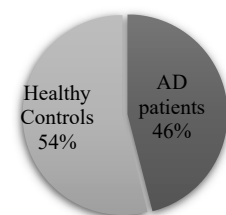


Figure 3. Distribution of the population by clinical condition (healthy/pathological).

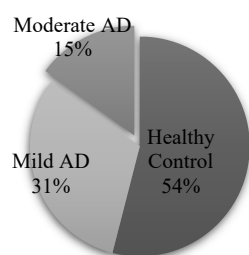


Figure 4. Distribution of the population by degree of AD.

2.2.2. Human Interviewer

The recordings of the samples were carried out by one member of the investigation team using an Intel Core i5, 12 GB RAM, 250 Gb SSD laptop, where different recordings have been stored. Furthermore, Audacity[®] software has been used to carry out the recording and the Ozone Rage ST [102] headset with microphone. When conducting the interview, each of the subjects has been asked or encouraged to talk about what they want, about a free topic, with the objective of achieving a spontaneous speech recording that varies between 30 s and 2 min.

2.2.3. Automatic Interviewer

The obtained samples of the automatic interviewer are gained by the Prognosis software. The main objective of this software is to make recordings based on the “automatic interviewer”. In order to obtain these samples of induced speech and to incite the subjects to speak, a series of videos have been used as stimulus, taking into account that the common denominator of all the participants is age and that they are all over 65 years old. In the Portal Memoria Digital de Canarias of the University Library of the ULPGC [103] there is a wide repertoire of reports and old news, among others, that we have used to provoke reminiscences in the subjects. Currently, there are numerous studies that demonstrate the effectiveness of the use of multimedia tools, such as videos, to make subjects who suffer from dementia or AD communicate or express themselves more easily [104–107], and they are also especially interesting when they are personalized or provoke reminiscences of moments lived. Based on this, we have expressly selected from the repository those videos that reflect past times and that, therefore, we consider that could somehow awaken memories in the participants.

The program is divided into three phases (see Figure 5). The first phase corresponds to the registration of the participant’s data. The interviews are anonymous but some participant’s data are recollected: degree of AD (control, mild, moderate, or severe), sex and age.

The second stage corresponds to the recording of a sustained vowel (/aa/). The software invites the speaker, through a video, to produce a sustained vowel. The software shows a short motivational video where one of the investigators of the project explains how the recording is going to be and how the subjects should act. Finally, as indicated in this video, the subject must pronounce the vowel /aa/ in a sustainable way for 8 s, which will be recorded automatically. Recordings of sustained vowels have not been specifically used for this work, although they are reserved in the database for future studies.

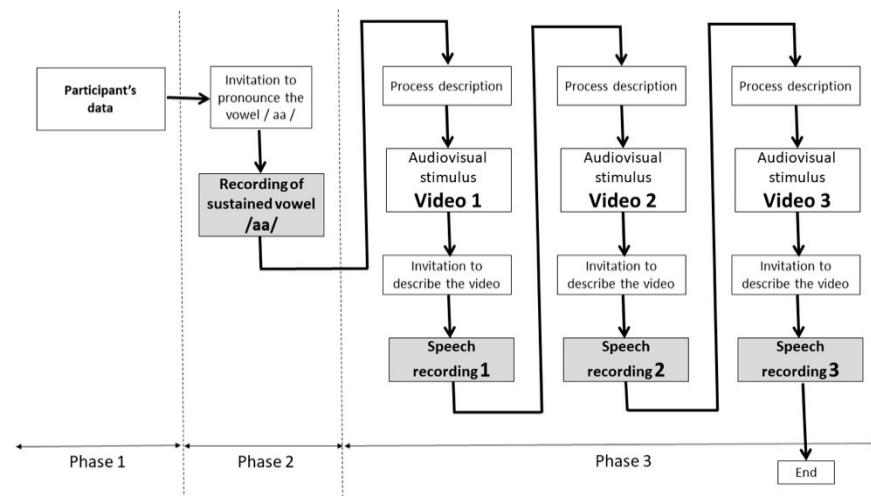


Figure 5. Flow chart describing the software PROGNOSIS.

In the third phase, three samples of induced speech are recorded for the speaker. For the recording of each sample, firstly the software reproduces a video inviting the speaker to describe the video that it is going to reproduce (see Figure 6). Next, the software reproduces a video, which is expected to generate fond memories of childhood or youth in the speaker (see Figure 7), and later the software will play another video inviting the speaker to describe the video. The description made by the speaker is automatically recorded. This recording has a duration of 30 s. It is expected to be able to generate reminiscences in the speaker which generate interest in the speaker by sharing their experiences and memories. Three videos have a duration of approximately between 30 s and 2 min, and they are randomly selected from a video repository provided by the software.



Figure 6. Frame of the video explaining the process. Instructions.

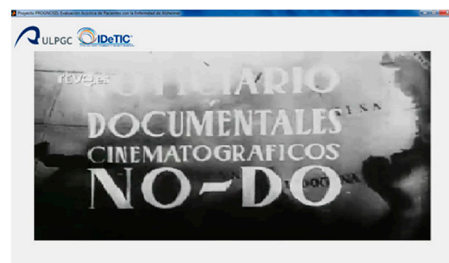


Figure 7. Frame of one of the videos. Newcastle NO-DO.

Once each recorded period is finished, the software is also designed to detect if the collected sample collected contains audio from the participant or if, by contrast, only silence has been recorded. In case the participant has not spoken during the recording, a short video will be shown again where the same team researcher indicates this situation and encourages the subject to speak. Since in that case no sample would have been obtained,

this deficiency is compensated by playing a new video. Once the recordings are finished, the researcher reappears on screen and thanks the subject for participating in the study.

3. Results

The descriptive statistics of each measure have been studied for each of the populations (HC and AD), differentiating in the methodology for the interview (human and automatic interviewer). The results are show in Table 3.

Table 3. Descriptive statistical results: mean values (μ) and standard deviation (σ).

Variable	HC		AD	
	Human Interviewer μ (σ)	Automatic Interviewer μ (σ)	Human Interviewer μ (σ)	Automatic Interviewer μ (σ)
\bar{t}_S	1.79 (0.53)	1.8 (0.76)	1.42 (0.39)	1.36 (0.45)
$\sigma_{t_S}^2$	1.6 (1.32)	1.33 (1.40)	0.93 (0.47)	0.9 (1.01)
$\tilde{\mu}_{t_{S3}}$	0.81 (0.56)	0.49 (0.59)	0.9 (0.46)	0.47 (0.64)
$Kurt_{t_S}$	3.3 (1.52)	2.3 (1.1)	3.37 (1.41)	2.35 (1.17)
Ind_{t_S}	0.71 (0.1)	0.7 (0.12)	0.48 (0.14)	0.47 (0.18)

Base on the obtained data in characteristics extraction process, Table 4 shows boxplots of samples analyzed for each of the five variables under study.

Table 4. Boxplots for the five temporary variables.

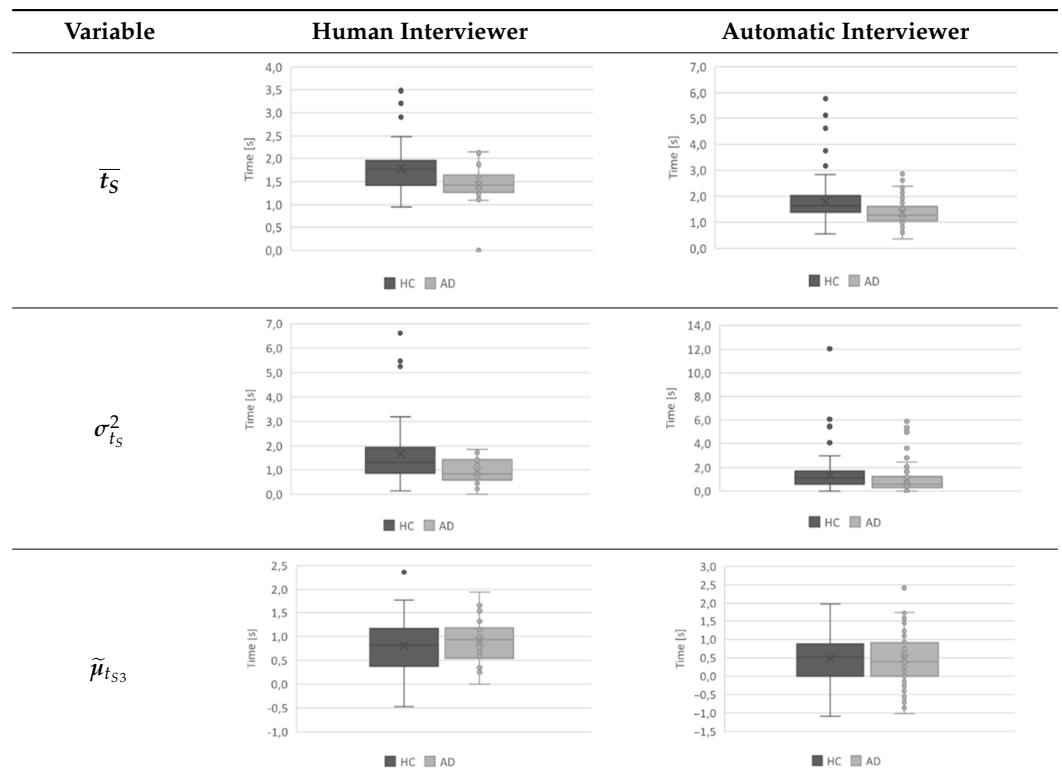
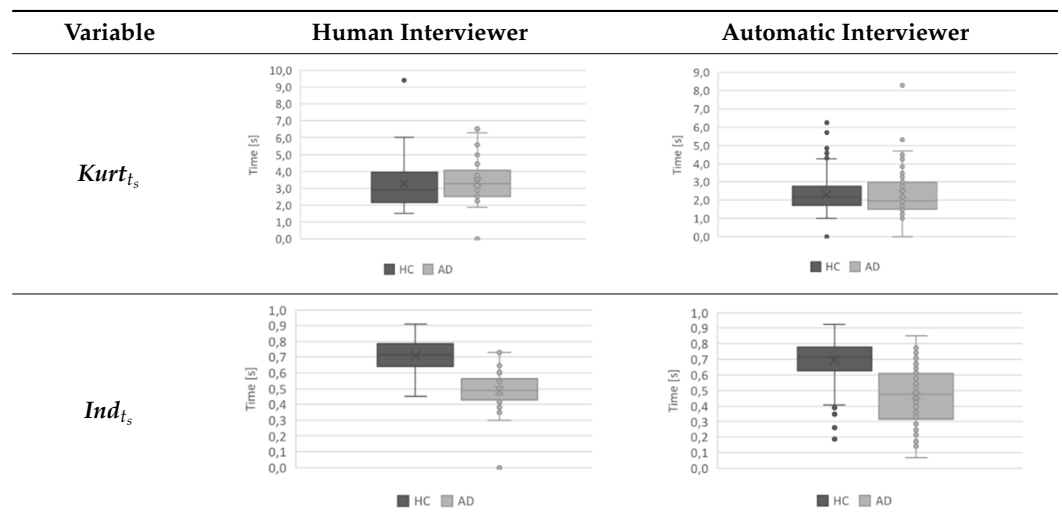


Table 4. Cont.



In previous graphs it is possible to see the discriminative capacity of some of the variables under study, such as, for example, the variable Ind_{t_s} . Others, however, are less discriminatory.

Subsequently, in order to know with each interviewer which variables are capable of discriminating AD from HC, a non-parametric analysis based on the Wilcoxon Summation Test has been made. The results are shown in Table 5.

Table 5. Results of Wilcoxon Summation Test.

Variable	Human Interviewer		Automatic Interviewer	
	z	Prob z	z	Prob z
\bar{t}_S	3.05	0.002	5.77	0
$\sigma_{t_S}^2$	2.79	0.005	3.94	0
$\tilde{\mu}_{t_{S3}}$	-0.57	0.565	0.47	0.637
$Kurt_{t_s}$	-0.80	0.418	0.65	0.515
Ind_{t_s}	5.87	0	9.64	0

In this case, the probability of z must be less than 0.05%, or the error below 95%, so that the result of any of the variables studied can be accepted as discriminatory. In other words, this implies that a variable will be considered discriminatory when the null hypothesis that there is no difference between the study variables is rejected.

As can be seen in Table 5, the variables \bar{t}_S , $\sigma_{t_S}^2$ and Ind_{t_s} confirm that there is a difference between HC and AD subjects, since the probability of z is less than 0.05%, for both methodologies.

4. Discussion

Databases are a fundamental aspect of any research since only on the basis of them we can develop our experimental studies and analyses. In this work, we have located and classified a series of databases related to language use in AD patients in order to understand better their role and current situation in the field. From a thorough review of the state of the art, we have found, in addition to a shortage in the number of localized databases, a great diversity in how the recordings of the subjects are made. We have, for example, aspects related to the recording process, to the automation of the interviews or to the linguistic tasks performed by the subjects that differ much from one study to another. Although aspects like those mentioned above influence the recordings, they are not the only ones; there are many other variables as the language, the environment or

simply the pre-processing methods used which make each database obtainable under different conditions.

However, it should be mentioned that there is a certain inclination as regards the type of linguistic task performed, and up to 80% of the cases analyzed use spontaneous speech in their recordings, understanding spontaneous speech as those tasks in which questions are asked to the subject and give a limited, relatively long time to express themselves freely. Nevertheless, we do not have clear evidence as to whether other linguistic tasks, such as reading, could provide us with more or less information or even be complementary to each other. On the other hand, we have been able to observe that only 18% of the located databases correspond to longitudinal studies. This type of study is of special interest because it allows the analysis of the time variable on the samples, which is undoubtedly a reflection of the progressive deterioration of language that these patients suffer. Among the databases located, it should be noted that only one of them automated the interview process with the subject by means of computerized avatars. However, we are not aware of how this fact can effectively affect the final results.

The Cross-Sectional Alzheimer Prognosis R2019 database has been created in order to find out how the automating process of interviews affects the taking of samples and the results of subsequent statistical analyses. This database consists of two types of recordings: spontaneous speech (human interviewer) and induced speech (automatic interviewer). In this sense, it is worth highlighting two main advantages of the database used. In the first place, it is a database in which two clear methodologies are collected to record the voice for the purposes indicated in this study. We have the same subject who participates in both types of recordings. The second advantage resides in the potential of these registers which, based on the pertinent analyzes carried out, allow us to discover to what extent these methodologies are validated, the automatic one in particular.

A descriptive statistical analysis and a non-parametric statistical study of the different variables (\bar{t}_S , $\sigma_{t_S}^2$ and Ind_{t_S}) has been realized, which allow characterizing the different sound fragments of a speech sample.

As a first step, by means of the descriptive statistical analysis, we have probed that for both kinds of interviewer, durations of the different sound fragments of a speech sample, and their variability, in HC subjects have been always higher than those related to AD patients. In a second round, the Wilcoxon Summation Test has been applied to these variables (\bar{t}_S , $\sigma_{t_S}^2$ and Ind_{t_S}) in order to verify whether the variables have discriminatory capacity to differentiate between HC and AD subjects, and the dependence on the methodology used in the interviewer (human interviewer and automatic interviewer). As a result, the probabilities (z) obtained for each of these variables show that the proposed measures have discriminatory capacity independent of the methodology applied in the interviewer.

By its part, it can be seen from Table 4 that the Ind_{t_S} variable has a high discriminatory potential for the case of both automatic interviewer and human interviewer. For the rest of variables it can be seen that in all cases values of samples overlap to a greater or lesser extent. However, this fact does not apply to the Ind_{t_S} variable where, apart from not overlapping values, the mean for each population is clearly different regardless of the interviewer used. For the human interviewer there are approximately 0.8 s for the control subjects and about 0.55 s for AD patients. For its part, using the automatic interviewer we have a mean close to 0.7 s for HC subjects and a value of 0.5 s for each AD population. Regarding the dispersion of the data, it can be said that is greater in the case of the automatic interviewer than in the case of the human interviewer.

In any case, both interviewers offer good results. In the case of the automatic interviewer, there are fewer studies that allow us to quantify its effectiveness compared to the human one. The good results obtained in this first approach serve as a baseline for us to continue working in this line, applying different tests, such as the parametric ones, or analyzing other variables that could be not only temporary.

5. Conclusions

In this work we have carried out a search and review of the state of the art on the different strategies that have been followed so far in the recording of subjects for the linguistic analysis of AD. Currently, although the repository of localized databases is not as extensive as would be desired, we have been able to note the great diversity of strategies followed in the creation of these databases and the lack of a common criteria. This common criterion would imply knowing what kind of tasks carried out by the subject are more interesting, who, how, where and when these interviews take place and the possible influence of these factors on the recording. Up until now, there has been no clear evidence of which is more beneficial for the purpose: the detection and/or evolutionary control of AD. For its part, we have been able to verify that there is a significant lack of longitudinal databases. When recording, the contribution of longitudinal databases is particularly interesting, which would allow us to capture the progressive deterioration that these patients suffer.

Another important aspect is the automation of the recording process as such. Few works have focused on applying automatic strategies, at least applied to AD. This includes, for example, the automation of neuropsychological tests or patient interviews that are currently still administered manually. In this case, although there is evidence that at least one of the databases we have located conducts guided interviews with computer avatars, little is known about the benefits of automating these processes. Therefore, it seems interesting to check and go deeper into the extent to which, for the same subject, the parameters obtained could vary depending on the type of interviewer.

In this paper, two concepts have been defined to understand the specific benefits of these automatic techniques compared to their manual counterparts: induced speech (obtained from the automatic interviewer) and spontaneous speech (obtained from a human interviewer). Based on them, a database called Cross-Sectional Alzheimer Prognosis R2019 has been created, in which samples of both types have been taken. From the processing of these samples and their subsequent statistical analysis, the results obtained prove to be discriminatory regardless of the recording strategy, whether a person has been employed as an interviewer or whether the recordings have been obtained automatically from the Prognosis software. Although the results obtained are promising, the scarcity of this type of study implies a clear need to continue working on evidence that is increasingly strong and allows for a higher level of interpretation.

From the promising results obtained, the door is also opened to carry out a thorough study in which the Cross-Sectional Alzheimer Prognosis R2019 database is expanded, complemented and improved. As future lines, we propose the possibility of combining the analysis of speech raised in this work by expanding the database, not only to new participants, but also to new data that can likely be correlated with AD, such as an educational or pharmacological factor, as well as other data of a geographical and social nature.

It also seems interesting to be able to extend this type of study to new techniques proposed in the field of voice recognition, in which ideas based on the study of samples obtained in real noisy environments such as social gatherings, streets, cafes and restaurants are raised [108]. Likewise, an interesting line to take into account in this regard is given by the current challenges posed to address a voice recognition scenario capable of providing speech enhancement, speaker diarization and speech recognition modules, for example, by means of recognition modules based on multispeaker speech recognition for unsegmented recordings [109].

Mainly this study aims to open the door to an evaluation methodology based on non-invasive, objective, collaborative and easily replicable techniques. Added to that, it is also important to mention the high acceptance by patients as an important advantage. With a view to the future, all these conditions together would help to make real online remote evaluations of sample acquisition, which is useful not for medical diagnosis but for evolutionary control, pharmacological control or early detection of AD, among others.

Although the automatic methods that have been applied to date are obviously not sophisticated enough, it seems clear that, in the future, greater knowledge in this area could contribute to the scalability of this type of language test, which has been applied manually until now and with the economic, social or health limitations that this involves. There is also a wide field of study in which not only could examinations be automated from an acoustic point of view but also taking into consideration other cognitive aspects such as semantics or lexicon. In any case, deepening the automation of these processes would give more possibilities to new techniques, currently on the rise, such as Telecare.

Author Contributions: Conceptualization, J.B.A.-H., J.M.G.-B. and M.L.B.-P.; methodology, J.B.A.-H.; software, J.B.A.-H.; validation, J.B.A.-H., J.M.G.-B. and M.L.B.-P.; formal analysis, J.M.G.-B. and M.L.B.-P.; investigation, J.M.G.-B. and M.L.B.-P.; resources, J.B.A.-H.; data curation, J.M.G.-B. and M.L.B.-P.; writing—original draft preparation, M.L.B.-P.; writing—review and editing, J.M.G.-B. and M.L.B.-P.; visualization, J.M.G.-B.; supervision, J.B.A.-H., M.Á.F.-B. and C.M.T.-G.; project administration, J.B.A.-H.; funding acquisition, J.B.A.-H. and M.Á.F.-B. All authors have read and agreed to the published version of the manuscript

Funding: This study was funded by the Spanish Ministry of Science and innovation PID2019-109099RB-C41 Research Project and European Union FEDER program/funds.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of University of Las Palmas de Gran Canaria (protocol code CEIH-2014-01 and date of approval July 17, 2014).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Molinuevo Guix, J.L. Role of biomarkers in the early diagnosis of Alzheimer's disease. *Revista Española Geriatria Gerontol.* **2011**, *46*, 39–41.
2. Weller, J.; Budson, A. Current understanding of Alzheimer's disease diagnosis and treatment. *F1000Research* **2018**, *7*. [[CrossRef](#)] [[PubMed](#)]
3. Andersen, C.K.; Wittrup-Jensen, K.U.; Lolk, A.; Andersen, K.; Kragh-Sørensen, P. Ability to perform activities of daily living is the main factor affecting quality of life in patients with dementia. *Health Qual. Life Outcomes* **2004**, *2*, 52. [[CrossRef](#)] [[PubMed](#)]
4. Laske, C.; Sohrabi, H.R.; Frost, S.M.; López-De-Ipiña, K.; Garrard, P.; Buscema, M.; Dauwels, J.; Soekadar, S.R.; Mueller, S.; Linnemann, C.; et al. Innovative diagnostic tools for early detection of Alzheimer's disease. *Alzheimer's Dement.* **2015**, *11*, 561–578. [[CrossRef](#)]
5. Hane, F.T.; Robinson, M.; Lee, B.Y.; Bai, O.; Leonenko, Z.; Albert, M.S. Recent Progress in Alzheimer's Disease Research, Part 3: Diagnosis and Treatment. *J. Alzheimer's Dis.* **2017**, *57*, 645–665. [[CrossRef](#)] [[PubMed](#)]
6. Bäckman, L.; Jones, S.; Berger, A.-K.; Laukka, E.J.; Small, B.J. Cognitive impairment in preclinical Alzheimer's disease: A meta-analysis. *Neuropsychology* **2005**, *19*, 520–531. [[CrossRef](#)]
7. Deramecourt, V.; Lebert, F.; Debachy, B.; Mackowiak-Cordoliani, M.A.; Bombois, S.; Kerdraon, O.; Buée, L.; Maurage, C.-A.; Pasquier, F. Prediction of pathology in primary progressive language and speech disorders. *Neurology* **2009**, *74*, 42–49. [[CrossRef](#)]
8. Szatloczki, G. Speaking in Alzheimer's disease, is that an early sign? Importance of changes in language abilities in Alzheimer's disease. *Front. Aging Neurosci.* **2015**, *7*, 1–7. [[CrossRef](#)]
9. Meilan, J.J.; Martinez-Sanchez, F.; Carro, J.; Carcavilla, N.; Ivanova, O. Voice Markers of Lexical Access in Mild Cognitive Impairment and Alzheimer's Disease. *Curr. Alzheimer Res.* **2018**, *15*, 111–119. [[CrossRef](#)]
10. Nebes, R.D.; Brady, C.B.; Huff, F.J. Automatic and attentional mechanisms of semantic priming in alzheimer's disease. *J. Clin. Exp. Neuropsychol.* **1989**, *11*, 219–230. [[CrossRef](#)]
11. McKhann, G.M.; Knopman, D.S.; Chertkow, H.; Hyman, B.T.; Jack, C.R.; Kawas, C.H.; Klunk, W.E.; Koroshetz, W.J.; Manly, J.J.; Mayeux, R.; et al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dement. J. Alzheimer's Assoc.* **2011**, *7*, 263–269. [[CrossRef](#)] [[PubMed](#)]
12. Pulido, M.L.B.; Hernández, J.B.A.; Ballester, M.A.F.; Gonzalez, C.M.T.; Mekyska, J.; Smekal, Z. Alzheimer's disease and automatic speech analysis: A review. *Expert Syst. Appli.* **2020**, *150*, 113213. [[CrossRef](#)]
13. Kim, Y.; Lee, H.; Provost, E.M. Deep learning for robust feature generation in audiovisual emotion recognition. In Proceedings of the Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference, Vancouver, BC, Canada, 26–31 May 2013; pp. 3687–3691.
14. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436. [[CrossRef](#)] [[PubMed](#)]

15. Brabenec, L.; Mekyska, J.; Galaz, Z.; Rektorova, I. Speech disorders in Parkinson's disease: Early diagnostics and effects of medication and brain stimulation. *J. Neural Transm.* **2017**, *124*, 303–334. [CrossRef]
16. Khodabakhsh, A.; Yeşil, F.; Güner, E.; Demiroglu, C. Evaluation of linguistic and prosodic features for detection of Alzheimer's disease in Turkish conversational speech. *EURASIP J. Audio Speech Music. Process.* **2015**, *2015*, 189. [CrossRef]
17. Tanaka, H.; Adachi, H.; Ukita, N.; Kudo, T.; Nakamura, S. 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, Tokyo, Japan, 12–16 November 2016; pp. 261–265.
18. Rentoumi, V.; Paliouras, G.; Danasi, E.; Arfani, D.; Fragkopoulou, K.; Varlokosta, S.; Papadatos, S. Automatic detection of linguistic indicators as a means of early detection of Alzheimer's disease and of related dementias: A computational linguistics analysis. In Proceedings of the Cognitive Infocommunications (CogInfoCom), 8th IEEE International Conference, Debrecen, Hungary, 11–14 September 2017; pp. 33–38.
19. De Ipiña, K.L.; Alonso, J.B.; Solé-Casals, J.; Barroso, N.; Faúndez, M.; Ecay, M.; Travieso, C.; Ezeiza, A.; Estanga, A. Alzheimer disease diagnosis based on automatic spontaneous speech analysis. In Proceedings of the International Joint Conference on Computational Intelligence, IJCCI 2012, Barcelona, Spain, 5–7 October 2012; pp. 698–705.
20. Roy, D.; Pentland, A. Automatic spoken affect classification and analysis. In Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, Killington, VT, USA, 14–16 October 1996; pp. 363–367.
21. Lopez-De-Ipiña, K.; Alonso, J.; Travieso, C.; Eguiraun, H.; Ecay, M.; Ezeiza, A.; Barroso, N.; Martinez-Lage, P. Automatic analysis of emotional response based on non-linear speech modeling oriented to Alzheimer disease diagnosis. In Proceedings of the INES 2013—IEEE 17th International Conference on Intelligent Engineering Systems, San Jose, Costa Rica, 19–21 June 2013; pp. 61–64.
22. Alonso, J.B.; Cabrera, J.; Medina, M.; Travieso, C.M. New approach in quantification of emotional intensity from the speech signal: Emotional temperature. *Expert Syst. Appl.* **2015**, *42*, 9554–9564. [CrossRef]
23. Lopez-De-Ipiña, K.; Alonso, J.B.; Solé-Casals, J.; Barroso, N.; Henriquez, P.; Faundez-Zanuy, M.; Travieso, C.M.; Ecay-Torres, M.; Martinez-Lage, P.; Eguiraun, H. On Automatic Diagnosis of Alzheimer's Disease Based on Spontaneous Speech Analysis and Emotional Temperature. *Cogn. Comput.* **2013**, *7*, 44–55. [CrossRef]
24. Sohn, J.; Kim, N.S.; Sung, W. A statistical model-based voice activity detection. *IEEE Signal Process. Lett.* **1999**, *6*, 1–3. [CrossRef]
25. Hernández-Domínguez, L.; García-Canó, E.; Ratt, S.; Sierra-Martínez, G. 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, Berlin, Germany, 11 August 2016; pp. 10–15.
26. Vincze, V.; Gosztolya, G.; Tóth, L.; Hoffmann, I.; Sztalóczyki, G.; Bánréti, Z.; Pákáski, M.; Kálmán, J.; Erk, K.; Smith, N.A. Detecting Mild Cognitive Impairment by Exploiting Linguistic Information from Transcripts. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany, 7–12 August 2016; pp. 181–187.
27. Bhaduri, S.; Das, R.; Ghosh, D. Non-Invasive Detection of Alzheimer's Disease-Multifractality of Emotional Speech. *J. Neurol. Neurosci.* **2016**, *7*, 84. [CrossRef]
28. López-De-Ipiña, K.; Solé-Casals, J.; Eguiraun, H.; Alonso, J.; Travieso, C.; Ezeiza, A.; Barroso, N.; Ecay-Torres, M.; Martinez-Lage, P.; Beitia, B. Feature selection for spontaneous speech analysis to aid in Alzheimer's disease diagnosis: A fractal dimension approach. *Comput. Speech Lang.* **2015**, *30*, 43–60. [CrossRef]
29. Bertram, L.; McQueen, M.B.; Mullin, K.; Blacker, D.; E Tanzi, R. Systematic meta-analyses of Alzheimer disease genetic association studies: The AlzGene database. *Nat. Genet.* **2007**, *39*, 17–23. [CrossRef]
30. De La Mata, M. Metaanálisis del Sistema Olfatorio Como Diagnóstico Precoz en Parkinson y Alzheimer. *Revista de Biología* **2016**, *8*, 102–110.
31. Al-Hameed, S.; Benaissa, M.; Christensen, H. Simple and robust audio-based detection of biomarkers for Alzheimer's disease. In Proceedings of the 7th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT), San Francisco, CA, USA, 13 September 2016; pp. 32–36.
32. Zhou, L.; Fraser, K.C.; Rudzicz, F. Speech Recognition in Alzheimer's Disease and in its Assessment. In Proceedings of the Interspeech 2016, San Francisco, CA, USA, 12–18 September 2016; pp. 1948–1952.
33. Hernández-Domínguez, L.; Ratté, S.; Sierra-Martínez, G.; Roche-Bergua, A. Computer-based evaluation of Alzheimer's disease and mild cognitive impairment patients during a picture description task. *Alzheimer's Dement. Diagn. Assess. Dis. Monit.* **2018**, *10*, 260–268. [CrossRef] [PubMed]
34. López-De-Ipiña, K.; Martínez-De-Lizarduy, U.; Barroso, N.; Ecay-Torres, M.; Martínez-Lage, P.; Torres, F.; Faundez-Zanuy, M. Automatic analysis of Categorical Verbal Fluency for Mild Cognitive impairment detection: A non-linear language independent approach. In Proceedings of the Bioinspired Intelligence (IWOBI), 4th International Work Conference IEEE, San Sebastian, Spain, 10–12 June 2015; pp. 101–104.
35. AMI Corpus. 2006. Available online: <http://groups.inf.ed.ac.uk/ami/corpus/> (accessed on 26 May 2018).
36. Folstein, M.F.; Robins, L.N.; Helzer, J.E. The mini-mental state examination. *Arch. Gen. Psychiatry* **1983**, *40*, 812. [CrossRef] [PubMed]
37. Rockwood, K.; Graham, J.E.; Fay, S. Goal setting and attainment in Alzheimer's disease patients treated with donepezil. *J. Neurol. Neurosurg. Psychiatry* **2002**, *73*, 500–507. [CrossRef] [PubMed]

38. Weiner, J.; Frankenberg, C.; Telaar, D. Towards Automatic Transcription of ILSE—An Interdisciplinary Longitudinal Study of Adult Development and Aging. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), Portorož, Slovenia, 23–28 May 2016.
39. DementiaBank | TalkBank. 2007. Available online: <https://dementia.talkbank.org/access/> (accessed on 26 May 2018).
40. López-de-Ipiña, K.; Alonso, J.B.; Barroso, N.; Faundez-Zanuy, M.; Ecay, M.; Solé-Casals, J.; Travieso, C.M.; Estanga, A.; Ezeiza, A. 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*; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7657, pp. 407–414.
41. López-De-Ipiña, K.; Alonso, J.B.; Barroso, N.; Solé-Casals, J.; Ecay-Torres, M.; Martínez-Lage, P.; Zelarain, F.; Egiraun, H.; Travieso, C.M. Spontaneous speech and emotional response modeling based on one-class classifier oriented to Alzheimer disease diagnosis. In Proceedings of the XIII Mediterranean Conference on Medical and Biological Engineering and Computing, IFMBE Proceedings, Seville, Spain, 23–28 September 2013; Springer: Cham, Switzerland, 2014; Volume 41, pp. 571–574.
42. López-de-Ipiña, K.; Alonso, J.B.; Travieso, C.M.; Solé-Casals, J.; Egiraun, H.; Faundez-Zanuy, M.; Ezeiza, A.; Barroso, N.; Ecay-Torres, M.; Martínez-Lage, P.; et al. On the Selection of Non-Invasive Methods Based on Speech Analysis Oriented to Automatic Alzheimer Disease Diagnosis. *Sensors* **2013**, *13*, 6730–6745. [[CrossRef](#)]
43. Lopez-De-Ipiña, K.; Alonso-Hernández, J.; Solé-Casals, J.; Travieso-González, C.; Ezeiza, A.; Faundez-Zanuy, M.; Calvo, P.; Beitia, B. Feature selection for automatic analysis of emotional response based on nonlinear speech modeling suitable for diagnosis of Alzheimer's disease. *Neurocomputing* **2015**, *150*, 392–401. [[CrossRef](#)]
44. Lopez-De-Ipiña, K.; Martínez-De-Lizarduy, U.; Calvo, P.M.; Mekyska, J.; Beitia, B.; Barroso, N.; Estanga, A.; Tainta, M.; Ecay-Torres, M. Advances on Automatic Speech Analysis for Early Detection of Alzheimer Disease: A Non-linear Multi-task Approach. *Curr. Alzheimer Res.* **2018**, *15*, 139–148. [[CrossRef](#)]
45. De Lizarduy, U.M.; Salomón, P.C.; Vilda, P.G.; Torres, M.E.; de Ipiña, K.L. Alzumeric: A decision support system for diagnosis and monitoring of cognitive impairment. *Loquens* **2017**, *4*, 37. [[CrossRef](#)]
46. Proyectos Fundación CITA Alzheimer. 2017. Available online: <http://www.cita-alzheimer.org/investigacion/proyectos> (accessed on 26 May 2018).
47. Satt, A.; Sorrin, A.; Toledo-Ronen, O.; Barkan, O.; Kompatsiaris, I.; Kokonozi, A.; Tsolaki, M. Evaluation of Speech-Based Protocol for Detection of Early-Stage Dementia. In Proceedings of the Interspeech 2013, Lyon, France, 25–29 August 2013; pp. 1692–1696.
48. Tröger, J.; Linz, N.; Alexandersson, J.; König, A.; Robert, P. 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, Barcelona, Spain, 23–26 May 2017; ACM: New York, NY, USA, 2017; pp. 292–297.
49. Satt, A.; Hoory, R.; König, A.; Aalten, P.; Robert, P.H. Speech-based automatic and robust detection of very early dementia. In Proceedings of the Fifteenth Annual Conference of the International Speech Communication Association, INTERSPEECH-2014, Singapore, 14–18 September 2014; pp. 2538–2542.
50. König, A.; Satt, A.; Sorin, A.; Hoory, R.; Toledo-Ronen, O.; Derreumaux, A.; Manera, V.; Verhey, F.R.J.; Aalten, P.; Robert, P.H.; et al. Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. *Alzheimer's Dement. Diagn. Assess. Dis. Monit.* **2015**, *1*, 112–124. [[CrossRef](#)]
51. Mirzaei, S.; El Yacoubi, M.; Garcia-Salicetti, S.; Boudy, J.; Muvingi, C.K.S.; Cristancho-Lacroix, V. Automatic speech analysis for early Alzheimer's disease diagnosis. In Proceedings of the JETSAN 2017, 6e Journées d'Etudes sur la Télésant, Bourges, France, 31 May–1 June 2017; pp. 114–116.
52. Boyé, M.; Tran, T.M.; Grabar, N. NLP-Oriented Contrastive Study of Linguistic Productions of Alzheimer's and Control People. In Proceedings of the International Conference on Natural Language Processing. Advances in Natural Language Processing, Warsaw, Poland, 17–19 September 2014; pp. 412–424.
53. Luz, S. Longitudinal Monitoring and Detection of Alzheimer's Type Dementia from Spontaneous Speech Data. In Proceedings of the Computer-Based Medical Systems (CBMS), 2017 IEEE 30th International Symposium on IEEE, Thessaloniki, Greece, 22–24 June 2017; pp. 45–46.
54. Asgari, M.; Kaye, J.; Dodge, H. Predicting mild cognitive impairment from spontaneous spoken utterances. *Alzheimer's Dement. Transl. Res. Clin. Interv.* **2017**, *3*, 219–228. [[CrossRef](#)]
55. Sadeghian, R.; Schaffer, J.D.; Zahorian, S.A. Speech Processing Approach for Diagnosing Dementia in an Early Stage. In Proceedings of the Interspeech 2017, Stockholm, Sweden, 20–24 August 2017; pp. 2705–2709.
56. Mueller, K.D.; Kosciak, R.L.; Hermann, B.P.; Johnson, S.C.; Turkstra, L.S. Declines in Connected Language Are Associated with Very Early Mild Cognitive Impairment: Results from the Wisconsin Registry for Alzheimer's Prevention. *Front. Aging Neurosci.* **2018**, *9*, 437. [[CrossRef](#)] [[PubMed](#)]
57. Warnita, T.; Inoue, N.; Shinoda, K. Detecting Alzheimer's Disease Using Gated Convolutional Neural Network from Audio Data. *arXiv* **2018**, arXiv:1803.11344. Available online: <http://arxiv.org/abs/1803.11344> (accessed on 1 May 2018).
58. Wankerl, S.; Nöth, E.; Evert, S. 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 2017, Stockholm, Sweden, 20–24 August 2017.
59. Yancheva, M. Automatic Assessment of Information Content in Speech for Detection of Dementia of the Alzheimer Type. Master's Thesis, University of Toronto, Toronto, ON, Canada, 2016.

60. Abdalla, M.; Rudzicz, F.; Hirst, G. Rhetorical structure and Alzheimer's disease. *Aphasiology* **2018**, *32*, 41–60. [CrossRef]
61. Sirts, K.; Piguet, O.; Johnson, M. Idea density for predicting Alzheimer's disease from transcribed speech. *arXiv* **2017**, arXiv:1706.04473.
62. Thomas, C.; Kešelj, V.; Cercone, N.; Rockwood, K.; Asp, E. Automatic detection and rating of dementia of Alzheimer type through lexical analysis of spontaneous speech. In Proceedings of the IEEE International Conference Mechatronics and Automation, Niagara Falls, ON, Canada, 21 July–1 August 2005; Volume 3, pp. 1569–1574.
63. Carolinas Conversations Collection. About Who We Are. 2008. Available online: <http://carolinaconversations.musc.edu/about/who> (accessed on 26 May 2018).
64. Weiner, J.; Herff, C.; Schultz, T. Speech-Based Detection of Alzheimer's Disease in Conversational German. In Proceedings of the Interspeech 2016, San Francisco, CA, USA, 12–18 September 2016; pp. 1938–1942.
65. Meilán, J.J.G.; Martínez-Sánchez, F.; Carro, J.; Sánchez, J.A.; Pérez, E. Acoustic Markers Associated with Impairment in Language Processing in Alzheimer's Disease. *Span. J. Psychol.* **2012**, *15*, 487–494. [CrossRef] [PubMed]
66. Martínez-Sánchez, F.; Meilán, J.J.G.; Vera-Ferrandiz, J.A.; Carro, J.; Pujante-Valverde, I.M.; Ivanova, O.; Carcavilla, N. Speech rhythm alterations in Spanish-speaking individuals with Alzheimer's disease. *Aging Neuropsychol. Cogn.* **2016**, *24*, 418–434. [CrossRef] [PubMed]
67. Peraita, H.; Grasso, L. *Corpus lingüístico de definiciones de categorías semánticas de personas mayores sanas y con la enfermedad del Alzheimer*; Technical Report; Fundación BBVA: Bilbao, Spain, 2010.
68. Toledo, C.M.; Aluisio, S.M.; dos Santos, L.B.; Brucki, S.M.D.; Tres, E.S.; de Oliveira, M.O.; Mansur, L.L. Analysis of macrolinguistic aspects of narratives from individuals with Alzheimer's disease, mild cognitive impairment, and no cognitive impairment. *Alzheimer's Dement. Diagnosis, Assess. Dis. Monit.* **2018**, *10*, 31–40. [CrossRef]
69. Beltrami, D.; Calzà, L.; Gagliardi, G.; Ghidoni, E.; Marcello, N.; Favretti, R.R.; Tamburini, F. Automatic Identification of Mild Cognitive Impairment through the Analysis of Italian Spontaneous Speech Productions. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), Portorož, Slovenia, 23–28 May 2016; pp. 2089–2093.
70. Nasrolahzadeh, M.; Mohammadpoory, Z.; Haddadnia, J. A novel method for early diagnosis of Alzheimer's disease based on higher-order spectral estimation of spontaneous speech signals. *Cogn. Neurodynamics* **2016**, *10*, 495–503. [CrossRef]
71. Kato, S.; Homma, A.; Sakuma, T.T. Easy Screening for Mild Alzheimer's Disease and Mild Cognitive Impairment from Elderly Speech. *Curr. Alzheimer Res.* **2018**, *15*, 104–110. [CrossRef]
72. Graovac, J.; Kovacevic, J.; Lazetic, G.P. Machine learning-based approach to help diagnosing Alzheimer's disease through spontaneous speech analysis. In Proceedings of the Belgrade BioInformatics Conference, Belgrade, Serbia, 20–24 June 2016; p. 111.
73. Gosztolya, G.; Tóth, L.; Grósz, T.; Vincze, V.; Hoffmann, I.; Sztatlóczi, G.; Pákáski, M.; Kálmán, J. Detecting Mild Cognitive Impairment from Spontaneous Speech by Correlation-Based Phonetic Feature Selection. In Proceedings of the Interspeech 2016, San Francisco, CA, USA, 12–18 September 2016; pp. 107–111.
74. Toth, L.; Hoffmann, I.; Gosztolya, G.; Vincze, V.; Sztatloczki, G.; Banreti, Z. A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech. *Curr. Alzheimer Res.* **2018**, *15*, 130–138. [CrossRef] [PubMed]
75. Tóth, L.; Gosztolya, G.; Vincze, V.; Hoffmann, I.; Sztatloczki, G.; Biro, E.; Zsura, F.; Pakaski, M.; Kalman, J. Automatic Detection of Mild Cognitive Impairment from Spontaneous Speech Using ASR. In Proceedings of the Sixteenth Annual Conference of the International Speech Communication Association, Interspeech-2015, Dresden, Germany, 6–10 September 2015; pp. 2694–2698.
76. St-Pierre, M.-C.; Ska, B.; Béland, R. Lack of coherence in the narrative discourse of patients with dementia of the Alzheimer's type. *J. Multiling. Commun. Disord.* **2005**, *3*, 211–215. [CrossRef]
77. Malekzadeh, G.; Arsalan, G.; Shahabi, M. A comparative study on the use of cohesion devices by normal age persian natives and those suffering from Alzheimer's disease. *J. Med. Sci. Islam. Azad Univ. Mashhad* **2009**, *5*, 153–161.
78. Ahangar, A.; Morteza, S.; Fadaki, J.; Sehhati, A. The Comparison of Morpho-Syntactic Patterns Device Comprehension in Speech of Alzheimer and Normal Elderly People. *Zahedan J. Res. Med. Sci.* **2018**, *20*, e9535. [CrossRef]
79. Khodabakhsh, A.; Demiroglu, C. Analysis of Speech-Based Measures for Detecting and Monitoring Alzheimer's Disease. In *Data Mining in Clinical Medicine*; Humana Press: New York, NY, USA, 2015; pp. 159–173.
80. König, A.; Satt, A.; David, R.; Robert, P. O4-12-02: Innovative Voice Analytics for the Assessment and Monitoring of Cognitive Decline in People with Dementia and Mild Cognitive Impairment. *Alzheimer's Dement.* **2016**, *12*, P363. [CrossRef]
81. Aluisio, S.M.; Cunha, A.; Toledo, C.; Scarton, C. A computational tool for automated language production analysis aimed at dementia diagnosis. In Proceedings of the International Conference on Computational Processing of the Portuguese Language, XII, Demonstration Session, Tomar, Portugal, 13–16 July 2016.
82. Aluisio, S.; Cunha, A.; Scarton, C. Evaluating Progression of Alzheimer's Disease by Regression and Classification Methods in a Narrative Language Test in Portuguese. In Proceedings of the International Conference on Computational Processing of the Portuguese Language, Tomar, Portugal, 13–16 July 2016; pp. 109–114.
83. Nasrolahzadeh, M.; Mohammadpoori, Z.; Haddadnia, J. Analysis of mean square error surface and its corresponding contour plots of spontaneous speech signals in Alzheimer's disease with adaptive wiener filter. *Comput. Hum. Behav.* **2016**, *61*, 364–371. [CrossRef]

84. De Looze, C.; Kelly, F.; Crosby, L.; Vourdanou, A.; Coen, R.F.; Walsh, C.; Lawlor, B.A.; Reilly, R.B. Changes in Speech Chunking in Reading Aloud is a Marker of Mild Cognitive Impairment and Mild-to-Moderate Alzheimer's Disease. *Curr. Alzheimer Res.* **2018**, *15*, 828–847. [CrossRef]
85. Becker, J.T.; Boiler, F.; Lopez, O.L.; Saxton, J.; Mcgonigle, K.L. The Natural History of Alzheimer's Disease: Description of Study Cohort and Accuracy of Diagnosis. *Arch. Neurol.* **1994**, *51*, 585–594. [CrossRef]
86. DeVault, D.; Artstein, R.; Benn, G.; Dey, T.; Fast, E.; Gainer, A.; Georgila, K.; Gratch, J.; Hartholt, A.; Lhommet, M.; et al. SimSensei kiosk: A virtual human interviewer for healthcare decision support. In Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2014, Paris, France, 5–9 May 2014; Volume 2, pp. 1061–1068.
87. Huggins-Daines, D.; Kumar, M.; Chan, A.; Black, A.W.; Ravishankar, M.; Rudnicky, A.I. Pocketsphinx: A free, real-time continuous speech recognition system for hand-held devices. In Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing, ICASSP 2006, Toulouse, France, 14–19 May 2006; pp. 185–188.
88. Littlewort, G.; Whitehill, J.; Wu, T.; Fasel, I.; Frank, M.; Movellan, J.; Bartlett, M. The computer expression recognition toolbox (CERT). In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, FG, Santa Barbara, CA, USA, 21–25 March 2011; pp. 298–305.
89. Stone, M. Specifying Generation of Referring Expressions by Example. In Proceedings of the AAI Spring Symposium on Natural Language Generation in Spoken and Written Dialogue, Palo Alto, CA, USA, 24–26 March 2003; pp. 133–140.
90. Wiki—SAIBA—Mindmakers. Available online: <http://mindmakers.com/projects/SAIBA> (accessed on 1 April 2021).
91. Bickmore, T.; Schulman, D.; Shaw, G. DTask and litebody: Open source, standards-based tools for building web-deployed embodied conversational agents. In Proceedings of the IVA 2009, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Amsterdam, The Netherlands, 14–16 September 2009; Volume 5773 LNAI, pp. 425–431.
92. Poggi, I.; Pelachaud, C.; de Rosis, F.; Carofiglio, V.; de Carolis, B. *Greta. A Believable Embodied Conversational Agent*; Springer: Dordrecht, The Netherlands, 2005; pp. 3–25.
93. Schröder, M. The Semaine Api: Towards a Standards-Based Framework for Building Emotion-Oriented Systems. *Adv. Human-Computer Interact.* **2010**, *2010*, 1–21. [CrossRef]
94. Hartholt, A.; Traum, D.; Marsella, S.C.; Shapiro, A.; Stratou, G.; Leuski, A.; Morency, L.P.; Gratch, J. All together now: Introducing the virtual human toolkit. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Dordrecht, The Netherlands, 2013; Volume 8108 LNAI, pp. 368–381.
95. IBM. SimpleC Advancing Memory Care with IBM Watson and IBM Cloud Solutions. Available online: <https://www.ibm.com/case-studies/w796019n50088s93> (accessed on 15 June 2019).
96. Wilcoxon, F. Some rapid approximate statistical procedures. *Ann. N. Y. Acad. Sci.* **1950**, *52*, 808–814. [CrossRef]
97. Lazar, N.A. Basic Statistical Analysis. In *The Statistical Analysis of Functional MRI Data*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 1–36.
98. Sarhan, A.E. Estimation of the Mean and Standard Deviation by Order Statistics. *Ann. Math. Statist.* **1954**, *25*, 317–328. [CrossRef]
99. Groeneveld, R.A.; Meeden, G. Measuring Skewness and Kurtosis. *Statistician* **1984**, *33*, 391. [CrossRef]
100. Wonnacott, T.H.; Wonnacott, R.J. *Introductory Statistics*, 5th ed.; Part I Basic Probability and Statistics 1 The Nature of Statistics; John Wiley and Sons: Hoboken, NJ, USA, 1990.
101. Corder, G.; Foreman, D. *Nonparametric Statistics: A Step-by-Step Approach*; Wiley: Hoboken, NJ, USA, 2014.
102. Ozone. Auricular Rage ST—Ozone Gaming. Available online: <https://www.ozonegaming.com/es/product/rage-st> (accessed on 5 November 2019).
103. Universidad de Las Palmas de Gran Canaria. Memoria Digital de Canarias—mdC. Available online: <https://mdc.ulpgc.es/> (accessed on 11 December 2019).
104. Yasuda, K.; Kuwabara, K.; Kuwahara, N.; Abe, S.; Tetsutani, N. Effectiveness of personalised reminiscence photo videos for individuals with dementia. *Neuropsychol. Rehabil.* **2009**, *19*, 603–619. [CrossRef]
105. Gowans, G.; Campbell, J.; Alm, N.; Dye, R.; Astell, A.; Ellis, M. Designing a multimedia conversation aid for reminiscence therapy in dementia care environments. In Proceedings of the Conference on Human Factors in Computing Systems, Vienna Austria, 24–29 April 2004; pp. 825–836.
106. Davis, B.H.; Shenk, D. Beyond reminiscence: Using generic video to elicit conversational language. *Am. J. Alzheimer's Dis. Other Demen.* **2015**, *30*, 61–68. [CrossRef] [PubMed]
107. Irazoki, E.; A Garcia-Casal, J.; Sanchez-Meca, J.; Franco-Martin, M. Efficacy of group reminiscence therapy for people with dementia. Systematic literature review and meta-analysis. *Rev. Neurol.* **2017**, *65*, 447–456. [PubMed]
108. Gogate, M.; Dashtipour, K.; Hussain, A. Visual Speech in Real Noisy Environments (VISION): A Novel Benchmark Dataset and Deep Learning-based Baseline System. In Proceedings of the Interspeech 2020, Shanghai, China, 14–18 September 2020.
109. Watanabe, S.; Mandel, M.; Barker, J.; Vincent, E.; Arora, A.; Chang, X.; Khudanpur, S.; Manohar, V.; Povey, D.; Raj, D.; et al. CHiME-6 Challenge: Tackling Multispeaker Speech Recognition for Unsegmented Recordings. In Proceedings of the 6th International Workshop on Speech Processing in Everyday Environments (CHiME 2020), Barcelona, Spain, 4 May 2020.