# MULTI-TASK ESTIMATION OF AGE AND COGNITIVE DECLINE FROM SPEECH

*Yilin Pan*[1] *Venkata Srikanth Nallanthighal*[2,3] *Daniel Blackburn*[4]
*Heidi Christensen*[1] *Aki Härmä*[2]

[1]Department of Computer Science, University of Sheffield, UK
[2]Philips Research, Eindhoven, The Netherlands
[3]Centre for Language Studies (CLS), Radboud University Nijmegen
[4]Sheffield Institute for Translational Neuroscience (SITraN), University of Sheffield, UK
{yilin.pan@sheffield.ac.uk,aki.harma@philips.com}

## ABSTRACT

Speech is a common physiological signal that can be affected by both ageing and cognitive decline. Often the effect can be confounding, as would be the case for people at, e.g., very early stages of cognitive decline due to dementia. Despite this, the automatic predictions of age and cognitive decline based on cues found in the speech signal are generally treated as two separate tasks. In this paper, multi-task learning is applied for the joint estimation of age and the Mini-Mental Status Evaluation criteria (MMSE) commonly used to assess cognitive decline. To explore the relationship between age and MMSE, two neural network architectures are evaluated: a SincNet-based end-to-end architecture, and a system comprising of a feature extractor followed by a shallow neural network. Both are trained with single-task or multi-task targets. To compare, an SVM-based regressor is trained in a single-task setup. i-vector, x-vector and ComParE features are explored. Results are obtained on systems trained on the DementiaBank dataset and tested on an in-house dataset as well as the ADReSS dataset. The results show that both the age and MMSE estimation is improved by applying multi-task learning, with state-of-the-art results achieved on the ADReSS dataset acoustic-only task.

**Index Terms**: Multi-task learning, age estimation, cognitive decline estimation, SincNet, x-vector

## 1. INTRODUCTION

Home healthcare is becoming increasingly important in the post Covid-19 world and takes different forms ranging from conventional telehealth contact centers to autonomous home assistants based on spoken dialog systems. Spoken language can be used in home healthcare as one of the health sensing modalities, for example, for cognitive [1], mental, and respiratory conditions [2]. In a long care relationship, it is possible to collect longitudinal data, possibly over tens of years. In this process, the properties of speech of an individual may change over time due to ageing but changes may also occur as a consequence of an illness, like cognitive decline.

With ageing, the subsystems which make up the human speech production system undergo progressive physiological change affected by the decreasing rate and strength of muscle contraction [3], resulting in acoustic changes. Automatic age estimation can be regarded as either a regression (ageing is a continuous progress) or a classification task (consider each specific age or age-range as a class). For classification, the earlier studies were based on Perceptual Linear Prediction (PLP) and Mel-Frequency Cepstral Coefficients (MFCC) [4] as the input of SVM for the classification procedure. A Gaussian mixture model (GMM) based method was proposed for learning the age-specific information followed by an SVM for classification or regression [5]. The efficiency of $F\_0$ and formants, as well as the prosodic features has been demonstrated effective for the age estimation [6]. State-of-the-art approaches involve popular speaker embeddings like the i-vector [7] or x-vector [8] as the front-end features followed by a regression stage, like SVM based regression [9] or a shallow Neural Network [10–12] based regression. The result reported in [11] is 7.60 and 8.63 root mean square deviation (RMSE) for male and female respectively, as well as $4.92$ mean average error (MAE) on SRE10 [10].

Cognitive decline, associated with early signs of many neurodegenerative disorders, is caused by slow progressive loss of neurons in the central nervous system and can lead to an irreversible selective loss of brain functions [13], resulting in speech changes even decades before diagnosis. For automatic methods for acoustic-based cognitive decline detection, the performance of the typical pipeline system for cognitive decline depends on both the front-end acoustic feature and back-end detection stage. It has been found that in addition to paralinguistic acoustic feature sets [13], x-vector and i-vector are also efficient for pathological speech detection [14, 15]. In addition, inspired by the outstanding performance of deep neural networks used in numerous speech-based research areas, [16] proposed using a deep neural network for task-specific feature extractor learning and achieved superior performance.

The changes by ageing and cognitive decline exist on speech result from two independent processes but are highly correlated. [17] demonstrated that utilizing information such as age and education can improve the estimation of mini-mental status examination (MMSE), a commonly used set of questions for screening cognitive function. In addition, the previous research demonstrated that the acoustic features (e.g. speaking duration, $F\_0$, x-vector) can be utilized for diagnosing pathological speech and estimating age are similar. For example, [18] proposed that speech measures linked to Alzheimer's Disease (AD) are also associated with normal ageing.

Multi-task learning (MTL) has led to successes in many applications of machine learning, from natural language processing and speech recognition to computer vision and drug discovery [19]. The approach uses the correlation between related tasks to improve the performance of the system by learning the tasks in parallel. To achieve the benefits of addressing two individual tasks within one system, this paper addresses a novel problem of the simultaneous estimation of more than one longitudinal change in speech properties utilizing multi-task learning.

The contribution of the paper can be summarized as follows: (1). The results demonstrate that both an end-to-end based multi-task learning system and a pipeline based multi-task learning system are effective for estimating age and MMSE. (2). The analysis of the SincNet filters learned in the end-to-end system demonstrates the different information learned by the two single tasks and the multi-task system. (3). The multi-task pipeline system on the Interspeech 2020 Alzheimer's Dementia Recognition through Spontaneous Speech (ADReSS) [20] dataset achieved state-of-the-art results with acoustic-only features.

In the remainder of this paper, Section 2 presents the background. Section 3 introduces the experimental setup and the results are described in Section 4. Finally, the conclusions are given in Section 5.

## 2. BACKGROUND

To get an initial understanding of the acoustic changes caused by age and cognitive decline, a statistic analysis is carried out on the used datasets: the Trinity College Dublin Speaker Ageing (TCDSA) dataset [21], DementiaBank [22] and in-house collected dataset named *Intelligent Virtual Agent* (IVA), more details can be found in Section 3.3. Although this collective dataset contains different accents, recording environments and speech content, it can still provide some intuition for the correlations observed between the effect of age and MMSE on acoustic features.

First, the influence of age on the typical acoustic features is analyzed for people with or without cognitive decline. The features are extracted automatically using an open-source *myspsolution.praat* toolkit [23]. As shown in Figure 1, F0_median and speaking duration from people living with or without cognitive decline seems to have a weak inverse correlation trend when ageing in our data.

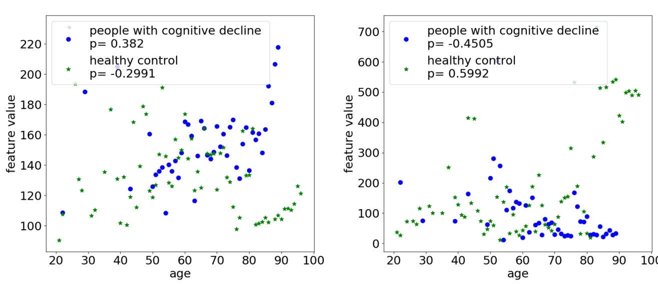
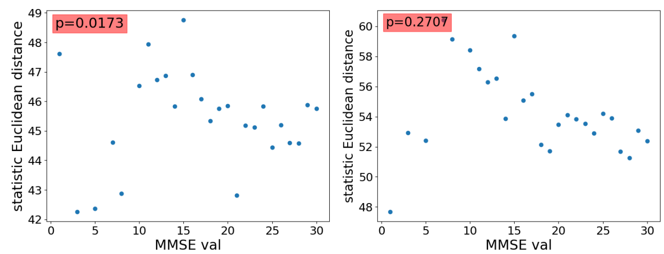



**Fig. 1**. The correlation between age and acoustic features (left: F0_median, right: speaking duration) with different cognitive status.

The relationship between age and MMSE is also explored by calculating the Euclidean distance between the x-vector of people diagnosed with different MMSE and healthy controls. To get the *anchor* (i.e., average) x-vector representing the healthy controls, only the x-vectors from the people in the TCDSA dataset not known to have any cognitive health issues are used for calculating the average representation. We hypothesise that the distance between the anchor x-vector and x-vectors averaged across speakers with a particular MMSE value is larger for lower MMSE values (indicating more severe cognitive decline cases).

In Figure 2(a), the plotted distance for each MMSE is found by averaging the distance between the anchor x-vector and the x-vectors of speakers with the corresponding MMSE. However, the values in Figure 2(a) has not taken the age into consideration. To analyse any correlation between age and MMSE values, multiple age-specific anchor x-vectors are calculated by averaging the x-vectors from people with the same age. The anchor x-vector corresponding to any missing age values is estimated by averaging the x-vectors of its neighbour ages. The average distance of the x-vectors for each MMSE value and healthy anchor age-specific x-vector is shown in Figure 2(b). By comparison, it is found that the relationship (Pearson's correlation) between healthy and people living with cognitive decline becomes stronger after taking age into consideration (increased from 0.0173 to 0.2707).




(a) age un-specific (b) age specific

**Fig. 2**. The Euclidean distance between the anchor x-vectors and x-vectors extracted from people with different MMSE values.

The relationship between the change caused by age and MMSE on speech encourages us to explore how their estimation might be improved by training a joint, single system, like a multi-task system. The mainstream architectures for age and MMSE estimation are pipeline and end-to-end systems. For the pipeline system, speaker embeddings like x-vector or i-vectors have achieved excellent performances for both age or MMSE estimation. On the other hand, the efficiency and interpretability of an end-to-end system named Sinc-CLA for neurodegenerative related disorder classification has been demostrated in [16]. In this paper, both of the two mainstream structures are demonstrated to be efficient for multi-task learning.

## 3. EXPERIMENTAL SETUP

### 3.1. End-to-end System

Previous studies [16] have shown that Sinc-CLA architecture has a good performance and interpretability in classifying recordings from people living with mild cognitive impairment, neurodegenerative disorders, or healthy controls. The multi-task Sinc-CLA system introduced in this paper is shown in Figure 3. The SincNet Layer and CNN layers are shared by the two tasks, but the bi-directional LSTM and its following layers are separately trained with a specific target (age or MMSE). The detailed description of each functional layers can be found in Section 3.4 of this paper and in [16].

### 3.2. Pipeline System

For constructing the pipeline system, the x-vector or i-vector speaker embeddings were adopted as the front-end features for age and

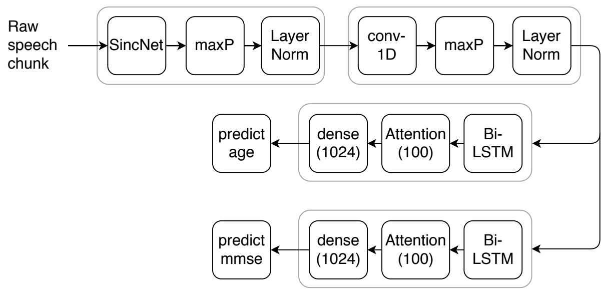



**Fig. 3**. The structure of the multi-task Sinc-CLA system for age and MMSE estimation.

MMSE estimation. To make use of the age information in the estimation of MMSE and likewise using the cognitive status to improve the age estimation, a multi-task shallow neural network comprised of two shared fully connected layers and one separated output layer were designed for the front-end feature regression.

### 3.3. Datasets

The target of our experiment is to estimate the age and MMSE for the IVA dataset with a system trained on the public available DementiaBank dataset. The IVA dataset was collected at the University of Sheffield's Department of Neurology at the Royal Hallamshire Hospital in the UK in a real clinical setting [1]. A *Digital Doctor* (or Intelligent Virtual Agent (IVA)) presented on a laptop asks a series of conversational questions and administers a series of verbal tests designed to mimic a neurologist-patient conversation. In our experiment, only the audio recordings from the participants that have the MMSE and age information are used, which consists of a total of $34$ recordings aging from 45 to 80. Further information about the data can be found in [1].

Likewise, the subset of recordings with both age and MMSE labeled were selected from DementiaBank. After selection, $459$ recordings from $286$ speakers aging from 46 to 95 were left. The $459$ recordings are separated into 5 folds [1] for cross-validation (CV) application. In 5-fold CV, $4$ folds were used for training, $1$ fold for validation (hyper-parameter optimization) and the recordings from the IVA dataset are used for testing (test set). The results presented in the paper are averaged across the $5$ results obtained from testing the test set on each of the systems corresponding to the $5$ folds.

To compare, we also applied our approach on the ADReSS dataset. To train the system, we divided the $108$ speakers in the training set into $9$ folds as in [20] and the result presented is the average across the result estimated by the 9 trained systems.

The Trinity College Dublin Speaker Ageing (TCDSA) dataset was designed primarily to investigate the effect of the ageing-related vocal change on speaker verification [21]. The main portion of this dataset contains speech recordings of $26$ adults ($15$ males and $11$ females) across a time span of between $25 - 58$ years per speaker. Among the $26$ speakers, three of them (Thatcher, Reagan and Neill) were diagnosed with mental health problems in their later years and the others are not known to have any cognitive health issues. Only the recordings from the people without any cognitive health issues were adopted for the analysis in Section 2. The detailed information about the used datasets is shown in Table 1.

[1]Partitioning of folds available on request

**Table 1**. Detailed information for datasets; HC is used to represent healthy controls, and CD is used to represent cognitive decline.

| Dataset Name | #Rec(HC vs. CD) | #Spk (HC vs. CD) | Age Range |
|---|---|---|---|
| TCDSA | 179 vs.71 | 23 vs. 3 | [19, 96] |
| ADReSS | 78 vs. 78 | 78 vs. 78 | [50, 79] |
| DB | 147 vs. 190 | 85 vs. 145 | [46, 90] |
| IVA | 5 vs. 44 | 5 vs. 40 | [26, 87] |

### 3.4. Evaluation Setting

The Kaldi Toolkit [2] is adopted for the x-vector and i-vector based speaker embedding extraction. The detailed information about the system setting can be found in [8]. To train the x-vector DNN extractor and total variability space for i-vector estimation with the Kaldi script, the combination of SRE (SRE04, SRE06 train set and SRE08) and SWBD (LDC2001S13, LDC2004S07, LDC98S75, LDC99S79 and LDC2002S06) is used [8]. In total, $141$k acoustic recordings were used. The trained DNN extractors and total variability space can map each recording in DementiaBank and IVA dataset into a $512$ dimension x-vector and a $600$ dimension i-vector respectively.

For the Sinc-CLA system, the parameter setting of each layer is the same as in [16], except for the loss function, which is the MAE in the current regression system. While training, the mini-batch size is set to $80$ and the epoch is set to $100$. For regression, both age and MMSE is normalized to the [0,1] range before estimation. To train the system, each recording is cut into multiple 2-second chunks and assigned a label corresponding to its normalized age or MMSE value. The predicted value for the test recording is the average of the estimated value of all the corresponded chunks from that recording.

The single-task shallow neural network is comprised of two fully-connected dense layers with 64 units and a 1-unit output layer. The output layer of the multi-task shallow neural network are two separate 1-unit dense layers for each regression task. All hidden layers use leaky-ReLU non-linearities. To train the system, $rmsprop$ is applied as the optimizer with a learning rate of $0.01$. While training, the batch size is set to $80$ and the epoch is set to $300$. For the two multi-task systems, the weight of age based MAE and MMSE based MAE share the same weights when added together as the loss criteria for parameter tuning.

## 4. RESULTS

### 4.1. SVM based Regression

As the baseline system, x-vector, i-vector and the ComParE statistic features (6373-dimension including energy, spectral, MFCC, and voicing related low-level descriptors (LLDs)) extracted by OpenSMILE are regressed with SVM for age or MMSE estimation. The results are shown in Table 2. In our experiment, RMSE is utilized as the criteria for comparing the performance of the baseline and proposed approaches.

Though some previous research demonstrated that x-vector can provide better or similar performance in various pathological related research tasks [14], our experimental results show that i-vector achieves a better result for MMSE estimation ($5.03$ RMSE) compared with x-vector and ComParE in Table 2.

[2]https://github.com/kaldi-asr/kaldi

**Table 2**. The results from the SVM based regression.

| Target | Feature type | RMSE |
|---|---|---|
| Age | ComParE | 5.23 |
| | i-vector | 4.99 |
| | x-vector | **4.83** |
| MMSE | ComParE | 5.34 |
| | i-vector | **5.03** |
| | x-vector | 5.36 |

### 4.2. Multi-task Learning Results

#### *4.2.1. End-to-end system*

The results of the end-to-end system trained for single-task and multi-task targets are shown in Table 3. Both the average RMSE and standard deviation RMSE over the 5 folds are shown in the Table.

**Table 3**. Result from single-task and multi-task Sinc-CLA network.

| Target | Task type | RMSE |
|---|---|---|
| Age | single | 5.17±(0.32) |
| | multi | 5.40±(0.21) |
| MMSE | single | 4.44±(0.21) |
| | multi | 4.43±(0.14) |

By comparing the results for the same task from single and multi-task systems within Table 3, it is found that multi-task learning can improve the RMSE of the MMSE estimation from $4.44$ to $4.43$, but causes a small decline in the age estimation.

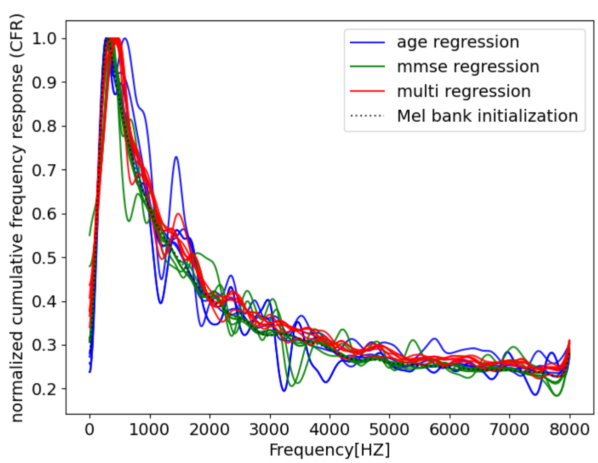



**Fig. 4**. The learned Normalized Cumulative Frequency Response from three systems.

To illustrate the critical information for age and MMSE estimation learned by the SincNet filters, the cumulative frequency responses (CFRs) of the SincNet from the three systems are plotted in Figure 4. Lines in the same colour correspond to the same regression task trained for the 5-fold CV. The black line is the initialized Mel bank CFRs. By comparing with the Mel bank CFRs, it is found the different fold-specific CFRs appear to be less variable for the multi-task regression than for the single task system.Considering the results shown in Table 3, MMSE estimation is improved at a cost of a small decrease of age estimation, how to balance the multi-task learning to improve each single task at the same time should be considered in the future.

#### *4.2.2. Pipeline System*

The results from the single-task and multi-task neural network based systems are shown in Table 4. By comparing the results from x-vectors, it is found that both the age and MMSE estimation can be improved from $4.89$ to $4.64$ (age) and from $4.50$ to $4.35$ (MMSE) when utilizing multi-task learning. By comparing with the results in Table 3 and Table 2, it is found that x-vector based multi-task learning can achieve the best performance. In addition, the result of i-vectors is consistent with our expectation that multi-task learning performs better than single-task learning, though not as well as the result from SVM based regression in Table 2. Comparing the performance of x-vector and i-vector under the same evaluation setting proves the efficiency of x-vector with the shallow neural network.

**Table 4**. The results with speaker embedding features on single-task/multi-task pipeline system estimation.

| Target | Task type | RMSE (x-vector) | RMSE (i-vector) |
|---|---|---|---|
| Age | single | 4.89±(0.08) | 5.88±(0.39) |
| | multi | **4.64**±(0.11) | 5.47±(0.29) |
| MMSE | single | 4.50±(0.09) | 8.32±(0.42) |
| | multi | **4.35**±(0.22) | 7.25±(0.37) |

#### *4.2.3. Results on ADReSS*

Next, proposed x-vector based shallow neural network was used to the ADReSS dataset MMSE estimation task. Similar to the previous results, multi-task learning can improve the estimation of age and MMSE: the RMSE values obtained on MMSE estimation is $5.85$ with multi-task learning, compared with the baseline $6.14$ shared in [24] and $5.92$ in [25] (acoustic features only).

## 5. CONCLUSIONS

This paper presented a multi-task method by utilizing both the end-to-end system and shallow neural network based pipeline system for estimating the MMSE and age for the people living with or without cognitive decline. The result from the in-house IVA dataset demonstrated that applying multi-task learning techniques using x-vectors and a Sinc-CLA architecture can achieve better results than the single-task architecture and SVR based pipeline systems. Furthermore, we also demonstrated the efficiency of x-vector based shallow neural network on the ADReSS dataset, which achieved a state-of-the-art result using acoustic features only.

## 6. ACKNOWLEDGEMENTS

This work is supported under the European Union's H2020 Marie Skłodowska-Curie programme TAPAS (Training Network for PAthological Speech processing; Grant Agreement No. 766287).

## 7. REFERENCES


[1] Bahman Mirheidari, Daniel Blackburn, Ronan O'Malley, Traci Walker, Annalena Venneri, Markus Reuber, and Heidi Christensen, "Computational cognitive assessment: Investigating the use of an intelligent virtual agent for the detection of early signs of dementia," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, United Kingdom, 2019, IEEE, pp. 2732–2736.

[2] Venkata Srikanth Nallanthighal, Aki Härmä, and Helmer Strik, "Speech breathing estimation using deep learning methods," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, 2020, IEEE, pp. 1140–1144.

[3] Finnian Kelly and Naomi Harte, "Effects of long-term ageing on speaker verification," in *European Workshop on Biometrics and Identity Management*. Springer, 2011, pp. 113–124.

[4] Davood Mahmoodi, Hossein Marvi, Mehdi Taghizadeh, Ali Soleimani, Farbod Razzazi, and Marzieh Mahmoodi, "Age estimation based on speech features and support vector machine," in *Third Computer Science and Electronic Engineering Conference (CEEC)*, Athens, Greece, 2011, IEEE, pp. 60–64.

[5] Gil Dobry, Ron M Hecht, Mireille Avigal, and Yaniv Zigel, "Supervector dimension reduction for efficient speaker age estimation based on the acoustic speech signal," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 1975–1985, 2011.

[6] Werner Spiegl, Georg Stemmer, Eva Lasarcyk, Varada Kolhatkar, Andrew Cassidy, Blaise Potard, Stephen Shum, Young Chol Song, Puyang Xu, Peter Beyerlein, et al., "Analyzing features for automatic age estimation on cross-sectional data," in *Tenth Annual Conference of the International Speech Communication Association*, Brighton, United Kingdom, 2009.

[7] Najim Dehak, Patrick J Kenny, Réda Dehak, Pierre Dumouchel, and Pierre Ouellet, "Front-end factor analysis for speaker verification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 4, pp. 788–798, 2010.

[8] David Snyder, Daniel Garcia-Romero, Gregory Sell, Daniel Povey, and Sanjeev Khudanpur, "X-vectors: Robust DNN embeddings for speaker recognition," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, Canada, 2018, IEEE, pp. 5329–5333.

[9] Seyed Omid Sadjadi, Sriram Ganapathy, and Jason W Pelecanos, "Speaker age estimation on conversational telephone speech using senone posterior based i-vectors," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, 2016, IEEE, pp. 5040–5044.

[10] Pegah Ghahremani, Phani Sankar Nidadavolu, Nanxin Chen, Jesús Villalba, Daniel Povey, Sanjeev Khudanpur, and Najim Dehak, "End-to-end deep neural network age estimation.," in *Proc. Interspeech*, Hyderabad, India, 2018, pp. 277–281.

[11] Shareef Babu Kalluri, Deepu Vijayasenan, and Sriram Ganapathy, "A deep neural network based end to end model for joint height and age estimation from short duration speech," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, United Kingdom, 2019, IEEE, pp. 6580–6584.

[12] Anna Fedorova, Ondřej Glembek, Tomi Kinnunen, and Pavel Matějka, "Exploring ann back-ends for i-vector based speaker age estimation," in *Sixteenth Annual Conference of the International Speech Communication Association*, Dresden, Germany, 2015.

[13] Fasih Haider, Sofia De La Fuente, and Saturnino Luz, "An assessment of paralinguistic acoustic features for detection of Alzheimer's dementia in spontaneous speech," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 2, pp. 272–281, 2019.

[14] Sebastião Quintas, Julie Mauclair, Virginie Woisard, and Julien Pinquier, "Automatic Prediction of Speech Intelligibility based on X-vectors in the context of Head and Neck Cancer," in *Proc. Interspeech*, Shanghai, China, 2020.

[15] José Vicente Egas López, László Tóth, Ildikó Hoffmann, János Kálmán, Magdolna Pákáski, and Gábor Gosztolya, "Assessing Alzheimer's disease from speech using the i-vector approach," in *International Conference on Speech and Computer*. Springer, 2019, pp. 289–298.

[16] Yilin Pan, Bahman Mirheidari, Zehai Tu, Ronan O'Malley, Traci Walker, Annalena Venneri, Markus Reuber, Daniel Blackburn, and Heidi Christensen, "Acoustic feature extraction with interpretable deep neural network for neurodegenerative related disorder classification," *Proc. Interspeech*, 2020.

[17] Ziyang Fu, Fasih Haider, and Saturnino Luz, "Predicting minimental status examination scores through paralinguistic acoustic features of spontaneous speech," in *42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 5548–5552.

[18] Juan José G Meilán, Francisco Martínez-Sánchez, Juan Carro, Dolores E López, Lymarie Millian-Morell, and José M Arana, "Speech in Alzheimer's disease: Can temporal and acoustic parameters discriminate dementia?," *Dementia and Geriatric Cognitive Disorders*, vol. 37, no. 5-6, pp. 327–334, 2014.

[19] Sebastian Ruder, "An overview of multi-task learning in deep neural networks," *arXiv preprint arXiv:1706.05098*, 2017.

[20] Saturnino Luz, Fasih Haider, Sofia de la Fuente, Davida Fromm, and Brian MacWhinney, "Alzheimer's Dementia Recognition through Spontaneous Speech: The ADReSS Challenge," in *Proc. Interspeech*, Shanghai, China, 2020.

[21] Finnian Kelly, Rahim Saeidi, Naomi Harte, and David A van Leeuwen, "Effect of long-term ageing on i-vector speaker verification," in *Proc. Interspeech*, Singapore, 2014.

[22] James T Becker, François Boiler, Oscar L Lopez, Judith Saxton, and Karen L McGonigle, "The natural history of Alzheimer's disease: description of study cohort and accuracy of diagnosis," *Archives of Neurology*, vol. 51, no. 6, pp. 585–594, 1994.

[23] Yannick Jadoul, Bill Thompson, and Bart De Boer, "Introducing parselmouth: A python interface to praat," *Journal of Phonetics*, vol. 71, pp. 1–15, 2018.

[24] Saturnino Luz, Fasih Haider, Sofia de la Fuente, Davida Fromm, and Brian MacWhinney, "Alzheimer's dementia recognition through spontaneous speech: The ADReSS challenge," *arXiv preprint arXiv:2004.06833*, 2020.

[25] Muhammad Shehram Shah Syed, Zafi Sherhan Syed, Margaret Lech, and Elena Pirogova, "Automated screening for Alzheimer's dementia through spontaneous speech," .