Dementia Classification using Acoustic Descriptors Derived from Subsampled Signals

Ayush Triapthi, Rupayan Chakraborty, Sunil Kumar Kopparapu
TCS Research and Innovation-Mumbai, INDIA
{t.ayush, rupayan.chakraborty, sunilkumar.kopparapu}@tcs.com

***Abstract*—Dementia is a chronic syndrome characterized by deteriorating cognitive functions, thereby impacting the person's daily life. It is often confused with decline in normal behavior due to natural aging and hence is hard to diagnose. Although, prior research has shown that dementia affects the subject's speech, but it is not studied which frequency bands are being affected, and up to what extent, that in turn might influence identifying the different stages of dementia automatically. This work investigates the acoustic cues in different subsampled speech signals, to automatically differentiate Healthy Controls (HC) from stages of dementia such as Mild Cognitive Impairment (MCI) or Alzheimer's Disease (AD). We use the Pitt corpus of DementiaBank database, to identify a set of features best suited for distinguishing between HC, MCI and AD speech, and achieve an F-score of 0.857 which is an absolute improvement of 2.8% over the state of the art.**

***Index Terms*—Dementia, classification, feature reduction, Alzheimer's disease, mild cognitive impairment**

I. Introduction

Dementia is a broad category of brain diseases that causes a long-term, gradual decrease in the ability to think and remember. With the progression of the disease, the cognitive function of a person declines beyond what is expected due to natural aging. Alzheimer's disease is the most common cause of dementia, while Parkinson's disease, Huntington's disease, Frontotemporal lobar degeneration and stroke being other minor causes. It affects memory, thinking, orientation, comprehension, calculating, learning capacity, language and judgment; thereby making it hard for the person to remember important information, solve problems, plan the day and other daily activities. Alzheimer's disease (AD) accounts for 60% to 70% of cases of dementia worldwide [1]. Manual diagnosis of AD severity through a series of cognitive tests such as the mini mental state examination (MMSE) [2] is the gold standard for assessment of progression of the disease. It is performed by specialized neurologists and geriatricians. Instrumentation based assessment techniques such as examination of cerebrospinal fluid and magnetic resource imaging of the brain have been employed for diagnosing dementia [3]. However, these methods are invasive and cause pain and discomfort to the patients. Hence, it is required to develop methods which are non-invasive and simple to administer while maintaining a trade-off between comfort to a patient and ease of assessment by a clinician.

In previous studies, both linguistic and acoustic parameters have been employed for proper diagnosis of dementia. The use of manual transcriptions to extract linguistic features for dementia classification has been explored in the past. In [4], the authors explored an Artificial Neural Network based dementia classification by using Bag of Words representation. Recent works have also focused on using transcriptions obtained from Automatic Speech Recognition (ASR) engines to further obtain linguistic parameters [5], [6]. A combination of manual transcription and ASR based automatically obtained transcription has also been explored for this purpose [7]. The use of an intelligent virtual agent for detecting early signs of dementia using a combination of acoustic and linguistic features was proposed by the authors in [8]. N-gram based approaches have also been used for automatic detection of AD in the past literature [9], [10]. All these approaches use linguistic features in one form or another. Clearly, these approaches are language dependent and hence disadvantageous. On the other hand, acoustic features provide a language independent framework for dementia classification [11]. This idea was put forward by the authors in [12], where they proposed a system that uses a cognitive-task based framework in addition to other non-verbal features to assess predementia. The performance of acoustic features was found to be comparable to a system using manual transcriptions [13]. High accuracies have been reported for distinguishing between Healthy Control (HC) and AD speech [14]. However, upon introducing more than one stage of dementia i.e. MCI in addition to AD, confusion increases and thus accuracy reduces significantly [9].

In this paper, we propose a method to classify dementia into one of the three classes namely, Healthy Control (HC), Mild Cognitive Impairment (MCI) or Alzheimer's Disease (AD) by using acoustic features only. Towards this, we first subsample the signal by low-pass filtering the speech signal with equally spaced cut-off frequencies. Then we obtain a set of low level descriptors (LLDs) corresponding to the low-pass filtered (LPF) signals at the segment level (i.e. 20ms). To extract the relevant information regarding how those LLDs vary at supra-segmental level, we extract high level descriptors (HLDs) by using statistical functionals (upto 4^{th} order) on top of the LLDs. Further, HLD features obtained from each of the subsampled signals are concatenated, both preceded and followed by a correlation based feature selection, to ensure the reduction of intra- and inter-band redundant information. The selected features thereby obtained are employed for the purpose of dementia classification. Experimenting with patient excerpts from a clinician-participant conversations taken from

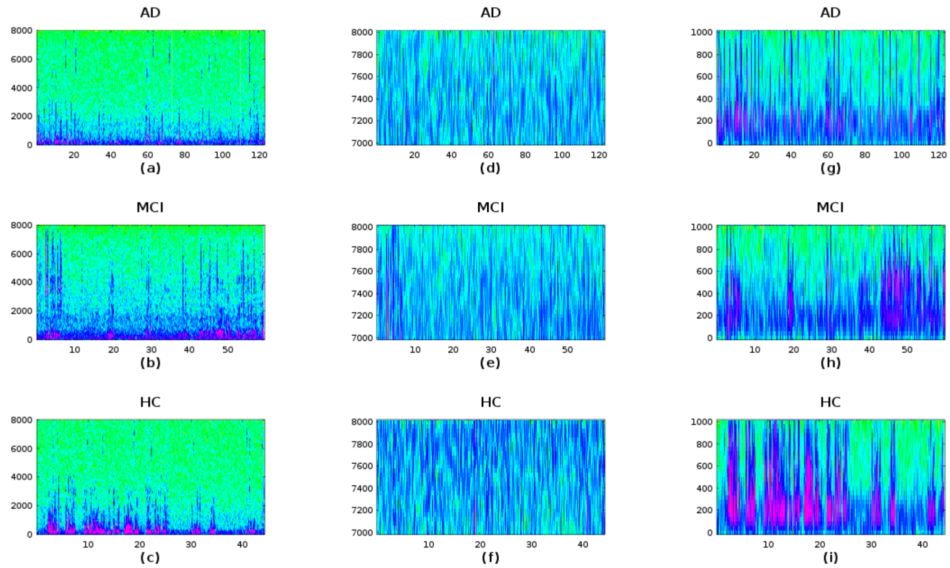

Fig. 1. Illustration of the significance of low frequency information in detection of dementia. (a)-(c); (d)-(f) and (g)-(i) represents spectrogram of the original speech signal; high-pass filtered speech signal with a cut-off frequency 7 kHz and low-pass filtered speech signal with a cut-off frequency 1 kHz respectively for AD, MCI and HC speech.

Pitt corpus of DementiaBank database, and for 3 class dementia classification task (i.e. HC, MCI, AD), we observed F-Score of 0.794, so far the best score reported in the literature to the best of our knowledge. Moreover, we conducted experiments for 3 binary classification tasks (HC-MCI, AD-MCI, AD-HC) individually, and also for combining them for 3 class classification task (as done in [21]), by using our proposed features. All these experiments clearly indicate that our proposed method outperforms (by a margin of around 3%) all the previously reported approaches. Further, to investigate the usefulness of our proposed features, we observed that band-limited speech less than 3 kHz was contributing significantly for dementia classification, with half ($\approx 47\%$) of the total number of selected relevant features.

The rest of the paper is organized as follows. Section II describes the motivation behind extracting the proposed set of features. The algorithmic steps involved in the feature extraction process are explained in section III. In Section IV, we explain the experimental setup. Section V we analyze the results obtained by using the proposed method, while we conclude in section VI.

II. Motivation

The objective of this paper is classification of a given speech utterance into one of the three classes - HC, MCI and AD. Earlier work has focused on extracting features from the speech signal sampled at a frequency of 16kHz which contains spectral imformation in the range of 0 to 8kHz. Low frequency information in speech is attributed to the voice source related information and has been used in previous studies to analyze disordered speech. The emergence of voice low tone to high tone ratio (VLHR) (defined as the power ratio of low frequency to high frequency energy obtained by dividing the voice spectrum with a specific cut-off frequency), was motivated by the interest that it captures the low frequency information in speech. Disordered speech is often characterized by introducing pole-zero pairs in the low frequency region. Motivated by all these observations, in this paper we explore feature extraction after low-pass filtering the speech signal.

Figure 1 represents the spectrogram of the original speech signal (left column) and their high-pass (middle column) and low-pass filtered (right column) version. It is clearly visible that the low frequency contains important characteristics to distinguish between the three classes. On the other hand, there is no significant difference in the high frequency spectrograms for different dementia levels. Therefore, we propose a methodology wherein feature extraction is performed not only on the original speech sample (sampled at 16 kHz), but also for a low-pass filtered (LPF) signal. For this purpose, we filter the original signal by choosing equally spaced cut-off frequencies in the range of 1 kHz to 7 kHz before feature extraction. We then apply a correlation based feature selection to identify the best set of features tailored to best discriminate between the three dementia classes.

III. Feature Extraction Methodology

The proposed set of features are extracted in two steps. In the first step, a set of low level descriptors (LLDs) are extracted from the signal while in the second step computing statistical functionals corresponding to the LLDs is performed. The process is then repeated for LPF signals with different cut-off frequencies. Since the speech utterances in the dataset are noisy, we denoise the signal at the front-end. Consider $S(\omega)$, $X(\omega)$ and $N(\omega)$ to be the spectral domain representation of noisy $s_n(t)$, denoised $s(t)$ and noise signal $n(t)$. Then the spectral subtraction process is defined as,

$$|\acute{X}(\omega)| = |S(\omega)| - \upsilon|\hat{N}(\omega)| \quad (1)$$

where $|\acute{X}(\omega)|$ is the estimate of the original signal spectra, and $|\hat{N}(\omega)|$ is the time averaged noise spectra, which we can get by using minimum mean-square error based optimal estimation [15]. Time domain signal is restored by converting the spectra that we get after combining $|\acute{X}(\omega)|$ with the phase of $S(\omega)$. The restored signal $s(t)$ is used for further processing. The denoised signal is then passed through a set of low-pass filters, each designed for a specific cut-off frequency ω_c. The ω_c is varied from 1 kHz to 7 kHz and the corresponding filtered signal being represented by $s_f(t)$ with f varying from 1 to 8 where, $f = 8$ represents the original denoised signal $s(t)$.

Further, a set of features is extracted from the set of LPF signals obtained and is represented as $\phi_i(s_f(t))$, where "i" denotes different LLDs such as Mel Frequency Cepstral Coefficients (MFCCs), Zero Crossing Rate (ZCR), Fast Fourier Transform (FFT) magnitude etc. extracted from the signal $s_f(t)$. In order to obtain a set of high level descriptors (HLDs), number of statistical functionals are applied on these extracted LLDs. We denote $\psi_j(\phi_i(s_f(t)))$ as the HLD extraction process. where, ψ_j is a statistical function where different values of "j" denotes different functionals such as mean, standard deviation, kurtosis, skewness etc.

By performing the above mentioned analysis, for a signal $s_f(t)$, we obtain a feature vector

$$\mathbb{F}_f = [\psi_j(\phi_i(s_f(t)))]_{i=1,j=1}^{M,N} \quad (2)$$

where M is the number of LLDs and N is the number of HLDs. Subsequently, we apply a correlation based feature reduction technique (ζ) that evaluates the worth of a subset of features by considering the individual predictive ability of each feature along with the degree of redundancy between them. The reduced set of features are represented by $\mathbb{F}_f^{'} = \zeta(\mathbb{F}_f)$.

Finally, we concatenate the reduced set of features obtained from each of the LPF signals and subsequently apply feature reduction function again to obtain the proposed set of features, $F = [\zeta[\mathbb{F}_f^{'}]]_{f=1}^{8}$.

IV. Experimental Setup

A. Data

We use the Pitt corpus [16] from the DementiaBank dataset, collected at University of Pittsburgh School of Medicine for all our experiments. It comprises of clinician-patient audio interviews, manual transcripts and subjective assessment of the patients cognitive state as a longitudinal study over the span of four years. Data corresponding to four different linguistic-cognition tasks namely, picture description, fluency, recall and sentence construction is provided in the dataset. In this work, we use the audio recordings and subjective assessment from the picture description task, which is a verbal description of the Boston Cookie Theft picture. It was recorded from people in the age group 49 to 90 years with different levels of dementia, as well as from the healthy (HC) subjects with an age range 46 to 81 years. During the interviews, patients were asked to discuss everything they could see happening in the cookie theft picture.

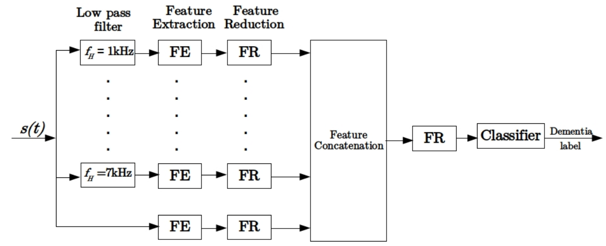

Fig. 2. Overview of the proposed Dementia classification system

We consider a total 613 samples, comprising of 242 HC, 116 MCI and 255 AD samples from 67 HC, 19 MCI and 168 AD participants. We use the speaker timing information provided in the transcripts, to remove the clinician turns from the recordings, retaining only the participant speech.

B. Feature Extraction and Classification

Figure 2 depicts the complete proposed dementia classification system. The raw speech signal is denoised by using the spectral subtraction technique as mentioned in section III by using the PRAAT toolkit [17]. We used the openSMILE toolkit [18] configured to use the large openSMILE emotion feature set (emolarge) [19], which gives a set of 6552 acoustic features, given an input audio file. The feature set consists of 56 LLDs and their first and second order time-derivatives along with 39 statistical functionals applied to these LLDs.

As described in Section II, we extract the features not only for the original speech signal but also for the low-pass filtered speech signal with equally spaced cut-off frequencies in the range 1 kHz to 7 kHz. Further, we use the attribute evaluator CfsSubsetEval with the search method BestFirst specified in the Weka toolkit [20] for feature reduction. The reduced feature set obtained from the different LPF speech signals and the original signal is then concatenated and another feature reduction using the same specifications is applied on top of the concatenated feature set. We thus obtain an 87 dimension vector corresponding to an audio file which we further use for dementia classification. We evaluate the efficacy of the proposed features to classify dementia by using 10 fold validation. Five different classifiers, namely, Random Forest, Multilayer Perceptron, Sequential Minimal Optimization based Support Vector Classifier, Logistic Regression and BayesNet, are used on the same set of extracted features. Apart from classifying by using a 3-class classifier, we also train 3 binary classifiers for the same 3 class classification task, to investigate the classification performance using the proposed feature set along the lines of [21].

TABLE I
CLASSIFICATION PERFORMANCE USING FEATURES OBTAINED FROM ORIGINAL SPEECH SIGNAL

Classifier	Precision	Recall	F-Score
Random Forest (RF)	**0.775**	**0.775**	**0.762**
BayesNet (BN)	0.755	0.749	0.748
Support Vector Machine (SVM)	0.672	0.680	0.662
Logistic Regression (LR)	0.640	0.649	0.644
Multi Layer Perceptron (MLP)	0.746	0.752	0.748

TABLE II
CLASSIFICATION PERFORMANCES USING FEATURES OBTAINED FROM LPF SPEECH SIGNALS (AT DIFFERENT CUT-OFF FREQUENCIES), BEFORE AND AFTER FEATURE REDUCTION USING RANDOM FOREST CLASSIFIER

	Before Feature Reduction			After Feature Reduction		
Feature	Precision	Recall	F-Score	Precision	Recall	F-Score
$\mathbb{F}_1$	0.592	0.582	0.577	0.675	0.670	0.665
$\mathbb{F}_2$	0.614	0.599	0.593	0.663	0.661	0.656
$\mathbb{F}_3$	0.613	0.597	0.592	0.682	0.674	0.669
$\mathbb{F}_4$	0.607	0.594	0.586	0.668	0.659	0.653
$\mathbb{F}_5$	0.607	0.612	0.604	0.673	0.664	0.657
$\mathbb{F}_6$	0.611	0.594	0.589	0.674	0.657	0.651
$\mathbb{F}_7$	0.599	0.584	0.578	0.677	0.659	0.654
$\mathbb{F}_8$	0.69	0.677	0.667	0.775	0.775	0.762

TABLE III
CLASSIFICATION PERFORMANCE OF DIFFERENT CLASSIFIERS USING OUR PROPOSED FEATURES

Classifier	Precision	Recall	F-Score
Random Forest	0.765	0.762	0.751
BayesNet	**0.798**	**0.794**	**0.794**
Suport Vector Machine	0.692	0.697	0.691
Logistic Regression	0.670	0.669	0.669
Multi Layer Perceptron	0.722	0.728	0.724

V. RESULTS AND ANALYSIS

A. Results

To set the baseline accuracies, we use the emo-large feature set from the original signal for classifying the speech into one of the three categories - HC, MCI or AD. After extracting the feature set, we use a correlation based feature reduction technique to reduce the number of features to 50. Table I shows the performance obtained by using different standard classifiers. The best score was obtained by using Random Forest (RF) classifier as an F-score[1] of 0.762. The RF classifier's performances of the different sets of features obtained by low-pass filtering are tabulated in Table II. As explained in section II, the LPF signal contains voice source information and thus contributes significantly in dementia classification. It is also observed that the features obtained from the LPF signal alone are not sufficient to reliably predict a label for dementia. Hence, we retain the features obtained from different frequency filtered signals and concatenate them together for further classification.

[1]Overall metric (Precision/Recall/F-Score) refers to the weighted average of the metric computed across three classes (HC/MCI/AD). Hence, overall F-Score need not necessarily be equal to the harmonic mean of overall Precision and Recall.

TABLE IV
BINARY CLASSIFICATION PERFORMANCE USING BAYESNET

Classification task	Class	Precision	Recall	F-Score
HC-MCI	HC	0.916	0.901	0.908
	MCI	0.800	0.828	0.814
	Overall	**0.878**	**0.877**	**0.878**
AD-MCI	AD	0.888	0.875	0.881
	MCI	0.733	0.759	0.746
	Overall	**0.840**	**0.838**	**0.839**
AD-HC	AD	0.918	0.831	0.872
	HC	0.838	0.921	0.878
	Overall	**0.879**	**0.875**	**0.875**

TABLE V
CLASSIFICATION ACCURACIES FOR 3 CLASS (COMBINING BINARY CLASSIFIERS) USING OUR PROPOSED FEATURES

Class	Precision	Recall	F-Score
HC	0.904	0.859	0.881
MCI	0.776	0.689	0.729
AD	0.854	0.937	0.893
Overall	**0.859**	**0.859**	**0.857**

Table III shows the results obtained by the proposed set of features by using different classifiers. The best result is obtained by using a BayesNet classifier with an F-score of 0.794. We observe a significant improvement over the baseline where the best obtained F-score was 0.762. This indicates that the proposed set of features is able to discriminate a person's cognitive state in a better way. Further, we use a system of three, two-class classifiers as proposed in [22]. Table IV shows the precision, recall and F-score for each of the two-class BayesNet classifiers using the proposed set of features. The lower F-score for the MCI class may be attributed to the low number of samples available for this class (almost half as compared to HC and AD). Further, we fuse the posterior probabilities of the two-class classifiers to reach a class decision for a given test utterance. The results obtained by fusing 3 binary classifiers are presented in Table V. We obtain an F-score of 0.857, which shows an absolute improvement of 9.5% over the baseline. The confusion matrix for the proposed dementia identification system is depicted in Figure 3.

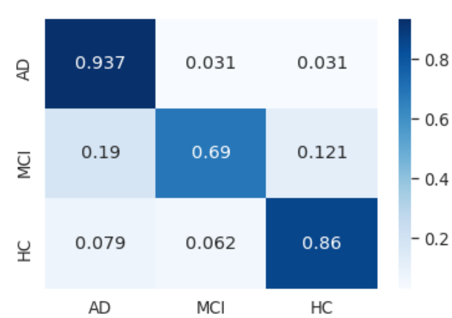

Fig. 3. Class confusion matrix for the system (combining 3 binary classifiers) that uses our proposed features

TABLE VI

CONTRIBUTION OF LOW LEVEL DESCRIPTORS OBTAINED FROM LPF SIGNALS AND THE ORIGINAL SIGNAL AT DIFFERENT CUTOFF FREQUENCIES TOWARDS THE PROPOSED FEATURE VECTOR

Descriptor/Feature Set	$\mathbb{F}_1^{'}$	$\mathbb{F}_2^{'}$	$\mathbb{F}_3^{'}$	$\mathbb{F}_4^{'}$	$\mathbb{F}_5^{'}$	$\mathbb{F}_6^{'}$	$\mathbb{F}_7^{'}$	$\mathbb{F}_8^{'}$	**Total**
VoiceProb	2	0	0	0	0	0	0	0	**2**
F0env	1	1	0	0	0	0	0	0	**2**
F0	0	0	0	0	1	0	0	0	**1**
logEnergy sma	0	0	0	0	0	0	0	1	**1**
ZCR	2	0	1	0	1	1	0	0	**5**
fftmag_SpectralRolloff25	1	0	0	0	0	2	0	1	**4**
fftmag_SpectralRolloff50	0	0	2	1	0	0	0	0	**3**
fftmag_SpectralRolloff75	1	0	1	0	0	0	0	0	**2**
fftmag_SpectralRolloff90	0	0	0	1	1	0	0	5	**7**
fftmag_spectralMinPos	1	0	0	0	0	0	0	2	**3**
fftmag_spectralMaxPos	1	0	3	0	1	0	0	5	**10**
fftmag_melSpec	5	3	2	0	1	0	1	3	**15**
MFCC	3	1	0	0	1	1	3	4	**13**
MFCC delta	1	1	1	1	2	1	0	0	**7**
MFCC delta delta	2	4	1	1	1	1	1	1	**12**
Total	**20**	**10**	**11**	**4**	**9**	**6**	**5**	**22**	**87**

B. Analysis

As mentioned in Section IV-B, we obtain a 87 dimension feature vector for a given audio file. The number of features contributing to the final feature vector from individual LPF speech signals and the original signal is given in Table VI. It can be seen that the features extracted from the original signal contribute the most to the final feature vector. Subsequently, features obtained from LPF signal at ω_c = 1 kHz have the second highest contribution. This is expected as low frequency contains vocal tract information which is adversely affected by a person's cognitive inability.

We analyze the contribution of different descriptors obtained from the emo-large feature set as seen in table VI. FFT Magnitude comes out as the most prominent desciptors contributing to around 50% of the feature set while, MFCCs contribute around 36% of the feature set. From the table, we can also infer that dementia affects the fundamental frequency and zero crossing rate (ZCR) of the signal, when compared to a healthy speech sample.

VI. CONCLUSION

Alzheimer's disease (AD) is the most frequent cause of dementia and the number of people affected is on the rise in an aging society. AD can easily be mistaken with senile dementia that is commonly developed throughout aging process. This makes the diagnosis of AD at an early stage, more challenging. Speech has proven to be quite useful in providing an indication of a person's cognitive state. In this work we propose an effective technique of identification of dementia by classifying it into one of the three categories - AD, MCI and HC. In addition to the traditional method of obtaining a feature set from the raw speech signal, we propose to use features by low-pass filtering the signal. Combination of descriptors from different LPF speech signals and the original signal improves the performance of dementia classification. We obtain an F-score of 0.857 which is a significant improvement over the existing dementia classification systems. The major contribution of our paper is the use of bandlimited acoustic features, which is not only an important step towards building a language independent dementia identification system but it is also seen to better discriminate the three classes in our study.

REFERENCES

[1] https://www.who.int/news-room/fact-sheets/detail/dementia, World Health Organization, September 2019.

[2] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "Mini-mental state: A practical method for grading the cognitive state of patients for the clinician," Journal of Psychiatric Research, vol. 12, no. 3, pp. 189198, 1975.

[3] K. Bettens, F. L.Wuyts, and K. M. Lierde, "Instrumental assessment of velopharyngeal function and resonance: A review," Journal of Communication Disorders, 2014.

[4] P. Klumpp, J. Fritsch and E. Noeth, "ANN-based Alzheimers disease classification from bag of words," in Speech Communication; 13th ITG-Symposium, 2018

[5] R. Sadeghian, J. D. Schaffer, and S. A. Zahorian, "Speech processing approach for diagnosing dementia in an early stage," in INTERSPEECH, 2017, pp. 27052709.

[6] G. Gosztolya, L. Tth, T. Grsz, V. Vincze, I. Hoffmann,G. Szatlczki, M. Pkski, and J. Klmn, "Detecting mild cognitive impairment from spontaneous speech by correlation-based phonetic feature selection," in INTERSPEECH, 2016, pp. 107111.

[7] J. Weiner, M. Engelbart, and T. Schultz, "Manual and automatic transcriptions in dementia detection from speech," in INTERSPEECH, 2017.

[8] B. Mirheidari, D. Blackburn, R. O'Malley, T. Walker, A. Venneri, M. Reuber, H. Christensen, "Computational cognitive assessment: investigating the use of an intelligent virtual agent for the detection of early signs of dementia," in ICASSP, 2019, pp. 2732-2736.

[9] C. Thomas, V. Keselj, N. Cercone, K. Rockwood, and E. Asp, "Automatic detection and rating of dementia of Alzheimer type through lexical analysis of spontaneous speech," in IEEE International Conference Mechatronics and Automation, 2005, vol. 3, 2005, pp. 15691574.

[10] S. Wankerl, E. Nth, and S. Evert, "An N-gram based approach to the automatic diagnosis of Alzheimers disease from spoken language," in INTERSPEECH, 2017

[11] A. Satt, R. Hoory, A. Knig, P. Aalten, and P. H. Robert, "Speech based automatic and robust detection of very early dementia," in INTERSPEECH, 2014.

[12] A. Knig, 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 Alzheimers disease," Alzheimers and Dementia: Diagnosis, Assessment and Disease Monitoring, vol. 1, no. 1, pp. 112124, 2015.

[13] J. Weiner, M. Angrick, S. Umesh, and T. Schultz, "Investigating the effect of audio duration on dementia detection using acoustic features," in INTERSPEECH, 2018, pp. 23242328.

[14] S. Al-Hameed, M. Benaissa, and H. Christensen, "Simple and robust audio-based detection of biomarkers for Alzheimers disease," in 7th SLPAT, 2016, pp. 3236.

[15] T. Gerkmann and R. C. Hendriks, "Unbiased mmse-based noise power estimation with low complexity and low tracking delay," IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, no. 4, pp. 13831393, 2012.

[16] J. T. Becker, F. Boller, O. L. Lopez, J. Saxton, and K. L. McGonigle, "The natural history of Alzheimers disease: Description of study cohort and accuracy of diagnosis," vol. Archives of Neurology, 51(6), pp. 585594, 1994.

[17] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer [computer program]. version 6.0.39," 2018, http://www.praat.org/.

[18] "opensmile," http://www.audeering.com/research/opensmile.

[19] E. Florian, W. Felix, G. Florian, and B. W. Schuller, "Recent developments in openSMILE, the Munich open-source multimedia feature extractor," in ACM Multimedia, 2013, pp. 835838.

[20] E. Frank, M. A. Hall, and I. H. Witten, "The Weka workbench. Online appendix for data mining: Practical machine learning tools and techniques," 2016, Morgan Kaufmann, Fourth Edition.

[21] C. Bhat and S.K. Kopparapu "Identification of Alzheimers Disease using Non-linguistic Audio Descriptors" in 27th European Signal Processing Conference (EUSIPCO) 2019.

[22] D. M. J. Tax and R. P. W. Duin, "Using two-class classifiers for multi-class classification," in Object recognition supported by user interaction for service robots, 2002, vol.2, pp. 124127.