Simple and robust audio-based detection of biomarkers for Alzheimer's disease

Sabah Al-Hameed [1], *Mohammed Benaissa* [1], *Heidi Christensen* [2]

[1] Department of Electronic and Electrical Engineering, University of Sheffield, United Kingdom
[2]Department of Computer Science, University of Sheffield, United Kingdom
{ssaal-hammed1, m.benaissa, heidi.christensen}@sheffield.ac.uk

Abstract

This paper demonstrates the feasibility of using a simple and robust automatic method based solely on acoustic features to identify Alzheimer's disease (AD) with the objective of ultimately developing a low-cost home monitoring system for detecting early signs of AD. Different acoustic features, automatically extracted from speech recordings, are explored. Four different machine learning algorithms are used to calculate the classification accuracy between people with AD and a healthy control (HC) group. Feature selection and ranking is investigated resulting in increased accuracy and a decrease in the complexity of the method. Further improvements have been obtained by mitigating the effect of the background noise via pre-processing. Using DementiaBank data, we achieve a classification accuracy of 94.7% with sensitivity and specificity levels at 97% and 91% respectively. This is an improvement on previous published results whilst being solely audio-based and not requiring speech recognition for automatic transcription.

Index Terms: Dementia, feature extraction, feature selection procedure, de-noising, classification.

1. Introduction

Recent statistics show an increase of the elderly population around the world according to Alzheimer's Disease International [1] and a relatively high percentage of those will go on to develop dementia [2], [3]. Dementia is used as an umbrella term to describe symptoms of brain disease damaging the cells and neuron synapses caused by e.g. Alzheimer's disease. Dementia symptoms include cognitive decline (affecting amongst other things memory and the person's speech and language), limited motor control, abnormal behavior, loss of memory and judgment, apathy and at a late stage losing the ability to speak [2].

Currently, there is no powerful tool that gives a reliable diagnosis of dementia; rather, the patient has to go through a series of cognitive tests conducted by a professional neurologist for assessments. This process can be very challenging for the patient and involves a certain amount of anxiety and stress. Especially in the case of the early stage detection, complementary tests include the analysis of samples of cerebrospinal fluid taken from the brain and a magnetic resonance brain imaging test [4], [5]. Such methods are invasive, bring discomfort to the patients, are relatively costly and require a significant amount of effort and time.

Finding lightweight, noninvasive diagnostic and/or screening tools, that can be used in the comfort of peoples' homes and inform this process, is therefore of interest. This could be in the form of wearable sensors or incorporated in existing intelligent home technology. This paper describes a relatively simple audio-based tool for detecting biomarkers of dementia in a person's speech.

Changes in speech and language patterns offer valuable clues to the detection of dementia as the speech production process starts in the left hemisphere of the brain [6] and any decline in speech capabilities might indicate the presence of e.g. Alzheimer's disease. Several studies investigated the use of speech-based features for the detection of dementia providing a noninvasive and inexpensive tool that does not require extensive infrastructure or the presence of medical equipment [7], [8]. Automated speech and language analysis methods are potentially powerful tools, especially when using machine learning algorithms capabilities to evaluate the features extracted from the speech. Many methods rely on relatively computationally heavy processing involving speech recognition and the use of natural language processing techniques to achieve some degree of speech understanding at the linguistic level [9]. This makes them unsuitable as low-cost home-based solution and means they are expensive to port to new languages. The alternative solution presented in this paper investigates audio-only processing to address this challenge.

We propose a simple automated method for detecting/screening AD at an early stage. The proposed method is solely based on acoustic features and therefore would only require simple readily available audio technology that can be adapted to suit patient requirement either in terms of being portable or/and wearable. We also explore the performance of different classification techniques applied to a number of acoustic features automatically extracted from the speech recordings obtained from DementiaBank [10]. Finally, we investigate the effect of pre-processing and noise reduction on the performance of the proposed method.

The rest of the paper is organized as follows. Section 2 describes the background. Section 3 describes the experiment setup. Section 4 explores the machine learning. Section 5 presents the results. Finally, section 6 presents the conclusions and future work.

2. Background

Several publications have demonstrated the potential of speech based approaches to identifying dementia. Jarrold et al [11] distinguish between different types of dementia by combining two profiles of features related to acoustic and lexical features collected from 9 controls and 39 patients who have been diagnosed with different types of dementia. Features-based profiles were extracted from structured interviews and used as

10.21437/SLPAT.2016-6

input to a machine learning algorithm. A score of 88% was achieved by using a multi-layer perceptron algorithm.

Orimaye et al [12] proposed a diagnostic method to identify people with AD using nine syntactic and eleven lexical features extracted from transcribed audio files from the DementiaBank dataset. They used a sample size of 242 files for both healthy older people and people with AD. They explored four different machine learning classification algorithms, achieving a 74% classification accuracy using a support vector machine (SVM) classifier with 10% cross-validation.

López et al [8], [13] investigated using features called Emotional Temperature derived from the speech along with acoustic features from 20 healthy subjects and 20 people suffering from dementia. This was done in an attempt to evaluate the importance of the emotions encapsulated in the spontaneous speech and they showed promising results when attempting to differentiate different stages of the disease.

Furthermore König et al [14] conducted an experiment of using four short cognitive vocal tasks with a number of participants divided into three groups: healthy control (HC), people with Mild Cognitive Impairment (MCI) and people with Alzheimer (AD). Their method included pre-processing, analyzing the data and feature extraction from the speech recordings. They were able to distinguish between HC and MCI with an accuracy of 79%, between HC and AD patients with an accuracy of 87%, and between those with MCI and AD with 80% accuracy.

Recently and similar to our work, Fraser et al [15] studied the potential of using linguistic features to identify Alzheimer's disease. They used speech recordings along with their manually transcribed files derived from the DementiaBank data set. They chose 240 speech recordings belonging to a group of 167 people identified as probably or possibly having AD and 233 samples from 97 subjects with no memory complaint. In total, a set of 370 acoustic, lexical and semantic features were extracted and they then applied two machine learning classification algorithms and obtained a highest accuracy of 92% in distinguishing between HC subjects and AD patients using the top 25 ranked features. Although they obtained promising results, their method relies on the accuracy of manually transcribed files, whereas a real system would need the added complexity of a speech recognizer to compute all three types of features. It is unclear how the results of Fraser's system are affected when having to rely on erroneous transcripts from an automatic speech recognizer.

Key aspects of our proposed method compared to state of art are listed as follows:

- Feasible for application in real time and in a range of environments (home/clinic) since our results have been evaluated in the presence of high levels of background noise.
- Higher classification accuracy; outperforming the recent highest score in [15] using the top 20 ranked features
- Robustness: as high classification accuracy is maintained when using higher numbers of features and even when using all 263 features.
- Not reliant on speech recognition to transcribe the audio, so the method could potentially be language independent.

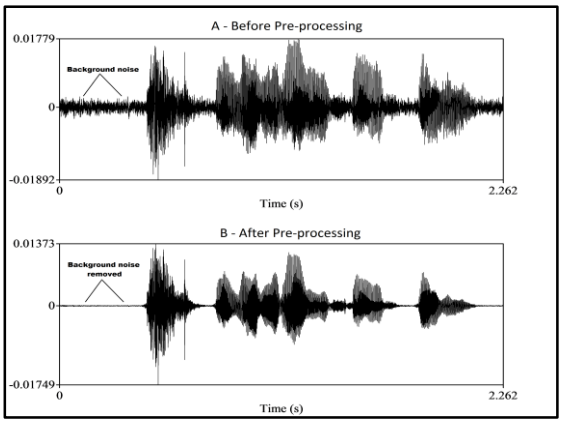

Figure 1. *Speech sample before (A) and after (B) the pre-processing step.*

3. Experimental setup

3.1. **Data set**

We utilized the DementiaBank data set [10], a free access large existing database for Alzheimer's and related dementia diseases collected during longitudinal study conducted by the University of Pittsburgh School of Medicine. A verbal description of the Boston Cookie Theft picture was recorded from people with different types of dementia with an age span from 49 to 90 years as well as from elderly HC subjects with an age range from 46 to 81 years. During the interviews, patients were given the picture and were told to discuss everything they could see happening in the picture. The speech samples were collected through annual visits from the majority of the participants and were transcribed using the CHAT transcription format Mac Whinney [16]. We consider the same sample size used by Fraser et al [15] with a total of 473 recording from 97 HC controls having 233 speech samples and the rest from 167 AD patients diagnosed as possible or probable AD.

3.2. **Pre-Processing**

The first step of the pre-processing is background noise reduction, as the DementiaBank data contains a high level of background noise. Effective de-noising is important to enable accurate features extraction. The spectral noise gating method using version 2.1.1 of the Audacity(R) recording and editing software [17] was applied to the audio without sacrificing the overall speech quality. Initial experimentation was carried out to examine the overall performance with and without the presence of the background noise as shown in Figure (1).

Next, using Praat [18], the instructor utterance was removed from the recordings and the audio files were converted from MP3 to mono wave; the sampling frequencies were kept unchanged.

3.3. **Features extraction**

In our study we focused on extracting acoustic features only and investigating the effectiveness of these features in detecting dementia at an early stage. This would avoid relying on the need for manually transcribed files or indeed the problems around achieving reliable speech recognition results in challenging far-field acoustic conditions.

Table 1. *Summary of all the features.*

#	Features set	Description
1	First Group (24) features	Task completion time
		Pitch variation features (mean, median, STD, Min and Max)
		Mean periods and STD periods
		Fraction of locally unvoiced frames and degree of voice breaks
		Jitter: (local, local-absolute, the relative average perturbation (rap), five-point perturbation quotient (ppq5) and the average absolute difference (ddp).
		Shimmer: (local, local-dB, three-point amplitude perturbation (apq3), five-point amplitude perturbation quotient (apq5), eleven-point amplitude perturbation quotient (apq11) and the average absolute difference (dda).
		Mean of autocorrelation
		Mean noise-to-harmonics ratio
		Mean harmonics-to-noise ratio
2	Second Group (17) features	Max, mean, median and STD of speech segment length >=0.4 sec
		No. of pauses (pause length of >=1ms are considered)
		Total speech & silent durations for the segments >= 0.4 sec
		Max, mean, median and STD of silent segment length >=0.4 sec
		Total silent length >=0.4 sec. including the pauses
		Number of speech and silent segments >=0.4 sec.
		Mean and STD of pauses and total duration of the pauses
3	Third Group (222) features	26 Spectral centroid coefficients
		26 Filter bank energy coefficients
		First 42 MFCC coefficients and their skewness, kurtosis, mean with kurtosis and skewness of the mean

Table (1) summarizes all 263 features extracted. The first group of features includes the pitch statistics, mean and standard deviation of periods, degree of voice breaks, fraction of locally unvoiced frames, and the voice quality measures including harmonic-to-noise ratio, mean of autocorrelation and noise-to-harmonic ratio. Various features related to jitter and shimmer scales were also extracted in accordance with [19], [8] using Praat [18].

The second group of features was derived by applying machine classification algorithms to identify speech/non-speech segments. This is done by windowing the audio files into 40ms frames with 50% overlapping window. For each frame we calculate the short time energy, zero crossing rate and the correlation coefficients. The three measures with labeled frames are used to train and build a voice activity detection (VAD) classifier using predefined frame samples randomly selected from the data. Next we used the VAD to label each frame for the rest of the audio files. The results from the VAD classifier gives us duration statistics for speech/silent regions with the amount of pauses presented in the recordings [20].

The last group of features includes the Mel Frequency Cepstral Coefficients extracted using the method mentioned by [21], including: the first 42 MFCC coefficients and their skewness, kurtosis, means and kurtosis and skewness of the means) previously used by [15] in addition to the first 26 coefficients for both filter bank energies and spectral centroid.

Table 2. *Top 20 rank features as automatically selected by the Weka attribute selection function.*

#	Features	Rank – Weight
1.	MFCC2	82.241
2.	Kurtosis -MFCC30	81.606
3.	Mean-MFCC30	81.606
4.	Skewness - MFCC2	80.972
5.	Mean-MFCC16	80.126
6.	Filter bank energy 22	79.069
7.	Spectral centroid -C14	79.069
8.	MFCC30	77.801
9.	Kurtosis -MFCC16	77.589
10.	Filter bank energy 2	77.589
11.	Filter bank energy 24	77.167
12.	MFCC1	76.532
13.	Filter bank energy 15	76.052
14.	Kurtosis -MFCC2	73.995
15.	Filter bank energy 20	72.304
16.	Filter bank energy 13	65.961
17.	No. of silent segments	61.522
18.	Fraction of locally unvoiced frames	59.830
19.	Minimum silent segments length	57.928
20.	Median pitch	49.48

4. Classification

4.1. **Automatic classification**

We used the capability and accuracy of the automated machine learning algorithms to measure the potential of the acoustic features to distinguish between AD patients and HC subjects. We applied four different classifiers: Bayesian Networks (BN), Trees-Random Forest (RF), AdaboostM1 (AB) and Meta- Bagging (MB). We used the Weka [22] software for running the experiments, with k-fold cross validation, in which we randomly divide the data into K equal-sized parts. We leave out part k, fit the model to the other K-1 parts (combined), and then obtain predictions for the left-out kth part. This is done in turn for each part k= 1, 2,..K [23], and then the results were averaged to obtain the final result. In our study we used k=10 as a cross validation.

4.2. **Feature selection**

Due to the variety and high number of features extracted as well as supporting the idea of simplicity, we applied a feature

Table 3. *Shows the performance under different running configurations.*

#	Machine Learning Algorithm	1st Configuration: 263 features	2nd Configuration: Top 22 features	3rd Configuration: Pre-processing with 263 features	4th Configuration: Pre-processing with top 20 features
		263 features	Accuracy %	Accuracy %	Accuracy %
1.	Bayes Net	89.64	91.75	**93.66**	**94.71**
2.	Meta-Bagging	90.27	**92.38**	93.65	92.6
3.	Random forest	**90.90**	91.96	91.96	92.8
4.	AdaBoost M1	82.87	85.83	91.96	91.75

selection technique. This is used to rank the features, to explore the lowest number of features that provides the best classification accuracy, and to avoid overfitting the data. For the unprocessed data (with the presence of background noise), this function automatically selected the top 22 features based on their ranks, whilst 20 features were selected when working with the files that had been pre-processed. Table (2) lists the top (20) features automatically selected by Weka using the built-in attribute selection technique function.

5. Results

We used the four machine learning algorithms stated in section 4.1 to achieve the final results in four different configurations resulting from using pre-processing or not, and using the full (263) or the reduced features sets we also calculate the sensitivity and specificity for the highest score achieved for the 2nd and the 4th configurations.

Table (3) lists the accuracies obtained for the four different configurations. The highest classification accuracy achieved was 94.71% using the (BN) classifier, running under the fourth configuration followed by configuration three with 93.66% using (BN) classifier, while configurations two and one score 92.38% and 90.90% using (MB) and (RF) classifiers respectively.

By adopting a pre-processing step and extracting fewer, better quality features for the classifiers, the highest accuracy was achieved.

The sensitivity and specificity for the 2nd configuration was 92.00%. Only 19 patients from 240 and 17 HC subjects from 233 were incorrectly classified, but when comparing with the 4th configuration, only 7 AD patients were incorrectly classified making the sensitivity level at 97.00%. However the specificity of the 4th configuration was slightly reduced to 91.00% (only 21 HC were misclassified)

Our results reveal two important facts: first, the majority of the features have the potential to identify dementia even when all the features have been utilized by the classifiers (93.66% classification accuracy using the full 263 features compared to 94.71% when feature selection is used). This is in contrast to what had been reported by [15], as their results showed a sharp drop off in the case of using all of the features (from 92.01% classification accuracy with feature selection down to 79% without).

The first group of features measures the perturbation of the fundamental frequency reflecting the defects on vocal folds closing and opening times. This is, captured by the shimmer and jitter parameters as they measure the differences of amplitude and cycles of consecutive periods. Also it is known that AD patients produce more noise in their speech due to the fluctuations in the airflow, caused by incomplete vocal fold closure than do healthy subjects. This is measured by the harmonics to noise ratio (HNR) feature, previously demonstrated by [24], [14]. Pauses and number of silent segments are more prevalent in AD patients as they tend to shorten the speech segments in contrast to HC subjects. This is because the AD patients most of the time find that talking requires much effort and concentration. The MFCC features, although they are well-known as standards in speech recognition systems, capture important separation between the two groups as they relate to the articulators (libs and tongue) control ability, that is decreased in AD patients [25].

Secondly our proposed method is robust and very capable of identifying dementia patients from healthy subjects even in the presence of significant background noise. These facts support our proposition for using only acoustic features for automatic detection and/or screening of AD at a low cost and within the home environment.

6. Conclusions and future work

Speech and language impairment serve as a strong evidence for Alzheimer's disease detection and it can be used to indicate its severity over the time [26].

In our study, for the same data set (based on short speech recordings from a picture description task.), but using only acoustic features, higher accuracy results were obtained, in distinguishing between HC subjects and AD patients, than those reported in the most recent state of the art [15].

Furthermore, we used acoustic features derived automatically from the speech recordings without the addition of any lexical or syntactic features that rely on complex speech recognition technology as in [9].

In this paper, we proposed a simple high accuracy automated method that can be used in the clinic and/or at home to guide the diagnosing and/or screening of dementia by using just speech. In the future, we plan to investigate more features and to test the performance of our method with different datasets to classify between neurodegenerative dementia patients and people with functional memory disorders. The analysis will be applied to the conversations between the neurologists and patients during their visit to the memory clinic.

7. References

[1] C. F. Martin Prince, Renata Bryce, "World Alzheimer Report - The benefits of early diagnosis and intervention World Alzheimer Report," *Alzeheimer's Dis. Int.*, p. 72, 2011.

[2] J. Berger, "The age of biomedicine: current trends in traditional subjects," *J. Appl. Biomed.*, vol. 9, no. 2, pp. 57–61, 2011.

[3] V. Taler and N. a Phillips, "Language performance in Alzheimer's disease and mild cognitive impairment: a comparative review.," *J. Clin. Exp. Neuropsychol.*, vol. 30, no. 5, pp. 501–556, 2008.

[4] C. Laske, H. R. Sohrabi, S. M. Frost, K. López-De-Ipiña, P. Garrard, M. Buscema, J. Dauwels, S. R. Soekadar, S. Mueller, C. Linnemann, S. A. Bridenbaugh, Y. Kanagasingam, R. N. Martins, and S. E. O'bryant, "Innovative diagnostic tools for early detection of Alzheimer's disease," *Alzheimer's Dement.*, vol. 11, no. 5, pp. 561–578, 2015.

[5] K. S. Santacruz and D. Swagerty, "Early diagnosis of dementia," *Am Fam Physician*, vol. 63, no. 4, pp. 703–713,717–718, 2001.

[6] B. Klimova Q1 and K. Kuca, "Speech and language impairments in dementia – a mini review," *J. Econ. Financ. Adm. Sci.*, pp. 1–7, 2016.

[7] R. Bucks, S. Singh, J. Cuerden, and G. Wilcock, "Analysis of spontaneous, conversational speech in dementia of Alzheimer type: Evaluation of an objective technique for analysing lexical performance," *Aphasiology*, vol. 14, no. July 2015, pp. 71–91, 2000.

[8] K. López-de-ipi, M. Ecay-torres, P. Martinez-lage, and B. Beitia, "Feature selection for spontaneous speech analysis to aid in Alzheimer ' s disease diagnosis : A fractal dimension approach," vol. 30, pp. 43–60, 2014.

[9] D. Hakkani-Tur, D. Vergyri, and G. Tur, "Speech-based automated cognitive status assessment," *Proc. Interspeech 2010*, no. September, pp. 258–261, 2010.

[10] "Dementia Bank." [Online]. Available: https://talkbank.org/DementiaBank/. [Accessed: 10-Dec-2015].

[11] W. Jarrold, B. Peintner, D. Wilkins, D. Vergryi, C. Richey, M. L. Gorno-Tempini, and J. Ogar, "Aided diagnosis of dementia type through computer-based analysis of spontaneous speech," *Proc. Work. Comput. Linguist. Clin. Psychol. From Linguist. Signal to Clin. Real.*, pp. 27–37, 2014.

[12] S. O. Orimaye and K. J. Golden, "Learning Predictive Linguistic Features for Alzheimer ' s Disease and related Dementias using Verbal Utterances," *Proc. Work. Comput. Linguist. Clin. Psychol. From Linguist. Signal to Clin. Real.*, pp. 78–87, 2014.

[13] K. López-de-Ipiña, J. B. Alonso, N. Barroso, M. Faundez-Zanuy, M. Ecay, J. Solé-Casals, C. M. Travieso, A. Estanga, and A. Ezeiza, "New approaches for Alzheimer's disease diagnosis based on automatic spontaneous speech analysis and emotional temperature," *Ambient Assist. Living Home Care. Lect. Notes Comput. Sci.*, vol. 7657, pp. 407–414, 2012.

[14] A. König, A. Satt, A. Sorin, R. Hoory, O. Toledo-Ronen, A. Derreumaux, V. Manera, F. Verhey, P. Aalten, P. H. Robert, and R. David, "Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease," *Alzheimer's Dement. Diagnosis, Assess. Dis. Monit.*, vol. 1, no. 1, pp. 112–124, 2015.

[15] K. C. Fraser, J. A. Meltzer, and F. Rudzicz, "Linguistic features identify Alzheimer's disease in narrative speech," *J. Alzheimer's Dis.*, vol. 49, no. 2, pp. 407–422, 2015.

[16] J. R. Booth, B. Mac Whinney, and Y. Harasaki, "Developmental differences in visual and auditory processing of complex sentences," *Child Dev.*, vol. 71, no. 4, pp. 981–1003, 2000.

[17] "Audacity® is free, open source, cross-platform software for recording and editing sounds." [Online]. Available: http://www.audacityteam.org/. [Accessed: 15-Jan-2016].

[18] P. Boersma and D. Weenink, "Praat: doing phonetics by computer." [Online]. Available: http://www.fon.hum.uva.nl/praat/. [Accessed: 05-Jan-2016].

[19] J. J. G. Meilán, F. Martínez-Sánchez, J. Carro, D. E. López, L. Millian-Morell, and J. M. Arana, "Speech in Alzheimer's disease: can temporal and acoustic parameters discriminate dementia?," *Dement. Geriatr. Cogn. Disord.*, vol. 37, no. 5–6, pp. 327–334, 2014.

[20] B. Roark, M. Mitchell, J. P. Hosom, K. Hollingshead, and J. Kaye, "Spoken language derived measures for detecting mild cognitive impairment," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 19, no. 7, pp. 2081–2090, 2011.

[21] J. I. Godino-Llorente, P. Gómez-Vilda, and M. Blanco-Velasco, "Dimensionality reduction of a pathological voice quality assessment system based on gaussian mixture models and short-term cepstral parameters," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 10, pp. 1943–1953, 2006.

[22] "Weka 3: Data Mining Software in Java." [Online]. Available: http://www.cs.waikato.ac.nz/ml/weka/. [Accessed: 01-Feb-2016].

[23] P. Taylor, R. R. Picard, and R. D. Cook, "Cross-Validation of Regression Models Cross-Validation of Regression Models," vol. 79, no. April 2013, pp. 37–41, 2012.

[24] D. E. L. Juan José G. Meilán, Francisco Martínez-Sánchez, Juan Carro, "Speech in alzheimer's disease: Can temporal and acoustic parameters discriminate dementia?," *Dement. Geriatr. Cogn. Disord.*, vol. 37, no. 5–6, pp. 327–334, 2014.

[25] A. Tsanas, M. a. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of Parkinsons disease," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 5, pp. 1264–1271, 2012.

[26] M. Yancheva, K. Fraser, and F. Rudzicz, "Using linguistic features longitudinally to predict clinical scores for Alzheimers disease and related dementias." in *6th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT)*, 2015, p. 134.