DETECTING ALZHEIMER'S DISEASE BY ESTIMATING ATTENTION AND ELICITATION PATH THROUGH THE ALIGNMENT OF SPOKEN PICTURE DESCRIPTIONS WITH THE PICTURE PROMPT

A PREPRINT

Bahman Mirheidari
Department of Computer Science
University of Sheffield
United Kingdom
b.mirheidari@sheffield.ac.uk

Yilin Pan
Department of Computer Science
University of Sheffield
United Kingdom
yilin.pan@sheffield.ac.uk

Traci Walker
Department of Human Communication Sciences
University of Sheffield
UK
traci.walker@sheffield.ac.uk

Markus Reuber
Academic Neurology Unit, Royal Hallamshire Hospital
University of Sheffield
United Kingdom
m.reuber@sheffield.ac.uk

Annalena Venneri
Sheffield Institute for Translational Neuroscience
University of Sheffield
United Kingdom
a.Venneri@sheffield.ac.uk

Daniel Blackburn, Annalena Venneri
Sheffield Institute for Translational Neuroscience
University of Sheffield
United Kingdom
d.blackburn@sheffield.ac.uk

Heidi Christensen
Department of Computer Science
University of Sheffield
United Kingdom
heidi.christensen@sheffield.ac.uk

October 2, 2019

ABSTRACT

Cognitive decline is a sign of Alzheimer's disease (AD), and there is evidence that tracking a person's eye movement, using eye tracking devices, can be used for the automatic identification of early signs of cognitive decline. However, such devices are expensive and may not be easy-to-use for people with cognitive problems. In this paper, we present a new way of capturing similar visual features, by using the speech of people describing the Cookie Theft picture - a common cognitive testing task - to identify regions in the picture prompt that will have caught the speaker's attention and elicited their speech. After aligning the automatically recognised words with different regions of the picture prompt, we extract information inspired by eye tracking metrics such as coordinates of the area of interests (AOI)s, time spent in AOI, time to reach the AOI, and the number of AOI visits. Using the DementiaBank dataset we train a binary classifier (AD vs. healthy control) using 10-fold cross-validation and achieve an 80% F1-score using the timing information from the forced alignments of the automatic speech recogniser (ASR); this achieved around 72% using the timing information from the ASR outputs.

***K*eywords** clinical applications of speech technology · pathological speech · dementia detection

1 Introduction

The number of people living with dementia has increased significantly in recent years and the economic impact on the society is huge. According to [1] every 3 seconds a person develops dementia in the world. There are around 50 million people living with dementia in total and it is estimated to rise to 152 million by 2050; the current cost of dementia is approximated about a trillion US dollar a year.

The treatments are most effective in the early stage of the disease before dementia has developed . However, the process of diagnosing this disorder is very complex, mostly due to overlapping symptoms with normal ageing and low accuracy of existing cognitive screening tools. The currently available tests for stratifying (screening) people with cognitive complaints, are based on pen-and-paper testing and lack sensitivity or specificity especially early in the disease process. It is therefore highly desirable to build a cheap and reliable automatic screening tool to identify people at risk of developing dementia. Ideally, such a tool could be used in a person's own home, without the need for an examiner to be present.

Speech and language are affected early in dementia (e.g. by losing vocabulary, simplifying syntax/semantics, overusing semantically empty [2, 3, 4]). The Boston Cookie Theft picture description task [5] is a widely used cognitive test which was originally designed to capture decline in spontaneous speech and language of people with aphasia. However, it has been shown that it could help in identify people with dementia as well [6, 7]. In this task, the subjects are asked to look at a picture (the *prompt*) and describe everything happening in the picture.

These tests are then normally scored based on the utterances that are generated in the description, and by counting things like the number of simple/complex and complete/incomplete sentences. However, there is no part of the assessment scoring that takes into account the picture prompt itself, i.e., things like which areas of the picture the person might have been looking at, in which order. However, it is known that cognitive decline may affect a person's eye movement and attention [8]. Since the generated language is a direct result of the person looking at the the picture prompt, tracking the attention and elicitation path on the picture might reveal informative data to help identify early signs of dementia. Accurate tracking of eye movement requires an eye tracking devices. However, routinely using eye tracking devices in memory clinics or for home-based testing would be an impractical, expensive and inconvenient solution. This paper proposes an approach to capturing information about eye movement path on a picture prompt without needing eye tracking equipment.

The approach uses the automatic transcription of the picture description (using automatic speech recognition (ASR)) to identify the area of interests (AOIs) in the picture prompt, that may have caught the attention of the individual and elicited their speech at that particular point in time. A number of features will be used to support the estimation. The extracted features are then used to train a classifier to identify dementia and compared with using features designed to capture information directly from the speech itself.

In the remainder of the paper, Section 2 presents the background, Section 3 presents our proposed approach, Section 4 describes the experimental setup, and finally, results and conclusions are presented in Sections 5 and 6.

2 Background

Eye tracking technology has been used in a variety of applications ranging from learning assistants [8, 9], Human Computer Interface (HCI) design [10, 11], mobile phone applications [12], assistive technology for disabled people [13], and diagnosing mental/memory diseases (autism, Mild Cognitive Decline (MCI), Parkinson's and Alzheimer's disease (AD), etc.)[14].

According to a survey by [15], eye movement difficulties might be a sign of cognitive decline and altered eye movements indicate visual-spatial and executive function problems, although same behaviour can be seen in aged subjects as well. Eye tracking can be done in a scenario based (e.g. reading) or exploration task or a more naturalistic situation (e.g identifying objects in daily life). In scenario based tasks, the AD participants exhibited longer fixation duration than the Healthy Controls (HCs), and in naturalistic tasks, the AD patients had fewer overall fixations as well as a lower proportion of relevant fixations. Based on a study by [16] eye tracking devices could assess cognitive functions of people with AD. The AD subjects showed a longer pro-saccade[1] latency and more anti-saccade[2] error rates compared to the healthy controls (HC). In a longitudinal eye tracking study on AD subjects vs. HCs, [17] found out that the AD individuals had a slower saccade reaction time compared to the HC group, however, over a 12 month period, the reaction time for both groups did not deteriorate.

[1] in a pro-saccade task, eyes first focus on a dot in the centre of a screen and then turn the gaze to a target stimulus as it appears.

[2] in an anti-saccade task, eyes first focus on a dot in the centre of a screen, but they will be asked to turn their gaze in the opposite direction of a stimulus.

The 'Cookie Theft' picture description task was originally part of the Boston Diagnostic Aphasia Examination [18] in which the examiner shows the Cookie Theft line drawing to individuals asking them to describe everything that is happening in the picture. The Pitt Corpus - DementiaBank [19] is a collection of audio recordings and transcripts of 104 HC, 208 AD and 85 other diagnosis participants describing the Cookie Theft picture [3]. The study was carried out between 1983 and 1988 and many of the participants had several repeated sessions during the study, providing a longitudinal corpus. A few years after gathering the data, the diagnoses of the subjects were reviewed again and the final decisions were made by the specialist neurologists. The corpus is free to download from their website and use for research purposes. In recent years, there has been a number of studies based on it, mostly aimed at detecting dementia using the audio, text and language processing technologies.

In [20], the corpus was used to extract a wide range of features (over 477 lexico-syntactic, acoustic, and semantic) and train an AD/HC classifier. They achieved an 81% classification accuracy using all features and 92% when they selected 40 informative features from the their feature set. [21] used GloVe word vectors (representing the meaning of words) to make 10 common clusters of the words (only nouns and verbs) in the training set of the dementia group as well as 10 common cluster of words in the HC (each cluster representing a topic or related words, e.g. C0: window, floor, curtains, plate, kitchen, D0: cookie, cookies, cake, baking, apples). Then, using the average scaled distance between the words of a given transcript from the test set and the created clusters, they extracted 20 semantic features. In addition they calculated the 'idea density' as the number of topics mentioned in the transcript divided by the number of words, and the 'idea efficiency' as the number of topics mentioned in the transcript divided by the length of recording. They gained 80% accuracy and 80% F1 score when they combined all their features with the lexicosyntactic and acoustic features from [22]. [23] achieved a range of accuracies between 83% and 92% for a number of different classifiers trained on 263 acoustic-only features extracted from the audio files of DementiaBank, thereby avoiding the need to use ASR. Using the 20 most informative features their best classifier (BayesNet) achieved 95%. In their follow-up research they included the MCI group for classification [24] and gained an accuracy rate between 89.2% and 92.4% for the pairwise classifications. [25] chose 25 HC transcriptions to make reference Information Content Units (ICUs) (representing the main information in the picture including the actors, objects, actions and places e.g. woman/lady/mother/mommy make up a group named mother, and little/young-boy/kid make another group (boy). These reference ICUs were then used to estimate the coverage of informativeness of a description, i.e., how much did the main ICUs cover. They extracted the coverage and a number of linguistic features to train an SVM classifier. The binary classifier achieved an 81% F1 score (excluding the 25 reference HCs and the subjects with MCI). Note that all of these studies used the manual transcriptions and did not used the ASRs.

In addition to the conventional classifiers, a number of authors have tried using Deep Neural Network (DNN) based classifiers to detect dementia from the DementiaBank corpus. In [26] we used the predefined GloVe word vectors to make a sequence classification, combining Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM). The classifier achieved a 75.6% accuracy, however, replacing the manual utterances with the automatic transcripts produced by the ASR (45% WER), the accuracy dropped to 62.3%. [27] trained three different DNN models to classify dementia: LSTM, 2 Dimensional CNN, and a mix of LSTM and CNN (CNN-LSTM). They used the manual transcriptions and the part of speech (POS) tags provided for the words in the utterances. Using only the words in the transcriptions, CNN-LSTM model could outperform the two other models with an 84.9% accuracy. However, the best performance achieved a 91.1% accuracy when they used the POS tagged information to trained the CNN-LSTM model.

3 Tracking attention on picture prompt

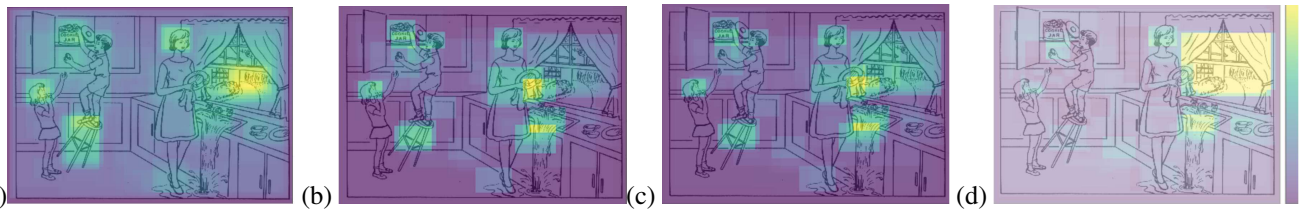
Unlike approaches described in the previous section, this paper presents a way to make use of the output of the ASR to estimate what people might have been looking at (their AOIs or attention) as they are describing the Cookie Theft picture. First, a number of reference AOIs were identified manually on the picture, corresponding to the important actors, objects and actions (nouns and verbs), e.g., boy, girl, mother, cookie, grab, fall, and wash. As we divide data into train and test sets, this information should be filtered based on the train set data, i.e., the words not included in the train set were removed from the reference AOIs. Thus the filtered reference AOIs indicates the location and size of the AOI for a given word. After training an ASR on the training set data, we can drive the timing information for each word in an utterance, i.e., the forced alignments give the estimated start and end times for the words in the training set, and the output of the ASR (decoding latices) provides the timing information for the test set words in a similar way.

Eye tracking features are often related to *saccades*, *fixations* and *gaze points* where gaze points are the points on a screen that our eyes directly stare at, fixations are the clusters of continuous and close gaze points (the pauses of eyes movements) over AOI on a screen, and saccades are rapid eye movements between fixations or the transitions from one

[3] the latest figures taken from their website:
https://dementia.talkbank.org/access/English/Pitt.html

(a) (b)

Figure 1: Scanpath for the patient with AD 005-2 (a) vs. the health control 631-0 (b).



(a) (b) (c) (d)

Figure 2: Heatmaps of the area of interests (AOIs) (a) manually assinged as the references, (b) produced automatically for the AD group, (c) produced automatically for the HC group, the difference between the AD and HC groups.

gaze point to another [28, 29]. Fixation can be represented by the coordinates (x,y) and duration (time spent) normally in milliseconds. Scanpath is a way to visualise attention tracking, in which fixation is represented by a circle with a radius of the duration of fixation and a saccade is shown by a line connecting two fixations [29]. Figure 1 shows the scanpaths for the descriptions given by an individual with AD ((a), participant 005-2) and a healthy control ((b), participant 631-0). Participant 005-2 said: "OK, He's **falling** off a **chair**. She's uh **running** the **water** over. Can't see anything else. No, OK, She's No.". The important words (in bold) are identified by circles on the picture and the arrows show the transition from one AOI to another, representing the order of attention as well. Therefore, looking at the picture we can estimate which areas of the pictures were viewed by the participant over time. The utterance of the healthy control 631-0 is:"The **kids** are in the **cookies**. The **stool** is **falling** over, The **mother**'s **spilling** the **water** and also **drying** the **dishes** and the **wind** might be **blowing** the **curtains**. And, the **water**'s **running**. Uh I can't tell that's tell is anything going on **outside** or not. I guess that's all I see that's not very many". As can be seen, from the estimated scanpaths, the HC participant gave more details of the picture, visiting more reference AOIs (note that the AOI "Water" was visited twice by the participant).

We hypothesise that the people with cognitive decline might not able to identify some of the important AOIs in the picture. Heatmaps [4] represent the viewing behaviour and distribution of attention for the individuals watching a scene on a screen [30]. Figure 2 shows (a) the heatmaps of the AOIs which manually assigned AOIs on the picture, (b) the heatmaps produced automatically for the AD participants using the ASR and the reference AOIs, (c) the automatic heatmaps for the HC groups, and the differences between the two groups. The heatmaps show that the AD groups, similarly to the HC group, could also identify many important AOIs, however, the differences are mostly on the part of the picture which shows a window and outside view including a path, trees and neighbours houses. Thus there could be some AOIs which were overlooked by the AD participants and did not catch their attention.

4 Experimental setup

Data: DementiaBank contains 551 transcriptions/audio files from 98 HC (241 recordings) and 195 people with dementia (310 recordings) (including AD, other types of dementia and MCI). However, the diagnosis of a number of the participants were changed over the longitudinal study, i.e., some changed from MCI to AD, some from HC to MCI. For this evaluation, we have chosen to exclude those participants. This resulted in us using a total of 257 participants with stable diagnoses for the HC/AD classification experiments (89 HC participants - 215 recordings; 168

[4] two dimensional graphical representation of data using colours, normally the warmer the colour (e.g. yellow warmer than blue) the higher the value.

AD participants - 249 recordings) [5]. For training the ASR, all of the 551 recordings were used (diagnostic class is not relevant) in a 10-fold cross-validation approach. Additionally, we supplemented with two data sets recorded in-house: a collection of interviews with doctors (more than 64 hours of speech) [26], and a much smaller collection of Cookie Theft description collected using an Intelligent Virtual Agent (IVA - 'digital doctor') [32]. We refer to this data set as IVA and DementiaBank as Dem for the remaining parts of the paper. Note that only 33 out of 76 IVA participants were diagnosed as AD or HC (17 vs. 16). Table 1 gives more details of the data sets.

Table 1: Information about the datasets used for training the ASRs, including Len.:the total length in hours/mins, Utts.:number of utterances, Spks.:number of speakers, and Avg. Utts.:Average utterance length in seconds.

Dataset(No)	Len.	Utts.	Spks.	Avg Utts.
Dem (551)	8h 34m	6737	293	4.6s
Dr intvws (295)	64h 21m	39184	736	5.9s
IVA (76)	1h 15m	497	76	9.11s

Features: For each AOI in the picture, we extract a number of features inspired by the eye tracking studies: the x and y coordinates of the centre point and the radius of the AOI, time spent in the AOI (estimated using length of the uttered word), time to approach the AOI (estimated using word start time), number of visits to the AOI, transition time (similar to saccade time) and total length of the pauses (silence) made up to the current uttered word. Table 2 summarises the extracted features. In addition to the AOI features, two other feature types were extracted, Age of Acquisition (AoA, the average age we normally learn a word for the first time [33]) and word vector features GloVe [34]. We include the mean and standard deviation of AoA. The predefined GloVe word vectors are in 300 dimensions, however, using the Principal Component Analysis (PCA), the dimensions of the vectors were reduced to 7. Table reftab:features summarises all of the features. For each feature in the table the mean, standard deviation, minimum and maximum were calculated.

5 Results

The Kaldi ASR [35] toolkit was used for training the ASRs. We followed the GMM/HMM based (SAT training) and the hybrid GMM/DNN (TDDN-LSTM) recipes. For the language models, we trained in-domain 3/4 grams with the KN/Turing smoothing. The HMM/GMM SAT training for the Dem dataset resulted in an average 64.2% WER, while using the LSTM-TDNN recipe the average **41.6%** WER was achieved. For the IVA dataset a 33.8% WER was gained by the DNN based ASR (50.5% using SAT). For classifications, a Logistic Regression (LR) classifier was used using a 10-fold speaker independent cross-validation approach (none of the recordings of the speakers in the testing set are seen in the training set).

[5] Note that [31] worked on the same dataset, but with 222 HC and 257 AD recordings.

Table 2: Extracted features: AOI: Area of interest. AoA: Age of Acquisition. WV: Word vectors

Feature	Type	Description
$x\_coordinate$	AOI	x coordinate of the AOI
$y\_coordinate$	AOI	y coordinate of the AOI
$radius$	AOI	radius of AOI circle
$time\_spent$	AOI	time spent on the AOI
$time\_to\_approach$	AOI	time to approach the AOI
$number\_of\_visits$	AOI	number of visits to the AOI
$transition\_time$	AOI	transition time from previous to current AOI
$pause\_length$	AOI	the total pause length up to the current AOI
$mean\_aoa$	AoA	mean of the AoA for a word
std_aoa	AoA	standard deviation of the AoA for a word
$wv_1, ..., wv_7$	WV	GloVe word vector representation

Table 3: LR classification results using forced alignment. All: all features; Mix: Dem+IVA; Ac: Accuracy; Rc: Recall; Pr: Precision; F1: F1-measure;

Train/Test	Feature	Ac%	Rc%	Pr%	F1%
Dem/Dem	WV	71.9	72.7	71.4	71.0
Dem/Dem	AoA	63.2	63.2	62.8	62.0
Dem/Dem	AOI	77.3	78.1	76.8	76.5
Dem/Dem	AOI+AoA	78.8	79.2	78.7	77.9
Dem/Dem	AOI+WV	79.2	79.7	78.7	78.3
Dem/Dem	All	**80.8**	**81.1**	**80.1**	**79.9**
IVA/IVA	All	65.8	62.5(	63.3	60.2
Mix/Dem	All	79.6	80.2	79.1	78.8
Mix/IVA	All	65.8	62.5	52.5	54.3

Table 4: LR classification results using the outputs of the ASRs and all features. Mix: Dem+IVA; Ac: Accuracy; Rc: Recall; Pr: Precision; F1: F1-measure;

Train/Test	Ac%	Rc%	Pr%	F1%
Dem/Dem	**73.1**	**73.4**	**72.4**	**72.2**
IVA/IVA	62.5	62.5	60.0	56.8
Mix/Dem	72.3	72.0	71.4	70.9
Mix/IVA	70.0	70.0	65.0	63.5

5.1 Classification results using the forced alignments

Table 3 shows the accuracy (Ac), recall (Rc), precision (Pr) and F1-score of the classifiers when using the forced alignment timing information. For the Dem dataset using only WV, only AoA and only AOI, 71%, 62% and 76.5% F1-scores were achieved respectively; on their own the picture prompt based AOI features were the most discriminative for the AD/HC classification. As we added other feature types, the performance of the classifier improved, and the best performance was achieved by combining all features, resulting in an F1-score around 80%. For the IVA dataset, using all the features, a 60.2% F1-score was gained. Training the classifiers on a combination of the two datasets (the last two rows) not only did not improve the performance, but somewhat decreased the performance especially for IVA dataset.

5.2 Classification results using the ASR outputs

Table 4 shows the classification performance using all features extracted from the outputs of the ASRs (the words, the start time and the length of the words). As can be expected when comparing to Table 3, the performance of the classifiers decreases, but not considerably. The F1-measure for Dem/Dem decreased around 8% from around 80% to 72% , for IVA/IVA around 3% from 60% to 57%. Mixing datasets, however, improved the performance for IVA from 54% to around 64%, but deteriorated for Dem (from 79% to 71%).

6 Conclusions

This paper presented an approach to aligning verbal picture descriptions with their picture prompt in order to estimate the attention and elicitation paths on the picture. Features, inspired by studies on the use of eye tracking devices for detecting cognitive decline in people with Alzheimer's disease, are then extracted. These features are then used to train a classifier achieving an 80% F1-score when using information produced from the forced alignment of ASR and 72% F1-score when we directly use the output of the ASR. In future work, we plan to investigate the use of additional features to improve the classification and apply a DNN based sequence classifier.

7 Acknowledgements

This work is supported by the MRC Confidence in Concept Scheme and the European Union's H2020 Marie Skłodowska-Curie programme TAPAS (Training Network for PAthological Speech processing; Grant Agreement No. 766287).

References

[1] C Patterson. World Alzheimer report 2018: The state of the art of dementia research: New frontiers. *Alzheimer's Disease International (ADI): London, UK*, 2018.

[2] Julian Appell, Andrew Kertesz, and Michael Fisman. A study of language functioning in Alzheimer patients. *Brain and Language*, 17(1):73–91, 1982.

[3] K A Bayles and A W Kaszniak. *Communication and cognition in normal aging and dementia*. Taylor & Francis Ltd London, 1987.

[4] H E Hamilton. *Conversations with an Alzheimer's patient: An interactional sociolinguistic study*. Cambridge, England: Cambridge University Press, 1994.

[5] Harold Goodglass and Edith Kaplan. *The assessment of aphasia and related disorders*. Lea & Febiger, 1972.

[6] Elaine Giles, Karalyn Patterson, and John R Hodges. Performance on the boston cookie theft picture description task in patients with early dementia of the alzheimer's type: missing information. *Aphasiology*, 10(4):395–408, 1996.

[7] Katrina E Forbes-McKay and Annalena Venneri. Detecting subtle spontaneous language decline in early Alzheimer's disease with a picture description task. *Neurological sciences*, 26(4):243–254, 2005.

[8] Meng-Lung Lai, Meng-Jung Tsai, Fang-Ying Yang, Chung-Yuan Hsu, Tzu-Chien Liu, Silvia Wen-Yu Lee, Min-Hsien Lee, Guo-Li Chiou, Jyh-Chong Liang, and Chin-Chung Tsai. A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational research review*, 10:90–115, 2013.

[9] Tamara van Gog and Halszka Jarodzka. Eye tracking as a tool to study and enhance cognitive and metacognitive processes in computer-based learning environments. In *International handbook of metacognition and learning technologies*, pages 143–156. Springer, 2013.

[10] Sushil Chandra, Greeshma Sharma, Saloni Malhotra, Devendra Jha, and Alok Prakash Mittal. Eye tracking based human computer interaction: Applications and their uses. In *2015 International Conference on Man and Machine Interfacing (MAMI)*, pages 1–5. IEEE, 2015.

[11] Asier Lopez-Basterretxea, Amaia Mendez-Zorrilla, and Begoña Garcia-Zapirain. Eye/head tracking technology to improve hci with ipad applications. *Sensors*, 15(2):2244–2264, 2015.

[12] Kyle Krafka, Aditya Khosla, Petr Kellnhofer, Harini Kannan, Suchendra Bhandarkar, Wojciech Matusik, and Antonio Torralba. Eye tracking for everyone. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2176–2184, 2016.

[13] Vidas Raudonis, Rimvydas Simutis, and Gintautas Narvydas. Discrete eye tracking for medical applications. In *2009 2nd International Symposium on Applied Sciences in Biomedical and Communication Technologies*, pages 1–6. IEEE, 2009.

[14] Katarzyna Harezlak and Pawel Kasprowski. Application of eye tracking in medicine: a survey, research issues and challenges. *Computerized Medical Imaging and Graphics*, 65:176–190, 2018.

[15] Jessica Beltrán, Mireya S García-Vázquez, Jenny Benois-Pineau, Luis Miguel Gutierrez-Robledo, and Jean-François Dartigues. Computational techniques for eye movements analysis towards supporting early diagnosis of alzheimer's disease: a review. *Computational and mathematical methods in medicine*, 2018, 2018.

[16] Naomi Kahana Levy, Michal Lavidor, and Eli Vakil. Prosaccade and antisaccade paradigms in persons with alzheimer's disease: a meta-analytic review. *Neuropsychology review*, 28(1):16–31, 2018.

[17] Trevor J Crawford. The disengagement of visual attention in alzheimer's disease: a longitudinal eye-tracking study. *Frontiers in aging neuroscience*, 7:118, 2015.

[18] Harold Goodglass and Edith Kaplan. *Boston diagnostic aphasia examination booklet*. Lea & Febiger, 1983.

[19] 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*, 51(6):585–594, 1994.

[20] M Yancheva, K Fraser, and F Rudzicz. Using linguistic features longitudinally to predict clinical scores for Alzheimer's disease and related dementias. *6th Workshop on Speech and Language Processing for Assistive Technologies*, 2015.

[21] Maria Yancheva and Frank Rudzicz. Vector-space topic models for detecting alzheimer's disease. In *Proc 54th Annual Meeting of the Association for Computational Linguistics*, volume 1, pages 2337–2346, 2016.

[22] Kathleen C Fraser and Graeme Hirst. Detecting semantic changes in alzheimer's disease with vector space models. In *Proc Processing of Linguistic and Extra-Linguistic Data from People with Various Forms of Cognitive/Psychiatric Impairments*. Linköping University Electronic Press, 2016.

[23] Sabah Al-Hameed, Mohammed Benaissa, and Heidi Christensen. Simple and robust audio-based detection of biomarkers for alzheimer's disease. *7th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT)*, pages 32–36, 2016.

[24] Sabah Al-Hameed, Mohammed Benaissa, and Heidi Christensen. Detecting and predicting alzheimer's disease severity in longitudinal acoustic data. In *Proc International Conference on Bioinformatics Research and Applications*, pages 57–61. ACM, 2017.

[25] Laura Hernández-Domínguez, Sylvie Ratté, Gerardo Sierra-Martínez, and Andrés Roche-Bergua. Computer-based evaluation of Alzheimer's disease and mild cognitive impairment patients during a picture description task. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 10:260–268, 2018.

[26] Bahman Mirheidari, Daniel Blackburn, Traci Walker, Annalena Venneri, Markus Reuber, and Heidi Christensen. Detecting signs of dementia using word vector representations. *Proc. Interspeech 2018*, pages 1893–1897, 2018.

[27] Sweta Karlekar, Tong Niu, and Mohit Bansal. Detecting linguistic characteristics of alzheimer's dementia by interpreting neural models. *arXiv preprint arXiv:1804.06440*, 2018.

[28] Dario D Salvucci and Joseph H Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 71–78. ACM, 2000.

[29] Tanja Blascheck, Kuno Kurzhals, Michael Raschke, Michael Burch, Daniel Weiskopf, and Thomas Ertl. State-of-the-art of visualization for eye tracking data. In *Proceedings of EuroVis*, volume 2014, 2014.

[30] Agnieszka Aga Bojko. Informative or misleading? heatmaps deconstructed. In *International conference on human-computer interaction*, pages 30–39. Springer, 2009.

[31] Yilin Pan, Bahman Mirheidari, Markus Reuber, Annalena Venneri, Daniel Blackburn, and Heidi Christensen. Automatic hierarchical attention neural network for detecting alzheimer's disease. unpublished, 2019.

[32] Bahman Mirheidari, Daniel Blackburn, Kirsty Harkness, Annalena Venneri, Markus Reuber, Traci Walker, and Heidi Christensen. An avatar-based system for identifying individuals likely to develop dementia. In *Proc INTERSPEEH*. ISCA, 2017.

[33] Katrina E Forbes-McKay, Andrew W Ellis, Michael F Shanks, and Annalena Venneri. The age of acquisition of words produced in a semantic fluency task can reliably differentiate normal from pathological age related cognitive decline. *Neuropsychologia*, 43(11):1625–1632, 2005.

[34] Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proc EMNLP*, pages 1532–1543, 2014.

[35] Daniel Povey, Arnab Ghoshal, Gilles Boulianne, Lukas Burget, Ondrej Glembek, Nagendra Goel, Mirko Hannemann, Petr Motlicek, Yanmin Qian, Petr Schwarz, Jan Silovsky, Georg Stemmer, and Karel Vesely. The kaldi speech recognition toolkit. In *Proc IEEE Workshop on Automatic Speech Recognition and Understanding*. IEEE, 2011.